Morse-Smale Regression

Samuel Gerber, University of Utah
Oliver Rübel, Lawrence Berkeley National Laboratory
Peer-Timo Bremer, Lawrence Livermore National Laboratory
Valerio Pascucci, University of Utah
Ross T. Whitaker, University of Utah

December 14, 2011

Abstract

This paper introduces a novel partition-based regression approach that incorporates topological information. Partition-based regression typically introduce a quality-of-fit-driven decomposition of the domain. The emphasis in this work is on a topologically meaningful segmentation. Thus, the proposed regression approach is based on a segmentation induced by a discrete approximation of the Morse-Smale complex. This yields a segmentation with partitions corresponding to regions of the function with a single minimum and maximum that are often well approximated by a linear model. This approach yields regression models that are amenable to interpretation and have good predictive capacity. Typically, regression estimates are quantified by their geometrical accuracy. For the proposed regression, an important aspect is the quality of the segmentation itself. Thus, this paper introduces a new criterion that measures the topological accuracy of the estimate. The topological accuracy provides a complementary measure to the classical geometrical error measures and is very sensitive to over-fitting. The Morse-Smale regression is compared to state-of-the-art approaches in terms of geometry and topology and yields comparable or improved fits in many cases. Finally, a detailed study on climate-simulation data demonstrates the application of the Morse-Smale regression. Supplementary materials are available online and contain an implementation of the proposed approach in the R package *msr*, an analysis and simulations on the stability of the Morse-Smale complex approximation and additional tables for the climate-simulation study.

1 Introduction

In the visualization community, the Morse-Smale complex has been successfully applied to visualize structures in data sets from medicine (Carr et al., 2004), to physics (Laney et al., 2006; Bremer et al., 2010; Gerber et al., 2010) and material science (Gyulassy et al., 2007). Gerber et al. (2010) use the Morse-Smale complex to build abstract two dimensional representations for high dimensional scalar functions. This paper builds on the basic idea in (Gerber et al., 2010) and exploits the Morse-Smale complex for partition based regression.

Partition-based regression provides a flexible trade-off between the simplicity and interpretability of parametric models and the predictive capabilities of non-parametric methods. A classical

approach to partition-based regression are regression trees, e.g. CART (Breiman et al., 1984). In regression trees, the domain is split recursively into multiple partitions and low-order parametric models are fitted within each partition. Splits are introduced based on the improvement of a cost function using a discrete search over possible splitting points. The cost function is formulated to yield a model with smallest possible prediction error. Typically, a large regression tree is constructed which is subsequently pruned using cross-validation to decide whether to keep or discard partitions. Several variations on this approach exist. CART originally proposed a piecewise constant model, treed regression (Alexander and Grimshaw, 1996) proposed linear models, and MARS (Friedman, 1991) uses regression splines of multiple orders for fitting in each partition. SUPPORT (Chaudhuri et al., 1994) introduces splits based on significance testing and a look-ahead procedure to control tree size with polynomial models in each partition. BART (Chipman et al., 2010) introduces a Bayesian formulation over the tree structure to guide splitting and control tree size.

The proposed Morse-Smale regression is, in spirit, similar to the principal Hessian direction (PHD) tree regression (Li et al., 2000). Li et al. (2000) introduced an alternative viewpoint for constructing the domain decomposition. Instead of focusing on a quality of fit measure for splitting, the goal is to capture the geometry of the regression surface, i.e. the main target is an interpretable and exploratory data model. Splitting directions are based on curvature, i.e. the regression surface is split orthogonal to the direction of highest curvature. Since the regression surface is not a priori known, principal Hessian directions are used to estimate the direction of highest curvature. This procedure is recursively applied to each partition starting from the complete domain. This ultimately results in partitions that are well approximated by linear models.

The Morse-Smale complex regression is, as the PHD approach, geometrically motivated. However, instead of focusing on splitting across high curvature directions, the Morse-Smale regression uses the segmentation of the domain induced by the Morse-Smale complex of the regression surface. The Morse-Smale complex segments the domain into regions of uniform gradient flow such that the interior of each partition does not contain any critical points. Thus, each partition is associated with exactly one minimum and maximum located on the boundary of the partition. This suggests to fit simple parametric models within each partition. As for PHD regression, the regression surface is not a priori known and the Morse-Smale complex is approximated directly from the data sample.

In most tree-based regression methods, the splitting criteria is a greedy, local operation that does not necessarily lead to a good global fit or meaningful partitions. The Morse-Smale complex considers the global structure of the regression surface for partitioning. This leads to topologically

meaningful partitions and can, for some regression problems, improve the fit. However, the regression will not adapt to features in the data that do not introduce additional partitions. Another difference to other partition-based regression schemes is that the number of partitions is limited. The maximal number of partitions introduced by the Morse-Smale complex is directly related to the number of extrema in the data. This can help to mitigate overfitting.

The focus on a topologically meaningful partition of the domain suggest to consider the quality of the domain decomposition. This paper proposes to measure differences in segmentations based on the F-measure (van Rijsbergen, 1979). Thus, the Morse-Smale complex introduced by a regression estimate can be compared to the ground truth segmentation (if available) and provides a quantitative measure of the topological accuracy of a regression estimate. This introduces a measure that is very sensitive to overfitting. In the case where a high frequency approximation follows the true underlying function closely, the RMSE can be arbitrarily small. However, the artificially introduced extrema can result in a very different Morse-Smale segmentation.

Section 2 describes the Morse-Smale complex, highlights properties useful for partition-based regression and introduces the computational procedure to approximate the Morse-Smale complex from a finite sample. The Morse-Smale complex partition-based regression is introduced in Section 3. Section 4 presents experimental results on synthetic and real-world data that demonstrate the properties of the Morse-Smale regression analysis and compares it to state-of-the-art approaches. Section 5 applies the Morse-Smale regression to explore the parameter space of the CAM (Community Atmosphere Model) climate simulation (McCaa et al., 2004). The proposed regression approach, as well as the work in (Gerber et al., 2010) , is implemented in the R package *msr* (Gerber et al., 2011). The R package is documented in detail in (Gerber and Potter, 2011).

2   The Morse-Smale Complex

In Morse theory, the Morse-Smale complex relates the number and connectivity of critical points of a function $f : \mathcal{M} \to \mathbf{R}$ to the topology of the manifold $\mathcal{M}$. This work is not interested in the information on the topology of $\mathcal{M}$ but exploits the Morse-Smale complex to glean information about the geometry of f. This section briefly introduces the main concepts and highlights properties relevant for the proposed regression. For a treatment of the topological aspects of the Morse-Smale complex the literature by Morse (1925) and Milnor (1963) is recommended.

Let $\mathcal{M}$ be a smooth, compact, p-dimensional manifold. A smooth function $f : \mathcal{M} \mapsto \mathbb{R}$ is *Morse* if for all critical points x of f the Hessian matrix $Hf(x)$ is not singular. An *integral line* $\lambda : \mathbb{R} \mapsto \mathcal{M}$ is a curve in $\mathcal{M}$ with $\frac{d\lambda}{ds}(s) = \nabla f(\lambda(s))$. Define $\mathrm{src}(\lambda) = \lim_{s \to -\infty} \lambda(s)$ and $\mathrm{sink}(\lambda) = \lim_{s \to \infty} \lambda(s)$

as source and sink of the integral line, respectively. Note, src and sink are both, by definition, critical points of f. Define the ascending and descending (also referred to as stable and unstable) manifolds of a critical point x as

$$A(x) = \{\lambda : \text{src}(\lambda) = x\} \tag{1}$$

$$D(x) = \{\lambda : \text{sink}(\lambda) = x\} . \tag{2}$$

A Morse function f is Morse-Smale if the ascending and descending manifolds intersect transversally only. The Morse-Smale complex is the set of intersections $A(x_i) \cap D(x_j)$ over all critical points x_i, x_j. The Morse-Smale complex includes regions (i.e., sub-manifolds of $\mathcal{M}$) of dimensions 0 through p and partitions $\mathcal{M}$. The 0 through $p - 1$ dimensional components of the Morse-Smale complex delineate the boundaries of the p-dimensional regions, i.e. the partitions on $\mathcal{M}$.

