

<https://doi.org/10.1038/s43856-026-01611-9>

# Networked SIRS model with Kalman filter state estimation for epidemic monitoring in Europe



Atte Aalto <sup>1</sup>✉, Daniele Proverbio <sup>2</sup>, Giulia Giordano <sup>2</sup>, Alexander Skupin <sup>1,3,4</sup> & Jorge Gonçalves <sup>1,5</sup>

## Abstract

**Background** Metapopulation models, which consider epidemic spread across interconnected regions, can provide more accurate epidemic predictions compared to isolated models for the corresponding regions. Still, their added complexity and data requirements raise questions about their tangible benefits over simpler, localized models.

**Methods** We develop and validate two networked compartmental metapopulation models for predicting influenza-like illness across Europe: a detailed network-based model, including international travel dynamics, and a simpler mean-field model, aggregating average regional data. The network is constructed using public mobility data and complemented with population densities at border regions. Incidence data of influenza-like illnesses from 28 countries are integrated using an Extended Kalman filter.

**Results** We show that networked epidemic models effectively capture epidemic dynamics across regions and epidemic phases. The models enable accurate forecasts, missing data imputation, and actionable insights: network models outperform isolated models in forecasting epidemic progression, particularly during critical periods such as wave onsets and peaks, and maintain reliability in scenarios with missing data.

**Conclusions** The findings unveil and quantify the advantages of metapopulation models for epidemic forecasting in interconnected regions, and pave the way to the integration of mobility and epidemic surveillance to improve the monitoring and prediction of spreading diseases.

## Plain language summary

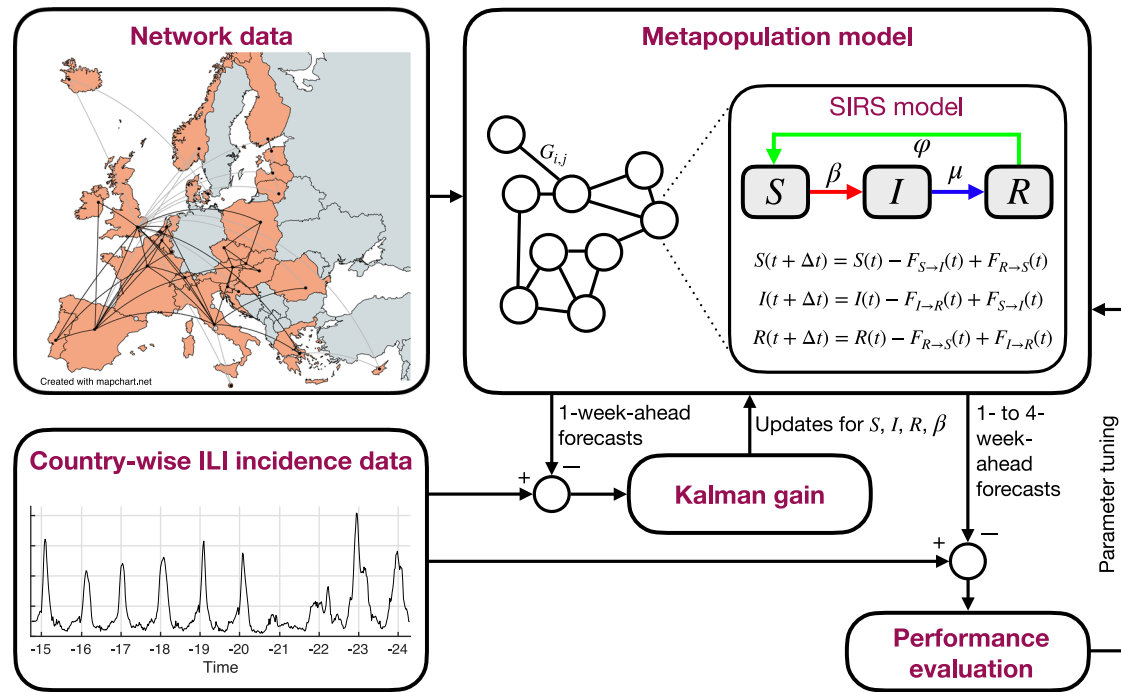
Flu and other respiratory illnesses spread quickly across countries, especially in regions like Europe where people travel frequently. To better understand and predict these patterns, we built two mathematical models that predict how flu-like illnesses move between countries using information about travel dynamics and national illness reports. These models were tested using data from 28 European countries and found that they can track and forecast outbreaks more accurately than models that look at each country on its own. The models were especially good at predicting when new waves of illness would start and peak, and they could also fill in gaps when some countries had missing or delayed data. These results show that combining mobility and health data can improve how we monitor and anticipate disease spread, helping public health agencies respond faster and more effectively.

Modern epidemiology increasingly relies on mathematical models to integrate empirical data, perform advanced analytics, and provide quantitative predictions that help forecast and control the spread of infectious diseases<sup>1,2</sup>. These models not only predict epidemic progression and future scenarios, but also support the development of early warning systems<sup>3</sup>, the assessment of the impact and effectiveness of restrictions<sup>4</sup>, and the estimation of unreported cases<sup>5</sup> or cryptic transmission<sup>6,7</sup>. Hence, advancing forecasting methodologies is crucial to enable timely and effective public health responses and healthcare resource allocation, as testified by initiatives like Influcast<sup>8</sup>, Respicast<sup>9,10</sup>, and similar CDC initiatives<sup>11</sup> aimed at providing platforms for integrating and benchmarking diverse forecasting models for influenza-like illnesses (ILI).

Traditional epidemiological models often focus on individual countries or regions, with the aim of predicting the overall spread of infectious diseases based on available data<sup>12</sup>, but without capturing the broader dynamics of interconnected regions. In contrast, metapopulation models<sup>13–16</sup>, based on the theory of epidemic spreading on complex networks<sup>17–20</sup> and multi-patch mathematical models<sup>21–23</sup>, can predict the spatiotemporal progression of infectious diseases across large interconnected areas by incorporating mobility networks<sup>24–26</sup> or distance-based networks<sup>16</sup>.

Recent studies have successfully employed network-based metapopulation models to forecast real-world epidemic dynamics, including COVID-19 across Italian regions<sup>27,28</sup>, influenza across U.S. states<sup>29,30</sup>, and dengue in China's Guangdong province<sup>31</sup>. In fact, considering mobility at

<sup>1</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg. <sup>2</sup>Department of Industrial Engineering, University of Trento, Trento, Italy. <sup>3</sup>Department of Physics and Material Sciences, University of Luxembourg, Belvaux, Luxembourg. <sup>4</sup>Department of Neurosciences, University of California, San Diego, CA, USA. <sup>5</sup>Department of Plant Sciences, University of Cambridge, Cambridge, UK. ✉e-mail: [atte.aalto@uni.lu](mailto:atte.aalto@uni.lu)



**Fig. 1 | Model scheme.** The network data panel shows daily fluxes above 10,000 with black lines, and the two strongest connections for each country with gray lines.

any given spatial granularity (e.g., between regions or states) has been shown to provide additional information compared to only considering region-wise incidence<sup>30,32,33</sup>. Despite the significant potential of metapopulation models, a critical question persists: given the same availability and spatial resolution of data for a given disease, do metapopulation models offer improvements in forecasting accuracy, and for which aspects of epidemic dynamics, compared to isolated models<sup>34</sup>?

Here we investigate whether a network-based metapopulation model improves epidemic forecasts for ILI over multiple seasons in Europe. Using empirical data on European mobility, we construct two models of increasing complexity: a mean-field model focuses on the epidemic evolution within a country and the average incidence in Europe, requiring relatively few data; a detailed network-based model accounts for inter-country travel dynamics and requires additional data on the network structure of mobility between countries. To estimate and predict disease dynamics, both models use an Extended Kalman Filter (EKF), an automated calibration method, also employed in alternative versions by earlier epidemiological studies<sup>5,29–31,34–37</sup>, to fruitfully integrate mechanistic models with empirical data and provide forecasts. The performance of the models is characterized by comparison against corresponding country-specific models, during 10 years of weekly ILI incidence data. These country-specific models were among the top-performing methods in the Respcast dashboard during the season 2023–2024<sup>10</sup>. Our results reveal and quantify the benefits of metapopulation models for epidemic forecasting in interconnected regions, offering probabilistic forecasts and scenario modeling at a continental scale. Our findings show that augmenting localized information with network-based data improves prediction accuracy, particularly for key periods such as wave onset, and resilience to missing data. Moreover, our framework can be adapted to other diseases and geographic scales, and thus offers a versatile tool for global public health applications.

## Methods

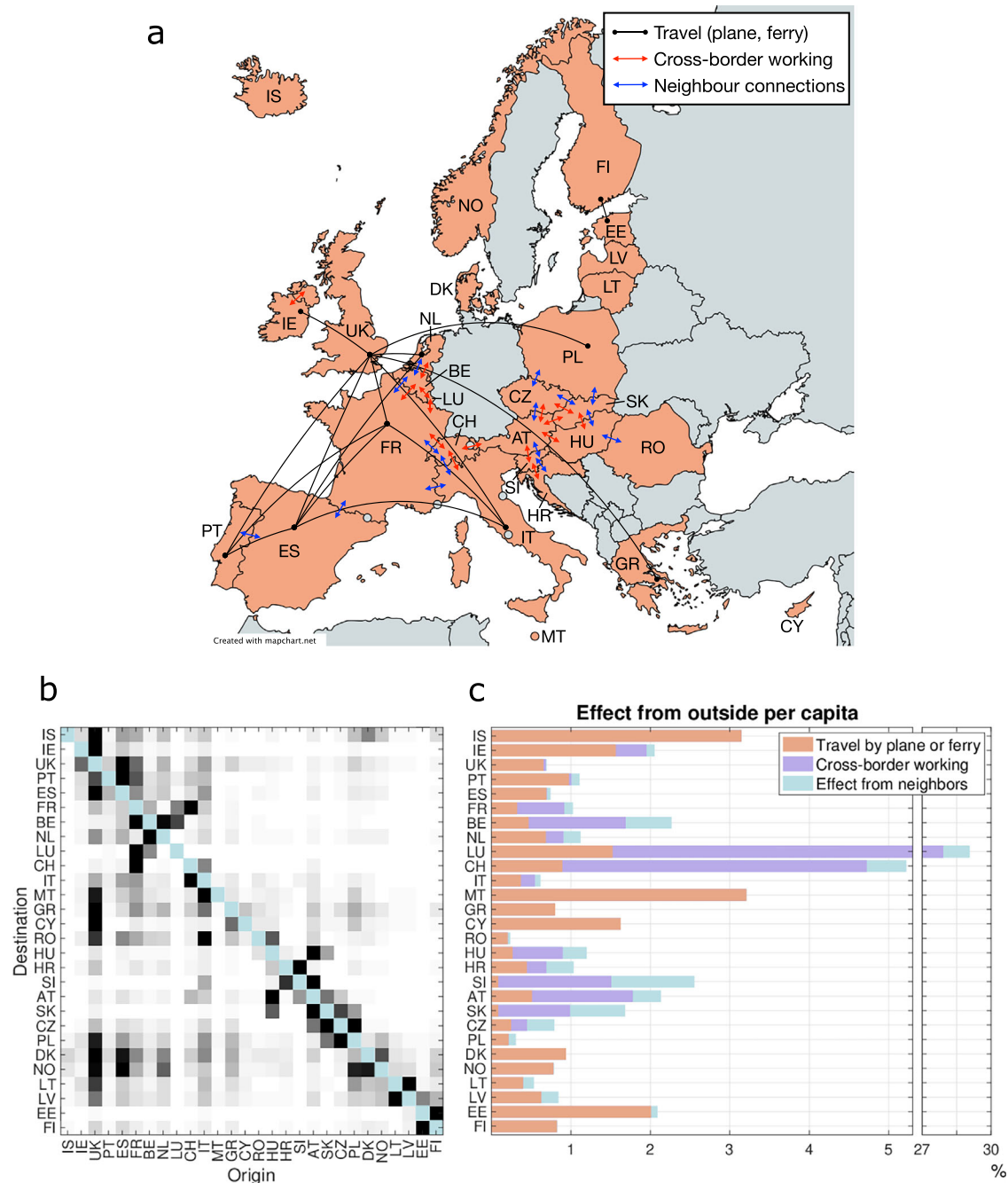
This section first outlines the construction of the mobility network; then, it introduces the dynamical model used to simulate the spread of ILI across 28 European countries. The model is based on a stochastic discrete-time Susceptible-Infectious-Removed-Susceptible (SIRS) framework, which describes the time evolution of the disease spread in

