Geostatistical disaggregation of polygon maps of average crop yields by area-to-point kriging

Brus, D. J., Boogaard, H., Ceccarelli, T., Orton, T. G., Traore, S., & Zhang, M.

Geostatistical disaggregation of polygon maps of average crop yields by area-to-point kriging

D.J. Brus[a,1], H. Boogaard[b], T. Ceccarelli[b], T.G. Orton[c,d], S. Traore[e], M. Zhang[f]

[a]*Biometris, Wageningen University and Research, PO Box 16, 6700 AA Wageningen, the Netherlands*
[b]*Alterra, Wageningen University and Research, PO Box 32, 6700 AA Wageningen, the Netherlands*
[c]*The University of Queensland, School of Agriculture and Food Sciences, St Lucia, Queensland 4072, Australia*
[d]*Department of Environment and Science, Ecosciences Precinct, GPO Box 5078, Brisbane, Queensland 4001, Australia*
[e]*Centre Regional Aghrymet (Aghrymet) Boulevard de Université 425, 11011 Niamey, Niger*
[f]*Institute of Remote Sensing and Digital Earth - Chinese Academy of Science (IRSA), Beijing, 100101, China*

Abstract

Crop yield data are often available as statistics of areas, such as administrative units, generated by national agricultural surveys and censuses. This paper shows that such areal data can be used in area-to-point kriging (ATP kriging) to estimate the crop yield at the nodes of a fine grid that discretizes the study area, so that a more detailed map of the crop yield is obtained. The theory behind ATP kriging is explained, and illustrated with a one-dimensional simulation study and two real-world case studies. Vegetation, precipitation, temperature and soil data were used as potential covariates in the spatial trend part of the geostatistical model. ATP kriging requires the

*Corresponding author
Email address: dick.brus@wur.nl (D.J. Brus)

covariogram at point support, which can be recovered from the areal data by restricted maximum likelihood. The standard errors of the estimated variogram parameters can then be obtained by the Fisher information matrix. The average yields of only 17 administrative units in Shandong province (China) were not enough to obtain reliable estimates of the covariogram at point support. Also the ranges of the regional averages of the covariates were very narrow, so that the model must be extrapolated in the largest part of the study area. We were more confident about the covariogram parameters estimated from 45 provinces in Burkina Faso. We conclude that ATP kriging is an interesting method for disaggregation of spatially averaged crop yields. Contrary to other downscaling methods ATP kriging is founded on statistical theory, and consequently provides estimates of the precision of the disaggregated yields. Shortcomings are related to the uncertainty in the estimated covariogram parameters, as well as to the extrapolation of the model outside the range of the regional means of the covariates. Opportunities for future advancements are the use of modelled yields as covariates and the introduction of expert knowledge at different levels. For the latter a Bayesian approach to ATP kriging can be advantageous, introducing prior knowledge about the model parameters, as well as accounting for uncertainty about the model parameters.

Keywords:

yield gap, aggregated data, uncertainty

---

1. Introduction

Global change processes raise new estimation problems challenging conventional statistical methods. New problems require, for instance, recovering information from available aggregate agricultural statistics, and other available evidence, through disaggregation or downscaling methods (Fischer et al., 2006). There is a broad range of applications requiring such spatially downscaled statistics and foremost, crop area, yield or production data, which have been summarized for instance in You et al. (2014). These include food security, climate change, livestock production systems, technical change, ecosystem service valuation. For instance, in the context of yield gap analysis (van Ittersum et al., 2013) there is the need to evaluate the difference between actual yield (usually with reference to official statistics) and yield potential (usually obtained as the outcome of crop modelling). In general, applications generating spatially explicit gridded data respond to the need of adequately accounting for the geographical distribution of environmental, management and socio-economic conditions. This is regarded as a pre-requisite for more effective policies and interventions aimed at improving rural well-being, and for revealing untapped opportunities and shaping spatially-explicit responses to such opportunities (You et al., 2014).

For generating gridded maps Goerlich and Cantarino (2013) distinguish between 'bottom-up' and 'top-down' approaches. For a 'bottom-up' approach adequate individual georeferenced data must be available. In 'bottom-up' approaches for generating gridded estimates of crop production, the product of crop areas and yields, remote sensing techniques are increasingly used. As to yields, current methods include direct estimation of proxies to yields, such

as total biomass, vegetation indices and more complex yield indices. Indirect methods envisage for instance the assimilation of auxiliary variables derived from remote sensing in crop models. However, the ability of current methods for estimating crop yields is limited for many crops and geographies (Lobel, 2013).

In the case of 'top-down' methods only areal unit data are available and disaggregation techniques should be used. Most countries in the world have such aggregated data, but only at national and, at the most, at sub-regional levels. Production statistics are generated from national agricultural surveys and censuses. Their sampling frameworks however, usually limit the spatial units at which statistics can be reported within acceptable levels of statistical confidence. Therefore, a spatial disaggregation approach is sought which attempts to generate allocations of crop production at finer scales, possibly down to the scale of individual grid units. In other words, such methods try to resolve one of the major analytical weaknesses of regional and global agricultural studies, the inability to objectively downscale production statistics into spatial units such as agro-ecological zones or watersheds, and down to units (e.g. gridded products) having spatial resolutions finer than the original reporting units.

Spatial disaggregation methods that are relevant for our purposes include areal interpolation from simple area weighting to binary or poly-categorical dasymetric disaggregation, see Gallego et al. (2011) and Goerlich and Cantarino (2013), methods based on cross-entropy (You et al., 2014), statistical and geo-statistical methods, among which are kriging methods. The methods above have been applied to several application fields and variables, including

population densities and cropped area. However, with the exception of the cited cross-entropy method, there have been so far no applications related to the down-scaling of crop yields or production. Kriging methods that disaggregate observations of the spatial means of subareas into predictions at points are referred to as area-to-point (ATP) kriging methods. The use of spatially averaged data for spatial prediction of the values at points (e.g. the nodes of a fine discretisation grid covering the study area) by ATP kriging has received much interest in the scientific literature since 2000, see e.g. Gotway and Young (2002); Kyriakidis (2004); Kyriakidis and Yoo (2005); Gotway and Young (2007); Goovaerts (2008, 2011); Orton et al. (2012).

The predictions obtained with ATP kriging are coherent, also referred to as mass-preserving or pycnophylactic (Kyriakidis, 2004). This means that the average of point predictions within any arbitrary area with known spatial mean is equal to that spatial mean. This is a desirable property when the areal data can be assumed errorless observations of the spatial means, think for instance of the values of pixels of remotely sensed images.

The theory of ATP kriging is well established, and its potentials have been shown in many application areas, for instance in soil science (Schirrmann et al., 2012; Brus et al., 2014), spatial socio-economic studies (Nagle, 2010), disease mapping (Lin et al., 2014) and environmental health studies (VoPham et al., 2016). We are not aware of papers explaining how this statistical technique can be used for spatial disaggregation of polygon maps of average crop yields. Therefore, the aim of this paper is to draw the attention of agronomist to this technique, to explain the basics of ATP kriging as in a tutorial, and to illustrate it with a simulation study and two real-world case

studies.

2. Theory

As a stepping stone for explaining how values at points can be predicted from averages of blocks by ATP kriging, we first explain how values at points can be predicted from measurements at points by point kriging. Strictly speaking the points need not be infinitely small units but can also be small areas, think of pixels. What is essential in point kriging is that the size and geometry (referred to as the support) of the measurement units equals that of the prediction units.

2.1. Point kriging

In geostatistics the value of our variable of interest Z at a location $\mathbf{s}$ is modeled as the sum of the expected value, μ, and a random error (residual) at that location, $\epsilon(\mathbf{s})$:

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

The model is extended with a description of the probability distribution of the residuals. It is assumed that the residuals have a normal distribution with zero mean and a constant variance σ^2. Contrary to classical statistics, in geostatistics the residuals at any pair of locations are not assumed independent. The covariance of the residuals is modeled by a parametric function of the length (and direction) of the vector separating two locations.

A slightly more complicated model is obtained by replacing the expected value μ by a linear combination of covariates related to the variable of interest, think of remote sensing imagery such as a vegetation index, or rainfall

estimates:

$$Z(\mathbf{s}) = \sum_{k=0}^{p} \beta_k x_k(\mathbf{s}) + \epsilon(\mathbf{s}), \tag{2}$$

with β_k the regression coefficient for covariate x_k, $x_k(\mathbf{s})$ the value of covariate x_k at location $\mathbf{s}$, and p the number of covariates. By convention $x_0(\mathbf{s}) = 1$ so that β_0 is an intercept. In this model the expectation is not a constant, as before, but varies in space as the covariates show spatial variation. The non-constant expectation is referred to as the spatial trend. This trend component models the large-scale spatial structures. The small-scale spatial structure not accounted for by the spatial trend is modeled as a random effect, by the covariance of the residuals.

We note here that when we refer to a covariate as being on point support, we mean that it is extracted from a map of that covariate at a particular point. However, that map could itself represent some attribute at a larger spatial support. For instance, one covariate could come from a digital elevation model, produced on a 10-m grid, with the value for each pixel representing the average elevation over that grid cell, while another covariate could be related to climate, with a map available on a much coarser scale, each pixel of which would represent the average conditions within perhaps 5-km grid cells.

