Meteorologische Zeitschrift, Vol. **24**, No. 1, 15–37 (published online January 13, 2015)

Evaluation of European regional reanalyses and downscalings for precipitation in the Alpine region

Francesco A. Isotta[1*], Raphaela Vogel[1,2] and Christoph Frei[1]

[1]Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland
[2]Now at Max Planck Institute for Meteorology, Hamburg, Germany

(Manuscript received January 31, 2014; in revised form August 28, 2014; accepted October 23, 2014)

Abstract

Datasets of the observed weather and climate on a regular grid are an important requisite for environmental and climate change research. In this paper we evaluate several European-scale regional reanalyses, developed in the EU project EURO4M, and compare them to existing datasets popularly used in applications today. The evaluation focuses on precipitation in the region of the European Alps, which, due to the marked topographic imprints, poses a challenging test bed. Utilizing a set of statistical indicators, we examine how the datasets represent the spatial pattern, annual cycle, frequency distribution and interannual variations. The evaluated datasets encompass new model-based regional reanalyses (UKMO and HIRLAM) and downscaling datasets (MESAN and MESCAN), one global reanalysis (ERA-Interim), and three station-based interpolation datasets (E-Obs, CRU, GPCC). The reference (APGD) is a gridded interpolation derived from very dense rain-gauge observations.

The evaluation provides insight into the relative strengths and limitations of the various datasets and construction concepts. The new model-based regional reanalyses show richer spatial variations in the climatology of daily precipitation compared to the driving global reanalysis (ERA-Interim), and they correct for several unrealistic spatial features seen in global and continental interpolation datasets. But they also show biases, shifts in regional anomalies and inaccuracies in mountain-valley contrasts. Common to both regional reanalyses are overestimates of mean precipitation and wet day frequency, and an underestimate of the frequency of heavy precipitation. The 4DVar UKMO reanalysis shows a better space-time coherency with APGD compared to the 3DVar HIRLAM reanalysis. The accuracy of datasets that explicitly use rain-gauge observations (interpolation and downscaling) is found to be very heterogeneous, depending on the density of available station time series. In regions of high station density, downscaling procedures effectively correct for biases in the underlying reanalysis and reduce RMSE and SEEPS errors by a factor of 2 to 5. In areas where E-Obs and the downscaling datasets have similar input of station data the skill of the two methodologies is comparable. Interannual variations of monthly mean precipitation are highly correlated (> 0.9) with those of the reference for all European-scale datasets and even for meso-beta scale subregions. Our evaluations illustrate that the new regional reanalyses provide a valuable data resource also in a region with complex topography, but many applications will have to involve bias correction and downscaling procedures based on direct observations, ideally at high spatial resolution.

Keywords: evaluation, regional reanalyses, grid dataset, downscaling, Alpine region, daily precipitation, mesoscale

1 Introduction

Datasets that provide spatially comprehensive information of the observed weather and climate at the earth's surface are a key input for modelling natural processes (e.g. Monestiez et al., 2001; Zappa, 2008; Machguth et al., 2009), for planning tasks concerned with natural resources and natural hazards (e.g. Greminger, 2003; Tveito et al., 2005; Weingartner et al., 2007; Holzkämper et al., 2012), for the monitoring of climate variations of the past (e.g. Brunetti et al., 2006; Auer et al., 2007; Schrier et al., 2013) and for the study of potential impacts of future climate change (e.g. Blenkinsop et al., 2008; Ott et al., 2013; Köplin et al., 2014; Bosshard et al., 2014).

There is an increasing resource of high-resolution datasets suitable for such applications. Many of these rely on the spatial interpolation of observations at weather stations onto a regular grid, eventually supplemented by satellite and radar data (e.g. Paulat et al., 2008; Erdin et al., 2012; Frei et al., submitted). Due to restrictions in national data policies, in Europe, many of these interpolation datasets are constrained onto national territories (e.g. Perry and Hollis, 2005; Dolinar, 2006; Perčec Tadić, 2010; Brunetti et al., 2012), or larger geographic entities from which data was collected for specific projects (e.g. Rubel and Hantel, 2001; Rauthe et al., 2013; Isotta et al., 2014). There exist datasets based on station data only that extend over the entire European continent or even the globe (Haylock

---

*Corresponding author: Francesco A. Isotta, Federal Office of Meteorology and Climatology MeteoSwiss, Operation Center 1, P.O. Box 257, 8058 Zurich-Airport, Switzerland, e-mail: francesco.isotta@alumni.ethz.ch

DOI 10.1127/metz/2014/0584

Gebrüder Borntraeger Science Publishers, Stuttgart, www.borntraeger-cramer.com

et al., 2008; Becker et al., 2013; Harris et al., 2014). Large improvements have been made in the volume of data accessible for these datasets (e.g. Klok and Klein Tank, 2008; Schneider et al., 2014), but the density of station series is partly still limited, and measuring conventions are heterogeneous. This compromises the effective resolution and regional consistency in these datasets, especially in areas of complex physiography (e.g. Hijmans et al., 2005; Hofstra et al., 2009). Also the highly variable density of stations prevents from using versatile interpolation methods, which exploit the observational data extensively where it is abundant.

Global reanalysis datasets such as those from the NOAA/NCEP (Saha et al., 2010), the ECMWF (Dee et al., 2011), NASA (Rienecker et al., 2011) and the JMA (Onogi et al., 2007) provide an alternative source of information on surface climate over decadal time scales. Data assimilation systems based on physico-dynamical models offer physically consistent information in space and time, sub-daily time resolution and a more comprehensive set of surface climate variables (including fluxes of energy and water that are not measured directly). Still, the spatial resolution of current global reanalyses limits their direct use in applications requiring high-resolution forcing data. Therefore, downscaling procedures have been pursued that enhance the resolution of global reanalyses regionally by either dynamical downscaling with regional climate models (e.g. Vidale et al., 2003; Suklitsch et al., 2008; Soares et al., 2012) or by combination with high-resolution surface observations through objective analysis (e.g. Jansson et al., 2007; Vidal et al., 2010).

More recently, techniques of regional model-based data assimilation are being pursued at climate time scales (e.g. Mesinger et al., 2006; Hahmann et al., 2010; Luhamaa et al., 2010). There is a high potential of these systems to represent mesoscale patterns and extreme events more accurately than the global reanalyses or downscaling procedures mentioned above. This is because advanced data assimilation techniques allow for the assimilation of observation types not integrated into global reanalyses, and because the much finer model resolutions allow better exploitation of observations where they are abundant.

Within the EU project EURO4M (European Reanalysis and Observations for Monitoring) the needs in high-resolution datasets for climate monitoring and environmental research were addressed specifically for the domain of Europe. For this purpose, the project has developed and brought together continental scale datasets from all the mentioned categories. In particular, existing climate datasets from interpolation of station data have been extended and improved, and new regional reanalyses were developed using regional model-based data assimilation.

The present study is part of EURO4M and aims at comparing and evaluating existing and new datasets from the above categories with regard to the representation of the daily precipitation climate in the region of the European Alps. Datasets of particular interest here are two new model-based mesoscale reanalyses based on, respectively, a 3DVar data assimilation with the HIRLAM model (Undén et al., 2002) and a 4DVar data assimilation with the UK Met Office North Atlantic and European model (Bush et al., 2006). Furthermore, two datasets will be examined that were derived by further downscaling the HIRLAM reanalysis with an objective analysis of high-resolution surface observations (MESAN and MESCAN, Häggmark et al., 2000). For comparison, our evaluation also includes the global reanalysis that was used to drive the regional reanalyses (ERA-Interim, Dee et al., 2011) as well as two global and one continental interpolation dataset (GPCC, Becker et al., 2013; CRU, Harris et al., 2014; E-Obs, Haylock et al., 2008).

The general purpose of this evaluation is to provide insight into the relative strengths and limitations of the various datasets in a region of complex topography. Specific consideration will be given to the meso-beta scale patterns and the statistics of daily precipitation, including the frequency of intense precipitation. These characteristics are essential for several applications of the new regional reanalyses and improvement over existing coarser resolution datasets may be expected particularly in these elements (e.g. Bukovsky and Karoly, 2007). The global datasets (ERA-Interim, GPCC, CRU) are included here primarily to illustrate the added value of the new reanalyses. Moreover, comparison of the downscaling procedures MESAN and MESCAN against the driving regional reanalysis (HIRLAM) will point to the added value of integrating additional surface observations in a posteriori downscaling. Interestingly, our study domain encompasses regions with many and regions with little additional observations used for downscaling, so that the benefit can be examined in its dependence on data density. Finally, our evaluation also examines the reproduction of intra-seasonal to interannual variations in order to assess the suitability of the datasets for climate monitoring. The various datasets span a range of spatial resolutions, from 50–80 km for the global datasets down to 5 km for the downscaling datasets. The present evaluation is conducted at three distinct resolutions (5, 22 and 55 kilometers), in order to ensure scale consistency and to portray the performance of the datasets over a range of spatial scales.

Our study region extends over a $1200 \times 700\,\mathrm{km}^2$ domain (2–17.5 ° E, 43–49 ° N) and comprises the entire mountain range of the European Alps, adjacent flatland and several smaller-scale hill ranges (see Figure 1). This is a challenging test area for regional reanalyses. Its precipitation climate is influenced by mid-latitude disturbances, local convection, the moisture source of the Mediterranean and several orographic dynamical phenomena (Schär et al., 1997). As a result, the region exhibits strong topographic imprints on precipitation and pronounced regional variations in the seasonal cycle and the distribution of heavy precipitation (see e.g. Isotta et al., 2014). Even though the present evaluation covers

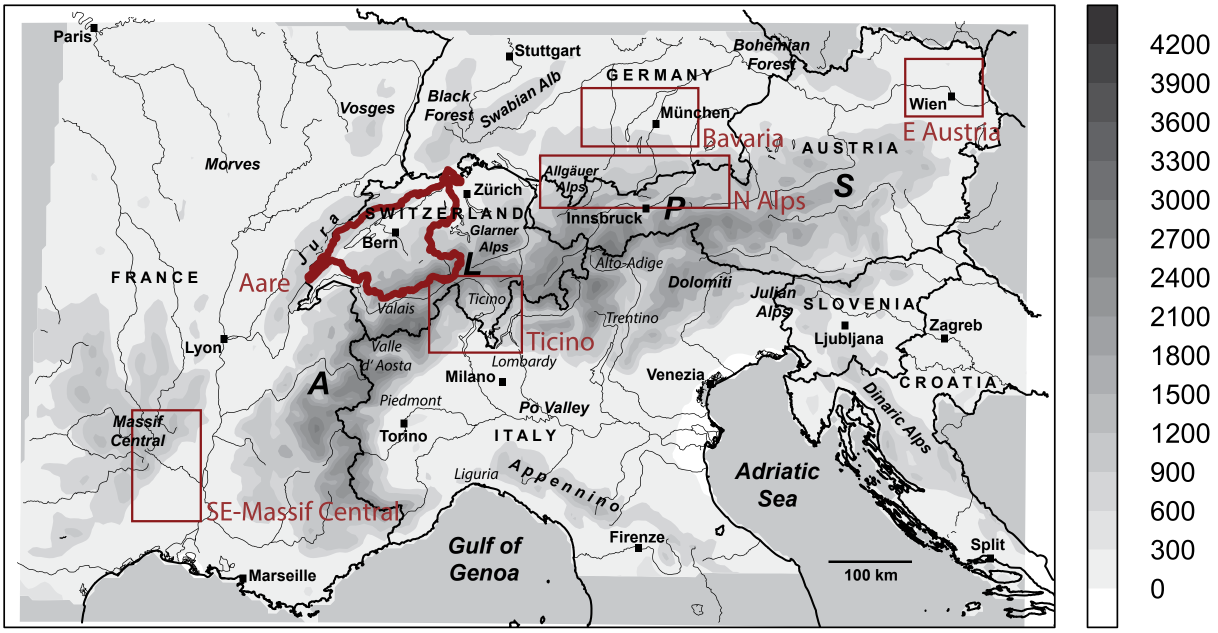

Figure 1: Map of the study domain with topographic height (grey-shading, in meters above sea level), the subregions used in our analyses (frames and Aare catchment) and geographic names used in the text. The domain spans over 1200 km in the east-west direction (2–17.5 ° E) and over 700 km in the north-south direction (43–49 ° N).

only a small portion of the area, for which the evaluated datasets have been constructed (Europe), the results may give valuable insights into their potential and limitations, considering that complex physiography and short-range climate variations are met in many other parts of the continent.

Compared to other high-mountain areas, the Alpine region is one with particularly dense in-situ rain-gauge observations. This data resource is distributed between many different national and regional services. While only a limited portion of it was available for the construction of the evaluated datasets, this resource is extensively exploited in the Alpine Precipitation Grid Dataset (APGD, Isotta et al., 2014), the reference dataset used in this evaluation. APGD is a spatial interpolation derived from all available high-resolution networks of the Alpine countries, incorporates 5500 quality-controlled rain-gauge measurements on average per day and is constructed on a grid with 5 km spacing.

Both the Alpine region and earlier versions of APGD have been considered in several previous evaluation studies with observation datasets and numerical models. These show, for example, that many of the large-scale interpolation datasets lack mesoscale variations even at scales nominally resolved by the grid spacing (Rubel and Rudolf, 2001; Hofstra et al., 2009). Also, systematic underestimates in the frequency of intense precipitation were observed as a result of the smoothing inherent to interpolation with coarse station data (Hofstra et al., 2009; Turco et al., 2013). It seems that station density is the main factor for interpolation quality in this region, more decisive even than differences in the interpolation method (see also Hofstra et al., 2008). On the other hand, the evaluation of weather forecasting models and regional climate models in the Alps have also pointed to biases in the representation of the Alpine precipitation climate by mesoscale models (e.g. Cherubini et al., 2002; Suklitsch et al., 2008; Smiatek et al., 2009). These biases can vary between seasons and are connected to errors in the frequency distribution (e.g. Frei et al., 2003). It is possible that such model biases compromise the quality of a model-based reanalysis, unless they are corrected through data assimilation (e.g. Bukovsky and Karoly, 2007). It will be interesting to learn how these challenges affect the accuracy of current large-scale observation datasets in a direct comparison of both categories, interpolation datasets and regional reanalyses.

The outline of the paper is as follows. The evaluated datasets and the reference dataset are introduced in sections 2 and 3 respectively. Section 4 details the evaluation procedure. Section 5 describes and discusses the results in subsections devoted to the different aspects of the evaluation. Finally, the results are summarized in Section 6.

2 Evaluated datasets

The present evaluation examines eight different datasets from three different categories. Three datasets (GPCC, CRU and E-Obs) were generated by statistical interpolation of in-situ rain-gauge observations and will hereafter be denoted simply as *interpolation datasets*. The station data available over the Alpine region varies

Table 1: Overview of datasets.

