Lecture 1: Inverse Crimes, Model Discrepancy and Statistical Error Modeling

Erkki Somersalo

Case Western Reserve University

Cleveland, OH

Inverse Problems in the Bayesian Framework

Two classical ways of *understanding* probability:

- *Frequentist's definition*: Probability can be understood in terms of frequencies in repeated experiments
- *Bayesian definition:* Probability is a subject's expression of degree of belief.

Bayesian probability = subjective probability

Bayesian Subjective Probability

Bayesian probability:

- Expresses a subject's level of belief
- Asserts that randomness is not the object's but the *subject's* property
- May be subjective, but needs to be defendable ("Dutch book argument")

Note: Subjective is *not* the same as arbitrary.

Subjective Probability and Inverse Problems

Basic principles:

- If a value of a variable is not known, the variable is modeled as a random variable
- The information about the distribution of values is encoded in probability distributions
- From the point of view of modeling, it is immaterial whether the lack of information is contingent (imperfect measurement device, insufficient sampling of data) or fundamental (quantum physical description of an observable)

Notations:

- Variables: x, y, b, θ ...
- Random variables: X, Y, B, Θ ...
- Realizations: $X = x$, $Y = y$, $B = b$, $\Theta = \theta$, ...

Modeling error

"All models are wrong; some are useful" (George EP Cox)

Noise model:

$$y = f(x) + n, \quad n = \text{"noise"}$$

where y, x and n are realizations of Y, X and N.

- The variable Y represents observed data ("reality")
- The variable X represent the model variable, may or may not represent a physical quantity
- Reality and model are **not** identical, therefore ...
- The noise N must account for measurement noise, but **also** for the discrepancy between the model and the reality.

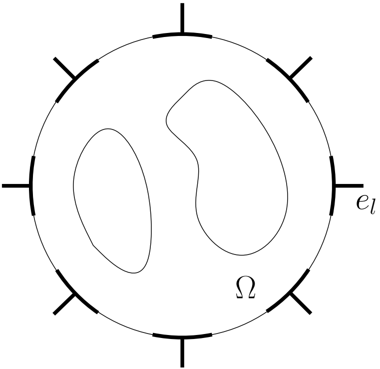
Modeling error

- *Model discrepancy* = the difference between the model and "reality".
- The fallacy of identifying model and reality is sometimes referred to as the **inverse crime**.
- To quantify the model discrepancy, we replace here "reality" with a "pretty good" model.

A "pretty good model"

- is a computational model that to best of our understanding approximates the reality
- may be computationally expensive
- is in practice replaced by a less expensive, and less accurate model.

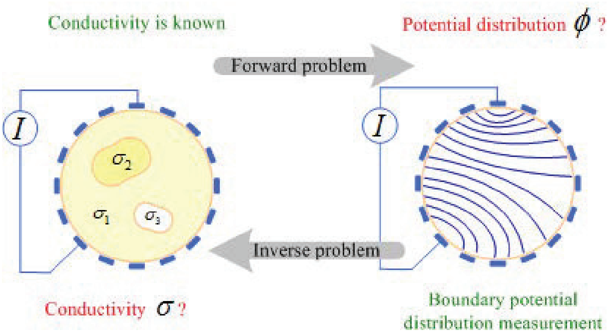
Model Case: Electrical Impedance Tomography (EIT)



[1]

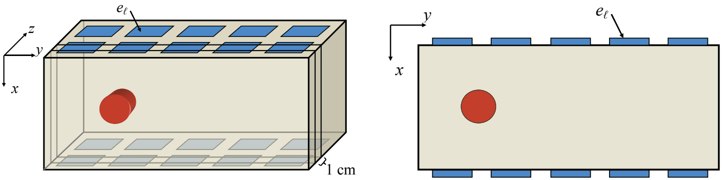
[1] Cheney M, Isaacson D and Newell J, SIAM Rew **41** 1999

