Hyper-parameter optimization for improving the performance of localization in an iterative ensemble smoother

Xiaodong Luo, NORCE, Norway;
William C. Cruz, UiS, Norway;
Xinlei Zhang, Chinese Academy of Sciences, China;
Heng Xiao, University of Stuttgart, Germany

Outline

- Background
- Parameterized correlation-based localization
- Numerical examples
- Summary

Iterative ensemble smoother (IES)

- Basic iterative ensemble smoother (IES) without localization:

$$m_j^{i+1} = m_j^i + K\left(d^o - g\left(m_j^i\right)\right), j = 1, 2, \ldots, N_e$$

  - $d^o \in R^{p\times 1}$: p-dimensional observed field data
  - $m \in R^{n\times 1}$: n-dimensional reservoir model
  - i: iteration index
  - j: ensemble member index
  - g: reservoir simulator (ignoring model errors)
  - K: Kalman-gain matrix
  - N_e: ensemble size

Localization in the IES

- IES often runs with a small ensemble
  - Spurious correlations (sampling errors)
  - Rank deficiency
  - Degraded data assimilation (history matching) performance

- Localization often adopted for improved performance

- IES with (Kalman-gain) localization

$$m_j^{i+1} = m_j^i + (T \circ K)\left(d^o - g\left(m_j^i\right)\right), j = 1, 2, \ldots, N_e$$

  - $T \in [0,1]$: localization (or tapering) matrix
  - $T \circ K$: elementwise product between T and K

Localization in the IES

$$m_j^{i+1} = m_j^i + (T \circ K)\left(d^o - g\left(m_j^i\right)\right), j = 1, 2, \ldots, N_e$$

- Essential question: how to construct the tapering matrix T?

- Here the focus on correlation-based localization

Correlation-based localization

- Additional notations:

$$m_j^{i+1} = m_j^i + (T \circ K)\left(d^o - g\left(m_j^i\right)\right)$$

  - $\left[m_j^i\right]_s$: the s-th model variable of m_j^{i+1}
  - $\left[g\left(m_j^i\right)\right]_k$: the k-th element of $g\left(m_j^i\right)$
  - $T \equiv [t_{sk}]$
  - t_{sk}: element of T on the s-th row and the k-th column

- In correlation-based localization:

$$t_{sk} = h_\theta(\rho_{sk})$$

  - ρ_{sk}: sample correlation between the ensembles $\left[m_j^i\right]_s$ and $\left[g\left(m_j^i\right)\right]_k$, for $j = 1, 2, \ldots, N_e$
  - h_θ: tapering function, parameterized by a set of hyper-parameters θ

Correlation-based localization

- Example 1*:

$$t_{sk} = h_\theta(\rho_{sk})$$

$$h_\theta(\rho_{sk}) = h(|\rho_{sk}| > \theta) = \begin{cases} 1, if\ |\rho_{sk}| > \theta \\ 0, otherwise \end{cases} \quad \text{(Heaviside function)}$$

  - θ: threshold value (hard-thresholding)
  - θ empirically chosen, as in, e.g., Luo et. al*
  - Referred to as empirical tuning strategy hereafter

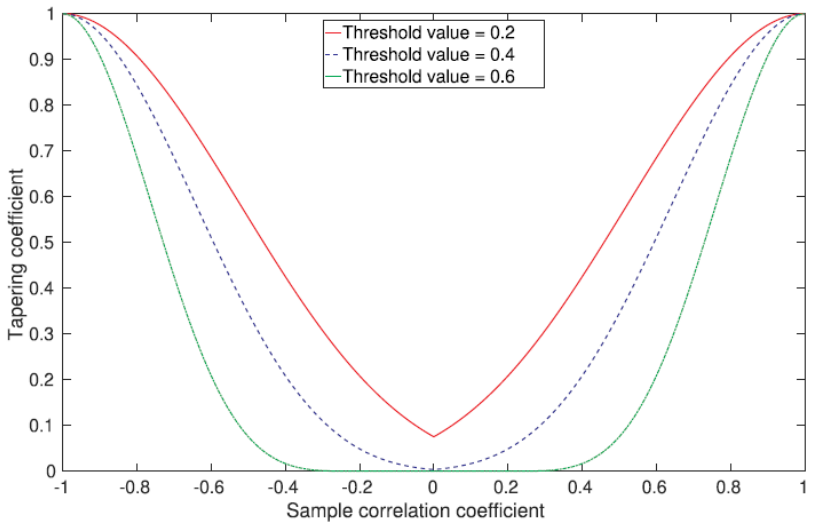
*Luo, X., Bhakta, T. and Naevdal, G., 2018. Correlation-based adaptive localization with applications to ensemble-based 4D-seismic history matching. *SPE Journal*, *23*(02), pp.396-427.

