Uncertainty in Continuous Scatterplots, Continuous Parallel Coordinates, and Fibers

Boyan Zheng and Filip Sadlo

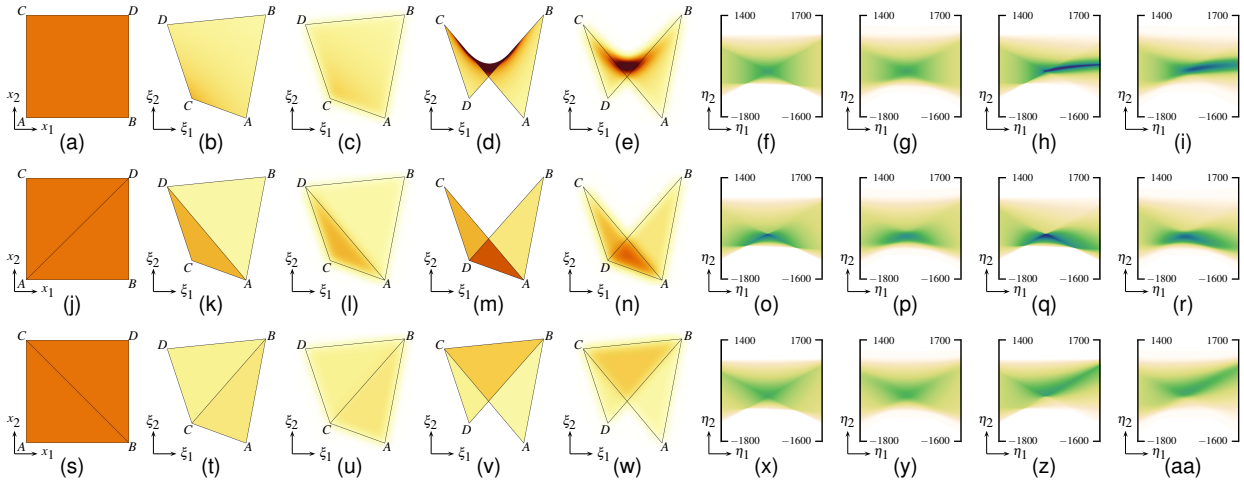

Fig. 1. Bivariate Square dataset. Original geometry (upper row), geometry subdivided along A–D (middle row), and along B–C (bottom row). Swapped values of C and D in columns four, five, eight, and nine. First column: geometry; second and fourth column: traditional (without uncertainty) continuous scatterplots [3]; third and fifth column: uncertain continuous scatterplots using our sampling-based approach; column six and eight: traditional (without uncertainty) continuous parallel coordinates [19]; columns seven and nine: our uncertain continuous parallel coordinates. All continuous scatterplots in this paper are color-coded by accumulated density (yellow (low) to red (high)), and all continuous parallel coordinates color-coded by accumulated density (yellow (low) to blue (high)).

Abstract—In this paper, we introduce uncertainty to continuous scatterplots and continuous parallel coordinates. We derive respective models, validate them with sampling-based brute-force schemes, and present acceleration strategies for their computation. At the same time, we show that our approach lends itself as well for introducing uncertainty into the definition of fibers in bivariate data. Finally, we demonstrate the properties and the utility of our approach using specifically designed synthetic cases and simulated data.

Index Terms—Multivariate data, uncertainty visualization, uncertain continuous scatterplots, uncertain continuous parallel coordinates, uncertain fibers.

1 Introduction

There is hardly any problem in research and engineering that does not include multivariate data. In some of these cases, the data are given in scattered representation, i.e., as a collection of samples without continuity or a domain. In such cases, statistical analysis is a primary approach for obtaining insights, among which scatterplots, which map bivariate data to the coordinates of the plane, are very simple but often very useful. Since scatterplots of more than two attributes would become volumetric and thus hard to present and investigate, parallel coordinates, which map each attribute to a distinct parallel axis, have become a primary technique for visualizing multivariate data with attribute count higher than two.

For a long time, scatterplots and parallel coordinates have also been employed for multivariate samples from continuous data, such as fields defined on a continuous domain. This, however, is susceptible to sampling-induced artifacts, and would require very high sampling density to reveal the true data distribution. This is where continuous scatterplots (CSP) and continuous parallel coordinates (CPC) play their strengths, by efficiently providing the respective representations that would be obtained if the sample count would reach infinity.

With the advent of modern measurement and simulation techniques, as well as increased requirements on reliability and interpretability in data analysis, uncertainty has become an important component and an active field of research in visualization. Continuous scatterplots, continuous parallel coordinates, as well as fibers (the generalization of isocontours to multivariate data) have, however, not yet been extended to cope with uncertain data. It is the focus of this work to contribute the extension of these three visualization concepts to uncertain multivariate data, and thus to pave the way for a more reliable and integrative data analysis.

The main contributions of this work are:

- extension of continuous scatterplots to uncertain data,
- two algorithms for their accelerated computation, whereof
- one without restrictions on the type of input grid,
- extension of continuous parallel coordinates to uncertain data,
- extension of fibers to uncertain data.

2 Related Work

Due to their rather straightforward nature and very wide use, it is hard to determine who introduced scatterplots. Parallel coordinates, in contrast, have been introduced comparably recently by Inselberg [20] in

- *Boyan Zheng is with Heidelberg University, Germany. E-mail: boyan.zheng@iwr.uni-heidelberg.de.*
- *Filip Sadlo is with Heidelberg University, Germany. E-mail: sadlo@uni-heidelberg.de.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

1985. Continuous scatterplots [3] were introduced by Bachthaler and Weiskopf in 2008, followed by two approaches for their accelerated computation [2, 4]. The extension of parallel coordinates to continuous domains was accomplished by Heinrich and Weiskopf [19] with their continuous parallel coordinates concept. Based on that, Lehmann and Theisel focused on features in continuous scatterplots and continuous parallel coordinates [25, 26], and Zheng et al. [42] investigated equivalence in bivariate 2D fields. Heinrich et al. [18] investigated the progressive splatting of these plots. None of these works included uncertainty of the multivariate data, as we do.

Once acquired with a reasonable model, it is quite simple to add uncertainty as an additional "channel" to existing data representations, and many application domains see strong needs and challenges in the visualization of uncertain continuous data [27]. Extending existing visualization methods for uncertain data, however, is far beyond straightforward application [6], similar to the problem of visualizing ensembles. Otto et al. [28, 29] presented vector field topology for 2D and 3D uncertain vector fields. Pfaffelmoser et al. visualized the variability of isosurfaces and gradients, and global correlation structures in uncertain scalar fields [32–34]. Pöthkow et al., on the other hand, presented several works on feature extraction from uncertain scalar and vector fields [31, 35–37]. For ensembles, Ferstl et al. presented works for analyzing the spatial variability of isocontours in ensemble scalar fields and of streamlines in ensemble flow fields [15, 16]. Favelier et al. and Günther et al., on the other hand, focused on the variability of critical points in ensembles [13, 17]. However, all these works focus on visualizing uncertainty in the spatial domain, whereas our work focuses on its visualization in the data domain in terms of continuous scatterplots and continuous parallel coordinates. Feng et al. presented approaches for adding uncertainty to plots [14], however, based on non-continuous scatterplots and non-continuous parallel coordinates, not their continuous counterparts, and not closely related to ours.

Finally, there are several previous works on the visualization of isocontours in uncertain scalar fields [33, 35, 37]. On the other hand, Carr and colleagues presented fiber surfaces [8, 24, 39] for generalizing isosurfaces to bivariate data, however without uncertainty. Our work can be seen at the interface between these works.

3 Mathematical Model

3.1 Model Without Uncertainty

For introduction and self-contained presentation, we first summarize the mathematical models of the previous works on continuous scatterplots [3] and continuous parallel coordinates [19]. Both plot types require the input data to be a continuous mapping $\boldsymbol{\tau}\colon \mathbb{R}^n \to \mathbb{R}^m$ from the n-dimensional spatial domain Ω to the m-dimensional data domain Ψ.

Bachthaler and Weiskopf [3] formulate the construction of *continuous scatterplots* as finding a density function κ defined in the data domain:

$$\kappa\colon \mathbb{R}^m \to \mathbb{R}, \quad \boldsymbol{\xi} \longmapsto \kappa(\boldsymbol{\xi}), \quad \text{with } \boldsymbol{\xi} \in \Psi. \tag{1}$$

Under the assumptions that a mass density s is given in the spatial domain,

$$s\colon \mathbb{R}^n \to \mathbb{R}, \quad \mathbf{x} \longmapsto s(\mathbf{x}), \quad \text{with } \mathbf{x} \in \Omega \tag{2}$$

(which is typically uniform, i.e., $s(\mathbf{x}) \equiv 1$), that the mapping $\boldsymbol{\tau}$ does not change the number of sample points, and that the sample points have mass, Bachthaler and Weiskopf derive the mathematical model of continuous scatterplots by mass conservation.

The mass M of a volume $D \subset \Omega$ in the spatial domain is thus $M = \int_D s(\mathbf{x})\,\mathrm{d}^n\mathbf{x}$. Defining $\Delta := \boldsymbol{\tau}(D)$, i.e., the respective volume in the data domain, mass conservation implies

$$\int_{\Delta=\boldsymbol{\tau}(D)} \kappa(\boldsymbol{\xi})\,\mathrm{d}^m\boldsymbol{\xi} = \int_D s(\mathbf{x})\,\mathrm{d}^n\mathbf{x} = M\,. \tag{3}$$

For the case $m = n$, where the transformation theorem for integrals can be applied, this leads to

$$\kappa(\boldsymbol{\tau}(\mathbf{x})) = \frac{s(\mathbf{x})}{|\det(\nabla\boldsymbol{\tau})(\mathbf{x})|}\,, \tag{4}$$

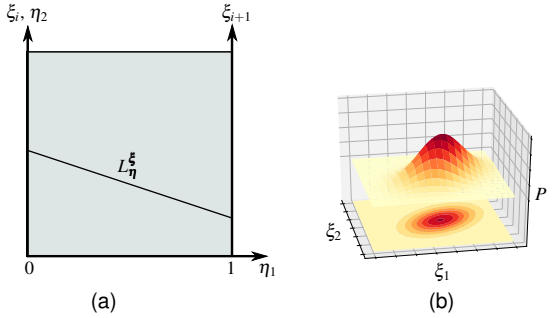

Fig. 2. Independent 2D parallel coordinates domain (shaded, spanned by η_1 and η_2) (a), and an example for a bivariate normal distribution (b).

with $\nabla\boldsymbol{\tau}$ being the $n \times n$ Jacobian matrix.

