PUBLIC HEALTH

Forecast skill assessment of an operational continental heat-cold-health forecasting system: New avenues for health early warning systems

Marcos Quijal-Zamorano[1,2]*, Desislava Petrova[1], Èrica Martínez-Solanas[1,3], François R. Herrmann[4], Xavier Rodó[1,5], Jean-Marie Robine[6,7], Marc Marí-Dell'Olmo[8,9,10], Hicham Achebak[1,11], Joan Ballester[1]

More than 110,000 Europeans died as a result of the record-breaking temperatures of 2022 and 2023. A new generation of impact-based early warning systems, using epidemiological models to transform weather forecasts into health forecasts for targeted population subgroups, is an essential adaptation strategy to increase resilience against climate change. Here, we assessed the skill of an operational continental heat-cold-health forecasting system. We used state-of-the-art temperature-lag-mortality epidemiological models to transform bias-corrected ensemble weather forecasts into daily temperature-related mortality forecasts. We found that temperature forecasts can be used to issue skillful forecasts of temperature-related mortality. However, the forecast skill varied by season and location, and it was different for temperature and temperature-related mortality due to the use of epidemiological models. Overall, our study demonstrates and quantifies the forecast skill horizon of heat-cold-health forecasting systems, which is a necessary step toward generating trust among public health authorities and end users.

INTRODUCTION

Every year, more than 300,000 premature deaths are related to ambient temperatures in Europe (*1*). In the current climate, the impact of cold on human mortality surpasses that of heat by a factor of 10 (*2*), but in the absence of mitigation and adaptation actions, climate change projections point to the reversal of the seasonality of temperature-related mortality (TRM), with rapidly increasing deaths due to extreme heat (*3*). To reduce the negative consequences of climate change on human health (*4*), European societies have already started to implement adaptation measures (*5*–*7*). Thus, after the extreme heat in the summer of 2003 that caused 70,000 excess deaths in western Europe (*8*), several national and regional meteorological and public health agencies started to design and implement heat early warning systems.

However, the death toll arising from the record-breaking temperatures in 2022 and 2023, with more than 110,000 heat-related deaths in the continent (*1*, *9*, *10*), emphasized again the need to further strengthen existing emergency and resilience plans, including the development of a new generation of heat-cold-health early warning systems (*11*). Existing operational systems generally rely only on physical data from operational weather forecasts to activate public health actions, and therefore, they account neither for the real impacts on the exposed populations nor for the inequalities in vulnerability to heat and cold across different population subgroups, which is key to protect vulnerable populations (*12*). Current protocols, methodologies, and warning-issuing criteria largely vary from one country to another, but the vast majority of them are actually entirely based on regional thresholds of temperature and, in a few cases, on other climate variables (e.g., humidity) (*13*), combined indices (e.g., apparent temperature) (*13*), or even thermophysiological parameters (*14*).

Determining the forecast skill horizon of any operational forecasting system is key to issue early warnings that generate trust among end users. Over the past decades, numerical weather prediction (*15*) has shifted from a deterministic (single numerical integration) to a probabilistic approach (ensembles of numerical integrations with perturbed initial conditions). This change has allowed a substantial improvement of forecast skill horizons beyond the previously thought 2-week limit arising from the chaotic nature of the atmosphere (*16*, *17*). To issue early warnings directly targeting the impacts of temperatures on human health, epidemiological models need to be used to transform weather forecasts into forecast of health-related variables, e.g., mortality or morbidity (*18*–*20*). It yet remains to be explored whether this transformation preserves the skill horizon found in weather forecasting systems for physical variables such as temperature. This skill assessment of health-related forecasts requires the transdisciplinary integration of weather forecasts and techniques with health datasets and epidemiological models to develop and evaluate health early warning systems directly targeting impact-based health variables (*21*).

The overarching goal of the present study is to analyze and explain the differences in the forecast skill of temperature and TRM forecasts in Europe, both focusing on winter and summer months, as a necessary step toward the release of Forecaster.Health (https://forecaster.health/), an operational, open-access, continental heat-cold-health early warning system (*22*, *23*).

RESULTS

Figure 1A shows the cumulative exposure-response association accumulated over the range of lags in southern, western, and central

[1]ISGlobal, Barcelona, Spain. [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [3]Sub-Directorate General of Surveillance and Response to Public Health Emergencies, Public Health Agency of Catalonia, Generalitat of Catalonia, 08005 Barcelona, Spain. [4]Division of Geriatrics, Department of Rehabilitation and Geriatrics, Geneva University Hospitals and University of Geneva, Thônex, Switzerland. [5]ICREA, Barcelona, Spain. [6]Institut National de la Santé et de la Recherche Médicale (INSERM), Montpellier, France. [7]École Pratique des Hautes Études, Paris, France. [8]Agència de Salut Pública de Barcelona (ASPB), Pl. Lesseps 1, 08023 Barcelona, Spain. [9]Institut de Recerca Sant Pau (IR SANT PAU), Sant Quintí 77-79, 08041 Barcelona, Spain. [10]Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Av. Monforte de Lemos, 3-5, 28029 Madrid, Spain. [11]Inserm, France Cohortes, Paris, France.
*Corresponding author. Email: marcos.quijal@isglobal.org

Europe, and fig. S1 shows the association for the ensemble of 147 European regions here analyzed (see Materials and Methods). The relative risk (RR) of death is found to monotonically increase for temperatures above and below the minimum mortality temperature (MMT), which in the ensemble of regions analyzed, is always found between percentiles 78 and 93 of the respective distribution of daily regional temperatures (fig. S2). The cold part of the association therefore covers a wider range of temperatures, although the increase in the RR per 1°C is nonlinear and much steeper in the warm tail. Figure 1 (B and C) additionally depicts the lag-response association at the 1st and 99th temperature percentiles, respectively, showing how the RR at extreme cold and warm temperatures is differently distributed along the range of lags. Thus, the risk of death from warm temperatures is immediate and acute and generally does not last for more than 5 days, while the effect of cold temperatures starts 1 or 2 days later, and it is distributed along a wider range of lags of a few weeks. Figure 1 (D to F) depicts the regional differences in the MMT and the RR at the 1st and 99th temperature percentiles, respectively. The regional MMTs are latitudinally distributed, i.e., the warmer the climate of a region, the warmer its optimum temperature, showing the effect of long-term adaptation to climatological conditions (Fig. 1D) (*5*, *24*). Moreover, the risk of death is relatively homogeneous for cold temperatures (Fig. 1E), while a latitudinal pattern emerges for heat with higher RR values in southern Europe (Fig. 1F).

We used the regional epidemiological associations in fig. S1 to transform the daily time series of observed and forecast temperatures into a time series of observed and forecast TRM values at lead times of up to 15 days. Figure 2 shows the associations and the square of the Pearson correlations (R^2) between daily observations and ensemble mean forecasts aggregated over the ensemble of regions at different lead times. Correlations are found to monotonically decay as a function of the forecast lead time, and in all cases, they are higher in winter than in summer. The difference in correlation between winter and summer is found to increase with the lead time and to be generally higher for TRM. The rather small correlation differences between winter and summer temperatures at short lead times are found to be amplified by the decay of the forecast skill at larger lead times [cf. R^2 values in Fig. 2 (A to C)] as well as by the added complexity and diversity of the regional epidemiological associations transforming temperatures into TRM values [cf. Fig. 2 (A to C with D to F)].

