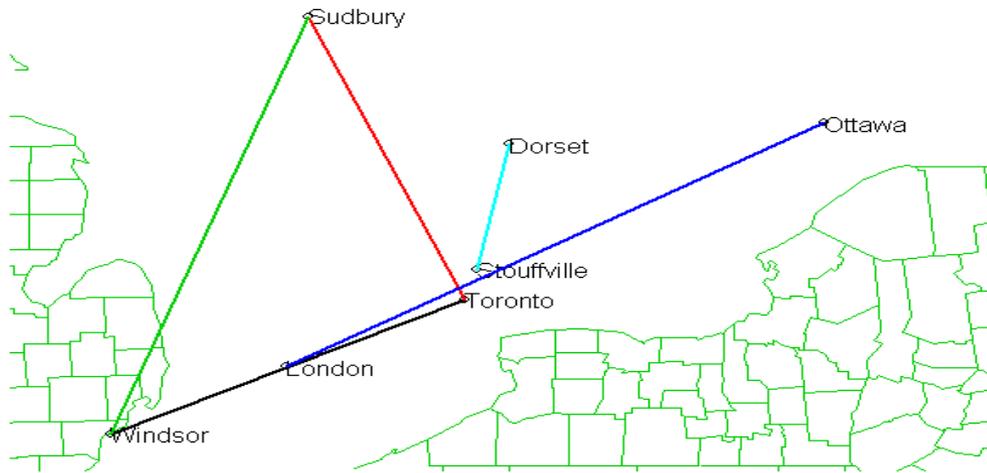


Nonstationary spatial covariance modeling through spatial deformation

**Paul D. Sampson --- Wendy Meiring
Univ of Washington --- U.C. Santa Barbara**

this presentation derived from that presented at the
Pan-American Advanced Study Institute on
Spatio-Temporal Statistics
Búzios, RJ, Brazil
June 16-26, 2014

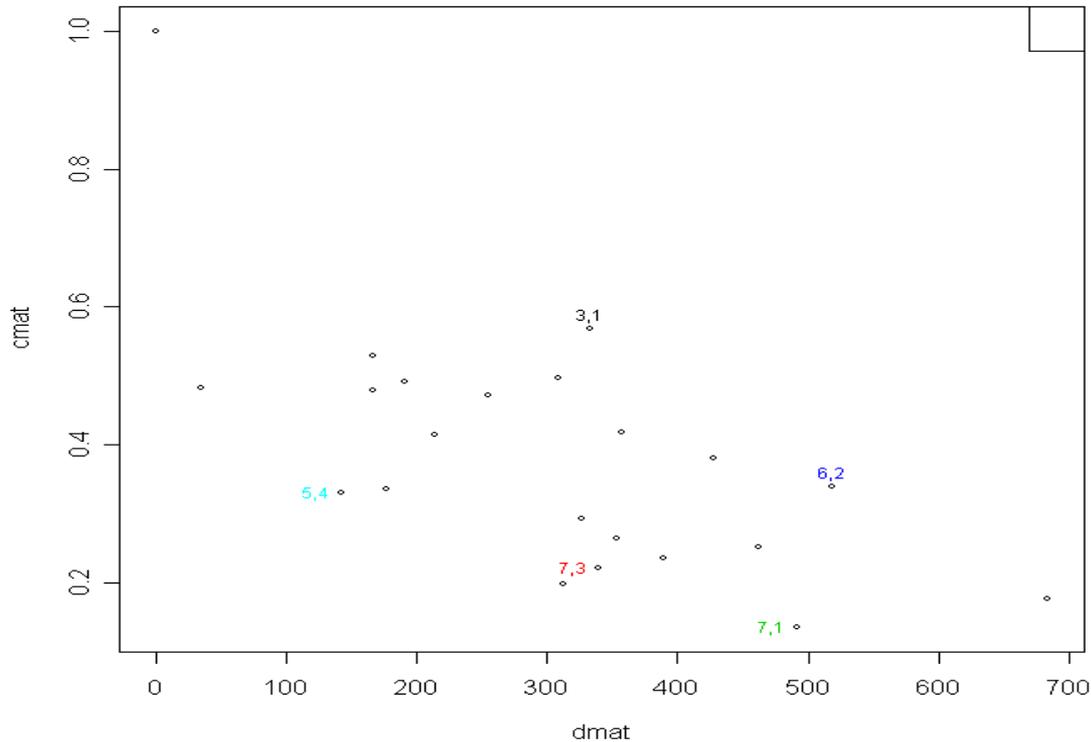
Correlation vs Distance for Ontario Ozone Data



(from Le and Zidek, UBC, Vancouver, CA)

Apparent anisotropy

in this plot of correlation vs distance



Perspective: 1st and 2nd order stationarity is almost never a realistic assumption for any environmental monitoring data, except at small spatial scales.

Objectives for approaches to nonstationary spatial covariance modeling.

- **Characterizing spatially varying (locally stationary) anisotropic structure.**
- **Scientific understanding/representation of covariance structure—not just a method of providing covariances for kriging.**

Capable of:

- ❖ reflecting effects of known explanatory environmental processes such as transport/wind, topography, point sources
- ❖ modeling effects of known explanatory environmental processes

Objectives (cont.)

- Application to purely spatial problems and/or problems with data sampled irregularly in space and time
- Application in context of dynamic models for space-time structure
- Application to “large” problems/data sets

❖ **Diagnostics for local and large-scale correlation structure:**

- is the spatial structure “right”
- is the nature/degree of nonstationarity (smoothness) right?

❖ **Evaluation of uncertainty in estimation (interpolation) of spatial covariance structure**

❖ **Incorporation in an approach to spatial estimation accounting for uncertainty in estimation of (parameters of) spatial covariance structure**

Selected classes of methods:

- **Spatial deformation models** (Sampson & Guttorp, Damian, Perrin, Meiring, Monestiez, Schmidt & O'Hagan, *Fouedjio*, ...)
- **Process convolution models** (Higdon, Swall & Kern, Calder, ...; Paciorek & Schervish, Risser & Calder)
- **Kernel/smoothing methods** (Fuentes, ...)
- **Models with covariates** (Reich et al., Schmidt et al.)
- **Basis function methods, including EOF, Karhunen-Loeve, and wavelets** (Nychka, Wikle, Pintore & Holmes, ...)
- **MDS-related dimension expansion** (*Bornn* et al.)
- See “Constructions for Nonstationary Spatial Processes”, Chap 9 in 2010 Handbook of Spatial Statistics, eds. Gelfand, Diggle, Fuentes, Guttorp.
- This week we will review some basics and then present our approaches to spatial deformation models. Next week we will discuss other models with a focus on process convolution models and a new R package for these models.
- Note: There is also a spectral approach to nonstationary spatial processes that provides a test for nonstationarity in terms of an assessment of interaction between location and frequency.

Review: Descriptive characteristics of (stationary) spatial covariance expressed in a variogram

The spherical correlation

$$\rho(\mathbf{v}) = \begin{cases} 1 - 1.5\mathbf{v} + 0.5\left(\frac{\mathbf{v}}{\phi}\right)^3; & \mathbf{h} < \phi \\ 0, & \text{otherwise} \end{cases}$$

Corresponding variogram

nugget \longrightarrow $\tau^2 + \frac{\sigma^2}{2} \left(3 \frac{t}{\phi} - \left(\frac{t}{\phi} \right)^3 \right); \quad 0 \leq t \leq \phi$

sill \longrightarrow $\tau^2 + \sigma^2; \quad t > \phi \longleftarrow$ range

Geometric anisotropy

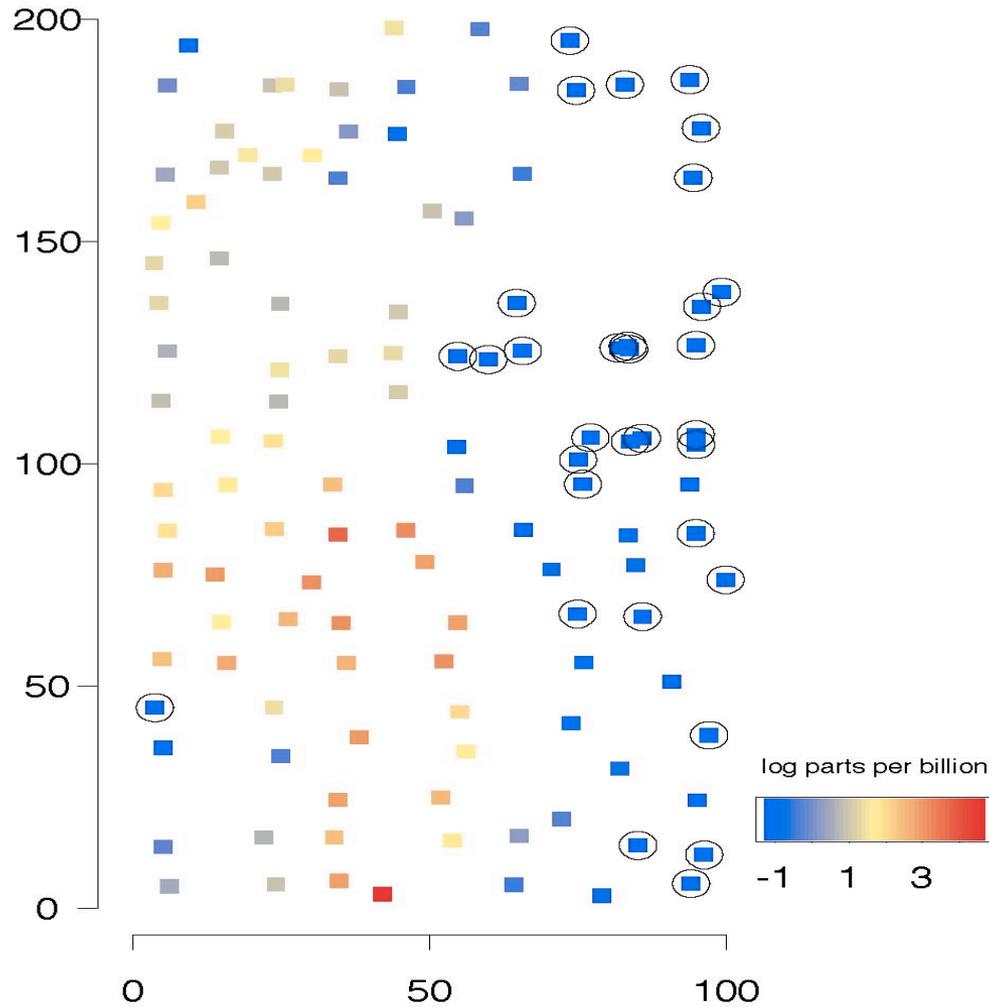
- **If** $C(x, y) = C(\|x - y\|)$ **we have an** *isotropic covariance* **(circular isocorrelation curves).**
- **If** $C(x, y) = C(\|Ax - Ay\|)$ **for a linear transformation** **A**, **we have** *geometric anisotropy* **(elliptical isocorrelation curves).**
- **General nonstationary correlation structures are typically** *locally geometrically anisotropic*.

Nonstationary spatial covariance:

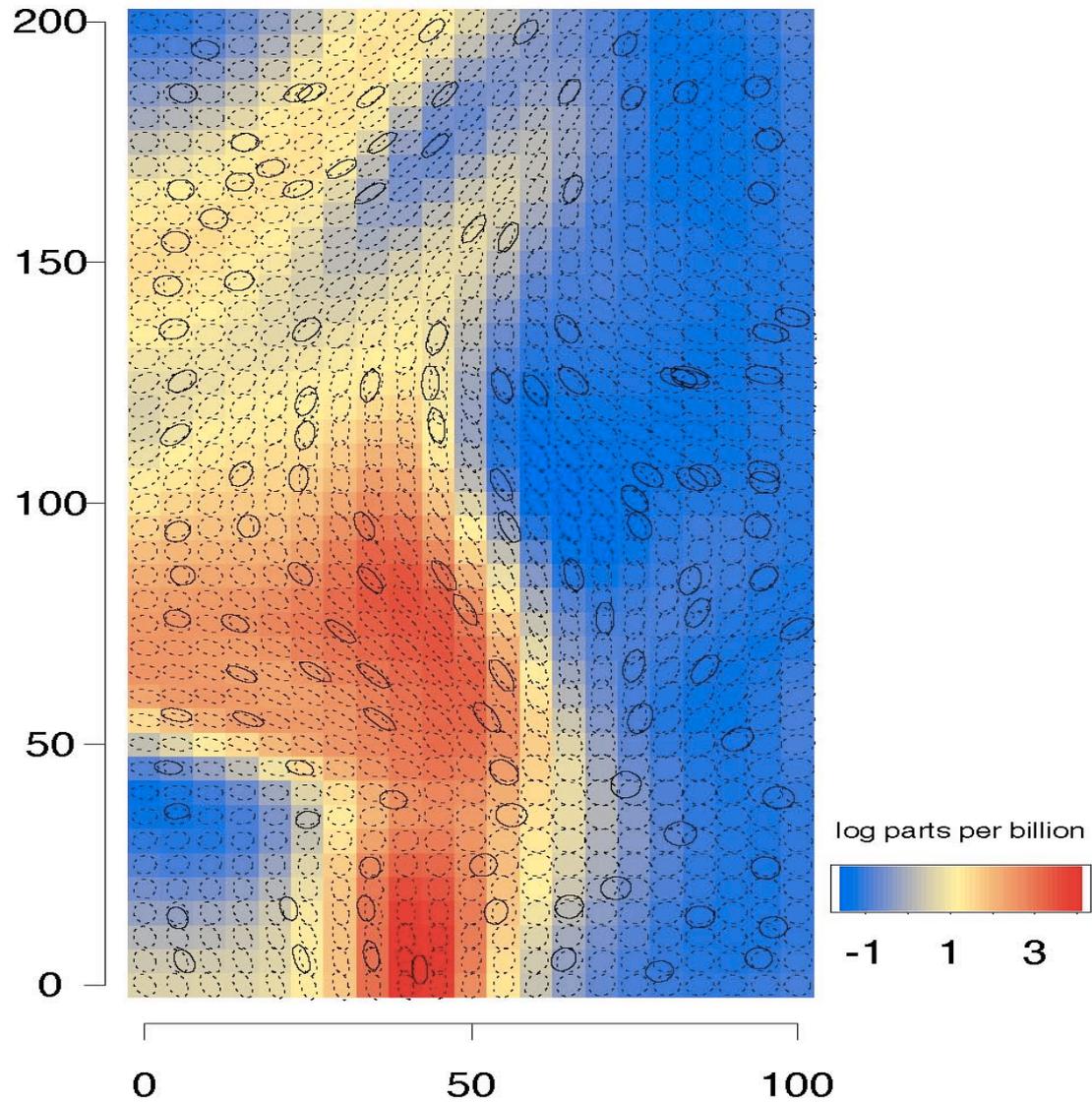
Basic idea: the parameters of a local variogram model---nugget, range, sill, and anisotropy---vary spatially.