Figure 1(a) shows the Morse-Smale complex on a function with four maxima and nine minima. The Morse-Smale complex decomposition guarantees that within each partition the function has no critical points and exactly one local minimum and maximum are located on the boundary. Thus, in many cases, the function within a partition is well approximated by a linear model, as illustrated in Figure 1(b).

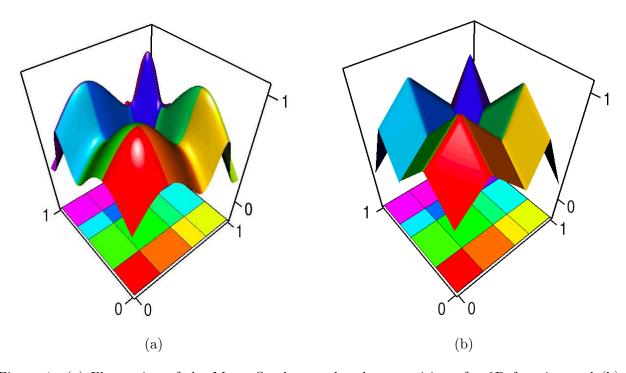
(a)   (b)

Figure 1: (a) Illustration of the Morse-Smale complex decomposition of a 2D function and (b) piecewise linear model fit.

In principle, higher-order parametric or nonparametric models could be fitted in each partition. However, this work restricts attention to linear models. One advantage of fitting linear models is the ease of interpretation of each partition. In general, it is difficult to extract meaningful insight from non-axis aligned partitions. However, the Morse-Smale complex induced segmentation yields topologically meaningful entities that facilitate interpretation in terms of watershed regions and their difference in slopes, as well as number and location of extrema. Linear models also guarantee that no extremal points are introduced in the interior of a partition, as dictated by the Morse-Smale complex.

2.1 Persistence Simplification and Hierarchical Partitions

The Morse-Smale complex introduces a measure of the *significance* of each extremal point, referred to as *persistence*. Persistence is a measure of the amount of change in the function f required to remove a critical point and, thus, merge or cancel two critical points. Note, persistence describes the significance of an extremal point in geometric terms and not in the statistical sense of a hypothesis test.

Let x_i be the critical points of f. Define $s(x_i)$ as the set of critical points that have a direct integral line connecting to x_i. Let $n(x_i) = \arg\min_{x_j \in s(x_i)} \|f(x_i) - f(x_j)\|$. The associated persistence of a critical point x_i is then defined as $p(x_i) = \|f(x_i) - f(n(x_i))\|$. This definition is motivated by the amount of change of f in L_∞ required such that the critical point pair $(x_i, n(x_i))$ is merged into a single critical point or canceled as illustrated in Figure 2. Thus, for extrema, persistence is the minimal difference between the extrema and its directly connected saddle points.

Recursively removing the critical point with minimal persistence leads to a nested series of successively simplified Morse-Smale complicies, also called a filtration (Edelsbrunner et al., 2006; Chazal et al., 2009). At each level, some of the partitions induced by the Morse-Smale complex are merged into a single partition until the Morse-Smale partitioning consists of only a single partition (i.e., the entire input domain).

The concept of persistence introduces an intrinsic notion of scale; a hierarchy of segmentations with each subsequent segmentation subsuming the previous one. In the context of regression, the persistence hierarchy of the Morse-Smale complex corresponds to a hierarchy of increasingly flexible models with decreasing persistence; from a single linear model at the highest persistence to a model with multiple extrema.

The persistence itself can be used to select a segmentation at an appropriate scale, e.g. if the noise level is known, partitions introduced by extrema with persistence lower than the noise
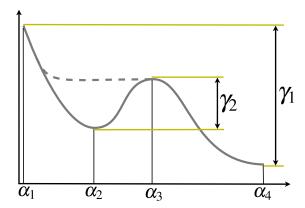
Figure 2: At the highest persistence level the Morse-Smale complex of the solid gray function has 3 partitions $[\alpha_1, \alpha_2]$, $[\alpha_2, \alpha_3]$ and $[\alpha_3, \alpha_4]$. The dashed gray line indicates the change required to remove critical point pair (α_2, α_3) with persistence γ_2 resulting in a Morse-Smale complex with a single partition $[\alpha_1, \alpha_4]$.

level can be ignored. However, for regression the model fit at each persistence level provides an alternative approach to select an appropriate scale. At each level, the fitted model can be analyzed and compared to the model at the subsequent persistence level, which subsumes the current model. Thus, model selection is simplified to a linear search over the models induced by the hierarchy of segmentations. This approach to model selection is described in detail in Section 3.4.

For model selection, the restriction to linear models is important since it guarantees that the fit does not introduce extrema within a partition. For higher order models or non-parametric fits it would be difficult — there is no explicit boundary representation for the partitions — to determine whether additional extrema are introduced within a partition. Thus, a higher-order model fit does not indicate if the Morse-Smale complex at a given persistence level still contains significant extrema within partitions.

2.2 Computation of the Morse-Smale Complex

The construction of the Morse-Smale complex on discrete structures (i.e., a set of points with connectivity information) has received much attention especially in two and three dimensions (Edelsbrunner et al., 2003; Carr et al., 2003; Gyulassy et al., 2007). However, for regression, the Morse-Smale complex needs to be estimated from unorganized scattered data (i.e., a set of points without neighborhood information) in higher dimensions, which was only recently considered (Chazal et al., 2009). The algorithm described here follows closely the work by Gerber et al. (2010) .

The definition of the Morse-Smale complex in terms of ascending and descending manifolds leads to a direct algorithm. Given a data set $X = \{x_1, \ldots, x_n\}$ and associated function values $Y = \{f(x_1), \ldots, f(x_n)\}$, possibly corrupted by noise, the source and sink for each data point x_i, and thus the corresponding partition, is determined by following the gradient at x_i. This direct approach requires to estimate the gradient at x_i — essentially an estimate of the regression surface — and to trace for each point the integral line that passes through it. Requiring the regression surface apriori defeats the purpose of using the Morse-Smale complex to build a partition-based regression. However, to partition the domain, the only relevant information is the source and sink of each point and not the complete integral line. This section describes an algorithm to compute the source and sink for each point x_i without first estimating the regression surface.

Gerber et al. (2010) approximate the domain via a k nearest neighbor graph. This leads to Algorithm 1 to approximate the integral lines. The algorithm follows paths of steepest ascent and descent based on the connectivity of the graph; essentially a variant of the quick-shift algorithm (Vedaldi and Soatto, 2008).

