Journal of Hydrology: Regional Studies 15 (2018) 184–202

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Spatial interpolation of climate variables in Northern Germany—Influence of temporal resolution and network density

C. Berndt*, U. Haberlandt

Institute of Hydrology and Water Resources Management, Leibniz Universität Hannover, Appelstr. 9A, 30167, Hanover, Germany

A R T I C L E I N F O

Keywords:
Climate data
Geostatistics
Kriging
Interpolation

A B S T R A C T

Study region: Region in Lower Saxony (North Germany) covered by the measuring range of the weather radar device located near Hanover (approx. 50.000 m^2).

Study focus: This study investigates the performance of various spatial interpolation techniques for climate variables. Meteorological observations are usually recorded as site-specific point information by weather stations and estimation accuracy for unobserved locations depends generally on station density, temporal resolution, spatial variation of the variable and choice of interpolation method. This work aims to evaluate the influence of these factors on interpolation performance of different climate variables. A cross validation analysis was performed for precipitation, temperature, humidity, cloud coverage, sunshine duration, and wind speed observations. Hourly to yearly temporal resolutions and different additional information were considered.

New hydrological insights: Geostatistical techniques provide a better performance for all climate variables compared to simple methods Radar data improves the estimation of rainfall with hourly temporal resolution, while topography is useful for weekly to yearly values and temperature in general. No helpful information was found for cloudiness, sunshine duration, and wind speed, while interpolation of humidity benefitted from additional temperature data. The influences of temporal resolution, spatial variability, and additional information appear to be stronger than station density effects. High spatial variability of hourly precipitation causes the highest error, followed by wind speed, cloud coverage and sunshine duration. Lowest errors occur for temperature and humidity.

1. Introduction

Climate or weather information more generally is usually recorded as site-specific point information by meteorological stations. However, the modelling of many processes in hydrology or environmental science requires areal input data, or in many cases, data for unobserved locations is needed. Spatial interpolation techniques are a reliable approach in order to estimate climate information for unobserved locations from nearby measurements.

Many different techniques have been proposed for various climate information, while the estimation performance depends not only on the selected interpolation method, but also on other factors like station network configuration, temporal data resolution, spatial variability of the variable, and whether a useful additional information can be incorporated into the interpolation procedure.

First investigations towards the issue of rainfall interpolation were carried out by Thiessen (1911), who used polygons drawn around the locations of rain gauges on a station network map in order to obtain an estimation of rainfall based on the nearest

* Corresponding author.
E-mail address: berndt@iww.uni-hannover.de (C. Berndt).

https://doi.org/10.1016/j.ejrh.2018.02.002
Received 17 October 2017; Received in revised form 1 February 2018; Accepted 3 February 2018
Available online 20 February 2018

neighbouring station. Shepard (1968) proposed a technique, wherein the estimate is computed as a weighted average of adjacent rain gauges. The impact of each station is defined according to the inverse of its distance from the location to be estimated. Geostatistical methods like Kriging allow the consideration of the spatial correlation of adjacent observations for the estimation of unknown locations. Several studies reported that Ordinary Kriging can outperform simpler approaches (Dirks et al., 1998; Phillips et al., 1992; Tabios and Salas, 1985). Moreover, methods based on spline fitting have been applied (Hutchinson, 1998a). The fitting of three-variate splines even allows for the incorporation of elevation data (Hutchinson, 1998b). The consideration of elevation as additional information resulted in a significant improvement of interpolation performance at certain time scales. Annual or mean annual precipitation was predicted using regression (Daly et al., 1994; Nalder and Wein, 1998) and geostatistics (Hevesi et al., 1992a,b; Lloyd, 2005; Martínez-Cob, 1996). A study by Goovaerts (2000) reported that the incorporation of elevation in various Kriging methods can outperform linear regression as well as univariate interpolation for the estimation of monthly and annual rainfall.

Quantitative precipitation estimates from weather radar was proven to be useful additional information in the interpolation of short-term rainfall. However, radar data tends to be strongly biased (Seo et al., 1999) and is prone to a variety of different measuring errors, for instance a variation in the relationship between reflected energy and rainfall intensity, changes in precipitation particles before reaching the ground, anomalous beam propagation, attenuation, and clutter. Haberlandt (2007) as well as Verworn and Haberlandt (2011) applied Kriging with External Drift in addition to Indicator Kriging with External Drift for hourly data and achieved an improvement compared with univariate techniques. Cokriging was applied by Krajewski (1987) and provided slightly better estimation compared to using rain gauges only. A further technique to combine radar and rain gauge data is the so called Conditional Merging approach reported by Ehret (2003). According to Sinclair and Pegram (2005), it can efficiently reduce bias and error variance of quantitative precipitation estimates. Goudenhoofdt and Delobbe (2009) compared different merging algorithms using daily rainfall data and preferred geostatistical techniques over univariate rain gauge interpolation and radar data adjustment techniques like mean field bias correction (Smith and Krajewski, 1991). Overall, Kriging with External Drift performed best. Berndt et al. (2014) reported that Conditional Merging outperforms Kriging with External Drift and Indicator Kriging with External Drift for temporal resolutions from 10 min to 360 min.

Spline fitting (Price et al., 2000) as well as regression-based approaches (Nalder and Wein, 1998) taking into account the elevation were also applied for the estimation of air temperature and provided a better estimation quality than simple methods. Other techniques are based on a linear lapse rate (Dodson and Marks, 1997). Stahl et al. (2006) compared different approaches based on lapse rates and spatial interpolation They found that a combination of computing a regression based lapse rate and inverse-distance weighting performed best. Hudson and Wackernagel (1994) applied Kriging with External Drift for January averages of temperature and concluded that the incorporation of topography results in a substantial improvement of interpolation performance compared to univariate interpolation of station data. Other climate variables such as wind speed, humidity, and sunshine duration are less often studied regarding spatial interpolation issues. However, interpolation is often performed if exhaustive data sets are generated for a certain region (Jeffrey et al., 2001; Li et al., 2014).

Most of the previous work focused in improving the interpolation performance for one specific temporal resolution of a specific climate variable. A comparison of interpolation performances for different climate variables is difficult to find in the literature. Only some studies compare the interpolation performance among different station densities (Goudenhoofdt and Delobbe, 2009; Krajewski, 1987; Nanding et al., 2015; Yoon et al., 2012) and even fewer among different time scales (Bárdossy and Pegram, 2013; Dirks et al., 1998), although network density is considered to have a strong impact on the estimation accuracy. Additionally, the spatial variability of certain climate information depends significantly on the accumulation time. A combined evaluation of all influencing factors in order to provide a guidance for the choice of interpolation method depending on study area, climate variable, network configuration, temporal resolution, and intended data use is not available. This paper aims at evaluating the different impacts on the interpolation performance for a region in North Germany. Geostatistical techniques as well as simple methods are considered in the cross validation experiments that were performed here for observations of precipitation, temperature, relative humidity, cloud coverage, sunshine duration, and wind speed. Different station network density scenarios as well as temporal data resolutions ranging from 1 h to 1 year were taken into account.

The paper is organised as follows. Chapter 2 contains a brief description of all interpolation techniques. Geostatistical approaches were considered since they are widely applied in hydrology and environmental science. Simpler techniques, specifically Nearest Neighbour and Inverse-Distance Weighting are included as a standard for comparison. Chapter 3 describes the study region as well as data and their pre-processing. The cross validation strategy considering different network densities and the performance assessment of spatial estimation are presented in Ch. 4. All results of the analysis are shown and discussed in Ch. 5, while the main findings and conclusions are summarised in Ch. 6.

2. Interpolation methods

2.1. Simple interpolation techniques

The nearest neighbour interpolation technique (NN), also known as the Thiessen polygon method (Thiessen, 1911), is a basic interpolation approach that is often used for the spatial interpolation of rainfall data. It can be easily applied for the interpolation of other meteorological variables as well. Each location to be estimated within the regarded domain is simply assigned with the closest available observation.

Inverse-distance weighting (Shepard, 1968) is able to account for a simple spatial dependency in the interpolation of point observations. It does not require an a priori investigation of spatial variability for the regarded variable, in contrast to the more sophisticated geostatistical approaches. The estimation of each unknown location is calculated according to:

$$Z^*(u_0) = \sum_{i=1}^{4} \lambda_i Z(u_i) \tag{1}$$

The estimate Z^* at location u_0 is calculated as a linear combination of the closest four measurements. One station is taken into account for each quadrant based on u_0 (northeast, southeast, southwest, northwest). The weights λ_i are calculated according to the inverse of the squared distance of the corresponding station from u_0.

$$\lambda_i = \frac{\frac{1}{d(u_0, u_i)^2}}{\sum_{i=1}^{4} \frac{1}{d(u_0, u_i)^2}}, \tag{2}$$

with $d(u_0, u_i)$ representing the distance between the locations u_0 and u_i.

2.2. Geostatistical approaches

This section gives a short summary of the main concepts and basic assumptions used in geostatistical interpolation methods. For further reading regarding the mathematical as well as statistical background the reader is referred to geostatistical textbooks, for instance Goovaerts, (1997) and Isaaks and Srivastava (1990). All computations shown in this section were performed using the Geostatistical Software Library (GSLIB) by Deutsch and Journel (1992).

2.2.1. Variogram estimation

All geostatistical interpolation techniques need information about the spatial persistence of the variable prior to performing any estimations for unknown locations. The investigation is carried out by calculating the empirical variogram and fitting a theoretical model. In order to achieve this, the difference in point pair data values is investigated depending on their spatial distance:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(u_i) - z(u_j)]^2, \tag{3}$$

with $\gamma(h)$ being the variogram value of the distance class h, $N(h)$ representing the number of available point pairs belonging to h, as well as $z(u_i)$ and $z(u_j)$ being measured values of the variable at the locations u_i and u_j separated approximately by h. This spatial dependency measure is widely applied in geostatistics. Due to reasons of simplification, the semivariogram is referred to as variogram henceforth.

Theoretically, a variogram model needs to be fitted for each time step of the variable that is interpolated. However, previous research by Verworn and Haberlandt (2011) showed that the choice of variogram has only a minor impact on the performance when continuous time series are interpolated. They found that the use of event-specific averaged variograms delivers a similar interpolation performance compared to using an individual variogram for each time step. Due to this, average variograms were used here as well. A standardisation with the spatial variance was carried out for each time step prior to averaging:

$$\gamma_{avg}(h) = \frac{1}{n} \sum_{i=1}^{n} \frac{\gamma_i(h)}{Var(z_t)}. \tag{4}$$

In this equation n is the number of time steps, $\gamma_i(h)$ the variogram value for distance class h and of time step i and $Var(z_t)$ is the variance of observations z_t for time step t.

The exponential variogram model was used for all climate variables:

$$\gamma_h = c_0 + c\left[1 - \exp\left(-\frac{h}{a}\right)\right], \tag{5}$$

where a is the range, c the partial sill and c_0 the nugget effect.