Look at some pictures of applications from methodology publications.

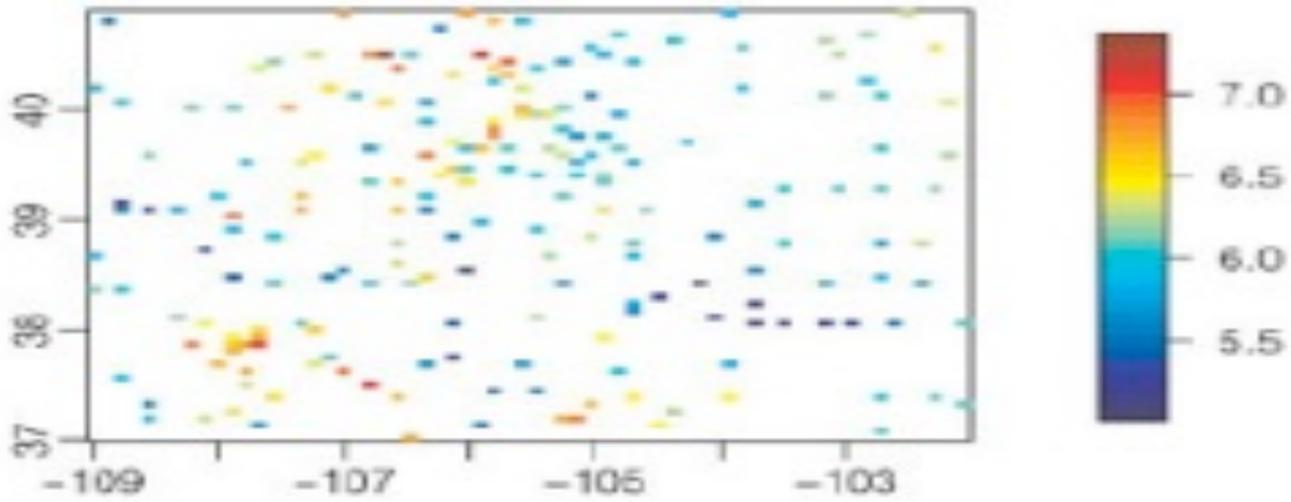
Swall & Higdon. Process convolution approach,
Soil contamination example --- Piazza Rd site.



Swall & Higdon. Process convolution approach,
Posterior mean and covariance kernel ellipses.



Paciorek & Schervish, 2006 –
Colorado 1981 annual precip (log)



Paciorek & Schervish, 2006 –
kernels (ellipses of constant Gaussian density) representing
estimated correlation structure

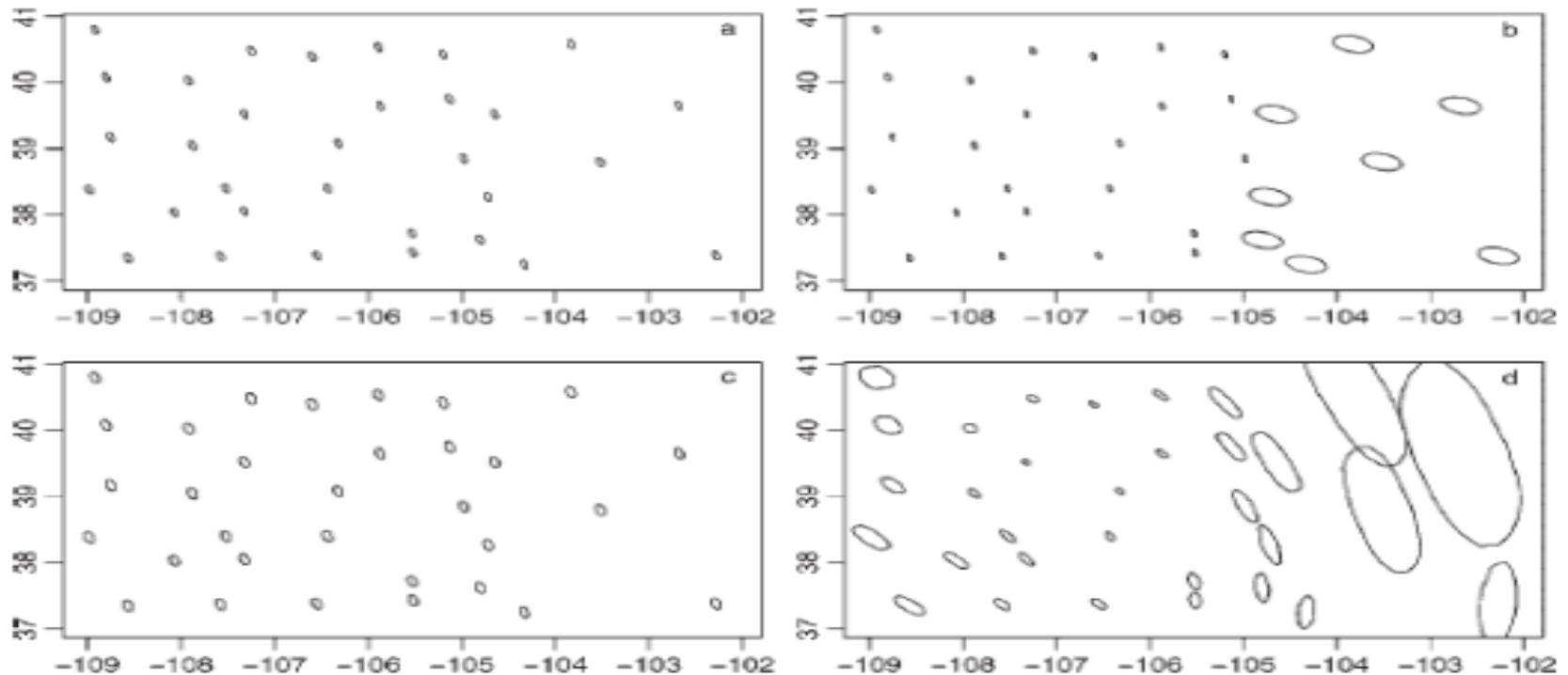


Figure 4. Kernels (ellipses of constant probability density of Gaussian densities) representing the estimated correlation structure for (a) stationary kriging, (b) nonstationary kriging based on two regions, (c) the fully Bayesian stationary GP model, and (d) the nonstationary GP model. For the Bayesian models, the ellipse-like figures are the posterior means of constant probability density ellipse values at a sequence of angles, $0, \dots, 2\pi$

The deformation idea

In the geometric anisotropic case, write

$$C(x, y) = C(\|f(x) - f(y)\|)$$

where $f(x) = Ax$. This suggests using a general nonlinear transformation

$$f : R^2 \rightarrow R^d$$

“G-plane” \rightarrow “D-space”

Usually $d = 2$ or 3 .

We do not want f to fold.

Remark: Originally introduced as a **multidimensional scaling problem**: find Euclidean representation with intersite distances monotone in spatial dispersion, $D(x, y)$

Space-time Model with Spatial Deformation

Damian et al., 2000 (*Environmetrics*), 2003 (*Journal of Geophysical Research*)

$$Z(x, t) = \mu(x, t) + \nu(x)^{1/2} H_t(x) + \varepsilon(x, t)$$

$\mu(x, t)$ spatio-temporal trend
parametric in time; mv spatial process

$\nu(x)$ temporal variance at x ,
log-normal spatial process

$\varepsilon(x, t)$ msmt error and short-scale variation
 $N(0, \sigma_\varepsilon^2)$, independent of $H_t(x)$

$H_t(x)$ mean 0, var 1, 2nd-order cont. spatial process
 $C(x, y) = \text{Cov}(H_t(x), H_t(y)) \xrightarrow{x \rightarrow y} 1$.

$$\text{Cov}(Z(x, t), Z(y, t)) = \begin{cases} \sqrt{\nu(x)\nu(y)}C(x, y) & x \neq y \\ \nu(x) + \sigma_\varepsilon^2 & x = y \end{cases}$$

Model (cont.)

$H_t(x)$ mean 0, var 1, 2nd-order cont. spatial process

$$\text{Cov}(H_t(x), H_t(y)) \xrightarrow{x \rightarrow y} 1.$$

$$\text{Cor}(H_t(x), H_t(y)) = \rho_\theta(\|f(x) - f(y)\|)$$

$f : G \rightarrow D$ smooth, bijective
(Geographic \rightarrow Deformed plane)

$\rho_\theta(d)$ isotropic correlation function
in a specified parametric family
(exponential, power exp, Matern)

i.e. The correlation structure of the spatial process is an (isotropic) function of Euclidean distances between site locations after a bijective transformation of the geographic coordinate system.

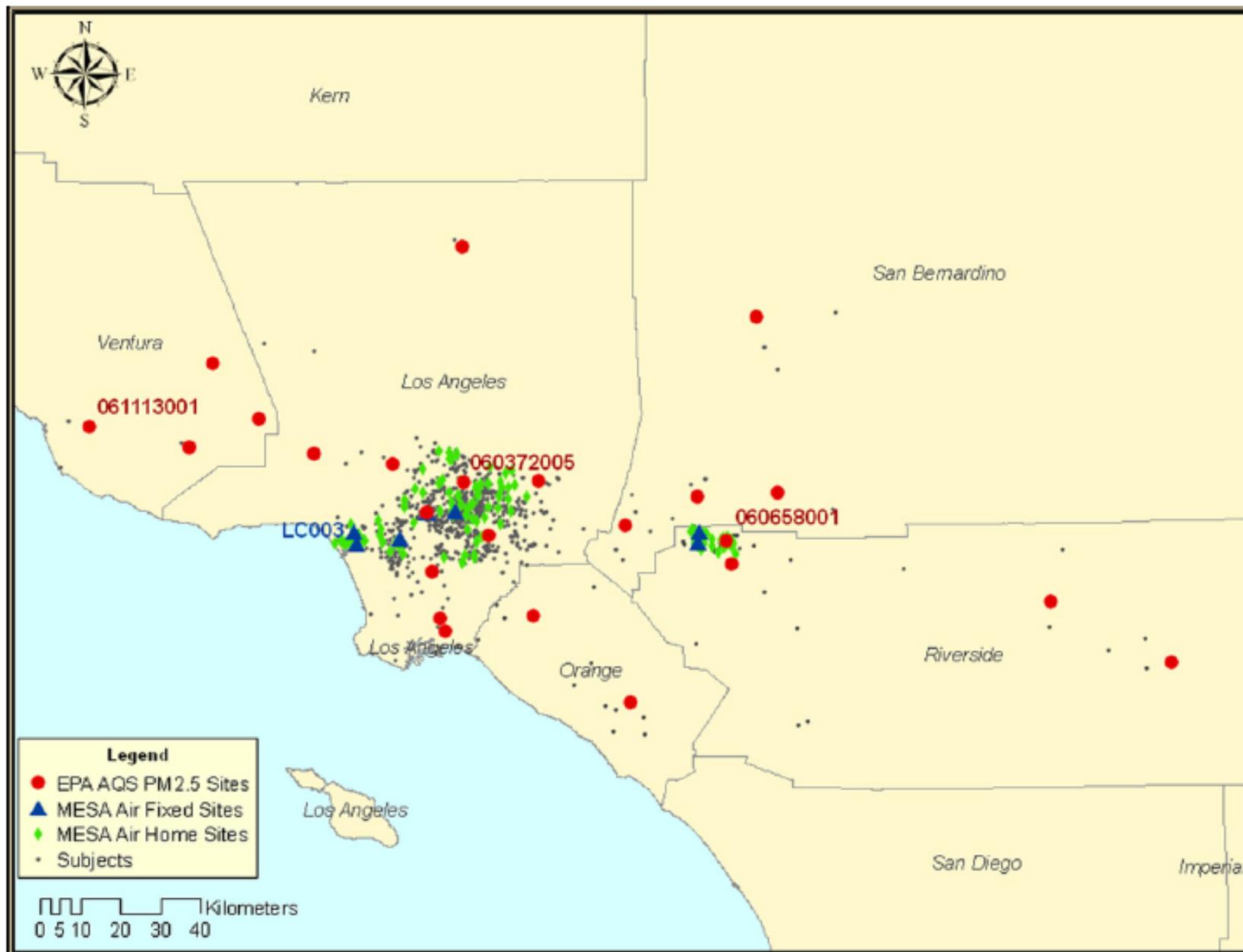
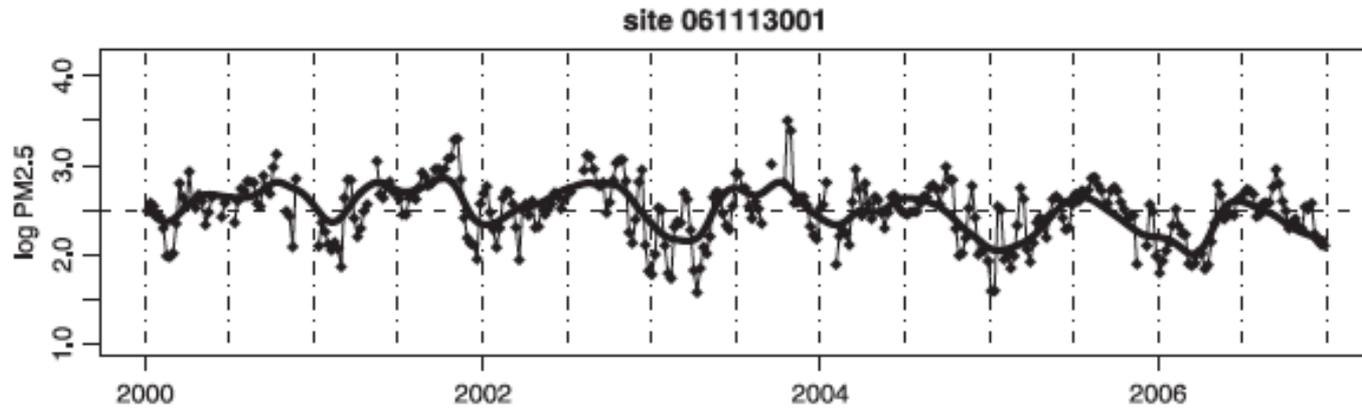
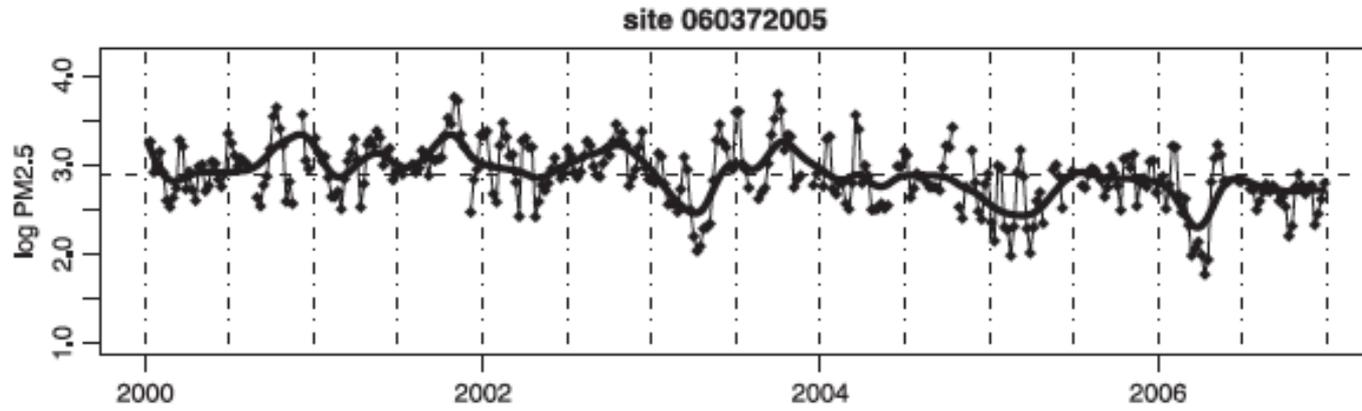
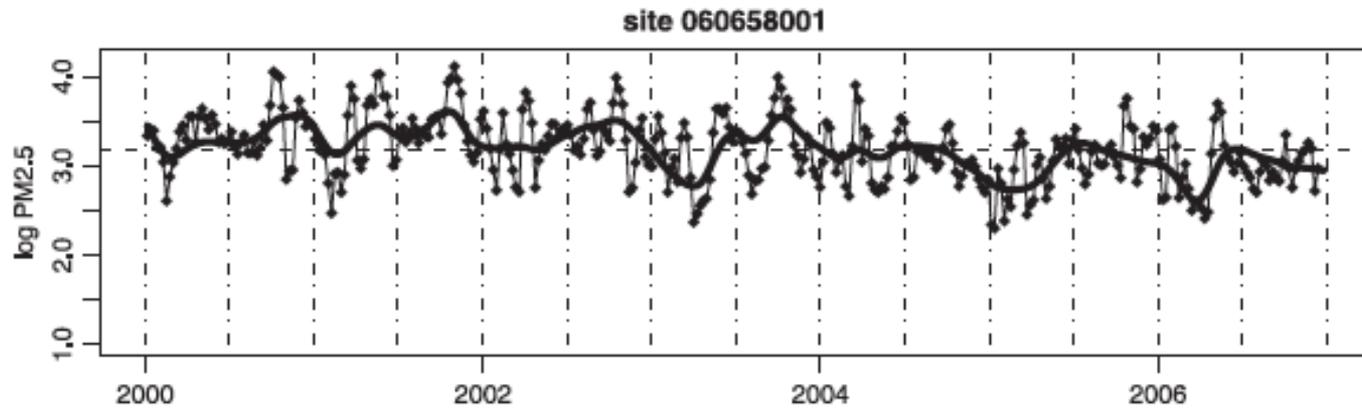