Algorithm 1 Compute Morse-Smale complex partitions

$\{(x_i, f(x_i))\}_i^n\}$ { Data set with observations x_i and function value $f(x_i)$}
$\mathrm{adj}(x_i) = \{x_j : x_i \in \mathrm{knn}(x_j), x_j \in \mathrm{knn}(x_i)\}$ {Adjacencies of k-nearest neighbor graph}
$p_a(x_i) = \arg\max_{x_j \in \mathrm{adj}(x_i)} f(x_j) - f(x_i)$ {Direction of steepest ascent}
$p_d(x_i) = \arg\max_{x_j \in \mathrm{adj}(x_i)} f(x_i) - f(x_j)$ {Direction of steepest descent}
for $i = 1 \to n$ **do** {Approximate integral lines for each data point}
   $x_a = x_i$
   while $p_a(x_a) \neq x_a$ **do** {Find source for x_i}
     $x_a = p_a(x_a)$
   end while
   $x_d = x_i$
   while $p_d(x_d) \neq x_d$ **do** {Find sink for x_i}
     $x_d = p_d(x_d)$
   end while
   $\mathrm{partition}(x_i) = (x_a, x_d)$ {Assign x_i to partition with maximum x_a and minimum x_d}
end for
return $\mathrm{partition}(\cdot)$ {Data structure with partition assignments}

Algorithm 1 assigns each point x_i to a p-dimensional component, i.e., a partition, of the Morse-Smale complex. The result is a set of l partitions $\mathcal{C} = \{C_1, \ldots, C_l\}$ such that $\bigcup_i C_i = \{x_i\}_1^n$ and $C_j \cap C_i = \emptyset \; \forall i \neq j$. Computing the persistence-based hierarchy of the Morse-Smale complex, requires an approximation of saddle point values between neighboring partitions. This is achieved by inspecting the points of the edges in nearest neighbor graph that cross partition boundaries. Saddle point values are then defined as the minimal and maximal value of the function values

of the vertices of crossing edges. This leads to Algorithm 2 for computing the persistence of an extremal point and Algorithm 3 to merge an extrema pair. Applying Algorithm 3 recursively to

---

Algorithm 2 Compute persistence of an extremal point x

$C_i, i = 1 \ldots l_x$ {Set of partitions that contain extrema x}
$p_x = \infty$ {Persistence of extrema x}
for $i = 1 \to l_x$ **do**
  for $x_1 \in C_i$ **do**
    $t_x = 0$
    for $x_2 \in C_j$ **do**
      if $x_1 \in \text{adj}(x_2)$ or $x_2 \in \text{adj}(x_1)$ **then** {Is (x_1, x_2) an edge?}
        if $\text{partition}(x1) \neq \text{partition}(x2)$ **then** {Is edge crossing partitions?}
          $\Delta = \max(|f(x) - f(x_1)|, |f(x) - f(x_2)|)$
          if $t_x < \Delta$ **then** {Update closest saddle?}
            $t_x = \Delta$
          end if
        end if
      end if
    end for
  end for
  if $t_x < p_x$ **then**
    $p_x = t_x$
    $p = (\text{extrema}(C_i), \text{extrema}(C_j), p_x)$ {Store extrema pair and its persistence}
  end if
end for
return p

---

the currently lowest persistence extrema pair yields the persistence hierarchy of Section 2.1. Note that after running Algorithm 3 the persistences need to be updated since one of the extrema is merged to an extrema with a higher/lower (for maxima/minima) function value.

For a Morse function $f : \mathcal{M} \to \mathbf{R}$ and probability measures μ on $\mathcal{M}$ such that $\mu(X) = 0$ for $X = \{x : f(x) = c\}$, i.e., level sets of f have μ-measure zero, Algorithm 1 applied to a random sample of μ will always find paths ending in local minima and maxima. However, the choice of the number of nearest neighbors affect the Morse-Smale complex approximation. A large number of nearest neighbors can act as a smoothing of f and remove some of the effects of noise. However, too many nearest neighbors increase the potential of shortcuts that result in the merge of two valid partitions. Too few nearest neighbors increase the potential of introducing *artificial* extrema, either due to noise or connecting only to a set of nearest neighbors that are not representative of the directional derivatives of f.

The supplemental material "Stability of The Morse-Smale Complex" analyzes the stability of the Morse-Samle complex approximation with respect to noise, number of nearest neighbors

Algorithm 3 Merge partitions

$x_1 = \text{extrema}(C_i), x_2 = \text{extrema}(C_j)$ {Extrema pair to merge ordered such that $f(x_1) < f(x_2)$}
  for $i = 1 \to n$ **do** {Update partitions containing extrema x_1, x_2}
    $(x_a, x_d) = \text{partition}(x_i)$ {Maximum and minimum of the partition of x_i}
    if x_1, x_2 are maxima **then**
      if $x_1 == x_a$ **then** {Update partition?}
        $\text{partition}(x_i) = (x_2, x_d)$
      end if
    else
      if $x_2 == x_d$ **then** {Update partition?}
        $\text{partition}(x_i) = (x_a, x_1)$
      end if
    end if
  end for
  return $\text{partition}(\cdot)$

---

and dimensionality. The analysis shows that the proposed Morse-Smale complex algorithms are effective for large signal to noise ratios only. That is, if the smallest true persistence of f is an order of magnitude larger than the standard deviation of the noise, then the Morse-Smale complex computation typically provides, for large enough sample sizes, a good approximation to the true segmentation.

3   Morse-Smale Complex Regression

This work employs a linear regression fit withing each partition and considers — following the SUPPORT (Chaudhuri et al., 1994) model — two approaches: a piece-wise linear model (approach I), and a sum of weighted linear models (approach II) with weights based on estimating partition probabilities. Approach I enables the statistical validation of the Morse-Smale complex in terms of differences in slope and intercept of the partitions. In approach II, some of the capability of distinguishing different partitions is sacrificed to obtain a better predictive model. Thus, the Morse-Smale regression serves two purposes, either a detailed analysis of the segmentation or a novel approach to predictive modeling that, for some regression problems, can improve performance.

The proposed regression scheme is based on the geometry of the data. Rescaling effects the k nearest neighbor graph and thus, the Morse-Smale complex computation. Hence, in practice it is important to scale the variables such that they are commensurate. The proposed method cannot deal with categorical variables unless a metric commensurate with all other variables is introduced on the categories.

9

3.1 Partition Prediction

The segmentation by the Morse-Smale complex approximation provides the partition assignments, i.e., labels, for the input data, but not the complete domain. This information is sufficient to fit a linear model in each partition and infer information about differences in partitions. However, prediction or estimation of the partition probabilities in approach II requires a segmentation that covers the complete domain.

The approach in this paper is to build a probabilistic classifier from the partition assignments introduced by the Morse-Smale complex. The goal is to estimate $P(\mathcal{C}|X)$, the probability distribution over the set of partitions $\mathcal{C} = \{C_1, \ldots, C_l\}$, given the location X. In this work, two ways for constructing the partition probabilities are considered. The first approach estimates the probability distribution $P(X|C_i)$ with a kernel density estimator over the observations in partition i

$$p_i(x) = \frac{1}{\|C_i\|} \sum_{z \in C_i} K(x, z) . \tag{3}$$

Then, Bayes theorem yields the partition probabilities $P(C_i|X = x) = \frac{p_i(x)P(C_i)}{\sum_{j=1}^{l} p_j(x)P(C_j)}$ with $P(C_i) = \frac{\|C_i\|}{\sum_{j=1}^{l} \|C_j\|}$. The second approach employs a support vector machine (SVM) to build a multi-class classifier using the one-against-one approach described in (Hsu and Lin, 2002). The SVM is trained on the partition assignments of the Morse-Smale complex. To estimate partition probabilities from the classifier, a logistic regression is fitted to the decision values of the SVM, as described in Platt (1999).