The R package NLME (Pinheiro et al., 2016) was used for obtaining all parameters by an automatic fitting procedure. A least-squares approach was used while the same weight was assigned to each distance class.

2.2.2. Ordinary Kriging

Ordinary Kriging (OK) is the most common and most frequently applied interpolation technique in geostatistics. It is classified as a univariate approach, i.e. it only allows for the consideration of one data source and no additional information can be taken into account. An important assumption is that the expected value of the underlying random function is equal within the entire domain. The OK estimate is generally unbiased and calculated as follows:

$$Z^*(u_0) = \sum_{i=1}^{n} \lambda_i Z(u_i),$$

where λ_i is the weight of each of the n adjacent observations taken into account. The weights are obtained by solving the kriging system:

$$\begin{aligned} &\sum_{j=1}^{n} \lambda_j \gamma(u_i - u_j) + \mu = \lambda\gamma(u_i - u_0) \quad \text{for} \quad i= 1,\dots,n \\ &\qquad\qquad \sum_{j=1}^{n} \lambda_j = 1 \end{aligned}, \tag{6}$$

here μ is a Lagrange multiplier.

2.2.3. Kriging with external drift

Kriging with External Drift (KED) allows the incorporation of one or more additional variables that are used as background information for the interpolation of the primary variable. KED assumes that the expected value of the random function is linearly related with m additional variables $Y_k(u)$, $k = 1, \dots, m$:

$$E[Z(u)|Y_1(u), Y_2(u), .., Y_m(u)] = b_0 + \sum_{k=1}^{m} b_k Y_k(u), \tag{7}$$

where $b_0, b_1, \dots, b_m$ are unknown constants.

The same estimator as for OK is used for computing the KED estimates. However, the kriging system for determining the weights is changed as follows:

$$\begin{aligned} &\sum_{j=1}^{n} \lambda_j \gamma(u_i - u_j) + \mu_0 + \sum_{k=1}^{m} \mu_k Y_k(u_i) = \lambda\gamma(u_i - u_0) \quad \text{for} \quad i= 1,\dots,n \\ &\qquad\qquad \sum_{j=1}^{n} \lambda_j = 1 \\ &\qquad\qquad \sum_{j=1}^{n} \lambda_j Y(u_j) = Y(u) \end{aligned}, \tag{8}$$

where n is he number of neighbours, m is the number of additional variables Y_k and μ_k are the $m + 1$ Lagrange multipliers.

Theoretically, the variogram for KED should be estimated from the residuals $Z(u) - m(u)$. However, this is usually not a simple procedure since neither the residuals nor the trend $m(u)$ is known a priori. As it was also performed by Haberlandt (2007), the experimental variograms for KED were inferred by a simplified approach, i.e. using only the observations $Z(u)$. Delrieu et al. (2014) compared three different variograms for the merging of rain gauge and radar data using KED. They found that the use of rainfall variograms, as they are also required for OK, results in a similar interpolation performance like the use of residual variograms obtained by applying the method of Velasco-Forero et al. (2009). Only a pure nugget effect variogram resulted in a significant worsening of interpolation performance.

The KED procedure is applied for each time step independently when time series are interpolated. The coefficients $b_0, b_1, \dots, b_m$ of Eq. (7) will thus vary in space and time, which allows the consideration of a space-time variable relationship between the primary variable and any additional information.

2.2.4. Conditional merging

Conditional Merging (CM) is a specific approach for merging radar and rain gauge data. It was first described by Ehret (2003) and later adapted for simulated rainfall fields by Sinclair and Pegram (2005). According to Berndt et al. (2014) it outperformed other techniques for the interpolation of 10 min to 360 min rainfall accumulations. The CM estimate is computed according to the following stepwise procedure:

(1) Rain gauge data (point information) and gridded radar data are available.
(2) Rain gauge observations are interpolated by OK onto the radar grid.
(3) Radar pixel information are extracted for all rain gauge locations.
(4) Pixel information at gauge locations are interpolated by OK onto the radar grid
(5) The deviation between interpolated radar pixel information (step 4) and the original radar field (step 1) is computed at each grid point.
(6) The deviation grid (step 5) is added to the interpolated rain gauge field from step 2.

3. Study region and data

3.1. Study region and observation networks

The study region is located within the 128 km range of the radar station Hanover in Lower Saxony, North Germany. Fig. 1 shows the location of the study area. It is identical to the measuring range of the radar device located in Hanover with an area of approx. 50.000 m^2 and covers a large area of the German federal state Lower Saxony as well as the entire federal city state of Bremen.

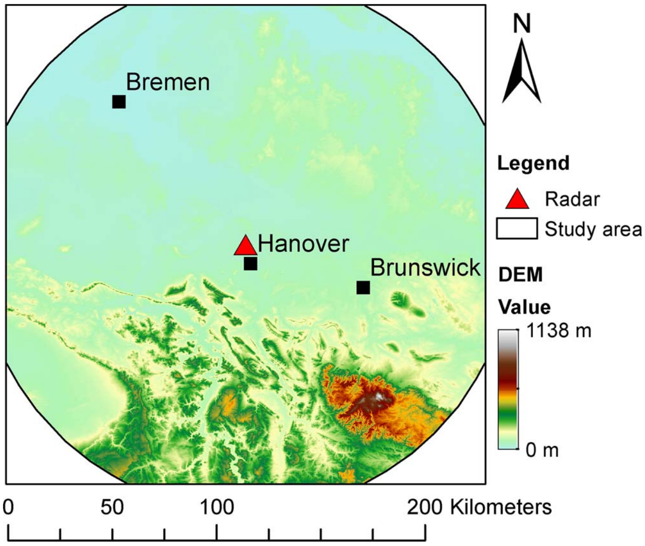

Fig. 1. Study area.

Moreover, small areas of neighbouring federal states are included.

The northern part of the study area can be characterised as entirely flat, being located in the North German Plain. The Harz Mountains are found towards the southeast of the study area and have a maximum elevation of 1141 m.a.s.l., according to the digital elevation model of the German Federal Agency for Cartography and Geodesy (BKG), which has been used in this study.

Table 1 contains minima and maxima of long-term averages for all meteorological variables considered in this study. The ranges of the variables were determined from the daily station data provided by the German Weather Service (Deutscher Wetterdienst, DWD). For each station, the long-term average (temperature, relative humidity, sunshine duration, wind speed, cloudiness) or the annual sums (rainfall) were computed. Minimum as well as maximum value within the study area are reported.

The DWD operates different nation-wide station networks, i.e. point related weather information is recorded at various locations. The time period from 2008 to 2013 was used for the cross validation investigations of this study (see Section 4). This time span provided a relative constant number of stations for all meteorological variables except for the cloud coverage recordings. The analysis of cloud coverage was conducted using a shorter investigation period, i.e. the years 2009, 2010 and 2011. The data of those stations, which do not cover the complete study period, were not taken into account. However, stations with time series that contained missing values for single time steps were considered in the evaluations. Table 2 contains a summary of all climate variables with the respective number of stations. All data was obtained from the Climate Data Center of the DWD. The observation networks are presented in Fig. 2. A different number of stations is available for each meteorological variable.

3.2. Data pre-processing

Daily rainfall observations were converted to longer time scales by accumulation to weekly, monthly and annual rainfall sums. The daily observations of other climate variables were averaged for the same time scales. These long-term accumulations and averages were used together with the hourly recordings for the cross validation analyses.

Radar data of the C-band instrument at Hanover were provided as raw reflectivities with an azimuth resolution of 1° and a time discretisation of 5 min (dx-product of the DWD). The pre-processing is only briefly summarised here. The same procedure as in Berndt et al. (2014) was applied.

A clutter correction was performed and the radar reflectivities were converted to rainfall intensities using a constant Z-R-relationship according to:

Table 1
Maximum and minimum value within the study area for long-term station averages of all meteorological variables.

Variable	Minimum	Maximum
Annual rainfall [mm]	500	1700
Temperature [∘C]	3.5	10.0
Relative humidity [%]	73	90
Sunshine duration [h]	3.8	4.8
Wind speed [m/s]	1.5	4.5
Cloud coverage [Okta]	4.0	6.1

Table 2
Number of available stations for the investigation period.

Climate variable	Abbreviation	Number of stations	
		Hourly	Daily
Precipitation	PCP	92	202
Mean temperature	TAV	39	39
Relative humidity	HUM	38	38
Wind speed	WVE	25	25
Sunshine duration	SUN	25	25
Cloud coverage	CLD	18	38

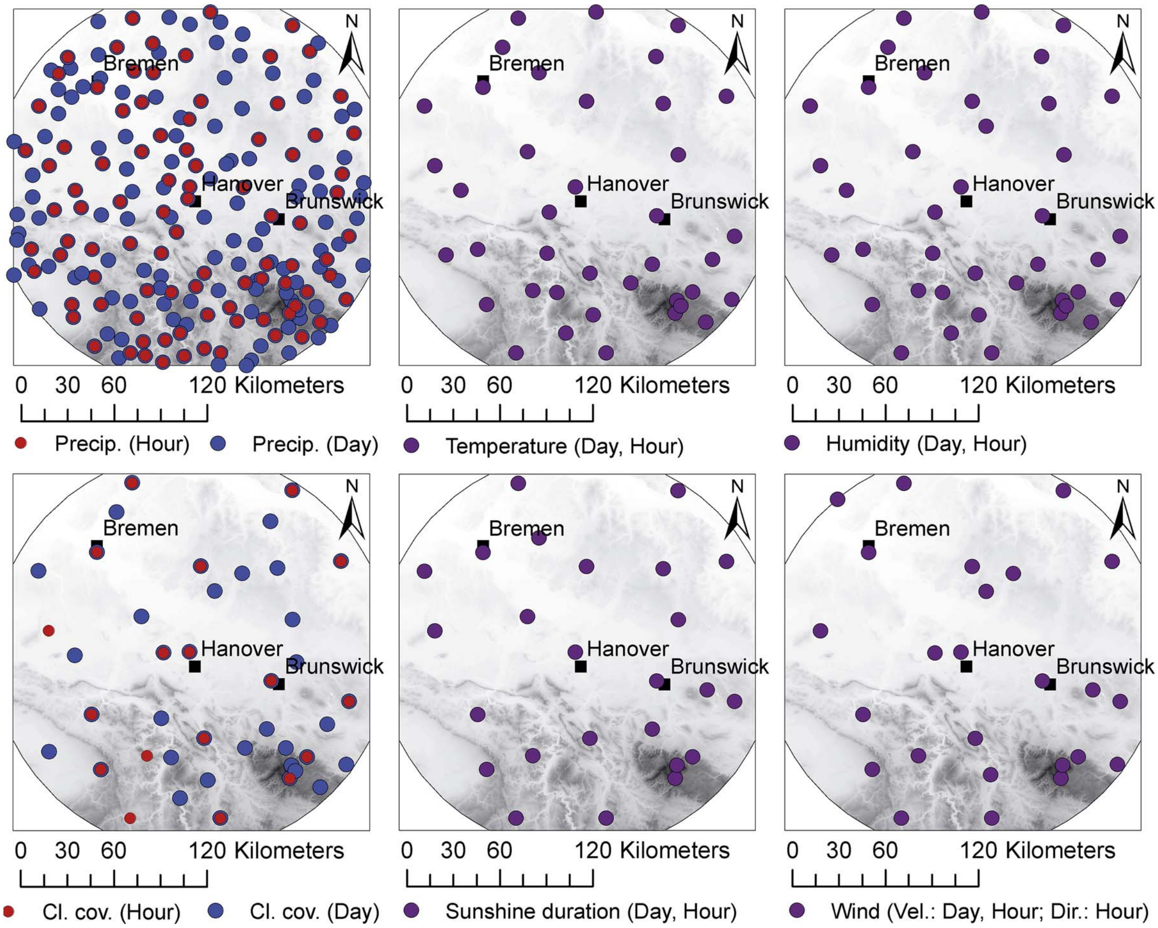

Fig. 2. Meteorological observation networks.

$$Z = aR^b, \tag{9}$$

where Z is the reflectivity in mm^6m^{-3} and R the rainfall intensity in mm/h. The parameters were set to $a = 256$ and $b = 1.42$ according to the Standard-DWD-relationship (Riedl, 1986; Seltmann, 1997).

