Received: 7 December 2022 | Accepted: 27 March 2023

DOI: 10.1111/2041-210X.14106

Methods in Ecology and Evolution

RESEARCH ARTICLE

Uncertainty of spatial averages and totals of natural resource maps

Alexandre M. J.-C. Wadoux[1] | Gerard B. M. Heuvelink[2,3]

[1]Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia

[2]Soil Geography and Landscape group, Wageningen University & Research, Wageningen, The Netherlands

[3]ISRIC—World Soil Information, Wageningen, The Netherlands

Correspondence
Alexandre M. J.-C. Wadoux
Email: alexandre.wadoux@sydney.edu.au

Handling Editor: Phil J Bouchet

Abstract

1. Global, continental and regional maps of concentrations, stocks and fluxes of natural resources provide baseline data to assess how ecosystems respond to human disturbance and global warming. They are also used as input to numerous modelling efforts. But these maps suffer from multiple error sources and, hence, it is good practice to report estimates of the associated map uncertainty so that users can evaluate their fitness for use.
2. We explain why quantification of uncertainty of spatial aggregates is more complex than uncertainty quantification at point support because it must account for spatial autocorrelation of the map errors. Unfortunately, this is not done in a number of recent high-profile studies. We describe how spatial autocorrelation of map errors can be accounted for with block kriging, a method that requires geostatistical expertise. Next, we propose a new, model-based approach that avoids the numerical complexity of block kriging and is feasible for large-scale studies where maps are typically made using machine learning. Our approach relies on Monte Carlo integration to derive the uncertainty of the spatial average or total from point support prediction errors. We account for spatial autocorrelation of the map error by geostatistical modelling of the standardized map error.
3. We show that the uncertainty strongly depends on the spatial autocorrelation of the map errors. In a first case study, we used block kriging to show that the uncertainty of the predicted topsoil organic carbon in France decreases when the support increases. In a second case study, we estimated the uncertainty of spatial aggregates of a machine learning map of the above-ground biomass in Western Africa using Monte Carlo integration. We found that this uncertainty was small because of the weak spatial autocorrelation of the standardized map errors.
4. We present a tool to get realistic estimates of the uncertainty of spatial averages and totals of natural resource maps. The method presented in this paper is essential for parties that need to evaluate whether differences in aggregated environmental variables or natural resources between regions or over time are statistically significant.

KEYWORDS
block kriging, change of support, geostatistics, machine learning, mapping spatial aggregation, quantile regression forest

1 | INTRODUCTION

Large-scale mapping of concentrations, stocks and fluxes of natural resources is important for many purposes. Global, continental and regional maps of biophysical variables provide baseline data to assess how ecosystems respond to human disturbance and global environmental processes. They are used by modellers as well as end-users (Kullberg & Moilanen, 2014; Mokany et al., 2020; Schmidt-Traub, 2021). These maps also support climate change research. For instance, global maps of above-ground and below-ground carbon stocks and fluxes are crucial to evaluate whether there is a net positive or negative land-based emission of carbon into the atmosphere. In past years, a number of maps have been generated to inform and track ecosystem changes at continental and global scale. Some recent examples are maps of soil bacteria (Delgado-Baquerizo et al., 2018), above-ground biomass (AGB; Baccini et al., 2012), above- and below-ground biomass carbon density (Spawn et al., 2020) and plant biomass (Ma et al., 2021), forest cover change (Hansen et al., 2013), landcover change (Song et al., 2018) and soil organic carbon concentration and stocks (Padarian et al., 2022; Poggio et al., 2021).

Natural resource maps suffer from multiple error sources that affect their quality so that it is common practice to report an estimate of the map uncertainty. If maps are made using IPCC guidelines (Eggleston et al., 2006) then map uncertainty is often estimated by propagating the input uncertainties associated with the activity data and emission/removal factors. This uncertainty propagation approach, which ignores model structural uncertainty, was, for example, used in Harris et al. (2021) to map atmospheric greenhouse gas fluxes from global forest lands to obtain approximations of the lower and upper limits of a 95% prediction interval. Ahlström et al. (2012) estimated the global spatio-temporal carbon balance and obtained the associated uncertainty by simulating the carbon balance with a global vegetation model that used output from 18 different climate models. A similar approach was taken by Berthelot et al. (2005), where uncertainty in the simulated land carbon cycle response was quantified by running a global terrestrial carbon cycle model with inputs derived from 14 different ocean and atmosphere general circulation models. Alternatively, some studies have used bootstrapping strategies to quantify uncertainty. For instance, Saatchi et al. (2011) reported the uncertainty of forest carbon stock maps using a MaxEnt model and bootstrapping, whereas in Cook-Patton et al. (2020) global maps of carbon accumulation rates with uncertainty were obtained by fitting 100 random forest models on 100 bootstrap samples of the measurements, and by computing the pixel-wise standard deviation of the 100 model predictions. Mapping methods that explicitly quantify prediction uncertainty also exist, such as kriging (Webster & Oliver, 2007) and quantile regression methods (e.g. quantile regression forest [QRF], Meinshausen, 2006, as used in Baccini et al., 2012). Hengl et al. (2014) used regression kriging to map soil carbon stocks obtained at point locations along with an estimate of the uncertainty assessed by the point kriging variances. Poggio et al. (2021) quantified prediction uncertainty of global maps of basic soil properties using QRFs. These studies show that many solutions exist to quantify the uncertainty of spatial predictions, although some do not capture all uncertainty.

For mapping with machine learning, there is uncertainty in the measurements used to fit the model, uncertainty due to the use of a limited training dataset and uncertainty caused by the fact that the covariates explain only part of the spatial variation of the response variable. Of these three, the first is often ignored (one assumes that the training data have no measurement errors), but for a study where it is included see Wadoux et al. (2019) or Van der Westhuizen et al. (2022). The second uncertainty source is usually assessed using bootstrapping, while the third is characterized by the residual variance and is in most cases the main source of uncertainty. While many studies report an estimate of the uncertainty, this is often done with techniques that report a confidence interval (e.g. with bootstrapping the training data) instead of a prediction interval. A confidence interval is likely to grossly underestimate the overall uncertainty because it only considers the second source of uncertainty mentioned above. A prediction interval accounts for both the second and third source of uncertainty and is, therefore, wider than a confidence interval. In a recent study, McRoberts et al. (2022) proposed an uncertainty assessment that accounts for both sampling and residual uncertainty, which correspond to the second and third sources of uncertainty. Hereafter we consider uncertainty caused by the second and third sources.

Many studies also report on spatial averages or totals of the mapped variable, either for the whole study area or for geographies within it, such as national or bioclimatic domains. Averages and totals are informative with respect to international coordinated efforts to assess trajectories and for climate change research. Baccini et al. (2012), for instance, reported the total carbon stored in above-ground live woody vegetation by country and sub-region. They did so by calculating the sum of all carbon stock pixel values within the area and compared their estimates with existing products. Santoro et al. (2021) made global maps of AGB and provided country-specific statistics of forest area and total AGB, to enable comparison with country-specific data from an FAO FRA survey of 2010. Another example is Wiesmeier et al. (2011) in which total soil organic carbon, total carbon, total nitrogen and total sulphur stocks were estimated for different land use units in the Xilin River Catchment in China. More recently, the web platform https://soilsrevealed.org/ was launched and supports spatial aggregation of soil organic carbon stocks to countries and other geographies. Generally, spatial aggregation through summing or averaging is straightforward and

performed by simply adding up or averaging the spatial predictions over the area of interest.

The question addressed in this paper is how to obtain the uncertainty of spatial aggregates. While this is highly relevant because it is required for significance testing, such as when carbon accounting projects need to demonstrate that carbon sequestration in a region was successful over a given period of time, it is more difficult than computing the spatial aggregate of the predictions. For instance, it is incorrect to estimate the uncertainty of a spatial average by averaging the uncertainties at all points over which the average is computed. This strongly overestimates the uncertainty of the spatial aggregate, because it ignores the fact that map errors partially cancel out. In fact, the uncertainty of a spatial aggregate strongly depends on the degree of spatial autocorrelation of the map errors and hence this must be taken into account when the uncertainty of a spatial average or total is computed. While this fact is well-known to geostatisticians (see for instance Isaaks & Srivastava, 1989, chapter 8), it is largely disregarded by the global mapping community. For instance, Lugato et al. (2014) ignored spatial autocorrelation of the map error when computing the uncertainty of aggregated soil organic carbon maps for NUTS2 regions in Europe. Harris et al. (2021) and Plaza et al. (2018) assumed map errors to be spatially uncorrelated when the uncertainty of global forest carbon fluxes and totals of element stocks in drylands were derived. As we will show in the next section, this assumption leads to an underestimation of the actual uncertainty because it means that practically all map errors cancel out when aggregation is done over a large number of points.

