European Journal of Agronomy 97 (2018) 48–59

Contents lists available at ScienceDirect

European Journal of Agronomy

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

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

D.J. Brus[a,*], 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

ARTICLE INFO

Keywords:
Yield gap
Aggregated data
Uncertainty

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 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.

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 prerequisite 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

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

https://doi.org/10.1016/j.eja.2018.05.003
Received 22 September 2017; Received in revised form 4 May 2018; Accepted 10 May 2018
Available online 26 May 2018
1161-0301/

'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 downscaling 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, 2007), Kyriakidis (2004), Kyriakidis and Yoo (2005), 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