

Kai Pöthkow Christoph Petz Britta Weber Hans-Christian Hege





Institut P', CNRS - Université de Poitiers - ENSMA, Poitiers, September 21, 2011

# Outline

- Elementary Motivation for Considering Uncertainties
- The Bigger Picture
- · Feature Extraction in Uncertain Data
- · Iso-Contours in Uncertain Data
- Roads to Sharp Formulations, Future of Uncertainty Visualization

Elementary Motivation	

# Numbers are Uncertain

All floating point numbers are afflicted with uncertainties

- → 'true values' are not known exactly
- Measurements values are uncertain

due to non-perfect instruments, intrinsic randomness,  $\dots$ 

Simulation results are uncertain

due to uncertain parameters, inaccurate models,

inaccurate numerics, rounding errors

Strictly spoken:

Numbers

without indication of **error bars or even PDF** are worthless!

(fortunately: error bounds are often implicitly available by foreknowledge)

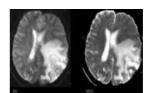
# Example: Uncertainties in Biomedicine

- Physical view
  - Quantitative science
  - No numbers without errors; some variables are random / considered as ...
  - Uncertainties need to be specified
- Biological view
  - Huge variability of biological systems
  - Large variability of biological states
  - Regulation: intervals instead numbers
  - Many relevant variables are inaccessible
- Technical view
  - Measurements are riddled with error

5

# Imagine, you are a Surgeon ...

- Patient: a human with a severe problem; his/her future well-being depends on you
- · You have a highly responsible task
- You have to infer information from images, take decisions and perform actions
- You have to **rely** on data acquisition data processing data visualization



Glioblastoma of the brain.

Displacement of nerves by the tumor.



## Uncertainty in Visualization (More General)

All data with continuous range are affected by uncertainty.

- Conclusions have to be drawn from uncertain information
- · This is the rule, not an exception
- · Error estimation should be ubiquitous
  - → Visualization tools should show
    - which information is reliable
    - which information is uncertain

7

# Data Analysis Point of View

In data analysis we tend

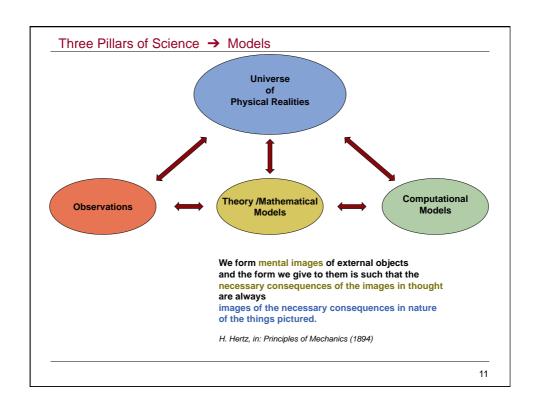
- to reconstruct fine details at the resolution limit and at marginal SNRs
- · to extract complex features at the information limit
- · to depict untrustworthy information...

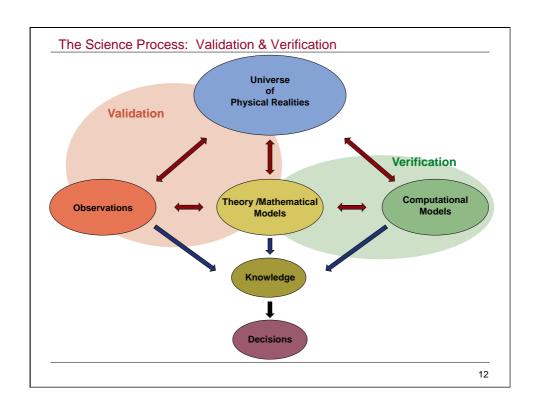
Visualization methods should

- sense these limits
- · discern which information is reliable / which not
- convey visually the degree of trustworthiness of information

# Previous Work (Few Samples ONLY) Statistical graphics: well developed Few 2D and 3D visualization methods exist However: most 2D and 3D visualization methods still do not consider uncertainty A. Pang. C. Wittenbrink, S. Lodha

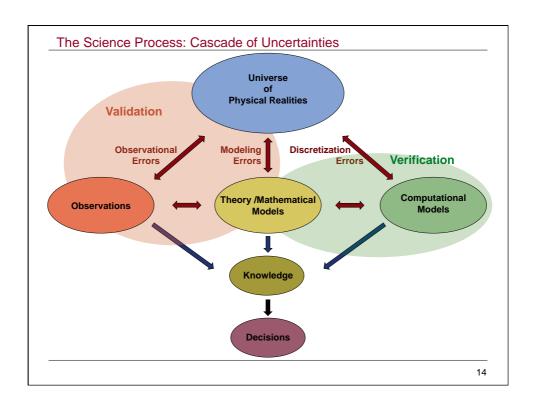
The Bigger Picture		





# Creation of Theory, Knowledge

- David Hume (1711 1776)
  - Problem of induction: theories can not be validated
- Karl Popper (1902 1994)
  - Scientific theory can only be invalidated by contrary experimental evidence
  - Experimental observations are intrinsically interwoven into the scientific method
  - Scientific research is hypothesis-driven
  - Principle of falsification: only such hypotheses are legitimate that could be refuted by experimental evidence



# Creation of Theory, Knowledge

# Thomas Bayes' (1702-1761) ideas re-emerge:

- Bayesian analysis of computational predictions: uncertainty arises from
  - Model parameters m
    - Begin with priori joint PDF  $\rho_{\it M}(m)$ , describing what we know at the beginning
    - Calibrate via Bayesian inference: update the PDF to make the theory (model) consistent with particular observations (data)
  - Experimental observations d
    - Observational data d also will have uncertainties; represent by PDF  $\rho_{D}(d)$
  - Theoretical model  $\theta$ 
    - Consider theory  $\theta$  as conditional probability distribution  $\theta(d \mid m)$  (,likelihood') that maps the parameters m to a probability distribution of the outputs d

15

# Creation of Theory, Knowledge

• Bayes's theorem: Characterizes what is known about the model parameters expressed as the posterior PDF  $\sigma(m \mid d)$  of the parameters m conditioned on the data d

$$\sigma (m \mid d) = \frac{\rho_M(m) \ \theta(d \mid m)}{\rho_D(d)}$$

- Statistically infers the posterior distribution  $\sigma(\bullet \mid d)$  of parameters m that fit the theoretical model to the observations d
- Key to validation and ultimately to prediction with quantifiable uncertainties.

