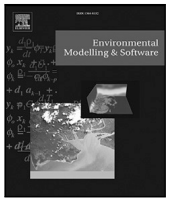
Environmental Modelling and Software 156 (2022) 105498

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Sequential and batch data assimilation approaches to cope with groundwater model error: An empirical evaluation

Katherine H. Markovich [a,*], Jeremy T. White [b], Matthew J. Knowling [c,d]

[a] INTERA Incorporated, Albuquerque, NM, USA
[b] INTERA Incorporated, Boulder, CO, USA
[c] School of Civil, Environmental and Mining Engineering, Faculty of Engineering, Computer and Mathematical Sciences, The University of Adelaide, Adelaide, Australia
[d] School of Agriculture, Food and Wine, Faculty of Sciences, Engineering and Technology, The University of Adelaide, Adelaide, Australia

ARTICLE INFO

Dataset link: https://github.com/kmarkovich-intera/Freyberg_DA/releases/tag/v2.0.0

Keywords:
Groundwater decision support modeling
Data assimilation
Model error

ABSTRACT

Groundwater model data assimilation (DA) aims to reduce uncertainty in simulated outcomes of interest to resource management while minimizing the potential for predictive bias. Sequential DA, which can estimate model states along with properties and stresses dynamically in time, offers a potentially powerful alternative to batch DA (i.e., history matching) for reducing bias in decision-relevant predictions in the presence of incorrect model structure and/or processes. This study evaluates the ability of batch and sequential DA approaches to history match and forecast simulated quantities in the presence of model error using a novel ensemble-based paired complex–simple approach that enables the incorporation of stochastic uncertainty and a statistical evaluation of predictive bias. Our findings have implications for groundwater decision support modeling as they underscore the pitfalls of fixing parameters and forcing variables *a priori* and present a proof of concept for using adjustable model states to cope with model error.

1. Introduction

Groundwater models are, by necessity and by design, a simplified representation of the infinitely complex real-world hydrogeological systems. This simplification and abstraction may arise from issues of spatial or temporal scale, where the discretization of a modeled process is too coarse to capture fine-scale governing dynamics or the upscaling of properties smooths out temporal and/or spatial heterogeneity. Or the simplification and abstraction may refer to mathematical representation of physical processes, whether they be physically motivated or empirical schemes. The cumulative effect of these well-documented limitations in groundwater models is referred to as model error or model misspecification, and this discrepancy gives rise to structural noise in the simulated results (Doherty and Welter, 2010). Many workers have sought to address the degree to which this abstraction and simplification produces deleterious effects on predictions (see Gupta et al. (2012), White et al. (2014) and references therein).

The primary consequences of using simplified models for making predictions are parameter compensation (Clark and Vrugt, 2006a; Doherty and Christensen, 2011) and resulting potential for predictive bias. Parameter compensation describes the tendency for parameters or parameter components which are not sensitive to the information content of the observations (i.e, null space parameters) to be erroneously adjusted during model calibration (Doherty and Christensen, 2011; Watson et al., 2013). This occurs because the model structure—including discretization and process representation—is not sufficient to represent the information contained in the assimilated observations, and signal components corresponding to this insufficiency in the observations are incorrectly assimilated. In many cases, the propensity for parameter compensation may be "visible" as prior-data conflict (Evans and Moshonov, 2006). Prior-data conflict arises from the (simple) model's inability to simulate the proper range and/or characteristics of the hydrologic state at a given observation location and/or time. For predictions of management interest that are sensitive to these erroneously adjusted parameters, parameter compensation leads to predictive bias, which compromises the model's fidelity for decision making despite undergoing rigorous calibration and uncertainty analysis. Particularly concerning is that parameter compensation is often invisible to the modeler (White et al., 2014; Knowling et al., 2019).

