Water Resources Research

Supporting Information for

Data Assimilation and Online Parameter Optimization in Groundwater Modeling using Nested Particle Filters

M. Ramgraber[1,2], C. Albert[3], M. Schirmer[1,2]

[1]Department Water Resources and Drinking Water, Swiss Federal Institute of Aquatic Science and Technology (Eawag)

[2]Centre d'hydrogéologie et de géothermie (CHYN), University of Neuchâtel

[3]Department Systems Analysis, Integrated Assessment and Modelling, Swiss Federal Institute of Aquatic Science and Technology (Eawag)

Contents of this file

Additional Supporting Information (Files uploaded separately)

Introduction

This supporting document contains several auxiliary figures illustrating the functioning of the meander-based field generator or detailing the pseudo-code for the algorithm used in this study. We furthermore provide videos visualizing the optimization of the parameter and state distributions through time for all field generators and random seeds. Supporting tables detail the probability and nature of mutation operations used for updating the hyperparameters, as well as various model specifications.

Initialization:
At initialization, generate N_θ sets of hyperparameters and use the field generator to retrieve corresponding parameter fields. Similarly, generate ensembles of N_x state particles from a suitable prior.

For each cycle c from 1 to m:

For each θ-particle $\theta_c^{n_\theta}$, n_θ from 1 to N_θ:

Mutate θ-particles

$$\theta_c^{(n_\theta)} \sim k\left(\theta_c^{(n_\theta)} | \theta_{c-1}^{(n_\theta)}\right)$$

Hyperspace:
- mutate hyperparameters
- map into full parameter space

Go through sub-steps

For each sub-step u from 1 to L, introducing cycle-dependent subscript $z = (c-1)L + u$:

Propagate x-particles

For each x-particle $x_z^{(n_\theta, n_x)}$, n_x from 1 to N_x:

Propagate states to next available observation with deterministic model, add error

$$x_z^{(n_\theta, n_x)} = model\left(x_{z-1}^{(n_\theta, n_x)}, \theta_c^{(n_\theta)}\right) + \mathcal{J}\epsilon_x, \text{ with } \epsilon_x \sim \mathcal{N}(0, \sigma_{model}^2) \text{ and } \mathcal{J} = (1, \dots, 1)^T$$

For each observation $y_z^{(o)}$, o from 1 to N_{obs}:

Determine likelihood

$$l_z^{(n_\theta, n_x, o)} = \frac{1}{\sqrt{2\pi\sigma_{obs}^2}} exp\left(-\frac{\left(x_z^{(n_\theta, n_x, o)} - y_z^{(o)}\right)^2}{2\sigma_{obs}^2}\right)$$

Calculate un-normalized x-particle weights

$$W^{(i,n)} = \prod_{o=1}^{N_{obs}} l_z^{(n_\theta, n_x, o)}$$

Normalize x-particle weights

For each x-particle $x_z^{(n_\theta, n_x)}$, n_x from 1 to N_x:

$$w^{(n_\theta, n_x)} = \frac{W^{(n_\theta, n_x)}}{\sum_{n=1}^{N_x} W^{(n_\theta, n_x)}}$$

Resample x-particles

For each x-particle $x_z^{(n_\theta, n_x)}$, n_x from 1 to N_x:

Individually sample ancestral particle index from multinomial distribution

$$a \sim \mathcal{M}\left(1, \left(w^{(n_\theta, 1)}, \dots, w^{(n_\theta, N_x)}\right)\right)$$

Inherit state vector from ancestor

$$x_z^{(n_\theta, n_x)} = x_z^{(a, n_x)}$$

Calculate un-normalized θ-particle weights

For each θ-particle $\theta_c^{(n_\theta)}$, i from 1 to N_θ:

$$W^{(n_\theta)} = \sum_{n_x=1}^{N_x} \prod_{u=1}^{L} \prod_{o=1}^{N_{obs}} l_z^{(n_\theta, n_x, o)}$$

Normalize θ-particle weights

For each θ-particle $\theta_c^{(n_\theta)}$, n_θ from 1 to N_θ:

$$w^{(n_\theta)} = \frac{W^{(n_\theta)}}{\sum_{i=1}^{N_\theta} W^{(n_\theta)}}$$

Resample θ-particles

For each θ-particle $\theta_c^{(n_\theta)}$, n_θ from 1 to N_θ:

Sample ancestral particle index from multinomial distribution

$$a \sim \mathcal{M}\left(1, \left(w^{(1)}, \dots, w^{(N_\theta)}\right)\right)$$

Inherit θ-particle trajectory from ancestor

$$\theta_c^{(n_\theta)} = \theta_c^{(a)}$$

Hyperspace:
- Inherit ancestor's hyperparameters

Inherit inner particle filter from ancestor

$$x_z^{(n_\theta, \cdot)} = x_z^{(a, \cdot)}$$

Figure S 1. Pseudo-code for the nested particle filter algorithm as employed in this study

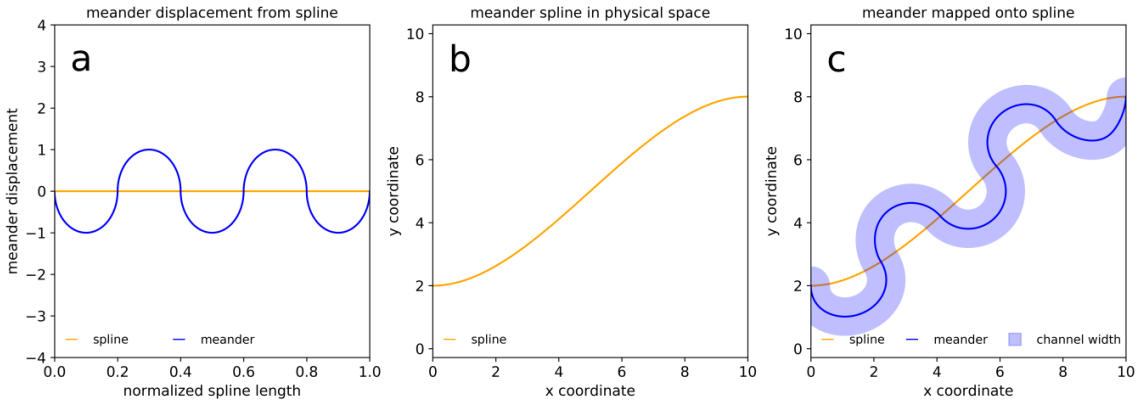

Figure S 2. Conceptual steps of the meander generator: the meander displacement relative to a parent spline is determined, based on hyperparameters 'number of meander turns' and 'phase shift' (a); the parent spline is created in physical space, based on hyperparameters 'start point', 'end point', and their 'first derivatives' (b); meander displacement is multiplied by 'meander width' and mapped onto the parent spline; adherence to 'meander facies' is determined based on hyperparameter 'channel width' (c).

