UNIVERSITÀ DI TRENTO

Doctoral School in
Agrifood and Environmental Sciences

Daniele Andreis

Hyper-resolution monitoring and modelling of the water budget for improving water management

Supervisors
Prof. Riccardo Rigon
Prof. Giuseppe Formetta

37 Cycle, 2026

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisors, Prof. *Riccardo Rigon* and Prof. *Giuseppe Formetta*, for their guidance, availability, and constructive feedback throughout this PhD work. I also thank *Marialaura Bancheri* and *Andrea Antonello* for the collaboration and valuable discussions that contributed to the research presented in this thesis.

I thank all those who interacted with this work in professional contexts, and I am grateful to the GEOframe group for insightful exchanges and technical support.

Finally, I would like to thank my family and my partner for their patience, encouragement, and constant support during this journey.

Statement of contributions

The three studies presented in this thesis were led by the author and developed in collaboration with other researchers. Study 1 was conducted in collaboration with G. Formetta, M. Bancheri, and R. Rigon; Study 2 with G. Formetta and R. Rigon; and Study 3 with A. Antonello, G. Formetta, and R. Rigon. Authorship and author order will be formally agreed upon when the corresponding journal manuscripts are prepared.

Declaration on the use of generative AI

Generative artificial intelligence tools were used for language editing, including the identification of grammatical and typographical errors and suggestions to improve the clarity and readability of the text. All suggested revisions were critically reviewed by the author. The scientific content, methodological choices, analyses, interpretations, and conclusions remain the author's own, and the author assumes full responsibility for the content of the thesis.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

LIST OF TABLES ix

LIST OF FIGURES xi

LIST OF ACRONYMS xv

SUMMARY xvii

CHAPTER 1. Introduction 1
1.1 Climate change in the Alps 4
1.2 Hydrological modeling: selection, evaluation, and scale-related issues 6
1.2.1 Multi-source observations for model evaluation 9
1.3 Scale and spatial discretization in hydrological modeling 10
1.3.1 Hyper-resolution scale 13
1.3.2 Meteorological forcing and spatial variability in semi-distributed models 14
1.4 Objectives and structure of the thesis 15
1.4.1 Methodological approach and thesis structure 17
1.4.2 Experimental strategy 18
1.4.3 Modeling framework and software contributions 19
1.4.4 Model structure 22
1.4.5 Thesis outline 23

CHAPTER 2. The Noce River basin: study area, monitoring network, and data 25
2.1 Study area: the Noce River basin 25
2.1.1 The upper basin: the Malé and Vermiglio sub-catchments 28
2.2 Monitoring network and observational data 29
2.3 Simulation periods 31

CHAPTER 3. Methodological framework 33
3.1 Partition of the basin into Geomorphic Units 34
3.1.1 Constant-threshold delineations 35
3.1.2 Mixed-resolution (hybrid) delineation 36
3.2 Meteorological forcing at the Geomorphic Unit scale 36
3.2.1 Geostatistical interpolation and automatic semivariogram estimation 37
3.2.2 Validation of the interpolated fields 39

3.3 The GEOframe semi-distributed hydrological model 41
3.4 Calibration and validation protocol 44
3.5 Uncertainty estimation 46
3.6 Multi-source evaluation of the water budget 46
3.6.1 Snow compartment 47
3.6.2 Root-zone water content: Cosmic-Ray Neutron Sensing 48
3.6.3 Root-zone and surface soil moisture: satellite and reanalysis products 50
3.6.4 Spatial diagnostics of snow, root-zone, and groundwater storage 52

CHAPTER 4. Software development 53
4.1 The GEOframe modeling framework 54
4.2 Mixed-resolution input builder 54
4.3 The kriging component 56
4.4 The rainfall-runoff model and its numerical core 59
4.5 Python tooling for data preparation and diagnostics 60
4.6 Maintenance and infrastructure of the framework 61
4.7 Reproducibility: FAIR principles and the DARTHs blueprint 62
4.8 Operational prototype for the Noce River basin 64

CHAPTER 5. Mixed-resolution analysis of the water budget of an Alpine basin 69
Scope and original contributions 69
5.1 Introduction 70
5.2 Experimental setup 75
5.3 Results and discussion 79
5.3.1 Hydrological model calibration and validation against measured discharge 80
5.3.2 Water balance at the GU scale 84
5.3.3 Modeled snow assessment 88
5.3.4 Modeled soil moisture assessment 90
5.4 Conclusion 94

CHAPTER 6. Effects of spatial representation of meteorological data on different hydrological processes in semi-distributed hydrological models 97
Scope and original contributions 97
6.1 Introduction 98
6.2 Experimental setup 103
6.3 Results and discussion 107
6.3.1 Semivariogram and leave-one-out cross-validation to test kriging results 107
6.3.2 Evaluation of the different approaches on multiple hydrological processes 114
6.4 Conclusion 136

CHAPTER 7. Assessing spatial discretization impacts on semi-distributed hydrological simulations 140
Scope and original contributions 140
7.1 Introduction 141

7.2 Experimental setup **145**
7.2.1 Scale-focused calibration experiment 148
7.3 Results and discussion **149**
7.3.1 Discharge calibration and validation 150
7.3.2 Validation of snowpack and root-zone dynamics 160
7.3.3 Calibration experiments results 165
7.4 Conclusion **169**

CHAPTER 8. General synthesis and future perspectives **171**
8.1 Cross-study synthesis **172**
8.2 Main contributions **173**
8.3 Limitations and scope **174**
8.4 Future research directions **179**
8.4.1 Process representation and model-structure development 179
8.4.2 Calibration strategy and use of auxiliary information 181
8.4.3 Further development and short-range prediction of soil moisture 181
8.5 Concluding remarks **183**

APPENDIX A. Complete formulation of the hydrological model **185**
A.1 Rain/snow separation **185**
A.2 Snow compartment **186**
A.3 Canopy **187**
A.4 Root zone **188**
A.5 Runoff **189**
A.6 Groundwater **190**

APPENDIX B. Geostatistical interpolation: formulation and semivariogram models **191**

APPENDIX C. Physiographic and climatic characterization of the Noce basin **195**
C.1 Climate and data **195**
C.2 Physiographic and environmental setting of the Noce basin **195**

APPENDIX D. Supplementary tables and figures **199**
D.1 Study 1: model schematic and mapped outputs **199**
D.1.1 Model output 200
D.2 Study 2: configurations, semivariogram diagnostics and rainfall-runoff metrics **212**
D.2.1 Semivariogram and LOOCV 212
D.2.2 Rainfall-runoff 220
D.3 Study 3: generalization, metric ranges and cumulative discharge **228**

APPENDIX E. Computational cost of the model configurations **234**
E.1 Mixed-resolution configurations (Chapter 5) **235**
E.2 Interpolation and simulation cost by forcing strategy (Chapter 6) **235**
E.3 Cost as a function of the discretization scale (Chapter 7) **237**

APPENDIX F. Open research: data, code, and workflows **238**
F.1 Data sources and access conditions **238**
F.2 Study-specific OSF deposits **239**
F.2.1 Study 1: mixed-resolution model configuration 239
F.2.2 Study 2: meteorological forcing and GU representation 239
F.2.3 Study 3: spatial discretization 239
F.2.4 Repository structure and documentation 239
F.3 Source code **240**
F.4 License and citation **241**
F.5 Acknowledgments of data providers **242**

REFERENCES **243**

LIST OF TABLES

5.1 Characteristics of CFG1 and CFG2 by sub-catchment 77

5.2 Root-zone and groundwater storage changes, summer 2021 to 2022 low flow 87

5.3 Snow-cover detection metrics for the two configurations (CFG1 and CFG2). 88

5.4 Results of the field campaign and calibration of the CRNS probes. 90

6.1 Runoff coefficient by forcing method and spatial discretization 116

6.2 Malé performance metrics across configurations 120

6.3 Vermiglio performance metrics across configurations 121

6.4 Discharge agreement from the ECP analysis (r and RMSE) 127

6.5 Snow-cover performance metrics across configurations (MODIS vs model) 132

6.6 Root-zone soil moisture validation metrics (MERIDA vs model) 134

7.1 GU area and elevation ranges by spatial configuration 147

7.2 Parameters and search ranges for the second calibration 149

7.3 Streamflow performance at Vermiglio across spatial scales 151

7.4 KGE decomposition at Vermiglio 152

7.5 Streamflow performance at Malé across spatial scales 156

7.6 KGE decomposition at Malé 157

7.7 Classification summary by scale 163

7.8 Best-run streamflow performance at Vermiglio across spatial scales 166

A1 Symbols for the rain/snow separation module. 185

A2 Symbols for the snow compartment. 187

A3 Symbols for the canopy module. 188

A4 Symbols for the root-zone module. 189

A5 Symbols for the runoff module. 189

A6 Symbols for the groundwater module. 190

C1 Summary of temperature statistics (°C) for different stations in the basin. 195

C2 Summary of the spatio-temporal resolution of the datasets used in this thesis. 198

D1 Observed and simulated discharge volumes by hydrological year (Study 1) 202

D2 Characteristics of the model configurations (Study 2) 212

D3 Rainfall-runoff metrics at Malé (RMSE, PBIAS) (Study 2) 220

D4 Rainfall-runoff validation metrics at Vermiglio (RMSE, PBIAS) (Study 2) 220

D5 Malé total discharge-volume errors across configurations (Study 2) 222

D6 Generalization drop from calibration to validation at Vermiglio (Study 3) 228

D7 Generalization loss in streamflow performance from calibration to validation at Malé (Study 3) 228

D8 Maximum and minimum KGE values by spatial configuration (Study 3) 229

E1 Computational performance of CFG1–CFG3 235

E2 Computational requirements of the tested configurations during calibration 236

E3 Computational requirements of the 2017–2023 simulation runs 236

E4 Computational requirements at the Malé stream gauge 237

LIST OF FIGURES

2.1 Noce basin overview map and monitoring network 26

2.2 Map of the study basin 29

3.1 Extended Petri Net representation of the hydrological model 43

4.1 Activity diagram of the automated kriging workflow 58

4.2 Use-case diagram of the kriging component in OMS3 59

4.3 Web interface of the operational prototype 66

4.4 Information and diagnostics for a selected GU 67

5.1 GU maps for CFG1 and CFG2 76

5.2 GU subdivision and elevation differences (CFG1 vs CFG2) 78

5.3 Discharge at Vermiglio: observed vs modeled (CFG1/CFG2) 80

5.4 Discharge at Malé: observed vs modeled (CFG1/CFG2) 80

5.5 Water balance for GU 4829 (monthly vs weekly) 85

5.6 MODIS vs model snow presence (Brier score, CFG1/CFG2) 89

5.7 TRES1 CRNS vs model root-zone storage (CFG1/CFG2) 91

5.8 TRES2 CRNS vs model root-zone storage (CFG1/CFG2) 92

5.9 Root-zone storage correlation with Copernicus (CFG1 vs CFG2) 93

6.1 Experimental configurations and modeling workflow 105

6.2 LOOCV performance metrics for temperature and precipitation 110

6.3 Monthly forcing statistics across GUs for the nine configurations 118

6.4 Hydrographs at Malé for the nine model configurations 119

6.5 Brier score by elevation band for snow cover (NEAR, CEN, GRIDW) 138

6.6 Soil-moisture RMSE by elevation band 139

7.1 Experimental design: GU discretizations 146

7.2 Discharge at Vermiglio gauge across spatial scales 153

7.3 Discharge at Malé gauge across spatial scales 155

7.4 Brier score by elevation band across spatial scales 162

7.5 Soil-moisture correlation (r) by elevation band across spatial scales 164

C1 Lithological map of the Noce basin 197

D1 Discharge uncertainty at Vermiglio (CFG1 vs CFG2) (Study 1) 200

D2 Discharge uncertainty at Malé (CFG1 vs CFG2) (Study 1) 200

D3 Cumulative discharge for CFG1 and CFG2 (Study 1) 201

D4 Box-plots of monthly volume errors (Study 1) 203

D5 CFG2 water balance for GUs derived from CFG1 GU 4829 (Study 1) 204

D6 Annual water balance for the Noce basin (Study 1) 205

D7 SWE in 2021 vs 2022 for the whole basin (CFG1) (Study 1) 205

D8 SWE in 2021 vs 2022 for the whole basin (CFG2) (Study 1) 206

D9 Root-zone water content anomaly (2021 vs 2022) (Study 1) 206

D10 Basin-mean snow water equivalent (Study 1) 207

D11 Basin-mean root-zone water content (Study 1) 207

D12 Basin-mean groundwater storage (Study 1) 208

D13 Spatial distribution of simulated snow water equivalent (Study 1) 209

D14 Spatial distribution of simulated root-zone water storage (Study 1) 210

D15 Spatial distribution of simulated groundwater storage (Study 1) 211

D16 Temperature variogram models (OK vs DK) (Study 2) 212

D17 Precipitation variogram models (OK vs DK) (Study 2) 213

D18 Seasonal temperature variogram models for OK (Study 2) 214

D19 Seasonal temperature variogram models for DK (Study 2) 215

D20 Seasonal precipitation variogram models for OK (Study 2) 216

D21 Seasonal precipitation variogram models for DK (Study 2) 217

D22 Temperature LOOCV performance by season (Study 2) 218

D23 Temperature LOOCV performance by elevation band (Study 2) 218

D24 Precipitation LOOCV performance by season (Study 2) 219

D25 Precipitation LOOCV performance by elevation band (Study 2) 219

D26 Cumulative discharge volumes at the Malé stream gauge for the nine model configurations (Study 2) 221

D27 Monthly discharge-volume error at the Malé stream gauge (Study 2) 223

D28 Scatter plots at the Malé stream gauge for the nine model configurations (Study 2) 224

D29 Annual water balance for the nine model configurations (Study 2) 225

D30 Monthly mean AET across GUs (CEN vs GRIDW) (Study 2) 226

D31 Soil-moisture correlation by elevation band (Study 2) 227

D32 Cumulative discharge volumes across spatial configurations at Vermiglio (Study 3) 230

D33 Cumulative discharge volumes across spatial configurations at Malé (Study 3) 231

D34 Annual water-balance components across spatial configurations at Malé (Study 3) 232

D35 Variability of root-zone, runoff, and groundwater parameters across spatial scales (Study 3) 233

D36 Variability of snow-related parameters across spatial scales (Study 3) 233

LIST OF ACRONYMS

AET	Actual Evapotranspiration
ARF	Areal Reduction Factor
CEN	Centroid-based Forcing Evaluation
CFG1	Configuration 1
CFG2	Configuration 2
CRNS	Cosmic-Ray Neutron Sensing
DEM	Digital Elevation Model
DK	Detrended Kriging
ECP	Empirical Conditional Probability
ERA5	ECMWF Reanalysis v5
ERM	Embedded Reservoir Model
FN	False Negative
FP	False Positive
GRIDW	Grid-weighted Forcing Evaluation
GU	Geomorphic Unit
GU*	Geomorphic GU Delineation
HRU	Hydrological Response Unit
KGE	Kling–Gupta Efficiency
LOOCV	Leave-One-Out Cross-Validation
MAE	Mean Absolute Error
MERIDA	MEteorological Reanalysis Italian DAtaSet
MODIS	Moderate Resolution Imaging Spectroradiometer
NDSI	Normalized Difference Snow Index
NEAR	Nearest weather station-Forcing Evaluation

NSE	Nash–Sutcliffe Efficiency
OAT	Overall Accuracy Test
OK	Ordinary Kriging
OMS	Object Modeling System
PBIAS	Percent Bias
PET	Potential Evapotranspiration
r	Pearson Correlation Coefficient
R^2	Coefficient of Determination
RC	Runoff Coefficient
RMSE	Root Mean Square Error
SCA	Snow Cover Area
SG	Stream-gauge-based Delineation
SSM	Surface Soil Moisture
SWE	Snow Water Equivalent
TN	True Negative
TP	True Positive
WRF-ARW	Weather Research and Forecasting Model – Advanced Research WRF

ABSTRACT

The hydrologic regime of Alpine regions is undergoing rapid change under global warming, characterized by an alternation of wet and dry years, and even by floods and droughts within the same year. In this context, accurate catchment-scale water balance estimation is essential. To manage water withdrawals and the competing demands of hydropower, agriculture, domestic use, tourism, and industry under such variability becomes difficult and requires new tools and multiscale answers. Therefore, this work investigates scale issues in Alpine catchments and implements new modeling components. The multiscale approach is based on the GEOframe system and its space partitioning into Geomorphic Units (GU), which is mixed with a grid approach. This has practically implied writing a tool to increase the resolution of sub-catchment extraction, tools for grid information averaging over the GU, and adaptation of many existing GEOframe tools. These improvements have also affected the kriging interpolator, whose new version accommodates the possibility of varying the semivariogram to capture the temporal variability of meteorological forcing at each time step. These developments also contributed to a broader refactoring of the GEOframe system, including bug fixes, the reorganization of its Git repositories and website, and improvements to the build system. The resulting code also supported the work carried out for the AdbPo project. The tools were extensively applied to the Noce River basin, an area of particular agricultural interest. The modeling and the analysis combined ground-based observations (weather stations and cosmic-ray neutron probes) with remote sensing products (snow cover and surface soil moisture). Summarizing the methodology, hourly semivariograms were used to estimate meteorological variables on a grid, subsequently aggregated as weighted area averages over the GUs and compared to the "centroid-based approach" formerly used, showing interesting improvements in simulating the river discharges. Besides, the Noce River basin was modeled at various resolutions, meaning with coarser and finer GUs. The multiscale approach also shows improved validation performance not only for discharge but

also for soil moisture and snow cover when GUs are smaller. The modeling infrastructure was implemented according to FAIR principles and respecting the DARTHs blueprint and is currently operational for water budget analysis.

CHAPTER 1. INTRODUCTION

Water plays a central role in sustaining ecosystems and supporting societies and is therefore fundamental to human well-being. Water availability and socio-economic development are closely intertwined through a complex and continuously evolving balance between supply and multiple demands. The same limited resource is required for domestic supply, agriculture, hydropower production, industry, and tourism, while sufficient water must also be preserved to sustain aquatic ecosystems. Its management is therefore intrinsically complex, as competition can arise among water uses, particularly when availability is insufficient to meet all demands (Wada et al., 2017). Managing these competing demands is further complicated by the different hydrological compartments in which water is stored and among which it is transferred, including snow, root-zone soil moisture, groundwater, and surface-water bodies. Water storage and availability in these compartments vary across space and time and depend on the interactions among precipitation, snow accumulation and melt, evapotranspiration, infiltration, recharge, and runoff (Abera et al., 2017; Wood et al., 2011). Moreover, human activities, including water abstraction, irrigation, reservoir regulation, and flow diversion, modify these fluxes and storages across multiple spatial and temporal scales (Wada et al., 2017). Water-management decisions must therefore account not only for the total amount of water available, but also for where it is stored, when it becomes available, and how its use in one part of the system may affect other users and ecosystems (Wada et al., 2017). This complexity is further amplified by climate change, which is altering water availability and demand and increasing pressure on water-management systems (Cammalleri et al., 2020; Wada et al., 2016).

In recent decades, the impacts of climate change have become increasingly evident, particularly through rising temperatures and the increasing frequency of extreme events such as floods and droughts. Mountain regions, which are characterized by high elevations,

steep climatic gradients, and complex ecosystems, are among the most sensitive, rapidly changing, and fragile environments. This is especially evident in the Alps (Beniston & Stoffel, 2014; Huss et al., 2017), where pronounced topographic variability and exposure to large-scale atmospheric circulation patterns make the region an effective "sentinel of environmental change" (Beniston & Stoffel, 2014). In such environments, physical and ecological responses to warming often become visible earlier and more clearly than in lowland areas (Beniston, 2006; Beniston & Stoffel, 2014).

Due to their central location in Europe, the Alps span several countries and strongly influence continental climate, ecosystems, hydrology, and socio-economic activities (Gobiet et al., 2014). For example, the major European rivers—including the Rhine, Rhône, Danube, and Po—originate in the Alps and provide freshwater to more than 170 million people downstream (Stephan et al., 2021; Viviroli et al., 2007a). In this sense, the Alpine arc plays a crucial hydrological role and functions as one of the main "water towers" of Europe (Beniston, 2006; Beniston et al., 2011; Dahlmann et al., 2022; Gobiet et al., 2014; Zierl & Bugmann, 2005). Consequently, to address the complex environmental challenges that modern hydrological models and tools must confront, the choice of spatial and temporal scale becomes crucial.

Several studies emphasize this strategic function for water supply, hydropower production, agriculture, and ecosystem services (Beniston et al., 2011; Viviroli et al., 2007a, 2011).

Consequently, climate change in the Alpine region has far-reaching implications that extend well beyond mountain areas, affecting economic sectors, infrastructure, and water security in downstream regions (Beniston et al., 2011; Gobiet et al., 2014; Immerzeel et al., 2020).

Over recent decades, several major droughts have affected Europe, demonstrating the vulnerability of interconnected natural and socio-economic systems to water scarcity and making the associated management challenges particularly evident. Current drought-related

economic losses in the European Union and the United Kingdom are estimated at approximately €9 billion per year, expressed in 2015 values. At a global warming level of 3 °C above pre-industrial levels, these losses could increase to approximately €17.3 billion per year if current socio-economic conditions are maintained (Cammalleri et al., 2020).

Among these events, the 2022 drought was particularly severe and spatially extensive (Toreti et al., 2022). Agriculture is generally the most directly affected sector through reduced soil-water availability, irrigation restrictions, declining crop yields, and higher production costs (Cammalleri et al., 2020). In the Po River basin, agricultural production was severely affected, particularly for water-intensive crops (Bonaldo et al., 2023). The event was exacerbated by an exceptionally severe winter snow deficit, which contributed to reduced summer streamflow and coincided with the introduction of emergency water-use restrictions across the basin (Avanzi et al., 2024). In the Po River Delta, the exceptionally low river discharge allowed salt water to intrude more than 30 km inland, constraining irrigation withdrawals and threatening agricultural soils, freshwater ecosystems, and water-supply sources (Autorità di Bacino Distrettuale del Fiume Po [ADBPo], 2022; Bonaldo et al., 2023).

However, drought impacts extend well beyond agriculture and hydropower production. Low river levels can disrupt inland navigation and industrial supply chains. Ademmer et al. (2023) estimated that 30 consecutive days of low water levels along the Rhine reduce German industrial production by about 1%. Water scarcity and high river temperatures can also constrain thermoelectric generation because sufficient water is required for plant cooling. In France, these conditions limited nuclear power generation during the summer of 2022, although the associated losses represented only a small fraction of total annual nuclear production (Électricité de France [EDF], 2022). Drought can also directly affect the population by reducing water availability, deteriorating water quality, and increasing pressure on drinking-water supply systems. Consequently, local authorities may be required to restrict non-essential domestic water uses (Cammalleri et al., 2020).

At the time of writing, Europe is again experiencing widespread drought conditions. During summer 2026, warning or alert conditions affected large parts of western and central Europe, including northern Italy (Toreti et al., 2026). In August 2026, all the main monitoring sections of the Po River were classified as experiencing extreme hydrological drought. Persistent saline intrusion in the delta constrained or prevented water abstraction, while an increasing number of municipalities reported difficulties in drinking-water supply and were advised to restrict non-essential uses of the public water supply (ADBPo, 2026).

Taken together, these events show how water scarcity can propagate through interconnected natural and socio-economic systems and increase the complexity of water-management decisions. Streamflow observations alone provide only a partial representation of water availability because they do not explicitly characterize how water is partitioned among snow, soil moisture, groundwater, evapotranspiration, and surface-water storage (Wagner et al., 2025). An integrated assessment of the main water-budget components is therefore needed to quantify the amount, location, and timing of water availability for different users and ecosystems (Abera et al., 2017; Wood et al., 2011). In response to this need, hydrological research is increasingly combining in situ observations, remote sensing, and modeling to represent water-budget dynamics at spatial and temporal scales relevant to water management (Bierkens et al., 2015; Wood et al., 2011). These advances can provide more directly actionable information for water allocation, drought preparedness, and the evaluation of trade-offs among competing water uses (Loucks, 2023; Wada et al., 2017).

1.1 Climate change in the Alps

Observed and projected changes in temperature and precipitation in the Alps are already substantial and are expected to intensify in coming decades. Since the late nineteenth century, the Alps have warmed by about 2 °C, at roughly twice the rate observed in many other Northern Hemisphere regions (Corradini et al., 2023; Dumont et al., 2025; Gobiet et al., 2014). Projections indicate continued warming throughout the twenty-first century,

with magnitude depending on emissions and season (Dumont et al., 2025; Gobiet et al., 2014; Kotlarski et al., 2023).

Precipitation changes are expected to be dominated by shifts in seasonality and spatial patterns rather than uniform changes in annual totals (Gobiet et al., 2014). Regional scenarios generally suggest wetter winters and drier summers over large parts of the Alps, with strong spatial variability and substantial inter-model spread (Beniston, 2006; CH2018, 2018; Dumont et al., 2025; Fischer et al., 2022; Gobiet et al., 2014; Zierl & Bugmann, 2005). Extreme events are also projected to intensify, including exceptionally warm summers and milder winters (Gobiet et al., 2014; Zierl & Bugmann, 2005).

These trends are already affecting Alpine hydrology and related hazards. Warming and snow drought are accelerating glacier retreat, with projections indicating that most remaining glacier ice may disappear by the end of the century (Beniston & Stoffel, 2014; Beniston et al., 2011; Gobiet et al., 2014; Huss et al., 2017). Permafrost degradation further destabilizes steep slopes, increasing the likelihood of rockfalls, slope failures, and debris flows (Beniston, 2006; Beniston & Stoffel, 2014). Representing these processes in a hydrological model requires an explicit treatment of the coupled water and energy budget of freezing soils, for which energy-conserving formulations and their implementation in physically based land-surface schemes are now available (Dall'Amico et al., 2011; Endrizzi et al., 2014; Wani et al., 2026). Snow cover is projected to decline markedly, especially at low and mid elevations, weakening the seasonal storage function of the cryosphere and shifting the timing of meltwater contributions (Beniston et al., 2011; Dall'Amico et al., 2025; Gobiet et al., 2014; Kotlarski et al., 2023; Zierl & Bugmann, 2005). As a consequence, winter discharge is expected to increase and summer discharge to decrease, with earlier peak flows driven by earlier snowmelt (Zierl & Bugmann, 2005). Snowmelt-related droughts, such as the 2022 event in the Po basin, are therefore expected to become more frequent (Avanzi et al., 2024; Brunner et al., 2023).

More broadly, climate change is projected to intensify drought processes (meteorological,

soil-moisture, hydrological, and snow-related), increasing the frequency and persistence of low-flow conditions and pressure on agriculture, hydropower, ecosystems, and water supply (Avanzi et al., 2024; Brunner et al., 2023; Gobiet et al., 2014; Roudier et al., 2016; Stephan et al., 2021; Viviroli et al., 2011). In parallel, heavier precipitation extremes may increase flood risk, especially in winter, and reduced snow storage may contribute to more rapid, "flashy" catchment responses to rainfall (Gobiet et al., 2014; Huss et al., 2017; Zierl & Bugmann, 2005).

These changes have direct socio-economic implications, including challenges for reservoir and hydropower management, increasing reliance on artificial snowmaking in winter tourism, rising drought stress in agriculture and forestry, and higher exposure to mountain hazards (Beniston & Stoffel, 2014; Corradini et al., 2023; Dahlmann et al., 2022; Gaudard et al., 2014; Gobiet et al., 2014). Overall, the Alpine region is undergoing climatic and hydrological transformations that affect multiple components of the water cycle and require increasingly adaptive management. This, in turn, strengthens the need for hydrological modeling tools that can represent key processes, explore scenarios, and support decision-making in complex mountain environments.

As discussed in the Special Issue edited by V. Kumar et al. (2025), modeling approaches contribute across a wide range of applications, from groundwater vulnerability assessment and improved flow modeling to nature-based solutions and drought early-warning systems.

1.2 Hydrological modeling: selection, evaluation, and scale-related issues

Hydrological models are therefore central to addressing Alpine challenges. However, the diversity of available modeling tools raises a practical question: how should one select a model that is suitable for a given purpose? As discussed by Horton et al. (2022), model selection should follow principles of parsimony and adequacy: models should be no more complex than required by the research question and the information content of available data, while still capturing the processes relevant at the target scales. Existing approaches span a

spectrum from lumped to distributed formulations, and from conceptual representations to more physically based descriptions of catchment behavior.

In parallel, open-source modeling frameworks have gained increasing relevance because they facilitate systematic experimentation and transparent reporting of modeling choices. Frameworks such as FLEX (Fenicia et al., 2008; Kavetski & Fenicia, 2011), SUMMA (Clark et al., 2015b), SuperflexPy (Dal Molin et al., 2021), and GEOframe (Abera et al., 2017; Bancheri, 2017; Formetta et al., 2014a; Serafin et al., 2018) provide modular and extensible environments that support the development and comparison of alternative model structures and hypotheses. In this context, transparency (e.g., inspectable code, explicit configurations, and traceable workflows) is a key enabler of reproducibility, although reproducibility ultimately also depends on data provenance, parameter choices, and the completeness of reporting.

Beyond specific applications, it is also important to consider how hydrological information is produced and used. Brunner et al. (2021) argue that forecasting, projection, and frequency analyses are frequently driven by a single class of extreme event (drought or flood) and may overlook transitions between regimes and related non-stationarities. They highlight methodological gaps and propose improvements along the entire chain—from data and process understanding, through modeling, to the representation of human-water interactions.

Within this broader perspective, hydrological modeling increasingly intersects with hydroinformatics. As argued by Loucks (2023), combining tools of different nature within a multidisciplinary approach can improve the capacity of models to support decision-making, particularly when uncertainty and competing objectives must be explicitly acknowledged. This view also emphasizes that modeling should engage with sustainability (including economic and political feasibility) and with normative dimensions such as ethics, equity, and the protection of fundamental rights. These ideas are closely connected to the development of digital-twin concepts for hydrological and socio-hydrological systems (Rigon et al.,

2022).

Despite these advances, two challenges remain particularly critical and motivate the analyses presented in this thesis:

1. **Reliability and robustness of model evaluation beyond calibration.** Hydrological predictions are affected by a chain of uncertainties and errors, including observational errors, uncertainty in meteorological forcing, parameter uncertainty, and structural model error (Todini & Biondi, 2017). In this context, *reliability* refers to the credibility of model outputs for a specified purpose, whereas *robustness* refers to the stability of model performance under changing conditions (e.g., different hydroclimatic regimes, periods, or plausible perturbations of inputs and parameters). Calibration can improve agreement with observed discharge, but it cannot remove structural inadequacies, and "optimal" parameter sets may not be unique or physically interpretable—an issue closely related to equifinality (K. Beven, 2006). Moreover, widely used objective functions (e.g., Nash–Sutcliffe Efficiency, Kling–Gupta Efficiency) may emphasize specific hydrograph features while neglecting others, and can therefore be misleading if used in isolation (Cinkus et al., 2023; Clark et al., 2021; Lamontagne et al., 2020; Santos et al., 2018). For these reasons, robust evaluation increasingly relies on multiple and independent sources of information (multi-variable and/or multiscale evaluation). A representative example is the use of remotely sensed snow cover data to constrain and evaluate snow-related processes (Parajka & Blöschl, 2006).

2. **Representing hydrological heterogeneity across scales.** Hydrological processes exhibit strong spatial and temporal heterogeneity, and understanding how process behavior emerges across scales remains a central scientific challenge (e.g., Blöschl et al., 2019). This issue has direct operational implications, because stakeholders require actionable information at spatial and temporal scales relevant for flood and drought adaptation, water allocation, irrigation scheduling, and hydropower optimization (e.g., Wada et al., 2017). Addressing scale-related limitations therefore requires

both improved process representations and an observational basis that spans multiple processes and resolutions. Consequently, it becomes essential to integrate data from multiple sources and to exploit modeling approaches that can explicitly support multiscale and multi-process hypotheses.

The next step is therefore to discuss the different data sources that can be used to force, calibrate, and—critically—evaluate hydrological models across processes and scales.

1.2.1 Multi-source observations for model evaluation

Recent research on model evaluation and validation highlights the benefits of exploiting multiple and independent observational datasets to improve both the reliability and the robustness of hydrological simulations (Biondi et al., 2012; Koltsida & Kallioras, 2022; Moriasi et al., 2007). The motivation is that discharge alone often provides insufficient information to constrain internal states and fluxes: different model structures or parameter sets may reproduce streamflow with similar skill while compensating errors across processes (e.g., snow dynamics, soil-water storage, and evapotranspiration). Integrating observations that directly inform these processes therefore supports more diagnostic evaluation and reduces the risk of equifinality-driven interpretations.

In data-scarce mountain regions, remote sensing products are increasingly valuable because they offer spatially distributed information that complements sparse ground networks. For instance, Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover products from the Terra and Aqua satellites provide near-daily observations at approximately 500 m spatial resolution, supporting regional snow cover monitoring and model evaluation (e.g., Sirguey et al., 2008). Cloud contamination remains a key limitation, commonly addressed through temporal filtering and gap-filling approaches (e.g., Parajka & Blöschl, 2006; Şorman et al., 2007). Similarly, satellite-derived soil moisture datasets (e.g., Copernicus services) provide information on near-surface wetness at regional scales and can be used to assess whether simulated soil-moisture dynamics reproduce plausible temporal variability

and spatial patterns (Dorigo et al., 2017).

Despite the growing role of satellite products, ground-based measurements remain essential for process-level validation and for addressing uncertainties that remote sensing cannot fully resolve. Cosmic-Ray Neutron Sensing (CRNS) provides non-invasive, field-scale estimates of soil moisture with a footprint on the order of 10^2 m (radius) (Zreda et al., 2008, 2012) and sub-daily temporal resolution (Franz et al., 2020), helping bridge the gap between point measurements and satellite retrievals. CRNS has been applied in snow hydrology and agricultural water management (Brogi et al., 2023; Franz et al., 2020; Schattan et al., 2017) and integrated with hydrological models across multiple spatial scales (Franz et al., 2013, 2020; Scheiffele et al., 2025). Overall, the increasing availability of heterogeneous observations strengthens the empirical basis for hydrological modeling and supports evaluation strategies that go beyond streamflow.

Having identified scale and heterogeneity as key challenges addressed in this thesis, this section examines these issues in greater detail, focusing on the relationships among process, observation, and modeling scales and on how spatial heterogeneity is represented through different model discretizations.

1.3 Scale and spatial discretization in hydrological modeling

Beyond purely technical aspects of numerical implementation and model calibration, hydrological modeling is constrained by more fundamental questions of scale. The relationship between process representations, data support, and model discretization remains an open problem and is directly linked to several of the unresolved questions formulated by Blöschl et al. (2019), particularly Problem 6, which asks whether general laws exist to describe how dominant hydrological processes change with scale. A useful starting point is the distinction between process (or intrinsic) scale, observation (or data) scale, and modeling (or working) scale (Blöschl & Sivapalan, 1995). Hydrological processes such as infiltration, snow accumulation, and groundwater recharge operate over characteristic spatial

and temporal scales that may differ substantially from the support of available measurements and from the scales at which models are designed. The resulting scale mismatch introduces an additional layer of complexity: fine-scale variability must be aggregated or parameterized, information is lost, and the underlying process behavior is not always preserved, particularly in highly heterogeneous or topographically complex basins (Blöschl & Sivapalan, 1995; Seyfried & Wilcox, 1995).

One of the main questions is therefore how to represent heterogeneity and variability across scales. The representative elementary area (REA) concept proposed by Wood et al. (1988) and further investigated by Blöschl et al. (1995) suggests that beyond a characteristic length scale, catchment heterogeneity can be treated statistically, such that non-spatial descriptors (e.g., means, variances, probability distributions) are sufficient to parameterize model behavior. At scales smaller than the REA, explicit spatial patterns of rainfall, topography, and soils may become influential and may need to be represented more explicitly. Seyfried and Wilcox (1995) also show that the deterministic versus stochastic classification varies with the model intrinsic scale: patterns that appear random at one scale may appear organized at another. Importantly, neither the REA nor any deterministic length scale is a universal constant; both depend on catchment characteristics, dominant processes, and on the space-time structure of the forcing (Blöschl & Sivapalan, 1995; Seyfried & Wilcox, 1995).

Different hydrological model families embed different assumptions about scale. Lumped conceptual models treat the catchment as a single unit, where parameters represent integrated behavior at the basin scale and small-scale heterogeneity is implicitly absorbed into effective parameter values. Semi-distributed conceptual models (SDMs) introduce a coarser representation of spatial variability by discretizing the basin into sub-catchments or Geomorphic Units (GUs), while typically assuming internal homogeneity within each unit. Fully distributed, grid-based models and large-scale land-surface or Earth system models aim to resolve spatial patterns more explicitly, but must still rely on effective parameterizations

to represent sub-grid heterogeneity. As a result, they remain subject to scale-related issues, including equifinality and structural uncertainty (K. J. Beven, 2001; Clark et al., 2015a, 2017).

Across model families, the key challenge is to design model structures and parameterizations that are consistent with the scales at which both processes and data operate. One strategy to cope with scale effects is to seek parameterizations that are approximately scale invariant, in the sense that functional relationships between model parameters and physiographic predictors can be transferred across multiple spatial resolutions with limited re-calibration. This idea is formalized in the Multiscale Parameter Regionalization (MPR) framework, which links high-resolution predictors (e.g., geophysical data) to grid-scale "effective parameters" that account for sub-grid variability (R. Kumar et al., 2013; Samaniego et al., 2010). Applications of MPR indicate that such an approach can reduce resolution-dependent inconsistencies in simulated streamflow over large domains (Mizukami et al., 2017; Samaniego et al., 2010).

Scale also plays a central role in semi-distributed conceptual models, where spatial discretization is commonly based on GUs. In many implementations, GUs are defined as aggregations controlled by user-defined thresholds, and changes in GU resolution or aggregation rules can modify the resulting "effective" parameters and potentially alter the dominant processes expressed by the model. Field examples reported by Seyfried and Wilcox (1995) emphasize how hydrological response and the apparent nature of spatial variability change with increasing support area and contributing area, reinforcing the idea that threshold-based discretizations impose a specific scale on runoff generation and connectivity. In many SWAT-like applications, commonly used HRU thresholds have been shown to have minor impacts on water yield and streamflow, but to substantially affect simulated sediment and nutrient loads (Chiang et al., 2015; Her et al., 2015).

Building on these scale considerations, this thesis aims to move the investigation toward finer spatial supports to better represent heterogeneity and short-term hydrological dynamics,

while explicitly testing the trade-offs introduced by discretizations and forcing choices.

1.3.1 Hyper-resolution scale

To address the challenges in modern hydrological modeling (e.g., strong spatial heterogeneity, rapid hydrological dynamics, data limitations, computational costs, equifinality, and difficulties in validating internal model states), the choice of scale is crucial. In this thesis, *hyper-resolution* refers to spatial resolutions finer than 1 km^2 and to hourly time steps, which are capable of capturing spatial heterogeneity and rapid hydrological dynamics. Hyper-resolution land-surface and hydrological modeling has been proposed as a major challenge at the global scale (Bierkens et al., 2015; Wood et al., 2011), but it is also highly relevant at regional and basin scales (Mascaro et al., 2015; Wada et al., 2016).

Traditional coarse-resolution models, sometimes combined with limited hydro-meteorological data, are often unable to capture processes that become evident only at finer scales, and model outputs can even become misleading for some applications. By contrast, hyper-resolution models can represent physical processes in greater detail and thus support a better understanding of hydrological dynamics and their socio-economic impacts (Wada et al., 2016; Wood et al., 2011). The use of high resolution is fundamental, for example, for flash floods and other events whose spatio-temporal scale is smaller than the intrinsic model scale, for resolving the dynamics of key states (e.g., soil moisture), and for improving the representation of topography, compartment interactions, and anthropogenic impacts (Wada et al., 2016).

Alongside these advantages, hyper-resolution modeling poses technical challenges: simulations can be computationally expensive; data availability at the desired resolution is often limited; higher detail does not necessarily guarantee higher accuracy; and increasing resolution can make calibration more complex, potentially enhancing parameter equifinality (K. J. Beven & Cloke, 2012; Bierkens et al., 2015; Wada et al., 2016; Wood et al., 2011). It also increases the need to validate not only discharge at the outlet but also the spatio-

temporal evolution of internal components (soil moisture, snow, subsurface storage, lateral flows), often requiring multi-source observational datasets (Mascaro et al., 2015).

These motivations are particularly relevant in Alpine catchments, which exhibit strong climatic and geomorphological gradients over short distances, widespread hydraulic infrastructure, and rapid hydrological responses during intense precipitation events and snowmelt periods (Bierkens et al., 2015; Wood et al., 2011). At the same time, increasing resolution is not a sufficient solution: fully distributed setups at tens to hundreds of meters and hourly time steps can be computationally demanding, and their effective use is limited by the availability and consistency of input and evaluation data, especially for subsurface properties and human interventions (K. J. Beven & Cloke, 2012; Mascaro et al., 2015).

These considerations motivate the analysis presented in the following sections, which focuses on how meteorological forcing representation and spatial discretization choices control model behavior at fine scales, and on whether the additional complexity is justified in terms of performance and process realism.

1.3.2 Meteorological forcing and spatial variability in semi-distributed models

Meteorological variables are fundamental drivers of environmental processes and are essential for a wide range of applications. In hydrology, precipitation and temperature play particularly critical roles in controlling the processes underlying rainfall-runoff modeling, and their magnitude, quality, and spatial distribution strongly affect model performance and outcomes (Bardossy & Das, 2008; K. Beven & Westerberg, 2011; K. J. Beven, 2001; De Lavenne et al., 2016; Wilby et al., 2017; Zhu et al., 2018). This requirement is especially important in Alpine catchments, where complex topography produces pronounced spatial variability across multiple spatial and temporal scales (Evin et al., 2024; Ledesma & Futter, 2017; Vaze et al., 2011). Precipitation is characterized by strong spatio-temporal variability, especially at high temporal resolutions (Berndt & Haberlandt, 2018; Emmanuel et al., 2012; Hu et al., 2019), whereas temperature generally exhibits lower spatial variability and is thus

easier to represent in Alpine contexts (Berndt & Haberlandt, 2018; Orlowsky & Seneviratne, 2014).

In semi-distributed models, where the GU is often used as the basic computational unit, meteorological inputs are commonly assigned to each GU either from the nearest weather station or from the GU centroid (Jiang et al., 2024; Ledesma & Futter, 2017; Morlot et al., 2023). In both cases, however, there is a risk of spatial misrepresentation in regions with challenging topography and large elevation ranges, leading to biased inputs and a poor representation of spatial heterogeneity.

To overcome these limitations, an alternative is to use gridded meteorological fields derived from interpolation of ground-station measurements, remote sensing datasets, or reanalysis products (Evin et al., 2024; Jiang et al., 2024; Ledesma & Futter, 2017; Loritz et al., 2021; Mankin et al., 2025; Poblete et al., 2020; Vu et al., 2012). The relative influence on model performance of gridded versus point-based forcing is not trivial and depends on catchment state and structure and on event characteristics (Loritz et al., 2021). Moreover, model structural error plays a crucial role, because parameter estimates obtained during calibration often compensate for structural deficiencies and input uncertainty (Todini & Biondi, 2017).

In summary, meteorological grids should be interpreted primarily as an intermediate representation of spatially varying forcing fields (and their uncertainty), rather than as an implicit definition of the hydrologic model resolution.

1.4 Objectives and structure of the thesis

The previous sections highlight that climate-driven changes in Alpine hydrology are increasing the need for modeling tools that are both scientifically credible and operationally useful. Two issues are particularly critical in this context. First, model skill assessed only against outlet discharge may be insufficient to constrain internal states and fluxes, making robust evaluation increasingly dependent on multi-variable and multi-source information.

Second, hydrological processes and observations operate across a hierarchy of spatial and temporal scales, so discretization choices and forcing representations can strongly influence model behavior in heterogeneous mountain catchments. These challenges motivate the need for scale-aware and transparent modeling workflows that explicitly test alternative hypotheses while remaining computationally feasible and consistent with the available data.

Despite the recognized importance of scale, the implications of discretizations and meteorological forcing representation are often insufficiently constrained in Alpine basins, especially within GU-based semi-distributed frameworks (Evin et al., 2024; Ledesma & Futter, 2017; Vaze et al., 2011). This thesis addresses this gap through two research questions that explicitly separate model resolution from forcing representation:

1. **RQ1 – Scaling behavior and mixed configurations.** How do discretizations at different spatial scales influence hydrologic predictability and process representation, including the partitioning of water-budget components, and can mixed-resolution configurations improve the trade-off between realism and computational efficiency? (Blöschl et al., 2019)

2. **RQ2 – Spatial representation of meteorological forcing and scaling.** How important are meteorological forcing patterns for runoff generation and how do they influence all components of the water budget at the catchment scale, and how does their influence vary across sub-catchments of different sizes? In particular, does the impact of precipitation and temperature representation depend on the GU delineation process, i.e., on the spatial scale at which GUs are defined? (K. J. Beven & Hornberger, 1982; Lobligeois et al., 2014; Loritz et al., 2021)

For both research questions, reliability and robustness are assessed beyond calibration and outlet discharge by using multiple independent sources of information for model evaluation.

The two research questions address the challenges introduced above from complementary

perspectives. While scale and heterogeneity provide the primary scientific focus of this thesis, model evaluation provides the methodological framework through which both questions are investigated. In this way, scale-dependent modeling hypotheses are assessed not only in terms of predictive performance, but also in terms of their effects on internal processes and water-budget partitioning.

Accordingly, this thesis investigates how spatial discretizations and meteorological forcing strategies affect hydrological simulations in Alpine basins, with a specific focus on resolutions that are fine enough to represent strong gradients and rapid dynamics (hourly time steps and, where feasible, kilometer-scale information). The analyses quantify trade-offs between increased spatial detail, computational cost, and the robustness of model evaluation beyond streamflow.

1.4.1 Methodological approach and thesis structure

To address the research questions outlined above, the methodological approach combines multiscale hydrological modeling, multi-source observations, and process-based model evaluation.

This thesis uses the Noce River basin (Trentino, northeastern Italy) as a test case, a large Alpine catchment characterized by strong elevation-driven climatic gradients and heterogeneous hydrological processes. The basin is also socio-economically strategic—supporting hydropower production, intensive irrigated agriculture, and tourism—so changes in water availability and model reliability have direct operational relevance.

In this work, model resolution is defined by the spatial discretization into GUs. We adopt a semi-distributed conceptual model in which the basin is represented as a set of geomorphologically delineated units obtained using a contributing-area threshold (i.e., a minimum drainage area for unit or channel initiation). This choice fixes the plan-view size of model elements and imposes an explicit characteristic scale for hillslope representation, while implicitly aggregating land use and soil variability within each unit. Routing follows the

drainage network represented as a directed tree graph, with fluxes propagating downstream along the graph.

1.4.2 Experimental strategy

The analyses follow a common workflow. The available input data are first checked for consistency and, where required, spatially interpolated before being used to run the hydrological model. Model parameters are calibrated exclusively against observed discharge. Other observations, including snow-cover, soil-moisture, and CRNS data, are subsequently used to evaluate the simulated hydrological compartments independently of streamflow. Where necessary, parameters associated with the processing or interpretation of these observations, such as snow-cover thresholds, are estimated separately and do not form part of the hydrological-model calibration. Model evaluation extends beyond conventional streamflow-performance metrics to include water-budget components, cumulative volumes, spatial patterns, and temporal dynamics, using maps and time-series diagnostics as appropriate.

Among the strategies adopted to investigate the effects of spatial representation on hydrological modeling, we design a variable-resolution multiscale strategy for an Alpine basin. GUs are first delineated at a basin-wide coarse scale and then refined at a finer scale in selected areas to support targeted investigations, while maintaining feasible computational costs (S. Kumar & Ojha, 2024; B. Li et al., 2024; Samaniego et al., 2017). This approach also aligns with "model-everywhere" perspectives that aim for comprehensive yet efficient environmental modeling solutions (K. Beven, 2025; Rigon et al., 2022).

Meteorological forcing is treated with a complementary two-level design. Rather than using only centroid-based assignments, the workflow performs hourly geostatistical interpolation on a regular grid, with semivariograms estimated automatically at each time step (Haberlandt, 2007; Lucas et al., 2022). Grid values are subsequently aggregated as area-weighted averages over GUs. In this way, forcing fields retain spatial structure at an intermediate (grid) scale while GUs remain the computational units for hydrologic sim-

ulations, providing some advantages commonly associated with hyper-resolution forcing without adopting a fully grid-cell-based hydrologic scheme (Bierkens et al., 2015; Wood et al., 2011).

A key interpretive point adopted throughout the thesis is that meteorological grids should be interpreted primarily as an intermediate representation of spatially varying forcing fields (and their uncertainty), rather than as an implicit definition of hydrologic model resolution (Ledesma & Futter, 2017). Gridded forcing is used to represent spatial structure before aggregation to GUs.

Model evaluation integrates multiple independent data sources. Streamflow observations are used for calibration, while MODIS snow-cover products (Hall & Riggs, 2021; Hall et al., 2006; Parajka & Blöschl, 2006), Copernicus soil moisture (Das et al., 2019), and two field-scale CRNS observations installed and calibrated within the domain (Franz et al., 2020; Katya et al., 2020) support process-based validation across scales. The analysis focuses on a period characterized by multiple high-flow extreme events and an exceptionally long drought with persistent low-flow conditions (Avanzi et al., 2024; Cavaleri et al., 2019; Ginesta et al., 2023; Toreti et al., 2022). Accordingly, the thesis follows a structured workflow: model definition, compilation of available datasets, calibration, experimental design, and multi-dataset validation of several hydrologic compartments.

1.4.3 Modeling framework and software contributions

To address these questions in a reproducible and systematic manner, this thesis relies on a modular modeling framework that enables controlled experiments across spatial configurations and forcing strategies. This thesis investigates scale-related issues—from the intrinsic model scale to the scale of physical processes—using GEOframe (Abera et al., 2017; Formetta et al., 2014a) as both a modeling framework and an experimental environment. GEOframe is a framework based on the Object Modeling System (OMS) that supports modular and reproducible environmental modeling workflows (David et al., 2013, 2016). In

this thesis, GEOframe provides the computational environment in which alternative spatial configurations and meteorological forcing strategies can be implemented and compared.

Geomorphological preprocessing and drainage-network extraction are performed with Horton Machine (Rigon et al., 2006), which supplies terrain-derived information to delineate computational units and define the routing structure. The modeling chain is then executed within GEOframe, including geostatistical interpolation (kriging) and the radiation and evapotranspiration modules (Abera et al., 2014; Bancheri et al., 2018; Bottazzi et al., 2021; Formetta et al., 2016), together with the rainfall-runoff model originally developed by Bancheri (2017). GEOframe is especially suited for mountainous Alpine environments thanks to its modular design and its ability to integrate multiple data sources across different spatial and temporal scales. For example, Abera et al. (2017) applied the model provided by GEOframe to estimate water budget components in pre-Alpine basins, highlighting the importance of accounting for spatial variability in precipitation and evapotranspiration. Morlot et al. (2023) applied GEOframe for modeling snow, soil moisture, and river discharge evolution of the Adige River basin, considering also artificial reservoirs and human influence in the water budget. Arnone et al. (2023) used GEOframe to simulate peak flow generation in a small mountainous catchment, incorporating high temporal resolution meteorological data necessary to capture intense storms. More recently, Azimi et al. (2023) improved upon these methods by integrating enhanced snow modeling techniques and validating results with remote sensing data and proposing uncertainty estimation tools.

During this work, several components were improved and new modules were developed. To support the scientific analyses, the thesis delivers two substantial software contributions: i) the kriging workflow was extensively refactored and redesigned to improve maintainability, extensibility, and code reuse (i.e., following widely adopted object-oriented design principles) (Sarcar, 2022). This redesign enables full automation of the interpolation pipeline, including the automatic estimation and fitting of a theoretical semivariogram at each time step and the end-to-end execution of the time-stepped workflow; ii) the hydrologic

model was refactored and stabilized through targeted bug fixes, and the numerical core was upgraded by rewriting the adaptive Runge–Kutta solver used for ordinary differential equation (ODE) integration (Kavetski & Clark, 2010). These developments support reproducible experimentation and are freely available in the repository of the GEOframe group (geoframecomponents, n.d.).

Beyond the two components that were redesigned for the experiments of this thesis, the work carried out during the doctorate contributed to a broader refactoring of the GEOframe system. Several algorithms were rewritten, a number of long-standing bugs were identified and fixed, and the code base was consolidated: the Git repositories of the GEOframe organization and the project website were reorganized so that components, documentation, and examples can be located and cited unambiguously, and the build system was modernized to make compilation and dependency resolution reproducible across components. Within the interpolation chain, the efficiency of the kriging component was substantially improved, which is what makes hourly, time-step-wise semivariogram estimation over multi-year simulations affordable in practice. Part of these developments were carried out in support of, and were tested within, a project involving the Po River Basin Authority (AdbPo), so that the same code base that produces the results of this thesis is also in use in an operational, third-party context.

The whole modeling infrastructure was built following the FAIR principles—findable, accessible, interoperable, and reusable (Wilkinson et al., 2016)—and, more specifically, according to the blueprint for participatory Digital eARth Twin Hydrology systems (DARTHs) proposed by Rigon et al. (2022); the ideas behind it, and their gradual refinement, are also documented in the group's open notebook (Rigon, n.d.). In practice this means that models are treated as composable, documented, and citable software components rather than as monolithic executables, and that a simulation is considered a scientific result only when the reader can retrieve its inputs, re-run it, and obtain the same outputs. Accordingly, the input datasets that can be redistributed, together with the simulation workflows (OMS3

simulation files, parameter files, and scripts used for pre- and post-processing) and the model outputs presented in this thesis, are published in open repositories, as detailed in Appendix F. Some input datasets could not be made publicly available because of licensing or access restrictions imposed by the respective data providers. A direct consequence of this approach is that the modeling chain did not remain an experimental prototype: by the end of this work, it had become operational in the Noce River basin, running with current meteorological forcing and producing water-budget estimates that can support water-management activities.

1.4.4 Model structure

Parameters are calibrated at the sub-catchment scale: all GUs upstream of a given stream gauge share the same parameter set under a common calibration strategy, including sequential calibration approaches (Nkiaka et al., 2018). In contrast, each GU is represented by spatially uniform GU-specific forcings and descriptors: meteorological inputs (precipitation, air temperature, radiation) and derived variables (e.g., potential evapotranspiration) are expressed as a single representative value (or time series) per GU. Inside each GU, the model is composed of several compartments to represent key hydrological processes (canopy, snow, root-zone, runoff, and groundwater). We use the term GU for consistency with common practice; however, our units are geomorphologically delineated and parametrically lumped, and therefore do not coincide with the physiographic concept, the Hydrological Response Unit (HRU), originally formulated by Flügel (1995).

GUs should be viewed as scale-dependent statistical filters rather than physically unique units, and semi-distributed conceptual models are expected to exhibit structural scale dependence and pronounced equifinality. This raises a key question for the present work: to what extent can such models provide scale-consistent simulations and parameter sets, and how robust are their predictions to the chosen discretizations based on contributing-area thresholds?

1.4.5 Thesis outline

The thesis comprises three studies, each intended to form the basis of a journal article. It is organized so that the methodological framework and material common to the three studies are presented only once, while each study is developed as a self-contained scientific contribution that builds on this shared foundation.

- **Chapter 2** – *The Noce River basin: study area, monitoring network, and data*. This chapter describes the hydro-climatic setting of the basin and the observational datasets used throughout the thesis.

- **Chapter 3** – *Methodological framework*. This chapter presents the methods shared across the three studies, including spatial discretization, meteorological forcing, hydrological modeling, calibration, and multi-source model evaluation.

- **Chapter 4** – *Software development*. This chapter describes the principal software developments produced during the doctorate and the practices adopted to support reproducibility and reuse.

- **Chapter 5** – *Mixed-resolution analysis of the water budget of an Alpine basin* (Study 1). This study compares constant- and mixed-resolution configurations and evaluates their effects on the simulated water budget.

- **Chapter 6** – *Effects of spatial representation of meteorological data on different hydrological processes in semi-distributed hydrological models* (Study 2). This study evaluates alternative representations of meteorological forcing and their effects on simulated hydrological processes.

- **Chapter 7** – *Assessing spatial discretization impacts on semi-distributed hydrological simulations* (Study 3). This study examines how model spatial resolution affects streamflow performance and simulated internal states.

- **Chapter 8** – *General synthesis and future perspectives*. This chapter integrates the findings of the three studies, discusses their limitations and operational implications, and identifies future research directions.

The appendices are global rather than chapter-specific: Appendix A collects the complete formulation of the hydrological model, Appendix B the geostatistical formulation and the semivariogram families, Appendix C the physiographic and climatic characterization of the basin, Appendix D the supplementary tables and figures of the three studies, Appendix E the computational-cost tables, and Appendix F the open-research statement with the location of data, code, and workflows.

CHAPTER 2. THE NOCE RIVER BASIN: STUDY AREA, MONITORING NETWORK, AND DATA

The three studies that constitute the core of this thesis (Chapters 5–7) share the same region of interest and, to a large extent, the same observational basis. However, they consider different portions of the basin and cover different simulation periods. To avoid repetition, this chapter provides a common description of the basin, its monitoring system, and the observational datasets. The individual studies refer to this chapter and specify their respective study domains and simulation periods.

2.1 Study area: the Noce River basin

The study area encompasses the entire Noce River basin, located in the Province of Trento (Trentino), northeastern Italy, at the southern edge of the Alps. From upstream to downstream, the basin extends across two major valleys, Val di Sole and Val di Non. The Noce River eventually flows into the Adige River, which has the third-largest drainage basin in Italy (Figure 2.1). Owing to its extent and abundant water resources, the Noce basin plays a major role in the hydrology of Trentino and is central to regional water management.

The basin exhibits characteristic Alpine features and a pronounced elevation gradient, ranging from 210 m asl near the outlet at Mezzolombardo to 3769 m asl at the Cevedale peak. It covers a total area of 1367 km^2 and has a mean elevation of 1624 m asl. Mean annual precipitation is approximately 1170 mm/year, whereas mean annual temperature is 3.9 °C (Bellin et al., 2016).

Because of this broad elevation range, climatic conditions vary substantially across the basin. The upper basin has an Alpine climate, represented by the Presena station, where the mean annual temperature is −1.4 °C and monthly mean temperatures range from −9.6 °C in winter to 8.9 °C in summer. These conditions result in long, cold winters, substantial

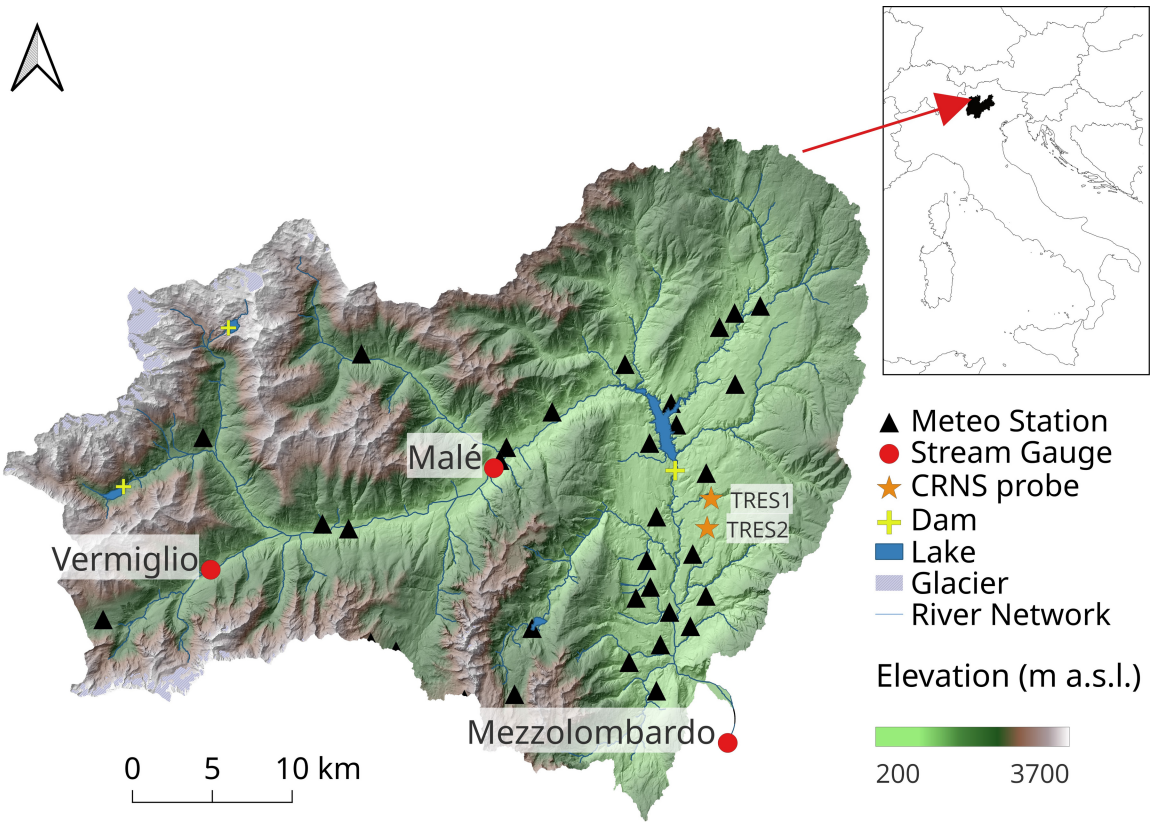

Figure 2.1 – The Noce River basin, showing the river network, lakes, dams, stream gauges, meteorological stations, and CRNS probes. Elevation is represented by a color ramp ranging from green at lower elevations to white at higher elevations. A hillshade overlay is used to enhance the representation of the terrain.

seasonal snow accumulation, and short, cool summers. Several small glaciers are also present at high elevations. At intermediate elevations, the Cles station represents subalpine conditions, with a mean annual temperature of approximately 8.5 °C, cold winters with moderate snowfall, and cool to mild summers. Near the basin outlet, Mezzolombardo is characterized by a subcontinental climate, with a mean annual temperature of approximately 12.6 °C, cold winters with occasional snowfall, and warm to hot summers, during which monthly mean temperatures can reach 24.8 °C. Appendix Table C1 summarizes temperature statistics for these three stations, which represent the main elevation zones of the basin.

The hydrological regime also varies considerably from upstream to downstream because of both the basin's physiographic and climatic heterogeneity and the spatially uneven influence of human water use and regulation. In its upper part, runoff is strongly influenced

by seasonal snowmelt and glacier melt from the Ortles–Cevedale and Adamello–Presanella mountain groups. The karstic Brenta Massif introduces additional complexity into subsurface water storage and drainage processes. Further details on the geology, soils, and vegetation of the Noce basin are provided in Appendix C.

In the lower part of the basin, the flow regime is increasingly affected by water abstractions, hydropower infrastructure, and reservoir operations. Indeed, the basin's abundant water resources have supported extensive hydropower development. Numerous dams constructed during the twentieth century have substantially altered the natural flow regime (Majone et al., 2016). The main hydropower system has a nominal capacity of 113.9 MW (Majone et al., 2016). In addition, approximately 100 small hydropower plants, each with an installed capacity below 3 MW, contribute a combined capacity of approximately 21 MW (Bellin et al., 2016). The Santa Giustina Dam, the largest in the region, plays a central role in flow regulation and energy production (Maran et al., 2023; Stucchi et al., 2023). Its reservoir has a storage capacity of 183×10^6 m^3, and the associated hydropower plant has a nominal output of approximately 40 MW (Bellin et al., 2016; Majone et al., 2016; Stucchi et al., 2023).

Agriculture is another major activity, particularly in the lower basin, where more than 3000 farms and more than 6000 hectares of cultivated land make a substantial contribution to the local economy and employment (ISPAT – Istituto di Statistica della Provincia Autonoma di Trento, n.d.). Apple orchards are the dominant crop, whereas pastures are common at higher elevations. Most orchards are equipped with drip-irrigation systems, and agricultural districts depend on both surface water and groundwater to support crop production and livestock farming.

Irrigation demand is particularly high during summer because of increased evapotranspiration and crop water requirements. Many orchards do not directly benefit from upstream snowmelt or glacier melt and therefore depend on irrigation consortia, which operate artificial storage ponds to mitigate seasonal water shortages (Zambotti, 2020). During severe

droughts, water may also need to be pumped from lower sources, including the Santa Giustina reservoir. The allocation and management of water rights are therefore essential for meeting agricultural demand while balancing competing water uses (Zambotti, 2020).

The basin is also a major tourist destination, attracting visitors for both winter and summer recreation. Seasonal increases in population further intensify freshwater demand for domestic and recreational purposes (Favargiotti et al., 2022).

2.1.1 The upper basin: the Malé and Vermiglio sub-catchments

Chapters 6 and 7 focus on the upper part of the Noce basin, upstream of the Malé stream gauge (Figure 2.2). This delineation excludes the highly regulated lower reaches and the Santa Giustina reservoir, allowing the mountain hydrology of the basin to be investigated with reduced influence from major reservoir operations.

The resulting catchment covers approximately 463 km^2, with elevations ranging from 684 m to 3769 m asl and a mean elevation of approximately 2000 m asl. Its headwaters include glacierized areas in the Adamello–Presanella and Ortles–Cevedale mountain groups, while seasonal snow accumulation at high elevations contributes substantially to runoff during the warm season. Together with the pronounced elevation gradient, these features result in marked spatial and seasonal variability in the hydrological response. Compared with the characterization of the entire Noce basin presented above, the climatic conditions of the upper basin are represented by the Presena and Cles stations, whereas Mezzolombardo, located in the lower part of the Noce basin, falls outside the study domain.

The study area is located immediately upstream of the Santa Giustina reservoir and the intensively cultivated lower valley, where apple orchards represent the dominant agricultural activity and irrigation constitutes an important water demand. Consequently, runoff generated in the upper basin is relevant not only to the local hydrological regime but also to downstream water management, particularly during periods of water scarcity. Although the major downstream regulation is excluded from the study domain, the upper basin is itself

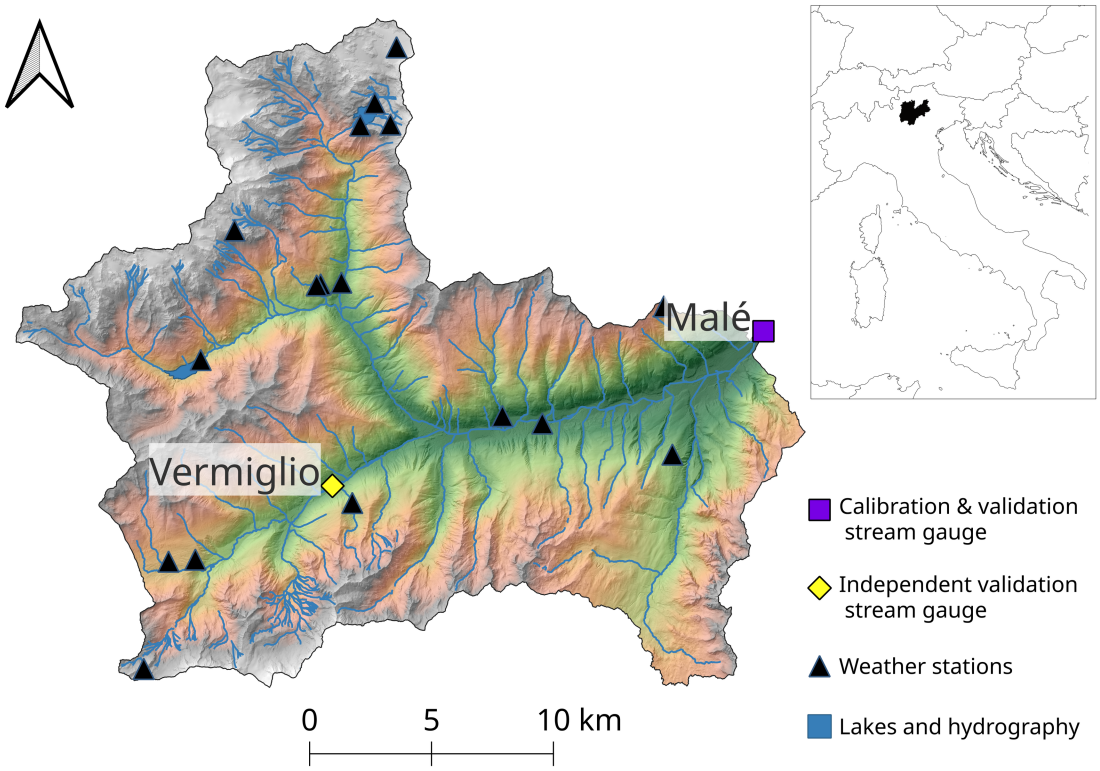

Figure 2.2 – Map of the study basin showing the river network, lakes, meteorological stations located within the basin, and stream gauges. Stream gauges used for calibration and validation are represented by purple squares, whereas gauges used exclusively for independent validation are represented by yellow diamonds. The inset shows the location of the study area within Italy and the Province of Trento.

affected by several small hydropower structures and water-withdrawal intakes (Bellin et al., 2016; Majone et al., 2016).

A more detailed physiographic and climatic characterization of the study area is provided in Appendix C.

2.2 Monitoring network and observational data

The Noce basin is well instrumented, with extensive meteorological and hydrometric monitoring covering a broad range of elevations and environmental conditions. Ground-based observations were complemented by remote-sensing and reanalysis products to evaluate hydrological compartments beyond streamflow and to extend the model evaluation

over larger spatial domains. However, not all observational records span the full simulation periods, and some contain internal data gaps.

Meteorological observations were obtained from MeteoTrentino, operated within the Civil Protection system of the Province of Trento (MeteoTrentino, n.d.), and from Fondazione Edmund Mach (Fondazione Edmund Mach, 2022). Both monitoring networks have expanded since the 1990s and together provide broad spatial and elevation coverage of the Noce basin. Overall, 49 stations record both precipitation and temperature, while 5 additional stations record temperature only. The observations are available at hourly or sub-hourly temporal resolutions. To ensure consistency among the studies, all meteorological time series were standardized to an hourly time step, with temperature averaged over each hour and precipitation accumulated hourly.

Streamflow data were provided by the Civil Protection of Trentino, Ufficio Dighe (Ufficio Dighe, n.d.), which operates three stream gauge stations within the basin. Listed from upstream to downstream, these gauges are Vermiglio, Malé, and Mezzolombardo. They provide discharge observations at an hourly resolution.

In addition to conventional hydro-meteorological observations, the analysis of the entire basin used hourly soil-moisture observations from two Cosmic-Ray Neutron Sensing (CRNS) probes. The sensors were installed through a collaboration with Fondazione Edmund Mach, with support from the Tres Consortium and Rete2000. They are located in Val di Non, within the area managed by the Tres Consortium (Figure 2.1), and use the FINapp F3 sensor model (Gianessi et al., 2024; Stevanato et al., 2019). The two probes, designated TRES1 and TRES2, provide spatially integrated estimates of soil moisture based on the cosmic-ray neutron sensing method introduced by Zreda et al. (2008). These observations were used to evaluate the simulated dynamics of soil-water storage. This study marks the first use of these sensors in the region for model evaluation of soil moisture storage dynamics.

Remote-sensing products were used to evaluate the spatial and temporal dynamics of

hydrological variables that cannot be adequately characterized using streamflow observations alone.

Snow-cover dynamics were assessed using daily maps derived from the MODIS Terra and MODIS Aqua products, which have a spatial resolution of 500 m (Hall & Riggs, 2021; Hall et al., 2006). Observations affected by cloud cover were excluded. In all three studies, the resulting cloud-free records were used for qualitative and quantitative evaluation of the seasonal evolution of snow accumulation and melt.

In the first study, soil-moisture variability was assessed using the Copernicus Surface Soil Moisture (SSM) product because adequate observations were available over the areas considered in the analysis (Das et al., 2019). This product has a spatial resolution of approximately 1 km (1°/112) and a temporal sampling interval ranging from approximately 1.4 to 4 days.

By contrast, Copernicus SSM could not be used in the two studies focusing on the upper basin because of the complex topography and higher elevations of the study area. Satellite-derived surface-soil-moisture products provide inadequate spatial coverage in these areas, with observations generally unavailable or unreliable above approximately 1900 m asl (Brocca et al., 2013). To overcome this limitation, soil-moisture estimates were obtained from the MEteorological Reanalysis Italian DAtaSet (MERIDA) at depths of 5 cm, 25 cm, 70 cm, and 150 cm (Bonanno et al., 2019). MERIDA is a high-resolution reanalysis dataset obtained by dynamically downscaling the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) with the Advanced Research core of the Weather Research and Forecasting model (WRF-ARW) and has a horizontal resolution of 0.07°, corresponding to approximately 7 km.

2.3 Simulation periods

The three studies use two overlapping simulation periods, determined primarily by data availability. The shorter period used in the first study is entirely contained within the longer

period adopted for the two studies focusing on the upper basin.

The first study covers the period from 1 January 2017 to 15 September 2022. Most observational datasets were available over this period, whereas the CRNS observations only became available at the end of May 2022.

The two studies focusing on the upper basin cover the longer period from 1 August 2015 to 31 December 2023.

Both simulation periods include Storm Vaia in October 2018 (Cavaleri et al., 2019), Storm Alex in October 2020 (Cavaleri et al., 2022; Ginesta et al., 2023), and major drought conditions in 2019 and 2022 (Avanzi et al., 2024; Toreti et al., 2022). They therefore encompass contrasting hydrological conditions, including major high-flow events and the exceptional snow-drought and low-flow conditions of 2022. This temporal coverage allows the model configurations to be evaluated under a broad range of hydrological conditions rather than only during average years.

CHAPTER 3. METHODOLOGICAL FRAMEWORK

This chapter presents the methodological framework and analysis procedures adopted throughout the thesis, following the methodological sequence outlined in Section 1.4.2. The modeling experiments and software developments presented in this thesis were carried out within GEOframe (Abera et al., 2014, 2017; Formetta et al., 2014a), an open-source environment comprising modular and interoperable components for a wide range of hydrological applications. Built on the Object Modeling System version 3 (OMS3) (David et al., 2013, 2016), GEOframe enables atomic modules to be combined into tailored workflows, facilitating component reuse and adaptation to specific modeling objectives.

