CrossMark

ORIGINAL PAPER

Daily precipitation grids for Austria since 1961—development and evaluation of a spatial dataset for hydroclimatic monitoring and modelling

Johann Hiebl[1] · Christoph Frei[2]

Abstract Spatial precipitation datasets that are long-term consistent, highly resolved and extend over several decades are an increasingly popular basis for modelling and monitoring environmental processes and planning tasks in hydrology, agriculture, energy resources management, etc. Here, we present a grid dataset of daily precipitation for Austria meant to promote such applications. It has a grid spacing of 1 km, extends back till 1961 and is continuously updated. It is constructed with the classical two-tier analysis, involving separate interpolations for mean monthly precipitation and daily relative anomalies. The former was accomplished by kriging with topographic predictors as external drift utilising 1249 stations. The latter is based on angular distance weighting and uses 523 stations. The input station network was kept largely stationary over time to avoid artefacts on long-term consistency. Example cases suggest that the new analysis is at least as plausible as previously existing datasets. Cross-validation and comparison against experimental high-resolution observations (WegenerNet) suggest that the accuracy of the dataset depends on interpretation. Users interpreting grid point values as point estimates must expect systematic overestimates for light and underestimates for heavy precipitation as well as substantial random errors. Grid point estimates are typically within a factor of 1.5 from in situ observations. Interpreting grid point values as area mean values, conditional biases are

reduced and the magnitude of random errors is considerably smaller. Together with a similar dataset of temperature, the new dataset (SPARTACUS) is an interesting basis for modelling environmental processes, studying climate change impacts and monitoring the climate of Austria.

1 Introduction

Spatially comprehensive datasets of essential climate variables are an important source of information for many monitoring, modelling and planning tasks. Knowledge on the spatial distribution of observed precipitation, for example, is of primary interest for river runoff management, hydropower generation, drinking water supply, agriculture and natural hazard prevention. Often, applications build on ready-made grid datasets used as input to environmental models, such as for modelling river runoff and drought risk (e.g. Blöschl et al. 2013; Haslinger et al. 2015; Parajka et al. 2015), for simulating vegetation and crop growth (e.g. Kapeller et al. 2012) or for calculating snow cover and glacier mass balance (e.g. Huss et al. 2008; Olefs et al. 2013). Requirements on the characteristics and quality of grid datasets vary between applications. However, with the growing interest into longer term climatic variations (possibly trends) and their effects on the environment, temporal coverage and long-term consistency become more and more important. Today, many applications call for grid data that satisfies high standards in long-term consistency while still offering the high spatial (kilometer scale) and temporal (daily) resolution necessary to model relevant processes.

Construction of precipitation grid datasets that accommodate both the requirements on long-term consistency and high resolution is faced with the challenge of finding a decent compromise: On one hand, reproduction of the high short-range variability of precipitation calls for a high station density to

✉ Johann Hiebl
johann.hiebl@zamg.ac.at

1 Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Hohe Warte 38, 1190 Vienna, Austria

2 Federal Office of Meteorology and Climatology (MeteoSwiss), Postfach 257, 8058 Zürich-Flughafen, Switzerland

sample the small-scale distribution comprehensively and to reliably estimate the statistics of precipitation-topography relationships by spatial interpolation procedures (e.g. Bénichou and Le Breton 1987; Daly et al. 1994; Prudhomme and Reed 1998; Gottardi et al. 2012; Mergili and Kerschner 2015). On the other hand, variations of the station network over time may introduce temporal inconsistencies in a grid dataset (e.g. Hofstra et al. 2010; Becker et al. 2013; Frei 2014), which could be avoided by interpolation from a temporally invariant network of high-quality stations (e.g. Schmidli et al. 2002; Schöner and Dos Santos Cardoso 2004; Hamlet and Lettenmaier 2005). The associated loss of information will, however, compromise the resolution of fine-scale precipitation patterns and topographic effects on the precipitation climate.

For the territory of Austria, several precipitation grid datasets have been constructed in the past with different spatiotemporal characteristics: (1) For the StartClim dataset, Schöner and Dos Santos Cardoso (2004) applied the AUREHLY (Analyse utilisant le relief pour l'hydrométéorologie) procedure and residual kriging to about 70 continuous high-quality station series. The dataset has daily resolution and extends over the period 1948–2007. The AUREHLY method (Bénichou and Le Breton 1987) is a regression-based interpolation technique that uses elevation as well as principle components of the neighbourhood topography as predictors. (2) Hasenauer et al. (2003) applied the Daymet method (Thornton et al. 1997), which combines local multiple linear regression with a Gaussian distance weighting. The dataset is primarily intended for ecosystem modelling over Austria, has daily time resolution, starts in 1960 and has been updated until 2012. The underlying station network varies from about 190 to 300 stations. (3) The GPARD-1 dataset (Hofstätter et al. 2015) was constructed by interpolating the ratio between daily measurements and a climatological background for 223 stations. For the interpolation, a smooth surface was fitted approximating the scattered data, with emphasis on smooth but anisotropic gradients in the optimisation. An adapted PRISM PRISM (Parameter-elevation Relationships on Independent Slopes Model) approach (Daly et al. 1994, 2008) using as many as 1400 stations was employed as climatological background. The dataset extends over the period from 1951 to 2006. (4) The HYRAS dataset (Rauthe et al. 2013) was constructed for the territory of Germany and the surrounding hydrological catchments, including large parts of Austria. It is based on up to 6200 stations (most of which in Germany), extends over the period 1951–2006 and was constructed by a combination of multi-linear regression with topogeographic predictors and inverse distance weighting. (5) The Alpine Precipitation Grid Dataset (APGD; Isotta et al. 2014), a pan-Alpine dataset, is based on about 5500 observations and covers the period 1971–2008. It was created with a combination of PRISM for the long-term monthly mean and a modified version of SYMAP for daily relative anomalies (Shepard 1984; see also Frei and Schär 1998). (6) The pan-European grid dataset E-OBS (Haylock et al. 2008) exhibits a spatial resolution of 0.25°, starts in 1950 and is regularly updated. The analysis combines climatological mean fields estimated by thin-plate smoothing splines (Hutchinson 1998) with daily anomaly fields determined by kriging (see also Hofstra et al. 2008). E-OBS uses over 2300 stations, 18 of which are located within Austria. (7) Unlike the so far mentioned datasets, INCA (Integrated Nowcasting through Comprehensive Analysis; Haiden et al. 2011) is not solely based on station data but also integrates information from a model-based data assimilation. INCA is the operational analysis and nowcasting system of the Austrian meteorological service ZAMG (Zentralanstalt für Meteorologie und Geodynamik) and covers many more climate variables. The analysis for precipitation uses inverse distance squared weighting of up to 1100 station measurements and, for the period since 2006 integrates climatologically scaled radar data. INCA is subject to considerable variations in data input and several changes in algorithms. Unless otherwise mentioned (i.e. expect for datasets (5) and (6)), all datasets have a grid spacing of 1 km.

Despite these many data sources, a precipitation grid dataset for Austria, suitable for applications with high requirements in temporal consistency, spatial resolution and near real-time availability, is missing. Such a dataset is highly desirable, given the interest into decadal variations of the precipitation climate and the need for monitoring these continuously. The aim of this paper is to present the development of a new grid dataset that can fill this gap. The dataset focusses on daily precipitation totals, exhibits a grid spacing of 1 km over Austria, extends back to 1961 and will be continuously updated. Its creation takes advantage of methodical developments and insights in the interpolation of precipitation in mountain regions. The approach builds on the common two-step practice with climatological background fields and daily anomaly fields. For the development of the climatological background fields, we utilise topographic predictors at several space scales carefully selected with exploratory cross-validation experiments. To increase the reliability of estimates at high elevation, the estimation of the climatological background fields also integrates data from totaliser devices. A particular ambition in our development is to ensure a high level of long-term consistency by minimising adverse effects from a temporally varying station network. To this end, a largely invariable network of input stations is used. Some efforts have been made in a comprehensive evaluation of the resulting dataset, to derive a range of summary statistics on the accuracy of the dataset and to point to limitations that may be critical to consider by users in quantitative applications.

The presented precipitation dataset accompanies an existing grid dataset of daily minimum and maximum air temperature on the same grid, over the same period and with similarly high requirements in long-term consistency (Hiebl

and Frei 2016). Together, these datasets constitute the official Austrian spatial climate dataset SPARTACUS (Spatiotemporal Reanalysis Dataset for Climate in Austria).

2 Data

The derivation of daily precipitation fields is effected in the familiar two-step approach with separate procedures for the mean monthly climatology (background fields) and the daily relative anomaly fields (e.g. Haylock et al. 2008; Becker et al. 2013; see later Sect. 3). Accordingly, two different observational datasets had to be created: one with mean monthly precipitation sums (Sect. 2.1) and one with daily precipitation sums (Sect. 2.2). In both cases, a compromise had to be made between maximum spatial coverage and continuous temporal coverage. This choice is outlined in the subsections below. A daily precipitation sum is defined here as the precipitation depth accumulated between 7 a.m. of the respective day and 7 a.m. of the following day, consistent with the standard reading time in Austria.

2.1 Mean monthly precipitation sums

Mean monthly precipitation sums form the basis of our climatological background fields. They were derived from daily station data obtained from several data providers. From all available stations, only those were considered further, which covered at least 20 years within the period 1961–2014. This selection resulted in 251 daily precipitation series from ZAMG, 879 from the Austrian provincial hydrographical services and 1365 from institutes in neighbouring countries. The 30 years of 1977–2006 had the best data coverage in this resulting dataset, and this period was, therefore, chosen as a reference period for our climatological background. To ensure approximate consistency of the resulting station normals, monthly averages for stations that did not fully cover the reference period were adjusted. The adjustment was estimated by regressing, individually, the relative anomalies of the incomplete records against five high-correlated stations extending over the entire period since 1961.

As additional data source, monthly precipitation data from 119 totalisers were included in our climatological dataset. In the topographically complex terrain of Austria, totalisers offer valuable information at high elevations, which are underrepresented by conventional rain gauge networks (see also Gottardi et al. 2012). Moreover, in the high Alpine areas of Austria, totalisers were found to provide more reliable observations of monthly precipitation sums compared to conventional ombrometers (Auer et al. 2000).

Altogether, our climatological dataset encompasses 2614 station normals for each calendar month (Fig. 1). One

thousand one hundred thirty conventional stations are within the national borders of Austria. Eight hundred fifty five of these cover a continuous 30-year period, while the remaining 275 stations cover 24.6 years on average. The effective reference period 1977–2006 is covered completely by 697 stations, and the remaining 433 station series cover 21.3 years on average.

In addition to the operational quality control conducted at each of the data providing institutions, we tested all data series for the possibility that interrupted operations were erroneously coded by zero values. To this end, cases with continuous zero reports over a full calendar month were compared to neighbouring stations. The zero reports were substituted by missing values, unless there was at least one station with less than 3 mm (monthly total) and at least three stations with less than 10 mm within a 50-km distance of the test station. This procedure identified 920 suspicious station months that were flagged.

In our dataset, no corrections were made for the measurement bias induced by wind deflection, wetting and evaporation. The bias has been quantified by Sevruk (1985) and Richter (1995) in the Alpine region (Switzerland and southern Germany, respectively) to range between about 7% at low-elevation or protected stations and up to 25% at wind-exposed high-elevation stations on an annual average. In situations with solid precipitation and high wind speed, the measurement error may, however, clearly exceed 50%. As a consequence, wind-exposed locations at high elevations are more affected, and the measurement bias is generally larger in winter compared to summer.

