Clim Dyn (2010) 35:841–858
DOI 10.1007/s00382-009-0698-1

The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data

Nynke Hofstra · Mark New · Carol McSweeney

Received: 6 February 2009 / Accepted: 22 October 2009 / Published online: 8 November 2009

Abstract We study the influence of station network density on the distributions and trends in indices of area-average daily precipitation and temperature in the E-OBS high resolution gridded dataset of daily climate over Europe, which was produced with the primary purpose of Regional Climate Model evaluation. Area averages can only be determined with reasonable accuracy from a sufficiently large number of stations within a grid-box. However, the station network on which E-OBS is based comprises only 2,316 stations, spread unevenly across approximately 18,000 0.22° grid-boxes. Consequently, grid-box data in E-OBS are derived through interpolation of stations up to 500 km distant, with the distance of stations that contribute significantly to any grid-box value increasing in areas with lower station density. Since more dispersed stations have less shared variance, the resultant interpolated values are likely to be over-smoothed, and extreme daily values even more so. We perform an experiment over five E-OBS grid boxes for precipitation and temperature that have a sufficiently dense local station network to enable a reasonable estimate of the area-average. We then create a series of randomly selected station sub-networks ranging in size from four to all stations within the E-OBS interpolation search radii. For each sub-network realisation, we estimate the grid-box average applying the same interpolation methodology as used for E-OBS, and then evaluate the effect of network density on the distribution of daily values, as well as trends in extremes indices. The results show that when fewer stations have been used for the interpolation, both precipitation and temperature are over-smoothed, leading to a strong tendency for interpolated daily values to be reduced relative to the "true" area-average. The smoothing is greatest for higher percentiles, and therefore has a disproportionate effect on extremes and any derived extremes indices. For many regions of the E-OBS dataset, the station density is sufficiently low to expect this smoothing effect to be significant and this should be borne in mind by any users of the E-OBS dataset.

N. Hofstra · M. New · C. McSweeney
School of Geography and the Environment,
University of Oxford, South Parks Road,
Oxford OX1 3QY, UK

Present Address:
N. Hofstra (✉)
Environmental Systems Analysis Group,
Wageningen University, P.O. Box 47,
6700 AA Wageningen, The Netherlands
e-mail: nynke.hofstra@wur.nl

1 Introduction

Gridded climate data are important for many reasons, including the evaluation of climate model outputs and detection of trends in mean climate and climate extremes (e.g., Haylock et al. 2008). While remotely sensed products, reanalysis, station and gridded station data have all been used for these purposes, the first three have several disadvantages. Satellite and, for precipitation, radar data have complete coverage, but can have significant spatio-temporal biases and only cover a relatively short period of time (Gerstner and Heinemann 2008; New et al. 2001; Reynolds 1988). Reanalysis data, which are derived from numerical weather prediction hindcasts with assimilated observations, have the advantage that they are gridded and readily available (Hanson et al. 2007), but the data are only comparable to observations from stations after 1979 in the case of temperature, because during late 1978 the observing system used for reanalysis data improved strongly,

with, among others, better satellite temperature and humidity soundings (Simmons et al. 2004). For precipitation, reanalyses exhibit large errors and systematic biases. Indeed, Hanson et al. (2007) find that reanalysis data underestimate precipitation and temperature extremes significantly. Nonetheless, reanalysis data have been used for regional climate model (RCM) evaluation (e.g. in Kjellström and Ruosteenoja 2007 in an analysis of RCM performance in the Baltic Sea region). The direct observations of climate which are most readily available are from meteorological stations, but there is a mismatch of scale between 'point' station observations and the areal average output of climate models (Chen and Knutson 2008).

Gridded data derived from interpolation or area-averaging of station observations are often used to overcome the scale mismatch between climate models and station observations. In some areas the station density is so high that only stations within each grid box are used to estimate the grid-box area-average. Some studies use these dense areas for the evaluation of RCMs (e.g., Beniston et al. 2007; Buonomo et al. 2007; Huntingford et al. 2003; Jones and Reid 2001; Semmler and Jacob 2004), while for other areas, RCMs have been evaluated using gridded data developed with much sparser station networks. An example of climate model evaluation with gridded data developed with sparse station networks is the study by Christidis et al. (2005), who use the gridded global daily temperature dataset developed by Caesar et al. (2006) for the evaluation of their general circulation model (GCM). Until recently, European-wide high-resolution daily gridded data for Europe did not exist, despite a need for such data, as explained by Santos et al. (2007), who were forced to use reanalysis and station data for their evaluation of precipitation outputs of GCMs.

Haylock et al. (2008) describe a new European high-resolution gridded dataset of daily surface temperature and precipitation for 1950–2006 that has been developed as part of the EU-funded ENSEMBLES project, termed the E-OBS dataset. One of the primary purposes of E-OBS has been to facilitate the evaluation of RCMs used in the ENSEMBLES project (Hewitt and Griggs 2004), with a particular emphasis on assessment of the ability of RCMs to simulated daily climate and daily extremes. E-OBS is available on at four resolutions; 0.5° and 0.25° regular longitude–latitude grids, and 0.44° and 0.22° rotated-pole grids (Haylock et al. 2008). These correspond to the RCM resolutions defined for comparison and evaluation purposes by the ENSEMBLES project; therefore E-OBS is designed to permit comparison between model and observational data at the RCM native resolution.

The aim of the E-OBS data is to represent the daily areal average in each grid-box, which is equal to the average of a sufficiently large number of stations within the grid-box. Since outputs of RCMs are generally considered to be areal averages (Chen and Knutson 2008; Osborn and Hulme 1997) they would then be comparable to the E-OBS area-average estimates. However, the station network used to develop E-OBS is variable in space and time (Haylock et al. 2008). In addition, the station network is relatively sparse, comprising only 2,316 stations over Europe. While this is an increase of an order of magnitude compared to the data availability before the ENSEMBLES project started (Klok and Klein Tank 2009), there are approximately 18,000 0.22° grid-boxes over the European land area, so only a small fraction of grid boxes contain even one station. These two issues (variability of the density of the station network and sparseness of the network) can potentially affect the resultant grid box area-average estimates, in three main ways.

The first possible influence concerns the temporal variance shared between stations that contribute to a grid-box estimate, and which is retained in the areal average. When fewer than a sufficiently large number of stations within the grid-box are used to estimate the average, the variance of the area-average is likely to be larger than the true variance, and the estimate will not be a "true" areal average, but something in between a point estimate and an areal average.

The second potential effect on the areal average results from the practice of using stations outside a grid-box for the estimation of the areal average. Stations further away from each other have a lower shared variance (e.g., Hofstra and New 2009; Osborn and Hulme 1997). When stations outside a grid-box are used to estimate the grid-box areal average, this can result in reduced variance compared to the variance of the true grid box average. Although most interpolation methods apply a higher weight to stations closer to the grid-box, this "smoothing" will be less marked for grid-boxes with stations that are nearby, but a further consequence is that the degree of variance smoothing changes with station density across the grid domain.

The third potential effect is that over- or under-smoothing of the variance implies that daily extremes are influenced more than the mean. This is important if the data are being used to evaluate the ability of RCMs to simulate daily extremes, and could also result in an underestimation of trends in indices of extremes, such as that undertaken by Alexander et al. (2006).

All three of these effects are expected to be more important for precipitation than for temperature, because precipitation is a discontinuous variable in space and time. However, for E-OBS, the station density is lower for temperature than for precipitation, so temperature may be more affected than expected. In the case of precipitation, some studies have described methods to obtain "true" areal average information, either as parameters of their probability distribution (McSweeney 2007) or as return values

(Booij 2002; Fowler et al. 2005), which are then used for evaluation of GCM or RCM outputs. However, we are unaware of any studies that focus on the scaling issues in gridded station data themselves.

Haylock et al. (2008) briefly explored some of these scale issues in a comparison of extremes in stations and E-OBS. They show that the reduction factor (the proportional decrease in the return value for precipitation) and the reduction anomaly (the proportional decrease in the anomaly for maximum temperature) decrease in all extremes higher than the 75th percentile for precipitation and the 90th percentile for maximum temperature, and that these reduction factors are variable in space. Haylock et al. (2008) conclude that the interpolation methodology smoothes the intensity of the extremes and that the E-OBS data should therefore be suitable for evaluation of RCMs, but they do not evaluate the extent to which over- or under-smoothing occurs.

To improve the understanding of issues of scale within E-OBS, the objective of this study is to assess the extent to which the interpolation procedure used for E-OBS produces true areal averages. We focus on the influence of station network density on (1) the shape of the distribution of gridded daily climate variables and (2) the size and trends of extreme indices calculated from the gridded data. In the sections that follow, we first describe the general method with which we assess the influence of station density, and then we describe the results for the distribution and extremes.

2 Data and methods

As we do not have true area-averages to evaluate against, we select grid boxes from E-OBS that contain or are adjacent to a sufficiently large number of precipitation and temperature stations to obtain a reasonable estimate of the true area-average. Since there are approximately 18,000 0.22° land grid-boxes in E-OBS and only 2,316 stations, there are only five grid boxes that have sufficient stations for our purposes for precipitation and temperature, respectively. Table 1 and Fig. 1 give an overview of the location of the grid-boxes, which are different for precipitation and temperature, and show the stations in or nearby the grid-boxes used to calculate area-averages, as well as all stations within the search radius used in the interpolation, which is 450 km for precipitation and 500 km for temperature.

We use the rotated pole grid for our evaluation, using both 0.22° and 0.44° grid-boxes so we can explore any differences in smoothing at these different scales (e.g., Chen and Knutson 2008).

