Spatial Statistics based Prior Generation for Ocean Experimental Design

Vikram V. Garg

Abstract

Modern ocean modeling is reliant on data assimilation to inform critical, incompletely known model parameters. Acquisition of relevant data from the ocean is an expensive, time consuming and potentially risky exercise. In the absence of a rigorous, quantitative framework for data acquisition, it is unclear where data needs to be measured and at what time. Using tools from statistical experiment design, in combination with the simulation capabilities of ocean models, we can develop a framework for ocean experimental design. This document discusses the use of R-INLA, a Bayesian spatial modeling software, to represent prior knowledge of target ocean parameters, using spatially varying, nonstationary distributions via the state of the art R-INLA statistical toolkit.

Keywords: Experimental Design, Ocean Modeling, Spatial Statistics, Covariance Modeling

1 Mathematical Formulation

1.1 Expected Utility of an Experiment

We begin in a general setting, with a deterministic forward model $\mathbf{G} : \mathbb{R}^{n_\xi} \times \mathbb{R}^{n_C} \rightarrow \mathbb{R}^{n_y}$, which maps the parameters $\boldsymbol{\xi} \in \mathbb{R}^{n_\xi}$ and experimental conditions $\mathcal{C} \in \mathbb{R}^{n_C}$ into the observables $\mathbf{y} \in \mathbb{R}^{n_y}$ measured by the experiment,

$$\mathbf{y} = \mathbf{G}(\boldsymbol{\xi}, \mathcal{C}) \tag{1}$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is a sample space, $\mathcal{F}$ is a σ-field and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. The vector of parameters $\boldsymbol{\xi} : \Omega \rightarrow \mathbb{R}^{n_\xi}$ is thus a random variable, with potentially infinite dimension. Following [Huan and Marzouk, 2013], we have the following expression for the expected utility of an experiment at given experimental conditions $\mathcal{C}$,

$$U(\mathcal{C}) = \int_{\mathcal{Y}\times\boldsymbol{\xi}} (\ln(p[\mathbf{y}|\boldsymbol{\xi}, \mathcal{C})] - \ln[p(\mathbf{y}|\mathcal{C})])\ (p(\mathbf{y}|\boldsymbol{\xi}, \mathcal{C})p(\boldsymbol{\xi})d\boldsymbol{\xi}d\mathbf{y}) \tag{2}$$

This utility function represents the expected information gain in the parameter(s) $\boldsymbol{\xi}$ for given experimental conditions $\mathcal{C}$. The experimental conditions $\mathcal{C}^*$ for which the utility function is maximized represents the optimal experimental design,

$$\mathcal{C}^* = \underset{\mathcal{C}\in\mathbb{R}^{n_C}}{\arg\max}\, U(\mathcal{C}) \tag{3}$$

1.2 Arctic Argo Setting

For designing Argo experiments [Riser et al., 2016], we use MITgcm [Adcroft et al., 2008] as the forward model. The parameter we seek to design an experiment to learn is the ocean's isopycnal diffusivity [Forget et al., 2015], while the experimental conditions we seek to optimize are the initial positions of the Argo floats. In the future we can include a larger set of ocean model parameters such as the vertical diffusivity, to locate positions that optimally inform a combination of parameters.

The observables are the time series of the temperature and salinity measured by floats along their trajectories. The motion of such floats in the ocean can be chaotic, with high sensitivity to initial location. This chaotic behavior would in turn imply that the utility function is highly irregular, with no meaningful optima. Such behavior would necessitate averaging the utility function on some large length scale. However, the coarse resolution of MITgcm setup imparts diffusion to the flow and results in an averaged flow field which is relatively non-chaotic.

In addition to locating optimal locations for floats, the computational model of the ocean allows us to explore questions regarding the choice of observables that maximize information gain, and the length of time-series used for constructing the observables, thus helping guide future device design.

