Mathematical Geology, Vol. 32, No. 4, 2000

Elicited Data and Incorporation of Expert Opinion for Statistical Inference in Spatial Studies[1]

Subhash R. Lele[2] and Abhik Das[3]

Spatial data are often sparse by nature. However, in many instances, information may exist in the form of "soft" data, such as expert opinion. Scientists in the field often have a good understanding of the phenomenon under study and may be able to provide valuable information on its likely behavior. It is thus useful to have a sensible mechanism that incorporates expert opinion in inference. The Bayesian paradigm suffers from an inherent subjectivity that is unacceptable to many scientists. Aside from this philosophical problem, elicitation of prior distributions is a difficult task. Moreover, an intentionally misleading expert can have substantial influence on Bayesian inference. In our experience, eliciting data is much more natural to the experts than eliciting prior distributions on the parameters of a probability model that is a purely statistical construct. In this paper we elicit data, i.e., guess values for the realization of the process, from the experts. Utilizing a hierarchical modeling framework, we combine elicited data and actual observed data for inferential purposes. A distinguishing feature of this approach is that even an intentionally misleading expert proves to be useful. Theoretical results and simulations illustrate that incorporating expert opinion via elicited data substantially improves the estimation, prediction, and design aspects of statistical inference for spatial data.

KEY WORDS: Bayesian inference, elicited prior, hierarchical models, honesty parameter, kriging, optimal sequential sample design.

INTRODUCTION

Spatial data are often sparse by nature, and in many applications, the studies are frequently characterized by lack of sufficient data on the quantity of interest (Journel, 1986; Kulkarni, 1984). Limited amount of data leads to a relatively flat likelihood surface that is not very informative. One way out of this problem is to augment the available data by incorporating other available sources of information. In many instances, hard data on an attribute such as pollutant concentration, or presence/absence of species, is difficult to come by. However, a wealth of information

[1]Received 5 October 1998; accepted 1 December 1998.
[2]Department of Mathematical Sciences, University of Alberta, Edmonton, T6G 2G1 Canada. e-mail: slele@ulberta.ca
[3]Statistics Research Division, Research Triangle Institute, 6110 Executive Blvd., Suite 420, Rockville, Maryland 20852.

0882-8121/00/0500-0465$18.00/1

may exist in the form of "soft" data such as expert opinion about whether pollutant concentration exceeds a certain threshold or not (Journel, 1986; Kulkarni, 1984). Further, scientists in the field often have a good understanding of the phenomenon under study and can provide valuable information on its likely behavior. It will thus be useful to have a sensible mechanism that incorporates such soft information or expert opinion in the process of inference.

One of the standard approaches for incorporating expert opinion is the Bayesian paradigm (Steffey, 1992; DeGroot, 1988; Genest and Zidek, 1986; von Mises, 1943). Unfortunately, scientists in the field are reluctant to use the Bayesian paradigm for a variety of reasons (Royall, 1997; Mayo, 1996; Dennis, 1996; Efron, 1986). It is the subjectivity inherent in specifying a prior that is most troubling to the scientists. For example, Efron (1986) observes that "subjectivism . . . has failed to make much of a dent in scientific statistical practice" because "strict objectivity is one of the crucial factors separating scientific thinking from wishful thinking." A related problem is that the prior provided by the expert may be intentionally misleading. For example, in environmental sampling for pollution investigations, an industry expert may want the sampling effort to be concentrated on the top of a hill in the study area where there is little chance of finding any contamination. An environmental activist, on the other hand, may want it to be exclusively concentrated at the bottom of the hill, where a high degree of contamination may be expected. Dennis (1996) discusses other instances in ecological applications where subjective priors elicited from intentionally misleading experts could have serious public policy implications. He argues that "the Bayesian philosophy of science is scientific relativism" where "truth is subjective" (Dennis, 1996). This "scientific relativism" is unacceptable to many scientists. Moreover, Royall (1997) and Journel (1986) have demonstrated that the standard Bayesian practice of expressing lack of prior information as a uniform distribution can also be patently misleading.

Even if we set aside the philosophical objections to the Bayesian formulation, quantification of prior belief in the form of a probability distribution on the parameter space, is a complex problem (Garthwaite and Dickey, 1992). West (1988) observes that ". . . it is often (or rather, always) difficult to elicit a full distribution with which an expert is totally comfortable." Why is it so difficult to elicit a prior distribution? It is important to recognize that the concept of a prior probability distribution on the parameters of a statistical model is a statistical construct that is hard for most scientists to visualize. It is more natural for an expert to think in terms of the process under study and not in terms of statistical distributions over a parameter space. The sensible approach, then, is to ask the expert to provide guess values for observable data, not a prior probability distribution.

Kulkarni (1984) and Journel (1986) have suggested methods that modify the Bayesian framework to enable the use of expert opinion in the form of elicited data. They formulate the problem in the following manner. Uncertainty about the unknown process under study is characterized by its probability distribution, conditioned by available expert information. This posterior probability distribution is estimated by combining both expert opinion and actual observations. Inferences about the process are based on this distribution. Suppose we have actual data $Z(x_i)$ for a process $Z(x)$ observed at n locations x_i, $i = 1, \ldots, n$. Similarly, we denote expert information in the form of elicited data, as $E(x_j)$, which is available for m locations, $j = 1, \ldots, m$. Then, the posterior c.d.f. $F(.)$ is estimated by

$$F^*(c) = \sum_{i=1}^{n} \lambda_i I[Z(x_i) \leq c] + \sum_{j=1}^{m} \gamma_j G_j(c)$$

where $G(.)$ denotes the prior probability distribution, I's are indicator functions, and the λ's and γ's are weights assigned to different data points. The prior distribution $G(.)$ characterizes the expert opinion, and it has to be constructed from the expert data $E(x_j)$, $j = 1, \ldots, m$. One way to do this is to assume a suitable functional form for $G(.)$, and then estimate the parameters of this distribution from the given expert data (Kulkarni, 1984). Once $G(.)$ is available, a point predictor for any unobserved location x_p is the conditional expectation associated with the posterior c.d.f. (Journel, 1986) given by

$$\hat{Z}(x_p) = \int_{R(Z)} z \cdot dF^*(z(x_p))$$

where the integration is over $R(Z)$, the range of all possible values of Z. Note that the weights λ_i, $i = 1, \ldots, n$, and γ_j, $j = 1, \ldots, m$, in $F^*(.)$ are considered optimal in the best linear unbiased prediction sense, and determined by standard kriging methods (Cressie, 1993, p. 105), so that the prediction error for $Z(x_p)$ is minimized.

In this paper we utilize and expand on the ideas of Journel and Kulkarni to augment our knowledge about the process under study by the use of data elicited from experts. However, instead of the paradigm used by Kulkarni (1984) and Journel (1986), we use hierarchical models to formulate the problem of combining observed data and guess values obtained from experts. In a spatial data setup, information may be elicited from experts in the form of observable data, for the whole study area. Hard data would generally be available for a limited number of locations in the same area. The problem then essentially becomes that of drawing inferences about the underlying process from a combination of the two likelihoods, one provided by the elicited data, and one provided by the observed data. This approach is similar in spirit to Efron's method (1996) of assessing information concerning one set of data by combining information from several experiments. He considered the following problem. Suppose K independent experiments are conducted, yielding data $\underline{x}_k$, $k = 1, \ldots, K$. These data are generated from K different likelihood functions $L_k(\theta_k \mid \underline{x}_k)$, where θ_k are the parameters of interest, $k = 1, \ldots, K$. The statistical problem consists in combining these K likelihoods in order to make inferences on any one of the θ_k's. Empirical Bayes likelihood theory is used to combine all these likelihoods to get an interval estimate for some θ_k (Efron, 1996).

The situation that we are concerned with in this paper is, however, different from that considered by Efron in one crucial aspect. In our approach, the two sets of data (expert guesses and actual observations) cannot be assumed independent, as they relate to the same underlying phenomenon. A hierarchical Empirical Bayes type model still provides the basis for combining these two kinds of information, with the caveat that it should allow for dependence between the two sets of data. This dependence reflects the credibility of the expert, and can be taken into account in the inferential process. In this way, even a misleading expert opinion can be informative. For instance, if we find that data elicited from an expert are negatively correlated with real data, this is useful information that can be used to suitably adjust our inferences.