EIT: Forward Problem vs. Inverse Problem



Motivation: EIT Beyond Imaging

EIS combined with mammographic tomosynthesis

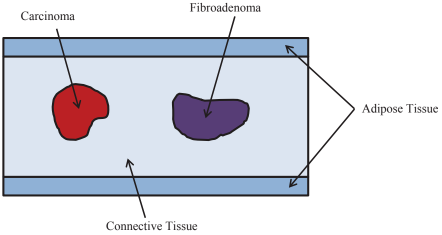


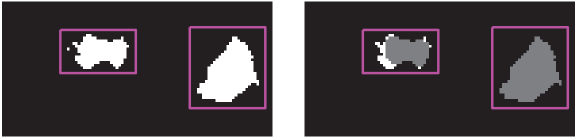
The impedance spectrum below 10 kHz of benign and malignant lesions are significantly different[2].

[2] Jossinet J (1996) Variability of impedivity in normal and pathological breast tissue. Med. Biol. Eng. Comput. 34:346-50.

Quantitative Imaging with Structural Prior

[3]





[3] McGivney D, Calvetti D and Somersalo E (2012) Quantitative imaging with electrical impedance spectroscopy. Phys. Med. Biol. **57** 7289.

PDE Model for EIT

Conservation of charge:

$$\nabla \cdot (\sigma \nabla u) = 0 \quad \text{in } \Omega.$$

Current feed through electrodes e_ℓ, $1 \leq \ell \leq L$,

$$J_\ell = \int_{e_\ell} \sigma \frac{\partial u}{\partial n} dS, \quad \sum_{\ell=1}^{L} J_\ell = 0.$$

$$\sigma \frac{\partial u}{\partial n}\bigg|_{\partial\Omega \setminus \cup e_\ell} = 0.$$

Electrode voltages with contact impedances z_ℓ,

$$\left(u + z_\ell \sigma \frac{\partial u}{\partial n} \right) \bigg|_{e_\ell} = U_\ell.$$

Variational Form

$$\mathscr{B}((v, V), (u, U)) = \int_\Omega \sigma \nabla v \cdot \nabla u dx + \sum_{\ell=1}^{L} \frac{1}{z_\ell} \int_{e_\ell} (v - V_\ell)(u - U_\ell) dS = \sum_{\ell=1}^{L} V_\ell J_\ell,$$

with ground condition,

$$\sum_{\ell=1}^{L} U_\ell = 0, \tag{1}$$

Discretize using finite elements.

FEM Approximation

$$u(x) = \sum_{j=1}^{N} u_j \psi_j(x), \quad \sigma(x) = \sigma_0 + \sum_{j=1}^{K} \theta_j \chi_j(x).$$

Define

$$x = \begin{bmatrix} \mathbf{u} \\ U \end{bmatrix} \in \mathbb{R}^{N+L}.$$

FEM approximation of the variational form:

$$\mathrm{A}_\theta x = y = \begin{bmatrix} \mathbf{0} \\ J \end{bmatrix}.$$

Observation Model

Given the current vector $J \in \mathbb{R}^L$, measure the voltage vector $U \in \mathbb{R}^L$:

$$b = U + e = \mathrm{B}x + e = \mathrm{B}\mathrm{A}_\theta^{-1} y + e,$$

where e is the observation noise, and

$$\mathrm{B} = \begin{bmatrix} 0 & \mathrm{I}_L \end{bmatrix}.$$

Concisely: For a given fixed current feed J,

$$b = f^N(\theta) + e, \quad e \sim \mathcal{N}(0, \mathrm{C}).$$

Inverse Problem in the Bayesian Setting

Prior density:

$$\theta \sim \mathscr{N}(0, \Gamma).$$

Likelihood with Gaussian observation noise:

$$b \mid \theta \sim \mathscr{N}(f^N(\theta), \mathrm{C}).$$

Posterior density:

$$\pi(\theta \mid b) \propto \exp\left(-\frac{1}{2}\|b - f^N(\theta)\|_{\mathrm{C}}^2 - \frac{1}{2}\|\theta\|_{\Gamma}^2\right),$$

where we use the notation $\|z\|_{\mathrm{M}}^2 = z^{\mathrm{T}}\mathrm{M}^{-1}z$.