Correlation-based localization

- Example 2 (AutoAdaLoc)*:

$$t_{sk} = h_\theta(\rho_{sk})$$

$$h_\theta(\rho_{sk}) = h_{GC}\left(\frac{1 - |\rho_{sk}|}{1 - \theta}\right)$$



- h_{GC}: Gaspari-Cohn (GC) function
- θ: threshold value (soft-thresholding)
- θ chosen based on statistical analyses, before data assimilation starts;
- Referred to as prior tuning strategy hereafter

*Luo, X. and Bhakta, T., 2020. Automatic and adaptive localization for ensemble-based history matching. *Journal of Petroleum Science and Engineering*, *184*, p.106559.

Outline

- Background
- Parameterized correlation-based localization
- Numerical examples
- Summary

Hyper-parameters in correlation-based localization

$$t_{sk} = h_{GC}\left(\frac{1 - |\rho_{sk}|}{1-\theta}\right) \Rightarrow t_{sk} = h_{GC}\left(\frac{1 - |\rho_{sk}|}{\ell_k}\right)$$

- $\theta \equiv \{\ell_k, k = 1,2,\dots,p\}$: localization length scales varying for each observation data point (θ in the same size as field data d^o)
- Hyper-parameters can also vary over each model variable and each data point (i.e., $\theta \equiv \{\ell_{sk}, s = 1,2,\dots,n;\ k = 1,2,\dots,p\}$), but maybe too many to be practical (a lighter choice: $\theta \equiv \{\ell_s, s = 1,2,\dots,n\}$, i.e., varying over each model variable)
- Localization called parameterized AutoAdaLoc (P-AutoAdaLoc) scheme hereafter
- θ iteratively updated during data assimilation
- Referred to as posterior tuning strategy hereafter

Continuous hyper-parameter optimization (CHOP)*

- CHOP designed to estimate (an ensemble of) algorithmic hyper-parameters $\{\theta_j^i\}_{j=1}^{N_e}$ in a generic model update formula $m_j^{i+1} = f(m_j^i, \theta_j^i | d^o)$

- Containing as a special case an IES with parameterized localization

$$m_j^{i+1} = f\left(m_j^i, \theta_j^i \middle| d^o\right) \equiv m_j^i + (T(\theta_j^i) \circ K)\left(d^o - g\left(m_j^i\right)\right)$$

*Luo, X. and Xia, C.A, 2022. Continuous Hyper-parameter OPtimization (CHOP) in an ensemble Kalman filter. Frontiers in Applied Mathematics and Statistics, 2022, 8, p. 1021551.

Continuous hyper-parameter optimization (CHOP)

- Two-step update procedure: Given ensembles of $\{m_j^i\}_{j=1}^{N_e}$ and $\{\theta_j^i\}_{j=1}^{N_e}$ at the i-th iteration step

  - Step 1: model update (through an IES): ← existing in the original IES algorithm

$$m_j^{i+1} = m_j^i + \left(T\left(\theta_j^i\right) \circ K\right)\left(d^o - g\left(m_j^i\right)\right), j = 1, 2, \dots, N_e$$

K computed with respect to $\{m_j^i\}_{j=1}^{N_e}$ and $\{g(m_j^i)\}_{j=1}^{N_e}$

  - Forward simulations to obtain $\{g(m_j^{i+1})\}_{j=1}^{N_e}$: ← existing in the original IES algorithm

  - Step 2: hyper-parameter update (also through an IES): ← new component

$$\theta_j^{i+1} = \theta_j^i + \left(T\left(\theta_j^i\right) \circ \widetilde{K}\right)\left(d^o - g\left(m_j^{i+1}\right)\right), j = 1, 2, \dots, N_e$$

$\widetilde{K}$ computed with respect to $\{\theta_j^i\}_{j=1}^{N_e}$ and $\{g(m_j^{i+1})\}_{j=1}^{N_e}$

Continuous hyper-parameter optimization (CHOP)*

$$\theta_j^{i+1} = \theta_j^i + \left(T\left(\theta_j^i\right) \circ \widetilde{K}\right)\left(d^o - g\left(m_j^{i+1}\right)\right), j = 1, 2, \ldots, N_e$$

- Optimization criterion: $\left\{\theta_j^{i+1}\right\}_{j=1}^{N_e}$ chosen to minimize/reduce average data mismatch between $\left\{g\left(m_j^{i+1}\right)\right\}_{j=1}^{N_e}$ and d^o

- CHOP converted to a normal parameter estimation problem (and solved by IES)

For those who are interested: $\left\{\theta_j^{i+1}\right\}_{j=1}^{N_e}$ approximate solution to the following minimum-average cost (MAC) problem

$$\min_{\left\{\theta_j^{i+1}\right\}_{j=1}^{N_e}} \frac{1}{N_e}\sum_j L\left(\theta_j^{i+1}\right)$$