Geomorphological preprocessing and drainage-network extraction are performed with Horton Machine (Rigon et al., 2006). The resulting terrain-derived information is used to delineate the computational units and define the routing structure, thereby providing the spatial basis for the subsequent simulations. The GEOframe modeling chain combines geostatistical interpolation of meteorological forcing through kriging, radiation and evapotranspiration modules (Abera et al., 2014; Bancheri et al., 2018; Bottazzi et al., 2021; Formetta et al., 2016), and the rainfall-runoff model formulation originally developed by Bancheri (2017).

Previous applications demonstrate the suitability of this framework for hydrological analyses in mountainous environments and across different spatial and temporal scales. Abera et al. (2017) used GEOframe to estimate water-budget components in pre-Alpine basins, emphasizing the importance of accounting for the spatial variability of precipitation and evapotranspiration. At a larger spatial scale, Morlot et al. (2023) simulated the evolution of snow, soil moisture, and river discharge across the Adige River basin while also accounting for reservoirs and anthropogenic influences on the water budget. Arnone et al. (2023) applied GEOframe to simulate peak-flow generation in a small mountainous catchment using high-

temporal-resolution meteorological data to represent intense storm events. Finally, Azimi et al. (2023) extended these applications by integrating enhanced snow-modeling methods, evaluating the simulations against remote-sensing observations, and introducing tools for uncertainty estimation.

The GEOframe modeling framework and Horton Machine toolkit introduced above were jointly used to implement the hydrological modeling workflow for the Noce River basin, with Horton Machine supporting geomorphological preprocessing and GEOframe providing the hydrological simulation components. The resulting model outputs were evaluated against multiple observational sources, including in situ streamflow measurements and remote-sensing products. Accordingly, the following sections describe the spatial discretization of the basin into Geomorphic Units (GUs) (Section 3.1), the estimation of meteorological forcing at the GU scale (Section 3.2), the GEOframe semi-distributed hydrological model (Section 3.3), the calibration and validation framework (Section 3.4), the uncertainty analysis (Section 3.5), and the multi-source evaluation of model outputs (Section 3.6).

In summary, this chapter focuses on the methodological principles shared by the three studies, whereas study-specific configurations, parameterization choices, and calibration and validation periods are described in the respective *Experimental setup* subsection of the corresponding chapter. Software components developed or extended within this thesis are described separately in Chapter 4.

3.1 Partition of the basin into Geomorphic Units

This thesis uses a semi-distributed hydrological model in which the basin is represented as an interconnected, organized collection of GUs. Each GU corresponds to a non-overlapping, spatially explicit polygon; together, the GUs form a complete tessellation of the basin. Each GU is treated as an elementary control volume characterized, at each time step, by one representative value for each model state and flux (Formetta et al., 2014a, 2014b). Accordingly, each input variable, such as temperature, precipitation, and poten-

tial evapotranspiration, is represented by one time series per GU. Model states and fluxes are likewise computed separately for each GU. The topological connections among GUs reproduce the drainage network and determine the downstream aggregation of simulated fluxes (Abera et al., 2014; Formetta et al., 2014a). The river network is represented by GEOframe as an acyclic tree graph. Several methods can be used to partition a basin. Three families of GU partitions are considered throughout the thesis:

1. **stream-gauge-based (SG)** partitions, in which gauge locations define the GU outlets and the basin is divided into the corresponding non-overlapping incremental contributing areas, without imposing an additional geomorphological area threshold. This is the coarsest semi-distributed representation considered in Chapter 6;

2. **constant-threshold** partitions, in which a single contributing-area threshold is applied over the whole domain, producing units of comparable characteristic area (Chapters 6 and 7);

3. **mixed-resolution** partitions, in which a coarse basin-wide threshold is locally refined in selected sub-regions (Chapter 5).

Because SG units are defined directly by stream-gauge locations, the following subsections provide additional methodological details and references for the other two delineation approaches.

3.1.1 Constant-threshold delineations

Constant-threshold delineation is performed using the Horton Machine toolkit (Rigon et al., 2006), which provides a set of utilities for geomorphological and hydrological analyses. Specifically, the Horton Machine is used to extract the drainage network and delineate the corresponding contributing areas from a Digital Elevation Model (DEM). GU delineation is controlled by a contributing-area threshold. Due to the geomorphological constraints of the

delineation algorithm, the resulting GU areas do not exactly match the nominal contributing-area threshold. Therefore, a tolerance of $\pm 20\%$ relative to the nominal threshold was allowed during the delineation procedure.

3.1.2 Mixed-resolution (hybrid) delineation

The mixed-resolution strategy allows the spatial discretization to be refined only in selected parts of the basin while retaining a coarser representation elsewhere. Starting from a basin-wide delineation obtained with a fixed contributing-area threshold, selected GUs are subdivided using a smaller threshold. This approach increases spatial detail where finer representation is required without imposing the same computational resolution over the entire catchment. The hydrological connectivity between the resulting GUs is preserved through the river-network graph. As the standard GEOframe delineation chain supports only a single contributing-area threshold, a dedicated software component was developed to generate mixed-resolution configurations, as described in Section 4.2.

3.2 Meteorological forcing at the Geomorphic Unit scale

Each GU requires one representative time series for each meteorological input variable. Therefore, point observations or gridded data must be transferred to the GU scale. Three strategies are considered for temperature and precipitation: i) NEAR (Nearest weather station-Forcing Evaluation), which assigns to each GU the observations from the meteorological station closest to its centroid; ii) CEN (Centroid-based Forcing Evaluation), which interpolates each variable at the GU centroid using a geostatistical method and assumes that the centroid estimate represents the entire GU; and iii) GRIDW (Grid-weighted Forcing Evaluation), which represents each GU using the area-weighted mean of the gridded values associated with the cells intersecting the GU. The grid may originate from an existing product or from the interpolation of ground observations.

3.2.1 Geostatistical interpolation and automatic semivariogram estimation

Whereas the NEAR method directly assigns observations from the nearest station to each GU centroid, the CEN and GRIDW approaches require estimates at locations where no measurements are available. These estimates were obtained using kriging, a geostatistical interpolation technique based on the concept of semivariance, which quantifies spatial variability (autocorrelation) between pairs of observations (Equation 3.1).

$$\gamma(h) = \frac{1}{2}\,\mathbb{E}\Big[\big(Z(s) - Z(s+h)\big)^2\Big] \tag{3.1}$$

where $\gamma(h)$ is the semivariance, h is the lag or separation distance, $Z(s)$ is the value of the variable at location s, and $\mathbb{E}[\cdot]$ denotes the expectation operator. In practice, semivariances are calculated for all available station pairs, grouped into lag-distance bins, and averaged within each bin to obtain an experimental semivariogram. A parametric semivariogram model is then fitted to these experimental estimates and used to determine the kriging weights for prediction at unsampled locations.

For bounded semivariogram models, spatial dependence is commonly described using the nugget, sill, and range. The nugget represents the discontinuity between the zero semivariance at the origin and the limiting semivariance at an infinitesimally small non-zero lag; it may reflect measurement error and variability below the sampling resolution. The total sill is the plateau approached or reached at large distances, whereas the range describes the characteristic distance over which spatial dependence persists. For models that approach the sill asymptotically, such as the exponential and Gaussian models, a practical range may be defined as the distance at which a specified proportion of the sill, commonly 95%, is reached. Linear and power models may be unbounded and therefore do not necessarily have a finite sill or range (Camana & Deutsch, 2019; Ly et al., 2013; Skøien & Blöschl, 2006). Beyond the range distance, spatial autocorrelation is no longer present.

A useful metric for assessing spatial dependence, according to the definition proposed by

Cambardella et al. (1994), is the nugget-to-sill ratio (NSR), defined as the ratio between the nugget and the total sill: values below 0.25 indicate strong dependence, values between 0.25 and 0.75 moderate dependence, and values above 0.75 weak dependence. These thresholds were used as heuristic interpretive criteria rather than universal limits.

Since a large number of variograms need to be computed—potentially one for each time step—an automated workflow was implemented within the GEOframe framework and is illustrated in the use-case and activity diagrams (Figures 4.2 and 4.1). The automated workflow comprises three stages: i) compute the experimental variogram from each time step and station network and fit a theoretical model; ii) assess the goodness-of-fit of the fitted model; and iii) predict values at unsampled locations via kriging.

Here, the first step—the estimation of the theoretical variogram—is described in greater detail, as it represents one of the new features introduced in the software. For each time step, an experimental variogram is computed and then fitted to obtain a theoretical model. If detrending is requested, the elevation-dependent trend is removed if and only if a significant (linear) relationship with elevation is detected. At this stage, the number of bins can be specified, which defines the distance classes. For each bin, the midpoint represents the distance, while the semivariance corresponds to the average of all pairwise semivariance values falling within that class. Several candidate models are tested for fitting, including linear, exponential, power, spherical, and Gaussian functions. Model fitting is performed using the Apache Commons Math 3.6.1 library (The Apache Software Foundation, 2023), which provides a robust nonlinear least-squares solver based on the Levenberg–Marquardt algorithm (Watanabe et al., 2015). In this evaluation all bins of the experimental variogram have equal weights to ensure that each class contributes equally to the fit. A variogram is accepted if its relative error is below a predefined threshold, and the model with the minimum relative error is selected. When none of the candidate models satisfied the fitting criterion for a given time step, a global semivariogram was used as a fallback. This global model was fitted using semivariance estimates aggregated over the full time series and applying the

same model-selection procedure used for the time-specific semivariograms. The output, for each time step, contains the minimum information required to reconstruct the variogram in the kriging module: nugget, range, sill, type of variogram, whether detrending was applied, whether a backup value was used, whether a logarithmic transformation was applied and—if detrended—the intercept and slope of the linear interpolation. This automated procedure ensures consistent variogram estimation across time steps, enabling the dynamic interpolation of meteorological variables. After this step, the output is used as input in the kriging module. The module can operate either as ordinary kriging (OK), which represents spatial dependence using the planar coordinates and the fitted semivariogram, or as detrended kriging (DK), in which a significant elevation-dependent trend is first removed, the residuals are kriged, and the estimated trend is subsequently restored at each prediction location. The module also supports localized prediction, in which only the nearest stations are included in the kriging system for a given prediction location, although the semivariogram is estimated using all available stations. In this thesis, DK predictions were obtained using the 10 nearest stations.

The complete geostatistical formulation—the kriging system variants, and the five semivariogram families that the automatic procedure can select from—is reported in Appendix B; the software architecture that implements it is described in Section 4.3.

3.2.2 Validation of the interpolated fields

The predictive performance of the interpolated meteorological fields was assessed using the Leave-One-Out Cross-Validation (LOOCV) method (Aidoo et al., 2015; Bancheri et al., 2018; Efron, 1982; Isaaks & Srivastava, 1989; Martin & Simpson, 2003). LOOCV is a well-established approach in geostatistics and has been extensively employed for validating temperature and precipitation interpolation methods (Bancheri et al., 2018; Benavides et al., 2007; Brunetti et al., 2014; Jewell & Gaussiat, 2015; Ly et al., 2013). This procedure consists of iteratively running the kriging, each time removing one sampling point from

the predictor dataset and estimating its time series values. The goodness of fit was then evaluated by comparing the estimated output with the observed values. The most common performance metrics were calculated, including Pearson correlation coefficient (r), the coefficient of determination (R^2), the root mean square error (RMSE), the mean absolute error (MAE), and the bias as the mean error. In addition, two other indices are evaluated as complementary summary metrics of interpolation performance: the Nash–Sutcliffe efficiency index (NSE) (Nash & Sutcliffe, 1970) and the Kling–Gupta efficiency index (KGE) (Gupta et al., 2009). This index simultaneously accounts for correlation, bias, and variability, providing a more balanced assessment of interpolation performance (Kling et al., 2012; Towner et al., 2019). For this work, the theoretical semivariogram was evaluated at each time step, and a LOOCV was performed for both variables—temperature and precipitation—using the DK and localized prediction model options. In this context, the sampling points correspond to the weather stations within the basin.

After assessing interpolation performance through LOOCV, the final temperature and precipitation fields were generated using the selected semivariogram models, detrended kriging, and localized prediction based on the 10 nearest stations. In the CEN approach, values were interpolated at the GU centroids, directly providing the representative value for each GU. In contrast, in the GRIDW approach, interpolation was first performed at the grid-cell level and then aggregated using an area-weighted mean for each GU. Similarly, shortwave and longwave radiation were estimated following Formetta et al. (2016), and potential evapotranspiration was computed using the Priestley–Taylor method (Bottazzi et al., 2021; Priestley & Taylor, 1972). For the NEAR method, as described above, the meteorological forcing was assigned using the values measured at the station closest to each GU centroid. In this case, the precipitation Areal Reduction Factor (ARF) was fixed at 1, so that the precipitation measured at the nearest station was transferred to the GU without attenuation. ARFs are often related to the characteristics of individual precipitation events or defined for specific design events. Because GEOframe is applied here as a continuous

hydrological model over a multi-year period, using a single ARF for the entire precipitation series would constitute an arbitrary choice, whereas event-specific ARFs would require the identification and separate characterization of individual events. Setting ARF = 1 also keeps the approach simple and transparent, isolates the effect of using the nearest observations rather than spatially interpolated precipitation fields, and is consistent with the decision not to alter precipitation through the rainfall and snowfall correction factors ($\alpha_r = \alpha_s = 1$; Appendix A).

To diagnose the effects of the alternative precipitation fields, a moving annual runoff coefficient was calculated over a one-year rolling window. Observed discharge was integrated over each window and divided by the upstream catchment area to express runoff as an equivalent water depth. The runoff coefficient was then calculated as the ratio between this runoff depth and the basin-mean precipitation accumulated over the same window as in Equation 3.2:

$$RC = \frac{Q_{\text{annual}}}{P_{\text{annual}}} \tag{3.2}$$

where Q_{annual} and P_{annual} are, respectively, the observed runoff depth and basin-mean precipitation depth accumulated over the same one-year window, both derived from daily sums. Because the same observed runoff series was used for all forcing configurations, differences in RC reflect only differences among the precipitation estimates.

3.3 The GEOframe semi-distributed hydrological model

The GEOframe hydrological modeling system (Bancheri, 2017; Formetta et al., 2014a; Morlot et al., 2023) enables comprehensive simulation of the hydrological cycle by dynamically integrating different model components. Its modular structure allows flexible configurations tailored to specific study objectives.

For each GU, precipitation and temperature are assigned using the forcing strategy

described in Section 3.2. Shortwave and longwave radiation are estimated using the GEOframe radiation tools (Formetta et al., 2016), and potential evapotranspiration (PET, ET_p) is calculated using the Priestley–Taylor method (Bottazzi et al., 2021; Priestley & Taylor, 1972).

The hydrological structure adopted throughout the thesis follows that proposed by Bancheri (2017) and is illustrated in Figure 3.1 using the Extended Petri Net methodology (Bancheri et al., 2019). In this representation, state variables—liquid water in snow (S_w), frozen water in snow (S_i), water stored in the canopy (S_c), water content in the root zone (S_{rz}), water stored in the runoff compartment (S_r), and groundwater storage (S_{gw})—are depicted as circles (places). Fluxes—the rain (P_r) and snow (P_s) fractions of precipitation, freezing (F) and melting (M), canopy throughfall (T_c), actual evapotranspiration from the canopy (ET_c), non-infiltrated water (R), groundwater recharge (R_g), actual evapotranspiration from the root zone (ET_{rz}), and the runoff (Q_r) and groundwater (Q_g) discharge components—are shown as squares (transitions). Dark-bordered triangles represent external forcing variables, namely observed precipitation (P), temperature (T), and potential evapotranspiration (ET_p).

The model includes five main compartments, each representing a specific hydrological domain and color-coded in Figure 3.1 for clarity. Precipitation (P) enters the snow module, where rain–snow partitioning is performed and both rainfall and snowfall are corrected using phase-specific precipitation correction factors accounting for gauge undercatch (Formetta et al., 2014b). These processes are coupled with a temperature-index approach that accounts for both melting and freezing (Asaoka & Kominami, 2013; Formetta et al., 2014b; Hock, 2003). Meltwater and rainfall then enter the canopy compartment, where plant interception is represented through a small nonlinear reservoir with two outputs: evapotranspiration and throughfall (Aston, 1979; Rutter et al., 1971).

Throughfall enters the root zone, where it is partitioned between infiltration and surface runoff following the approach described by Moore (1985). Infiltrated water is stored in a

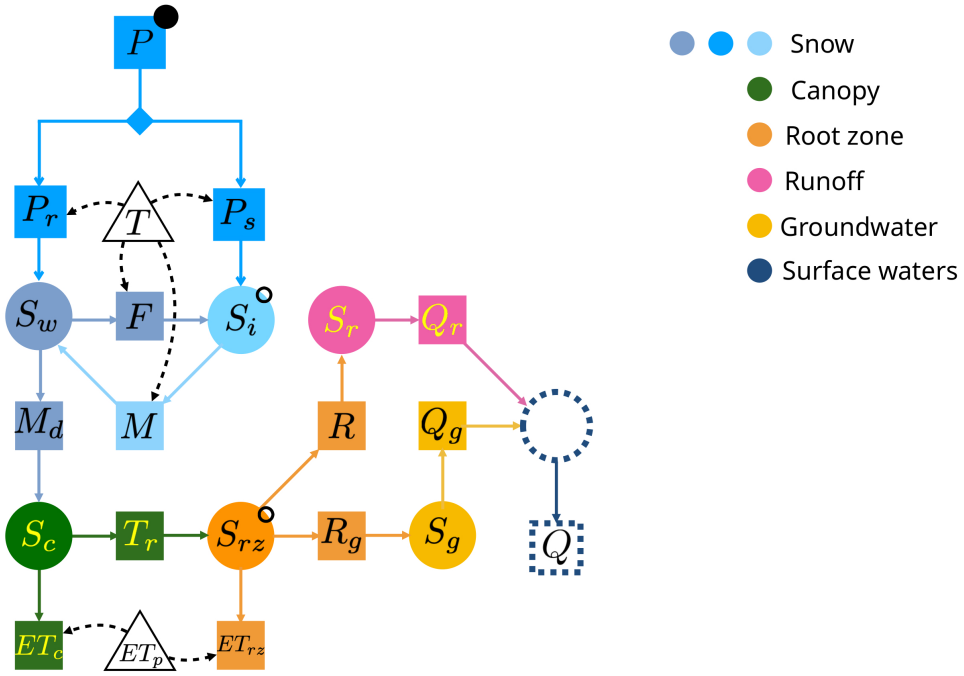

Figure 3.1 – Extended Petri Net representation of the hydrological model. Rain–snow partitioning and the temperature-index snow processes are shown in light blue; the canopy compartment in green; the root-zone compartment in orange; the runoff compartment in purple; the groundwater compartment in yellow; and the sum of the runoff and groundwater discharge components in dark blue.

nonlinear reservoir that produces two outflows: actual evapotranspiration and groundwater recharge. Water routed to the runoff and groundwater compartments ultimately generates the two discharge components at the outlet. Total actual evapotranspiration is calculated as the sum of the canopy and root-zone contributions. Some state variables are marked with a dark ring in Figure 3.1 to indicate their use in model validation. Detailed definitions of the state variables, fluxes, parameters, and governing equations are provided in Appendix A.

Within each controlled comparison, the hydrological model structure was held constant so that differences among configurations could be interpreted primarily in relation to the methodological factor being varied, such as spatial discretization or forcing representation. For every spatial configuration, the same hydrological model structure is applied independently to each GU, irrespective of its spatial scale, using the corresponding calibration parameters. The only exception is represented by GUs containing dam reservoir areas, for which a specific dam module is used. All GUs, including those characterized by different spatial scales (e.g., mixed-scale cases), are simulated using the same uniform time step of 1 h. Using a common computational time step avoids temporal inconsistencies among GUs and ensures consistent flux calculations and aggregation. The discharge generated by each GU is subsequently aggregated according to the river network graph.

3.4 Calibration and validation protocol

Model parameters are estimated using LUCA (Let Us Calibrate), the calibration workflow integrated in OMS3 (Hay & Umemoto, 2007). The Kling–Gupta efficiency (KGE) (Gupta et al., 2009) is used as the principal objective function for discharge calibration. Where observations at nested stream gauges are available, a sequential calibration (SC) strategy (Nkiaka et al., 2018) is adopted, proceeding from upstream to downstream so that parameters estimated for an upstream calibration domain can be retained while calibrating the downstream domain.

To support a fair comparison among alternative spatial or forcing configurations, calibra-

tion experiments use consistent optimization settings within each study, including common initial parameter values, parameter bounds, and computational budgets. Parameters that could directly compensate for the methodological factor under investigation may be fixed when required by the experimental design. The resulting parameter-sharing strategy is defined for each study according to its spatial configuration and calibration domain.

Model performance is assessed over independent validation periods using KGE together with complementary goodness-of-fit metrics, including the Nash–Sutcliffe efficiency (NSE) (Nash & Sutcliffe, 1970) and NSE calculated from logarithmically transformed discharge (NSE_{log}) (Krause et al., 2005). The stream gauges used as calibration and validation targets, the spin-up, calibration, and validation periods, the parameters held fixed, and the parameter-sharing scheme are specified in the dedicated *Experimental setup* subsection of each study chapter.

In addition to these performance metrics, discharge-volume diagnostics were used to characterize model behavior more thoroughly. Cumulative observed and simulated discharge volumes at the basin outlet were compared over the calibration and validation periods to identify systematic volume biases and temporal divergence between the two series. Seasonal variations in model performance were further investigated using box-plots of the monthly relative discharge-volume errors.

Although the model is formulated to preserve water-balance closure, the closure and partitioning of the simulated water balance were also examined using paired-bar plots. For each aggregation period, the first bar represents precipitation input, whereas the second stacked bar shows its partitioning into actual evapotranspiration, discharge, and storage change. Storage change may be either positive or negative. When its sign is retained, the stacked components reproduce the precipitation input according to $P \simeq \mathrm{AET} + Q + \Delta S$, so the two bars have, except for a negligible water-balance residual, the same total height. These diagnostics were generally produced for each hydrological year, running from 1 October to 30 September of the following year, and aggregated over the entire upstream basin, with all

components expressed as equivalent water depths (mm) through basin-area normalization. The same analysis can also be performed for individual GUs and at different temporal scales, including hourly, daily, weekly, and monthly periods. For each period, ΔS is calculated as the difference between storage at the end and beginning of the aggregation interval.

3.5 Uncertainty estimation

To complement and to overcome the limitations of scalar performance metrics such as KGE (Cinkus et al., 2023; Clark et al., 2021), model uncertainty is assessed using the Empirical Conditional Probability (ECP) method proposed by Azimi et al. (2023). The simulated discharges were divided into classes, and the empirical distribution of each class was evaluated and compared against the measured data. The reliability of the model was assessed as the distance between the most probable simulated discharge (mode) and the observed discharge. In this thesis, this correspondence was further quantified as the correlation and RMSE between the mode values and the simulation as described in the supplementary material of Azimi et al. (2023).

3.6 Multi-source evaluation of the water budget

To evaluate the model's ability to reproduce various hydrological processes, its outputs were compared with satellite observations and data from Cosmic-Ray Neutron Sensing (CRNS). The primary focus of this assessment is on the model's capability to represent snow evolution and soil moisture dynamics across both spatial and temporal scales. This evaluation helps to identify the model's strengths and limitations, thereby improving understanding of its behavior and increasing confidence in its overall performance. Evaluation based solely on discharge may conceal compensating errors among internal water-budget components, even when scalar performance metrics such as KGE indicate satisfactory streamflow performance (Clark et al., 2021; Lamontagne et al., 2020; Santos et al., 2018). Therefore, discharge-based evaluation was complemented with independent data sources describing snow and

soil-moisture dynamics.

3.6.1 Snow compartment

To assess the model's capability to reproduce snow dynamics, we compared the snow water equivalent (SWE) modeled with GEOframe, calculated using the equations reported in Appendix A for the snow domain, with MODIS snow cover area data. The available MODIS data consist of daily rasters of the Normalized Difference Snow Index (NDSI). These optical satellite products are affected by cloud cover, resulting in data gaps under overcast conditions (Gafurov & Bárdossy, 2009; Gyawali & Bárdossy, 2022; Härer et al., 2018). To reduce these gaps, the Aqua (Hall & Riggs, 2021) and Terra (Hall et al., 2006) products were merged as a first preprocessing step (Gafurov & Bárdossy, 2009). Because GEOframe provides modeled SWE at the GU scale, the MODIS NDSI data were aggregated to the same spatial units. For each GU and time step, the median NDSI was calculated using a zonal statistic that included all MODIS pixels intersecting the GU polygon, including boundary-crossing pixels that were only partially contained within it. Because SWE and NDSI represent different physical quantities and are not directly comparable, both time series were converted into binary snow-presence/absence series. This transformation is based on defined thresholds for both NDSI and SWE (Parajka & Blöschl, 2006, 2008; Tong et al., 2020): if the value falls below the selected threshold, it denotes the absence of snow; otherwise, snow presence is inferred. When study-specific calibration was performed, the thresholds were selected by minimizing the Brier score (Brier, 1950) over the calibration period. This score ranges from 0 to 1, where 0 indicates perfect agreement between the two series and 1 represents complete disagreement. This comparative analysis quantifies the concordance between observed and modeled snow presence, providing a robust validation metric for assessing the model's performance in simulating snow dynamics throughout the study area. The binary series generated from the NDSI and SWE thresholds are then compared using both the Brier score (Gyawali & Bárdossy, 2022) and overall accuracy

(Tong et al., 2020). The threshold is not a fixed value but can vary with season and elevation (Tong et al., 2020). Since this variability is difficult to account for, many applications adopt a single value (Härer et al., 2018; Morlot et al., 2023). When study-specific calibration was not performed, snow presence was defined using thresholds of NDSI ≥ 40 and SWE $\geq 5\,\mathrm{mm}$ (Dawson et al., 2018; Di Marco et al., 2020; Härer et al., 2018).

3.6.2 Root-zone water content: Cosmic-Ray Neutron Sensing

As stated in Fatima et al. (2024) and Katya et al. (2020), CRNS data are compared with the model's output in terms of soil moisture dynamics. A CRNS probe records the near-surface epithermal-to-fast neutron count rate, expressed as counts per hour (CPH). Because these neutrons are moderated primarily by environmental hydrogen, their count rate is inversely related to soil water content (Zreda et al., 2008). The conversion from raw CPH to soil water content involved three stages: correction of the neutron counts for atmospheric and cosmic-ray effects, local estimation of the reference count rate N_0, and application of the calibrated neutron-to-water-content relationship as presented in Joint et al. (2017). First, to account for variations in atmospheric pressure, absolute humidity, and cosmic-ray intensity, the raw neutron count rate, CPH, was corrected using three multiplicative factors (f_P, f_W, and f_I) (Franz et al., 2012; Joint et al., 2017; Schrön et al., 2017; Zreda et al., 2012).

The corrected neutron count N is obtained by applying atmospheric and cosmic-ray correction factors to the raw neutron counts CPH, as expressed in Equation 3.3:

$$N = CPH\, f_P\, f_W\, f_I \tag{3.3}$$

The correction factors account respectively for variations in atmospheric pressure, atmospheric water vapor, and incoming cosmic-ray intensity, and are defined in Equations 3.4–3.6.

The pressure correction factor is given by

$$f_P = \exp\big[\beta_P\,(P - P_0)\big], \qquad \beta_P = 0.0072\,\frac{1}{\mathrm{hPa}}, \tag{3.4}$$

where P (hPa) is the measured atmospheric pressure, P_0 is a reference (mean) pressure, and β_P is the barometric coefficient accounting for neutron attenuation by air density. The correction for atmospheric water vapor is expressed as:

$$f_W = 1 + \beta_W\,(\rho_v - \rho_{v,0}), \qquad \beta_W = 0.0054\,\frac{\mathrm{m}^3}{\mathrm{g}}, \tag{3.5}$$

where ρ_v (g/m^3) is the absolute humidity, $\rho_{v,0}$ is a reference humidity, and β_W accounts for the moderation of neutrons by atmospheric water vapor. Finally, variations in incoming cosmic-ray intensity are corrected as:

$$f_I = \frac{I_0}{I}, \tag{3.6}$$

where I is the instantaneous cosmic-ray intensity measured at a reference neutron monitor, and I_0 is its long-term mean value.

Second, a soil-sampling campaign was conducted to estimate the gravimetric soil water content representative of the CRNS footprint and sensing depth for local calibration (Wahbi et al., 2018). Finally, the Desilets formula (Equation 3.7)

$$N_0 = N\left(\frac{a_0}{sm + soc + lw} + a_1\right)^{-1} \tag{3.7}$$

is applied for calibration (Desilets et al., 2010; Zreda et al., 2008, 2012). In the calibration the weight of the biomass water is small compared to the other sources of water (e.g., soil water) typically captured by cosmic-ray probes and it varies slowly over time. Therefore, its effect on neutron counts can be considered negligible for the purpose of the present

application (Desilets et al., 2010; Franz et al., 2012), and its approximately stationary contribution was assumed to be incorporated into the site-specific calibration of N_0. In Equation 3.7, N_0 is the reference CPH in dry conditions, N represents the measured CPH corrected and normalized as described in the first step, sm is the soil moisture obtained through gravimetric measurements during the soil sampling campaign, soc is the soil organic content (Venturelli et al., 2012), and lw is the lattice water, estimated as in J. Dong and Ochsner (2018). The remaining terms are constants as described in Pelowitz et al. (2005): $a_0 = 0.0808$, $a_1 = 0.372$. Finally, after calibration, the equation is inverted to obtain the volumetric water content θ as in Equation 3.8:

$$\theta = \left(\frac{a_0}{N/N_0 - a_1} - a_2 \right) \cdot \frac{\rho_b}{\rho_w} \tag{3.8}$$

where ρ_b is the dry-soil bulk density, ρ_w is the density of water, and $a_2 = 0.115$. When both densities are expressed in g/cm^3, ρ_w is approximated as 1 g/cm^3.

Due to the noisy nature of CRNS signals, a Savitzky–Golay filter is applied to the soil water content (SWC) values to smooth the CRNS hourly time series (Franz et al., 2020). As the resulting SWC has a variable depth that does not reach the maximum root depth, an exponential filter is applied—following Katya et al. (2020) and Peterson et al. (2016)—to extend the values to the root zone depth, which is obtained from Venturelli et al. (2012). The SWC is then transformed into millimeters of water and normalized (De Santis & Biondi, 2018). The processed CRNS series was compared with the simulated root-zone water storage of the GU containing the sensor using Pearson's correlation coefficient. Because the CRNS footprint is substantially smaller than the GU, this comparison evaluates temporal co-variation rather than strict spatial equivalence.

3.6.3 Root-zone and surface soil moisture: satellite and reanalysis products

In addition to the comparison with CRNS observations, simulated soil-moisture dynamics are compared with the Copernicus Surface Soil Moisture product with 1 km spatial

sampling (SSM1km). A direct comparison is not straightforward because SSM is expressed as a relative degree of saturation ranging from 0 to 100%, whereas GEOframe expresses root-zone water storage in mm. Furthermore, SSM mainly represents the upper few centimeters of soil (approximately 0–5 cm), rather than the entire root zone. The analysis follows several steps. First, encoded SSM values are extracted from the valid pixels within each GU. These values are then multiplied by a scale factor of 0.5, as specified in the product documentation (European Commission Directorate-General Joint Research Centre, 2017). This scaling converts the stored integer values, ranging from 0 to 200, into relative surface-soil-moisture values ranging from 0 to 100% saturation. Subsequently, a separate SSM time series is generated for each GU by calculating, at each time step, the zonal median of all valid pixels intersecting the GU polygon, including boundary-crossing pixels that are only partially contained within it. The analysis was restricted to GUs with mean elevations below 1900 m asl to reduce the influence of snow cover and frozen-soil conditions (Brocca et al., 2013). An exponential filter was then applied to the selected surface-moisture proxy to derive an index representative of root-zone temporal dynamics (Albergel et al., 2008). The filtered values are then normalized following the approach in De Santis and Biondi (2018) to convert them into millimeters of water, and subsequently aggregated to daily mean values. The processed satellite data and model outputs are compared using Pearson correlation coefficients. Since the actual root-zone depth remains unknown, the time series is divided into two periods. The first period is used to calibrate the exponential filter and determine its characteristic time-length parameter, which accounts for several soil-moisture processes, including the effective depth of the soil layer. The second period is used for validation (Albergel et al., 2008). The dynamics of soil moisture in Alpine regions are difficult to assess using satellite data due to complex topography and snow cover. For this reason, the MERIDA reanalysis (Bonanno et al., 2019) was used, which provides volumetric soil moisture (θ) at four nominal soil-layer depths: 5, 25, 70, and 150 cm. Time series were prepared for each nominal depth and for every GU in each spatial configuration considered

in the corresponding study. For each nominal depth and GU, a time series was obtained by calculating the median of the MERIDA grid-cell values whose centers fell within the GU. Since the GEOframe model expresses soil water content as total water storage (mm) rather than by layer depth, a calibration procedure was required to make the two datasets comparable. The reanalysis root-zone soil moisture (θ_{RZ}) was derived as a weighted combination of the four MERIDA layers, with weights obtained by maximizing the Pearson correlation between observed and target time series (Paulik et al., 2014). To ensure vertical consistency between the model output and the reanalysis data, a vertical scaling procedure was applied (Albergel et al., 2008). Finally, to further harmonize the datasets and remove systematic bias, a linear scaling bias correction was applied, normalizing the model soil moisture (θ_{MO}) to the long-term mean and variance of θ_{RZ} (Albergel et al., 2008; De Santis & Biondi, 2018).

3.6.4 Spatial diagnostics of snow, root-zone, and groundwater storage

Because the units are spatially explicit, every state and flux of the budget can be mapped. Throughout the thesis the diagnostics are therefore presented in two complementary forms: as time series at the locations where observations exist (stream gauges, CRNS sites) and as maps of the simulated quantities over the complete set of GUs, which reveal where changes in spatial discretization or meteorological forcing modify the simulated distribution of water. These maps provide spatial diagnostics of model behavior but do not, by themselves, constitute validation of the simulated spatial patterns.

Maps of snow water equivalent, root-zone storage, and groundwater storage are used systematically in Chapter 5, and elevation-band aggregations of the same quantities in Chapters 6 and 7.

CHAPTER 4. SOFTWARE DEVELOPMENT

The GEOframe hydrological modeling framework used in this thesis has been developed collaboratively over several years by a community of developers and users. At the beginning of this doctoral work, GEOframe already comprised a mature collection of model components and the software infrastructure required for their integration and application. The work presented in this thesis built upon this existing framework by introducing improvements and specific extensions required for the experiments described in the following chapters. Some of these developments were carried out jointly with other contributors, consistently with the collaborative nature of the GEOframe project. Continuous exchanges among developers and feedback from users applying the framework under different conditions also helped identify and resolve issues that emerged only in specific applications, particularly in complex basin configurations.

Within this broader context, the software activities undertaken during the doctorate—including the development of new components, extensions to existing components, infrastructure improvements, and bug fixes—constitute a research output in their own right. They are therefore presented together in this chapter rather than being distributed among the individual experimental studies.

Section 4.1 introduces GEOframe as the modeling framework and computational environment used throughout the thesis. Sections 4.2–4.5 describe, respectively, the mixed-resolution input builder, the redesigned kriging component, the rainfall-runoff model and its numerical core, and the Python tools developed for data preparation and model diagnostics. Section 4.6 presents the maintenance, build, and distribution infrastructure, whereas Section 4.7 discusses how the software and data-management choices support reproducibility, the FAIR principles, and the DARTHs blueprint. Finally, Section 4.8 describes the operational prototype developed for the Noce River basin.

4.1 The GEOframe modeling framework

To address the research questions systematically and reproducibly, the experiments required a computational environment in which spatial discretizations, meteorological forcing strategies, and individual model components could be modified while keeping the remaining elements of the modeling workflow unchanged. GEOframe, as introduced in Chapter 3, was therefore used both as a hydrological modeling framework and as an experimental environment for investigating scale-related effects.

GEOframe is built upon OMS3, which supports the integration of modular model components into reproducible environmental modeling workflows (David et al., 2013, 2016). Within this environment, alternative spatial configurations and meteorological forcing strategies can be implemented and compared through controlled numerical experiments while retaining a common representation of the remaining hydrological processes.

4.2 Mixed-resolution input builder

The objective is to generate a mixed-resolution basin configuration in which selected GUs are subdivided into smaller computational units, whereas the remainder of the basin retains its original coarser discretization (Section 3.1.2). Producing such a configuration requires more than the local delineation of additional GUs. The coarse and refined units must be combined into a single, coherent routing topology and accompanied by the complete directory and file structure required to run GEOframe simulations. This structure includes the geomorphological attributes associated with each unit and the topological connections involving lakes.

For a basin represented using a uniform target resolution, these inputs are generated through the standard `GeoFrameInputBuilder` workflow. Starting from the delineated drainage network and sub-catchments, this workflow organizes the spatial information into the directory and file structure required by GEOframe and constructs the final basin

topology. However, the standard delineation procedure applies a single target contributing-area threshold across the entire basin. In the applications considered here, this generally produces GUs whose areas remain within approximately $\pm 20\%$ of the prescribed target area, but it does not allow individual parts of the basin to be selectively refined.

To enable local refinement, a dedicated Horton Machine module, `GeoFrameInputBuilderWithSubBasin`, was developed by extending the `HMModel` base class. The module coordinates the existing `GeoFrameInputBuilder` and `NetNumbering` components and can construct the initial coarse-resolution configuration in two ways. It can execute the standard `GeoFrameInputBuilder` workflow or read an existing configuration from a user-specified directory. The user then provides the identifiers of the GUs to be refined and the finer contributing-area threshold to be applied within them.

For each selected GU, `NetNumbering` is rerun using the finer threshold to extract the local drainage network and delineate the corresponding smaller sub-catchments. The `GeoFrameInputBuilder` workflow is then executed for the refined portion of the basin to generate the required geomorphological attributes and simulation-input files. Finally, these outputs are merged with the unchanged coarse-resolution configuration. Unit identifiers and upstream–downstream relationships are reconciled to ensure that the refined units are correctly connected to the downstream coarse network and that the lake-related connections defined in the original topology are preserved.

The resulting output is a single, topologically consistent, and reproducible mixed-resolution configuration that can be used directly in GEOframe simulations, without manually modifying the directory structure or individual input files. This module enables the mixed-resolution spatial design adopted in Chapter 5 and is released as a separate software module, as documented in Appendix F.

4.3 The kriging component

The kriging component was extensively refactored and redesigned to improve its maintainability, extensibility, and code reuse, following established object-oriented design principles (Sarcar, 2022). The redesigned architecture supports the interpolation of both point and raster data, the automatic estimation and selection of theoretical variogram models, and leave-one-out cross-validation within a common computational workflow.

The module is organized into several subpackages under the root package `org.geoframe.blogpost.kriging`. The main class, `Kriging`, is an abstract class that implements the common logic of the kriging algorithm. It is extended by two specialized classes: `KrigingPointCase`, which handles vector-based data represented as collections of points, and `KrigingRasterCase`, which handles raster data. Each class belongs to a dedicated subpackage.

The common computational workflow is defined in the abstract class, whereas specialized behavior is implemented through hook methods and data-provider interfaces. In particular, the `InterpolationDataProvider` interface supplies the spatial data required by the interpolation. Its specialized implementations are the `RasterInterpolationDataProvider` for raster data and the `VectorInterpolationDataProvider` for vector data. These implementations allow data to be read from different sources, whereas the point- and raster-based kriging classes manage the outputs in the corresponding formats.

The `kriging.variogram` package contains the components used to calculate the experimental variogram and fit the candidate theoretical models. The `ExperimentalVariogram` class uses the observed values and corresponding station locations to calculate the experimental variogram. The classes representing the available theoretical models are contained in the `variogram.theoretical.model` package, whereas the algorithms used to estimate their parameters are implemented in `variogram.theoretical.model.curvefitter`.

At each time step, the available theoretical models are fitted to the experimental variogram and compared according to the prescribed error criterion. The best-performing model is then selected for the interpolation. If none of the time-step-specific fits satisfies the acceptance criterion, the workflow uses a global variogram as a fallback. The package also provides the `GlobalParameterEvaluator`, located in `variogram.theoretical`, which can be called from the OMS3 console to evaluate the global theoretical variogram.

The `primarylocation` package contains the logic used to manage the meteorological stations. This includes a statistical test to determine whether the variable exhibits a significant relationship with elevation and methods for restricting the interpolation to the n stations closest to each target location. Besides reducing the computational cost, this station-selection option can reduce the risk of ill-conditioned kriging matrices, particularly when the available stations provide spatially redundant information.

The overall computational workflow and the principal interactions available to the user are summarized, respectively, by the activity and use-case diagrams in Figures 4.1 and 4.2.

Three outcomes of the redesign are particularly relevant to the applications presented in this thesis. First, the automatic calculation of the experimental variogram, fitting and comparison of the candidate theoretical models, and fallback to a global variogram enable long time series to be processed without the manual selection of a variogram for each time step (Section 3.2.1).

Second, the computational cost of the interpolation chain was substantially reduced by restructuring the solution of the kriging system and the management of station subsets. These improvements made the hourly interpolation of multi-year datasets over the spatial configurations considered here computationally feasible. The corresponding computational costs are reported in Appendix E.

Third, the leave-one-out cross-validation component was redesigned within the same framework, allowing the predictive performance of the interpolated fields to be evaluated using the same kriging implementation. At each time step, each station is sequentially

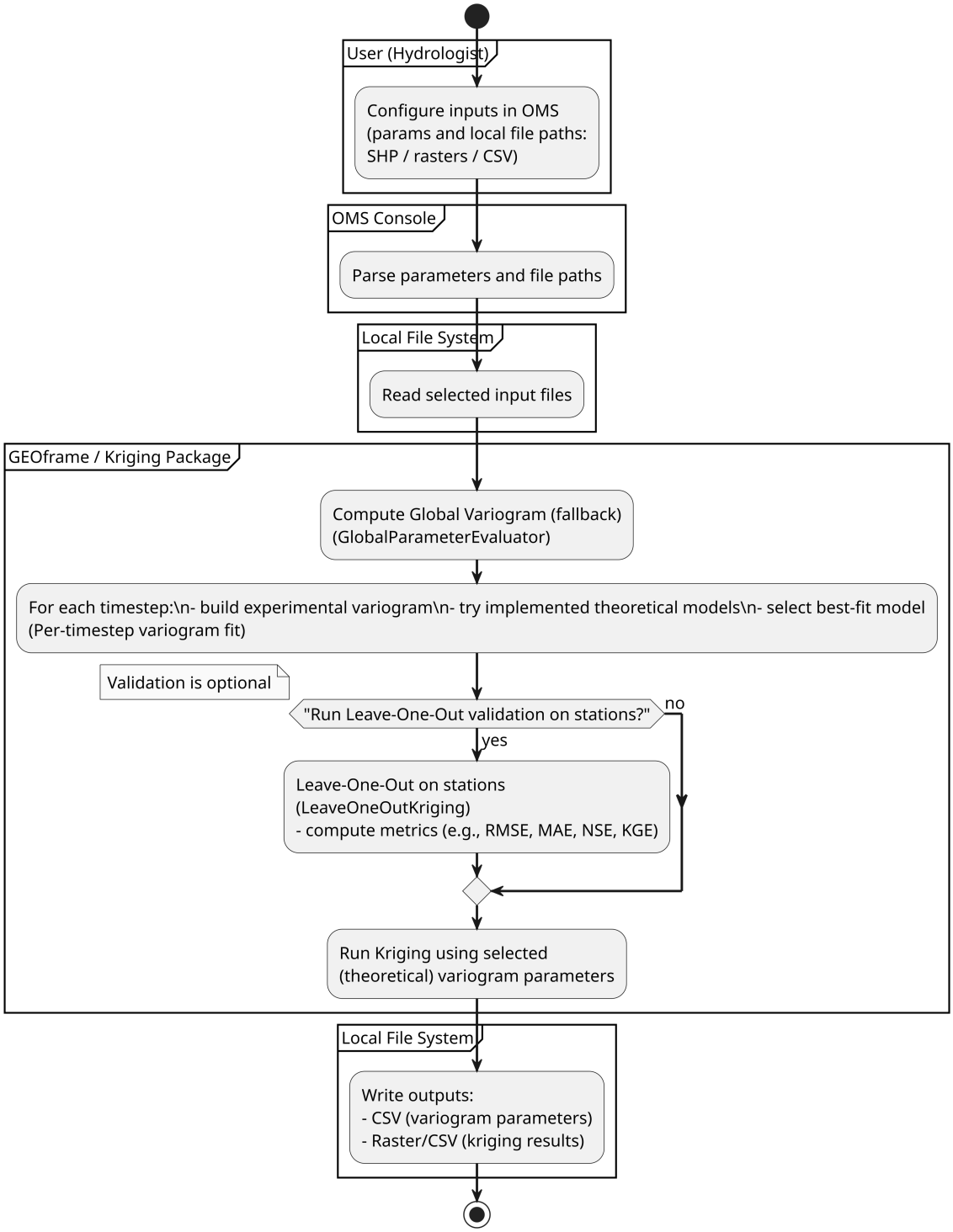

Figure 4.1 – Activity diagram of the automated kriging workflow. At each time step, the procedure computes an experimental semivariogram, fits a theoretical model, and performs spatial interpolation and leave-one-out cross-validation. When the time-specific fitting procedure fails, the global variogram is used as a fallback.

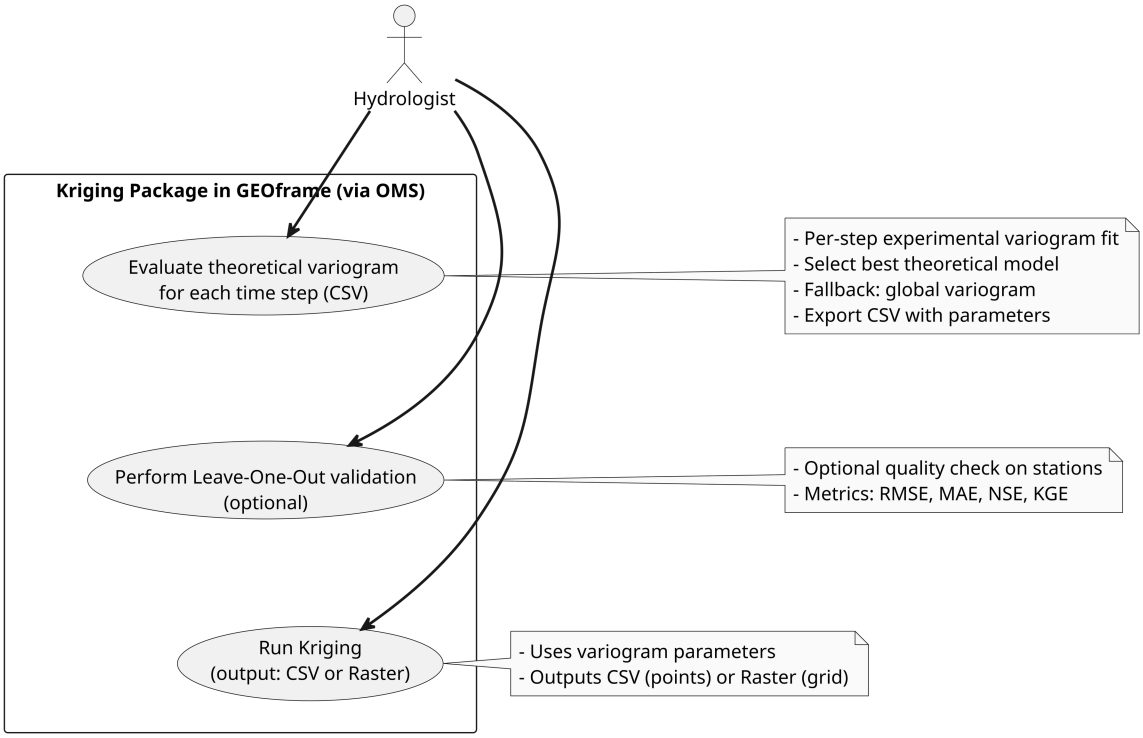

Figure 4.2 – Use-case diagram of the kriging component within the OMS3 framework, illustrating its main functionalities and their interactions.

excluded from the kriging system, and its value is estimated using the remaining stations. The variogram is fitted using all the observations available at that time step, including the observation from the station being evaluated. Given that the observational network comprises more than 40 stations, the influence of an individual station on the fitted variogram is expected to be limited. The resulting predictions are organized as a matrix in which the rows represent the time steps and the columns represent the stations subjected to cross-validation.

4.4 The rainfall-runoff model and its numerical core

The implementation of the Embedded Reservoir Model (ERM), the rainfall-runoff model used in this thesis, extends the model described by Bancheri (2017). During this work, the model was refactored and stabilized through targeted bug fixes. Additional constraints were introduced to prevent negative storages and preserve the water balance of the individual model compartments, and the numerical core was upgraded by implementing a fourth-order

Runge–Kutta method with adaptive step size.

The constraints on the compartment fluxes ensure that the total outflow does not exceed the water available in storage. When the available water is insufficient to sustain all potential outgoing fluxes, actual evapotranspiration is given priority, whereas percolation and the other outgoing fluxes are reduced or suppressed. This allocation prevents the numerical solution from producing negative storage values while maintaining consistency among storage changes, incoming fluxes, and outgoing fluxes. The adaptive integration scheme adjusts the internal time step according to the local dynamics of each reservoir, improving the numerical stability of the model under different hydrological conditions.

The relevance of these developments is not merely computational. Numerical approximation errors in conceptual hydrological models can affect the simulated fluxes and the calibration objective function and may subsequently be compensated for by changes in the calibrated parameter values (Kavetski & Clark, 2010). Such compensation can impair parameter interpretation and confound comparisons among model configurations. In the tests performed during this work, the numerical mass-balance residuals calculated separately for the individual compartments ranged approximately from 10^{-15} to 10^{-9} in the storage units used by the model. By preventing negative storages and reducing numerical mass-balance errors to negligible levels, the revised implementation reduces the likelihood that the differences among the configurations investigated in Chapters 5–7 arise from numerical artifacts. These differences can therefore be more confidently interpreted in relation to the alternative spatial representations.

4.5 Python tooling for data preparation and diagnostics

In addition to the Java components, reproducible workflows for data preparation, quality control, and model-output analysis were developed as Python Jupyter notebooks and as reusable utilities within the `GEOframePy` package. The notebooks contain operations specific to the experiments presented in this thesis, whereas functions of more general use

were incorporated into `GEOframePy` so that they could be reused across different modeling applications.

`GEOframePy` was already part of the GEOframe software ecosystem and was extended during this work. Several functions initially implemented within individual notebooks were reorganized and transferred to the library, reducing code duplication and improving their accessibility to other workflows. New utilities were also introduced to represent and manage basin topology using the NetworkX graph library (Hagberg et al., 2008). These utilities support operations such as network traversal and the subdivision of the drainage network into sub-catchments, allowing the basin topology to be directly inspected and manipulated within Python workflows.

4.6 Maintenance and infrastructure of the framework

Alongside the development of the components required for the experiments, part of the work involved maintaining and consolidating the broader GEOframe software infrastructure. These activities included rewriting selected algorithms, identifying and correcting long-standing bugs, and reorganizing the repositories and documentation associated with the framework. The rainfall-runoff model modified during this thesis and the newly developed kriging component were also applied and tested in a separate research activity involving the Po River Basin Authority (AdbPo). This provided an opportunity to evaluate these components beyond the experiments presented in this thesis and to identify issues arising in a different application context.

The Git repositories of the GEOframe organization (geoframecomponents, n.d.) and the project website were reorganized to improve the discoverability, documentation, and distribution of the individual components. Where available, dedicated repositories, versioned releases, and component-specific documentation make it possible to identify and reference the software used in a particular model configuration.

As part of this broader effort, the GEOframe developer community began migrating

selected components to Maven (Apache Software Foundation, n.d.), an effort to which this thesis contributed. A build system automates the compilation, testing, packaging, and management of software dependencies. Although the previous infrastructure was primarily based on Gradle (Gradle Inc., n.d.), Maven was progressively adopted to provide a standardized build and distribution workflow. The transition is still ongoing, and both systems are currently supported to maintain compatibility with existing components.

The adoption of Maven also enabled selected GEOframe components to be published on Maven Central as versioned software artifacts (Sonatype, n.d.).[1] This distribution approach allows other projects to reference specific component versions explicitly rather than relying on locally compiled or manually distributed copies of the source code. It therefore simplifies dependency management, improves the traceability of the software environment used for model simulations, and facilitates the reuse of individual components within the broader GEOframe community.

Building on the migration to Maven-based distribution, a broader redesign of the GEOframe workflow was initiated and is currently being developed within the Horton Machine project. This effort aims to streamline the simulation chain by consolidating the data used and produced for a basin into a single GeoPackage database. The database can contain the vector-based geographical information of the basin, meteorological-station metadata and observations, GU-level input time series, calibrated parameter sets, and, for each model run, separate tables containing the simulated state variables and fluxes. By storing model inputs, parameters, and outputs in a structured and portable format, this approach is intended to improve the reproducibility and traceability of the simulations.

4.7 Reproducibility: FAIR principles and the DARTHs blueprint

The software and data-management choices described above support the reproducibility of the experiments presented in this thesis. This is particularly important because

[1] https://central.sonatype.com/namespace/org.geoframe

comparisons among model configurations can be independently assessed only when the corresponding input data, software versions, parameter sets, and simulation workflows can be identified and accessed. The adopted approach is aligned with the FAIR principles (Wilkinson et al., 2016) and with the blueprint for participatory Digital eARth Twin Hydrology systems (DARTHs) proposed by Rigon et al. (2022). The development of these practices within the GEOframe community is also documented in the group's open notebook (Rigon, n.d.).

In practical terms, these principles are supported as follows:

- **Findable.** The software components, datasets, and simulation materials are organized in identifiable repositories, with versioned releases and persistent identifiers provided where available. The resources associated with this thesis are listed in Appendix F.
- **Accessible.** The code and the material required to reproduce the experiments are made available through the repositories identified in Appendix F, subject to the licenses and access conditions associated with the individual resources.
- **Interoperable.** The model components communicate through OMS3 annotations and exchange commonly used data formats, including OMS time-series files, GeoTIFF rasters, and shapefiles. This modular structure facilitates the integration and replacement of components that implement compatible interfaces.
- **Reusable.** Refactoring, documentation, versioned releases, and reusable preprocessing and post-processing utilities facilitate the application of the software components in modeling configurations other than those examined in this thesis.

Accordingly, the published research material includes not only the source code but also the available input datasets, OMS3 simulation and parameter files, preprocessing and post-processing scripts, and model outputs required to reconstruct the experimental workflows.

4.8 Operational prototype for the Noce River basin

Beyond the experimental applications presented in this thesis, the resulting modeling chain was used to develop an operational prototype for the Noce River basin. Its configuration was informed by the findings presented in Chapters 5–7 and incorporates the GEOframe developments described in Section 4.6. The prototype uses a spatial configuration generated with a contributing-area threshold of 9 km^2, adopts a simplified representation of the basin in which dams and lakes are not explicitly represented, and is executed once per day. Meteorological forcings are assigned to the GUs using the CEN configuration described in Section 3.2, in which the value interpolated at the centroid of each GU is assumed to represent the entire unit.

At the beginning of each execution, the workflow first inspects the current state of the GeoPackage database to verify the consistency and temporal coverage of the stored records. It determines the most recent time step for which a complete set of meteorological observations is available and identifies any missing intervals. This initial check establishes which observations must be retrieved and the starting date and time of the data request. The required observations are then retrieved from the online data services operated by MeteoTrentino (MeteoTrentino, n.d.) and Fondazione Edmund Mach (FEM) (Fondazione Edmund Mach, 2022) and stored in the dedicated weather-station data table.

After the acquisition, the workflow checks the temporal continuity and completeness of the updated records. If missing observations are detected, the acquisition procedure is repeated in an attempt to recover them. The hydrological simulation is started only after data completeness has been verified. If any gap persists, the entire daily update is skipped rather than advancing the model with incomplete meteorological forcings. The workflow can subsequently be restarted once the missing observations have been recovered. These checks prevent incomplete data transfers or temporary data-access problems from affecting the continuity and consistency of the simulated model states.

Once the input data have been validated, the workflow performs the required preprocessing steps. The meteorological observations are spatially interpolated using the automatic variogram-selection procedure, and the interpolated values are evaluated at the centroid of each GU according to the CEN configuration. Solar radiation and potential evapotranspiration are then calculated for each GU. Finally, the resulting forcings are used to advance the rainfall-runoff model. The simulation produces spatially distributed estimates of water-budget components and hydrological state variables, including snow water equivalent, root-zone and groundwater storage, actual evapotranspiration, and discharge.

All simulated state variables and fluxes are stored in the GeoPackage database. At the end of each successful execution, the final values of the state variables are also retained as restart conditions and reloaded at the beginning of the subsequent run. This warm-start procedure avoids repeating the complete model spin-up, reduces the computational cost of each update, and allows the simulated basin conditions to be advanced continuously as new observations become available.

To make the information stored in the GeoPackage accessible to the web application, a web service was developed using FastAPI, a Python framework for building web APIs (Ramírez, n.d.). The service exposes a REST API, namely an application programming interface based on the Representational State Transfer architectural style introduced by Fielding (2000). An API defines a set of endpoints, each corresponding to an address accessible through a specific URL, through which a software client can request a particular resource. In this implementation, the REST API receives HTTP requests, translates them into queries to the GeoPackage database, and returns the requested information in a structured format.

This architecture separates data storage from data presentation: the web application does not need to access the GeoPackage directly or reproduce its internal database queries. Moreover, the API is not limited to supporting the website. The same endpoints could also be used by other applications, authorized external users, or automated machine-to-machine

services.

A dedicated web application was developed to explore the data made available through the API. Its landing page displays an interactive map of the basin, including the GU boundaries, meteorological stations, and stream gauges (Figure 4.3). Selecting one of these elements opens a contextual information panel on the right-hand side of the interface. For each meteorological station, the panel reports its name and elevation and displays the available time series of air temperature and precipitation, as well as radiation where available. For each stream gauge, it displays the corresponding observed discharge time series.

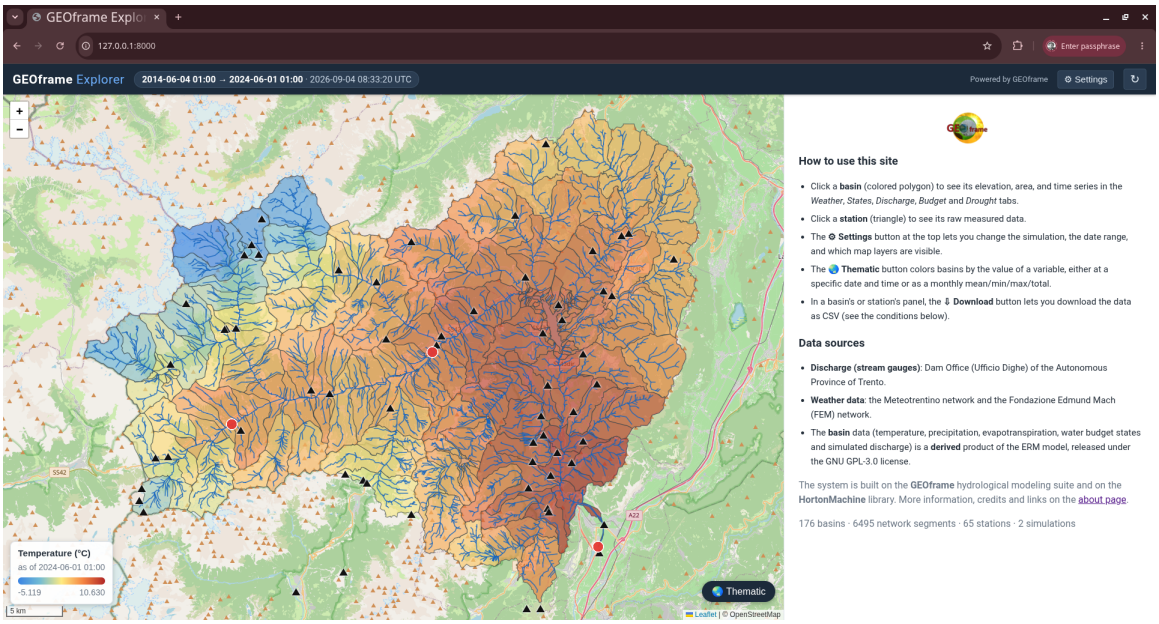

Figure 4.3 – Web interface of the operational prototype for the Noce River basin. The interactive map displays the GU boundaries, meteorological stations, and stream gauges. Selecting a map element opens the corresponding information and available time series in the panel on the right-hand side.

When a GU is selected, the upper part of the right-hand panel summarizes its principal geographical and topographical attributes, including minimum, mean, and maximum elevation, area, channel length, and the sky-view factor evaluated at its centroid. It also identifies the upstream contributing GUs and the downstream GU according to the basin topology (Figure 4.4).

The lower part of the panel contains a series of tabs through which the user can explore

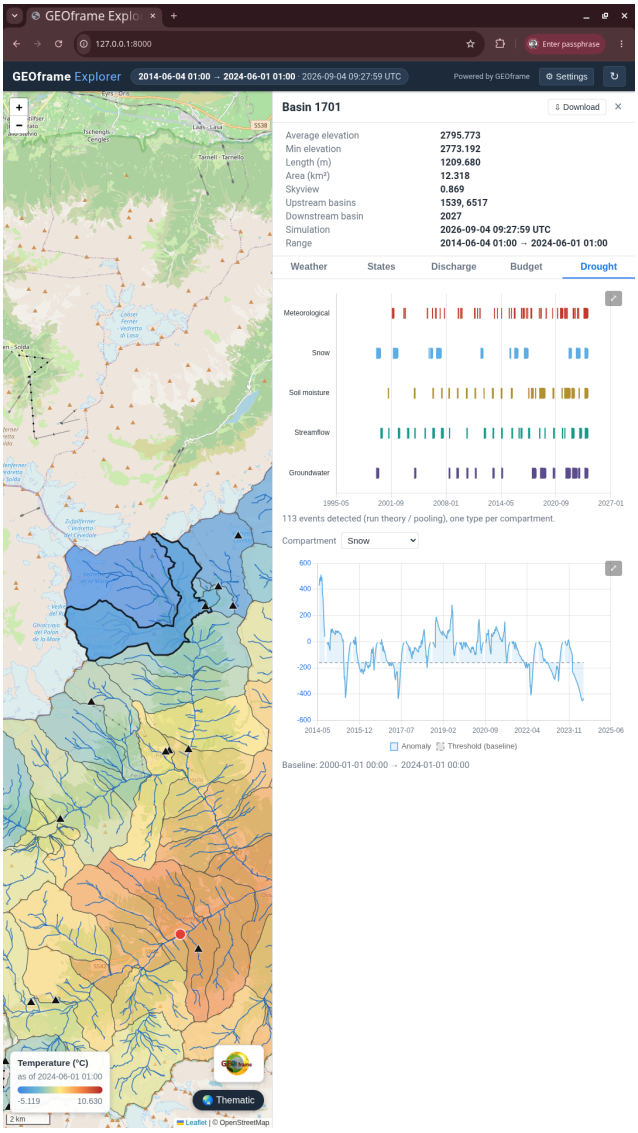

Figure 4.4 – Detailed view of a selected GU in the web application. The upper part of the right-hand panel reports its principal geographical and topographical attributes and its connections with upstream and downstream units. The lower part provides access to meteorological forcings, simulated state variables and internal fluxes, discharge, water-budget diagnostics, and drought diagnostics.

different groups of information. The meteorological-forcing tab displays air temperature, precipitation, solar radiation, and potential evapotranspiration. The state tab presents the simulated state variables and internal hydrological fluxes, while the discharge tab displays the simulated discharge at the outlet of the selected GU. The final two tabs provide water-budget and drought diagnostics.

The water-budget and drought views adapt the visualization and analysis procedures described by Rigon (2026). The water-budget view jointly represents the fluxes entering and leaving the selected GU and the corresponding changes in storage, allowing the consistency of the simulated water balance to be inspected. The drought view identifies meteorological and compartment-specific drought periods and displays the corresponding anomaly time series.

The drought diagnostics are based on the continuous model simulation covering the period 2000–2025. The years 2000–2020 are used as an independent reference baseline for calculating the climatologies, anomalies, and drought thresholds applied to both the meteorological variables and the simulated hydrological compartments. Conditions from 2021 onward are evaluated against this reference. The baseline could not be extended further into the past because only one or two meteorological stations were available before 2000. This spatial coverage was considered insufficient to characterize meteorological variability across the basin and generate spatially representative forcing fields for a consistent retrospective simulation.

The web application is currently accessible only from within the internal network and requires authentication. Within this setting, the prototype demonstrates the technical feasibility of applying the modeling chain in a routinely updated context. Nevertheless, it should not yet be regarded as a fully operational water-management system. Further developments are required in areas such as validation, automated failure recovery, system monitoring, production deployment, and access management before it can reliably support operational decision-making.

CHAPTER 5. MIXED-RESOLUTION ANALYSIS OF THE WATER BUDGET OF AN ALPINE BASIN

This chapter presents Study 1 and introduces the mixed-resolution (hybrid) modeling approach used throughout the thesis to improve process representation in stakeholder-critical areas while keeping the overall setup computationally efficient.

Scope and original contributions

Climate change is rapidly reshaping water availability in the European Alps, increasing the need for effective water-resource management and adaptation strategies. At the same time, new observational datasets, including satellite-derived products and ground-based measurements from Cosmic-Ray Neutron Sensing (CRNS), provide opportunities to evaluate hydrological processes beyond discharge alone. Exploiting these complementary observations requires flexible hydrological models capable of representing water-budget components at the spatial scales relevant to both process understanding and water management.

This study introduces a mixed-resolution (hybrid) discretization for semi-distributed hydrological modeling, together with the software developed to generate it (Section 4.2). Instead of adopting a single characteristic scale for the entire basin, the approach combines a coarse basin-wide partition with selective high-resolution refinement in stakeholder-critical areas, such as irrigated agricultural valleys and areas surrounding artificial reservoirs.

The study makes three main contributions. First, the refinement is not based on a nested grid but on a topologically consistent re-delineation of Geomorphic Units (GUs), so that the coarse and refined portions of the basin remain part of a single connected routing structure and model configuration. Second, the conventional and mixed-resolution configurations are compared not only in terms of discharge but also through the components and spatial distribution of the water budget. CRNS measurements are used to evaluate root-zone

water content, while MODIS observations are used to evaluate snow cover and Copernicus products to evaluate surface soil moisture. The mixed-resolution configuration improves the simulation of snow cover, reducing the median Brier score from 0.20 to 0.14, and increases the correlations with both CRNS and satellite soil-moisture observations. Third, the analysis includes the 2022 Alpine snow drought, allowing the value of additional spatial resolution to be assessed under particularly challenging conditions for water-resource management.

These results show that selectively increasing spatial detail can improve the representation of the internal partitioning of water even when changes in the outlet hydrograph are limited. The approach therefore provides a computationally efficient basis for applications ranging from drought-mitigation planning to climate-impact assessment and water-resource management in complex Alpine environments.

5.1 Introduction

Water resources in Alpine regions are undergoing unprecedented changes driven by climate variability and increasing anthropogenic pressures (Immerzeel et al., 2010; Viviroli et al., 2007b). The European Alps, like many mountainous regions worldwide, are experiencing significant alterations in hydrological regimes due to climate change (e.g., Mastrotheodoros et al., 2020), with far-reaching implications for ecosystems, water resource management, and human well-being (Beniston, 2012).

To better capture and understand the complex spatiotemporal dynamics of mountainous catchment water budgets, hyper-resolution hydrological models (with spatial scales equal to or smaller than 1 km^2 and hourly or sub-hourly time steps) integrated with hydrological multi-process in situ and satellite observations (e.g., An et al., 2025; Dall'Amico et al., 2025; van Jaarsveld et al., 2025) offer promising avenues for more accurate simulation and improved water use strategies. The pronounced topographical and geomorphological heterogeneity of mountainous regions, such as Alpine landscapes, results in substantial spatial variability in meteorological and hydrological conditions, necessitating the use of

high-resolution modeling approaches (Carletti et al., 2022). However, the scarcity of in situ hydro-meteorological observations (sparse weather and discharge measurement stations in Alpine terrain with complex topography) in these remote and often inaccessible regions poses a significant challenge to the deployment of fully distributed models. This creates a persistent tension between the desired model complexity and the limitations of available data, underscoring the need for innovative methodological solutions. Addressing these challenges requires hydrological models that can effectively operate across multiple spatial and temporal scales to accommodate the inherent heterogeneity of Alpine hydrological processes.

Understanding the causes of spatial heterogeneity in hydrological processes, such as rainfall, snow, runoff, evaporation, and subsurface water flow, remains an open scientific question (e.g., Blöschl et al., 2019). This in turn reflects one of the major challenges in hydrology, i.e., providing reliable information at spatial and temporal scales that are meaningful for stakeholders implementing flood or drought adaptation strategies and managing water resources (e.g., Wada et al., 2017), including applications such as agricultural scheduling or optimizing hydropower production. Overcoming limitations in current representations of hydrological processes in regional water management models is essential. To address this challenge, it is crucial to collect data from multiple sources, across various processes and spatial resolutions and use them in advanced multi-scale/multi-process hydrological models.

In recent years, the scientific community has made significant progress in improving the quality and spatial-temporal resolution of hydrological data from multiple sources, including both ground-based and satellite observations.

Recent research into model validation techniques using multiple observational datasets (Biondi et al., 2012; Koltsida & Kallioras, 2022; Moriasi et al., 2007) offers promising avenues for improving model reliability (i.e., consistent agreement between simulations and observations across multiple metrics and time periods) and robustness (i.e., the ability of the model to maintain acceptable performance under different hydro-climatic conditions).

In particular, remote sensing data are proving indispensable for hydrological modeling in data-scarce mountain regions. For example, MODIS snow cover products from the Terra and Aqua satellites provide daily observations with approximately 500 m spatial resolution, making them especially useful for regional snow cover monitoring (e.g., Sirguey et al., 2008). These products have demonstrated high accuracy, around 95%, when compared to ground-based measurements, although cloud cover remains a challenge, often addressed through gap-filling techniques (e.g., Gascoin et al., 2015; Parajka & Blöschl, 2006; Şorman et al., 2007). Similarly, remotely sensed soil moisture datasets, such as those provided by the European Union's Copernicus Earth Observation Programme, supply critical information on surface soil conditions at regional scales, with a typical grid-cell area of about 1 km^2 (Dorigo et al., 2017), supporting improved model input and calibration in areas with limited ground observations.

Despite the growing importance of remote sensing data in hydrological modeling, ground-based measurements remain indispensable for accurate model validation of the state variables. Among these, CRNS stands out for its ability to provide non-invasive, field-scale estimates of soil moisture, with a measurement footprint of approximately 100 m in radius (Zreda et al., 2008, 2012) and sub-daily temporal resolution (Franz et al., 2020). This intermediate spatial scale effectively bridges the gap between traditional point-based measurements and broader-scale remote sensing observations (see Table C2). Due to these advantages, CRNS technology has been recently applied across various domains. In snow hydrology, it has been used to estimate snow depth and snow water equivalent (SWE) (Schattan et al., 2017), while in agricultural water management, it supports the estimation of soil water content (Brogi et al., 2023; Franz et al., 2020). Additionally, CRNS data have been successfully integrated with hydrological models at multiple spatial scales (Franz et al., 2013, 2020; Scheiffele et al., 2025). Notably, recent studies have demonstrated the utility of CRNS measurements in validating model outputs within catchment-scale hydrological models (Fatima et al., 2024; Katya et al., 2020), highlighting their potential to enhance

model accuracy in complex environments.

To fully leverage these advancements in data availability and better understand changes in multiple hydrological processes at high spatial resolutions, it is essential to develop and refine hydrological models capable of integrating diverse data sources. Such models can provide multi-process insights at fine spatial scales, supporting more effective and adaptive water management strategies.

Advancements in flexible modeling frameworks (e.g., Fenicia et al., 2011; Formetta et al., 2014a) have opened new opportunities for implementing scale-adaptive approaches that strategically allocate computational resources to areas where they are most needed. Furthermore, the complex interactions among snow dynamics, soil moisture, and runoff necessitate multi-component modeling frameworks that go beyond traditional discharge-centric approaches, enabling a more comprehensive simulation of the complete water budget. Open-source modeling frameworks that offer flexible, process-based approaches (often employing structures based on GUs) are increasingly seen as valuable tools for tackling these challenges by combining various methods. Systems like FLEX (Fenicia et al., 2008; Kavetski & Fenicia, 2011), SUMMA (Clark et al., 2015b), SuperflexPy (Dal Molin et al., 2021), and GEOframe (Abera et al., 2017; Bancheri, 2017; Formetta et al., 2014a; Serafin et al., 2018) provide the flexibility, scalability, reproducibility, and transparency (i.e., the ability to inspect and critically evaluate the model implementation and code together with the underlying conceptual formulation) needed to develop and test different modeling hypotheses.

Among these, GEOframe stands out as especially well-suited for mountainous Alpine environments, thanks to its modular design (where modeling processes are implemented as interchangeable components) and ability to integrate multiple data sources across different spatial and temporal scales. In the past, different configurations of the components were used, for example, by Abera et al. (2017) to estimate water budget components in pre-Alpine basins, highlighting the importance of accounting for spatial variability in precipitation

and evapotranspiration. Morlot et al. (2023) applied GEOframe for modeling snow, soil moisture, and river discharge evolution of the Adige River basin, considering also artificial reservoirs and human influence on the water budget. Arnone et al. (2023) used GEOframe to simulate peak flow generation in a small mountainous catchment, incorporating high temporal resolution meteorological data (5-minute intervals) necessary to capture intense storms. More recently, Azimi et al. (2023) improved upon these methods by integrating enhanced snow modeling techniques and validating results with remote sensing data.

However, these studies have generally used uniform spatial discretizations throughout the modeling domain, which may overlook opportunities to optimize computational efficiency by adjusting resolution according to terrain complexity and dominant hydrological processes. Research on mixed spatial resolution modeling in various environmental fields has shown promise for balancing computational efficiency with accurate process representation. Although focused on different hydrological domains, recent advances in hydrologic model scaling (S. Kumar & Ojha, 2024), multiscale process analysis (Samaniego et al., 2017), and hybrid modeling (B. Li et al., 2024), as well as multiscale snowdrift-permitting snowpack simulations (Marsh et al., 2024) and flexible multi-physics snow modeling including forest canopy processes (Essery et al., 2025), indicate that mixed-resolution approaches could greatly enhance water budget estimation in topographically complex areas. These methods align with the idea of "model everywhere" frameworks (K. Beven, 2025; Rigon et al., 2022), which strive to deliver comprehensive yet efficient environmental modeling solutions.

