Spatial Statistics 14 (2015) 91–113

Contents lists available at ScienceDirect

Spatial Statistics

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

Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania

Tim Appelhans [a,*], Ephraim Mwangomo [a,b], Douglas R. Hardy [c], Andreas Hemp [d], Thomas Nauss [a]

[a] *Environmental Informatics, Philipps University of Marburg, Germany*

[b] *Tanzania National Parks, Arusha, Tanzania*

[c] *Climate System Research Centre and Department of Geosciences, University of Massachusetts Amherst, USA*

[d] *Plant physiology, Bayreuth University, Germany*

ARTICLE INFO

Article history:
Received 31 October 2014
Accepted 28 May 2015
Available online 17 June 2015

Keywords:
Spatial interpolation
Machine learning
Air temperature
Kriging
Cubist
Cross-validation

ABSTRACT

Spatially high resolution climate information is required for a variety of applications in but not limited to functional biodiversity research. In order to scale the generally plot-based research findings to a landscape level, spatial interpolation methods of meteorological variables are required. Based on a network of temperature observation plots across the southern slopes of Mt. Kilimanjaro, the skill of 14 machine learning algorithms in predicting spatial temperature patterns is tested and evaluated against the heavily utilized kriging approach. Based on a 10-fold cross-validation testing design, regression trees generally perform better than linear and non-linear regression models. The best individual performance has been observed by the stochastic gradient boosting model followed by Cubist, random forest and model averaged neural networks which except for the latter are all regression tree-based algorithms. While these machine learning algorithms perform better than kriging in a quantitative evaluation, the overall visual interpretation of the resulting air temperature maps is ambiguous. Here, a combined Cubist and residual kriging approach can be considered the best solution.

* Corresponding author.
E-mail address: tim.appelhans@gmail.com (T. Appelhans).

http://dx.doi.org/10.1016/j.spasta.2015.05.008

1. Introduction

For biodiversity and ecosystem research, climate conditions are a major explanatory variable (e.g. Sala, 2000; Hawkins et al., 2003; Currie et al., 2004) and a common demand of biodiversity researchers is to get plot-scale information on the weather and climate conditions. The problem is complicated by the fact that the world wide weather station network is at its minimum or virtually absent in the regions of the world's biodiversity hotspots (see Myers et al., 2000) and that regional climate model simulations or satellite observations are generally to coarse (or less frequent) to fully meet the demands of the biodiversity community regarding plot-scale observations. So basically, the question to be addressed is how we can provide accurate weather status information for biodiversity and ecosystem research which often demands site specific information for certain intensively investigated research plots and area-wide information on a landscape scale.

The common solution for this problem is the installation of individual stations by the respective researchers working in the area. However, due to funding and man power restrictions, the stations can generally not be installed on each of the biodiversity observation plots (e.g. Fries et al., 2009). But even if this would be possible, area-wide weather and climate information is mandatory if the individual plot-based findings are to be transferred to the landscape scale. This necessitates the application of either spatial interpolation or downscaling techniques which potentially provide spatially high resolution weather datasets based on individual station observations or medium to high-resolution climate model simulations and/or meteorological satellite observations. Even though the latest CORDEX climate model runs for e.g. the African domain (Panitz et al., 2014) or the brand new microclim dataset (Kearney et al., 2014) have a resolution of 0.22° and 0.17° respectively, the grid edge length at the equator of roughly 24.5 km and accordingly 15 km is still to coarse to cover the local to landscape scale patterns relevant for many research approaches (e.g. the Mt. Kilimanjaro region is represented in the microclim dataset by 12 grid cells). Similar limitations hold true for the application of satellite observations which drop beyond about a 1 km by 1 km resolution as soon as the temporal resolution is daily or better. For the Kilimanjaro region as for any other mountain system the situation is further complicated because of the modification of meso-scale weather by micro-topographic site conditions (Loeffler et al., 2006). Hence, spatial interpolation or prediction techniques are still required to derive the high resolution meteorological fields demanded by functional biodiversity research.

An amplitude of studies exists on the utilization of different spatial interpolation methods for meteorological parameters across different regions. Until recently and neglecting quite specialized approaches like PRISM (Daly, 2006; Daly et al., 2007) or DAYMET (Thornton et al., 1997), these methods could be divided into simpler methods like distance weighting (e.g. Lennon and Turner, 1995; Willmott and Matsuura, 1995; Nalder and Wein, 1998) or polynomial interpolations (e.g. Tabios and Salas, 1985; Ashraf et al., 1997; Goodale et al., 1998; Xia et al., 1999; Ninyerola et al., 2000) and the more advanced geostatistical interpolation techniques of kriging and splining. The various forms of kriging (Krige, 1951) use linear weighting combinations at the known data points to predict the parameter of interest at points where no measurements are available (e.g. Holdaway, 1996; Ashraf et al., 1997; Diodato, 2005). Compared to kriging where the statistical model has to be subjectively selected, splining, i.e. fitting splines to the known data points is less dependent on the underlying statistical model but on the other hand it requires regularly spaced input data (Hulme et al., 1995; Hutchinson, 1995; Price et al., 2000; Xia et al., 2001).

Several evaluation studies of such geostatistical approaches have been carried out by various authors indicating that kriging produces generally more accurate results than other interpolation techniques (e.g. Ashraf et al., 1997; Goovaerts, 2000; Apaydin et al., 2004; Ustrnul and Czekierda, 2005; Chen et al., 2007; Hofstra et al., 2008). Even if kriging does not perform best with respect to typical validation indices (e.g. mean square error), the resulting interpolation fields might be more plausible than the ones from other techniques (e.g. Collins and Bolstad, 1996). For applications in regions with highly complex topography, different ancillary data sources like digital elevation models or land-use classifications have been used (e.g. Jarvis and Stuart, 2001; Hasenauer, 2003; Stahl et al., 2006; Baltas, 2007; Daly et al., 2007; Guler et al., 2007; Di Luzio et al., 2008) which are especially important for temperature interpolation techniques (e.g. Vicente-Serrano et al., 2003) or the derivation of sheltering

factors used to interpolate wind speed (Ryan, 1977; Worlen et al., 1999). For a comprehensive review of geostatistical interpolation methods, see Li and Heap (2008).

With increasing computation power, machine learning approaches have extended the group of statistical interpolation techniques for environmental datasets during the last couple of years. However, most of these studies focus on biodiversity patterns. For example, Attorre et al. (2013) predicted proper niche areas for tree and reptile species of Socotra Island, Yemen, Melini (2013) used such methods for tree species distribution in the Tuscany region, Italy and Siaulys and Bucas (2012) predicted benthic invertebrates distribution in the south-eastern Baltic sea. Thereby, the non-parametric random forest method performed best (see also Pal, 2005; Peters et al., 2007; Polishchuk et al., 2009; Liess et al., 2012) although – if considered at all – ordinary kriging results are also quite robust for e.g. predicting deep sea biodiversity (Li et al., 2011b, 2012; Li, 2013) and random forest and kriging can also be combined into one approach (e.g. Li et al., 2011a).

Regarding meteorological parameters, machine learning-based approaches are sometimes used for satellite rainfall retrievals (e.g. Behrangi et al., 2009; Kühnlein et al., 2014), the downscaling of model or satellite observations (e.g. Snell et al., 2000; Tripathi et al., 2006; Ho et al., 2014) and still rarely for spatial prediction (e.g. Pozdnoukhov et al., 2009) also more studies exist which use such approaches for (short-term) forecasting (e.g. Tasaduqq and Rehman Bubshait, 2002; Abdel-Aal, 2004; Ustaoglu et al., 2008; Dombayc and Glc, 2009; Pozdnoukhov et al., 2009; Radhika and Shashi, 2009; Smith et al., 2009; Paniagua-Tineo et al., 2011; Ortiz-Garca et al., 2012). The study from Pozdnoukhov et al. (2009) also showed that the kriging variants with some kind of external drift are superior to some neuronal network derivatives when applied to air temperature interpolations in Switzerland.

To broaden the study basis with a special focus on machine learning algorithms and to show the performance of not just one but many groups of machine learning or geostatistical approaches in highly complex terrain and under the same boundary conditions, we present a comprehensive analysis of the skill of 14 different models for the spatial prediction of monthly air temperature in the Mt. Kilimanjaro region and test them against the well known kriging with external drift as a geostatistical reference method.

2. Study region

The study is part of the research unit ‘Kilimanjaro ecosystems under global change: Linking biodiversity, biotic interactions and biogeochemical ecosystem processes’ which is funded by the German research foundation (DFG). Based on the extensive works of A. Hemp (e.g. Hemp, 2005, 2006), the research unit has been established in 2010 to investigate responses of local biodiversity patterns, ecosystem functioning and biogeochemical processes to climate and local land-cover change.

The design is implemented for a comparative analysis of ecosystem functioning across an elevational and disturbance gradient. The elevational gradient encompasses six near-natural habitats starting from the savannah over the lower mountain forest, Ocotea forest, Podocarpus forest and Erica bush/forest zone to the apline Helichrysum region. The disturbance gradient is organized orthogonal to the elevation zones and adds another six habitats comprised by maize fields, coffee plantations, grasslands, the Chagga home gardens which form some kind of sustainable agroforestry system as well as the disturbed (i.e. logged, burned) versions of the Ocotea and Podocarpus forest zones. Within each of these 12 habitats, five replica sites have been selected across the southern slopes of Mt. Kilimanjaro summing up to a total of 60 study plots with an elevation span between 867 m a.s.l (lowest savannah plot) to 4,550 m a.s.l. (highest Helichrysum plot). The individual plot locations have been selected manually based on local expertise of A. Hemp.

Fig. 1 gives an overview over the plot locations which span across 2,182 square kilometres. All of these plots have been equipped in 2011 and 2012 with at least a basic climate station recording air temperature and air humidity and many of the plots host additional sensors for e.g. solar radiation, precipitation, fog or wind speed. As to be expected, not all stations delivered continuous data recordings up until now due to malfunctions or vandalism although most of the plots have been visited on a monthly basis since then. For a handful of plots, we also decided not to reinstall a station because of repeated vandalism or theft at some date between 2011 and today. In addition to these stations, we also utilized a University of Massachusetts station located close to Kibo summit at 5776 m a.s.l.

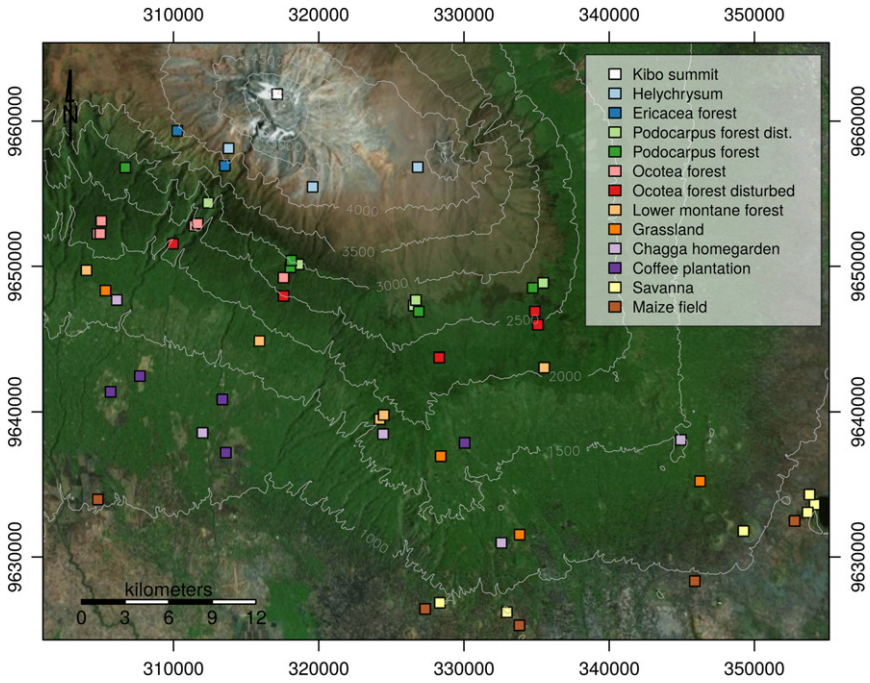

Fig. 1. Overview of the location of the research plots across the southern slopes of Mt. Kilimanjaro (projection for all maps is UTM 37 south, EPSG 32737).

which has been installed by Douglas Hardy. For more information on the research unit, please visit the project's web site at http://www.kilimanjaro.biozentrum.uni-wuerzburg.de/.

