Multi-fidelity experimental design for ice-sheet simulation

Anonymous Author(s)
Affiliation
Address
email

1 Introduction

Computer simulations are an essential tool in many scientific fields from molecular dynamics [1] to aeronautics [2]. In glaciology, future predictions of sea level change require input from ice sheet models, which represent flow of the ice, mass changes due to accumulation of snow on the ice surface, and loss of ice through melting under floating ice shelves and iceberg calving. Ideally, these models need to be run over large geographical areas such as the Antarctic Ice Sheet and at sufficiently high resolution to accurately represent the dynamics of the ice flow [3; 4]. Moreover, due to uncertainties in the forcings and the parameter choices for such models, many different realisations of the model are needed to capture uncertainty in sea level change [5]. For these reasons, producing robust probabilistic forecasts from an ensemble of model simulations over regions of interest can be extremely expensive for many ice sheet models.

At the core of many numerical simulations lie differential equations approximated by numerical solvers that discretize the domain in time or space. While the accuracy of these solvers improves when step size is reduced, practitioners need to choose a size that is computationally feasible. The computational costs of running a simulator are intrinsically related to the accuracy of the model, leading to the question of finding the optimal trade-off between costs and accuracy. Multi-fidelity experimental design (MFED) is a strategy that models the high-fidelity output of a simulator by combining information from various resolutions in an attempt to minimize the computational costs of the process and maximize the accuracy of the posterior [6; 7; 8]. Intuitively, for some problems, running a simulator with a low resolution might be sufficient to characterize its behaviour since the underlying dynamics might not involve high-frequency effects.

In this paper, we present an application of MFED to an ice-sheet simulator [9] and demonstrate potential computational savings by modelling the relationship between spatial resolutions. We also analyze the behavior of MFED strategies using theoretical results from sub-modular maximization.
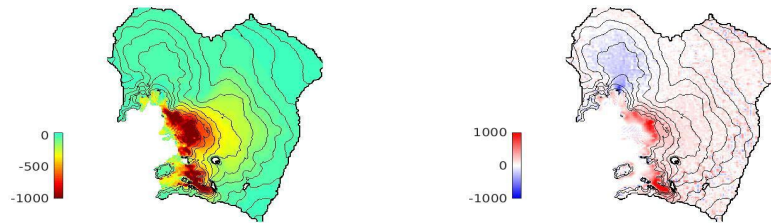

Figure 1: Simulation results using the WAVI ice sheet model over the Amundsen Sea Sector of the West Antarctic Ice Sheet (details in Appendix 5.3.2). Left: modelled ice thickness change (in metres) over 100 years of a WAVI simulation at 3km resolution. Right: the difference in modelled ice thickness at t=100 years between a 10 km and a 3km simulation.

2 Method

The goal of MFED is to model the behavior of a simulator $f : \theta \times \tau \to \mathbb{R}^D$ in a cost-efficient manner such that $\sum_i c(\theta_i, \tau_i) \leq C$. $\theta \in \mathbb{R}^N$ represents the parameters studied (e.g. temperature), $\tau \in [0, \infty)^M$ the fidelity of the simulator (0 representing the highest fidelity), c the non-negative cost-function of the simulator and C the total computational budget. In the context of the ice-sheet simulator, the resolution τ represents the spatial resolution used (1-10km). The fidelity parameter balances the accuracy of the simulator with the computational costs (discretization factor). The Bayesian approach to experimental design is to infer a surrogate model of a simulator f and then select points such as to minimize the model's uncertainty over the posterior distribution [10]. There exist two important elements to define in the strategy: the model and the objective.

