SAND2018-7606C

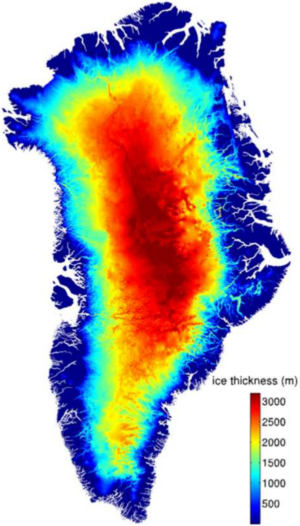
Optimal Bayesian Design of Borehole Locations for Inferring Past Ice Sheet Surface Temperatures

Andrew D. Davis, **Xun Huan**, Patrick Heimbach, Youssef M. Marzouk

Massachusetts Institute of Technology
Sandia National Laboratories

July 10, 2018

Motivation



Source: (left) National Snow & Ice Data Center, (middle) earth-chronicles.com, (right) Andrew D. Davis

Motivation

Ice sheets impact future global mean sea level

- Greenland: $\approx$ 6 meters   Antarctica: $\approx$ 60 meters

Source: National Geographic

Model calibration from field observations

Generally, we need to. . .

- develop and improve physics-based models
- calibrate models with observation data
- predict with quantified uncertainty

Past surface temperature:

- An indication of past climate and geological events
- Inferred from borehole data
- Drill, insert, equilibrate, measure

$\Rightarrow$ Overall process 2-4 years!

Some measurements are more useful than others

Borehole data acquisition is:

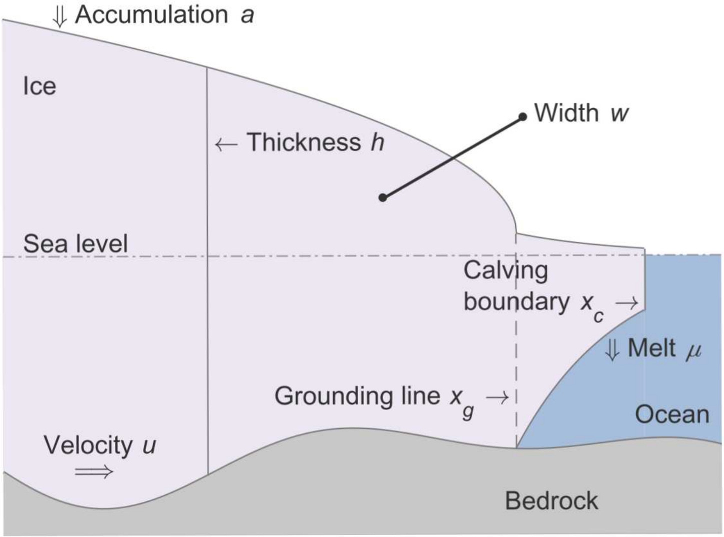
- time-consuming
- expensive
- dangerous

Goal of Optimal Experimental Design:

Select borehole locations that maximize the value of data for inferring past surface temperature

Outline

Anatomy of an ice stream



Three-part physical model

Three-part flowline model:

1. Compute grounded thickness $h(x,t)$ and velocity $\mathbf{u}(x,t)$ from stress and mass balance equations [Dupont 05, Hindmarsh 12, Jamieson 12, Pegler 13, Pegler 16]
2. Select a calving law and solve for the terminus $x_c(t)$ [Nick 10, Schoof 17]
3. Describe temperature $T(x,z,t)$ using advection-diffusion
   - Dirichlet boundary condition at the surface (assumed constant in x)

$$\dot{T} + \kappa\nabla^2 T + \mathbf{u}\cdot\nabla T = 0$$
$$T(x, h(x,t), t) = T_s(t;\theta)$$

Three-part physical model

Three-part flowline model:

1. Compute grounded thickness $h(x, t)$ and velocity $\mathbf{u}(x, t)$ from stress and mass balance equations [Dupont 05, Hindmarsh 12, Jamieson 12, Pegler 13, Pegler 16]
2. Select a calving law and solve for the terminus $x_c(t)$ [Nick 10, Schoof 17]
3. Describe temperature $T(x, z, t)$ using advection-diffusion
   - Dirichlet boundary condition at the surface (assumed constant in x)

$$\dot{T} + \kappa \nabla^2 T + \mathbf{u} \cdot \nabla T = 0$$
$$T(x, h(x,t), t) = T_s(t; \theta)$$

Forward problem: $\theta \to T(x, z, t)$

Inverse problem: noisy $T(\xi, :, t_f) \to \theta$

Optimal experimental design: *best* location ξ^* for noisy $T(\xi^*, :, t_f) \to \theta$

Parameterization of the surface temperature

Surface temperature field...