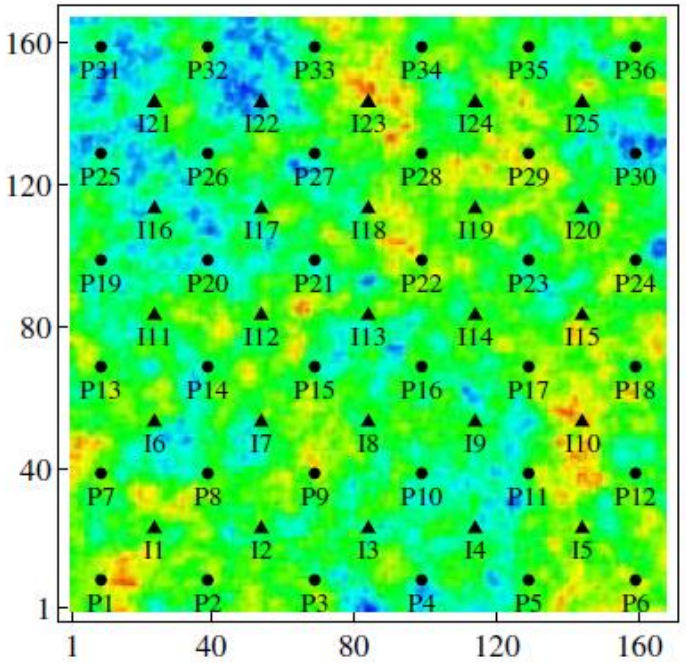
$$L\left(\theta_j^{i+1}\right) = \frac{1}{2}\left(d^o - g\left(f\left(m_j^i, \theta_j^{i+1}|d^o\right)\right)\right)^T C_d^{-1}\left(d^o - g\left(f\left(m_j^i, \theta_j^{i+1}|d^o\right)\right)\right) + \frac{\gamma}{2}\left(\theta_j^{i+1} - \theta_j^i\right)^T C_\theta^{-1}\left(\theta_j^{i+1} - \theta_j^i\right)$$

*Luo, X., Cruz, W., Zhang, X. L., & Xiao, H. Hyper-Parameter Optimization for Improving the Performance of Localization in an Iterative Ensemble Smoother. Preprint, *available at SSRN 4388296*.

Outline

- Background
- Parameterized correlation-based localization
- Numerical examples
- Summary

2D case: multiple 5spots (M-5Spots)*



Experimental settings	
Model information	167 x 167 36 producers + 25 injectors Uncertain parameters: PERMX
Production data used for history matching	WOPR, WWPR, WBHP, WWIR total number = 1098
HM algorithm	IES with AutoAdaLoc vs. IES with P-AutoAdaLoc Ensemble size: 100
Initial ensemble of localization length scales	i.i.d samples from the uniform distribution on the interval [0.23, 0.43] (manual choice)

*Chen, Y. and Oliver, D.S., 2010. Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Computational Geosciences*, *14*(4), pp.579-601.

2D case: multiple 5spots (M-5Spots)

Table 2: Data mismatch (DM), root mean squared error (RMSE) and ensemble spread in the M-5Spots case.

	Initial ensemble	Final ensemble (AutoAdaLoc)	Final ensemble (P-AutoAdaLoc)
DM (mean $\pm$ STD)	$(1.7091 \pm 0.3517) \times 10^7$	$(1.7896 \pm 0.5328) \times 10^5$	$\mathbf{(7.2535 \pm 1.1881) \times 10^4}$
RMSE (mean $\pm$ STD)	1.6647 ± 0.0707	1.3811 ± 0.0357	$\mathbf{1.2550 \pm 0.0340}$
Spread	1.1864	0.9293	**0.5819**

2D case: multiple 5spots (M-5Spots)