	Dataset	Resolution		Period used	Method	References
		spatial (approx.)	temporal			
Interpolation dataset	APGD	5 km	daily	1989–2003, 2008	PRISM for means, modified SYMAP for anomalies	Isotta et al., 2014
	E-Obs	25 km	daily	1989–2003, 2008	thin-plate splines for means, kriging for anomalies	Haylock et al., 2008; Klok and Klein Tank, 2008
	GPCC	55 km	monthly	1989–2003	SPHEREMAP	Becker et al., 2013; Schneider et al., 2014
	CRU	55 km	monthly	1989–2003	Thin-plate splines for means, triangulated linear interpolation for anomalies	Harris et al., 2014
Model-based reanalysis	HIRLAM	22 km	daily	1989–2003, 2008	3DVAR data assimilation	Undén et al., 2002
	UKMO	12 km	daily	2008	4DVAR data assimilation	Renshaw et al., 2013
	ERA-INT	75 km	daily	1989–2003, 2008	4DVAR data assimilation	Dee et al., 2011
Downscaling	MESAN	5 km	daily	1989–2003, 2008	Optimal interpolation	Häggmark et al., 2000; Jansson et al., 2007
	MESCAN	5 km	daily	2008	Optimal interpolation	Häggmark et al., 2000; Taillefer, 2002

considerably between them, with CRU building on coarser coverage compared to GPCC and E-Obs. The second category encompasses three datasets produced by model-based data assimilation systems, one of which covers the globe (ERA-Interim) and two that are regional and cover Europe (HIRLAM and UKMO). These will be denoted as *global and regional reanalyses* in the following. The regional reanalyses are both driven by boundary conditions from ERA-Interim. As a third group we consider two datasets (MESAN and MESCAN) that were produced by downscaling a regional reanalysis and objective assimilation of additional rain-gauge observations. The HIRLAM reanalysis served as the common basis for these downscaling datasets. MESAN and MESCAN are methodologically very similar and use station data from the ECA&D archive (Klok and Klein Tank, 2008), which also serves as input for E-Obs. Over France and some other countries (outside the Alps), MESAN and MESCAN also use additional high-resolution rain-gauge data.

All the evaluated datasets extend over the European continent and North Africa or over the entire globe. Table 1 gives an overview of the main characteristics of the datasets, including spatial and temporal resolutions. Additional detail and references on each of the datasets is provided in the following subsections. For the present study, we simply use the section of these datasets over the Alpine region and for the period 1971–2008, as far as available.

2.1 Interpolation datasets

2.1.1 E-Obs

The E-Obs dataset is a daily European land-only gridded dataset (Haylock et al., 2008) that was established in the ENSEMBLES project and was extended in EURO4M. Here we use version 7 on the 0.22 ° rotated pole grid, i.e. with a grid spacing of about 25 km over the Alps. E-Obs covers several surface climate variables, extends over the period 1950–2012 and is widely used for climate model evaluation and climate monitoring at European scale (e.g. Lenderink, 2010; Nikulin et al., 2011; Schrier et al., 2013). E-Obs builds on quality-controlled rain-gauge data of an updated and extended version of the European Climate Assessment Dataset ECA&D (Klok and Klein Tank, 2008). The precipitation data utilized is not corrected for measurement bias (Neff, 1977). The interpolation was done in a first step for the monthly precipitation totals with a three-dimensional thin-plate splines algorithm (Hutchinson, 1995), then for the daily proportions by a combination of indicator and ordinary kriging (Haylock et al., 2008). Hofstra et al. (2009) find E-Obs to have larger random errors and biases in the region of the Alps than elsewhere, likely due to the sparse station database available (only about 350 rain-gauges) and larger precipitation variability in this region.

2.1.2 GPCC

The Global Precipitation Climatology Centre (GPCC, http://gpcc.dwd.de), operated by the German Meteorological Service (DWD), has established and regularly updates a suite of gridded precipitation datasets at monthly time resolution over global land areas (Becker et al., 2013). In this study, we use the Full Data Reanalysis V6, provided on a regular longitude-latitude grid with a spacing of 0.5 ° degrees, i.e. approximately 55 km. The dataset covers the full study period (1971–2008) considered here. Data sources for the GPCC dataset are various GTS and CLIMATE archives and additional data supplied bilaterally by national meteorological services (Schneider et al., 2014). A semi-automatic and source-specific quality control system is applied to the data before analysis (Schneider et al., 2014). To compensate for systematic measurement errors, GPCC has employed an adjustment to the interpolated monthly fields (Schneider et al., 2014). Data density over the Alpine region is comparable to that of E-Obs, except for Germany, where GPCC has access to the high-resolution rain-gauge data of DWD. The spatial interpolation for GPCC is accomplished with a spherical adaptation of Shepard's angular distance weighting scheme (Shepard, 1984; Willmott et al., 1985).

2.1.3 CRU

This monthly gridded global land-only dataset constructed by the Climatic Research Unit (CRU, http://www.cru.uea.ac.uk) at the University of East Anglia aims at providing information on long-term climate variations at continental space scales since 1901 (Mitchell and Jones, 2005; Harris et al., 2014). In this study we use version CRU TS 3.20 for the period 1971–2008. The dataset is provided on a 0.5 ° longitude-latitude grid (approximately 55 km grid spacing). The spatial interpolation builds on a climatology (1961–1990) derived by three-dimensional thin-plate splines (Hutchinson, 1995; New et al., 1999) and a triangular linear interpolation of monthly anomalies (Harris et al., 2014). The number of stations integrated in CRU is smaller than in GPCC (see also Schneider et al., 2014). For Europe as a whole there are between about 1200 and 500 station anomalies at the beginning and end of our study period respectively (Harris et al., 2014) but the background climatology was estimated with a denser coverage (New et al., 1999). The CRU precipitation grids were calculated without correction for measurement bias.

2.2 Reanalysis datasets

2.2.1 HIRLAM

The HIRLAM regional reanalysis is a model-based data assimilation produced by the Swedish Meteorological and Hydrological Institute using the hydrostatic limited-area model HIRLAM (Undén et al., 2002). For the purpose of this reanalysis the model is configured on a rotated-pole grid with 0.2 ° spacing (approximately 22 km) and with 60 vertical levels over the European continent. It is driven with lateral boundary conditions from the global reanalysis ERA-Interim (see below). The reanalysis adopts a three-dimensional variational data assimilation (3DVAR) in a six-hourly cycle (Gustafsson et al., 2001; Lindskog et al., 2001) with observations taken from the ECMWF MARS archive. The large-scale dynamics in the model interior are constrained to that of the driving global reanalysis by penalizing for large-scale vorticity differences (Dahlgren and Gustafsson, 2012). For this study, the six-hourly output was aggregated to daily precipitation totals consistent with the daily observation times in the Alpine region.

2.2.2 UKMO

The regional reanalysis of the UK Met Office was newly developed in EURO4M, and data was only available from one year (2008) for this study (Renshaw et al., 2013). UKMO is a four-dimensional variational data assimilation (4DVAR) with the Met Office Unified Model (Bush et al., 2006), run at 0.11 ° resolution (approx. 12 km) and with 70 vertical levels, driven by ERA-Interim, in six-hourly cycles. The 4DVAR assimilation is run at 24 km. The model domain extends over all of Europe including the Mediterranean and parts of North Africa. In addition to observations assimilated into ERA-Interim (ECWMF MARS archive), the 4DVAR system also assimilates visibility and cloud fraction (Rawlins et al., 2007).

2.2.3 ERA-Interim (ERA-INT)

The global reanalysis ERA-Interim of the European Centre for Medium-Range Weather Forecasts (ECMWF, www.ecmwf.int) is documented in Dee et al. (2011). It has a resolution of approximately 75 km (0.7 °) and, hence, is the nominally coarsest of all datasets considered. ERA-Interim is based on a 12-hourly 4DVAR cycle. The dataset extends over the full period considered in this study. Clearly, we are not expecting ERA-Interim to reproduce meso-beta scale climate patterns in the Alpine region. It is still interesting to consider it here for illustrating the added value of higher resolution datasets. Evaluations of ERA-Interim (and its predecessor ERA-40) for precipitation and the hydrological cycle at larger space scales have been conducted, for example, by Zolina et al. (2004) and Lorenz and Kunstmann (2012).

2.3 Downscaling datasets

2.3.1 MESAN

MESAN is an operational mesoscale analysis system of selected meteorological parameters developed at the Swedish Meteorological and Hydrological Institute (Häggmark et al., 2000; Jansson et al., 2007). It assimilates additional rain-gauge observations by optimal interpolation (OI, e.g. Daley, 1991; Kalnay, 2003), using, as background, a downscaled version of the HIRLAM reanalysis on a 5 km grid and isotropic structure functions.

The analysis is performed at a three-hourly, six-hourly or daily time scale, depending on the type of observation, namely conventional rain-gauge observations as well as remote sensing data from radar and satellite. Daily precipitation totals were taken from the ECA&D archive (Klok and Klein Tank, 2008, the same as for E-Obs) and from the high-resolution rain-gauge networks of France and Sweden. As a result, the observational input into MESAN was very heterogeneous over the Alps, with much higher station densities over France, Germany and Slovenia than elsewhere. For the present study, data from MESAN was available for the period 1989–2003 and for the year 2008.

2.3.2 MESCAN

MESCAN is the name given to the developments of the Météo-France surface analysis system CANARI (Taillefer, 2002) during the EURO4M project. Like MESAN, MESCAN is based on the OI algorithm and both systems use the same observational datasets and downscaled HIRLAM reanalysis, but with different features of the host system, such as a different model orography, the usage of the observational data base (Saarinen, 2004), the selection of the observations and the spatial quality control. Moreover it has a modified statistical model with, among others, a different background error correlation function (structural function, implemented as in the Canadian precipitation analysis Project, Mahfouf et al., 2007) and shorter correlation length compared to MESAN. For the present study, MESCAN data was available for one year (2008) only.

3 Reference dataset

The dataset used as reference in this evaluation is itself an interpolation dataset, derived by statistical interpolation of rain-gauge observations. In comparison to the evaluated datasets, the Alpine Precipitation Grid Dataset APGD (Isotta et al., 2014) is based on much denser in-situ data. It integrates data from all available national and regional high-resolution networks of the Alpine region (i.e. the domain depicted in Fig. 1), with more than 8500 time series in total and around 5500 observations on average per day over the period 1971–2008. The rain-gauge density varies between 0.4 and 1.4 stations per 100 km^2, with denser coverage over Austria, Germany, Slovenia and Switzerland and coarser coverage over Croatia and Italy (see Fig. 3 of Isotta et al., 2014). A large proportion of the stations dispose of long records.

APGD is a gridded analysis of Alpine daily precipitation on a 5-km grid (see Table 1). The spatial analysis was produced with a combination of PRISM (Daly et al., 1994 and 2002) for the long-term monthly mean (see also Schwarb et al., 2001) and a modified version of SYMAP (Shepard, 1984; see also Frei and Schär, 1998) for the relative daily deviations from climatology. All observations underwent extensive quality testing, manual and automatic, checking for coding problems and spatial consistency. APGD is an updated, improved and higher resolution version of a previous Alpine precipitation dataset (Frei and Schär, 1998) that was used in various earlier evaluation studies (e.g. Frei et al., 2003; Hofstra et al., 2008; Suklitsch et al., 2008).

The APGD is used here as reference because of the amount of observational information integrated. But, like the other datasets, it is affected by errors and uncertainties (for a detailed discussion see Section 4 of Isotta et al., 2014). In particular, the following error sources shall be born in mind when interpreting the results of this evaluation: The rain-gauge data used here are subject to systematic errors due to the measurement undercatch (Neff, 1977; Yang et al., 1999). This is not corrected for and it likely results in an underestimate of true precipitation by APGD. Sevruk (1985) and Richter (1995) estimate the systematic measurement error in the Alps to range from about 7 % (5 %) over the flatland regions in winter (summer) to 30 % (10 %) above 1500 m MSL in winter (summer). Note that all other interpolation and downscaling datasets considered in this study, except GPCC, are affected by measurement biases too. The interpolation method used in APGD is designed to estimate an area average precipitation value in the neighbourhood of the grid points. In general this estimation also involves measurements outside the pixels of the underlying grid and, hence, the spatial precipitation fields of APGD represent area averages of a larger domain than the nominal 5-km grid pixels. The fine grid spacing is of advantage for climatological mean fields where topographic effects on the distribution are eminent and can be estimated using physiographic predictors (Schwarb, 2000). But for daily precipitation fields the effective resolution of APGD is coarser than the grid spacing, in the order of 10–20 km (typical inter-station distance). It can be even larger in areas with coarser networks such as in Northern Italy (Isotta et al., 2014). As a result, there is a tendency for APGD to overestimate low quantiles of daily precipitation (and the frequency of wet days), and to underestimate high quantiles (and the frequency of intense precipitation), compared to an analysis with a true 5-km resolution. The magnitude of bias from this resolution mismatch is difficult to estimate, but an upper limit can be estimated when grid-point estimates are compared against independent point measurements. Isotta et al. (2014) conducted such a cross-validation and found that biases at high quantiles are about 8 % in data dense regions and about 20 % in a region with very coarse station coverage. At low quantiles the bias is small (a few percent) and only evident in summer. Clearly, these numbers are upper limits for the 5-km resolution and the effect is likely negligible when aggregating APGD up to a resolution of 22 km or larger.

4 Evaluation procedure

4.1 Parameters, scores and domains

The main focus of the present evaluation is on statistics of daily precipitation and their meso-beta scale patterns. To this end we have calculated and compared a large number of the precipitation indices proposed by the ETCCDI (Klein Tank et al., 2009). For simplicity we present and discuss a selection only. The choice is made to convey the most characteristic features of and differences between the datasets.

The statistics to be presented include mean annual precipitation, the frequency of wet days and the 95 % quantile of daily precipitation. A threshold of 1 mm was chosen to separate dry and wet days. Furthermore, a large number of empirical quantiles were estimated to describe the frequency distribution of daily precipitation. For this purpose the daily grid-point values within pre-defined sub-domains were pooled. Finally, we compare time series of monthly precipitation anomalies from the long-term mean in order to investigate the reproduction of intra-seasonal to inter-annual variations, and, hence, to assess the reliability of the datasets for climate monitoring. Some elements of this evaluation will be undertaken with a subset of the datasets only, because of limited time resolution or temporal extent in some of the datasets.

