Bayesian Analysis (2010) **5**, Number 3, pp. 557–582

Predicting Vertical Connectivity Within an Aquifer System

Margaret Short*, Dave Higdon†, Laura Guadagnini‡, Alberto Guadagnini§ and Daniel M. Tartakovsky¶

Abstract. The subsurface environment beneath the Municipality of Bologna, Italy, is comprised of a series of alluvial deposits which constitute large and productive aquifer systems. These are separated from the shallow, free surface aquifer by a low permeability barrier called aquitard Alpha. The upper aquifer contains water that shows relevant contamination from industrial pollutants. The deep aquifers are relatively pristine and provide about 80% of all groundwater used for drinking and industrial purposes in the area of Bologna. Hence, it is imperative that planners understand where along aquitard Alpha there exists potential direct connection between the upper and the deep aquifers, which could lead to contamination of the city's key water supply well fields.

In order to better assess the existence of preferential flow paths between these aquifer systems, we carry out a statistical analysis in which the aquitard is represented as a bivariate spatial process, accounting for dependence between the two spatial components. The first process models its effective thickness. The second process is binary, modeling the presence or absence of direct vertical connections between the aquifers. This map is then cross referenced with other forms of data regarding the hydrology of the region.

Keywords: Markov chain Monte Carlo, Gaussian process, subjective likelihood, spatial model

1 Introduction

The key geological units beneath the Municipality of Bologna, Italy, host large and productive aquifers, which are separated vertically from the upper, free surface aquifer by a low permeability barrier called aquitard Alpha (Figure 1). The upper aquifer contains water that shows evidence of contamination by industrial pollutants. The lower aquifers are relatively pristine and provide about 80% of all groundwater used for drinking and industrial purposes in the area of Bologna. Therefore, the integrity of

*Department of Statistics, University of Alaska Fairbanks, Fairbanks, AK, http://www.dms.uaf.edu/~mshort/

†Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM, mailto:dhigdon@lanl.gov

‡Dipartimento di Ingegneria Idraulica, mbientale, Infrastrutture Viarie e Rilevamento, Politecnico di Milano, Milano, Italy, mailto:laura.guadagnini@polimi.it

§Dipartimento di Ingegneria Idraulica, mbientale, Infrastrutture Viarie e Rilevamento, Politecnico di Milano, Milano, Italy, mailto:alberto.guadagnini@polimi.it

¶Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA, mailto:dmt@ucsd.edu

 DOI:10.1214/10-BA522

aquitard Alpha is crucial for the health of the lower aquifers.

Relatively fast migration of contaminants between the upper and deep aquifers is possible through highly permeable inclusions (e.g. fractures, cracks, lenses of coarse material, etc.) embedded in aquitard Alpha. The ability to ascertain the locations and spatial extent of such inclusions which constitute preferential flow paths for contaminants is paramount for ensuring the safe and environmentally sound exploitation of groundwater resources in the region.

The goal of the analysis presented here is to assess the likelihood that Alpha contains such preferential pathways for contaminants. Our analysis makes use of sedimentological data, spatial modeling and expert judgment. It leads to the construction of a spatial map depicting the probability of occurrence of vertical connections between the upper ad lower aquifers. These connections are typically associated with discontinuities within aquitard Alpha. The latter are in turn related to the occurrence of large fractions of coarse material mixed with fine particle sizes within Alpha or small (local) thicknesses of the aquitard. As an ancillary result, we provide an estimate of the spatial distribution of the fine materials (clay and silt) within Alpha, which act as the primary barrier to the vertical flow of contaminants. While the construction of this map is not our primary goal, we show that it is a necessary step for the assessment of the occurrence of local vertical connections between aquifers.

The spatial modeling uses two-dimensional representations. We are not attempting to pinpoint the precise location (trajectory) of preferential pathways through the aquitard, which would be the case for a three-dimensional reconstruction, but rather to infer the existence of such pathways and their (two-dimensional) spatial location. The available data provide *indirect* information about connectivity between the aquifers. Our modeling approach handles this indirectness through a subjective likelihood whose form is guided by expert opinion.

The remainder of this section discusses in greater detail the study region and the motivation for this study. Section 2 describes the data and introduces terminology that will be used through the remainder of this paper. Section 3 spells out the primary statistical model that was developed for this data. Section 4 discusses results and cross-references with other types of hydrogeologic information. An alternative modeling approach is outlined in Section 5. We end with some concluding remarks.

1.1 Description of the site and motivation of the study

The city of Bologna lies on the alluvial plain of the Reno river, in the Emilia Romagna region of Northern Italy. This study focuses on a nearby area (Figure 1a) of about $50km^2$, which contains three major well fields whose combined yield accounts for about 80% of the municipality's groundwater supply. The most important environmental problem suffered by the municipality of Bologna is groundwater contamination. In recent decades, industrial development has produced many sources of pollutants; these include organohalogenatic compounds such as perchloroethylene (used as an industrial solvent, for example, in dry-cleaning) and nitrates (mainly associated with agricultural

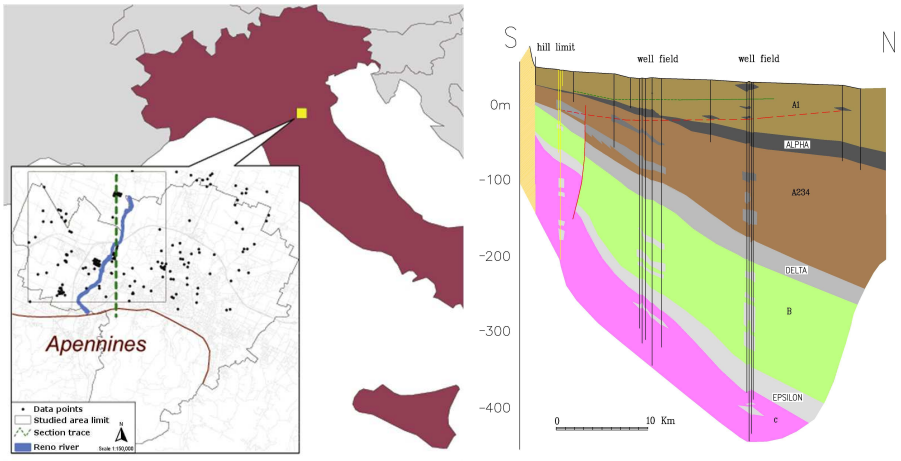

Figure 1: Locations of the municipality of Bologna, the study area (the square region in the insert) and data points within the study area and its vicinity (a). A conceptual representation of the main geological units in the vertical cross-section of the study area (b). The wells shown in (b) are used by the municipality to pump drinking water; to avoid a preferential flow path through the aquitards, the well bores are sealed. Vertical units are meters. Our focus is on the analysis of the composition of an aquitard, Alpha, that separates two aquifers in the Bologna aquifer system.

activities).

The geologic structure of the region's subsurface has been the subject of numerous investigations (e.g., (Ricci Lucchi 1984; Ricci Lucchi et al. 1982; Francavilla et al. 1980; Amorosi and Farina 1994b,a, 1995)). These and other studies reveal that the Reno river alluvial fan in the area has a wedge shape, increasing in thickness from south to north (Figure 1b). The alluvial fan rests on sea clayey deposits, which are rich with saline water. The water-bearing alluvial deposits in the Bologna area are more than $300m$ thick (Francavilla et al. (1980)) and can be subdivided into three large-scale geological units (also known as depositional cycles) denoted by A, B, and C in Figure 1b. These depositional cycles, each about $100 - 150m$ thick, are separated by the clayey deposits denoted Delta and Epsilon, which form flow barriers and are commonly referred to as aquitards. Depositional cycle A can be further subdivided into two major subunits (denoted by A1 and A234 in Figure 1b), which are separated by the aquitard called Alpha.

The alluvial deposits A234 and B form large and productive aquifer systems, which are heavily used by the municipality of Bologna as a major source of fresh water. Groundwater from the upper aquifer A1 shows signs of local contamination. Aquitard Alpha plays a major role as the natural barrier between the upper contaminated aquifer system A1 and the deeper aquifer systems A234 and B; its reconstruction is the primary objective of this study.

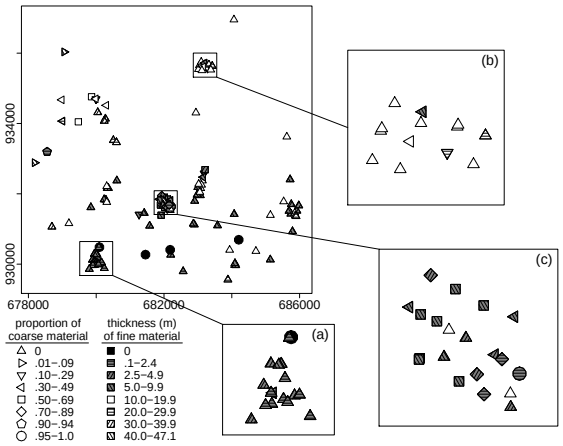

Figure 2: Spatial distribution of well core samples corresponding to aquitard Alpha. Shape indicates proportion of coarse material (sand and gravel) in the well core. Color and hatching indicate the thickness of fine material (silt and clay). A dark circle thus indicates the greatest potential for contamination to pass from upper aquifer to lower: little or no fine material, and a high proportion of coarse material. Of the 123 core samples, 85 contained no coarse material.

