ORIGINAL ARTICLE

Mapping depth to Pleistocene sand with Bayesian generalized linear geostatistical models

Luc Steinbuch1,2,3 | Dick J. Brus4 | Gerard B. M. Heuvelink1,2

1Soil Geography and Landscape Group, Wageningen University and Research, Wageningen, the Netherlands
2ISRIC–World Soil Information, Wageningen, the Netherlands
3Wageningen Environmental Research, Wageningen University and Research, Wageningen, the Netherlands
4Biometris, Wageningen University and Research, Wageningen, the Netherlands

Correspondence
Luc Steinbuch, Soil Geography and Landscape Group, Wageningen University and Research, Wageningen, the Netherlands.
Email: luc.steinbuch@wur.nl

Abstract
Spatial soil applications frequently involve binomial variables. If relevant environmental covariates are available, using a Bayesian generalized linear model (BGLM) might be a solution for mapping such discrete soil properties. The geostatistical extension, a Bayesian generalized linear geostatistical model (BGLGM), adds spatial dependence and is thus potentially better equipped. The objective of this work was to evaluate whether it pays off to extend from a BGLM to a BGLGM for mapping binary soil properties, evaluated in terms of prediction accuracy and modelling complexity. As motivating example, we mapped the presence/absence of the Pleistocene sand layer within 120 cm from the land surface in the Dutch province of Flevoland, using the BGLM implementation in the R-package geoRglm. We found that BGLGM yields considerably better statistical validation metrics compared to a BGLM, especially with – as in our case – a large (n = 1,000) observation sample and few relevant covariates available. Also, the inferred posterior BGLM parameters enable the quantification of spatial relationships. However, calibrating and applying a BGLM is quite demanding with respect to the minimal required sample size, tuning the algorithm, and computational costs. We replaced manual tuning by an automated tuning algorithm (which eases implementing applications) and found a sample composition that delivers meaningful results within 50 h calculation time. With the gained insights and shared scripts spatial soil practitioners and researchers can – for their specific cases – evaluate if using BGLGM is feasible and if the extra gain is worth the extra effort.

Highlights
- Does adding spatial correlation to a Bayesian GLM for mapping a binary soil variable pay off?
- We aim to make spatial Bayesian hierarchical modelling accessible for pedometricians.
- Most hierarchical models work well when enough observations are provided, even without covariates.
Including spatial correlation might sometimes be worth the extra effort and computational costs.

KEYWORDS
Bayesian statistics, digital soil mapping, generalized linear model, geoRglm, model-based geostatistics, spatial statistics

1 | INTRODUCTION

In many soil mapping applications, spatial variables are continuous and can, perhaps after a transformation (Müller, 2007), conveniently be modelled using Gaussian spatial models. However, some spatial soil applications involve discrete count variables or categorical variables (e.g., Kempen, Brus, & Heuvelink, 2012; Malone, Minasny, & McBratney, 2017). In other cases, the end-user is only interested in information relative to threshold values (e.g., Lark & Ferguson, 2004). By definition, discrete and categorical variables cannot be transformed to normality. With such variables, non-linear methods such as indicator kriging (Journel, 1983) deliver reasonable results, but are suboptimal in the case of a trend (Papritz, 2009) and lack model-based consistency, which implies, among others, that they are unsuited to facilitate kriging with change of support (Emery & Ortiz, 2004).

If environmental covariates are available for the whole area of interest related to the target variable, using a generalized linear model (GLM) might be a solution for mapping discrete soil properties. GLM assumes that the actual observations are a realization of a discrete random process, such as Bernoulli, binomial or Poisson. Given this assumption, GLMs use, for each location, a transformation of a virtual continuous variable – the linear predictor – to the distribution parameter. For instance, the logit transformation links the linear predictor to the probability of success for a Bernoulli process. The linear predictor is the linear combination of covariate values (including an intercept) scaled by regression parameters (Myers, Montgomery, & Vining, 2002, section 4.2).

However, when applied to spatial data, GLM ignores any spatial dependencies, other than those contained in the covariates. The geostatistical extension, a generalized linear geostatistical model (GLGM), adds spatial dependence to the model and linear predictor, by means of a spatially correlated Gaussian field (Diggle & Ribeiro, 2007). Therefore, compared to a GLM, a GLGM is more generic and flexible, and thus potentially better equipped, provided that its spatial dependence parameters can be estimated well. However, GLGMs also require more skills of the modeller (as calibrating spatial parameters is less straightforward than calibrating regression parameters) and are computationally more expensive.

Another possible extension is application of Bayesian statistics. Bayesian statistics demands explicit incorporation of pre-observation knowledge, even if this means an explicit definition of our ignorance (Lindley, 2004). It also considers all model parameters to be stochastic quantities, thus parameter uncertainty is taken into account (McElreath, 2016). Both Bayesian properties are applied in soil mapping by, for example, Steinbuch, Brus, and Heuvelink (2018) and Poggio, Gimona, Spezia, and Brewer (2016). Both properties together enable the construction of hierarchical models in a convenient and statistically sound way (Gelman, Carlin, Stern, & Rubin, 2013), which can be very useful in spatial statistics (Banerjee, Carlin, & Gelfand, 2004). In our context, both GLMs and GLGMs have Bayesian extensions, abbreviated to BGLM and BGLGM, respectively. A BGLGM is a spatial implementation of a Bayesian hierarchical model (Diggle & Ribeiro, 2007).

The objective of this work was (a) to make spatial Bayesian hierarchical modelling accessible and understandable for pedometricians, and (b) to evaluate whether it pays off to extend from a BGLM to a BGLGM for mapping binary soil properties, when evaluated in terms of prediction accuracy, modelling effort and computational costs. We pay due attention to software implementation issues and provide scripts in Appendix S1, because especially a BGLGM cannot easily be programmed from scratch, thus we need to rely on existing software libraries and functions.

As motivating example, we map the depth of the Pleistocene cover sand layer relative to the surface in the Dutch province of Flevoland. This depth is defined as a Bernoulli variable obtained by determining whether the depth is above or below 1.2 m.

2 | THEORY

We model a spatial process for all locations \( s \in A \), where \( A \subset \mathbb{R}^2 \) is the study area. As a first modelling step, we
state that at every $s$, $z(s)$ is a binary realization (zero or one) of the Bernoulli distribution:

$$Z(s) \sim \text{Bernoulli}(p(s)), \quad (1)$$

where parameter $p(s) \in [0,1]$ indicates the probability of success. We further assume that $Z$ is an independent process, that is, $Z(s)$ and $Z(s')$ are statistically independent for all $s, s' \in A, s \neq s'$.