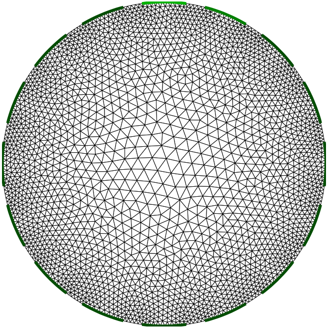
Discretization Error

- The FEM approximation converges to the solution of the PDE in $H^1(\Omega)$ as $N \to \infty$.
- A dense mesh increases the computational burden of solving the inverse problem

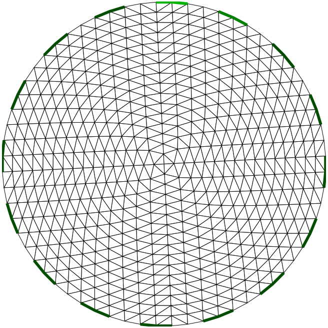
$\Rightarrow$ Trade-off between accuracy and computational complexity.

Fine vs. Coarse Mesh

Fine Mesh



Coarse Mesh



Number of elements: $N_{\mathrm{e}} = 4579$, $n_{\mathrm{e}} = 1024$
Number of nodes: $N_{\mathrm{n}} = 2675$, $n_{\mathrm{n}} = 545$

Design Guidelines

We assume to have

- A forward solver in both fine and coarse mesh
- An inverse solver in the coarse mesh only

Goal: A high precision inverse solver

- Requiring few forward solutions in the fine mesh
- Requiring few inverse solutions in the coarse mesh
- Producing an estimate of the conductivity with quantified uncertainty.

Numerical Approximation Error

Fine mesh vs. coarse mesh model:

$$\mathsf{A}_\theta^N x = y^N, \quad \mathsf{A}_\theta^n x = y^n,$$

where $\mathsf{A}^N \in \mathbb{R}^{N\times N}$, $\mathsf{A}^n \in \mathbb{R}^{n\times n}$, $n < N$.
Noiseless observation models:

$$b^N = \mathsf{B}^N \big(\mathsf{A}_\theta^N\big)^{-1} y^N = f^N(\theta), \quad b^n = \mathsf{B}^n \big(\mathsf{A}_\theta^n\big)^{-1} y^n = f^n(\theta),$$

Noisy observation in terms of the coarse mesh model:

$$b = f^N(\theta) + e = f^n(\theta) + \big\{f^N(\theta) - f^n(\theta)\big\} + e.$$

Modeling Error

Coarse grid model:

$$\boxed{b = f^n(\theta) + m + e,}$$

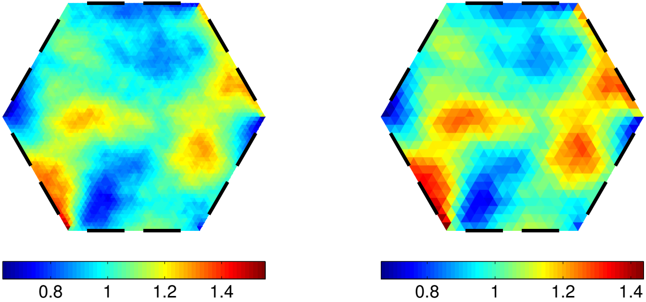
where the modeling error m is defined as

$$\boxed{m = F^{N,n}(\theta) = f^N(\theta) - f^n(\theta).}$$

Observations:

- For a known θ, the modeling error can be evaluated
- As θ is unknown, in the Bayesian framework, m must be modeled as a random variable M
- The statistics of M depends on the random variable Θ.

