Covariance scaling: Theory and applications to ensemble-based history matching

Paulo Ranazzi [1] Xiaodong Luo [2] Marcio Sampaio [1]

[1]Departamento de Engenharia de Minas e de Petróleo, Escola Politécnica, Universidade de São Paulo [2]Norwegian Research Centre (NORCE), Bergen, Norway

Introduction

Motivation: Enhance the estimation accuracy of sample covariance matrices to reduce the effect of spurious correlations in ensemble-based history matching, especially when the true covariance structures are unknown.

Proposal: This work introduces a novel approach, called covariance scaling, to mitigate sampling errors in sample covariance matrices. This approach aims to find the optimal regularization parameter that minimizes the difference between a true covariance and its sample estimate. In contrast to other similar methods in the literature, such as the covariance shrinkage method, covariance scaling can be applied to improve the estimate of **a covariance-like matrix in an arbitrary shape**, including the cross-covariance matrix in the calculation of a Kalman gain, which is of particular interest to ensemble-based methods.

The covariance scaling method

Given a true covariance matrix $\mathbf{C} \in \mathbb{R}^{Y\times Y}$, its sample estimate $\hat{\mathbf{C}}$ from N samples, and a regularization parameter $\gamma \in [0,1]$, we seek a solution to the following minimization problem:

$$\min_{\gamma} \mathbb{E}\left[\left\|\gamma\hat{\mathbf{C}} - \mathbf{C}\right\|_{\mathrm{F}}^2\right], \tag{1}$$

where $\|\bullet\|_{\mathrm{F}}^2$ is the squared Frobenius norm, and $\|\mathbf{A}\|_{\mathrm{F}}^2 \equiv \mathrm{tr}(\mathbf{A}\mathbf{A}^{\mathrm{T}})$ for a matrix $\mathbf{A}$. It can be shown that the **"theoretical"** solution to Eq. 1 is given by

$$\gamma = \frac{\mathrm{tr}\left(\mathbf{C}\mathbf{C}^{\mathrm{T}}\right)}{\mathrm{tr}\left(\mathbf{C}\mathbf{C}^{\mathrm{T}}\right) + \frac{1}{N}\left[\mathrm{tr}\left(\mathbf{C}\mathbf{C}^{\mathrm{T}}\right) + \mathrm{tr}^2\left(\mathbf{C}\right)\right]}. \tag{2}$$

By partitioning the full covariance matrix $\mathbf{C}$ as

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_{\mathrm{mm}} & \mathbf{C}_{\mathrm{md}} \\ \mathbf{C}_{\mathrm{md}}^{\mathrm{T}} & \mathbf{C}_{\mathrm{dd}}\end{bmatrix}, \tag{3}$$

we can define an optimization problem similar to Eq. 1, but with the cross-covariance matrix $\mathbf{C}_{\mathrm{md}}$ as the target of approximation:

$$\min_{\gamma_{\mathrm{md}}} \mathbb{E}\left[\left\|\gamma_{\mathrm{md}}\hat{\mathbf{C}}_{\mathrm{md}} - \mathbf{C}_{\mathrm{md}}\right\|_{\mathrm{F}}^2\right]. \tag{4}$$

Based on the following identities:

$$\mathbb{E}\left[\hat{\mathbf{C}}_{\mathrm{md}}\right] = \mathbf{C}_{\mathrm{md}}, \tag{5}$$

$$\mathbb{E}\left[\mathrm{tr}\left(\hat{\mathbf{C}}_{\mathrm{md}}\hat{\mathbf{C}}_{\mathrm{md}}^{\mathrm{T}}\right)\right] = \mathrm{tr}\left(\mathbf{C}_{\mathrm{md}}\mathbf{C}_{\mathrm{md}}^{\mathrm{T}}\right) + \frac{1}{N}\left[\mathrm{tr}\left(\mathbf{C}_{\mathrm{md}}\mathbf{C}_{\mathrm{md}}^{\mathrm{T}}\right) + \mathrm{tr}\left(\mathbf{C}_{\mathrm{mm}}\right)\mathrm{tr}\left(\mathbf{C}_{\mathrm{dd}}\right)\right], \tag{6}$$

it can be shown that the **"theoretical"** solution to Eq. 4 is given by:

$$\gamma_{\mathrm{md}} = \frac{\mathrm{tr}\left(\mathbf{C}_{\mathrm{md}}\mathbf{C}_{\mathrm{md}}^{\mathrm{T}}\right)}{\mathrm{tr}\left(\mathbf{C}_{\mathrm{md}}\mathbf{C}_{\mathrm{md}}^{\mathrm{T}}\right) + \frac{1}{N}\left[\mathrm{tr}\left(\mathbf{C}_{\mathrm{md}}\mathbf{C}_{\mathrm{md}}^{\mathrm{T}}\right) + \mathrm{tr}\left(\mathbf{C}_{\mathrm{mm}}\right)\mathrm{tr}\left(\mathbf{C}_{\mathrm{dd}}\right)\right]}. \tag{7}$$

Approximating the optimal covariance scaling

Since $\mathbf{C}_{\mathrm{md}}$ is typically unknown, in a practical implementation, the "theoretical" solution γ_{md} needs to be calculated using the sample estimate $\hat{\mathbf{C}}_{\mathrm{md}}$. To this end, we first re-write Eq. 7 as

$$\gamma_{\mathrm{md}} = \frac{N}{N+1+\phi_{\mathrm{md}}}; \quad \phi_{\mathrm{md}} = \frac{\mathrm{tr}\left(\mathbf{C}_{\mathrm{mm}}\right)\mathrm{tr}\left(\mathbf{C}_{\mathrm{dd}}\right)}{\mathrm{tr}\left(\mathbf{C}_{\mathrm{md}}\mathbf{C}_{\mathrm{md}}^{\mathrm{T}}\right)}. \tag{8}$$

Similarly, define $\hat{\phi}_{\mathrm{md}} = \mathrm{tr}\left(\hat{\mathbf{C}}_{\mathrm{mm}}\right)\mathrm{tr}\left(\hat{\mathbf{C}}_{\mathrm{dd}}\right) / \mathrm{tr}\left(\hat{\mathbf{C}}_{\mathrm{md}}\hat{\mathbf{C}}_{\mathrm{md}}^{\mathrm{T}}\right)$, then using the iterative estimation idea in [1], we replace ϕ_{md} by $\hat{\phi}_{\mathrm{md}}/\gamma_{\mathrm{md},k}$ to account for sampling errors, resulting in the following recursive sequence:

$$\gamma_{\mathrm{md},k+1} = \frac{N}{N+1+\frac{\hat{\phi}_{\mathrm{md}}}{\gamma_{\mathrm{md},k}}}. \tag{9}$$

Then we obtain the **"practical"** solution

$$\hat{\gamma}_{\mathrm{md}} = \lim_{k\to\infty}\gamma_{\mathrm{md},k} = \begin{cases}0, & \text{if} \quad N < \hat{\phi}_{\mathrm{md}} \\ \frac{N-\hat{\phi}_{\mathrm{md}}}{N+1}, & \text{if} \quad N \geq \hat{\phi}_{\mathrm{md}}\end{cases} = \max\left(0, \frac{N-\hat{\phi}_{\mathrm{md}}}{N+1}\right) \tag{10}$$

Extension to the localization problem

The idea behind Eq. 9 can be similarly extended to the localization problem, e.g.,

$$\min_{\mathbf{P}} \quad \mathbb{E}\left[\left\|\mathbf{P}\circ\hat{\mathbf{C}} - \mathbf{C}\right\|_{\mathrm{F}}^2\right]; \quad \rho_{ij} \equiv [\mathbf{P}]_{i,j}. \tag{11}$$

The "**theoretical**" solution to Eq. 11 is [2]:

$$\rho_{ij} = \frac{N}{N+1+\frac{c_{ii}c_{jj}}{c_{ij}^2}}; \quad N \equiv \text{ensemble size}. \tag{12}$$

Using the idea of recursive sequence, the "**practical**" solution is given by:

$$\hat{\rho}_{ij} = \max\left(0, (N-\hat{c}_{ii}\hat{c}_{jj}/\hat{c}_{ij}^2)/(N+1)\right) = \max\left(0, (N-1/\hat{r}_{ij}^2)/(N+1)\right), \tag{13}$$

where $\hat{r}_{ij}$ represents the sample correlation between the i-th model variable and j-th model variable or simulated data point.