To evaluate the effect of different station network densities on the gridded estimates, we randomly sample from the full station network within a grid-box search radius to create "sub-networks" of different size, ranging from four up to 250 stations (henceforth termed INT-4 to INT-250, although not all grid boxes have as many as 250 stations within the search radius). For each sub-network size, one hundred random sub-samples are created. We use each sub-network to estimate the grid-box areal average using the same approach as that used for E-OBS (see below). We also calculate an interpolated estimate based on the full station network (INT-ALL) and a simple average based on the stations within, or very close to, the grid-box, which serve as our estimate of the true spatial average (termed AVG). As we use the grid-boxes with the highest station density, AVG is the closest we can get to a true areal average.

We fit the gamma and Gaussian distribution to the realisations of interpolated daily values so we can evaluate the influence of network density on the full distribution of daily precipitation and temperature, respectively. Similarly, to evaluate the influence of network density and

Table 1 Details of the selected grid-boxes

	Latitude	Longitude	Stations within the 0.22° grid-box	Extra stations within the 0.44° grid-box	Stations within the search radius
Precipitation					
1. Italy	45.908	11.144	11	8	172
2. South Ireland	51.766	−8.761	9	7	258
3. East Ireland	53.185	−6.533	8	11	314
4. Luxembourg	49.697	6.139	7	3	252
5. Netherlands	52.555	5.748	5	2	207
Minimum/maximum temperature (first number before the/is for minimum temperature, number behind the/is for maximum temperature)					
1. Alps	47.132	9.678	4	1/2	139/152
2. Netherlands	52.047	5.160	3	0	147/146
3. Birmingham, UK	52.422	−1.967	3	1	108/106
4. Edinburgh, UK	55.868	−3.252	3	1	156/153
5. Ireland	52.047	5.160	3	0	123/121

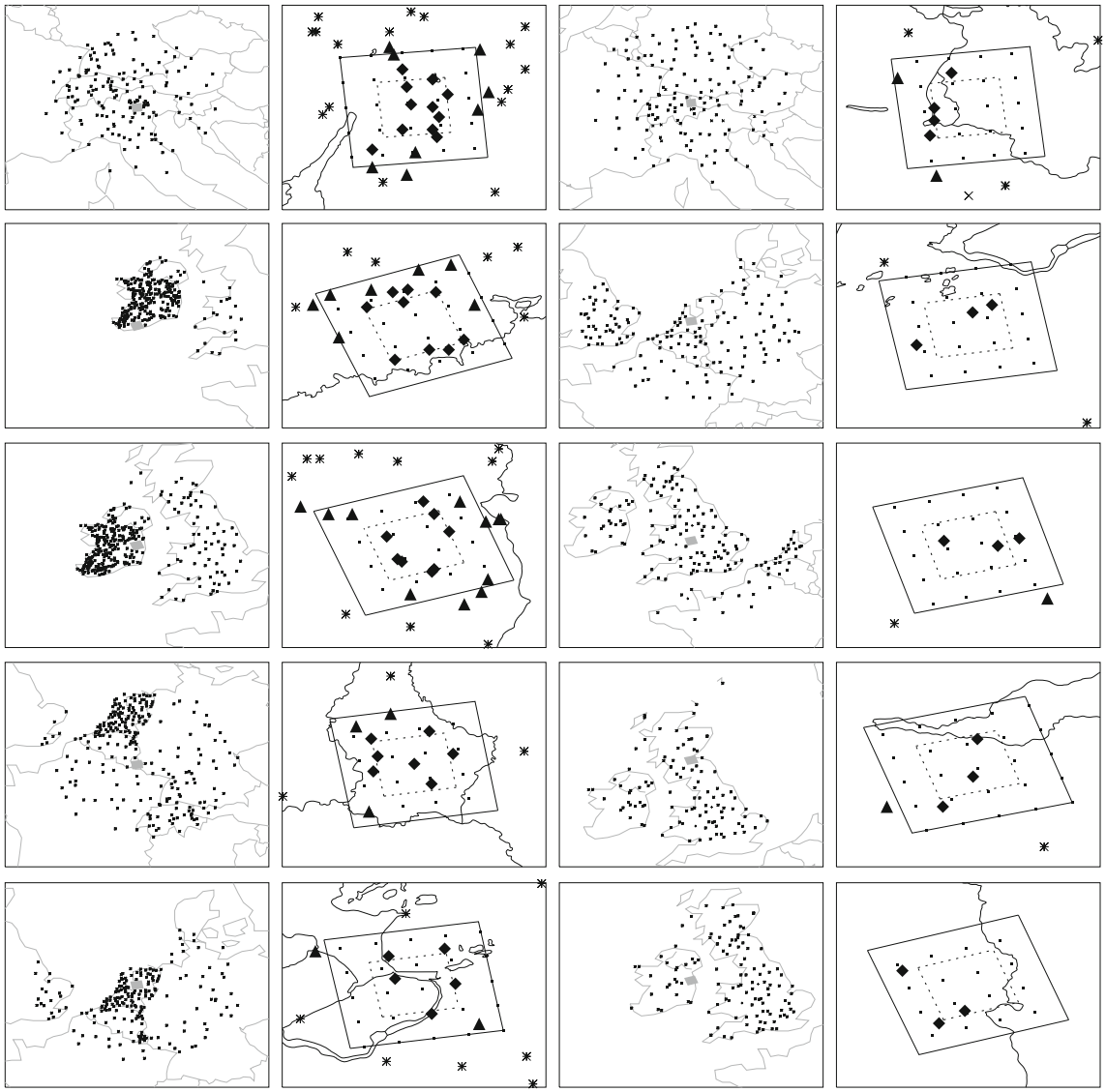


Fig. 1 Stations (*asterisks*) within the search radius (*first* and *third columns*) and within or nearby the grid-box (*second* and *fourth* columns) for five different grid-boxes for precipitation (*first* and *second column*) and five different grid-boxes for temperature (*third* and *fourth column*). These grid-boxes correspond to those in Table 1. Grid boxes in the first and third columns are highlighted by *grey quadrangles* in the middle of the figures. In columns 2 and 4, the *dotted quadrangles* are the 0.22° grid-boxes, the *solid squares* the 0.44° grid-boxes. The *diamonds* are the stations within or nearby the 0.22 degree grid, *triangles* the stations within or nearby the 0.44° grid-box. *Crosses* are stations that do have a value for maximum temperature, but not for minimum temperature and plusses do have a value for minimum temperature, but not for maximum temperature. The *small squares* show the 0.1° grids within the grid-boxes

interpolation on over- or under-smoothing of extremes, we calculate a range of extreme indices for each realisation.

In the remainder of this section we describe the station data, explain the E-OBS interpolation method and fitting of distributions, and finally define the extremes indices that we evaluate.

2.1 Station data

We make use of the same station dataset that underpins the E-OBS gridded data, which were collated by the Royal Netherlands Meteorological Institute (KNMI) in collaboration with over 50 partners from European countries, as part of the European Climate Assessment and Data (ECA&D) (Klok and Klein Tank 2009). In total, there are 2,316 stations, most of which have precipitation data, with a smaller fraction having both precipitation and temperature data. For our study, we only use station data that fall within the interpolation search radii of the grid-boxes, and that also have less than 31% missing data (Fig. 1). The density of the station network is variable in space and time and by using stations with more than 69%

data we ensure that even if only four stations are selected, there will be sufficient data available for interpolation on a large proportion of days and that any analysis is not overly affected by sub-sampling from a changing master network. The data have been subject to an automated quality control by KNMI (Klok and Klein Tank 2009) so most potentially erroneous outliers have been removed. However, in the process of our analysis, we discovered some problems with potential duplicate stations in Ireland and Italy, where a few stations with slightly different latitude and longitude have the same values for most of the time. These issues have been reported to KNMI and they will be addressed before the next version of E-OBS. We use the full period of the dataset, 1950–2006 for our analysis, except when we assess the effect network density and smoothing on trends in extremes, where we use the 1961–1990 period.

2.2 Interpolation

We employ the interpolation procedure used to produce E-OBS (Haylock et al. 2008), but apply it to each of the realisations of station networks of different density. For any one sub-network realisation, the 15 closest stations are selected. If there are fewer than 15 stations in the search radius (450 and 500 km for precipitation and temperature, respectively) all stations are selected. If there a fewer than four stations available, the interpolation is not undertaken, and the grid value for that day is set to missing.

For the E-OBS dataset, daily values are interpolated in a three-phase process (Haylock et al. 2008). Monthly values are first interpolated using thin-plate splines onto a 0.1° grid. Thereafter daily anomalies are interpolated separately onto the same 0.1° grid using ordinary kriging for temperature, or for precipitation, indicator kriging (Barancourt and Creutin 1992) to determine the state (wet or dry), followed by ordinary kriging for precipitation amount at locations whose state is wet. The 0.1° monthly and daily anomaly fields are then combined and averaged to obtain area-average estimates for the E-OBS 0.25 (0.22) and 0.5 (0.44) cartesian (rotated pole) grids.

For this study, we only interpolate daily anomalies from the multiple sub-networks, as it has been shown elsewhere (Haylock et al. 2008) that any smoothing of gridded daily values and extremes occurs at this stage of the interpolation rather than after combination with monthly values. We employ the global variograms defined for E-OBS rather than defining a new variogram for this test case. Each realisation of daily anomalies is then combined with the existing E-OBS monthly fields and 0.22° and 0.44° area-averages are calculated, to enable distributions to be fitted and extremes indices to be computed. Results for the 0.25° and 0.50° grids are virtually identical to the 0.22° and 0.44° grids, so we only report results for the latter here.

