ENVIRONMETRICS
Environmetrics 2006; **17**: 517–525
Published online 22 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/env.789

Estimating soil properties in heterogeneous land-use patches: a Bayesian approach

Jacob J. Oleson[1]*,†, Diane Hope[2], Corinna Gries[2] and Jason Kaye[3]

[1]*Department of Biostatistics, The University of Iowa, Iowa City, IA, U.S.A.*
[2]*Global Institute of Sustainability, Arizona State University, Tempe, AZ, U.S.A.*
[3]*Department of Crop and Soil Sciences, Penn State University, University Park, PA, U.S.A.*

SUMMARY

Cities provide unique opportunities for integrating humans into ecology. Using data from a socio-ecological inventory of metropolitan Phoenix, Arizona, we explore the contribution of human-related variables to explaining observed variation in soil nitrate-N (NO_3–N) and total carbon (C) concentrations across the city, agricultural fields, surrounding desert, and mixed regions. Conventional modeling approaches in such a setting would lead to examination of spatial relationships over the entire study area or on subsets of the data independently. However, the spatial relationships for NO_3–N and C may be different in each of these regions. Here we estimate the correlation coefficients for influential variables toward soil NO_3–N and C across the entire region, while at the same time accounting for potentially differing spatial patterns in each of these regions. Soil NO_3–N shows markedly greater spatial autocorrelation in the desert regions, while the soil C shows varying amounts of spatial relationships in the different regions. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: hierarchical Bayes; MCMC; regression; spatial correlation; LTER

1. INTRODUCTION

The Central Arizona-Phoenix Long Term Ecological Research (CAP-LTER) program was established specifically to understand the functioning of an urban ecosystem (Kaiser, 2001). The CAP study region is a spatially heterogeneous mixture of various urban (e.g. residential, commercial, industrial) land uses, agricultural, and undeveloped Sonoran desert situated on a large alluvial plain with remnant mountains providing the topographic template upon which the city has developed.

In order to best quantify key ecological characteristics of such a large, complex study site, a randomized tessellation stratified design was employed, by superimposing a grid subdivided into $4 \times 4\,\text{km}^2$ squares on the study area. This scheme produced 462 sampling units across the entire area,

*Correspondence to: J. J. Oleson, Department of Biostatistics, The University of Iowa, Iowa City, IA, 52242-1009 U.S.A.
†E-mail: jacob-oleson@uiowa.edu

Contract/grant sponsor: National Science Foundation; contract/grant number: DEB-0423704.
Contract/grant sponsor: Central Arizona—Phoenix Long Term Ecological Research (CAPLTER).

Received 28 June 2005
Accepted 21 January 2006

over half of which fell outside the current urban core area. A random sample was assigned to every square inside the urban core and every third square outside of the urban core ($n = 206$). Access was denied to two of these sites and at ten more sites there are missing data for our variables of interest leaving $n = 194$.

These sites can be broken into five land use categories: urban ($n_{\text{urban}} = 89$), desert ($n_{\text{desert}} = 66$), agriculture ($n_{\text{agr}} = 22$), near roads and highways ($n_{\text{highway}} = 11$), and a mixed class ($n_{\text{mixed}} = 6$) indicating that more than one of the land use types were found in a survey plot. There are small sample sizes for near roads and mixed, thus we combine these two categories to produce a new mixed category with 17 sites ($n_{\text{mixed}} = 17$) (Figure 1). Soils were sampled using a hand-coring device at four random locations within a $30 \times 30\,\text{m}^2$ field plot and the four samples combined to give one sample for each per site as detailed in Hope *et al*. (2005).

We are interested in concentrations of nitrate-N (NO_3–N) and total carbon (C) in soils in the top 10 cm of the soil profile measured in mg/kg dry weight of soil. Nitrate is the dominant form of available inorganic N in these soils, which, along with water, co-limits plant growth in southwestern desert ecosystems (Schlesinger *et al*., 1996) and therefore represents an important 'bottom-up' control of ecosystem structure and dynamics.

