Positional Uncertainty of Isocontours: Condition Analysis and Probabilistic Measures

Kai Pöthkow and Hans-Christian Hege, *Member*, IEEE

Abstract—Uncertainty is ubiquitous in science, engineering and medicine. Drawing conclusions from uncertain data is the normal case, not an exception. While the field of statistical graphics is well established, only a few 2D and 3D visualization and feature extraction methods have been devised that consider uncertainty. We present mathematical formulations for uncertain equivalents of isocontours based on standard probability theory and statistics and employ them in interactive visualization methods. As input data, we consider discretized uncertain scalar fields and model these as random fields. To create a continuous representation suitable for visualization we introduce interpolated probability density functions. Furthermore, we introduce *numerical condition* as a general means in feature-based visualization. The condition number—which potentially diverges in the isocontour problem—describes how errors in the input data are amplified in feature computation. We show how the average numerical condition of isocontours aids the selection of thresholds that correspond to robust isocontours. Additionally, we introduce the *isocontour density* and the *level crossing probability field*; these two measures for the spatial distribution of uncertain isocontours are directly based on the probabilistic model of the input data. Finally, we adapt interactive visualization methods to evaluate and display these measures and apply them to 2D and 3D data sets.

Index Terms—Uncertainty, probability, isolines, isosurfaces, numerical condition, error analysis, volume visualization.

---

1 INTRODUCTION

NUMERICAL quantities with continuous range, like scalar variables $y \in \mathbb{R}$, can be measured only with finite precision. Therefore, their exact "true" values are unknown, i.e., all measured values of such quantities are afflicted with some uncertainty. In favorable cases, this uncertainty is small, but it is always present. This is true both for deterministic and random variables. While deterministic variables take infinitely precise values that practically cannot be determined, random variables take intrinsically random values, regardless of the measurement. Thus, the "true" values of measurands are unknowable and observations are only interpretable when additionally to the measured values also their uncertainties are expressed. A measured result, therefore, should *always* include two entities: the measured value and some indication of its uncertainty [1], [2].

Almost all numerical data is affected by uncertainty because it comes either from measurements or from numerical computations that are based on empirical data. Error estimation and analysis of error propagation, therefore, is ubiquitous, and drawing conclusions from uncertain data is the normal case, not an exception. The most common way to represent, analyze, and deal with uncertainty is to employ methods from probability theory and statistics, see, e.g., [3], [4], [5].

The majority of tables and 2D-graphs in publications in science and engineering express uncertainty or provide error estimates. Evidently uncertainty is an important part of the information that has to be represented in order to avoid erroneous interpretation of the data. However, the majority of 2D and 3D visualizations and feature extraction methods still ignore errors and uncertainties of data, as well as, their propagation through the different stages of data analysis [6].

One major area of visual data analysis is the visualization of scalar fields, either by direct display or by extraction and depiction of topological or geometrical features. We are interested *how uncertainty affects such features*. The most considered features of scalar fields are *isocontours* (we use this term to denote isolines, isosurfaces and higher dimensional counterparts), or, more general, level sets. We focus in this paper on isocontours, although the developed concepts can be applied to many other features in scalar, vector, and tensor fields.

The uncertainty related to position and shape of isocontours has been discussed previously [7], [8], [9], [6]. However, none of these approaches were interpretable in terms of probability theory/statistics or fuzzy theory. In this paper, we establish a mathematical basis for modeling uncertainty of scalar fields and then address the computation and depiction of uncertain equivalents to isocontours. We aim at a formulation based on probability theory, since this leads to quantifiable and easily interpretable results. As mathematical model for uncertain scalar fields we use random fields. To describe the propagation of errors from the input data to computed features, we introduce condition analysis to feature-based visualization. Using the numerical condition, we assess the sensitivity of the isocontour problem relative to perturbations of the scalar field. Furthermore, we show how average condition numbers can aid the selection of thresholds that correspond to robust isocontours.

Methods to investigate level sets in parameter-continuous random fields have been developed in probability theory. However, for the intended applications these are not

---

- *The authors are with the Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Berlin, Germany. E-mail: {poethkow, hege}@zib.de.*

appropriate. Instead, we work on the level of probability density functions (PDFs) and interpolate these between sample points. We define the *isocontour density* (ICD) and the *level crossing probability field* (LCP) that quantify positional uncertainty of isocontours. Their results depend on both the uncertainty inherent in the scalar field and the nature of isocontour extraction. Both functions match our intuitive conception of fuzzy isocontours in $\mathbb{R}^d$. We employ these definitions in interactive visualization methods for uncertain isolines and isosurfaces of 2D and 3D scalar fields, respectively. In this paper, we have purposefully chosen a fairly general mathematical formulation since concepts are introduced that can be applied to a much wider range of feature extraction problems than just the isocontour problem.

In the next section, we review related work from data visualization literature and beyond. A probabilistic formulation of uncertain input data is given in Section 3. In Section 4 we introduce condition analysis to feature extraction, apply it to the isocontour problem and discuss average condition numbers. In Section 5 we establish mathematical models for uncertain isocontours. In Section 6 we describe the visualization methods that are applied to 2D and 3D. Example data sets from engineering, medicine and climate research are analyzed in Section 7. Our results are discussed and compared to previous work in Section 8.

2 RELATED WORK

The representation of uncertainty in 2D and 3D visualizations has been identified to be a major challenge in data visualization [6]. A variety of reasons for the occurrence of uncertainty, as well as, data representations and several methods for uncertainty visualization were presented in [10] and [11]. The uncertainty in physical measurements and common procedures of uncertainty evaluation are discussed, e.g., in [12]. Emphasis lies on statistical methods for data analysis and related concepts from probability theory. Alternative representations of uncertain data are multivalue or ensemble data sets. Their manipulation by special operators and possible visualization methods are described by Luo et al. [13]. A second alternative is the representation by *fuzzy sets* that were introduced by Zadeh [14] as an extension of the classical concept of crisp sets. Membership functions are used to describe degrees of membership of elements of fuzzy sets. Intervals can be considered as specific fuzzy sets with full degree of membership. Data in fuzzy representation can be processed using interval arithmetic. Fuzzy sets are used in many disciplines to model various kinds of uncertainty ranging from classifications to uncertain measurements [15].

Uncertainty visualization techniques for several different data types and applications have been introduced. In direct volume rendering uncertainty can be depicted by texture or color [16]. In vector and flow visualization uncertainty was represented by special glyphs, color mapping, and textures [17], [18], [19]. Otto et al. [20] generalized the concept of vector field topology to uncertain vector fields and described a method to obtain uncertain topological segmentations. Li et al. [21] used ellipsoid models to represent positional uncertainty and envelope models to expose

trajectory uncertainty of astronomical objects. Kniss et al. [22] developed a volume rendering approach that uses probabilistic classification. It allows the user to investigate the uncertainty of segmentations and defer the classification decision to the rendering stage. Uncertainty of tissue classification in medical volume data was also visualized by animation using fuzzy time-dependent transfer functions for direct volume rendering [23].

The uncertainty of the position, size, and shape of surfaces has been addressed in multiple publications. Pang et al. [10] presented *fat surfaces* that use two surfaces to enclose the volume in which the true (but unknown) surface lies. Reconstructions of surfaces from point cloud data have a degree of uncertainty at a given point that depends on the position of that point relative to the sample points in the neighborhood. Pauly et al. [24] presented a formulation of likelihood and confidence maps which allows to quantify and visualize that uncertainty.

Grigoryan and Rheingans [7] used point primitives for rendering uncertain surfaces: A large number of points were randomly displaced along the isosurface normals whose maximum distance varied according to the uncertainty. Isosurfaces in uncertain data were also proposed by Johnson and Sanderson [6] where a combined volume and surface rendering was applied to display surface uncertainty. Rhodes et al. [9] used color and texture mapping on isosurfaces to indicate areas of high data uncertainty. The uncertainties of the surface's position and shape are not visualized by this approach. Note that there was no specification of the mathematical model describing the uncertainty in [6] and [9]. Kindlmann et al. [8] used the magnitude of the flowline curvature as a measure for isosurface uncertainty and mapped this to surface color. Djurcilov et al. [25] presented contour lines that were stippled in areas of high uncertainty and continuously drawn elsewhere. Zehner et al. [26] augment the isosurfaces extracted from geological data with additional geometry indicating the spatial area being crossed by the isosurface with a certain confidence.

MacEachren et al. [27] review several approaches for representing uncertainty related to geospatial data and for improving decision making with uncertainty. They also present seven key research challenges in visualizing uncertainty. Torre Zuk [28] presented a theoretical framework to aid the development and qualitative evaluation of visualizations that support reasoning under uncertainty.

The term "uncertainty" is used with different meanings and for addressing different problems, even in the narrow field of data visualization. For instance, Jänicke et al. [29] and Wang et al. [30] attribute some "average uncertainty" or "local statistical complexity" to spatial or spatiotemporal domains to characterize *spatial variance* of data values, and thereby to identify significant parts of data sets. In this paper, however, statistical parameters are attributed to each voxel to express the uncertainty of data values due to *measurement errors* and other sources of uncertainty.

3 UNCERTAIN DATA

All data based on a measurement that is not just simple counting is uncertain. This section gives a brief overview of different manifestations and causes of errors or uncertainties

in science and engineering, and their mathematical modeling. New contributions to data visualization are the modeling with random fields and interpolated probability distributions. For general references in probability theory and statistics, see, e.g., [3], for random fields, e.g., [31], and for uncertainty in measurements, e.g., [4].

3.1 Uncertainty in Physical Measurements

In general, data uncertainty consists of several parts. Systematic and random errors occur in all measurements. Rounding and discretization lead to additional uncertainty. It is important to distinguish between the terms *error* and *uncertainty*. In a measurement where significant random errors *can* occur, an observed value may, by chance, be very close to the true value. In this case the error is low, while the uncertainty (assessed by repeated measurements) is high [1].

3.1.1 Systematic Errors

An error that always occurs in the same way and extent if a measurement of some quantity is repeated, is called *systematic error*. It can be additive (constant shift) or a multiplicative (deviation of constant percentage). Reasons for systematic errors can be calibration deficits, environmental conditions, or too simple models of the measured quantity. As the errors are constant over repeated measurements it is not possible to detect and eliminate them with only one measurement procedure. Sometimes it is possible to compare results of multiple measuring devices or methods, and thereby to minimize the systematic error. In general, it is assumed that most values have an unknown systematic error that is impossible to eliminate completely.

3.1.2 Random Errors

At each repetition of a measurement *random errors* affect the results differently and randomly. Multiple observations can be statistically analyzed and the quantity can be described by statistical parameters. If the quantity has a deterministic value, i.e., is not a result of a stochastic process, several reasons can lead to random fluctuations; this includes interference of the environment with the measurement process, like background noise that occurs when, e.g., temperature, humidity or vibrations influence the measurement.