2.3 Distributions

To quantify the influence of interpolation under different station densities on the distribution of our climate variables, we fit the gamma distribution to precipitation data and Gaussian distributions to the temperature data. The gamma distribution is fitted using the Thom (1958) maximum likelihood method, after which the shape (α) and scale (β) parameters are used to obtain the mean ($\alpha \times \beta$) and the variance ($\alpha \times \beta^2$) of the distribution. Since the gamma distribution is only fitted for days with precipitation (>0.5 mm), we also calculate the dry day probability. For temperature we use the method of moments (Wilks 2006) to fit the Gaussian distribution to the anomalies from the monthly mean (to remove the seasonal cycle of temperature), and then analyse the variance (for temperature, the mean is not strongly affected—not shown). To determine how well the chosen distributions fit the data, we use quantile–quantile (Q–Q) plots. For precipitation, minimum temperature and maximum temperature for each grid-box we have plotted Q–Q plots for selected time series (not shown). For precipitation the fitted gamma distribution fits the data of the stations better than the interpolated data or the average of the station data within the grid-boxes. In many cases of the former the distribution fits the data well until at least the 99th percentile. However, in the case of the gridded data the fitted distribution begins to diverge from the data from the 90th percentile onwards, which may be due to possible smoothing of the interpolated data. However, the deviation from the fitted distribution is so small that we assume that the gamma distribution fits well until at least the 95th percentile. For temperature the results for maximum temperature are generally slightly better than for minimum temperature. In all cases the distribution fits the data well until at least the 95th percentile.

2.4 Extreme indices and trends

To investigate the influence of interpolation and station density on the extremes of precipitation and temperature and on the trends in these extremes, we calculate a subset of the standard extremes indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI):[1]

- R20 mm: the annual number of days with daily precipitation larger than or equal to 20 mm

[1] ETCCDI meets under the auspices of the joint WMO Commission for Climatology, CLIVAR, Joint-WMO-IOC Technical Commission for Oceanography and Marine Meteorology.

- RX1 day: the annual maximum precipitation on one single day
- R95p: the sum of the precipitation on all days on which precipitation is larger than the 95th percentile
- TXx: the annual maximum of the daily maximum temperature
- TNx: the annual maximum of the daily minimum temperature
- SU: summer days, the annual number of days on which the daily maximum temperature is above 25°C
- FD: frost days, the annual number of days on which the daily minimum temperature is below 0°C

These indices are relatively easy to interpret compared to some of the other indices, and also test a range of generic types of indices: percentiles, annual maxima and number of days above or below a threshold. All indices are determined using the FClimDex programme (available from http://cccma.seos.uvic.ca/ETCCDMI/).

A number of approaches for trend analysis of extremes have been used. Linear regressions trends have been used by e.g. Klein Tank and Können (2003) and Groisman et al. (2005). Alexander et al. (2006) use the nonparametric Kendall's tau based slope estimator, which is considered to be less sensitive to outliers. Here we employ the more widely used ordinary least-squares linear regression approach. We use the period 1961–1990 because the station network density is highest in this period. Using different periods will give different trends, but we are not interested in the trends as such, but in the differences in trend between the interpolations based on different network densities, so any single analysis period will serve to elucidate these differences; similarly, an alternative more robust trend estimator could have been used, and different analyses periods, these factors are not critical to our objectives.

Inhomogeneous data could result in biased trends. Many stations used for the E-OBS interpolation, as well as the E-OBS dataset itself, have potential inhomogeneities (Hofstra et al. 2009). However, since in this study we are only interested in the influence of interpolation and station density on the extremes and on trends in these extremes, any inhomogeneities will not affect our specific aims. However, the values of calculated trends from this study should not be seen as real trends for certain areas, as they may have been influenced by inhomogeneities.

3 Results

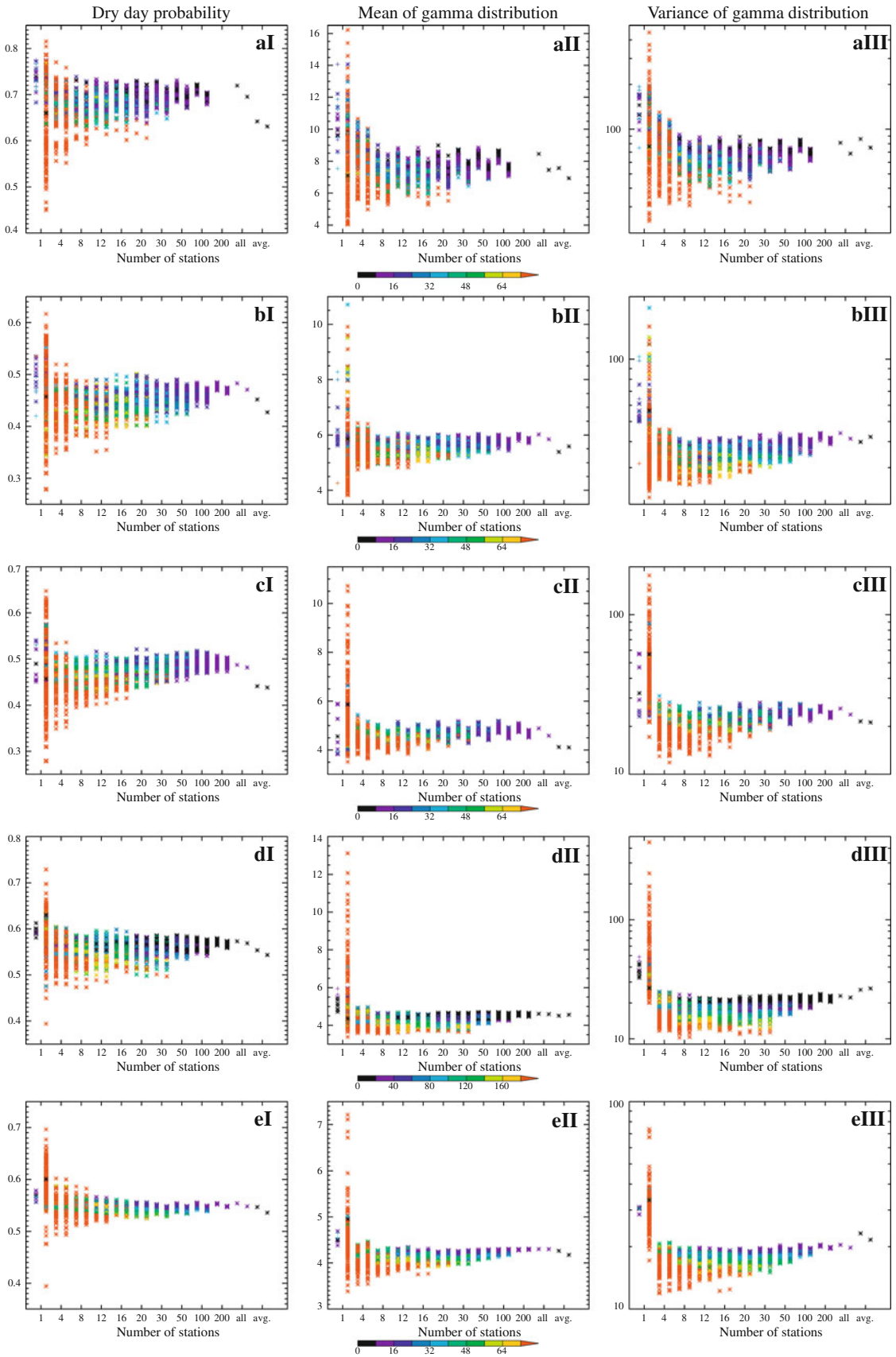
3.1 Distribution of daily estimates

Figure 2 shows, for precipitation, the dependence on station density of dry day probability, mean and variance of the gamma distribution for wet-day precipitation amount for the five grid-boxes analysed. We also show results for (1) each individual station, (2) an interpolation based on all available stations, as would occur in E-OBS, and (3) the simple average calculated using the stations within and adjacent to the grid box. We note that the patterns for seasonal data (not shown) are the same (though with different absolute values) as those for annual data shown in Fig. 2.

Fig. 2 Dry day probability (**I**) and mean (**II**) and variance (**III**) of the gamma distribution for the grid-boxes in Italy (**a**), South Ireland (**b**), East Ireland (**c**), Luxembourg (**d**) and the Netherlands (**e**). The *colour* represents the mean distance of the closest two stations that contributed to the interpolation to the centre of the grid-boxes. For each plot the first column represents the stations within the 0.22° grid-box (*asterisks*) and within the 0.44° grid-box (*plus symbol*). In the right part of the first column all stations within the search radius are represented. The second to eleventh column pairs represent all combinations of stations for INT-4 to INT-ALL, column pair 12 is AVG. The left of the column combinations is the value for the 0.22° grid, right for the 0.44° grid-box. Missing values for grid-boxes **a** and **e** are due to lack of stations within the search radius ▶

A similar general pattern emerges for each test case, where the spread in parameters is largest for small networks, narrowing and approaching the estimated area-average as the number of stations available to sample from increases. As might be expected, the spread of parameters at individual stations is even larger. The spread of the interpolated realisations is mainly directed towards lower values than the area-average, indicating that the tendency is for both the mean and variance of daily precipitation to be underestimated, and that the dry-day probability is also underestimated. We also show (using colour) the average distance of the two nearest stations used in the interpolation; for any sub-network size, when the nearest stations are more dispersed, the degree in mismatch between the interpolated and area-average values tends to be larger. This indicates that less dense networks increase the likelihood that more distant stations will be influencing the interpolation, and that the interpolation of daily values are smoothed to lower values.

In all the precipitation grids, the dry day probability of AVG is lower than that of INT-ALL, and also the realisations with smaller networks that have nearby stations, contrary to the expectation that use of additional stations outside the grid box during interpolation would increase the number of interpolation days with rainfall. A possible reason for this relates to the use of indicator kriging to estimate the dry/wet state at each 0.1° grid point prior to averaging. This could result in a greater dry day probability on the 0.22° and 0.44° grids overall compared to the simple averaging used to estimate AVG. Also, the use of a global variogram and a wet/dry threshold may result in biases at any single grid box.


