# Response to Reviewer

**Journal:** Earth System Science Data

**Manuscript:** essd-2026-467, "A 30-Year 1-km Daily Precipitation and Air Temperature Dataset for the Po River District (Italy)"

**Authors:** H. Salehi et al.

Dear Editor, dear Reviewer,

Thank you for this careful and constructive review. The revised dataset (version 2.0; the submitted dataset is version 1.0) uses a two-part precipitation scheme: wet-day occurrence and wet-day amount are interpolated separately. Temperature uses the anomaly method with local lapse rates. Unless stated otherwise, the evaluation metrics reported for our dataset refer to version 2.0 and are computed over 1991–2020 from a full leave-one-out (LOO) cross-validation, in which each station is removed both from the daily interpolation and from its own climatological normals.

The response to Reviewer 1 gives more detail on the method and evaluation. The numerical results are consistent across both response letters.

## General comments

### Comment 1:

*I think it would be useful to take a more critical look at the limitations of the dataset. The datasets have strengths, but particularly in Chapter 7, as well as highlighting these strengths, its limitations could also be explained a little more clearly. The description of such limitations helps to prevent inappropriate use and guide the user. There is also no detailed critical description of the interpolation method used (e.g. smoothing, …), which is quite widespread and for which there is a considerable body of literature. The limitations should also be described in greater detail in the results section, for example in the analysis of Figure 8, that in my opinion should be expanded with analyses that provide a better understanding of the nature of the datasets. Another example is the beginning of Chapter 3, where emphasis is placed on the number of stations and time steps available; however, it is only by comparing this with Figure 2 that one realises that many of these stations have a rather limited duration. This also has a fundamental impact on the actual resolution (as opposed to the nominal resolution of 1 km), which is extremely important to explain in detail.*

*In general, I feel that a more detailed analysis of the results is really necessary, given that the purpose of this publication is to present new datasets and thus serve as a reference for all those who use this data. At present, a great deal of essential information is missing, and I do not feel that I am being provided with a comprehensive analysis of the data's quality and characteristics.*

**Response.** We agree. We now discuss these limitations in three places:

- in the method description (Sect. 4);
- in the results (Sect. 6);
- in a rewritten Usage notes and limitations section (Sect. 7), which is mirrored in the dataset README.

Interpolation limits (Sect. 4.3). Kriging smooths the field between stations. For daily precipitation this has a known consequence: the wet area spreads, light precipitation is spread over too many days, and the intensity

of wet days falls. Wet-day frequency is therefore overestimated and wet-day intensity underestimated, even when the total is unbiased.

This is why the revised dataset interpolates occurrence and amount separately. The same principle underlies E-OBS (Haylock et al., 2008; Cornes et al., 2018). The section also describes:

- the reduction of variance by kriging (see the response to the comment on Sect. 3.2);
- the loss of local contrasts within single events;
- the dependence of the effective resolution on the station network.

Record length and network change (Sect. 3). The station count alone can be misleading, so we now report the record lengths explicitly. Of the 1636 precipitation stations:

- the median record covers 16.7 years of valid data;
- 388 stations have fewer than 10 years;
- 485 have at least 20 years;
- 187 have at least 28 years.

The mean number of stations reporting per day grows from 369 in 1991 to 1203 in 2020. A new panel in Fig. 2 shows the distribution of record lengths.

Actual versus nominal resolution (Sects. 5.2 and 6.2). We used the kriging weights to quantify the spatial support of each estimate. Specifically, we measured the distance at which the weights fall to a given fraction of the largest weight. The characteristic weight-based spatial support of the daily fields is about 8–20 km for precipitation and 11–25 km for temperature, and it is coarser in the 1990s than in the 2010s, following the growth of the network. Maps by decade are distributed with the dataset.

We now describe the two components separately:

- the climatological (normal) layer is represented on the 1-km grid and contains terrain-dependent spatial structure at that scale;
- the daily anomaly component superimposed on it has a characteristic weight-based support of about 8–20 km for precipitation and 11–25 km for temperature.