This spatial process $Z = \{Z(s), s \in A\}$ is observed at $n$ locations $s_i$, $i = 1, 2, \ldots, n$, providing observation vector $\mathbf{z} = [z(s_1) z(s_n)]^T$, being a realization of $Z = [Z(s_1) Z(s_n)]^T$. We model $p(s)$, and thus $Z(s)$, initially with a generalized linear model (GLM), as explained in the following section. Next, we extend this GLM to a Bayesian generalized linear model (BGLM) and a Bayesian generalized linear geostatistical model (BGLGM), the two models to be compared in this research.

### 2.1 | BGLM

#### 2.1.1 | Generalized linear model (GLM)

We base the explanation of GLM on textbooks such as Myers et al. (2002). The logit transformation projects $p(s)$ to the mathematically more convenient $y(s)$:

$$y(s) = \log \left( \frac{p(s)}{1 - p(s)} \right) = \logit(p(s)), \quad (2)$$

while the inverse logit function derives $p(s)$ from $y(s)$:

$$p(s) = \frac{\exp(y(s))}{\exp(y(s)) + 1} = \frac{1}{1 + \exp(-y(s))} = \logit^{-1}(y(s)). \quad (3)$$

The unobservable variable $y(s)$, $s \in A$ represents the “signal” over the study area. The vector $\mathbf{y} = [y(s_1) y(s_n)]^T$ indicates the signal on the observation locations. In a GLM, $\mathbf{y}$ is modelled as a linear predictor:

$$\mathbf{y} = \mathbf{X} \beta. \quad (4)$$

Here, $\mathbf{X}$ indicates the $k \times n$ design matrix, containing leading ones and covariate values at all observation locations. The vector of $k$ regression coefficients, including the intercept, is given as $\beta$. Later we discuss how we can calibrate this model using the vector of observations $\mathbf{z}$. Note that in the GLM defined above (but not in the other models considered next), all randomness is captured in Equation (1).

#### 2.1.2 | Bayesian extension: BGLM

To include parameter uncertainty, we extend the GLM by considering $\beta$ a stochastic parameter vector (but note that in this work we assume neither $s$ nor $\mathbf{X}$ to be uncertain). To work conveniently with stochastic parameters we shape our models using Bayesian statistics – we assume that the reader is familiar with concepts such as Bayes’ rule, prior (conjugate or otherwise), conditional distributions, likelihood, posterior, the posterior predictive distribution and Markov Chain Monte Carlo (MCMC) simulation; for example textbooks, see Banerjee et al. (2004), Gelman et al. (2013) or McElreath (2016). In the following, we introduce a Bayesian GLM (BGLM).

The posterior density of $\beta$, conditional on the observations, is given by Bayes’ rule:

$$f(\beta|\mathbf{z}) \propto f(\mathbf{z}|\beta)f(\beta), \quad (5)$$

with $f(\mathbf{z}|\beta)$ the posterior probability distribution of $\beta$ conditional on $\mathbf{z}$, containing all our knowledge after taking the observations. The prior $f(\beta)$ contains our pre-observation knowledge; it might be a mathematical abstraction of our assumed ignorance. The likelihood of the observations conditional on model parameter $f(\mathbf{z}|\beta)$ follows directly from the Bernoulli probability mass function and can, while incorporating Equations (3) and (4), be expressed as:

$$f(\mathbf{z}|\beta) = \prod_{i=1}^{n} p_i^{1} \{1 - p_i\}^{1-z_i} = \prod_{i=1}^{n} \logit^{-1}(X\beta)^{z_i} \{1 - \logit^{-1}(X\beta)\}^{1-z_i}. \quad (6)$$

In Section 3.1 we elaborate on the choice of $f(\beta)$ and on how we implemented an algorithm to infer $f(\beta|\mathbf{z})$.

#### 2.1.3 | Prediction with a BGLM

To derive a prediction map of the probabilities of success, we first define $n'$ prediction locations $s_i, i = n + 1, \ldots, n + n'$. Typically, these are the nodes of a grid covering the study area, for which all covariates must be available. The prediction probabilities at the prediction locations result from back-transforming the signal at these locations using Equation (3):

$$p' = \logit^{-1}(\mathbf{y}^*). \quad (7)$$

We therefore require the vector $\mathbf{y}^* = [y(s_{n+1}) y(s_{n+n'})]^T$ of signals at prediction locations.
In contrast to GLM, in a BGLM $y^*$ is stochastic. The posterior probability distribution, referred to as the “posterior predictive”, equals:

$$f(y^*|z) = \int f(y^*, \beta|z)\beta$$

$$= \int f(y^*|\beta, z)f(\beta|z)\beta$$

$$= \int f(y^*|\beta)f(\beta|z)\beta.$$  \hspace{1cm} (8)

Note that the last identity holds because $y^*$ is completely characterized by the covariates and $\beta$, given $\beta$, $y^*$ and $z$ are independent. In Section 3.1 we show the implementation of the BGLM prediction.

### 2.2 | BGLM

#### 2.2.1 | Generalized linear geostatistical model (GLGM)

In a generalized linear geostatistical model (GLGM) (Banerjee, Carlin, & Gelfand, 2015; Diggle & Ribeiro, 2007), Equation (4) is extended by adding a spatially correlated random effect:

$$Y = X\beta + U.$$  \hspace{1cm} (9)

Here, $U$ is a vector of random variables $U(s_i), i = 1, \ldots, n$, taken from a stochastic spatial process $U$, which is modelled as a zero mean, stationary Gaussian random field defined over $A$. Note that we use again $Y$, and its realization $y$, for the signal, although they are modelled differently in a GLM and GLGM.

The spatial structure of $U$ in this research characterized by an exponential correlation function:

$$c(\phi, h, \eta^2) = \begin{cases} 1 & h = 0 \\ \frac{1}{1 + \eta^2} \exp \left( -\frac{h}{\phi} \right) & h > 0 \end{cases},$$  \hspace{1cm} (10)

where $h$ indicates the Euclidean distance between locations $s$ and $s + h \in A$, and $\phi$ is in geostatistical context called the “range parameter” or “distance parameter”. The “nugget to partial sill ratio” $\eta^2$ is given by:

$$\eta^2 = \frac{\tau^2}{\sigma^2},$$  \hspace{1cm} (11)

where parameter $\tau^2$ captures short-distance variation, that is, the “nugget”, while $\sigma^2$ (“partial sill” in geostatistics) is the variance of $U$, minus the variance of $U$ already captured by $\tau^2$. Note that when $\tau^2 = 0, \eta^2 = 0$ and the correlation function Equation (10) on very short distances approximates one.

In this research, we made simplifying choices regarding the spatial covariance function $\sigma^2 c(\phi, h, \eta^2)$, such as not to consider spatial anisotropy ($c$ is a function of Euclidean distance only, not of direction) and not to consider other covariance functions.