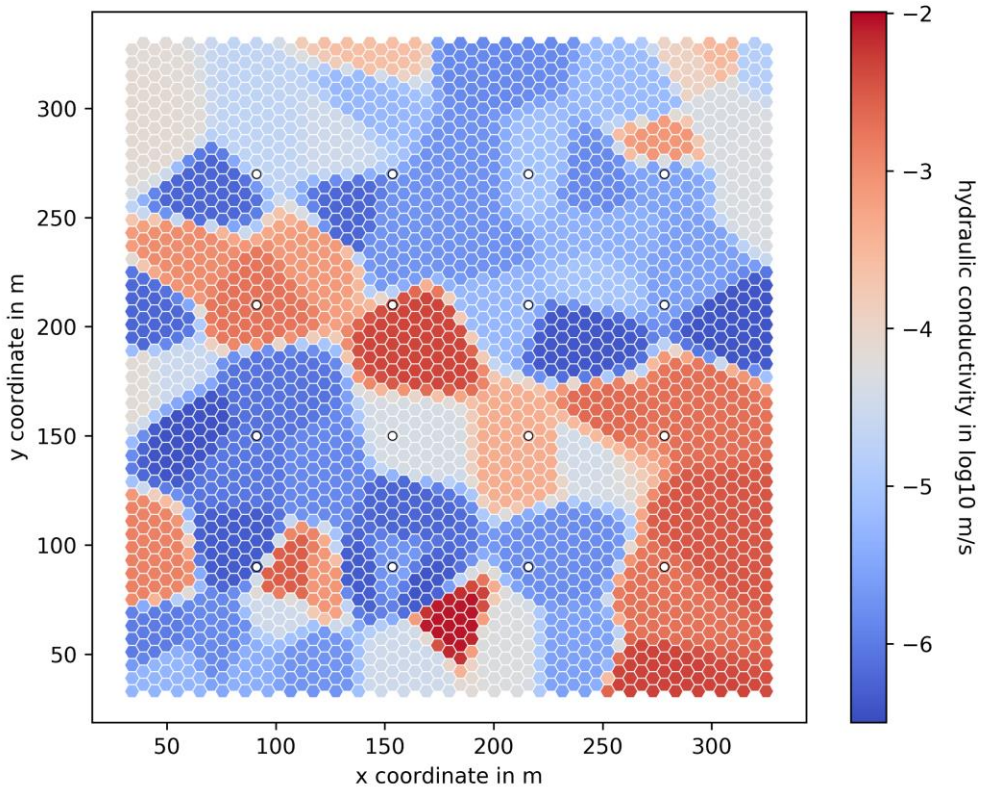

Figure S 3. Ensemble average of hydraulic conductivity for the hybrid nested particle (node-based scenario) at the end of the assimilation period. As in the nested particle filter scenarios, parameter uncertainty has collapsed.

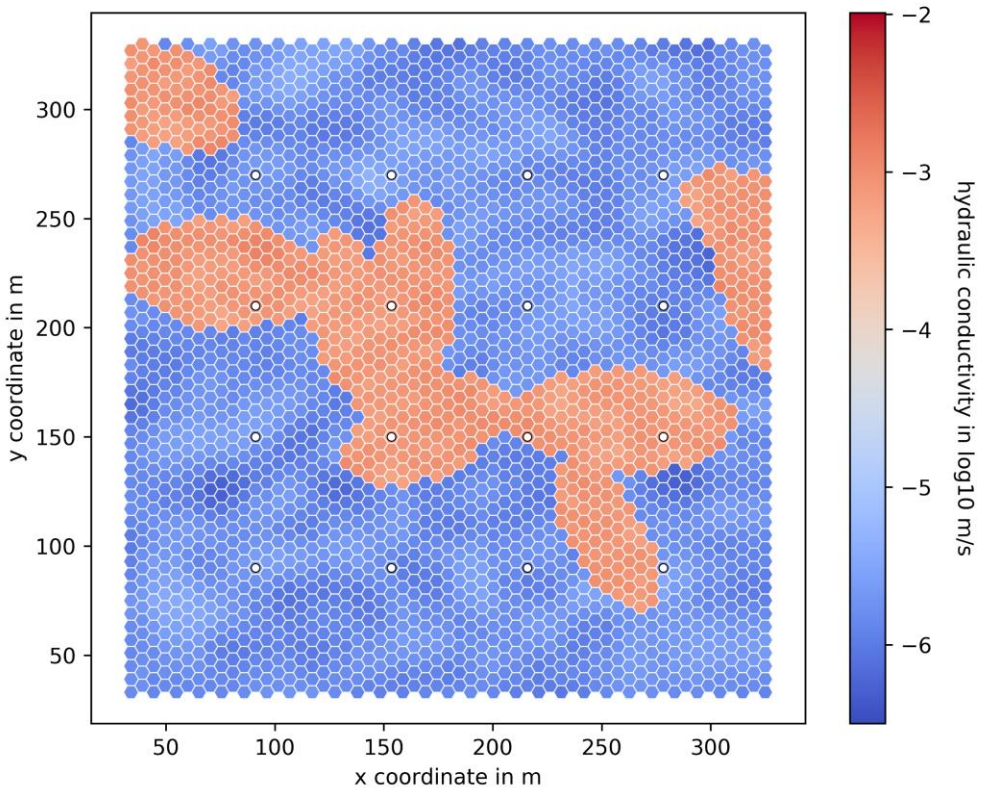

Figure S 4. Ensemble average of hydraulic conductivity for the hybrid nested particle (lens-based scenario) at the end of the assimilation period. As in the nested particle filter scenarios, parameter uncertainty has collapsed.

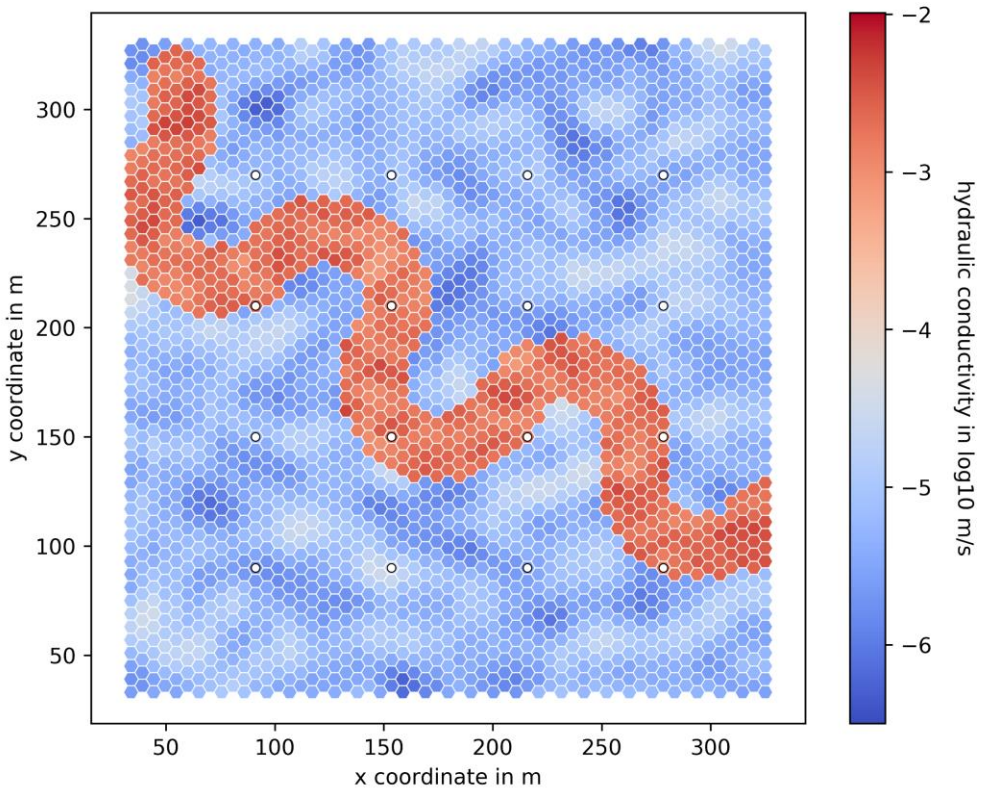

Figure S 5. Ensemble average of hydraulic conductivity for the hybrid nested particle (meander-based scenario) at the end of the assimilation period. As in the nested particle filter scenarios, parameter uncertainty has collapsed.