Dry day probability
Mean of gamma distribution
Variance of gamma distribution
aI
aII
aIII
bI
bII
bIII
cI
cII
cIII
dI
dII
dIII
eI
eII
eIII
Number of stations
1 4 8 12 16 20 30 50 100 200 all avg.
0 16 32 48 64
0 40 80 120 160

The mean and variance of AVG are noticeably different from those for INT-ALL in some of the grid boxes analysed here. In three cases, the mean of AVG is less than INT-ALL (with the remaining two being similar) and in four cases the variance of AVG is larger. With only five grid boxes analysed it is difficult to arrive at a firm conclusion as to whether this tendency of reduced (elevated) variance (mean) is a pervasive feature of the full E-OBS interpolated dataset. However, Hofstra et al. (2009) shows that the E-OBS data have a negative bias over a large part of the domain when they are compared to gridded data that have been developed with much denser station networks and are deemed to be closer to "true" areal averages. For the variance, the reduced variance for interpolation is expected, because stations from relatively far away have been used for the interpolation and these stations do not share as much variance as the nearby stations used to estimate AVG.

Another general pattern in Fig. 2 is that there is a reduction in all three parameters in the 0.44° grid compared to the 0.22° grid. This reduction is expected, as a quadrupling of the grid area will result in less shared variance across the grid. The reduction is stronger in some grid-boxes, such as the Italian one, and only marginal in others, such as the one in Luxembourg.

Figure 3 shows the results for the same analysis applied to the variance of minimum and maximum temperature (mean temperature results are similar and not shown here). As for precipitation, the spread of interpolated averages is larger for smaller network sizes, and tends to be underestimated relative to AVG, indicating that the variance is smoothed. An exception is the Alpine grid box (Fig. 3a), where the spread for different network sizes straddles the AVG value; this suggests that in areas in complex terrain such as the Alps, the daily values can be over or undersmoothed depending on the specific stations used in the interpolation. At the Irish and Edinburgh grid boxes, the variance is increased for 0.44 grid boxes; this may be because the area covered by the larger boxes encompass more land away from the coast, where temperature variance would be expected to be higher.

We also briefly studied the influence of network density on interpolation to the individual 0.1° grid-points that are then used to create the 0.44° and 0.22° area-averages. The underlying rationale of the E-OBS scheme is that the 0.1° grid estimates approximate a network of stations, which are then averaged together to obtain the grid box area-average values. Ideally, one would therefore expect a reduction in variance when moving from individual 0.1° points to larger area-averages. However, if a network is sparse, it is likely that there will already be marked variance reduction at the 0.1° grid points. This was confirmed by Haylock et al. (2008), who find a reduction in the 10 year return level for precipitation in E-OBS at 0.1° grid points compared to the return periods in the underlying station data. For all our test grid boxes, we find a general reduction in variance at the 0.1° points as the station density decreases, apart from a few of the realisations with a network of only four stations; in these cases, the stations by chance have high shared variance and this, combined with the higher variance expected when only four stations are interpolated, is translated to the grid points. These effects are illustrated in Fig. 4, for south Ireland. Figure 4 also shows that the variance at the 0.1° points is similar to that of the 0.22° and 0.44° area-averages, suggesting that the majority of smoothing occurs during the interpolation to 0.1° grid points.

To evaluate how the differences in wet/dry probability, mean and variance explored in Figs. 2 and 3 translate into differences in actual precipitation and temperature, we plot percentiles from the fitted distributions (Figs. 5, 6). For each station, each interpolated time series and AVG we calculate the 5th, 25th, 50th, 75th and 95th percentiles, and for the stations and INT-4–INT-100 plot the variation in these percentiles across the 100 realisations as box-and-whisker plots. Broadly, the distribution of the percentiles reflect the same pattern as that observed for mean and variance in Figs. 2 and 3. INT-ALL is smaller in the 95th percentile than AVG in almost all cases, except for precipitation in both Irish grid-boxes (in the south of Ireland only the 0.22° resolution) for the reasons explained above.

The spread in precipitation and temperature is smaller in the less extreme percentiles. For smaller percentiles the difference between AVG and INT-ALL becomes smaller or even reverses. The spread also reduces when more stations are used for the interpolation. However, for precipitation this reduction is stronger than for temperature. The difference between the whiskers for the 95th precipitation percentile can be as large as 13 mm in the case of INT-4 of Italy. That is a difference of 51% from the mean 95th percentile of INT-4. For the other grid-boxes the difference is 15–31%. For temperature the difference between the 5th and 95th percentile of the 95th temperature anomaly percentile of INT-4 is around 1°C. In the case of minimum temperature this is around 23% of the 95th percentile of INT-4, in the case of maximum temperature around 20%.

3.2 Extreme trends

The effect of station network size and interpolation on trends in extreme indices is shown in Figs. 7 and 8 for precipitation and temperature, respectively. As with the previous analysis, the spread in the range of trends between realisations decreases as the sub-network density increases, due to the increased likelihood of any one of the larger sub-networks having more stations nearby the target grid box.

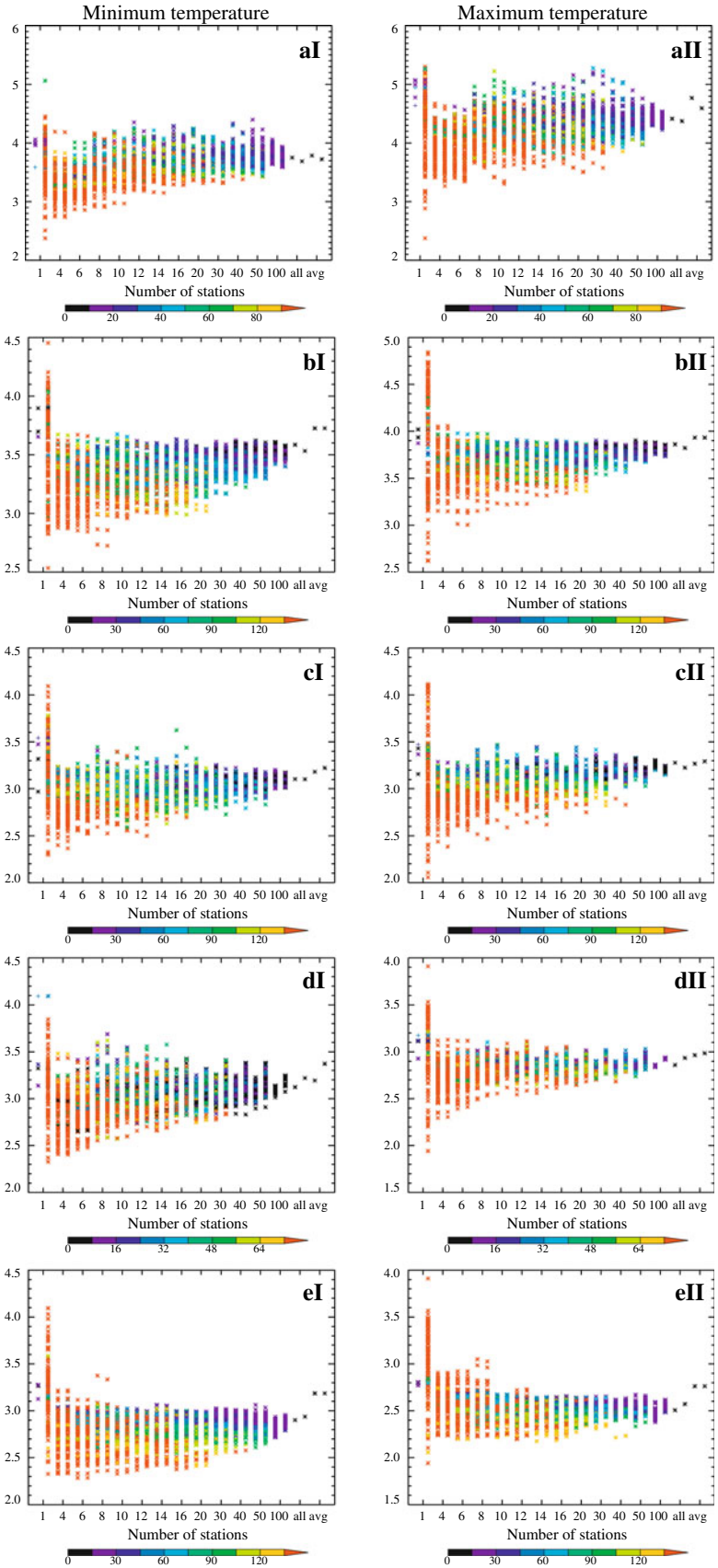


Fig. 3 Variance of minimum temperature (**I**) and maximum temperature (**II**) for grids in the Alps (**a**), the Netherlands (**b**), Birmingham (**c**), Edinburgh (**d**) and Ireland (**e**). On the *x*-axis are 15 column pairs. The left part of the first column depicts the stations in the 0.22° grid-box (*asterisks*) and 0.44° grid-box (*plus symbol*), the right part depicts all stations within the search radius. Columns 2–14 are INT-4–INT-ALL and column 15 is AVG. The left part of these columns shows the results for the 0.22° grid-box and the right part for the 0.44° grid-box. The *colours* represent the mean distance of the closest two stations that have contributed to the interpolation to the centre of the grid

