Predicting missing values in spatio-temporal remote sensing data

Gerber, Florian ; de Jong, Rogier ; Schaepman, Michael E ; Schaepman-Strub, Gabriela ; Furrer, Reinhard

Abstract: Continuous, consistent, and long time-series from remote sensing are essential to monitoring changes on Earth's surface. However, analyzing such data sets is often challenging due to missing values introduced by cloud cover, missing orbits, sensor geometry artifacts, and so on. We propose a new and accurate spatio-temporal prediction method to replace missing values in remote sensing data sets. The method exploits the spatial coherence and temporal seasonal regularity that are inherent in many data sets. The key parts of the method are: 1) the adaptively chosen spatio-temporal subsets around missing values; 2) the ranking of images within the subsets based on a scoring algorithm; 3) the estimation of empirical quantiles characterizing the missing values; and 4) the prediction of missing values through quantile regression. One advantage of quantile regression is the robustness to outliers, which enables more accurate parameter retrieval in the analysis of remote sensing data sets. In addition, we provide bootstrap-based quantification of prediction uncertainties. The proposed prediction method was applied to a Normalized Difference Vegetation Index data set from the Moderate Resolution Imaging Spectroradiometer and assessed with realistic test data sets featuring between 20% and 50% missing values. Validation against established methods showed that the proposed method has a good performance in terms of the root-mean-squared prediction error and significantly outperforms its competitors. This paper is accompanied by the open-source R package gapfill, which provides a flexible, fast, and ready-to-use implementation of the method.

DOI: https://doi.org/10.1109/TGRS.2017.2785240

Predicting Missing Values in Spatio-Temporal Remote Sensing Data

Florian Gerber, Rogier de Jong, Michael E. Schaepman, *Senior Member, IEEE*, Gabriela Schaepman-Strub, and Reinhard Furrer

***Abstract*— Continuous, consistent, and long time-series from remote sensing are essential to monitoring changes on Earth's surface. However, analyzing such data sets is often challenging due to missing values introduced by cloud cover, missing orbits, sensor geometry artifacts, and so on. We propose a new and accurate spatio-temporal prediction method to replace missing values in remote sensing data sets. The method exploits the spatial coherence and temporal seasonal regularity that are inherent in many data sets. The key parts of the method are: 1) the adaptively chosen spatio-temporal subsets around missing values; 2) the ranking of images within the subsets based on a scoring algorithm; 3) the estimation of empirical quantiles characterizing the missing values; and 4) the prediction of missing values through quantile regression. One advantage of quantile regression is the robustness to outliers, which enables more accurate parameter retrieval in the analysis of remote sensing data sets. In addition, we provide bootstrap-based quantification of prediction uncertainties. The proposed prediction method was applied to a Normalized Difference Vegetation Index data set from the Moderate Resolution Imaging Spectroradiometer and assessed with realistic test data sets featuring between 20% and 50% missing values. Validation against established methods showed that the proposed method has a good performance in terms of the root-mean-squared prediction error and significantly outperforms its competitors. This paper is accompanied by the open-source R package *gapfill*, which provides a flexible, fast, and ready-to-use implementation of the method.**

***Index Terms*— Alaska, gap filling, imputation, interpolation, Moderate Resolution Imaging Spectroradiometer Normalized Difference Vegetation Index (MODIS NDVI), quantile regression, R gapfill, TIMESAT, uncertainty.**

I. Introduction

REMOTE sensing data are used to study a wide range of Earth surface processes. The derived data sets have the advantage of extensive spatial and temporal coverage. A less attractive feature is that they often contain low-quality or missing values. In particular, data sets derived from optical satellite sensors partially show clouds instead of the Earth's surface. As a consequence, the usability of the data is limited. For example, proper monitoring of continuous vegetation change can be inhibited (see Arctic [1], Amazon [2], and general [3]). To deal with this issue, low-quality and missing values are often excluded from the analysis or they are replaced with predictions from a variety of prediction methods (also called "gap filling methods" or "imputation methods"). Many prediction methods exploit the temporal correlation of the data [4], [5], but only little attempt has been undertaken to date to exploit the spatio-temporal nature of the data [6]. In this paper, we introduce a new spatio-temporal prediction method and assess its performance.

A. Existing Methods to Predict Missing Values

A selection of methods to replace missing values in remote sensing data sets is presented in Table I. They are grouped into methods that exploit the temporal, spatial, or spatio-temporal dependence structure of the data. Furthermore, software implementations and distribution under open-source licenses as well as eventual provisioning of uncertainties for the predicted values are listed in Table I. It should be noted that some methods focus on the prediction of missing values and are applied before data analysis (e.g., gapfill-MAP), while others are designed to analyze the data and return predictions for missing values as a byproduct (e.g., TIMESAT).

B. Novelty and Outline of the Proposed Method

The methods listed in Table I are well founded and capable of predicting missing values in remote sensing data sets. However, we discovered at least one of the following drawbacks in each of the methods that we studied: 1) predictions partially fail, especially when missing values are within large gaps; 2) lack of quantification of prediction uncertainties; 3) nonavailability of well-documented open-source software to support the usage and further developments of the method; 4) nonscalability to large data sets, especially through the choice of methodology that prevents effective parallelization; and 5) low speed because of computationally expensive methods that require large-scale storage.

The novelty of the proposed prediction method is that it overcomes all of these drawbacks. More precisely, it avoids drawback: 1) by predicting all missing values separately based on dynamically chosen subsets, which allows the method to

Manuscript received September 27, 2016; revised February 3, 2017, June 28, 2017, and December 3, 2017; accepted December 14, 2017. This work was supported by the University of Zürich Research Priority Program on Global Change and Biodiversity. *(Corresponding author: Reinhard Furrer.)*

F. Gerber is with the Department of Mathematics, University of Zürich, 8057 Zürich, Switzerland.

R. de Jong and M. E. Schaepman are with the Remote Sensing Laboratories, Department of Geography, University of Zürich, 8057 Zürich, Switzerland.

G. Schaepman-Strub is with the Department of Evolutionary Biology and Environmental Studies, University of Zürich, 8057 Zürich, Switzerland.

R. Furrer is with the Department of Mathematics, University of Zürich, 8057 Zürich, Switzerland, and also with the Department for Computational Science, University of Zürich, 8057 Zürich, Switzerland (e-mail: reinhard.furrer@math.uzh.ch).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2017.2785240

TABLE I

OVERVIEW OF PUBLISHED METHODS TO PREDICT MISSING VALUES IN REMOTE SENSING DATA SETS. IF A SOFTWARE IMPLEMENTATION OF THE METHOD IS AVAILABLE, THE PROGRAMMING LANGUAGE IS INDICATED IN THE COLUMN "LANGUAGE." OPEN-SOURCE SOFTWARE IS MARKED WITH A ⋆ SYMBOL. THE COLUMN NAME "PU" ABBREVIATES "PREDICTION UNCERTAINTY" AND THE ✓SYMBOL IN THAT COLUMN INDICATES WHETHER THE METHOD PROVIDES UNCERTAINTY QUANTIFICATION OF THE PREDICTED VALUES

	Method	Details	Language	PU	Reference
Temporal	BFAST	Breaks For Additive Season and Trend	R ⋆	–	[7], [8]
	CACAO	Consistent Adjustment of Climatology to Actual Observations	–	–	[9]
	DBEST	Detecting Breakpoints and Estimating Segments in Trend	R ⋆	–	[10]
	HANTS	Harmonic ANalysis of Time-Series	Matlab, Fortran	–	[11]
	LOESS	LOcally wEighted Scatterplot Smoothing	IDL	–	[12]
	neural networks	neural networks combined with Savitzky-Golay functions	–	–	[13]
	splines	spline models with elevation as additional variable	GRASS GIS ⋆	–	[14]
	splines and Fourier	linear and spline models combined with Fourier analysis	quickBASIC	–	[15]
	TIMESAT	Savitzky-Golay, asymmetric Gaussian, double logistic functions	Matlab, Fortran	–	[16]–[18]
	TiSeG	Time-Series Generator for MODIS data	IDL	–	[19]
Spatial	co-kriging	spatial variogram models	–	✓	[20]–[22]
	GNSPI	Geostatistical Neighborhood Similar Pixel Interpolator	–	–	[23], [24]
Spatio-temporal	EOF	Empirical Orthogonal Functions		✓	[25]
	GAM and kriging	Generalized Additive Model and kriging	GRASS GIS, R ⋆	✓	[26]
	gapfill	quantile regressions fitted to spatio-temporal subsets	R, C++ ⋆	✓	[27]
	gapfill-MAP	prediction method used in the Malaria Atlas Project	Python, C ⋆	✓	[28]
	Kalman-filtering	Monte Carlo version of Kalman-filtering	–	✓	[29]
	linear regression	linear regressions fitted to spatio-temporal subsets	–	–	[30], [31]
	SSA	Singular Spectrum Analysis	R ⋆	–	[6], [32]
	space-time kriging	spatio-temporal variogram models	–	✓	[33], [34]

deal with small and large gaps; 2) by including statistical considerations, which enable the quantification of prediction uncertainties; 3) because it is available as open-source R package *gapfill* featuring well-structured R and C++ code, unit tests, and documentation; 4) by choosing a parallelizable method and implementation; and 5) by relying on algorithms that allow high computing speed.

The proposed method performs the following four steps [see Fig. 1(a)] for each missing value.

1) *Extract Subset:* Iteratively select a large enough spatio-temporal neighborhood of the target value, i.e., the considered missing value to predict. We call that neighborhood the "prediction set."
2) *Rank Images:* Calculate a score for each image in the prediction set based on valuewise comparisons between all images in the prediction set. (An image refers to all values of the data set observed at one point in time.) Rank the images in the prediction set based on their scores and index the images by their ranks.
3) *Estimate Quantile:* For all images in the prediction set that have an observed value at the spatial location of the target value, determine to which empirical quantile that value corresponds relative to all values of the image. Take the mean of the resulting empirical quantile levels obtained from all images and use that as a target quantile level τ.
4) *Quantile Regression:* Regress all values in the prediction set on their associated image ranks using τ-quantile regression. And finally, predict the target value based on the fitted quantile regression.