Using this model the value of the variable of interest at a target location $\mathbf{s}_0$ is predicted by

$$\hat{Z}(\mathbf{s}_0) = \sum_{k=0}^{p} \hat{\beta}_k x_k(\mathbf{s}_0) + \sum_{i=1}^{n} \lambda_i \left[Z(\mathbf{s}_i) - \sum_{k=0}^{p} \hat{\beta}_k x_k(\mathbf{s}_i) \right], \tag{3}$$

with $\hat{\beta}_k$ the estimated regression coefficient, n the number of sampling locations, and λ_i the weight attached to the residual at sampling location $\mathbf{s}_i$.

The first component of this predictor is the estimated expectation at the new location using the covariate values at this location and the estimated regression coefficients, and the second component is a weighted sum of the residuals at the sampling locations.

The question now is how to compute the weights λ_i. These weights are derived by minimizing the variance of the prediction error under the constraint that the prediction is unbiased. It can be shown that an unbiased prediction is obtained when the sum of the weights equals 1 ($\sum_{i=1}^{n} \lambda_i = 1$), and when for all p covariates the weighted sum of the covariate values at the sampling locations equals the covariate value at the target location ($\sum_{i=1}^{n} \lambda_i x_k(\mathbf{s}_i) = x_k(\mathbf{s}_0)$ for all $k = 1 \cdots p$). The constrained minimization problem can be redefined into an unconstrained minimization problem as follows. Each of the $q = p+1$ constraints mentioned above is multiplied by a constant. These constants, referred to as Lagrange multipliers, are unknown and must be estimated. The resulting terms are added to the variance of the prediction error, leading to a new minimization criterion. Setting the partial derivatives of this criterion with respect to the weights and the Lagrange multipliers to 0, leads to a set of $n + q$ equations (Webster and Oliver, 2007):

$$\begin{array}{l} \sum_{j=1}^{n} \lambda_j \operatorname{Cov}(\mathbf{s}_i, \mathbf{s}_j) + \sum_{k=0}^{p} \nu_k x(\mathbf{s}_j) = \operatorname{Cov}(\mathbf{s}_i, \mathbf{s}_0) \ , \quad i = 1, \ldots, n \\ \sum_{i=1}^{n} \lambda_i x_k(\mathbf{s}_i) = x_k(\mathbf{s}_0) \ , \quad k = 0, \ldots, p \end{array} \ , \qquad (4)$$

with $\operatorname{Cov}(\mathbf{s}_i, \mathbf{s}_j)$ the covariance between sampling points $\mathbf{s}_i$ and $\mathbf{s}_j$, $\operatorname{Cov}(\mathbf{s}_i, \mathbf{s}_0)$ the covariance between sampling point $\mathbf{s}_i$ and target point $\mathbf{s}_0$, and $\nu_k, k = 0, \ldots, p$ Lagrange multipliers. Note that the covariances are covariances of residuals, i.e. of the data minus the spatial trend component (Eq. 2). It is

convenient to represent this system of equations in matrix notation:

$$\begin{bmatrix} \mathbf{C_{ss}} & \mathbf{X_s} \\ \mathbf{X_s^T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} \mathbf{c_{s,s_0}} \\ \mathbf{x_{s_0}} \end{bmatrix}, \tag{5}$$

with $\mathbf{C_{ss}}$ the $n \times n$ matrix of residual covariances between the sampling points, $\mathbf{c_{s,s_0}}$ the $n \times 1$ vector of residual covariances between the sampling points and the target point, $\mathbf{X_s}$ the $n \times q$ matrix of covariate values at the sampling points, $\mathbf{x_{s_0}}$ the $q \times 1$ vector of covariate values at the target point, $\mathbf{0}$ the $q \times q$ matrix of zeroes, $\boldsymbol{\lambda}$ the $n \times 1$ vector with weights, and $\boldsymbol{\nu}$ the $q \times 1$ vector with Lagrange multipliers.

2.2. Area-to-point kriging

Now we consider the case where the observations consist of averages of blocks. These block averages, $\bar{Z}(\mathcal{B}_i)$ for block $\mathcal{B}_i$ of the m data blocks, are defined in terms of a point-support variable, $Z(\mathbf{s})$, by

$$\bar{Z}(\mathcal{B}_i) = \frac{1}{|\mathcal{B}_i|} \int_{\mathbf{s} \in \mathcal{B}_i} Z(\mathbf{s}) \mathrm{d}\mathbf{s}, \tag{6}$$

with $|\mathcal{B}_i|$ the surface area of block $\mathcal{B}_i$. Combining this definition with that of the point-support variable (Eq. 2), a statistical model can be written for the areal-support data

$$\bar{Z}(\mathcal{B}_i) = \frac{1}{|\mathcal{B}_i|} \int_{\mathbf{s} \in \mathcal{B}_i} \sum_{k=0}^{p} \beta_k x_k(\mathbf{s}) + \epsilon(\mathbf{s}) \mathrm{d}\mathbf{s}. \tag{7}$$

This expression can be re-arranged to give

$$\bar{Z}(\mathcal{B}_i) = \sum_{k=0}^{p} \beta_k \bar{x}_k(\mathcal{B}_i) + \bar{\epsilon}(\mathcal{B}_i), \tag{8}$$

where $\bar{x}_k(\mathcal{B}_i) = \frac{1}{|\mathcal{B}_i|}\int_{\mathbf{s}\in\mathcal{B}_i} x_k(\mathbf{s})\mathrm{d}\mathbf{s}$ is the block $\mathcal{B}_i$ average of the covariate x_k, and $\bar{\epsilon}(\mathcal{B}_i) = \frac{1}{|\mathcal{B}_i|}\int_{\mathbf{s}\in\mathcal{B}_i} \epsilon(\mathbf{s})\mathrm{d}\mathbf{s}$ the block average of the residuals. Recall that the distribution of the residuals on point support was assumed normal, with zero mean, constant variance σ^2 and with covariance between any pair of locations modelled by a parametric function of their separation vector. Based on this assumption, the statistical properties of $\bar{\epsilon}(\mathcal{B}_i)$ can be derived. It too is normally distributed with mean zero. The covariance for $\bar{\epsilon}(\mathcal{B}_i)$ and $\bar{\epsilon}(\mathcal{B}_j)$ is (Kyriakidis and Yoo, 2005)

$$\mathrm{Cov}(\bar{\epsilon}(\mathcal{B}_i), \bar{\epsilon}(\mathcal{B}_j)) = \frac{1}{|\mathcal{B}_i||\mathcal{B}_j|}\int_{\mathbf{s}\in\mathcal{B}_i}\int_{\mathbf{t}\in\mathcal{B}_j} C(\mathbf{s},\mathbf{t})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t}, \tag{9}$$

where $C(\mathbf{s},\mathbf{t})$ is the point-support covariance function applied for points $\mathbf{s}$ and $\mathbf{t}$ that sweep blocks $\mathcal{B}_i$ and $\mathcal{B}_j$, respectively. The covariance for $\mathcal{B}_i$ and $\mathcal{B}_i$ gives the variance; note that this variance is not the same for each block, due to their different sizes and geometries. Also, the covariance between a point-support variable and areal-support variable can be calculated as (Kyriakidis and Yoo, 2005)

$$\mathrm{Cov}(\epsilon(\mathbf{s}), \bar{\epsilon}(\mathcal{B}_j)) = \frac{1}{|\mathcal{B}_j|}\int_{\mathbf{t}\in\mathcal{B}_j} C(\mathbf{s},\mathbf{t})\mathrm{d}\mathbf{t}. \tag{10}$$

In practice, all of the above integrals can be approximated by summations over a large number of points that discretize the blocks. For example,

$$\bar{x}_k(\mathcal{B}_i) = \frac{1}{|\mathcal{B}_i|}\int_{\mathbf{s}\in\mathcal{B}_i} x_k(\mathbf{s})\mathrm{d}\mathbf{s} \approx \frac{1}{L_i}\sum_{l=1}^{L_i} x_k(\mathbf{s}_{i,l}), \tag{11}$$

where $\mathbf{s}_{i,l}$ is the i^{th} of the L_i points that discretize block $\mathcal{B}_i$. Eqs 7 and 8 provide a model to relate point-support values of the covariates to block-support data of the primary variable.