Where meaningful, we summarize the correspondence between evaluated datasets and reference by quantitative scores. For this purpose we choose the conventional root-mean-square error (RMSE) and the Stable Equitable Error in Probability Space (SEEPS, Rodwell et al., 2010). SEEPS is a three-category score that values the distinction between dry, light and intense precipitation. Compared to RMSE, the categorical nature of SEEPS places less emphasis on quantitative correspondence and avoids that the score is dominated by intense precipitation. In SEEPS, light and intense precipitation events are defined respectively as the lower two-thirds and the upper third of all wet days. The threshold (quantile) is defined individually for evaluated and reference data and, hence, SEEPS is not sensitive to biases in the frequency distribution of the evaluated datasets. Just as for the RMSE, a small value of SEEPS means better correspondence.

In some parts of our evaluation, the comparison between evaluated data and reference is made grid-point by grid-point (at comparable resolutions, see below). This brings a risk that higher resolution datasets are excessively penalized for small misplacements in their precipitation patterns (e.g. Ebert, 2008). Several of our analyses will therefore be conducted either at a coarser than the nominal resolution of the evaluated datasets or over climatologically and topographically distinct subdomains of the Alps. The subdomains are depicted in Figure 1 and their climatological characteristics are discussed in Isotta et al. (2014). One of the domains was defined as a medium size river catchment (Aare without Reuss and Limmat, Switzerland, 11'900 km^2). It represents conditions with both high elevations and plains, typical when considering catchment mean precipitation.

4.2 Evaluation at different spatial resolutions

Statistics of daily precipitation are sensitive to spatial resolution (e.g. Mearns et al., 1995; Osborn and Hulme, 1997; Tustison et al. 2001; Frei et al., 2003). To avoid inconsistencies we compare datasets with different original resolutions only after rescaling them to the same resolution and onto a common grid. Three different resolutions will be considered, at 0.5 ° (~ 55 km), 0.2 ° (~ 22 km) and 0.05 ° (~ 5 km), using the grid structures of GPCC (longitude-latitude grid), HIRLAM and MESAN (rotated pole grids), respectively, as common grids. In most cases, the rescaling is done from higher to coarser resolution. This upscaling is carried out with the original daily fields by averaging, day-by-day, over all grid points of the fine resolution contained in a grid pixel of the coarser resolution grid. In some cases, it will be interesting to compare datasets at fine resolution, when a coarse resolution dataset is crudely downscaled. This was done with a simple nearest neighbor interpolation, which was also adopted to re-grid the APGD dataset (originally developed on an ETRS kilometer grid) onto the 5 km rotated pole grid used for dataset comparisons (i.e. the MESAN grid coordinate system).

The daily precipitation statistics are calculated finally from the rescaled datasets on the standard grids. As a result, fields of the wet-day frequency, for example, are available on three different grids for the high-resolution datasets MESAN, MESCAN and APGD, and an informative evaluation of the wet-day frequency in HIRLAM (22 km) is made by comparison with the frequency determined from the upscaled APGD.

We have briefly tested the sensitivity of the correspondence scores (RMSE and SEEPS) upon the regridding of the daily fields from MESCAN and APGD onto the MESAN grid, used as the standard at high resolution. To this end we calculated the scores in three different versions, using, in turn, the three native grid structures as common grid. The results for RMSE are shown in Table 2. Estimates of RMSE differ typically by 5 % or less between the different grids. The largest differences are found in regions SE-Massif Central (SEMC) and Ticino, regions with particularly strong precipitation gradients (see later). Nevertheless, the uncertainties in RMSE arising from the re-gridding seem to be clearly smaller than the variations of RMSE between regions and between datasets. Similar results were found for SEEPS. We conclude that the re-gridding does not seriously affect the results of our evaluation.

Table 2: Root mean square error of daily precipitation (mm per day) for MESAN (upper rows) and MESCAN (bottom rows) when compared to APGD (year 2008 only, 5 km resolution). Results are averages over sub-regions (see Figure 1) and were calculated when using the APGD, MESAN and MESCAN grids as a common standard respectively. SEMC: SE-Massif Central.

		Region				
Dataset	Grid	Alps	SEMC	Ticino	Bavaria	Aare
MESAN	APGD	4.00	4.66	8.27	1.74	3.54
	MESAN	4.00	4.46	8.14	1.73	3.58
	MESCAN	4.00	4.97	8.35	1.76	3.55
MESCAN	APGD	3.93	2.67	6.94	1.58	3.29
	MESAN	3.94	2.59	6.94	1.59	3.32
	MESCAN	3.89	2.72	6.98	1.62	3.30

5 Results

This section presents and discusses results of our evaluation both by visual comparison and quantitative analysis. We begin by comparing mean precipitation, its annual cycle, and selected statistics of daily precipitation (wet-day frequency, 95 % quantile, Section 5.1), proceed to the distribution functions of daily precipitation (Section 5.2) and compare results for a particular case of heavy precipitation (Section 5.3). Section 5.4 discusses scores of the correspondence between the datasets for daily precipitation. Finally, the ability of the datasets to reproduce the intra-seasonal to inter-annual variations is examined in Section 5.5. Our description makes use of geographical notions for which the reader is referred to Figure 1.

The APGD dataset will serve as reference but in our interpretations we take into account that this dataset itself has errors and uncertainties (see previously Section 3). Where meaningful, results will be discussed for several resolutions (out of 5 km, 22 km and 50 km) and always ensuring consistency in spatial resolution between evaluated and reference datasets (see Section 4).

The time period considered is partly constrained by the limited temporal coverage in some of the evaluated datasets. Therefore, apart from a long reference period (15 years 1989–2003), some of our analyses will be conducted for a single year only (2008), for which all datasets are available. Tests revealed that the pattern of the considered statistical parameters, as well as the characteristic differences between the datasets are fairly robust between the single-year and multi-year periods. Despite the limited temporal coverage in some of our analyses, the results can be considered representative of the climatological behaviour of the datasets.

5.1 Mean precipitation and daily precipitation statistics

5.1.1 Mean precipitation

Figure 2 depicts mean annual precipitation at all three resolutions. The results provide a first overview of all datasets considered in this evaluation. The reference period differs between the left-hand column (1989–2003) and the remaining columns (2008), to ensure maximum comparability for datasets with limited temporal coverage. The reference dataset (APGD) is shown at all three resolutions (first row). Not surprisingly, the two versions at coarser resolution reproduce smooth versions of the distribution with much smaller peak values and less prominent mesoscale patterns, notably for the dry inner-Alpine conditions, the wet anomalies along the southern Alpine rim and the anomalies associated with the smaller-scale hill ranges (Black Forest, Vosges).

The three global datasets (ERA-INT, GPCC and CRU) reveal quite different representations of annual mean precipitation in the Alps. Patterns range from a smooth moist anomaly (ERA-INT) over a pronounced regional moist maximum (CRU) to a structured distribution with mesoscale anomalies (GPCC). GPCC reproduces the characteristic distribution of APGD most accurately, with signatures of the moist anomaly extending along the northern rim of the Alps, two regional anomalies along the southern rim, dryness in the interior of the ridge and patterns associated with the Massif Central and the small-scale hill ranges (Vosges, Black Forest). Some discrepancies to APGD are found in the Apennine mountains, south-eastern France and for the spatial extent of the inner-Alpine dry anomaly (Trentino in Northern Italy). ERA-INT and CRU miss out many of these mesoscale patterns. While ERA-INT tends to underestimate domain mean precipitation, CRU exaggerates the magnitude of the Alpine precipitation anomaly. It is to be noted that the resolution of ERA-INT (75 km) is coarser than the 55-km reference considered here (panel a) and that the topography of the underlying global model may be even coarser, which may explain some of the discrepancies. As regards CRU, a possible explanation is that the three-dimensional spline interpolation used for the background climatology overestimates the height dependence of precipitation and produced maximum precipitation in the region of maximum elevation (Hutchinson, 1995; New et al., 1999).

The European datasets at 22-km resolution (Figure 2, middle column) reveal more pronounced spatial variations and much larger peak values, compared to the global datasets. The regional reanalyses (HIRLAM and UKMO) clearly capture the characteristic mesoscale pattern, with moist anomalies along the rims, separated by dryer conditions in the interior. Also, anomalies associated with details of the topography can be well discerned and UKMO reproduces the location of the two prominent moist anomalies (Ticino and Julian Alps) particularly well. However, both reanalyses tend to overestimate mean precipitation. HIRLAM also overestimates precipitation contrasts (see e.g. Jura, Vosges, Black Forest), whereas UKMO has a more widespread overestimate also over flatland (see e.g. Po valley, Southern Germany). These overestimates seem to be larger than the expected underestimate in APGD due to measurement bias (see Section 3). Visually there is an impression

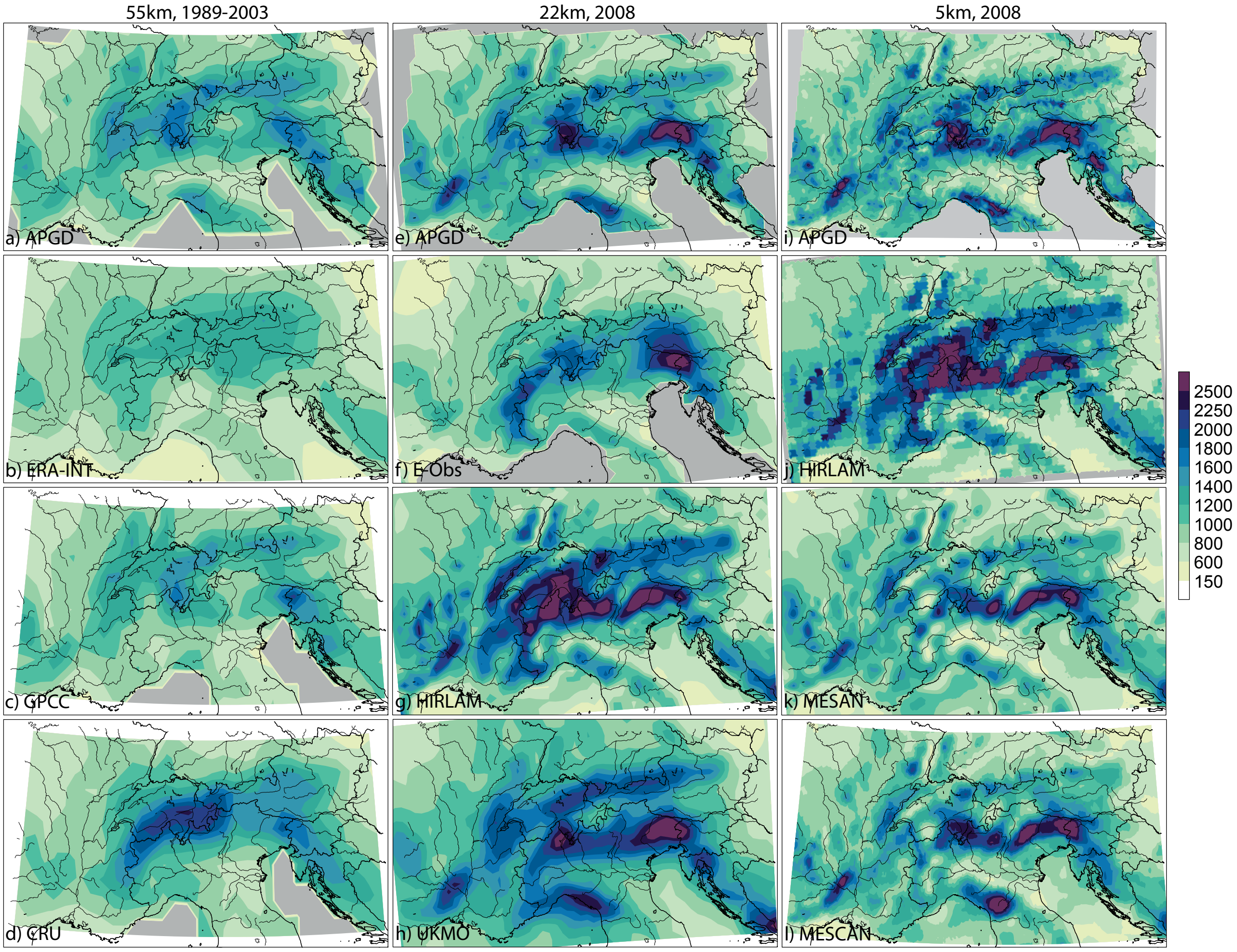

Figure 2: Mean annual precipitation (mm per year). Left column: 55-km resolution for the period 1989–2003. a) APGD; b) ERA-INT; c) GPCC; d) CRU. Middle column: 22-km resolution for the year 2008. e) APGD; f) E-Obs; g) HIRLAM; h) UKMO. Right column: 5-km grid resolution for the year 2008. i) APGD; j) HIRLAM; k) MESAN; l) MESCAN.

that UKMO is considerably smoother than APGD (and HIRLAM) at the 22-km resolution. This is not merely an artefact of the rescaling. The distribution on the native UKMO grid (12 km resolution, not shown) leaves a similar impression. A smoother model topography and the reduced resolution of the 4DVAR data assimilation may be reasons for this (see Section 2).

E-Obs captures certain elements of the mesoscale Alpine precipitation pattern (e.g. the moist anomaly in the Julian Alps) but it lacks several of the characteristic anomalies, including for example the northern moist anomaly and the pattern at the Massif Central. Over the western Alps, precipitation anomalies are placed at maximum elevations, which may have similar reasons like for CRU, given the methodological similarities in the interpolation of the climatology (see Section 2). It is interesting to note that E-Obs compares much better with APGD in Germany and Slovenia, where the stations used are much denser, while major discrepancies to APGD go along with regions of poor station coverage (e.g. in France and Austria, see http://eca.knmi.nl/).