Therefore, we can classify the proposed method within all other methods. It belongs to the group of spatio-temporal methods, exploiting the characteristic tendency of spatial coherence as well as the temporal seasonal dynamics of remote sensing data sets in order to predict missing values. Similar to the methods described in [28], [30], and [31], this method predicts the missing values in a data set based on spatio-temporal subsets around the missing values.

II. PREDICTION METHOD

We start this section with an illustration of the proposed method (see Section II-A), such that the main ideas become immediately clear. In the following, the complete formal description (see Section II-B) is given in order to explain details, which are important for potential extensions and modifications of the method.

A. Illustration With a Test Data Example

For illustration, we consider the test data set shown in Fig. 1(b), which is based on the Moderate Resolution Imaging Spectroradiometer Normalized Difference Vegetation Index (MODIS NDVI) product MOD13A1. This product comprises several data layers (variables) containing values on a common regular grid in space and time with a resolution of 500 m, covering 16 days. One of the layers consists of NDVI values and another one indicates the reliability of each value based on quality assignments [35]. The selected data set has a spatial extent of $n_x \cdot n_y = 21 \cdot 21$ values and consists of 16 images having $n_s = 4$ seasonal indices (the days 145, 161, 177, and 193 of the year) observed over $n_a = 4$ years (2004–2007). In total, the data set has $n_x n_y n_s n_a = 7056$ (observed or missing) values. The 1603 ($\approx 23\%$) values shown in black are flagged as less than "good quality" by the quality

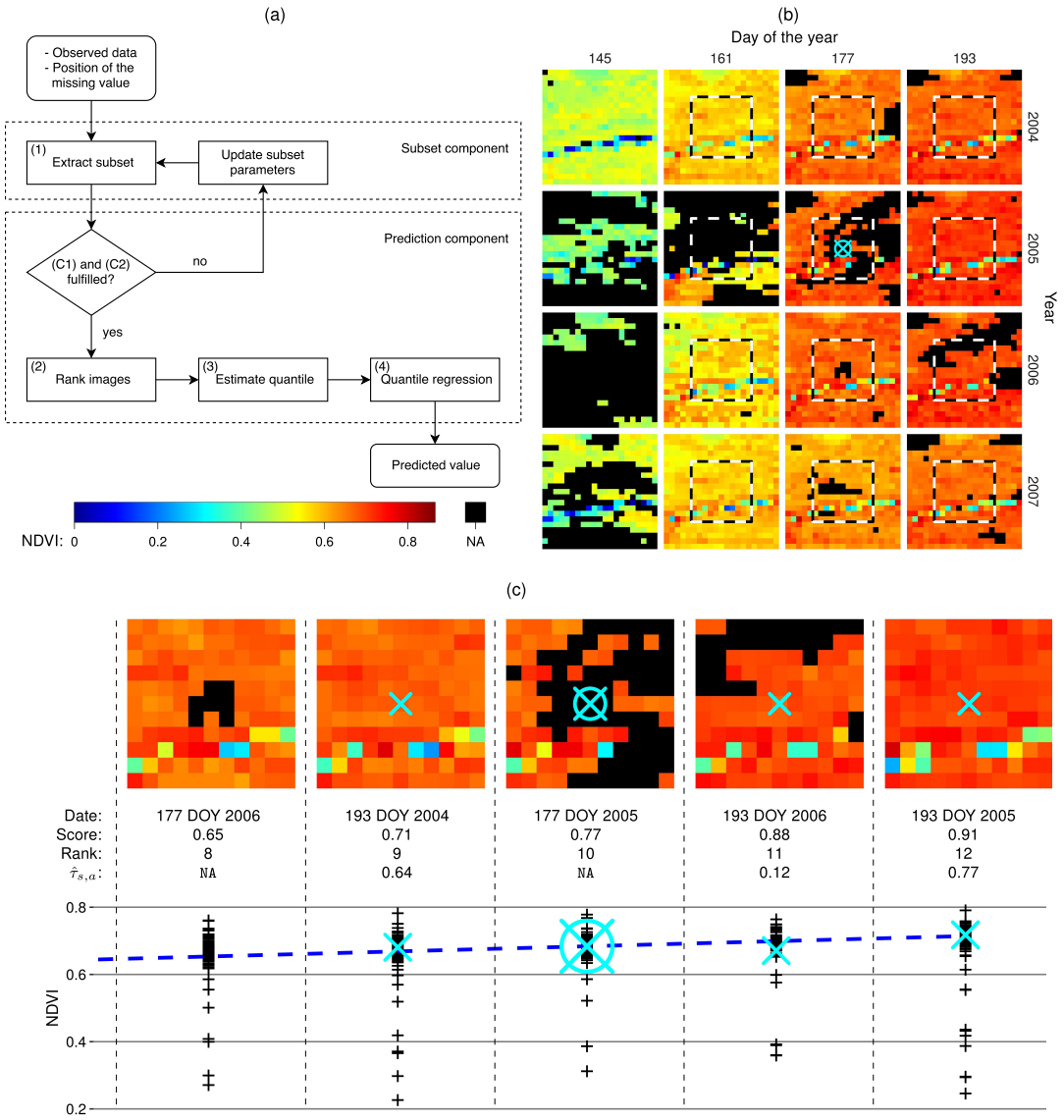


Fig. 1. (a) Flow diagram of the proposed method showing the subset and prediction components as well as the involved steps 1–4 used to predict one missing value. (b) Example MODIS NDVI data set featuring values for 21×21 spatial locations and 4×4 time points. The 1603 missing values are shown in black. The prediction of the target value ($\otimes$-symbol) is based on the subimages of the prediction set (indicated by dashed squares). (c) Five subimages of the prediction set ranked 8–12 and a scatter plot of the observed NDVI values (y-axis) and the rank of the subimages (x-axis) are shown. The scatter plot also shows the fitted line of the quantile regression (dashed line) and the predicted value ($\otimes$-symbol). The $\hat{\tau}_{s,a}$ values are estimates from the empirical quantiles of the values marked with $\times$-symbols relative to their respective subimages.

layer of the data set and we treat them as missing. The proposed method predicts each missing value separately and applies steps 1–4 summarized in Fig. 1(a) to each of them. In the following, we illustrate the prediction of the missing value highlighted by the $\otimes$-symbol in the image of day 177 in the year 2005 [see Fig. 1(b)]. We refer to that value as the "target value."

In step 1, we select the prediction set, which is a sufficiently large spatio-temporal neighborhood of the target value. The prediction set for the considered target value is indicated by the dashed squares [see Fig. 1(b)]. It consists of 12 smaller subimages contained in the test data set. To select the prediction set, we iteratively increase the spatial extent of the neighborhood until enough subimages thereof are nonempty and the subimage containing the missing value has enough observed values. In the remaining steps 2–4, the target value is predicted from the prediction set.

In step 2, we first score each subimage in the prediction set. The score of a subimage is defined as the proportion of values in the subimage that are larger than all values in the prediction set with the same spatial location. This score induces a ranking of the subimages, i.e., the subimage with the smallest score has rank one, and so on. In Fig. 1(c), the subimages ranked 8–12 and their scores are shown.

To explain step 3, we consider the bottom row of Fig. 1(c), which displays the observed NDVI values of the shown subimages (y-axis) plotted against the ranks of the subimages (x-axis) by a scatter plot. The $\times$-symbols indicate values that

have the same spatial location as the target value. Their vertical positions in the scatter plot provide a visual estimation of their empirical quantile relative to the values of the corresponding subimages. If possible, we estimate those empirical quantiles for each subimage [indicated as $\hat{\tau}_{s,a}$ in Fig. 1(c)]. Then, the empirical quantile of the target value is estimated as the mean of all estimated empirical quantiles of all subimages.

In the final step 4, we fit a quantile regression to the observed values of the prediction set. The regression uses an intercept and the associated ranks of the subimages as linear predictors. As quantile of interest, we use the estimated empirical quantile of the target value. In the scatter plot of Fig. 1(c), the fitted quantile regression line (dashed line) and the corresponding predicted target value ($\otimes$-symbol) are shown.

B. Formal Description

In this section, we formalize the proposed prediction method illustrated in Section II-A. We store the observed remote sensing data in the $n_x \times n_y \times n_s \times n_a$ array Z, where n_x and $n_y \in \mathbb{N}$ are the spatial extents of the images, $n_s \in \mathbb{N}$ is the number of images within a year, and $n_a \in \mathbb{N}$ is the number of years. The values of Z are either observed values in $\mathbb{R}$ or missing (denoted with NA). We use the square bracket notation to select subsets of arrays. For example, $Z[x, y, s, a]$ denotes one value of Z, where the indices $x \in \{1, \ldots, n_x\}$ and $y \in \{1, \ldots, n_y\}$ describe the spatial location, the index $s \in \{1, \ldots, n_s\}$ the season (position within a year), and the index $a \in \{1, \ldots, n_a\}$ the year. In addition, we write $k_1{:}k_2$ with $k_1, k_2 \in \mathbb{N}$ as an abbreviation for the integer sequence $k_1, \ldots, k_2$. With that, we can specify array subsets, e.g., $Z[1{:}3, 1, 1, 1]$ selects three values of Z. We use the dot notation to indicate all valid indices. For example, $Z[\cdot, \cdot, \cdot, \cdot]$ equals Z and $Z[\cdot, \cdot, 1, 1]$ denotes all values of Z with seasonal index 1 of year 1, and hence a spatial image.