This study builds upon previous work by implementing a novel mixed-resolution approach to hydrological modeling, emphasizing the calibration and validation of individual hydrological cycle components. We explore how varying the spatial resolution of GUs influences the accuracy of different water budget components, with a particular focus on snow dynamics and soil moisture processes.

Using diverse data sources, including streamflow measurements, MODIS snow cover products, Copernicus soil moisture data, and field-scale CRNS observations, we assess

model performance across scales and processes. Additionally, we examine how the developed mixed-resolution method improves the quantification of extreme events such as low-flow and high-flow periods observed in the analyzed catchment. Our specific objectives are to:i) develop and implement a mixed-resolution modeling framework that optimizes spatial discretizations based on terrain features and dominant hydrological processes; ii) evaluate the impact of spatial resolution on the simulation of key water budget components, especially snow accumulation and melt, as well as soil moisture dynamics; iii) assess model performance through comparisons with multiple independent datasets covering various spatial scales; iv) enhance understanding of Alpine water budgets while providing methodological innovations applicable to similar mountainous regions worldwide. The chapter is organized as follows:i) Section 5.2 describes the experimental setup, while the study area and data are presented in Chapter 2 and the common methods and the hydrological modeling framework in Chapter 3; ii) Section 5.3 presents the results and their validation across various processes and available datasets, and discusses the implications of our findings for understanding Alpine hydrological processes; iii) Section 5.4 contains conclusions and recommendations for future research.

5.2 Experimental setup

The basin, monitoring network, and datasets are described in Chapter 2. The analysis considers the entire Noce basin down to Mezzolombardo over the period 1 January 2017–15 September 2022.

Two spatial configurations, referred to as CFG1 and CFG2, are compared to assess the effect of model spatial discretization. CFG1 represents the reference configuration and is obtained by delineating the entire basin using a constant contributing-area threshold of $10\,\mathrm{km}^2$. CFG2 is derived from CFG1 using a mixed-resolution strategy, in which selected GUs are locally refined using a contributing-area threshold of $1\,\mathrm{km}^2$, while the remainder of the basin retains the $10\,\mathrm{km}^2$ threshold. The GU selection procedures for CFG1 and CFG2

are presented in Section 3.1.1 and Section 3.1.2.

The refinement was designed to increase spatial detail in areas representative of different geomorphological and hydrological conditions. Five GUs upstream of each of the three main stream gauges—Vermiglio, Malé, and Mezzolombardo—were manually selected for refinement. The Vermiglio area represents high-elevation mountain terrain, including glacierized areas; the Malé area represents intermediate-elevation and transitional hydrological conditions; and the Mezzolombardo area represents lower-elevation, cultivated regions characterized by flatter terrain and agricultural water use. The selected GUs in CFG1 were subsequently subdivided into smaller units in CFG2 using the 1 km^2 threshold.

Meteorological forcing variables were assigned to each GU using the CEN approach (Section 3.2), whereby the spatially interpolated value at the GU centroid is used as the representative forcing for that unit. Consequently, the finer discretization in CFG2 affects not only the representation of topography and hydrological properties, but also the spatial sampling of the meteorological forcing fields.

Figure 5.1 illustrates the spatial organization of CFG1 and CFG2 and highlights the areas selected for local refinement.

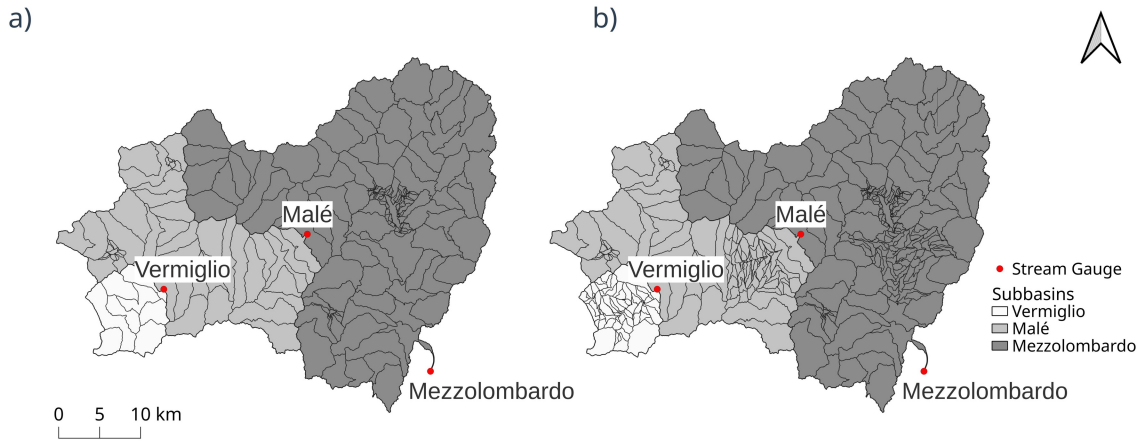

Figure 5.1 – The GUs of the basin. Panel (a) shows the CFG1 configuration, where all GUs are extracted using a 10 km^2 threshold. Panel (b) shows CFG2, which uses a mixed threshold, with some GUs derived with a threshold of 1 km^2. In both cases, the grayscale color ramp indicates the sub-catchment divisions.

The effects of the local refinement on GU geometry and elevation are summarized in Table 5.1. For each of the three refined areas, the table compares the five GUs selected in CFG1 with the smaller GUs resulting from their subdivision in CFG2, reporting the number of units, their average area, and the minimum, mean, and maximum elevations of their centroids. Local refinement substantially increases the number of GUs in the selected

Table 5.1 – Summary of the characteristics of the two configurations (CFG1 and CFG2), divided by sub-catchment. The table compares the characteristics of the original (CFG1) and the resulting refined (CFG2) GUs.

Sub-catchment	**Configuration**	**GUs**	**Mean GU area** [km^2]	**Min Elev.** [m]	**Avg Elev.** [m]	**Max Elev.** [m]
Vermiglio	CFG1	5	10.18	1662	2032	2348
	CFG2	43	1.18	1331	2035	2683
Malé	CFG1	5	11.57	1378	1510	1756
	CFG2	47	1.23	894	1500	2397
Mezzolombardo	CFG1	5	12.05	485	715	1006
	CFG2	57	1.06	340	712	1611

For each sub-catchment and configuration, the table reports the number of GUs, their average area, and the elevation statistics (minimum, average, and maximum) based on the centroid altitude. For CFG1, the number indicates GUs that are later split in CFG2; for CFG2, the number represents the resulting smaller GUs.

areas while reducing their average spatial extent from approximately 10 km^2 in CFG1 to approximately 1 km^2 in CFG2. The finer partition also resolves a broader range of centroid elevations, particularly in areas characterized by strong topographic gradients. Through the CEN forcing approach, this increased spatial detail also results in a finer sampling of the spatial variability of the meteorological forcing.

Figure 5.2 provides a more detailed illustration of this effect. One representative GU from each refined area was selected in CFG1: GU 5561 in Vermiglio, GU 3854 in Malé, and GU 4645 in Mezzolombardo. These GUs are subdivided into 14, 13, and 19 GUs, respectively, in CFG2. Panels (a)–(c) show the resulting spatial subdivisions, while panel (d) compares the mean elevation of each original CFG1 GU with those of the corresponding refined GUs in CFG2.

The catchment is divided into three sub-catchments, each terminating at a stream gauge: Vermiglio, Malé, and Mezzolombardo. Model parameters are calibrated sequentially for the two upstream sub-catchments, Vermiglio and Malé. Calibration based on discharge

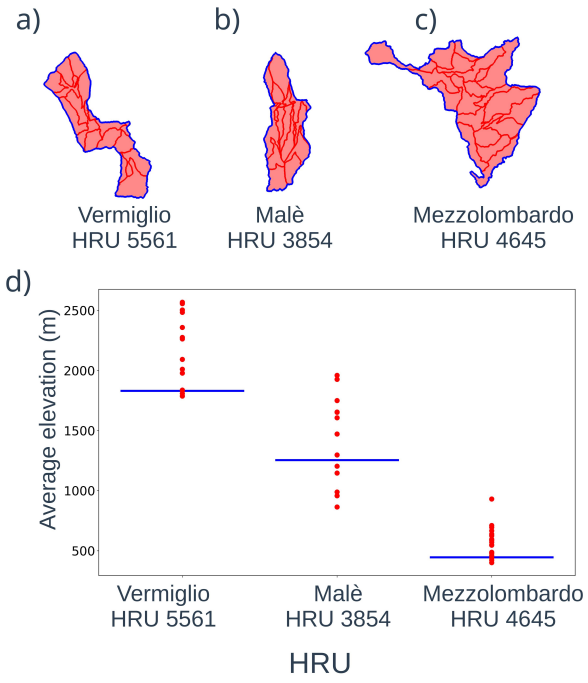

Figure 5.2 – An example illustrating the distribution of average elevation differences across sub-catchments and configurations. In panels (a), (b), and (c), one GU is selected from each sub-catchment (Vermiglio, Malé, and Mezzolombardo) to show how a single GU in CFG1 is subdivided into multiple GUs in CFG2. In panel (d), the average elevation of the GU in CFG1 is represented as a blue line, while the average elevations of the corresponding GUs in CFG2 are depicted as red dots.

is not performed for the downstream Mezzolombardo sub-catchment because reservoir regulation introduces substantial uncertainty in the observed flow regime (Avesani et al., 2021). The parameters calibrated for Malé are therefore transferred to the downstream sub-catchment. Although the Santa Giustina reservoir is represented in the model, the timing and magnitude of reservoir releases are unknown; consequently, discrepancies between simulated and observed discharge at Mezzolombardo cannot be attributed solely to model performance. Evaluation in this part of the basin therefore focuses primarily on other hydrological variables.

The simulation period is divided into spin-up, calibration, and validation periods. The spin-up period extends from 1 January to 31 March 2017 and is used to stabilize the initial model states. Model calibration is performed from 1 April 2017 to 31 December 2019, while the period from 1 January 2020 to 15 September 2022 is used for validation. All 18

model parameters were calibrated during the calibration period.

The hydrological model structure and sequential calibration protocol are presented in Sections 3.3 and 3.4.

For the evaluation of the snow compartment, the snow-presence thresholds applied to simulated SWE and MODIS snow-cover observations were calibrated following the procedure described in Section 3.6.1.

Simulated root-zone water content is first evaluated against measurements from the CRNS probes following the procedure described in Section 3.6.2. For GUs with mean elevations below 1900 m asl, Copernicus surface soil moisture (SSM) is used as an independent dataset for evaluating simulated root-zone soil moisture as described in Section 3.6.3. Since the actual root-zone depth is unknown, the first three years of observations (April 2017–April 2020) are used to calibrate the filter and estimate its characteristic time-length parameter, which implicitly accounts for the soil-moisture dynamics associated with the effective soil depth (Albergel et al., 2008). The resulting values are subsequently aggregated to daily means.

Apart from spatial discretization, CFG1 and CFG2 share the same hydrological model structure, meteorological interpolation and assignment method, parameter bounds, and calibration protocol. Their comparison therefore evaluates the direct and indirect effects of spatial discretization, including changes in the representation of topography and hydrological properties, the spatial sampling of meteorological forcing, and the calibrated parameter values.

5.3 Results and discussion

The model results are evaluated in terms of discharge performance and the representation of snow and soil moisture.

5.3.1 *Hydrological model calibration and validation against measured discharge*

The model was calibrated and validated using discharge data from two upstream hydrometers in the basin: Vermiglio and Malé. Figures 5.3 and 5.4 present the calibration and validation results for configurations CFG1 and CFG2 for Vermiglio and Malé, respectively. Results are shown both as time series comparing measured and simulated discharge, and as goodness-of-fit indices.

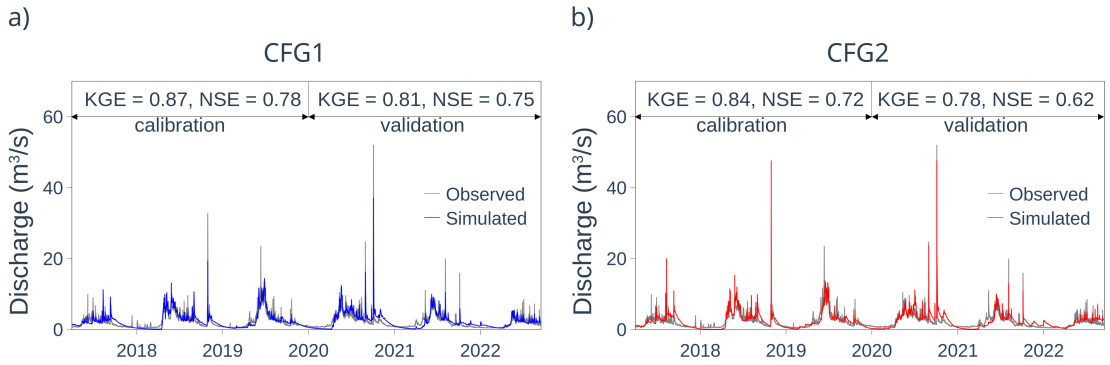

Figure 5.3 – Comparison of observed and modeled discharge at Vermiglio for two configurations: (a) CFG1, with observed data shown in light gray and modeled data in blue; (b) CFG2 with observed data in light gray and modeled data in red. Both plots include calibration and validation periods, with corresponding KGE and NSE values provided for model performance assessment.

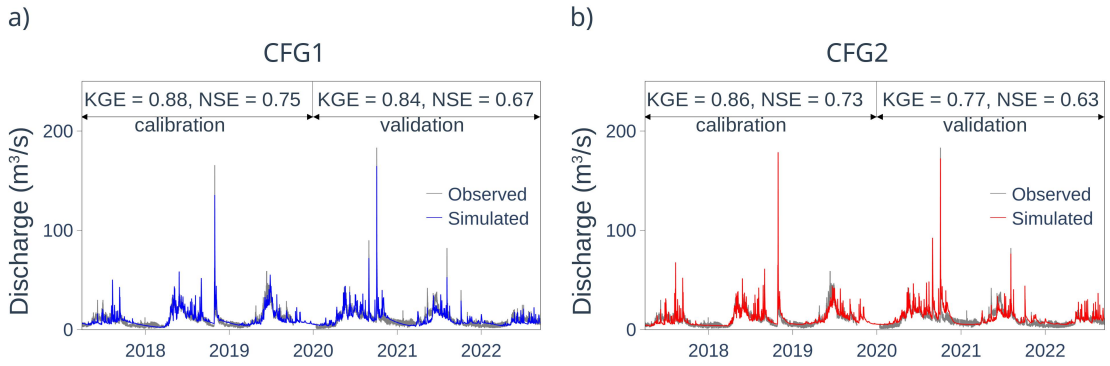

Figure 5.4 – Comparison of observed and modeled discharge at Malé for two configurations: (a) CFG1, with observed data shown in light gray and modeled data in blue; (b) CFG2 with observed data in light gray and modeled data in red. Both plots include calibration and validation periods, with corresponding KGE and NSE values provided for model performance assessment.

Both configurations achieved a KGE value of 0.75 or higher, indicating good model performance according to Kling et al. (2012). A slight deterioration in performance (i.e.,

lower KGE) was observed when transitioning from CFG1 to CFG2. However, this decline was limited to a maximum of 8% during the Malé validation period, and in other cases, the reduction was less than 5%.

The consistency of results is confirmed when performance is evaluated using other indices such as the NSE, which is more sensitive to high flows, and NSE_{log}, which better assesses low-flow simulation ability (Moriasi et al., 2007). The maximum observed performance reduction in terms of KGE between the calibration and validation periods was 10% for CFG2 at Malé, with smaller reductions in other scenarios.

The discharge time series shows two major peaks, each corresponding to one of the storm events discussed in Chapter 2, one occurring during the calibration period and the other during validation. These events, which took place at the end of October 2018 and 2020 at relatively high elevations, added complexity to precipitation estimation and the separation of rain and snow. Additionally, small variations in snowmelt parameters significantly influenced the total discharge volume, either increasing or decreasing it.

In general, CFG2 performs better in 2018. At Vermiglio, the KGE was 0.39 for CFG1 and 0.76 for CFG2; at Malé, it was 0.73 for CFG1 and 0.71 for CFG2. In 2020, performance shifted slightly, with CFG1 outperforming CFG2 at Malé, though the differences were less pronounced. Specifically, at Vermiglio, the KGE was 0.04 for CFG1 and 0.60 for CFG2, while at Malé it was 0.78 for CFG1 and 0.67 for CFG2. Overall, CFG2 yielded notably better performance at Vermiglio, where orographic effects are more significant, while differences at Malé were smaller.

Other studies focusing on this basin, particularly Bellin et al. (2016) and Majone et al. (2016), used NSE as the objective function and applied a daily timestep. These studies generally achieved higher NSE values, except at the Vermiglio station in Majone et al. (2016). Bellin et al. (2016) reports values consistently above 0.8 in the calibration period, while Majone et al. (2016) reported an NSE of 0.69 for Vermiglio and 0.93 for Malé during calibration, and 0.69 for Vermiglio and 0.91 for Malé during the validation period.

The observed differences in model fit highlight the increased sensitivity of hourly-scale hydrological modeling to input biases and more complex hydrological processes (Koltsida et al., 2023).

The ECP analysis indicates that the model is generally reliable. The distance between the highest-probability estimate and the observed discharge is typically small, as shown in Figures D1 and D2. The correlation between these values is strong, especially in the Vermiglio sub-catchment, where it exceeds 0.9 for both configurations. In Malé, the correlation is 0.91 for CFG1 and 0.88 for CFG2, though some noise is observed in the highest-probability discharge for CFG2.

The simulation period includes two major low-flow events affecting the area in 2019 and 2022, as reported in Toreti et al. (2022). The model successfully reproduced most features of these events, although CFG1 showed some difficulties at Vermiglio. Specifically, for the summer of 2019 (June–August), the KGE was 0.42 and 0.79 for CFG1 and CFG2, respectively, at Vermiglio; and 0.83 and 0.64 at Malé. For the 2022 summer event, the KGE was -0.19 and 0.67 at Vermiglio, and 0.45 and 0.75 at Malé for CFG1 and CFG2, respectively.

While the two events differ substantially, CFG1 consistently overestimated minor peak flows and underestimated baseflows in the Vermiglio sub-catchment. This behavior is likely due to CFG1's coarser discretization (only 8 GUs), making it more sensitive to precipitation input. Moreover, changes in precipitation and temperature estimation, as well as in model configuration, had a more pronounced impact on discharge in Vermiglio than in Malé, mostly due to the increased number of GUs, from 8 in CFG1 to 46 in CFG2 (see Table 5.1).

CFG2 more accurately reproduced the basin's hydrological behavior, particularly during the 2022 drought. As shown by Avanzi et al. (2024), in the adjacent Po basin and across northern Italy, this drought was preceded by an SWE drought, marked by quasi-stationary SWE levels, a pronounced minimum SWE volume, and an elevated snow line (Koehler et al., 2022). Similar conditions were observed in the Noce basin and aligned with model outputs,

as illustrated in Appendix Figures D7 and D8. This SWE anomaly contributed to reduced summer discharge, lower groundwater recharge, and diminished root-zone water content.

Alongside the traditional performance metrics, additional insights can be obtained by analyzing cumulative discharge volumes and the associated volume errors. The models generally tend to overestimate discharge, particularly during the validation period. This behavior is more pronounced for CFG2 at the Malé stream gauge, whereas at Vermiglio both configurations exhibit a similar behavior, as shown in Appendix Figure D3, where the observed cumulative discharge is compared with the simulated cumulative discharge for CFG1 (blue) and CFG2 (red) in panels (a)–(d).

To further investigate these differences, discharge volumes were also aggregated by hydrological year (Appendix Table D1), excluding incomplete hydrological years. This analysis reveals that the long-term cumulative divergence is less pronounced than suggested by the continuous cumulative curves. Although a general tendency toward discharge overestimation is still evident, annual volume errors vary from year to year and can also become negative, indicating partial compensation between years. With the exception of a single hydrological year, in which the annual volume error exceeds approximately 12%, the annual errors generally remain below 7% at both stream gauges.

Inspection of the cumulative discharge curves indicates that the divergence mainly originates from two recurring features. The first is a slight overestimation of the recession following peak-flow events, which causes simulated cumulative volumes to progressively exceed the observations. The second is a persistent winter discrepancy, particularly evident at the Malé stream gauge, where the observed discharge exhibits a periodic signal that is not reproduced by the model and may be related to processes not represented in the current model formulation (e.g., anthropogenic factors such as flow regulation). The seasonal origin of these cumulative discrepancies is further confirmed by the monthly volume-error box-plots (Figure D4), which reveal systematic biases concentrated in specific months rather than uniformly distributed throughout the year. The two stream gauges exhibit distinct

seasonal patterns, likely reflecting the different elevation ranges and hydrological regimes of the corresponding catchments. At the Vermiglio stream gauge, the model generally overestimates discharge during the autumn and early winter (snow accumulation period), whereas it tends to underestimate discharge during the snowmelt season, with relatively small biases during summer. In contrast, at the Malé stream gauge the largest positive biases occur during winter, while the model tends to underestimate discharge during summer. These contrasting seasonal behaviors suggest that the cumulative volume differences mainly arise from season-dependent processes rather than from a persistent long-term model drift.

5.3.2 Water balance at the GU scale

As the model provides a set of time series outputs for each GU and for each hydrological process (domains), it is possible to evaluate the water balance at the GU level. Thanks to the hourly resolution of the data, outputs can be aggregated to coarser time steps. For example, Figure 5.5 shows the water balance represented as a bar plot for GU 4829 in CFG1, with data aggregated at both monthly and weekly scales. In these plots, actual evapotranspiration is calculated as the sum of evapotranspiration from the canopy and root-zone domains. Water storage variation is computed as the combined change in snow, root-zone, and groundwater storage. Outflow includes both surface runoff and groundwater discharge.

Figure 5.5(a) displays the water balance on a monthly scale for the year 2022. It highlights:i) the generally low magnitudes of all water balance components compared with previous years and ii) minimal variations in storage, due to a drought that began months earlier, during the winter of 2021/2022. In that winter, storage variations were negative, contrasting with 2021, when they were positive due to significant snowfall.

Figure 5.5(b) illustrates the weekly-scale water balance. While the overall trends are similar to the monthly view, notable differences emerge due to scale effects. For example, in October 2021, the variation in storage appears negligible when aggregated monthly, but becomes clearly noticeable at the weekly scale. A similar pattern is evident in autumn 2022.

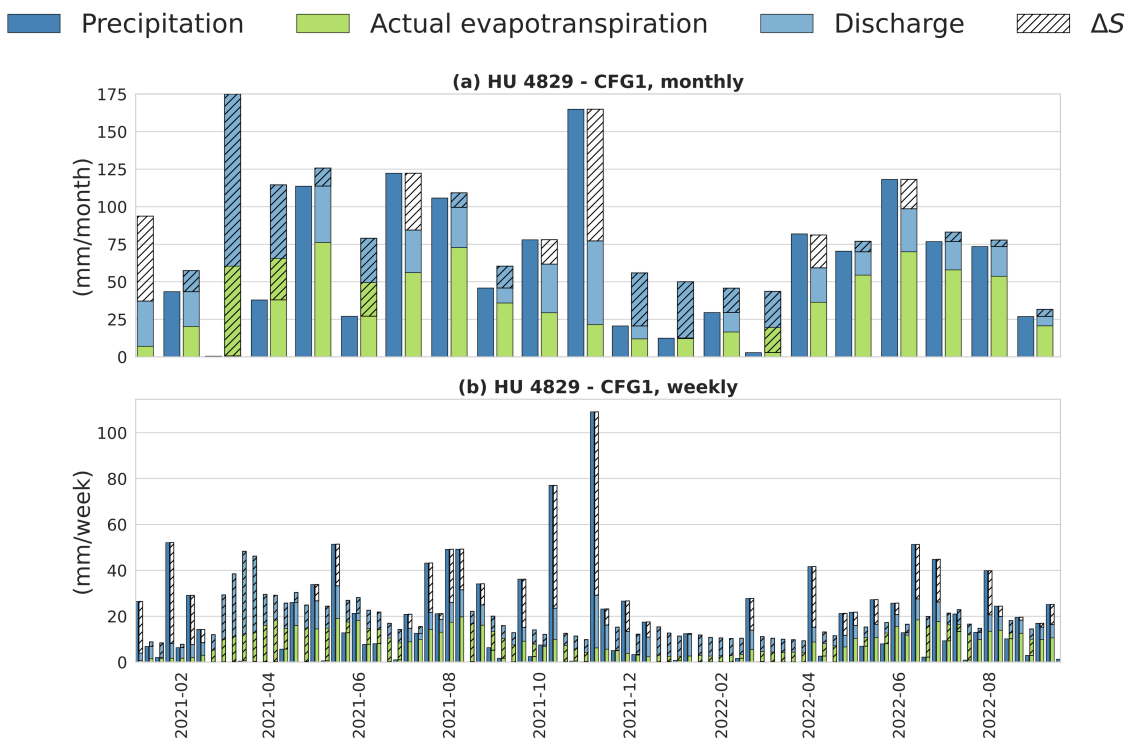

Figure 5.5 – Example of water balance for the GU with ID 4829, located in the last part of the basin downstream of the Santa Giustina reservoir, in the Mezzolombardo sub-catchment. The bar plot reports actual evapotranspiration in green, the variation in total stored water in orange, and the outflow in blue. Panel (a) shows the monthly balance and panel (b) the weekly balance.

The same water balance analysis was carried out for CFG2, where some GUs from CFG1 are spatially disaggregated into finer-scale GUs. As a result, instead of a single water balance, multiple localized balances are produced. Figure D5 presents the water balance bar plots for the eight GUs in CFG2 that spatially correspond to GU 4829 in CFG1.

Although the trends across these GUs are generally consistent, differences emerge in the magnitude of the balance components, especially at the extremes. For instance, GU 4829004 and GU 4829033 consistently show the highest or lowest values for key variables. The maximum monthly evapotranspiration ranges from 45 mm in GU 4829012 to 60.2 mm in GU 4829004. Similarly, storage variation maximum absolute value spans from 121 mm in GU 4829005 to 159 mm in GU 4829033. In contrast, outflow is higher in GU 4829033 (118 mm) than in GU 4829004 (98 mm). These variations are primarily

influenced by temperature and aspect for evapotranspiration, and by temperature alone for storage, especially in the snow compartment. This is particularly evident during heavy autumn precipitation events, where rain and snowfall are partitioned based on temperature. Differences in outflow reflect the combined effects of evapotranspiration and water storage processes.

The analysis becomes more robust when the water balance is evaluated at the hydrological-year scale and aggregated over the entire basin, as shown in Appendix Figure D6. The water balance components, grouped as described above, are summed over all GUs upstream of the Malé stream gauge and normalized by the basin area for each hydrological year (October–September). For each hydrological year, the first bar represents the total precipitation, whereas the second stacked bar reports the partitioning of water into actual evapotranspiration (green), discharge (light blue), and storage change (transparent with diagonal hatching). The storage change may be either positive or negative, and its value is reported above the stacked bar. Panels (a)–(d) show the results for the different model configurations. At this temporal and spatial scale, the annual water balance is dominated by discharge, followed by actual evapotranspiration, whereas the annual storage change represents the smallest contribution. Nevertheless, the sign and magnitude of the storage change provide useful information on the interannual wetting and depletion of basin water storage. The figure also highlights the evolution of the 2022 drought. A negative storage change is already observed during the 2021 hydrological year, indicating a depletion of basin water storage before the drought peak. In contrast, the 2022 hydrological year shows a slightly positive storage change, mainly due to the autumn precipitation. However, both the precipitation input and the output fluxes (discharge and actual evapotranspiration) are considerably lower than in the other hydrological years.

We also analyzed the impact of the 2022 low-flow period on changes in mean root-zone and groundwater storage compared to the summer of 2021. These values for the Vermiglio, Malé, and Mezzolombardo sub-catchments, under CFG1 and CFG2, are summarized in

Table 5.2. The results indicate that, in both configurations, the effects of drought on groundwater recharge and root-zone water storage were more pronounced at higher elevations, especially up to the Vermiglio stream gauge.

Basin-mean time series of SWE, root-zone water content, and groundwater storage are shown in Appendix Figures D10–D12. The Appendix additionally shows the spatial distributions of the same variables at representative winter and summer dates (Figs. D13–D15).

Table 5.2 – Changes in mean root-zone water storage and groundwater storage from summer 2021 to the 2022 low-flow period for the Vermiglio, Malé, and Mezzolombardo sub-catchments under CFG1 and CFG2.

Sub-catchment	**Configuration**	**Root Zone**	**Groundwater**
Vermiglio	CFG1	-20%	-14%
	CFG2	-18%	-23%
Malé	CFG1	-7%	-14%
	CFG2	0%	-4%
Mezzolombardo	CFG1	-11%	-12%
	CFG2	-8.6%	-1.7%

Comparison of values during the summer months of 2022 compared to summer 2021, for the Vermiglio, Malé, and Mezzolombardo sub-catchments, under both CFG1 and CFG2 configurations.

The actual evapotranspiration (AET) reported by Tovazzi (2013) for the years 2001 to 2006 ranged from 250 to 380 mm/year. In comparison, our results for CFG1 range from 316 to 490 mm/year, and for CFG2 from 263 to 431 mm/year. Although these values are not directly comparable due to differences in time periods and climate variability, they are within a reasonable range. As noted by Morlot et al. (2023), AET has shown a positive trend, which supports the reliability of the current values.

The lowest AET was observed in 2022, coinciding with the drought period, when root-zone water content remained low for an extended time. This evapotranspiration deficit is illustrated in Figures 5.5 and D5. For example, in April 2021, abundant precipitation led to increased root-zone water storage (see Figure D9 in the appendix), resulting in significantly higher evapotranspiration compared to April 2022.

5.3.3 Modeled snow assessment

The initial phase of the snow analysis focuses on comparing MODIS NDSI data with model-derived SWE to evaluate the thresholds for snow presence or absence. The calibrated threshold values are 38 for NDSI and 6 mm for model-derived SWE. Using these thresholds, binary series outcomes were analyzed to assess several performance metrics, including false negatives, false positives, overall accuracy, and the Brier score on a global scale (see Table 5.3). The Brier score ranges from 0, indicating perfect agreement, to 1, representing the poorest predictive performance (Brier, 1950).

Table 5.3 – Snow-cover detection metrics for the two configurations (CFG1 and CFG2).

Configuration	**True Pos.**	**False Pos.**	**True Neg.**	**False Neg.**	**Accuracy**	**Median Brier**
CFG1	0.30	0.027	0.49	0.18	0.79	0.20
CFG2	0.27	0.060	0.57	0.10	0.84	0.14

This table reports classification metrics, including true positives, false positives, true negatives, false negatives, overall accuracy, and the median Brier score.

Figure 5.6 presents the Brier score evaluation for each GU. The box-plots categorize values into elevation classes, which are computed to be equinumerous, ensuring each class contains the same number of elements. The map highlights the differences between the two GU configurations.

Numerous studies (Morlot et al., 2023; Parajka & Blöschl, 2006, 2008; Tong et al., 2020) have demonstrated that determining an appropriate NDSI threshold for snow presence or absence is a complex task. A detailed study over the Austrian region (Tong et al., 2020) found that this threshold varies significantly with both time (i.e., across seasons) and space (i.e., across elevation bands). Nevertheless, many applications adopt a fixed threshold; for instance, Härer et al. (2018) and Morlot et al. (2023) use a single value. The threshold identified here, set at 38, closely aligns with the conventional threshold of 40 as reported by Härer et al. (2018). Similarly, the SWE threshold of 6 mm was selected to ensure accurate snow cover delineation, closely approximating the 5 mm used in previous research (Dawson et al., 2018; Di Marco et al., 2020).

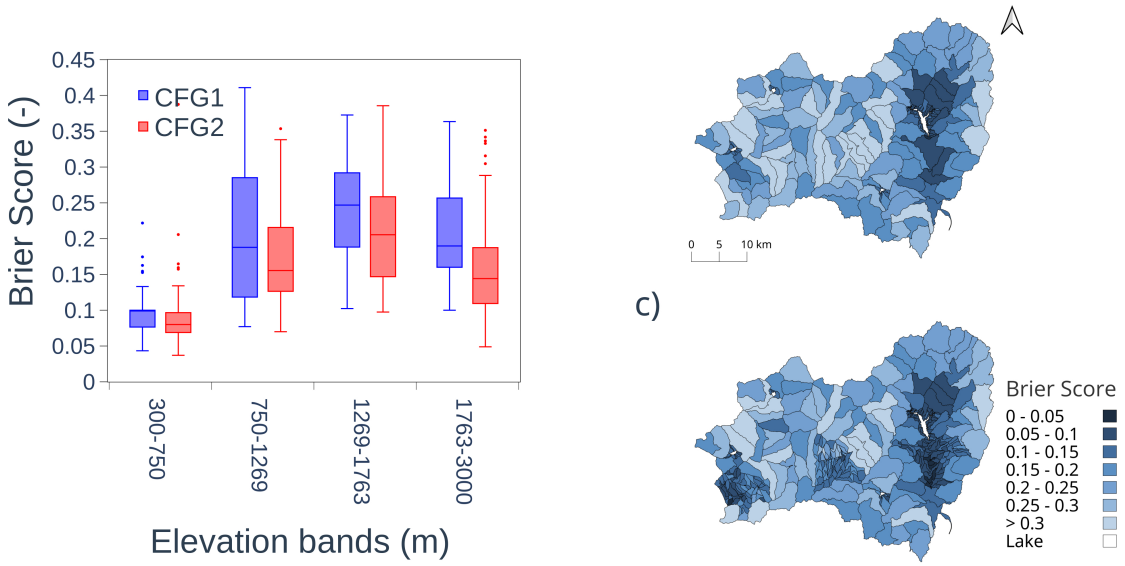

Figure 5.6 – Comparison between the snow presence/absence estimated from MODIS and from the model in terms of the Brier score. (a) Box-plots of the Brier score of each GU for CFG1 (blue) and CFG2 (red), split by elevation bands with an equal number of GUs in each band. (b) Map of the Brier score for CFG1. (c) Map of the Brier score for CFG2. Brier score: 0 = perfect; higher values = worse.

When comparing the two configurations (see Table 5.3), CFG2 generally outperforms CFG1 in terms of both the Brier score and overall accuracy. The global statistics, particularly the Brier score, show improved performance with CFG2. These results are comparable to those of a related study (Gyawali & Bárdossy, 2022), where model snow parameters were directly calibrated to reduce the Brier score, significantly enhancing model performance.

CFG2 also shows better results in overall accuracy, with scores of 0.79 for CFG1 and 0.84 for CFG2. Similar findings are reported in other studies; for example, Tong et al. (2020) achieved an accuracy index of 0.90 when comparing the same MODIS product to actual snow measurements. The model exhibits a slight underestimation of snow presence compared to MODIS, though this effect is less pronounced in CFG2, leading to fewer false negatives in this configuration.

An analysis of model performance across elevation bands reveals a clear pattern. At lower elevations, the accuracy score is higher, primarily due to the limited number of snow days. At mid-elevations, accuracy decreases as the snowpack becomes more ephemeral,

posing challenges for the model. In this range, true positive and true negative rates are balanced, with penalties mostly arising from false negatives, 0.22 in CFG1 and 0.15 in CFG2. At higher elevations, where colder temperatures support snow accumulation, model performance improves significantly. As illustrated in Figures D7 and D8, CFG1 shows a higher maximum SWE value during winter but a shorter snow season length compared to CFG2.

5.3.4 Modeled soil moisture assessment

The CRNS-measured count rates are converted into soil moisture values using the calibration process described in Section 3.6.2. Table 5.4 shows the optimal parameters used for the conversion (based on Equation 3.7), derived from neutron counts and gravimetrically retrieved soil moisture during a specific field campaign carried out for this work. These values are consistent with those reported in other studies using the same sensor (Gianessi et al., 2024).

Table 5.4 – Results of the field campaign and calibration of the CRNS probes.

Probe Name	**Bulk Density** [g/cm^3]	**Reference Count Rate**	**Count Rate**
TRES1	1.2	1419	2516
TRES2	1.4	1372	2241

The table shows the bulk density obtained from soil sampling, the count rate corrected and normalized for solar activity, and the reference count rate calculated using Equation 3.7.

Figure 5.7 presents the comparison between modeled and CRNS-derived soil moisture for TRES1. Results are shown for CFG1 and CFG2 as scatter plots (Figures 5.7a and 5.7c) and as time series (Figures 5.7b and 5.7d).

Figure 5.8 shows the same comparisons for TRES2. The model successfully captures the temporal dynamics of soil moisture as measured by CRNS across both configurations (CFG1 and CFG2) and for both TRES1 and TRES2. In terms of correlation, CFG2 demonstrates a higher Pearson correlation coefficient (r) than CFG1, with an average increase of 9% when moving from CFG1 to CFG2.

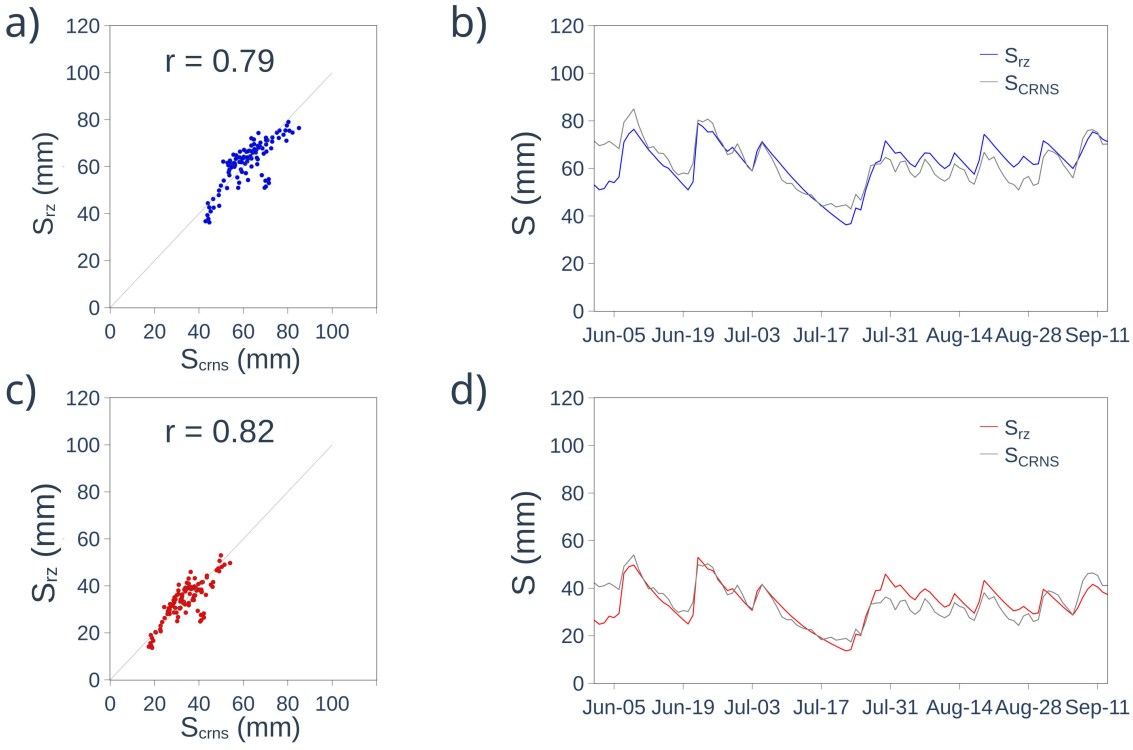

Figure 5.7 – Relation between the water storage estimated with the TRES1 CRNS probe and the modeled root-zone water storage. (a) Pearson correlation of CFG1, (b) time series of CFG1, (c) Pearson correlation of CFG2, and (d) time series of CFG2.

Both the model and CRNS exhibit similar magnitudes in their variation range, effectively capturing root-zone soil moisture dynamics. The comparison between modeled and CRNS-derived root-zone water content produced reasonable correlation coefficients (ranging from 0.66 to 0.82). This performance is especially notable considering that: (i) the model is applied over a large and heterogeneous area with varying land use, land cover, and soil types, and (ii) no explicit tuning of the root-zone model parameters to match CRNS measurements was performed. In comparison, Katya et al. (2020) reported a correlation coefficient of 0.96 between modeled soil water content and CRNS values. However, their study used a lumped conceptual model over a smaller basin ($\sim$10 km^2) with uniform soil and land cover types and included five separate sampling campaigns for CRNS calibration.

Differences between modeled root-zone soil moisture and CRNS measurements may also be influenced by irrigation near the sensor locations, for which no information was

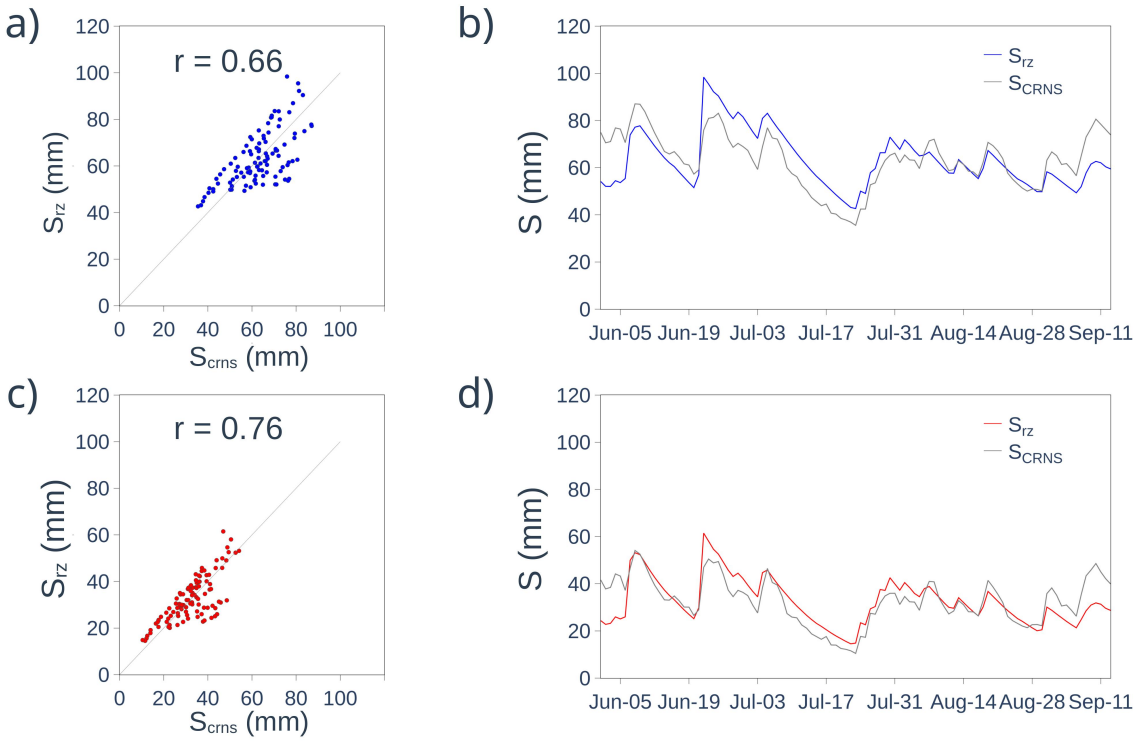

Figure 5.8 – Relation between the water storage estimated with the TRES2 CRNS probe and the modeled root-zone water storage. (a) Pearson correlation of CFG1, (b) time series of CFG1, (c) Pearson correlation of CFG2, and (d) time series of CFG2.

available. However, the primary vegetation in the area is orchard-based, irrigated with drip systems that use minimal water and operate for a limited duration (maximum of 2 hours per day). As reported by D. Li et al. (2019), the effect of such irrigation on CRNS measurements is comparable to the natural statistical fluctuations in CRNS counts, thus allowing for reliable comparison with model results.

Analysis across the spatial configurations indicates that CFG2 outperforms CFG1. This improvement may be attributed to CFG2's higher resolution and more detailed precipitation input, which enhance the modeling of root-zone moisture dynamics, where precipitation is the primary driver.

For the Copernicus SSM, the estimated mean correlation coefficient (minimum; maximum) across all basins is 0.28 (0.12; 0.57) for CFG1 and 0.36 (0.12; 0.58) for CFG2. Figure 5.9a illustrates these results grouped by elevation classes, with CFG1 in blue and

CFG2 in red. The classes were designed to be equinumerous, ensuring each contains approximately the same number of elements. Spatial variations in the correlation coefficient are shown in Figures 5.9b and 5.9c for CFG1 and CFG2, respectively.

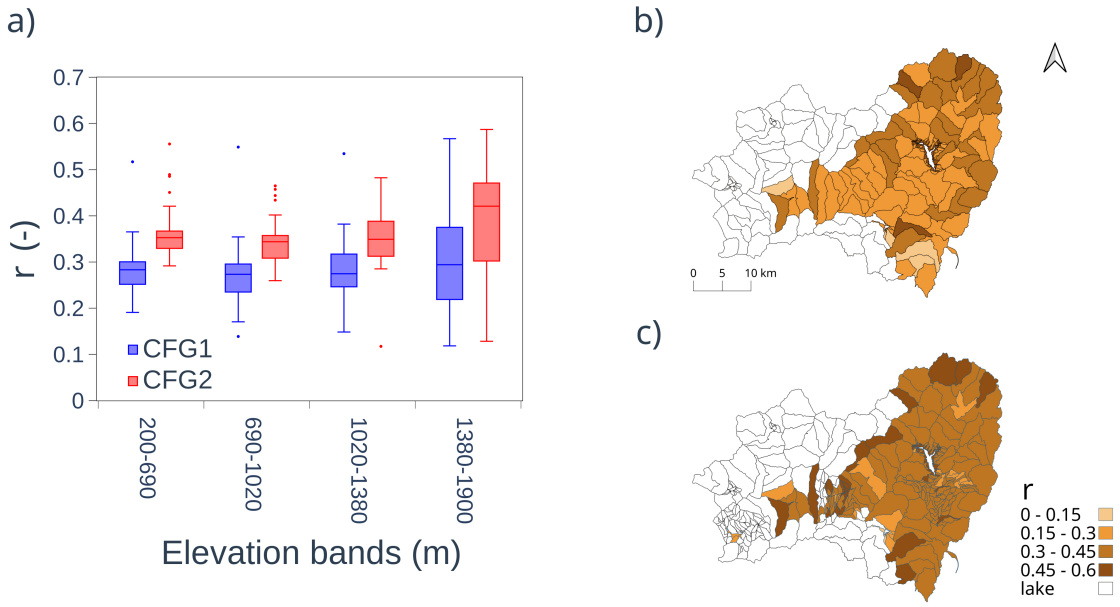

Figure 5.9 – Comparison between modeled and Copernicus root-zone water storage. (a) Boxplots of the correlation coefficient for each GU, grouped by elevation bands, for CFG1 (blue) and CFG2 (red). Spatial variation of the correlation coefficient for CFG1 (b) and CFG2 (c).

In this context, the transition from CFG1 to CFG2 shows noticeable improvements. The enhanced median correlation in CFG2 suggests that refinements in parameterization and the spatial distribution of precipitation in this configuration lead to better alignment with satellite-based observations. The correlation coefficients obtained in this study are consistent with, but slightly lower than, those from previous research (Brocca et al., 2013; Morlot et al., 2023), which compared modeled and satellite-derived soil moisture products over the same or nearby areas. However, those studies considered larger catchments, used daily time steps, relied on different datasets (e.g., Advanced Scatterometer, ASCAT), and covered significantly longer simulation periods (up to 38 years). It is important to note that Morlot et al. (2023) found the highest correlation values in the flat, low-elevation area downstream of the confluence of the Noce and Adige Rivers.

5.4 Conclusion

In this study, we applied a semi-distributed hydrological model within the GEOframe framework to an Alpine region, using two different configurations of GUs. The first configuration employed a fixed spatial resolution, delineating GUs with a uniform area threshold. The second used a mixed spatial resolution, refining selected GUs in hydrologically significant zones by applying a smaller area threshold. Both configurations were calibrated using observed discharge data and validated against multiple independent datasets: (i) hourly soil water content estimates derived from ground-based CRNS, and (ii) satellite products, including snow-covered area from MODIS and Copernicus SSM. The multi-source validation demonstrates that incorporating higher spatial detail enhances the model's ability to reproduce key hydrological dynamics, such as snow cover and root-zone soil moisture, while maintaining an acceptable level of discharge performance. Notably, the validation of root-zone water content against CRNS data is particularly valuable, as the hourly resolution captures short-term soil moisture dynamics more effectively than coarser temporal datasets. This study demonstrates that strategically increasing spatial resolution can significantly enhance hydrological modeling accuracy without incurring proportional increases in computational cost. This approach presents a promising path for improving water budget estimation in data-scarce mountainous regions.

This research offers an improved framework for Alpine hydrological modeling. Instead of applying uniformly high spatial resolution across entire basins, which is often computationally prohibitive, the study shows that targeted refinement of critical areas (CFG2) yields substantial improvements in process representation while maintaining overall model efficiency. Furthermore, process-specific improvements (e.g., snow dynamics and soil moisture) can be achieved without substantially degrading streamflow simulations or increasing model instability. From a multi-process model evaluation perspective (snow and soil moisture dynamics), our study demonstrated that multi-scale validation using CRNS and satellite

observations significantly strengthens confidence in model outputs, bridging the gap between measurements at different spatial scales. Furthermore, this study also underscores that different hydrological processes respond distinctly to changes in spatial resolution. Snow accumulation and melt, along with root-zone soil moisture dynamics, benefit markedly from finer discretization. In contrast, streamflow simulations remain robust across varying resolutions. The model simulations allowed an in-depth analysis of the 2022 drought, which revealed critical insights into the resilience of Alpine hydrological systems. Notably, elevation-dependent vulnerability was observed, with higher-altitude regions showing greater sensitivity to hydro-meteorological anomalies. This spatially explicit understanding is essential for climate adaptation planning in mountain environments. However, the study also highlights some limitations in the hydrological data and the model used. First, validation using remote sensing data is challenged by temporal mismatches: while model outputs and CRNS data are hourly, MODIS and Copernicus products provide data at daily or multi-day intervals. Second, differences in sensing depth complicate comparisons: satellite products reflect moisture in the top $\sim$5 cm of soil, CRNS measurements represent a variable depth (typically several decimeters), and the model simulates root-zone moisture without a fixed physical depth. To enable comparison, an exponential filter was applied to harmonize the datasets, introducing an additional layer of uncertainty. Finally, uncertainties in the spatial distribution of meteorological inputs also affect results. Most weather stations are located at lower elevations, reducing the representativeness of interpolated data in higher-altitude zones, which can impact snow accumulation and melt estimates. Additionally, temperature interpolation showed sensitivity to GU size: smaller GUs in complex terrain better captured elevation-dependent variability, influencing critical processes such as snow dynamics and evapotranspiration.

The limitations of this study are discussed together with those of the other two studies in Section 8.3. The computational cost of the two configurations, and of the finer CFG3 used as a reference, is reported in Appendix E; the supplementary uncertainty and mapped-output

figures are in Appendix D.

CHAPTER 6. EFFECTS OF SPATIAL REPRESENTATION OF METEOROLOGICAL DATA ON DIFFERENT HYDROLOGICAL PROCESSES IN SEMI-DISTRIBUTED HYDROLOGICAL MODELS

This chapter presents the second of the three studies comprising the thesis. It introduces an automated GEOframe component for semivariogram-based interpolation of meteorological data and investigates how forcing representation and Geomorphic Unit delineation affect the simulation of different hydrological processes in an Alpine catchment.

Scope and original contributions

The spatial variability of meteorological inputs strongly influences hydrological-model performance, particularly in Alpine environments characterized by rapid changes in topography and climate. Representing this variability requires decisions both about how meteorological observations are spatially interpolated and about how the resulting fields are assigned to the spatial units of the hydrological model.

On the instrumental side, this study introduces an automated GEOframe component that computes an empirical semivariogram and fits a theoretical model at each time step. When the time-specific fit fails, the procedure applies a documented fallback based on a global variogram. The component also includes leave-one-out cross-validation, ensuring that the interpolated fields are evaluated using the same procedure through which they are generated. Its computational implementation allows hourly meteorological fields to be interpolated over eight years and across multiple spatial configurations (Section 4.3). This time-varying treatment is particularly relevant for Alpine precipitation, whose spatial structure may differ substantially between convective and stratiform events and may therefore not be adequately represented by a single variogram.

On the methodological side, the study separates two related questions: how accurately

the meteorological fields are reconstructed, evaluated geostatistically and independently of discharge; and how their spatial representation affects the hydrological simulations. Nine configurations are examined by combining three forcing strategies—nearest station, interpolation at the GU centroid, and grid-weighted averaging—with three GU delineations. The configurations are calibrated using identical initial parameter values, parameter bounds, and computational budgets, ensuring that differences in performance are not attributable to unequal optimization effort.

The results indicate that the configuration combining the most spatially detailed forcing representation with the finest GU delineation provides the most stable discharge performance and the closest agreement with remotely sensed snow-cover observations. Simpler configurations nevertheless produce reasonable results, demonstrating the overall robustness of the modeling framework. By considering both meteorological interpolation and model discretization, the analysis distinguishes their respective effects on different hydrological processes rather than providing only an aggregate ranking of model performance.

6.1 Introduction

Meteorological variables are fundamental drivers of environmental processes and are essential for a wide range of applications, from scientific research to practical uses such as climate analysis, water resource management, and ecosystem studies. In hydrology, precipitation and temperature play particularly critical roles in controlling the processes underlying rainfall-runoff modeling. Their magnitude, quality, and spatial distribution strongly affect model performance and outcomes, making the availability of high-quality data essential (Bardossy & Das, 2008; K. Beven & Westerberg, 2011; De Lavenne et al., 2016; Wilby et al., 2017; Zhu et al., 2018).

This requirement is especially important in Alpine catchments, where complex topography produces pronounced spatial variability that strongly influences hydrological modeling across multiple spatial and temporal scales (Evin et al., 2024; Ledesma & Futter, 2017;

Vaze et al., 2011). Precipitation, in particular, is a complex phenomenon influenced by multiple factors and characterized by strong spatio-temporal variability, especially at high temporal resolutions (Berndt & Haberlandt, 2018; Emmanuel et al., 2012; Hu et al., 2019). Its impact on model performance is therefore critical. In contrast, temperature—which in Alpine contexts directly affects all snow-related processes—generally exhibits lower spatial variability and is thus easier to represent (Berndt & Haberlandt, 2018; Orlowsky & Seneviratne, 2014).

In semi-distributed models, where the Geomorphic Unit (GU) is often used as the basic computational unit, two main questions arise: i) how can the spatial variability of meteorological forcings be represented and evaluated?; ii) how does this variability influence hydrological model performance? Traditionally, meteorological inputs are assigned to each GU either from the nearest weather station or from the GU centroid (Jiang et al., 2024; Ledesma & Futter, 2017; Morlot et al., 2023). In both cases, however, there is a risk of spatial misrepresentation, especially in regions with challenging topography and large elevation ranges. For instance, a weather station located at the upper or lower elevation boundary of GUs, or a centroid situated in a narrow valley or depression, may not be representative of the entire area, leading to biased data and a poor representation of spatial heterogeneity. To overcome these limitations, an alternative is to use gridded meteorological fields (Evin et al., 2024; Jiang et al., 2024; Ledesma & Futter, 2017; Loritz et al., 2021; Mankin et al., 2025; Poblete et al., 2020; Vu et al., 2012) derived from interpolation of ground-station measurements, remote sensing datasets, or reanalysis products that combine multiple data sources using advanced statistical techniques. The relative influence on the model performance of these methods, gridded versus point, is not trivial, as reported by Loritz et al. (2021), and depends on several factors, including catchment state and structure, and the precipitation event characteristics. In some cases, the role of spatial information in the inputs may be attenuated—or even dissipated—by hydrological processes, and simpler lumped models driven by single-point data can outperform more complex

distributed models. In other cases, studies report significant improvements when gridded inputs are used. Moreover, model structural error also plays a crucial role: parameter estimates obtained during calibration often compensate for structural deficiencies and input uncertainty, meaning that the optimal parameters are not necessarily the true ones (Todini & Biondi, 2017). Most studies evaluating the spatial distribution of meteorological forcings examine how different representations—particularly gridded versus point-based inputs—influence hydrological model outcomes across spatial scales and resolutions. In some cases, multiple models are applied to compare their responses to alternative forcing strategies. However, the majority of these studies are conducted at daily or monthly temporal scales (Ledesma & Futter, 2017; Tuo et al., 2016; Vaze et al., 2011; Vu et al., 2012) and rely primarily on the NSE for discharge calibration. Spatial resolutions are typically on the order of several kilometers. Only a limited number of studies explore finer spatial and temporal resolutions or adopt alternative performance metrics and complementary validation datasets. An example is Loritz et al. (2021), who employed the KGE using meteorological forcing at a 1 km $\times$ 1 km resolution.

Based on this context, the present study addresses the following research questions:

- Is there an effective method for evaluating meteorological forcings that preserves more information—including spatial variability—while producing values that are truly representative at the GU scale?

- How does the effect of this forcing strategy depend on the spatial scale of Geomorphic Units (e.g., delineation threshold or aggregation level), and how consistently does it propagate across model outputs—streamflow as well as internal fluxes and state variables?

In this study, we address these questions by testing alternative meteorological forcing strategies and evaluating their effects across multiple model outputs. Specifically, we investigate how the GEOframe framework (Abera et al., 2017; Formetta et al., 2014a,

2014b)—a collection of tools for hydrological modeling—can be used and configured to i) prepare gridded inputs and ii) to compare the outcomes of a semi-distributed model obtained using gridded inputs with those obtained using GU centroid data and nearest-station data.

We assess whether adopting more computationally demanding approaches to improve the spatial representation of meteorological forcings is sustainable and justifiable in terms of model performance. Although several gridded datasets of meteorological variables (e.g., temperature, precipitation, evapotranspiration) are already available, typically produced through different interpolation techniques or by combining ground-based and remote sensing data, they often suffer from limitations such as coarse temporal resolution and/or delayed publication. These drawbacks limit their suitability for near-real-time applications such as forecasting and now-casting, motivating the need for alternative, on-demand approaches.

Among the techniques commonly used to generate spatially continuous meteorological fields, geostatistical methods—and in particular kriging—are widely applied. Originally developed in mining engineering in the 1950s (Cressie, 1990; Krige, 1951), kriging has since been extensively adopted in environmental sciences (Hengl, 2007; Webster & Oliver, 2007), and is frequently used—sometimes in more advanced variants—for estimating precipitation (Berndt & Haberlandt, 2018; Berndt et al., 2014; Formetta et al., 2024; Lucas et al., 2022) and temperature (Berndt & Haberlandt, 2018; Brunetti et al., 2014; S. Li et al., 2020; Wu & Li, 2013). Owing to its flexibility, kriging is well suited both for predicting values at unsampled locations and for producing gridded fields. In this study, we apply a consistent kriging setup to derive both gridded meteorological forcings and point-based estimates at GU centroids.

Kriging is based on the concept of the semivariogram, which accounts for the spatial autocorrelation between pairs of observed values. In hydrological applications, semivariograms are often computed at each time step, and for long or high-frequency datasets, this results in thousands of semivariograms, making automation essential (Haberlandt, 2007;

Lucas et al., 2022). The GEOframe kriging module has been under continuous development since its introduction (Formetta et al., 2014a) and subsequent improvements and extensions (Bancheri et al., 2018). It supports both the ordinary kriging (OK) and detrended kriging (kriging with external drift or regression kriging, DK) and includes several options such as the logarithmic transformation and the localized prediction model. It has proven robust in various applications (Azimi et al., 2023; Bouabdelli et al., 2025; Dall'Amico et al., 2025; D'Amato & Rigon, 2025; Morlot et al., 2023).

In earlier versions, the kriging workflow required multiple user-driven steps to evaluate the best semivariogram: i) compute the experimental variogram, ii) fit a theoretical model in a separate module, and iii) select the best theoretical semivariogram by comparing candidates. For long or hourly series, this became impractical and often led to using variograms calibrated on aggregated periods. In this study, the kriging package was enhanced to automate variogram estimation. In addition, a new implementation of the module to compute the validation based on the leave-one-out method was developed. The package has also been refactored to improve maintainability and flexibility, following SOLID (Single responsibility, Open/closed, Liskov substitution, Interface segregation, and Dependency inversion) design principles (Sarcar, 2022) and is now covered by an automated test suite.

Before conducting rainfall-runoff model experiments, we therefore perform a dedicated assessment of the robustness of the kriging module, focusing in particular on semivariogram generation and prediction validity through leave-one-out cross-validation applied to the meteorological station network. The validated kriging framework is then employed in hydrological modeling experiments to evaluate meteorological forcing represented both at GU centroids and as area-weighted averages derived from a finer $500\,\mathrm{m} \times 500\,\mathrm{m}$ gridded field, considering three different GU delineation strategies. Discharge is subsequently calibrated using the Kling–Gupta Efficiency (KGE), and validated using KGE and additional performance metrics.

Moreover, following recent recommendations on the integration of complementary data

sources (Wagner et al., 2025), model validation is extended beyond discharge to include remote sensing products such as snow cover and surface soil moisture.

6.2 Experimental setup

This chapter focuses on the upper Noce basin down to Malé (Section 2.1.1) over the period 1 August 2015–31 December 2023.

The experimental design was based on the combination of three GU delineations and three meteorological forcing representation approaches, resulting in nine model configurations. The three GU delineations were selected to represent contrasting levels of spatial discretization. The objective of this study was to evaluate the influence of different meteorological forcing representations under markedly different spatial discretizations, rather than to identify an optimal number of GUs or perform a comprehensive sensitivity analysis of the discretization level. This experimental design makes it possible to assess whether the relative performance of the forcing representation methods remains consistent across substantially different GU configurations.

Three GU delineations were considered, as described in Section 3.1. The first, named SG (stream gauge), consists of two GUs defined so that a stream gauge is located at the outlet of each GU, providing the coarsest spatial discretization. The other two, named GU20 and GU10, were derived using the constant-threshold geomorphological approach described in Section 3.1.1, with contributing-area thresholds of 20 km^2 and 10 km^2, respectively.

The three delineations consist of 2, 30, and 51 GUs, respectively. The number of GUs for each configuration and their main characteristics are presented in Table D2 (Appendix D).

All three meteorological forcing representation approaches described in Section 3.2 were considered: NEAR, CEN, and GRIDW. These approaches were applied to precipitation, air temperature, solar radiation, and potential evapotranspiration. For CEN and GRIDW, temperature and precipitation were interpolated using detrended kriging with local predictions based on the ten nearest stations, as described in Section 3.2.1. For GRIDW, the interpolated

variables were estimated over a regular grid with a spacing of 500 m and aggregated to each GU using an area-weighted mean. No areal reduction factor was applied to precipitation ($\mathrm{ARF} = 1$).

Combining the three GU delineations with the three meteorological forcing representation approaches resulted in nine configurations: SG-NEAR, SG-CEN, SG-GRIDW, GU20-NEAR, GU20-CEN, GU20-GRIDW, GU10-NEAR, GU10-CEN, and GU10-GRIDW.

The resulting combinations are summarized in Figure 6.1. The columns represent the meteorological forcing estimation methods, whereas the rows correspond to the GU delineation approaches. The first column illustrates the NEAR method, in which meteorological inputs are assigned from the nearest weather station (black triangles) to each GU centroid (black circles); the corresponding centroid–station connections are shown as orange lines. Note that in the NEAR case, in particular in the GU20 and GU10 configurations, several GUs can share the same weather station and therefore receive identical meteorological input data. The CEN maps show the GU centroids where the meteorological variables are estimated, whereas the GRIDW maps display the regular grid points (small black circles) where the variables are estimated before being aggregated to each GU.

For each configuration, precipitation, air temperature, radiation, and potential evapotranspiration (PET) were evaluated at the GU level and used for model calibration and validation, enabling a comparative assessment of model performance.

The overall workflow—from input preparation through calibration and validation for the nine configurations—is summarized in Figure 6.1b. The figure schematically represents the sequence of operations carried out in this study, from the definition of the nine configurations to the evaluation of the meteorological variables and the calibration and validation of the rainfall-runoff model.

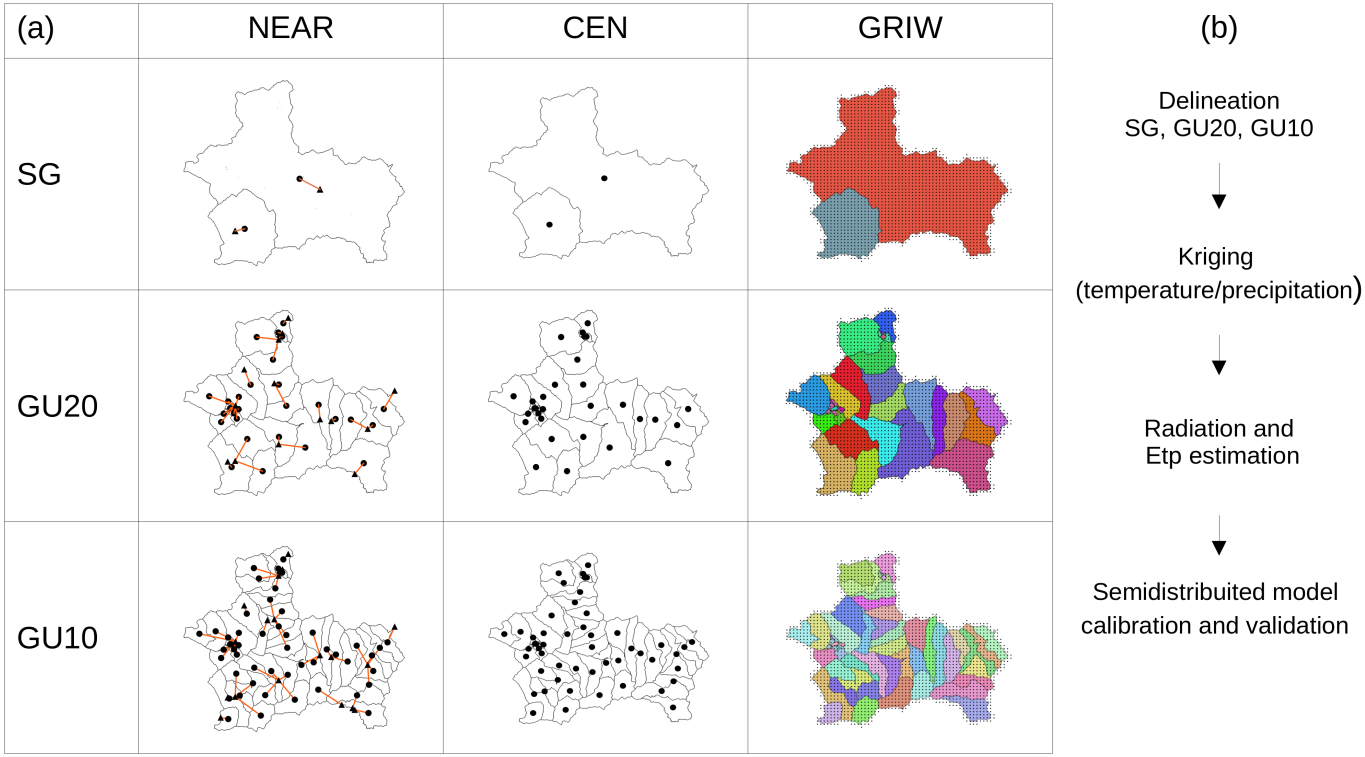

Figure 6.1 – Overview of the configurations adopted in this study. (a) Matrix of the nine modeling configurations obtained by combining the three GU delineation methods (SG, GU20, and GU10; rows) with the three approaches for estimating meteorological forcing (NEAR, CEN, and GRIDW; columns). The resulting combinations are SG-NEAR, SG-CEN, SG-GRIDW, GU20-NEAR, GU20-CEN, GU20-GRIDW, GU10-NEAR, GU10-CEN, and GU10-GRIDW. For the NEAR approach, black circles indicate the GU centroids, black triangles represent the meteorological stations, and orange lines connect each centroid to its nearest station (only the stations effectively used are shown). For the CEN approach, black circles indicate the GU centroids, where meteorological variables are estimated. For the GRIDW approach, small black circles represent the regular grid points where meteorological variables are estimated before being aggregated to each GU. (b) Workflow of the modeling framework, from GU delineation to the estimation of precipitation, air temperature, solar radiation, and potential evapotranspiration, followed by rainfall-runoff simulation using the hydrological model.

Each configuration was calibrated independently at the Malé stream gauge using identical initial parameter values, parameter bounds, and computational budgets. Discharge observations at the upstream Vermiglio stream gauge were used for spatial validation, providing an evaluation at an internal location not involved in calibration. The simulation period was divided into spin-up, calibration, and validation periods. The spin-up period extended from 1 August 2015 to 1 January 2017 and was used to stabilize the initial model states. Model calibration was performed from 1 January 2017 to 30 October 2019, while the period from 1 July 2020 to 31 December 2023 was used for validation, because of gaps in the observed discharge record between the two periods.

Not all model parameters were calibrated; some were fixed to limit the ability of the model to compensate for differences in the meteorological forcing. In particular, parameters that directly interact with variables estimated on the grid or at GU centroids were kept constant. The two adjustment coefficients for rainfall and snowfall measurements were set to 1 (i.e., no correction), preventing these parameters from modifying the precipitation estimates and artificially reducing differences among the evaluated configurations. Since potential evapotranspiration was computed using the interpolated temperature fields, temperature within the hydrological model primarily affects the snow component. The melting temperature parameter was therefore fixed at 0 °C to prevent the model from compensating for differences in temperature forcing through the snowmelt process.

Snow water equivalent (SWE) simulations were additionally evaluated against MODIS snow-cover observations. Snow presence or absence in the MODIS data was identified using the conventional NDSI threshold of 40 (Härer et al., 2018), whereas simulated snow cover was identified using an SWE threshold of 5 mm (Dawson et al., 2018; Di Marco et al., 2020). These thresholds were used to obtain binary classifications of snow presence and absence for the comparison between simulated SWE and MODIS-derived snow cover as explained in Section 3.6.1.

Simulated root-zone water storage was additionally evaluated against the corresponding

estimates from the MERIDA dataset, providing an independent assessment of simulated subsurface-water dynamics following the procedure described in Section 3.6.3.

6.3 Results and discussion

The results and discussion presented in this chapter are organized into two main parts. The first concerns the improvements to the GEOframe kriging procedure and includes: i) the analysis of the semivariograms and ii) the performance of the LOOCV validation to assess the interpolation accuracy. The second part addresses the core research question, namely the comparison of different approaches for generating the atmospheric forcing fields used in the rainfall-runoff model (NEAR, CEN and GRIDW), while also considering the GU delineation (SG, GU20 and GU10).

6.3.1 Semivariogram and leave-one-out cross-validation to test kriging results

The first step in the analysis of the GEOframe-kriging automated procedure focused on the semivariogram outputs over the study period. For both precipitation and temperature, the semivariogram interpolation yielded satisfactory results at each time step; in fact, no backup values were needed to complete the series. The behavior of the two variables, however, differs because of their distinct spatial patterns. Precipitation, in particular, is characterized by a large proportion of zero records. For temperature, more than 70 000 records were analyzed. In most cases (96%), the algorithm identified a clear trend with elevation, while in the remaining 4% no evident trend was found. The Gaussian (21%) and spherical (22%) models were the least frequent, whereas the linear (29%) and exponential (28%) models occurred with comparable frequency.

Figure D16 (Appendix) reports a summary of the semivariograms obtained for temperature for both OK and DK. In each plot, the shaded area represents the interquartile range (25%–75%), while the solid line is the medoid defined as the theoretical model among all that is closest to the median values of nugget, sill, and range across the ensemble, using

least-squares distance. Unlike a simple mean, the medoid is an actual fitted model, chosen because it minimizes the overall deviation from the central tendency, thus providing a representative curve for the distribution. For all models, detrending reduces both nugget and sill and shortens the ranges, indicating that elevation plays a key role in explaining spatial variability (Daly et al., 2002; Hofstra et al., 2008). Analyzing the behavior by season (Figure D18 and Figure D19 in the Appendix), winter exhibits higher sill values, indicating greater heterogeneity, whereas spring and summer, in the detrended case, show lower sill values and shorter ranges (Vicente-Serrano et al., 2003). The high variability observed in winter is mainly due to the lapse rate, which is influenced by temperature inversions, snow cover, slope exposure, and cold-air pooling (Rolland, 2003).

For temperature, the nugget-to-sill ratio (NSR) indicated strong spatial dependence for the exponential (0.11 for OK and 0.08 for DK) and Gaussian models (0.09 for OK and 0.17 for DK), while the linear (0.43 for OK and 0.34 for DK) and spherical models (0.27 for OK and 0.28 for DK) showed moderate dependence.

For precipitation, kriging was performed only on records where precipitation was observed at more than one station, under the conditions that i) not all values were identical (to avoid trivial cases) and ii) at least two different values were recorded. A total of about 14 000 records were analyzed. Among these, a trend was identified in 52% of cases, while no trend was detected in 48%. The Gaussian model was the most frequent (34%), followed by the linear (28%), exponential (19%), and spherical (19%) models. The Appendix reports semivariogram results for precipitation in the same way as for temperature: Figure D17 provides a general overview, showing both OK and DK cases, while Figure D20 and Figure D21 present the seasonal results for OK and DK, respectively, with the medoid model and the interquartile range (25%–75%) for the nugget, sill and range. Across all models, nugget values remain close to zero, indicating limited small-scale noise. Sill values are generally low in the non-detrended case, whereas detrending enhances spatial heterogeneity, especially for the exponential model during autumn and summer. Gaussian,

linear, and spherical semivariograms exhibit stable ranges (mostly below 30 km), in line with the localized character of precipitation events. Both nugget and sill are typically higher in summer than in winter, highlighting the influence of convective rainfall (Berndt et al., 2014). Conversely, the exponential model shows larger ranges after detrending, suggesting residual large-scale spatial components not fully accounted for by the elevation trend.

For precipitation, NSR showed a strong spatial dependence for all the semivariograms: exponential (0.000004 for OK and 0.000003 for DK), Gaussian (0.051 for OK and 0.025 for DK), linear (0.14 for OK and 0.02 for DK), and spherical (0.0 for OK and 0.0 for DK).

Using the semivariograms evaluated in the previous section, the LOOCV was applied to the network of weather stations to assess the interpolation performance for temperature and precipitation.