homogeneous and well-mixed populations<sup>38–41</sup>. The model is extended to incorporate network effects to capture both local epidemic dynamics within each country and the influence of cross-border interactions. The full pipeline integrating the model with empirical data through an EKF is schematically represented in Fig. 1.

### Network generation

To construct the metapopulation model, we first developed a mobility network representing population flows between European countries; see Fig. 2a. The network was built using publicly accessible data sources, ensuring consistency and comprehensive coverage. Alternatively, the network could be inferred from epidemic data, but this has been shown to be a severely ill-posed problem<sup>42</sup>. Each country was modeled as a node, with links representing three key types of travel flows: air and ferry travel, cross-border commuting, and other forms of land-based movement between neighboring countries. These components account for diverse modes of mobility, providing a realistic depiction of cross-border interactions. Here, we provide an overview of the network structure. An alternative to our travel-based network would be a distance-based network, such as the “gravity model”<sup>16</sup>. Supplementary Fig. 2 compares our network to the gravity network, highlighting that the gravity model underestimates connections based on popular flight (long-distance) destinations while overestimating mobility from certain large countries such as Italy or Poland. When possible, it is thus more effective to directly use empirical data to set the mobility matrix.

**Travel network.** Data for plane travels in 2023 were obtained from Eurostat<sup>43</sup>; only for UK, we used data from 2019 as the most recent available data before the COVID-19 pandemic. The flight data are given quarterly, and we averaged them over the whole year. Some missing value imputation is done based on data from different quarters between the countries in question, accounting for overall quarterly differences. Ferry traffic data were gathered from the statistics of the Finnish Port Association for Finland<sup>44</sup> and from the UK Government, Department for Transport<sup>45</sup>. While COVID-19 disrupted air and ferry travel patterns, we used a fixed network for the entire period based on either pre- or post-pandemic data. Future studies may embed dynamical data flows, when available for the whole network.



**Fig. 2 | The constructed network.** **a** The countries included in the network are highlighted in light red. The 15 busiest travel connections (by plane or ferry on any route between the indicated countries) are indicated with black lines (>10,000 passengers/day/direction). Red arrows show the 15 busiest cross-border worker flows (>3000 workers), and blue arrows show the 15 strongest neighbor connections of other types. **b** Visualization of the network matrix  $G$ . For visualization, each row

has been normalized to have maximum value one. Darker color indicates stronger effect. **c** The effect from the outside for each country, calculated as the row sums of the network matrices  $y_1 G^{(1)}$ ,  $y_1 G^{(2)}$ , and  $y_2 G^{(3)}$  shown as percentages of the respective countries' populations. Only effects from other countries included in the network are considered.

**Cross-border commuting network.** Cross-border worker flows were extracted from reports by the European Commission's Directorate-General for Employment, Social Affairs and Inclusion<sup>46</sup> for the EU and EFTA areas, and the Economic and Social Research Institute<sup>47</sup> for the UK-Ireland border. Commuter flows between non-neighboring regions were excluded, as they are typically captured within the air and ferry travel data. It was assumed that cross-border workers actually cross borders on 60% of days (thus excluding weekends, bank holidays and average leave periods). Cross-border workers were counted in the network as flows in both directions, between their country of residence and

the country of work, as they were assumed to commute daily and thus be able to carry an infection in either direction.

**Construction of neighbor graph.** The neighbor graph lumps together all other travels across the land borders of different countries in the absence of available data. For each NUTS 3 region<sup>48</sup> that shares a border with another country, we measured the length of the border segment between the region and the neighboring country, and weighted that with the population density of the region<sup>49</sup>. Weighted segments were then added up for the whole border between two countries, including the border

regions of both countries so as to have a symmetric network matrix. The numbers obtained in this way are considered as dimensionless units and later scaled by parameter fitting. An example for this procedure for the Czechia–Slovakia border is shown in Supplementary Fig. 1.

This procedure provides a simple estimation of unregistered movements, but it has some shortcomings. For example, it looks at only one side of the border at a time. In reality, there is likely more traffic across the border if both sides of the border are densely populated. For example, Vienna and Bratislava, the capitals of Austria and Slovakia, are very close to each other. Although our procedure predicts fairly high traffic across the border, it might still be under-estimated. Another simplification is the omission of geographical specificities. For example, the borders between France and Italy and France and Spain are on mountainous regions. Even though both sides of these borders are densely populated, the traffic across these borders might be over-estimated by our procedure. Having better curated datasets may further improve this estimation in future applications.

Some outlier flows appear by this procedure. These outliers arise due to densely populated regions, with narrow or protruding shape, that share a relatively long border with a neighboring country. In addition, some cities constitute their own regions in the NUTS 3 classification, and if these cities lie at the border, they may cause outliers as well. These outliers are dealt with either by joining the corresponding regions with (larger) neighboring regions or, in few cases, by manually modifying the border length. Details on handling outliers are described in Supplementary Note 1.

**Network structure and parameters.** Our network matrix integrates the three mobility components, weighted by tuning parameters  $\gamma_1$  and  $\gamma_2$ , to account for the relative contributions of different forms of travel. The population fluxes corresponding to air-and-ferry and commuting flows, denoted by matrices  $G^{(1)}$  and  $G^{(2)}$ , were combined and scaled by a tuning parameter  $\gamma_1$  to account for varying durations of cross-border trips and for travel-associated infection risks, while the land-based movement component, denoted by  $G^{(3)}$ , was scaled by a different tuning parameter  $\gamma_2$  since  $G^{(3)}$  does not readily correspond to actual fluxes and thus requires proper dimensionalization and rescaling. The resulting network matrix

$$G = \gamma_1 (G^{(1)} + G^{(2)}) + \gamma_2 G^{(3)} \quad (1)$$

captures the influence of mobility on epidemic dynamics, enabling detailed modeling of interconnected populations. The complete network is visualized in Fig. 2b as the heatmap of the row-wise normalized adjacency matrix  $G_{i,j} / \max(G_{i,\cdot})$  between countries of origin and destination (normalization done only for visualization).

**Regional impacts of the network.** In almost all countries, the contribution of traveling people to the spread of ILI diseases is estimated to be less than 5% (Fig. 2c). An exception is Luxembourg, where cross-border workers represent an important fraction of the effective population. For island countries, or those where the major urban areas are close to shores (e.g., Iceland, UK, or Finland), the largest effect is due to air and ferry travel.

By combining diverse mobility data sources into a unified framework, our network generation approach provides a foundational framework to simulate disease spread across Europe. The methodology is flexible and can be extended to other countries or regions as new data become available.

**Mean-field approximation.** To explore the trade-off between data availability and model complexity, we implemented a simpler “mean-field” version of the model, where the detailed network matrix  $G$  is replaced with

$$\hat{G}_{i,j} = \gamma \frac{N_i N_j}{\sum_{k=1}^{28} N_k}, \quad (2)$$

where  $\gamma$  is a tuning parameter and  $N_i$  is the population of country  $i$ . This approach is less granular, but requires fewer data and is suitable for scenarios with incomplete mobility information. The heatmap for the mean-field network matrix is shown in Supplementary Fig. 2.

**Individual models.** The performances of the network-based and mean-field models were compared against country-specific isolated models that do not consider cross-border interactions, that is, such models do not account for a network matrix  $G$ .

### SIRS model at each node of the network

To simulate disease dynamics at each node of the network, we employ a discrete-time SIRS model:

$$\begin{aligned} S(t + \Delta t) &= S(t) - F_{S \rightarrow I}(t) + F_{R \rightarrow S}(t) \\ I(t + \Delta t) &= I(t) + F_{S \rightarrow I}(t) - F_{I \rightarrow R}(t) \\ R(t + \Delta t) &= R(t) + F_{I \rightarrow R}(t) - F_{R \rightarrow S}(t) \end{aligned} \quad (3)$$

where  $F_{X \rightarrow Y}(t)$  is a shorthand notation for the flow from compartment  $X$  to compartment  $Y$  at time  $t$ , with unit of 1 week (we will later omit the time-dependency). The model’s time step is 1 day,  $\Delta t = 1/7$ . The model satisfies the conservation law

$$S(t + \Delta t) + I(t + \Delta t) + R(t + \Delta t) = S(t) + I(t) + R(t).$$

In the country-specific SIRS model, the fluxes between compartments are given by  $F_{S \rightarrow I}(t) = \beta(t)S(t)I(t)/N$ ,  $F_{I \rightarrow R}(t) = \mu I(t)$ , and  $F_{R \rightarrow S}(t) = \varphi R(t)$  where  $\mu = 0.06$  and  $\varphi = \ln(2)/60 = 0.0116$ . Parameter  $\mu$  is set to a small value because, to guarantee stability of the state estimation, the time scale of the model dynamics should be close to the time scale of data collection. Parameter  $\varphi$  corresponds to immunity half life of 60 days: ILI may be caused by many different viruses, with varying cross-immunity properties, so this parameter should be considered as an effective parameter rather than a reflection of the human immune system. In the metapopulation network model,  $M$  countries with populations  $N_i$  ( $i = 1 \dots M$ ) are represented as nodes in the network, with epidemic dynamics described by SIRS equations (3). The regions are connected by the mobility network matrix  $G$  defined in (1). Its entries  $G_{i,j}$  represent effective flows of individuals from region  $j$  to region  $i$  per day. The flows are called “effective” since we will estimate global parameters  $\gamma_1$  and  $\gamma_2$  in (1) multiplying the network matrix based on forecasting performance. In the network model, the number of new infections in country  $i$  in 1 day, that is, the flux from the susceptible population  $S_i(t)$  to the infectious population  $I_i(t)$  at time  $t$ , is given by

$$F_{S_i \rightarrow I_i}(t) = \frac{\beta_i(t)S_i(t)}{N_i} \left( I_i(t) + \sum_{j=1}^M G_{i,j} \frac{I_j(t)}{N_j} \right). \quad (4)$$

The flux depends on the infectious population within the country  $i$  itself, but also other countries:  $I_j(t)/N_j$  is the fraction of infected individuals in country  $j$ ; hence assuming perfect mixing of populations, the flux of infected individuals from country  $j$  to country  $i$  is  $G_{i,j}I_j(t)/N_j$ . We assume that these flows are symmetric:  $G_{i,j} = G_{j,i}$ . This assumption is a slight simplification of a more general conservation law stating that the population of each country remains constant, which would require  $\sum_i G_{i,j} = \sum_i G_{j,i}$  for all  $j$ . We do not differentiate between people traveling to another region and becoming infected there, and infectious people traveling and infecting others at the destination region.

