

Reviewer 1:

The manuscript addresses an important and timely topic, namely the development of impact-based forecasts for heat-related mortality. Overall, the manuscript clearly achieves its stated objectives and is well written, with a logical structure that makes the analysis easy to follow. I suggest some minor revisions, mainly related to clarification of certain aspects, as outlined below.

Thank you for taking the time to review our manuscript and your constructive feedback. Please find our point by point proposed revisions below.

1. The introduction clearly provides a clear overview of the motivation and research gap, it may benefit from a brief explanation of what is meant by “data-driven” versus “physics-based” weather forecasts. Since the journal attracts an interdisciplinary audience, including researchers working on climate impacts and hazards, a short clarification of these approaches could improve accessibility for readers who may be less familiar with weather forecasting.

Thank you for this comment. We propose revising the introduction to include the following from line 24.

Currently, two main frameworks exist for generating forecasts. Physics-based forecasts refer to weather forecasts issued from numerical weather prediction models. These models develop predictions by solving numerical equations representing the physics of the atmosphere, as well as other components of our earth system such as oceans, and have been the core framework for producing weather predictions for the past decades. Data-driven forecasts refer to weather forecasts where techniques from the field of statistics and artificial intelligence have been used to train weather models based on large amounts of climate data, typically in the form of reanalysis datasets such as ERA5. The rapid development of these data-driven weather models contrasts the so called ‘quiet revolution’ of traditional, physics-based numerical weather prediction (NWP) models (Bauer et al., 2015).

2. Section 2.1: the source of the reanalysis is not specified when first introduced (line 55), and only later in the paragraph it becomes clear that ERA5 is used. For clarity, the specific dataset (ERA5) could be mentioned at the beginning of the paragraph.

Thank you for highlighting this, we will clarify the use of ERA5 on line 55.

3. Section 2.1: The authors state that the meteorological data were aggregated to daily resolution. It would be useful to clarify the choice of daily mean temperature rather than daily maximum temperature (Tmax), which is commonly used in studies examining heat-related health impacts.

Daily mean temperature was used in the study of Masselot et al (2023) on which we base our results, and has also been used extensively for heat-related mortality studies (Vicedo-Cabrera et al. 2021; Ballester et al. 2023; Janos et al 2025, e.g.).

4. Section 2.1: The authors also state:

"We also use 2m temperature from archived forecasts from two different types of weather prediction models, one physical model (IFS HRES cycle 48r1), and one data-driven model (AIFS single v1)." It would be helpful if the authors briefly explained why these specific models were selected.

We propose adding the following subclause in the highlighted passage:

'..., which were the model versions available at the time this analysis was performed'.

5. Section 2.1: The manuscript states that the analysis focuses on a selection of European cities. However, the specific cities included are not clearly listed or illustrated. Providing a table or a map showing their spatial distribution across Europe would improve transparency and help readers understand the geographical area of the analysis.

We propose clarifying the geographical focus adding a list of the cities to the appendix (referred to in Section 2.2.) and as follows for the sentence on lines 58-59:

Specifically, this study focuses on a selection of 854 cities across 30 European countries further detailed in Section 2.2.

6. Section 2.2:

The epidemiological framework relies on exposure-response functions from the multi-city analysis of Masselot et al. (2023), which is based on **all-cause mortality data**. It would be helpful to state this explicitly, as the estimated impacts therefore correspond to temperature-related mortality across all causes rather than specific cause-of-death categories (e.g., cardiovascular or respiratory mortality which are the most common causes that are used and studied in health impact studies). A brief clarification regarding this would help.

We propose amending section 2.2 to state explicitly that we focus on all-cause mortality.

7. Section 3: In lines 111–112 the authors state that the continental-scale assessment uses population-weighted averages. However, the population dataset used to compute these weights is not specified. For transparency and reproducibility, the authors could clarify the population data source (dataset and spatial aggregation).

Thank you for highlighting this, we apologise for the missing information. We propose clarifying the city-level population data source on lines 111-112 as that provided together with the temperature-mortality associations by Masselot et al. (2023).