In the case of $m < n$, for which $\det(\nabla\boldsymbol{\tau})$ does not exist and the transformation theorem for integrals does not apply, Bachthaler and Weiskopf adapt Equation 4 accordingly. Since the preimage of a point $\boldsymbol{\xi}$ in the data domain is a fiber $\boldsymbol{\tau}^{-1}(\boldsymbol{\xi})$ in the spatial domain with dimensionality $(n-m)$, the density $\kappa(\boldsymbol{\xi})$ at $\boldsymbol{\xi}$ can be computed by integrating the density over that fiber, i.e.,

$$\kappa(\boldsymbol{\xi}) = \int_{\boldsymbol{\tau}^{-1}(\boldsymbol{\xi})} \frac{s(\mathbf{x})}{|\mathrm{Vol}(\nabla\boldsymbol{\tau})(\mathbf{x})|}\,\mathrm{d}^{(n-m)}\mathbf{x}\,, \tag{5}$$

where $|\mathrm{Vol}(\nabla\boldsymbol{\tau})(\mathbf{x})|$ is the volume spanned by the partial derivatives of $\boldsymbol{\tau}$, which are restricted to parameter variations in the normal space $\boldsymbol{\tau}^{-1}_{\mathrm{normal}(\mathbf{x})}$ [3, Section 3.3].

The construction of *continuous parallel coordinates* is derived from the mapping between the 2D data domain and the 2D parallel coordinates domain [19]. Each consecutive pair of axes ξ_i and ξ_{i+1} of a parallel coordinates plot spans an independent 2D parallel coordinates system (shaded area in Figure 2a) in a Cartesian coordinate system with axes η_1 and η_2. Due to the point–line duality in parallel coordinates [20–22], a point $\boldsymbol{\xi} = (\xi_i, \xi_{i+1})^\top$ in the 2D data subdomain Ψ_i (with $(\xi_i, \xi_{i+1})^\top \in \Psi_i$) equals a line segment $L^{\boldsymbol{\xi}}_{\boldsymbol{\eta}}\colon \eta_2 = (\xi_{i+1} - \xi_i)\eta_1 + \xi_i$ in the independent 2D parallel coordinates domain, which can be formulated as

$$\eta_2 = \mathbf{n} \cdot \boldsymbol{\xi}\,, \tag{6}$$

using $\mathbf{n} := (1 - \eta_1, \eta_1)^\top$. This implies that the mapping between a point $\boldsymbol{\eta} = (\eta_1, \eta_2)^\top$ in the independent 2D parallel coordinates domain and the corresponding line $L^{\boldsymbol{\eta}}_{\boldsymbol{\xi}}$ in the 2D data domain depends on η_1 only. Let now $\varphi\colon \mathbb{R}^2 \to \mathbb{R}, \boldsymbol{\eta} \longmapsto \varphi(\boldsymbol{\eta})$ be the density at a point $\boldsymbol{\eta}$ in parallel coordinates [19]. Based on mass conservation, the mass of an interval Θ of η_2 computes

$$\int_\Theta \varphi(\boldsymbol{\eta})\,\mathrm{d}\eta_2 = \int_{\Delta_i} \kappa(\boldsymbol{\xi})\,\mathrm{d}^2\boldsymbol{\xi} = M\,, \tag{7}$$

with $\Delta_i \subset \Psi_i$. Using that, Heinrich and Weiskopf express the density $\varphi(\boldsymbol{\eta})$ at a point $\boldsymbol{\eta}$ in the 2D parallel coordinates domain as

$$\varphi(\boldsymbol{\eta}) = \int_{L^{\boldsymbol{\eta}}_{\boldsymbol{\xi}}} \frac{1}{\|\mathbf{n}\|}\,\kappa\left(\mathbf{L}^{\boldsymbol{\eta}}_{\boldsymbol{\xi}}(t)\right)\,\mathrm{d}t\,, \tag{8}$$

with $\mathbf{L}^{\boldsymbol{\eta}}_{\boldsymbol{\xi}}(t)$ being the arc-length (t) parametrized line $L^{\boldsymbol{\eta}}_{\boldsymbol{\xi}}$ [19].

3.2 Model With Uncertainty

If uncertainty is introduced to the input data, then a point in the spatial domain always links to a point with an uncertain position in the data domain, or, in other words, a distribution. Obviously, the above models introduced by Weiskopf and colleagues need to be adapted accordingly.

For modeling the uncertainty, we use the basic idea by Otto et al. [28, 29], who represent 2D (3D) uncertain vector fields as 4D (6D)

scalar fields. The first two (three) dimensions of this scalar field represent the original domain, whereas the second two (three) dimensions represent the value. The scalar entry at the respective 4D (6D) coordinate stores the probability for the vector field values to be within a respective bin. We simply extend their representation to uncertain n-dimensional m-variate fields $\tilde{\boldsymbol{\tau}}\colon \mathbb{R}^n \to \mathbb{R}^m$. That is, $\tilde{\boldsymbol{\tau}}$ transforms to a scalar field $P(\mathbf{x};\boldsymbol{\xi}) \geq 0$ defined on a $(n+m)$-dimensional domain Γ, with $\int_{\mathbb{R}^m} P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}^m\boldsymbol{\xi} = 1$, for all $\mathbf{x} \in \Omega$. Notice, that $P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}^m\boldsymbol{\xi}$ represents the probability that, at the space location $\mathbf{x}$, the field $\tilde{\boldsymbol{\tau}}$ has some value within the range $[\xi_1, \xi_1 + \mathrm{d}\xi_1] \times \cdots \times [\xi_m, \xi_m + \mathrm{d}\xi_m]$.

Using this formulation, the density $s(\mathbf{x})$ at a certain position $\mathbf{x}$ in the spatial domain is

$$s(\mathbf{x}) = \int_{\mathbb{R}^m} P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}^m\boldsymbol{\xi} = 1\,. \tag{9}$$

To construct continuous scatterplots of our uncertain input data, $\tilde{\boldsymbol{\tau}}(D)$ is not a fixed volume Δ in the data domain anymore, but a distribution with infinite range. Furthermore, the scatterplot has to reflect the entire dataset. Thus, plugging Equation 9 into Equation 3 and substituting D with Ω gives

$$\int_{\mathbb{R}^m} \kappa(\boldsymbol{\xi})\,\mathrm{d}^m\boldsymbol{\xi} = \int_{\Omega}\int_{\mathbb{R}^m} P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}^m\boldsymbol{\xi}\,\mathrm{d}^n\mathbf{x}\,. \tag{10}$$

Swapping the order of integration on the right-hand side leads to

$$\int_{\mathbb{R}^m} \kappa(\boldsymbol{\xi})\,\mathrm{d}^m\boldsymbol{\xi} = \int_{\mathbb{R}^m}\int_{\Omega} P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}^n\mathbf{x}\,\mathrm{d}^m\boldsymbol{\xi}\,, \tag{11}$$

which, by identifying the integrand, simplifies to

$$\kappa(\boldsymbol{\xi}) = \int_{\Omega} P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}^n\mathbf{x}\,. \tag{12}$$

From that, we construct the continuous parallel coordinates with uncertainty (UCPC) by substituting Equation 12 into Equation 7:

$$\varphi(\boldsymbol{\eta}) = \frac{\mathrm{d}M}{\mathrm{d}\eta_2} = \frac{\mathrm{d}}{\mathrm{d}\eta_2}\int_{\Psi_i}\int_{\Omega} P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}^n\mathbf{x}\,\mathrm{d}^2\boldsymbol{\xi}\,. \tag{13}$$

Notice that, in this equation, both $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are two-dimensional.

4 Implementation

4.1 Continuous Scatterplots With Uncertainty

So far, we have provided a generic mathematical model for computing continuous scatterplots with uncertainty for arbitrary dimensionality n of the spatial domain and dimensionality m of the data domain. However, in our implementation, we only focus on the common cases of $n = m = 2$ and $n = 3$, $m = 2$.

In our implementation, we assume that the uncertainty of the bivariate input field is given by a bivariate normal distribution, as illustrated in Figure 2b. Thus, in the spatial domain, at every position $\mathbf{x}$, there is a bivariate normal distribution with probability density function

$$P(\xi_1,\xi_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\; e^{-g/\left(2(1-\rho^2)\right)}\,, \tag{14}$$

where

$$g = \frac{(\xi_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(\xi_1-\mu_1)(\xi_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(\xi_2-\mu_2)^2}{\sigma_2^2}\,, \tag{15}$$

with $(\mu_1,\mu_2)^\top$ being the mean, $(\sigma_1,\sigma_2)^\top$ the standard deviation,

$$\rho = \mathrm{cor}(\xi_1,\xi_2) = \frac{V_{12}}{\sigma_1\sigma_2}, \quad |\rho| < 1\,, \tag{16}$$

being the correlation of ξ_1 and ξ_2, and V_{12} the covariance of ξ_1 and ξ_2 [23, 41]. We assume that the spatial domain is given in discretized form, as a grid consisting of triangles or rectangles in 2D, or tetrahedra or hexahedra in 3D, that the data (μ_1, μ_2, σ_1, σ_2) are given in node-based representation, and that ρ is constant.

According to Bursal [7], since the mean $(\mu_1,\mu_2)^\top$ has the same unit as the standard deviation $(\sigma_1,\sigma_2)^\top$, it is advantageous particularly for physical quantities to interpolate mean and standard deviation, instead of interpolating mean and variance. Schlegel et al. [38], on the other hand, state that if the standard deviation would be interpolated linearly, variance would not necessarily need to be interpolated quadratically. This is because the covariance ρ influences the effect of the interpolation of variance. However, according to Bursal, in some cases, e.g., in a Wiener process, the variance at a certain point in time is better interpolated linearly. Thus, in the applications of our concept provided in this paper, we interpolate the standard deviation, not the variance.