Each element of the $n \times n$ correlation matrix $C(\phi, h, \eta^2)$ is given by the correlation function Equation (10) corresponding to the distance between two observation locations. With $C(\phi, h, \eta^2)$, we can extend Equation (9) with spatial parameters:

$$Y = X\beta + U$$

$$\sim \text{MVN}(X\beta, \sigma^2 C(\phi, h, \eta^2)).$$  \hspace{1cm} (12)

For notational convenience, in what follows we often merge spatial and regression parameters into one parameter vector $\theta = (\beta, \sigma^2, \phi, \eta^2)$, and we will often indicate $C(\phi, h, \eta^2)$ by $C(\phi)$ or $C$.

#### 2.2.2 | Bayesian extension: BGLGM

In the Bayesian extension of the GLGM we consider $\theta$ stochastic. $Y$ is stochastic as well, due to $\beta$ and the spatial signal $U$. We therefore need their joint distribution. The joint posterior density is, according to Bayes’ rule and the chain rule of conditional probability, proportional to:

$$f(y, \theta|z) \propto f(z|y, \theta) f(y, \theta)$$

$$= f(z|y) f(y|\theta),$$  \hspace{1cm} (13)

with $f(y|\theta)$ the proportional density of the signal conditional on the parameters. Note that $f(z|y, \theta) = f(z|y)$ because conditional on $y, z$ and $\theta$ are independent. Equation (13) is often referred to as a “hierarchical model”, referring to the three related levels of probability distributions (Banerjee et al., 2004).

Equation (13) shows three terms to be inferred or defined: (a) the likelihood of the observations, conditional on the signal, which follows from Equations (1) and (3) and is much like Equation (6):

$$f(z|y) = \prod_{i=1, n} \beta_i \{1 - p_i\}^{1-z_i}$$

$$= \prod_{i=1, n} \logit^{-1}(y_i) \{1 - \logit^{-1}(y_i)\}^{1-z_i},$$  \hspace{1cm} (14)

(b) the likelihood of the signal, conditional on the parameters, which is given by the MVN probability density distribution:
prediction on new locations. In the next section we first extend the above theory to the prior of $\theta$ by integrating out the signal:

$$f(\theta) = \int f(y, \theta) dy.$$  

(16)

In the next section we first extend the above theory to prediction on new locations.

### 2.2.3 Prediction with a BGLM

To predict the signal $y'$ conditional on $z$ within a BGLM context, we formulate the posterior predictive distribution as an integral over both the parameters $\theta$ and the signal at the observation locations $y$:

$$f(y'|z) = \int \int f(y', \theta, y|z) d\theta dy$$

$$= \int \int f(y'| \theta, y, z)f(\theta, y|z) d\theta dy$$

(17)

$$= \int \int f(y'| \theta, y)f(\theta, y|z) d\theta dy.$$

Note that $y'$ and $z$ are conditionally independent given $\theta$ and $y$. In the final expression of Equation (17), we can consider $f(\theta, y|z)$ (the joint posterior density of parameters and signal as provided by Equation (13)) as weights to derive $f(y'|z)$ from $f(y'| \theta, y)$. The other component of the integrand, $f(y'| \theta, y)$, the distribution of a Gaussian response at prediction locations given observations and model parameters, boils down to simple kriging. This is because $y'$ and $y$ are part of the joint multivariate normal distribution:

$$\begin{bmatrix} y | \beta, \sigma^2, \tau^2, \phi \\ y' | \beta, \sigma^2, \tau^2, \phi \end{bmatrix} \sim MVN \left( \begin{bmatrix} X \\ X^* \end{bmatrix} \beta, \sigma^2 \begin{bmatrix} C(\phi) & C^*(\phi) \\ C^*(\phi)^T & C^{**}(\phi) \end{bmatrix} \right),$$

(18)

where $C^*(\phi)$ indicates the correlations between the signals at observation and prediction locations and $C^{**}(\phi)$ the correlation matrix of $y^*$, both of which can be calculated using Equation (10) and the geographic distance between the relevant locations. The simple kriging predictor, for a given realization of model parameters and signal, equals a Gaussian distribution (Searle, 1997, p. 47) with mean:

$$\hat{y}^* = X^* \beta + C^*(\phi)^T C(\phi)^{-1} (y - X\beta)$$  

(19)

and with prediction error variance:

$$\text{var} \left( y^* - \hat{y}^* \right) = \sigma^2 \left( 1 - C^*(\phi)^T C(\phi) \right)^{-1} C^*(\phi).$$

(20)

In theory, one could calculate both the prediction probability density as provided above and the corresponding weight factor $f(\theta, y|z)$ in Equation (17) for every possible combination of parameters and signal at the observation locations, and numerically integrate the outcomes to infer the posterior predictive distribution $f(y'|z)$. However, even if the space spanned by the model parameters and signal $y$ would be discretized by a coarse grid, this would be prohibitive because of the high dimension of the parameter-signal space. In Section 3.2 we discuss how to relieve the burden of computational costs and show how the actual inference of posterior parameters and posterior predictive was carried out in this research.

## 3 IMPLEMENTATION

### 3.1 BGLM implementation

We chose to apply a prior $f(\beta)$ indicating our pre-observation total ignorance about $\beta$. Because $\beta$ is a centrality parameter, this prior might be considered “flat” or “improper uniform” and can from a mathematical point of view be left out from Equation (5). This means that the posterior $\beta$ for a BGLM is proportional to the likelihood (expressed as a function of $\beta$) as derived for GLM. According to Kutner et al. (2005, section 14.5) and Myers et al. (2002, sections 4.3 and 4.4.1) this likelihood proportionally approaches an MVN distribution in the case of large sample sizes, where the mean of this MVN distribution can be estimated by a maximum likelihood estimator using the iteratively reweighted least squares algorithm, and the corresponding variances and covariances are estimated via the Fisher information matrix. For these calculations, we applied the algorithm as implemented in the base function glm in the statistical programming language R (R Core Team, 2017). The above mathematical and numerical steps are illustrated in Figure 1. We obtain a prediction $p^*$ by substituting the maximum likelihood estimate of $\beta$ in Equation (4), with $X$ replaced by $X^*$ and the result of that in Equation (7).
3.2 | BGLGM implementation

For BGLGM inference and prediction we shaped our approach around several functions from the geoRglm R package, which is developed to calibrate generalized linear spatial models and thus enable two-dimensional spatial prediction of binomial- and Poisson-distributed variables of interest (Christensen & Ribeiro Jr, 2002, 2015). These functions numerically approximate the posterior distribution of signal and model parameters by simulating a large sample from the posterior. However, sampling from a posterior can be computationally expensive, especially if – as in our case – the posterior contains a signal of size \( n \) (the number of observations), making it necessary to sample from a high-dimensional combined parameter and signal space. To lower computational costs, several extensions and simplifications were implemented by Christensen, Roberts, and Sköld (2006), to be discussed in the following subsections together with several choices we made for this research. Appendix S1 elaborates on the applied mathematics. The corresponding workflow, showing analytical and numerical processing steps between the boxed representations, is illustrated in Figure 2. Our embedding of geoRglm follows in Section 3.2.3.

