MS No.: egusphere-2026-973:
Answers to Community Comment

Original comments are in italics, **authors' response is in bold**.
All revisions described in this response have been fully implemented in the clean revised manuscript, and all line-number references correspond exclusively to the clean revised manuscript.

Community comment #1: Yves Tramblay

I read this manuscript with interest. It proposes a classification of drought types in Europe, both historically and in future ISIMIP2b simulations using the mHM hydrological model. However, looking at supplementary figure S1, I am very surprised by the choice of validation basins for the mHM simuations, that constitute the core of the study. No justification is given on the criteria to select these stations for model validation, even though there are now large river discharge databases in Europe that would allow for much more comprehensive validation. As an example, weekly discharge in 17,130 basins in Europe: do Nascimento, T.V.M., Rudlang, J., Höge, M. t al. EStreams: An integrated dataset and catalogue of streamflow, hydro-climatic and landscape variables for Europe. Sci Data 11, 879 (2024). https://doi.org/10.1038/s41597-024-03706-1. There is also no mention of the size of the basins considered in this work, while the results suggest a homogeneous coverage of the whole of Europe. However, with data such as EOBS at 0.125° spatial resolution, which is also not homogeneous in quality across Europe, it is obvious that this data is more tailored for large basins. The stations shown in Figure S1 reveal a strong location bias towards Central Europe. However, this happens to be the region with the highest density of stations in EOBS data (see for example https://doi.org/10.1002/joc.7269 and https://doi.org/10.1029/2017JD028200, notably Figure 1 of the latter). In fact, this figure S1 shows that there are only 7 basins in France and 11 in Spain (with some nested within the Ebro), and none in Italy or Greece. However, the results place a strong emphasis on Mediterranean regions. I find this problematic because we see low performances for some basins in Spain. It seems to me that this validation should be reinforced to give more confidence in the results, particularly in the basins of southern Europe where it is well known that, in addition to climatic influences, there are various factors that play a role in the reduction of low flows: Vicente-Serrano, S. M., Kenawy, A. E., Peña-Angulo, D., Lorenzo-Lacruz, J., Murphy, C., Hannaford, J., Dadson, S., Stahl, K., Noguera, I., Fraquesa, M., Fernández-Duque, B., & Domínguez-Castro, F. (2025). Forest expansion and irrigated agriculture reinforce low river flows in southern Europe during dry years. Journal of Hydrology, 653, 132818. https://doi.org/10.1016/j.jhydrol.2025.132818.

Authors' response:
We thank Dr. Yves Tramblay for this important and constructive Community Comment. We agree that the original validation figure did not clearly explain how the validation basins were selected, which could have led to a misleading impression of the model's spatial representativeness. In particular, the earlier presentation did not sufficiently distinguish between the evaluation of the fine-resolution mHM simulations and the more restricted evaluation of the coarser 0.5° setup. This distinction is

important because evaluating streamflow at coarse spatial resolution is only meaningful for basins whose observed upstream area is sufficiently well represented by the model grid.

To address this issue, we substantially revised the model evaluation, see Figure R1. The main evaluation is now performed at the 0.125° model resolution, using an updated validation network of 1 183 gauges from GRDC and a feasible subset of EStreams for which daily discharge records could be retrieved and processed. All retained stations have at least 30 years (representing a climatological period) of daily streamflow observations during 1970–2025, and their observed basin areas differ by less than 15% from the simulated upstream areas. This revised evaluation, therefore, provides a much more comprehensive assessment of mHM performance at the resolution used for the fine-resolution E-OBS-driven simulations.

The selection procedure was as follows. Applying the minimum record-length criterion of 30 years for 1970–2025 reduced the set of European GRDC stations from 3 032 to 1 721 basins. We then retained only stations for which the observed basin area differed by less than 15% from the simulated upstream area at 0.125° model resolution. This resulted in 1 176 GRDC stations. To address the reviewer's concern regarding basin size, we now also report the distribution of observed basin areas for the retained validation stations. The retained 0.125° validation basins cover a wide range of basin sizes, with observed basin areas ranging from 250 to 658 400 km^2, a median of 1 682 km^2, and an interquartile range of 596–6 284 km^2.