The proposed method predicts each missing value in Z independently of other missing values. Therefore, it is sufficient to describe the prediction of one missing value $Z[x_T, y_T, s_T, a_T]$, where the subscript "T" stands for "target value." To simplify notation, we assume that $Z[x_T, y_T, s_T, a_T]$ is far away from the boundaries of Z. To obtain a flexible and user-friendly implementation of the method, we split it into a subset and a prediction component [see Fig. 1(a)]. The subset component selects a subset of Z around $Z[x_T, y_T, s_T, a_T]$ according to a neighborhood search scheme detailed in Section II-B1. The subsequent prediction component decides whether it is possible to predict $Z[x_T, y_T, s_T, a_T]$ based on that subset. If so, $Z[x_T, y_T, s_T, a_T]$ is predicted as described in Section II-B2; otherwise, the procedure returns to the subset component and updates the subset parameters to extract a larger subset. This iterative algorithm is repeated until the prediction component decides that it is possible to predict $Z[x_T, y_T, s_T, a_T]$ based on the selected subset [see Fig. 1(a)].

1) Subset Component: To be more specific about the search strategy for suitable subsets [step 1 in Fig. 1(a)], we define the following function, which returns a subset of Z around the target value $Z[x_T, y_T, s_T, a_T]$:

$$\begin{aligned} f(Z, i) &= f(Z; i, x_T, y_T, s_T, a_T, \lambda_x, \lambda_y, \lambda_s, \lambda_a) \\ &= Z[(x_T - (\lambda_x + i)){:}(x_T + \lambda_x + i), \\ &\qquad (y_T - (\lambda_y + i)){:}(y_T + \lambda_y + i), \\ &\qquad (s_T - \lambda_s){:}(s_T + \lambda_s), (a_T - \lambda_a){:}(a_T + \lambda_a)]. \end{aligned} \quad (1)$$

The parameters $\lambda_x, \lambda_y, \lambda_s, \lambda_a \in \mathbb{N}$ are tuning parameters, which define the initial size of the subset. The values of the tuning parameters used in this paper are given in Section III-C. The index i is set to 0 initially and is increased by one whenever the selected subset is rejected by the prediction component. By increasing i, the subset selected by $f(Z; i)$ grows in the spatial dimension but not in the temporal dimension.

2) Prediction Component: The first task of the prediction component is to decide whether it is possible to predict the target value $Z[x_T, y_T, s_T, a_T]$ based on the subset selected by $f(Z; i)$. We require that the subset fulfills the following criteria.

- *C1:* The subset must contain at least θ_1 nonempty subimages.
- *C2:* The subimage in the subset containing the target value $Z[x_T, y_T, s_T, a_T]$ must have at least θ_2 observed values.

θ_1 and θ_2 are again tuning parameters. If C1 and C2 are fulfilled, we call the subset selected by $f(Z; i)$ the prediction set and denote it by P. For further description, let P be an $n_x^p \times n_y^p \times n_s^p \times n_a^p$ array and let x_t, y_t, s_t, a_t be the indices of the target value relative to P, i.e., $P[x_t, y_t, s_t, a_t] = Z[x_T, y_T, s_T, a_T]$.

In step 2, we rank the subimages in the prediction set P based on a scoring algorithm. The underlying assumption of the scoring algorithm is that the subimages have similar but potentially shifted distributions of values. The algorithm scores each subimage separately and the score of a subimage is defined as the proportion of values in the subimage that are larger compared with the values at the same spatial coordinates in all other subimages. Missing values are automatically corrected for, since the proportions are calculated based on observed values only. Pseudocode 1 describes in detail how the score of a subimage is obtained. Note that the mean function in the pseudocodes is assumed to return the mean of the observed values. The scoring algorithm (Pseudocode 1) is repeated for all subimages in P. Note that some subimages may receive an NA as a score, e.g., because the subimage consists of NA values only, and we exclude subimages with NA scores from the further analysis. The subimages are ranked by increasing score, i.e., the subimage with the smallest score is assigned rank $r = 1$, and so on. With that, each value in P has an associated rank r. We close this part by mentioning that scoring of images containing missing values is the subject of current research [36].

In step 3, we estimate the target quantile level τ_{s_t,a_t} relative to $P[\cdot, \cdot, s_t, a_t]$, i.e., the subimage containing the target value. If all values of $P[\cdot, \cdot, s_t, a_t]$ were observed, we could simply estimate the empirical cumulative distribution function (ECDF) $\hat{F}_{s_t,a_t}(\cdot)$ from $P[\cdot, \cdot, s_t, a_t]$ and set $\hat{\tau}_{s_t,a_t} = \hat{F}_{s_t,a_t}(P[x_t, y_t, s_t, a_t])$ to obtain an estimate of the

Pseudocode 1 Score the Subimage $P[\cdot, \cdot, s_k, a_k]$ Relative to the Other Subimages in P

Input: P, s_k, a_k, n_x^p, n_y^p, n_s^p, n_a^p
Output: Score of the subimage $P[\cdot, \cdot, s_k, a_k]$
Define R as a $n_s^p \times n_a^p$ matrix
$R[\cdot, \cdot] \leftarrow$ NA
Define M as a $n_x^p \times n_y^p$ matrix
for $s \in \{1, \ldots, n_s^p\}$ and $a \in \{1, \ldots, n_a^p\}$ **do**
  $M[\cdot, \cdot] \leftarrow$ NA
  for $x \in \{1, \ldots, n_x^p\}$ and $y \in \{1, \ldots, n_y^p\}$ **do**
    if $P[x, y, s_k, a_k]$ and $P[x, y, s, a]$ are not NA **then**
      if $P[x, y, s_k, a_k] > P[x, y, s, a]$ **then**
        $M[x, y] \leftarrow 1$
      else
        $M[x, y] \leftarrow 0$
      end if
    end if
  end for
  $R[s, a] \leftarrow \text{mean}(M)$
end for
return mean(R)

Pseudocode 2 Estimate the Quantile τ_{s_t,a_t} of the Target Value $P[x_t, y_t, s_t, a_t]$ Relative to $P[\cdot, \cdot, s_t, a_t]$

Input: P, ν, x_t, y_t, n_x^p, n_y^p, n_s^p, n_a^p
Output: $\hat{\tau}_{s_t,a_t}$
Define V as a $n_s^p \times n_a^p$ matrix
$V[\cdot, \cdot] \leftarrow$ NA, $j \leftarrow 0$, $D \leftarrow P[x_t, y_t, \cdot, \cdot]$
while the number of observed values in $D < \nu$ **do**
  $j \leftarrow j+1$, $D \leftarrow P[(x_t - j):(x_t + j), (y_t - j):(y_t + j), \cdot, \cdot]$
end while
Define A as $(2j+1) \times (2j+1) \times n_s^p \times n_a^p$ array
$A[\cdot, \cdot, \cdot, \cdot] \leftarrow$ NA
for $s \in \{1, \ldots, n_s^p\}$ and $a \in \{1, \ldots, n_a^p\}$ **do**
  Estimate the ECDF $F_{s,a}(\cdot)$ from $P[\cdot, \cdot, s, a]$
  for $x \in \{1, \ldots, 2j+1\}$ and $y \in \{1, \ldots, 2j+1\}$ **do**
    $p \leftarrow P[x_t - (j+1) + x, y_t - (j+1) + y, s, a]$
    if p is not NA **then**
      $A[x, y, s, a] \leftarrow \hat{F}_{s,a}(p)$
    end if
  end for
  $V[s, a] \leftarrow \text{mean}(A[\cdot, \cdot, s, a])$
end for
return mean(V)

target quantile level. However, by definition, $P[x_t, y_t, s_t, a_t]$ is missing, and hence, we rely on the algorithm formalized in Pseudocode 2 to estimate τ_{s_t,a_t} from all other subimages in P. The algorithm first uses the observed values in P to estimate an ECDF $\hat{F}_{s,a}(\cdot)$ for each subimage. Then, it estimates $\hat{\tau}_{s,a} = \hat{F}_{s,a}(P[x_t, y_t, s, a])$ for each subimage (i.e., for all $s \in \{1, \ldots, n_s^p\}$ and $a \in \{1, \ldots, n_a^p\}$) and sets $\hat{\tau}_{s_t,a_t}$ to the mean of all those $\hat{\tau}_{s,a}$ values. Note that some $\hat{\tau}_{s,a}$ values may be NA because of missing $P[x_t, y_t, \cdot, \cdot]$ values. We require at least $\nu \in \mathbb{N}$ observed values in $P[x_t, y_t, \cdot, \cdot]$, where ν is the last tuning parameter. If this requirement is not met, the ECDFs $\hat{F}_{s,a}$ are evaluated in a spatial neighborhood of (x_t, y_t).

Finally, we fit a quantile regression [37], [38] to all observed values of P using an intercept and their associated rank r as linear predictors [see step 4 in Fig. 1(a)]. Quantile regression fits a regression line for an arbitrary but fixed quantile level τ (as opposed to ordinary regression where the conditional mean is modeled). This makes it robust to outliers, which are common in remote sensing data sets. For the quantile level τ, the regression is formalized as

$$Q(\tau \mid r) = \beta_0(\tau) + \beta_1(\tau) r \tag{2}$$

where $\beta_0(\tau)$ and $\beta_1(\tau) \in \mathbb{R}$ are the coefficients. We set $\tau = \hat{\tau}_{s_t,a_t}$ and fit the $\hat{\tau}_{s_t,a_t}$-quantile regression model to obtain the estimates $\hat{\beta}_0(\hat{\tau}_{s_t,a_t})$ and $\hat{\beta}_1(\hat{\tau}_{s_t,a_t})$. The prediction of the target value is then

$$\begin{aligned} \hat{Z}[x_T, y_T, s_T, a_T] &= \hat{P}[x_t, y_t, s_t, a_t] \\ &= \hat{\beta}_0(\hat{\tau}_{s_t,a_t}) + \hat{\beta}_1(\hat{\tau}_{s_t,a_t}) r_t \end{aligned} \tag{3}$$

where r_t denotes the rank associated with $P[\cdot, \cdot, s_t, a_t]$, i.e., the subimage containing the target value.

