Complex Region Spatial Smoother (CReSS)

L. A. S. SCOTT-HAYWARD, M. L. MACKENZIE, C. R. DONOVAN, C. G. WALKER, AND E. ASHE

Conventional smoothing over complicated coastal and island regions may result in errors across boundaries, due to the use of Euclidean distances to measure interpoint similarity. The new Complex Region Spatial Smoother (CReSS) method presented here uses estimated geodesic distances, model averaging, and a local radial basis function to provide improved smoothing over complex domains. CReSS is compared, via simulation, with recent related smoothing techniques, Thin Plate Splines (TPS), geodesic low rank TPS (GLTPS), and the Soap film smoother (SOAP). The GLTPS method cannot be used in areas with islands and SOAP can be hard to parameterize. CReSS is comparable with, if not better than, all considered methods on a range of simulations. Supplementary materials for this article are available online.

Key Words: Geodesic distance; Local radial basis; Model averaging; Thin plate splines.

1. INTRODUCTION

Every year hundreds of surveys are carried out to monitor spatial and temporal trends in abundance and distribution of animals and, in particular, develop management frameworks for individual species. As an example, here we consider the 2006 survey of inshore waters around San Juan Island, Washington State (USA) and adjacent Canadian waters with a view to designating Marine Protected Areas for killer whales. For this survey (and similar land- or water-based surveys) to be an effective tool for conservation, accurate maps of animal densities are required.

Conventional kriging (Cressie 1990) and smoothers, such as Thin Plate Splines (TPS; Harder and Desmarais 1972), are currently employed to construct many density surfaces. However, these smoothing methods do not always produce good estimates for highly variable animal densities over complex topographies. The West coast of Canada/USA (Figure 1) or the man-made palm developments off the coast of Dubai (Figure 1) are examples of such
topographies. These represent unique ecosystems where long-term monitoring is under-
way, and where accurate mapping of animal densities and environmental data is of great
scientific interest.

Both TPS and kriging are often used for modeling spatial data and tend to allow the
spatial estimate to change smoothly with Euclidean (straight line) distance. This is not
always realistic as a straight path between points may cross an exclusion zone. For example,
a straight path between the locations of two abundance observations of a marine species
may cross a coastline. This can result in “leakage” in the model predictions, where high or
low densities in one body of water can have undue influence across boundaries into another.
This results in over- or underprediction in certain regions, which is purely an artifact of the
distance measure.

Rathbun (1998) and Curriero (2007) both proposed using a non-Euclidean metric with
kriging methods. However, Curriero noted that a nonpositive-definite covariance matrix
may arise in the non-Euclidean setting, making these methods untenable. He provided
conditions when certain non-Euclidean distances can be used in conventional kriging, but
as spline-based approaches can be used for all distance metrics we focus instead on these
types of methods. There are recent alternatives to the TPS method, which are designed to
respect complex boundaries. Finite Element L-Splines (FELS; Ramsay 2002) use a mesh
that is constrained to the domain and the observed points within it. The FELS approach has
been shown to be a marked improvement over TPS (Ramsay 2002).

A more recent alternative, the geodesic low-rank Thin Plate Spline (GLTPS; Wang and
Ranalli 2007) method, involves a mixed model representation of the TPS basis and uses
local neighborhoods around points to estimate geodesic interpoint distances (see Section
2.3 for more details). The amount of leakage that is permitted by GLTPS can be small
if the size of the neighborhood chosen is also small, but there is no explicit constraint to
prevent leakage across boundaries. This method also requires that a grid is chosen prior

Figure 1. (a) Killer whale group size off the West coast of the USA/Canada. Red dots are killer whale group
size, with the size of the dots representing relative group size. The range of the group sizes is 1–50 individuals.
(b) Palm Jumeirah in the Persian Gulf off the coast of Dubai (McMorrow 2004).
to modeling, to which the solution is sensitive. Additionally, GLTPS uses global basis functions, meaning individual points can be influential over the entire surface.

A recent alternative to conventional TPS is SOAP film smoothing (SOAP; Wood, Bravington, and Hedley 2008), which uses a specific basis function to model the domain interior, alongside a cyclic penalized cubic regression spline to model each boundary. This method respects boundaries but employs at least two tuning parameters: one global parameter for the interior, which may struggle to approximate surfaces with spatially varying complexity, and one for each boundary.

In this article, we propose a new general smoothing method, the Complex Region Spatial Smoother (CReSS), which respects boundaries and employs both global and local smoothers. This method implements a more accurate estimate of the geodesic distance than Wang and Ranalli (2007), allows the choice of a local or global radial basis function, and uses model averaging for model selection. In contrast to the TPS and GLTPS methods, CReSS uses points on the domain boundary to estimate geodesic distance between points. This explicitly constrains connections between points to be within the domain, allowing more accurate distances to be determined, even for sparse datasets. Although the main focus here is spatial smoothing, all methods considered may use additional covariates in practice. This article is organized as follows. In Section 2, we present the modeling framework and method description for all methods implemented in this article. We begin with a brief explanation of TPS (Section 2.1) before describing SOAP and GLTPS (Sections 2.2 and 2.3, respectively). In Section 2.4, the CReSS method is introduced. Two simulation studies are used to compare the performance of the TPS, GLTPS, SOAP, and CReSS methods. The first study (Section 3.1) uses the, now-standard, horseshoe region (Ramsay 2002), while the second study (Section 3.2) is loosely based on the palm developments off the coast of Dubai. Section 4 describes the application of CReSS and other methods to killer whale data around the West coast of Canada/USA, followed by a discussion of the relative merits of CReSS and the TPS, GLTPS, and SOAP methods for this application (Section 5). A FELS comparison has not been explicitly included in this article as our simulation conclusions were very similar to those of Wang and Ranalli (2007) and Wood, Bravington, and Hedley (2008). Specifically, GLTPS and SOAP both clearly outperform the FELS method. However, there has been no comparison of GLTPS and SOAP to date, so this is undertaken here.