Thereafter, a transformation from polar coordinates to Cartesian coordinates was performed. All non-clutter observation points were interpolated on a 1 km × 1 km grid by using inverse-distance weighting. The gridded rainfall intensities in mm/h obtained by application of the Z-R-relationship were converted into the corresponding 5 min rainfall depths and a spatio-temporal smoothing filter was applied in the same way as in Berndt et al. (2014). In the last step, the smoothed 5 min radar grids were accumulated to all required temporal resolutions.

4. Cross validation strategy

The interpolation performance was assessed by applying “leave-one-out” cross validation. This method is based on a simple

Table 3
Station density scenarios for interpolation performance evaluation.

Variable	Temp. res.	17 Stat.	24 Stat.	36 Stat.	56 Stat.	70 Stat.	90 Stat.	200 Stat.
PCP	1 h	✓	✓	✓	✓	✓	✓	–
PCP	1 d – 1 yr	✓	✓	✓	✓	✓	✓	✓
TAV	1 h – 1 yr	✓	✓	✓	–	–	–	–
HUM	1 d – 1 yr	✓	✓	✓	–	–	–	–
WVE	1 h – 1 yr	✓	✓	–	–	–	–	–
SUN	1 h – 1yr	✓	✓	–	–	–	–	–
CLD	1 h	✓	–	–	–	–	–	–
CLD	1 d – 1 yr	✓	✓	✓	–	–	–	–
Stations per 10.000 m^2:		3.58	5.05	7.79	11.78	14.73	18.98	42.08

principle: a successive estimation of all sampled locations is performed by using all other stations while always excluding the sample value at the regarded location.

4.1. Network density scenarios and temporal resolutions

The cross validation experiments were performed for a wide range of network densities and temporal resolutions. The time scales considered here are based on the temporal resolution of the observed data. However, coarser temporal resolutions were analysed as well in order to obtain a comparison of interpolation performance among different time scales.

The network density scenarios were defined according to the number of available stations for the investigation period. Cross validation using scenarios containing the same number of stations, for instance 17 and 24, allows a comparison of interpolation performance among all climate variables. Table 3 contains an overview about all network density scenarios taken into account.

For each network density scenario, the stations were selected randomly from the total available number of time series for each climate variable. Due to the variability of available stations among the climate variables, the network density scenarios could not be defined in a way so that the exact same stations are used for each meteorological information. In order to achieve a better comparability of interpolation performance among different climate information, the cross validation analysis was based on ten realisations of each network density scenario, i.e. the performance evaluation was carried out for ten different random subsamples that were drawn from the entire set of stations.

For each realisation, the random selection of stations is designed in a way that the 17 stations of the lowest network density scenario are present in all other network scenarios as well. The computation of the performance criteria (see Section 4.2) only takes into account observed and estimated values at these 17 locations. However, depending on the network density scenario, more stations might have been used for calculating the estimates. This slightly modified cross validation procedure allows the comparison of interpolation performances for different density scenarios. Using a different number of stations for calculating the interpolation performance of each network density scenario, i.e. the entire number of stations in the scenario, leads to a variation in the validation sample and impairs the comparability among the density scenarios.

4.2. Performance assessment

The following performance measures were used to compare the estimations Z^* and observations Z for the n locations:

The simple bias criterion

$$\mathrm{Bias} = \frac{1}{n}\sum_{i=1}^{n}\left[Z^*(u_i) - Z(u_i)\right], \tag{10}$$

The root mean square normalised with the average of the observations

$$\mathrm{RMSE} = \frac{1}{\bar{Z}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[Z^*(u_i) - Z(u_i)\right]^2}, \tag{11}$$

and the RVar coefficient, which indicates the preservation of variance of observed information

$$\mathrm{RVar} = \frac{\mathrm{Var}[Z^*(u_i)]}{\mathrm{Var}[Z(u_i)]}. \tag{12}$$

5. Analyses and results

5.1. Variogram inference

Fig. 3 shows all theoretical variogram models that were inferred for this study using always all available stations for each climate

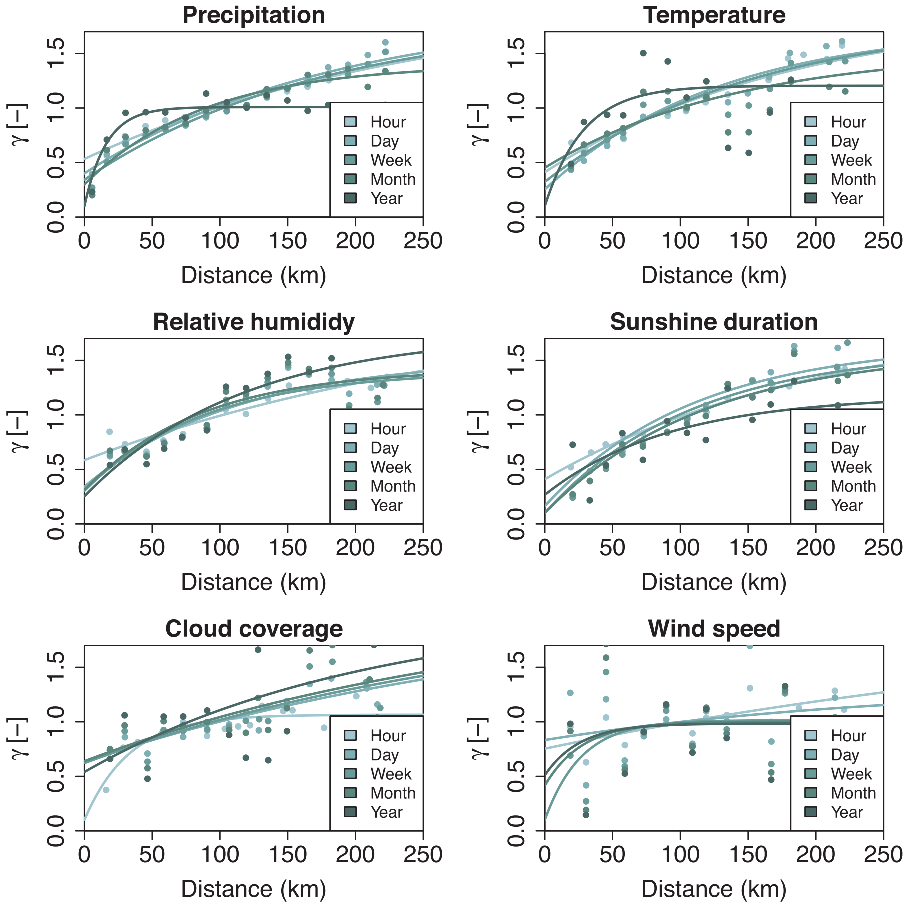

Fig. 3. Experimental variograms and fitted theoretical models for all climate variables and temporal resolutions.

Table 4
Parameters of the exponential variogram model for all temporal resolutions of each climate variable.

Precipitation	c_0 [−]	c [−]	a_{eff} [m]	Temperature	c_0 [−]	c [−]	a_{eff} [m]
Hour	0.53	1.50	259,915	Hour	0.41	1.50	187,094
Day	0.40	1.50	186,280	Day	0.30	1.50	141,143
Week	0.34	1.50	176,863	Week	0.32	1.50	153,716
Month	0.30	1.11	88,965	Month	0.45	1.09	141,324
Year	0.10	0.91	16,211	Year	0.30	0.91	34,559

Rel. humidity	c_0 [−]	c [−]	a_{eff} [m]	Sunshine dur.	c_0 [−]	c [−]	a_{eff} [m]
Hour	0.58	1.5	315,585	Hour	0.41	1.50	207,469
Day	0.34	1.18	112,967	Day	0.16	1.50	110,546
Week	0.30	1.10	88,222	Week	0.09	1.50	105,682
Month	0.31	1.11	85,854	Month	0.04	1.50	103,387
Year	0.25	1.50	117,198	Year	0.27	0.91	92,205

Cloud cov.	c_0 [−]	c [−]	a_{eff} [m]	Wind speed	c_0 [−]	c [−]	a_{eff} [m]
Hour	0.10	0.97	33,301	Hour	0.75	1.31	500,000
Day	0.64	1.50	359,219	Day	0.83	0.52	258,859
Week	0.62	1.50	327,490	Week	0.10	0.91	26,815
Month	0.64	1.50	318,020	Month	0.42	0.59	27,898
Year	0.54	1.50	210,505	Year	0.51	0.47	26,026

variable (see Table 2). The variogram models are grouped by climate variable and the variograms for different temporal resolutions are correspondingly shown in the same panel. All parameters of the exponential model are shown in Table 4. An identical automatic fitting procedure was used for all variograms (see Section 2.2.1). A maximum value of 1.5 was used as the upper limit of the partial sill c in order to allow a stable fitting.

For precipitation, only time steps with an average of 0.1 mm or higher were taken into account for the calculation of empirical variograms. This ensures that time steps with no rainfall are omitted. There is a general decrease of the nugget c_0 with increasing time step. In addition, the results show a relationship between time scale and variogram range: the lower the temporal resolution of the data, the lower the effective range a_{eff} of the variogram model. The range obtained for the annual temporal resolution is particularly low. This surprising behaviour might be caused by the automatic fitting procedure that has a strong impact on the resulting nugget parameter. Moreover, the exponential variogram model is highly adjustable for low distance classes and accordingly the third data point (Distance = 30 km, γ = 0.96) is affecting the shape of the model in a substantial manner.