2 Sedimentological Data

A total of 39 logs of geognostical boreholes and 183 well logs were used to characterize the subsurface structure of the Reno river alluvial fan within the municipality of Bologna. Their locations are shown by the dots in Figure 1a. Of these, 123 well logs fall within the area under investigation — the region identified by the dashed-line square in the insert of Figure 1a — and are used in this analysis. From each of these well logs, two attributes of aquitard Alpha were extracted: its local thickness and the volumetric fraction (percentage) of the embedded coarse materials. These are reported as the cumulative thickness of fine materials (silt and clay) and the fraction (0.0–1.0) occupied by coarse materials (sand and gravel), respectively. Figure 2 depicts spatial locations of the well core data over the study region.

The data show that aquitard Alpha is composed mainly of fine (silty-clayey) material with local interbedding of coarse (sand and gravel) material. The dominance of the fine material defines the ability of aquitard Alpha to act as a natural barrier, which prevents flow and migration of contaminants between the upper aquifer and the deep groundwater reservoirs. The inclusions of coarse material are highly permeable and can effectively act as locations of preferential flow paths in the otherwise impermeable aquitard.

Aquitard Alpha is most likely to pose environmental problems (i.e., to display contaminant pathways) at locations where it is thin and contains relatively large fractions of the coarse geomaterial. From each of the $n = 123$ well core samples the bivariate

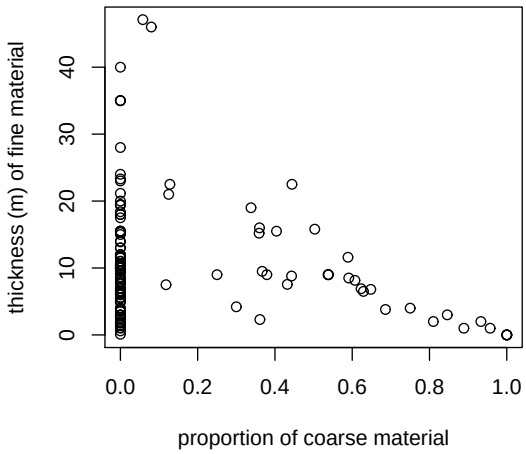

Figure 3: Proportion of coarse material (sand and gravel) vs. the thickness of impermeable material (silt and clay) in the well core samples of aquitard Alpha.

measurement

$$y(s_i) = \begin{pmatrix} y_T(s_i) \\ y_F(s_i) \end{pmatrix} = \begin{pmatrix} \text{thickness of fine material at } s_i \\ \text{fraction of coarse material at } s_i \end{pmatrix}, \quad i = 1, \ldots, n \tag{1}$$

was recorded. The data are shown spatially in Figure 2 and as a scatterplot in Figure 3. The shape and orientation of the plotting symbols indicate the fraction of coarse material; the color (darkness) and hatching indicate the thickness of the fine material. The measurements showing aquitard Alpha to be locally thin as well as having a high proportion of coarse material suggest possible discontinuities within the impervious matrix of the aquitard. Sites of Figure 2 where this occurs are depicted using plotting symbols that are both round – indicating little or no fine material – and dark – indicating a high proportion of coarse material. Likewise, these same environmentally critical locations correspond to points in the lower right corner of Figure 3.

3 Statistical modeling

The statistical modeling for this application requires us to specify a spatial prior model for aquitard Alpha as well as a likelihood which determines how the observed data informs about the aquitard. In this section we describe our final model, which explicitly models the thickness of aquitard Alpha. This model uses a subjective likelihood, constructed with the aid of hydrogeologists involved with this analysis. The resulting inference relies heavily on this construction.

3.1 Spatial prior model for aquitard Alpha

Over the study region $\mathcal{S}$ we focus on two features of aquitard Alpha: the total thickness of the fine materials, and the presence or absence of contaminant pathways through the aquitard. We specify a prior bivariate spatial process $(z_T(s), z_P(s))$, $s \in \mathcal{S}$, for aquitard Alpha. The first component, $z_T(s)$, which we call clay thickness, denotes the total thickness of impermeable material (clay and silt) within Alpha. It is a positive, continuous field over the spatial region $\mathcal{S}$. The binary process $z_P(s)$ we call permeability. It denotes the presence ($z_P(s) = 1$) or absence ($z_P(s) = 0$) of contaminant pathways through aquitard Alpha at spatial location s. Both of these fields are non-Gaussian, but are constructed using latent Gaussian process models.

The clay thickness is derived by taking the positive part of a standard Gaussian process $u_T(s)$ so that

$$z_T(s) = \begin{cases} u_T(s) & \text{if } u_T(s) \geq 0 \\ 0 & \text{if } u_T(s) < 0. \end{cases}$$

The underlying Gaussian process $u_T(s)$ has a mean function $\mu_0 + \mu_1 s_2$ that depends on the north-south spatial coordinate s_2. The random part of u_T is constructed using a discrete representation given in Higdon (2002), which we define as follows. Let $\mathcal{S}$ be the spatial region defined by the study region in this application, and let $x_1^T, \ldots, x_K^T$ be iid $N(0, \lambda_T^{-1})$ random variables ("knot values") associated with sites $w_1, \ldots, w_K \in \mathcal{S}$. The knot locations w_j are taken to be an equally spaced 25×25 array over the study region $\mathcal{S}$. The spatial process $u_T(s)$ for $s \in \mathcal{S}$ is constructed using the representation

$$u_T(s) = \mu_0 + \mu_1 s_2 + \sum_{j=1}^{K} x_j^T k(s - w_j; \sigma_T),$$

where we take $k(\cdot; \sigma_T)$ to be a circular, bivariate normal density with standard deviation σ_T. The precision parameter λ_T controls the precision of the x_j^T's, which in turn controls the marginal variance of the $u_T(s)$ process. A $\Gamma(a_T = 1, b_T = .001)$ prior is specified for λ_T. The parameter μ_0 is given a $U(0, 40)$ prior; μ_1 is given a $U(-30, 30)$ prior. Finally, the kernel width parameter σ_T controls the range of spatial dependence for the $u_T(s)$ process. We specify a $U(.04, 1)$ prior distribution for σ_T, after rescaling $\mathcal{S}$ to the unit square, $[0, 1] \times [0, 1]$. The lower bound of 0.04 is equal to the minimum knot spacings in the underlying knot locations. Allowing σ_T to take values significantly lower than this will result in "dead" regions in $u_T(s)$ that take values near the mean value μ_0. If smaller values of σ_T are needed, a finer grid of knot locations may be used. For the application here, the 25×25 grid is sufficient. Note that increasing the knot density has very little effect, since λ_T can be correspondingly increased, leaving the induced distribution for the latent process $u_T(s)$ essentially unchanged.

This discrete representation allows the thickness field to be controlled by μ_0, μ_1, and the K knot values x_T. This is particularly useful for the MCMC approach used to explore the complicated posterior distribution resulting from this application (see Equation (5)).

Conditional on the clay thickness field $z_T(s)$, the binary permeability field $z_P(s)$ is constructed in a similar fashion using a latent, mean 0, Gaussian process $u_P(s)$ according to the rule

$$z_P(s) = \begin{cases} 1 & \text{if } \beta_0 + \beta_1 z_T(s) + u_P(s) \geq 0 \\ 0 & \text{if } \beta_0 + \beta_1 z_T(s) + u_P(s) < 0. \end{cases} \tag{2}$$

Conditional on $z_T(s)$, this is an example of a clipped Gaussian random field model from De Oliveira (2000), where the mean function depends on the clay thickness $z_T(s)$. It can also be considered a slight generalization of the latent multivariate probit model given in (Agresti (1990),Johnson and Albert (1999)). Unconditionally, this formulation results in a bivariate non-Gaussian spatial model. The process $u_P(s)$ is defined in a manner analogous to that of $u_T(s)$:

$$u_P(s) = \sum_{j=1}^{K} x_j^P k(s - w_j; \sigma_P),$$

where the knot locations are the same as those for $u_T(s)$. Here the underlying knot values $x_1^P, \ldots, x_K^P$ are given independent $N(0, \lambda_P^{-1})$ distributions a priori. The precision parameter λ_P is fixed so that the resulting Gaussian process $u_P(s)$ has a nearly constant marginal variance of 1. The variance is not exactly constant over $\mathcal{S}$ due to the discrete representation. However, the fluctuations over $\mathcal{S}$ are small enough to be negligible. Also, the kernel $k(\cdot; \sigma_P)$ is controlled by its own scaling parameter σ_P. As with its clay thickness counterpart, σ_P is assigned a $U(.04, 1)$ prior.