The goal of this paper is to show how expert opinion or elicited data can be incorporated to augment likelihood inference in the spatial context. Key features of our approach are the elicitation of observable data (and not prior distributions) from the expert, and the ability to validate it, so that inferences can be calibrated in a suitable manner. The paper is organized as follows. In the next section we present a general modeling framework for augmenting observed data by incorporating data elicited from experts. In the following three sections, we apply this model for different data structures and explore the effect of adding expert data on different aspects of statistical inference, including estimation of model parameters, prediction of unobserved values and designing of optimal sampling schemes. We conclude with some general observations.

A HIERARCHICAL MODEL FOR THE ELICITED DATA

In this section we propose a hierarchical modeling approach that provides a framework for the augmentation of observed data with data elicited from experts. We couch our discussion in terms of the finite population setup, which is natural for spatial data. Suppose there are N locations indexed by $1, 2, \ldots, N$, in a study area D. Let the true values of the quantity of interest at these locations be denoted by $\underline{Y} = (Y_1, Y_2, \ldots, Y_N)$. The Y's may be numeric, categorical or binary. For example, we may be interested in the amount of pollutant contamination (numeric values), level of pollutant contamination—low, medium, or high (categorical values)—or presence/absence of particular species (binary values). Suppose we have only one expert. We ask the expert to provide his/her best guess for the values of $\underline{Y}$. Let the expert guess values, the elicited data, be denoted by $\underline{E} = (E_1, E_2, \ldots, E_N)$. These may, in turn, be numeric, categorical or binary. For instance, if the expert is prepared to guess exact values for the amount of pollutant contamination, then we have numeric expert data. On the other hand, it may be much easier for an expert to provide categorical (high, medium, or low contamination) or binary values (high or low) over the whole study area, rather than coming up with exact numbers. We assume the following relationship between the truth $\underline{Y}$ and expert guess about the truth, $\underline{E}$:

- The vector $\underline{Y}$ is a realization from a distribution $f(\underline{y};\underline{\theta})$, indexed by unknown parameters $\underline{\theta}$.
- Given $\underline{Y} = \underline{y}$, expert opinion $\underline{E}$ has a distribution $g(\underline{e} \mid \underline{y}; \eta)$, where η is an unknown parameter that characterizes the dependence between $\underline{Y}$ and $\underline{E}$.

The parameter η plays a central role in the inferential process. It captures the degree of dependence between the truth and expert guess, thus providing a mechanism to evaluate the sincerity of an expert. Information from a good expert closely mimics reality, and this should be reflected in a high positive value of η. Conversely, a high negative value for η indicates misleading expert opinion. If η is very small, expert opinion is purely random, bearing no relation to reality. Thus, η may be thought of as an "honesty parameter," which tells us whether the expert is credible. Suppose we have sampled the process at n locations, with the sampled values denoted by $\underline{Y}_s = (Y_1, Y_2, \dots, Y_n)$, and the unknown values denoted by $\underline{Y}_{ns} = (Y_{n+1}, \dots, Y_N)$, $n < N$. The sampled observations, along with the expert guesses, enable us to estimate the honesty parameter η. This estimate helps in suitably calibrating any information provided by an expert, so that proper inferences can be made.

Notice that the model presented here is very general. We do not assume any particular distributional structure for either $\underline{Y}$ or $\underline{E}$. This enables different kinds of expert information to be combined with real data. For instance, while the actual response may be continuous, elicited data could be either continuous or discrete. In the following sections we explore specific instances of particular distributions for $\underline{Y}$ and $(\underline{E} \mid \underline{Y})$ and their implications for estimation and prediction, as well as sampling designs.

A SIMPLE SITUATION: NORMAL–NORMAL HIERARCHY

In this section we apply the hierarchical modeling framework in the relatively simple situation of Gaussian independence to gain some insight on the inferential aspects of this setup. This facilitates explicit derivation of theoretical properties concerning the effect of the honesty parameter η on estimation and prediction. Consider the following regression setup:

- Let Y_i be normally distributed with mean $\underline{x}_i\underline{\beta}$ and variance σ^2, $i = 1, 2, \dots, N$, independently, where $\underline{x}_i$'s are known covariates for location i and $\underline{\beta}$, the unknown vector of regression coefficients.
- $(\underline{E}_i \mid \underline{Y}_i)$ is independently normal with mean ηY_i, variance $\sigma^2(1 - \eta^2)$, $i = 1, 2, \dots, N$, where η is the unknown correlation coefficient between Y_i and E_i (1)

Thus, marginally, E_i's have an independent normal distribution with mean $\eta \underline{x}_i \underline{\beta}$ and variance σ^2, $i = 1, 2, \ldots, N$. As both the real and elicited data are normally distributed, the honesty parameter η is just the correlation coefficient between these two. It would be close to 1 for a good expert, -1 for a misleading expert, and nearly 0 for a random expert. In the following subsections we derive some relevant theoretical properties for the inferential aspects of this normal-normal mixture setup.

Estimation of Parameters

First, we consider estimation of the model parameters, which should be based on the observed sample of hard data $\underline{Y}_s = (Y_1, Y_2, \ldots, Y_n)$ and the expert supplied elicited data $\underline{E} = (E_1, E_2, \ldots, E_N)$. The likelihood function for parameters ($\underline{\beta}$, σ^2, η) is thus given by

$$L(\beta, \sigma^2, \eta \mid \underline{Y}_s, \underline{E}) = \prod_{i=1}^{n} [f(Y_i \mid x_i, \beta, \sigma^2) f(E_i \mid Y_i, \eta, \sigma^2)] \times \prod_{i=n+1}^{N} f(E_i \mid x_i, \beta, \eta, \sigma^2) \tag{2}$$

where distributions of Y_i, $(E_i \mid Y_i)$ and E_i are as given above. Let $\hat{\beta}$ denote the MLE based on ($\underline{Y}_s$, $\underline{E}$) and $\tilde{\beta}$ denote the MLE based on $\underline{Y}_s$ only. Standard calculations show that (see Das, 1998, for details) in these two cases, the Fisher Information for the estimates of β, given the other parameters, are

$$I(\hat{\beta} \mid \sigma^2, \eta) = \frac{1}{\sigma^2} X_s' X_s + \frac{\eta^2}{\sigma^2} X_{ns}' X_{ns} [1 - f(\beta, \sigma^2, \eta)]$$

where

$$f(\beta, \sigma^2, \eta) = \frac{[(N+n)\beta^2(1-\eta^2) X_{ns}' X_{ns}]/\sigma^2}{(N+n)\left[n + \frac{\beta^2}{\sigma^2}\{X_s' X_s + (1-\eta^2) X_{ns}' X_{ns}\} + \frac{2n\eta^2}{1-\eta^2}\right] - \frac{2n^2\eta^2}{1-\eta^2}} \tag{3}$$

and

$$I(\tilde{\beta} \mid \sigma^2) = \frac{1}{\sigma^2} X_s' X_s \tag{4}$$

where X_s and X_{ns} are the portions of the design matrix X, corresponding to the observed data $\underline{Y}_s$ and the unobserved data $\underline{Y}_{ns}$, respectively. Observe that,

$I(\hat{\beta}\,|\,\sigma^2, \eta = 0) = I(\tilde{\beta}\,|\,\sigma^2)$. This is natural because, when $\eta = 0$, the expert is useless. On the other hand, $I^{-1}(\hat{\beta}\,|\,\sigma^2, \eta = \pm 1) = \sigma^2(X'X)^{-1}$, where X is the covariate information corresponding to all N locations. This is equivalent to observing $\underline{Y} = (\underline{Y}_s, \underline{Y}_{ns})$ at all N locations. Further, note that the denominator of the function $f(\beta, \sigma^2, \eta)$ in (3) can be written as

$$\frac{1}{\sigma^2}[(N+n)\beta^2(1-\eta^2)X_{ns}'X_{ns}] + \frac{1}{\sigma^2}(N+n)\beta^2 X_s' X_s + \frac{2Nn\eta^2}{1-\eta^2}$$

which implies that $f(\beta, \sigma^2, \eta) < 1$. Since $(1/\sigma^2)X_{ns}'X_{ns} > 0$, it then follows from (3) and (4) that

$$I(\hat{\beta}\,|\,\sigma^2, \eta) > I(\tilde{\beta}\,|\,\sigma^2), \quad \forall \eta \neq 0$$