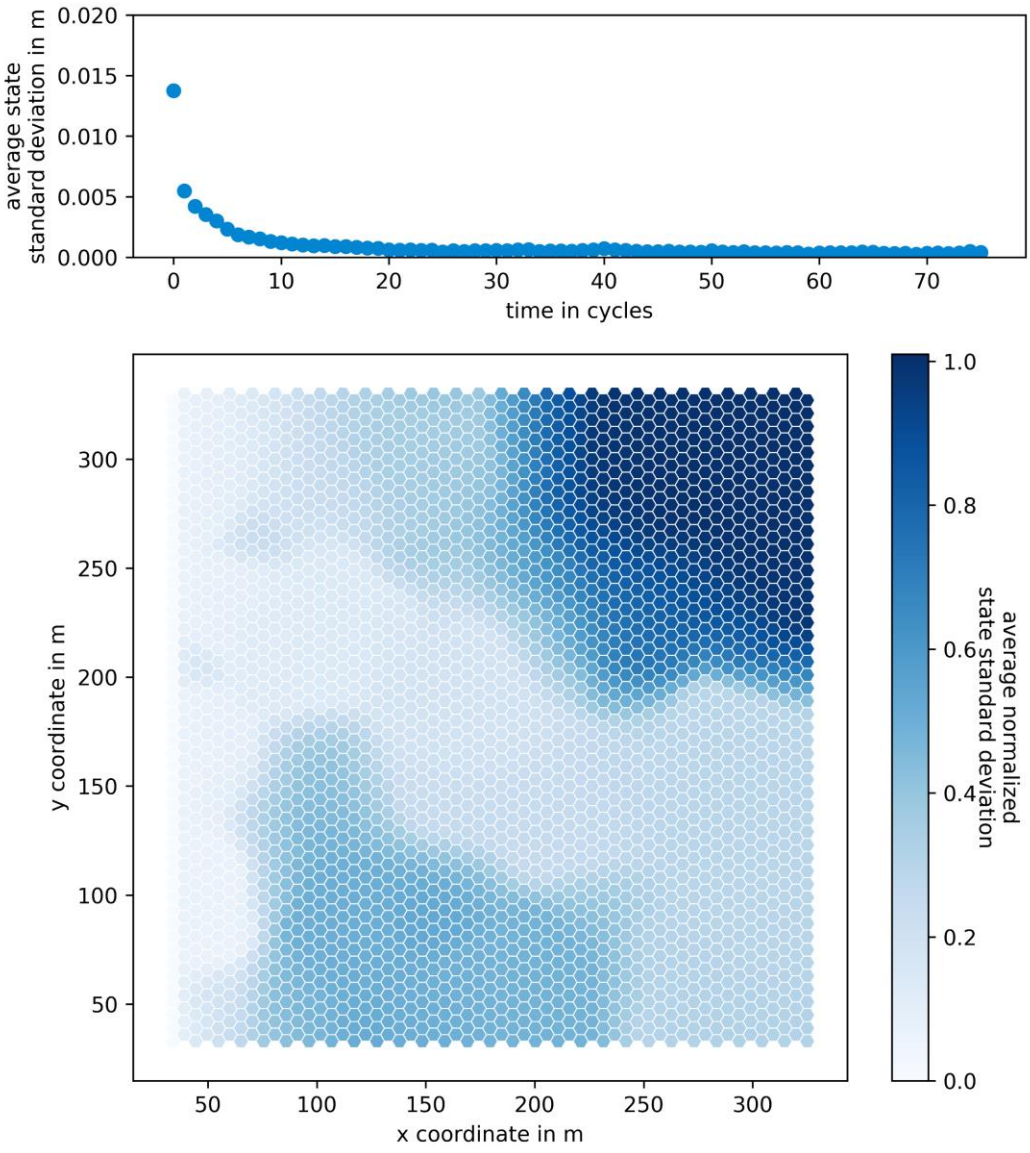

Figure S 6. Development of the standard deviation of the inner EnKFs' state uncertainties averaged across all inner filters and grid cells over time (top). The lower subplot shows the normalized average of the inner EnKFs' spatial standard deviation distribution. Results are shown for the node-based scenario of the hybrid nested particle filter.

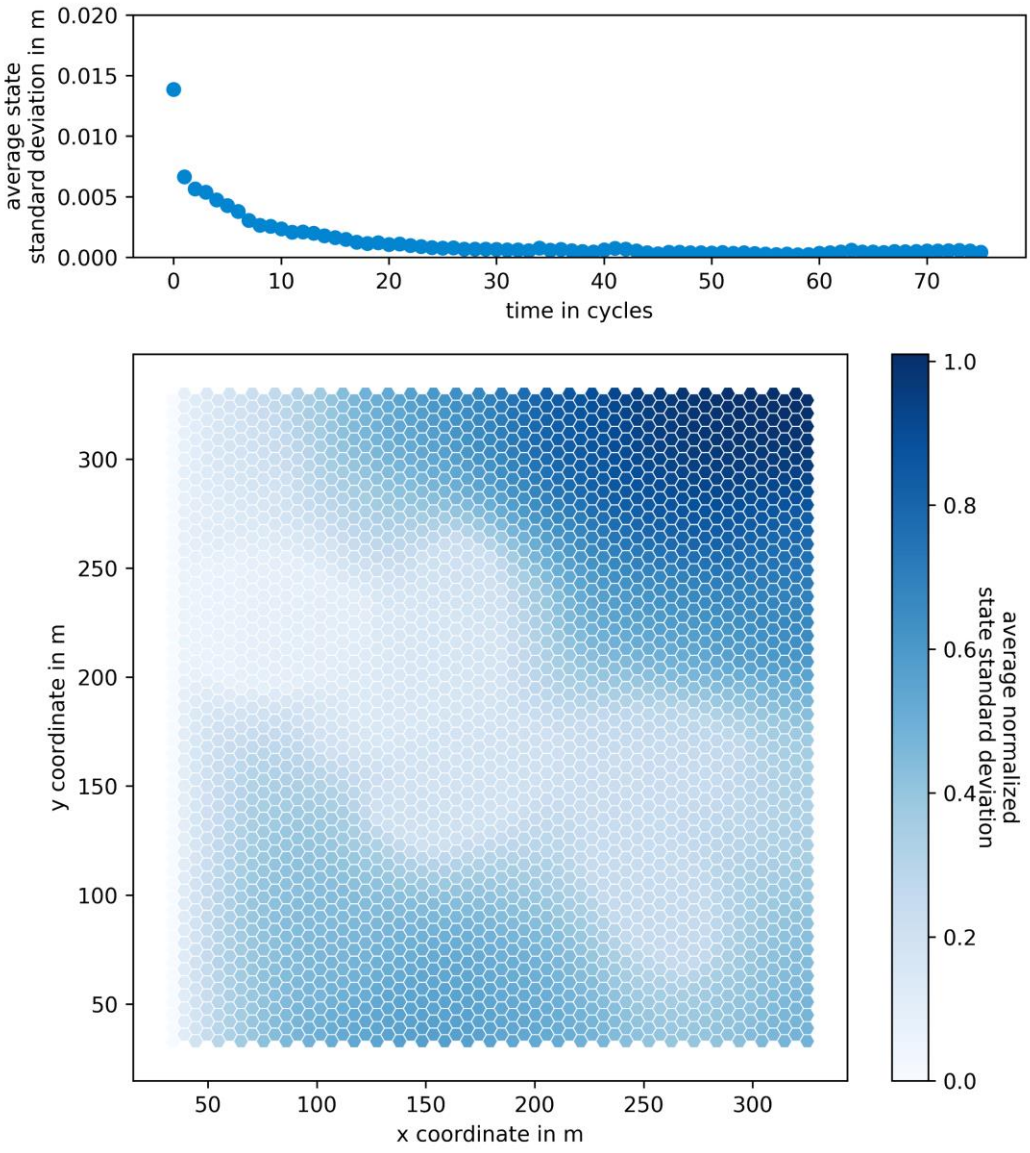

Figure S 7. Development of the standard deviation of the inner EnKFs' state uncertainties averaged across all inner filters and grid cells over time (top). The lower subplot shows the normalized average of the inner EnKFs' spatial standard deviation distribution. Results are shown for the lens-based scenario of the hybrid nested particle filter.