In equation (2), β_0 controls the mean connectivity (permeability) probability; β_1 controls the dependence between the clay thickness and connectivity fields. We expect that β_1 will be negative, so that greater clay thickness at spatial location s makes it less likely that there is a preferential pathway through aquitard Alpha (i.e. $z_P(s) = 1$). However, we specify wide $U(-15, 15)$ priors for both β_0 and β_1 to allow the well core data to inform about these parameters.

Figure 4 helps visualize how this model works, with the top row depicting aspects of the clay thickness portion of the model ($u_T(s)$ and $z_T(s)$) and the bottom row for the permeability components of the model ($u_P(s)$ and $z_P(s)$). The left column shows the grids used to create the two processes, $u_P(s)$ and $u_T(s)$. Independent normal variates are located at the grid locations that are shown in the left hand column of plots; these knot values are convolved by a normal kernel whose standard deviation is given by the circles in these plots (for $u_T(s)$ in the top row, $u_P(s)$ in the bottom row). The center column depicts realizations for $u_T(s)$ and $u_P(s)$, obtained by convolving knot values by the kernels. The final column of this figure shows the corresponding induced realization for the two dependent fields, $z_T(s)$ and $z_P(s)$. Note that the realization of $z_T(s)$ is obtained by truncating the negative values in $u_T(s)$; locations where clipping has occurred are painted red (flat gray). The realization of $z_P(s)$ is binary, constructed by thresholding a linear combination of $z_T(s)$ and $u_P(s)$. For this realization, values for the parameters $x_T = (x_1^T, \ldots, x_K^T)^T$, $x_P = (x_1^P, \ldots, x_K^P)^T$, σ_T, σ_P, μ_0, μ_1, β_0, and β_1 were taken from a posterior realization used in the eventual analysis of aquitard Alpha. This figure is discussed further in Section 4.

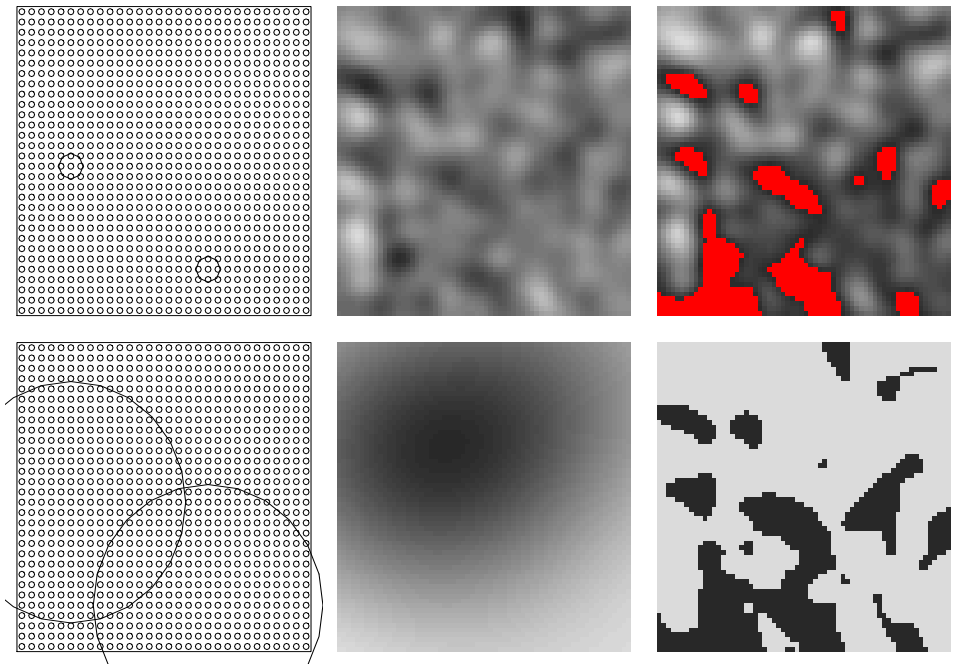

Figure 4: Construction of the bivariate spatial model. The top row refers to the clay portion of the process (u_T and z_T), the bottom row to the permeability portion of the process (u_P and z_P). Independent normal variates are located at the grid locations shown in the left hand column of plots; these knot values are convolved by a normal kernel whose standard deviation is given by the overlaid circles. The induced fields $u_T(s)$ and $u_P(s)$ are shown in the middle columns. The right hand columns show the two spatial fields: the clay thickness $z_T(s)$ and the binary connectivity field $z_P(s)$. The red (flat gray) portion of the top right image shows where $z_T(s) = 0$, i.e. where clay thickness is 0m. The dark portion of the bottom right image shows where $z_P(s)$ is 1, i.e. where aquitard Alpha is permeable.

3.2 Likelihood specification

The physical measurements $(y_T(s_i), y_F(s_i))$, $i = 1, \ldots n$, give the thickness and proportion (fraction) of coarse material of aquitard Alpha at the $n = 123$ spatial locations. However, we wish to infer about the connectivity of Alpha over the entire study area $\mathcal{S}$, which is never directly observed. The data inform about the underlying processes $z_T(s)$ and $z_P(s)$ through the likelihood, which factors into two terms: one involving the clay thickness, and one involving the connectivity field.

The first term is a standard likelihood for a spatial model where the observed clay depth is a noisy version of the true depth. The observed clay depth is not expected to be exact, since the well core data typically carries a fair bit of uncertainty. In addition,

we expect this noise term to absorb some of the small scale variability in the actual clay depth. Taking $y_T = (y_T(s_1), \ldots, y_T(s_n))^T$ to be the n-vector of clay depth observations, and $z_T = (z_T(s_1), \ldots, z_T(s_n))^T$ to be the clay depth process at the observation sites, this component of the likelihood becomes

$$L_T(y_T|z_T, \lambda_y) \propto \lambda_y^{n/2} \exp\left\{-\frac{\lambda_y}{2}(y_T - z_T)^T(y_T - z_T)\right\},$$

where the parameter λ_y controls the observation precision. Note that this is not a multivariate truncated normal distribution, but rather a multivariate normal distribution in which the means are non-negative. This is a modeling choice that allows for the possibility of contiguous regions having clay thickness $z_T(s)$ equal to 0. In addition, it greatly facilitates computations.

The second component of the likelihood determines how the observed data inform about the binary connectivity process $z_P(s)$. Here it is standard to specify a sampling model for which observations $y(s_i) = (y_T(s_i), y_F(s_i))^T$, $i = 1, \ldots, n$, are independent given the underlying spatial field, so that

$$L_P(y|z_P) = \prod_{i=1}^{n} L_P(y(s_i)|z_P(s_i)), \tag{3}$$

where $y = (y(s_1)^T, \ldots, y(s_n)^T)^T$ and $z_P = (z_P(s_1), \ldots, z_P(s_n))^T$. Similar binary classification applications from spatial and image applications typically specify a normal sampling model where the mean of $y(s)$ depends on the state of $z_P(s)$ (Besag et al. 1991; Hurn 1998). However, the nature of the bivariate measurements taken here, along with their spread – evident in Figure 3 – are incompatible with a normal sampling model for the $y(s_i)$'s.

We assume the product form of the sampling model in equation (3); this, together with the fact that each $z_P(s_i)$ is binary, means that the likelihood depends only on the ratios

$$r(y(s_i)) = \frac{L_P(y(s_i)|z_P(s_i) = 1)}{L_P(y(s_i)|z_P(s_i) = 0)}, \quad i = 1, \ldots, n,$$

since (3) can be rewritten as

$$L_P(y|z(s)) = \prod_{i=1}^{n} r(y(s_i))^{z_P(s_i)} L_P(y(s_i)|z_P(s_i) = 0) \propto \prod_{i=1}^{n} r(y(s_i))^{z_P(s_i)}. \tag{4}$$

Hence the data only inform about $z_P(s)$ through the specification of the *odds map* $r(y(s))$.

It remains to specify $r(y)$ as a function of possible bivariate outcomes $y(s) \in \{[0, \infty) \times [0, 1]\}$. In fact, we need only consider thicknesses between 0 and 50 meters since it is unlikely aquitard Alpha is thicker than 50 meters within the study region. In specifying the odds map, it is convenient to interpret $r(y(s))$ as the odds that the underlying $z_P(s) = 1$ at the spatial location s, independent of spatial information encoded into the prior for $z_P(s)$.

We rely on expert knowledge from hydrogeologists familiar with the study region to determine the odds map $r(y(s))$. We model $\log r(y(s))$ as a linear function of the observed clay thickness $y_T(s)$ and the observed proportion of coarse material in Alpha $y_F(s)$,

$$\log(r(y(s)) = \gamma_0 + \gamma_1 y_T(s) + \gamma_2 y_F(s). \tag{5}$$

The expert knowledge is used to specify a line for which the odds are equal to 1, and a slope which determines how quickly the odds change as the data move away from the $r(y(s)) = 1$ line. This, equivalently, determines γ_0,γ_1, and γ_2. Considerations for determining this map include the following points.

