Synthetic Weather Time Series

Simulation Phase Report

A. Burton[1], C. Harpham[2], P. Jones[2], C. Goodess[2],
C.G. Kilsby[3]

September 2019

1. Enviro Techno Insights Ltd, Durham, UK
2. University of East Anglia, Norwich, UK
3. Newcastle University, Newcastle upon Tyne, UK

Table of Contents

Introduction

The Universities of Newcastle (NCL) and East Anglia (UEA) have been advised that Bayer AG (BAG) and the University of Copenhagen are collaborating on a scientific research project which started in August 2017. The aim of the project being to explore possibilities for optimization of pesticide application dates taking into account soil moisture and weather forecasts so as to reduce surface water contamination via drainage systems. The optimization will be carried out primarily using a mathematical modelling approach which requires long time series of synthetic weather data. Whilst not being actual weather observations or forecasts, these synthetic weather data will consist of model output which represents a wide range of locally relevant plausible weather scenarios.

The aim of the Simulation Phase Collaboration Agreement (effective date 15/12/18) between BAG, NCL and UEA is the generation of synthetic weather series using weather generator software (RainSim and CRU-WG) developed by the UK Universities. Enviro Techno Insights Ltd (ETI) contributed as a subcontractor to NCL.

This report describes the generation of these synthetic weather time series for the selected European locations. Two locations were selected, Eelde, in the Netherlands, and Chemnitz, in Germany. Whilst the agreement included an option for a third location, this was not requested.

Specification

The key properties required of the synthetic data were that they should be similar to the synthetic control series in the Rasmussen et al. (2018) paper. Also, for an application to evaluate pesticide leaching risk, the most important data would be the hourly rainfall, especially the extreme events.

Methodology

For both locations, the preparation of the synthetic weather time series followed a similar methodology which is outlined here.

Observed weather data provides a sample representing a location’s weather properties. Typically, observed weather data undergo quality assurance checks as part of the archival process. These go some way to address potential limitations in the observational record such as measurement errors, missing data, recording errors and processing errors, and may infill missing data with plausible values. In general, the characteristics of the weather may differ in the future from those recorded, there may be systematic biases in observations (e.g. rainfall under-catch due to strong winds) or missing data may have occurred in a systematic way so introducing biases in the observations.

The weather datasets provided for each location were assumed to be a reliable representation of the actual weather that occurred at the study sites. However, some checks were carried out to identify missing and spurious data in the observed datasets. It is not important to remove all spurious values as the weather models are typically fitted to a sample of 20-30 years of observations, so reducing the influence of occasional non-systematic errors. Extreme events, such as intense rainfall, can however, be disproportionately influential, and so extreme outliers are checked for reasonableness.

The checks carried out would not, however, be sufficient to identify most individual erroneous values, systematic biases in the observational record, or changes to the observed record introduced during the archival process.

Annual and seasonal anomaly plots were produced, to investigate how precipitation and temperature have changed over time, according to the observed dataset. Here, “anomaly” is used to denote the difference between a value of a variable and the annual or seasonal mean for that variable. Following this the period of observations to be used in the study was identified and the observed datasets formatted for input to the weather models.

The observed rainfall was then characterised by estimating, for each calendar month, a set of daily and hourly statistics as input data for the weather generation model for the selected study period. These statistics are described as the hourly variance, proportion dry and skewness coefficient; the daily mean, proportion dry, variance, correlation and skewness coefficient; and the 28-day (672-hour) variance. First, the time series of rainfall data was partitioned by calendar month and then accumulated to the required aggregation period (either hourly, daily or 28-days). The daily variance (for a given calendar month) is then the variance of the set of daily accumulations (for that month). Hourly and 28-day variances are similarly derived. The *proportion dry* statistics refer to the proportion of the set of aggregated data considered ‘dry’, where here a dry period is one in which the accumulation is strictly less than 0.2mm. Daily auto-correlation is the correlation of daily rainfall with that of the following day for days falling within the specified calendar month. Thus a high correlation would suggest that each day’s rainfall is strongly related to the adjacent days’ rainfall. Alternatively, a zero value of correlation suggests little connection between rainfall values on adjacent days. The skewness coefficient is the third order central moment of the aggregated data, standardised using the standard deviation cubed. This gives a measure of asymmetry of the distribution of rainfall, with (for example) a positive skewness coefficient indicating that outlying higher values are more common than similarly outlying lower values.

A single site Neyman Scott Rectangular Pulses (NSRP) rainfall model (e.g. see Burton et al., 2008) was then fitted to the estimated observed rainfall statistics. The fitting procedure used a numerical optimization approach to minimize the difference between the observed and the expected properties of the synthetic time series. Weights were set on the various observed statistics, to focus on the accuracy of the key properties required in the synthetic series and a correction procedure was used to reduce biases introduced by analytical approximations.

Following the fitting, the rainfall model was used to generate an ensemble of 100 30-year synthetic hourly rainfall time-series, and the corresponding daily aggregated time series. Plots of the annual cycle of the observed and simulated statistics were then prepared and compared. Extreme value plots comparing observed and simulated annual maximum daily and hourly rainfall were also prepared.

The remaining weather variables could then be simulated at daily resolution using a conditional multivariate autoregressive approach (often more simply referred to as a ‘weather generator’) that preserves both inter-weather-variable relationships and the seasonal weather cycle (Jones et al., 2016). The daily variables used in this study (in addition to precipitation) were minimum temperature, maximum temperature, vapour pressure, sunshine hours and wind speed. Derived variables were also calculated. These included relative humidity, diffuse radiation, direct radiation and reference potential evapotranspiration (PET). Here PET was calculated using the Penman-Monteith method (see details in Jones et al., 2016).

For each location, the conditional autoregressive model was fitted to the observed weather dataset, then the synthetic daily rainfall time-series was used to condition the simulation of the ensemble of 100 30-year daily synthetic weather datasets using the fitted model. Validation plots comparing the observed weather properties with those in the simulated ensemble were then prepared and evaluated.

First location: Eelde, the Netherlands

Weather Observations

The Eelde meteorological station is located at 53.125° N 6.585° E in the Netherlands. This station is at the airport for Groningen. The meteorological data was obtained from the KNMI website[1]. This data may be reused, noting that KNMI was not involved in and has not provided any endorsement of the derived works[2].

The data included both daily and hourly time series of weather observations for the period 1957 - 2019. The rainfall data was provided in units of 0.1mm with an indicator for a trace of rainfall. Two rainfall time series were available, one at daily and one at hourly resolution. The daily values appeared consistent with the daily aggregation being midnight to midnight.

Analysis of the observed data

Figure 1 shows annual temperature averages and annual precipitation totals for Eelde for 1957-2018. Figure 2 shows seasonal precipitation totals and Figure 3 seasonal temperature averages. In these and subsequent time series plots, the smooth curve is a 10-year Gaussian filter, which highlights changes on the decadal timescale. Also in these plots, the horizontal line is the 1961-90 average with warmer years/seasons shown in red and cooler years in blue. Wetter years/seasons are shown in green and drier years/seasons shown in brown.

Precipitation trends at Eelde indicate little long-term change in annual and seasonal totals over the 1957 to 2018 period. Temperature, in contrast, indicates almost all years since 1988 were warmer than the 1961-90 average. The only three years that were cooler were 1993, 1996 and 2010. Seasonally, temperatures have generally been warmer since 1988, more so in winter, spring and summer, than in the autumn season.

Typically climatological studies use a 30-year period, so here the latest 30-years of weather data, 1989 - 2018, was selected as a basis for this study. The following analyses all relate to this period (unless otherwise stated). From Figures 1 and 3, this choice of the 30-year period means that the averages of the generated temperatures will reflect this selected period, and not the 1961-90 average.

The hourly rainfall dataset did not contain any indicators of missing data. A lack of missing data seems unusual in such an observed dataset as there are many ways in which observations can be interrupted, or errors introduced. This suggests that either some records of the ‘observed’ series have been estimated or ‘corrected’ in some way, or possibly that missing data has been incorrectly recorded as zero rainfall.

1 http://projects.knmi.nl/klimatologie
2 https://www.knmi.nl/copyright translated by Bayer.

Here it was assumed that a record of zero rainfall was an *actual* observation of no rainfall and that the data correctly represented the underlying weather properties at Eelde. Where trace values of rainfall were indicated (< 0.05mm) these were set to zero rainfall prior to further analysis. Following these preliminary assumptions, the average annual rainfall total for 1989 to 2018 was 802 mm/year.