Further, the probability $\hat{P}$ of a bivariate normal distribution over a rectangular region Υ computes

$$\hat{P}_\Upsilon(\xi_1,\xi_2) = \iint_{\Upsilon} P(\xi_1,\xi_2)\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2\,. \tag{17}$$

4.1.1 Acceleration

Equation 17 can be simplified according to Didonato et al. [11], using a linear transformation with

$$w = \frac{\frac{\xi_1-\mu_1}{\sigma_1} - \rho\,\frac{\xi_2-\mu_2}{\sigma_2}}{\sqrt{1-\rho^2}}\,, \quad \text{and} \quad z = \frac{\xi_2-\mu_2}{\sigma_2}\,. \tag{18}$$

Equation 17 can then be rewritten to

$$\begin{aligned}\hat{P}_\Upsilon(\xi_1,\xi_2) = \hat{F}_\Pi(w,z) &= \iint_{\Pi} F(w,z)\,\mathrm{d}z\,\mathrm{d}w \\ &= \iint_{\Pi} \frac{1}{2\pi} e^{-\frac{1}{2}\left(w^2+z^2\right)}\,\mathrm{d}z\,\mathrm{d}w\,.\end{aligned} \tag{19}$$

$F(w,z)$ represents a probability density function of a bivariate normal distribution with: $\rho = 0$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\mu_1 = 0$, and $\mu_2 = 0$.

This means, if we select an infinitesimal rectangular range $[\xi_1, \xi_1 + \mathrm{d}\xi_1] \times [\xi_2, \xi_2 + \mathrm{d}\xi_2]$ at $(\xi_1,\xi_2)^\top$, based on Equation 19, the probability of the bivariate normal distribution over this range computes

$$P(\xi_1,\xi_2)\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2 = F(w,z)\,\mathrm{d}z\,\mathrm{d}w\,. \tag{20}$$

Based on Equation 17, the probability $\hat{P}_\Upsilon(\xi_1,\xi_2)$ in a certain (not uncertain) rectangular region $\Upsilon\colon [\min^\Upsilon\xi_1, \max^\Upsilon\xi_1] \times [\min^\Upsilon\xi_2, \max^\Upsilon\xi_2]$, with $\min^A b := \min_A b$, and $\max^A b := \max_A b$, can be computed as

$$\begin{aligned}\hat{P}_\Upsilon(\xi_1,\xi_2) &= \iint_{\Upsilon} P(\xi_1,\xi_2)\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2 \\ &= \int_{\min^\Upsilon\xi_2}^{\max^\Upsilon\xi_2}\int_{\min^\Upsilon\xi_1}^{\max^\Upsilon\xi_1} P(\xi_1,\xi_2)\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2\,.\end{aligned} \tag{21}$$

If $\rho = 0$, the rectangular region Υ corresponds to the rectangular region $\Pi\colon [w_{\min}, w_{\max}] \times [z_{\min}, z_{\max}]$, computed from Equation 18. Based on Equation 19, the probability $\hat{F}_\Pi(w,z)$ of region Π can be computed by

$$\begin{aligned}\hat{P}_\Upsilon(\xi_1,\xi_2) = \hat{F}_\Pi(w,z) &= \int_{w_{\min}}^{w_{\max}}\int_{z_{\min}}^{z_{\max}} F(w,z)\,\mathrm{d}z\,\mathrm{d}w \\ &= \int_{w_{\min}}^{w_{\max}}\int_{z_{\min}}^{z_{\max}} \frac{1}{2\pi} e^{\frac{-\left(w^2+z^2\right)}{2}}\,\mathrm{d}z\,\mathrm{d}w\,.\end{aligned} \tag{22}$$

However, if $\rho \neq 0$, the corresponding region Π to compute $\hat{F}_\Pi(w,z)$ will be a parallelogram. But if Π is small enough, one can still compute the probability $\hat{F}_\Pi(w,z)$ by approximating this region Π as rectangular, which still enables the application of Equation 22. Based on this approximation, we finally derive Equation 17 to

$$\begin{aligned}\hat{P}_\Upsilon(\xi_1,\xi_2) \approx &\int_{-\infty}^{w_{\min}}\int_{-\infty}^{z_{\min}} F(w,z)\,\mathrm{d}z\,\mathrm{d}w + \int_{-\infty}^{w_{\max}}\int_{-\infty}^{z_{\max}} F(w,z)\,\mathrm{d}z\,\mathrm{d}w \\ &- \int_{-\infty}^{w_{\min}}\int_{-\infty}^{z_{\max}} F(w,z)\,\mathrm{d}z\,\mathrm{d}w - \int_{-\infty}^{w_{\max}}\int_{-\infty}^{z_{\min}} F(w,z)\,\mathrm{d}z\,\mathrm{d}w\,,\end{aligned} \tag{23}$$

Algorithm 1 UCSP: Sampling-based approach

Require: *uncertain bivariate data* $\tilde{\boldsymbol{\tau}}$ *, defined on data grid* G
procedure UCSP_SAMPLING($\tilde{\boldsymbol{\tau}}$)
  for all *pixel* $\Upsilon \in \Psi$ **do**
    $M(\Upsilon) = 0$
    $\Upsilon.range = [\min^{\Upsilon}\xi_1, \max^{\Upsilon}\xi_1] \times [\min^{\Upsilon}\xi_2, \max^{\Upsilon}\xi_2]$
    for all *(2D or 3D) sampling pixel/voxel* $sc \in \Omega$ **do**
      for all *cell* $c \in G$ **do**
        if $sc.random \in c$ **then**
          interpolate $\mu_1, \mu_2, \sigma_1, \sigma_2$ *at* $sc.random$
          $[w_{\min}, w_{\max}] \times [z_{\min}, z_{\max}] \leftarrow \Upsilon.range$
          $\hat{P}_{\Upsilon}(\xi_1, \xi_2) \leftarrow \int_{w_{\min}}^{w_{\max}} \int_{z_{\min}}^{z_{\max}} F(w,z)\,\mathrm{d}z\,\mathrm{d}w$
          $M(\Upsilon) = M(\Upsilon) + \hat{P}_{\Upsilon}(\xi_1, \xi_2) \cdot \lambda(sc)$
          mark sc in Ω: $sc.\hat{P} \leftarrow \hat{P}_{\Upsilon}(\xi_1, \xi_2)$
          break
        end if
      end for
    end for
    update UCSP by marking Υ : $\kappa(\Upsilon) \leftarrow M(\Upsilon)/\lambda(\Upsilon)$
  end for
end procedure

which can be computed based on Donnelly's algorithm [12,30], which obtains the lower-left tail values of the bivariate normal distribution. Equation 23 enables accelerated computation of $\hat{P}_{\Upsilon}(\xi_1, \xi_2)$. For further acceleration, values that exceed three standard deviations away from the mean can be omitted.

Next, let us present two approaches (Sections 4.1.2 and 4.1.3) for computing continuous scatterplots with uncertainty.

4.1.2 Sampling-Based Approach

Our first approach is based on massive sampling, why we also denote it the brute-force approach. Due to its simplicity, we also used it for providing a ground-truth for our more complex, but also more efficient approach described in Section 4.1.3. Based on Equation 12, the mass of a pixel (or cell) Υ of a 2D uncertain continuous scatterplot (UCSP) can be formulated as

$$\begin{aligned} M(\Upsilon) = \iint_{\Upsilon} \kappa(\boldsymbol{\xi})\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2 &= \iint_{\Upsilon} \int_{\Omega} P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}^n\mathbf{x}\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2 \\ &= \int_{\Omega} \iint_{\Upsilon} P(\mathbf{x};\boldsymbol{\xi})\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2\,\mathrm{d}^n\mathbf{x} \\ &= \int_{\Omega} \hat{P}_{\Upsilon}(\mathbf{x};\xi_1,\xi_2)\,\mathrm{d}^n\mathbf{x}, \end{aligned} \tag{24}$$

where $\hat{P}_{\Upsilon}(\mathbf{x};\xi_1,\xi_2)$ is the probability that values at $\mathbf{x}$ lie within $[\min^{\Upsilon}\xi_1, \max^{\Upsilon}\xi_1] \times [\min^{\Upsilon}\xi_2, \max^{\Upsilon}\xi_2]$. Then, the density of cell Υ is

$$\kappa(\Upsilon) = \frac{M(\Upsilon)}{\lambda(\Upsilon)}, \tag{25}$$

with $\lambda(\Upsilon)$ being the area of cell Υ. In this approach, many samples $\mathbf{x}$ are required to obtain the uncertain continuous scatterplot at sufficient accuracy. On the other hand, in our approach, the input grid can be non-simplicial, whereas the previous approaches by Bachthaler and Weiskopf to compute continuous scatterplots without uncertainty allowed only simplicial grids [3], or allowed for non-simplicial grids by employing subdivision in the data domain [4], which, however, would make it hard to include uncertainty. The sampling-based approach is detailed in Algorithm 1, with *sc.random* giving a random point within sampling cell *sc*, and $\lambda(sc)$ measuring the area/volume of *sc*.

Since every pixel Υ can be computed independently, this algorithm lends itself well for parallelization. To avoid systematic error, we employ a random sampling strategy, i.e., we use Monte Carlo integration for computing the density $\kappa(\Upsilon)$. Rectilinear input grids substantially accelerate finding the grid cell c of the input grid that contains a respective random position in the spatial domain, and also accelerate interpolation of means and standard deviations. A further way to accelerate the approach is to omit those random samples $\mathbf{x}$, that lie in data grid cells c whose pixel Υ is outside the enlarged bounding box of $\boldsymbol{\tau}(c)$, defined by three times the largest standard deviation of all nodes of c in the data domain Ψ. However, since a high sampling density in the spatial domain is required for good results, non-rectilinear input grids tend to cause high computational cost using our sampling-based approach.

