Supplementary material to "Fast Methods for Spatially Correlated Multilevel Functional Data"

Ana-Maria Staicu,

Department of Statistics, North Carolina State University,

2311 Stinson Drive Raleigh, NC 27695-8203, USA *email*: staicu@stat.ncsu.edu,

Ciprian M. Crainiceanu,

Department of Biostatistics, Johns Hopkins University,

615 N. Wolfe St. Baltimore, MD 21205, USA. *email*: ccrainic@jhsph.edu

and

Raymond J. Carroll,

Department of Statistics, Texas A&M University,

College Station, TX 77843-3143, USA. *email*: carroll@stat.tamu.edu

Appendix A: Computational Complements

A.1 Matrix inversion

Let Σ_{dr} denote the variance of the $\sum_{i=1}^{M_{dr}} N_{dri} \times 1$ vector $\mathbb{Y}_{dr}$. Similarly, let ϵ_{dr} be the $\sum_{i=1}^{M_{dr}} N_{dri} \times 1$ vector of regression errors. Define $\xi_{dr} = (\xi_{dr,1}, ..., \xi_{dr,K_1})^{\mathrm{T}}$ and define $\zeta_{dr} = (\zeta_{dr1,1}, ..., \zeta_{dr1,K_2}, \zeta_{dr2,1}, ..., \zeta_{drM_{dr},K_2})^{\mathrm{T}}$ Define $\Sigma_\xi = \mathrm{cov}(\xi_{dr}) = \mathrm{diag}(\lambda_1^{(1)}, ..., \lambda_{K_1}^{(1)})$, $\Sigma_\beta = \mathrm{diag}(\lambda_1^{(2)}, ..., \lambda_{K_2}^{(2)})$ and $\Sigma_\zeta = \mathrm{cov}(\zeta_{dr}) = I \otimes \Sigma_\beta$.

Let $B_{dr}^{\mathrm{T}} = (\Phi_{dr1}^{(1)T}, \ldots, \Phi_{drM_{dr}}^{(1)T})$ be the $\sum_{i=1}^{M_{dr}} N_{dri} \times K_1$ matrix with elements $\{\phi_1^{(1)}(t), ..., \phi_{K_1}^{(1)}(t)\}$, where the arguments for t match those of the corresponding row of $\mathbb{Y}_{dr}$ and let $\mathbb{B}_{dr} = \mathrm{diag}(\Phi_{dr1}^{(2)}, \ldots, \Phi_{drM_{dr}}^{(2)})$ be the $\sum_{i=1}^{M_{dr}} N_{dri} \times K_2 M_{dr}$ matrix of $\phi_l^{(2)}(t)$'s. Let 1_{dri} be the $N_{dri} \times 1$ vector of ones, and $\mathbb{E}_{dr} = \mathrm{diag}(1_{dr1}, ..., 1_{drM_{dr}})$. Finally let $\mathbb{U}_{dr}$ be the $M_{dr} \times 1$ vector of $\{U(\Delta_{dri}) : i = 1, \ldots M_{dr}\}$

and $\Sigma_{U,dr}$ be the $M_{dr} \times M_{dr}$ variance covariance matrix of $\mathbb{U}_{dr}$. Then our model is $\mathbb{Y}_{dr} = \mu_{dr} + B_{dr}\xi_{dr} + \mathbb{B}_{dr}\zeta_{dr} + \mathbb{E}_{dr}\mathbb{U}_{dr} + \epsilon_{dr}$, where μ_{dr} is the vector obtained by stacking $\mu_d(t_{drip})$ over p's and i's. Clearly,

$$\Sigma_{dr} = \text{cov}(\mathbb{Y}_{dr}) = \sigma_\epsilon^2 I + B_{dr}\Sigma_\xi B_{dr}^T + \mathbb{B}_{dr}\Sigma_\zeta \mathbb{B}_{dr}^T + \mathbb{E}_{dr}\Sigma_{U,dr}\mathbb{E}_{dr}^T. \tag{S.1}$$

Here we show how using matrix and determinant equalities used in (Harville, 1977) can simplify both the inversion of and the determinant calculation of Σ_{dr}. For simplicity of notation only, since there will be no confusion, we drop the indices and write $M_{dr} = M$ and $N_{dri} = N$, etc. Thus, we denote $\Phi_{dr}^{(1)} = \Phi^{(1)}$ and $\Phi_{dri}^{(2)} = \Phi^{(2)}$, $\mathbb{E} = \mathbb{E}_{dr}$, $B_{dr} = B$ and $\mathbb{B}_{dr} = \mathbb{B}$.

