Theor Appl Climatol (2013) 112:99–111
DOI 10.1007/s00704-012-0716-9

ORIGINAL PAPER

Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models

Juha Aalto · Pentti Pirinen · Juha Heikkinen · Ari Venäläinen

Received: 11 May 2012 / Accepted: 3 July 2012 / Published online: 22 July 2012

Abstract The Finnish Meteorological Institute has calculated statistics for the new reference period of 1981–2010. During this project, the grid size has been reduced from 10 to 1 km, the evaluation of the interpolation has been improved, and comparisons between different methods has been performed. The climate variables of interest were monthly mean temperature and mean precipitation, for which the spatial variability was explained using auxiliary information: mean elevation, sea percentage, and lake percentage. We compared three methods for spatial prediction: kriging with external drift (KED), generalized additive models (GAM), and GAM combined with residual kriging (GK). Every interpolation file now has attached statistical key figures describing the bias and the normality of the prediction error. According to the cross-validation results, GAM was the best method for predicting mean temperatures, with only very small differences relative to the other methods. For mean precipitation, KED produced the most accurate predictions, followed by GK. In both cases, there was notable seasonal variation in the statistical skill scores. For the new reference period and future interpolations, KED was chosen as the primary method due to its robustness and accuracy.

J. Aalto (✉) · P. Pirinen · A. Venäläinen
Finnish Meteorological Institute,
BOX 503, 00101 Helsinki, Finland
e-mail: juha.aalto@fmi.fi

P. Pirinen
e-mail: pentti.pirinen@fmi.fi

A. Venäläinen
e-mail: ari.venalainen@fmi.fi

J. Heikkinen
Finnish Forest Research Institute Metla,
Vantaa, Finland
e-mail: juha.heikkinen@metla.fi

1 Introduction

Many biotic and abiotic processes are determined by climatic conditions and require long-term climatic datasets for research (Fronzek et al. 2006; Araújo and Luoto 2007; Kullmann 2010). The spatial extent of climate data can be a problem because the weather station network is often sparse and is not deployed on a regular grid. In addition, climate data may not be available where it is most needed. Spatially and temporally continuous gridded datasets are important in many applications such as in agriculture, water resources, soil sciences, and ecological studies (Clifton and Neuman 1982; Robertson 1987; Goovaerts 1999; Miller and Franklin 2002; Liu et al. 2006). One important use of gridded data is the verification of regional climate models (Jylha et al. 2004; Barnett et al. 2005).

Various statistical methods have been developed for predicting the spatial distribution of variables of interest. Geostatistics is concerned with producing continuous maps from noisy point data and taking advantage of the spatial association of the observations. Spatial autocorrelation is a very common feature for geographical variables and datasets, and means that observation close to each other are more likely to be similar (Legendre 1993). The aim is to predict values for locations with no observations based on the spatial autocorrelation between observations, possibly including some explanatory variables (Goovaerts 1997). Geographical information systems provide effective tools for collecting these predictors from different digital sources for use in spatial prediction.

Geostatistical methods include ordinary kriging, universal kriging, co-kriging, and indicator kriging (Goovaerts 1999). Kriging rely on an assumption of a linear relationship between the response and explanatory variables. In many cases, natural processes show a curvilinear rather than linear relationship (e.g., Leathwick et al. 2006; Hjort and Luoto 2010), and the possibility of relaxing the assumptions about linearity to let the data determine the shape of the response is often appreciated (Kammann and Wand 2003). Generalized additive models (GAM) are a semiparametric extension of generalized linear models (GLM) (Hastie and Tibshirani 1990), with the benefit of no strict parametric form, allowing a more flexible response (Hastie and Tibshirani 1984). GAM combined with geostatistical kriging methods are also known as *geoadditive models* (Kneib et al. 2009). Other modern statistical modeling methods, such as artificial neural networks, can also be combined with kriging to obtain better spatial predictions (Rizzo and Dougherty 1994; Lin and Chen 2004). Nongeostatistical methods include trend-surface analysis (Chorley and Haggett 1965), inversed distance weighted (IDW) interpolation (Lu and Wong 2008), and spline interpolation (Hong et al. 2005).

The Finnish Meteorological Institute (FMI) has calculated statistics for the new reference period of 1981–2010. At the same time, an older kriging algorithm from a `FORTRAN` environment has been replaced with one from the `R` statistical programming environment (R Development Core Team 2011), and the spatial resolution of the gridded datasets has been reduced from 10 to 1 km. The interpolation has been improved, with comparisons between different interpolation methods and statistics on the performance of each interpolation provided. Effort has also been made on the parametrization of variogram models, especially in the case of mean precipitation.

`R` is an open source programming environment for statistical computing and analyses (R Development Core Team 2011). With `R`, it is possible to perform a variety of statistical (e.g., linear and nonlinear modeling, geostatistics, time-series analysis) and graphical techniques. A major benefit is that it is highly extensible via different `packages`.

The object of this paper is to compare three different methods for spatial prediction: kriging with external drift (KED), GAM, and GAM combined with residual kriging (GK). The ability of kriging to make spatial predictions has been recognized in many studies (Voltz and Webster 1990; Laslett 1994; Hofstra et al. 2008). Some authors (Hutchinson 1998; Hong et al. 2005) have ignored the spatially correlated errors in their studies or have used nongeostatistical methods; an example is IDW for the residuals (Perry and Hollis 2005). Our aim, therefore, is to also compare the relative performance of deterministic and geostatistical methods in spatial prediction. In addition, we present key statistics for the new reference period of 1981–2010.

2 Study area and data

2.1 Study area

Finland is situated at high latitudes, between 60 ° and 70 °N, in the northeast of the Eurasian continent, close to the Atlantic Ocean (Fig. 1). The mean surface temperature varies significantly across the country; from a 1971–2000 mean of 5–6 °C in the south to a −3 °C for Lapland in the north, with a mean for the entire country of ca. 1.7 °C, calculated from gridded values. Precipitation also shows a north–south trend, ranging from an annual accumulation of over 700 mm in the south to less than 400 mm in the north and a mean value of 590 mm for the entire country for the 1971–2000 period (Drebs et al. 2002). The climate is substantially milder than elsewhere at the same latitudes due to the effects of the North Atlantic current providing heat to northern areas (Tikkanen 2005). Elevations for the country range from sea level to 1,310 m according to the digital elevation

Fig. 1 Location of Finland

model. A spatial grid with a resolution of 1 × 1 km is constructed for the entire country, containing a total of 336,897 points.

2.2 Observations

The Finnish network of observing stations is distributed relatively evenly across the country, except for northern Finland. The number of stations providing data varied from more than 500 to less than 200, depending on the date and variable being interpolated (Fig. 2). The observed values were obtained from FMI's climate database, and all available values were used to provide the best possible spatial coverage. Different stations were used for the verification of temperature and precipitation.