Notice that this approach can also be used for computing continuous scatterplots without uncertainty, since input data without uncertainty is a special case in our model with $\sigma_1 = \sigma_2 = 0$.

4.1.3 Convolution-Based Approach

In cases where the computational cost to compute uncertain continuous scatterplots using the sampling-based approach is high, we provide here an alternative approach that does not require massive sampling. Instead, it is based on convolution. Our approach builds on that by Bachthaler and Weiskopf [3], and is thus only applicable to bivariate data defined on simplicial grids. If the given grids are not simplicial and one still wants a fast preview, one can decompose the input cells to simplices, however, including the approximation error induced by such decomposition. For exact results of non-simplicial grids, our sampling-based approach would need to be used instead.

Bachthaler and Weiskopf presented how to compute the density κ for a given mapping $\boldsymbol{\tau}$. Since $\boldsymbol{\tau}$ satisfies mass conservation, for an input grid consisting of triangles T, the contribution of a triangle $\boldsymbol{\tau}(T)$ in the data domain to the total density at a point $(\xi_1,\xi_2)^{\top} \in \boldsymbol{\tau}(T)$ is

$$\kappa(\xi_1,\xi_2;\boldsymbol{\tau}(T)) = \frac{\lambda(T)}{\lambda(\boldsymbol{\tau}(T))}, \tag{26}$$

with triangle area $\lambda(T)$ in the spatial domain, and triangle area $\lambda(\boldsymbol{\tau}(T))$ in the data domain.

If the input grid consists of tetrahedra Te, every tetrahedron $\boldsymbol{\tau}(Te)$ is projected to the 2D data domain according to Bachthaler and Weiskopf's approach [3]. There, the contribution of a tetrahedron $\boldsymbol{\tau}(Te)$ to the total density at a point $(\xi_1,\xi_2)^{\top} \in \boldsymbol{\tau}(Te)$ depends on the Euclidean distance between the front face ($\boldsymbol{\tau}_f^{-1}$) and back face ($\boldsymbol{\tau}_b^{-1}$) of $\boldsymbol{\tau}^{-1}(\xi_1,\xi_2;\boldsymbol{\tau}(Te))$ in Te in the spatial domain and the gradient of ξ_1 and ξ_2 in Te in the spatial domain, which formulates as

$$\kappa(\xi_1,\xi_2;\boldsymbol{\tau}(Te)) = \frac{\left\| \boldsymbol{\tau}_f^{-1}(\xi_1,\xi_2;\boldsymbol{\tau}(Te)) - \boldsymbol{\tau}_b^{-1}(\xi_1,\xi_2;\boldsymbol{\tau}(Te)) \right\|}{\left\| \frac{\partial \xi_1}{\partial \mathbf{x}} \times \frac{\partial \xi_2}{\partial \mathbf{x}} \right\|}. \tag{27}$$

The constant gradient $\nabla g_i = \partial \xi_i / \partial \mathbf{x}$ of each variable ξ_i in a tetrahedron is computed from its four nodes $\mathbf{x}_i$ with $i \in 1,\ldots,4$, using the averaging-based approach by Correa et al. [9] defined as

$$\begin{bmatrix} (\mathbf{x}_2-\mathbf{x}_1)^{\top} \\ (\mathbf{x}_3-\mathbf{x}_1)^{\top} \\ (\mathbf{x}_4-\mathbf{x}_1)^{\top} \end{bmatrix} \nabla g_i = \begin{bmatrix} \boldsymbol{\tau}_{\xi_i}(\mathbf{x}_2) - \boldsymbol{\tau}_{\xi_i}(\mathbf{x}_1) \\ \boldsymbol{\tau}_{\xi_i}(\mathbf{x}_3) - \boldsymbol{\tau}_{\xi_i}(\mathbf{x}_1) \\ \boldsymbol{\tau}_{\xi_i}(\mathbf{x}_4) - \boldsymbol{\tau}_{\xi_i}(\mathbf{x}_1) \end{bmatrix}. \tag{28}$$

The left-hand side consists of a 3×3 matrix times a 3-vector, and the right-hand side is a 3-vector of scalar differentials of the variable ξ_i. Thus, ∇g_i can be directly solved for nondegenerate tetrahedra.

According to Bachthaler and Weiskopf [3], the density at a point $(\xi_1,\xi_2)^{\top}$ in Ψ can be constructed by linear superposition of the contributions $\kappa(\xi_1,\xi_2;\boldsymbol{\tau}(c))$ from all cells $c \in G$, with input grid G, based on the linear model from Equation 3 as

$$\kappa(\xi_1,\xi_2;\boldsymbol{\tau}(\Omega)) = \sum_{c \in G} \kappa(\xi_1,\xi_2;\boldsymbol{\tau}(c)). \tag{29}$$

If no uncertainty is involved, every point in the spatial domain links to only one point with a fixed position in the data domain. However, if there is uncertainty, such a point represents a density according to a

Algorithm 2 UCSP: Convolution-based approach (triangle grid)

Require: *uncertain bivariate data* $\tilde{\boldsymbol{\tau}}$ *, defined on triangle grid* G
1: **procedure** UCSP_CONVOLUTIONTRI($\tilde{\boldsymbol{\tau}}$)
2: **for all** *pixel* $\Upsilon \in \Psi$ **do**
3: **for all** *triangle* $T \in G$ **do**
4: **if** $\Upsilon.center \in \boldsymbol{\tau}(T)$ **then**
5: *interpolate* σ_1, σ_2 *at* $\boldsymbol{\tau}^{-1}(\Upsilon.center; \boldsymbol{\tau}(T))$
6: $\kappa \leftarrow \frac{\lambda(T)}{\lambda(\boldsymbol{\tau}(T))}$
7: *build convolution kernel* k *based on* σ_1, σ_2
8: *convolve* κ *using* k
9: *update UCSP by accumulating this result*
10: **end if**
11: **end for**
12: **end for**
13: **end procedure**

Algorithm 3 UCSP: Convolution-based approach (tetrahedral grid)

Require: *uncertain bivariate data* $\tilde{\boldsymbol{\tau}}$ *, defined on tetrahedral grid* G
1: **procedure** UCSP_CONVOLUTIONTET($\tilde{\boldsymbol{\tau}}$)
2: **for all** *pixel* $\Upsilon \in \Psi$ **do**
3: **for all** *tetrahedron* $Te \in G$ **do**
4: **if** $\Upsilon.center \in \boldsymbol{\tau}(Te)$ **then**
5: *interpolate* $\sigma_{1_f}, \sigma_{2_f}$ *at front* $\boldsymbol{\tau}_f^{-1}(\Upsilon.center; \boldsymbol{\tau}(Te))$
6: *interpolate* $\sigma_{1_b}, \sigma_{2_b}$ *at back* $\boldsymbol{\tau}_b^{-1}(\Upsilon.center; \boldsymbol{\tau}(Te))$
7: $\kappa \leftarrow \frac{\|\boldsymbol{\tau}_f^{-1}(\Upsilon.center; \boldsymbol{\tau}(Te)) - \boldsymbol{\tau}_b^{-1}(\Upsilon.center; \boldsymbol{\tau}(Te))\|}{\|\frac{\partial \xi_1}{\partial \mathbf{x}} \times \frac{\partial \xi_2}{\partial \mathbf{x}}\|}$
8: *divide* κ *by number* $|M|$ *of used segments*
9: **for all** *segment* $m \in M$ *along* $\boldsymbol{\tau}_f^{-1} - \boldsymbol{\tau}_b^{-1}$ **do**
10: *interpolate* σ_1, σ_2 *at* $m.center$
11: *build kernel* k *from* σ_1, σ_2
12: *convolve* κ *using* k
13: *update UCSP by accumulating this result*
14: **end for**
15: **end if**
16: **end for**
17: **end for**
18: **end procedure**

probability distribution (in our case, the bivariate normal distribution) centered at the fixed position (Figure 2b). Thus, the role of uncertainty is to "blur" the density in the data domain, more precisely, to blur the density obtained without uncertainty in the data domain by a convolution kernel that corresponds to the bivariate normal distribution given by the uncertainty at that point in the spatial domain. Thus, a continuous scatterplot with uncertainty, $\kappa(\tilde{\boldsymbol{\tau}}(\Omega))$, defined on the entire domain, can be considered a superposition of many such "blurred" results, which can be formulated as

$$\kappa(\tilde{\boldsymbol{\tau}}(\Omega)) = \sum_{(\xi_1,\xi_2)^\top \in \boldsymbol{\tau}(\Omega)} \left(\sum_{c \in G} \kappa(\xi_1, \xi_2; \boldsymbol{\tau}(c)) * k \right), \tag{30}$$

with convolution operator $*$, and bivariate normal distribution kernel k constructed by interpolating mean $(\mu_1, \mu_2)^\top$ and standard deviation $(\sigma_1, \sigma_2)^\top$ in c for position $\boldsymbol{\tau}^{-1}(\xi_1, \xi_2; \boldsymbol{\tau}(c))$. See Algorithms 2 and 3 for details.

4.1.4 Uncertainty Indication

Inspired by gradient plots, we not only visualize the uncertain continuous scatterplots with color-coding (see, e.g., Figure 7b and 7f), but additionally display the outline of its certain (mean) counterpart by a superimposed white line (see, e.g., Figure 7d and 7h). This helps interpretation w.r.t. uncertainty [10].

4.2 Continuous Parallel Coordinates With Uncertainty

For continuous parallel coordinates with uncertainty, we focus on the two cases $n = 2, m \geq 2$ and $n = 3, m \geq 2$. In the computation, cases with $m > 2$ need to be split into independent cases with $n = m = 2$ and $n = 3$, $m = 2$. That means, the dimension of the data (sub)domain Ψ_i is 2 or decomposed to 2 (Section 3.1).