To invert Σ_{dr}, we repeatedly use the formula

$$(A^{-1} + CB^{-1}C^{\mathrm{T}})^{-1} = A - AC(C^{\mathrm{T}}AC + B)^{-1}C^{\mathrm{T}}A. \tag{S.2}$$

Define $S = (\sigma_\epsilon^2 I + B\Sigma_\xi B^{\mathrm{T}} + \mathbb{B}\Sigma_\zeta \mathbb{B}^{\mathrm{T}})^{-1}$. Then $\Sigma_{dr}^{-1} = (S^{-1} + \mathbb{E}\Sigma_{U,dr}\mathbb{E}^{\mathrm{T}})^{-1} = S - S\mathbb{E}(\mathbb{E}^{\mathrm{T}}S\mathbb{E} + \Sigma_{U,dr}^{-1})^{-1}\mathbb{E}^{\mathrm{T}}S$. Note that $\mathbb{E}^{\mathrm{T}}S\mathbb{E} + \Sigma_{U,dr}^{-1}$ is $M \times M$. To compute S without taking a large matrix inverse, denote by $J = (\sigma_\epsilon^2 I + \mathbb{B}\Sigma_\zeta \mathbb{B}^{\mathrm{T}})^{-1}$. Then $S = (J^{-1} + B\Sigma_\xi B^{\mathrm{T}})^{-1} = J - JB(B^{\mathrm{T}}JB + \Sigma_\xi^{-1})^{-1}B^{\mathrm{T}}J$. The matrix $B^{\mathrm{T}}JB + \Sigma_\xi^{-1}$ is $K_1 \times K_1$, however, still J is the inverse of a $MN \times MN$ matrix. Fortunately, note that $\sigma_\epsilon^2 I + \mathbb{B}\Sigma_\zeta \mathbb{B}^{\mathrm{T}} = \text{diag}(\sigma_\epsilon^2 I_N + \Phi^{(2)}\Sigma_\beta \Phi^{(2)T}, \ldots, \sigma_\epsilon^2 I_N + \Phi^{(2)}\Sigma_\beta \Phi^{(2)T})$, where I_N is the $N \times N$ dimensional matrix. Hence, $J = \text{diag}\left\{(\sigma_\epsilon^2 I_N + \Phi^{(2)}\Sigma_\beta \Phi^{(2)T})^{-1}, \ldots, (\sigma_\epsilon^2 I_N + \Phi^{(2)}\Sigma_\beta \Phi^{(2)T})^{-1}\right\}$. Using (S.2) once again, we write

$$(\sigma_\epsilon^2 I_N + \Phi^{(2)}\Sigma_\beta \Phi^{(2)T})^{-1} \;\; = \;\; \sigma_\epsilon^{-2} I_N - \sigma_\epsilon^{-2}\Phi^{(2)}(\Phi^{(2)T}\Phi^{(2)} + \sigma_\epsilon^2 \Sigma_\beta^{-1})^{-1}\Phi^{(2)T}.$$

Hence: (1) we can compute J with nothing more than $K_2 \times K_2$ inversion, (2) we can compute S with nothing more than $K_1 \times K_1$ and $K_2 \times K_2$ inversion and (3) we can compute Σ_{dr}^{-1} with nothing more than $K_1 \times K_1$, $K_2 \times K_2$ and $M \times M$ inversion.

A.2 Matrix determinant

Here is use the formula

$$|A + CBC^{\mathrm{T}}| = |A||I + C^{\mathrm{T}}A^{-1}CB|. \tag{S.3}$$

Of course, $|\Sigma_{dr}| = |S^{-1} + \mathbb{E}\Sigma_{U,dr}\mathbb{E}^{\mathrm{T}}| = |S^{-1}||I + \Sigma_{U,dr}^{1/2}\mathbb{E}^{\mathrm{T}}S\mathbb{E}\Sigma_{U,dr}^{1/2}|$, where $I + \Sigma_{U,dr}^{1/2}\mathbb{E}^{\mathrm{T}}S\mathbb{E}\Sigma_{U,dr}^{1/2}$ is $M \times M$. However, $|S^{-1}| = |J^{-1}||I + \Sigma_{\xi}^{1/2}B^{\mathrm{T}}JB\Sigma_{\xi}^{1/2}|$, where $I + \Sigma_{\xi}^{1/2}B^{\mathrm{T}}JB\Sigma_{\xi}^{1/2}$ is $K_1 \times K_1$. Finally, remember that $J^{-1} = \sigma_\epsilon^2 I + \mathbb{B}\Sigma_\gamma\mathbb{B}^{\mathrm{T}} = \mathrm{diag}(\sigma_\epsilon^2 I_N + \Phi^{(2)}\Sigma_\beta\Phi^{(2)T}, \ldots, \sigma_\epsilon^2 I_N + \Phi^{(2)}\Sigma_\beta\Phi^{(2)T})$. Hence $|J^{-1}| = \det\{\sigma_\epsilon^2 I_N + \Phi^{(2)}\Sigma_\beta\Phi^{(2)T}\}^M$, which involves the determinant of a $N \times N$ matrix.

It follows, that we can compute $|\Sigma_{dr}|$ without large matrix inversions or the need to take determinants of large matrices.

References

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320–338.