2.2 Daily precipitation sums

The selection of stations for the daily interpolation step is motivated by our ambition for long-term consistency and near real-time updating. To accommodate the first of these requirements, stations were included only, if their series covered at least 95% of all days during the period 1961–2014. This constraint was satisfied by 115 stations from ZAMG plus 408 stations from the provincial hydrographical services. The second of the above requirements implied restricting to stations from neighbouring countries that are accessible via SYNOP exchange. Altogether the daily dataset involves 566 station series, whereof 523 are located within Austria (Fig. 1).

The level of long-term homogeneity that can be expected in Austrian daily precipitation series was, among other things, assessed by Auer et al. (2010). Their results suggest that, for about 30% of the series, tests do provide some evidence for inhomogeneities. This must be considered a lower bound of the fraction of affected stations, given that there are limitations in break detection. Closer inspection of example series showed that the correction for breaks is capable of trend reversal in precipitation sums and indices. In the construction of
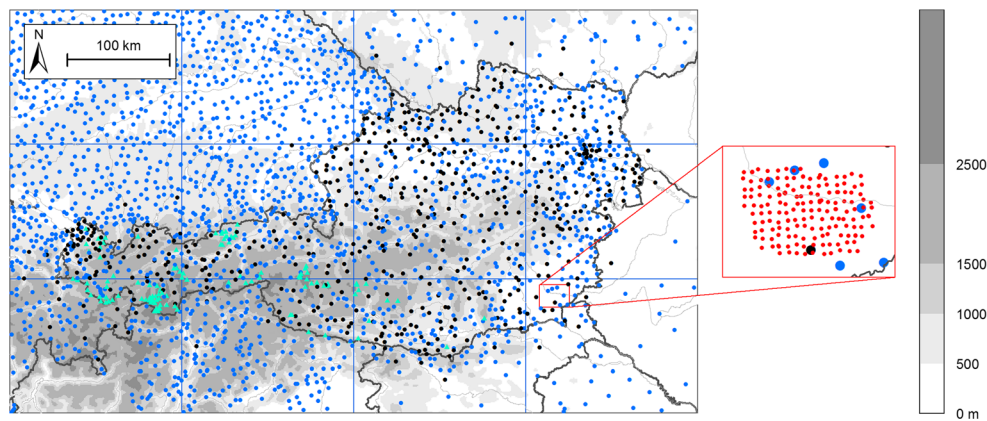
Fig. 1 Station networks used for the generation of the precipitation grid dataset. Data from all depicted stations (*black* and *blue dots*, *light blue triangles*) were used for the interpolation of mean monthly precipitation (background fields). Data from a station subset (*black dots only*) were employed for the interpolation of daily precipitation sums. *Dots* denote conventional rain gauges; *triangles* denote totalisers. Subregions for the interpolation of mean monthly precipitation sums are divided by *blue lines*. The stations of WegenerNet used for evaluation (region highlighted by *red frame*) are indicated by *red dots*. *Grey contours* denote terrain altitude (in metres above sea level)

our dataset, stations and periods, for which homogenised data are available, are included (Auer et al. 2010). Up to 2015, this applies to 4% of the observational station data used here. These efforts imply that inhomogeneities may still be present at least in roughly 25% of the stations incorporated.

Further efforts to improve data quality were made: (1) Where possible, gaps in most recent daily data are filled by aggregation of values from hourly and subhourly data sources. (2) Gross errors within this dataset were identified and suspicious values removed. Again, possibly erroneously coded "dry periods" were checked and eliminated. Moreover, a spatial consistency check was conducted to identify isolated dry and wet reports following suggestions of Scherrer et al. (2011).

Figure 1 reveals uneven station distribution over the study region. Sparser data coverage is obvious over parts of Salzburg and Styria. Moreover, there is a clear overrepresentation of low-elevation regions. Only 2.5% (13 stations) of the available observations are at altitudes above 1500 m a.s.l., whereas 21% of the area of Austria is at elevations above this threshold.

3 Interpolation

The daily precipitation fields of our grid dataset were constructed in a two-step process:

- Derivation of fields of climatological mean precipitation for each calendar month (background fields). The climatology represents conditions of the 30-year reference period 1977–2006 and was obtained by kriging with external drift (KED), using a set of topographic predictors (Sect. 3.1).
- For each day, calculation of relative anomalies of station observations with respect to the climatology of the respective calendar month. Fields of these anomalies were then derived by spatial interpolation, using an adapted version of the angular distance weighting algorithm SYMAP. Finally, multiplication of the anomaly field with the pertinent background field (Sect. 3.2).

As mentioned in the introduction, the two-step approach is common to many procedures of spatial precipitation analysis (e.g. Haylock et al. 2008; Rauthe et al. 2013; Isotta et al. 2014; Hofstätter et al. 2015). Details within each of the two steps may, however, vary considerably. The main purpose of separating the climatological from the daily timescale is to reduce systematic errors that may result from the non-representative vertical distribution of the observing stations (see, e.g. Masson and Frei 2014). Imprints of the topography on precipitation at scales not resolved by the station network are difficult to identify on a day-by-day basis due to large spatial variations. A climatological reference in the daily interpolation imposes such fine-scale patterns from the long-term mean. Even though these patterns may be of limited representativity for a particular day, the procedure ensures that the resulting daily dataset encapsulates major fine-scale effects and, hence, is consistent with an estimate of the climatology.

The grid of our spatial analysis is a regular 1-km mesh in the metric coordinate system ETRS89/Austria Lambert. The digital elevation model adopted in our analysis is from the global Shuttle Radar Topography Mission (SRTM) model (Farr et al. 2007), with elevations on the metric grid obtained by averaging the original 3″ (90 m) resolution SRTM values over the 1-km grid pixels.

3.1 Mean monthly background fields

The derivation of the climatological background fields for our precipitation analysis over Austria is based on KED (Schabenberger and Gotway 2005; Diggle and Ribeiro

2007). KED is a geostatistical interpolation model consisting of a deterministic component, which describes the target variable as a linear combination of known spatial fields (hereafter denoted as "covariates" or "external drift") and a stochastic component, which models the residuals as a random Gaussian field with a parametric spatial covariance function. KED is related to regression kriging and is a popular framework for interpolation in climatology, due to its high flexibility and the possibility to integrate auxiliary spatial information (e.g. Goovaerts 2000; Hengl et al. 2007; Perčec-Tadić 2010; Aalto et al. 2012; Frei et al. 2015).

Masson and Frei (2014) have examined different configurations of KED for the interpolation of seasonal mean precipitation in a north–south section across the Alps that also contained parts of Austria. Their experiments showed that the inclusion of informative spatial covariates and the modelling of spatial covariance, both, lead to a considerable reduction of interpolation errors. Moreover, model configurations with sets of covariates that encompass several space scales of the underlying topography were in advantage over configurations with only one, the 1-km scale. In our application of KED, we closely follow the procedure described in Masson and Frei (2014), but we reinvestigate the specification of a set of covariates that is informative generally over all of Austria. Again, the experiments conducted for that purpose are similar to those of Masson and Frei (2014), and results are described further below.

To allow for varying relationships of mean precipitation with topography, KED is applied in this study individually for each calendar month and separately for 12 similar-sized rectangular subregions (see Fig. 1). The number of stations per subregion ranges from 95 to 467 with an average of 310. The results for the subregions are then combined by gradual merging over a 40-km boundary zone. To better comply with the assumptions of Gaussian residuals and stationary variance, KED is applied with square root-transformed data, followed by a numerical back-transformation of results (see, e.g. Erdin et al. 2012).

For the stochastic component, we choose an exponential semi-variogram with a range and sill parameter plus a nugget effect. The nugget effect is included to account for measurement uncertainties and short-range variability that may be related to the local environment of the measurement device (see also Goovaerts 2000; Perčec Tadić 2010; Masson and Frei 2014). The nugget to sill ratios in individual subregions vary from 6 to 22% in February to 3 to 17% in August and seem to depend on station density and topographic-climatic complexity of the subregions. All parameters, those of the semi-variogram and the drift coefficients, are simultaneously estimated by restricted maximum likelihood using the R package geoR (see Diggle and Ribeiro 2007).

To identify an informative (and parsimonious) configuration of the deterministic component of KED, an exploratory experiment with several different sets of topographic covariates is performed. The testing is done similarly to that in Masson and Frei (2014) but for all the rectangular subdomains (see Fig. 1). Topographic elevation as well as east–west and north–south gradients of the topography, all at several different space scales, are considered as constituents in different sets for the external drift. Candidate covariates at scales other than the original 1-km resolution are obtained from spatially filtered versions of the elevation field, derived using a Gaussian filter with different window widths. The sets of covariates that are compared in this experiment are listed in Table 1 (column 1). Covariate sets with elevation and gradients are labelled as *eg*, sets with only elevation as *e*. The scales included in the set are indicated by integer numbers separated by hyphen, e.g. the set *eg1-9* includes elevation and two gradient fields (north–south, east–west), all three of them at the original 1-km scale and at the smoothed 9-km scale. The number of covariates in this example is $3 \times 2 = 6$ (Table 1, column 2). In cases when several scales were included, the finer scale fields are taken to be the residuals from the coarser scale fields in order to avoid colinearity in the resulting combined set. Scales considered in our experiment range from the original 1-km up to a 17-km smoothing.

Table 1 summarises results obtained with each of the different configurations of the external drift. The numbers in columns 3 and 4 are the error scores used in Masson and Frei (2014) (see their Eqs. (4) and (5)), namely *relative bias* (ratio of the average of predicted versus the average of observed values) and the *relative mean root transformed error* (rel. MRTE, ratio of mean squared error versus spatial variance calculated with root-transformed values) from a station-by-station cross-validation. Columns 5 and 6 represent values of the AIC and BIC, common likelihood-based model selection criteria, where smaller numbers indicate better performance. All the numbers listed are averages of the scores from the individual experiments in each subregion and each calendar month.

In summary, the results of our exploratory analysis (Table 1) suggest that among all elevation-only, single-scale covariate sets (*e1* to *e17*), a smoothing at the 5-km scale (*e5*) delivers the best result. A further reduction of interpolation errors can be achieved when combining the 5-km elevation with the shortest scale elevation covariate (*e1-5*). Combination of three smoothing levels (e.g. *e1-5-9*) does not reduce interpolation errors further. If topographic gradients are included in the external drift (*eg1* to *eg9*), errors at all smoothing levels are reduced compared to the pertinent elevation-only model. Finally, including elevation and gradients at several smoothing levels (e.g. *eg1-5*) seems to be of advantage and—in most cases—improves the scores of the single-scale models. Again, models with three scales are at best marginally better than models with two scales. This is not too surprising, considering that even coarser space scales begin to be resolved by the station network

Table 1 Error measures and model selection criteria for interpolation experiments with kriging with external drift using different sets of topographic covariates

Experiment	Number of predictors	Rel. bias (B)	Rel. MRTE (E)	Akaike information criterion (AIC)	Bayesian information criterion (BIC)
e1	1	1.00	0.071	1848	1867
e3	1	1.00	0.067	1828	1847
e5	1	1.00	0.067	1824	1843
e7	1	1.00	0.069	1833	1852
e9	1	1.00	0.072	1846	1865
e11	1	1.00	0.076	1860	1879
e13	1	1.00	0.078	1871	1890
e15	1	1.00	0.081	1883	1902
e17	1	1.00	0.084	1894	1913
e1-3	2	1.00	0.068	1826	1849
e1-5	2	1.00	0.067	1820	1843
e1-9	2	1.00	0.067	1822	1845
e1-3-5	3	1.00	0.067	1821	1848
e1-5-9	3	1.00	0.067	1820	1846
eg1	3	1.00	0.071	1846	1872
eg3	3	1.00	0.066	1820	1847
eg5	3	1.00	0.065	1815	1841
eg9	3	1.00	0.072	1841	1867
eg1-3	6	1.00	0.066	1818	1856
eg1-5	**6**	**1.00**	**0.065**	**1811**	**1849**
eg1-9	6	1.00	0.067	1813	1851
eg1-3-5	9	1.00	0.066	1810	1860
eg1-5-9	9	1.00	0.065	1806	1856

The numbers listed represent an average over all the subregions (blue lines in Fig. 1). The selected experiment is highlighted in bold type. See Sect. 3.1 for details

explicitly. Among all our experiments, the external drift with elevation and gradients at the 1 and the 5-km space scale (eg1-5) shows the smallest error measures on average. A stratification of the analysis in Table 1 by calendar month revealed that this set of covariates has the smallest rel. MRTE and AIC throughout the year. Also, BIC was smallest except in summer. As for the variation in space, best performing drift models differed between subregions with no contiguous geographical pattern. Quality measures were likely non-robust given the limited number of cross-validation samples in individual subregions. Based on these analyses, we decided to choose the set eg1-5, encompassing six covariate fields in total, in all the KED models for all the subregions and calendar months. A geographically uniform drift model was also preferred to avoid excessive discontinuities of the predictive model at the subregion borders. Note that the coefficients are estimated individually per subregion and month, so that the deterministic component of KED, despite the structural constraint, is flexible to vary between subregions and seasons.