The value at the target point-location is predicted by

$$\hat{Z}(\mathbf{s}_0) = \sum_{k=0}^{p} \hat{\beta}_k x_k(\mathbf{s}_0) + \sum_{i=1}^{m} \lambda_i \left[\bar{Z}(\mathcal{B}_i) - \sum_{k=0}^{p} \hat{\beta}_k \bar{x}_k(\mathcal{B}_i) \right], \tag{12}$$

with m the number of observed blocks. The first component is equal to the first component of the point kriging predictor, but the second component differs. It equals a weighted average of the m residuals of the blocks. These residuals are computed as the difference between the observed means and a linear combination of the averages of the predictors of the blocks. Optimal weights are obtained by minimizing the variance of the prediction error and requiring that the prediction must be unbiased. Unbiasedness is now guaranteed by constraining the sum of the weights sum to 1 and so that $\sum_{i=1}^{n} \lambda_i \bar{x}_k(\mathcal{B}_i) = x_k(\mathbf{s}_0)$ for all $k = 1 \cdots p$. The optimal weights can be obtained by solving

$$\begin{bmatrix} \overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}} & \overline{\mathbf{X}}_{\mathcal{B}} \\ \overline{\mathbf{X}}_{\mathcal{B}}^{\mathrm{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{c}}_{\mathcal{B}\mathbf{s}_0} \\ \mathbf{x}_{\mathbf{s}_0} \end{bmatrix}, \tag{13}$$

with $\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}$ the $m \times m$ matrix of average residual covariances between the blocks, $\overline{\mathbf{c}}_{\mathcal{B}\mathbf{s}_0}$ the $m \times 1$ vector of average residual covariances between the blocks and the target point, and $\overline{\mathbf{X}}_{\mathcal{B}}$ the $m \times q$ matrix of average covariate values for the blocks. Comparing Eqs. 13 and 5 shows that the point-wise covariances in Eq. 5 have been replaced by average covariances between the blocks and the target point in Eq. 13. In practice the average covariance between a block $\mathcal{B}_i$ and the target point $\mathbf{s}_0$ can be approximated by selecting a large number, say S, points fully randomly from the block $\mathcal{B}_i$, computing the covariance between each of the S points and the target point, and averaging the S covariances at point support. The average covariance between two

blocks can be approximated similarly, by random selection of S points from both blocks, and forming S pairs of points. This shows that we must know the covariogram (covariance function) or variogram at point support. The next subsection explains how this covariogram can be estimated from the areal data.

Given the covariogram at point support, the variance of the prediction error equals

$$\mathrm{Var}\left(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right) = \mathrm{Cov}(0) - \boldsymbol{\lambda}^{\mathrm{T}}\overline{\mathbf{c}}_{\mathcal{B}\mathbf{s}_0} - \boldsymbol{\lambda}^{\mathrm{T}}\overline{\mathbf{X}}_{\mathcal{B}}\,\boldsymbol{\nu}, \tag{14}$$

with $\mathrm{Cov}(0)$ the covariance at distance 0, referred to as the *a priori* variance.

2.3. Estimating the covariogram parameters and regression coefficients

A hurdle in ATP kriging is the calibration of the covariogram (or variogram) at point support. Kyriakidis (2004) and Kyriakidis and Yoo (2005) suggest likelihood methods to infer the point support variogram. Goovaerts (2008) proposed to estimate this variogram by an iterative method-of-moments method. The method seeks the point-support variogram model by minimizing the difference between the theoretically regularized variogram model and the method-of-moments variogram model fitted to the areal data. More recently Orton et al. (2012) presented estimation of the covariogram at point support by restricted maximum likelihood (REML). Assuming a multivariate normal distribution of the residuals, the restricted log-likelihood multiplied by -2 equals

$$-2\ln\ell(\mathbf{Z} \mid \boldsymbol{\theta}) = \text{constant} + \ln \mid \overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}} \mid + \ln \mid \overline{\mathbf{X}}_{\mathcal{B}}^{\mathrm{T}}\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}{}^{-1}\overline{\mathbf{X}}_{\mathcal{B}} \mid + \bar{\mathbf{z}}^{\mathrm{T}}\mathbf{T}_{\mathcal{B}\mathcal{B}}\bar{\mathbf{z}}, \tag{15}$$

with $\mathbf{Z}$ the (unobserved) values of the response variable at point locations, $\boldsymbol{\theta}$ the vector with covariogram parameters, $\bar{\mathbf{z}}$ the $m \times 1$ vector with averages

of the response variable for the blocks, and

$$\mathbf{T}_{\mathcal{B}\mathcal{B}} = \overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}{}^{-1} - \overline{\mathbf{C}}_{\mathcal{B},\mathcal{B}}{}^{-1}\overline{\mathbf{X}}_{\mathcal{B}}\left[\overline{\mathbf{X}}_{\mathcal{B}}^{\mathrm{T}}\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}{}^{-1}\overline{\mathbf{X}}_{\mathcal{B}}\right]^{-1}\overline{\mathbf{X}}_{\mathcal{B}}^{\mathrm{T}}\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}{}^{-1}. \tag{16}$$

The restricted log-likelihood is not a function of the regression coefficients but of the covariogram parameters only. The regression coefficients are integrated out, i.e. the restricted log-likelihood is the log of the expected value over all possible values of the regression coefficients.

With an exponential variogram, which has two parameters, the distance parameter ϕ and the sill σ^2, minimization of the negative restricted log-likelihood is a one-dimensional problem. This is because for any given value of ϕ, the value of σ^2 that minimizes the negative restricted log-likelihood can be computed by

$$\frac{\bar{\mathbf{z}}^{\mathrm{T}}\mathbf{T}_{\mathcal{B}\mathcal{B}}\bar{\mathbf{z}}}{n-q}, \tag{17}$$

in which the covariance matrix $\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}$ in $\mathbf{T}_{\mathcal{B}\mathcal{B}}$ (Eq. 16) is computed with a sill of one.

The variance of the covariogram parameters can be estimated by the inverse of the Fisher information matrix (Pardo-Igúzquiza and Dowd, 2001). For REML estimation of the covariogram parameters, the ijth element of the Fisher information matrix is given by (Zimmerman, 2006)

$$\frac{1}{2}\mathrm{tr}\left(\mathbf{T}_{\mathcal{B}\mathcal{B}}\frac{\delta\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}}{\delta\theta_i}\mathbf{T}_{\mathcal{B}\mathcal{B}}\frac{\delta\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}}{\delta\theta_j}\right). \tag{18}$$

Given the REML estimates of the covariogram parameters the regression coefficients are estimated by Generalized Least Squares (GLS):

$$\hat{\boldsymbol{\beta}} = \left(\overline{\mathbf{X}}^{\mathrm{T}}\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}{}^{-1}\overline{\mathbf{X}}\right)^{-1}\overline{\mathbf{X}}^{\mathrm{T}}\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}{}^{-1}\overline{\mathbf{z}}. \tag{19}$$

The variance of the estimated regression coefficients equals

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \left(\overline{\mathbf{X}}^{\mathrm{T}}\overline{\mathbf{C}}_{\mathcal{B}\mathcal{B}}{}^{-1}\overline{\mathbf{X}}\right)^{-1}. \tag{20}$$

This variance is for the estimated parameters of the covariogram; in other words uncertainty about the covariogram parameters is not accounted for in this estimated variance of the regression coefficients.

2.4. Extrapolation

A potential risk of using covariates in ATP kriging is extrapolation of the model. Extrapolation occurs when for one or more covariates the values at a prediction location is outside the range of the covariate values in the data used for estimating the model parameters. This is a well-known problem in linear regression. In ATP kriging this problem can even be more serious because the covariate data are averages of regions, having smaller ranges than individual covariate values at points. A simple approach to avoid extrapolation would be to check whether for all covariates the covariate values are within the ranges of the covariates in the calibration data. However, this univariate approach is insufficient. Think of two correlated covariates, and a prediction location with a value for the first covariate just below the maximum of that covariate in the calibration data, and for the second covariate a value just above the minimum of that covariate in the calibration data. When correlation between the two variables is strong, the probability that this combination of covariate values is not present in the calibration data is large, and as a consequence there is still a risk of extrapolation. This is referred to as hidden extrapolation. A superior, multivariate approach is to compute the scaled distance of a prediction point to the centre of the cloud of

calibration data, in the space spanned by the covariates (Montgomery et al., 2001):

$$d = \mathbf{x}_0^{\mathrm{T}} \left[\overline{\mathbf{X}}_{\mathcal{B}}^{\mathrm{T}} \overline{\mathbf{X}}_{\mathcal{B}} \right]^{-1} \mathbf{x}_0 \tag{21}$$

with $\mathbf{x}_0$ a vector of length q with the covariate values at a prediction location. This distance is then compared with the maximum scaled distance of the calibration data to its centre, which can be computed by

$$d_{\max} = \max\{\mathrm{diag}(\overline{\mathbf{X}}_{\mathcal{B}}^{\mathrm{T}} \left[\overline{\mathbf{X}}_{\mathcal{B}}^{\mathrm{T}} \overline{\mathbf{X}}_{\mathcal{B}} \right]^{-1} \overline{\mathbf{X}}_{\mathcal{B}})\} \tag{22}$$

We use $d - d_{\max}$ as a measure for the extrapolation. For points with $d - d_{\max} > 0$, the model is extrapolated and the ATP kriging predictions and their kriging standard deviations must be interpreted with care, i.e. we actually are more uncertain about the crop yield than indicated by the kriging standard deviation, because we rely on a linear relation beyond the domain of the model.