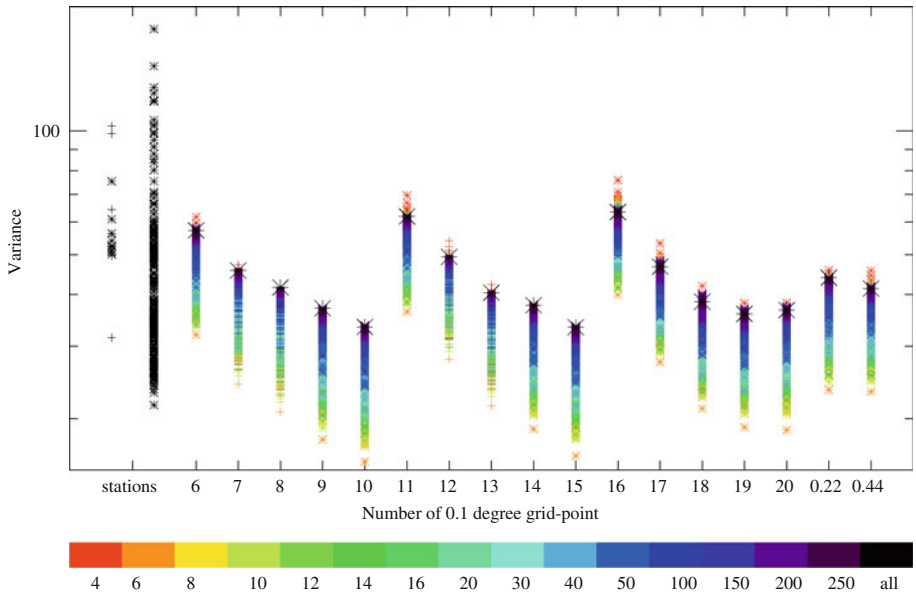


Fig. 4 Comparison of variance of precipitation at stations, when interpolated to 0.1° points, and at 0.22° and 0.44° area-averages, for the south Ireland example. The first two columns of the *x*-axis show the values for stations located within the 0.44 and 0.22 grid boxes. Columns numbered 6 through 20 represent 0.1° grid points (grid-box 1–5 have no data as they are over the sea, see Fig. 1). The last two columns are the area-averaged data for the 0.2 and 0.44 grid boxes. Grid points 7, 8, 12 and 13 contribute to the 0.22° grid-box and are denoted by *plus symbol*, and all other grid points are denoted by *asterisks*. Values arising from different sub-network densities (INT-4–INT-250) are represented by the colour (*red–black* for increasing density) and INT-ALL is depicted with the large *asterisks*

In our examples, the direction of the trend is mostly consistent between realisations, for example R20 mm for the Luxembourg grid boxes (Fig. 7dI). For this grid-box half of the stations within the search radius actually have a negative trend, but the trend is positive for almost all realisations of INT-4–INT-ALL and also at the local stations. However, for precipitation sometimes there are more differences in the trends, due to the differing trend in precipitation indices within the grid-boxes. For example, for RX1 day in Luxembourg (Fig. 7dIII) the interpolated data from sparse networks, until INT-50, mainly have positive trends, whereas the trend in all except for one station in the grid-box is negative.

Of note is that the trend in AVG is often smaller than the trend in INT-ALL. However, there are exceptions, such as R95p for east Ireland (Fig. 7cII). For temperature, differences between AVG and INT-ALL are generally small, suggesting that trends in temperature extremes in the E-OBS gridded have a higher likelihood of matching the true trends.

Figures 7 and 8 do not show a very clear influence of the average distance to the two closest stations on the differences in trend. For precipitation some trends for some grid-boxes do seem to show an influence (for example, RX1 day at the south Ireland and Luxembourg grid boxes and R95p at the east Ireland box). For other indices for other grids the results are not as pronounced. For temperature the influence of the average distance to the two closest stations seems slightly more apparent, mainly in FD and SU (Fig. 8aII, bI, bII, cI).

These results, especially for precipitation have two implications for the E-OBS dataset. First, as the E-OBS station network varies in density spatially, there is greater likelihood that trends in the gridded data in data-sparse areas will not be the true local trend. Second, as the network density in any geographical area tends to vary in time, trends in any one area may reflect the appearance and disappearance of specific stations, rather than the true local trend.

When we compare the trends in the 0.44° and 0.22° grid boxes, the larger grid boxes tend to have smaller trends, indicating that the bigger extremes are reduced more than more moderate ones when expressed as areal averages of the individual 0.1° points. This is illustrated in Fig. 9a, which shows the SU index for Birmingham. In years where the SU is relatively large, the index is smoothed more for the 0.44° areal average, resulting in a reduced trend compared to that for the 0.22° grid box. A similar pattern might be expected when comparing time series from individual stations and AVG in Fig. 9, but the difference in magnitude of the trends at individual stations within a grid-box is large, preventing any firm conclusions. Figure 9 also shows a different feature that is important for indices calculated as a count over a fixed threshold, such as SU (Fig. 9) and

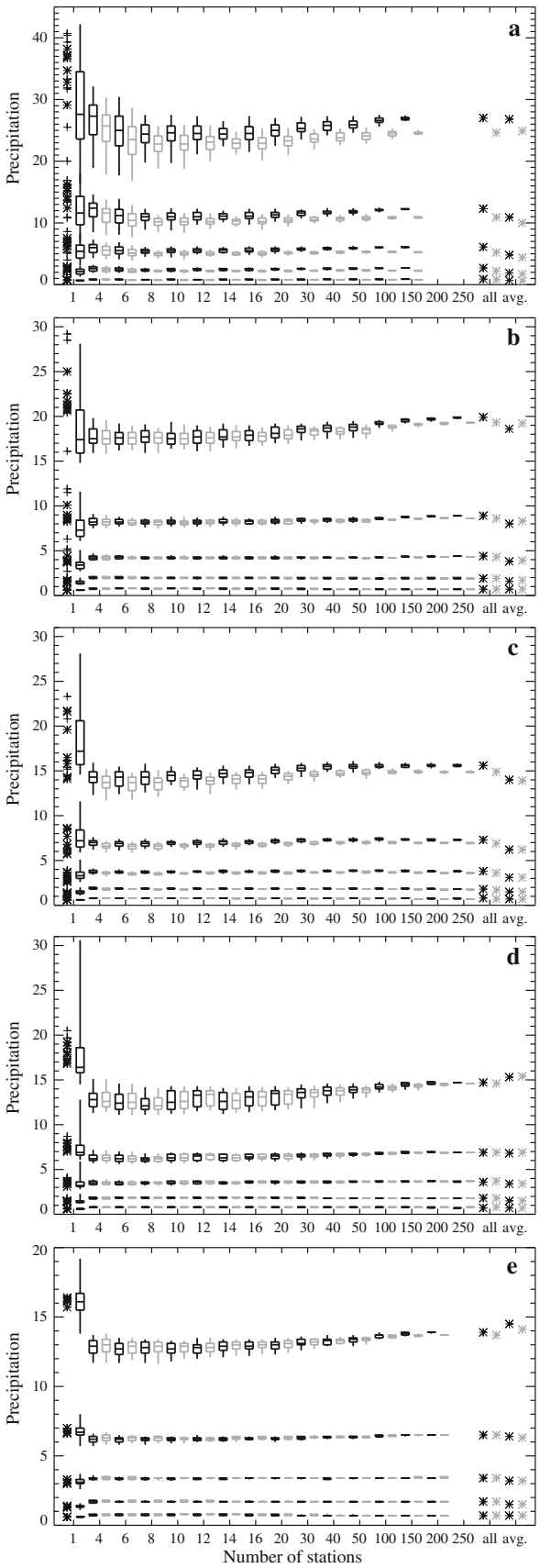


Fig. 5 Dependence on different percentiles of the fitted Gamma distribution for precipitation on network density. The five precipitation cases, **a**–**e**, are the same as in Fig. 2. The values for the 0.22° grid-boxes are *black* and the ones for the 0.44° grid-boxes *grey*. We show box plots for the 5th, 25th, 50th, 75th and 95th percentile. Each box plot shows the median and the 25th and 75th percentiles (*box*) and the 5th and 95th percentiles (*vertical lines*) for each 100 sub-network realisations. *Stars* in the left column are the percentiles for stations within the 0.44 and 0.22 gird boxes, and *stars* in the end columns are percentiles from the single realisations using INT-ALL and AVG

R20 mm (not illustrated). Area averages in years where the count at individual stations is low tend to have an index of zero, as the area average falls below the threshold. Therefore, if the variable of interest (e.g. the number of summer days for SU) shows an increase (or decrease) over time, years with a zero index will tend to be clustered near the beginning (end) of the time series, resulting in an exaggerated trend, compared to the station trends.

4 Summary and conclusions

In this paper we explored the extent to which interpolation of station networks of varying density over or under-smoothes gridded daily climate estimates, and also the extent to which trends in the indices of extreme climate are affected. This analysis was applied to both precipitation and maximum and minimum temperature. Our approach was to select 0.22° and 0.44° rotated pole grid boxes from the E-OBS gridded data that contain sufficient stations to reliably estimate the grid-box area-average. We then randomly selected 100 combinations of sub-networks of varying size, from 4 to 250 from the full set of E-OBS stations that fell within the interpolation search radius for each grid box. The selected sub-networks were then used to estimate grid-box area-average values using the same interpolation method used for the construction of the E-OBS gridded data product. In the first part of the study we then estimated and analysed the gamma distribution (in the case of precipitation) or Gaussian distribution (in the case of temperature) for all stations individually, each realisation of the interpolation under different sub-network densities, the interpolation using all stations and the simple average of the station series within the grid-box. In the second part of the study we calculate extremes indices and analyse the dependence of trends in these indices to variations in sub-network densities.

