I.E.2 Continuous-by-continuous interactions (including 2D spatial smoothing) RWC Chapter 13

Two examples (RWC 13.1)

- Number of scallops caught off Long Island NY
  - counts are made at specific x-y coordinates (latitude & longitude)
- Incidence of AIDS in Italian MSM
  - predictors: - calendar year
    - age at diagnosis

In technical terms these problems are identical

BUT there's an important difference

- AIDS: - the two predictors are on meaningful scales
- Scallops: - the coordinate axes have no inherent meaning; they are purely conventional and could be replaced with no loss

→ Affects the choice of basis functions

Basis #1 Tensor product basis (RWC 13.2)

- s, t continuous predictors: $y_i = f(s_i, t_i) + \varepsilon_i$

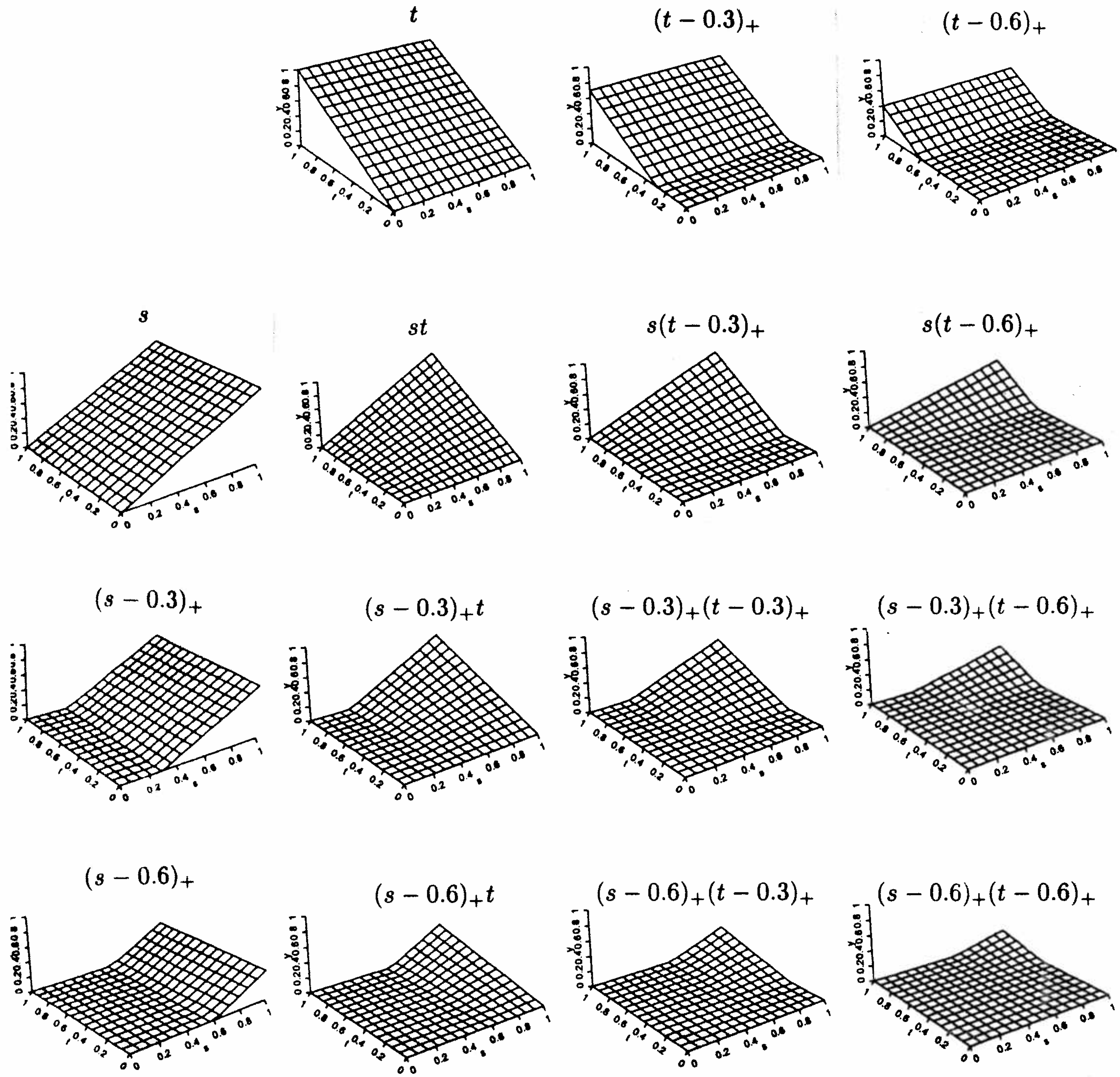
- The natural extension of the truncated-lines basis in 1 Dimension is:

$$
\begin{aligned}
y_i = \beta_0 &+ \beta_s s_i + \sum_{k=1}^{K^s} u_k^s (s_i - \kappa_k^s)_+ \\
&+ \beta_t t_i + \sum_{k=1}^{K^t} u_k^t (t_i - \kappa_k^t)_+
\end{aligned}
\quad \text{additive part (main effects)}
$$

$$
\begin{aligned}
&+ \gamma s_i t_i + \sum_{k=1}^{K^s} v_k^s s_i (t_i - \kappa_k^t)_+ \\
&+ \sum_{k=1}^{K^t} v_k^t t_i (s_i - \kappa_k^s)_+
\end{aligned}
\quad \text{interaction part.}
$$

$$
+ \sum_{k=1}^{K^s} \sum_{k'=1}^{K^t} v_{kk'}^{st} (s_i - \kappa_k^s)_+ (t_i - \kappa_{k'}^t)_+ + \varepsilon_i
$$

For $K_1^s = K_1^t = 0.3$ & $K_2^s = K_2^t$, this basis is:



IE 2/3 2/17/08

Drawback of tensor-product basis: It depends on the original coordinate axes. Results will change if you use axes with different orientations.

Basis #2 Radial basis functions (RWC 13.2)

- We saw a less general version of these in Sec IC.
- More general form: $C(\|(s_i, t_i) - (K_k^s, K_{k'}^t)\|)$, where:
  - $C(\cdot)$ is a univariate function $\mathbb{R}^+ \to \mathbb{R}$
  - $\|\cdot\|$ is a distance measure
- The value of this basis function for (i, k, k') depends only on the distance $\|\cdot\|$ from (s_i, t_i) to $(K_k^s, K_{k'}^t)$
- With a radial basis, the solution is invariant to axis rotations.

I.E.2/4 2/17/08

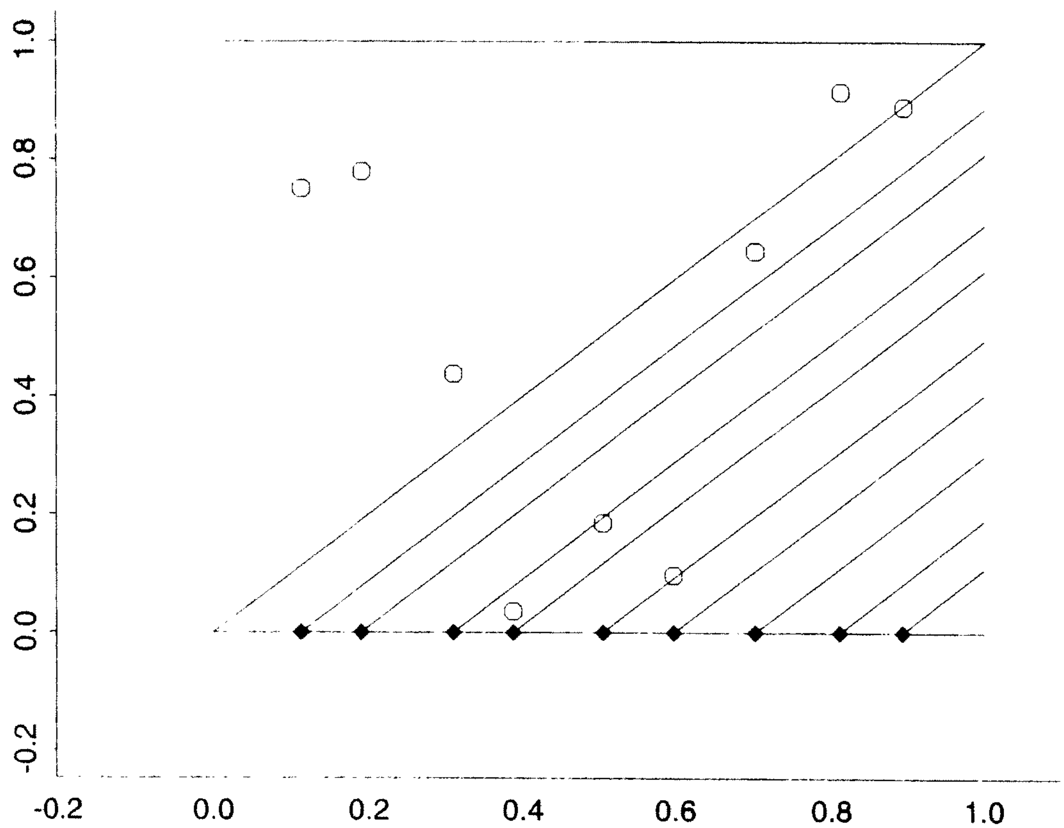
General Radial Smoothing (RWC 13.4)