Hope *et al*. (2005) reported results for NO_3–N and NH_4–N data previously. In their study, each of the response variables (NO_3–N and NH_4–N) were modeled using 12 predictor variables chosen to represent the main geophysical, geographic, and human characteristics of the study site. The original

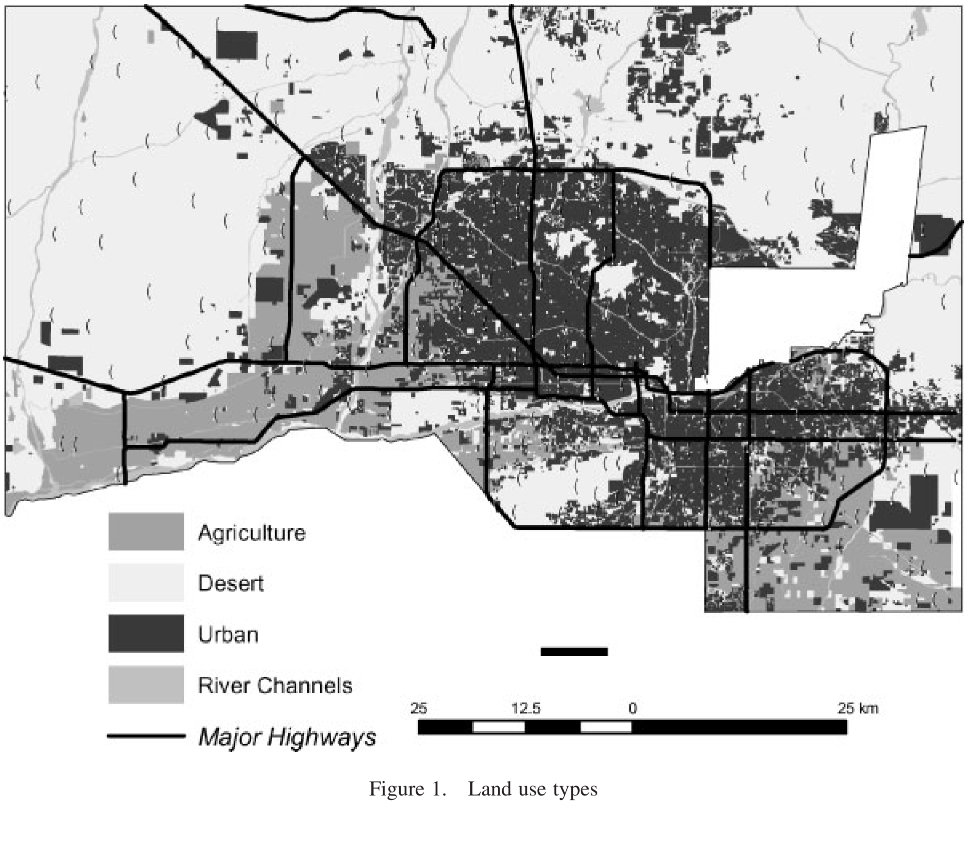

Figure 1. Land use types

predictor variables were site latitude, site longitude, number of years since last land use change, an indicator variable for if the land had ever been used in agriculture, elevation, distance from urban center, slope of the land, per cent impervious surface on the plot, per cent lawn cover on the plot, an indicator variable showing type of irrigation, median family income, and population density of the area. They examined the response variables on the site as a whole (all 194 sites) as well as for desert ($n = 66$) and urban ($n = 89$) separately. Backwards elimination was primarily used to determine influential variables. The results of their analysis are presented in Section 3 for NO_3–N. We have used the same technique for C in Section 4 for a comparison standard. The residuals were examined for spatial autocorrelation between the sites. If spatial correlation was found, this relationship was included in the error structure. If no spatial correlation was found, a standard multiple regression analysis was performed.

