Mandatory Critical Points of 2D Uncertain Scalar Fields

David Günther[1], Joseph Salmon[1], and Julien Tierny[2]

[1]Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, Paris, France, [2]CNRS LTCI, Télécom ParisTech, Paris, France
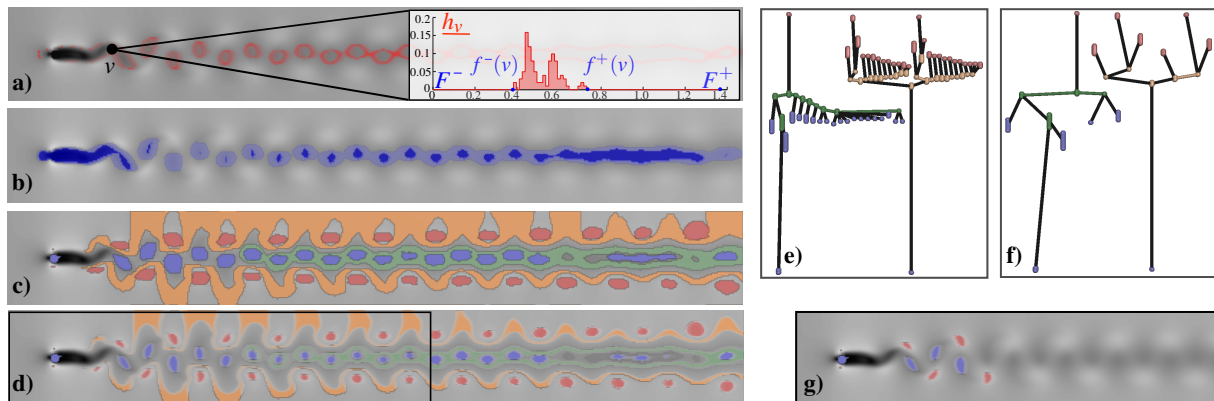
Figure 1: *Mandatory critical points of the velocity magnitude of the* uncertain *Kármán vortex street. a) Each vertex v is assigned with a histogram h_v estimating its probability density function. The shades of red show the point-wise probability for the isovalue 0.85. b) The support of h_v is visualized by the lower (light blue) and upper (dark blue) bound fields. c) depicts the mandatory critical points (blue: minimum, green: join saddle, yellow: split saddle, red: maximum), d) illustrates the spatial uncertainty within the components. e) shows the mandatory join/split tree, and f) and g) the simplified visualization.*

Abstract

This paper introduces a novel, non-local characterization of critical points and their global relation in 2D uncertain scalar fields. The characterization is based on the analysis of the support of the probability density functions (PDF) of the input data. Given two scalar fields representing reliable estimations of the bounds of this support, our strategy identifies mandatory critical points*: spatial regions and function ranges where critical points have to occur in any realization of the input. The algorithm provides a global pairing scheme for mandatory critical points which is used to construct* mandatory join *and* split trees*. These trees enable a visual exploration of the common topological structure of all possible realizations of the uncertain data. To allow multi-scale visualization, we introduce a simplification scheme for mandatory critical point pairs revealing the most dominant features. Our technique is purely combinatorial and handles parametric distribution models and ensemble data. It does not depend on any computational parameter and does not suffer from numerical inaccuracy or global inconsistency. The algorithm exploits ideas of the established join/split tree computation. It is therefore simple to implement, and its complexity is output-sensitive. We illustrate, evaluate, and verify our method on synthetic and real-world data.*

Categories and Subject Descriptors (according to ACM CCS):   F.2.2 [Analysis Of Algorithms And Problem Complexity]: Nonnumerical Algorithms and Problems—Computations on discrete structures

1. Introduction

In industry and scientific research, measurements and numerical simulations are necessary to understand underlying technical, physical or chemical processes. Results are of-ten scalar fields, which could represent: the temperature in a closed system, the geometric height of a surface, or the velocity of a fluid. The data acquisition could suffer from inaccuracies and noise might be introduced. Simulations, on

the other hand, suffer from a limited floating-point precision and quantification artifacts. Moreover, input parameters may not be known with certainty yielding parameter studies. In summary, uncertainty [GUM08] is introduced into the data. Each vertex of the numerical domain is no longer assigned to a single value but a probability density function (PDF) describing the scalar value of this point.

When facing an uncertain scalar field, the user is left with the difficult task of understanding the data set as a whole. Individual realizations (e.g., mean) or observations (ensemble data) – which can be numerous – are not necessarily representative of the global process. The understanding of the data in its entirety requires to identify the common denominator of all realizations or observations: features, such as critical points, which are shared by all of them.

In this paper, we introduce a novel, non-local, and combinatorial definition of *mandatory critical points* (Sec. 3) which is independent of the underlying distribution model. The mandatory critical points are globally consistent (Sec. 4) and occur in any realization. Therefore, they can be interpreted as the *common topological denominator* of the realizations of the uncertain data. Due to the uncertainty, a mandatory critical point is represented by a *critical component* (Fig. 1c) and a *critical interval*. It is guaranteed that for any realization at least one critical point of a given type is present in the prescribed region and takes a critical value in the prescribed interval. The global nature of our combinatorial algorithm allows to derive the notion of *mandatory join* and *split* trees (Sec. 4). These trees describe the expected structure of the mandatory critical points. A *progressive simplification* (Sec. 4) of mandatory critical points is introduced to help the user differentiate small- and large-scale structures (Fig. 1d,g). Our algorithm does not depend on any computational parameter and its time complexity is output-sensitive, which makes it well suited for large ensemble data. Experiments on real-life data-sets (Sec. 5) demonstrate the ability of our technique to interactively explore recurring and important features of uncertain scalar fields.

1.1. Related work

Uncertainty in visualization is an active field of research [PWL97, MRH*05, PRJ12]. In particular, several recent approaches propose direct visualization techniques to represent the distribution of the uncertainty in the data [PGA13] or the positional uncertainty of level sets [PH11, PRW11, PH13, PP13]. However, these techniques do not allow for a precise extraction of features such as critical points which are shared by all realizations. In the case of certain scalar fields, topological data analysis [Mor34, Ree46, Sma61] has proven to be useful due to its capability to extract robustly critical points and to characterize their global relation [CSA00, PSBM07, TGSP09, GBHP08, GRWH12]. Also, persistent homology [ELZ02] and robustness can distinguish features which persist under small perturbations [BEKP10,BEMP13]. However, there are only few

approaches addressing the generalization of these notions to uncertain data, as discussed next.

Ensemble data is given by capturing the same phenomenon with k different observations. The similarity of the observations can be visualized by analyzing the correlation of their gradient field [STS06, NNN11]. A first topological insight in the correlated components of a function $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}^k$ is given by the notion of Jacobi sets [EH04] but their computation requires that $k \leq d$, which limits their applicability to small ensemble data-sets. Jacobi sets were used to generalize the notion of Reeb graph to Reeb spaces [EHP08] but these suffer from the same restriction. Hüttenberger et al. [HHC*13] use the notion of Pareto optimality to extract the most probable extremal structures. In particular, an extremal point is identified if it locally dominates its neighbors in the observations. Although this work introduces a connectivity between extremal points, it addresses neither the notion of saddle point nor multi-scale analysis. Several approaches propose to extend the contour tree to uncertain scalar data by computing the contour trees of each observation and merging them into a single tree with graph-matching [WZ13] or overlap-based [Kra10] heuristics. These heuristic-based techniques do not provide guaranteed predictions regarding the regions and function intervals for critical point appearance. Carr and Duke [CD13] combine contour trees into an abstract graph called the joint contour net. While it captures the union of topological information of all observations, reading their common denominator, i.e., the intersection of topological information, out of this graph remains an open question.