In subsequent figures, we explore how these regional epidemiological transformations differently modify the forecast skill horizon,

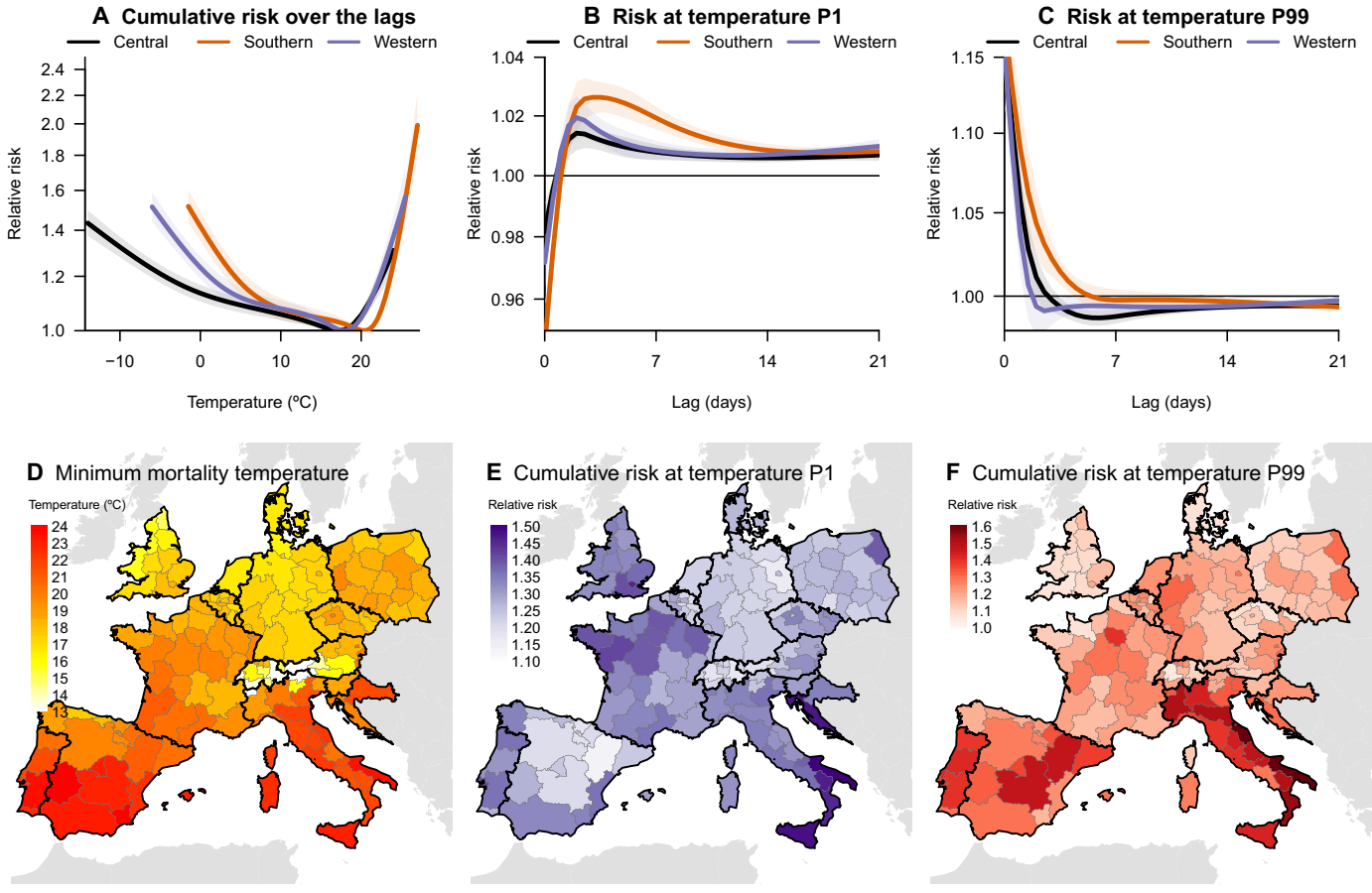

Fig. 1. Observed temperature-mortality associations (1998–2012). (**A** to **C**) RR (unitless) of death accumulated over the range of lags (A) and for percentiles 1 (P1) (B) and 99 (P99) (C) of the daily temperature distribution. These RRs were predicted for central (Austria, Switzerland, Czechia, Germany, Denmark, Croatia, Poland, and Slovenia; in black), southern (Spain, Italy, and Portugal; in red), and western (Belgium, France, Luxembourg, the Netherlands, and England and Wales; in blue) Europe. (**D**) Regional MMT (in degrees Celsius) and (**E** and **F**) the regional RR (unitless) of death for percentiles 1 (E) and 99 (F) of the daily temperature distribution.