3. Data sets and methods

3.1. Spatial interpolation methods

The 14 statistical algorithms used for this study are taken from R's caret package (Kuhn, 2014) and encompass a variety of regression based algorithms. An overview of the individual models is given in Table 1. For details about the individual implementations, please refer to the respective documentations of the R packages mentioned in the right most column. In addition, the kriging implementation of R's automap package (Hiemstra et al., 2008) has been used as baseline spatial interpolation reference.

The statistical and machine learning models of Table 1 can roughly be grouped into (i) linear and spline models, (ii) nonlinear models and (iii) regression trees. A detailed description of the individual models is beyond the scope of this article. Therefore, we refer to e.g. Kuhn and Johnson (2013) from which the following summary is largely taken, too.

Models of the first category generally try to minimize the sum of the squared errors either with a focus on bias or variance. While linear regression models (*glm*, *gam*, *pcr*, *pls*, *svmLinear*, see Table 1 for the abbreviations) can generally only be applied for linear relationships between predictor and response variables, the non-linear regression models do not have this restriction and the form of the relationship does not need to be known a prior.

K-nearest neighbours approaches (*knn*) are one example of the potentially non-linear methods. Their prediction is solely based on the distance of the predictor variables to the closest training group known to the model and an amplitude of different models exists to compute that distance. Neural networks (*nnet*) achieve their predictive power by modelling the target variable using a hidden layer of variables which results from a linear combination of some to all of the predictors. If multiple networks

are averaged (*avNNet*), the approach is comparable to the idea of random forest which also averages over many individual decision trees. Multivariate adaptive regression models (*earth*) are similar to neural networks but they break the feature space into discrete sections and fit a linear model to each of the ranges. Maybe the most robust algorithm of this group with respect to some outliers in the training data is the family of support vector machines (*svmRadial*). This robustness results from using squared or absolute residuals during the model fitting procedure depending on the actual value of the individual residual (i.e. squared if the value is small, absolute if the value is large). Hence, extreme outliers and optimally predicted values have only minor influence on the model fit.

The third group of prediction methods encompasses various kinds of regression trees. They all have in common that they split the (diverse) training dataset into (homogeneous) sub-groups which likely have the same response value. While conditional inference trees (*ctree*) are comprised of only on tree, the other models used within this study combine generally weak individual classifiers to derive a highly skilled model. For stochastic gradient boosting (*gbm*), this combination is realized by adding the individual models which are developed within each iteration. In random forest (*rf*), the combination is based on the mean across a large number of individual trees. Each tree is quite independent from the others since bagging (i.e. the sampling of n test samples for the computation of n models from the entire available data sample) is combined with a randomization of the predictor variables used at each node. In contrast to random forest and the other algorithms described above, cubist (*cubist*) does not retrieve one final model but a set of rules associated with sets of multi-variate models. The actual prediction model for a specific set of predictor variables is selected depending on the rule which best fits the predictors.

Since kriging is a *de facto* standard in air temperature interpolation, R's *automap* implementation of universal kriging is used within this study, too. A distinctive characteristic of the automap package is the self-searching variogram feature which does not require a-priori information.

3.2. Data sets

The spatial interpolation of monthly mean air temperature values has been performed on a regular grid of 30 m by 30 m and is based on predictors derived from

- mean monthly air temperature observations from the individual stations,
- topographic information derived from a digital elevation model with a horizontal resolution of 30 m by 30 m and
- mean monthly values of the normalized difference vegetation index (NDVI) from satellite observations which are downscaled to a resolution of 30 m by 30 m.

The 30 m by 30 m resolution is in the order of the spatial extension of the research plots and can be regarded as just sufficient for the differentiated provision of site-specific temperature information.

The basis for the first group of spatial predictor variables are monthly mean air temperature values from the individual research plots which have been recorded in five minute intervals. These datasets are operationally quality corrected using the julendat software package[1] which will shortly be replaced by a more parallelled and enhanced Java-based implementation.[2]

Monthly air temperature means have only been computed for locations, where the five minute records are available for at least 20 days. From that subset, we included only those months between 2011 and 2014 in the present study, where the mean monthly air temperature could be computed for at least 25 stations. This leaves us with a total dataset of 32 months starting in 02/2012 and ending in 12/2014 (with continuous data availability since 02/2013) for which 25 to 49 stations are available (30 stations or more are available for 27 months and 44 are available on average). For these months, the dataset was extended by the temperature records from Kibo summit (see Section 2). Since these records also show gaps, the respective time series has been gap-filled by using R's *forecast* package by Hyndman (2015).

[1] http://github.com/environmentalinformatics-marburg/julendat.

[2] See software on www.environmentalinformatics-marburg.de for an update.

Table 1
Statistical algorithms included in R's caret package and the kriging library used for this study. Grouping partly follows Kuhn and Johnson (2013).

Abbr.	Model	R package
Linear models		
glm	Generalized Linear Model	base
gam	Generalized Additive Model using Splines	mgcv
pcr	Principal Component Analysis	pls
pls	Partial Least Squares	pls
svmLinear	Support Vector Machines with Linear Kernel	kernlab
Nonlinear models		
avNNet	Model Averaged Neural Network	nnet
earth	Multivariate Adaptive Regression Spline	earth
knn	k-Nearest Neighbours	base
nnet	Neural Network	nnet
svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab
Regression trees		
cubist	Cubist	Cubist
ctree	Conditional Inference Tree	party
gbm	Stochastic Gradient Boosting	gbm, plyr
rf	Random Forest	randomForest
Geostatistical model		
kriging	Universal kriging	automap

Except for kriging, the spatial prediction methods used within this study do not get the mean monthly air temperature observations from the individual stations directly. Instead, a single regionally averaged mean temperature value for each month is used as one of the predictors. Since the number of available stations differ per month, we do not use the arithmetic mean of the monthly mean values but take the intercept of a linear regression function fitted to the observed temperature observations at the individual available stations and their respective height a.s.l. Hence, the averaged temperature is the temperature one would expect at sea level given an averaged vertical temperature gradient.

Appelhans et al. (submitted for publication) identified a condensation level with quite distinct temperature gradients below and above about 2,300 m a.s.l.. Therefore, two more temperature-based predictors have been computed which are the regionally averaged monthly air temperatures for the upper and lower plot locations. The computation of these regionally averaged values have been computed analogous to the regional mean value except that for the upper locations, only stations above 2,300 m a.s.l. and for the lower locations, only stations below this level are considered.

Other studies have additionally made use of remotely sensed land surface temperatures (LST) from platforms such as MODIS (e.g. Kilibarda et al., 2014), however, given extremely high cloud contaminations at Mt. Kilimnajaro, this is impossible at monthly scales.

The second group of predictors encompasses information on elevation, hill slope, aspect, and the sky-view factor which are computed for each of the grid cells of the digital elevation model (DEM). The DEM is derived from digitized topographical maps by J. A. Onginjo, C. Lambrecht and A. Hemp. While the elevation can directly be read from the DEM, the other predictor variables have been computed using the respective modules in SAGA GIS (http://www.saga-gis.org/).

For the last predictor, maximum composites of the satellite-derived Normalized Difference Vegetation Index (NDVI) are computed for each month. The NDVI is surely the most utilized indicator for vegetation greenness and phenological status and is computed from the difference between the near infra-red and red reflectance divided by their sum. It ranges between -1 and 1 with values above about 0.2 indicating green vegetation. Larger values indicate denser vegetation which in turn has a direct influence on boundary layer dynamics which affect the local temperature. Non-vegetated surfaces show lower values around 0 and water surfaces are negative. The NDVI is preferred in this study over a land-cover classification since the static nature of the latter does not account for different phrenological phases, stand densities etc. The monthly mean NDVI is bi-linearly interpolated from 250 m by 250 m to 30 m by 30 m from a combined GIMMS (Tucker et al., 2005) and Aqua-MODIS time series which has been computed following Appelhans et al. (2015).

Table 2
Overview of the predictors used for the spatial prediction of monthly air temperature.

Abbr.	Predictor	Models
Ta	Monthly mean temperature at the stations	Kriging only
Ta_reg	Regional averaged monthly mean temperature	All but kriging
Ta_reg_upr	As above but above 2,300 m a.s.l.	All but kriging
Ta_reg_lwr	As above but below 2,300 m a.s.l.	All but kriging
dem	DEM-based elevation	All
slp	DEM-based hill slope	All
asp	DEM-based hill aspect	All
svf	DEM-based sky-view factor	All
ndvi	Monthly means of NDVI	All
ndvi_tr	Transformed monthly means of NDVI	All

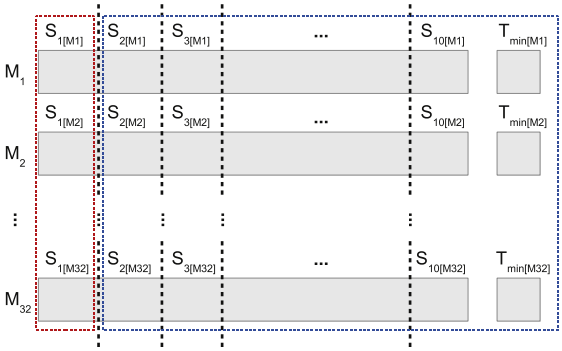

Fig. 2. Schematic overview of the dataset structure used to train and test the individual spatial prediction models.

In the end, this results in a dataset which links each available monthly air temperature observation at a given station (i.e. the response variable) with (i) the regional temperatures computed based on all available stations within that month, (ii) the DEM-based information for the location of this station and (iii) the NDVI value of the respective month at this location (see Table 2). Prior to be used in the different spatial interpolation methods, all predictor variables are centred and scaled by their standard deviations. For linear models, the NDVI is additionally transformed by $sin(\pi\, NDVI)$ to account for the non-linear relationship with height with a maximum in the forest belt between 1,800 m and 3,000 m a.s.l.

3.3. Experiment setting

Given the available dataset of an average of 44 stations over 32 months, we did not select a fixed portion of the data as independent test set but estimated the final performance of the individual spatial prediction methods using a stratified modification of a 10-fold cross validation. The resulting dataset structure is illustrated in Fig. 2. The rows M1 to M32 indicate the individual 32 months for which sufficient data is available. Each horizontal combination of a long and a short bar represents the total available dataset for the respective month as summarized at the end of the preceding chapter.