Mean annual precipitation for the datasets with highest resolution (MESAN and MESCAN, 5 km) are depicted in the right-hand column of Figure 2. The 5-km version of HIRLAM (obtained by nearest-neighbor downscaling, see Section 4) is included for comparison. The large resolution step from 22 to 5 kilometers reveals a further structuring of the pattern. Both downscaling datasets show precipitation anomalies associated with individual mountain massifs. Compared to HIRLAM, which provided the background to these OI systems, there is a generally better correspondence with APGD. Notably the overestimate by HIRLAM is substantially reduced. However, the degree of adjustment varies regionally. In France, MESAN and MESCAN show a spatially refined pattern, which is in close agreement with APGD, except for a somewhat smoother appearance. The agreement is not too surprising considering that the high-resolution station data available here contributes a bulk of added information. However, in Switzerland, MESAN and MESCAN largely replicate the spatial distribution of HIRLAM without adding

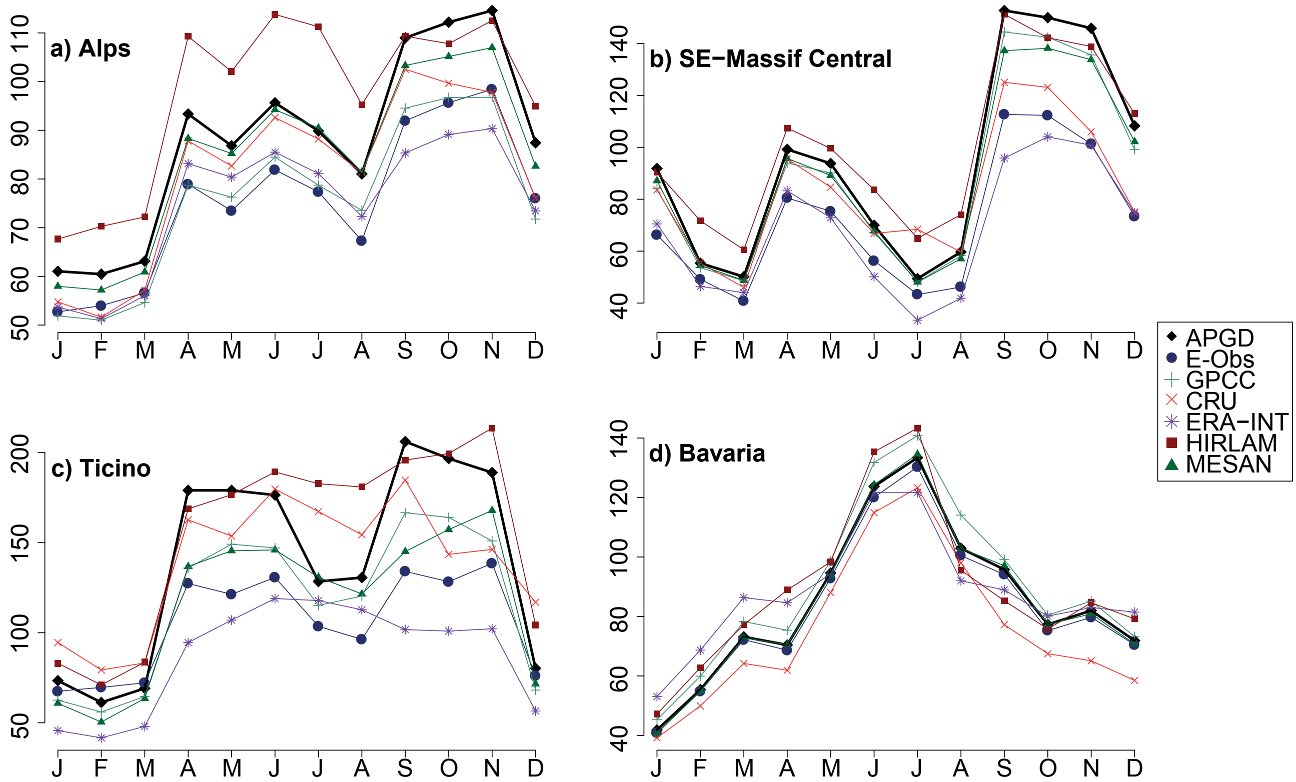

Figure 3: Mean monthly precipitation (1989–2003) showing differences in the annual cycle. Results are depicted for APGD (black) and six evaluation datasets (color and symbol, see legend). Numbers refer to averages over the entire Alpine region (a), and three subregions (South East Massif Central (b), Ticino (c) and Bavaria (d), see Fig. 1). The domain-mean values were derived from the 22-km versions of all datasets. Note the different scales between panels.

much detail (except for the correction of bias). Like in HIRLAM, the precipitation contrasts in the Aare catchment are overemphasized and the dryness of the Valais and upper Inn valleys are too broad. A similar lack in small-scale variability is found for Northern Italy. It seems that the coarse station network available in these regions has limited the ability of the downscaling methods to correct for errors in the background field and to compensate for the lack of spatial resolution. Between the two downscaling datasets the results are very similar. Differences can be noted over the Apennines and the Massif Central.

More insight into how the various datasets represent mean precipitation is provided in Figure 3, which depicts the annual cycle averaged over the Alps and over three subdomains. The comparison is restricted to datasets with a multi-year coverage. Generally, the main features are well captured by all datasets. Notably, they reproduce the distinct courses with a summer (convective) maximum at the northern Alpine rim (Bavaria) and two seasonal maxima (spring and autumn) in the southern regions (SE-Massif Central and Ticino). The underestimation by E-Obs and ERA-INT is evident in all domains (except Bavaria) and around the year. In E-Obs, the underestimation is particularly large at the southern rim (Fig. 3c) and relates to the poor representation of the prominent mesoscale anomalies there (see also Figure 2). In ERA-INT the underestimates are of comparable magnitude throughout the year. GPCC follows the annual cycle of the reference closely but underestimates its amplitude along the southern rim (Fig. 3c). The overestimate by HIRLAM is particularly from winter, spring and summer. In Ticino, this dataset (as well as CRU and ERA-INT) misses the intermediate reduction of mean precipitation in summer. Finally, the MESAN downscaling dataset reproduces the annual cycle very well and, hence, has corrected for much of the overestimates in HIRLAM. In Ticino, however, MESAN also underestimates the amplitude of the seasonal variation.

5.1.2 Wet-day frequency

In the Alpine region wet days are more frequent over the ridge and over the northern and western foreland of the Alps in Germany and France (APGD, Figure 4a). Along the French Mediterranean coast, the Po Valley and Eastern Austria the frequency is only about half that over the ridge. Lower frequencies are also found in interior parts of the Alps. This general pattern is qualitatively evident in all evaluated datasets, but with considerable quantitative discrepancies and mismatches in spatial detail.

E-Obs has a very smooth distribution with several of the topography-related patterns missing or misplaced. Over France, where the interpolation has to bridge large distances between stations, wet-day frequency is generally overestimated.

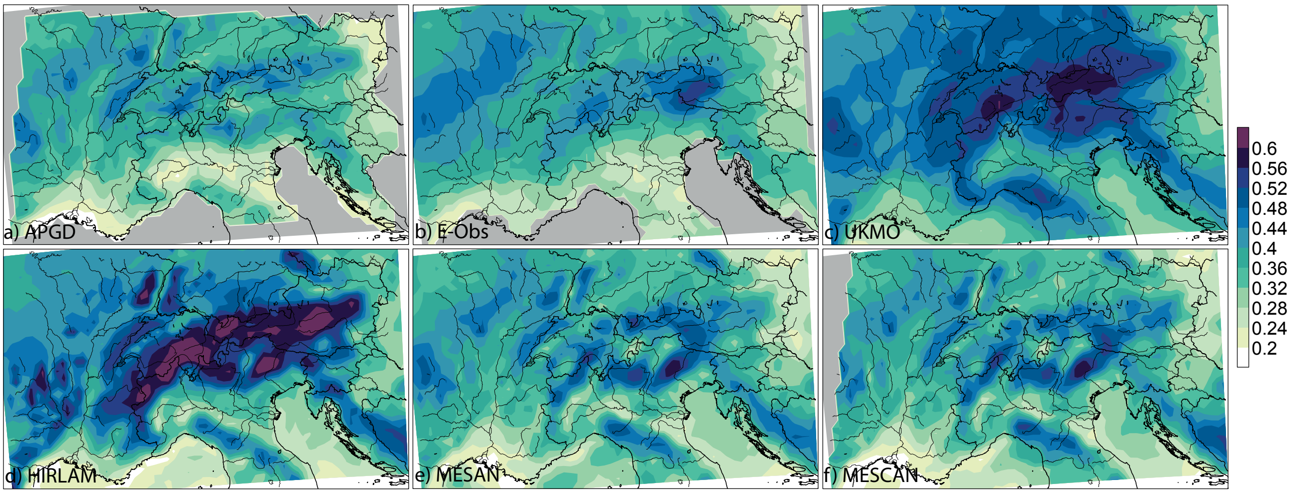

Figure 4: Annual frequency of wet days (≥ 1 mm, fraction) in 2008 at 22-km resolution. a) APGD; b) E-Obs; c) UKMO; d) HIRLAM; e) MESAN; f) MESCAN.

The two regional reanalyses (HIRLAM and UKMO) overestimate the frequency of wet days considerably over the entire domain. A similar overestimate with even larger frequencies was found for ERA-INT (not shown). In terms of spatial distribution, UKMO agrees well with APGD but has a much smoother variation than would be expected at this resolution. On the other hand, HIRLAM overestimates spatial variability, notably the contrasts between high and low elevations. These main spatial characteristics of the two regional reanalysis were found similarly for mean precipitation (Section 5.1.a).

MESAN and MESCAN (Fig. 4e, f) clearly benefit from the analysis of additional rain-gauge observations. Frequency and distribution of wet days are very different from HIRLAM that has served as background. Particularly good correspondence with the pattern of APGD is found over France, Germany and Slovenia, where a dense station network was available for downscaling. Residual signals from HIRLAM are, however, clearly discernable in other regions, notably in Northern Italy, where MESAN and MESCAN overestimate and misplace the gradient between Po valley and the Southern Alpine rim.

A similar analysis for individual seasons (not shown) reveals that the characteristic discrepancies of all evaluation datasets are similar in all seasons. Quantitatively, the overestimate in wet day frequency by the regional reanalysis HIRLAM is particularly large in spring and summer, the more convective seasons. A similar result is found for UKMO, however, this comparison may be less robust due to only one year of available data. As regards E-Obs, the overestimation in the Alps is much stronger in the winter half-year.

5.1.3 95 % quantile

A reliable reproduction of the occurrence of intense and heavy precipitation is crucial in many applications of climate datasets. Here we investigate this element in terms of the 95 % quantile, i.e. the daily precipitation amount that is exceeded on average every 20 days. The quantiles were determined grid-point by grid-point and, hence, are representative of intense area-average precipitation (not point values). Consistency in spatial scale is established by upscaling higher resolution datasets (Section 4). Results from the 22-km versions of the datasets and for year 2008 are depicted in Figure 5. The distribution in APGD (Fig. 5a) is very similar to that in a long-term climatology (see Isotta et al., 2014). Large values are found along the northern and southern Alpine rim, particularly in two regions to the south (Ticino and Julian Alps) that are well known for experiencing heavy precipitation frequently. The south-eastern slopes of the Massif Central and the Ligurian coast (Apennines) are also affected frequently.

Signatures of this characteristic pattern are evident in all datasets. The prominent anomalies along the southern rim and many other small-scale anomalies (Jura, Vosges and Black Forest), all of which are missed by the coarser resolution ERA-INT (not shown), are clearly evident in the regional reanalyses (HIRLAM, UKMO) and in the downscaling datasets (MESAN and MESCAN). Discrepancies to APGD are mostly with regard to the exact location (e.g. Ticino and Julian Alps anomalies shifted in HIRLAM), the spatial extent (anomalies in UKMO are broader) and relative amplitude (Massif Central in UKMO) of the anomalies. The downscaling datasets (MESAN and MESCAN) add spatial information over France and correct for the overestimate of HIRLAM over Southern Germany. However, the Ticino anomaly is excessively reduced and, particularly in MESAN, the quantile is distinctly underestimated there.

E-Obs shows more substantial differences to APGD than the regional reanalysis and downscaling datasets. Several of the mesoscale anomalies are missing and there is a tendency for underestimates generally over the domain. These discrepancies resemble those of mean precipitation (see Fig. 2) and may be similarly inter-

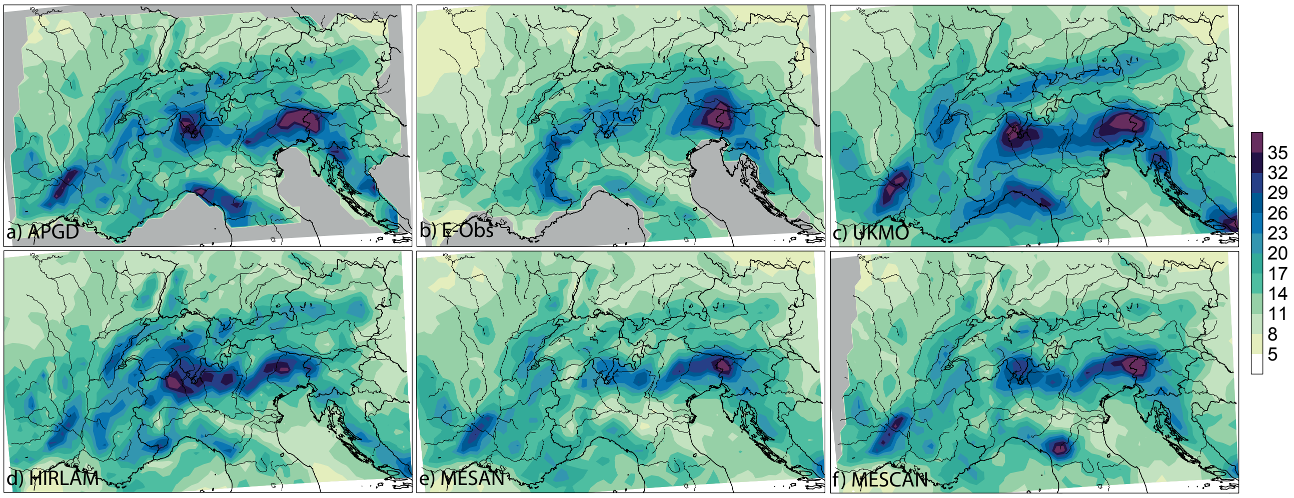

Figure 5: 95 % quantile of daily precipitation (mm) for 2008 at 22-km resolution. a) APGD; b) E-Obs; c) UKMO; d) HIRLAM; e) MESAN; f) MESCAN.

