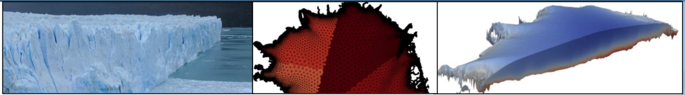
SAND2023-12646C

# Computationally efficient estimation of the extreme event probability of the mass loss of Greenland and Antarctic ice sheets



Center for Computing Research

*Presented by:*

K. Liegeois, M. Perego & G. Stadler

SIAM CSE 2023

# Outline

- Introduction:
  - Land ice simulation and UQ big picture;
  - Extreme event probabilities;
- Optimization approach;
  - Computation of the most likely point;
  - Importance sampling;
- Numerical results;
  - Implementation and test problems;
  - Probabilities and performance;
- Conclusions and future work.

# Unlikely events with high impacts

Unlikely events with high impacts:

- Are difficult to quantify with standard approaches;
  events with small probability require a large number of samples to be evaluated precisely;
- Can have severe consequences (even if they are rare);
  large tsunamis are very rare but can lead to high losses.

When unlikely events have high impacts, it is important to have a precise estimation of their probabilities.

# Unlikely events with high impacts

Unlikely events with high impacts:

- Are difficult to quantify with standard approaches;
  events with small probability require a large number of samples to be evaluated precisely;
- Can have severe consequences (even if they are rare);
  large tsunamis are very rare but can lead to high losses.

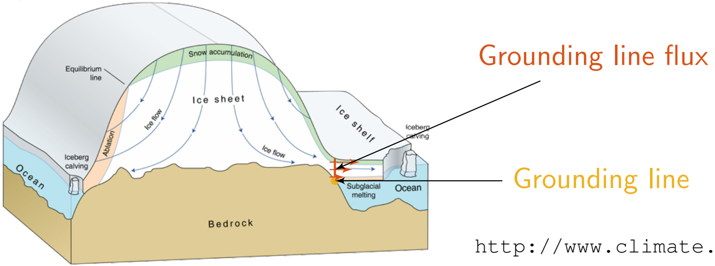
When unlikely events have high impacts, it is important to have a precise estimation of their probabilities.

Example:


In Infinity War, Dr Strange simulated 14 million samples and saw only one case where the Avengers win; the event was extremely unlikely but had severe consequences.

# Unlikely events with high impacts for land ice simulation

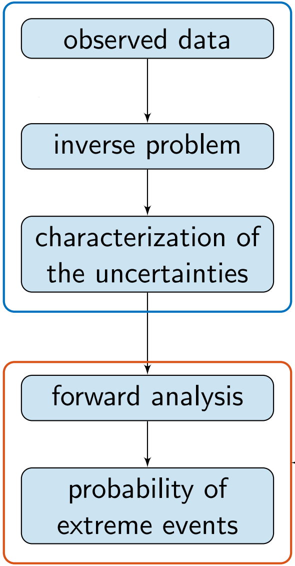
In the case of land ice simulation, this will allow us to estimate smaller probabilities for different levels of sea level rise due to land ice mass loss.

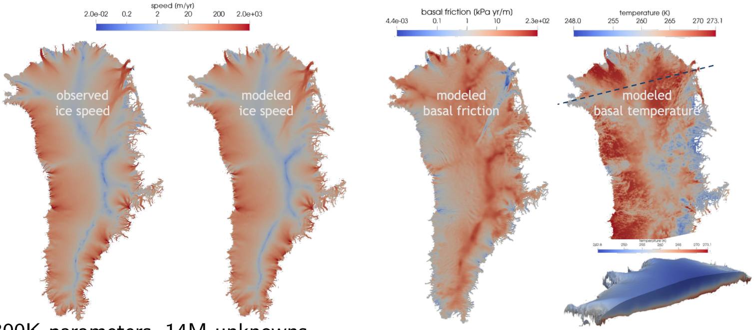



http://www.climate.be

Random parameter: the basal friction between the ice sheet and the bedrock (random field),
Unlikely event of interest: to have a flux at the grounding line above a certain threshold.