- We'll do this in <u>1</u> Dimension (like RWC); generalizing to $d > 1$ dimensions is "trivial" (to RWC!)

- Broadly speaking - and don't forget that I know little about this topic - this strikes me as a lot of ad-hoc tinkering, that happens to start with a given model (pp. 249-250) but happily "kluges" that model to achieve properties that seem desirable
  - I don't think there's anything wrong with this approach (unlike certain purist Bayesians)
  - <u>BUT</u> it would seem to suggest that many other paths <u>could</u> be taken and might be fruitful

    $\Rightarrow$ Thesis topics!!

Start with a truncated-line basis for 1-D p-spline:



Define $X = [1 \quad x_i]_{1 \le i \le n}$ $\quad Z = \left[(x_i - x_j)_+\right]_{1 \le i, j \le n}$ Note: knots = unique x_i

P-spline is $\hat{y} = X\hat{\beta} + Z\hat{u}$ where $(\hat{\beta}, \hat{u})$ solve

$$\underset{\beta, u}{\operatorname{argmin}} \left(\|y - X\beta - Zu\|^2 + \lambda^2 u'u \right)$$

The EBLUP of $y = X\beta + Zu + \varepsilon$ $\quad \operatorname{cov}\begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \left[\begin{array}{c|c} \sigma_u^2 I & 0 \\ \hline 0 & \sigma_\varepsilon^2 I \end{array}\right]$