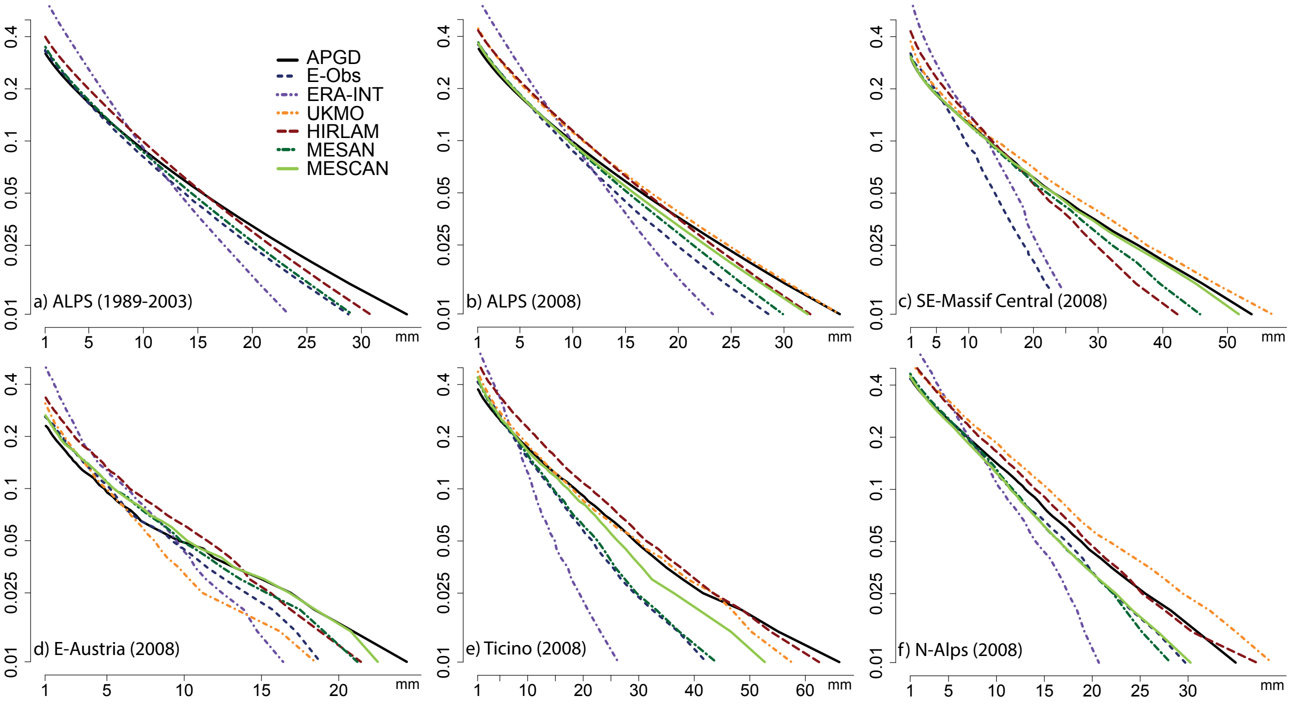

Figure 6: Frequency (y-axis, fraction of days, log-scale) at which daily precipitation amounts (values at 22-km resolution grid points) exceed a threshold (x-axis, mm). Results are plotted for different time periods (a) 1989–2003, (b–f) 2008, and different subdomains (see Figure 1): (a) and (b) whole Alps; (c) South East Massif Central; (d) East Austria; (e) Ticino; (f) Northern Alps. UKMO and MESCAN are displayed in panels (b)–(f) only. Note the different scales (x-axes).

preted as a consequence of the coarse station network, the attendant smoothing and the exaggeration of height gradients by the interpolation method.

A similar analysis for the four seasons (not shown) reveals that major features of the seasonal variation in heavy precipitation (high quantiles in autumn, low quantiles in winter) are found in all datasets. HIRLAM also reproduces main characteristics of the seasonal variation in the mesoscale pattern. In the one-year dataset available for UKMO there is a sign of substantial overestimation in autumn over Italy, but this comparison is uncertain given the short data series and high quantile considered. E-Obs underestimates the 95 % quantiles along the southern rim in all seasons, but particularly so in autumn and spring.

5.2 Distribution function

The frequency distribution of daily precipitation offers further insight into the error structure of the various datasets. Figure 6 depicts these distributions for several subdomains of the Alps by pooling the corresponding grid points (see Figure 1). Comparability between the various datasets is established by upscaling higher-resolution datasets to the 22-km standard grid before calculating the distributions (see Section 4). The results for ERA-INT are not strictly comparable (because of coarser resolution) but are added for comparison.

There are marked differences in the frequency distribution between the domains owing to the spatial variation in precipitation intensity. For example, the regions Ticino and SE-Massif Central exhibit distributions with much longer tails (note the different scales between the panels). All regional datasets reproduce these characteristic differences between subdomains, corroborating the finding of the previous subsection that major spatial variations in precipitation intensity in the Alps are captured.

A discrepancy common to many of the datasets is the tendency to overestimate the frequency of light and underestimate that of heavy precipitation (see e.g. panels a and b). The overestimate at small thresholds is particularly marked for the model-based regional reanalyses (HIRLAM and UKMO). This is not merely the result of too frequent drizzle; the overestimate is still evident at moderate precipitation thresholds. This bias is related to the overestimate of mean precipitation and wet-day frequency in these datasets (see Figure 2 and 3). On the other hand, the underestimation of high intensities is visible in all the dataset types. E-Obs underestimates the frequency after about 10 mm distinctly (note the log scale) and in all domains. In SE-Massif Central the frequency decrease is even steeper than that for the coarser resolution ERA-INT. An obvious explanation for this is the smoothing involved with the interpolation in data poor regions.

Underestimates of the frequency of intense events are also found for the downscaling datasets (MESAN more so than MESCAN, see panels b, e and f). In some domains this underestimation is even more substantial than in HIRLAM (panel e, f). This has similar reasons like in E-Obs, because the underestimation is strongly reduced in region SE-Massif Central where these datasets build on high-resolution station data. The result that MESCAN has a generally more long-tailed frequency distribution than MESAN, despite similar input data, may be attributable to a different form and the assumption of a shorter length scale in the structure function for background errors (35 km in MESCAN versus 240 km in MESAN), and, hence, more local corrections and less smoothing.

The UKMO regional reanalysis shows the best correspondence in the frequency of intense precipitation with APGD. Clear underestimates are present in one domain only (panel d), while there is a tendency for overestimates in others (panel f). This result is somewhat counterintuitive with the generally smoother spatial patterns (see Fig. 1, 2, 3). It seems that the lack in spatial variability in this reanalysis does not go along with a reduction of peak values.

5.3 Case study of an extreme precipitation event

To illustrate the degree of correspondence between the various observation datasets we briefly illustrate precipitation distributions for a particular case of heavy precipitation (Figure 7). The case consisted of two episodes (11–17 and 20–22 May 1999) with intense and enduring rainfall, succeeding shortly one after the other. The first – a quasi-stationary frontal system – affected parts of the northern Alpine rim and adjacent foreland regions in Switzerland, Austria and Southern Germany (Fig. 7a). The second – moist Mediterranean air masses lifted over a blocked cold-air layer – was even more intense and slightly shifted eastwards (Fig. 7e). In combination with rapid snowmelt, the two rainfall episodes caused flooding and costly damages particularly in eastern Switzerland, but also in Germany and Austria (Bundesamt für Wasser und Geologie, 2000).

The variations between datasets are quite different in the two episodes. In the first case, location and intensity of the precipitation peaks differ considerably. HIRLAM misses the core of the heavy precipitation system entirely. MESAN partly corrects for this, but the seven-day total is still underestimated by more than 30 %. In this episode, E-Obs shows the best correspondence with APGD. It correctly positions the peak intensity, however, again underestimates the totals. For the second period, the estimated fields are surprisingly similar. Still, the three evaluated datasets underestimate the peak intensity. In this case HIRLAM shows the best agreement with APGD. The 22-km version of MESAN, however, shows a clear underestimate of the peak intensity.

The depicted cases summarize impressions that we gained from a visual comparison of the studied datasets for a series of heavy precipitation cases in the Alps. Namely, that the accuracy of one single dataset can vary considerably from case to case and that a statistical assessment, such as that of the present study, may or may not be representative of the relative accuracy of datasets in a particular case.

5.4 Scores

Results from a quantitative comparison of daily precipitation in the evaluated datasets against APGD are shown in Figure 8 and 9. Figure 8 depicts the spatial distribution of root mean square error (RMSE, mm per day) and the Stable Equitable Error in Probability Space (SEEPS, Rodwell et al., 2010, dimensionless). The scores were determined from the daily precipitation series of year 2008 for each gridpoint of the 22-km versions of the datasets (see Section 4). Figure 9 compares mean values of the two scores averaged over subdomains (see Fig. 1) and stratified between winter half-year (defined here as the period from October to March) and summer half-year (April–September).

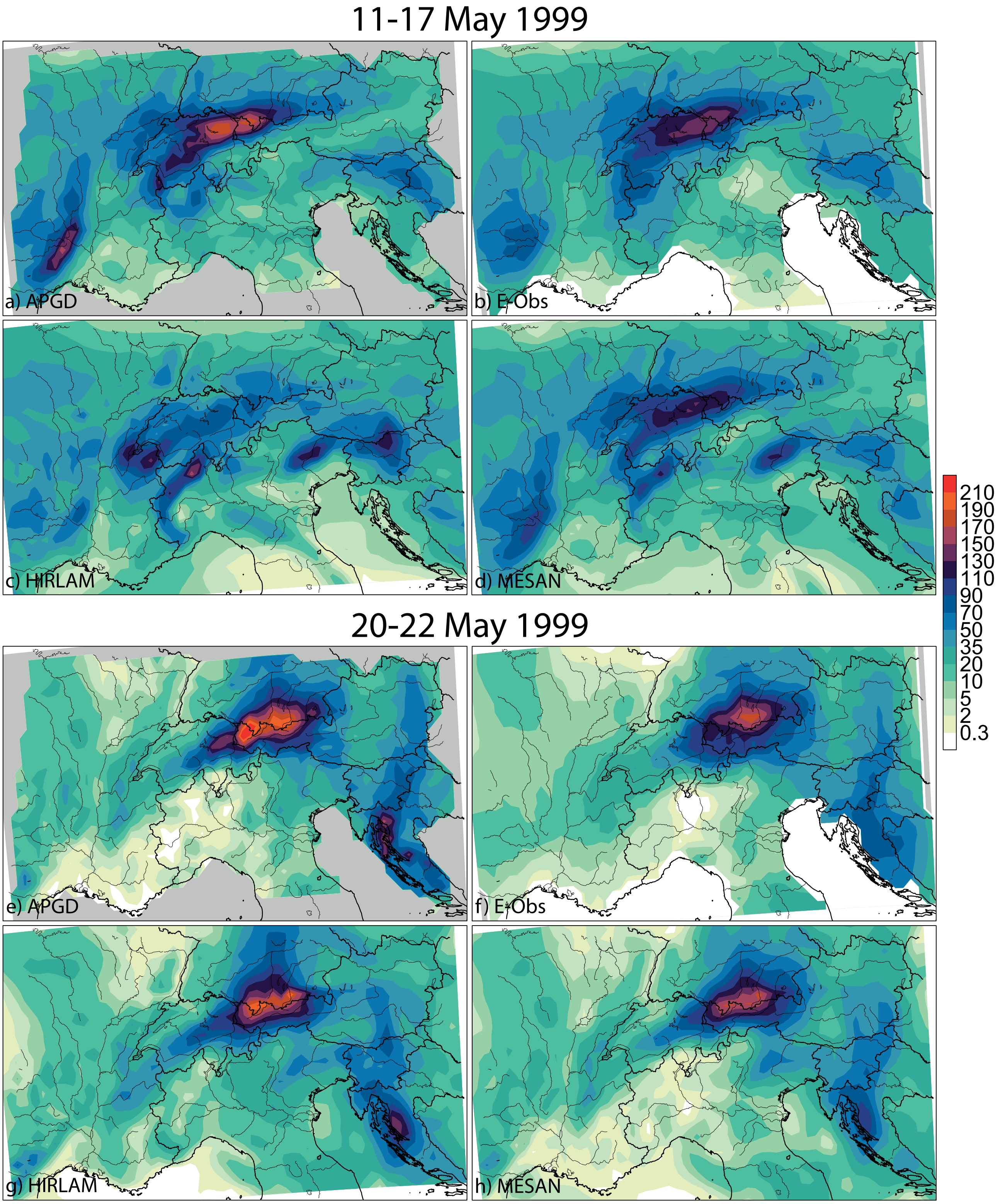

Figure 7: Precipitation sum (mm) for 11–17 May 1999 (a–d) and 20–22 May 1999 (e–h). Results for 22-km resolution datasets. a,e) APGD; b,f) E-Obs; c,g) HIRLAM; d,h) MESAN.

For most of the datasets the spatial distribution is different between the two scores (Fig. 8). RMSE tends to be larger in areas with a more vigorous precipitation climate (Massif Central, Ticino, Julian Alps, cf. Fig. 5), because it weights similar relative errors more heavily at high compared to low intensities. On the other hand, SEEPS tends to show more evenly distributed patterns due to its implicit calibration to the local climate (categories defined by quantiles). RMSE is considered here primarily to quantify the magnitude of typical local

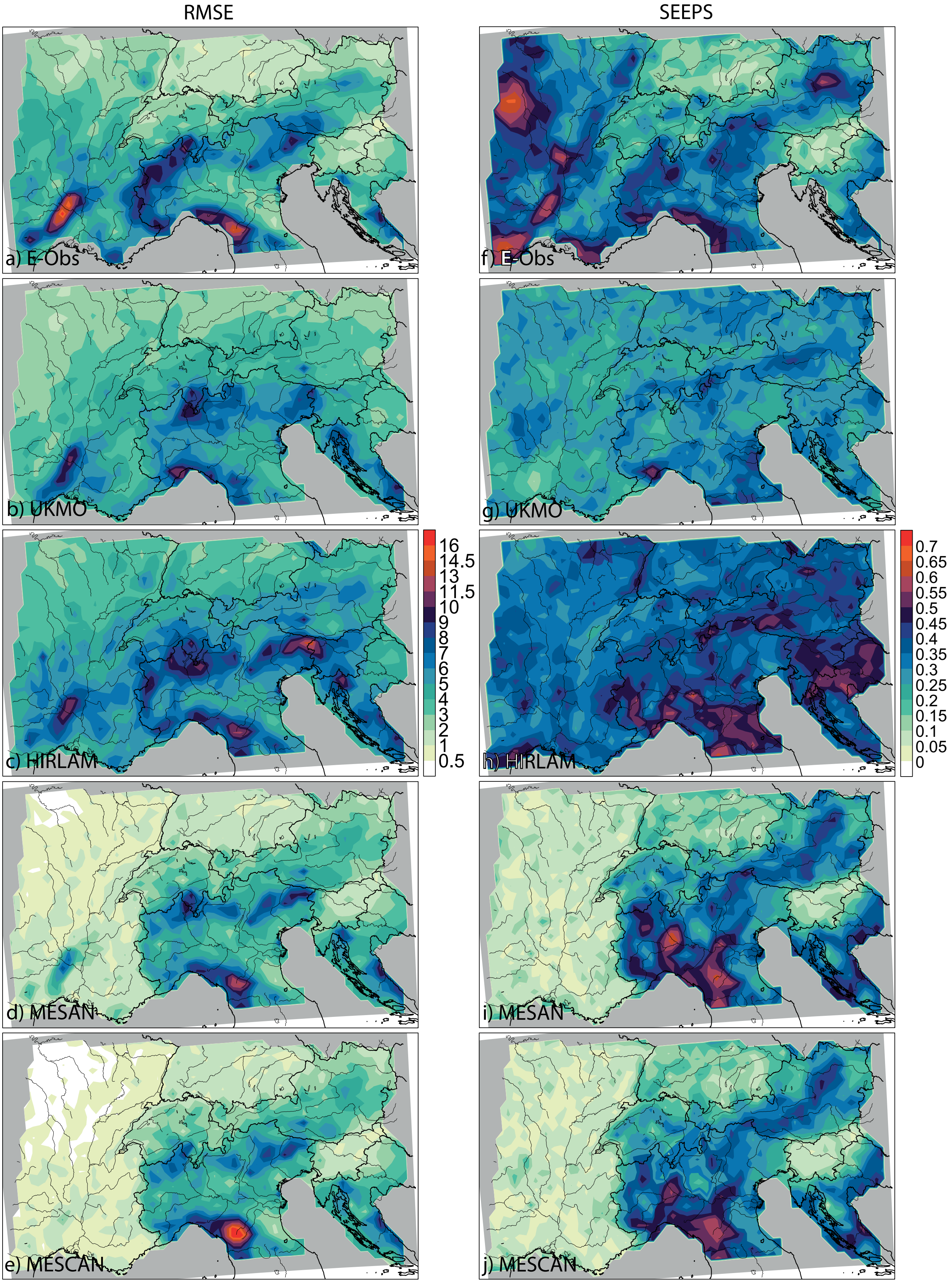

Figure 8: RMSE (mm/d, left column) and SEEPS (non-dimensional, right column) of the evaluated datasets when compared to APGD. Results are for daily precipitation in 2008 using the 22-km versions of the datasets. a, f) E-Obs; b, g) UKMO; c, h) HIRLAM; d, i) MESAN; e, j) and MESCAN.