3.1.3 Other Reasons for Uncertainty

Other sources of uncertainty lie in the measurement devices and the computers that process the data. The precision of a result is always limited, for instance by the number of bits of floating point variables. Quantization or discretization of continuous phenomena to sample point or grid based representations lead to uncertainty as well [10].

3.2 Mathematical Model for Uncertain Spatial Data

We do not distinguish between "raw" measured data and data computed from measured data: conceptually we consider the possibly complex computation as part of the measurement process. In the following, we assume occurrence of *additive* errors only, i.e., we assume that an observed value can be written as:

$$\text{observed value} = \text{true value} + \text{systematic error}$$
$$+ \text{ random error.}$$

Thus, we do not consider *multiplicative* errors. Furthermore, as systematic errors have to be dealt with in a highly application-specific way, we assume that systematic errors have been minimized and are negligible. Hence, we consider the simplified case

$$\text{observed value} = \text{true value} + \text{random error.}$$

3.2.1 Discretely Sampled Uncertain Spatial Data

The true values at all positions $\mathbf{x} \in I^*$ of the domain $I^* \subseteq \mathbb{R}^d$ are assumed to be described by a continuous unknown function $h : I^* \to \mathbb{R}, \mathbf{x} \mapsto h(\mathbf{x})$. The variable $\mathbf{x}$ may embody space, time and other parameters. In practice the function $h(\mathbf{x})$ can only be measured at a finite set of n points $\mathbf{x}_i \in I$, $i \in \{1, 2, \ldots, n\}$, where $I \subset I^*$. We assume that the points form a regular grid and are close enough to exhibit a sampling rate that exceeds the Nyquist rate.

To model the uncertainty, the observational errors at all positions $\mathbf{x}_i$ are regarded as random variables $Z_i := Z(\mathbf{x}_i)$. These random variables form a *parameter-discrete random field* [31] written as $\{Y_{\mathbf{x}_i} : \mathbf{x}_i \in I\}$ with

$$Y_{\mathbf{x}_i} = h(\mathbf{x}_i) + Z_i. \tag{1}$$

The lth observation is regarded as the lth realization of the random field

$$Y_{\mathbf{x}_i}^{(l)} = h(\mathbf{x}_i) + Z_i^{(l)}. \tag{2}$$

We assume that the sample means $\overline{Y_{\mathbf{x}_i}^{(\cdot)}}$ converge to the true values $h(\mathbf{x}_i)$ in a series of many observations and thus,

$$\mathrm{E}(Z_i) = 0 \quad \text{for all } i \in \{1, 2, \ldots, n\}. \tag{3}$$

All random variables Z_i with $i \in \{1, 2, \ldots, n\}$ are assumed to be described by PDFs $\varphi_i(t)$ such that

$$\mathrm{E}(Z_i) = \int_{-\infty}^{\infty} \varphi_i(t)\, t\, \mathrm{d}t = 0 \tag{4}$$

holds. From this it follows that

$$\mathrm{E}(Y_{\mathbf{x}_i}) = h(\mathbf{x}_i) \tag{5}$$

and the random variables $Y_{\mathbf{x}_i}$ have the PDFs $f_i(t) := \varphi_i(t - h(\mathbf{x}_i))$. The cumulative distribution function(CDF)

$$F_i(a) = \int_{-\infty}^{a} f_i(t)\, dt, \tag{6}$$

provides the *probability* that a realization of $Y_{\mathbf{x}_i}$ is less than or equal to a. The spread of the random values at a point $\mathbf{x}_i$ can be characterized by the standard deviation

$$\sigma_i = \sqrt{\mathrm{E}(Y_{\mathbf{x}_i} - \mathrm{E}(Y_{\mathbf{x}_i}))^2}$$

or its square σ_i^2, the variance.

Possible distributions f_i are, e.g., uniform or normal distributions. The normal (or Gaussian) distribution plays a fundamental role in applications, because it represents the distribution of random variables in many natural phenomena, as well as, the distribution of measured values of deterministic quantities. A theoretical explanation of this fact is provided by the central limit theorem, see, e.g., [3].

The distributions f_i can be acquired in various ways depending on the input data and the quantities one is

interested in. The unknown expected value $E(Y_{\mathbf{x}_i})$ can be estimated by the *arithmetic mean*; the standard deviation σ_i by the *sample standard deviation* s_i. Sometimes, for example if the distributions are unknown or nonparametric, more advanced statistical methods such as bootstrapping, jackknifing, or other resampling techniques are needed to analyze the data [3], [32].

The random field $\{Y_{\mathbf{x}_i} : \mathbf{x}_i \in I\}$ is completely characterized if the covariances $C(\mathbf{x}_i, \mathbf{x}_j) = E((Y_{\mathbf{x}_i} - \mu_i)(Y_{\mathbf{x}_j} - \mu_j))$ with $\mu_k = E(Y_{\mathbf{x}_k})$ are known, too. The covariance often depends only on the distance $\gamma = \|\mathbf{x}_i - \mathbf{x}_j\|$, i.e., mappings $\gamma \mapsto \bar{C}(\gamma)$ are considered.

3.3 Uncertain Data in a Continuous Spatial Domain

Many visualization techniques and feature extraction methods require functions as input that are defined in a continuous domain. In order to apply standard visualization techniques, one needs to specify for uncertain data what happens in regions between sample points. One possibility is to extend the parameter-discrete random field $\{Y_{\mathbf{x}_i} : \mathbf{x}_i \in I\}$ to a *parameter-continuous* random field $\{Y_{\mathbf{x}} : \mathbf{x} \in \mathbb{R}^d\}$ with properties like, e.g., covariances between arbitrary locations that smoothly interpolate the covariances between discrete locations.

3.3.1 Level Crossings in Continuous Random Fields

The analysis of level sets in parameter-continuous random fields is an area of active research in mathematics, see, e.g., [33]. This research is triggered by applications in natural sciences, in which randomness is required to describe certain phenomena, see, e.g., [31]. Depending on its covariance, a random field might be rather nonsmooth. Its smoothness is directly determined by the differentiability of the covariance $\bar{C}$ at distance 0.

Given a random field and a level set $\Omega \subset \mathbb{R}^d$ that corresponds to some threshold θ. Then, the following quantities are of interest: for random fields from $\mathbb{R}^d$ to $\mathbb{R}^d$, the number of level crossings N_B in some subset $B \subset \mathbb{R}^d$, and for random fields from $\mathbb{R}^d$ to $\mathbb{R}^{d'}$ with $d > d'$, the geometric measure (length, area, volume, ...) A_B of the intersection $\Omega \cap B$. Then N_B and A_B itself are random variables, but currently there is no known way to compute their distribution for nontrivial situations. A tool to understand the distributions are Rice formulas; they allow to express the expectation values and higher moments of N_B and A_B as integrals over a function that depends on the joint distribution of the random field and its derivative [34].

For Gaussian random fields it can be shown 1) that the level sets are continuous, but in general are nondifferentiable, and 2) that their Hausdorff dimension is larger than $d - 1$ [35]. Obviously, these objects do not represent what we are interested in: Instead of approximating level sets of the underlying field, here level sets are considered whose properties are largely determined by covariances of measurement errors. A more adequate mathematical model is described in the next section.

3.3.2 Spatial Interpolation of PDFs

As described earlier, we consider as input uncertain data represented by a parameter-discrete random field $\{Y_{\mathbf{x}_i} : \mathbf{x}_i \in I\}$ with parameter set $I \subset \mathbb{R}^d$, state space $S = \mathbb{R}$, and random variables Z_i that can be described by PDFs $f_i(t)$. In the following, we additionally assume that the probability
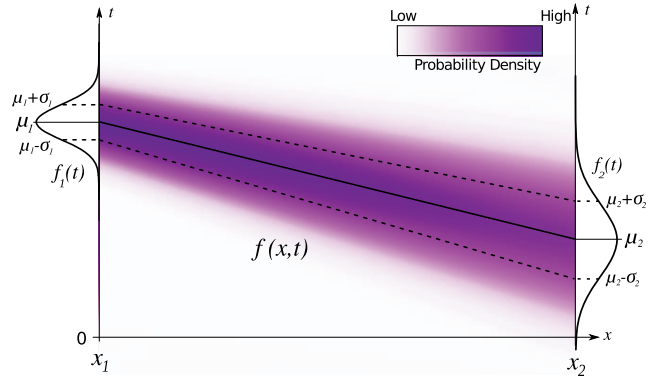

Fig. 1. Linear interpolation between two normal distributions given at two sample points x_1 and x_2.

distributions for all $Y_{\mathbf{x}_i}$ are of the same type, e.g., are *all* uniform or are *all* normal. We further assume that the PDFs can be functionally represented by their expected values $\mu_i = E(Y_{\mathbf{x}_i})$ and a finite number of $\hat{m} - 1$ central moments $E((Y_{\mathbf{x}_i} - \mu_i)^m)$, where $m \in \{2, \ldots, \hat{m}\}$.

To build a continuous extension we consider a continuous parameter set $I^* \subseteq \mathbb{R}^d$ that contains all sample points $\mathbf{x}_i$. Given that the PDFs at the sample points are all of the same type, the most natural assumption is that the PDFs between the sample points, and thus in all $\mathbf{x}$ of I^*, are also of that type. Therefore, we extend the discrete model defined at the sample points $\mathbf{x}_i$, $i \in \{1, 2, \ldots, n\}$ to a continuous model in the whole domain I^* by 1) *interpolating* the expected values μ_i and the m-th roots of the central moments $\nu_{m,i} = E((X - \mu)^m)^{\frac{1}{m}}$ in parameter space and 2) inserting these interpolated values in the PDFs. This is a rather general method; it can be applied to all kinds of PDFs that can be parametrized by their moments. (Note that simple blending of PDFs between grid points would in general not preserve the type of the distribution; for instance blending two normal distributions would yield a bimodal distribution.)

For the functions $\mu(\mathbf{x})$ and $\nu_m(\mathbf{x})$ with $\mathbf{x} \in I^*$ all sorts of interpolants can be used. For simplicity we use linear tensor product interpolation in this paper.

As an example, assume that the random variables Y_i are normally distributed $f_i(t) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp(-\frac{1}{2}(\frac{t - \mu_i}{\sigma_i})^2)$ with $t \in S$. The interpolated PDF is then obtained by substituting the expected values μ_i and standard deviations σ_i by interpolants $\mu(\mathbf{x})$ and $\sigma(\mathbf{x}) \equiv \nu_2(\mathbf{x})$:

$$f(\mathbf{x}, t) = \frac{1}{\sigma(\mathbf{x})\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right)^2\right). \qquad (7)$$

Fig. 1 shows an example in one dimension with linear interpolation between two sample points x_1 and x_2.

4 ISOCONTOURS AND THEIR NUMERICAL CONDITION

4.1 Basics

Before we investigate uncertainty propagation of isocontour extraction, crisp (or "certain") level sets and isocontours are revisited. A *scalar field* is a function $y : \mathbb{R}^d \to \mathbb{R}$ with $\mathbf{x} \mapsto y(\mathbf{x})$ and its *gradient* is the vector field $\nabla y(\mathbf{x}) = (\frac{\partial y(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial y(\mathbf{x})}{\partial x_n})$, where $x_1 \ldots x_n$ are the Cartesian coordinates.