In the case of mean annual temperature, the fitting of the theoretical model is clear and simple for the hourly and daily time scale. However, the spatial persistence of monthly and annual data is not as obvious. Although there is a general increase of variogram value with increasing distance, the variogram points scatter heavily around the theoretical model. It can be concluded that the spatial persistence of annual mean temperature is lower than the spatial persistence of temperature averages for shorter periods. The corresponding range parameter also indicates a stronger spatial dependency for short aggregation periods. In contrast to rainfall, there is no clear influence of the temporal resolution on the nugget effect c_0. The lower parameter values of the partial sill c for monthly and annual data might result from the automatic fitting in combination with the indistinctive behaviour of the experimental variogram.

A clear spatial persistence is present in the recordings of relative humidity. However, the variograms tend to have relatively high nugget values, in particular for the temporal resolution of 1 h. The variogram obtained for hourly data has a higher range a_{eff} compared to the variograms for other temporal resolutions. It appears that the mean slope of the variogram is increasing with increasing time interval, i.e. the annual humidity has a steeper incline of variogram value than the hourly humidity.

Due to the high number of missing cloudiness values, usually occurring at night, a threshold of available cloud coverage recordings was established in order to filter out time steps with too few observations. Twelve non-missing recordings had to be available for a time step to be taken into account. Like with temperature, the spatial persistence is more obvious and clearer for higher temporal resolutions. The variogram parameters obtained for the hourly temporal resolution differ significantly from the parameters for other time scales. The nugget effect c_0 and the range a_{eff} are significantly lower. However, only the variogram point obtained for the shortest distance class causes this behaviour. It might be explained by the different number of stations for hourly data and the implication that less point pairs are available due to this. When comparing the variogram parameters from daily to annual temporal resolution, it is noticed that there is a decrease in range and nugget effect.

For sunshine duration, a threshold for the mean of the observations of 0.1 min and 0.1 h was established for the hourly and all other temporal resolutions, respectively. The idea was to leave out the time steps with no sunshine at all in order to compute a more accurate variogram. The night time steps that are entirely without sun are removed hence by this constraint. The spatial persistence of sunshine duration is clear and distinct for rather high temporal resolutions. In the case of larger time scales, in particular for the annual scale, there are higher deviations of the experimental points from the theoretical model. The decrease of spatial persistence with decreasing temporal resolution appears to be less strong compared to other climate variables. A relatively high nugget effect c_0 is seen for a temporal resolution of 1 h. Daily, weekly and monthly average sunshine durations show low nugget values, while the nugget value obtained for the annual time scale is slightly higher. A lower interpolation quality of hourly sunshine duration is expected due to this.

In the case of wind speed, the experimental variograms show an unclear behaviour for lower temporal resolutions and a decent spatial persistence for hourly data. Among all meteorological observations, wind speed shows the weakest relationship between distance and similarity of recorded values. For weekly, monthly and annual data, an objective manual fitting of the theoretical model was not possible, since no clear behaviour is seen. The cross validation of wind speed is carried out using the obtained variogram parameters regardless of the poor fitting result.

5.2. Interpolation performance

A separate analysis of interpolation performance was performed for each climate variable. NN, InvD, OK and KED using different additional information were considered in the evaluations. CM, the specific merging technique for rain gauge and radar data, is only present in the evaluations for precipitation. In general, the cross validation analysis was performed for ten realisations of each station density scenario and the performance criteria were averaged. Different additional information was used as a background variable in KED and the cross validation result is only shown for selected additional information in the following figures. However, all numerical cross validation results are shown in corresponding tables, i.e. also the additional information that could not improve the interpolation performance in comparison to OK is presented. The numerical values of all performance criteria are averaged over all network density scenarios for each temporal resolution.

5.2.1. Precipitation

The cross validation results of precipitation are shown in Fig. 4. The performance criteria were computed for all time steps of each temporal resolution, however only the average performance calculated from all time steps with a mean station rainfall higher than 1.0 mm is shown. Radar data was used successfully in several studies and is therefore taken into account here as one of the additional

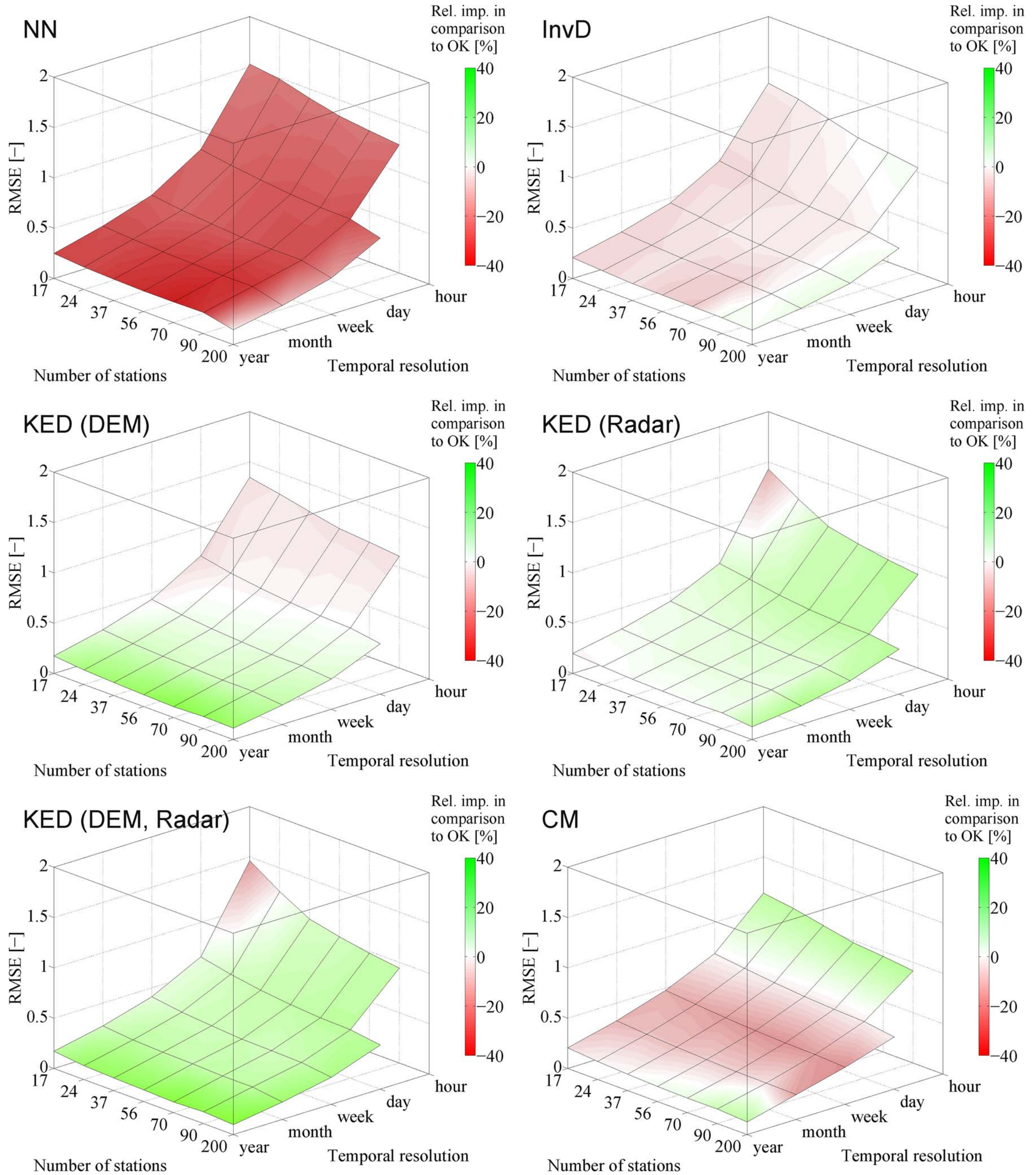

Fig. 4. Interpolation performance of precipitation using NN, InvD, KED (DEM), KED (Radar), KED (Radar, DEM) and CM.

information being used in KED and CM. Due to errors in radar data and the absence of a thorough attenuation correction, it is expected that the benefit of incorporating radar information might be restricted to certain accumulation times. Moreover, a digital elevation model (DEM) was taken into account as it can aid spatial interpolation.

The RMSE performance is plotted on the vertical axis while the horizontal axes contain the information about temporal resolution and station density. In addition, the surface colour illustrates the relative improvement of interpolation quality with respect to OK. There is a general decrease of interpolation performance with increasing temporal resolution and network density. NN performs worse than OK for all combinations of temporal resolution and network density, the relative difference to OK ranges from −17% (1h) to −35% (1a). The decline in interpolation quality is less strong for the network density of 200 stations. InvD performs similarly to OK, i.e. for most network density scenarios and temporal resolutions the interpolation performance is around −1% to −5%. However, for the 200 stations scenario and for the hourly temporal resolution of the scenarios with 90 and 70 stations InvD performs slightly better than OK.

Using the DEM as the only additional information in KED can improve the interpolation performance, especially at the annual time scale, whereas the RMSE of KED is around 12% to 30% lower than the RMSE of OK. The improvement of monthly interpolation performance is somewhat weaker. It ranges from approximately 5% to 12%. On the weekly time scale, there is only a minor benefit of using the DEM, while on the daily time scale the interpolation performance of KED is already slightly lower than those of OK for most network densities. Only for the scenario using 200 stations and the scenario using 90 stations is a minimal improvement seen. In the case of hourly data, the KED interpolation performance is approx. 3% lower. The results of KED based on radar data are different. For the annual temporal resolution, there is only a significant improvement of interpolation performance for the scenario consisting of 200 stations. The other annual network densities show either a minor decrease or increase of interpolation performance. The monthly and weekly time scales show each a slight improvement of around 5% for all network densities, except for 200 stations. Here, an improvement of around 12% is achieved. The daily temporal resolution is improved by approx. 6% (17 stations) to approx. 12% (200 stations). In the case of hourly data, the maximum benefit of using KED with radar can be quantified to 13% in comparison with OK. However, there is a strong decrease in interpolation performance of approx. 10% for the network density of 17 stations.

KED using the additional information of both, DEM and Radar, results in an improvement of interpolation performance compared to OK for almost all combinations of temporal resolutions and network density scenarios. Only for a low station density together with hourly data, it performs worse than OK. KED using both additional information behaves for low temporal resolutions like KED using the elevation only. For high temporal resolutions it is similar to KED using radar only. However, the performance of KED (Radar, DEM) with an hourly temporal resolution is somewhat lower than the corresponding interpolation performance of KED (Radar).