Our results show clearly that the majority of smoothing occurs at the first stage of the interpolation process—where fine scale grid point values are estimated using kriging—rather than at the stage of calculating a larger (0.5°, 0.44°, 0.25°, 0.22°) area-average. Previous work comparing the

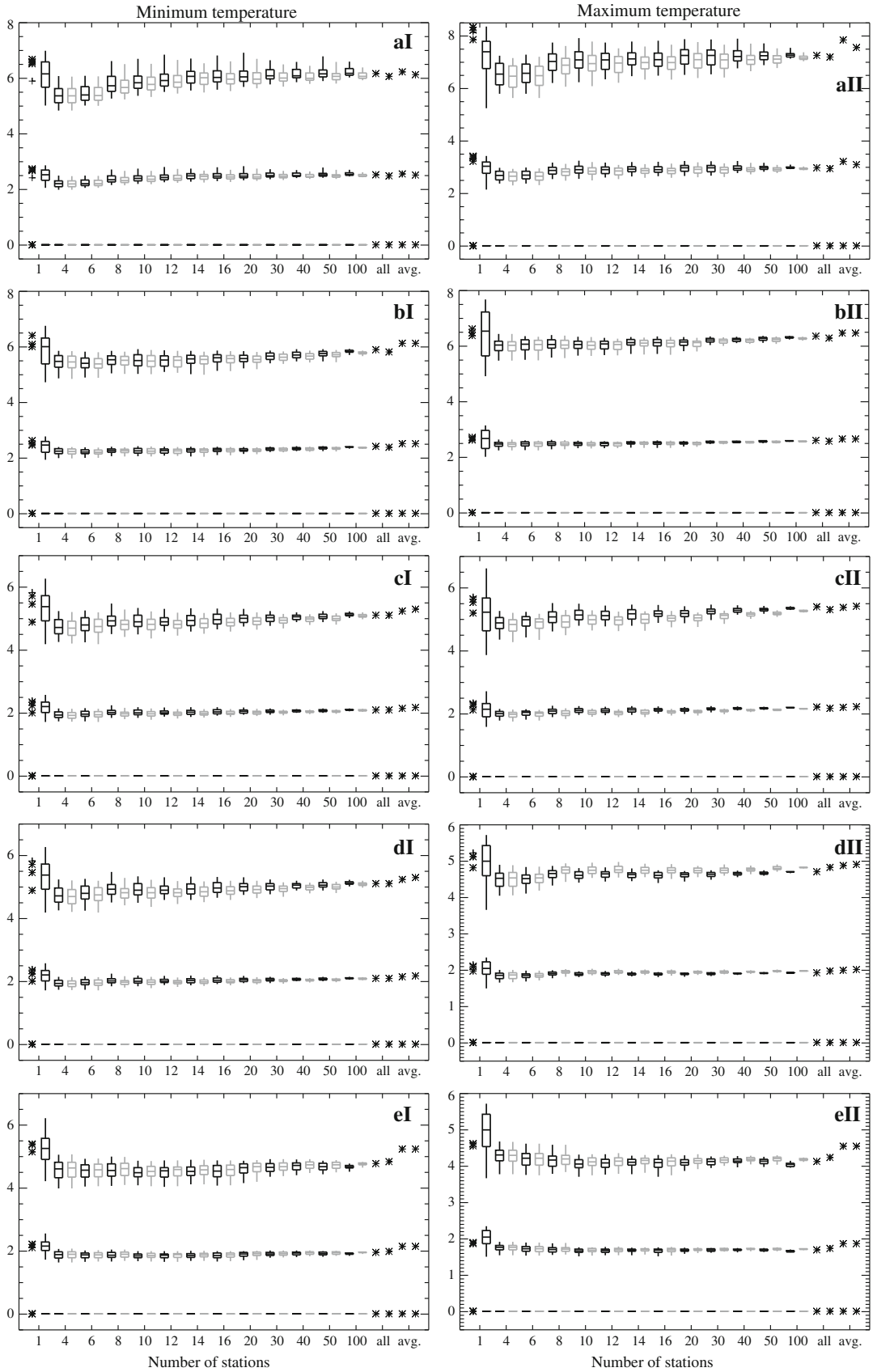


Fig. 6 Same as Fig. 5, but for minimum temperature (*left*) and maximum temperature (*right*) in the same order of the grids as in Fig. 3

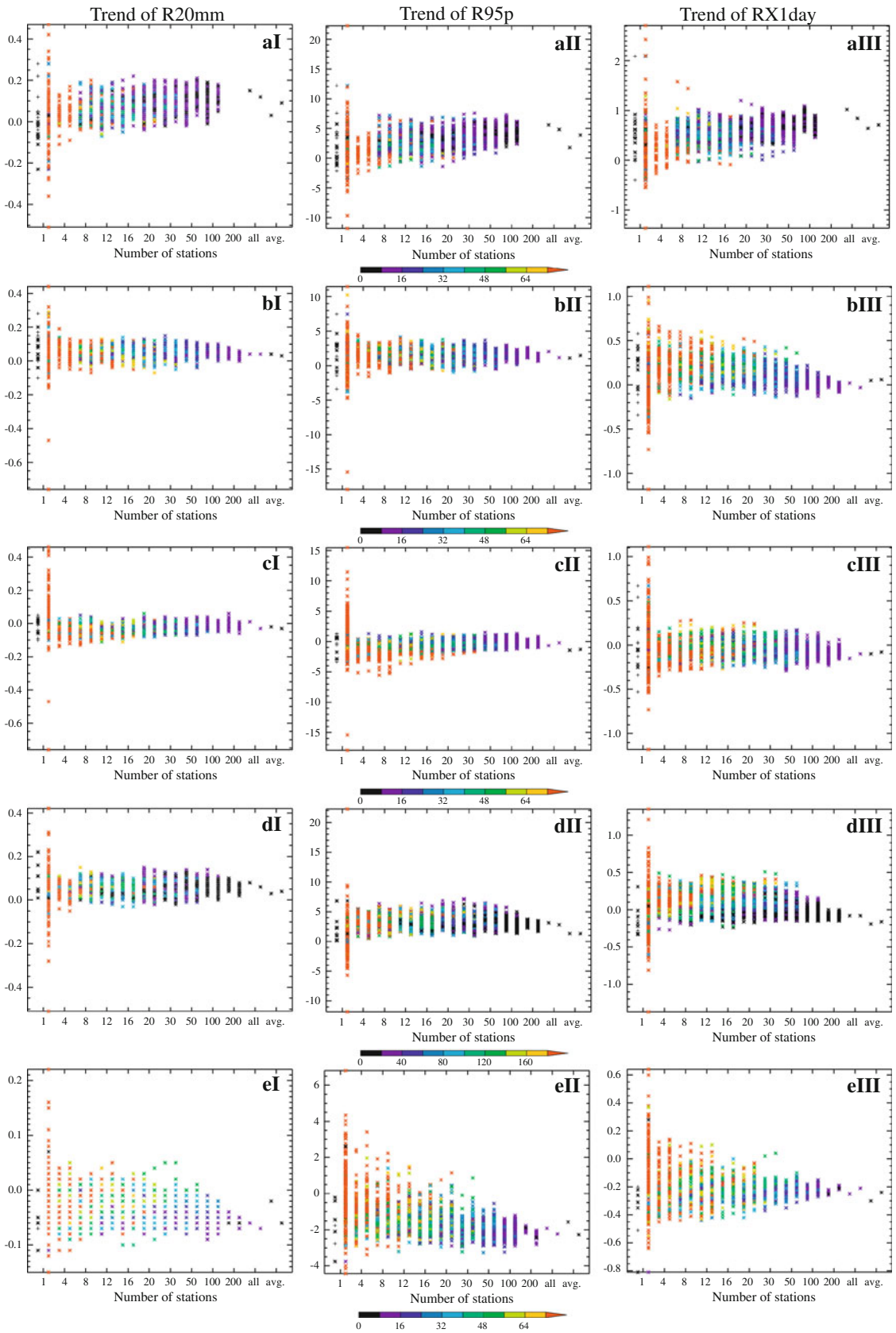


Fig. 7 Trend in the indices R20 mm (**I**), R95p (**II**) and RX1 day (**III**) for the same grid-boxes as Fig. 2. The *x*-axes and *colours* are also the same as in Fig. 2