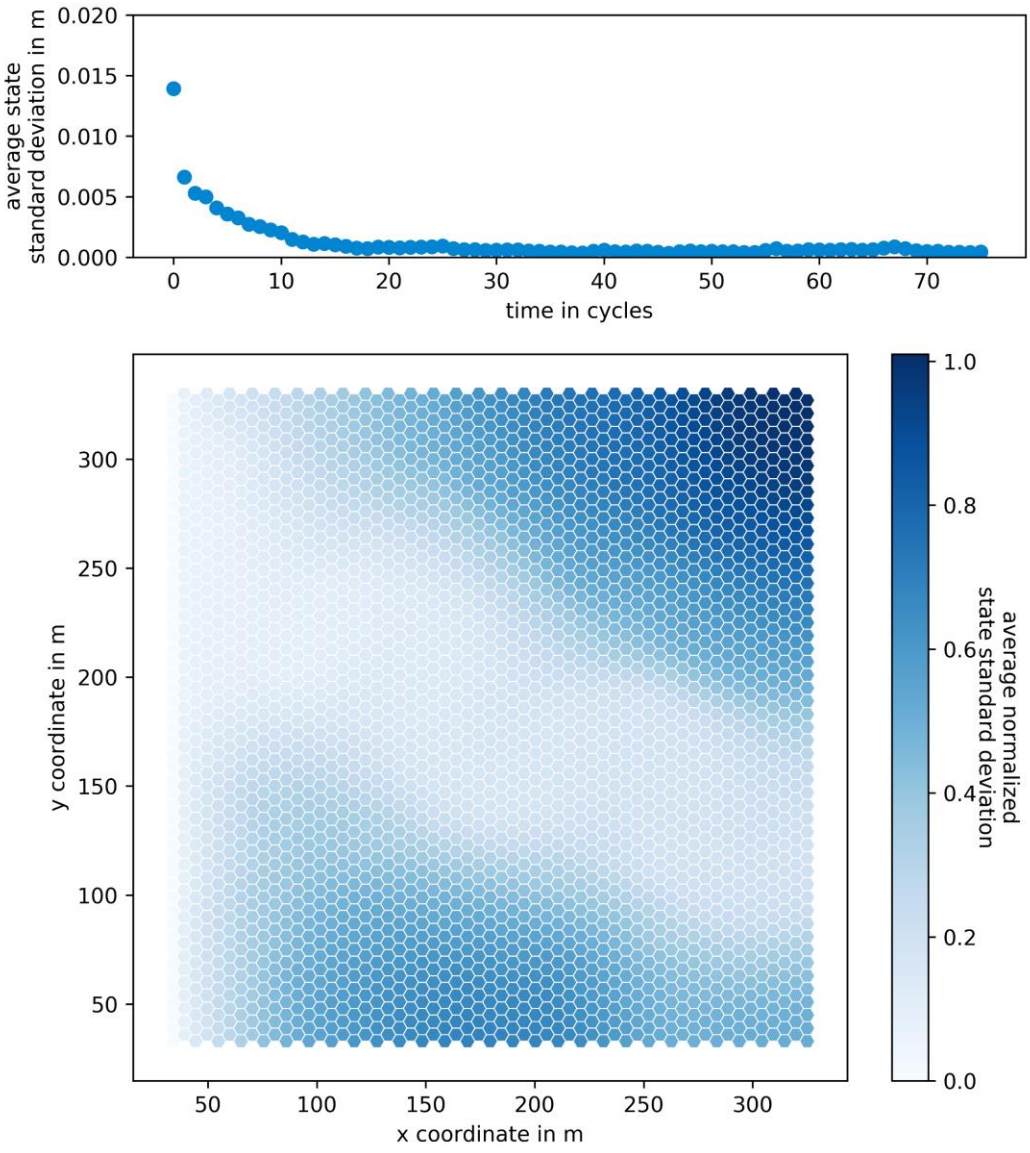

Figure S 8. Development of the standard deviation of the inner EnKFs' state uncertainties averaged across all inner filters and grid cells over time (top). The lower subplot shows the normalized average of the inner EnKFs' spatial standard deviation distribution. Results are shown for the meander-based scenario of the hybrid nested particle filter.