Appendix B: Simulation Studies

B.1 Setup

Simulations studies were conducted to assess the practical performance of the proposed estimation procedure. We generated data from model (3.2), where $\mu_d(t) \equiv 0$, $\lambda_1^{(1)} = 2$, $\lambda_k^{(1)} = 0$, $k \geq 2$, $\lambda_\ell^{(2)} = 2(1/2)^{\ell-1}$ for $l = 1, 2$ while $\lambda_\ell^{(2)} = 0$ for $\ell \geq 3$. We assumed that the covariance function at each level is derived from the eigenfunctions $\phi_1^{(1)}(t) = \sqrt{2}\cos(4\pi t)$ at level 1 and $\phi_1^{(2)}(t) = \sqrt{3}(2t-1)$ and $\phi_2^{(2)}(t) = \sqrt{5}(6t^2 - 6t + 1)$ at level 2. Note that the level 1 eigenfunctions and the level 2 eigenfunctions are not mutually orthogonal. We generated: (a) the FPC scores from $\xi_{dr,1} \sim \mathrm{Normal}(0, \lambda_1^{(1)})$ and $\zeta_{dri,l} \sim \mathrm{Normal}(0, \lambda_\ell^{(2)})$ for $\ell = 1, 2$, (b) the spatial component from a Gaussian process $\{U_{dr}(\Delta) : \Delta \in [0, L]\}$ with zero-mean, variance $\sigma_U^2 = 1$ and autocorrelation function $\rho(\Delta)$ to be described below, and (c) the measurement error from $\epsilon_{dri}(t) \sim \mathrm{Normal}(0, \sigma_\epsilon^2)$. We used $D = 2$ groups, $R_d = 12$ subjects per group, $M_{dr} = 20$ units per subject located at various locations $\{\Delta_{dri} : i = 1, \ldots M_{dr}\}$ given by two different designs to be discussed next and $N_{dri} = 31$ subunits equally spaced in $[0, 1]$. The first design considers the locations to be independently and identically generated from the uniform distribution on $[0, L]$, where $L = 15,000$ microns, the actual length of the rat tissue in the colon carcinogenesis data. The second design uses the exact crypt locations from the colon carcinogenesis data, see Figure 1 of Baladandayuthapani, et al. (2008). The actual

locations used are reproduced in Figure S.1 For each design the Gaussian process was generated according to three types of correlation functions: (a) $\rho_1(\Delta) = \exp(-6|\Delta|/1000)$, (b) $\rho_2(\Delta) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(\Delta/\phi)^{\kappa}K_{\kappa}(\Delta/\phi)$, where $K_{\kappa}(\cdot)$ is the modified Bessel function (Stein, 1999) with parameters $\kappa = 1.5$ and $\phi = 120$, and (c) $\rho_3(\Delta) = \frac{1}{2}\cos(\Delta/60)/(1+|\Delta|/100) + \frac{1}{2}\exp(-|\Delta|/800)$; the correlation functions (b) and (c) were considered in Li, et al. (2007). The three correlation functions differ in few aspects. First, the correlation range, in the sense discussed in Section 4.2, is different for across the three functions: ρ_1 and ρ_2 decay to zero much faster than ρ_3. Second, the functions differ in the monotonicity behavior: ρ_1 and ρ_2 are monotone functions, while ρ_3 is not. Finally, the degree of smoothness at the value $\Delta = 0$ varies across the types of correlations considered, with ρ_2 being the smoothest function at $\Delta = 0$, and ρ_3 not being differentiable at $\Delta = 0$.

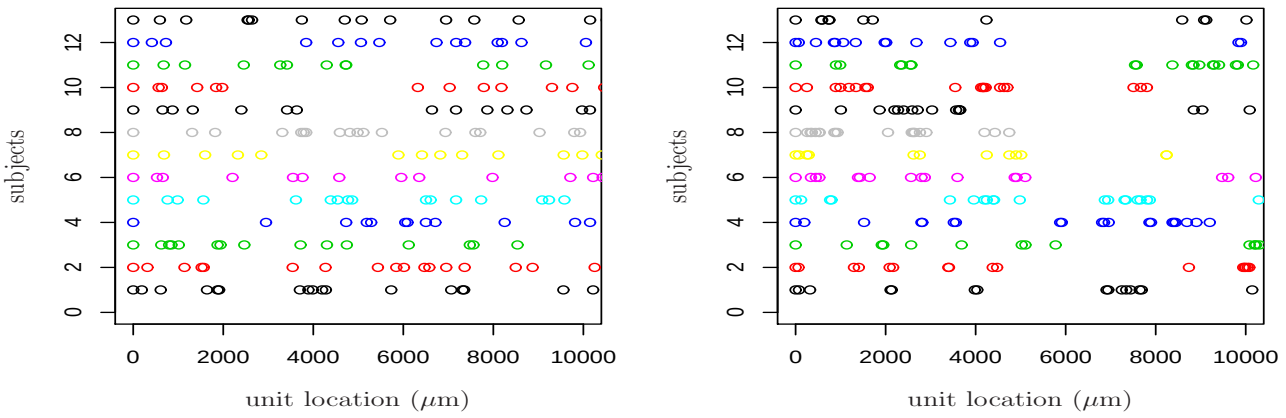

Figure S.1: Simulation study: the spatial location of the units used in the uniform design (left panel) and in the colon carcinogenesis study design (right panel) for 12 subjects.

We generated 1000 data sets for each scenario, both with measurement error so that $\sigma^2_{\epsilon} = 1$ and with no measurement error, $\sigma^2_{\epsilon} = 0$. For each data set, we used the algorithm described in Section 4 to estimate all the model components, including the number of principal components and the eigenfunctions. The following subsections elaborate our findings.