8. Section 3.2: Figure 6 (“QQ plot of population-weighted mean AF forecast bias vs temperature...”) appears to be either mislabeled or incorrectly described. The figure is referred to as a QQ plot in both the caption and the text (lines 151–159), yet the axes represent temperature and AF forecast bias rather than quantiles of two distributions. Consequently, the reference to deviations from a diagonal line is unclear. The authors may wish to clarify or revise the description of the figure.

Thank you for pointing out this. We agree with the reviewer that the figure is not technically a qq-plot since it is centered around 0 and not a diagonal axis representing the ground truth (which is p.d. always 0 here, since we are looking at bias from the ground definition). The figure shows the AF forecast bias vs temperature (for each quantile), so we propose to remove the QQ-plot terminology, and rephrase the label to “population-weighted mean AF forecast bias vs temperature for lead time 1 day (a), 3 days (b), 5 days (c) and 7 days (d). Green denotes the physics-based forecasts, purple denotes the data-driven forecasts (a–d). The black horizontal line denotes a bias of 0.”

9. Section 3.2: It could also be informative to distinguish heatwave days from typical summer days and assess forecast performance separately, as this may provide additional insight into how well physics-based and data-driven forecasts capture the attributable fraction of heat-related mortality during extreme heat conditions. I understand that this may not be possible in this type which focuses only on the summer of 2024 with limited data, so this is just a suggestion for future work including multi-year analysis.

Thank you for the suggestion, we propose adding this as a point to the discussion on potential further extensions on lines 220- 221.

... or to focus specifically on heatwaves.

10. Section 4: The continental-scale population-weighted analysis provides a useful summary of forecast performance across Europe. However, this approach may also hide some regional differences in forecast skill. Given that temperature–mortality relationships vary substantially across Europe (Masselot et al. (2023)), future work could explore city-level or regional patterns of forecast performance. In addition, examining cause-specific mortality rather than all-cause mortality may help identify where data-driven and physics-based forecasts perform better or worse under different climatic and vulnerability conditions. While this point is briefly mentioned in the discussion, it could benefit from stronger emphasis.

Thank you for highlighting these points, we propose expanding the sentence on lines 221-222 as follows:

A further extension of the work presented here would be to focus on specific subregions within Europe, as well as regions beyond Europe, particularly those that have been highlighted as being especially vulnerable to climate change.

Regarding cause-specific mortality, we would like to highlight that this study was based on temperature-mortality associations published by Masselot et al. (2023). We are not aware of a similar database for cause-specific mortality, but would warmly welcome openly accessible data of such nature. We suggest expanding lines 223-224 as follows:

The association between heat and health outcomes beyond all cause mortality is an active area of research within the health community, and we warmly welcome efforts to increase accessibility to this data.

References:

Bauer, P., Thorpe, A. & Brunet, G. The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55 (2015). <https://doi.org/10.1038/nature14956>

- Masselot P, Mistry M, Vanoli J et al. Excess mortality attributed to heat and cold: a health impact assessment study in 854 cities in Europe. *The Lancet Planetary Health* **7**, e271–e281 (2023). DOI: 10.1016/S2542-5196(23)00023-2
- Vicedo-Cabrera, A.M., Scovronick, N., Sera, F. *et al.* The burden of heat-related mortality attributable to recent human-induced climate change. *Nat. Clim. Chang.* **11**, 492–500 (2021). <https://doi.org/10.1038/s41558-021-01058-x>
- Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R.F. *et al.* Heat-related mortality in Europe during the summer of 2022. *Nat Med* **29**, 1857–1866 (2023). <https://doi.org/10.1038/s41591-023-02419-z>
- Janoš, T., Quijal-Zamorano, M., Shartova, N. *et al.* Heat-related mortality in Europe during 2024 and health emergency forecasting to reduce preventable deaths. *Nat Med* **31**, 4065–4074 (2025). <https://doi.org/10.1038/s41591-025-03954-7>