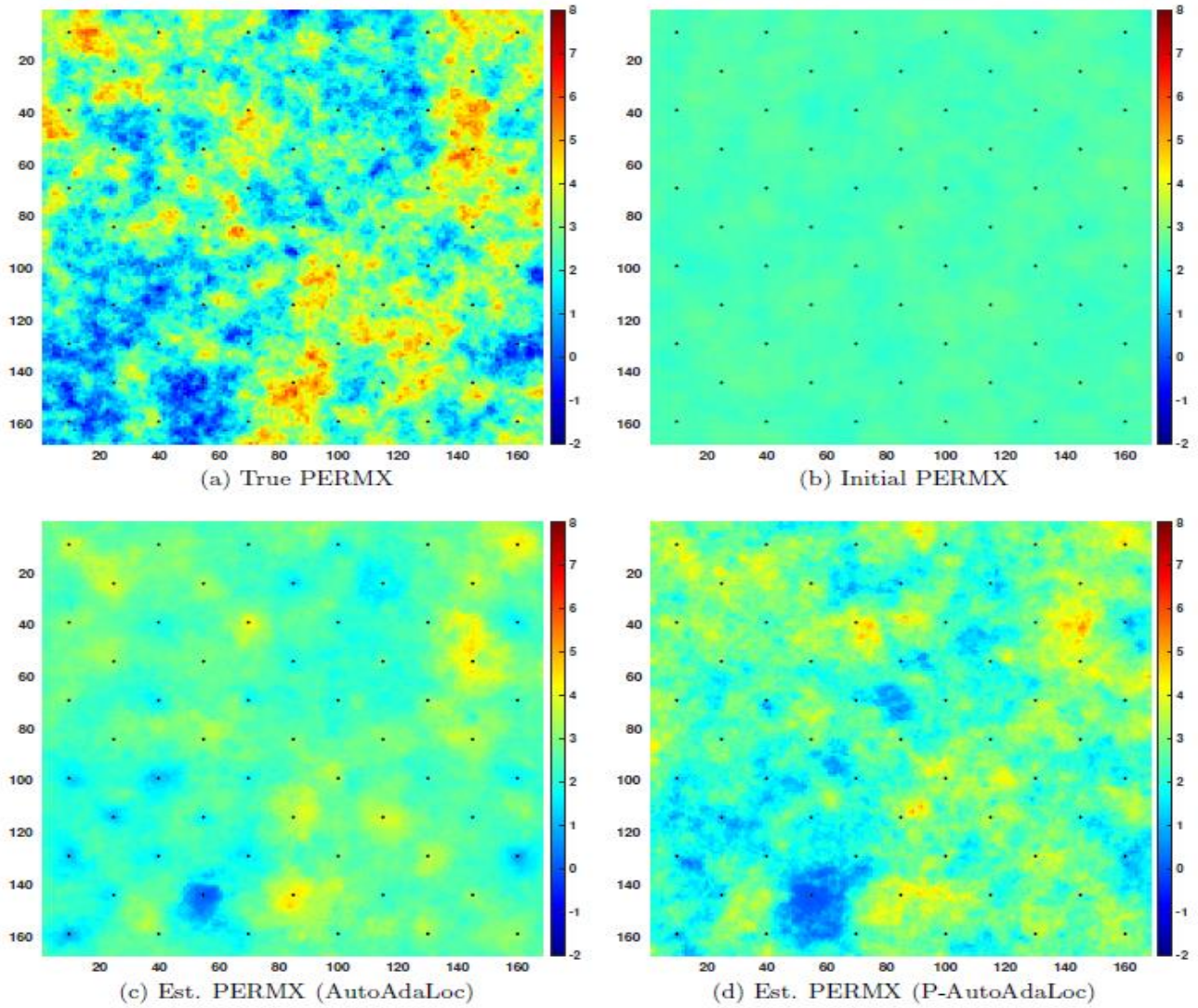
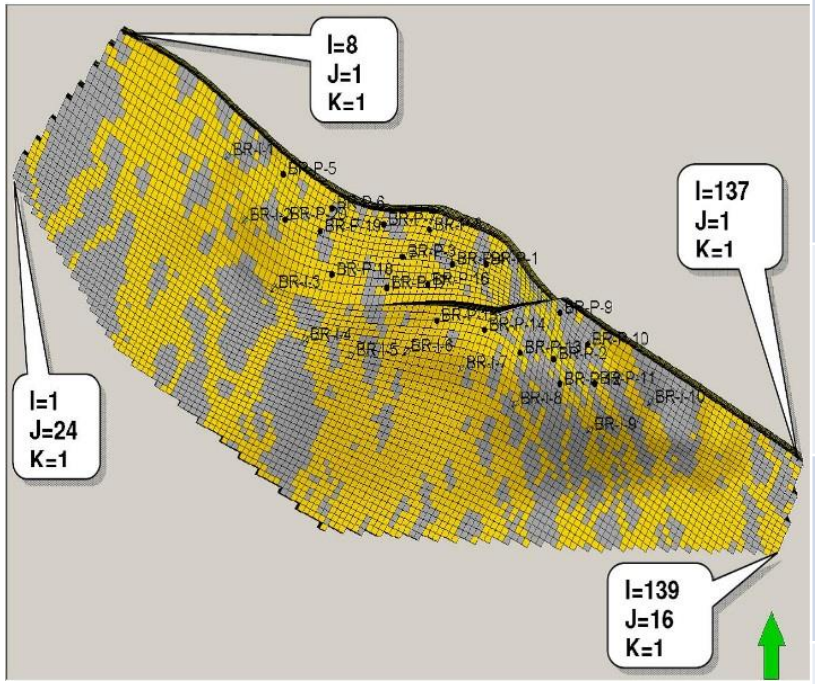


Figure 5: PERMX maps (in the natural logarithmic scale) with respect to (a) the reference model; (b) the mean of the initial ensemble; (c) – (d) the means of the final estimated (est.) ensembles obtained by the IES algorithm with the AutoAdaLoc and the P-AutoAdaLoc schemes, respectively, in the M-5Spots case. In all the maps, the small dots indicate the locations of wells.