Hope *et al*. (2005) describe that spatial correlations were not found between sites for any of the response variables when only the urban sites were explored. This likely results from a wide variety of land uses and influences on soil chemistry across the Phoenix metropolitan area. Spatial correlations were found for NO_3–N and C in the desert sites while small sample sizes restrain agricultural, transportation, and mixed sites from being studied separately. Due to the marked differences between urban and desert sites, it is not surprising that there was no significant spatial correlation when examining the entire region using only a single correlation structure. However, we know there are significant spatial correlations in some subareas of the region (i.e., in the desert sites). Therefore, we sought to improve upon simply pooling the data together and checking for spatial correlation among all the sites. We do this by proposing a model that estimates the influence of the covariates over the region as a whole, while allowing for separate spatial covariance structures in the different regions. With this approach, we are able to use the distinct information from each of the regions, while retaining the factors that influence the region as a whole. We will re-analyze the NO_3–N of Hope *et al*. (2005) and compare our results to those in the aforementioned paper. We will perform a similar analysis for C and execute a similar comparison.

We will use an urban region, a desert region, an agriculture region, and the class we call mixed that combines the transportation and the original mixed class. This type of model will require a relatively large number of parameters. A Bayesian hierarchical model offers many advantages for this structure. It accounts for the variability of all parameters in the model and provides a coherent framework in which to incorporate scientific reasoning and experience explicitly in the model. For example, we believe that there is more spatial correlation among the desert sites than for the urban sites. This was true for NO_3–N and we expect to see similar relationships in C. The Bayesian approach to spatial statistics has seen a lot of recent additions to the literature. For a very good overview of such techniques see Banerjee *et al*. (2004) and Wikle and Royle (2004).

In Section 2, we introduce the Bayesian model for our analysis. We then use this model in Section 3 for the NO_3–N data. A similar model is used in Section 4 for C. We make concluding remarks in Section 5.

2. MODELING SPATIAL CORRELATION FROM DISTINCT AREAS

We want a model to estimate the regression coefficients while allowing for differing spatial correlations in distinct regions. To allow for this we set up a Bayesian hierarchical model. Using the flexibility of a hierarchical Bayesian model, we can incorporate these differing spatial structures into one model.

We first model the data or likelihood. Let Y_s be the response of interest at each spatial location s for $s = 1, \ldots, n$ where $\mathbf{Y} = (Y_1, \ldots, Y_n)'$. We assume this response variable, Y_s, is normally distributed with mean μ_s and variance δ_s. Let $\mu = (\mu_1, \ldots, \mu_n)'$ and let $\mu = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}$ such that

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}, \Delta)$$

with Δ being a diagonal matrix where the first n_1 elements are δ_1, the next n_2 elements being δ_2, the next n_3 elements consist of δ_3 and the remaining n_4 elements being δ_4. $\mathbf{X}$ is our $n \times (k+1)$ design matrix that will contain the covariates of interest and a vector of 1s for the intercept. $\boldsymbol{\beta}$ is the $(k+1) \times 1$ vector of regression coefficients. $\mathbf{Z}$ is the $n \times 1$ random component that allows for spatial relationships between sites.

In introducing additional notation, we also write that $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \mathbf{Y}_3', \mathbf{Y}_4')'$ breaks $\mathbf{Y}$ into its' four regions and are $n_j \times 1$ for $j = 1, \ldots, 4$. $\mathbf{Z} = (\mathbf{Z}_1', \mathbf{Z}_2', \mathbf{Z}_3', \mathbf{Z}_4')'$ breaks $\mathbf{Z}$ into its' four regions and are $n_j \times 1$ for $j = 1, \ldots, 4$. $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \mathbf{X}_3', \mathbf{X}_4')'$ breaks $\mathbf{X}$ into its' four regions where $\mathbf{X}_j$ is $n_j \times (k+1)$.

In the next stage of the hierarchical model, we assign priors to these processes. We have selected the explanatory variables in $\mathbf{X}$ that we believe will have a significant impact on $\mathbf{Y}$. The specific explanatory variables for NO_3–N and C will be discussed in their respective sections later in the article. We will assume no additional information toward these variables and assign $\boldsymbol{\beta}$ the non-informative prior $\beta_i = 1$ for $i = 0, \ldots, k$.