Fig. 1. Monitoring sites and subject home locations in the Los Angeles region.



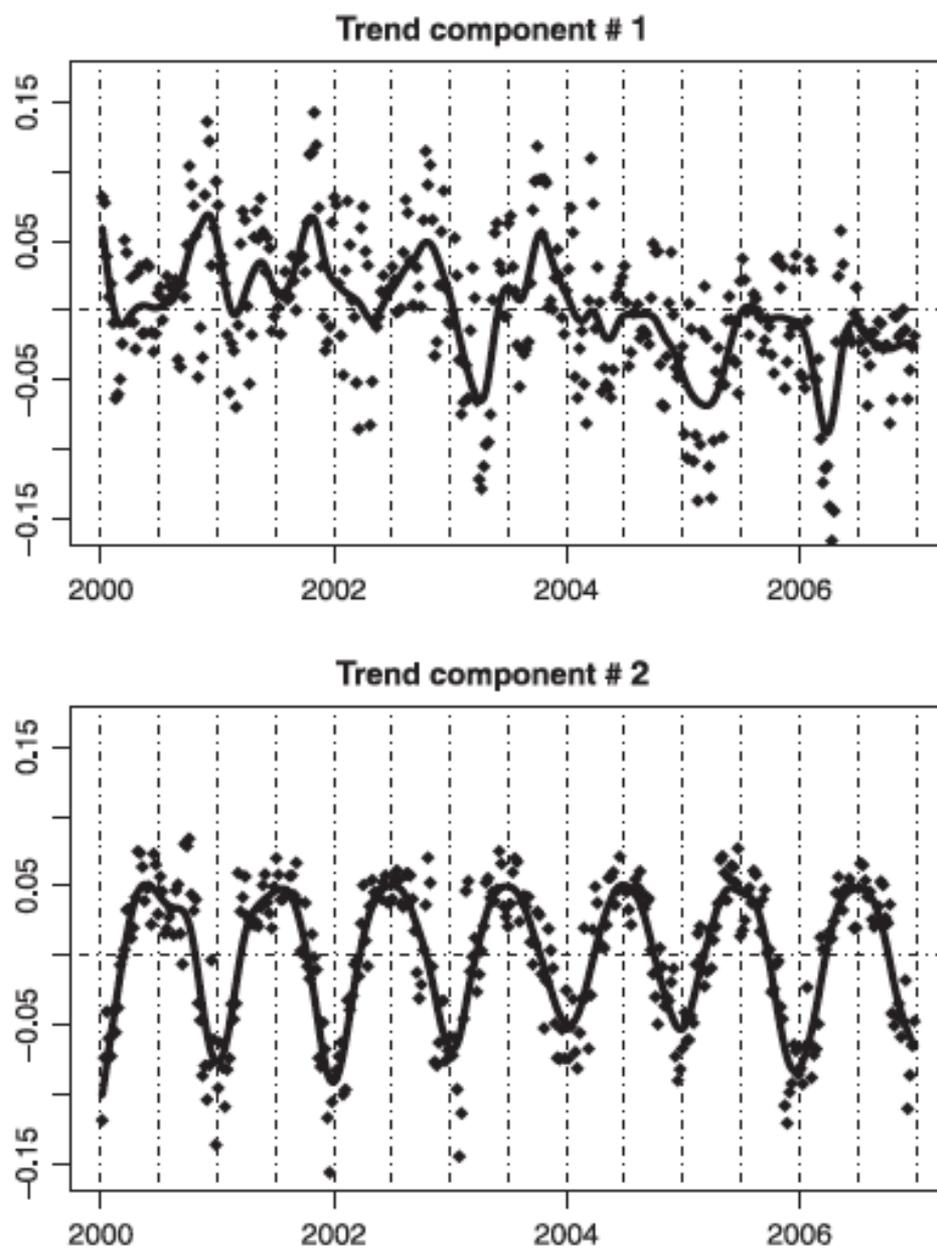


Fig. 4. Smoothed empirical orthogonal basis functions for log-transformed two-week average concentrations of PM_{2.5} in the Los Angeles area. The first smoothed component explains 27.5% of the variation in the matrix of log-transformed 2-week average PM_{2.5} concentrations while the second component explains only 9.4% of the variation.

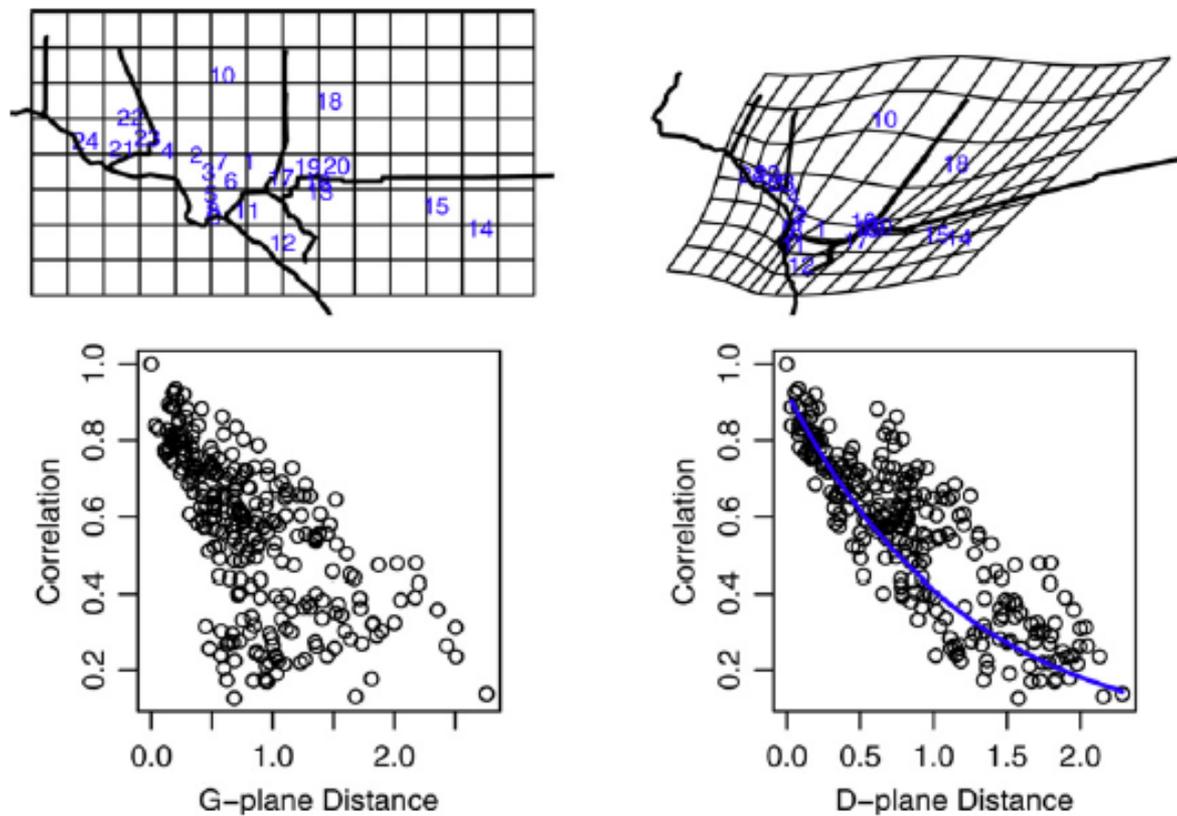
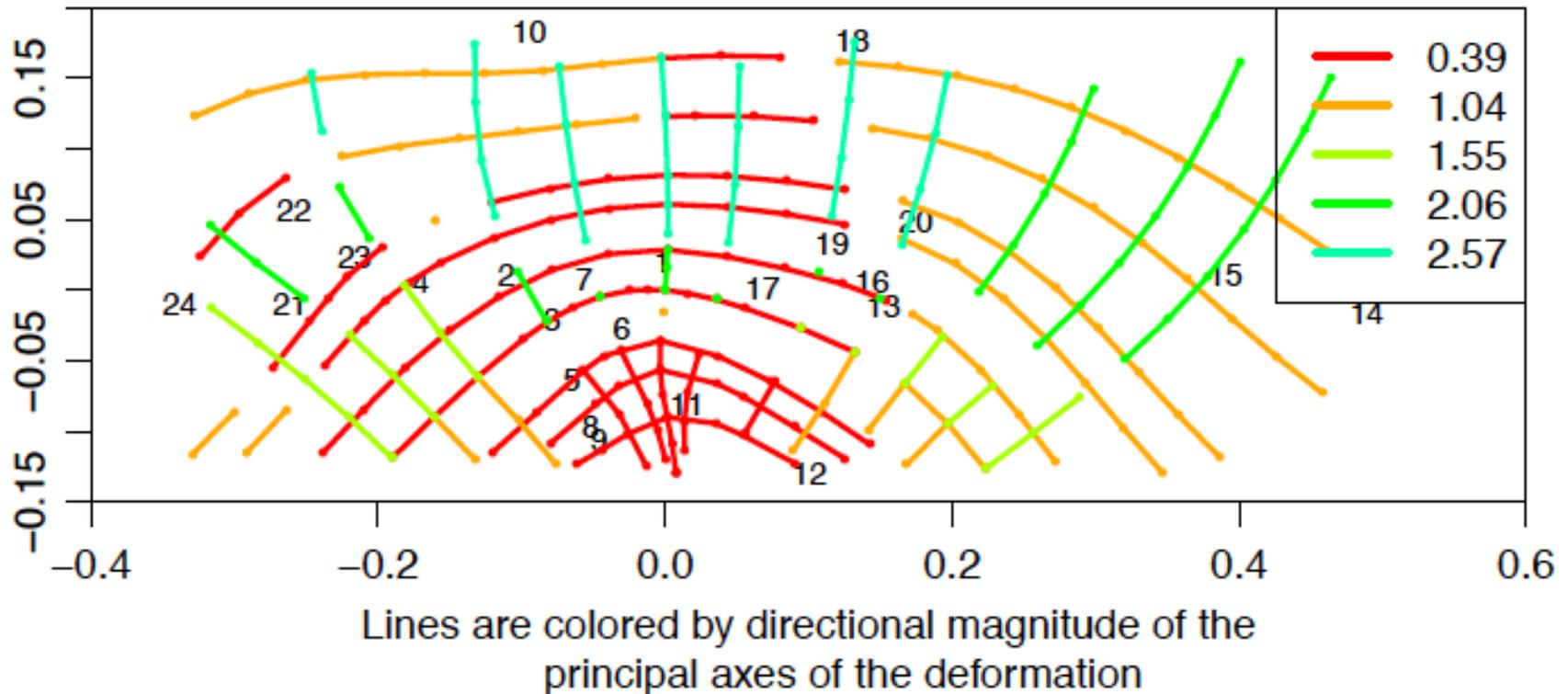


Fig. 7. Spatial structure of the spatio-temporal residuals before (left) and after (right) transformation using the Sampson–Guttorp method to account for nonstationarity.