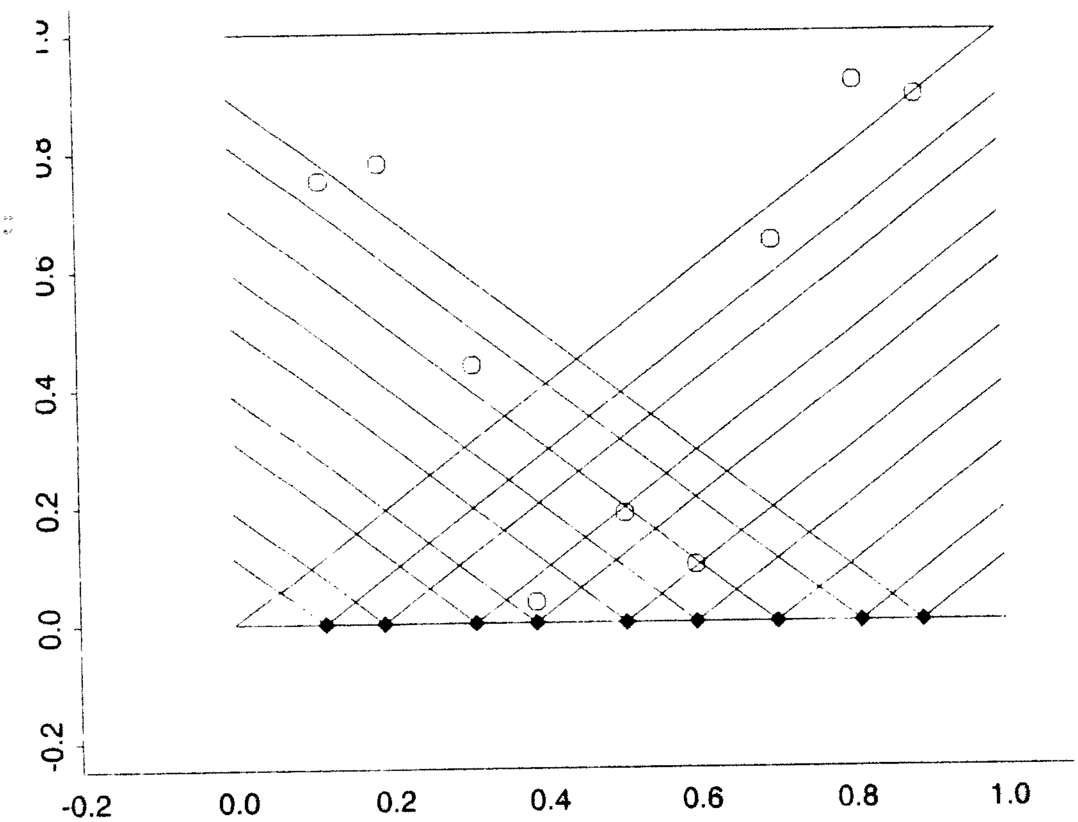
for $\lambda^2 = \dfrac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_u^2}$

IE 2/6 2/17/08

Now transform this basis to

$$[X \ \ Z_R] = [X \ \ Z] L \qquad (\text{such an } L \ \exists)$$

so X is unchanged and $Z_R = [|x_i - x_j|]_{1 \le i,j \le n}$



The new fitted values are $\hat{y} = X\hat{\beta}_R + Z_R \hat{u}_R$ solving

$$\underset{\beta_R, u_R}{\operatorname{argmin}} \left(\|y - X\beta_R + Z_R u_R\|^2 + \underbrace{\lambda^2 [\beta, u] L' D L \begin{bmatrix} \beta \\ u \end{bmatrix}} \right) \quad D = \begin{bmatrix} 0_{2\times 2} & 0 \\ 0 & I \end{bmatrix}$$

Z_R is radially symmetric; penalty is not, and doesn't extend to >1 dimension easily.

IE 2/7 2/17/08

Adhoc adjustment #1:

change the penalty to $\lambda u' Z_R u$

Now $\hat{y} = X\hat{\beta} + Z_R \hat{u}$ comes from solving

$$\underset{\beta, u}{\operatorname{argmin}} \left(\|y - X\beta - Z_R u\|^2 + \lambda u' Z_R u \right)$$

Good luck! This is a thin-plate spline $\left\{ \begin{array}{l} \text{lots of} \\ \text{literature} \end{array} \right.$

Bad luck!? This can't be written as a mixed linear model (MLM) because that would require

$$\operatorname{cov}(u) = G = \sigma_u^2 Z_R^{-1}$$

and Z_R is not necessarily positive definite.

(Easy counterexample: $x_i = 1, 2, \ldots, 100$: Z_R has 99 negative eigenvalues)

<u>ALSO</u> We want low-rank smoothers, unlike these full-rank smoothers

This leads to ...

I E 2/8 2/17/08

Ad hoc adjustment options #2a (RWC 13.4.1)

OR #2b (RWC 13.4.2)

OR #2c (RWC 13.4.3)

- I'll go right to RWC 13.4.4, which is option #2b combined with reducing the smoother's rank.