Analyzing directly the PDFs is an alternative to the above techniques. Otto et al. [OGHT10] extend the notion of continuous stream lines and critical points to uncertain vector fields. In contrast to this global approach, Petz et al. [PPH12] propose a local definition of an uncertain critical point. These techniques are restricted to Gaussian models. Moreover, they face numerical inaccuracies, which can yield global inconsistencies and a costly Monte-Carlo integration. Bhatia et al. [BJB*12] introduce the concept of edge maps to visualize error propagation in streamline integration. Szymczak [Szy13] propose to compute stable features in terms of (pseudo) Morse sets to analyze error affected vector fields. Both techniques depend on a refinement/subdivision parameter which affects the output, the running times, and the memory consumption. Since they address vector data, these techniques do not provide predictions on the function values of critical points, which prevents value-uncertainty analysis.

1.2. Contributions

This paper makes the following new contributions:

1. A non-local characterization of mandatory critical points;
2. A combinatorial algorithm for their extraction;
3. A tree representation capturing their global relation;
4. A mandatory critical point pair simplification scheme.

2. Preliminaries

This section describes our formal setting and presents the concepts of scalar field topology used in the rest of the paper.

2.1. Data Representation

Let $\Omega \subset \mathbb{R}^2$ be a regular grid. We call a *planar uncertain scalar field F* a function that maps each $v \in \Omega$ to a random variable $F(v)$. The input is given at each $v \in \Omega$ by the bounds of the support of a function $h_v : \mathbb{R} \to \mathbb{R}$, which faithfully describes the PDF of $F(v)$, i.e., $\int_{\mathbb{R}} h_v = 1$. Specifically, the bounds are given by two scalar functions $f^-, f^+ : \Omega \to \overline{\mathbb{R}}$ which we call *lower (f^-)* and *upper (f^+) bound fields* (Fig. 1d). This representation is generic in the sense that it can represent the bounds of parametric models and observed random variables found in ensemble data.

We call a *realization field* a scalar field $g : \Omega \to \mathbb{R}$ that maps each vertex of Ω to a realization of its random variable $F(v)$ given h_v. For example, the mean or the observations of an ensemble are instances of a realization field. The goal of our approach is to capture the common *topological denominator* of any realization g. It is based on the key property that h_v admits compact support in practice. In the worst case, this support is the interval $[F^-, F^+]$ with $F^- = \min_{v \in \Omega} f^-(v)$ and $F^+ = \max_{v \in \Omega} f^+(v)$. However, much tighter bounds are observed in practice (Fig. 1d). In the following, we assume finite f^- and f^+. In the case no bound is provided but an estimator of the PDF h_v is available, e.g., a histogram, we estimate f^- and f^+ by the support of this estimator.

The key property which we exploit in the rest of the paper is that f^- and f^+ are point-wise nested, i.e., $f^-(v) < f^+(v)$.

2.2. Critical Points of a Realization Field

We assume that any realization g behaves monotonically within each quad of Ω. Then, the critical points of g occur only on the vertices of Ω. This can be obtained by subdividing each quad of Ω into two triangles yielding a piecewise-linear (PL) manifold $\mathcal{S}$. In the following, we interpret Ω as a triangulated surface $\mathcal{S}$ and g as a PL scalar field. We additionally assume that g takes distinct values for all $v \in \mathcal{S}$, which can be obtained by simulation of simplicity [EM90].

Given an isovalue $i \in \mathbb{R}$, the *sub-level set $L_g^-(i)$* of g is $L_g^-(i) = \{p \in \mathcal{S} \mid g(p) \leq i\}$. Symmetrically, the *sur-level set $L_g^+(i)$* is $L_g^+(i) = \{p \in \mathcal{S} \mid g(p) \geq i\}$. As i increases or decreases continuously in $\mathbb{R}$, new connected components appear in $L_g^-(i)$ or $L_g^+(i)$. We call the birth vertices of these components *minima* and *maxima*, respectively. The vertices at which connected components of $L_g^-(i)$ and $L_g^+(i)$ merge as i changes are called *join* and *split saddles*, respectively. The structural relations between these critical points are captured by the notions of *join* and *split trees* [CSA00].

The *join tree $\mathcal{T}(g)$* is a 1-dimensional simplicial complex obtained by contracting each connected component of the
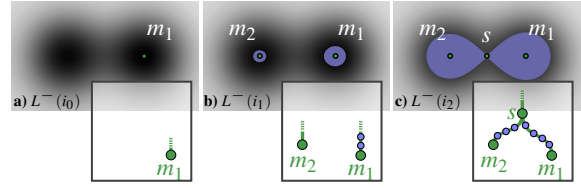

Figure 2: *Sub-level sets (blue) of a realization g with iso-values $i_0 < i_1 < i_2$. Connected components are born at i_0 a) and i_1 b) (minima m_1 and m_2 (green)), and merge at i_2 c) (join saddle s (green)). The other vertices of L_g^- (blue) are mapped to valence-2 vertices in the join tree (blue, insets).*

sub-level set to a point, see Fig. 2. Since there is a bijection between the vertices of $\mathcal{S}$ and $\mathcal{T}(g)$, we use the same notation for a vertex in $\mathcal{T}(g)$ or $\mathcal{S}$. Minima of g and its global maximum are mapped to valence-1 vertices in $\mathcal{T}(g)$ while saddles where k connected components of L_g^- merge are mapped to valence-$(k+1)$ vertices. All other vertices are mapped to valence-2 vertices (Fig. 2). A *super-arc (v_a, v_b)* is a directed connected path in $\mathcal{T}(g)$ from v_a to v_b with $g(v_a) < g(v_b)$ such that v_a and v_b are the only non-valence-2 vertices of the path. The split tree is defined symmetrically by considering the sur-level sets L_g^+ of g. Given a planar domain, there can be one pair of saddles mapped to valence-2 vertices in the join and split tree [TP12]. These are due to the boundary component of $\mathcal{S}$ and are missed in the above description, but are usually of minor interest in practice.

3. Mandatory Critical Points

In this section, we introduce the notion of mandatory critical points and we present a combinatorial algorithm to extract them. In the rest of the paper, we focus on mandatory minima and join saddles. Mandatory maxima and split saddles are treated symmetrically. According to Sec. 2, we denote the uncertain scalar field by F and a realization field by g.

The key intuition of our characterizations is that the sub-level sets of any realization g are nested in between these of f^+ and f^-, i.e., $L_{f^+}^- \subset L_g^- \subset L_{f^-}^-$. A minimum of g *can* occur if a component of $L_{f^-}^-$ appears but *must* occur if a component of $L_{f^+}^-$ appears (Fig. 3). A similar intuition holds for join saddles by identifying merge events in $L_{f^-}^-$ and $L_{f^+}^-$ (Fig. 5).

To identify these events, we make use of the join trees of f^- and f^+ both to identify the mandatory critical points and to extract their critical component and critical interval. In particular, we use $\mathcal{T}(f^+)$ to identify the mandatory vertices where topological events must have happened and $\mathcal{T}(f^-)$ to extract the corresponding critical components and intervals.