2. METHODS

We investigate here methods for selecting an appropriate Generalized Additive Model (GAM; Hastie and Tibshirani 1990) to describe the relationship between a single response variable, \( y \), and a set, \( x \), of data points consisting of spatial coordinates, \( x_i = [x_{i1}, x_{i2}] \in \mathbb{R}^2, i = 1, \ldots, n \). A GAM consists of systematic and stochastic components, which are determined via the Generalized Linear Model (GLM; McCullagh and Nelder 1989). The systematic component models the expected response \( \mu \) as a function of \( x \):

\[
g(\mu) = \eta = s(x)
\]

with \( g \) and \( \eta \) being the usual link function and linear predictor, respectively. The distribution of the stochastic response \( Y \), given the covariates, is assumed to be of the Exponential family.
We concern ourselves here in particular with the specification of the linear predictor $\eta_i = s(x_i)$, where the two-dimensional surface $s$ will be approximated by a linear combination of basis functions. The surface approximation, $\hat{s}$, can then be used for prediction or to explain the systematic process generating the observations.

The remainder of this section describes TPS, two existing complex region methods (GLTPS and SOAP), and a new hybrid approach (CReSS).

2.1 THIN PLATE SPLINES (TPS)

TPS are generalizations of smoothing splines, providing a flexible smooth function in multiple dimensions (Harder and Desmarais 1972; Green and Silverman 1994). Only two-dimensional, penalized, low-rank thin plate regression splines are considered here, where the number, $T$, of underlying basis functions is less than the number of observations, $n$. A low-rank TPS requires some decisions regarding the number and location of basis functions each of which can be determined by the location of a corresponding “knot” point in $\mathbb{R}^2$.

TPS used in this article are of the form in Hastie and Tibshirani (1990) and implemented using the mgcv package in R (Wood 2003; R Development Core Team 2009). TPS can be used to estimate the smooth surface, $s$, by finding the function that minimizes

$$\|y - s\|^2 + \lambda J(s),$$

where $y$ is the vector of $y_i$ response data, $s$ is the corresponding vector of evaluations of $s$, $\|\cdot\|$ is the Euclidean norm, and $J(s)$ is a penalty functional measuring the wiggliness of $s$ (for example, that in Wood 2003). The smoothing parameter, $\lambda$, which controls the trade-off between fitting the response data and function smoothness, is typically chosen by generalized cross-validation (GCV).

Given a set of knots, and the corresponding basis functions, $b_t$ with parameters $\theta$, the equation for the fitted smooth surface $\hat{s}$ at a coordinate $x_i$ is

$$\hat{s}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \sum_{t=1}^T \hat{\delta}_t b_t(\theta),$$

where $\hat{\beta}$'s and $\hat{\delta}$'s are the estimated coefficients. Once the coefficients are estimated, $\hat{s}$ may be calculated for any value of $x$ (not just at data locations) using this equation.

For TPS the smooth surface $s$ uses radial basis functions:

$$b_t(d_{it}) = d_{it}^2 \log d_{it}. \quad (2)$$

Variable $d_{it}$ represents the distance, in this case Euclidean, between the $t$th knot and $i$th datum (Harder and Desmarais 1972; Green and Silverman 1994).

2.2 SOAP FILM SMOOTHING

SOAP and TPS use the same GAM framework for model fitting but the underlying bases are different in the two methods (Wood, Bravington, and Hedley 2008; Wood 2010). The idea of SOAP comes from dipping a wire loop into a bucket of soapy water. The outer boundary of the surface is represented by the wire, and the smooth surface by the soap.
film. The height of the soap film is the value of the smooth surface at that point. To move away from a flat surface, the soap film is distorted toward each data point but at the same time the overall surface tension is minimized. To achieve this, two sets of basis functions are used: one for the interior region of interest (soap film) and one for the boundary (wire). Therefore, knots must be specified both for the bases in the interior and for bases on every boundary edge. This leads to the major disadvantage of this approach: basis setup can be computationally expensive and currently there is little guidance available regarding how best to choose the number and location of these knots. Furthermore, SOAP does not discriminate between a boundary that has model significance (such as a coastline) and an arbitrary one determining the edge of the area of interest. Nevertheless, SOAP has been shown to perform better than both TPS and FELS and avoids the issue of leakage (Wood, Bravington, and Hedley 2008), but has yet to be compared with GLTPS, which we describe next. For a more detailed description of the SOAP method, see Wood, Bravington, and Hedley (2008).

2.3 Geodesic Low-Rank Thin Plate Splines (GLTPS)

GLTPS (Wang and Ranalli 2007) is fit within a mixed model framework using a modified version of low-rank TPS, where an estimated geodesic distance is used to measure the similarity between all observations and knot locations. The geodesic distance between two points, $x_i$ and $x_j$ in a region $A$, is the length of the shortest path between $x_i$ and $x_j$ that lies within $A$.