An alternative to a gridded map of ellipsoids for local anisotropy is a “biorthogonal grid” which integrates the principal axes of the local affine derivative of the deformation.

Bgrid: CA PM2.5; 24 sites
lambda=2, warps all



Back to the model:

**The spatial deformation f encodes the nonstationarity:
spatially varying local anisotropy.**

We model this in terms of observation sites x_1, x_2, \dots, x_N as a pair of thin-plate splines:

$$f(x) = c + \mathbf{A}x + \mathbf{W}^T \sigma(x)$$

$c + \mathbf{A}x$ Linear part: global/large scale anisotropy $c_{2 \times 1}, \mathbf{A}_{2 \times 2}$

$\mathbf{W}^T \sigma(x)$ Non-linear part, *decomposable into components of varying spatial scale*: $\mathbf{W}_{N \times 2}, \sigma(x)_{N \times 1}$

$$\sigma(x) = \begin{bmatrix} \sigma(x - x_1) \\ \vdots \\ \sigma(x - x_N) \end{bmatrix} \quad \sigma(h) = \begin{cases} \|h\|^2 \log(\|h\|) & \|h\| > 0 \\ 0 & \|h\| = 0 \end{cases}$$

Lots of model parameters! $f : \{c, \mathbf{A}, \mathbf{W}\}, \mu, \theta, \sigma_\varepsilon^2, v : \{\tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2\}$

More on the **equations** of the thin-plate spline

$$f(x) = (f_1(x), f_2(x))^T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

minimizing "bending energy" subject to interpolation constraints

$$f_j(x_i) = \xi_{ij}, \quad 1 \leq i \leq N; \quad j = 1, 2,$$

is an equation of the form

$$f(x) = c + \mathbf{A}x + \mathbf{W}^T \sigma(x)$$

where the coefficients \mathbf{W} satisfy $\mathbf{1}^T \mathbf{W} = 0$, $\mathbf{X}^T \mathbf{W} = 0$.

I.e. the columns W_1 and W_2 of \mathbf{W} are vectors in the subspace spanned by $\{1, X_1, X_2\} : \mathbf{V} = \{v \in \mathbb{R}^N : v^T \mathbf{1} = 0, v^T X_1 = 0, v^T X_2 = 0\}$.

The system of equations for computation of a thin-plate spline is

$$\begin{bmatrix} \Xi \\ 0 \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{\mathbf{S}} & \mathbf{1} & \mathbf{X} \\ \mathbf{1}^T & 0 & 0 \\ \mathbf{X}^T & 0 & 0 \end{bmatrix}}_{\Gamma} \begin{bmatrix} \mathbf{W} \\ c^T \\ \mathbf{A}^T \end{bmatrix}, \quad \text{where } \tilde{\mathbf{S}} \text{ is } N \times N \text{ with elements}$$

$$\tilde{\mathbf{S}}_{ij} = \sigma(x_i - x_j), \quad \text{and the "bending energy" is } J(f) = \text{tr}(\mathbf{W}^T \tilde{\mathbf{S}} \mathbf{W})$$

Theoretical properties of the deformation model

Identifiability

Perrin and Meiring (1999): Let

$$D(x, y) = \gamma(\|f(x) - f(y)\|), \quad (x, y) \in R^n \times R^n$$

If (1) f and f^{-1} are differentiable in R^n

(2) $\gamma(u)$ is differentiable for $u > 0$

then (f, γ) is unique, up to a scaling for γ

and a homothetic transformation for f

(rotation, scaling, reflection)

Implementation 1. Weighted least squares

Consider observations at sites x_1, \dots, x_n . Let \hat{C}_{ij} be the empirical covariance between sites x_i and x_j .

Minimize

$$(\theta, f) \mapsto \sum_{i,j} w_{ij} \left(\hat{C}_{ij} - C(|f(x_i) - f(x_j)|; \theta) \right)^2 + \lambda J(f)$$

where $J(f)$ is a penalty for non-smooth transformations, such as the **bending energy**

$$J(f) = \iint \left[\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right] dx dy$$

When f is computed as a thin-plate spline, the minimization above can be considered in terms of the deformed coordinates, $\xi_i = f(x_i)$, or the parameters of the analytic representation of the thin-plate spline, $\{c, \mathbf{A}, \mathbf{W}\}$

Implementation 2. Bayesian

Likelihood:

$$L(S | \Sigma) = (2\pi |\Sigma|)^{-(T-1)/2} \exp \left\{ -\frac{T}{2} \text{tr} \Sigma^{-1} S \right\}$$

Nonlinear part: Bending energy Prior:

$$p(\mathbf{W}) \propto \exp \left(-\frac{1}{2\tau} \sum_{i=1}^2 \mathbf{w}_i' \tilde{\mathbf{S}} \mathbf{w}_i \right)$$

Linear part:

- fix two points in the G-D mapping
- put a (proper) prior on the remaining two parameters

Posterior computed using Metropolis-Hastings

Can get idea of reasonable values for τ parameter for the prior by simulating random deformations from the prior.

***** See closely related approach of Alex Schmidt using a Gaussian process prior.**

Computation

Metropolis-Hastings algorithm for sampling from the highly multidimensional posterior. (Naïve implementation not very well behaved due to correlation among a very large number of parameters.)

Given estimates of D-plane locations, $f(x_i)$, the transformation is extrapolated to the whole domain using thin-plate splines. (Visualization and diagnostics.)

Predictive distributions for

- (a) temporal variance at unobserved sites,**
- (b) the spatial covariance for pairs of observed and/or unobserved sites,**
- (c) the observation process at unobserved sites.**

Problems with the WLS and Bayesian computational approaches

There are serious practical problems with the approaches to deformation mapping presented here.

- They are computationally intensive, involving constrained or regularized optimization of approximately 2 parameters per spatial monitoring site. Large problems are not practical.
- Whether parameterized in terms of the coefficients W of the radial basis functions ($d^2 \log d$), or the coordinates of the D-plane representation,
 - the WLS or likelihood objective functions are likely to have multiple local maxima,
 - in the case of Bayesian estimation by MCMC, the parameters are highly correlated, making convergence of the Markov Chain problematic

Implementation 3. Reduced rank thin-plate spline mappings via partial warps

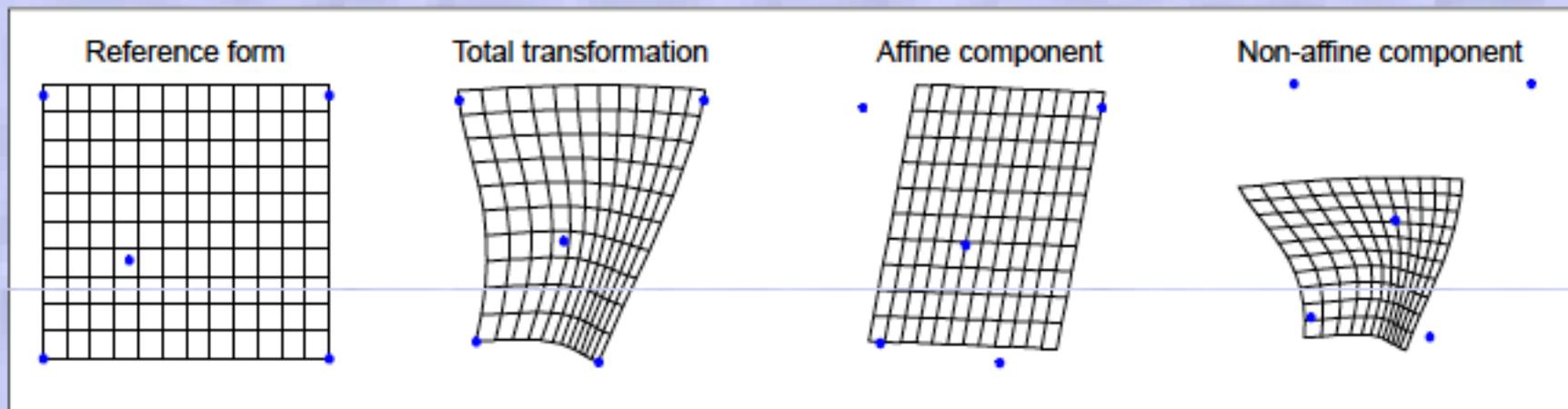
A more efficient and practical approach is to

- reparameterize the spline in terms of coefficients of a set of *orthogonal spatial basis functions*
- reduce the dimension of the problem by selecting/fitting a subset of the basis functions. We do this using an L1 penalty (instead of the TPS bending energy).

Thin-plate spline deformations were introduced in *morphometrics* (shape analysis) by Bookstein (1986), where he also proposed the decomposition of deformations (warps) according to “principal warps” derived from eigenvectors of the bending energy matrix.

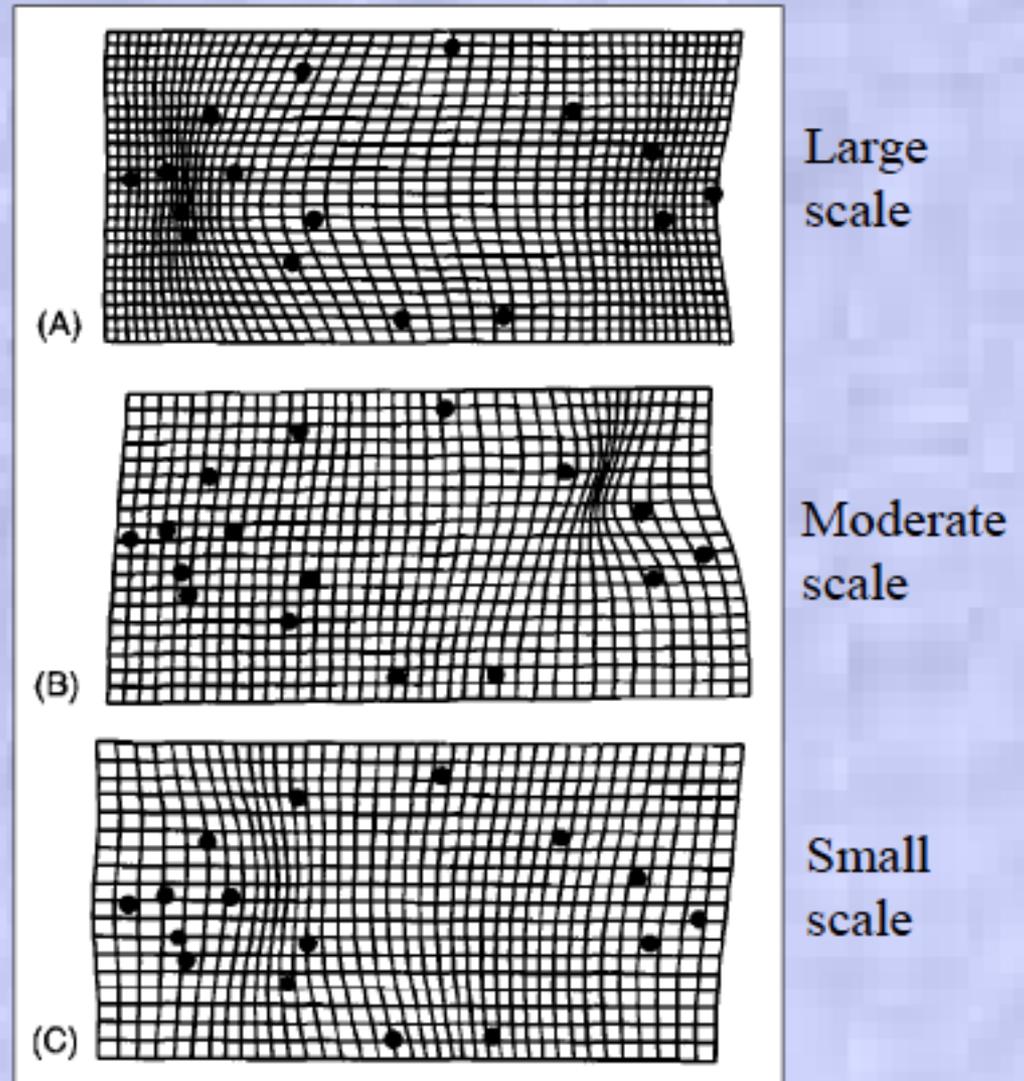
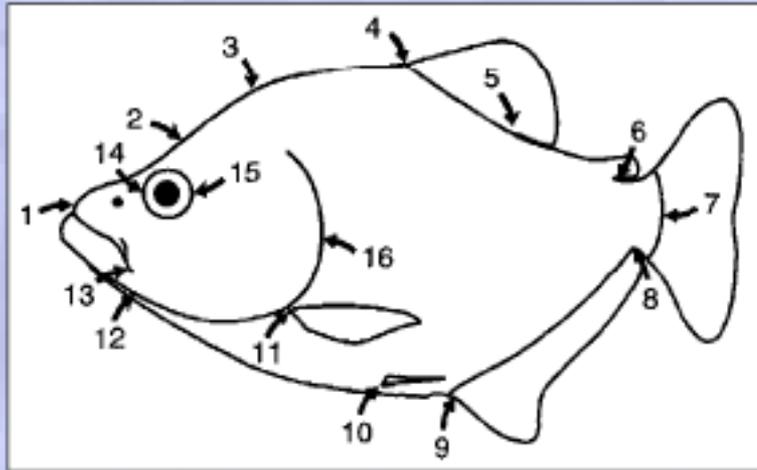
Partial warps

- Recall: thin-plate spline decomposes shape difference into *global* and *local* components:
 - *Uniform, affine component* is a tilted plane viewed in perspective.
 - *Non-uniform, non-affine component* characterizes regional deformations (warping of the thin plate).
 - Characterized by bending-energy matrix.
 - *Total deformation* is sum of the two components.



- *Non-uniform* portion of deformation:
 - Describes changes that vary in nature and extent across the organism.
 - Can be further decomposed into set of *orthogonal* components: *partial warps*.
 - Eigenvectors of bending-energy matrix.
 - Characterize changes at progressively smaller, more localized spatial scales:
 - 1st partial warp describes change at largest scale (lowest bending energy).
 - 2nd partial warp describes change at smaller scale.
 - Etc.
 - For k landmarks, can calculate $k-3$ partial warps.