3.1. Mandatory Minima

Definition 1 A *mandatory minimum M* is a minimal connected component $C \subset \mathcal{S}$ with a minimal interval $I \subset \mathbb{R}$ such that any g admits at least one minimum m in C with $g(m) \in I$. We call C its *critical component* and I its *critical interval*.
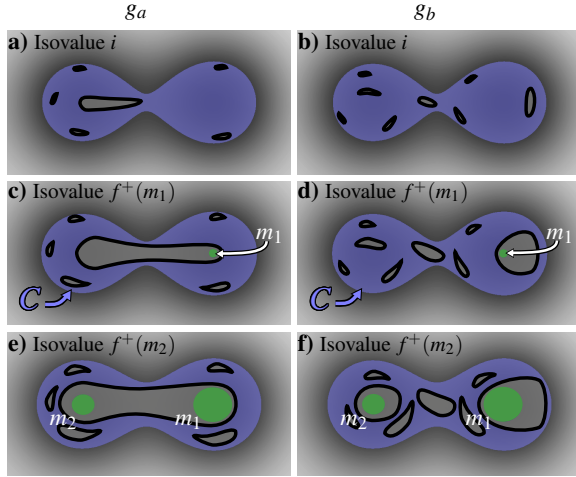
Figure 3: *Sub-level sets of f^- (blue) and f^+ (green) at three different isovalues $i < f^+(m_1) < f^+(m_2)$ with m_1 and m_2 being minima of f^+. The sub-level sets of two realizations g_a and g_b are illustrated as gray regions.*

3.1.1. Characterization of Mandatory Minima

Given an isovalue i, we denote $A^- \subset L_{f^-}^-(i)$ a connected component of the sub-level set of f^- (Fig. 3a,b, blue region). By definition of $L_{f^-}^-(i)$, there holds $f^-(v) \leq i$ for all $v \in A^-$, and we have $f^-(v) \leq g(v)$ for all $v \in \mathcal{S}$ and any g. Hence, there is a realization $\tilde{g}$ such that there is $v \in A^-$ with $\tilde{g}(v) = f^-(v) \leq i$. Let $A_{\tilde{g}} = \{v \in A^- \mid g(v) \leq i\}$. Then $A_{\tilde{g}} \neq \emptyset$, and $A_{\tilde{g}}$ consists of at least one connected component implying the existence of at least one minimum of $\tilde{g}$. Fig. 3a,b show two examples. We call A^- an *admissible component* for sub-level set realization at isovalue i.

Let m_1 be a minimum of f^+ (Fig. 3c,d). Since $g(m_1) \leq f^+(m_1)$ for any g, m_1 has to be in the sub-level set of g at $f^+(m_1)$. We call m_1 a *mandatory vertex* at isovalue $f^+(m_1)$. Since f^- and f^+ are nested, m_1 is located in $L_{f^-}^-(f^+(m_1))$.

Let C be the admissible component at isovalue $f^+(m_1)$ containing m_1. Let us assume that m_1 is the only minimum of f^+ in C (Fig. 3c,d). With the above discussion, L_g^- can be composed of an arbitrary number of connected components in C (Fig. 3a,b, gray regions). Since $g(m_1) \leq f^+(m_1)$, there is exactly one connected component of L_g^- which contains m_1 (Fig. 3c,d). Hence, any g admits at least one minimum in C. We call C the critical component of a mandatory minimum M, i.e., a connected region where at least one minimum of g has to appear. The critical interval I of M is given by the isovalue where C is born with respect to f^- up to $f^+(m_1)$, i.e., $I = [\min_{v \in C} f^-(v), f^+(m_1)]$.

Let C' be the admissible component at isovalue i containing m_1 with $i < f^+(m_1)$ and $C' \subsetneq C$. Before the isovalue $f^+(m_1)$, there is a realization $\tilde{g}$ such that $L_{\tilde{g}}^- \cap C'$ is empty, i.e., $\tilde{g}$ does not admit any minimum in C'. However, any g must admit at least one minimum arbitrarily located within
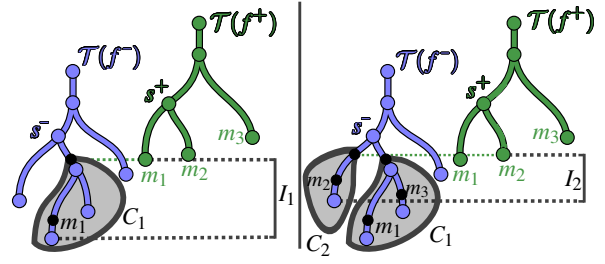

Figure 4: *Minima m_k of $\mathcal{T}(f^+)$ are traversed by increasing value of f^+ (m_1, m_2, m_3). For each m_k, the critical component C_k corresponds to the sub-tree of $\mathcal{T}(f^-)$ containing m_k and rooted at the isovalue $f^+(m_k)$. The component is valid if it contains no previously processed minimum of $\mathcal{T}(f^+)$.*

C at the isovalue $f^+(m_1)$. Thus, C describes a critical component of minimal size. Consider the case that m_1 and m_2 are two minima of f^+ located in C with $f^+(m_1) < f^+(m_2)$ (Fig. 3e,f). Since $g(m_1) \leq f^+(m_2)$ and $g(m_2) \leq f^+(m_2)$, $L_g^-(f^+(m_2))$ includes m_1 and m_2 for any g. Then, there is $\tilde{g}$ such that $L_{\tilde{g}}^-(f^+(m_2))$ consists of only one connected component which contains both minima (Fig. 3e). Thus, the existence of m_2 in C does not necessarily imply the existence of an additional minimum in g. Hence, the existence of several minima of f^+ in C do not introduce topological but geometrical constraints to the sub-level sets of g (Fig. 3e,f).

3.1.2. Extraction of Mandatory Minima

We now introduce a combinatorial algorithm to extract the mandatory minima of F (Fig. 4). The minima of $\mathcal{T}(f^+)$ are visited by increasing value of f^+. For each minimum m_k of f^+, the component C_k (Fig. 3c,d) is computed by extracting the sub-tree of $\mathcal{T}(f^-)$ containing m_k and rooted at the isovalue $f^+(m_k)$. The component C_k describes a valid mandatory minimum if it contains no previously processed minimum of $\mathcal{T}(f^+)$ and if m_k is not included in any previously computed critical component. For example, m_1 and m_2 in Fig. 4 yield the creation of two mandatory minima M_1 and M_2 with components C_1 and C_2. This is not the case for m_3 since it is included in the previously computed critical component C_1. The critical interval I_k is given by $[\min_{v \in C_k} f^-(v), f^+(m_k)]$. The output is a list of mandatory minima M_k with their critical component C_k and interval I_k.

3.2. Mandatory Join Saddles

Definition 2 A *mandatory join saddle* S is a minimal connected component $C \subset \mathcal{S}$ with a minimal interval $I \subset \mathbb{R}$ such that any g admits at least one join saddle s in C with $g(s) \in I$. We call C its *critical component* and I its *critical interval*.