LOOCV was applied to temperature data using about 3.8 million observations from 54 weather stations. The global performance was as follows: r = 0.99, R^2 = 0.98, RMSE = 1.37 °C, MAE = 0.98 °C, Bias = -0.0035, Spearman = 0.99, NSE = 0.98, KGE = 0.99, and PBIAS = -0.05. Overall, the model performance is high, with only a few exceptions. The KGE consistently remains very high, the RMSE is satisfactory, and the MAE is low, indicating good predictive accuracy. The bias is negligible, which suggests that the model predicts temperature correctly without systematic distortion. The obtained values are satisfactory, especially when compared with other publications that cover a wider area on a daily time-step and over a climatic series, which also includes this study region, for example, Crespi et al. (2021) reported an MAE of 1.5 °C, an RMSE of about 1.9 °C, and a correlation coefficient (r) of 0.97. In comparison, the model by Brunetti et al. (2014) performs better than the results of this study; for regression kriging they obtained an RMSE in the range 0.76 °C to 1.11 °C and an MAE between 0.59 °C and 0.86 °C, although these results refer to monthly temperature over the whole of northern Italy. The performance is comparable to, or even better than, that reported by Bancheri et al. (2018) where the previous version of the kriging module—based on a single theoretical model—achieved an NSE of 0.93 for

temperature in LOOCV. In our case, an NSE of 0.98 was obtained, indicating a substantial improvement. Similar results were found by Berndt and Haberlandt (2018).

Figure 6.2 shows the distribution of performance metrics obtained from the leave-one-out cross-validation (LOOCV) of the kriging interpolation, presented as box-plots by station. The figure is organized into two main rows: the upper row refers to air temperature and the lower row to precipitation. Each row includes four panels: (a, e) the coefficient of determination (R^2) and KGE; (b, f) the RMSE; (c, g) the MAE; and (d, h) the bias. When examined at the station level, the results are generally satisfactory. In terms of KGE, only five stations fall below the threshold of 0.75 (values above 0.75 are considered good, while those between 0.50 and 0.75 are acceptable), and only two of these stations have values lower than 0.50. It is worth noting that the station with ID 364 shows a negative KGE (-7) and a very high PBIAS (≈ 800) and is therefore not shown in the plot. This behavior results from numerical issues: both the observed and estimated mean air temperature values are close to zero, which makes the KGE and PBIAS unstable. Under such conditions, even small variations can lead to large fluctuations in these metrics (Santos et al., 2018).

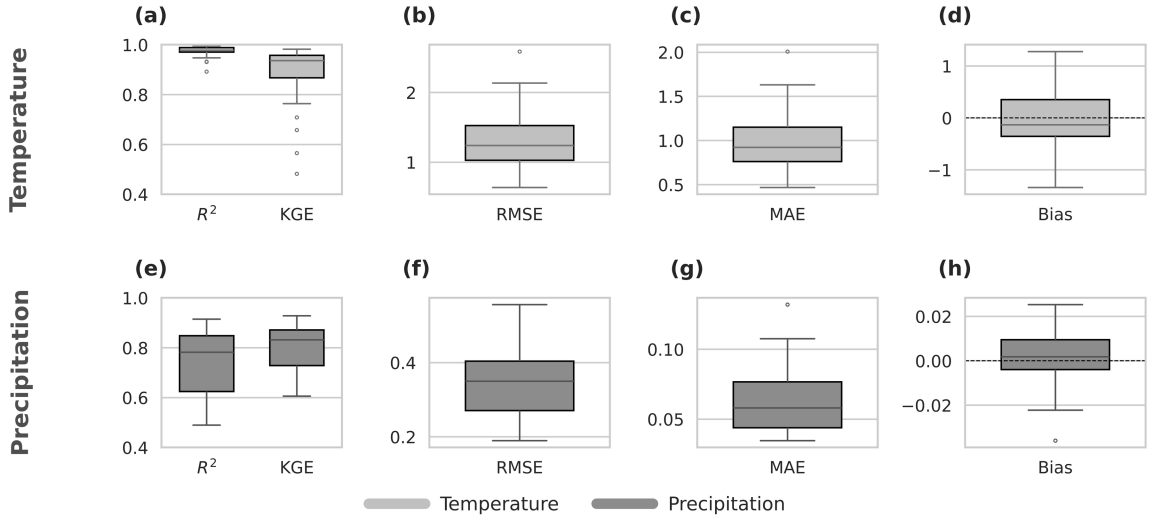

Figure 6.2 – Box-plots of performance metrics obtained through leave-one-out cross-validation (LOOCV) for all meteorological stations. The upper row (a–d, light gray) refers to air temperature, and the lower row (e–h, dark gray) to precipitation. Panels show (a, e) coefficient of determination (R^2) and Kling–Gupta Efficiency (KGE), (b, f) root mean square error (RMSE), (c, g) mean absolute error (MAE), and (d, h) bias.

Figure D23 (Appendix) presents the same set of performance metrics as Figure 6.2, but stratified by elevation. In this case, each panel contains four box-plots, corresponding to the four elevation classes considered. The elevation ranges were defined so as to include a comparable number of stations in each class. This representation allows the evaluation of model performance across different altitude bands, highlighting that the two elevation bands 200 m to 650 m asl and 650 m to 974 m asl perform better due to the fact that most of the highest-elevation stations are located near the boundary of the domain.

Figure D22 (Appendix) presents the same set of performance metrics as Figure D23 but stratified by season. In this case, each panel contains four box-plots corresponding to the four climatological seasons: winter (DJF: December–January–February), spring (MAM: March–April–May), summer (JJA: June–July–August), and autumn (SON: September–October–November). This representation highlights the seasonal variability in performance, with winter generally showing the lowest accuracy across most metrics. This reduced performance is mainly related to the semivariogram structure: as noted above, the winter semivariogram exhibits higher variance, which negatively affects the interpolation results. The high temperature variability in winter is due to winter inversions, snow presence, slope exposure, and also cold-air pooling (Gubler et al., 2011; Rolland, 2003).

For precipitation, based on about 3.3 million observations from 45 weather stations, the model yielded the following performance metrics: r = 0.87, R^2 = 0.76, RMSE = 0.35 mm, MAE = 0.06 mm, Bias = 0.002 mm, Spearman = 0.66, NSE = 0.75, KGE = 0.85, and PBIAS = 1.51. The performance comparison can be made with Berndt and Haberlandt (2018), who tested several interpolation methods for hourly data; OK yielded a bias of 0.007 mm. The model presented in this study outperformed these values, although it was based on both OK and DK. The RMSE is also comparable with the event in Foehn et al. (2018). In comparison with Bancheri et al. (2018), where only one event was computed for two stations with NSE values ranging from 0.91 to 0.35, the present case also shows values ranging from 0.91 to 0.39 across all stations.

Figure 6.2, in the bottom row, shows four panels related to the precipitation: the first panel shows the coefficient of determination (R^2) and the KGE, with median value of 0.8 and a minimum value of 0.6 (above the 0.5 threshold usually considered acceptable), indicating generally good agreement between observed and interpolated values. The second panel reports the distribution of RMSE, and the third shows the MAE, both highlighting relatively low errors and limited systematic deviations. The last panel presents the bias, with most stations close to zero, although a few outliers reveal local over- or underestimation.

Figure D25 (Appendix) provides a similar representation as in Figure 6.2, but highlights the differences by elevation, where the stations are grouped into four bins with the same number of stations in each bin. The performance decreases with increasing altitude, leading to higher errors at higher elevations and a wider spread of the error distributions, as indicated by the larger boxes and whiskers in the plots. The decrease in performance with increasing altitude can be partly explained by boundary effects, since most high-elevation stations are located at the edge of the study area, where kriging is known to perform less reliably. In addition, the sparse and uneven distribution of stations at high altitudes, combined with stronger local variability of precipitation, further contributes to the higher errors observed. Figure D24 (Appendix) is similar to the previous one but divided by season: winter (DJF: December–January–February), spring (MAM: March–April–May), summer (JJA: June–July–August), and autumn (SON: September–October–November). The worst performances are observed in winter, mainly due to the higher spatial heterogeneity of precipitation and the frequent occurrence of mixed rain–snow events, which are difficult to capture by the interpolation. By contrast, the highest RMSE values occur in summer, largely due to convective storms, which generate localized and highly variable precipitation that is poorly represented by the available station network (Cavalleri et al., 2024; Suklitsch et al., 2011). In autumn, the interpolation performs comparatively better, as precipitation is mainly driven by large-scale events, which are more spatially coherent and therefore easier to capture by the kriging procedure. In both cases (temperature and precipitation), stations at the edge of

the study area are more strongly affected by boundary effects, which are known to reduce the reliability of geostatistical interpolation (Berezowski et al., 2016; Verdin et al., 2016), particularly in the north of the domain.

Alongside boundary effects, some methodological limitations should be considered when interpreting the rainfall interpolation results. Although the LOOCV statistics indicated an overall satisfactory predictive performance, these results should not be interpreted as evidence that kriging can fully reproduce the spatial structure of rainfall fields. Kriging is a well-established and practical approach for deriving continuous rainfall surfaces from point rain-gauge observations, but it is also known to produce smooth and continuous estimates. This aspect is particularly relevant for precipitation, whose spatio-temporal patterns are highly variable and difficult to reconstruct from point observations, especially in data-scarce catchments or during highly dynamic small-scale rainstorms (Grundmann et al., 2019), intermittency and zero-rainfall areas embedded within rainy regions (P. Kumar & Foufoula-Georgiou, 1994), as well as localized high-intensity structures associated with convective rain cells (Rebora & Ferraris, 2006). Under these conditions, the smoothing effect of kriging may attenuate local maxima and sharp spatial gradients, leading to biased or unrealistic local estimates (Yamamoto, 2005). Moreover, Ly et al. (2013) highlighted that the performance of geostatistical rainfall interpolation depends on the choice of the variogram model, the type of precipitation event, e.g., convective versus stratiform rainfall, and the density and configuration of the rain-gauge network. Therefore, the kriged rainfall fields used in this study should be interpreted as practical catchment-scale estimates, suitable for supporting the analysis of the hydrological response, rather than as a detailed reconstruction of fine-scale rainfall processes. This practical use of kriging-based rainfall interpolation in hydrological applications is consistent with previous studies (Bancheri et al., 2018; Haberlandt, 2007; Verworn & Haberlandt, 2011).

6.3.2 Evaluation of the different approaches on multiple hydrological processes

This section evaluates how the different spatial representations affect both the meteorological forcing supplied to the model and the resulting hydrological simulations. First, the three forcing representation approaches (NEAR, CEN, and GRIDW) are compared to quantify differences in the meteorological inputs associated with the GUs. Second, the combined effects of forcing representation and GU delineation are assessed by comparing the hydrological outputs of the nine model configurations.

6.3.2.1 Effects on the rainfall-runoff model inputs

The comparison between the NEAR, CEN, and GRIDW evaluation methods highlights systematic differences in the meteorological input variables, which then affect the rainfall-runoff model. Figure 6.3 reports the monthly mean values of air temperature, cumulative potential evapotranspiration, and cumulative precipitation across GUs, for both the SG and GU* configurations. For each variable, monthly statistics were computed by averaging the basin-aggregated values across all GUs.

In the case of potential evapotranspiration and precipitation, monthly cumulative means were used, while air temperature was expressed as the monthly mean of basin-averaged values. Vertical black bars represent the variability of GU-averaged values for each corresponding month, expressed as the range between the 97.5th and 2.5th percentiles. Generally, the NEAR and CEN configurations produced comparable estimated values of air temperature, whereas the GRIDW configurations tended to produce lower temperatures in all cases. This difference does not necessarily translate into systematic variations in potential evapotranspiration. For example, in the GU10 configuration, lower temperatures in GRIDW are partly compensated by the representation of incoming radiation, which plays a dominant role in PET estimation. In contrast, in the SG configuration, the expected relationship between lower temperature and lower PET is clearly visible. The higher temperatures in NEAR and CEN arise from the fact that the values from the nearest station or the estimated

centroid do not always represent the average elevation of the GU, whereas the GRIDW approach accounts for the full elevation range within the unit. Radiation, on the other hand, is strongly influenced by local orography, and therefore PET differences vary on a case-by-case basis. Within the GU10 delineation, for instance, some GUs show large internal variability in radiation exposure: in GU 3928, the sky-view factor—defined as the fraction of the sky dome visible from a point, ranging from 1 (completely open sky) to 0 (no sky visible)—equals 0.32 at the centroid but 0.82 when averaged over the entire GU. This example illustrates how local shading effects can substantially alter the amount of incoming radiation.

As shown by the black distribution lines, for GU20 and GU10 delineation, the mean values of PET are nearly identical among configurations, although the variability—represented by the vertical black bars (97.5th–2.5th percentile range)—is generally larger in NEAR and CEN. This indicates a higher spatial dispersion among GUs when the NEAR and CEN methods are used to evaluate PET.

For precipitation, the differences are smaller overall, though the GU configuration shows a slightly different distribution pattern between the three forcing methods. Despite these relatively small differences, NEAR consistently receives less precipitation than CEN and GRIDW across all delineations and throughout the year.

To better understand the impact of precipitation estimates, the runoff coefficient (RC; Section 3.6) was calculated using observed rather than simulated discharge, according to Equation 3.2. The same observed discharge time series was used for all configurations, providing a common reference for comparing the precipitation estimates obtained with the different forcing representation approaches. Consequently, differences in RC among configurations reflect differences in the estimated precipitation rather than differences in the simulated hydrological response.

The comparison across the different configurations highlights systematic differences as summarized in Table 6.1. The NEAR configurations, due to their lower precipitation

estimates than the other two approaches, produce the highest runoff coefficients. Among the CEN configurations, the SG-CEN configuration shows the highest value, which can be explained by the fact that only two centroids contribute to the calculation, and these are strongly influenced by a limited number of nearby stations with large interpolation weights. Overall, in both catchments the CEN configurations tend to overestimate precipitation compared with the GRIDW configurations. As expected, GRIDW cases produce very similar runoff coefficient values, since all configurations represent a form of double mean aggregation.

Table 6.1 – Runoff coefficient (RC) for the different meteorological forcing methods and spatial discretizations.

Discretization	**NEAR**	**CEN**	**GRIDW**
SG	0.68	0.59	0.53
GU10	0.63	0.55	0.53
GU20	0.65	0.55	0.54

6.3.2.2 Effects on the rainfall-runoff model outputs

Using the aforementioned procedure, the rainfall-runoff model was calibrated as described in the Methods section. All configurations were calibrated using an equal number of steps (2×800) to ensure comparable conditions, and validated over the same time periods. As expected, the results differ substantially in terms of performance: the simpler configuration (SG) performs worse when evaluated with the same number of steps. However, since this configuration involves only two GUs, the required computational time is considerably lower. This implies that, if the comparison were based on equal wall-clock time rather than equal steps, the SG configuration could achieve more iterations and potentially improve its performance. Nevertheless, for consistency we preferred to use the same number of steps across all cases.

Figure 6.4 illustrates the simulated (red) and observed (blue) discharge at the Malé outlet for the nine configurations. The shaded gray area corresponds to the calibration period,

while the unshaded section represents the validation period.

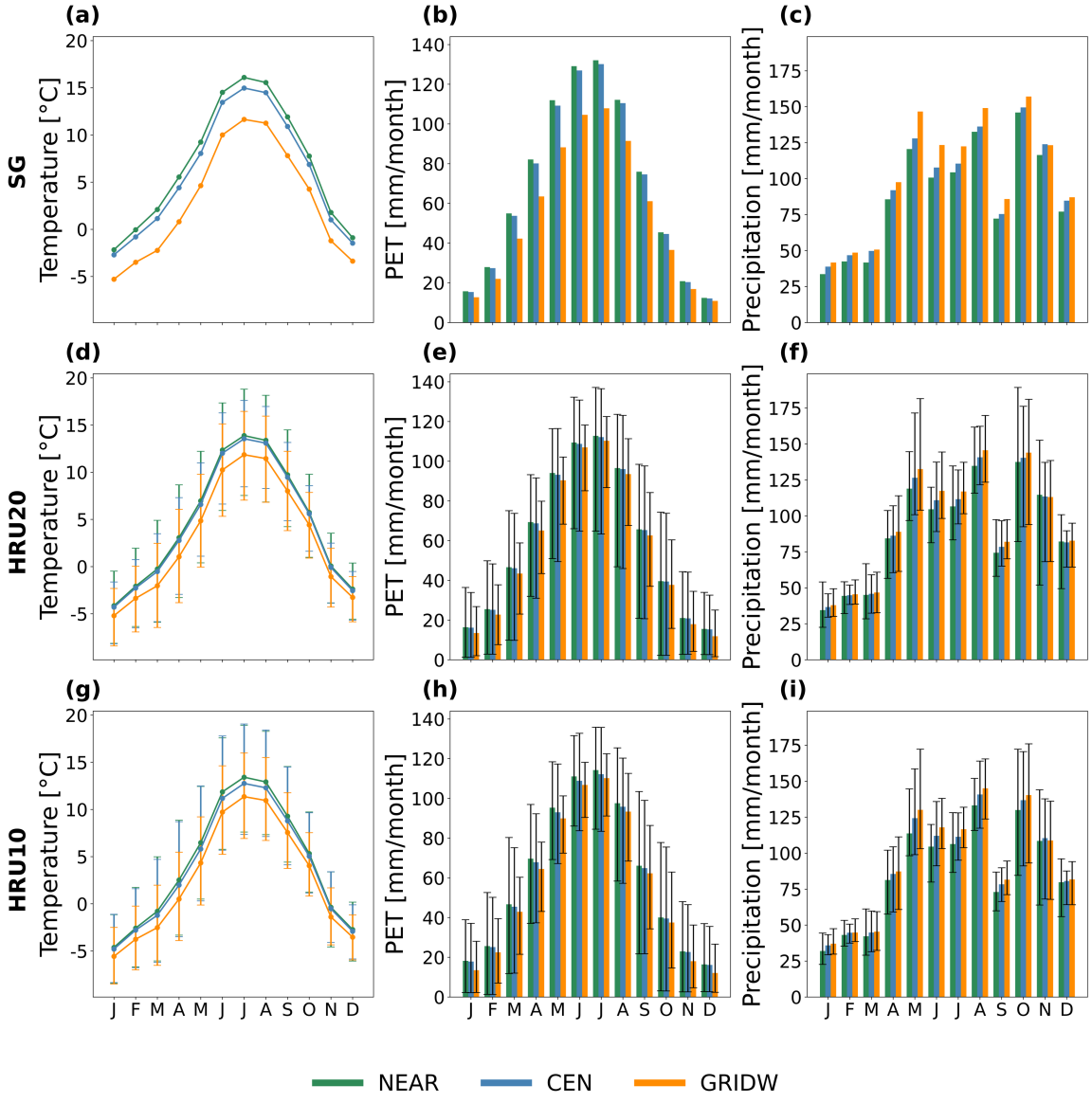

Figure 6.3 – Monthly statistics across all GUs for the nine configurations. The upper row (a–c) refers to the SG-based configurations, the middle row (d–f) to the GU20 and the lower row (g–i) to the GU10-based configurations. Panels (a, d, g) show mean air temperature, calculated as the monthly mean of basin-averaged values. Panels (b, e, h) show cumulative potential evapotranspiration, and panels (c, f, i) show cumulative precipitation, both obtained as monthly cumulative means across all basins. Vertical bars indicate variability between basins for the GU20 and GU10 cases (97.5–2.5 percentile range). Green denotes NEAR-based forcing, blue denotes CEN-based forcings, and orange denotes GRIDW-based forcings.

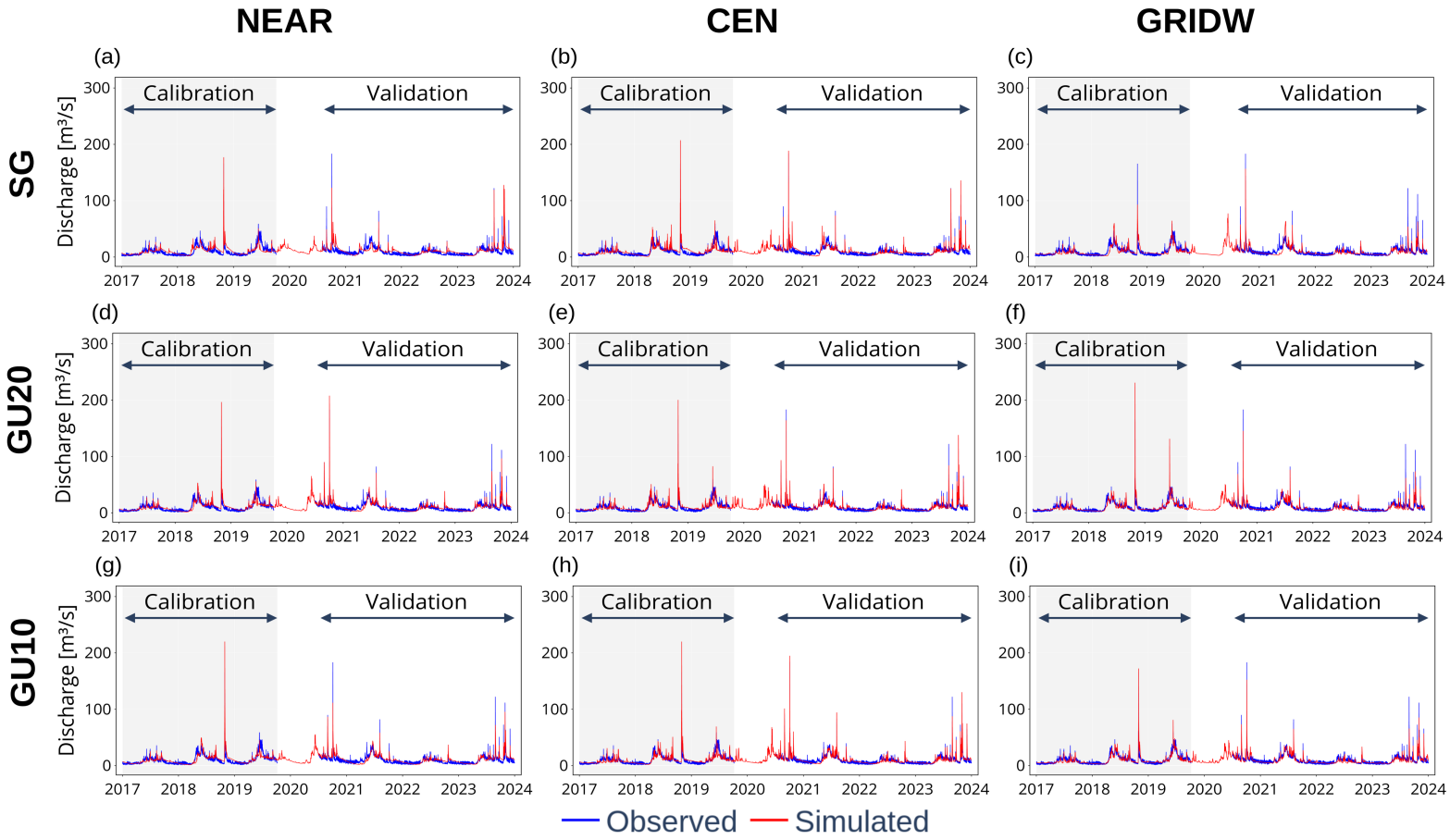

Figure 6.4 – Simulated and observed hydrographs at the Malé stream gauge for the nine model configurations. Panels (a)–(i) correspond to: (a) SG-NEAR, (b) SG-CEN, (c) SG-GRIDW, (d) GU20-NEAR, (e) GU20-CEN, (f) GU20-GRIDW, (g) GU10-NEAR, (h) GU10-CEN, and (i) GU10-GRIDW. Rows correspond to the GU delineation method (SG, GU20, and GU10), whereas columns represent the method used to estimate the meteorological forcings (NEAR, CEN, and GRIDW). Discharge is expressed in cubic meters per second (m^3/s). Blue lines represent the observed discharge and red lines the simulated discharge. The arrows indicate the calibration (light-gray shaded area) and validation periods.

Table 6.2 reports the values of the main efficiency metrics (KGE, NSE, and NSE_{log}), separated into calibration and validation periods, while Table D3 in the Appendix presents the error-based metrics (RMSE, PBIAS).

Table 6.2 – Performance metrics for discharge value at the Malé stream gauge for the nine model configurations. Rows correspond to the combinations of GU delineation (SG, GU20 and GU10) and meteorological forcing evaluation (NEAR, CEN, GRIDW). Columns report KGE, NSE, and NSE_{log} for both calibration and validation periods.

	Calibration (2017-01-01–2019-10-30)			Validation (2020-07-01–2023-12-31)		
Configuration	KGE	NSE	NSE_{log}	KGE	NSE	NSE_{log}
SG-NEAR	0.69	0.35	0.09	0.67	0.41	0.24
SG-CEN	0.73	0.47	0.31	0.78	0.66	0.25
SG-GRIDW	0.81	0.64	0.42	0.79	0.63	0.46
GU20-NEAR	0.78	0.56	0.47	0.80	0.71	0.46
GU20-CEN	0.81	0.64	0.20	0.82	0.73	0.29
GU20-GRIDW	0.87	0.76	0.59	0.85	0.74	0.59
GU10-NEAR	0.77	0.56	0.52	0.79	0.69	0.50
GU10-CEN	0.79	0.59	0.39	0.80	0.77	0.51
GU10-GRIDW	0.90	0.80	0.62	0.86	0.76	0.60

At the Vermiglio stream gauge, validation was carried out, and the results are reported in Table 6.3 for the efficiency metrics and in Table D4 in the Appendix for error metrics.

Overall, the model is able to reproduce the observed discharge with satisfactory accuracy, with most configurations falling within the acceptable range for KGE and NSE (Kling et al., 2012; Moriasi et al., 2007; Towner et al., 2019), though the skill in reproducing low flows (NSE_{log}) varies across configurations.

Comparing the calibration (2017–2019) and validation (2020–2023) periods reveals distinct behaviors among the different model configurations. The GU20-GRIDW and GU10-GRIDW configurations maintain consistently high performance from calibration to validation, with only minor decreases in KGE (2–4%), NSE (3–5%), and virtually no change in NSE_{log} (0–3%). This stable performance indicates good robustness and little evidence of overfitting.

Table 6.3 – Performance metrics for discharge at the Vermiglio stream gauge for the nine model configurations. Rows correspond to the combinations of GU delineation (SG, GU20 and GU10) and meteorological forcing evaluation (NEAR, CEN, GRIDW). Columns report KGE, NSE, and NSE_{log} computed over the independent validation period (2017-01-01 / 2023-12-31).

Configuration	KGE	NSE	NSE_{log}
SG-NEAR	0.54	0.50	0.33
SG-CEN	0.70	0.62	0.34
SG-GRIDW	0.66	0.57	0.46
GU20-NEAR	0.70	0.55	0.34
GU20-CEN	0.60	0.52	0.29
GU20-GRIDW	0.65	0.61	0.51
GU10-NEAR	0.73	0.54	0.34
GU10-CEN	0.74	0.66	0.49
GU10-GRIDW	0.76	0.75	0.69

For the CEN forcing, both the GU20-CEN and GU10-CEN configurations perform better during validation than during calibration. KGE increases slightly (about 1%), while NSE improves by 14% and 31%, respectively. Similarly, NSE_{log} increases by 45% for GU20-CEN and by 31% for GU10-CEN, indicating an improved representation of low flows during the validation period.

Despite relying on the simplest precipitation assignment, the GU20-NEAR and GU10-NEAR configurations also show improved performance during validation. KGE increases by approximately 3% in both cases, while NSE improves by 27% and 23%, respectively. In contrast, NSE_{log} remains essentially unchanged, with decreases of only 2% and 4%. Among the SG configurations, SG-GRIDW is the most stable, exhibiting only marginal differences between calibration and validation. SG-CEN shows a marked improvement in validation NSE (+40%) relative to calibration, whereas NSE_{log} decreases by 19%, indicating that low-flow simulation remains more challenging. Finally, SG-NEAR represents the weakest-performing configuration. KGE decreases slightly (-3%) from calibration to validation, whereas NSE increases by 17%. Although NSE_{log} also increases, from 0.09 to 0.24, it remains substantially lower than that of the interpolation-based precipitation forcings.

The effect of spatial discretization (SG versus the GU20 and GU10 delineations) is clearly evident. Both the GU20 and GU10 delineations generally outperform their SG counterparts across all performance metrics, with the most pronounced improvements observed for NSE. Depending on the precipitation forcing, the relative increase in NSE ranges from 25% to 60% during calibration and from 17% to 68% during validation, highlighting the substantial benefit of a finer spatial discretization for rainfall-runoff simulations.

Within each delineation, the influence of the precipitation forcing (NEAR, CEN, and GRIDW) is also evident, although it is less pronounced than the effect of spatial discretization. For the SG delineation, model performance consistently improves from NEAR to CEN and then to GRIDW, with the largest gains observed during calibration. The same ordering is maintained during validation, although the performance gap between the two interpolation-based forcings (CEN and GRIDW) becomes smaller. For both the GU20 and GU10 delineations, GRIDW consistently provides the highest overall performance. The relative performance of NEAR and CEN, however, depends on the evaluation metric: CEN generally yields slightly higher KGE and NSE values, whereas NEAR achieves higher NSE_{log} values, indicating a better representation of low-flow conditions. This behavior is consistent for both the GU20 and GU10 delineations, suggesting that once a finer spatial discretization is adopted, the choice of precipitation forcing mainly affects the simulation of low flows rather than the overall hydrograph.

For the Vermiglio stream gauge, evaluated over the period 2017–2023, the results show generally satisfactory performance. Among all configurations, GU10-GRIDW provides the best overall performance, with KGE, NSE, and NSE_{log} values of 0.76, 0.75, and 0.69, respectively, and is the only configuration achieving satisfactory performance for all three metrics. GU10-CEN also performs well, with KGE, NSE, and NSE_{log} values of 0.74, 0.66, and 0.49, respectively. GU10-NEAR provides comparable KGE (0.73) but lower NSE (0.54) and NSE_{log} (0.34), indicating reduced skill in reproducing low flows.

The GU20 delineation exhibits a different behavior. GU20-NEAR achieves the highest

KGE (0.70) and NSE (0.55), whereas GU20-GRIDW provides the highest NSE_{log} (0.51), indicating an improved representation of low flows. GU20-CEN shows the lowest performance among the GU20 forcing configurations.

Among the stream-gauge (SG) configurations, SG-CEN achieves the highest KGE (0.70) and NSE (0.62), whereas SG-GRIDW provides the highest NSE_{log} (0.46), again indicating a better representation of low-flow conditions. Finally, SG-NEAR, representing the simplest forcing assignment, exhibits the lowest overall performance, with KGE, NSE, and NSE_{log} values of 0.54, 0.50, and 0.33, respectively.

The deterioration observed when moving from Malé to Vermiglio can be attributed to the change in basin characteristics. The upstream portion of the basin is located at higher elevations and is more strongly influenced by glaciers and peak flows, which increases hydrological variability and makes model reproduction more challenging.

Comparing Malé with Vermiglio, the CEN configurations generally deteriorate less than the GRIDW and NEAR configurations, although this behavior depends on the spatial discretization. In particular, SG-CEN shows only minor decreases in KGE and NSE and even improves in NSE_{log}, whereas GU10-CEN also maintains relatively high performance across all metrics. In contrast, GU20-CEN exhibits a more pronounced deterioration. By contrast, SG-GRIDW undergoes larger losses in KGE and NSE, whereas GU10-GRIDW remains the best-performing configuration at the Vermiglio stream gauge. The NEAR configurations also show greater sensitivity to the transition from Malé to Vermiglio, particularly GU10-NEAR, which exhibits noticeable reductions in NSE and NSE_{log}.

Notably, the GU20 configurations exhibit a much larger deterioration in performance than the GU10 configurations when transferred to the Vermiglio stream gauge. A similar, although less pronounced, behavior is also observed for the SG configurations. These results suggest that the parameter set calibrated for the GU20 delineation transfers less successfully to the Vermiglio stream gauge, whereas the GU10 and SG parameterizations exhibit better generalization across catchments. To better understand the behavior of the

different configurations, a cumulative discharge-volume analysis was performed (Figure D26 in Appendix). In the figure, the observed cumulative discharge is shown in blue, whereas the simulated cumulative discharge is shown in red. The NEAR configurations exhibit a progressive divergence between the simulated and observed cumulative volumes, with the simulated discharge systematically exceeding the observations, particularly during the validation period. This discrepancy is reduced by the CEN approach, especially for the GU10 configuration, while the GRIDW approach minimizes it across all configurations. These findings are consistent with the total discharge-volume errors reported in Table D5. Figure D27 shows the distribution of the monthly discharge-volume errors for the different model configurations. Overall, the results indicate a tendency to underestimate discharge volumes during the summer months and to overestimate them during winter. The GRIDW configurations exhibit a distinct seasonal behavior, with a slight overestimation during the snowmelt period. Among all months, July generally shows the smallest volume errors. It should be noted that the model was developed to enforce water-balance closure at every simulation time step. Consequently, the residual water-balance error corresponds only to numerical rounding errors. This can be observed in Figure D29, which presents the annual water-balance components for each of the nine model configurations using hydrological years (October–September). In each panel, the dark-blue bar represents precipitation, while the adjacent stacked bar represents the partitioning of precipitation into actual evapotranspiration (green), discharge (light blue), and the variation in total water storage (ΔS, hatched orange). The signed value of ΔS is reported above each stacked bar, where negative values indicate a depletion of basin water storage. As expected from mass conservation, the total height of the stacked bar closely matches the precipitation bar for every hydrological year and configuration. The annual water-balance plots clearly highlight the severe drought affecting hydrological year 2022, which originated during 2021. Reduced snowfall and prolonged precipitation deficits resulted in lower water availability, thereby limiting both actual evapotranspiration and streamflow. Most configurations also indicate a net depletion

of basin water storage during this period, although the magnitude and temporal distribution of ΔS differ among forcing strategies. Most CEN and GRIDW configurations exhibit a pronounced negative storage variation during hydrological year 2021, followed by a smaller depletion in 2022. In contrast, the NEAR configurations display a different behavior. The GU10-NEAR configuration shows relatively small storage variations in both years, whereas SG-NEAR exhibits a large storage depletion in 2021 followed by a slightly positive storage change in 2022.

Figure D30 (Appendix) presents the monthly mean actual evapotranspiration (AET) aggregated over all basins for SG, GU20 and GU10 delineations. Green, blue and orange bars represent respectively the mean values obtained with NEAR, CEN and GRIDW-based forcings, while the error bars indicate the 2.5th–97.5th percentile range across basins. The NEAR configuration systematically produces lower estimates of AET compared to the CEN and GRIDW configurations. This difference is particularly marked for the SG delineation in panel (a), especially during the summer months, where the comparison with potential evapotranspiration reveals the opposite pattern. The effect of the scale configuration is evident when comparing the SG and GU20 or GU10 delineations. In the SG delineation, the NEAR and CEN configurations exhibit higher PET but lower AET than GRIDW, whereas this contrast becomes much less pronounced in the GU20 or GU10 delineation. The underlying limiting factor is the reduced availability of water in the root zone in the NEAR and CEN configurations and the concurrent overestimation of discharge (Table D3 and Table D4 in Appendix), both of which constrain actual evapotranspiration. In particular, the limited root-zone water storage in the NEAR configuration results from the substantially lower precipitation estimates compared with the other configurations. In the GU20 and GU10 delineations, the AET estimates are much more similar among the three forcing representations, although the NEAR configuration still yields the lowest values.

In comparison with previous work, Tovazzi (2013) reported cumulative AET values ranging from 250 to 380 mm/year for the period 2001–2006. In contrast, for the period

2017–2023, our simulations yield mean values: approximately 218 mm/year for SG-NEAR, 354 mm/year for SG-CEN, 481 mm/year for SG-GRIDW, 299 mm/year for GU20-NEAR, 359 mm/year for GU20-CEN, 470 mm/year for GU20-GRIDW, 322 mm/year for GU10-NEAR, 423 mm/year for GU10-CEN, and 483 mm/year for GU10-GRIDW. Across all delineations, the NEAR forcing consistently produces the lowest AET values, CEN yields intermediate values, and GRIDW results in the highest AET. The influence of spatial discretization is less pronounced but generally leads to higher AET estimates when moving from the SG to the GU20 and GU10 delineations, particularly for the NEAR and CEN forcings. Consequently, only the NEAR configurations and GU20-CEN remain within or below the range reported by Tovazzi (2013), whereas all GRIDW configurations and GU10-CEN exceed its upper bound. It should be noted that the values reported by Tovazzi (2013) refer to an earlier period, whereas more recent studies have emphasized the presence of a positive long-term trend in evapotranspiration. Mallucci et al. (2019) quantified an increase of about 2 mm/year in AET, while Morlot et al. (2023) also reported an upward trend in evapotranspiration across the Adige basin, of which the Noce catchment is a tributary. These findings suggest that the higher AET values obtained for several configurations should not necessarily be interpreted as an overestimation relative to the earlier estimates of Tovazzi (2013). Rather, they are consistent with the documented long-term increase in evapotranspiration and with the improved representation of water availability provided by the semi-distributed model configurations.

From the ECP analysis, the agreement between measured and modeled discharge, expressed in terms of the Pearson correlation coefficient and the corresponding RMSE, is summarized in Table 6.4. Overall, the GRIDW configurations achieved the highest correlation coefficients and the lowest RMSE values, followed by the CEN configurations, whereas the NEAR configurations showed the lowest agreement. Among the three spatial discretizations, the GU20 and GU10 configurations generally outperformed the SG configuration. In particular, GU10-GRIDW provided the best overall agreement with the observations in

terms of both Pearson correlation and RMSE, although its performance was only marginally better (RMSE) than that of GU20-GRIDW.

Table 6.4 – Agreement between measured and modeled discharge obtained from the ECP analysis, expressed as the Pearson correlation coefficient (r) and the corresponding RMSE.

Configuration	r	**RMSE** (m^3/s)
SG-NEAR	0.70	5.66
SG-CEN	0.79	4.85
SG-GRIDW	0.81	4.70
GU20-NEAR	0.83	4.48
GU20-CEN	0.85	4.16
GU20-GRIDW	0.88	3.88
GU10-NEAR	0.82	4.53
GU10-CEN	0.85	4.10
GU10-GRIDW	0.88	3.75

These results suggest that the most relevant improvement in model reliability is achieved by increasing the spatial discretization from SG to GU20 and GU10. The effect of the precipitation forcing is comparatively smaller: while the interpolated forcings (CEN and GRIDW) consistently outperform the NEAR approach, the difference between CEN and GRIDW is relatively modest, particularly for the GU20 and GU10 delineations. Overall, these findings indicate that spatial discretization is the primary driver of the observed improvement in model reliability, whereas the choice of precipitation forcing plays a secondary, although still measurable, role.

In summary, the GU20 and GU10 delineations consistently outperform the SG configuration during both the calibration and validation periods at the Malé stream gauge, while SG-NEAR consistently exhibits the lowest performance among the tested configurations. The differences between GU20 and GU10 are generally small, indicating that both delineations provide a comparable representation of the hydrological response under the calibration conditions.

The main differences between the two GU-based delineations emerge when the parameter sets calibrated at the Malé stream gauge are used in validation of the upstream Vermiglio

stream gauge. In this case, the GU10 delineation generally maintains higher performance than GU20 across the tested precipitation forcings, indicating a more robust spatial transfer of the calibrated parameterization. No systematic deterioration associated with a specific forcing method can be identified, suggesting that the observed differences are primarily related to the spatial discretization rather than to the meteorological forcing evaluation approach.

Considering the comparable performance achieved at the calibration site, the better overall performance at the upstream validation stream gauge, and its finer spatial representation, the GU10 delineation represents the most suitable compromise among the tested configurations.

Overall, the results can be considered satisfactory for all configurations, especially considering that some parameters were kept fixed during calibration. In particular, the assumption of no measurement undercatch in precipitation and snow (with the adjustment coefficients set to 1) and the melting temperature fixed at 0 °C may have limited the model's ability to compensate for errors in the meteorological inputs or for intrinsic model uncertainties.

Beyond predictive performance assessed using discharge-based metrics, the practical value of the alternative spatial configurations also depends on their computational requirements and intended application. Table E2 summarizes the computational requirements of the calibration procedure. All calibrations were performed on the High-Performance Computing (HPC) cluster of the University of Trento under comparable conditions and using the same software environment. The SG configuration was allocated only two CPU cores because it consists of two GUs, whereas the GU20 and GU10 configurations were each allocated 22 cores to exploit their greater degree of available parallelism.

Despite the larger computational allocation, calibration wall-clock time increased substantially with spatial refinement, from approximately 6 h for SG to about 66 h and 82 h for GU20 and GU10, respectively. Compared with SG, these values correspond to increases by

factors of approximately 11 and 14. The difference was even greater in terms of aggregate CPU time, which increased from 18 CPU-hours for SG to 370 and 754 CPU-hours for GU20 and GU10, corresponding to approximately 21 and 42 times the computational effort required by SG. Consequently, although all configurations were calibrated using the same number of iterations, their computational requirements differed markedly.

Had the comparison instead been based on the same wall-clock-time budget, the SG configuration could have completed a substantially larger number of calibration iterations within the available time. Under an equal computational budget expressed in CPU-hours, the difference in the number of iterations that could be assigned to SG would have been even greater. Although additional iterations would not necessarily guarantee better final performance, they would provide the simpler configuration with a larger optimization budget and therefore a greater opportunity to identify an improved parameter set. Nevertheless, the same number of calibration iterations was retained in all cases to compare the configurations under a consistent calibration protocol.

Computational cost is not determined solely by the number of GUs. Because runoff routing follows the basin topology, downstream GUs cannot be evaluated until their upstream contributing units have been solved. These dependencies limit the degree of parallelism that can be effectively exploited, particularly when several GUs belong to the same sequential flow path. Similar considerations apply to individual simulation runs (Table E3), although the absolute computational costs are considerably smaller than for the complete calibration procedure. For a single multiyear simulation, wall-clock time increased from approximately 1 min for SG to 5 min for GU20 and 8 min for GU10, while the corresponding CPU time increased from 4 min to 52 min and 1 h 22 min, respectively.

The reported simulation and calibration times do not include the computational cost of preparing the meteorological inputs. In the NEAR approach, precipitation and temperature are assigned from the nearest meteorological station to each GU, while radiation and potential evapotranspiration are subsequently evaluated at the GU scale. By contrast, the

CEN and GRIDW approaches require spatial interpolation within the meteorological-input workflow. CEN performs the interpolation only at the GU centroids, corresponding to 2, 30, and 51 interpolation points for SG, GU20, and GU10, respectively. GRIDW instead performs the interpolation at approximately 5800 grid points before aggregating the resulting values to the GU scale. Differences in the computational cost of the complete workflow are therefore greater than those represented by the simulation and calibration times alone, particularly for GRIDW.

The increase in computational cost from SG to GU20 and GU10 was not accompanied by a proportional improvement in discharge simulation performance. If the objective is limited to reproducing discharge at the catchment outlet, SG therefore provides the most advantageous compromise between performance and model-execution cost. However, the cost–benefit assessment depends on the intended application. Because SG contains only two GUs, each simulated state variable and flux is represented by only two spatial values over the entire catchment. This configuration consequently provides limited spatial detail for applications requiring distributed information on variables such as snow water equivalent, soil moisture, evapotranspiration, or groundwater storage. The finer configurations provide greater spatial differentiation of these variables, although greater spatial detail does not necessarily imply greater accuracy.

The practical relevance of computational cost also depends on the operational workflow and simulation horizon. The individual simulation times reported in Table E3 refer to complete multiyear simulations, which are repeated many times during calibration. Calibration is generally performed offline and repeated only periodically, whereas an operational forecasting system subsequently performs much shorter simulations, typically covering the latest observations and a forecast horizon of a few days. Provided that the model states are carried forward and a long initialization period is not rerun at each update, the absolute differences in execution time among the configurations become much less consequential in daily operations, even if their relative computational costs remain different. The calibra-

tion cost is therefore important during model development and recalibration but may be amortized over many subsequent operational forecasts.

From an operational cost–performance perspective, the preferred configuration therefore depends on the available computational resources and intended model outputs. For discharge-oriented applications under strong computational constraints, SG-GRIDW provides a favorable compromise, offering satisfactory performance with relatively low model-execution costs, although its limited spatial detail and the preprocessing requirements of GRIDW must be considered. GU20 represents an intermediate compromise among computational cost, spatial detail, and model performance, although its spatial transferability to the independent upstream validation location requires further improvement. GU10 provides the best overall performance and the smallest degradation at the independent validation gauge, making it the most robust of the tested configurations, but it also has the highest computational cost.

6.3.2.3 Snow cover

As the basin lies in the Alps and has a high average elevation, the snow compartment represents an important aspect of the overall system. To better understand the reliability of the model, the presence/absence snow series derived from MODIS satellite observations were compared with those extracted from the SWE values simulated by the model. Table 6.5 reports the values of the performance metrics, namely false positives, false negatives, true positives, true negatives, the Overall Accuracy Test (OAT), and the Brier score, for the nine configurations. These metrics allow a comparison between the different configurations providing a quantitative assessment of the model's capability to reproduce snow presence/absence.

Comparing the performance achieved by the different configurations, the effect of spatial discretization emerges as the primary driver of model performance. Both GU20 and GU10 consistently outperform the SG configuration in terms of OAT and median Brier score, regardless of the forcing evaluation method. Differences among the forcing evaluation

Table 6.5 – Summary of performance metrics for the nine configurations. Each metric is computed by comparing the binary snow presence/absence series derived from MODIS observations with those simulated by the model for all sub-catchments. False positives (FP), false negatives (FN), true positives (TP), and true negatives (TN) were obtained by direct comparison between simulated and observed series. The Overall Accuracy Test (OAT) represents the proportion of correctly classified time steps, i.e., $(\mathrm{TP} + \mathrm{TN})/\mathrm{total}$, while the median Brier score summarizes the accuracy of binary predictions as the median of the Brier values across all sub-catchments.

Configuration	FP	FN	TP	TN	OAT	Brier (median)
SG-NEAR	0.038	0.137	0.326	0.500	0.826	0.169
SG-CEN	0.050	0.087	0.375	0.488	0.863	0.135
SG-GRIDW	0.132	0.015	0.447	0.405	0.852	0.147
GU20-NEAR	0.106	0.047	0.385	0.461	0.846	0.130
GU20-CEN	0.035	0.078	0.354	0.531	0.886	0.104
GU20-GRIDW	0.106	0.021	0.411	0.461	0.873	0.129
GU10-NEAR	0.112	0.047	0.369	0.472	0.841	0.136
GU10-CEN	0.093	0.036	0.381	0.490	0.871	0.118
GU10-GRIDW	0.072	0.033	0.384	0.512	0.896	0.099

methods are less consistent, although the interpolated approaches generally provide the best results. In particular, GU20-CEN and GU10-GRIDW achieve the highest OAT values (0.886 and 0.896, respectively), while GU10-GRIDW yields the lowest median Brier score (0.099). These results suggest that refining the spatial discretization consistently improves predictive performance, whereas the benefits of the forcing evaluation method are less systematic, with more sophisticated approaches not always yielding the best results. Figure 6.5 presents the distribution of the Brier score across common 500 m elevation bands for the GU20 and GU10 configurations. The overall patterns are consistent with the results reported in Table 6.5. Within the GU10 configuration, GRIDW consistently achieves the lowest median Brier score across all elevation bands, followed by CEN and NEAR. For GU20, the relative performance of CEN and GRIDW is less systematic, with CEN slightly outperforming GRIDW in some elevation bands, while both interpolated approaches generally yield lower Brier scores than NEAR.

The results also indicate a clear improvement in model performance with increasing elevation, a trend that is consistent across configurations. With respect to classification errors,

GU20-CEN records the lowest false positive rate (0.035), while SG-GRIDW provides the lowest false negative rate (0.015). Although GU10-GRIDW does not achieve the best result in either metric individually (false positives = 0.072, false negatives = 0.033), it represents the most balanced compromise between the two. This balance explains its superior overall performance, as confirmed by both the OAT and Brier scores.

To evaluate the consistency in capturing snow dynamics, the model performance was compared with values reported in the literature for similar applications of hydrological models in Alpine regions. Tong et al. (2020) obtained better performance in terms of OAT, but their analysis relied on station-based comparisons and the determination of optimal thresholds under different conditions (e.g., seasons, elevation, land cover). Similarly, Gyawali and Bárdossy (2022) reported improved results, as they calibrated their model using the Brier score as the objective function, achieving values ranging from 0.095 to 0.064. Compared with our results, the number of false negatives is generally lower (except in the SG-CEN configuration), while their model performed better in terms of false positives. Finally, Morlot et al. (2023) showed that, when comparing data at equivalent elevations, the Brier scores are comparable to those obtained in the present study. The overall results are consistent with those reported in the literature, particularly for the GU20 and GU10 configurations.

6.3.2.4 Soil moisture

As the basins are located inside the Alps, with relatively high elevation and complex topography, and given the limited availability of in situ soil moisture data in the region, the validation of soil moisture was carried out using the MERIDA reanalysis product.

Table 6.6 reports the main performance metrics, namely Pearson's correlation coefficient (r) and RMSE, obtained by comparing the model simulated water content with the rescaled MERIDA estimates. Since the SG configuration includes only two spatial units, the reported median values should be interpreted with caution. The SG configurations show relatively

similar performance across the three forcing evaluation methods. In contrast, both GU20 and GU10 achieve their best performance with the NEAR method, while the correlation generally decreases from NEAR to CEN and GRIDW. GU10-NEAR provides the highest Pearson correlation ($r = 0.508$), whereas GU20-NEAR yields the lowest median RMSE (0.0146). Overall, these results indicate that the NEAR method performs better for the GU-based discretizations, although the configuration associated with the highest correlation does not coincide with that producing the lowest RMSE.

Table 6.6 – Summary of the root-zone soil moisture validation metrics for the nine model configurations. Simulated root-zone soil moisture was compared with the MERIDA product over the validation period for each GU. Pearson's correlation coefficient (r) evaluates the agreement in temporal dynamics, whereas the Root Mean Square Error (RMSE) quantifies the magnitude of the simulation error. Reported values correspond to the median of the GU-level metrics within each configuration.

Configuration	Pearson r (median)	RMSE (median)
SG-NEAR	0.434	0.0218
SG-CEN	0.485	0.0163
SG-GRIDW	0.436	0.0154
GU20-NEAR	0.457	0.0146
GU20-CEN	0.387	0.0172
GU20-GRIDW	0.365	0.0164
GU10-NEAR	0.508	0.0153
GU10-CEN	0.426	0.0164
GU10-GRIDW	0.397	0.0170

In particular, Figure 6.6 shows the distribution of the RMSE values across common 500 m elevation bands. Panel (a) presents the GU20 configuration, while panel (b) shows the GU10 configuration; the SG configuration is omitted because it comprises only two GUs. Within each elevation band, the NEAR, CEN, and GRIDW forcing evaluation methods are represented by green, blue, and orange box-plots, respectively.

As shown in Figure 6.6 and Figure D31, the performance of all forcing evaluation methods exhibits a clear dependence on elevation. For both GU20 and GU10, Pearson's correlation generally decreases with increasing elevation, while RMSE shows the oppo-

site behavior, indicating a gradual deterioration of the agreement between simulated and MERIDA-derived root-zone soil moisture at higher elevations. This pattern reflects the influence of elevation and topographic complexity, which cause a decrease in performance at higher altitudes, as also reported by X. Dong et al. (2022).

Within both spatial discretizations, the NEAR method consistently provides the highest correlation and the lowest RMSE across most elevation bands. The differences between CEN and GRIDW are smaller and vary with elevation, with neither method showing a systematic advantage over the other. Overall, the trends are consistent between GU20 and GU10, although the larger number of GUs in the GU10 configuration provides a more detailed representation of the elevation-dependent variability.

MERIDA is a relatively recent dataset covering the Italian territory. Few studies have been published using it, while numerous studies based on ERA5 have compared reanalysis soil moisture with ground measurements or satellite-derived surface soil moisture (X. Dong et al., 2022; Ling et al., 2021; Pasik et al., 2023; Tian & Zhang, 2023; Zheng et al., 2024). The correlations reported in these studies vary depending on the product against which they are compared (other reanalysis product, remote sensing data or ground-based data) and the region (topography, vegetation), as for example arid or frozen areas generally show lower correlation values (X. Dong et al., 2022; Zheng et al., 2024). For instance, X. Dong et al. (2022) compared results with cosmic-ray neutron sensors (CRNS) and found a mean Pearson correlation of 0.74, though many sites were below 0.5. Ling et al. (2021) analyzed monthly series and obtained a maximum correlation of 0.538 with ground measurements. Tian and Zhang (2023) reported a value of 0.64 but with a high bias compared to ground data. Zheng et al. (2024) proposed an exponential model using satellite-derived surface soil moisture, achieving a correlation of 0.54 in the first layer, which degraded to a mean value of 0.28. In general, the correlations are lower than those obtained from direct measurements. However, this is justifiable considering the higher spatial heterogeneity and complex topography of the studied areas, the dependence of model outputs on calibration parameters (therefore, on

values depending on the entire basin), and the influence of the discrepancy in the spatial scale and depth (Peng et al., 2021; Zheng et al., 2024).

The median correlation values are consistent with other studies comparing satellite-derived soil moisture (filtered exponentially) with model outputs, such as Morlot et al. (2023), for similar elevation classes, limited to 1900 m asl due to data reliability constraints (Brocca et al., 2013). It is important to note, however, that this comparison does not represent a strict validation against observed data, but rather a qualitative consistency check with an independent reanalysis product. The results should therefore be interpreted as an indication of the model's ability to reproduce spatial and temporal soil-moisture patterns in a complex Alpine environment, rather than a direct accuracy assessment.

6.4 Conclusion

In this study, nine basin representation configurations were compared using the outputs of a semi-distributed hydrological model. The aim was to assess the influence of three different methods for evaluating meteorological forcings and three approaches for delineating the Geomorphic Unit structure. To achieve this, an automated kriging procedure was implemented to compute semivariograms and generate spatially distributed meteorological fields at each time step.

As expected, the most complex configurations—based on grid evaluation of the meteorological forcings combined with the two geomorphological-threshold delineations (GU20 and GU10)—outperformed the simpler solutions based on stream-gauge delineation. Both configurations proved to be the most stable in terms of simulated discharge at the calibration gauge during the validation period. At the independent validation gauge, characterized by a small contributing area and markedly different hydrological conditions due to its proximity to the mountain peaks, the GU10 configuration exhibited a smaller performance degradation than GU20. Overall, these results indicate that combining a finer spatial discretization with a grid-based evaluation of the meteorological forcings improves both predictive performance

and model robustness.

The comparison with remote sensing and reanalysis data also confirmed the higher reliability of both threshold-based configurations, which achieved the best agreement with the remote sensing data and reanalysis data. Overall, all configurations, including the simplest ones, produced reasonable results. Considering their expected lower performance and the limitations imposed by the experimental design—such as the fixed number of calibration iterations and the use of fixed parameters—these outcomes can still be regarded as satisfactory.

The scale effect associated with the GU delineation method is clearly evident, whereas the effect of the meteorological evaluation method becomes more pronounced—particularly during the validation phase—in the geomorphological-threshold delineation.

The computational demand of the grid-based configurations, which is the main practical constraint identified by this study, is quantified in Appendix E; the remaining limitations are discussed in Section 8.3 and the perspectives they open in Section 8.4. The semivariogram and cross-validation figure sets and the complete rainfall-runoff metric tables are in Appendix D.

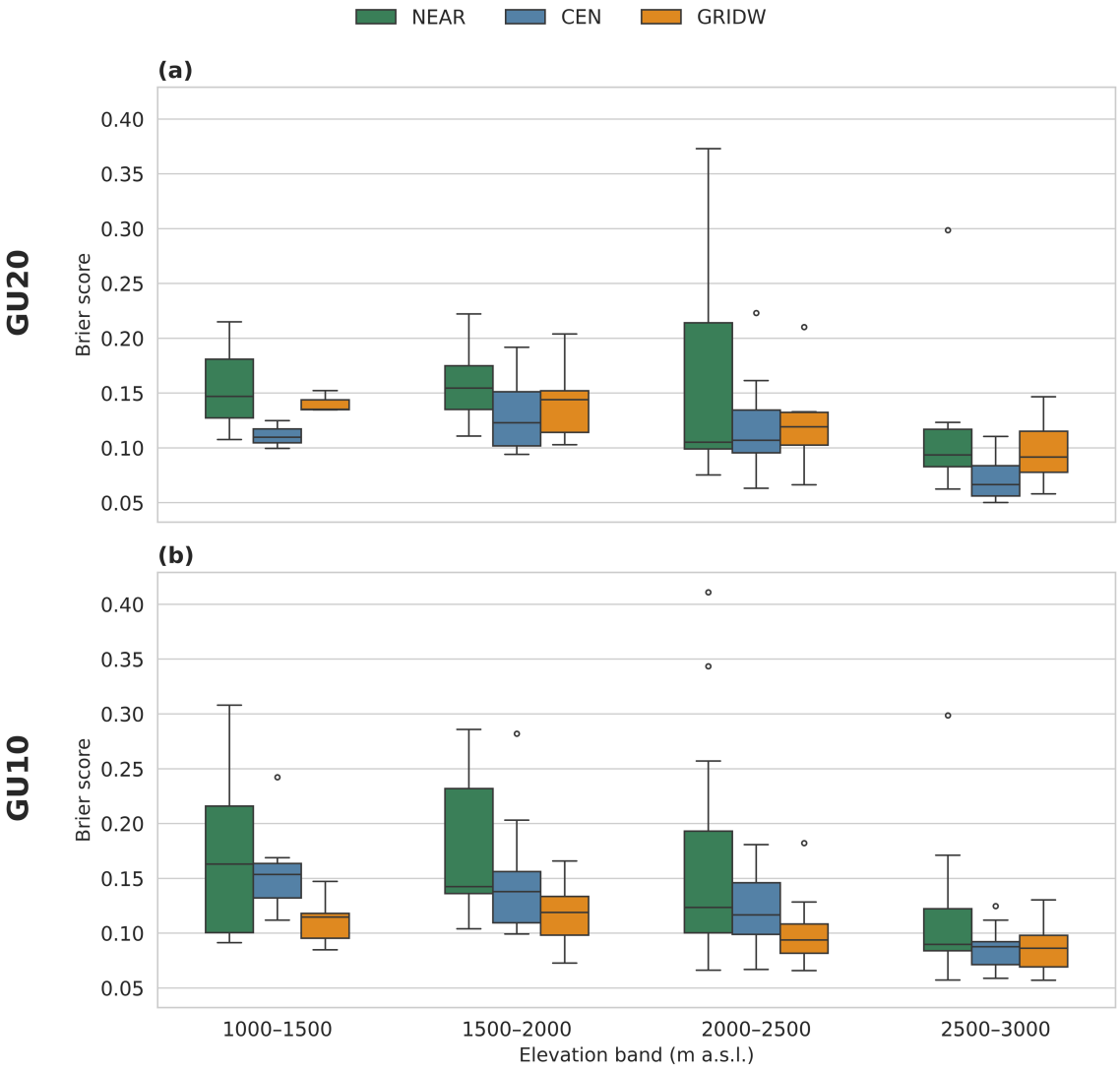

Figure 6.5 – Distribution of the Brier score across common 500 m elevation bands for the GU20 (a) and GU10 (b) configurations, for the snow cover presence/absence modeled versus MODIS. Elevation bands are identical in both panels to enable a direct comparison between the two spatial discretizations. Box-plots summarize the performance of the NEAR, CEN, and GRIDW forcing evaluation methods.

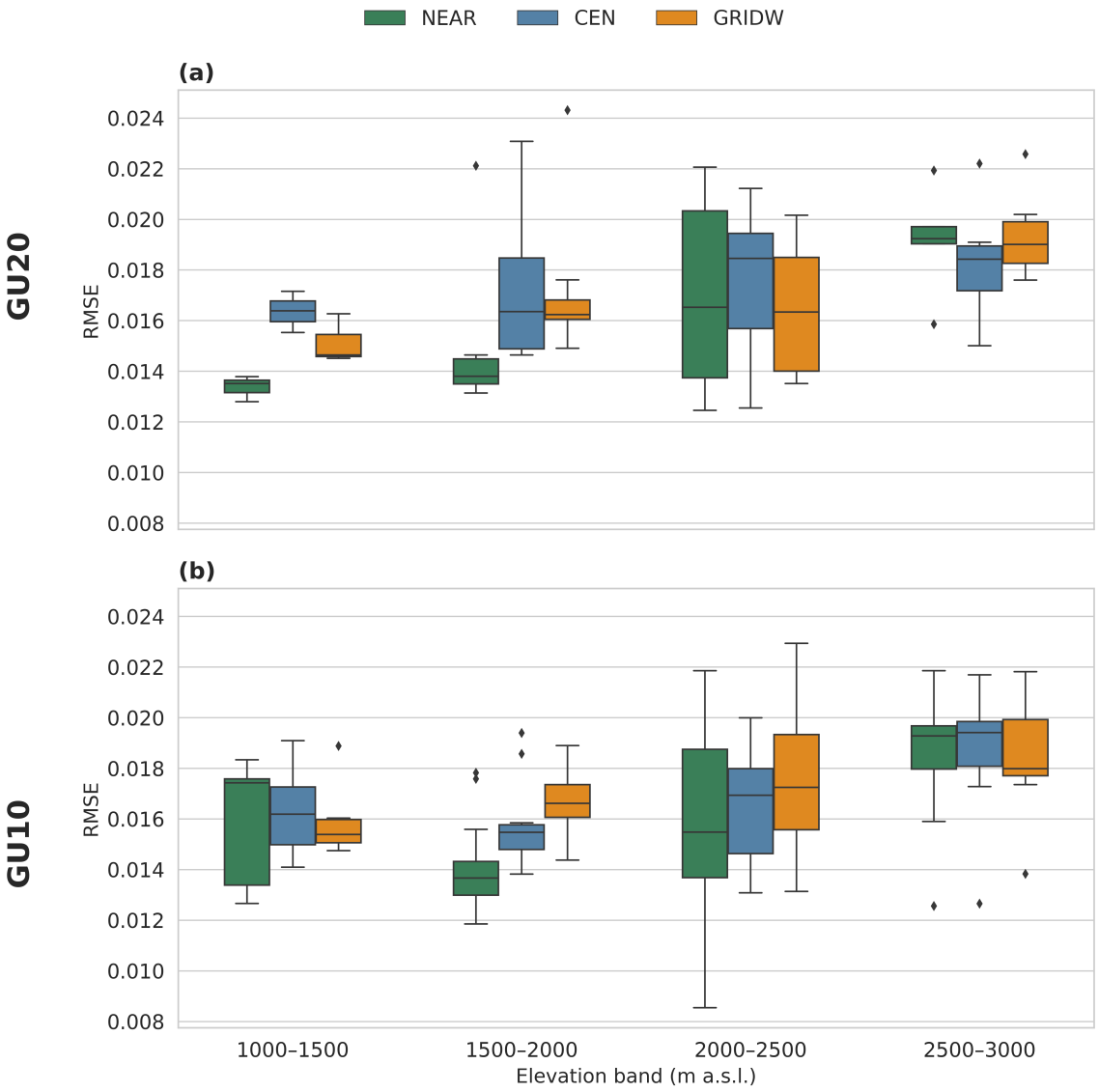

Figure 6.6 – Box-plots of the Root Mean Square Error (RMSE) between simulated root-zone water content and the MERIDA product across common elevation bands of 500 m. Panel (a) shows the GU20 configuration, while panel (b) shows the GU10 configuration. Using identical elevation bands in both panels enables a direct comparison between the two spatial discretizations. Within each elevation band, the box-plots compare the NEAR, CEN, and GRIDW forcing evaluation methods. Lower RMSE values indicate better model performance.

CHAPTER 7. ASSESSING SPATIAL DISCRETIZATION IMPACTS ON SEMI-DISTRIBUTED HYDROLOGICAL SIMULATIONS

This chapter presents the third of the three studies comprising the thesis. It quantifies how the spatial discretization of Geomorphic Units (GUs) influences streamflow performance, internal hydrological functioning, parameter behavior, and computational cost in a semi-distributed rainfall-runoff model under a common model structure and forcing strategy.

Scope and original contributions

Spatial scale remains a persistent challenge in hydrological modeling because catchment behavior emerges from a continuum of interacting processes, whereas models necessarily operate on discrete spatial units. In semi-distributed models, GU delineation determines how landscape heterogeneity is represented and can therefore affect both predictive performance and the attribution of hydrological processes through scale-dependent parameterization and closure assumptions.

This study provides the most comprehensive analysis of spatial discretization conducted in the thesis. Eight configurations of the same basin are generated using contributing-area thresholds ranging from 90 km^2 to 3 km^2, corresponding to between 6 and 145 GUs. The model structure, meteorological forcing strategy, and calibration protocol are kept consistent across configurations so that the effects associated with spatial discretization can be evaluated under comparable conditions. The analysis considers three aspects: streamflow performance, the partitioning of internal fluxes and water-balance components, and scale-dependent parameter behavior and compensation.

The study makes three main contributions. First, the use of eight discretizations spanning more than one order of magnitude makes it possible to examine whether systematic scale-dependent patterns persist across intermediate configurations rather than relying only on a

comparison between two or three spatial resolutions. Second, the evaluation extends beyond discharge at two stream gauges to include internal model states. The snow compartment is evaluated against MODIS snow-cover observations, while root-zone water content is compared with the MERIDA reanalysis across elevation bands. This reveals whether changes in discharge performance are accompanied by physically consistent changes in the internal representation of the basin. Third, a controlled calibration experiment is performed at every scale using a deliberately reduced number of free parameters. This reduces the dimensionality of the calibration problem and limits parameter-compensation effects, allowing scale-dependent parameter tendencies to be examined more clearly.

The results show that changes in GU discretization produce non-trivial differences in both streamflow performance and internal hydrological functioning, while calibrated parameters exhibit systematic scale-dependent tendencies. Discharge performance and the plausibility of internal states do not necessarily improve together. Moreover, beyond a certain level of refinement, the gain in discharge performance is limited to a few hundredths of KGE, whereas computational cost increases several-fold. The analysis therefore clarifies the trade-offs among spatial detail, model performance, computational efficiency, and process interpretability, while also providing an empirical assessment of parameter transferability across scales.

7.1 Introduction

Scale is a key concept in hydrology, as in any discipline where modeling is required. Following the framework proposed by Blöschl and Sivapalan (1995), different notions of scale can be distinguished, including i) the process scale (the characteristic scale of dominant dynamics), ii) the observation scale (the spatial and temporal support of measurements), and iii) the modeling (working) scale (the scale implied by numerical representation and parameterization). In this chapter we focus on the intrinsic model scale, i.e., the modeling scale implied by how the model represents the study system in space and time.

In reality, hydrological behavior emerges from a near-continuous range of interacting scales, from pore and plot processes to hillslope organization and catchment response. A model, by contrast, is necessarily a discretization in space and time. This can create a mismatch between the scale of natural variability and the scale at which model equations are applied, requiring scale mediation through parameterizations and closure assumptions (K. J. Beven, 2001; Blöschl & Sivapalan, 1995; Sivapalan et al., 2003). Because many hydrological processes are non-linear, aggregation is not guaranteed to preserve the relevant dynamics, and the effective response can depend on discretization choices and on how sub-grid variability is represented (Blöschl & Sivapalan, 1995; Sivapalan et al., 2003). Related approaches such as the Representative Elementary Watershed (REW) framework make scale mediation and closure explicit by formulating conservation laws on discrete watershed elements (Reggiani et al., 1998, 1999, 2000).

This mismatch is not a purely technical detail: it is part of a broader, unresolved challenge in hydrological science. Community assessments have emphasized that the relationship between process representation, parameterization, and scale remains a central open problem, directly affecting both predictability and transferability of hydrological models across settings and applications (Blöschl et al., 2019). In other words, scale choices are not only a matter of computational convenience; they can shape what the model can learn from the available data.

Changing spatial discretizations while keeping the model structure fixed can also lead to parameters with different meanings and values (Blöschl, 1999; Sivapalan et al., 2003). This issue is particularly relevant for conceptual and semi-distributed models, where calibrated parameters should be interpreted as effective parameters that depend on model structure, forcing and observation errors, calibration data, and the chosen objective function (K. Beven, 1989; K. Beven & Binley, 1992; K. J. Beven, 2001). Consequently, different discretizations may require different parameter sets, and direct comparisons of parameter values across scales can be problematic. This connects to the long-standing debate on parameter stability

and on parameter behavior under changing modeling choices, including spatial scale (Blöschl et al., 2019).

The impact of spatial discretizations on hydrological modeling has not only theoretical but also practical implications, because discretization affects computational costs, data requirements, calibration effort, and the robustness of predictions used for operational and planning decisions (Merz et al., 2022; Puche et al., 2025). Several studies have tested how changing the number and/or size of sub-catchments, GUs, or grid cells influences simulated discharge and other signatures (Dehotin & Braud, 2008; Puche et al., 2025; Vázquez et al., 2002; Wildemeersch et al., 2014). In parallel, the question of parameter transferability across scales has been addressed through parameter regionalization approaches such as Multiscale Parameter Regionalization (MPR), which links parameters to physiographic descriptors to reduce scale-dependent re-calibration (R. Kumar et al., 2010, 2013; Samaniego et al., 2010). At the same time, increasing spatial resolution or model complexity is not guaranteed to improve performance or reduce uncertainty; therefore, added detail should be justified and evaluated using diagnostic, process-relevant criteria beyond outlet discharge (Gupta et al., 2008; Merz et al., 2022).

In this study we do not seek a single "correct" scale. Instead, we test which discretization is coherent with dominant hydrological processes and with the information content of the available data. We isolate discretization effects under controlled conditions, run the model at an hourly time step, and analyze not only outlet discharge but also internal behavior and parameter tendencies across scales.

Specifically, we keep the meteorological forcing strategy fixed by driving all configurations with the same gridded, spatially averaged precipitation and temperature fields. We also keep model structure, parameter bounds, and calibration strategy unchanged across configurations. We then generate eight GU configurations by varying a single, physically interpretable control: the contributing-area threshold used to derive the drainage network. This produces GU setups with target areas ranging from 90 to 3 km^2. Because the model

is run at hourly resolution, the experiment is sensitive to event dynamics, storage-release timing, and routing effects that are often smoothed out at daily scales.

We evaluate discretization effects in three complementary ways. First, we quantify changes in outlet performance and robustness across flow regimes. Each configuration is calibrated by maximizing the KGE computed on discharge, and is evaluated using the NSE and its logarithmic form NSE_{log} to assess both high-flow and low-flow behavior. Second, we analyze how discretization affects internal flux partitioning and water-balance components. Third, we test whether discretization induces consistent parameter-compensation patterns (equifinality) and whether these tendencies support an empirical cross-scale parameter transfer strategy with limited loss of performance (Gupta & Nearing, 2014; Gupta et al., 2008).

Within this framework, we address the following research questions:

1. How sensitive is the performance of a semi-distributed hydrological model to changes in spatial discretization?

2. How does spatial discretization affect internal flux partitioning and water-balance components?

3. Do changes in spatial discretization lead to consistent tendencies in parameter compensation (equifinality), and do these tendencies support an empirical cross-scale parameter transfer strategy with limited loss of performance?

We implement the experiment in the GEOframe framework (Abera et al., 2017; Formetta et al., 2014a), an open-source, component-based environment that supports reproducible hydrological analyses with transparent implementations and publicly available code (geoframecomponents, n.d.). In this study, we use the conceptual semi-distributed rainfall-runoff model described by Bancheri (2017), with minor bug fixes and adaptations for the present application.

Model evaluation is not limited to discharge at the outlet. In line with diagnostic evaluation concepts (Gupta et al., 2008), we also assess internal states. In particular, we evaluate snow dynamics by comparing simulated snow water equivalent (SWE) and/or snow storage against remotely sensed snow cover area (SCA) derived from MODIS products (Hall & Riggs, 2021; Hall et al., 2006). We additionally analyze root-zone water storage dynamics to investigate how discretization affects subsurface storage and release behavior.

7.2 Experimental setup

The domain is the upper Noce basin down to Malé (Section 2.1.1), over 1 August 2015–31 December 2023.

The focus of this experiment is to investigate the effect of spatial discretization scale on rainfall-runoff model behavior. The experimental design was therefore based on a fixed GU delineation procedure, in which the spatial resolution was controlled by varying the contributing-area threshold from 90 km^2 to 3 km^2. Eight configurations were considered: three coarse configurations with contributing-area thresholds of 90 km^2, 60 km^2, and 30 km^2, followed by five finer configurations ranging from 15 km^2 to 3 km^2 in steps of 3 km^2.

All GUs were generated using the constant-threshold delineation procedure described in Section 3.1.1. The resulting spatial configurations are shown in Figure 7.1, while their main characteristics are reported in Table 7.1.

The number of GUs increases from 6 in the 90 km^2 configuration to 145 in the 3 km^2 configuration, while the median GU area decreases from approximately 88 km^2 to 3 km^2. The range of GU-mean elevations also changes across configurations because the delineation procedure resolves the basin topography differently at each scale.