2 Methods and Algorithms

2.1 Numerical Approximation of the Utility Function

Following [Huan and Marzouk, 2013], we define the estimator $\hat{U}(\mathcal{C})$ for the utility function in (2) as follows,

$$\hat{U}(\mathcal{C}) = \frac{1}{N_{out}} \sum_{i=1}^{N_{out}} \left\{ \ln[p(\mathbf{y}^i|\boldsymbol{\xi}^i_{out}, \mathcal{C})] - \ln[p(\mathbf{y}^i|\mathcal{C})] \right\} \tag{4}$$

The joint samples $\left\{\boldsymbol{\xi}^i_{out}, \mathbf{y}^i\right\}_{i=1}^{N_{out}}$, are obtained from the prior density $p(\boldsymbol{\xi})$ and the conditional distribution (likelihood function) $p(\mathbf{y}|\boldsymbol{\xi} = \boldsymbol{\xi}^i_{out}, \mathcal{C})$. The evidence $p(\mathbf{y}^i|\mathcal{C})$ can be approximated as,

$$p(\mathbf{y}^i|\mathcal{C}) = \int_{\boldsymbol{\xi}} p(\mathbf{y}^i|\boldsymbol{\xi}, \mathcal{C}) p(\boldsymbol{\xi})\, d\boldsymbol{\xi} \approx \frac{1}{N_{in}} \sum_{j=1}^{N_{in}} p(\mathbf{y}^i|\boldsymbol{\xi} = \boldsymbol{\xi}^j_{in}, \mathcal{C}) \tag{5}$$

where the samples $\{\boldsymbol{\xi}^j_{in}\}_{j=1}^{N_{in}}$ are again drawn from the prior density $p(\boldsymbol{\xi})$. The combined expression for the estimator $\hat{U}(\mathcal{C})$ is then,

$$\hat{U}(\mathcal{C}) = \frac{1}{N_{out}} \sum_{i=1}^{N_{out}} \left\{ \ln[p(\mathbf{y}^i|\boldsymbol{\xi}^i_{out}, \mathcal{C})] - \ln \left[\frac{1}{N_{in}} \sum_{j=1}^{N_{in}} p(\mathbf{y}^i|\boldsymbol{\xi} = \boldsymbol{\xi}^j_{in}, \mathcal{C}) \right] \right\} \tag{6}$$

2.1.1 Sample Reuse

At the cost of a small increase in bias [Huan, 2010], one may reuse the samples generated for the outer sum $\{\boldsymbol{\xi}^i_{out}\}_{i=1}^{N_{out}}$ for the evaluation of the evidence in (5). Then we have $N_{in} = N_{out} = N_s$, $\{\boldsymbol{\xi}^i_{in}\}_{i=1}^{N_{in}} = \{\boldsymbol{\xi}^i_{out}\}_{i=1}^{N_{out}} = \{\boldsymbol{\xi}^i\}_{i=1}^{N_s}$ the expression for the estimator $\hat{U}(\mathcal{C})$ in (6)

becomes,

$$\hat{U}(\mathcal{C}) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left\{ \ln[p(\mathbf{y}^i|\boldsymbol{\xi}^i, \mathcal{C})] - \ln \left[\frac{1}{N_s} \sum_{j=1}^{N_s} p(\mathbf{y}^i|\boldsymbol{\xi} = \boldsymbol{\xi}^j, \mathcal{C}) \right] \right\} \tag{7}$$

The above Eq (7) is the basis for our approximation of the expected utility in the ocean modeling setting.

2.1.2 Flow Computation

The calculation of the utility function via the Monte Carlo method is computationally challenging, due to repeated evaluations of the likelihood function $p(\mathbf{y}^i|\boldsymbol{\xi}^i, \mathcal{C})$. The float trajectories, which determine the observations $\mathbf{y}^i$, depend on the computed flow, which in turn depends on the input parameter parameter $\boldsymbol{\xi}^i$.

2.2 Utility Function Optimization

Once a suitable method for computing the expected utility at an arbitrary design condition is chosen, one can move on to the problem specified by (3). Since gradients are usually available, various gradient based stochastic optimization strategies can be used [Spall, 1998, Nelder and Mead, 1965].

In the specific case of the ocean modeling setting we are considering, a relatively small number of design conditions $\{\mathcal{C}^m\}_{m=1}^{N_m}$ can be probed to locate a maximum. This removes the need for applying gradient based strategies, although these and nested evaluation algorithms can be considered later.

2.3 Algorithms and Complexity

Algorithm 1 details a step-by-step method to estimate the design conditions given by (3).

Let \$ denote the cost of a computation. The total cost of Algorithm 1 can be expressed as,

---

Algorithm 1 Given a reference parameter value $\boldsymbol{\xi}^{ref}$, a vector of initial float positions $\{\mathcal{C}^m\}_{m=1}^{N_m}$, a prior distribution $p(\boldsymbol{\xi})$ and likelihood function $p(\mathbf{y}|\boldsymbol{\xi}, \mathcal{C})$, obtain an optimal float position $\mathcal{C}^*$ as per (3).