Expanded Fig. 8 analysis (Sect. 6.4). We now split mean annual precipitation into wet-day frequency and wet-day intensity. We show both quantities by elevation for the dataset, the stations, and five reference products (Fig. R2-1 and Fig. R2-2; see also Comment 2). At 933 in-domain stations, the LOO errors of the station climatologies are:

| | **LOO median bias** | **LOO median absolute error** |
|---|---|---|
| Mean annual precipitation | +0.3 % | 5.8 % |
| Wet-day frequency (≥ 1 mm) | −0.8 % | 3.7 % |
| Wet-day intensity | +2.9 % | 5.6 % |

The error increases with elevation: the LOO error of mean annual precipitation is 5.0 % below 500 m and 11.4 % above 2000 m. Only 30 of the stations are above 2000 m. All values are relative to gauge observations and are not corrected for gauge undercatch, which is largest for solid precipitation at high elevation. Both limitations are stated in Sect. 7.

We also ran two 50 % deletion experiments (30 realisations each; Sect. 6.3): temporal missingness, in which station-days are removed and the normals kept, and network sparsity, in which stations are removed and the normals rebuilt. The results are reported in our response to Reviewer 1 (Comment 2.3). For temperature, the anomaly method is the more robust. For precipitation, the daily errors of the two-part scheme increase

about as much as those of direct interpolation, while its wet-day frequency, wet-day intensity and, in a four-year check, its station-mean bias are less affected by the missing data.

## Comment 2:

*In Chapter 6, in my view, only two datasets for evaluation have been selected – datasets for which one might already expect, a priori, that they perform less well. On the other hand, many datasets developed over the last few decades and focusing on this part of Europe – often developed in Italy – which would be extremely useful for an evaluation, have been ignored for evaluation (and are not even cited, which also makes a revision of the introduction necessary). One example is the ARCIS dataset (www.arcis.it), and there are various other examples (https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.5217 and others). It would certainly also be useful to utilise the datasets cited, such as the "High-resolution Alpine datasets [...] by Isotta et al. (2014)", to assess precipitation.*

**Response.** We agree. We revised the Introduction to include these products. The evaluation (Sect. 6.4) now uses five daily reference products:

- Manara et al. (2026; IH-GAR, 30″);
- ARCIS (Pavan et al., 2019; 5 km);
- APGD (Isotta et al., 2014; current MeteoSwiss release RapdD, 5 km; 1991–2019);
- EMO-1arcmin (Salamon et al., 2026);
- E-OBS v31.0e (Cornes et al., 2018; 0.1°).

We also use the 1961–1990 normals of Crespi et al. (2018) and Brunetti et al. (2014) for the climatological comparison.

Each product was evaluated on its native grid against the same 933 in-domain stations (≥ 10 years of data) used for our dataset. We compared mean annual precipitation, wet-day frequency and wet-day intensity (Table R2-1, Fig. R2-1 and Fig. R2-2):

**Table R2-1.** Station-median bias / median absolute error (%) of the 1991–2020 station climatologies at 933 stations.

| Product | Annual precipitation | Wet-day frequency | Wet-day intensity |
|---|---|---|---|
| This dataset (LOO) | +0.3 / 5.8 | −0.8 / 3.7 | +2.9 / 5.6 |
| This dataset (gridded, in-sample) | +1.2 / 4.4 | −0.7 / 2.8 | +3.5 / 4.3 |
| Manara et al. (2026) | +0.5 / 4.8 | +1.3 / 3.4 | −0.2 / 3.7 |
| ARCIS | +1.4 / 6.1 | +5.6 / 6.5 | −4.0 / 5.4 |
| APGD | +3.4 / 7.1 | +7.1 / 7.9 | −3.5 / 5.8 |
| EMO-1arcmin | +0.6 / 7.1 | +14.0 / 14.0 | −12.6 / 12.9 |
| E-OBS | −5.9 / 11.1 | +2.6 / 5.8 | −7.5 / 9.4 |

Five results stand out:

1. ARCIS, APGD, EMO-1arcmin and E-OBS overestimate wet-day frequency and underestimate wet-day intensity. In EMO-1arcmin the two errors largely cancel in the annual total (+0.6 %). Manara et al. (2026) is nearly unbiased overall. Its sign reverses above 2000 m, however: frequency −6.4 %, intensity +6.1 %. ARCIS also underestimates frequency above 2000 m (−4.4 %).
2. Our dataset is the only product in Table R2-1 that overestimates intensity, slightly (+2.9 % LOO). Its frequency bias is within ±2.2 % in all elevation bands, both in-sample and in LOO.
3. In the lowlands (below 1000 m), Manara et al. (2026) has a lower error than the LOO error of our dataset: 4.2–4.6 % against 5.0–6.7 % for annual precipitation. In the 1500–2000 m band, ARCIS (9.2 %) and Manara et al. (2026) (9.6 %) are slightly lower than our dataset (9.9 %).