A recent evolution in how best to undertake data assimilation (DA) for decision support modeling is the idea that models should strive for an appropriate balance of simplicity complexity (Doherty and Moore, 2020; Hunt et al., 2007). That is, they should be simple enough to be

* Corresponding author.
E-mail address: kmarkovich@intera.com (K.H. Markovich).

https://doi.org/10.1016/j.envsoft.2022.105498
Received 20 July 2022; Received in revised form 9 August 2022; Accepted 10 August 2022
Available online 26 August 2022

computational cheap and numerically stable, yet possess appropriately high-dimensional and diverse parameterization to receive information contained in observations and/or to incorporate appropriate stochasticity and uncertainty in decision-relevant predictions. This push for a increased level of parameterization has been enabled by the advent of ensemble-based methods for inverse problems, which effectively eliminate the computational constraint on free parameters in estimating the Jacobian (Chen and Oliver, 2013; White, 2018; White et al., 2021). With enough spatial coverage and diverse types of parameters, the relative influence of null space components on any one type of parameter or location in the model is minimized while the ability for various parameter types distributed across the domain to receive information from observations is enhanced.

The success of this model simplification approach, defined here as computational savings and uncertainty reduction achieved with minimal predictive consequences, depends on the type of decision relevant prediction and form of simplification. For example, Knowling et al. (2019) showed that a reduction in horizontal (parameter) resolution produced generally similar first and second moments of the flow and spatially integrated transport predictions compared to a highly parameterized model. They found that this did not hold for concentration predictions, which are known to be sensitive to local-scale properties and stresses. Similarly, White et al. (2020c) and Knowling et al. (2020) showed that reducing resolution in the vertical direction results in acceptable flow model predictions but pronounced biases in transport predictions. Moore and Doherty (2021) explored the consequences of only using steady-state calibration to reduce predictive uncertainty and found that it effectively reduced uncertainty for heads predictions via conditioning of the hydraulic conductivity field near data assimilation locations, but that artificiality of the steady-state condition and error incurred through observation averaging led to substantial predictive biases elsewhere in the model domain. That the conclusion of each of these studies is "it depends" suggests that we are far from a generalizable DA approach for balancing model complexity, uncertainty reduction, and computational efficiency across the range of decision support modeling contexts.

To date, the majority of groundwater modeling studies that perform DA for parameter estimation and uncertainty analysis assimilate data in a "batch" fashion. That is, the groundwater model is run to completion through the entire historic period, and all observations are assimilated at once, *post hoc*. For the purposes of this study, we refer to this approach as batch DA, but note that it is identical in practice to history matching and calibration. The inherent assumption and algorithmic requirement of batch DA is that the model structure and processes are correct, and the only sources of model input uncertainty or receptacles for information are the unknown properties and stresses. A major advantage of batch DA is the ability to assimilate long-term, low frequency information, since all observations are assimilated at once. However, only allowing properties and stresses to be uncertain introduces the potential for parameter compensation in batch DA through the assimilation of structural noise generated from model error.

An alternative to batch DA and potential strategy for coping with error in groundwater modeling decision support exists: sequential DA (Evensen, 1994). Sequential DA, most commonly implemented in hydrogeologic applications as Kalman filtering methods, allows for a more flexible and dynamic approach to assimilation and forecast frequency, ranging from near-real time (Hendricks Franssen and Kinzelbach, 2008; Hendricks Franssen et al., 2011) to seasonal assimilation and forecast cycles (Huang et al., 2017). A key functionality provided by sequential DA methods is the ability to estimate system state along with properties and stresses at user-defined temporal frequencies. Within a sequential DA framework, the DA algorithm takes control of the simulated states to advance the model through time, whereas in batch DA, the model itself controls the simulated states and advances through time. Implicit in state estimation is that states are uncertain, and this state uncertainty is a more flexible and appropriate receptacle