For the mean-field network model, instead, we substitute  $G$  defined in Eq. (2) into (4) to get

$$\hat{F}_{S_i \rightarrow I_i}(t) = \beta_i(t)S_i(t) \left( \frac{I_i(t)}{N_i} + \gamma \frac{\sum_{j=1}^M I_j(t)}{\sum_{j=1}^M N_j} \right) \quad (5)$$



where the second term inside the parentheses does not depend on the target country  $i$ , but is the average prevalence over all included countries, modulated by the parameter  $\gamma$ . This model ignores geographical and other detailed information on the traffic between countries. For the country-specific models, we set  $G_{ij} = 0$  to recover the isolated node-specific SIRS model.

For all models, the other fluxes  $F_{I \rightarrow R}(t)$  and  $F_{R \rightarrow S}(t)$  are as in the isolated models, and do not assume dependency on travels.

## Stochastic model

Towards an implementation of the EKF, a stochastic version of the networked SIRS model (4) is obtained based on the method described in ref. 37. We follow the derivation of a Langevin equation<sup>50</sup>, that is, we pose that each susceptible person has a probability (in the non-networked version, for simplicity)  $\beta(t)I(t)/N$  to become infected at time  $t$ . The number of new infections on a day at time  $t$  is then binomially distributed with mean  $\frac{\beta(t)I(t)}{N}S(t)$  and variance  $\frac{\beta(t)I(t)}{N}\left(1 - \frac{\beta(t)I(t)}{N}\right)S(t)$ . Assuming  $\frac{\beta(t)I(t)}{N} \ll 1$ , the term with  $\left(\frac{\beta(t)I(t)}{N}\right)^2$  is omitted from the variance. Further assuming  $\frac{\beta(t)I(t)}{N}S(t)$  is large enough, the binomial distribution can be well approximated by the normal distribution with the same mean and variance.

As a result, each transition between compartments is accompanied by a stochastic white noise process modulated by a multiplicative state-dependent term, which scales as the square root of the transition rate. However, the derivation of the stochastic model implicitly assumes a homogeneous and well-mixed population. This assumption leads to severely underestimated noise levels. To account for this underestimation, the variance terms are multiplied by a tuning factor  $\kappa_1$ . In addition, a second term  $\kappa_2 N_i^2$  is added to the variance of the transition from  $S$  to  $I$ , that does not depend on the epidemic state. This can be interpreted as import of cases from countries that are not included in the network; a second reason behind its introduction is technical: during summer periods, many countries stop the ILI monitoring and reporting, hence the model tends to a state where there are no new cases emerging. Then, also the uncertainty on the epidemic state eventually converges to zero. When reporting is restarted, the Kalman filter can mainly make updates on the  $\beta$ -parameter due to its unfounded certainty on the model variables. This leads to poor projections in the beginning of a new epidemic season. The term  $\kappa_2 N_i^2$  is introduced to mitigate this problem by providing inertia to uncertainties at the beginning of a new season. The final stochastic networked SIRS model is:

$$\begin{aligned} S(t + \Delta t) = S(t) - F_{S \rightarrow I} + F_{R \rightarrow S} - \sqrt{\kappa_1 F_{S \rightarrow I}} w_1(t) \\ + \sqrt{\kappa_1 F_{R \rightarrow S}} w_3(t) - \sqrt{\kappa_2 N w_4(t)} I(t + \Delta t) = I(t) + F_{S \rightarrow I} \\ - F_{I \rightarrow R} + \sqrt{\kappa_1 F_{S \rightarrow I}} w_1(t) - \sqrt{\kappa_1 F_{I \rightarrow R}} w_2(t) \\ + \sqrt{\kappa_2 N w_4(t)} R(t + \Delta t) = R(t) + F_{I \rightarrow R} - F_{R \rightarrow S} \\ + \sqrt{\kappa_1 F_{I \rightarrow R}} w_2(t) - \sqrt{\kappa_1 F_{R \rightarrow S}} w_3(t) \end{aligned} \quad (6)$$

where  $w_k(t)$  are mutually independent Gaussian white noise processes with unit variance.

## Dynamics of transmission parameter

The transmission parameters  $\beta_i(t)$  in (4) and (5) are modeled as time-varying variables influenced by both local and network-level factors. Changes in  $\beta_i(t)$  can reflect shifts in population behavior, interventions, seasonal effects, or varying transmissibility of different viral strains. The dynamics of  $\beta_i(t)$  are governed by

$$\beta_i(t + \Delta t) = \beta_i(t) + k_1 \left[ k_2 \frac{\sum_j G_{ij} \beta_j(t)}{\sum_j G_{ij}} + (1 - k_2) k_3 \mu - \beta_i(t) \right] + w_\beta(t), \quad (7)$$

where  $w_\beta(t)$  is Gaussian noise with constant variance  $\sigma_\beta$  and the term inside the brackets acts as a nudge towards a reference value that is a combination

of the weighted average of  $\beta_j$ 's in the neighbors of country  $i$  and a “grounding value”  $k_3 \mu$  corresponding to a basic reproduction number  $R_0 = k_3$ . The tuning parameters  $k_1$ ,  $k_2$ , and  $k_3$  all have their own roles:  $k_1$  modulates the strength of the nudging towards the reference value,  $k_2$  is the weight between the two components of the reference value, and  $k_3$  scales the grounding value. In isolated models,  $k_2$  is zero. In the absence of data, the  $\beta$ -parameters tend to the grounding value  $k_3 \mu$ , and, consequently, the system would eventually converge to an endemic steady state, which is (for the isolated models)  $S = N/k_3$ ,  $I = \frac{\phi(k_3 - 1)}{(\phi + \mu)k_3} N$ , and  $R = \frac{\mu(k_3 - 1)}{(\phi + \mu)k_3} N$ .

In contrast to the model in ref. 37, we include deterministic dynamics for the  $\beta$ -parameters. The reason for the inclusion is threefold. Firstly, due to missing data in many countries during the summer periods, the Kalman filter does not make significant updates to the  $\beta_i$  parameters during these periods. However, it is reasonable to assume that, before seasonal re-emergence of ILI, the effective transmission approaches similar values each year, without retaining the value reached at the end of the previous wave. The deterministic dynamics for  $\beta_i$  first takes into account this seasonality. Secondly, some effects modeled by the time-varying  $\beta$ -parameters are not country-specific, such as seasonal effects and transmissibility of a prevailing viral strain. Thirdly, by allowing the country-specific  $\beta_i$  to be influenced by the parameters of the neighbors in the network, the parameter estimates become more stable. This inclusion of network effects in the  $\beta$ -parameter dynamics differs significantly from previous models<sup>30,31,34</sup> as it allows regional waves to influence neighboring countries, thus enhancing the model's ability to anticipate epidemic patterns (as seen in Supplementary Fig. 7).

## Extended Kalman filter implementation

To estimate epidemic parameters and make accurate predictions, the network-based SIRS model was coupled with an EKF<sup>51,52</sup>. The EKF integrates real-world observations with the model dynamics, providing robust state estimation and allowing for dynamic adjustments as new data become available. The EKF requires an underlying dynamical model (in our case, the stochastic networked SIRS model), its output, covariance matrices for state noise and measurement noise, and measurement data (Fig. 3), provided as weekly aggregated case numbers. The EKF works recursively involving a prediction step using the model, and then an update of the state vector based on discrepancies between predictions and observations, ensuring that the model evolution is close to real-world dynamics. The algorithm is presented in Supplementary Note 2.

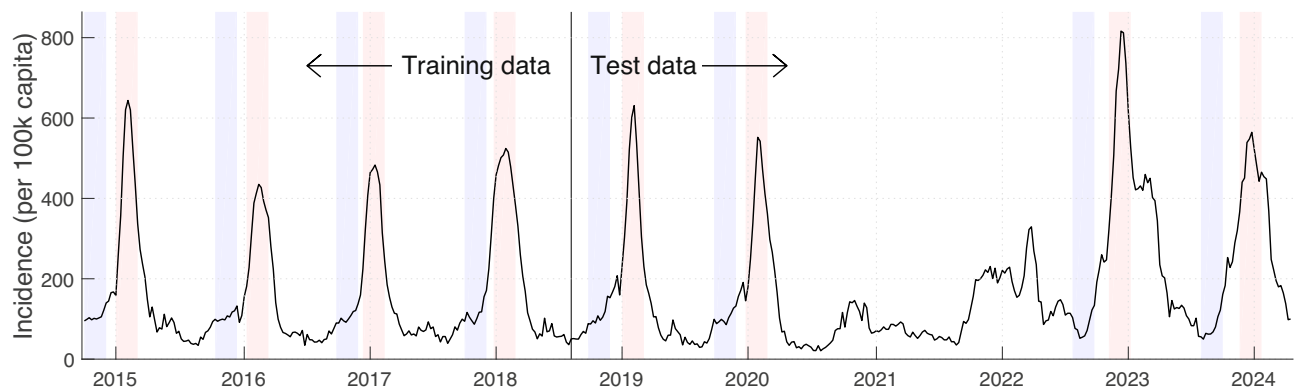
To apply the EKF into the networked system, the state of the system for each country is represented with a 3-dimensional state vector, consisting of  $S_i(t)$ ,  $I_i(t)$ , and  $\beta_i(t)$ . The  $R_i(t)$  compartments are not included, since they can be recovered using the conservation law,  $R_i(t) = N_i - S_i(t) - I_i(t)$ . The 3-dimensional state vectors for all 28 countries are concatenated into one 84-dimensional state vector  $x$ , for which we have a stochastic model

$$x(t + \Delta t) = f(x(t)) + W(t).$$

$[f_1(x), \dots, f_{28}(x)]^T$  country by country through equations (3), (4), and (7). The state noise vector  $W(t)$  includes the contributions of all noise terms in the stochastic model (6), and from the stochastic model we can obtain its covariance matrix  $Q$  as a block diagonal matrix whose  $i^{\text{th}}$  block is given by

$$Q_i = \begin{bmatrix} \kappa_1(F_{S_i \rightarrow I_i} + F_{R_i \rightarrow S_i}) + \kappa_2 N_i^2 & -\kappa_1 F_{S_i \rightarrow I_i} - \kappa_2 N_i^2 & 0 \\ -\kappa_1 F_{S_i \rightarrow I_i} - \kappa_2 N_i^2 & \kappa_1(F_{S_i \rightarrow I_i} + F_{I_i \rightarrow R_i}) + \kappa_2 N_i^2 & 0 \\ 0 & 0 & \sigma_\beta \end{bmatrix}. \quad (8)$$