- assumed independent of x but changes with t
- modeled using a **Gaussian process** [Rasmussen 06]

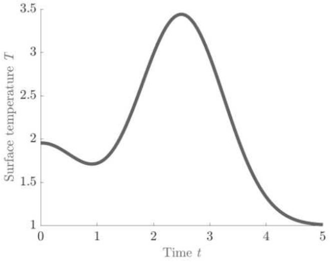
$$T_s(t;\theta) \approx \sum_{k=1}^{k_t} \sqrt{\theta_k} q_k(t) \zeta_k$$

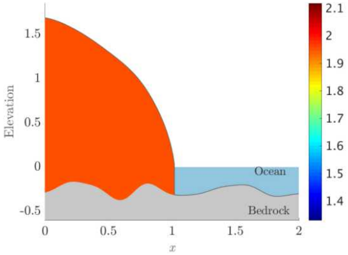
$$\text{with kernel} \quad K(t,t') = \exp\left(-\frac{\| t - t' \|_2^2}{2\ell^2}\right)$$

- k_t: number of terms after truncation (~ 10 in our study)
- K: covariance kernel (square-exponential; hyperparams fixed for now)
- θ_k: k^{th} eigenvalue of K
- $q_k(t)$: k^{th} eigenfunction of K
- ζ_k: $\stackrel{\text{iid}}{\sim} \mathcal{N}(0,1)$

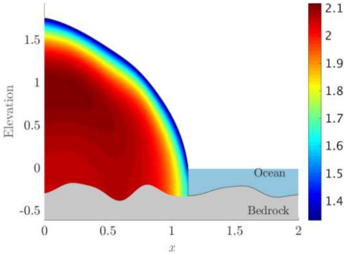
Sample modeling results

Surface temperature:





Initial condition (steady state)



Observable snapshot

Outline

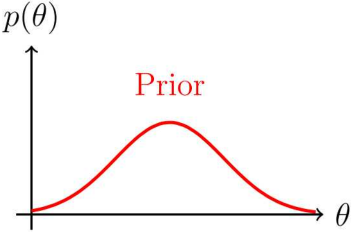
1. Physical model
2. Optimal Bayesian Experimental Design
3. Results

Bayesian inference

Characterize uncertainty using **Bayes' Theorem**:

$$\frac{\overbrace{p(\theta\quad)}^{prior}}{}$$

θ — parameters δ — noisy data

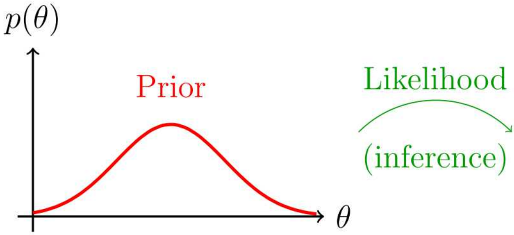


Bayesian inference

Characterize uncertainty using **Bayes' Theorem**:

$$\frac{\overbrace{p(\delta|\theta\quad)}^{\textit{likelihood}}\overbrace{p(\theta\quad)}^{\textit{prior}}}{}$$

θ — parameters δ — noisy data

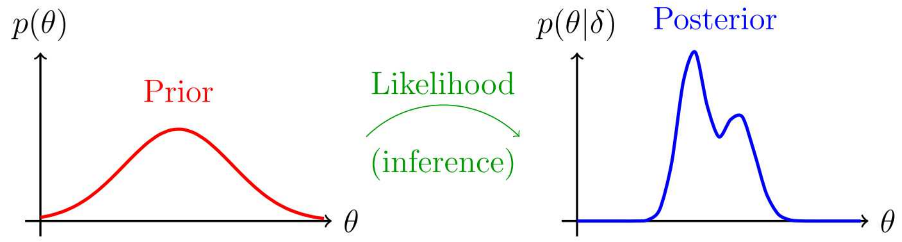


Bayesian inference

Characterize uncertainty using **Bayes' Theorem**:

$$\overbrace{p(\theta|\delta)}^{\textit{posterior}} = \frac{\overbrace{p(\delta|\theta)}^{\textit{likelihood}}\overbrace{p(\theta)}^{\textit{prior}}}{\underbrace{p(\delta)}_{\textit{evidence}}}$$