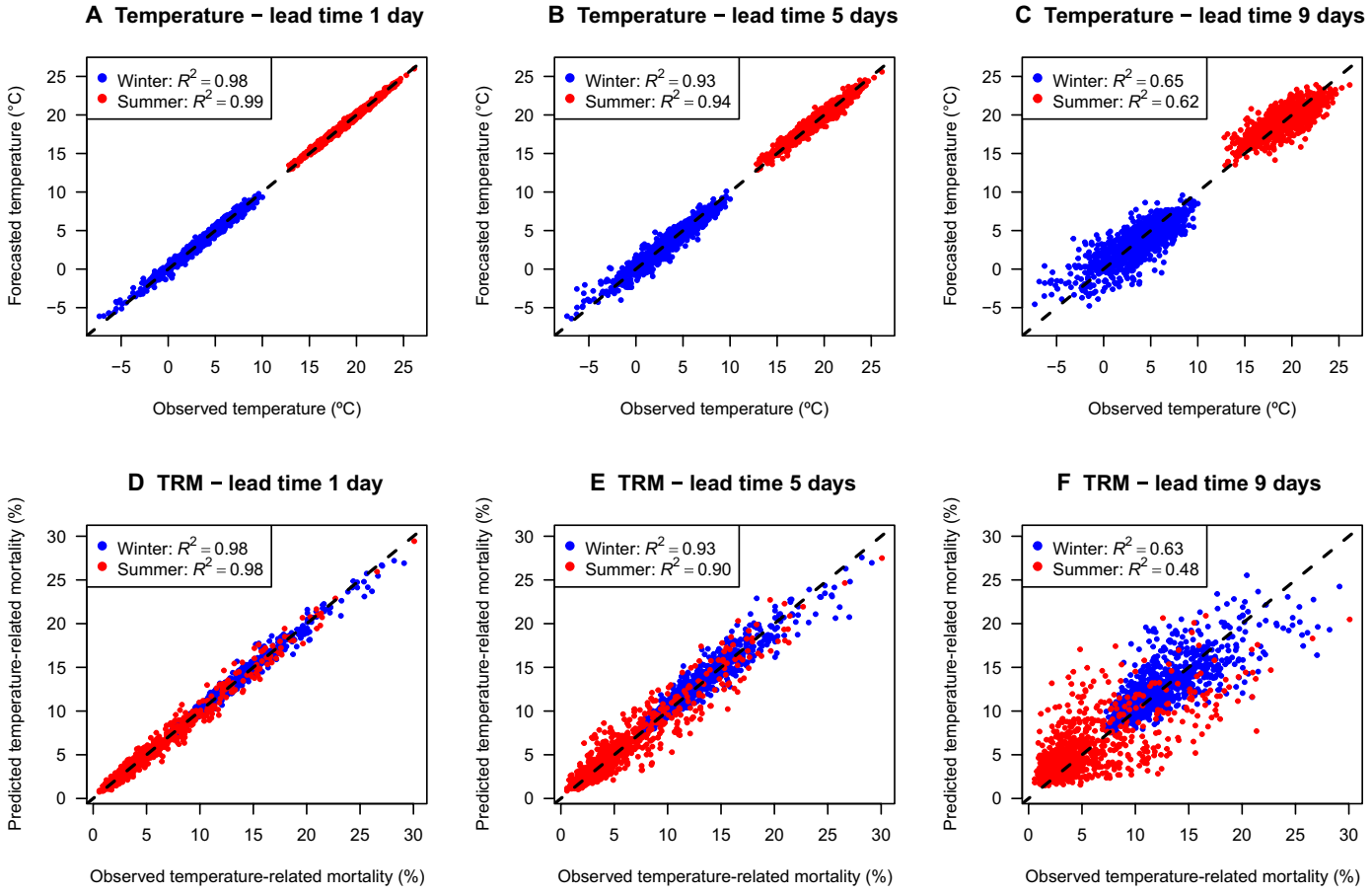

Fig. 2. Association between daily temperature (in degrees Celsius) or TRM (in %) observations and forecasts (2009–2020). Each point represents a couple of daily winter (December to February, in blue) or summer (June to August, in red) observed and ensemble mean forecast temperatures (**A** to **C**) or TRMs (**D** to **F**) aggregated over the ensemble of European regions at lead times of 1 [(A) and (D)], 5 [(B) and (E)], and 9 [(C) and (F)] days. The square of the Pearson correlation (R^2) is shown in the top left corner (all correlations are $P < 0.001$).

so that the range of lead times with skillful TRM forecasts is to a large extent dependent on the season and location. We evaluated the forecast skill of temperature and TRM forecasts by computing the ranked probability skill score (RPSS) (see Materials and Methods), a probabilistic measure accounting for the ensemble of forecast members and its uncertainty. Figure 3 shows the RPSS of temperature and TRM forecasts aggregated for the ensemble of regions as a function of the forecast lead time, and fig. S3 shows the corresponding values for each individual region. As previously seen in the correlations of Fig. 2, the forecast skill monotonically decays with the lead time, being generally higher in winter than in summer, and higher for temperature than for TRM. We note that these two basic characteristics of the forecast skill are generally found in most of the regions (fig. S3). We found that the forecast skill horizon, here defined as the lead time at which the RPSS is equal to 0.2, is almost the same for temperature and TRM in winter (10.7 and 10.5 days, respectively), while in summer, it is higher for temperature (9.5 days) than for TRM (8.3 days). We note, however, that the RPSS in winter is greater than zero at least at a lead time of 15 days for both temperature and TRM and, in summer, beyond 14 days for temperature and around 11.5 days for TRM.

Figure 4 explores the spatial heterogeneity of the forecast skill horizon, as well as the loss of forecast skill due to the regional epidemiological transformations, here quantified as the difference in the forecast skill horizon of temperature minus that of TRM. Again, the forecast skill horizon is found to be generally higher in winter than in summer and for temperature than for TRM. In winter, the forecast skill horizon of temperature is between 2 and 3 days higher in northeastern Europe than in southern and western Europe (Fig. 4A), possibly due to the more persistent, and thus predictable, winter conditions of the continental climate, characterized by the Eurasian snow cover (*25*) or through the direct effect of the polar vortex (*26*, *27*). Similar values and spatial distribution are found for TRM in winter (Fig. 4B), with generally no loss of forecast skill due to the regional epidemiological transformations (i.e., regional forecast skill horizon differences generally between −0.2 and +0.8 days; see Fig. 4C). Instead, in summer, forecast skill horizon values are found to be spatially homogeneous for temperature but with some changes in the case of TRM (Fig. 4, D and E). Moreover, there is a clear loss of forecast skill due to the regional epidemiological transformations in summer, with a difference of up to 2 days in several regions in western and central Europe (Fig. 4F). The loss of forecast skill is smaller

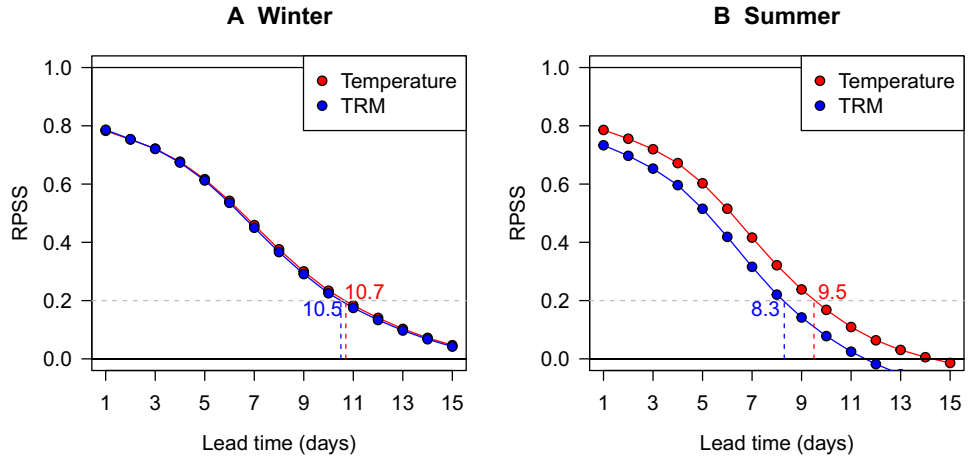

Fig. 3. Decay of the forecast skill as a function of the forecast lead time for the ensemble of European regions (2009–2020). The forecast skill of winter [December to February, (**A**)] and summer [June to August, (**B**)] temperature (red) and TRM forecasts (blue) is measured with the RPSS. The forecast skill horizon is defined as the lead time at which the RPSS is equal to 0.2.

