When to Segment: Topology-Aware Energy-Based Diagnostics for Structural Model Uncertainty

Xana Verte
Axiolyze Lab

January 14, 2026

Abstract

In many applied modeling settings, performance degradation is addressed by rebuilding or retuning predictive models in response to detected drift. While effective in some cases, such interventions may obscure the underlying cause of instability, particularly when degradation arises from latent structural heterogeneity rather than gradual distributional change. In this work, we propose a topology-aware, energy-based diagnostic framework for identifying when segmentation of a predictive model is warranted.

Throughout the paper, we use the term "topology-aware" in a loose but standard sense, referring to methods that explicitly respect neighborhood structure encoded by a graph, rather than to formal tools from topological data analysis such as persistent homology.

Our approach treats segmentation not as a primary modeling objective, but as a diagnostic hypothesis about structural uncertainty in the relationship between covariates and outcomes. By combining local geometric structure in feature space with discrepancies in target behavior, we construct an energy functional whose minimization reveals stable zones of structural conflict. The resulting segmentation highlights regions where a global model fails systematically, while avoiding overreaction to noise.

The framework admits both frequentist and Bayesian interpretations: as a regularized risk minimization problem with complexity penalties, or as an implicit evidence-based assessment of competing structural hypotheses. Through controlled synthetic experiments, we illustrate how segmentation may emerge, stabilize, or dissolve as data are progressively corrupted by scale changes and noise, reflecting conditions commonly encountered in real-world actuarial and industrial applications.

1 Introduction

Predictive models deployed in production environments are routinely monitored for performance degradation. Common first-line diagnostics include marginal drift detection, univariate or low-dimensional distributional comparisons, and stability checks on summary statistics. These tools are essential for quality control, yet they operate primarily in simple projections of the data and offer limited insight into higher-order structural changes.

When such diagnostics indicate deterioration, the prevailing response is often to rebuild or retune the model. In practice, this can lead to frequent and costly model updates, sometimes performed even when performance later recovers without intervention. This pattern raises a fundamental question: *when does observed instability reflect transient noise, and when does it signal a persistent structural mismatch that justifies segmentation or redesign of the model?*

This work is motivated by applied scenarios in which model degradation was not caused by gradual drift along individual covariates, but by the emergence of regions in feature space where the relationship between predictors and target variables became inconsistent with a global modeling assumption. In such cases, neither univariate diagnostics nor routine hyperparameter adjustments provided reliable guidance. Segmentation of the model was a plausible organizational and engineering response, yet lacked principled diagnostic support.

We argue that segmentation should be treated as a *second-line diagnostic decision*, rather than a default corrective action. Specifically, segmentation is warranted only when there is evidence of stable structural heterogeneity that cannot be explained by noise or by smooth variation captured within a single model.

To this end, we propose a topology-aware energy-based diagnostic framework that integrates geometric information from the feature space with discrepancies observed in the target behavior. The goal is not to optimize predictive accuracy per se, but to assess whether the data support the hypothesis that a single global model is structurally inadequate. From a modeling perspective, this diagnostic question can be interpreted either in frequentist terms, as a regularized assessment of structural adequacy, or in Bayesian terms, as an implicit comparison of competing structural hypotheses under a complexity-penalizing prior.

Although the framework produces a partition of the dataset into zones, segmentation is not treated as an end in itself. Instead, zone formation is used as a diagnostic hypothesis: the emergence of stable, target-inconsistent regions indicates obstacles to a single global predictive model, while the absence of such regions is itself a meaningful outcome. In this sense, the proposed procedure should be viewed as a diagnostic tool rather than a clustering method aimed at uncovering latent groups.

2 Related Work

The proposed framework intersects several established lines of research, including drift detection and model monitoring, manifold and topology-aware learning, graph-based segmentation, and structural diagnostics for heterogeneous data. This section reviews the most relevant directions and clarifies the positioning of the present work among them.

2.1 Drift Detection and Model Monitoring

A large body of literature addresses the problem of detecting changes in data distributions over time, commonly referred to as concept drift or dataset shift [7, 4]. Classical approaches focus on marginal feature distributions, target distributions, or model residuals, using statistical tests, divergence measures, or summary statistics.

