

Two new approaches to smoothing over complex regions

David Lawrence Miller

Mathematical Sciences
University of Bath

Modelling complex environmental spatial and temporal data,
Bath

Outline

Smoothing over complex regions

Intro

Solutions

Domain morphing with the Schwarz-Christoffel transform

Details

Simulation Results

Multidimensional Scaling

Details

Simulation Results

Conclusions

Outline

Smoothing over complex regions

Intro

Solutions

Domain morphing with the Schwarz-Christoffel transform

Details

Simulation Results

Multidimensional Scaling

Details

Simulation Results

Conclusions

Smoothing in 2 dimensions

- ▶ Have some geographical region and wish to find out something about the biological population in it.
- ▶ Response is eg. animal distribution, wish to predict based on (x, y) and other covariates eg. habitat, size, sex, etc.
- ▶ This problem is relatively easy if the domain is simple.



Smoothing with penalties

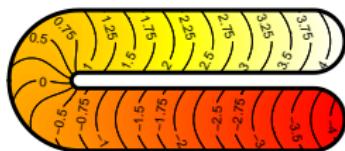
- ▶ Objective function takes the form:

$$\sum_{i=1}^n (z_i - f(x_i, y_i; \theta))^2 + \lambda \int_{\Omega} Pf(x, y; \theta) d\Omega$$

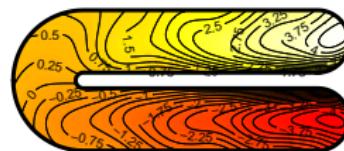
- ▶ f is the function you want to estimate, made up of some combination of basis functions.
- ▶ P is some squared derivative penalty operator, usually $P = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})^2$.
- ▶ This can be generalized to an additive model or GAM.

Smoothing over complex domains

- ▶ Smoothing of complex domains makes this a lot more difficult.
- ▶ Problem of leakage.
- ▶ Euclidean distance doesn't always make sense.
- ▶ Models need to incorporate information about the intrinsic structure of the domain.



(modified) Ramsay test function



Thin plate spline fit

Possible solutions to leakage problems

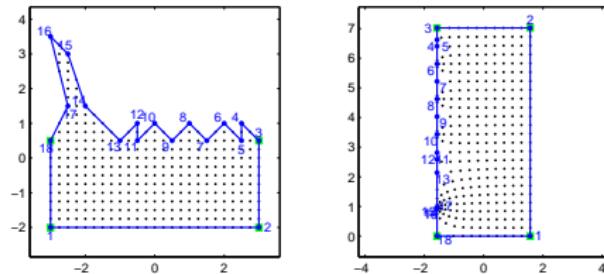
- ▶ FELSPLINE (Ramsay, (2002).)
- ▶ Domain morphing (Eilers (2006).)
- ▶ Within-area distance (Wang and Ranalli, (2007).)
- ▶ Soap film smoothers (Wood *et al.* (2008).)

Why morph the domain?

- ▶ Takes into account within-area distance.
- ▶ Gives a known domain that is easier to smooth over.
- ▶ Potentially less computationally intensive.

However:

- ▶ Don't maintain isotropy - distribution of points odd.
- ▶ Not clear what this does to the smoothness penalty.



Outline

Smoothing over complex regions

Intro

Solutions

Domain morphing with the Schwarz-Christoffel transform

Details

Simulation Results

Multidimensional Scaling

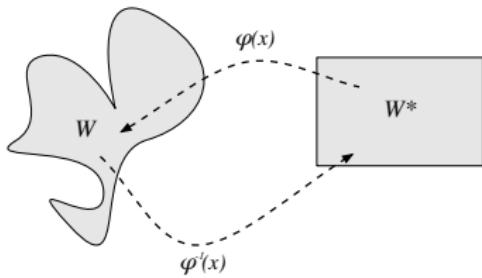
Details

Simulation Results

Conclusions

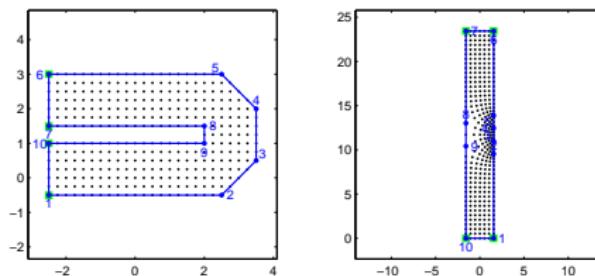
The Schwarz-Christoffel transform

- ▶ Take a polygon in some domain W and morph it to a new domain W^* .
- ▶ Do this by starting at the new domain and working back to the polygon.
- ▶ Can draw a polygonal bounding box around some arbitrary shape.