3.2 Approach I: Piece-wise Linear Model

The first approach uses hard boundaries between the partitions and estimates a linear regression in each partition

$$\hat{f}_l(x) = a_j + b_j x \quad \text{if } x \in C_j, \text{ j=1} \ldots \text{l}, \tag{4}$$

resulting in a discontinuous piece-wise linear fit. This can be modeled as a multiple linear regression with the partition index as a categorical variable to capture interactions between function variables and partitions, i.e., per partition differences in slope. Thus, an ANOVA can be used to provides insight into the significance and differences of the partitions.

The prediction of unseen data points involves two steps in this case. First, the partition ID of the data point has to be predicted as described in section 3.1. Then, the function value is estimated with the linear model corresponding to the predicted partition.

3.3 Approach II: Additive Weighted Linear Model

The second approach estimates a sum of weighted linear models, with weights based on a probabilistic interpretation of the segmentation by the Morse-Smale complex, i.e., a soft assignment of each point based on the probability of being in each partition.

$$\hat{f}_w(x) = \sum_{j=1}^{l} w_j(x)(a_j + b_j x) \tag{5}$$

with $w_j(x) = P(C = C_i | X = x)$ estimated as described in section 3.1. Note that if $w_j(x)$ is C^m continuous then the resulting fit will also be C^m continuous. This formulation leads to a linear model with $(d+1)k$ coefficients, i.e., each partition introduces d columns for the weighted points $w_j(x)x$ plus an additional column $w_j(x)$ for the intercept. As in the piece-wise linear case, prediction requires first the estimation of the partition probabilities $w_j(x)$ for evaluation of the model.

3.4 Hierarchical Model Selection

For noisy observations, the MS-complex is very likely to over-segment the region due to the noise that can introduce additional extrema. Furthermore, the nearest-neighbor based computation can introduce artificial extrema if all of the neighbors connect exclusively to observations with higher/lower function values. Thus, an important aspect of the proposed regression is to infer the correct persistence level at which the partitions of the Morse-Smale complex are meaningful. The persistence, as described in Section 2.1, introduces a hierarchy of segmentations based on the geometrical significance of the associated extrema. Thus, the challenge is to select a segmentation at the right persistence level to avoid over- or under-fitting.

Let k_i be the number of partitions at persistence level i, starting with $i = 1$, the highest persistence yields $k_1 = 1$ partition (a single minimum and maximum) to lowest persistence level l with $k_l = K$ partitions. Thus, the number of partitions is given by the selected persistence level of the Morse-Smale complex. This provides a natural approach to select the persistence level, i.e., the number of partitions, for the Morse-Smale regression. Decreasing persistence levels introduce regression models of increasing complexity. Figure 3 illustrates the persistence hierarchy by showing the piece-wise linear models resulting from different persistence levels on a function with four peaks. No sophisticated search or optimization strategies are required since only a linear sequence of models, with each subsequent model subsuming the previous one, is considered. Thus, model selection approaches, such as cross-validated mean squared error, Akaike/Bayesian

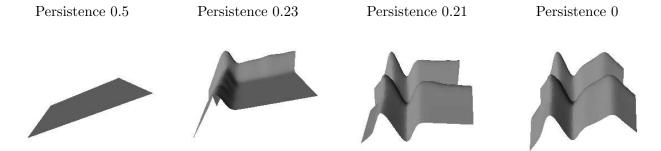Persistence 0.5   Persistence 0.23   Persistence 0.21   Persistence 0

Figure 3: A hierarchy of regression models induced by the persistence simplification of the Morse-Smale complex. Starting at the highest persistence, with a single minimum and maximum, on the left, to multiple extrema, at zero persistence, on the right.

information criterion (AIC/BIC) or adjusted R^2, can be employed in a sequential fashion. In this work, cross-validated mean squared error is used for model selection in approach II and ANOVA and BIC for approach I. The ANOVA analysis provides the significance of the introduced partitions at each level (sequential F-ratio test) as well as information on differences in partitions.

The Morse-Smale regression considers a very restrictive set of admissible functions (limited by the geometry of f plus noise), i.e., there is a fixed number of functions, defined by the number of persistence levels, for selection. The Morse-Smale regression is not immune to over-fitting; noise can introduce partitions not supported by the geometry of f. However, the linear hierarchy of models simplifies model selection and the restrictive class of functions reduces over-fitting.

3.5   Measuring Topological Accuracy

In many scenarios, a topological accurate fit is important. One approach to measure agreement in topology is to quantify how much the segmentation induced by the Morse-Smale complex of the estimated function agrees with the Morse-Smale segmentation of the true function. Various measures for comparing segmentations have been proposed. Here, the F-Measure (van Rijsbergen, 1979) is employed.

The F-Measure summarizes precision P and recall R. Here, the F-measure is used to describe the similarity of two domain decompositions. Let $X = \{x_1, \ldots, x_n\}$ be a sample with a ground truth domain decomposition $T = \{T_1, \ldots, T_m\}$ and estimated domain decomposition $C = \{C_1, \ldots, C_l\}$. For each class pair T_i and C_i, the precision and recall are defined as $P(T_i, C_i) = \frac{|T_i \cap C_i|}{|C_i|}$ and $R(T_i, C_i) = \frac{|T_i \cap C_i|}{|T_i|}$. The F-Measure is then defined as

$$F(T_i, C_i) = 2 \frac{P(T_i, C_i) R(T_i, C_i)}{P(T_i, C_i) + R(T_i, C_i)} \tag{6}$$

and the F-measure over all partitions is

$$F(T, C) = \frac{1}{n} \sum_{i}^{\max l,m} |T_i| F(T_i, C_i) \tag{7}$$

with $C_i = \emptyset$, $\forall i > l$ and $T_i = \emptyset$, $\forall i > m$. The F-Measure is dependent on the partition co-assignments, i.e., a permutation of partition labels yields a different F-Measure. Hence, to compute the F-Measure in equation (7), all possible combinations of label co-assignments are considered, and the highest F-Measure is retained.

To compare the Morse-Smale complex of an estimate $\hat{f} : \Omega \to \mathbf{R}$ to the Morse-Smale complex of the ground truth $f : \Omega \to \mathbf{R}$, a dense sample X of the domain Ω is randomly drawn (e.g. uniform on Ω) or deterministically generated (e.g., a grid). Now, the ground truth decomposition T on X is assigned based on the true Morse-Smale complex of f. If it is not possible to construct the ground truth Morse-Smale complex of f, T is computed by applying Algorithm 1 to $(X, f(X))$. The domain composition C, i.e., the Morse-Smale complex of $\hat{f}$, is computed by applying Algorithm 1 to $(X, \hat{f}(X))$. Finally, the F-measure is computed for the two label assignments T and C on X. Note, that this approach can also be used to compare two estimates $\hat{f}_1$ and $\hat{f}_2$.

An alternative approach is to compare the segmentations introduced during fitting a partition-based regression to the Morse-Smale complex of the estimate or the ground truth. This focuses on the topological quality of the segmentation induced by a particular partition-based regression scheme, i.e., whether the domain decomposition is topologically meaningful. Since previous partition based regression schemes do not focus on topological meaningful partitions, the comparison in Section 4.5 restricts attention to comparing the Morse-Smale complex of the estimate (instead of the partitions used for fitting) to the ground truth.

It is not feasible to build an algorithm that optimizes this criterion directly because, almost always, the ground truth Morse-Smale complex is unknown. However, the topological accuracy provides another tool to illustrate the behavior of different regression schemes on known test functions or to compare the topology of two different estimates.