Obtain the state and tracer vectors $\mathbf{U}^{ref}$, $\mathbf{V}^{ref}$, T^{ref} and S^{ref} for the reference parameter value $\boldsymbol{\xi}^{ref}$ using MITgcm.
for Each initial position: $m = 1, m \leq N_m$ **do**
    Use $\mathbf{U}^{ref}$ and $\mathbf{V}^{ref}$ to obtain the float trajectory $\{f^{m,t}\}_{t=1}^{N_t}$ for initial postion $\mathcal{C}^m$
    Obtain the reference observables $y^{m,ref} \leftarrow \{T(f^{m,t}), S(f^{m,t})\}_{t=1}^{N_t}$
end for
Obtain samples $\{\boldsymbol{\xi}^i\}_{i=1}^{N_s}$ from $p(\boldsymbol{\xi})$
for Each sample $\boldsymbol{\xi}^i$: $i = 1, i \leq N_s$ **do**
    Use MITgcm offline with $\boldsymbol{\xi} = \boldsymbol{\xi}^i$ to compute $\tilde{T}^i$ and $\tilde{S}^i$
    for Each initial position: $m = 1, m \leq N_m$ **do**
        $\tilde{\mathbf{y}}^{m,i} \leftarrow \{\tilde{T}^i(f^{m,t}), \tilde{S}^i(f^{m,t})\}_{t=1}^{N_t}$
        Use $\tilde{\mathbf{y}}^{m,i}$ to sample $p(\mathbf{y}|\boldsymbol{\xi} = \boldsymbol{\xi}^i, \mathcal{C})$ and obtain $\mathbf{y}^{m,i}$
    end for
end for
for Each initial position: $m = 1, m \leq N_m$ **do**
    $\hat{U}^m \leftarrow 0$
    for Each sample $\boldsymbol{\xi}^i$: $i = 1, i \leq N_s$ **do**
        $E^i \leftarrow 0$
        for Each sample $\boldsymbol{\xi}^j$: $j = 1, j \leq N_s$ **do**
            $E^i = E^i + p(\mathbf{y}^{m,i}|\boldsymbol{\xi} = \boldsymbol{\xi}^j, \mathcal{C}^m)$
        end for
        $E^i \leftarrow \frac{E^i}{N_s}$
        $\hat{U}^m = \hat{U}^m + ln[p(\mathbf{y}^{m,i}|\mathcal{C}^m)] - ln[E^i]$
    end for
    $\hat{U}^m \leftarrow \frac{\hat{U}(m)}{N_s}$
end for
Return $\arg\max\{\hat{U}^m\}_{m=1}^{N_m}$

(1 $\times$ \$MITgcm) + (1 $\times$ \$$N_s$ samples from $p(\boldsymbol{\xi})$) + (N_m $\times$ \$Offline Float Propagation) + (N_s $\times$ \$Offline Tracer Calculation).

3 Covariance Modeling

The experimental design methodology proposed in Section 2 requires the specification of a prior density on the measured parameter(s) $\boldsymbol{\xi}$, which in our case is the infinite dimension turbulent mixing viscosity D_{kr}. We assume that the prior density for D_{kr} can be described as a Gaussian Process (GP). This GP has a given mean field $\mu_{D_{kr}}$, but the structure of its covariance field $\text{Cov}_{D_{kr}}(\mathbf{x}, \mathbf{y})$ remains to be determined.

The covariance fields of GP's are characterised by properties of stationarity, isotropy, smoothness and periodicity. It is natural to assume that $\text{Cov}_{D_{kr}}$ is smooth and displays no specific periodicity. However, further assuming stationarity and isotropy proves to be rather restrictive and non-physical. For example, assuming the classical squared exponential covariance structure leads to realizations of D_{kr} that are either too uncorrelated or over correlated.

We can conclude that the assumptions of stationarity and isotropy are not admissible for D_{kr}. Indeed, it is easy to see that the stationarity assumption is invalid, since expert opinion informs us that D_{kr} has local regions of high correlation (near ocean surface and the deep bottom for example), which are essentially uncorrelated from each other. Any assumption of an isotropic D_{kr} also wont hold since turbulent flows usually show clear directional dependence.