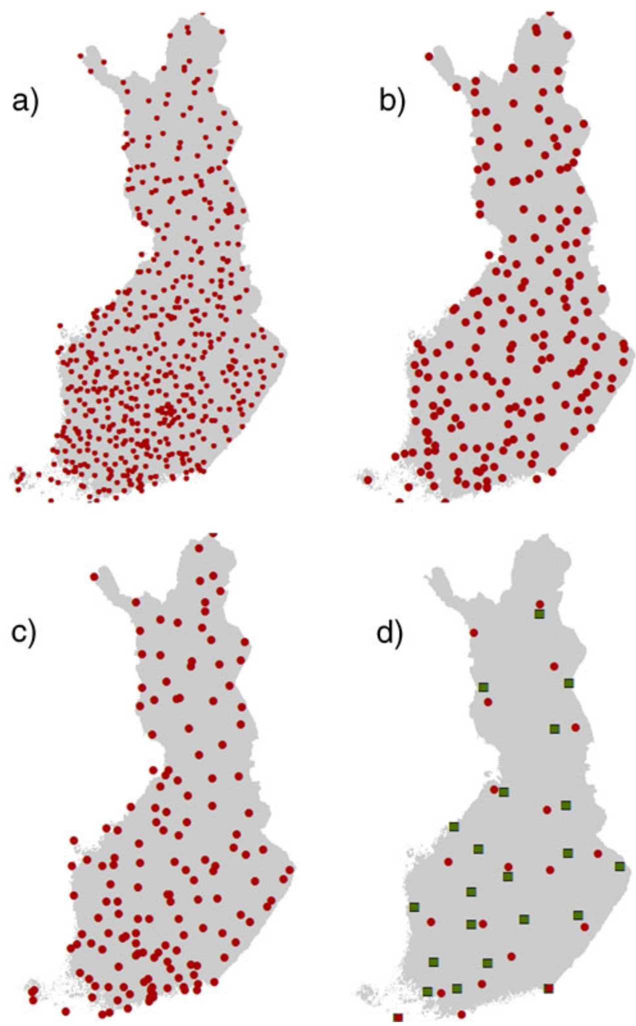

Fig. 2 Examples of the observation stations in Finland. **a** Precipitation observations in August 1985 ($n = 563$). **b** Precipitation observations in August 2010 ($n = 207$). **c** Temperature observations in August 2010 ($n = 148$). **d** Stations used for verification of temperature ($n = 20$, *red circles*) and of precipitation ($n = 22$, *green squares*)

2.3 Explanatory variables

Several explanatory variables were used to construct the gridded datasets of monthly mean temperature and precipitation (Fig. 3). A fine-resolution digital elevation model (DEM) is essential since the topography affects many climate variables; temperature decreases with altitude at approximately 6.5 °C/km (Rolland 2002), and precipitation can be dependent on the orographic effect, where the topography forces air upwards, causing moisture to condense (Adam et al. 2006). The DEM used here has a horizontal resolution of 25 × 25 m (National Land Survey of Finland 2011), and, to correspond to the modeling resolution of 1 km, a 1 km mean elevation, e, was calculated with the `ArcGis 10 aggregate` function, using a cell factor of 40. A smoother elevation surface was used for precipitation to remove possible small-scale spurious topographical variability, calculated with the ArcGis 10 `focal mean` function and a 10 × 10 km moving kernel. It is also possible to derive other explanatory variables from the DEM, such as slope, aspect, curvature, and annual radiation, but these were not considered to be relevant at the modeling scale used (Daly 2006). A square root transformation to elevation was conducted to normalize the distribution.

Finland has many lakes with different sizes and depths, and, because of the heat capacity of water bodies, it is important to incorporate their impact in the spatial models (Solantie 1976; Holdaway 1996). A parameter lake percent, l, describes the percentage surface area of lakes on a 1 km^2 scale and was calculated from Corine 2006 landcover data (spatial resolution of 25 × 25 m) with the ArcGis 10 `focal mean` command with a 6 × 6 km moving kernel. The impact of lakes was assumed to extend 3 km from the lake shore. Similarly, a parameter sea percent, s, was calculated using a moving kernel of 20 × 20 km, with the impact assumed to extend to 10 km inland. Lake percent was not included when interpolating mean precipitation values. In addition, the interpolation routine included x- and y-coordinates.

3 Methods

3.1 Gridding methodology

3.1.1 Kriging

Kriging is a spatial interpolation method that, in its simplest form, predicts a value for a location with no data based on the spatial autocorrelation of observed

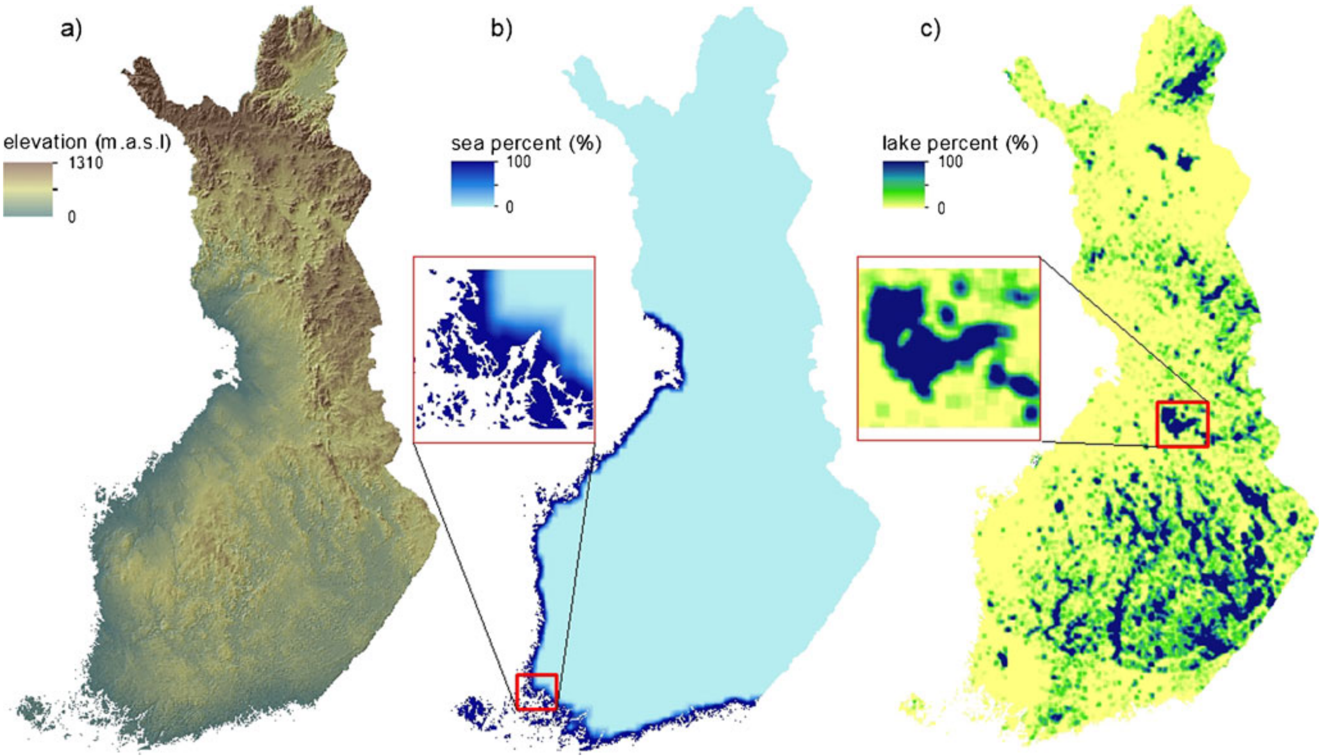

Fig. 3 Explanatory variables used for interpolations. **a** Mean elevation. **b** Sea percentage. **c** Lake percentage