3.2.1 | Choice of BGLGM priors: Restrict nugget and range parameter

We assume that the marginal prior \( \phi \), the marginal prior \( \tau^2 \) and the combined prior of \( \beta \) and \( \sigma^2 \) are independent:

\[
f(\theta) = f(\phi) f(\tau^2) f(\beta, \sigma^2).
\]

We write the joint prior for \( \beta \) and \( \sigma^2 \) as a product of conditional and marginal densities:

\[
f(\beta, \sigma^2) = f(\beta | \sigma^2) f(\sigma^2).
\]

Next, following the ideas of Diggle and Ribeiro (2007), we assign as combined prior a multivariate normal (MVN) and inverse scaled chi-squared (\( \chi^2_{\text{ScI}} \)) distribution, respectively:

\[
\begin{align*}
f(\beta | \sigma^2) &= \text{MVN}(\xi_0, \sigma^2D_0) \\
f(\sigma^2) &= \chi^2_{\text{ScI}}(\nu_0, \varsigma^2_0).
\end{align*}
\]

From this we derive the distribution of \( \beta \) and \( \sigma^2 \) conditional on \( y \) and \( \phi - f(\beta, \sigma^2 | y, \phi) \) which also has a \( \text{MVN}\chi^2_{\text{ScI}} \) distribution; the proof is provided in Appendix S1. Because \( f(\beta, \sigma^2) \) and \( f(\beta, \sigma^2 | y, \phi) \) have the same type of distribution, they are “conjugate” and in this context, we consider \( f(\beta, \sigma^2) \) a “conjugate prior”. The “hyper-priors” \( \xi_0, D_0, \nu_0 \) and \( \varsigma^2_0 \) define the distribution of the prior; subscript 0 indicates that a variable is a prior or hyperprior parameter.

For pragmatic reasons, \( \eta^2 \) is restricted to a fixed plug-in value. This means that we should apply external information about \( \eta^2 \) or try out different values and compare results. Furthermore, there are some computational costs involved for each possible value of \( \phi \). Therefore, the prior of \( \phi \) is restricted to a discrete set of equally distanced values, thus forcing every possible \( \phi \) in any calculation to be one of those values. For these values for \( \phi \), related internal variables are precalculated and stored.

Within the above-indicated constraints, we chose in our research a low-informative prior \( f(\beta) \) for the regression coefficients, by stating that \( D_0^{-1} = 0 \). This is a limit situation that causes \( f(\beta | \sigma^2) \) to become improper. For \( f(\sigma^2) \) and \( f(\phi) \), we followed the reasoning of Berger, de Oliveira, and Sans'o (2001) for Gaussian spatial models, which provides a reference prior composed of \( f(\sigma^2) = 1/\sigma^2 \) (which equals \( \nu_0 = 0 \), in this case a limit situation

\[
\begin{align*}
\text{Posterior} &\propto \text{likelihood; Gaussian} \\
\text{Choose flat prior} &\Rightarrow \text{neglect prior} \\
\text{Apply iteratively reweighted least squares algorithm} \\
\text{Maximum likelihood for } \beta &\Rightarrow \text{Calculate variance based on information matrix} \\
\text{Predict using covariates} &\Rightarrow \text{Variance} \\
\text{Posterior as a parameterized Gaussian distribution} &\Rightarrow \text{Formal definition posterior: } f(\beta | z)
\end{align*}
\]

FIGURE 1 Workflow to infer the posterior of \( \beta \) and predict the probability of success using a Bayesian generalized linear model (BGLM). Our implementation has been built from R base components.
for the $\chi^2_{Sc}$ distribution whereby $f(\sigma^2)$ becomes improper) and a proper prior for $\phi$, depending on the observation locations and related covariate values. An example of this reference prior $f(\phi)$ follows in the case study, and among others in Figure 4. We fixed the prior for $\tau^2$ (and thus $\eta^2$) to an arbitrary zero, because we have no other information about its possible value nor its distribution, and exploring multiple values was outside the scope of this research.

3.2.2 | Integrate out regression coefficients and variance; sample from posterior range parameter and signal

Computational costs can be substantially reduced by integrating out the regression coefficients $\beta$ and variance of the spatial signal $\sigma^2$ (see Appendix S1), rather than being part of numerically approximating the posterior. This “integrating out” means that these parameters, as such, disappear from the equation, while their influence remains, expressed in the relations between the remaining variables.

The remaining $\phi, y$ parameter and signal space is sampled using Markov chain Monte Carlo (MCMC) (Brooks, Gelman, Jones, & Meng, 2011). In each iteration the MCMC algorithm alternates between a proposal for $\phi$ and a proposal for $y$. For the one-dimensional update in $\phi$, a Metropolis algorithm is applied, meaning that the proposal distribution is symmetric. For the $n$-dimensional block update in signal space, the Langevin-Hastings algorithm (LH) is applied (Christensen et al., 2006). LH is a special case of Metropolis-Hastings, with faster convergence: whereas with Metropolis the number of iterations to reach convergence is proportional to the dimension of the combined parameter-signal space (i.e., it is $O(n)$; see Chivers & Sleightholme, 2015, for an explanation of the used “big O” notation), LH converges proportionally to the cubic root of the dimensionality ($O(n^{1/3})$) (Roberts & Rosenthal, 1998). It does so by using information captured in the spatial correlation matrix $C$ and in the observations $z$, to form a proposal probability gradient field (Christensen et al., 2006).

3.2.3 | Tuning the proposal distributions, chain phases

The proposal distributions for the MCMC iteration steps of $\phi$ and $y$ are, respectively: (a) a normal distribution, scaled and rounded to the earlier defined discrete values for $\phi$, and (b) a multivariate normal distribution scaled.
according to the LH algorithm (Christensen et al., 2006). Each proposal distribution is scaled by a single variance parameter, that is, $\gamma_\phi$ and $\gamma_Y$, respectively. These scale settings should force the acceptance rates to be as close as possible to the ideal rates: 0.44 for a single-parameter Metropolis algorithm (Rosenthal, 2011) and 0.57 for LH (Christensen et al., 2006).

Manually tuning $\gamma_\phi$ and $\gamma_Y$ proved to be a very cumbersome process, because (a) a test run to assess a tuning setting can take up to half an hour of computation time and we need to assess several settings of two scaling parameters combined, and (b) sometimes at the very beginning of a chain a low value for $\gamma_Y$ is needed to achieve an acceptance rate for $Y$ larger than zero (in other words, to start exploring the parameter-signal space), whereas soon afterwards this value has to be increased. To overcome both issues we used a self-tuning algorithm that searches values for $\gamma_\phi$ and $\gamma_Y$, forcing acceptance rates near the optimal values as given above. Because the chain continues while trying out different settings during the first several thousand iterations, this approach supports starting up.