Next we model the random component $\mathbf{Z} = (Z_1, \ldots, Z_n)'$. In $\mathbf{Z}$ we allow for spatial correlations between sites. We break $\mathbf{Z}$ into four regions, $\mathbf{Z} = (\mathbf{Z}_1', \mathbf{Z}_2', \mathbf{Z}_3', \mathbf{Z}_4')'$, to specify that there will be a different spatial relationship in each of the regions. For each region, let Z_{s_j} for $s_j = 1, \ldots, n_j$ and $j = 1, \ldots, 4$ each follow a normal distribution with a mean of zero and variance σ_j^2. The correlation between sites is $\mathrm{Corr}(Z(s_j), Z(s_j')) = k_{\theta_j}(\|s_j - s_j'\|)$. The correlation function depends on parameter θ_j and the distance between locations s_j and s_j'. We will assume that sites from differing sections are not related over space. Thus, $k_{\theta_j}(\|s_j - s_j'\|) = 0$ for $j \neq k$ with $j, k = 1, \ldots, 4$. The soil content in the urban sites has changed drastically over time and we do not foresee any relationship between these sites and the desert, agricultural, or mixed sites. Similarly, the soil content in the agricultural sites is assumed to have changed enough from the desert sites, that we do not feel we need to include a relationship. For the remaining combinations between mixed and agriculture as well as mixed and desert, this assumption is based on modeling convenience.

A common correlation model is the single parameter exponential function given by

$$k_{\theta_j}(\|s - s'\|) = e^{-\|s-s'\|/\theta_j}$$

Many other correlation functions can be considered (see Cressie, 1993). For our purposes, this one parameter model works well. In addition, estimates of the spatial parameters are not the main goal of this analysis. We are using this to account for the variability and to better estimate the $\boldsymbol{\beta}$ parameters. Under this model, the correlation between two points decays exponentially as the distance between those points increases. The rate of decay is controlled by θ_j. We will assume this fits each of the four regions as the rates of decay are denoted by θ_j for $j = 1, \ldots, 4$.

Now that we have modeled the mean structure of $\mathbf{Y}$, we turn to the variance. The common conjugate prior for the variance of a normal distribution is the inverse gamma distribution (Gelman *et al.*, 1995). We choose a proper, but relatively flat prior and let $\delta_j \sim \mathrm{IG}(2.01, 1.01)$ which has a mean of 1 and a variance of 100 for $j = 1, \ldots, 4$.

Finally, we set the hyperprior distributions. The σ_j^2 with $j = 1, \ldots, 4$ are the variances of normal distributions and we assume each of them to be $\text{IG}(2.01, 1.01)$ as well. For the spatial dependence parameter θ_j for $j = 1, \ldots, 4$ Berger *et al*. (2001) suggest a reference prior since many common improper priors yield an improper posterior distribution. We have some prior knowledge in this case, thus we define a proper prior on θ_j. We let θ_j follow a gamma distribution with parameters specified in the analysis sections.

Our goals under this model are to obtain estimates of the regression coefficient $\boldsymbol{\beta}$ and determine the amount of spatial autocorrelation within each of the four regions. To obtain the posterior densities, we make use of Markov chain Monte Carlo (MCMC) techniques. In particular we use Gibbs sampling (Gelfand and Smith, 1990) and the Metropolis–Hastings (M–H) algorithm (Chib and Greenberg, 1995). See Gelman *et al*. (1995), Gilks *et al*. (1996), and Robert and Casella (1999) for a full description of MCMC techniques. For these algorithms, we require the full conditional distributions of the parameters. If this distribution is of a known form, then they are easily sampled from directly via the Gibbs sampler. If, however, they are not of a known form, we rely on the random walk M–H algorithm. The full conditional distributions for this model can be found in the Appendix. All programming and plotting was performed in R and is available from the first author upon request.