3.2.1. Characterization of Mandatory Join Saddles

Let M_k with $k \in \{1,2\}$ be two mandatory minima with their critical components C_k. Let $m_k \in C_k$ be their mandatory vertices (Fig. 5a,b). Note that m_k are minima of f^+. The com-
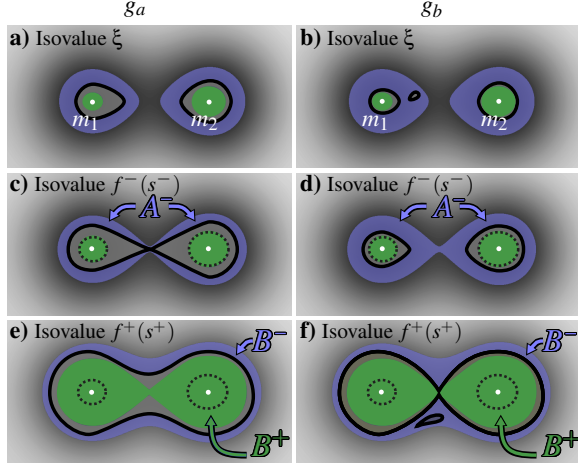
Figure 5: *Sub-level sets of f^- (blue) and f^+ (green), and of two realizations g_a and g_b (gray regions, a)-f)) for the isoval-ues $\xi < f^-(s^-) < f^+(s^+)$. The set of vertices R is enclosed by the dashed lines c)-f). Two components of a realization can merge from the isovalue $f^-(s^-)$ c),d) but must merge by the isovalue $f^+(s^+)$ e),f).*

mon saddle ancestors of m_k in $\mathcal{T}(f^-)$ and $\mathcal{T}(f^+)$ are de-noted by s^- and s^+ (Fig. 4). Without loss of generality, we assume s^- and s^+ are simple saddles. Let A^- be the con-nected component of $L_{f^-}^-(f^-(s^-))$ containing s^-. Since C_1 and C_2 are disconnected (Sec. 3.1.2), we have $s^- \notin C_k$ and $C_k \subset A^-$. Thus, m_k must be contained in A^- (Fig. 5c,d).

We characterize mandatory join saddles in three steps. We analyze the sub-level set of g before $f^-(s^-)$ (step 1), after $f^-(s^-)$ (step 2), and after $f^+(s^+)$ (step 3).

Step 1 (Fig. 5a,b). Let $\xi < f^-(s^-)$ such that there are two connected components $A_\xi^-(m_1), A_\xi^-(m_2) \subset A^-$ of $L_{f^-}^-(\xi)$ each of which containing m_k, respectively. Since $f^-(m_k) < f^-(s^-)$, the sets $A_\xi^-(m_k)$ always exist. Following Sec. 3.1.1, any realization g must also admit at least two connected components of $L_g^-(\xi)$ which contain m_1 and m_2, respectively.

Step 2 (Fig. 5c,d). Since A^- consists of only one con-nected component, there is a realization $\tilde{g}$ for which a sin-gle connected component of $L_{\tilde{g}}^-(f^-(s^-))$ includes m_1 and m_2 (Sec. 3.1.1). $f^-(s^-)$ describes the earliest moment where a join saddle for any g in A^- can occur. Hence, it cannot occur in the set of *finished vertices* $R = \{v \in A^- \mid f^+(v) < f^-(s^-)\}$.

Step 3 (Fig. 5e,f). Let $B^- \subset L_{f^-}^-(f^+(s^+))$ and $B^+ \subset L_{f^+}^-(f^+(s^+))$ be the connected components of the sub-level set of f^- and f^+ containing s^+. Since f^- and f^+ are nested, we have $B^+ \subset B^-$ and $m_1, m_2 \in B^+$. Note that we also have $A^- \subset B^-$. For each $v \in B^+$, we have $f^+(v) \leq f^+(s^+)$ and, hence, $g(v) \leq f^+(v) \leq f^+(s^+)$ for any g. It follows that $B^+ \subset L_g^-(f^+(s^+))$. Since B^+ is connected, there is exactly one connected component B_g of $L_g^-(f^+(s^+))$ with $B^+ \subset B_g$. By definition, we have $g(v) \leq f^+(s^+)$ for each $v \in B_g$. Since
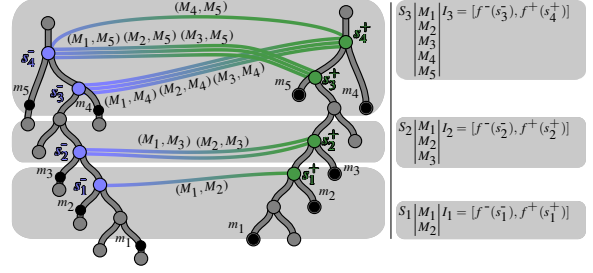

Figure 6: *Left: Bipartite graph G between the saddles S^- of $\mathcal{T}(f^-)$ (left) and S^+ of $\mathcal{T}(f^+)$ (right). Each connected com-ponent of G yields a unique mandatory n-join saddle (top to bottom: $n = 2, 1, 1$). Right: For each S_k, the list $\mathcal{M}_k$ of its mandatory minima and its critical interval I_k is maintained.*

$B^+ \subset B^-$, it follows that $B_g \subset B^-$. Hence, there is exactly one connected component of $L_g^-(f^+(s^+))$ that contains m_1 and m_2, and which is contained in B^-.

Overall, the admissible component C' where a join sad-dle can occur is given by $C' = B^- \setminus R$. Due to the presence of split saddles of f^- and f^+, the removal of R in C' might yield disconnected regions. Thus, we restrict C' to the connected component C containing s^- and s^+. The mandatory join sad-dle S that merges the mandatory minima M_k is defined by the critical component C and interval $I = [f^-(s^-), f^+(s^+)]$.

In summary, at isovalue ξ (step 1) any g must have two connected components of L_g^- each of which containing m_1 and m_2. At isovalue $f^-(s^-)$, there exists a realization $\tilde{g}$ for which one connected component of $L_{\tilde{g}}^-$ includes both m_1 and m_2. At isovalue $f^+(s^+)$ (step 3), any g admits one connected component of L_g^- including m_1 and m_2. Hence, C describes a critical component of minimal size in which a join saddle s must appear by the isovalue $f^+(s^+)$ (Fig. 5e,f).

We want to emphasize that the sub-level set component of g which initially contains m_1 at isovalue $f^+(m_1)$ con-tains at least one minimum in the critical component C_1 (Sec. 3.1.1). The same holds for m_2. Thus, there is a minima-pair (m_a, m_b) for any g with $m_a \in C_1$ and $m_b \in C_2$ such that their join saddle is located in C of S and takes a value in I.

3.2.2. Enumeration of Mandatory Join Saddles

We present now a strategy to enumerate all mandatory join saddles. This enumeration is used to construct the mandatory join tree and to simplify mandatory critical points.

To enumerate all mandatory join saddles, we enumerate all pairs of mandatory minima. According to Sec. 3.2.1, a mandatory join saddle is uniquely identified by its cor-responding ancestor saddles s^- and s^+. Any two pairs of mandatory minima (M_i, M_j) and (M_k, M_ℓ) which use the same pair (s^-, s^+) characterize the same mandatory saddle (Fig. 6). This yields duplicated entries, handled as follows.