C. Prediction Uncertainties

Uncertainty estimates of the predicted values are essential when using them to derive conclusions in further analyses. Statistical theory provides ways to quantify uncertainty through prediction intervals. Possible approaches applicable to the proposed method are bootstrap and cross validation. However, both approaches are computationally expensive and inaccurate if the underlying assumptions about the data are not met.

We study the magnitude of the uncertainties introduced by steps 1–4 [see Fig. 1(a)] separately by constructing prediction intervals. We assess the uncertainties introduced in step 1 by running the proposed method with all possible initial sizes of the spatial subset. Then, we measure the variability in those predictions with a 90% prediction interval for each missing value derived from the 5% and the 95% quantile of the predicted values. The uncertainty of step 2 is assessed via permutations of the ranks of the subimages in the prediction set. We calculate the predicted values for all permutations and again construct a 90% prediction interval for each missing value by considering the 5% and the 95% quantile of the predictions. For step 3, we derive a 90% prediction interval by quantifying the variability in the estimated quantiles stored in V of Pseudocode 2. This variability is again summarized with the interval given by the 5% and the 95% quantiles. Finally, the uncertainty introduced in step 4 is assessed by calculating a 90% prediction interval based on bootstrap methods implemented in the R package *quantreg* [39].

Estimates of the prediction uncertainties could be derived by combining the uncertainties of all four previously described steps. However, doing so in a meaningful way is not straightforward due to possible interactions and elimination effects among them. Nevertheless, we combine the uncertainties from

steps 2 and 3 in one prediction interval. The lower bound of the prediction interval is obtained by fitting a quantile regression for the quantile level set to the 5% quantile of the values in V of Pseudocode 2. Then, all values in the prediction set P are predicted using that model fit and the 5% quantile of all those predictions is taken as lower bound of the interval. The upper bound is constructed similarly but uses 95% quantiles instead of 5% quantiles. An evaluation of the properties of that interval is given in Section IV-B.

D. Software Implementation

The proposed method is implemented in the programming languages R and C++ and is available as open-source R package *gapfill* at https://cran.r-project.org/package=gapfill [27]. *gapfill* features a flexible design allowing the user to optimize the prediction for specific data sets. It has separate subset and prediction functions corresponding to the subset and prediction components [see Fig. 1(a)]. This enables independent modification of those components and is useful to construction of new prediction methods with a little effort. Examples using an MODIS data set and a detailed documentation of the package are available in the reference manual [27] and in Section S1 in the supplementary material.

By design, the method is straightforward to parallelize, because it predicts each missing value independently of the others. To enable parallel execution, *gapfill* relies on tools from the R package *foreach* [40], which allows the user to choose between an OpenMP [41] and an MPI [42] back end depending on the architecture of the available computer. This makes the method scalable in the sense that *gapfill* can exploit the resources of both small and large computers. Hence, the method can be used to predict large amounts of missing values typical of remote sensing data with minimal effort.

III. Setup for Evaluation

Four test data sets are constructed to investigate the predicted values and uncertainties. In addition, we compare the accuracy of the predicted values against those of two competing prediction methods.

A. Selected MODIS Data

We consider the MODIS satellite product MOD13A1, which is part of the MODIS vegetation index product MOD13 [43]. MOD13A1 is a land surface product and is based on eight-day MODIS Level-2G surface reflectance data, which have been further composited to obtain the final resolution of 16 days and 500 m [44], [45]. The NDVI layer of MOD13A1 can be used to describe vegetation activity [46]. Moreover, we use the "pixel reliability" layer to set NDVI values flagged as less than "good data" to NA [35].

To evaluate the prediction method, NDVI data from the years 2004 to 2009 are considered and restricted to the region of northern Alaska, as shown in Fig. 2. Due to the high latitudes ranging from 66° north to more than 71° north, the NDVI values exhibit a strong seasonal component. That is reflected in both the NDVI values and the number of available values classified as "good data." Especially during wintertime, little data are available because of missing sunlight and snow cover. Therefore, we restrict the analysis to the seasonal period starting on the day of the year (DOY) 145 (about May 24) and ending on DOY 257 (about September 13). The data set then covers eight dates with observations each year and consists of at least 30% "good data" per DOY.

The MOD13A1 data are downloaded in six spatial tiles and merged to one single image per considered point in time using the R package *MODIS* [47], which interfaces the MODIS reprojection tool [48]. Furthermore, the data are transformed from the sinusoidal to the geographic map projection (WGS84). The R packages *raster* [49], *sp* [50], [51], *fields* [52], *lattice* [53], and *ggplot2* [54] are used to handle and visualize the data.

B. Test Data Sets Based on MODIS Data

Next, we construct four tests data sets based on the previously described subset of MOD13A1. Using real data, as opposed to simulated data, has the advantage that the data come close to the use-case of interest. To construct the four test data sets, we consider the data from the spatial region labeled with "Data" in Fig. 2. The "Data" region has a spatial extent of 100×100 values and is selected such that the resulting data set (referred to as "data subset") has a relatively few missing values (about 12%). Moreover, the data subset reflects typical features of NDVI data sets in high latitudes. Two of these features are the latitudinal gradient manifesting itself through lower NDVI values in the northern regions and low NDVI values that are caused by surface water. Note that the temporal extent of the data subset remains unchanged and has eight seasonal time points per year for six years.

The four test data sets are then constructed by artificially setting NDVI values of the data subset to NA. To mimic realistic spatio-temporal distributions of missing values, setting values to NA is performed according to the distributions of missing values observed at other locations of the Alaska data set (as opposed to selecting values randomly). We choose the distribution of missing values observed at the regions of the rectangles indicated by "20%," "30%," "40%," and "50%" in Fig. 2. This selection leads to test data sets with fractions of 20%, 30%, 40%, and 50% missing values. The four test data sets are shown in Figs. S1–S5 in the supplementary material.

For the four test data sets, we know the true values for most missing values and call them validation values. Hence, we can apply a prediction method to the test data sets and compare the predicted values to those validation values. To measure prediction accuracy, we use the root-mean-squared prediction error (RMSPE) and the mean absolute prediction error (MAPE) defined as $\left(\sum_{i=1}^{n}(\hat{z}_i - z_i)^2/n\right)^{1/2}$ and $\sum_{i=1}^{n}\left|\hat{z}_i - z_i\right|/n$, respectively. Here, n is the number of predicted values $\hat{z}_i$, and z_i denote the corresponding validation values [55].

In addition, a data set consisting of the entire spatial extent of northern Alaska, as shown in Fig. 2, was compiled. While the temporal dimensions of that data set remained unchanged, the size of the images is increased to

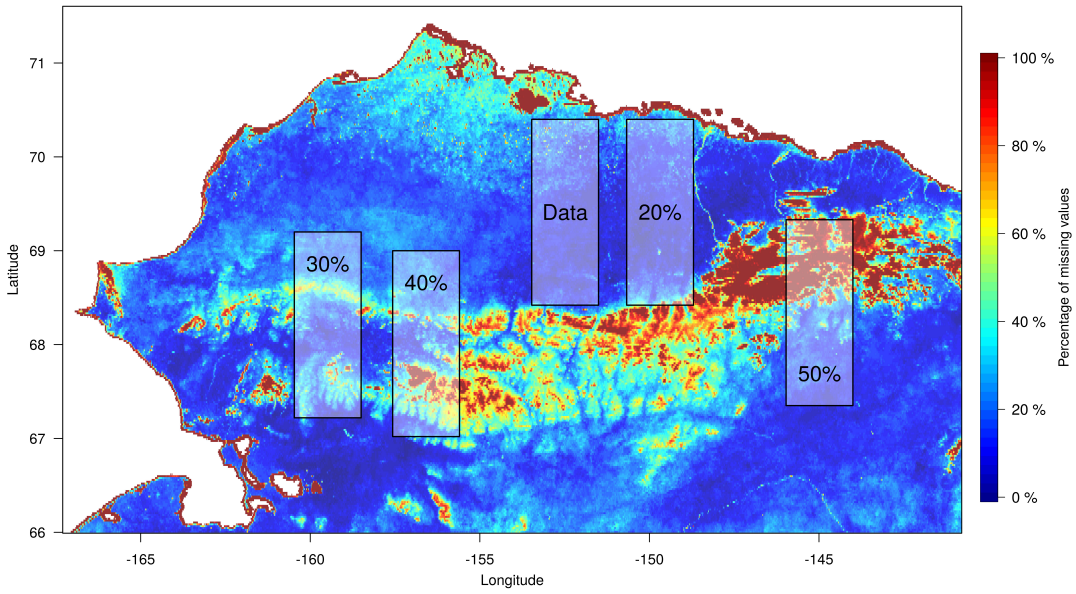


Fig. 2. Map of the study region of northern Alaska. Colors: percentage of missing values of all 48 considered time points. To construct test data sets, we consider the NDVI values of the region labeled with "Data," which has a spatial extent of 100×100 values. That subset exhibits a relatively few (about 12%) missing values. Test data sets with 20%, 30%, 40%, and 50% missing values are obtained by artificially removing values from the "Data" region according to the distribution of missing values observed at the regions labeled with "20%," "30%," "40%," and "50%," respectively. Note that the regions are shown as rectangles, as opposed to squares, because of the chosen geographic map projection.

271 819 values. 3 696 691 (28%) of the $271\,819 \cdot 8 \cdot 6 \approx 13 \cdot 10^6$ values in that data set are missing. The data set is shown in Fig. S13 in the supplementary material.

C. Competitors

In addition to the performance measurements of the proposed method referred to as *gapfill*, we also compare the accuracy of its predicted values against those from two competing prediction methods. The first competitor is presented in [28] and uses two different prediction algorithms depending on the amount of missing values. We choose this method, because it is a relatively new and promising spatio-temporal prediction method, which is available as open-source Python and C code (github.com/malaria-atlas-project/modis-gapfilling). We use a code version downloaded on August 15, 2015 (git commit c83776c) and refer to that software as "gapfill-MAP."