Constructing uncertain continuous parallel coordinates plots can be split into the construction of several independent 2D parallel coordinate systems, each computed from a 2D continuous scatterplot [19]. We base our approach on the *scattering* algorithm mentioned by Heinrich and Weiskopf [19]. Our resulting algorithm is almost identical to our sampling-based approach for obtaining uncertain continuous scatterplots (Section 4.1.2). The only difference is to replace Line 17 in Algorithm 1 with

$$\textit{draw } \mathbf{L}_{\boldsymbol{\eta}}^{\boldsymbol{\xi}}(\Upsilon) \textit{ with density } \alpha \textit{ according to } \kappa(\Upsilon)\,.$$

In our case, $\mathbf{L}_{\boldsymbol{\eta}}^{\boldsymbol{\xi}}(\Upsilon)$ is a *trapezoid* in the independent 2D parallel coordinates domain, with one of its bases on axis ξ_i spanning the interval $[\min^\Upsilon \xi_i, \max^\Upsilon \xi_i]$, and the other base on axis ξ_{i+1} spanning $[\min^\Upsilon \xi_{i+1}, \max^\Upsilon \xi_{i+1}]$. Based on mass conservation with respect to that trapezoid, the density α is given as

$$\alpha = \frac{2\kappa(\Upsilon)\lambda(\Upsilon)}{\left((\max^\Upsilon \xi_i - \min^\Upsilon \xi_i) + (\max^\Upsilon \xi_{i+1} - \min^\Upsilon \xi_{i+1})\right) \|\overline{\xi_i, \xi_{i+1}}\|}\,, \tag{31}$$

with $\|\overline{\xi_i, \xi_{i+1}}\|$ being the distance between the ξ_i-axis and the ξ_{i+1}-axis in η_1-direction, which defaults to 1 (Figure 2a). The point density φ in the parallel coordinates domain can be constructed implicitly by superposition of these trapezoids, which can be computed by accumulating the density of the trapezoids within a small interval along axis η_2 according to Equation 13. With this approach, one can also generate uncertain continuous parallel coordinates plots directly from existing uncertain continuous scatterplots.

4.3 Fibers with Uncertainty

Fibers are the preimage of $\boldsymbol{\xi}$ in the spatial domain (Section 3.1) and represent the intersection of the isocontours of the components of $\boldsymbol{\xi}$. If $\boldsymbol{\xi}$ is two-dimensional, fibers f of $(\xi_1, \xi_2)^\top$ can be formulated as

$$f(\xi_1, \xi_2) = \boldsymbol{\tau}^{-1}(\xi_1, \xi_2) = \boldsymbol{\tau}^{-1}(\xi_1) \cap \boldsymbol{\tau}^{-1}(\xi_2)\,. \tag{32}$$

Accordingly, fibers $\tilde{f}$ of the uncertain field $\tilde{\boldsymbol{\tau}}$ are

$$\tilde{f}(\xi_1, \xi_2) = \tilde{\boldsymbol{\tau}}^{-1}(\xi_1, \xi_2)\,. \tag{33}$$

In this case, however, a formulation by means of an intersection of sets, as in Equation 32 is not possible, as the preimages $\tilde{\boldsymbol{\tau}}^{-1}(\xi_1)$ and $\tilde{\boldsymbol{\tau}}^{-1}(\xi_2)$ are distributions, not sets, and thus $\tilde{\boldsymbol{\tau}}^{-1}(\xi_1, \xi_2)$ not a set in the spatial domain either, but a probability distribution $\hat{P}(\mathbf{x}; \xi_1, \xi_2)$ in $\mathbb{R}^2$. $\hat{P}(\mathbf{x}; \xi_1, \xi_2)$ represents the probability that values at $\mathbf{x}$ lie within $[\xi_1, \xi_1 + \mathrm{d}\xi_1] \times [\xi_2, \xi_2 + \mathrm{d}\xi_2]$. Thus, Algorithm 1 lends itself to computing fibers with uncertainty. For each position given in the 2D data domain, we can compute the corresponding uncertain fibers using Lines 3–16 of Algorithm 1 (Figure 8).

To enable effective exploration of (uncertain) continuous scatterplots, we additionally propose "range-fibers", which are the preimages of ranges of Ψ instead of points $\boldsymbol{\xi} \in \Psi$. Range fibers are simply obtained by increasing the pixel-size region in Lines 3–16 of Algorithm 1 to the range of interest, if $\rho = 0$. In case $\rho \neq 0$, Algorithm 1 can be easily adapted accordingly. Figure 9 demonstrates the utility of range fibers for analyzing the impact of uncertainty in the spatial domain.

5 Results

We demonstrate the utility and the properties of our approaches using uncertain bivariate and trivariate examples of increasing complexity. We start with a simple quad in 2D, to demonstrate our different approaches, the role of uncertainty, and implications (Section 5.1). We then proceed from $n = 2$ to $n = 3$, by using an uncertain bivariate cube

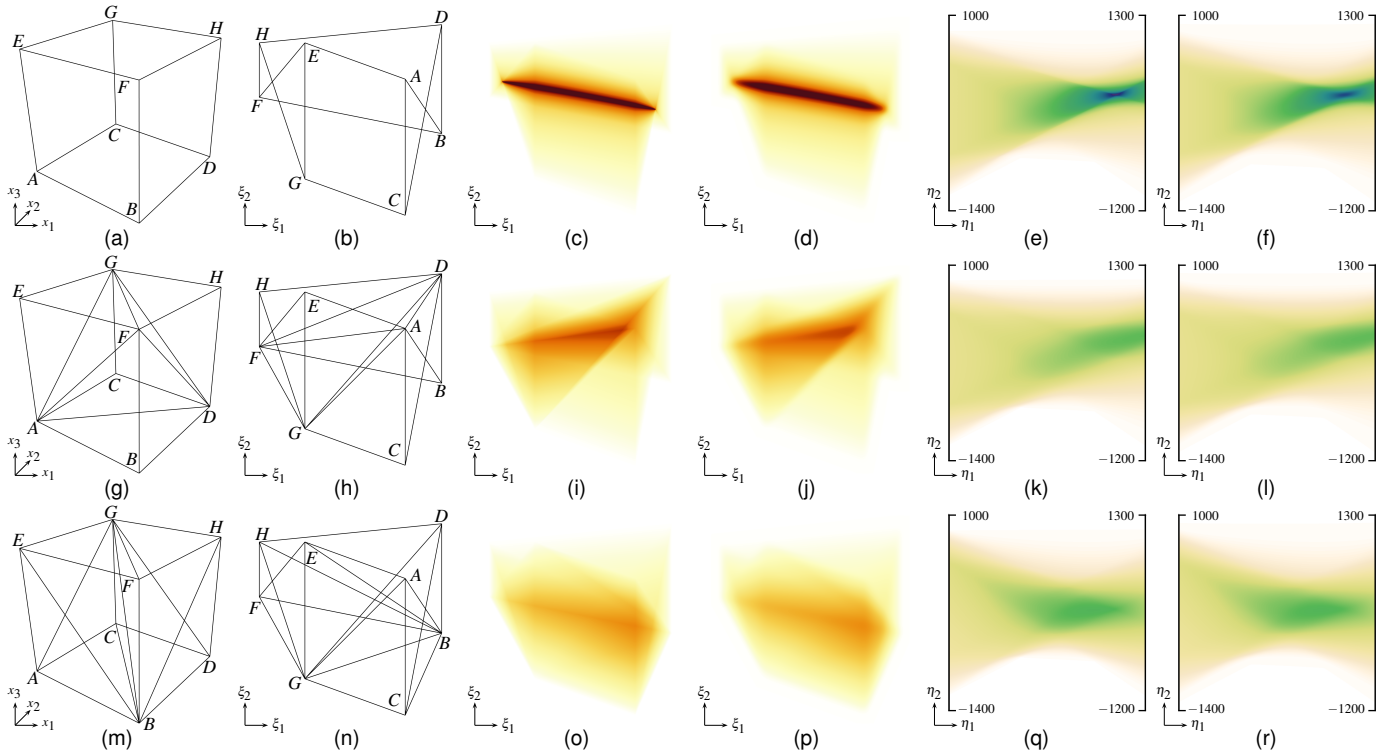

Fig. 3. Bivariate Cube example. Top row: original geometry; middle row: 5-tetrahedron subdivision; bottom row: 6-tetrahedron subdivision. Geometry in the spatial domain (first column), and geometry in the data domain (second column). Traditional continuous scatterplot (third column) and traditional continuous parallel coordinates (fifth column). Our uncertain continuous scatterplot (column four), and our uncertain parallel coordinates (column six) are similar to their traditional counterparts due to the low uncertainty. Notice, however, the rather strong impact of the subdivision.

(Section 5.2). Having established an understanding of our approaches, we then move to simulated meteorological data, also starting with a case with $n = 2$ (Section 5.3), followed by a case with $n = 3$ (Section 5.4). Finally, we demonstrate our approach for visualizing fibers in uncertain data, using the same examples for $n = 2$ and $n = 3$ (Section 5.5).

5.1 Bivariate Square

Let us start with a synthetic example with $m = n = 2$, consisting of a single square cell. Figure 1a shows the original grid, comprising four vertices A, B, C, and D, with respective bivariate values $(400,-1000)^\top$, $(800,1100)^\top$, $(-700,-600)^\top$, and $(-1200,900)^\top$. The uncertainty is set to be uniform and isotropic as $\sigma_1 = \sigma_2 = 100$, $\rho = 0$, for all four vertices. The upper row of Figure 1 shows the results for this case. To investigate the influence of subdivision into simplices, we subdivide the quad into two triangles, and show the respective results in the lower rows of Figure 1. As an additional alternative, we use the same data positions but swap the values of C and D, which we show in columns four, five, eight, and nine. The first column of Figure 1 shows the geometry of the examples, whereas the second and fourth column shows the traditional (without uncertainty) continuous scatterplot [3]. The third and fifth column shows our uncertain continuous scatterplot using our sampling-based approach. In analogy, column six and eight shows the traditional (without uncertainty) continuous parallel coordinates plot [19], whereas column seven and nine shows our uncertain continuous parallel coordinates plot. Notice that we do not show results for our convolution-based approach, since it produces results identical to those of the sampling-based approach for the triangulated dataset. The convolution-based approach differs only in terms of execution times, which are available in Table 1 and document its efficiency.