Figure 2 illustrates the selected predictors for a map section in central Austria. The graphs reveal the highly different space scales included in the two components. Both of the covariate sets involve patterns at scales that are not explicitly resolved

by the station network. The distribution of long-term mean precipitation for two example months (February and August) as deduced by KED using these covariates is depicted in Fig. 3 (left column). Both of the examples reveal fine-scale patterns introduced by the covariates, yet with a slightly different appearance, evinced, for example, in a stronger elevation dependency in February compared to August.

3.2 Daily anomaly fields

The spatial analysis of daily precipitation anomalies (percentage of the daily sum with respect to the climatological mean in the pertinent calendar month) builds on the angular distance weighting scheme SYMAP (Shepard 1984). SYMAP has frequently been applied for spatial precipitation analysis at monthly and daily timescales (e.g. Legates and Willmott 1990; Xie et al. 1996; Frei and Schär 1998; Hamlet and Lettenmaier 2005; Becker et al. 2013). Here, an adapted version of the algorithm is applied, which enhances the flexibility when the density of station data varies across space or over time (Frei and Schär 1998). In this modification, the search radius is adjusted to the local station density, by selecting, among a sequence of possible radii, one that includes at least
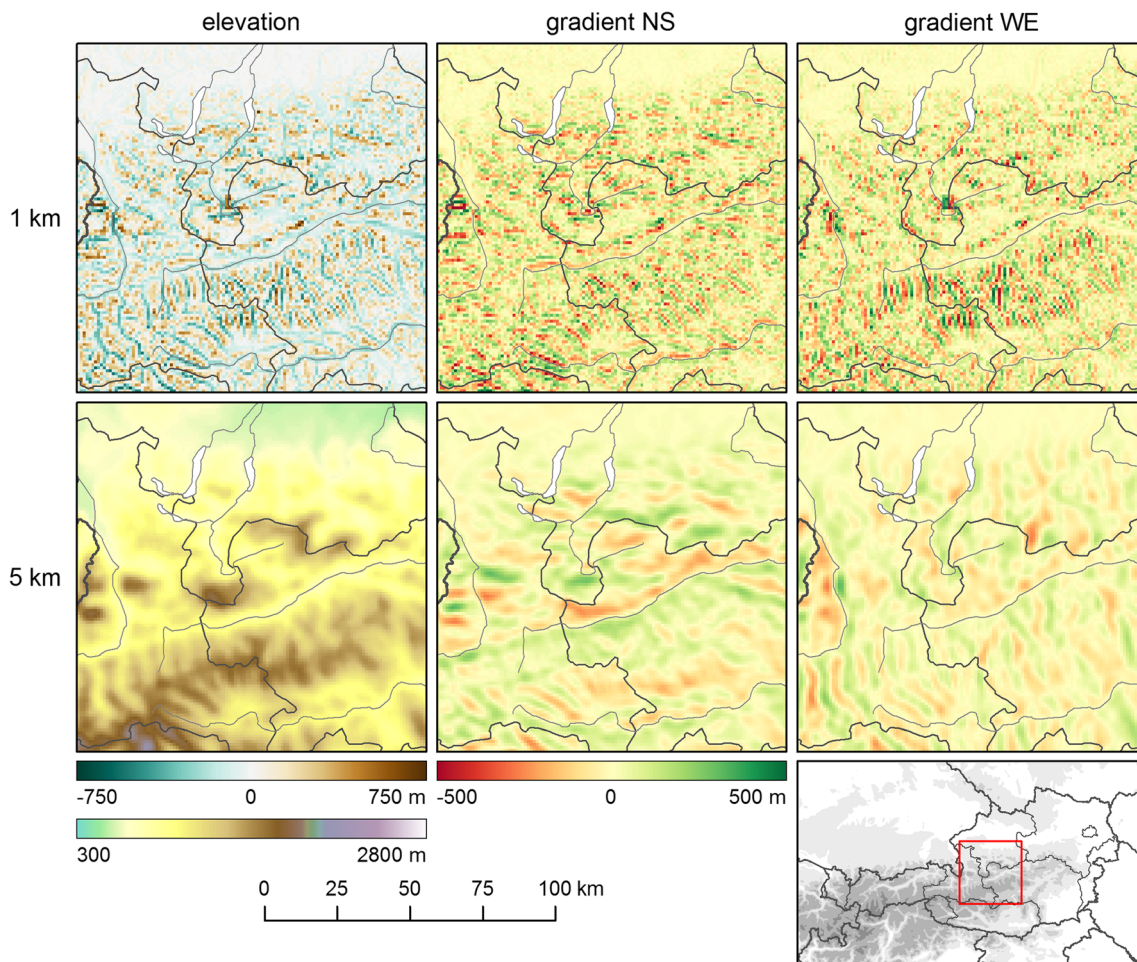
Fig. 2 Topographic covariates (elevation, north–south gradient and west–east gradient) at two different space scales (at *1* and *5 km* in the *top* and *bottom row*, respectively) chosen as external drift of KED in the interpolation of mean monthly precipitation. The *panels* show conditions over a section of the study region in central Austria (see *inset*). Note that the 1-km covariate is defined as a residual with respect to the pertinent 5-km fields

3 but not more than 30 stations. The factor for the relative strength of angular weighting in relation to distance weighting is set to 3, implying that the angular weight is not allowed to attain more than the threefold value of the distance weight. From the interpolated field of relative anomalies, the final distribution of the daily precipitation sum is obtained by multiplication with the monthly background climate field. Note that SYMAP does not formally account for the intermittence of daily precipitation. Dry and wet observations are not distinguished methodologically. This, and the limitations of the station network to resolve small-scale pattern, will lead to smoothing effects that will be further investigated in Sect. 5.2.

For illustration, Fig. 3 depicts the relative anomaly fields and the pertinent fields of daily precipitation sums for two exemplary days: one with stratiform precipitation along the northern edge of the Alps (03 February 2010) and the second for a summer day with a large convective cell in the interior of the Alps (26 August 2009). Note that the imprint of fine-scale patterns from the climatology varies between the two cases.

Many of the local maxima/minima in the stratiform case are found in areas between stations. They represent systematic topographic enhancement estimated in the climatology. In contrast, the amplitude and shape of the local maximum in the convective case are mostly determined by the station values themselves with the background climatology having a more modest influence.

4 Example results

The new precipitation grid dataset for Austria encompasses about 21,000 fields of daily precipitation sum covering the period since 1961. In addition, monthly, seasonal and annual precipitation sums as well as anomalies with respect to the reference period 1961–1990 have been calculated from the daily fields. The dataset is updated on a daily basis. Together with the already existing dataset of daily temperature (Hiebl and Frei 2016), the new results constitute a data source that is
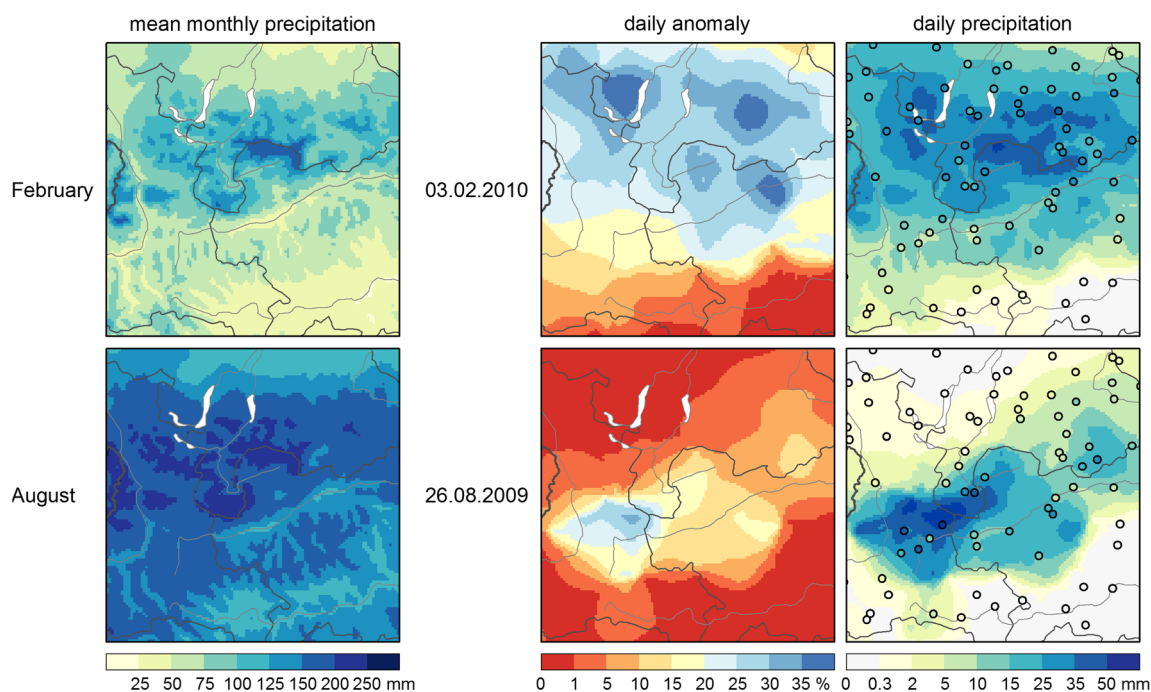
Fig. 3 The background fields (mean monthly precipitation in millimeter) for February (*top*) and August (*bottom*) (*left column*). Two examples of daily anomaly fields (percent of mean monthly precipitation) for a February day (*top*) and an August day (*bottom*) (*center column*). The final fields of daily precipitation for the two example days (*right column*). Circle symbols represent the original station observations. The maps show the same section as that in Fig. 2

suitable for applications, where long-term consistency is essential. The SPARTACUS dataset provides the basis for spatial climate monitoring at ZAMG.

The purpose of this section is to illustrate selected results of our new grid dataset and to provide a qualitative comparative assessment. In Sect. 4.1, results from the present analysis are compared to some of the other grid datasets for Austria. Section 4.2 will illustrate an application, where the new dataset is explored for a climatology and trends of two prominent indices of daily precipitation. A quantitative evaluation of the dataset will be described in Sect. 5.

4.1 Daily precipitation

Figure 4 compares examples of daily precipitation fields from SPARTACUS with pertinent fields from other precipitation grid datasets, including StartClim, Daymet, GPARD-1, E-OBS and INCA, all of which were shortly introduced in Sect. 1. The cases were selected because of their distinct meteorological situation without considering the possible plausibility of the fields beforehand. The cases are briefly discussed individually.