Under- or overestimation of the uncertainty of spatial aggregates can have serious consequences. It means that end-users will be too confident or not confident enough about the aggregated products. This is a problem for a number of applications, such as when the maps are used to inform policy-makers and coordinate international efforts (e.g. to mitigate climate change). It also leads to unrealistic results in spatial uncertainty propagation applications, when the map is used as input to an environmental model (e.g. with Earth System models). To date, only few studies (e.g. Kros et al., 2012; Lesschen et al., 2007) used methodologies that explicitly account for spatial autocorrelation of the map error during the aggregation process. Geostatistical modelling and prediction (e.g. Kempen et al., 2019) combined with block kriging (Cressie, 2015) is a preferred solution to map environmental variables and estimate the uncertainty of aggregated variables in presence of autocorrelated error. However, the use of block kriging for large-scale and global applications is limited because it is computationally demanding. It also requires geostatistical expertise, which many map makers do not have. Further, global and large-scale maps are nowadays often made using machine learning (see, for example, the studies of Ma et al., 2021; Poggio et al., 2021; Saatchi et al., 2011), which does not account for spatial autocorrelation and so cannot quantify the uncertainty of spatial aggregates (Heuvelink & Webster, 2022).

The objectives of this paper are to (1) show that the uncertainty of spatial aggregates strongly depends on the spatial autocorrelation of the map errors; (2) explain and illustrate how spatial autocorrelation can be properly accounted for in block kriging and (3) propose and test an alternative model-based approach that accounts for spatial autocorrelation of map errors to quantify uncertainty in spatial averages and totals, one that avoids the numerical complexity of block kriging and is feasible for large-scale machine learning studies. In the following three sections, we discuss each of these objectives and use a synthetic dataset and two real-world case studies to illustrate the methodology.

2 | UNCERTAINTY OF SPATIAL AGGREGATES

Consider a biological, ecological or environmental variable $z = \{z(\mathbf{s}), \mathbf{s} \in D\}$ in a study area D. In practice we can only approximate z by a map $\hat{z} = \{\hat{z}(\mathbf{s}), \mathbf{s} \in D\}$ and so at each location $\mathbf{s}$ there is a map error $e(\mathbf{s}) = z(\mathbf{s}) - \hat{z}(\mathbf{s})$. These errors are usually not known (otherwise we would eliminate them by replacing $\hat{z}$ by $\hat{z} + e$). Instead, we treat $e(\mathbf{s})$ as a realization of a random variable $\varepsilon(\mathbf{s})$, which is fully characterized by a probability distribution $F(\mathbf{s})$. Kriging, QRFs and Monte Carlo uncertainty propagation are examples of techniques that produce these map error probability distributions. In this way, we can quantify the map accuracy at each location $\mathbf{s}$, for instance by the mean and variance of $\varepsilon(\mathbf{s})$ or by a prediction interval (e.g. the difference between the 0.95 and 0.05 quantiles of $F(\mathbf{s})$).

The probability distribution of $\varepsilon(\mathbf{s})$ for any $\mathbf{s} \in D$ depends on many factors, such as the degree of spatial variation of the response variable, the density and locations of training data, the explanatory power of the covariates, and the mapping method. It also depends on the *support* of the measurements and predictions, that is, the area or volume over which a measurement or prediction is made. For instance, soil samples can have 'point' support, that is a volume of soil of the order of $1\,dm^3$ at a point location or can have a bigger support, such as when a composite soil sample is taken within a block of $5\,m \times 5\,m$ or $10\,m \times 10\,m$ at a specific depth interval, typically the 0–20 or 0–30 cm topsoil. Likewise, in ecology the average AGB could, for example, be measured on a spatial support of a $10\,m \times 10\,m$ plot or for a circular area with 5 m radius. In mining, the measurement support typically is a volumetric unit such as a drill hole core or a rock chip. While sensu stricto these supports are not points because they have a non-zero area or volume, they are small compared to the extent of the study area and often considered as points. Note that the variability of measured properties largely depends on their support. For instance, the variability of soil properties may differ greatly between single soil samples and composite soil samples, especially when the micro-scale soil variation is large (Webster & Oliver, 2007, section 4.8). It is more work to collect a composite soil sample than a single soil sample but the advantage is that micro-scale variation is eliminated. Note also that in mapping, usually the support of the predictions is equal to that of the measurements. Thus, a prediction $\hat{z}(\mathbf{s})$ refers to the value that one expects to get if one would measure at $\mathbf{s}$ in the same way that the training data were obtained. Likewise, $\varepsilon(\mathbf{s})$ refers to the difference between the predicted and true value

at **s** at the measurement support. This is important, because we will see below that the variance of $\varepsilon(\mathbf{s})$ strongly depends on the support.

2.1 | Uncertainty of spatial averages illustrated with synthetic examples

In practice, predictions are often required at a support larger than that of the measurements. For instance, we may wish to map the AGB for entire forest stands, management zones or regions. We refer to a support that is much larger than that of the point measurements as the 'block' support, which can be a square grid cell but also an irregularly shaped region. A change of support has an effect on prediction and prediction uncertainty (Heuvelink, 1998, section 2.5). For prediction, moving from point to block support is easy when the aggregation process is linear, such as when we predict the arithmetic mean of the point support values within the block:

$$\hat{z}_B = \frac{1}{|B|}\int_{\mathbf{s}\in B}\hat{z}(\mathbf{s})\mathrm{d}\mathbf{s} \approx \frac{1}{n}\sum_{i=1}^{n}\hat{z}(\mathbf{s}_i), \quad (1)$$

where $B \subseteq \mathcal{D}$ is the block and n is the number of points used to discretize the block and approximate the integral. While the difference between point and block predictions is often modest, spatial averaging typically leads to a substantial decrease of the uncertainty. This is demonstrated by computing the variance of the block-average map error:

$$\mathrm{var}\left(\frac{1}{|B|}\int_{\mathbf{s}\in B}\varepsilon(\mathbf{s})\,\mathrm{d}\mathbf{s}\right) = \frac{1}{|B|^2}\int_{\mathbf{s}\in B}\int_{\mathbf{u}\in B}\mathrm{cov}(\varepsilon(\mathbf{s}),\varepsilon(\mathbf{u}))\,\mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{s}, \quad (2)$$

where $\mathrm{cov}(\varepsilon(\mathbf{s}),\varepsilon(\mathbf{u}))$ is the covariance of the map errors at locations **s** and **u**.

If we discretize the block B into a finite number of n points as before then the variance of the block-averaged map error is estimated by

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon(\mathbf{s}_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(\varepsilon(\mathbf{s}_i)) + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{cov}(\varepsilon(\mathbf{s}_i),\varepsilon(\mathbf{s}_j)), \quad (3)$$

The discretization involves an approximation error which becomes smaller as n increases. Equation (3) shows the importance of the covariance term: if the correlation between $\varepsilon(\mathbf{s}_i)$ and $\varepsilon(\mathbf{s}_j)$ is zero for all $\mathbf{s}_i \neq \mathbf{s}_j$ (i.e. no spatial autocorrelation of map errors) then the variance reduces to zero; if the correlation is one for all $\mathbf{s}_i$ and $\mathbf{s}_j$ then there is no reduction of uncertainty. This is graphically illustrated for a one-dimensional case in Figure 1.

Let us also illustrate the variance reduction effect with spatial stochastic simulation in a two-dimensional synthetic case. On a grid of size 100 cells × 100 cells, we simulated three cases of a stationary isotropic normally distributed random field, characterized by a correlation function representing map errors with a weak, moderate and strong spatial autocorrelation. Spatial autocorrelation was characterized by an exponential model (Webster & Oliver, 2007) with three parameters: a range parameter, a nugget-to-sill ratio and a sill value. We used a range parameter of 30 units in all cases and a nugget-to-sill ratio equal to two-thirds for weak, one-third for moderate and zero for strong spatial autocorrelation. We generated 500 possible realities for each of the three cases using unconditional sequential Gaussian simulation (Webster & Oliver, 2007, chapter 12). The mean of the random field was set to zero and the variance (i.e. the sill of the variogram) to one in all cases. Each of the three rows in Figure 2a shows four example realizations for each case. Next, we computed the spatial average for each of the 500 realities per case. Figure 2b shows boxplots of the 500 spatial averages for each case. The box plots clearly show that the spatial average of the map error is smallest for the weak spatial autocorrelation case and largest for the strong spatial autocorrelation case. Note that in case of zero spatial autocorrelation (i.e. a pure nugget variogram), the boxplot would have nearly collapsed to a single value close to zero. In that case, the standard deviation of the spatial average error would be 100 times smaller than that of the point support error, as is also evident from Equation (3).