θ — parameters $\qquad \delta$ — noisy data

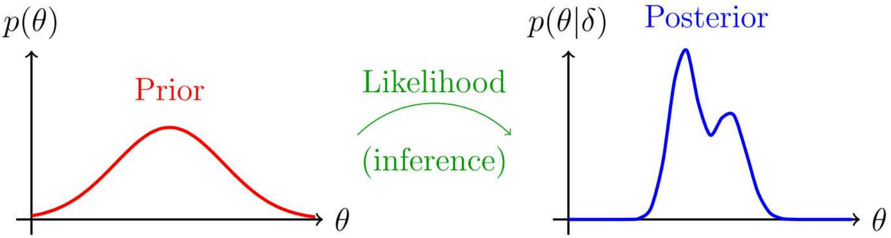


Bayesian inference

Characterize uncertainty using **Bayes' Theorem**:

$$\overbrace{p(\theta|\delta,\xi)}^{\textit{posterior}} = \frac{\overbrace{p(\delta|\theta,\xi)}^{\textit{likelihood}}\,\overbrace{p(\theta|\xi)}^{\textit{prior}}}{\underbrace{p(\delta|\xi)}_{\textit{evidence}}}$$

θ — parameters δ — noisy data ξ — design (experimental conditions)



Expected information gain

Expected **Kullback-Leibler (KL)** divergence between posterior and prior ($\iff$ **mutual information** between data and parameters) [Lindley 56]

$$U(\xi) = \mathbb{E}_{\delta|\xi}\left[D_{\mathrm{KL}}(p(\theta|\delta,\xi)||p(\theta|\xi))\right]$$

$$= \int_{\delta}\left[\int_{\theta}\ln\left[\frac{p(\theta|\delta,\xi)}{p(\theta)}\right]p(\theta|\delta,\xi)\,d\theta\right]p(\delta|\xi)\,d\delta$$

$$\approx \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[p(\delta^{(i)}|\theta^{(i)},\xi)\right]-\ln\left[p(\delta^{(i)}|\xi)\right]\right\}$$

$$p(\delta^{(i)}|\xi) \approx \frac{1}{M}\sum_{j=1}^{M}p(\delta^{(i)}|\theta^{(i,j)},\xi)$$ (nested Monte Carlo!) [Ryan 03]

U — expected utility

θ — parameters of interest

δ — noisy data

ξ — design variables

Expected information gain

Expected **Kullback-Leibler (KL)** divergence between posterior and prior ($\iff$ **mutual information** between data and parameters) [Lindley 56]

$$
\begin{aligned}
U(\xi) &= \mathbb{E}_{\delta|\xi}\left[D_{\mathrm{KL}}(p(\theta|\delta,\xi)||p(\theta|\xi))\right] \\
&= \int_\theta \int_\delta \{\ln[p(\delta|\theta,\xi)] - \ln[p(\delta|\xi)]\}\, p(\delta|\theta,\xi)\, d\delta\, p(\theta)\, d\theta \\
&\approx \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[p(\delta^{(i)}|\theta^{(i)},\xi)\right] - \ln\left[p(\delta^{(i)}|\xi)\right]\right\}
\end{aligned}
$$

$$
p(\delta^{(i)}|\xi) \approx \frac{1}{M}\sum_{j=1}^{M} p(\delta^{(i)}|\theta^{(i,j)},\xi) \quad \text{(nested Monte Carlo!)}
$$

[Ryan 03]

U — expected utility

θ — parameters of interest

δ — noisy data

ξ — design variables

Expected information gain

Expected **Kullback-Leibler (KL)** divergence between posterior and prior ($\iff$ **mutual information** between data and parameters) [Lindley 56]

$$
\begin{aligned}
U(\xi) &= \mathbb{E}_{\delta|\xi}\left[D_{\mathrm{KL}}(p(\theta|\delta,\xi)||p(\theta|\xi))\right] \\
&= \int_\theta \int_\delta \{\ln[p(\delta|\theta,\xi)] - \ln[p(\delta|\xi)]\}\, p(\delta|\theta,\xi)\, d\delta\, p(\theta)\, d\theta \\
&\approx \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[p(\delta^{(i)}|\theta^{(i)},\xi)\right] - \ln\left[p(\delta^{(i)}|\xi)\right]\right\}
\end{aligned}
$$

$$
p(\delta^{(i)}|\xi) \approx \frac{1}{M}\sum_{j=1}^{M} p(\delta^{(i)}|\theta^{(i,j)},\xi)
$$