The second competitor, the TIMESAT software, is chosen, because it is well established. The main purpose of TIMESAT is to smooth time-series of remote sensing data and to estimate seasonal parameters. TIMESAT treats the values of each sampled location separately and, hence, does not exploit the spatial dependence in the data. Several authors report that the smoothed time-series from TIMESAT can be used to predict missing values [9], [56], [57]. TIMESAT is a closed-source software implemented in Fortran and comes with a MATLAB [58] interface featuring a graphical user interface. The software (version 3.2) and its documentation are available at web.nateko.lu.se/timesat/.

gapfill, gapfill-MAP, and TIMESAT have several tuning parameters, which influence the prediction process and the accuracy of the predictions. Although we tried to find good parameter configurations for each competitor, results may improve with other settings. Nevertheless, the presented comparisons give a solid overview of the performance.

For *gapfill*, the chosen tuning parameters are $\lambda_x = \lambda_y = 10$, $\lambda_s = 1$, $\lambda_s = 5$, $\theta_1 = 5$, $\theta_2 = 25$, and $\nu = 2$. An example R-code showing how to execute the *gapfill* software is available in Listing S1 in the supplementary material. gapfill-MAP has 16 tuning parameters, which are difficult to interpret because of missing documentation. Important parameters are those controlling the search of informative values and the "despeckle" algorithm (see the Python code for more information). We used the parameters given in Listing S2 in the supplementary material. For TIMESAT, the tuning parameters are described in the software manual [59]. We choose to fit a "double logistic" smoothing function, which is recommended for NDVI values in high latitudes with many missing values [60] and which provides the most satisfying results for our test data sets. The complete configuration file is shown in Listing S3 in the supplementary material.

IV. Results

A. Prediction Accuracy

We apply the proposed prediction method to the test data sets described in Section III-B. It returns predictions for all missing values in all test data sets. Images of the resulting completed data sets are shown in Figs. S6–S9 in the supplementary material and Fig. 3 (left). A visual examination of the predicted values reveals that they have the expected spatial distribution, including small-scale features such as, e.g., the band of low NDVI values crossing the images from the west to the east. Moreover, the predicted values do not show any artificially introduced patterns. Time-series for three spatial locations of the test data set with 40% missing values are shown

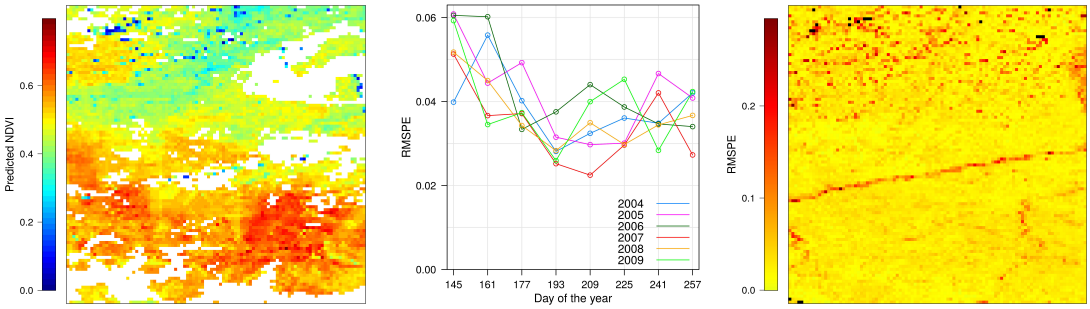


Fig. 3. Predictions and accuracy measurements for the test data set with 40% missing values. (Left) Predicted NDVI values for the day 177 of the year 2006. For that image, 2 335 of 10 000 values were observed (shown in white). (Middle) RMSPE for the indicated dates. (Right) Spatial distribution of the RMSPE. 19 RMSPE values are missing (shown in black), because the corresponding locations have observed values at all considered time points. Note that the left and right images have an extent of 100×100 values and each value corresponds to a pair of latitude (y-axis) and longitude (x-axis) coordinates.

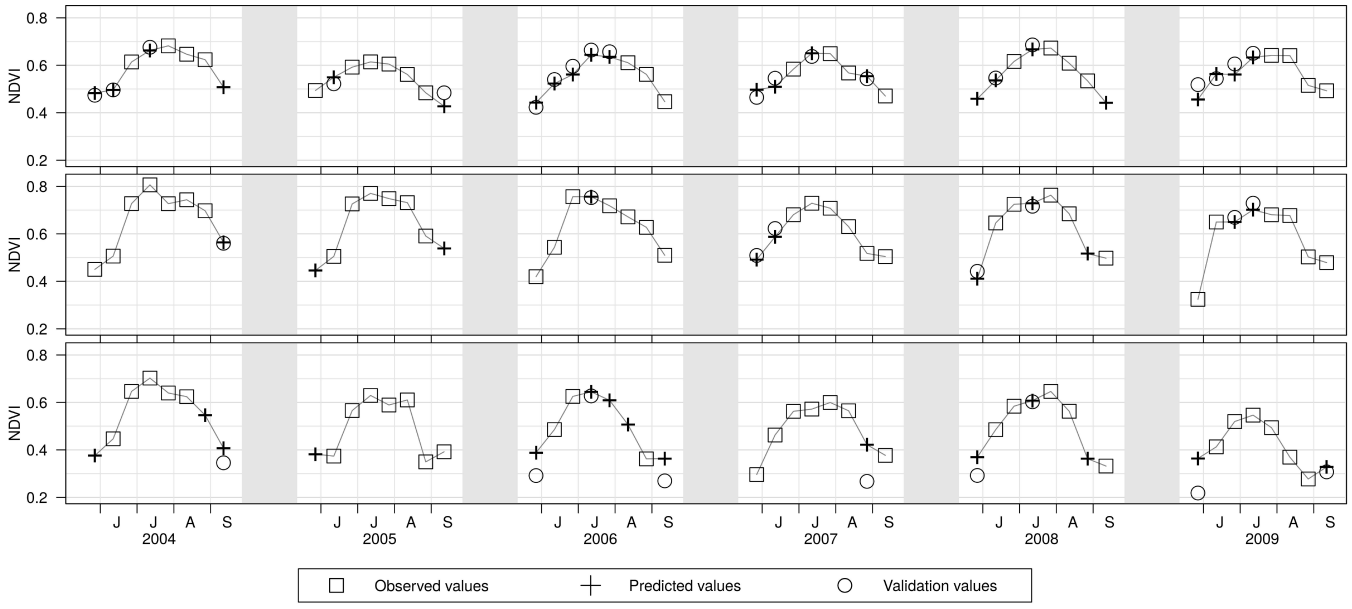


Fig. 4. Time-series for three spatial locations of the test data set with 40% missing values. The spatial locations is selected, such that they represent locations with (top) large, (middle) average, and (bottom) low NDVI values. If validation values are available, they are displayed with circles. The gray areas represent the time span from mid-September to mid-May with low vegetation activity.

in Fig. 4. The spatial locations are selected such that typical time-series with large, average, and low NDVI values result. The predicted values follow the expected seasonal curve and match the validation values well. Low NDVI values are more difficult to predict [see Fig. 4 (bottom)] as they have larger uncertainties (see Fig. S1 in the supplementary material).

To assess the temporal variation of the prediction accuracy, the mean RMSPEs per time point for the test data set with 40% missing values are shown in Fig. 3 (middle). The figure shows that the RMSPEs are larger for early DOYs. This is in accordance with the observation that values at the beginning of the season are more likely to be missing and exhibit larger variability compared with values observed at the end of the season (see Fig. S1 in the supplementary material). Fig. 3 (right) shows the spatial distribution of the mean RMSPEs, which resembles the spatial distribution of the temporal variation in the data (see Fig. S1 (top-right) in the supplementary material). This is expected, because values observed at locations with large temporal variability in the NDVI values are more difficult to predict.

Another way to study prediction accuracy is to plot the validation values against the predicted values, as shown in Fig. 5 (top). Most of the validation values are between 0.3 and 0.8. In that interval, the predicted values are scattered around the diagonal (red line) indicating that they are near the validation values on average. Predictions for values below 0.3 have lower accuracy. This is in accordance with the observation that those values tend to have larger variance over time and are, therefore, more difficult to predict. As expected, the deviation of the predicted values from the validation values increases with larger percentages of missing values. This can also be seen in Fig. 5 (bottom), where the histograms of the prediction errors (predicted minus validation values) show a wider distribution with increasing percentages of missing values. While the medians of the error differences are located at zero, the distributions of the differences are positively skewed (skewness between 2.17 and 2.3), which reflects an increased prediction uncertainty for low NDVI values. The standard deviation of the distributions of differences is between 0.041 and 0.042 for the test data set with 20%, 30%, and 40% missing values,