Thus, whether the expert is good or misleading, he helps improve the inference. A plot of $I^{-1}(\hat{\beta}\,|\,\sigma^2, \eta)$, a theoretical measure similar to MSE, vs. η (Fig. 1) shows that it is symmetric around $\eta = 0$. The quantity $I^{-1}(\hat{\beta}\,|\,\sigma^2, \eta)$ becomes smaller as the dependence between observed and elicited data gets stronger in either direction. Also observe that additional information gained from elicited data is an increasing

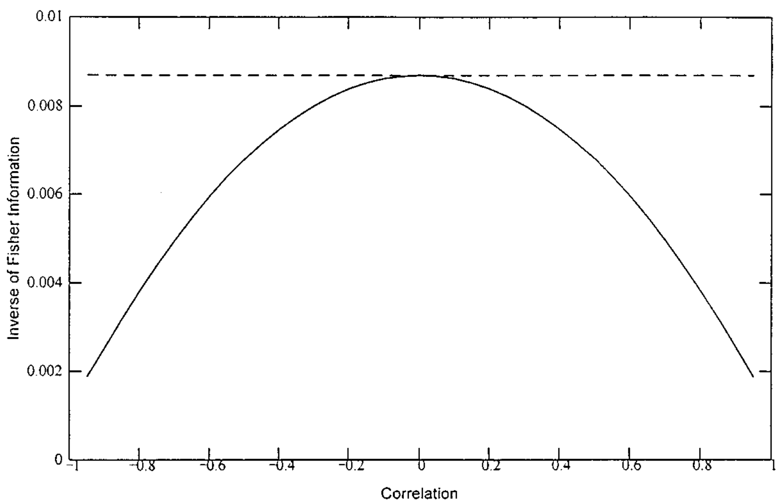

Figure 1. Plot of inverse of Fisher Information for the estimator of β with elicited data (solid line) and without elicited data (dashed line). This quantity, when based on expert opinion in the form of elicited data, gets smaller as correlation between real and elicited data differs from zero in either direction. Even the elicited data from an intentionally misleading expert (negative correlation) is useful.

function of $X'_{ns}X_{ns}$, which is large when the sample size n is small. Using expert opinion is thus most profitable when very little observed data are available (n is very small) and the covariates for the unobserved responses are substantial (i.e., $X'_{ns}X_{ns}$ is large). As more of actual data are observed (larger n), there is less use for any additional elicited data.

The theoretical results in this section indicate that using good or misleading data elicited from experts could significantly improve the estimation of parameters. However, we have introduced an additional parameter η into the model. What burden does the estimation of η place on the estimation of β, which is the parameter of interest, in a small sample situation? To answer this question, we performed simulations to investigate actual performance of the MLE's of β in terms of mean squared errors, both with and without expert opinion. We chose $\beta = 2$, $\sigma^2 = 1$, $N = 100$, and $n = 5$, with η varying from -1 to 1 in steps of 0.05.

Figure 2A shows a plot of the observed mean squared errors for β (MSE, averaged over 1,000 simulations) for the two competing scenarios (with and without elicited data) over the range of η. Estimates utilizing elicited data are always closer to the truth than those based solely on the observed data, unless the elicited data is characterized by small values of η, and does not add to our state of knowledge about β. More importantly, both good and misleading experts are equally useful. The similarity of this plot to Figure 1 where the theoretical inverse of the Fisher Information is plotted against $\eta \in (-1, 1)$ implies that MSE's produced by the simulations are close to their theoretical values. This indicates that soft data elicited from experts generally improves parameter estimates.

Prediction of Unobserved Values

The goal of a spatial analysis may not be estimation *per se*, but prediction at unsampled locations. It is equally important to provide precise measures of accuracy for such predicted values (i.e., prediction intervals). Here we investigate how the incorporation of expert opinion helps in this aspect of inference.

Note that under model (1),

$$E(Y_i \mid \underline{Y}_s\underline{E}) = E(Y_i \mid E_i) = X_i\beta + \eta(E_i - \eta X_i\beta)$$

and

$$\mathrm{var}(Y_i \mid \underline{Y}_s, \underline{E}) = \mathrm{var}(Y_i \mid E_i) = \sigma^2(1-\eta^2), \quad i = n+1, \ldots, N \tag{5}$$

Thus, given data $(\underline{Y}_s, \underline{E})$, the predictor for an unknown response Y_i would be

$$\hat{Y}_i = X_i\hat{\beta} + \hat{\eta}(E_i - X_i\hat{\eta}\hat{\beta}) \text{ with estimated prediction error } \widehat{\mathrm{var}}(\hat{Y}_i) = \hat{\sigma}^2(1-\hat{\eta}^2),$$
$$\text{and 90\% prediction intervals given by } \hat{Y}_i \pm 1.68 \times \sqrt{\widehat{\mathrm{var}}(\hat{Y}_i)}. \tag{6}$$

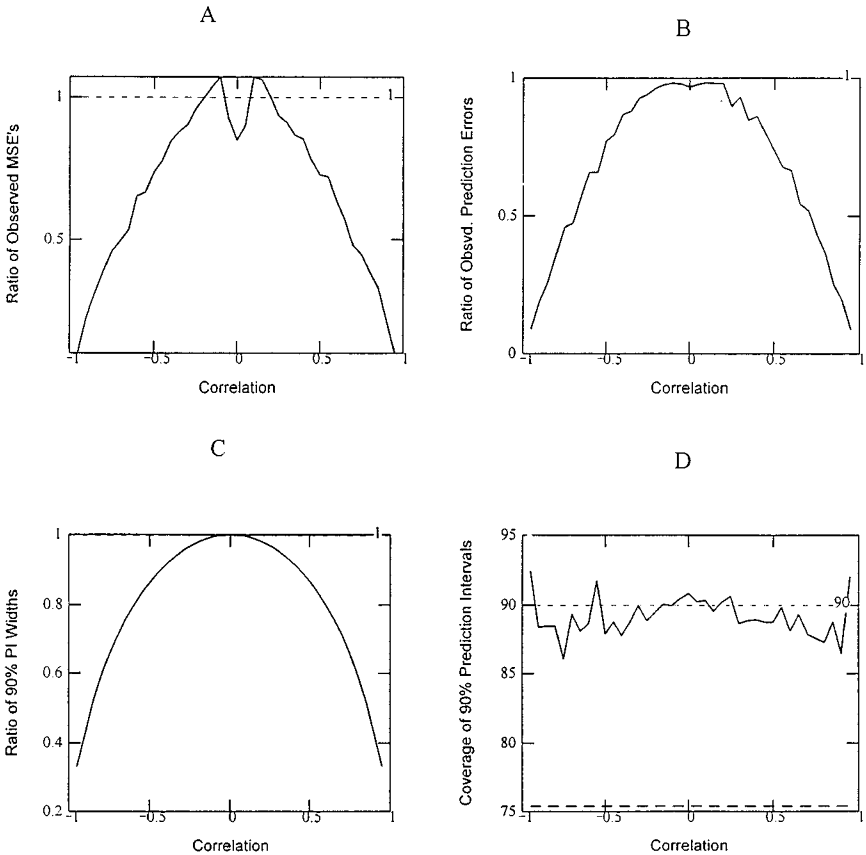

Figure 2. Simulation results for normal normal mixture, independent case: A, ratio of observed MSE($\hat{\beta}$) with elicited data, to MSE($\tilde{\beta}$), calculated without elicited data; B, prediction error; C, 90% prediction interval width (a dashed line indicates the value 1, where the two scenarios would have similar results); D, 90% prediction interval coverage, solid line for elicited data and dashed line for no elicited data. Estimation and prediction error for elicited data scenario gets smaller as correlation between real and expert data differs from zero in either direction. Observe that elicited data from an intentionally misleading expert is also useful. Also notice the similarity of these plots with theory.