Hence, we have to model $\text{Cov}_{D_{kr}}(\mathbf{x}, \mathbf{y})$ as a nonstationary GP. This is the objective of

our next section.

3.1 Specifying and sampling a nonstationary covariance model for D_{kr}

We model the uncertain D_{kr} field as,

$$D_{kr} = 10^{\mu_{D_{kr}} + \Delta D_{kr}(\mathbf{x};\omega)} \tag{8}$$

where $\mu_{D_{kr}}$ is a scalar mean exponent, and the additional exponent $\Delta D_{kr}(\mathbf{x};\omega)$ is a spatially varying random field. The random perturbation $\Delta D_{kr}(\mathbf{x};\omega)$ is constrained to lie in the interval,

$$\Delta D_{kr}(\mathbf{x};\omega) \in [-\infty, 3] \tag{9}$$

and the mean exponent $\mu_{D_{kr}} = -5$. Therefore, D_{kr} lies in the interval,

$$D_{kr} \in [0, 10^{-2}] \tag{10}$$

The spatial stochastic model for $\Delta D_{kr}(\mathbf{x};\omega)$ is determined by a stochastic partial differential equation (SPDE) as specified in [Lindgren et al., 2011]. The generalized second order SPDE is written as,

$$(\kappa(\mathbf{x}) - \Delta)(\tau(\mathbf{x})u(\mathbf{x})) = \mathcal{W}(\mathbf{x}),\ \mathbf{x} \in \Omega \tag{11}$$

For the stationary version of this SPDE ($\kappa(\mathbf{x})$ and $\tau(\mathbf{x})$ constant), the solution covariance

is a generalized Matern kernel of the form,

$$\text{cov}(u(\mathbf{0}), u(\mathbf{x})) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\kappa\|\mathbf{x}\|\right)^{\nu} K_{\nu}\left(\kappa\|\mathbf{x}\|\right) \tag{12}$$

where $\sigma^2 = \frac{\Gamma\nu}{\Gamma(\alpha)(4\pi)^{\frac{d}{2}}\kappa^{2\nu}\tau^2}$. Although we use a nonstationary SPDE to model $\Delta D_{kr}(\mathbf{x};\omega)$, the covariance expression for the stationary model given above illustrates the roles of key SPDE parameters κ and τ in the properties of the resulting covariance.

We only need to specify two parameters of the general covariance solution of the SPDE in [Lindgren et al., 2011]. These parameters, denoted as $\kappa(\mathbf{x})$ and $\tau(\mathbf{x})$, represent the inverse of the pointwise correlation length and the pointwise marginal variance respectively. In our model, we postulate that both these parameters depend only on the pointwise bathymetry gradient. Regions with high bathymetry gradients have more active flow, hence low correlation lengths and high marginal variances, and large $\kappa(\mathbf{x})$ and small $\tau(\mathbf{x})$ accordingly. Regions with flatter bathymetry have quieter, more regular flow, hence high correlation lengths and low marginal variance, and and small $\kappa(\mathbf{x})$ and large $\tau(\mathbf{x})$ accordingly.

Our models for κ and τ are,

$$\kappa(\mathbf{x}) = \kappa_m e^{c_\kappa \|\nabla b(\mathbf{x})\|} \tag{13}$$

$$\tau(\mathbf{x}) = \tau_m e^{c_\tau \|\nabla b(\mathbf{x})\|} \tag{14}$$

The constants (κ_m, c_κ) and (τ_m, c_τ) can be determined using extremal values for the correlation lengths (L_{min},L_{max}) and marginal variances (V_{max},V_{min}) which correspond to the minimum and maximum values of the bathymetry gradient norm. To relate the correlation

lengths and κ, we employ the empirical relation [Lindgren and Rue, 2015],

$$\rho = \frac{\sqrt{8\nu}}{\kappa} \tag{15}$$

where ρ is the correlation length and ν is the order of the SPDE we employ (in our case, ν = 2).

In the model we use, we set the correlation length and marginal variance extremes as follows,

$$\begin{aligned} L_{max} &= 100km,\ L_{min} = 10km \\ \tau_{max} &= 0.1,\ \tau_{min} = 0.0001 \end{aligned} \tag{16}$$

The correlation length limits imply that where the bathymetry gradient is small, disturbances travel for upto a 100km before dissipating, whereas where the bathymetry varies rapidly disturbances travel only upto 10km. The values of τ_{max} and τ_{min} were determined by trial and error to achieve samples which lay in the range for $\Delta D_{kr}(\mathbf{x};\omega)$ as prescribed by (9).