(nested Monte Carlo!) [Ryan 03]

U — expected utility

θ — parameters of interest

δ — noisy data

ξ — design variables

Expected information gain

Expected **Kullback-Leibler (KL)** divergence between posterior and prior ($\Longleftrightarrow$ **mutual information** between data and parameters) [Lindley 56]

$$
\begin{aligned}
U(\xi) &= \mathbb{E}_{\delta|\xi}\left[D_{\mathrm{KL}}(p(\theta|\delta,\xi)||p(\theta|\xi))\right] \\
&= \int_\theta \int_\delta \{\ln[p(\delta|\theta,\xi)] - \ln[p(\delta|\xi)]\}\, p(\delta|\theta,\xi)\, d\delta\, p(\theta)\, d\theta \\
&\approx \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[p(\delta^{(i)}|\theta^{(i)},\xi)\right] - \ln\left[p(\delta^{(i)}|\xi)\right]\right\}
\end{aligned}
$$

$$
p(\delta^{(i)}|\xi) \approx \frac{1}{M}\sum_{j=1}^{M} p(\delta^{(i)}|\theta^{(i,j)},\xi)
$$

(nested Monte Carlo!) [Ryan 03]

Find optimal design: $\xi^* = \underset{\xi}{\operatorname{argmax}}\ U(\xi)$

Computationally prohibitive: $N = M = 10^3$ requires 10^6 forward evaluations

Another perspective on the expected utility

$$U(\xi) = \int_\theta \underbrace{\int_\delta \{\ln[p(\delta|\theta,\xi)] - \ln[p(\delta|\xi)]\}\, p(\delta|\theta,\xi)\, d\delta}_{f(\xi,\theta)}\, p(\theta)\, d\theta$$

$$\approx \frac{1}{N_1}\sum_{i_1=1}^{N_1} \underbrace{\frac{1}{N_2}\sum_{i_2=1}^{N_2}\left\{\ln\left[p(\delta^{(i_2)}|\theta^{(i_1)},\xi)\right] - \ln\left[p(\delta^{(i_2)}|\xi)\right]\right\}}_{\hat{f}(\xi,\theta)}$$

Another perspective on the expected utility

$$U(\xi) = \int_\theta \underbrace{\int_\delta \{\ln[p(\delta|\theta,\xi)] - \ln[p(\delta|\xi)]\} \, p(\delta|\theta,\xi)\, d\delta}_{f(\xi,\theta)} \, p(\theta)\, d\theta$$

$$\approx \frac{1}{N_1}\sum_{i_1=1}^{N_1} \underbrace{\frac{1}{N_2}\sum_{i_2=1}^{N_2} \left\{ \ln\left[p(\delta^{(i_2)}|\theta^{(i_1)},\xi)\right] - \ln\left[p(\delta^{(i_2)}|\xi)\right]\right\}}_{\hat{f}(\xi,\theta)}$$

with $\theta^{(i_1)} \sim p(\cdot)$, $\delta^{(i_2)} \sim p(\cdot|\theta,\xi)$ and an importance sampling estimate

$$p(\delta^{(i_2)}|\xi) \approx \sum_{j=1}^{M} p(\delta^{(i_2)}|\theta^{(i_2,j)},\xi) w^j \quad \text{with} \quad \theta^{(j)} \sim q(\cdot)$$

Another perspective on the expected utility

$$U(\xi) = \int_\theta \underbrace{\int_\delta \{\ln[p(\delta|\theta,\xi)] - \ln[p(\delta|\xi)]\}\, p(\delta|\theta,\xi)\, d\delta}_{f(\xi,\theta)}\, p(\theta)\, d\theta$$

$$\approx \frac{1}{N_1}\sum_{i_1=1}^{N_1} \underbrace{\frac{1}{N_2}\sum_{i_2=1}^{N_2}\left\{\ln\left[p(\delta^{(i_2)}|\theta^{(i_1)},\xi)\right] - \ln\left[p(\delta^{(i_2)}|\xi)\right]\right\}}_{\hat{f}(\xi,\theta)}$$

with $\theta^{(i_1)} \sim p(\cdot)$, $\delta^{(i_2)} \sim p(\cdot|\theta,\xi)$ and an importance sampling estimate

$$p(\delta^{(i_2)}|\xi) \approx \sum_{j=1}^{M} p(\delta^{(i_2)}|\theta^{(i_2,j)},\xi)w^j \quad \text{with} \quad \theta^{(j)} \sim q(\cdot)$$