Feature Extraction in Uncertain Data  Data Visualization	
Data Visualization	
Two fundamental cases:	
Visualization of »raw« data	
Visualization of »features«	

# Visualizing Uncertain »Features«

#### Features = entities

- that can be computed from the input data
- that characterize certain aspects of the input data

## Example: features in fields

	Spatial Features	Non-Spatial Features
Scalar	iso-surfaces	min/max; histogram
	ridges	entropy
	MS complex	
Vector	critical points	histogram on S <sup>2</sup> xR
	integral curves	
	topological skeleton	
Tensor	eigenvector field lines	tensor DF

19

# Visualizing Uncertain »Features«

- · Feature-based visualization
  - Definition of feature
  - Preprocessing of raw data: reconstruction, regularization → error propagation
  - Extraction of features
  - Display of features
- → error propagation
- → error propagation
- · How to deal with uncertain features?

# Example: Uncertain Iso-Surfaces

# Questions:

- How do the errors propagate?
   What is the resulting spatial uncertainty?
  - Assume a sample grid.
  - What is the probability of a level crossing in a given cell element?
     (edge, face, voxel, ...)
- · How to depict the spatial uncertainty?

21

# Iso-Contours in Uncertain Data

Positional Uncertainty of Iso-Contours: Condition Analysis and Probabilistic Measures

Kai Pöthkow, Hans-Christian Hege

IEEE Trans Comp Graph Vis, 17:10, Oct 2011, pp. 1393-1406

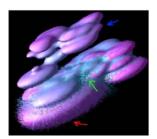
**Probabilistic Marching Cubes** 

Kai Pöthkow, Britta Weber, Hans-Christian Hege

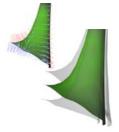
Computer Graphics Forum, 30:3, May 2011, pp. 931-940

Iso-Contours in Uncertain Data: Previous Work

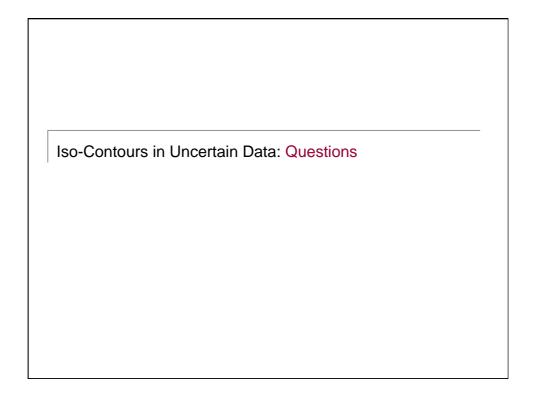
# **Previous Work**

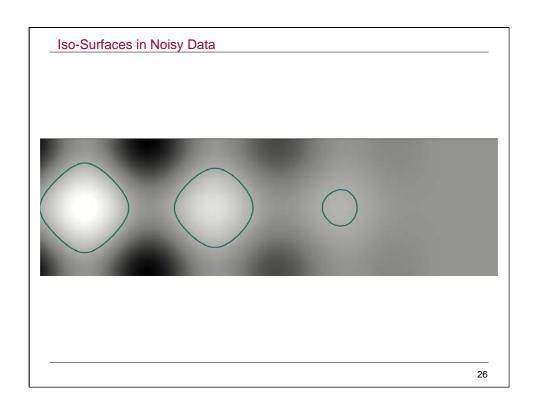


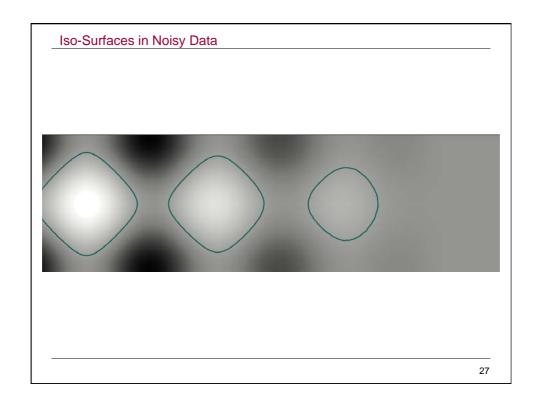
Grigoryan & Rheingans:
Probabilistic Surfaces: Point Based
Primitives to Show Surface
Uncertainty,
IEEE Trans Vis Comput Graph 10(5),
pp. 564-573, 2004