- Partial warps 1–3 for Fink's ontogenetic piranha data:



Recall the algebra of thin-plate splines, driven by the the matrix $\tilde{\mathcal{J}}$ containing the terms $\sigma(x_i - x_j) = (x_i - x_j)^2 \log(x_i - x_j)$

Partial warps can be computed as follows:

1. Compute the upper $n \times n$ component of the inverse of the “ Γ ” matrix of the system of linear equations of the thin-plate spline. This is the bending energy matrix B .
1. Compute the eigenvectors of B , the principal warps, $g_j, j = 1, \dots, n$
1. Partial warps are linear combinations of of these (spatial) eigenvectors, $\sum \beta_j g_j$, where β_j is a 2-vector of coefficients for the elements of the 2D mapping.
2. The image coordinates for the thin-plate spline in terms of partial warps is $Y = c + Ax + (\sum \beta_j g_j) \sigma(x)$
 where $\sigma(x) = (\sigma(x - x_1), \dots, \sigma(x - x_n))'$

The β s replace the coefficients W in the previously introduced **equations** of the thin-plate spline: (these eqns were shown above)

$$f(x) = (f_1(x), f_2(x))^T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

minimizing "bending energy" subject to interpolation constraints

$$f_j(x_i) = \xi_{ij}, 1 \leq i \leq N; j = 1, 2,$$

is an equation of the form

$$f(x) = c + \mathbf{A}x + \mathbf{W}^T \sigma(x)$$

where the coefficients \mathbf{W} satisfy $\mathbf{1}^T \mathbf{W} = 0$, $\mathbf{X}^T \mathbf{W} = 0$.

I.e. the columns W_1 and W_2 of \mathbf{W} are vectors in the subspace

spanned by $\{1, X_1, X_2\}$: $\mathbf{V} = \{v \in \mathbb{R}^N : v^T \mathbf{1} = 0, v^T X_1 = 0, v^T X_2 = 0\}$.

The system of equations for computation of a thin-plate spline is

$$\begin{bmatrix} \Xi \\ 0 \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{\mathbf{S}} & \mathbf{1} & \mathbf{X} \\ \mathbf{1}^T & 0 & 0 \\ \mathbf{X}^T & 0 & 0 \end{bmatrix}}_{\Gamma} \begin{bmatrix} \mathbf{W} \\ c^T \\ \mathbf{A}^T \end{bmatrix}, \text{ where } \tilde{\mathbf{S}} \text{ is } N \times N \text{ with elements}$$

$\tilde{\mathbf{S}}_{ij} = \sigma(x_i - x_j)$, and the "bending energy" is $J(f) = \text{tr}(\mathbf{W}^T \tilde{\mathbf{S}} \mathbf{W})$

Implementation 3. Reduced rank thin-plate spline mappings via partial warps

$$L(S, \Sigma) = (2\pi|\Sigma|)^{-(T-1)/2} \exp\left\{-\frac{T}{2} \text{tr}(\Sigma^{-1}S)\right\}$$

where Σ is a function of the parameters c , A , and β in the equation for the TPS:

$$Y = c + Ax + \left(\sum \beta_j g_j\right) \sigma(x)$$

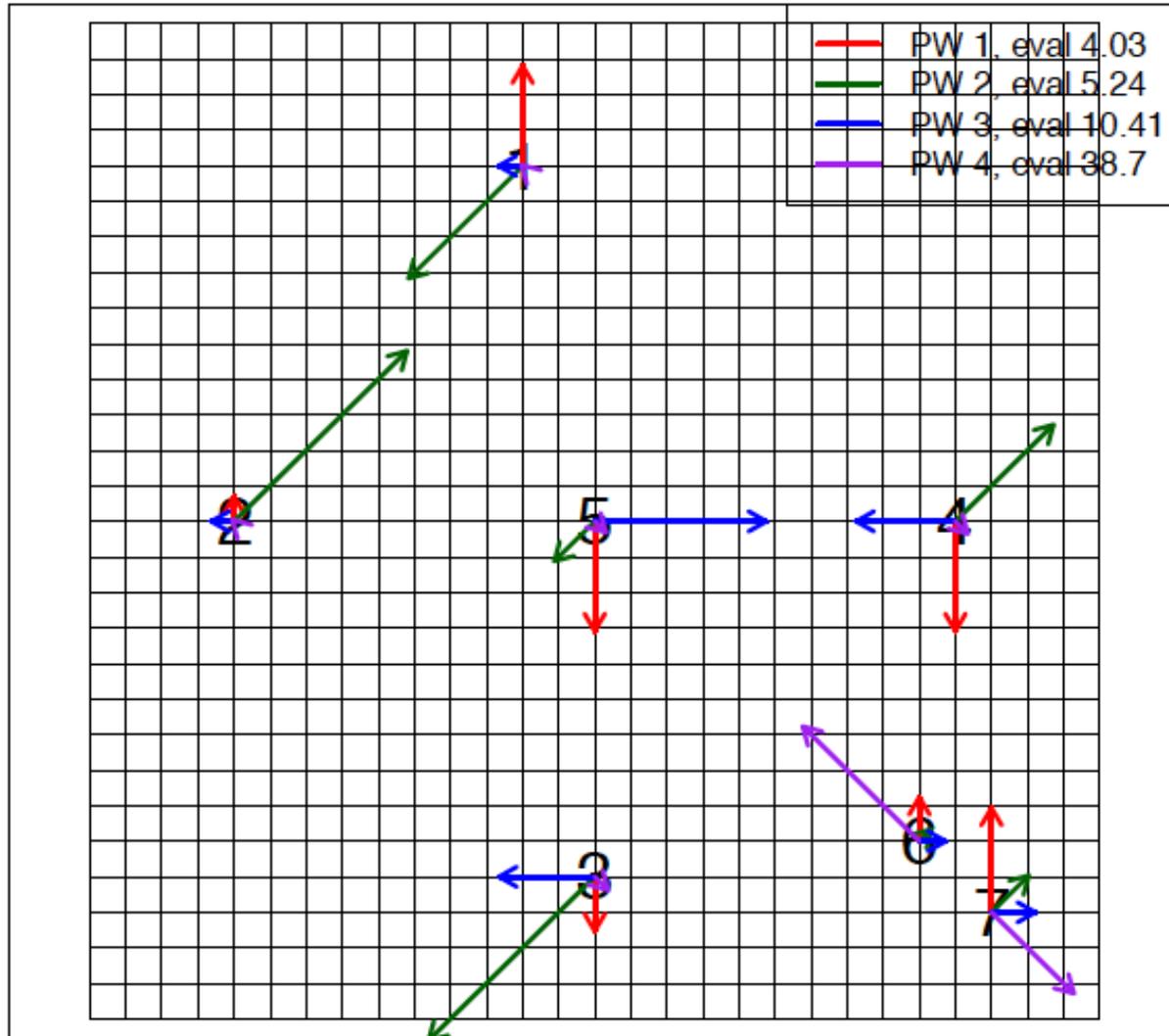
We optimize $L(S, \Sigma) + \lambda \|\beta\|_1$.

We effectively reduce the dimensionality of the solution by removing partial warps corresponding to eigenvectors g_j with coefficients shrunk to zero.

Following are a series of plots to illustrate

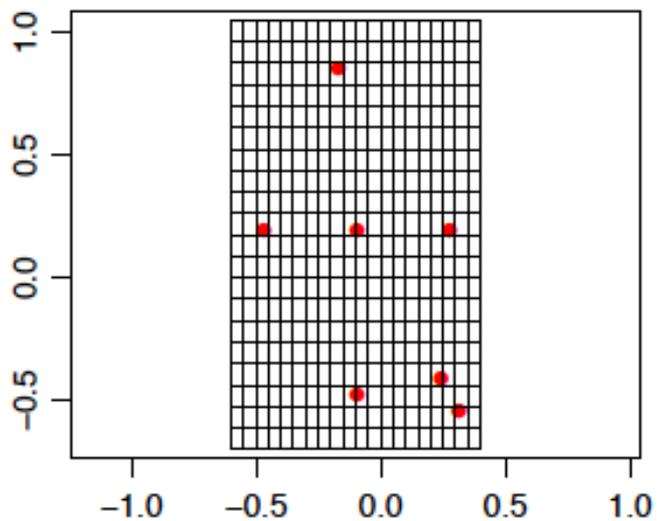
- **the definition of the eigenvectors of the bending energy matrix for a configuration of 7 points**
- **Affine and partial warps corresponding to the above eigenvectors, with each warp illustrated**
 - **for deformations in the ‘x’ and ‘y’ directions separately, and**
 - **for two different coefficient multipliers (‘scale’)**

Principal Warps: Bending energy matrix eigenvectors

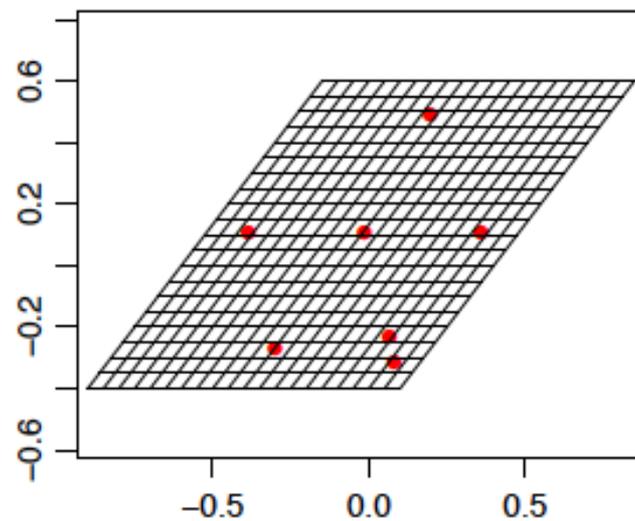


Evect 1: 0.45, 0.11, -0.24, -0.49, -0.49, 0.2, 0.46
Evect 2: -0.4, 0.6, -0.58, 0.34, -0.14, 0.04, 0.13
Evect 3: -0.11, -0.1, -0.42, -0.43, 0.75, 0.11, 0.19
Evect 4: -0.01, -0.01, 0.09, 0.09, 0.08, -0.81, 0.57

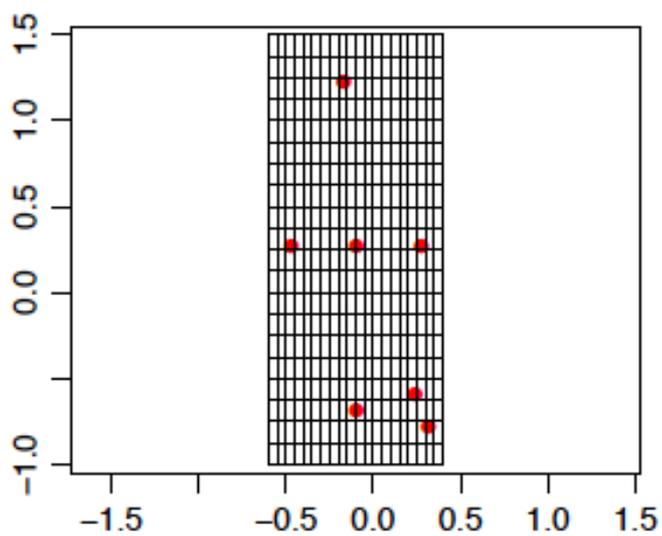
Affine warp 'y'



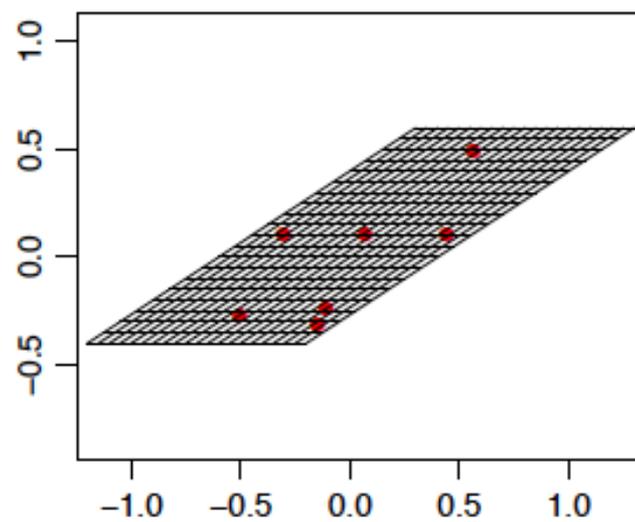
Affine warp 'x'



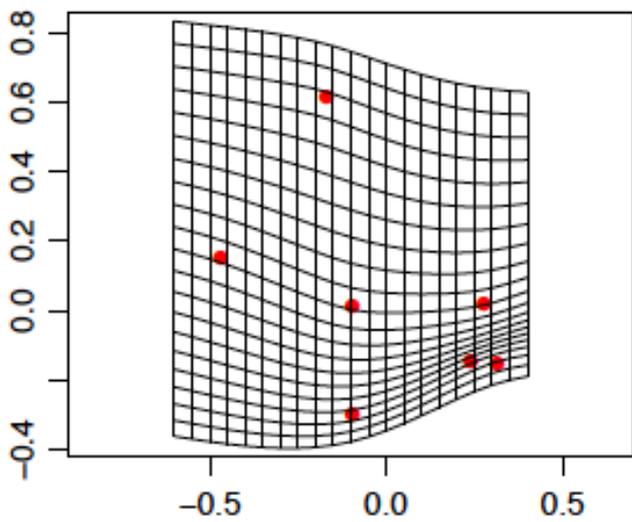
Affine warp 'y', scale 2



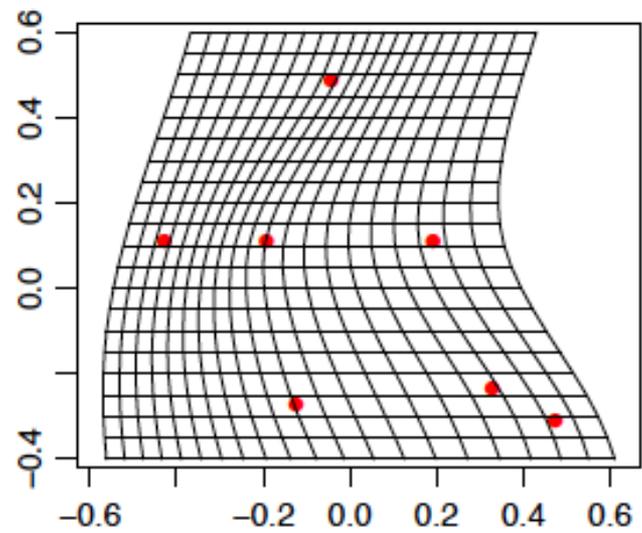
Affine warp 'x', scale 2



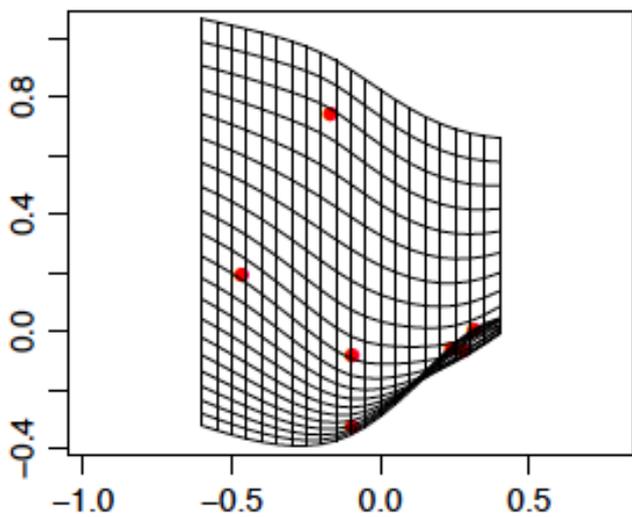
PW 1 'y'