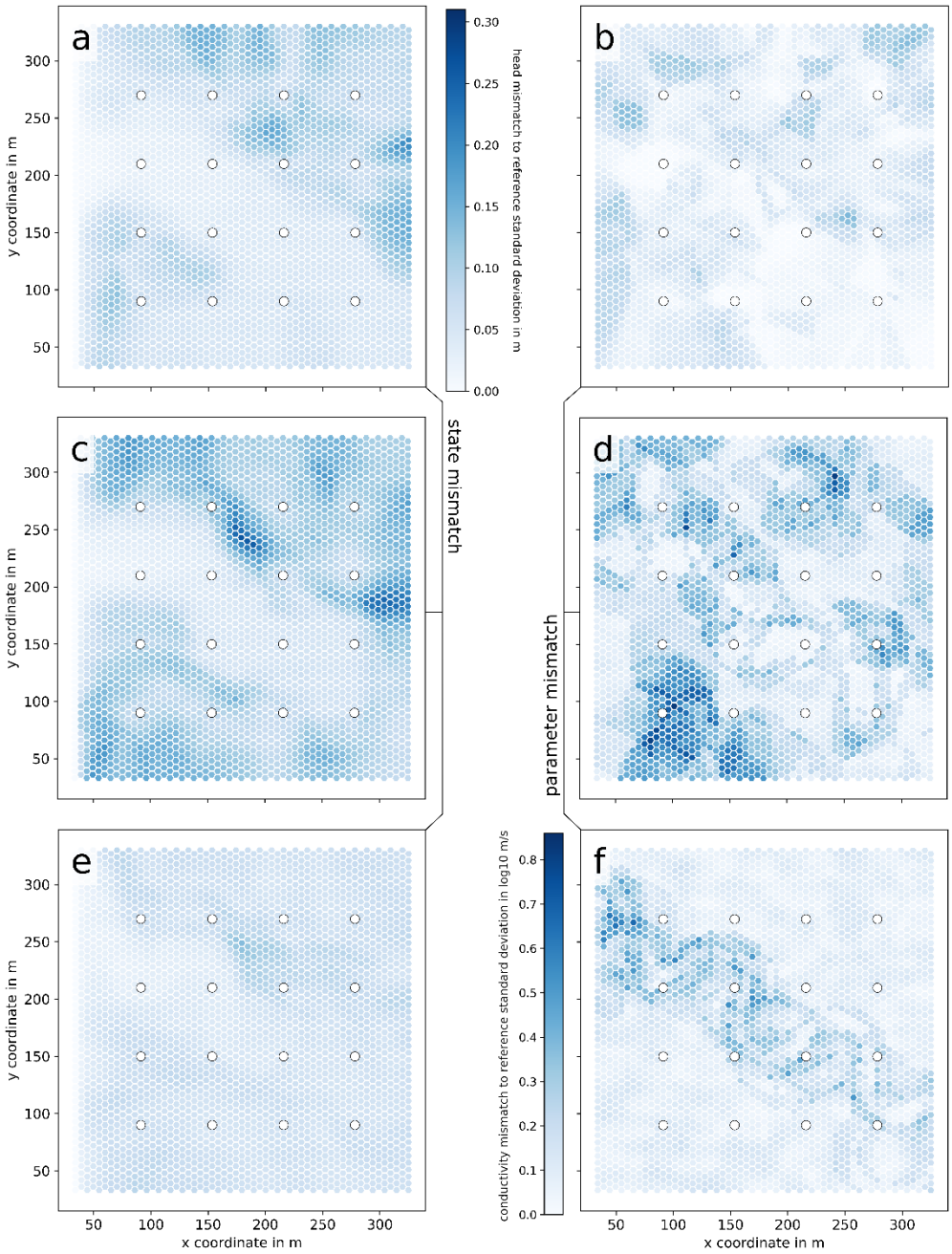

Figure S 9. Standard deviation of the mismatch between the synthetic reference and ensemble means at the end of the data assimilation period for hydraulic heads (a, c, e) and hydraulic conductivities (b, d, f) across ten random seeds. Results are shown for the node-based (a, b), lens-based (c, d), and meander-based (e, f) hyperparameterizations.

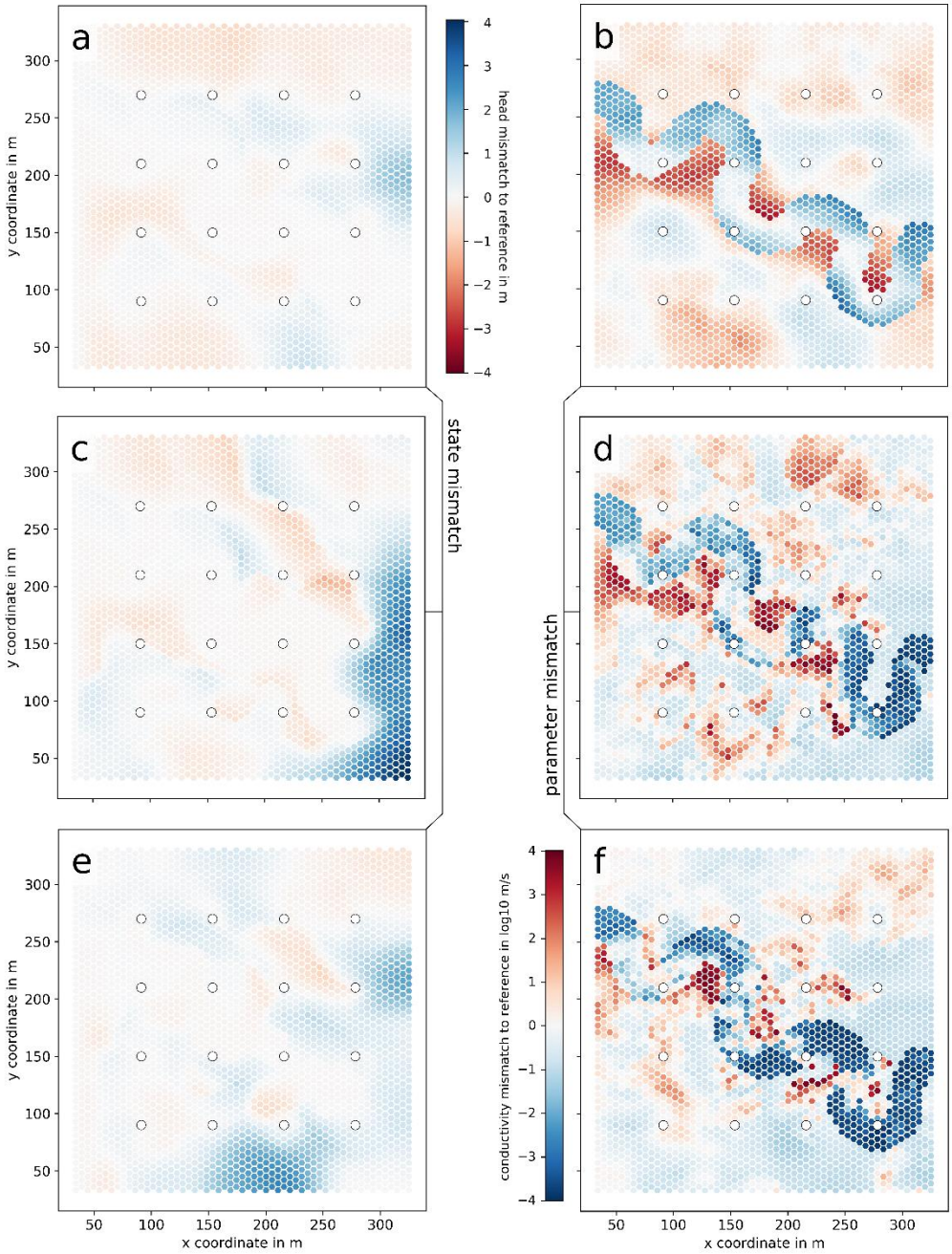

Figure S 10. Standard EnKF mismatch between the synthetic reference and ensemble mean at the end of the data assimilation period for hydraulic heads (a, c, e) and hydraulic conductivities (b, d, f). Results are shown for the node-based (a, b), lens-based (c, d), and meander-based (e, f) hyperparameterizations. Mind that the relation of colors to quantities is reversed between the state mismatch (a, c, e) and the parameter mismatch (b, d, f) columns. This was a deliberate choice to visually underline the common relation of parameter underestimation to state overestimation, and vice versa.