Schwarz-Christoffel algorithm

- ▶ Start with a rectangle.
- ▶ Add vertices.
- ▶ Iteratively deform by changing angles.
- ▶ Continue until the new shape is identical to the polygon.
- ▶ We then have a function for the mapping, $\varphi(x, y)$.
- ▶ $\varphi(x, y)$ is a conformal mapping.

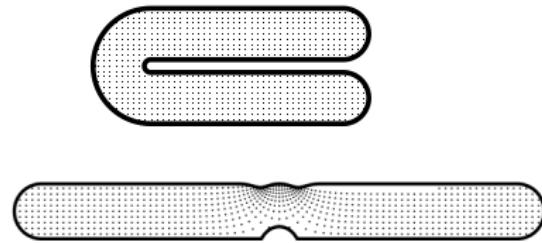


The mapping

- ▶ Use a bounding box around the horseshoe.

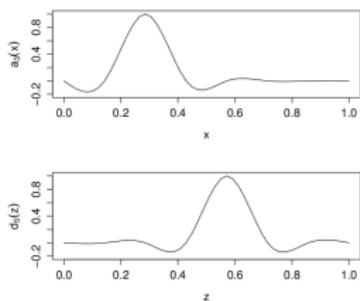


- ▶ Morphing the horseshoe shape still gives a slightly odd domain however, we are still doing better than before.

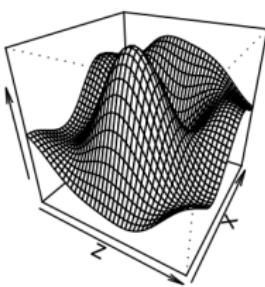
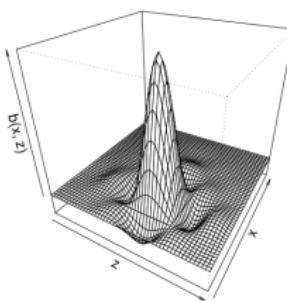


Results

- ▶ Fit using both thin plate regression splines and P-splines with a comparison to the soap film (current best.)
- ▶ MSE used to compare over a grid of prediction points.
- ▶ Transformation method does well on the horseshoe.



Tensor product



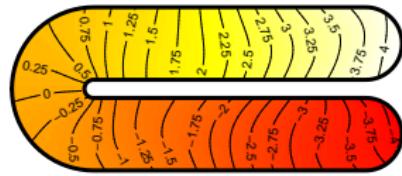
Thin plate

Horseshoe plots

Truth



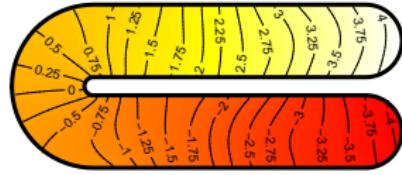
SC+PS



SC+TPRS

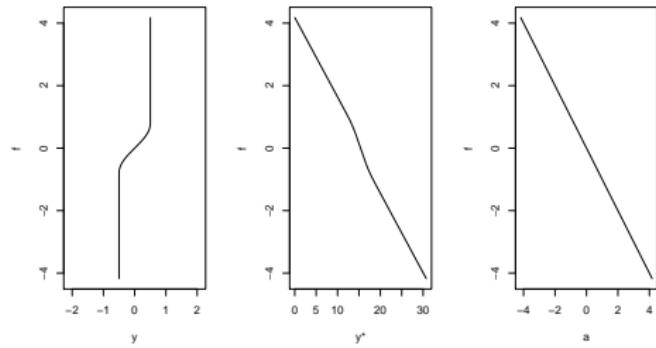
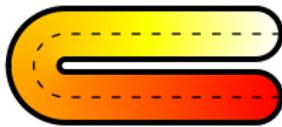


Soap film



Why does it do well?