4. Above 2000 m, the LOO error of our dataset is 11.4 %, against 14.0–14.9 % for Manara et al. (2026), ARCIS, APGD and EMO-1arcmin and 30.1 % for E-OBS.
5. Part of the evaluation stations are also inputs to the reference products, so the errors of those products are likely optimistic. This makes the comparison conservative for our dataset.

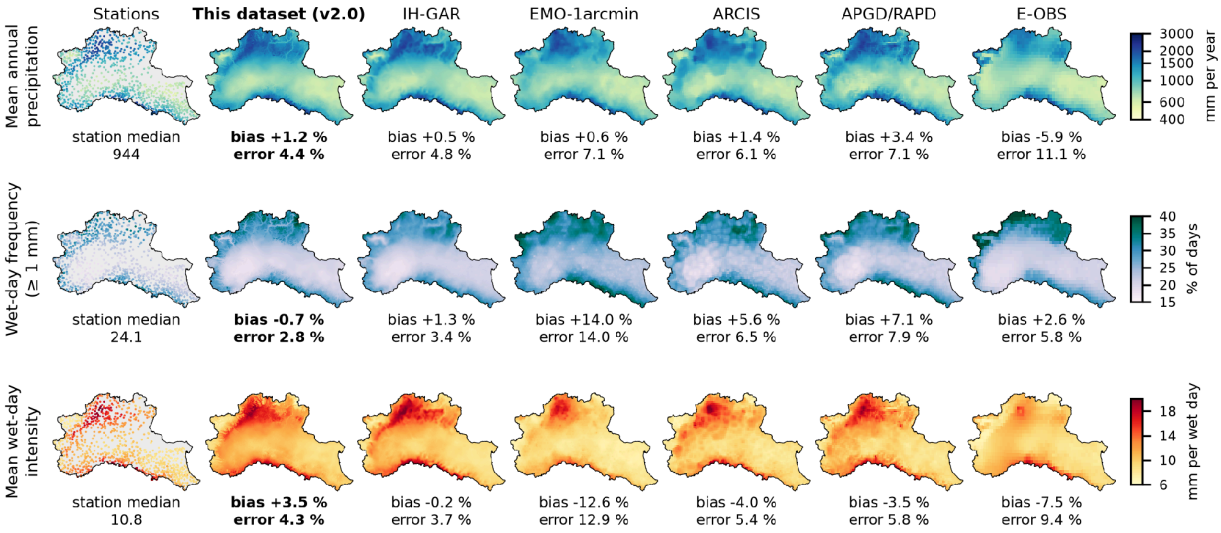



**Figure R2-1.** Mean annual precipitation, wet-day frequency (≥ 1 mm) and mean wet-day intensity, 1991–2020, of the station climatologies, this dataset and five reference products, each on its native grid (APGD: 1991–2019). Bias and error: median and median absolute value of the relative difference between the gridded value and the station climatology at the 933 in-domain stations with at least 10 years of data. Product labels: IH-GAR = Manara et al. (2026); APGD/RAPD = APGD, MeteoSwiss release RapdD.