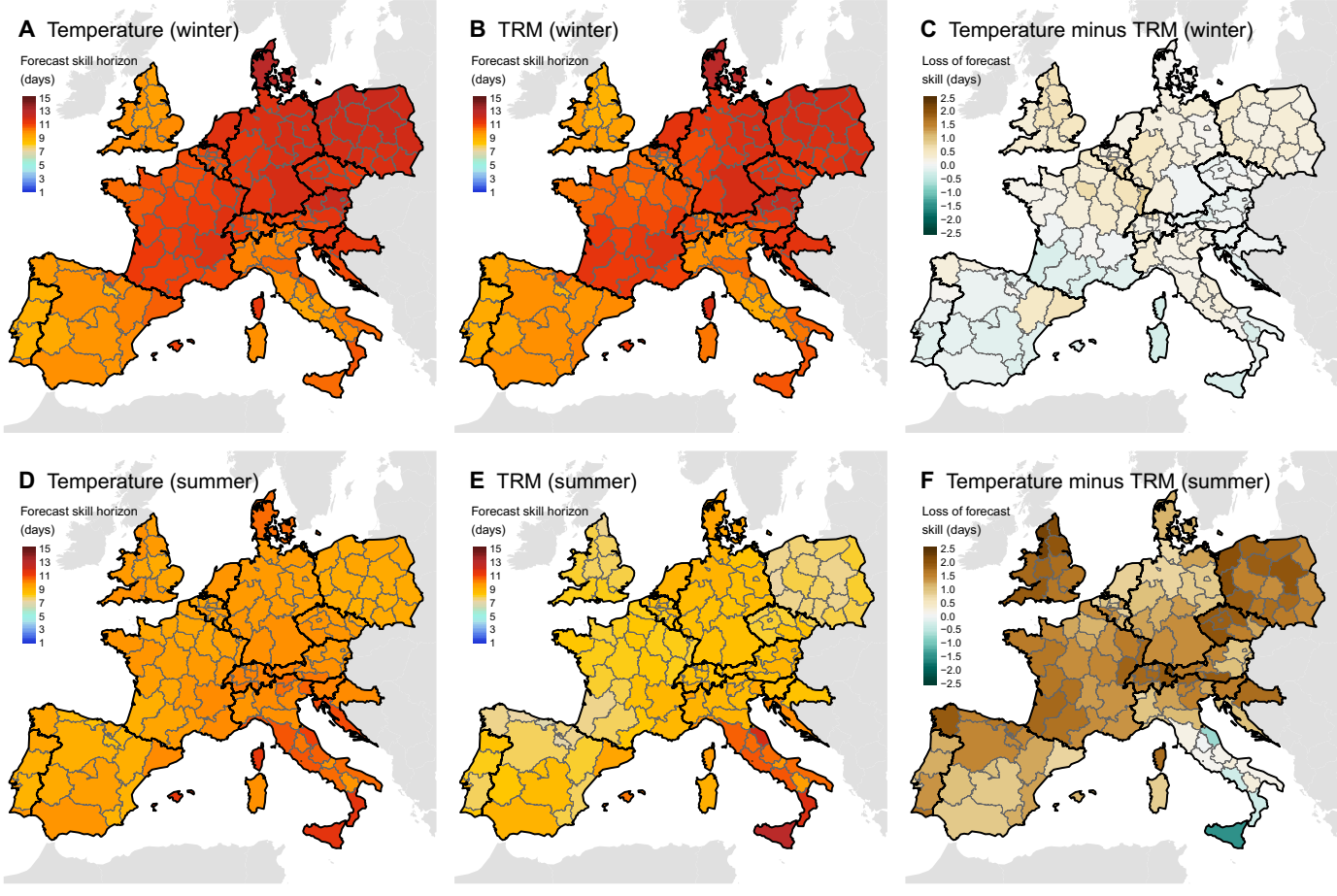

Fig. 4. Forecast skill horizon and loss of forecast skill due to the epidemiological transformations (2009–2020). The forecast skill horizon is defined as the lead time at which RPSS is equal to 0.2 and the loss of forecast skill due to the regional epidemiological transformations as the difference in the forecast skill horizon of temperature and TRM. The regional forecast skill horizon is shown for winter [December to February, (**A** and **B**)] and summer [June to August, (**D** and **E**)] temperature [(A) and (D)] and TRM forecasts [(B) and (E)]. (**C** and **F**) Loss of forecast skill in winter and summer, respectively.

in the Mediterranean basin, which is the region with the highest vulnerability to heat (cf. Figs. 1F and 4F).

Last, Fig. 5 shows how the regional loss of forecast skill is explained by the shape of the regional epidemiological associations. On the one hand, in summer, we found a clear association between the loss of forecast skill and the MMT percentile (MMTP; Fig. 5B), i.e., the percentile of the daily temperature distribution corresponding to the MMT. The regional MMTPs are always found between percentiles 78 and 93 (fig. S2), that is, at temperatures that are generally observed within the summer season. This means that temperatures are colder than the MMT in a fraction of summer days, and consequently, in these days, the TRM increases with decreasing (not increasing) temperatures below the MMT (see the asymmetric V-shaped curves in Fig. 1A). In other words, the colder is a regional MMTP, the more infrequent is the occurrence of regional temperatures colder than the MMT in summer, and therefore, the stronger and more monotonically increasing is the epidemiological association transforming the summer temperatures into TRM values, and consequently, the smaller is the amplification of errors in the temperature forecasts by the epidemiological transformations. On the other hand, in winter, the association between the regional MMTP and the regional loss of forecast skill is weak (Fig. 5A). This is possibly explained not only by the lower spatial heterogeneity in the loss of forecast skill (Fig. 4C) but also by the asymmetric V-shaped curve of the epidemiological associations (Fig. 1A), with the MMT never observed in winter.

DISCUSSION

Our study represents a comprehensive forecast skill assessment of an operational continental heat-cold-health forecasting system. We showed that the forecast skill of our system monotonically decays as a function of the lead time at a higher pace in summer than in winter and for TRM than for temperature. We found that temperature forecasts can be used to issue skillful forecasts of TRM at lead times beyond 15 days in winter and beyond 11 days in summer. Nonetheless, with a more conservative criterion, i.e., RPSS greater than or equal to 0.2, we concluded that we can generally issue skillful forecasts of temperature at a lead time of 10.7 days in winter and 9.5 days in summer and skillful forecasts of TRM at a lead time of 10.5 days in winter and 8.3 days in summer. We also described how the regional MMTP explains the regional loss of forecast skill between the temperature and TRM forecasts in summer. Thus, we showed that temperatures from operational weather forecasts can be successfully transformed into skillful forecasts of TRM accounting for the real impacts of temperatures on human health. Nonetheless, the forecast skill horizon of TRM is differently modified by the regional epidemiological association factors, such as the season and location.

We must emphasize that TRM is a variable that cannot be directly observed or measured, and therefore, it was estimated here with epidemiological models. Despite our efforts to collect mortality time series records, health registries do not have a specific code for TRM, as temperature generally acts as an additional factor to a wide range of existing baseline comorbidities. Therefore, the forecast skill assessment of TRM involves comparing estimates (here referred to as "observations") and forecasts derived from the same epidemiological model (*28*). We must, however, emphasize that exploring how the nonlinear epidemiological transformation of the bias-corrected temperature forecasts determines the forecast skill of TRM forecasts with regard to estimated TRM observations is crucial to demonstrate and quantify the forecast skill horizon of heat-cold-health forecasting systems. This forecast skill assessment is, in turn, a necessary step toward generating trust among public health authorities and end users.