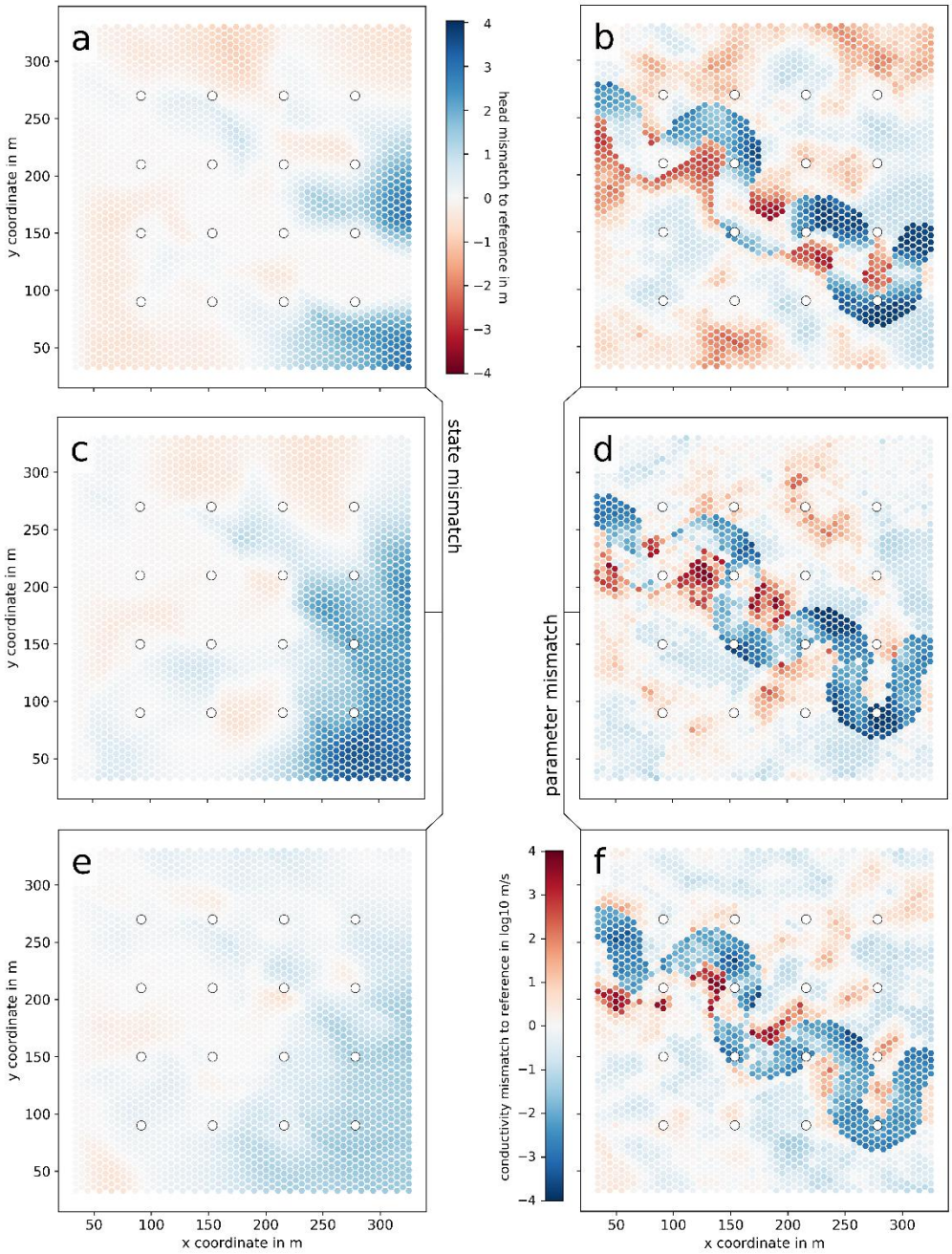

Figure S 11. GA-EnKF mismatch between the synthetic reference and ensemble mean at the end of the data assimilation period for hydraulic heads (a, c, e) and hydraulic conductivities (b, d, f). Results are shown for the node-based (a, b), lens-based (c, d), and meander-based (e, f) hyperparameterizations. Mind that the relation of colors to quantities is reversed between the state mismatch (a, c, e) and the parameter mismatch (b, d, f) columns. This was a deliberate choice to visually underline the common relation of parameter underestimation to state overestimation, and vice versa.

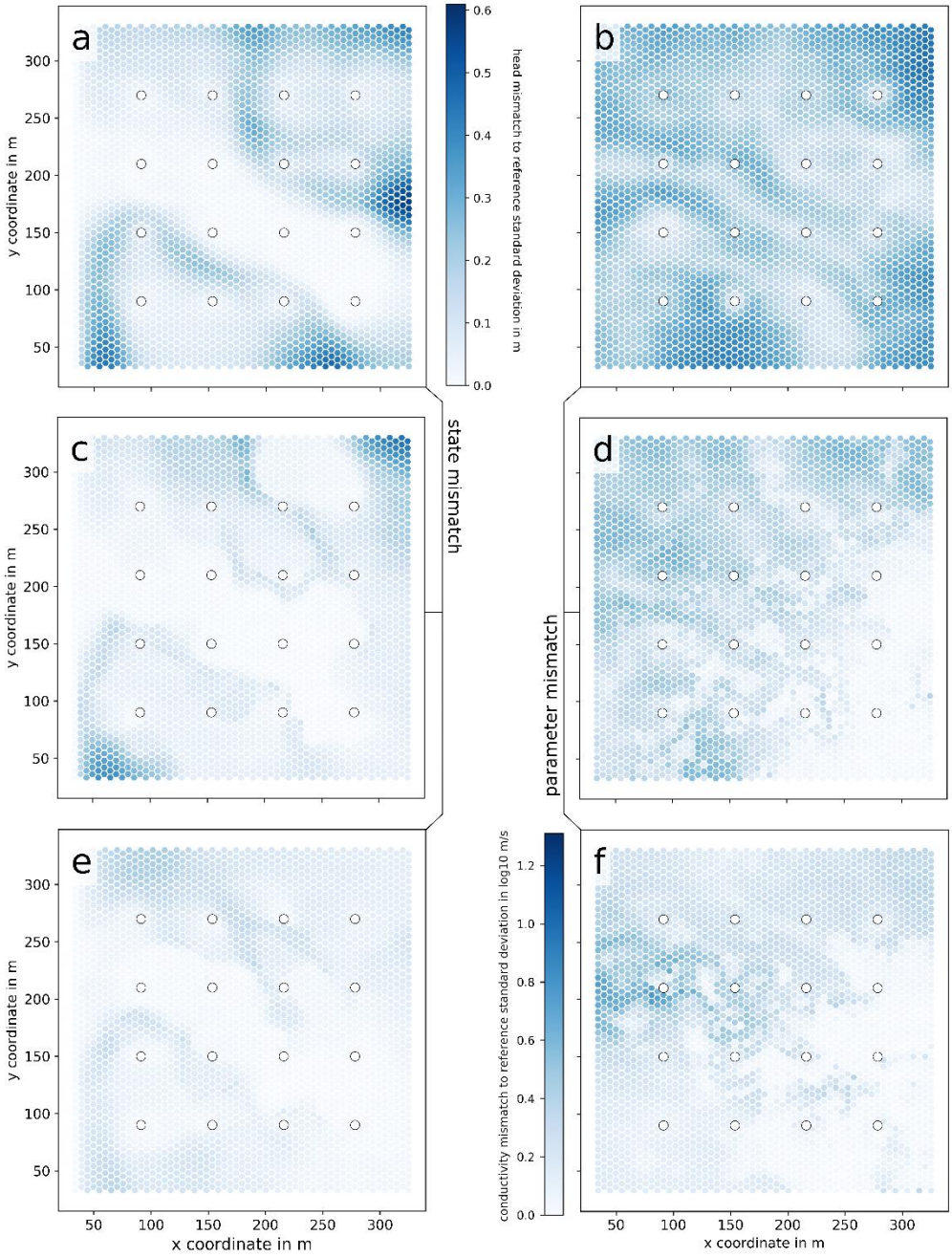

Figure S 12. Standard deviation of the mismatch between the synthetic reference and ensemble of the standard EnKF at the end of the data assimilation period for hydraulic heads (a, c, e) and hydraulic conductivities (b, d, f). Results are shown for the node-based (a, b), lens-based (c, d), and meander-based (e, f) hyperparameterizations.