The stratified modification of the 10-fold sampling procedure first selects one station per available habitat randomly for each month and groups the results into a minimum training set $T_{min[Mm]}$ ($m = [1 \ldots 32]$). This is indicated by the small block at the right end of each row which contains 12 data pairs on average. After the separation of this sample, each of the m monthly remaining datasets is randomly divided into 10 samples $S_{s[Mm]}$ ($s = [1 \ldots 10]$) indicated by the dashed vertical lines which contain another 3 data pairs per sample on average.

The dashed red and blue lines additionally illustrate the division into testing (red) and training (blue) dataset during the first of the ten final performance estimates. For all methods except kriging, the first training dataset encompasses the samples $S_{2[Mm]}$ to $S_{10[Mm]}$ and $T_{min[Mm]}$ for all months $m = 1$

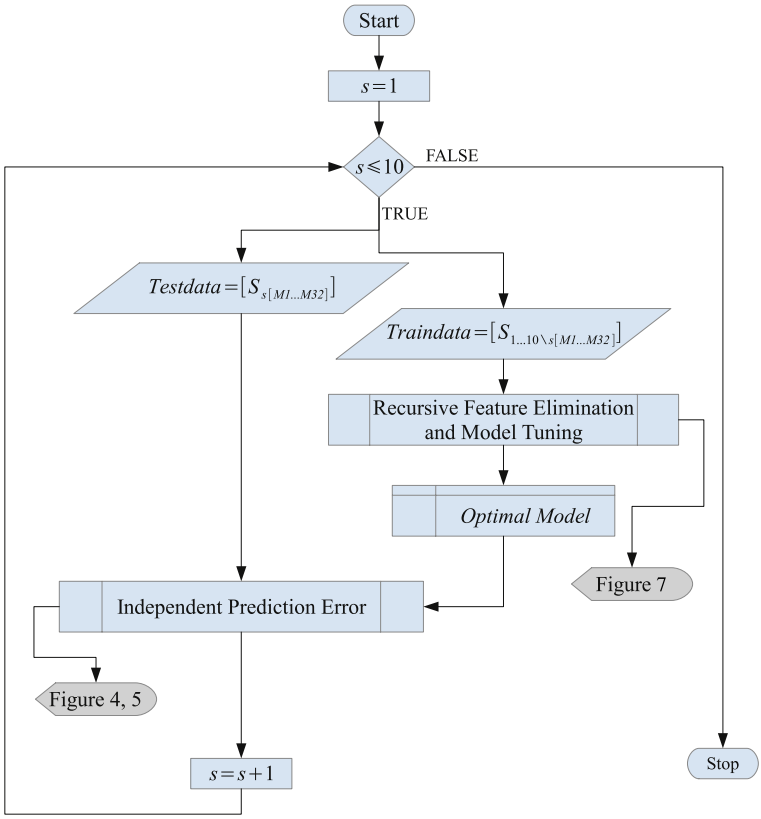

Fig. 3. Schematic overview of the recursive feature elimination and model tuning procedure.

to $m = 32$ and the corresponding testing dataset is comprised by $S_{1[Mm]}$ again for all months m. The test error for the first of the ten folds is then the average of all errors computed for each individual response value in $S_{1[M1]}$ to $S_{1[M32]}$. In terms of point pairs, the training data sample contains 1045 and the test data sample 105 observations on average.

For kriging, the procedure is slightly different because the model training and the test error has to be computed within the individual month. Therefore, the samples $S_{2[M1]}$ to $S_{10[M1]}$ and $T_{min[M1]}$ are used to build a kriging model which performance is tested using $S_{1[M1]}$ and so forth. As for the other models, the test error for the first fold is the average over all errors computed for $S_{1[M1]}$ to $S_{1[M32]}$.

Switching from a dataset to a data flow perspective, this procedure is illustrated in the left hand path of Fig. 3. The resulting performance estimates for each of the ten folds are visualized in Fig. 4 in Section 4.

The right hand side shows the model training procedure performed within each of these ten "outer" cross-validation loops. The training procedure is based on the *rfe* function of the caret package and consists of a combined recursive variable elimination and model tuning routine. The recursive feature elimination starts with all available predictors listed in Table 2. Based on another 10-fold cross validation computed on the training dataset of the respective "outer" loop (not shown in the figure), the model performance is estimated and predictors are ranked according to their importance determined by their influence on the related root mean square error. The least important predictor is eliminated from the dataset and the elimination procedure starts again until only one predictor is left. Within each of these elimination loops, model parameters are tuned within another 10-fold cross validation (also not shown in the figure). Hence, for each of the seven recursive feature selection runs which are required for the eight predictor variables used in this study, two nested 10-fold cross

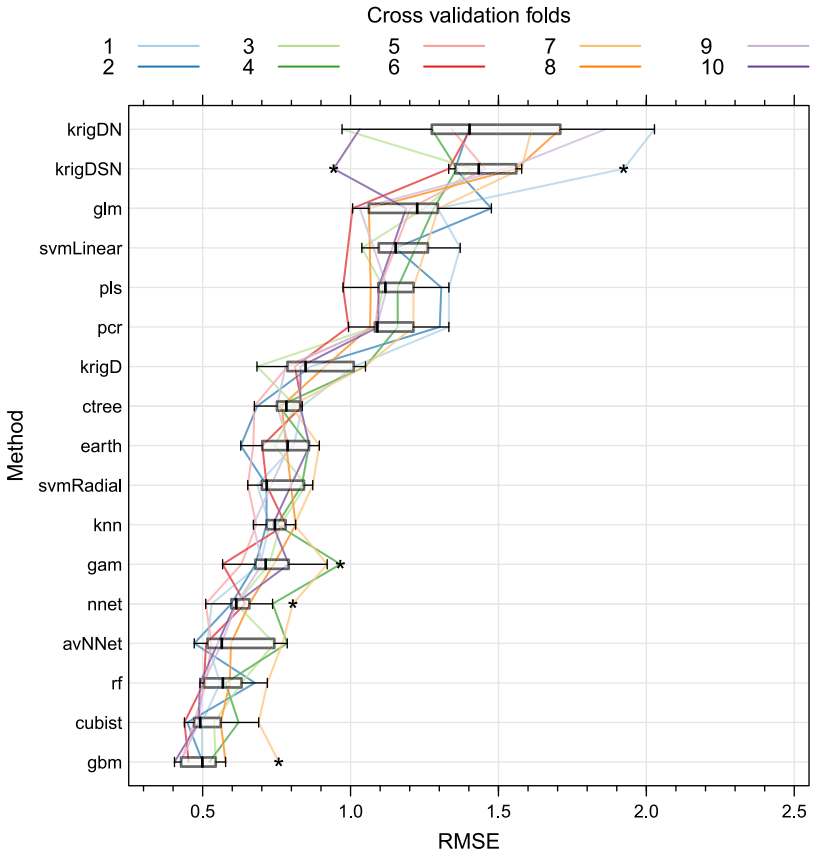

Fig. 4. Summary of the results in terms of the root mean square error (RMSE) from the 10-fold cross validation for each of the 17 spatial interpolation methods. The methods are sorted by their predictive skill from top (fair) to bottom (good). For model abbreviations please refer to Table 1. Kriging with elevation only is labelled krigD, kriging with elevation and NDVI is labelled krigDN and the one additionally including the sky view factor is labelled krigDSN.

validation runs are computed. Table 3 gives an overview of the model parameters modified during the tuning step.

3.4. Variable selection, model tuning and performance estimation using cross-validation

By using the feature elimination procedure, one can easily identify the model that shows the smallest mean cross-validation test error. This model and the corresponding predictor set could be regarded as optimal but we follow a more conservative selection approach by also considering the error range of this model which is defined by the standard deviation of its cross-validation errors. Our optimum model and corresponding predictor set is subsequently selected by identifying the model with a mean test error that is just below the upper limit of this error range but uses less predictors. This model is finally used as optimal model for the estimation of the independent prediction error within the actual "outer" cross validation loop by applying the model to the actual test sample (i.e. 105 observations on average). The results are visualized in Fig. 4 and will be presented in the next chapter.

While the overall structure of Fig. 3 remains valid for kriging, too, the feature selection approach is different mainly due to computing time considerations. Feature selection is performed on a set of five randomly selected months. For each of these months, all available observations (i.e. the long and short grey bar in 2) are used to compute the model on the 30 m by 30 m grid. The minimum and maximum

Table 3
Overview of the parameters considered during the model tuning. Notations of type $x \ldots y, z$ give the range (x, y) and step size (z).

Model	Tuning
glm	No special tuning
gam	Smoothing: GCV.Cp, GACV.Cp, REML, P-REML, ML, P-ML
	Penalization of a parameter to zero: do/do not allow
pcr	Max. 3 components, no special tuning
pls	Max. 3 components, no special tuning
svmLinear	Cost of constraints violation: 0.25…1, 0.25
avNNet	Units in hidden layer: 10…20, 5
	Weight decay: 0, 0.001, 0.1
	Bag for each repeat
earth	Terms in pruned model: 3…18, 3
	Interaction degree: 1…3, 0.5
knn	Neighbours considered: 3…8, 1
nnet	Units in hidden layer: 10…20, 5
	Weight decay: 0, 0.001, 0.1
svmRadial	Scale of laplace distribution: 0.25…1, 0.25
	Cost of constraints violation: 0.25…1.25, 0.25
cubist	Number of committee models: 20…80, 20
	Neighbours considered: 3…9, 3
ctree	Split criterion: 0.99, 0.50, 0.01
gbm	Variable interaction depth: 1…7, 2
	Number of trees: 100…1000, 100
	Shrinkage of tree expansion: 0.01, 0.1
rf	Sampled predictors at node: 2…8, 1
kriging	No explicit tuning

value of the resulting temperature map is then compared to the minimum and maximum value of the mean monthly temperatures recorded within this month and used as proxy for the error of the kriging model. This selection procedure is repeated for various combinations of external drift variable sets (e.g. elevation only, elevation and NDVI, elevation and slope etc.). For the final performance estimation, we finally selected the kriging model with the smallest error which turns out to be the one just using elevation as external drift and the two model configurations which include the NDVI as external drift and have the smallest error of this group (i.e. elevation, NDVI and sky-view factor as well as elevation, NDVI and aspect). The latter where included to double check if additional information to the elevation actually has a negative influence in the much more comprehensive final performance test.

4. Results and discussion

4.1. Quantitative evaluation of the spatial prediction methods using 10-fold cross validation

The results of the model performance evaluation are given in Fig. 4. Each row shows the results for one method. Each vertically spanning coloured line is the result of one of the 10 cross-validation folds with the same training and testing dataset used and applied to each of the models as described in the previous section.

Measuring the prediction accuracy by the root mean square error, the geostatistical kriging approach using only the elevation as external drift (*krigD*) performs significantly better than the other versions included in this final test. As mentioned earlier, this has already been expected from the variable selection procedure. It has a median error of just above 0.85 K with 50% of the errors ranging between about 0.8 and 1.0 K.