3. Simulation study

This section illustrates ATP-kriging with a simple simulation study. We simulated data using a statistical model, details of which are given below. The simulated data are used as reality, i.e. as errorless values at points of the variable of interest. Areas are then defined, and for each area the simulated values at all points within that area is used to compute the average of that area. These averages are subsequently used to recover again the values at the points by ATP-kriging. The ATP-kriging predictions can be compared with the simulated values to compute the prediction errors.

We simulated values at 2000 equally spaced nodes (with spacing of 0.05 distance units) along a transect of length 100 distance units (Fig. 1). At each node a pair of observations is simulated, consisting of the value of the variable of interest z_i and a correlated covariate x_i. Values of covariate x were simulated with an exponential variogram with a sill of 2 and a distance parameter of 10 distance units (practical range: 30 distance units). A constant of 10 was added to the simulated x values. Then residuals ϵ were simulated using an exponential variogram with a sill of 0.5 and a practical range of 5 units. The simulated ϵ values were added to the simulated x values to give the simulated values of the variable of interest z. The correlation coefficient of the simulated x and z values was 0.85. The transect was split into 10 sections of equal length. The means of all simulated z values within sections were computed; these are shown in Figure 1 as a stepwise function.

We then predicted the simulated values of the z in two ways, using the section means of z only, without using the covariate x (ATP kriging without trend), and using the section means of z together with the simulated x values (ATP kriging with trend). The variogram used in kriging is the same as used in simulation. More specific, the variogram of z used in prediction equals the sum of the variogram of x and the variogram of ϵ.

The upper figure in Fig. 1 shows the results for ATP kriging without trend. The sharp boundaries of the stepwise line are smoothed by the ATP kriging. The kriging predictions roughly follow the simulated z values, but the prediction errors generally are quite large. The mean squared prediction error (MSE) equals 0.625. The width of the prediction interval is somewhat larger near the boundaries of the sections as compared to the centres. Intu-

itively this makes sense as near the boundaries we are in a transition zone.

The predictions obtained with ATP kriging with trend are much less smooth and show more detail. Prediction errors are much smaller than with ATP kriging without trend since it exploits the correlation with the covariate: the MSE equals 0.162.

Both for ATP kriging without trend and ATP kriging with trend, for all sections the average of the predictions at the nodes within this section is exactly equal to the section mean (result not shown), i.e. predictions are mass-preserving.

4. Case study

4.1. Data

The case study takes into consideration two different areas: the Province of Shandong in China, and the entire Burkina Faso. For Shandong yield statistics for the 17 districts are available for winter wheat and maize. These are official figures collected from Shandong Statistical Yearbook from 2000-2013 provided by Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (RADI, CAS), and the long term mean covers the period 2000-2013.

For Burkina Faso yield statistics refer to the 45 provinces and were obtained from AGRHYMET and the long term value is also based on the years 2000-2013.

Figure 2 show the actual average yields in ton ha^{-1} of grain maize and winter wheat in Shandong province of China, and of millet and sorghum in

Burkina Faso. Table 1 shows the minimum, maximum and average of the regional mean yields.

Besides data on the average crop yields we have maps with covariate data. Four sets of covariate data can be distinguished: 1. Vegetation data 2. Precipitation data 3. Temperature data and 4. Soil data. As to the first group we considered vegetation related parameters at a spatial resolution of approximately 1 by 1 km: for example FAPAR (Gobron et al., 2006), as derived from SPOT-VGT and cumulated over the crop cycle. The second and third group include crop driving variables namely rainfall, radiation and temperature as derived from CHIRPS and ECMWF ERA-Interim, also cumulated over the crop cycle. CHIRPS and ECMWF ERA-Interim data variables have a spatial resolution of respectively around 5 by 5 km and 25 by 25 km, respectively. In addition we selected some dekad specific variables derived from SPOT-VGT, CHIRPS and ECMWF ERA-Interim. These are variables not cumulated over the crop cycle but summarizing conditions for a specific 10-day period within the crop cycle. This selection was based on the performance of these dekad-specific covariates in a statistical crop forecast analysis. Data of ECMWF ERA-Interim, SPOT-VGT and CHIRPS were collected from the MARS project (Micale and Genovese, 2004; de Wit et al., 2010; Meroni et al., 2013), see http://marswiki.jrc.ec.europa.eu/agri4castwiki/index.php/Meteorological_data_from_ECMWF_mo and http://marswiki.jrc.ec.europa.eu/agri4castwiki/index.php/SPOT-VEGETATION, and U.S. Geological Survey (Funk et al., 2015). All covariate data of the first three groups have been aggregated over the years 2000-2013.

Afterwards data were spatially aggregated from their original resolution

to the administrative regions for which the yields are given. The spatial aggregation was done based on a land cover map only including (the share of) the pixels / grid cells under cropland (GLCshare, class = 2).

Tables 1 - 4 in the supplement list the set of vegetation and climate data used as potential covariates for wheat and maize in Shandong, and sorghum and millet in Burkina Faso from which we selected a subset that served as predictors in the model. As to the fourth group, the soil data, the following variables were included: coarse fragments (vol. % > 2mm), sand (mass %), silt (mass %), clay (mass %), bulk density (kg dm^{-3}), total available water capacity (cm m^{-1}), cation exchange capacity (CEC; cmolc kg^{-1} of fine earth fraction), pH (measured in water), organic carbon (OC; g kg^{-1}), total N (g kg^{-1}) and C/N ratio. The values apply to the soil depth: 0-20 cm. Data were derived from the WISE30SEC version 1.0 soil database (Batjes, 2015) and aggregated to the administrative regions. Finally, quantitative covariates were centered to zero means and scaled to standard deviations of 1, so that the importance of covariates can be evaluated on the basis of the absolute values of their associated regression coefficients.

4.2. Implementation of statistical methods

This section describes how we selected a model, more specifically how we selected a subset of covariates from the full list of all covariates. Besides, this section describes how we implemented the ATP kriging in practice.

4.2.1. Model selection and calibration

Model selection boils down to selecting a combination of covariates that serve as predictors x_k in Eq. 12. To reduce the total number of possible mod-

els the covariates are grouped into four subsets: vegetation, precipitation, temperature and soil covariates, see Tables with covariates in supplement. We only considered models with a maximum of one covariate per group. So we fitted all possible models with four covariates (one from each group), three covariates (a covariate from one group is missing), *et cetera*. Models were fitted by maximizing the log-likelihood, accounting for spatial correlation of the data. The fitted models were ranked on the basis of Akaike Information Criterion (AIC)

$$AIC = -2ln(\hat{L}) + 2k \tag{23}$$

with $\hat{L}$ the maximized loglikelihood and k the number of model parameters (regression coefficients and covariogram parameters). The equation shows that there is a penalty for the number of parameters, so that overfitting through inclusion of many covariates in the model as predictors, is avoided.

The best models in terms of AIC were then fitted by restricted maximum likelihood (REML) (Lark et al., 2006). Each entry of the matrix with mean covariances within and between regions, $\overline{\mathbf{C}}_{\mathcal{BB}}$ in Eq. 15, was estimated from $200 \times 200 = 40,000$ randomly selected pairs of points. For each pair the covariance is computed, the average of which is used as an estimate of the mean covariance. In principle, the model with the lowest AIC was selected. However, we also looked at the sign of the estimated regression coefficients. In case this sign did not make sense from an agronomic point of view, this model was discarded. The deviance was minimized by differential evolution using R package DEoptim (Ardia et al., 2012).

399 *4.2.2. ATP kriging*

400 For predicting the values at points by ATP kriging, given the estimated
401 covariogram parameters and regression coefficients, we wrote an R script.

402 *4.2.3. Leave-one-out cross-validation*

403 Unfortunately we did not have observed yields at point support, so that
404 we cannot validate the predicted yields at point support. What we can do
405 is validate the predictions at the area support. We did this by leave-one-
406 out cross-validation. The regional averages of the yield are left out one-by-
407 one. The average yields of the remaining regions are used, together with the
408 full-coverage maps of the covariates, to predict the yield at all grid nodes
409 discretizing the region that is left out. The average of these predictions at
410 point support are then compared with the reported regional average yield.

411 *4.3. Results of statistical analysis*

412 *4.3.1. Selected models*

413 The results of the statistical analysis are given in this section and an
414 agronomic interpretation is attempted. In case a model contains multiple
415 covariates such agronomic interpretation should be done with care when the
416 covariates are correlated. For instance, when two covariates are positively
417 correlated and both covariates individually have a positive effect on the yield,
418 then the sign of the coefficient associated with one of the covariates can even
419 become negative.

420 *Winter wheat and grain maize in Shandong.* For winter wheat in China
421 (Shandong) the model with the lowest AIC contains the covariates sand,
422 CRAIN-CH-17 (rain cumulated over the period 1 January - 20 June), and

TMAX-EC-15 (maximum temperature in the period 21 - 31 May) as covariates (Table 2). For sand and CRAIN-CH-17 the sign was negative, while for TMAX-EC-15 it was positive. The negative sign for sand indicates the lower capacity of coarse textured soils to retain water and nutrients. Winter wheat is fully irrigated and therefore normally water shortage is not a problem. But excessive precipitation could harm the crop (e.g. due to water logging, increased disease pressure, nutrient leaching, etc.). This might explain the negative sign for CRAIN-CH-17. Although the variable represents a cumulated amount of rain over the growing period, the negative sign may also indicate direct damages to the crop caused by excessive rainfall during flowering and before harvest. Warm and sunny weather is favorable for ripening, explaining the positive sign for TMAX-EC-15.