Model A commonly used model in MFED is to define a probability distribution over a space of functions distributed according to a Gaussian process such that $p(f) = GP(\mu(\cdot); k(\cdot, \cdot))$ where $\mu(\cdot)$ is the mean function and $k(\cdot, \cdot)$ is the covariance function. The computational axis τ can be modeled as a discrete set of fidelities [6] or as a continuous one. In the discrete case, one strategy is to find a stationary relationship between the fidelities such as $f(\theta, 0) = \sin(f(\theta, 1)) + f(\theta, 2)$. However, in the context of physical simulators, it seems unlikely that there exists such stationary relationships between fidelities. Furthermore, the number of models would explode due to the continuous nature of the fidelities. By modelling f as a joint continuous space, the relationship between points of varying fidelities can be obtained through the covariance matrix. We consider kernels of the exponential quadratic families parameterized by a lengthscale β.

Cost-adjusted utility In experimental design, the aim is to reduce the uncertainty over our posterior distribution. In this work, we consider maximizing the reduction in conditional integrated variance (CIV) [11] such that the utility of adding a point $x_i = \{\theta_i, \tau_i\}$ is given by

$$U(x, X) = \int \sigma_X(\theta, \tau = 0)d\theta - \int \sigma_{x \cup X}(\theta, \tau = 0)d\theta. \tag{1}$$

where σ_X is the variance of a GP fitted to the set of points X. The conditional on $\tau = 0$ is due to the fact that we only care about the uncertainty over the highest fidelity. The aim of MFED is to reduce the uncertainty over the posterior while minimizing the computational budget used ($\sum_i c(x_i)$). Optimally reasoning about the budget allocation is an NP-hard combinatorial problem [12] as illustrated by the knapsack problem which consists of selecting items with assigned weights and value such as to maximize the cumulative value under a total weight constraint. In special cases (discussed in section 3), a reasonable approximation can be obtained using a cost-adjusted strategy such that each point is selected to maximize the cost-utility ratio $\frac{U(x)}{c(x)}$. This strategy was proposed in [6] and the related work are discussed in Appendix 5.1.

Uncertainty over objective One challenge is that CIV is dependent on the hyper-parameters of the GP and is unknown a priori. In particular, if we consider kernels from the exponential quadratic family, the lengthscale over the computational axis will have a large effect on the utility. Selecting a new observation that would maximize the CIV is thus challenging as the true lengthscale β^* is unknown. To alleviate this issue, we learn a posterior distribution over β at each time step and choose a specific β based on a upper-confidence bound parameter ν. In Section 3, we discuss the role of ν in trading-off exploration exploitation along the computational axis.

Algorithm 1 MF-UCB

Require: Confidence parameter ν, prior over lengthscale $\widetilde{p}(\beta)$, budget C
    while $C > 0$ **do**
        Select β_i such that $\widetilde{p}(\beta \geq \beta_i) \leq \nu$
        $x_i \leftarrow \text{argmax}_x \frac{U_{\beta_i}(x)}{c(x)}$
        Fit GP model with posterior distribution $\widetilde{p}(\beta)$ on $X \cup x_i$.
        $C \leftarrow C - c(x_i)$
    end while

---

3 Algorithm analysis

In this section, we analyze the performance of the cost-adjusted strategy from Section 2. First, we discuss the regret incurred by using the cost-adjusted greedy approximation. Then we analyze the *identification error* that captures the uncertainty existing over the optimal lengthscale. We showcase the existing underlying exploration-exploitation trade-off and its relationship with the confidence parameter ν introduced in Section 2.

MFED can be analyzed as a set-maximization problem if the input space is discretized. Formally, the goal is to find a set of points $\{x_i\} \in X$ that maximizes a utility set function $U : 2^X \to \mathbb{R}$ while satisfying a cost constraint $\sum_i c(x_i) \le C$. As a reminder, the algorithms described in Section 2 selects greedily the next point by maximizing the cost-adjusted variance reduction to form the set X^G. However, due to the uncertainty over the lengthscale, the algorithm cannot select the optimal sample x_i^G with respect to the greedy strategy and we call this cost-normalized error ϵ (definition 2).