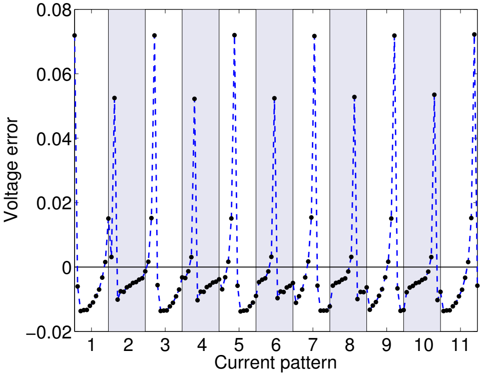
Modeling Error is Highly Structured

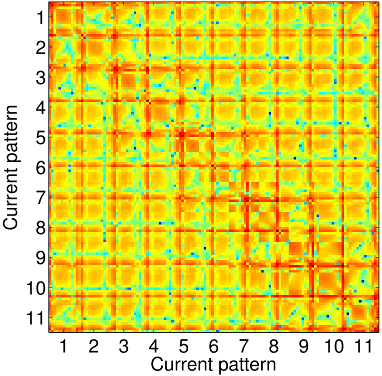


$N_{\mathrm{e}} = 4\,704$ $N_{\mathrm{e}} = 1\,176$

Mean and Covariance of the Modeling Error

Why not inflate the variance of e to mask the modeling error?





- Modeling error is not zero mean, and exhibits high level cross-talk between channels.
- Dominant noise with high quality data.

Solution in the Bayesian Framework[4]

$$M = F^{N,n}(\Theta) = f^N(\Theta) - f^n(\Theta).$$

If $\Theta \sim \pi_\Theta$ then $M \sim \pi_M$, where

$$\pi_M = F^{N,n}_* \pi_\Theta, \quad \text{(push-forward)},$$

$$\mathrm{P}\{M \in A\} = \int_{(F^{N,n})^{-1}(A)} \pi_\Theta(\theta) d\theta = \int_A F^{N,n}_* \pi_\Theta(m) dm,$$

or, assuming that $F^{N,n}$ is a diffeomorphism,

$$F^{N,n}_* \pi_\Theta(m) = \left| \frac{\partial (F^{N,n})^{-1}}{\partial m}(m) \right| \pi_\Theta((F^{N,n})^{-1}(m)).$$

[4] Kaipio JP and Somersalo E (2007) Statistical inverse problems: discretization, model reduction and inverse crimes. J. Comp. Appl. Math. **198** (2007) 493–504.

Enhanced error model

- Approximate π_M by a Gaussian density,

$$\pi_M \sim \mathcal{N}(\overline{m}, \Sigma),$$

- Estimate the mean $\overline{m}$ and covariance Σ using the prior density of Θ,
- Neglect the interdependency of Θ and M in the coarse model (not necessary[5]).

[5] Calvetti D and Somersalo E (2005) Statistical compensation of boundary clutter in image deblurring. Inverse Problems **21**: 1697–1714.

Off-line Prior Sampling Approximation

1. Generate a sample of realizations of $\theta \sim \pi_{\text{prior}} = \mathscr{N}(0, \Gamma)$,

$$\{\theta_1, \theta_2, \ldots, \theta_K\},$$

2. Compute a sample of model error vectors,

$$m_\ell = F^{N,n}(\theta_\ell), \quad 1 \leq \ell \leq K,$$

   and sample mean and covariance,

$$\overline{m} = \frac{1}{K} \sum_{\ell=1}^{K} m_\ell, \quad \Sigma = \frac{1}{K} \sum_{\ell=1}^{K} (m_\ell - \overline{m})(m_\ell - \overline{m})^{\mathsf{T}}.$$

3. Compute the posterior estimates of θ from the *enhanced error model*,

$$b = f^n(\theta) + E, \quad E \sim \mathscr{N}(\overline{m}, \mathsf{C} + \Sigma),$$

   where it is assumed that E is independent of θ.