While effective as first-line monitoring tools, such methods typically operate on low-dimensional projections of the data and do not directly address structural heterogeneity in the conditional relationship between covariates and targets. In particular, they provide limited guidance on whether observed degradation reflects transient noise, gradual drift, or persistent structural regimes that would justify changes in model structure.

The present work is complementary to drift detection. Rather than replacing standard monitoring tools, it addresses a second-order diagnostic question: whether the assumption of a single global predictive model remains structurally valid given observed local inconsistencies in the conditional distribution $F \mid X$.

2.2 Manifold Learning and Topology-Aware Methods

Methods from manifold learning and nonlinear dimensionality reduction, such as diffusion maps and related spectral techniques, aim to uncover low-dimensional geometric structure in high-dimensional data [2, 10]. These approaches have been extended to clustering and segmentation tasks by exploiting spectral properties of neighborhood graphs.

Related work in topological data analysis (TDA) studies global and local topological features of data using tools such as persistent homology. These methods provide powerful invariants for detecting holes, intersections, and other nontrivial geometric structures.

However, most topology-aware methods focus primarily on the geometry of the feature space and do not explicitly incorporate discrepancies in target behavior. In contrast, the proposed framework integrates geometric locality with outcome-driven disagreement. Topological concepts are employed implicitly through graph locality and connectivity, without relying on explicit TDA constructions. Geometry serves as a scaffolding for diagnostics rather than as an objective in itself. Neighborhood graphs used in such methods often rely on generic similarity measures, including mixed-type distances for heterogeneous data [5].

2.3 Graph-Based Segmentation and Energy Minimization

Graph-based segmentation using energy functionals has a long history in computer vision and clustering, including Potts models, Mumford–Shah-type formulations [8], and related variational approaches such as active contours [1].

These approaches typically balance data fidelity, smoothness across edges, and complexity penalties to obtain coherent partitions.

The present work adopts an energy-based perspective but differs in both intent and construction. The energy functional is not optimized to produce a final predictive partition, but to diagnose whether stable structural boundaries exist at all. Edge weights are derived from discrepancies in target behavior rather than purely geometric similarity.

The contribution does not lie in introducing a new class of energy functions or optimization techniques, but in the construction of the data-consistency term under low signal-to-noise conditions and in the dual-frame perspective that explicitly separates geometric locality in X from inconsistency in $F \mid X$.

2.4 Structural Heterogeneity and Regime Detection

Several specialized algorithms aim to detect intersecting manifolds or regime boundaries by analyzing local tangent spaces, curvature, or density variations, including early work on manifold intersection detection [3]. Early work by Deutsch and Medioni, among others, explicitly targets the detection of manifold intersections through geometric cues.

Related ideas also appear in regime detection and mixture modeling, particularly in time series analysis, where data are assumed to arise from multiple latent regimes with distinct statistical properties [6, 9]. Such methods often rely on explicit parametric assumptions, strong signal-to-noise ratios, or well-separated regimes.

While conceptually related, these approaches are frequently sensitive to noise and assume relatively clean structure. The present framework is designed for low signal-to-noise, heterogeneous tabular data, where geometric signals alone are insufficient and must be combined with outcome-based diagnostics to assess structural validity.

2.5 Positioning of the Present Work

In summary, this work does not propose a new clustering or dimensionality reduction algorithm. Instead, it introduces a diagnostic framework that synthesizes ideas from topology-aware learning, graph-based segmentation, and statistical model validation to address a practical modeling question: when does segmentation represent a justified structural response, rather than an overreaction to noise?

By framing segmentation as a diagnostic hypothesis and providing both frequentist and Bayesian interpretations of the resulting energy landscape, the approach bridges methodological traditions while remaining grounded in applied modeling constraints. Segmentation is treated as evidence of structural incompatibility with a global model, not as a modeling objective in itself.

To further clarify this positioning, Table 1 provides a high-level conceptual comparison with related methodological directions.