CM performs best at fine temporal resolutions: the hourly RMSE values are approx. 13% to 14% lower than those of OK. Also for low station densities, an improvement in comparison to OK can be achieved. CM and OK deliver a relatively similar interpolation performance for the daily temporal resolution, i.e. for some station densities, CM performs slightly worse. In contrast to KED (Radar), CM performs worse than OK at the weekly, monthly and annual time scale. Only for 90 and 200 stations, there is a slightly better interpolation quality of CM. It is assumed that this different behaviour of CM in comparison to KED results from the more direct use of radar data implemented in the CM method. KED uses the radar images in combination with station data to obtain the weights required for calculating the rainfall estimate based on adjacent station recordings. In contrast, CM computes a spatial pattern of radar rainfall and adds it directly to the OK estimate. Another difference between CM and KED (Radar) can be observed from the hourly interpolation performance. The RMSE values of CM are relatively constant over all network density scenarios, whereas KED (Radar) causes a much higher decline of interpolation quality when fewer stations are used. This behaviour might also be explained by the difference in the implementation of radar data. The network density scenarios with few stations are not able to capture the skew and non-Gaussian spatial distribution of short-time rainfall sums. The more direct use of radar information (CM) can therefore achieve a better performance than the estimation based on a linear combination of neighbouring points, in which radar is only used to determine the weights (KED). In contrast to the findings of Goudenhoofdt and Delobbe (2009), the interpolation performance of CM was not better than that of OK for daily data. However, it could be confirmed that KED using Radar outperforms OK and CM.

Table 5 contains the RMSE, Bias and RVar interpolation performance for all interpolation methods and temporal resolutions. The criteria are averaged over all network densities available for the corresponding temporal resolution. Since the Bias criterion is given in mm, the values are increasing with an increase of accumulation time. For low temporal resolutions, the application of geostatistical

Table 5
Average interpolation performance (Bias, RMSE, RVar) for hourly to annual precipitation over all station density scenarios.

Method	Add. Inf.	Criterion	Temporal resolution				
			Hour	Day	Week	Month	Year
NN	–	Bias [mm]	0.006	0.038	0.170	0.694	8.779
		RMSE [−]	1.390	0.742	0.461	0.335	0.217
		RVar [−]	0.955	0.968	0.974	1.051	1.146
InvD	–	Bias [mm]	0.007	0.031	0.136	0.550	6.918
		RMSE [−]	1.181	0.623	0.384	0.273	0.174
		RVar [−]	0.594	0.624	0.619	0.628	0.658
OK	–	Bias [mm]	0.007	0.028	0.121	0.474	5.440
		RMSE [−]	1.178	0.606	0.376	0.263	0.168
		RVar [−]	0.291	0.368	0.373	0.381	0.388
EDK	DEM	Bias [mm]	0.008	0.028	0.083	0.284	2.527
		RMSE [−]	1.212	0.612	0.363	0.243	0.133
		RVar [−]	0.327	0.485	0.537	0.611	0.736
EDK	Radar	Bias [mm]	0.025	0.058	0.183	0.438	1.554
		RMSE [−]	1.109	0.546	0.357	0.254	0.166
		RVar [−]	0.486	0.609	0.603	0.523	0.464
EDK	Radar, DEM	Bias [mm]	0.027	0.059	0.146	0.304	−0.499
		RMSE [−]	1.131	0.546	0.344	0.233	0.129
		RVar [−]	0.529	0.719	0.761	0.769	0.827
CM	Radar	Bias [mm]	0.028	0.081	0.234	0.259	−0.563
		RMSE [−]	1.019	0.615	0.420	0.279	0.164
		RVar [−]	0.607	0.771	0.948	0.902	0.609

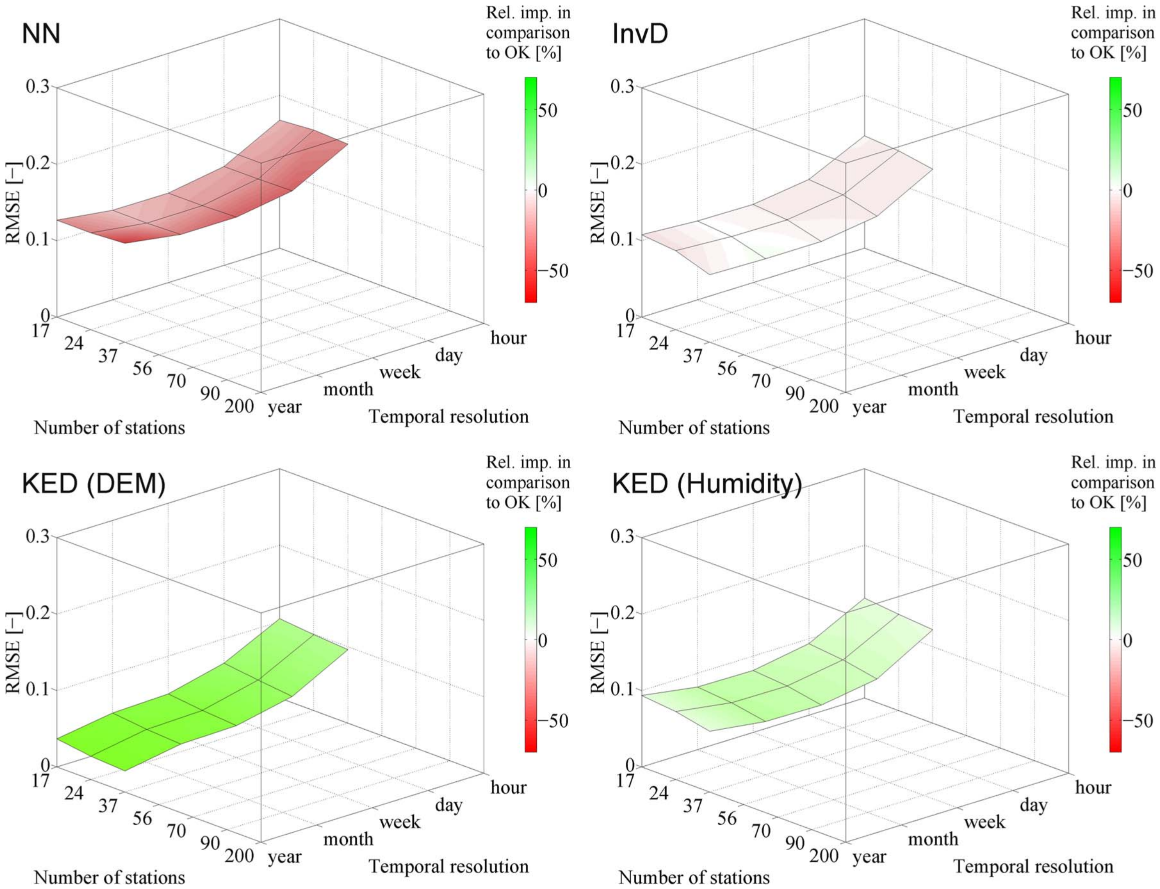

Fig. 5. Interpolation performance of temperature using NN, InvD, KED (DEM), KED (Humidity).

techniques produces a lower Bias than using simple methods. For hourly and daily rainfall accumulations, the Bias of radar-based methods is somewhat higher compared to no radar data use. OK shows a lower preservation of the observed variance in comparison to KED using the DEM for low temporal resolutions and in comparison to KED and especially CM incorporating radar data for high temporal resolutions. The smoothing effect of OK is also stronger than the one of InvD, since 12 neighbouring stations are used to calculate the rainfall estimates. InvD uses only four adjacent rain gauges, the closest in each quadrant starting from the location of the estimate. In general, NN preserves the variance very well, but delivers the worst interpolation performance according to the other criteria.

5.2.2. Temperature

For mean temperature, all available time steps were used in the performance evaluation. The problem where many time steps have zero rainfall recordings does not occur with temperature recordings. Many studies reported that an estimation of temperature could benefit from taking into account the site specific altitude. As such, the DEM was used within the temperature interpolation as well. Furthermore, gridded information of relative humidity was used as an additional variable. The evaluations in Fig. 5 shows that NN performs worse than OK for all temporal resolutions and network densities. The decrease in performance compared to OK ranges from −13% to −50%. InvD performs only slightly worse than OK for most combinations of temporal resolution and station density, i.e. the maximum decrease of interpolation performance is around 6%. For monthly values, there is an improvement of approx. 5% when 24 stations are used.

The cross validation results of KED using the elevation and KED using interpolated grids of relative humidity are shown in Fig. 5. The humidity grids were interpolated using OK. KED achieves a good interpolation performance for both secondary variables. The estimation quality of mean temperature improves by almost 30% for hourly data to almost 70% for annual averages when the DEM is used. In the case using relative humidity as the additional information, the improvement ranges from 8% to 23%. Generally, the effect of network density on the interpolation performance is hardly visible in the plots, i.e. there is no significant decrease when the network density is reduced from 37 to 17 stations. The interpolation depends strongly on the temporal resolution, however for some interpolation techniques there is no clear behaviour, i.e. the interpolation performance does not always increase with decreasing temporal resolution, as it happens for rainfall. KED using elevation as additional information displays by far the best interpolation performance and delivers a continuous improvement of RMSE with decreasing temporal resolution. Great improvements are

Table 6
Average interpolation performance (Bias, RMSE, RVar) for hourly to annual mean temperature over all station density scenarios.

Method	Add. Inf.	Criterion	Temporal resolution				
			Hour	Day	Week	Month	Year
NN	–	Bias [°C]	−0.216	−0.161	−0.175	−0.162	−0.225
		RMSE [−]	0.169	0.130	0.118	0.118	0.128
		RVar [−]	1.063	1.282	1.419	1.394	1.720
InvD	–	Bias [°C]	−0.093	−0.084	−0.095	−0.085	−0.123
		RMSE [−]	0.142	0.106	0.096	0.096	0.100
		RVar [−]	0.616	0.726	0.773	0.753	0.875
OK	–	Bias [°C]	−0.031	0.003	−0.004	0.008	−0.042
		RMSE [−]	0.138	0.103	0.094	0.098	0.095
		RVar [−]	0.349	0.420	0.414	0.361	0.494
EDK	DEM	Bias [°C]	0.039	0.027	0.014	0.027	−0.015
		RMSE [−]	0.100	0.063	0.046	0.044	0.032
		RVar [−]	0.583	0.683	0.716	0.684	0.837
EDK	HUM	Bias [°C]	0.051	0.037	0.025	0.038	0.000
		RMSE [−]	0.098	0.064	0.051	0.052	0.073
		RVar [−]	0.636	0.701	0.710	0.657	0.559

especially seen at the annual and monthly time scale.

Table 6 contains the Bias, RMSE and RVar interpolation performance of mean temperature. The abbreviations from Table 2 are used if the respective climate variable was used as additional information in KED. The Bias of the geostatistical techniques is in general lower than the Bias of NN and InvD. OK always delivers the best interpolation performance in terms of this criterion, however, OK also creates, due to the rather high number of neighbouring stations taken into account, a very smooth surface of the interpolated variable. This is indicated by the lowest preservation of variance expressed by the RVar criterion. InvD and the two KED implementations achieve a similar interpolation performance according to this measure. NN generates a higher variance than the one present in the observed values. It is assumed that this is caused by the weather station located on the Brocken, the highest peak of the Harz Mountains. This station has by far the highest altitude (1141 m) with two neighbouring stations having elevations of around 600 m or less. In the case of NN cross validation, the temperature observation at the Brocken station is taken as the estimate for both adjacent stations with a significantly lower altitude while only the temperature observation of one of the lower altitude stations is used as the estimate for the Brocken location. Due to the smoothing effect or the additional information taken into account, this phenomenon is not as severe as for the other interpolation techniques.