values (Goovaerts 1997). This means that two closely neighboring points are more likely to have similar values than two data points farther apart. Kriging was originally developed for mining purposes based on core samples (Matheron 1963; Cressie 1990). The model of spatial variability can be expressed as (Høst 1999)

$$Z(s) = \mu(s) + \epsilon(s), \tag{1}$$

where $Z(s)$ is the predicted value at some location, $\mu(s)$ is the deterministic function describing the trend component (*drift*) of $Z(s)$, and $\varepsilon(s)$ denotes the stochastic locally varying but spatially dependent residuals.

The simplest form, *ordinary kriging*, is based on the assumption of a constant mean $\mu(s)$ (Goovaerts 1998). Many regionalized variables, however, show a trend structure when analyzed spatially. This violates the assumption of a constant mean and the trend must be removed before performing kriging. For universal kriging (UK), and KED, the unknown mean can vary spatially but must be estimated from the data. UK is a linear prediction with a nonconstant mean structure, utilized when the explanatory variables include coordinates (Journel and Rossi 1989); kriging with auxiliary covariates such as elevation is referred to as kriging with external drift (Goovaerts 2000). The spatial autocorrelation (a small-scale random variation) is modeled with a semivariogram (Goovaerts 1997),

$$2\gamma(h) = \mathrm{Var}(Z(s_i) - Z(s_i + h)), \tag{2}$$

with the usual assumption of intrinsic stationarity, implying that the variance on the right-hand side is only dependent on the vector difference h. We assume, in addition, that the process is isotropic, i.e., the autocorrelation is only dependent on the distance between observations. From the observations, the empirical semivariogram can then be estimated as

$$\hat{\gamma}(\tilde{h}_j) = \frac{1}{2N_h}\sum_{i=1}^{N_h} E(Z(s_i) - Z(s_i + h))^2, \forall h \in \tilde{h}_j, \tag{3}$$

where $Z(s_i)$ and $Z(s_i + h)$ are sample data pairs at a distance h and $\tilde{h}_j$ is the distance between observations. In theory, kriging is an exact interpolator, so that the observed and the predicted value should be the same at the observing location. In practice, this is not the case due to the nugget effect, which represents the measurement error or variation on the microscale. Hence, kriging predictions rely heavily on the appropriate choice of variogram model (Goovaerts 1997).

The main advantages of kriging, relative to nongeostatistical interpolation methods, are that the level of spatial smoothing is defined objectively through the variogram estimated from the observations and that an uncertainty estimate is provided for the predicted values in the form of a variance taking into account the spatial autocorrelation. The semivariogram model is required to specify the expectations under which the kriging variance is calculated. Kriging can be performed within a defined local neighborhood (Walter et al. 2001), but in this work, we used global kriging, using all observations for spatial prediction.

Kriging aims to produce a "best linear unbiased prediction" for an unknown location. It is linear since the estimated values are weighted linear combinations of the available data, unbiased because the mean of the error is 0, and it aims to minimize the variance of the errors (Cressie 1990). When using UK or KED, the trend is substracted, and kriging is then performed with the residuals. The `R` package, `gstat`, does this automatically. However, this can be done separately in two steps with an approach termed regression-kriging (Hengl et al. 2003, 2004). The advantage of regression-kriging is that the parametric trend model can be replaced with a more flexible non-parametric model, while still retaining the benefits of kriging.

The variogram describes how observations are related with distance (and direction). We used an exponential model (Bivand et al. 2008) for the variogram,

$$\gamma(h) = c(1 - e^{-h/a}), \tag{4}$$

where c is the partial sill representing the variance of the random field, and a is the range parameter describing the distance over which observations are still autocorrelated. Initialization parameters are required for `gstat`, but the fitting of the variogram model is performed automatically by minimizing the weighted sum of the squared errors (Bivand et al. 2008),

$$\sum_{j=1}^{p} \omega_j(\gamma(h) - \hat{\gamma}(h))^2, \tag{5}$$

with $\gamma(h)$, the value according to the parametric model is minimized. The weights are given as N_j/h_j^2.

3.1.2 The calibration of the kriging models

Kriging interpolations were performed in the `R` statistical programming environment (R Development Core Team 2011) with package `gstat` (Pebesma 2004). The trend surfaces were fitted using the least-squares method, and the same terms were used for all months. The final trend model for monthly temperature was

$$Z(x, y, e, l, s) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2 + \beta_6 e + \beta_7 l + \beta_8 s, \tag{6}$$

and, for monthly precipitation,

$$Z(x, y, e, s) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2 + \beta_6 e + \beta_7 s, \tag{7}$$

where x and y are the coordinates in meters (Euref Fin TM35 coordinate reference system), e is the mean elevation, l is the lake percentage, and s is the sea percentage. Potential anisotrophy of the spatial autocorrelation was investigated for some of the months, but no evidence of such a behavior was found. With the mean temperature, the range was fixed to 80 km to avoid singular matrix errors. This often happens when all of the variogram model parameters cannot be estimated properly. It can result from duplicate observations in the dataset or perfectly correlated variables in the regression model (Bivand et al. 2008). The distance 80 km has been used for previous interpolations and has proved to be a reasonable estimate of the range of the spatial autocorrelation.

For monthly mean precipitation, we used information from previous months to obtain the range parameter for the variogram model

$$m = \text{mean}(\text{stdev}(RR1961 - 2010)), \tag{8}$$

$$\text{Range} = 120,000 - \text{stdev}(RRmon)/m \times 40,000, \tag{9}$$

where *RR1961–2010* is the total rainfall of the season 1961–2010 for every month and *RRmon* is the observed monthly mean precipitation. Equation 9 essentially states that when the standard deviation of the monthly precipitation is large, the range parameter (the distance of autocorrelated observations) decreases and vice versa. In practice, the range parameter takes values from 40 to 120 km, and empirical evidence has suggested that this is the interval that most of the range parameters were estimated automatically. These values were also chosen for mapping purposes to produce realistic interpolations without overly spotted or smoothed maps, attempting to capture whether the precipitation was frontal or convective in nature, instead of using the same range parameter for every month. For instance, a large standard deviation in monthly precipitation is assumed to correspond to precipitation of a more convective precipitation, with more small-scale spatial variability, and a smaller range parameter. A nugget