Let S^- and S^+ be the list of join saddles of $\mathcal{T}(f^-)$ and $\mathcal{T}(f^+)$ associated to mandatory minima pairs. Let G be a
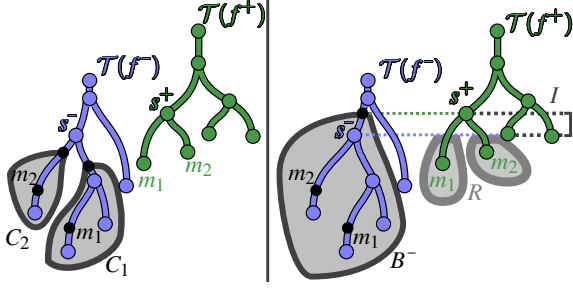
Figure 7: *Mandatory join saddle extraction. Left: Given a pair of mandatory minima (M_1, M_2), the common ancestor of their mandatory vertices $m_{1,2}$ are extracted in $\mathcal{T}(f^-)$ and $\mathcal{T}(f^+)$. Right: B^- is given by the sub-tree of $\mathcal{T}(f^-)$ rooted at the vertex obtained by walking up from s^- to $f^+(s^+)$.*

bipartite graph such that each saddle of S^- and S^+ is represented by a node (Fig. 6). The edges of G link a saddle from S^- to a saddle of S^+ if these saddles are used by a mandatory minimum pair (M_i, M_j). Each unique edge (s_k^-, s_k^+) of G represents a unique mandatory join saddle S_k.

Since the uncertainty can be spread out arbitrarily across the domain, the size of the support of a PDF at a vertex v can vary across $\mathcal{S}$. This might result in nested mandatory saddles (Fig. 6, top). A mandatory join saddle S_k can completely hide another S_ℓ in terms of critical component and interval, i.e., $C_\ell \subset C_k$ and $I_\ell \subset I_k$. In such a case, we represent S_k and S_ℓ by a single mandatory 2-join saddle. Given a realization g, this mandatory 2-join saddle describes two join saddles of g that have to appear in C_k and I_k. This construction is generalized to *mandatory n-join saddles*, where n is called the *multiplicity* of the mandatory saddles. Mandatory saddles are nested if they belong to the same connected component of G. The saddle S_k with the largest critical component and interval is selected as a representative. Each S_k is associated with its list of minima $\mathcal{M}_k$ which define its component in G. The multiplicity of S_k is given by the number of saddles of S^- in this component. In Fig. 6, the bottom component containing $\{s_1^-, s_1^+\}$ yields the mandatory saddle S_1 of multiplicity one. The top component containing $\{s_3^-, s_3^+, s_4^-, s_4^+\}$ yields S_3 of multiplicity two.

3.2.3. Extraction of Mandatory Join Saddles

We present the overall algorithm for computing mandatory join saddles (Fig. 6,7). For each pair of mandatory minima (M_i, M_j), the corresponding join saddles of $\mathcal{T}(f^-)$ and $\mathcal{T}(f^+)$ are identified and an edge is added to G (Fig. 6, left). Given a connected component of G, a mandatory saddle S_k is created (Fig. 6, right). Let s^- and s^+ be the lowest and highest saddles of this component. The critical interval I_k is given by $I_k = [f^-(s^-), f^+(s^+)]$ (Fig. 7). The critical component C_k of S_k is given by walking up $\mathcal{T}(f^-)$ from s^- up to $f^+(s^+)$ (Fig. 7) and considering the connected subset of $B^- \setminus R$ (Sec. 3.2.1). Finally, the list $\mathcal{M}_k$ of mandatory minima that S_k joins is maintained (Fig. 6, right).
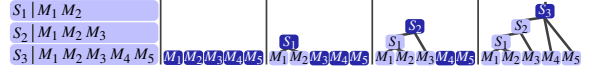

Figure 8: *Mandatory join tree construction (left to right). Saddles are progressively added to the tree (top to bottom, left). A union-find data-structure tracks the added edges, enabling to identify the non-visited nodes to which a saddle node must be linked, given its minima list. At each step, visited nodes are depicted in light blue.*

3.3. Computational Complexity

In this section, we analyze the computational complexity of our algorithm. We denote the number of vertices of $\mathcal{S}$ by n and the number of mandatory extrema by e. The computation of the join trees is bounded by $O(n \log(n))$ [CSA00]. In Sec. 3.1.2, all minima of f^+ are processed in increasing order. Thus, the worst-case complexity of this step is $O(n \log(n))$. Since the mandatory minima are disjoint, the computation of the critical components is done in $O(n)$. Hence, the worst-case complexity for the computation of mandatory extrema is $O(n \log(n))$. Enumerating the list of extrema pairs in Sec. 3.2.3 needs $O(e^2)$ steps. For each pair, the computation of the common ancestors s^- and s^+ needs $O(n)$ steps. Thus, the computation of the pairing has a worst-case complexity of $O(ne^2)$. The overall complexity of our algorithm is therefore $O(n \log(n) + ne^2)$.

4. Visualization of Mandatory Critical Points

In this section, we show how the mandatory critical points can be used for interactive and multi-scale visualizations.

4.1. Visualization of Mandatory Critical Components

Each mandatory critical point P is visualized by displaying its critical component C on the domain, with a specific color according to its type (minimum, join saddle, split saddle, or maximum). More insights can be gleaned by analyzing the critical interval I of each P. In the following, we assume that not only the support bounds of each PDF is given as input but also the PDF estimators. For each vertex $v \in C$, the probability that $F(v) \in I$ is depicted as an opacity transfer function giving visual hints on the spatial uncertainty of P.

4.2. Mandatory Join Tree Construction

The algorithm for the mandatory join saddle computation maintains for each join saddle S_k the list $\mathcal{M}_k$ of mandatory minima it joins (Fig. 6, right). We exploit this information to visualize the global structure of the mandatory critical points in a tree representation, called the *mandatory join tree*. Its computation (Fig. 8) is an adaptation of the join tree algorithm [CSA00] to our setting. In particular, the mandatory saddles S_k are processed in increasing order of $|\mathcal{M}_k|$. This ordering is motivated by the following fact. In a scalar field, the closer a join saddle is to the root, the more minima are contained in its sub-level set component. Based on this order, the mandatory join tree is constructed as follows.

Data set	n	N	e	s	f^-, f^+ (sec)	E.C. (sec)	S.C. (sec)	**Overall** (sec)	O.V. (sec)	R.V. (sec)
Vortex Street	33,345	100	47	*35*	0.1	< 0.1	< 0.1	**0.1**	0.2	0.5
Heated Cylinder	197,633	25	8	*5*	0.4	0.1	< 0.1	**0.5**	0.5	4.7
Synthetic	262,144		14	12	0.5	0.9	0.1	**1.5**	0.9	0.9
Sea Surface Height	1,036,800	31	860	*235*	2.1	0.6	18.0	**20.7**	10,073.2	76,760.7
Mars	1,476,993		995	*784*	4.0	1.5	130.3	**135.8**	81,480.6	> 86,400.0

Table 1: *Running times (in seconds). The size of the domains is given by the number of vertices. We use the following notation: n (number of vertices in the domain), N (number of fields in an ensemble if applicable), e (number of mandatory extrema), s (number of mandatory saddles, italic if multiple saddles are present), f^- and f^+ (computation of the join/split trees of f^-, f^+), E.C. and S.C. (computation of the mandatory extrema and saddles). The last two columns report the computation time for the verification of a single realization field (O.V. observation fields, R.V. random realizations from the PDFs).*

The mandatory join tree T^- is initialized with nodes representing the saddles and minima. Edges are added between the first saddle S_1 and its minima. These minima are marked as visited (Fig. 8, light blue). A union-find data-structure [GF64, CLRS01] tracks the added edges. This enables to find the non-visited nodes to which a remaining saddle S_k should be linked, based on its list $\mathcal{M}_k$ (Fig. 8). In contrast to the join tree (Sec. 2), we do not augment T^- with valence-2 vertices. T^- only includes the mandatory critical points and the edges linking them. Let S_k and S_l be two saddles linked by an edge in T^-. If S_k and S_l overlap in terms of critical component and interval, the corresponding join saddles s_k and s_l in a realization g may admit an arbitrary order. Thus, s_k and s_l may switch their positions in $\mathcal{T}(g)$. We flag each edge of T^- as *switchable* in such a case.