Compared to the kriging results, all linear methods except *gam* perform worse while all non-linear and tree based models perform better. Starting at least with neural networks (*nnet*), the largest RMSE values of the best performing models lie in the range of the lowest RMSE values of the *krigD* or even outperform these results entirely. Among these, averaging methods, especially when combined with weak learners seem to perform better. This is in accordance with many studies that generally propose

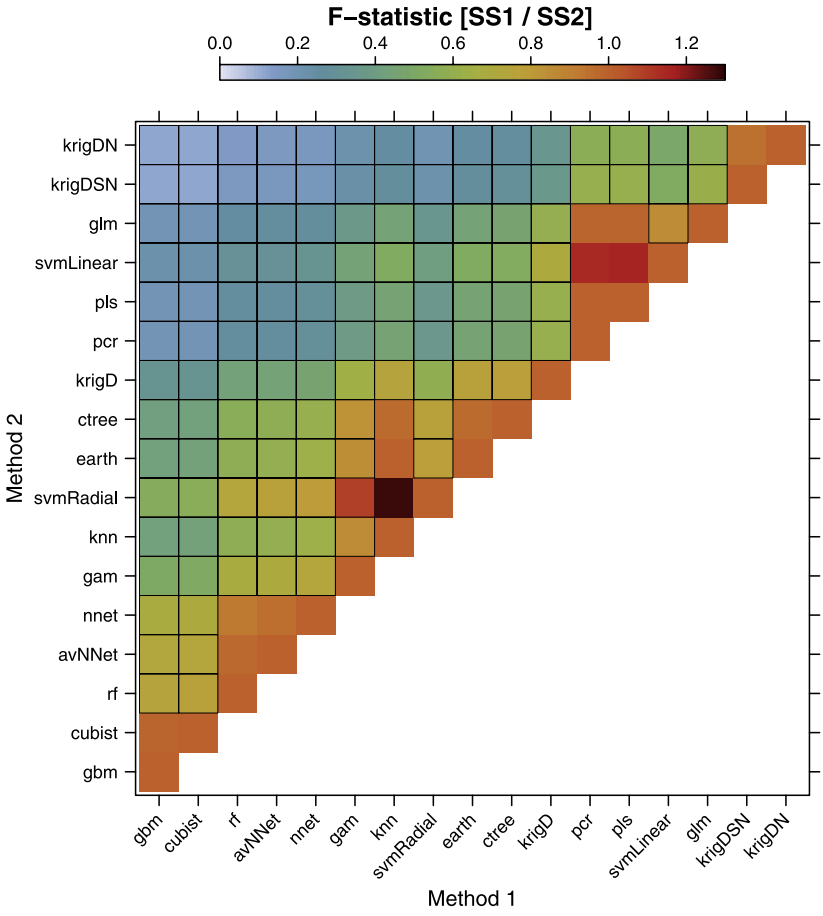

Fig. 5. F-statistic comparison of all 17 spatial interpolation methods. Colours denote F-statistic (sum of squares of method 1/sum of squares of method 2), black squares indicate statistically significant differences in F-statistic. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

variations of regression tree ensembles for the interpolation of environmental data sets (see overview in Section 1).

Regarding statistically significant differences between the models, Fig. 5 provides an overview of the F-statistic between measured and modelled values for each model combination. The F-statistic is expressed as the ratio of the sum of squares of method 1 (x-axis) and method 2 (y-axis). The black squares denote statistical significance, i.e., method 1 performs significantly better than method 2 if surrounded by a black square. It is obvious that there is no significant difference between the top two models *gbm* and *cubist*. These are followed by a block of *rf*, *AvNNet* and *nnet* which all perform equally well. Another large block is denoted by the strictly linear models *pcr*, *pls*, *svmLinear* and *glm* for which no differences can be found.

A look on the performance of the spatial interpolation methods for the individual habitats reveals that the error of the five best performing models is largest on grassland patches while especially the linear models and the kriging variants have quite large problems in predicting the air temperature in the alpine zone (i.e. the helichrysum areas). The latter is likely a consequence of both the very limited amount of data from this area (only two measurements are available to compute the RMSE in this habitat) and a highly non-linear transition of the temperature gradients from the upper forest boundary to the helichrysum sites combined with height levels ($>$4,000 m a.s.l) which are more and

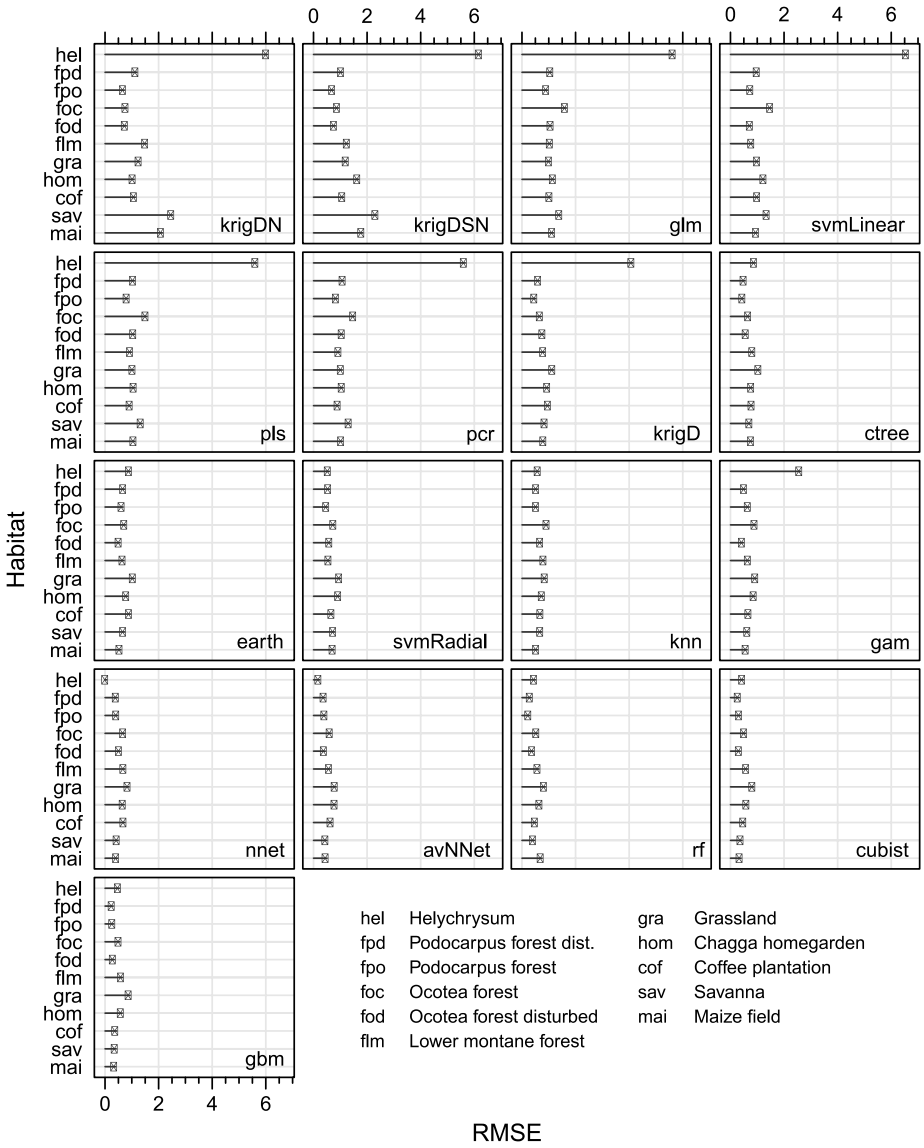

Fig. 6. Mean of the root mean square error (RMSE) from the 10-fold cross validation for each of the 15 spatial interpolation methods as a function of habitat type. The methods are sorted by their predictive skill from top (good) to bottom (fair) based on the performance visualized in Fig. 4. For model abbreviations please refer to Table 1. Kriging with elevation only is labelled krigD, kriging with elevation and NDVI is labelled krigDN and the one additionally including the sky view factor is labelled krigDSN. Please note that results for the helichrysum zone are based on only two measurement locations per month on average.

more influenced by advective air-masses in mid-tropospheric wind fields than local boundary layer conditions (see Fig. 6).

In general, the exact reason for the performance of most of the methods included in this study is hidden to a large extend by their machine learning paradigm. Looking into the predictor importance of the individual methods however reveals some insights. Fig. 7 shows the overall predictor importance for the individual optimal models derived in each of the ten cross-validation runs. Except for the support vector machines and neural networks, the elevation (*dem*) is the most important variable of

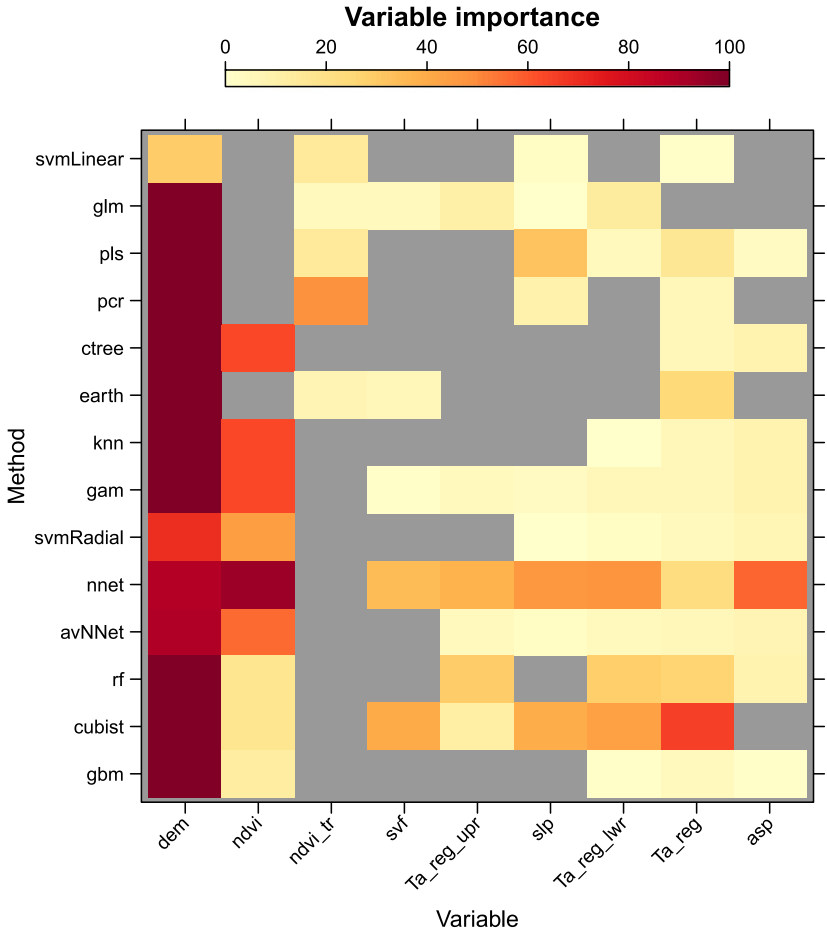

Fig. 7. Importance of individual predictor variables for each method. Grey colours indicate that the predictor has not been used in any of the ten individual optimal models derived in each of the ten "outer" cross-validation runs. Colours give the variable importance in the individual cross-validation runs multiplied by the number of cross-validation runs it has been used within and normalized by ten. As an example, elevation (*dem*) has a value of 100 for most of the methods which indicates that elevation has also been the most important variable within these models (scaled value of 100) and that it has been used within each of the ten cross-validation runs ($100 \cdot 10/10 = 100$). The predictor abbreviations are given in Table 2 and the methods are sorted according to their quantitative performance from bottom to top (see Fig. 4). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

each optimal model from each method and for each of the cross-validation runs. NDVI values, either untransformed (*ndvi*) for non-linear or transformed *(ndiv_tr)* for linear models are also important and used by every method although with quite large variations in the individual importance. All other variables are only used by a subset of methods and with very different importance values. However, at least one topography related parameter (i.e. *svf* , *slp* or *asp*) is always included. The same is true for the mean regional temperature based predictors where at least one of Ta_{reg}, *Ta_reg_upr* or *Ta_reg_lwr* is used by every method and at least two are used except by *earth*, *ctree*, *pcr* and *svmLinear*. This is meaningful from a meteorological point of view since

- air temperature changes over the course of the year independently of any other local factor which is accounted for by the regionally averaged, individual monthly air temperatures,
- air temperature is a non-linear function of elevation,

- air temperature is a function of the land cover status which directly influences the boundary layer conditions and is mapped by the NDVI and
- air temperature is a function of the radiative transfer geometry which in turn is highly influenced by the morphology (i.e. the slope, aspect or sky view factor).