The observed case numbers are also assumed to be affected by uncertainties. We assume that every infectious person has a probability  $c_i(t)$  to be detected at week  $t$  and the detections are independent. Then, the number of detected cases is binomially distributed with mean  $c_i(t)I_i(t)$  and variance  $(1 - c_i(t))c_i(t)I_i(t)$ . The model-predicted number of detected cases



**Fig. 3 | Average ILI incidence across Europe.** Influenza-like illness (ILI) incidence data were obtained from the “Respicast” project<sup>9</sup>, which consolidates data from the ECDC’s European Respiratory Virus Surveillance Summary (ERVISS)<sup>53</sup> and the World Health Organization’s FluID global influenza program<sup>54</sup>. The dataset spans ten epidemic seasons (2014–2024). The figure shows the population-weighted average ILI incidence across Europe. Blue-shaded areas indicate wave onset periods,

while red-shaded areas mark epidemic peaks. These critical periods were used for evaluating forecasting performance in subsequent analyses. The data were aggregated from weekly reports and adjusted for population differences across reporting countries. The figure highlights the seasonal ILI dynamics and the variation in timing and intensity of epidemic waves.

for country  $i$  is therefore given by

$$\hat{y}_i(t) = c_i(t)I_i(t) \quad (9)$$

and the full observation vector  $\hat{y}(t) = [\hat{y}_1(t), \dots, \hat{y}_{28}(t)]^\top$  is obtained as  $\hat{y}(t) = C(t)x(t)$  where  $C(t)$  is a  $28 \times 84$  matrix whose element  $(i, 3(i-1) + 2)$  is  $c_i(t)$  with zeros elsewhere. The measurement noise covariance matrix  $U$  for the full observation vector is diagonal with elements

$$U_i(t) = (\rho_1(1 - c_i(t))y_i(t) + \rho_2 N_i)u_i \quad (10)$$

where the first term inside the parentheses corresponds to the variance of the binomial distribution (using the observed number  $y_i(t)$  as a proxy for  $c_i(t)I_i(t)$ ). As we did with the stochasticity of new infections, here we also introduce a second, state-independent noise term whose variance is proportional to the population size  $N_i$  to enforce a lower bound to the variance during times of low incidence.  $\rho_1$  and  $\rho_2$  are tuning parameters for the different noise terms, and  $u_i$  is a country-specific coefficient measuring the noise level in the data, defined by

$$u_i = \frac{1}{499} \sum_{t=1}^{499} \frac{(y_i(t) - \bar{y}_i(t))^2}{\bar{y}_i(t)},$$

where  $\bar{y}_i(t)$  denotes the 5-week moving average of  $y_i$ , with the moving window centered at  $t$ . The coefficient  $u_i$  reflects country-specific differences in data collection and reporting practices. It compares the data variance over the entire dataset covering 499 weeks around a moving average to the variance of the binomial distribution, which is approximated by the moving average.

The parameter  $c_i(t)$  in the model output adjusts for under-reporting. The ratio of detected and total cases is captured by  $\frac{c_i(t)\Delta t}{\mu}$ . In SIR-type models, the initial exponential growth rate of a new wave and the herd immunity level are tightly connected. However, if only a fraction of the infections are detected, this has no effect on the observed initial exponential growth rate but it directly affects the amplitude of the upcoming (detected) wave. Without knowledge of this fraction, the link between initial growth rate and wave amplitude is broken, which precludes (model-based) projections going beyond short-term extrapolation. Therefore, the  $c_i(t)$  parameters are crucial for modeling the wave amplitude, and we have developed an adaptive scheme for their estimation.

As a small remark, note that  $I_i(t)$  in (9) corresponds to ILI prevalence, whereas the data are ILI incidences. In the CoWWAn<sup>37</sup> model and for the

Respicast projections (during epidemic season 2023–2024), we used a slightly more complex observation model with an additional “counter” variable for the daily or weekly new cases precisely corresponding to disease incidence. However, this model implicitly assumed that all observations were from infections occurring on the same day or week, whereas the current model (9), where case numbers are simply obtained from the size of the  $I$ -compartment, assumes that an infection can be detected and counted at any time when the infected individual is in the  $I$ -compartment. We observed that forecasting performance was just as good with this simpler model warranting its implementation here.

After every Kalman update, the model (without noise) is run 4 weeks forward in time to produce incidence projections for  $\tau = 1, 2, 3, 4$  weeks ahead, denoted by  $\hat{y}_i(t + \tau|t)$ . That is,  $\hat{y}_i(t + \tau|t)$  is a projection for week  $t + \tau$  done using data up to week  $t$ .

### Adaptive estimation of $c_i(t)$

**Initial values.** The initial values for the ratios are based on the observed epidemic dynamics for the first two epidemic waves in the data. The discounted case numbers are calculated as  $Y_i(t; 1 - \nu) = \sum_{\tau=0}^t (1 - \nu)^{t-\tau} y_i(\tau)$  where  $\nu = \frac{1}{1/\mu + 1/\phi}$  represents the average flow rate through the  $I$  and  $R$  compartments (that is,  $1/\nu$  is the average time from infection to becoming susceptible again). Then

$$c_i(1) = \alpha_1 \frac{\max_{t \in [T_{0,i}, T_{0,i} + 103]} Y_i(t; 1 - \nu)}{N_i} \quad (11)$$

where  $T_{0,i}$  denotes the first week when data are available for country  $i$ . Here  $\alpha_1$  is a tuning parameter.

**Upward update.** If a too small value  $c_i(t)$  is used for a country during an epidemic wave, the  $S$ -compartment will become exhausted. When this happens, more precisely, when  $S_i(t) < 0.4N_i$  for some country, this is deemed unrealistic, and a parameter adaptation is triggered. The parameter  $c_i(t)$  is retroactively increased, and the last 12 weeks are re-run with the higher parameter value. Further upward updates are not allowed during the re-run weeks.

The coefficient used for increasing  $c_i(t)$  is determined as follows. The discounted true case numbers and 1-week ahead forecasts are calculated:

$$Y_i(t; 0.8) = \sum_{\tau=1}^t 0.8^{t-\tau} y_i(\tau), \quad \hat{Y}_i(t; 0.8) = \sum_{\tau=1}^t 0.8^{t-\tau} \hat{y}_i(\tau).$$

The parameter  $c_i(t)$  is then increased by multiplying it with the coefficient

$$\max \left\{ 1.1, \left( \frac{Y_i(t; 0.8)}{\hat{Y}_i(t; 0.8)} \right)^{\alpha_2} \right\} \quad (12)$$

where the exponent  $\alpha_2$  is a tuning parameter used to link case numbers with the parameter  $c_i(t)$ . The parameter is increased retroactively in a piecewise linear fashion, that is,  $c_i$  has a linearly increasing phase from  $t - 12$  to  $t - 6$  from the old value to the increased value. From week  $t - 5$  to  $t + 15$ , the parameter remains at the new higher value. From week  $t + 16$  to  $t + 26$  the parameter linearly decreases to a value that is  $c_i(t - 12) + \alpha_3(c_i(t) - c_i(t - 12))$ , that is,  $\alpha_3 \in (0, 1)$  denotes a share of the parameter update that is kept for the next epidemic wave. This parameter is used only when an upward update is triggered; we observed very few triggering events during the training data period. Hence,  $\alpha_3$  was estimated directly from data as illustrated in Supplementary Fig. 3.

Once the parameter has been updated, the last 12 weeks are re-run with the new parameter value. Note that the  $\tau$ -week ahead projections  $\hat{y}_i(t + \tau|t)$  are only generated when week  $t$  is passed for the first time in the algorithm to ensure causality of the projections.

**Downward update.** Unlike for too small  $c_i(t)$ , there are no clear signs for too high value, except for overshooting projections, in particular, during times of high incidence. Therefore, downward updates are only done on pre-determined times between epidemic waves (that is, in summer). At such time  $T$ , the performance for the previous year is analyzed for all countries. To check for overshooting projections, for each country, we find the weeks from the previous year  $T - 51, \dots, T$  during which the incidence has been on the highest 30th percentile during that year (denote this set of weeks by  $T_i(T)$ ). The  $c_i(t)$  is then reduced by multiplication by a factor

$$\min \left\{ 1, \left( \frac{\sum_{t \in T_i(T)} \sum_{\tau=1}^4 y_i(t + \tau)}{\sum_{t \in T_i(T)} \sum_{\tau=1}^4 \hat{y}_i(t + \tau|t)} \right)^{\alpha_2} \right\}. \quad (13)$$

The reduction is done in a piecewise linear way, such that the linear descent takes place on weeks  $T + 1$  to  $T + 6$ .

## Statistics and reproducibility

Travel data to construct the network model have been sourced from public mobility repositories or estimated using population census, as described in the section “Network generation”.

ILI incidence data originates from the ECDC’s European Respiratory Virus Surveillance Summary (ERVISS)<sup>53</sup> and the World Health Organization’s (WHO) FluID global influenza program<sup>54</sup>. Since the study is based on publicly available and fully anonymized aggregate data, and because ILI does not constitute a sensitive topic, an ethical approval was not required<sup>55</sup>. The dataset spans ten epidemic seasons, from week 40 of 2014 to week 16 of 2024. Incidences are reported as weekly cases per 100,000 inhabitants. To enable integration into the metapopulation model, these incidences were scaled to estimate the actual number of detected infections for each country. For the United Kingdom, data were aggregated from England, Wales, Scotland, and Northern Ireland to align with available mobility data, which were reported for the whole UK.

All data were processed using the model-based approach described above, integrating them into the EKF, informed by stochastic dynamical epidemic models. The model parameters and the noise content are fitted and calibrated as described below. The complete pipeline is available on the project Gitlab repository (see “Data and Code availability”).

## Performance evaluation

For performance evaluation, we define the  $\tau$ -week ahead forecasts  $\hat{y}_i(t + \tau|t) = c_i(t + \tau)\hat{I}_i(t + \tau|t)$  where  $\hat{I}_i(t + \tau|t)$  is obtained by simulating the

noise-free model forward in time for  $\tau$  weeks, starting from the Kalman filter state estimate after the update on week  $t$ .