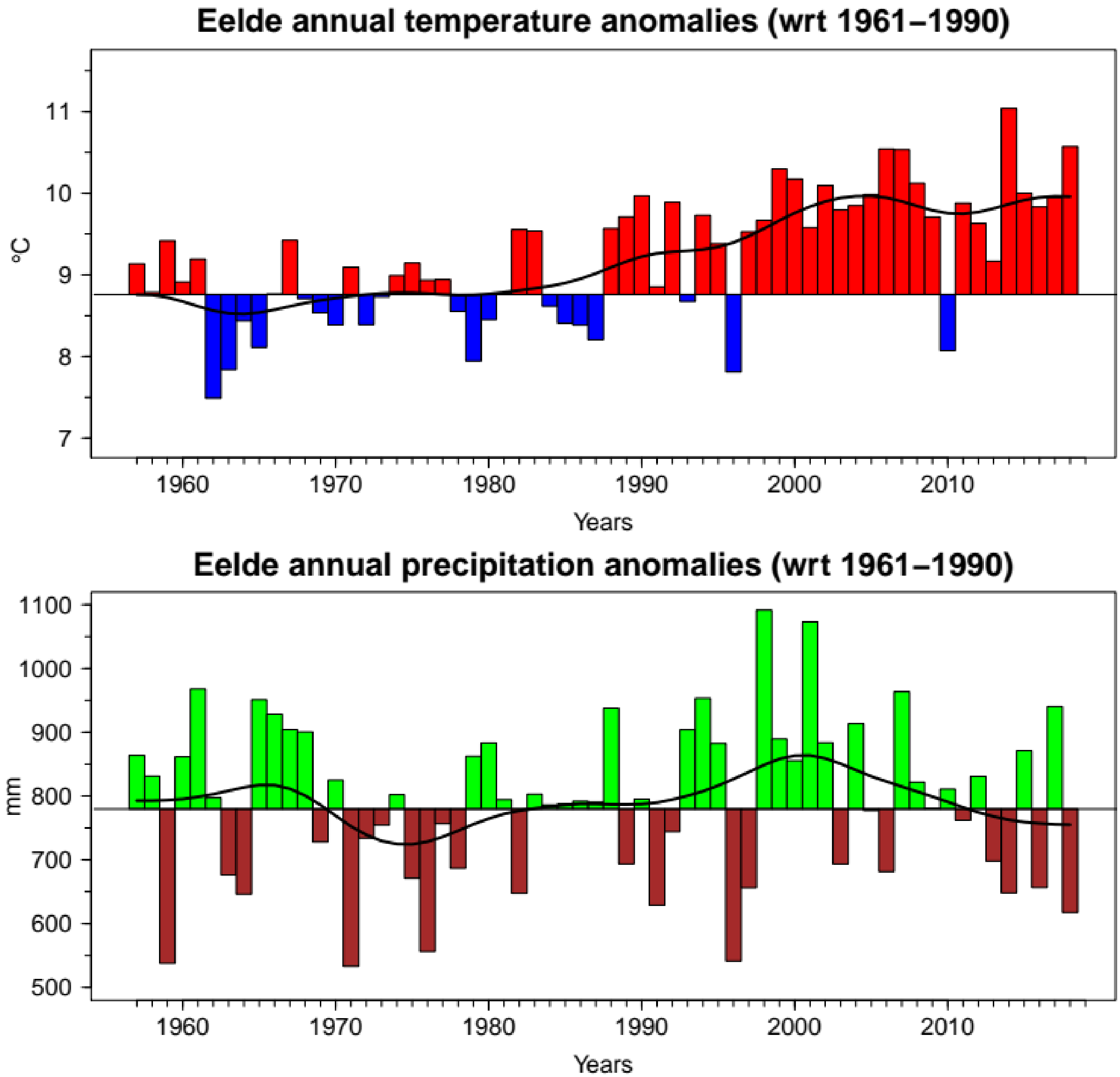

Figure 1. Plot of the annual temperature and precipitation anomalies for the period 1957-2018 with respect to a 1961-1990 baseline (horizontal line). Warmer years are shown in red and cooler years in blue. Wetter years are shown in green and drier years in brown. The smooth curve is a 10-year Gaussian filter.

Preparation of the Synthetic Weather Time series

The NSRP rainfall model was fitted to the daily and hourly rainfall statistics estimated from the observed hourly dataset. Once fitted, the model was used to generate an ensemble of 100 30-year hourly rainfall time-series. The statistics plots (Figure 4) compare the statistics of the daily and hourly observed and simulated rainfall for the Eelde station. In each case: the diamonds show the sample estimate of the observed rainfall data and the red lines indicate the mean and typical variation exhibited across the synthetic ensemble. Differences between the mean synthetic statistic and the observed statistic shown in each case comprise a combination of sample variability, fitting approximation and biases. Overall, the synthetic data are a reasonable match to the observed rainfall properties.

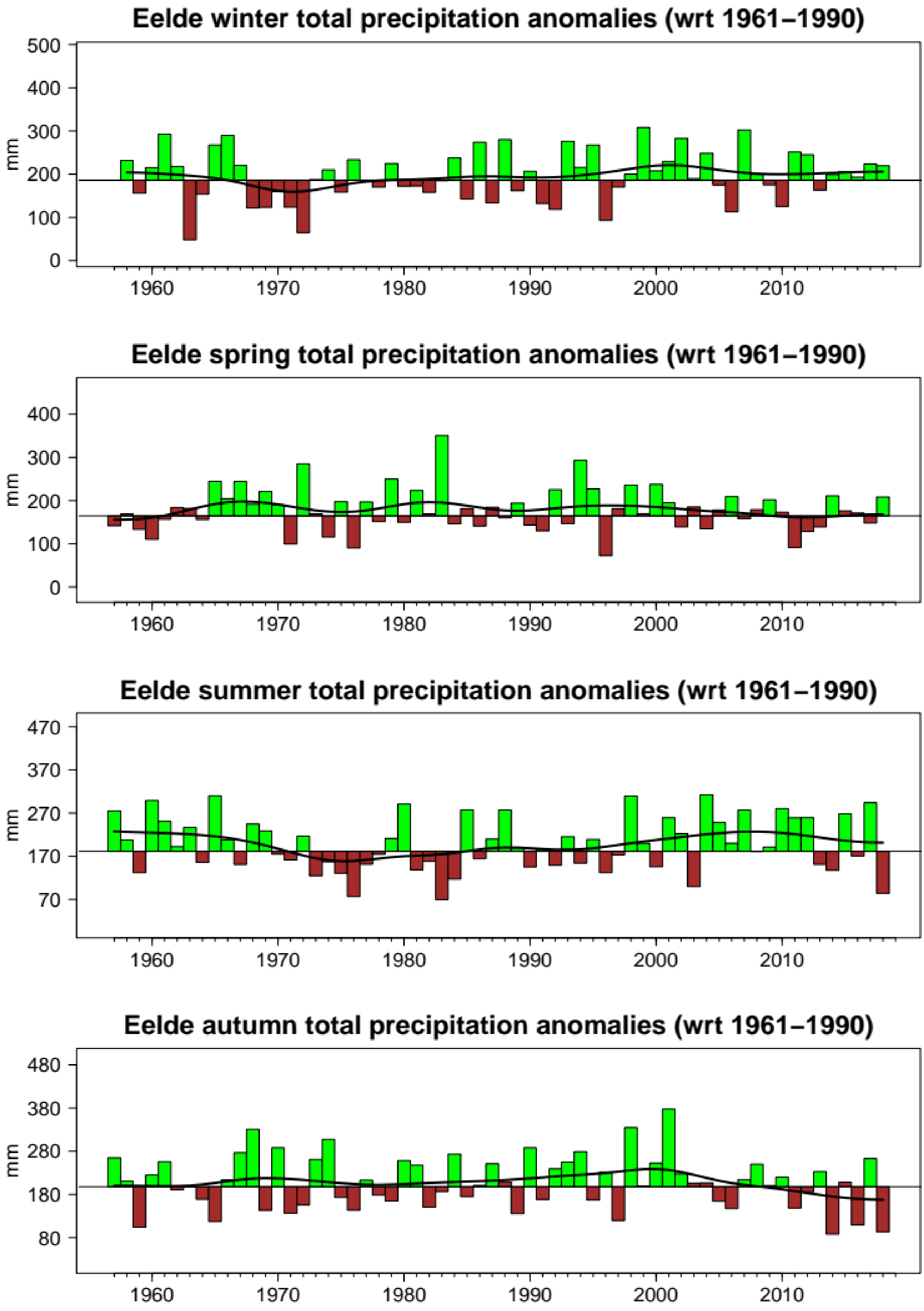

Figure 2. Plot of the seasonal precipitation anomalies for the period 1957-2018 with respect to a 1961-1990 baseline (horizontal line). Wetter seasons are shown in green and drier seasons are shown in brown. The smooth curve is a 10-year Gaussian filter.