3. SOIL NITRATE-N (NO_3–N) RESULTS

Hope *et al*. (2005) used a log transformation on NO_3–N concentration data to ensure constant variance. We do the same in this model. Their regression analysis for the entire study region found: (1) latitude ($p = 0.014$); (2) an indicator variable for whether the site had ever been used in agriculture ($p = 0.001$); (3) the population density ($p = 0.000$); (4) the per cent of impervious surface ($p = 0.009$); and (5) per cent of lawn cover ($p = 0.029$) to be significant at the $\alpha = 0.05$ level using a multiple regression model for soil nitrate. No spatial autocorrelation was found in the errors. We chose our predictor variables to coincide with this analysis.

Similarly, a regression analysis was performed when modeling soil nitrate on only the urban sites. Again, there was no spatial autocorrelation in the errors. A third regression model was used when modeling soil nitrate on only the desert sites. Here spatial autocorrelation was found among the desert sites to approximately 17 000 m. Even though they found no spatial correlation between sites for the entire study region, there was correlation when confined to a particular subarea. Thus, a model such as that proposed in Section 2 is appropriate for this data set. We would like to find the significant regression coefficients for the site as a whole, while allowing subareas to have differing spatial relationships.

In our Bayesian analysis of this data, a Bayesian model selection technique should be instituted. Bayes factor is one common model selection technique to capture the change in the odds in favor of one model versus another. Kass and Wasserman (1995) state that the Bayes information criterion (BIC) is a good approximation of -2 log Bayes factor. Using the same 12 initial predictor variables, we estimated BIC for all possible subsets. The model with the five predictors mentioned above had one of the four smallest BIC values with the BIC values being indistinguishable. We will thus proceed with the same five predictor variables using the modeling in Section 2 and compare results from the two approaches.

The Markov chain was run for 20 000 iterations at which point convergence was judged to have occurred. This convergence happened within 1 000 cycles, we ran the 'burn-in' to 20 000 to be conservative. This was repeated with five different starting values, and the results were impacted

Table 1. Regression parameter estimates for NO_3–N

Variable	Regression estimates	Model estimates	95% credible limits
Intercept	70.77	50.795 (26.119)	(0.055, 102.374)
Latitude	−2.06	−1.459 (0.777)	(−2.998, 0.050)
Ever in ag	0.91	0.839 (0.321)	(0.218, 1.412)
Population density	4.00 E(−4)	3.84 E(−4) (1.08 E(−5))	(1.71 E(−4), 5.94 E(−4))
% Impervious	−0.011	−0.013 (0.005)	(−0.023, −0.003)
Cover lawn	−0.017	−0.016 (0.009)	(−0.033, 0.001)

minutely. We then ran the chain for 20 000 more iterations. Estimates of parameters are based on all 20 000 converged samples.

Viewing the variogram of the residuals from Hope *et al*. (2005), we define proper priors to use for the θ parameters. We are not concerned with the estimates of these parameters, as $\boldsymbol{\beta}$ is the primary concern. We use this parameter to account for the additional variability in the model. The starting values for θ_{urban}, θ_{agr}, θ_{mixed} was 10 000. The prior distribution was Gamma (2.7778, 0.0005555), which yields a mean of 5 000 and standard deviation of 3000. The starting value for θ_{desert} was 17 000 with a prior distribution of Gamma (32.1111, 0.0018889) which yields a mean of 17 000 and standard deviation of 3000.

Since our primary interest is in the regression coefficients, we construct 95 per cent credible intervals for the estimates $\hat{\boldsymbol{\beta}}$. The estimates and corresponding standard deviations of $\boldsymbol{\beta}$ are shown in Table 1 along with the coefficients from the simple linear regression model. The 95 per cent credible intervals for latitude and per cent lawn cover now include zero. This is likely due to including the extra information by using the four different regions. Spatial relationships are now included in the model and thus latitude would likely add very little above this. The desert has very little lawn cover, agricultural sites would have a substantial amount of 'lawn' (i.e., grass or other sort green vegetation) cover and the urban core would be mixed. Incorporating the heterogeneous variation, this additional information again would add relatively little to the model. While the remaining variables ever used in agriculture, population density, and percent impervious surface are still all significant, the values of each of the coefficients has changed slightly. We note that latitude and lawn cover had the highest *p*-values from the previous analysis.

