# Estimation of confidence regions for random excursion sets with application to large-scale ice-sheet simulations

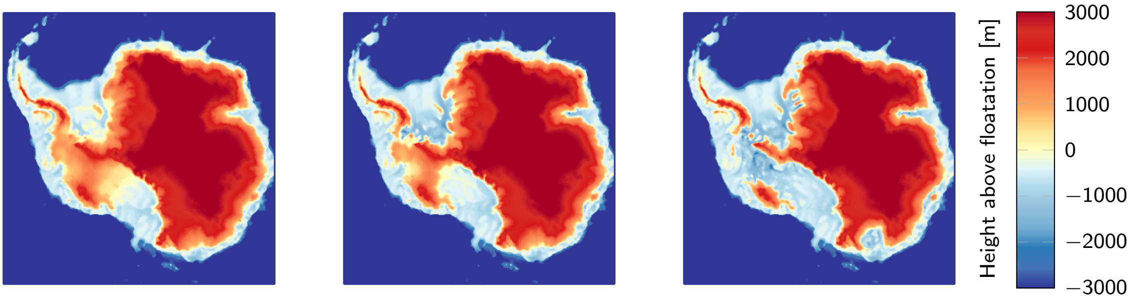
**Kevin Bulthuis**[1,2], F. Pattyn[2], and M. Arnst[1]

[1] Computational and Stochastic Modeling, Université de Liège, Belgium
[2]Laboratory of Glaciology, Université Libre de Bruxelles, Belgium

*UNCECOMP 2019* Hersonissos, Greece June 25, 2019

# Introduction

- Motivation: Building risk-assessment maps for the projected retreat of the Antarctic ice sheet in the presence of uncertainties.
- Challenges:
  - How to quantify the uncertainty in the retreat of the Antarctic ice sheet: excursion sets of spatially non-homogeneous random fields.
  - Computational ice-sheet models have generally a high computational cost.
  - Ice sheets can exhibit abrupt and irreversible retreats (risk of instabilities).




Projection 1

Projection 2

Projection 3

# Outline

(1) Introduction

(2) Confidence regions for excursion sets

(3) Implementation of the problem of quantile estimation for confidence regions

(4) Application: Uncertainty quantification in the retreat of the Antarctic ice sheet

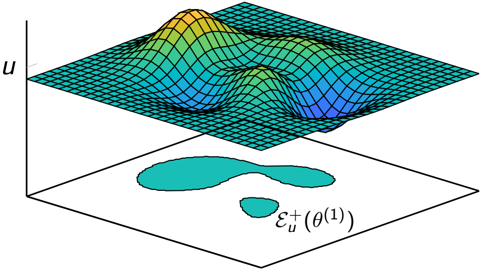
(5) Conclusion

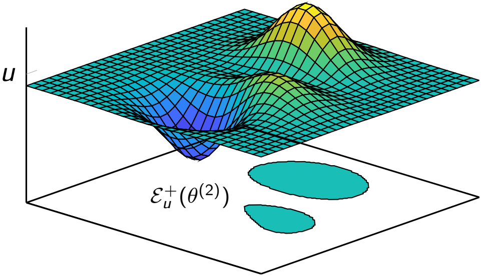
# Problem setting

- Let $\{Y(\mathbf{x}), \mathbf{x} \in D\}$ be a random field defined on a probability space $(\Theta, \mathfrak{U}, \mathbb{P})$, indexed by $D \subset \mathbb{R}^d$ $(d \geq 1)$, with values in $\mathbb{R}$ and with continuous paths almost surely.
- Positive excursion set:

$$\mathcal{E}_u^+ : \Theta \to \mathfrak{F}; \theta \mapsto \mathcal{E}_u^+ = \{\mathbf{x} \in D : Y(\mathbf{x}) \geq u\},$$

  where $u$ is a threshold of interest. The set $\mathcal{E}_u^+$ defines a random closed set in $\mathbb{R}^d$.
- Objective: Characterise the variability/uncertainty in the random closed set $\mathcal{E}_u^+$.