For grain maize in Shandong the model with the lowest AIC has CRAD-EC-28, CRAIN-EC-28, CFAPAR-28 and sand as predictors. The cumulated radiation (in this case cumulated over the period 1 July - 10 October, the growing season for maize in the area) has a positive effect on the biomass production and yield. Grain maize mainly grows on summer rainfall. Thus, a positive correlation with rainfall is expected. The coefficient for CFAPAR-28 equals -0.0800, so the effect of this covariate is small compared to that of the other two covariates. The negative sign can only be explained by the positive correlation between CFAPAR-28 and CRAD-EC-28.

For winter wheat and grain maize in Shandong the fitted distance parameters of the exponential covariograms are small (1.85 and 0.72 km, respectively; see Figure 1 in supplement), and the fitted sill parameters are very large, especially for maize (Table 2). The standard errors of the estimated

covariogram parameters, as obtained by the inverse of the Fisher information matrix, are very large when related to their estimated values, especially for wheat (Table 2). The very large uncertainty about these parameters is in accordance with the scarce data, consisting of the average crop yields for 17 districts only. For maize also the standard errors of the regression coefficients are very large, which can be explained by the large value for the estimated sill.

In REML estimation of the variogram we assume that the residuals at point support have multivariate normal distribution. This assumption cannot be checked because we do not have point support yield data. What we can do is look at the residuals of the mean yields of the regions. If this residual distribution is not normal, this suggests that the assumption of normal point-support residuals may not be valid. However, if the residual distribution is normal, this is not a proof for a normal distribution of the residual yield at point support. Q-Q plots of the residuals for grain maize and winter wheat show that the residuals of the mean yield are nicely normally distributed (see Figure 2 in supplement). The Shapiro-Wilk test statistic for maize equals 0.953 with a p-value of 0.512, and for wheat 0.972 with a p-value of 0.858.

Sorghum and millet in Burkina Faso. In the case of sorghum in Burkina Faso we selected a model with TMIN-EC-21 (minimum temperatures over the period 21 - 31 July), pH, CFAPAR-27 and CRAIN-EC-27 (FAPAR and rainfall cumulated over the growing period i.e. dekads 13-27, indicatively from 1 May to 30 September) as predictors. This was not the model with the lowest AIC (see excel file in supplement for the ten best models). The models with smaller AIC were discarded for agronomic reasons. All predictors except

pH in the selected model have positive signs. The negative sign for pH can be explained by the negative correlation between pH and OC ($r = -0.72$). In general terms, lower pH values lead to accumulation of organic matter. Higher organic matter facilitates higher yields, especially when it is scarce as it is usually the case in sub-tropical environments. The other predictors have a clear agronomic significance, specifically in relation to the growing conditions in Burkina Faso. In the first place, the more rainfall over the whole growing cycle, the higher the yield. Also, relatively low, suboptimal minimum temperatures can reduce the biomass growth. The FAO crop requirement database (http://ecocrop.fao.org) indicates, for varieties adapted to sub-tropical conditions, a minimum temperature (day and night) of 22 degrees throughout the season. At dekad 21 the southern regions of Burkina Faso have sub-optimal temperature conditions, below 22 degrees (19-20 degrees), while the northern zones are well above 22 degrees (23-24 degrees).

In the case of millet we did not select a model with covariates. The model that gave the lowest AIC has two covariate, CRAD-EC-27 (cumulated radiation over the growing season) and FAPAR-23, both negatively correlated with yield. This relation does not have a straightforward agronomic interpretation. None of the top 40 models was acceptable from an agronomic viewpoint (negative sign for coefficients associated with FAPAR, radiation or precipitation). Millet, probably due to the marginal conditions in which it is cultivated, deserves further investigations on its growth limiting factors and the accuracy of the available data to adequately represent them.

For sorghum the fitted distance parameter of the exponential covariogram was 21 km (see Figure 1 in supplement), and the estimated sill parameter

was 0.092 ton^2 ha^{-2} (Table 2). For millet the fitted covariogram parameter are 73 km (see Figure 1 in supplement) and 0.102 ton^2 ha^{-2}. The much larger value for the distance parameter compared to sorghum is because all spatial structure in the yield of millet is captured by the covariogram, whereas for sorghum part of the spatial structure is explained by the covariates. Compared to the covariogram parameters of the models for grain maize and wheat the relative standard errors of the covariogram parameters are considerably smaller.

Q-Q plots of the residuals of the mean yields for millet and sorghum show that the assumption of a normal distribution might be violated (see Figure 2 in supplement). Based on the Shapiro-Wilk test the null hypothesis of a normal distribution is rejected for millet ($W = 0.926$, p-value 0.0066), but not so for sorghum ($W = 0.969$, p-value 0.26).

4.3.2. Disaggregated yield maps

We decided not to make maps with disaggregated yields of winter wheat and grain maize in Shandong province for two reasons: 1. the high uncertainty about the model parameters, see previous section, and 2. to predict the yield at points, the model must be extrapolated in a very large part of the area (see Figure 4 in supplement).

Figures 3 and 4 show the ATP kriging predictions of the actual yields of sorghum and millet. Areas not cultivated are masked out based on the same land cover map and classes described for the spatial aggregation. The map of sorghum in Burkina Faso shows more spatial detail than the polygon map with average yield due to the use of covariates in the prediction. In the map for sorghum sharp transitions can be seen. This can be explained by the use

of the soil covariate pH in the prediction. This soil covariate is represented by a polygon map: all pixels within a polygon have the same value for the soil covariate (Figure 5). The map with predicted yield of millet is a smoothed version of the polygon map with average yields for the 45 regions in Burkina Faso. No covariates were used in the prediction, which explains the smooth surface.

For both crops in Burkina Faso the averages of the disaggregated yields are nearly equal to the average of the reported mean yields of the 45 regions (Table 1). Also the 45 regional averages of the predictions at points are very close to the reported means (see Figure 3 in supplement). The averages of the predictions are not exactly equal to the reported mean yields because the discretization points used to calculate average covariances and average values of fixed effects were not exactly the same as the prediction grid points (if the same points were used, then the relationship should in theory be exact). The ranges of the disaggregated yields are wider than the reported regional mean yields.

For both millet and sorghum, the kriging standard errors, computed as the square root of the kriging variances, are the smallest in the centre of the regions and increase towards their boundaries. For sorghum the standard error is also a function of the covariates for the spatial trend. Broadly speaking, the more extreme the covariate values at a target point compared to the average covariate values of the regions, the larger the kriging standard error. The large standard errors of predicted sorghum, say > 0.32, correspond with areas with high pH values (pH > 7.5), see Figure 5. The area where the geostatistical model of sorghum is extrapolated corresponds with

areas with high (> 7.5) or low (< 5.5) pH (Figure 5). For millet no covariate was used in disaggregation by ATP kriging, so in this case there is no risk of extrapolation.

Results of the leave-one-out cross-validation are shown in Figure 6. The correlation between the averages of the point-predictions and the reported mean yields are moderately strong: $r = 0.68$ and 0.76, for millet and sorghum, respectively. The root mean squared errors of the predictions equal 0.162 and 0.149 t ha^{-1} for millet and sorghum, respectively.

5. Discussion

ATP kriging is founded on statistical theory, and consequently also provides estimates of the precision of the disaggregated yields. This cannot be derived from the other downscaling methods indicated in the Introduction. Therefore this is the added value of the present application to the downscaling of crop yields. Variables derived from high resolution data sets such as numerical weather models (e.g. ECMWF), satellites (e.g. SPOT-VGT, CHIRPS) or soil databases are easily accommodated as covariates in ATP kriging, increasing the detail of the yield maps resulting from the disaggregagation method. There are however, a number of limitations and problems related to the application of ATP-kriging which are discussed now.

In both case studies no point data of crop yields were available. Ideally, data on the target support are available, so that they can be used in calibrating the model. In REML estimation of the covariogram a multivariate normal distribution is assumed, see section 2.3. This assumption can only be checked if we have crop yield data at the target support. Besides, if we have

crop yield data at the target support, these data can be used for validation, see Brus et al. (2014) for an example. In the absence of cop yield data at the target support, a critical evaluation of the geostatistical model by agronomy experts is of utmost importance.

In this research we treated the average crop yields of the regions as errorless data. In practice these averages are often regional means, estimated from data collected in national agricultural surveys. In this case we are uncertain about the regional means of crop yields. if their uncertainty could be quantified by the variance of the estimated mean, then it can be accounted for in ATP kriging, as shown by Orton et al. (2012) and Brus et al. (2014). The ATP kriging predictions then are not mass-preserving anymore, i.e. the average of disaggregated point-predictions in a region are not equal to the regional mean.

