AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR016994

On the use of spatially distributed, time-lapse microgravity surveys to inform hydrological modeling

Sebastiano Piccolroaz[1], Bruno Majone[1], Francesco Palmieri[2], Giorgio Cassiani[3], and Alberto Bellin[1]

[1]Department of Civil, Environmental, and Mechanical Engineering, University of Trento, Trento, Italy, [2]Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Trieste, Italy, [3]Department of Geosciences, University of Padova, Padova, Italy

Key Points:
- Microgravity data constrain subsurface water storage in hydrological models
- Combining streamflow and microgravity data rules out improper conceptual models
- Indirect measures of state variables constrain inversion of hydrological data

Supporting Information:
- Supporting Information S1

Correspondence to:
S. Piccolroaz,
s.piccolroaz@unitn.it

Citation:
Piccolroaz, S., B. Majone, F. Palmieri, G. Cassiani, and A. Bellin (2015), On the use of spatially distributed, time-lapse microgravity surveys to inform hydrological modeling, *Water Resour. Res.*, *51*, 7270–7288, doi:10.1002/2015WR016994.

Received 30 JAN 2015
Accepted 6 AUG 2015
Accepted article online 11 AUG 2015
Published online 7 SEP 2015

Abstract In the last decades significant technological advances together with improved modeling capabilities fostered a rapid development of geophysical monitoring techniques in support of hydrological modeling. Geophysical monitoring offers the attractive possibility to acquire spatially distributed information on state variables. These provide complementary information about the functioning of the hydrological system to that provided by standard hydrological measurements, which are either intrinsically local or the result of a complex spatial averaging process. Soil water content is an example of state variable, which is relatively simple to measure pointwise (locally) but with a vanishing constraining effect on catchment-scale modeling, while streamflow data, the typical hydrological measurement, offer limited possibility to disentangle the controlling processes. The objective of this work is to analyze the advantages offered by coupling traditional hydrological data with unconventional geophysical information in inverse modeling of hydrological systems. In particular, we explored how the use of time-lapse, spatially distributed microgravity measurements may improve the conceptual model identification of a topographically complex Alpine catchment (the Vermigliana catchment, South-Eastern Alps, Italy). The inclusion of microgravity data resulted in a better constraint of the inversion procedure and an improved capability to identify limitations of concurring conceptual models to a level that would be impossible relying only on streamflow data. This allowed for a better identification of model parameters and a more reliable description of the controlling hydrological processes, with a significant reduction of uncertainty in water storage dynamics with respect to the case when only streamflow data are used.

1. Introduction

Building hydrological models able to accurately reproduce hydrological fluxes at a multiplicity of temporal and spatial scales, as more frequently required in applications, is still a serious challenge. The main difficulty is in the inability to adequately characterize the spatial variability of hydrological fluxes, due to the large heterogeneity that characterizes the Earth's subsurface [see e.g., *Rubin*, 2003]. Since this variability cannot be assessed with the detail needed to represent the real, but unknown, spatial distribution of hydrological fluxes, simplifying hypotheses on the spatial distribution of the hydraulic parameters are typically introduced, which makes simulations highly uncertain. Traditionally, this difficulty has been addressed by using simplified conceptual models, which require the definition of lumped parameters with values determined through suitable inference procedures with streamflow as the most used, and often the only, observational variable [see, e.g., *Beven*, 2011].

Since the works of *Duan et al*. [1992] and *Gupta et al*. [1998] the need to consider the inherent multiobjective nature of inference and the importance of assessing, and possibly minimizing, the epistemic error associated to the conceptual model (i.e., the uncertainty that stems from a lack of knowledge about a given phenomenon) emerged clearly together with a growing concern about the representativity of consolidated believes [see e.g., *Kirchner*, 2006]. In particular, *Efstratiadis and Koutsoyiannis* [2010] and *Beven and Westerberg* [2011] indicated that epistemic uncertainty may be related to the noninformativeness (i.e., limited effectiveness) of observational data in conditioning model parameters and structure. Furthermore, *Efstratiadis and Koutsoyiannis* [2010] and *Linde* [2014] evidenced that, in some cases, the additional information provided by the adoption of a multiobjective framework may lead to rejection of an inappropriate conceptual model, which would appear as proper against a single criterion (i.e., the classical comparison of observed and simulated water discharges).

Indeed, several studies revealed the utility of conditioning hydrological models on multiple variables, possibly state variables, in order to reduce uncertainties and improve prediction capabilities of the models. Additional data on hydrological variables other than streamflow, such as evapotranspiration [*Immerzeel and Droogers*, 2008] and snow water equivalent [*Parajka and Blöschl*, 2008], and most importantly on state variables, i.e., soil water content [*Demarty et al*., 2005] and groundwater levels [*Khadam and Kaluarachchi*, 2004], have been shown to be very effective in constraining hydrological models. The growing need to monitor hydrological processes in the shallow subsurface bursted in the last two decades the development of hydro-geophysical techniques as a valuable tool for field-scale investigations capable to reach a level of detail not achievable with traditional techniques (for a recent review, see e.g., *Binley et al*. [2015]). This technological breakthrough prompted a wealth of applications in subsurface hydrology and in monitoring hydrological fluxes in the Earth's critical zone, chiefly at scales ranging from a few meter to hundred meters. Conversely, applications are more difficult and less frequent at the catchment scale, where the trade-off between resolution and volume of investigation still hampers the applications [*Rubin et al*., 1998; *Hubbard and Rubin*, 2006; *Binley et al*., 2010].

Incorporating soft data, in the form of hydrological indexes or empirical criteria, which reflect both modeler expertise and experimental knowledge of the system [see e.g., *Yadav et al*., 2007; *Zhang et al*., 2008; *Wagener and Montanari*, 2011], may provide additional information useful to capture relevant hydrological fluxes that cannot be resolved by calibrating on streamflow only [see e.g., *Seibert and McDonnell*, 2002]. However, given that soft data are only indirectly related to state variables, their inclusion should be considered with caution, because they may mask and reduce the influence of primary data. In fact, soft data and hydrological indexes cannot replace field measurements since effective constraining of parameter estimates to physically plausible ranges can only be achieved when observations on state variables are available [*Gupta et al*., 1999]. In this perspective we explore here how the use of microgravity measurements, as an integral measure of changes in the total water storage (including soil moisture, groundwater and snowpack), may improve the identification of the main processes occurring in a complex Alpine catchment.

The gravitational acceleration at any location on the Earth's surface depends chiefly on the elevation and the mass distribution around (beneath) the observation point. The high sensitivity and repeatability achieved by modern relative gravity meters and the improvements in electronic technology allow the detection of extremely small gravity variations (of the order of a few μGal, where 1 $\mu Gal = 10^{-8}\,\mathrm{m\,s^{-2}}$) [e.g., *Torge*, 1989; *Blakely*, 1995,]. Signals of this magnitude can be linked to phenomena associated with crustal deformations in seismic or volcanic areas [*Rymer*, 1994], exploitation of geothermal or hydrocarbon fields [*Sofyan et al*., 2015], and water storage fluctuations [*Hector et al*., 2013]. In subsurface hydrology, early studies showed that suitably designed microgravity networks can detect small variations of the terrestrial gravity field associated with water table variations at the regional scale [see e.g., *Montgomery*, 1971; *Giorgetti et al*., 1987; *Pool and Eychaner*, 1995; *Pool and Schmidt*, 1987; *Pool*, 2008; *McClymont et al*., 2012]. Other studies [*Lambert and Beaumont*, 1977; *Ferguson et al*., 2007; *Davis et al*., 2008] showed that microgravity measurements can be used to detect seasonal groundwater storage changes, thereby providing valuable information for water resources management. Overall, these studies showed that repeated microgravity measurements can be used to obtain relevant information on temporal changes of water storage in the Earth's subsurface at a scale useful for hydrological applications [*Chapman et al*., 2008; *Creutzfeldt et al*., 2008; *Hasan et al*., 2008; *Creutzfeldt et al*., 2010; *Christiansen et al*., 2011a, 2011b; *Herckenrath et al*., 2012].

In this paper we discuss the value of repeated microgravity measurements in the identification of the conceptual model for hydrological modeling at the catchment scale. To this purpose we assimilate microgravity measurements into a distributed hydrological model of an Alpine catchment, and assess their value in the identification of the conceptual model of the hydrological system. The overall objective is to show the advantage of assimilating data from multiple sources in order to reduce the risk of obtaining good fitting of streamflow data with the wrong conceptual model [see e.g., *Kirchner*, 2006].

The paper is organized as follows: Section 2 presents a description of the study area followed, in section 3, by a short description of the hydrological and gravity models, and parameters identification procedure. The main results on the identification of the correct conceptual model and on the uncertainty assessment of hydrological variables are presented in section 4, and finally conclusions are drawn in section 5.

2. Study Area and Field Campaigns

The study area is the Vermigliana catchment, a typical small Alpine catchment in the South-Eastern Alps, Italy (see Figure 1). At the confluence with the Noce river, a tributary of the Adige river (the second longest river in Italy; for details, see *Chiogna et al.* [2015]), the drainage area is of 104 km^2, which reduces to 79 km^2 at the upstream streamflow gauging station located near the village of Vermiglio. Catchment's morphology is complex (Figure 2) with elevation ranging from 950 m a.s.l. to 3558 m a.s.l., and several steep lateral valleys and ridges. The upper part of the catchment is characterized by snow and ice-capped mountains with steep rocky slopes, whereas in the lower part the valley shows a U-shaped profile with steep forested slopes, and a flat and narrow valley bottom. The geology is composed mainly of paragneiss, ortogneiss and schists, and large alluvial fans occupy the valley floor at the confluence of the main tributaries [*Dal Piaz et al.*, 2007]. A few stratigraphic surveys are available for the main valley (information accessible via the WebGIS service of the Province of Trento, http://www.territorio.provincia.tn.it), which indicate a paragneiss bedrock at about 5–20 m depth.

Climate is continental-alpine, with a mean annual temperature of 2.9°C and a mean annual precipitation of 1440 mm/yr that, in winter, falls as snow. Streamflow data recorded at the Vermiglio gauging station (see Figure 2) show a typical nivo-glacial regime with significant seasonal patterns [*Chiogna et al.*, 2014]: long term observed streamflow averages are 4.6 m^3/s and 1.2 m^3/s during warm (from April to September) and cold (from October to March) seasons, respectively. Nine meteorological stations, recording precipitation and air temperature, are located within and in close proximity of the catchment. These stations are operated by the meteorological office of the Province of Trento (http://www.meteotrentino.it) and data are available with daily resolution since 1921, which increases to hourly frequency starting from late 70s.

2.1. Microgravity Surveys

In May 2009 a microgravity network composed of 53 nodes was established within the zone shown in Figure 1, which includes the Vermigliana catchment. Six microgravity surveys were conducted in the period June 2009 to May 2011, with the objective of monitoring spatial and temporal variations of subsurface water storage. An outcrop of unfractured metamorphic rock, rigid and stable over time, was identified within the catchment and established as reference station. Due to its geological and morphological characteristics, this formation is expected to undergo negligible seasonal water storage fluctuations, thus allowing to disregard temporal gravity changes at the reference location. Measurements were taken using a LaCoste-Romberg D-018 gravimeter equipped with a ZLS (Zero Length Spring) feedback. Before and after each field campaign, the instrumental calibration function was verified against a calibration line established between the absolute stations of Brunico and Novacella (Province of Bolzano), observing no significant deviations.