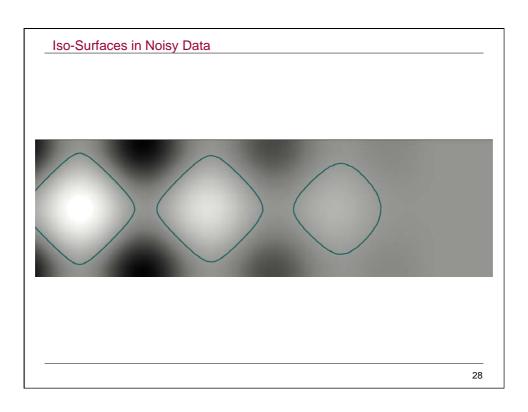


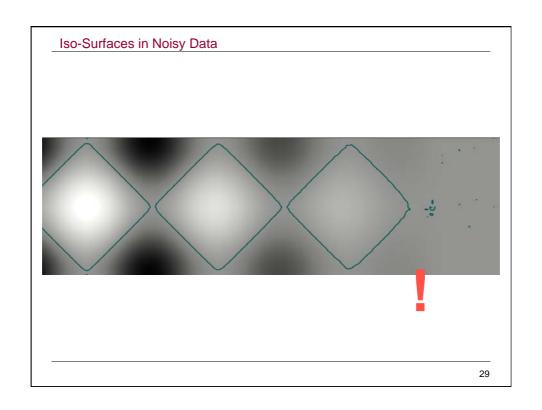
Zehner, Watanabe & Kolditz: Visualization of Gridded Scalar Data with Uncertainty in Geosciences, Comp. & Geosci 36(10), pp.1268-1275, 2010

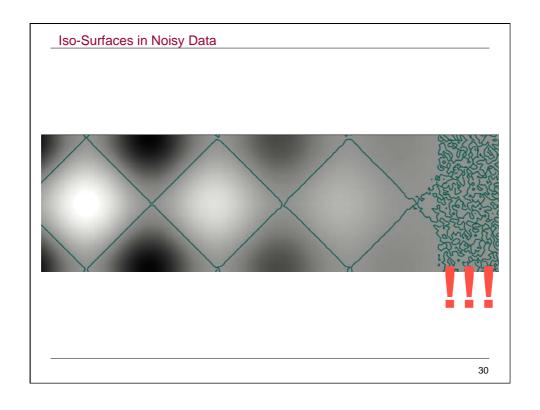












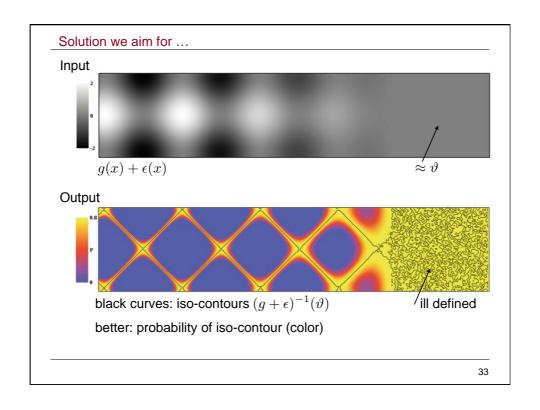
# Questions

- How can we adequately model errors and uncertainty in scalar fields?
- What is the resulting spatial uncertainty of iso-contours?

31

# Questions

- How can we adequately model errors and uncertainty in scalar fields?
- What is the resulting spatial uncertainty of isocontours?
- What is the impact of spatial correlation?



Iso-Contours in Uncertain Data: Sensitivity Analysis

# Sensitivity of Results to Changes in the Data

Problem: from x compute  $\rho$ 

Condition number  $\kappa_{abs}$ : gives a measure of how sensitive the solution of a problem is

to perturbations in the input data.

 $\kappa_{abs}$  is the smallest number with

 $\kappa_{abs} \geq \frac{\|\rho(x) - \rho(x + \epsilon)\|}{|\epsilon|}$ 

Common simplification: linearization. If ho is differentiable then

$$\kappa_{abs} = \|\rho'(x)\|$$

35

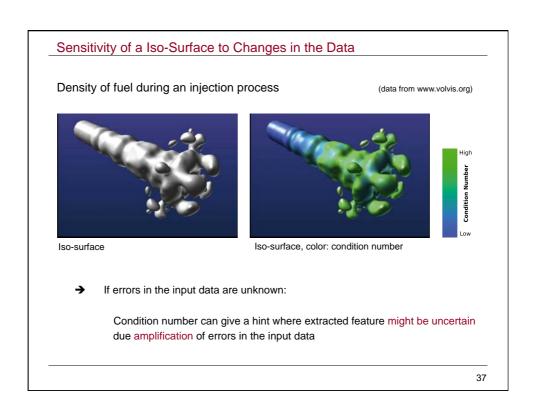
# **Error Amplification**

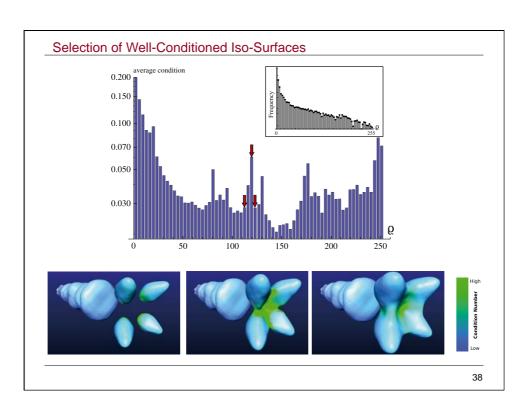
How do iso-contours change if input data (scalar field g) is changed ? Condition number of iso-surface computation  $g^{-1}(\vartheta)=x$ :

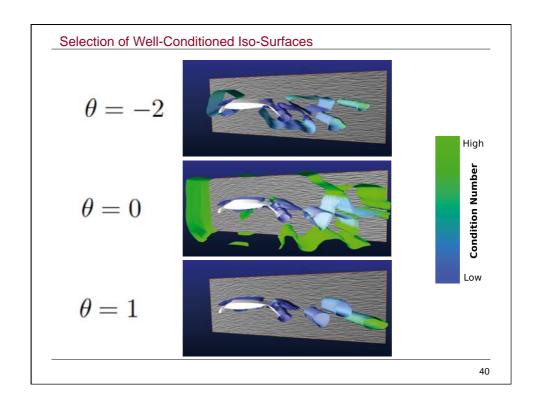
$$\kappa_{abs} = \frac{1}{\|\nabla g(x)\|}$$