The hardest part of the application of ATP kriging is the estimation of the (residual) covariogram at the support of the prediction units (target support). With real points, i.e. infinitely small areal units as target support, the nugget parameter of the covariogram cannot be estimated from the data alone (Truong et al., 2014). This is because the contribution of the nugget parameter to the matrix with mean covariances within and between regions ($\overline{\mathbf{C}}_{\mathcal{BB}}$ in Eq. 15) tends to zero when the number of discretisation points of a region tends to infinity.[1] Also, the uncertainty about the distance parameter of the covariogram is large, especially with a few, large regions. This uncertainty about the covariogram parameters is not accounted for in the

[1]With target supports larger than points, for instance square grid cells of 1 km^2, the nugget parameter would not disappear from the diagonal elements.

predictions of the yield at points. Especially the kriging standard deviations are sensitive to the covariogram parameters. Therefore the kriging standard deviations should be used as a relative measure of uncertainty, not so much as an absolute measure. As a follow-up of this research we are testing a Bayesian approach to ATP kriging. Advantages of a Bayesian approach are that prior (expert) knowledge about the regression coefficients associated with the covariates is easily accommodated as shown by (Truong et al., 2014), and that uncertainty about the model parameters is accounted for in the uncertainty distribution of the predictions.

Another point of concern is the extrapolation of the model when covariates are used as predictors in the model. The model is calibrated on spatial averages of crop yields and covariates. Due to the averaging of the covariates, the range of covariate values becomes smaller than of the underlying covariate values at points. As a consequence, the domain of the model is smaller than that of a model calibrated on the point data. Broadly speaking, the fewer the number of regions with spatial averages of crop yields, the narrower the range of average covariate values, the smaller the domain of the geostatistical model. Apart from the requirements on estimation of the covariogram, this sets a lower limit to the number of regions to be used in ATP kriging.

Expert knowledge on the relation between crop yields and the covariates was used for grouping the covariates. More expert knowledge should also be included for instance by setting plausible yield ranges, avoiding that predicted values go beyond physical yield limits or maximum attainable yield levels. Such knowledge, especially from national and local experts for the relevant

crops, is also important when evaluating more in depth the spatial patterns resulting from the disaggregation of crop yields. In future other covariates and especially modelled yields could be tested. To the extent that key crop yield stresses are captured by the models, these could further help estimating actual yields.

6. Conclusions

- ATP kriging has potentials for disaggregation of spatially averaged crop yields. The advantage over other downscaling methods is that it is founded on statistical theory, and consequently also provides estimates of the precision of the disaggregated yields.

- ATP kriging requires the covariogram (or variogram) at the target support, which can be recovered from the area data by ML or REML. An advantage of ML and REML estimation of the covariogram over the deconvolution approach is that the uncertainty about the estimated covariogram parameters can be estimated by the Fisher information matrix.

- Uncertainty about the covariogram parameters is not accounted for in conventional ATP kriging. This sets a lower limit to the required number of regions. For Shandong with average crop yields for 17 regions only the standard errors of the variogram parameters were very large; For Burkina Faso with crop yield data of 45 regions, our uncertainty was considerably smaller.

- Extrapolation of the model can be a serious problem in ATP kriging,

especially when the number of regions is small, so that the range of regional averages of covariate values is much smaller than the range of the covariate values at the target support

Acknowledgements

This study was supported by the SIGMA European Collaborative Project (FP7-ENV-2013 SIGMA-Stimulating Innovation for Global Monitoring of Agriculture and its Impact on the Environment in support of GEOGLAM-project). The authors wish to thank VITO for providing and processing SPOT-VGT data and the JRC MARS project for providing downscaled ECMWF ERA-interim data.

References

Ardia, D., Mullen, K. M., Peterson, B. G., and Ulrich, J. (2012). *DEoptim: Differential Evolution in R*. version 2.2-1.

Batjes, N. (2015). World soil property estimates for broad-scale modelling (wise30sec, ver. 1.0).

Brus, D. J., Orton, T. G., Walvoort, D. J. J., Reijneveld, J. A., and Oenema, O. (2014). Disaggregation of soil testing data on organic matter by the summary statistics approach to area-to-point kriging. *Geoderma*, 226-227:151–159.

de Wit, A., Baruth, B., Boogaard, H., van Diepen, K., van Kraalingen, D., Micale, F., te Roller, J., Supit, I., and van der Wijngaart, R. (2010). Using ERA-INTERIM for regional crop yield forecasting in Europe. *Climate Research*, 44:41 – 53.

Fischer, G., Ermolieva, T., Ermoliev, Y., and van Velthuizen, H. (2006). Spatial recovering of agricultural values from aggregate information: Sequential downscaling methods. *International Journal of Knowledge and Systems Sciences*, 3(1):1 – 6.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., LauraHarrison, Hoell, A., and Michaelson, J. (2015). The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes. *Scientific Data*, 2.

Gallego, F. J., Batista, F., Rocha, C., and Mubareka, S. (2011). Disaggregating population density of the European Union with CORINE land cover. *International Journal of Geographical Information Science*, 25(12):2051–2069.

Gobron, N., Aussedat, O., and Pinty, B. (2006). JRC-FAPAR at 250 m resolution Algorithm Theoretical Basis Document.

Goerlich, F. J. and Cantarino, I. (2013). A population density grid for Spain. *International Journal of Geographical Information Science*, 27(12):2247–2263.

Goovaerts, P. (2008). Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences*, 40(1):101–128.

Goovaerts, P. (2011). A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties. *European Journal of Soil Science*, 62(3):371–380.

689 Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial
690 data. *Journal of the American Statistical Association*, 97(458):632–648.

691 Gotway, C. A. and Young, L. J. (2007). A geostatistical approach to linking
692 geographically aggregated data from different sources. *Journal of compu-*
693 *tational and graphical statistics*, 16(1):115–135.

694 Kyriakidis, P. (2004). A geostatistical framework for area-to-point spatial
695 interpolation. *Geographical Analysis*, 36(3):259–289.

696 Kyriakidis, P. C. and Yoo, E. (2005). Geostatistical prediction and simulation
697 of point values from areal data. *Geographical Analysis*, 37:124–151.

698 Lark, R. M., Cullis, B. R., and Welham, S. J. (2006). On spatial prediction
699 of soil properties in the presence of a spatial trend: the empirical best
700 linear unbiased predictor (E-BLUP) with REML. *European Journal of*
701 *Soil Science*, 57:787–799.

702 Lin, W.-C., Lin, Y.-P., Wang, Y.-C., Chang, T.-K., and Chiang, L.-C. (2014).
703 Assessing and mapping spatial associations among oral cancer mortality
704 rates, concentrations of heavy metals in soil, and land use types based on
705 multiple scale data. *International Journal of Environmental Research and*
706 *Public Health*, 11(2):2148–2168.

707 Lobel, D. B. (2013). The use of satellite data for crop yield gap analysis.
708 *Field Crops Research*, 143:56–64.

709 Meroni, M., Atzberger, C., Vancutsem, C., Gobron, N., Baret, F., Lacaze,
710 R., Eerens, H., and Leo, O. (2013). Evaluation of agreement between

space remote sensing SPOT-VEGETATION FAPAR time series. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4):195 – 1962.

Micale, F. and Genovese, G. (2004). Methodology of the mars crop yield forecasting system. vol. 1. meteorological data collection, processing and analysis. Technical Report Publication EUR 21291 EN, Office for Official Publications of the EU, Luxembourg.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Nagle, N. (2010). Geostatistical smoothing of areal data: Mapping employment density with factorial kriging. *Geographical Analysis*, 42(1):99–117.

Orton, T. G., Saby, N. P. A., Arrouays, D., Walter, C., Lemercier, B., Schvartz, C., and Lark, R. M. (2012). Spatial prediction of soil organic carbon from data on large and variable spatial supports. I. Inventory and mapping. *Environmetrics*, 23(2):129–147.

Pardo-Igúzquiza, E. and Dowd, P. A. (2001). Variance-covariance matrix of the experimental variogram: Assessing variogram uncertainty. *Mathematical Geology*, 33(4):397–419.

Schirrmann, M., Herbst, R., Wagner, P., and Gebbers, R. (2012). Area-to-point kriging of soil phosphorus composite samples. *Communications in Soil Science and Plant Analysis*, 43(7):1024–1041.

Truong, P., Heuvelink, G., and Pebesma, E. (2014). Bayesian area-to-point kriging using expert knowledge as informative priors. *International Journal of Applied Earth Observation and Geoinformation*, 30(1):128–138.

van Ittersum, M., Cassman, K., Grassini, P., Wolf, J., Tittonell, P., and Hochman, Z. (2013). Yield gap analysis with local to global relevance - a review. *Field Crops Research*, 14:4 – 17.

VoPham, T., Hart, J. E., Bertrand, K. A., Sun, Z., Tamimi, R. M., and Laden, F. (2016). Spatiotemporal exposure modeling of ambient erythemal ultraviolet radiation. *Environmental Health*, 15(1):111.

Webster, R. and Oliver, M. A. (2007). *Geostatistics for Environmental Scientists, Second Edition.* Wiley, Chichester.