- For the very small local thickness of aquitard Alpha (< 10m), even a very large fraction of fine material does not guarantee the spatial continuity (impermeability) of the aquitard.
- An intermediate thickness guarantees the continuity only for a relatively small fraction of coarse material. This is because coarse materials within a clearly identifiable aquitard manifest themselves as a sequence of interbedding structures, so that a porous pathway through the aquitard is possible.
- A very large fraction of coarse materials tends to be indicative of local discontinuities, unless the thickness of an aquitard is very large (> 20m).

With these points in mind, we see that $\gamma_1 < 0$ (the odds of permeability decrease as clay thickness increases) and that $\gamma_2 > 0$ (the odds of permeability increase as the proportion of coarse material increases). These are consistent with our best determination of this function, shown in the central frame of Figure 7 and given by $(\gamma_0, \gamma_1, \gamma_2) = (2.30, -.461, 4.61)$.

We digress briefly to discuss this likelihood, as it is somewhat subjective. Our rationale in developing its general form was driven, in part, by the following. First the modeling of the clay thickness process is relatively straightforward, and the strong north-south trend in clay thickness provides a considerable amount of information about the permeability of aquitard Alpha. We therefore choose to include it in our model. Secondly, we consider it sensible and in fact desirable to use the bivariate observations (y_T, y_F) to inform about z_P, because doing so allows us to incorporate expert opinion in tying these together through the log-odds map. We believe these strengths outweigh the complexity of the model, as well as the fact that the clay thickness observations are being used twice: once in specifying $L_T(y_T|z_T)$ and again in specifying $L_P(y|z_P)$, recalling that z_P is partly determined by z_T.

To help solidify these ideas, let us examine how the spatial priors and likelihood fit together, by describing how one might simulate a data set, $y = (y(s_1), y(s_2), \ldots, y(s_n))$. Assuming known values for the model parameters (knot values, precision parameters, etc.), we construct a realization of the spatial prior processes, $z_T(s)$ and $z_P(s)$ by convolving knot values by the kernel, as illustrated in Figure 4. Observations $y(s_i)$ can then be simulated, first by simulating y_T, the clay thickness, from its marginal distribution,

then simulating from y_F's distribution given y_T. It is straightforward to verify that the marginal density for $y_T(s_i)$ is given by

$$f(y_T|z_T(s_i), z_P(s_i)) \propto \exp\left(-\frac{\lambda_y}{2}(y_T - z_T)^2\right) \times g(y_T|z_P) \times I(0 \leq y_T \leq 50),$$

where

$$g(y_T|z_P) = \begin{cases} \log\left(\frac{1+\exp(\gamma_0+\gamma_1 y_T+\gamma_2)}{1+\exp(\gamma_0+\gamma_1 y_T)}\right) & \text{if } z_P = 1 \\ \gamma_2 - \log\left(\frac{1+\exp(\gamma_0+\gamma_1 y_T+\gamma_2)}{1+\exp(\gamma_0+\gamma_1 y_T)}\right) & \text{if } z_P = 0 \end{cases}$$

The effect of multiplying $g(y_T)$ against the density $\exp(-\frac{\lambda_y}{2}(y_T - z_T)^2)$ can be seen in Figure 5, which shows the marginal densities $f(y_T) = f(y_T|z_T, z_P)$ under four scenarios ($z_P \in \{0, 1\}$ and $z_T \in \{5m, 25m\}$). In each panel, the values of γ_0, γ_1, and γ_2 are those from our final model, and λ_y is set equal to its posterior median. In this figure, the factor $g(y_T)$ has little effect when z_T and z_P are aligned similarly (z_T large and $z_P = 0$, or z_T small and $z_P = 1$). If, however, z_T is small and $z_P = 0$, i.e. little clay yet impermeable, multiplying $g(y_T)$ against $\exp(-\frac{\lambda_y}{2}(y_T - z_T)^2)$ increases the marginal mean of y_T, which is as we desire; and when z_T is large and $z_P = 1$, the multiplication by $g(y_T)$ decreases the marginal mean of y_T, also the desired effect.

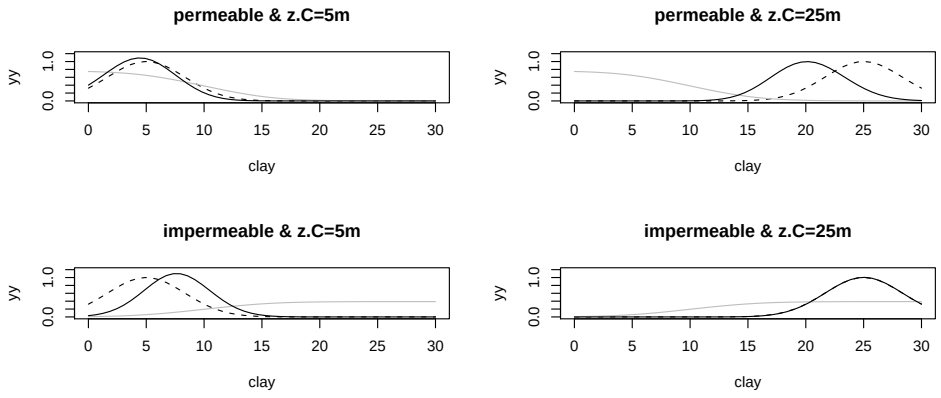

Figure 5: The marginal density $f(y_T)$ is plotted in each panel (solid line), along with the density $\exp(-\frac{\lambda_y}{2}(y_T - z_T)^2)$ (dashed line); the gray curve shows the ratio of these, denoted in the text by $g(y_T)$. The panels illustrate four scenarios, in which $z_P \in \{0, 1\}$, and $z_T \in \{5m, 25m\}$ (corresponding to estimated mean clay thickness at southern and northern ends of the study region). The values of γ_0, γ_1, and γ_2 are those from our final model, and λ_y is set equal to its posterior median. The factor $g(y_T)$ has little effect when z_T is large and $z_P = 0$ (clay thickness is large and Alpha is not permeable), and when z_T is small and $z_P = 1$ (little clay and Alpha is permeable). If, however, $z_P = 0$ and z_T is small, the marginal mean of y_T is greater than z_T, which indicates that the model prefers larger observations y_T in accordance with Alpha being impermeable.

Once simulated values for y_T have been obtained, the fraction of coarse material, $y_F(s_i)$, $i = 1, \ldots n$, can be simulated from its conditional density,

$$f_i(y_F) \propto \begin{cases} p_i(y_F) \times I(0 \le y_F \le 1) & \text{if } z_P(s_i) = 1 \\ (1 - p_i(y_F)) \times I(0 \le y_F \le 1) & \text{if } z_P(s_i) = 0 \end{cases},$$

where

$$p_i(y_F | y_T(s_i) = y_T) = \frac{\exp(\gamma_0 + \gamma_1 y_T + \gamma_2 y_F)}{1 + \exp(\gamma_0 + \gamma_1 y_T + \gamma_2 y_F)} \times I(0 \le y_F \le 1).$$

One can verify that this method of simulating data is consistent with the likelihood description provided above.

Returning to the main thread of our discussion, we note that the data provide essentially no information about the values of γ_0, γ_1 or γ_2, as we convinced ourselves by examining output from lengthy MCMC runs (plots omitted). Thus we eventually restricted our analysis to several combinations of these parameters, as shown in the various frames of Figure 7. The hydrogeologists believe these cases bound the plausible relationships between the data and the connectivity field. In particular, combinations of the γs lying far outside this range of possibilities resulted in estimated maps considered completely implausible – in their judgment, roughly 1/4 of the study region is believed consistent with vertical connections between the relevant aquifers. In Section 4, we investigate sensitivity of results to our choice of odds map.

3.3 Posterior distribution

After specifying common, independent gamma priors for the precisions λ_y, λ_T and λ_P, and common uniform priors for the kernel width parameters σ_T and σ_P, the resulting posterior distribution has the form:

$$\begin{aligned} \pi(\boldsymbol{\theta}|y) \quad \propto \quad & \lambda_y^{\frac{n}{2}} \exp\left\{-\frac{\lambda_y}{2}||y_T - z_T||^2\right\} \times \prod_{i=1}^{n} r(y(s_i))^{z_P(s_i)} \\ & \times \exp\{-.001\lambda_y\} \\ & \times \lambda_T^{\frac{K}{2}} \exp\{-\frac{1}{2}\lambda_T||x_T||^2\} \times \exp\{-.001\lambda_T\} \\ & \times \lambda_P^{\frac{K}{2}} \exp\{-\frac{1}{2}\lambda_P||x_P||^2\} \times \exp\{-.001\lambda_P\} \\ & \times I[.04 \le \sigma_T, \sigma_P \le 1] \times I[-15 \le \beta_0, \beta_1 \le 15] \\ & \times I[0 \le \mu_0 \le 40] \times I[-30 \le \mu_1 \le 30], \end{aligned} \tag{6}$$

where $\boldsymbol{\theta} = (\lambda_y, x_T, \lambda_T, x_P, \lambda_P, \beta_0, \beta_1, \mu_0, \mu_1, \sigma_T, \sigma_P)$.