# Condition number can diverge!

- κ<sub>abs</sub> is the amplification factor for absolute errors
- No surprise: only for regular values is  $g^{-1}(\vartheta)$  a smooth d-1 manifold
- For iso-contour algorithms g needs to be a Morse function
- But even if |g| is a Morse function: problems occur for  $\|\nabla g\|\gtrsim 0$







Iso-Contours in Uncertain Data: Random Field Model

# Uncertain Iso-contours

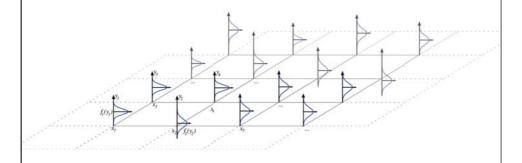
- Scalar field  $g\!:\!M\to\mathbb{R}$  on compact domain  $M\subset\mathbb{R}^d$
- Consider level sets  $\{x \in M | x = g^{-1}(\vartheta)\}$
- Discretely sampled on nodes  $\{\mathbf{x}\}_{i\in I}$  with values  $\ \{Y\}_{i\in I}$
- Let  $\{Y\}_{i\in I}$  be random variables

```
distributions f_i means \mu_i=\mathrm{E}(Y_i) variances \sigma_i^2=\mathrm{E}(Y_i-\mu_i)^2 co-variances \mathrm{Cov}(Y_i,Y_j)=\mathrm{E}((Y_i-\mu_i)(Y_j-\mu_j))
```

• True values of field  $g(\mathbf{x})$  unknown; assume:  $g(\mathbf{x}_i) = \mu_i$ 

# Gaussian Random Field

Discrete random field = multivariate Gaussian RV



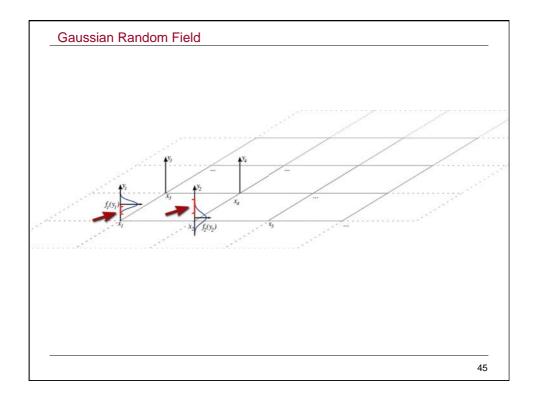
43

# Gaussian Random Field

Discrete random field = multivariate Gaussian RV

$$\mathbf{Y} \sim \mathcal{N}_n(\mu, \Sigma) \qquad \mu = [\mathbf{E}(Y_1), \mathbf{E}(Y_2), \dots, \mathbf{E}(Y_n)] \\ \Sigma = [\mathbf{Cov}(Y_i, Y_j)]_{i=1, 2, \dots, n; j=1, 2, \dots, n}.$$

$$\mathbf{Y}(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu)^T \Sigma^{-1}(\mathbf{y} - \mu)\right)$$



# Gaussian Random Field

Marginalization:

$$\int_{-\infty}^{\infty} dy_{m+1} \dots \int_{-\infty}^{\infty} dy_n \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu)\right)$$

$$= \frac{1}{(2\pi)^{m/2} \det(\tilde{\Sigma})^{1/2}} \exp\left(-\frac{1}{2} (\tilde{\mathbf{y}} - \tilde{\mu})^T \tilde{\Sigma}^{-1} (\tilde{\mathbf{y}} - \tilde{\mu})\right)$$

$$=: f_{\tilde{\mathbf{Y}}}(y_1, \dots, y_m)$$

where  $\tilde{\mathbf{Y}}$  is the reduced random vector and  $\tilde{\mathbf{y}}$ ,  $\tilde{\mu}$  and  $\tilde{\Sigma}$  are the quantities  $\mathbf{y}$ ,  $\mu$  and  $\Sigma$  with n-m columns/rows deleted that correspond to the marginalized variables  $y_{m+1} \cdots y_n$ 

# Gaussian Random Field

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} \qquad \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix} \qquad \begin{bmatrix} \operatorname{Cov}_{1,1} & \cdots & \operatorname{Cov}_{1,m} & \cdots & \operatorname{Cov}_{1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \operatorname{Cov}_{m,1} & \cdots & \operatorname{Cov}_{m,m} & \cdots & \operatorname{Cov}_{m,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_n \end{bmatrix} \qquad \begin{bmatrix} \mu_1 \\ \mu_m \end{bmatrix} \qquad \begin{bmatrix} \operatorname{Cov}_{1,1} & \cdots & \operatorname{Cov}_{1,m} & \cdots & \operatorname{Cov}_{m,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \operatorname{Cov}_{n,1} & \cdots & \operatorname{Cov}_{n,m} & \cdots & \operatorname{Cov}_{n,n} \end{bmatrix}$$

# Local marginal distribution

$$\tilde{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ V \end{bmatrix}$$

$$\tilde{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}$$

$$\tilde{\Sigma} = \begin{bmatrix} \text{Cov}_{1,1} & \cdots & \text{Cov}_{1,m} \\ \vdots & \ddots & \vdots \\ \text{Cov}_{m,1} & \cdots & \text{Cov}_{m,m} \end{bmatrix}$$

47

# Probabilities of Classes of Realizations

Constrain  $m \leq n$  RV  $Y_i$  to subsets  $S_i$  .

Re-order RV such that constrained ones are the first  $\,m\,$  ones.