- Continental cyclone, 06 August 1985 (Fig. 4a): A pronounced low-pressure area over Central Europe leads to heavy rainfall throughout Austria. All five available datasets reproduce the general pattern in a similar manner. E-OBS does not reproduce as much regional detail as the

other datasets due to its coarser grid spacing. Daymet and StartClim exhibit very smooth structures. Both datasets are subject to border effects, namely a decrease of precipitation sums in some border-near regions, probably related to the lack of stations outside the study region. GPARD-1 (shown) and SPARTACUS provide very similar patterns. Finer, strongly topography-related structures are visible in GPARD-1.

- Orographic lifting at the northern edge of the Alps ("Nordstau"), 31 July 1977 (Fig. 4b): A classical Vb-track cyclone (Van Bebber 1891) causes heavy precipitation from Vorarlberg to western Lower Austria leading to flooding and mudslides. In E-OBS (v. 12.0, shown), the area with heavy precipitation is confined to the western part of the Northern Alps. All other datasets extend this area further to the east. Daymet shows a very sharp transition between areas with heavy and with no precipitation, probably due to the binary precipitation (on/off) model in the interpolation and gross errors in the station data. StartClim predicts a stronger spread of precipitation to the south of the Alpine main crest compared to the other analyses. GPARD-1 and SPARTACUS exhibit similar patterns again, but GPARD-1 estimates a stronger increase of precipitation with height, resulting in more extreme precipitation sums at high altitudes.

- Orographic lifting at the southern edge of the Alps ("Südstau"), 01 September 1965 (Fig. 4c): This case represents a meteorological condition, where gradual rise of
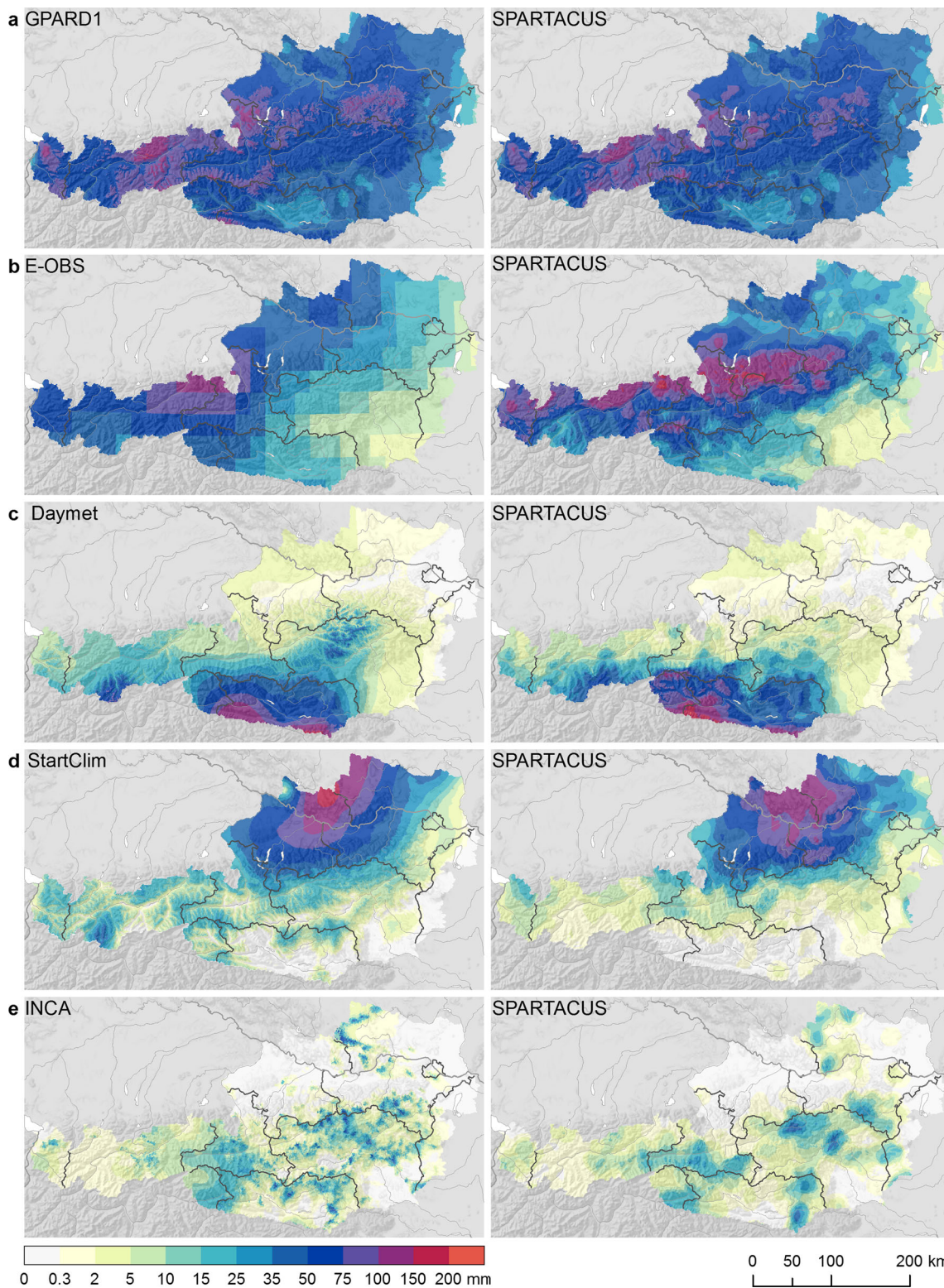
Fig. 4 Comparison of daily precipitation analyses in Austria as derived by SPARTACUS and several existing gridded datasets (in millimetre). The depicted cases represent distinct meteorological situations. **a** Continental cyclone, 06 August 1985. **b** Orographic lifting at the northern edge of the Alps, 31 July 1977. **c** Orographic lifting at the southern edge of the Alps, 01 September 1965. **d** Alpine cut-off low, 07 August 2002. **e** Convective activity, 23 July 2006

humid warm air over previously advected cold air led to very large precipitation amounts along the southern border of Austria. The case caused catastrophic flooding in Carinthia and East Tyrol. Areas on the northern side of

the Alps remain mostly dry. E-OBS concentrates the area with highest precipitation amounts along the south-eastern part of Austria's southern border (to the Karawanks range) and shows less precipitation along the south-western part (in the Carnic Alps), in stark contrast to the distributions in other datasets. The distributions from GPARD-1 and SPARTACUS are very similar. Daymet (shown) exhibits an overall quite similar pattern to SPARTACUS but has heavier precipitation in central parts of Austria (northern Styria). Areas with more pronounced topographic fine-scale structure and with stronger horizontal gradients co-exist in Daymet. StartClim simulates mostly dry conditions in East Tyrol, clearly misinterpreting the situation in comparison to the other datasets.

- Alpine cut-off low, 07 August 2002 (Fig. 4d): A high-level trough moving across Austria and the upglide of warm and humid air upon cooler air beneath cause very heavy precipitation especially in Upper and Lower Austria. In combination with a second event 4 days later, the case led to catastrophic flooding. In this case and taking into account the coarser resolution, E-OBS is in good agreement with the other datasets. Daymet again exhibits strong horizontal gradients. StartClim (shown) estimates stronger increases of precipitation with altitude in western and southern Austria compared to the other datasets. GPARD-1 and SPARTACUS draw a detailed and consistent picture, with SPARTACUS being somewhat smoother.

- Convective activity, 23 July 2006 (Fig. 4e): A flat pressure distribution and unstably layered warm air characterise this hot summer day with scattered thunderstorms. The analysis from INCA (available from 2006 onwards, shown) provides a richly structured pattern with many cells of high-intensity convective rainfalls. The high spatial resolution seen in INCA is due to the incorporation of radar data. The characteristics of the pattern are quite variable though. The distribution is much smoother in inner Alpine regions, a possible reason of which may be that the visibility of the radar is shadowed at low elevations there. Other examples of the INCA precipitation analysis show artificial radial structures close to radar locations. Again, these are related to the difficulties of radar-based precipitation estimation in complex terrain. Due to its coarse resolution and the limited station data available, E-OBS does not recognise the small-scale character of the precipitation pattern. Also, Daymet and StartClim exhibit strongly smoothed patterns, though less so compared to E-OBS. StartClim misses important details such as the rainfall activity in the north of Austria. GPARD-1 and SPARTACUS reproduce the pattern seen in INCA, with all the larger scale features incorporated, but with a smoother appearance. Isolated small convective cells like that in Upper Austria (see INCA) are not registered by the station network and are, therefore, missing in all station-only-based analyses. It is unclear, whether all the small-scale features in the radar composite, evident in INCA, were truly associated with rainfall at the ground.

4.2 Climate analysis

Here, we illustrate the utility of the developed grid dataset for climatological analyses by investigating the spatial distribution and long-term variation of two popular indices that characterise the climate of daily precipitation. These are the wet day frequency, defined as the annual number of days with at least 1 mm of precipitation, and the wet day intensity, defined as the average daily precipitation sum on all wet days of the year (see also Klein Tank et al. 2009).

The occurrence of wet days (Fig. 5a) reveals remarkable spatial variations that contribute to the climatic diversity of this small country. Along the northern rim of the Alps and, in parts, also along the main Alpine crest (especially the Hohe Tauern range), the mean wet day frequency is as large as 180 days per year, corresponding to a wet day every second day on average. This is not true for parts of inner Alpine Tyrol (the Ötztal Alps), where valleys are particularly dry. The most pronounced horizontal gradient is found from the summits of the Hohe Tauern (forming the border between the state of Salzburg with Carinthia and East Tyrol) towards the nearby dry southern valleys. The smallest number of wet days is recorded in north-eastern Austria, where, locally, there are only 75 wet days on average per year, corresponding to one day out of five only.

The distribution of wet day intensity looks quite different (Fig. 5c). Average intensities are small (4 to 8 mm) in the northern and eastern forelands as well as in the Ötztal Alps. Values are larger (8 to 14 mm) along the northern rim of the Alps, and maximum values (14 to 20 mm) are observed in the Southern Alps (Carnic Alps and Karawanks). Together, the analyses for the two indices reveal a contrast between a "frequent-moderate" precipitation climate along the Northern Alps and an "episodic-heavy" precipitation climate along the Southern Alps. These climatological differences have implications for many geological, hydrological, ecological and glaciological processes, and the SPARTACUS dataset provides a profound basis to incorporate these high-frequency (daily) climate characteristics in modelling studies of these processes.