Wang and Ranalli (2007) estimated the geodesic distance via a distance network. The vertex set of this network consists of the data point locations, with edges included between every pair of vertices. The edges are labeled with distance “weights” as follows. For each vertex $v$, the edges joining $v$ to each of its $k$ closest neighbors are labeled with the Euclidean distance to that vertex. All unlabeled edges are then given infinite weight in the network. Floyd’s algorithm (Floyd 1962) is used to determine the shortest distance through this network between all pairs of vertices. This shortest distance is used to approximate the geodesic distance between all pairs of points. Wang and Ranalli (2007) recommended using the smallest $k$ for which there are no infinite values in the shortest path distance matrix (all points can be reached from every other point).

While the GLTPS technique has been shown to perform better than both TPS and FELS, it does not preclude the shortest distance between two points crossing a boundary. The choice of $k$ is fixed for the entire surface and represents a trade-off between accuracy and computational expense. Ideally, $k$ is small so that in areas where the exclusion area between boundaries is small, the possibility and extent of leakage is also small. However, if $k$ is too small, the points in the network may be poorly connected. Furthermore, if $k$ is too big, then points are connected directly by Euclidean distance and boundaries will be breached.

To alleviate problems associated with relatively small $k$, Wang and Ranalli (2007) used a predefined grid over the region. The finer the grid, the greater the likelihood of the distances between points converging on the true geodesic distance. While this provides a lower likelihood of leakage, grid resolution is constrained by computational resources in practice. Notably, Wang and Ranalli (2007) did not explicitly include the boundary points in the grid.
Owing to leakage and plotting artifacts created using distances calculated by Wang and Ranalli (2007), we use an alternative method of calculation of geodesic distance (Section 2.4). Thus, the modeling framework of GLTPS is assessed without being compromised by geodesic distance calculations.

2.4 **Complex Region Spatial Smoother (CReSS)**

CReSS contains elements similar to the three methods described previously. Like TPS and SOAP, it uses the GAM framework. It uses a different basis to TPS, but the type is still a radial function. As with GLTPS, a geodesic distance metric is used. However, unique to CReSS, a model averaging approach is adopted, which has proven very successful in mapping animal densities in complex regions (the application that motivated its development). We describe here in more detail the main components of the CReSS method, beginning with the method of estimating the geodesic distance.

2.4.1 Improved Geodesic Distance Estimation. Before fitting a model using the CReSS method, we must first calculate geodesic distances between data locations and knot locations. As explained in Section 2.3, to determine the geodesic distance between points, GLTPS constructs a network with vertices at the points. In CReSS, we also build a network to estimate geodesic distances, but our vertex set includes the corners of the boundary polygons and the knot locations, as well as the data points. We use polygons to identify the boundary of the exclusion zone, for example, a coastline. This boundary is defined by one or more polygons, the vertices of which are included in the network to accommodate the calculation of the geodesic distance between the pairs of data points. Thus, the edge set is made up of all line segments between distinct pairs of vertices, with the length of all edges that do not cross the exclusion zones being calculated using the Euclidean norm, and all other edges being assigned infinite length. In the case when the edge between two data points is infinite, the geodesic distance is calculated using noninfinite edges (see example in Figure 2). Floyd’s algorithm is used to determine the shortest distance through the network between all pairs of data points, knot points, and the boundary points (which GLTPS does not explicitly include). The use of polygons means the estimation of geodesic distance using this method is as accurate as the definition of the exclusion area polygons.

2.4.2 Basis Structure. When choosing a basis, the behavior of many radial basis functions near the boundaries is a cause for concern (Fornberg et al. 2002). A test region (supplementary Figure 1), which shows a triangular exclusion region, is used to show the global nature of TPS, using one of the usual TPS basis functions [supplementary Figure 2(a)]. The values of the TPS basis increase with distance from each knot location often leading to errors at the edges of the region of interest (Fornberg et al. 2002), which give rise to pronounced edge effects. These effects are exaggerated when non-Euclidean distances are used, since the furthest distance from a knot point may no longer be at the edge of the plot and the radial pattern may no longer be guaranteed if distances are modified to accommodate boundaries. In some cases, the basis is distorted, leading to areas of reinforcement where large distances compound [supplementary Figure 2(b)]. This could make some surfaces difficult to approximate, and prediction errors (mean squared error, MSE) are typically greatest where reinforcement occurs [supplementary Figure 2(d)]. A local
radial basis can accommodate spatially varying complexity more easily than a globally defined radial TPS basis. The local basis restricts the distance from each knot over which the basis is effective and reduces the likelihood of reinforcement occurring [supplementary Figure 2(c)].

Thus, the CReSS replaces global radial basis function $b_t(\theta)$, [Equation (2)] with $b_t(h,r) = \exp(-h/r^2)$, where $r$ dictates the decay of this Exponential function with distance (Rathbun 1998), and thus its local nature. Notably $h$ indicates a geodesic distance that in practice will be between some $t$th knot and $i$th data location, indexed accordingly $h_{ti}$. Parameter $r$ takes values such that if $r$ is small, that model will have a set of local basis functions and if $r$ is large, that model will have a set of global basis functions. The exact values of $r$ are dependent upon the range and units of the spatial covariates. We have removed the planar parts of Equation (1) since a linear trend in $x_1$ or $x_2$ could be based on unrealistic Euclidean distance. Thus, the equation for the smooth surface $s$ at point $x_i$, 

![Figure 2. An example of graph construction using the CReSS method, where the gray areas represent exclusion zones, filled circles represent the polygon vertices, and lines represent edges. (a) is the Euclidean distance between two points (open circles), (b) is the distance network created by CReSS for these two points (edges with infinite weight have been omitted), and (c) is the geodesic distance between the two points using only the edges shown in (b).](image-url)
2.4.3 Model Averaging. Rather than just use a single best model, we find the relative merits of a set of models and average the results. For each model in the set, we change the number of knot locations and/or the size of parameter \( r \).