The posterior distribution is sampled using standard single site MCMC. Random walk Metropolis updates are used for the majority of the parameter components. Because the full conditionals for the precision parameters λ_y and λ_T are gamma, these parameters can be updated using Gibbs updates.

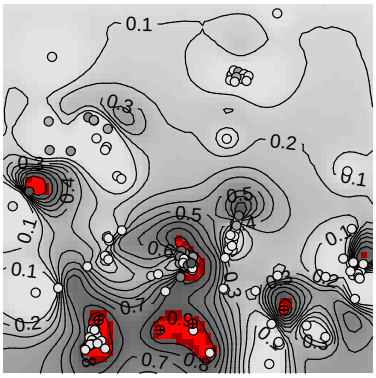

Figure 6: Posterior mean estimates for binary connectivity process $z_P(s)$. Bright red denotes the spatial region for which the probability of connectivity is greater that .95; dark red denotes area for which the probability of connectivity is greater than .90. The nominal odds map corresponding to the central frame of Figure 7 was used.

The MCMC was carried out by running four chains of length 10^6 and thinning by 250, for a posterior sample of size 16,000; each chain required roughly three days on a 2.5GHz quad-core desktop PC. This procedure results in an effective sample size that is no worse than 400 for any of the parameter components. In fact, the effective sample size for most parameters was far greater than 400; the mixing of (just) the mean parameters μ_0 and μ_1 is quite slow. The resulting posterior mean estimate for $z_P(s)$ is shown in Figure 6.

As the sampling is quite challenging, we devote the remainder of this subsection to a discussion of some of the issues that arose. As with many other spatial models fitted using a convolution approach, particularly those in which the knot values must be sampled via Metropolis or Metropolis-Hastings steps, it takes a considerable amount of time to run the MCMC code; each iteration is slow (at best, around 4-5 iterations per second) and many iterations are required for the MCMC to converge. In fact, convergence for the knot values x_P and for most of the knot values x_T is rapid. A small subset of the knot values x_T is problematic. In several locations in the study region, clay thickness observations obtained in tightly clustered areas differed significantly from one another, suggesting a fairly rough underlying process. The use of Gaussian kernels in our convolution approach is best suited for smooth underlying processes, with two resulting problems for fitting model to data:

The clay thickness kernel width attempts to shrink as much as its prior distribution allows, and thus its sampled values are tightly clustered near the minimum value allowed by the prior. At one point we experimented with a 30×30 grid of knot locations, and still the kernel widths shrank to the smallest width we allowed (1/30, recalling that the study area was rescaled to correspond to a unit square).

Secondly, knot values corresponding to locations near the apparent clay discontinu-

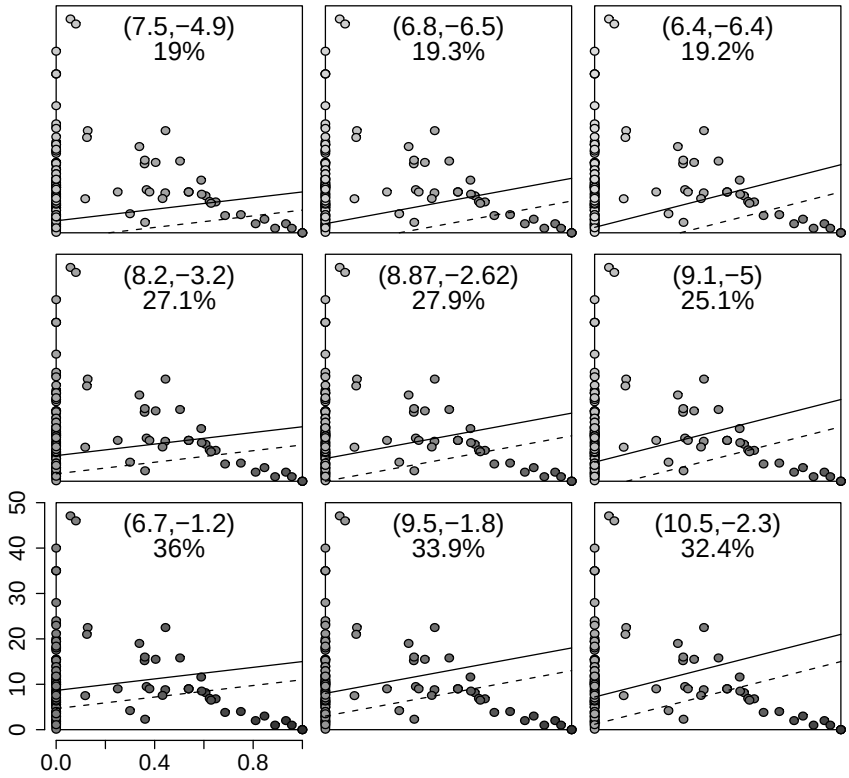

Figure 7: Odds maps based on expert judgment. The central figure corresponds to the best estimate; the remaining figures show other plausible odds maps. The solid line corresponds to data for which the odds of connectivity is 1. The dotted line corresponds to an odds of 10. For each odds map, the posterior proportion of the study region that is permeable is given. Above that are the posterior median estimates for (β_0, β_1).

ities converge slowly, exhibiting extremely high lag 1 autocorrelations. As part of our standard approach to fitting models with MCMC, we begin all our chains with an adjustment phase in which the Metropolis proposal widths for all sampled parameters are periodically recalculated so that the acceptance rates are in the 30-60% range. Individual knot values are given their own proposal widths, and the problematic locations have far narrower proposal widths than their neighbors, yet the problems persist. Correlated Metropolis-Hastings proposals might help (a proposed increase in one knot value could be accompanied by an equal-sized decrease in the value at an adjoining knot); but based on our attempts in earlier models to carry this out, we are not optimistic it would help here.

We mention that these issues with fitting the clay thickness portion of our model also arose in our separate attempts to fit a model for clay thickness only. We doubt the

problems here are caused by or are particularly influenced by our attempts to model simultaneously the proportion of coarse material.

In our efforts to construct an appropriate model for clay thickness – in particular, to handle bumpiness (i.e. small scale variation) without dramatically increasing the number of knots – we considered a likelihood function in which the clay thickness observations were correlated, using an exponential covariance function; in a clay-thickness-only model (results omitted), this resulted in marginally wider kernels, but had no noticeable effect in our bivariate-process model. We also attempted a model with two resolution scales, adding a much denser grid of knot locations over the subregion that we believed was causing the worst of our problems (inset (c) of Figure 2). This was ineffective, likely because the discontinuities in clay thickness appear in several other tightly clustered observation groupings as well.

4 Results

Figure 8 shows the posterior median for the clay thickness of aquitard Alpha. In general it is thinner near the south edge of the study region, as expected. Inference regarding the clay thickness field $z_T(s)$ is fairly insensitive to the choice of odds map since nearly all of the information regarding this field is contained in the thickness component of the likelihood $L_T(y_T(s)|z_T(s), \lambda_y)$.

The posterior mean for the binary connectivity field is given in Figure 9. Here the estimates vary depending on which odds map is used. Also given in Figure 7 are the posterior median estimates for β_0 and β_1 as well as the total proportion of the study site where possible interconnections between upper and deep aquifers are revealed under each of the 9 different formulations.

The reconstructions of connected regions show an increased probability of connectedness (permeability) for the aquitard in the southern portion of the study region. The magnitude of this probability varies depending on the odds map specified. All of the analyses show potential for discontinuity (permeability) in Alpha in the middle of the western side of the region. In Figure 9, dark red (flat medium gray) regions denote permeability probability greater than 0.90; bright red (dark gray) regions denote probability greater than .95.