4 Simulation Studies

This section focuses on controlled experiments using explicit functions described in Section 4.1. The goal is to demonstrate and compare the properties of the Morse-Smale regression to other state-of-the-art regression approaches (Section 4.2). The experiments explore the sensitivity of the persistence-based model selection (Section 4.3), compare prediction performance (Section 4.4) and analyze topological accuracy (Section 4.5).

4.1 Data Sets

The simulation study is based on the following three functions.

Four-peaks is a two-dimensional, additive separable function of four Gaussian peaks

$$f(x) = \frac{1}{2}\left(e^{-(x_1-\frac{1}{4})^2/0.3^2} + e^{-(x_2-\frac{1}{4})^2/0.3^2} + e^{-(x_1-\frac{3}{4})^2/0.1^2} + e^{-(x_2-\frac{3}{4})^2/0.1^2}\right). \tag{8}$$

The function has four maxima and nine minima, and the Morse-Smale decomposition of f consists of 16 rectangular partitions, as illustrated in Figure 1.

For the experiments, two versions of f are used. First, the domain is uniformly sampled on $[0,1] \times [0,1]$. Second, let R be the 2D, 45-degree rotation matrix, then the domain is uniformly sampled on $R([0,1] \times [0,1])$ and $f_r(x) = f(R^{-1}x)$, i.e., the same function, but the domain rotated by 45 degrees. In this case, the split directions of piece-wise regression methods, such as MARS and regression trees, are no longer aligned with features of f.

Diagonal is a d dimensional function $g_{p,d}(x) : [0, \frac{1}{\sqrt{d}}]^d \mapsto \mathbf{R}$:

$$g_{p,d}(x) = -0.5cos(2\sigma p <x, v> 4\pi)e^{-\frac{||X||^2-<x,v>}{\sigma^2}} \tag{9}$$

with $\sigma = 0.5\sqrt{d}$ and v the unit length vector along the diagonal. Thus $g_{p,d}(x)$ is a cosine along the diagonal of the d-dimensional hyper-cube with exponentially decreasing magnitude of the cosine orthogonal to the diagonal. The function has $p-2$ maxima and $p-1$ minima. The Morse-Smale complex of $g_{p,d}$ consists of p partitions defined by $p-1$ hyper-planes orthogonal to the diagonal.

Optimization energy: This function is based on a real-world camera-estimation problem. Given two images with point correspondences, the goal is to estimate the translation and rotation of two calibrated cameras. This problem can be formulated as a minimization of the total squared *algebraic error* (Hartley and Zisserman, 2003):

$$h(R,t) = f(E) = \sum_i (x_i^T E x_i')^2, \tag{10}$$

with $x_i = [x_{i1} \ x_{i2} \ 1]^T$ and $x_i' = [x_{i1}' \ x_{i2}' \ 1]^T$ being corresponding points on the image plane defined in the respective camera coordinates. The *essential matrix* $E = [t]_\times R$ is a 3×3 rank-2 matrix. In this formulation, the translation between the two cameras is described by the unit vector t, and the relative camera orientation is defined by the orthogonal rotation matrix R. Both t and R are expressed in the coordinate frame of x. Due to the formulation of the problem, E is guaranteed to

have only 5 degrees of freedom: 3 to describe the rotation and 2 to determine the translation up to scale. Hence, h defines a 5D manifold embedded in 9D space. For more detailed information on the definition of this problem, see the manuscript by Lindstrom and Duchaineau (2009).

4.2 Methods and Parameter Tuning

The experiments compare against MARS, boosted trees, SVM regression, and kernel regression. The SVM regression and kernel regression approaches are not interpretable models but can yield, with appropriate tuning, fits with very good prediction performances. The boosted regression trees have been cited as the best *out-of-the-box* method (Hastie et al., 2009). Finally, MARS targets, as the proposed approach, interpretable models with good prediction capacity.

For MARS, the R package earth (Milborrow, 2009) is used with a 10-fold cross validation and hand tuning of the threshold, degree, and maximal number of basis functions. For the SVM regression, the R package e1071 (Dimitriadou et al., 2010) provides an interface to the libsvm library (Chang and Lin, 2001), and the parameters are tuned using the provided tuning facilities in the package. The R package mboost (Hothorn et al., 2010) provides an implementation of gradient boosted regression trees using conditional inference trees (Hothorn et al., 2005) as base learners. For the experiments, multiple boosted trees were fitted, with different number of boosting iterations and tree depths, and the best fit is selected based on cross-validation. For the kernel regression, the locally polynomial regression (Loess) implemented in the basic R statistics package is used with bandwidth selection based on Silverman's rule of thumb (Silvermann, 1986).

The Morse-Smale computation has two free parameters: the number of nearest neighbors and the persistence level. The number of nearest neighbors is set to $5d$, with d the number of variables. For the Morse-Smale regression approach I (msI, piece-wise linear), these are the only parameters to be set. For approach II (msII, additive weighted linear), the partition probabilities are either predicted with a kernel density estimation (msII-kd) or an SVM (msII-svm). For the kernel density estimation, the bandwidth is set as the average k-th nearest-neighbor distance. For the SVM, the parameters are selected automatically through tuning facilities provided in the R SVM package.

4.3 Morse-Smale Persistence Selection

This section evaluates the sensitivity of the persistence-based model selection described in Section 3.4 with respect to noise, nearest-neighbor parameter and dimensionality. Here, the piece-wise linear model is employed and the persistence level is selected based on the model of the persistence hierarchy with the maximal adjusted R^2. The experiments compare the number of minima and

maxima identified by the Morse-Smale computation and persistence selection to the ground truth number of maxima and minima.

Table 1 shows results fort varying number of nearest neighbors and sample sizes on the four peaks function. The Morse-Smale complex computation and persistence selection is relatively stable with respect to the number of nearest neighbors. The results highlight the observation on the proposed algorithm in Section 2.2: For few nearest neighbors, to many extrema are identified since a small number of nearest neighbors is more likely to introduce artificial extrema. This is because (a) there is a smaller chance of introducing *shortcuts* that smooth out the effect of noisy observation and (b) even without noise, the observations can be arranged such that none of the nearest neighbors of a particular observation will introduce a path to an increasing/decreasing observation and thus, is labeled as a maximum/minimum. An increased number of partitions, in turn, can yield better fitting models. A linear model does not do justice to the curved shape of the function within a partition and an additional partition can better approximate this curved shape. With more nearest neighbors, no artificial extrema are introduced and the model selection correctly selects the partition corresponding to the lowest persistence Morse-Smale complex.

	9 Minima				4 Maxima			
	200	500	1000	2000	200	500	1000	2000
10	9.3 (1.6)	9.2 (1.6)	9.8 (1.6)	10.2 (1.7)	4.5 (0.5)	4.6 (0.5)	4.6 (0.5)	4.7 (0.5)
25	8 (0)	9 (0)	9 (0)	9 (0)	4.1 (0.1)	4 (0)	4 (0)	4 (0)
50	3 (0)	8 (0)	9 (0)	9 (0)	3.9 (0.1)	4 (0)	4 (0)	4 (0)

Table 1: Accuracy of the Morse-Smale persistence selection using varying numbers of nearest neighbors k on the four peaks function (9 minima and 4 maxima). The table reports the mean and standard deviation of the number of extrema selected over 100 runs.