To improve spatial coverage in southern Europe, we also explored adding additional basins from the EStreams dataset. EStreams is an important and valuable resource for large-sample hydrology in Europe, particularly because it helps identify streamflow records and the corresponding national or regional data sources. For the present model evaluation, however, stations also had to satisfy our record-length and basin-area consistency criteria, and daily discharge records had to be retrieved and harmonised in a form suitable for direct comparison with the model simulations.

Within the scope of this revision, we were able to retrieve and process EStreams-linked daily discharge data from the Trentino and Emilia-Romagna regions of Italy. Among the 49 stations available in these two regions, seven met both validation requirements: at least 30 years of observations and a basin-area mismatch of less than 15%. The final fine-resolution validation network, therefore, comprises 1 183 stations (see Figure R1).

The model evaluation at 0.5° resolution is now treated separately and only as a consistency check for the coarser model configuration used in the ISIMIP-based climate-change analysis. For this coarse-resolution evaluation, we applied the same basin-area consistency criterion, which reduced the validation network to 149 stations. These retained 0.5° validation basins have observed basin areas ranging from 941 to 658 400 km^2, with a median of 22 639 km^2 and an interquartile range of 4 758–62 723 km^2. This filtering is necessary because many smaller or topographically complex basins cannot be represented reliably at 0.5° resolution without additional basin-specific delineation and routing checks. We there-