Meteorological forcing was assigned using GRIDW in all configurations, following the procedure described in Section 3.2. Temperature and precipitation were estimated over the same regular grid with a spacing of 500 m using GEOframe kriging, while radiation and potential evapotranspiration were computed using the GEOframe modeling chain, on the

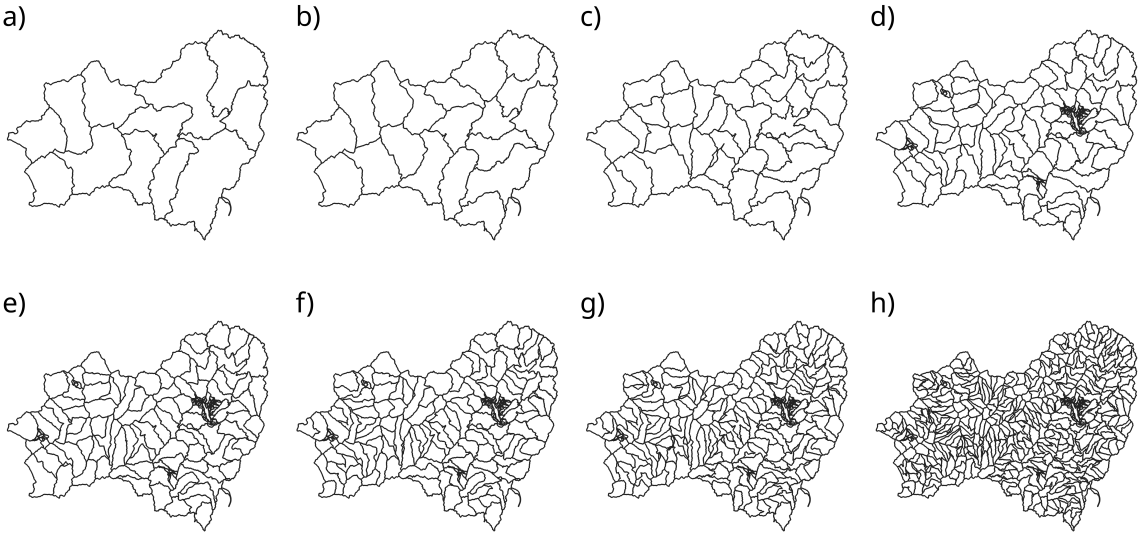

Figure 7.1 – Experimental design showing the eight GU-based spatial configurations obtained by varying the contributing-area threshold from 90 km^2 to 3 km^2 (i.e., 90 km^2, 60 km^2, 30 km^2, 15 km^2, 12 km^2, 9 km^2, 6 km^2 and 3 km^2). Each panel reports the corresponding GU delineation used for the semi-distributed model setup.

grid. For each spatial configuration, the gridded meteorological fields were aggregated to the corresponding GUs using an area-weighted mean. Thus, all configurations used the same underlying meteorological fields and forcing-generation procedure, although the GU-level forcing time series varied with the spatial aggregation imposed by each delineation.

The same hydrological model structure, described in Appendix A, was used for all spatial configurations. Each configuration was calibrated independently using identical initial parameter values and parameter bounds, following the procedures described in Sections 3.4 and 3.6. This common experimental protocol prevents differences in performance from being caused by unequal optimization effort and allows the effects associated with spatial discretization to be evaluated under comparable conditions. To limit parameter compensation, the precipitation-correction factors were fixed to unity, the melting temperature was fixed at 0 °C, and the exponents of the nonlinear reservoirs were fixed to 1.

Calibration was performed sequentially (SC) from upstream to downstream. First, the parameters of the upstream catchment draining to Vermiglio were calibrated. The resulting

Table 7.1 – GU characteristics for the eight spatial configurations. The table reports the number of GUs, the median and maximum GU areas, and the median, minimum, and maximum values of GU-mean elevation.

Configuration	Number of GUs	A_{med} [km^2]	A_{max} [km^2]	z_{med} [m]	z_{min} [m]	z_{max} [m]
90 km^2	*6*	88.3881	124	2196	1656	2481
60 km^2	*7*	78.0835	85	2012	1683	2481
30 km^2	*14*	27.8621	70	2009	1349	2861
15 km^2	*37*	13.7585	30	2121	1369	2996
12 km^2	*42*	11.4466	30	2129	1315	2996
9 km^2	*55*	8.711	17	2050	1292	2996
6 km^2	*72*	6.4527	15	2078	1131	3096
3 km^2	*145*	3.0492	6	2029	908	3238

parameter set was then kept fixed while the larger catchment draining to Malé was calibrated. Consequently, for each spatial discretization, two parameter sets were obtained: one for the upstream Vermiglio catchment and one for the remaining portion of the modeled basin. Within each calibration domain, all GUs shared the same parameter set.

The simulation period was divided into a spin-up period (1 August 2015–1 January 2017) and a calibration period (1 January 2017–30 October 2019). At the Vermiglio stream gauge, the validation period extended from 30 October 2019 to 31 December 2023, whereas at the Malé stream gauge it extended from 30 March 2020 to 31 December 2023 because of gaps in the observed discharge record.

In the snow compartment, the state variable was evaluated against MODIS Terra and Aqua daily NDSI products (Hall & Riggs, 2021; Hall et al., 2006). Snow presence/absence was derived by thresholding both remote sensing and model output (NDSI > 40; SWE $>$ 5 mm) (Dawson et al., 2018; Di Marco et al., 2020; Härer et al., 2018), using the method described in Section 3.6.1.

For the root zone, MERIDA was used as a reference dataset to compare the modeled water content, as described in Section 3.6.3.

7.2.1 Scale-focused calibration experiment

In addition to the full calibration described above, a dedicated scale-focused calibration experiment was designed to better isolate scale effects and limit parameter compensation for the upstream sub-catchment draining to the Vermiglio stream gauge. In this experiment, a subset of parameters that can absorb variability and potentially mask scale-dependent behavior was held fixed across all runs and discretizations. Compared to the previous calibration, which used 12 free parameters, this experiment calibrated only 10 parameters.

The remaining calibrated parameters and their a priori bounds are reported in Table 7.2, while parameter definitions are provided in Appendix A. This reduced-parameter setup prevents flexible parameters from compensating for changes in discretization, making scale-induced differences in model behavior more directly interpretable.

To reduce degrees of freedom and mitigate equifinality, the following parameters were fixed: i) precipitation-correction and threshold parameters; ii) all nonlinear reservoir exponents (set to 1, i.e., linear reservoirs); and iii) canopy-related parameters, effectively removing canopy calibration degrees of freedom (with p acting as the canopy partition coefficient).

Model parameters were calibrated using Particle Swarm Optimization (PSO). For each spatial discretization, 10 independent calibration runs were performed, each initialized with a different random seed. Considering the eight discretizations investigated in this study, this corresponds to 80 calibration runs per stream gauge. This ensemble approach reduces dependence on any single initialization and enables robust cross-scale comparisons of both performance metrics and parameter values.

To ensure comparability across discretizations, the following elements were kept identical across all calibrations and configurations: PSO hyperparameters, objective function (KGE), parameter bounds, and input datasets.

Note that the partitioning between infiltration and runoff also includes a nonlinear term, pB_{soil}. This parameter appears in the storage-capacity distribution approach of Moore

Table 7.2 – Model parameters, initial values, and search ranges used in the second calibration experiment to investigate parameter scaling across spatial configurations. Parameters with zero range were held fixed.

Parameter	Initial value	Range	Unit
α_m	1.5	[0.5, 10.0]	mm/(°C d)
α_f	0.8	[0.001, 1.0]	mm/(°C d)
α_l	0.2	[0.01, 0.3]	–
$S_{rz,\max}$	200.0	[80.0, 350.0]	mm
g	0.005	[0.00001, 0.02]	–
pB_{soil}	2.0	[0.5, 3.0]	–
$S_{ru,\max}$	80.0	[30.0, 120.0]	mm
c	0.4	[0.1, 3.0]	–
$S_{gw,\max}$	1000.0	[200.0, 2500.0]	mm
e	0.002	[0.00005, 0.005]	–

(1985), in which the catchment is represented by a probability distribution of local soil-moisture storage capacities. Runoff is generated when rainfall exceeds the local storage capacity, and the resulting capacity distribution is used to determine the partition coefficient α between runoff and infiltration, where α depends on the precipitation (M_d), the state and the maximum storage of the root zone (S_{rz}, $S_{rz,max}$) and the parameter pB_{soil}. In this formulation, pB_{soil} controls the nonlinearity by acting as the exponent of a Pareto-type distribution.

7.3 Results and discussion

Results from the scale-dependency modeling experiment described in the previous section are reported and interpreted by progressing from overall predictive skill based on streamflow comparisons at the gauges to more diagnostic, process-oriented evidence. Specifically: i) model performance is first assessed against streamflow observations through calibration and independent validation, using objective metrics and hydrograph comparisons; ii) snow dynamics are analyzed and compared against independent snow cover area (SCA) data; iii) soil-moisture variability is evaluated using the MERIDA product to assess

the plausibility of simulated root-zone dynamics across configurations; and iv) repeated calibrations with a reduced subset of free parameters are performed to investigate parameter sensitivity and stability across spatial scales.

7.3.1 Discharge calibration and validation

Discharge is compared at the two stream gauges. Starting from Vermiglio, the results are summarized in Table 7.3, while the diagnostic KGE components (correlation, bias, and variability) are reported in Table 7.4. In the KGE decomposition, the bias ratio $\beta = \mu_s/\mu_o$ measures systematic errors in mean discharge (volume), with $\beta > 1$ indicating overestimation and $\beta < 1$ underestimation. The variability ratio $\alpha = \sigma_s/\sigma_o$ measures the relative amplitude of fluctuations, with $\alpha > 1$ indicating overly variable simulations and $\alpha < 1$ overly damped variability.

During the calibration period, all spatial configurations achieve high skill with a KGE always above 0.8 and with negligible bias (PBIAS within $\pm 2\%$) in the finest configuration, increasing in the coarser configurations. However, validation performance decreases consistently as the spatial resolution coarsens, exhibiting a systematic positive bias (PBIAS $\approx$ 12–15% for the finer configurations, increasing to approximately 21–23% for the coarser configurations).

Consistently with the KGE results, r increases as the discretization threshold decreases, in both calibration and validation, indicating an improved ability to reproduce the temporal dynamics of the observed discharge at finer discretizations. A different behavior emerges for the other KGE components. The variability ratio α is generally greater than 1, indicating that the simulated discharge is more variable than the observed discharge. Lower α values are generally found for the finer configurations; during calibration, the finest configuration is the only one characterized by $\alpha < 1$. The bias ratio β indicates an overestimation of mean discharge in most cases, with the exceptions of the 15 km^2, 6 km^2, and 3 km^2 configurations during calibration.

Table 7.3 – Streamflow performance at the Vermiglio station for each spatial configuration (90 km^2 to 3 km^2). Metrics refer to discharge-based evaluation for the calibration and independent validation periods. For readability, the table reports KGE, NSE, NSE_{log}, and PBIAS.

	Calibration				Validation			
Scale	KGE	NSE	NSE_{log}	PBIAS [%]	KGE	NSE	NSE_{log}	PBIAS [%]
90	0.811	0.654	-0.060	6.569	0.646	0.453	-0.973	23.228
60	0.816	0.652	0.100	5.198	0.636	0.481	-0.651	21.617
30	0.819	0.665	0.015	6.456	0.651	0.465	-0.865	23.006
15	0.881	0.766	0.561	-1.67	0.719	0.591	0.324	13.23
12	0.884	0.775	0.569	1.25	0.740	0.576	0.323	15.48
9	0.902	0.815	0.612	1.28	0.755	0.659	0.328	15.30
6	0.900	0.806	0.665	-1.67	0.772	0.671	0.466	12.13
3	0.914	0.828	0.618	-0.34	0.792	0.694	0.376	13.39

The finest discretization (3 km^2) provides the best overall validation performance (KGE = 0.792, NSE = 0.694). Interestingly, the 6 km^2 configuration yields the highest NSE_{log} (0.466), indicating a superior ability to represent low-flow conditions and recession limbs compared to both coarser and finer scales.

One of the main challenges at the Vermiglio stream gauge was the 2022 drought year. For instance, under the 3 km^2 configuration, the hydrological year October 2021–October 2022 achieved a markedly lower KGE (0.45), whereas performance improved substantially in October 2022–October 2023 (KGE $\approx$ 0.80). This contrast suggests that model skill at Vermiglio is strongly affected by extreme dry conditions, when discharge becomes more sensitive to input biases and to the representation of storage and recession processes.

To assess whether model parameters are transferable across spatial configurations, we performed an additional test on the 15 km^2 setup at the Vermiglio stream gauge. Specifically, we ran the 15 km^2 model using parameter sets calibrated for other configurations. The resulting KGE values were: i) using the 90 km^2 parameter set: KGE = 0.81 (calibration) and 0.65 (validation); ii) using the 60 km^2 parameter set: KGE = 0.81 (calibration) and 0.64 (validation); iii) using the 30 km^2 parameter set: KGE = 0.81 (calibration) and 0.65 (validation); iv) using the 12 km^2 parameter set: KGE = 0.88 (calibration) and 0.70 (validation);

Table 7.4 – Decomposition of the Kling–Gupta Efficiency (KGE) into correlation (r), bias ratio (β), and variability ratio (α) for calibration and validation across spatial configurations for discharge at Vermiglio stream gauge.

	Calibration			Validation		
Scale	r	β	α	r	β	α
90	0.825	1.070	1.031	0.752	1.302	1.110
60	0.825	1.055	1.021	0.748	1.275	1.174
30	0.832	1.069	1.020	0.758	1.299	1.108
15	0.882	0.984	1.010	0.787	1.152	1.145
12	0.886	1.013	1.016	0.797	1.183	1.052
9	0.906	1.013	1.024	0.834	1.181	1.106
6	0.902	0.984	1.010	0.833	1.138	1.107
3	0.914	0.997	0.997	0.852	1.155	1.064

v) using the 9 km^2 parameter set: KGE = 0.75 (calibration) and 0.70 (validation); vi) using the 6 km^2 parameter set: KGE = 0.79 (calibration) and 0.70 (validation); vii) using the 3 km^2 parameter set: KGE = 0.60 (calibration) and 0.65 (validation).

Compared to the baseline 15 km^2 calibration (KGE = 0.881; Table 7.3), transferring parameter sets calibrated at finer scales to the 15 km^2 model leads to a progressive loss of performance—negligible for 12 km^2 (KGE $\approx$ 0.88) but substantial for 3 km^2 (KGE = 0.60)—indicating limited parameter transferability across scales and suggesting scale-dependent compensatory effects. Moving to coarser scales, the degradation in performance during the calibration period remains relatively moderate, whereas substantially poorer performance is observed during validation, further confirming the limited transferability of parameter sets across spatial scales.

Figure 7.2 presents eight panels (a–h) corresponding to the 90, 60, 30, 15, 12, 9, 6, and 3 km^2 spatial configurations at the Vermiglio stream gauge.

Streamflow performance at the Malé station is summarized in Table 7.5, while the diagnostic KGE components are reported in Table 7.6. During calibration, all spatial configurations achieve high skill (KGE $\approx$ 0.85–0.90; NSE $\approx$ 0.70–0.79), with limited bias (PBIAS between -0.2 and 3.3%) and near-ideal values of the bias and variability ratios

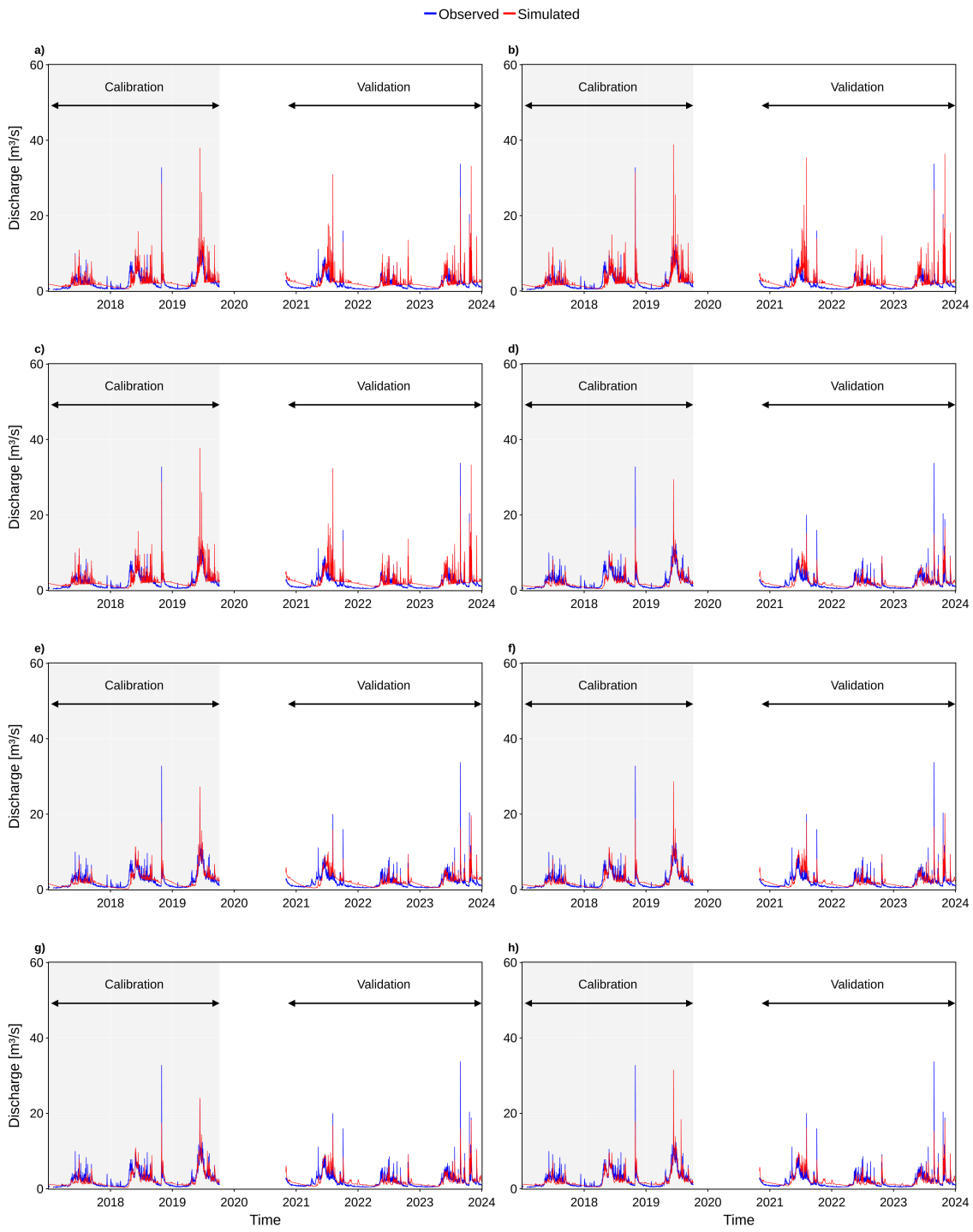

Figure 7.2 – Observed and simulated discharge at the Vermiglio stream gauge for the full study period (August 2015–December 2023). Measured discharge is shown in blue and model-simulated discharge in red. The record includes an initial spin-up phase starting in August 2015, a calibration period from January 2017 to October 2019, and an independent validation period from October 2019 to December 2023. The eight panels compare model performance across spatial configurations: (a) 90 km^2, (b) 60 km^2, (c) 30 km^2, (d) 15 km^2, (e) 12 km^2, (f) 9 km^2, (g) 6 km^2, and (h) 3 km^2.

($\beta \approx 1$ and $\alpha \approx 1$). The three coarsest configurations (90, 60, and 30 km^2) show slightly lower skill (KGE $\approx$ 0.85; NSE $\approx$ 0.70–0.71) than the finer ones (KGE $\approx$ 0.88–0.90; NSE $\approx$ 0.77–0.79), with the 12 km^2 configuration providing the highest calibration KGE and NSE.

In independent validation, two distinct behaviors emerge. The 60 and 90 km^2 configurations show a marked deterioration (KGE $\approx$ 0.57; NSE $\approx$ 0.56), driven by a large positive bias (PBIAS $\approx$ 27%; $\beta \approx 1.38$) and amplified variability ($\alpha \approx 1.40$). From 30 km^2 downward, skill remains high (KGE $\approx$ 0.82–0.88; NSE $\approx$ 0.69–0.79), indicating good transferability to an unseen period. For these configurations, correlation stays strong ($r \approx$ 0.84–0.90), suggesting that timing and seasonal dynamics are largely preserved, and the main degradation is associated with a moderate positive bias (PBIAS $\approx$ 4–8%; $\beta > 1$) and, in some configurations, amplified variability, particularly for the 6 km^2 configuration ($\alpha \approx 1.14$). Among the tested discretizations, the 9 km^2 configuration provides the best overall validation performance and a balanced KGE decomposition, whereas 12 km^2 exhibits the largest positive bias among the configurations of 30 km^2 and finer, despite the highest validation NSE.

Overall, sensitivity to spatial scale at Malé is limited once the GU area is 30 km^2 or smaller: these configurations perform well, and differences are primarily expressed through bias and variability terms rather than through timing errors. By contrast, the two coarsest configurations (60 and 90 km^2) show a substantial loss of skill in validation.

Figure 7.3 presents eight panels (a–h) corresponding to the 90, 60, 30, 15, 12, 9, 6, and 3 km^2 spatial configurations.

A comparative analysis between Vermiglio and Malé reveals distinct patterns in how spatial scale dictates model performance and robustness. While calibration skill is high at both sites, the generalization capability differs significantly. Vermiglio, a smaller headwater catchment, exhibits high sensitivity to spatial discretizations, with a clear performance gain at the finest resolution (3 km^2) and a substantial performance drop in validation (ΔKGE $\approx$ -0.12 to -0.18). This suggests that in headwater catchments, explicitly resolving small-

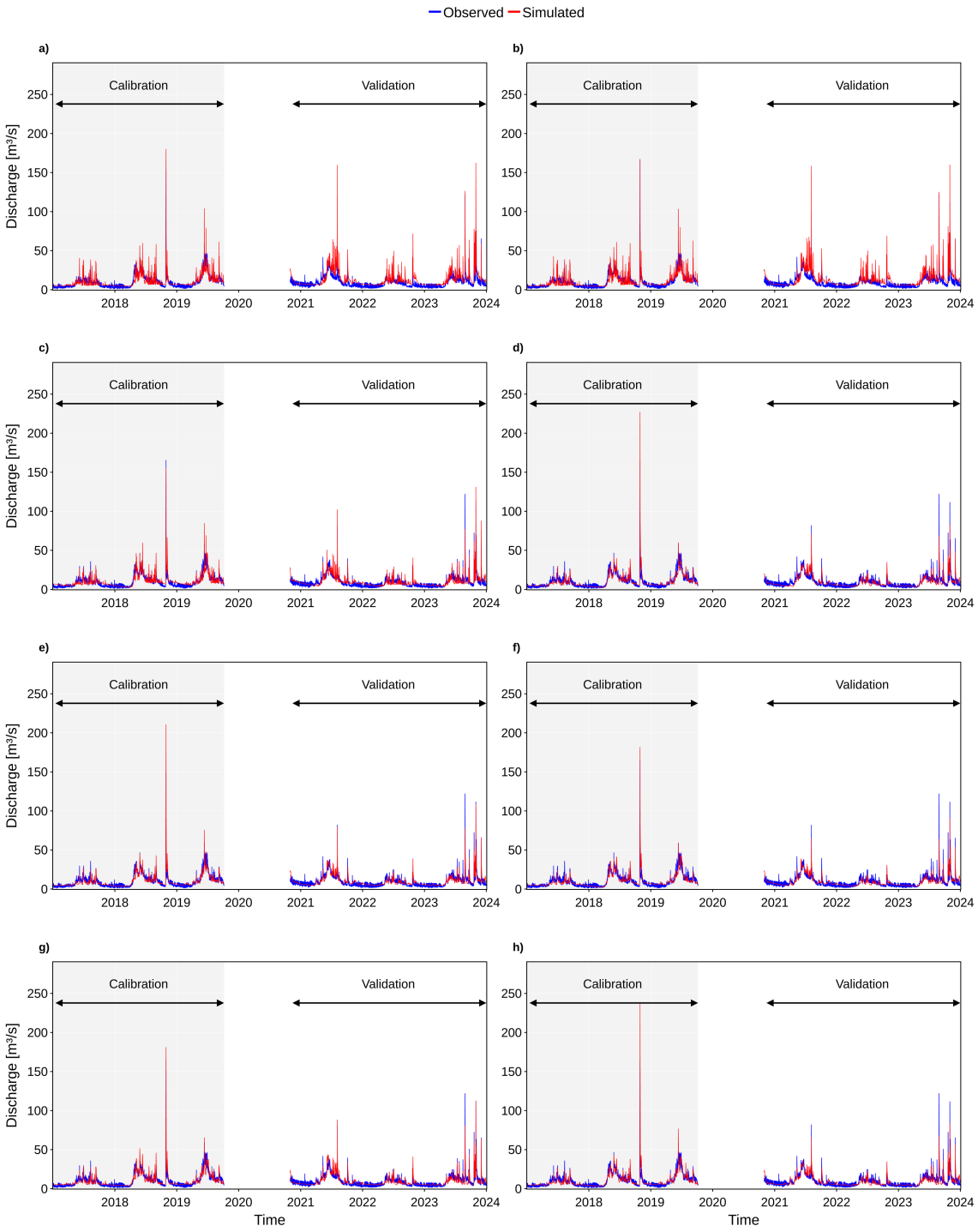

Figure 7.3 – Observed and simulated discharge at the Malé stream gauge for the full study period (August 2015–December 2023). Measured discharge is shown in blue and model-simulated discharge in red. The record includes an initial spin-up phase starting in August 2015, a calibration period from January 2017 to October 2019, and an independent validation period from October 2019 to December 2023. The eight panels compare model performance across spatial configurations: (a) 90 km^2, (b) 60 km^2, (c) 30 km^2, (d) 15 km^2, (e) 12 km^2, (f) 9 km^2, (g) 6 km^2, and (h) 3 km^2.

Table 7.5 – Streamflow performance at the Malé station for each spatial configuration (90 km^2 to 3 km^2). Metrics refer to discharge-based evaluation for the calibration and independent validation periods. For readability, the table reports KGE, NSE, NSE_{log}, and PBIAS.

	Calibration				Validation			
Scale	KGE	NSE	NSE_{log}	PBIAS [%]	KGE	NSE	NSE_{log}	PBIAS [%]
90	0.853	0.703	-0.097	0.656	0.572	0.555	-0.097	27.46
60	0.854	0.707	0.004	1.01	0.569	0.555	0.004	27.63
30	0.852	0.714	0.209	3.26	0.820	0.686	0.209	6.43
15	0.884	0.777	0.657	1.66	0.829	0.753	0.616	5.46
12	0.896	0.792	0.585	-0.22	0.857	0.793	0.456	8.05
9	0.890	0.782	0.592	1.38	0.877	0.775	0.571	4.53
6	0.883	0.772	0.522	1.80	0.831	0.792	0.508	5.27
3	0.888	0.785	0.627	0.79	0.851	0.767	0.612	3.85

scale heterogeneity is critical for capturing runoff generation processes and mitigating systematic biases.

Conversely, the downstream Malé station demonstrates greater robustness, characterized by minimal generalization loss for configurations of 30 km^2 and finer (ΔKGE ≈ -0.01 to -0.06) and moderate sensitivity to GU size, with a marked deterioration from calibration to validation only in the 60 and 90 km^2 coarse configurations (ΔKGE ≈ -0.28). The intermediate 9 km^2 configuration provides optimal stability here, indicating that the larger drainage area acts as a filter, buffering local-scale errors and reducing the need for hyper-resolution. This contrast highlights a trade-off: as the catchment scale increases, the aggregation of fluxes tends to mask the benefits of finer spatial discretizations, whereas smaller basins remain strictly dependent on high-resolution forcing and process representation to maintain predictive skill outside the calibration period.

The results obtained at the Malé stream gauge are consistent with the main message of the previous chapter: performance is strongly influenced by modeling choices (e.g., discretization type), whereas changes in spatial setup tend to produce more moderate variations once a coherent workflow is adopted. In the previous chapter, the best-performing configuration (GU10-GRIDW) reached high predictive skill at Malé (calibration: KGE $\approx$

Table 7.6 – Decomposition of the Kling–Gupta Efficiency (KGE) into correlation (r), bias ratio (β), and variability ratio (α) for calibration and validation across spatial configurations for discharge at Malé.

	Calibration			Validation		
Scale	r	β	α	r	β	α
90	0.854	1.007	0.982	0.834	1.377	1.395
60	0.854	1.010	0.993	0.836	1.382	1.403
30	0.856	1.034	1.009	0.839	1.068	1.054
15	0.887	1.017	1.021	0.871	1.058	1.109
12	0.897	0.998	0.993	0.898	1.088	1.063
9	0.891	1.014	1.002	0.887	1.047	1.022
6	0.885	1.018	1.012	0.892	1.056	1.135
3	0.890	1.008	1.021	0.878	1.040	1.083

0.90, NSE $\approx$ 0.80; validation: KGE $\approx$ 0.86, NSE $\approx$ 0.76), while other configurations showed a wider spread (e.g., SG-CEN and SG-GRIDW). The scale-focused experiment presented here yields validation skill values that fall within a comparable range for configurations of 15 km^2 and finer (validation KGE $\approx$ 0.83–0.88; NSE $\approx$ 0.75–0.79) with slightly better results for NSE_{log} in the 3 km^2 configuration.

Overall, a slightly different scale-dependent behavior emerges at the two stream gauges. Differences in contributing area may partly explain this response, together with differences in the dominant hydrological processes. Vermiglio represents a smaller, higher-elevation catchment, where snow and glacier melt play a more relevant role and the representation of elevation, and consequently temperature, may therefore become particularly important. At Vermiglio, the most evident improvement occurs between the 30 and 15 km^2 configurations: validation KGE increases from 0.651 to 0.719, while PBIAS decreases from about 23% to about 13% and β from 1.299 to 1.152. A different transition is observed at Malé, where the most pronounced improvement occurs at a coarser scale, between 60 and 30 km^2. Validation KGE increases markedly from 0.569 to 0.820, accompanied by a reduction in PBIAS from 28% to 6% and in β from 1.382 to 1.068. The variability ratio also approaches unity, decreasing from $\alpha = 1.403$ to 1.054.

These differences are also evident in the cumulative discharge volumes (Figure D32 and Figure D33). At Vermiglio, simulated and observed cumulative volumes remain relatively close during calibration, whereas a progressive positive divergence develops during validation, indicating an increasing overestimation of discharge volume. This divergence is reduced when moving from 30 to 15 km^2 and toward the finer configurations. At Malé, the scale effect is more evident: the 90 and 60 km^2 configurations show a pronounced cumulative overestimation during validation, whereas from 30 km^2 toward the finer configurations the simulated cumulative volumes remain substantially closer to the observations. The cumulative-volume behavior therefore supports the changes observed in PBIAS and β, highlighting that the main scale-dependent transitions are also associated with a reduction in the overall discharge-volume error. Further insight is provided by the water-balance analysis at the hydrological-year scale at the Malé stream gauge (Figure D34). The 90 and 60 km^2 configurations exhibit a different partitioning of the water-balance components compared with the finer configurations, characterized by lower actual evapotranspiration and larger contributions from storage change and discharge, particularly during the first two hydrological years. Since the model formulation enforces water-balance closure, changes in the representation of one component necessarily affect the partitioning among the remaining components. The lower actual evapotranspiration simulated by the two coarsest configurations therefore leaves a larger fraction of the available water to be partitioned between storage and discharge, consistent with the larger discharge volumes and positive bias observed for these configurations.

Although finer configurations generally outperform the coarsest ones at both gauges, the improvement is not monotonic and, importantly, the most pronounced transition in terms of discharge performance occurs at different spatial scales: between 60 and 30 km^2 at Malé and between 30 and 15 km^2 at Vermiglio. The fact that this transition occurs at a coarser discretization for the larger contributing area at Malé suggests that the scale at which additional spatial refinement becomes beneficial may depend on catchment size

and, indirectly, on the hydrological characteristics represented at different spatial scales. Further refinement beyond these transitions produces comparatively smaller improvements, suggesting diminishing benefits once a sufficiently detailed spatial representation is reached. This observation provides a useful starting point for further investigation across a larger number of catchments and spatial scales. In particular, future research could investigate adaptive refinement strategies, starting from a coarse discretization and progressively refining only within the range where the largest performance gains occur, such as from 60 to 30 km^2 at Malé and from 30 to 15 km^2 at Vermiglio. These findings have implications beyond the theoretical assessment of scale effects, as they also highlight a practical trade-off between model performance and computational cost. Once a sufficiently detailed discretization is reached, the marginal improvement in discharge performance may become small relative to the additional computational demand. Quantifying the point at which further refinement is no longer computationally advantageous is beyond the scope of this analysis and depends on the available computational resources and modeling objective. Nevertheless, the computational cost observed here illustrates the relevance of this trade-off. For example, for Malé, the computationally most demanding of the two study areas, refining the discretization from 30 to 15 km^2 increases the runtime of a complete simulation from approximately 3.5 to 9 minutes under the same computational setup. While this difference is relatively small for a single simulation, it becomes substantial during calibration, where each optimization round may require hundreds of model evaluations. Thus, identifying a spatial discretization beyond which performance gains become marginal may provide not only hydrological insight but also a practical criterion for balancing model accuracy and computational efficiency.

A comparison with earlier studies on the same basin further supports this interpretation. Bellin et al. (2016) and Majone et al. (2016) calibrated at a daily timestep using NSE as the objective function and generally reported higher NSE values, particularly at Malé, where NSE ≈ 0.93 in calibration and ≈ 0.91 in validation were reported (Majone et al., 2016). The lower NSE values obtained here are therefore plausible given the hourly timestep and

the associated sensitivity to input biases, timing errors, and short-term process dynamics that are partially smoothed at the daily scale. Overall, differences in timestep and objective function must be considered when comparing absolute performance levels across studies, while the relative ranking of modeling choices remains informative for interpreting scale effects.

7.3.2 Validation of snowpack and root-zone dynamics

The snow evaluation starts by comparing MODIS NDSI with model-derived SWE to define binary snow/no-snow conditions. The selected thresholds are 40 for NDSI and 5 mm for model-derived SWE. Based on these thresholds, we derived binary outcomes and computed global performance metrics, including false positives (FP), false negatives (FN), overall accuracy (OAT), and the Brier score (Table 7.7). The Brier score spans from 0 (perfect agreement) to 1 (poorest probabilistic performance) (Brier, 1950).

Across the eight spatial configurations (90 to 3 km^2), performance improves consistently as the spatial resolution becomes finer. The 3 km^2 configuration achieves the lowest Brier score (0.102) and the highest overall accuracy (0.898), while the coarsest configuration (90 km^2) shows the weakest performance (Brier = 0.170; OAT = 0.830). Intermediate scales (6 km^2 to 12 km^2) cluster closely, with only minor differences in global metrics, suggesting that most of the gain is realized at the finest resolution. The largest deterioration occurs above 30 km^2, where the Brier score increases from 0.114 for the 30 km^2 configuration to approximately 0.165 and 0.170 for the 60 and 90 km^2 configurations, respectively.

These findings are in line with studies that explicitly calibrate snow-related parameters to reduce the Brier score and thereby improve agreement with satellite-derived snow information (Gyawali & Bárdossy, 2022). In terms of accuracy, our OAT values (0.830–0.898) are comparable to the range reported in the literature for MODIS-based snow validation, with some studies reaching values around 0.90 under favorable conditions (Tong et al., 2020).

The global statistics also reveal a trade-off between error types. At coarser scale (90 km^2),

FN is lower but FP is higher, indicating a tendency to classify snow presence more often (fewer misses, more false alarms). Notably, as the discretization size increases, the fraction of false negatives decreases, whereas the fraction of false positives increases.

At finer scale (3 km^2), both the Brier score and OAT improve, pointing to a more balanced probabilistic performance overall. Finally, the elevation-band box-plots in Figure 7.4 confirm the same pattern: at most elevation classes, the 3 km^2 configuration exhibits lower median Brier values and a generally tighter spread, while coarser configurations tend to show higher scores, consistent with the global ranking observed in Table 7.7.

Also in this analysis, which considers the entire basin upstream of Malé, a marked improvement emerges between the 60 and 30 km^2 discretizations, consistently with the transition identified in the discharge analysis. The Brier score decreases from 0.165 at 60 km^2 to 0.114 at 30 km^2, while the overall accuracy increases from 0.835 to 0.886. This improvement is mainly associated with a substantial reduction in the fraction of false positives, from 0.150 to 0.084, although the fraction of false negatives increases from 0.016 to 0.030. Beyond this transition, further spatial refinement results in smaller and non-monotonic changes, although the finest 3 km^2 configuration ultimately achieves the lowest Brier score (0.102) and the highest overall accuracy (0.898).

Compared with the configuration-based results reported in the previous chapter, the scale-aggregated snow classification metrics show broadly consistent performance levels. In particular, the best results obtained here (e.g., OAT close to 0.90 and Brier close to 0.10 at the finest scales) are very similar to those achieved by the best configuration previously identified (GU10-GRIDW; OAT = 0.896 and Brier = 0.103). At the same time, some alternative configurations in the previous chapter (notably the SG-based setups) exhibited slightly lower overall accuracy and higher Brier scores, which are comparable to the degradation observed when moving toward the coarsest spatial scale. Overall, this suggests that differences attributable to scale and to configuration are of similar magnitude, and that the snow classification skill is relatively robust provided that an adequate discretization/forcing

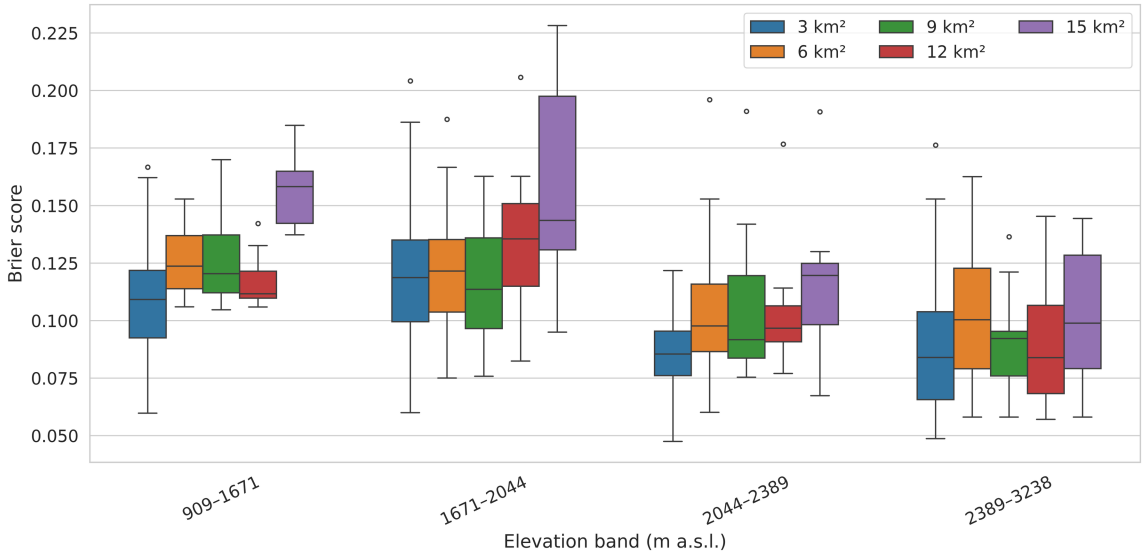

Figure 7.4 – Brier score distribution across elevation bands for the five spatial configurations (3, 6, 9, 12, and 15 km^2). Values obtained comparing the binary snow presence/absence series derived from MODIS observations with those simulated by the model. Elevation bands are defined using common quantile-based classes derived from all configurations to enable a consistent comparison across scales. Boxes show the interquartile range (IQR) with the median, and whiskers indicate data spread. The Brier score measures the accuracy of probabilistic predictions: lower values indicate better performance, whereas values approaching 1 denote poor performance.

strategy is adopted.

A complementary view is provided by the box-plots in Figure 7.4, which consider only the finer configurations, as the coarser ones contain too few GUs to provide a meaningful distribution. The median Brier scores are comparable to those obtained in the previous chapter (Figure 6.5), and the spread of the distributions is similar to that of the best configuration identified there (GU10-GRIDW), with the lowest values reaching approximately 0.05 and no GU exceeding about 0.23. Unlike the NEAR-based configurations of the previous chapter, no pronounced upper tail is observed, indicating that the scale-based configurations do not produce a subset of GUs or elevation bands with markedly poorer snow classification skill. Since the elevation bands are defined differently in the two analyses (quantile-based here, 500 m bands in the previous chapter), this comparison is qualitative.

As the basin is located at high elevation, no in situ or remote sensing data for soil moisture are available. Therefore, the root-zone soil moisture was evaluated against the

Table 7.7 – Summary of classification outcomes and performance metrics, for snow presence/absence, aggregated (sample-size weighted) across elevation bands for each spatial configuration. N is the total number of evaluated samples (sum of n over elevation bands); finer spatial configurations contain more spatial units and therefore yield larger N. FP/FN/TN/TP are the fractions of false positives, false negatives, true negatives, and true positives (FP+FN+TN+TP $\approx$ 1). OAT is overall accuracy (OAT = TP + TN). The Brier score evaluates probabilistic predictions: lower values indicate better performance, while values approaching 1 indicate poor performance.

Scale [km^2]	N	FP	FN	TN	TP	OAT	Brier
90	7936	0.156	0.014	0.411	0.419	0.830	0.170
60	9414	0.150	0.016	0.431	0.404	0.835	0.165
30	18988	0.084	0.030	0.509	0.378	0.886	0.114
15	41480	0.113	0.020	0.462	0.405	0.868	0.132
12	47776	0.080	0.031	0.503	0.385	0.888	0.112
9	63890	0.078	0.033	0.510	0.379	0.889	0.111
6	84116	0.086	0.028	0.500	0.386	0.885	0.115
3	163893	0.062	0.040	0.527	0.370	0.898	0.102

independent MERIDA reanalysis product. Figure 7.5 summarizes the distribution of the Pearson correlation coefficients for the GUs across four elevation bands for the five finer spatial configurations (15, 12, 9, 6, and 3 km^2). Overall, median correlations are comparable across scales and do not indicate a clear, systematic dominance of one configuration over the others. The finest configuration (3 km^2) exhibits a noticeably wider interquartile range and longer tails, which is consistent with the substantially larger number of GUs (and thus higher within-configuration heterogeneity) compared to coarser setups.

Across all configurations, the box-plots show a marked degradation of correlation with increasing elevation, suggesting that model–reanalysis agreement decreases in high-altitude terrain. This behavior is consistent with previous findings that reanalysis-based soil-moisture skill is strongly modulated by physiographic complexity (e.g., topography, vegetation, and frozen/snow-affected conditions), with lower correlations often reported in challenging environments (X. Dong et al., 2022; Zheng et al., 2024). More broadly, published correlations from studies comparing reanalysis soil moisture with ground observations or satellite-derived products span a wide range depending on the reference dataset, temporal aggregation, and

regional characteristics (X. Dong et al., 2022; Ling et al., 2021; Pasik et al., 2023; Tian & Zhang, 2023; Zheng et al., 2024).

The magnitude of the median correlations is broadly consistent with previous studies comparing modeled soil moisture with satellite-derived estimates (often after temporal filtering), although the values reported here are slightly lower (Brocca et al., 2013; Morlot et al., 2023). This difference is not unexpected, as many published comparisons involve different target variables, depths, spatial supports, and reference datasets (e.g., satellite surface products or in situ measurements), and not all studies directly evaluate model-derived root-zone states against large-scale gridded products. Importantly, this exercise should not be interpreted as a strict validation against observations; rather, it provides a qualitative assessment of whether the model reproduces plausible spatial and temporal soil-moisture patterns when confronted with an independent large-scale reanalysis product (Peng et al., 2021; Zheng et al., 2024).

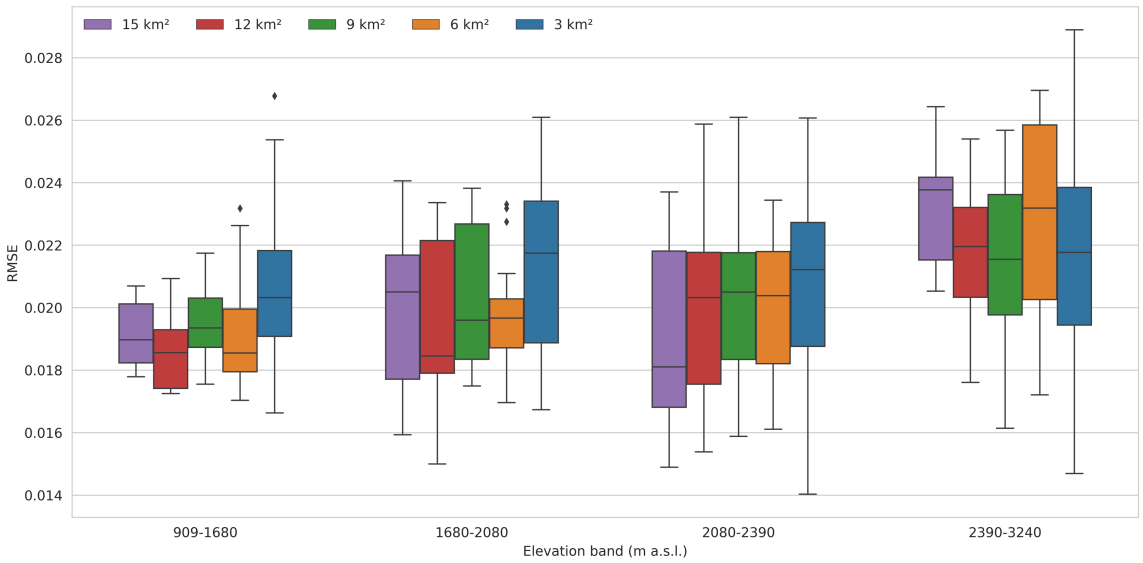

Figure 7.5 – Pearson correlation coefficient (r) between modeled root-zone water content and MERIDA soil-moisture estimates across elevation bands for the five spatial configurations (3, 6, 9, 12, and 15 km^2). Elevation bands are defined using common quantile-based classes derived from all configurations to enable a consistent comparison across scales. Boxes show the interquartile range (IQR) with the median, and whiskers indicate data spread. Pearson's r quantifies the strength of linear association between the two time series: values close to 1 indicate strong positive co-variation (similar temporal dynamics), values near 0 indicate weak association, and negative values indicate opposite co-variation.

It is crucial to stress that MERIDA provides a reanalysis-based estimate rather than direct observations, and therefore this comparison cannot be considered a strict validation in the classical sense. Reanalyses combine model physics with data assimilation and are affected by structural assumptions, forcing uncertainties, and smoothing effects that may reduce local variability, especially in complex terrain (X. Dong et al., 2022; Zheng et al., 2024). In addition, several sources of uncertainty limit the interpretability of the goodness-of-fit metrics: i) a mismatch in spatial support (the model GUs differ from the reanalysis grid cells) (Peng et al., 2021; Zheng et al., 2024), ii) a mismatch in effective depth and representativeness (the model provides only the root-zone water content, not a depth), and iii) elevation- or season-dependent errors related to topographic complexity, vegetation, and snow/frozen-soil conditions, which are known to reduce soil-moisture skill in mountainous or cryospheric environments (X. Dong et al., 2022; Pasik et al., 2023; Zheng et al., 2024). Furthermore, the model is calibrated against basin-scale discharge, so internal states (storage) are only indirectly constrained: due to equifinality, alternative parameter sets could reproduce discharge equally well but with different soil-moisture dynamics. Therefore, the comparison with MERIDA should be interpreted as a qualitative consistency check with an independent reanalysis product, rather than a direct measure of soil-moisture accuracy.

7.3.3 Calibration experiments results

The second calibration experiment on the Noce basin was performed to obtain an ensemble of ten independent calibration sets for each spatial configuration, using a reduced number of free parameters, with the aim of better understanding parameter behavior. Although the model allows up to 18 parameters to be calibrated, only 10 were considered free in this experiment to reduce the dimensionality of the calibration problem and limit compensatory effects among parameters. Accordingly, some model components were simplified; for instance, all reservoirs were represented using a linear formulation, as described in Section 7.2.1.

Table 7.8 reports the best-performing run, in terms of calibration KGE, among the

Table 7.8 – Best-performing (maximum-KGE) streamflow metrics at the Vermiglio station for each spatial configuration (90 km^2 to 3 km^2). For each configuration, values refer to the single best run among the 10 independent calibrations. Reported metrics (KGE, NSE, PBIAS, and NSE_{log}) are computed on discharge for the calibration and independent validation periods.

	Calibration				Validation			
Scale	KGE	NSE	PBIAS [%]	NSE_{log}	KGE	NSE	PBIAS [%]	NSE_{log}
90	0.819	0.705	6.61	0.602	0.632	0.574	19.88	0.331
60	0.803	0.668	5.11	0.361	0.654	0.540	19.23	0.013
30	0.833	0.697	4.84	0.410	0.700	0.595	18.63	0.004
15	0.855	0.726	4.00	0.473	0.710	0.626	17.70	0.100
12	0.853	0.706	-0.81	0.017	0.685	0.565	23.84	-1.532
9	0.868	0.735	0.71	0.066	0.657	0.586	27.88	-2.263
6	0.871	0.743	-1.09	0.257	0.703	0.649	23.56	-1.126
3	0.870	0.738	0.16	0.296	0.659	0.641	27.69	-0.396

ten independent calibrations performed for each spatial configuration. Calibration KGE shows a general tendency to increase toward finer discretizations, from approximately 0.80–0.82 for the coarsest configurations to approximately 0.87 for the finest ones, although the improvement is not strictly monotonic. This scale dependence is less evident during validation, where KGE ranges between 0.632 and 0.710 without a consistent improvement toward finer configurations.

Despite the substantial reduction in the number of free parameters, the best runs retain satisfactory calibration KGE values, indicating that a more parsimonious parameterization can reproduce the observed discharge dynamics reasonably well. Validation performance is lower and cannot be considered fully satisfactory; nevertheless, the obtained KGE values may be regarded as acceptable in the context of this exploratory experiment. In particular, the Vermiglio station represents the upper, partly glacierized portion of the basin, where snow accumulation and melt, glacier contributions, and precipitation measurement uncertainties complicate streamflow simulation. Moreover, the restricted parameterization excluded some parameters that directly affect the simulated water volumes, including α_r and α_s, which account for the correction of rainfall and snowfall inputs and associated gauge-undercatch effects. The model therefore had limited flexibility to compensate for possible biases

in the meteorological forcing and catchment water balance. However, considering the complete ensemble rather than only the best-performing run reveals a substantial spread in calibration KGE at most spatial scales, with some finer configurations also producing relatively poor-performing parameter sets. Spatial refinement therefore appears to increase the attainable calibration performance, but does not systematically reduce the variability among independent calibration runs. This variability should also be interpreted in light of the deliberately reduced computational effort adopted for this exploratory experiment, which used fewer optimization iterations than the main calibration procedure. Increasing the number of iterations and assessing convergence more systematically should therefore be considered in future experiments.

The ensemble was then used to investigate whether the calibrated parameters exhibit systematic changes with spatial scale. Substantial parameter variability is observed across the independent calibration runs, with several parameters approaching the imposed bounds (Figures D36 and D35). The figures show, for each spatial configuration, the distribution of the calibrated parameters using the median and 10th–90th percentiles. This variability is consistent with the presence of equifinality: parameter sets with different combinations of values can produce similar discharge performance. Consequently, the parameter set associated with the maximum KGE represents only one possible solution. Nevertheless, examining the best-performing sets provides a first indication of how parameter values may change with spatial scale.

Focusing on the free parameters of the best-performing calibration for each configuration, no parameter exhibits a strictly monotonic relationship with spatial scale, although some scale-dependent patterns emerge. A particularly marked change is observed for e, which is on the order of 10^{-3} for the two coarsest configurations (1.16×10^{-3} at 90 km^2 and 1.01×10^{-3} at 60 km^2), but decreases by approximately one order of magnitude from 30 km^2 onwards, remaining within 5.1×10^{-5}–9.0×10^{-5} across the finer configurations. The melt factor α_m shows a less regular but overall increasing tendency toward finer discretizations, from 0.88

at 90 km^2 to 1.97 at 3 km^2, with intermediate fluctuations. In contrast, pB_{soil} does not show a consistent scale-dependent trend, remaining close to 0.5 for several configurations but exhibiting larger values at 30 km^2 (0.76) and, particularly, at 15 km^2 (2.60). Storage-related parameters similarly display substantial non-monotonic variability across scales.

The ensemble analysis further shows that some parameters are correlated, suggesting compensatory behavior. A relevant example is the inverse relationship between $S_{rz,\max}$ and g, which emerges across the spatial configurations, although with varying strength. Other relatively strong associations occur only at specific scales and are not consistent across all configurations. These interactions indicate that apparent scale-dependent changes in individual parameters should be interpreted cautiously, as changes in one parameter may be offset by adjustments in others. Given the limited number of independent calibrations performed for each configuration, these correlations should be regarded as exploratory rather than conclusive. Overall, the results do not support a simple monotonic relationship between individual parameter values and spatial scale and remain consistent with the persistence of equifinality despite the reduced dimensionality of the parameter space.

The experiment can be refined by revisiting which parameters are fixed versus free. For instance, a test in which the maximum storage volume was fixed while keeping the exponent of the nonlinear reservoir formulation free yielded the best performance; in the 3 km^2 configuration, KGE reached 0.92, comparable to the calibration with 12 free parameters. This preliminary result suggests that reducing the number of free parameters does not necessarily imply a corresponding loss of model performance and that the selection of which parameters remain free may be more important than their total number. Future work will aim to reduce the number of free parameters to better isolate specific behaviors, increase the ensemble size and/or focus on "behavioral" runs with very high KGE, and automate the workflow to compute additional diagnostics for each calibration run (e.g., Brier score against MODIS-based products and Pearson correlation against MERIDA or other remote-sensing datasets).

7.4 Conclusion

Studying how parameters vary with spatial scale is challenging because several parameters tend to co-vary and compensate for each other, making individual parameter trends difficult to interpret. Future work should therefore adopt a more targeted experimental design, for example by fixing a larger subset of parameters. However, this requires defensible fixed values, which are not always straightforward to define (e.g., pB_{soil}, which is linked to a probabilistic formulation). Incorporating additional constraints from remote sensing across all parameter sets could support ensemble selection beyond discharge-based metrics alone (i.e., avoiding reliance on KGE as the sole criterion). Future developments should also extend the evaluation downstream and include independent soil-moisture information, either from remote sensing products or, where available, ground-based measurements.

Scale issues in hydrological modeling have a long history and remain a cutting-edge topic, both theoretically—highlighted among the "unresolved problems in hydrology" by Blöschl et al. (2019)—and in practice, for example in the push toward hyper-resolution modeling and the persistent challenge of parameter transferability across scales.

In this chapter, we investigated how model performance and internal consistency vary across multiple basin configurations defined by the spatial scale of the GUs within the GEOframe framework for an Alpine catchment. To enable a controlled comparison, we kept the model structure unchanged and forced all configurations with the same gridded meteorological inputs, ensuring that spatial variability in forcing was represented consistently. Under this setup, we derived parameter sets for each configuration—both with all parameters calibrated freely and with selected subsets held constant—and compared them to assess scale effects and parameter transferability.

As expected, finer GU discretizations generally improved discharge performance, most notably at the upstream Vermiglio gauge (high calibration KGE). However, these gains were not consistently retained in validation, with a marked degradation during the 2022 drought,

which the model struggled to reproduce. In contrast, the downstream Malé gauge showed more stable performance across configurations. Across model variants, the calibration setup based on linear reservoirs achieved good calibration skill but exhibited a stronger loss of robustness in validation, particularly for low flows as indicated by NSE_{log}, reinforcing that parameter transferability across scales remains sensitive to hydroclimatic extremes.

Although performance improves at finer spatial resolution, the discharge gain—often of only a few hundredths of KGE—may not justify the added complexity when discharge is the sole objective. In this context, it is noteworthy that both stream gauges exhibit a marked transition beyond which further refinement provides comparatively smaller improvements: around 30 km^2 at Malé and 15 km^2 at Vermiglio. This suggests the presence of a catchment-dependent scale beyond which the marginal benefit of additional spatial refinement progressively decreases. Importantly, finer resolution also improved the realism of other compartments, especially snow dynamics (e.g., snow presence/absence against MODIS SCA). By contrast, we found no clear evidence of improvement in root-zone water storage, likely because the available reference in Alpine regions relies on reanalysis products, which may limit the detectability of scale-dependent differences.

Finally, the compensatory behavior among model parameters discussed above remains a key limitation in isolating parameter-specific responses to spatial scale. Future work should therefore build on the proposed multi-constraint framework, combining discharge with independent information from remote sensing and ground observations, to better distinguish genuine scale-dependent parameter behavior from calibration-induced compensation. Such an approach would provide a more robust basis for assessing whether parameterizations can be transferred across spatial scales and for identifying the level of spatial detail that is both hydrologically informative and computationally justified.

The generalization drop between calibration and validation, the range of KGE per configuration and the computational requirements of the eight configurations are tabulated in Appendices D and E. The limitations of the analysis are discussed in Section 8.3.

CHAPTER 8. GENERAL SYNTHESIS AND FUTURE PERSPECTIVES

Climate change is already affecting the European Alps, increasing pressure on water resources and strengthening the need for accurate and reliable hydrological models. In Alpine catchments, where heterogeneity and complexity are the rule rather than the exception, scale becomes a central modeling concern: representing processes and meteorological inputs at sufficiently fine resolution can improve realism, but it also increases computational costs and may introduce additional sources of uncertainty. Therefore, a careful trade-off is required.

This thesis investigates a semi-distributed hydrological model with spatially aggregated (effective) parameters, with the objective of understanding how model behavior responds to changes in spatial scale and identifying an appropriate balance between predictive performance, parsimony, and internal consistency. The model was calibrated exclusively against observed discharge and subsequently evaluated against discharge over an independent validation period. Selected internal states and water-budget components were then compared with complementary information sources, including remote-sensing products and CRNS-derived soil-moisture estimates (Clark et al., 2011; Fenicia et al., 2011; Gupta et al., 2008; Ten Berge et al., 2025; Wagner et al., 2025). Using this approach, the thesis compares multiple basin configurations, including constant-resolution and mixed-resolution (hybrid) discretizations, alternative GU delineations, and different meteorological forcing strategies across several spatial scales.

Overall, the results demonstrate that scale and forcing choices can affect not only outlet performance but also—crucially—the internal coherence of simulated states and the partitioning of fluxes. These findings support scale-aware and purpose-driven model configuration in Alpine catchments, extending evaluation beyond discharge to include the consistency of selected internal states and water-budget components.

8.1 Cross-study synthesis

The three studies do not identify a single configuration that is optimal for every modeling objective. In the first study, the mixed-resolution configuration did not consistently improve discharge performance, but provided a more informative representation of selected internal water-budget components. Its main value therefore lies in targeting additional spatial detail in areas where internal states and fluxes are themselves relevant outputs, rather than in refining the entire basin solely to improve discharge at the outlet.

In the second study, GU10-GRIDW generally provided the strongest overall discharge performance and snow-cover agreement, but required substantially greater computational effort. Its advantage did not extend uniformly to all internal diagnostics, as the NEAR forcing produced the closest agreement with MERIDA-derived root-zone soil-moisture dynamics for the GU-based delineations. Moreover, the much simpler SG configurations already provided satisfactory discharge simulations. In particular, SG-CEN represents a parsimonious option when overall discharge performance and computational efficiency are prioritized, whereas SG-GRIDW provides a better representation of low flows.

The third study provided the clearest evidence of diminishing returns from increasing spatial resolution. Refinement produced modest improvements during calibration but no monotonic gains in either temporal or spatial validation. Below characteristic GU areas of approximately 30 km^2 at Malé and 15 km^2 at Vermiglio, further refinement did not produce consistent validation improvements and, in some cases, increased discharge-volume bias and reduced low-flow skill. These values represent empirical thresholds for the model structure, forcing data, calibration strategy, and study basin considered here and should not be interpreted as universally transferable limits.

Taken together, the results indicate that spatial refinement is beneficial when it introduces information that can be effectively used by the model or when spatially distributed states are relevant to the intended application. Conversely, increasing spatial resolution alone

does not guarantee improved predictive performance. Once the available observations and model structure can no longer constrain the additional spatial detail, further refinement may primarily increase computational demand and sensitivity to calibration choices.

8.2 Main contributions

This thesis provides three complementary contributions to scale-aware hydrological modeling in Alpine catchments. Because model parameters and internal states are conceptual and should be interpreted as effective quantities rather than as uniquely identifiable physical variables (K. Beven, 2006; Blöschl & Sivapalan, 1995), their evaluation must account for the type and spatial support of the available observations. In particular, for internal states that are not directly comparable with measurements, such as root-zone storage, the assessment emphasizes temporal dynamics, characteristic response times, and event timing rather than absolute magnitudes. Within this interpretative framework, the main contributions of the thesis are:

- **A systematic assessment of spatial-discretization and meteorological forcing choices.** The thesis examines how constant and mixed-resolution configurations, alternative GU delineations, and different meteorological forcing strategies affect discharge performance, basin water-balance partitioning, and the behavior of selected internal states. This provides a consistent basis for evaluating spatial configuration choices across different characteristic scales and modeling objectives.

- **A multi-source evaluation showing that the value of spatial complexity is purpose-dependent.** The model was calibrated exclusively against discharge, while independent discharge data and complementary information sources, including CRNS-derived soil-moisture estimates and remote-sensing products, were used to evaluate its predictive performance and selected internal states. The results show that finer spatial discretization and more detailed forcing representation do not produce uniformly

better performance. Their value depends on the variable considered, the evaluation criterion, and the intended application. Additional spatial complexity may therefore be justified when it provides relevant information on internal dynamics, spatial patterns, or water-budget partitioning that cannot be inferred from outlet discharge alone, but it should not be assumed to improve model performance automatically.

- **A parsimonious strategy for targeting spatial detail where it is most informative.** The thesis investigates mixed-resolution configurations in which Geomorphic Units are refined only within selected sub-regions, such as high-elevation or managed agricultural areas. This hybrid approach provides a practical compromise between spatial detail and computational cost and supports purpose-driven configuration choices for variables such as snow water equivalent and root-zone storage. More generally, the results provide practical guidance for selecting a model configuration that remains computationally feasible while being consistent with the intended scientific or operational application.

The software contributions are presented separately in Chapter 4, as they constitute an independent outcome of the thesis rather than an implementation detail of the three studies.

8.3 Limitations and scope

This thesis investigates how meteorological forcing strategies and spatial discretization affect the outputs of a rainfall-runoff model, with particular attention to discharge, water-balance partitioning, and the consistency of selected internal states. The results should be interpreted in light of the following limitations.

Study-area specificity. The analyses were conducted in an Alpine catchment characterized by steep elevation gradients, seasonal snow dynamics, and heterogeneous geology. Parameter transferability was examined within the Noce basin using two complementary strategies. In the first and third studies, parameters were calibrated sequentially from upstream to

downstream (first at Vermiglio, then at Malé with the upstream parameters held fixed). In the second study, the parameter sets calibrated at the Malé stream gauge were applied without recalibration to the upstream Vermiglio sub-catchment for spatial validation. In all cases, the calibrated parameters were subsequently transferred to the farther downstream portion of the basin, where suitable observations for an independent calibration were not available. These procedures evaluate spatial parameter transfer within a hydrologically connected basin; however, they do not demonstrate transferability to catchments with different environmental and hydrological characteristics. A key next step is therefore to test the model and the selected configurations in other basins, where differences in climate, topography, geology, human regulation, and data availability may affect the robustness and generality of the conclusions.

Observation network density and spatial representativeness. The quality of gridded forcings and their uncertainty depend on station density and complex topography. In data-scarce sub-regions, interpolation uncertainty can remain substantial and may influence both calibration outcomes and spatial patterns of simulated states and fluxes.

Nature of computational units. The modeling units are spatially contiguous GUs derived from the drainage network and a contributing-area threshold. They are not physiographic HRUs in the strict sense of Flügel (1995) and can include substantial internal heterogeneity (elevation, land cover, and associated energy-balance conditions). The model therefore targets effective, spatially aggregated behavior rather than locally realistic process dynamics within each unit.

Kriging workflow limitations (smoothing, edge effects, variogram selection, and numerical conditioning). The interpolation workflow is subject to limitations that may affect the reliability of the gridded forcings. First, although kriging provides spatially continuous meteorological forcings that are practical for catchment-scale hydrological modeling,

rainfall fields obtained through deterministic kriging should not be interpreted as detailed reconstructions of fine-scale rainfall processes. The smoothing inherent in the method may attenuate local intensity maxima and sharp spatial gradients and is not designed to preserve rainfall intermittency and lacunarity, such as zero-rainfall areas embedded within an event (Yamamoto, 2005). Second, edge effects can arise near domain boundaries, where the spatial configuration of stations is typically asymmetric and the neighborhood of informing observations is truncated; this may induce biased fields and increased uncertainty at the edges of the domain and in poorly monitored sub-regions. Future work should therefore consider mitigation strategies such as using buffered interpolation domains and including external stations where available. Third, the current variogram-fitting strategy is primarily driven by goodness-of-fit criteria, selecting the model that best fits the empirical semi-variogram without explicit physical constraints (e.g., bounds or priors linked to expected correlation scales and nugget behavior). Therefore, fitted theoretical variogram models may be statistically adequate but physically implausible under certain conditions. A natural extension is to introduce constrained or regularized variogram estimation (e.g., physically plausible parameter bounds and/or weak priors) to improve interpretability across time steps and regimes. Finally, the kriging system may become numerically fragile for specific station configurations (e.g., clustered sensors, near-collinearity, or strong redundancy), potentially leading to ill-conditioned covariance matrices and unstable kriging weights. At present, no dedicated diagnostic test is systematically applied to detect and handle such ill-conditioning (e.g., conditioning metrics, regularization, or adaptive neighborhood strategies).

Snow and ice process representation. Seasonal snow and glacier-melt dynamics are represented using a parsimonious temperature-index approach with spatially uniform melt parameters. This supports robust basin-scale calibration but cannot reproduce known spatial variability in melt drivers (e.g., radiative effects, longwave radiation emitted by the forest canopy, wind redistribution/sublimation, and differences among seasonal snow, firn, and

glacier ice) (e.g., Ellis et al., 2010; Essery et al., 2025; Link & Marks, 1999; Marsh et al., 2024; Musselman & Pomeroy, 2017; Musselman et al., 2017). Process attribution in glacierized or high-elevation zones should therefore be interpreted cautiously, particularly under warming climate conditions.

Hydrogeological complexity and groundwater. The studied basin includes heterogeneous geological settings (e.g., crystalline bedrock, karstic carbonates, and valley sediments), which can strongly affect storage and low-flow dynamics (e.g., Chen et al., 2023; Hayashi, 2020; Somers & McKenzie, 2020). The groundwater component is conceptual and does not explicitly represent hydrogeology-driven spatial variability or inter-unit groundwater exchanges. Because independent groundwater observations are limited, groundwater parameters are calibrated jointly with the other model parameters using discharge alone. Simulated groundwater storage and fluxes should therefore be interpreted as effective model representations rather than as direct estimates of physical aquifer properties.

Calibration constraints, equifinality, and identifiability. Although multiple datasets are used for evaluation, calibration is constrained exclusively by discharge, so parameter equifinality remains possible. Because the model is semi-distributed, increasing the number of GUs does not increase the number of calibrated parameters, which remains constant across spatial configurations. In the controlled scale experiment, only 10 of the 18 model parameters were calibrated, while the reservoir exponents were fixed to unity to reduce calibration dimensionality and compensatory effects. Although this design helped isolate the effect of spatial discretization, it may also have limited model flexibility and contributed to the deterioration observed in validation. However, because the same constraints were applied to all configurations, their role cannot be determined without a dedicated sensitivity analysis.

Snow-cover information versus SWE magnitude. Snow-cover products (e.g., MODIS) can be uncertain in forested terrain due to canopy masking and mixed pixels; moreover, snow cover is not a direct measure of snow water equivalent (SWE). Agreement in snow-cover dynamics therefore does not guarantee correct snow-storage magnitude (e.g., Gascoin et al., 2015; Parajka & Blöschl, 2006; Şorman et al., 2007).

Forcing uncertainty and precipitation undercatch. Precipitation measurement bias, including potential undercatch, remains an important source of uncertainty. Although the model includes precipitation-correction factors, these were fixed to unity in the experiments designed to isolate the effects of forcing representation and spatial discretization. Consequently, residual precipitation biases may propagate into simulated runoff, evapotranspiration, and storage partitioning. When correction factors are calibrated, they may instead compensate for errors in other processes, complicating the interpretation of model parameters and internal fluxes.

Soil-moisture support mismatch (model states vs CRNS and satellite products). The modeled root-zone storage and observation-based soil-moisture estimates are not directly comparable because they represent different quantities and different spatial and vertical supports. In particular, the model does not explicitly define a measurement-equivalent soil depth for its root-zone (or soil storage) state, whereas in situ sensors and CRNS observations have a characteristic footprint and an effective sensing depth, typically ranging from a few centimeters to several tens of centimeters depending on soil properties and moisture conditions. Satellite soil-moisture retrievals introduce additional limitations: they represent near-surface conditions, may be affected by vegetation and surface roughness, and, in this study, are not available daily but only every few days (approximately 1.4 to 4 days; Chapter 2), which limits their ability to capture fast wetting and drying events. Therefore, comparisons should emphasize temporal dynamics and event timing (possibly after temporal aggregation or filtering) rather than absolute magnitudes, and should explicitly acknowledge

representativeness and scaling uncertainties.

8.4 Future research directions

The GEOframe framework provides a modular basis for transparent and reproducible hydrological workflows. In this thesis, the methodological and computational components were documented and shared where licensing and data-access conditions permitted. GEOframe has also been applied in previous studies addressing a range of hydrological objectives (Abera et al., 2017; Bancheri, 2017; Bouabdelli et al., 2025; Formetta et al., 2014a; Morlot et al., 2023). Nevertheless, the framework requires continuous maintenance and updates, as well as new tools to address emerging questions and to improve existing results.

The issues analyzed in this thesis have two main implications: i) a theoretical implication, related to improving process representation and reducing structural limitations of the current modeling chain; and ii) an applied implication, aimed at consolidating and extending the operational, near-real-time modeling pipeline and, ultimately, enabling short-range forecasting.

In both cases, further progress requires targeted extensions of GEOframe components and of the surrounding data-management and automation infrastructure.

8.4.1 Process representation and model-structure development

From a theoretical and process-modeling perspective, a key priority is to overcome structural limitations of the current modeling chain. In particular, we identify three priority directions: i) Represent deeper flow pathways. At present, the model represents local conceptual soil and groundwater storages, while transfers between GUs occur primarily through discharge routing along the drainage network; deep lateral flow and fractured-bedrock exchanges are not explicitly accounted for. However, in mountain catchments, geology and bedrock structure can strongly control subsurface flow paths, storage, and discharge recession behavior, and may lead to partial decoupling between subsurface connectivity and

surface topographic flow directions (Chen et al., 2023; Hayashi, 2020; Somers & McKenzie, 2020). A natural extension is therefore to introduce a deeper storage/transfer domain and explicit exchange fluxes between shallow and deep compartments (e.g., percolation to a deep reservoir, fracture-enabled lateral transport, and return flow to streams). This would provide a more realistic representation of hydrological memory and baseflow persistence, and would support a more defensible attribution of runoff generation mechanisms across seasons and hydrological regimes. ii) Adopt a more physically informed snow model (radiation/aspect) and reduce temperature-only fragility. A second priority concerns snow processes. Many operationally convenient approaches rely on temperature-index (temperature-only) melt formulations, but such methods can be overly sensitive to warming and can miss key radiative forcing interactions that control melt rates and their spatial variability (Comola et al., 2015; Musselman & Pomeroy, 2017). A pragmatic intermediate step is to adopt an enhanced temperature-index formulation that incorporates potential solar radiation and topographic controls (slope/aspect), as proposed by radiation-index approaches (Hock, 1999) and by the distributed radiation-index method of Cazorzi and Dalla Fontana (1996). Basin-scale implementations that discretize radiation classes can retain computational efficiency while improving spatial realism (Zaramella et al., 2019). Within GEOframe, this can be implemented as (a) a native hybrid snow module (temperature index + radiation index) or (b) a coupling strategy where snow states (e.g., SWE) are provided by a more physically based external model. iii) Couple external snow physics when needed, and expand forcing/evaluation targets. An attractive pathway is to couple GEOframe with a physically based snow model (e.g., GEOtop) and integrate snow outputs (e.g., SWE) into the hydrological model, as explored by Wani et al. (2025). In parallel, while this thesis focused on temperature, precipitation, and potential evapotranspiration, both forcing and evaluation could be extended to additional variables and diagnostic targets relevant for snow and energy balance (e.g., shortwave/longwave radiation, wind, humidity) and to observation-based proxies of internal states (e.g., snow cover fraction from remote sensing). These additions

would help reduce equifinality and strengthen process realism.

8.4.2 Calibration strategy and use of auxiliary information

From a methodological perspective, a central question is how the chosen calibration strategy influences model behavior and, consequently, the conclusions drawn from model comparisons. This includes quantifying the sensitivity of inferred results to calibration design choices (e.g., parameter bounds, training period selection, weighting of errors, and treatment of different flow regimes). Another priority is to move beyond single-objective metrics (such as KGE) toward multi-objective calibration. In a multi-objective setting, complementary signatures and variables are balanced explicitly, which helps reveal trade-offs and reduces the risk that good performance at the outlet is achieved through compensating errors among internal processes. At the same time, the increasing availability of auxiliary information sources (satellite products and reanalyses) opens opportunities to better constrain internal states; however, these data must be used with care due to spatio-temporal scale mismatches, retrieval uncertainty, and representativeness limitations. For example, remote sensing provides observation-based proxies such as snow-covered area (and, in some contexts, SWE-related indicators), which can be incorporated into calibration and evaluation alongside discharge to reduce equifinality and improve the physical consistency of the simulated water balance.

In future work, it would be useful to investigate how the number of free calibration parameters affects model performance and to identify which individual parameters have the greatest influence. Particular attention should be given to whether these choices explain the deterioration observed when moving from calibration to temporal and spatial validation.

8.4.3 Further development and short-range prediction of soil moisture

Building on the operational prototype described in Section 4.8, the next development phase should focus on consolidating the reliability of the existing modeling chain. Priorities

include more robust input and output handling, automated detection and recovery from failures, systematic monitoring of incoming observations and model executions, and comprehensive end-to-end testing. These improvements would strengthen the transition from a research prototype to a system capable of supporting routine applications.

Reliability could be further improved by introducing explicit validation checks at each stage of the continuous simulation chain. Before preprocessing, these checks should verify chronological ordering, duplicated or missing time steps, the expected temporal resolution, the availability of all required variables, and the physical plausibility of the observations. After spatial interpolation, the workflow should verify that a forcing value has been generated for every GU and time step, record the number and spatial distribution of the stations used, and flag the use of fallback variograms or other exceptional conditions. After model execution, automated checks should identify missing or non-finite outputs, negative or implausible storage values, discontinuities in the simulated states, and water-balance residuals exceeding a predefined numerical tolerance. The final model states should be accepted as restart conditions only after all these checks have been passed.