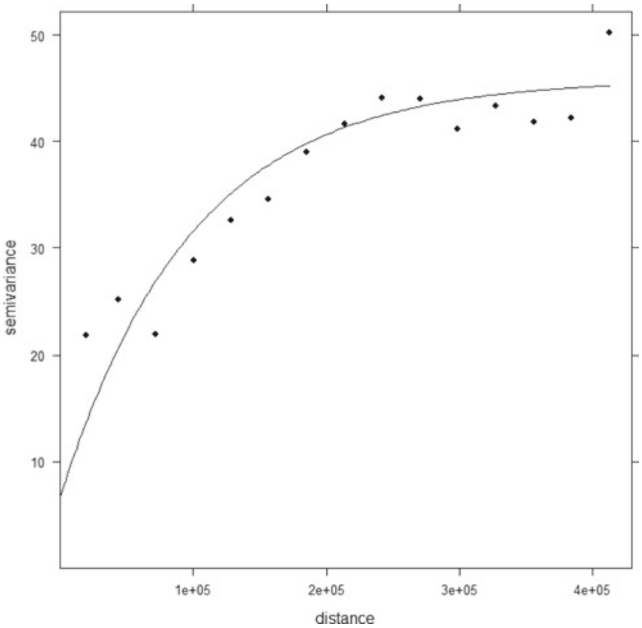

Fig. 4 Variogram for January 2010. Standard deviation of the monthly precipitation (nugget) is 6.7, indicating frontal precipitation, and the range parameter is 99 km

parameter was taken from the standard deviation of monthly mean precipitation, as higher variation could presumably lead to larger measurement errors. In Figs. 4 and 5, examples of two variograms are shown,

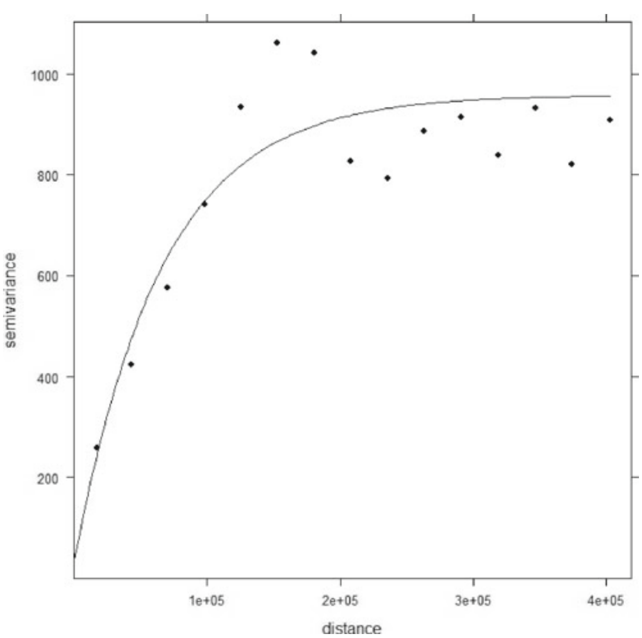

Fig. 5 Variogram for August 2008. Standard deviation of the monthly precipitation (nugget) is 35, indicating convective precipitation, and the range parameter is 66 km

representing the cases from January 2010 and August 2008. No fixed nugget was used for mean temperatures.

3.1.3 Generalized additive models

Kriging with external drift assumes linearity between the response and the covariates, but this assumption rarely holds true in environmental studies. Linear regression models are not always flexible enough to describe the true process– environment relationship (Yee and Mitchell 1991). GAM are semiparametric extensions of GLM (McCullagh and Nelder 1989) where no parametric coefficients are estimated for predictors. GAM allow a nonlinear relationship between the response and explanatory variables (Hastie and Tibshirani 1990). The probability distribution of the response should be known, which is why these models are referred as semiparametric. In GAM, the combination of predictors are related to response variables through a link function, which allows a transformation to linearity and any predictions to be maintained within the range of the response variable. GAM have the form (Hastie and Tibshirani 1990; Wood and Augustin 2002)

$$g(\mu) = \beta_0 + s_1(x_1) + s_2(x_2) + ... + s_k(x_k), \tag{10}$$

where $g(\mu)$ is the link function, β_0 is a constant, s_k is the smoothing parameter to be estimated, and x_k is the explanatory variable.

Nonlinear and smoothing regression techniques, such as GAM, are data-driven and do not force the shape of the response to any parametric form. An important step when fitting GAM is to choose an appropriate level of smoothing through limiting the degrees of freedom; the variation in the data can be captured correctly while avoiding the danger of overfitting (Guisan et al. 2002; Wood 2004). One benefit over linear models is that GAM take into account the shape of the response with no higher order terms required. This flexibility comes at a cost, the iterative smoothing parameter estimation can be computationally heavy, especially with large datasets (Leathwick et al. 2006). GAM aim to minimize the generalized cross validation (GCV) criterion (Wood 2011),

$$nD/(n - \mathrm{DoF})^2, \tag{11}$$

where D is the deviation, n the number of data points, and DoF is the effective degrees of freedom in the model. By combining the GAM-modeled trend with geostatistics, we obtain *geoadditive models* (Kammann and Wand 2003; Kneib et al. 2009). The large-scale trend is modeled with GAM, and kriging is then performed on the spatially autocorrelated residuals. The final interpolation result is a combination of both elements (Ginsbourger et al. 2009) referred to as GK. An example is shown by Brown et al. (2002), who combined geostatistics with GAM for land cover change modeling.

3.1.4 Calibration of GAM

GAMs were fitted using the `R` statistical programming environment with package `mgcv` (Wood 2011). The same model terms were used for all 360 months, even though automatic term selection is available within the `mgcv` package. The maximum number of degrees of freedom for explanatory variables (the level of smoothing) was restricted to 4. The xy-interaction term was modeled using the `te` function, designed specially for anisotropic interactions (Wood 2006). The response variables were expected to be normally distributed with a canonical `identity`-link function. Ordinary kriging was applied to the residuals from GAM, and, in order to obtain the final GK prediction, the estimated trend and residuals were combined into one final interpolation using residual variogram models parametrized as described in Section 3.1.2.