Theorem 1 ([13]). *Let X^G be the greedy set, X^* the optimal set maximizing the monotone submodular function U, $\epsilon(x)$ the cost-normalized error and $c(x)$ the cost function. Then,*

$$U(X^G) > \underbrace{\left(1 - e^{-1}\right) U(X^*)}_{\textit{greedy approximation error}} - \underbrace{\sum_{i=1}^{L} \epsilon_i(x_i^G) c(x_i^G)}_{\textit{identification error}} \tag{2}$$

Greedy approximation The greedy error arises from the fact that finding the right combination of samples under a cost-constrained objective can be shown to be NP-hard (variant of the knapsack problem [14]). However, if we assume that the utility function U is sub-modular, the greedy approximation which sequentially selects points that maximize the utility-cost ratio defined in Section 2 is guaranteed to be within a reasonable distance from the true answer as per Theorem 1. Sub-modular functions (definition 1) are set functions with the special property that the value of adding an element to the set decreases as the size of the input set increase (diminishing returns). The conditional integrated variance is an example of a sub-modular monotone function (proof in [15]).

Identification error The identification error represents the uncertainty over the lengthscale which prevents the algorithm from selecting the greedy set X^G. As the posterior distribution over the lengthscale converges, the step-wise identification error converges to 0 since the correct model has been identified. The rate of convergence ϵ could be derived from the rate of convergence of the hyper-parameters of a GP under some assumptions [16]. However, the greedy strategy does not take into account the improvement over the lengthscale's uncertainty. To illustrate this, we study the following example where we can either sample a point x_0 with a cost of 2 or two points x_1, x_2 with a cost of 1 (illustrated in Figure 2). The greedy approximation compares the utility of $U(x_0)$ and $U(x_0, x_1)$ and makes the implicit assumption that the error on both actions are the same $2\epsilon(x_0) = \epsilon(x_1) + \epsilon(x_2)$. However, the error ϵ changes after the first time step due to the GP being fitted on new data and $\epsilon_1(x_2)$ may be smaller than $\epsilon_0(x_2)$ where ϵ_i denotes the error at time step i. We call $\Delta_i(x) = \epsilon_i(x) - \epsilon_{i+1}(x)$ the exploration bonus. This implies that MFED methods are underweighting lower fidelities by not taking into account the improvement over the lengthscale. One way to alleviate this issue, is to encourage exploration by selecting an upper bound on the posterior distribution over β. Practically, this can be done by increasing the confidence parameter ν. Intuitively, the values of ν and Δ should be correlated, where, reduction in identification error are encouraged using the exploration parameter. Selecting a confidence parameter close to $\nu = 0.5$ leads to a greedy behavior over the high fidelities as opposed to high ν which leads to more exploratory behavior. There exists a balance between two objectives; one, is learning the relationships between the fidelities and the second one, is maximizing the utility.
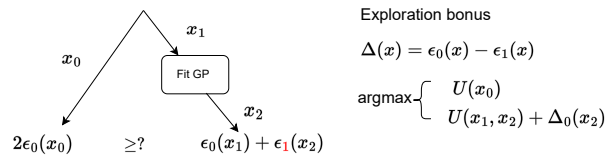

Figure 2: Example demonstrating the existing exploration-exploitation bonus.

4 Experiment

We consider the case of applying MFED to an ice sheet model [17]. The goal is to predict the contribution to the sea level from a specified region of Antarctica using the WAVI simulator. However, due to the unknown nature of some of the forcing parameters, we produce a probabilistic output over a varying set of input parameters. Specifically, we analyze the behavior of WAVI when varying the melting rate under the floating part of the ice sheet (the ice shelves). The melting rate is obtained using a parametrization of the basal melt rate described in [18; 19] (see Appendix 5.3).