Then, from (6), we see that for the perfect expert ($\eta = 1$), we simply use his value E_i as the predictor at the ith location. For the totally misleading expert ($\eta = -1$), we go in the exact opposite direction, with $-E_i$ as the predicted value. In both these cases prediction is perfect, as the prediction error is 0 [from (6)]. Further, elicited data from a random expert ($\eta = 0$) is ignored, and the standard linear predictor based only on the observed data, is used to predict Y_i. In that case, prediction error is the same as when there is no elicited data. Thus, the prediction error is very small when there is a high degree of dependence (positive or negative) between elicited data and the reality. It follows that prediction intervals based on these prediction errors should also be very precise for high absolute values of η.

We verified the prediction performance of our approach for utilizing elicited data through simulations. We fixed $\beta = 2$, $\sigma^2 = 1$, $N = 100$, and $n = 5$, while varying η from -1 to 1 in steps of 0.05. We fixed one of the nonsampled locations where prediction is desired. Let us call the corresponding value Y_p. At each iteration of the simulation, we used $(\underline{Y}_s, \underline{E})$ to predict Y_p, checked if the true value is covered by the prediction interval and also calculated $(\hat{Y}_p - Y_p)^2$. Similarly, we predicted Y_p using only $\underline{Y}_s$. We remind the reader that because this is a simulation study, the true value Y_p is *known*.

Figure 2B shows a plot of the observed prediction errors (values averaged over 1,000 simulations) based on observed and elicited data and observed data alone—for different values of η. Predictions that utilize data elicited from experts are very close to truth when there is strong correlation (positive or negative) between observed and elicited data, and perform no worse than the conventional predictors when elicited data is of little value (η close to 0).

Expert opinion expressed through elicited data also improves the coverage and precision of prediction intervals. A plot of the coverage properties under the two scenarios (expert and no expert) against different values of η (Fig. 2D) shows that actual coverage of nominal 90% prediction intervals using elicited data (mean of 89.1%) are consistently closer to their nominal value than prediction intervals based only on the observed data (mean of 75.4%). This correct coverage does not come at the cost of precision; which is clear from Figure 2C, where we plot average widths of the 90% prediction intervals for the two competing scenarios, against different values of η. As Equation (6) suggests, prediction interval width is smallest for an informative expert, when there is high dependence, and largest for a noninformative expert ($\eta \approx 0$). At that point it is very similar to prediction interval widths based solely on the observed data. Note that, even though prediction intervals based solely on observed data are generally wider than those based on observed and elicited data, coverage for the former is much less than the latter. This is because point predictors in the former case are quite inferior to those based on elicited data, and consequently, wider prediction intervals cannot compensate for their wrong centering values.

A couple of points on the honesty parameter η are worth noting here. First, from Equations (5) and (6) we can see that if η is estimated wrongly, there could be problems with both point and interval prediction of Y. Fortunately, since η captures the relationship between Y and E, its estimation draws on both these sources of information. Thus, even though we only have five observed data points ($n = 5$), we still have the whole population of elicited data ($N = 100$) available, and this helps in proper estimation of η. This is true for all the simulations presented in this paper, and in fact, η was uniformly well estimated in all these situations. A second related issued could be misspecification of the form of $g(\eta)$, the link function that postulates how Y and E are related. Thus, model (1) may not properly characterize the relationship between Y and E. In practice, we should protect against such problems of model misspecification by applying simple exploratory tools such as scatter plots of $(\underline{Y}_s, \underline{E})$ for building the hierarchical model for elicited data. Further, we should also do some model checking by studying diagnostic plots of residuals from the data fitted to this model. These graphical explorations would be useful in finding a proper form of the honesty parameter η for combining observed and elicited data.

Results in this section suggest strongly that expert opinion available in the form of elicited data over the entire study area can significantly improve estimation of parameters and predictions at unobserved locations. Not only are the point predictions better, the combination of observed and elicited data produces prediction intervals that are precise and have close to nominal coverage properties.

GENERALIZATION TO DEPENDENT DATA

In this section we generalize the ideas of the previous section to dependent data, and investigate how augmentation of observed data with elicited data may help in making inferences. This is important because most spatially distributed data are correlated. In the following, we modify the hierarchical model in (1) to incorporate spatial dependence:

- Let Y_i be normally distributed with mean $\underline{x}_i\underline{\beta}$, variance σ^2 and corr$(Y_i, Y_j) = \gamma^{d(i,j)}$, $(i, j) = 1, 2, \ldots, N$, where γ is a measure of spatial dependence, $d(i, j)$ is the distance between locations indexed i and j, $\underline{x}_i$'s are covariates for location i and $\underline{\beta}$, the vector of regression coefficients.
- $(E_i \mid Y_i)$ are independent normal with mean ηY_i and unit variance ($i = 1, 2, \ldots, N$), η being the correlation coefficient between Y_i and E_i. (7)

Notice that model (7) for $(E_i \mid Y_i)$ is slightly different, and probably more realistic, as compared to model (1). Here the correlation structure for the Y's assumes that spatial dependence among locations decreases with distance. The honesty parameter η has the same implications as before—it is positive for good experts, negative for misleading experts, and close to zero for random experts. Note that given observed and elicited data $(\underline{Y}_s, \underline{E})$, parameters $(\underline{\beta}, \sigma^2, \gamma, \eta)$ in model (7) can be estimated using method of maximum likelihood.

Estimation and Prediction: Simulations

Because of the introduction of dependence in the model, unlike Section 3, theoretical derivation of the Fisher Information for parameters $(\underline{\beta}, \sigma^2, \gamma, \eta)$ are mathematically intractable. We thus report the results of a simulation study. Recall that in Section 3 we have shown that simulations faithfully reproduce theoretical conclusions.

The simulation framework is similar to Section 3.1, except that now we choose $\gamma(=0.5)$, and generate $(\underline{Y}, \underline{E})$ under model (7). Figure 3A plots the ratio of observed MSE's for β (average of 1,000 simulations) in the two competing scenarios with and without elicited data, against η. Addition of a spatial dependence parameter does not appear to change the bottom line: estimates using expert opinion in the form of elicited data are almost always closer to their true values than those based solely on observed data, unless the expert information is random ($\eta \approx 0$). In Figures 3B and 3C we show similar plots for observed MSE's of σ^2 and γ, respectively. The conclusions are the same. Incorporation of expert information via elicited data, whether good or misleading, improves estimates of all the parameters, and even when it is just noise ($\eta \approx 0$), these estimates are not worse than MLE's based exclusively on the observed data $\underline{Y}_s$.

Next, we investigate the effect of expert data on prediction performances. It can be shown that under model (7), for any i that does not belong to the data Y_s,

$$E(Y_i \mid \underline{Y}_s, \underline{E}) = x_i\beta + (\Sigma_{i,n} \quad \eta\Sigma_{i,N}) \begin{pmatrix} \Sigma_{n,n} & \eta\Sigma_{n,N} \\ \eta\Sigma_{N,n} & I_N + \eta^2\Sigma_{N,N} \end{pmatrix}^{-1} \begin{pmatrix} \underline{Y}_s - \underline{x}_s\beta \\ \underline{E} - \eta\underline{x}\beta \end{pmatrix}$$

and

$$\mathrm{var}(Y_i \mid \underline{Y}_s, \underline{E}) = \sigma^2 - (\Sigma_{i,n} \quad \eta\Sigma_{i,N}) \begin{pmatrix} \Sigma_{n,n} & \eta\Sigma_{n,N} \\ \eta\Sigma_{N,n} & I_N + \eta^2\Sigma_{N,N} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{i,n} \\ \eta\Sigma_{i,n} \end{pmatrix} \qquad (8)$$

where $\Sigma_{N,N} = ((\sigma^2\gamma^{d(i,j)}))$ is the N by N variance—covariance matrix for $\underline{Y}$.