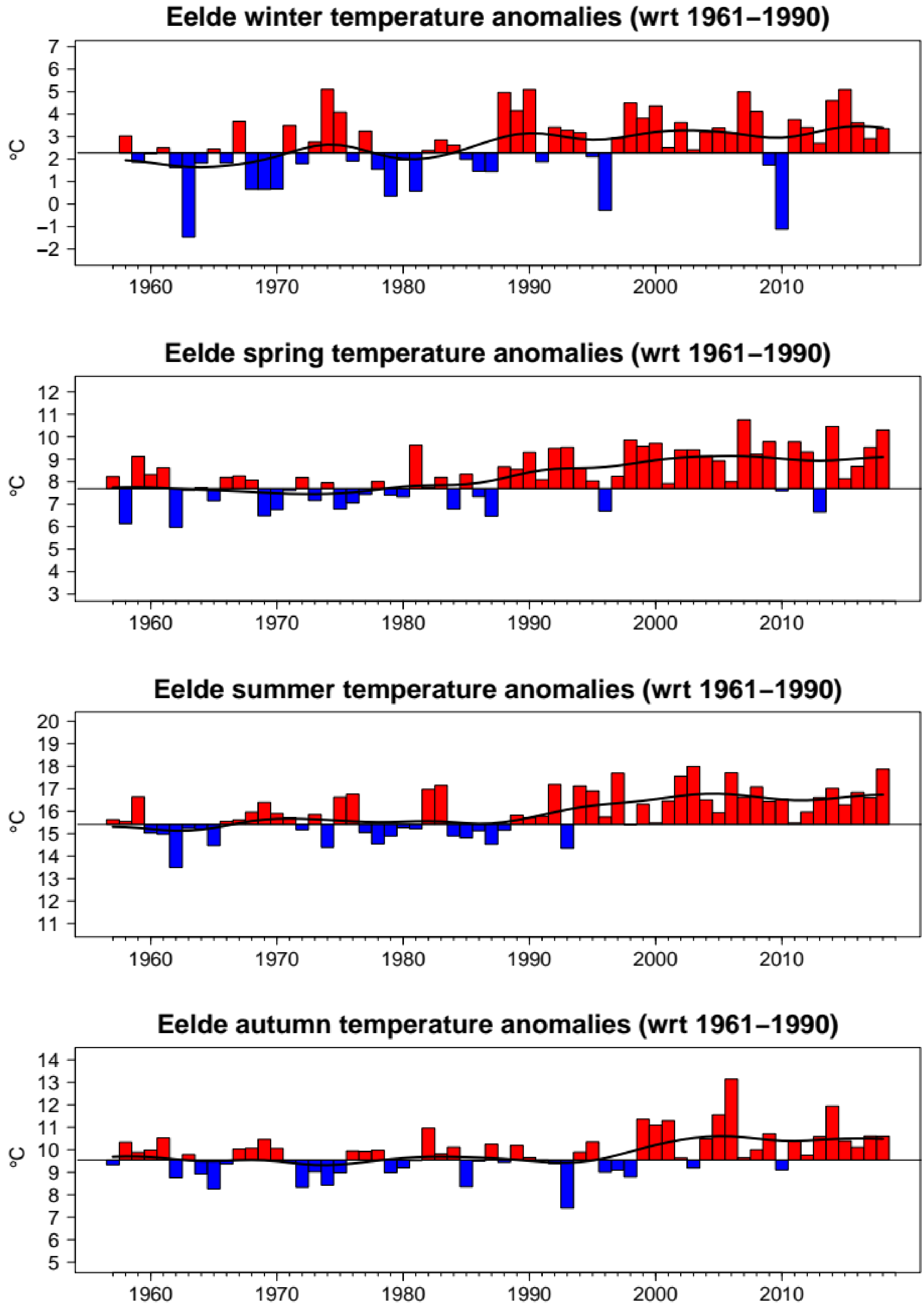

Figure 3. Plot of the seasonal temperature anomalies for the period 1957-2018 with respect to a 1961-1990 baseline (horizontal line). Warmer seasons are shown in red and cooler seasons in blue. The smooth curve is a 10-year Gaussian filter.

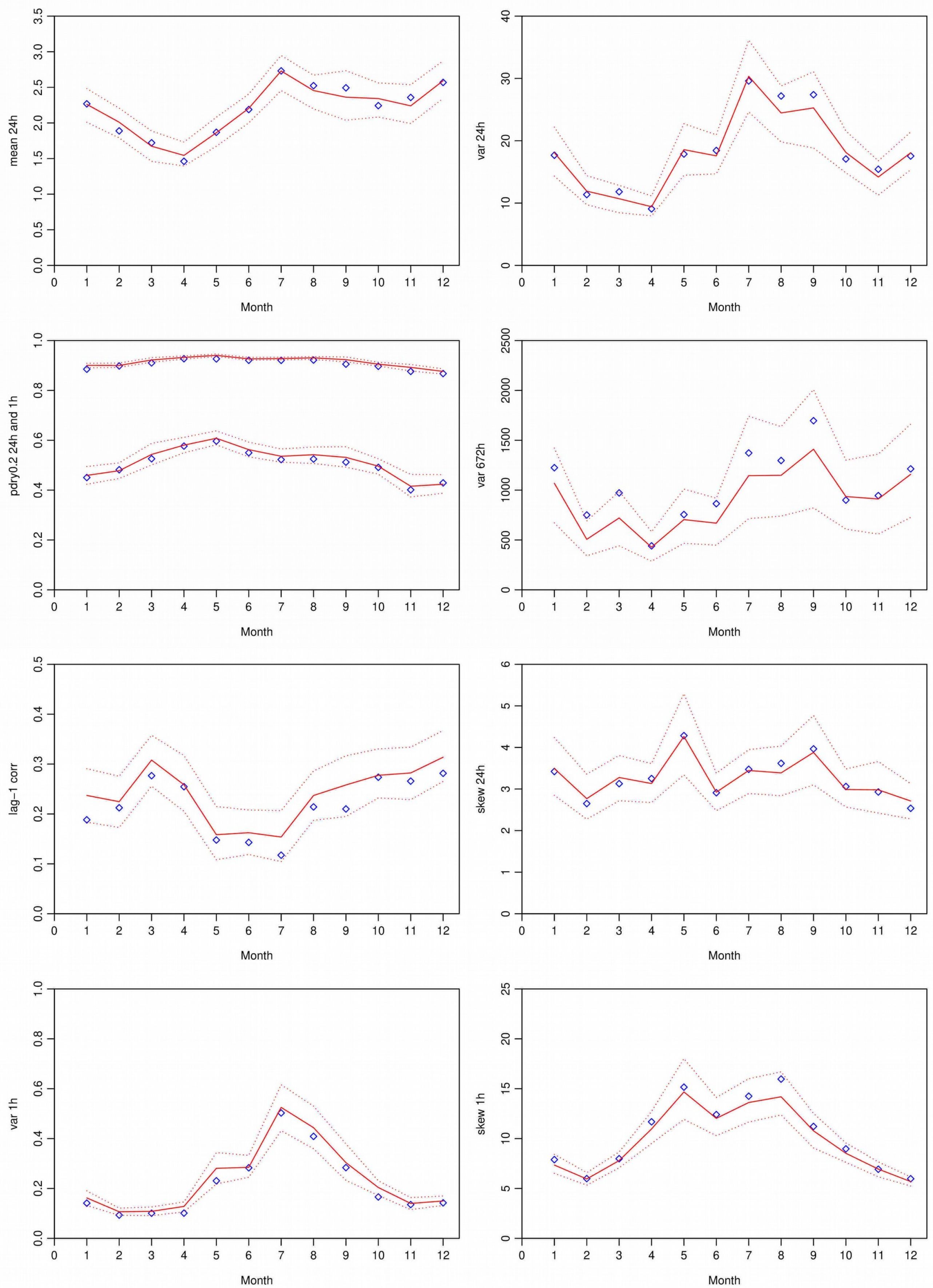

Figure 4. The observed and ensemble simulated rainfall statistics for Eelde by calendar month. Plots and units are: daily mean (mm), daily variance (mm^2), proportion dry (pdry) (-) daily (lower) and hourly (upper), 28-day variance (mm^2), daily autocorrelation (-), daily skewness coefficient (-), hourly variance (mm^2) and hourly skewness coefficient (-). The blue diamonds indicate the observed statistics. The red lines indicate the mean and the 10th and 90th percentiles across the 100 member ensemble of 30-year synthetic time series.

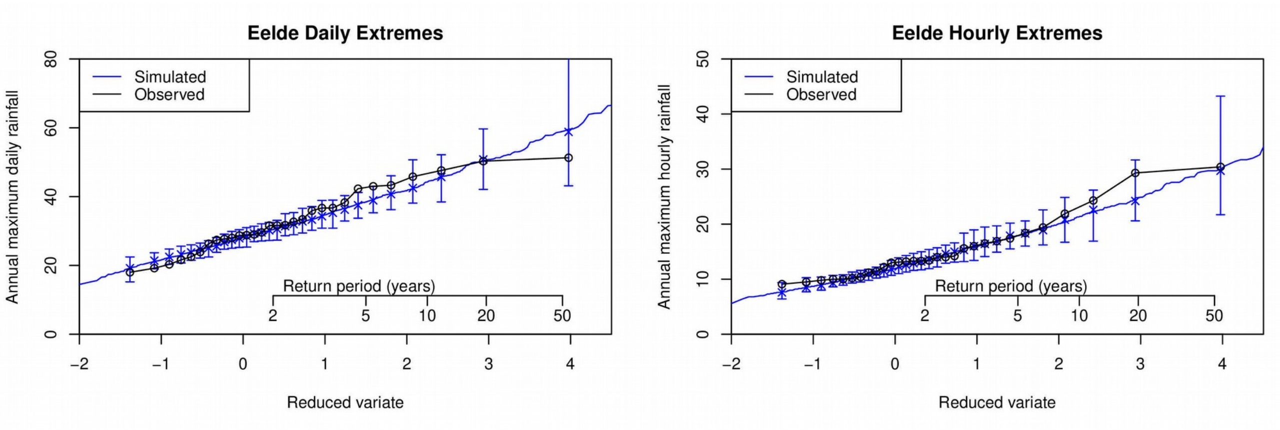

Figure 5. Extreme value plots showing observed and simulated annual maximum daily and hourly rainfall (mm). The x-axis relates to the return period which is 1 / (frequency of exceedance). The observed extremes are plotted as black circles and lines. A synthetic 1000-year rainfall time series from the rainfall model was similarly analysed and plotted as the blue curve. 30-year subsamples of this series were then used to estimate the simulated variability at each plotting position (blue): the cross shows the median and the bar indicates the estimated 5 – 95 percentile range.