The posterior distribution favors a rather short spatial correlation distance for the clay thickness field $z_T(s)$, and a large correlation distance for the process $u_P(s)$ used to build the connectivity field $z_P(s)$. This can be seen in the leftmost panels in Figure 4, which correspond to a single (yet fairly typical) posterior realization. The correlation distance for $z_T(s)$ is driven almost entirely by the clay thickness portion of the model. Extensive experimentation with a stand-alone model for clay thickness (using u_T and z_T as defined in this paper) resulted in an estimated value of σ_T that is indistinguishable from that obtained here. Evidently the roughness in $z_T(s)$ provides virtually all of the roughness called for in the permeability field $z_P(s)$, resulting in extremely smooth realizations for $u_P(s)$. Note that the north-south mean trend in $z_T(s)$ gives some larger

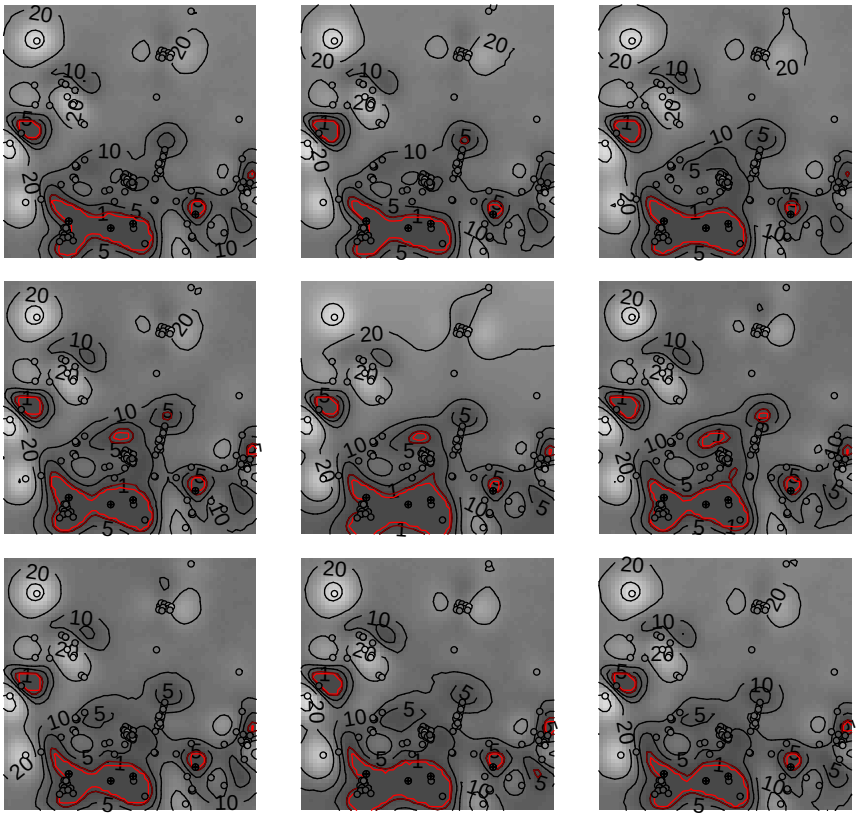

Figure 8: Posterior median estimates for the clay thickness field $z_T(s)$. The nine posterior maps correspond the odds maps given in Figure 7. The center panel corresponds to our preferred odds map; note that this estimate is somewhat insensitive to the choice of odds map. The red (medium gray) contours in the southern and west-central areas of the study region correspond to an estimated clay thickness of 0.1m.

scale dependence to $z_P(s)$.

Residuals for the fitted model are shown in Figure 10. For clay thickness, the residuals are simply $y_T(s_i) - \hat{z}_T(s_i)$, $i = 1, \ldots, n$, where $\hat{z}_T(s_i)$ denotes the posterior mean of the thickness field at s_i. For connectivity, the residual is defined to be the difference between the inferred probability from the local well core that the aquitard is permeable at s_i, i.e. $p(s_i) = r(s_i)/(1+r(s_i))$, and the posterior probability that $z_P(s_i) = 1$. The posterior mean for these two fields appears to match the data adequately.

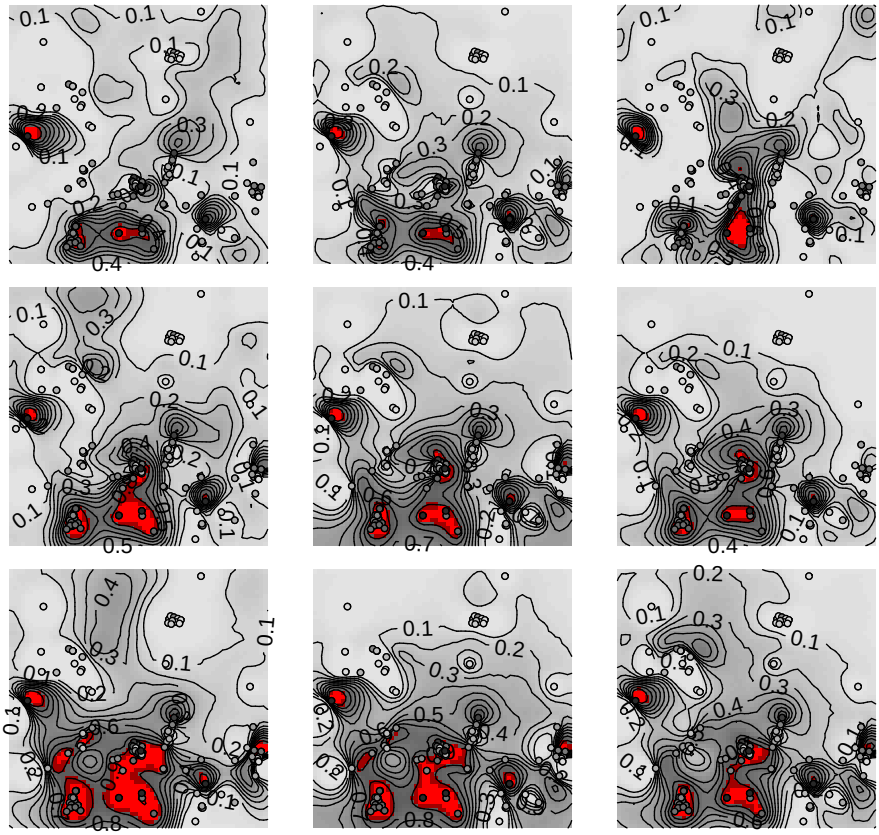

Figure 9: Pointwise posterior probability of connectivity for aquitard Alpha. The nine posterior maps correspond to the posterior expectation of $z_P(s)$ using the corresponding odds maps given in Figure 7. The center panel corresponds to our preferred odds map.

4.1 Comparing the reconstruction to other information sources

This analysis is complicated by the fact that the sedimentological data do not give direct information on whether or not aquitard Alpha is permeable at any particular spatial location. In this section we compare the estimated connectivity field with information from piezometer readings from other municipal data sources. In addition, we also see how results obtained in the previous section compare with qualitative reconstructions which utilize sedimentological information along with the well cores used for the statistical analysis.

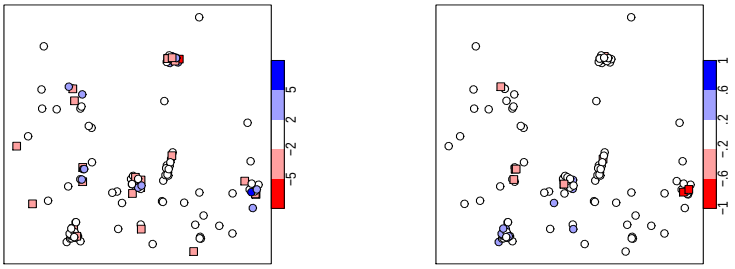

Figure 10: Residuals from the posterior mean for $z_T(s_i)$ and $z_P(s_i)$. For clay thickness (left panel), the residuals are $y_T(s_i) - \hat{z}_T(s_i)$, $i = 1, \ldots, n$. For connectivity (right panel) the residual is defined to be the difference between the inferred probability that the aquitard is permeable at s_i, $p(s_i) = r(s_i)/(1 + r(s_i))$, and the posterior probability that $z_P(s_i) = 1$. A red color (square symbol) indicates a negative residual.

Comparison to piezometer readings

If aquitard Alpha allows a connection between the upper and lower aquifers, then the water pressure (hydraulic head) should be the same in both aquifers at the location of this connection. This can be ascertained by piezometers which are capable of measuring hydraulic head in the upper and lower aquifers at (approximately) the same locations. Even though the regional network of Bologna comprises a large number of piezometers (Regione Emilia Romagna (1998)), only two pairs of piezometers satisfy this requirement. Their locations are denoted by the symbols 1 and 2 in Figure 11. We analyzed the average difference between hydraulic heads in the upper and lower aquifers $\triangle h$ at these two locations over the period of 1999 – 2000. The first pair of piezometers (symbol 1, designated by 4030P and 4028P in the data base of the municipality of Bologna) gives $\triangle h = 27.53m$. The second pair (symbol 2, designated by 5261Pa and 5261Pb in the data base of the municipality of Bologna) gives $\triangle h = 35.37m$. Such large differences in hydraulic heads, combined with the remaining piezometric readings, strongly support the continuity of aquitard Alpha at these two locations. This is consistent with the central reconstruction of Figure 9 (as well as the others) which gives a posterior probability of .08 and .41 at those two locations. These data can also be incorporated into the data analysis by enforcing the condition $z_P(s) = 0$ at these two locations. The resulting connectivity estimate – computed via importance sampling using the original MCMC sample – is given in the right hand frame of Figure 11.