B.2 Spatial covariance estimation

Here we illustrate the performance of our proposed k-nearest neighbor estimator of the correlation/covariance functions. The selection of k is important for the estimation accuracy of the spatial covariance function. More precisely, large values of k lead to over-smooth covariance estimator, while smaller values of k favor a rougher estimator. We propose a penalized procedure based on cross-validation prediction error to select k.

Our criterion is inspired by Li et al. (2007) who suggested a modification of the cross-validation procedure of Diggle and Verbyla (1998), based on prediction error to choose the optimal bandwidth for a kernel variogram estimator. Specifically, define the cross-validation prediction error by:

$$\mathrm{CV}(k) := \sum_{d=1}^{D} \sum_{r=1}^{R_d} \sum_{\Delta_{dr,ij} \le \Delta^*} \sum_{s,t} \{\widehat{K}_r^Y(t, s, \Delta_{dr,ij}) - \nu(t, s, \Delta_{dr,ij})\}^2, \tag{S.4}$$

where $\nu(t, s, \Delta_{dr,ij}) = \{Y_{dri}(t, \Delta_{dri}) - Y_{drj}(t, \Delta_{drj})\}\{Y_{dri}(s, \Delta_{dri}) - Y_{drj}(s, \Delta_{drj})\}/2$ and $\widehat{K}_r^Y(t, s, \Delta_{dr,ij})$ is the k-nearest neighbor estimator using neighboring size k, as defined in (4.1) with all the information on the rth subject left out. We propose to choose k as the value which minimizes the $\mathrm{CV}(k)$, under the constraint that $k^2 \ge c$, where $c \ge 0$ is a tuning parameter controlling the smoothness of the covariance estimators. For $c = 0$, our optimization criterion simplifies to minimizing $\mathrm{CV}(k)$, a criterion similar to the procedure proposed by Li et al. (2007). For large c, the procedure encourages large values k, and therefore over-smooth covariance estimators. The optimization problem can be written in the Lagrangian formulation as a penalized optimization $\widehat{k} := \mathrm{argmin}_k\{CV(k) - \lambda k^2/2\}$, where $\lambda > 0$ is the Lagrange multiplier. This cross-validation criterion uses the prechosen cutoff Δ^* and assumes no covariance structure of the model.

The penalty λ and the smoothing parameter k are among the zeros of the derivative of the function $\mathrm{CV}_\lambda(k) := CV(k) - \lambda k^2/2$. In practice we use a two-step procedure 1) first choose $\lambda = \lambda^o$ as the mean of the ratio of lagged differences and corresponding neighboring size $\{CV_\lambda(k+1) - CV_\lambda(k)\}/k$, for a large enough grid of k's; and 2) select the value k which minimizes $\mathrm{CV}_{\lambda^o}(k)$. We found this simple criterion to work well in the simulations and application. In fact, in the simulation study, we observed minimal differences in the bias and mean integrated square error of the covariance estimator using values of k obtained with our optimization criterion versus other reasonable choices of k. For example, similar accuracy results are reported if the neighboring size

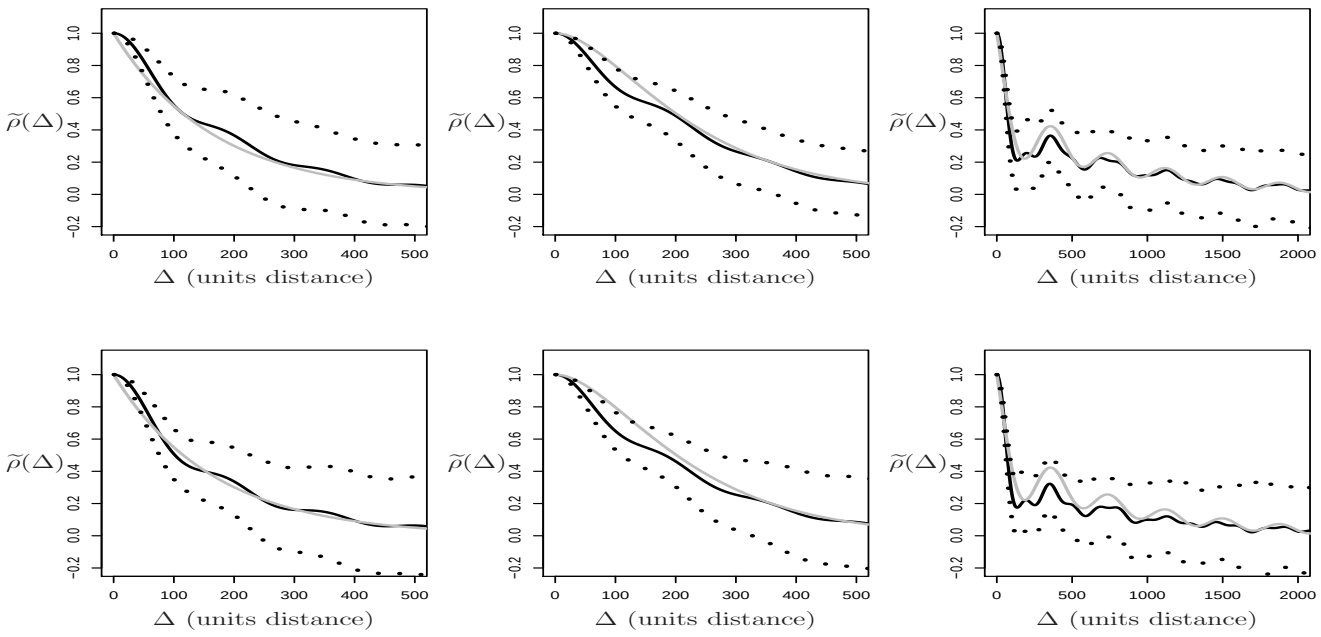

Figure S.2: The mean of the estimated correlation functions by k-nearest neighbor with positive semi-definite adjustment (solid lines) along with their pointwise 90% confidence interval in the case of the uniform design (top panel) and the CCS design (bottom panel) of the units location. The true correlation functions (grey line) are ρ_1 (left), ρ_2 (middle) and ρ_3 (right).