3.2 Evaluation

Probably, the most important part of any interpolation task is to evaluate the concordance between predictions and reality. Twenty weather stations were chosen as verification stations for mean temperature (22 for precipitation) based on their spatial coverage (Fig. 2d). The verification stations were not included in the original predictions, allowing true independent cross-validation. Three statistical measures were calculated to assess the performance: root mean squared error (RMSE), mean absolute difference (MAD), and Pearson's correlation coefficient (COR) (Hofstra et al. 2008). According to Wilmott (1982), RMSE and MAD provide good overall measures of model performance, as they summarize the mean error in the same units as the observed and predicted values. RMSE is defined here as

$$\mathrm{RMSE} = \sqrt{n^{-1}\sum_{i=1}^{n}(O_i - P_i)^2}, \tag{12}$$

where O_i is the observation and P_i the predicted value. RMSE is sensitive to outliers and can be used as indicator of the magnitude of extreme errors (Price et al. 2000). MAD is less sensitive to extreme values and is defined here as

$$\mathrm{MAD} = n^{-1}\sum_{i=1}^{n}|O_i - P_i|. \tag{13}$$

For every monthly interpolation datafile created by the KED and GK routine, there is additional information attached about the skill of the prediction obtained from the `gstat` inbuilt cross-validation `krige.cv` function. This function performs *leave-one-out* cross-validation which can be used for validating monthly KED and GK interpolations. Standardized prediction error (SPE) is also calculated;

$$\mathrm{SPE} = \frac{(O_i - P_i)}{\sqrt{\mathrm{Var}_i}}, \tag{14}$$

where Var_i is the kriging variance.The standard deviation of SPE, σ_{SPE}, is ideally close to one indicating that the magnitude of the kriging variance is correct, and the variogram model is realistic (Bivand et al. 2008). Also included in the output datafiles are the lower (2.5 %) and upper (97.5 %) quantiles of SPE, which, if close to −2 and 2, indicating normally distributed prediction errors as assumed. These measures provide a quick overview of the success of the interpolation routine.

4 Results

4.1 Validation of monthly interpolations

An example of the statistical measures included with every monthly KED output file is presented in Table 1. For both monthly mean temperature and monthly mean precipitation, the errors are small and correlation coefficients are high, and SPE appears to be normally distributed as expected. KED slightly underestimates the kriging variance for mean temperature, with $\sigma_{\mathrm{SPE}} > 1$, whereas the kriging variance of mean precipitation is slightly overestimated.

Table 1 Example of statistical measures included with the KED output file for September 2010

	Temperature	Precipitation
σ_{SPE}	1.03	0.85
2.5 % quantile	−1.84	−1.93
97.5 % quantile	1.88	1.40
RMSE	0.29	8.3
MAD	0.04	0.63
COR	0.99	0.91

Table 2 Interpolation results for monthly mean temperature over the period of 1981–2010

	KED	GAM	GK	Baseline
RMSE	0.36	0.32	0.34	3.18 °C
MAD	0.07	0.05	0.06	
COR	0.98	0.99	0.98	

Table 4 Minimum and maximum skill score for monthly mean temperature during 1981–2010

		KED	GAM	GK
RMSE	Min	0.19	0.14	0.15
	Max	0.88	0.91	0.87
MAD	Min	2.5e-4	6.94e-05	7.98e-5
	Max	0.27	0.22	0.27
COR	Min	0.80	0.83	0.76
	Max	0.998	0.999	0.999

4.2 Validation for 1981–2010

The following statistics have been calculated for the entire reference period of 1981–2010. The results are summarized in Tables 2 and 3. The baseline is the mean of the monthly observations over the 30-year period.

For monthly mean temperature, the results of the different interpolation methods show very small differences. GAM displays the smallest RMSE (0.32) and MAD (0.05). All three methods provide similar high correlation coefficients between observed and predicted values (≥ 0.98). For monthly mean precipitation, KED returned the lowest RMSE (9.06), and GAM the lowest MAD (1.93). The highest correlation coefficient was produced by KED (0.84).

Tables 4 and 5 display the minimum and maximum value for every skill score, from which it is possible to asses the variability of the performance of each interpolation method. For example, Table 4 shows that the range of correlation coefficients for mean temperature is smallest for GAM, indicating consistent prediction; for mean precipitation, however, the range of RMSE values shown by GAM is very large, and the smallest range of correlation values is displayed by KED.

4.3 Temporal differences

Figure 6 reveals that there are clear seasonal patterns in the performance of the interpolation methods. Monthly mean temperatures display the largest RMSE in winter months and smallest in late summer–early autumn (Fig. 6a). GAM had the lowest median RMSE for 11 out of 12 months. For MAD, GAM produced low errors throughout all months, while KED and GK displayed more error during the summer months (Fig. 6c). Correlation coefficients for monthly temperatures were very high for all months (Fig. 6e), even taking into account a slight dip in performance exhibited in July where the mean of the lowest correlation was still well over 0.95.

Clear seasonal patterns in the performance are also seen for monthly mean precipitation. Summer months display the largest RMSE values (Fig. 6b) and lowest correlations (Fig. 6f) presumably due to the high spatial variation in convective precipitation. The largest errors and lowest correlations were found for GAM, and there was little difference between the output from KED and GK throughout the seasons. However, the results for the MAD skill score show that GAM is still competitive in terms of bias.

Table 3 Interpolation results for monthly mean precipitation over the period of 1981–2010

	KED	GAM	GK	Baseline
RMSE	9.06	10.52	9.36	51.0 mm
MAD	2.09	1.93	2.12	
COR	0.84	0.78	0.83	

4.4 The new reference period of 1981–2010

The annual mean temperature and precipitation for the new reference period of 1981–2010 is presented in Fig. 7. Monthly KED interpolations are based on station values, and the final maps for 1981–2010 are the grid-averaged values for the entire period. The mean surface temperature in Finland for 1981–2010 is 2.1 °C, and the mean annual precipitation is 607 mm. As a comparison, in 1971–2000 the mean temperature was 1.7 °C, calculated from the grid values, giving an increase of 0.38 °C ($p < 2.2 \times 10^{-16}$ from Mann–Whitney test). The mean annual precipitation for the last reference period was 588 mm, so rainfall has increased by

Table 5 Minimum and maximum skill score for monthly mean precipitation during 1981–2010

		KED	GAM	GK
RMSE	Min	1.70	1.66	1.77
	Max	27.0	32.00	24.60
MAD	Min	0.03	0.00	0.00
	Max	9.06	9.38	8.15
COR	Min	0.46	0.12	0.40
	Max	0.98	0.97	0.99