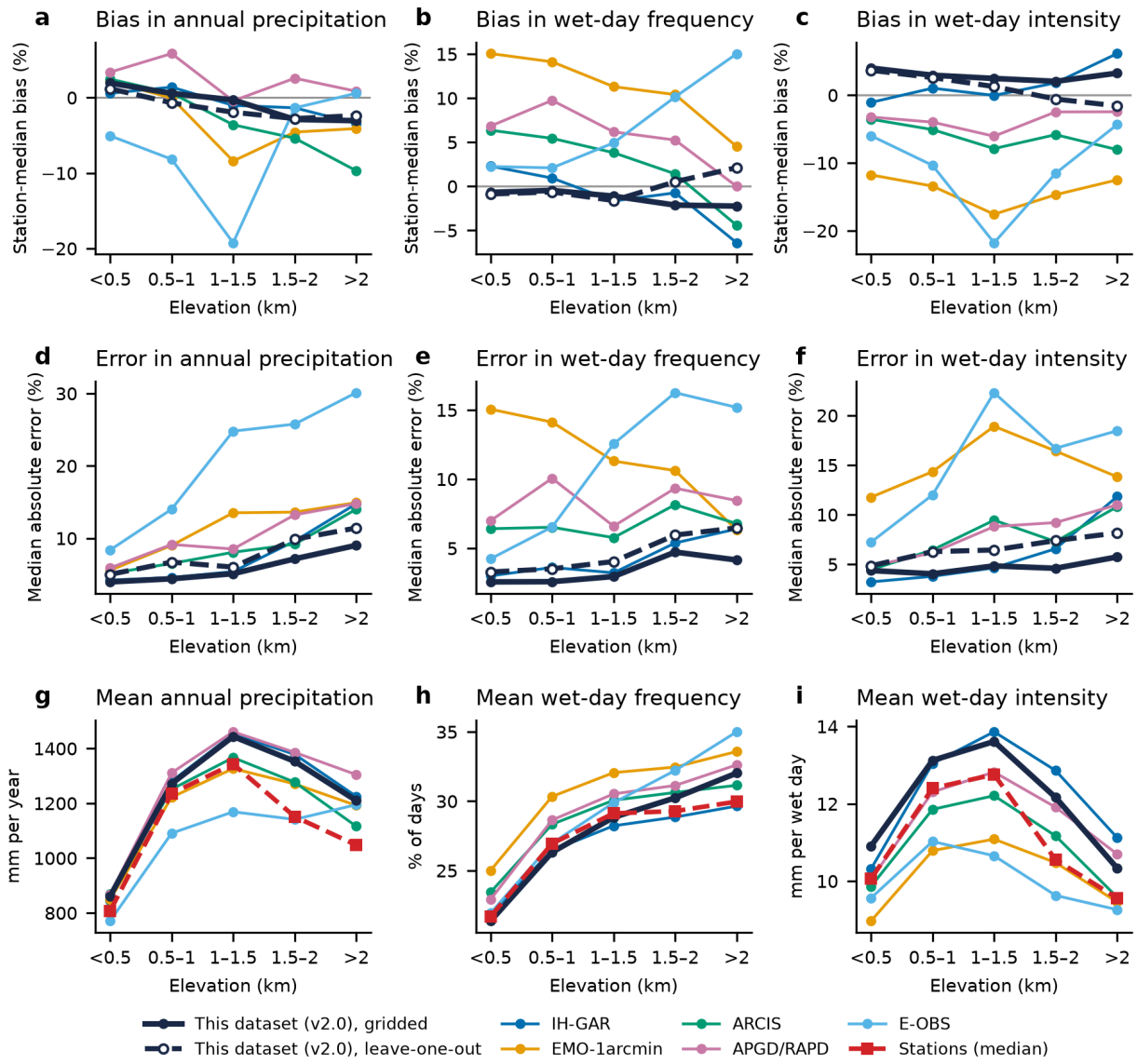



**Figure R2-2.** Station-median bias (a–c), median absolute error (d–f) and mean values (g–i) of annual precipitation, wet-day frequency and wet-day intensity by elevation band, for this dataset (gridded and leave-one-out) and five reference products, at the 933 in-domain stations (514, 229, 95, 64 and 30 stations per band). Panels g–i show area means of each gridded product over all cells of the band; the red line is the median of the station climatologies, which do not represent the band area at high elevation. Product labels: IH-GAR = Manara et al. (2026); APGD/RAPD = APGD, MeteoSwiss release RapdD.

**Temperature.** Daily mean temperature was compared on the reference grids with ARCIS (Tmax+Tmin)/2 (Pavan et al., 2026), EMO-1arcmin and E-OBS, and the normals with the 1961–1990 climatology of Brunetti et al. (2014).