Setup & analysis A pre-defined ensemble of runs over a varying range of melt rate parameters and resolutions (5 and 10km) were obtained. We analyze the behavior MF-UCB would have had in this constructed dataset. In Figure 3, we show that depending on some input parameters of WAVI (whether melt is allowed within partially floating cells), the lower fidelities can be either informative or not. When partial melting is considered, it becomes significantly more cost-efficient to query 10km resolutions, even though the most information could be gained by querying the high fidelity. In contrast, when partial melting is not considered, querying lower resolutions is not favourable, as the model has identified that the fidelities do not share information. CIV ensures that the lower fidelity is discarded in this case, which is reflected in the cost-utility surface of Figure 3. Concretely, the shape of the utility surface is modulated by the lengthscale learned (details in Appendix 5.4) by the GP (high lengthscale when partial melt is enabled, low otherwise). We also display the difference in utility surface when considering the cost-adjusted utility. The cost of running the higher resolution (5km) is approximately ten times more expensive than the 5km one which scales the utility in favour of the 10km resolution with partial melt.
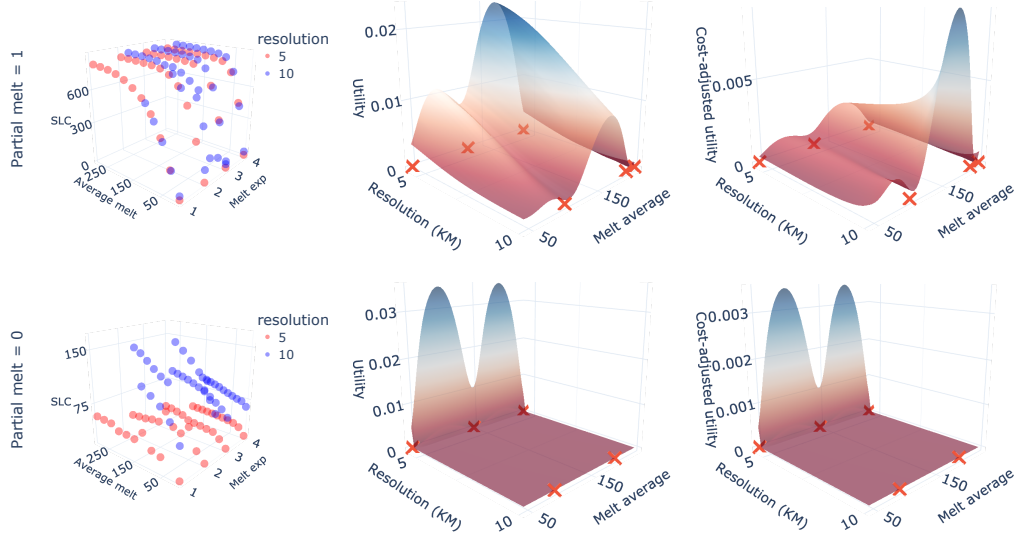

Figure 3: The left column shows the data distribution for varying average melt rate, melting exponent and resolution (5 and 10km). The middle column plots the variance reduction surface obtained when selecting a point after fitting a GP on 6 data points. The right columns displays the cost-adjusted utility. On the first row where partial melt is enabled, the lower resolution (10km) is informative and is a better cost-adjusted choice. On the contrary, when partial melt is disabled, the lower resolution is not informative and the algorithms learns to represent this aspect in the utility surface.

Discussion In this work, we investigate when the low-cost low-resolution simulations of an ice-sheet model can be used to predict the outcome of high-cost high-resolution runs. Our experiments summarized the information of the simulator into a single metric (SLC). However, this masks different ice-sheet behaviors that could prove informative. The next step of this project is to run the simulators based on the suggestions of MF-UCB on higher resolution (1-5km) and compare the performance of MFED using varying ν as well as other MFED methods. We hope to demonstrate in the future that MFED can be routinely incorporated in the glaciologists toolbox to obtain a cost-efficient probabilistic description of the change in sea level over a varying set of input parameters.

References

[1] Scott A Hollingsworth and Ron O Dror. Molecular dynamics simulation for all. *Neuron*, 99 (6):1129–1143, 2018.

[2] Domenico Quagliarella and Matteo Diez. An open-source aerodynamic framework for benchmarking multi-fidelity methods. In *AIAA AVIATION 2020 FORUM*, page 3179, 2020.