Direction	Primary goal	Typical signal	Output	Role of segmentation
Drift detection	Detect distributional change	Marginals, residuals	Alarm / score	Not explicit
Manifold learning / TDA	Recover geometric structure	Geometry of X	Embedding / topology	Primary objective
Graph-based segmentation	Partition data	Similarity, smoothness	Final segmentation	Optimization goal
Regime-switching / mixtures	Model heterogeneity	Parametric $F \mid X$	Fitted regimes	Assumed structure
This work	Diagnose structural validity	Local inconsistency in $F \mid X$	Diagnostic segmentation	Hypothesis to be validated

Table 1: Conceptual comparison of the proposed framework with related methodological approaches. The proposed method treats segmentation as diagnostic evidence rather than as a modeling objective.

Accordingly, the proposed framework is best viewed as a second-line structural diagnostic tool for low signal-to-noise tabular data, positioned between first-line drift monitoring and full model restructuring.

3 Problem Formulation

3.1 Segmentation as a Diagnostic Hypothesis

We consider a dataset consisting of covariates $X \in \mathbb{R}^d$ and target variables F, observed over discrete time periods. A global predictive model implicitly assumes that the conditional relationship between X and F is structurally homogeneous across the domain of interest.

Segmentation introduces an alternative hypothesis: that the data are better described by multiple regimes or zones, each governed by a locally coherent relationship between predictors and outcomes. Importantly, segmentation increases model complexity and operational burden, and therefore should be justified by sufficiently strong evidence.

In this work, segmentation is not treated as an optimization goal, but as a *diagnostic outcome*: an indication that structural uncertainty in the model cannot be ignored.

From a Bayesian perspective, this diagnostic view can be interpreted as a comparison between competing structural hypotheses with different effective dimensionalities, where the absence of segmentation corresponds to a parsimonious prior preference for a single global model.

3.2 Problem Setting and Inputs

We consider observations (X_i, F_i), where X_i denotes a (possibly mixed-type) covariate vector and F_i is a target variable of interest. The goal is to detect regions of the covariate space where a single global predictive relationship between X and F fails.

The output of the diagnostic procedure is a partition of selected landmark observations into a small number of zones, together with diagnostic quantities that explain and support the resulting segmentation. Importantly, the procedure may also return the trivial single-zone solution, indicating insufficient evidence to justify structural separation.

4 Dual-frame Graph-based Framework

The proposed framework decomposes the diagnostic task into two interacting but conceptually distinct components: a geometric frame, which encodes locality and comparability in the covariate space, and a target frame, which captures local inconsistencies in the conditional behavior of the target given the covariates. Together, these frames provide a structured basis for assessing whether segmentation is warranted.

4.1 Geometry Frame: Locality in X

To assess structural heterogeneity, we construct a graph-based representation of the feature space that captures local geometry without assuming a global manifold structure. Nodes correspond to representative points (e.g., landmarks), and edges encode neighborhood relationships based on a suitable distance metric.

This representation serves two purposes:

- It localizes comparisons of target behavior to nearby regions in feature space.
- It enables the identification of boundaries where local relationships change abruptly.

The framework is deliberately agnostic to the specific choice of distance or neighborhood construction, provided it yields a stable notion of local proximity.

A sparse graph $G = (V, E)$ is constructed over landmark points using a local distance in X-space (e.g. Gower distance with k-nearest neighbors). The graph encodes local comparability but is not used directly for clustering. Its role is to provide a topological scaffold on which diagnostic comparisons of target behavior are performed.

4.2 Target Frame: Local Inconsistency in $F \mid X$

While the geometry frame defines *where* comparisons are made, the target frame defines *what* is compared. Specifically, it captures discrepancies in the conditional behavior of the target variable F among neighboring observations in the covariate space.

We adopt an energy-based diagnostic perspective, in which segmentation is driven by the accumulation of local inconsistencies in $F \mid X$, rather than by geometric separation alone.