- Area-mean differences are −0.25 °C (ARCIS), −0.08 °C (EMO-1arcmin) and −0.28 °C (E-OBS), and the median daily correlation is 0.99 for all three.
- With Brunetti et al. (2014), the spatial correlation of the annual normals is 0.987 and the elevation gradients are −5.30 and −5.33 °C km$^{-1}$. The mean difference of +0.72 °C reflects the warming between the two periods.

More detail is given in our response to Reviewer 1 (Comments 1.7 and 3.1).

The district-mean annual series are highly correlated across products (Table R2-2, Comment 3). The expanded Fig. 8 compares the mean annual fields of all five products.

## Comment 3:

*I am not convinced that the tests carried out in Chapter 5.1 – which demonstrate the temporal consistency of the dataset, albeit within the context of changes in the various scores used – are sufficient to then conclude, at the end of Chapter 7, that the dataset is suitable for trend analysis. The data from the stations are not necessarily homogeneous, and in any case, a change in the density of measurements can create artificial trends.*

**Response.** We agree and have removed the statement that the dataset is suitable for trend analysis. The station records were not homogenised, and the density of the network roughly triples over the period. The revised Usage notes explain that the dataset is intended for hydrological applications and for the description of climatological conditions and variability. They also state that trend estimates require homogenised records and a fixed network.

We still report three checks so users can judge the temporal consistency of the dataset.

**(a) Trends at the long-record stations.** At the 171 stations with at least 28 complete years, we compared least-squares trends in the observed series with those in the LOO predictions at the same stations. Because each station is withheld, the prediction depends only on the surrounding network, which changes in time.

- Annual precipitation: the median trend is −0.74 % per decade observed and −0.57 % per decade predicted. The median difference is −0.02 % per decade.
- Wet-day frequency: the median difference is +0.15 percentage points per decade.

At the median, we found no systematic artificial trend caused by network change at these long-record stations. Local trend errors, however, remain substantial:

- annual precipitation: correlation 0.50, median absolute difference 1.9 % per decade;
- wet-day frequency: correlation 0.18.

The dataset should therefore not be used to derive local trends.

**(b) District-mean annual series (Table R2-2).** Trends are Theil–Sen slopes; significance is tested with the Mann–Kendall test. None of the precipitation trends is significant, in this dataset or in the references ($p \geq 0.4$).

**Table R2-2.** District-mean annual precipitation: correlation with the reference datasets and Theil–Sen trends (% per decade).

| Reference | Years | Correlation | Trend, reference (% per decade) | Trend, this dataset (% per decade) |
|---|---|---|---|---|
| Manara et al. (2026) | 1991–2020 | 0.999 | +1.1 | +0.8 |
| ARCIS | 1991–2020 | 0.998 | +0.7 | +0.8 |
| APGD | 1991–2019 | 0.996 | +2.3 | +1.5 |
| E-OBS | 1991–2020 | 0.988 | +0.5 | +0.8 |
| EMO-1arcmin | 1991–2020 | 0.952 | −3.4 | +0.8 |

**(c) Stability of the wet-day frequency error.** The same network effect is visible in the reference products. On the 81 stations with data in all three decades, the frequency error is stable in ARCIS (+4.0, +4.6, +3.6 %) and Manara et al. (2026) (+1.9, +3.0, +2.2 %). It decreases in EMO-1arcmin (+16.6 → +4.5 %) and increases in E-OBS (+4.1 → +9.7 %). This supports the reviewer's concern that changes in the input network can create artificial changes in gridded products. We now cite this explicitly in Sect. 7 (Fig. R2-3).