# UQ big picture and flow chart



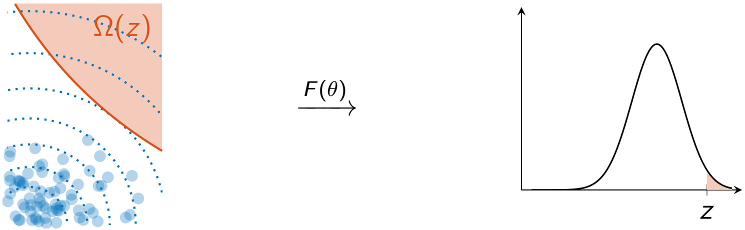

300K parameters, 14M unknowns.
Initialization: 10 hours on 2k cores on NERSC Cori (Haswell),
The optimization is constrained by the coupled velocity-temperature solvers.

The work presented in this talk.

# Introduction: Extreme event probability estimation

Given a $n$-variate Gaussian variable $\theta \sim \mathcal{N}_n(0, \mathrm{I})$ and a $F$ parameter-to-event map (involved PDE solve), quantity of interest:

$$F : \theta \sim \mathcal{N}_n(0, \mathrm{I}) \rightarrow \mathbb{R}$$




Target: Estimate the measure of extreme event sets for $z \gg 0$

$\Omega(z) := \{\theta : F(\theta) \geq z\}$ i.e., compute $\mathbb{P}(F(\theta) \geq z)$ when $\mathbb{P}(F(\theta) \geq z) \ll 1$.

Tong, S., Vanden-Eijnden, E., & Stadler, G. (2021). Extreme event probability estimation using PDE-constrained optimization and large deviation theory, with application to tsunamis. Communications in Applied Mathematics and Computational Science, 16(2), 181-225.

# Extreme event probabilities using optimization theory

The used strategy relies on finding $\theta^\star(z)$ the most likely point above the threshold which can be computed by solving the PDE-constrained optimization problem:

$$\theta^\star(z) = \underset{\theta \in \Omega(z)}{\arg\min}\, I(\theta),$$

where $I(\theta) = \frac{1}{2}\|\theta\|^2$ for $\theta \sim \mathcal{N}_n(0, \mathrm{I})$ as $pdf(\theta) = c\,\exp(-I(\theta))$.

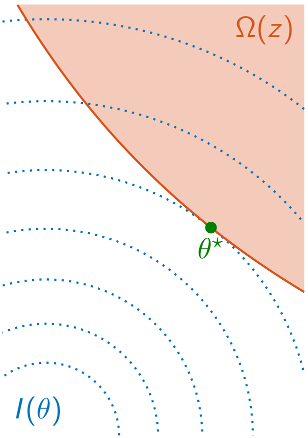
Then, the probability can be approximated as follows:

$$\mathbb{P}(F(\theta) \geq z) \approx C_0(z) \exp\left(-I\left(\theta^\star(z)\right)\right), \qquad \text{as } z \to \infty,$$

where $C_0(z)$ is a sub-exponential prefactor.

The method relies on 2 steps:

- Compute the most likely point $\theta^\star$,
- Compute the prefactor $C_0(F(\theta^\star))$.



# Computation of the most likely point

Under some assumptions, the minimizer over $\Omega(z)$ is reached on $\partial\Omega(z)$ and the inequality constraint is now active:

$$\theta^\star(z) = \arg\min_{\theta \in \partial\Omega(z)} I(\theta).$$

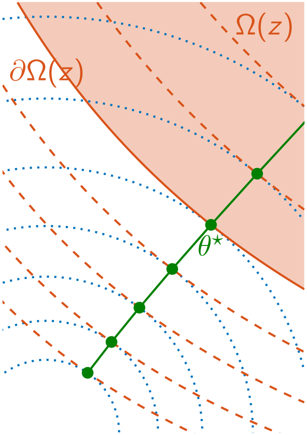
We used a quadratic penalty method as follows:

$$\theta^\star(z) = \arg\min_{\theta} I(\theta) + \alpha\, (F(\theta) - z)^2\,,$$

where $\alpha$ is a penalty weight that should be large enough such that $F(\theta^\star) \approx z$.