The mandatory join tree can be visualized with a planar layout [Aub04, HSCS11]. In particular, a mandatory critical point is represented by a vertical bar whose length depicts its interval, and switchable edges are colored. From a practical point of view, an edge of T^- represents a connected path of super-arcs in $\mathcal{T}(g)$ for any g. Thus, T^- describes the topological connection between critical points of any g, which provides visual hints of its expected topological structure. The mandatory split tree T^+ is constructed symmetrically.

4.3. Simplification of Mandatory Critical Points

In this section, we introduce a procedure to discriminate between small- and large-scale structures represented by mandatory critical points. This strategy is inspired by the notion of topological persistence [ELZ02]. For each mandatory critical point P, we note f_P^- and f_P^+ the bounds of its critical interval $I = [f_P^-, f_P^+]$. For each mandatory join saddle S, the list of its mandatory minima $\mathcal{M}$ is traversed (Fig. 6, right). A mandatory critical point pair (M, S) is created for each minimum $M \in \mathcal{M}$ and we evaluate the metric: $d(M, S) = |f_S^+ - f_M^-|$. Note that a mandatory minimum M can be represented in multiple pairs. Given a user-defined threshold t, the global list of pairs is traversed by increasing d-value. A pair (M, S) with $d(M, S) < t$ is marked as *simplified* and removed from the visualization only if both M and S have not already been simplified. After all pairs have been processed, the mandatory join/split trees are reconstructed by considering the non-simplified mandatory critical points.

5. Experimental Results

In this section, we present experimental results obtained with ensemble data sets and scalar fields with prescribed error bounds, and provide empirical verifications of our predictions. The results are obtained with a C++ implementation on an i7 CPU (2.93GHz). Tab. 1 shows running times of the individual steps. Note that the timings depend on the topological complexity, i.e., the number of mandatory extrema. Critical components are colored as follows: minima (blue), join saddles (green), split saddles (yellow), maxima (red).

5.1. Verification and Performance

We verified our prediction of mandatory critical points and their global relation with two test procedures. Let g be an arbitrary realization. In Test 1, we checked if there is a critical point of g of correct type located in the critical component of a mandatory critical point. We also checked if its critical value is in the predicted critical interval. In Test 2, we verified the global pairing of mandatory critical points as follows. Let $M_1, M_2 \in \mathcal{M}$ be two extrema in the list $\mathcal{M}$ of a mandatory saddle S (Sec. 3.2.3). Let $C_{1,2}$ and $I_{1,2}$ be the critical components and intervals of the mandatory extrema $M_{1,2}$. We collected all extrema-saddle triplets (m_1, m_2, s) of g with $m_1 \in C_1, g(m_1) \in I_1, m_2 \in C_2, g(m_2) \in I_2$, and s being the common ancestor of m_1 and m_2 in the join/split tree of g. The prediction of our algorithm is correct if the set of triplets is not empty, and s and its critical value fall in the critical component and interval of S. We applied this verification to all mandatory extrema-saddle triplets.

Test 1 and 2 were used in a twofold manner. Firstly, we verified the observed data, i.e., the topological structure of each individual observation. Scalar fields with prescribed error bounds were handled as ensemble data containing the lower and upper bounds. All data sets presented in Tab. 1 passed successfully Test 1 and 2. Secondly, we generated 5000 random realizations based on the PDFs and verified these fields. Uniform distributions were assumed for scalar fields with prescribed error bounds. Note that random realizations usually result in a much more complex topological configuration due to the induced noise. Thus, the verification for the Sea Surface Height and the Martian data was not completed (more than 20 hours per realization, see Tab. 1). The global topological structure of all remaining data sets
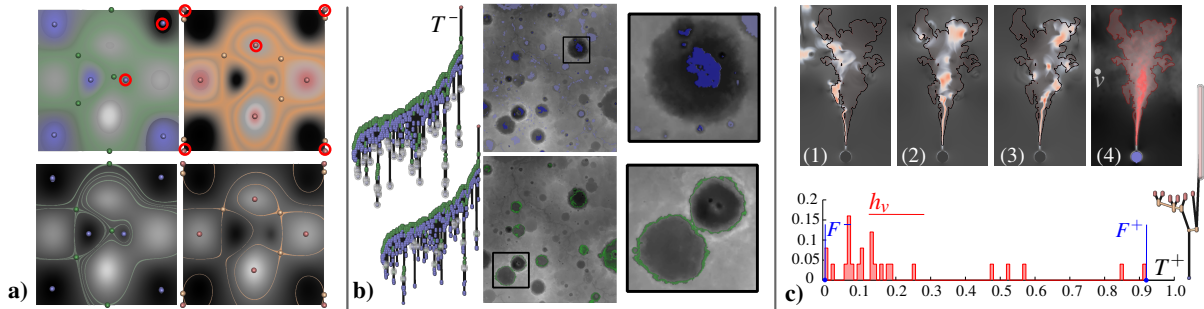
Figure 9: *Illustration of mandatory critical points and mandatory join/split trees: a) A synthetic scalar field with global error bounds (20% (top), 0.5% (bottom)). The critical points of the synthetic field (ground truth) are shown with colored spheres. Red circles indicate extrema of the synthetic field which are not extracted as mandatory features. b) User-selected critical points in a Martian elevation map using the mandatory join tree. Zoom-ins illustrate their geometrical structure and spatial uncertainty. c) Magnitude of velocity fields caused by a heated cylinder. The most dominant mandatory maximum is depicted with a black line (1-3). The shades of red illustrate the spatial uncertainty within this region (4). The PDF of a $v \in \mathcal{S}$ is estimated by a histogram.*

of Tab. 1 was verified as correct. The verification requires more computational effort than the algorithm itself (Tab. 1). This is especially the case for random realization due to the induced noise. In all of our experiments, the construction of the mandatory join/split tree and its simplification took less than 5 ms. This enables an interactive exploration and visualization once the initial set of mandatory critical points is computed (as showcased in the accompanying video).

5.2. Experiments

Kármán Vortex Street. Fig. 1 illustrates our algorithmic pipeline on the uncertain Kármán vortex street obtained with Gerris [Pop03]. The simulation has been run for a fixed duration with varying viscosity parameters yielding an ensemble data set which represents the magnitude of the velocity field. In such an experiment, the resulting point-wise histograms h_v (Fig. 1a) do not necessarily follow a parametric distribution. Based on f^- and f^+ (Fig. 1b), our algorithm computes the mandatory critical points (Fig. 1c). The geometry of the critical components enable to apprehend the spatial uncertainty of the topological features of the flow. In particular, the shape of the critical components of the mandatory minima varies along the street: thinner and more elongated regions are observed at its right extremity. This indicates a larger spatial uncertainty away from the obstacle. A refined visualization of the spatial uncertainty is obtained by the point-wise likelihood of appearance (Sec. 4.1), see Fig. 1d. The mandatory join and split trees (Fig. 1e) capture nicely the periodicity and the symmetry of the vortex street. The length of the vertical bars in the trees visualizes the value-uncertainty of the mandatory critical points. A progressive simplification of mandatory critical points (Fig. 1f) reveals the most dominant structures (Fig. 1g). Interestingly, the trees show that these are also the most uncertain value-wise.