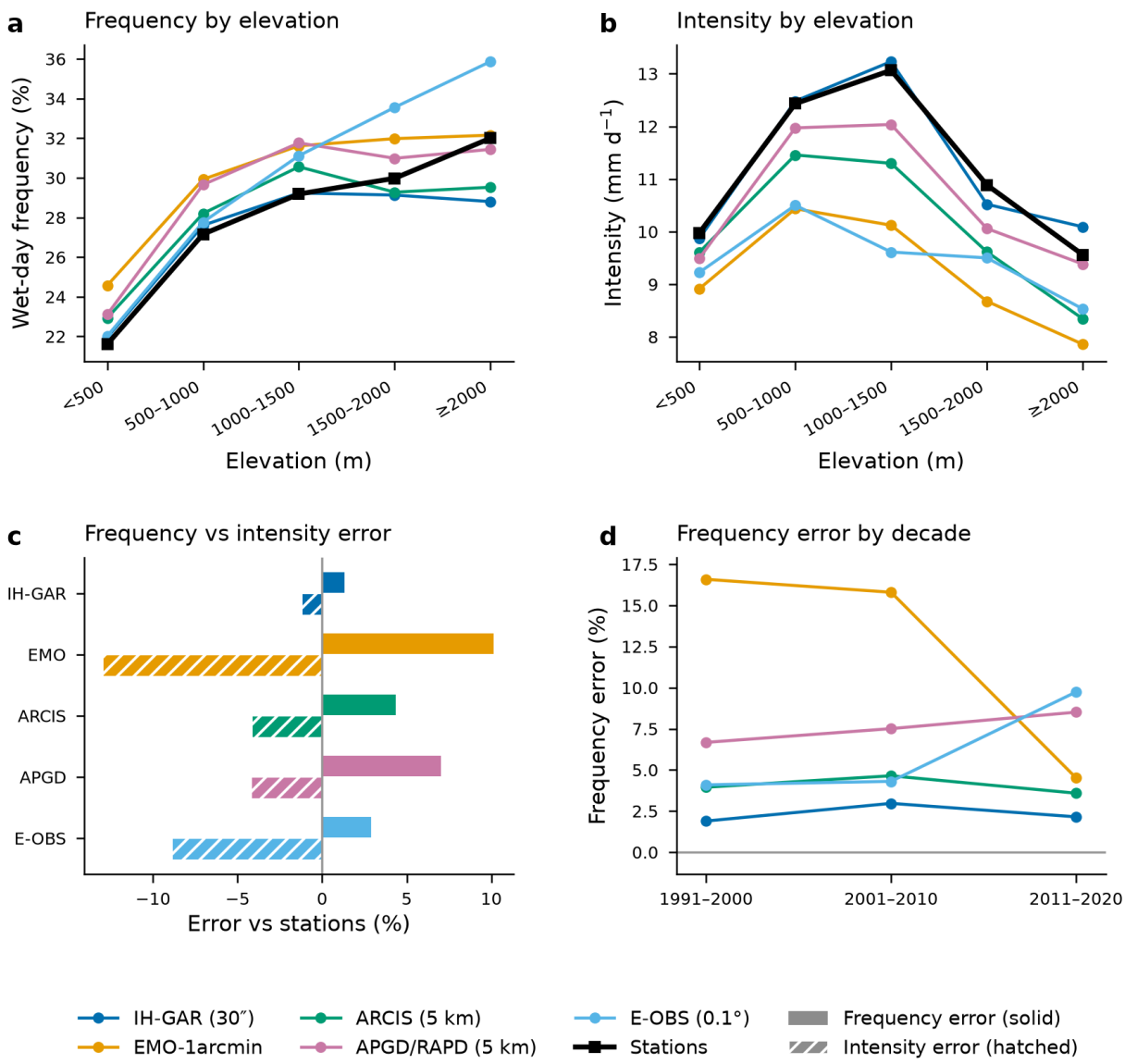



**Figure R2-3.** Wet-day frequency and intensity of five reference products compared with the stations (1991–2020; in-domain stations with at least 5 years of data; common days only): (a) frequency and (b) intensity by elevation band; (c) relative error of frequency (solid) and intensity (hatched) for all stations; (d) frequency error by decade on the 81 stations with data in all three decades. Product labels: IH-GAR = Manara et al. (2026); APGD/RAPD = APGD, MeteoSwiss release RapdD.

## Specific comments

**L33 (Isotta et al., 2014).** Revised: *"APGD (Isotta et al., 2014), recently extended to 1971–2019 by MeteoSwiss (RapdD), does not cover the 1991–2020 normal period completely, as the year 2020 is missing."*

**L41–44.** The repetition has been removed.

**Fig. 2.** The axes now show easting and northing in km in ETRS89-LAEA (EPSG:3035), with no offset factor. A panel with the distribution of record lengths has been added (Comment 1).