Data acquisition was conducted according to the classical step method, which consists in moving from station to station following a multiple occupation sequence, so that each station is occupied at least three times (1-2-1-2-3-2-3-4-3-. . .). In this way, the Lacoste-Romberg drift behavior could be tracked with accuracy and could be modeled with a polynomial fitting [e.g., *Chapman et al.*, 2008]. The stations, in conformity of multiples of 3 and 5, have been tied to form a strongly structured network where each point is connected to at least 4 other stations. The precise locations of the microgravity stations have been geo-referenced by means of the differential GPS technique, and successive positioning has been eased by the installation of sheet metal plates. Each field campaign lasted for about 1 week, therefore variations of water storage at smaller temporal scales cannot be resolved.

The gravity data were tide corrected using the *Tamura* [1987] procedure, and the drift was removed through polynomial interpolation, under the hypothesis that the signal is composed by the superposition of a deterministic drift and an erratic component [*Blecha*, 2002]. Notice that barometric corrections were not applied, since changes of barometric pressure were always less than 3 hPa between two successive occupations of the same station. After correction, the data set was adjusted through the least-square procedure that is discussed in details in Appendix A.

3. Methods

3.1. Hydrological Model

In the present work, a modified version of the distributed modeling system GEOTRANSF (for a description of the hydrological kernel, see *Majone et al.* [2015]) (see also A. Bellin, et al., GEOTRANSF: A continuous

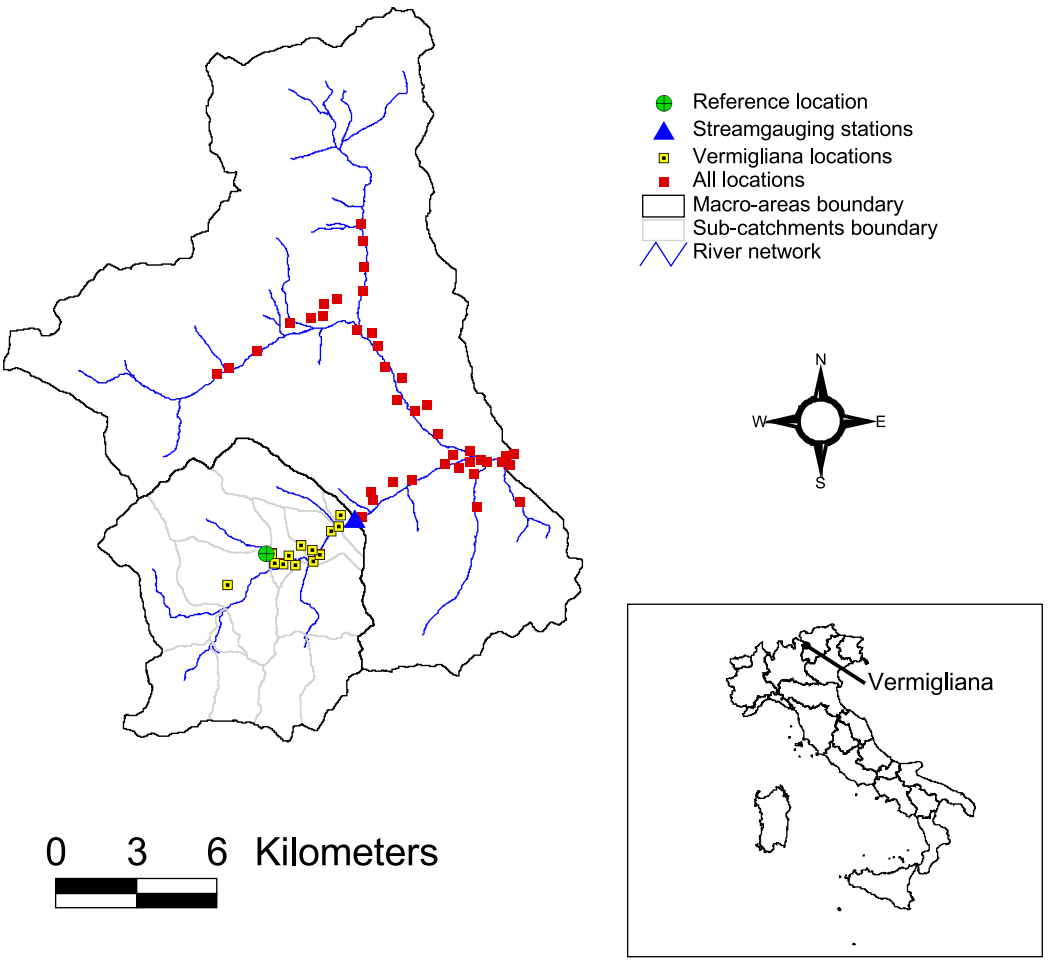

Figure 1. Map of the catchment monitored during the microgravity campaigns with indicated the locations belonging to the microgravity network. The boundaries of the Vermigliana catchment are drawn, together with the subdivision in 13 subcatchments (grey lines), and the position of the stream gauging station (blue triangle). The inset shows the location of the Vermigliana catchment within the Italian territory.

coupled hydrological and water resources management model, submitted to *Environmental Modelling & Software*, 2015) was adopted for simulating the hydrological cycle of the Vermigliana catchment. GEOTRANSF is built around two geomorphological units: the subcatchment and the channel. The former unit identifies a portion of the catchment where hillslope processes dominate. The latter is the base element composing the river network that connects the selected subcatchments to the control section, where streamflow is simulated. A mass balance equation is written for each subcatchment under the assumption that streamflow at its outlet depends nonlinearly upon water storage [*Kirchner*, 2009; *Majone et al*., 2010].

In order to make the model sensitive to water redistribution within the drainage area, the catchment has been subdivided into two interconnected elements, the hillslope and the valley bottom, roughly corresponding to convex and concave areas, respectively (see Figure 3a). In this way, the conceptual model identifies two storing mechanisms related to identifiable morphological characteristics [*Winter*, 2001; *Savenije*, 2010; *Gharari et al*., 2011; *Nobre et al*., 2011; *Gao et al*., 2014]. Hillslope is identified as the area with high average slope where runoff is dominated by shallow subsurface flow and groundwater storage is small to negligible. On the other hand, the valley bottom area, which in the work of *Savenije* [2010] is included in the broader "wetland" landscape classification, is the area where water storage occurs chiefly as groundwater. Note that, in general, a subcatchment is not necessarily formed of both landscape units (see the schematic in Figure 3).

In the present study, landscape classification was performed by using the following mixed criterion. The cells belonging to the hillslope unit were first identified as those characterized by slope and elevation higher than

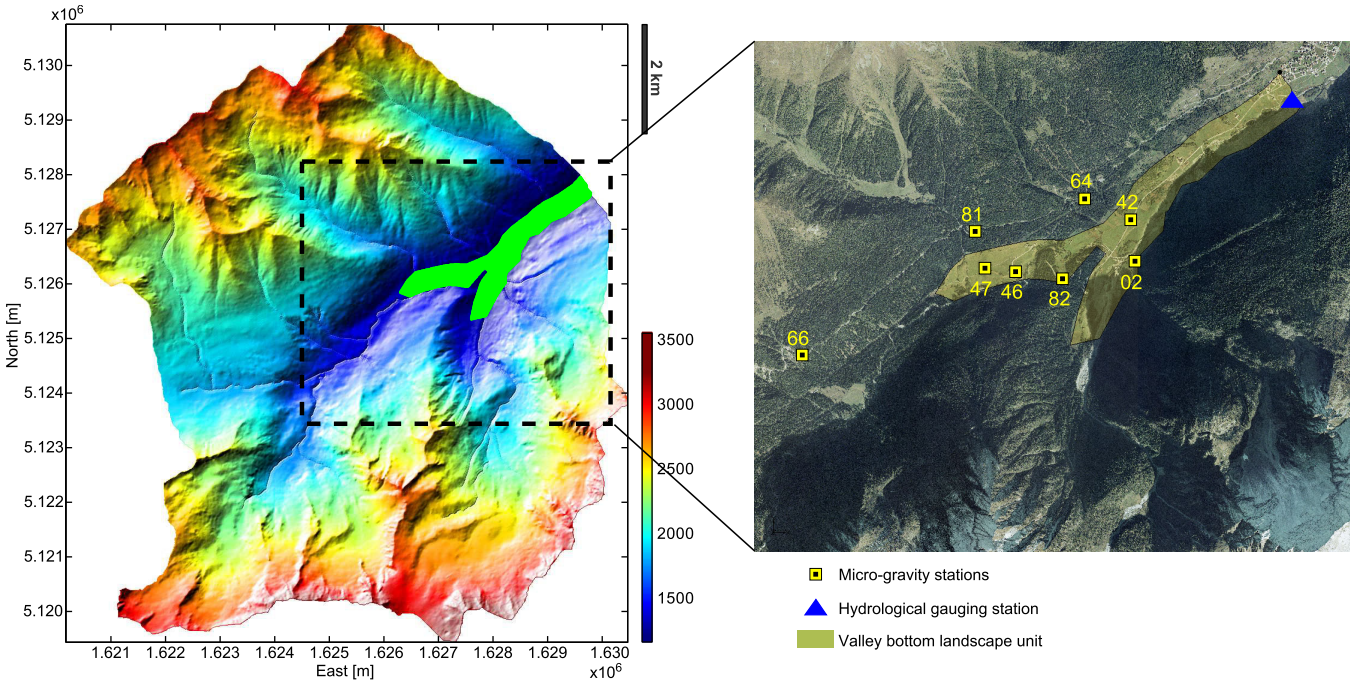

Figure 2. (left) DEM of the Vermigliana catchment at 10 m resolution, position of microgravity and stream gauging stations used in this work, and (right) delimitation of the bottom valley landscape unit.

30% and 1400 m a.s.l., respectively. This delimitation was successively refined by visual inspection of orthophoto images, and by comparison of the resulting delimitation with geophysical surveys (i.e., GPR and ERT) that identified the depth of the bedrock along a number of cross sections of the valley bottoms.

The valley bottom unit identified with this procedure is shown in Figure 2, and has a surface area of about 1.8 km^2, corresponding to 2.2% of the whole catchment area. An analysis of the influence that the selection of the two landscape units may have on model results will be presented in section 4.2.3.

The setup of the conceptual hydrological model follows from the landscape classification presented above, thus separating between processes occurring at hillslope and valley bottom areas. The structure of the model is outlined in Figure 3b with the list of calibration parameters shown in Table 1, while water balance equations are detailed in the supporting information. Besides this general version of the model, hereafter referred to as M1, a second simpler model, M2, has been considered, in which the separation between the two landscape units is removed. This simpler model serves as a reference in order to highlight the utility of complementary data (i.e., microgravity time-lapse measurements) in improving the conceptual model, as will be discussed in section 4.1.