# Confidence regions for excursion sets

- The uncertainty in random sets may be characterised using the concept of confidence regions [Bolin and Lindgren, 2015; French and Hoeting, 2015].
- A closed set $C_\alpha^{\text{out}} \in \mathfrak{F}$ is an outer confidence region for $\mathcal{E}_u^+$ with probability at least $\alpha$ if

$$\mathbb{P}\left(\mathcal{E}_u^+ \subseteq C_\alpha^{\text{out}}\right) \geq \alpha.$$

- An open set $C_\alpha^{\text{in}} \in \mathfrak{L}$ is an inner confidence region for $\mathcal{E}_u^+$ with probability at least $\alpha$ if

$$\mathbb{P}\left(C_\alpha^{\text{in}} \subset \mathcal{E}_u^+\right) \geq \alpha.$$

$\mathcal{E}_u^+ \subset C_\alpha^{\text{out}}, C_\alpha^{\text{in}} \subset \mathcal{E}_u^+$ | $\mathcal{E}_u^+ \not\subset C_\alpha^{\text{out}}, C_\alpha^{\text{in}} \subset \mathcal{E}_u^+$ | $\mathcal{E}_u^+ \subset C_\alpha^{\text{out}}, C_\alpha^{\text{in}} \not\subset \mathcal{E}_u^+$

# Parametric family of sets

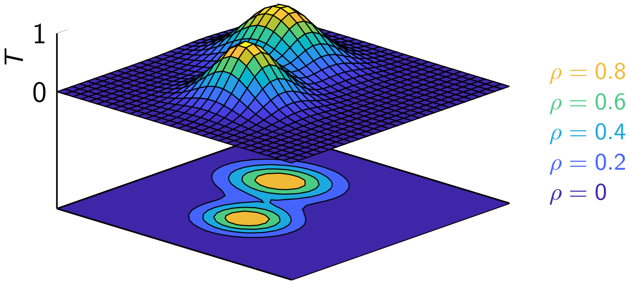
- Interest is directed towards finding the largest inner confidence region:

$$C_\alpha^{\text{in}} = \arg\max_{S \in \mathfrak{L}} |S| \text{ subject to } \mathbb{P}(S \subset \mathcal{E}_u^+) \geq \alpha.$$

  This optimisation problem does not in general have a unique solution.
- The family of sets $\mathfrak{L}$ is restricted to a parametric family of nested sets

$$T_\rho = \{\mathbf{x} \in D : T(\mathbf{x}) > \rho, \rho \in (0,1)\},$$

  where $T : D \to [0,1]$ is referred to as a membership function.



# Evaluation of a confidence region in a parametric family of sets

1) Determine a membership function $T$ for the random field:

- $T$ should (1) quantify the difference between the random field and the threshold $u$ and (2) account for the uncertainty in the random field [French and Hoeting, 2015].
- Examples for $T$ based on first-order statistical descriptors of the random field:

$$T_1(\mathbf{x}) = \mathbb{P}(Y(\mathbf{x}) \geq u),\ T_2(\mathbf{x}) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\mathbb{E}[Y(\mathbf{x})] - u}{\sqrt{2\mathbb{V}[Y(\mathbf{x})]}}\right)\right),\ T_3(\mathbf{x}) = \frac{1}{2}\left(1 + \frac{\mathbb{E}[Y(\mathbf{x})] - u}{\sqrt{\mathbb{E}[(Y(\mathbf{x}) - u)^2]}}\right).$$

2) Solve an optimisation problem:

- The optimal threshold $\rho^*$ satisfies

$$\rho^* = \inf_{\rho \in (0,1)} \rho \text{ subject to } \mathbb{P}(T_\rho \subset \mathcal{E}_u^+) \geq \alpha.$$

- The evaluation of the inclusion probability can be computationally expensive unless the problem is restricted to simple random fields.