is fixed at $k = 40$ for all the simulation scenarios considered.

The spatial covariance estimator is not positive semi-definite, which could have serious implications when estimating the other components of the model. To correct this problem we followed Li, et al. (2007) and used an adjustment based on the spectral density (Christakos, 1984). Our simulations indicate that this correction decreases the variance of the covariance estimator. We denote by $\widetilde{\nu}(\Delta)$ the positive semi-definite adjusted estimator of $\nu(\Delta)$.

Figure S.2 gives the mean of the adjusted correlation estimators $\widetilde{\rho}(\Delta) = \widetilde{\nu}(\Delta)/\widetilde{\nu}(0)$ along with their 90% pointwise confidence intervals, for the two designs of the unit locations. The corresponding figure for the covariancefunctions is given in Figure S.5. Both the correlation function and the covariance function (results not shown) are very nearly unbiased and suggest somewhat smaller variability for the estimator in the uniform design than in the actual design in the colon carcino genesis study. The estimated variance, $\widehat{\sigma}^2_U = \widetilde{\nu}(0)$, was also nearly unbiased for the different designs and correlation functions; boxplots are available in Figure S.6.

In addition to variability we studied the integrated squared errors (ISE) of the correlation and covariance functions estimators. As previously remarked by Li, et al. (2007) or Hall and

Table S.1: Mean integrated squared error of the correlation and covariance estimators (positive semi-definite adjusted), when $\sigma_\epsilon^2 = 1$.

Method	Uniform design			CCS design		
	$\rho_1(\Delta)$	$\rho_2(\Delta)$	$\rho_3(\Delta)$	$\rho_1(\Delta)$	$\rho_2(\Delta)$	$\rho_3(\Delta)$
Correlation	10.28	6.42	7.36	10.43	9.36	9.96
Covariance	8.74	7.57	9.57	10.87	11.97	16.05

Patil (1994), and as seen by us, the positive semi-definiteness adjustment of the covariance estimator significantly improves the ISE. The numerical results agree with the above findings on the variability of the estimators; see Table S.1.

B.3 Multilevel functional estimation

B.3.1 Eigenfunctions and eigenvalues

Our methodology performs remarkably well at recovering the true eigenfunctions, in all scenarios we considered. For brevity, we present results only for two scenarios corresponding to the two spatial designs for the unit locations and correlation function $\rho_1(\Delta)$. Figure S.3 shows the estimated eigenfunctions for the case when the noise standard deviation is $\sigma_\epsilon = 1$; similar results are obtained in the case of no measurement error. The estimation methods correctly identify the different levels of variation.

These results, and many other unreported simulation results, indicate that the estimation quality of the true basis functions at each level is very little affected by the spatial structure (units location sampling design or correlation function). Moreover, estimation accuracy varies slightly with the choice of Δ^* for the level 1 eigenfunctions. In practice we choose Δ^* so that it is reasonable to assume that observations at distance Δ^* apart or larger are uncorrelated. For the simulation study we used different values of Δ^* corresponding to different correlation types. In the case of ρ_1 and we took $\Delta^* = 1,000$ and the estimated eigenfunctions are presented in Figure S.3.