For model selection, we use frequentist model averaging (Buckland, Burnham, and Augustin 1997; Claeskens and Hjort 2009), with AIC\(_c\) (Hurvich and Tsai 1989) for model weights. AIC\(_c\) is a small sample Akaike information criterion (AIC) and Burnham and Anderson (2002) recommended it be used when the ratio \( n/k < 40 \), where \( n \) is the sample size and \( k \) is the total number of estimated regression parameters (including the intercept and \( \sigma^2 \)). For values of this ratio > 40 AIC and AIC\(_c\) converge. Other information criteria may be substituted for AIC\(_c\).

The full set of models is limited to models with \( \Delta \text{AIC}_c < 10 \) since Burnham and Anderson (2002) suggested that a model with \( \Delta \text{AIC}_c < 10 \) shows no empirical support for that model. This reduced set of models is the candidate model set \( (M) \), and the relative merits of these models is found by calculating weights (Buckland, Burnham, and Augustin 1997; Claeskens and Hjort 2009) using

\[
w_m = \frac{\exp\left(-\frac{1}{2} \Delta_m\right)}{\sum_{m=1}^{M} \exp\left(-\frac{1}{2} \Delta_m\right)},
\]

where \( m = 1, \ldots, M \) and \( \Delta \) is the difference in AIC\(_c\) between model \( m \) and the best model (lowest AIC\(_c\)). If the Bayesian information criterion is used, Equation (3) becomes the Schwarz (1978) approximation of the Bayes’ factor. Predictions are made for all models in model set \( M \) and their weights, \( w_m \), are used to calculate a weighted sum of predictions to get an overall prediction. To calculate predictions, we calculate the geodesic distance between prediction locations and knot locations, using the method described in Section 2.4.1, and use them to make the bases.

We consider a range of knot sets where each set contains a different number of knots and a range of values for parameter \( r \). For example, 10 knot sets \( (\tau = 10) \) and 5 \( r \)'s \( (R = 5) \) results in 50 possible models. Of these models, the candidate knot set, \( M \), is some number \( \leq 50 \). Figure 3 gives an overview of the CReSS algorithm.

2.4.4 Choice of Knots and \( r \). Knot placement in this article follows Wang and Ranalli (2007) by using a space-filling design, such as that of John et al. (1995) from the FIELDS package (Furrer et al. 2010). The knot sets therefore represent a range of different numbers of knots, whose locations in each case are determined by the space-filling algorithm. The knots for all methods were generated from the observed data so any simulation run comparing methods has the same knot choices.
CReSS:

**Inputs**
- spatial coordinates,
- geodesic distance matrix (data to knot locations),
- \( \tau \) knot sets,
- and basis ‘range’ parameter \( r \) (\( R \) values in total considered)

**Model Fitting**
- \( \tau \times R \) candidate models calculated
  
  for \( j \) in 1: \( \tau \)
    
    for \( k \) in 1:R
      
      Calculate locally radial basis functions for given \( r \) and knot set, \( j \)
      
      Fit models for given \( k \) and \( j \) using maximum-likelihood as per GAM
      
      Retrieve AIC\(_c\) score (or other fit statistic)
  

**Model Selection**
- Calculate model weights for models with \( \Delta AIC_c < 10 \) (\( M \) models, \( M \leq \tau \times R \))

**Model Prediction**
- Calculate weighted sums of predictions (using AIC\(_c\) weights) from \( M \) models using geodesic distances from prediction locations to knot locations.

Figure 3. Pseudocode outlining the structure of CReSS.

To find a range of values for \( r \), we take a sequence of 50 values on the log scale so that there are more smaller values than larger ones. The sequence is defined using \( r_{\text{min}} \) and \( r_{\text{max}} \):

\[
\left( r_{\text{min}} = \sqrt{\frac{h_{\text{max}}}{21}}, r_{\text{max}} = \sqrt{\frac{h_{\text{max}}}{3e^{-7}}} \right),
\]

where \( h_{\text{max}} \) is the maximum distance between data locations. The minimum \( r \) is chosen such that it is small enough to allow local patterns but not so small that the noise is overfitted and the maximum \( r \) allows the model to accommodate a plane. Additional details on the formulas for \( r_{\text{min}} \) and \( r_{\text{max}} \) can be found in the supplementary material.

### 2.5 Model Performance

Two measures were employed to evaluate the performance of each of the methods: the estimation bias, \( \hat{B} \) [Equation (4)] and MSE [Equation (5)]. Each of the performance measures was estimated by simulation, whereby \( P \) data realizations were generated from a known underlying function and known model of noise.

The estimation bias was assessed at each of \( N \) locations within the domain. So the bias for the \( j \)th location \( x_j \) (\( j = 1, \ldots, N \)) was

\[
\hat{B}_j = P^{-1} \sum_{p=1}^{P} \hat{z}_p(x_j) - z^*(x_j), \tag{4}
\]

where \( \hat{z}_p(x_j) \) is the method’s estimate of the true value, \( z^*(x_j) \), for realization \( p \). MSE considers differences in predictions to the underlying function and is calculated for out-of-set prediction locations. These are locations unseen by the fitting process, meaning the points \( x_j \) in Equation (5) were not those used to estimate \( \hat{z} \). MSE is calculated for each
replicate, \( p \):

\[
\widehat{\text{MSE}}_p = N^{-1} \sum_{j=1}^{N} \left\{ \hat{z}_p(x_j) - z^*(x_j) \right\}^2
\]  

(5)

with terms previously defined.