In addition to providing partial validation of our results, the piezometer data helps suggest another and perhaps more intuitive way our model might have arisen:

- Had we directly observed $z_P(s)$ so that the data $y_B(s_i)$ were 1's (permeable) and 0's (not permeable), we could have created a simple binary spatial model where $u_P(s)$ is a mean 0 gaussian process with some spatial dependence. The resulting

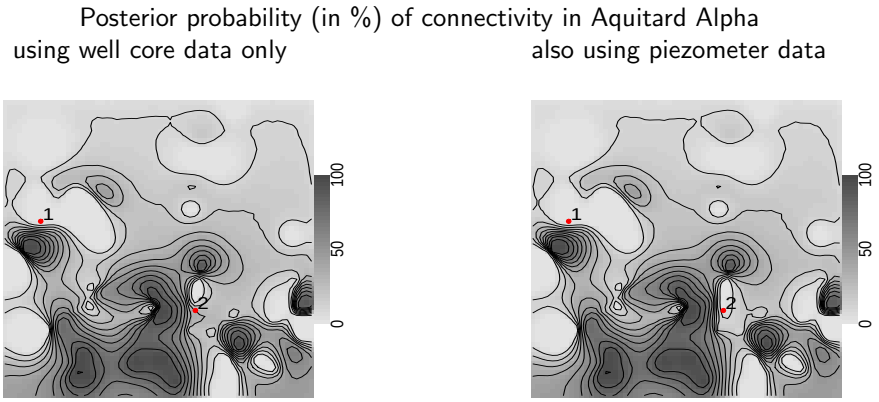

Figure 11: Posterior mean estimate of the connectivity field along with the spatial locations of the piezometers (denoted by the numerals 1 and 2). The left hand figure is based only on the well core data; the right hand figure also incorporates information from the piezometers, both of which suggest impermeability (no connectivity) at those locations.

formulation would be:

$$\begin{aligned} L(y_B|z_P(s)) &\propto \prod_{i=1}^{n} I[z_P(s_i) = y_B(s_i)] \\ z_P(s) &= I[u_P(s) > 0] \\ u_P(s) &\sim GP(0, C_P(s, s')) \end{aligned}$$

- Now, if the clay thickness process $z_T(s)$ were measured (with error) at spatial locations, we have a standard likelihood for the thickness data $y_T(s)$. We build in dependence between $z_P(s)$ and $z_T(s)$ and let the data determine the strength of this dependence, by adjusting the mean of the latent GP:

$$\begin{aligned} L(y_B|z_P(s)) &\propto \prod_{i=1}^{n} I[z_P(s_i) = y_B(s_i)] \\ z_P(s) &= I[u_P(s) > 0] \\ u_P(s) &\sim GP(\beta_0 + \beta_1 z_T(s), C_P(s, s')) \\ L(y_T|z_T(s)) &\propto \exp(-\frac{\lambda_y}{2}||y_T - z_T||^2) \\ z_T(s) &= \max(0, u_T(s)) \\ u_T(s) &\sim GP(\mu_0 + \mu_1 s_2, C_T(s, s')). \end{aligned} \tag{7}$$

This is where we would find ourselves if our data consisted of clay thickness and piezometer data measured jointly at multiple locations. In this case, the formulation is rather straightforward. Here the dependence of $z_P(s)$ on $z_T(s)$ is

appropriate. In fact, from a modeling standpoint, it is a bad idea not to account for this dependence. Knowing the thickness of the aquitard at s should inform us about its permeability at s.

- This formulation becomes more unwieldy because we have to replace the likelihood $L(y_B|z_P(s))$, in which the $z_P(s_i)$ are measured directly, with a weaker, subjective likelihood based on the bivariate data $y(s_i) = (y_T(s_i), y_P(s_i))$, i.e. the thickness of fine material and proportion of coarse material in the core sample. The formulation now becomes:

$$\begin{aligned} L(y|z_P(s)) &\propto \prod_{i=1}^{n} r(y(s_i))^{z_P(s_i)} \\ z_P(s) &= I[u_P(s) > 0] \\ u_P(s) &\sim GP(\beta_0 + \beta_1 z_T(s), C_P(s, s')) \end{aligned}$$

The quantities $L(y_T|z_T(s)), u_T(s)$ and $z_P(s)$ are specified as above. The likelihood terms $r(y(s_i))^{z_P(s_i)}$ allow the possibility that $z_P(s_i)$ may be either 1 or 0, and the data, $(y_T(s_i), y_P(s_i))$, determine the relative odds of these two outcomes through the specification of γ_0, γ_1 and γ_2. This results in the form of our final model.

Comparison to qualitative geological cross-sections

The geological data set consists of the complete 123 well logs (stratigraphic columns), from which the sedimentological data described in Section 2 have been extracted. By supplementing this with the knowledge of the dynamics of depositional processes over the geological time scale, qualitative, expert-based reconstructions of geological cross-sections are produced, as we discuss momentarily. Six of these are shown in Figure 12. In these cross-sections, the dark color indicates the fine material inclusions into the coarse material depositional structures. The central figure shows the posterior mean connectivity field, along with the spatial locations of these cross-sections. Darker colors indicate greater probability of vertical connectivity between the aquifers. The red ovals indicate the approximate location and extent of aquitard Alpha.

Construction of the cross sections is performed upon identifying boreholes which are more-or-less aligned and then connecting ("correlating") segments between the boreholes where similar materials are found. These connections are constrained by the knowledge of depositional sequence. For instance, if one knows that it is unlikely that an aquitard deepens along a given direction (e.g. N-S), then he/she does not connect regions of fine materials that are at different elevations in two wells. These reconstructions are not performed by any formal interpolation techniques, but rather on the basis of the information just described.

Cross-sections S1–S4 located in the southern part of the study area show little or no fine materials at the depths where aquitard Alpha should be. In each of these cases, the cross-sections are consistent with our final reconstructed map in showing a high probability of connectivity (permeability) between the aquifers immediately above and below Alpha.

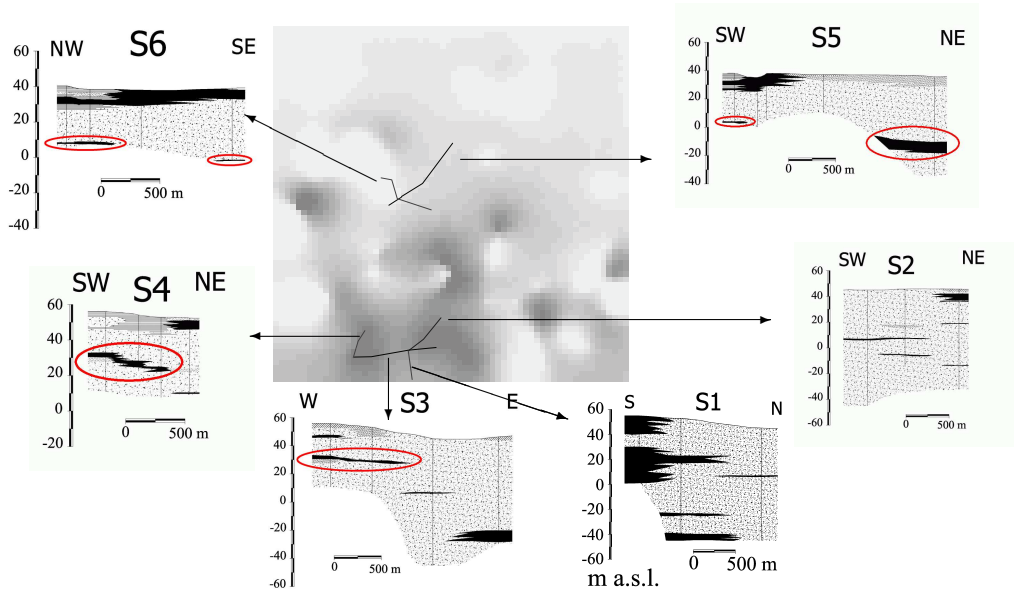

Figure 12: Qualitative geological sections. The posterior mean estimate of the connectivity field (center) along with qualitatively estimated cross-sections of the geology. These qualitative cross-sections were constructed using the well cores described in Section 2 along with consideration of the dynamics of depositional processes over the geological time scale. Dark regions in the cross-sections denote fine-scale (low-conductivity) inclusions and/or aquitard bodies. In the central connectivity map, darker regions denote areas where aquitard Alpha is likely to be more permeable, thus offering a less effective separation between the upper and lower aquifers. Red ovals indicate the approximate location and extent of aquitard Alpha.

The northern cross-sections S5 and S6 show very noticeable dark regions in the shallow depths where aquitard Alpha is located. Again, this is consistent with our reconstruction, which estimates a small probability of connectivity in Alpha near those locations. In all cross-sections, S1–S6, the vertical extent of the reconstruction is limited to the depths of the well cores used for that figure. Thus the seeming gaps in aquitard Alpha in S5 and S6 are due to missing, or inadequate, data; in all wells in the northern part of the study region which reach sufficient depths there *is* evidence of Alpha. Contrast this with the gaps seen in S1–S4.