The points $\mathbf{x}_c \in \mathbb{R}^d$, where $\|\nabla y(\mathbf{x}_c)\| = 0$ are called *critical points* of f. A *level set* of f for constant level $\theta \in \mathbb{R}^d$ is defined as

$$\Omega = \{\mathbf{x}_s \in \mathbb{R}^d : y(\mathbf{x}_s) = \theta\} \tag{8}$$

or $\Omega = y^{-1}(\theta)$.

In computer graphics and visualization, *isocontours* and *isosurfaces* are extracted from scalar fields using, e.g., marching squares or marching cubes [36] algorithms. These methods only find $(d-1)$-dimensional intersections of y with a given isovalue θ. The resulting isocontours do not necessarily contain all points of the true level set. In particular, they do not contain the complete set of critical points $\mathbf{x}_c$ with $y(\mathbf{x}_c) = \theta$ if these points form a d-dimensional plateau.

4.2 Numerical Condition

Let us consider a problem where, we have to compute some quantity ("feature") $\rho \in \mathbb{R}^m$ from input $\alpha \in \mathbb{R}^n$. Since α is not exactly known, we should instead consider an input set D that contains all perturbed inputs $\tilde{\alpha}$, i.e., instead of a pointwise mapping $\alpha \mapsto \rho(\alpha)$ a set-valued mapping $\rho : D \to E = \rho(D)$. The effect of perturbations of input data on the output quantities—also called the *condition* of a problem (ρ, α)—can be expressed by some measure of a ratio of output versus input sets [37].

A *condition number* describes the sensitivity of a solution for a given problem to perturbations of the input data, *independently* of the algorithm and the character of the perturbations. Let $\| \cdot \|$ be norms in $\mathbb{R}^m$ and $\mathbb{R}^n$. If a perturbation ε distorts the input α to $\alpha + \varepsilon$ then the *absolute normwise condition* of a problem (ρ, α) is defined (in a first order approximation) as the smallest number $\kappa_{abs} \geq 0$ with the property that there is a real number $\delta > 0$ so that for all $0 < \|\varepsilon\| < \delta$ the inequality

$$\|\rho(\alpha) - \rho(\alpha + \varepsilon)\| \leq \kappa_{abs} \|\varepsilon\| \tag{9}$$

holds. A problem is said to be well-conditioned if κ_{abs} is low and it is said to be ill-conditioned if κ_{abs} is high. The exact meaning of low and high depends on the problem at hand.

If we assume that $\rho(\alpha)$ is (totally) differentiable, because of the mean value theorem the condition number can be calculated through the derivative:

$$\kappa_{abs} = \|\nabla_\alpha \rho(\alpha)\|, \tag{10}$$

where ∇_α is the gradient in parameter space $\mathbb{R}^n$.

The condition number is a dimensionful quantity. Its unit of measurement is [unit of output data] per [unit of input data]. The quantity relates the errors of (in general) different dimensions; loosely speaking, κ_{abs} can be seen as the amplification factor of input errors. Its absolute size, therefore, depends on the units used. Condition numbers in different data sets, therefore, can be compared only if the same units of measurements are used in each data set.

4.3 Well and Ill-Conditioned Isocontours

4.3.1 Numerical Condition of the Isocontour Problem

To determine, which parts of an isocontour are the result of well or ill-conditioned computation, we determine the absolute normwise condition for the isocontour problem. By means of the condition number, we can assess where potential errors and uncertainty are attenuated or amplified even if *no* information about the uncertainty of the data is available. Additionally the isovalues leading to contours that, on average, are well or ill-conditioned are of interest.
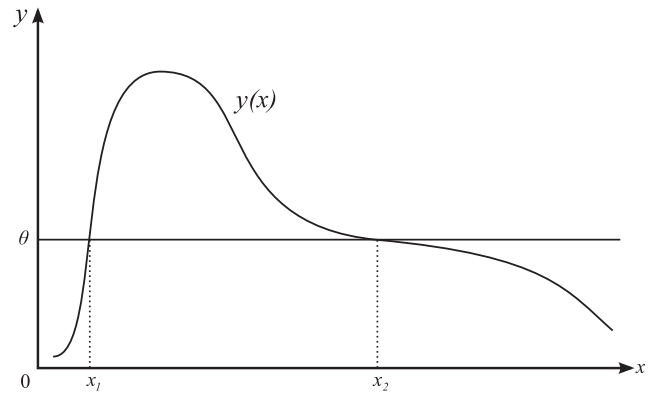

Fig. 2. Calculation of the intersection point x_1 of $y(x)$ with threshold θ is well-conditioned, while calculation of x_2 is ill-conditioned, i.e., the position of x_1 is less sensitive to perturbations of y than the position of x_2.

This information can be used to aid the selection of thresholds for isocontour extraction or segmentation.

To extract isocontours from a scalar field, we have to solve $y(\mathbf{x}) = \theta$ to find the level crossing points. If the gradient $\nabla y(\mathbf{x})$ is invertible, then, according to the inverse function theorem, we can write $y^{-1}(\theta) = \mathbf{x}$. The derivative satisfies

$$\|(y^{-1})'(\theta)\| = \|\nabla y(\mathbf{x})\|^{-1}. \tag{11}$$

Thus, the absolute normwise condition of the problem (y^{-1}, θ) is

$$\kappa_{abs}(\mathbf{x}) = \|\nabla y(\mathbf{x})\|^{-1}. \tag{12}$$

Fig. 2 shows an example of a 1D function. The calculation of x_1 is well-conditioned while the calculation of x_2 is ill-conditioned. In other words, the position of the level crossing at x_1 is less prone to perturbations of y than the position of x_2. With (12) we can see that in case of a plateau in y, where the norm of the gradient is zero, the extraction of an isocontour is an ill-posed problem. The condition number κ_{abs} provides an estimate for the propagation of input error to positional error of the isocontour.

4.3.2 Average Condition Numbers

To investigate, which of the isocontours are well- or ill-conditioned we compute average condition numbers with respect to the possible isovalues. For that we define the total condition of the isocontour $y^{-1}(\theta)$

$$\hat{\kappa}_{abs}(\theta) = \int_{y^{-1}(\theta)} \kappa_{abs}(\mathbf{x}) \, d\omega$$

and use the area (or length)

$$a(\theta) = \int_{y^{-1}(\theta)} 1 \, d\omega$$

of that isocontour to calculate the average condition number. We are interested in the condition of isocontours regardless of their size. So we choose to weight the average condition by the average area because $\hat{\kappa}_{abs}(\theta)$ scales with the size of the isocontour. For the isosurfaces in the interval $[\theta_1, \theta_2] \subset \mathbb{R}^d$ the *average condition number* is

$$\bar{\kappa}_{abs}([\theta_1, \theta_2]) = \frac{\int_{\theta_1}^{\theta_2} \hat{\kappa}_{abs}(t) \, dt}{\int_{\theta_1}^{\theta_2} a(t) \, dt}. \tag{13}$$
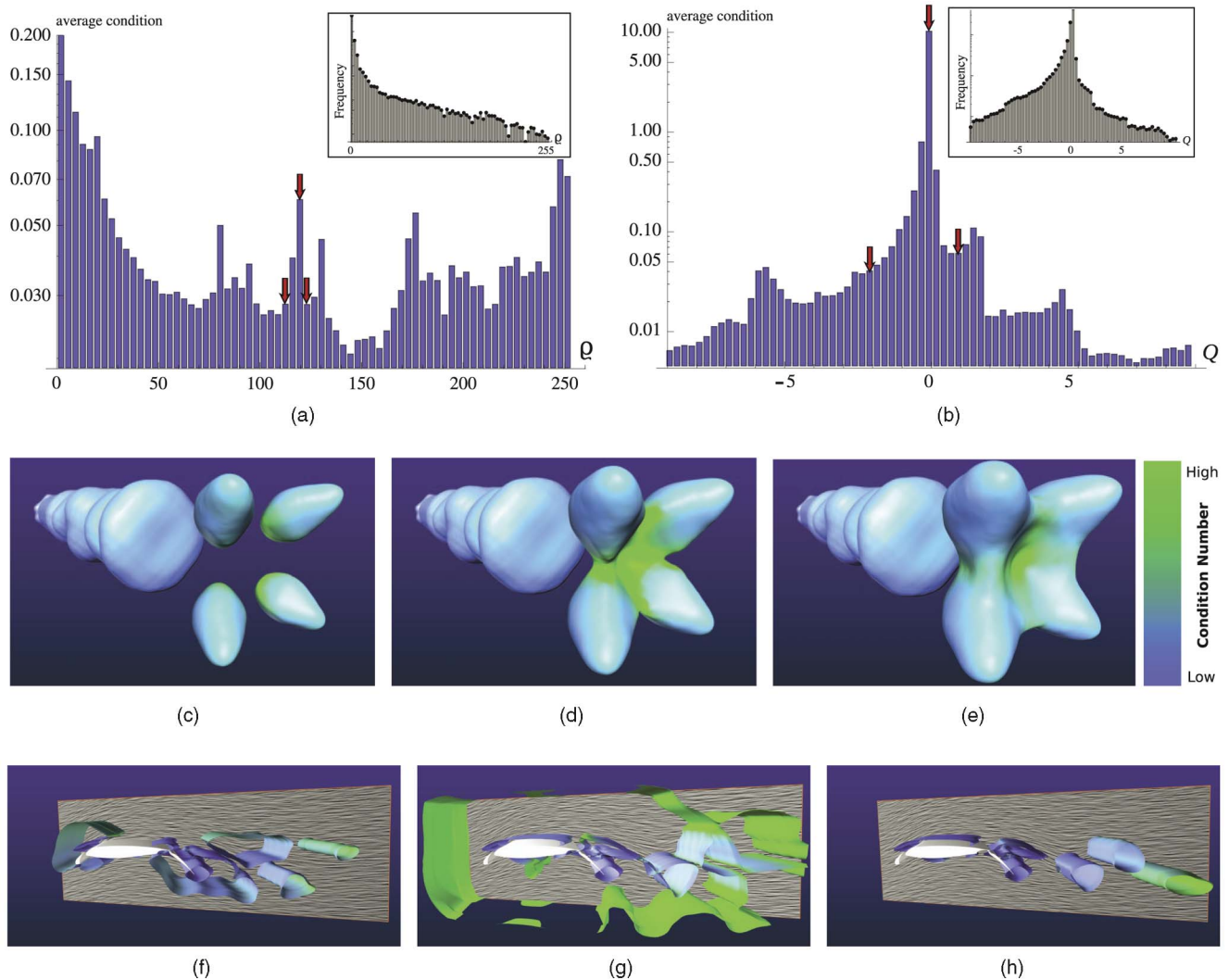
Fig. 3. The average numerical condition is shown for isosurfaces in the fuel data set (a) and the Q field of a flow data set (b). For comparison the histograms are included. (c)-(e) show isosurfaces for $\theta = 113, \theta = 119$, and $\theta = 125$ in the fuel data set, (f)-(h) show isosurfaces for $\theta = -2, \theta = 0$, and $\theta = 1$ in the Q field. The isosurfaces in (d) and (g) have relatively high average condition numbers and should be considered less reliable than the other surfaces.