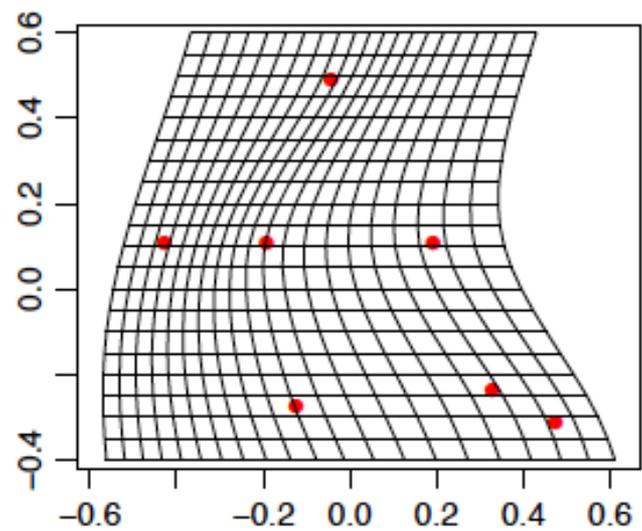
PW 1 'x'



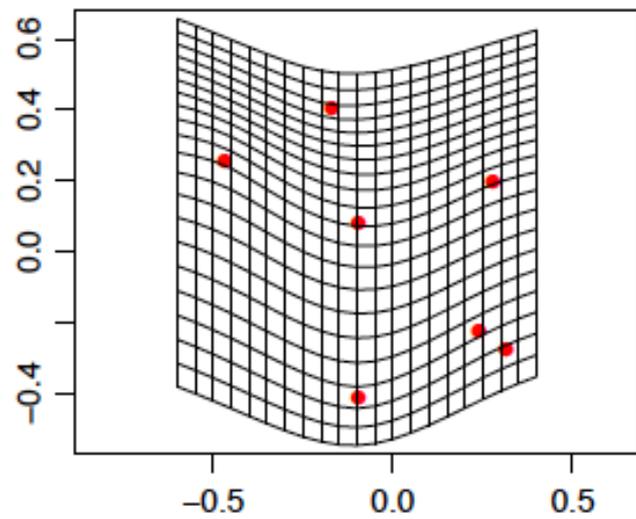
PW 1 'y', scale 2



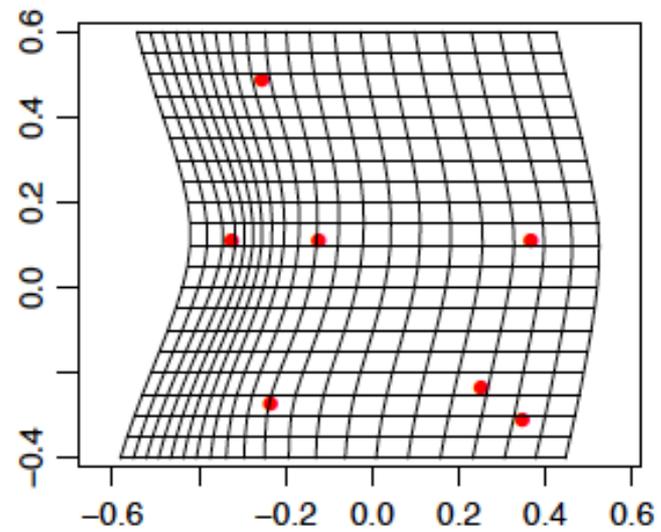
PW 1 'x', scale 2



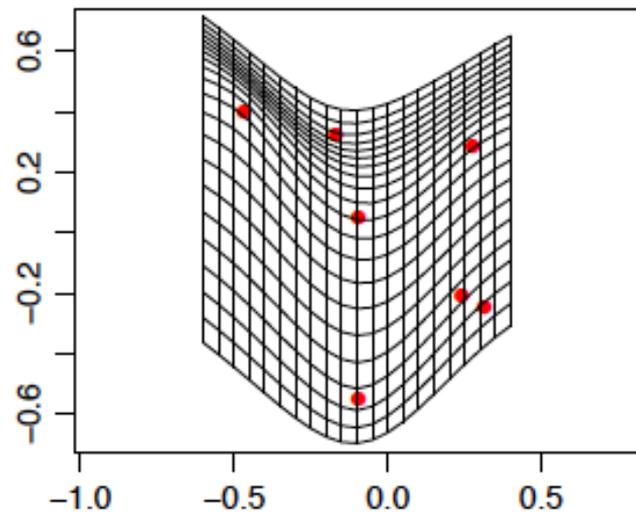
PW 2 'y'



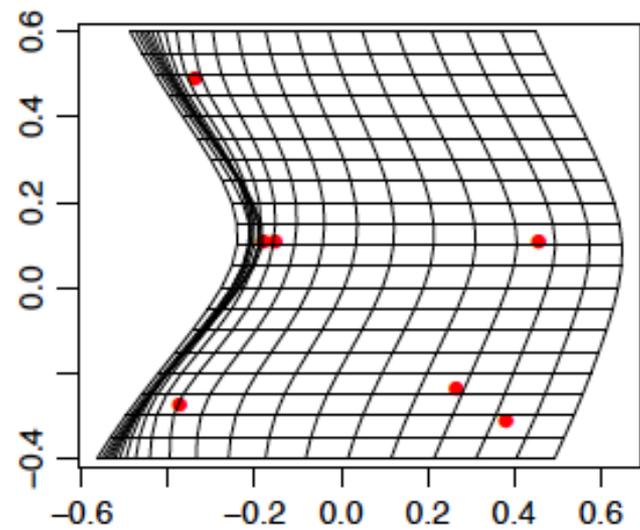
PW 2 'x'



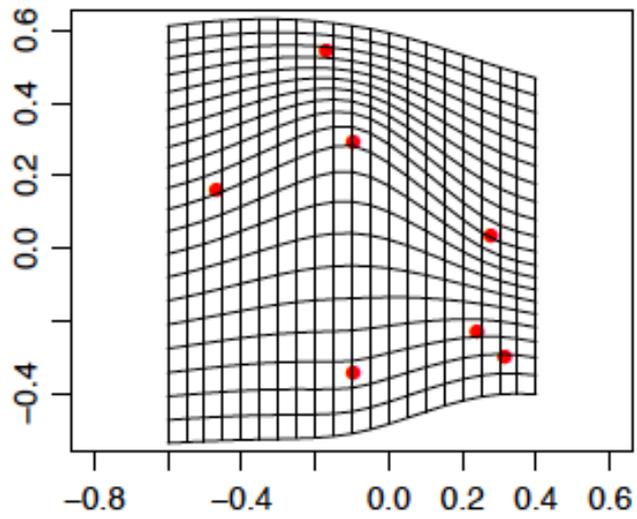
PW 2 'y', scale 2



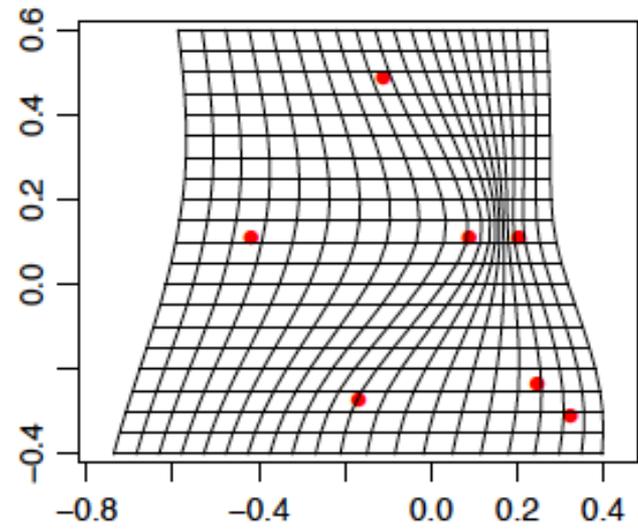
PW 2 'x', scale 2



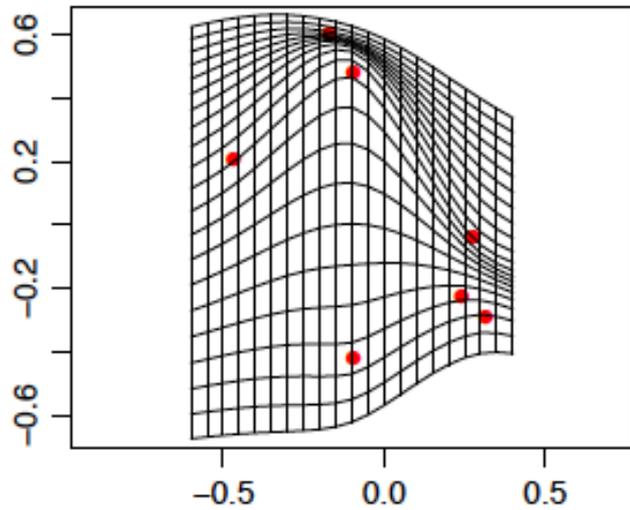
PW 3 'y'



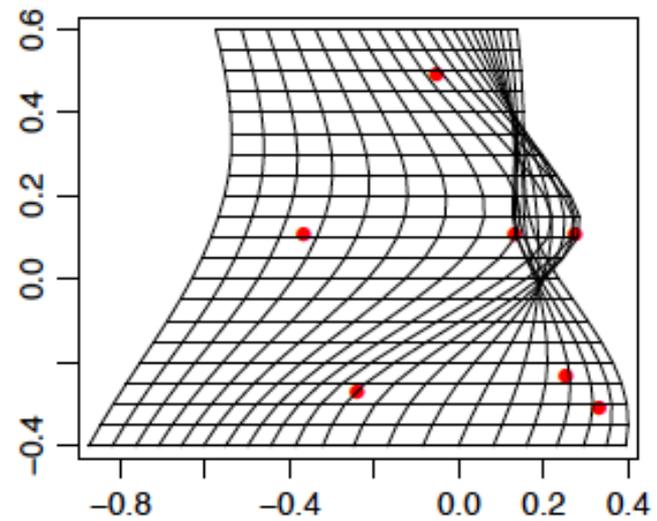
PW 3 'x'



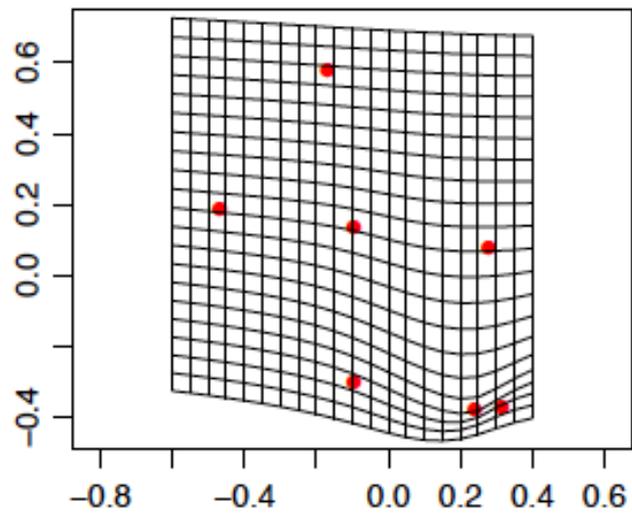
PW 3 'y', scale 2



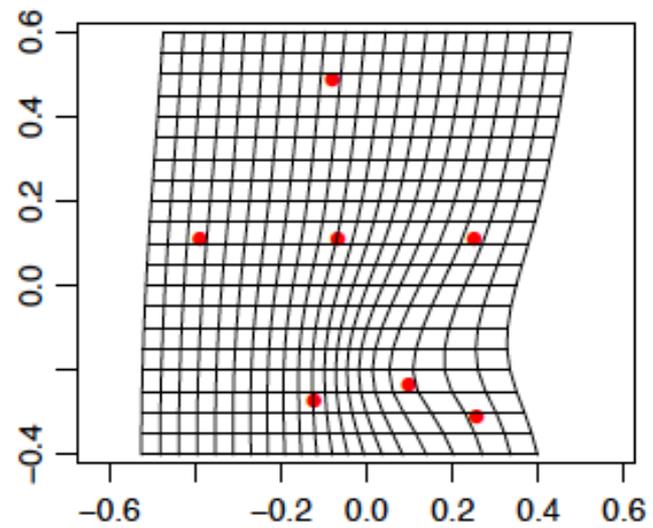
PW 3 'x', scale 2



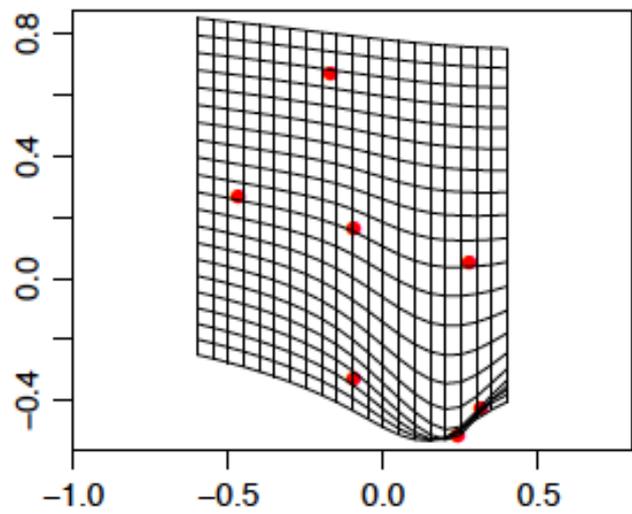
PW 4 'y'