[3] Frank Pattyn, Christian Schoof, Laura Perichon, RCA Hindmarsh, Ed Bueler, B De Fleurian, Geoffroy Durand, Olivier Gagliardini, Rupert Gladstone, Dan Goldberg, et al. Results of the marine ice sheet model intercomparison project, mismip. *The Cryosphere*, 6(3):573–588, 2012.

[4] Frank Pattyn, Laura Perichon, Gaël Durand, Lionel Favier, Olivier Gagliardini, Richard CA Hindmarsh, Thomas Zwinger, Torsten Albrecht, Stephen Cornford, David Docquier, et al. Grounding-line migration in plan-view marine ice-sheet models: results of the ice2sea mismip3d intercomparison. *Journal of Glaciology*, 59(215):410–422, 2013.

[5] Isabel J Nias, Stephen L Cornford, Tamsin L Edwards, Noel Gourmelen, and Antony J Payne. Assessing uncertainty in the dynamical ice response to ocean warming in the amundsen sea embayment, west antarctica. *Geophysical Research Letters*, 46(20):11253–11260, 2019.

[6] Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, Damien Marquis, and Emmanuel Vazquez. Sequential design of multi-fidelity computer experiments: maximizing the rate of stepwise uncertainty reduction. *Technometrics*, 64(2):199–209, 2022.

[7] John D Jakeman, Sam Friedman, Michael S Eldred, Lorenzo Tamellini, Alex A Gorodetsky, and Doug Allaire. Adaptive experimental design for multi-fidelity surrogate modeling of multi-disciplinary systems. *International Journal for Numerical Methods in Engineering*, 123(12): 2760–2790, 2022.

[8] Xianliang Gong and Yulin Pan. Multi-fidelity bayesian experimental design to quantify extreme-event statistics. *arXiv preprint arXiv:2201.00222*, 2022.

[9] Alex Bradley and Robert Arthern. Wavi. jl: A fast, flexible, and friendly modular ice sheet model, written in julia. In *AGU Fall Meeting Abstracts*, volume 2021, pages C11A–03, 2021.

[10] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

[11] Alex Gorodetsky and Youssef Marzouk. Mercer kernels and integrated variance experimental design: connections between gaussian process regression and polynomial approximation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):796–828, 2016.

[12] Silvano Martello and Paolo Toth. *Knapsack problems: algorithms and computer implementations*. John Wiley & Sons, Inc., 1990.

[13] Matthew Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. *Advances in Neural Information Processing Systems*, 21, 2008.

[14] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[15] Andreas Krause, H Brendan McMahan, Carlos Guestrin, and Anupam Gupta. Robust submodular observation selection. *Journal of Machine Learning Research*, 9(12), 2008.

[16] Cheng Li. Bayesian fixed-domain asymptotics: Bernstein-von mises theorem for covariance parameters in a gaussian process model. *arXiv preprint arXiv:2010.02126*, 2020.

[17] Robert J Arthern, Richard CA Hindmarsh, and C Rosie Williams. Flow speed within the antarctic ice sheet and its controls inferred from satellite observations. *Journal of Geophysical Research: Earth Surface*, 120(7):1171–1188, 2015.

[18] Lionel Favier, Nicolas C Jourdain, Adrian Jenkins, Nacho Merino, Gaël Durand, Olivier Gagliardini, Fabien Gillet-Chaulet, and Pierre Mathiot. Assessment of sub-shelf melting parameterisations using the ocean–ice-sheet coupled model nemo (v3. 6)–elmer/ice (v8. 3). *Geoscientific Model Development*, 12(6):2255–2283, 2019.

[19] Paul R Holland, Adrian Jenkins, and David M Holland. The response of ice shelf basal melting to variations in ocean temperature. *Journal of Climate*, 21(11):2558–2572, 2008.

[20] Jerome L Myers. Fundamentals of experimental design. 1972.

[21] Howard J Seltman. Experimental design and analysis, 2012.

[22] V Roshan Joseph. Space-filling designs for computer experiments: A review. *Quality Engineering*, 28(1):28–35, 2016.