Federer's Coarea formula [38], cf. also [39]

$$\int_{\mathbb{R}^d} \int_{y^{-1}(t) \cap I^*} q(\mathbf{x}) \, \mathrm{d}\,\omega \, \mathrm{d}t = \int_{I^*} q(\mathbf{x}) \, \|\nabla y(\mathbf{x})\| \, \mathrm{d}V, \qquad (14)$$

where $q : I^* \to \mathbb{R}^d$ is any function defined on the same domain I^* as y, allows us to express integrals over level sets as integrals over the domain.

Restricting the thresholds to the interval $[\theta_1, \theta_2]$, we can rewrite (13) using integrals over

$$V_t = \{\mathbf{x} \in \mathbf{R^d} \mid \theta_1 \le \mathbf{y}(\mathbf{x}) \le \theta_2\} :$$

$$\bar{\kappa}_{abs}([\theta_1, \theta_2]) = \frac{\int_{V_t} 1 \, \mathrm{d}V}{\int_{V_t} \|\nabla y(\mathbf{x})\| \, \mathrm{d}\mathbf{V}}. \qquad (15)$$

This formulation is numerically more convenient than integrals over surfaces, since calculating the sum over potentially diverging values of $\kappa_{abs}(\mathbf{x})$ can lead to overflows or accuracy issues. It also saves us from having to generate many isosurfaces. Instead, we approximate the result of (15)

by computing sums over the discretized scalar field and its gradient magnitude field.

4.3.3 Examples

Fig. 3 shows plots with average condition numbers $\bar{\kappa}_{abs}$ for small ranges of possible isovalues and isosurfaces with the condition mapped to surface color for two scalar fields given on uniform grids. Arrows indicate the used isovalues in the plots. On (spatial) average, the isocontours for isovalues at the minima of $\bar{\kappa}_{abs}$ have the best numerical condition while those at the maxima have the worst average numerical condition in each data set. The selection of thresholds is often based on histograms so, for comparison, we also depict the histograms in the respective subfigures.

Our first example is the fuel data set (freely available at http://www.volvis.org), describing the density ϱ of fuel during an injection process. The plot showing average condition numbers in Fig. 3a was computed using 70 disjoint equally sized intervals on the range $(0, 255)$. It indicates that isovalues close to zero yield diverging condition numbers. Additionally there are several less significant peaks. These peaks are not present in the

histogram which means that the average condition plot is better suited for the selection of reliable thresholds. In Figs. 3c, 3d, and 3e isosurfaces for $\theta = 113, \theta = 119$ and $\theta = 125$ in the fuel data set are shown. The isosurface close to a saddle point Fig. 3d is relatively ill-conditioned.

Fig. 3b represents average condition numbers of isosurfaces in the Q field (Okubo-Weiss parameter [40]) of a smoothed vector field. We used a single timestep from a simulation of flow around an airfoil. Again the average condition numbers were computed using 70 disjoint equally sized intervals on the range $(-9, 9)$. The plot exhibits several peaks, some of which are not visible in the histogram. In Figs. 3f and 3h isosurfaces for $\theta = -2, \theta = 1$ in the Q field enclose areas of high strain and high vorticity, respectively. Isosurfaces for $\theta \approx 0$ (similar to Fig. 3g) have previously been used to show the separate regions of dominant strain and vorticity [41], [42]. However, the isovalue $\theta = 0$ is a possibly problematic choice because it yields the most ill-conditioned isosurface in the whole data set because parts of the level set lie in areas with very low (or zero) gradient magnitude.

Generally speaking, all isosurfaces close to critical points, plateaus in particular, or other areas of low gradient magnitude ("near-plateaus") are on average relatively ill-conditioned.

These results can be related to those of Bajaj et al. [43] and Pekar et al. [44] who identified the significant isosurfaces, specifically those that separate homogenous regions in a data set. They show that a high average (or total) gradient magnitude is a good criterion for isosurfaces to be considered to be the most "meaningful" ones in a data set. Empirically, we found out that isovalues corresponding to surfaces with maximal average gradient magnitude also correspond to those with the best average numerical condition.

5 UNCERTAIN ISOCONTOURS

We are interested in equivalents to isocontours in uncertain data. Obviously the position and shape of an isocontour is not precisely defined in this case. It is our aim to quantify the positional uncertainty of isocontours with respect to a given isovalue from the probability distributions of the input data.

Because of the practical relevance of the Gaussian distribution all examples and explicit formulas in the following are provided for this case. However, the formulas are valid for arbitrary probability distributions, unless stated differently, and in many cases also explicit formulas can be easily derived.

5.1 Isocontour Density

We consider isocontours in $I^* \subset \mathbb{R}^d$. Using an interpolated PDF $f(\mathbf{x}, \mathbf{t})$ we compute a spatial density of an isocontour by simply evaluating this function at a given *isovalue* $\theta \in \mathbb{R}^d$ for all points $\mathbf{x} \in \mathbf{I}^*$. For that we introduce a function

$$g_\theta(\mathbf{x}) := f(\mathbf{x}, t = \theta), \qquad (16)$$

which we call ICD. It provides the probability density with respect to θ at position $\mathbf{x}$ which is a measure for the spatial distribution of the uncertain isocontour. The quantity $g_\theta(\mathbf{x})\, \mathrm{d}t$ is the probability that the true but unknown function takes a value in the interval $[\theta, \theta + \mathrm{d}t]$ at position $\mathbf{x}$. Note that f is a normalized PDF for a specific $\mathbf{x}$ with respect to the *state space* S (i.e., in t-direction). Thus, the values of $g_\theta(\mathbf{x})$ are probability densities with respect to S, not I^*.
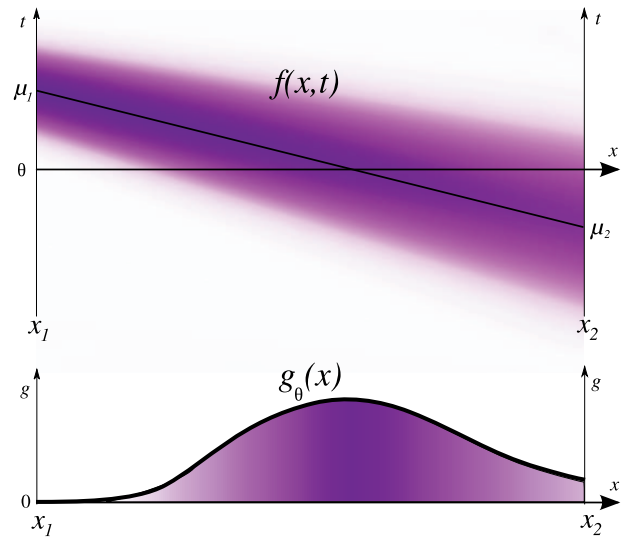

Fig. 4. Example graph for the ICD in a 1D grid cell. The values $g_\theta(x)$ are computed from the interpolated PDF $f(x, t)$ with respect to threshold θ.

For normally distributed data and given interpolants $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ the ICD follows directly from (7):

$$g_\theta(\mathbf{x}) = \frac{1}{\sigma(\mathbf{x})\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\theta - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right)^2\right). \qquad (17)$$

An example graph of $g_\theta(\mathbf{x})$ for the 1D case between two sample points x_1, x_2 is given in Fig. 4.

Due to the continuous interpolation the values of $g_\theta(\mathbf{x})$ at the grid points only depend on the probability distributions that are given there: $g_\theta(\mathbf{x}_i) = f_i(\theta)$. That means $g_\theta(\mathbf{x})$ is a *continuous* function. As $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are piecewise linear functions, $g_\theta(\mathbf{x})$ is nondifferentiable at the grid boundaries and differentiable within the grid cells.

For the limit $\sigma(\mathbf{x}) \to 0$ we expect $g_\theta(\mathbf{x})$ to return a *crisp* isocontour. If we consider $\lim_{\sigma(\mathbf{x}) \to 0} g_\theta(\mathbf{x})$ and use the following definition of the Dirac delta distribution for $x \in \mathbb{R}^d$ [45]:

$$\delta(x) = \lim_{m \to 0} \frac{1}{m\sqrt{\pi}} \exp\left(-\left(\frac{x}{m}\right)^2\right), \qquad (18)$$

with support $\{0\}$, then

$$\lim_{\sigma(\mathbf{x}) \to 0} g_\theta(\mathbf{x}) = \delta(\theta - \mu(\mathbf{x})). \qquad (19)$$

So, if $\sigma(\mathbf{x})$ vanishes (i.e., no uncertainty is considered) then the support of $g_\theta(\mathbf{x})$ is identical to the set of θ-level-crossings of $\mu(\mathbf{x})$, i.e., the level set

$$\Omega = \{\mathbf{x}_s \in I^* : \mu(\mathbf{x}_s) = \theta\}. \qquad (20)$$

For a symmetric distribution like the Gaussian distribution, at a fixed spatial position $x_c \in I^* \subseteq \mathbb{R}^d$ (1D case) the interpolated PDF $f(x_c, t)$ takes its maximum at $t = \mu(x_c)$. From this it follows that, if $\sigma(x)$ is constant, $g_\theta(x)$ takes its *local maxima* at the points $x_s \in \Omega$.

The function $g_\theta(x)$ can be related to the numerical condition of isocontour extraction. Recall that (9) describes the relation of perturbations of the input data to the perturbation of the result. How does $g_\theta(x)$ propagate the

uncertainty represented by the standard deviation of the input data?

Assuming that an interval $[x_j, x_k] \subset I^*$ contains a level-crossing $\mu(x) = \theta$ and that $\sigma(x)$ is constant ($\sigma_j = \sigma_k$), the *inflection points* x_a, x_b of $g_\theta(x)$ can be calculated using elementary calculus. If $[x_a, x_b] \subseteq [x_j, x_k]$, we can consider the distance between x_a and x_b to be a measure of the spread of $g_\theta(x)$. With regard to the fact that the inflection points of a normal distribution are located one standard deviation from the mean we define the spread of $g_\theta(x)$ as

$$s_{g_\theta} = \frac{1}{2} |x_a - x_b|. \tag{21}$$

A simple calculation shows that

$$\frac{1}{2} |x_a - x_b| = \sigma_{j,k} \frac{1}{|\mu_j - \mu_k|}. \tag{22}$$

As we assume linear interpolation, $|\mu_j - \mu_k|$ is the derivative of $\mu(x)$ with respect to x in the interval $[x_j, x_k]$. From this it follows that

$$s_{g_\theta} = \kappa_{abs} \, \sigma_{j,k}, \tag{23}$$

where κ_{abs} is the numerical condition for the calculation of level-crossings of $\mu(x)$ (cf. 12). This shows that the ICD propagates the uncertainty of the input data proportionally to the condition number.