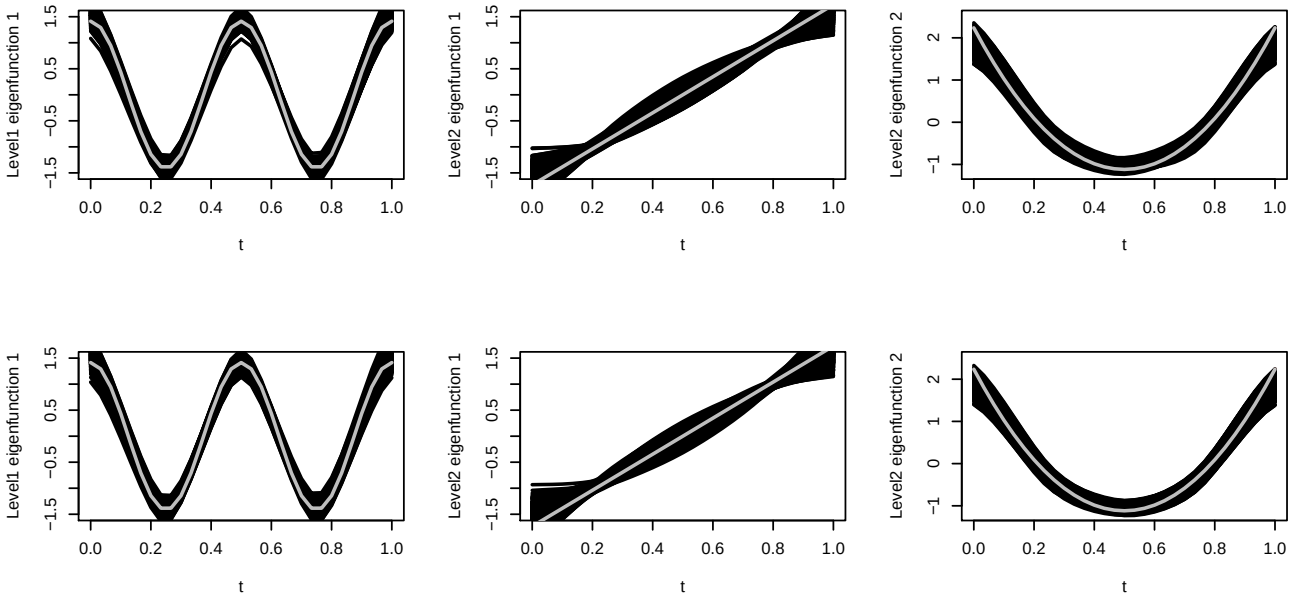

Figure S.3: The estimated (black) versus the true (grey) eigenfunctions at level 1 and level 2 in 1000 simulations, when the data uses crypt locations given by a uniform design (top panel) and the colon cancer design (bottom panel), the correlation function $\rho_1(\Delta)$ and an error variance $\sigma_\epsilon^2 = 1$.

We now turn to the estimation of the eigenvalues. The presence of measurement error does not affect the eigenvalue estimation in any of the scenarios considered. As in the estimation of eigenfunctions, the estimated eigenvalues are very little affected by the spatial design or by the correlation structure of the spatial process involved. There are small biases in the estimated eigenvalues, and large variance for the eigenvalues estimates at level 1; see Figure S.7. This is not surprising, given the relatively moderate sample size used. Further simulation studies, unreported here, indicate that these small problems disappear with increased sample size.

B.3.2 Functional Principal Component scores

FPC scores are important because they provide level-specific functional predictions; for example, in our application they quantify the effect of the cell relative depths on the level of the biological marker p27. Here we present the root mean square error of the estimated FPC scores for all scenarios considered, when the measurement error is $\sigma_\epsilon^2 = 1$. Table S.2 summarizes the results based on 1000 simulations.

Interestingly, the type of spatial correlation has minimal effect on root mean squared error.

Table S.2: Root mean squared error of the FPC scores for different correlation types and design scenarios, when $\sigma^2_\epsilon = 1$.

FPC scores	Uniform design			CCS design		
	$\rho_1(\Delta)$	$\rho_2(\Delta)$	$\rho_3(\Delta)$	$\rho_1(\Delta)$	$\rho_2(\Delta)$	$\rho_3(\Delta)$
Level 1: component 1	0.405	0.405	0.405	0.406	0.406	0.406
Level 2: component 1	0.227	0.227	0.228	0.228	0.228	0.230
Level 2: component 2	0.222	0.222	0.224	0.225	0.225	0.229

Similar patterns of results are obtained in the case without measurement error: the main difference being that the root mean squared error for the functional principal component scores is of smaller magnitude; for instance the root mean squared error is equal to 0.403 for the estimated components at level 1 and to 0.147, 0.148 for the estimated components at level 2, in the case of colon carcinogenesis design for the unit locations and correlation $\rho_1(\Delta)$.

Finally, we compare the performance of the estimated noise variance σ^2_ϵ across the different scenarios. The estimates have negligible bias for the two unit locations designs and three correlation functions. Figure S.4 displays the boxplots of $\widehat{\sigma}^2_\epsilon$.

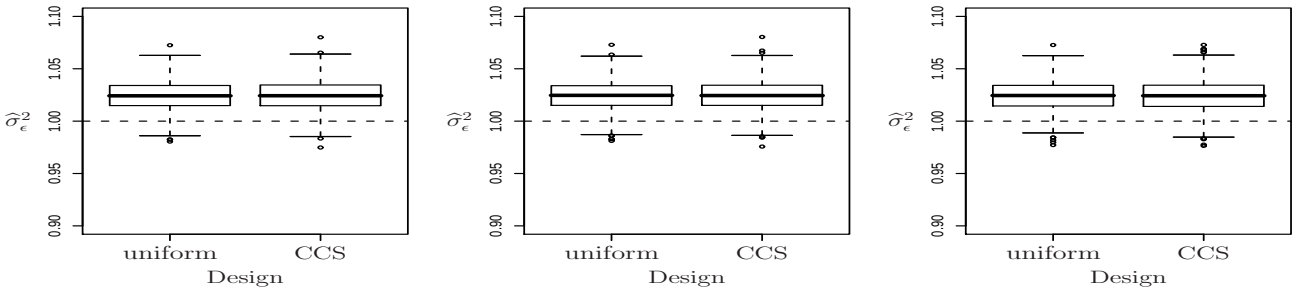

Figure S.4: Simulation study: boxplots of the estimated variance of the measurement error, $\widehat{\sigma}^2_\epsilon$, for the correlation functions ρ_1 (left), ρ_2(middle) and ρ_3(right), and the two designs for the crypt locations.