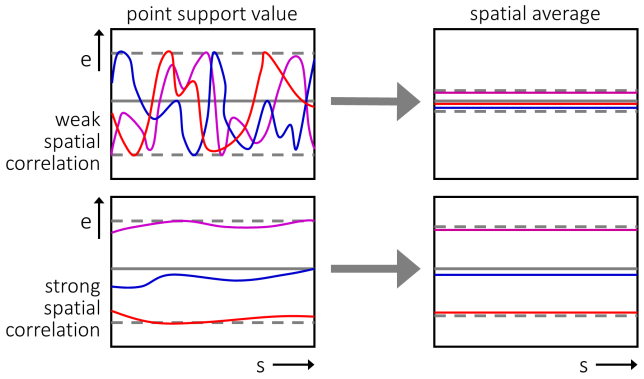

FIGURE 1 Spatial averaging of map error (*y*-axis) for three realizations (coloured lines) along a spatial transect (*x*-axis), in case of weak spatial autocorrelation (top) and strong spatial autocorrelation (bottom). Dashed lines represent the error range.

3 | SPATIAL AGGREGATION WITH BLOCK KRIGING

Geostatistical modelling and prediction with block kriging is a well-developed theory and thoroughly described in standard textbooks (Goovaerts, 1997; Webster & Oliver, 2007). In this section, we briefly summarize block kriging as a means to predict block averages of a target variable from point measurements and obtain the associated prediction uncertainty. We also explain how block kriging of map errors may be used to estimate the uncertainty of spatial averages of maps that were not made using kriging.

3.1 | Block kriging of a target variable

Geostatistical prediction of block averages and quantification of the associated prediction uncertainty starts by defining a geostatistical model of a target biological, ecological or environmental variable as:

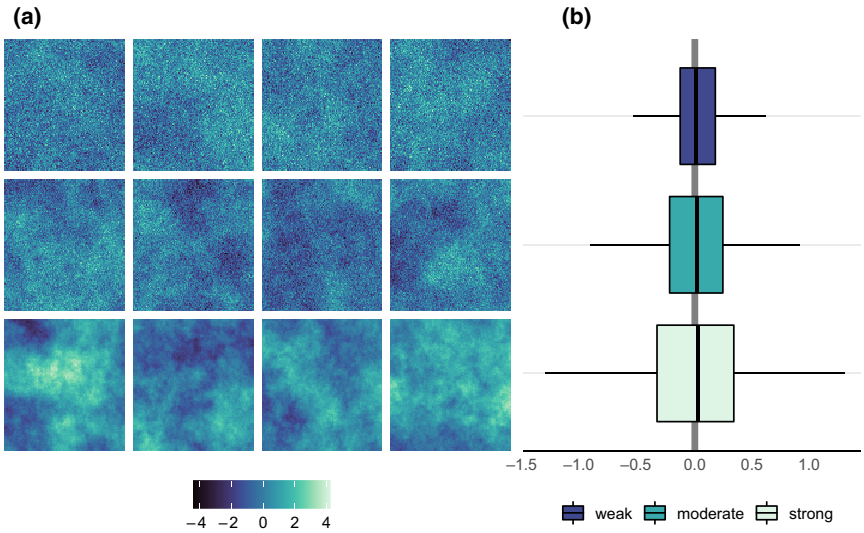

FIGURE 2 Example of four out of 500 simulated realities of a map error for the weak (top row), moderate (middle row) and strong (bottom row) spatial autocorrelation cases (a) and boxplots of the spatial averages of the 500 realities (b).

$$Z(\mathbf{s}) = m(\mathbf{s}) + \delta(\mathbf{s}), \tag{4}$$

where m is the trend or drift of the target variable and δ is a zero-mean stochastic residual, whose spatial variability is characterized by the semivariance $\gamma(\mathbf{s},\mathbf{s}') = \frac{1}{2}\mathrm{E}\left[\left(\delta(\mathbf{s})-\delta(\mathbf{s}')\right)^2\right]$, which must be defined for all combinations of locations $\mathbf{s}$ and $\mathbf{s}'$ in $\mathcal{D}$. The trend is typically modelled as a linear function of environmental covariates in the form $m(\mathbf{s}) = \sum_{k=0}^{K}\beta_k f_k(\mathbf{s})$, where the β_k are regression coefficients and the f_k spatially distributed covariates (Hengl et al., 2004). Note that it is common to define $f_0(\mathbf{s}) \equiv 1$ for all $\mathbf{s}$ so that β_0 represents the intercept of the regression. Alternatively, the trend may be taken as the outcome of a mechanistic or machine learning model (for an example, see Section 3.2). For simplicity we will assume here that $m(\mathbf{s})$ is constant, but the theory easily extends to the case of a non-constant trend. In addition, we will also assume that δ is second-order stationary, which means that it has constant variance and that the semivariance only depends on the separation distance between locations, that is $\gamma(\mathbf{s},\mathbf{s}') = \gamma(\mathbf{s}-\mathbf{s}')$. The function that defines the semivariance as a function of separation distance is known as the variogram.

Let there be n measurements $z(\mathbf{s}_i)$ ($i = 1, \ldots, n$) of the target variable. Ordinary block kriging predicts the average of Z in block B as a weighted average of these measurements:

$$\hat{z}(B) = \sum_{i=1}^{n}\lambda_i z(\mathbf{s}_i), \tag{5}$$

where λ_i is a weight associated with measurement $z(\mathbf{s}_i)$. The weights are chosen such that the expected squared prediction error, also known as the block kriging variance, is minimized, under a condition of unbiasedness. The weights are obtained by solving a linear system of $n+1$ by $n+1$ equations, which depend on the variogram. Any geostatistical study must therefore model the variogram before kriging.

The block kriging variance is derived as follows:

$$\begin{aligned}\mathrm{var}\left(\hat{Z}(B)-Z(B)\right) &= \mathrm{E}\left[\left\{\hat{Z}(B)-Z(B)\right\}^2\right] \\ &= 2\sum_{i=1}^{n}\lambda_i\overline{\gamma}(\mathbf{s}_i,B) - \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\mathbf{s}_i-\mathbf{s}_j) - \overline{\gamma}(B,B),\end{aligned} \tag{6}$$

where $\overline{\gamma}(\mathbf{s}_i,B)$ is the average of the semivariance between location $\mathbf{s}_i$ and all locations in block B, $\gamma(\mathbf{s}_i-\mathbf{s}_j)$ is the semivariance between location $\mathbf{s}_i$ and $\mathbf{s}_j$, and $\overline{\gamma}(B,B)$ is the within block variance, that is the average of all semivariances between all paired locations in the block.

Values predicted by block kriging are usually slightly smoother than those obtained with point kriging and the block kriging variance is generally much smaller than the point kriging variance. Thus, predictions of spatial averages are less uncertain than predictions at points, as already noted in Section 2. This is also apparent from Equation (6), which has a term that subtracts the within-block variance. This term is absent in case of point kriging (see Oliver & Webster, 2015, section 4.2). Equation (6) shows that the block kriging variance is much smaller than the point kriging variance in a case where the within-block variance is large, that is when there is weak spatial autocorrelation (see Figures 1 and 2).

Block kriging requires the within-block variance, which involves a double integration. This may become computationally intensive for large-scale applications where the number of measurements used to discretize the block is large. This is further discussed with a solution for large-scale applications in Section 4.

3.2 | Application to topsoil organic carbon mapping in France

We used data from the French Soil Monitoring Network (RMQS; Saby et al., 2020) composed of 2084 topsoil (0–30 cm) samples with values of organic carbon content (g kg^{-1}). The RMQS dataset is based on a systematic grid sampling design with a grid spacing of 16 km, covering mainland France. A composite soil sample was obtained at

each sampling location by mixing 25 individual soil cores collected according to an unaligned design within a 20×20 m area. Organic carbon data were obtained by the dry combustion method. We further collected a set of five spatially exhaustive covariates: elevation (in meters Rabus et al., 2003), slope (percent), long-term average precipitation (mm) and temperature (°C) from the set of BioClim variables (Fick & Hijmans, 2017), and the MODIS SWIR 2 (Lyapustin et al., 2018) long-term mean. All covariates were either aggregated or resampled with bilinear interpolation to have the same spatial resolution with grid cells of 250 m × 250 m and projected to the Lambert 93 coordinate system.