The execution procedure should also ensure that an unsuccessful run cannot overwrite or modify the states produced by the most recent valid simulation. Model outputs and restart states should therefore be committed to the database only after successful completion and validation of the entire workflow. Repeating the same execution should not create duplicated records or inconsistent results. Each run should also be associated with a unique identifier and with information describing the input-data period, parameter set, model configuration, and software version used. If delayed or corrected observations become available, the system should be able to restart from the latest valid checkpoint and repeat the affected portion of the simulation chain. Finally, the workflow should be tested under deliberately adverse conditions, including missing, delayed, duplicated, or malformed observations, unavailable stations, interrupted model runs, and incomplete database writes.

Once the operational chain has been sufficiently validated, it could be extended to

produce short-range predictions over periods of a few days by combining the most recently persisted model states with meteorological forecasts. The primary prediction target would be root-zone soil moisture because of its relevance to agricultural drought monitoring and water-management applications. Predictions could be evaluated against independent observations, including CRNS estimates and measurements from in situ soil-moisture probes. Tensiometers could provide complementary information on soil-water status, although they measure soil-water potential rather than water content; a direct comparison would therefore require an appropriate site-specific water-retention relationship.

The evaluation should account for the differences in spatial and vertical support between the modeled quantities, CRNS estimates, and point-scale measurements discussed in Section 8.3. Particular attention should be given to temporal dynamics, wetting and drying patterns, event timing, forecast lead time, and the persistence of anomalously dry or wet conditions. These aspects are more informative for the intended application than agreement in absolute soil-moisture values alone.

The web application could be extended to support these predictive uses. The interface should clearly distinguish observed, simulated, and forecast periods and report the initialization time and prediction horizon of each forecast. Additional status indicators could show the latest successful data acquisition and model execution, the completeness of the input records, unavailable stations, and failed processing steps. Comparative views of observations, simulations, and forecasts, together with information on prediction uncertainty, would make the results easier to evaluate. Controlled data-download functions and documentation of the API endpoints would further facilitate the use of the system by researchers and other authorized users.

8.5 Concluding remarks

Scale is often treated as a secondary modeling choice or implicitly assumed, yet it can strongly shape both seasonal water-balance partitioning and event-scale response. Starting

from the widespread (but not universally valid) assumption that finer spatial detail is always better, this thesis analyzed the trade-off between increasing resolution and computational cost, showing that additional complexity is not automatically beneficial and may even amplify sensitivity to modeling choices (Jakeman & Hornberger, 1993; Merz et al., 2022). The controlled scale experiment showed that finer discretization can improve calibration without producing corresponding improvements in temporal or spatial validation. This indicates that increasing spatial detail beyond the level supported by the model structure, available observations, and calibration strategy may add complexity without improving predictive robustness.

The work provides a documented computational workflow and practical guidance for navigating this trade-off. Complexity is most justified when the objectives extend beyond outlet discharge to internal consistency and state dynamics (e.g., SWE and root-zone storage), whereas parsimonious or hybrid strategies can preserve diagnostic value while keeping operational costs tractable. Ultimately, the value of scale-aware modeling lies in choosing the simplest representation that remains fit for purpose while explicitly recognizing the limits imposed by model structure and by the available observations.

APPENDIX A. COMPLETE FORMULATION OF THE HYDROLOGICAL MODEL

This appendix provides the complete formulation of the semi-distributed hydrological model used consistently in the experiments presented in Chapters 5–7. The model builds upon the formulation originally presented by Bancheri (2017), which has been subsequently improved and extended as described in this thesis. The complete formulation is reported here to provide a common reference for the different applications and to facilitate the comparison of the spatial representations investigated. An overview of the model structure is provided in Section 3.3.

The appendix is organized following the progression of water through the main model compartments. Starting from precipitation and its partitioning into rainfall and snowfall, the formulation proceeds through the snow compartment, the canopy compartment, the root-zone compartment, which also controls infiltration, the runoff generation component, and finally the groundwater compartment. Each section describes one of these components and reports the corresponding equations used in the model formulation.

A.1 Rain/snow separation

Precipitation is partitioned into rainfall and snowfall as:

$$\begin{aligned} P_r(t) &= \alpha_r \, P(t) \left(\frac{1}{\pi} \arctan\left(\frac{T(t) - T_s}{m_1} \right) + \frac{1}{2} \right), \\ P_s(t) &= \alpha_s \left(P(t) - P_r(t) \right). \end{aligned} \tag{A.1}$$

Table A1 – Symbols for the rain/snow separation module.

Symbol	Meaning	Unit
$P(t)$	Precipitation at time t	mm/h
$P_r(t)$	Rainfall at time t	mm/h
$P_s(t)$	Snowfall at time t	mm/h
$T(t)$	Air temperature at time t	°C
T_s	Threshold air temperature	°C
α_r	Adjustment coefficient for rain measurements	–
α_s	Adjustment coefficient for snow measurements	–
m_1	Smoothing degree parameter	°C

A.2 Snow compartment

Snowmelt and freezing model (temperature-index method; Formetta et al., 2014b). The maximum liquid water content in the snowpack is:

$$L_{\max}(t) = \alpha_l \, S_w(t). \tag{A.2}$$

The solid and liquid water storages evolve as:

$$\begin{aligned} \frac{dS_w(t)}{dt} &= P_s(t) + F(t) - M(t), \\ \frac{dL_w(t)}{dt} &= P_r(t) - F(t) + M(t). \end{aligned} \tag{A.3}$$

Solid and liquid water dynamics are defined as:

$$\begin{aligned} M_d(t) &= \max\big(0,\ L(t) - L_{\max}(t)\big), \\ L_w(t) &= \min\big(L(t),\ L_{\max}(t)\big), \\ \mathrm{SWE}(t) &= S_w(t) + L_w(t). \end{aligned} \tag{A.4}$$

Table A2 – Symbols for the snow compartment.

Symbol	Meaning	Unit
$P_r(t)$	Rainfall at time t	mm/h
$P_s(t)$	Snowfall at time t	mm/h
$T(t)$	Air temperature	°C
T_s	Threshold air temperature	°C
α_m	Melt factor	mm/(h °C)
α_f	Freezing factor	mm/(h °C)
α_l	Coefficient for maximum liquid water	–
$M(t)$	Melt	mm/h
$F(t)$	Freezing water	mm/h
$S_w(t)$	Solid water storage	mm
$L_w(t)$	Liquid water storage	mm
$L_{\max}(t)$	Maximum liquid water content	mm
$\mathrm{SWE}(t)$	Snow Water Equivalent	mm

A.3 Canopy

The canopy storage is governed by:

$$\frac{dS_c(t)}{dt} = (1-p)\, M_d(t) - D(t) - \mathrm{ET}_c(t), \tag{A.5}$$

with canopy evapotranspiration:

$$\mathrm{ET}_c(t) = \max\left(0,\ \mathrm{ET}_p(t)\ \min\left(1,\ \frac{S_c(t)}{S_{c,\max}}\right)\right). \tag{A.6}$$

The total throughfall is:

$$T_r(t) = D(t) + p \cdot M_d(t). \tag{A.7}$$

Table A3 – Symbols for the canopy module.

Symbol	Meaning	Unit
$S_c(t)$	Canopy storage at time t	mm
$T_r(t)$	Throughfall at time t	mm/h
$\mathrm{ET}_c(t)$	Canopy evapotranspiration	mm/h
p	Free throughfall coefficient	–
$D(t)$	Drainage from canopy	mm/h
$M_d(t)$	Melting discharge	mm/h
$S_{c,\max}$	Maximum canopy storage	mm
$\mathrm{ET}_p(t)$	Potential evapotranspiration	mm/h

A.4 Root zone

The root-zone storage evolves as:

$$\frac{dS_{rz}(t)}{dt} = (1 - \alpha(t))\, T_r(t) - R_g(t) - \mathrm{ET}_{rz}(t), \tag{A.8}$$

where root-zone evapotranspiration is:

$$\mathrm{ET}_{rz}(t) = \min\left(1,\ \frac{4}{3}\frac{S_{rz}(t)}{S_{rz,\max}}\right)\ \big(\mathrm{ET}_p(t) - \mathrm{ET}_c(t)\big), \tag{A.9}$$

and recharge is parameterized as:

$$R_g(t) = g\, S_{rz}(t)^k. \tag{A.10}$$

The partitioning coefficient $\alpha(t)$ between the root zone and surface runoff follows the storage-capacity distribution approach of Moore (1985), in which the catchment is represented by a probability distribution of local soil-moisture storage capacities. The coefficient depends on the water input, on the root-zone storage $S_{rz}(t)$ and its maximum $S_{rz,\max}$, and on the parameter pB_{soil}, which controls the nonlinearity of the partitioning by acting as the exponent of a Pareto-type distribution.

Table A4 – Symbols for the root-zone module.

Symbol	Meaning	Unit
$\alpha(t)$	Partitioning coefficient between root zone and surface runoff	–
pB_{soil}	Exponent of the Pareto-type storage-capacity distribution	–
$S_{rz}(t)$	Root-zone storage	mm
$S_{rz,\max}$	Maximum root-zone storage	mm
$R_g(t)$	Recharge to groundwater	mm/h
$\mathrm{ET}_{rz}(t)$	Root-zone evapotranspiration	mm/h
k	Recharge exponent	–
g	Recharge coefficient	$\mathrm{mm}^{1-k}/\mathrm{h}$

A.5 Runoff

The runoff reservoir is:

$$\frac{dS_r(t)}{dt} = \alpha(t)\, T_r(t) - Q_r(t), \tag{A.11}$$

with discharge:

$$Q_r(t) = c\, S_r(t)^d. \tag{A.12}$$

Table A5 – Symbols for the runoff module.

Symbol	Meaning	Unit
$S_r(t)$	Runoff storage	mm
$Q_r(t)$	Runoff discharge	mm/h
d	Runoff exponent	–
c	Runoff coefficient	$\mathrm{mm}^{1-\mathrm{d}}/\mathrm{h}$
$S_{ru,\max}$	Maximum runoff storage	mm

A.6 Groundwater

The groundwater reservoir evolves as:

$$\frac{dS_g(t)}{dt} = R_g(t) - Q_g(t), \tag{A.13}$$

with outflow:

$$Q_g(t) = e\, S_g(t)^f. \tag{A.14}$$

Table A6 – Symbols for the groundwater module.

Symbol	Meaning	Unit
$S_g(t)$	Groundwater storage	mm
$Q_g(t)$	Groundwater outflow	mm/h
f	Baseflow exponent	–
e	Baseflow coefficient	$\mathrm{mm}^{1-\mathrm{f}}/\mathrm{h}$
$S_{gw,\max}$	Maximum groundwater storage	mm

APPENDIX B. GEOSTATISTICAL INTERPOLATION: FORMULATION AND SEMIVARIOGRAM MODELS

This appendix provides the mathematical background for the interpolation procedure introduced in Section 3.2.1 and implemented in the software described in Section 4.3. The main equations and their matrix formulation are reported to complement the description provided in the main text.

The method adopted is kriging, a geostatistical interpolation technique used to estimate the value of a field at unsampled locations by exploiting the spatial autocorrelation among sampled locations. If we call $z(\mathbf{x}_\alpha)$ the observed, known value at a sampled site $\mathbf{x}_\alpha$, and $z(\mathbf{x})$ the unknown value at the unsampled site $\mathbf{x}$, these observations can be regarded as realizations of the random variables $Z(\mathbf{x}_\alpha)$ and $Z(\mathbf{x})$, respectively. We can also think of kriging as an estimator that expresses the unknown value as a linear combination of the sampled ones, where the weights assigned to each sampled value depend on the spatial correlation structure. The estimator is constructed to satisfy an unbiasedness condition and to minimize the estimation variance.

As described in Section 3.2.1, the spatial dependence is represented by the semivariogram. An empirical semivariogram is first computed from the sampled observations and then fitted with a theoretical model, which is a parametric function, allowing its evaluation at any separation distance. Using this semivariogram function, the matrix and the known terms of the kriging linear system are populated. The solution of the system provides the weights needed to estimate the value at the unsampled point.

In situations where the field exhibits large-scale spatial trends (for example, an elevation-dependent gradient as in a temperature field), most of the variability can be directly explained by these dependencies. To address this, a detrended kriging approach may be used: the deterministic trend is estimated and removed from the sampled data, kriging is applied to

the residuals, and the final estimate at the unsampled location is obtained by adding the trend back to the kriged residual field.

The full formulation used in this study, directly derived from Bancheri et al. (2018) and following the same notation, is reported below.

The kriging estimator at the unsampled location $\mathbf{x}$ is defined as:

$$\widehat{Z}(\mathbf{x}) - m(\mathbf{x}) = \sum_{\alpha=1}^{N} \lambda_\alpha \left(Z(\mathbf{x}_\alpha) - m(\mathbf{x}_\alpha) \right), \tag{B.1}$$

where $m(\mathbf{x})$ and $m(\mathbf{x}_\alpha)$ are the expected values of the random variables, and λ_α are the kriging weights.

The weights are chosen to minimize the estimation variance:

$$\arg\min_{\{\lambda_\alpha\}} \sigma^2(\mathbf{x}) = \arg\min_{\{\lambda_\alpha\}} \mathrm{Var}\left[\widehat{Z}(\mathbf{x}) - Z(\mathbf{x}) \right], \tag{B.2}$$

and are subject to the unbiasedness constraint:

$$\mathbb{E}\left[\widehat{Z}(\mathbf{x}) - Z(\mathbf{x}) \right] = 0. \tag{B.3}$$

For ordinary kriging, this implies:

$$\sum_{\alpha=1}^{N} \lambda_\alpha = 1. \tag{B.4}$$

With a few algebraic steps, we obtain a linear system that can be written in compact form as:

$$\mathbf{A}\,\mathbf{u} = \mathbf{b}, \tag{B.5}$$

where, in ordinary kriging, one common representation is:

$$\begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{1} \\ \mathbf{1}^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma} \\ 1 \end{bmatrix}. \tag{B.6}$$

Here, $\boldsymbol{\Gamma}$ is the matrix containing the semivariogram values between all sampled locations, $\boldsymbol{\Lambda}$ is the vector of unknown weights, $\boldsymbol{\gamma}$ contains the semivariogram values between the unsampled point and each sampled point, $\mathbf{1}$ is an N-vector of ones, and μ is the Lagrange multiplier enforcing the unbiasedness constraint. Solving the system provides the kriging weights and, consequently, the estimate at the unsampled location.

The theoretical semivariograms implemented are the linear, exponential, power, spherical, and Gaussian models. Their mathematical formulations are reported below.

Linear semivariogram

$$\gamma(h) = \begin{cases} c_0 + b\,h, & h \leq a, \\ c_0 + c, & h > a. \end{cases} \tag{B.7}$$

Exponential semivariogram

$$\gamma(h) = c_0 + c\left(1 - \exp\left(-\frac{h}{a}\right)\right). \tag{B.8}$$

Power semivariogram

$$\gamma(h) = c_0 + b\,h^p. \tag{B.9}$$

Spherical semivariogram

$$\gamma(h) = \begin{cases} c_0 + c\left[\frac{3}{2}\left(\frac{h}{a}\right) - \frac{1}{2}\left(\frac{h}{a}\right)^3\right], & h \leq a, \\ c_0 + c, & h > a. \end{cases} \tag{B.10}$$

Gaussian semivariogram

$$\gamma(h) = c_0 + c\left(1 - \exp\left(-\left(\frac{h}{a}\right)^2\right)\right). \tag{B.11}$$

In all models, h denotes the separation distance, $\gamma(h)$ the semivariogram value at distance

h, c_0 the nugget, c the partial sill, a the range parameter, b a scaling coefficient, and p the exponent of the power model.

APPENDIX C. PHYSIOGRAPHIC AND CLIMATIC CHARACTERIZATION OF THE NOCE BASIN

This appendix complements the physical description of the study area provided in Chapter 2 with additional information on the basin and its monitoring network. It reports the main characteristics of three monitoring stations representative of different elevation ranges within the basin, together with a description of the geological setting and the dominant tree species. These elements provide additional context for the physical characterization of the study area and the interpretation of the hydrological processes discussed throughout the thesis.

C.1 Climate and data

Table C1 – Summary of temperature statistics (°C) for different stations in the basin.

	Temperature [°C]				
Weather station	**Annual**			**Monthly**	
	Min	**Mean**	**Max**	**Min**	**Max**
Presena	−25.2	−1.4	20.0	−9.6	8.9
Cles	−12.9	8.5	39.5	−1.1	23.5
Mezzolombardo	−13.7	12.6	40.6	1.2	24.8

Values are based on data from 2012 to 2024. The stations are located at various altitudes and are representative of different climate types. The first station, Presena, is near the eponymous glacier at 2715 m asl. The second station is located in the middle of the basin, in the village of Cles, at an elevation of 665 m asl. The last station is situated in Mezzolombardo, near the basin's outlet, on the valley floor at 204 m asl.

C.2 Physiographic and environmental setting of the Noce basin

Geologically, in the upper part of the basin, bedrock is dominated by intrusive rocks of the Adamello batholith (mainly tonalites and granodiorites with subordinate mafic bodies) in the south-west (Passo Tonale and Presena), and by metamorphic rocks of the Ortles–

Cevedale unit (paragneiss, orthogneiss and mica schists) in the northern part of the basin (Val di Peio, Val di Rabbi). Along the eastern margin of the middle basin, the Brenta Dolomites form a prominent carbonate massif composed predominantly of Mesozoic dolomitic rocks. Downstream, the Noce incises a glacially overdeepened Alpine valley cut into metamorphic basement and Mesozoic carbonate formations, locally mantled by thick Quaternary glacial and glaciofluvial deposits that form wide valley floors and terraces, particularly in Val di Sole and Val di Non.

Soils reflect this strong geomorphological and lithological gradient. On the steep slopes of the upper and middle parts of the basin, soils are generally shallow, coarse-textured and well drained, developed on crystalline and dolomitic bedrock, with high skeletal content and limited profile development (Aberegg et al., 2009; Egli et al., 2004; Favilli et al., 2010; Sartori, 2009). In contrast, valley bottoms, glaciofluvial terraces and the alluvial plain host deeper, more evolved, well-drained glaciofluvial and alluvial soils with loamy to sandy-loam textures and high gravel content, which sustain intensive agriculture, especially apple orchards in Val di Non and vineyards in Rotaliana (Mezzolombardo) (Bassetti & Borsato, 2007; Boaga et al., 2013; GuildSomm, 2016).

Vegetation and land cover follow the altitudinal and climatic gradient from the glacier region to the intensively cultivated valley bottoms. In the upper basin (above 1800 m asl), subalpine conifer forests and alpine grasslands dominate, with larch (*Larix decidua*), Swiss stone pine (*Pinus cembra*) and mountain pine (*Pinus mugo*); near the treeline, forests become more open, with sparse larch and pine stands interspersed with meadows and scree (Favilli et al., 2010; ISPRA, 2013). At middle elevations (1000 m–1800 m asl), continuous coniferous forests dominated by Norway spruce (*Picea abies*), with Scots pine (*Pinus sylvestris*), silver fir (*Abies alba*) and local beech (*Fagus sylvatica*), cover most valley sides (Hilmers et al., 2019; Nocentini, 2009). Along the valley floor, riparian belts of alder (*Alnus glutinosa*, *A. incana*), willow (*Salix* spp.) and poplar (*Populus* spp.) fringe the Noce and its tributaries, forming typical Alpine riparian forest communities. In the mid- and lower basin, natural

forests are progressively replaced, especially in the wider valley bottoms of Val di Sole, Val di Non and the Piana Rotaliana, by a cultural mosaic of permanent hay meadows, intensively managed apple orchards and vineyards, together with remnant patches of broadleaf forest (Assandri, 2015).

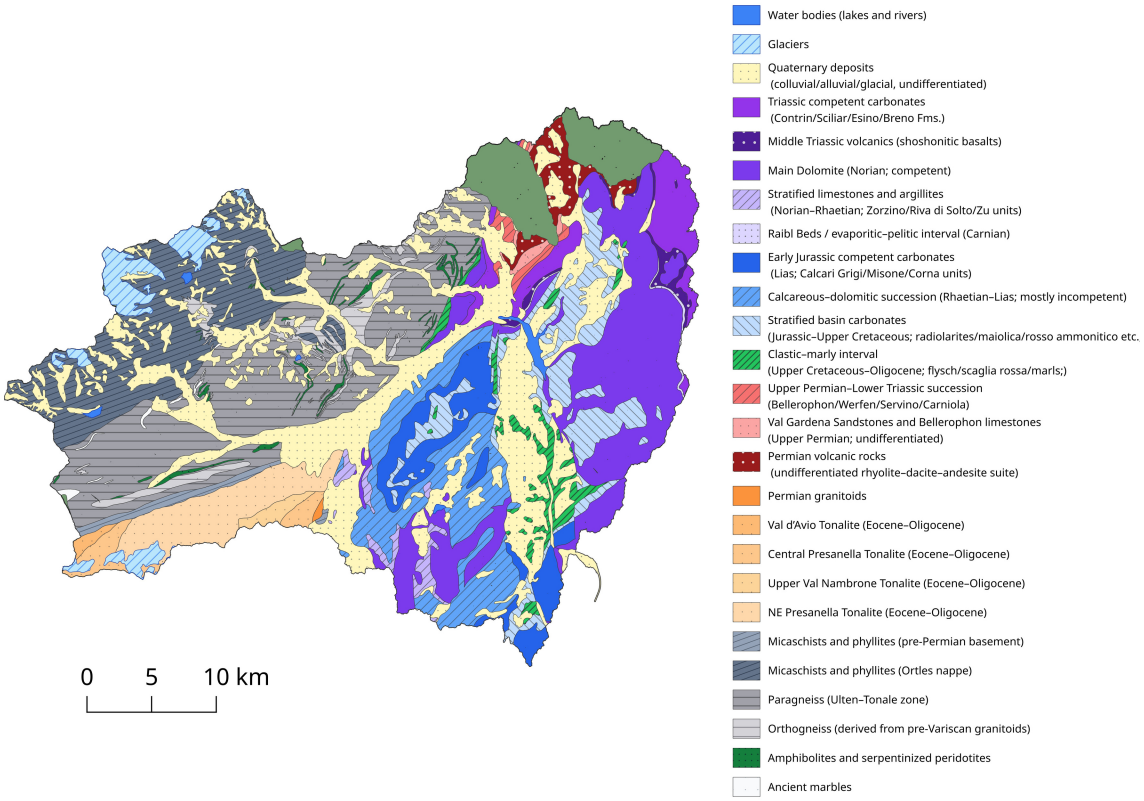

Figure C1 – Lithological map of the Noce basin (extracted from Trentino map), adapted and color-modified. Originally created by A. Bosellini, A. Castellarin, G.V. Dal Piaz, and M. Nardin, the map was derived from a critical review of geological mapping from the mid-1980s. Based on the data available at that time, a lithological and structural map was produced to serve as a foundational tool for subsequent detailed studies. The map was initially drafted at a 1:100,000 scale and later digitized and printed at a 1:200,000 scale in the first quarter of 1999. The legend is based on major lithological divisions, distinguishing stratigraphic intervals based on lithological uniformity and their deformational behavior (Bosellini et al., 2023).

Table C2 – Summary of the spatio-temporal resolution of the datasets used in this thesis.

Method	**Variable**	**Spatial Resolution**	**Temporal Resolution**	**Depth**
CRNS	Soil Moisture	$\sim$100 m	Hourly	$\sim$20 cm
Copernicus	Soil Moisture	$\sim$1 km	$\sim$Weekly	$\sim$5 cm
MODIS	Snow (presence/absence)	500 m	Daily	–
MERIDA	Soil Moisture	$\sim$7 km	Hourly	4 layers (5, 25, 70, 150 cm)

Further details on the datasets are discussed in the text.

APPENDIX D. SUPPLEMENTARY TABLES AND FIGURES

D.1 Study 1: model schematic and mapped outputs

This section provides supplementary material supporting the analysis presented in Chapter 5.

D.1.1 Model output

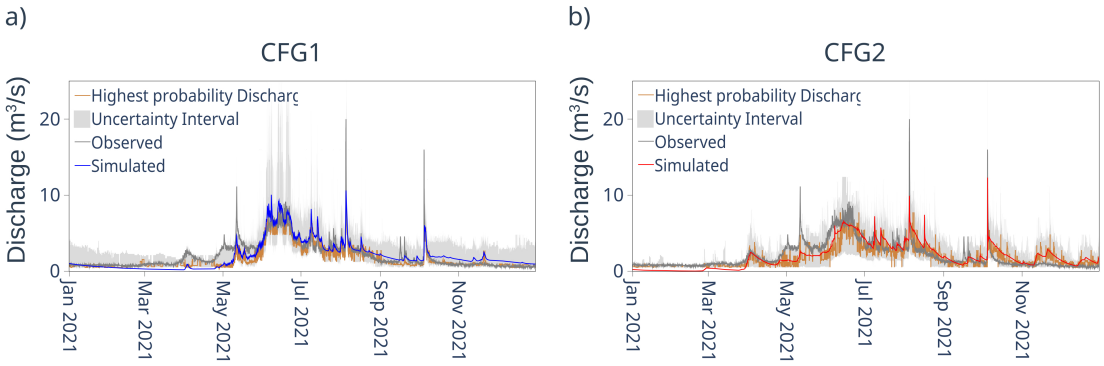

Figure D1 – Uncertainty analysis for discharge at Vermiglio. (a) CFG1, with observed data shown in light gray and modeled data in blue; (b) CFG2, with observed data in light gray and modeled data in red. Both plots include the uncertainty interval in gray.

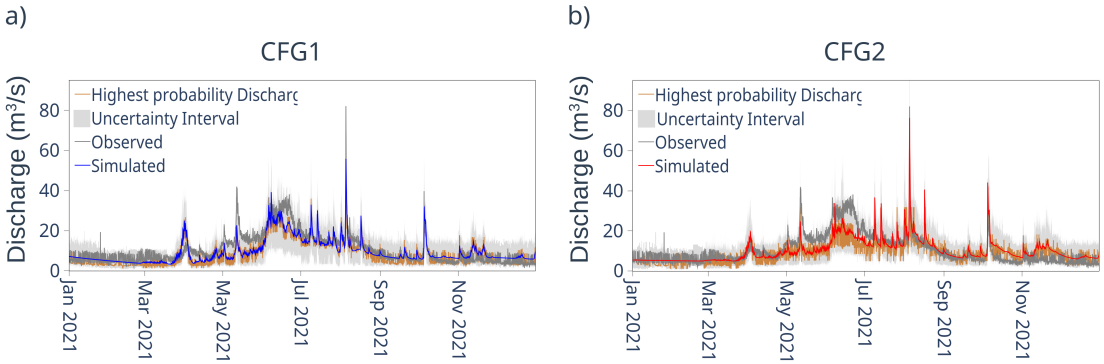

Figure D2 – Uncertainty analysis for discharge at Malé. (a) CFG1, with observed data shown in light gray and modeled data in blue; (b) CFG2, with observed data in light gray and modeled data in red. Both plots include the uncertainty interval in gray.

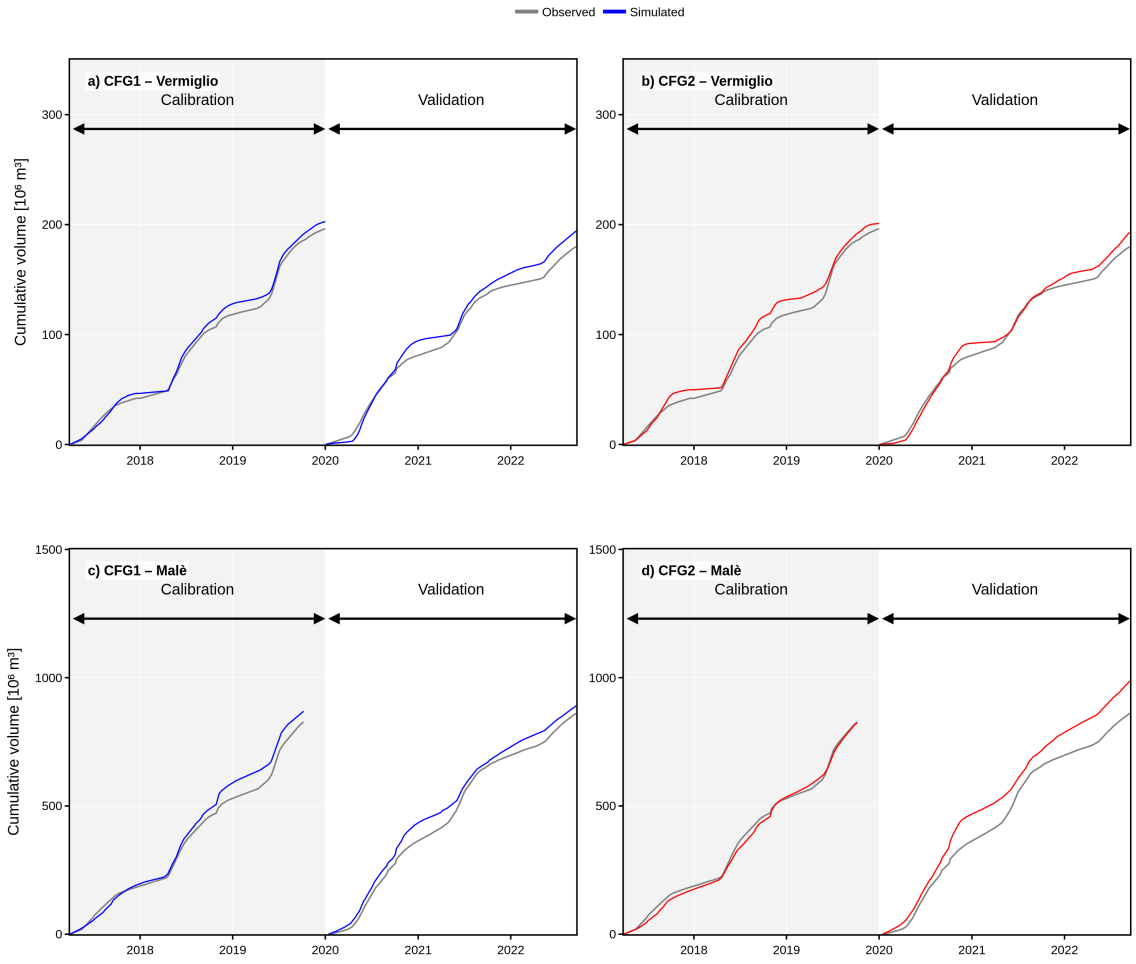

Figure D3 – Observed and simulated cumulative discharge during the calibration and validation periods. Panels (a) and (b) show the Vermiglio stream gauge for CFG1 and CFG2, respectively, whereas panels (c) and (d) show the Malé stream gauge for CFG1 and CFG2, respectively. The observed cumulative discharge is shown in gray, while the simulated cumulative discharge is shown in blue for CFG1 and in red for CFG2.

Table D1 – Observed and simulated discharge volumes aggregated by hydrological year. Only complete hydrological years are reported.

Station	CFG	Hydrological year	V_{obs} ($10^6\,\mathrm{m}^3$)	V_{sim} ($10^6\,\mathrm{m}^3$)	Error (%)
Vermiglio	CFG1	2018	68.42	73.18	6.95
	CFG1	2019	80.62	79.45	-1.45
	CFG1	2020	76.23	80.13	5.11
	CFG1	2021	72.12	75.88	5.22
	CFG2	2018	68.42	71.88	5.05
	CFG2	2019	80.62	74.77	-7.26
	CFG2	2020	76.23	77.31	1.41
	CFG2	2021	72.12	70.61	-2.10
Malé	CFG1	2018	304.03	343.48	12.98
	CFG1	2019	364.28	376.33	3.31
	CFG1	2021	380.35	363.76	-4.36
	CFG2	2018	304.03	305.48	0.48
	CFG2	2019	364.28	378.24	3.83
	CFG2	2021	380.35	379.48	-0.23

Years 2017 and 2022 are omitted because they do not cover complete hydrological years. The 2020 hydrological year at the Malé stream gauge is also excluded due to incomplete data coverage.

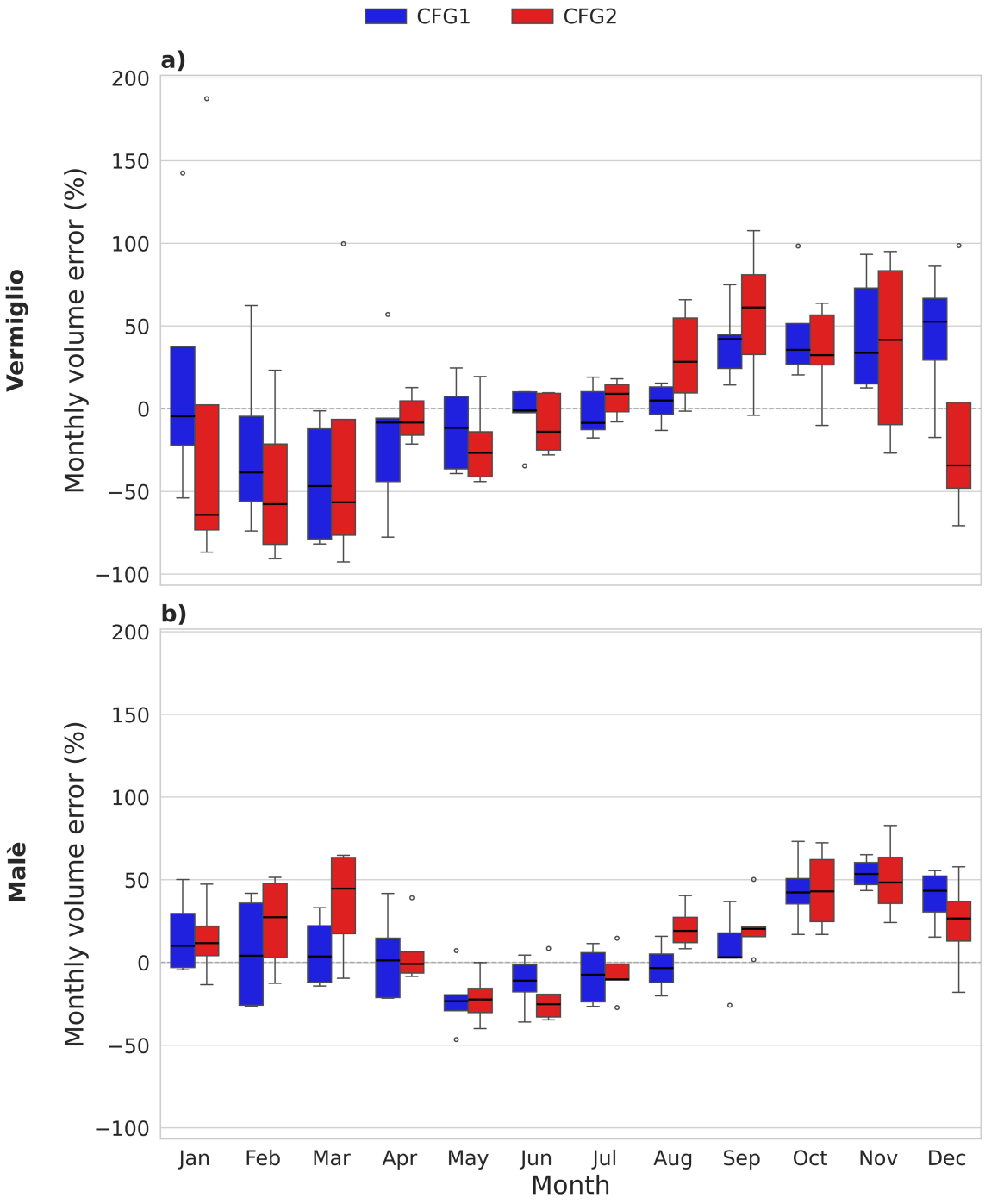

Figure D4 – Box-plots of monthly discharge-volume percentage errors computed from aggregated monthly discharge volumes. CFG1 is shown in blue and CFG2 in red. Panel (a) refers to the Vermiglio stream gauge, whereas panel (b) refers to the Malé stream gauge.

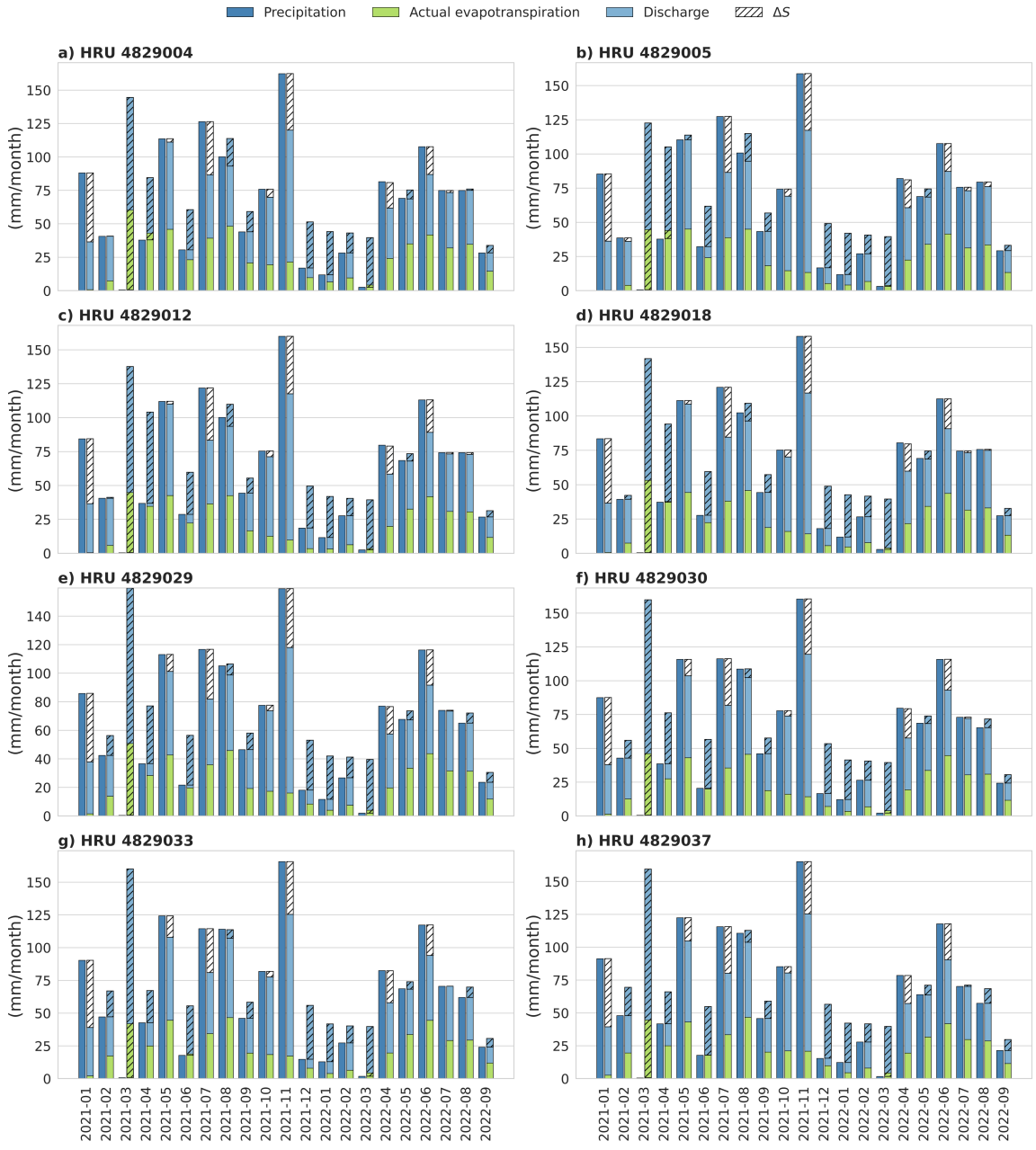

Figure D5 – Example of the water balance for the GUs in CFG2, derived from GU 4829 in CFG1. For each time step, precipitation input is shown as a dark-blue bar, while the corresponding water balance components are represented by a stacked bar including actual evapotranspiration (green), outflow (light blue), and the change in total water storage (transparent with diagonal hatching). Positive storage changes are stacked above the other components, whereas negative storage changes overlap them. Panels (a)–(h) show the individual GUs.

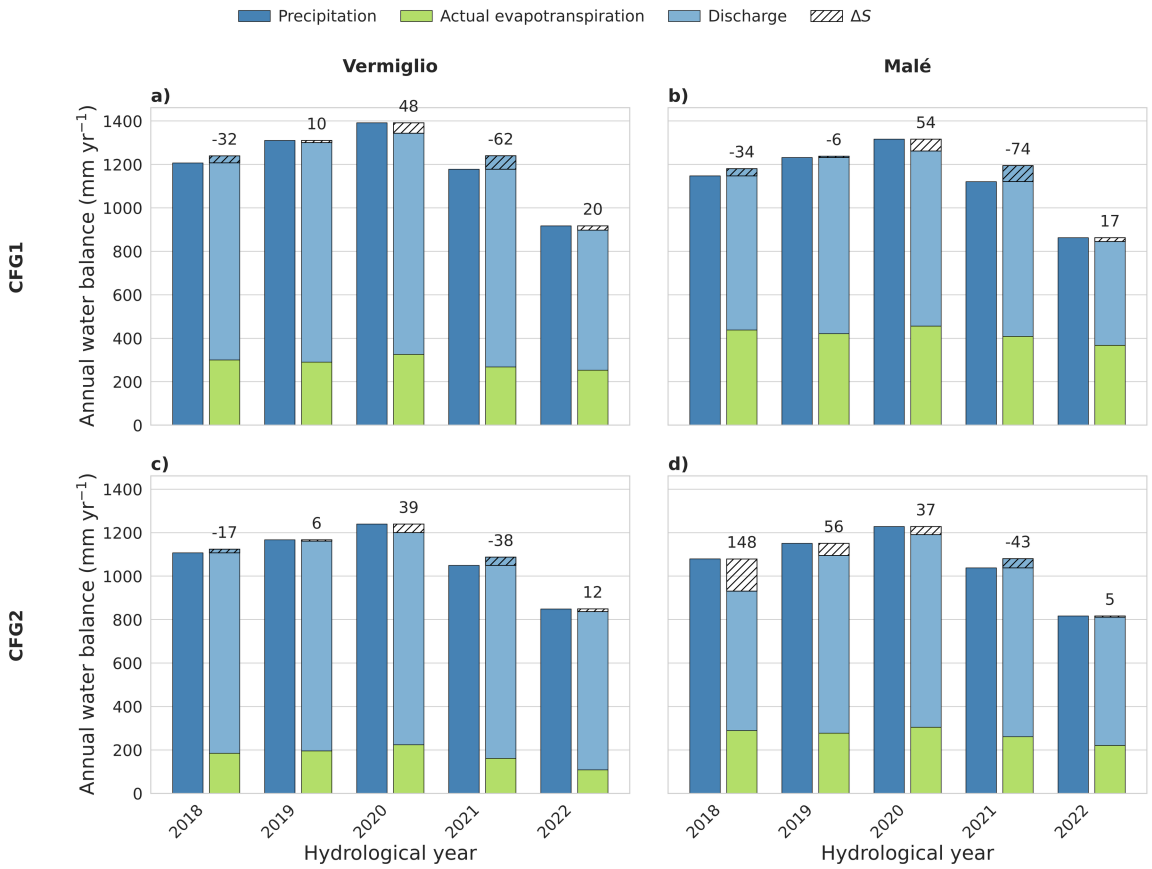

Figure D6 – Annual water balance for the entire Noce basin, upstream of Malé, reported for each hydrological year and obtained by summing the water balance components of all GUs. For each hydrological year, precipitation input is shown as a dark-blue bar, while the corresponding water balance components are represented by a stacked bar including actual evapotranspiration (green), outflow (light blue), and the change in total water storage (transparent with diagonal hatching). Positive storage changes are stacked above the other components, whereas negative storage changes overlap them. The value of the annual storage change is reported above each bar.

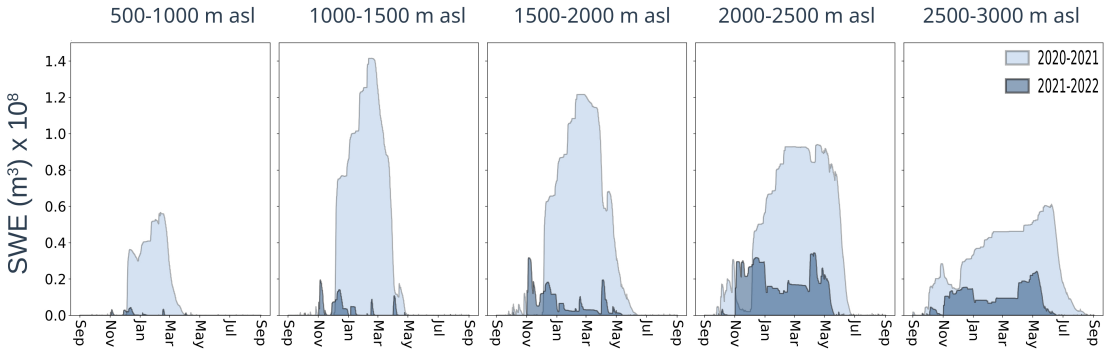

Figure D7 – SWE in 2021 compared with 2022, CFG1, in m^3, for the whole basin.

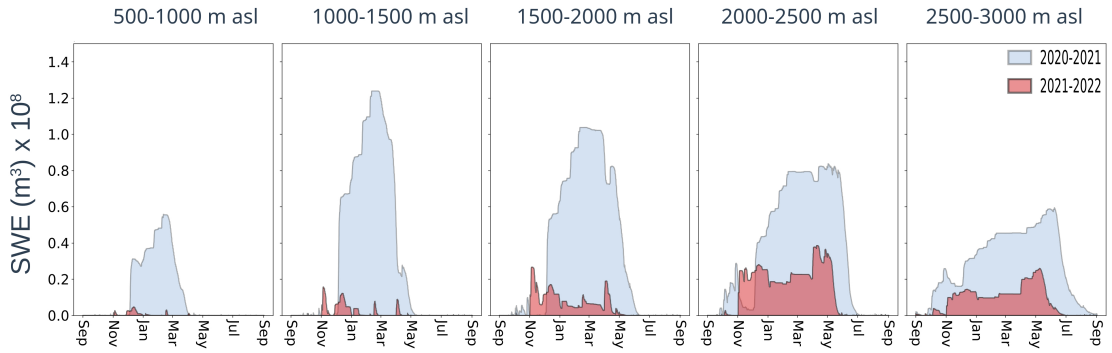

Figure D8 – SWE in 2021 compared with 2022, CFG2, in m^3, for the whole basin.

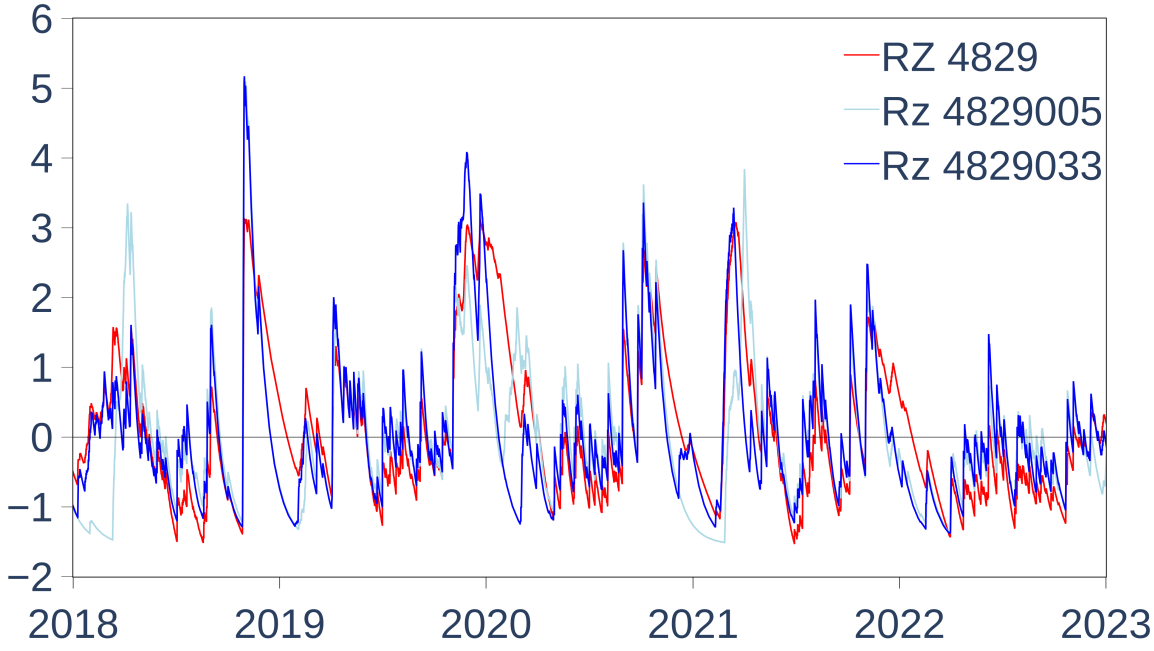

Figure D9 – Anomaly in water content in the root zone, comparing 2021 with 2022. The red line represents the value for GU 4829, while the blue lines show two GUs derived from GU 4829 in CFG2: 4829005 and 4829033. The anomaly is calculated relative to the mean value of the study period and normalized by the standard deviation.

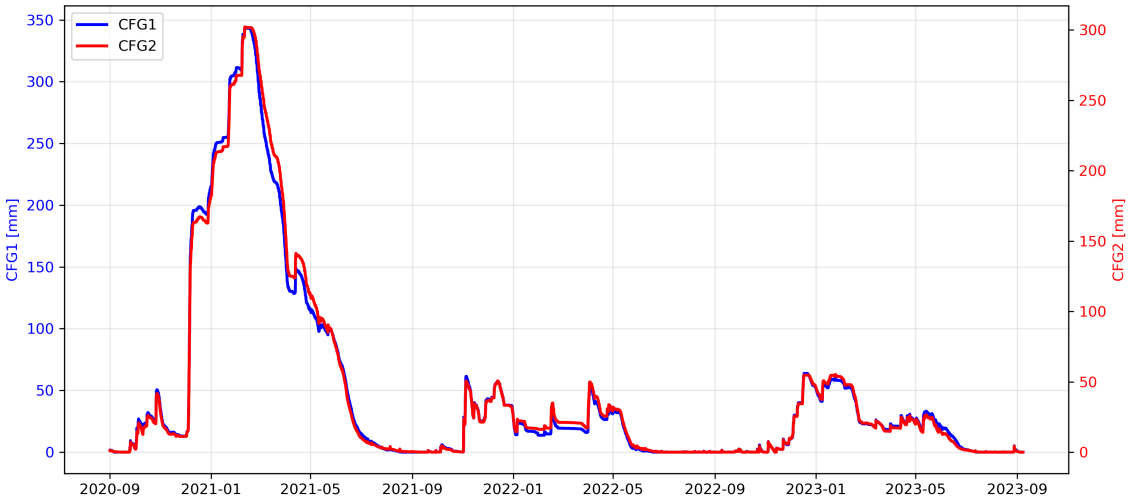

Figure D10 – Basin-mean snow water equivalent (SWE; mm) time series, averaged over all Geomorphic Units (GUs). Blue: CFG1; red: CFG2.

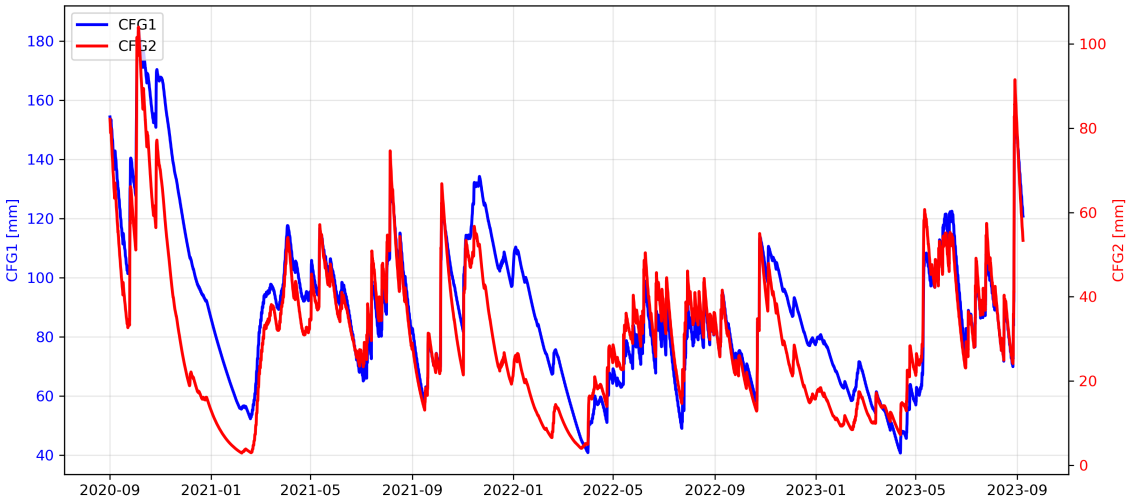

Figure D11 – Basin-mean root-zone water content (mm) time series, averaged over all Geomorphic Units (GUs). Blue: CFG1; red: CFG2.

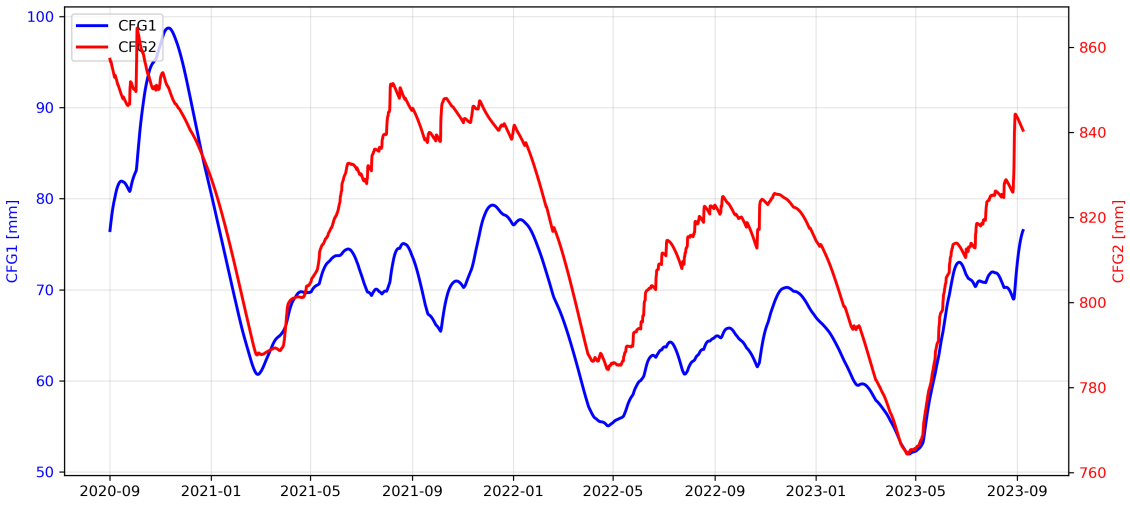

Figure D12 – Basin-mean groundwater storage (mm) time series, averaged over all Geomorphic Units (GUs). Blue: CFG1; red: CFG2.

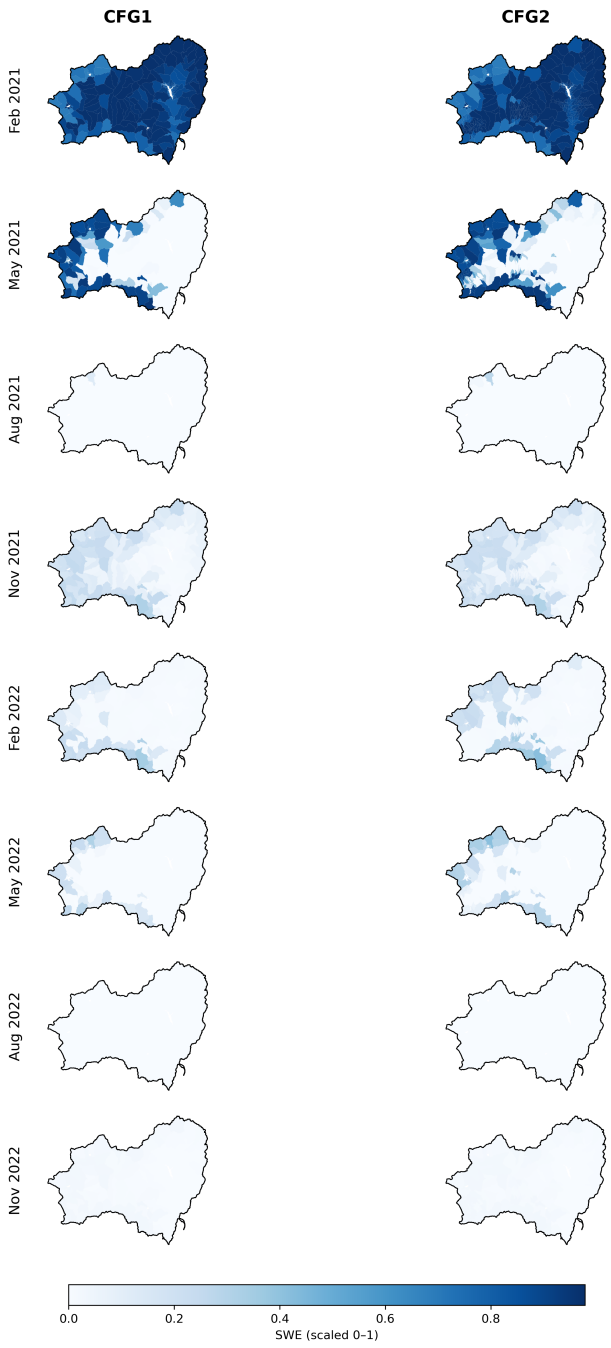

Figure D13 – Spatial distribution of simulated snow water equivalent (SWE) for representative winter and summer dates (panels), comparing CFG1 and CFG2. Values are normalized to the $[0, 1]$ range using min–max scaling within each map (0 = basin minimum, 1 = basin maximum) to emphasize spatial patterns.

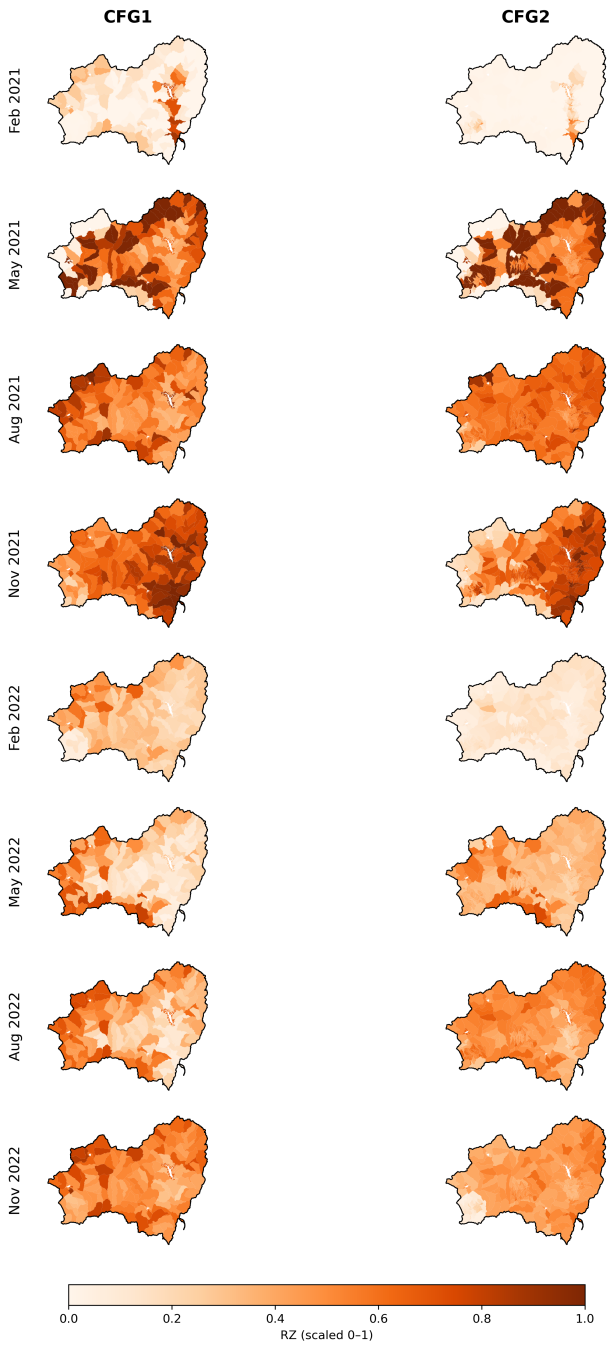

Figure D14 – Spatial distribution of simulated root-zone water storage for representative winter and summer dates (panels), comparing CFG1 and CFG2. Values are normalized to the $[0, 1]$ range using min–max scaling within each map (0 = basin minimum, 1 = basin maximum) to emphasize spatial patterns.

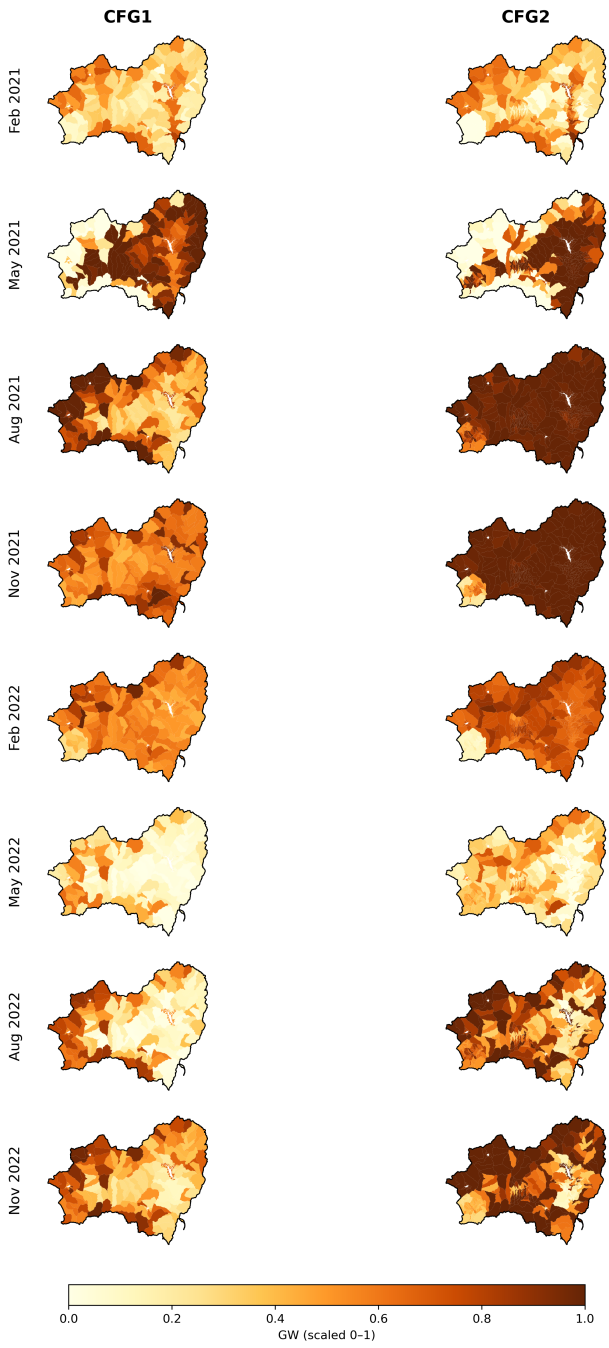

Figure D15 – Spatial distribution of simulated groundwater storage for representative winter and summer dates (panels), comparing CFG1 and CFG2. Values are normalized to the $[0, 1]$ range using min–max scaling within each map (0 = basin minimum, 1 = basin maximum) to emphasize spatial patterns.

D.2 Study 2: configurations, semivariogram diagnostics and rainfall-runoff metrics

This section provides supplementary material supporting the analysis presented in Chapter 6.

Table D2 – Characteristics of the three model configurations. The table reports the number of Geomorphic Units (GUs), together with the characteristic GU area and elevation. For the SG configuration, which consists of only two GUs, the values of both GUs are reported. For GU20 and GU10, the median GU area and median GU elevation are reported.

Configuration	No. of GUs	Characteristic GU area [km^2]	Characteristic GU elevation [m asl]
SG	2	79 / 384	987 / 2718
GU20	30	18	2162
GU10	51	8.9	2121

D.2.1 Semivariogram and LOOCV

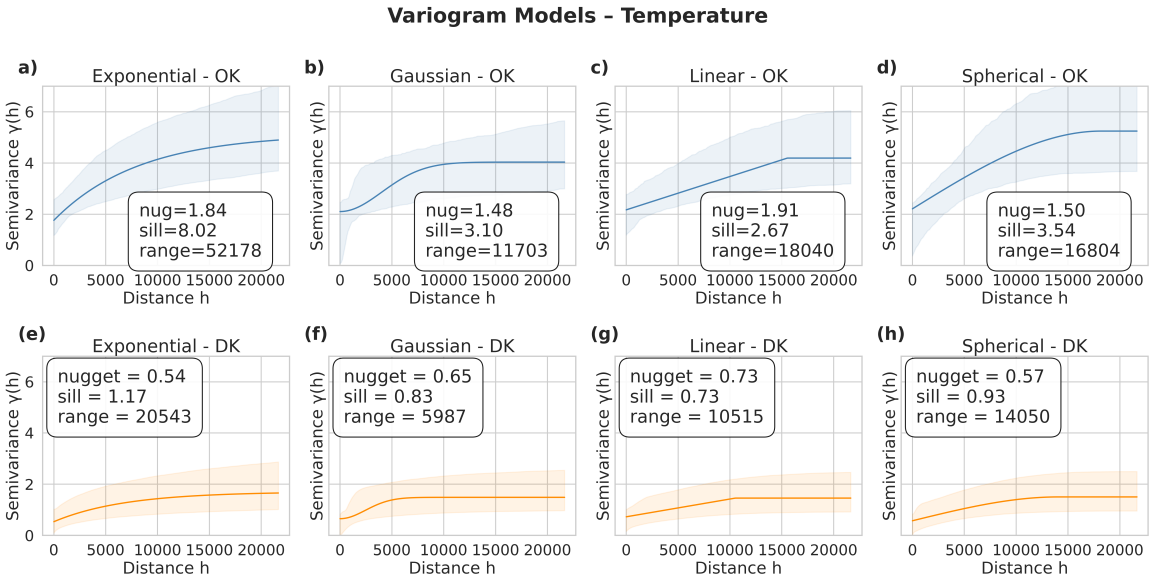

Figure D16 – Theoretical variogram models for temperature. Panels (a–d) correspond to the OK (Ordinary Kriging) case, while panels (e–h) correspond to the DK (Detrended Kriging) case. Each panel shows a different theoretical model (Exponential, Gaussian, Linear, and Spherical). The central black curve represents the medoid, i.e., the fitted variogram curve that is closest to the pointwise median of the ensemble. Shaded areas indicate the 25th–75th percentile band of the fitted variograms, representing the interquartile range of the ensemble variability. Text boxes report the nugget, sill, and range values of the medoid for each configuration.

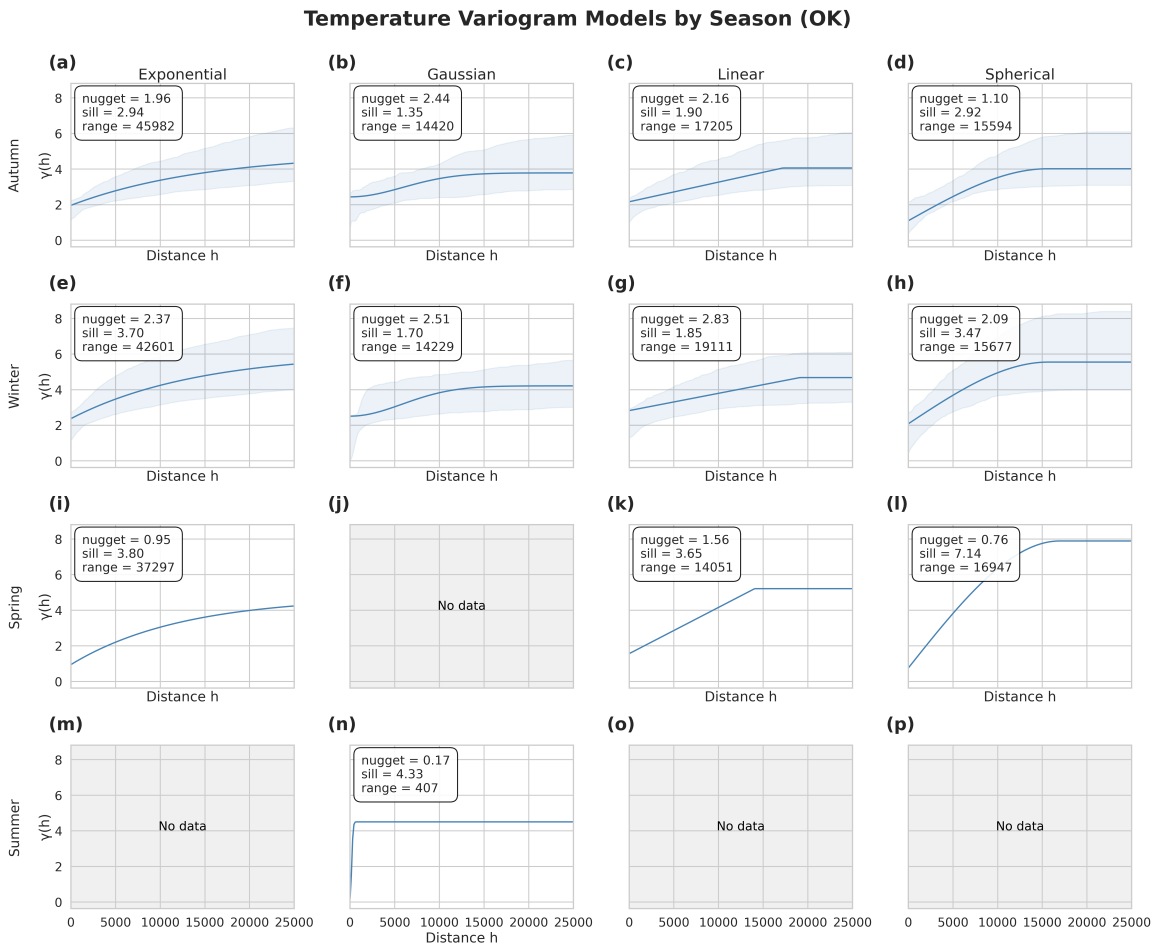

Figure D17 – Theoretical variogram models for precipitation. Panels (a–d) correspond to the OK (Ordinary Kriging) case, while panels (e–h) correspond to the DK (Detrended Kriging) case. Each panel shows a different theoretical model (Exponential, Gaussian, Linear, and Spherical). The central black curve represents the medoid, i.e., the fitted variogram curve that is closest to the pointwise median of the ensemble. Shaded areas indicate the 25th–75th percentile band of the fitted variograms, representing the interquartile range of the ensemble variability. Text boxes report the nugget, sill, and range values of the medoid for each configuration.

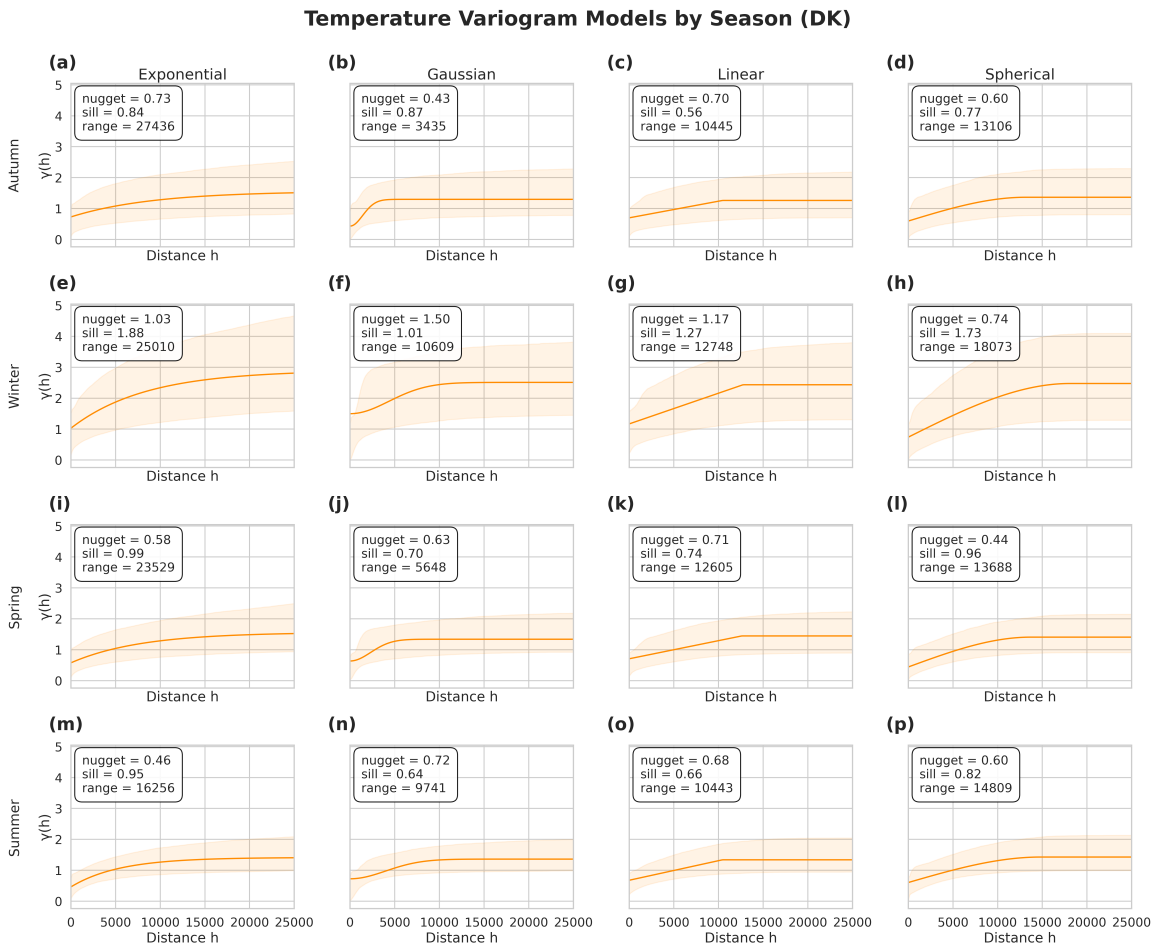

Figure D18 – Theoretical variogram models for temperature by season for OK. Panels (a–d), (e–h), (i–l), and (m–p) correspond to autumn, winter, spring, and summer, respectively. Columns correspond to the different theoretical models (Exponential, Gaussian, Linear, and Spherical). In each panel, the central black curve represents the medoid, i.e., the fitted variogram curve closest to the pointwise median of the ensemble. Shaded areas indicate the 25th–75th percentile band of the fitted variograms, representing the interquartile range of ensemble variability. Text boxes report the nugget, sill, and range values of the medoid for each configuration.