3.2. Gravity Model

Changes in the local gravity field due to temporal variations of water mass stored within the catchment have been modeled by coupling the hydrological model described in section 3.1 to a forward gravity model. The catchment volume has been divided into a number of prisms with a squared base of size equal to the DEM resolution, (10 m in the present work, Figure 2), and the same number of vertical layers as in the hydrological model, including the snowpack (Figure 3b). Following *Hasan et al.* [2008], we assumed that within each prism a change in aquifer thickness, dictated by the hydrological model, results in a vertical translation of water table, which is assumed parallel to topographic surface. The total effect of water storage variation is derived as the sum of gravity changes associated to all layers of each prism, thus accounting for the effect of the three main water storage components: soil moisture in the rhizosphere and in the vadose zone, groundwater, and snowpack.

For a generic prismatic subvolume, the gravity changes caused by water storage variations (Δg, $(m\,s^{-2})$) can be expressed through the closed-form solution of Newton's Law proposed by *Nagy* [1966], which reads as follows:

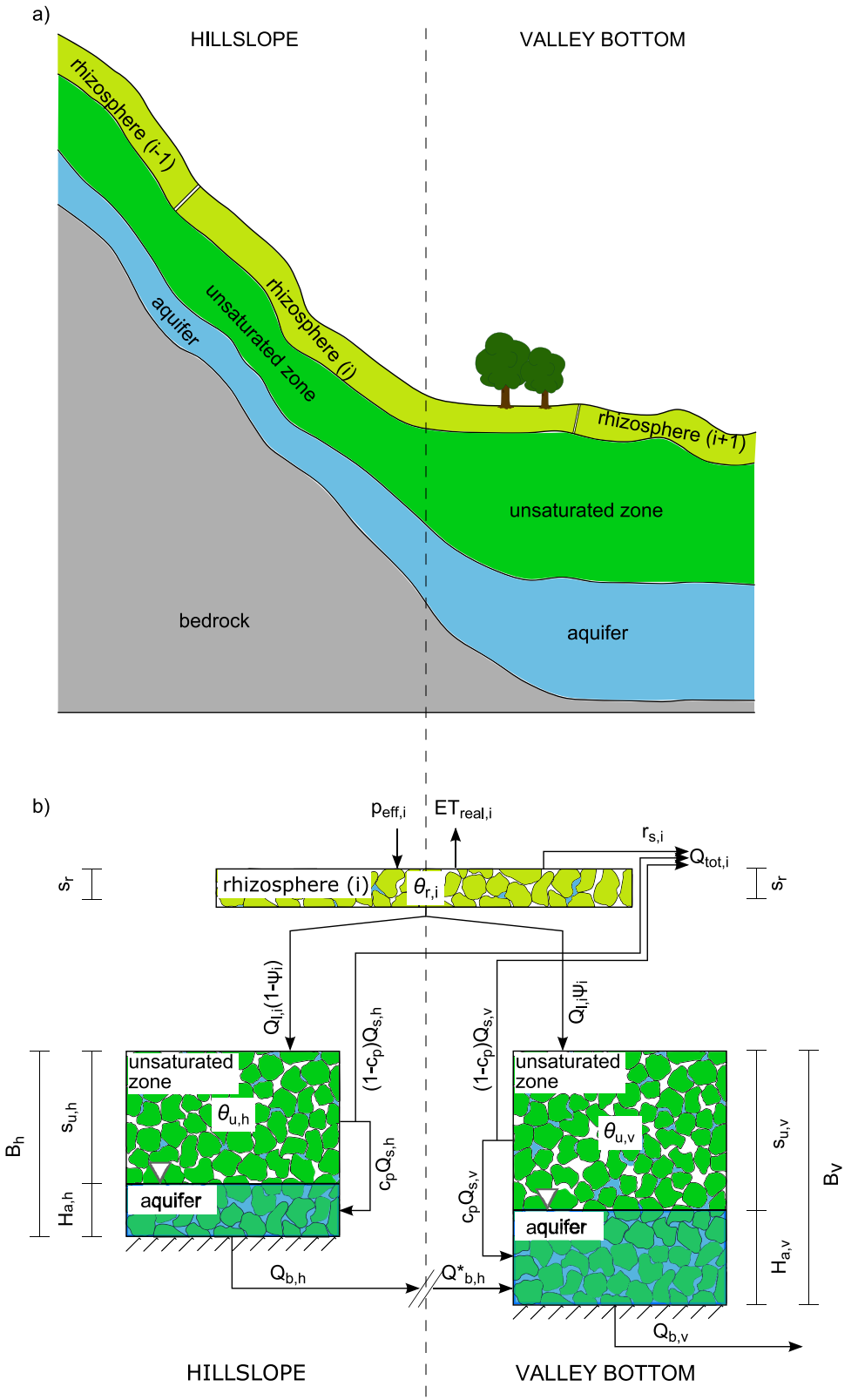

Figure 3. Schematic representation of the hydrological conceptual model M1: (a) subdivision of the catchment into hillslope and valley bottom landscape units, and into three vertical layers (rhizosphere, vadose zone and aquifer); (b) water fluxes in the modified coupled surface-groundwater version of GEOTRANSF (description of each flux term is reported in the supporting information).

Table 1. List of the Calibration Parameters With Their Range of Variation[a]

Model Component	Symbol	Unit	Range of Variation	M1	M2[b]	M1($\alpha=0$)	M1($\alpha=0.5$)
Snow	T_s	($°C^{-1}$)	−1.00 to 1.40	✓	✓	1.07	0.89
Snow	T_{sm}	($°C^{-1}$)	1.40–4.00	✓	✓	1.44	1.62
Snow	c_m	($mm°C^{-1}\ d^{-1}$)	2.00–8.00	✓	✓	2.78	2.95
Evapotranspiration	θ_{lim}		5.00E-2–4.00E-1	✓	✓	3.51E-1	9.50E-2
Rainfall excess	c_s		1.00–5.00	✓	✓	2.25	2.26
Rainfall excess	c_a		2.00E-1–1.00	✓	✓	5.49E-1	5.86E-1
Subsurface flow	q_{max}	($m^3\ s^{-1}\ km^{-2}\ m^{-1}$)	1.00E-1–1.10	✓	✓	5.79E-1	4.90E-1
Subsurface flow	m		1.00E-3–6.00E-2	✓	✓	3.23E-2	3.31E-2
Subsurface flow	c_p		2.00E-2–5.00E-1	✓	✓	1.33E-1	6.35E-2
Groundwater	$k_{b,h}$	(d)	1.00E1.5–1.00E3.5	✓	✓	329	222
Groundwater	$k_{b,v}$	(d)	100–225	✓	–	124	191
Groundwater	B_h^a	(m)	5.00E-1–10.0	✓	–	1.41	9.38E-1
Groundwater	B_v	(m)	10.0–30.0	✓	✓	10.7	16.6
Groundwater	w_h	(m)	1.00E-4–3.16-3	✓	–	4.09E-4	2.07E-4

[a]Symbols "✓" and "–" indicate whether the parameter is included or not in the model. The last two columns report the optimal parameters of the model M1 for the objective functions (2) with $\alpha = 0$ and $\alpha = 0.5$, respectively. The parameters of the model M1 are the following: T_s and T_{sm}: temperature thresholds for snow precipitation and snow melting; c_m: snow melting factor; θ_{lim}: field capacity of the rhizosphere; c_s and c_a: parameters of the rainfall excess model; q_{max} and m: parameters of the nonlinear reservoir mimicking the dynamics of the unsaturated zone; c_p: partition coefficient of subsurface flow; $k_{b,h}$ and $k_{b,v}$: mean residence time of hillslope and valley bottom aquifers; B_h and B_v: thickness of hillslope and valley bottom lower reservoirs; w_h: groundwater flow velocity in the hillslope. A detailed description of the model, including the above parameters, is provided in the supporting information.

[b]In model configuration M2, B_h has been allowed to range from 5.00E-1 to 30.0, which is the entire range of variation of B as assumed in model configuration M1.

$$\Delta g = \gamma \Delta \rho \left| \left| x \log (y+d) + y \log (x+d) - z \arctan \left(\frac{x\,y}{z\,d} \right) \right|_{x_2}^{x_1} \right|_{y_2}^{y_1} \Bigg|_{z_2}^{z_1}, \quad (1)$$

where γ is the universal gravitational constant ($6.67 \times 10^{-11}\ Nm^2kg^{-2}$), $\Delta\rho$ is the density change in the subvolume at time t with respect to a reference time t_{ref}, $d = \sqrt{x^2 + y^2 + z^2}$, and x_i, y_i and z_i, $i = 1, 2$ are the corner coordinates of the prism in a Cartesian reference system, whose origin coincides with the location of the microgravity meter (see Figure 4b).

Equation (1) takes different forms depending on which hydrological state variable is considered [e.g., *Leirião et al.*, 2009]. Gravity variations in the rhizosphere are only caused by changes in water content $\Delta\theta_r = \theta_r(t) - \theta_r(t_{ref})$ between times t and t_{ref}, such that $\Delta\rho = \rho_w \Delta\theta_r$, where ρ_w is the density of water. Conversely, gravity changes due to snow are driven by the evolution of snowpack thickness due to accumulation and melting, with $z_{1,s} = z(t_{ref})$ and $z_{2,s} = z(t)$ being the vertical coordinate of the snowpack surface at time t_{ref} and t, respectively. Here the effect of changes in snow density due to metamorphism is neglected, therefore $\Delta\rho$ has not the strict meaning of a density change, but it is assumed constant and equal to 150 $kg\,m^{-3}$ (i.e., the density of damp new snow).

In the subsurface, both changes in water content and geometry should be considered. The thickness of the two subsurface layers below the rhizosphere changes in time in dependence of the water table position (see Figure 4b). The saturated layer is bounded from above by the water table, such that $z_{1,a} = z(t_{ref})$ and $z_{2,a} = z(t)$ are the depths of the water table at time t_{ref} and t, respectively. The overlying unsaturated layer extends from the water table to the lower limit of the rhizosphere, $z_{2,u}$, which is fixed in time. In particular, $z_{1,u}$ is given by the minimum between water table depth at times t_{ref} and t: $z_{1,u} = \min(z_{1,a}, z_{2,a})$. In addition, $\Delta\rho = \rho_w[\theta_u(t) - \theta_u(t_{ref})]$ in the unsaturated zone, and $\Delta\rho = \rho_w v_a$ in the portion of soil affected by water table variation (i.e., between $z_{1,a}$ and $z_{2,a}$, see Figure 4b). Note that the portion of the aquifer not affected by water table fluctuation between time t_{ref} and t does not contribute to gravity variations, since $\Delta\rho = 0$ and geometry does not change. Similar considerations apply also to the snowpack. The term v_a represents the volumetric air content which, in this specific case, is the volumetric fraction of soil that is depleted/filled as a consequence of water table rise or fall between times t_{ref} and t, and is defined as $v_a = \phi - \theta_u(t)$ for a declining water table, and $v_a = \phi - \theta_u(t_{ref})$ for a rising water table.