In summary, most of the errors are in a range of 0.4–0.7 K which is comparable to expected sensor accuracies of about 0.5 K which result from instrumental and systematic errors. In this light, these result are very encouraging although the true error range might be a little larger since cross-validation and/or random sub-sampling approaches tend to slightly overestimate model performance. Aside from that, the quantitative error analysis at individual locations is only one aspect of the performance estimation since it does not account for spatial interpolation artefacts or unreasonable spatial shifts off the validation locations. Therefore, a qualitative evaluation follows in the next chapter.

4.2. Qualitative evaluation of the spatial interpolation methods using visual assessment

Collins and Bolstad (1996) already stated that kriging gives the best visually verified results although the quantitative performance indicators might not be the best within a group of evaluated models. The obvious reason for such a contradiction lies in the fact that quantitative estimates based on cross-validation approaches do only consider the error at the available locations while the overall interpolation pattern can be highly biased in between these locations (e.g. initial temperature decrease with height with an exaggerated increase in the vicinity of the next known station location). In order to get a qualitative estimate of the overall interpolation quality, we computed monthly mean air temperature maps for each month in 2014 using the best four methods identified in the previous section (i.e. *gbm*, *cubist*, *rf* and *avNNet*) and the two kriging versions using only elevation as external drift (*krigD*) and using elevation, NDVI and the sky view factor (*krigDSN*).

Since for each method not one but ten individual optimal models have resulted from the cross-validation approach, we trained each method described in this chapter again using the entire available dataset (i.e. no test sample has been excluded as it was the case before). Hence, the models resulting from this training procedure and as a consequence also the related spatial predictions are slightly different to any of the ten corresponding optimal models of the cross-validation study though the error estimates can still be regarded as of the same magnitude.

For the kriging models, this approach implies that each of the 12 months in 2014 has been used individually to retrieve the corresponding 12 monthly mean air temperature maps. The corresponding variograms of the final *kriD* models are shown in Fig. 8. Especially the main rainy season (February–June) exhibits comparably low *range* and *sill* values indicating that spatial temperature variability is generally reduced during these months.

For the four machine learning methods, the training dataset consisted of all available data of the 32 months and the final model was then applied to the predictor values valid for the individual months in 2014. As predictors, only those variables have been used, which have also been used by the optimal models resulting from the cross validation runs described in the previous section and visualized in Fig. 7. Model tuning has been performed as described in Section 3.3.

In order to reduce the number of figures and for an easier visual assessment, we do not present the results of the individual months but aggregate the monthly air temperature maps to a mean annual air temperature map for 2014. The reader is referred to the supplementary kml dataset for the individual monthly temperature maps.

Fig. 9 shows the resulting interpolation patterns from the individual models in terms of the mean annual air temperatures for 2014. All models except *gbm* show a generally decreasing temperature pattern with elevation and the kriging models predict lower temperatures at the summit region. The latter is an intrinsic consequence of the averaging nature of the depicted models which prohibits a prediction of the actual observed minimum and maximum values.

The temperature decrease with elevation is quite disturbed in the *gbm* model which shows a warm elevation belt in the homegardens region (1,000 to 1,500 m a.s.l.) and also the most pronounced temperature variations inside the forest belt (1,800 to 3,000 m a.s.l.). This likely indicates a strong sensitivity of the model regarding the NDVI dataset which is also evident in the kidney-shaped forest

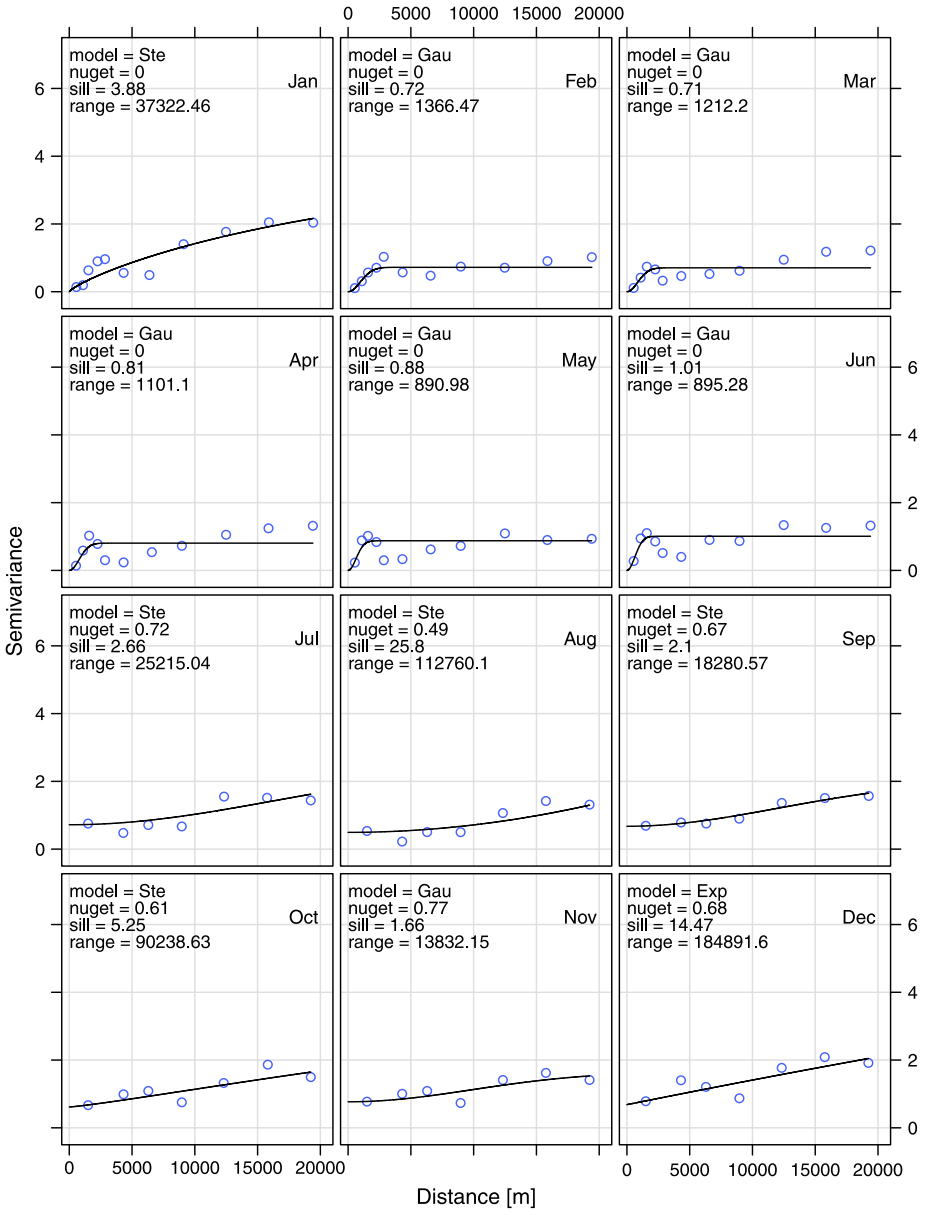

Fig. 8. Variograms for the twelve final monthly *krigD* models for 2014. Values for *nugget*, *range* and *sill* along with the variogram model that was used are given in the top left corner of each panel. Months are shown in the top right corner of each panel. Model abbreviations are: *"Ste": Matern, M. Stein's parameterization*, *Gau: Gaussian*, *Exp: Exponential*.

patch located below 1,000 m a.s.l. at the southern border of the map (close to coordinates 9,630,000 North and 320,000 East, forest area is called RAU). Except of course for *krigD*, the latter pattern is also evident in the other models but with different signs and amplitudes. From this group, *avNNet* is the model which seems to decrease air temperature as a function of denser vegetation much more than *cubist* or *rf*. This is not only evident in the RAU forest patch or along the lower boundary of the forest belt which runs just below the 2,000 m contour line but also in the huge forest clearings at the

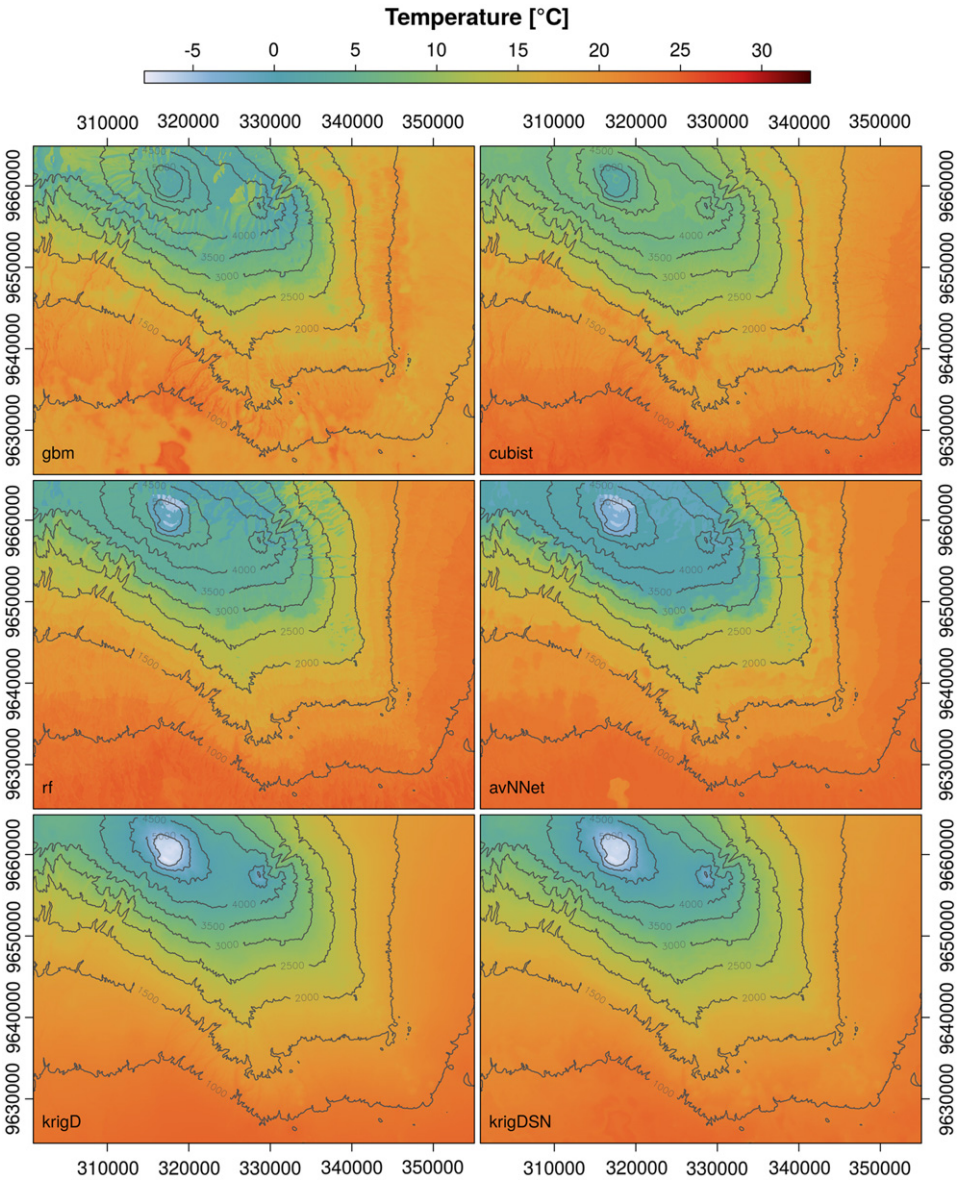

Fig. 9. Mean annual temperature maps for 2014 aggregated from monthly air temperature maps computed by (a) stochastic gradient boosting (*gbm*), (b) cubist (*cubist*), (c) random forest (*rf*), (d) model averaged neural networks (*avNNet*), (e) universal kriging with elevation as external drift (*krigD*) and (f) universal kriging with elevation, NDVI and sky view factor as external drift (*krigDSN*).

northern edge of the study area (just above the 2,000 m contour line between 330,000 and 340,000 East) which are considerably warmer.