Conceptually, the diagnostic energy consists of three components:

- A **data term**, reflecting within-zone consistency of target behavior;
- A **boundary term**, penalizing disagreements between neighboring nodes assigned to different zones, weighted by the severity of target discrepancies;

- A **complexity penalty**, discouraging unnecessary proliferation of zones.

Minimizing this energy yields a segmentation that balances explanatory gain against structural complexity. Importantly, the shape of the energy landscape itself provides diagnostic information: flat regions indicate ambiguity or noise-dominated settings, while sharp changes or elbows suggest meaningful structural transitions.

4.3 Dual Interpretation

The proposed energy-based formulation admits two complementary interpretations.

From a **frequentist perspective**, the energy corresponds to a regularized objective function. Introducing segmentation reduces residual structural inconsistency in $F \mid X$, but incurs a penalty proportional to the number of zones. The balance between these terms reflects a trade-off between explanatory adequacy and model complexity.

From a **Bayesian perspective**, the same energy can be interpreted as a negative log-posterior over structural hypotheses. The data and boundary terms correspond to a likelihood capturing local disagreement in $F \mid X$, while the complexity penalty acts as an implicit prior favoring parsimonious structural explanations.

These interpretations are mathematically compatible and lead to identical computational procedures. The dual framing allows the method to be situated naturally within both statistical and probabilistic modeling traditions, without altering its algorithmic core.

4.4 Energy-based Segmentation Algorithm

This section describes the concrete algorithmic realization of the proposed diagnostic framework. The goal is not to optimize predictive performance directly, but to explore the space of possible segmentations and assess whether any of them provide stable and interpretable reductions in structural inconsistency between covariates and targets.

The algorithm operates on a graph-based representation of the data and produces a sequence of candidate segmentations with increasing structural complexity. Rather than selecting a single partition a priori, it exposes the full energy–complexity trade-off for diagnostic inspection.

4.5 Notation and energy definition

Let $G = (V, E)$ denote a locality graph constructed over a set of n landmark observations. Each node $i \in V$ corresponds to a landmark point (X_i, F_i). A segmentation is represented by zone labels

$$z_i \in \{1, \ldots, K\}, \quad i \in V,$$

where K denotes the number of zones.

Each zone k is associated with parameters ϕ_k summarizing the local behavior of the target variable $F \mid X$ within that zone. The specific form of ϕ_k depends on the application and may range from simple moment-based summaries to fitted local predictive or distributional models.

We define an energy functional of the form

$$E(z, \phi) \;=\; \sum_{i=1}^{n} \ell_i\big(\phi_{z_i}\big) \;+\; \lambda \sum_{(i,j)\in E} w_{ij}\, \mathbf{1}[z_i \neq z_j] \;+\; \alpha\, K.$$

The first term is a *data-consistency term* measuring how well observation i agrees with the target behavior of its assigned zone. The second term is a *boundary penalty* that discourages separating neighboring observations with similar target behavior; edge weights w_{ij} encode local disagreement in $F \mid X$. The final term penalizes the number of zones and controls structural complexity.

From a frequentist perspective, this energy corresponds to a regularized objective balancing explanatory adequacy against model complexity. From a Bayesian perspective, the same functional can be interpreted as a negative log-posterior, with the complexity term acting as an implicit prior favoring parsimonious structural explanations. Importantly, both views lead to the same computational procedure.

4.6 Edge weights and disagreement metrics

The quality of the diagnostic segmentation depends critically on how local disagreement in target behavior is quantified. Direct computation of rich distributional discrepancies for all candidate segmentations is often computationally infeasible.

To address this, we adopt a two-phase strategy.

In *Phase A*, surrogate disagreement measures $\tilde{w}_{ij}$ are computed for each edge $(i, j) \in E$ using low-complexity summaries of local target behavior, such as moment differences, residual contrasts, or simplified predictive discrepancies. These surrogates are designed to be inexpensive and robust, and are used solely to guide the exploration of candidate partitions.

In *Phase B*, for a restricted set of candidate segmentations, edge weights w_{ij} are recomputed using richer discrepancy measures, such as energy distance, Hellinger distance between fitted mixtures, or posterior predictive divergence in Bayesian models. This refinement step is used for validation and interpretation rather than for large-scale search.