Trends in the two precipitation indices over the period 1961–2014 are assessed in a grid point by grid point trend analysis of the annual index values. In the case of the number of wet days, this assessment is made using logistic regression, an extension of classical least-squares regression for binomial count data (see, e.g. McCullagh and Nelder 1989; Frei and Schär 2001). The null hypothesis "probability of wet days has
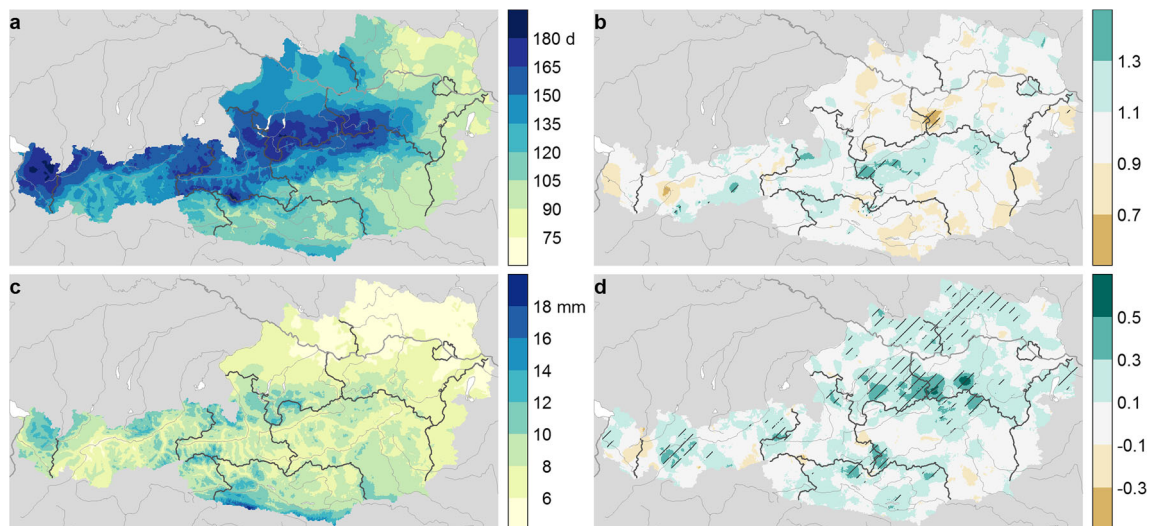
Fig. 5 a Mean annual number of wet days (≥1 mm) during the period 1961–1990. b Long-term trend (1961–2014) in annual number of wet days, expressed as the odds ratio of a logistic regression of the annual counts. c Mean annual precipitation intensity (mean precipitation on wet days) during the period 1961–1990. d Long-term trend (1961–2014) in annual precipitation intensity, expressed as the Theil-Sen linear trend slope. Hatching in b and d denotes areas, where trends are significant, using a meta-test controlling the false discovery rate at 5% (see Wilks 2016)

not monotonically changed" was tested at the 5% significance level, correcting for overdispersion. The magnitude of the trend is expressed as the ratio of odds (the fraction of wet day to dry day probability) between the end and the beginning of the period. In the case of wet day intensity, the trend analysis was made with the non-parametric Mann-Kendall trend test (e.g. Mann 1945; Kendall 1948) again at the 5% significance level. The magnitude of change is expressed by the Theil-Sen slope (change in millimetre per 54 years; Theil 1950; Sen 1968). The set of p values at the grid points were subjected to a meta-test controlling the false discovery rate (see Benjamini and Hochberg 1995; Wilks 2016).

Results for the annual frequency of wet days (Fig. 5b) show a fairly scattered picture of increases and decreases that are statistically significant in only small contiguous regions. The proportion of statistically significant grid points is small (2.0%, i.e. similar to the significance level), and hence, the results should not be interpreted as evidence for a long-term change in this index. Some of the patches may also be due to residual inhomogeneities in the station data.

The trend analysis for annual wet day intensity (Fig. 5d) suggests that this index has increased in the study period over large parts of Northern Austria. The trend is statistically significant for a considerable fraction of the grid points in this region (16.4% over all of Austria). For the rest of the country, there are no larger contiguous regions with positive or negative trends. More detailed analyses show that the regional increase in wet day intensity is evident in all seasons, except winter, and is especially pronounced in autumn along the northern rim of the Alps.

There are many more precipitation indices for which a climatology and trend analysis would be of interest. The two examples illustrate the utility of the new grid dataset for such analyses. Constructing the dataset from an almost continuous station dataset has helped to avoid unrealistic patches of large-amplitude trends, often seen in other grid datasets and resulting from variations in the underlying station density over time. Still, care should be exercised in the interpretation of small-scale structures on a trend map, considering that only few of the station series incorporated in our dataset are homogenised for instrumental changes and station relocations. However, significant trends over larger scale contiguous regions, i.e. involving many stations, are unlikely pure artefacts from residual inhomogeneities in the underlying station data.

5 Evaluation

In this section, we thoroughly evaluate the precipitation grid dataset in order to obtain a quantitative understanding of the dataset's accuracy and to illustrate its potential and limitations for users. In accordance with the two main steps of the production (cf. Sect. 3), we conduct leave-one-out cross-validations both for the construction of the background fields (mean monthly precipitation, Sect. 5.1) and that of the daily precipitation fields, which constitute the final grid dataset (Sect. 5.2).

A general caveat of leave-one-out cross-validation is that it quantifies error statistics for the special case when grid point estimates are interpreted as precipitation at the point scale, the scale of the verifying measurements. In fact, this interpretation

is rather uncommon. Most users will interpret grid point estimates as area means (typically over a grid pixel) or will even spatially aggregate grid points to form catchment mean values. For these more common interpretations, cross-validation errors are rather pessimistic accuracy measures. Therefore, in addition to classical cross-validation, we quantitatively compare the grid dataset to independent observation data from an experimental station network in a small section of our study region. The very dense spacing of stations in that network allows an evaluation at scales larger than the point scale (Sect. 5.3).

5.1 Mean monthly background fields

The accuracy of the 12 mean monthly background fields (Sect. 3.1) is quantified by systematic leave-one-out cross-validation, restricted to the subset of 1249 Austrian stations. In this evaluation, error statistics are calculated based on the logarithm of ratios between predicted (y_i) and observed (o_i) mean monthly precipitation sums. Log transformation ensures that multiplicative underestimates and overestimates are equally penalised. Back-transformation allows the statistics to be interpreted as relative errors. The *bias*

$$B = \exp\left(\frac{1}{N}\Sigma_i\left[\log\left(\frac{y_i}{o_i}\right)\right]\right) \tag{1}$$

informs about the systematic error component. The *root mean squared error fraction*

$$\text{RMSF} = \exp\left(\sqrt{\frac{1}{N}\Sigma_i\left[\log\left(\frac{y_i}{o_i}\right)\right]^2}\right) \tag{2}$$

can be regarded as the "average multiplicative error". Root mean squared error fraction (RMSF) will usually be dominated by random errors, except in areas with large biases. For both, B and RMSF, a value of 1 indicates optimal performance.

Evaluation results stratified by altitude band and climatological season are listed in Table 2. Biases are very small.

Across all categories, values are within ±2%. This result is particularly satisfying for the topmost elevation band, where systematic errors can easily be incurred from inaccurate or non-representative models of the small-scale precipitation topography relationship (see, e.g. Masson and Frei 2014). RMSF takes a value of 1.11 when averaged over all stations and seasons implying an average relative error of 11%. Errors are typically smaller in summer and autumn (1.07 and 1.09, respectively) compared to winter and spring (1.15 and 1.11). The spatial patterns of mean monthly precipitation are more gradual in summer compared to the pronounced altitudinal variations in winter. Larger measurement errors in winter may also contribute to this contrast. Throughout the four seasons, there is a gradual increase of RMSF from low to high altitude bands, with smallest values for stations below 500 m a.s.l. in summer (1.06) and largest values for stations above 1500 m a.s.l. in winter (1.23). The larger spatial variance in the precipitation climate within the Alps and at high altitudes and the decrease of station density with elevation are factors responsible for this remarkable contrast in interpolation accuracy.

5.2 Daily fields

The actual daily gridded precipitation dataset (Sect. 3.2) is evaluated here in slightly more detail to uncover some of the systematic effects of spatial interpolation that users should be aware of. The leave-one-out cross-validation was calculated for a period of 10 years (2005–2014, 3652 days) and for the subset of 523 stations within Austrian borders. Several aspects of the correspondence will be examined, including the mismatch between dry or wet conditions (intermittence), typical quantitative errors during wet episodes as well as errors in reproducing long-term variations.

Figure 6 shows box plots of the relative interpolation error (y_i/o_i, the ratio of predicted versus observed precipitation) as a function of precipitation intensity. For this purpose, the ratios are binned into classes defined in terms of quantiles of the distribution of observed daily precipitation (x-axis). Ratios are only

Table 2 Cross-validation errors for the mean monthly background fields

Altitude band	n	B					RMSF				
		DJF	MAM	JJA	SON	Year	DJF	MAM	JJA	SON	Year
<500 m	475	1.00	1.00	1.00	1.00	1.00	1.10	1.07	1.06	1.06	1.07
500–1000 m	443	1.02	1.01	1.00	1.01	1.01	1.14	1.10	1.06	1.09	1.10
1000–1500 m	185	1.02	1.01	1.00	1.01	1.01	1.19	1.14	1.08	1.11	1.14
>1500 m	146	0.99	1.00	1.00	1.01	1.00	1.23	1.18	1.12	1.14	1.17
All	1249	1.01	1.01	1.00	1.01	1.01	1.15	1.11	1.07	1.09	1.11

Bias (B) and root mean squared error fraction (RMSF), stratified by altitude bands and seasons. n is the number of stations per altitude band. (See Sect. 5.1 for details)
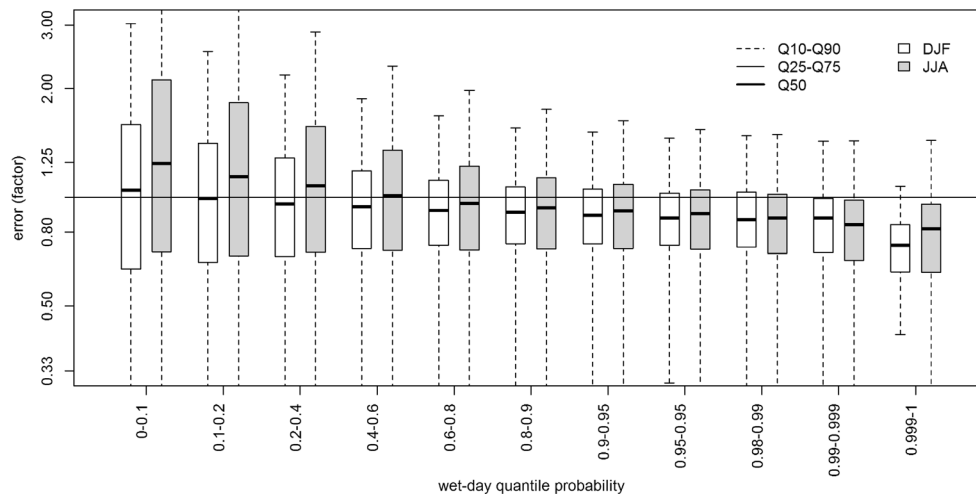
Fig. 6 Box plots of the interpolation error inferred from a systematic leave-one-out cross-validation. Errors are expressed as ratio between interpolated (at the location of the station) and observed (at the stations) values. The sample of errors is stratified into bins of precipitation intensity, which are defined in terms of quantiles. The sample includes only cases with estimated and observed precipitation ≥1 mm per day. Box plots represent the median (*bold line*), the interquartile range (*box*) and the 10–90% quantile range (*whisker*) of the error distribution. Results are shown separately for winter (*DJF*) and summer (*JJA*)

considered when estimated and observed precipitation is larger or equal 1 mm per day. The *y*-axis is log-transformed to ensure symmetry for the same relative overestimates and underestimates. Offsets of the boxes against relative errors of 1 (no error) indicate tendencies for systematic underestimation and overestimation in the respective intensity class. The vertical extent of the boxes is a measure of error spread in the respective class. Results are shown separately for winter and summer.