3.2 Sampling nonstationary prior covariance using R-INLA

The R SPDE solver package INLA [Lindgren and Rue, 2015] is used to generate samples with the covariance structure prescribed by the model described in the previous section. R-INLA creates a Finite Element mesh, and nearest neighbor interpolation is used to calculate $\kappa(\mathbf{x})$ and $\tau(\mathbf{x})$ values at the nodes of the mesh. INLA then rapidly generates samples with the prescribed covariance structure.

Figure 1 shows plots of the land and sea masks, the bathymetry gradient magnitude

and the derived hyperparameters $\kappa(\mathbf{x})$ and $\tau(\mathbf{x})$. We point out here that the values of κ and τ are extremely large in the land area (white shading), which implies that these regions have extremely low correlation lengths and low marginal variances. This ensures that any perturbations in these regions do not propagate and pollute the sample values in the water region. We observe that the values of κ and τ track the magnitude of the bathymetry gradient, regions with high gradient have low correlation lengths (high κ) and high marginal variances (low τ) and vice versa. Figure 2 shows a sample generated by INLA. The land

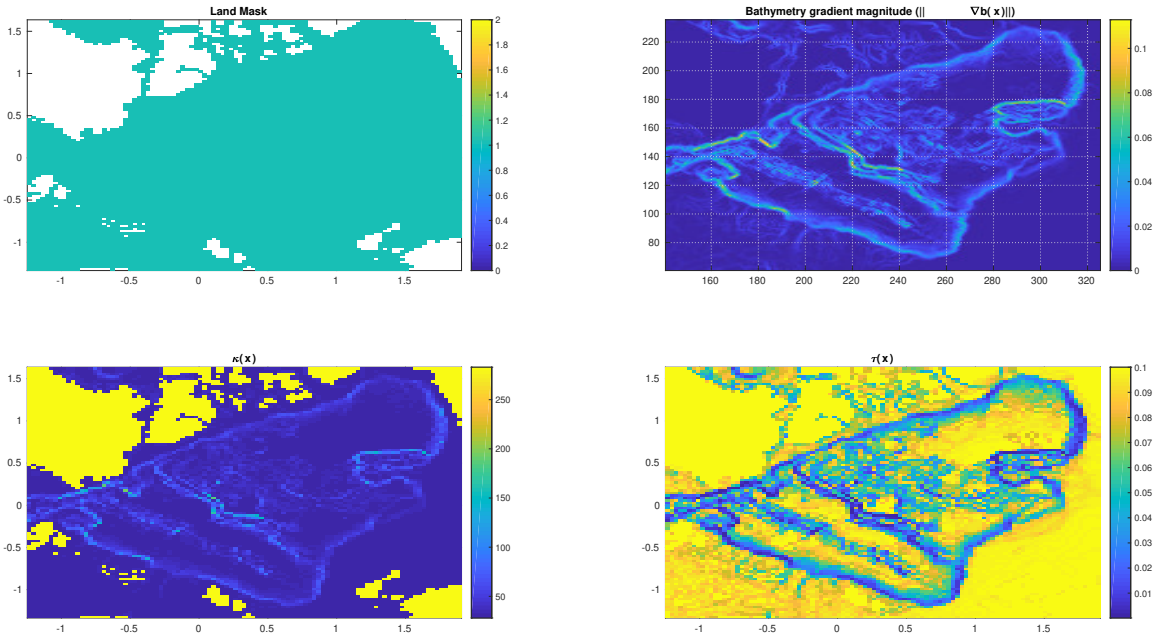

Figure 1: Clockwise: Plots of the Land Mask, bathymetry gradient norm $\|\nabla \mathbf{x}\|$, correlation length parameter $\kappa(\mathbf{x})$ and marginal variance parameter $\tau(\mathbf{x})$.

regions have virtually no $\Delta D_{kr}(\mathbf{x};\omega)$ perturbation as we desired. One can clearly see that regions with high bathymetry gradients have large but localized (uncorrelated) spikes in the $\Delta D_{kr}(\mathbf{x};\omega)$. In contrast, regions with a relatively flat bathymetry have smoother

$\Delta D_{kr}(\mathbf{x};\omega)$, with a larger degree of continuity.