Our model incorporates additional information by accounting for additional spatial variables. The simple linear regression model does not account for spatial correlation in the errors and has a variance of 2.26. We do not expect to see a spatial relationship among the urban sites. The posterior estimate of the variance parameter (and standard deviations) from the urban sites is $\hat{\delta}_{\text{urban}} = 1.76$ (0.69). We did expect a relationship among the desert sites, which would likely lower the model error for the desert sites. The posterior estimate is $\hat{\delta}_{\text{desert}} = 1.03$ (0.23). The sample sizes for the remaining areas are quite small and produce large estimates for the variance parameters at $\hat{\delta}_{\text{agr}} = 1.04$ (0.56) and $\hat{\delta}_{\text{mixed}} = 2.02$ (1.45).

We use four regions to account for differing levels of spatial variation and obtain posterior means and standard deviations $\hat{\sigma}^2_{\text{urban}} = 0.94$ (0.66), $\hat{\sigma}^2_{\text{desert}} = 0.48$ (0.25), $\hat{\sigma}^2_{\text{agr}} = 0.90$ (0.58), and $\hat{\sigma}^2_{\text{mixed}} = 1.86$ (1.57). We also obtained estimates of the posterior means and standard deviations $\hat{\theta}_{\text{urban}} = 3731$ (2624), $\hat{\theta}_{\text{desert}} = 17265$ (3101), $\hat{\theta}_{\text{agr}} = 4616$ (2801), and $\hat{\theta}_{\text{mixed}} = 6362$ (3766). We see here that there are clearly differences between the land use types. In particular, the range of spatial relationship for NO_3–N is much larger in the desert than in the other land use regions.

4. TOTAL CARBON (C) RESULTS

Initial examinations at the empirical semi-variogram suggest slight spatial correlation when examining all the sites. There was no spatial auto-correlation when using only the urban sites. There was spatial autocorrelation when viewing only the desert sites. Even though there was spatial autocorrelation using all the sites, we will benefit by breaking this into four regions of differing spatial correlations.

The frequentist approach used for soil nitrate-N was employed to incorporate one spatial correlation error structure for the entire study region. This analysis found: (1) if the land was ever used for agriculture ($p = 0.000$); (2) distance from the urban center ($p = 0.000$); (3) amount of vegetation cover ($p = 0.002$); (4) per cent clay in the soil ($p = 0.028$); (5) if irrigation other than drip or flood is used ($p = 0.001$) to be significant at the $\alpha = 0.05$ level. The estimate of the sill was 0.029, the nugget effect = 0.004, and the range = 2765.

Using the same 12 initial predictor variables, we estimated BIC for all possible subsets. The model with the five predictors mentioned above had one of the three smallest BIC values, all of which were indistinguishable. We will thus proceed with these same predictors using the modeling in Section 2 and compare the results of the two approaches.

In our analysis, the Markov chain was run for 100 000 iterations at which point convergence was judged to have occurred. This convergence happened within 30 000 cycles, we again ran the ‘burn-in’ to 70 000 to be conservative. This was repeated with five different starting values, and the results were impacted minutely. We then ran the chain for 40 000 more iterations. Estimates of parameters are based on all 40 000 converged samples. Convergence was much slower for C than it was for NO_3–N. This is likely due to relatively small $\hat{\theta}$ values. In particular, the estimates of the standard deviation for $\hat{\theta}_{\text{agr}}$ and $\hat{\theta}_{\text{mixed}}$ are larger than the means. Also, the standard deviation estimates for $\hat{\theta}_{\text{urban}}$ and $\hat{\theta}_{\text{desert}}$ are large relative to their means.

The starting value for θ_{urban} was 1000 with a prior distribution of Gamma (0.25, 0.00025) which yields a mean of 1000 and a standard deviation of 2000. The starting value for θ_{desert} was 3000 with a prior distribution of Gamma (2.25, 0.00075) which yields a mean of 3000 and standard deviation of 2000. The starting values for θ_{agr} and θ_{mixed} was 1000. The prior distribution was Gamma (0.1111, 0.0001111), which yields a mean of 1000, and standard deviation of 3000.