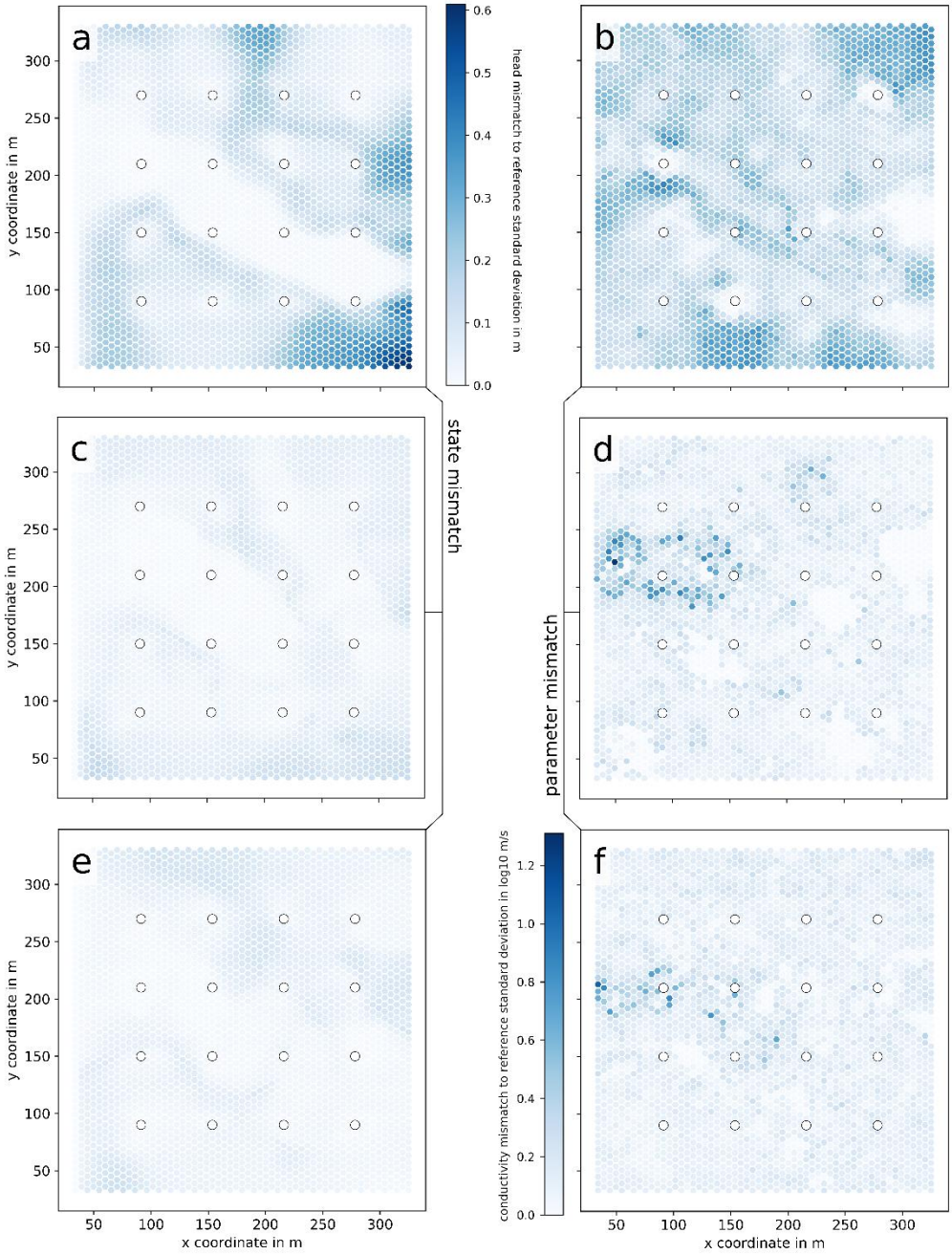

Figure S 13. Standard deviation of the mismatch between the synthetic reference and ensemble of the GA-EnKF at the end of the data assimilation period for hydraulic heads (a, c, e) and hydraulic conductivities (b, d, f). Results are shown for the node-based (a, b), lens-based (c, d), and meander-based (e, f) hyperparameterizations.

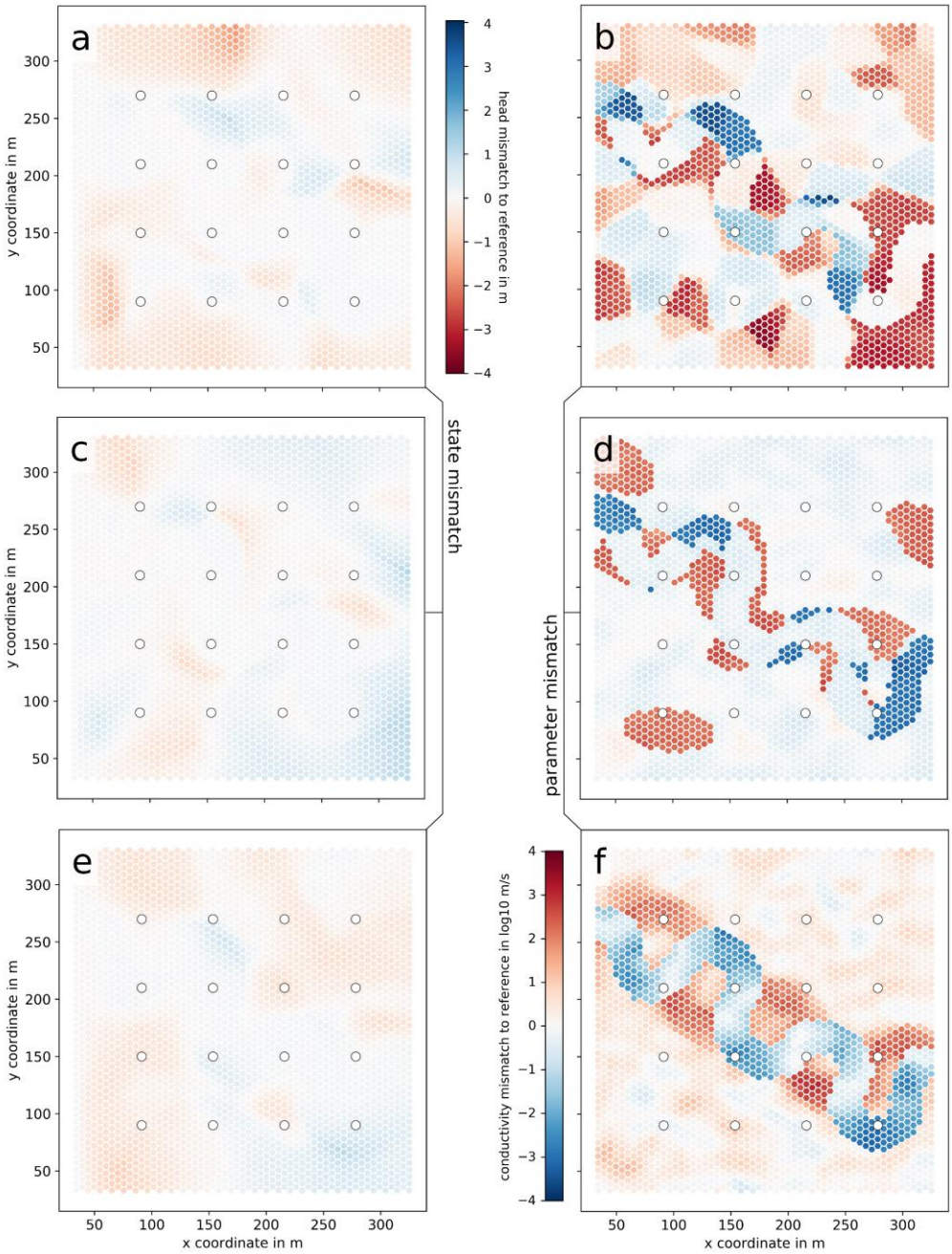

Figure S 14. Mismatch between the synthetic reference and ensemble mean at the end of the data assimilation period for hydraulic heads (a, c, e) and hydraulic conductivities (b, d, f) for the hybrid nested particle filter with inner EnKFs. Results are shown for the node-based (a, b), lens-based (c, d), and meander-based (e, f) hyperparameterizations. Mind that the relation of colors to quantities is reversed between the state mismatch (a, c, e) and the parameter mismatch (b, d, f) columns. This was a deliberate choice to visually underline the common relation of parameter underestimation to state overestimation, and vice versa.