Regarding the emphasis of the models on topographic features, it seems that it decreases from *gbm* over *cubist* and *rf* to *avNNet* which is apparent from the valley structures. The consideration of aspect as predictor variable by *gbm*, *rf* and *avNNET* is evident especially at the north facing slopes of the summit region which are not different to the south facing ones in the other three models.

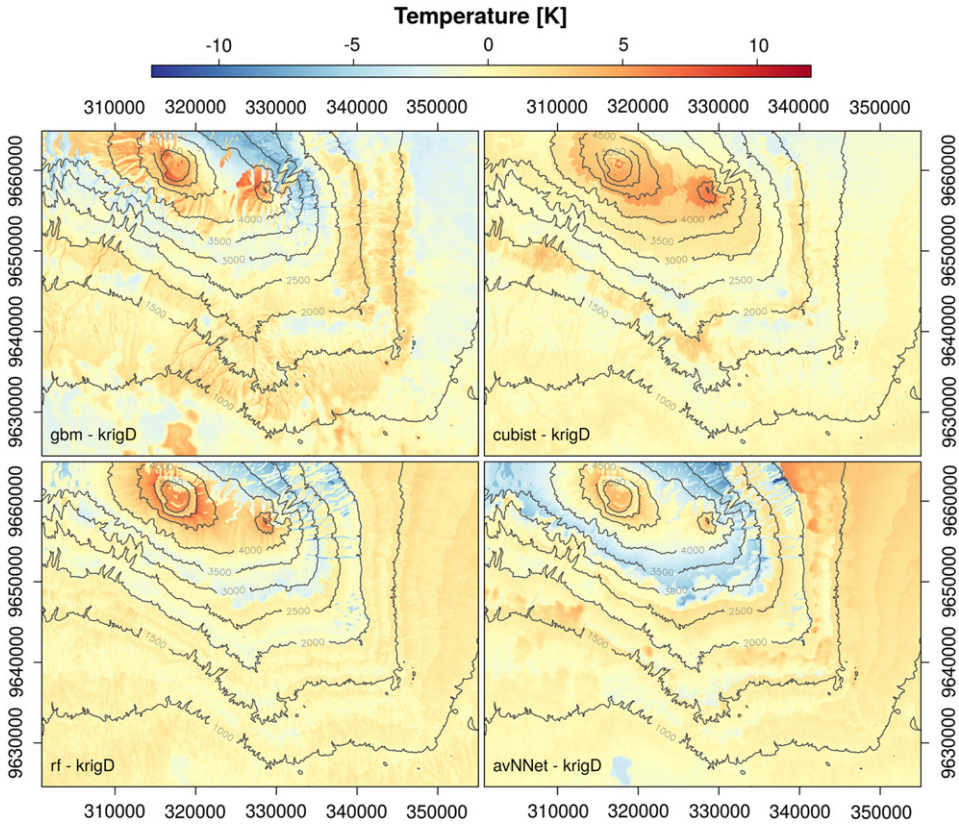

Fig. 10. Differences in the mean annual temperature maps for 2014 from Fig. 9 between (a) stochastic gradient boosting (*gbm*), (b) cubist (*cubist*), (c) random forest (*rf*), (d) model averaged neural networks (*avNNet*) and the universal kriging with elevation as external drift (*krigD*).

To visually enhance the differences between the spatial interpolation models, we computed the mean annual temperature difference based on the maps in Fig. 9 between each of the four machine learning models and kriging with elevation as external drift. In general, the resulting patterns shown in Fig. 10 confirm the findings above. Especially the cooling of north-facing slopes and the warmer summit regions are directly evident.

Two additional features are worth noting. The first one is a rather regular pattern of alternating warmer and colder belts in the *rf* and *avNNet* model results which cannot be explained by atmospheric physics and reduces the applicability of the results for other studies to a certain degree. The second one is a zonal temperature increase along the 2,000 m contour line in the *cubist* results which follows the lower border of the mountain forest zone. We will discuss potential reasons for this later in connection with Fig. 13.

From a thermodynamic view, the internal structure of the atmospheric boundary layer is influenced by the underlying land cover characteristics which should be noticeable and clearly visible in interpolated temperature maps. In light of the visual characteristics of the different approaches summarized above, the magnitude of these vegetation related influences is, however, questionable at least in some areas. Therefore, we adopted the approach taken by Sanabria et al. (2013) where the residuals from machine learning predictions made at the observation points are spatially interpolated using ordinary kriging. The residual interpolation is shown in Fig. 11. It indicates that the forest belt is generally over-predicted, i.e., too warm, while neighbouring regions above and below are under-predicted by *cubist* and to a much stronger extend by *gbm*. For the other two models, *rf* seems to be

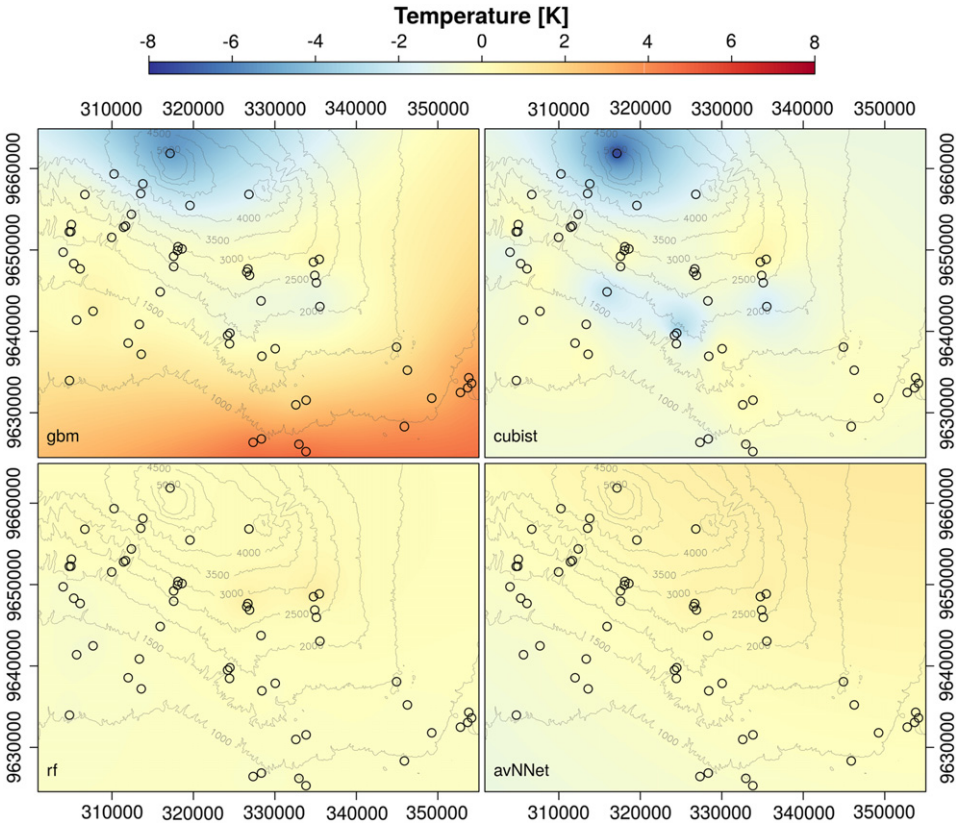

Fig. 11. Map of the interpolated residuals of (a) stochastic gradient boosting (*gbm*), (b) cubist (*cubist*), (c) random forest (*rf*) and (d) model averaged neural networks (*avNNet*). The residuals have been computed at each of the available observations locations and interpolated using kriging on a monthly basis. The map shows the average for the year 2014 over the twelve monthly interpolations for each model. Black circles indicate all of the potentially available observation sites (i.e. the station network already shown in 1).

fitted quite well to the observations with just some under-predictions in the upper forest belt region between the 2,500 m and 3,000 m contour line. The *avNNet* model on the other hand clearly depicts a systematic drift from south-west (over-prediction) to north-east (under-prediction).

The "adjustment" maps of Fig. 11 are simply added to the temperature maps of Fig. 9 in the following which results in the residual adapted interpolations of the four machine learning models that are shown in Fig. 12. As one can expect from the residual maps, the resulting temperature patterns for the adjusted random forest model *rfOK* and the average neural network model *avNNetOK* do not change much and the temperature decrease with elevation which has a rather band-type pattern that leads to the alternating warmer and colder belts in Fig. 10 is still notable. Regarding the adjustment of the stochastic gradient boosting model, *gbmOK* now also shows the general temperature decrease with elevation also the RAU forest area is still notable as a much too warm artefact. From our point of view, the cubist model (*cubistOK*) is the one which gains most additional quality from the residual kriging approach. It already made a good visual impression before and both the summit region as well as the region along lower forest border which both used to be too warm have been corrected although the latter is still slightly evident.