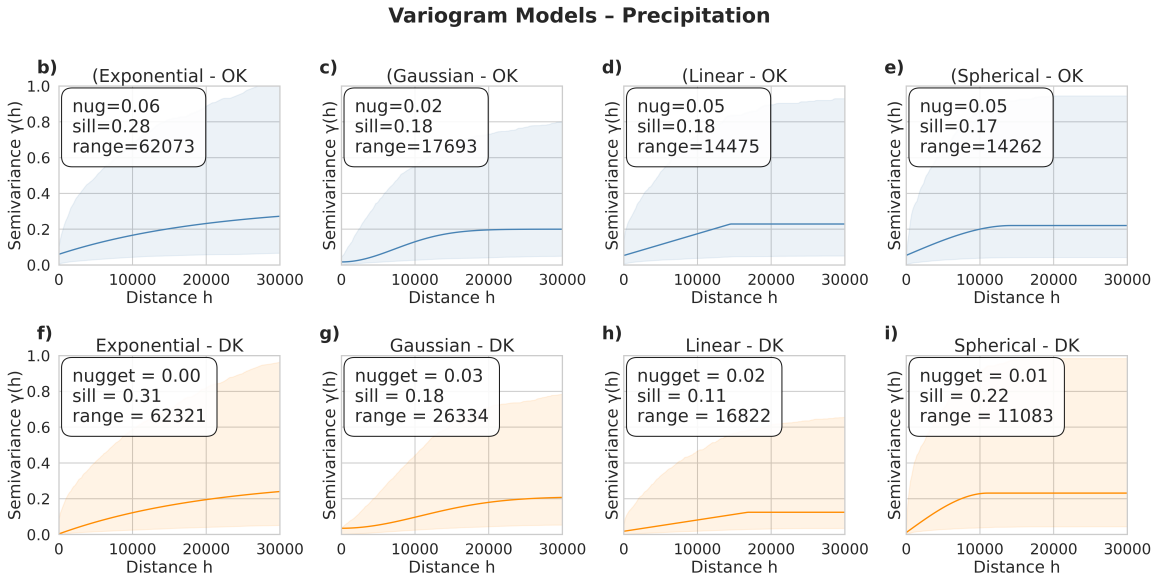

Figure D19 – Theoretical variogram models for temperature by season for DK. Panels (a–d), (e–h), (i–l), and (m–p) correspond to autumn, winter, spring, and summer, respectively. Columns correspond to the different theoretical models (Exponential, Gaussian, Linear, and Spherical). In each panel, the central black curve represents the medoid, i.e., the fitted variogram curve closest to the pointwise median of the ensemble. Shaded areas indicate the 25th–75th percentile band of the fitted variograms, representing the interquartile range of ensemble variability. Text boxes report the nugget, sill, and range values of the medoid for each configuration.

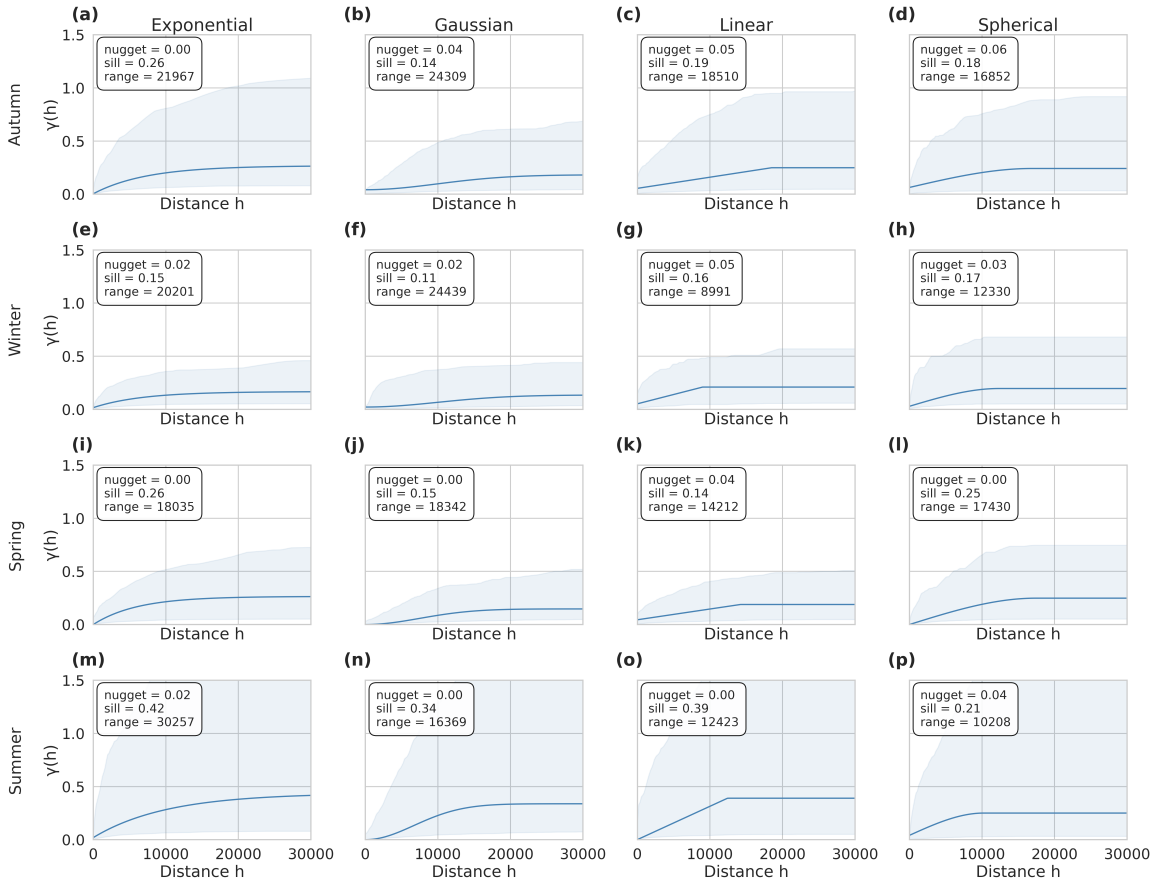

Figure D20 – Theoretical variogram models for precipitation by season for OK. Panels (a–d), (e–h), (i–l), and (m–p) correspond to autumn, winter, spring, and summer, respectively. Columns correspond to the different theoretical models (Exponential, Gaussian, Linear, and Spherical). In each panel, the central black curve represents the medoid, i.e., the fitted variogram curve closest to the pointwise median of the ensemble. Shaded areas indicate the 25th–75th percentile band of the fitted variograms, representing the interquartile range of ensemble variability. Text boxes report the nugget, sill, and range values of the medoid for each configuration.

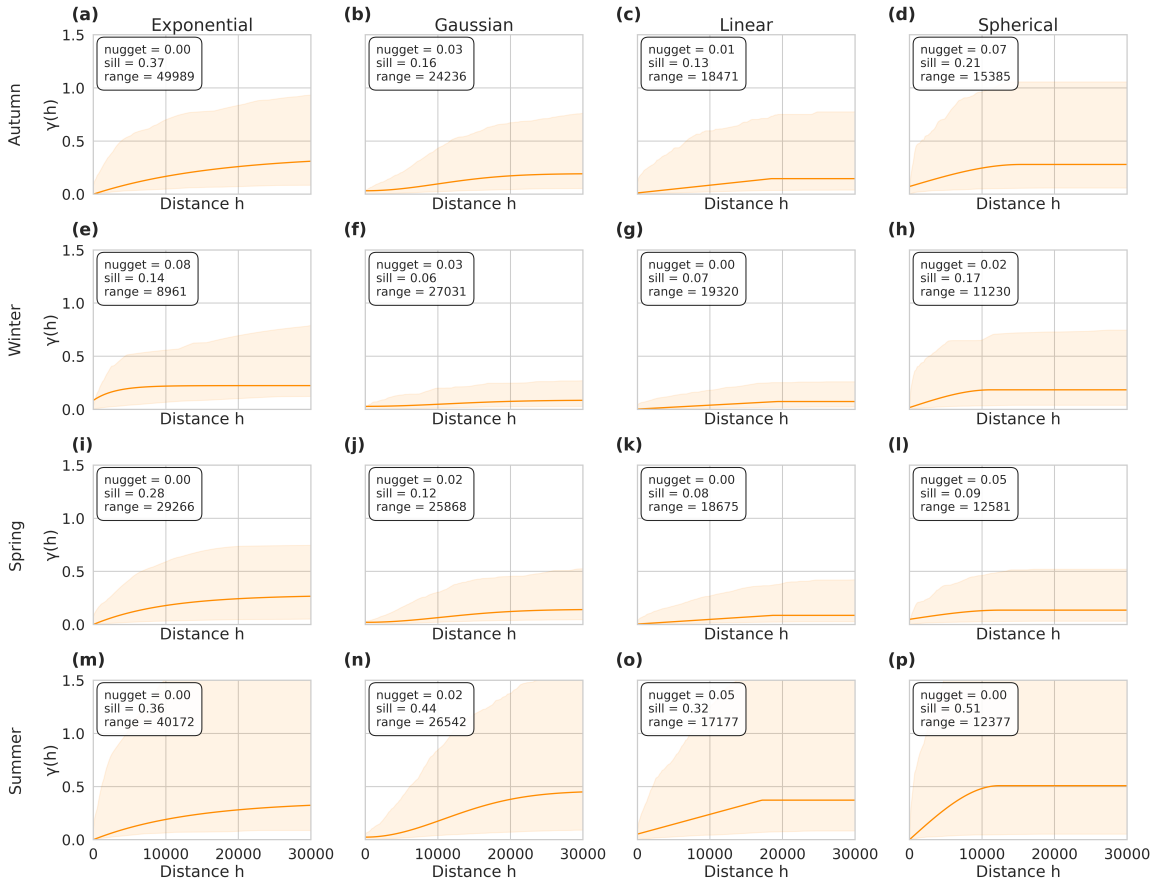

Figure D21 – Theoretical variogram models for precipitation by season for DK. Panels (a–d), (e–h), (i–l), and (m–p) correspond to autumn, winter, spring, and summer, respectively. Columns correspond to the different theoretical models (Exponential, Gaussian, Linear, and Spherical). In each panel, the central black curve represents the medoid, i.e., the fitted variogram curve closest to the pointwise median of the ensemble. Shaded areas indicate the 25th–75th percentile band of the fitted variograms, representing the interquartile range of ensemble variability. Text boxes report the nugget, sill, and range values of the medoid for each configuration.

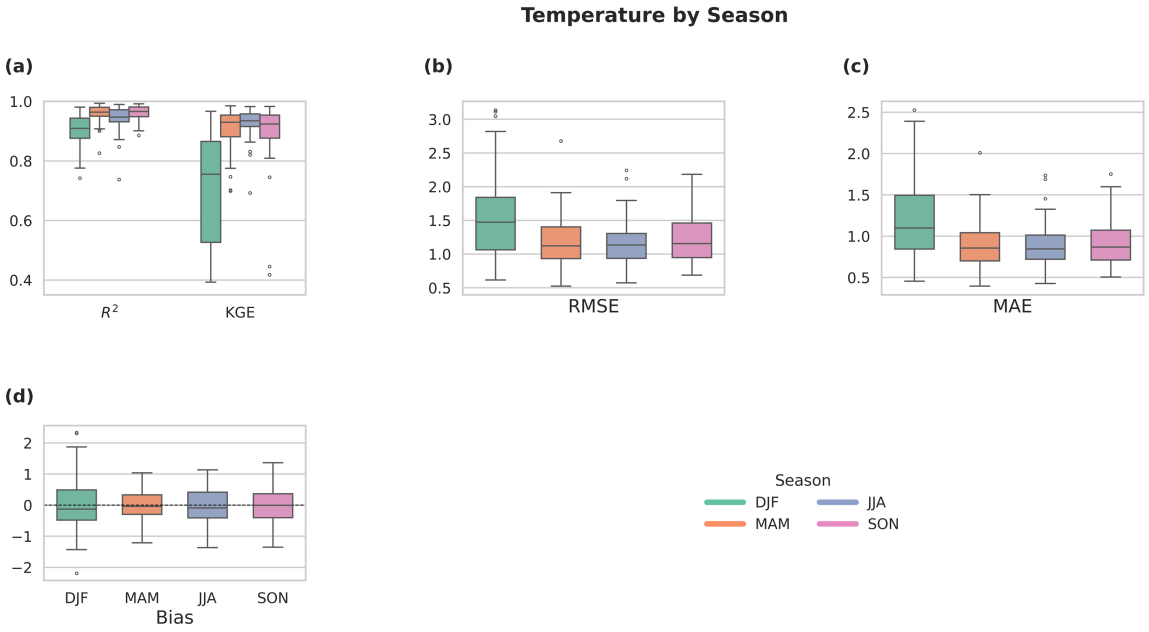

Figure D22 – Box-plots of model performance metrics (R^2, KGE, RMSE, MAE, and Bias) for temperature, obtained through leave-one-out cross-validation (LOOCV). Each box represents the distribution of station-specific metrics within a given season: DJF (December–January–February), MAM (March–April–May), JJA (June–July–August), and SON (September–October–November). The results illustrate seasonal variations in model performance across all meteorological stations.

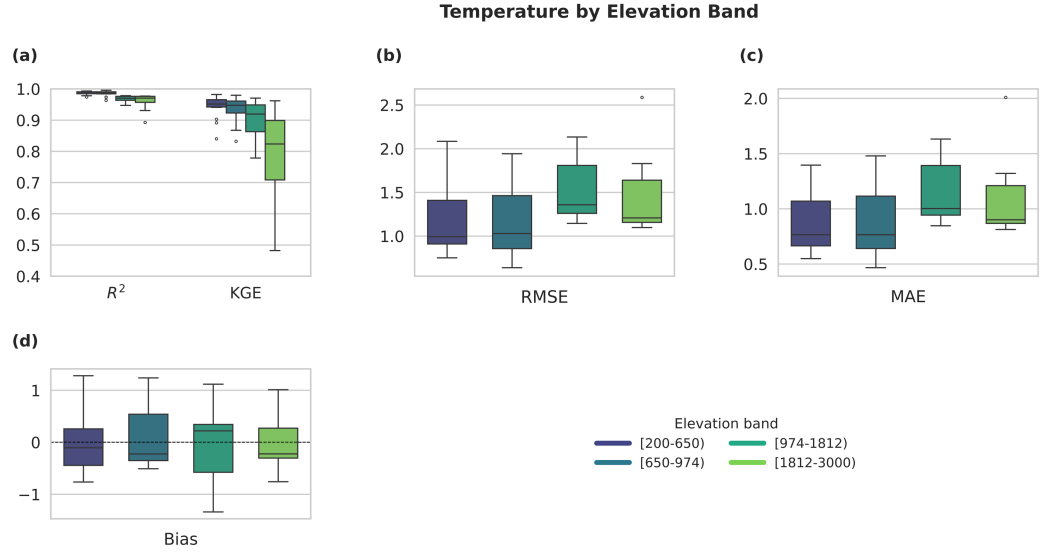

Figure D23 – Box-plots of model performance metrics (R^2, KGE, RMSE, MAE, and Bias) for temperature, obtained through leave-one-out cross-validation (LOOCV) and aggregated by elevation bands. Each box represents the distribution of station-specific metrics within an elevation class, defined using equal-frequency binning based on station altitude. The comparison highlights how model performance varies with elevation across all stations.

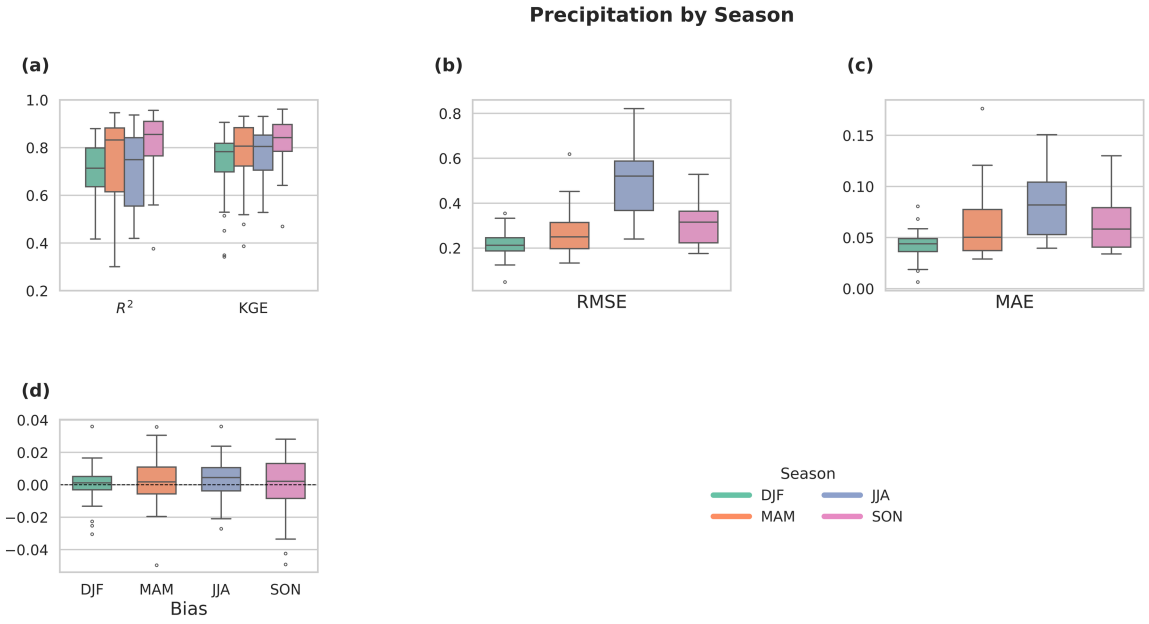

Figure D24 – Box-plots of model performance metrics (R^2, KGE, RMSE, MAE, and Bias) for precipitation, obtained through leave-one-out cross-validation (LOOCV). Each box represents the distribution of station-specific metrics within a given season: DJF (December–January–February), MAM (March–April–May), JJA (June–July–August), and SON (September–October–November). The results illustrate seasonal variations in model performance across all meteorological stations.

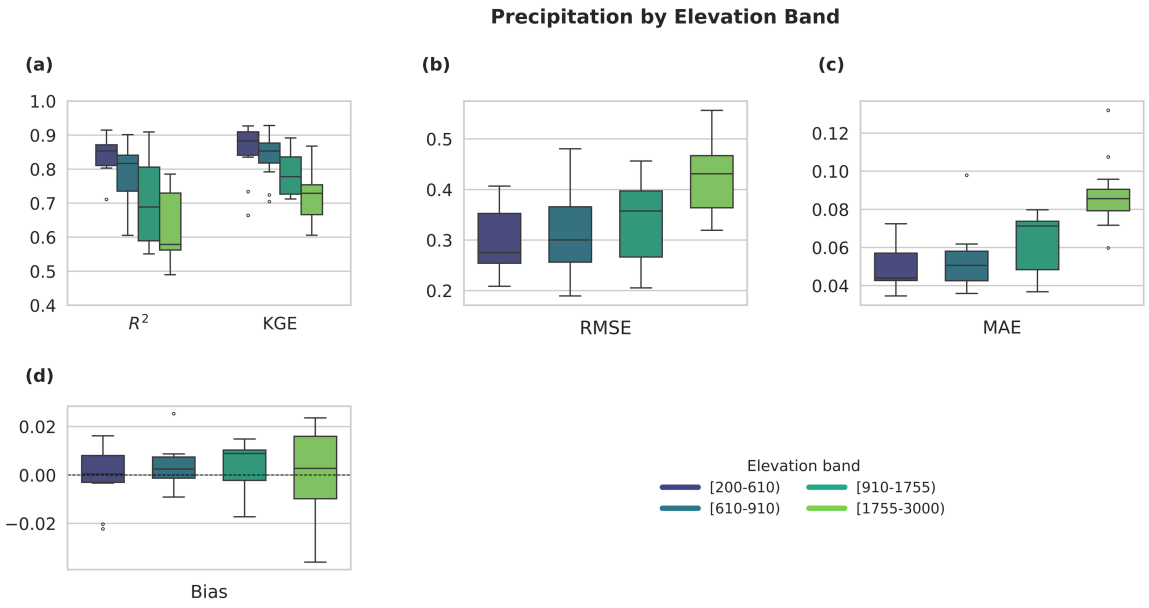

Figure D25 – Box-plots of model performance metrics (R^2, KGE, RMSE, MAE, and Bias) for precipitation, obtained through leave-one-out cross-validation (LOOCV) and aggregated by elevation bands. Each box represents the distribution of station-specific metrics within an elevation class, defined using equal-frequency binning based on station altitude. The comparison highlights how model performance varies with elevation across all stations.

D.2.2 Rainfall-runoff

Table D3 – Error-based metrics (RMSE, PBIAS) of the rainfall-runoff model, relative to measured versus modeled discharge, at the Malé stream gauge. Values are reported separately for the calibration and validation periods.

	Calibration (2017-01-01–2019-10-30)		Validation (2020-07-01–2023-12-31)	
Configuration	RMSE [m³/s]	PBIAS [%]	RMSE [m³/s]	PBIAS [%]
SG-NEAR	6.00	7.46	6.18	13.44
SG-CEN	6.07	5.38	5.47	6.67
SG-GRIDW	5.07	-1.70	6.36	2.54
GU20-NEAR	5.32	4.77	4.44	10.58
GU20-CEN	4.97	4.92	4.25	8.41
GU20-GRIDW	4.03	-0.79	4.06	1.62
GU10-NEAR	5.59	5.99	4.58	12.74
GU10-CEN	5.33	-0.17	4.89	3.12
GU10-GRIDW	3.74	-1.56	4.28	1.99

Table D4 – Error-based metrics (RMSE, PBIAS) of the rainfall-runoff model, relative to measured versus modeled discharge, at the Vermiglio stream gauge. Metrics are computed over the independent validation period (2017-01-01–2023-12-31).

Configuration	RMSE [m³/s]	PBIAS [%]
SG-NEAR	2.01	15.65
SG-CEN	1.58	5.97
SG-GRIDW	1.73	2.44
GU20-NEAR	1.52	5.87
GU20-CEN	1.84	5.91
GU20-GRIDW	1.65	0.91
GU10-NEAR	1.45	2.51
GU10-CEN	1.44	-2.98
GU10-GRIDW	1.22	4.70

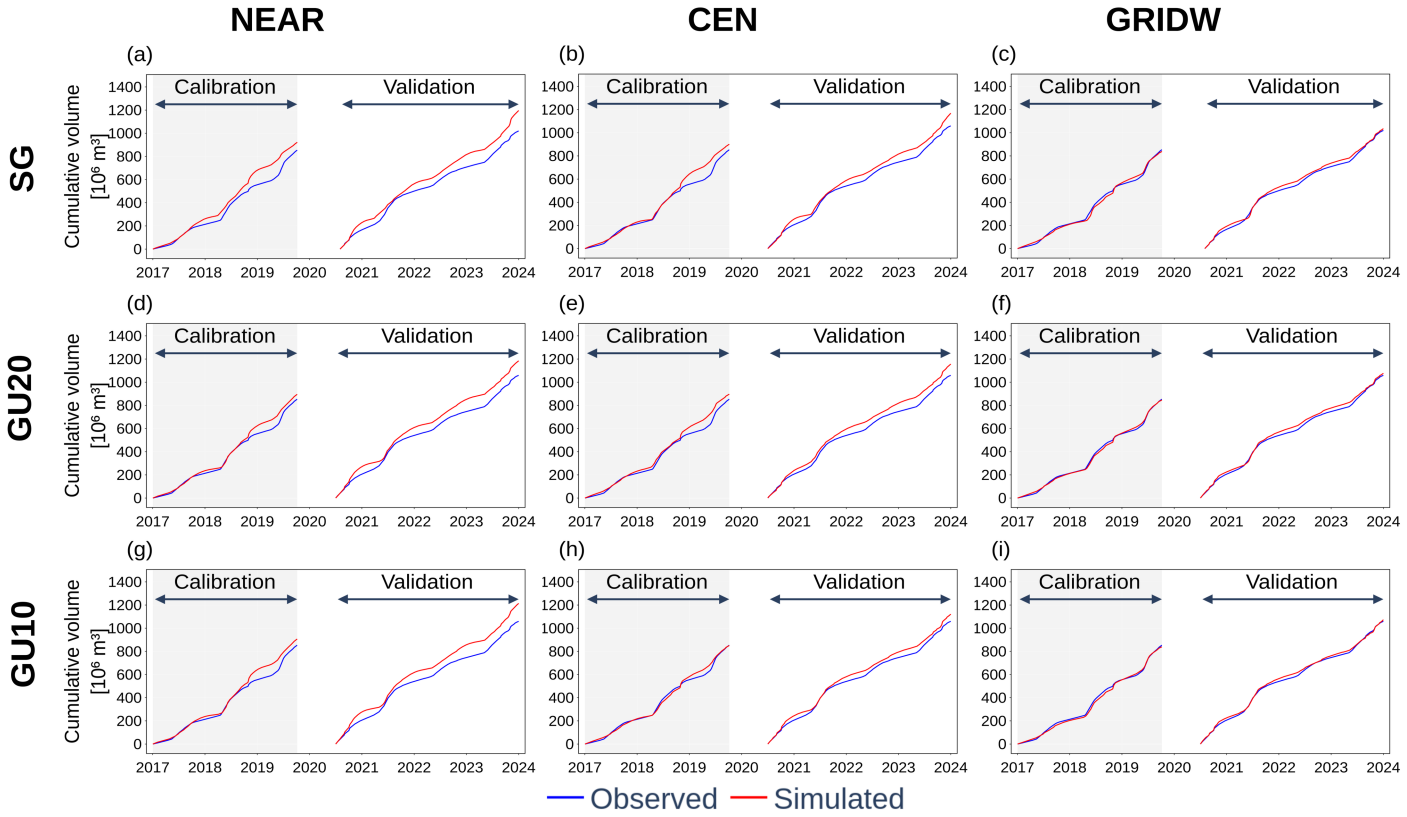

Figure D26 – Observed and simulated cumulative discharge volumes at the Malé stream gauge for the nine model configurations. Panels (a)–(i) correspond to: (a) SG-NEAR, (b) SG-CEN, (c) SG-GRIDW, (d) GU20-NEAR, (e) GU20-CEN, (f) GU20-GRIDW, (g) GU10-NEAR, (h) GU10-CEN, and (i) GU10-GRIDW. Rows correspond to the GU delineation method (SG, GU20, and GU10), whereas columns represent the method used to estimate the meteorological forcings (NEAR, CEN, and GRIDW). The calibration period (2017–2019) and the validation period (2020–2023) are shown separately within each panel. Blue lines represent the observed cumulative discharge volumes and red lines the simulated cumulative discharge volumes.

Table D5 – Total discharge-volume error (in %) at the Malé stream gauge for the nine model configurations. Rows correspond to the combinations of GU delineation (SG, GU20 and GU10) and meteorological forcing evaluation (NEAR, CEN, GRIDW). Columns report the relative error in cumulative discharge volume for the calibration and validation periods. Positive values indicate overestimation of the observed discharge volume, whereas negative values indicate underestimation.

	Calibration (2017-01-01–2019-10-30)	Validation (2020-07-01–2023-12-31)
Configuration	Volume error (%)	Volume error (%)
SG-NEAR	8.10	17.30
SG-CEN	5.72	11.32
SG-GRIDW	-1.69	1.54
GU20-NEAR	5.05	12.16
GU20-CEN	5.21	9.91
GU20-GRIDW	-0.76	1.35
GU10-NEAR	6.41	14.82
GU10-CEN	-0.18	5.83
GU10-GRIDW	-1.54	0.91

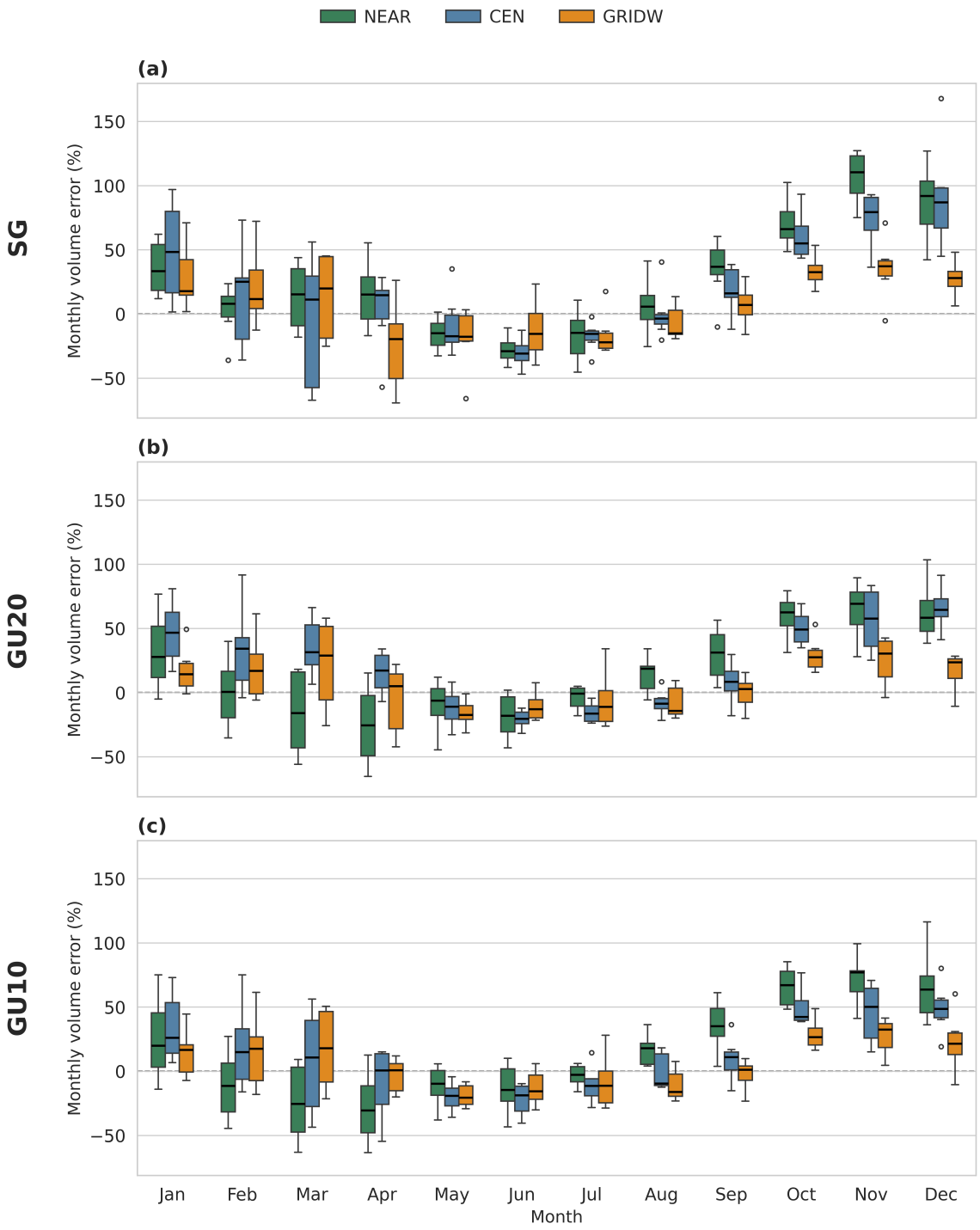

Figure D27 – Monthly discharge-volume error (%) at the Malé stream gauge for the nine model configurations. Box-plots summarize the distribution of monthly discharge-volume errors over the entire simulation period. The central line indicates the median, the box spans the interquartile range (25th–75th percentiles), the whiskers extend to 1.5 times the interquartile range, and points denote outliers. Panels (a), (b), and (c) correspond to the SG, GU20, and GU10 discretizations, respectively. Colors identify the meteorological forcing evaluation method: green for NEAR, blue for CEN, and orange for GRIDW.

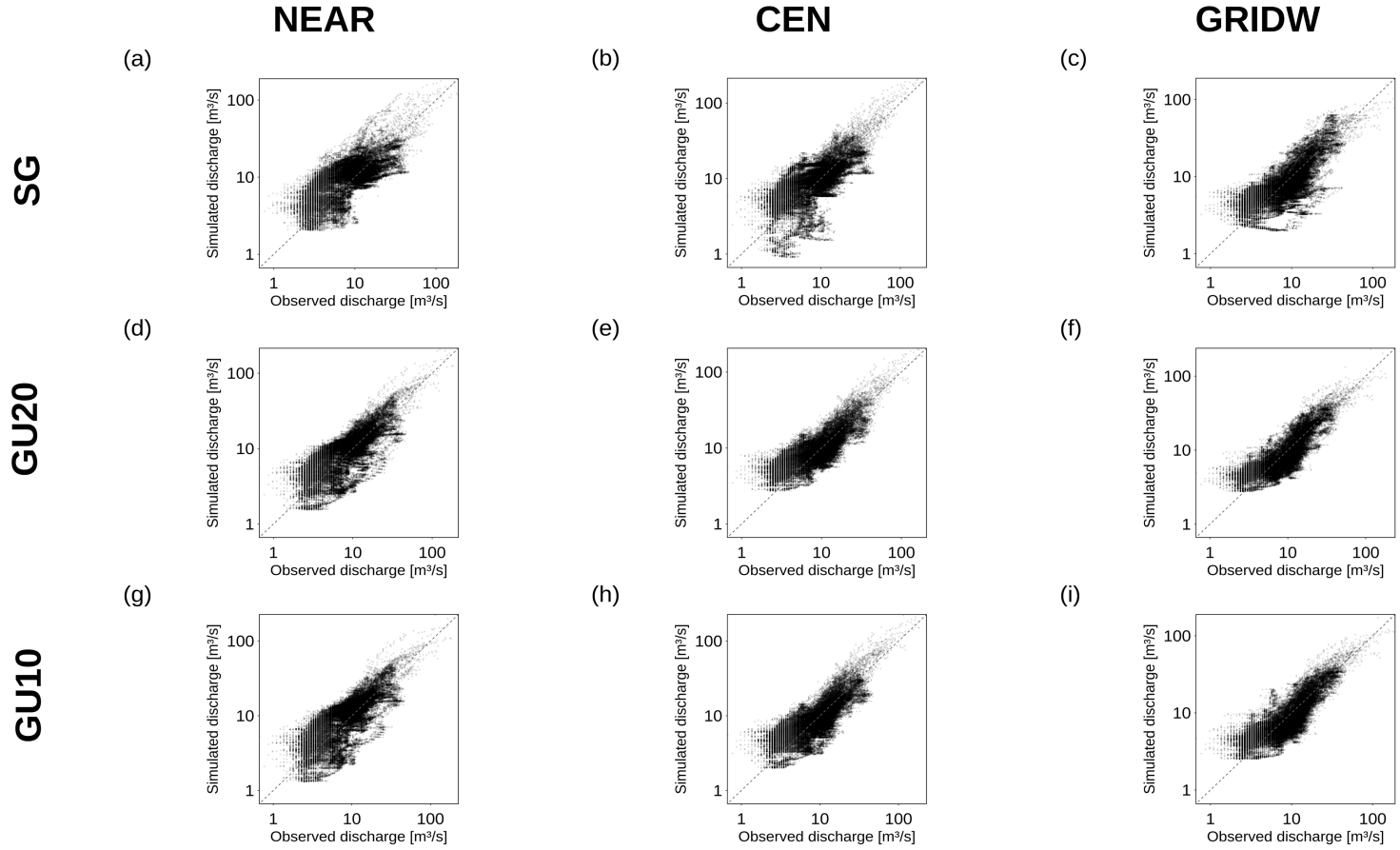

Figure D28 – Scatter plots of simulated versus observed discharge at the Malé stream gauge for the nine model configurations. Panels (a)–(i) correspond to: (a) SG-NEAR, (b) SG-CEN, (c) SG-GRIDW, (d) GU20-NEAR, (e) GU20-CEN, (f) GU20-GRIDW, (g) GU10-NEAR, (h) GU10-CEN, and (i) GU10-GRIDW. Rows correspond to the GU delineation method (SG, GU20, and GU10), whereas columns represent the method used to estimate the meteorological forcings (NEAR, CEN, and GRIDW). Calibration and validation periods are shown separately within each panel. The solid line represents the 1:1 relationship.

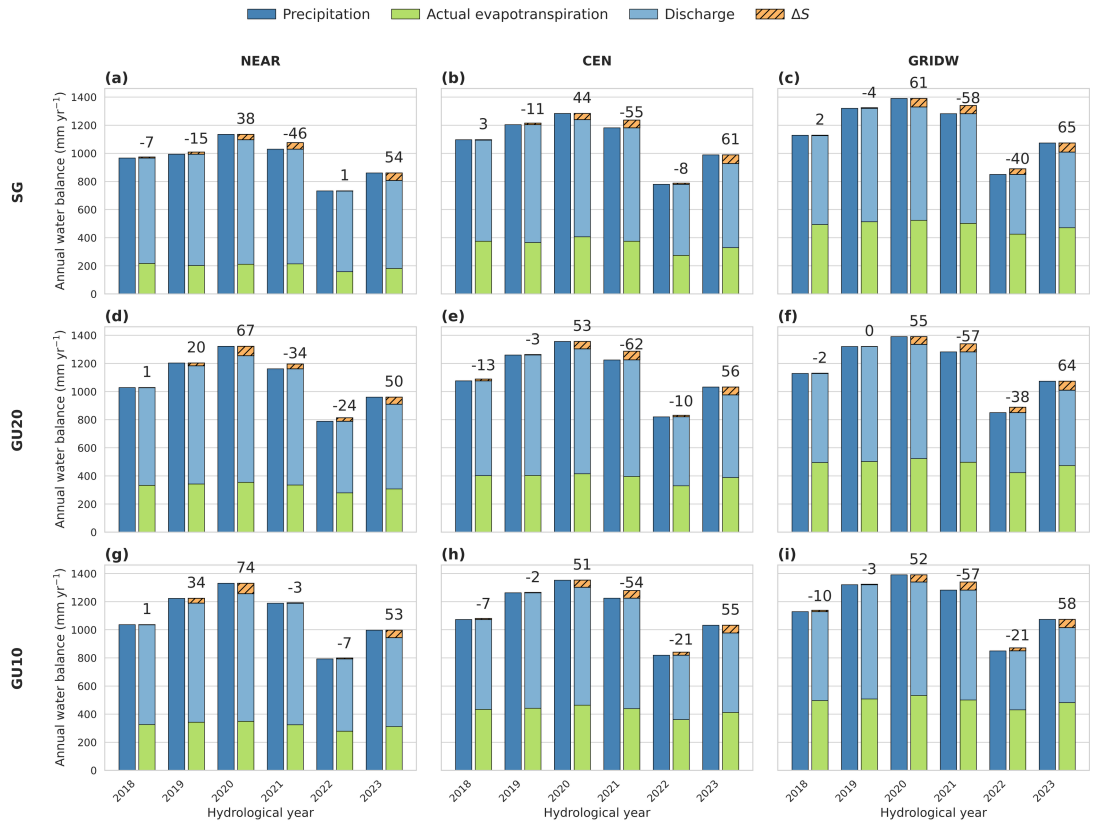

Figure D29 – Annual water-balance components for the nine model configurations at the Malé stream gauge. Panels (a)–(i) correspond to: (a) SG-NEAR, (b) SG-CEN, (c) SG-GRIDW, (d) GU20-NEAR, (e) GU20-CEN, (f) GU20-GRIDW, (g) GU10-NEAR, (h) GU10-CEN, and (i) GU10-GRIDW. Rows correspond to the GU delineation method (SG, GU20, and GU10), whereas columns represent the method used to estimate the meteorological forcings (NEAR, CEN, and GRIDW). Values are aggregated over hydrological years, defined from 1 October to 30 September and labeled according to the year in which they end. Dark-blue bars represent precipitation, while the adjacent stacked bars show actual evapotranspiration in green, discharge in light blue, and the change in total water storage (ΔS) in hatched orange. The numerical label associated with the storage component reports the signed value of ΔS in millimeters, with positive values indicating storage accumulation and negative values indicating storage depletion.

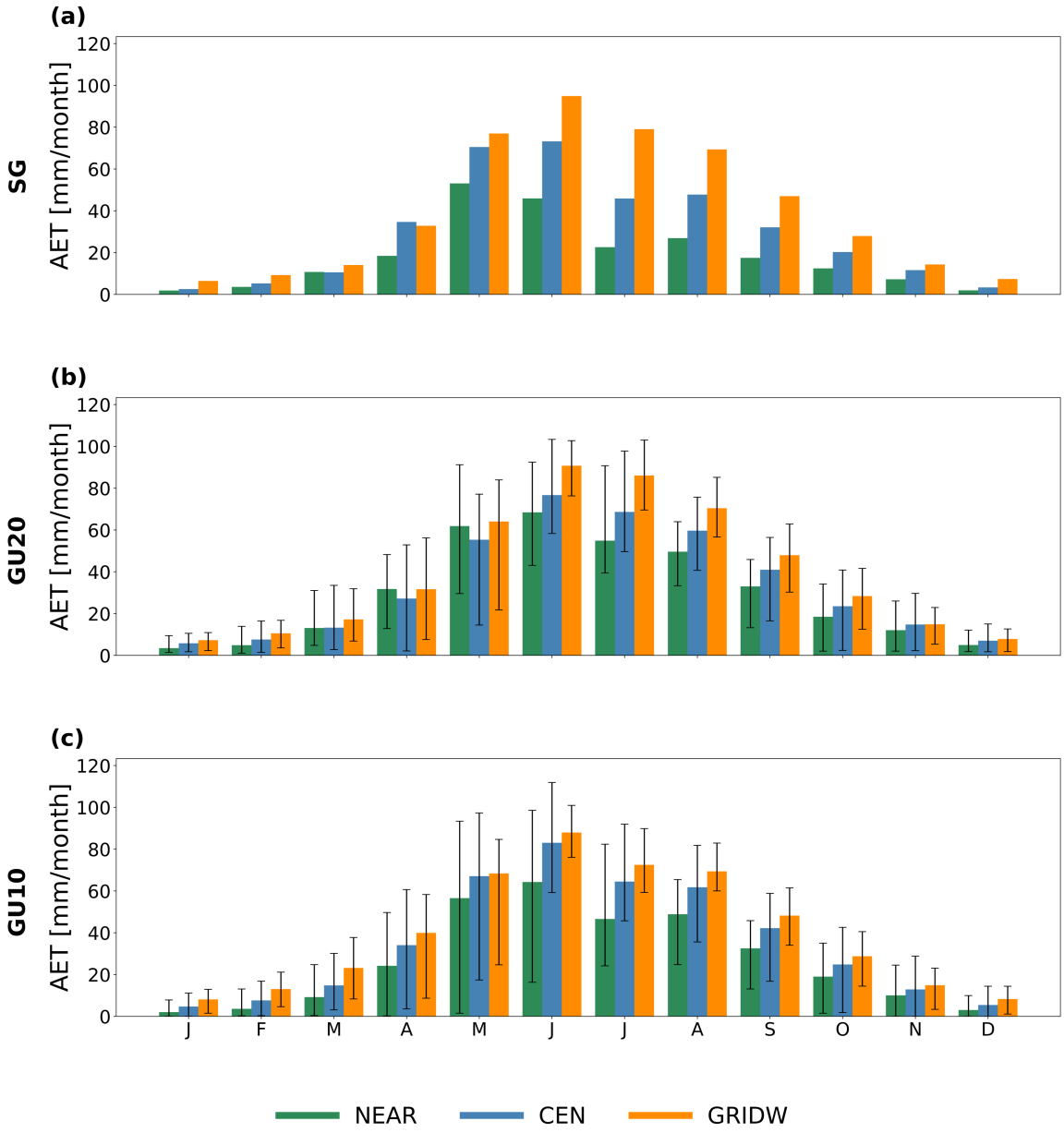

Figure D30 – Monthly mean actual evapotranspiration (AET) aggregated over all GUs. The bars represent the mean values obtained with CEN (blue) and GRIDW (orange) input configurations, while the error bars indicate the 2.5th–97.5th percentile range across GUs. Panel (a) refers to the SG configuration, panel (b) to the GU20 configuration and panel (c) to the GU10 configuration.

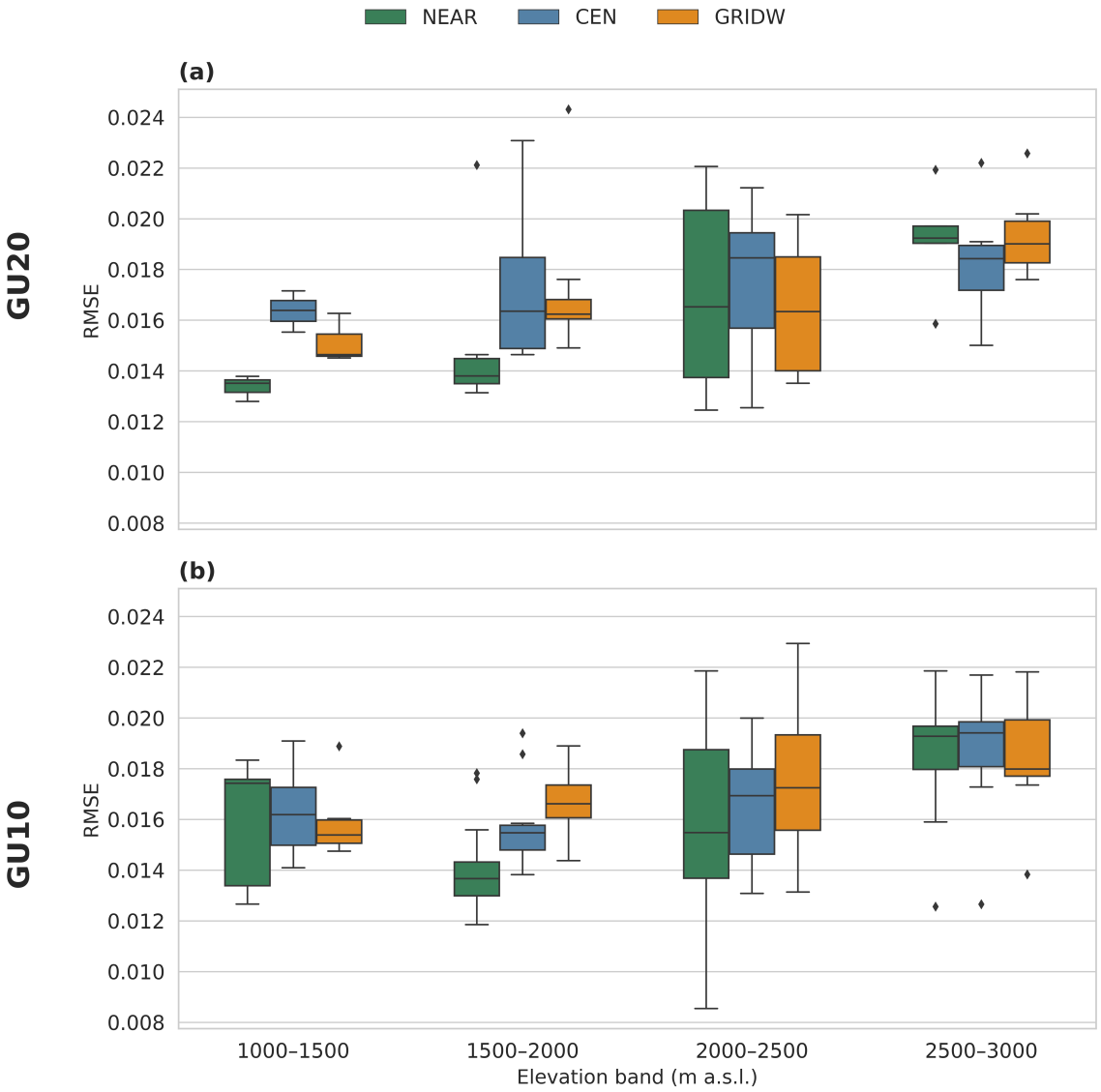

Figure D31 – Box-plots of Pearson's correlation coefficient (r) between simulated root-zone water content and the MERIDA product across common elevation bands of 500 m. Panel (a) shows the GU20 configuration, while panel (b) shows the GU10 configuration. Using identical elevation bands in both panels enables a direct comparison between the two spatial discretizations. Within each elevation band, the box-plots compare the NEAR, CEN, and GRIDW forcing evaluation methods. Higher correlation values indicate better agreement between simulated and reference (MERIDA) temporal dynamics.

D.3 Study 3: generalization, metric ranges and cumulative discharge

This section provides supplementary material supporting the analysis presented in Chapter 7.

Table D6 – Changes in key streamflow-performance metrics from calibration to validation (validation minus calibration), highlighting generalization loss across spatial configurations at the Vermiglio stream gauge.

Scale [km^2]	ΔKGE	ΔNSE	ΔRMSE [m^3/s]	ΔPBIAS [%]
90	-0.165	-0.201	-0.913	16.659
60	-0.180	-0.171	-0.751	16.419
30	-0.168	-0.200	-0.880	16.550
15	-0.161	-0.175	0.298	14.90
12	-0.144	-0.199	0.225	14.22
9	-0.147	-0.157	0.243	14.02
6	-0.128	-0.135	0.213	13.80
3	-0.122	-0.134	0.196	13.73

Table D7 – Changes in key streamflow-performance metrics from calibration to validation (validation minus calibration), highlighting generalization loss across spatial configurations at the Malé stream gauge.

Scale [km^2]	ΔKGE	ΔNSE	ΔRMSE [m^3/s]	ΔPBIAS [%]
90	-0.281	-0.148	0.000	26.804
60	-0.285	-0.152	0.000	26.620
30	-0.032	-0.028	0.000	3.170
15	-0.055	-0.025	0.223	3.80
12	-0.039	0.001	-0.056	8.27
9	-0.014	-0.007	-0.166	3.15
6	-0.052	0.020	-0.048	3.47
3	-0.037	-0.017	0.071	3.06

Table D8 – Maximum and minimum Kling–Gupta efficiency (KGE) values obtained across the ten independent calibration runs for each spatial configuration.

Scale	**KGE**$_{\max}$	**KGE**$_{\min}$
90 km^2	0.819	0.567
60 km^2	0.803	0.566
30 km^2	0.833	0.733
15 km^2	0.855	0.601
12 km^2	0.853	0.601
9 km^2	0.868	0.638
6 km^2	0.871	0.739
3 km^2	0.870	0.627

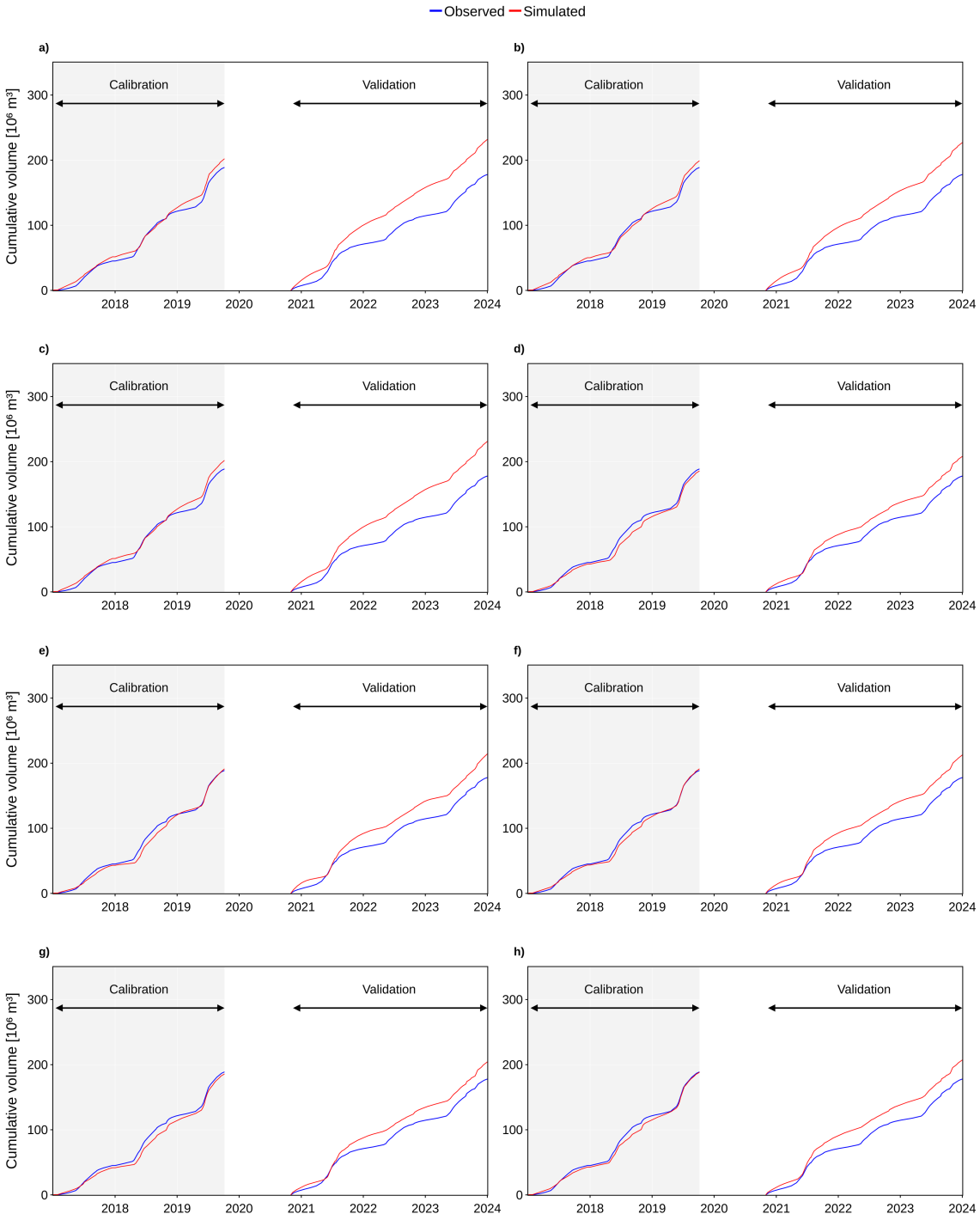

Figure D32 – Cumulative observed and simulated discharge volumes at the Vermiglio stream gauge for the calibration and validation periods across the eight spatial discretization configurations: (a) 90 km^2, (b) 60 km^2, (c) 30 km^2, (d) 15 km^2, (e) 12 km^2, (f) 9 km^2, (g) 6 km^2, and (h) 3 km^2. The separation between calibration and validation periods is indicated in each panel. Differences between the observed and simulated cumulative curves highlight the progressive accumulation of discharge-volume errors.

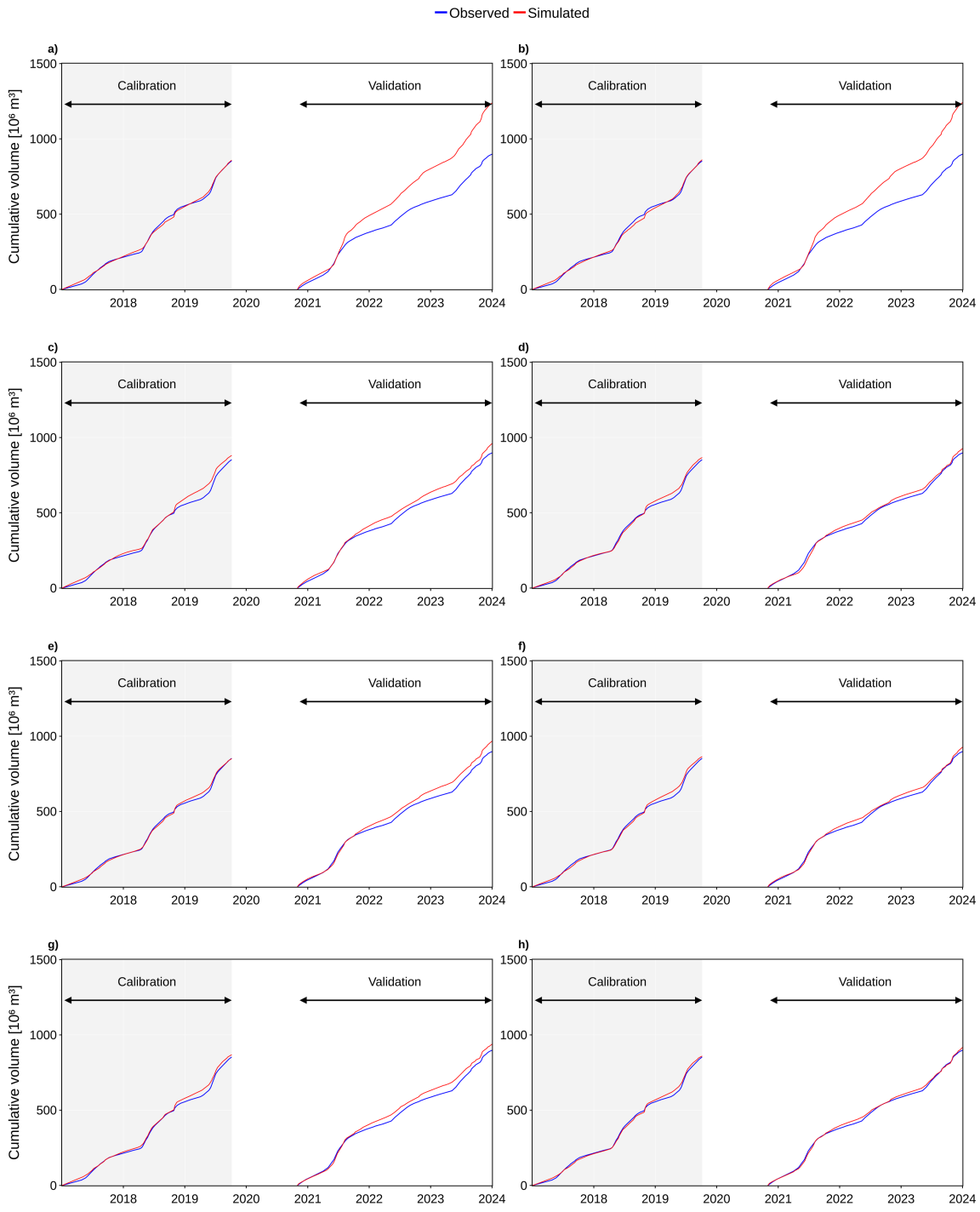

Figure D33 – Cumulative observed and simulated discharge volumes at the Malé stream gauge for the calibration and validation periods across the eight spatial discretization configurations: (a) 90 km^2, (b) 60 km^2, (c) 30 km^2, (d) 15 km^2, (e) 12 km^2, (f) 9 km^2, (g) 6 km^2, and (h) 3 km^2. The separation between calibration and validation periods is indicated in each panel. Differences between the observed and simulated cumulative curves highlight the progressive accumulation of discharge-volume errors.

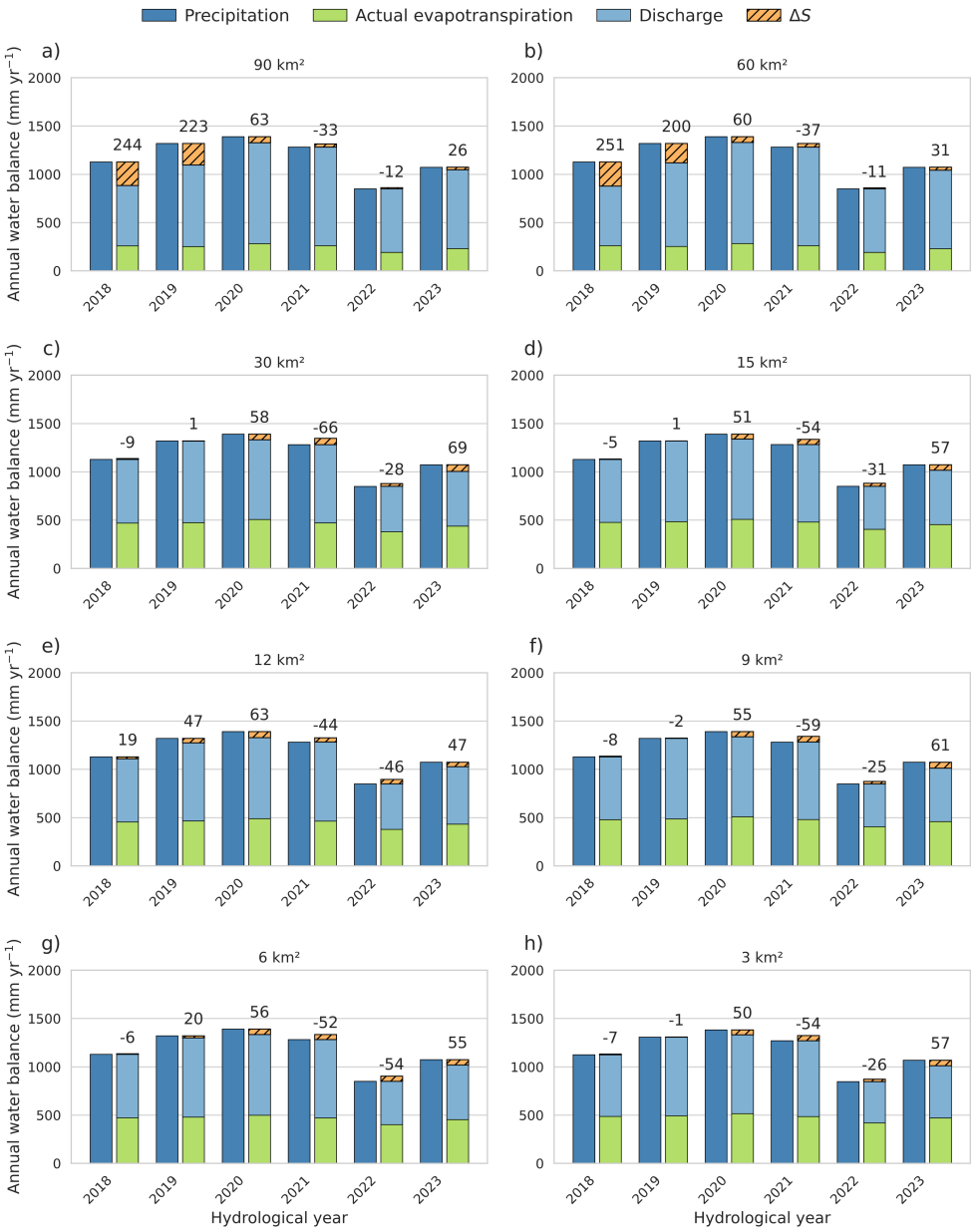

Figure D34 – Annual water-balance components for the catchment upstream of the Malé stream gauge across the eight spatial configurations. Panels (a)–(h) correspond to contributing-area thresholds of (a) 90 km^2, (b) 60 km^2, (c) 30 km^2, (d) 15 km^2, (e) 12 km^2, (f) 9 km^2, (g) 6 km^2, and (h) 3 km^2. Values are aggregated over hydrological years, defined from 1 October to 30 September and labeled according to the year in which they end. Dark-blue bars represent precipitation, while the adjacent stacked bars show actual evapotranspiration in green, discharge in light blue, and the change in total water storage (ΔS) in hatched orange. The numerical label associated with the storage component reports the signed value of ΔS in millimeters, with positive values indicating storage accumulation and negative values indicating storage depletion. When ΔS is negative, the hatched segment extends downward from the top of the combined actual-evapotranspiration and discharge components to the net water-balance total.

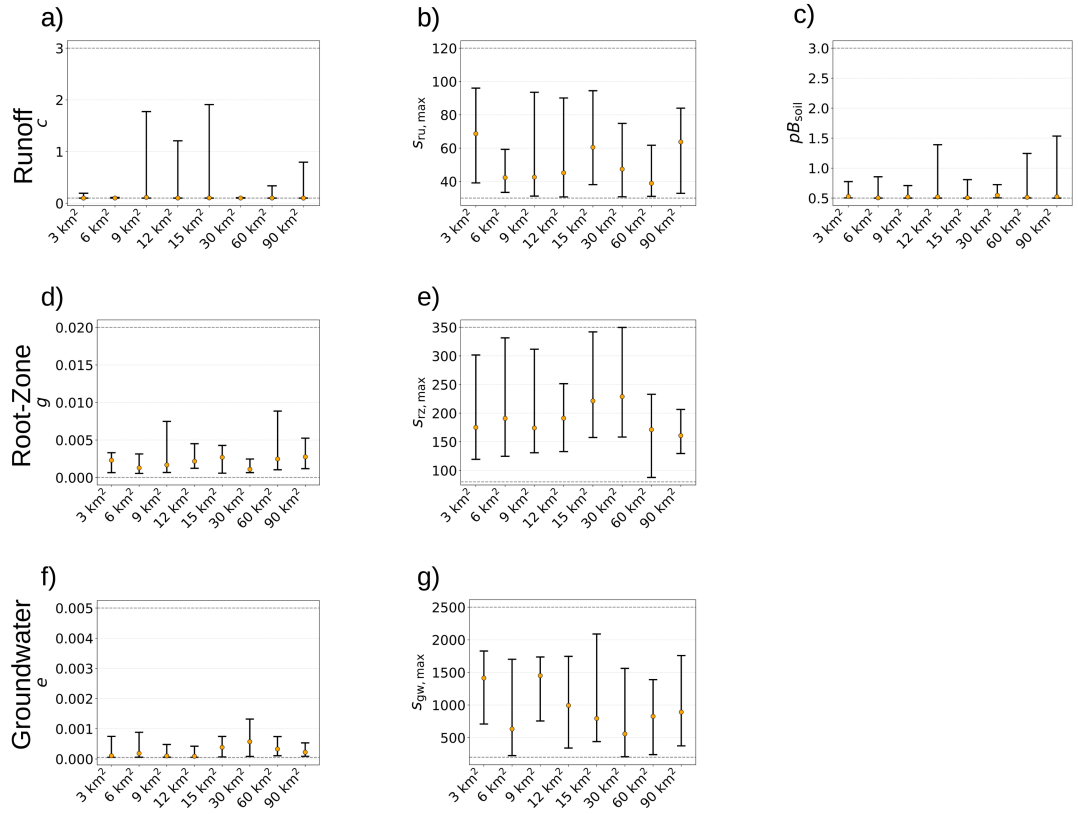

Figure D35 – Variability of root-zone, runoff, and groundwater parameters across spatial configurations (90 km^2 to 3 km^2), based on ten independent calibration runs per configuration. The first row of panels reports root-zone parameters (reservoir coefficient, maximum storage, and pB_{soil}), while the subsequent rows report runoff and groundwater parameters (nonlinear-reservoir coefficient and exponent). For each configuration, blue bars span the 10th–90th percentile range across the ten calibrations, and the orange marker indicates the median value.

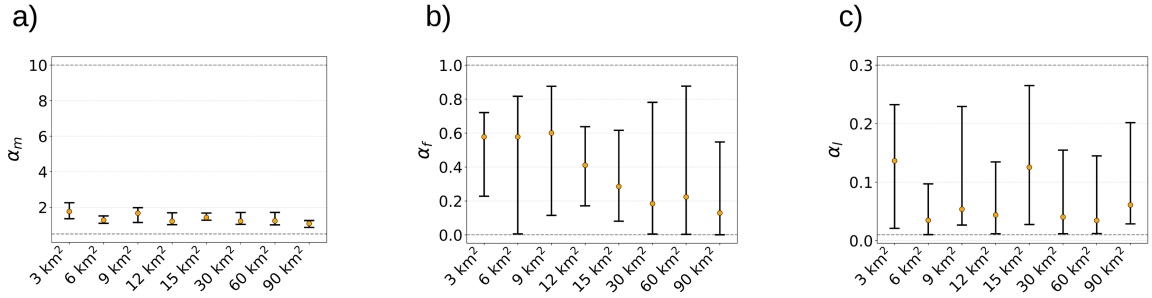

Figure D36 – Variability of snow-related parameters across spatial configurations (90 km^2 to 3 km^2) based on ten independent calibration runs per configuration. The three panels show (a) α_l, (b) α_f, and (c) α_m. For each configuration, blue bars span the 10th–90th percentile range across the ten calibrations, while the orange marker indicates the median value.

APPENDIX E. COMPUTATIONAL COST OF THE MODEL CONFIGURATIONS

In real-world applications, the computational resources required for model calibration and simulation can represent an important limiting factor. Model configurations should therefore not be evaluated solely in terms of predictive performance, nor should their comparison be restricted to differences in skill metrics. Computational requirements and efficiency also need to be considered, particularly when the additional computational cost associated with a more complex configuration is substantial relative to the improvement in model performance.

This appendix summarizes the computational resources required by the different model configurations considered in this thesis. All experiments were performed on the High Performance Computing (HPC) cluster of the University of Trento. The reported values should, however, be interpreted as indicative rather than as absolute measures of computational efficiency. Direct comparison between configurations is not straightforward because execution time is strongly affected by the structure of the computational workflow and by the degree of parallelism that can be achieved. Furthermore, while resource requests were explicitly specified for each job, actual consumption may occasionally exceed these allocations without a straightforward mechanism for direct monitoring or control.

In particular, the river network can be represented as a directed acyclic graph (DAG) which, in the present case of a single outlet, follows a tree-like structure. Computations proceed from the upstream branches toward the outlet, and a downstream node can only be evaluated once the calculations of all its contributing upstream nodes have been completed. Consequently, the number of nodes that can be processed concurrently varies throughout the simulation and is strongly controlled by the topology of the river network. While several independent upstream branches can be evaluated in parallel, the available level of parallelism generally decreases as the computation progresses downstream toward the outlet. As a

result, the allocated CPU cores cannot necessarily be fully utilized throughout the entire simulation, and the relationship between the number of allocated cores and wall-clock time is therefore not linear.

The following tables report the main statistics used to characterize the computational requirements of each configuration. These include wall-clock time, CPU time, mean and maximum RAM usage, together with the number of allocated CPU cores and the maximum requested memory. Both allocated and effectively used resources are reported because the optimal allocation depends on the characteristics of each configuration. For example, for the SG configuration with only two GUs, allocating more than four CPU cores would provide little or no computational benefit. The reported statistics therefore provide an indication of both the absolute computational demand and the efficiency with which the allocated HPC resources are utilized.

E.1 Mixed-resolution configurations (Chapter 5)

Table E1 – Summary of computational performance for the three model configurations (CFG1–CFG3). All simulations were executed on the HPC cluster of the University of Trento. Each configuration was run five times using a single compute node with 25 CPU cores and 350 GB of allocated memory (`select=1:ncpus=25:mem=350GB`), and a maximum wall-clock time of 6 hours as specified in the PBS job script. Configuration CFG3 corresponds to the finest spatial discretization, based on a GU area threshold of 1 km^2.

Configuration	RAM mean [GB]	RAM max [GB]	Wall-clock [hh:mm:ss]
CFG1	107.82	140.11	00:25:10
CFG2	146.07	183.91	00:58:00
CFG3	358.89	359.34	04:48:00

E.2 Interpolation and simulation cost by forcing strategy (Chapter 6)

The grid-based strategy (GRIDW) requires interpolation at every grid node and time step and dominates the cost of the whole chain; this is the main practical constraint identified in Chapter 6.

Table E2 – Summary of the computational requirements of the three model configurations during calibration. The reported values for GU20 and GU10 are averaged over the corresponding calibrations, as the computational cost mainly depends on the number of GUs. The table reports the number of GUs, the number of allocated CPU cores, the allocated memory, the mean memory usage, the mean wall-clock execution time, and the mean CPU time. CPU time is the value reported by the PBS scheduler; since the scheduler does not strictly confine the processes to the allocated cores, CPU time does not necessarily correspond to the product of wall-clock time and allocated CPU cores. Different HPC resource allocations were adopted for each configuration according to model complexity. The SG configuration was executed using a lightweight HPC setup, whereas GU20 and GU10 required substantially more computational resources. Therefore, the reported execution times and CPU times should be interpreted as indicators of the computational cost of each workflow rather than as a direct comparison of HPC efficiency. All calibrations were performed on the HPC cluster of the University of Trento.

Configuration	**GUs**	**CPUs**	**Allocated RAM [GB]**	**Mean RAM [GB]**	**Mean wall-clock time [hh:mm]**	**Mean CPU time [hh:mm]**
SG	2	2	35	6.33	06:02	18:30
GU20	30	22	120	57.70	65:43	370:26
GU10	51	22	120	55.66	81:49	754:14

Table E3 – Summary of the computational requirements for the 2017–2023 simulation runs. The table reports the number of GUs, the allocated CPU cores, the allocated memory, the mean memory usage, the mean wall-clock execution time, and the mean CPU time. The reported values are averaged over two simulation runs for each configuration. All simulations were performed on the HPC cluster of the University of Trento.

Configuration	**GUs**	**CPUs**	**Allocated RAM [GB]**	**Mean RAM [GB]**	**Mean wall-clock time [hh:mm]**	**Mean CPU time [hh:mm]**
SG	2	2	35	10.58	00:01	00:04
GU20	30	22	120	75.28	00:05	00:52
GU10	51	22	120	85.62	00:08	01:22

E.3 Cost as a function of the discretization scale (Chapter 7)

Table E4 – Summary of the computational requirements of the model configurations with different spatial resolutions at the Malé stream gauge. The table reports the spatial resolution, the number of CPU cores, the memory usage reported by PBS, the wall-clock execution time, and the total CPU time. CPU time is the value reported by the PBS scheduler; since the scheduler does not strictly confine the processes to the allocated cores, CPU time does not necessarily correspond to the product of wall-clock time and allocated CPU cores. The 9 km^2 configuration terminated with a non-zero exit status and should therefore be interpreted with caution. All simulations were performed on the HPC infrastructure of the University of Trento.

Resolution [km^2]	**CPUs**	**Memory used [GB]**	**Wall-clock time [hh:mm:ss]**	**CPU time [hh:mm:ss]**
90	25	9.29	00:02:23	00:13:09
60	25	17.49	00:02:14	00:12:35
30	25	27.41	00:03:25	00:24:14
15	25	55.12	00:08:56	02:23:35
12	25	45.25	00:10:07	02:37:22
9	25	62.45	00:17:05	12:26:20
6	30	117.72	00:43:54	22:23:55
3	25	128.69	01:14:49	13:33:10

Read together, the three sets of measurements support the trade-off statement made in Chapters 7 and 8: refining the discretization increases memory and wall-clock demand several-fold, while the corresponding gain in discharge skill, beyond an intermediate scale, is of the order of a few hundredths of KGE.