The box plot reveals a characteristic feature in the error structure that is common to most interpolation procedures, namely a tendency for small precipitation intensities (left hand boxes) to be overestimated and for large precipitation intensities (right hand boxes) to be underestimated. This conditional bias is a manifestation of the smoothing effect of spatial interpolation. A local precipitation maximum (minimum) is likely to be underestimated (overestimated). The effect may be quantitatively significant, in that grid point values tend to underestimate heavy precipitation by about 20% (boxes on the right). The conditional biases are larger in summer compared to winter, particularly at low intensities. The more constrained spatial extent of convective precipitation invokes more serious "smearing-out" effects. It is important for users to bear in mind that the SPARTACUS precipitation grid dataset (like most other interpolation-based datasets, see, e.g. Isotta et al. 2014) is conditionally biased when grid point values are interpreted as point estimates. Associated biases in the frequency distribution may be relevant in applications. It may be somewhat appeasing, though, that the magnitude of the conditional biases reduces when grid point estimates are interpreted as area mean values rather than point values.

In the following, the accuracy of the daily grid dataset is summarised by four skill measures that characterise different aspects of the correspondence between grid point estimates and station observations in the cross-validation set. The first two focus on the distinction between dry and small events versus moderate and intense events. For this binary evaluation, a threshold of 3 mm is chosen. As *frequency bias*, we define the ratio of predicted versus observed frequencies of events above this threshold. It quantifies systematic errors in the occurrence of moderate and intense precipitation without regard to spatiotemporal correspondence. The latter is quantified by means of the *Hanssen-Kuipers* (HK) *discriminant*, a frequently used skill score for binary deterministic forecast evaluation (see, e.g. Wilks 2011). HK is sometimes referred to as Pierce skill score (see, e.g. Hogan and Mason 2012). HK can take values between −1 and 1, with HK = 1 representing a perfect binary forecast and HK < 0 representing forecasts with less skill than a purely random forecast.

Table 3 lists the two summary measures of the binary evaluation as a function of altitude band and season. The frequency bias reveals that grid point estimates tend to slightly overestimate the frequency of events above the 3-mm threshold. This reflects the conditional bias at low precipitation intensities, previously seen in Fig. 6. (Note that 3 mm of daily precipitation roughly corresponds to the 30% quantile of events larger than 1 mm in Austria.) Consistent with the interpretation provided previously, the overestimate is slightly larger in summer than in winter. The frequency bias decreases with elevation turning into a minor underestimate above 1500 m above sea level. The spatiotemporal correspondence of events above the 3-mm threshold attains values of HK around 0.84. Values for all seasons and altitude bands are above or near 0.8, indicating a high level of correspondence, far above the score HK = 0 of a random forecast.

The two other statistics of this evaluation focus on the accuracy of the grid point estimates for occurrences above

Table 3 Cross-validation statistics for the daily precipitation grid dataset

Altitude band	n	Frequency bias					Hanssen-Kuipers				
		DJF	MAM	JJA	SON	Year	DJF	MAM	JJA	SON	Year
<500 m	215	1.03	1.04	1.07	1.04	1.05	0.83	0.84	0.83	0.87	0.85
500–1000 m	207	1.03	1.04	1.07	1.04	1.05	0.83	0.83	0.82	0.87	0.84
1000–1500 m	88	1.00	1.02	1.04	1.02	1.02	0.81	0.81	0.81	0.86	0.83
>1500 m	13	0.94	0.97	1.01	0.98	0.98	0.78	0.77	0.80	0.82	0.80
All	523	1.02	1.04	1.06	1.03	1.04	0.83	0.83	0.82	0.87	0.84

Frequency bias and Hanssen-Kuipers discriminant for threshold exceedance of 3 mm per day, stratified by altitude bands and seasons. n is the number of stations per altitude band. (See Sect. 5.2 for details)

the 3-mm threshold. To this end, B and RMSF (as given in Eqs. (1) and (2)) are calculated for daily precipitation amounts (restricted to $o_i \geq 3$ mm per day). Table 4 lists these statistics. As for the bias, there is a slight underestimation, which likely reflects the conditional underestimate at the upper tail (see Fig. 6) but is somewhat compensated by the frequency overestimate near the threshold (see Table 3). Despite these small

systematic errors, values of the RMSF indicate partly considerable random errors. Typical average multiplicative errors are in the order of 1.5. Clearly, the high values are partly due to the large number of days with precipitation only slightly above 3 mm, where minor absolute errors can be large in relative terms. The magnitude of relative errors clearly decreases with precipitation intensity (see Fig. 6). Quite plausibly, the RMSF

Fig. 7 Station-wise bias (B, **a**) and root mean squared error fraction (RMSF, **b**) for days with precipitation ≥3 mm
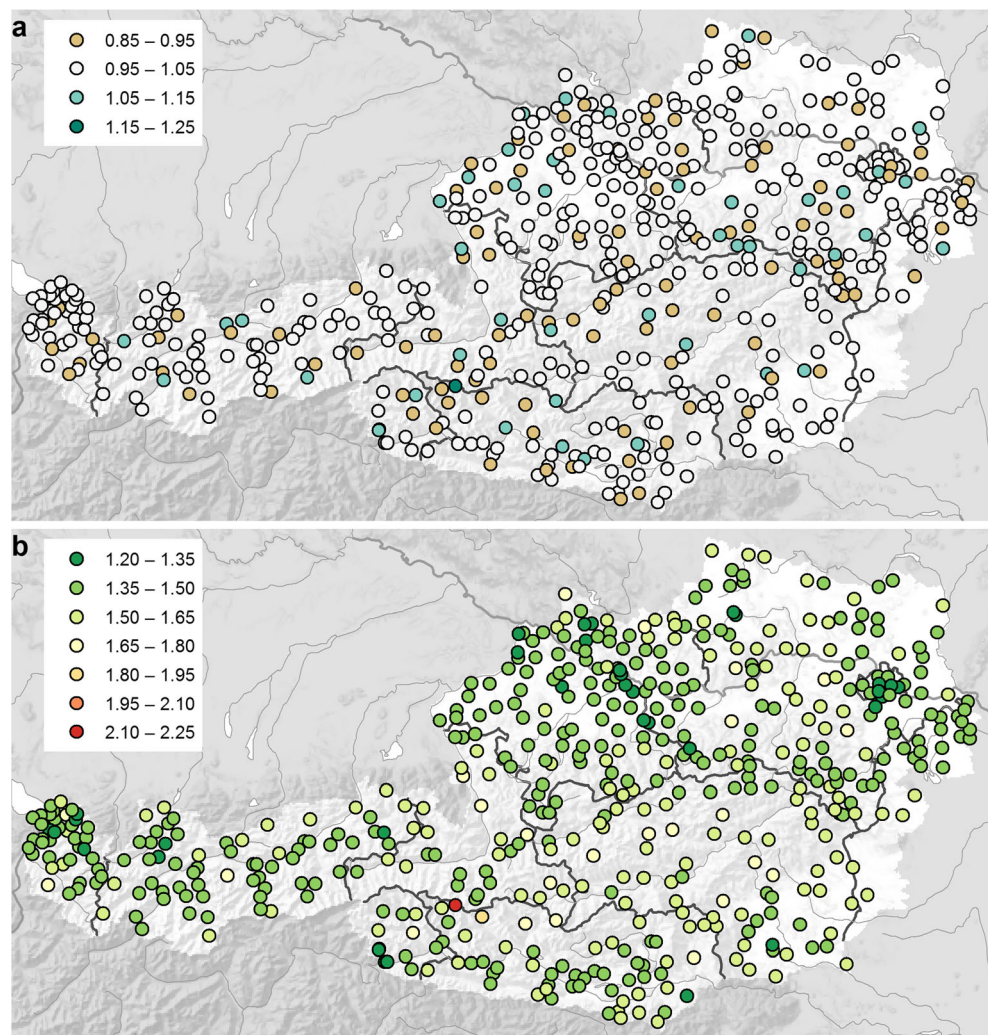
Table 4 Cross-validation statistics for the daily precipitation grid dataset

Altitude band	n	B					RMSF				
		DJF	MAM	JJA	SON	Year	DJF	MAM	JJA	SON	Year
<500 m	215	0.98	0.99	0.99	0.99	0.99	1.37	1.43	1.55	1.39	1.45
500–1000 m	207	0.97	0.98	0.99	0.99	0.98	1.42	1.45	1.55	1.42	1.48
1000–1500 m	88	0.98	0.97	0.99	0.98	0.98	1.49	1.47	1.56	1.46	1.50
>1500 m	13	0.98	1.00	1.01	1.03	1.01	1.61	1.58	1.56	1.57	1.58
All	523	0.98	0.98	0.99	0.99	0.99	1.42	1.45	1.55	1.42	1.48

Bias (B) and root mean squared error fraction (RMSF) constrained to precipitation events exceeding a threshold of 3 mm per day, stratified by altitude bands and seasons. n is the number of stations per altitude band. (See Sect. 5.2 for details)

increases from low to high altitudes and is larger in the convective compared to the other seasons.

A station-wise analysis of B and RMSF, again using the 3-mm threshold, is depicted in Fig. 7. As for the bias (upper panel), there is no obvious clustering of stations with positive or negative values. Values are smaller than 15% (i.e. $0.85 \leq B \leq 1.15$) for all, except one station (Sonnblick 1.17). For RMSF on the other hand (bottom panel), the pattern implies that the error is a function of regional station density and topographic complexity. For example, comparatively small error values (RMSF near 1) are found in flat regions with dense station spacing like the Vienna Basin, northern Burgenland and central Upper Austria. But also in the Klagenfurt Basin, the Tyrolean Oberland and the Bregenz Forest, the large number of stations reflects in relatively higher accuracy of the grid dataset. In contrast, larger errors are found along the northern border of Tyrol, in Salzburg and Styria as well as Upper Carinthia, where the station network is comparatively coarser. Again, station Sonnblick, the highest station in the network, exhibits the largest relative errors (RMSF = 2.18), much larger than any other station. The reasons for this are unclear, but error statistics for this station may be confounded by systematic measurement biases and relatively larger random measurement errors at this wind-exposed and snow-rich station. Clearly, the statistics do not seem to be representative for other mountain top stations.

To further evaluate the accuracy of the interpolation method in discriminating dry and wet conditions (i.e. the intermittence of precipitation), Table 5 lists two measures that quantify the level of mismatch. The *false alarm index* (FAI) is the fraction of mistakenly predicted wet conditions relative to the number of observed wet days. The *missing index* (MI) is the fraction of mistakenly predicted dry conditions relative to the number of observed wet days. Here, a threshold of 1 mm is chosen for the distinction. Table 5 stratifies the error measures by altitude band and season again. It shows that the degree of mismatch is in the order of 8 to 15% relative to the occurrence of wet days. Overprediction of wet conditions (FAI, typically 0.15) is larger than underprediction (MI, typically 0.08),

indicating that the precipitation grid dataset has a tendency for slightly too large rainfall regions. There are only small seasonal and altitudinal differences in these error scores; most notably, the rate of missing wet days is comparatively larger in winter and at high altitudes. One reason for a too large spatial extent of rainy areas is the smoothing inherent to this and many other interpolation methods. This tendency is related to the excessive frequency of small precipitation amounts seen in Fig. 6. It has been argued that the level of mismatch could be reduced by an explicit account of intermittence in interpolation schemes (e.g. Barancourt and Creutin 1992; Haylock et al. 2008). However, adequate interpretation of the figures in Table 5 should take into account that there are scale inconsistencies when comparing an interpolation (grid-pixel mean) against point measurements, which contribute to larger FAI than MI (see, e.g. Osborn and Hulme 1997). The numbers may therefore be considered as upper bounds of the true bias made by the interpolation.