Predicted values for unknown Y_i's are thus given by substituting estimates for the corresponding parameter values in the conditional expectation in (8). Prediction intervals are based on the estimated prediction error. This is obtained by plugging in the corresponding parameter estimates in the conditional variance in (8). On the other hand, when no expert opinion is available,

$$E(Y_i \mid \underline{Y}_s, \underline{E}) = x_i\beta + \Sigma_{i,n}\Sigma_{i,n}^{-1}(\underline{Y}_s - \underline{x}_s\beta)$$

and

$$\mathrm{var}(Y_i \mid \underline{Y}_s, \underline{E}) = \sigma^2 - \Sigma_{i,n}\Sigma_{n,n}^{-1}\Sigma_{n,i}. \qquad (9)$$

We can similarly use (9), plugging in the corresponding estimates based on observed data, to get point predictions and prediction intervals when no expert is available.

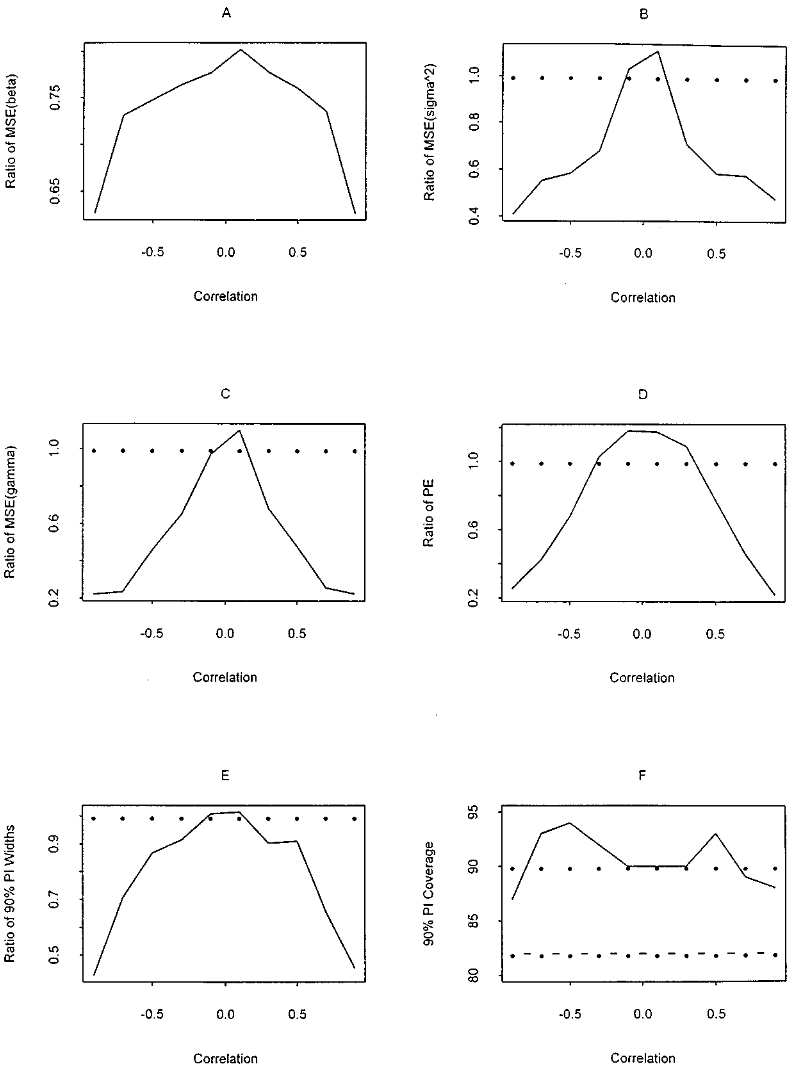

Figure 3. Simulation results for normal normal mixture, dependent case: A, ratio of observed MSE($\hat{\beta}$) with elicited data, to MSE($\tilde{\beta}$), calculated without elicited data; B, MSE(σ^2); C, MSE(γ); D, observed prediction error; E, 90% prediction interval width (a dashed line indicates the value 1, where the two scenarios would have similar results); F, 90% prediction interval coverage, solid line for elicited data and dashed line for no elicited data. Estimation and prediction error for elicited data scenario gets smaller as correlation between real and expert data differs from zero in either direction. Observe that elicited data from an intentionally misleading expert is also useful.

Figure 3D plots observed prediction errors (averaged over 1,000 simulations) based on both observed and elicited data and observed data alone—for different values of η. We see that the use of elicited data can greatly improve predictions, as long as the expert is informative, i.e., $|\eta|$ is high, though, for $\eta \in (-0.2, 0.2)$, point predictions are slightly worse than those based solely on the observed data. However, notice that coverage of 90% prediction intervals (Fig. 3F) for elicited data (mean of 90.6%) is much closer to the nominal value than that for observed data alone (mean of 82%). Also, prediction intervals based on elicited data are narrower than those based on observed data (Fig. 3E). Since precise prediction intervals are generally more important from a policy perspective than point predictions, these results again indicate that using elicited data in addition to observed data remains profitable even when the simple independence model in Section 3 is generalized to take account of dependence structures in the data. Simulations show that we get improved parameter estimates, as well as better predictions, as a result.

Sampling Designs

The preceding discussions have shown that augmentation of observed data with elicited data generally improves estimation and prediction. In this subsection, while confining ourselves to the same modeling framework as in (7), we deal with another aspect of inference—sampling design. In resource or pollution investigations, it is both financially and operationally impossible to obtain data at all spatial locations on the site under study. Good sampling strategy is thus of immense practical importance.

Sequential Design Algorithm

The primary goal in spatial studies is often prediction of unknown values. For instance, in pollution investigations, the quantity of interest is the amount of contamination at each of the unobserved locations. The aim in prediction sampling is to select a set of n locations such that based on these n observations and the assumed model, the rest of the $N - n$ values can be predicted with minimal prediction error. However, this selection is a computationally difficult task (Christakos and Killam, 1993; Benedetti and Palma, 1995; Lee and Ellis, 1996). An intuitively appealing and computationally manageable, although not necessarily optimal (Cressie, 1988; Cressie, Gotway, and Grondona, 1990), approach is to solve the problem sequentially. Suppose values at n sites have been already observed. Then the $(n + 1)$st site is added so that based on these $(n + 1)$ observations rest of the $(N - n - 1)$ unknowns can be predicted with least prediction error. This process is continued until the desired sample size is achieved.

Expert Opinion in Sample Design

In sequential sampling designs, selection of the first sample is critical and tricky. Usually, one just selects a simple random sample of size n and proceeds from there. In our formulation, however, one can use expert opinion in the form of elicited data to select the first sample. Of course, elicited data is used throughout the process as well.

Selection of the First Sample

Note that, under model (7),

$$(\underline{Y} \mid \underline{E}) \sim N[\boldsymbol{X}\beta + \eta\Sigma(I + \eta^2\Sigma)^{-1}(\underline{E} - \eta\boldsymbol{X}\beta),\, \Sigma - \eta^2\Sigma(I + \eta^2\Sigma)^{-1}\Sigma] \tag{10}$$

where Σ is as in (8). To obtain var$(Y_i \mid \underline{E})$, $i \notin s$, we need to obtain estimates of (σ^2, γ, η). Unfortunately, the marginal distribution of $\underline{\mathrm{E}}$ is:

$$\underline{\mathrm{E}} \sim N(\eta\boldsymbol{X}\beta,\, I + \eta^2\Sigma) \tag{11}$$

making only $(\eta\underline{\beta}^2, \eta^2\sigma^2)$ estimable. However, notice that var$(\underline{Y} \mid \underline{E})$ can be written as $\sigma^2[I - \sigma^{*2}\overline{R}(I + \sigma^{*2}R)^{-1}]R$, where $\beta^* = \eta\beta$, $\sigma^{*2} = \eta^2\sigma^2$ and R is the correlation matrix corresponding to Σ. Hence the *ranking* of locations in terms of their prediction errors depends on $[I - \sigma^{*2}R(I + \sigma^{*2}R)^{-1}]R$, which is an estimable quantity. We thus include that location i in the sample which has the largest value for the (i, i)th element of the estimated preediction error matrix $[I - \sigma^{*2}R(I + \sigma^{*2}R)^{-1}]R$, $i = 1, 2, \ldots, N$. Selection of the rest of the sample locations is straightforward. Given $(\underline{Y}_s, \underline{E})$, and the distribution of $(\underline{Y}_i \mid \underline{Y}_s, \underline{E})$, $i \notin s$ (8), we estimate the parameters $(\underline{\beta}, \sigma^2, \gamma, \eta)$ and choose that location that has the largest variability (prediction error), i.e., the one for which the estimated variance in (8) is the largest.