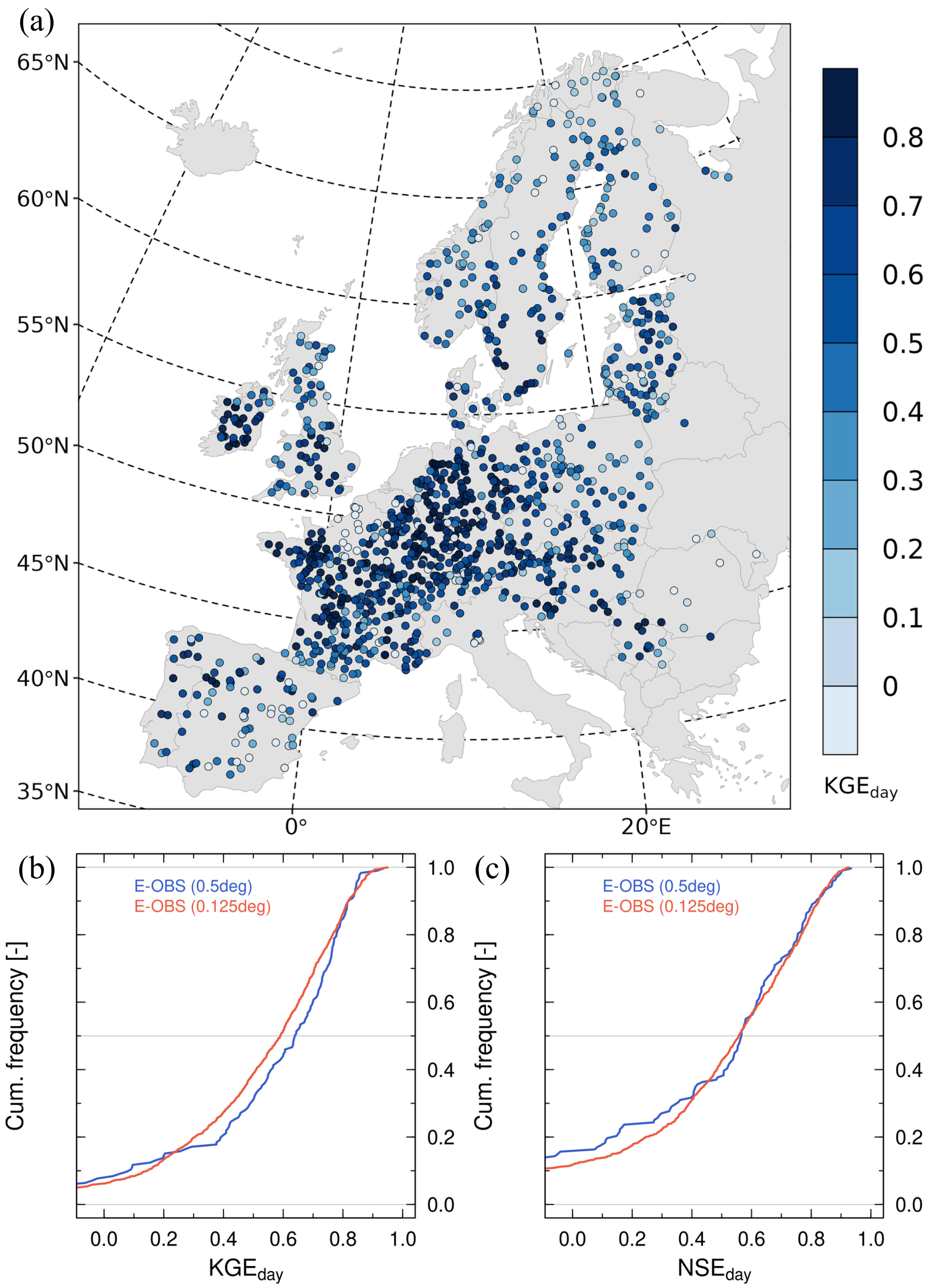

Figure R1: **Evaluation of daily streamflow simulations from mHM driven by E-OBS during 1970–2025. Panel (a) shows the spatial distribution of Kling-Gupta Efficiency (KGE) values for the fine-resolution 0.125° simulations. The validation network includes 1 183 stations from GRDC and feasible EStreams stations, each with at least 30 years of observations and an observed basin area differing by less than 15% from the simulated upstream area. Panels (b) and (c) show cumulative distributions of daily KGE and Nash-Sutcliffe Efficiency (NSE), respectively, for the 0.125° and 0.5° simulations. The 0.5° evaluation is restricted to 149 basins that satisfy the same basin-area consistency criterion and is used only as a consistency check for the coarser-resolution model configuration.**

fore avoid interpreting the 0.5° evaluation as a spatially comprehensive validation across Europe.

The revised Supplementary Fig. S1 (here, Figure R1) now shows the spatial distribution of Kling-Gupta Efficiency (KGE) values for the 0.125° evaluation network, together with cumulative distributions of KGE and Nash-Sutcliffe Efficiency (NSE) for both model resolutions. The median daily KGE is approximately 0.58 at 0.125° resolution and 0.63 at 0.5° resolution. These results indicate reasonable overall performance for a continental-scale application across the retained validation basins. The NSE-based evaluation shows a consistent pattern, supporting the robustness of the model-performance assessment.

It is also important to note that the mHM simulations were not calibrated individually for each validation basin. Instead, we used a single, spatially consistent parameter set across the European modelling domain. Individual basin-wise calibration would likely improve performance at many gauging stations, but it would introduce additional degrees of freedom, increase the risk of over-parameterisation, and reduce the transferability of the model setup for large-scale historical and climate-change applications. The purpose of the evaluation is therefore not to demonstrate locally optimised performance at each station, but to assess whether a consistent continental-scale model configuration provides reasonable skill across a large and independent validation network.

At the same time, we agree that the observational validation coverage remains spatially uneven. Although the revised validation network is much larger than in the original submission, gaps remain in parts of southern and eastern Europe, particularly Italy, the Balkans, and the Mediterranean region. We now state this limitation explicitly in the revised manuscript and avoid implying spatially uniform confidence across Europe. In particular, regional results in observationally underrepresented regions are interpreted more cautiously.

For completeness, we provide further detail on our attempt to use EStreams-linked data to extend the validation network in regions with sparse GRDC coverage. As noted above, EStreams is an important and valuable resource for large-sample hydrology in Europe. For the present model evaluation, however, stations had to satisfy two additional requirements: (i) at least 30 years of daily discharge observations during 1970–2025, and (ii) sufficient agreement between the observed basin area and the simulated upstream area. These criteria are necessary to ensure a meaningful comparison between observed and simulated discharge.