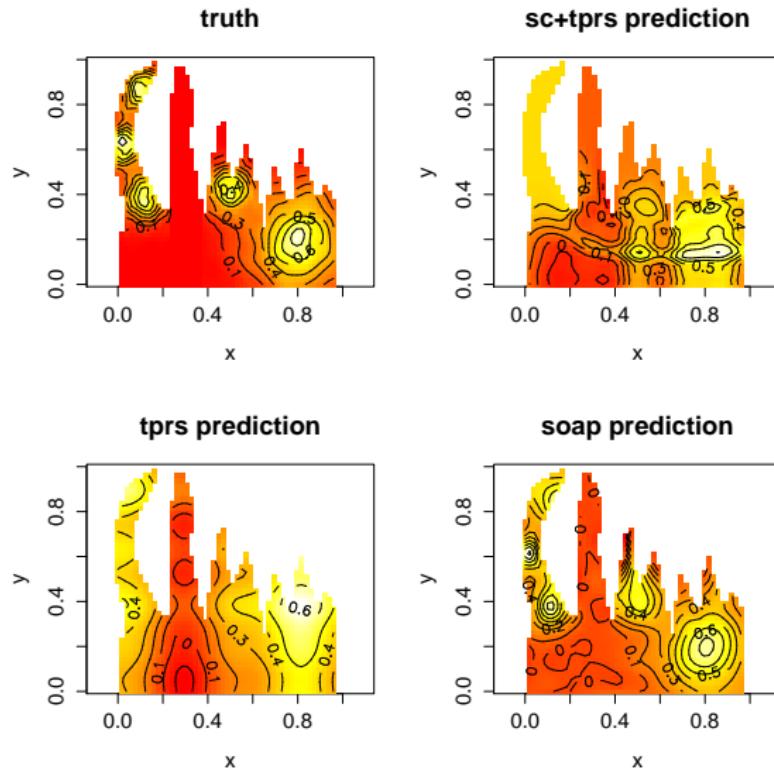
- ▶ Looking at line plots, we can see the difference in gradient.
- ▶ SC method seems to approximate the gradient better.



Problems

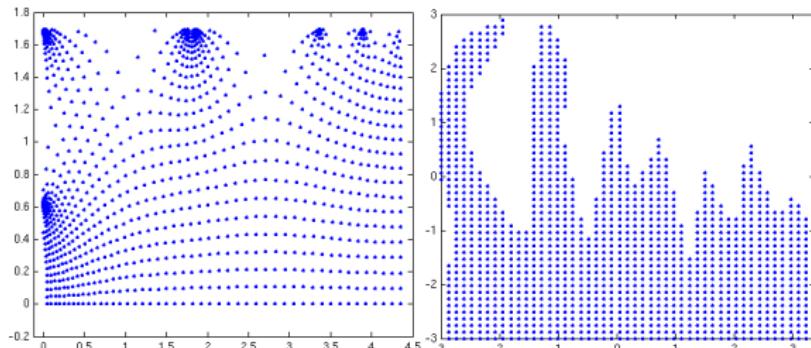
- ▶ Implementation is Matlab+R.
- ▶ Problems with more realistic situations:
 - ▶ Weird artifacts.
 - ▶ Arbitrary selection of vertices.
 - ▶ Morphing of domain appears to cause features to be smoothed over.

A more realistic domain



A more realistic domain - what's happening?

- ▶ Weird “crowding” effect.
- ▶ Different with each vertex choice. All bad.



Outline

Smoothing over complex regions

Intro

Solutions

Domain morphing with the Schwarz-Christoffel transform

Details

Simulation Results

Multidimensional Scaling

Details

Simulation Results

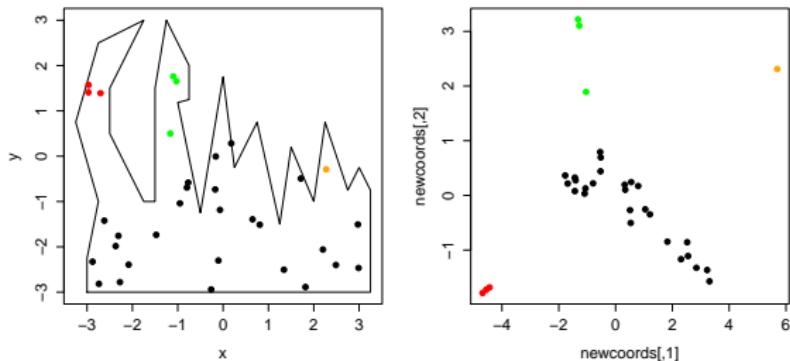
Conclusions

Multidimensional scaling and within-area distances

- ▶ Idea: use MDS to arrange points in the domain according to their “within-domain distance.”
- ▶ First need to find the within area distances.
- ▶ Can then smooth over the new points.