The used strategy is the following:

- Select an increasing sequence $z_1, \ldots, z_m$ of quantity of interest,
- For a given $z_i$, solve the corresponding optimization problem using $\theta^\star(z_{i-1})$ as the initial guess,
- Deduce the sequence $\theta^\star(z_1), \ldots, \theta^\star(z_m)$.



# Computation of the prefactor with a sampling strategy

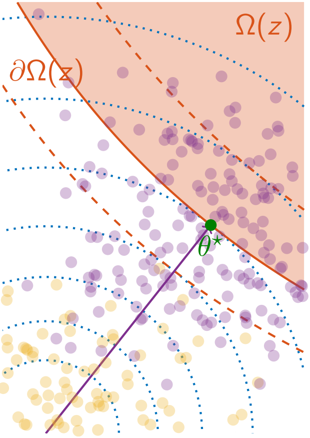
One strategy is to use an Importance Sampling (IS) strategy:

- to draw $N$ random samples $\theta_1, \ldots, \theta_N$ from the initial distribution,
- for a given value of $z_i$:
  - shift the samples: $\tilde{\theta}_k = \theta_k + \theta^\star(z_i)$,
  - evaluate $F$ for all the $N$ shifted samples,
  - evaluate:

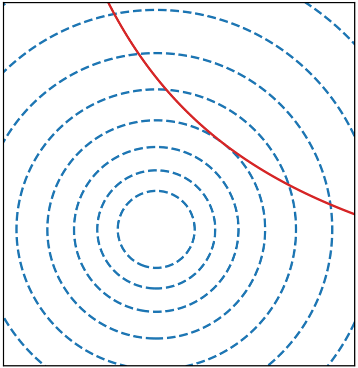
$$P_N^{IS}(z) = e^{-I(\theta^\star)} \frac{1}{N} \sum_{k=1}^{N} \left[ \mathbb{1}_{\Omega(z)} \left(\tilde{\theta}_k\right) \exp\left( - \left(\tilde{\theta}_k - \theta^\star\right)^\top \theta^\star \right) \right],$$

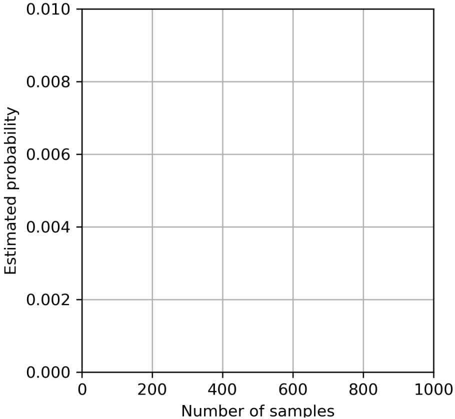
  where $z$ can be different from $F(\theta^\star)$.

- Advantages: $\mathbb{E}\left[P_N^{IS}(z)\right] = P(z)$ and variance in $1/N$,
- Challenges: this approach requires $N \times m$ evaluations of $F$ where $m$ is the number of $z$ values.