Probability of constrained realization:

$$\operatorname{Prob}\left(Y_{1} \in S_{1}, \dots, Y_{m} \in S_{m}\right) =$$

$$\int_{S_{1}} dy_{1} \dots \int_{S_{m}} dy_{m} \int_{\mathbb{R}} dy_{m+1} \dots \int_{\mathbb{R}} dy_{n} f_{\mathbf{Y}}(y_{1}, \dots, y_{n})$$

For Gaussian distribution:

$$\int_{S_1} \mathrm{d}y_1 \dots \int_{S_m} \mathrm{d}y_m \ f_{\tilde{\mathbf{Y}}}(y_1, \dots, y_m)$$



# Level Crossing Probabilities

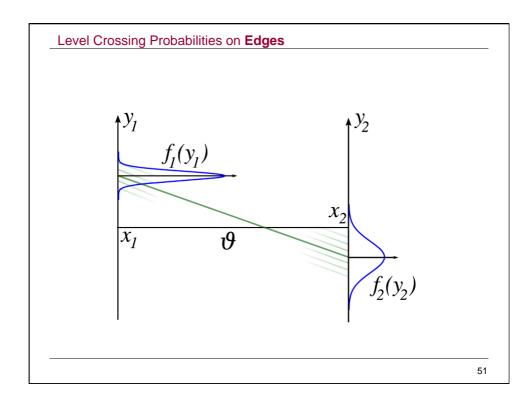
Assume  $\,C^0\,$  interpolant  $\,g\,$  for any realization (= grid function) which takes its extreme values at the sample points.

Consider grid cell  $\,c\,$  with indices  $\,\tilde{I}\in I.$ 

Cell c crosses  $\vartheta$ -level of  $g_{\{y\}}$  if and only if not all differences  $(y_i-\vartheta)_{i\in \tilde{I}}$  have the same sign.

Level crossing probability  $\operatorname{Prob}_c(\vartheta\operatorname{-crossing})$ : Integrate  $\{Y\}_{i\in \tilde{I}}$  over sets  $\{y_j\in \mathbb{R}\ |\ y_j\geq \vartheta\}$  and  $\{y_i\in \mathbb{R}\ |\ y_i\leq \vartheta\}$ 

Alternatively:



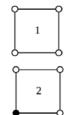
# Level Crossing Probabilities on **Edges**

Edge with bivariate Gaussian RV  $\mathbf{Y} = [Y_1, Y_2]$ 

$$\begin{aligned} &\operatorname{Prob}_{c}(\vartheta\text{-crossing}) = \\ &= \operatorname{Prob}(Y_{1} \leq \vartheta, Y_{2} > \vartheta) + \operatorname{Prob}(Y_{1} > \vartheta, Y_{2} \leq \vartheta) \\ &= \int\limits_{y_{1} \leq \vartheta} \operatorname{d}y_{1} \int\limits_{y_{2} > \vartheta} \operatorname{d}y_{2} \ f_{\mathbf{Y}}(y_{1}, y_{2}) + \int\limits_{y_{1} > \vartheta} \operatorname{d}y_{1} \int\limits_{y_{2} \leq \vartheta} \operatorname{d}y_{2} \ f_{\mathbf{Y}}(y_{1}, y_{2}) \end{aligned}$$



4 Cases (after Symmetry Reduction) △ Corresponding Integrals



$$\mathbf{P}_{\vartheta,1} = \int dy_1 \int dy_2 \int dy_3 \int dy_4 f_{\mathbf{Y}}(y_1, y_2, y_3, y_4)$$
$$(y_1 > \vartheta \land y_2 > \vartheta \land y_3 > \vartheta \land y_4 > \vartheta)$$

$$P_{\vartheta,2} = \int dy_1 \int dy_2 \int dy_3 \int dy_4 f_{\mathbf{Y}}(y_1, y_2, y_3, y_4)$$
$$(y_1 \le \vartheta \land y_2 > \vartheta \land y_3 > \vartheta \land y_4 > \vartheta)$$

$$\mathbf{P}_{\vartheta,3} = \int \mathbf{d}y_1 \int \mathbf{d}y_2 \int \mathbf{d}y_3 \int \mathbf{d}y_4 f_{\mathbf{Y}}(y_1, y_2, y_3, y_4)$$
$$(y_1 \le \vartheta \land y_2 \le \vartheta \land y_3 > \vartheta \land y_4 > \vartheta)$$

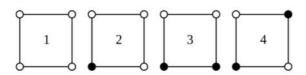
$$\mathbf{P}_{\vartheta,4} = \int \mathbf{d}y_1 \int \mathbf{d}y_2 \int \mathbf{d}y_3 \int \mathbf{d}y_4 f_{\mathbf{Y}}(y_1, y_2, y_3, y_4)$$
$$(y_1 \leq \vartheta \wedge y_2 > \vartheta \wedge y_3 \leq \vartheta \wedge y_4 > \vartheta)$$

53

# Level Crossing Probabilities on Rectangular Cells, ...

Types of integrals  $\stackrel{\triangle}{=}$  symmetry-reduced Marching cubes cases.