# Estimating confidence regions as a problem of quantile estimation

- We recast the optimisation problem

$$\rho^* = \inf_{\rho \in (0,1)} \rho \text{ subject to } \mathbb{P}(T_\rho \subset \mathcal{E}_u^+) \geq \alpha$$

as an equivalent generalised problem of quantile estimation:

$$\rho^* = \inf \{\rho \in (0,1) : F_\chi(\rho) \geq \alpha\} \equiv q_\chi(\alpha),$$

where $F_\chi$ is the cdf and $q_\chi$ the generalised quantile function of the random variable

$$\chi = \sup_{\mathbf{x} \in \left(\mathcal{E}_u^+\right)^c} T(\mathbf{x}).$$

- Proof:

$$\mathbb{P}\left(T_\rho \subset \mathcal{E}_u^+\right) \geq \alpha$$
$$\mathbb{P}\left(\left(\mathcal{E}_u^+\right)^c \subset (T_\rho)^c\right) \geq \alpha$$
$$\mathbb{P}(T(\mathbf{x}) \leq \rho, \mathbf{x} \in \left(\mathcal{E}_u^+\right)^c) \geq \alpha$$
$$\mathbb{P}(\sup_{\mathbf{x} \in \left(\mathcal{E}_u^+\right)^c} T(\mathbf{x}) \leq \rho) \geq \alpha.$$

# Outline

# Evaluation of confidence regions: Computational aspects

- Let $\{Y(\mathbf{x}), \mathbf{x} \in D\}$ be the response of a stochastic computational model that depends on an $\mathbb{R}^n$-valued random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$.
- Evaluation of a first-order statistical descriptor $T$ of the random field:
  - For each $\mathbf{x} \in D$, build an approximation $\widehat{T}(\mathbf{x})$ of $T(\mathbf{x})$ using standard nonintrusive methods for uncertainty quantification (Monte Carlo sampling, spectral expansions, kriging, ...).
- Estimation of the $\alpha$-quantile of the random variable $\chi$:
  - Monte Carlo sampling method:

$$\hat{q}_\chi^\nu(\alpha) = \inf a : \widehat{F}_\chi^\nu(a) \geq \alpha,$$

    where

$$\widehat{F}_\chi^\nu(a) = \frac{1}{\nu} \sum_{l=1}^{\nu} I(\chi^{(l)} \leq a)$$

    is the sample distribution function built on the i.i.d. samples $\{\chi^{(l)}, 1 \leq l \leq \nu\}$.
  - Monte Carlo sampling may be intractable for computational models with a high computational cost and extreme quantiles.

# Quantile estimation based on a surrogate model

a) Polynomial chaos expansion of the response of the stochastic model:

1) Polynomial chaos representation of the random field:

$$Y(\mathbf{x}, \boldsymbol{\xi}) \approx Y^p(\mathbf{x}, \boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=0}^{p} y^p_{\boldsymbol{\alpha}}(\mathbf{x})\, \psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}).$$

2) Estimation of the excursion set:

$$\mathcal{E}_u^+ \approx \mathcal{E}_u^{+,p} = \left\{\mathbf{x} \in D : Y^p(\mathbf{x}) \geq 0\right\}.$$

3) Surrogate model for $\chi$:

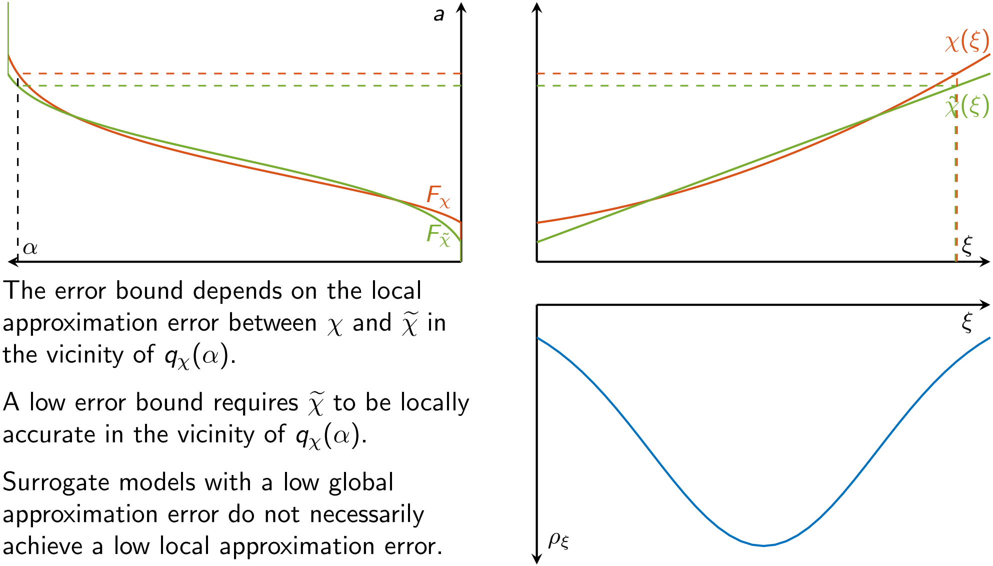
$$\chi(\boldsymbol{\xi}) \approx \sup_{\mathbf{x} \in \left(\mathcal{E}_u^{+,p}\right)^{\mathrm{c}}} T(\mathbf{x}).$$

b) Polynomial chaos expansion of the random variable $\chi$:

$$\chi(\boldsymbol{\xi}) = \sup_{\mathbf{x} \in \left(\mathcal{E}_u^{+}\right)^{\mathrm{c}}} T(\mathbf{x}) \approx \chi^p(\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=0}^{p} \chi^p_{\boldsymbol{\alpha}}\, \psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}).$$

- This approach is enabled by the reformulation of the optimisation problem for confidence regions as a problem of quantile estimation of a random variable.

# Quantile estimation based on a surrogate model: Approximation error




- The error bound depends on the local approximation error between $\chi$ and $\widetilde{\chi}$ in the vicinity of $q_\chi(\alpha)$.
- A low error bound requires $\widetilde{\chi}$ to be locally accurate in the vicinity of $q_\chi(\alpha)$.
- Surrogate models with a low global approximation error do not necessarily achieve a low local approximation error.

# Quantile estimation based on a hybrid approach

- Following Li and Xiu (2010), we build a hybrid surrogate model $\widetilde{\chi}^{\mathrm{h}}$ by correcting the surrogate model $\widetilde{\chi}$ with $\chi$ in the vicinity of $q_{\chi}(\alpha)$:

$$\widetilde{\chi}^{\mathrm{h}}(\boldsymbol{\xi}) = \widetilde{\chi}(\boldsymbol{\xi}) I\left(|\widetilde{\chi}(\boldsymbol{\xi}) - q_{\chi}(\alpha)| > \gamma\right) + \chi(\boldsymbol{\xi}) I\left(|\widetilde{\chi}(\boldsymbol{\xi}) - q_{\chi}(\alpha)| \leq \gamma\right).$$

Let $\chi : \mathbb{R}^n \to \mathbb{R}$ be a measurable continuous function, $\widetilde{\chi}$ be a surrogate model of $\chi$ and $\widetilde{\chi}^{\mathrm{h}}$ be a hybrid surrogate model with $\gamma$ that satisfies

$$\mathbb{P}\left(|\widetilde{\chi} - \chi| > \gamma\right) \leq \epsilon$$

for some $\epsilon \geq 0$. Then, the quantile function $q_{\widetilde{\chi}^{\mathrm{h}}}$ satisfies

$$|q_{\widetilde{\chi}^{\mathrm{h}}}(\alpha) - q_{\chi}(\alpha)| = \mathcal{O}(\epsilon).$$

- The threshold $\gamma$ to achieve an error control of $\epsilon$ is given by $\gamma = \frac{1}{\epsilon^{1/p}} \|\chi - \widetilde{\chi}\|_{\mathbb{L}^p}$.
- Algorithm: Evaluate iteratively $\chi$ at a set of new points around the estimated quantile in the parameter space until a desired level of accuracy is achieved.