For the self-tuning algorithm we applied a proportional-integral (PI) feedback controller for each scale setting. Such controllers are used in many industrial processes, and are known for their robustness with respect to process modelling uncertainties (Aström & Murray, 2008) and easy mathematical description. Thus, in our research the complete MCMC chain consisted of three phases: (a) the tuning phase, where $\gamma_\phi$ and $\gamma_Y$ are tuned and exploration of the parameter-signal space starts, (b) the warm-up phase, where the chain arrives in an appropriate subspace of the parameter-signal space (note that, according to Gelman et al. (2013), the expression “warm-up” is a better analogy than the often used expression “burn-in” for this phase), and (c) the production phase, where the chain explores the parameter and signal space and thus samples the posterior. During phases 2 and 3, $\gamma_\phi$ and $\gamma_Y$ are constant. The chain part from phase 3 is thinned, and ultimately used for inference and prediction.

### 3.2.4 Inferring posterior and posterior predictive

As illustrated in Figure 2, each iteration of the thinned chain enables the sampling of one value of the marginal posterior $f(\beta | z)$; all those values together provide the sample describing this marginal posterior as an empirical distribution. Exactly the same holds for $f(\sigma^2 | z)$. The process to arrive at the posterior predictive is similar: for each thinned chain iteration, a corresponding prediction distribution is calculated, and from this distribution one two-dimensional set of values is simulated. Again, all values together constitute the spatial posterior predictive as empirical distribution $f(y^* | z)$, which can be inverse logit transformed to $f(z^* | z)$. In Appendix S1, all this is explained in more detail; for example, how the chosen priors and chain iteration values influence those three sampling functions. The marginal posterior $f(\phi | z)$ as empirical distribution is already captured in the thinned chain values.

### 4 CASE STUDY: DEPTH OF PLEISTOCENE SAND LAYER IN FLEVOLAND

The Dutch province of Flevoland consists of two land reclamations (or polders): the Noordoostpolder (which translates literally to “North-East polder”) reclaimed in 1943, and the Flevopolder, reclaimed in two parts, 1957 and 1968, respectively. The total land area is ca. 1,400 km$^2$. The Noordoostpolder encapsulates two former islands: Urk (founded on boulder clay deposited in the Saalien ice age) and Schokland (founded on Holocene peat).

During the last ice age (Weichselien), cover sand, an aolian deposit with a median grain size of 105–210 μm (Koster, 2009), was deposited in this area. During the Holocene, the cover sand was covered by marine deposits (silt and clay) and with peat. The depth below the land surface of this Pleistocene cover sand layer now ranges from 0 to over 16 m.

The provincial government and the regional water board are interested in the soil composition until the Pleistocene sand layer (Brouwer, de Vries, & Walvoort, 2018). This is because unripened clay and peat might be present above the Pleistocene substrate, which has properties that change under the influence of air. These changing properties cause a subsiding ground level (on some spots with over 2 cm year$^{-1}$), while the Pleistocene sand layer is stable. For example, in the municipality of Zeewolde, subsidence as a result of ripening soils causes problems with road maintenance. Also, spatially variable subsidence causes changes in surface water flows, which both the water board and farmers have to anticipate. A map of the depth of the Pleistocene sand layer might help to locate potential problems with land subsidence. The map might also be helpful in the foundation of construction work (such as buildings, roads and railroads).

### 4.1 Soil data and covariates

In 2018, the soil in the province of Flevoland was sampled at 1,507 points (Brouwer et al., 2018) selected by
spatial coverage sampling (Walvoort, Brus, & de Gruijter, 2010) over the area of interest. At each location, the depth of the Pleistocene cover sand layer below the land surface was determined. At some locations the cover sand was not encountered within the augering depth, which was at least 1.20 m. At these locations we have so-called right-censored observations of the cover sand depth. We transformed the continuous depth data into indicators, having value 1 if the depth of the cover sand exceeds 1.20 m, and 0 elsewise.

We used two covariates in raster format, the first being elevation as derived from the digital elevation model (DEM) of the Netherlands, more specifically the LiDAR based Dutch AHN2 (PDOK, 2020). We used the AHN2 version with 5-m resolution and void filling, and resampled (using the mean) it to a 25-m square grid. We chose this covariate because we expected it to be related to the cover sand depth. For the calibration locations, the DEM ranged from −5.4 to 8.6 m, with a mean of −3.6 m, relative to the reference sea level.
The second covariate is the estimated thickness of the Holocene layer according to the three-dimensional Dutch Digital Geological Model (Dinoloket, 2020; Gunnink, Maljers, van Gessel, Menkovíc, & Hummelman, 2013). This covariate is a legacy map of the depth to the Pleistocene cover sand. Depending on the quality of the legacy map, we expect this covariate to be a good candidate for modelling and mapping our Bernoulli variable. For the calibration locations, the thickness of the Holocene layer ranges from 0 to 8.6 m, with mean 2.9 m.

After cleaning the data (the original 1,507 observations and the covariates) and defining 500 observation locations as “validation locations”, exactly 1,000 calibration locations remain, of which 782 have value “1”, that is, depth to Pleistocene sand > 1.20 m. The removed observations contain two observations at exactly the same location as other observations and five locations without full covariate coverage. The calibration and statistical validation locations, and the covariates are shown in Figure 3. Note that urban areas, swamps, waterbodies and areas where the Pleistocene sand layer is known to be at the surface were not or less intensively sampled because they are outside the area of interest (Brouwer et al., 2018).

### 4.2 Research approach

We compared four regression models: no covariates (indicated by “empty regression model”), elevation as covariate, thickness of the Holocene as covariate, and both elevation and thickness of the Holocene as covariates (“full regression model”). All four regression models have an intercept. For each regression model, we used a BGLM and a BGLGM approach. To facilitate convergence assessment (further explained in Subsection 4.2.1) each BGLM was fitted twice. Thus, we calibrated 12 models (four BGLMs and eight BGLGMs), and assessed their performance by statistical validation for comparison; the research goal here was to compare the added value of a BGLM versus BGLM in the context of different regression models. We selected the models without covariates and the models with both covariates to show the posterior distributions of model parameters and in prediction over the whole area of interest.

In the case of a BGLM, we aggregated the data by constructing pairs of points based on minimal separation distance, to reduce computing time. The sum of the two Bernoulli variables was treated as a binomial variable of two trials, located at a virtual observation point halfway between the two points of a pair. The binomial distribution is a straightforward generalization of the Bernoulli distribution shown in the theory sections of this work, allowing for several “draws” or observations per location instead of one; the number of observations at a given point is denoted by \( m \) or \( m_i \), not to be confused with the unit of distance “m”. In our case, \( m = 2 \) for all virtual points. The covariate values at the individual points were averaged and assigned to the virtual observation point.

We ran the warm-up phase and production phase with 5 million iterations each, and thinned the production phase to 50 iterations for posterior inference and statistical validation. This strong thinning was needed for computational feasibility.