5.2   Level-Crossing Probability Field

A different approach to calculate spatial distributions of isocontours is described in this section. Again, we consider a parameter-discrete random field $\{Y_{\mathbf{x}i} : \mathbf{x}_i \in I\}$ as a model for the uncertain input data (described in Section 3.2.1). For simplicity we consider a 1D parameter set first.

Let $\mathbf{x}_j, \mathbf{x}_k \in I$, be adjacent sample points with associated random variables Y_j, Y_k and PDFs $f_j(t), f_k(t)$. If we assume monotonic interpolants between the realizations then we have at least one crossing of the constant level line $y = \theta$. The probability for a level-crossing along the line segment $[\mathbf{x}_j, \mathbf{x}_k]$ then is

$$\begin{aligned} prob_{[\mathbf{x}_j, \mathbf{x}_k]}(\theta \text{ crossed}) &= prob(Y_j \leq \theta) \, prob(Y_k \geq \theta) \\ &\quad + prob(Y_j \geq \theta) \, prob(Y_k \leq \theta). \end{aligned} \tag{24}$$

The probabilities on the right side of this equation can easily be calculated by the cumulative distribution functions $F_{(\cdot)}(\theta)$ that provide the probability that $Y_{(\cdot)}$ less or equal to θ. So, the level crossing probability can be expressed as

$$\begin{aligned} prob_{[\mathbf{x}_j, \mathbf{x}_k]}(\theta \text{ crossed}) &= F_j(\theta) \, (1 - F_k(\theta)) \\ &\quad + (1 - F_j(\theta)) \, F_k(\theta). \end{aligned} \tag{25}$$

With that it is possible to calculate a level-crossing probability for each grid cell. However, it is desirable to get a probability for each point in a continuous domain.

Using the interpolation scheme introduced in Section 3.3 it is possible to compute a CDF from interpolated PDF $f(\mathbf{x}, \mathbf{t})$ and thus, subdivide the grid cells in order to compute probabilities for subsegments, i.e., $prob_{[\mathbf{x}, \mathbf{x}+\Delta\mathbf{x}]}(\theta \text{ crossed})$. We are then able to find the limit

$$P_\theta(x) = \lim_{\Delta\mathbf{x}\to 0} prob_{[\mathbf{x}, \mathbf{x}+\Delta\mathbf{x}]}(\theta \text{ crossed}), \tag{26}$$

which evaluates to LCP

$$P_\theta(\mathbf{x}) = 2 \, F_{\mathbf{x}}(\theta) \, (1 - F_{\mathbf{x}}(\theta)), \tag{27}$$

where $F_{\mathbf{x}}(\theta) = \int_{-\infty}^{\theta} f(\mathbf{x}, t) \mathrm{d}t$ is the CDF of the interpolated PDF. Equation (27) gives the probability that for two independent realizations y_a and y_b of a random variable distributed according to the interpolated PDF $f(\mathbf{x}, t)$ one of them is greater or equal to θ while the other is less or equal to θ.

To show the range of function values we consider P_θ as a function of $F_{\mathbf{x}}$ in the interval $[0, 1]$. Obviously P_θ is not negative. Because $P_\theta = 2(F_{\mathbf{x}} - F_{\mathbf{x}}^2)$, we can write

$$\frac{\mathrm{d}P_\theta}{\mathrm{d}F_{\mathbf{x}}} = 2(1 - 2F_{\mathbf{x}}).$$

From $\mathrm{d}P_\theta/\mathrm{d}F_{\mathbf{x}} = 0$ and $F_{\mathbf{x}} \in [0, 1]$ it follows, that the function $P_\theta(\mathbf{x})$ takes a maximum at $F_{\mathbf{x}}^{(max)} = \frac{1}{2}$. Thus, the maximum value of P_θ is $P_\theta(\frac{1}{2}) = \frac{1}{2}$, and therefore, $P_\theta(\mathbf{x}) \in [0, \frac{1}{2}]$.

If $P_\theta(\mathbf{x})$ is maximal then

$$F_{\mathbf{x}}^{(max)}(\theta) = \frac{1}{2} = 1 - F_{\mathbf{x}}^{(max)}(\theta)$$

holds. This means that $prob(Y_{\mathbf{x}} \leq y_\theta) = prob(Y_{\mathbf{x}} \geq y_\theta) = \frac{1}{2} = prob(Y_{\mathbf{x}} \leq t_{\frac{1}{2}}) = prob(Y_{\mathbf{x}} \geq t_{\frac{1}{2}})$ if $t_{\frac{1}{2}}$ is the *median* of the PDF $f_{\mathbf{x}}(y)$. Thus, $P_\theta(\mathbf{x})$ takes its maximum if the isovalue θ is equal to the *median* of the interpolated PDF $f(\mathbf{x}, t)$.

For normally distributed input data LCP this evaluates to

$$P_\theta(\mathbf{x}) = \frac{1}{2} \left(1 - \mathrm{Erf}\left(\frac{\mu(\mathbf{x}) - \theta}{\sqrt{2} \, \sigma(\mathbf{x})} \right) \right), \tag{28}$$

where Erf is the *error function*.

5.3   Comparison

While the ICD maps points $\mathbf{x} \in I^*$ to probability densities, LCP maps points to probabilities. The cardinality of the range of $g_\theta(\mathbf{x})$ depends on $f(\mathbf{x}, \mathbf{t})$. The density values $g_\theta(\mathbf{x})$ are relatively low in case of large values of $\sigma(\mathbf{x})$ and high in case of small values of $\sigma(\mathbf{x})$ because f is normalized with respect to the state space. As we have shown in Section 5.2 the range of $P_\theta(\mathbf{x})$ is always $[0, \frac{1}{2}]$ and does not depend on the spread of the input distributions. The positions of the maxima of $g_\theta(\mathbf{x})$ and $P_\theta(\mathbf{x})$ are not necessarily identical. Comparisons for three different pairs of input distributions are shown in Fig. 5.

The ICD can be computed directly for data with analytically defined PDFs. For LCP we need CDFs which for the major practical cases are also available, either as closed analytical expressions or as fast numerical approximations. For instance, for normally distributed input data the Erf function has to be computed; an efficient and accurate approximation is presented in [46]. If such approximations are used, both approaches are suitable for interactive real-time visualization of uncertain spatial data.

We chose to use the LCP exclusively in the remainder of this paper, because a probability field is more easily interpretable than probability densities with respect to the state space. The fact that the range of $P_\theta(\mathbf{x})$ does not depend on the input data also simplifies the implementation.
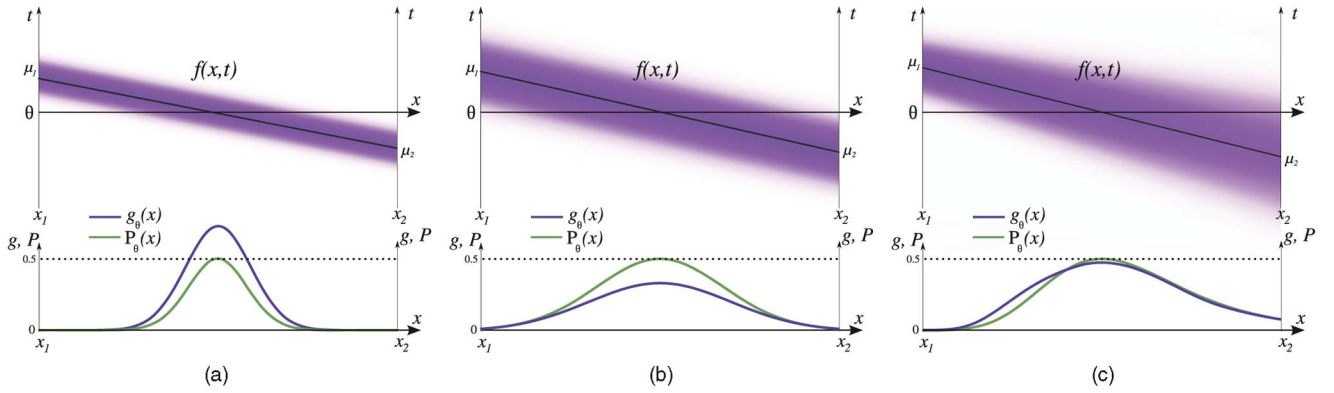
Fig. 5. Comparison between the ICD $g_\theta(x)$ and the LCP $P_\theta(x)$ for different Gaussian input distributions in 1D. The probability density $f(x,t)$ is indicated by a colormap.

6 VISUALIZATION METHODS

The definitions in the previous section are used as a basis for the design of interactive visualization methods that depict uncertain isolines and isosurfaces, and not only give an impression of the uncertainty, but also depict quantitative measures.

The pure display of the distribution in 3D data can be difficult to interpret (cf. Fig. 7). In order to improve 3D impression, we can depict both the crisp isocontours $\mu^{-1}(\theta)$ of the mean field and the spatial distribution via the LCP.

To uncover the *reason* for the spatial distribution of an uncertain isocontour, both numerical condition κ_{abs} and the standard deviation $\sigma(\mathbf{x})$ of the input distributions can be depicted simultaneously in two separate images. The numerical condition on crisp isosurfaces is shown by a colormap. In addition to that, the values of the standard deviation $\sigma(\mathbf{x})$ can be indicated by the length of line glyphs, see Fig. 11d.

For 2D input data, the LCP can easily be displayed as a 2D distribution using a colormap. For each texel the LCP has to be evaluated with respect to a given isovalue θ. The visualization of uncertain isosurfaces in real time is realized using GPU-based ray casting [47]. Instead of employing user-defined transfer functions, the LCP is evaluated and the values are mapped to *transparency* and *color* which then are used in a emission-absorption model for ray integration. For combined volume and surface rendering the intersection of the ray and $\mu^{-1}(\theta)$ is tested in each step of the ray integration.

7 RESULTS

In the following, we apply the visualization methods to various data sets. Using a nonoptimized implementation we achieve frame rates between 5 and 25 fps on an Intel Xeon X5550 2.66 GHz system with a GeForce GTX 285 GPU.

7.1 Fuel Data Set

As an example, an uncertain isoline in a slice of the fuel data set is depicted in Fig. 6. We assume normal distributions as input data and artificially set the standard deviation to be constant $\sigma = 5$ (about 2 percent of the range of values in the data set) because no information about the uncertainty was available. The LCP is displayed using a colormap and in combination with the crisp isoline $\mu^{-1}(\theta)$. In Fig. 7

isosurfaces in the same data set are shown. The condition number mapped to surface color (left column), a volume rendering of the LCP (middle column) and a combined rendering of the LCP and the mean surface are displayed for isovalues $\theta = 90$ and $\theta = 22$. The combined rendering (right column) improves 3D impression of the visualization compared to the volume rendering alone. The uncertain isosurfaces in the second row reveal higher position uncertainty compared to the first row due to higher condition numbers in the respective areas. The fact that position uncertainty inside a closed mean surface is occluded when surface and volume rendering are combined can be met by the user by placing clipping planes to make the interior of the mean surface visible, cf. Fig. 8a. A close-up is shown in Fig. 8b. For comparison, the LCP considering uniform instead of normal distributions as input is displayed in Fig. 8c.