With the topsoil OC data and their matching values of environmental covariates, we built a block regression kriging model composed of a linear trend and spatially correlated residuals. We fitted a Matèrrn function to the sample variogram using the Method of Moments (Lark, 2000). The fitted variogram had a nugget of 212 $(g\,kg^{-1})^2$, a partial sill of 91 $(g\,kg^{-1})^2$, a smoothness parameter of 0.5 and an effective range of 96 km. These values are similar to those from previous studies on OC mapping in France (e.g. Mulder et al., 2016, supplementary material D). Note that here we assumed isotropy so that the separation distance reduces to Euclidean distance. Predictions were then made at different spatial supports, for points (i.e. the same support as the measurements), aggregated to the 98 mainland French departments, to the 12 French regions and to the whole of mainland France. For each of these cases, we also computed the block kriging variance of the spatial averages.

The block predictions shown in Figure 3 are equal to the average of all point predictions within the block, which means that at national support the prediction is simply the average of all predictions in mainland France. The most important message from Figure 3 is that the prediction uncertainty decreases substantially when the support increases. For instance, the mean relative error (i.e. the kriging standard deviation divided by the prediction and multiplied by 100%) at point, departmental, regional and national support is 74%, 11%, 5% and 1%, respectively.

4 | UNCERTAINTY OF SPATIAL AVERAGES WITHOUT KRIGING

Block kriging is computationally demanding because it requires solving a linear system of equations and computation of the within-block variance. For large datasets and global applications, block kriging is challenging or might not be possible, but one still needs to take spatial autocorrelation of the map error into account to get a realistic estimate of the uncertainty of spatial averages and totals. It is not only computational problems that hinder the use of block kriging. Other obstacles are that it requires geostatistical expertise and that it makes stringent assumptions about the statistical properties of the stochastic residual of Equation (4), which may not always be realistic.

4.1 | Proposed approach

To circumvent the problems associated with block kriging, we propose the following spatial aggregation method, which also quantifies the uncertainty of block averages and totals. The method assumes that the point support prediction errors are quantified by a probability distribution or standard deviation. It accounts for non-stationary variance of the point prediction errors but assumes that their spatial autocorrelation is stationary. It needs measurements of the target variable to assess spatial autocorrelation of map errors, either using a separate independent dataset or through a cross-validation approach. The method consists of eight steps:

1. Quantify the prediction uncertainty at point support for all locations in the study area. For instance, these may be characterized by the quantiles of the conditional distribution as obtained using QRFs or by conformal prediction (Shafer & Vovk, 2008).
2. Compute the point support prediction error standard deviation at each measurement and prediction location from its conditional distribution. When using QRF this can easiest be done by sampling N times from the conditional distribution and computing the sample standard deviation. The sample size N should be sufficiently large. In practice, a number between 200 and 500 should be sufficient.
3. At every measurement location, divide the prediction error (i.e. the difference between the measurement and the prediction) by the prediction error standard deviation obtained in step 2 to obtain the standardized prediction error. Note that the prediction errors is obtained at validation locations, that is, the error is derived using a new set of independent data or with cross-validation of the training data. In other words, the prediction at a measurement location should be made using a model that was not trained with that measurement.
4. Estimate a correlation function from the standardized prediction errors at measurement locations, for instance, by fitting a variogram. In that case the correlation function is derived from the variogram using the identity $\rho(\mathbf{h}) = (\text{sill} - \gamma(\mathbf{h}))/\text{sill}$ (Webster & Oliver, 2007, section 4.1), where ρ is the correlation function, sill is the estimated sill of the variogram and $\mathbf{h}$ is spatial distance. Note that here we assume that the standardized prediction errors are second-order stationary, that is we assume that the correlation between the standardized prediction errors at two locations only depends on the Euclidean distance between the locations. This is a less stringent assumption than assuming that the prediction errors are second-order stationary.
5. To obtain the variance of the prediction error of the spatial average, which is equal to the variance of the spatial average of the point support prediction errors, we first note that it is given by:

$$\begin{aligned}\operatorname{var}\left(\frac{1}{|B|}\int_{\mathbf{s}\in B}\varepsilon(\mathbf{s})\,d\mathbf{s}\right) &= \frac{1}{|B|^2}\int_{\mathbf{s}\in B}\int_{\mathbf{u}\in B}\operatorname{cov}(\varepsilon(\mathbf{s}),\varepsilon(\mathbf{u}))d\mathbf{s}d\mathbf{u}\\ &= \frac{1}{|B|^2}\int_{\mathbf{s}\in B}\int_{\mathbf{u}\in B}\sigma(\mathbf{s})\cdot\sigma(\mathbf{u})\cdot\rho(|\mathbf{s}-\mathbf{u}|)d\mathbf{s}d\mathbf{u},\end{aligned} \tag{7}$$

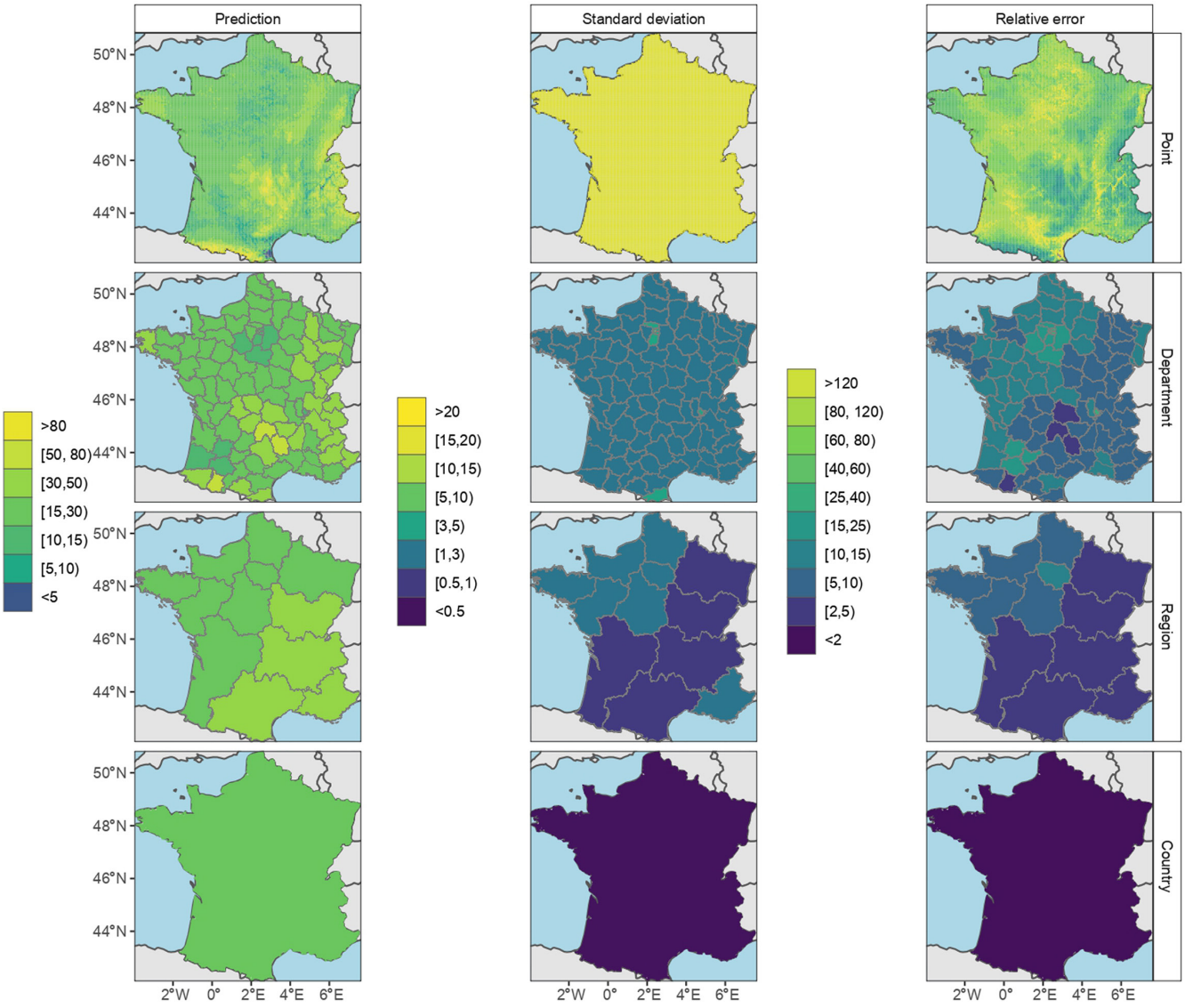

FIGURE 3 Topsoil organic carbon maps of kriging prediction ($g\,kg^{-1}$), kriging standard deviation ($g\,kg^{-1}$) and relative error (kriging standard deviation divided by prediction and multiplied by 100) for mainland France at point, departmental, regional and national support.