- Triple-nested, but no new cost in integration
- Avoids surrogates that involve data δ (high-dimensional)
- Inner loops produce **noisy estimates** $\hat{f}(\theta,\xi) \approx f(\theta,\xi)$

Adaptive surrogate for Bayesian evidence

Given noisy estimates $\{\hat{f}_i(\theta_i, \xi_i)\}_{i=1}^{S}$ with known variance, we compute a **local polynomial surrogate** using regression

$$\eta(\theta, \xi) = \underset{m \in \mathcal{P}}{\operatorname{argmin}} \sum_{i=1}^{S} \left(m(\theta_i, \xi_i) - \hat{f}_i(\theta_i, \xi_i) \right)^2 \mathcal{K}(\theta_i, \xi_i; \theta, \xi)$$

Adaptive surrogate for Bayesian evidence

Given noisy estimates $\{\hat{f}_i(\theta_i, \xi_i)\}_{i=1}^{S}$ with known variance, we compute a **local polynomial surrogate** using regression

$$\eta(\theta, \xi) = \underset{m \in \mathcal{P}}{\text{argmin}} \sum_{i=1}^{S} \Big(m(\theta_i, \xi_i) - \hat{f}_i(\theta_i, \xi_i) \Big)^2 \mathcal{K}(\theta_i, \xi_i; \theta, \xi)$$

The error bound can be shown to be

$$\| \eta(\theta, \xi) - f(\theta, \xi) \| \leq C_S \Delta^{p+1}$$

Replacing f with η introduces a bias of $C_S \Delta^{p+1}$; choose S such that

$$\begin{aligned} [\text{Monte Carlo variance}] &\sim [\text{Surrogate bias}]^2 \\ C_{MC} N^{-1} &\sim C_s^2 \Delta^{2(p+1)} \end{aligned}$$

Adaptive surrogate for Bayesian evidence

Given noisy estimates $\{\hat{f}_i(\theta_i, \xi_i)\}_{i=1}^S$ with known variance, we compute a **local polynomial surrogate** using regression

$$\eta(\theta, \xi) = \underset{m \in \mathcal{P}}{\operatorname{argmin}} \sum_{i=1}^{S} \Big(m(\theta_i, \xi_i) - \hat{f}_i(\theta_i, \xi_i) \Big)^2 \mathcal{K}(\theta_i, \xi_i; \theta, \xi)$$

The error bound can be shown to be

$$\|\eta(\theta, \xi) - f(\theta, \xi)\| \leq C_S \Delta^{p+1}$$

Replacing f with η introduces a bias of $C_S \Delta^{p+1}$; choose S such that

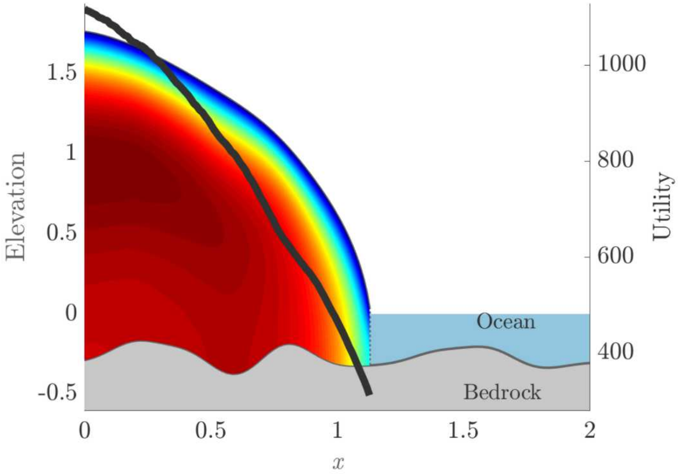
$$\begin{aligned} [\text{Monte Carlo variance}] &\sim [\text{Surrogate bias}]^2 \\ C_{MC} N^{-1} &\sim C_s^2 \Delta^{2(p+1)} \end{aligned}$$

- Regression averages noise from $\hat{f}$ so we can choose smaller N_2, M
- Choosing $S \approx 50$, $N_2 = M = 10$ requires $SN_2M \approx 5000 \ll 10^6$ model evaluations