# Outline

(1) Introduction

(2) Confidence regions for excursion sets

(3) Implementation of the problem of quantile estimation for confidence regions

(4) Application: Uncertainty quantification in the retreat of the Antarctic ice sheet

(5) Conclusion

# Ice-sheet simulation: Problem setting

Forcing

$a_s(\xi_1), a_w(\xi_1, \xi_2)$

Ice-sheet model

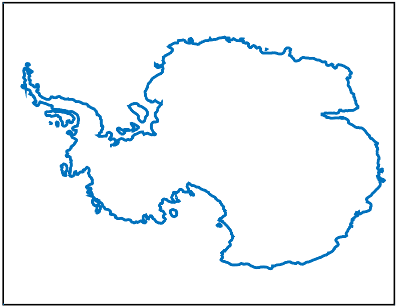
$$\left\{ Y(\mathbf{x}) = h(\mathbf{x}, \boldsymbol{\xi}) + \frac{\rho_w}{\rho_i} b(\mathbf{x}, \boldsymbol{\xi}), \mathbf{x} \in D \right\}$$

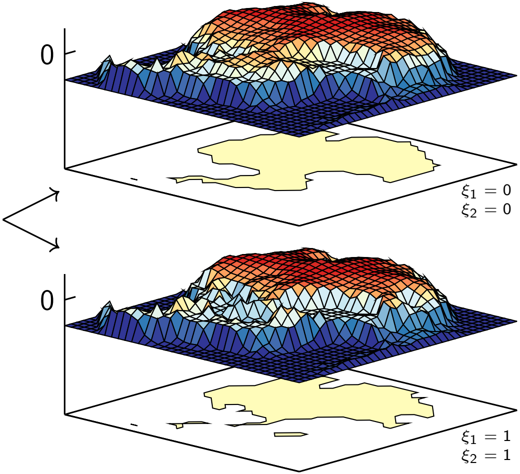
Excursion set

$$\mathcal{E}_0^+ = \{\mathbf{x} \in D : Y(\mathbf{x}) \geq 0\}$$

$\xi_1 \sim \mathcal{U}(0,1)$

$\xi_2 \sim \mathcal{U}(0,1)$ $\longrightarrow$





# Reference solution with Monte Carlo sampling

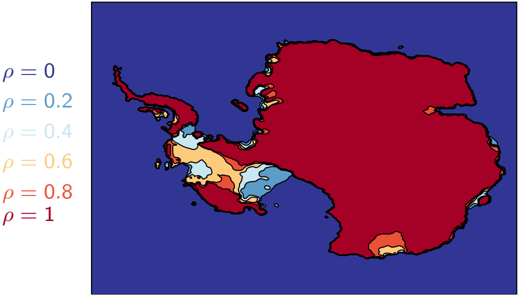
- We build a reference solution using $\nu = 5000$ Monte Carlo samples.
- Membership function:

$$\hat{T}(\mathbf{x}) = \frac{1}{\nu}\sum_{l=1}^{\nu} I(Y(\mathbf{x}) \geq 0).$$

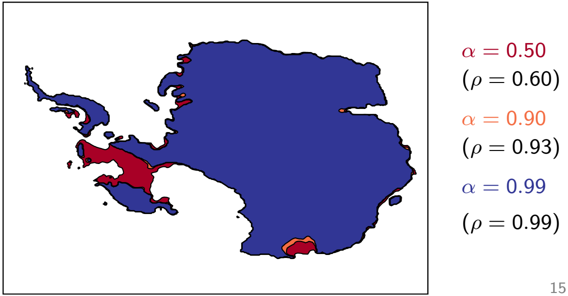
- Quantile estimation:

$$\hat{q}^{\nu}_{\chi}(\alpha) = \inf a : \widehat{F}^{\nu}_{\chi}(a) \geq \alpha.$$