In 2D: 4 distinct cases (1 non-crossing, 3 crossing)



In 3D: 15 distinct cases (1 non-crossing, 14 crossing)

In 4D: 223 distinct cases (1 non-crossing, 222 crossing)

In nD: use Polya's counting theory

# Level Crossing Probabilities - Simplified

# of cases (i.e. integrals) with level crossings grows with dimension ...

Better exploit  $\operatorname{Prob}_c(\vartheta\operatorname{-crossing}) = 1 - \operatorname{Prob}_c(\vartheta\operatorname{-non-crossing})$ 



only 2 cases without level crossings
→ only 2 integrals for all dimensions!

e.g. for square cells in 2D:

$$\begin{aligned} \text{Prob}_{c}(\vartheta\text{-crossing}) &= \\ 1 - \int dy_{1} \int dy_{2} \int dy_{3} \int dy_{4} f_{\mathbf{Y}}(y_{1}, y_{2}, y_{3}, y_{4}) \\ & \stackrel{(y_{1} \leq \vartheta \wedge y_{2} \leq \vartheta \wedge y_{3} \leq \vartheta \wedge y_{4} \leq \vartheta)}{\vee (y_{1} > \vartheta \wedge y_{2} > \vartheta \wedge y_{3} > \vartheta \wedge y_{4} > \vartheta)} \end{aligned}$$

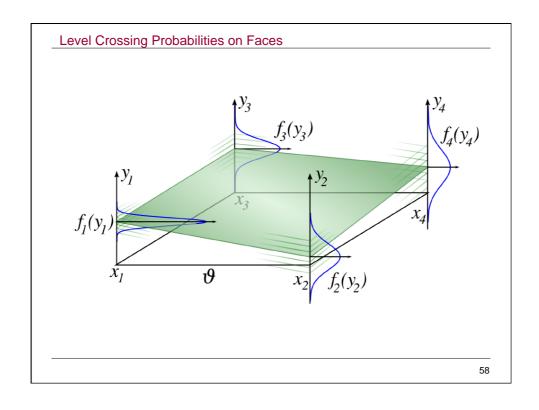
But dimension of integrals still = # vertices of geometric object!

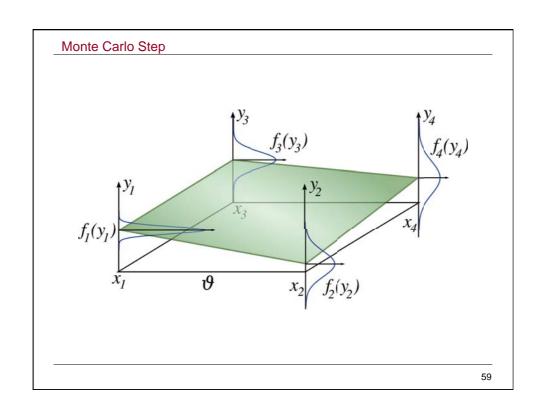
55

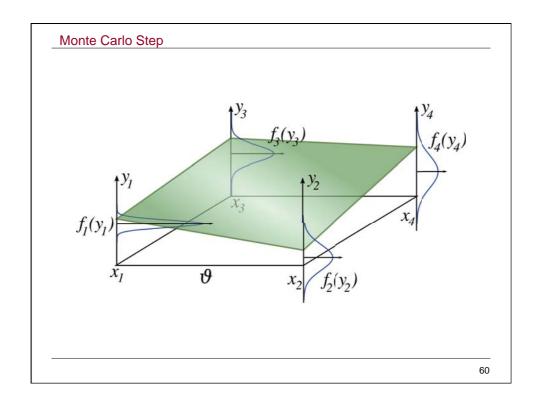
Iso-Contours in Uncertain Data: Algorithm & Implementation

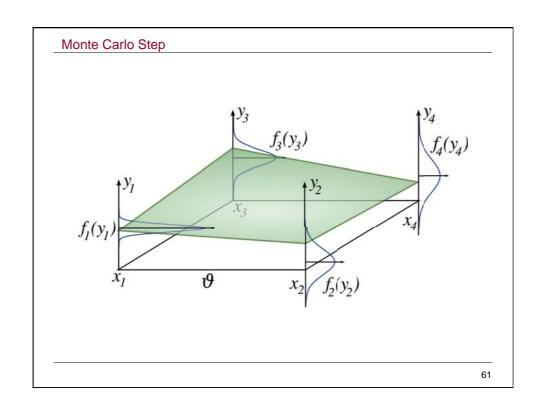
# Algorithm & Implementation

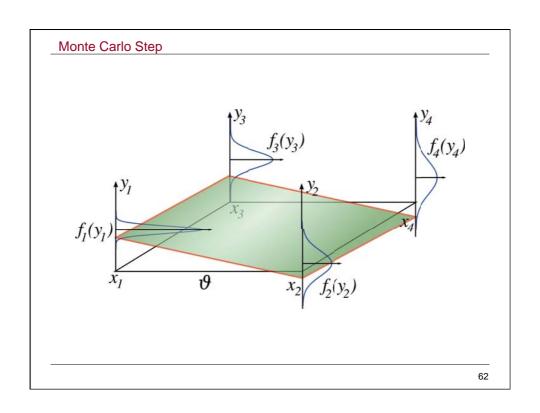
- Preprocessing
  - Estimate  $\hat{\mu}_i$  for all sample points
  - Estimate  $\widehat{\mathrm{Cov}}_{i,j}$  for all 2- or 3-cells
- For a given iso-value  $\boldsymbol{\vartheta}$ 
  - Estimate crossing probabilities using Monte Carlo integration