[23] Zongyu Geng, Feng Yang, Xi Chen, and Nianqiang Wu. Gaussian process based modeling and experimental design for sensor calibration in drifting environments. *Sensors and Actuators B: Chemical*, 216:321–331, 2015.

[24] Felipe AC Viana, Gerhard Venter, and Vladimir Balabanov. An algorithm for fast optimal latin hypercube design of experiments. *International journal for numerical methods in engineering*, 82(2):135–156, 2010.

[25] Alexander IJ Forrester, András Sóbester, and Andy J Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the royal society a: mathematical, physical and engineering sciences*, 463(2088):3251–3269, 2007.

[26] Shifeng Xiong, Peter ZG Qian, and CF Jeff Wu. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1):37–46, 2013.

[27] Julien Villemonteix, Emmanuel Vazquez, and Eric Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534, 2009.

[28] Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, 2012.

[29] Jialin Song, Yuxin Chen, and Yisong Yue. A general framework for multi-fidelity bayesian optimization with gaussian processes. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3158–3167. PMLR, 2019.

[30] Jian Wu, Saul Toscano-Palmerin, Peter I Frazier, and Andrew Gordon Wilson. Practical multi-fidelity bayesian optimization for hyperparameter tuning. In *Uncertainty in Artificial Intelligence*, pages 788–798. PMLR, 2020.

[31] Kirthevasan Kandasamy, Gautam Dasarathy, Jeff Schneider, and Barnabás Póczos. Multi-fidelity bayesian optimisation with continuous approximations. In *International Conference on Machine Learning*, pages 1799–1808. PMLR, 2017.

[32] Xu He, Rui Tuo, and CF Jeff Wu. Optimization of multi-fidelity computer experiments via the eqie criterion. *Technometrics*, 59(1):58–68, 2017.

[33] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[34] Tze Leung Lai, Herbert Robbins, et al. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[35] Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability*, 3:71–104, 2014.

[36] Peter Fretwell, Hamish D Pritchard, David G Vaughan, Jonathan L Bamber, Nicholas E Barrand, R Bell, C Bianchi, RG Bingham, Donald D Blankenship, G Casassa, et al. Bedmap2: improved ice bed, surface and thickness datasets for antarctica. *The Cryosphere*, 7(1):375–393, 2013.

[37] Romain Millan, Eric J Rignot, Mathieu Morlighem, Anders A Bjork, Jeremie Mouginot, and Michael Wood. Bathymetry and retreat of southeast greenland glaciers from operation ice-bridge and ocean melting greenland data. In *AGU Fall Meeting Abstracts*, volume 2017, pages C14A–02, 2017.

[38] Thomas Slater, Andrew Shepherd, Malcolm McMillan, Alan Muir, Lin Gilbert, Anna E Hogg, Hannes Konrad, and Tommaso Parrinello. A new digital elevation model of antarctica derived from cryosat-2 altimetry. *The Cryosphere*, 12(4):1551–1562, 2018.

[39] Benjamin Smith, Helen A Fricker, Nicholas Holschuh, Alex S Gardner, Susheel Adusumilli, Kelly M Brunt, Beata Csatho, Kaitlin Harbeck, Alex Huth, Thomas Neumann, et al. Land ice height-retrieval algorithm for nasa's icesat-2 photon-counting laser altimeter. *Remote Sensing of Environment*, 233:111352, 2019.

6

[40] Jeremie Mouginot, Eric Rignot, Bernd Scheuchl, and Romain Millan. Comprehensive annual ice sheet velocity mapping using landsat-8, sentinel-1, and radarsat-2 data. *Remote Sensing*, 9 (4):364, 2017.

[41] GPy. GPy: A gaussian process framework in python. http://github.com/SheffieldML/GPy, since 2012.

[42] Andrei Paleyes, Mark Pullin, Maren Mahsereci, Cliff McCollum, Neil Lawrence, and Javier González. Emulation of physical processes with emukit. In *Second Workshop on Machine Learning and the Physical Sciences, NeurIPS*, 2019.