Table 2 shows results for different sample sizes and with different levels of Normal distributed noise $\mathcal{N}(0, \sigma)$ for a fixed $k = 25$ number of nearest neighbors. With noisy observations, the increased number of extrema for $\sigma = 0.1$ with growing sample size is again due to a better fit using too many partitions to model non-linearities. The same behavior manifests for $\sigma = 0.25$ but delayed in the number of observations. This indicates that the proposed model selection scheme prevents over-fitting for small sample sizes by using less partitions.

Table 3 shows results on the diagonal function with increasing dimension d, using a fixed number of samples $n = 2000$ and $k = 5d$ nearest neighbors for the Morse-Smale complex computation. In low-dimensions, additional artificial extrema again yield a better fitting model so that the model selection approach identifies too many extrema. This behavior is especially pronounced in the diagonal function with strongly articulated saddle points in the corners orthogonal to the diagonal. For higher dimension, the number of extrema reduces to one minimum and two maxima. This

	9 Minima				4 Maxima			
	200	500	1000	2000	200	500	1000	2000
0	8.8 (0.2)	9 (0)	9 (0)	9(0)	4 (0)	4 (0)	4 (0)	4 (0)
0.1	8 (1.1)	9.6 (1.1)	10.2 (1)	10.8 (1.1)	4.2 (0.5)	4.5 (0.5)	5.3 (0.4)	5.5 (0.5)
0.25	4.2 (1.5)	6.2 (1.4)	6.8 (1.5)	9 (1.3)	2.7 (0.7)	4.3 (0.7)	4.7 (0.7)	5.1 (0.7)

Table 2: Accuracy of the Morse-Smale persistence selection using varying levels of Normal noise with $\sigma \in \{0, 0.1, 0.25\}$ on the four peaks function (9 minima and 4 maxima). The table reports the mean and standard deviation of the number of extrema selected over 100 runs.

is expected since the probability of sampling the corner partitions tends to zero with increased dimensionality.

	3 Minima				
	2	5	10	50	100
0	4 (0.5)	1 (0)	1 (0)	1 (0)	1 (0)
0.1	5.8 (1.3)	1 (1.2)	1 (1.2)	1 (1.2)	1.8 (1.2)
0.25	4.7 (1.8)	1 (1.8)	1.1 (1.8)	1.5 (1.8)	2 (1.8)
	2 Maxima				
	2	5	10	50	100
0	3.8 (2.1)	7.1 (2)	5.8 (2.1)	5.9 (2.2)	5.2 (2.1)
0.1	3.7 (1.5)	5.1 (1.6)	4.8 (1.4)	4.2 (1.5)	2.2 (1.3)
0.25	3.2 (1.4)	4.1 (1.3)	3.8 (1.3)	1.8 (1.2)	2.6 (1.3)

Table 3: Accuracy of the Morse-Smale persistence selection using varying levels of Normal noise with $\sigma \in \{0, 0.1, 0.25\}$ on the diagonal function $g_{(4,d)}$ for varying dimension $d = \{2, 5, 10, 50, 100\}$ (3 minima and 2 maxima). The table reports the mean and standard deviation of the number of extrema selected over 100 runs.

4.4   Prediction Performance

The following experiments analyze and compare the prediction errors of the Morse-Smale regression — using the additive weighted linear model and an SVM for partition prediction — with MARS, boosted trees, SVM regression, and Loess. The mean squared error of the estimate on a sample of the function, possibly contaminated with Normal noise, of size $n = \{128, 256, 512, 1024\}$ is computed on a test set of size 10000.

n	Fourpeaks					Fourpeaks 45° rotated				
	msII-kd	MARS	tree	Loess	SVM	msII-kd	MARS	tree	Loess	SVM
128	0.028	0.003	0.006	0.006	0.022	0.026	0.023	0.030	0.007	0.021
256	0.014	0.001	0.003	0.002	0.019	0.008	0.025	0.024	0.004	0.019
512	0.003	0.001	0.002	0.002	0.013	0.002	0.021	0.022	0.002	0.014
1024	0.002	0.001	0.002	0.002	0.011	0.002	0.021	0.022	0.002	0.012

Table 4: Mean squared error for various sample sizes on four-peaks function with axis and non-axis aligned features.

Table 4 and 5 shows the results on the four-peaks function. MARS and the boosted tree outperform the other approaches in the case of axis aligned features for small sample sizes. With larger samples size, the Morse-Smale regression performs as well as the other partition-based approaches. For non-axis aligned features the Morse-Smale regression outperforms the other partition-based approaches.

n	Fourpeaks					Fourpeaks 45° rotated				
	msII-kd	MARS	tree	Loess	SVM	msII-kd	MARS	tree	Loess	SVM
128	0.024	0.011	0.012	0.092	0.024	0.027	0.038	0.037	0.067	0.022
256	0.024	0.008	0.009	0.010	0.019	0.018	0.025	0.024	0.013	0.019
512	0.009	0.004	0.005	0.005	0.014	0.009	0.021	0.022	0.008	0.018
1024	0.005	0.004	0.005	0.005	0.013	0.005	0.020	0.021	0.004	0.015

Table 5: Mean squared error for various sample sizes on four-peaks function with axis and non-axis aligned features with Normal noise $\sigma = 0.15$ added.

Table 6 shows results on the diagonal function with increasing dimension. Here, MARS and the boosted tree perform badly since a large number of axis-aligned splits is required to approximate the diagonally aligned features. The MS regression performs almost as well as the SVM and Loess regression. The MS regression correctly identifies 4 partitions in the 2D and 4D cases. In higher dimension the probability of an observation in the corner partitions drops sharply and the MS regression detects typically only the two partitions in the center. Note that MARS introduces many partitions to capture the diagonal split and renders interpretation of partitions impossible.

Table 7 shows results for the optimization energy. The Morse-Smale regression performs similarly to the other two partition-based approaches and has some gains in performance with increasing sample size.

4.5 Topological Accuracy

The following experiments analyze and compare the topological accuracy, as described in section 3.5 of the Morse-Smale regression with MARS, boosted trees, SVM regression, and Loess. For the F-measure based accuracy comparisons the Morse-Smale complex is computed at 10% persistence

d	msII-svm	MARS	tree	Loess	SVM
2	0.003	0.021	0.037	0.001	0.01
4	0.003	0.043	0.051	0.003	0.01
6	0.005	0.041	0.046	-	0.01
8	0.004	0.040	0.039	-	0.01

Table 6: Mean squared error on diagonal function with Normal $\sigma = 0.15$ noise added on sample of size 2000 for increasing dimensions d.

n	msII-svm	MARS	tree	SVM
128	0.022	0.021	0.020	0.013
256	0.017	0.012	0.016	0.005
512	0.008	0.011	0.016	0.002
1024	0.004	0.011	0.015	0.002

Table 7: Mean squared error on optimization energy.

| | Fourpeaks | | | | | Fourpeaks 45° rotated | | | | |
n	msII-kd	MARS	tree	Loess	SVM	msII-kd	MARS	tree	Loess	SVM
128	0.38	0.71	0.05	0.51	0.25	0.39	0.35	0.05	0.51	0.27
256	0.67	0.84	0.09	0.61	0.36	0.64	0.41	0.06	0.59	0.32
512	0.81	0.84	0.11	0.72	0.37	0.81	0.36	0.05	0.71	0.33
1024	0.83	0.84	0.15	0.82	0.42	0.82	0.40	0.07	0.77	0.35

Table 8: F-measure for various sample sizes on four-peaks function with axis and non-axis aligned features and Normal noise $\sigma = 0.15$ added.

d	msII-svm	MARS	tree	Loess	SVM
2	0.88	0.63	0.05	0.57	0.99
4	0.85	0.57	0.53	0.18	0.97
6	0.91	0.52	0.43	-	0.94
8	0.91	0.43	0.31	-	0.91

Table 9: F-measure on diagonal function with different dimensionality on a sample of size 2000 with Normal noise $\sigma = 0.15$ added.