We also used a Wilcoxon signed rank test (Wilcoxon 1945) to see if MSE scores for TPS, GLTPS, and CReSS were significantly different to SOAP (the most recent method).

### 3. SIMULATION STUDIES

We consider CReSS’s performance using two simulation studies and compare it with TPS, GLTPS, and SOAP. The first simulation employs the horseshoe (Ramsay 2002; Wang and Ranalli 2007; Figure 4, Section 3.1) and the second is inspired by a land reclamation project in the Persian Gulf near the coast of the United Arab Emirates (Figure 8, Section 3.2). Both are examples of areas with irregular shaped boundaries and sharp changes in the response across these boundaries, though the latter shows more complexity.

#### 3.1 HORSESHOE SIMULATION

The established shape/surface for evaluating complex topography smoothing methods is the horseshoe developed by Ramsay (2002; Figure 4). The horseshoe varies smoothly from approximately 4 to \(-4\) from the right-hand end of the top arm to the right-hand end of the lower arm. Three test cases were generated, for evaluation of the different methods, by adding a normal errors noise term with standard deviation 0.05, 1, and 5 to the function values (called low, medium, and high noise, respectively) and randomly

![Figure 4. The underlying function on the horseshoe region (Ramsay 2002).](image-url)
Table 1. Mean MSE scores and standard deviation (SD) for all methods at all noise levels on the horseshoe simulation. An “*” indicates that the MSE results of a method are significantly better than SOAP \((p < 0.05,\text{ Wilcoxon signed rank test})\), a “†” indicates that the results for SOAP are significantly better.

| Method | Low noise | Medium noise | High noise |
|--------|-----------|--------------|------------|
|        | Mean      | SD           | Mean       | SD          | Mean       | SD          |
| TPS    | 0.24608†  | 1.69 × 10⁻²  | 0.2925†    | 0.0224      | 1.163†     | 0.285       |
| GLTPS  | 0.00062†  | 1.54 × 10⁻⁴  | 0.0261*    | 0.0064      | 0.365      | 1.198       |
| SOAP   | 0.00055   | 7.68 × 10⁻⁵  | 0.0294     | 0.0114      | 0.458      | 0.358       |
| CReSS  | 0.00071†  | 2.11 × 10⁻⁴  | 0.0292     | 0.0105      | 0.357*     | 0.299       |

sampling \(n = 600\) points from each noisy surface. Predictions were obtained on a grid of \(N = 3584\) points. For the GLTPS method, the geodesic distances calculated using code by Wang and Ranalli (2007) were poor and led to artifacts. Therefore, the same geodesic distances were used for both GLTPS and CReSS and calculated using the method in Section 2.4. SOAP is constructed using a cyclic penalized cubic regression spline (40 knots) to estimate the unknown boundary values (Wood, Bravington, and Hedley 2008). There is no guidance for boundary knot allocation, so the authors use the same as Wood, Bravington, and Hedley (2008). Additionally, for CReSS, parameter \(r\) took 50 values between 0.58 (local basis) and 4873 (global basis) for basis calculation. All methods use a choice of 10–100 knots (in steps of 5) generated using the space filling algorithm (number of knot sets, \(\tau\), is 19). As per author’s recommendations, model selection was performed using GCV for TPS and SOAP, AIC for GLTPS, and \(AIC_c\) for weights’ calculation for CReSS.

The number of simulation realizations, \(P\), was set to be 100. The simulation parameters for all methods are detailed in supplementary Table 1.

CReSS, SOAP, and GLTPS all perform substantially better than TPS for this function at all noise levels (Table 1). Consistent with other analyses (Ramsay 2002; Wang and Ranalli 2007; Wood, Bravington, and Hedley 2008), the main error for TPS is along the inner edges of the two arms, while GLTPS, SOAP, and CReSS show their greatest error in the elbow region (Figures 5–7). At low noise, the elbow region seems more problematic for SOAP since the errors radiate along the arms [Figure 5(c)]. The range of the estimation bias, \(\hat{B}\), is comparable for complex region methods, but slightly lower for CReSS at high noise (Figures 5–7). CReSS also has the best mean MSE score (and smallest variance) at high noise and performs significantly better than SOAP (Table 1). At other noise levels the methods are very similar with SOAP marginally, but significantly, better at low noise and GLTPS significantly better than SOAP at medium noise. The actual differences between methods are so small at low noise that they are practically insignificant given the very small MSEs in question. A single model tended to be chosen at low noise using CReSS, but for high noise many more were chosen (based on \(AIC_c\)) and averaged. CReSS also describes the increasing function along the arms better than SOAP or GLTPS at high noise (Figure 7).

### 3.2 Palm Simulation

The simulated palm region is inspired by the palm structures in the Persian Gulf off the coast of Dubai (Figures 8 and 1). The upper-hat-shaped island and edge pieces represent the
outer breakwater with two channels, while the inner segment represents a palm leaf with three fronds on each side. The surface is constructed using the definitions in supplementary Table 2 and the zones in Figure 8, and created to test the performance of the method when the function changes greatly across small exclusion areas.

The function varies smoothly from approximately \(-40\) to \(110\) and contains an island, which can give rise to the reinforcement issue outlined in Section 2.4. Three test cases were
Figure 7. Bias for high noise ($\sigma = 5$), (a) TPS, (b) GLTPS, (c) SOAP, and (d) CReSS. The heavy weight black line is for zero bias and the finer contours are in increments of 0.1 except for plot (a) where the increment is 0.2. For ease of viewing, plot (a) has a different color/contour scale to (b), (c), and (d).