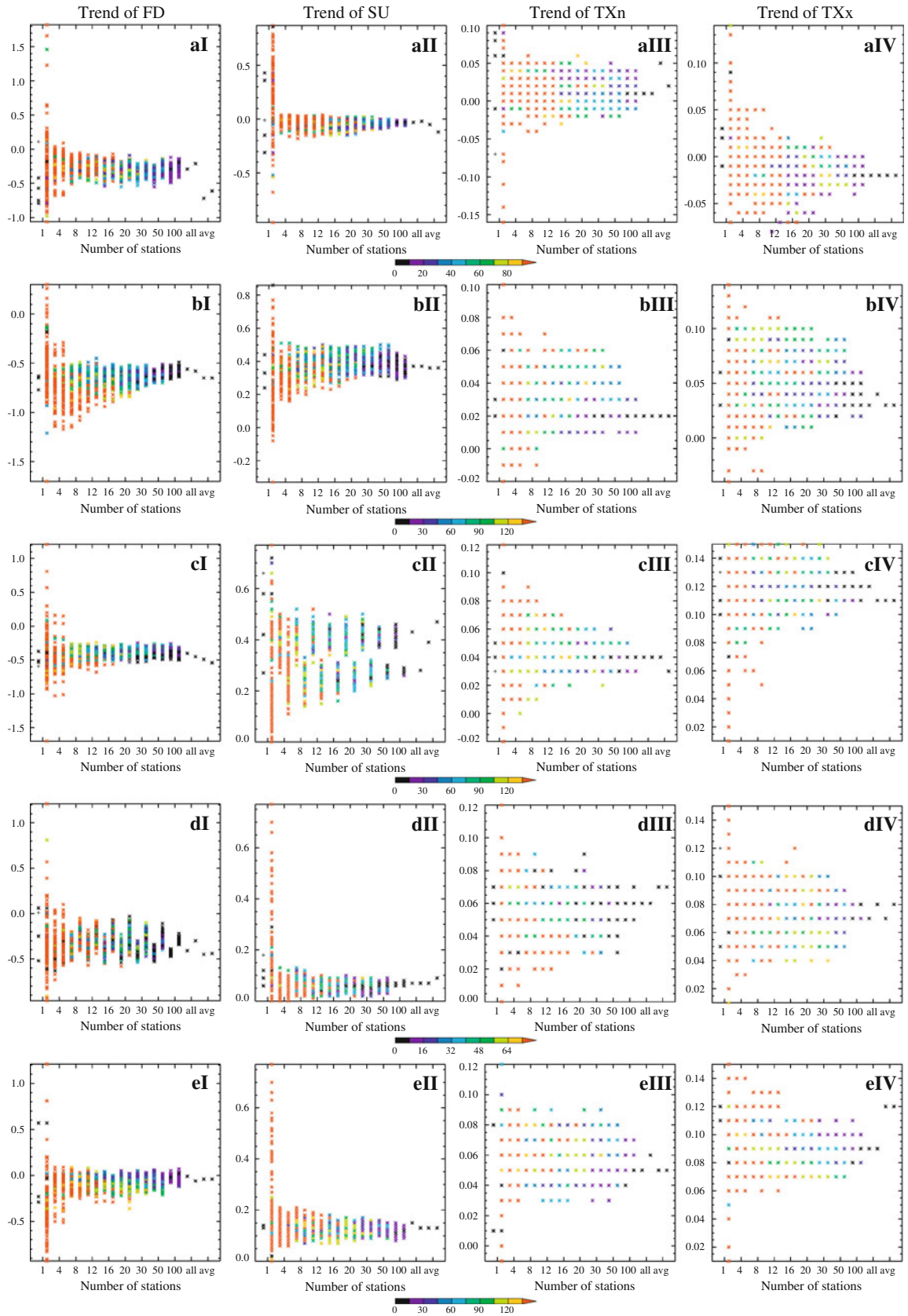


Fig. 8 See Fig. 7, but for the indices FD (**I**), SU (**II**), TNx (**III**) and TXx (**IV**) for the same grid-boxes as Fig. 3. Trends in TNx and TXx are so small that rounding of the numbers results in many realisations having the same values

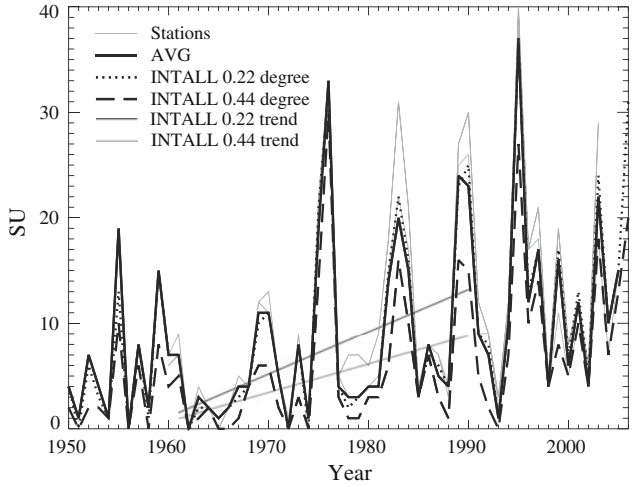


Fig. 9 The number of annual summer days in time for the grid-box close to Birmingham. *Thin grey lines* are individual station series, the thick black line is AVG, the *dotted black line* the series of the 0.22° grid-box for INT-ALL and the *thick dashed line* the series of the 0.44° grid-box for INT-ALL. *Medium-thick dark grey* and *grey lines* are the trend lines for the 0.22° and 0.44° grid-boxes for INT-ALL, respectively. Trend lines are estimated for the period 1961–1990 only

ability of several candidate interpolation schemes for the E-OBS dataset construction showed that the primary constraint on interpolation skill as station network density and the different interpolation schemes has very similar levels of skill (Hofstra et al. 2008). Thus we expect the dependence of smoothing on kriging to be a generic issue for all interpolation methods.

The results of the analysis of the distributions of daily values are similar for both precipitation and temperature. The variance and, in case of precipitation, also the mean of INT-ALL are generally smaller than AVG, which suggests over-smoothing due to the fact that in most cases stations from outside the grid-box—which have lower shared variance—have been used to estimate the interpolated areal average,. Only in case of the dry day probability there are suggestions that the indicator-kriging procedure results in too many dry days for most grid boxes.

The range of estimated area-averages based on interpolations of less dense networks is large and mainly directed towards lower variance and, in the case of precipitation, also dry day probability and mean, again a result of the lower shared variance of stations further away from the grid-box. When four stations are used for the

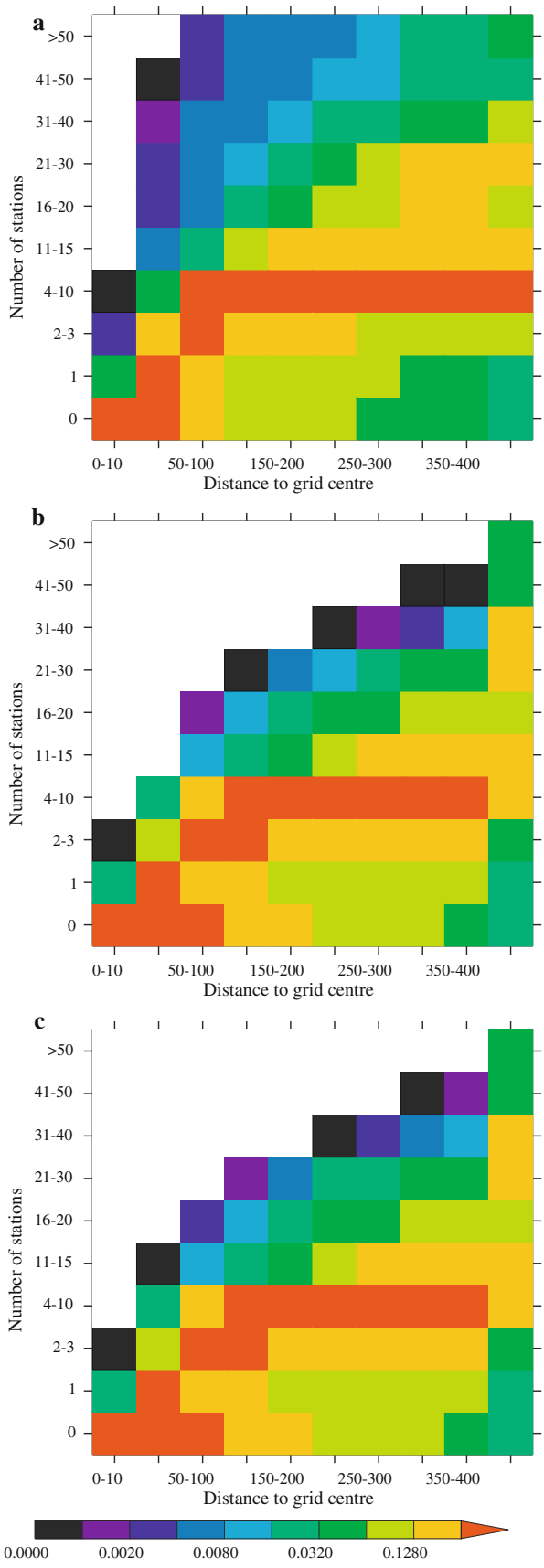


Fig. 10 The proportion of 0.22° grid-boxes of the E-OBS dataset that have a given number of stations (*y*-axis) within a given distance to the grid-centre (*x*-axis) for precipitation (*top*), minimum temperature (*middle*) and maximum temperature (*bottom*). All stations with data have been included, not just the ones with more than 69% of data available that were used in this study ▶

interpolation, the 95th percentiles of precipitation and temperature, range 51% compared to the mean value for the former and 23% for the latter variable. Significant, but smaller differences also occur for other percentiles. As a rule, the interpolated percentiles are reduced compared to those for the area-average derived from stations in the grid box, suggesting that over-smoothing occurs across the full distribution, albeit to a greater extent for higher percentiles.

The analysis of trends in extremes indices shows that the spread in trend between realisations is generally smaller when more stations have been used for the interpolation. The sign of the trend of the realisations generally agrees with the sign of the stations within the grid-boxes, although there are some exceptions. This would suggest that in most cases trends at an individual grid box are consistent with a wider regional trend. The strength of any trend is generally lower for the 0.44° than the 0.22° grid boxes, which is caused by the fact that larger extremes are preferentially smoothed more than smaller extremes for larger area-averages, likely because the magnitude of these strong extremes are shared by fewer stations in smaller areas

In some specific areas local conditions produce exceptions to the general results summarised above. For example, the variation of precipitation for INT-ALL for the Irish grid-boxes (only the 0.22° box for south Ireland) is larger than AVG, suggesting that the data are not smoothed enough. For temperature the difference between the 0.44° grid-box compared to the 0.22° grid-box is reversed; there is an increase in the temperature visible for the grids in Ireland and close to Edinburgh. In this case the exception is probably realistic, as the 0.44° grid represents a different area with larger variance than the 0.22° grid. In the Alps the spread of area-average values tends to straddle AVG suggesting that the general smoothing observed at the other grid boxes may be more random in an area with complex topography.

This analysis shows that the network density can introduce biases in the mean and variance of the E-OBS grid values compared to those expected for the true area-averages. In general, both the mean and variance of daily precipitation and, to a lesser extent, temperature of E-OBS are reduced through interpolation unless the network density is extremely high. The degree of over-smoothing is greater in the more extreme percentiles, and in general for less dense station networks. The same pattern of over-smoothing is also reflected in the extremes, where the largest extremes are smoothed more for less dense networks, which may influence the magnitude of the trends in extreme indices. The trends are also influenced by the number of stations used for the interpolation, indicating that the analysis of extreme trends in the E-OBS data may suffer from the changes over time in the station network used for the interpolation. When fewer stations have been used the difference in smoothing between the 0.44° grid-boxes and 0.22° grid-boxes is much smaller than when more stations have been used. This is because more smoothing occurs during interpolation to 0.1° grid points when the network is sparse, and this is then inherited by both 0.44° and 0.22° area-averages. Therefore, in areas of the E-OBS grid where the station network is sparse, we advise to use the 0.44° data rather than the 0.22° data, because the latter will be more over-smoothed than the 0.44° data.