where $\sigma(\mathbf{s})$ and $\sigma(\mathbf{u})$ are the standard deviations of the point support prediction errors at $\mathbf{s}$ and $\mathbf{u}$, as obtained in step 2, and $\rho(|\mathbf{s}-\mathbf{u}|)$ is the value of the correlation function of the standardized prediction error at the separation distance between $\mathbf{s}$ and $\mathbf{u}$, obtained in step 4.

6. Evaluation of the double integral in Equation (7) can be challenging, particularly because the block B itself is two-dimensional. However, an effective way of evaluating it is through Monte Carlo integration (Robert & Casella, 2004, chapter 3). While other numerical integration algorithms usually evaluate the integrand at a regular grid, Monte Carlo integration randomly chooses points at which the integrand is evaluated. Thus, we sample locations $\mathbf{s}$ and $\mathbf{u}$ independently from a uniform distribution defined over B, compute the integrand of Equation (7) for this pair of locations, store the outcome, repeat the procedure many times and take the average of all outcomes. One must take care to use a sufficiently large Monte Carlo sample size, for which trial and error approaches can be used. In this paper, we used a Monte Carlo sample size of 10,000.
7. If the uncertainty of the block total instead of the block average is required then the result of step 6 must be multiplied by the square of the size of the study area, that is, by $|B|^2$.
8. Both for the block average and block total it is useful to take the square root of the variance and compute and plot the standard deviation of the prediction errors at block support because these are easier to interpret and have the same measurement units as the block predictions. Since the prediction error at block support is approximately normally distributed because of the central limit theorem, the lower and upper limits of a prediction interval of the block average or total can easily be computed, by subtracting and adding the product of the standard deviation and an appropriate z-score to the predicted block average or total. For ratio variables, these block support standard deviations may also be divided by the block support predictions to obtain a measure of the relative error (i.e. coefficient of variation).

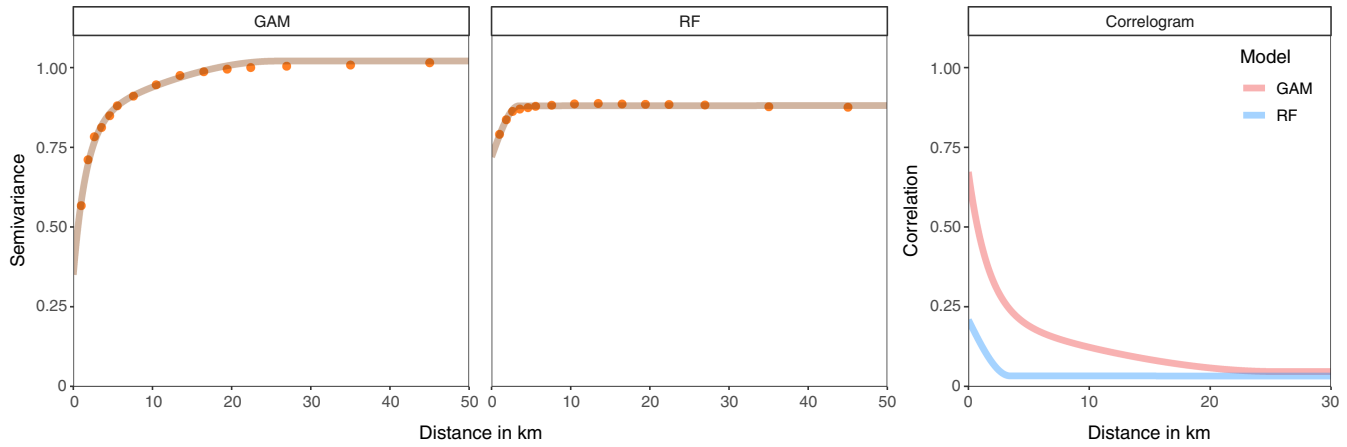

FIGURE 4 Sample and fitted variograms of generalized additive model (GAM) and regression forest (RF) standardized prediction errors and correlograms obtained from the fitted variograms. Note that we zoomed in to short distances because these are most relevant.

4.2 | Application to aboveground biomass mapping in Western Africa

We used a large dataset ($n = 59{,}867$) of tropical AGB in the Congo basin in Western Africa (Ploton et al., 2020). The dataset has average AGB estimates in $Mg\,ha^{-1}$ with block support of 1 km × 1 km. It has a minimum value of 14.3, a median of 286.1, a mean of 290.1 and a maximum value of 695.1 $Mg\,ha^{-1}$. The estimates are based on measured trees in field plots, which are standardized using a computation scheme detailed in Ploton et al. (2020). We further collected a set of spatially exhaustive environmental covariates representing mean climate conditions (annual mean temperature, precipitation, mean cloud cover frequency, water vapour pressure, solar radiation and evapotranspiration), climate seasonality (temperature and precipitation seasonality, standard deviation of monthly water vapour pressure and of monthly solar radiation), extreme climate conditions (mean temperature of driest and warmest quarter, precipitation of wettest and driest month, aridity index), topography (elevation and slope), soil (topsoil clay and sand contents, carbon concentration and stocks), MODIS long-term average reflectance (RED, NIR, SWIR 1–2) and vegetation (diversity of enhanced vegetation index, primary productivity). The sources of these covariates are provided in the supplementary data in Wadoux et al. (2021). All covariates were projected to the World Mercator system (EPSG: 3395), cropped to the extent of the measurements and either aggregated or resampled with bilinear interpolation to have the same resolution with grid cells of 1 km × 1 km. Hereafter, only areas corresponding to moist forests in the ESA CCI Land Cover dataset (ESA CCI, 2016) were considered.

With the values of AGB and their matching values of environmental covariates we applied the methodology described in Section 4.1. We fitted two models: a generalized additive model (GAM) and a QRF model (Meinshausen, 2006). The GAM model was fitted with cubic regression splines as a smoothing term. The QRF model was fitted with 250 trees with other hyperparameters held to their default value. Results of a 10-fold cross-validation showed that both models had little or no prediction bias (mean error close to zero), a root mean squared error of 74.09 and 64.18 $Mg\,ha^{-1}$, and a model efficiency coefficient (Janssen & Heuberger, 1995) of 0.37 and 0.54 for GAM and RF, respectively. In the R programming language (R Core Team, 2022) we used the packages mgcv (Wood, 2022) and gratia (Simpson & Singmann, 2022) for GAM and ranger (Wright & Ziegler, 2017) for QRF. For ranger the stochastic simulation was done using what = function(x) sample in the predict function. We estimated the prediction error standard deviation by sampling 1000 times from the GAM and QRF conditional distributions. The term 'conditional' means that the distributions are conditional to the training data. In other words, we derive a probability distribution of the target value at prediction locations such that it accounts for the information at training locations. Next, we standardized the prediction errors at the 59,857 measurement locations by dividing each prediction error by the associated prediction error standard deviation obtained by 10-fold cross-validation. We fitted nested variogram models (Wackernagel, 2003, chapter 15). For RF, we used two spherical functions whereas for GAM the nested variogram consisted of one spherical function and two exponential functions. For both RF and GAM variograms the sills and nuggets were fitted manually, and the ranges were estimated by unweighted ordinary least-squares. The two variograms were then converted to a correlation function. This completed step 4 of the method presented in Section 4.1.

To estimate the standard deviation of the prediction error of the spatial average (steps 5–8 in Section 4.1), we implemented a Monte Carlo numerical integration algorithm to estimate the double integral in Equation (7). To estimate the required Monte Carlo sample size, we used a trial-and-error approach where the AGB block standard deviation values for a given sample size were plotted against AGB block standard deviation values obtained with the same sample size, but using a different seed (after Nol et al., 2010, figure 7). We tested an increasing number of runs from 100 to 10,000. We considered 10,000 to be sufficient to obtain a stable outcome because the results were close to the 1:1 lines in a scatter plot.