(i.e. persistence set to 10% of the function range) on a sample of 10000 points and compared to a ground truth assignment of the same sample.

Table 8 shows the topological accuracy of the different regression methods on the four-peaks function. MARS and regression trees model f well and are well-behaved under the influence of noise but do not produce a valid fit for the rotated function. The piece-wise constant fit of the boosted tree results in a low topological accuracy even under 10% persistence simplification. The SVM regression produces a stable fit under all conditions. However, all features show significantly lower persistence, which explains the low F-Measure under 10% persistence simplification. The Morse-Smale regression and Loess successfully reproduce the true topology of f. Regions of high curvature within a partition may not be captured by the Morse-Smale regression due to the employed piece-wise linear fit. However, the topology of the fit is not affected by this behavior.

Table 9 shows the results for the diagonal function with increasing dimension. For MARS, boosted trees and in particular Loess, the topology of the fit is unstable and does not capture the true topology of the function in any case. In the case of MARS, this behavior is caused by various false, artificially introduced critical points with $> 10\%$ relative persistence introduced by the regression. The Morse-Smale and the SVM regression accurately captures the underlying topology. The poor results of Loess are due to an overly smooth estimate which combines several partitions at the 10% persistence level.

Parameter	Description
cldfrc_rhminh	Minimum RH for high stable cloud formation
cldfrc_rhminl	Minimum RH for low stable cloud formation
cldopt_rliqice	Liquid drop size over sea ice
cldopt_rliqland	Liquid drop size over land
cldopt_rliqocean	Liquid drop size over ocean
cldsed_ice_stokes_fac	Ice Stokes factor scaling fall speed
cldwat_capnc	Cloud particle number density over cold land/ocean
cldwat_capnsi	Cloud particle number density over sea ice
cldwat_capnw	Cloud particle number density over warm land
cldwat_conke	Stratiform precipitation evaporation efficiency
cldwat_icritc	Threshold for cold ice autoconversion
cldwat_icritw	Threshold for warm ice autoconversion
cldwat_r3lcrit	Critical radius where liquid conversion begins
hbdiff_ricr	Critical Richardson number for boundary layer
hkconv_c0	Shallow convection precipitation efficiency parameter
hkconv_cmftau	Time scale for consumption rate of shallow CAPE
zmconv_alfa	Initial cloud downdraft mass flux
zmconv_c0	Deep convection precipitation efficiency parameter
zmconv_dmpdz	Parcel fractional mass entrainment rate
zmconv_ke	Environmental air entrainment rate
zmconv_tau	Time scale for consumption rate of deep CAPE

Table 10: Abbreviations and short description of the climate simulation parameters.

5 Case-Study: Climate Simulations

This case-study describes a Morse-Smale regression analysis of data from the Uncertainty Quantification Strategic Initiative of the Lawrence Livermore National Laboratory with application to climate science. This initiative currently collects data from climate simulations — using a recent version of the Community Atmosphere Model (CAM) (McCaa et al., 2004) — for sampling-based uncertainty quantification and sensitivity analysis. One goal of this effort is to gain an understanding of the behavior of the climate model with respect to the parameters. Alternatively, a faithful *black box* prediction model can be used as a surrogate to avoid costly simulations for new parameter settings. In this context, it is important that no artificial topological features are introduced that would lead to extrema not supported by the climate simulation.

To date, climate scientists have collected a sample of 1197 simulation runs. Within this sample, the scientist varied 21 input parameters (see Table 10) describing various aspects of the atmospheric physics for each simulation. The input parameters are normalized so that the input domain defines a 21-dimensional unit hyper-cube. For each simulation, 45 spatio-temporal outputs are recorded. The outputs are aggregated over space (averages over the entire globe or the northern/southern hemisphere) and time (annual or seasonal averages), resulting in a total of 945 different outputs.

The following analysis is based on the global annual long-wave flux ϕ, a measure of the total

amount of thermal radiation leaving the planet. Thus, the model to estimate is $\phi = f(x)$ with $\phi \in [219.2, 246.6]$ and $x \in [0, 1]^{21}$. There is no noise term since the simulations are deterministic. ANOVA and regression tables mentioned in the text without explicit references are in the supplemental material.

Summary of the Morse-Smale complex of ϕ. The Morse-Smale complex of ϕ estimated from the 1197 simulation runs with $k = 100$ nearest neighbors and no persistence simplification, consists of 7 extremal points (3 maxima $\phi = 246.6, 246.4, 240.9$ and 4 minima $\phi = 220.7, 219.7, 219.3, 219.2$) and 10 partitions. The persistences and associated number of partitions are $(2.58, 1$ partition$), (1.29, 2), (0.91, 4), (0.54, 6), (0.40, 10)$. The persistences are relatively small compared to the function range of 27.37. However, note that the saddle points are approximated by edges connecting two partitions (see section 2.2) which can over/underestimate the true saddle point. At the lowest persistence level, with 10 partitions, very small partitions including only 4 and 10 observations are introduced.

For discussion of regression models at different persistence levels let w_i denote the additive weighted linear Morse-Smale regression (approach II with kernel density estimation) at persistence level i with $i = 1$ corresponding to the Morse-Smale complex at the highest persistence level. Similarly, let l_i be the piece-wise linear model at the corresponding persistence level.

Predictive models with the additive weighted Morse-Smale regression. With a 10-fold cross-validation the additive weighted linear model achieves the best fit when all partitions are included (model w_6 with 10 partitions). If the Bayesian information criterion (BIC) is used for model selection, the Morse-Smale complex with 8 partition is best (model w_5). Using 900 observations to fit the model results in a mean test error on the remaining 297 observations over 10 runs of 0.81 and 1.24 for w_6 and w_5, respectively. A MARS model achieves a slightly better fit with a mean squared test error of 0.68. For comparison, a linear model has a mean squared test error of 2.61.

Interpretation with the piece-wise linear model. The differences and the significance of the introduced partitions are examined by an ANOVA of $l_1, \ldots, l_6$ (see Table 11). The ANOVA indicates that the segmentation at persistence level 5 is significant. This matches the choice of w_5 when using the BIC. A Normal Q-Q and residual plot of the l_5 model show no significant deviations and indicate that the parametric tests are feasible. The Morse-Smale complex at persistence level 5 has 4 minima and 2 maxima, with 8 partitions containing $303, 163, 91, 63, 282, 133, 76$, and 43

	Res.Df	RSS	Δ Df	Sum of Sq	F	Pr($>$F)
l_1	1175	3180.62				
l_2	1153	2983.57	22	197.05	3.93	0.0000
l_3	1109	2728.17	44	255.40	2.55	0.0000
l_4	1065	2535.15	44	193.02	1.92	0.0003
l_5	1021	2326.90	44	208.26	2.08	0.0001
l_6	1007	2292.55	14	34.34	1.08	0.3739

Table 11: ANOVA table for the piece-wise linear models $l_1, \ldots, l_6$ with $1, 2, 4, 6, 8$ and 10 partitions.

observations.

The final analysis examines the properties of the Morse-Smale complex partitioning at persistence level five. This analysis could be applied to any partition based regression, however, in classical tree-based regression the domain is split based on fit and does not necessarily represent a meaningful topological unit.