5.2 Bivariate Cube

We now increase from $n = 2$ to $n = 3$, by investigating a single synthetic bivariate cube cell. Figure 3a shows the basic grid, containing eight vertices A, B, C, D, E, F, G, and H, with respective bivariate values $(400,500)^\top$, $(800,-100)^\top$, $(400,-1000)^\top$, $(800,1100)^\top$, $(-700,900)^\top$, $(-1200,300)^\top$, $(-700,-600)^\top$, and $(-1200,900)^\top$. Also here, we set uncertainty to be uniform and isotropic, in this case $\sigma_1 = \sigma_2 = 30$, $\rho = 0$, for all eight vertices. The upper row of Figure 3 shows results for this case. To also investigate the influence of subdivision into simplices, a procedure still quite often applied at the cost of introduced deviations, we subdivide the cube into five tetrahedra and into six tetrahedra (middle and bottom row of Figure 3), respectively. The first column of Figure 3 shows the geometry in the spatial domain, whereas the second column shows their geometry in the data domain. The third column shows the traditional (without uncertainty) continuous scatterplot and the fifth column the traditional (without uncertainty) continuous parallel coordinates plot, whereas our uncertain continuous scatterplot and uncertain parallel coordinates plot is shown columns four and six. All continuous scatterplots are computed using our sampling-based approach.

By comparing the plots with uncertainty and without uncertainty, we can see that, since the uncertainty is set lower compared to the Bivariate Square dataset, the "blurring" effect is weaker. This effect is not only apparent in the continuous scatterplots with uncertainty but also in the continuous parallel coordinates plots with uncertainty.

5.3 2D ERA5 Hourly Data

ERA5 is the fifth-generation ECMWF atmospheric reanalysis of the global climate, which includes an uncertainty estimate that provides guidance on the accuracy of the products [1]. Here, we take the data on February 02, 2020, at 00:00, at pressure level 350 hPa. The original grid is a rectangular grid with resolution 719×360. We visualize specific humidity q, temperature t, and relative vorticity vo. The means of these variables in the spatial domain are shown in the top row of Figure 5, whereas the corresponding standard deviations are shown in the bottom row. The uncertainty estimation for ERA5 is based on physical considerations using an ensemble of data assimilations (EDA) system [1]. Since EDA does not account for correlated errors, $\rho = 0$.

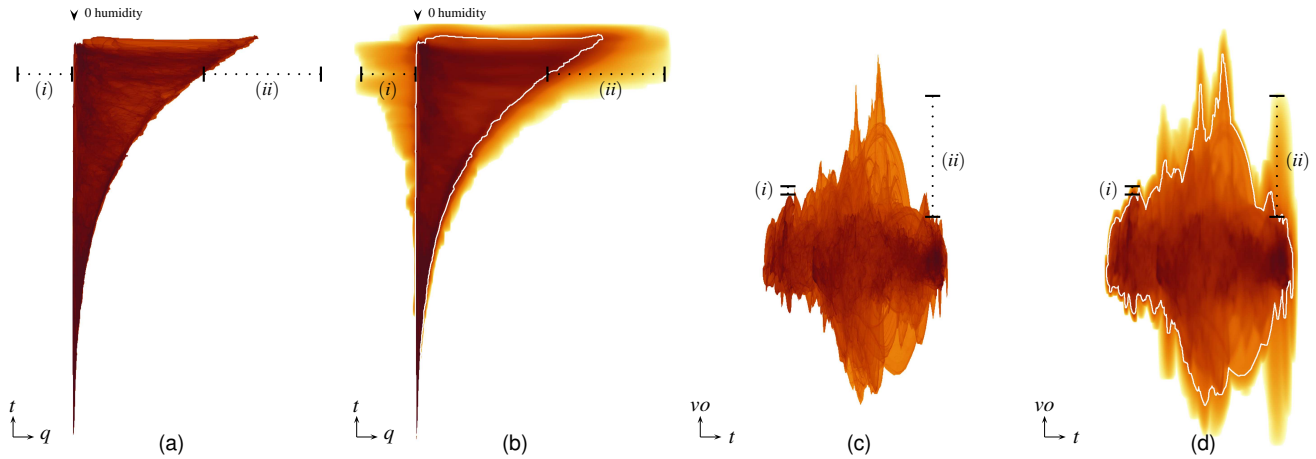

Fig. 4. Continuous scatterplots of 2D ERA5 Hourly dataset, of specific humidity q and temperature t, without (a) and with (b) uncertainty, and of temperature t and relative vorticity vo, without (c) and with (d) uncertainty. White outlines of certain (mean) continuous scatterplots for reference.

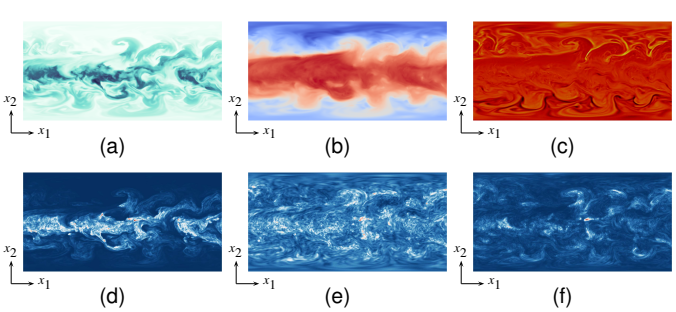

Fig. 5. 2D ERA5 Hourly dataset. Top row: mean values in the spatial domain. (a) Mean of specific humidity q, using a white (low) to dark blue (high) color scale. (b) Mean of temperature t, using a blue (low) to red (high) color scale. (c) Mean of relative vorticity vo, using a black (low) to yellow (high) color scale. Bottom row: column-wise corresponding standard deviations, also in the spatial domain, using a blue (low) over white (medium) to red (high) color scale.

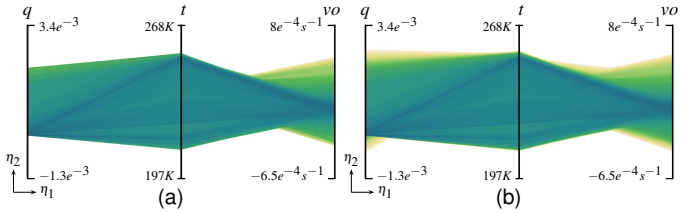

Fig. 6. Continuous parallel coordinates plot of 2D ERA5 Hourly dataset, showing quantities from Figure 4, without (a) and with (b) uncertainty.

Figure 4 shows the respective continuous scatterplots for specific humidity and temperature, without (a) and with (b) uncertainty, and correspondingly for temperature and relative vorticity (c)–(d). Figure 6, on the other hand, shows these quantities in continuous parallel coordinates plots without (a) and with (b) uncertainty.

As we can see in Figure 4d, the error range at (ii) is larger than the error range at (i), which means that if the temperature reaches this high level, the uncertainty of relative vorticity becomes high. In Figure 4b, we can observe how the uncertainty of humidity varies along with the temperature. One can also observe in Figure 4b that, though at the same temperature level, the uncertainties of humidity are different. As is visible at (i) and (ii), at the chosen temperature, the error range differs between the left and the right side. A closer inspection with respect to the zero-axis of humidity, reveals that the blurred region at (i) in Figure 4b represents unphysical negative humidity, which reveals limitations with the uncertainty-based representation of ensembles and could guide respective improvements of EDA.

5.4 3D ERA5 Hourly Data

Here, instead of taking the data from only one pressure level, as in Section 5.3, we take the entire data from 37 pressure levels with vertical coverage from 1000 hPa to 1 hPa. Thus, the resolution of this rectangular grid is $719 \times 360 \times 37$. The top two columns of Figure 7 show continuous scatterplots for specific humidity q and temperature t, and relative vorticity vo and temperature t—without uncertainty in columns one and three, and with uncertainty in columns two and four. In the bottom row, we provide a view on all three quantities using continuous parallel coordinates plots—in traditional form without uncertainty in columns one and three, and with uncertainty using our approach in columns two and four. The first two columns have been obtained by our convolution-based approach, whereas the last two columns with our sampling-based approach.

In Figure 7d and 7h, we mark some locations that we want to investigate and compare with respect to uncertainty. We can see that at these chosen locations, U_q, the error range of specific humidity, U_t, the error range of temperature, and U_{vo}, the error range of relative vorticity relate $U_q > U_{vo} > U_t$. From Figure 7b, we can see that in range (i), where the temperature is high, the error range on the left (low q) is larger than on the right (high q). This is opposite to the observation from Figure 4b, and is caused by the fact that in this case, all pressure levels have been considered. Since the blurred region caused by uncertainty reaches even more negative values of humidity as in Figure 4, we conclude that the unphysical issue with the uncertainty-based representation of ensembles is even stronger away from medium heights over the ground in this model.

5.5 Fibers in ERA5 Hourly Data

As a final example, we investigate fibers in uncertain 2D and 3D bivariate fields. Our results have been obtained by the approach presented in Section 4.3. We select four sample positions in the data domain of the 2D and 3D ERA5 Hourly datasets, depicted in the first column of Figure 8. The fibers without uncertainty corresponding to these sample positions are shown in the second column of Figure 8. These have been generated using the topology toolkit [5, 40]. Compare the uncertain fibers obtained with our method, shown in the third column.

Our uncertain fibers in Figure 8c and 8f, reveal regions around the equator which have high uncertainty in relative vorticity, and regions near the Arctic that exhibit low uncertainty in relative vorticity.

From the zoomed views in Figure 8j and 8n, we can see that for the sample position where the uncertainty is larger, the fibers with uncertainty are less opaque in the volume rendering, which means the probabilities $\hat{P}$ are lower. In those zoomed views, one can also see the structure of varying $\hat{P}$ along the uncertain fibers.