4.7 Two-phase optimization procedure

The algorithm follows an agglomerative strategy operating on the locality graph. Starting from singleton zones, zones are iteratively merged to minimize a surrogate energy

$$\tilde{E}(z) = \sum_{(i,j)\in E} \tilde{w}_{ij}\, \mathbf{1}[z_i \neq z_j] + \alpha\, K.$$

This procedure produces an ordered sequence of candidate partitions, along with the full energy path as a function of K. The role of surrogate metrics at this stage is intentionally limited: false positives are acceptable and filtered out during refinement, while false negatives are mitigated by conservative thresholding and neighborhood aggregation.

Rather than selecting a single optimum automatically, the algorithm retains multiple candidate segmentations corresponding to stable regions or elbows in the energy trajectory. For these candidates, the refined energy $E(z)$ is evaluated using richer disagreement measures.

If no candidate segmentation yields a meaningful or stable improvement over the global model ($K = 1$), the algorithm explicitly returns the unsegmented solution, treating abstention from segmentation as a valid diagnostic outcome.

4.8 Temporal consistency

When data are available across multiple time slices, the diagnostic procedure can be applied independently to each slice. Resulting segmentations are compared using overlap measures such as the adjusted Rand index.

Zones that persist across time are interpreted as stable structural obstacles to a single global model. In contrast, zones that appear only sporadically under increased noise or scale changes are treated as artifacts rather than actionable signals. Temporal consistency thus provides an additional layer of validation for segmentation decisions.

4.9 Algorithmic summary

Algorithm 1 summarizes the full topology-aware energy-based segmentation procedure, including landmark selection, graph construction, surrogate screening, refinement, and diagnostic output.

Algorithm 1 Topology-Aware Energy-Based Segmentation

```
1: Input: Covariates X = {X_i}_{i=1}^n, targets F = {F_i}_{i=1}^n, number of
   landmarks m, neighborhood size k, complexity penalty α, surrogate
   disagreement w̃, refined disagreement w (optional)
2: Output: Zone labels z_i for landmarks, diagnostic metadata
3: Select landmarks L ⊂ {1, ..., n}, |L| = m
4: Construct kNN graph G = (V, E) over {X_i}_{i∈L}
5: Compute surrogate edge weights w̃_ij
6: Initialize each landmark as a singleton zone
7: Iteratively merge zones to minimize surrogate energy
8: Select candidate partitions along energy path
9: if refinement enabled then
10:    Recompute w_ij and evaluate refined energy
11: end if
12: Select final partition or return global model
```

5 Diagnostics and Validation Principles

The proposed method is not intended to produce a segmentation by default. Instead, segmentation is treated as a diagnostic outcome that must be justified by multiple, mutually consistent signals.

This section describes diagnostic principles used to assess whether a detected partition corresponds to a meaningful structural obstacle to a global predictive model, rather than a noise-driven artifact.

From a Bayesian perspective, these diagnostics can also be interpreted as posterior checks on the plausibility of competing structural hypotheses, providing qualitative evidence in favor of or against increased structural complexity.

5.1 Energy path analysis

During agglomerative minimization, the full energy path

$$E(K), \quad K = m, m-1, \ldots, 1$$

is recorded.

Diagnostics include:

- identification of energy elbows or plateaus;
- analysis of marginal energy reduction ΔE per merge;
- detection of regimes where further fragmentation yields diminishing returns.

A selected value of K is interpretable only if the energy decrease up to K is substantial relative to the complexity penalty.

In Bayesian terms, pronounced elbows in the energy trajectory correspond to regions where posterior support shifts from favoring a single global structure to favoring multiple structural explanations. Flat energy paths indicate weak evidence and correspondingly diffuse posterior mass over segmentations.

5.2 Spatial localization of cut edges

Edges with high disagreement weights that connect different zones (*cut edges*) are examined in the covariate space X.

A meaningful segmentation is characterized by:

- concentration of cut edges in localized regions of X;
- alignment of cut locations with geometrically ambiguous or overlapping areas.