The coefficients from the mixed model, estimates and corresponding standard deviations of $\boldsymbol{\beta}$ are shown in Table 2. The 95 per cent credible intervals for the coefficients with distance from urban center, vegetation cover, percent clay, and other irrigation all contain zero (2). By adding the additional sources of spatial correlation, we have seemingly accounted for the effects of these variables. Distance from urban center is a distance that should be accounted for in the spatial distances. Vegetation cover

Table 2. Regression parameter estimates for C

Variable	Regression estimates	Model estimates	95% credible limits
Intercept	0.875	0.852 (0.142)	(0.573, 1.129)
Ever in ag	0.149	0.165 (0.074)	(0.019, 0.310)
Distance from urban center	−3.25 E(−6)	−3.22 E(−6) (2.68 E(−6))	(−8.42 E(−06), 2.08 E(−06))
Vegetation cover	0.002	0.002 (0.002)	(−0.001, 0.005)
% Clay	0.005	0.004 (0.005)	(−0.005, 0.013)
Other irrigation	0.105	0.074 (0.087)	(−0.096, 0.243)

differs in each of the four specified regions. Allowing the variance to be different in each of the regions diminishes its' contribution. This is similar for the per cent clay variable. Additionally, there is no irrigation in the desert and only one irrigation type (flood) is not used in agricultural regions. In other words, the effect of these variables are due to marked differences between sub-regions themselves, rather than due to continuous variation in those variables across the site as a whole.

The posterior estimate of the variance parameter (and standard deviations) from the urban sites is $\hat{\delta}_{\text{urban}} = 0.066$ (0.013). The posterior estimate is $\hat{\delta}_{\text{desert}} = 0.071$ (0.016). The sample sizes for the remaining areas are quite small and produce large estimates for the variance parameters at $\hat{\delta}_{\text{agr}} = 0.160$ (0.057) and $\hat{\delta}_{\text{mixed}} = 0.198$ (0.081).

We use four regions to account for differing levels of spatial variation and obtain posterior means and standard deviations $\hat{\sigma}^2_{\text{urban}} = 0.095$ (0.028), $\hat{\sigma}^2_{\text{desert}} = 0.080$ (0.020), $\hat{\sigma}^2_{\text{agr}} = 0.165$ (0.061), and $\hat{\sigma}^2_{\text{mixed}} = 0.199$ (0.082).

We also obtained estimates of the posterior means and standard deviations $\hat{\theta}_{\text{urban}} = 16782$ (9854), $\hat{\theta}_{\text{desert}} = 6030$ (3674), $\hat{\theta}_{\text{agr}} = 3189$ (5292), and $\hat{\theta}_{\text{mixed}} = 1583$ (2706). The relative imprecision is likely due to the small sample sizes for a spatial statistical problem.

5. CONCLUSIONS

This modeling framework has allowed us to better account for the variability in the data. In particular, NO_3–N and total C have differing degrees of spatial relationships as well as differing variances in the subregions. By accounting for this, we give a more accurate portrayal of factors influencing these NO_3–N and C. In this work, we find the variables that effect the soil content as a whole while we maintain the importance of the individual subregions.

In addition, these analyses broadly confirm the results of previous results soil NO_3–N concentrations across the CAP region (Hope *et al*., 2005), as well as those for soil C. However, use of the Bayesian approach detailed here allows a more detailed and accurate interpretation of the data in several ways. First, the Bayesian models better fit/explain the variance in the data set, because they better represent the subregional differences seen across CAP. Second, the Bayesian results indicate where variables are important because they differ between land use types (i.e., defineable subregions) rather than as a result of region-wide continuous variation in the data.

This survey is to be conducted again in 2005. There is still much that could be done when examining the soil contents. Model selection is an important factor for this study. A Bayesian model selection technique that works in the MCMC chain can be incorporated into a future analysis that takes the spatial correlation structure into account. In addition, the current model assumes independent processes for the individual land types. A new model could include multivariate relationships between them.