- Define $Z_C = \left[C(|x_i - \kappa_k|) \right]_{\substack{1 \le k \le K}}^{1 \le i \le n}$ — $C(r) = r$ gives basis on previous pages.

  for knots $\kappa_1, \dots, \kappa_K$

  - C may have unknown parameters

- Fit $y = X\beta + Z_C u + \varepsilon \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$

  $$\mathrm{cov}(u) = \sigma_u^2 \left(\Omega_K^{-1/2}\right)\left(\Omega_K^{-1/2}\right)'$$

  for $\Omega_K = \left[C(|\kappa_k - \kappa_{k'}|) \right]_{1 \le k, k' \le K}$

Notation slightly different from RWC!

To get $\Omega_K^{-1/2}$: Ω_K has SVD $U \operatorname{diag}(\underset{\sim}{d}) V^T$, U, V are $\perp$, $d_i \ge 0$

$$\Omega_K^{-1/2} = U \left[\operatorname{diag}(\underset{\sim}{d})\right]^{-1/2} V^T$$

IE 2/9 2/17/08

Final step: Re-parameterize the random effects with design matrix $Z = Z_C \Omega_K^{-1/2}$ giving:

$$y = X\beta + Zu + \varepsilon \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$$
$$u \sim N(0, \sigma_u^2 I_K)$$

recall $Z_C = [C(|x_i - \kappa_k|)]_{n \times K}$

$\Omega_K = [C(|\kappa_k - \kappa_{k'}|)]_{K \times K}$

To extend this to higher (≥ 2) dimensions,

replace $|x_i - \kappa_k|$ with $\|x_i - \kappa_k\|$

and $|\kappa_k - \kappa_{k'}|$ with $\|\kappa_k - \kappa_{k'}\|$

In terms of the technical machinery, it really is trivial.

IE 2/10 2/17/08

You still have to choose the function $C(\cdot)$, which determines Z (RE design matrix)

- Approx thin-plate splines are obtained with X having polynomials and $C(\underset{\sim}{r}) = \begin{cases} \|\underset{\sim}{r}\|^{2m-d} & d \text{ odd} \\ \|\underset{\sim}{r}\|^{2m-d} \log\|\underset{\sim}{r}\| & d \text{ even.} \end{cases}$

- RWC also mention the Matérn class, of which the two simplest C functions are

$$C(\underset{\sim}{r}) = \begin{cases} \exp(-\|\underset{\sim}{r}\|/\rho) & \nu = \tfrac{1}{2} \text{ in Matérn class} \\ \exp(-\|\underset{\sim}{r}\|/\rho)(1 + \|\underset{\sim}{r}\|/\rho) & \nu = \tfrac{3}{2} \end{cases}$$

You also have to choose knot locations.
- Rectangular lattice is wasteful if data aren't on a rectangular grid
- RWC like space-filling knot selection (13.4.6) (seems very reasonable.

- IE 2/11 2/17/08

Hodges (ignorant) opinions: I don't understand

- the love affair with Matérn covariances
- what on earth $\tilde{Z} = Z_C \Omega_K^{-1/2}$ looks like.
  - and how much it depends on the choice of C
  - just within the Matérn class, how much does it depend on ν and ρ?

(*)

$$C(\tilde{s}) = \begin{cases} \exp(-\|\tilde{s}\|/\rho) & \nu = 1/2 \\ \exp(-\|\tilde{s}\|/\rho)(1 + \|\tilde{s}\|/\rho) & \nu = 3/2 \end{cases}$$