Figure 4 shows the fitted nested variograms and corresponding correlation functions for GAM and RF standardized prediction

errors. Both cases showed a weak spatial autocorrelation of the standardized map errors. There was more spatial autocorrelation in the GAM error than in the RF error. The GAM variogram had a nugget of 0.35 and partial sills of 0.49, 0.05 and 0.18, and range parameters of 2.1, 7800 and 21 km for two exponential and one spherical function, respectively. The RF variogram had a nugget of 0.72 and partial sills of 0.16 and 0.03, and range parameters of 3.4 and 2245 km.

Figure 5 shows maps of the block standard deviation for the two models and various spatial aggregation levels (i.e. 10 km × 10 km, 50 km × 50 km and 100 km × 100 km blocks, as well as for the six countries in the basin). Spatial gaps occur because only results for moist forests are shown; larger square blocks were only included if their center was a moist forest. Table 1 shows the values of the total AGB and uncertainty for the six countries and the Congo basin.

Figure 5 shows that the prediction uncertainty decreases with increasing block support. For instance, while the standard deviations are between 10 and 41 Mg ha^{-1} for blocks of 10 km × 10 km, the standard deviations are always smaller than 20 Mg ha^{-1} for country averages. Table 1 shows that GAM predicts higher values of total AGB than RF and that its uncertainty is also larger than that of RF. The difference in prediction is likely predominately caused by the different assumptions made by the GAM and RF models. For instance, GAM assumes that the response variable is a linear combination of smooth functions of the explanatory variables, while RF is a tree-based model that is much more flexible in modelling non-linear relations between the response and explanatory variables. The uncertainty of the predicted total AGB for the entire Congo basin is 1713 Tg for RF and 2365 Tg for GAM.

5 | DISCUSSION

We explained that the uncertainty of spatial aggregates critically depends on the spatial autocorrelation of the map errors, and we described methods that account for this when predicting spatial averages and totals with associated uncertainty. We illustrated the methods with two real-world case studies. In the first case, we used kriging to obtain block prediction of OC at points, for French administrative units of increasing sizes, and for the whole of France. We found that the prediction uncertainty decreases when the support increases. This result is not new and corroborated by the existing literature (e.g. Szatmári et al., 2021). In the second case, we introduced a methodology to compute the uncertainty of spatial averages without kriging. The methodology is applicable to any mapping model that quantifies the point support uncertainty, including machine learning. We found that the standard deviations of GAM were higher

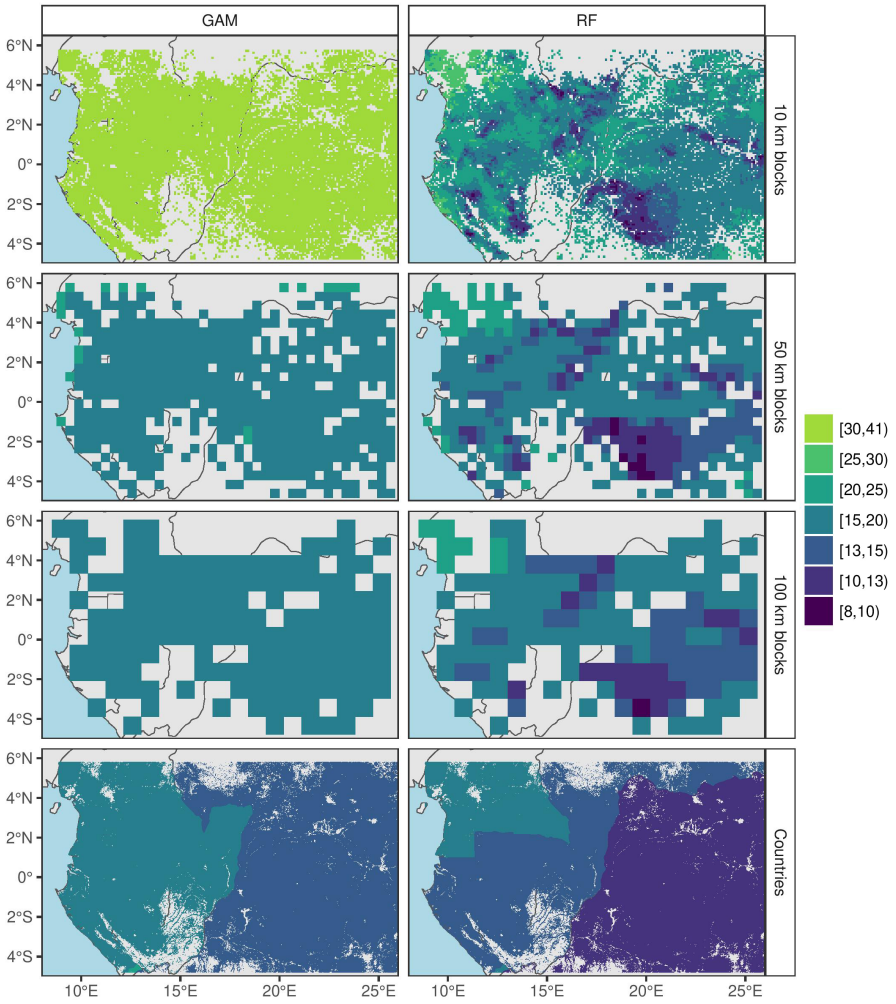

FIGURE 5 Prediction error standard deviation maps (Mg ha^{-1}) of the average above-ground biomass for various spatial aggregation levels, using the generalized additive model (GAM) and regression forest (RF) models.

TABLE 1 Total above-ground biomass and uncertainty (prediction ± standard deviation, in Teragram) obtained with the two prediction models and the methodology described in Section 4.1.

	RF	GAM
Angola	47 ± 3	36 ± 4
Cameroon	6066 ± 337	7205 ± 304
Central African Republic	1835 ± 98	1950 ± 112
Congo	6279 ± 301	6691 ± 330
Democratic Republic of the Congo	20,260 ± 933	23,288 ± 1110
Gabon	670 ± 41	667 ± 36
Equatorial Guinea	6707 ± 332	6954 ± 342
Total Congo basin	41,866 ± 1713	46,791 ± 2365

Abbreviations: GAM, generalized additive model; RF, regression forest.

than those of RF when mapping AGB in the Congo basin. This was no surprise because GAM explained less of the AGB spatial variation than RF. The correlograms of the standardized prediction errors revealed correlation in the map errors. There was also decreasing uncertainty for aggregation at larger support. The uncertainty of the totals at the support of the country and for the whole Congo basin was small because the correlation in the map errors was weak, which resulted in having most uncertainty cancelling out when averaging. Note that in this paper we quantified uncertainty by standard deviations but that prediction intervals can easily be computed from them because the block averages and totals are approximately normally distributed. Thus, the methodology also supports significance testing.

The method presented in Section 4.1 incorporates a non-stationary variance because it operates on standardized errors. The method is therefore more flexible than second-order stationary models that assume a constant variance (e.g. Wadoux et al., 2018), but it should be noted that the assumption that spatial correlation only depends on the separation distance between locations is still quite strong. In case of large datasets, flexibility could be further enhanced by allowing the correlation function to vary between regions, for example, by letting the parameters of $\rho(\mathbf{h})$ vary spatially. This would however require sufficient point data in each of the regions. In a case where the correlation length (i.e. the variogram range) is small, only short-distance pairs would be needed, because the correlation is zero for all distances larger than the range. This suggests that additional field work to collect data in undersampled regions is doable in many circumstances, also for large regions. In case of no or limited point data $\rho(\mathbf{h})$ may be derived with expert knowledge, for which protocols exist (Truong et al., 2013). However, expert judgement is no substitute for real data and collecting ground truth data to quantify the correlation function of standardized map errors is preferred. Note also that while our procedure quantifies the uncertainty of the block aggregate properly because it accounts for spatial autocorrelation of point support prediction errors, it does not exploit the spatial autocorrelation to reduce the uncertainty by residual kriging, as was done in Section 3.1. This means that the Monte Carlo integration method leads to larger uncertainties than block kriging, which is the price paid for using a computationally much more feasible method.