7.2 Medical Volume Data

Medical volume data from CT scanners usually does not contain explicit uncertainty information, but the amount of noise in the scans can be used as an estimate. An approach that was also used in [48] to compute the signal to noise ratio is based on analysis of homogenous subsets of images (i.e., areas with constant signal). In case no area with constant signal is available in a data set, alternative methods such as single image SNR estimation [49] should be applied.
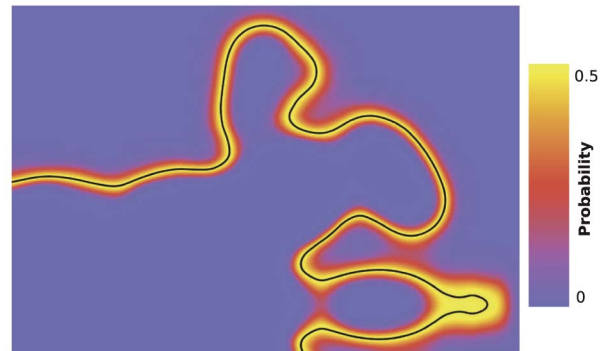

Fig. 6. The probability $P_\theta(\mathbf{x})$ (LCP) for an uncertain isocontour in a slice of the fuel data set is displayed using a colormap in combination with the crisp isoline $\mu^{-1}(\theta)$ of the mean field (black).
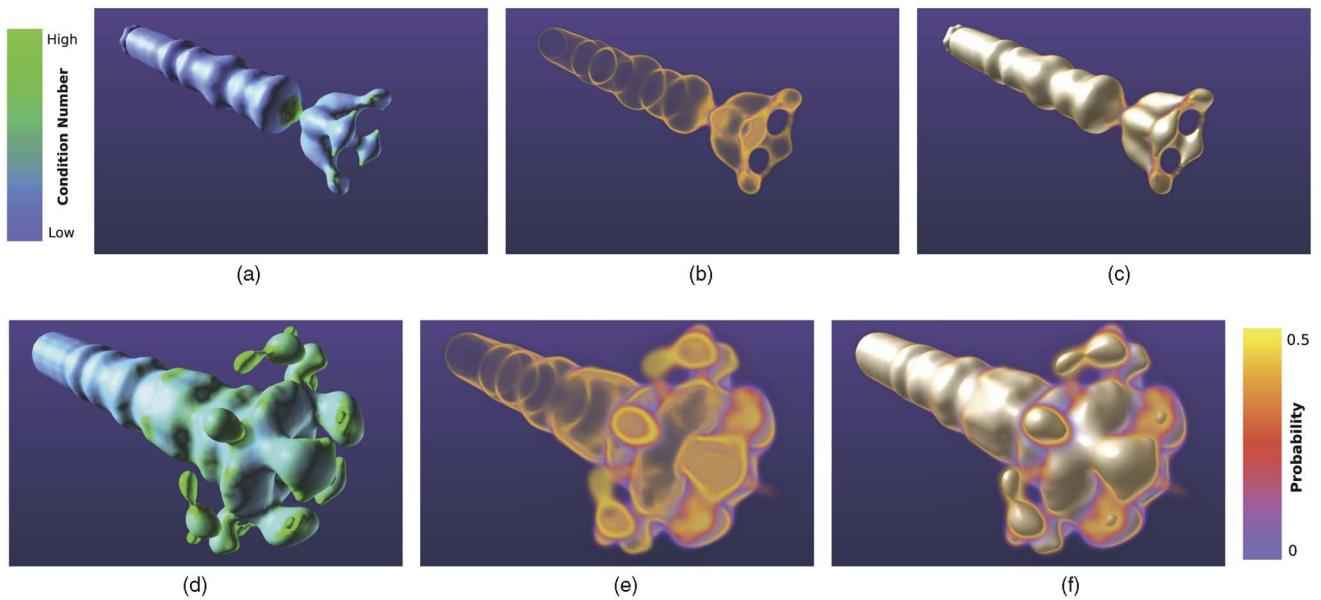
Fig. 7. Isosurfaces of the fuel data set with a fixed (artificial) standard deviation $\sigma = 5$. The condition number for $\mu^{-1}(\theta)$ is color mapped in the left column. The LCP is depicted by volume rendering in the middle column. The right column displays combined renderings of the LCP and the mean surface. The first row of images use isovalue $\theta = 90$ and reveal low positional uncertainty; the second row ($\theta = 22$) uncovers higher positional uncertainty due to higher condition numbers.

As an example, we considered a data set used for planning an implant in the middle ear of a patient. Here the size and number of connected air pockets is important. We estimated the standard deviation of the noise from areas that contain air only. The noise was approximately normally distributed. We denoised the CT scan using a median filter and considered this as $\mu(\mathbf{x})$. The estimated standard deviation ($\sigma \approx 13.4$) is used to display an uncertain isosurface that depicts air pockets in the middle ear (shown in Fig. 9). This shows that the topology of the isosurface is not clearly defined. The mean isosurface has several distinct parts that are connected by relatively large LCP values. Thus, the number, size, and shape of air pockets are highly uncertain. This also implies that a segmentation of these areas by thresholding can lead to erroneous results.
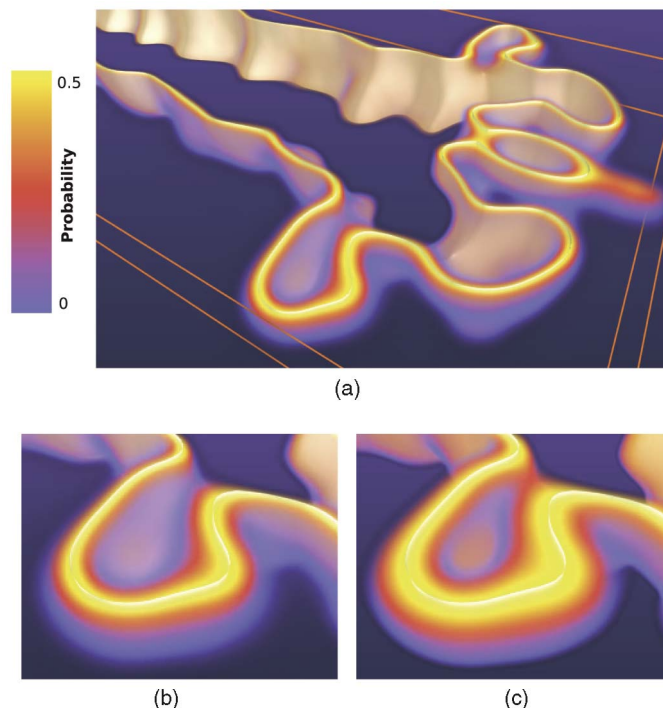

Fig. 8. Uncertain isosurface bounded by two clipping planes to display $P_\theta(\mathbf{x})$ inside the mean surface (a). Normal distributions are considered as input data in (a) and the closeup (b). When considering uniform distributions we obtain the spatial distribution depicted in (c).
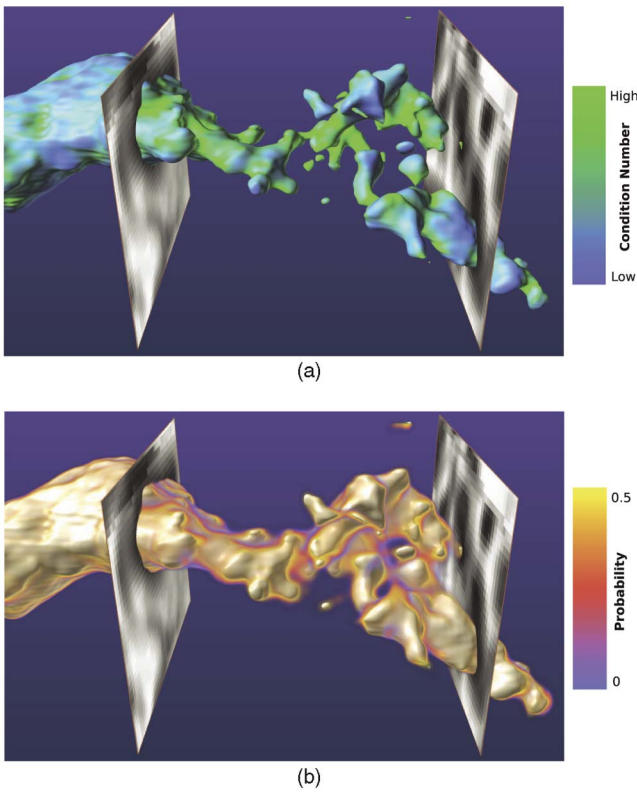
Fig. 9. Isosurface with numerical condition mapped to color (a) and an uncertain isosurface in a CT scan of the middle ear region (b).

7.3 Simulated Climate Data

Uncertainty in climate simulations is often represented by *ensembles* which contain multiple results for the simulated quantities. As an example, we use daily average hindcast data from the DEMETER project [50] where the results of seven different climate models and nine different sets of simulation parameters each constitute ensembles with 63 members. This way, the simulation accounts for both model uncertainty and uncertainty of the input data.

To prepare for the extraction of uncertain isocontours, the data in a *temperature field* ensemble for February 20th, 2000 is analyzed statistically. For each location $\mathbf{x}$ we compute the ensemble mean $\mu(\mathbf{x})$ and standard deviation $\sigma(\mathbf{x})$ (see Figs. 10a and 10b). Again, we model the uncertainty by normal distributions. In Fig. 10c the

histogram of temperatures at a single location, but considering all ensemble members, is shown.

From these fields, we are able to extract uncertain isolines (isotherms) using the LCP. Figs. 11a, 11b, and 11c depict uncertain isolines for $-25°C$, $0°C$, and $25°C$, respectively, which reveal highly varying position uncertainty around the crisp ensemble mean isoline (black).

From DEMETER not only temperatures at the two-meter level but also temperatures at the pressure levels 850 hPa, 500 hPa, and 200 hPa above the earth's surface are available. We analyze the ensembles for these levels in the same manner as the two-meter temperature field and use all results to construct a volume data set with pressure mapped to the third coordinate. This volume is used for the extraction of uncertain isotherm surfaces. In Fig. 11d the numerical condition for the crisp mean surface is displayed by a colormap for isovalue $\theta = 25°C$ and the values of $\sigma(\mathbf{x})$ are indicated by the length of line glyphs at that surface. The corresponding uncertain isotherm surface is shown in Fig. 11e.