Changes in these soil contents will be analyzed over time. The idea of the CAP-LTER study in general is to see how these factors are changing over time and thus a (small) time component will begin to be implemented.

APPENDIX: MCMC ALGORITHM

The full conditional distributions for our model are given below, $[\Omega|\bullet]$ denotes the full conditional distribution of Ω given all other parameters. Let $j = 1, \ldots, 4$.

(i) $[\boldsymbol{\beta} \mid \bullet] \sim$ is distributed as $N[(\mathbf{X}'\Delta^{-1}\mathbf{X})^{-1}\mathbf{X}'\Delta^{-1}(\mathbf{Y}-\mathbf{Z}), (\mathbf{X}'\Delta^{-1}\mathbf{X})^{-1}]$.

(ii) $[\delta_j \mid \bullet] \sim$ is distributed as Inverse Gamma$(\alpha + \frac{n_j}{2}, \beta + \frac{1}{2}(\mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{Z}_j)'(\mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{Z}_j))$.

(iii) $[\mathbf{Z}_j \mid \bullet] \sim$ is distributed as $N[(\frac{1}{\delta_j}\mathbf{I} + \frac{1}{\sigma_j^2}\mathbf{K}_{\theta_j}^{-1})^{-1}(\frac{1}{\delta_j}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})), (\frac{1}{\delta_j}\mathbf{I} + \frac{1}{\sigma_j^2}\mathbf{K}_{\theta_j}^{-1})^{-1}]$. $\mathbf{Z}$ may be sampled from directly. It will have a mean vector consisting of the means of the four $\mathbf{Z}_j$'s and a covariance matrix that is block diagonal with the four blocks being of the covariance forms above.

(iv) $[\sigma_j^2 \mid \bullet] \sim$ is distributed as Inverse Gamma$(\alpha + \frac{n_j}{2}, \beta + \frac{1}{2}\mathbf{Z}_j'\mathbf{K}_{\theta_j}^{-1}\mathbf{Z}_j)$.

(v) $[\theta_j \mid \bullet]$ is $\propto \theta_j^{\alpha-1}e^{-\beta\theta_j}|\mathbf{K}_{\theta_j}|^{-1/2}\exp\{-\frac{1}{2\sigma_j^2}\mathbf{Z}_j'\mathbf{K}_{\theta_j}^{-1}\mathbf{Z}_j\}$.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. DEB-0423704, Central Arizona—Phoenix Long-Term Ecological Research (CAP LTER). Any opinions, findings, and conclusions or recommendation expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

REFERENCES

Banerjee S. Carlin BP, Gelfand AE. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall: New York, USA.

Berger JO, De Oliveira V, Sans B. 2001. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association* **96**: 1361–1374.

Chib S, Greenberg E. 1995. Understanding the Metropolis–Hastings algorithm. *The American Statistician* **49**: 327–335.

Cressie NAC. 1993. *Statistics for Spatial Data* (Revised Edition). Wiley-Interscience: New York, USA.

Gelfand AE, Smith AFM. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409.

Gelman A, Carlin J, Stern H, Rubin D. 1995. *Bayesian Data Analysis*. Chapman & Hall: New York, USA.

Gilks WR, Richardson S, Spiegelhalter DJ. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London, UK.

Hope D, Zhu W, Gries C, Oleson JJ, Kaye J, Grimm N, Baker L. 2005. Spatial variation in soil inorganic nitrogen across an arid urban ecosystem. *Urban Ecosystems* **8**: 253–273.

Kaiser J. 2001. An experiment for all seasons. *Science* **293**: 624–627.

Kass RE, Wasserman L. 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* **90**: 928–934.

Robert CP, Casella G. 1999. *Monte Carlo Statistical Methods*. Springer-Verlag: New York, USA.

Schlesinger WH, Raikes JA, Hartley AE, Cross AF. 1996. On the spatial pattern of soil nutrients in desert ecosystems. *Ecology* **77**: 364–374.

Wikle CK, Royle JA. 2004. Spatial statistical modeling in biology. In *Encyclopedia of Life Support Systems*. EOLSS Publishers Co. Ltd. to appear