5.2.3. Relative humidity

Due to the relationship of temperature to dew point, it is expected that the use of temperature information could help for the interpolation of relative humidity. Moreover, the DEM, interpolated grids of precipitation as well as the number of wet 5 min time steps computed from radar were all taken into account for the evaluations. OK was applied for the interpolation of rainfall grids, since no interpolation method using either radar or the DEM could achieve a consistent improvement of interpolation performance for all temporal resolutions and network densities. The preparation of the temperature grids was carried out using KED, since the incorporation of the DEM improves the interpolation performance considerably as previously stated.

Fig. 6 illustrates the RMSE cross validation results, wherein the performance criteria were computed for all time steps of each temporal resolution and then averaged in order to determine the mean interpolation performance. As for temperature, all available time steps were used for the performance evaluation since there are no time steps with only zero recordings for all stations. The evaluations show that NN and InvD perform worse than OK for almost all temporal resolutions and network densities. The NN interpolation performance is around 22% to 40% worse and there is a particular decline for the network density scenario consisting of 37 stations. InvD performs up to 8% worse than OK. There is only a minor improvement of approx. 3% for annual data when 37 stations are used. The DEM and interpolated grids of temperature and rainfall were used as additional information. Moreover, the number of 5 min time steps with rainfall was computed from each radar grid point for each time step for all temporal resolutions and incorporated in KED as well. For relative humidity, the interpolation performance of KED using elevation improves only for some combinations of station density and temporal resolution in comparison to OK. The maximum improvement (approx. 15%), illustrated by the green colour shading occurs for annual data using 17 stations. The decline in interpolation performance observed for some combinations of station density and temporal resolution reaches a maximum of 5%. KED using OK interpolated rainfall data can improve the interpolation performance only at the annual time scale. An improvement of 18% to 25% in comparison to OK is achieved for annual data, while the interpolation performance declines by around 1% to 17% for the other temporal resolutions. KED using temperature grids delivers a consistent improvement of interpolation performance for all temporal resolutions and station density scenarios. The improvement ranges from 9% to 20%. The estimation could not be improved by using the number of rainy time steps as an additional information in KED. Besides the additional information displayed in Fig. 6, sunshine duration and wind velocity were used in combination with KED. The incorporation of both additional variables could not improve the RMSE interpolation performance in comparison to OK either.

Table A1 (Supplementary Material) contains the Bias, RMSE and RVar interpolation performance for relative humidity. It is also

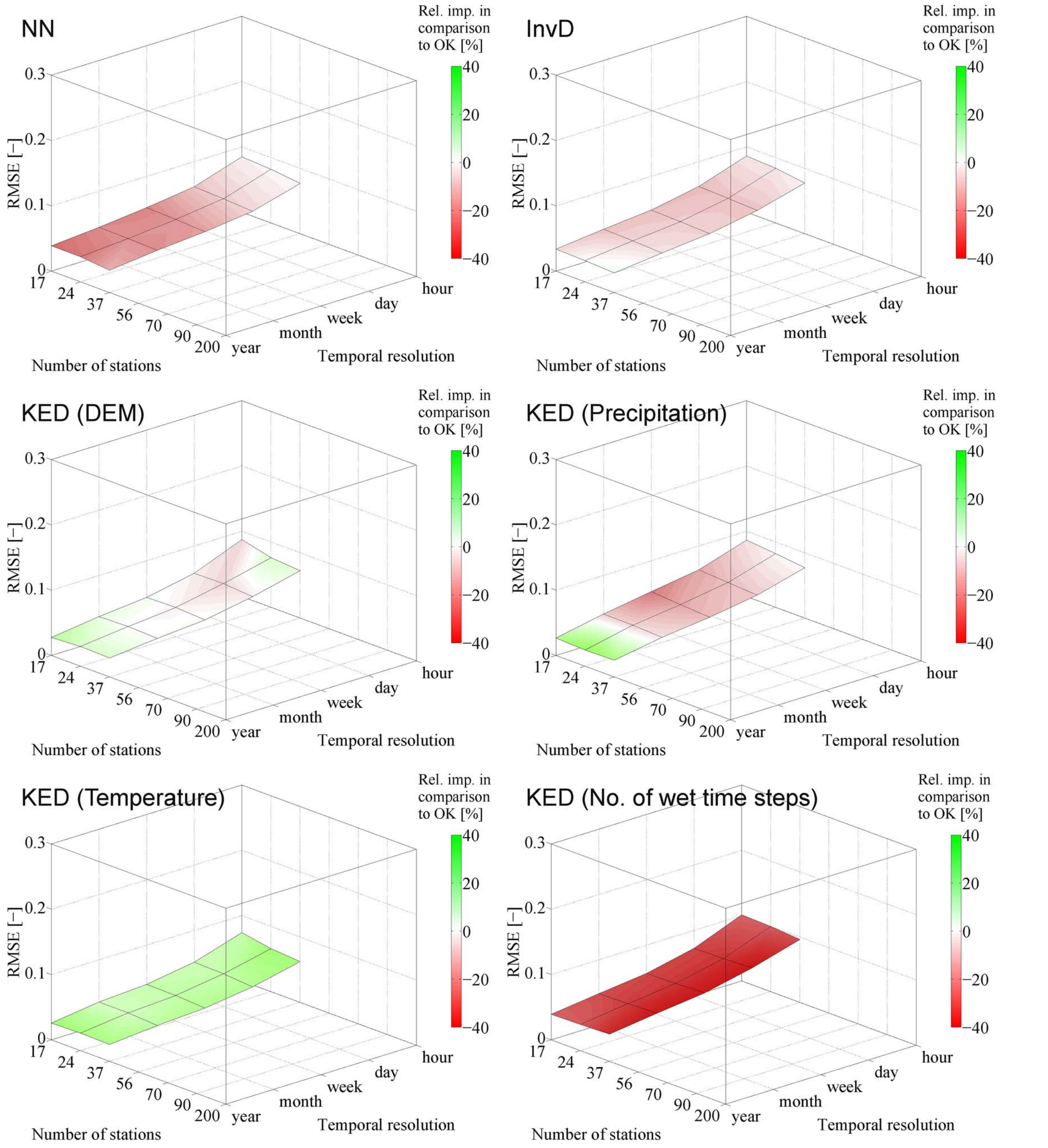

Fig. 6. Interpolation performance of relative humidiy using NN, InvD, KED (DEM), KED (Precipitation), KED (Temperature) and KED (No. of wet time steps).

seen here, that the simple interpolation methods NN and InvD generate a higher bias than the geostatistical approaches. OK again causes the highest reduction of variance, while KED using temperature grids achieves a good variance preservation.

5.2.4. Cloud coverage, sunshine duration and wind speed

Several additional information were tested for the interpolation of cloud coverage, sunshine duration and wind velocity. None of the variables could improve the performance in comparison to OK and thus Figs. 7–9 contain only the results obtained for NN and InvD. For all the variables, the simple NN technique performs significantly worse than OK. For cloud coverage and sunshine duration, the decline of interpolation accuracy ranges from approx. 21% to 36%, while NN performs around 39% to 60% worse for wind speed. InvD performs slightly worse than OK for sunshine duration as well as cloud coverage. The maximum decline of interpolation performance is approx. 11% for cloud coverage and approx. 5% for sunshine duration. In the case of wind speed, the interpolation performance is reduced around 29% to 42%. Here, it seems implausible that the network density scenario based on 17 stations has a better interpolation performance than the scenario using 24 stations. This irregular behaviour is explained by the cross validation

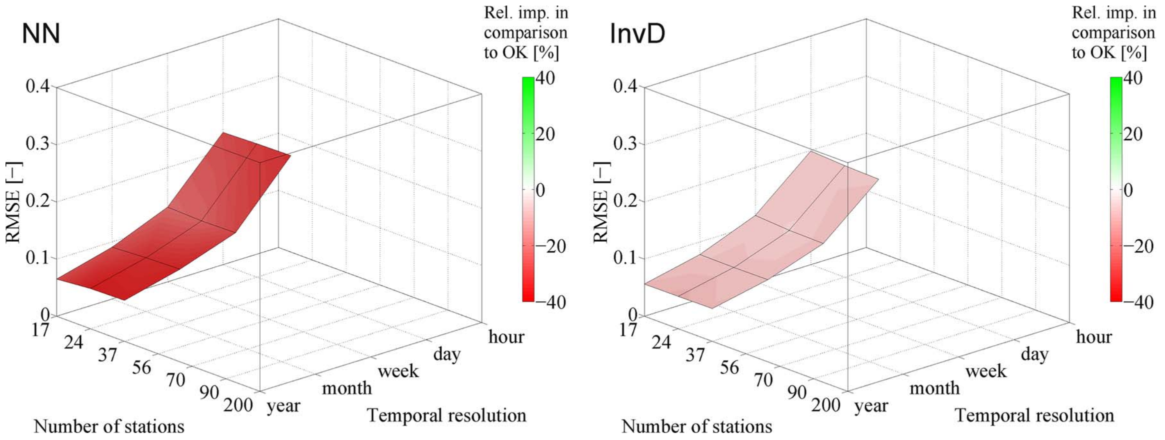

Fig. 7. Interpolation performance of cloud coverage using NN and InvD.

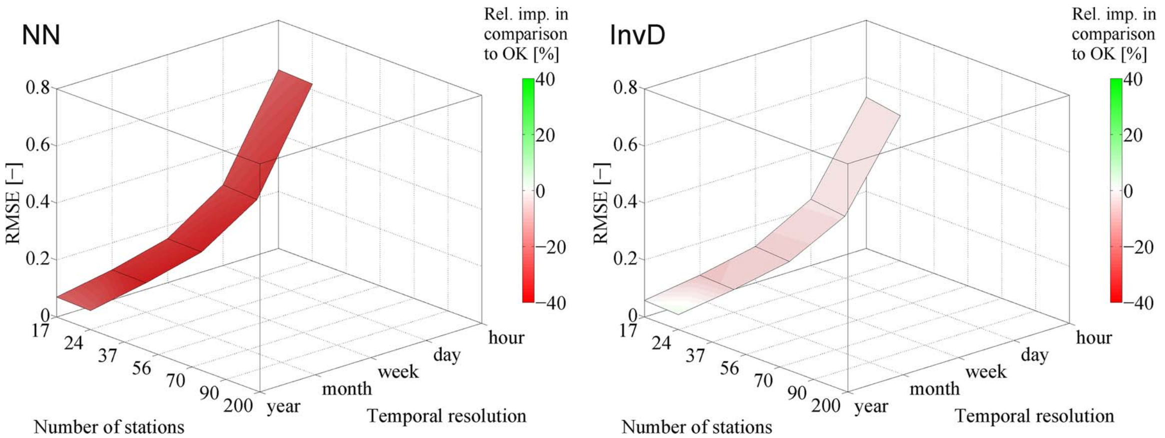

Fig. 8. Interpolation performance of sunshine duration using NN and InvD.