Comparing Figs. 11d and 11e we see that low values of $\sigma(\mathbf{x})$ with high values of κ_{abs} and high values of $\sigma(\mathbf{x})$ with low values of κ_{abs} both can result in similar amount of position uncertainty. Fig. 12 shows an uncertain isotherm surface for $0°C$. This example shows that volume rendering of the LCP reveals also structural information that crisp isosurfaces do not display: possible topological changes of crisp isosurfaces are indicated as well as ridges of P_θ.

8 DISCUSSION

Compared to previous approaches to isosurface uncertainty the methods in this paper differ in regard to mathematical modeling and visualization methods.

8.1 Modeling and Computation

Rhodes et al. [9] do not use an error model, but assume that uncertainty is somehow quantified ("error value") and provided in the data set. Grigoryan and Rheingans [7] computed the position of point primitives (depicting a probabilistic surface) by multiplying an uncertainty value, a random number and a user-defined scale factor. When applying the method to data from tumor-growth simulations the authors use the error estimation from the simulation as input, while for clinical data they extract crisp isosurfaces and
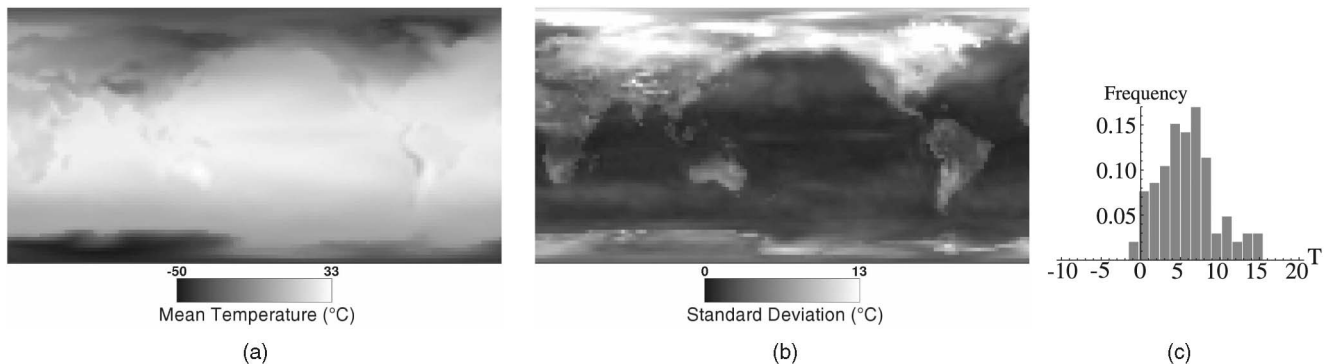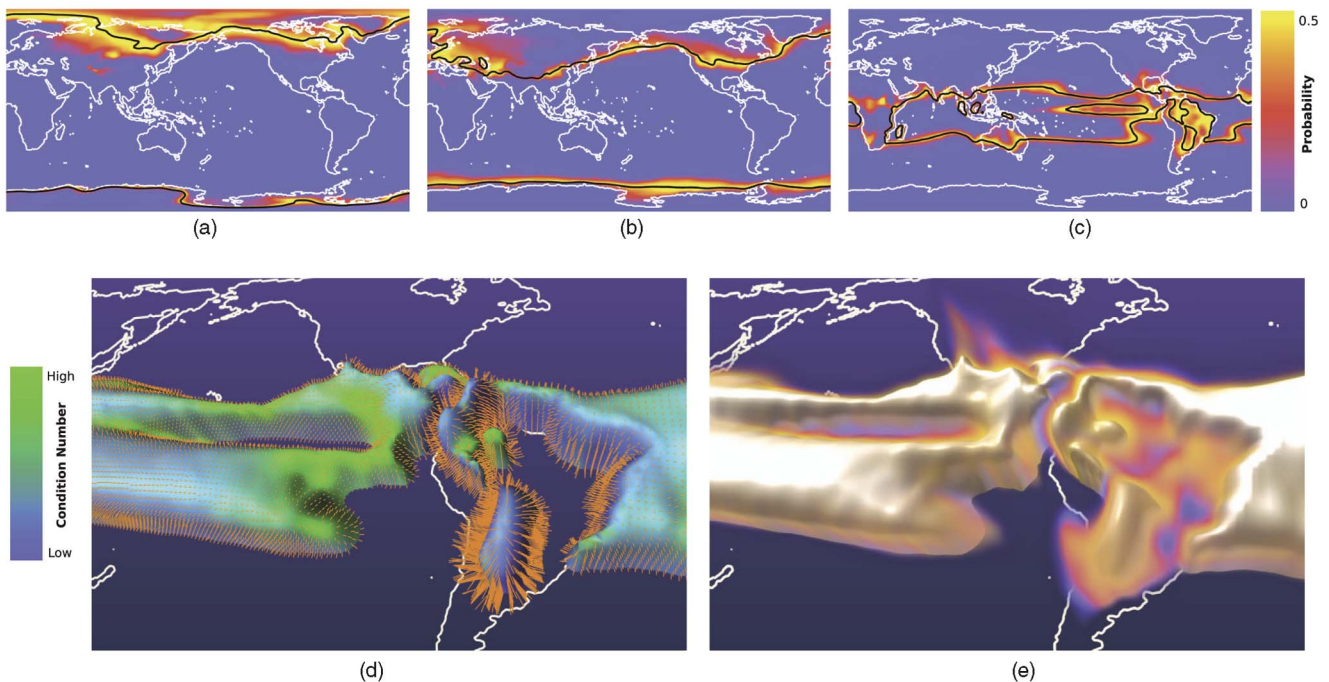

Fig. 10. Simulated climate data: The ensemble mean $\mu(\mathbf{x})$ for the temperature field at 2 m altitude is shown in (a), the standard deviation $\sigma(\mathbf{x})$ in (b), a histogram for the ensemble values at a single location in (c).

Fig. 11. Uncertain isotherms lines for $-25°C$, $0°C$, and $25°C$ are shown in (a)-(c), respectively. The crisp mean isoline is drawn in black. On the mean isotherm surface ($\theta = 25°C$) the condition numbers are mapped to color and the values of $\sigma(\mathbf{x})$ are depicted by the length of line glyphs in (d). The resulting uncertain isosurface is displayed in (e).

use the inverse density gradient magnitude at the surface points as an uncertainty measure. Their model does not consider probability distributions.

As we have shown in Section 4.2 the inverse gradient magnitude at the points of an isocontour is the absolute normwise numerical condition. The position uncertainty of an isocontour, however, depends on the product of the condition number and the uncertainty of the input data. The ICD and LCP can be computed from data sets with uncertainty given by arbitrary probability distributions (with a finite number of moments) and in arbitrary-dimensional parameter spaces. The results of these functions are probabilities and probability densities, which means, that our methods quantify uncertainty in a statistically well-founded way.

8.2 Uncertainty Visualization

The "error value" used by Rhodes et al. [9] is mapped to crisp isosurfaces by color or texture for visualization but the authors do not address positional uncertainty. Grigoryan and Rheingans [7] propose a heuristic to show the position uncertainty of surfaces. Clouds of point primitives are placed along the normals of crisp surfaces. An evaluation shows that this approach is more suitable to clarify position
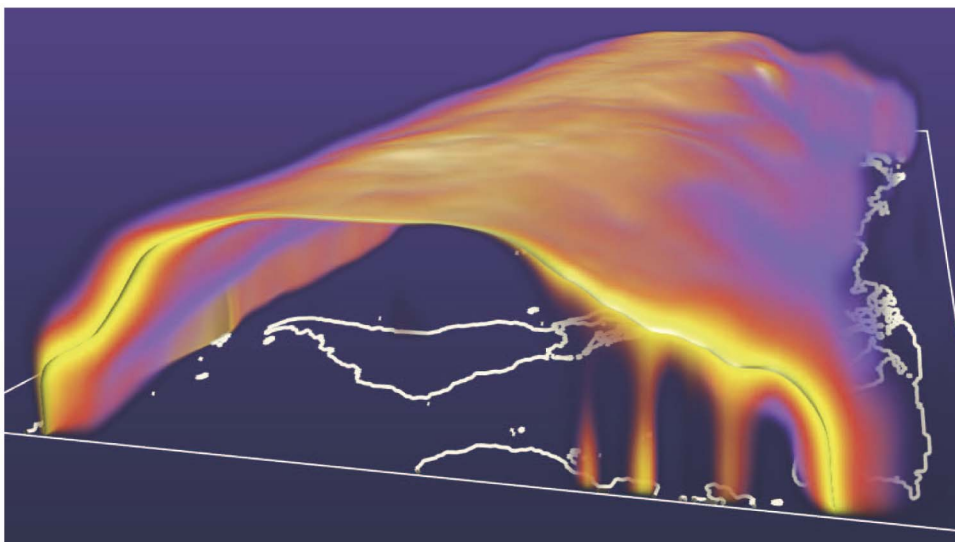

Fig. 12. Uncertain isotherm surface for $0°C$. Note the ridges extending downwards to the earth's surface that are not indicated by the crisp isosurface.

uncertainty compared to color coding. Zuk [28] points out that the user-defined scale factor in [7] may lead to an arbitrary *Lie Factor* in the visualizations.

The visualization methods presented in Section 6 are based on $P_\theta(\mathbf{x})$ and thus the output is a function of the input data only, i.e., there are no user-defined parameters. Difficulties in interpreting visualizations arise in areas of an image where the mean surface is normal to the camera vector. There, the amount of positional uncertainty is difficult to recognize. This can usually be worked around by the user by interactively changing the viewpoint.

The uncertainty aggregated from the application of the visualization algorithms themselves is not considered in this paper.

The separate display of the numerical condition κ_{abs} and the standard deviation $\sigma(\mathbf{x})$ of the data at the surface points helps the user to understand *why* a given uncertain isosurface has a specific spatial distribution. A visualization technique for depicting this information simultaneously in a single image would be useful.

9 CONCLUSIONS AND FUTURE WORK

We introduced the concept of numerical condition to feature-based visualization. Specifically, we applied it to isocontour extraction to examine its sensitivity to uncertainties in the scalar field. The average numerical condition of isocontours has been shown to aid the selection of thresholds that correspond to robust isocontours.

Furthermore, we established a model for uncertain spatial data based on stochastic quantities from which we derived statistically founded measures for the spatial distribution of uncertain isocontours. For this, we spatially interpolate probability density functions between the grid points and define two quantities, the *isocontour density* and the *level crossing probability field*. These definitions match our intuitive conception of fuzzy isocontours.

The LCP is employed in interactive visualization methods for uncertain 2D and 3D volume data. Uncertain isolines are displayed as spatial distributions in 2D using color mapping. For depiction in 3D the introduced quantitative measures are used as procedural transfer functions in GPU assisted ray casting. No preprocessing is necessary and interactive frame rates are achieved. These methods were applied to data sets from engineering, medicine, and climate research.