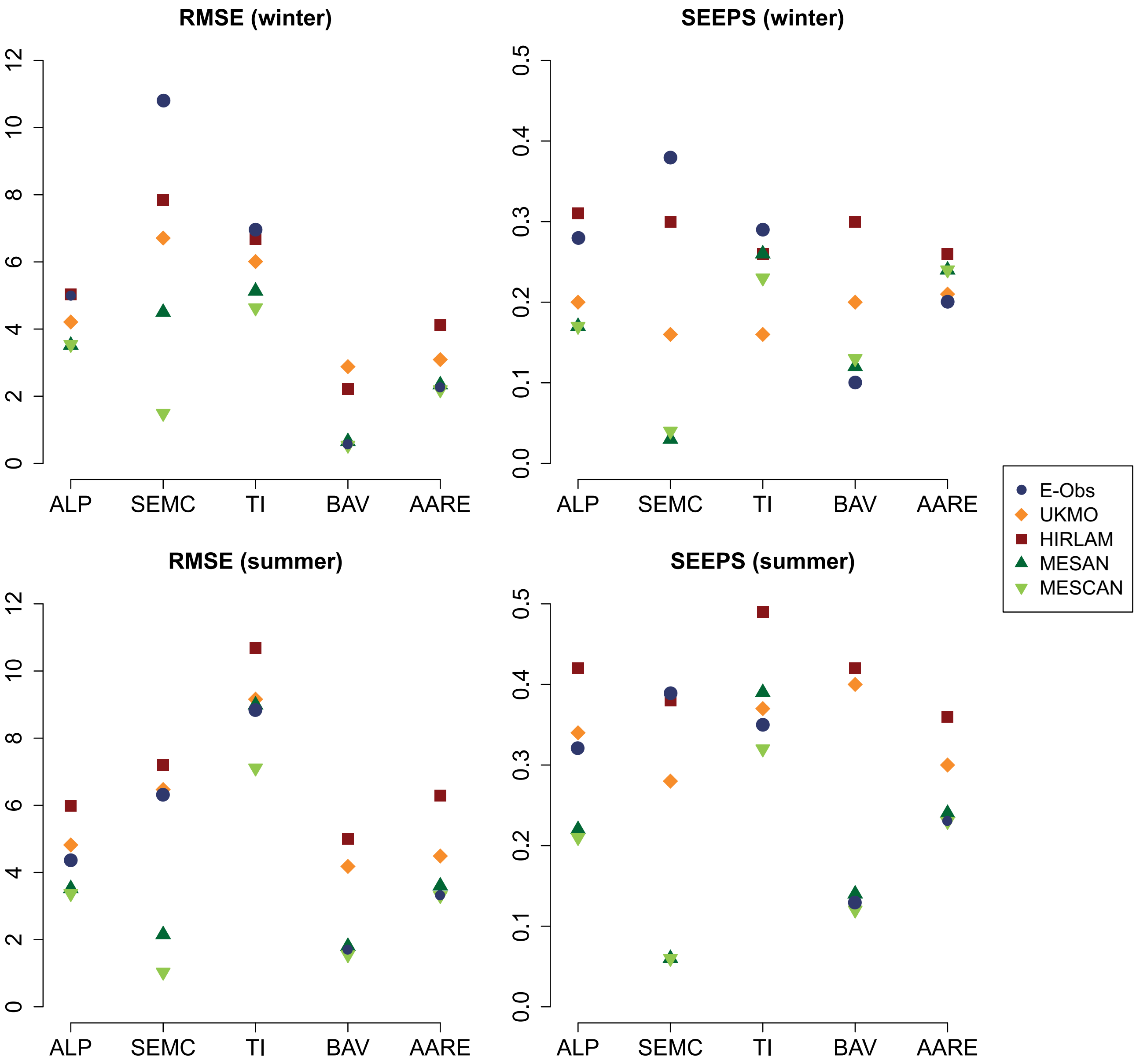

Figure 9: RMSE (mm/d, left column) and SEEPS (right column) for winter half-year (October–March; upper row) and for summer half-year (April-September; lower row) in 2008. The scores were calculated for the 22-km resolution datasets. Values are averages of the scores over five subdomains (see Fig. 1; ALP: whole domain, SEMC: SE-Massif Central, TI: Ticino, BAV: Bavaria, AARE: Aare river catchment).

errors, while SEEPS is used for comparing the degree of correspondence between different regions of the domain (see also Rodwell et al., 2010).

For the European interpolation dataset (E-Obs) and the downscaling datasets (MESAN and MESCAN) the density of input data seems to be a key factor of dataset reliability. In E-Obs this manifests in clear contrasts of SEEPS between Germany and Slovenia on the one hand, and France and Italy on the other hand (see Fig. 8, and regions BAV and SEMC in Fig. 9). These are exactly the regions with particularly dense and coarse station coverage respectively in the underlying ECA&D dataset (see Section 2). In MESAN and MESCAN, where high-resolution observations were also used over France (on top of the ECA&D data), the SEEPS score replicates the pattern of data density too and shows pronounced contrast across country borders, such as the French-Italian border (Fig. 8). Both downscaling datasets reveal very small values both for RMSE and SEEPS over France (region SEMC in Fig. 9), indicating very close correspondence with APGD. This is not surprising considering that over this part of the domain MESAN and MESCAN used the same high-resolution station data like APGD, except for differences in the data quality procedures.

The two regional reanalyses (UKMO, HIRLAM) show a much more evenly distributed SEEPS score (Fig. 8). There is an indication that correspondence with APGD is relatively poor over an elongated region of the inner Alpine dry climate (see also Fig. 3). This is likely a result of the somewhat coarser effective resolution of the reanalyses and difficulties to capture precipitation processes in this topographically complex region. UKMO has smaller errors than HIRLAM throughout the domain, with RMSE being typically 10–20 % smaller (Fig. 9). While the difference in RMSE may, to some extent, be influenced by the varying biases of the two reanalyses (see also Figs. 2 and 4), the systematic difference in SEEPS also points to a better reproduction of the temporal course of precipitation in UKMO.

Compared to E-Obs the two model-based reanalyses have smaller error scores in areas where E-Obs disposes of a very low station data input, such as France and parts of Northern Italy (see Fig. 8, regions SEMC and TI in Fig. 9). However, for much of the remaining areas, the European interpolation dataset corresponds better with APGD (see Germany in Fig. 8; regions BAV, and AARE in Fig. 9).

The two downscaling datasets provide clearly improved 22-km precipitation fields compared to the original regional reanalysis (HIRLAM, see Fig. 8). This is evident in areas with dense station coverage (e.g. domains SEMC and BAV in Fig. 9), where RMSE and SEEPS scores are smaller by a factor of 2 to 5. In the other regions improvements are smaller but still measureable (Fig. 9). An exception to this is the Apennine region, where HIRLAM has slightly smaller RMSE and SEEPS (Fig. 8). Comparison of the downscaling datasets against E-Obs reveals that MESAN and MESCAN rarely improve over the interpolation of station data alone when the observational data input is comparable (i.e. everywhere except France, Fig. 8). Only in Ticino (see Fig. 9) there is a measurable improvement over E-Obs. Clearly, MESAN and MESCAN offer higher spatial resolution than E-Obs, but at the coarser scale considered here, the gain from using a regional reanalysis (HIRLAM) as background is not clearly measureable.

Between MESAN and MESCAN there are only small differences in the scores, but errors are somewhat smaller for MESCAN (Figs. 8 and 9). This may be related to the generally stronger underestimate of high intensities by MESAN (see Figs. 5 and 6) and it may have its origin in the longer correlation range used for the background error function in MESAN, which is also evident in generally smoother precipitation fields (MESAN vs. MESCAN) as was verified in several case studies.

5.5 Intra-seasonal to inter-annual variations

In the Alpine region, the variations of precipitation within and between years can be considerable. Departures from the long-term monthly mean typically exceed the range of +/−25 % in more than half of the months, even when considering averages over mesoscale subregions. Several studies have also reported on long-term variations and trends in parts of the region (e.g. Schmidli et al., 2002; Brunetti et al., 2006; Efthymiadis et al., 2006; Zolina et al., 2008). In this context, the evaluated datasets examined in this study could provide a basis for monitoring the precipitation climate in the Alps and for studying its dynamical relationships (e.g. Klein Tank et al., 2002; Quadrelli et al., 2001). In view of such applications we briefly test the various datasets for their representation of intra-seasonal to inter-annual variations over the recent past.

For this purpose we have derived monthly time series of precipitation anomalies (ratios between actual monthly totals and long-term means of the calendar month) in several subdomains and separately for each dataset. The common 22-km grid was used to ensure consistent calculation of subdomain means, with higher- and lower-resolution datasets re-scaled (see Section 4). APGD is again used as reference. The evaluation is restricted to datasets covering the full 15 years (1989–2008), i.e. UKMO and MESCAN are omitted.

The evaluation of the temporal anomaly series is presented in the form of Taylor diagrams in Figure 10, comparing the correlation (angular direction) and the temporal standard deviation (radial direction) of the datasets with respect to APGD (circle on x-axis, Taylor, 2001).

For temporal anomalies over the entire Alpine region (Fig. 10a) all the datasets provide very consistent time series. Correlations with APGD all exceed 0.95. Also the magnitude of the variations is accurately captured, except that ERA-INT underestimates the standard deviation significantly, likely due to the limited resolution of mesoscale patterns. Larger differences between the datasets, however, emerge for meso-beta scale subdomains, where resolution and density of input data are more critical (Fig. 10b–d). Best correspondence with APGD is found for GPCC and MESAN (all subdomains) as well as for E-Obs in areas with dense station coverage (Bavaria). The correlations for CRU, HIRLAM and ERA-INT are clearly smaller and these datasets also underestimate variability.

In the mountainous regions Ticino and SE-Massif Central, monthly precipitation anomalies are often strongly influenced by a few events of intense precipitation. Errors in their spatial extent and amplitude, due to limited station coverage and smoothing, affect the accuracy of the time series. This is reflected, for example, in the substantial improvement of MESAN over E-Obs in SE-Massif Central. It is interesting that MESAN also improves on E-Obs in Ticino, despite similar data density and limited correlation of its background dataset (HIRLAM). It is unclear whether methodological differences in the spatial interpolation/downscaling are responsible for this.

In summary, the present evaluation indicates that most of the evaluated datasets provide quite consistent information regarding the temporal evolution of monthly anomalies. Even the coarse resolution global dataset GPCC can provide accurate regional monthly anomaly series. However, there is a tendency to underestimate variability, notably over the smaller domains. Datasets using station observations directly (interpola-

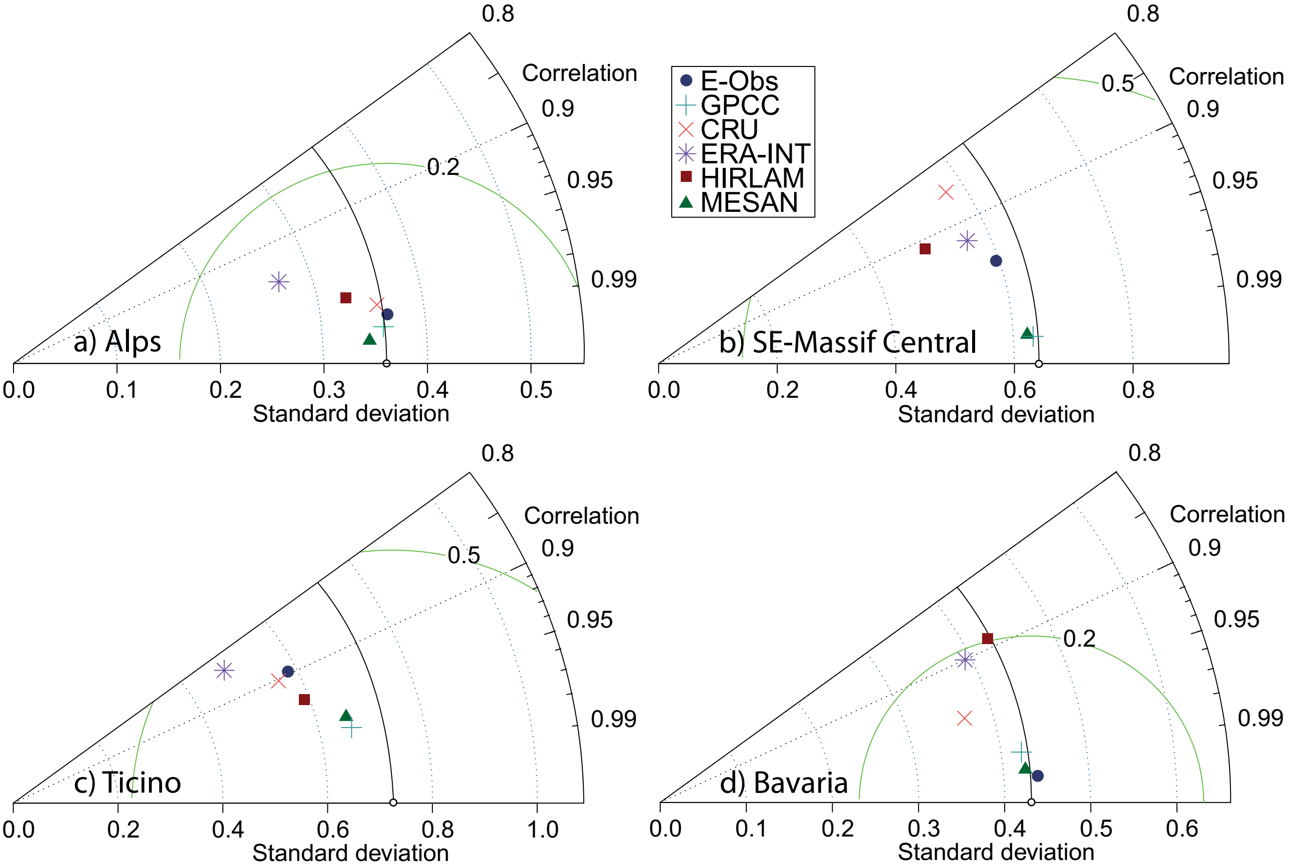

Figure 10: Taylor diagrams of the variation in relative monthly precipitation anomalies over the 15 years 1989–2003. APGD is used as reference (small circle on x-axis). The angular axis denotes correlation with APGD, the radial direction represents the standard deviation of the relative anomalies (dimensionless). a) whole Alpine domain; b) SE-Massif Central; c) Ticino; d) Bavaria (see Fig. 1 for details of the regions). The domain-mean anomalies were calculated from the 22-km resolution versions of the datasets.

tion or downscaling) are observed to be more accurate than the model-based regional reanalyses.