3D case: Brugge benchmark



Grid geometry of the Brugge field

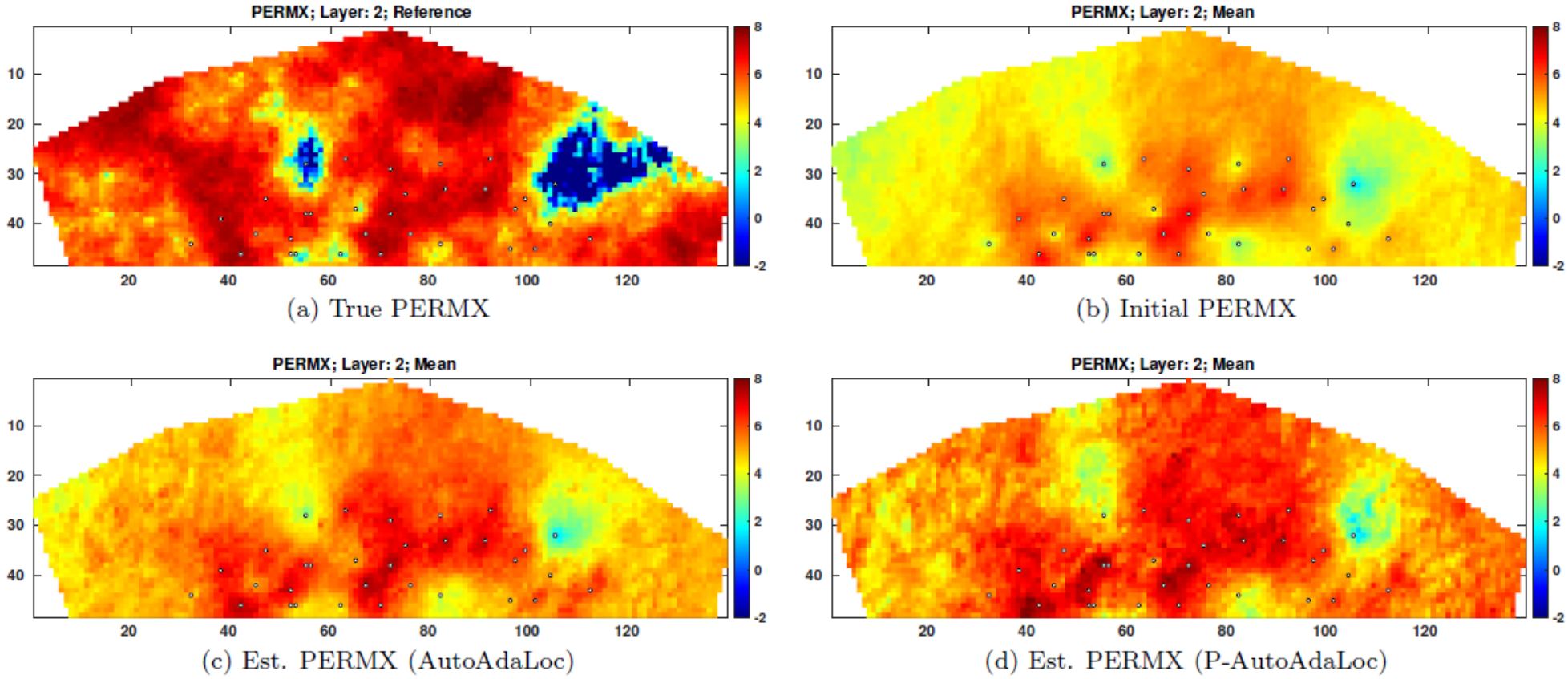
Experimental settings	
Model information	139 x 48 x 9 20 producers + 10 injectors Uncertain parameters: PERMX, PERMY, PERMZ, PORO
Production data used for history matching	WOPR, WWCT, WBHP total number = 1400
HM algorithm	IES with AutoAdaLoc vs. IES with P-AutoAdaLoc Ensemble size: 103
Initial ensemble of localization length scales	i.i.d samples from the uniform distribution on the interval [0.23, 0.43] (manual choice)

3D case: Brugge benchmark

Table 3: Data mismatch (DM), root mean squared error (RMSE) and ensemble spread in the Brugge benchmark case.

	Initial ensemble	Final ensemble (AutoAdaLoc)	Final ensemble (P-AutoAdaLoc)
DM (mean $\pm$ STD)	$(0.3623 \pm 1.4900) \times 10^{10}$	$(0.9481 \pm 1.9730) \times 10^{7}$	$\mathbf{(3.9842 \pm 7.0173) \times 10^{5}}$
Total RMSE (mean $\pm$ STD)	1.5450 ± 0.3362	1.2610 ± 0.1663	$\mathbf{1.1645 \pm 0.1046}$
RMSE of PERMX (mean $\pm$ STD)	1.6585 ± 0.3827	1.3498 ± 0.1982	$\mathbf{1.2399 \pm 0.1205}$
RMSE of PERMY (mean $\pm$ STD)	1.6612 ± 0.3794	1.3546 ± 0.1959	$\mathbf{1.2444 \pm 0.1197}$
RMSE of PERMZ (mean $\pm$ STD)	2.0077 ± 0.4096	1.6426 ± 0.1937	$\mathbf{1.5282 \pm 0.1307}$
RMSE of PORO (mean $\pm$ STD)	0.0302 ± 0.0033	0.0298 ± 0.0031	$\mathbf{0.0259 \pm 0.0018}$
Spread	0.8661	0.6308	**0.5435**