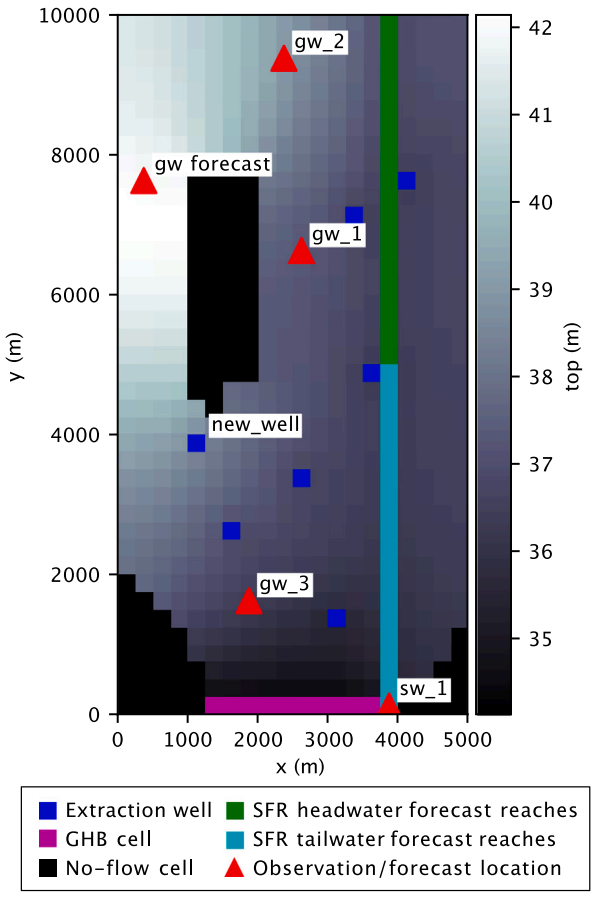


Fig. 1. The lower-resolution "simple" model domain discretized with 40 rows and 20 columns.
Source: Modified from White et al. (2020b).

for structural noise contained in observations (i.e., simulated water levels are a more appropriate receptacle for missing signal components in water level observations than hydraulic conductivity or porosity). Sequential DA methods can therefore reduce the burden of structural noise on static properties by assimilating this noise into the states through the sequential DA process (Clark and Vrugt, 2006b; Kim et al., 2021). Put bluntly, batch DA may never be completely free of predictive bias due to being tied to an imperfect model structure, regardless of the level of complexity. Sequential DA offers a potentially powerful strategy for reducing bias in decision-relevant predictions in the presence of model error.

In this work, we evaluate the outcomes of using batch and sequential DA approaches in a hypothetical case where a simple model structure is in error with respect to the more complex, "truth" model from which the observations available for assimilation are generated. This evaluation of batch and sequential DA approaches in the presence of model error is an important and novel contribution to the practice of groundwater decision support modeling. Additionally, this work advances the powerful paired complex–simple analysis of Doherty and Christensen (2011) into an empirical paired complex–simple analysis to enable the incorporation of stochastic uncertainty and a statistical evaluation of predictive bias. The model error in this empirical paired complex–simple analysis is explicitly represented by two scenarios, one where the simple model is of coarser spatial (both vertical and horizontal) and temporal resolution compared to the complex model. The other scenario builds on the reduced resolution and includes the additive effect of model process error by treating historic groundwater pumping as a fixed, "known" quantity rather than an uncertain one. Performance