The observed and simulated extreme rainfall properties are compared in Figure 5, an extreme value plot. This compares the observed and simulated annual maximum daily and hourly rainfall amounts. The x-axis is a transformed time (expected time between exceedances) axis, with selected return periods indicated on the internal axis. The observed extremes are plotted based on the 30-years of observations. A 1000-year synthetic hourly rainfall series was simulated and divided into an ensemble of 30-year series (the same length as the observations). Analysis of the extremes of this ensemble provides an indication of the uncertainty of the value simulated at each plotting position, shown using error bars in Figure 5. Whilst the rainfall model is not fitted directly to the extreme value distribution, it can be seen that both daily and hourly extreme rainfall properties are in excellent agreement.

Figures 6 and 7 show a comparison of the 100 30-year simulations with the 1989-2018 observations for a number of metrics. The weather generator (WG) is based on half months (see details in Jones et al., 2016), as this better simulates the annual cycle of air temperatures. The x-axis scale of half months means, for example, that 13 and 14 represent the first and second half of July. The red ranges indicate the two standard deviation (SD) ranges of the various metrics, with the blue cross marking the 1989-2018 average of the observations. The two SD range means that the range encompasses 95% of the distribution of the 100 simulations, so the top/bottom of the red bar is 97.5%/2.5%. All the blue crosses are within this range, but a few are close to the upper or the lower limit of the range.

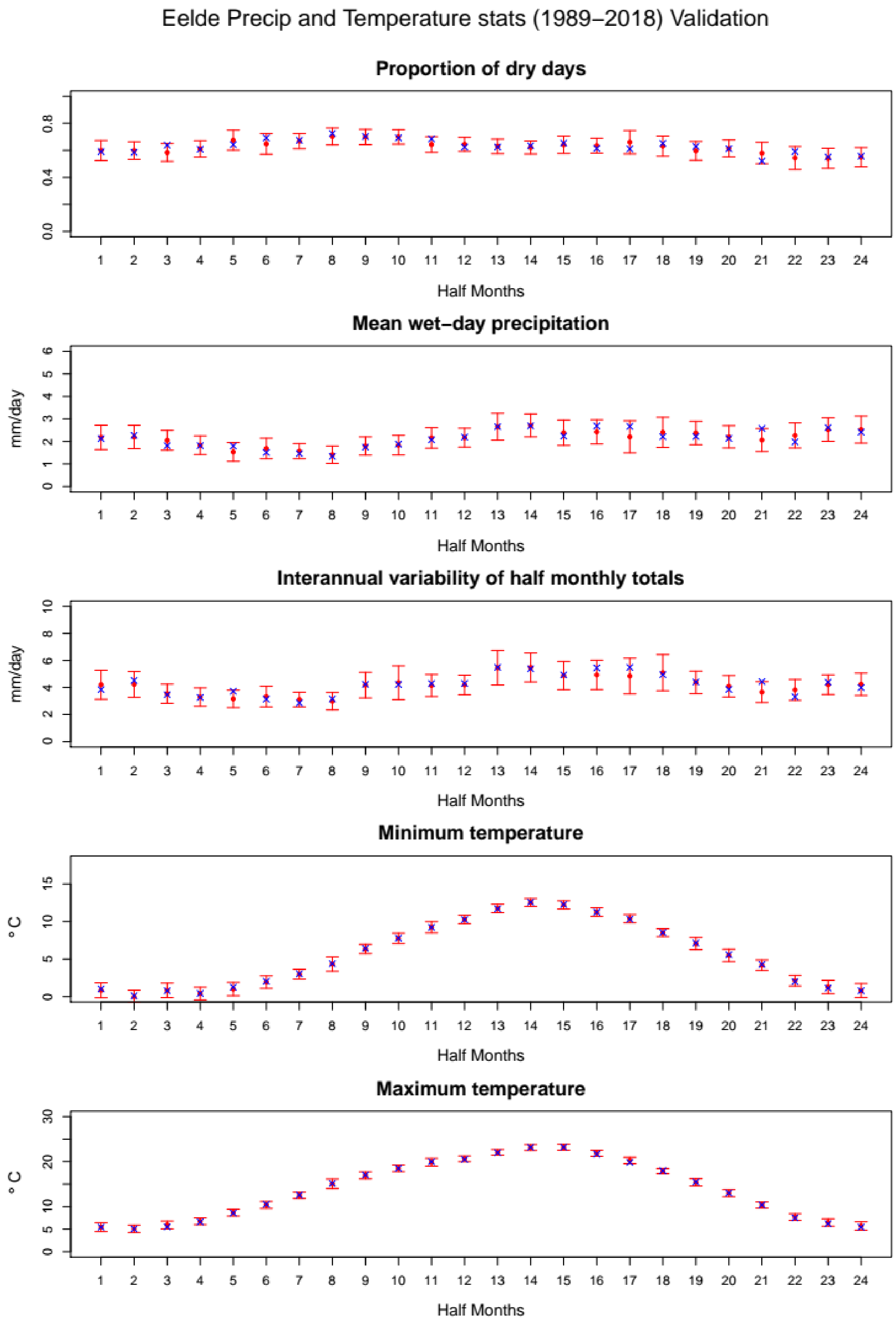

Figure 6. Daily precipitation and temperature validation plots comparing the annual cycle of observations (blue crosses) with the simulated properties (red bars). The x-axis scale indicates half months (e.g. 13 and 14 represent the first and second half of July). The red bars indicate the two standard deviation ranges of the simulated properties across the 100 simulation ensemble.

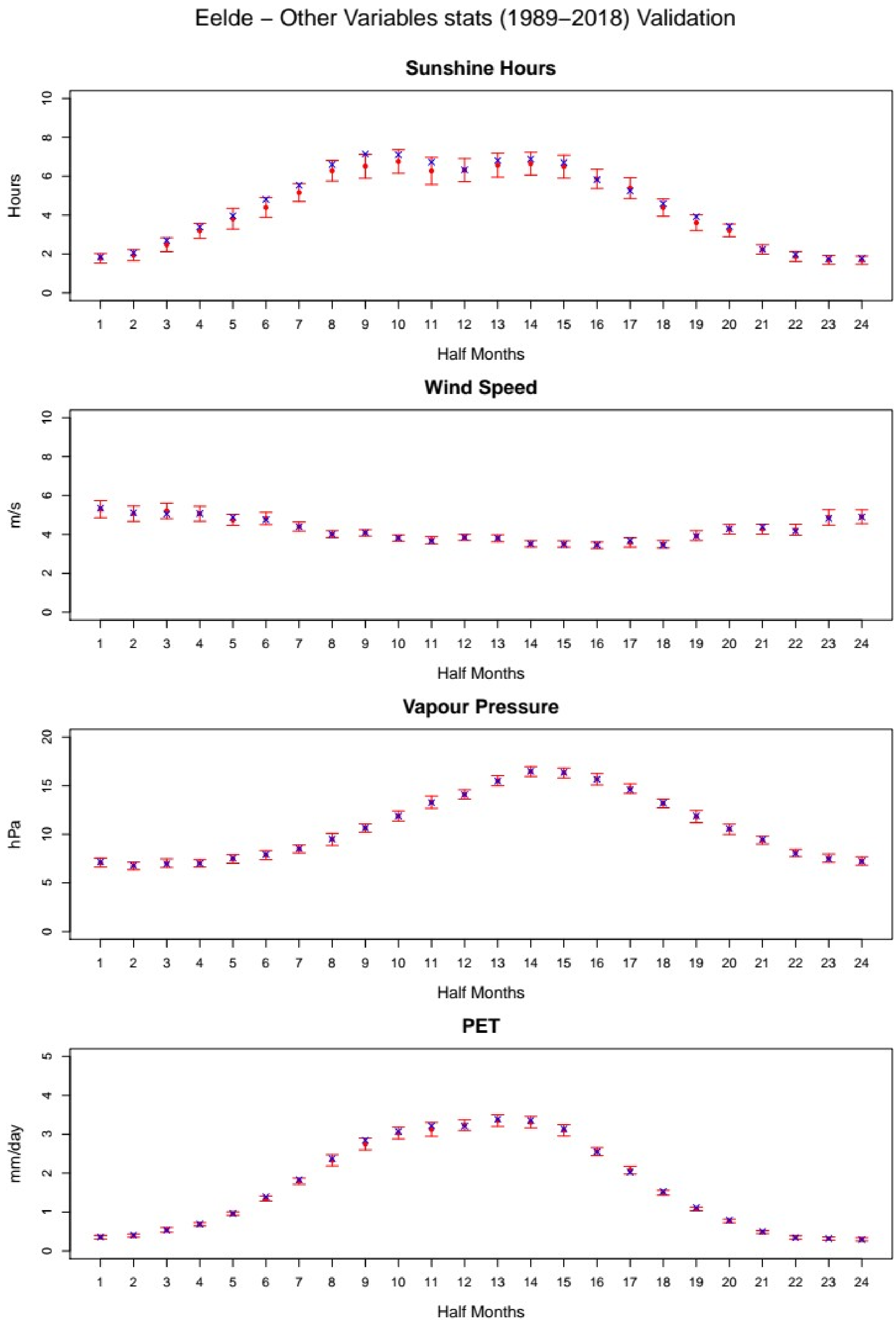

Figure 7. Daily sunshine, wind, vapour pressure and potential evapotranspiration (PET) validation plots comparing the annual cycle of observations (blue crosses) with the simulated properties (red bars). The x-axis scale indicates half months (e.g. 13 and 14 represent the first and second half of July). The red bars indicate the two standard deviation ranges of the simulated properties across the 100 simulation ensemble.

Second location: Chemnitz, Germany

Weather Observations
The Chemnitz weather station (station number 853) is located in Saxonia in Germany. The associated meta-data indicates that since 1976 it has been located at 50.7913° N 12.8720° E at an altitude of 418m (Deutscher Wetterdienst, 2019b).