In this context, EStreams helped us identify potentially relevant additional stations and data sources. However, incorporating these data into the present pan-European validation was not straightforward, as in many regions, daily discharge records are not available as a single, harmonised, ready-to-use time series archive. Instead, the records often need to be retrieved from national or regional data providers, sometimes station by station, and then harmonised before they can be used consistently in the model evaluation. Below, we summarise the main practical constraints encountered during our attempt to extend the validation network in regions where GRDC coverage remains limited. This summary is not meant as a criticism of EStreams, but rather documents why only a limited number of additional stations could be incorporated within the scope of this revision.

- Portugal
  The linked national data portal could not be accessed reliably during our retrieval attempt. In addition, the available workflow appears to require station-wise or small-batch retrieval rather than a single harmonised bulk download of daily discharge time series. This prevented a systematic incorporation of Portuguese stations within the present revision. `https://snirh.apambiente.pt/index.php?idMain=2&idItem=1`

- Spain
  Daily streamflow data are available through the national hydrological database, but records have to be retrieved through separate hydrological regions. This makes a complete and consistent acquisition of long daily time series relatively time-intensive for a pan-European validation exercise. `https://ceh.cedex.es/anuarioaforos/demarcaciones.asp`

- Italy
  Daily streamflow observations are distributed across several regional data providers, with different access procedures, formats, and levels of automation. We were able to retrieve and process data from Emilia-Romagna and Trentino, but other regions would require additional manual retrieval, data requests, or provider-specific processing. For example, data from Piemonte, Sardegna, Toscana, and Umbria require station-wise retrieval; Lombardia and Valle d'Aosta require request-based access; and some provider links were unavailable during our retrieval attempt. We therefore incorporated only the feasible stations from Emilia-Romagna and Trentino in this revision. Among the 49 stations available from these two regions, seven satisfied both validation requirements. `https://simc.arpae.it/dext3r/`; `https://www.floods.it/public/DatiStorici.php`; `https://idro.arpalombardia.it/it/map/sidro/#`; `https://www.arpa.piemonte.it/rischi_naturali/snippets_arpa_graphs/map_meteoweb/?rete=stazione_meteorologica`; `https://www.sardegnaambiente.it/index.php?xsl=611&s=21&v=9&c=93749&na=1&n=10`; `http://www.sir.toscana.it/consistenza-rete`; `https://annali.regione.umbria.it`; `https://presidi2.regione.vda.it/str_dataview_download`

- Croatia
  EStreams provides a useful route to the Croatian data source, including web-scraping support. However, retrieving and processing daily records for the large number of potentially relevant stations would require additional automated processing and quality control. This was beyond what could be completed consistently within the present revision. `https://hidro.dhz.hr/`

- Greece
  Daily streamflow observations could not be incorporated because access requires manual retrieval for individual stations and/or submission of data requests through the provider portal. This prevented a systematic extension of the validation network for Greece within this revision.

- Bosnia and Herzegovina
  The available discharge records are provided through hydrological yearbooks in PDF format. Their use would require additional extraction and conversion into machine-readable daily time series before they could be harmonised with the other validation data. `https://www.fhmzbih.gov.ba/latinica/HIDRO/godisnjaci.php`

- Slovenia
  The data source provides access to station records, but the retrieval appears to require manual station-wise downloads. A complete incorporation of Slovenian stations would therefore require additional retrieval and harmonisation work. `https://vode.arso.gov.si/hidarhiv/`

Overall, the revised evaluation addresses the concern raised in the Community Comment by: (i) clarifying the basin-selection criteria; (ii) reporting the distribution of basin sizes represented in the validation network; (iii) expanding the fine-resolution validation network to more than 1 100 gauges with at least 30 years of observations; (iv) distinguishing clearly between the main 0.125° model evaluation and the restricted 0.5° consistency check; (v) clarifying that the simulations use a single spatially consistent parameter set rather than basin-wise calibration; and (vi) explicitly acknowledging the remaining spatial gaps in observational validation coverage. We believe these changes provide a clearer and more transparent assessment of mHM performance and its limitations.