To alleviate the burden of computing the geometrical term, the above forward gravity model has been applied by using equation (1) only for the elements that satisfy the condition $d_c^2/(\Delta x^2 + \Delta y^2 + \Delta z^2) < 4$ [*Leirião et al.*, 2009], where d_c is the distance between the cell's center and the gravimeter, and Δx, Δy, Δz,

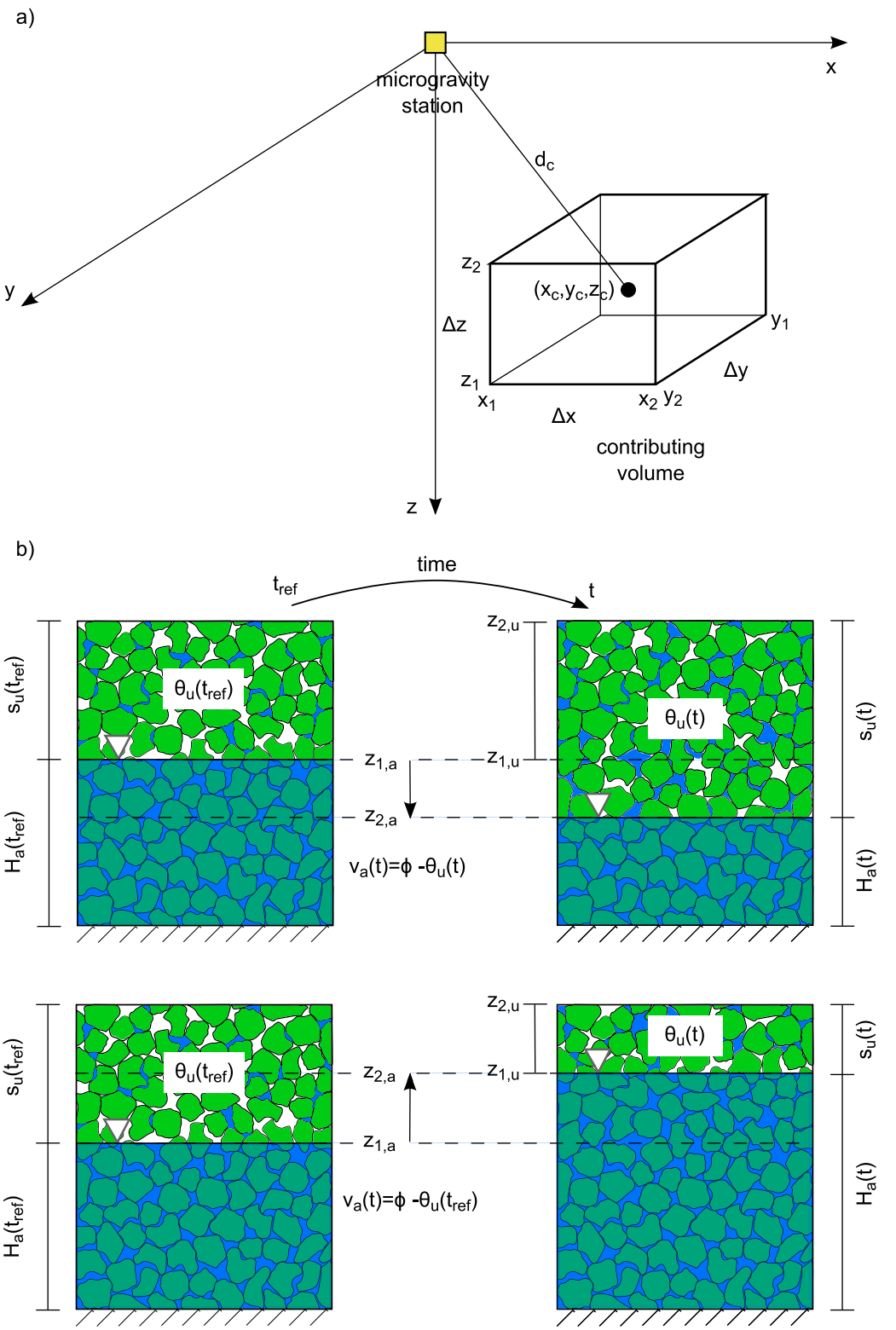

Figure 4. Gravity model: (a) generic prismatic subvolume and Cartesian coordinate system of the gravity model; (b) schematic of the two subsurface layers below the rhizosphere with indicated the reference quantities used to evaluate the gravity change associated to falling and rising water table conditions.

are the side lengths of the prisms (Figure 4a). In the remaining of the computational domain the approximated solution proposed by *MacMillan* [1958] has been used in place of equation (1). To further reduce the execution time, the code has been parallelized by using OpenMP directives. Finally, on the basis of a preliminary sensitivity analysis, the computational domain was limited to the area within a radius of 5 km from the gravimeter's location, which leads to an error of the order of $1E-3$ μGal, therefore largely acceptable for the purposes of the present study.

3.3. Calibration Approach

The main objective of the present work is to evaluate the benefit in terms of model identifiability of including microgravity measurements as additional observational data supplementing streamflow measurements. This is in line with the tenet that additional data are more effective in constraining model parameters if they are of different type with respect to the currently used data [*McLaughlin and Townley*, 1996; *Madsen et al.*, 2010]. In this context, optimal parameters of the hydrological model have been inferred within a multiobjective framework, in which the efficiency index KGE, proposed by *Gupta et al.* [2009], has been used to define two distinct metrics of model performance: one providing a measure of model capabilities in simulating observed streamflow (KGE_{hydro}), and the other in simulating observed gravity change (KGE_{gravi}). Following a standard approach, these two metrics have been aggregated into a single objective function KGE_{tot}:

$$KGE_{tot} = (1-\alpha)KGE_{hydro} + \alpha KGE_{gravi}\,, \tag{2}$$

with the weighting factor α ranging from 0 to 1. The two metrics in equation (2) are defined as follows:

$$KGE = 1-\sqrt{(\beta-1)^2+(\gamma-1)^2+(r-1)^2}\,, \tag{3}$$

where $\beta=\mu_s/\mu_o$ is the ratio between the means of the simulated (μ_s) and observed (μ_o) time series, $\gamma=\sigma_s/\sigma_o$ is the ratio between the corresponding standard deviations, and $r=cov_{so}/(\sigma_s\sigma_o)$ is the linear correlation coefficient between observations and simulations (with cov_{so} being the covariance between simulated and observed time series). The optimization of KGE_{tot} has been performed by implementing the Particle Swarm Optimization (PSO) algorithm, which is an evolutionary and self-adaptive search optimization technique proposed by *Kennedy and Eberhart* [1995] for applications in information technology, and successively applied to a variety of different fields, including hydrology [e.g., *Gill et al.*, 2006; *Castagna and Bellin*, 2009; *Majone et al.*, 2010].

Table 1 lists the calibration parameters of the hydrological model with their range of variation. We note that the hypercube search space explored by PSO has 14 and 11 dimensions in models M1 and M2, respectively. The range of variation of calibration parameters has been chosen on the basis of previous hydrological studies conducted with GEOTRANSF, for the parameters θ_{lim}, c_s, and c_p [*Majone et al.*, 2012; *Graveline et al.*, 2014; *Majone et al.*, 2015], or have been derived from the WebGIS service of the Province of Trento (e.g., ϕ, B_h, B_v). We applied the same prior ranges of variability for all parameters that the two models have in common, with the only exception of the total thickness of the lower reservoir (B). In M2 this parameter is defined only for one landscape unit, and has been allowed to span the entire range of variation of B as defined in model configuration M1.

4. Results

Numerical simulations have been performed with a daily time step in the period from 1 January 2000 to 30 May 2011, when streamflow data were available at the Vermiglio gauging station. The time window considered for model calibration extents from 1 January 2009 to 30 May 2011, thus including the time interval covered by microgravity surveys (i.e., from 5 June 2009 to 30 Ma 2011, see section 2.1). Since the daily simulation time step is larger than the residence time of water particles in the river network, flow routing has been neglected in the streamflow generation process.

Between 15 December 2009 and 4 February 2010 streamflow measurements showed significantly high discharge values, of the same order of magnitude as those observed in summer during snow melting. Since these fluctuations occurred in the absence of relevant rainfalls and at low temperatures (mean, minimum and maximum temperatures at Mezzana, 905 m a.s.l. were -4.1°C, -11.5°C and 2.4°C, respectively), it is

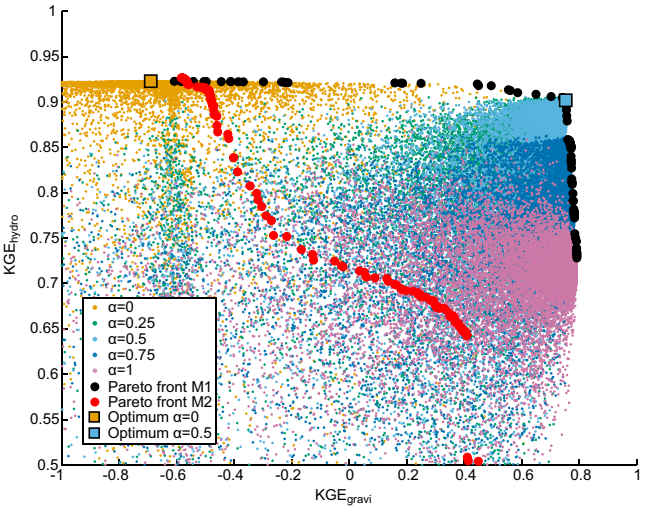

Figure 5. Comparison between Pareto fronts obtained for model configurations M1 and M2. The dominated space of solutions for model M1, and the best solutions associated to $\alpha = 0$ and $\alpha = 0.5$ are also shown.

likely that they are anomalies in the data recorded by the ultrasonic device measuring water level, and for this reason this period has been excluded from calibration.

For each microgravity station, temporal variations of the gravity field have been evaluated following the procedure described in section 3.2, assuming the date of the second survey (conducted between 22 September 22 and 2 October 2009, see Table A1) as reference time for simulated snow and aquifer depths, and water content in the rhizosphere and unsaturated zones. This campaign is used as reference because precipitation was minimal in the period preceding, as well as during, the survey, and consequently no significant variations of storage are expected (i.e., almost stationary conditions). In addition, the forward gravity response was not corrected by removing the variation at the reference station, because it has been neglected according to the careful selection of its location (see section 2.1).

We selected 8 stations out of the 53 gravimetric stations belonging to the monitoring network, all within the Vermigliana catchment. Some other stations are located in this catchment, but too close to its boundaries to be considered totally independent on the water storage changes in the nearby areas.

4.1. Analysis of Pareto Fronts

Ideally, the aim of a multiobjective optimization would be to identify a unique model and a unique set of optimal parameters that simultaneously optimize all metrics quantifying the consistency of the model to observational data. In practice, trade-offs between the metrics exist, mainly because epistemic and parametric uncertainties combine with measurement errors to make possible the situation in which the model is not optimal for some, if not all, metrics [*Schoups et al*., 2005a]. An effective way to deal with this identification problem is the analysis of the Pareto front (or nondominated front), whose shape helps in identifying model structural errors and highlighting model limitations [e.g., *Gupta et al*., 1998; *Schoups et al*., 2005a,2005b; *Efstratiadis and Koutsoyiannis*, 2010; *Rye et al*., 2012]. In this section, the compatibility of the models M1 and M2 to the observational data is evaluated by comparing their Pareto fronts.