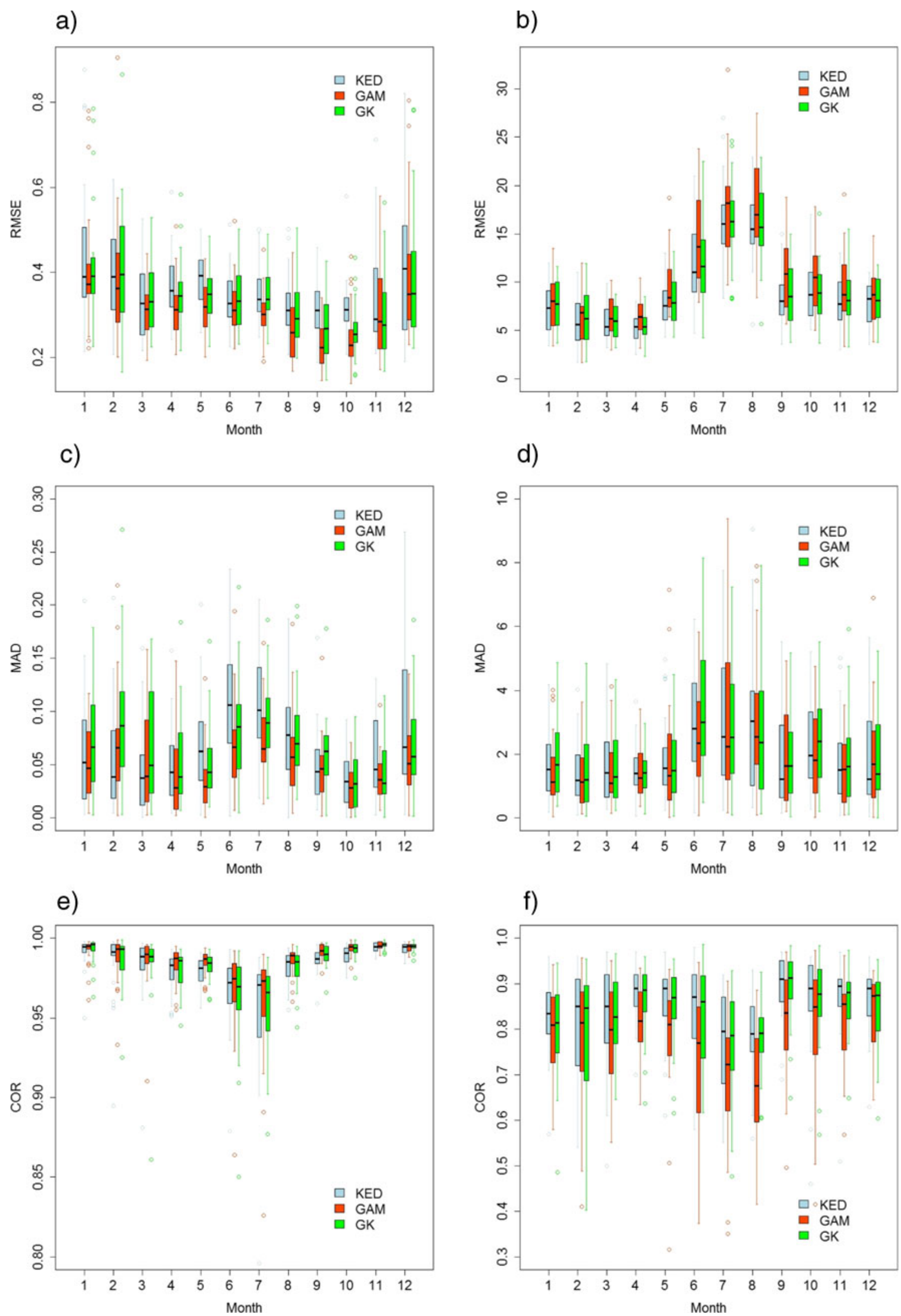

Fig. 6 Boxplots for each validation skill score (*RMSE*, *MAD*, and *COR*). *Left column* is for monthly mean temperature (**a**, **c**, **e**) and *right column* for monthly mean precipitation (**b**, **d**, **f**)

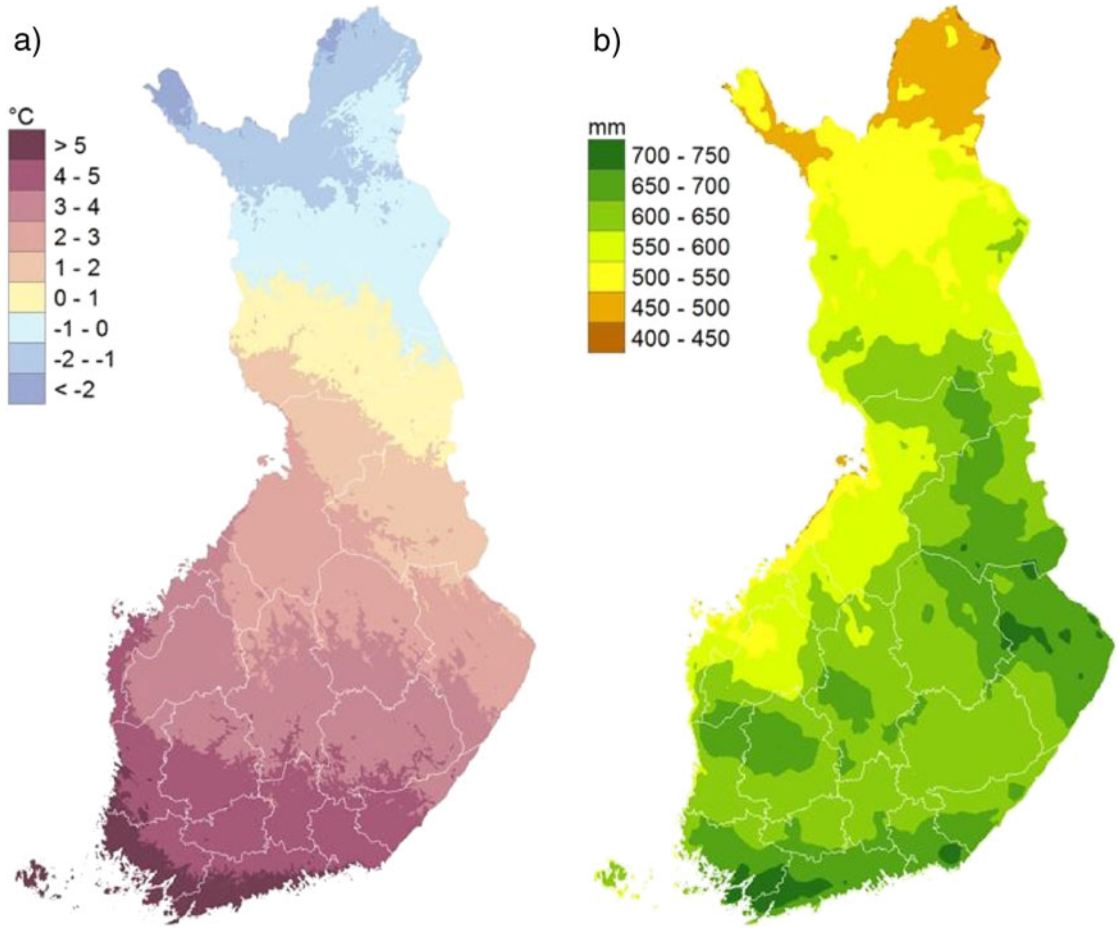

Fig. 7 Map of reference values of mean temperature (**a**) and mean precipitation (**b**) for the new reference period of 1981–2010, calculated by using KED from monthly gridded values to provide a grid-averaged value for the entire period

3 % ($p < 2.2 \times 10^{-16}$) on average. The most notable increase in temperatures has been in eastern Lapland, and precipitation has increased in both northern and eastern Lapland as well as in western Finland.

5 Discussion

The choice of interpolation method depends on data type, desired accuracy, area of interest, computation capacity, and the spatial scale used. The quality of the spatial prediction is also subject to the spatial patterns of the observations (Brus and Heuvelink 2007). Hong et al. (2005) states that thin-plate smoothing techniques (such as GAM) are superior to other methods for interpolating climate data. They interpolated monthly mean temperatures and precipitation using ANUSPLIN software. Hutchinson (1998) has also interpolated rainfall using thin-plate splines. Our results imply that without taking into consideration the spatially autocorrelated residuals, it is not possible to accurately interpolate monthly mean precipitation. Spatial autocorrelation does not play an important role for monthly temperatures presumably due to the more stochastic nature of rainfall. When interpolating monthly mean temperatures, the use of more flexible GAM produced more accurate predictions than geostatistical methods.