Prior and MAP Estimate

Prior model

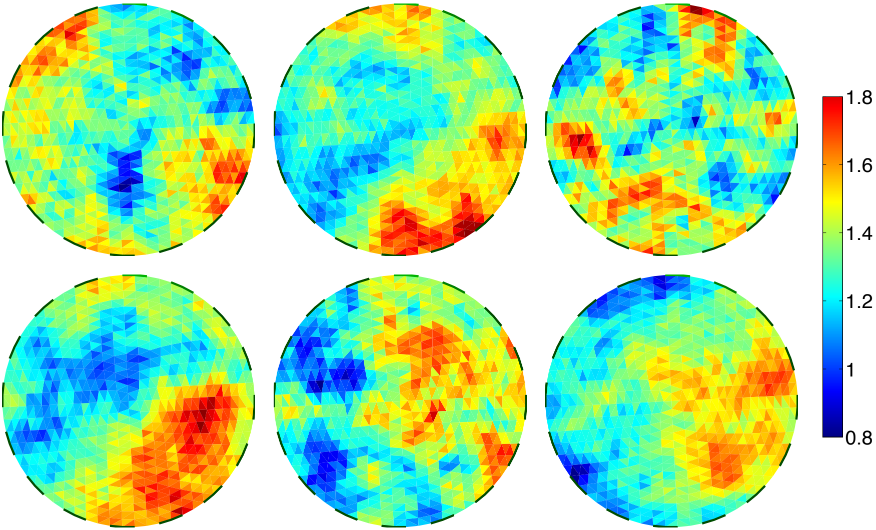
$$\theta \sim \mathscr{N}(0, \Gamma), \quad \Gamma_{j\ell} = \gamma \exp\left(-\frac{|x_j - x_\ell|}{\lambda}\right),$$

where $\lambda > 0$ is the correlation length, γ is the prior marginal pixel variance.
MAP estimate:

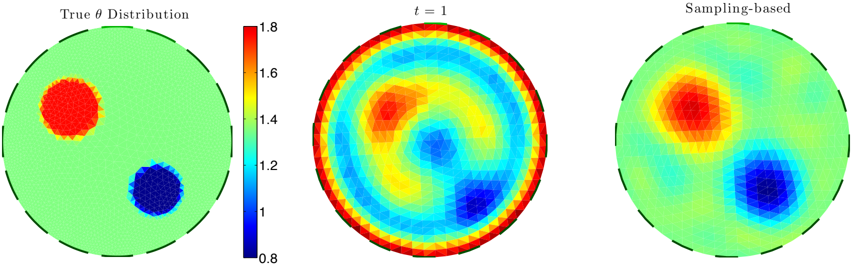
$$\widehat{\theta} = \operatorname{argmin}\left\{\|b - \overline{m} - f^n(\theta)\|^2_{C+\Sigma} + \|\theta\|^2_\Gamma\right\},$$

solved by using a Gauss-Newton optimization.

Draws from Prior Density



MAP estimates



Enhanced error model computed with $K = 2\,500$ draws from the prior density.

Variance Reduction

Given $\pi(\theta) = \pi_{\text{prior}}$, we used

$$\pi_M = F_*^{N,n}\pi_{\text{prior}} \approx \mathcal{N}(\overline{m}, \Sigma),$$

to obtain the update

$$\pi_{\text{prior}} \to \pi_{\text{post}}(\theta \mid b; \pi_M, f^n).$$

Question: *Using updated information about θ, can we effectively update the error model leading to an updated posterior for θ?*

Bayesian Approach: Updating Beliefs

Iterative algorithm:

- Initial belief = prior: Set $\pi_\Theta^0(\theta) = \pi_{\text{prior}}(\theta)$, set $j = 0$.
- Estimate $\pi_M^j = F_*^{N,n}\pi_\Theta^j$.
- Estimate the posterior density,

$$\pi_\Theta^{j+1}(\theta) = \pi_{\text{post}}(\theta \mid b; \pi_M^j, f^n).$$

- Update $j \leftarrow j + 1$ and iterate from 2.

Effective implementation: Ensemble Kalman Filtering (EnKF)[6]

[6] Evensen G 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J Geophys Res 99 10143 – 10162

Bayesian Filtering and Dynamic Inverse Problems

Sequential observations:

$$B_t = g\big(\Theta_t\big) + E_t, \quad t = 1, 2, \ldots$$

State evolution model:

$$\Theta_{t+1} = G(\Theta_t) + V_{t+1}, \quad t = 0, 1, \ldots$$

Update

$$\pi(\theta_t \mid D_t) \rightarrow \pi(\theta_{t+1} \mid D_t) \rightarrow \pi(\theta_{t+1} \mid D_{t+1}),$$

where

$$D_t = \big\{b_1, b_2, \ldots, b_t\big\}.$$

Ensemble Kalman Filtering (EnKF)

1. Generate a prior sample,

$$\mathscr{S}_0 = \{\theta_0^1, \theta_0^2, \ldots, \theta_0^k\}, \quad \theta_\ell \sim \pi_{\rm prior}$$

   and set $t = 0$.