5 An alternative modeling approach

One of the reviewers suggested an alternative approach we might have taken in modeling these data. In this section, we briefly describe this approach, along with our assessment of its strengths and weaknesses. The ultimate assessment of the relative merits of this alternative requires an actual fitting of the model to this data. This is beyond the scope of this paper.

Define two latent processes, $z_T(s)$ (clay thickness process, the same as $z_T(s)$ elsewhere in this paper) and $z_{\text{coarse}}(s)$, reflecting the proportion of coarse material. The data are just noisy measurements of these fields at the n spatial measurement locations. The critical latent unknown, namely connectivity, could then be a third process, $z_P(s)$ that is related to a functional, $g(z_T(s), z_{\text{coarse}}(s))$, of the values of $z_T(s)$ and $z_{\text{coarse}}(s)$. The function g would be a modification of the subjective odds models (5), substituting latent process values for the observations, such as $g(z_T, z_{coarse}) = \gamma_0 + \gamma_1 z_T + \gamma_2 z_{\text{coarse}}$.

As in any prior elicitation, one could use real examples of well measurements to elicit the prior on connectivity as a function of thickness and proportion of coarse material (as has been done here), and then in the actual model have the probit (or logit) of the probability of connectivity be $g(z_T, z_{\text{coarse}})$. This is in keeping with standard statistical formulations. Figuring out the posterior for connectivity would then be a matter of calculating a functional of well-defined latent processes, $z_T(\cdot)$ and $z_{coarse}(\cdot)$. These latent processes are related directly to the analogous observables in a standard modeling fashion. So the elicitation of g would then play out in the posterior as a subjective functional rather than a subjective likelihood.

The suggestion of formulating the spatial distribution of $z_P(s)$ as a function of $g(z_T(s), z_{\text{coarse}}(s))$ is appealing since it directly connects $z_T(s)$ and $z_{\text{coarse}}(s)$ to the observations. However linking $z_P(s)$ to these (partially) observed processes seems, to us, to be more complicated than it first appears. Recall that we need $z_P(s)$ to be a binary spatial process over the continuous 2-d space $\mathcal{S}$. So having $g(z_T(s), z_{coarse}(s))$, which gives a pointwise probability of connectivity at s, is not sufficient to produce binary, realizations of $z_P(s)$. One obvious candidate is a spatial extension of the standard probit model:

$$\begin{aligned} u_P(s) &= \gamma_0 + \gamma_1 z_T(s) + \gamma_2 z_{\text{coarse}}(s) + \epsilon(s) \\ \epsilon(s) &\sim GP(0, C_\epsilon(s, s')) \\ z_P(s) &= I[u_P(s) > 0]. \end{aligned}$$

As stated by the reviewer who suggested this alternative approach, the advantage here is that the data inform about z_{coarse} in a standard way. The price paid is that the model now requires three spatial processes to be estimated: $z_T(s)$, $z_{\text{coarse}}(s)$, and $\epsilon(s)$.

Note that with this approach, we must still use expert judgment to determine γ_0, γ_1 and γ_2 as we did with our approach. Furthermore, expert judgement is also required to specify $\epsilon(s)$. Setting $\epsilon(s) \equiv 0$, makes $z_P(s)$ a deterministic function of $z_T(s)$ and $z_{\text{coarse}}(s)$; we feel this does not allow enough flexibility in using the experts' judgment. Hence we think the approach we've presented in this paper is sensible, and we doubt there are substantial gains in simplicity to be had by switching to this suggested alternative approach.

6 Discussion

This analysis has identified locations in the study region for which there is cause for concern regarding the ability of aquitard Alpha to protect the deep aquifers system from contaminants. Although the quantitative results of this analysis depend on the subjective odds map, the qualitative nature of the estimated permeable inclusions is fairly stable for a range of odds maps shown in Figure 7. The southern region, along with the central part of the western edge of the study area, shows potential for connectivity between the upper and lower aquifers. The likelihood specification is typically somewhat subjective in spatial, binary classification problems – Besag et al. (1991), Hurn (1997) and Higdon (1998) give examples of this. We note out that a linear odds map could have been induced by assuming the data arise from two normal populations with different means, but a common covariance, as in linear discriminant analysis. However, in this application, our focus is on how the odds change with the data, not the means of these two normal distributions. More recently Rappold et al. (2007) use a subjective likelihood to identify the mixing layer boundary in an oceanographic application.

This study led us to a new, non-Gaussian, bivariate spatial model that links a continuous, non-negative field with a binary field. We originally tried a single binary spatial model for the permeability field, but its behavior in the northern region of the study area, where Alpha is thick, was not satisfactory. We feel that coupling the clay thickness and permeability gives more realistic results.

We used normal kernels in the convolution representation of both latent fields, $u_T(s)$ and $u_P(s)$, for computational reasons. The nonstandard model formulations of Section 3.1, as well as the need for predictions over a fine grid of spatial locations, require that $z_T(s)$ and $z_P(s)$ can be quickly produced given the model parameters. The convolution construction of the latent processes answers both of these needs. The resulting processes $u_T(s)$ and $u_P(s)$ are, to a close approximation, stationary zero mean Gaussian processes with a Gaussian covariogram. A limitation of such representations is that they do not attempt to model very small scale variation. Since the residuals and the small kernel widths for $z_T(s)$ hint that small scale variation may be present, future research will focus on developing practical ways to incorporate this effect in such nonstandard models.

References

Agresti, A. (1990). *Categorical Data Analysis*. John Wiley & Sons. 563

Amorosi, A. and Farina, M. (1994a). "Sequenze deposizionali nei depositi alluvionali quaternari del primo sottosuolo nell'area ad est di Bologna." In *Proceedings of the 1st European Congress on Regional Geological Cartography and Information Systems*, volume 5, 35–54. 559

— (1994b). "Stratigrafia della successione quaternaria continentale della pianura Bolognese mediante corelazione di dati di pozzo." In *Proceedings of the 1st European Congress on Regional Geological Cartography and Information Systems*, volume 5, 16–34. 559

— (1995). "Large-scale architecture of a thrust-related alluvial complex from subsurface data: the Quaternary succession of the Po Basin in the Bologna area (Northern Italy)." *Giornale di Geologia*, 57: 3–16. 559

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics (with discussion)." *Annals of the Institute of Statistical Mathematics*, 43: 1–59. 565, 579

De Oliveira, V. (2000). "Bayesian Prediction of Clipped Gaussian Random Fields." *Computational Statistics & Data Analysis*, 34(3): 299–314. 563

Francavilla, F., D'Onofio, S., and Toni, G. (1980). "Hydrogeological, structural, paleoecological and stratigraphical characteristics of the Reno alluvial fan (Bologna)." *Quaderni Ist. Ric. Acque C.N.R.*, 51: 81–95. 559

Higdon, D. (1998). "Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications." *Journal of the American Statistical Association*, 93(442). 579

— (2002). "Space and space-time modeling using process convolutions." In Anderson, C., Barnett, V., Chatwin, P. C., and El-Shaarawi, A. H. (eds.), *Quantitative Methods for Current Environmental Issues*, 37–56. London: Springer Verlag. 562

Hurn, M. (1997). "Difficulties in the Use of Auxiliary Variables in Markov Chain Monte Carlo Methods." *Statistics and Computing*, 7: 35–44. 579

— (1998). "Confocal Fluorescence Microscopy of Leaf Cells: An Application of Bayesian Image Analysis." *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 47: 361–377. 565

Johnson, V. E. and Albert, J. H. (1999). *Ordinal Data Modeling*. Springer-Verlag Inc. 563

Rappold, A., Lavine, M., and Lozier, S. (2007). "Subjective Likelihood for the Assessment of Trends in the Ocean's Mixed-Layer Depth." *Journal of the American Statistical Association*, 102(479): 771–780. 579

Regione Emilia Romagna (1998). "Riserve idriche sotterranee della Regione Emilia Romagna (Groundwater resources of the Emilia Romagna region)." Technical report, Firenze. 574

Ricci Lucchi, F. (1984). "Flysh, molassa, clastic deposits: traditional and innovative approaches to the analysis of north Apeninic basins." In *Cento Anni di Geologia Italiana*, volume Giubilare 1 centenario, Societa Geologica Italiana, 279–295. 559

Ricci Lucchi, F., Colalongo, M. L., Cremonini, G., Gasperi, G., Iaccarino, S., Papani, G., Raffi, S., and Rio, D. (1982). "Sedimentary and palaeogeographic evolution of the apenninic margin." In *Guida alla geologia del margine appenninico padano*, volume Guide geologiche regionali, Societa Geologica Italiana, 17–46. 559

Acknowledgments

The authors are grateful for helpful comments provided by the reviewers.

Part of the work, which includes stratigraphic and sedimentological data assimilation, was performed under the auspices of the European Commission (contract EVK1-CT-1999-00041-W-SAHaRA). Financial support by the Marie Curie Initial Training Network "Towards Improved Groundwater Vulnerability Assessment (IMVUL)" is gratefully acknowledged.