Membership function

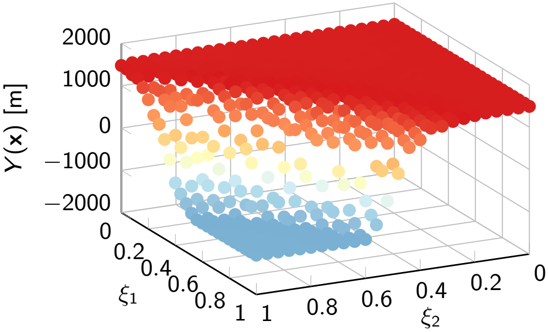


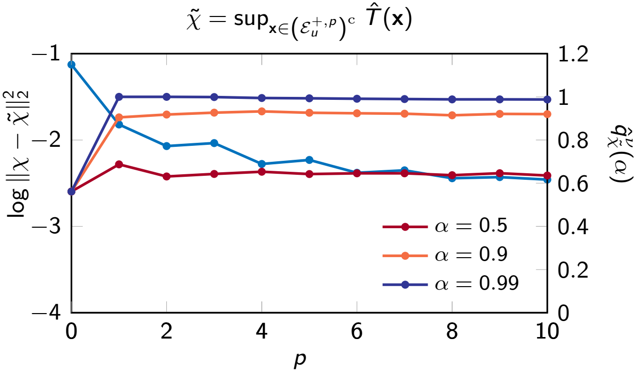

Confidence regions



# Polynomial chaos expansion of the response of the stochastic model

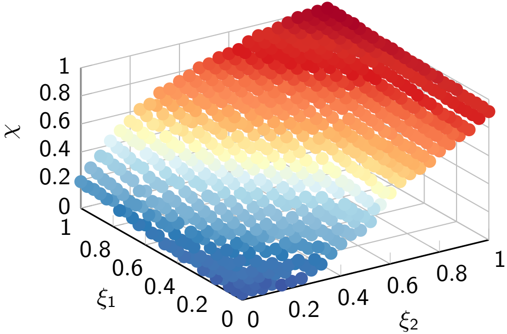
- The local response may exhibit a sharp discontinuity in the presence of instability.
- A polynomial chaos approximation leads to a poor approximation of the response.

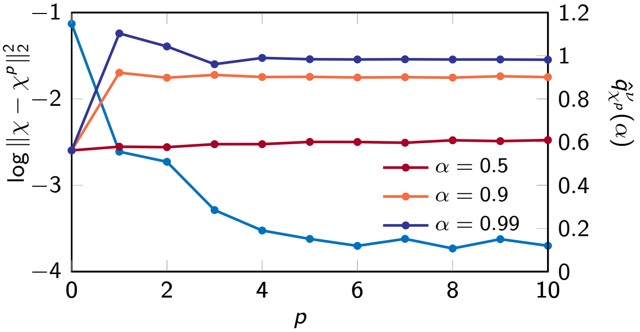


# Polynomial chaos expansion of the random variable $\chi$

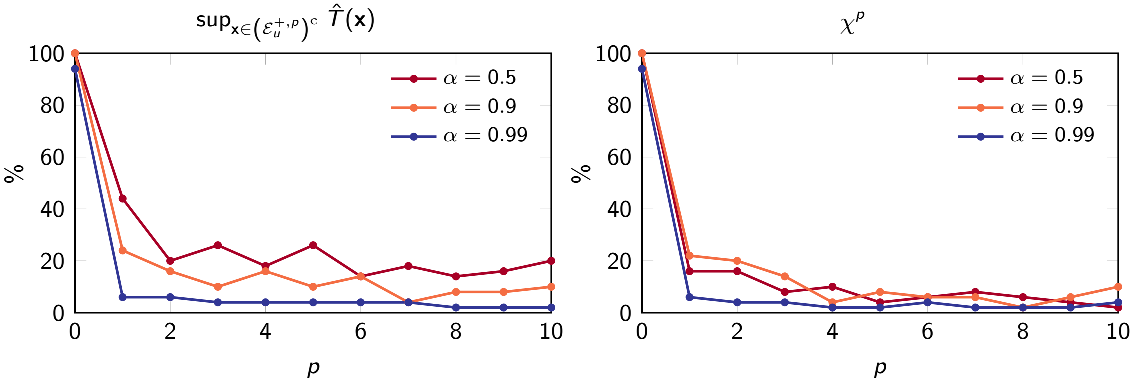
- The random variable $\chi$ exhibits a smoother behaviour than the random field (global averaging).
- A polynomial chaos approximation of $\chi$ provides a more accurate surrogate model.