The total forecasting error is obtained by summing up country-wise errors over all the 28 considered countries:

$$E(\mathbf{T}) = \sum_{i=1}^{28} E_i(\mathbf{T}), \text{ where } E_i(\mathbf{T}) = \sum_{\tau=1}^4 \sum_{t+\tau \in \mathbf{T}} \left| \sqrt{y_i(t + \tau)} - \sqrt{\hat{y}_i(t + \tau|t)} \right|. \quad (14)$$

Here,  $\mathbf{T}$  denotes the time interval or a collection of time intervals used for evaluation. The square root is used as transformation to stabilize variance, so as to balance the performance metric between times of high and low incidence, and between countries with larger and smaller population size. Note that weeks  $t$  where data for country  $i$  are missing for weeks  $t - 3, \dots, t$ , are excluded from  $E_i$ .

As a reference to benchmark our predictions, we simply use the latest data point as a forecast for the next 4 weeks:

$$E_i^{(\text{ref})}(\mathbf{T}) = \sum_{\tau=1}^4 \sum_{t+\tau \in \mathbf{T}} \left| \sqrt{y_i(t + \tau)} - \sqrt{\bar{y}_i(t)} \right|$$

where  $\bar{y}_i(t) = y_i(t)$  if these data exist, or equals the value of the previous existing data point, if  $y_i(t)$  does not exist. We measure how much all different models, used for evaluation, improve the overall predictions w.r.t. the reference forecast, that is,

$$\text{Performance gain} = 1 - \frac{\sum_{i=1}^{28} E_i(\mathbf{T})}{\sum_{i=1}^{28} E_i^{(\text{ref})}(\mathbf{T})}, \quad (15)$$

calculated separately for the training data period and the test data period. In addition, to evaluate performance during the wave onset and peak periods, the errors  $E_j$  are collected over time intervals covering the onsets of all epidemic waves (except for 2020–2022, which were heavily affected by the COVID-19 pandemic). Here, we no longer considered the split between training and testing data. These time intervals were determined by identifying weeks  $W$  when the European-wide incidence crossed 120, and taking the interval from  $W - 8$  to  $W + 1$ . Additionally, we collect all time intervals around the peak incidence, identified by weeks  $\hat{W}$  with the highest incidence during each epidemic season, and take the interval  $\hat{W} - 5$  to  $\hat{W} + 4$  (see Fig. 3).

To evaluate probabilistic forecasts, 1–4-week ahead forecasts were generated every other week. They were evaluated using the weighted interval score (WIS), commonly used to evaluate probabilistic forecasts in epidemiology<sup>8,9,56</sup>, that measures both the accuracy and confidence on the forecasts by penalizing for the width of the obtained confidence intervals. Since the WIS scales directly with the scale of the data, we normalize the scores for each country by the average peak incidence of the epidemic waves in the respective country over the 10-year period considered.

## Parameter fitting

There are 11 tuning parameters in the network model, 10 in the mean-field model, and 8 in the isolated models (Supplementary Table 1), fitted with simulated annealing running the full EKF pipeline and using the total forecasting error  $E(\mathbf{T})$  defined in (14) as the cost function (Supplementary Table 2). Since the different models have different numbers of fitted parameters, the data are divided into training and test data for a fair comparison. Data from the first four epidemic seasons (2014–2018) are used for parameter fitting, that is,  $\mathbf{T} = \{1, \dots, 202\}$  in the cost function (until week 32 in 2018, see Fig. 3). The remaining six seasons (2018–2024), that is,  $\mathbf{T} = \{203, \dots, 499\}$ , are used for evaluations. To ensure convergence, 15 simulated annealing runs with randomized starting points were run. Details of the simulated annealing are presented in Supplementary Note 4.

Note that the cost function  $E(T)$  for parameter fitting is based on the noise-free forecasts  $\hat{y}_i(t + \tau|t)$ . If the covariance matrices  $Q$  and  $U$  and the initial uncertainty covariance  $P_0$  in the Kalman filter (see the algorithm in Supplementary Note 2) are multiplied by any positive constant, then the estimation error covariance  $P_t$ , state prediction error covariance  $\tilde{P}_t$ , and the output prediction error covariance  $S_t$  will be multiplied by the same constant at all times  $t$ . In the Kalman gain matrix  $\tilde{P}_t C(t) S_t^{-1}$ , however, the constant will cancel out due to the inverse of  $S_t$ . As a consequence, the state estimates  $\hat{x}(t)$  and the noise-free forecasts  $\hat{y}_i(t + \tau|t)$  remain unchanged by such covariance scaling (see Supplementary Note 3 for details). It is therefore not possible to identify all covariance parameters at once by fitting the noise-free forecasts to data, but only their relative magnitudes. Without loss of generality, we therefore fix  $\sigma_\beta$  in  $Q_i$  in (8) to  $\Delta t \mu^2$  to have a grounding value for the other covariance terms that are scaled by tuning parameters  $\kappa_1$  and  $\kappa_2$  in (8) and  $\rho_1$  and  $\rho_2$  in (10). Since the covariance terms can only be obtained up to constant multiplication by the parameter fitting procedure, a final noise calibration step will be carried out to find the proper (relative) scaling factor for the covariances.

### Noise calibration for probabilistic forecasts

The best way to produce probabilistic forecasts is to simulate trajectories from the full stochastic model and gather statistics from these simulations. However, as discussed above, the parameter fitting procedure only reveals the relative strengths of different noise components, but not the absolute noise level. For noise calibration, we therefore introduce a coefficient  $K$  used to multiply all noise terms in the model to control the uncertainty in the stochastic projections. As explained above, such noise scaling has no effect on the Kalman filter state estimate, since it does not affect the relative noise contributions. Hence, the parameters obtained by the fitting procedure can be used for probabilistic forecasts after the scaling. We simulate 1000 trajectories forward in time, initialized from a random initial state drawn from  $\mathcal{N}(\hat{x}(t), K^2 P_t)$ . Trajectories are then simulated forward in time using the stochastic model (6)–(7), where all noise terms  $w_j$  for  $j = 1, \dots, 4$  and  $w_\beta$  are scaled with  $K$ . Both state-dependent and state-independent measurement noise corresponding to (10), again multiplied by  $K$ , is then added to the projected case numbers.

To find a good value for  $K$ , we generated 4-week ahead probabilistic forecasts on 58 different weeks covering the last two epidemic seasons (every week during high incidence and every second week during low incidence). Statistical coverage plots were then created by generating several different percentiles for the projections, and then calculating the frequency of occurrences where the real incidence (for forecast weeks 1–4) was above each percentile (Supplementary Fig. 5). This was repeated with several different values of  $K$ . Note that while the precise choice of  $\sigma_\beta$  does not influence the end results, having a reasonable value helps evaluate the sanity of the calibration results. With  $\sigma_\beta = \Delta t \mu^2$ , a change of one standard deviation in  $\beta_i$  in a time unit (a week) corresponds to the basic reproduction number  $R_0 = \beta_i / \mu$  changing by 1. With such choice,  $K$  is expected to not deviate much from 1 (in orders of magnitude).

Due to nonlinear dynamics, the median of the 1000 trajectories does not coincide with the noise-free projection ( $K = 0$ ) that was used for tuning the method. To account for potential bias, we also tried a scheme where all the simulated percentiles were adjusted by adding the difference of the noise-free projection and the median of the noisy simulations. From Supplementary Fig. 5, it can be observed that this adjustment indeed helps mitigate bias in the forecasts. Following the calibration result,  $K = 0.85$  was selected as the reference value for the adjusted probabilistic forecasts.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Results

### Including network effects improves forecasting performance

Similarly to earlier works such as Influcast<sup>8</sup> and Respicast<sup>9</sup>, we evaluated forecasting performance using 1–4-week-ahead predictions for ILI across

Europe. Forecasting performance was assessed from the adjusted average error between model predictions and actual data. The performance gain defined in (15) was quantified as the percentual reduction in error compared to a reference forecast, which uses the most recent observed incidence as the future prediction. This evaluation was conducted over 10 years of data, with separate analyses for training (first four epidemic seasons) and test datasets (remaining six epidemic seasons) to foster a fair comparison between different models with different numbers of fitted parameters, and for critical periods, namely wave onsets and epidemic peaks (as identified in Fig. 3).

The performance of each model—network model, mean-field model, and isolated models for 1–4-week-ahead forecasts—is shown in Fig. 4. The network and mean-field models consistently outperformed isolated models across all periods, highlighting the benefit of using trans-regional models across connected regions. The augmented models also prove useful to predict key moments of epidemic progression: peaks and onsets. Around epidemic peaks, the network and mean-field models achieved up to a 25% improvement in forecasting performance. Wave onset remains more difficult to forecast; nonetheless, in these periods, the network model provides the best results, clearly outperforming both the mean-field and isolated models. Knowing the mobility patterns among neighboring countries is thus particularly valuable for anticipating wave onsets.