Synthetic Evaluation. Fig. 9a shows the mandatory critical points of a synthetic scalar field with two global error bounds based on the function range (top: 20%, bottom:

0.5%). We additionally show the critical points (spheres) of the synthetic field as ground truth. With 20% error only few critical points are detected as mandatory. Missing critical points are highlighted by red circles. The global topological structure of the scalar field is identified as mandatory in case of 0.5% error. As the error bound width decreases, f^- and f^+ converge to the synthetic field. In practice, the critical components of mandatory critical points tend to describe extrema points and saddle contours. This indicates a capturing of topological changes of certain level sets in the limit.

Martian Elevation Map. Fig. 9b shows the interactive exploration of mandatory critical points in an elevation map of a Martian region [HP03] (http://pds-geosciences.wustl.edu/). This data set has a spatial resolution of ∼460 m per pixel and we consider an elevation error of 1% (∼40 m). The spatial organization of craters is of special interest helping to understand properties of celestial bodies such as age and origin. Based on the mandatory join tree, the user can explore interactively the most dominant mandatory minima to identify the deepest craters. The mandatory join saddles enable a navigation of the rims of these craters. Our technique provides assessments regarding the position and height uncertainties of the craters.

Heated Cylinder. Fig. 9c shows a challenging example for our technique. It represents the flow generated by a heat source with varying viscosity. In this unstable process, the flow trajectories drastically change from one observation to another (Fig. 9c (1-3)). This instability is reflected by a large support in the histograms h_v, which sometimes cover the entire histogram range (Fig. 9c). Despite the high variability of the observations, our approach still identifies several mandatory critical points including a dominant mandatory maximum (Fig. 9c (4), red colored region). As shown in Fig. 9c (1-3), all the observations admit at least one maximum of velocity magnitude located in the predicted region. Interestingly, they all also admit a maximum in the early stage of
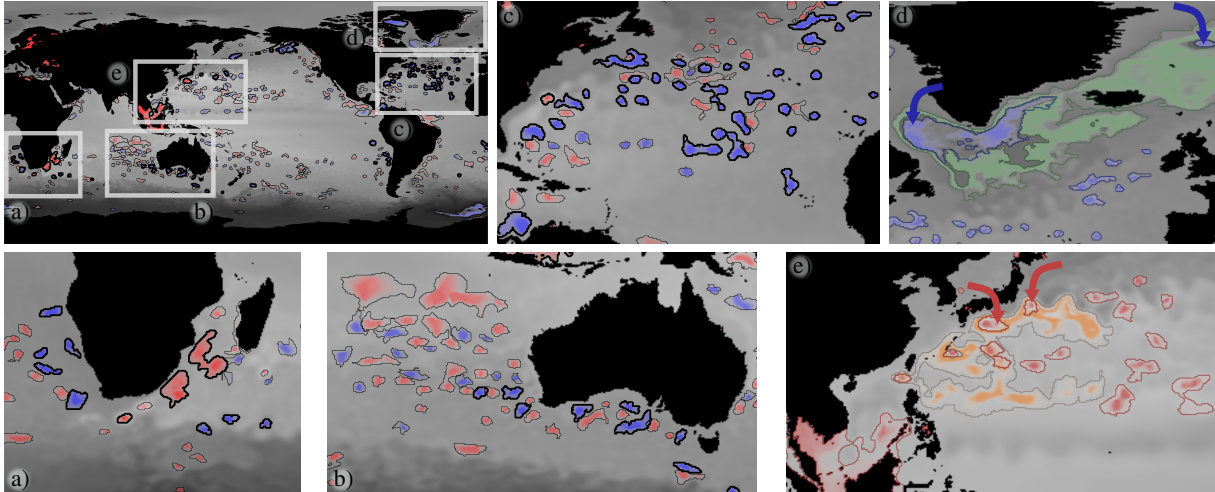
Figure 10: *Mandatory critical points on the Sea Surface Height data set. Our simplification scheme allows the user to distinguish more dominant features in the data (top left), marked with bold boundaries. This drives the selection of regions of interest (a, b, c) that can be further inspected with zoom-ins. Each of these reveal the local structure of a major sea stream. The visualization of the mandatory saddles (d, e) helps understanding the relation between the extrema (arrows) found in the data.*

the stream. This common behavior is captured by the cone-shaped structure of the dominant mandatory maximum.

Sea Surface Height. Fig. 10 shows the mandatory critical points of the Sea Surface Height (SSH) computed from 31 observations (http://ecco2.jpl.nasa.gov/products/, Jan. 2012). The algorithm identifies regions where minima and maxima exist over the entire month. This enables to apprehend the global structure of the ensemble. Local inspection provides further insights in this structure (Fig. 10d-e). Our simplification allows to identify dominant critical points (shown with bold boundaries). This drives a selection of regions of interest (Fig. 10a-c) that can be further inspected with zoom-ins. In particular, they reveal that the dominant extrema of SSH are located along the major sea streams: along South-Africa (a), Australia (b), and in the Gulf Stream (c). This simplification helps the user overcome the amount of information given in this complex ensemble data set.

5.3. Limitations

The key assumption of our approach is that the PDFs of the input admit a finite support at each vertex of the domain and that a reliable estimation of this support is given. When constructing these PDFs from ensemble data-sets, the support bounds (and the resulting critical components and intervals) may be over-estimated in the presence of outliers. Also, PDFs with infinite support can only be processed in an approximated way. The pairing of mandatory critical points takes $O(n e^2)$. Hence, only the critical components and intervals of mandatory extrema can be computed for highly complex data sets. As the uncertainty decreases, mandatory extrema shrink to points. Mandatory saddles, however, tend to converge to saddle contours. Thus, the prediction accuracy of the exact position of a saddle point is restricted by the

spacial extent of this contour. In practice, this implies that saddles have a larger extent than extrema in the presence of uncertainty, which may challenge their visual interpretation.

6. Conclusion

In this paper, we introduced a combinatorial technique which allows to extract the common topological denominator of uncertain scalar fields. To do so, we introduced the notion of mandatory critical points. These describe regions and intervals where critical points and values have to happen in any realization of the input, which help the users apprehend the spatial and value uncertainties of topological features. Thanks to this notion, we described new interactive visualization techniques for uncertain data. In particular, we introduced mandatory join/split trees to reveal the global relation of mandatory critical points. We also presented a simplification strategy which enables multi-scale visualizations.

For the special case where the PDF support width is constant, we suspect a more efficient algorithm could be derived. An improvement for the general case, i.e., spatial variability of the uncertainty, is more challenging but could be beneficial to the processing of highly complex data. An investigation of the relation between the proposed simplification strategy and topological persistence could provide deeper theoretical insights of the topological structure of uncertain data. Finally, a natural direction for future work is the extension of this approach to volumetric data-sets. However, as the dimension increases, new types of saddle points appear in the bound fields and these are not captured by the join/split trees.