#### 4.2.1 Chain convergence assessment for a BGLGM

As we are not aware of any easy to use MCMC convergence diagnostics for high-dimensional sample spaces, we applied the between-chain approach (Gelman & Shirley, 2011): for each BGLM under investigation, we started two chains, with different random starting values for \( \phi \) and \( Y \), and different pseudo random number seeds, forcing the proposals of each iteration to be different between the two chains. We assessed the convergence by visually comparing the posterior densities of \( \phi \), \( \beta \) and \( \sigma^2 \) of the two chains, as well as the statistical validation metrics (i.e., measures of the quality of the spatial predictions). Because the model parameters and the statistical validation metrics depend on the signal, we assume that this approach, although indirectly, provides sufficient information to assess the convergence of inference of both \( \phi \) and the signal. For illustration, we also present some MCMC traceplots and parameter correlations in Appendix S1.

#### 4.2.2 Statistical validation

As already indicated, we randomly divided the original observations into a calibration set (originally \( n = 1,000 \), for a BGLM aggregated to 500) and a validation set (\( n_v = 500 \)). As a statistical validation metric for continuous predictions \( p_v \) – probabilities between (and including) zero and one – in combination with binary validation observations \( z_v \) – zero or one – we used the Brier score (Spiegelhalter, 2019):

\[
\text{Brier score} = \frac{1}{n_v} \sum_{i=1}^{n_v} (p_{vi} - z_{vi})^2,
\]

which is analogue to the general mean squared error statistic.
Based on an arbitrary probability threshold of 0.5, we also computed confusion matrices for each calibrated model, each matrix showing the numbers of true positives, false positives, true negatives and false negatives. We summarized these confusion matrices by the overall accuracy:

$$\text{overall accuracy} = \frac{\text{true positives} + \text{true negatives}}{n_v}$$

(25)

and by the balanced accuracy:

$$\text{balanced accuracy} = \frac{1}{2} \left( \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} + \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} \right),$$

(26)

which is the average of the fraction of correctly predicted positives and the fraction of correctly predicted negatives (Brodersen, Ong, Stephan, & Buhmann, 2010).

To investigate the improvement by the recent soil survey and our models as compared to the legacy map (used as a covariate in some of our models), we also calculated the validation metrics for this legacy map. We transformed the continuous variable thickness of the Holocene (which is equal to the depth to the Pleistocene cover sand) according to the legacy map into an indicator applying the depth threshold of 1.20 m.

### 4.2.3 | Prediction

For prediction, we chose to use all available information (thus, use the full regression model). We resampled the covariate rasters to a prediction mask resolution of 250 × 250 m. In the case of a BGLM, we used the predict function based on a glm object. For a BGLGM, we found that the calculation time of the prediction (additional to calculating the MCMC itself) was substantial. This prediction calculation time appeared to be roughly proportional to $O(n_p^2 n_t)$, with $n_p$ the number of prediction locations and $n_t$ the number of remaining MCMC samples after thinning. We separated the prediction mask into a batch of 10 small prediction masks with ca. 2,500 pixels each, and predicted each small mask with a separate run, additional to the single run mentioned in Section 4.2 to infer the posterior parameters. All runs were provided with the same random generator seed and calibration data. Within each run we thinned the MCMC production phase to five samples; again, this quite rigid approach in separation and thinning was needed for computational feasibility. For each of those five samples, we calculated a spatial prediction as depicted in Figure 2. For each prediction location, we selected the median of the five predictions and finally merged the small prediction masks into the complete prediction map.

### 4.3 | Results

#### 4.3.1 | General performance and computational costs

Comparing the calculation approaches, we found that a BGLM always and swiftly produces an answer. However, we had initially some problems getting a BGLGM to work at all. A first hurdle was that we obtained MCMC chains with acceptance rates very close to zero or one, meaning that the chain did not explore or very slowly explored the parameter and signal space. This was due to difficulties with finding workable MCMC proposal settings $\gamma_0$ and $\gamma_1$, and probably also the need for different proposal settings during the initial iterations, compared to the subsequent iterations. Those problems were solved for all but a few of our many trials (not shown, and tested with both real-world and simulated data) using the tuning algorithm as described in Section 3.2.3. Given a proper working MCMC algorithm with reasonable acceptance rates, the second hurdle was to produce meaningful results, such as convergence to reasonable posteriors. We found that enough data were needed to achieve convergence, and that the (detrended) data needed to be sufficiently spatially structured; we also found that convergence problems due to a weak spatial structure could be solved by increasing the MCMC chain length.

The BGLMs produced answers within a few seconds calculation time, whereas each single BGLGM run had a calculation time of 30–50 h, using contemporary computer technology optimized for speed, one processing core and sufficient working memory.
For elaborating on the results and a more detailed understanding of the BGLM algorithm, we show and discuss in Section A7 in Appendix S1 MCMC traceplots and the MCMC parameter correlations.

4.3.2 | Posterior parameters

Figures 4 and 5 show posteriors of regression and spatial parameters are shown for the empty regression model and the full regression model, respectively, both using the BGLM and the BGLGM models. The regression coefficient distributions for elevation are around zero for both the models. In the case of the BGLGM, the distributions of $\beta$ are wider compared to those obtained with the BGLM. In the empty BGLM regression model the posterior of $\beta_0$ is even almost bi-modal, with one peak at $\beta_0 = 0$ and one at $\beta_0 = 3$. In the case of the full regression BGLGM models, the posteriors of $\beta_1$ and $\beta_2$ have more probability mass away from zero than that obtained with the BGLM. The posterior of spatial parameter $\sigma^2$ has quite some probability mass at higher values in the case of the empty regression model. With the full regression model, this probability mass moves closer to zero. The median of the posterior of $\sigma^2$ for the two BGLMs with the empty regression model equals 47 and 57, whereas for the two BGLMs with the full regression model these are 6.8 and 5.4. Posterior $\phi$ shows the same movement: the medians are 29 and 30 km for the empty regression models and 8.3 and 7.0 km for the full regression models.

**FIGURE 4** Posterior parameter distributions of a Bayesian generalized linear model (BGLM) and a Bayesian generalized linear geostatistical model (BGLGM) with empty regression model (intercept $\beta_0$ only). The spatial parameters $\sigma^2$ and $\phi$ indicate variance and range parameter, respectively. Note that the BGLM model only has a regression coefficient (no spatial parameters). Note also that prior $\sigma^2$ cannot be normalized because it is improper; in other words: its surface under the curve cannot be scaled to one. For each BGLGM, the posteriors of two separate runs (different in random seed) are given, allowing exploration of their convergence.
### 4.3.3 | Statistical validation

The validation metrics are given in Table 1 and the confusion matrices in Figure 6. Note that by its definition, the balanced accuracy in the case of a no covariates model, which equals “just taking the mean”, is exactly 0.5. On the contrary, because almost 80% of the calibration data have value 1 (see Section 4.1), a BGLM model based on taking the mean of the observations results in a spatially constant prediction larger than 0.5, and thus an overall accuracy of around 0.8 and also around 20% false positives; this is visualized with the confusion matrices.