Figure 8. The underlying function on the simulated palm region. The letters refer to the regions in supplementary Table 2 used to construct the function.
generated, for evaluation of the different methods, by randomly choosing $n = 500$ points from the surface and adding a normal errors noise term with mean zero and standard deviations 0.5, 9, and 50 to the function values. The noise was added such that the signal-to-noise ratio (SNR) is similar to that of the horseshoe simulation. The SNR was calculated using $\frac{\text{var}(y)}{\text{var}(y - y_n)}$, where $y$ is the underlying function and $y_n$ is the underlying function with noise added. Predictions were obtained on $N = 2518$ points. After some trial and error to choose “good” values, SOAP was fitted using 50 knots for the outer boundary and 40 for the island. Since there is no guidance for boundary knot selection and the default (10 knots) is too small in many situations, these knot numbers were found to give the best results after an extensive nonexhaustive search. In keeping with the simulation study in Section 3.1, parameter $r$ for the CReSS method took 50 values between 1.03 and 8579. For all methods, a choice of 10–100 knots (in increments of 5, $\tau = 19$) was allowed and finalized using GCV (TPS and SOAP), AIC (GLTPS), and AICc (CReSS). In the case of SOAP these knots are for the interior soap basis. A table of full simulation settings can be found in supplementary Table 3.

Of the methods trialed, the CReSS method exhibited the lowest MSE scores at low and medium noise (Table 2) for this region and was significantly better than SOAP at medium and high noise ($p < 0.05$). Unsurprisingly, TPS has the worst fit to the data and underlying function across all noise levels (Table 2). The lack of fit of TPS becomes more pronounced as noise increases and is mainly due to leakage across the island, where the difference in underlying function values is greatest [Figures 9(a), 10(a), and 11(a)]. At high noise there is also some evidence of leakage down through the palm fronds from the hotspot at the top. As expected, there was no evidence of leakage for GLTPS, SOAP, or CReSS, however, all methods (including TPS) struggled to model the high and low function values to the left of the stem (Figures 9–11). These errors may be due to lack of coverage by the data points here or an inflexibility in knot number and/or placement.

GLTPS shows a good numerical fit (Table 2), particularly at high noise, but exhibits some problems visually due to artifacts (striations) to the upper right and left of the island. These are particularly apparent on the prediction plot for a single realization with medium noise [Figure 12(a)]. We consider that this is due to reinforcement issues arising from the global function basis and thus while giving reasonable MSE scores, the errors are clearly concentrated spatially. SOAP deals very well with the breakwater though there is some evidence of errors on the ends of the upper fronds and, as noise increases, on the upper edge

| Method | Low noise | Medium noise | High noise |
|--------|-----------|--------------|------------|
|        | Mean    | SD | Mean | SD | Mean | SD |
| TPS    | 97.48†  | 7.42 | 101.95† | 7.92 | 213.47† | 33.43 |
| GLTPS  | 10.51   | 9.95 | 24.18† | 4.78 | 131.45* | 29.54 |
| SOAP   | 21.47   | 81.52 | 22.70 | 3.73 | 188.24 | 51.33 |
| CReSS  | 7.66     | 2.70 | **21.93** | **5.48** | **163.80** | **43.41** |

The bold scores represent the best average for each statistic at each noise level.

Table 2. Mean MSE scores and standard deviation (SD) for all methods at all noise levels for the palm simulation using 500 data points. An “*” indicates that the MSE results of a method are significantly better than SOAP ($p < 0.05$, Wilcoxon signed rank test), a “†” indicates that the results for SOAP are significantly better.
of the central palm shape [Figure 11(c) and supplementary Figure 6(c)]. CReSS respects all the boundaries, keeping the high values under the breakwater and the low values above it without leakage. However, CReSS exhibits some negative bias just under the breakwater and just above the central palm shape [Figures 9(d)–11(d) and supplementary Figures 4(d)–6(d)]. These are two areas where perhaps the radial nature of CReSS struggles to approximate the striations of the underlying function. While, CReSS performs significantly better than SOAP at medium and high noise, on average, at low noise CReSS performs better than SOAP, but not significantly so due to the highly variable performance of SOAP. It approximates the true surface either well (as good or better than CReSS) or very badly (much worse than CReSS). Numerically GLTPS performs best at high noise but we believe the local concentration of errors and associated artifacts make GLTPS a poor choice.

The methods were also compared on the palm simulation surface when data are sparse. One hundred data points of the three different noise levels were taken from the surface. After trial and error, SOAP was constructed with 10 knots for both the outer boundary and for the island. In keeping with the two previous simulation studies, parameter $r$, for the CReSS method, took 50 values between 1.03 and 8579. For all methods a choice of 10–75 knots was allowed ($\tau = 14$). Full simulation settings can be found in supplementary Table 3. The results for $n = 100$ are much more conclusive. CReSS was significantly the best performing method at all noise levels ($p < 0.05$; Table 3). Due to the skewed nature of the results, particularly for SOAP and GLTPS, the median MSE is also included. At low
Figure 10. Absolute bias to the underlying function for medium noise ($\sigma = 9$), (a) TPS, (b) GLTPS, (c) SOAP, and (d) CReSS. The transparency of the dots represents the scale of the absolute bias, where white is 0–10 units, and black is a bias of approximately 50–55 units, except in the case of TPS, where black represents a bias of 50–82 units.

noise, the results for SOAP and GLTPS are surprisingly variable and GLTPS continued to show plotting artifacts like those in Figure 12(b). Bias plots for the data sparse simulation can be found in supplementary Figures 7–9. Similar results can be drawn from these plots compared with the data rich plots, however, it is interesting to note that SOAP exhibits an increase in bias around the boundary possibly due to a smaller number of knots here.