You, L., Wood, S., Wood-Sichra, U., and Wu, W. (2014). Generating global crop distribution maps: from census to grid. *Agricultural Systems*, 127:53 – 60.

Zimmerman, D. L. (2006). Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, 17:635–652.

Table 1: Minimum, maximum and mean of reported regional mean yields and of disaggregated yields in ton ha^{-1}.

	Minimum	Mean	Maximum
Reported regional mean yield			
Wheat	4.55	5.66	6.58
Maize	5.57	6.61	7.43
Millet	0.526	0.902	1.60
Sorghum	0.705	1.05	1.69
Disaggregated yield			
Millet	0.458	0.931	1.76
Sorghum	0.377	1.07	2.38

Table 2: Estimated model parameters. In parantheses: standard errors of estimated model parameters.

	Grain maize, Shandong, China	
Regresssion coefficients	Intercept	6.66 (0.139)
	CFAPAR-28	-0.0547 (0.209)
	CRAIN-EC-28	0.152 (0.336)
	CRAD-EC-28	0.408 (0.365)
	sand	-0.228 (0.166)
Variogram	distance parameter (km)	0.72 (1.55)
	sill (ton^2 ha^{-2})	720 (2760)
	Winter wheat, Shandong, China	
Regresssion coefficients	Intercept	5.69 (0.0747)
	CRAIN-CH-17	-0.361 (0.0993)
	TMAX-EC-15	0.180 (0.0959)
	sand	-0.332 (0.0786)
Variogram	distance parameter (km)	1.85 (13.0)
	sill (ton^2 ha^{-2})	37.1 (489)
	Millet, Burkina Faso	
Regresssion coefficients	Intercept	0.962 (0.0848)
Variogram	distance parameter (km)	73.0 (26.2)
	sill (ton^2 ha^{-2})	0.102 (0.0255)
	Sorghum, Burkina Faso	
Regresssion coefficients	Intercept	1.060 (0.0366)
	CFAPAR-27	0.106 (0.0741)
	CRAIN-EC-27	0.0830 (0.113)
	TMIN-EC-21	0.127 (0.0725)
	pH	-0.112 (0.0444)
Variogram	distance parameter (km)	21.1 (10.1)
	sill (ton^2 ha^{-2})	0.0921 (0.0392)

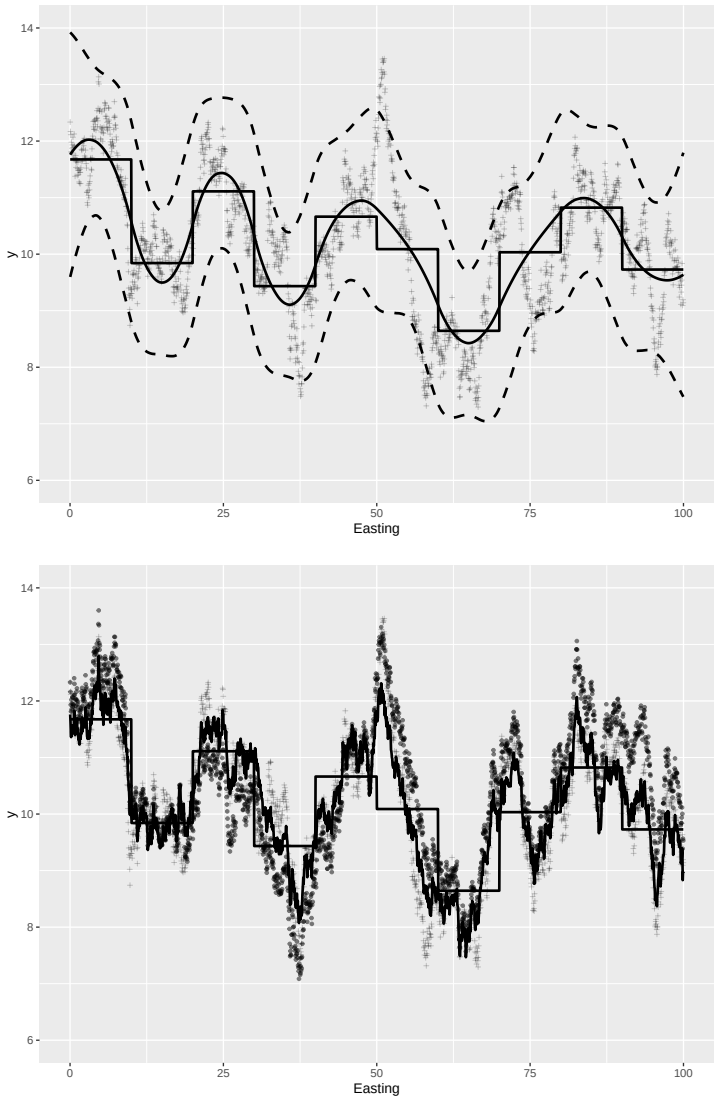

Figure 1: 1D illustration of ATP kriging without (upper figure) and with trend (lower figure). Simulated values of the variable of interest z (unobserved) are represented by the + symbols, simulated values of covariate x (lower figure) by dots. The data used in ATP kriging (without trend) are the means of z for the 10 sections (the stepwise line), and in ATP kriging with trend the means of z for the 10 sections plus the simulated x values. The solid lines represent the ATP kriging predictions of z. The dashed lines (upper figure) are the bounds of the 95% prediction interval

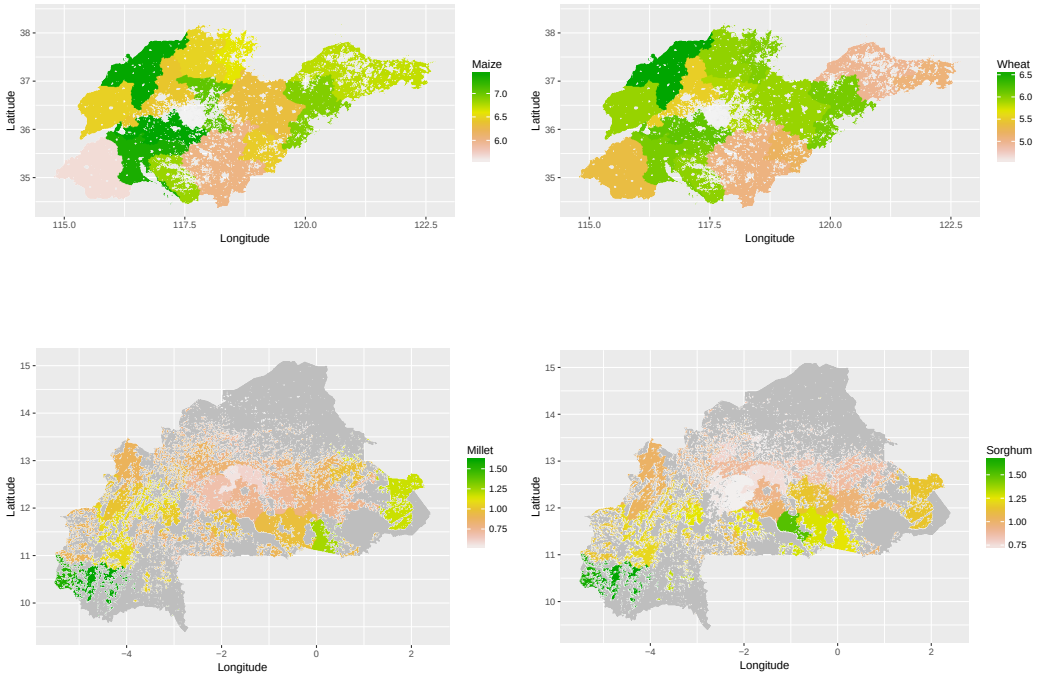

Figure 2: Actual average yield in ton/ha of grain maize and winter wheat in 17 regions of Shandong province (China), and of millet and sorghum in 45 regions in Burkina Faso. Masked by non-arable land of GLCshare (all classes except 2)