Appendix C: Additional Figures

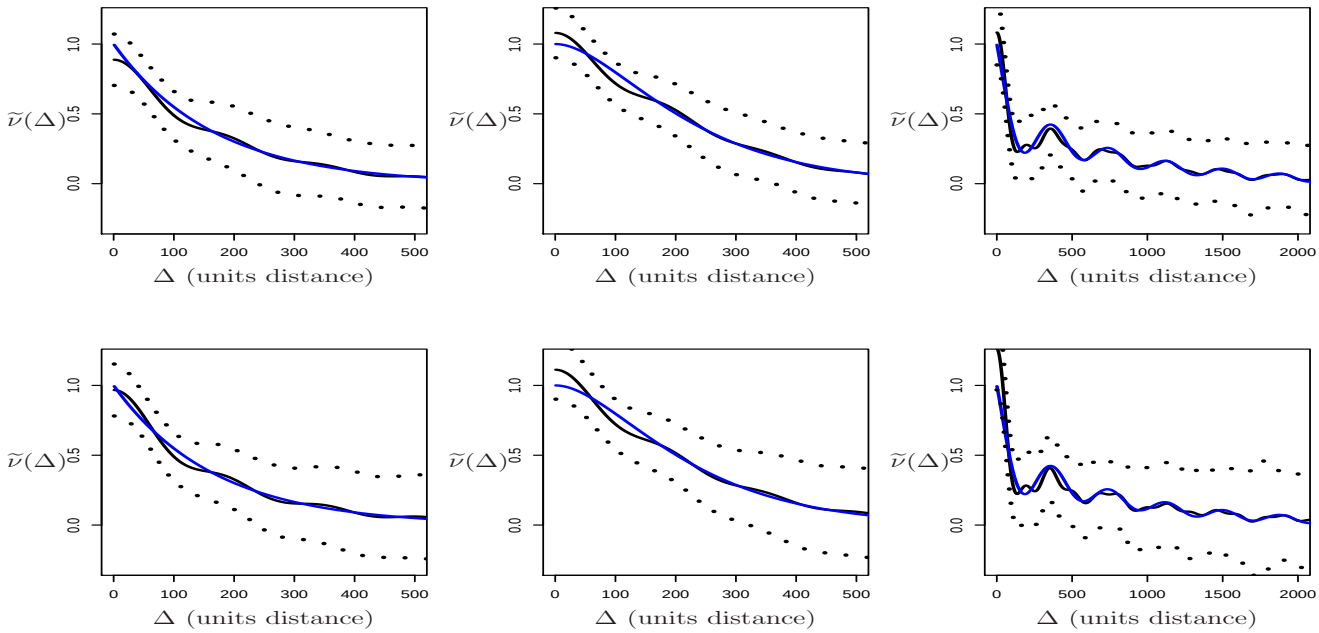

Figure S.5: Simulation study: the mean of the adjusted estimates of the covariance functions along with their pointwise 90% confidence interval in the case of the uniform design (top panel) and the CCS design (bottom panel) of the units location. The true covariance functions (blue line) are $\nu_1 = \sigma_U^2 \rho_1$ (left), $\nu_2 = \sigma_U^2 \rho_2$ (middle) and $\nu_3 = \sigma_U^2 \rho_3$ (right); the estimates are obtained using k-nearest neighbor with positive semi-definite adjustment (solid lines).

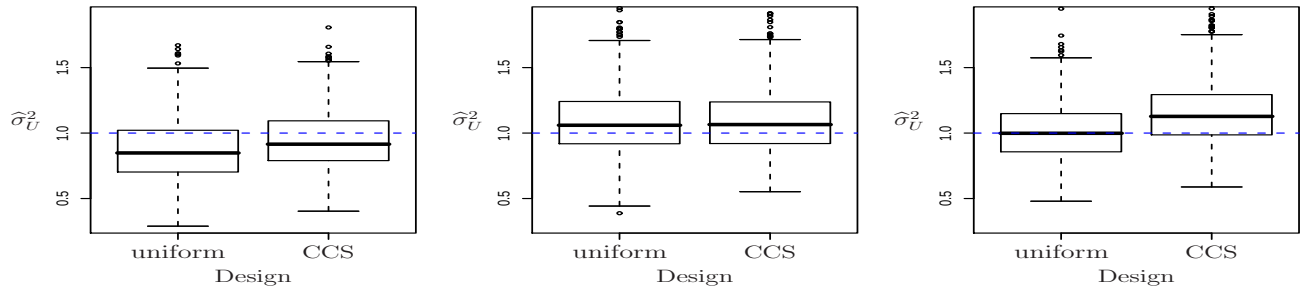

Figure S.6: Simulation study: boxplots of the estimated variance of the spatial process U_{dr} for the three correlation functions ρ_1 (left panel), ρ_2 (middle panel) and ρ_3 (right panel) in the case of uniform and CCS spatial design.