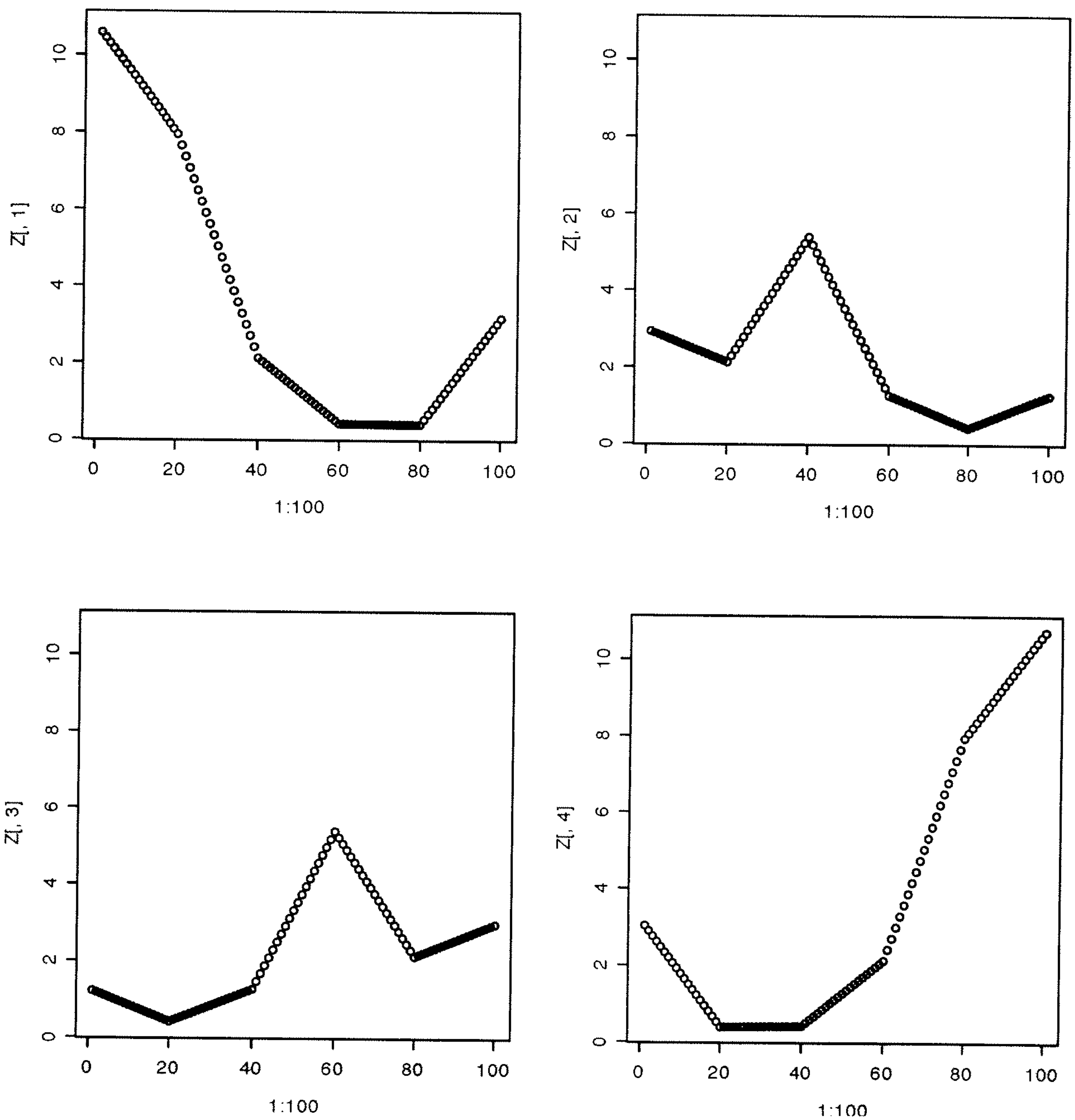
This would make a good class project, perhaps using an approach like the following, which we saw earlier as a way of understanding what the CAR-distributed random effect does. (NOTE! The spline literature probably has some of this already.)

IE 2/12 2/17/08

Dumbest, simplest case:

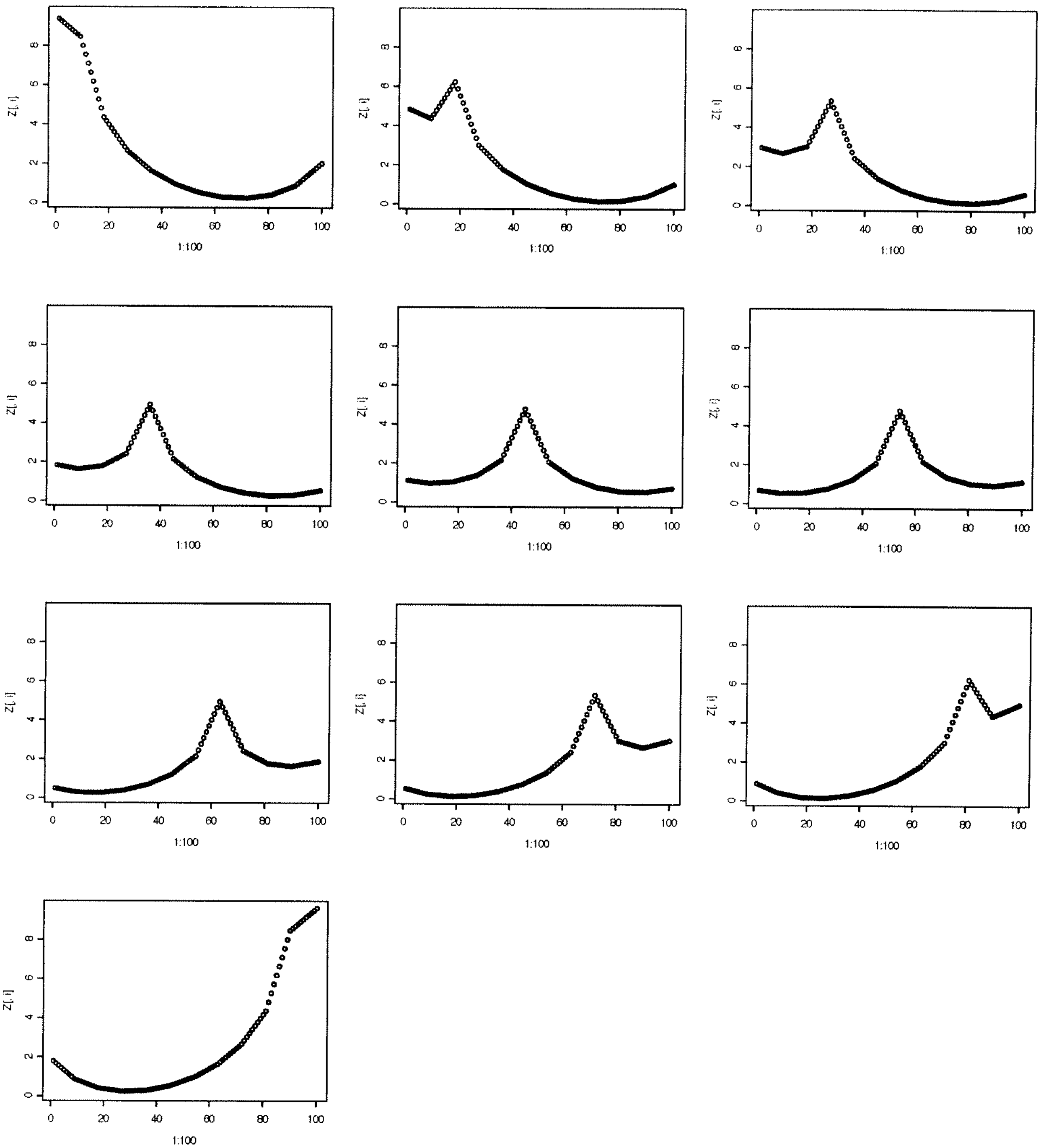
- 1 dimension, $C(r) = |r|$
- $\{x_i\} = \{1, 2, \dots, 100\} - 50.5$
- Knots: $\{K_1 = 20, K_2 = 40, K_3 = 60, K_4 = 80\} - 50.5$

Here are plots of the 4 columns of $Z = Z_C \Omega_K^{-1/2}$:



IE 2/13 2/17/08

Here's the analogous picture for 10 knots



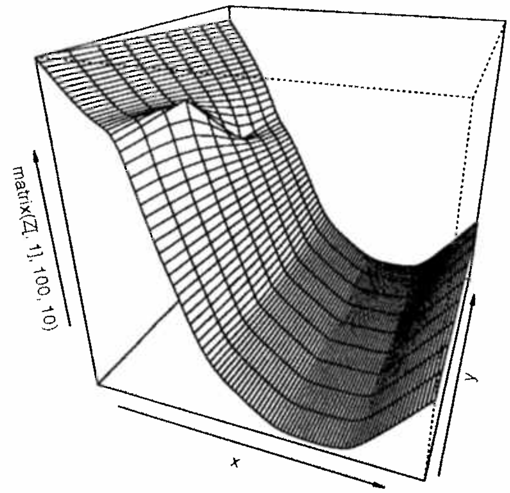
IE 2/13a   2/18/08

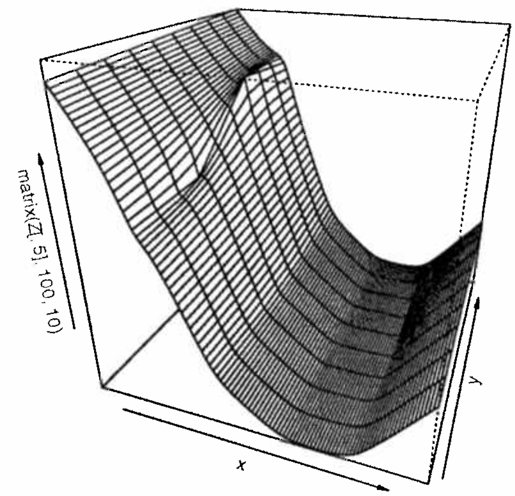
Dumbest, simplest 2-D case

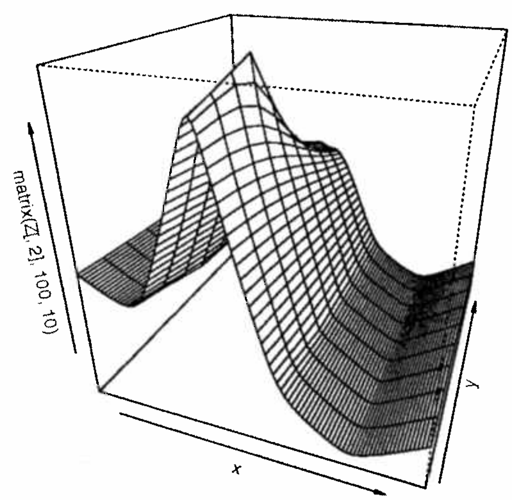
- $C(\tilde{s}) = \|\tilde{s}\|$, Euclidean distance
- $\begin{cases} x_i : \{1, 2, \ldots, 100\} - 50.5 \\ y_i : \{1, 2, \ldots, 10\} - 5.5 \end{cases}$

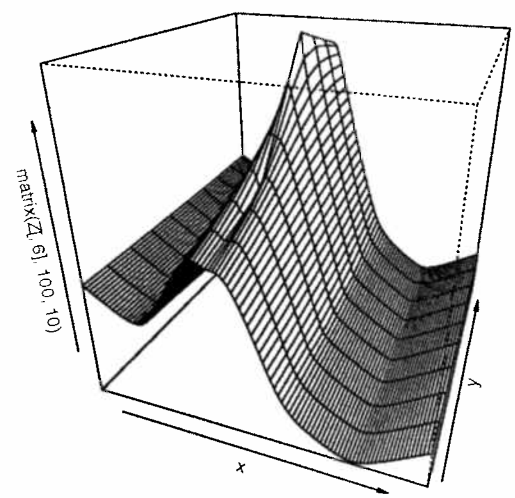
- Knots: (20*(1:4) , 1+3*(1:2)) − (50.5, 5.5)

IE 2/14 2/17/08









IE 2/14a 2/18/08

A possible class project:

- Compare a CAR smooth to a radial-basis smooth in a case where both could be used
  - e.g. GMST data
  - MN counties, with county centroids used as $\underset{\sim}{x}_i$ for radial basis fit.

- fit using REML

- fit using Bayes where both fits have the same prior on DF (!)

IL 2/15 2/17/08