Ensemble Kalman Methods for Large-Scale Geophysical Inverse Problems and Optimal Experimental Design

Alex de Beer [1]   Elvar K. Bjarkason [2]   Michael Gravatt [3]
Ruanui Nicholson [3]   John P. O'Sullivan [3]   Michael J. O'Sullivan [3]
Oliver J. Maclaren [3]

[1]School of Mathematics and Statistics, University of Sydney
[2]Graduate School of International Resource Sciences, Akita University
[3]Department of Engineering Science and Biomedical Engineering, University of Auckland

Joint Meeting of the NZMS, AustMS and AMS, December 2024

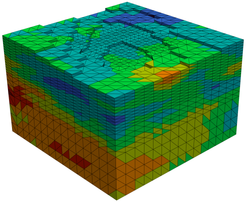
Overview

Geothermal Reservoir Modelling

Computer modelling is an important mechanism for learning about the structure of geothermal systems and predicting the effect of proposed management plans.

Challenges associated with geothermal reservoir modelling include:

- High-dimensional parameter spaces
- Long simulation times (hours to weeks)
- Frequent simulation failures



Bayesian Inference

We consider problems of the form

$$\boldsymbol{y} = \mathcal{G}(\boldsymbol{\theta}) + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}$ denotes the observations, $\mathcal{G}$ denotes the forward model, $\boldsymbol{\theta} \sim \pi_0(\boldsymbol{\theta})$ denote unknown parameters, and $\boldsymbol{\epsilon} \sim \pi_\epsilon(\boldsymbol{\epsilon})$ is an additive error term.

Assuming $\boldsymbol{\theta} \perp \boldsymbol{\epsilon}$, the density of $\boldsymbol{y} \,|\, \boldsymbol{\theta}$ inherits the density of the error. Using Bayes' theorem, the posterior density of $\boldsymbol{\theta} \,|\, \boldsymbol{y}$ is

$$\underbrace{\pi^y(\boldsymbol{\theta})}_{\text{posterior}} \propto \underbrace{\pi_\epsilon(\mathcal{G}(\boldsymbol{\theta}) - \boldsymbol{y})}_{\text{likelihood}} \underbrace{\pi_0(\boldsymbol{\theta})}_{\text{prior}}.$$

Ensemble Methods

Ensemble methods iteratively update a set of *interacting particles* to produce an approximation to the posterior.

Ensemble methods are:

- **Derivative free:** gradients are estimated using the ensemble
- **Efficient:** they can produce accurate results using $\mathcal{O}(10^3)$ simulations
- **Embarrassingly parallel:** the ensemble can be run simultaneously at each iteration

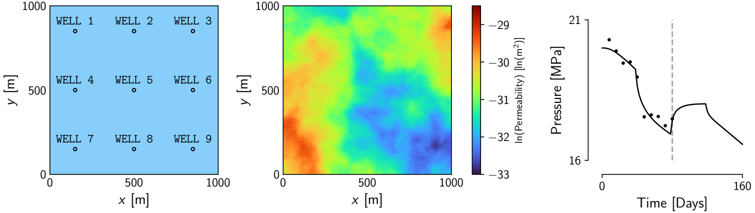
Widely-used ensemble methods include:

- **Ensemble Randomised Maximum Likelihood** (Chen and Oliver 2012, 2013)
- **Ensemble Smoother with Multiple Data Assimilation** (Emerick and Reynolds 2013)
- **Ensemble Kalman Inversion** (Iglesias *et al*. 2013; Iglesias and Yang 2021)
- **Ensemble Kalman Sampler** (Garbuno-Inigo *et al*. 2020a, 2020b)

Oil Reservoir Benchmark: Problem

Forward Problem

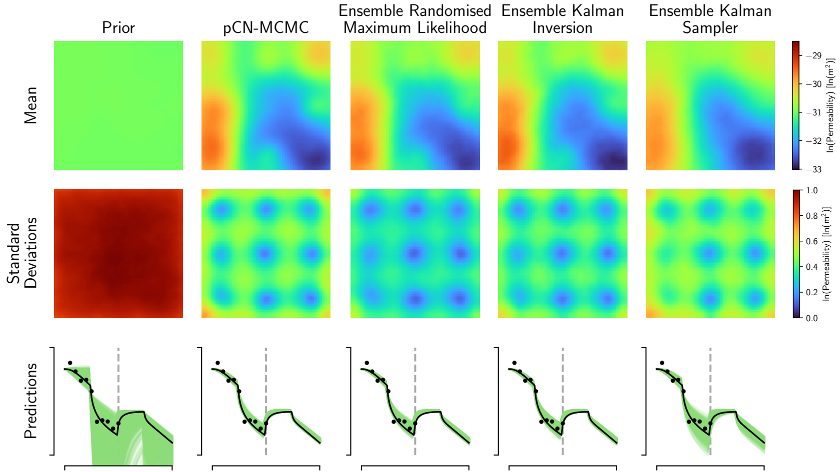
$$
\begin{aligned}
c\phi \frac{\partial p}{\partial t} - \frac{1}{\mu} \nabla \cdot (e^{u} \nabla p) &= \sum_{i=1}^{n_w} q_i \delta_{\boldsymbol{x}_i}(\boldsymbol{x}), && \boldsymbol{x} \in \Omega, t \in (0, T], \\
-e^{u} \nabla p \cdot \boldsymbol{n} &= 0, && \boldsymbol{x} \in \partial\Omega, t \in (0, T], \\
p &= p_0, && \boldsymbol{x} \in \Omega, t = 0.
\end{aligned}
$$