While our study demonstrates that operational weather forecasts can be transformed into skillful forecasts of TRM at very long lead times, there is room for additional improvement. We postprocessed the forecasts with BC-30 as a trade-off between simplicity and forecast refinement, but other more advanced bias-correction methods, such as nonhomogeneous Gaussian regression, would further improve the forecast skill by considering the changes in the spread of the ensembles (*29*, *30*). Reforecasts are forecasts run for past decades that are produced specifically with the same model systems used to produce the actual forecasts (*30*). We acknowledge that the use of training datasets based on reforecasts for the postprocessing of the forecast ensembles would further improve the forecast skill in comparison with the use of the last 30 days of operational forecasts

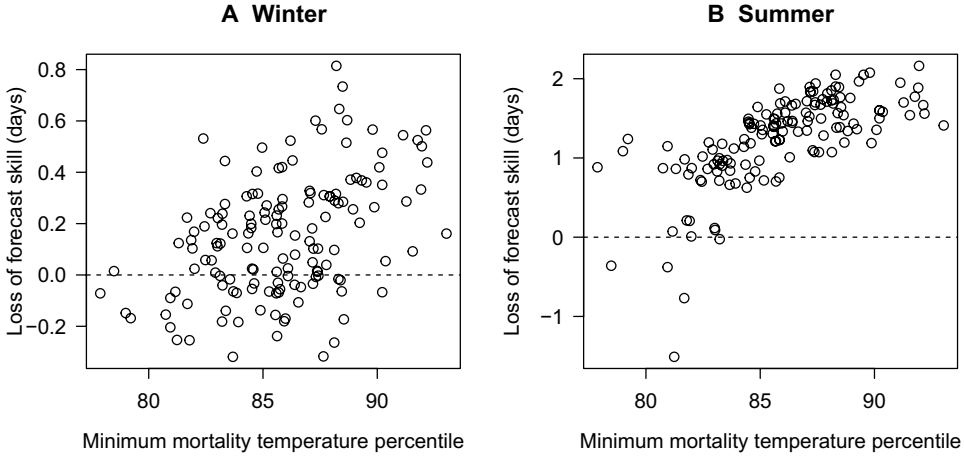

Fig. 5. Factors of the epidemiological associations explaining the regional differences in loss of forecast skill. The loss of forecast skill due to the regional epidemiological transformations is defined as the difference in the forecast skill horizon of temperature and TRM. (**A**) Relationship between the percentile of the MMT and the loss of forecast skill in winter; (**B**) the same association in summer.

(*29*, *30*), and therefore, the forecast skill horizon values here found should be considered as a lower bound.

From a health perspective, future implementations of health early warning systems should incorporate the uncertainty of the epidemiological associations. In addition, our predictability assessment for a large ensemble of European regions is not necessarily representative of other continents, particularly in data-scarce regions in low- and middle-income countries. In these regions, we would crucially need to take into account the lack of availability of mortality data (*1*, *31*), as well as the varying performance of global weather forecast systems in these areas. Even with these possible improvements, our study showed that the loss of forecast skill (of the order of 0 to 2 days) is found to be small compared to the overall forecast skill horizon of weather forecasts (8 to 11 days), which implies that the forecast skill of operational health early warning systems is, to a very large extent, influenced by the forecast skill of the temperature forecasts and, to a much lesser extent, by the epidemiological models. In other words, our results indicate that any further improvement in weather (and subseasonal-to-seasonal climate) forecasting would automatically turn into an increase in the forecast skill horizon of operational health early warning systems.

We demonstrated that the skill assessment of impact-based health early warning systems should account for both the forecast skill of the weather variables (e.g., temperature) and the corresponding epidemiological association with the health variable (e.g., mortality), both being crucial to determine when derived action measures are activated. This result emphasizes the pressing need to develop a new interdisciplinary field at the interplay between research and innovation, which combines weather and climate forecasting with epidemiology and social sciences (*22*, *32*), as recently done with the release of an operational continental heat-cold-health forecasting system of TRM (https://forecaster.health/) (*23*). This emerging domain of science would transform the existing forecast skill of the physical variables of the Earth system into a range of early warning systems of health indicators directly targeting vulnerable populations at a range of spatiotemporal scales. The forecast skill of these innovative tools would be assessed, as we here did, by comparing the forecast skill horizon of weather and climate forecasts with that of impact-based health forecasts. The World Health Organization highlights the importance of providing effective climate services and health early warning systems for saving lives in vulnerable communities around the world (*12*). Our methodological framework could be generally applied to any social determinant of health (early warning systems by sex, age, socioeconomic level, or comorbidities), variable (mortality, morbidity, and occupational health), or lead time (from days to months) as long as health data are available to calibrate epidemiological models specifically targeting these groups and variables. The development of these innovative tools, supported by a rigorous assessment of their forecast skill, as we demonstrated here, would generate trust among public health authorities and end users and, at the same time, increase resilience and strengthen our early adaptation response to climate change.

MATERIALS AND METHODS

Mortality data

We used a spatiotemporally homogeneous daily regional mortality database for the period 1998–2012 (*2*). The dataset consisted of $N = 58{,}784{,}430$ counts of deaths from 147 contiguous NUTS [Nomenclature of Territorial Units for Statistics, version of 2013 (*33*)] regions in 16 European countries, representing their entire urban and rural population of 420 million people. These countries are Austria (9 NUTS2 regions), Belgium (11 NUTS2), Switzerland (7 NUTS2), Czechia (8 NUTS2), Germany (16 NUTS1), Denmark (1 NUTS1), Spain (16 NUTS2), France (22 NUTS2), Croatia (2 NUTS2), Italy (21 NUTS2), Luxembourg (1 NUTS3), the Netherlands (1 NUTS0), Poland (16 NUTS2), Portugal (5 NUTS2), Slovenia (1 NUTS1), and the United Kingdom (10 NUTS1, regions in England and Wales only). Further details are provided in previously published studies (*24*, *34*, *35*).

Temperature data

We obtained gridded (0.1° × 0.1°) daily mean 2-m temperatures from E-OBS (version 29.0) (*36*) for the period 1998–2020, which we consider here as a proxy for observations. We also obtained gridded (interpolated to 0.4° × 0.4°) 6-hourly forecasts of 2-m temperature for the period 2009–2020 from the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble prediction system (Atmospheric model Ensemble 15-day forecast, Set III–ENS) (*37*) archived in the THORPEX Interactive Grand Global Ensemble (TIGGE) database (*38*–*40*). We retrieved the $M = 50$ perturbed ensemble members, initialized at 00:00 UTC across forecast lead times +6, +12, +18, …, +360 hours, covering lead times from 1 to 15 days. We derived the regional daily temperatures by computing the temperature average of grid cell data within the regional boundaries of each of the 147 regions analyzed. For the forecasts, after deriving the regional 6-hourly forecasts first, we averaged them to obtain the daily mean temperature forecast for lead times 1 (average of +6-, +12-, +18-, and +24-hour lead times) to 15 days (average of +342-, +348-, +354-, and +360-hour lead times).