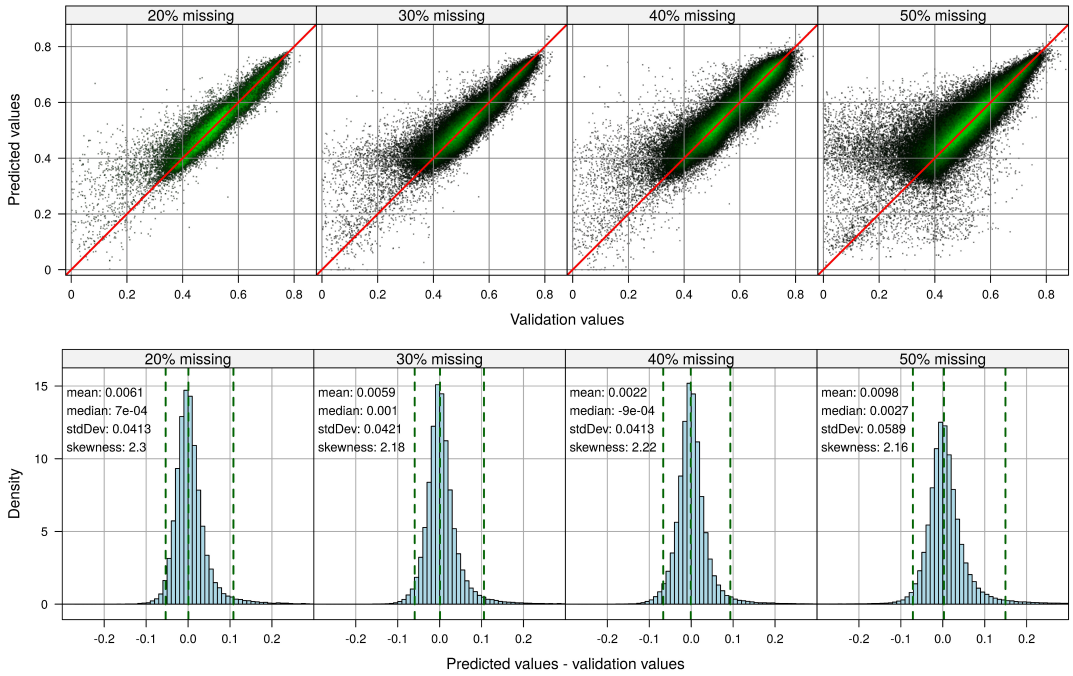


Fig. 5. Accuracy of the *gapfill* predictions for the four test data sets. (Top) Scatter plots of the predicted values (y-axis) and the validation values (x-axis). Green color shading: regions with a large density of points. Light green color shading: 100 overlaying points. (A similar figure for the gapfill-MAP and the TIMESAT method is given in Fig. S16 in the supplementary material.) (Bottom) Histograms of the differences between the predicted and the validation values. Dashed lines: 2.5%, 50%, and 97.5% quantiles.

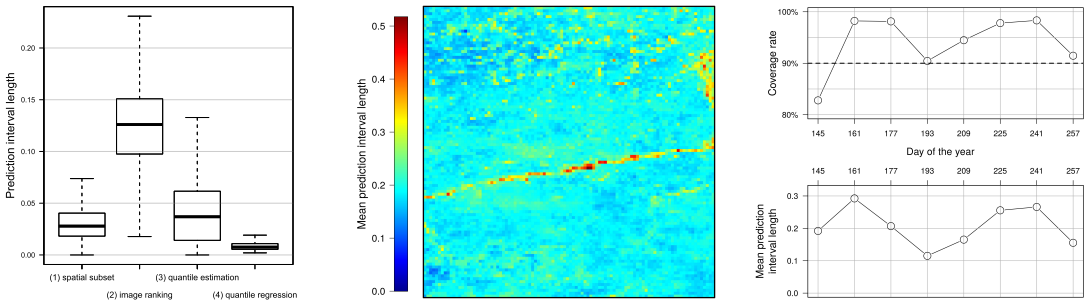


Fig. 6. (Left) Uncertainty contribution from the four steps of the prediction method. (Middle) Spatial distribution of the mean 90% prediction interval width for the test data set with 40% missing values. (Right) Corresponding coverage rates and mean prediction interval widths per DOY.

and increases to 0.059 for the test data set with 50% missing values. This increase could be due to an increased amount of missing low NDVI values in the test data set with 50% missing values.

In addition, *gapfill* is applied to the 48 images covering the entire spatial region of northern Alaska (see Fig. 2). To predict the $\approx 3.7 \times 10^6$ missing values of that data set, 80 cores of an Intel Xeon CPU E7-2850 at 2 GHz run for 10 h. The images of the predicted values are shown in Fig. S14 in the supplementary material. Again, all missing values are predicted and a visual inspection of the data does not reveal any artificially introduced patterns.

B. Uncertainty Assessment

The widths of the prediction intervals, corresponding to the four main steps of the prediction method, summarize their uncertainty contributions. Fig. 6 (left) shows the summary statistics of these widths as boxplots, revealing that the sorting step 2 (Pseudocode 1) introduced the largest uncertainties, followed by the estimation of the quantile of step 3 (Pseudocode 2). To investigate the properties of the 90% prediction interval combining the uncertainties from steps 2 and 3, the spatial distribution of the mean prediction interval widths is shown in Fig. 6 (middle). It exhibits a similar spatial distribution as the standard deviation estimated for each spatial location of the test data set (see Fig. S1 (top-right) in the supplementary material) and the spatial distribution of the average RMSPEs [see Fig. 3 (right)]. Since the seasonal variability of the prediction interval widths is larger, compared with the interannual variability, we only show the former in Fig. 6 (bottom-right). It has a U-shape, which is also observed in the distributions of the missing values (see Fig. S1 in the supplementary material) with some deviations early and late in the season, i.e., the values of DOY 145 and 257. These deviations might be caused by the fact that we only consider a part of the seasonal cycle, and, hence, have less information at the boundaries thereof. The overall coverage rate of the prediction interval for that

TABLE II

PREDICTED VALUES OF THE FOUR TEST DATA SETS OBTAINED WITH *gapfill*, GAPFILL-MAP, AND TIMESAT ARE SUMMARIZED IN TERMS OF THE NUMBER AND PERCENTAGE OF SUCCESSFULLY PREDICTED VALUES AND THE RMSPE $\times 10^3$. TO GET COMPARABLE RESULTS, THE RMSPEs OF *gapfill* ARE ALSO GIVEN FOR THE SUBSETS OF SUCCESSFULLY PREDICTED VALUES FROM GAPFILL-MAP ($\text{RMSPE}_{\text{MAP}}$) AND TIMESAT ($\text{RMSPE}_{\text{T}}$)

	gapfill				gapfill-MAP		TIMESAT	
	#predicted	RMSPE	$\text{RMSPE}_{\text{MAP}}$	RMSPE_{T}	#predicted	RMSPE	#predicted	RMSPE
20%	92'822 (100%)	41.80	42.06	41.10	90'307 (97%)	45.00	59'948 (65%)	83.43
30%	147'827 (100%)	42.54	42.39	37.09	146'686 (99%)	45.54	42'892 (29%)	71.43
40%	192'456 (100%)	41.34	40.98	36.41	169'998 (88%)	42.49	31'279 (16%)	71.93
50%	240'326 (100%)	59.58	44.94	37.24	134'540 (56%)	45.61	14'127 (6%)	86.09

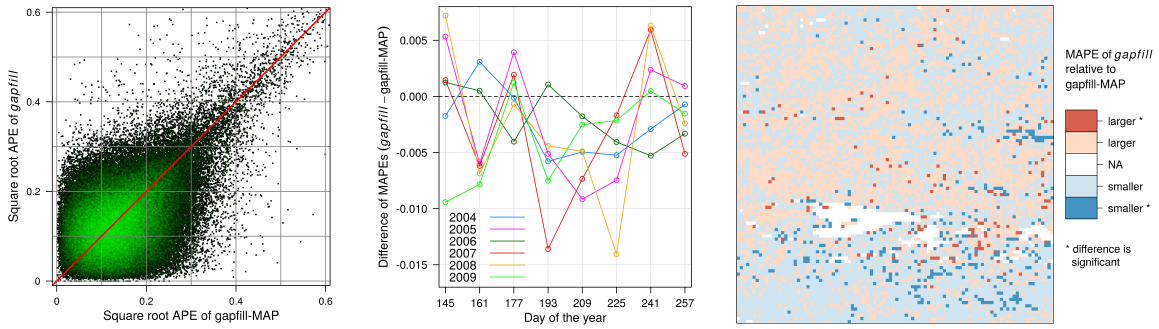


Fig. 7. Comparison of the APEs of the predictions of *gapfill* and gapfill-MAP for the test data set with 40% missing values. (Left) Scatter plot of the square root APEs of *gapfill* (y-axis) and gapfill-MAP (x-axis). The green color shading indicates regions with a large density of points (light green color shading corresponds to 25 overlaying points). (Middle) Difference of the MAPEs of *gapfill* and gapfill-MAP for the indicated dates. (Right) Spatial comparison of the MAPEs. Colors indicate for each of the 100×100 spatial locations of the test data set whether the MAPE of *gapfill* or gapfill-MAP is larger. Dark colors indicate that the differences are significant (Wilcoxon tests, $\alpha = 0.05$). For the 385 values shown white, no comparison was made, because all values for the corresponding locations were observed or gapfill-MAP did not return predictions for them.

data set is 93%, i.e., 93% of the validation values lie within the corresponding prediction intervals. Hence, the prediction uncertainty is slightly overestimated on average. The average coverage rate per DOY is shown in Fig. 6 (right).

C. Comparison With Gapfill-MAP and TIMESAT

First, we compare for all three methods the number of successfully predicted values and their percentages relative to the numbers of missing values in the test data sets (column "# predicted" of Table II). While *gapfill* returned predictions for all missing values of the four test data sets, gapfill-MAP and TIMESAT partially returned NAs as predicted values. The numbers of NAs in the predictions seem to increase with the numbers of NAs in the test data sets and they represent a considerable proportion (up to 94%) of the predictions from TIMESAT. The large amount of missing values in the TIMESAT predictions might be explained by the uneven spatial distribution of missing values in the test data sets, which implies that TIMESAT had to process some time-series with a large proportion of missing values. Another difficulty for TIMESAT might be that the test data sets only contain 8 of the 24 time points of the seasonal cycle, because the quality of the values from the remaining 16 winter time points was too low.

The RMSPEs for *gapfill* are between 41.34×10^{-3} and 42.54×10^{-3} for the test data sets with 20%, 30%, and 40% missing values, and increase to 59.58×10^{-3} for the test data set with 50% missing values (see Table II). This increase can be explained by the spatio-temporal distribution of missing values, which is different in all test data sets and impacts the difficulty of the prediction task. gapfill-MAP and TIMESAT have more problems in predicting the missing values of the test data set with 50% missing values, as indicated by the decrease in the proportion of successfully predicted values (see Table II). The prediction accuracy in terms of the RMSPE can only be calculated for successfully predicted values. To make the RMSPEs of gapfill-MAP and TIMESAT comparable to those from *gapfill*, we calculated the RMSPEs of *gapfill* relative to the subsets of successfully predicted values from gapfill-MAP and TIMESAT, and denote them with $\text{RMSPE}_{\text{MAP}}$ and RMSPE_{T}, respectively. According to the RMSPEs given in Table II, the *gapfill* predictions are the most accurate ones for all test data sets. The MAPEs show a similar pattern as the RMSPEs (Table S1 in the supplementary material). Both the RMSPEs and MAPEs of *gapfill* are significantly smaller than those from gapfill-MAP and TIMESAT for all test data sets (Wilcoxon tests [61], all p-values $< 10^{-15}$).