Inverse Problem

Estimate the reservoir log-permeability structure, $u = u(\boldsymbol{x})$, using noisy measurements of the reservoir pressure, $p = p(\boldsymbol{x}, t)$.

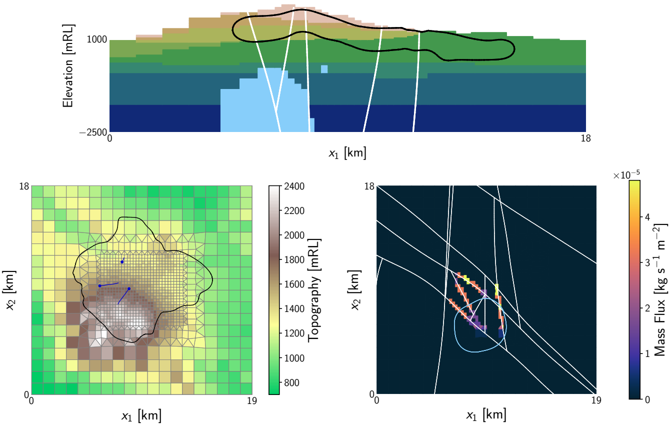
Oil Reservoir Benchmark: Results



Geothermal Case Study

Forward Problem

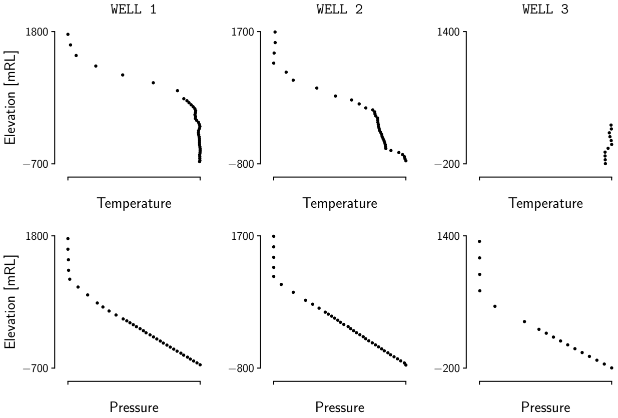
$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega} M_m \,\mathrm{d}\boldsymbol{x} = -\int_{\partial\Omega} \boldsymbol{F}_m \cdot \boldsymbol{n}\,\mathrm{d}\boldsymbol{\sigma} + \int_{\Omega} q_m \,\mathrm{d}\boldsymbol{x}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega} M_e \,\mathrm{d}\boldsymbol{x} = -\int_{\partial\Omega} \boldsymbol{F}_e \cdot \boldsymbol{n}\,\mathrm{d}\boldsymbol{\sigma} + \int_{\Omega} q_e \,\mathrm{d}\boldsymbol{x}.$$



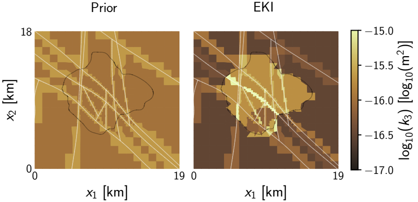
Geothermal Case Study

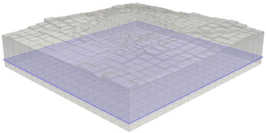
Inverse Problem

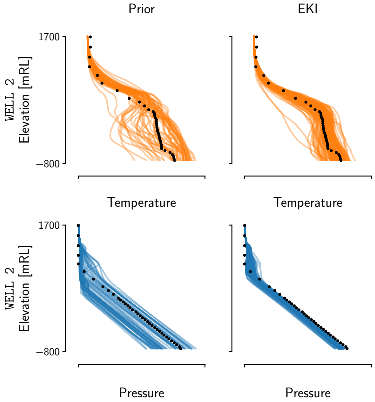
Estimate the full anisotropic reservoir permeability structure using noisy downhole temperature and pressure measurements.



Geothermal Case Study[1]







[1]For additional results: A de Beer *et al.* (2024). *Ensemble Kalman inversion for geothermal reservoir modelling*. arXiv: 2410.09017.