Simulations

We compare the above scheme with simple random sampling without replacement (SRSWOR). Notice that the ordinary sequential sampling scheme where no elicited data is used (Cressie, 1993, p. 314) corresponds to $\eta = 0$.

Results of the simulations are visually presented in Figures 4A–4C. In Figure 4A we plot average values (over 100 simulations) of the observed prediction errors based on sequentially designed samples (with and without elicited data) and SRSWOR, over different values of η. Sequential designs perform much

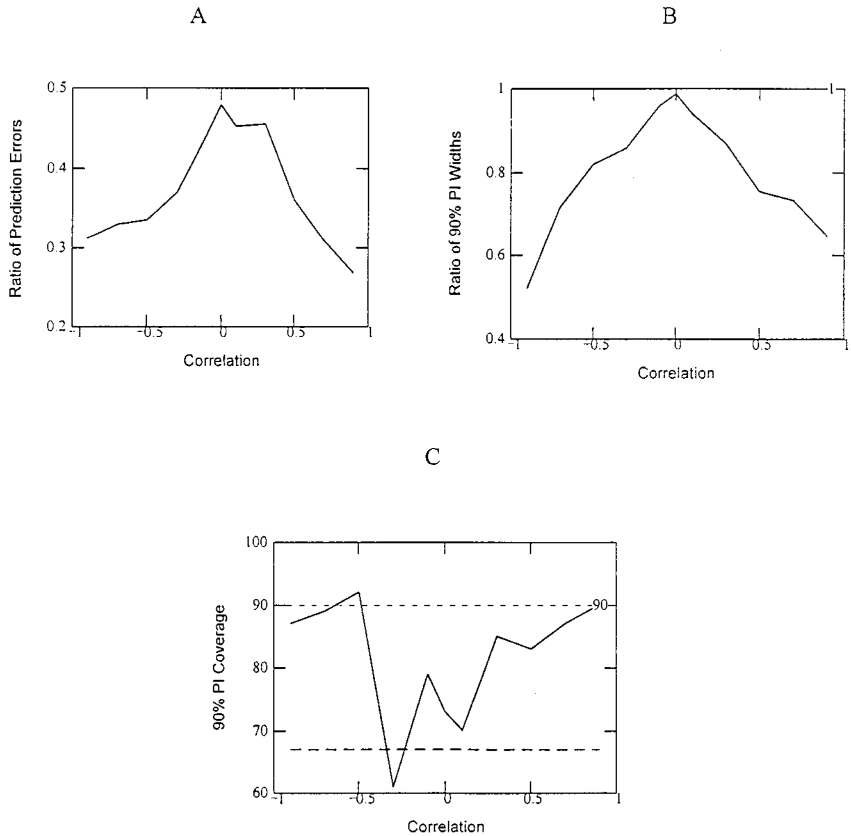

Figure 4. Simulation results from sample design for normal normal mixture: A, ratio of observed prediction error from sequentially designed samples using elicited data, to that from SRSWOR; B, 90% prediction interval width (a dotted line indicates the value 1, where sequential samples with elicited data and SRSWOR, are similar); C, 90% prediction interval coverage. Both in terms of prediction error and coverage, sequential designs are better than SRSWOR. Within sequential designs, prediction error for elicited data scenario gets smaller as correlation between real and expert data differs from zero in either direction.

better than the SRSWOR design. Even the worst sequential design had an observed prediction error that was 50% lower than the simple random samples. The plot further shows that use of an informative expert ($\eta \neq 0$) produces a better sample. Observed prediction errors are highest for a noninformative expert ($\eta = 0$), which is the same as ordinary sequential sampling with no elicited data, and fall off quite rapidly as η gets farther from 0 in either direction. Even a moderately informative expert ($\eta = \pm 0.5$) produces a sample that reduces prediction error by more than 30% over ordinary sequential sampling ($\eta = 0$). Figure 4C shows that prediction intervals with better coverage are produced by ordinary sequential samples ($\eta = 0$) as compared to simple random samples (73% vs. 67%). Moderately informative experts ($0 < |\eta| \leq 0.5$) improve this still further (78.33%), and coverage gets close to nominal as experts become more informative (88.25% for $0.5 < |\eta| \leq 1$). In Figure 4B we see that improved coverage does not come with less precision. In fact, prediction interval width is smallest for a sequential sample based on informative elicited data, and largest for noninformative elicited data. Prediction interval widths for ordinary sequential samples based solely on the observed data are very similar to those produced by simple random samples.

To summarize the discussion of this section, we note that incorporation of expert opinion, in terms of elicited data, improves estimation, prediction, and sampling designs. An extremely important feature of our approach, which addresses the concerns voiced by critics of the Bayesian paradigm (Royall, 1997; Mayo, 1996; Dennis, 1996; Efron, 1986) and needs to be emphasized strongly, is that elicited data from an intentionally misleading expert can be usefully incorporated into our analysis. This is in stark contrast with the effect of elicited prior from an intentionally misleading expert in a Bayesian setting. Unless one has large amounts of observed data (at which point Bayesian analysis is less influential anyway), such a misleading prior has large influence on the Bayesian inferences.

NORMAL DATA WITH BINARY EXPERT OPINION

In this section we consider the situation where, although real data are continuous and normally distributed, expert opinion is elicited in the form of more imprecise binary information. For example, in a pollution investigation where data consist of amounts of contamination at different locations, it may be much easier for an expert to provide binary values (high or low contamination) over the whole study area, rather than coming up with exact numbers.

We now introduce the model that provides a basis for combining actual continuous data and elicited binary data:

- Let Y_i be normally distributed with mean $\underline{x}_i\underline{\beta}$, variance σ^2 and corr(Y_i, Y_j) $= \gamma^{d(i,j)}$, $(i, j) = 1, 2, \ldots, N$, where γ is a measure of spatial dependence, *d(i,j)* is the distance between locations indexed *i* and *j*, $\underline{x}_i$'s are covariates for location *i* and $\underline{\beta}$, the vector of regression coefficients.
- Given Y_i, conditionally,

$$(E_i \mid Y_i) \sim \text{Bernoulli}\left(\frac{e^{\eta(y_i - c)}}{1 + e^{\eta(y_i - c)}}\right) \tag{12}$$

  independently, $i = 1, \ldots, N$, where the log odds ratio η characterizes the dependence between $\underline{Y}$ and $\underline{E}$ and the threshold c is a known constant.

Note that, as the elicited data here has a Bernoulli distribution, the honesty parameter η is the log odds, which are positive for a good expert who mimics reality, negative for a misleading expert, and close to 0, for a random expert.