5 Appendix

5.1 Related work

Experimental design: There exists rich literature that studies the optimal placement of samples in experimental design [20; 21]. The two main strategies are to either use a space-filling design [22] or a model-based strategy [23]. With the space-filling strategy, the goal is to find a set of points that maximize the distance between points [24]. In contrast, the model-based strategy casts a (probabilistic) model over the space studied and chooses points to minimize its uncertainty over the model [23]. In model-based experimental design, the selection strategy is often divided into two phases. The first one focuses on finding points that are useful for learning the model. The second phase focuses on maximizing some defined utility functions such as the integrated variance.

Multi-fidelity experimental design: Algorithm in multi-fidelity experimental can be divided into several categories based on whether they use a model and how the fidelities are analyzed. For example, [25] proposed a model-based two-stage process where several simulations are run at low fidelity and then use a sequential strategy to select high fidelity points. Another example is [26] which uses hypercubes for sampling and doubles the fidelity until a satisfying criterion is reached. In this paper, we build our algorithm from the maximum rate of uncertainty reduction of [6; 27; 28]. The main algorithmic difference lies in the fact that we select the lengthscale based on the upper bound ν of a learned distribution over the lengthscale. The justification for this modification comes from the fact that selecting the median lengthscale underestimates lower fidelities as explained in Section 3. Practically, this provides an exploratory phase into the algorithms that are often hard-coded by researchers. We also provide theoretical insights into the performance of the algorithm using sub-modular function.

Multi-fidelity Bayesian optimization: In multi-fidelity bayesian optimization, the goal is to find the minima of a function by leveraging lower fidelities [29; 30]. The concept of multi-fidelity is more popular in bayesian optimization than in experimental design. The knapsack analysis for multi-fidelity while originating from experimental design has been used to understand multi-fidelity bayesian optimization behavior. There have been several works using a cost-adjusted strategy in conjunction with continuous modelling over the computational axis [31; 32]. The main difference between BO and experimental design lies in the utility function maximized. In BO, the only thing that matter is the value of the output of the simulator. This is in contrast to experimental design which is agnostic to the output value of the simulator.

Online learning: Online learning [33] is a field that focuses on analyzing the behavior of decision-making strategies. The simplest setting considers comparing the cumulative reward of a discrete set of actions where the rewards are sampled from a Gaussian with an unknown mean. In this setting, several algorithms such as upper confidence bound [34] can be guaranteed to have an optimal regret. Optimality in this setting means that the regret grows at a logarithmic rate. The essential insight behind UCB is optimism in the face of uncertainty principle. Practically, this means that the algorithm will over-estimate arms with high uncertainty to encourage exploration. A similar strategy is used in our algorithm where the lengthscale is selected in an optimistic fashion (lower fidelity) until proven wrong. We suspect similar optimality results than online learning could be obtained in our setting.

Sub-modular maximization: The analysis performed in this field falls under the umbrella of combinatorial optimization. The main idea is to take variants of combinatorics problems such as knapsack and endow them with stronger assumptions (such as sub-modularity) to achieve reasonable approximations. The work developed by [35] is particularly relevant for two reason: one, they analyze a noisy setting which enables the incorporation of uncertainty over the lengthscale. Second, many of their applications are on experimental design which enables us to leverage some of their theoretical results concerning the integrated variance. Finally in several papers, they also make a similar connection than us to online learning [15].

5.2 Sub-modular function

Definition 1 (Sub-modular function)**.** *For every $X, Y \subseteq \Omega$ with $X \subseteq Y$ and $\forall x \in \Omega \setminus Y$, we have that*

$$U(X \cup x) - U(X) \geq U(Y \cup x) - U(Y) \tag{3}$$

Definition 2 ([13])**.** *For a cost function c and a sub-modular function U, we define the additive error $\epsilon(x)$ as:*

$$\epsilon(x) = max_y \frac{U(X \cup y)}{c(y)} - \frac{U(X \cup x)}{c(x)} \tag{4}$$

Caveat theorem 1 The caveat in this bound is that it only applies to specific T that are given by applying the greedy strategy. Practically, you take one element at a time and add up the τ giving you the T for which the theorem holds. Extensions can be found in [citation] to relax this assumption.