Bayesian Optimal Experimental Design (OED)

Drilling geothermal wells costs millions of dollars. Where should we drill to learn the most about a geothermal system?

We can formulate this within the framework of OED, in which we aim to identify a measurement plan that minimises the expectation of a function of the resulting posterior distribution.

The Sensor Placement Problem

$$\min_{\boldsymbol{w} \in \{0,1\}^{n_s}} \Psi(\boldsymbol{w}), \qquad \text{s.t. } \|\boldsymbol{w}\|_0 = k.$$

This is a challenging binary optimisation problem. We use a greedy approach to solve it.

Design Criteria

We consider several design criteria:

- A-Optimality: $\psi_{\mathsf{A}}(\boldsymbol{y}, \boldsymbol{w}) := \mathrm{tr}[\boldsymbol{C}^{y,w}]$,
- D-Optimality: $\psi_{\mathsf{D}}(\boldsymbol{y}, \boldsymbol{w}) := -\mathcal{D}_{\mathsf{KL}}(\pi^{y,w} \parallel \pi_0)$,
- Normality: $\psi_{\mathsf{N}}(\boldsymbol{y}, \boldsymbol{w}) := \mathcal{D}_{\mathsf{KL}}(\pi^{y,w} \parallel \pi_{\mathsf{N}}^{y,w})$.

We aim to *minimise* the expectations of each of these.

Design Criteria

We compute approximations of posteriors using ensemble methods, and use sample average approximation to estimate the expectation of each criterion:

$$\mathbb{E}_{\theta,y|w}[\psi_\mathrm{A}] \approx \hat{\Psi}_\mathrm{A}^{n_y}(\boldsymbol{w}) := \frac{1}{n_y}\sum_{i=1}^{n_y} \mathrm{tr}[\hat{\boldsymbol{C}}^{y_i,w}],$$

$$\mathbb{E}_{\theta,y|w}[\psi_\mathrm{D}] \approx \hat{\Psi}_\mathrm{D}^{n_y}(\boldsymbol{w}) := \frac{1}{n_y}\sum_{i=1}^{n_y} \left(\frac{1}{J}\sum_{j=1}^{J} \log\left(\frac{\pi_0(\boldsymbol{\theta}_j)}{\hat{\pi}^{y_i,w}(\boldsymbol{\theta}_j)} \right) \right),$$

$$\mathbb{E}_{\theta,y|w}[\psi_\mathrm{N}] \approx \hat{\Psi}_\mathrm{N}^{n_y}(\boldsymbol{w}) := \frac{1}{n_y}\sum_{i=1}^{n_y} \left(\frac{1}{J}\sum_{j=1}^{J} \log\left(\frac{\hat{\pi}^{y_i,w}(\boldsymbol{\theta}_j)}{\hat{\pi}_\mathrm{N}^{y_i,w}(\boldsymbol{\theta}_j)} \right) \right).$$

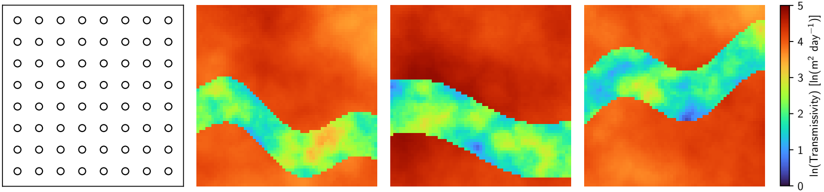
Where required, we compute a kernel density estimate of the posterior using the ensemble.

Aquifer Case Study Part I: Problem

Inverse Problem

Estimate the transmissivity of a steady-state aquifer based on measurements of the hydraulic head.

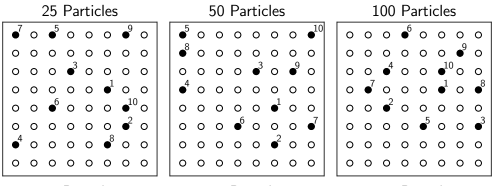
We consider an 8×8 grid of possible sensor locations.



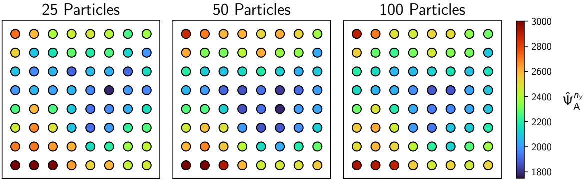
We aim to compute A-optimal designs using EKI.