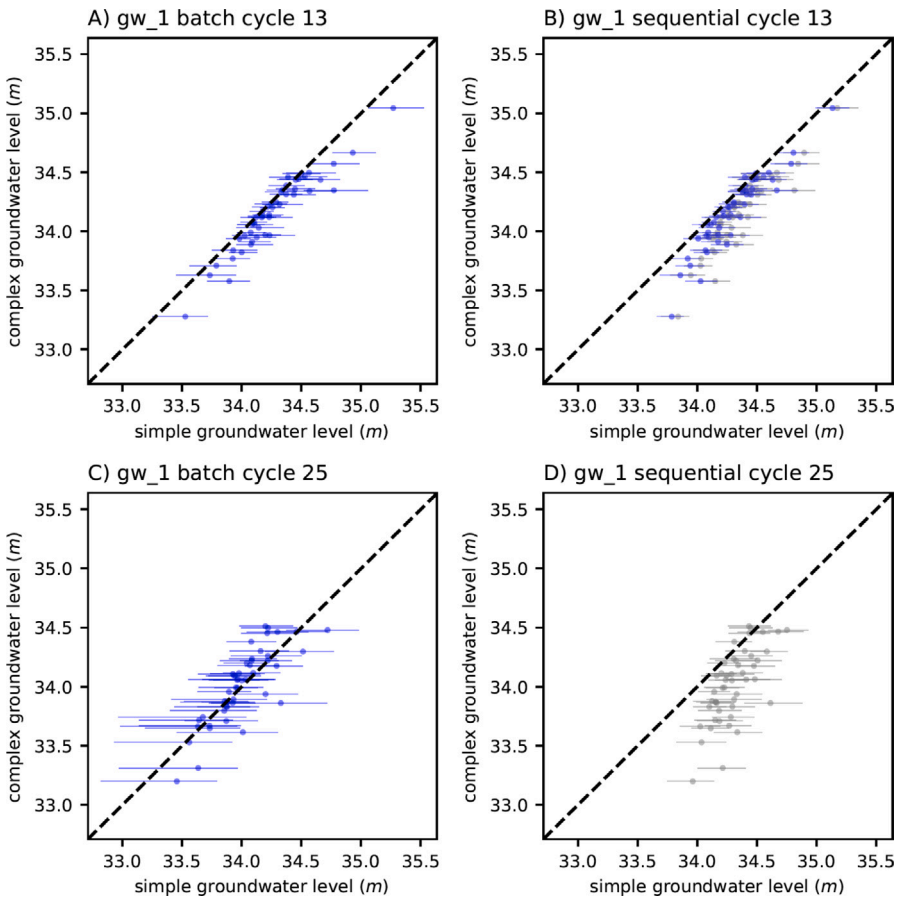


Fig. 2. Posterior (blue) and open-loop (gray) simple model outputs for the GW_1 observation location using batch (left panel) and sequential (right panel) DA for assimilation cycle 13 (top panel) and open-loop cycle 25 (bottom panel).

of the two DA approaches are evaluated through a comparison of simple models' ability to history match and forecast decision-relevant predictions. We hypothesize that the joint parameter-state estimation ability of sequential DA is able to better cope with model error by subsuming structural noise that would otherwise be absorbed by static properties in batch DA.

2. Methods

We evaluate the sequential and batch DA strategies using PESTPP-DA (Alzraiee et al., 2021), a newly developed generic DA tool to join the suite of modeling decision support and uncertainty analysis programs in PEST++ (White et al., 2021). Full details on the theory, implementation, and capabilities of PESTPP-DA can be found in Alzraiee et al. (2021), and we provide a brief overview here. PESTPP-DA can assimilate data in both batch and sequential modes, where for sequential DA, the model simulation is divided into user-defined "cycles" or a period of time for which applied forcings (i.e., boundary conditions) and parameters are constant and/or discrete observations are available for assimilation. The MODFLOW stress period concept, which naturally lends itself to the cycle concept in PESTPP-DA, allows for a high level of flexibility for designing the assimilation cycle length and frequency. PESTPP-DA can be run in three modes: state estimation, parameter estimation, and, the focus of this study, joint state-parameter estimation.

In joint state-parameter estimation mode, model states (e.g., heads) are considered uncertain along with model properties and forcings (e.g., hydrogeologic properties and boundary condition fluxes); each cycle has an estimated initial state and a simulated final state, where the final state becomes the initial state for the subsequent cycle. Importantly, during each cycle, the model is run using the estimated initial state prior to data assimilation, and the simulated output of that initial run is referred to as a "one-step ahead" (one-step ahead) forecast. This forecast is significant because it reveals the near-term predictive power of sequential DA, where the posterior state from the previous assimilation cycle combined with model physics through a forward run, can provide a simulated quantity that closely predicts observed quantities prior to any data assimilation.

As with other programs of the PEST++ suite, PESTPP-DA follows a Bayesian framework, where a prior distribution is defined for all uncertain parameters based on field measurements, expert knowledge, and the literature. Within the solution scheme of PESTPP-DA, the likelihood is proportional to the inverse of a weighted $\mathcal{L}_2$ norm of innovations (e.g., the objective function), which represents the degree of agreement between observations and their simulated equivalents for each assimilation cycle. PESTPP-DA updates all parameters (and states, if doing joint state-parameter estimation) with the option of using a standard Kalman update, an iterative Kalman update known as multiple data assimilation (MDA, Emerick and Reynolds, 2013), or an iterative ensemble smoother based on the Gauss–Levenberg–Marquardt scheme (Chen and Oliver, 2013; White, 2018). The posterior ensembles of parameters and predictions represent uncertainty in these model components after DA. See Alzraiee et al. (2021) for a complete discussion of the PESTPP-DA solution schemes. Importantly, both (iterative) solution schemes involve the use of ensembles to approximate the cross-covariance between parameters (and states) and simulated equivalents to observations. Worth adding, however, is that the sequential nature of PESTPP-DA (if running in sequential mode) means that the iterative solution schemes are applied for each user-defined assimilation cycle,