The grid-boxes selected for this study all had a very dense station network in and around the grid-box, whereas most of the grid boxes in the E-OBS dataset are surrounded by a much sparser station network. Figure 10 summarises the distribution of station density for the entire E-OBS dataset. We do this by counting the number of stations falling within a series of search radii of increasing length, on a grid-box by grid-box basis and then counting the proportion of grid boxes that have a given number of stations within each search radius. The majority of grid-boxes have between 4 and 15 stations available within the maximum search radius (4 is the minimum amount of stations for which the interpolation has been carried out), but a very small fraction of stations have that many stations within 50 km of the grid-centre. In our analysis we have shown that area-averages derived from sparse networks (INT-4–INT-16) result in a high likelihood that the true area-average is incorrectly estimated, and that this will tend to be an underestimate. Thus, a large part of the E-OBS gridded data are expected to be over-smoothed and the degree of smoothing will vary over time with variations in the station network.

Even though the E-OBS dataset has been developed for the evaluation of RCM outputs and, therefore, this paper focuses on that field of study, the data can of course be used for many other analyses. Haylock et al. (2008) list studies for which the interpolated data is important, including monitoring of climate change, assessment of patterns of coherent variability and impact studies that use the climate data as driving data or for calibration. The high likelihood of over-smoothing should be borne in mind during any application of the E-OBS dataset. For example, if the precipitation data are used for rainfall-runoff modelling the runoff outputs will be influenced by the underestimated larger percentiles of precipitation, which may cause an underestimation of flooding.

In this study, we have explored the issues of scaling and smoothing in one dataset of gridded daily data. Many more daily gridded data have been developed with different station network densities and interpolation methods, where the issue of scaling has not been assessed (e.g., Caesar et al. 2006; Feng et al. 2004; Van der Groot and Orlandi 2003; Hewitson and Crane 2005; Piper and Stewart 1996; Rubel et al. 2004). Further, the E-OBS dataset itself could be

further evaluated, for example, to identify areas and time-periods that produce values closer to true areal averages than others. Scaling issues in the evaluation of climate models have only just started to receive more attention in climate science (e.g., Booij 2002; Chen and Knutson 2008; Fowler et al. 2005; McSweeney 2007). It is clear that the gridded data have biases that need to be considered in such evaluations, as in some cases mismatches between climate model output and observations may be partly due to inaccuracies in the observational data. It may be more prudent to evaluate RCMs against only those observational grid boxes that satisfy certain "station density" criteria, as for example Beniston et al. (2007), Buonomo et al. (2007), Huntingford et al. (2003), Jones and Reid (2001) and Semmler and Jacob (2004) have done in the past. More remains to be done to ensure that over-smoothed gridded data do not result in an over-smoothing RCM being incorrectly selected as the best performing one in an evaluation of climate models.

Acknowledgments We would like to thank all institutes (see Appendix 1 of Klok and Klein Tank (2009)) that made meteorological station data available for the study. This study was funded by the EU project ENSEMBLES (WP 5.1 contract GOCE-CT-2004-50539). NH is also funded by the Dutch Prins Bernhard Cultuurfondsbeurs and the Dutch talentenbeurs.

References

Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Rupa Kumar K, Revadekar J, Griffiths G, Vincent L, Stephenson DB, Burn J, Aguilar E, Brunet M, Taylor M, New M, Zhai P, Rusticucci M, Vazquez-Aguirre JL (2006) Global observed changes in daily climate extremes of temperature and precipitation. J Geophys Res 111:D05109. doi:10.1029/2005JD006290

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

Beniston M, Stephenson DB, Christensen OB, Ferro CAT, Frei C, Goyette S, Halsnaes K, Holt T, Jylhä K, Koffi B, Palutikof JP, Schöll R, Semmler T, Woth K (2007) Future extreme events in European climate: an exploration of regional climate model projections. Clim Change 81:71–95. doi:10.1007/s10584-006-9226-z

Booij MJ (2002) Extreme daily precipitation in Western Europe with climate change at appropriate spatial scales. Int J Climatol 22:69–85. doi:10.1002/joc.715

Buonomo E, Jones RG, Huntingford C, Hannaford J (2007) On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations. Q J Meteorol Soc 133:65–81. doi:10.1002/qj.13

Caesar J, Alexander L, Vose R (2006) Large-scale changes in observed daily maximum and minimum temperatures: creation and analysis of a new gridded dataset. J Geophys Res 111:D05101. doi:10.1029/2005JD006280

Chen C-T, Knutson T (2008) On the verification and comparison of extreme rainfall indices from climate models. J Clim 21:1605–1621. doi:10.1175/2007JCLI1494.1

Christidis N, Stott PA, Brown S, Hegerl GC, Caesar J (2005) Detection of changes in temperature extremes during the second half of the 20th century. Geophys Res Lett 32:L20716. doi:1031029/2005GL023885

Feng S, Hu Q, Qian W (2004) Quality control of daily meteorological data in China, 1951–2001: a new dataset. Int J Climatol 24:853–870. doi:10.1002/joc.1047

Fowler HJ, Ekström M, Kilsby CG, Jones PD (2005) New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate. J Hydrol 300:212–233. doi:10.1016/j.jhydrol.2004.06.017

Gerstner E-M, Heinemann G (2008) Real-time areal precipitation determination from radar by means of statistical objective analysis. J Hydrol 352:296–308. doi:10.1016/j.jhydrol.2008.01.016

Groisman PY, Knight RW, Easterling DR, Karl TR, Hegerl GC, Razuvaev VN (2005) Trends in intense precipitation in the climate record. J Clim 18:1326–1350

Hanson CE, Palutikof JP, Livermore MTJ, Barring L, Bindi M, Corte-Real J, Durao R, Giannakopoulos C, Good P, Holt T, Kundzewicz Z, Leckebusch GC, Moriondo M, Radziejewski M, Santos J, Schlyter P, Schwarb MC, Stjernquist I, Ulbrich U (2007) Modelling the impact of climate extremes: an overview of the MICE project. Clim Change 81:163–177

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

Hewitson BC, Crane RG (2005) Gridded area-averaged daily precipitation via conditional interpolation. J Clim 18(1):41–57

Hewitt CD, Griggs DJ (2004) Ensembles-based predictions of climate changes and their impacts. EOS 85:566–568

Hofstra N, New M (2009) Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over Europe. Int J Clim 29:1872–1880. doi:10.1002/joc.1819

Hofstra N, Haylock M, New M, Jones PD, Frei C (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, Haylock M, New M, Jones PD (2009) Testing E-OBS European high-resolution gridded dataset of daily precipitation and surface temperature. J Geophys Res (in press). doi:10.1029/2009JD011799

Huntingford C, Jones RG, Prudhomme C, Lamb R, Gash JHC, Jones DA (2003) Regional climate-model predictions of extreme rainfall for a changing climate. Q J Meteorol Soc 129:1607–1621. doi:10.1256/qj.02.97

Jones PD, Reid PA (2001) Assessing future changes in extreme precipitation over Britain using regional climate model integrations. Int J Climatol 21:1337–1356. doi:10.1002/joc.667

Kjellström E, Ruosteenoja K (2007) Present-day and future precipitation in the Baltic Sea region as simulated in a suite of regional climate models. Clim Change 81:281–291. doi:10.1007/s10584-006-9219-y

Klein Tank AMG, Können GP (2003) Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. J Clim 16:3665–3680

Klok EJ, Klein Tank AMG (2009) Updated and extended European dataset of daily climate observations. Int J Climatol 29:1182–1191. doi:10.1002/joc.1779

McSweeney CF (2007) Daily rainfall variability at point and areal scales: evaluating simulations of present and future climate. University of East Anglia, Norwich, p 256

New M, Todd M, Hulme M, Jones PD (2001) Precipitation measurements and trends in the twentieth century. Int J Climatol 21:1899–1922. doi:10.1002/joc.680

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

Piper SC, Stewart EF (1996) A gridded global data set of daily temperature and precipitation for terrestrial biospheric modeling. Global Biogeochem Cycles 10(4):757–782

Reynolds RW (1988) A real-time global sea surface temperature analysis. J Clim 1:75–86

Rubel F, Brugger K, Skomorowski P, Kottek M (2004) Daily and 3-hourly quantitative precipitation estimates for ELDAS, edited. Biometeorology Group, University of Veterinary Medicine, Vienna, 32 p

Santos JA, Corte-Real J, Ulbrich U, Palutikof JP (2007) European winter precipitation extremes and large-scale circulation: a coupled model and its scenarios. Theor Appl Climatol 87:85–102. doi:10.1007/s00704-005-0224-2

Semmler T, Jacob D (2004) Modeling extreme precipitation events—a climate change simulation for Europe. Glob Planet Change 44:119–127. doi:10.1016/j.gloplacha.2004.06.008

Simmons AJ, Jones PD, da CostaBechtold V, Beljaars ACM, Kållberg PW, Saarinen S, Uppala SM, Viterbo P, Wedi N (2004) Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. J Geophys Res 109:D24115. doi:10.1029/2004JD005306

Thom HCS (1958) A note on the gamma distribution. Mon Weather Rev 86:117–122

Van der Groot E, Orlandi S (2003) Technical description of interpolation and processing of meteorological data in CGMS, pp 16

Wilks DS (2006) Statistical methods in the atmospheric sciences, 2nd edn. Elsevier, Burlington, p 627