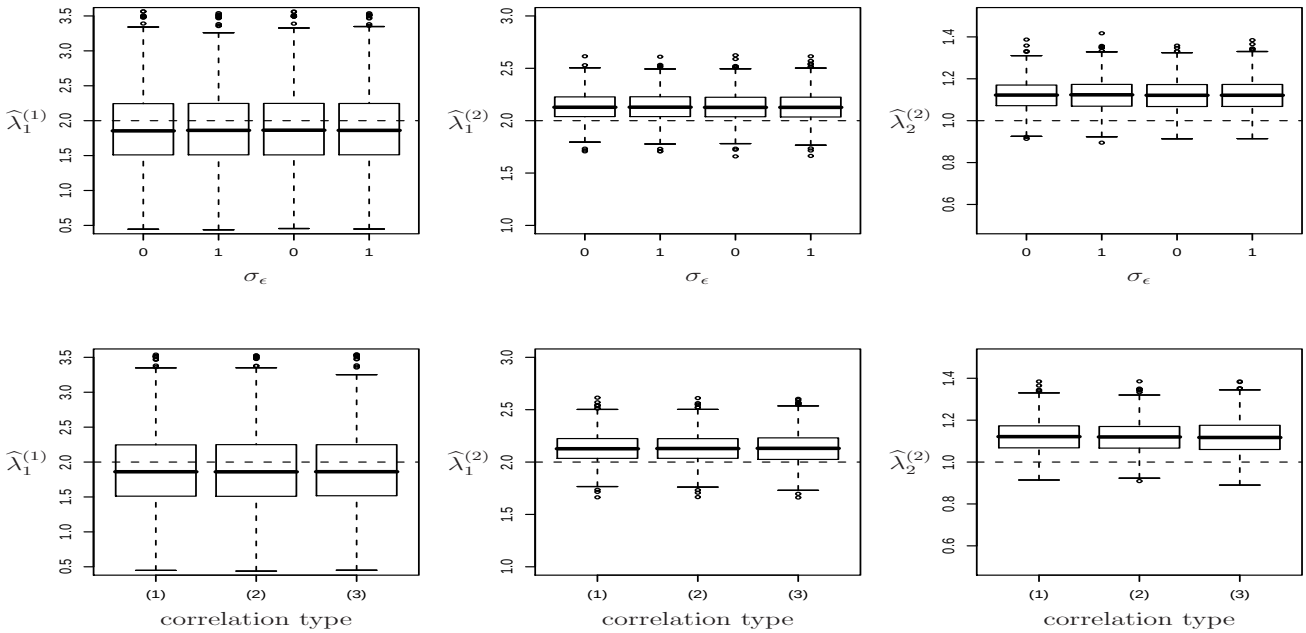

Figure S.7: Simulation study: boxplots of the estimated eigenvalues at each hierarchical level. The top panel corresponds to the correlation function $\rho_1(\Delta)$ under the uniform design (left) and CCS design (right) for two magnitudes of measurement error $\sigma_\epsilon = 0$ and $\sigma_\epsilon = 1$. The bottom panel corresponds to correlation functions ρ_1 in (1), ρ_2 in (2) and ρ_3 in (3), under CCS design and $\sigma_\epsilon = 1$.

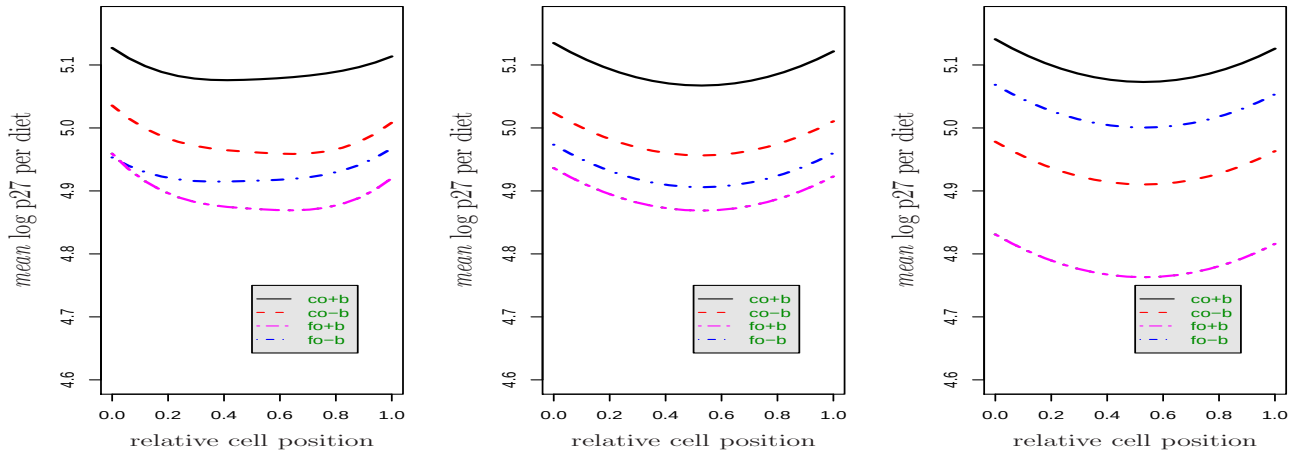

Figure S.8: The estimated mean functions for the 4 diet groups by: (left panel) penalized spline smoothing under a working independence assumption, (middle panel) ordinary least squares quadratic estimation along with the independence assumption and (right panel) weighted least squares quadratic estimation, accounting for the dependence considered by the model.