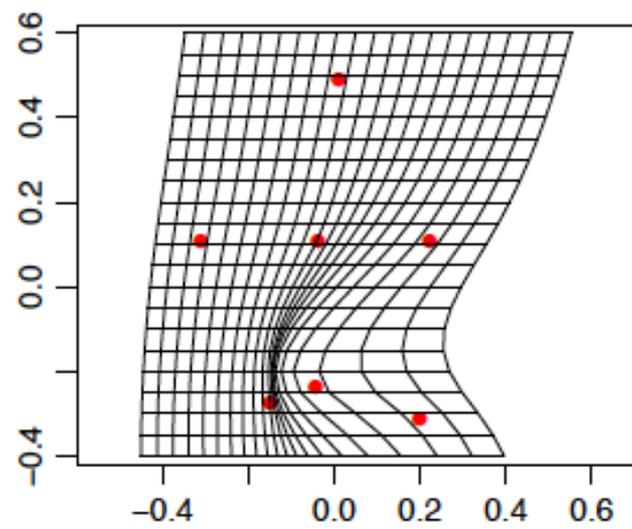
PW 4 'x'



PW 4 'y', scale 2

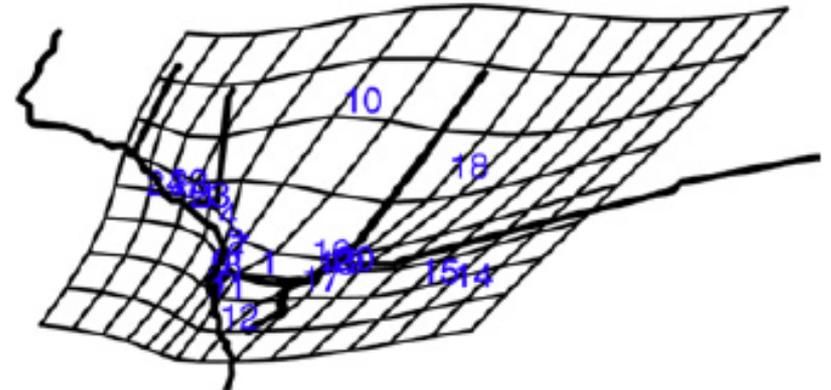
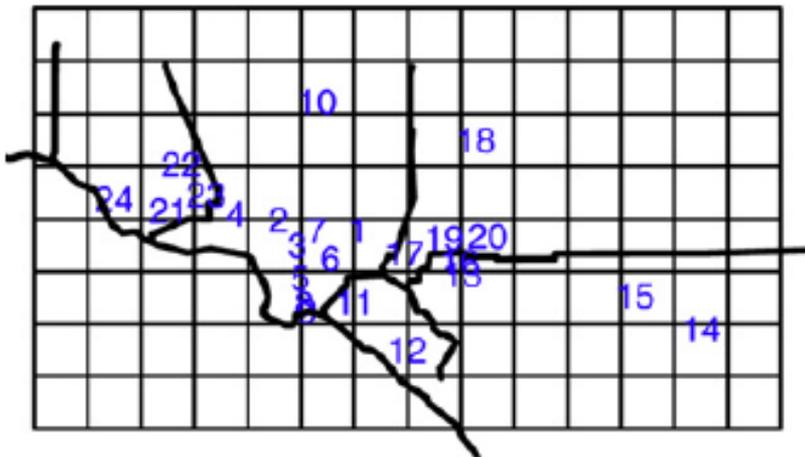


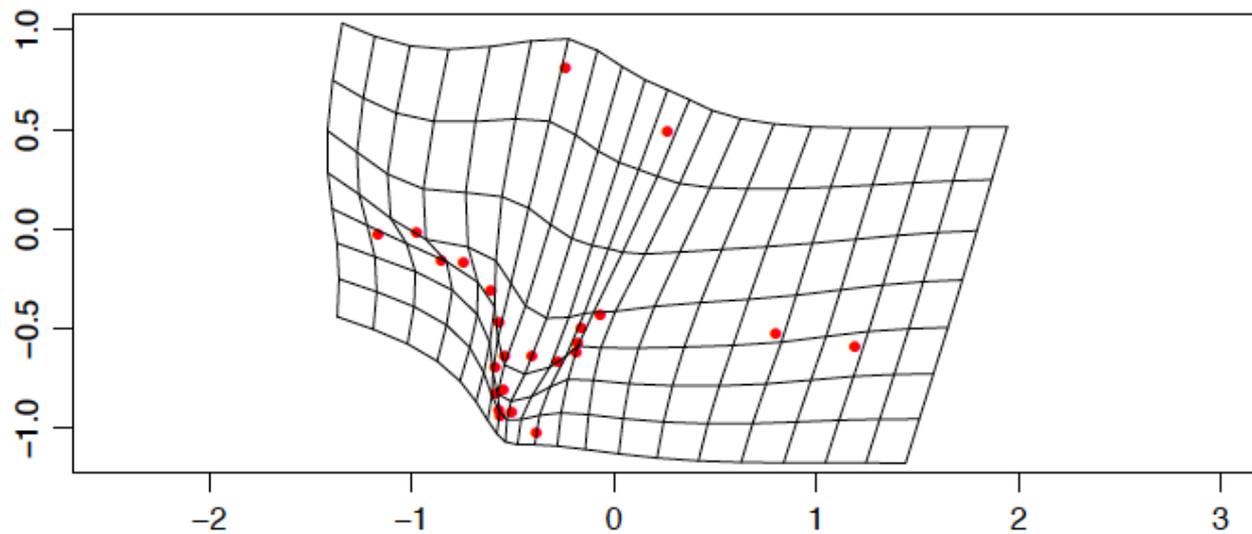
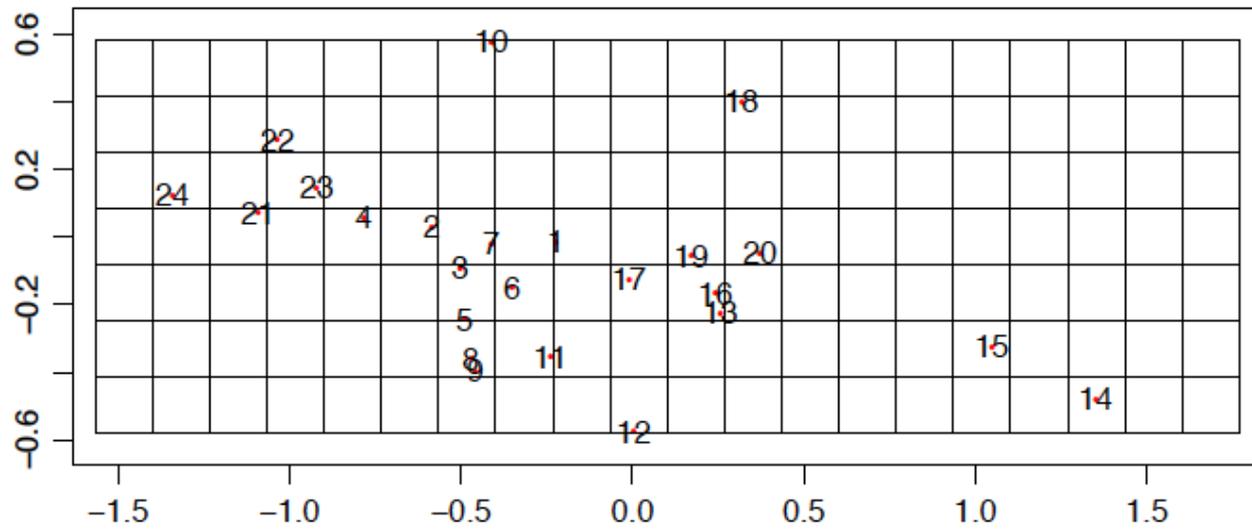
PW 4 'x', scale 2

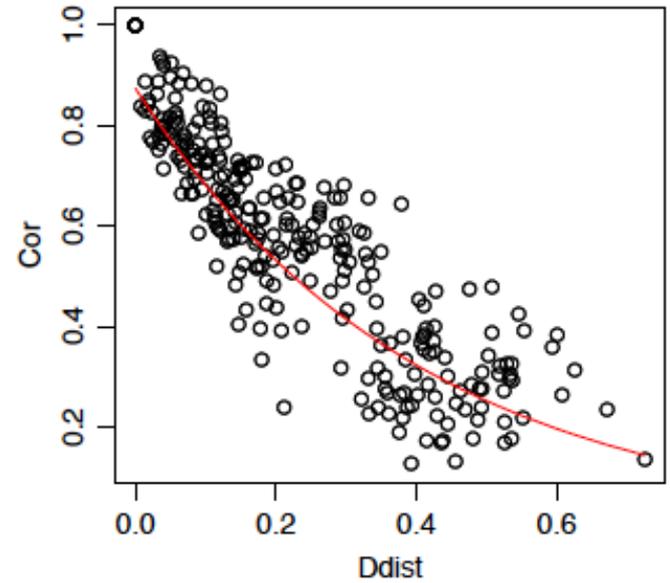
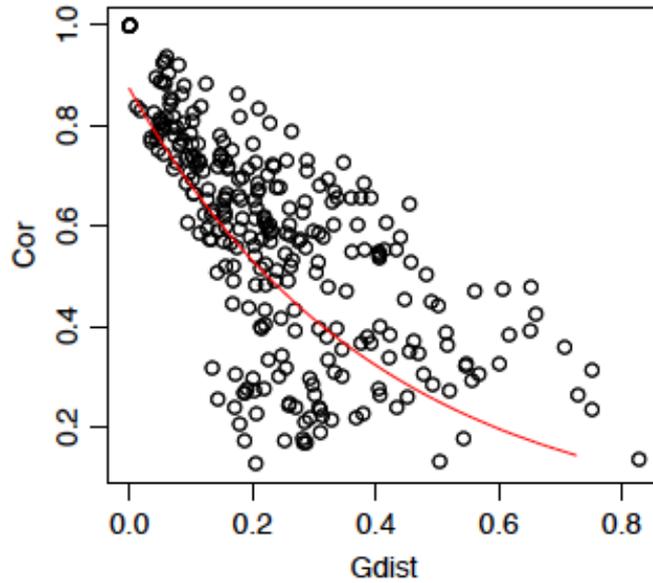
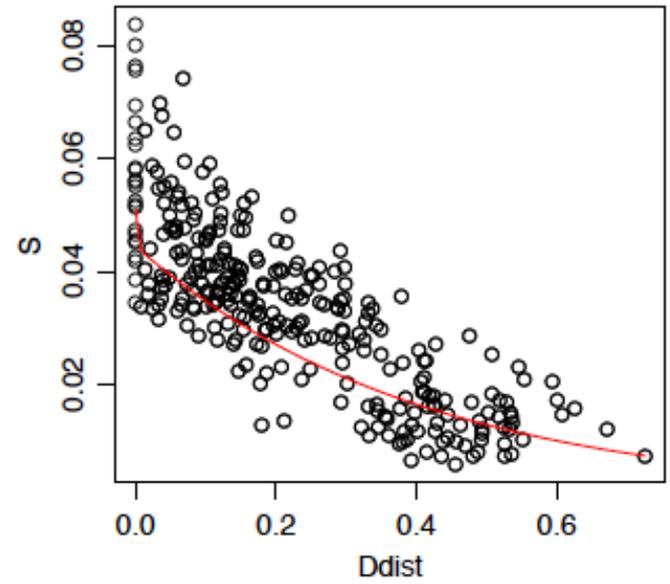
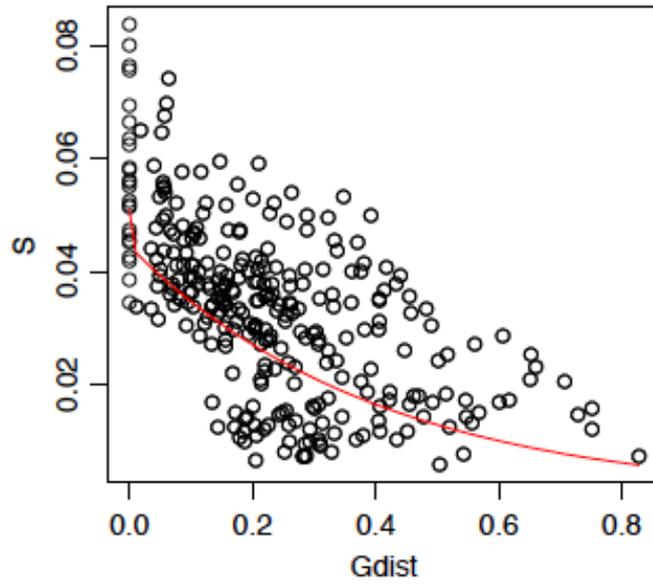


Return to the application to PM2.5 data at 24 sites in the region of southern CA around Los Angeles and Riverside. Analysis based on time series of about 150 2-week average concentrations from 2000 through 2006.