Bayer obtained weather data from the Deutscher Wetterdienst[3] (DWD) Climate Data Center web-service[4] and provided it to NCL, UEA and ETI for use in this project. According to the associated metadata, freely accessible data from the Climate Data Center may be reused without restriction providing that the source reference is indicated, as described in the DWD terms of use.

The data include both daily rainfall times series from 1882-2018 and hourly rainfall time series observed from September 1995 – 2018 (i.e. ~ 23 years) and is provided with units of mm, a resolution of 0.1mm and a missing data indicator. The metadata appears to indicate that daily observations were accumulated to: 0700 German legal time (GZ) until 1990; 0730 GZ from 1991 until March 2001; and 0550 UTC from April 2001 until present (Deutscher Wetterdienst, 2019a)[5].

The Chemnitz weather observations include all those necessary for use in the weather generator model. These are maximum and minimum temperature, precipitation, sunshine hours, wind speed and vapour pressure.

Analysis of the observed data
Precipitation trends for Chemnitz are shown annually in Figure 8 and seasonally in Figure 10 for the period 1951 to 2018. Over this 68-year period there is little long-term trend in precipitation totals. Temperature trends (annually in Figure 8 and seasonally in Figure 9) in contrast indicate long-term warming with only two years since 1988 being cooler than the 1961-90 average. These years were 1996 and 2010. The temperature series for Chemnitz compare very favourably with those for Eelde.

Typically climatological studies use a 30-year period, but for this location only ~23 years of hourly data are available. From the perspective of characterising the hourly rainfall properties, the 23-years provides the best available observations of the Chemnitz hourly rainfall properties. From a climatic perspective, the use of only 23-years of data is not ideal, but seems to provide a reasonable compromise of using the available data in a straightforward manner when compared with the introduction of a more complex estimation procedure and the additional assumptions necessary to estimate the properties of a longer observation period. Therefore the 1995-2018 period was selected as the basis of this study, and the following analyses all relate to this period (unless otherwise stated). As noted for Eelde, this is a warmer period than 1961-90, but it does include the two cold years of 1996 and 2010.

Following the preliminary processing, the average annual total rainfall was estimated from the *daily* rainfall dataset to be 742 mm (for the period 1996-2018).

Missing data records only occur in the hourly rainfall observations from 2004 suggesting that prior to this date, they were not recorded *as missing*. It was also found that many

3 https://www.dwd.de
4 The Climate Data Centre is located at https://www.dwd.de/EN/ourservices/cdcftp/cdcftp.html
5 Translation provided by Bayer.

hourly records (~1,400 or about 0.7% of the dataset) were simply omitted from the dataset (i.e. the records *were* missing rather than simply *indicating* missing data). After inserting the omitted records and labelling them as missing, the total number of missing hours in the record was found to be ~1,500. About 40% of days with missing records were found to occur in the year 1998, and after further analysis, each month in the period December 1997 – August 1998 was found to have between 9% and 16% of data missing. Such a coherent period with a relatively high proportion of missing data may indicate the introduction of sample biases into the observed dataset due to the missing data.

To investigate sample bias arising from missing data, the daily statistics from aggregated hourly data were compared with those from the available daily record (for the same time period). Different thresholds of acceptance / rejection of a month of rainfall data were considered. The difference between the daily statistics determined from the daily and the hourly datasets was found to be significantly reduced by rejecting any month in the hourly dataset in which the missing data exceeded 10%. In particular this threshold was found to remove most of the months in the period December 1997 – August 1998. The analysis suggested that the daily dataset has been corrected in some manner (or independently observed from the hourly data). It also suggested that the hourly data was systematically biased by the missing data, and that removing months in which missing data exceeded 10% could reduce these systematic errors. Consequently, months with more than 10% missing data were rejected from subsequent analyses and the remaining data used to characterise the hourly rainfall. This reduced the original 23 years and four months of hourly data by eight months.

Preparation of the Synthetic Weather Time series

The single site NSRP rainfall model was fitted to the estimated observed daily and hourly rainfall statistics estimated from the observed hourly dataset. The fitted model was used to generate an ensemble of 100 23-year hourly rainfall time-series, the length chosen to match the period of hourly observations. The ensemble statistics plots (Figure 11) compare the statistics of the daily and hourly observed and simulated rainfall for the Chemnitz station. In each case: the diamonds show the sample estimate of the observed rainfall data and the red lines indicate the mean and typical variation exhibited by the synthetic ensemble. Differences between the mean synthetic statistic and the observed statistic shown in each case comprise a combination of sample variability, fitting approximation and biases. There is again a good match between the synthetic data and the observed rainfall properties.

As for Eelde, the observed and simulated extreme rainfall properties are compared with an extreme value plot (Figure 12). A preliminary comparison of the daily rainfall extremes estimated from both the daily and the hourly datasets did not suggest any systematic bias. Therefore observed rainfall extremes were analysed for the period 1996 – 2018 assuming that important events were not omitted due to missing data. The observed extremes are plotted based on the 23-years of observations. A 1000-year synthetic hourly rainfall series was simulated and divided into an ensemble of 23-year series (the same length as the observations). Analysis of the extremes of this ensemble provides an indication of the uncertainty of the value indicated at each plotting position. This variability is shown using error bars in Figure 12. Whilst the rainfall model is not fitted to the extreme value distribution, it can again be seen that both daily and hourly extreme rainfall properties are in excellent agreement.

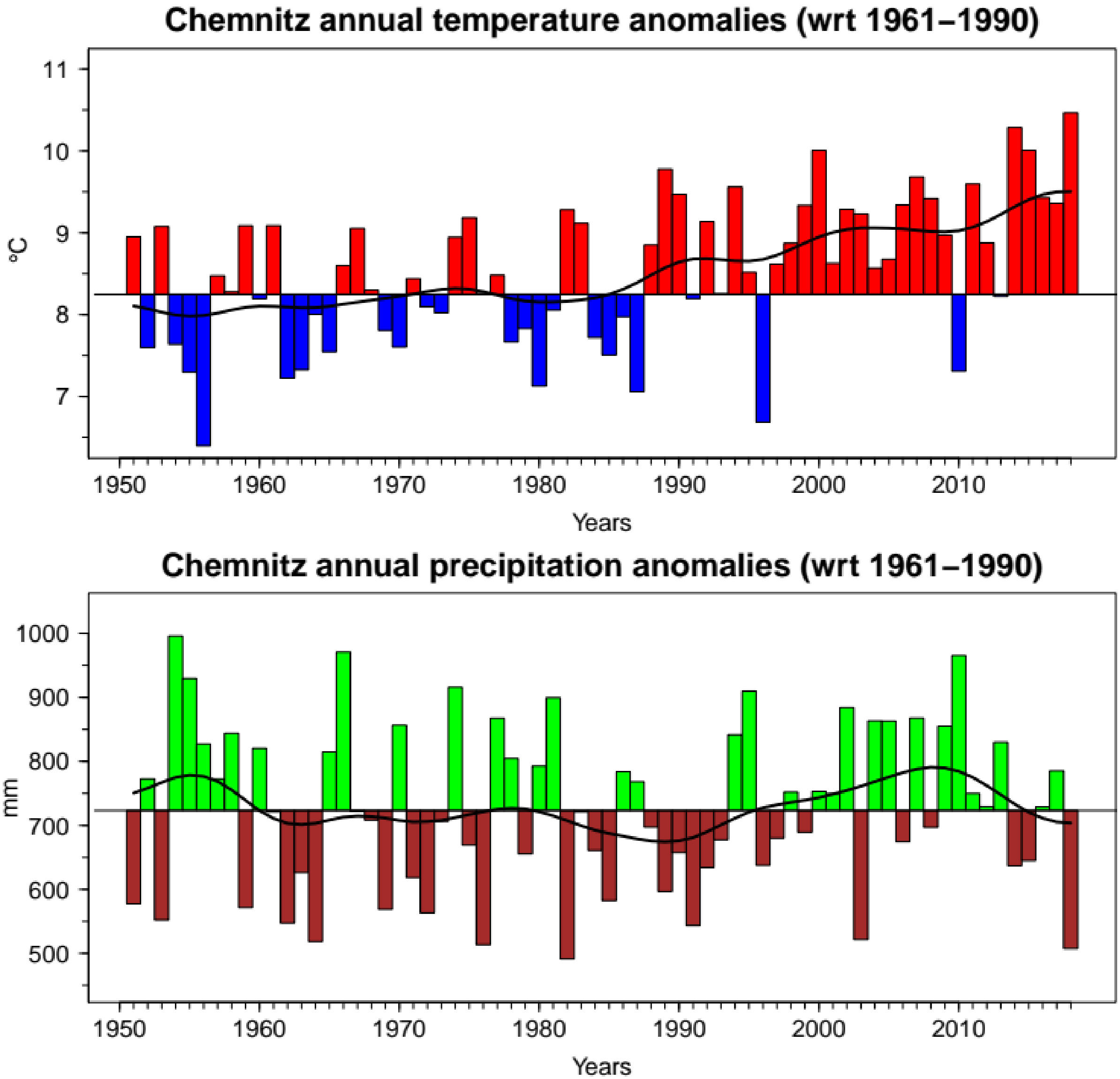

Figure 8. Plot of the annual anomalies of temperature and precipitation for the period 1951-2018 with respect to a 1961-1990 baseline (horizontal line). Warmer years are shown in red and cooler years in blue. Wetter years are shown in green and drier years in brown. The smooth curve is a 10-year Gaussian filter.