Our approach is suitable for all applications which acquire data using standard uncertainty estimation and for which the assumptions stated in Sections 3 and 5 are appropriate. Goals for further development are the generalization of our methods to different data representations, e.g., nonparametric distributions, and the consideration of correlation structures, where applicable. Furthermore, we will apply the concept of condition numbers to other features in scalar, vector, and tensor fields. Another topic for future work is the evaluation of the effects of uncertainty in isocontour visualizations on inference.

ACKNOWLEDGMENTS

REFERENCES

[1] B. Taylor and C. Kuyatt, *Guidelines for Expressing and Evaluating the Uncertainty of NIST Experimental Results,* NIST Technical Note 1297, 1994.

[2] *ISO/IEC Guide 98-3:2008 Uncertainty of Measurement—Part 3: Guide to the Expression of Uncertainty in Measurement (GUM),* Int'l Organization for Standardization, 2008.

[3] W. Feller, *Introduction to Probability Theory and Its Applications.* John Wiley and Sons, 1968 and 1971.

[4] P. Fornasini, *The Uncertainty in Physical Measurements.* Springer, 2008.

[5] I. Lira, *Evaluating the Measurement Uncertainty: Fundamentals and Practical Guidance.* Inst. of Physics Publishing, 2002.

[6] C.R. Johnson and A.R. Sanderson, "A Next Step: Visualizing Errors and Uncertainty," *IEEE Computer Graphics and Applications,* vol. 23, no. 5, pp. 6-10, Sept./Oct. 2003.

[7] G. Grigoryan and P. Rheingans, "Point-Based Probabilistic Surfaces to Show Surface Uncertainty," *IEEE Trans. Visualization and Computer Graphics,* vol. 10, no. 5, pp. 564-573, Sept./Oct. 2004.

[8] G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Möller, "Curvature-Based Transfer Functions for Direct Volume Rendering: Methods and Applications," *Proc. IEEE Visualization 2003 Conf.,* pp. 513-520, Oct. 2003.

[9] P.J. Rhodes, R.S. Laramee, R.D. Bergeron, and T.M. Sparr, "Uncertainty Visualization Methods in Isosurface Rendering," *Proc. Eurographics 2003, Short Papers,* pp. 83-88, 2003.

[10] A.T. Pang, C.M. Wittenbrink, and S.K. Lodha, "Approaches to Uncertainty Visualization," *The Visual Computer,* vol. 13, no. 8, pp. 370-390, 1997.

[11] H. Griethe and H. Schumann, "The Visualization of Uncertain Data: Methods and Problems," *SimVis,* T. Schulze, G. Horton, B. Preim, and S. Schlechtweg, eds., pp. 143-156, SCS Publishing House e.V., Mar. 2006.

[12] P. Fornasini, *The Uncertainty in Physical Measurements: An Introduction to Data Analysis in the Physics Laboratory.* Springer, 2008.

[13] A. Luo, D. Kao, and A. Pang, "Visualizing Spatial Distribution Data Sets," *Proc. Symp. Data Visualization 2003 (VISSYM '03),* pp. 29-38, 2003.

[14] L. Zadeh, "Fuzzy Sets," *Information Control,* vol. 8, pp. 338-353, 1965.

[15] W.A. Lodwick, *Fuzzy Surfaces in GIS and Geographical Analysis: Theory, Analytical Methods, Algorithms and Applications.* CRC Press, 2008.

[16] S. Djurcilov, K. Kim, P. Lermusiaux, and A. Pang, "Visualizing Scalar Volumetric Data with Uncertainty," *Computers and Graphics,* vol. 26, no. 2, pp. 239-248, 2002.

[17] C.M. Wittenbrink, A.T. Pang, and S.K. Lodha, "Glyphs for Visualizing Uncertainty in Vector Fields," *IEEE Trans. Visualization and Computer Graphics,* vol. 2, no. 3, pp. 266-279, Sept. 1996.

[18] R.P. Botchen, D. Weiskopf, and T. Ertl, "Texture-Based Visualization of Uncertainty in Flow Fields," *Proc. IEEE Visualization 2005 Conf.,* pp. 647-654, 2005.

[19] T. Zuk, J. Downton, D. Gray, S. Carpendale, and J. Liang, "Exploration of Uncertainty in Bidirectional Vector Fields," *Proc. SPIE and IS&T Conf. Electronic Imaging, Visualization and Data Analysis 2008,* p. 68090B, 2008.

[20] M. Otto, T. Germer, H.-C. Hege, and H. Theisel, "Uncertain 2D Vector Field Topology," *Computer Graphics Forum,* vol. 29, pp. 347-356, 2010.

[21] H. Li, C.-W. Fu, Y. Li, and A. Hanson, "Visualizing Large-Scale Uncertainty in Astrophysical Data," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1640-1647, Nov./Dec. 2007.

[22] J.M. Kniss, R.V. Uitert, A. Stephens, G.-S. Li, T. Tasdizen, and C. Hansen, "Statistically Quantitative Volume Visualization," *Proc. IEEE Visualization 2005 Conf.,* pp. 287-294, Oct. 2005.

[23] C. Lundstrom, P. Ljung, A. Persson, and A. Ynnerman, "Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1648-1655, Nov./Dec. 2007.

[24] M. Pauly, N. Mitra, and L. Guibas, "Uncertainty and Variability in Point Cloud Surface Data," *Proc. Eurographics Symp. Point-Based Graphics,* pp. 77-84, 2004.

[25] S. Djurcilov and A. Pang, "Visualizing Sparse Gridded Data Sets," *IEEE Computer Graphics and Applications,* vol. 20, no. 5, pp. 52-57, Sept./Oct. 2000.

[26] B. Zehner, N. Watanabe, and O. Kolditz, "Visualization of Gridded Scalar Data with Uncertainty in Geosciences," to be published in *Computers and Geoscience,* 2010.

[27] A. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler, "Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know," *Cartography and Geographic Information Science,* vol. 32, no. 3, pp. 139-161, 2005.

[28] T. Zuk, "Visualizing Uncertainty," PhD thesis, Univ. of Calgary, Apr. 2008.

[29] H. Jänicke, A. Wiebel, G. Scheuermann, and W. Kollmann, "Multifield Visualization Using Local Statistical Complexity," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1384-1391, Nov./Dec. 2007.

[30] C. Wang, H. Yu, and K.-L. Ma, "Importance-Driven Time-Varying Data Visualization," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1547-1554, Nov./Dec. 2008.

[31] R.J. Adler, J.E. Taylor, and K.J. Worsley, *Applications of Random Fields and Geometry,* Preprint, Technion-Israel Inst. of Technology, 2009.

[32] D.S. Moore, G.P. McCabe, and M.J. Evans, *Introduction to the Practice of Statistics.* W. H. Freeman & Co., 2005.

[33] R.J. Adler and J. Taylor, *Random Fields and Geometry.* Springer, 2007.

[34] J.-M. Azaïs and M. Wschebor, *Level Sets and Extrema of Random Processes and Fields,* ch. 6. John Wiley & Sons, 2009.

[35] R.J. Adler, *The Geometry of Random Fields.* John Wiley and Sons, 1981.

[36] W.E. Lorensen and H.E. Cline, "Marching Cubes: A High Resolution 3D Surface Construction Algorithm," *ACM SIGGRAPH Computer Graphics,* vol. 21, no. 4, pp. 163-169, 1987.

[37] P. Deuflhard and A. Hohmann, *Numerical Analysis in Modern Scientific Computing: An Introduction.* Springer, 2003.

[38] H. Federer, *Geometric Measure Theory.* Springer, 1969.

[39] C.E. Scheidegger, J.M. Schreiner, B. Duffy, H. Carr, and C.T. Silva, "Revisiting Histograms and Isosurface Statistics," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1659-1666, Nov./Dec. 2008.

[40] J.C.R. Hunt, "Vorticity and Vortex Dynamics in Complex Turbulent Flows," *Canadian Soc. for Mechanical Eng. Trans.,* vol. 11, no. 1, pp. 21-35, 1987.

[41] J. Sahner, T. Weinkauf, N. Teuber, and H.-C. Hege, "Vortex and Strain Skeletons in Eulerian and Lagrangian Frames," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 5, pp. 980-990, Sept./Oct. 2007.

[42] Y.-H. Ryu and J.-J. Baik, "Flow and Dispersion in an Urban Cubical Cavity," *Atmospheric Environment,* vol. 43, no. 10, pp. 1721-1729, 2009.

[43] C.L. Bajaj, V. Pascucci, and D.R. Schikore, "The Contour Spectrum," *Proc. IEEE Visualization 1997 Conf.,* pp. 167-174, 1997.

[44] V. Pekar, R. Wiemker, and D. Hempel, "Fast Detection of Meaningful Isosurfaces for Volume Data Visualization," *Proc. IEEE Visualization 2001 Conf.,* pp. 223-230, 2001.

[45] R.P. Kanwal, *Generalized Functions: Theory and Technique.* Birkhäuser, 1998.

[46] S. Winitzki, "A Handy Approximation for the Error Function and Its Inverse," http://homepages.physik.uni-muenchen.de/Winitzki/erf-approx.pdf, 2008.

[47] J. Krüger and R. Westermann, "Acceleration Techniques for GPU-Based Volume Rendering," *Proc. IEEE Visualization 2003 Conf.,* pp. 287-292, 2003.

[48] M. Firbank, A. Coulthard, R. Harrison, and E. Williams, "A Comparison of Two Methods for Measuring the Signal to Noise Ratio on MR Images," *Physics in Medicine and Biology,* vol. 44, no. 12, pp. N261-N264, 1999.

[49] J. Thong, K. Sim, and J. Phang, "Single-Image Signal-to-Noise Ratio Estimation," *Scanning,* vol. 23, no. 5, pp. 328-336, 2001.

[50] T.N. Palmer et al., "Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER)," Technical Memorandum, European Centre for Medium-Range Weather Forecasts, 2004.

Kai Pöthkow received the MS degree in computer science and ethnomusicology from Freie Universität Berlin, Germany, in 2009. Since 2006, he is with the Scientific Visualization Department of Zuse Institute Berlin (ZIB), where he is currently working toward the PhD degree. His research interests include uncertainty visualization, music visualization, and volume rendering.

Hans-Christian Hege is head of the Visualization and Data Analysis Department at Zuse Institute Berlin (ZIB). After studying physics and mathematics, he performed research in computational physics and quantum field theory at Freie Universität Berlin (1984-1989). Then, he joined ZIB, initially as a scientific consultant for high-performance computing and then as head of the Scientific Visualization Department, which he started in 1991. His group performs research in visual data analysis and develops visualization software such as Amira and Biosphere3D. He is also the cofounder of Mental Images (1986), Indeed-Visual Concepts (1999) (now Visage Imaging), and Lenné3D (2005). He taught as a guest professor at Universitat Pompeu Fabra, Barcelona, and as honorary professor at the German Film School (University for Digital Media Production). His research interests include many branches in visual computing and applications in natural sciences, life sciences, and engineering. He is a member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.