# Hybrid approach: validation test

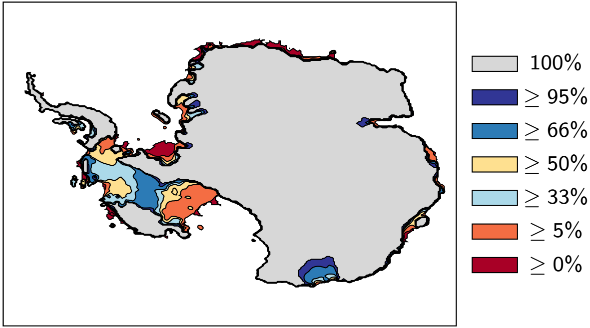
- The efficiency of the hybrid method is measured as the percentage of evaluations of the computational model required to determine the reference quantile.
- The efficiency of the hybrid method is improved for a surrogate model based on a polynomial chaos expansion of $\chi$.



# Risk-assessment map

- We build a risk-assessment map for the retreat of the Antarctic ice sheet by superimposing confidence regions with different levels of probability.
- Confidence regions give insight into the most vulnerable regions to instabilities and the impact of uncertainties on the retreat of the Antarctic ice sheet.

Probability to remain grounded



# Outline

# Conclusion

- Uncertainty in excursion sets:
  - Confidence regions provide a useful way to represent the uncertainty in excursion sets;
  - Confidence regions are estimated in a parametric family of nested sets;
  - Estimating a confidence region may be recast as a problem of quantile estimation.
- Implementation:
  - Quantile estimation based on a surrogate model requires the surrogate model to be locally accurate in the vicinity of the quantile;
  - We used a multifidelity approach in which the computational model is only evaluated in the vicinity of the quantile.
- Application:
  - Surrogate models based on a polynomial chaos expansion of the computational model perform poorly in the presence of instability and abrupt behaviours;
  - A surrogate model for the random variable $\chi$ may perform better than a surrogate model built on the computational model;
  - Confidence regions allow to draw risk-assessment maps.

# Estimation of confidence regions for random excursion sets with application to large-scale ice-sheet simulations

**Kevin Bulthuis**[1,2], F. Pattyn[2], and M. Arnst[1]

[1] Computational and Stochastic Modeling, Université de Liège, Belgium
[2]Laboratory of Glaciology, Université Libre de Bruxelles, Belgium

*UNCECOMP 2019* Hersonissos, Greece June 25, 2019

# References

- R. Adler and J. Taylor. Random Fields and Geometry. Springer Monographs in Mathematics, 2007.
- D. Bolin and F. Lindgren. Excursion and contour uncertainty regions for latent Gaussian models. Journal of the Royal Statistical Society: Series B, 2015.
- K. Bulthuis et al. Uncertainty quantification of the multi-centennial response of the Antarctic Ice Sheet to climate change. The Cryosphere, 2019.
- G. Enss et al. Nonparametric quantile estimation based on surrogate models, IEEE Transactions on Information Theory, 2016.
- J. French and J. Hoeting. Credible regions for exceedance sets of geostatistical data. Environmetrics, 2015.
- J. Li and D. Xiu. Evaluation of failure probability via surrogate models. Journal of Computational Physics, 2010.
- I. Molchanov. Theory of Random Sets. Springer London, 2017,
- F. Pattyn. Sea-level response to melting of Antarctic ice shelves on multi-centennial time scales with the fast Elementary Thermomechanical Ice Sheet model (f.ETISh v1.0). The Cryosphere, 2017.

# Acknowledgement

The first author, Kevin Bulthuis, would like to acknowledge the Belgian National Fund for Scientific Research (F.R.S.-FNRS) for its financial support (F.R.S.-FNRS Research Fellowship).

FREEDOM TO RESEARCH