Once the analysis of the fitted model properties was complete, the required synthetic time series were generated. These comprised 100 samples, each of 30-years in length, of synthetic hourly rainfall.

Figures 13 and 14 show a comparison of the 100 30-year simulations with the 1995-2018 observations for a number of metrics. The red ranges indicate the two standard deviation (SD) ranges of the various metrics, with the blue cross marking the 1995-2018 average of the observations. The two SD range means that the range encompasses 95% of the distribution of the 100 simulations each of 30 years, so the top/bottom of the red bar is 97.5%/2.5%. As for Eelde, the blue crosses are within this range, but a few are close to the upper or the lower limit of the range. The more continental nature of the Chemnitz climate (compared to Eelde) is evident in the greater range in the summer half months of the inter-annual variability of half-monthly totals and in the mean wet day precipitation totals. This probably stems from some summers having many thunderstorms and some having markedly fewer.

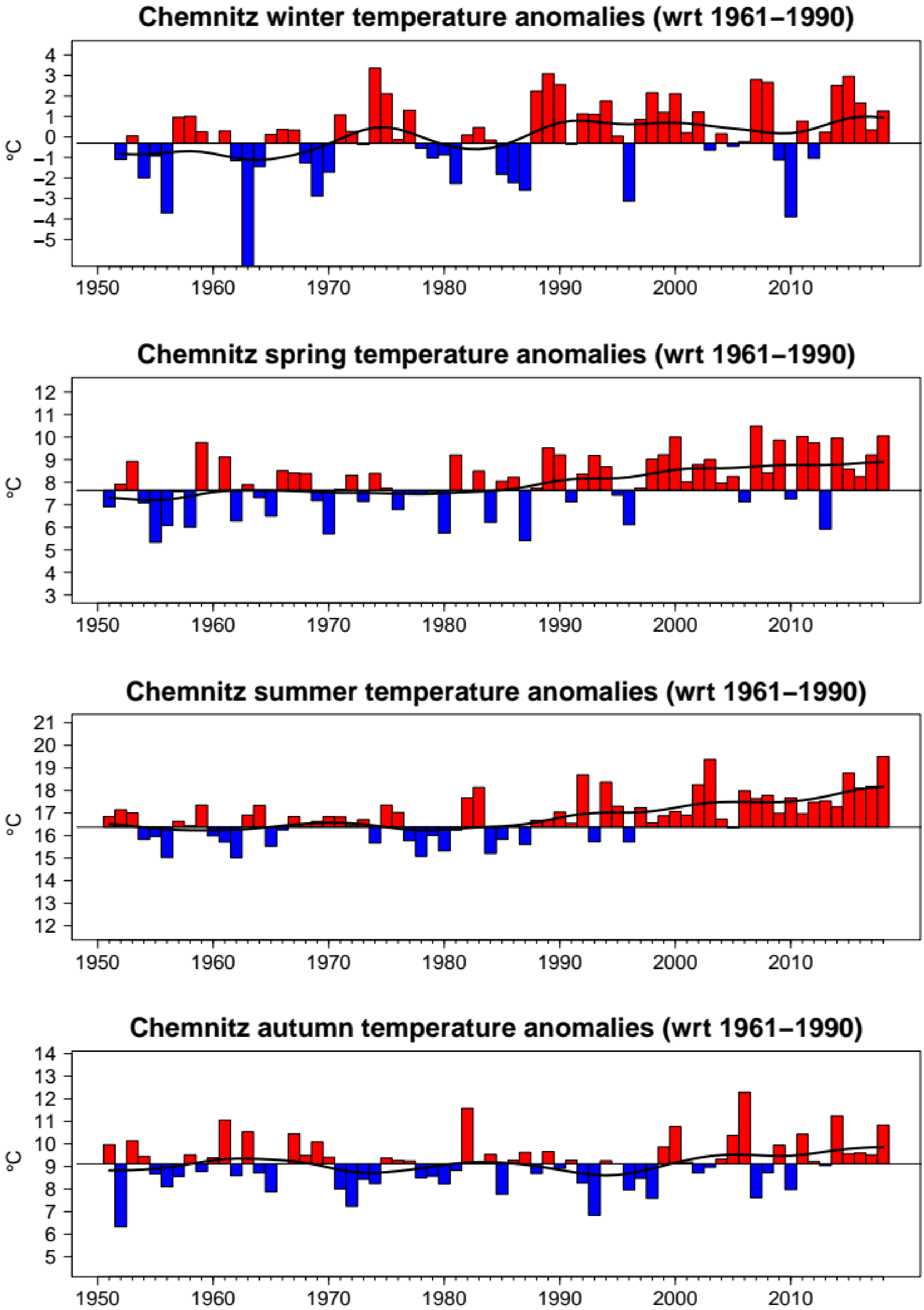

Figure 9. Plot of the seasonal temperature anomalies for the period 1951-2018 with respect to a 1961-1990 baseline (horizontal line). Warmer seasons are shown in red and cooler seasons in blue. The smooth curve is a 10-year Gaussian filter.

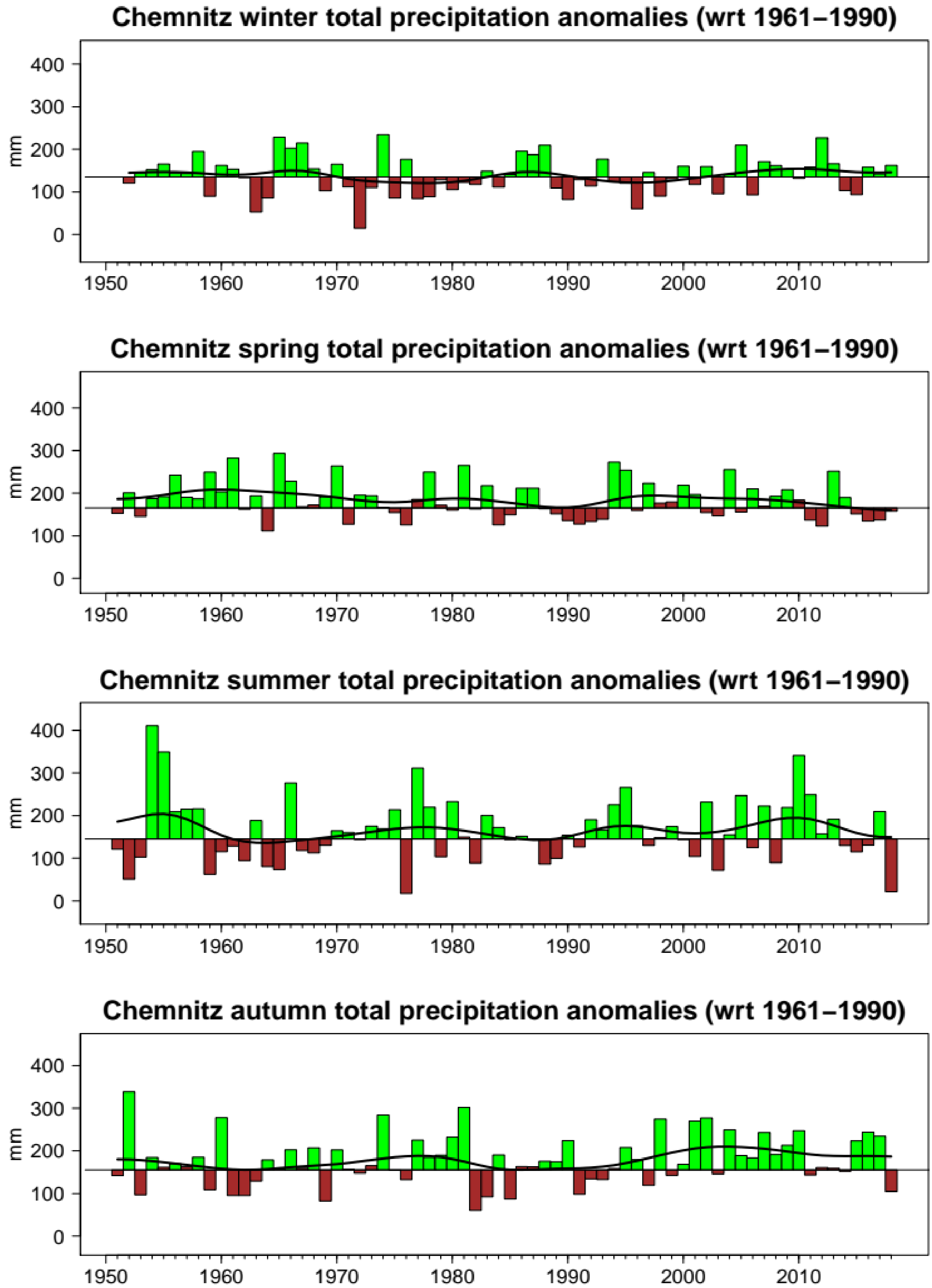

Figure 10. Plot of the seasonal precipitation anomalies for the period 1951-2018 with respect to a 1961-1990 baseline (horizontal line). Wetter seasons are shown in green and drier seasons in brown. The smooth curve is a 10-year Gaussian filter.