Diffuse or uniformly distributed cut edges indicate over-segmentation or noise-driven artifacts.

From a probabilistic viewpoint, localized cut edges suggest coherent regions where competing structural explanations concentrate, while diffuse boundaries reflect broadly distributed uncertainty rather than genuine structural breaks.

5.3 Boundary load per zone

For each zone, a boundary load is computed as the total weight of cut edges incident to the zone, optionally normalized by zone size.

Zones with substantial boundary load are interpreted as necessary to accommodate strong local disagreements. Zones with near-zero boundary load are candidates for merging.

In Bayesian terms, zones with high boundary load correspond to regions with strong posterior tension between local and global structural explanations, while low boundary load indicates weak evidence for maintaining separation.

5.4 Between- and within-zone variability of the target

To verify that zones differ in a meaningful way with respect to the target variable F, between-zone and within-zone variability measures are computed.

Depending on the target type, this may involve:

- variance decomposition;
- deviance or likelihood-based comparisons;
- nonparametric tests such as PERMANOVA.

A valid segmentation should increase between-zone variability while reducing within-zone heterogeneity.

From a Bayesian perspective, this corresponds to posterior predictive checks: well-supported segmentations induce clearly distinguishable predictive distributions across zones, while unsupported ones fail to separate posterior predictive behavior.

5.5 Improvement of local predictive models

Global predictive models are compared against collections of zone-specific local models.

Diagnostics focus on:

- improvement in predictive scores (e.g. deviance, log-likelihood);
- robustness of improvements across folds or time slices;
- absence of systematic overfitting in small zones.

If local models do not outperform the global model in a stable manner, the segmentation is rejected.

In Bayesian terms, this comparison can be interpreted as an informal model comparison: segmentation is supported only if local models achieve consistent posterior predictive gains relative to the global model.

5.6 Comparison with geometry-only clustering

Results are compared with geometry-based clustering methods, such as spectral clustering applied to the locality graph.

Interpretation guidelines include:

- geometry $= 1$, energy > 1: conflict arises from $F \mid X$, not geometry;

- geometry > 1, energy $\approx$ geometry: structural geometric separation dominates;
- geometry > 1, energy $= 1$: geometric separation not supported by the target.

This comparison clarifies whether segmentation is driven by predictive disagreement or purely geometric effects.

5.7 Temporal stability

When data are available across time slices, segmentations are compared using overlap measures such as the adjusted Rand index.

Persistent zones across time are interpreted as structural obstacles, while transient zones appearing only under increased noise or scale changes are treated as artifacts.

From a Bayesian perspective, temporal persistence can be viewed as repeated posterior support for the same structural hypothesis across comparable datasets, while instability reflects sensitivity to noise rather than genuine structure.

5.8 Validation with hidden variables (synthetic data)

In synthetic experiments, recovered zones are compared against latent variables and regime indicators.

Validation focuses on:

- alignment with true regimes;
- diversity of zone assignments in overlapping regions;
- failure modes under increasing noise and scaling.

These experiments are used to probe the limits of detectability rather than to optimize performance. In Bayesian terms, they illustrate how posterior support for segmentation degrades as signal-to-noise conditions worsen.

6 Controlled Diagnostic Protocol

This section describes a controlled diagnostic protocol designed to examine the qualitative behavior and internal consistency of the proposed framework under progressive degradation of structural signals. The goal is not empirical validation or performance benchmarking, but to clarify how diagnostic signals emerge, stabilize, or dissolve as the underlying structure becomes weaker or increasingly obscured by noise.

Real-world datasets rarely provide ground truth regarding structural heterogeneity. Observed changes in predictive performance may arise from transient noise, scale effects, or genuine structural mismatch, and these factors are often entangled. To disentangle them conceptually, we consider synthetic constructions in which specific sources of degradation can be introduced in a controlled and interpretable manner.

From a Bayesian perspective, the protocol can be viewed as probing how posterior support shifts between competing structural hypotheses — a single global model versus segmented alternatives — as signal-to-noise conditions deteriorate. Persistent support for the global hypothesis under increasing variability is treated as a meaningful diagnostic outcome, rather than a failure of the method.