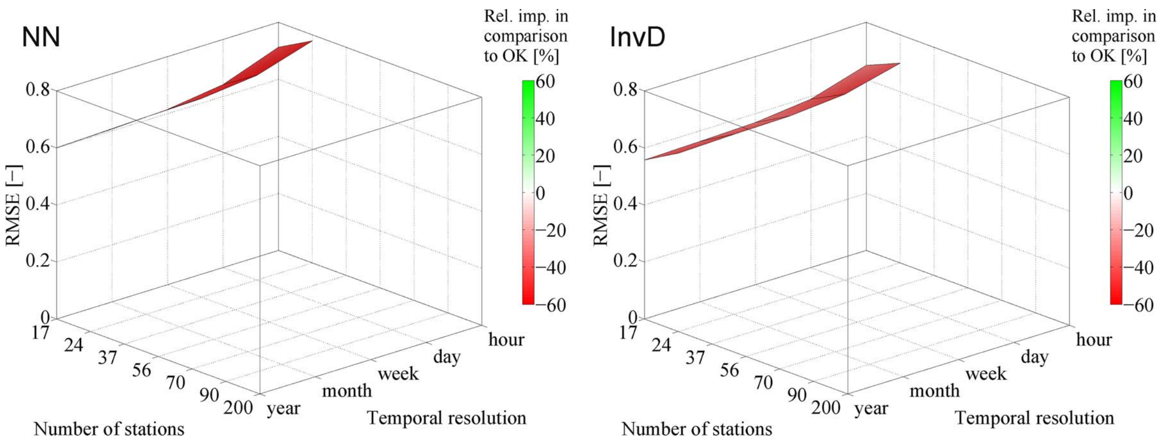

Fig. 9. Interpolation performance of wind speed using NN and InvD.

setup of this study in combination with the high exposure of the weather station situated on the Brocken Mountain. The performance evaluation was conducted for ten realisations of each station density scenario, while the stations were selected randomly for the realisations of each scenario. The Brocken station is considered in all ten realisations of the 24 stations scenario and only in seven realisations of the scenario using 17 stations. A significantly higher interpolation error for this particular station causes a reduction of interpolation performance for the concerned realisations and affects the corresponding average.

Bias, RMSE and RVar interpolation performances are shown in Tables A2–A4 (Supplementary Material). Geostatistical

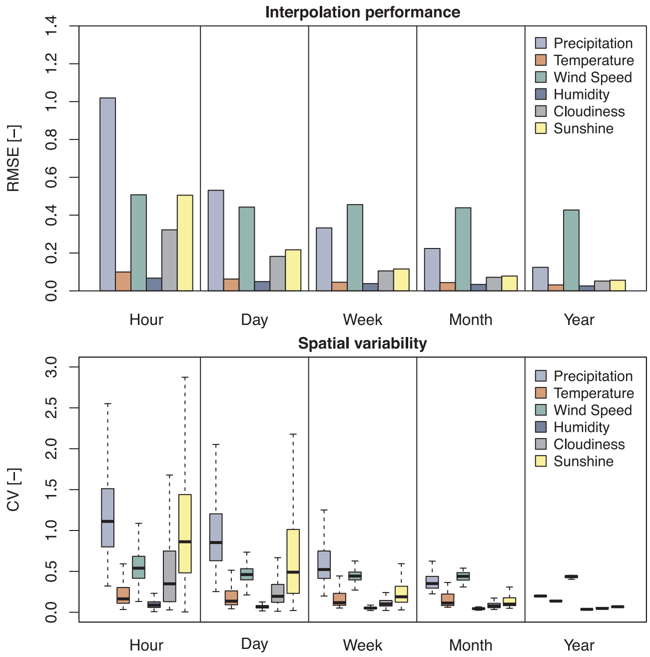

Fig. 10. Interpolation performance for the best interpolation technique (top) and spatial variability (bottom) compared among all climate variables and temporal resolutions.

interpolation techniques deliver a lower bias than NN and InvD in most cases. For cloud coverage and sunshine duration, the smoothing effect of OK causes a strong reduction of the spatial variance, while only NN can preserve it in its entirety. For wind data, NN and InvD cause an overestimation of the spatial variance. It is assumed that this phenomenon is also caused by the exposed location of the Brocken station. The observed value of this gauge is used as an estimate for two neighbouring stations in the same way as it was the case in the cross validation of temperature data.

5.3. Comparison among all climate variables

The comparison of interpolation performance among climate variables is carried out for the best interpolation methods found in Section 5.2. The results are shown in the upper panel of Fig. 10. The panel at the bottom contains the coefficient of variation that was computed for each time step of the different meteorological observations and their temporal resolutions. The RMSE performance criterion is used to compare the interpolation performance among the different climate variables. This comparison approach is considered valid, since the RMSE is standardised with the mean of the observations for each climate variable. It is carried out despite the fact that there is a natural upper limit of the observation range of sunshine duration, cloud coverage and relative humidity.

In case of the hourly time scale, the interpolation of precipitation generates the worst result, with an RMSE of approx. 1.0. A good interpolation performance was achieved for relative humidity and mean temperature, i.e. the RMSE is lower than 0.1 for both. Wind speed, cloud coverage and sunshine duration have a medium interpolation performance. The interpolation performance of precipitation, sunshine duration and cloud coverage improves significantly when examined at longer time periods. For mean temperature as well as relative humidity, there is a less significant improvement. In contrast to all other variables, the estimation of wind speed does not improve appreciably when longer time periods are interpolated. On the weekly time scale, the interpolation of wind speed is already slightly worse than the interpolation of precipitation. In terms of annual performance, wind speed interpolation gives by far the worst result. It is assumed that wind is in general strongly affected by local conditions, in particular by the topography that can cause either shielding or amplification effects depending on wind direction. Relative humidity has the lowest interpolation error for all temporal resolutions, whereas the error of the mean temperature interpolation is slightly higher. The boxes in the bottom panel of Fig. 10 contain the coefficients of variation that were computed for each time step of the corresponding climate variable and temporal resolution. Outlier time steps are removed to enable a better illustration of the spatial variability. For rainfall, the CV was only computed for time steps that have an observation mean value higher than or equal to 1.0 mm. It is clear that the interpolation performance is strongly linked to the spatial variability of each climate variable. Moreover, variance decreases significantly with a decrease of temporal resolution. In general, it is clear that the spatial interpolation of fine temporal resolution rainfall data is the most challenging task. Rainfall observations with high temporal resolution have a high spatial variability and the expected interpolation error is therefore much higher. The interpolation error of humidity and temperature is in general the lowest. In particular, for temperature interpolation, the DEM offers reliable additional information that can improve the cross validation result.

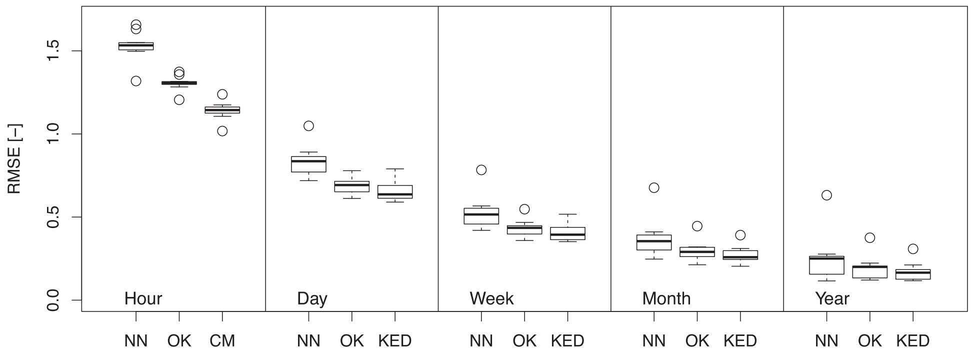

Fig. 11. Interpolation performance range obtained for precipitation from ten realisations of the 17 stations scenario.

5.4. *Influence of station selection*

This section contains information about the effect of the random selection of stations on the interpolation performance. In order to find out whether the location of the measurement is important, the RMSE cross validation performance obtained for the ten realisations of each density scenario using 17 stations is displayed in box plots. The results of each meteorological variable are presented in a separate figure for all temporal resolutions. The scenario of 17 stations is available for all climate information and therefore allows a thorough comparison among all meteorological information.

Fig. 11 contains the box plots obtained for the interpolation of hourly to annual rainfall sums, while only selected methods are displayed for each temporal resolution. NN and OK are always shown. These simple univariate techniques are used as a standard of comparison to the best multivariate approach. NN represents the simplest possible basic interpolation technique, while OK is the most basic univariate geostatistical interpolation method. CM is displayed for hourly data since it has shown the best interpolation performance. For all other temporal resolutions, KED, which uses both elevation and radar data, is shown as the third method. There is a significant variation in interpolation performance among the random station selections for each temporal resolution and interpolation method. NN generates a higher variation compared to the geostatistical methods. OK and KED generate a similar variation in the error for daily and weekly data. In particular, one specific realisation causes a high RMSE for the daily, weekly, monthly and annual time scale. It is marked as an outlier for all methods and is explained by the high elevation and the corresponding high amounts of orographic precipitation of the gauge situated on the Brocken mountain. Only one scenario contained this station, since the random selection was drawn from 200 stations in total.

The variation of temperature interpolation performance is shown in Fig. 12. The KED method presented here used the DEM as the additional information. The variation in KED interpolation performance is significantly lower than the variation in NN and OK interpolation performance. NN causes the highest variation, while OK generates a medium range of interpolation performances. The distribution of NN and OK interpolation performance appears to be skewed to the right, i.e. a relatively good interpolation performance is achieved for most realisations of each temporal resolution, while some realisations have significantly higher errors. NN and OK do not consider elevation and therefore it is assumed that the weather station on the Brocken mountain, which is not considered in all realisations, causes this behaviour.

Fig. A1 (Supplementary Material) shows the variation in interpolation performance of relative humidity. It is clearly seen that NN creates the highest variation, while the cross validation of geostatistical techniques, in particular KED using temperature, were able to provide a relatively robust assessment of interpolation quality. A specific reason for the outlier caused by KED for hourly data could not be identified.