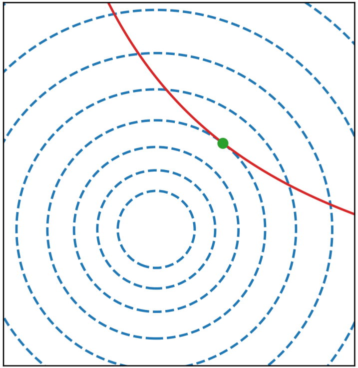


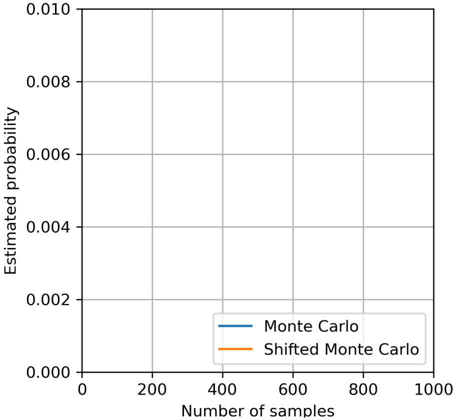
# Monte Carlo strategy



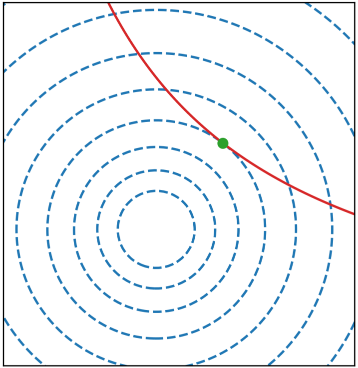


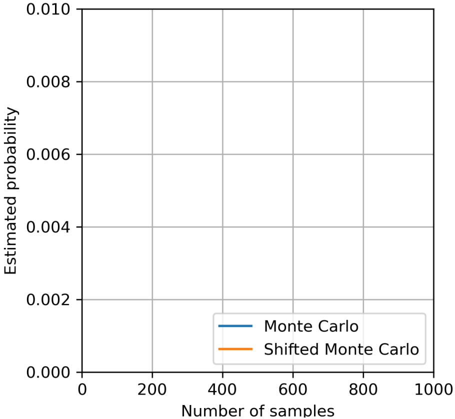
# Optimization strategy with shifted Monte Carlo sampling





# Optimization strategy with shifted Monte Carlo sampling





# Numerical strategies and used software

- FE software: Albany,
- PDE constrained optimizer: algorithm: trust region, software: ROL,
- Non-linear solver: algorithm: Newton solver, software: NOX,
- Linear solver: algorithm: GMRES, software: Belos,
- Preconditioner: algorithm: Schwarz, software: FROSch,
- Preconditioner: algorithm: multigrid, software: MueLu,
- First and second derivative computation: algorithm: automatic differentiation (AD), software: Sacado,
- Reduced Hessian and Gradient vector product computed using ROL and AD.

Albany

ROL
RAPID OPTIMIZATION LIBRARY

TRILINOS

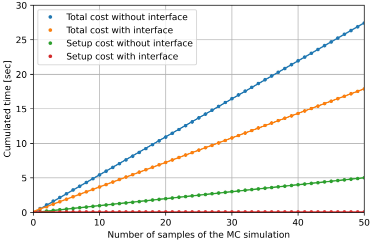
# PyAlbany: A Python interface to Albany

PyAlbany allows to easily and quickly:

- Use Albany without C++ or bash knowledge (convenient for students),
- Prototype applications that require multiple Albany solves,
- Enable fast pre-processing and post-processing in Python,
- Use Python as a glue language to couple Albany with other software:
  - For UQ methods (PyDakota),
  - For machine learning (TensorFlow, Keras, Scikit-learn),
  - For plotting (Matplotlib, Paraview).

```python
import numpy as np
from numpy.random import default_rng
from PyAlbany import Utils

parallelEnv = Utils.createDefaultParallelEnv()

# Create an Albany problem:
filename = "input_scalar.yaml"
paramList = Utils.createParameterList(
    filename, parallelEnv
)

problem = Utils.createAlbanyProblem(paramList, parallelEnv)

parameter_map_0 = problem.getParameterMap(0)
parameter_0 = Utils.createVector(parameter_map_0)

parameter_0_view = parameter_0.getLocalView()

N = 200
p_min = -2.
p_max = 2.

# Generate N samples randomly chosen in [p_min, p_max]:
rng = default_rng()
p = rng.uniform(p_min, p_max, N)
QoI = np.empty((N,))

# Loop over the N samples and evaluate the quantity of interest:
for i in range(0, N):
    parameter_0_view[0] = p[i]
    parameter_0.setLocalView(parameter_0_view)
    problem.setParameter(0, parameter_0)

    problem.performSolve()

    response = problem.getResponse(0)
    QoI[i] = response.getLocalView()[0]
```

Listing 3: Monte Carlo example.