Table 3. Mean, median (med), and standard deviation (SD) of MSE scores for all methods at all noise levels for the palm simulation using 100 data points. A "∗" indicates that the MSE results of a method are significantly better than SOAP ($p < 0.05$, Wilcoxon signed rank test), a "†" indicates that the results for SOAP are significantly better.

| Method  | Low noise | Medium noise | High noise |
|---------|-----------|--------------|------------|
|         | Mean (med) | SD           | Mean (med) | SD           | Mean (med) | SD           |
| TPS     | 130 (119) | 36.6         | 152† (142) | 40.5         | 499† (152) | 477         |
| GLTPS   | 1745 (121) | 5926         | 144* (78.4) | 363          | 383* (191) | 343         |
| SOAP    | 1806 (99.3) | 11207        | 141 (107)  | 181          | 521 (348)  | 423         |
| CReSS   | 36.6∗ (34.0) | 13.3         | 77.1∗ (73.9) | 24.8         | 345∗ (312) | 138         |

The bold scores represent the best average for each statistic at each noise level.
Figure 11. Absolute bias to the underlying function for high noise ($\sigma = 50$), (a) TPS, (b) GLTPS, (c) SOAP, and (d) CReSS. The transparency of the dots represents the scale of the absolute bias, where white is 0–10 units, and black is a bias of approximately 60–62 units, except in the case of TPS, where black represents a bias of approximately 60–80 units.

Figure 12. Example predictions for a single realization with medium noise, (a) TPS, (b) GLTPS, (c) SOAP, and (d) CReSS.
4. KILLER WHALE ANALYSIS: MAPPING KILLER WHALE GROUP SIZE OFF THE WEST COAST OF THE CANADIAN/USA BORDER

In this section, we apply the methods described in this article to killer whale group size data collected in the area shown in Figure 1. This region of this study contains a complex area of coastline with 15 islands and the data consist of \( n = 763 \) geographically referenced group size estimates from 1 to 50 killer whales. The data are overdispersed and therefore modeled as a GAM using a quasi-Poisson error distribution and a log link. However, GLTPS was fitted using Poisson errors with a log link as no quasi-Poisson distribution is available for this method (a mixed model). Greater interpretability might be gleaned by including other covariates, such as depth or chlorophyll, however, the two-dimensional smooth or geographical location serves as a good exemplar here.

For all methods, knot sets of 10–60 knots (\( \tau = 11 \)) were chosen randomly from a subsample of 200 space-filled knots. The reduced sample (not all 763 available knot locations) prevented overpopulating the “arms” of the data and thus having too many knots where there is no data support. SOAP was particularly difficult to parameterize with knot choices needed for all 16 boundaries (1 outer box and 15 islands). There is no guidance for selection so, after various combinations tried, the best choice was 40 knots for the outer box and 5 for all remaining polygons. GCV was used to choose knots for TPS and SOAP, while QAIC \(_c\) (Lebreton et al. 1992) governed model averaging for CReSS and Residual Sums of Squares (RSS) was used for GLTPS. For CReSS, parameter \( r \) took 50 values between 0.22 and 1846. Predictions for all methods were made on the scale of the response (animal group size) using a grid of \( N = 4799 \) points.

All of the methods showed a decent fit to the data using RSS scores, with SOAP having the best fit to the data, and CReSS having the best QAIC \(_c\) score (Table 4). However, even very minor extrapolation presented some interesting problems. TPS and GLTPS both show edge effects from the use of a global basis function, but the range of group size predictions for TPS is much closer to the data (Figure 13). GLTPS also shows some effect of reinforcement errors in the very center of the plot. SOAP is notably poor for extrapolation in this region, with values of infinity predicted in the center of the plot (Figure 13). To see if SOAP could extrapolate a small distance from the data, the prediction area was restricted to a nonconvex bounding polygon close-fitting to the data. No values of infinity were predicted but six points had predictions > 10\(^3\), which is well outside the data range (supplementary

| Method | Knot number | RSS    | QAIC \(_c\) |
|--------|-------------|--------|-------------|
| TPS    | 60          | 46169  | 788         |
| GLTPS  | 40          | 46911  | –           |
| SOAP   | 60          | 34487  | 844         |
| CReSS  | 45          | 41836  | 104         |

Table 4. Knots chosen, RSS, and QAIC \(_c\) scores for the killer whale data. Note that the QAIC \(_c\) score for CReSS is a weighted average.
Figure 13. Plots of predictions of group size for the killer whale data using (a) TPS, (b) GLTPS, (c) SOAP, and (d) CReSS. The range is wider for GLTPS and SOAP to accommodate higher group size predictions. Contours are in increments of 5 animals with bold lines every 20. Gray dots represent points with group size >100 and black dots represent points where a value of infinity was predicted. Color plots may be found in Supplementary Figure 10.

Figure 11). The best CReSS models were two locally acting models with \( k = 45 \) and \( r = 0.38 \) and 0.46. CReSS fits the data better than TPS and GLTPS and predicts a surface that would seem more realistic. Hotspots are produced in areas where there are large group sizes in the data and the surface declines to approximately zero elsewhere.