3D case: Brugge benchmark

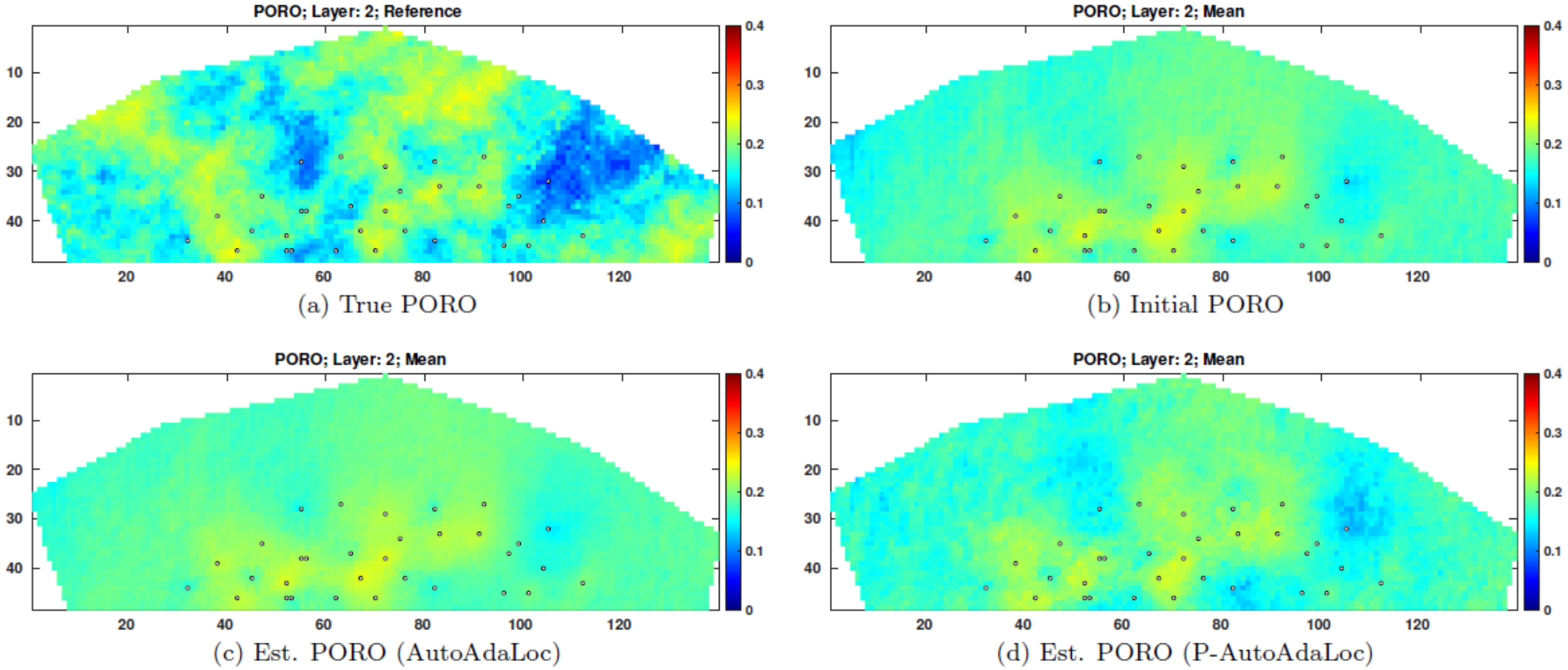


(a) True PERMX (b) Initial PERMX

(c) Est. PERMX (AutoAdaLoc) (d) Est. PERMX (P-AutoAdaLoc)

Figure 8: PERMX maps (in the natural logarithmic scale) on Layer 2 of (a) the reference model; (b) the mean reservoir model of the initial ensemble; (c) – (d) the mean reservoir models of the final estimated (est.) ensembles obtained by the IES algorithm with the AutoAdaLoc and the P-AutoAdaLoc schemes, respectively, in the Brugge benchmark case. In all the maps, the small dots indicate the locations of wells.