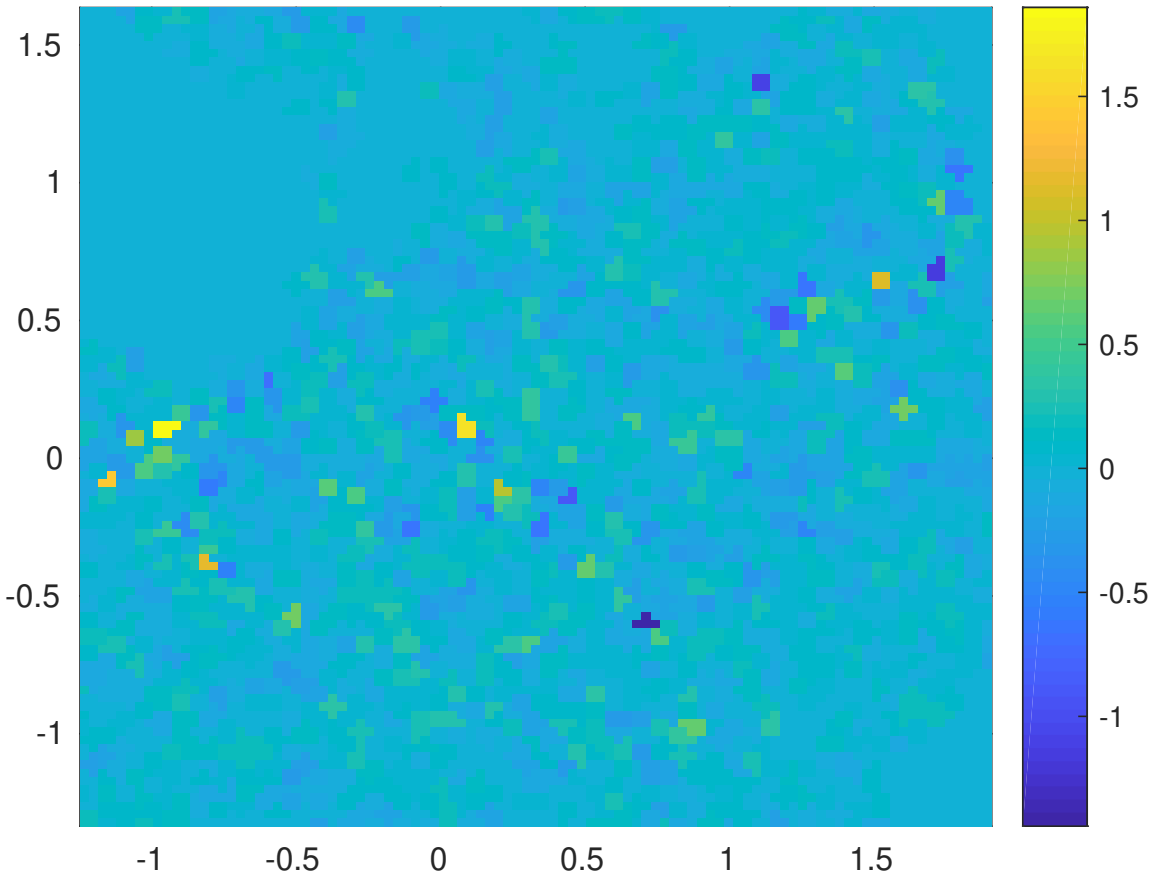

Figure 2: $\Delta D_{kr}(\mathbf{x};\omega)$ sample realization from R-INLA. Regions with high bathymetry gradients have large, uncorrelated spikes. Regions with gentler bathymetry have smoother, more correlated spatial variation.

3.3 3-D Projection of R-INLA samples

From R-INLA, we obtain 2-d samples reflecting the values $\Delta D_{kr}(\mathbf{x};\omega)$ at the interface of the land and water. We need to develop a full 3-d $\Delta D_{kr}(\mathbf{x};\omega)$, since D_{kr} is a 3-d field in

MITgcm. Since the MITgcm has no cliffs, cells with same (x, y) coordinates but varying z coordinates, each (x, y) location has a unique interfacial $z_{\text{interface}}$. We simply project the $\Delta D_{kr}(\mathbf{x}; \omega)(x, y)$ to the associated $(x, y, z_{\text{interface}})$ cell, leaving the other non-interface cells unchanged.