Fig. 13 provides a more detailed look on the kriging-adjusted cubist result. It shows scaled profiles of *cubist*, *krigD* and *cubistOK* that are plotted for cross-sections along three major climbing routes (Mweka, Maua and Marangu). Temperature observations of research sites in the close vicinity of these

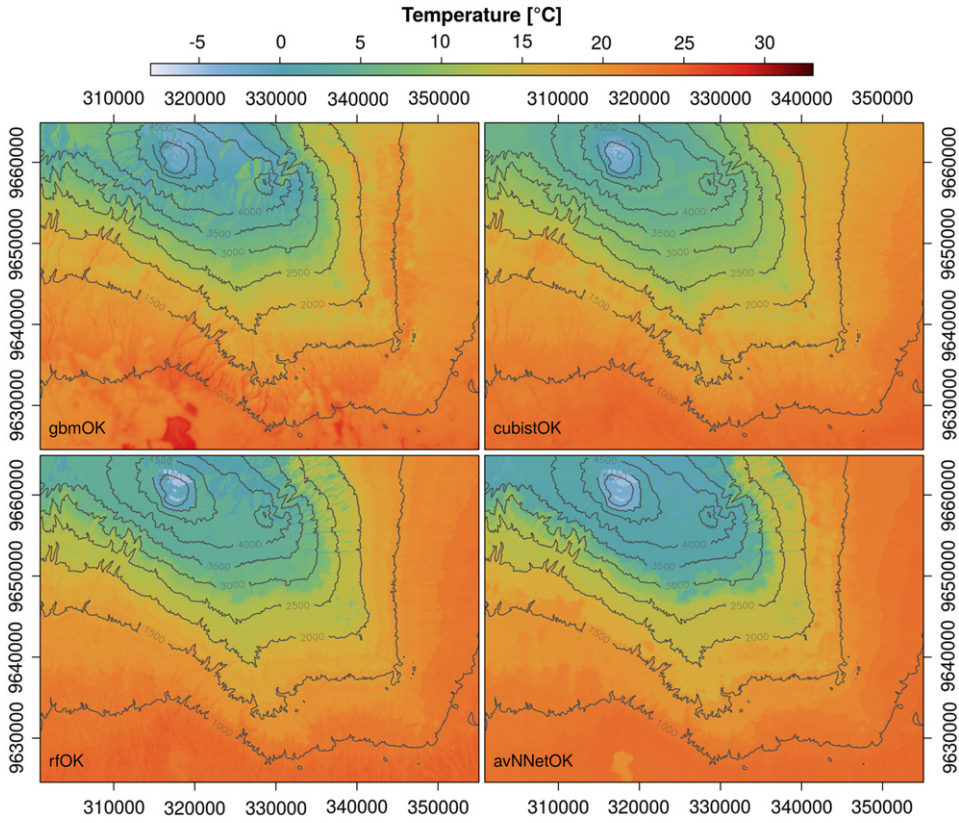

Fig. 12. Mean annual temperature maps for 2014 as in Fig. 9 but corrected by the interpolated residual maps using ordinary kriging from Fig. 11 for (a) stochastic gradient boosting (*gbm*), (b) cubist (*cubist*), (c) random forest (*rf*) and (d) model averaged neural networks (*avNNet*).

trails are also shown (black dots). Additionally, the standardized profile of the NDVI (averaged over the 12 monthly NDVI datasets) as well as the elevation profile along these cross-sections is provided.

General characteristics of the cubist profiles are such that (i) temperature increases in response to the NDVI increase at the lower montane forest border (at elevations between 1,600 m a.s.l. and 1,800 m a.s.l.), (ii) temperatures also increase at the upper forest border at approximately 3,000 m a.s.l. and (iii) flattening of the terrain also seems to cause temperatures to increase. It may seem contradictory that temperatures increase in response to both increases as well as decreases in NDVI (lower/upper forest border, respectively) and we cannot rule out that it just happens by chance. However, this may well be a realistic scenario resulting from the fact that the 2 m air temperature in the forest belt represents within-stand conditions, whereas outside the forest it is more representative of the ambient atmosphere.

Regarding the kriging based adjustment, it becomes evident that the major temperature jumps along the profiles that are apparent in the *cubist* curve are generally decreased. The relative variations within the curve basically remain unchanged. This is generally good news, as it means that we do not loose the vegetation related patterns, they are simply adjusted to better fit the available observations.

5. Conclusions

In this study we wanted to estimate the performance of a variety of (machine learning) algorithms for the spatial prediction of air temperature values on a monthly basis. Therefore we have computed

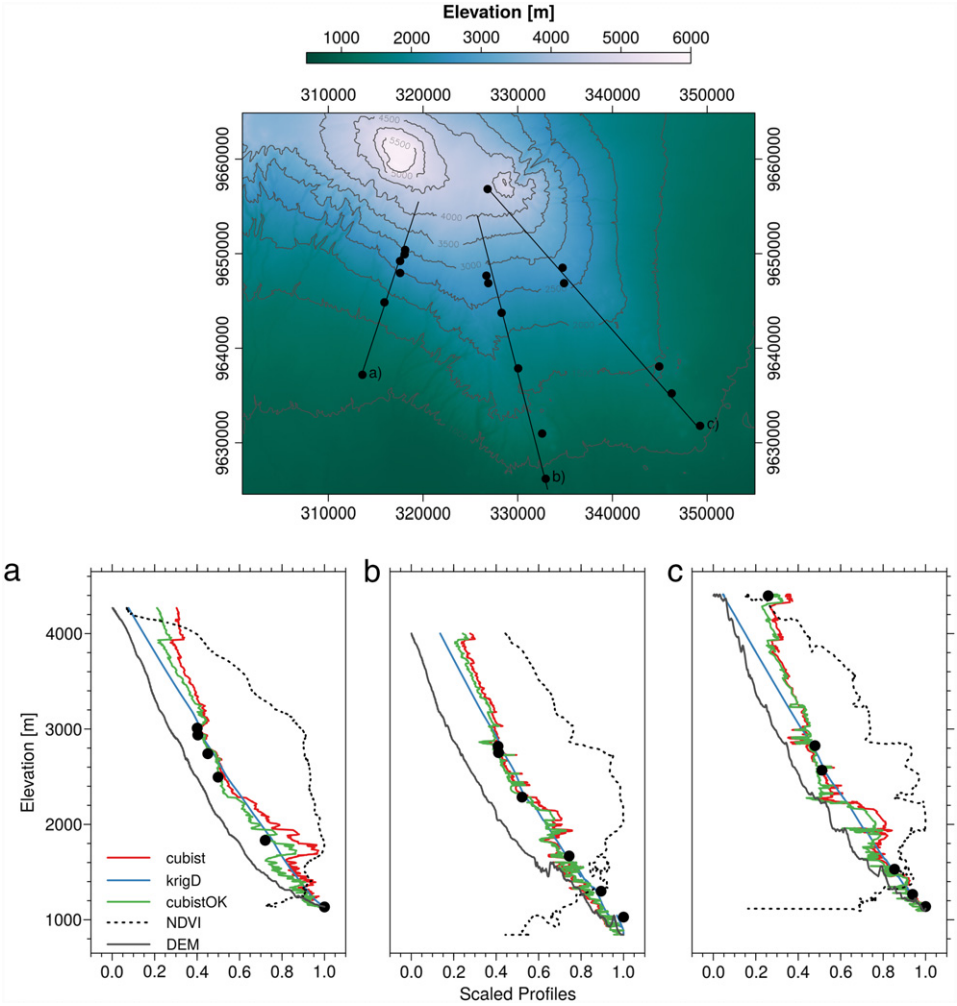

Fig. 13. Profiles of the predicted mean annual air temperature for 2014 along (a) Mweka, (b) Maua and (c) Marangu route based on cubist (red) and kriging with elevation as external drift (blue) which are both shown in Fig. 9, and the combined cubist and residual kriging approach (green) from Fig. 12. Profiles of the NDVI (dotted) and the elevation (black) are taken from the respective raster datasets. For the NDVI profile, the monthly datasets have been averaged. Black points represent the location of temperature observations in the vicinity of the transects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a 10-fold cross validation study based on 32 months of observational data with 46 observations pairs on average for each month. Prior to the performance estimate, a recursive variable selection and model tuning has been computed for each of the 10 cross-validation runs using another nested cross-validation approach.

In light of both the quantitative 10-fold cross-validation statistics form 4.1 as well as general characteristics derived through the qualitative visual inspection in Section 4.2, we are confident that cubist is able to produce more reliable spatial estimates than the other methods, including universal kriging. However, it needs to be mentioned that for optimizing the kriging interpolation we have focused on variable selection and not on other particular tuning options as we have simply used the self-optimizing algorithm from the R automap package for this. Hence, there should be scope for some

improvement regarding the kriging approach although the results indicate that this improvement must be quite significant to outperform the best machine learning methods.

Especially, the post-processing of the cubist-based temperature maps using the residuals interpolated with ordinary kriging is promising, as the land-cover related patterns remain nearly unchanged while the overall estimates are adjusted towards the observations. On the one hand this means that the cubist interpolation patterns are brought closer to those obtained by kriging which is desirable especially in locations where cubist might put too little emphasis on the influence of elevation or where the averaging nature of the method prevents it from reaching the observed extremes. On the other hand the ordinary kriging of the residuals produces rather smooth adjustments which ensures that the identified land-cover induced patterns are not obliterated.

Acknowledgements

This work was carried out in the frame work of the research unit 1246 "Kilimanjaro ecosystems under global change: Linking biodiversity, biotic interactions and bio-geochemical ecosystem processes" funded by the German Research Foundation (DFG, funding id Ap 243/1-2, Na 783/5-1, Na 783/5-2). We would also like to thank Jimmy Ndyamkama and the rest of our field assistant team who are constantly looking after the installations and facilitate data collection in the field. D.R. Hardy acknowledges support from the US National Science Foundation and NOAA, through NSF grants 0402557 and 9909201 to the University of Massachusetts. Last but not least we would like to thank the two reviewers of this paper for their valuable and detailed comments and suggestions.

References

Abdel-Aal, R., 2004. Hourly temperature forecasting using abductive networks. Eng. Appl. Artif. Intell. 17 (5), 543–556.
Apaydin, H., Sonmez, F., Yildirim, Y., 1911. Spatial interpolation techniques for climate data in the GAP region in Turkey. Clim. Res. 28, 31–40.
Appelhans, T., Detsch, F., Nauss, T., 2015. remote: Empirical orthogonal teleconnections in R. J. Statist. Soft. 65 (10), 1–19.
Appelhans, T., Mwangomo, E., Otte, I., Detsch, F., Nauss, T., Hemp, A., Eco-meteorological characteristics of the southern slopes of Mt. Kilimanjaro, Tanzania. Int. J. Climatol. (submitted for publication).
Ashraf, M., Loftis, J.C., Hubbard, K., 1997. Application of geostatistics to evaluate partial weather station networks. Agricult. Forest Meteorol. 84 (34), 255–271.
Attorre, F., De Sanctis, M., Farcomeni, A., Guillet, A., Scepi, E., Vitale, M., Pella, F., Fasola, M., 2013. The use of spatial ecological modelling as a tool for improving the assessment of geographic range size of threatened species. J. Nat. Conserv. 21, 48–55.
Baltas, E., 2007. Spatial distribution of climatic indices in northern Greece. Meteorol. Appl. 14 (1), 69–78.
Behrangi, A., Hsu, K.-l., Imam, B., Sorooshian, S., Huffman, G.J., Kuligowski, R.J., 2009. PERSIANN-MSA: A precipitation estimation method from satellite-based multispectral analysis. J. Hydrometeorol. 10 (6), 1414–1429.
Chen, D., Gong, L., Xu, C.-Y., Halldin, S., 2007. A high-resolution, gridded dataset for monthly temperature normals (1971–2000) in Sweden. Geogr. Ann. Ser. A Phys. Geogr. 89A (4), 249–261.
Collins, F.C., Bolstad, P., 1996. A comparison of spatial interpolation techniques in temperature estimation. Online-publication at http://goo.gl/utj6e7.
Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guegan, J.-F., Hawkins, B.a., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E., Turner, J.R.G., 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecol. Lett. 7 (12), 1121–1134.
Daly, C., 2006. Guidelines for assessing the suitability of spatial climate data sets. Int. J. Climatol. 26 (6), 707–721.
Daly, C., Smith, J.W., Smith, J.I., McKane, R.B., 2007. High-resolution spatial modeling of daily weather elements for a catchment in the oregon cascade mountains, United States. J. Appl. Meterol. Climatol. 46 (10), 1565–1586.
Di Luzio, M., Johnson, G.L., Daly, C., Eischeid, J.K., Arnold, J.G., 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. J. Appl. Meterol. Climatol. 47 (2), 475–497.
Diodato, N., 2005. The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain. Int. J. Climatol. 25 (3), 351–363.
Dombayc, O., Glc, M., 2009. Daily means ambient temperature prediction using artificial neural networks method: a case study of turkey. Renew. Energy 34.
Fries, A., Rollenbeck, R., Göttlicher, D., Nauß, T., Homeier, J., Peters, T., Bendix, J., 2009. Thermal structure of a megadiverse Andean mountain ecosystem in southern Ecuador and its regionalization. Erdkunde 63 (4), 321–335.
Goodale, C., Aber, J., Ollinger, S., 1998. Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model. Clim. Res. 10 (1), 35–49.
Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. 228, 113–129.
Guler, M., Cemek, B., Gunal, H., 2007. Assessment of some spatial climatic layers through GIS and statistical analysis techniques in Samsun Turkey. Meteorol. Appl. 14 (2), 163–169.
Hasenauer, H., 2003. Validating daily climate interpolations over complex terrain in Austria. Agricult. Forest Meteorol. 119 (1–2), 87–107.