Following the conventional aggregation approach [see e.g., *Madsen*, 2000, 2003], optimization is performed separately five times with the weight α of the objective function (2) that varies between 0 (only streamflow data are considered) and 1 (only microgravity data are considered) with steps of 0.25. For each α value, the PSO algorithm has been run with a population of 100 particles and for 500 iterations. These settings have been shown to provide a satisfactory approximation of the continuous Pareto space.

Figure 5 shows the comparison between the Pareto fronts obtained with the models M1 and M2. The entire set of dominated solutions of the model M1 (i.e., solutions not belonging to the Pareto front) is also presented. Substantial differences can be observed in the shape of the two Pareto fronts, suggesting different capabilities of the two models to use the information contained in the two data sets. The Pareto front associated to M1 shows two branches almost parallel to the metric axes, which intersect at the optimal value for $\alpha = 0.5$, where $KGE_{hydro} = 0.90$ and $KGE_{gravi} = 0.75$. In this situation, KGE_{gravi} can be improved with a slight reduction of KGE_{hydro}, implying little trade-off between the two metrics, which is symptomatic of an appropriate model conceptualization. Conversely, the shape of the Pareto front belonging to model M2 calls for a significant trade-off between the two metrics. With this model, in fact, it is impossible to identify a parameter set resulting in values of KGE_{hydro} and KGE_{gravi} both close to their maximum value. In other words, with M2 high values of KGE_{hydro} can be obtained only with a set of parameters leading to negative KGE_{gravi}, which

implies that good streamflow reproduction is obtained with a model that offers a very poor description of groundwater storage. To obtain an acceptable, though suboptimal, simulation of groundwater dynamics (i.e., $KGE_{gravi} = 0.41$) a dramatic reduction of KGE_{hydro} to 0.64 should be accepted. This trade-off between the two metrics, together with the low values of KGE_{gravi} generally achieved with the model M2, is indicative of a poor conceptualization of this model. In addition, both Akaike (AIC) and Bayesian (BIC) information criteria [*Laio et al*., 2009], which penalize models with more parameters, resulted in favor of M1 for both hydrological and microgravity metrics obtained with α=0.5, as expected from the shape of the two Pareto front shown in Figure 5.

A first consideration that can be drawn from this exercise is the advantage of combining data from multiple sources in a multiobjective approach to identify improper or incomplete conceptualizations. The inadequacy of model M2 would not have been detected using only streamflow data, coherently with the fact that parameter calibration in a single-objective framework has often enough degrees of freedom to mask untenable model conceptualizations [*Schoups et al*., 2005a]. Also a comparative evaluation of model appropriateness would have been impossible, since in this case both models show high performance, with no significant differences in the efficiency indexes: optimal value of KGE_{hydro} equal to 0.92 and 0.93 for model M1 and M2, respectively. Conversely, the inclusion of microgravity data resulted in a better constraint of the inversion procedure and in an improved capability to identify limitations of concurring conceptual models, to a level that would be impossible relying only on streamflow data. Furthermore, we note how microgravity data are able to describe a large part of the catchment dynamics (much more than streamflow data alone), as suggested by the fact that for $\alpha = 1$ model M1 provides good *KGE* values in reproducing both streamflow and gravity changes (0.73 and 0.79, respectively). Finally, the results also indicate that a landscape classification as introduced in model M1 is an essential requirement in order to obtain a robust and trustworthy partition of water fluxes between compartments (i.e., hillslope and valley bottom areas) of a topographically complex alpine catchment.

4.2. Uncertainty Assessment

In the light of the above considerations, the following analyses have been carried out only with the conceptual model M1, focusing the discussion on the quantification of the improvements in modeling capabilities that can be achieved by combining streamflow data with microgravity data.

In order to address this issue, we estimated and compared parameters and model output uncertainty for two significant cases: 1) only streamflow data are available for model calibration ($\alpha = 0$), and 2) both streamflow and microgravity data are used for parameters inference with a balanced aggregated objective function (α=0.5). The latter case has been chosen because it explores the central part of the Pareto front where the break point is positioned (Figure 5), which is the combination that optimizes both efficiency metrics.

Among the solutions obtained during the exploration of the parameters space, only those with KGE_{tot} differing less the 1% from the maximum value have been retained for the uncertainty analysis, as suggested in previous studies [e.g., *Madsen*, 2000]. In the present study, the maximum values of KGE_{tot} are 0.92 and 0.83 for $\alpha = 0$ and $\alpha = 0.5$, respectively (see the squared symbols in Figure 5), resulting in 6695 and 3269 parameter sets. This is preferable than selecting a fixed percentage of the total number of simulations, because the adopted searching algorithm shows an intrinsic tendency to cluster around the optimal solution.

As a consequence of the clustering effect, the a posteriori pdf of the parameters, computed by weighting each parameter's value by its performance index [se e.g., *Beven*, 2011; *Vrugt et al*., 2009], is accurate around the optimal point, but biased away from it, because of the exclusion of unlikely, yet possible, parameter combinations [*Rubin et al*., 2010]. This is in some sense suboptimal, but still acceptable since a model that behaves poorly in proximity of the optimal point is not expected to provide a better approximation away from it, and our analysis is intended to assess the benefit of including microgravity data rather than identifying the interval of confidence with the highest possible level of accuracy. In the light of these considerations, we decided to focus the analysis on the 100% confidence bands resulting from the retained solutions, and we introduce the average distance between the upper and lower limits of the confidence interval of a predicted state variable or calibrated parameter, $\bar{d}$, as a metric of uncertainty. As a side comment, we note that despite a standard operating procedure to perform uncertainty analysis of predictions is still debated in the hydrological literature, the procedure adopted here is acceptable to provide a quantification of the

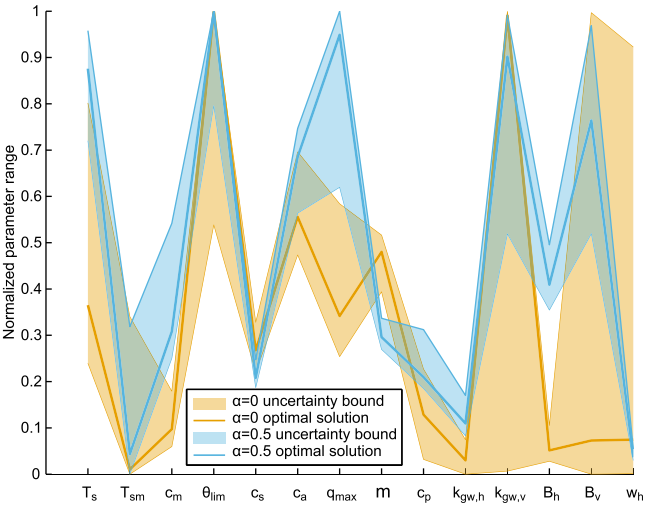

Figure 6. Normalized range the parameters (shaded areas) for both $\alpha = 0$ and $\alpha = 0.5$, obtained by retaining the parameter sets with an aggregated objective function differing less than 1% from the optimal solutions. Solid lines indicate the optimal parameter sets for both $\alpha = 0$ and $\alpha = 0.5$. See Table 1 and supporting information for the description of parameters.

advantages obtainable by combining different sources of information in a multiobjective framework, which is the main scope of this work.

4.2.1. Model Parameters

Figure 6 shows the range of variability of the parameters associated to the retained solutions for both $\alpha = 0$ and $\alpha = 0.5$, together with the corresponding two optimal parameter sets (i.e., the two sets corresponding to the solutions depicted by square symbols in Figure 5). The values of the parameters are normalized with respect to their range (see Table 1), such as to make them ranging from 0 to 1. The dimensions of the searching domain have been fixed by means of preliminary simulations conducted with a reduced number of particles and increasing parameter ranges, such as to minimize the probability of excluding from the searching domain combinations of parameters leading to behavioral solutions [*Beven and Binley*, 1992]. In both $\alpha = 0$ and $\alpha = 0.5$ cases, confidence bands are generally well distributed between 0 and 1, indicating a proper choice of their prior range of variation, although for a few parameters optimality was located close to the boundary of the searching domain.

As shown in Figure 6 the majority of the parameters have a confidence band that is tighter for $\alpha = 0.5$ than for $\alpha = 0$, as indicated by $\bar{d}$, which is about 45% smaller in the former case, on average. This effect is particularly pronounced for parameters controlling hydrological fluxes in the valley bottom unit ($k_{gw,v}$, Bv, and wh) thus supporting the conclusion that for $\alpha = 0.5$ the model is constrained by microgravity measurements to reproduce correctly water storage changes, which are significant in this portion of the catchment. As a consequence, these parameters are better constrained than by streamflow only and vary within a narrower range. Figure 6 shows also that for most of the parameters the two cases lead to non overlapping confidence bands, and to different values at the optimum. This is in accordance with the concept of multiobjective equivalence of parameter sets along a Pareto Front [*Gupta et al.*, 1998], and allows to identify to what degree each parameter is sensitive to different aggregated objective functions. Significant is the case of parameters that control water exchanges between unsaturated and saturated zones (e.g., qmax, Bh, and Bv), as well as snow accumulation and melting processes (i.e., Ts and cm).

4.2.2. Water Storage Dynamics

In the present section we focus on the dynamics of the total water volume stored within the catchment, which includes both subsurface and snowpack storage. The analysis has been performed on water storage variations evaluated relatively to the simulated conditions of the second microgravity survey, centered on 27 September 2009. Optimal solutions and the associated confidence bands, obtained as envelop of all the retained solutions, are shown in Figures 7a and 7b for hillslope and landscape units. Both cases $\alpha = 0$ and $\alpha = 0.5$ have been considered.

In the hillslope unit, the simulated water volume variations are insensitive to α. Water storage shows a remarkable seasonal pattern, characterized by a period of accumulation between December and May, followed by a rapid depletion phase between May and late July due to snow melting. During the rest of the year, the total water volume stored in the system is at its annual minimum and does not show significant variations, except for minor fluctuations due to summer and autumn precipitation events. When streamflow data are coupled with microgravity measurements, the amplitude of the confidence band reduces on average by about 26%, with respect to the case in which only streamflow data are used. Furthermore, in both cases the confidence interval is significantly smaller (from 3 to 5 times, for $\alpha = 0$ and $\alpha = 0.5$, respectively) than the standard deviation of water volume fluctuations associated with the best solution, suggesting a good correspondence between the conceptual model M1 and the investigated processes.

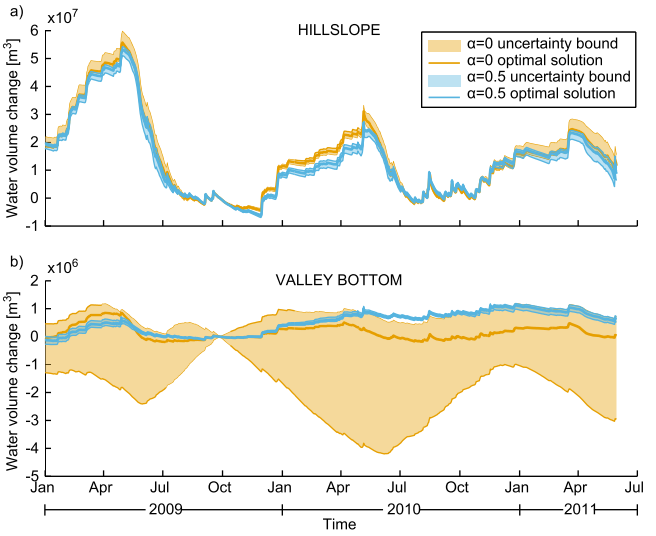

Figure 7. Confidence bands (shaded areas) and optimal solutions (solid lines) of modeled changes of water volume stored in the system for $\alpha = 0$ and $\alpha = 0.5$, and for (a) hillslope and (b) valley bottom units. Confidence bands refer to all solutions for which the aggregated objective functions differ by less than 1% from the optimal solutions. Changes are calculated relatively to 27 September 2009.