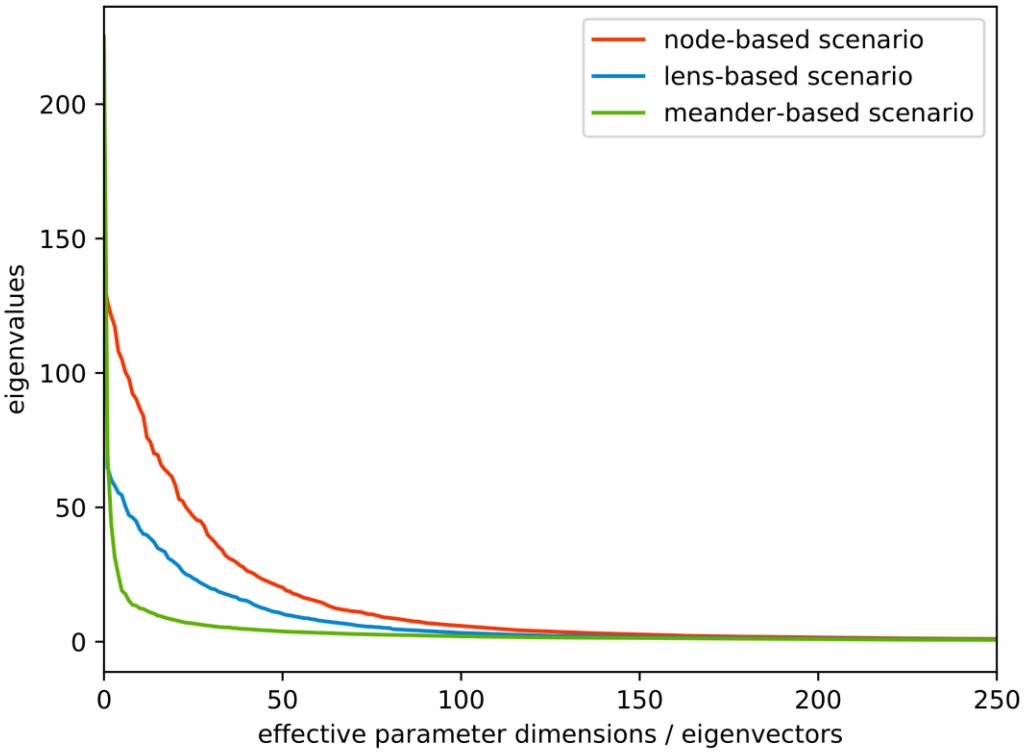

Figure S 15. Eigenvalues of the covariance matrices of the initial parameter ensembles for the three different geological characterizations. The node-based case has the largest effective dimensionality (the most non-zero eigenvalues), followed by the lens-based scenario. The meander-based scenario, perhaps unsurprisingly, has the fewest effective dimensions.

Table S1. Variables used in model discretization and filter setup.

Model variable	Value
Cell count	2850
Hexagon radius	3 m
Cell top elevation	10 m
Cell bottom elevation	-10 m
Recharge period length	25 d
Recharge mean	1.07E-9 m/s
Total simulation time	750 d
Filter variable	**Value**
State measurement frequency	every 24 h
N_θ	200
N_x	5 per parameter particle
L	10
σ_M	0.0001 m
σ_{obs}	0.02 m

Table S2. Random mutation operations for the node-based field generator.

Chance	Random operation
10%	Add or remove a random node, chance dependent on whether node count is below or above the target count of 50 nodes. New nodes are placed randomly and assigned a random hydraulic conductivity.
5%	Remove a random node and add a new one.
15%	Move a random node within a user-specified radius, here 50 m.
35%	Change a random node's log hydraulic conductivity by adding a random value drawn from a standard normal distribution. Log hydraulic conductivities are bounded between -2.5 and -6.
15%	Adopt hydraulic conductivity of a random node within a radius of 100 m.
20%	Switch hydraulic conductivity of two random nodes within a radius of 100 m of each other.

Table S3. Random mutation operations for the lens-based field generator.

Chance	Random operation
10%	Add or remove a random lens, chance dependent on whether the lens count is below or above target count of 12 lenses. New lenses are placed randomly and assigned random rotation, size, and aspect.
20%	Remove a random lens and add a new one.
30%	Move a random lens within a user-specified radius, here 50 m.
5%	Change a random lens's size by adjusting the length of its primary axis, bounded between 75 m and 90 m.
15%	Change a random lens's rotation.
5%	Change a random lens's aspect between primary and secondary axis, bounded between 1.75 and 2.25.
15%	Change hydraulic conductivity of one of the facies. Draw a value from a standard normal distribution, then generate a new Gaussian parameter field with the specified

	mean, isotropic spatial correlation, and an amplitude of 1. Log hydraulic conductivities are bounded between -2.5 and -6.

Table S4. Random mutation operations for the meander-based field generator.

Chance	Random operation
10%	Move start or end point of the meander direction spline.
15%	Change first derivative of start or end of the meander direction spline, bounded between -1 and 1.
15%	Adjust number of meander turns, bounded between 5 and 9.
20%	Adjust channel width, bounded between 5 and 50 % of the meander 'wavelength'.
20%	Adjust meander phase shift.
20%	Change hydraulic conductivity of one of the facies. Draw a value from a standard normal distribution, then generate a new Gaussian parameter field with the specified mean, isotropic spatial correlation, and an amplitude of 1. Log hydraulic conductivities are bounded between -2.5 and -6.

Movie S1. Type or paste caption here (upload your movie(s) to AGU's journal submission site and select, "Supporting Information (SI)" as the file type. Following naming convention: ms01.

Movie S 1. Evolution of expected parameter field for random seed 0, node-based scenario.

Movie S 2. Evolution of expected parameter field for random seed 1, node-based scenario.

Movie S 3. Evolution of expected parameter field for random seed 2, node-based scenario.

Movie S 4. Evolution of expected parameter field for random seed 3, node-based scenario.

Movie S 5. Evolution of expected parameter field for random seed 4, node-based scenario.

Movie S 6. Evolution of expected parameter field for random seed 5, node-based scenario.

Movie S 7. Evolution of expected parameter field for random seed 6, node-based scenario.

Movie S 8. Evolution of expected parameter field for random seed 7, node-based scenario.

Movie S 9. Evolution of expected parameter field for random seed 8, node-based scenario.

Movie S 10. Evolution of expected parameter field for random seed 9, node-based scenario.

Movie S 11. Evolution of expected parameter field for random seed 0, lens-based scenario.

Movie S 12. Evolution of expected parameter field for random seed 1, lens-based scenario.

Movie S 13. Evolution of expected parameter field for random seed 2, lens-based scenario.

Movie S 14. Evolution of expected parameter field for random seed 3, lens-based scenario.

Movie S 15. Evolution of expected parameter field for random seed 4, lens-based scenario.

Movie S 16. Evolution of expected parameter field for random seed 5, lens-based scenario.

Movie S 17. Evolution of expected parameter field for random seed 6, lens-based scenario.

Movie S 18. Evolution of expected parameter field for random seed 7, lens-based scenario.

Movie S 19. Evolution of expected parameter field for random seed 8, lens-based scenario.

Movie S 20. Evolution of expected parameter field for random seed 9, lens-based scenario.

Movie S 21. Evolution of expected parameter field for random seed 0, meander-based scenario.

Movie S 22. Evolution of expected parameter field for random seed 1, meander-based scenario.

Movie S 23. Evolution of expected parameter field for random seed 2, meander-based scenario.

Movie S 24. Evolution of expected parameter field for random seed 3, meander-based scenario.

Movie S 25. Evolution of expected parameter field for random seed 4, meander-based scenario.

Movie S 26. Evolution of expected parameter field for random seed 5, meander-based scenario.

Movie S 27. Evolution of expected parameter field for random seed 6, meander-based scenario.

Movie S 28. Evolution of expected parameter field for random seed 7, meander-based scenario.

Movie S 29. Evolution of expected parameter field for random seed 8, meander-based scenario.

Movie S 30. Evolution of expected parameter field for random seed 9, meander-based scenario.