3D case: Brugge benchmark



(a) True PORO (b) Initial PORO

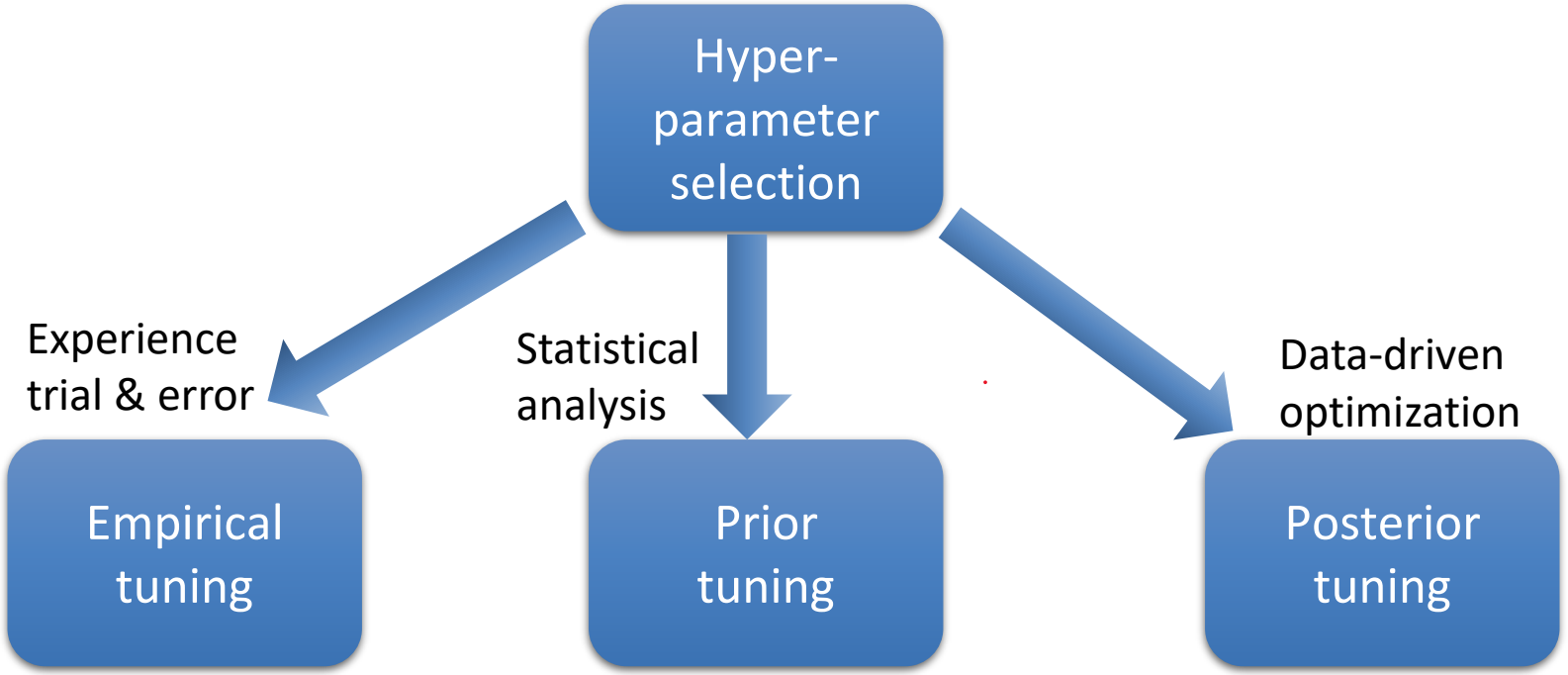
(c) Est. PORO (AutoAdaLoc) (d) Est. PORO (P-AutoAdaLoc)

Figure 9: As in Figure 8, but for PORO map on Layer 2 of the Brugge benchmark case.

Outline

- Background
- Parameterized correlation-based localization
- Numerical examples
- Summary

Summary



Example:	Correlation-based localization with hard-thresholding (Luo et. al 2018)	AutoAdaLoc (Luo & Bhakta 2020)	P-AutoAdaLoc (current work)

Q&A

The author(s) acknowledge partial financial supports from

- *The 4DSeis project, which is funded by industry partners, Equinor Energy AS, Lundin Energy Norway AS, Repsol Norge AS, Shell Global Solutions International B.V., TotalEnergies EP Norge AS, and Wintershall Dea Norge AS, as well as the Research Council of Norway (project number: 295002).*

- *The NCS2030 petro-center, which is funded by the Research Council of Norway (project number 331644) and industry partners AkerBP, DNO, Equinor, Halliburton, Okea, SLB, Vår Energy and Wintershall DEA.*

Continuous hyper-parameter optimization (CHOP)

Algorithm 1 Pseudo-code of the IES algorithm equipped with the original AutoAdaLoc localization scheme. The texts in red highlight places where differences take place, in comparison to the pseudo-code in Algorithm 2.