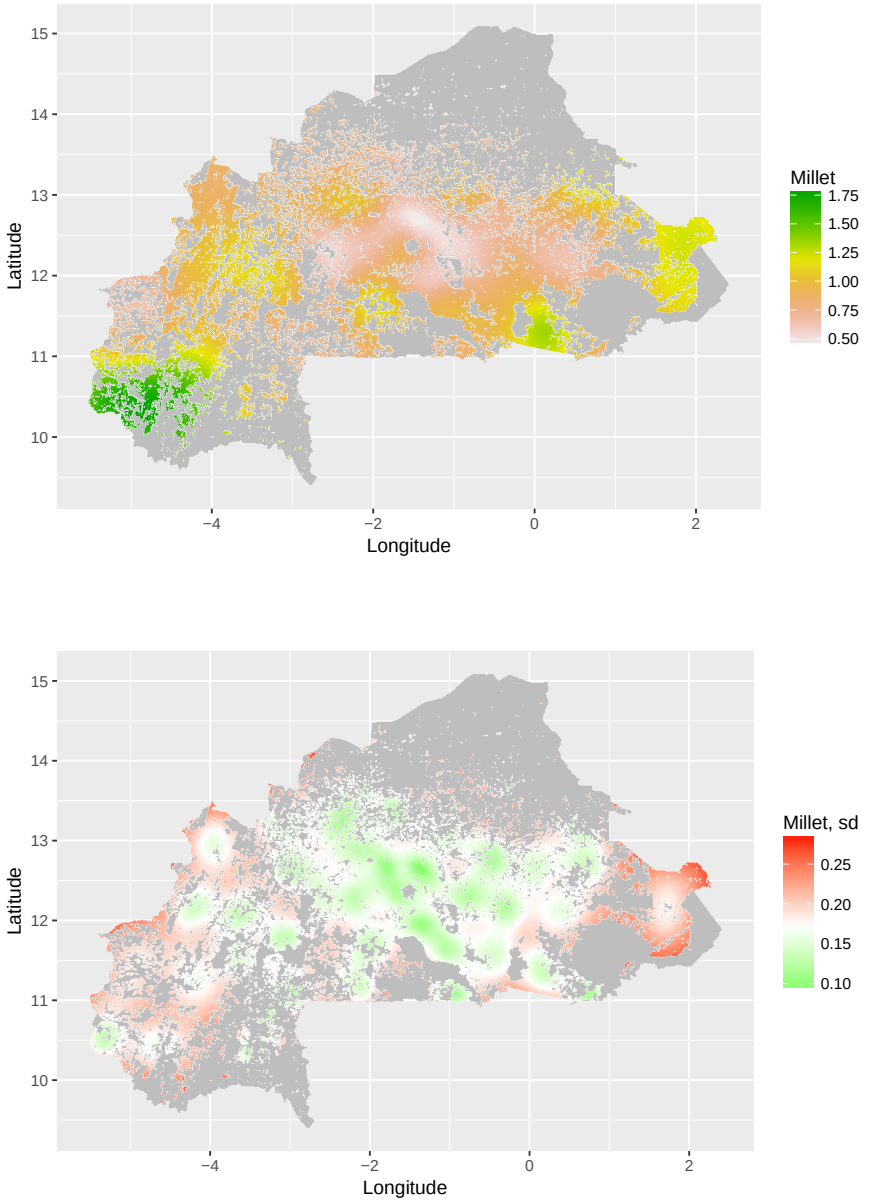

Figure 3: Predicted yield and kriging standard deviation for millet in Burkina Faso. Masked by non-arable land of GLCshare (all classes except 2)

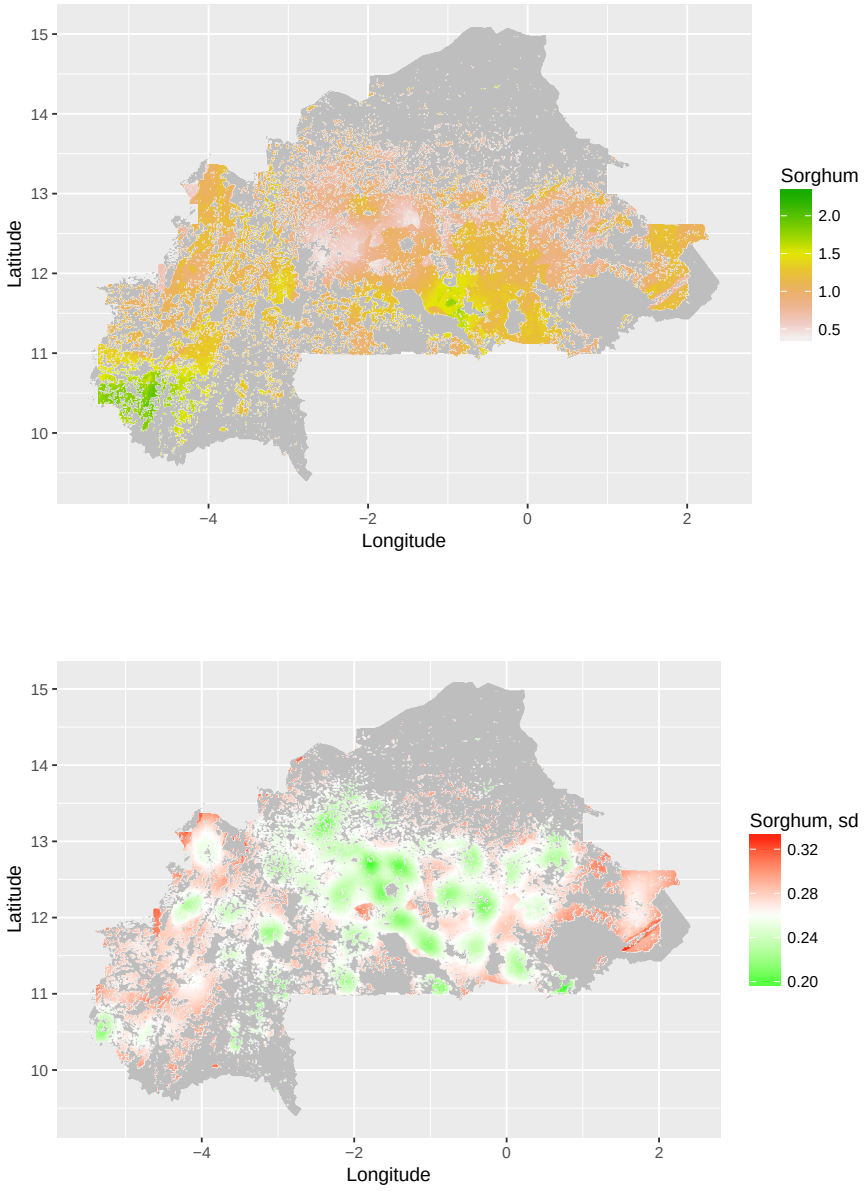

Figure 4: Predicted yield and kriging standard deviation for sorghum in Burkina Faso. Masked by non-arable land of GLCshare (all classes except 2)

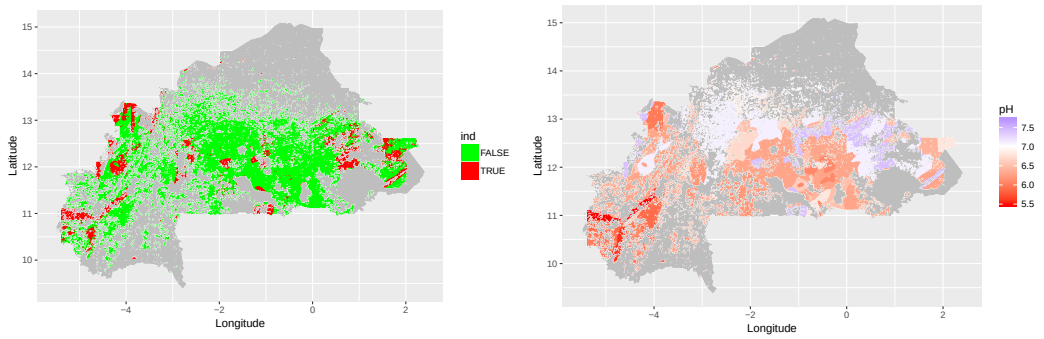

Figure 5: Area where model for sorghum is extrapolated, and map of pH

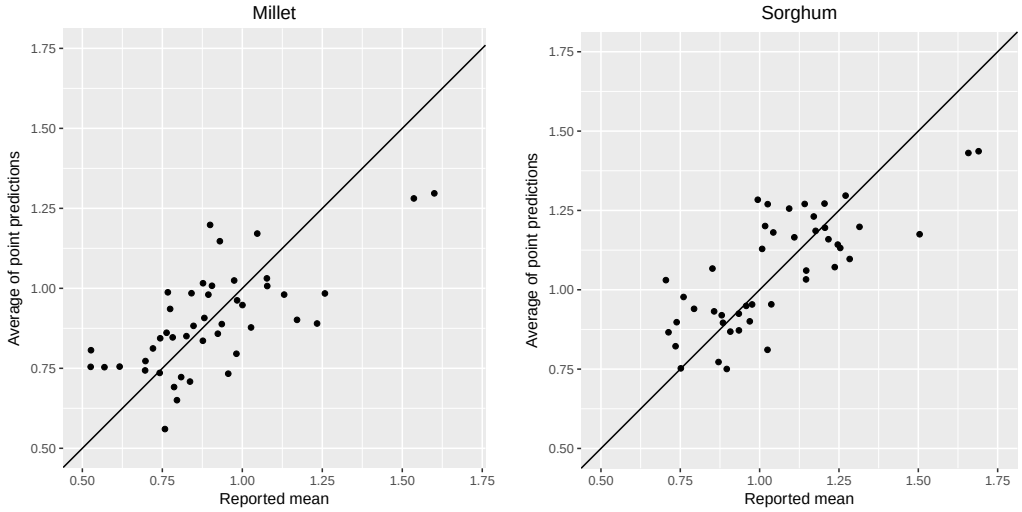

Figure 6: Scatterplot of average of predicted yield at point support versus reported mean yield per province for millet and sorghum in Burkina Faso, obtained by leave-one-out cross-validation.