Bias correction of temperature forecasts

We postprocessed the ensemble of temperature forecasts to bias-correct them, so that the temperature forecasts are aligned with the temperature observations used in the epidemiological models. By doing so, we also improved the skill of the temperature and TRM forecasts (*41*). We applied a state-of-the-art, widely used in operational forecasting, bias-correction method considering the most recent $N = 30$ pairs of observations and forecasts with respect to each forecast start date (BC-30) (*30*). Thus, for any given region r, observation date or forecast start date s, and forecast lead time l (taking values from 1 to 15 days), we calculated the correction c of the ensemble member forecasts as

$$c(r,s,l) = \frac{1}{N}\sum_{n=1}^{N}\left[o(r,s-n) - f_{\text{ave}}(r,s-n-l+1,l)\right]$$

where $o(r, s - n)$ and $f_{\text{ave}}(r, s - n - l + 1, l)$ are the pairs of temperature observations and ensemble mean forecasts for all cases in the training dataset, respectively. We then added this correction individually to each of the $M = 50$ ensemble members to obtain the ensemble of bias-corrected temperature forecasts.

Epidemiological model

We used a time series quasi-Poisson regression model in combination with a distributed lag nonlinear model in each region to derive estimates of region-specific temperature-lag-mortality risks (*42*). The equation is as follows

$$\log[E(\text{mort})] = \text{intercept} + S(\text{time}, 8\ \text{df per year}) + \text{dow} + \text{cb}$$

where mort denotes the daily time series of mortality counts, E corresponds to its expected value, S is a natural cubic spline of time with 8 df/year to adjust for the seasonal and long-term trends, dow corresponds to a categorical variable to control for the day of the week, and cb is the cross-basis function produced by the distributed lag nonlinear model combining the exposure-response and lag-response associations (*42*). The exposure-response association was modeled with a natural cubic spline, with three internal knots placed at the 10th, 75th, and 90th percentiles of the observed distribution of daily regional temperatures during 1998–2012. The lag-response association was modeled with a natural cubic spline, with three internal knots placed at equally spaced intervals in the log scale, with a maximum lag of 21 days, and an intercept. We then performed a multivariate multilevel meta-analysis, modeling dependencies of regions within countries through structured random effects (*43*). The fitted meta-analytical model was used to derive the best linear unbiased predictions of the cumulative temperature-mortality associations in each region, from which we estimated the regional MMT, which was used as the baseline temperature of the regional association. All these modeling choices have been widely tested and used in the epidemiological literature (*2*, *44*).

Temperature-related mortality

The daily regional time series representing the fraction of deaths attributable to nonoptimal temperatures (*45*), here referred to as TRM, was calculated as

$$\mathrm{TRM}(d) = \frac{\mathrm{RR}\left[T(d)\right] - 1}{\mathrm{RR}\left[T(d)\right]}$$

where d represents a given date, and RR[$T(d)$] is the RR at temperature $T(d)$, compared to the baseline MMT, as defined by the respective regional cumulative temperature-mortality association. We transformed every daily regional time series of temperature (i) observations and (ii) bias-corrected ensemble member forecasts into daily regional time series of TRM (i) estimations and (ii) ensemble member forecasts. For the sake of simplicity, we here refer to the TRM estimations as TRM observations throughout the manuscript, although we acknowledge the fact that TRM is a variable that cannot be directly observed or measured.

Temperature and TRM forecast skill

We quantified the forecast skill by calculating a modified version of the ranked probability score (RPS) (*46*). This probabilistic assessment of the forecast skill considers forecast uncertainty, in contrast to deterministic approaches (*41*). We applied the exact same approach separately to the ensemble of temperature and the ensemble of TRM forecasts.

First, for both temperature and TRM, we calculated the RPS for any given region r, observation date or forecast start date s, and forecast lead time l (from 1 to 15 days) as

$$\mathrm{RPS}(r,s,l) = \frac{1}{M}\sum_{m=1}^{M} \left| o(r,s) - f_m(r, s-l+1, l) \right|$$

where $o(r, s)$ and $f_m(r, s - l + 1, l)$ are the pairs of observation and forecast ensemble member m. Then, we averaged the RPS across all winter (December to February) or summer (June to August) days during the period 2009–2020, to obtain the mean RPS in each region and forecast lead time, here represented as $\overline{\mathrm{RPS}}(r, l)$.

We calculated the skill of a reference temperature or TRM forecast consisting of the climatology of the observations in the 23-year reference period 1998–2020. By construction, these climatological forecasts provide no valuable forecast information and are thus used as a reference baseline to $\overline{\mathrm{RPS}}(r, l)$. The RPS for the climatological forecasts, here represented as $\mathrm{RPS}_{\mathrm{CL}}$, was calculated as

$$\mathrm{RPS}_{\mathrm{CL}}(r,d) = \frac{1}{K}\sum_{k=1}^{K} \left| o\left(r, d_k^{'}\right) - o(r,d) \right|$$

where $o(r, d)$ are the observations for any given region r and date d, and $d_k^{'}$ refers to the set of days on the same calendar day than d within the 23-year reference period 1998–2020 ($K = 23$). Leap days were compared not only to other leap days but also to all the 28 February and 1 March in other years. In the same way, we calculated the mean skill of the climatological forecasts in each region for the winter and summer seasons, i. e., $\overline{\mathrm{RPS}_{\mathrm{CL}}}(r)$.

The resulting indicator we used for evaluating the forecast skill for each region r and forecast lead time l is the RPSS (*47*), defined as follows

$$\mathrm{RPSS}(r,l) = 1 - \frac{\overline{\mathrm{RPS}}(r,l)}{\overline{\mathrm{RPS}_{\mathrm{CL}}}(r)}$$

High RPSS values represent high forecast skill, with a maximum value of 1 when forecast errors are infinitely smaller than the corresponding climatological forecasts. Positive RPSS values indicate forecast errors smaller than those of a purely climatological forecast, and negative RPSS values indicate that forecasts provide no valuable forecast skill. Nonetheless, we here used a stricter threshold to define the forecast skill horizon as the lead time (in days) with a regional RPSS value of 0.2 (*16*, *48*). Last, we defined the regional loss of forecast skill due to the epidemiological transformations as the difference of the regional forecast skill horizon of temperature minus the regional forecast skill horizon of TRM.

Supplementary Materials

This PDF file includes:

Figs. S1 to S3

REFERENCES AND NOTES

1. J. Ballester, K. R. van Daalen, Z.-Y. Chen, H. Achebak, J. M. Antó, X. Basagaña, J.-M. Robine, F. R. Herrmann, C. Tonne, J. C. Semenza, R. Lowe, The effect of temporal data aggregation to assess the impact of changing temperatures in Europe: An epidemiological modelling study. *Lancet Reg. Health Eur.* **36**, 100779 (2024).
2. È. Martínez-Solanas, M. Quijal-Zamorano, H. Achebak, D. Petrova, J. M. Robine, F. R. Herrmann, X. Rodó, J. Ballester, Projections of temperature-attributable mortality in Europe: A time series analysis of 147 contiguous regions in 16 countries. *Lancet Planet. Health* **5**, e446–e454 (2021).
3. M. Quijal-Zamorano, È. Martínez-Solanas, H. Achebak, D. Petrova, J. M. Robine, F. R. Herrmann, X. Rodó, J. Ballester, Seasonality reversal of temperature attributable mortality projections due to previously unobserved extreme heat in Europe. *Lancet Planet. Health* **5**, e573–e575 (2021).
4. M. Marí-Dell'Olmo, L. Oliveras, L. E. Barón-Miras, C. Borrell, T. Montalvo, C. Ariza, I. Ventayol, L. Mercuriali, M. Sheehan, A. Gómez-Gutiérrez, J. R. Villalbí, Climate change and health in urban areas with a Mediterranean Climate: A conceptual framework with a social and climate justice approach. *Int. J. Environ. Res. Public Health* **19**, 12764 (2022).
5. H. Achebak, D. Devolder, J. Ballester, Trends in temperature-related age-specific and sex-specific mortality from cardiovascular diseases in Spain: A national time-series analysis. *Lancet Planet. Health* **3**, e297–e306 (2019).
6. H. Achebak, D. Devolder, V. Ingole, J. Ballester, Reversal of the seasonality of temperature-attributable mortality from respiratory diseases in Spain. *Nat. Commun.* **11**, 2457 (2020).