Despite the efforts for high-quality and long-term stability of the input data, the quality of the final grid dataset may still be affected by residual inhomogeneities in the station data and density. Here, we assess this aspect by examining the variation of mean annual precipitation between two well-separated 5-year periods with anomalously dry (1982–1986) and wet (1998–2002) conditions in Austria. The difference serves as a simple test case of long-term variation. The test is undertaken for the subset of 44 homogenised Austrian station series. Differences for this subset are calculated both from the observations and from the predictions of the interpolation method leaving out the data of the test stations in turn. The correspondence between predicted and observed differences is quite good (Fig. 8). The average error of the predicted difference (5.2%) is clearly smaller than the average difference over all stations (about 15%). There is a respectable correlation ($R^2 = 0.54$) suggesting that the larger scale spatial pattern of this difference is reasonably reproduced by the prediction. At individual stations, though, errors may be large. The largest discrepancy is found for a mountain summit station (Villacher Alpe). The reasons for this outlier are not clear; other mountain stations are not similarly biased. We conclude from this

Table 5 Cross-validation statistics on the misclassification of dry and wet conditions by the interpolation method

Altitude band	n	FAI					MI				
		DJF	MAM	JJA	SON	Year	DJF	MAM	JJA	SON	Year
<500 m	215	0.16	0.16	0.17	0.14	0.16	0.09	0.08	0.07	0.07	0.08
500–1000 m	207	0.16	0.16	0.16	0.14	0.15	0.08	0.07	0.06	0.06	0.07
1000–1500 m	88	0.16	0.16	0.14	0.13	0.15	0.10	0.09	0.07	0.08	0.08
>1500 m	13	0.11	0.12	0.12	0.12	0.12	0.14	0.11	0.08	0.10	0.11
All	523	0.16	0.16	0.16	0.14	0.15	0.09	0.08	0.07	0.07	0.08

A threshold of 1 mm per day is employed for the definition of dry and wet days. Results are stratified by altitude bands and seasons. n is the number of stations per altitude band. (See Sect. 5.2 for details)

FAI false alarm index, *MI* missing index

analysis that pronounced long-term variations of precipitation should be reasonably represented by our grid dataset, in what regards the typical magnitude of the variation and its larger scale spatial pattern. Smaller scale detail may, however, be subject to residual inhomogeneities and sampling limitations by the station network. Care should, therefore, be exercised in interpreting small-scale features in maps of trends and other diagnostics of interdecadal variation.

5.3 Evaluation using WegenerNet high-resolution data

The Wegener Center for Climate and Global Change of the University of Graz operates an experimental climate observation network, the WegenerNet (Kirchengast et al. 2014) around the town of Feldbach in south-eastern Styria (Fig. 1). The network measures meteorological parameters at 151 evenly distributed sites with a typical spacing of 1.4 km, i.e. at a resolution much

Fig. 8 Relative difference in precipitation between the 5-year periods 1982–1986 and 1998–2002. Results are compared between observations at 44 homogenised high-quality stations across Austria (*x-axis*, reference set) and predictions using the interpolation method but leaving out data of the respective test station (*y-axis*, test set)

denser than current operational climate networks. In this section, precipitation observations from WegenerNet are used for evaluating the SPARTACUS precipitation grid dataset with respect to area mean precipitation.

In our evaluation, we use 1-day precipitation totals over a 9-year period (2007–2015), aggregated from the original measurements at 5-min intervals. Till 2013, the WegenerNet measurements were affected by heated sensor problems during snowfall, which meant that part of the data had to be excluded from the evaluation. Almost one third of the days (31.5%) in the period was flagged or missing. It is also relevant to note that the WegenerNet precipitation data is subject to systematic underestimation. Sungmin et al. (2016) estimate a systematic difference of 10 to 13% between the measurement devices of the WegenerNet and those used at conventional climate stations. This is supported by the observation that the space-time average over all WegenerNet stations is 12% smaller than the time mean of the only conventional climate station (Bad Gleichenberg) in the experimental domain. (The latter station was the only one used for the interpolation of the grids.) Therefore, we followed the recommendations of Sungmin et al. (2016) and applied constant correction factors for different WegenerNet stations and periods.

In the following, the difference between SPARTACUS grid data and WegenerNet observations are quantified for area mean precipitation over squares with side lengths of 1, 2, 4, 8 and 16 km. For this purpose, the station observations falling into the pertinent squares of the SPARTACUS grid were spatially averaged. At the 1-km scale, this comparison still involves the somewhat unsatisfactory scale mismatch between grid point estimates and point observations common to cross-validation. But at larger scales, the observational reference becomes more consistent with the space scale at which users commonly interpreted grid data. At the 8-km scale, for example, the observational area mean value is inferred by averaging over typically 29 WegenerNet stations, which, unless there is only a small fraction of the area with rain and except for the measurement biases, may be considered a fairly reliable estimate of the true area mean.
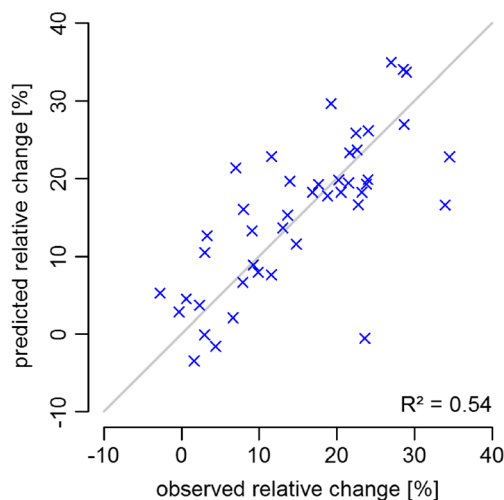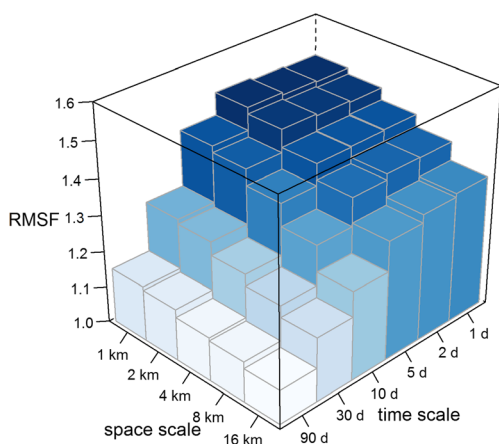
Fig. 9 Interpolation error (root mean squared error fraction (RMSF)) as inferred from the comparison of SPARTACUS grid dataset with observations at the experimental measurement network WegenerNet. RMSF is calculated for precipitation events with ≥3 mm and is displayed at different space scales (side length of a square in kilometers) and time scales (aggregate in days)

Figure 9 depicts the RMSF (see Eq. 2) for the area mean estimates. Results are shown as a function of space scale (side length of square) and for six different time scales (number of days). Only those days/areas were included in the calculation of RMSF when predicted and observed values were larger or equal to 3 mm per day. Clearly, RMSF values are becoming markedly smaller with increasing space and increasing time scale. At the 1-day time scale, the typical relative error (RMSF) decreases from 1.49 at the 1-km/point scale to 1.36 at the 16-km space scale, implying a 27% reduction in interpolation error. (Probably, this decrease is underestimated here because of the partly corrected systematic offset of the WegenerNet and conventional observations which is largely independent of scale.) A similar error reduction can be found at all time scales. Even more remarkable is the reduction in error towards longer time scales. At the 1-km resolution, RMSF decreases from 1.49 at the 1-day time scale to 1.15 at the 90-day time scale, implying an error reduction by 69%. All these results suggest that the error statistics inferred by cross-validation and at the 1-day timescale (depicted in Figs. 6 and 7, and listed in Tables 3 and 4) should be considered pessimistic estimates of the accuracy of the SPARTACUS dataset for applications that will utilise this dataset at larger space scales than single points and at larger time scales than single days.

6 Conclusions

Several precipitation grid datasets have been developed for the territory of Austria in the past. Limitations in long-term consistency, restrictions in the effective or nominal resolution and/or the lack of an operational and near real-time updating have restricted the utility of these existing datasets for many modern applications. In this study, we have presented a new dataset aimed at filling these gaps. The SPARTACUS grid dataset of daily precipitation has a grid spacing of 1 km; it covers the entire territory of Austria, extends back till 1961 and is regularly updated. A special feature of its development is a careful selection of the station data that are incorporated into the different production steps, to warrant, both, high spatial resolution and stable coverage of input data over time. The dataset is particularly devoted for users addressing questions of interannual to interdecadal variations and change in a number of disciplines (e.g. agriculture, hydrology, water resources, hydropower) and for the climate monitoring and operational public information services at ZAMG.

In terms of methodology, the dataset adopts the traditional two-tier analysis with separate interpolations for the mean monthly precipitation (background fields, kriging with external drift) and for the daily relative anomalies (angular distance weighting, SYMAP). Particularly noteworthy elements of the analysis scheme are the inclusion of totaliser data to fill in data gaps at high elevations, the utilisation of a multi-scale set of topographic predictors and a square root data transformation to improve compatibility with statistical assumptions. Example cases suggest that the new analysis is at least as plausible as previously existing datasets.

It is important that users of the new dataset are aware of limitations and uncertainties in the new dataset and that efforts are made in examining implications for the application at hand. Firstly, all measurements used are affected by a systematic bias, the gauge undercatch due to wind deflection and wetting that causes the spatial analyses to generally underestimate precipitation between a few percent in summer up to several 10% at wind-exposed and snow-rich sites at high elevation. Moreover, inhomogeneities may be present, despite stationarity of station network and efforts into data quality. As a result, patterns of long-term variations that are of small spatial scale should be interpreted with reservations as these may be from one single station only. Residual inhomogeneities are also likely to be found in high-frequency statistics (e.g. wet day frequency, threshold exceedance). Users interpreting the grid points of the dataset as point estimates must expect conditional biases and considerable random errors. The former manifests in an overestimation of light and moderate precipitation intensities and an underestimation of intense and heavy precipitation. The conditional bias is up to 20% but markedly depends on station density and season. In terms of random error, grid point estimates are typically within a factor of 1.5 from an in situ (point) observation, larger even in data sparse regions, at high elevations and in summer, smaller over flatlands and in autumn and winter. If grid point values are interpreted as area mean values instead of point estimates, the magnitude of errors is considerably smaller, e.g. some 25% smaller for an area mean estimate over a 16 × 16 km square.

Regarding longer term (interdecadal) variations, tests let us conclude that the present grid dataset does reproduce the typical magnitude and the larger scale spatial pattern of variations of notable amplitude. Care should, however, be exercised in interpreting small-scale features in maps of trends and other diagnostics of interdecadal variation. They may be compromised by residual inhomogeneities and limitations in representativity of the station network.

At present, the SPARTACUS precipitation dataset encompasses about 21,000 daily grids (since 1961). It is continuously updated in near real - time. At present, however, some of the input data (the component from the hydrographic services) is only available with a delay of several months, which implies that near real-time analyses will be preliminary and a final production can only be distributed after 5 months. In tandem with a similar grid dataset of temperature (Hiebl and Frei 2016), the SPARTACUS datasets constitute state-of-the-art and official daily spatial climate analyses for Austria.