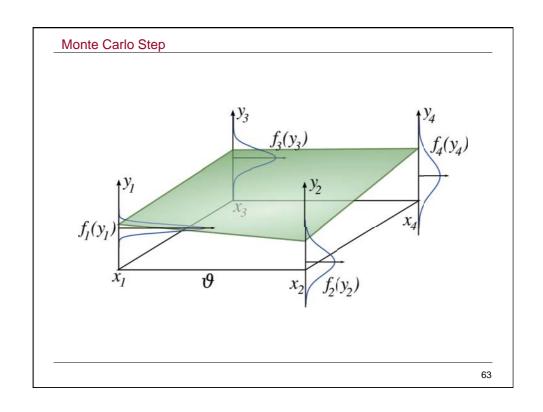


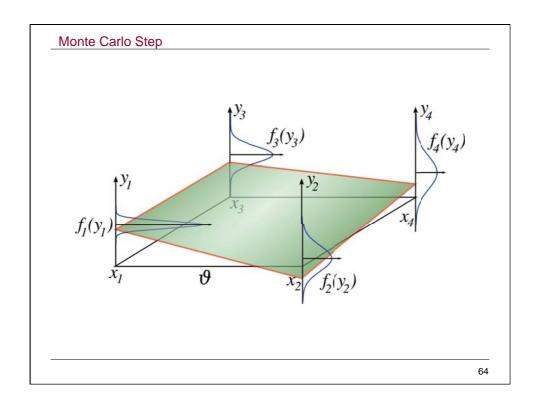


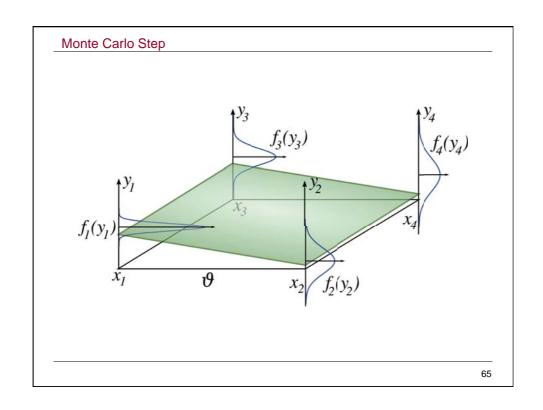


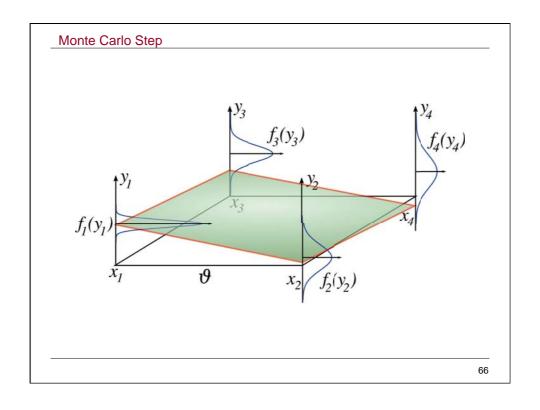






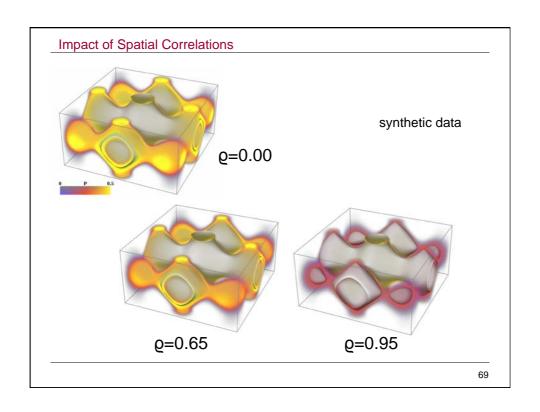


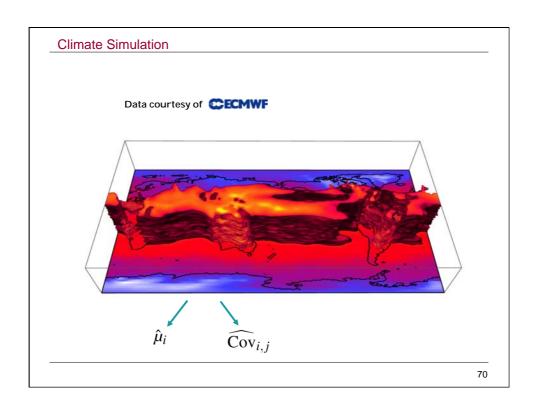


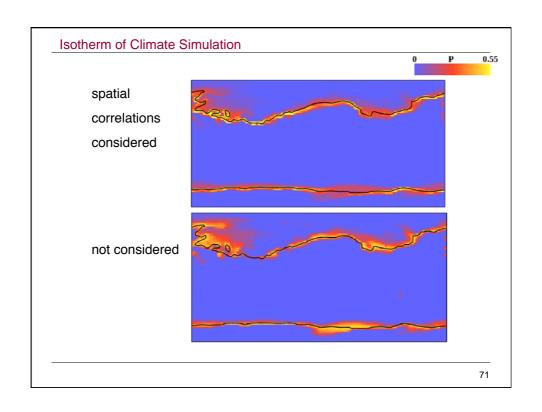


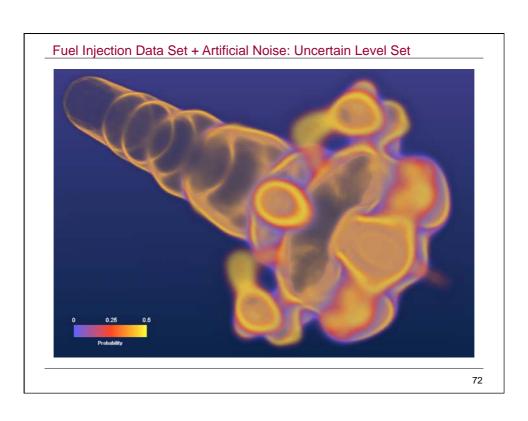
# for each cell c { $L_c \leftarrow \text{CholeskyDecomposition}(\Sigma_c)$ $\# \text{crossings} \leftarrow 0$ $\text{for } 1 \dots \# \text{samples } \{$ $\mathbf{y} \leftarrow \text{random numbers } y_1 \dots y_m \sim \mathcal{U}(0,1)$ $\mathbf{y} \leftarrow \text{BoxMullerTransform}(\mathbf{y})$ $\mathbf{y} \leftarrow L_c \, \mathbf{y} + \mu_c$ $\text{if}(\text{crossing}_{\vartheta}(\mathbf{y})) \ \# \text{crossings} \leftarrow \# \text{crossings} + 1$ $\}$ $\text{Prob}_c \leftarrow \# \text{crossings} / \# \text{samples}$ $\}$