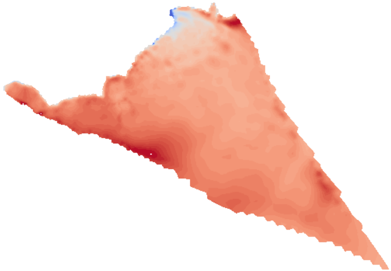



Liegeois, K., Pergeo, M., & Hartland, T. (2023). PyAlbany: A Python interface to the C++ multiphysics solver Albany. Journal of Computational and Applied Mathematics, 425, 115037.

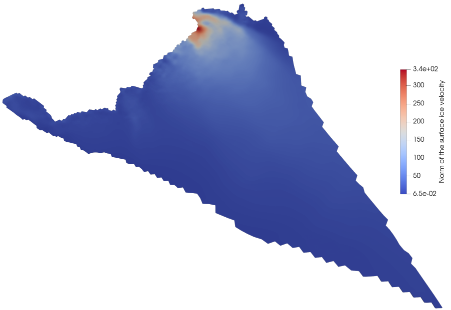
# Land ice test problem: Humboldt glacier

Mean case:

Basal friction:



Surface ice velocity:




Courtesy of
T. Hillebrand.

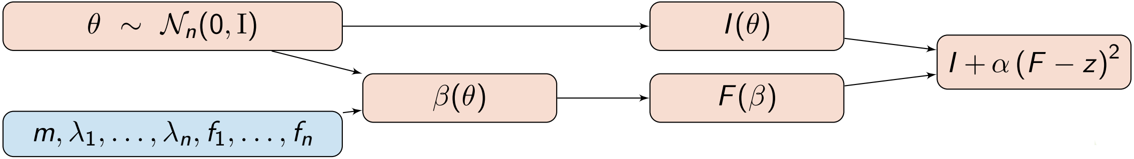
The velocity is faster if the friction is smaller, the quantity of interest is the flux at the grounding line; we expect the extreme events to be associated to smaller basal friction values.

# Land ice test problem: Humboldt glacier

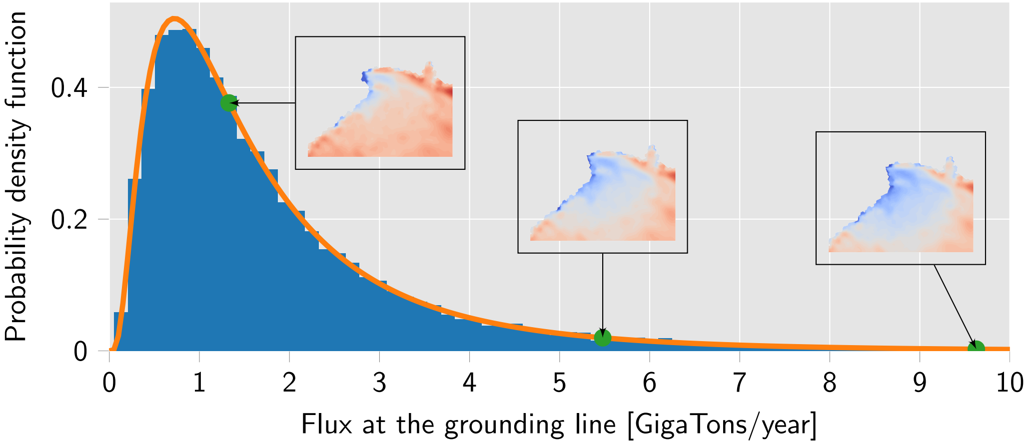
- Random parameter: the basal friction represented using a log-normal random field and a KL expansion with 24 modes:

$$\begin{aligned}\beta(\boldsymbol{x},\theta) &= e^{\log(m(\boldsymbol{x}))+\sum_{i=0}^{n}\sqrt{\lambda_i}\varepsilon_i(\theta)f_i(\boldsymbol{x})},\\ \int_\Omega \mathrm{Cov}(\boldsymbol{x},\boldsymbol{x}')f_i(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' &= \lambda_i\, f_i(\boldsymbol{x}),\\ \mathrm{Cov}(\boldsymbol{x},\boldsymbol{x}') &= \sigma\, e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|}{\ell}},\end{aligned}$$

  with $\sigma = 0.1$, $\ell = 50\,\mathrm{km}$, and $n = 24$. Those modes allow to capture 99% of the variance.
- Quantity of interest: flux at the grounding line.
- PDE: steady state first order Stokes equation, Blatter-Pattyn model.

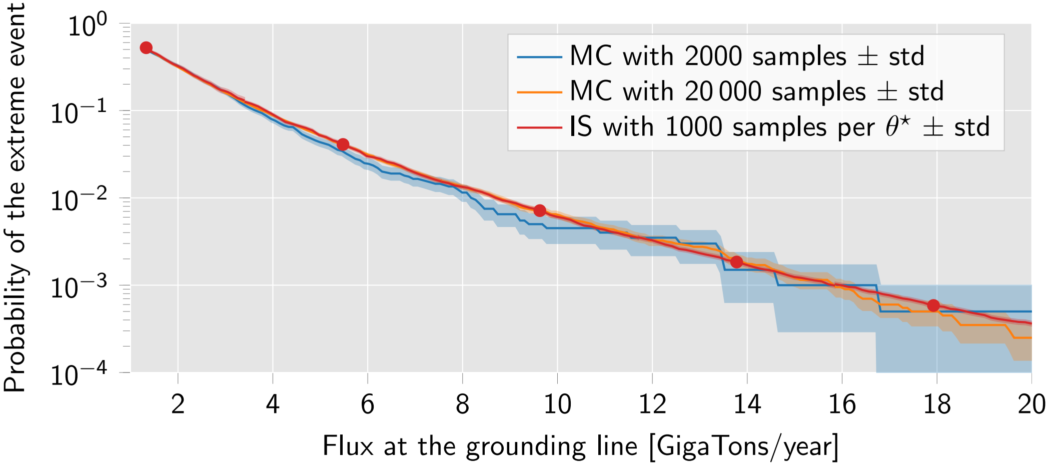


# Quantity of interest and random samples




Histogram of 20 000 samples, the orange curve is a log-normal distribution that fits the data the best.

# Probability of the extreme events




For smaller values of $z$, all the methods are consistent. When increasing $z$, the MC approach needs more and more samples to be consistent with the importance sampling.

# Performance

Average cost per simulation (measured on Skylake):

| | Wall-clock time | Relative cost |
|---|---|---|
| To compute $F(\theta)$ | $c_F = 13$ sec | 1 |
| To compute $\theta^\star(z)$ | $c_{\theta^\star} = 2257$ sec | 174 |

Expected cost per method:

| | $\mathbb{E}[c]$ | $\mathbb{E}[c/c_F]$ |
|---|---|---|
| Monte Carlo | $N_{MC}\, c_F$ | $N_{MC}$ |
| Importance Sampling per $\theta^\star$ | $c_{\theta^\star} + N_{IS}\, c_F$ | $174 + N_{IS}$ |

Comparison of Monte Carlo with $N_{MC} = 20\,000$ and importance sampling with $N_{IS} = 1000$:

$$\frac{N_{MC}}{174 + N_{IS}} = \frac{20\,000}{1174} = 17,$$

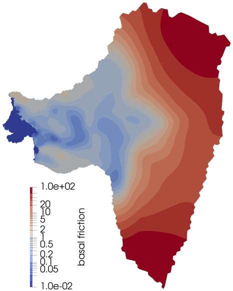
the importance sampling is 17 times faster.

The more extreme the event, the more samples will be required for the Monte Carlo method and the more efficient the importance sampling will be compared to standard Monte Carlo.