Finally, we chose three regions in uncertainty-dominated ranges (located outside of the white mean outline), with two boxes (red and green) in regions where the specific humidity is negative (Figure 9a),

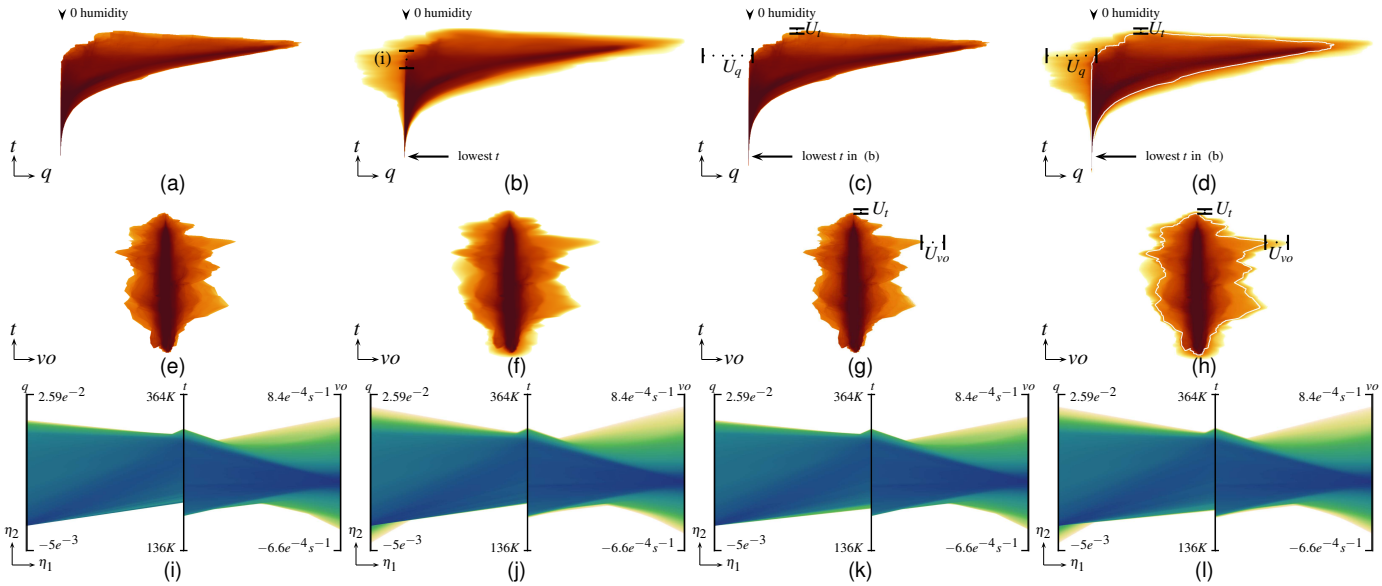

Fig. 7. 3D ERA5 Hourly dataset. First column: continuous scatterplots and continuous parallel coordinates plot without uncertainty, computed with the traditional approaches [3, 19]. Using our convolution-based approach, we obtain respective results with uncertainty, as shown in the second column. Using our sampling-based approach provides the results without and with uncertainty shown in the third and fourth column, respectively.

and visualize their range-fibers, consisting only of uncertainty-induced fibers (Figure 9b). We can locate the widespread red/green range-fibers, which caused the unrealistic humidity/temperature combinations within the red/green boxes. In comparison, the range-fiber belonging to the blue box in Figure 9a, is not apparently unphysical and is more compact.

6 Discussion

Let us first discuss the performance measurements, provided in Table 1. For the sampling-based approach, we subdivided every rectangular cell of the data grid uniformly at subdivision resolution R_c, and in each resulting subcell, we chose a random position to interpolate the uncertain data in terms of mean and standard deviation.

In some cases, the convolution-based approach is slower than the sampling-based approach, which results from the fact that Algorithms 2 and 3 cannot be run on the GPU in parallel in our implementation. That is, we cannot update the result in Ψ in parallel due to the read-write lock on graphics memory. For this reason, in our CUDA implementation, instead of convolving every pixel's certain density according to the uncertainty and accumulating the results, we evaluate the accumulated influence of all blurred results to a given pixel. For this reason, the runtime of the implementation is higher, in some cases.

In our evaluation, we focused on uncertain continuous scatterplots instead of uncertain continuous parallel coordinates plots, because uncertain continuous parallel coordinates plots tend to suffer from occlusion during the projection process, and therefore uncertainty structure is typically better revealed in the continuous scatterplots.

We observe that not only in simple datasets the decomposition into simplices affects the analysis, as shown in Figures 1 and 3. As is revealed in Figure 7 at the example of temperature, its span is reduced by the decomposition, and therefore we strongly suggest to use the sampling approach in case of bivariate data given on rectilinear grids.

For the case $m = 3$, our implementation can be straightforwardly extended from 2D to 3D by using a trivariate Gaussian distribution. In cases where the uncertainty is not given by a parametric model but, e.g., given by ensembles, our uncertain continuous scatterplots, parallel coordinates, and (range-)fibers can be considered a superposition of certain continuous scatterplots, parallel coordinates, and (range-)fibers computed from these ensembles. If, on the other hand, the uncertainty model is parametric but based on other probability density functions, our approach can be easily adapted, provided that the parameters can be interpolated in the spatial domain.

Table 1. Performance results, evaluating the method (continuous scatterplots (CSP) and continuous parallel coordinates plots (CPC)), different choices for the subdivision resolution R_c for each cell of the original data grid, influence of the resolution R_{CSP} of the resulting CSP, and the resolution R_{CPC} of the resulting CPC. All measurements and results were obtained on a GeForce GTX 1070 GPU with 8 GB VRAM and CUDA 8.

dataset		2D ERA5 Hourly				3D ERA5 Hourly			
variables		q and t		t and vo		q and t		t and vo	
uncertainty		without	with	without	with	without	with	without	with
1. CSP (sampling)	R_c	40×40	40×40	40×40	40×40	20×20×20	10×10×10	20×20×20	10×10×10
	R_{CSP}	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k
	Time	12.50 s	1837.43 s	14.40 s	2123.24 s	1574.60 s	27350.00 s	2242.40 s	45216.30 s
2. CSP (convolution)	R_{CSP}	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k
	Time	36.56 s	91.36 s	38.75 s	87.18 s	1604.20 s	10191.40 s	4765.10 s	44370.80 s
3. Fibers	R_c	5×5	5×5	5×5	5×5	5×5×5	5×5×5	5×5×5	5×5×5
	Time	—	—	—	≈ 0.08 s	—	≈ 12.45 s	—	—
4. CPC (from 1.)	R_{CPC}	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k	1k×1k
	Time	110.35 s	120.16 s	112.70 s	118.20 s	113.94 s	119.78 s	133.07 s	139.10 s

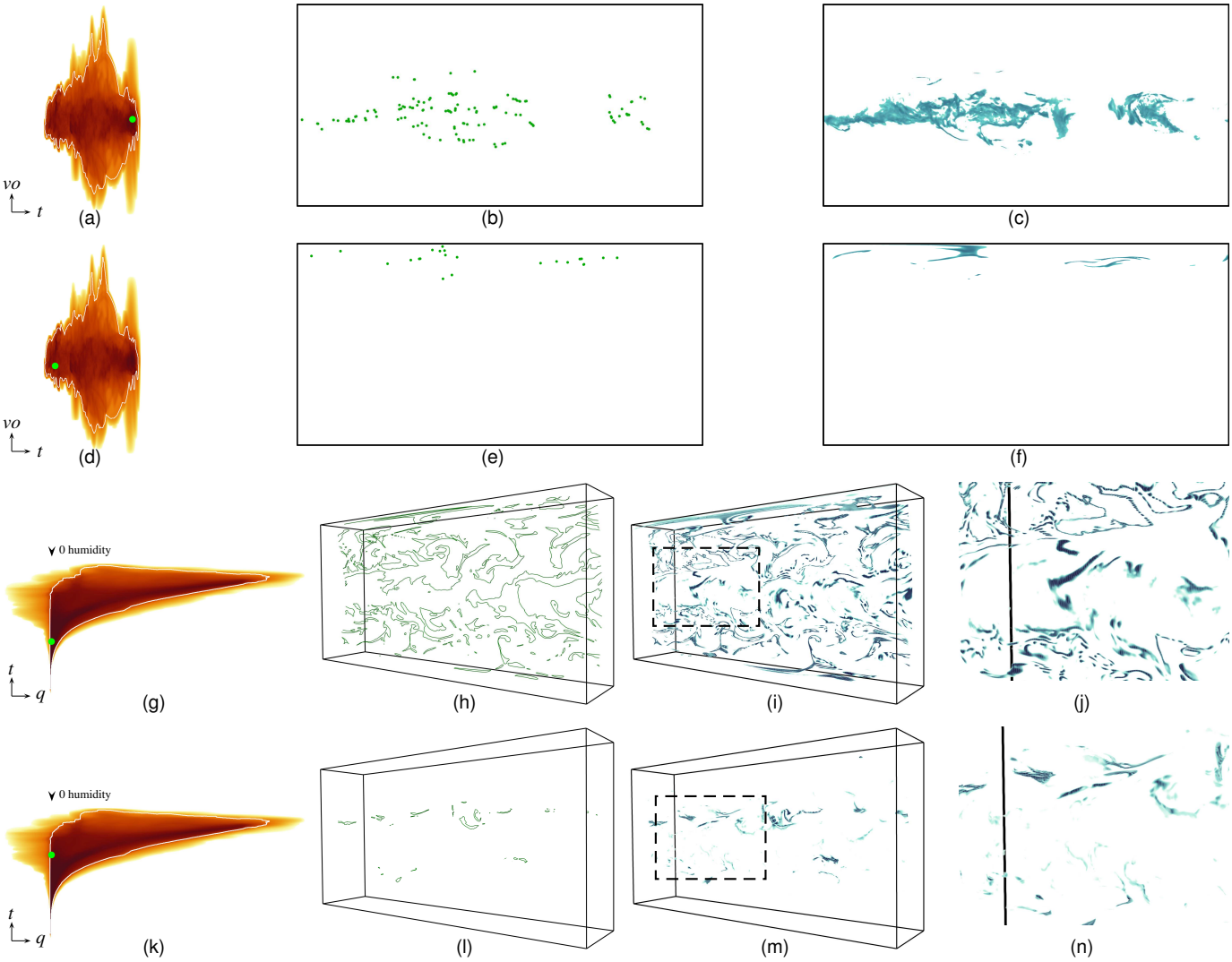

Fig. 8. Fibers in 2D ERA5 Hourly dataset (two top rows), and in 3D ERA5 Hourly dataset (two bottom rows). First column: four sample positions (green dots) in the data domain. Traditional fibers (without uncertainty) generated by these sample positions are shown in the second column. Fibers with uncertainty generated with our method are shown in the third column. Zoomed view of boxed region from (i) and (m) shown in (j) and (n). Fibers with uncertainty are color-coded by probability $\hat{P}$, using a white (low) to dark blue (high) color scale, revealing additional structure.