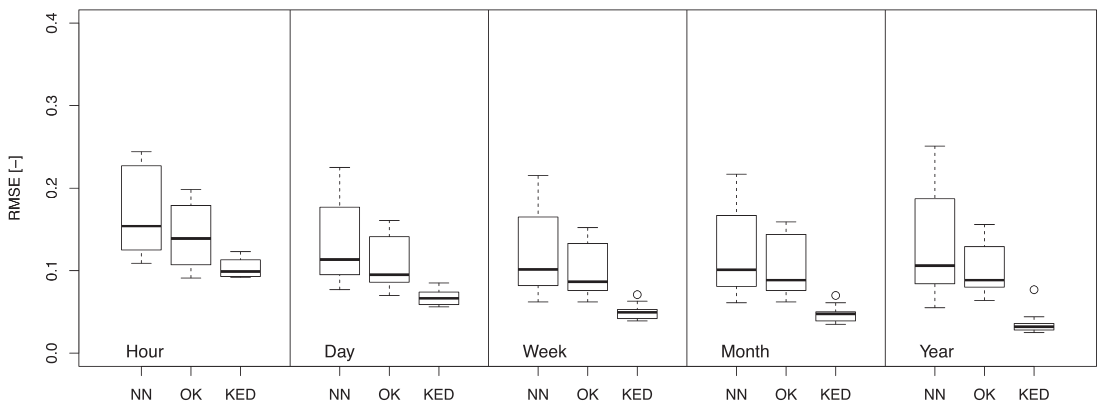

Fig. 12. Interpolation performance range obtained for temperature from ten realisations of the 17 stations scenario.

The results for cloud coverage are shown in Fig. A2 (Supplementary Material). The variation is relatively low compared to the meteorological observations shown before. Nevertheless, the application of OK leads to a slightly lower spread of interpolation performances.

The range of sunshine duration interpolation performance is shown in Fig. A3 (Supplementary Material). The variations seem to be even lower than those obtained for cloud coverage. It is assumed that the limited number of available stations additionally contributes to the low variation. Random selections were drawn from 25 stations only. Again, OK causes a marginally lower variation of interpolation performance among the ten realisations compared to NN.

The distribution of wind speed interpolation performance, seen in Fig. A4 (Supplementary Material), is highly skewed for all interpolation methods and temporal resolutions. A relatively poor performance is reached for most realisations, while some random selections, which did not take into account the Brocken station, achieve a much better interpolation quality. The range of interpolation performance is by far the highest, when compared to other meteorological observations. It is assumed that wind velocity is strongly affected by the local topographic conditions. Realisations that contain only locations with similar local topography therefore achieve the best interpolation performance. A spatial interpolation using the entire set of available stations does not necessarily yield an optimal interpolation quality.

6. Summary and conclusions

This study investigated the performance of different interpolation techniques for various climate variables observed by weather stations. Simple interpolation techniques (NN, InvD) and more sophisticated geostatistical approaches (OK, KED, CM) were taken into account. For each climate variable, different additional information based on topography, other measurements and other factors were used within the multivariate interpolation techniques. Cross validation experiments based on different temporal resolutions and station density scenarios were implemented in order to determine the interpolation performance for each variable. The main findings and conclusions can be summarised as follows:

1. KED using radar as the additional information can improve the interpolation performance of rainfall in comparison to OK, particularly for fine temporal resolutions of 1 h and 1 day. In the case of low station density, KED may perform worse than OK for high temporal resolutions. KED incorporating the DEM is especially helpful for long accumulation times but cannot achieve an improvement for high temporal resolutions. CM delivers the best performance for hourly data but performs worse than OK and KED for temporal resolutions of 1 day or lower.
2. KED using the DEM performs significantly better than OK for temperature data for all temporal resolutions and station densities.
3. KED using temperature grids delivered the best interpolation performance for relative humidity. All combinations of station density and temporal resolution could be improved compared to OK. The incorporation of precipitation grids could only achieve an improvement at the annual time scale.
4. No useful additional information was found for the interpolation of cloud coverage, sunshine duration, or wind data. The application of OK resulted in the best interpolation performance for all station densities and temporal resolutions.
5. The simple approaches NN and InvD cannot reach the interpolation performance of OK for most climate variables. Only for precipitation, InvD performs similarly well as OK if a very dense station network is available.
6. The influence of the random station selection on the interpolation performance varies strongly on the climate variable that is interpolated. In particular, for wind speed, a strong impact of the station selection was observed.
7. Moreover, the interpolation performance depends generally on temporal resolution, station density and the specific spatial variability of the climate information. The influences of temporal resolution and spatial variability appear to be higher than the influence of station density.
8. Precipitation with a high temporal resolution shows the highest spatial variability and thus the worst interpolation performance. For all meteorological variables except wind speed, the spatial variability decreases with decreasing temporal resolution. The best estimation accuracy is achieved for relative humidity and temperature.

All results were obtained from the interpolation of continuous time series, i.e. the analysis of single events or short time step sequences might lead to different findings, in particular when radar is taken into account for rainfall interpolation. The study area is characterised by rather few topographic elevations. For wind data in particular, it is assumed that the interpolation performance might decrease even further if a more mountainous area is investigated. The interpolation performance of temperature and relative humidity seem to be rather robust, due to the consideration of the DEM and the low spatial variability.

Conflict of interest

None.

Acknowledgements

The authors would like to thank the German Weather Service for providing the meteorological observations and Ross Pidoto for correcting an early version of the manuscript. This research was partly funded by the German Environment Agency (RESTNI, project number: 3710422182) and the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (KliBiW, no project number available). The article publication was funded by the Open Access fund of Leibniz Universität Hannover. Finally, the reviewer and the editor are gratefully acknowledged for their contributions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ejrh.2018.02.002.

References

Bárdossy, A., Pegram, G., 2013. Interpolation of precipitation under topographic influence at different time scales. Water. Resour. Res. 49, 4545–4565.
Berndt, C., Rabiei, E., Haberlandt, U., 2014. Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. J. Hydol. 508, 88–101.
Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteorol. 33, 140–158.
Delrieu, G., Wijbrans, A., Boudevillain, B., Faure, D., Bonnifait, L., Kirstetter, P.-E., 2014. Geostatistical radar-raingauge merging: a novel method for the quantification of rain estimation accuracy. Adv. Water Resour. 71, 110–124.
Deutsch, C.V., Journel, A.G., 1992. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press.
Dirks, K.N., Hay, J.E., Stow, C.D., Harris, D., 1998. High-resolution studies of rainfall on Norfolk Island: part II: Interpolation of rainfall data. J. Hydol. 208, 187–193.
Dodson, R., Marks, D., 1997. Daily air temperature interpolated at high spatial resolution over a large mountainous region. Climate Res. 8, 1–20.
Ehret, U., 2003. Rainfall and Flood Nowcasting in Small Catchments Using Weather Radar, 1st ed. Mitteilungen. Eigenverlag des Instituts für Wasserbau der Universität Stuttgart, Stuttgart.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.
Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydol. 228, 113–129.
Goudenhoofdt, E., Delobbe, L., 2009. Evaluation of radar-gauge merging methods for quantitative precipitation estimates. Hydrol. Earth Syst. Sci. 13, 195–203.
Haberlandt, U., 2007. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. J. Hydol. 332, 144–157.
Hevesi, J.A., Flint, A.L., Istok, J.D., 1992a. Precipitation estimation in mountainous terrain using multivariate geostatistics: part II: isohyetal maps. J. Appl. Meteorol. 31, 677–688.
Hevesi, J.A., Istok, J.D., Flint, A.L., 1992b. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis. J. Appl. Meteorol. 31, 661–676.
Hudson, G., Wackernagel, H., 1994. Mapping temperature using kriging with external drift: theory and an example from Scotland. Int. J. Climatol. 14, 77–91.
Hutchinson, M.F., 1998a. Interpolation of rainfall data with thin plate smoothing splines: part I: two dimensional smoothing of data with short range correlation. J. Geogr. Inf. Decis. Anal. 2, 139–151.
Hutchinson, M.F., 1998b. Interpolation of rainfall data with thin plate smoothing splines: part II: analysis of topographic dependence. J. Geogr. Inf. Decis. Anal. 2, 152–167.
Isaaks, E.H., Srivastava, R.M., 1990. An Introduction to Applied Geostatistics. Oxford University Press, New York.
Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ. Modell. Softw. 16, 309–330.
Krajewski, W.F., 1987. Cokriging radar-rainfall and rain gage data. J. Geophys. Res.–Atmos. 92, 9571–9580.
Li, T., Zheng, X., Dai, Y., Yang, C., Chen, Z., Zhang, S., Wu, G., Wang, Z., Huang, C., Shen, Y., Liao, R., 2014. Mapping near-surface air temperature, pressure, relative humidity and wind speed over Mainland China with high spatiotemporal resolution. Adv. Atmos. Sci. 31, 1127–1135.
Lloyd, C.D., 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. J. Hydol. 308, 128–150.
Martínez-Cob, A., 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. J. Hydol. 174, 19–35.
Nalder, I.A., Wein, R.W., 1998. Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. Agr. For. Meteorol. 92, 211–225.
Nanding, N., Rico-Ramirez, M.A., Han, D., 2015. Comparison of different radar-raingauge rainfall merging techniques. J. Hydroinform. 17, 422–445.
Phillips, D.L., Dolph, J., Marks, D., 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. Agr. For. Meteorol. 58, 119–141.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2016. Package 'nlme'. Available at: http://cran.r-project.org/web/packages/nlme/nlme.pdf.
Price, D.T., McKenney, D.W., Nalder, I.A., Hutchinson, M.F., Kesteven, J.L., 2000. A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. Agr. For. Meteorol. 101, 81–94.
Riedl, J., 1986. Radar-Flächenniederschlagsmessung. Promet 20–23.
Seltmann, J.E.E., 1997. Radarforschung im DWD: Vom Scan zum Produkt. Promet 26, 32–42.
Seo, D.-J., Breidenbach, J.P., Johnson, E.R., 1999. Real-time estimation of mean field bias in radar rainfall data. J. Hydol. 223, 131–147.
Shepard, D., 1968. A Two-Dimensional Interpolation Function for Irregularly-Spaced Data. ACM Press, pp. 517–524.
Sinclair, S., Pegram, G., 2005. Combining radar and rain gauge rainfall estimates using conditional merging. Atmos. Sci. Lett. 6, 19–22.
Smith, J.A., Krajewski, W.F., 1991. Estimation of the mean field bias of radar rainfall estimates. J. Appl. Meteorol. 30, 397–412.
Stahl, K., Moore, R.D., Floyer, J.A., Asplin, M.G., McKendry, I.G., 2006. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agr. For. Meteorol. 139, 224–236.
Tabios, G.Q., Salas, J.D., 1985. A comparative analysis of techniques for spatial interpolation of precipitation. JAWRA J. Am. Water Resour. Assoc. 21, 365–380.
Thiessen, A.H., 1911. Precipitation averages for large areas. Mon. Weather Rev. 39, 1082–1089.
Velasco-Forero, C.A., Sempere-Torres, D., Cassiraga, E.F., Gómez-Hernández, J.J., 2009. A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. Adv. Water Resour. 32, 986–1002.
Verworn, A., Haberlandt, U., 2011. Spatial interpolation of hourly rainfall – effect of additional information, variogram inference and storm properties. Hydrol. Earth Syst. Sci. 15, 569–584.
Yoon, S.-S., Phuong, A.T., Bae, D.-H., 2012. Quantitative comparison of the spatial distribution of radar and gauge rainfall data. J. Hydrometeorol. 13, 1939–1953.