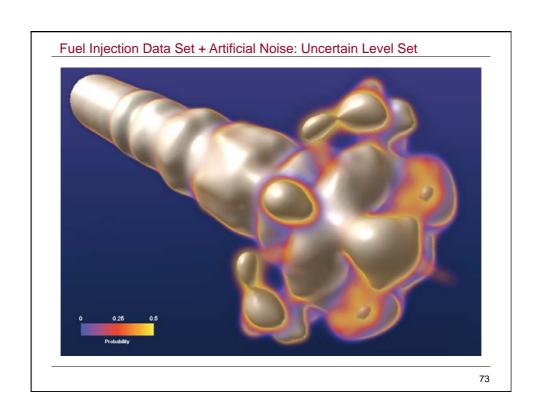
Iso-Contours in Uncertain Data: Results

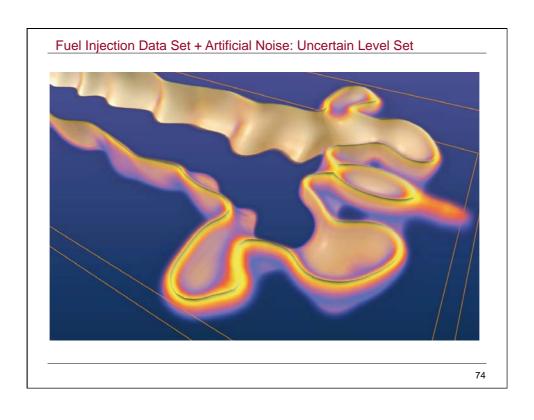


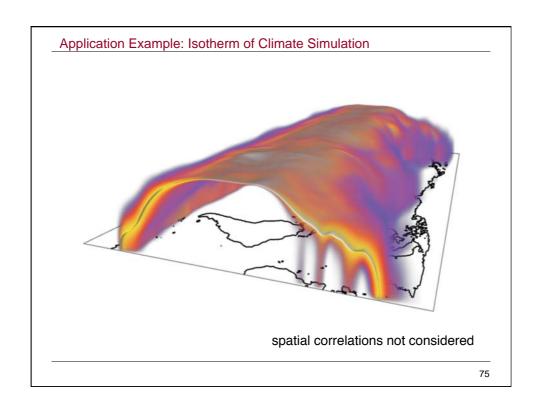


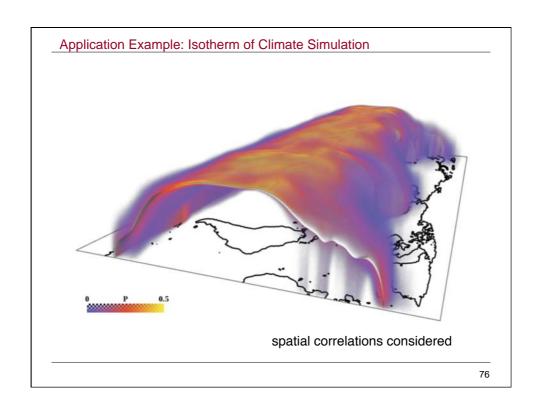












Roads to Sharp Formulations - Future of Uncertainty Vis

# We need to understand ...

Types of uncertainty

Aleatoric

experimenter: results differ, each time she/he runs an experiment modeller: does not foresee the possibility to reduce uncertainty

- Epistemic

modeller:

experimenter: we could in principle know, but don't know in practice sees the possibility to reduce uncertainty by gathering more data or be refining models

Uncertainty representations

- Intervals

interval computing

- Fuzzy numbers, sets

soft computing

- Probabilities, PDFs

probability theory, statistics

Reasoning under uncertainty, decision support

- Formal reasoning

statistical inference

uncertainty in Al

- Defuzzification, decision taking

→ risk & decision theory

# We need to understand and further develop ...

Uncertainty quantification in modelling and simulation

- Estimate parameter uncertainty
- · Develop statistical / fuzzy models
- Analyse uncertainty propagation
- · Perform sensitivity analysis and dimensional reduction
- Develop methods for defuzzification
- Develop tools to support in decision making

79

# We need to understand and further develop ...

# **Uncertainty Visualization**

- · UQ in the visualization pipeline
- · Fuzzy analogues of crisp features, uncertainty of features
- Visual mapping of uncertain data and fuzzy features
- Evaluation of uncertainty representations (perception, cognition)
- Visual support for data processing techniques: data aggregation, ensemble analysis, ...
- · Visual support of defuzzification
- · Visual support in decision making
- · Evaluation of uncertainty VIS / VA systems

# Conclusion

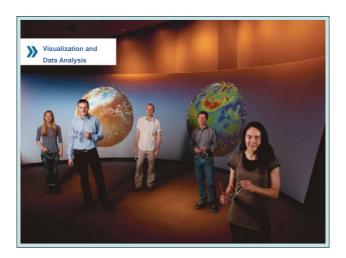
- Major tasks in uncertainty visualization
  - Uncertainty quantification in visualization pipeline
  - Visual mapping of uncertain data and fuzzy features
  - Support in decision making
- · Uncertain features
  - Condition numbers, sensitivity analysis
  - Probabilistic formulation

81

# Conclusion

- Uncertain iso-surfaces
  - reveal information not visible before
- · Assumption of certain distribution law
  - arbitrary number of realizations possible
  - more details than with limited number of realizations
- Advantage of not computing crisp iso-surfaces:
  - no regularity requirements (Morse, non-Morse)
  - no special cases in algorithm for degenerate cases
- Most important research questions:
  - visual mapping
  - non-Gaussian random fields

# Thank you very much for your attention!



www.zib.de/visual