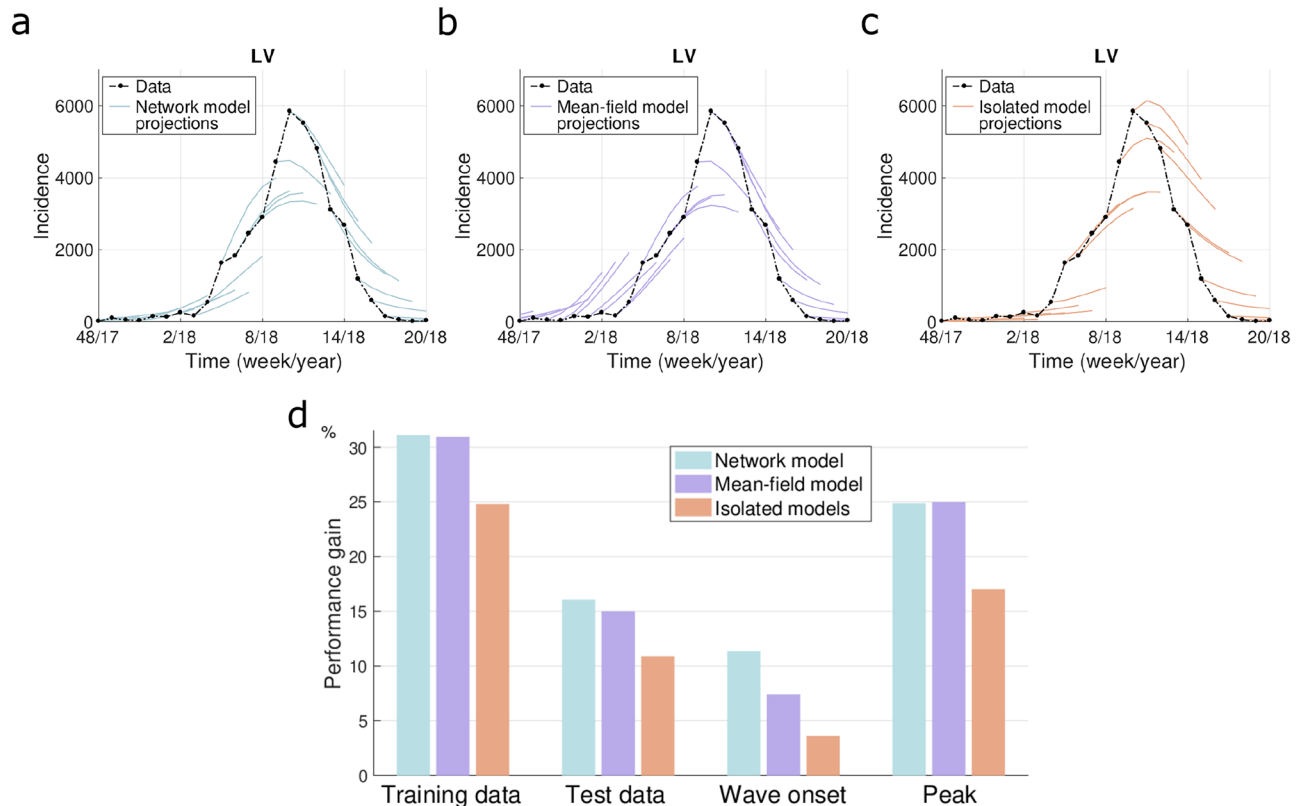
Performance gains varied by region and epidemic season (Fig. 5). The network model provided the highest gains in countries with strong cross-border interactions, such as Austria and Slovakia. In countries with a larger population size, performance tends to be slightly better overall, while the difference between the network model and isolated models is higher in countries with a smaller population size and with higher effect from the outside (see Supplementary Fig. 10). Predictions were more accurate during pre-pandemic seasons (2018–2019) compared to COVID-affected years, where reduced mobility and reporting inconsistencies impacted performance: since the mobility network is assumed to be static, it was less accurate when travel restrictions were in place. Also, the adaptive estimation scheme for the detection parameters  $c_i(t)$  may be inadequate in extreme situations. Future improvements could incorporate a dynamic network evolution to reflect real-time changes in mobility patterns.

We also fitted model parameters using all 10 years of data, to check whether providing longer time series improved model predictions. The performance did not change considerably, and the same drop in the performance gain for test data period compared to training data period could still be observed. This confirms that the pandemic was the main cause of this performance drop, rather than potential overfitting. A similar performance was obtained even when only the first 2 years of data were used for parameter fitting. This is indicative of the parsimony of our model, and suggests that some country-specific parameters might be fitted based on forecasting performance without risk of overfitting. Results of these experiments are in Supplementary Fig. 7.

### Including network effects improves forecasting with missing data

Delays or missing values in reporting data are fairly common<sup>9</sup>. It is therefore crucial for an epidemic model to provide robust forecasts, even with missing data. Our network-based models leverage information from interconnected regions to impute missing data and provide accurate predictions, even when substantial portions of the dataset are unavailable. Here we study simplified missing data scenarios with randomly deleted data entries. While this setup does not exactly correspond to the observed reporting delays, it nevertheless indicates differences in performance for the different models in the presence of missing data. We simulated missing data scenarios by randomly deleting 20%, 40%, 60%, and 80% of the ILI data across all countries. Data that followed a period of at least 3 weeks of missing data were “protected” from deleting. Each scenario was repeated 20 times to account for variability. Forecasting performance was evaluated by the average forecasting error, obtained from the total error  $E(T)$  defined in (14) by normalizing by the number of summands, throughout the 10 years considered,  $T = \{1, \dots, 499\}$ . In addition, we measured how well the deleted data could be reconstructed





**Fig. 4 | Forecasting results of the different models. a, b, c** Comparison of 1–4-week-ahead forecasts by the network model, mean-field model, and isolated model, respectively for the 2017–2018 epidemic season in Latvia. The network model accurately captures the wave onset, outperforming the other models, which either underestimate or misalign the timing of the initial wave. In this epidemic season, the epidemic wave reached the Baltic states later than elsewhere in Europe (cf. Fig. 3),

which explains the differences in the forecasts. Supplementary Fig. 4 shows the estimated  $R_0(t) = \beta(t)/\mu$  from different models for Latvia. **d** Performance gain of each model compared to the reference forecast, across training and test data (all countries), as well as during wave onset and peak periods. These results highlight the largely superior predictive accuracy of the network model and the mean-field model compared to isolated models, particularly during critical epidemic phases.

using all available data until the reconstruction time. To enable direct comparison with forecasts, the same error metric is used, which is obtained by summing up (and normalizing by the number of summands) the reconstruction errors over all deleted data entries:

$$E_{\text{rec}} = \sum_{\text{Data } (i,t) \text{ deleted}} \left| \sqrt{y_i(t)} - \sqrt{\hat{y}_i(t|t)} \right|.$$

As shown in Fig. 6a, the forecasting performance of the network-based model is robust to increasing percentages of missing data: prediction errors are rising only modestly. For example, with 40% data missing, the error increased by just 11.6%, compared to 15.1% for isolated models. The network model thus shows its ability to utilize inter-country dynamics to compensate for missing information. The reconstructed data closely aligned with true infection trends. Notably, reconstruction errors remained significantly lower than forecasting errors.

Moreover, the performance gain of the network model over the reference forecast increased with higher deletion levels, rising from 23 to 27.5% as data availability decreased from 100 to 40% (Fig. 6b), while the performance gain of the isolated models remained roughly constant at around 17%. The network model is also more consistent in its gain than the mean-field and isolated models, as appreciated by considering the variance in Fig. 6b. For reference, Fig. 6c shows an example for three epidemic seasons in Austria with 60% of the data deleted.

Our findings demonstrate the value of integrating information from neighboring countries, particularly in scenarios with incomplete datasets, and hence the utility of network-based models for improving forecasts and

imputing missing data, which is particularly relevant for epidemic monitoring in regions with uneven surveillance or reporting practices.

### Probabilistic forecasts

Uncertainty quantification provided by the EKF allows the network-based model to provide valuable insight into the potential future trajectory of epidemics, accounting for inherent uncertainties and generating confidence intervals and probabilistic scenarios to guide decision-making.

Fig. 7a shows the normalized WIS for the 1–4-week ahead forecasts. The scores are first averaged over the entire 10-year period and then normalized by country with the average peak height over the 10 years. The statistics over all countries are shown by box plots. Supplementary Fig. 5 shows statistical coverage plots for the probabilistic forecasts, while Supplementary Fig. 6 shows separate box plots for each epidemic season. Fig. 7b shows examples of probabilistic forecasts for France during the 2023–2024 epidemic season. The model provides point estimates and different confidence intervals, capturing the uncertainty in the disease progression. The forecasts demonstrate the ability of the network model to anticipate epidemic trends, including wave peaks and durations, while offering actionable ranges for public health planning.

### Insights into model mechanisms and interpretability

Inspecting the results of parameter fitting in Supplementary Table 2, we notice that  $k_1$  is relatively big, strongly driving  $\beta$  towards the reference value (see (7)), in particular with the network models, where the reference value is heavily influenced by the network ( $k_2 \approx 0.8$  in all network cases). Some parameters do not seem to be very well identifiable, resulting in large differences in the parameter values when the training is done using the first 4



**Fig. 5 | Detailed forecasting results. a** Performance gain for each country over the entire 10-year period. **b** Season-by-season performance gains.

years of data or all 10 years of data, while the performance did not change significantly. In particular,  $\gamma_2$  changes considerably in the network model, indicating that the inclusion of the neighbor network is not very important for forecasting performance.

Sensitivity analysis for the fitted parameters is provided in Supplementary Figs. 8 and 9 for the network model, also revealing a fairly weak sensitivity to changes in  $\gamma_2$ . In addition, the results are not very sensitive to  $\kappa_2$  and  $\rho_2$  that regulate the state-independent noise terms in the covariances  $Q_i$  in (8) and  $U_i$  in (10). These terms were introduced to provide a lower bound for covariances during times of low incidence and thereby improve responsiveness at the onset of a new wave. In the isolated models, the values of these parameters are significantly higher, and indeed, setting  $\kappa_2 = 0$  makes the performance gain for the wave onset periods drop from 3.61 to 0.98% and setting  $\rho_2 = 0$  makes it drop to 1.48%. Setting both  $\kappa_2 = \rho_2 = 0$  would make the EKF unstable for isolated models, since the measurement prediction error covariance  $S_t$  becomes at least nearly singular during times of low incidence (see the Algorithm in the Supplementary Note 2).

The mechanistic design of the model allows for a detailed inspection of the factors driving its performance. To better understand these dynamics, we conducted additional experiments, whose results are shown in Supplementary Fig. 7. Firstly, the network model improves performance more for countries where the outside effect is larger (Supplementary Fig. 10c). To investigate whether this result is merely spurious, we ran the method after scaling the network matrices  $G^{(1)} + G^{(2)}$ , and  $G^{(3)}$  row-wise to a fixed percentage of the populations of the respective countries. The forecasting performance decreased slightly compared to the network model, suggesting that the network model accurately captures the true effect of mobility. Secondly, the inclusion of network effects in the dynamics of the transmission parameter  $\beta$  was found to play a crucial role in enhancing forecasting capabilities. When this effect was removed by setting  $k_2 = 0$  in (7), the model's performance dropped markedly. Finally, we tested a hybrid model where the mean-field network informed only the  $\beta$ -parameter dynamics,