2. Propagate the sample,

$$\widehat{\theta}_{t+1}^j = G(\theta_t^j) + v_{t+1}^j, \quad j = 0, 1, \ldots, k$$

   Compute the empirical mean and covariance, $\overline{\theta}_{t+1}$, $\overline{\Gamma}_{t+1}$.

3. Parametric bootstrap of the data b_{t+1},

$$b_{t+1}^j = b_{t+1} + w_{t+1}^j, \quad w_{t+1}^j \sim \mathscr{N}(0, \mathsf{C}), \quad 1 \leq j \leq k.$$

4. Update the sample $\mathscr{S}_t \to \mathscr{S}_{t+1}$,

$$\theta_{t+1}^j = \mathrm{argmin}\{\|b_{t+1}^j - f^n(\theta)\|_{\mathsf{C}}^2 + \|\theta - \widehat{\theta}_{t+1}^j\|_{\overline{\Gamma}_{t+1}}^2\}.$$

Application to Modeling Error Update

- Observation model: Update the likelihood,

$$B_t = f^n(\Theta) + M_t + E, \quad M_t \sim F_*^{N,n}\pi_\Theta^t,$$

while the realization is

$$B_t = b.$$

- Propagation model

$$\Theta_{t+1} = \Theta_t.$$

Iterative Updating of Modeling Error

1. Initialize: Draw a sample of size k from the prior density,

$$\mathscr{S}_k^0 = \{\theta_1^0, \ldots, \theta_k^0\}, \quad \theta_j^0 \sim \pi_{\mathrm{prior}}.$$

Set $\overline{m} = 0 \in \mathbb{R}^m$, $\Sigma = 0 \in \mathbb{R}^{m \times m}$. Set $t = 0$.

2. Generate bootstrap data,

$$b_\ell = b - \overline{m} + w_\ell, \quad w_\ell \sim \mathscr{N}(0, \mathsf{C} + \Sigma), \quad 1 \leq \ell \leq k.$$

3. Update the sample,

$$\mathscr{S}_k^{t+1} = \{\theta_1^{t+1}, \ldots, \theta_k^{t+1}\},$$

$$\theta_j^{t+1} = \mathrm{argmin}\{\|b_j - f^n(\theta)\|^2_{\mathsf{C}+\Sigma} + \|\theta - \widehat{\theta}_j^t\|^2_\Gamma\}, \quad \widehat{\theta}_j^t \sim \pi_{\mathrm{prior}}, \quad 1 \leq j \leq k.$$

Iterative Updating of Modeling Error

4. Compute the new model error sample,

$$m_j^{t+1} = F^{N,n}(\theta_j^{t+1}), \quad 1 \leq j \leq k,$$

and the corresponding mean and covariance,

$$\overline{m}^{t+1} = \frac{1}{k}\sum_{j=1}^{k} m_j^{t+1}, \quad \Sigma^{t+1} = \frac{1}{k}\sum_{j=1}^{k}(m_j^{t+1} - \overline{m}^{t+1})(m_j^{t+1} - \overline{m}^{t+1})^{\mathsf{T}}.$$

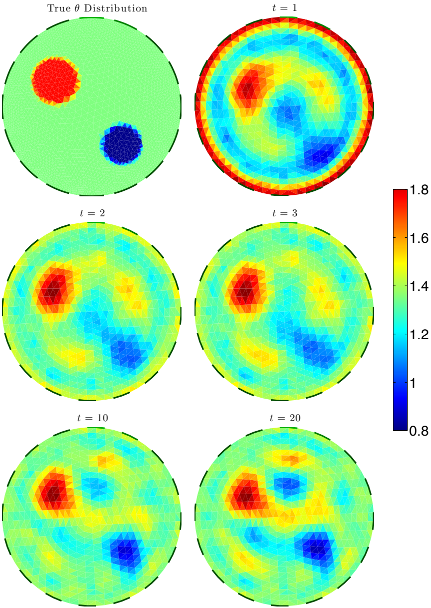
5. Update the cumulative model error mean and covariance,