### 5. CONCLUSIONS

In this article, we present a novel hybrid of three techniques used in spatial modeling, which has been developed with the goal of accurately mapping animal densities in complex regions. This spatial smoother, CReSS, incorporates locally varying basis functions and an improved geodesic distance estimate in a model averaging framework to accommodate local smoothing requirements; alleviate reinforcement problems around exclusion zones; and yield more stable results. The model averaging framework means CReSS is reasonably...
robust, in terms of its sensitivity to small changes in particular model parameters such as the basis range parameter or the number of knots.

In comparison with GLTPS, CReSS can be used for domains both with and without islands and is more applicable to general statistical models. Furthermore, CReSS is both easier to parameterize than SOAP, particularly with increasing boundary loops, and deals more naturally with boundaries of the survey region that are not delineated by a coastline.

We have attempted to provide the best results from all methods for comparison, frequently trialing many parameterizations—many more than would be considered in general use. CReSS is very simple to implement and is comparable with, and often better than, other methods in all examples shown.

The issue of knot selection is common to all the methods seen in this article. Here, the knots are chosen using a space filling algorithm that does not necessarily allow surface flexibility in the areas where it is most required, particularly for the sparse killer whale test dataset. In the one-dimensional case, a new Spatially Adaptive Local Smoothing Algorithm (Walker et al. 2010) has proved to be very effective for knot placement and is currently being adapted for use in two dimensions and incorporated into the CReSS method.

SUPPLEMENTARY MATERIALS

Supplementary tables and figures referenced in Sections 2–4 are available at the web link given. We have also provided code for calculation of geodesic distance, and implementation of the CReSS method on the palm simulation.

ACKNOWLEDGMENTS

The authors gratefully acknowledge NERC for funding and John Harwood for advice.

REFERENCES

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997), “Model Selection: An Integral Part of Inference,” Biometrics, 53, 603–618. [347]

Burnham, K. P., and Anderson, D. R. (2002), Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.), New York: Springer. [347]

Claeskens, G., and Hjort, N. L. (2009), Model Selection and Model Averaging: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press. [347]

Cressie, N. (1990), “The Origins of Kriging,” Mathematical Geology, 22, 239–252. [340]

Curriero, F. (2007), “On the Use of Non-Euclidean Distance Measures in Geostatistics,” Mathematical Geology, 38, 907–927. [341]

Floyd, R. W. (1962), “Algorithm 97: Shortest Path,” Communications of the ACM, 5, 345. [344]

Fornberg, B., Driscoll, T., Wright, G., and Charles, R. (2002), “Observations on the Behavior of Radial Basis Function Approximations Near Boundaries,” Computers & Mathematics with Applications, 43, 473–490. [345]
Furrer, R., Nychka, D., and Sain, S. (2010), FIELDS: Tools for Spatial Data, R package version 6.3. Available at http://CRAN.R-project.org/package=fields. [347]

Green, P. J., and Silverman, B. W. (1994), Nonparametric Regression and Generalised Linear Models (A Roughness Penalty Approach), England: Chapman and Hall. [343]

Harder, R. L., and Desmarais, R. N. (1972), “Interpolation Using Surface Splines,” Journal of Aircraft, 9, 189–191. [340,343]

Hastie, T. J., and Tibshirani, R. J. (1990), Generalized Additive Models, London: Chapman and Hall. [342,343]

Hurvich, C. M., and Tsai, C.-L. (1989), “Regression and Time Series Model Selection in Small Samples,” Biometrika, 76, 297–307. [347]

John, P. W. M., Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1995), “Minimax Distance Designs in Two-Level Factorial Experiments,” Journal of Statistical Planning and Inference, 44, 249–263. [347]

Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992), “Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach With Case Studies,” Ecological Monographs, 62, 67–118. [357]

McCullagh, P., and Nelder, J. A. (1989), Generalized Linear Models (2nd ed.), London: Chapman and Hall. [342]

McMorrow, B. (2004), “Palm Jumeirah Photograph.” Available at http://www.pbase.com/bmcmorrow. [341]

Ramsay, T. O. (2002), “Spline Smoothing Over Difficult Regions,” Journal of the Royal Statistical Society, Series B, 64, 307–319. [341,342,349,350]

Rathbun, S. L. (1998), “Spatial Modelling in Irregularly Shaped Regions: Kriging Estuaries,” Environmetrics, 9, 109–129. [341,346]

R Development Core Team. (2009), R: A Language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org. [343]

Schwarz, G. (1978), “Estimating the Dimension of a Model,” The Annals of Statistics, 6, 461–464. [347]

Walker, C., Mackenzie, M., Donovan, C., and O’Sullivan, M. (2010), “SALSA—A Spatially Adaptive Local Smoothing Algorithm,” Journal of Statistical Computation and Simulation, 81, 179–191. [359]

Wang, H., and Ranalli, M. G. (2007), “Low-Rank Smoothing Splines on Complicated Domains,” Biometrics, 63, 209–217. [341,342,344,347,349,350]

Wilcoxon, F. (1945), “Individual Comparisons by Ranking Methods,” Biometrics Bulletin, 1, 80–83. [349]

Wood, S. N. (2003), “Thin Plate Regression Splines,” Journal of the Royal Statistical Society, Series B, 65, 95–114. [343]

——— (2010), SOAP: Soap Film Smoothing, R package version 0.1-5. Available at: http://www.maths.bath.ac.uk/~sw283/simon/soap_0.1-5.tar.gz [343]

Wood, S. N., Bravington, M. V., and Hedley, S. L. (2008), “Soap Film Smoothing,” Journal of the Royal Statistical Society, Series B, 70, 931–955. [342,343,350]