6 Conclusions

We have evaluated and compared several existing and newly developed climate datasets that extend over the entire European continent (at least) and offer themselves for application in environmental research, climate monitoring and the study of climate change impacts. The suite of evaluated datasets encompasses rain-gauge-based interpolation datasets (E-Obs, CRU, GPCC), model-based regional (and one global) reanalyses (UKMO, HIRLAM and ERA-INT) and downscaling datasets (MESAN and MESCAN). The comparison is conducted for daily precipitation in the region of the European Alps. Their topographic complexity and climatic diversity provide an ideal test bed for datasets that promise to resolve meso-beta scale climatic features and heavy precipitation better than existing global and monthly datasets.

The reference is a gridded interpolation from high-resolution rain-gauge networks (1971–2008, Isotta et al., 2014). The APGD dataset stands out by uniquely dense (5500 stations on average) and spatially more homogenous data coverage than any of the evaluated datasets. Inaccuracies in the reference dataset are to be expected particularly at small scales, which is why the smallest of the considered scales (5 km) is used for visual comparison only. At the intermediate (22 km) and coarse (55 km) resolution, the reference dataset is expected to have smaller errors than the major discrepancies of the evaluated datasets that we describe here.

The new model-based regional reanalyses (UKMO and HIRLAM) show much richer and more realistic spatial variations in the climatology of mean Alpine precipitation and several statistics of daily precipitation than the driving global reanalysis (ERA-INT). They also correct for several unrealistic spatial features seen in global and continental interpolation datasets, deficiencies related to the limited station coverage available for these latter datasets.

In absolute values, however, the regional reanalyses exhibit biases and misplacements of regional anomalies that can result, locally, in discrepancies of similar or larger amplitude like those of the driving reanalysis or from interpolation datasets. The regional reanalyses overestimate mean precipitation in the Alps and they misrepresent mountain-valley and mountain-flatland contrasts (overemphasized in HIRLAM and underemphasized in UKMO). They also overestimate the frequency of wet days (light/moderate events) and underestimate that of intense precipitation. The 4DVar UKMO reanalysis underestimates intense events less and exhibits a better temporal correspondence (RMSE, SEEPS) with the reference dataset than the 3DVar HIRLAM reanalysis.

The accuracy of datasets that explicitly use in-situ rain-gauge observations (for interpolation or downscaling) is found to strongly depend on the station density available in their construction. As a result, their performance is very heterogeneous over the Alpine region. In areas with poor data coverage important mesoscale features are missing and there are biases in the frequency distribution (too frequent wet days, too small high quantiles) that relate to reduced effective resolution (larger search neighbourhood, see also Hofstra et al., 2009). In areas with dense station coverage, however, the continental interpolation dataset E-Obs is less biased and has better correspondence (RMSE and SEEPS) with the reference than the regional reanalyses. Likewise, when fed with dense station data, the two downscaling datasets MESAN and MESCAN largely correct for biases in the underlying regional reanalysis and reduce RMSE and SEEPS errors substantially (a factor of 2 to 5). In areas where E-Obs and the downscaling datasets have similar input of station data the skill of the two methodologies is comparable and it is unclear whether downscaling has profited from the usage of a regional reanalysis on top of the station data.

Our visual comparisons and reported error patterns (frequency distribution) suggest that all datasets do not fully resolve the spatial variations that one might expect to be resolved on the basis of their respective grid spacing. In model-based regional reanalyses this is due to a smooth topography, processes not resolved at the grid-spacing scale, and, in the case of UKMO, a data assimilation coarser than the model grid. In the interpolation and downscaling datasets it is the limited station network that compromises effective resolution (e.g. Ensor and Robeson, 2008; Hofstra et al., 2010; Turco et al., 2013). As a consequence, there is considerable uncertainty about and spatial heterogeneity in the scales effectively resolved in the datasets. On the one hand this causes difficulties in defining a strictly comparable reference dataset for the evaluation (e.g. Frei et al., 2003). On the other hand, several applications of such datasets – notably in hydrology and with extremes – are resolution dependent (e.g. Tetzlaff and Uhlenbrook, 2005; Lobligeois et al., 2013). Here, we recommend to presume that the datasets are of coarser resolution than the grid spacing, which would reduce impacts from the smoothing-related biases seen in our evaluation.

Despite biases and inaccuracies in the climatological patterns, the evaluated datasets provide very consistent results on intraseasonal to interannual variations of monthly mean precipitation. Correlations with the reference dataset are mostly larger than 0.9 when considering subdomains of 100 × 100 square kilometres or larger. Even the coarse resolution global dataset GPCC is found to provide accurate regional monthly anomaly series in Alpine subdomains. This is a promising result for the utility of the evaluated datasets as a basis for trans-European climate monitoring. What remains to be assessed is, however, how accurately the datasets can reproduce longer-term variations and trends, and how the underestimate of temporal variability, observed with most of the datasets, affects the statistical significance of the trends.

Our evaluations illustrate that regional reanalyses provide a valuable new data resource also in regions with a complex climatology. Yet, there are still considerable biases involved and applications will continue to involve bias corrections and downscaling procedures that rely on direct observations. The reliability of these post-processings as well as the accuracy of model-independent datasets for evaluation was found to depend critically on station density. Further developments would therefore substantially profit from improvements in the accessibility of high-resolution station data.

Acknowledgements

The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7/2007–2013) under grant agreement n ° 242093. We are grateful to the following persons and institutions for providing the evaluated datasets and for commenting on early results: Tomas Landelius and Per Undén (Swedish Meteorological and Hydrological Institute, HIRLAM and MESAN); Eric Bazile and Cornel Soci (Météo-France, MESCAN); Peter Jeremy, Richard Renshaw and Dale Barker (Met Office United Kingdom, UKMO); Anna Ghelli and Hans Hersbach (ECMWF, ERA-INT). We also acknowledge the EOBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://eca.knmi.nl). The GPCC dataset was provided by Deutscher Wetterdienst (http://gpcc.dwd.de), and the CRU dataset by the Climatic Research Unit, University of East Anglia (http://www.cru.uea.ac.uk). We also thank two anonymous reviewers for their valuable comments.

References

Auer, I., R. Böhm, A. Jurkovic, W. Lipa, A. Orlik, R. Potzmann, W. Schöner, M. Ungersböck, C. Matulla, K. Briffa, P. Jones, D. Efthymiadis, M. Brunetti, T. Nanni, M. Maugeri, L. Mercalli, O. Mestre, J.-M. Moisselin, M. Begert, G. Müller-Westermeier, V. Kveton, O. Bochnicek, P. Stastny, M. Lapin, S. Szalai, T. Szentimrey, T. Cegnar, M. Dolinar, M. Gajic-Capka, K. Zaninovic, Z. Majstorovic, E. Nieplova, 2007: HISTALP-historical instrumental climatological surface time series of the Greater Alpine Region. – Int. J. Climatol. **27**, 17–46. DOI:10.1002/joc.1377.

Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, M. Ziese, 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. **5**, 71–99. DOI:10.5194/essd-5-71-2013.

Blenkinsop, S., H.J. Fowler, I.G. Dubus, B.T. Nolan, J.M. Hollis, 2008: Developing climatic scenarios for pesticide fate modelling in Europe. – Environ. Poll. **154**, 219–231.

Bosshard, T., S. Kotlarski, M. Zappa, C. Schär, 2014: Hydrological climate-impact projections for the Rhine river: GCM-RCM uncertainty and separate temperature and precipitation effects. – J. Hydrometeor. **15**, 697–713. DOI:10.1175/JHM-D-12-098.1.

Brunetti, M., M. Maugeri, T. Nanni, I. Auer I, R. Böhm, W. Schoner, 2006: Precipitation variability and changes in the greater alpine region over the 1800–2003 period. – J. Geophys. Res. **111**, D11107. DOI:10.1029/2005JD006674.

Brunetti, M., G. Lentini, M. Maugeri, T. Nanni, C. Simolo, J. Spinoni, 2012: Projecting North Eastern Italy temperature and precipitation secular records onto a high-resolution grid. – Phys. Chem. Earth **40**, 9–22.

Bukovsky, M.S., D.J. Karoly, 2007: A brief evaluation of precipitation from the North American regional reanalysis. – J. Hydrometeorol. **8**, 837–846. DOI:10.1175/JHM595.1

Bundesamt für Wasser und Geologie, 2000: Hochwasser 1999, Analyse der Ereignisse. – Studienbericht **10**, 160 pp.

Bush, M., S. Bell, N. Christidis, R. Renshaw, B. Macpherson, C. Wilson, 2006: Development of the North Atlantic European Model (NAE) into an operational model. – Met Office Forecasting Research Technical Report **470**, http://www.metoffice.gov.uk/archive/forecasting-research-technical-report-470.

Cherubini, T., A. Ghelli, F. Lalaurette, 2002: Verification of precipitation forecasts over the Alpine region using a high-density observing network. – Wea. Forecast. **17**, 238–249.

Dahlgren, P., N. Gustafsson, 2012: Assimilating host model information into a limited area model. – Tellus A **64**, 15836. DOI:10.3402/tellusa.v64i0.15836

Daley, R., 1991: Atmospheric data analysis. – Cambridge University Press, United Kingdom, 457 pp. ISBN 0-521-38215-7.

Daly C., R.P. Neilson, D.L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. – J. Appl. Meteor. **33**, 140–158.

Daly C., W.P. Gibson, G.H. Taylor, G.L. Johnson, P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. – Climate Res. **22**, 99–113.

Dee, D., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A.P. McNally, B.M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. – Quart. J. Roy. Meteor. Soc. **137**, 553–597. DOI:10.1002/qj.828

Dolinar, M., 2006: Spatial interpolation of sunshine duration in Slovenia. – Meteor. Appl. **13**, 375–384. DOI:10.1017/S1350482706002362

Ebert, E.E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. – Meteor. Apps. **15**, 51–64.

Efthymiadis, D., P.D. Jones, K. Briffa, I. Auer, R. Böhm, W. Schönder, C. Frei, J. Schmidli, 2006: Construction of a 10-min gridded precipitation dataset for the Greater Alpine Region 1800–2003. – J. Geophys. Res. **111**, D01105. DOI:10.1029/2005JD006120.

Ensor, L.A., S.M. Robeson, 2008: Statistical Characteristics of Daily Precipitation: Comparisons of Gridded and Point Datasets. – J. Appl. Meteor. Climatol. **47**, 2468–2476. DOI:10.1175/2008JAMC1757.1

Erdin, R., C. Frei, H.R. Künsch, 2012: Data transformation and uncertainty in geostatistical combination of radar and rain gauges. – J. Hydrometeorol. **13**, 1332–1346.

Frei, C., C. Schär, 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. – Int. J. Climatol. **18**, 873–900.

Frei, C., J.H. Christensen, M. Déqué, D. Jacob, R.G. Jones, P.L. Vidale, 2003: Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. – J. Geophys. Res. **108**, D3 4124. DOI:10.1029/2002JD002287.

Frei, C., M. Willi, R. Stöckli, B. Dürr, submitted: Spatial analysis of sunshine duration in complex terrain by non-contemporaneous combination of station and satellite data. – Int. J. Climatol.

Greminger, P.J., 2003: Natural hazards and the Alpine Convention - Event analysis and recommendations. – Federal Office for Spatial Development (ARE), 53 pp.

Gustafsson, N., L. Berre, S. Hörnquist, Y. Huang, M. Lindskog, B. Navascues, K.S. Mogensen, S. Thorsteinsson, 2001: Three-dimensional variational data assimilation for a limited area model Part I: General formulation and the background error constraint. – Tellus **53 A**, 425–446.

Häggmark, L., I. Ivarsson, S. Gollvik, O. Olofsson, 2000: Mesan, an operational mesoscale analysis system. – Tellus **52A**, 2–20.

Hahmann, A.N., D. Rostkier-Edelstein, T.T. Warner, F. Vandenberghe, Y. Liu, R. Babarsky, S.P. Swerdlin, 2010: A Reanalysis System for the Generation of Mesoscale Climatographies. – J. Appl. Meteor. Climatol. **49**, 954–972.

Harris, I., P.D. Jones, T.J. Osborn, D.H. Lister, 2014: Updated high-resolution grids of monthly climatic observations. – Int. J. Climatol. **34**, 623–642. DOI:10.1002/joc.3711.

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones, M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. – J. Geophys. Res. **113**, D20119. DOI:10.1029/2009JD011799.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. – Int. J. Climatol. **25**, 1965–1978.

Hofstra, N., M. Haylock, M. New, P.D. Jones, C. Frei, 2008: The comparison of six methods for the interpolation of daily European climate data. – J. Geophys. Res. **113**, D21110. DOI:10.1029/2008JD010100.

Hofstra, N., M. Haylock, M. New, P.D. Jones, 2009: Testing E-Obs European high-resolution gridded data set of daily precipitation and surface temperature. – J. Geophys. Res. **114**, D21101. DOI:10.1029/2009JD011799.

Hofstra, N., M. New, C. McSweeney, 2010: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. – Climate Dyn. **35**, 841–858.

Holzkämper, A., P. Calanca, J. Fuhrer, 2012: Statistical crop models: predicting the effects of temperature and precipitation changes. – Climate Res. **51**, 11–21.

Hutchinson, M.F., 1995: Interpolation of mean rainfall using thin plate smoothing splines. – Int. J. Geogr. Inf. Syst. **9**, 385–403.