Contrary to the hillslope, groundwater dynamics in the valley bottom area is very sensitive to α. The first important difference between $\alpha = 0$ and α=0.5, is the amplitude of the confidence bands. For α=0.5 the dispersion around the optimal value is limited (the average amplitude of the confidence interval is one order of magnitude smaller than the standard deviation of the best solution). On the other hand, for $\alpha = 0$ the confidence band is remarkably wider (more than 4 times the standard deviation of the best solution), suggesting limited identifiability of model parameters, when inversion is based only on streamflow data. The average amplitude of the confidence band declines by 95%, from $\alpha = 0$ to $\alpha = 0.5$, which is a much stronger reduction than for the hillslope unit. The reason for this substantial difference is in the role of valley bottom in the long term water storage, which is properly captured only when gravity data are added, i.e., when additional information on storage dynamics are provided.

Water storage in the valley bottom, which varies seasonally for $\alpha = 0$, shows a tendency to accumulations for $\alpha = 0.5$. This is a consequence of the combined affect of a particularly wet summer in 2010 and the high storage capacity of the thick alluvial deposits of the valley bottom area that compensates for the depletion of surface accumulation occurring during snow melting. This compensation did not occur in the hillslope, since this portion of the catchment is dominated by a rapid transfer of water toward the valley bottom. Similar considerations can be drawn from Figure 8, which shows simulated water table fluctuations, both for hillslope and valley bottom units. In these cases, the narrowing of the confidence band achieved with α= 0.5 with respect to $\alpha = 0$ is about 39% for the hillslope and 93% for the valley bottom. Similarly to what observed in Figure 7, water accumulation in the valley bottom aquifer is observed only for α=0.5, while water table dynamics in the hillslope unit are similar for both $\alpha = 0$ and α=0.5. Finally, we note that for $\alpha = 0$ in some solutions the water table invaded the rhizosphere, eventually exfiltrating (as indicated by the straight horizontal segments in Figure 8), while for α=0.5 this condition did not occurred, coherently with what observed in the study area.

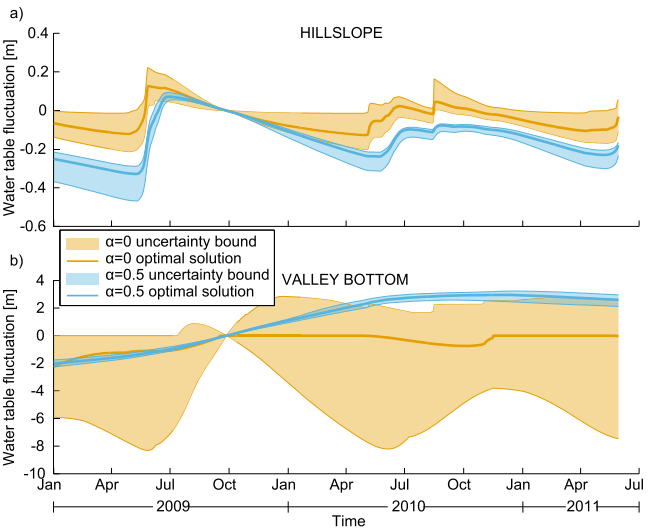

Figure 8. Confidence bands (shaded areas) and optimal solutions (solid lines) of modeled water table fluctuations for $\alpha = 0$ and $\alpha = 0.5$, and for (a) hillslope and (b) valley bottom units. Confidence bands refer to all solutions for which the aggregated objective functions differ by less than 1% from the optimal solutions. Fluctuations are calculated relatively to 27 September 2009.

4.2.3. Comparison Between Observations and Simulations

Figure 9 shows the comparison between observed and simulated streamflow time series, together with the confidence bands for both $\alpha = 0$ and α=0.5. In both cases the model reproduces with good accuracy the observed streamflow, with KGE_{hydro} being slightly higher for $\alpha = 0$

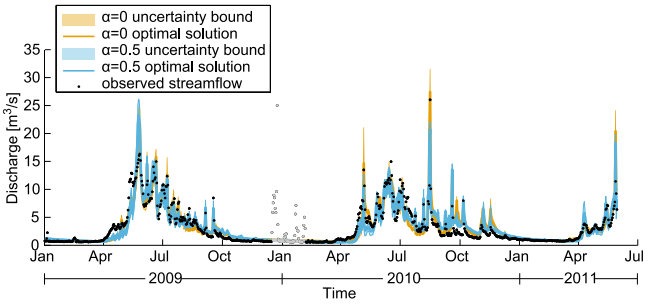

Figure 9. Optimal solutions (solid lines) and associate confidence bands of modeled streamflow for $\alpha = 0$ and $\alpha = 0.5$. Confidence bands refer to all solutions for which the aggregated objective functions differ by less than 1% from the optimal solutions.

than for α=0.5 (i.e., 0.92 against 0.90, respectively). Consistently with water storage dynamics analyzed in section 4.2.2, the amplitude $\bar{d}$ of the averaged confidence band reduces from 0.95 m^3/s to 0.78 m^3/s when microgravity data are included ($\alpha = 0.5$). This leads to the conclusion that microgravity data not only allow for a better representation of subsurface storage, which is important when modeling is extended to contaminant transport, but they also reduce uncertainty of streamflow.

Measured and modeled variations of the gravity field are compared in Figure 10a. The comparison is presented only for the optimal solution obtained with α=0.5 and for all gravity stations used in this work: three located in the hillslope area (station IDs 64, 66 and 81), and the remaining in the valley bottom unit (station IDs 02, 42, 46, 47 and 82, see Figure 2). Valley bottom stations show a good agreement between simulated and observed data, while hillslope stations show larger deviations. In a few stations the model underestimates gravity variations. We attribute this mismatch to the simplification, dictated by the lack of detailed stratigraphic data, of assuming the water table moving parallel to the ground surface. This does not allow for a perfect reproduction of the complex relationship between storage and gravity variations, which is unavoidable in a topographically complex alpine catchment like the Vermigliana.

Seasonal variations are well reproduced except for station 82, located at the border between the hillslope and the valley bottom units. Its position at the interface between two compartments may be the possible cause of this mismatch. Moreover, measurements collected during the first microgravity campaign show appreciable variability, most probably due to the rather intense precipitations occurred during the survey. For this reason, the second, and not the first, survey was assumed as reference, as already discussed in section 2.1.

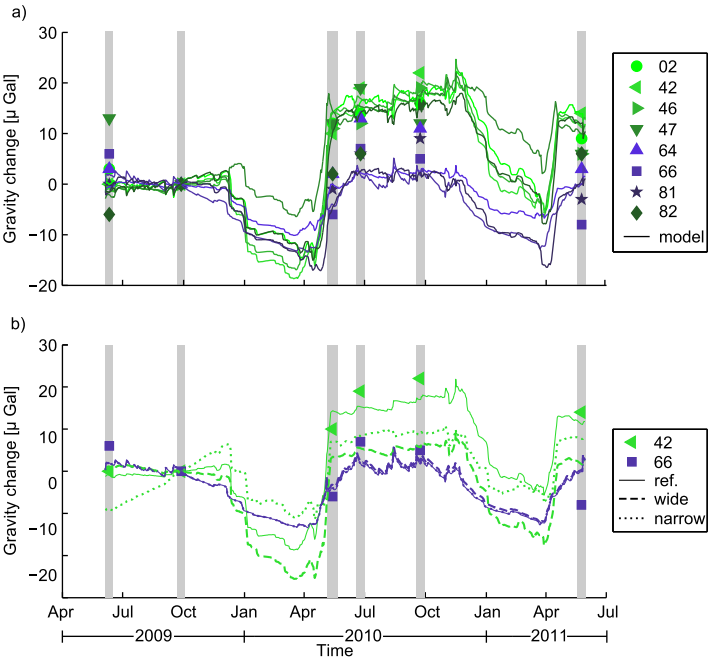

Figure 10. (a) Comparison between measured (symbols) and modeled (lines) gravity variations for gravity stations 02, 42, 46, 47, and 82 in the valley bottom, and 64, 66 and 81 in the hillslope area. Gravity changes are calculated with reference to the second field survey, and vertical bars indicate the survey periods. (b) Modeled gravity change for stations 42 and 66 as obtained for the reference configuration and for wider and narrower valley bottom areas.

We analyzed also the influence that the geometrical definition of the two landscape units may have on modeling of microgravity changes. To this purpose, we ran the model using the optimal set of parameters ($\alpha = 0.5$) but with different extents of the valley bottom unit. In particular, we considered both a narrower (i.e., 0.8 km^2, corresponding to the 1.1% of the catchment area) and a wider (i.e., 4.2 km^2, corresponding to the 5.3% of the catchment area) configuration, identified by imposing thresholds for slope and elevation equal to 20% and 1250 m, and 50% and 1550 m, respectively. Results are presented in Figure 10b for stations 42 and 66, which are located within the valley bottom and hillslope units, respectively, for the three examined configurations. Only the station 42 is sensitive to the extension of the valley bottom, while no significant differences are observed for station 66. Furthermore, a noticeable deterioration of KGE_{gravi} (evaluated for all considered stations) is observed, which decreases from 0.75 for the reference configuration, down to −0.01 and −0.23 for the wider and narrower configurations, respectively. Conversely, KGE_{hydro} is equal to 0.90 in all three cases, thereby suggesting that streamflow is insensitive to the extension of the valley bottom unit. Notice that a poor representation of storage processes in the valley bottom may not have a detectable impact on the modeled streamflow, as in the present case (see section 4.1), but most likely it will impair our ability to correctly model the amount of exploitable groundwater resources and solute transport at the catchment scale. This analysis provides evidence of model sensitivity to a suitable conceptualization of the catchment into different landscape units, and of the added value of microgravity measurements to guide the definition of the conceptual model and the identification of areas where water storage processes are significant.

5. Conclusions

The added value of coupling streamflow data and time-lapse microgravity measurements in constraining the simulation of water storage (the key state variable of hydrological modeling at the catchment scale) has been evaluated in the Vermigliana catchment, a topographically complex Alpine watershed. The joint inversion of hydrological and microgravity data has been performed by coupling a distributed hydrological model with a gravity model. The conceptualization of the hydrological model considered separately hillslope and valley bottom areas, with the latter providing much of the long term (i.e., seasonal) storage of subsurface water. In a simplified version of the model, called M2 to distinguish it from the full model indicated by M1, the separation between hillslope and valley bottom was removed. Both models were used in a multiobjective calibration procedure in which the Pareto fronts, obtained by changing the reciprocal weights of the objective functions of hydrological and microgravity data, evidenced significant limitations of model M2 with respect to model M1, thereby ruling out the simplest conceptual model. More importantly, the inversion showed that the inadequacy of model M2 is appreciable only when microgravity data are introduced. In fact, these two models provided very similar results when calibrated only with streamflow data, while microgravity changes were adequately captured only with model M1. This analysis provided a clear evidence of the value of microgravity data in improving the conceptualization of catchment functioning.