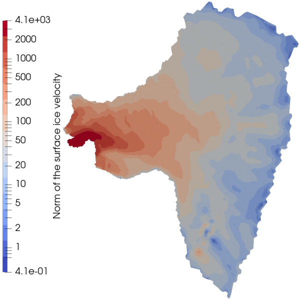
# Land ice test problem: Thwaites glacier

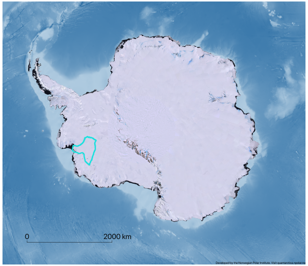
Mean case:

Basal friction:




Surface ice velocity:




courtesy of T. Hillebrand,
modified from Quantarctica

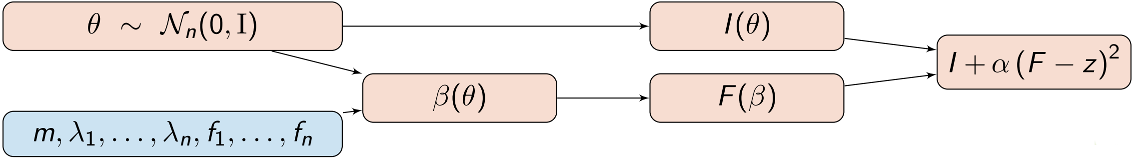
The velocity is faster if the friction is smaller, the quantity of interest is the flux at the grounding line; we expect the extreme events to be associated to smaller basal friction values.

# Land ice test problem: Thwaites glacier

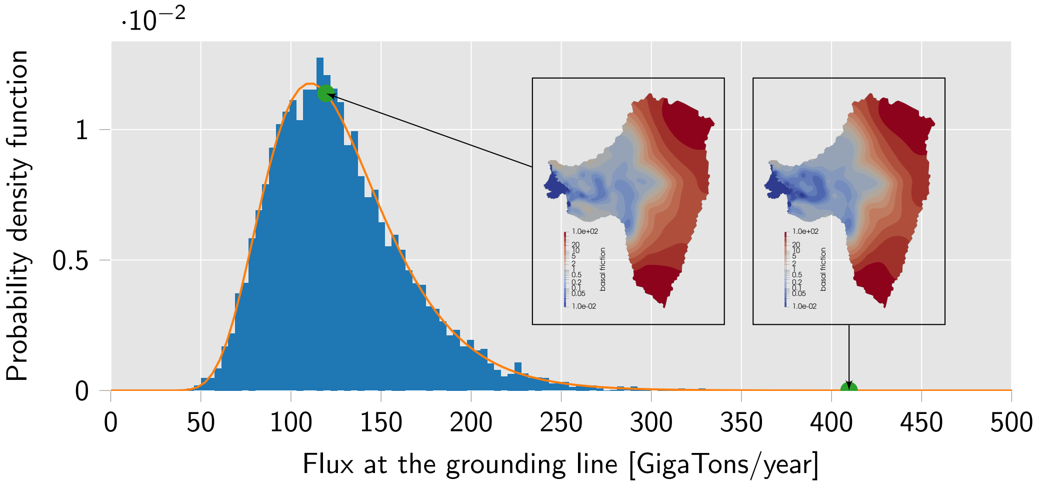
- Random parameter: the basal friction represented using a log-normal random field and a KL expansion with 50 modes:

$$
\begin{aligned}
\beta(\boldsymbol{x},\theta) &= e^{\log(m(\boldsymbol{x}))+\sum_{i=0}^{n}\sqrt{\lambda_i}\varepsilon_i(\theta)f_i(\boldsymbol{x})}, \\
\int_\Omega \mathrm{Cov}(\boldsymbol{x},\boldsymbol{x}')f_i(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' &= \lambda_i\, f_i(\boldsymbol{x}), \\
\mathrm{Cov}(\boldsymbol{x},\boldsymbol{x}') &= \sigma\, e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|}{\ell}},
\end{aligned}
$$

  with $\sigma = 0.3$, $\ell = 50\,\mathrm{km}$, and $n = 50$. Those modes allow to capture 99% of the variance.
- Quantity of interest: flux at the grounding line.
- PDE: steady state first order Stokes equation, Blatter-Pattyn model.