Isotta, F.A., C. Frei, V. Weilguni, M. Perčec Tadić, P. Lassègues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S.M. Ratto, M. Munari, S. Micheletti, V. Bonati, C. Lussana, C. Ronchi, E. Panettieri, G. Marigo, G. Vertačnik, 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

Jansson, A., C. Persson, G. Strandberg, 2007: 2D mesoscale re-analysis of precipitation, temperature and wind over Europe – ERAMESAN: Time period 1980–2004. – SMHI Reports Meteorology and Climatology **112**, 44 pp.

Kalnay, E., 2003: Atmospheric Modeling, Data Assimilation and Predictability. – Cambridge University Press, United Kingdom, 369 pp.

Klein Tank, A.M.G., J.B. Wijngaard, G.P. Koennen, R. Boehm, G. Demaree, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin, G. Müller-Westmeier, M. Tzanakou, S. Szalai, T. Pálsdóttir, D. Fitzgerald, S. Rubin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A.F.V. van Engelen, E. Forland, M. Mietus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar, J. Antonio López, B. Dahlström, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L.V. Alexander, P. Petrovic, 2002: Daily surface air temperature and precipitation dataset 1901–1999 for European Climate Assessment (ECA). – Int. J. Climatol. **22**, 1441–1453.

Klein Tank, A.M.G., F.W. Zwiers, X. Zhang, 2009: Guidelines on: Analysis of extremes in a changing climate in support of informed decisions for adaptation. – World Meteorological Organization, Report WCDMP-72, WMO-TD 1500, Geneva, Switzerland, 52 pp.

Klok, E.J., A.M.G. Klein Tank, 2008: Updated and extended European dataset of daily climate observations. – Int. J. Climatol. **29**, 1182–1191. DOI:10.1002/joc.1779

Köplin, N., B. Schädler, D. Viviroli, R. Weingartner, 2014: Seasonality and magnitude of floods in Switzerland under future climate change. – Hydrol. Process. **28**, 2567–2578. DOI:10.1002/hyp.9757

Lenderink, G., 2010: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations. – Climate Res. **44**, 151–166.

Lindskog, M., N. Gustafsson, B. Navascues, K.S. Mogensen, Y. Huang, X. Yang, L. Andrae, S. Thorsteinsson, J. Ranrakokko, 2001: Three-dimensional variational data assimilation for a limited area model Part II: Observation handling and assimilation experiments. – Tellus **53A**, 447–468.

Lobligeois, F., V. Anddréssian, C. Perrin, P. Tabary, C. Loumagne, 2013: When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation on 3620 flood events. – Hydrol. Earth Syst. Sci. Discuss. **10**, 12485–12536.

Lorenz, C., H. Kunstmann, 2012: The Hydrological Cycle in Three State of the Art Reanalyses: Intercomparison and Performance Analysis. – J. Hydrometeor. **13**, 1397–1420. DOI:10.1175/JHM-D-11-088.1

Luhamaa, A., K. Kimmel, A. Männik, R. Rõõm, 2011: High resolution re-analysis for the Baltic Sea region during 1965–2005 period. – Climate Dyn. **36**, 727–738.

Machguth, H., F. Paul, S. Kotlarski, M. Hoelzle, 2009: Calculating distributed glacier mass balance for the Swiss Alps from regional climate model output: A methodical description and interpretation of the results. – J. Geophys. Res. **114**, D19106. DOI:10.1029/2009JD011775.

Mahfouf, J.-F., B. Brasnett, S. Gagnon, 2007: A Canadian precipitation analysis (CaPA) project: Description and preliminary results. – Atmos.-Ocean **45**, 1–17. DOI:10.3137/ao.v450101

Mearns, L.O., F. Giorgi, C. McDaniel, 1995: Analysis of daily variability of precipitation in a nested regional climate model: Comparison with observations and doubled CO2 results. – Global Planet. Change **10**, 55–78.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P.C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E.H. Berbery, M.B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, W. Shi, 2006: North American Regional Reanalysis. – Bull. Amer. Meteor. Soc. **87**, 343–360. DOI:10.1175/BAMS-87-3-343

Mitchell, T.D., R.G. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. – Int. J. Climatol. **25**, 693–712. DOI:10.1002/joc.1181.

Monestiez, P., D. Courault, D. Allard, F. Ruget, 2001: Spatial interpolation of air temperature using environmental context: Application to a crop model. – Environ. Ecol. Stat. **8**, 297–309.

Neff, E.L., 1977: How much rain does a rain gage? – J. Hydrol. **35**, 213–220.

New, M., M. Hulme, P.D. Jones, 1999: Representing twentieth century space-time climate variability. Part 1: development of a 1961–90 mean monthly terrestrial climatology. – J. Climate **12**, 829–856. DOI: 10.1175/1520-0442(1999)012<0829:RTCSTC>2.0.CO;2.

Nikulin, G., E. Kjellström, G. Hansson, A. Ullerstig, 2011: Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. – Tellus **63A**, 41–55.

Onogi, K., J. Tsutsui, H. Koide, M. Sakamoto, S. Kobayashi, H. Hatsushika, T. Matsumoto, N. Yamazaki, H. Kamahori, K. Takahashi, S. Kadokura, K. Wada, K. Kato, R. Oyama, T. Ose, N. Mannoji, R. Taira, 2007: The JRA-25 Reanalysis. – J. Meteor. Soc. Japan **85**, 369–432.

Osborn, T.J., M. Hulme, 1997: Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. – J. Climate **10**, 1885–1908.

Ott, I., D. Duethmann, J. Liebert, P. Berg, H. Feldmann, J. Ihringer, H. Kunstmann, B. Merz, G. Schaedler, S. Wagner, 2013: High-Resolution Climate Change Impact Analysis on Medium-Sized River Catchments in Germany: An Ensemble Assessment, –J. Hydrometeor. **14**, 1175–1193. DOI10.1175/JHM-D-12-091.1

Paulat, M., C. Frei, M. Hagen, H. Wernli, 2008: A gridded dataset of hourly precipitation in Germany: its construction, application and climatology. – Meteorol. Z. **17**, 719–732.

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

Perry, M., D. Hollis, 2005: The generation of monthly gridded datasets for a range of climatic variables over the UK. – Int. J. Climatol. **25**, 1041–1054.

Quadrelli, R., M. Lazzeri, C. Cacciamani, S. Tibaldi, 2001: Observed winter Alpine precipitation variability and links with large-scale circulation patterns. – Climate Res. **17**, 275–284.

Rauthe, M., H. Steiner, U. Riediger, A. Mazurkiewicz, A. Gratzki, 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.

Rawlins, F., S.P. Ballard, K.J. Bovis, A.M. Clayton, D. Li, G.W. Inverarity, A.C. Lorenc, T.J. Payne, 2007: The Met Office global four-dimensional variational data assimilation scheme. – Quart. J. Roy. Meteor. Soc. **133**, 347–362. DOI:10.1002/qj.32

Renshaw, R., P. Jermey, D. Barker, A. Maycock, S. Oxley, 2013: EURO4M regional reanalysis system. – Forecast. Res., Technical Report **583**, 30 pp.

Richter, D., 1995: Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann-Niederschlagsmessers. – Bericht Deutschen Wetterdienstes **194**, 93 pp. (to be obtained from German Weather Service, Offenbach a.M., Germany).

Rienecker, M.M., M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu E, M.G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R.D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C.R. Redder, R. Reichle, F.R. Robertson, A.G. Ruddick, M. Sienkiewicz, J. Woollen, 2011: MERRA – NASA's Modern-Era Retrospective Analysis for Research and Applications. – J. Climate **24**, 3624–3648. DOI:10.1175/JCLI-D-11-00015.1.

Rodwell, M.J., D.S. Richardson, T.D. Hewson, T. Haiden, 2010: A new equitable score for verifying precipitation in numerical weather prediction. – Quart. J. Roy. Meteor. Soc. **136**, 1344–1363.

Rubel, F., M. Hantel, 2001: BALTEX 1/6-degree daily precipitation climatology 1996–1998. Meteorol. Atmos. Phys. **77**, 155–166.

Rubel, F., B. Rudolf, 2001: Global daily precipitation estimates proved over the European Alps. – Meteorol. Z. **10**, 407–418.

Saarinen, S., 2004: ODB User Guide. – Available from ECMWF, Reading, UK.

Saha, S., S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y.-T. Hou, H.-Y. Chuang, H.-M.H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. van Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. van den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J.-K. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C.-Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. W Reynolds, G. Rutledge, M. Goldberg, 2010: The NCEP Climate Forecast System Reanalysis. – Bull. Amer. Meteor. Soc. **91**, 1015–1057. DOI:10.1175/2010BAMS3001.1.

Schär, C., T.D. Davies, C. Frei, H. Wanner, M. Widmann, M. Wild, H.C. Davies, 1997: Current Alpine climate. – In Cebon P., et al. (Eds.): Views from the Alps: Regional Perspectives on Climate Change. – MIT Press, 20–72.

Schmidli, J., C. Schmutz, C. Frei, H. Wanner, C. Schär, 2002: Mesoscale precipitation variability in the region of the European Alps during the 20th century. – Int. J. Climatol. **22**, 1049–1074.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. – Theor. Appl. Climatol. **115**, 15–40. DOI:10.1007/s00704-013-0860-x.

Schrier, G. van der, E.J.M. van den Besselaar, A.M.G. Klein Tank, G. Verver, 2013: Monitoring European averaged temperature based on the E-Obs gridded dataset. – J. Geophys. Res. **118**, 5120–5135.

Schwarb, M., 2000: The Alpine precipitation climate: Evaluation of a high-resolution analysis scheme using comprehensive rain-gauge data. – Dissertation No 13911, ETH-Zürich, Switzerland.

Schwarb, M., C. Daly, C. Frei, C. Schär, 2001: Mean annual and seasonal precipitation in the European Alps 1971–1990. – Hydrological Atlas of Switzerland, available from University of Bern, Bern, Plates 2.6 and 2.7.

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

Shepard, D.S, 1984: Computer Mapping: The SYMAP Interpolation Algorithm. – In: Gaile G.L., C.J. Willmott (Eds): Spatial Statistics and Models, 133–145.

Smiatek, G., H. Kunstmann, R. Knoche, A. Marx, 2009: Precipitation and temperature statistics in high-resolution regional climate models: Evaluation for the European Alps. – J. Geophys. Res. **114**, D19107. DOI:10.1029/2008JD011353.

Soares, P.M.M., R.M. Cardoso, P.M.A. Miranda, J. de Medeiros, M. Belo-Pereira, F. Espirito-Santo, 2012: WRF high resolution dynamical downscaling of ERA-Interim for Portugal. – Climate Dyn. **39**, 2497–2522. DOI:10.1007/s00382-012-1315-2.

Suklitsch, M., A. Gobiet, A. Leuprecht, C. Frei, 2008: High resolution sensitivity studies with the regional climate model CCLM in the Alpine Region. – Meteorol. Z. **17**, 467–476.

Taillefer, F., 2002: CanariI-Technical documentation. – Technical documentation. Numerical Weather Prediction Research Group/ National Centre of Meteorological Research (Météo-France). – Available from Météo-France, Toulouse, France.

Taylor, K.E., 2001: Summarizing multiple aspects of model performance in a single diagram. – J. Geophys. Res. **106**, 7183–7192.

Tetzlaff, D., U. Uhlenbrook, 2005: Effects of spatial variability of precipitation for process-orientated hydrological modelling: results from two nested catchments. – Hydrol. Earth Syst. Sci. **9**, 29–41.

Turco, M., A. Zollo, C. Ronchi, C. de Luigi, P. Mercogliano, 2013: Assessing gridded observations for daily precipitation extremes in the Alps with a focus on northwest Italy. – Nat. Hazards Earth Syst. Sci. **13**, 1457–1468. DOI:10.5194/nhess-13-1457-2013.

Tustison, B., D. Harris, E. Foufoula-Georgiou, 2001: Scale issues in verification of precipitation forecasts. – - J. Geophys. Res. **106**, 11775–11784.

Tveito, O.E., I. Bjørdal, A.O. Skjelvåg, B. Aune, 2005: A GIS-based agro-ecological decision system based on gridded climatology. – Meteor. Appl. **12**, 57–68.

Undén, P., L. Rontu, H. Järvinen, P. Lynch, J. Calvo, G. Cats, J. Cuxart, K. Eerola, K. Fortelius, J.A. Garcia-Moya, C. Jones, G. Lenderink, A. McDonald, R. McGrath, B. Navascues, N.W. Nielsen, V. Ødegaard, E. Rodrigues, M. Rummukainen, R. Rõõm, K. Sattler, B.H. Sass, H. Savijärvi, B.W. Schreuer, R. Sigg, H. The, S. Tijm, 2002: HIRLAM-5 Scientific Documentation. – Hirlam, scientific report. http://www.hirlam.org/

Vidal, J.-P., E. Marin, L. Franchistéguy, M. Baillon, J.-M. Soubeyroux, 2010 : A 50-year high-resolution atmospheric reanalysis over France with the Safran system. – Int. J. Climatol. **30**, 1627–1644.

Vidale, P.L., D. Lüthi, C. Frei, S. Seneviratne, C. Schär, 2003: Predictability and uncertainty in a Regional Climate Model. – J. Geophys. Res. **108**, 4586. DOI:10.1029/2002JD002810.

Weingartner, R., D. Viviroli, B. Schädler, 2007: Water resources in mountain regions: A methodological approach to assess the water balance in a highland-lowland-system. – Hydrol. Process. **21**, 578–585.

Willmott, C.J., C.M. Rowe, W.D. Philpot, 1985: Small-Scale Climate Maps: A Sensitivity Analysis of Some Common Assumptions Associated with Grid-point Interpolation and Contouring. – Am. Cartographer **12**, 5–16.

Yang, D.Q., E. Elomaa, A. Tuominen, A. Aaltonen, B. Goodison, T. Gunther, V. Golubev, B. Sevruk, H. Madsen, J. Milkovic, 1999: Wind-induced precipitation undercatch of the Hellmann gauges. – Nord. Hydrol. **30**, 57–80.

Zappa, M., 2008: Objective quantitative spatial verification of distributed snow cover simulations – an experiment for entire Switzerland. – Hydrolog. Sci. J. **53**, 179–191. DOI:10.1623/hysj.53.1.179

Zolina, O., A. Kapala, C. Simmer, S.K. Gulev, 2004: Analysis of extreme precipitation over Europe from different reanalyses: a comparative assessment. – Glob. Planet. Chang. **44**, 129–161.

Zolina, O., C. Simmer, A. Kapala, S. Bachner, S. Gulev, H. Maechel, 2008: Seasonally dependent changes of precipitation extremes over Germany since 1950 from a very dense observational network. – J. Geophys. Res. **113**, D06110. DOI:10.1029/2007JD008393.