While the RMSPEs of *gapfill* are close to those of gapfill-MAP, the RMSPEs of TIMESAT are about two times higher compared with those of *gapfill* and gapfill-MAP. We, therefore, restrict the following comparison to the *gapfill* and gapfill-MAP methods and investigate their predicted values for the test data set with 40% missing values in more detail. Fig. 7 (left) shows a scatter plot of the square root absolute prediction errors (APEs) of *gapfill* (y-axis) and gapfill-MAP (x-axis). The square root transformation was chosen to facilitate the visual inspection of small differences. In Fig. 7, the scattering seems to be symmetric around the diagonal line and no clear pattern discriminates the methods. In Fig. 7 (middle), the differences

of the MAPEs of *gapfill* and gapfill-MAP are shown for all time points, i.e., each point in that middle corresponds to the difference between the MAPEs of *gapfill* and gapfill-MAP of one image. For 33 of the 48 time points, the *gapfill* method performed better, but no clear temporal pattern can be detected. Finally, a spatial comparison of the MAPEs is given in Fig. 7 (right). The colors for each of the 100×100 spatial locations of the test data set indicate whether the MAPEs of *gapfill* are larger or smaller compared with those of gapfill-MAP. For Fig. 7, the MAPE of a spatial location is calculated from the APEs of that location and all time points. Dark colors indicate that the differences are significant (Wilcoxon tests, $\alpha = 0.05$). For 372 (72.2%) of the total 508 values exhibiting significant differences, the *gapfill* method performed better. The *gapfill* predictions perform especially well in the southern region, where there is a higher concentration of significantly smaller *gapfill* MAPEs. However, the MAPEs of both methods are more similar in other regions. This could be due to the accumulation of missing values in the southern region (see Fig. S4 in the supplementary material). Images of the complete spatio-temporal distribution of the APEs of *gapfill* and gapfill-MAP as well as images of their difference are given in Figs. S10–S12 in the supplementary material.

V. Conclusion

The newly proposed prediction method for missing values in remote sensing data sets is convincing in many respects. First, the method was able to predict all missing values in data sets with large proportions of missing values (up to 50%). Second, the predicted values reconstruct the spatial and temporal patterns of NDVI data sets with many details. In realistic test data sets with known validation values, the method has returned accurate predictions with low RMSPEs and MAPEs. Third, in comparison with two established prediction methods (gapfill-MAP and TIMESAT), the proposed method had significantly lower RMSPEs for all considered test scenarios (Wilcoxon tests, all $p < 10^{-15}$). Fourth, the method provides prediction intervals, which quantify the prediction uncertainties based on statistical considerations.

When developing the method and its software implementation, a key priority was usability among practitioners working with remote sensing data. On the one hand, this influenced methodological choices in favor of techniques that are computationally efficient and parallelizable. On the other hand, the focus on potential users motivated us to develop a software implementation with several advantageous features. The software is available as open-source R package *gapfill* guaranteeing maximum transparency and making it easy to use and further develop the method. Moreover, *gapfill* contains C++ source code and relies on parallel computing infrastructure from the R package *foreach* enabling the users to fully exploit the available computing resources. All functions of *gapfill* are accompanied with documentation and unit tests to make them trustworthy.

We show the use of the proposed method for the prediction of missing values in MODIS NDVI data sets of the Arctic region. However, its use is not restricted to Arctic NDVI data as, e.g., shown by the validation experiment based on an MODIS NDVI data set from the Amazon region (see Supplementary Material). One of the current limitations is that the input data set has to be sampled on a regular grid in space and time. But such data sets are indeed common within and outside the field of remote sensing, and hence, for plenty of application types, *gapfill* may be a beneficial alternative to other methods.

Acknowledgment

The authors would like to thank Dr. H. Gibson for sharing the gapfill-MAP software and for answering related questions. They would also like to thank Dr. E. Furrer for the constructive discussions about the presentation of the material. They would also like to thank two anonymous reviewers for their comments and suggestions, which helped to improve the quality of the manuscript significantly.

References

[1] D. A. Stow *et al.*, "Remote sensing of vegetation and land-cover change in Arctic Tundra ecosystems," *Remote Sens. Environ.*, vol. 89, no. 3, pp. 281–308, 2004.

[2] G. P. Asner, "Cloud cover in Landsat observations of the Brazilian Amazon," *Int. J. Remote Sens.*, vol. 22, no. 18, pp. 3855–3862, 2001.

[3] G. Hmimina *et al.*, "Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements," *Remote Sens. Environ.*, vol. 132, pp. 145–158, May 2013.

[4] I. Garonna, R. de Jong, and M. E. Schaepman, "Variability and evolution of global land surface phenology over the past three decades (1982–2012)," *Global Change Biol.*, vol. 22, no. 4, pp. 1456–1468, 2016.

[5] M. A. White *et al.*, "Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006," *Global Change Biol.*, vol. 15, no. 10, pp. 2335–2359, 2009.

[6] J. V. Buttlar, J. Zscheischler, and M. D. Mahecha, "An extended approach for spatiotemporal gapfilling: Dealing with large and systematic gaps in geoscientific datasets," *Nonlinear Process. Geophys.*, vol. 21, no. 1, pp. 203–215, 2014.

[7] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 106–115, 2010.

[8] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor, "Phenological change detection while accounting for abrupt and gradual trends in satellite image time series," *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2970–2980, 2010.

[9] A. Verger, F. Baret, M. Weiss, S. Kandasamy, and E. Vermote, "The CACAO method for smoothing, gap filling, and characterizing seasonal anomalies in satellite time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1963–1972, Apr. 2013.

[10] S. Jamali, P. Jönsson, L. Eklundh, J. Ardö, and J. Seaquist, "Detecting changes in vegetation trends using time series segmentation," *Remote Sens. Environ.*, vol. 156, pp. 182–195, Jan. 2015.

[11] G. J. Roerink, M. Menenti, and W. Verhoef, "Reconstructing cloudfree NDVI composites using Fourier analysis of time series," *Int. J. Remote Sens.*, vol. 21, no. 9, pp. 1911–1917, 2000.

[12] Á. Moreno, F. J. García-Haro, B. Martínez, and M. A. Gilabert, "Noise reduction and gap filling of fAPAR time series using an adapted local regression filter," *Remote Sens.*, vol. 6, no. 9, pp. 8238–8260, 2014.

[13] A. Verger, F. Baret, and M. Weiss, "Near real-time vegetation monitoring at global scale," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3473–3481, Aug. 2014.

[14] M. Neteler, "Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data," *Remote Sens.*, vol. 2, no. 1, pp. 333–351, 2010.

[15] J. P. W. Scharlemann *et al.*, "Global data for ecology and epidemiology: A novel algorithm for temporal Fourier processing MODIS data," *PLoS ONE*, vol. 3, no. 1, p. e1408, 2008.

[16] P. Jönsson and L. Eklundh, "TIMESAT—A program for analyzing time-series of satellite sensor data," *Comput. Geosci.*, vol. 30, no. 8, pp. 833–845, 2004.

[17] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh, "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter," *Remote Sens. Environ.*, vol. 91, nos. 3–4, pp. 332–344, 2004.

[18] F. Gao *et al.*, "An algorithm to produce temporally and spatially continuous MODIS-LAI time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 60–64, Jan. 2008.

[19] R. R. Colditz, C. Conrad, T. Wehrmann, M. Schmidt, and S. Dech, "TiSeG: A flexible software tool for time-series generation of MODIS data utilizing the quality assessment science data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 3296–3308, Oct. 2008.

[20] Y. Zhu, E. L. Kang, Y. Bo, Q. Tang, J. Cheng, and Y. He, "A robust fixed rank Kriging method for improving the spatial completeness and accuracy of satellite SST products," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5021–5035, Sep. 2015.

[21] E. A. Addink, "A comparison of conventional and geostatistical methods to replace clouded pixels in NOAA-AVHRR images," *Int. J. Remote Sens.*, vol. 20, no. 5, pp. 961–977, 1999.

[22] R. E. Rossi, J. L. Dungan, and L. R. Beck, "Kriging in the shadows: Geostatistical interpolation for remote sensing," *Remote Sens. Environ.*, vol. 49, no. 1, pp. 32–40, 1994.

[23] J. Chen, X. Zhu, J. E. Vogelmann, F. Gao, and S. Jin, "A simple and effective method for filling gaps in landsat ETM+ SLC-off images," *Remote Sens. Environ.*, vol. 115, no. 4, pp. 1053–1064, 2011.

[24] X. Zhu, D. Liu, and J. Chen, "A new geostatistical approach for filling gaps in Landsat ETM+ SLC-off images," *Remote Sens. Environ.*, vol. 124, pp. 49–60, 2012.

[25] J. M. Beckers and M. Rixen, "EOF calculations and data filling from incomplete oceanographic datasets," *J. Atmos. Ocean. Technol.*, vol. 20, no. 12, pp. 1839–1856, 2003.

[26] L. Poggio, A. Gimona, and I. Brown, "Spatio-temporal MODIS EVI gap filling under cloud cover: An example in Scotland," *ISPRS J. Photogramm. Remote Sens.*, vol. 72, pp. 56–72, Aug. 2012.

[27] F. Gerber. (2017). *Gapfill: Fill Missing Values in Satellite Data R Package Version 0.9.5-3*. [Online]. Available: https://CRAN.R-project.org/package=gapfill

[28] D. J. Weiss, P. M. Atkinson, S. Bhatt, B. Mappin, S. I. Hay, and P. W. Gething, "An effective approach for gap-filling continental scale remotely sensed time-series," *ISPRS J. Photogramm. Remote Sens.*, vol. 98, pp. 106–118, Dec. 2014.