Estimation and Prediction

We now investigate the effect of augmenting information by incorporating binary data elicited from experts, on estimation and prediction. For that, we first note that the modeling framework in (12) belongs to the class of generalized linear mixed model (GLMM) (Breslow and Clayton, 1993). Parameter estimates for this GLMM are obtained by the Monte Carlo Newton Raphson methods (McCulloch, 1997). Note that predictions must be based on the conditional distribution $(Y_i \mid \underline{Y}_s, \underline{E}; \underline{\beta}, \sigma^2, \eta)$, $i \notin s$, which is difficult to evaluate analytically. In fact, a Monte Carlo approach can avoid such tedious calculations. We use the Metropolis Algorithm (Hastings, 1970) to generate a batch of realizations of Y_i from its conditional probability distribution $(Y_i \mid \underline{Y}_s, \underline{E}; \underline{\beta}, \sigma^2, \eta)$, $i \notin s$, after plugging in the parameter estimates. If we denote these values by $(Y_i^{(1)}, Y_i^{(2)}, \ldots, Y_i^{(B)})$, then, for a sufficiently large B,

$$E(Y_i \mid \underline{Y}_s, \hat{\underline{\beta}}, \hat{\sigma}^2, \hat{\eta}) \approx \frac{1}{B} \sum_{j=1}^{B} Y_i^{(j)}$$

is our point predictor for Y_i. Note that in a similar situation in Section 5.1, we used the maximum posterior as a point predictor. However, here we use the mean of the posterior distribution to see whether prediction performance is sensitive to the choice of a particular predictor. Nominal 90% prediction intervals for Y_i are approximated by the 5th and 95th percentile values of the empirical distribution function $(Y_i^{(1)}, Y_i^{(2)}, \ldots, Y_i^{(B)})$. Details are provided in Das (1998).

Simulation results presented in Figures 5A to 5D show that the introduction of spatial dependence does not appreciably change estimation and prediction properties. There is significant benefit from the incorporation of elicited data (both good and misleading), and the choice of the posterior mean instead of the posterior mode has no effect on the performance of the point predictor. (Note that to facilitate comparison with previous plots, the transformation $(e^n - 1)/(e^n + 1)$ has been applied to the η's in this and subsequent plots to restrict their values between -1 and 1.) A couple of points are worth mentioning. First, we note that for $\eta \approx 0$, the gain in estimation from using elicited data is somewhat greater here compared to the normal normal hierarchy (Sections 3 and 4). This is because, unlike the normal hierarchy, here η and β are both identifiable, and can be separately estimated based on $\underline{E}$ alone. Second,

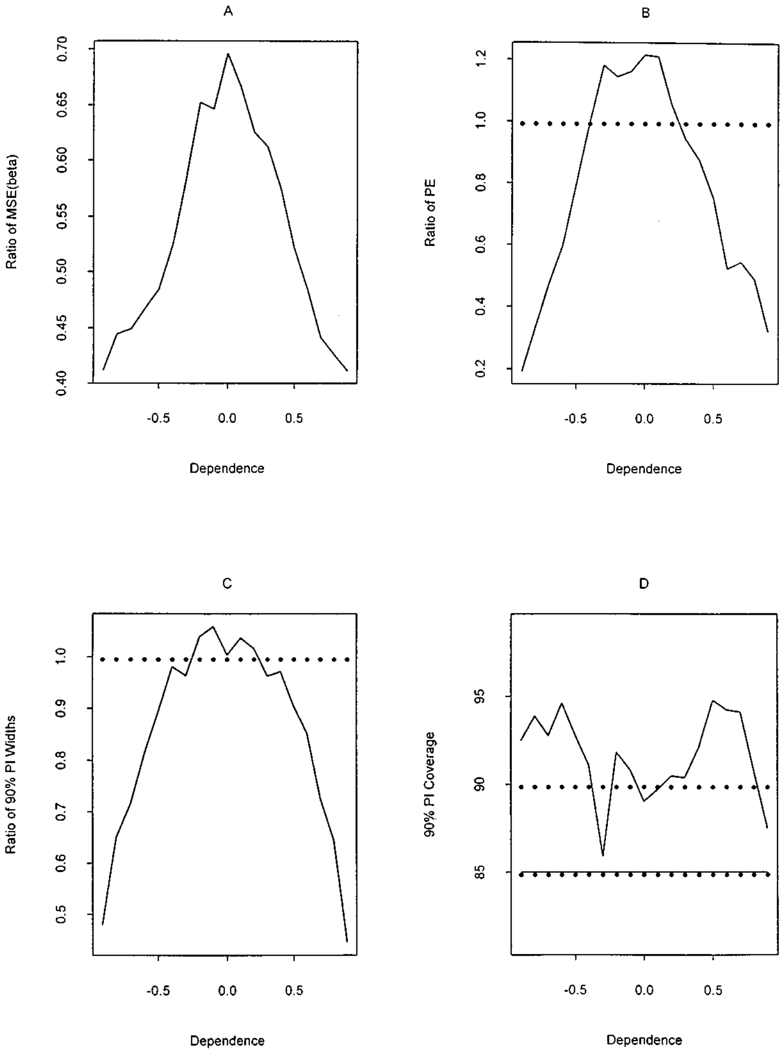

Figure 5. Simulation results for normal binary mixture, dependent case: A, ratio of observed MSE($\hat{\beta}$) with elicited data, to MSE($\tilde{\beta}$), calculated without elicited data; B, observed prediction error; C, 90% prediction interval width (a dotted line indicates the value 1, where the two scenarios would have similar results); D, 90% prediction interval coverage, solid line for elicited data and dashed line for no elicited data. To facilitate comparison with previous plots, the transformation $(e^{\eta} - 1)/(e^{\eta} + 1)$ has been applied to the η's in this and subsequent plots to restrict their values between -1 and 1. Estimation and prediction error for elicited data scenario gets smaller as correlation between real and expert data differs from zero in either direction. Observe that elicited data from an intentionally misleading expert is also useful.

results show that though elicited data greatly improves predictions, as long as the expert is informative (i.e., $|\eta|$ is high), for $\eta \in (-0.25, 0.25)$, point predictions are slightly worse than those based solely on the observed data. However, as before, coverage of 90% prediction intervals (Fig. 5D) for elicited data (mean of 89%) is closer to the nominal value than those based on observed data alone (mean of 85%). Moreover, prediction intervals are narrow for informative experts and similar to those based on observed data for random experts (Fig. 5C). Thus, the loss in precision for point predictions when $\eta \in (-0.25, 0.25)$ is compensated by improved prediction interval coverage. These results indicate that using binary expert opinion in addition to hard data improves estimates, as well as point and interval predictions even when it is in the relatively imprecise binary form, whereas the actual observations are continuous and spatially correlated.

Sampling Design

We now do a simulation study of sequential sampling designs based on expert opinion in the form of binary data. In this case, however, notice that $(Y_i \mid \underline{E}, \underline{Y}_s)$ is not a Gaussian random variable. So here, instead of including that location i in the sample, for which $\mathrm{var}(Y_i \mid \underline{E}, \underline{Y}_s)$ is the largest, we substitute $\mathrm{var}(Y_i \mid \underline{E}, \underline{Y}_s)$ by the width of the prediction interval for the ith location in the sequential sampling scheme described in the previous section.

Note that here η and β are both identifiable, and can be separately estimated based on $\underline{E}$ alone. To select the first sample, we first estimate the model parameters based on $\underline{E}$ alone. Then, we generate a batch of realizations of $Y_i, i = 1, \ldots, N$, from the conditional distribution $(Y_i \mid \underline{E})$ and get the Monte Carlo prediction intervals from this empirical distribution function, selecting that location for which width of the prediction interval is the largest. Selection of rest of the sample locations follows a similar procedure, except that once we have an initial sample s, we generate $Y_i, i \notin s$, from the conditional distribution $(\underline{Y}_i \mid \underline{Y}_s, \underline{E})$. Further details appear in Das (1998).

We present results of the simulations graphically in Figures 6A to 6C. Similar to the normal–normal mixture in Section 4.2, we see that sequential designs perform much better than ordinary SRSWOR. Moreover, the use of informative expert opinion produces a better sample; even for a moderately informative expert (transformed $\eta = \pm 0.5$) prediction error is reduced by more than 54%, with better coverage as compared to ordinary sequential design corresponding to $\eta = 0$ (92% vs. 84.7% for a 90% nominal level).

Results discussed in this section show that augmentation of observed data with even expert opinion, even in the imprecise form of binary data, can improve substantially both estimation and prediction, as well as construct sequential sample designs that perform better than ordinary sequential designs.