Two interesting observations:

1. Eq. 13 naturally induces a hard threshold for correlation-based localization, since $\hat{\rho}_{ij}$ will be set to 0 if $|\hat{r}_{ij}| \leq 1/\sqrt{N}$;
2. The threshold value, $1/\sqrt{N}$, corresponds to the (asymptotic) standard deviation of the sampling errors when the true correlation $r_{ij} = 0$ [3].

Numerical results in a 2D model

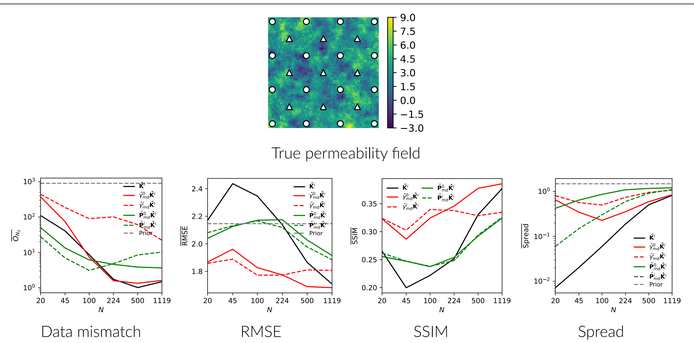

Figure 1. Performance measures versus ensemble size in the 2D toy model. Here, the black curves correspond the cases without any method to reduce the effect of spurious correlations, red lines to the cases with covariance scaling applied to the Kalman gain, green lines to the cases with Kalman gain localization, and gray dashed lines represent the prior mean values of the performance measures averaged over 10 different runs. Among the cases conducting scaling and localization on the Kalman gain, solid lines represent those where the regularization parameters were computed only with the prior ensembles, and the dashed lines stand for those where the regularization parameters were computed at each iteration step of the ESMDA.

Numerical results in the UNISIM-IV benchmark case

Two scenarios are considered:

- **Base:** Distance-based localization for petrophysical parameters
- **Scaling:** Distance-based localization for petrophysical parameters + covariance scaling for global parameters

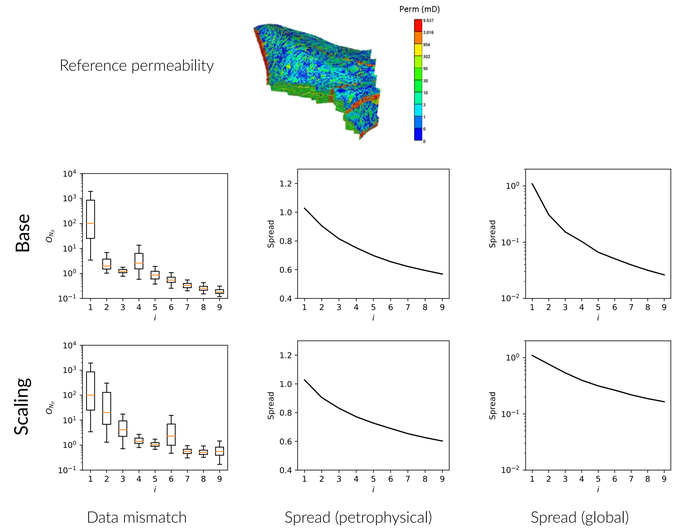

Figure 2. Performance measures with and without covariance scaling in the UNISIM-IV case.

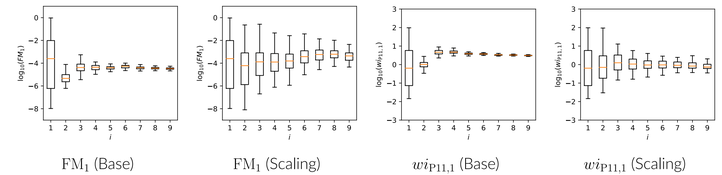

Figure 3. UNISIM-IV Fault multiplier index for Fault 1, and well index multiplier for producer well P11 (upper zone).

Acknowledgement

epic Research & Education · FAPESP · anp Agência Nacional do Petróleo, Gás Natural e Biocombustíveis · equinor · CMG · Universidade de São Paulo Escola Politécnica · LASG

References

[1] G. Letac and H. Massam, "All invariant moments of the wishart distribution," *Scandinavian Journal of Statistics*, vol. 31, no. 2, pp. 295–318, 2004.

[2] R. Furrer and T. Bengtsson, "Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants," *Journal of Multivariate Analysis*, vol. 98, pp. 227–255, Feb. 2007.

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

ranazzi@usp.br xluo@norceresearch.no marciosampaio@usp.br