**Sect. 3.1 — flagged values.** We agree that these counts should be reported. Quality control was applied in two stages, both of which are now described with counts in Sect. 3.3 of the revised manuscript.

Stage 1 — station and series screening (as in the submitted version). We applied this stage to the merged multi-source series for 1991–2020. It consisted of:

- record-length screening;
- physical range checks;
- for precipitation, a neighbour-based check of values ≥ 40 mm;
- for temperature, the day-to-day change test against the 10 nearest neighbours (Eqs. 1–3 of the submitted manuscript).

This stage removed:

- 61 precipitation stations (374 563 station-days) and 25 temperature stations (92 131 station-days);
- 19 642 values (0.21 %) at the retained precipitation stations and 165 185 values (1.84 %) at the retained temperature stations, almost all of them from the temperature consistency test.

For temperature, flagged values were inspected manually. The numbers above are net removals after that inspection. The number of reinstated values was not recorded, so we cannot report it.

*Stage 2 — station-day QC of the revised version (new, fully automatic).* An iterative procedure based on the leave-one-out errors of a first interpolation run was applied to the screened series. It uses the following tests (Sect. 3.3):

**Table R2-3.** Stage-2 quality control: station-days flagged by each test.

| Test | Variable | Station-days flagged |
|---|---|---|
| False zeros | Precipitation | 2238 |
| Isolated high values | Precipitation | 658 |
| Stuck sensors | Precipitation | 91 |
| Accumulations after gaps | Precipitation | 19 |
| Range / invalid entries | Precipitation | 7 |
| Spatial outliers of the debiased LOO error | Temperature | 3180 |
| Stuck sensors | Temperature | 1878 |
| Range / invalid entries | Temperature | 2 |

Stage 2 flagged 3013 precipitation (0.031 %) and 5060 temperature (0.057 %) station-days. All flagged values were set to missing. No values were reinstated, and there was no manual step. The code of this stage is released with the processing code, so the stage is fully reproducible. The list of flagged station-days, with the test that flagged each, is available from the authors on request, subject to the licences of the data providers.

Before stage 2, 934 precipitation and 382 temperature records were relabelled to the calendar day they cover (Sect. 3.2).

**L183.** Corrected: *"MAE is less sensitive to outliers than squared-error metrics such as RMSE, although it is not insensitive to them."*

**Sect. 3.2 and following — variability.** We agree that KGE alone can hide compensating errors. The revised manuscript reports its components separately for precipitation (Table R2-4) and temperature (Table R2-5), together with extremes.

**Table R2-4.** Precipitation: components of the KGE in leave-one-out cross-validation, station medians, 1991–2020 (r: correlation; α: ratio of standard deviations; β: ratio of means).

| Time scale | r | α | β | KGE |
|---|---|---|---|---|
| Daily | 0.916 | 0.958 | 1.004 | 0.856 |
| Monthly | 0.952 | 0.980 | 1.005 | 0.882 |
| Annual | 0.932 | 1.002 | 1.026 | 0.838 |

- These are medians of the station-wise values. The pooled values over all station-days are higher (KGE 0.887 for all 1627 stations); Reviewer 1's Table R4 uses pooled values on the common sample of 1580 stations available for all three versions (KGE 0.889).
- The annual β (1.026) is computed on complete years only, and therefore on a different set of station-years than the daily and monthly values.
- The daily variability is underestimated by about 4 % (α = 0.958), as expected from kriging. The underestimation decreases to 2 % at the monthly scale and vanishes at the annual scale.
- Extremes are underestimated more: the station-median ratio of the 99th percentile of daily precipitation is 0.95, and that of the mean annual maximum daily precipitation (Rx1day) is 0.93.
- Two further results from Table R2-1 are relevant here: the separate reproduction of frequency and intensity, and the absence of frequency–intensity compensation in our dataset.

For temperature, the ratio of means is not meaningful (temperatures in °C change sign), so Table R2-5 gives the correlation and the ratio of standard deviations. Daily and monthly values are also given as anomalies from the mean seasonal cycle of each station, so that the seasonal cycle does not dominate the statistics.