The value of microgravity data in reducing uncertainty was also evaluated. Confidence bands of the main hydrological variables (i.e., the total water volume changes, water table fluctuations, and streamflow) were calculated for model M1 considering results obtained relying on streamflow data only, and using the combination of streamflow and microgravity measurements. Results demonstrated a significant dispersion in the first case, while tighter confidence bands were achieved after adding microgravity data, with the reduction of uncertainty being larger in the valley bottom (e.g., more than 90% for water storage) than in the hillslope. This difference can be explained with the good constraining effect of microgravity measurements in the valley bottom area, due to its larger water storage capability and the prevalence of microgravity measurement points in this part of the catchment (due to logistic reasons). This overall improvement was obtained without significant deterioration of streamflow simulations, which also benefited from a narrowing of confidence bands. Moreover, parameter identifiability increased, as suggested by reductions of the confidence intervals associated to model parameters up to 45%.

The major practical limitation of the proposed approach lies in the difficulties in collecting accurate and spatially distributed gravity measurements in a complex Alpine catchment with steep slopes and deep valleys. Despite these difficulties, the analysis presented here showed that microgravity data provide significant

added value to hydrological modeling, resulting in an effective constraint of the conceptual model, and a reduction of both parameter and model uncertainty.

Appendix A: Microgravity Network Adjustment

The observation equation for a reading difference between two consecutive stations assumes the following form

$$g_i - g_j - \Delta g_{ij} = v_{ij}\,, \tag{A1}$$

where g_i and g_j are the unknown gravity values at stations i and j, Δg_{ij} is the measured gravity difference already corrected for Earth tide and instrumental drift, and v_{ij} is the (unknown) observational error.

Introducing approximate gravity values g_i^0 and g_j^0, equation (A1) can be rewritten as

$$(g_i^0 + x_i) - (g_j^0 + x_j) - \Delta g_{ij} = v_{ij}\,, \tag{A2}$$

where x_i and x_j are the unknown corrections of the trial values.

We consider now that equation (A1) holds also for the approximate gravity values with the observational error v_{ij} replaced by the error l_{ij} before the adjustment

$$g_i^0 - g_j^0 - \Delta g_{ij} = l_{ij}\,, \tag{A3}$$

therefore equation (A2) can be reduced to

$$x_i - x_j + l_{ij} = v_{ij}\,, \tag{A4}$$

or in matrix notation

$$AX + L = V\,, \tag{A5}$$

where A is the design ($n \times r$) matrix (n is the number of observation equations and r is the number of unknowns), X is the ($r \times 1$) vector of the unknowns to be estimated (i.e., the corrections to be applied to the approximate gravity values), L is the ($n \times 1$) vector of the errors before the adjustment, and V is the ($n \times 1$) vector of the residuals.

The optimal vector of the corrections X can be obtained by applying the least squares minimization concept to equation (A5), which leads to

$$X = -(A^T A)^{-1} A^T L\,. \tag{A6}$$

The a posteriori variance of unit weight referred to the whole network is given by

$$\sigma_0^2 = \frac{V^T V}{n - r}\,, \tag{A7}$$

whereas the variances of the adjusted gravity values for the i-th (i = 1,r) station is

$$\sigma_{ii}^2 = \sigma_0^2 (A^T A)^{-1} I\,, \tag{A8}$$

where I is the identity matrix. The results of the network adjustments for each survey are summarized in Table A1 (notice that this operation has been carried out on the entire network, i.e., 53 stations). No

Table A1. Survey Schedule and Network Adjustment Summary

Survey	Measurement Period	Number of Gravity Ties	Number of Equations	σ_0 (μGal)	min σ_{ii} (μGal)	max σ_{ii} (μGal)
1	05/06/2009 to 16/06/2009	139	403	4.9	1.7	2.6
2	22/09/2009 to 02/10/2009	137	401	4.7	1.6	2.5
3	06/05/2010 to 21/05/2010	133	400	4.5	1.5	2.4
4	19/06/2010 to 30/06/2010	133	400	4.4	1.5	2.4
5	17/09/2010 to 28/09/2010	133	401	4.6	1.6	2.5
6	18/05/2011 to 30/05/2011	134	402	4.1	1.4	2.2

significant differences can be noticed in the values of σ_0 and σ_{ii} calculated for the 6 field campaigns, indicating a good reliability of the measurements during the entire survey period. Moreover results suggest that, when considering the worst case (i.e., $\sigma_{ii}=2.6$ μGal), gravity variations larger than 5.2 μGal have a statistical significance of 95 %.

Acknowledgments
The authors are grateful to the Autonomous Province of Trento for providing hydrological data, and GPS and topographic benchmarks. Meteorological data can be downloaded from http://www.meteotrentino.it/, streamflow data are available upon request (uff.dighe@provincia.tn.it), while the microgravity data are available from the authors (contact: Francesco Palmieri fpalmieri@ogs.trieste.it). This research has been partially funded by European Union FP7 Collaborative Research Project GLOBAQUA (Managing the effects of multiple stressors on aquatic ecosystems under water scarcity, grant 603629) and by the Italian Ministry of Public Instruction, University and Research, Grant Number 2010JHF437 through the project PRIN 2010–2011 (Innovative methods for water resources management under hydroclimatic uncertainty scenarios, prot. 2010JHF437).

References

Beven, K. J. (2011), *Rainfall-Runoff Modelling: The Primer*, 2nd ed., John Wiley, Chichester, U. K., doi:10.1002/9781119951001.fmatter.
Beven, K. J., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.
Beven, K. J., and I. Westerberg (2011), On red herrings and real herrings: Disinformation and information in hydrological inference, *Hydrol. Processes*, *25*(10), 1676–1680, doi:10.1002/hyp.7963.
Binley, A., G. Cassiani, and R. Deiana (2010), Hydrogeophysics: Opportunities and challenges, *Bol. Geofis. Teor. Appl.*, *51*(4), 267–284.
Binley, A., S. S. Hubbard, J. A. Huisman, A. Revil, D. A. Robinson, K. Singha, and L. D. Slater (2015), The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales, *Water Resour. Res.*, *51*, 3837–3866, doi:10.1002/2015WR017016.
Blakely, R. J. (1995), *Potential Theory in Gravity and Magnetic Applications*, Cambridge Univ. Press, Cambridge, U. K.
Blecha, V. (2002), Corrections for time variations of gravity in microgravity surveys, paper presented at SEG International Exposition and 72nd Annual Meeting, Salt Lake City, Utah, October 6–11, 2002.
Castagna, M., and A. Bellin (2009), A Bayesian approach for inversion of hydraulic tomographic data, *Water Resour. Res.*, *45*, W04410, doi:10.1029/2008WR007078.
Chapman, D. S., E. Sahm, and P. Gettings (2008), Monitoring aquifer recharge using repeated high-precision gravity measurements: A pilot study in South Weber, Utah, *Geophysics*, *73*(6), WA83–WA93, doi:10.1190/1.2992507.
Chiogna, G., E. Santoni, F. Camin, A. Tonon, B. Majone, A. Trenti, and A. Bellin (2014), Stable isotope characterization of the vermigliana catchment, *J. Hydrol.*, *509*, 295–305, doi:10.1016/j.jhydrol.2013.11.052.
Chiogna, G., B. Majone, K. Cano Paoli, E. Diamantini, E. Stella, S. Mallucci, V. Lencioni, F. Zandonai, and A. Bellin (2015), A review of hydrological and chemical stressors in the Adige catchment and its ecological status, *Sci. Total Environ.*, doi:10.1016/j.scitotenv.2015.06.149, in press.
Christiansen, L., P. J. Binning, D. Rosbjerg, O. B. Andersen, and P. Bauer-Gottwein (2011a), Using time-lapse gravity for groundwater model calibration: An application to alluvial aquifer storage, *Water Resour. Res.*, *47*, W06503, doi:10.1029/2010WR009859.
Christiansen, L., S. Lund, O. B. Andersen, P. J. Binning, D. Rosbjerg, and P. Bauer-Gottwein (2011b), Measuring gravity change caused by water storage variations: Performance assessment under controlled conditions, *J. Hydrol.*, *402*(1–2), 60–70.
Creutzfeldt, B., A. Güntner, T. Klügel, and H. Wziontek (2008), Simulating the influence of water storage changes on the superconducting gravimeter of the geodetic observatory Wettzell, Germany, *Geophysics*, *73*(6), WA95–WA104, doi:10.1190/1.2992508.
Creutzfeldt, B., A. Güntner, S. Vorogushyn, and B. Merz (2010), The benefits of gravimeter observations for modelling water storage changes at the field scale, *Hydrol. Earth Syst. Sci.*, *14*, 1715–1730, doi:10.5194/hess-14-1715-2010.
Dal Piaz, G. V., et al. (2007), Note Illustrative della Carta Geologica d'Italia alla scala 1:50.000 Foglio 042 Malè, technical report, Prov. Auton. di Trento, Serv. Geol., Trento.
Davis, K., Y. Li, and B. M. (2008), Time-lapse gravity monitoring: A systematic 4d approach with application to aquifer storage and recovery, *Geophysics*, *73*(6), WA61–WA69, doi:10.1190/1.2987376.
Demarty, J., C. Ottlé, I. Braud, A. Olioso, J. P. Frangi, H. V. Gupta, and L. A. Bastidas (2005), Constraining a physically based soil-vegetation-atmosphere transfer model with surface water content and thermal infrared brightness temperature measurements using a multiobjective approach, *Water Resour. Res.*, *41*, W01011, doi:10.1029/2004WR003695.
Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall- runoff models, *Water Resour. Res.*, *28*(4), 1015–1031, doi:10.1029/91WR02985.
Efstratiadis, A., and D. Koutsoyiannis (2010), One decade of multi-objective calibration approaches in hydrological modelling: A review, *Hydrolog. Sci. J.*, *55*(1), 58–78, doi:10.1080/02626660903526292.
Ferguson, J. F., T. Chen, J. L. Brady, C. L. Aiken, and J. E. Seibert (2007), The 4d microgravity method for waterflood surveillance ii-gravity measurements for the Prudhoe bay reservoir, Alaska, *Geophysics*, *72*(2), I33–I43, doi:10.1190/1.2435473.
Gao, H., M. Hrachowitz, F. Fenicia, S. Gharari, and H. Savenije (2014), Testing the realism of a topography-driven model (flex-topo) in the nested catchments of the upper Heihe, China, *Hydrol. Earth Syst. Sci.*, *18*(5), 1895–1915, doi:10.5194/hess-18-1895-2014.
Gharari, S., M. Hrachowitz, F. Fenicia, and H. Savenije (2011), Hydrological landscape classification: Investigating the performance of hand based landscape classifications in a central European meso-scale catchment, *Hydrol. Earth Syst. Sci.*, *15*(11), 3275–3291, doi:10.5194/hess-15-3275-2011.
Gill, M. K., Y. H. Kaheil, A. Khalil, M. McKee, and L. Bastidas (2006), Multiobjective particle swarm optimization for parameter estimation in hydrology, *Water Resour. Res.*, *42*, W07417, doi:10.1029/2005WR004528.
Giorgetti, F., I. Marson, and F. Palmieri (1987), Ground water table changes detected by microgravity variations: An attempt to calculate the effective porosity, *Geol. Appl. Idrogeologia*, *22*, 17–24.
Graveline, N., B. Majone, R. V. Duiden, and E. Ansink (2014), Hydro-economic modeling of water scarcity under global change: An application to the Gállego river basin (Spain), *Reg. Environ. Change*, *14*(1), 119–132, doi:10.1007/s10113-013-0472-0.
Gupta, H., L. Bastidas, S. Sorooshian, W. Shuttleworth, and Z. Yang (1999), Parameter estimation of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,491–19,503.
Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751–763, doi:10.1029/97WR03495.
Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*(1-2), 80–91, doi:10.1016/j.jhydrol.2009.08.003.
Hasan, S., P. A. Troch, P. W. Bogaart, and C. Kroner (2008), Evaluating catchment-scale hydrological modeling by means of terrestrial gravity observations, *Water Resour. Res.*, *44*, W08416, doi:10.1029/2007WR006321.
Hector, B., L. Séguis, J. Hinderer, M. Descloitres, J.-M. Vouillamoz, M. Wubda, J.-P. Boy, B. Luck, and N. Moigne (2013), Gravity effect of water storage changes in a weathered hard-rock aquifer in West Africa: Results from joint absolute gravity, hydrological monitoring and geophysical prospection, *Geophys. J. Int.*, *194*(2), 737–750, doi:10.1093/gji/ggt146.