All BGLMs outperform all BGLMs according to all metrics. A BGLM extracts a bit more information out of the elevation covariate; we can see in all metrics that adding elevation to the BGLM empty regression model does not improve any performance, whereas for the BGLGM it does. Adding elevation to the regression
model, which already contains the thickness of the Holocene covariate, does not change any BGLM metric, and seems to slightly worsen any BGLGM metric.

There are minor differences in the validation statistics between the two different runs of the BGLGM models, with the largest difference being 0.003 (relative: ca. 5%) in the Brier score, 0.008 (< 1%) in overall accuracy and 0.004 (< 1%) in balanced accuracy.

The legacy map of the property of interest (thickness of the Holocene) performs better than the BGLMs without the thickness of the Holocene included as covariate, but is worse in all other cases.

Although all numbers presented in Table 1 point towards the same main conclusions as presented above, the three different metrics emphasize other differences between the models and model runs.
4.3.4 | Predictions

The predictions for the full models are presented in Figure 7. Figure 8 shows the same predictions for the detail area indicated by the orange rectangle in Figure 7. For the BGLM, we used run #1. In both maps, we can see the effect of the covariates elevation and thickness of the Holocene layer (compare with Figure 3), for example in the clearly distinctive spots with low probability in the very south-west of the Flevopolder. This may be an artefact in
the urban area of the city of Almere. The prediction by the BGLM is in general somewhat smoother, especially in the very south of the Flevopolder. The BGLGM allows predicted probabilities much closer to zero than does the BGLM, whereas in large parts of the Noordoostpolder the BGLGM predicts probabilities closer to one compared to the BGLM. In the north of the Flevopolder, the BGLGM map shows an almost circular pattern with low probabilities that is hardly visible in the BGLM. Relating the prediction to the observations (Figure 8), the BGLGM follows the observations in the calibration data more closely.

5 | DISCUSSION

5.1 | Mapping Pleistocene sand layer depth

Both prediction maps (Figure 7) confirm the general spatial pattern depicted on the covariate map thickness of the Holocene (Figure 3), showing several areas where the Pleistocene sand is close to the surface. The areas where the Pleistocene sand is deeper reflect old fluvial patterns, for example, the deltas formed by the ancestors of the current rivers Overijsselse Vecht and Gelderse IJssel (Brouwer et al., 2018) and probably also the Eem. The models predict also in the urban areas, the swampy areas, and areas otherwise excluded from the sampling design, lacking both calibration and validation data. One of these excluded parts is in the area just below the very north of the Noordoostpolder. Here, the predicted probability of Pleistocene cover sand deeper than 1.20 m correctly approaches zero. The maps indicate (although not conclusively) which areas are most susceptible to soil subsidence (Brouwer et al., 2018), so where a closer look into the soil properties would make sense. The same methodology could also be used to make a map of the probability of Pleistocene cover sand at the surface. The statistical validation data of these maps suggest that we should have more trust in Figure 7b than in Figure 7a. Also visually, it seems that the BGLGM prediction in Figure 7b follows the observations as provided in Figure 3 more closely. We should be careful with the low probability dots in the south-west of the Flevopolder as shown on both maps, because they are caused by the locally deviating covariate values (very probably artefacts) and not from the observation data.

5.2 | Posterior parameters

5.2.1 | Posterior regression parameters

The posteriors of the regression coefficients were calculated differently in the BGLM and BGLGM. For the BGLM, we used straightforward mathematics and a numerical algorithm converging within a limited number of iterations. This was possible because of our choice for a low-informative prior, which meant that we could borrow well-developed concepts based on non-Bayesian statistics (i.e., estimated maximum likelihood and the corresponding estimated variances). In the case of the BGLM, the posterior is a sample generated from an MCMC chain. We suppose that extending a BGLM model with variogram parameters and a signal so that a BGLGM model is obtained, allows for wider posteriors of the regression parameters and even for a non-Gaussian posterior; see \( \beta_0 \) in Figure 4. The empty regression model resulted in a very narrow BGLM posterior centred at the logit transformed mean of the observations, whereas the corresponding BGLGM posteriors explored a much wider range of values. We could not explain why the two posterior BGLGM modes consistently were around, but different from, the mode of the posterior BGLM. In the full model (Figure 5), the modes of the intercepts \( \beta_0 \) of the BGLM and BGLGM are about equal; more freedom for \( \beta \) results in larger modes for \( \beta_1 \) and \( \beta_2 \), suggesting that the BGLGM gets more information out of the covariates than the BGLM. This was already suggested by the statistical validation. Perhaps this is due to the covariance structure or to the hierarchical nature in the BGLGM.

5.2.2 | Posterior spatial parameters

The influence of the regression model on the spatial parameters, as shown in Section 4.3.2, indicates that the covariates explain part of the variation (i.e., the probability mass emphasizes lower values of \( \sigma^2 \) if covariates are included) and explain also part of the spatial structure at larger distances (i.e., the probability mass emphasizes lower values of \( \phi \) after including covariates). It would be interesting – in follow-up research – to explore if the resulting variogram slopes close to the origin are different when comparing the empty regression model with the full regression model. It would also be interesting to investigate the correlation between the spatial parameters \( \sigma^2 \) and \( \phi \) for a given regression model by exploring the joint posterior distribution \( f(\sigma^2,\phi | \mathbf{z}) \), which is, especially in a Bayesian setting like this, fairly easily to do.

5.3 | Statistical validation

The threshold of 0.5, used for computing the confusion matrices and the overall accuracies, is a fairly arbitrary choice. Additionally, the overall accuracy metric assumes that the penalties of a false-negative and a false-positive
5.4 Comparison of the models and alternative approaches

In our research the BGLM performed reasonably well (see Table 1). However, for this it leaned heavily on the thickness of the Holocene covariate, which is based on the 3D geological model as described in Gunnink et al. (2013). The covariate elevation, often included as covariate in digital soil mapping, hardly added predictive power to any of the models. We should also note that the DEM is already used in the 3D geological model.

Based on the validation statistics (Table 1), the BGLM clearly outperformed the BGLM. However, as indicated in Section 4.3.1 (and elaborated in Section A8 of Appendix S1), it was quite an effort to get the BGLM functioning properly, and when it worked there were serious computational costs involved, with an associated electrical energy footprint (Taffoni et al., 2019) and a waiting time of up to 50 h until results were available. For the BGLM, we followed the approach and implementation of Diggle and Ribeiro (2007) and Christensen (2002). Comparable approaches in the R-universe, such as \texttt{geoCount} (Jing & De Oliveira, 2015), \texttt{spBayes} (especially function \texttt{spMvGLM} [Finley, Banerjee, & Gelfand, 2015]), \texttt{geoBayes} (Evangelou & Roy, 2019), \texttt{PrevMap} (Giorgi & Diggle, 2017) and approximations such as \texttt{INLA} (Rue, Martino, & Chopin, 2009), eventually embedded in \texttt{geostatsp} (Brown, 2015), were outside our scope, but might be interesting to consider in future research and practical applications.