Outline

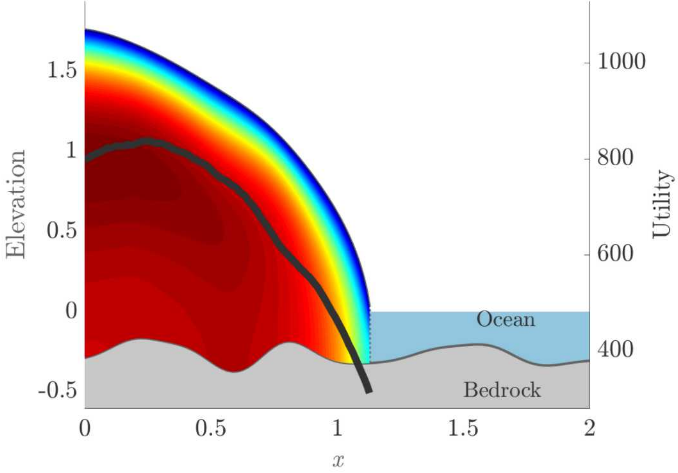
Expected information gain

$$U(\xi) = \mathbb{E}_{\delta|\xi}\left[D_{\mathrm{KL}}(p(\theta|\delta,\xi)||p(\theta|\xi))\right]$$



Expected information gain with inland hiking penalty

$$U(\xi) = \mathbb{E}_{\delta|\xi}\left[D_{\mathrm{KL}}(p(\theta|\delta,\xi)||p(\theta|\xi))\right] - 250(x_g - \xi)^2$$



Summary

Conclusions:

- Formulated borehole selection problem in a Bayesian setting and using optimal experimental design
- Constructed surrogate on information gain to significantly reduce computational costs from nested Monte Carlo
- Adapted surrogate based on bias-variance tradeoff
- Designed locations to maximize expected information gain on past surface temperature

Summary

Conclusions:

- Formulated borehole selection problem in a Bayesian setting and using optimal experimental design
- Constructed surrogate on information gain to significantly reduce computational costs from nested Monte Carlo
- Adapted surrogate based on bias-variance tradeoff
- Designed locations to maximize expected information gain on past surface temperature

Future work:

- Incorporate more complex physical model: geothermal heating, frictional heating, internal source/sinks
- Include more sophisticated cost/constraints within a 2D spatial setting
- Design multiple locations on single expedition

Acknowledgment

- Sandia National Laboratories[1]
- Massachusetts Institute of Technology
- U.S. Department of Energy
- National Science Foundation

[1] Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the presentation do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References I

T. K. Dupont & R. B. Alley.
Assessment of the importance of ice-shelf buttressing to ice-sheet flow.
Geophysical Research Letters, vol. 32, no. 4, pages 1–4, 2005.

Richard C. A. Hindmarsh.
An observationally validated theory of viscous flow dynamics at the ice-shelf calving front.
Journal of Glaciology, vol. 58, no. 208, pages 375–387, 2012.

Stewart S. R. Jamieson, Andreas Vieli, Stephen J. Livingstone, Colm Ó. Cofaigh, Chris Stokes, Claus Dieter Hillenbrand & Julian A. Dowdeswell.
Ice-stream stability on a reverse bed slope.
Nature Geoscience, vol. 5, no. 11, pages 799–802, 2012.

Dennis V. Lindley.
On a Measure of the Information Provided by an Experiment.
The Annals of Mathematical Statistics, vol. 27, no. 4, pages 986–1005, 1956.

F. M. Nick, C. J. Van Der Veen, A. Vieli & D. I. Benn.
A physically based calving model applied to marine outlet glaciers and implications for the glacier dynamics.
Journal of Glaciology, vol. 56, no. 199, pages 781–794, 2010.

Samuel S. Pegler, Katarzyna N. Kowal, Leonard Q. Hasenclever & M. Grae Worster.
Lateral controls on grounding-line dynamics.
Journal of Fluid Mechanics, vol. 722, pages 1–12, 2013.

Samuel S. Pegler.
The dynamics of confined extensional flows.
Journal of Fluid Mechanics, vol. 804, pages 24–57, 2016.

References II

Carl Edward Rasmussen & Christopher K. I. Williams.
Gaussian Processes for Machine Learning.
The MIT Press, Cambridge, MA, 2006.

Kenneth J. Ryan.
Estimating Expected Information Gains for Experimental Designs With Application to the Random Fatigue-Limit Model.
Journal of Computational and Graphical Statistics, vol. 12, no. 3, pages 585–603, 2003.

Christian Schoof, Andrew D. Davis & Tiberiu V. Popa.
Boundary layer models for calving marine outlet glaciers.
Cryosphere, vol. 11, no. 5, pages 2283–2303, 2017.