Require: Initial ensembles of reservoir models $\mathbf{M}^0 = \{\mathbf{m}_j^0\}_{j=1}^{N_e}$, simulated observations $\{\mathbf{g}(\mathbf{m}_j^0)\}_{j=1}^{N_e}$ and $\mathbf{g}(\overline{\mathbf{m}}^0)$; Initial regularization parameter γ^0

Construct the tapering matrix $\mathbf{T}$ ▷ cf. Eqs. 14 – 16
Iteration index $i \leftarrow 0$; IES not stopped $\leftarrow$ True
while IES not stopped **do**
  Calculate the Kalman-gain like matrix $\mathbf{K}^i$ for reservoir model update, e.g.,
    $\mathbf{K}^i = \mathbf{S}_m^i(\tilde{\mathbf{S}}_g^i)^T \left(\tilde{\mathbf{S}}_g^i(\tilde{\mathbf{S}}_g^i)^T + \gamma^i \mathbf{I}_p\right)^{-1}$ ▷ cf. Eqs. 4 – 9
  Update the reservoir models $\mathbf{m}_j^i$, e.g.,
    $\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \left(\mathbf{T} \circ \mathbf{K}^i\right)\left(\tilde{\mathbf{d}}_j^o - \tilde{\mathbf{g}}\left(\mathbf{m}_j^i\right)\right)$ ▷ cf. Eq. 18
  Run reservoir simulations to get the simulated observations $\{\mathbf{g}(\mathbf{m}_j^{i+1})\}_{j=1}^{N_e}$ and $\mathbf{g}(\overline{\mathbf{m}}^{i+1})$,
  Check data mismatch and update the value of γ^i ▷ cf. Luo et al. (2015)
  Apply stopping criteria to decide whether to stop the IES or not ▷ cf. Luo et al. (2015)
  $i \leftarrow i+1$
end while

Algorithm 2 Pseudo-code of the IES algorithm equipped with the P-AutoAdaLoc localization scheme. The texts in red highlight places where differences take place, in comparison to the pseudo-code in Algorithm 1.

Require: Initial ensembles of reservoir models $\mathbf{M}^0 = \{\mathbf{m}_j^0\}_{j=1}^{N_e}$, simulated observations $\{\mathbf{g}(\mathbf{m}_j^0)\}_{j=1}^{N_e}$ and $\mathbf{g}(\overline{\mathbf{m}}^0)$, localization length scales $\mathbf{L}^0 = \{\boldsymbol{\ell}_j^0\}_{j=1}^{Ne}$; Initial regularization parameter γ^0

Iteration index $i \leftarrow 0$; IES not stopped $\leftarrow$ True
while IES not stopped **do**
  Calculate the Kalman-gain like matrix $\mathbf{K}^i$ for reservoir model update, e.g.,
    $\mathbf{K}^i = \mathbf{S}_m^i(\tilde{\mathbf{S}}_g^i)^T \left(\tilde{\mathbf{S}}_g^i(\tilde{\mathbf{S}}_g^i)^T + \gamma^i \mathbf{I}_p\right)^{-1}$ ▷ cf. Eqs. 4 – 9
  for $j = 1, 2, \cdots, N_e$ **do**
    Construct the tapering matrix $\mathbf{T}\left(\boldsymbol{\ell}_j^i\right)$ with the set of length scales $\boldsymbol{\ell}_j^i$ ▷ cf. Eq. 19
    Update the reservoir models $\mathbf{m}_j^i$, e.g.,
      $\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \left(\mathbf{T}\left(\boldsymbol{\ell}_j^i\right) \circ \mathbf{K}^i\right)\left(\tilde{\mathbf{d}}_j^o - \tilde{\mathbf{g}}\left(\mathbf{m}_j^i\right)\right)$ ▷ cf. Eq. 18
  end for
  Run reservoir simulations to get the simulated observations $\{\mathbf{g}(\mathbf{m}_j^{i+1})\}_{j=1}^{N_e}$ and $\mathbf{g}(\overline{\mathbf{m}}^{i+1})$,
  Calculate the Kalman-gain like matrix $\mathbf{K}_\ell^i$ for hyper-parameter update, e.g.,
    $\mathbf{K}_\ell^i = \mathbf{S}_\ell^i(\tilde{\mathbf{S}}_g^{i+1})^T \left(\tilde{\mathbf{S}}_g^{i+1}(\tilde{\mathbf{S}}_g^{i+1})^T + \gamma^i \mathbf{I}_p\right)^{-1}$ ▷ cf. Eqs. 25 – 30
  for $j = 1, 2, \cdots, N_e$ **do**
    Construct the tapering matrix $\mathbf{T}_\ell\left(\boldsymbol{\ell}_j^i\right)$ with the set of length scales $\boldsymbol{\ell}_j^i$ ▷ similar to Eq. 19
    Update the set of localization length scale $\boldsymbol{\ell}_j^i$, e.g.,
      $\boldsymbol{\ell}_j^{i+1} = \boldsymbol{\ell}_j^i + \left(\mathbf{T}_\ell\left(\boldsymbol{\ell}_j^i\right) \circ \mathbf{K}_\ell^i\right)\left(\tilde{\mathbf{d}}_j^o - \tilde{\mathbf{g}}\left(\mathbf{m}_j^{i+1}\right)\right)$ ▷ cf. Eq. 31
  end for
  Check data mismatch and update the value of γ^i ▷ cf. Luo et al. (2015)
  Apply stopping criteria to decide whether to stop the IES or not ▷ cf. Luo et al. (2015)
  $i \leftarrow i+1$
end while