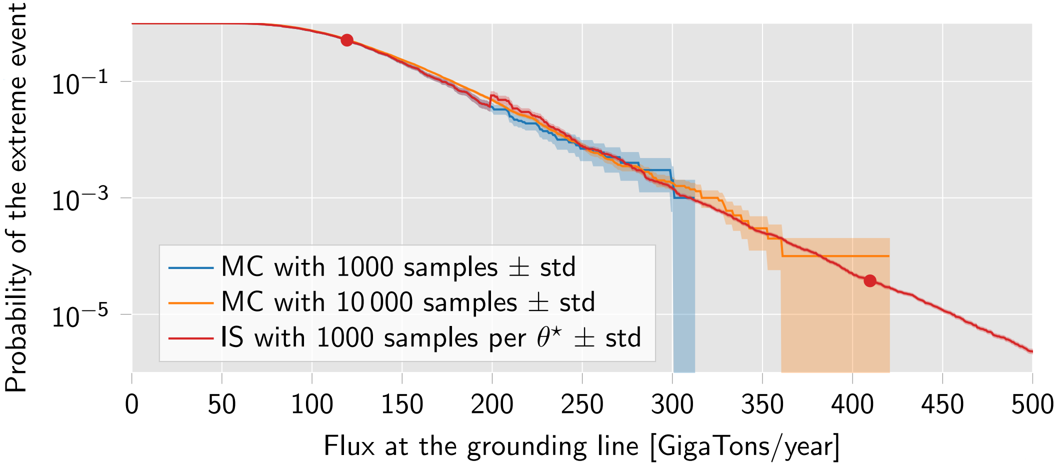


# Quantity of interest and random samples




Histogram of 10 000 samples, the orange curve is a log-normal distribution that fits the data the best.

# Probability of the extreme events




For smaller values of $z$, all the methods are consistent. When increasing $z$, the MC approach needs more and more samples to be consistent with the importance sampling.

# Performance

Average cost per simulation (measured on Skylake):

| | Wall-clock time | Relative cost |
|---|---|---|
| To compute $F(\theta)$ | $c_F = 39$ sec | 1 |
| To compute $\theta^\star(z)$ | $c_{\theta^\star} = 11842$ sec | 304 |

Expected cost per method:

| | $\mathbb{E}[c]$ | $\mathbb{E}[c/c_F]$ |
|---|---|---|
| Monte Carlo | $N_{MC}\, c_F$ | $N_{MC}$ |
| Importance Sampling per $\theta^\star$ | $c_{\theta^\star} + N_{IS}\, c_F$ | $304 + N_{IS}$ |

Comparison of Monte Carlo with $N_{MC} = 10\,000$ and importance sampling with $N_{IS} = 1000$:

$$\frac{N_{MC}}{304 + N_{IS}} = \frac{10\,000}{1304} = 7.66,$$

the importance sampling is 7 times faster.

The more extreme the event, the more samples will be required for the Monte Carlo method and the more efficient the importance sampling will be compared to standard Monte Carlo.

# Conclusions, current and future work

**Conclusions:**

- Discussion of the usage of optimization strategies to compute the probability of extreme events,
- Discussion of the implementation using open source libraries and software,
- Computation of the extreme event probabilities of high fluxes at the grounding line of the Humboldt and Thwaites glacier,
- Performance comparison of the proposed approach with the standard Monte Carlo method.

**Current work:**

- Move towards transient analysis.

**Future work:**

- Consider solving the constrained problem instead of using a quadratic penalty method,
- Consider larger problems,
- Use characterization of the uncertainties computed using the inverse problem,
- Deduce probability of extreme sea level rise due to land ice mass loss for the future.