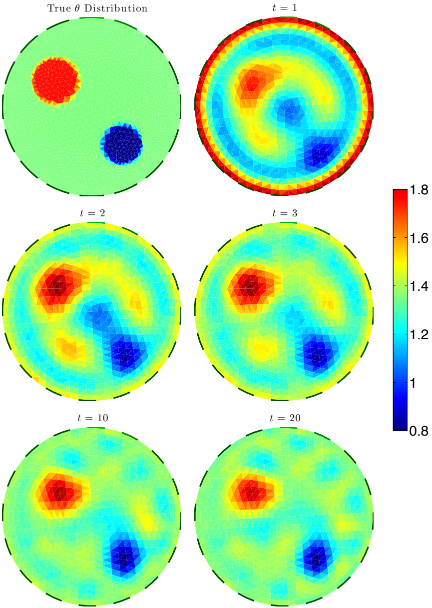
$$\overline{m}^{+} = \frac{T}{T+1}\overline{m} + \frac{1}{t+1}\overline{m}^{t+1},$$

$$\Sigma^{+} = \frac{t}{t+1}\Sigma + \frac{1}{t+1}\Sigma^{t+1} + \frac{t}{(t+1)^2}(\overline{m} - \overline{m}^{t+1})(\overline{m} - \overline{m}^{t+1})^{\mathsf{T}}.$$

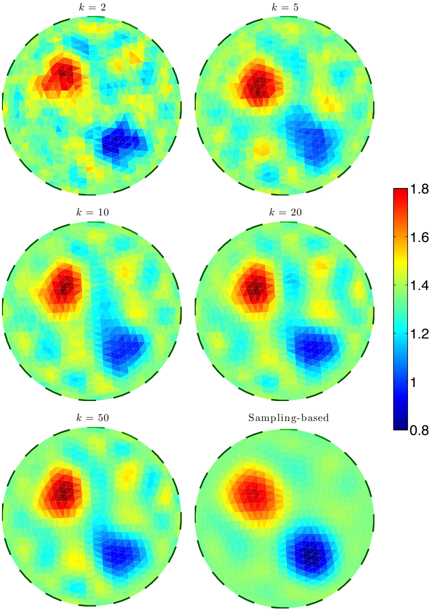
6. If the model error mean and covariance satisfy the convergence criterion, stop, else, increase t by one, set $\overline{m} = \overline{m}^{+}$ and $\Sigma = \Sigma^{+}$, and continue from Step (ii).

Convergence, $k = 5$ vs. $k = 20$





Effect of Sample Size



Additional References

- Calvetti D, Ernst O and Somersalo E 2014 Dynamic updating of numerical model discrepancy using sequential sampling. *Inverse Problems* **30**:14019.
- Calvetti D, McGivney D and Somersalo E 2012 Left and right preconditioning for electrical impedance tomography with structural information. Inverse Problems **28** 055015.
- Iglesias MA, Law KJH and Stuart AM 2013 Ensemble Kalman methods for inverse problems *Inverse Problems* **29** 045001.
- Arridge SR, Kaipio JP, Kolehmainen V, Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems* **22** 175-195.
- Kaipio JP and Somersalo E 2004 *Statistical and Computational Inverse Problems* (New York: Springer).
- Oliver DS, Reynolds AC and Liu N 2008 *Inverse Theory for Petroleum Reservoir Characterization and History Matching* Cambridge University Press, Cambridge.
- Vauhkonen M, Lionheart WR, Heikkinen LM, Vauhkonen PJ and Kaipio JP 2001 A MATLAB package for the EIDORS project to reconstruct two-dimensional EIT images. *Physiol Meas* **22** 107.