Overall, the results for both temperature and precipitation are notably good when compared to results reported by others (Price et al. 2000; Vicente-Serrano et al. 2003; Perry and Hollis 2005; Hong et al. 2005). Even so, it should be stated that Finland has relatively flat topography with the exception of northernmost Lapland, and it would be expected that the interpolation of monthly mean temperatures is a fairly straightforward task to perform. From cross-validation results, the largest residuals for both variables are located in northern Finland, where the topography is more complex, and the station network is more sparse. Errors were also larger near the borders, so accuracy of the predictions could be improved by adding observations from neighboring countries.

The performance of the different interpolation methods varied greatly in a temporal sense. For temperature, the largest RMSE values were obtained during winter months (Fig. 6a). This may be due to some missing explanatory variable, such as curvature or aspect, but these showed little or no relevance at the spatial scale used (Daly 2006). Also, the uncertainty associated with the prediction (the width of the boxplot bands) was substantially larger in winter for every method. All methods produced very high correlation coefficients throughout all months and, according in Fig. 6e, the highest coefficients with the smallest uncertainty are in winter, at the same time as with the largest RMSE values. Conversely, summer months show slightly lower correlations. This inconsistency of correlation coefficients and RMSE values is a result of minor nonlinearities in the relationship between observed and predicted values. A small systematic prediction error exists during summer months with every method, causing underestimation of the lowest and highest temperatures. Larger RMSE values occur in winter at the lower temperatures but are more evenly distributed, producing both higher correlation coefficients and relatively higher RMSE values. Also notable is that GAM produced low MAD values throughout all months, with KED and GK producing larger errors in summer months; this may be due to an inadequate trend model for KED, and a combination of complicated trend and residual kriging in the case of GK.

Figure 6b reveals that the highest RMSE values for monthly mean precipitation occur during summer due to convective processes. Precipitation is predominantly frontal in autumn and winter, with less associated small-scale variability, and is more straightforward to interpolate spatially, therefore plain GAM can produce fairly good results. As convective activity grows, with an associated increase in randomness, kriging methods are needed for accurate estimates. Figure 6f, displaying the lowest correlation coefficients during summer months, is essentially the inverse of Fig. 6b. All methods, but especially GAM, show a wide variation in correlation coefficient; the boxplot reveals that KED and GK maintain reasonably high correlation throughout the year, with KED being slightly better (Table 3). It is notable, that GAM produces the lowest MAD values for mean precipitation, indicating relatively good overall accuracy with larger range of errors (high RMSE and low COR).

The kriging predictions were more robust than GAM with interpolated values lying in a reasonable range. This is a very important feature of an interpolation method since the routine must produce stable and accurate predictions for all 360 months. The main challenge of this project was to create trend and variogram models that perform adequately at the very least, with a compromise between good overall prediction accuracy and reduced bias between observed and predicted values at any one particular observation location. Overall interpolation accuracy was selected to be more important, so there were no fixed nugget parameters in the variogram model for mean temperature; the variogram was calculated separately for every month. Because spatial autocorrelation is a real feature in geographical datasets, it is preferable to fix the variogram model parameters as little as possible, and the variogram model for mean precipitation adapted the nugget and range parameters according to the standard deviation of the monthly rainfall.

The accuracy of the interpolations could be improved if, for every month, an optimal trend model is constructed by minimizing criteria such as Akaike's information criteria, or GCV in the case of GAM. An automatic model term selection function exists in Rs interpolation packages for both linear models and GAM, useful for fitting large numbers of models. However, we selected a more hypothesis-based model containing the same terms for each month. The main explanatory variables used for interpolating monthly mean temperature and precipitation are x, y, and mean elevation (Goovaerts 2000; Price et al. 2000; Hong et al. 2005). Additional useful variables are sea percentage and lake percentage (Holdaway 1996; Vajda and Venäläinen 2003). At the spatial and temporal scale used here (1×1 km, 1 month), any additional auxiliary variables add little extra value to the predictions (Daly 2006). Predicting (monthly) mean temperatures at finer spatial and temporal resolutions could be improved by using variables such as curvature, aspect, forest cover, and soil cover. The use of remote-sensing data for input may also be useful for interpolating climate variables (Hengl et al. 2012).

Clearly, using a nonparametric method for spatial prediction has advantages, as seen in the case of monthly mean temperature. Combined with residual kriging, results can be improved further, and the method expanded to other climatological variables. However, nonparametric spatial modeling methods such as GAM are based on more complicated mathematical theory and the results can be difficult to interpret. GAM are very useful for modeling purposes to study the real shape of the response, and for finding possible threshold values (Yee and Mitchell 1991). Kriging with a linear trend model is more straightforward which can provide similar performance to nonparametric methods but with a simpler theoretical background.

6 Conclusions

According to the cross-validation statistics for the 30-year period, we have shown that there was little difference between the methods tested, especially for mean temperatures, where all three methods produced results with small errors and high correlation between observed and predicted values. GAM was the best overall method for predicting monthly mean temperature. The same interpolation methods showed larger differences for monthly mean precipitation, with KED and GK showing best overall performance. A strong seasonal variation in the performance of the different interpolation methods was displayed for both monthly mean temperature and precipitation.

It is possible to improve interpolation accuracy using nonparametric methods such as GAM for climate variables which show little or no spatial autocorrelation between observations (e.g., temperature). When combined with residual kriging, these methods can be extended to variables with spatially autocorrelated errors (e.g., precipitation). As predictors for interpolating the spatial variation of mean temperature and precipitation, we used mean elevation, sea percentage, and lake percentage. KED proved accurate and stable and was selected for interpolating monthly interpolations because of the robustness of kriging.

For the calculation of the new reference period of 1981–2010, the evaluation of the monthly interpolations has been improved and, in addition, statistical measures describing the quality of interpolation have been included in every interpolation file.

Acknowledgement We are grateful to Ewan O'Connor for improving the English of the manuscript.