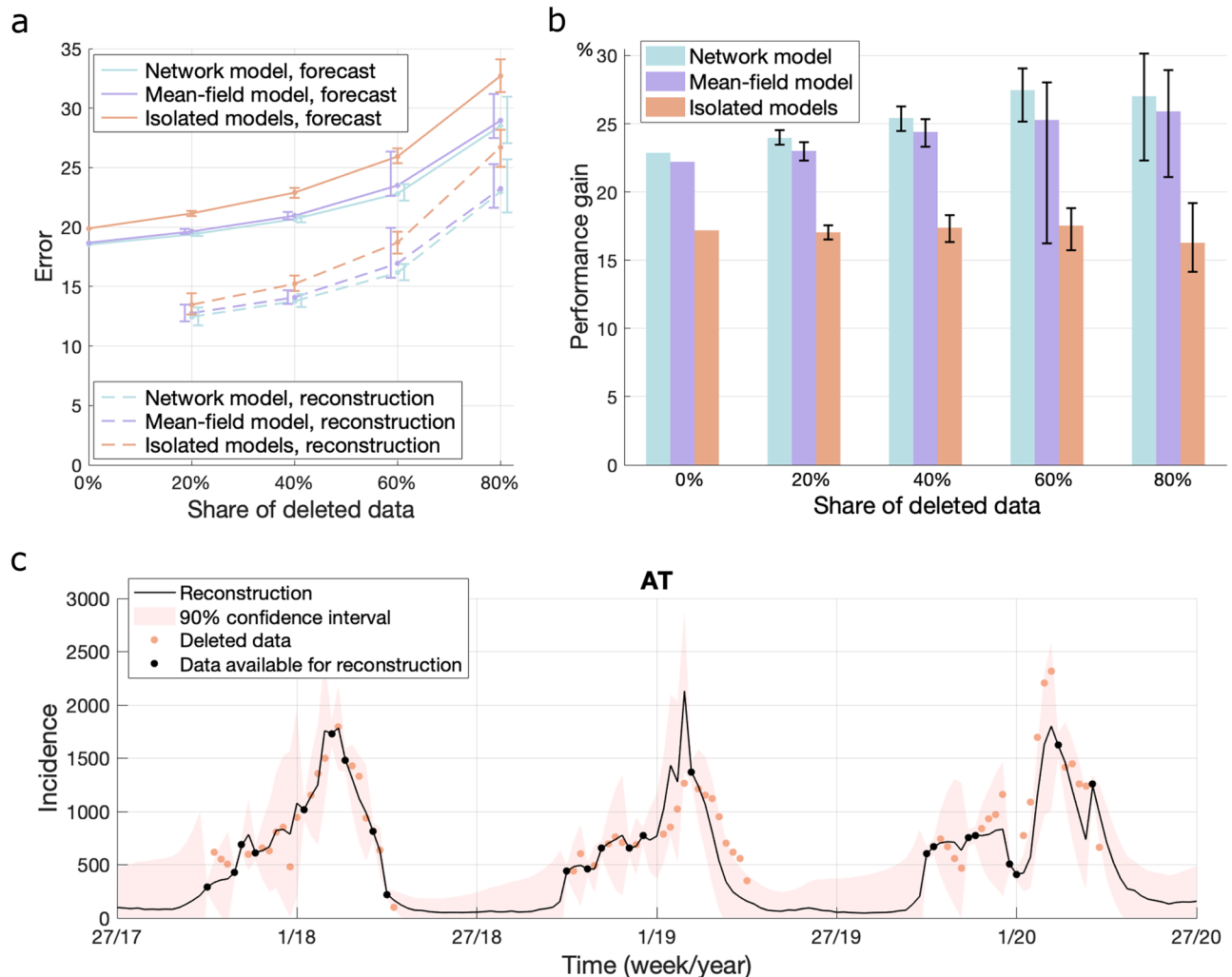
while the  $S \rightarrow I$  dynamics were governed by isolated models. The performance of this model was mostly comparable to the full network model. This observation suggests that including network effects into the  $\beta$ -parameter dynamics has a stronger impact than including the mobility component into the  $S \rightarrow I$  dynamics. This is also supported by a fairly small sensitivity to  $\gamma_1$  and  $\gamma_2$  parameters compared to the sensitivity to the  $k_j$  parameters in the  $\beta$ -dynamics. Note that  $\gamma_j$ -parameter magnitudes are normalized out in the  $\beta$ -dynamics (7), and therefore the sensitivity to  $\gamma_j$ -parameters does not reflect the relevance of the network component in (7).

Interestingly, the detailed network model outperformed all alternatives during wave onset periods, highlighting its strength in capturing the early dynamics of epidemic waves.

## Discussion

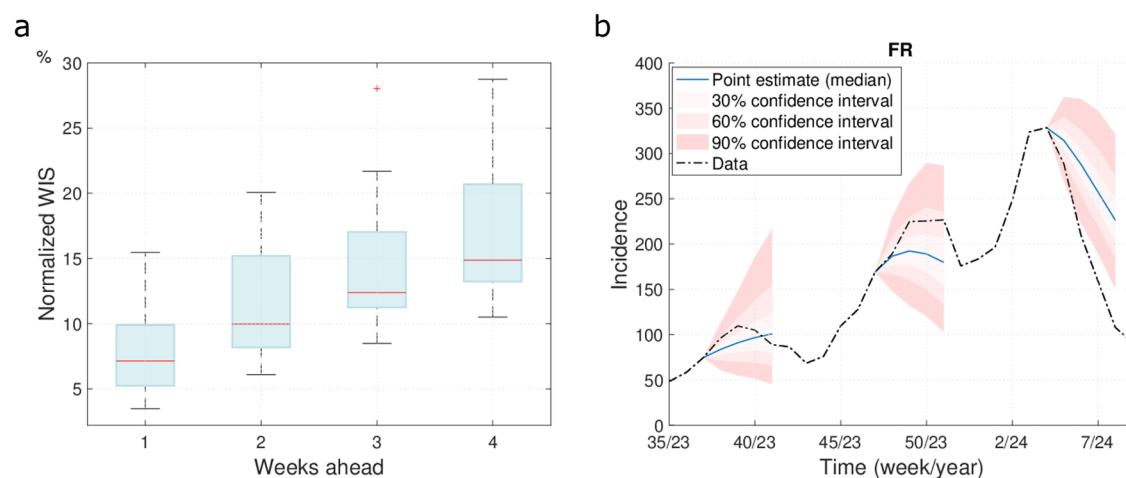
This work shows the advantages of a networked metapopulation SIRS model, integrated with an EKF, for forecasting different aspects of the dynamics of ILI across Europe. The model leverages empirical data on mobility and disease incidence to incorporate both local dynamics and cross-border interactions and thus improve forecasting accuracy. Comparative analyses show that the network-based model provides a mild advantage in forecasting epidemic progression, compared to mean-field models, but outperforms isolated and mean-field models during critical periods such as wave onsets and epidemic peaks. Additionally, the network model is more robust to missing data and provides reliable forecasts even with substantial data unavailability.

Our findings highlight that integrating data from several connected regions into network-based epidemic models improves accuracy and interpretability. In a probabilistic forecasting framework, the EKF enables the model to adapt dynamically to new data, providing real-time forecasts with quantified uncertainty that is critical for proactive responses. Our proposed framework is highly relevant for epidemic monitoring and regional and global public health decision-making, with several applications



**Fig. 6 | Performance of the models with different percentages of missing data.** Confidence intervals (error bars) show the 90th percentiles over different randomized evaluations. **a** Average error for different percentages of deleted data; the network and mean-field models consistently outperform the isolated models, the error remaining relatively low even with 80% missing data. **b** Performance gain

average with missing data compared to the reference forecast. **c** Example of data reconstruction for Austria over three epidemic seasons with 60% of the data deleted. The network model effectively reconstructs the missing data while maintaining confidence intervals. Observe the increasing width of the confidence interval in periods when data are missing.



**Fig. 7 | Evaluation and examples of probabilistic forecasts.** **a** Average (over time) weighted interval scores (WIS) for 1–4-week ahead forecasts, normalized by average epidemic peak heights in each country. The boxplots show the median (over the 28 countries) with the red line, the 50th percentile with the box, and the minimum and

maximum values with the whiskers, excluding outliers shown with red crosses. **b** Three 4-week-ahead probabilistic forecasts for France during the epidemic season 2023–2024.

in forecasting. By providing a spectrum of plausible epidemic trajectories, it supports scenario analysis and planning, including: (1) Early warnings: identifying potential wave onsets and peak periods with quantified uncertainty; (2) Resource allocation: informing healthcare capacity planning by projecting worst-case and best-case scenarios; (3) Scenario modeling: evaluating the impact of interventions, such as travel restrictions or vaccination campaigns, by simulating alternative futures.

The demonstrated benefits of network-based approaches are expected to influence the design of future epidemic forecasting systems, particularly in interconnected regions. Our results encourage further investigation and the adoption of similar frameworks for diseases such as COVID-19, dengue, and seasonal influenza (after appropriately incorporating adequate models into the Kalman filter pipeline). The reconstructed European mobility network can also serve as a foundation for further studies, fostering collaboration and data sharing among public health agencies.

Static network structures limit the accuracy of predictions during periods of rapid mobility changes, such as those caused by pandemics or natural disasters. Incorporating dynamic network adjustments, informed by real-time mobility data, could enhance model adaptability and performance. Our pipeline also implicitly assumes that data collection is consistent between epidemic seasons. The adaptive scheme for estimating the detection parameters  $c_i(t)$  may not be able to cope with drastic changes that are sometimes observed in the data. Any knowledge of changes can be manually incorporated to improve performance. Also, we did not fit parameters that are specific to countries or population demographics, based on forecasting performance. According to the country-specific parameter sensitivity analysis (Supplementary Fig. 9), doing so may lead to additional improvement.

Moreover, the model currently assumes homogeneous mixing within populations, which may oversimplify complex interactions, particularly in urban areas. It can be observed (Fig. 5a and Supplementary Fig. 10) that the performance is on average slightly better for countries with larger population than with smaller population (although the difference is mainly driven by the good performance for Italy and Spain), but the performance improvement of the network model compared to isolated models is generally smaller for larger populations. In countries with larger population, the data are averaged over larger numbers, and are typically more regular than in countries with smaller population. While regular data improve forecasting performance in general (like for Italy and Spain), there may be large regional differences within bigger countries, and more targeted forecasts could be more useful. Including network effects tends to improve the performance more significantly when the population size is small because the relative proportion of external (vs. internal) contributions to contagion (which depends on the amount of daily travelers across the borders, including cross-border workers) is likely to be larger in geographically small countries, which typically have a smaller population size; hence, the outcome does not simply depend on the population size, but on multi-pronged factors that implicitly account also for geographical size and intensity of exchanges with neighboring regions. While it is likely that the effect of mobility would be stronger with finer spatial granularity, developing interconnected models for given regions and studying the effect of spatial granularity are two distinct issues, and also the latter would be worth exploring in the future. Future work could integrate finer-grained data, such as inter-city mobility patterns, or other complementary data such as holiday schedules like in ref. 40, or humidity data like in refs. 5,29,34,40. Additionally, the absence of certain countries from the network, due to data unavailability, may introduce regional biases. Expanding public databases and standardizing data collection practices across Europe would address this issue.

Lastly, while the model performs well for ILI, its applicability to other diseases with different transmission dynamics, such as vector-borne or chronic diseases, remains to be explored. Adjusting the framework to accommodate diverse epidemiological characteristics would broaden its impact.

## Data availability

The 10 years of ILI incidence data (detected cases over time, with an average weekly reporting frequency) for 31 regions used for the results in this article originate from the Respicast platform<sup>10</sup>. To construct the network models, long-range mobility traffic data (number of flights and ferry travels) are publicly accessible from refs. 43–45; cross-border commuting information (flows over time) is publicly accessible at<sup>46,47</sup>; population densities of NUTS 3 regions are publicly available from refs. 48,49. A copy of the incidence data and the processed network components are provided along with the code. A separate figure source file is available online as Supplementary Data.

## Code availability

All code required to reproduce the results presented here is available at [github.com/uniluxembourg/lcsb/systems-control/epinetekf](https://github.com/uniluxembourg/lcsb/systems-control/epinetekf) together with a user guide<sup>57</sup>.