References

[Aub04] AUBER D.: Tulip - a huge graph visualization framework. *Graph Drawing Software* (2004), 105–126. 7

[BEKP10] BENDICH P., EDELSBRUNNER H., KERBER M., PATEL A.: Persistent homology under non-uniform error. In *Mathematical Foundations of Computer Science*, vol. 6281 of *Lecture Notes in Computer Science*. Springer, 2010, pp. 12–23. 2

[BEMP13] BENDICH P., EDELSBRUNNER H., MOROZOV D., PATEL A. K.: Homology and robustness of level and interlevel sets. *Homology, Homotopy and Applications 15*, 1 (2013), 51–72. 2

[BJB*12] BHATIA H., JADHAV S., BREMER P., CHEN G., LEVINE J., NONATO L., PASCUCCI V.: Flow visualization with quantified spatial and temporal errors using edge maps. *TVCG 18*, 9 (2012), 1383–1396. 2

[CD13] CARR H., DUKE D.: Joint contour nets: Computation and properties. In *Proc. of PacificVis* (2013). 2

[CLRS01] CORMEN T., LEISERSON C., RIVEST R., STEIN C.: *Introduction to Algorithms*, 2nd ed. McGraw-Hill Higher Education, 2001. 7

[CSA00] CARR H., SNOEYINK J., AXEN U.: Computing contour trees in all dimensions. In *Proc. of Symposium on Discrete Algorithms* (2000), pp. 918–926. 2, 3, 6

[EH04] EDELSBRUNNER H., HARER. J.: *Jacobi Sets of Multiple Morse Functions*. Cambridge Books Online, 2004. 2

[EHP08] EDELSBRUNNER H., HARER J., PATEL A. K.: Reeb spaces of piecewise linear mappings. In *Proc. of ACM Symp. on Comp. Geom.* (2008), pp. 242–250. 2

[ELZ02] EDELSBRUNNER H., LETSCHER D., ZOMORODIAN A.: Topological persistence and simplification. *Discrete & Computational Geometry 28* (2002), 511–533. 2, 7

[EM90] EDELSBRUNNER H., MUCKE E. P.: Simulation of simplicity: a technique to cope with degenerate cases in geometric algorithms. *ACM Trans. on Graph. 9* (1990), 66–104. 3

[GBHP08] GYULASSY A., BREMER P.-T., HAMANN B., PASCUCCI P.: A practical approach to Morse-Smale complex computation: scalabity and generality. *TVCG* (2008), 1619–1626. 2

[GF64] GALLER B., FISCHER M.: An improved equivalence algorithm. *Communications of the ACM 7* (1964), 301–303. 7

[GRWH12] GÜNTHER D., REININGHAUS J., WAGNER H., HOTZ I.: Efficient computation of 3D Morse-Smale complexes and persistent homology using discrete Morse theory. *The Visual Computer 28* (2012), 959–969. 2

[GUM08] ISO/IEC Guide 98-3:2008 uncertainty of measurement-part 3: Guide to the expression of uncertainty in measurement (GUM), 2008. 2

[HHC*13] HÜTTENBERGER L., HEINE C., CARR H., SCHEUERMANN G., GARTH C.: Towards multifield scalar topology based on pareto optimality. *Comput. Graph. Forum 32*, 3 (2013), 341–350. 2

[HP03] HYNEK B. M., PHILLIPS R. J.: New data reveal mature, integrated drainage systems on mars indicative of past precipitation. *Geology 31*, 9 (Sep. 2003), 757–760. 8

[HSCS11] HEINE C., SCHNEIDER D., CARR H., SCHEUERMANN G.: Drawing contour trees in the plane. *TVCG 17* (2011), 1599–1611. 7

[Kra10] KRAUS M.: Visualization of uncertain contour trees. In *IMAGAPP/IVAPP* (2010), Richard P., Braz J., (Eds.), INSTICC Press, pp. 132–139. 2

[Mor34] MORSE M.: *The Calculus of Variations in the Large*. No. v. 18 in Colloquium Publications - American Mathematical Society. American Mathematical Society, 1934. 2

[MRH*05] MACEACHREN A. M., ROBINSON A., HOPPER S., GARDNER S., MURRAY R., GAHEGAN M., HETZLER E.: Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science 32*, 3 (2005), 139–160. 2

[NNN11] NAGARAJ S., NATARAJAN V., NANJUNDIAH R.: A gradient-based comparison measure for visual analysis of multifield data. *CGF* (2011). 2

[OGHT10] OTTO M., GERMER T., HEGE H.-C., THEISEL H.: Uncertain 2D vector field topology. *CGF 29* (2010), 347–356. 2

[PGA13] POTTER K., GERBER S., ANDERSON E.: Visualization of uncertainty without a mean. *IEEE CGA 33* (2013), 75–79. 2

[PH11] PÖTHKOW K., HEGE H.-C.: Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *TVCG 17*, 10 (2011), 1393–1406. 2

[PH13] PÖTHKOW K., HEGE H.-C.: Nonparametric models for uncertainty visualization. *CGF 32* (2013), 131–140. 2

[Pop03] POPINET S.: Gerris: A tree-based adaptive solver for the incompressible euler equations in complex geometries. *J. Comp. Phys 190* (2003), 572–600. URL: http://gfs.sf.net/. 8

[PP13] PÄTZ T., PREUSSER T.: Segmentation of stochastic images using level set propagation with uncertain speed. *Journal of Mathematical Imaging and Vision* (2013), 1–21. 2

[PPH12] PETZ C., PÖTHKOW K., HEGE H.-C.: Probabilistic local features in uncertain vector fields with spatial correlation. *CGF 31*, 3pt2 (2012), 1045–1054. 2

[PRJ12] POTTER K., ROSEN P., JOHNSON C. R.: From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Uncertainty Quantification in Scientific Computing*, vol. 377. Springer, 2012, pp. 226–249. 2

[PRW11] PFAFFELMOSER T., REITINGER M., WESTERMANN R.: Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *CGF 30* (2011), 951–960. 2

[PSBM07] PASCUCCI V., SCORZELLI G., BREMER P. T., MASCARENHAS A.: Robust on-line computation of Reeb graphs: simplicity and speed. *TOG 26* (2007), 58.1–58.9. 2

[PWL97] PANG A. T., WITTENBRINK C. M., LODHA S. K.: Approaches to uncertainty visualization. *The Visual Computer 13*, 8 (1997), 370–390. 2

[Ree46] REEB G.: Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *Comptes-rend. de l'Acad. des Scien. 222* (1946), 847–849. 2

[Sma61] SMALE S.: On gradient dynamical systems. *The Annals of Mathematics 74* (1961), 199–206. 2

[STS06] SAUBER N., THEISEL H., SEIDEL H.: MultifieldGraphs: An approach to visualizing correlations in multifield scalar data. *TVCG 12* (2006), 917–924. 2

[Szy13] SZYMCZAK A.: Hierarchy of stable morse decompositions. *TVCG 19*, 5 (2013), 799–810. 2

[TGSP09] TIERNY J., GYULASSY A., SIMON E., PASCUCCI V.: Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees. *TVCG 15* (2009), 1177–1184. 2

[TP12] TIERNY J., PASCUCCI V.: Generalized topological simplification of scalar fields on surfaces. *TVCG* (2012). 3

[WZ13] WU K., ZHANG S.: A contour tree based visualization for exploring data with uncertainty. *International Journal for Uncertainty Quantification 3*, 3 (2013), 203–223. 2