Hawkins, B., Field, R., Cornell, H., 2003. Energy, water, and broad-scale geographic patterns of species richness. Ecology 84 (12), 3105–3117.
Hemp, A., 2005. Climate change-driven forest fires marginalize the impact of ice cap wasting on Kilimanjaro. Glob. Change Biol. 11 (7), 1013–1023.
Hemp, A., 2006. Vegetation of Kilimanjaro: hidden endemics and missing bamboo. Afr. J. Ecol. 44 (3), 305–328.
Hiemstra, P., Pebesma, E., Twenhöfel, C., Heuvelink, G., 2008. Real-time automatic interpolation of ambient gamma dose rates from the dutch radioactivity monitoring network. Comput. Geosci. 35 (8), 1711–1721. http://dx.doi.org/10.1016/j.cageo.2008.10.011.
Ho, H.C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., Henderson, S.B., 2014. Mapping maximum urban air temperature on hot summer days. Remote Sens. Environ. 154, 38–45.
Hofstra, N., Haylock, M., New, M., Jones, P., Frei, C., 2008. Comparison of six methods for the interpolation of daily, european climate data. J. Geogr. Res. 113.
Holdaway, M., 1996. Spatial modeling and interpolation of monthly temperature using kriging. Clim. Res. 06 (3), 215–225.
Hulme, M., Conway, D., Jones, P., Jiang, T., Barrow, E., Turney, C., 1995. Construction of a 1961–1990 European climatology for climate change modelling and impact applications. Int. J. Climatol. 15 (12), 1333–1363.
Hutchinson, M.F., 1995. Interpolating mean rainfall using thin-plate smoothing splines. In: 2nd International Conference/Workshop on Integrating GIS and Environmental Modelling. Breckenridge, CO, SEP, 1993. Int. J. Geogr. Inf. Syst. 9 (4), 385–403.
Hyndman, R.J., 2015. forecast: Forecasting functions for time series and linear models. R package version 5.8.
Jarvis, C.H., Stuart, N., 2001. A comparison among strategies for interpolating maximum and minimum daily air temperatures. part i: The selection of guiding topographic and land cover variables. J. Appl. Meterol. 40, 1060–1074.
Kearney, M.R., Isaac, A.P., Porter, W.P., 2014. microclim: Global estimates of hourly microclimate based on long-term monthly climate averages. Sci. Data 1, 1–9.
Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Perec Tadi, M., Bajat, B., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res.: Atmos. 119 (5), 2294–2313.
Krige, D., 1951. A statistical approach to some basic mine valuation problems on the witwatersrand. J. Chem. Metall. Min. Soc. South Africa 52 (6), 119–139.
Kuhn, M., 2014. caret: Classification and Regression Training. R package version 6.0-30.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer New York, New York, NY.
Kühnlein, M., Appelhans, T., Thies, B., Nauss, T., 2014. Improving the accuracy of rainfall rates from optical satellite sensors with machine learning — A random forests-based approach applied to MSG SEVIRI. Remote Sens. Environ. 141, 129–143.
Lennon, J., Turner, J., 1995. Predicting the spatial-distribution of climate — temperature in Great-Britain. J. Anim. Ecol. 64 (3), 370–392.
Li, J., 2013. Predicting the spatial distribution of seabed gravel content using random forest, spatial interpolation methods and their hybrid methods. In: 20th International Congress on Modelling and Simulation.
Li, J., Heap, A.D., 2008. A Review of Spatial Interpolation Methods for Environmental Scientists. Tech. Rep. 2008/023. Geoscience Australia, Canberra.
Li, J., Heap, A., Potter, A., Daniell, J.J., 2011a. Application of machine learning methods to spatial interpolation of environmental variables. Envir. Modell. Softw. 26, 1647–1659.
Li, J., Heap, A., Potter, A., Huang, Z., Daniell, J.J., 2011b. Can we improve the spatial predictions of seabed sediments? a case study of spatial interpolation of mud content across the southwest australian margin. Cont. Shelf Res. 31, 1365–1376.
Li, J., Potter, A., Huang, Z., Heap, A.D., 2012. Predicting Seabed Sand Content across the Australian Margin Using Machine Learning and Geostatistical Methods. Vol. Record 2012/048. Geoscience Australia, Canberra.
Liess, M., Glaser, B., Huwe, B., 2012. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. Geoderma 170, 70–79.
Loeffler, J., Pape, R., Wundram, D., 2006. The climatologic significance of topography, altitude and region in high mountains a survey of oceanic-continental differentiations of the scandes. Erdkunde 60 (1), 15–24.
Melini, D., 2013. A spatial model for sporadic tree species distribution in support of tree oriented silviculture. Ann. Silvic. Res. 37 (1), 64–68.
Myers, N., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403 (February), 853–858.
Nalder, I.A., Wein, R.W., 1998. Spatial interpolation of climatic normals: test of a new method in the canadian boreal forest. Agricult. Forest Meteorol. 92 (4), 211–225.
Ninyerola, M., Pons, X., Roure, J.M., 2000. A methodological approach of climatological modelling of air temperature and precipitation through gis techniques. Int. J. Climatol. 20 (14), 1823–1841.
Ortiz-Garca, E., Salcedo-Sanz, S., Casanova-Mateo, C., Paniagua-Tineo, A., Portilla-Figueras, J., 2012. Accurate local very short-term temperature prediction based on synoptic situation support vector regression banks. Atmos. Res. 107 (0), 1–8.
Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26 (1), 217–222.
Paniagua-Tineo, A., Salcedo-Sanz, S., Casanova-Mateo, C., Ortiz-Garca, E., Cony, M., Hernndez-Martn, E., 2011. Prediction of daily maximum temperature using a support vector regression algorithm. Renew. Energy 36 (11), 3054–3060.
Panitz, H.-J., Dosio, A., Büchner, M., Lüthi, D., Keuler, K., 2014. COSMO-CLM (CCLM) climate simulations over CORDEX-Africa domain: analysis of the ERA-Interim driven simulations at 0.44 and 0.22 resolution. Clim. Dynam. 42 (11–12), 3015–3038.
Peters, J., De Baets, B., Verhoest, N.E.C., Samson, R., Degroeve, S., De Becker, P., Huybrechts, W., 2007. Random forests as a tool for ecohydrological distribution modelling. Ecol. Modell. 207, 304–318.
Polishchuk, P.G., Muratov, E.N., Artemenko, A.G., Kolumbin, O.G., Muratov, N.N., Kuz'min, V., 2009. Application of random forest approach to qsar prediction of aquatic toxicity. J. Chem. Inf. Model 49, 2481–2488.
Pozdnoukhov, A., Foresti, L., Kanevski, M., 2009. Data-driven topo-climatic mapping with machine learning methods. Nat. Hazards 50, 497–518.
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. Agricult. Forest Meteorol. 101 (23), 81–94.

Radhika, Y., Shashi, M., 2009. Atmospheric temperature prediction using support vector machines. Int. J. Comput. Theory Eng. 1 (1), 1793–8201.
Ryan, B., 1977. Mathematical-model for diagnosis and prediction of surface winds in mountainous terrain. J. Appl. Meterol. 16 (6), 571–584.
Sala, O.E., 2000. Global biodiversity scenarios for the year 2100. Science 287 (5459), 1770–1774.
Sanabria, L.A., Qin, X., Li, J., Cechet, R.P., Lucas, C., 2013. Spatial interpolation of mcarthurs forest fire danger index across australia: Observational study. Environ. Modell. Softw. 50, 37–50.
Siaulys, A., Bucas, M., 2012. Species distribution modelling of benthic invertebrates in the south-eastern baltic sea. Baltica 25, 163–170.
Smith, B.A., Hoogenboom, G., McClendon, R.W., 2009. Artificial neural networks for automated year-round temperature prediction. Comput. Electron. Agric. 68 (1), 52–61.
Snell, S.E., Gopal, S., Kaufmann, R.K., 2000. Spatial interpolation of surface air temperatures using artificial neural networks: Evaluating their use for downscaling gcms. J. Clim. 13 (5), 886–895.
Stahl, K., Moore, R., Floyer, J., Asplin, M., Mckendry, I., 2006. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agricult. Forest Meteorol. 139 (3–4), 224–236.
Tabios, G., Salas, J., 1985. A comparative-analysis of techniques for spatial interpolation of precipitation. Water Resour. Res. 21 (3), 365–380.
Tasaduqq, S., Rehman Bubshait, K., 2002. Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. Renew. Energy 25, 545–554.
Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190 (34), 214–251.
Tripathi, S., Srinivas, V.V., Nanjundiah, R.S., 2006. Downscaling of precipitation for climate change scenarios: A support vector machine approach. J. Hydrol. 330, 621–640.
Tucker, C., Pinzon, J., Brown, M., Slayback, D., Pak, E., Mahoney, R., Vermote, E., El Saleous, N., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. Int. J. Remote Sens. 26 (20), 4485–4498.
Ustaoglu, B., Cigizoglu, H.K., Karaca, M., 2008. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Meteorol. Appl. 15 (4), 431–445.
Ustrnul, Z., Czekierda, D., 2005. Application of GIS for the development of climatological air temperature maps: an example from Poland. Meteorol. Appl. 12 (1), 43–50.
Vicente-Serrano, S., Saz-Sanchez, M., Cuadrat, J., 2003. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. Clim. Res. 24 (2), 161–180.
Willmott, C.J., Matsuura, K., 1995. Smart interpolation of annually averaged air temperature in the United States. J. Appl. Meteor. 34 (12), 2577–2586.
Worlen, C., Schulz, K., Huwe, B., Eiden, R., 1999. Spatial extrapolation of agrometeorological variables. Agricult. Forest Meteorol. 94 (3-4), 233–242.
Xia, Y., Fabian, P., Winterhalter, M., Zhao, M., 2001. Forest climatology: estimation and use of daily climatological data for bavaria, germany. Agricult. Forest Meteorol. 106 (2), 87–103.
Xia, Y., Winterhalter, M., Fabian, P., 1999. A model to interpolate monthly mean climatological data at Bavarian forest climate stations. Theor. Appl. Climatol. 64 (1–2), 27–38.