We illustrate below the fitted deformation and spatial correlation function based on maximum likelihood with an L1 constraint chosen 'by eye'. First, the fit in the published paper computed (with great effort!) by the Bayesian algorithm



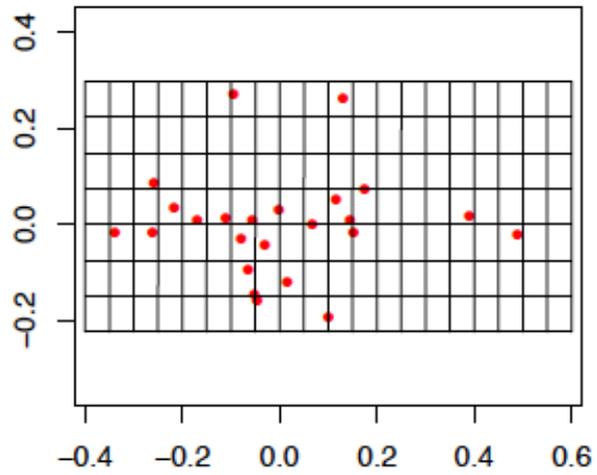




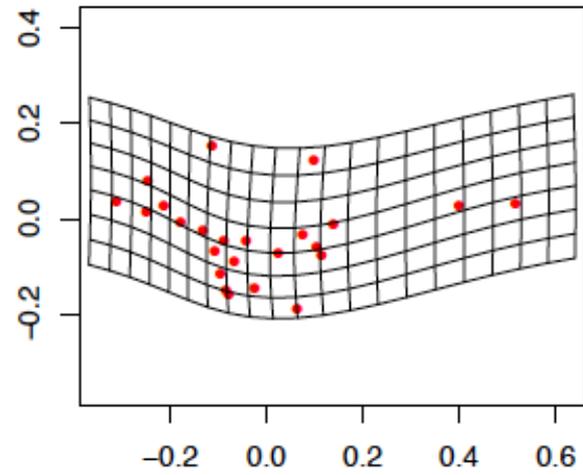
**Covariance (top) and
Correlation (bottom)
vs.
G-plane dist (left) and
D-plane dist (right)**

Examine the decomposition of the fitted deformation in terms of partial warps and the effect of the L1 penalty in zeroing out any contributions from all the higher bending energy (smaller spatial scale) warps.

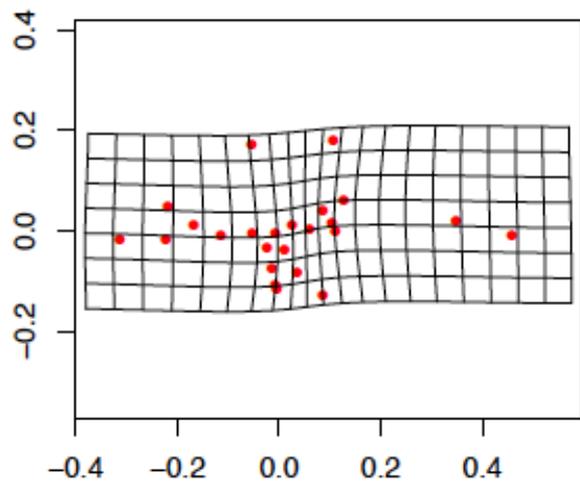
Affine, coefs = 0,0.49



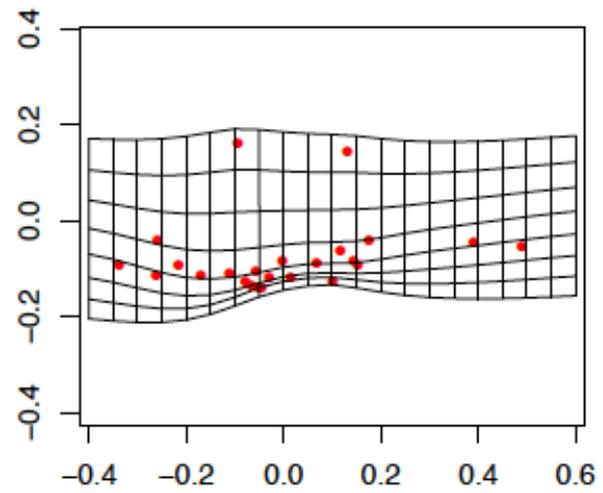
PW 2 0.21,0.34



PW 3 -0.69,0.16



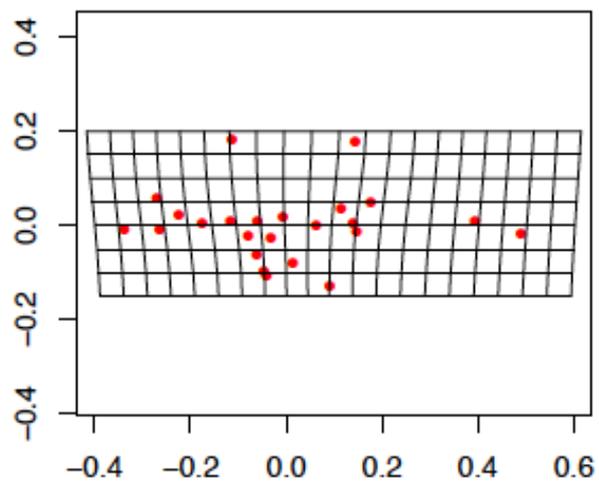
PW 4 0,-2.73



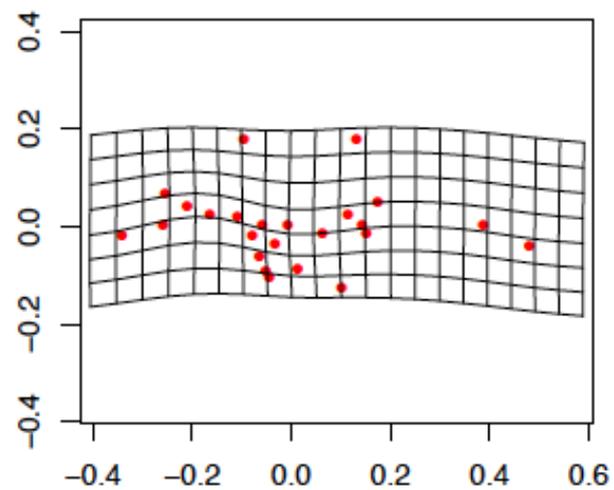
Among work to be considered:

- 1. Work to be done to facilitate choice of parameter for the L1 penalty, possibly in a Bayesian framework.**
- 2. Incorporate this deformation model in a full spatio-temporal model with mean structure.**
- 3. Further investigate and demonstrate the application to spatial only problems.**
- 4. Incorporate covariate in the partial warp modeling.**

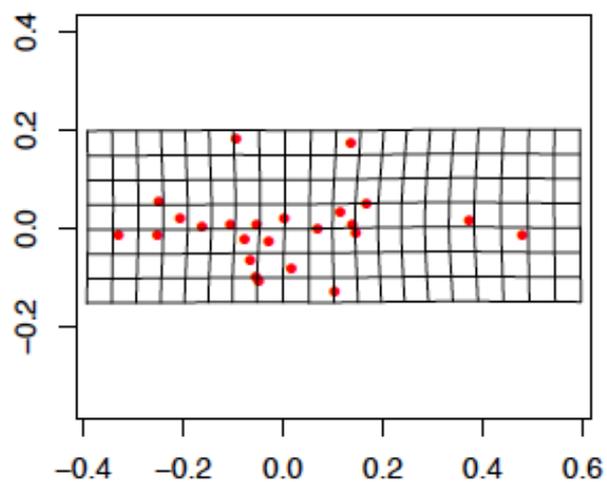
PW 5 -0.56,0



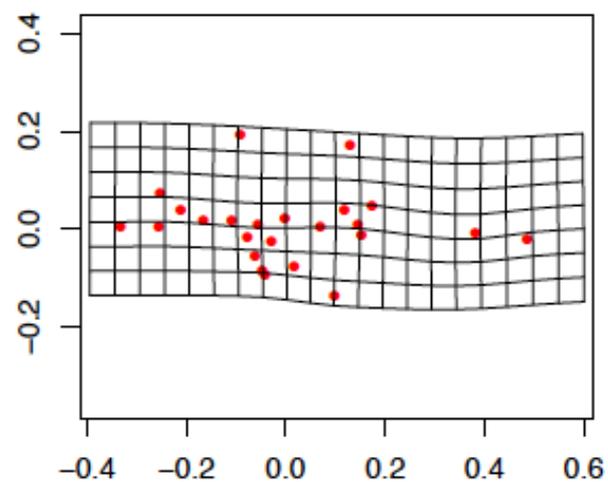
PW 6 -0.4,-1.18



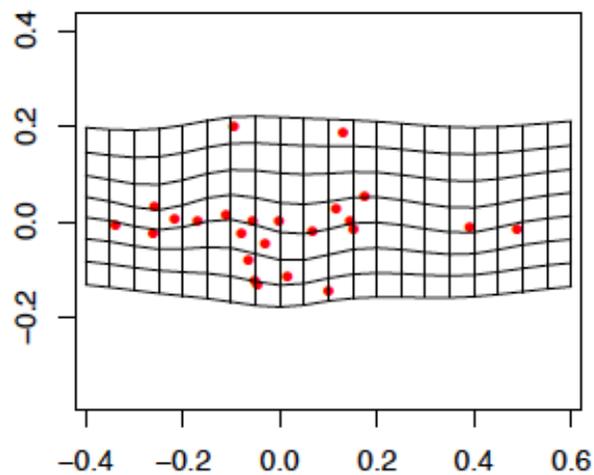
PW 7 -0.74,0.14



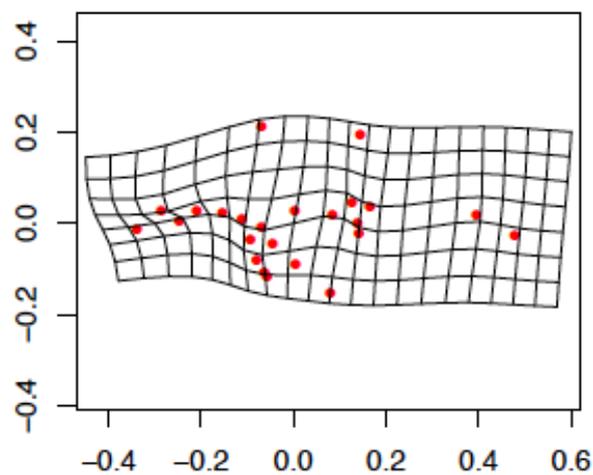
PW 8 0.41,0.99



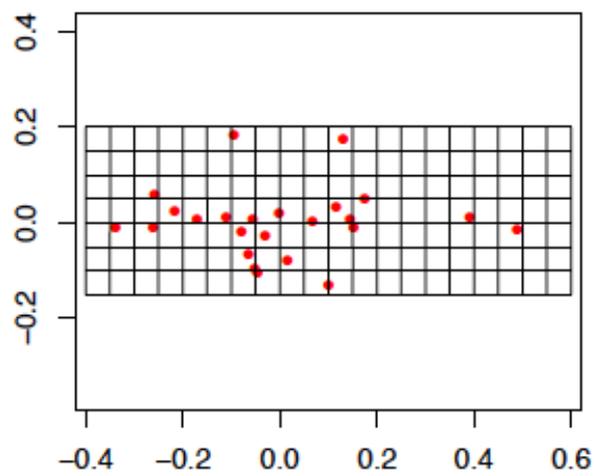
PW 9 0,-3.18



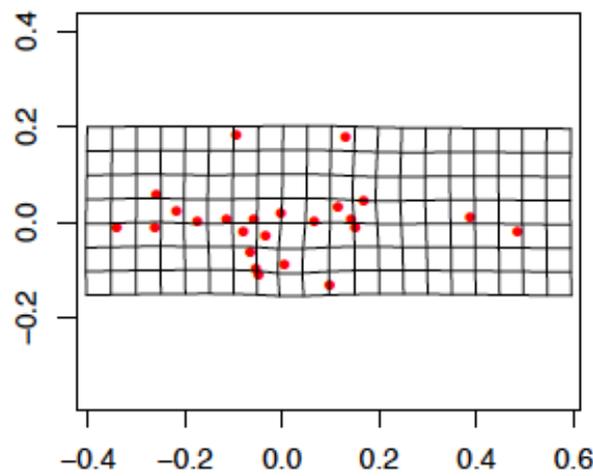
PW 10 -5.57,-6.19



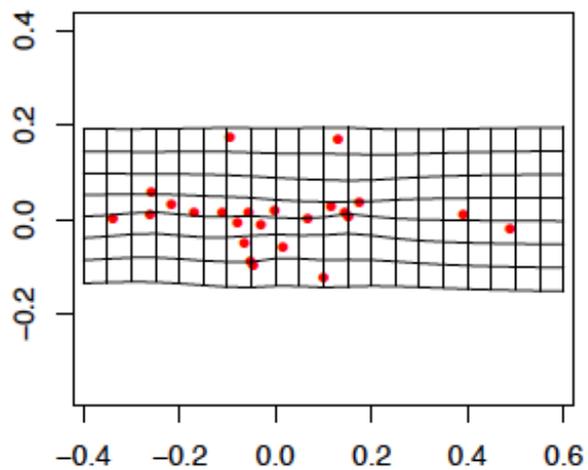
PW 11 0.03,0



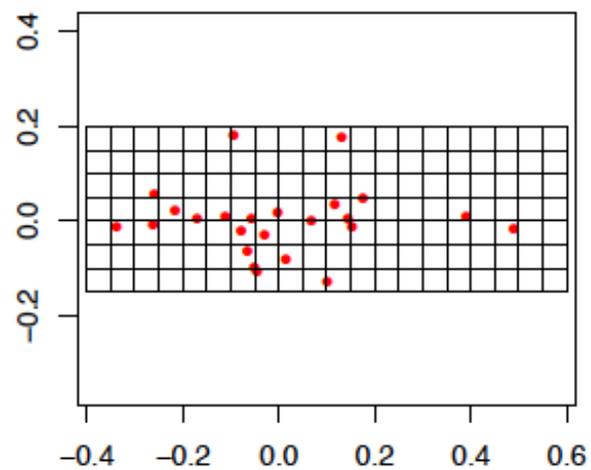
PW 12 1.81,1.36



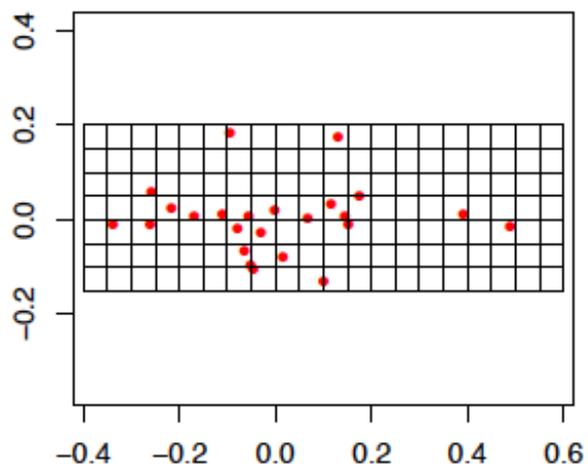
PW 13 0,-5.03



PW 14 0,0



PW 15 0,0



PW 16 0,0