6.1 Motivation

The protocol is motivated by practical diagnostic questions rather than by model optimization. In applied settings, practitioners often face ambiguity as to whether observed instability reflects noise amplification or the emergence of structurally distinct regimes. Standard first-line diagnostics are typically insensitive to such distinctions.

The controlled constructions considered here are intended to clarify when the proposed diagnostics indicate meaningful structural heterogeneity, when they remain inconclusive, and when they explicitly favor retaining a global modeling assumption.

6.2 Protocol: Time $\rightarrow$ Scale $\times$ Noise

Each dataset is organized into a sequence of discrete time slices. At the initial time point, the data exhibit clear structural properties with minimal noise and well-separated local behaviors. Subsequent slices apply controlled transformations designed to progressively degrade structural identifiability.

Specifically, we consider two classes of transformations:

- multiplicative scaling of the target variable, intended to simulate changes in exposure, volume, or magnitude without altering conditional structure;

- additive noise applied to covariates and/or targets, reducing the effective signal-to-noise ratio and blurring local distinctions.

These transformations are applied incrementally across time slices, producing a trajectory from clearly structured regimes toward ambiguous or noise-dominated settings. This design mirrors practical scenarios in which data quality or comparability degrades gradually rather than abruptly.

6.3 Expected Diagnostic Signals

Rather than reporting performance metrics, the protocol examines qualitative and structural diagnostic signals produced by the framework as conditions change. The following aspects are monitored:

- the stability or instability of the inferred number of zones across successive slices;
- the localization of high-energy boundaries in feature space and their persistence under increasing noise;
- relative changes in the adequacy of local predictive summaries compared to a global baseline;
- consistency of inferred segmentations across adjacent time slices.

Importantly, the absence of segmentation, or the collapse of previously detected zones under increased noise, is treated as a valid and informative diagnostic outcome. Such behavior indicates insufficient evidence to justify increased structural complexity.

6.4 Interpretation

From a frequentist perspective, the protocol illustrates how regularized structural diagnostics respond to decreasing signal strength and increasing variability. Flat or weakly informative energy landscapes correspond to noise-dominated regimes, while sharp transitions or stable elbows indicate structural signals that persist despite degradation.

From a Bayesian perspective, these observations correspond to posterior mass remaining concentrated on simpler structural hypotheses unless the data provide consistent and localized evidence to support segmentation. In this sense, the protocol highlights not only when segmentation is suggested, but also when it is explicitly discouraged.

The purpose of this controlled diagnostic examination is not empirical validation, but to clarify how the proposed framework behaves under interpretable and progressively degraded conditions. No claims are made regarding optimality, predictive superiority, or comparative performance relative to alternative methods. The constructions considered here serve solely as sanity checks for the internal logic of the framework and as guidance for its practical diagnostic interpretation.

7 Discussion and Limitations

The proposed framework is designed as a conservative diagnostic tool rather than as a segmentation method by default. Structural segmentation is treated as a hypothesis that must be supported by stable and interpretable diagnostic signals, rather than as an outcome to be optimized or enforced.

A key feature of the diagnostics is their ability to explicitly abstain from segmentation. When disagreement signals are weak, unstable, or inconsistent across validation criteria, the framework favors simpler structural explanations and returns the global model. This behavior is intentional and reflects uncertainty in the data rather than a failure of the method.

As structural signals degrade, for example due to increasing noise or reduced local sample sizes, diagnostic criteria become less decisive and energy differences between candidate segmentations flatten. In such regimes, the complexity penalty dominates, suppressing spurious segmentation and reinforcing parsimony. From a Bayesian perspective, this corresponds to posterior support remaining concentrated on simpler structural hypotheses as the likelihood becomes less informative.

Importantly, the framework is diagnostically insensitive to uniform scale transformations of the target variable. Changes in magnitude alone do not induce segmentation unless they are associated with genuine inconsistencies in conditional target behavior. This property reflects the diagnostic focus on structural conflict in $F \mid X$, rather than on absolute error levels.