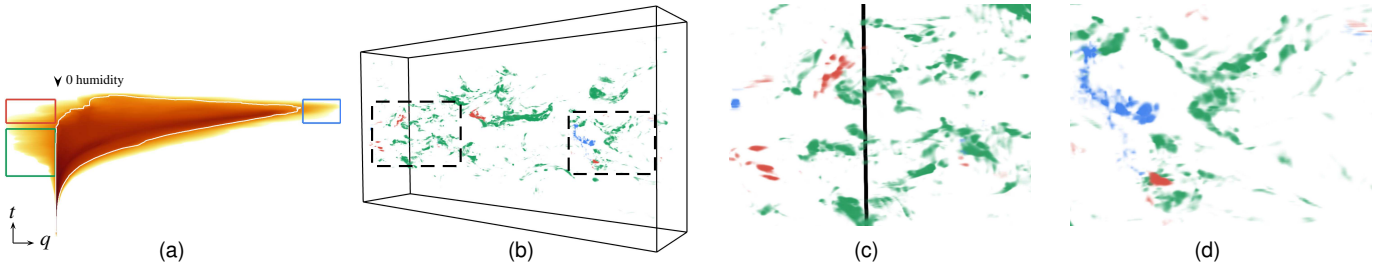

Fig. 9. Range-fibers in 3D ERA5 Hourly dataset. (a) Two selected regions (red, green) with negative specific humidity, and another region (blue) for comparison. Correspondingly colored uncertain range-fibers (b). Zoomed views of boxes in (b) reveal the compact distribution of blue part (c)–(d).

7 Conclusion

In this paper, we presented a mathematical model and two numerical approaches for extending continuous scatterplots to uncertain data. One of the numerical approaches served as ground truth for the development of the other, generally more efficient one. Based on that, we extended continuous parallel coordinates plots and (range-)fibers to uncertain data. We demonstrated the properties, the utility, and the interplay of our approaches using both synthetic and simulated data. Future work could research further acceleration, extension to cases with $m > 2$, and extension to different probability distributions.

Acknowledgments

We thank Kai Sdeo and the reviewers for their help. This work has been supported by DFG (Deutsche Forschungsgemeinschaft) grant GSC 220 in the German Universities Excellence Initiative at Heidelberg University, and the subproject A7 of the Transregional Collaborative Research Center SFB / TRR 165 "Waves to Weather" (www.wavestoweather.de) funded by the German Research Foundation (DFG).

REFERENCES

[1] ERA5 hourly data on pressure levels from 1979 to present. http://dx.doi.org/10.24381/cds.bd0915c6.

[2] S. Bachthaler, S. Frey, and D. Weiskopf. Poster: CUDA-accelerated continuous 2D scatterplots. 2009.

[3] S. Bachthaler and D. Weiskopf. Continuous scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1428–1435, 2008.

[4] S. Bachthaler and D. Weiskopf. Efficient and adaptive rendering of 2D continuous scatterplots. *Computer Graphics Forum*, 28(3):743–750, 2009.

[5] T. Bin Masood, J. Budin, M. Falk, G. Favelier, C. Garth, C. Gueunet, P. Guillou, L. Hofmann, P. Hristov, A. Kamakshidasan, C. Kappe, P. Klacansky, P. Laurin, J. Levine, J. Lukasczyk, D. Sakurai, M. Soler, P. Steneteg, J. Tierny, W. Usher, J. Vidal, and M. Wozniak. An Overview of the Topology ToolKit. In *Proceedings of Topological Methods in Data Analysis and Visualization*, 2019.

[6] G.-P. Bonneau, H.-C. Hege, C. Johnson, M. M. Oliveira, K. Potter, and P. Rheingans. Overview and State-of-the-Art of Uncertainty Visualization. In *Scientific Visualization: Uncertainty, Multifield, Biomedical, Scalable*, vol. 17 of *Mathematics and Visualization*, pp. 3–27. Springer, 2014.

[7] F. H. Bursal. On interpolating between probability distributions. *Applied Mathematics and Computation*, 77(2):213–244, 1996.

[8] H. Carr, Z. Geng, J. Tierny, A. Chattopadhyay, and A. Knoll. Fiber surfaces: Generalizing isosurfaces to bivariate data. *Computer Graphics Forum*, 34(3):241–250, 2015.

[9] C. D. Correa, R. Hero, and K. Ma. A comparison of gradient estimation methods for volume rendering on unstructured meshes. *IEEE Transactions on Visualization and Computer Graphics*, 17(3):305–319, 2011.

[10] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014.

[11] A. R. Didonato, M. P. Jarnagin, Jr., and R. K. Hageman. Computation of the integral of the bivariate normal distribution over convex polygons. *SIAM Journal on Scientific and Statistical Computing*, 1(2):179–186, 1980.

[12] T. G. Donnelly. Algorithm 462: Bivariate normal distribution. *Communications of the ACM*, 16(10):638, 1973.

[13] G. Favelier, N. Faraj, B. Summa, and J. Tierny. Persistence atlas for critical point variability in ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1152–1162, 2019.

[14] D. Feng, L. Kwock, Y. Lee, and R. M. T. II. Matching visual saliency to confidence in plots of uncertain data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):980–989, 2010.

[15] F. Ferstl, K. Brger, and R. Westermann. Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):767–776, 2016.

[16] F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Visual analysis of spatial variability and global correlations in ensembles of isocontours. *Computer Graphics Forum*, 35(3):221–230, 2016.

[17] D. Günther, J. Salmon, and J. Tierny. Mandatory critical points of 2D uncertain scalar fields. *Computer Graphics Forum*, 33(3):31–40, 2014.

[18] J. Heinrich, S. Bachthaler, and D. Weiskopf. Progressive splatting of continuous scatterplots and parallel coordinates. *Computer Graphics Forum*, 30(3):653–662, 2011.

[19] J. Heinrich and D. Weiskopf. Continuous parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1531–1538, 2009.

[20] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.

[21] A. Inselberg. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer, 2009.

[22] A. Inselberg and B. Dimsdale. Multidimensional lines II: Proximity and applications. *SIAM Journal on Applied Mathematics*, 54(2):578–596, 1994.

[23] J. F. Kenney and E. S. Keeping. Mathematics of Statistics, Part 2, 2nd ed. pp. 92, 202–205, 1951.

[24] P. Klacansky, J. Tierny, H. Carr, and Z. Geng. Fast and exact fiber surfaces for tetrahedral meshes. *IEEE Transactions on Visualization and Computer Graphics*, 23(7):1782–1795, 2017.

[25] D. J. Lehmann and H. Theisel. Discontinuities in continuous scatter plots. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1291–1300, 2010.

[26] D. J. Lehmann and H. Theisel. Features in continuous parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1912–1921, 2011.

[27] A. M. 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*, 32(3):139–160, 2005.

[28] M. Otto, T. Germer, H.-C. Hege, and H. Theisel. Uncertain 2D vector field topology. *Computer Graphics Forum*, 29(2):347–356, 2010.

[29] M. Otto, T. Germer, and H. Theisel. Uncertain topology of 3D vector fields. In *Proceedings of 2011 IEEE Pacific Visualization Symposium*, pp. 67–74, 2011.

[30] D. B. Owen. Tables for computing bivariate normal probabilities. *Ann. Math. Statist.*, 27(4):1075–1090, 1956.

[31] C. Petz, K. Pthkow, and H.-C. Hege. Probabilistic local features in uncertain vector fields with spatial correlation. *Computer Graphics Forum*, 31(3):1045–1054, 2012.

[32] T. Pfaffelmoser, M. Mihai, and R. Westermann. Visualizing the variability of gradients in uncertain 2D scalar fields. *IEEE Transactions on Visualization and Computer Graphics*, 19(11):1948–1961, 2013.

[33] T. Pfaffelmoser, M. Reitinger, and R. Westermann. Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *Computer Graphics Forum*, 30(3):951–960, 2011.

[34] T. Pfaffelmoser and R. Westermann. Visualization of global correlation structures in uncertain 2D scalar fields. *Computer Graphics Forum*, 31(3):1025–1034, 2012.

[35] K. Pöthkow and H. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1393–1406, 2011.

[36] K. Pöthkow and H.-C. Hege. Nonparametric models for uncertainty visualization. *Computer Graphics Forum*, 32(3):131–140, 2013.

[37] K. Pöthkow, C. Petz, and H.-C. Hege. Approximate level-crossing probabilities for interactive visualization of uncertain isocontours. *International Journal for Uncertainty Quantification*, 3, 2012.

[38] S. Schlegel, N. Korn, and G. Scheuermann. On the interpolation of data with normally distributed uncertainty for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2305–2314, 2012.

[39] J. Tierny and H. Carr. Jacobi fiber surfaces for bivariate Reeb space computation. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):960–969, 2017.

[40] J. Tierny, G. Favelier, J. A. Levine, C. Gueunet, and M. Michaux. The topology toolkit. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):832–842, 2018.

[41] E. T. Whittaker and G. Robinson. "Determination of the Constants in a Normal Frequency Distribution with Two Variables" and "The Frequencies of the Variables Taken Singly." in The Calculus of Observations: A Treatise on Numerical Mathematics, 4th ed. pp. 324–329, 1967.

[42] B. Zheng, B. Rieck, H. Leitte, and F. Sadlo. Visualization of equivalence in 2D bivariate fields. *Computer Graphics Forum*, 38(3):311–323, 2019.