**Table R2-5.** Temperature: correlation and ratio of standard deviations in leave-one-out cross-validation, station medians, 1991–2020.

| Time scale | r | α |
|---|---|---|
| Daily values | 0.994 | 0.996 |
| Daily anomalies | 0.966 | 0.977 |
| Monthly anomalies | 0.952 | 0.988 |
| Annual means | 0.948 | 0.933 |

- The daily and monthly variability of temperature is reproduced to within about 2 %. The median daily bias is −0.03 °C.
- The interannual variability of the annual means at stations is underestimated by about 7 %.
- The tails of the daily distribution are slightly compressed: the 1st percentile is 0.16 °C too warm and the 99th percentile 0.12 °C too cold (station medians).

The statement at L322 has been revised to: *"The dataset can be used to describe climate variability at monthly to annual scales, subject to the documented limitations: daily precipitation extremes are underestimated by about 5–7 % at stations and more for single events, and the interannual variability of annual mean temperature at stations by about 7 %. It is not intended for local trend detection."*

**Chapter 10 — access to station data.**

*I'm not sure how to access the station data. Is it available online? If anyone is interested, do they have to request them separately from each region?*

We thank the reviewer for raising this point. The station observations were provided to the Po River Basin District Authority (ADBPO) by the regional environmental protection agencies and meteorological services listed below, and the boundary stations outside the District were taken from the EEAR-Clim dataset (Bongiovanni et al., 2025). We are not the owners of these observations and are not permitted to redistribute them. They are, however, publicly available from each provider through its own data services, under the provider's terms of use, and users interested in the raw station records can obtain them there. In the revised manuscript, the Data Availability section names every data provider, the providers are acknowledged, and the EEAR-Clim dataset is cited with its DOI. The gridded dataset, which is the product presented in this paper, remains openly available on Zenodo under a CC BY 4.0 licence (https://doi.org/10.5281/zenodo.19207256).

## References


Brunetti, M., Maugeri, M., Nanni, T., Simolo, C., and Spinoni, J.: High-resolution temperature climatology for Italy: interpolation method intercomparison, Int. J. Climatol., 34, 1278–1296, https://doi.org/10.1002/joc.3764, 2014.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D.: An ensemble version of the E-OBS temperature and precipitation data sets, J. Geophys. Res.-Atmos., 123, 9391–9409, https://doi.org/10.1029/2017JD028200, 2018.

Crespi, A., Brunetti, M., Lentini, G., and Maugeri, M.: 1961–1990 high-resolution monthly precipitation climatologies for Italy, Int. J. Climatol., 38, 878–895, https://doi.org/10.1002/joc.5217, 2018.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res., 113, D20119, https://doi.org/10.1029/2008JD010201, 2008.

Isotta, F. A., Frei, C., Weilguni, V., et al.: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data, Int. J. Climatol., 34, 1657–1675, https://doi.org/10.1002/joc.3794, 2014.

Manara, V., et al.: A new daily high-resolution gridded precipitation dataset for the Italian Greater Alpine Region (1951–2023): spatio-temporal analysis and hydrological applications, J. Hydrol. Reg. Stud., 67, 103776, https://doi.org/10.1016/j.ejrh.2026.103776, 2026.

Pavan, V., Antolini, G., Barbiero, R., et al.: High resolution climate precipitation analysis for north-central Italy, 1961–2015, Clim. Dynam., 52, 3435–3453, https://doi.org/10.1007/s00382-018-4337-6, 2019.

Pavan, V., et al.: A new operational dataset of gridded minimum and maximum temperature over North-Central Italy 1991 to present, Clim. Serv., 43, 100689, https://doi.org/10.1016/j.cliser.2026.100689, 2026.

Salamon, P., Sperzel, T., Gomes, G. R., Radke-Fretz, M., Lemke, C.-D., Russo, C., Schweim, C., Zsoter, E., Dosio, A., Vomero, M., Ziese, M., and Grimaldi, S.: EMO-1: an improved version of the high-resolution multi-variable gridded meteorological dataset for Europe, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2025-723, in review, 2026.