Received: 6 March 2025; Accepted: 13 April 2026;

Published online: 08 May 2026

## References

1. Lutz, C. S. et al. Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. *BMC Public Health* **19**, 1–12 (2019).
2. Wang, Y., Wei, Z. & Cao, J. Epidemic dynamics of influenza-like diseases spreading in complex networks. *Nonlinear Dyn.* **101**, 1801–1820 (2020).
3. Proverbio, D., Kemp, F. & Gonçalves, J. Early warning of SARS-CoV-2 infection. In *Features, Transmission, Detection, and Case Studies in COVID-19* 13–24 (Elsevier, 2024).
4. Siegenfeld, A. F. & Bar-Yam, Y. The impact of travel and timing in eliminating COVID-19. *Commun. Phys.* **3**, 204 (2020).
5. Shaman, J., Karspeck, A., Yang, W., Tamerius, J. & Lipsitch, M. Real-time influenza forecasts during the 2012–2013 season. *Nat. Commun.* **4**, 2837 (2013).
6. Davis, J. T. et al. Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. *Nature* **600**, 127–132 (2021).
7. King, A. A., Ionides, E. L., Pascual, M. & Bouma, M. J. Inapparent infections and cholera dynamics. *Nature* **454**, 877–880 (2008).
8. Fiandrino, S. et al. Collaborative forecasting of influenza-like illness in Italy: the Influcast experience. *Epidemics* **50**, 100819 (2025).
9. Gozzi, N. et al. Performance evaluation of RespiCast ensemble forecasts for primary care syndromic indicators of viral respiratory disease in Europe. *medRxiv* 2025.10.30.25339155 (2025).
10. Respicast—European respiratory diseases forecasting hub. Accessed 23 September 2024. <https://respicast.ecdc.europa.eu/>.
11. Biggerstaff, M. et al. Results from the centers for disease control and prevention’s predict the 2013–2014 influenza season challenge. *BMC Infect. Dis.* **16**, 1–10 (2016).
12. Dehning, J. et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**, eabb9789 (2020).
13. Citron, D. T. et al. Comparing metapopulation dynamics of infectious diseases under different models of human movement. *Proc. Natl. Acad. Sci. USA*. **118**, e2007488118 (2021).
14. Apolloni, A., Poletto, C., Ramasco, J. J., Jensen, P. & Colizza, V. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theor. Biol. Med. Mod.* **11**, 1–26 (2014).
15. Ni, S. & Weng, W. Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks. *Phys. Rev. E* **79**, 016111 (2009).
16. Xia, Y., Bjørnstad, O. N. & Grenfell, B. T. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am. Nat.* **164**, 267–281 (2004).



17. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
18. Alutto, M., Cianfanelli, L., Como, G. & Fagnani, F. On the dynamic behavior of the network SIR epidemic model. *IEEE Trans. Control Netw. Syst.* **12**, 177–189 (2024).
19. Mistry, D. et al. Inferring high-resolution human mixing patterns for disease modeling. *Nat. Commun.* **12**, 323 (2021).
20. Sahneh, F. D., Scoglio, C. & Van Mieghem, P. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM T Net.* **21**, 1609–1620 (2013).
21. Bichara, D. & Iggidr, A. Multi-patch and multi-group epidemic models: a new framework. *J. Math. Biol.* **77**, 107–134 (2018).
22. Akuno, A. O. et al. Multi-patch epidemic models with partial mobility, residency, and demography. *Chaos Solit. Fractals* **173**, 113690 (2023).
23. Lahodny, G. E. & Allen, L. J. S. Probability of a disease outbreak in stochastic multipatch epidemic models. *Bull. Math. Biol.* **75**, 1157–1180 (2013).
24. Ruan, Z., Wang, C., Ming Hui, P. & Liu, Z. Integrated travel network model for studying epidemics: interplay between journeys and epidemic. *Sci. Rep.* **5**, 11401 (2015).
25. Harper, R. & Tee, P. Balancing capacity and epidemic spread in the global airline network. *Appl. Net. Sci.* **6**, 94 (2021).
26. Soriano-Paños, D. et al. Modeling communicable diseases, human mobility, and epidemics: a review. *Ann. Phys.* **534**, 2100482 (2022).
27. Della Rossa, F. et al. A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. *Nat. Commun.* **11**, 5106 (2020).
28. Pizzuti, C., Socievoli, A., Prasse, B. & Van Mieghem, P. Network-based prediction of COVID-19 epidemic spreading in Italy. *Appl. Net. Sci.* **5**, 1–22 (2020).
29. Pei, S., Kandula, S., Yang, W. & Shaman, J. Forecasting the spatial transmission of influenza in the United States. *Proc. Natl. Acad. Sci. USA.* **115**, 2752–2757 (2018).
30. Perini, M. et al. Modelling COVID-19 in the North American region with a metapopulation network and Kalman filter. *Epidemics* **50**, 100818 (2025).
31. Zeng, Q. et al. Dengue transmission dynamics prediction by combining metapopulation networks and Kalman filter algorithm. *PLOS Negl. Trop. Dis.* **17**, e0011418 (2023).
32. Gatto, M. et al. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl. Acad. Sci. USA.* **117**, 10484–10491 (2020).
33. Pluchino, A. et al. A novel methodology for epidemic risk assessment of COVID-19 outbreak. *Sci. Rep.* **11**, 5304 (2021).
34. Kramer, S. C., Pei, S. & Jeffrey, S. Forecasting influenza in Europe using a metapopulation model incorporating cross-border commuting and air travel. *PLoS Comput. Biol.* **16**, e1008233 (2020).
35. Cazelles, B. & Chau, N. P. Using the Kalman filter and dynamic models to assess the changing HIV/AIDS epidemic. *Math. Biosci.* **140**, 131–154 (1997).
36. Kalivianakis, M., Mous, S. L. J. & Grasman, J. Reconstruction of the seasonally varying contact rate for measles. *Math. Biosci.* **124**, 225–234 (1994).
37. Proverbio, D. et al. Model-based assessment of COVID-19 epidemic dynamics by wastewater analysis. *Sci. Total Environ.* **827**, 154235 (2022).
38. Proverbio, D. et al. Dynamical SPQIEIR model assesses the effectiveness of non-pharmaceutical interventions against COVID-19 epidemic outbreaks. *PLoS ONE* **16**, e0252019 (2021).
39. Giordano, G. et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **26**, 855–860 (2020).
40. Ben-Nun, M., Riley, P., Turtle, J., Bacon, D. P. & Steven, R. Forecasting national and regional influenza-like illness for the USA. *PLoS Comput. Biol.* **15**, e1007013 (2019).
41. Kemp, F. et al. Modelling COVID-19 dynamics and potential for herd immunity by vaccination in Austria, Luxembourg and Sweden. *J. Theor. Biol.* **530**, 110874 (2021).
42. Prasse, B. & Van Mieghem, P. Network reconstruction and prediction of epidemic outbreaks for general group-based compartmental epidemic models. *IEEE Trans. Netw. Sci. Eng.* **7**, 2755–2764 (2020).
43. Eurostat, Air passenger transport between reporting countries. Accessed 22 August 2024. [https://ec.europa.eu/eurostat/databrowser/view/avia\\_paocc\\_\\_custom\\_12644069/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/avia_paocc__custom_12644069/default/table?lang=en). [https://doi.org/10.2908/AVIA\\_PAOCC](https://doi.org/10.2908/AVIA_PAOCC).
44. Finnish Ports Association, statistics. Accessed 23 August 2024. <https://www.finnishports.fi/eng/statistics/trade-information/port-of-helsinki-top-in-europe/>.
45. Sea passenger statistics: All routes 2019, UK Government, Department for Transport. Accessed 23 August 2024. <https://assets.publishing.service.gov.uk/media/5faaa73e8fa8f5789445b397/sea-passenger-statistics-all-routes-2019.pdf>.
46. European Commission, Social Affairs Directorate-General for Employment, and Inclusion. Cross-border work in EU and EFTA countries. (2024).
47. McGuinness, S., Bergin, A. & Devlin, A. A study of cross-border working on the island of Ireland. *ESRI Working Paper* 781 (ESRI, 2024).
48. Eurostat, NUTS—Nomenclature of territorial units for statistics. Accessed 28 August 2024. <https://doi.org/10.2785/714519>.
49. Eurostat, Population density by NUTS 3 region. Accessed 28 August 2024. [https://ec.europa.eu/eurostat/databrowser/view/demo\\_r\\_d3dens/default/table](https://ec.europa.eu/eurostat/databrowser/view/demo_r_d3dens/default/table). [https://doi.org/10.2908/DEMO\\_R\\_D3DENS](https://doi.org/10.2908/DEMO_R_D3DENS).
50. Gillespie, D. The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306 (2000).
51. Kalman, R. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960).
52. Anderson, B.D.O. & Moore, J.B. *Optimal Filtering* (Prentice-Hall, 1979).
53. European Centre for Disease Prevention and Control. European respiratory virus surveillance summary (ERVIS). Accessed 23 September 2024. <https://erviss.org/> (2024).
54. World Health Organization, Global influenza programme, FluID. Accessed 23 September 2024. <https://www.who.int/teams/global-influenza-programme/surveillance-and-monitoring/fluid> (2024).
55. Council for International Organizations of Medical Sciences (CIOMS). *International Ethical Guidelines for Epidemiological Studies* (CIOMS, 2009).
56. Bracher, J., Ray, E. L., Gneiting, T. & Reich, N. G. Evaluating epidemic forecasts in an interval format. *PLoS Comput. Biol.* **17**, e1008618 (2021).
57. Aalto, A., Proverbio, D., Giordano, G., Skupin, A. & Gonçalves, J. EpiNetEKF: Code and data to accompany the article “Networked SIRS model with Kalman filter state estimation for epidemic monitoring in Europe”. In: Zenodo [Internet]. <https://doi.org/10.5281/zenodo.18671513> (2026).

## Acknowledgements

The authors wish to express their gratitude to the team behind the RespiCast hub. A.A. was supported by the Luxembourg Government through the CoVaLux programme (16954531). D.P. and G.G. acknowledge support from the European Union through the ERC INSPIRE grant (project number 101076926). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the European Research Council Executive Agency can be held responsible for them.

## Author contributions

A.A.: Conceptualization, Methodology, Formal analysis, Visualization, Investigation, Software, Validation, Data curation, Writing—Original draft, Writing—Review and editing. D.P.: Conceptualization, Methodology, Visualization, Validation, Writing—Original draft, Writing—Review and editing. G.G.: Conceptualization, Writing—Review and editing, Supervision, Funding acquisition. A.S.: Writing—Review and editing, Supervision, Funding acquisition. J.G.: Writing—Original draft, Writing—Review and editing, Supervision, Funding acquisition.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at

<https://doi.org/10.1038/s43856-026-01611-9>.

**Correspondence** and requests for materials should be addressed to Atte Aalto.

**Peer review information** *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026