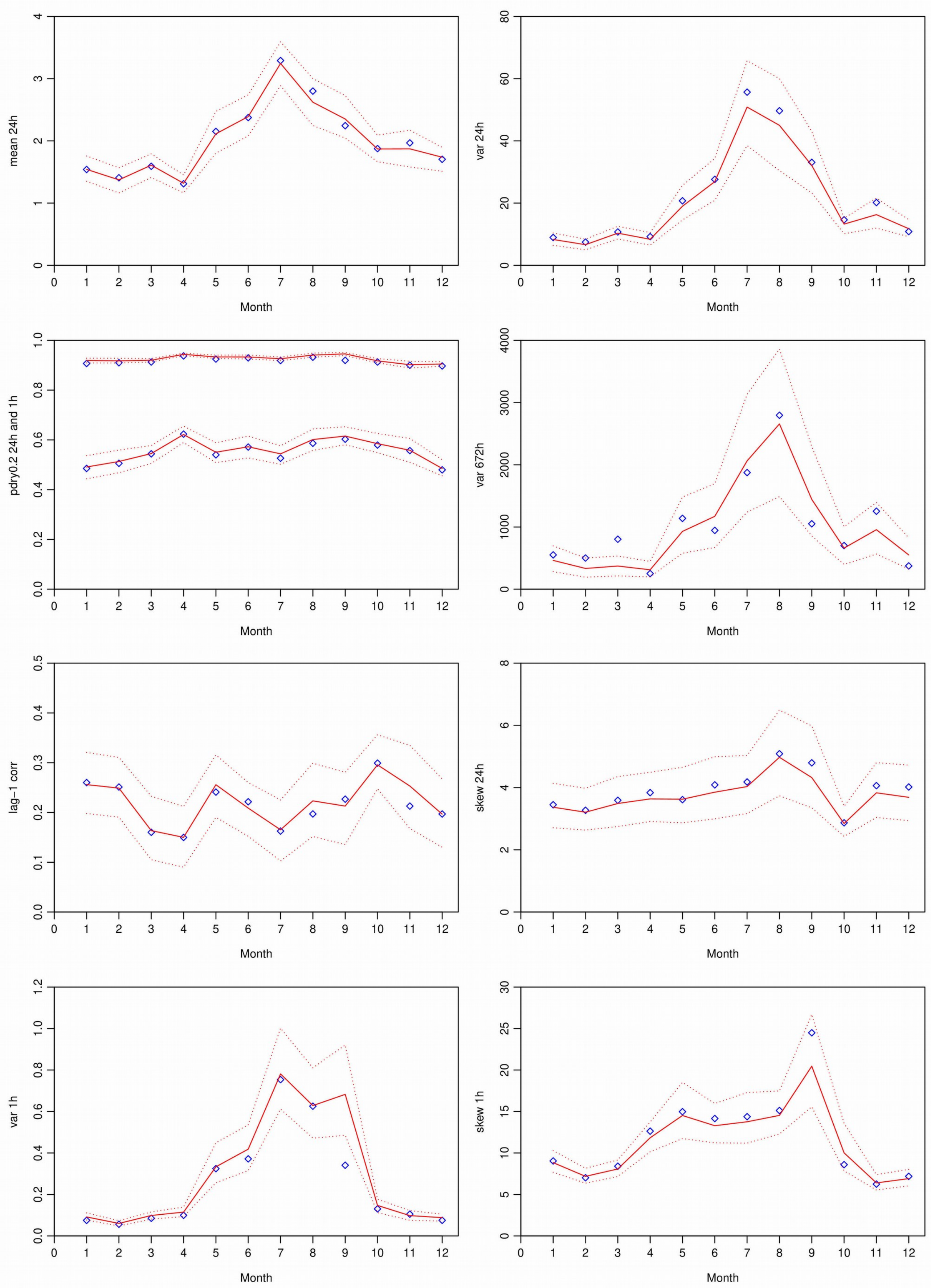

Figure 11. The observed and ensemble simulated rainfall statistics for Chemnitz by calendar month. Plots and units are: daily mean (mm), daily variance (mm^2), proportion dry (pdry) (-) daily (lower) and hourly (upper), 28-day variance (mm^2), daily autocorrelation (-), daily skewness coefficient (-), hourly variance (mm^2) and hourly skewness coefficient (-). The blue diamonds indicate the estimated observed statistics. The red lines indicate the mean and the 10^{th} and 90^{th} percentiles across a 100 member ensemble of 23-year synthetic time series.

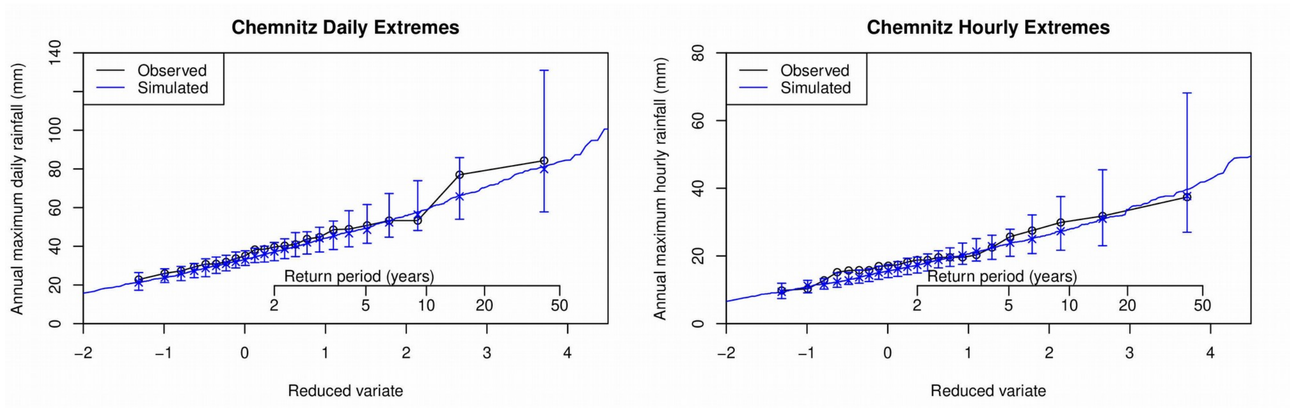

Figure 12. Extreme value plots showing observed and simulated annual maximum daily and hourly rainfall. The x-axis relates to the return period which is 1 / (frequency of exceedance). The observed extremes are plotted as black circles and lines. A synthetic 1000-year rainfall time series from the rainfall model was similarly analysed and plotted as the blue curve. 23-year subsamples of this series were then used to estimate the simulated variability at each plotting position (blue): the cross shows the median and the bar indicates the estimated 5 – 95 percentile range.

Synthetic Dataset Format

The synthetic rainfall dataset consists of 100 samples, each of 30-years of simulated rainfall. Nominally the hourly dataset begins at midnight on 1/1/3001. The format is single column with each value representing accumulations in mm. The year 3001 was selected to remind users of the synthetic data that the data *is* synthetic and does not correspond to a specific year (so for example it helps avoid the confusion that the data may be a weather forecast).

The daily dataset comprises 24-hour accumulations of the hourly data. The format is single column with each value representing accumulations in mm. Therefore each value in this dataset corresponds to the midnight to midnight period.

For the other daily weather variables, the data is space separated in 15 columns and can easily be read in by any programming language (e.g. C++, FORTRAN, R, Matlab etc.).

Column number:

1. Year
2. Month
3. Day
4. Day Count
5. Transition (for e.g. wet-wet to wet dry)
6. Rainfall (mm)
7. Temperature Minimum (^{o}C)
8. Temperature Maximum (^{o}C)
9. Vapour Pressure (hPa)
10. Relative Humidity (fraction of 1, multiply by 100 for %)
11. Wind Speed (m/s)
12. Sunshine Hours (hours)
13. Diffuse Radiation (kWh/m^2/day)
14. Direct Radiation (kWh/m^2/day)
15. PET (mm/day)