Compared to geometry-driven clustering approaches, the proposed diagnostics are orthogonal in spirit. Segmentation is triggered by disagreements in predictive or distributional behavior rather than by separation in feature space alone. As a result, the framework can detect structurally meaningful conflicts even in settings where geometric separation is weak or absent.

Temporal persistence, when comparable data slices are available, can be used as an auxiliary interpretational signal. Segmentations that are stable across time are more plausibly associated with persistent structural obstacles, while transient or unstable zones are treated with caution. Temporal consistency is therefore interpreted as supportive evidence rather than as a primary decision criterion.

7.1 Role of the complexity penalty

The complexity penalty parameter α plays a role analogous to model selection penalties in classical statistical learning. Rather than admitting a universally optimal choice, α is intended to be used diagnostically. Smaller values allow the framework to surface candidate structural inconsistencies, while larger values enforce conservative behavior and favor global explanations.

In practice, sensitivity analysis over a narrow range of α values, combined with inspection of energy–complexity trade-offs and stability considerations, provides a more informative assessment than selecting a single fixed value.

7.2 Computational considerations

The framework involves several computational trade-offs. In particular, refined disagreement measures based on rich distributional comparisons can be computationally expensive. For this reason, they are applied only to a restricted set of candidate segmentations identified during the surrogate screening phase.

The use of landmarks and sparse locality graphs further reflects a deliberate design choice to balance diagnostic sensitivity with computational feasibility. The framework is intended primarily for offline structural assessment rather than for real-time deployment.

7.3 Limitations and future directions

The proposed diagnostics require sufficient data within candidate zones to support reliable local comparisons. Rare regimes with very small sample sizes may remain undetectable. The method also depends on the construction of a meaningful locality graph; poor choices of distance metrics or neighborhood sizes can degrade diagnostic quality.

Several extensions are conceptually straightforward but are not explored in this work. These include richer Bayesian formulations with explicit priors over segmentations, bootstrap-based stability assessments, and more expressive topological diagnostics. We leave systematic empirical evaluation of these directions for future work.

8 Conclusion

We have introduced a topology-aware, energy-based diagnostic framework for assessing structural uncertainty in predictive models. The central idea of the framework is to treat segmentation not as a modeling objective, but as a diagnostic hypothesis about the validity of a single global predictive relationship.

By combining local geometric structure in the feature space with discrepancies in target behavior, the proposed diagnostics identify regions where global modeling assumptions systematically fail. At the same time, the framework is deliberately conservative: in the absence of stable structural signals, it favors simpler explanations and explicitly returns the global model.

The energy-based formulation admits both frequentist and Bayesian interpretations, providing a unified view of structural adequacy that is compatible with established modeling traditions. Importantly, these interpretations do not alter the computational procedure, but offer complementary perspectives on the same diagnostic evidence.

Through controlled synthetic experiments, we illustrated how segmentation emerges, stabilizes, or dissolves as data are progressively corrupted by scale changes and noise. These experiments reflect conditions commonly encountered in applied actuarial and industrial settings, where low signal-to-noise ratios complicate structural assessment.

Overall, the proposed framework is intended as a second-line diagnostic tool, positioned between first-line drift monitoring and full model restructuring. By formalizing when segmentation is warranted—and when it is not—the approach supports more deliberate and interpretable responses to model instability in complex, heterogeneous data.

References

[1] Tony F. Chan and Luminita A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 2001.

[2] Ronald R. Coifman and Stephane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 2006.

[3] Steven Deutsch and Gérard Medioni. Intersecting manifolds: Detection, segmentation, and labeling. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[4] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 2014.

[5] John C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 1971.

[6] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 1989.

[7] Jie Lu, Anjin Liu, Feng Dong, Fang Gu, João Gama, and Guangquan Zhang. Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[8] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 1989.

[9] Anne Samé, Faicel Chamroukhi, Gérard Govaert, and Patrick Aknin. Model-based clustering and segmentation of time series with changes in regime. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[10] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 2007.