APPENDIX F. OPEN RESEARCH: DATA, CODE, AND WORKFLOWS

In line with the FAIR principles (Wilkinson et al., 2016) and the DARTHs blueprint (Rigon et al., 2022), discussed in Section 4.7, the research materials supporting this thesis have been made publicly available wherever permitted by the applicable data licenses. Most of the input datasets are openly accessible and, together with the published simulation workflows, model configurations, source code, and model outputs, permit the reproduction of nearly all the analyses presented in this thesis. Only a limited subset of the observational data could not be redistributed because of licensing or data-ownership restrictions. In these cases, the corresponding repositories provide information on the original data source and, where applicable, the procedure required to obtain access.

F.1 Data sources and access conditions

Meteorological observations from the MeteoTrentino network can be obtained from the MeteoTrentino data portal (MeteoTrentino, n.d.). Data managed by Fondazione Edmund Mach are available through its data-access services or upon request for research purposes. Discharge observations can be obtained from the Ufficio Dighe della Provincia autonoma di Trento (Ufficio Dighe, n.d.), subject to the applicable access conditions. Data from any source for which redistribution permission was not granted, or whose licensing terms do not permit redistribution, are not included in the repositories.

The scripts used to retrieve and preprocess the MERIDA, MODIS snow-cover, and Copernicus surface-soil-moisture products are archived in a dedicated Open Science Framework (OSF) repository (Andreis et al., 2026c).

The sources and access conditions are documented in the relevant OSF repositories.

F.2 Study-specific OSF deposits

The research materials associated with the three studies are archived in separate OSF repositories.

F.2.1 Study 1: mixed-resolution model configuration

The materials supporting the mixed-resolution experiment presented in Chapter 5 are archived in a dedicated OSF repository (Andreis et al., 2025).

F.2.2 Study 2: meteorological forcing and GU representation

The materials supporting the comparison of meteorological-forcing strategies and GU configurations presented in Chapter 6 are archived in a dedicated OSF repository (Andreis et al., 2026b).

F.2.3 Study 3: spatial discretization

The materials supporting the spatial-discretization experiments presented in Chapter 7 are archived in a dedicated OSF repository (Andreis et al., 2026a).

F.2.4 Repository structure and documentation

Although their exact organization reflects the design of the corresponding study, the three deposits follow a broadly consistent structure. Where applicable, they contain:

- the raster and vector datasets used as inputs to, or generated by, the geomorphological analysis;
- a dedicated GEOframe directory structure for each model configuration, containing:
  - the spatial data representing each GU, including clipped digital elevation model and sky-view-factor rasters;

  - the meteorological forcing time series assigned to each GU according to the strategy being examined;
  - the simulated discharge and other water-budget flux and state-variable time series for each GU;

- the OMS3 simulation files and model-configuration files required to execute the experiments;
- the parameter files, including the calibrated parameter sets;
- the observational data used for calibration and evaluation, wherever their licenses permit redistribution; when redistribution is prohibited, the documentation identifies the original source and access procedure;
- the scripts used for data preparation, post-processing, and analysis;
- README files describing the repository structure, the principal contents, and the sequence required to reproduce the experiments.

Each deposit includes a README file or an accompanying OSF wiki describing its contents, directory structure, file-naming conventions, and the role of the main scripts. The documentation also identifies any external data that cannot be redistributed and provides links to the relevant source-code repositories. Where required to preserve the exact computational environment used in the thesis, the deposits additionally include or reference the corresponding compiled JAR files.

F.3 Source code

The source-code repositories for the main software components used or developed in this thesis are listed below. The specific tagged versions used in each study are documented in the corresponding research deposit.

- Embedded Reservoir Model (ERM): github.com/bubbobne/Embedded_Reservoir_Model-ERM;
- `GeoFrameInputBuilderWITHZOOM`: github.com/bubbobne/geoframe.zoom;
- kriging and semivariogram components: github.com/bubbobne/Krigings.

The GitHub repositories provide the source code and its development history, whereas the OSF deposits preserve the scripts, configuration files, parameter sets, and, where relevant, compiled versions used to produce the thesis results.

These components form part of the wider GEOframe system, whose repositories and documentation were reorganized during this work, as discussed in Section 4.6. The wider software infrastructure includes the GEOframe component collection (geoframecomponents, n.d.), the geomorphological and raster-processing tools provided by the Horton Machine (Rigon et al., 2006), and the Python tooling supporting the preparation, execution, and analysis of GEOframe simulations (Bancheri et al., 2024).

In addition, the source code for the GEOframe Explorer web application and RESTful API is available at github.com/bubbobne/geoframe-explorer.

F.4 License and citation

Unless otherwise specified, the materials produced as part of this thesis and made publicly available through the associated repositories are released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Different licensing or access conditions may apply to specific datasets, software, and third-party resources, as documented in the relevant OSF repositories. No registration is required to access the publicly available materials. Each resource should be cited using the DOI and citation information provided in the corresponding repository.

F.5 Acknowledgments of data providers

The author would like to thank the Consorzio di Miglioramento Fondiario di Tres and Agriduemila Hub Innovation S.r.l. for providing the CRNS probe data. Special thanks are also due to Fondazione Edmund Mach, particularly its Agro-Meteorology and Irrigation Unit, for its support and for providing high-resolution meteorological data.

REFERENCES

Abera, W., Antonello, A., Franceschi, S., Formetta, G., & Rigon, R. (2014). The uDig spatial toolbox for hydro-geomorphic analysis. In N. Clarke (Ed.), *Geomorphological techniques*. British Society for Geomorphology.

Abera, W., Formetta, G., Borga, M., & Rigon, R. (2017). Estimating the water budget components and their variability in a pre-alpine basin with JGrass-NewAGE. *Adv. Water Resour.*, *104*, 37–54.

Aberegg, I., Egli, M., Sartori, G., & Purves, R. (2009). Modelling spatial distribution of soil types and characteristics in a high alpine valley (val di sole, trentino, italy). *Studi Trentini di Scienze Naturali, Acta Geologica*, *85*, 39–50.

Ademmer, M., Jannsen, N., & Meuchelböck, S. (2023). Extreme weather events and economic activity: The case of low water levels on the Rhine River. *German Economic Review*, *24*(2), 121–144. https://doi.org/10.1515/ger-2022-0077

Aidoo, E. N., Mueller, U., Goovaerts, P., & Hyndes, G. A. (2015). Evaluation of geostatistical estimators and their applicability to characterise the spatial patterns of recreational fishing catch rates. *Fisheries Research*, *168*, 20–32.

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., & Martin, E. (2008). From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrol. Earth Syst. Sci. Discuss.*, *5*(3), 1603–1640.

An, H., Ouyang, C., & Chen, X. (2025). Real-time estimation of SMAP soil moisture in mountainous areas and its impact on rainfall-runoff simulation. *J. Hydrol. (Amst.)*, *660*(133487), 133487.

Andreis, D., Formetta, G., Bancheri, M., & Rigon, R. (2025). *Input and output data from multi-scale hru-based hydrological simulations comparing configurations with fixed and mixed resolution*. https://doi.org/10.17605/OSF.IO/5SZ8A

[dataset] Andreis, D., Formetta, G., & Rigon, R., (2026). *Assessing Spatial Discretization Impacts on Semi-Distributed Hydrological Simulations*. https://doi.org/10.17605/OSF.IO/UAY8G

[dataset] Andreis, D., Formetta, G., & Rigon, R., (2026). *Effects of spatial representation of meteorological data on different hydrological processes in semi-distributed hydrological models*. https://doi.org/10.17605/OSF.IO/J9ZVC

Andreis, D., Formetta, G., & Rigon, R. (2026c). *Remote sensing data pipeline*. OSF. https://doi.org/10.17605/OSF.IO/47DW9

Apache Software Foundation. (n.d.). Apache maven [Accessed: 2026-08-19].

Arnone, E., Zoratti, V., Formetta, G., Bosa, S., & Petti, M. (2023). Predicting peakflows in mountain river basins and data-scarce areas: A case study in northeastern italy. *Hydrol. Sci. J.*, *68*(3), 432–447.

Asaoka, Y., & Kominami, Y. (2013). Incorporation of satellite-derived snow-cover area in spatial snowmelt modeling for a large area: Determination of a gridded degree-day factor. *Annals of Glaciology*, *54*, 205–213. https://doi.org/10.3189/2013AoG62A218

Assandri, G. (2015). *Biodiversity conservation in permanent crops and grasslands: The case of intensive apple orchards in northern italy* [Doctoral dissertation, Università di Pavia]. Pavia, Italy. https://iris.unipv.it/retrieve/e1f104fb-9ba9-8c6e-e053-1005fe0aa0dd/Assandri_tesi%20dottorato_finale_ridotto.pdf

Aston, A. R. (1979). Rainfall interception by eight small trees. *J. Hydrol. (Amst.)*, *42*(3-4), 383–396.

Autorità di Bacino Distrettuale del Fiume Po. (2022, June 29). *Emergenza siccità nel distretto del Po: portate ancora molto basse, prelievo non ridotto e cuneo salino alla quota record di 30,6 km*. Retrieved August 24, 2026, from https://www.adbpo.it/wp-content/uploads/2022/06/Comunicato-Stampa-29_06_2022-1.pdf

Autorità di Bacino Distrettuale del Fiume Po. (2026, August 20). *Aggiornamento settimanale della situazione idrologica del 20 agosto 2026*. Retrieved August 24, 2026, from https://www.adbpo.it/aggiornamento-settimanale-situazione-idrologica-21-07-2026-2-4/

Avanzi, F., Munerol, F., Milelli, M., Gabellani, S., Massari, C., Girotto, M., Cremonese, E., Galvagno, M., Bruno, G., Morra di Cella, U., Rossi, L., Altamura, M., & Ferraris, L. (2024). Winter snow deficit was a harbinger of summer 2022 socio-hydrologic drought in the po basin, italy. *Communications Earth & Environment*, *5*(1), 64.

Avesani, D., Galletti, A., Piccolroaz, S., Bellin, A., & Majone, B. (2021). A dual-layer MPI continuous large-scale hydrological model including human systems. *Environmental Modelling & Software*, *139*, 105003.

Azimi, S., Massari, C., Formetta, G., Barbetta, S., Tazioli, A., Fronzi, D., Modanesi, S., Tarpanelli, A., & Rigon, R. (2023). On understanding mountainous carbonate basins of the mediterranean using parsimonious modeling solutions. *Hydrol. Earth Syst. Sci.*, *27*(24), 4485–4503.

Bancheri, M. (2017). *A flexible approach to the estimation of water budgets and its connection to the travel time theory* [Doctoral dissertation, Università degli studi di Trento]. Università degli studi di Trento.

Bancheri, M., Busti, N., Rigon, R., et al. (2024). Geoframepy: Python tools for the pre- and post-processing of geoframe simulations.

Bancheri, M., Serafin, F., Bottazzi, M., Abera, W., Formetta, G., & Rigon, R. (2018). The design, deployment, and testing of kriging models in GEOframe with SIK-0.9.8. *Geosci. Model Dev.*

Bancheri, M., Serafin, F., & Rigon, R. (2019). The representation of hydrological dynamical systems using extended petri nets (EPN). *Water Resour. Res.*, *55*(11), 8895–8921.

Bardossy, A., & Das, T. (2008). Influence of rainfall observation network on model calibration and application. *Hydrol. Earth Syst. Sci.*

Bassetti, M., & Borsato, A. (2007). Evoluzione geomorfologica della bassa valle dell'adige dall'ultimo massimo glaciale: Sintesi delle conoscenze e riferimenti ad aree limitrofe. *Studi Trentini di Scienze Naturali, Acta Geologica*, *82*, 31–42.

Bellin, A., Majone, B., Cainelli, O., Alberici, D., & Villa, F. (2016). A continuous coupled hydrological and water resources management model. *Environmental Modelling & Software*, *75*, 176–192.

Benavides, R., Montes, F., Rubio, A., & Osoro, K. (2007). Geostatistical modelling of air temperature in a mountainous region of northern spain. *Agricultural and Forest Meteorology*, *146*, 173–188.

Beniston, M. (2006). Mountain weather and climate: A general overview and a focus on climatic change in the alps. *Hydrobiologia*, *562*(1), 3–16.

Beniston, M. (2012). Impacts of climatic change on water and associated economic activities in the swiss alps. *J. Hydrol. (Amst.)*, *412-413*, 291–296.

Beniston, M., & Stoffel, M. (2014). Assessing the impacts of climatic change on mountain water resources. *Science of the Total Environment*, *493*, 1129–1137.

Beniston, M., Stoffel, M., & Hill, M. (2011). Impacts of climatic change on water and natural hazards in the alps: Can current water governance cope with future challenges? examples from the european "acqwa" project [Adapting to Climate Change: Reducing Water-related Risks in Europe]. *Environmental Science & Policy*, *14*(7), 734–743. https://doi.org/10.1016/j.envsci.2010.12.009

Berezowski, T., Szcześniak, M., Kardel, I., Michałowski, R., Okruszko, T., Mezghani, A., & Piniewski, M. (2016). CPLFD-GDPT5: High-resolution gridded daily precipitation and temperature data set for two largest Polish river basins. *Earth System Science Data*, *8*(1), 127–139. https://doi.org/10.5194/essd-8-127-2016

Berndt, C., & Haberlandt, U. (2018). Spatial interpolation of climate variables in Northern Germany—Influence of temporal resolution and network density. *Journal of Hydrology: Regional Studies*, *15*, 184–202. https://doi.org/10.1016/j.ejrh.2018.02.002

Berndt, C., Rabiei, E., & Haberlandt, U. (2014). Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. *Journal of Hydrology*, *508*, 88–101. https://doi.org/10.1016/j.jhydrol.2013.10.028

Beven, K. (1989). Changing ideas in hydrology—the case of physically-based models. *Journal of Hydrology*, *105*(1-2), 157–172. https://doi.org/10.1016/0022-1694(89)90101-7

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1-2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K. (2025). On the future of hydroecological models of everywhere. *Environ. Model. Softw.*, *188*(106431), 106431.

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Beven, K., & Westerberg, I. (2011). On red herrings and real herrings: Disinformation and information in hydrological inference. *Hydrological Processes*, *25*(10), 1676–1680. https://doi.org/10.1002/hyp.7963

Beven, K. J. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, *5*(1), 1–12. https://doi.org/10.5194/hess-5-1-2001

Beven, K. J., & Cloke, H. L. (2012). Comment on "hyperresolution global land surface modeling: Meeting a grand challenge for monitoring earth's terrestrial water" by eric f. wood et al. *Water Resources Research*, *48*(1), W01801. https://doi.org/10.1029/2011WR010982

Beven, K. J., & Hornberger, G. M. (1982). Assessing the effect of spatial pattern of precipitation in modeling stream flow hydrographs. *JAWRA Journal of the American Water Resources Association*, *18*(5), 823–829. https://doi.org/10.1111/j.1752-1688.1982.tb00078.x

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., … Wood, E. F. (2015). Hyper-resolution global hydrological modelling: What is next? "everywhere and locally relevant". *Hydrological Processes*, *29*(2), 310–320. https://doi.org/10.1002/hyp.10391

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., & Montanari, A. (2012). Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Phys. Chem. Earth (2002)*, *42-44*, 70–76.

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3–4), 251–290. https://doi.org/10.1002/hyp.3360090305

Blöschl, G. (1999). Scaling issues in snow hydrology. *Hydrological Processes*, *13*(14-15), 2149–2175. https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., … Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrol. Sci. J.*, *64*(10), 1141–1158.

Blöschl, G., Grayson, R. B., & Sivapalan, M. (1995). On the representative elementary area (REA) concept and its utility for rainfall–runoff modelling. *Hydrological Processes*, *9*(3-4), 313–330. https://doi.org/10.1002/hyp.3360090307

Boaga, J., Rossi, M., & Cassiani, G. (2013). Monitoring soil–plant interactions in an apple orchard using 3d electrical resistivity tomography. *Procedia Environmental Sciences*, *19*, 394–402. https://doi.org/10.1016/j.proenv.2013.06.045

Bonaldo, D., Bellafiore, D., Ferrarin, C., Ferretti, R., Ricchi, A., Sangelantoni, L., & Vitelletti, M. L. (2023). The summer 2022 drought: A taste of future climate for the Po Valley (Italy)? *Regional Environmental Change*, *23*, Article 1. https://doi.org/10.1007/s10113-022-02004-z

Bonanno, R., Lacavalla, M., & Sperati, S. (2019). A new high-resolution meteorological reanalysis italian dataset: MERIDA. *Quarterly Journal of the Royal Meteorological Society*, *145*, 1756–1779. https://doi.org/10.1002/qj.3530

Bosellini, A., Castellarin, A., Dal Piaz, G. V., & Nardin, M. (2023, November 8). *Carta litologica del Trentino* [Web page]. Provincia autonoma di Trento. Retrieved December 15, 2025, from https://www.provincia.tn.it/News/Approfondimenti/Carta-litologica-del-Trentino

Bottazzi, M., Bancheri, M., Mobilia, M., Bertoldi, G., Longobardi, A., & Rigon, R. (2021). Comparing evapotranspiration estimates from the GEOframe-prospero model with penman–monteith and priestley-taylor approaches under different climate conditions. *Water*, *13*(9), 1221.

Bouabdelli, S., Massari, C., Morlot, M., Castelli, M., & Formetta, G. (2025). High-resolution (1 km daily) potential evapotranspiration dataset over europe and the mediterranean region. https://doi.org/10.5194/egusphere-egu25-18052

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.*, *78*(1), 1–3.

Brocca, L., Tarpanelli, A., Moramarco, T., Melone, F., Ratto, S. M., Cauduro, M., Ferraris, S., Berni, N., Ponziani, F., Wagner, W., & Melzer, T. (2013). Soil moisture estimation in alpine catchments through modeling and satellite observations. *Vadose Zone J.*, *12*(3), 1–10.

Brogi, C., Pisinaras, V., Köhli, M., Dombrowski, O., Hendricks Franssen, H.-J., Babakos, K., Chatzi, A., Panagopoulos, A., & Bogena, H. R. (2023). Monitoring irrigation in small orchards with cosmic-ray neutron sensors. *Sensors (Basel)*, *23*(5).

Brunetti, M., Maugeri, M., Nanni, T., Simolo, C., & Spinoni, J. (2014). High-resolution temperature climatology for Italy: Interpolation method intercomparison. *International Journal of Climatology*, *34*(4), 1278–1296. https://doi.org/10.1002/joc.3764

Brunner, M. I., Götte, J., Schlemper, C., & Van Loon, A. F. (2023). Hydrological drought generation processes and severity are changing in the alps [e2022GL101776 2022GL101776]. *Geophysical Research Letters*, *50*(2), e2022GL101776. https://doi.org/10.1029/2022GL101776

Brunner, M. I., Slater, L., Tallaksen, L. M., & Clark, M. (2021). Challenges in modeling and predicting floods and droughts: A review. *WIREs Water*, *8*, e1520. https://doi.org/10.1002/wat2.1520

Camana, F. A., & Deutsch, C. V. (2019). The nugget effect [Accessed 2025-09-15]. http://geostatisticslessons.com/lessons/nuggeteffect

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., & Konopka, A. E. (1994). Field-scale variability of soil properties in central iowa soils. *Soil Science Society of America Journal*, *58*, 1501–1511. https://doi.org/10.2136/sssaj1994.03615995005800050033x

Cammalleri, C., Naumann, G., Mentaschi, L., Formetta, G., Forzieri, G., Gosling, S., Bisselink, B., De Roo, A., & Feyen, L. (2020). *Global warming and drought impacts in*

the EU (EUR 29956 EN). Joint Research Centre, European Commission. Luxembourg, Publications Office of the European Union. https://doi.org/10.2760/597045

Carletti, F., Michel, A., Casale, F., Burri, A., Bocchiola, D., Bavay, M., & Lehning, M. (2022). A comparison of hydrological models with different level of complexity in alpine regions in the context of climate change. *Hydrol. Earth Syst. Sci.*, *26*(13), 3447–3475.

Cavaleri, L., Barbariol, F., Bertotti, L., Besio, G., & Ferrari, F. (2022). The 29 october 2018 storm in northern italy: Its multiple actions in the ligurian sea. *Progress in Oceanography*, *201*, 102715. https://doi.org/10.1016/j.pocean.2021.102715

Cavaleri, L., Bajo, M., Barbariol, F., Bastianini, M., Benetazzo, A., Bertotti, L., Chiggiato, J., Davolio, S., Ferrarin, C., Magnusson, L., Papa, A., Pezzutto, P., Pomaro, A., & Umgiesser, G. (2019). The october 29, 2018 storm in northern italy – an exceptional event and its modeling. *Prog. Oceanogr.*, *178*, 102178.

Cavalleri, F., Lussana, C., Viterbo, F., Brunetti, M., Bonanno, R., Manara, V., Lacavalla, M., Sperati, S., Raffa, M., Capecchi, V., Cesari, D., Giordani, A., Cerenzia, I. M. L., & Maugeri, M. (2024). Multi-scale assessment of high-resolution reanalysis precipitation fields over italy. https://doi.org/10.48550/arXiv.2407.11517

Cazorzi, F., & Dalla Fontana, G. (1996). Snowmelt modelling by combining air temperature and a distributed radiation index. *Journal of Hydrology*, *181*(1-4), 169–187. https://doi.org/10.1016/0022-1694(95)02913-3

CH2018. (2018). *CH2018 – climate scenarios for switzerland. technical report* (tech. rep.) (Technical report for the Swiss climate scenarios CH2018). National Centre for Climate Services (NCCS). https://www.nccs.admin.ch/nccs/en/home/climate-change-and-impacts/swiss-climate-change-scenarios/technical-report.html

Chen, Z., Lucianetti, G., & Hartmann, A. (2023). Understanding groundwater storage and drainage dynamics of a high mountain catchment with complex geology using a semi-distributed process-based modelling approach. *Journal of Hydrology*, *625*, 130067. https://doi.org/10.1016/j.jhydrol.2023.130067

Chiang, L.-C., Chaubey, I., Gitau, M. W., & Arnold, J. G. (2015). The NHDPlus dataset, watershed subdivision, and SWAT hydrologic simulations. *Hydrological Sciences Journal*, *59*(10), 1768–1786. https://doi.org/10.1080/02626667.2013.870664

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., & Goldscheider, N. (2023). When best is the enemy of good – critical evaluation of performance criteria in hydrological models. *Hydrology and Earth System Sciences*, *27*(13), 2397–2411. https://doi.org/10.5194/hess-27-2397-2023

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., & Peters-Lidard, C. D. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, *21*(7), 3427–3440. https://doi.org/10.5194/hess-21-3427-2017

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., & Zeng, X. (2015a). Improving the representation of hydrologic processes in earth system models. *Water Resources Research*, *51*(8), 5929–5956. https://doi.org/10.1002/2015WR017096

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9), 2010WR009827. https://doi.org/10.1029/2010WR009827

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., & Rasmussen, R. M. (2015b). A unified approach for process-based hydrologic modeling: 1. modeling concept. *Water Resour. Res.*, *51*(4), 2498–2514.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., & Papalexiou, S. M. (2021). The Abuse of Popular Performance Metrics in Hydrologic Modeling. *Water Resources Research*, *57*(9), e2020WR029001. https://doi.org/10.1029/2020WR029001

Comola, F., Schaefli, B., Ronco, P. D., Botter, G., Bavay, M., Rinaldo, A., & Lehning, M. (2015). Scale-dependent effects of solar radiation patterns on the snow-dominated hydrologic response. *Geophysical Research Letters*, *42*(10), 3895–3902. https://doi.org/10.1002/2015GL064075

Corradini, P., Omizzolo, A., Rizzari, M., Corrado, F., Durbiano, E., La Riccia, L., Santoro, G., Nenert, B., Notarnicola, C., Crespi, A., Bertoldi, G., Tonelli, C., & Prèmier, V. (2023). *Report on the effects of climate change on the alpine space snow tourism destinations* (tech. rep.) (Deliverable D.1.1.1, Interreg Alpine Space). BeyondSnow project. https://www.alpine-space.eu/project/beyondsnow/

Crespi, A., Matiu, M., Bertoldi, G., Petitta, M., & Zebisch, M. (2021). A high-resolution gridded dataset of daily temperature and precipitation records (1980–2018) for Trentino-South Tyrol (north-eastern Italian Alps). *Earth System Science Data*, *13*(6), 2801–2818. https://doi.org/10.5194/essd-13-2801-2021

Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, *22*, 239–252.

Dahlmann, H., Stephan, R., & Stahl, K. (2022). Upstream-downstream asymmetries of drought impacts in major river basins of the european alps. *Frontiers in Water*, *4*, 1061991.

Dal Molin, M., Kavetski, D., & Fenicia, F. (2021). SuperflexPy 1.3.0: An open-source python framework for building, testing, and improving conceptual hydrological models. *Geoscientific Model Development*, *14*(11), 7047–7072.

Dall'Amico, M., Endrizzi, S., Gruber, S., & Rigon, R. (2011). A robust and energy-conserving model of freezing variably-saturated soil. *The Cryosphere*, *5*(2), 469–484. https://doi.org/10.5194/tc-5-469-2011

Dall'Amico, M., Tasin, S., Di Paolo, F., Brian, M., Leoni, P., Tornatore, F., Formetta, G., Wani, J. M., Rigon, R., & Roati, G. (2025). 30-years (1991-2021) snow water equivalent dataset in the po river district, italy. *Scientific Data*, *12*(1), 374.

Dall'Amico, M., Tasin, S., Di Paolo, F., Brian, M., Leoni, P., Tornatore, F., Formetta, G., Wani, J. M., Rigon, R., & Roati, G. (2025). 30-years (1991-2021) snow water equivalent dataset in the po river district, italy. *Sci. Data*, *12*(1), 374.

Daly, C., Gibson, W., Taylor, G., Johnson, G., & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. *Climate Research*, *22*, 99–113. https://doi.org/10.3354/cr022099

D'Amato, C., & Rigon, R. (2025). Elementary mathematics helps to shed light on the transpiration budget under water stress. *Ecohydrology*, *18*, e70009. https://doi.org/10.1002/eco.70009

Das, N. N., Entekhabi, D., Dunbar, R. S., Chaubell, M. J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P. E., Walker, J. P., Berg, A., Bosch, D. D., Caldwell, T., Cosh, M. H., Collins, C. H., Lopez-Baeza, E., & Thibeault, M. (2019). The SMAP and copernicus sentinel 1A/B microwave active-passive high resolution surface soil moisture product. *Remote Sens. Environ.*, *233*, 111380.

David, O., Ascough, J. C. I. I., Lloyd, W., Green, T. R., Rojas, K. W., Leavesley, G. H., & Ahuja, L. R. (2013). A software engineering perspective on environmental modeling framework design: The object modeling system. *Environmental Modelling & Software*, *39*, 201–213.

David, O., Markstrom, S. L., Rojas, K. W., Ahuja, L. R., & Schneider, I. W. (2016). The object modeling system. In *Agricultural system models in field research and technology transfer* (pp. 317–330). CRC Press.

Dawson, N., Broxton, P., & Zeng, X. (2018). Evaluation of remotely sensed snow water equivalent and snow cover extent over the contiguous united states. *J. Hydrometeorol.*, *19*(11), 1777–1791.

De Lavenne, A., Thirel, G., Andréassian, V., Perrin, C., & Ramos, M.-H. (2016). Spatial variability of the parameters of a semi-distributed hydrological model. *Proceedings of the International Association of Hydrological Sciences*, *373*, 87–94. https://doi.org/10.5194/piahs-373-87-2016

De Santis, D., & Biondi, D. (2018). A quality assessment of the soil water index by the propagation of ASCAT soil moisture error estimates through an exponential filter. *Int. J. Remote Sens.*, *39*(1), 232–257.

Dehotin, J., & Braud, I. (2008). Which spatial discretization for distributed hydrological models? Proposition of a methodology and illustration for medium to large-scale catchments. *Hydrology and Earth System Sciences*, *12*(3), 769–796. https://doi.org/10.5194/hess-12-769-2008

Desilets, D., Zreda, M., & Ferré, T. P. A. (2010). Nature's neutron probe: Land surface hydrology at an elusive scale with cosmic rays. *Water Resour. Res.*, *46*(11).

Di Marco, N., Righetti, M., Avesani, D., Zaramella, M., Notarnicola, C., & Borga, M. (2020). Comparison of MODIS and model-derived snow-covered areas: Impact of land use and solar illumination conditions. *Geosci. J.*, *10*(4), 134.

Dong, J., & Ochsner, T. E. (2018). Soil texture often exerts a stronger influence than precipitation on mesoscale soil moisture patterns. *Water Resour. Res.*, *54*(3), 2199–2211.

Dong, X., Lai, X., Wang, Y., Dong, W., Zhu, J., Dong, L., & Cen, S. (2022). Applicability evaluation of multiple sets of soil moisture data on the tibetan plateau. *Frontiers in Earth Science*, *10*, 872413. https://doi.org/10.3389/feart.2022.872413

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., … Lecomte, P. (2017). ESA CCI soil moisture for improved earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.*, *203*, 185–215.

Dumont, M., Monteiro, D., Filhol, S., Gascoin, S., Marty, C., Hagenmuller, P., Morin, S., Choler, P., & Thuiller, W. (2025). The European Alps in a changing climate: Physical trends and impacts. *Comptes Rendus. Géoscience*, *357*, 25–42. https://doi.org/10.5802/crgeos.288

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Society for Industrial; Applied Mathematics (SIAM). https://doi.org/10.1137/1.9781611970319

Egli, M., Mirabella, A., Mancabelli, A., & Sartori, G. (2004). Weathering of soils in alpine areas as influenced by climate and parent material. *Clays and Clay Minerals*, *52*(3), 287–303. https://doi.org/10.1346/CCMN.2004.0520304

Électricité de France. (2022). *Bilan de l'été 2022: Note de synthèse*. Électricité de France. Paris. Retrieved August 24, 2026, from https://www.edf.fr/sites/groupe/files/2022-11/EDF_Note%20de%20synth%C3%A8se_Bilan%20%C3%A9t%C3%A9%202022.pdf

Ellis, C., Pomeroy, J. W., Brown, T., & MacDonald, J. (2010). Simulation of snow accumulation and melt in needleleaf forest environments. *Hydrology and Earth System Sciences*, *14*, 925–940. https://doi.org/10.5194/hess-14-925-2010

Emmanuel, I., Andrieu, H., Leblois, E., & Flahaut, B. (2012). Temporal and spatial variability of rainfall at the urban hydrological scale. *Journal of Hydrology*, *430-431*, 162–172. https://doi.org/10.1016/j.jhydrol.2012.02.013

Endrizzi, S., Gruber, S., Dall'Amico, M., & Rigon, R. (2014). Geotop 2.0: Simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. *Geoscientific Model Development*, *7*(6), 2831–2857. https://doi.org/10.5194/gmd-7-2831-2014

Essery, R., Mazzotti, G., Barr, S., Jonas, T., Quaife, T., & Rutter, N. (2025). A flexible snow model (fsm 2.1.1) including a forest canopy. *Geoscientific Model Development*, *18*(12), 3583–3605. https://doi.org/10.5194/gmd-18-3583-2025

European Commission Directorate-General Joint Research Centre. (2017, January 1). *Surface soil moisture 2014-present (raster 1 km), europe, daily*. Version version 1.

Evin, G., Le Lay, M., Fouchier, C., Penot, D., Colleoni, F., Mas, A., Garambois, P.-A., & Laurantin, O. (2024). Evaluation of hydrological models on small mountainous catchments: Impact of the meteorological forcings. *Hydrology and Earth System Sciences*, *28*(1), 261–281. https://doi.org/10.5194/hess-28-261-2024

Fatima, E., Kumar, R., Attinger, S., Kaluza, M., Rakovec, O., Rebmann, C., Rosolem, R., Oswald, S. E., Samaniego, L., Zacharias, S., & Schrön, M. (2024). Improved representation of soil moisture processes through incorporation of cosmic-ray neutron count measurements in a large-scale hydrologic model. *Hydrol. Earth Syst. Sci.*, *28*(24), 5419–5441.

Favargiotti, S., Pasquali, M., Chioni, C., & Pianegonda, A. (2022). Water resources and health tourism in val di sole: Key elements for innovating with nature in the italian inner territories. *Sustain. Sci. Pract. Policy*, *14*(18), 11294.

Favilli, F., Cherubini, P., Collenberg, M., Egli, M., Sartori, G., Schoch, W., & Haeberli, W. (2010). Charcoal fragments of alpine soils as an indicator of landscape evolution during the holocene in val di sole (trentino, italy). *The Holocene*, *20*(1), 67–79. https://doi.org/10.1177/0959683609348850

Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. motivation and theoretical development. *Water Resour. Res.*, *47*(11).

Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L. (2008). Understanding catchment behavior through stepwise model concept improvement. *Water Resour. Res.*, *44*(1).

Fielding, R. T. (2000). *Architectural styles and the design of network-based software architectures* [Doctoral dissertation, University of California, Irvine]. https://ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation_2up.pdf

Fischer, A. M., Strassmann, K. M., Croci-Maspoli, M., Hama, A. M., Knutti, R., Kotlarski, S., Schaer, C., Schnadt Poberaj, C., Ban, N., Bavay, M., Beyerle, U., Bresch, D. N., Broennimann, S., Burlando, P., Casanueva, A., Fatichi, S., Feigenwinter, I., Fischer, E. M., Hirschi, M., … Zubler, E. M. (2022). Climate scenarios for switzerland CH2018 – approach and implications. *Climate Services*, *26*, 100288. https://doi.org/10.1016/j.cliser.2022.100288

Flügel, W.-A. (1995). Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using prms/mms in the drainage basin of the river bröl, germany. *Hydrological Processes*, *9*(3-4), 423–436. https://doi.org/10.1002/hyp.3360090313

Foehn, A., García Hernández, J., Schaefli, B., & De Cesare, G. (2018). Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. *Journal of Hydrology*, *563*, 1092–1110. https://doi.org/10.1016/j.jhydrol.2018.05.027

Fondazione Edmund Mach. (2022, August 10). Fondazione edmund mach [Accessed: 2024-4-3].

Formetta, G., Dallan, E., Borga, M., & Marra, F. (2024). Sub-daily precipitation return levels in ungauged locations: Added value of combining observations with convection permitting simulations. *Advances in Water Resources*, *194*, 104851. https://doi.org/10.1016/j.advwatres.2024.104851

Formetta, G., Antonello, A., Franceschi, S., David, O., & Rigon, R. (2014a). Hydrological modelling with components: A GIS-based open-source framework. *Environmental Modelling & Software*, *55*, 190–200.

Formetta, G., Bancheri, M., David, O., & Rigon, R. (2016). Performance of site-specific parameterizations of longwave radiation. *Hydrol. Earth Syst. Sci.*, *20*(11), 4641–4654.

Formetta, G., Kampf, S. K., David, O., & Rigon, R. (2014b). Snow water equivalent modeling components in NewAge-JGrass. *Geosci. Model Dev.*, *7*(3), 725–736.

Franz, T. E., Wahbi, A., Zhang, J., Vreugdenhil, M., Heng, L., Dercon, G., Strauss, P., Brocca, L., & Wagner, W. (2020). Practical data products from cosmic-ray neutron sensing for hydrological applications. *Frontiers in Water*, *2*.

Franz, T. E., Zreda, M., Ferre, T. P. A., Rosolem, R., Zweck, C., Stillman, S., Zeng, X., & Shuttleworth, W. J. (2012). Measurement depth of the cosmic ray soil moisture probe affected by hydrogen from various sources. *Water Resour. Res.*, *48*(8).

Franz, T. E., Zreda, M., Rosolem, R., Hornbuckle, B. K., Irvin, S. L., Adams, H., Kolb, T. E., Zweck, C., & Shuttleworth, W. J. (2013). Ecosystem-scale measurements of biomass water using cosmic ray neutrons. *Geophys. Res. Lett.*, *40*(15), 3929–3933.

Gafurov, A., & Bárdossy, A. (2009). Cloud removal methodology from MODIS snow cover product. *Hydrol. Earth Syst. Sci.*, *13*, 1361–73.

Gascoin, S., Hagolle, O., Huc, M., Jarlan, L., Dejoux, J.-F., Szczypta, C., Marti, R., & Sánchez, R. (2015). A snow cover climatology for the pyrenees from modis snow products. *Hydrology and Earth System Sciences*, *19*, 2337–2351. https://doi.org/10.5194/hess-19-2337-2015

Gaudard, L., Romerio, F., Dalla Valle, F., Gorret, R., Maran, S., Ravazzani, G., Stoffel, M., & Volonterio, M. (2014). Climate change impacts on hydropower in the swiss and italian alps. *Science of the Total Environment*, *493*, 1211–1221.

geoframecomponents. (n.d.). *Geoframe components (github organization)* [Accessed: 2025-12-13]. Retrieved December 13, 2025, from https://github.com/geoframecomponents/

Gianessi, S., Polo, M., Stevanato, L., Lunardon, M., Francke, T., Oswald, S., Ahmed, H., Tolosa, A., Weltin, G., Dercon, G., Fulajtar, E., Heng, L., & Baroni, G. (2024). Testing a novel sensor design to jointly measure cosmic-ray neutrons, muons and gamma rays for non-invasive soil moisture estimation. *Geosci. Instrum. Methods Data Syst.*, *13*(1), 9–25.

Ginesta, M., Yiou, P., Messori, G., & Faranda, D. (2023). A methodology for attributing severe extratropical cyclones to climate change based on reanalysis data: The case study of storm alex 2020. *Clim. Dyn.*, *61*(1), 229–253.

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., & Stoffel, M. (2014). 21st century climate change in the european alps—a review. *Science of the total environment*, *493*, 1138–1151.

Gradle Inc. (n.d.). Gradle user manual [Accessed: 2026-08-19].

Grundmann, J., Hörning, S., & Bárdossy, A. (2019). Stochastic reconstruction of spatio-temporal rainfall patterns by inverse hydrologic modelling. *Hydrology and Earth System Sciences*, *23*(1), 225–237. https://doi.org/10.5194/hess-23-225-2019

Gubler, S., Fiddes, J., Keller, M., & Gruber, S. (2011). Scale-dependent measurement and analysis of ground surface temperature variability in alpine terrain. *The Cryosphere*, *5*(2), 431–443. https://doi.org/10.5194/tc-5-431-2011

GuildSomm. (2016). Italy part iii: Northern italy. https://www.guildsomm.com/research/expert_guides/w/expert-guides/2624/italy-part-iii-northern-italy

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, *377*(1), 80–91.

Gupta, H. V., & Nearing, G. S. (2014). Debates—the future of hydrological sciences: A (common) path forward? using models and data to learn: A systems theoretic perspective on the future of hydrological science. *Water Resources Research*, *50*(6), 5351–5359. https://doi.org/10.1002/2013WR015096

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*, 3802–3813. https://doi.org/10.1002/hyp.6989

Gyawali, D. R., & Bárdossy, A. (2022). Development and parameter estimation of snowmelt models using spatial snow-cover observations from MODIS. *Hydrol. Earth Syst. Sci.*, *26*(12), 3055–3077.

Haberlandt, U. (2007). Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology*, *332*(1-2), 144–157. https://doi.org/10.1016/j.jhydrol.2006.06.028

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th python in science conference* (pp. 11–15).

Hall, D. K., & Riggs, G. A. (2021). MODIS/aqua snow cover daily L3 global 500m grid, version 61.

Hall, D. K., Riggs, G. A., & Salomonson, V. V. (2006). MODIS/terra snow cover 5-min L2 swath 500m, version 5.

Härer, S., Bernhardt, M., Siebers, M., & Schulz, K. (2018). On the need for a time- and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales. *Cryosphere*, *12*(5), 1629–1642.

Hay, L. E., & Umemoto, M. (2007, June 6). *Multiple-objective stepwise calibration using luca* (tech. rep.). U.S. Geological Survey.

Hayashi, M. (2020). Alpine hydrogeology: The critical role of groundwater in sourcing the headwaters of the world. *Groundwater*, *58*(4), 498–510. https://doi.org/10.1111/gwat.12965

Hengl, T. (2007). *A practical guide to geostatistical mapping of environmental variables*. Office for Official Publications of the European Communities.

Her, Y., Frankenberger, J. R., Chaubey, I., & Srinivasan, R. (2015). Threshold effects in HRU definition of the soil and water assessment tool. *Journal of the American Water Resources Association*, *51*(2), 504–522. https://doi.org/10.1111/jawr.12275

Hilmers, T., Avdagić, A., Bartkowicz, L., Bielak, K., Binder, F., Bončina, A., et al. (2019). The productivity of mixed mountain forests comprised of *Fagus sylvatica*, *Picea abies* and *Abies alba* across europe. *Forestry*, *92*(5), 512–524. https://doi.org/10.1093/forestry/cpz035

Hock, R. (1999). A distributed temperature-index ice- and snowmelt model including potential direct solar radiation. *Journal of Glaciology*, *45*(149), 101–111. https://doi.org/10.3189/S0022143000003087

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, *282*, 104–115. https://doi.org/10.1016/S0022-1694(03)00257-9

Hofstra, N., Haylock, M., New, M., Jones, P., & Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research: Atmospheres*, *113*(D21), 2008JD010100. https://doi.org/10.1029/2008JD010100

Horton, P., Schaefli, B., & Kauzlaric, M. (2022). Why do we have so many different hydrological models? a review based on the case of switzerland. *WIREs Water*, *9*, e1574. https://doi.org/10.1002/wat2.1574

Hu, Q., Li, Z., Wang, L., Huang, Y., Wang, Y., & Li, L. (2019). Rainfall Spatial Estimations: A Review from Spatial Interpolation to Multi-Source Data Merging. *Water*, *11*(3), 579. https://doi.org/10.3390/w11030579

Huss, M., Bookhagen, B., Huggel, C., Jacobsen, D., Bradley, R. S., Clague, J. J., Vuille, M., Buytaert, W., Cayan, D. R., Greenwood, G., et al. (2017). Toward mountains without permanent snow and ice. *Earth's Future*, *5*(5), 418–435.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., … Baillie, J. E. M. (2020). Importance and vulnerability of the world's water towers. *Nature*, *577*(7790), 364–369. https://doi.org/10.1038/s41586-019-1822-y

Immerzeel, W. W., van Beek, L. P. H., & Bierkens, M. F. P. (2010). Climate change will affect the asian water towers. *Science*, *328*(5984), 1382–1385.

Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction to applied geostatistics*. Oxford University Press.

ISPAT – Istituto di Statistica della Provincia Autonoma di Trento. (n.d.). ISPAT - annuario on-line [Accessed: 2024-4-1].

ISPRA. (2013). *Habitat 9420 – foreste alpine di* Larix decidua *e/o* Pinus cembra*. scheda per il monitoraggio e la valutazione ai sensi della direttiva habitat* (tech. rep.) (Documento tecnico nazionale per l'habitat 9420). Istituto Superiore per la Protezione e la Ricerca Ambientale. Roma, Italy. https://reportingdirettivahabitat.isprambiente.it/documents/monitoraggio_IV_report/habitat/9420m_mon.pdf

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, *29*(8), 2637–2649. https://doi.org/10.1029/93WR00877

Jewell, S. A., & Gaussiat, N. (2015). An assessment of kriging-based rain-gauge–radar merging techniques. *Quarterly Journal of the Royal Meteorological Society*, *141*(691), 2300–2313. https://doi.org/10.1002/qj.2522

Jiang, A., Zhang, W., Liu, X., Zhou, F., Li, A., Peng, H., & Wang, H. (2024). Improving hydrological process simulation in mountain watersheds: Integrating WRF model gridded precipitation data into the SWAT model. *Journal of Hydrology*, *639*, 131687. https://doi.org/10.1016/j.jhydrol.2024.131687

Joint, F. A. O., et al. (2017). *Cosmic ray neutron sensing: Use, calibration and validation for soil moisture estimation* (tech. rep.). Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture.

Katya, D.-P., Josie, G., E., W. M., Rafael, R., Lucile, V., Allan, L., & Chris, S. (2020). Opportunities and challenges in using catchment-scale storage estimates from cosmic ray neutron sensors for rainfall-runoff modelling. *Journal of Hydrology*.

Kavetski, D., & Clark, M. P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. impact of time stepping schemes on model analysis and prediction. *Water Resour. Res.*, *46*(10).

Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. application and experimental insights. *Water Resour. Res.*, *47*(11).

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, *424-425*, 264–277.

Koehler, J., Dietz, A. J., Zellner, P., Baumhoer, C. A., Dirscherl, M., Cattani, L., Vlahović, Ž., Alasawedah, M. H., Mayer, K., Haslinger, K., Bertoldi, G., Jacob, A., & Kuenzer, C. (2022). Drought in northern italy: Long earth observation time series reveal snow line elevation to be several hundred meters above long-term average in 2022. *Remote Sensing*, *14*(23), 6091.

Koltsida, E., & Kallioras, A. (2022). Multi-variable SWAT model calibration using satellite-based evapotranspiration data and streamflow. *Hydrology*, *9*(7), 112.

Koltsida, E., Mamassis, N., & Kallioras, A. (2023). Hydrological modeling using the soil and water assessment tool in urban and peri-urban environments: The case of kifisos experimental subbasin (athens, greece). *Hydrol. Earth Syst. Sci.*, *27*(4), 917–931.

Kotlarski, S., Gobiet, A., Morin, S., Olefs, M., Rajczak, J., & Samacoïts, R. (2023). 21st century alpine climate change. *Climate Dynamics*, *60*, 65–86. https://doi.org/10.1007/s00382-022-06303-3

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Proceedings of the 8th Workshop for Large Scale Hydrological Modelling - Oppurg 2004 - 8th workshop for Large-scale hydrological modelling, Oppurg, Germany, 11–13 November 2004*, *5*, 89–97.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, *52*, 119–139.

Kumar, P., & Foufoula-Georgiou, E. (1994). Characterizing Multiscale Variability of Zero Intermittency in Spatial Rainfall. *Journal of Applied Meteorology*, *33*(12), 1516–1525. https://doi.org/10.1175/1520-0450(1994)033<1516:CMVOZI>2.0.CO;2

Kumar, R., Samaniego, L., & Attinger, S. (2010). The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics. *Journal of Hydrology*, *392*(1-2), 54–69. https://doi.org/10.1016/j.jhydrol.2010.07.047

Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales. *Water Resources Research*, *49*(1), 360–379. https://doi.org/10.1029/2012WR012195

Kumar, S., & Ojha, R. (2024). A scaling-based model to describe temporal variability in soil hydraulic properties. *J. Hydrol. (Amst.)*, *636*(131265), 131265.

Kumar, V., Jahangeer, J., Singh, R., & Dikshit, P. K. S. (2025). Editorial: Advancement in hydrological modeling and water resources management for achieving sustainable development goals (sdgs). *Frontiers in Water*, *7*, 1599795. https://doi.org/10.3389/frwa.2025.1599795

Lamontagne, J. R., Barber, C. A., & Richard M. Vogel. (2020). Improved Estimators of Model Performance Efficiency for Skewed Hydrologic Data. *Water Resources Research*, *56*(9), e2020WR027101. https://doi.org/10.1029/2020WR027101

Ledesma, J. L., & Futter, M. N. (2017). Gridded climate data products are an alternative to instrumental measurements as inputs to rainfall–runoff models. *Hydrological Processes*, *31*(18), 3283–3293. https://doi.org/10.1002/hyp.11269

Li, B., Sun, T., Tian, F., Tudaji, M., Qin, L., & Ni, G. (2024). Hybrid hydrological modeling for large alpine basins: A semi-distributed approach. *Hydrol. Earth Syst. Sci.*, *28*(20), 4521–4538.

Li, D., Schrön, M., Köhli, M., Bogena, H., Weimar, J., Jiménez Bello, M. A., Han, X., Martínez Gimeno, M. A., Zacharias, S., Vereecken, H., & Hendricks Franssen, H.-J. (2019). Can drip irrigation be scheduled with cosmic-ray neutron sensing? *Vadose Zone J.*, *18*(1), 190053.

Li, S., Griffith, D. A., & Shu, H. (2020). Temperature prediction based on a space–time regression-kriging model. *Journal of Applied Statistics*, *47*(7), 1168–1190. https://doi.org/10.1080/02664763.2019.1671962

Ling, X., Huang, Y., Guo, W., Wang, Y., Chen, C., Qiu, B., Ge, J., Qin, K., Xue, Y., & Peng, J. (2021). Comprehensive evaluation of satellite-based and reanalysis soil moisture products using in situ observations over china. *Hydrology and Earth System Sciences*, *25*, 4209–4229. https://doi.org/10.5194/hess-25-4209-2021

Link, T. E., & Marks, D. G. (1999). Point simulation of seasonal snow cover dynamics beneath boreal forest canopies. *Journal of Geophysical Research: Atmospheres*, *104*(D22), 27841–27857. https://doi.org/10.1029/1998JD200121

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? an evaluation using 3620 flood events. *Hydrology and Earth System Sciences*, *18*(2), 575–594. https://doi.org/10.5194/hess-18-575-2014

Loritz, R., Hrachowitz, M., Neuper, M., & Zehe, E. (2021). The role and value of distributed precipitation data in hydrological models. *Hydrology and Earth System Sciences*, *25*(1), 147–167. https://doi.org/10.5194/hess-25-147-2021

Loucks, D. P. (2023). Hydroinformatics: A review and future outlook. *Cambridge Prisms: Water*, *1*, e10. https://doi.org/10.1017/wat.2023.10

Lucas, M. P., Longman, R. J., Giambelluca, T. W., Frazier, A. G., Mclean, J., Cleveland, S. B., Huang, Y.-F., & Lee, J. (2022). Optimizing Automated Kriging to Improve Spatial Interpolation of Monthly Rainfall over Complex Terrain. *Journal of Hydrometeorology*, *23*(4), 561–572. https://doi.org/10.1175/JHM-D-21-0171.1

Ly, S., Charles, C., & Degré, A. (2013). Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review.

Majone, B., Villa, F., Deidda, R., & Bellin, A. (2016). Impact of climate change and water use policies on hydropower potential in the south-eastern alpine region. *Sci. Total Environ.*, *543*(Pt B), 965–980.

Mallucci, S., Majone, B., & Bellin, A. (2019). Detection and attribution of hydrological changes in a large alpine river basin. *J. Hydrol.*, *575*, 1214–1229.

Mankin, K. R., Mehan, S., Green, T. R., & Barnard, D. M. (2025). Review of gridded climate products and their use in hydrological analyses reveals overlaps, gaps, and the need for a more objective approach to selecting model forcing datasets. *Hydrology and Earth System Sciences*, *29*(1), 85–108. https://doi.org/10.5194/hess-29-85-2025

Maran, L., Schneider, T., & Andreaus, M. (2023). A multi-period analysis of a water management arena in the italian alps, circa 1951–2007: The territorialisation of environmental concerns. *Accounting History*, 10323732231196939.

Marsh, C. B., Lv, Z., Vionnet, V., Harder, P., Spiteri, R. J., & Pomeroy, J. W. (2024). Snowdrift-Permitting Simulations of Seasonal Snowpack Processes Over Large Mountain Extents. *Water Resources Research*, *60*(8), e2023WR036948. https://doi.org/10.1029/2023WR036948

Martin, J. D., & Simpson, T. W. (2003). A study on the use of kriging models to approximate deterministic computer models. *Volume 2: 29th Design Automation Conference, Parts A and B*. https://doi.org/10.1115/DETC2003/DAC-48762

Mascaro, G., Vivoni, E. R., & Méndez-Barroso, L. A. (2015). Hyperresolution hydrologic modeling in a regional watershed and its interpretation using empirical orthogonal functions. *Advances in Water Resources*, *83*, 190–206. https://doi.org/10.1016/j.advwatres.2015.05.023

Mastrotheodoros, T., Pappas, C., Molnar, P., Burlando, P., Manoli, G., Parajka, J., Rigon, R., Szeles, B., Bottazzi, M., Hadjidoukas, P., & Fatichi, S. (2020). More green and less blue water in the alps during warmer summers. *Nat. Clim. Chang.*, *10*(2), 155–161.

Merz, R., Miniussi, A., Basso, S., Petersen, K.-J., & Tarasova, L. (2022). More Complex is Not Necessarily Better in Large-Scale Hydrological Modeling: A Model Complexity Experiment across the Contiguous United States. *Bulletin of the American Meteorological Society*, *103*(8), E1947–E1967. https://doi.org/10.1175/BAMS-D-21-0284.1

MeteoTrentino. (n.d.). Previsioni meteorologiche della provincia di trento [Accessed: 2024-4-3].

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., & Samaniego, L. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, *53*(9), 8020–8040. https://doi.org/10.1002/2017WR020401

Moore, R. J. (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.*, *30*(2), 273–297.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE*, *50*(3), 885–900.

Morlot, M., Rigon, R., & Formetta, G. (2023). Hydrological digital twin model of a large anthropized italian alpine catchment: The adige river basin. *J. Hydrol.*, 130587.

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change*, *7*(3), 214–219. https://doi.org/10.1038/nclimate3225

Musselman, K. N., & Pomeroy, J. W. (2017). Estimation of needleleaf canopy and trunk temperatures and longwave contribution to melting snow. *Journal of Hydrometeorology*, *18*(2), 555–572. https://doi.org/10.1175/JHM-D-16-0111.1

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I–a discussion of principles. *J. Hydrol.*, *10*(3), 282–290.

Nkiaka, E., Nawaz, N. R., & Lovett, J. C. (2018). Effect of single and multi-site calibration techniques on hydrological model performance, parameter estimation and predictive uncertainty: A case study in the logone catchment, lake chad basin. *Stoch. Environ. Res. Risk Assess.*, *32*(6), 1665–1682.

Nocentini, S. (2009). Structure and management of beech (*Fagus sylvatica* l.) forests in italy. *iForest – Biogeosciences and Forestry*, *2*, 143–153. https://doi.org/10.3832/ifor0499-002

Orlowsky, B., & Seneviratne, S. I. (2014). On the spatial representativeness of temporal dynamics at European weather stations. *International Journal of Climatology*, *34*(10), 3154–3160. https://doi.org/10.1002/joc.3903

Parajka, J., & Blöschl, G. (2006). Validation of MODIS snow cover images over austria. *Hydrol. Earth Syst. Sci. Discuss.*, *3*(4), 1569–1601.

Parajka, J., & Blöschl, G. (2008). The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. *J. Hydrol.*, *358*(3), 240–258.

Pasik, A., Gruber, A., Preimesberger, W., De Santis, D., & Dorigo, W. (2023). Uncertainty estimation for a new exponential-filter-based long-term root-zone soil moisture dataset from copernicus climate change service (c3s) surface observations. *Geoscientific Model Development*, *16*, 4957–4976. https://doi.org/10.5194/gmd-16-4957-2023

Paulik, C., Dorigo, W., Wagner, W., & Kidd, R. (2014). Validation of the ASCAT soil water index using in situ data from the International Soil Moisture Network. *International Journal of Applied Earth Observation and Geoinformation*, *30*, 1–8. https://doi.org/10.1016/j.jag.2014.01.007

Pelowitz, D. B., et al. (2005). MCNPXTM user's manual. *Los Alamos National Laboratory, Los Alamos*, *5*, 369.

Peng, J., Tanguy, M., Robinson, E. L., Pinnington, E., Evans, J., Ellis, R., Cooper, E., Hannaford, J., Blyth, E., & Dadson, S. (2021). Estimation and evaluation of high-resolution soil moisture from merged model and earth observation data in great britain. *Remote Sensing of Environment*, *264*, 112610. https://doi.org/10.1016/j.rse.2021.112610

Peterson, A. M., Helgason, W. D., & Ireson, A. M. (2016). Estimating field-scale root zone soil moisture using the cosmic-ray neutron probe. *Hydrol. Earth Syst. Sci.*, *20*(4), 1373–1385.

Poblete, D., Arevalo, J., Nicolis, O., & Figueroa, F. (2020). Optimization of Hydrologic Response Units (HRUs) Using Gridded Meteorological Data and Spatially Varying Parameters. *Water*, *12*(12), 3558. https://doi.org/10.3390/w12123558

Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.*, *100*(2), 81–92.

Puche, M., Troin, M., Fox, D., & Royer-Gaspard, P. (2025). Optimizing Spatial Discretization According to Input Data in the Soil and Water Assessment Tool: A Case Study in a Coastal Mediterranean Watershed. *Water*, *17*(2), 239. https://doi.org/10.3390/w17020239

Ramírez, S. (n.d.). *FastAPI*. Retrieved September 5, 2026, from https://fastapi.tiangolo.com

Rebora, N., & Ferraris, L. (2006). The structure of convective rain cells at mid-latitudes. *Advances in Geosciences*, *7*, 31–35. https://doi.org/10.5194/adgeo-7-31-2006

Reggiani, P., Hassanizadeh, S., Sivapalan, M., & Gray, W. G. (1999). A unifying framework for watershed thermodynamics: Constitutive relationships. *Advances in Water Resources*, *23*(1), 15–39. https://doi.org/10.1016/S0309-1708(99)00005-6

Reggiani, P., Sivapalan, M., & Hassanizadeh, S. M. (2000). Conservation equations governing hillslope responses: Exploring the physical basis of water balance. *Water Resources Research*, *36*(7), 1845–1863. https://doi.org/10.1029/2000WR900066

Reggiani, P., Sivapalan, M., & Majid Hassanizadeh, S. (1998). A unifying framework for watershed thermodynamics: Balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics. *Advances in Water Resources*, *22*(4), 367–398. https://doi.org/10.1016/S0309-1708(98)00012-8

Rigon, R. (2026, August 21). *Six notebooks to represent water budgets from compartment models (with EPNs inside)*. Retrieved September 5, 2026, from https://abouthydrology.blogspot.com/2026/08/five-notebooks-to-represent-water.html

Rigon, R. (n.d.). Abouthydrology — open research notebook of the geoframe group [last accessed 2026].

Rigon, R., Formetta, G., Bancheri, M., Tubini, N., D'Amato, C., David, O., & Massari, C. (2022). Hess opinions: Participatory digital earth twin hydrology systems (darths) for everyone–a blueprint for hydrologists. *Hydrology and Earth System Sciences*, *26*(18), 4773–4800.

Rigon, R., Ghesla, E., Tiso, C., & Cozzini, A. (2006). *The horton machine: A system for dem analysis* [E-book manual]. Retrieved December 30, 2025, from http://www.ing.unitn.it/dica/tools/download/Quaderni/Horton%20manual.pdf

Rolland, C. (2003). Spatial and Seasonal Variations of Air Temperature Lapse Rates in Alpine Regions. *Journal of Climate*, *16*(7), 1032–1046. https://doi.org/10.1175/1520-0442(2003)016<1032:SASVOA>2.0.CO;2

Roudier, P., Andersson, J., Donnelly, C., Feyen, L., Greuell, W., & Ludwig, F. (2016). Projections of future floods and hydrological droughts in europe under a+ 2 c global warming. *Climatic change*, *135*(2), 341–355.

Rutter, A. J., Kershaw, K. A., Robins, P. C., & Morton, A. J. (1971). A predictive model of rainfall interception in forests, 1. derivation of the model from observations in a plantation of corsican pine. *Agric. Meteorol.*, *9*, 367–384.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*, W05523. https://doi.org/10.1029/2008WR007327

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., Eisner, S., Müller Schmied, H., Sutanudjaja, E. H., Warrach-Sagi, K., & Attinger, S. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrol. Earth Syst. Sci.*, *21*(9), 4323–4346.

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, *22*(8), 4583–4591. https://doi.org/10.5194/hess-22-4583-2018

Sarcar, V. (2022). Understanding solid principles. In *Java design patterns* (pp. 3–64). Apress. https://doi.org/10.1007/978-1-4842-7971-7_1

Sartori, G. (Ed.). (2009). *Suoli degli ambienti alpini* [Includes chapters on Alpine soil types and catenae in Trentino]. Museo Tridentino di Scienze Naturali.

Schattan, P., Baroni, G., Oswald, S. E., Schöber, J., Fey, C., Kormann, C., Huttenlau, M., & Achleitner, S. (2017). Continuous monitoring of snowpack dynamics in alpine terrain by aboveground neutron sensing. *Water Resour. Res.*, *53*(5), 3615–3634.

Scheiffele, L. M., Munz, M., Francke, T., Baroni, G., & Oswald, S. E. (2025). Enhancing hectare-scale groundwater recharge estimation by integrating data from cosmic-ray neutron sensing into soil hydrological modeling. *Water Resour. Res.*, *61*(1).

Schrön, M., Köhli, M., Scheiffele, L., Iwema, J., Bogena, H. R., Lv, L., Martini, E., Baroni, G., Rosolem, R., Weimar, J., Mai, J., Cuntz, M., Rebmann, C., Oswald, S. E., Dietrich, P., Schmidt, U., & Zacharias, S. (2017). Improving calibration and validation of cosmic-ray neutron sensors in the light of spatial sensitivity. *Hydrol. Earth Syst. Sci.*, *21*(10), 5009–5030.

Serafin, F., Bancheri, M., David, O., & Rigon, R. (2018). On complex networks representation and computation of hydrological quantities. *International Congress on Environmental Modelling and Software*.

Seyfried, M. S., & Wilcox, B. P. (1995). Scale and the Nature of Spatial Variability: Field Examples Having Implications for Hydrologic Modeling. *Water Resources Research*, *31*(1), 173–184. https://doi.org/10.1029/94WR02025

Sirguey, P., Mathieu, R., Arnaud, Y., Khan, M. M., & Chanussot, J. (2008). Improving MODIS spatial resolution for snow mapping using wavelet fusion and ARSIS concept. *IEEE Geosci. Remote Sens. Lett.*, *5*(1), 78–82.

Sivapalan, M., Blöschl, G., Zhang, L., & Vertessy, R. (2003). Downward approach to hydrological prediction. *Hydrological Processes*, *17*(11), 2101–2111. https://doi.org/10.1002/hyp.1425

Skøien, J. O., & Blöschl, G. (2006). Scale Effects in Estimating the Variogram and Implications for Soil Hydrology. *Vadose Zone Journal*, *5*(1), 153–167. https://doi.org/10.2136/vzj2005.0069

Somers, L. D., & McKenzie, J. M. (2020). A review of groundwater in high mountain environments. *WIREs Water*, *7*, e1475. https://doi.org/10.1002/wat2.1475

Sonatype. (n.d.). The central repository documentation [Accessed: 2026-08-19].

Şorman, A. Ü., Akyürek, Z., Şensoy, A., Şorman, A. A., & Tekeli, A. E. (2007). Commentary on comparison of MODIS snow cover and albedo products with ground observations over the mountainous terrain of turkey. *Hydrol. Earth Syst. Sci.*, *11*(4), 1353–1360.

Stephan, R., Erfurt, M., Terzi, S., Žun, M., Kristan, B., Haslinger, K., & Stahl, K. (2021). An inventory of alpine drought impact reports to explore past droughts in a mountain region. *Natural Hazards and Earth System Sciences*, *21*(8), 2485–2501. https://doi.org/10.5194/nhess-21-2485-2021

Stevanato, L., Baroni, G., Cohen, Y., Fontana, C. L., Gatto, S., Lunardon, M., Marinello, F., Moretto, S., & Morselli, L. (2019). A novel cosmic-ray neutron sensor for soil moisture estimation over large areas. *Collect. FAO Agric.*, *9*(9), 202.

Stucchi, L., Bocchiola, D., Simoni, C., Ambrosini, S. R., Bianchi, A., & Rosso, R. (2023). Future hydropower production under the framework of NextGenerationEU: The case of santa giustina reservoir in italian alps. *Renewable Energy*, *215*, 118980.

Suklitsch, M., Gobiet, A., Truhetz, H., Awan, N. K., Göttel, H., & Jacob, D. (2011). Error characteristics of high resolution regional climate models over the alpine area. *Climate Dynamics*, *37*, 377–390. https://doi.org/10.1007/s00382-010-0848-5

Ten Berge, A. A., Booij, M. J., & Rientjes, T. H. (2025). Robustness of hydrological models for simulating impacts of climate change on high and low streamflow. *Journal of Hydrology*, *662*, 133734. https://doi.org/10.1016/j.jhydrol.2025.133734

The Apache Software Foundation. (2023). Apache commons math 3.6.1.

Tian, J., & Zhang, Y. (2023). Comprehensive validation of seven root zone soil moisture products at 1153 ground sites across china. *International Journal of Digital Earth*, *16*, 4008–4022. https://doi.org/10.1080/17538947.2023.2261902

Todini, E., & Biondi, D. (2017). Calibration, parameter estimation, uncertainty, data assimilation, sensitivity analysis, and validation. In V. P. Singh (Ed.), *Handbook of applied hydrology* (2nd ed., pp. 22-1–22-19). McGraw-Hill Education.

Tong, R., Parajka, J., Komma, J., & Blöschl, G. (2020). Mapping snow cover from daily collection 6 MODIS products over austria. *J. Hydrol.*, *590*, 125548.

Toreti, A., Bavera, D., Magni, D., de Jager, A., Maianti, P., Sedano, F., Thiemig, V., Henin, R., Tarnavsky, E., Seguini, L., Bussay, A., Niemeyer, S., Schaffhauser, T., & Salamon, P. (2026). *Drought in europe: July 2026: GDO analytical report* (JRC147969). Joint Research Centre, European Commission. Luxembourg, Publications Office of the European Union. https://doi.org/10.2760/2686847

Toreti, A., Bavera, D., Acosta Navarro, J., Cammalleri, C., de Jager, A., Di Ciollo, C., Hrast Essenfelder, A., Maetens, W., Magni, D., Masante, D., et al. (2022). Drought in europe august 2022. *Publications Office of the European Union, Luxembourg*, *10*, 264241.

Tovazzi, M. (2013). Bilanci idrici, agenzia provinciale per le risorse idriche e l'energia trento. *Provincia Autonoma di Trento*.

Towner, J., Cloke, H. L., Zsoter, E., Flamig, Z., Hoch, J. M., Bazo, J., Coughlan De Perez, E., & Stephens, E. M. (2019). Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. *Hydrology and Earth System Sciences*, *23*(7), 3057–3080. https://doi.org/10.5194/hess-23-3057-2019

Tuo, Y., Duan, Z., Disse, M., & Chiogna, G. (2016). Evaluation of precipitation input for SWAT modeling in Alpine catchment: A case study in the Adige river basin (Italy). *Science of The Total Environment*, *573*, 66–82. https://doi.org/10.1016/j.scitotenv.2016.08.034

Ufficio Dighe. (n.d.). Ufficio dighe.

van Jaarsveld, B., Wanders, N., Sutanudjaja, E. H., Hoch, J., Droppers, B., Janzing, J., van Beek, R. L. P. H., & Bierkens, M. F. P. (2025). A first attempt to model global hydrology at hyper-resolution. *Earth Syst. Dyn.*, *16*(1), 29–54.

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J.-M., Teng, J., & Viney, N. R. (2011). Conceptual Rainfall–Runoff Model Performance with Different Spatial Rainfall Inputs. *Journal of Hydrometeorology*, *12*(5), 1100–1112. https://doi.org/10.1175/2011JHM1340.1

Vázquez, R. F., Feyen, L., Feyen, J., & Refsgaard, J. C. (2002). Effect of grid size on effective parameters and model performance of the MIKE-SHE code. *Hydrological Processes*, *16*(2), 355–372. https://doi.org/10.1002/hyp.334

Venturelli, M. B., Sartori, G., & Parisi, A. (2012). La carta dei suoli delle valli del noce. *TERRA TRENTINA*, *58*(2), 36–38.

Verdin, A., Funk, C., Rajagopalan, B., & Kleiber, W. (2016). Kriging and Local Polynomial Methods for Blending Satellite-Derived and Gauge Precipitation Estimates to Support Hydrologic Early Warning Systems. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(5), 2552–2562. https://doi.org/10.1109/TGRS.2015.2502956

Verworn, A., & Haberlandt, U. (2011). Spatial interpolation of hourly rainfall – effect of additional information, variogram inference and storm properties. *Hydrology and Earth System Sciences*, *15*(2), 569–584. https://doi.org/10.5194/hess-15-569-2011

Vicente-Serrano, S., Saz-Sánchez, M., & Cuadrat, J. (2003). Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature. *Climate Research*, *24*, 161–180. https://doi.org/10.3354/cr024161

Viviroli, D., Archer, D. R., Buytaert, W., Fowler, H. J., Greenwood, G. B., Hamlet, A. F., Huang, Y., Koboltschnig, G., Litaor, M., López-Moreno, J. I., et al. (2011). Climate change and mountain water resources: Overview and recommendations for research, management and policy. *Hydrology and Earth System Sciences*, *15*(2), 471–504.

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007a). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water resources research*, *43*(7).

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007b). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resour. Res.*, *43*(7).

Vu, M. T., Raghavan, S. V., & Liong, S. Y. (2012). SWAT use of gridded observations for simulating runoff – a Vietnam river basin study. *Hydrology and Earth System Sciences*, *16*(8), 2801–2811. https://doi.org/10.5194/hess-16-2801-2012

Wada, Y., Bierkens, M. F. P., de Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., Konar, M., Liu, J., Müller Schmied, H., Oki, T., Pokhrel, Y., Sivapalan, M., Troy, T. J., van Dijk, A. I. J. M., van Emmerik, T., Van Huijgevoort, M. H. J., Van Lanen, H. A. J., Vörösmarty, C. J., Wanders, N., & Wheater, H. (2017). Human–water interface in hydrological modelling: Current status and future directions. *Hydrol. Earth Syst. Sci.*, *21*(8), 4169–4193.

Wada, Y., de Graaf, I. E. M., & van Beek, L. P. H. (2016). High-resolution modeling of human and climate impacts on global water resources. *Journal of Advances in Modeling Earth Systems*, *8*(2), 735–763. https://doi.org/10.1002/2015MS000618

Wagner, P. D., Duethmann, D., Kiesel, J., Pool, S., Hrachowitz, M., Ceola, S., Herzog, A., Houska, T., Loritz, R., Spieler, D., Staudinger, M., Tarasova, L., Thober, S., Fohrer, N., Tetzlaff, D., Wagener, T., & Guse, B. (2025). The unexploited treasures of hydrological observations beyond streamflow for catchment modeling. *WIREs Water*, *12*, e70018. https://doi.org/10.1002/wat2.70018

Wahbi, A., Heng, L., & Dercon, G. (2018). Cosmic ray neutron sensing: Estimation of agricultural crop biomass water equivalent. *Camb. Int. Law J.*

Wani, J. M., Bertoldi, G., Bozzoli, M., Andreis, D., & Rigon, R. (2025). Integrating snow-water equivalent simulated by a physically based model into a lumped model in an alpine catchment in italy [EGU25-12258, 27 Apr–2 May 2025]. *EGU General Assembly 2025*. https://doi.org/10.5194/egusphere-egu25-12258

Wani, J. M., D'Amato, C., & Rigon, R. (2026). The tricky water energy budget of freezing soil: A thermodynamic framework for understanding phase changes [in review]. *Geotechnical and Geological Engineering*.

Watanabe, H., Hayashi, K., Saeki, T., Maksyutov, S., Nasuno, I., Shimono, Y., Hirose, Y., Takaichi, K., Kanekon, S., Ajiro, M., Matsumoto, Y., & Yokota, T. (2015). Global mapping of greenhouse gases retrieved from GOSAT Level 2 products by using a kriging method. *Int. J. Remote Sens.*, *36*(6), 1509–1528.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. Wiley-Blackwell.

Wilby, R. L., Clifford, N. J., De Luca, P., Harrigan, S., Hillier, J. K., Hodgkins, R., Johnson, M. F., Matthews, T. K. R., Murphy, C., Noone, S. J., Parry, S., Prudhomme, C., Rice, S. P., Slater, L. J., Smith, K. A., & Wood, P. J. (2017). The 'dirty dozen' of freshwater science: Detecting then reconciling hydrological data biases and errors. *WIREs Water*, *4*(3), e1209. https://doi.org/10.1002/wat2.1209

Wildemeersch, S., Goderniaux, P., Orban, P., Brouyère, S., & Dassargues, A. (2014). Assessing the effects of spatial discretization on large-scale flow model performance

and prediction uncertainty. *Journal of Hydrology*, *510*, 10–25. https://doi.org/10.1016/j.jhydrol.2013.12.020

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., Van de Giesen, N., Houser, P., Jaffé, P., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., ... Whitehead, P. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring earth's terrestrial water. *Water Resources Research*, *47*(5), W05301. https://doi.org/10.1029/2010WR010090

Wood, E. F., Sivapalan, M., Beven, K., & Band, L. E. (1988). Effects of spatial variability and scale with implications to hydrologic modeling. *Journal of Hydrology*, *102*, 29–47. https://doi.org/10.1016/0022-1694(88)90090-X

Wu, T., & Li, Y. (2013). Spatial interpolation of temperature in the United States using residual kriging. *Applied Geography*, *44*, 112–120. https://doi.org/10.1016/j.apgeog.2013.07.012

Yamamoto, J. K. (2005). Correcting the Smoothing Effect of Ordinary Kriging Estimates. *Mathematical Geology*, *37*(1), 69–94. https://doi.org/10.1007/s11004-005-8748-7

Zambotti, A. (2020). Un nuovo modello per l'agricoltura in valle di non, fra risparmio idrico e tutela ambientale. *Econ. Agro-Aliment.*, (2), 1–16.

Zaramella, M., Borga, M., Zoccatelli, D., & Carturan, L. (2019). Topmelt 1.0: A topography-based distribution function approach to snowmelt simulation for hydrological modelling at basin scale. *Geoscientific Model Development*, *12*, 5251–5265. https://doi.org/10.5194/gmd-12-5251-2019

Zheng, Y., Coxon, G., Woods, R., Power, D., Rico-Ramirez, M. A., McJannet, D., Rosolem, R., Li, J., & Feng, P. (2024). Evaluation of reanalysis soil moisture products using cosmic ray neutron sensor observations across the globe. *Hydrology and Earth System Sciences*, *28*, 1999–2022. https://doi.org/10.5194/hess-28-1999-2024

Zhu, Z., Wright, D. B., & Yu, G. (2018). The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis. *Water Resources Research*, *54*(11), 8983–8998. https://doi.org/10.1029/2018WR023550

Zierl, B., & Bugmann, H. (2005). Global change impacts on hydrological processes in alpine catchments. *Water Resources Research*, *41*(2).

Zreda, M., Desilets, D., Ferré, T. P. A., & Scott, R. L. (2008). Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophys. Res. Lett.*, *35*(21).

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., & Rosolem, R. (2012). COSMOS: The COsmic-ray soil moisture observing system. *Hydrol. Earth Syst. Sci.*, *16*(11), 4079–4099.