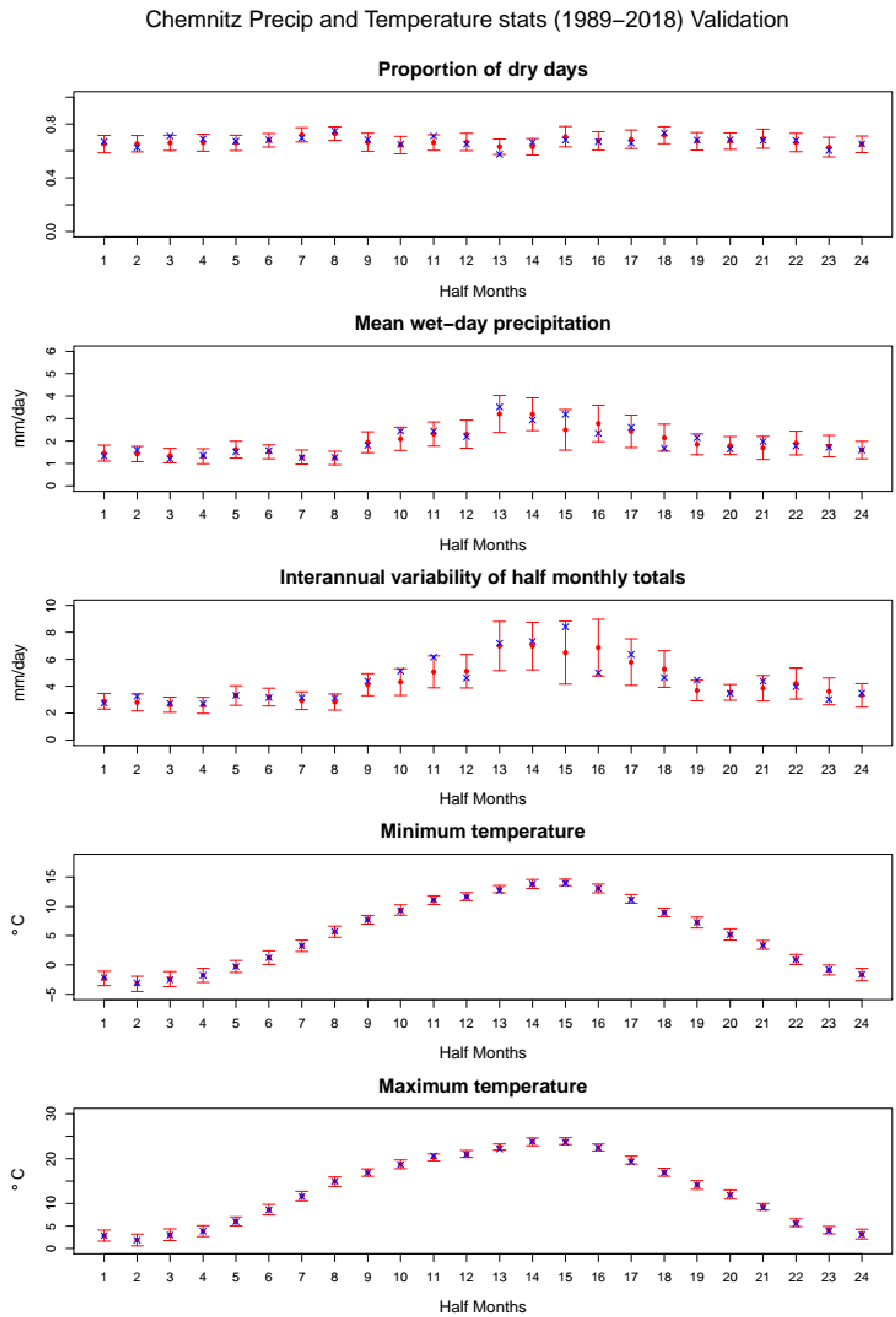

Figure 13. Daily precipitation and temperature validation plots comparing the annual cycle of observations (blue crosses) with the simulated properties (red bars). The x-axis scale indicates half months (e.g. 13 and 14 represent the first and second half of July). The red bars indicate the two standard deviation ranges of the simulated properties across the 100 simulation ensemble.

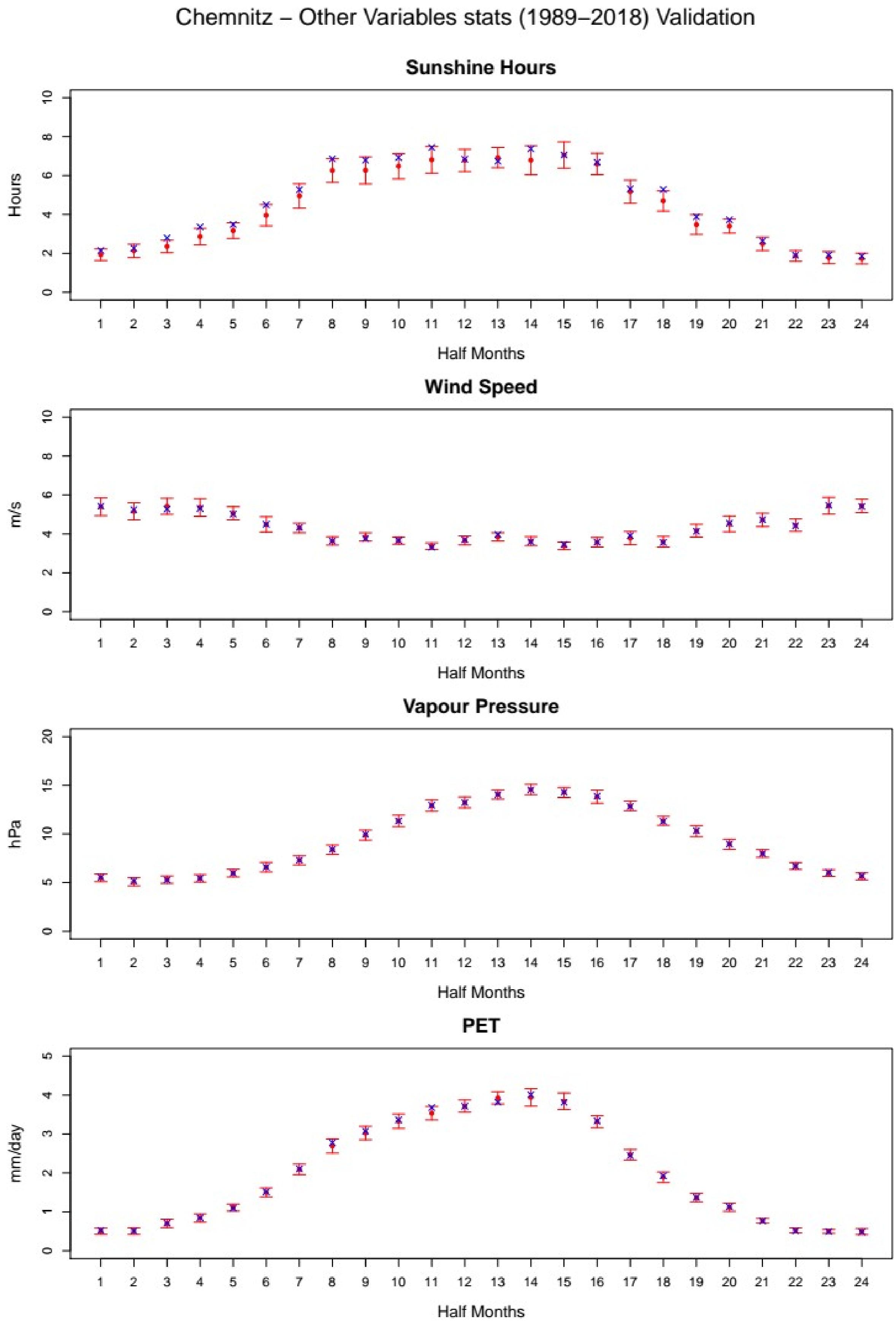

Figure 14. Daily sunshine, wind, vapour pressure and potential evapotranspiration (PET) validation plots comparing the annual cycle of observations (blue crosses) with the simulated properties (red bars). The x-axis scale indicates half months (e.g. 13 and 14 represent the first and second half of July). The red bars indicate the two standard deviation ranges of the simulated properties across the 100 simulation ensemble.

Summary

Synthetic ensemble weather generator datasets were prepared for the two selected European locations. In each case, the ensemble consisted of 100 30-year sets (ensemble members) of synthetic weather time series. These included hourly rainfall, and daily values for rainfall, minimum temperature, maximum temperature, vapour pressure, relative humidity, wind speed, sunshine hours, diffuse radiation, direct radiation and potential evapotranspiration.

Comparison of the properties of the synthetic ensemble weather datasets with those observed at the selected locations suggested that the synthetic data provide a reasonable match to the observed datasets for the various weather properties. Similarly, the extreme daily and hourly rainfall properties seem to be in excellent agreement with those observed.

Each ensemble dataset contains a long synthetic sample of consistently generated data suitable for the subsequent modelling of the implications of a wide range of plausible weather scenarios. Therefore some ensemble members exhibit synthetic events more extreme than those observed. The interpretation of this, however, is that the longer synthetic series provides an estimated representation of the sample variability of the observations rather than increasing the effective sample size beyond the ~30 years of observations. That is, increasing the length of the ensemble would not result in more accurate estimates of the implications of the actual weather. Nor should the synthetic data be used extrapolate to longer time periods (e.g. to estimate 100-year properties).

The synthetic ensemble datasets have been rigorously validated and provide a state of the art synthetic representation of the seasonal weather cycle and the extreme weather properties for each location as required. Single ensemble members may be unrepresentative, so each ensemble is designed to be used in its entirety or at least a representative sample of members should be used for the subsequent modelling.

Acknowledgements

The observed weather data for Eelde was obtained from the archive maintained by KNMI[6]. However, KNMI was not involved in and does not provide any endorsement of the derived works reported here. The observed weather data for Chemnitz, Germany, was provided by Deutscher Wetterdienst through their website[7]. This was analysed for its statistical properties which were used as described in this report.

6 http://projects.knmi.nl/klimatologie
7 http://www.dwd.de

References

Burton, A., Kilsby, C.G., Fowler, H.J., Cowpertwait, P.S.P., O'Connell, P.E. (2008) RainSim: A spatial-temporal stochastic rainfall modelling system, *Environmental Modelling & Software*, 23 (12), 1356-1369, doi:10.1016/j.envsoft.2008.04.003.

Deutscher Wetterdienst (2019a) *Datengewinnung*, in meta-data file named *Metadaten_Parameter_nieder_tag_00853.html*, generated 6/3/19.

Deutscher Wetterdienst (2019b) *Metadaten_Geographie_00853.txt,* a meta-data file dated 7/3/19

Jones, P.D., Harpham, C., Burton, A., and Goodess, C.M., (2016) Downscaling regional climate model outputs for the Caribbean using a weather generator, *International Journal of Climatology*, 36, 4141–4163, doi: 10.1002/joc.4624

Rasmussen*, S.B.,* Blenkinsop, S., Burton, A., Abrahamsen, P., Holm, P.E., Hansen, S., (2018) Climate change impacts on agro-climatic indices derived from downscaled weather generator scenarios for eastern Denmark, *European Journal of Agronomy*, 101, 222-238, doi: 10.1016/j.eja.2018.04.004.