5.3 Ice-sheet simulator

The simulator used in this paper is called WAVI (wavelet-based, adaptive-grid, vertically integrated ice sheet model) [17]. It is a vertically integrated, three dimensional ice sheet model which includes both the membrane stresses in the ice and the effects of vertical shear in order to simulate flow of both grounded and floating ice. A subgrid parameterisation is used to represent the movement of the grounding line (the boundary between the grounded and floating ice). Since ice retreat has shown to be sensitive to how melting is applied in cells containing the grounding line (e.g. Seroussi and Morlighem, 2018), a subgrid melting parameterisation is applied, where melt can be applied in proportion to the floating area of the cell (partial melt=1), or melt can only be prescribed in cells that are fully floating (partial melt=0). There is no calving in the WAVI model (the ice shelf front is fixed and cannot advance or retreat). For a complete introduction to the simulator refer to [17].

The model is initialised via data assimilation, inverting for basal drag and viscosity using accumulation, surface velocities and rates of change of surface elevation [17]. The data used to initialise the simulator for this experiment includes bathymetry from [36], modified to account for data from [37], DEM from Cryosat2 [38], updated from 2013 to 2017 using dh/dt, dh/dt from ICESat-2 [39], and surface velocities from annual MEaSUREs 16/17 [40]. Other data sets are as in [17]. The initial state thus corresponds to the ice sheet in approximately 2017. For the simulations included in this experiment, we run the model only over the sector of the West Antarctic Ice Sheet that drains into the Amundsen Sea, as shown in Figure 1.

While this information is enough to give some ideas of the data used, we intend to publish a full description and accompanying codebase at a later date once the work is more complete.

5.3.1 Parameters of interest

There exists many input parameters in the WAVI simulator that can be modified, such as parameters related to the prescription of basal melt rate, accumulation rate, and ice rheology. We restrict our analysis to a subset of parameters of major importance, namely those that control the basal melt forcing. In the absence of a coupled ice-ocean model, a basal melt parameterisation is used to prescribe the melt rate under the floating part of the ice sheet (the ice shelves). In these simulations, the basal melt is parameterised using an expression described in [18; 19], which is expanded to consider different values of the exponent, denoted m. The parameter γ^T is calibrated to set the average basal melt at the start of the simulation, $t = 0$. Simulations with different combinations of m, average melt and resolution are then run forward for 100 years of simulation time, and the output that we aim to emulate is the total cumulative sea level contribution (SLC) after 100 years. Two sets of simulations are performed: one with no melt in partially floating cells (partial melt=0), and one with proportional melt in partially floating cells (partial melt=1), as described above.

5.3.2 Details simulation Figure 1

In Figure 1, the melt parameters are set as $m = 2$, average melt at $t = 0$ years is 54 m/a, and there is no melt in partially grounded cells (partial melt=0). The thickness from the 10 km simulation is interpolated onto the 3km grid, for comparison. In both plots, elevation contours for the 3km run at $t = 100$ years are shown at 200 metre intervals.

5.4 Training GP

The kernel used is an RBF kernel with a lengthscale for each dimension of the data initialized at the value 1. We use HMC to learn a posterior distribution over the lengthscale parameter β. HMC is run for 500 steps and the last 250 are used to approximate the posterior distribution. For visualization purposes we train the lengthscale on only 6 datapoints which can result in somewhat unstable training. However, for the real application of MFED on ice sheet, the total number of runs will easily exceed 50. The codebase is developed using GPy [41] and Emukit [42]. We normalize the input and output of the simulator and add a small jitter to the resolution column to avoid numerical instabilities. We use an upper confidence bound ν of 0.9.