[29] R. Lguensat, P. Tandeo, R. Fablet, and R. Garello, "Spatio-temporal interpolation of sea surface temperature using high resolution remote sensing data," in *Proc. Oceans-St. John's*, 2014, pp. 1–4.

[30] J. C. de Oliveira, J. C. N. Epiphanio, and C. D. Rennó, "Window regression: A spatial-temporal analysis to estimate pixels classified as low-quality in MODIS NDVI time series," *Remote Sens.*, vol. 6, no. 4, pp. 3123–3142, 2014.

[31] J. C. de Oliveira and J. C. N. Epiphanio, "Noise reduction in MODIS NDVI time series data based on spatial-temporal analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2012, pp. 2372–2375.

[32] N. Golyandina and E. Osipov, "The 'Caterpillar'-SSA method for analysis of time series with missing values," *J. Stat. Planning Inference*, vol. 137, no. 8, pp. 2642–2653, 2007.

[33] L. Guo, L. Lei, Z. C. Zeng, P. Zou, D. Liu, and B. Zhang, "Evaluation of spatio-temporal variogram models for mapping Xco_2 using satellite observations: A case study in China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 376–385, Jan. 2015.

[34] Z. Zeng, L. Lei, S. Hou, F. Ru, X. Guan, and B. Zhang, "A regional gap-filling method based on spatiotemporal variogram model of CO_2 columns," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3594–3603, Jun. 2014.

[35] D. P. Roy, J. S. Borak, S. Devadiga, R. E. Wolfe, M. Zheng, and J. Descloitres, "The MODIS land product quality assessment approach," *Remote Sens. Environ.*, vol. 83, nos. 1–2, pp. 62–76, 2002.

[36] N. B. Shah, S. Balakrishnan, A. Guntuboyina, and M. J. Wainwright, "Stochastically transitive models for pairwise comparisons: Statistical and computational issues," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 934–959, Feb. 2017.

[37] D. P. McMillen, *Quantile Regression for Spatial Data* (Springer Briefs in Regional Science). Berlin, Germany: Springer, 2012.

[38] R. Koenker, *Quantile Regression*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[39] R. Koenker. (2015). *Quantreg: Quantile Regression, R Package Version 5.05*. [Online]. Available: http://CRAN.R-project.org/package=quantreg

[40] R. Analytics and S. Weston. (2015). *Foreach: Foreach Looping Construct for, R Package Version 1.4.3*. [Online]. Available: http://CRAN.R-project.org/package=foreach

[41] OpenMP Architecture Review Board. (2016). *OpenMP Application Program Interface, Version 4.5*. [Online]. Available: http://www.openmp.org

[42] MPI Forum. (2016). *Message Passing Interface (MPI) Forum*. [Online]. Available: http://www.mpi-forum.org

[43] C. O. Justice *et al.*, "An overview of MODIS land data processing and product status," *Remote Sens. Environ.*, vol. 83, nos. 1–2, pp. 3–15, 2002.

[44] K. Didan, A. B. Munoz, R. Solano, and A. Huete. (2015). *MODIS Vegetation Index User's Guide (MOD13 Series) Version 3.00*. [Online]. Available: http://vip.arizona.edu/documents/MODIS/MODIS_VI_UsersGuide_June_2015_C6.pdf

[45] W. J. D. van Leeuwen, A. R. Huete, and T. W. Laing, "MODIS vegetation index compositing approach: A prototype with AVHRR data," *Remote Sens. Environ.*, vol. 69, no. 3, pp. 264–280, 1999.

[46] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote Sens. Environ.*, vol. 83, nos. 1–2, pp. 195–213, 2002.

[47] M. Mattiuzzi. (2015). *MODIS: MODIS Acquisition and Processing R Package Version 0.10-18*. [Online]. Available: http://R-Forge.R-project.org/projects/modis/

[48] M. J. Dwyer and G. L. Schmidt, "The MODIS reprojection tool," in *Earth Science Satellite Remote Sensing*, J. J. Qu, W. Gao, M. Kafatos, R. E. Murphy, and V. V. Salomonson, Eds. Berlin, Germany: Springer, 2006, pp. 162–177.

[49] R. J. Hijmans. (2015). *Raster: Geographic Data Analysis and Modeling, R Package Version 2.3-0*. [Online]. Available: http://CRAN.R-project.org/package=raster

[50] R. S. Bivand, E. J. Pebesma, and V. Gómez-Rubio, *Applied Spatial Data Analysis With R*, 2nd ed. New York, NY, USA: Springer, 2013. [Online]. Available: http://www.asdar-book.org/

[51] E. J. Pebesma and R. S. Bivand, "Classes and methods for spatial data in R," *R News*, vol. 5, no. 2, pp. 9–13, 2005. [Online]. Available: http://CRAN.R-project.org/doc/Rnews/

[52] D. Nychka, R. Furrer, and S. Sain. (2016). *Fields: Tools for Spatial Data, R Package Version 8.4-1*. [Online]. Available: http://CRAN.R-project.org/package=fields

[53] D. Sarkar, *Lattice Multivariate Data Visualization With R*. New York, NY, USA: Springer, 2008. [Online]. Available: http://lmdvr.r-forge.r-project.org

[54] H. Wickham, *ggplot2 Elegant Graphics for Data Analysis*. New York, NY, USA: Springer, 2009. [Online]. Available: http://had.co.nz/ggplot2/book

[55] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geosci. Model Develop.*, vol. 7, no. 3, pp. 1247–1250, 2014.

[56] P. M. Atkinson, C. Jeganathan, J. Dash, and C. Atzberger, "Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology," *Remote Sens. Environ.*, vol. 123, pp. 400–417, Aug. 2012.

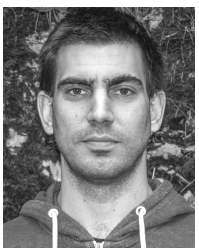
[57] J. N. Hird and G. J. McDermid, "Noise reduction of NDVI time series: An empirical comparison of selected techniques," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 248–258, 2009.

[58] *MATLAB Version 8.3.0 (R2014a)*, MathWorks, Inc., Natick, MA, USA, 2014.

[59] L. Eklundh and P. Jönsson. (2015). *TIMESAT 3.2 With Parallel Processing Software Manual*. [Online]. Available: http://web.nateko.lu.se/timesat/docs/ TIMESAT32_software_manual.pdf

[60] P. S. A. Beck, C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore, "Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 321–334, 2006.

[61] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.



Florian Gerber received the B.Sc. degree in mathematics from the University of Bern (UNIBE), Bern, Switzerland, in 2010, and the M.Sc. degree in biostatistics from the University of Zurich (UZH), Zürich, Switzerland, in 2013, where he is currently pursuing the Ph.D. degree with the Department of Mathematics under the supervision of Dr. R. Furrer.

During his studies, he was a Statistician with the Institute of Social and Preventive Medicine, UNIBE, and an External Statistical Consultant for a pharmaceutical company. Since 2013, he has been a Teaching Assistant in the Group of Dr. R. Furrer with the Department of Mathematics, UZH.

Rogier de Jong received the M.Sc. degree in earth science from Utrecht University, Utrecht, The Netherlands, and the Ph.D. degree in environmental science from Wageningen University, Wageningen, The Netherlands.

He joined the Remote Sensing Laboratories, University of Zurich, Zürich, Switzerland, as a Research Associate. Since 2015, he has been leading the Group on Remote Sensing of Dynamic Vegetation Systems. His research interests include remote sensing of vegetation, dynamics of ecosystems and land use, time-series analysis, and consumer-grade observation systems.

Michael E. Schaepman (M'05–SM'07) received the M.Sc. and Ph.D. degrees in geography from the University of Zurich (UZH), Zürich, Switzerland, in 1993 and 1998, respectively.

In 1999, he was a Postdoctoral Researcher with the Optical Sciences Center, The University of Arizona, Tucson, AZ, USA. In 2000, he was appointed as a Project Manager of an APEX Spectrometer with the European Space Agency. In 2003, he became the Full Chair of geoinformation science and remote sensing with Wageningen University, Wageningen, The Netherlands. In 2009, he was appointed as the Full Chair of remote sensing with UZH, where he is currently the Head of the Remote Sensing Laboratories, Department of Geography. He is the Director of the University Research Priority Program—Global Change and Biodiversity, and the Dean of the Faculty of Science. His research interests include computational earth sciences using remote sensing and physical models, with a particular focus on the land–atmosphere interface using imaging spectroscopy.

Gabriela Schaepman-Strub received the M.Sc. and Ph.D. degrees in geography from the University of Zurich (UZH), Zürich, Switzerland, in 1999 and 2004, respectively.

In 2001, she was a Visiting Scientist with Boston University, Boston, MA, USA. In 2003, she held a postdoctoral position at Wageningen University, Wageningen, The Netherlands. In 2009, she moved back to UZH as a Group Leader in spatial ecology and remote sensing. Her research interests include land surface–atmosphere interactions using remote sensing and radiative transfer models, with a particular focus on feedbacks of biodiversity changes to climate in the Arctic.

Reinhard Furrer received the Diploma degree in mathematics and the Ph.D. degree in statistics from the Swiss Federal Institute of Technology in Lausanne, Lausanne, Switzerland, in 1998 and 2002, respectively.

From 2002 to 2005, he was a Postdoctoral Fellow with the Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO, USA. From 2005 to 2009, he was an Assistant Professor with the Department of Mathematical and Computer Sciences, Colorado School of Mines, Golden, CO, USA. He is currently an Associate Professor with the Department of Mathematics and an Affiliate Faculty Member with the Department of Computational Science, University of Zurich, Zürich, Switzerland.

Dr. Furrer is a member of the American Statistical Society, the Institute of Mathematical Statistics, and the International Association of Mathematical Geology.