A decision on which method to use should depend on the available or feasible sample size and the available covariates, the final goal of the map and/or other deliverables such as the posteriors, uncertainty quantification and – more pragmatically – the available resources. Often, in digital soil mapping many covariates are used, some based on direct observation (such as DEM and its derivatives as generated by radar and vegetation data visible on multi-spectral satellite images) and some based rather on modelling and interpolation (such as climate data). In case few covariates related to the variable of interest are available, we need to have methods at hand that deliver reasonable mapping results based only on the sample itself. In such cases, indicator kriging, although statistically not completely sound, might be a good approach. If we want to add already existing knowledge not captured in covariates, a Bayesian approach might be a good option. If we, from a pedometrics’ viewpoint, are interested in a solid statistical description of binomial soil properties over space, an approach such as the BGLM might very well be worth the effort. Furthermore, the addition of an automated tuning phase as described in Section 3.2.3 makes both methodological research and practice-directed data processing much more feasible than manual tuning would allow.

6 CONCLUSIONS

Adding a geostatistical component to a Bayesian generalized linear model (BGLM) for mapping binary soil properties yielded considerably better statistical validation metrics in a case study with a large \((n = 1,000)\) observation sample and few relevant covariates available. However, the resulting Bayesian generalized linear geostatistical model (BGLM) is quite demanding with respect to sample size, tuning the algorithm and computational costs. In this study we focused on the implementation of the BGLM as provided in the R-package...
geoRglm, which we embedded in our own scripts. We replaced manual tuning by an automated tuning algorithm and found a sample composition (in size and in number of observations per location) that delivers meaningful results within 50 h calculation time. With the gained insights, spatial soil practitioners and researchers can – for their specific cases – better evaluate if using a BGLGM would be possible at all and if the extra gain would be worth the extra effort. The developed automated tuning algorithm (of which the code is available as part of an easy to run example script, see doi.org/10.5281/zenodo.5005828) makes implementation of a BGLGM in applications easier.

ACKNOWLEDGEMENTS
The data used in this research were provided by Wageningen Environmental Research.

CONFLICT OF INTEREST
There is no conflict of interest to declare.

AUTHOR CONTRIBUTIONS
The three authors together developed the research approach. Luc Steinbuch carried out the actual research, coded the algorithms around existing R-packages and wrote the manuscript and the additional material. Dick Brus and Gerard Heuvelink supported the development of the research (especially the statistical foundation as formulated in Sections 2 and 3, and in Appendix S1), supported interpretation of the results and provided feedback on the manuscript, both on the content and style.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Luc Steinbuch https://orcid.org/0000-0001-6484-0920
Dick J. Brus https://orcid.org/0000-0003-2194-4783
Gerard B. M. Heuvelink https://orcid.org/0000-0003-0959-9358

REFERENCES
Assel, M., Sjoberg, D. D., & Vickers, A. J. (2017). The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. *Diagnostic and Prognostic Research*, 1(1), 19. https://doi.org/10.1186/s41512-017-0020-3
Aström, K. J., & Murray, R. M. (2008). *Feedback systems: An introduction for scientists and engineers*. Princeton, New Jersey: Princeton University Press.
Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. Boca Raton: CRC Press. https://doi.org/10.1201/9780203487808
Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2015). *Hierarchical modeling and analysis for spatial data*. Boca Raton: CRC Press. https://doi.org/10.1201/b17115
Beaudette, D. (2020). Accuracy and uncertainty for categorical predictions. Retrieved from https://web.archive.org/web/20200906145909/http://ncss-tech.github.io/stats_for_soil_survey/chapters/9_uncertainty/class-accuracy-uncertainty.html
Accessed 6/15/2021
Berger, J. O., de Oliveira, V., & Sans'o, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456), 1361–1374. http://www.jstor.org/stable/3085905
Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In 2010 20th International Conference on pattern recognition (pp. 3121–3124). Los Alamitos, California: IEEE. https://doi.org/10.1109/ICPR.2010.764
Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo. Handbooks of modern statistical methods*. Boca Raton, FL, London, England, New York, NY: CRC Press.
Brouwer, F., de Vries, F., & Walvoort, D. (2018). *Basisregistratie Ondergrond (BRO) actualisatie bodemkaart: Herkartering van de bodem in Flevoland*. [Key Register of the Subsurface (BRO): Update of soil map: Soil remapping in Flevoland]. Technical Report WOt-Technical Report 143. Wageningen, The Netherlands: WUR. https://doi.org/10.18174/468672
Brown, P. E. (2015). Model-based geostatistics the easy way. *Journal of Statistical Software*, 63(12), 1–24. https://doi.org/10.18637/jss.v063.i12
Chivers, I., & Sleightholme, J. (2015). An Introduction to Algorithms and the Big O Notation. *Introduction to Programming with Fortran* (pp. 359–364). Cham: Springer. https://doi.org/10.1007/978-3-319-17701-4_23
Christensen, O. & Ribeiro Jr, P. (2015). Package “geoRglm”: A package for generalised linear spatial models. 1-48. Retrieved from https://cran.microsoft.com/snapshot/2016-04-14/web/packages/geoRglm/geoRglm.pdf Accessed 6/15/2021
Christensen, O. F. (2002). Methodology and applications in non-linear model-based geostatistics. (PhD thesis), Aalborg Denmark: Aalborg University Denmark.
Christensen, O. F., & Ribeiro, P., Jr. (2002). *geoRglm*: A package for generalised linear spatial models. *R News*, 2(2), 26–28. https://cran.r-project.org/doc/Rnews/Rnews_2002-2.pdf
Christensen, O. F., Roberts, G. O., & Sköld, M. (2006). Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1), 1–17. https://doi.org/10.1198/106186006X100470
Diggle, P., & Ribeiro, P. J. (2007). *Model-based geostatistics*. New York, NY: Springer. https://doi.org/10.1007/978-0-387-48536-2
Dinoloket. (2020) Digital geological model. Retrieved from https://www.dinoloket.nl/ Accessed 3/20/2017
Emery, X., & Ortiz, J. M. (2004). Shortcomings of multiple indicator kriging for assessing local distributions. *Applied Earth Science*, 113(4), 249–259. https://doi.org/10.1179/174327504X27242
Evangelou, E. & Roy, V. (2019). *geoBayes: Analysis of geostatistical data using Bayes and empirical Bayes methods*. R package...
SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Steinbuch, L., Brus, D. J., & Heuvelink, G. B. M. (2022). Mapping depth to Pleistocene sand with Bayesian generalized linear geostatistical models. *European Journal of Soil Science*, 73(1), e13140. [https://doi.org/10.1111/ejss.13140](https://doi.org/10.1111/ejss.13140)