3.4 Verification of R-INLA sample runs

Ocean circulation in Arctic is known to display a specific temperature and salinity profile. These observations are backed by extensive ocean data and high fidelity numerical simulations. Figure 3 shows typical averaged temperature and salinity profiles seen in the Arctic, with a layer of warm water at the very top, followed by a thick layer of cooler water from the Atlantic ocean and a warming trend further with depth. This mixed layer is produced by the interaction of the Arctic ocean's salinity with the hydrodynamics. Since each sample

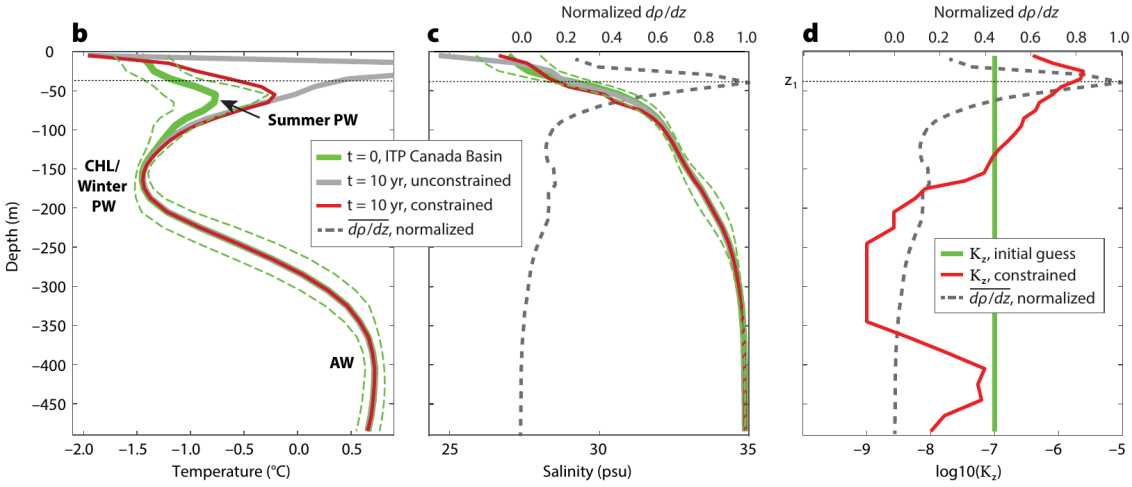

Figure 3: Baseline undersea temperature and salinity profiles for the Arctic.

run represents a perturbation of the turbulent heat transport, we expect to see changes in this profile. But to be physical, such changes have to be only quantitative, not qualitative. For example, the thickness of the cool middle layer can vary with sample, but it should never disappear. In other words, none of our $\Delta D_{kr}(\mathbf{x};\omega)$ samples should produce a bifurcation in the observed temperature profile. To avoid such situations, we can average the temperature profile for each sample run and reject those sample runs that produce profiles which differ qualitatively from the baseline, a simple form of rejection sampling.

References

[Adcroft et al., 2008] Adcroft, A., Campin, J., Dutkiewicz, S., Evangelinos, C., Ferreira, D., Forget, G., Fox-Kemper, B., Heimbach, P., Hill, C., Hill, E., et al. (2008). Mitgcm user manual.

[Forget et al., 2015] Forget, G., Ferreira, D., and Liang, X. (2015). On the observability of turbulent transport rates by argo: supporting evidence from an inversion experiment. *Ocean Science*, 11(5):839.

[Huan, 2010] Huan, X. (2010). *Accelerated Bayesian experimental design for chemical kinetic models*. PhD thesis, Massachusetts Institute of Technology.

[Huan and Marzouk, 2013] Huan, X. and Marzouk, Y. M. (2013). Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317.

[Lindgren and Rue, 2015] Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, 63(19).

[Lindgren et al., 2011] Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

[Nelder and Mead, 1965] Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.

[Riser et al., 2016] Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbéoch, M., Gilbert, D., Xu, J., Pouliquen, S., Thresher, A., et al. (2016). Fifteen years of ocean observations with the global argo array. *Nature Climate Change*, 6(2):145–153.

[Spall, 1998] Spall, J. C. (1998). Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on aerospace and electronic systems*, 34(3):817–823.