While our aim was to show that uncertainty of spatial aggregates cannot be assessed without accounting for the spatial autocorrelation of the map error, statistical validation of the computed uncertainty of the block averages or totals is also important. We recommend that this is done in future studies. Vaysse et al. (2017) performed a qualitative validation of the aggregated results by visual inspection of the aggregated soil map results and comparison with an existing soil map of the same property, but they did not evaluate the prediction uncertainty. A statistical validation at block support may also be obtained by taking a new, post-mapping composite/bulk sample in the 'block'. This is the preferred approach for validation. However, for applications where the block is very large (e.g. a bioclimatic region or a country), there might not be enough repetitions for validation of the prediction uncertainty. If the block is the entire study area there is only one measurement. This means that we cannot validate statistically whether the uncertainty was correctly assessed.

This paper took a model-based approach to quantify uncertainty of spatial averages and totals. However, it is important to note that the uncertainty may also be assessed with design-based inference (De Gruijter et al., 2006), if at least a sample collected with probability sampling is available from each block. Uncertainty estimates derived from design-based inference are attractive because they are model-free and design-unbiased. Two important conditions must be met to apply this approach. The first is that the inclusion probabilities of the sampling units that define the sample are deductible from the sampling design, and the second is that each unit or point in the population has a positive probability to be included in the sample. If one has a sample that satisfies these two conditions, inference about the target quantity (e.g. the average AGB or the average map error in a selected country) can be made with design-based inference (Brus et al., 2011).

Design-based statistical inference is usually used to estimate spatial averages or totals of environmental variables, such as the soil organic carbon stock or biomass, but it can also be used to estimate spatial averages or totals of map errors. For instance, the map average or total would be computed to obtain the estimate, while the associated uncertainty is estimated from a probability sample of map errors. For large supports design-based inference might be a more attractive option than model-based inference because it does not make model assumptions, but in practice, the requirements for a probability sample are often not satisfied for large-scale and global applications (with some exceptions, for example, in forestry and land cover mapping). In such a case, a model-based inference is the only valid option. For instance, there is no dataset of the AGB in the Congo basin obtained with probability sampling. The differences between model-based and design-based statistical inference for spatial mapping are discussed in Brus (2022). This text also explains model-assisted estimation, which is a design-based approach that benefits from model predictions.

Finally, we emphasize that the methodology described in this paper to obtain estimates of the uncertainty of spatial averages and totals is an improvement compared to the existing literature, where either perfect or complete absence of spatial autocorrelation of map errors is assumed (e.g. in Harris et al., 2021; Plaza et al., 2018). As already mentioned in the Introduction, such assumptions are highly unrealistic and will lead to improper estimates of the uncertainty of spatial averages and totals, preventing assessment of the statistical significance of differences between regions or changes over time. To give one example, in the verification steps of Monitoring, Reporting and Verification studies (IPCC, 2022; UNFCCC, 2014) parties need to evaluate if climate mitigation actions were successful by showing that a predicted increase of carbon stock for a region in a given time period is statistically significant. One valid approach for that is design-based statistical inference as explained above, but it requires sufficiently large probability samples at the start and end of the period, which may be too costly. The alternative is to take a model-based approach as done in this study, but then it is essential that uncertainties of spatial averages and totals are correctly derived, by accounting for spatial autocorrelation of map errors.

6 | CONCLUSION

We have shown that the uncertainty of spatial aggregates strongly depends on the spatial autocorrelation of the map errors. This usually is not accounted for in the literature. We conclude that:

- Ignoring spatial autocorrelation in map errors leads to serious underestimation of the uncertainty of spatial averages and totals.
- Solutions based on geostatistical modelling and block kriging are perfectly suited for uncertainty quantification of averages and totals, but such methods require geostatistical expertise, make stringent assumptions and are computationally inefficient for large-scale applications.
- Accounting for spatial autocorrelation in map errors is possible by mathematical integration of the variances and covariances of the point support prediction errors. We propose such integration over a spatial block or region using numerical Monte Carlo integration. We tested this method on a real-world case study and showed that it is feasible for large case studies.
- The method that we proposed does not require stationarity of the variance and allows non-Gaussian error distributions. With some modifications, the model-based methodology proposed here should also be applicable to categorical variables.
- We found that uncertainty of spatial averages decreased with increasing support size, particularly when the spatial autocorrelation of map errors is weak.
- In estimating the total aboveground biomass of our study area in the Congo basin, we found an average value of 41,866 Tg with a standard deviation of 1713 Tg (relative error of 4%) for the prediction made with the random forest algorithm.
- The method presented in this paper is an essential tool for parties that need to assess the accuracy of predictions of spatial averages and totals and that need to evaluate whether differences in aggregated environmental variables or natural resources between regions or over time are statistically significant.
- In future studies, the spatial aggregation method based on Monte Carlo integration could be further improved by relaxing the stationarity assumption in the correlation function and by including expert knowledge in the estimation of the correlation function.

AUTHOR CONTRIBUTIONS

Alexandre M. J.-C. Wadoux and Gerard B. M. Heuvelink conceived the ideas and designed methodology, collected the data and developed the methods. All authors contributed critically to the drafts and gave final approval for publication.

FUNDING INFORMATION

No funding was received to carry out this study.

ACKNOWLEDGEMENTS

Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest concerning the content of the manuscript.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14106.

DATA AVAILABILITY STATEMENT

Data used came from freely available data sources archived in https://doi.org/10.15454/AIQ9WS for the example of Section 3.2 (Saby et al., 2020), while the dataset used in Section 4.2 is available through figshare https://doi.org/10.6084/m9.figshare.11865450 with reference to Ploton et al. (2020).

ORCID

Alexandre M. J.-C. Wadoux https://orcid.org/0000-0001-7325-9716

Gerard B. M. Heuvelink https://orcid.org/0000-0003-0959-9358

REFERENCES

Ahlström, A., Schurgers, G., Arneth, A., & Smith, B. (2012). Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. *Environmental Research Letters*, *7*, 044008.

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P. S. A., Dubayah, R., Friedl, M. A., Samanta, S., & Houghton, R. A. (2012). Estimated carbon dioxide

emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, *2*, 182–185.

Berthelot, M., Friedlingstein, P., Ciais, P., Dufresne, J.-L., & Monfray, P. (2005). How uncertainties in future climate change predictions translate into future terrestrial carbon fluxes. *Global Change Biology*, *11*, 959–970.

Brus, D. J. (2022). *Spatial sampling with R*. CRC Press.

Brus, D. J., Kempen, B., & Heuvelink, G. B. M. (2011). Sampling for validation of digital soil maps. *European Journal of Soil Science*, *62*, 394–407.

Cook-Patton, S. C., Leavitt, S. M., Gibbs, D., Harris, N. L., Lister, K., Anderson-Teixeira, K. J., Briggs, R. D., Chazdon, R. L., Crowther, T. W., Ellis, P. W., Griscom, H. P., Herrmann, V., Holl, K. D., Houghton, R. A., Larrosa, C., Lomax, G., Lucas, R., Madsen, P., Malhi, Y., … Griscom, B. W. (2020). Mapping carbon accumulation potential from global natural forest regrowth. *Nature*, *585*, 545–550.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

De Gruijter, J., Brus, D. J., Bierkens, M. F. P., & Knotters, M. (2006). *Sampling for natural resource monitoring*. Springer Science & Business Media.

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K., & Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science*, *359*, 320–325.

Eggleston, H. S., Buendia, L., Miwa, K., Ngara, T., & Tanabe, K. (2006). *2006 IPCC guidelines for national greenhouse gas inventories*. Institute for Global Environmental Strategies (IGES) for the IPCC.

ESA CCI. (2016). *New release of 300 m global land cover and 150 m water products (v. 1.6. 1) and new version of the user tool (3.10) for download*. ESA CCI Land Cover Website.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302–4315.

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

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*, 850–853.

Harris, N. L., Gibbs, D. A., Baccini, A., Birdsey, R. A., De Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M. C., Herold, M., Houghton, R. A., Potapov, P. V., Suarez, D. R., Roman-Cuesta, R. M., Saatchi, S. S., Slay, C. M., Turubanova, S. A., & Tyukavina, A. (2021). Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change*, *11*, 234–240.

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G., Walsh, M. G., & Gonzalez, M. R. (2014). Soilgrids1km—Global soil information based on automated mapping. *PLoS One*, *9*, e105992.

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

Heuvelink, G. B. M. (1998). *Error propagation in environmental modelling with GIS*. CRC Press.