7. A. Fouillet, G. Rey, V. Wagner, K. Laaidi, P. Empereur-Bissonnet, A. Le Tertre, P. Frayssinet, P. Bessemoulin, F. Laurent, P. De Crouy-Chanel, E. Jougla, D. Hémon, Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave. *Int. J. Epidemiol.* **37**, 309–317 (2008).
8. J.-M. Robine, S. L. K. Cheung, S. Le Roy, H. Van Oyen, C. Griffiths, J.-P. Michel, F. R. Herrmann, Death toll exceeded 70,000 in Europe during the summer of 2003. *C. R. Biol.* **331**, 171–178 (2008).
9. J. Ballester, M. Quijal-Zamorano, R. F. Méndez Turrubiates, F. Pegenaute, F. R. Herrmann, J. M. Robine, X. Basagaña, C. Tonne, J. M. Antó, H. Achebak, Heat-related mortality in Europe during the summer of 2022. *Nat. Med.* **29**, 1857–1866 (2023).
10. E. Gallo, M. Quijal-Zamorano, R. F. Méndez Turrubiates, C. Tonne, X. Basagaña, H. Achebak, J. Ballester, Heat-related mortality in Europe during 2023 and the role of adaptation in protecting health. *Nat. Med.* **2024**, 10.1038/s41591-024-03186-1 (2024).
11. M. N. Mistry, A. Gasparrini, Real-time forecast of temperature-related excess mortality at small-area level: Towards an operational framework. *Environ. Res. Health* **2**, 035011 (2024).
12. G. R. McGregor, P. Bessmoulin, K. Ebi, B. Menne, *Heatwaves and Health: Guidance on Warning-System Development* (World Meteorological Organization, 2015).
13. A. Casanueva, A. Burgstall, S. Kotlarski, A. Messeri, M. Morabito, A. D. Flouris, L. Nybo, C. Spirig, C. Schwierz, Overview of existing heat-health warning systems in Europe. *Int. J. Environ. Res. Public Health* **16**, 2657 (2019).
14. A. Matzarakis, "The heat health warning system of DWD—Concept and lessons learned" in *Perspectives on Atmospheric Sciences*, T. Karacostas, A. Bais, P. T. Nastos, Eds. (Springer International Publishing, 2017), pp. 191–196.
15. D. S. Wilks, *Statistical Methods in the Atmospheric Sciences* (Academic Press, 2011), vol. 100.
16. R. Buizza, M. Leutbecher, *The Forecast Skill Horizon* (ECMWF, 2015); https://doi.org/10.21957/6g2wkoyb6.
17. E. N. Lorenz, Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963).
18. R. Lowe, J. Ballester, J. Creswick, J. M. Robine, F. R. Herrmann, X. Rodó, Evaluating the performance of a climate-driven mortality model during heat waves and cold spells in Europe. *Int. J. Environ. Res. Public Health* **12**, 1279–1294 (2015).
19. R. Lowe, M. García-Díez, J. Ballester, J. Creswick, J. M. Robine, F. R. Herrmann, X. Rodó, Evaluation of an early-warning system for heat wave-related mortality in Europe: Implications for sub-seasonal to seasonal forecasting and climate services. *Int. J. Environ. Res. Public Health* **13**, 206 (2016).
20. G. Masato, A. Bone, A. Charlton-Perez, S. Cavany, R. Neal, R. Dankers, H. Dacre, K. Carmichael, V. Murray, Improving the health forecasting alert system for cold weather and heat-waves in England: A proof-of-concept using temperature-mortality relationships. *PLOS ONE* **10**, e0137804 (2015).
21. A. Vaidyanathan, S. Saha, A. M. Vicedo-Cabrera, A. Gasparrini, N. Abdurehman, R. Jordan, M. Hawkins, J. Hess, A. Elixhauser, Assessment of extreme heat and hospitalizations to inform early warning systems. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5420–5427 (2019).
22. European Research Council Proof of Concept, Operational Heat-Health-Social Early Warning System (HHS-EWS) (2022); https://doi.org/10.3030/101069213.
23. J. Ballester, M. Beas-Moix, N. Beltrán-Barrón, R. Méndez Turrubiates, F. Peyrusse, M. Quijal-Zamorano, Forecaster.Health (2024); https://forecaster.health/.
24. J. Ballester, J. M. Robine, F. R. Herrmann, X. Rodó, Long-term projections and acclimatization scenarios of temperature-related mortality in Europe. *Nat. Commun.* **2**, 358 (2011).
25. C. I. Garfinkel, C. Schwartz, I. P. White, J. Rao, Predictability of the early winter Arctic oscillation from autumn Eurasian snowcover in subseasonal forecast models. *Clim. Dyn.* **55**, 961–974 (2020).
26. D. I. V. Domeisen, A. H. Butler, A. J. Charlton-Perez, B. Ayarzagüena, M. P. Baldwin, E. Dunn-Sigouin, J. C. Furtado, C. I. Garfinkel, P. Hitchcock, A. Y. Karpechko, H. Kim, J. Knight, A. L. Lang, E. P. Lim, A. Marshall, G. Roff, C. Schwartz, I. R. Simpson, S. W. Son, M. Taguchi, The role of the stratosphere in subseasonal to seasonal prediction: 1. Predictability of the stratosphere. *J. Geophys. Res. Atmos.* **125**, e2019JD030920 (2020).
27. D. I. V. Domeisen, A. H. Butler, A. J. Charlton-Perez, B. Ayarzagüena, M. P. Baldwin, E. Dunn-Sigouin, J. C. Furtado, C. I. Garfinkel, P. Hitchcock, A. Y. Karpechko, H. Kim, J. Knight, A. L. Lang, E. P. Lim, A. Marshall, G. Roff, C. Schwartz, I. R. Simpson, S. W. Son, M. Taguchi, The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *J. Geophys. Res. Atmos.* **125**, e2019JD030923 (2020).
28. D. B. Lobell, M. B. Burke, On the use of statistical models to predict crop yield responses to climate change. *Agric. For. Meteorol.* **150**, 1443–1452 (2010).
29. R. Hagedorn, Using the ECMWF reforecast dataset to calibrate EPS forecasts. ECMWF (2008); https://doi.org/10.21957/tlk9s1tzzr.
30. R. Hagedorn, R. Buizza, T. M. Hamill, M. Leutbecher, T. N. Palmer, *Comparing TIGGE Multi-Model Forecasts with Reforecast-Calibrated ECMWF ensemble forecasts* (ECMWF, 2012); https://doi.org/10.21957/9nfamplly.
31. K. L. Ebi, Weekly temperature data are sufficient to estimate exposure-response relationships: A boon for health adaptation in low-resource settings. *Lancet Reg. Health Eur.* **36**, 100811 (2024).
32. J. Ballester, R. Lowe, P. J. Diggle, X. Rodó, Seasonal forecasting and health impact models: Challenges and opportunities. *Ann. N. Y. Acad. Sci.* **1382**, 8–20 (2016).
33. Eurostat, *Regions in the European Union - Nomenclature of Territorial Units for Statistics - NUTS 2013/EU-28* (Publications Office of the European Union, 2015).
34. J. Ballester, X. Rodó, J. M. Robine, F. R. Herrmann, European seasonal mortality and influenza incidence due to winter temperature variability. *Nat. Clim. Change* **6**, 927–930 (2016).
35. J. Ballester, J. M. Robine, F. R. Herrmann, X. Rodó, Effect of the Great Recession on regional mortality trends in Europe. *Nat. Commun.* **10**, 679 (2019).
36. R. C. Cornes, G. van der Schrier, E. J. M. van den Besselaar, P. D. Jones, An ensemble version of the E-OBS temperature and precipitation data sets. *J. Geophys. Res. Atmos.* **123**, 9391–9409 (2018).
37. ECMWF, ECMWF: Set III–Atmospheric model Ensemble 15-day forecast (ENS) (2024); www.ecmwf.int/en/forecasts/datasets/set-iii.
38. P. Bougeault, Z. Toth, C. Bishop, B. Brown, D. Burridge, D. H. Chen, B. Ebert, M. Fuentes, T. M. Hamill, K. Mylne, J. Nicolau, T. Paccagnella, Y.-Y. Park, D. Parsons, B. Raoult, D. Schuster, P. S. Dias, R. Swinbank, Y. Takeuchi, W. Tennant, L. Wilson, S. Worley, The THORPEX interactive grand global ensemble. *Bull. Am. Meteorol. Soc.* **91**, 1059–1072 (2010).
39. R. Swinbank, M. Kyouda, P. Buchanan, L. Froude, T. M. Hamill, T. D. Hewson, J. H. Keller, M. Matsueda, J. Methven, F. Pappenberger, M. Scheuerer, H. A. Titley, L. Wilson, M. Yamaguchi, The TIGGE project and its achievements. *Bull. Am. Meteorol. Soc.* **97**, 49–67 (2016).
40. Y.-Y. Park, R. Buizza, M. Leutbecher, TIGGE: Preliminary results on comparing and combining ensembles. *Q. J. R. Meteorol. Soc.* **134**, 2029–2050 (2008).
41. K. Ylinen, O. Räty, M. Laine, Operational statistical postprocessing of temperature ensemble forecasts with station-specific predictors. *Meteorol. Appl.* **27**, e1971 (2020).
42. A. Gasparrini, B. Armstrong, M. G. Kenward, Distributed lag non-linear models. *Stat. Med.* **29**, 2224–2234 (2010).
43. F. Sera, A. Gasparrini, Extended two-stage designs for environmental research. *Environ. Health* **21**, 41 (2022).
44. A. Gasparrini, Y. Guo, M. Hashizume, E. Lavigne, A. Zanobetti, J. Schwartz, A. Tobias, S. Tong, J. Rocklöv, B. Forsberg, M. Leone, M. De Sario, M. L. Bell, Y.-L. L. Guo, C.-F. Wu, H. Kan, S.-M. Yi, M. de Sousa Zanotti Stagliorio Coelho, P. H. N. Saldiva, Y. Honda, H. Kim, B. Armstrong, Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *Lancet* **14**, 369–375 (2015).
45. A. Gasparrini, M. Leone, Attributable risk from distributed lag models. *BMC Med. Res. Methodol.* **14**, 55 (2014).
46. H. Hersbach, Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast.* **15**, 559–570 (2000).
47. A. P. Weigel, D. Baggenstos, M. A. Liniger, F. Vitart, C. Appenzeller, Probabilistic verification of monthly temperature forecasts. *Mon. Weather Rev.* **136**, 5162–5182 (2008).
48. F. Wetterhall, Towards sub-seasonal to seasonal forecasts for EFAS (2019); www.ecmwf.int/en/newsletter/160/meteorology/towards-sub-seasonal-seasonal-forecasts-efas.