The first analysis examines the partitions in terms of their mean flux ϕ with a pairwise t-test with Bonferroni correction of the ϕ-values per partition. The test shows a significant (Bonferroni corrected $p = 0.016$) difference of 1.3 between partition 1 (303 observations $E[\phi] = 234.1$) and partition 5 (282 observations $E[\phi] = 232.8$). This indicates that the Morse-Smale complex captures two parameter regions in which the climate simulation leads to statistically significant average outcomes.

An ANOVA decomposition of the l_5 model shows that the general linear trend covers a large part of the variation. Thus, the climate simulation with respect to ϕ is governed to a large part by a linear relationship to the input parameters. However, the number of significant parameters for the general linear trend l_5 is drastically reduced from l_1. The most important parameter is cldwat_icritc and has a slight, but none the less statistically significant, variation in slope from partition to partition.

The reduction in variance introduced by the different partitions is small but significant. This, together with the reduction in main (without interactions) linear parameters suggest that some parameters only have an influence in certain regions of the parameter space. The parameters cldwat_capnc, cldwat_icritw, and zmconv_tau change sign depending on partition. This is interesting since this implies that these parameters can have opposite effects depending on interactions with other parameters or for high versus low flux scenarios. However, the change in sign is only significant (t-test $p < 0.01$) for cldwat_capnc.

It is not possible to extract exactly which parameters interact, however, the Morse-Smale complex provides an exploratory tool to guide further analysis. For example, based on Table 2 in the supplemental material and the differences in mean flux of partition 1 and 5, the parameter

zmconv_tau could be influential in higher flux scenarios only. Another venue for exploration suggested by this analysis is that zmconv_tau could have interactions with parameter zmconv_dmpdz and cldwat_capnc, which significantly differ between partition 1 and 5, that negate its effect in partition 5.

Acknowledgments

Supplemental Materials

Stability of The Morse-Smale Complex: Analysis and simulations on the stability of the Morse-Smale complex approximation. (pdf)

CLimate Simulation Tables: Tables of regression models on climate simulation data. (pdf)

R package *msr*: An implementation of the proposed regression in R. The package is also available online `http://cran.r-project.org/web/packages/msr/` and includes scripts to generate the data sets for the numerical experiments in Section 4. (tar.gz)

References

Alexander, W. P. and S. D. Grimshaw (1996). Treed regression. *Journal of Computational and Graphical Statistics 5*(2), 156–175.

Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984). *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks.

Bremer, P.-T., G. Weber, V. Pascucci, M. Day, and J. Bell (2010). Analyzing and tracking burning structures in lean premixed hydrogen flames. *IEEE Trans. on Vis. and Comp. Graphics 16*(2), 248–260.

Carr, H., J. Snoeyink, and U. Axen (2003). Computing contour trees in all dimensions. *Comput. Geom. Theory Appl. 24*(3), 75–94.

Carr, H., J. Snoeyink, and M. van de Panne (2004). Simplifying flexible isosurfaces using local geometric measures. In *IEEE Visualization '04*, pp. 497–504. IEEE Computer Society.

Chang, C.-C. and C.-J. Lin (2001). *LIBSVM: a library for support vector machines.*

Chaudhuri, P., M. ching Huang, W. yin Loh, and R. Yao (1994). Piecewise-polynomial regression trees. *Statistica Sinica 4*, 143–167.

Chazal, F., L. Guibas, S. Oudot, and P. Skraba (2009). Analysis of scalar fields over point cloud data. In *SODA '09: Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1021–1030.

Chipman, H. A., E. I. George, and R. E. McCulloch (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat. 4*(1), 266–298.

Dimitriadou, E., K. Hornik, F. Leisch, D. Meyer, and A. Weingessel (2010). *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien.* gR package version 1.5-24.

Edelsbrunner, H., J. Harer, V. Natarajan, and V. Pascucci (2003). Morse-Smale complexes for piecewise linear 3-manifolds. In *Proc. 19th Sympos. Comput. Geom.*, pp. 361–370.

Edelsbrunner, H., D. Morozov, and V. Pascucci (2006). Persistence-sensitive simplification of functions on 2-manifolds. In *Proc. ACM Symposium on Computational Geometry (SOCG)*, pp. 127–134.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist. 19*(1), 1–141. With discussion and a rejoinder by the author.

Gerber, S., P.-T. Bremer, V. Pascucci, and R. Whitaker (2010, Oct.). Visual exploration of high dimensional scalar functions. *IEEE Transactions on Visualization and Computer Graphics 16*(6), 1271–1280.

Gerber, S. and K. Potter (2011). Data analysis with the morse-smale complex: The msr package for r. In *Journal of Statistical Software*, pp. to appear.

Gerber, S., K. Potter, and O. Ruebel (2011). *msr: Morse-Smale approximation, regression and visualization.* R package version 0.1.

Gyulassy, A., M. Duchaineau, V. Natarajan, V. Pascucci, E.Bringa, A. Higginbotham, and B. Hamann (2007). Topologically clean distance fields. *IEEE TVCG 13*(6), 1432–1439.

Gyulassy, A., V. Natarajan, V. Pascucci, and B. Hamann (2007). Efficient computation of Morse-Smale complexes for three-dimensional scalar functions. *IEEE TVCG 13*(6), 1440–1447.

Hartley, R. and A. Zisserman (2003). *Multiple View Geometry* (Second ed.). Cambridge University Press.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference and prediction* (2 ed.). Springer.

Hothorn, T., P. Buehlmann, T. Kneib, M. Schmid, and B. Hofner (2010). *mboost: Model-Based Boosting.* gR package version 2.0-4.

Hothorn, T., K. Hornik, and A. Zeileis (2005). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics 2006*, 651–674.

Hsu, C.-W. and C.-J. Lin (2002). A comparison of methods for multiclass support vector machines. *IEEE Transaction on Neural Networks 13*(2), 415–425.

Laney, D., P.-T. Bremer, A. Mascarenhas, P. Miller, and V. Pascucci (2006). Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE TVCG 12*(5), 1052–1060.

Li, K.-C., H.-H. Lue, and C.-H. Chen (2000). Interactive tree-structured regression via principal hessian directions. *Journal of the American Statistical Association 95*.

Lindstrom, P. and M. Duchaineau (2009, Mar.). Factoring algebraic error for relative pose estimation. Technical Report LLNL-TR-411194, Lawrence Livermore National Laboratory.

McCaa, J. R., M. Rothstein, B. E. Eaton, J. M. Rosinski, E. Kluzek, and M. Vertenstein (2004). *User's Guide to the NCAR Community Atmosphere Model (CAM 3.0).* Climate And Global Dynamics Division National Center For Atmospheric Research Boulder, Colorado, USA.

Milborrow, S. (2009). *earth: Multivariate Adaptive Regression Spline Models.* gR package version 2.4-0.

Milnor, J. (1963). *Morse Theory.* New Jersey: Princeton University Press.

Morse, M. (1925, July). Relations between the critical points of a real function of n independent variables. *Transactions of the American Mathematical Society 27*, 345–396.

Platt, J. C. (1999). Probabilities for sv machines. *Advances in Large Margin Classifiers*, 61–74.

Silvermann, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapmann and Hall.

van Rijsbergen, C. (1979). *Information retrieval* (second ed.). Butterworths.

Vedaldi, A. and S. Soatto (2008). Quick shift and kernel methods for mode seeking. In *Proc. European Conf. on Computer Vision*, pp. 705–718.