Heuvelink, G. B. M., & Webster, R. (2022). Spatial statistics and soil mapping: A blossoming partnership under pressure. *Spatial Statistics*, *50*, 100639.

IPCC. (2022). Summary for policymakers. In P. Shukla, J. Skea, R. Slade, A. A. Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, & J. Malley (Eds.), *Climate change 2022: Mitigation of climate change. Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change* (pp. 3–33). Cambridge University Press.

Isaaks, E. H., & Srivastava, M. R. (1989). *Applied geostatistics*. Oxford University Press.

Janssen, P. H. M., & Heuberger, P. S. C. (1995). Calibration of process-oriented models. *Ecological Modelling*, *83*, 55–66.

Kempen, B., Dalsgaard, S., Kaaya, A. K., Chamuya, N., Ruipérez-González, M., Pekkarinen, A., & Walsh, M. G. (2019). Mapping topsoil organic carbon concentrations and stocks for Tanzania. *Geoderma*, *337*, 164–180.

Kros, J., Heuvelink, G. B. M., Reinds, G. J., Lesschen, J. P., Ioannidi, V., & Vries, W. D. (2012). Uncertainties in model predictions of nitrogen fluxes from agro-ecosystems in Europe. *Biogeosciences*, *9*, 4573–4588.

Kullberg, P., & Moilanen, A. (2014). How do recent spatial biodiversity analyses support the convention on biological diversity in the expansion of the global conservation area network? *Natureza & Conservacao*, *12*, 3–10.

Lark, R. M. (2000). Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science*, *51*, 717–728.

Lesschen, J. P., Stoorvogel, J. J., Smaling, E. M. A., Heuvelink, G. B. M., & Veldkamp, A. (2007). A spatially explicit methodology to quantify soil nutrient balances and their uncertainties at the national level. *Nutrient Cycling in Agroecosystems*, *78*, 111–131.

Lugato, E., Panagos, P., Bampa, F., Jones, A., & Montanarella, L. (2014). A new baseline of organic carbon stock in european agricultural soils using a modelling approach. *Global Change Biology*, *20*, 313–326.

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). Modis collection 6 Maiac algorithm. *Atmospheric Measurement Techniques*, *11*, 5741–5765.

Ma, H., Mo, L., Crowther, T. W., Maynard, D. S., van den Hoogen, J., Stocker, B. D., Terrer, C., & Zohner, C. M. (2021). The global distribution and environmental drivers of aboveground versus belowground plant biomass. *Nature Ecology & Evolution*, *5*, 1110–1122.

McRoberts, R. E., Næsset, E., Saatchi, S., & Quegan, S. (2022). Statistically rigorous, model-based inferences from maps. *Remote Sensing of Environment*, *279*, 113028.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*, 983–999.

Mokany, K., Ferrier, S., Harwood, T. D., Ware, C., Di Marco, M., Grantham, H. S., Venter, O., Hoskins, A. J., & Watson, J. E. (2020). Reconciling global priorities for conserving biodiversity habitat. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 9906–9911.

Mulder, V. L., Lacoste, M., Richer-de Forges, A., & Arrouays, D. (2016). GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth. *Science of the Total Environment*, *573*, 1352–1369.

Nol, L., Heuvelink, G. B. M., Veldkamp, A., de Vries, W., & Kros, J. (2010). Uncertainty propagation analysis of an n2o emission model at the plot and landscape scale. *Geoderma*, *159*, 9–23.

Oliver, M. A., & Webster, R. (2015). *Geostatistical prediction: Kriging*. Springer.

Padarian, J., Stockmann, U., Minasny, B., & McBratney, A. B. (2022). Monitoring changes in global soil organic carbon stocks from space. *Remote Sensing of Environment*, *281*, 113260.

Plaza, C., Zaccone, C., Sawicka, K., Méndez, A. M., Tarquis, A., Gascó, G., Heuvelink, G. B. M., Schuur, E. A., & Maestre, F. T. (2018). Soil resources and element stocks in drylands to face global issues. *Scientific Reports*, *8*, 1–8.

Ploton, P., Mortier, F., Barbier, N., Cornu, G., Réjou-Méchain, M., Rossi, V., Alonso, A., Bastin, J.-F., Bayol, N., Benedet, F., Bissiengou, P., Chuyong, G., Demarquez, B., Doucet, J.-L., Droissart, V., Kamdem, N. G., Kenfack, D., Memiaghe, H., Moses, L., … Gourlet-Fleury, S. (2020). A map of African humid tropical forest aboveground biomass derived from management inventories. *Scientific Data*, *7*, 1–13.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). Soilgrids 2.0: Producing soil

information for the globe with quantified spatial uncertainty. *The Soil*, *7*, 217–240.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna. https://www.R-project.org/

Rabus, B., Eineder, M., Roth, A., & Bamler, R. (2003). The shuttle radar topography mission—A new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*, 241–262.

Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). Springer.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9899–9904.

Saby, N. P. A., Chenu, J.-P., Szergi, T., Csorba, A., Bertuzzi, P., Toutain, B., Picaud, C., Gay, L., & Creamer, R. (2020). *French RMQS soil profile and monitoring dataset with related management practices data*. https://doi.org/10.15454/AIQ9WS

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabile, V., Araza, A., De Bruin, S., Herold, M., Quegan, S., Rodrguez-Veiga, P., Balzter, H., Carreiras, J., Schepaschenko, D., Korets, M., Shimada, M., Itoh, T., Martínez, Á. M., Cavlovic, J., Gatti, R. C., … Willcock, S. (2021). The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth System Science Data*, *13*, 3927–3950.

Schmidt-Traub, G. (2021). National climate and biodiversity strategies are hamstrung by a lack of maps. *Nature Ecology & Evolution*, *5*, 1325–1327.

Shafer, G., & Vovk, V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research*, *9*, 371–421.

Simpson, G. L., & Singmann, H. (2022). *gratia: Graceful 'ggplot'-based graphics and other functions for GAMs fitted using 'mgcv'*. R package version 0.7.3. https://CRAN.R-project.org/package=gratia

Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., & Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, *560*, 639–643.

Spawn, S. A., Sullivan, C. C., Lark, T. J., & Gibbs, H. K. (2020). Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Scientific Data*, *7*, 1–22.

Szatmári, G., Pásztor, L., & Heuvelink, G. B. M. (2021). Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. *Geoderma*, *403*, 115356.

Truong, P. N., Heuvelink, G. B. M., & Gosling, J. P. (2013). Web-based tool for expert elicitation of the variogram. *Computers & Geosciences*, *51*, 390–399.

UNFCCC. (2014). *Handbook on measurement, reporting and verification for developing country parties*. United Nations Climate Change Secretariat Bonn.

Van der Westhuizen, S., Heuvelink, G. B. M., Hofmeyr, D. P., & Poggio, L. (2022). Measurement error-filtered machine learning in digital soil mapping. Spatial statistics 47, 100572. *Spatial Statistics*, *47*, 100572.

Vaysse, K., Heuvelink, G. B. M., & Lagacherie, P. (2017). Spatial aggregation of soil property predictions in support of local land management. *Soil Use and Management*, *33*, 299–310.

Wackernagel, H. (2003). *Multivariate Geostatistics: An introduction with applications* (3rd ed.). Springer.

Wadoux, A. M. J.-C., Brus, D. J., & Heuvelink, G. B. M. (2018). Accounting for non-stationary variance in geostatistical mapping of soil properties. *Geoderma*, *324*, 138–147.

Wadoux, A. M. J.-C., Heuvelink, G. B. M., De Bruin, S., & Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling*, *457*, 109692.

Wadoux, A. M. J.-C., Padarian, J., & Minasny, B. (2019). Multi-source data integration for soil mapping using deep learning. *The Soil*, *5*, 107–119.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. John Wiley & Sons.

Wiesmeier, M., Barthold, F., Blank, B., & Kögel-Knabner, I. (2011). Digital mapping of soil organic matter stocks using random forest modeling in a semi-arid steppe ecosystem. *Plant and Soil*, *340*, 7–24.

Wood, S. (2022). *mgcv: Mixed GAM computation vehicle with automatic smoothness estimation*. R package version 1.8-40. https://CRAN.R-project.org/package=mgcv

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*, 1–17.

How to cite this article: Wadoux, A.-C., & Heuvelink, G. B. M. (2023). Uncertainty of spatial averages and totals of natural resource maps. *Methods in Ecology and Evolution*, *14*, 1320–1332. https://doi.org/10.1111/2041-210X.14106