Aquifer Case Study Part I: Results

Designs using EKI with varying numbers of particles:

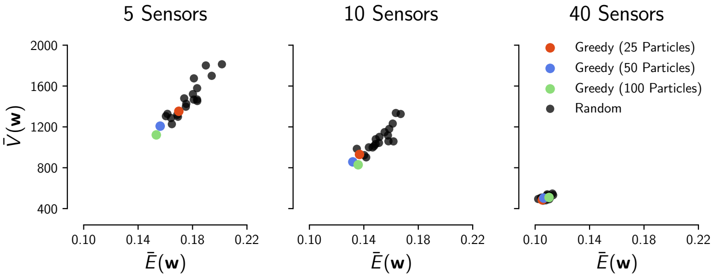


Estimates of Ψ_A at each location when selecting the first sensor:



Aquifer Case Study Part I: Validation

Results using $n_v = 20$ validation datasets and EKI with 200 particles:

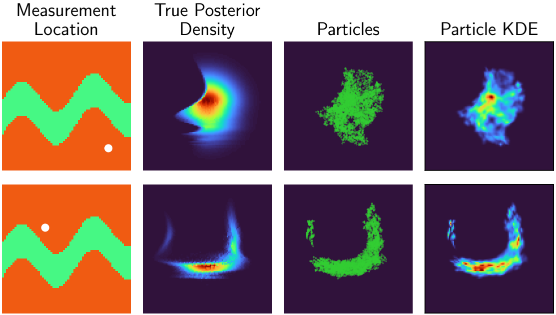


$$\underbrace{\bar{V}(\boldsymbol{w}) := \frac{1}{n_v}\sum_{i=1}^{n_v} \mathrm{tr}[\hat{\boldsymbol{C}}^{y_i^v,w}],}_{\text{A-Optimality}} \qquad \underbrace{\bar{E}(\boldsymbol{w}) := \frac{1}{n_v}\sum_{i=1}^{n_v} \frac{\|\hat{\boldsymbol{m}}^{y_i^v,w} - \boldsymbol{\theta}_i^v\|}{\|\boldsymbol{\theta}_i^v\|}}_{\text{Accuracy of Estimate}}$$

Aquifer Case Study Part II: Problem

We now consider the D-optimality and normality criteria.

Here, we use an ensemble method proposed by Reich and Weissmann (2021), which is similar to the ensemble Kalman sampler but uses "locally weighted" gradients.

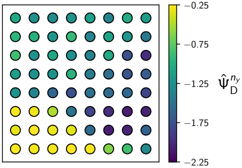


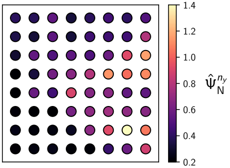
$\hat{\psi}_{\mathrm{D}} = -0.39$
$\hat{\psi}_{\mathrm{N}} = 0.23$

$\hat{\psi}_{\mathrm{D}} = -1.47$
$\hat{\psi}_{\mathrm{N}} = 1.22$

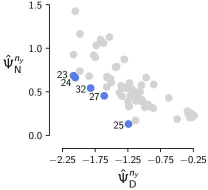
Aquifer Case Study Part II: Results

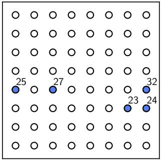
Objectives at each location when selecting a single sensor:





Some *nondominated* locations:





Conclusions and Future Work

Ensemble methods are flexible tools for solving geophysical inverse problems and OED problems.

Next steps include:

- Applying the OED methodology to real-world problems
- Pairing ensemble methods with dimension reduction techniques or surrogate models (Cleary *et al*. 2021; Dunbar *et al*. 2022)
- Using ensemble methods for goal-oriented OED

References

Chen, Y and Oliver, DS (2012). Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci.* **44**, 1–26.

Chen, Y and Oliver, DS (2013). Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* **17**, 689–703.

Cleary, E *et al.* (2021). Calibrate, emulate, sample. *J. Comput. Phys.* **424**, 109716.

de Beer, A *et al.* (2024). *Ensemble Kalman inversion for geothermal reservoir modelling*. arXiv: 2410.09017.

Dunbar, OR *et al.* (2022). Ensemble-based experimental design for targeting data acquisition to inform climate models. *J. Adv. Model. Earth Syst.* **14**, e2022MS002997.

Emerick, AA and Reynolds, AC (2013). Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **55**, 3–15.

Garbuno-Inigo, A *et al.* (2020a). Affine invariant interacting Langevin dynamics for Bayesian inference. *SIAM J. Appl. Dyn. Syst.* **19**, 1633–1658.

Garbuno-Inigo, A *et al.* (2020b). Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler. *SIAM J. Appl. Dyn. Syst.* **19**, 412–441.

Iglesias, M and Yang, Y (2021). Adaptive regularisation for ensemble Kalman inversion. *Inverse Probl.* **37**, 025008.

Iglesias, MA *et al.* (2013). Ensemble Kalman methods for inverse problems. *Inverse Probl.* **29**, 045001.

Reich, S and Weissmann, S (2021). Fokker–Planck particle systems for Bayesian inference: Computational approaches. *SIAM-ASA J. Uncertain. Quantif.* **9**, 446–482.