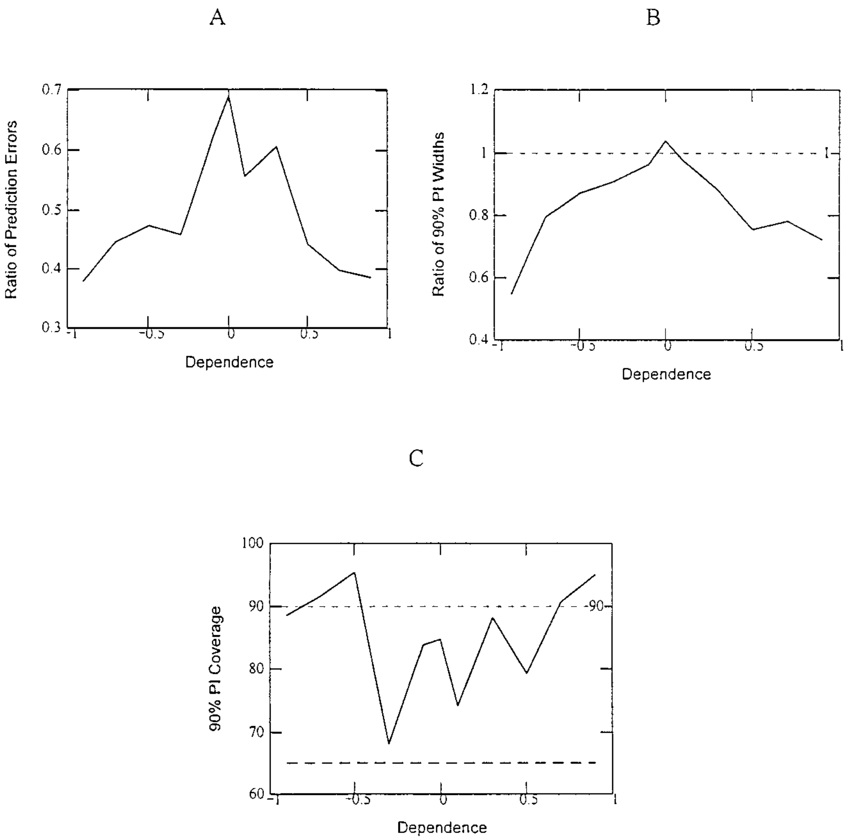

Figure 6. Simulation results for sample design for normal binary mixture: A, ratio of observed prediction error from sequentially designed samples using elicited data, to that from SRSWOR; B, 90% prediction interval width; (a dotted line indicates the value 1, where sequential samples with elicited data and SRSWOR are similar); C, 90% prediction interval coverage. Both in terms of prediction error and coverage, sequential designs are better than SRSWOR. Within sequential designs, prediction error for elicited data scenario gets smaller as correlation between real and expert data differs from zero in either direction.

DISCUSSION

The availability of expert opinion in most environmental studies is a potentially useful resource that can augment the observed data, which frequently is sparse. To incorporate expert opinion, we present an approach that is based on "elicited data." There are some key advantages to this approach:

(a) Scientists generally think in terms of the data generation process. Hence, eliciting data is, in our experience, more natural than eliciting a prior distribution on the parameters of a statistical model, which is purely a modeler's construct.

(b) Both intentionally misleading expert and a genuinely honest expert provide information about the underlying process. Elicited prior from an intentionally misleading expert can badly influence inferences in the Bayesian paradigm, whereas elicited data from such an expert proves equally useful in our approach.

A crucial issue, of course, is, Is it possible to elicit data in practice? We refer the reader to Journel (1986) and Kulkarni (1984), two geologists, not statisticians, for the practicality of such a proposal. We ourselves have been able to elicit data from the experts regarding presence/absence of mammal species in Montana. A detailed ecological analysis of these data is to be provided elsewhere.

When elicited data are obtained, we have illustrated, both theoretically and using simulations, that incorporating expert opinion via elicited data substantially improves estimation, prediction, and design aspects of statistical inference for spatial data. Incorporation of model selection procedures should reduce the sensitivity of this approach to model misspecification. More research is needed in this direction.

The modeling approach presented here for combining observed data with expert opinion is very general. We are not restricted to the particular scenarios considered in this paper. For instance, in many ecological studies the question centers on presence/absence of a species and its relation to habitat characteristics. Responses are then binary instead of continuous. Methods presented here should also be useful for this type of situation.

ACKNOWLEDGMENTS

This work was part of Abhik Das's Ph.D. dissertation and was partially supported by the U.S. Environmental Protection Agency's Climate Policy and Assessment Division, cooperative agreement CR823143, to Jonathan A. Patz, Principal Investigator. The authors wish to thank the editor and a referee for helpful suggestions.

REFERENCES

Benedetti, R., and Palma, D., 1995, Optimal sampling designs for dependent spatial units: Environmetrics, no. 6, p. 101–114.

Breslow, N. E., and Clayton, D. G., 1993, Approximate inference in generalized linear mixed models: Journal of the American Statistical Association, v. 88, no. 421, p. 9–25.

Christakos, G., and Killam, B. R., 1993, Sampling design for classifying contaminant level using annealing search algorithms: Water Resources Res., v. 29, no. 12, p. 4063–4076.

Cressie, N., 1988, Spatial prediction and ordinary kriging: Math. Geology, v. 20, no. 4, p. 405–421.

Cressie, N., 1993, Statistics for spatial data: John Wiley & Sons, New York, 900 p.

Cressie, N., Gotway, C. A., and Grondona, M., 1990, Spatial prediction from networks: Chemometrics and Intelligent Laboratory Systems, no. 7, p. 251–271.

Das, A., 1998, Topics in spatial statistics: Unpubl. doctoral dissertation, Johns Hopkins University, 157 p.

Dennis, B., 1996, Discussion: Should ecologists become Bayesians?: Ecological Applications, v. 6, no. 4, p. 1095–1103.

DeGroot, M. H., 1988, A Bayesian view of assessing uncertainty and comparing expert opinion: Journal of Statistical Planning and Inference, no. 20, p. 295–306.

Efron, B., 1986, Why isn't everyone a Bayesian? (with discussion): The American Statistician, v. 40, no. 1, p. 1–11.

Efron, B., 1996, Empirical Bayes methods for combining likelihoods (with discussion): Journal of the American Statistical Association, v. 91, no. 434, p. 538–565.

Garthwaite, P. H., and Dickey, J. M., 1992, Elicitation of prior distributions for variable selection problems in regression: Annals of Statistics, no. 20, p. 1697–1719.

Genest, C., and Zidek, J. V., 1986, Combining probability distributions: A critique and an annotated bibliography: Statistical Science, no. 1, p. 114–135.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, no. 57, p. 97–109.

Journel, A. G., 1986, Constrained interpolation and qualitative information—The soft kriging approach: Math. Geology, v. 18, no. 3, p. 269–286.

Kulkarni, R. B., 1984, Bayesian kriging in geotechnical problems, *in* Verly, G., ed., Geostatistics for Natural Resources Characterization Part 2: D. Reidel Publishing Company, Boston.

Lee, Y. M., and Ellis, J. H., 1996, Groundwater monitoring network design: Algorithms for nonlinear integer optimization: ASCE Journal of Environmental Engineering, v. 122, no. 6, p. 524–531.

Mayo, D. G., 1996, Error and growth of experimental knowledge: University of Chicago Press, Chicago, 494 p.

McCulloch, C. E., 1997, Maximum likelihood algorithms for generalized linear mixed models: Journal of the American Statistical Association, v. 92, no. 437, p. 162–170.

Royall, R. M., 1997, Statistical evidence: A likelihood paradigm: Chapman & Hall, London, 191 p.

Steffey, D., 1992, Hierarchical Bayesian modeling with elicited prior information: Communications in Statistics A, no. 21, p. 799–821.

von Mises, R., 1943, On the correct use of Bayes' formula: Annals of Mathematical Statistics, no. 14, p. 156–165.

West, M., 1988, Modelling expert opinion, *in* Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M., eds., Bayesian Statistics 3: Oxford University Press, New York, p. 493–508.