Acknowledgments: We appreciate the contribution of F. Peyrusse in the implementation of operational statistical postprocessing methods for ensemble forecasts, enhancing the quality of our study. We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (www.uerra.eu) and the Copernicus Climate Change Service and the data providers in the ECA&D project (www.ecad.eu). **Funding:** M.Q.-Z., D.P., H.A., and J.B. gratefully acknowledge funding from the European Union's Horizon 2020 and Horizon Europe research and innovation programmes under grant agreement nos. 865564 (European Research Council Consolidator Grant EARLY-ADAPT, www.early-adapt.eu/), 101069213 (European Research Council Proof-of-Concept HHS-EWS, https://forecaster.health/), and 101123382 (European Research Council Proof-of-Concept FORECAST-AIR). J.B. also acknowledges funding from the Spanish Ministry of Science and Innovation under grant agreement no. RYC2018-025446-I (programme Ramón y Cajal). J.-M.R. gratefully acknowledges funding from the EU Community Action Program for Public Health (grant agreement no. 2005114). H.A. acknowledges funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101065876 (MSCA Postdoctoral Fellowship TEMP-MOMO). X.R. thanks support from TipESM "Exploring Tipping Points and Their Impacts Using Earth System Models," funded by the European Union [grant agreement number: 101137673, DOI: 10.3030/101137673 (contribution no. 4)]. ISGlobal authors acknowledge support from the grant CEX2023-0001290-S funded by MCIN/AEI/10.13039/501100011033 and from the Generalitat de Catalunya through the CERCA Program. **Author contributions:** Conceptualization: M.Q.-Z. and J.B. Resources: F.R.H., X.R., J.-M.R., and J.B. Data curation: M.Q.-Z., D.P., F.R.H., X.R., J.-M.R., and J.B. Methodology: M.Q.-Z. and J.B. Investigation: M.Q.-Z. Visualization: M.Q.-Z. Writing—original

draft: M.Q.-Z. and J.B. Writing—review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Temperature data used underlying this article can be accessed by the corresponding institutions (www.ecmwf.int/en/forecasts/datasets/set-iii; www.ecad.eu/download/ensembles/download.php). Daily death count data were provided by the corresponding national statistical offices. The "dlnm" R package is available on CRAN (http://cran.r-project.org/), and codes on examples of applications of the distributed lag nonlinear models are available at www.ag-myresearch.com/r-code.html. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 6 February 2024
Accepted 9 October 2024
Published 13 November 2024
10.1126/sciadv.ado5286