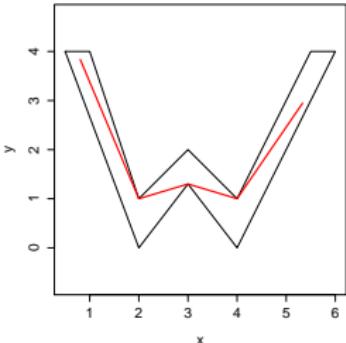
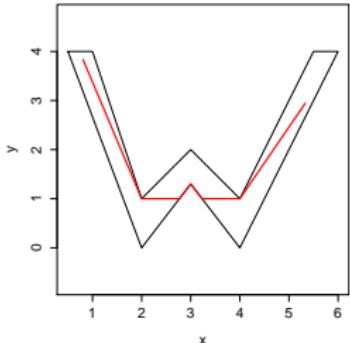
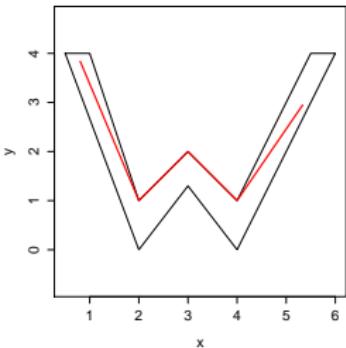
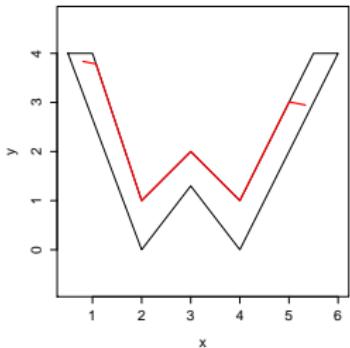
Multidimensional scaling refresher

- ▶ Double centre matrix of between point distances, D , (subtract row and column means) then find DD^T .
- ▶ Finds a configuration of points such that Euclidean distance between points in new arrangement is approximately the same as distance in the domain.
- ▶ Prediction outside the original data is fine.
- ▶ Already implemented in R by `cmdscale`.



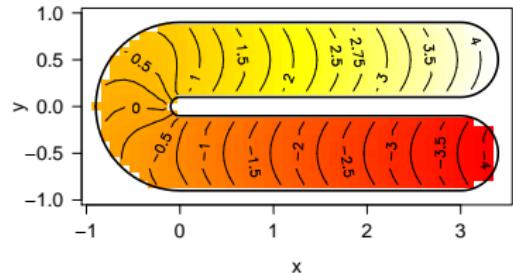
Finding within-area distances

- ▶ Use a new algorithm to find the within area distances.

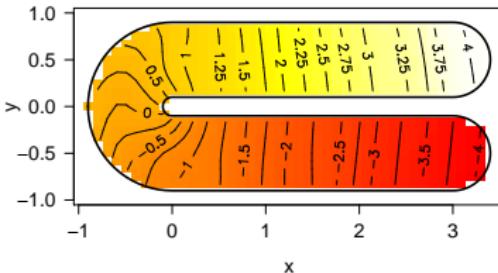


Ramsay simulations (low noise)

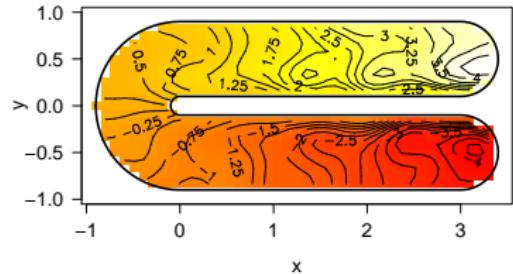
truth



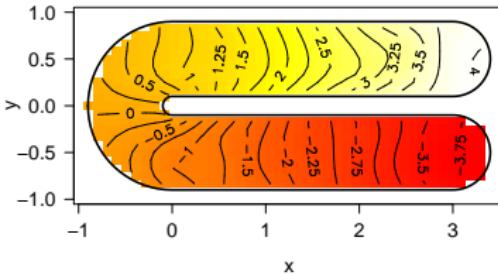
MDS



tprs

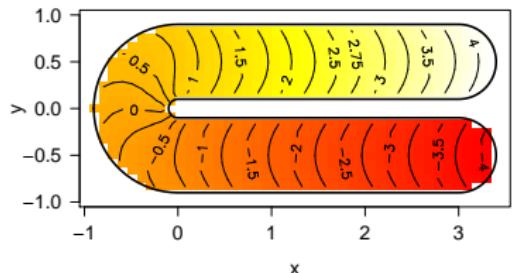


soap

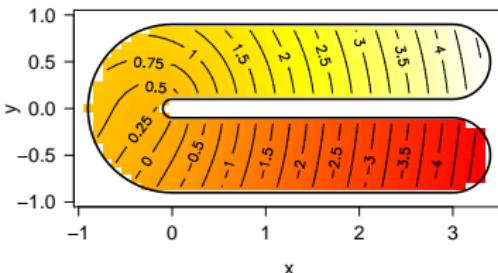


Ramsay simulations (high noise)