Herckenrath, D., E. Auken, L. Christiansen, A. A. Behroozmand, and P. Bauer-Gottwein (2012), Coupled hydrogeophysical inversion using time-lapse magnetic resonance sounding and time-lapse gravity data for hydraulic aquifer testing: Will it work in practice?, *Water Resour. Res.*, *48*, W01539, doi:10.1029/2011WR010411.

Hubbard, S. S., and Y. Rubin (2006), Chapter 14. Hydrogeological characterization using geophysical methods, in *The Handbook of Groundwater Engineering*, 2nd ed., edited by J. W. Delleur, pp. 1–52, CRC Press, Boca Raton, Fla.

Immerzeel, W., and P. Droogers (2008), Calibration of a distributed hydrological model based on satellite evapotranspiration, *J. Hydrol.*, *349*(3-4), 411–424, doi:10.1016/j.jhydrol.2007.11.017.

Kennedy, J., and R. Eberhart (1995), Particle swarm optimization, in *Proceedings of IEEE International Conference on Neural Networks*, Institute of Electrical & Electronics Engineering, University of Western Australia, Perth, Western Australia, pp. 1942–1948, doi:10.1109/ICNN.1995.488968.

Khadam, I., and J. Kaluarachchi (2004), Use of soft information to describe the relative uncertainty of calibration data in hydrologic models, *Water Resour. Res.*, *40*, W11505, doi:10.1029/2003WR002939.

Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, *42*, W03S04, doi:10.1029/2005WR004362.

Kirchner, J. W. (2009), Catchment as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, *45*, W02429, doi:10.1029/2008WR006912.

Laio, F., G. Di Baldassarre, and A. Montanari (2009), Model selection techniques for the frequency analysis of hydrological extremes, *Water Resour. Res.*, *45*, W07416, doi:10.1029/2007WR006666.

Lambert, A., and C. Beaumont (1977), Nano variations in gravity due to seasonal groundwater movements: Implications for the gravitational detection of tectonic movements, *J. Geophys. Res.*, *82*(2), 297–306, doi:10.1029/JB082i002p00297.

Leirião, S., X. He, L. Christiansen, O. B. Andersen, and P. Bauer-Gottwein (2009), Calculation of the temporal gravity variation from spatially variable water storage change in soils and aquifers, *J. Hydrol.*, *365*(3-4), 302–309, doi:10.1016/j.jhydrol.2008.11.040.

Linde, N. (2014), Falsification and corroboration of conceptual hydrological models using geophysical data, *WIREs Water*, *1*(2), 151–171, doi:10.1002/wat2.1011.

MacMillan, W. D. (1958), *Theoretical Mechanics: The Theory of the Potential*, Dover, N. Y.

Madsen, H. (2000), Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, *235*(3-4), 276 – 288, doi:10.1016/S0022-1694(00)00279-1.

Madsen, H. (2003), Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, *Adv. Water Resour.*, *26*(2), 205–216, doi:10.1016/S0309-1708(02)00092-1.

Madsen, H., A. van Griensven, and A. L. Højbjerg (2010), Guidance Report 1.2 Model calibration and validation in model-based water management, in *Modelling Aspects of Water Framework Directive Implementation*, edited by P. A. Vanrolleghem, vol. 1, pp. 93–188, IWA Publ., London, U. K.

Majone, B., A. Bertagnoli, and A. Bellin (2010), A non-linear runoff generation model in small alpine catchments, *J. Hydrol.*, *385*(1-4), 300–312, doi:10.1016/j.jhydrol.2010.02.033.

Majone, B., C. I. Bovolo, A. Bellin, S. Blenkinsop, and H. J. Fowler (2012), Modeling the impacts of future climate change on water resources for the Gàllego river basin (Spain), *Water Resour. Res.*, *48*, W01512, doi:10.1029/2011WR010985.

Majone, B., F. Villa, R. Deidda, and A. Bellin (2015), Impact of climate change and water use policies on hydropower potential in the southeastern Alpine region, *Sci. Total Environ.*, doi:10.1016/j.scitotenv.2015.05.009, in press.

McClymont, A. F., M. Hayashi, L. R. Bentley, and J. Liard (2012), Locating and characterising groundwater storage areas within an alpine watershed using time-lapse gravity, gpr and seismic refraction methods, *Hydrol. Processes*, *26*(12), 1792–1804, doi:10.1002/hyp.9316.

McLaughlin, D., and L. R. Townley (1996), A reassessment of the groundwater inverse problem, *Water Resour. Res.*, *32*(5), 1131–1161, doi:10.1029/96WR00160.

Montgomery, E. L. (1971), Determination of coefficient of storage by use of gravity measurements, PhD thesis, Univ. of Ariz. Tucson, Ariz.

Nagy, D. (1966), The gravitational attraction of a right rectangular prism, *Geophysics*, *31*(2), 362–371, doi:10.1190/1.1439779.

Nobre, A. D., L. A. Cuartas, M. Hodnett, C. D. Rennó, G. Rodrigues, A. Silveira, M. Waterloo, and S. Saleska (2011), Height above the nearest drainage—A hydrologically relevant new terrain model, *J. Hydrol.*, *404*(1-2), 13–29, doi:10.1016/j.jhydrol.2011.03.051.

Parajka, J., and G. Blöschl (2008), The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, *J. Hydrol.*, *358*(3-4), 240 – 258, doi:10.1016/j.jhydrol.2008.06.006.

Pool, D. R. (2008), The utility of gravity and water-level monitoring at alluvial aquifer wells in southern Arizona, *Geophysics*, *73*(6), WA49–WA59, doi:10.1190/1.2980395.

Pool, D. R., and J. H. Eychaner (1995), Measurements of aquifer-storage change and specific yield using gravity surveys, *Ground Water*, *33*, 425–432, doi:10.1111/j.1745-6584.1995.tb00299.x.

Pool, D. R., and W. Schmidt (1987), Measurement of ground-water storage change and specific yield using the temporal-gravity method near Rillito creek, Tucson, Arizona, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *97-4125*, *87*(4125), 30 pp.

Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, N. Y.

Rubin, Y., S. S. Hubbard, A. Wilson, and M. A. Cushey (1998), Chapter 10. Aquifer characterization, in *The Handbook of Groundwater Engineering*, edited by J. W. Delleur, pp. 1–68, CRC Press, Boca Raton, Fla.

Rubin, Y., X. Chen, H. Murakami, and M. Hahn (2010), A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields, *Water Resour. Res.*, *46*, W10523, doi:10.1029/2009WR008799.

Rye, C. J., I. C. Willis, N. S. Arnold, and J. Kohler (2012), On the need for automated multiobjective optimization and uncertainty estimation of glacier mass balance models, *J. Geophys. Res.*, *117*, F02005, doi:10.1029/2011JF002184.

Rymer, H. (1994), Microgravity change as a precursor to volcanic activity, *J. Volcanol. Geotherm. Res.*, *61*(3-4), 311–328, doi:10.1016/0377-0273(94)90011-6.

Savenije, H. H. G. (2010), Hess opinions "topography driven conceptual modelling (flex-topo)," *Hydrol. Earth Syst. Sci.*, *14*(12), 2681–2692, doi:10.5194/hess-14-2681-2010.

Schoups, G., J. W. Hopmans, C. A. Young, J. A. Vrugt, and W. W. Wallender (2005a), Multi-criteria optimization of a regional spatially-distributed subsurface water flow model, *J. Hydrol.*, *311*(1-4), 20–48, doi:10.1016/j.jhydrol.2005.01.001.

Schoups, G., C. Lee Addams, and S. M. Gorelick (2005b), Multi-objective calibration of a surface water-groundwater flow model in an irrigated agricultural region: Yaqui valley, Sonora, Mexico, *Hydrol. Earth Syst. Sci.*, *9*(5), 549–568, doi:10.5194/hess-9-549-2005.

Seibert, J., and J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.

Sofyan, Y., Y. Daud, J. Nishijima, Y. Fujimitsu, Y. Kamah, A. Yani, Y. Fukuda, and M. Taniguchi (2015), The first repeated absolute gravity measurement for geothermal monitoring in the kamojang geothermal field, Indonesia, *Geothermics*, *53*, 114–124, doi:10.1016/j.geothermics.2014.05.002.

Tamura, Y. (1987), A harmonic development of the tide-generating potential, *Bull. Inf. Marées Terr.*, *99*.

Torge, W. (1989), *Gravimetry*, Walter de Gruyter, Berlin.

Vrugt, J. A., C. J. F. ter Braak, H. V. Gupta, and B. A. Robinson (2009), Equifinality of formal (dream) and informal (glue) bayesian approaches in hydrologic modeling?, *Stochastic Environ. Res. Risk Assess.*, *23*(7), 1011–1026, doi:10.1007/s00477-008-0274-y.

Wagener, T., and A. Montanari (2011), Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, *Water Resour. Res.*, *47*, W06301, doi:10.1029/2010WR009469.

Winter, T. C. (2001), The concept of hydrologic landscapes, *J. Am. Water Resour. Assoc.*, *37*(2), 335–349, doi:10.1111/j.1752-1688.2001.tb00973.x.

Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, *30*(8), 1756–1774, doi:10.1016/j.advwatres.2007.01.005.

Zhang, Z., T. Wagener, P. Reed, and R. Bhushan (2008), Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, *Water Resour. Res.*, *44*, W00B04, doi:10.1029/2008WR006833.