References

Adam J, Clark E, Lettenmaier D, Wood E (2006) Correction of global precipitation products for orographic effects. J Climate 19:15–38

Araújo M, Luoto M (2007) The importance of biotic interactions for modeling species distribution under climate change. Glob Ecol Biogeogr 16:743–753

Barnett T, Zwiers F, Hegerl G, Allen M, Crowley T, Gillett N, Hasselmann K, Jones P, Santer B, Schcnur P, Stott K, Taylor K, Tett S (2005) Detecting and attributing external influences on the climate system: a review of recent advances. J Climate 27:1291–1314

Bivand R, Pebesma E, Còmez-Rubio V (2008) Applied spatial data analysis with R. Springer, New York

Brown D, Goovaerts P, Burnicki A, Li MY (2002) Stochastic simulation of land-cover change using geostatistics and generalized additive models. Photogramm Eng Remote Sensing 68(10):1051–1061

Brus D, Heuvelink G (2007) Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138:86–95

Chorley R, Haggett P (1965) Trend-surface mapping in geographical research. Trans Inst Br Geogr 37:47–67

Clifton P, Neuman S (1982) Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in souther Arizona. Water Resour Res 18(4):1215–1234

Cressie N (1990) The origins of kriging. Math Geol 22(3):239–252

Daly C (2006) Guidelines for assessing the suitability of spatial climate data sets. Int J Climatol 26:707–721

Drebs A, Nordlund A, Karlsson P, Helminen J, Rissanen P (2002) Climatological statistics of Finland 1971–2000. Climatological Statistics of Finland 2001. Finnish Meteorological Institute, Helsinki

Fronzek S, Luoto M, Carter T (2006) Potential effect of climate change on the distribution of palsa mires in subarctic Fennoscandia. Clim Res 32:1–12

Ginsbourger D, Dupuy D, Badea A, Carraro L, Roustant O (2009) A note on the choice and the estimation of kriging models for the analysis of deterministic computer experiments. Appl Stoch Models Bus Ind 25(2):115–131

Goovaerts P (1997) Geostatistics for natural resources evaluation. Applied Geostatistics Series, Oxford University Press

Goovaerts P (1998) Ordinary cokriging revisited. Math Geol 30(1):21–42

Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89:1–45

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

Guisan A, Edwards T, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol Model 157:89–100

Hastie T, Tibshirani R (1984) Generalized additive models. Technical report lcm02, Dept. of Statistics

Hastie T, Tibshirani R (1990) Generalized additive models, monographs on statistics and applied probability, vol 43. Chapman and Hall, New York

Hengl T, Heuvelink G, Stein A (2003) Comparison of kriging with external drift and regression-kriging. Technical note, ITC

Hengl T, Heuvelink G, Stein A (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120:75–93

Hengl T, Heuvelink G, Tadić M, Pebesma E (2012) Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. Theor Appl Climatol 107:265–277

Hjort J, Luoto M (2010) Novel theoretical insights into geomorphic process-environment relationships using simulated response curves. Earth Surf Processes Landf 36(3):363–371

Hofstra N, Haylock M, New M, Jones P, Frei C (2008) Comparison of six methods for the interpolation of daily, European climate data. J Geophys Res 113:D21,110

Holdaway M (1996) Spatial modelling and interpolation of monthly temperature using kriging. Clim Res 6:215–225
Hong Y, Nix H, Hutchinson M, Booth T (2005) Spatial interpolation of monthly mean climate data for China. Int J Climatol 25:1369–1379
Høst G (1999) Kriging by local polynomials. Comput Stat Data Anal 29:295–312
Hutchinson M (1998) Interpolation of rainfall data with thin plate smoothing. J Geogr Inf Decis Anal 2(2):139–151
Journel A, Rossi M (1989) When do we need a trend model in kriging? Math Geol 21(7):715–739
Jylha K, Tuomenvirta H, Ruosteenoja K (2004) Climate change projections for Finland during the 21st century. Boreal Environ Res (9):127–152
Kammann E, Wand M (2003) Geoadditive models. JR Stat Soc 52(1):1–18
Kneib T, Hothorn T, Tutz G (2009) Variable selection and model choice in geoadditive regression models. Biometrics 65(2):626–634
Kullmann L (2010) A richer, Greener and Smaller Alpine world: review and projection of warming-induced plant cover change in the Swedish Scandes. Ambio 39:159–169
Laslett G (1994) Kriging and splines: an empirical comparison of their predictive performance in some applications. J Am Stat Assoc 89(426):391–400
Leathwick J, Elith J, Hastie T (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modeling of species distributions. Ecol Model 199:188–196
Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74:1659–1673
Lin G, Chen L (2004) A spatial interpolation method based on radial basis function networks incorporating a semivariogram model. J Hydrol 288:288–298
Liu TL, Juang KW, Lee DY (2006) Interpolating soil properties using kriging combined with categorical information of soil maps. Soil Sci Soc Am J 70(4):1200–1209
Lu G, Wong D (2008) An adaptive inverse-distance weighting spatial interpolation technique. Comput Geosci 34(9):1044–1055
Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266
McCullagh P, Nelder J (1989) Generalized linear models, monographs on statistics and applied probability, vol 37, 2nd edn. Chapman and Hall, New York
Miller J, Franklin J (2002) Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. Ecol Model 157: 227–247
National Land Survey of Finland (2011) Terrain database. http://www.usc.fi/tutkimus/alat/geotieteet/paikkatieto/paituli
Pebesma E (2004) Multivariable geostatistics in S: the gstat package. Comput Geosci 30:683–691
Perry M, Hollis D (2005) The generation of monthly gridded datasets for a range of climatic variables over the UK. Int J Climatol 25:1041–1054
Price D, McKenney D, Nalder I, Hutchinson, Kesteven J (2000) A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. Agric For Meteorol 101:81–94
R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org, ISBN 3-900051-07-0
Rizzo D, Dougherty D (1994) Characterization of aquifer properties using artificial neural networks: neural kriging. Water Resour Res 30(2):483–497
Robertson G (1987) Geostatistics in ecology: interpolating with known variance. Ecology 68(3):744–748
Rolland C (2002) Spatial and seasonal variations of air temperature lapse rates in Alpine regions. J Climate 16:1032–1046
Solantie R (1976) The influences of lakes on meso-scale analysis of temperature in Finland. Ilmatieteen laitoksen tiedonantoja 30, p 130 (in Finnish)
Tikkanen M (2005) Climate. In: Seppälä M (ed) The physical geography of Fennoscandia. Oxford University Press
Vajda A, Venäläinen A (2003) The influence of natural conditions on the spatial variation of climate in Lapland, northern Finland. Int J Climatol 23:1011–1022
Vicente-Serrano S, Saz-Sánchez M, Cuadrat J (2003) Comparative analysis of interpolation methods on the middle Ebro Valley (Spain): application to annual precipitation and temperature. Clim Res 24:161–180
Voltz M, Webster R (1990) A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. J Soil Sci 41(3):473–490
Walter C, McBratney A, Douaoui A, Minasny B (2001) Spatial prediction of topsoil salinity in the Chelif Valley, Algeria, using local ordinary kriging with local variograms versus whole-area variogram. Aust J Soil Res 39(2):259–272
Wilmott C (1982) Some comments on the evaluation of model permormance. Bull Am Meteorol Soc 63(11):1309–1313
Wood S (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. J Am Stat Assoc 99:673–686
Wood S (2006) Generalized additive models: an introduction with R. Chapman & Hall, London
Wood S (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J R Stat Soc 73(1):3–36
Wood S, Augustin N (2002) GAMs with integrated model selection using penalized regression splines and applications to environmental modeling. Ecol Model 157:157–177
Yee T, Mitchell N (1991) Generalized additive models in plant ecology. J Veg Sci 2:587–602