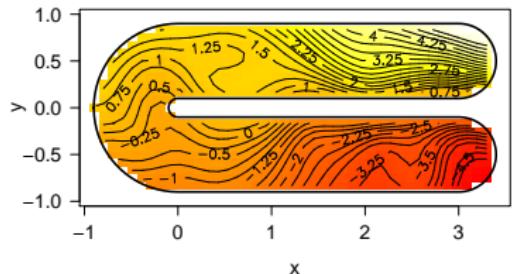
truth



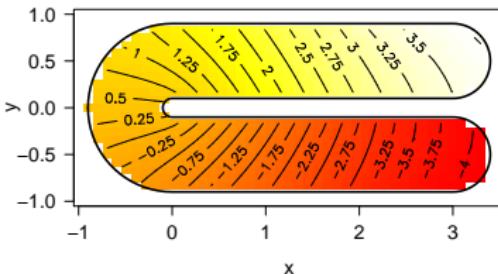
MDS



tprs

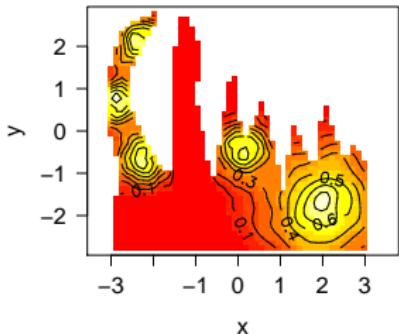


soap

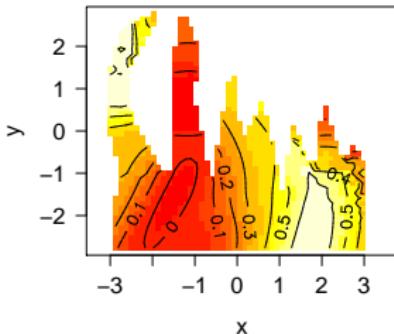


A different domain

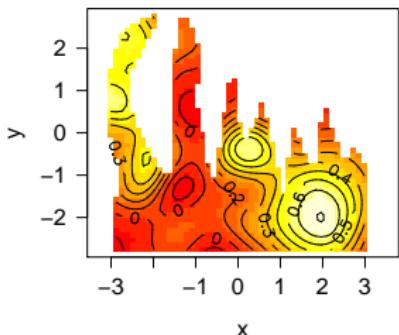
truth



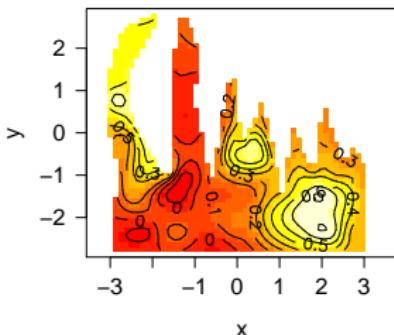
mds



tprs



soap



Outline

Smoothing over complex regions

Intro

Solutions

Domain morphing with the Schwarz-Christoffel transform

Details

Simulation Results

Multidimensional Scaling

Details

Simulation Results

Conclusions

Conclusions

- ▶ Seems that the S-C transform does not have much utility.
- ▶ MDS shows more promise, easier to transfer to higher dimensions.
- ▶ MDS does not impose strict boundary conditions so leakage still possible.
- ▶ After initial “transform” calculation, both methods only use the same computational time as a thin plate regression spline. (Soap is expensive.)

Further work

- ▶ Real data!
- ▶ Implement insertion of new points using method of Gower (1968).
- ▶ Current within-area distance algorithm is pure R, very slow. Need to re-write in C/Fortran.
- ▶ Pushing the data into more dimensions might be useful to separate points.

References

- ▶ S.N. Wood, M.V. Bravington, and S.L. Hedley. *Soap film smoothing*. *JRSSB*, 2008
- ▶ H. Wang and M.G. Ranalli. *Low-rank smoothing splines on complicated domains*. *Biometrics*, 2007
- ▶ T.A. Driscoll and L.N. Trefethen. *Schwarz-Christoffel Mapping*. Cambridge, 2002
- ▶ T. Ramsay. *Spline smoothing over difficult regions*. *JRSSB*, 2001
- ▶ P.H.C. Eilers. *P-spline smoothing on difficult domains*. University of Munich seminar, 2006
- ▶ C. Chatfield and A.J. Colins. *Introduction to multivariate analysis*. CRC, 1980
- ▶ J.C. Gower. *Adding a point to vector diagrams in multivariate analysis*. *Biometrika*, 1968.

Slides available at <http://people.bath.ac.uk/dlm27>