References

Aalto J, Pirinen P, Heikkinen J, Venäläinen A (2012) Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. Theor Appl Climatol 112:99–111. doi:10.1007/s00704-012-0716-9

Auer I, Potzmann R, Schöner W (2000) Welchen Beitrag leisten Totalisatoren für die Klimaforschung im Hochgebirge? In: 96.–97. Jahresbericht des Sonnblick-Vereines für die Jahre 1998–1999. Sonnblick-Verein, Vienna, pp 22–30

Auer I, Nemec J, Gruber C, Chimani B, Türk K (2010) HOM-START–homogenisation of climate series on a daily basis, an application to the StartClim dataset. Climate and Energy Fund of the Federal State, Vienna. http://www.zamg.ac.at/cms/de/dokumente/klima/dok_projekte/homstart/homstart-endbericht. Accessed 17 Jan 2017

Barancourt C, Creutin JD (1992) A method for delineating and estimating rainfall fields. Water Resour Res 28:1133–1144. doi:10.1029/91WR02896

Becker A, Finger P, Meyer-Christoffer A, Rudolf B, Schamm K, Schneider U, Ziese M (2013) A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. Earth Syst Sci Data 5:71–99. doi:10.5194/essd-5-71-2013

Bénichou P, Le Breton O (1987) Prise en compte de la topographie pour la cartographie des champs pluviométrique. La Météorologie 7:23–34

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 57:289–300. doi:10.2307/2346101

Blöschl G, Nester T, Komma J, Parajka J, Perdigão RAP (2013) The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. Hydrol Earth Syst Sci 17:5197–5212. doi:10.5194/hess-17-5197-2013

Daly C, Neilson RP, Phillips DL (1994) A statistical-topographical model for mapping climatological precipitation over mountainous terrain. J Appl Meteorol 33:140–158. doi:10.1175/1520-0450(1994)033<0140:ASTFM>2.0.CO;2

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int J Climatol 28:2031–2064. doi:10.1002/joc.1688

Diggle PJ, Ribeiro PJ (2007) Model-based geostatistics. Springer, New York

Erdin R, Frei C, Künsch HR (2012) Data transformation and uncertainty in geostatistical combination of radar and rain gauges. J Hydrometeorol 13:1332–1346. doi:10.1175/JHM-D-11-096.1

Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf D (2007) The Shuttle Radar Topography Mission. Rev Geophys 45:RG2004. doi:10.1029/2005RG000183

Frei C (2014) Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. Int J Climatol 34:1585–1605. doi:10.1002/joc.3786

Frei C, Schär C (1998) A precipitation climatology of the Alps from high-resolution rain-gauge observations. Int J Climatol 18:873–900. doi:10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9

Frei C, Schär C (2001) Detection probability of trends in rare events: theory and application to heavy precipitations in the Alpine region. J Clim 14:1568–1584. doi:10.1175/1520-0442(2001)014<1568:DPOTIR>2.0.CO;2

Frei C, Willi M, Stöckli R, Dürr B (2015) Spatial analysis of sunshine duration in complex terrain by non-contemporaneous combination of station and satellite data. Int J Climatol 35:4471–4790. doi:10.1002/joc.4322

Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J Hydrol 228:113–129

Gottardi F, Obled C, Gailhard J, Paquet E (2012) Statistical reanalysis of precipitation fields based on ground network data and weather patterns: application over French mountains. J Hydrol 432–433:154–167. doi:10.1016/j.jhydrol.2012.02.014

Haiden T, Kann A, Wittmann C, Pistotnik G, Bica B, Gruber C (2011) The integrated nowcasting through comprehensive analysis (INCA) system and its validation over the Eastern Alpine Region. Weather Forecast 26:166–183. doi:10.1175/2010WAF2222451.1

Hamlet AF, Lettenmaier DP (2005) Production of temporally consistent gridded precipitation and temperature fields for the Continental United States. J Hydrometeorol 6:330–336. doi:10.1175/JHM420.1

Hasenauer H, Merganicova K, Petritsch R, Pietscha SA, Thornton PE (2003) Validating daily climate interpolations over complex terrain in Austria. Agric For Meteorol 119:87–107. doi:10.1016/S0168-1923(03)00114-X

Haslinger K, Schöner W, Anders I (2015) Future drought probabilities in the Greater Alpine Region based on COSMO-CLM experiments—spatial patterns and driving forces. Meteorol Z. doi:10.1127/metz/2015/0604

Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J Geophys Res 113:D20119. doi:10.1029/2008JD010201

Hengl R, Heuvelink GBM, Rossiter DG (2007) About regression-kriging: from equations to case studies. Comput Geosci 33:1301–1315. doi:10.1016/j.cageo.2007.05.001

Hiebl J, Frei C (2016) Daily temperature grids for Austria since 1961—concept, creation and applicability. Theor Appl Climatol 124:161–178. doi:10.1007/s00704-015-1411-4

Hofstätter M, Jacobeit J, Homann M, Lexer A, Chimani B, Philipp A, Beck C, Ganekind M (2015) WETRAX—weather patterns, cyclone tracks and related precipitation extremes. Großflächige Starkniederschläge im Klimawandel in Mitteleuropa. Universität Augsburg, Augsburg. http://www.zamg.ac.at/cms/de/dokumente/klima/dok_projekte/wetrax/endbericht. Accessed 17 Jan 2017

Hofstra N, Haylock M, New M, Jones P, Frei C (2008) The comparison of six methods for the interpolation of daily European climate data. J Geophys Res 113. doi:10.1029/2008JD010100

Hofstra N, New M, McSweeney C (2010) The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. Clim Dynam 35:841–858. doi:10.1007/s00382-009-0698-1

Hogan RJ, Mason IB (2012) Deterministic forecasts of binary events. In: Jolliffe IT, Stephenson BD (ed) Forecast verification. A practitioner's guide in atmospheric sciences, 2nd edn. Wiley, Oxford, pp 31–59

Huss M, Farinotti D, Bauder A, Funk M (2008) Modelling runoff from highly glacierized alpine drainage basins in a changing climate. Hydrol Process 22:3888–3902. doi:10.1002/hyp.7055

Hutchinson MF (1998) Interpolation of rainfall data with thin plate smoothing splines, part I: two dimensional smoothing of data with short range correlation. J Geogr Inf Decis Anal 2:139–151

Isotta FA, Frei C, Weilguni V, Perčec Tadić M, Lassègues P, Rudolf B, Pavan V, Cacciamani C, Antolini G, Ratto SM, Munari M, Micheletti S, Bonati V, Lussana C, Ronchi C, Panettieri E, Marigo G, Vertačnik G (2014) 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. doi:10.1002/joc.3794

Kapeller S, Lexer MJ, Geburek T, Hiebl J, Schueler S (2012) Intraspecific variation in climate response of Norway spruce in the eastern Alpine range: selecting appropriate provenances for future climate. Forest Ecol Manag 271:46–57. doi:10.1016/j.foreco.2012.01.039

Kendall MG (1948) Rank correlation methods. Charles Griffin, London

Kirchengast G, Kabas T, Leuprecht A, Bichler C, Truhetz H (2014) WegenerNet: a pioneering high-resolution network for monitoring weather and climate. Bull Amer Meteor Soc 95:227–242. doi:10.1175/BAMS-D-11-00161.1

Klein Tank AMG, Zwiers FW, Zhang X (2009) Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. World Meteorological Organization, Geneva. http://eca.knmi.nl/documents/WCDMP_72_TD_1500_en_1.pdf. Accessed 17 Jan 2017

Legates DR, Willmott CJ (1990) Mean seasonal and spatial variability in global surface air temperature. Theor Appl Climatol 41:11–21

Mann HB (1945) Nonparametric tests against trend. Econometrica 13:245–259. doi:10.2307/1907187

Masson D, Frei C (2014) Spatial analysis of precipitation in a high-mountain region: exploring methods with multi-scale topographic predictors and circulation types. Hydrol Earth Syst Sci 18:4543–4563. doi:10.5194/hess-18-4543-2014

McCullagh P, Nelder JA (1989) Generalized linear models. Monographs on Statistics and Applied Probability 37. Chapman and Hall, London

Mergili M, Kerschner H (2015) Gridded precipitation mapping in mountainous terrain combining GRASS and R. Nor Geogr Tidsskr 69:2–17. doi:10.1080/00291951.2014.992807

Olefs M, Schöner W, Suklitsch M, Wittmann C, Niedermoser B, Neururer A, Wurzer A (2013) SNOWGRID – A new operational snow cover model in Austria. International Snow Science Workshop

Proceedings 2013:38–45. http://arc.lib.montana.edu/snow-science/item/1785. Accessed 17 Jan 2017

Osborn TJ, Hulme M (1997) Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. J Clim 10:1885–1908. doi:10.1175/1520-0442(1997)010<1885:DOARBS>2.0.CO;2

Parajka J, Blaschke AP, Blöschl G, Haslinger K, Hepp G, Laaha G, Schöner W, Trautvetter H, Viglione A, Zessner M (2015) Uncertainty contributions to low flow projections in Austria. Hydrol Earth Syst Sci Discuss 12:12395–12431. doi:10.5194/hessd-12-12395-2015

Perčec Tadić M (2010) Gridded Croatian climatology for 1961–1990. Theor Appl Climatol 102:87–103. doi:10.1007/s00704-009-0237-3

Prudhomme C, Reed DW (1998) Relationships between extreme daily precipitation and topography in a mountainous region: a case study in Scotland. Int J Climatol 18:1439–1453. doi:10.1002/(SICI)1097-0088(19981115)18:13<1439::AID-JOC320>3.0.CO;2-7

Rauthe M, Steiner H, Ulf R, Mazurkiewicz A, Gratzki A (2013) A Central European precipitation climatology—part I: generation and validation of a high-resolution gridded daily data set (HYRAS). Meteorol Z 22:235–256. doi:10.1127/0941-2948/2013/0436

Richter D (1995) Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann-Niederschlagsmessers. Deutscher Wetterdienst, Offenbach

Schabenberger O, Gotway CA (2005) Statistical methods for spatial data analysis. CRC Press, Boca Raton

Scherrer SC, Frei C, Croci-Maspoli M, van Geijtenbeek D, Hotz C, Appenzeller C (2011) Operational quality control of daily precipitation using spatio-climatological plausibility testing. Meteorol Z 20:397–407. doi:10.1127/0941-2948/2011/0236

Schmidli J, Schmutz C, Frei C, Wanner H, Schär C (2002) Mesoscale precipitation variability in the region of the European Alps during the 20th century. Int J Climatol 22:1049–1074. doi:10.1002/joc.769

Schöner W, Dos Santos Cardoso E (2004) Datenbereitstellung, Entwicklung von Regionalisierungstools und einer Schnittstelle zu den regionalen Klimamodellen. Zentralanstalt für Meteorologie und Geodynamik, Vienna

Sen PK (1968) Estimation of the regression coefficient based on Kendall's tau. J Amer Stat Assoc 63:1379–1389

Sevruk B (1985) Systematischer Niederschlagmessfehler in der Schweiz. Der Niederschlag in der Schweiz, Beiträge zur Geologischen Karte der Schweiz-Hydrologie 31:65–75

Shepard DS (1984) Computer mapping: the SYMAP interpolation algorithm. In: Gaile GL, Willmott CJ (eds) Spatial statistics and models. Springer, Dordrecht, pp 133–145

Sungmin O, Foelsche U, Kirchengast G, Fuchsberger J (2016) Validation and correction of rainfall data from the WegenerNet high density network in southeast Austria. J Hydrol. doi:10.1016/j.jhydrol.2016.11.049

Theil H (1950) A rank-invariant method of linear and polynomial regression analysis. Parts 1–3. Nederlands Akad Wetensch Proc 53:386–392, 521–525, 1397–1412

Thornton PE, Running SW, White MA (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. J Hydrol 190:214–251

Van Bebber WJ (1891) Die Zugstrassen der barometrischen Minima nach den Bahnenkarten der Deutschen Seewarte für den Zeitraum 1875–1890. Meteorol Z 8:361–366

Wilks DS (2011) Statistical methods in the atmospheric sciences. Academic Press, Oxford

Wilks D (2016) "The stippling shows statistically significant gridpoints": how research results are routinely overstated and over-interpreted, and what to do about it. Bull Amer Meteor Soc 97:2263–2273. doi:10.1175/BAMS-D-15-00267.1

Xie PP, Rudolf B, Schneider U, Arkin PA (1996) Gauge-based monthly analysis of global land precipitation from 1971 to 1994. J Geophys Res 101:19023–19034. doi:10.1029/96JD01553