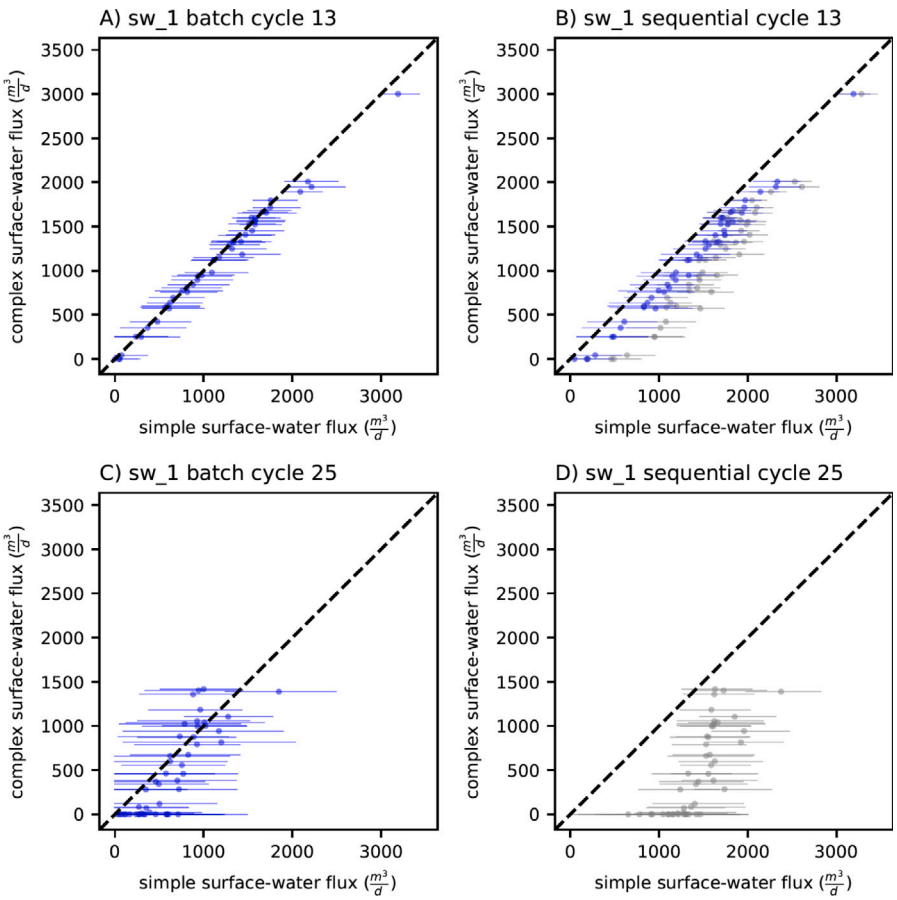


Fig. 3. Posterior (blue) and open-loop (gray) simple model outputs for the SW_1 streamflow observation location using batch (left panel) and sequential (right panel) DA for assimilation cycle 13 (top panel) and open-loop cycle 25 (bottom panel).

increasing the computational cost compared to batch assimilation using iterative ensemble methods, though still more efficient than the traditional finite-difference Jacobian of Gauss–Levenberg–Marquardt (GLM) algorithms (Doherty, 2003; White et al., 2020b) for high-dimensional inverse problems.

3. Model description

We perform the paired complex–simple model analysis using a version of the Freyberg (Freyberg, 1988; Hunt et al., 2020) model originally published in White et al. (2020b). The model, which is implemented in MODFLOW-6 (Langevin et al., 2017; Hughes et al., 2017), is a transient, multi-layer model that includes groundwater and surface-water processes. Model error is explicitly represented as resolution error by assimilating output from a finer spatial resolution, daily stress period version of the Freyberg model into a coarser resolution, monthly stress period version of the Freyberg model, following White et al. (2021). This is not unlike the situation of all groundwater models, which are invariably a simplification of the resolution and structure of real-world processes that generate our observations.

The lower-resolution, "simple" model has 40 rows, 20 columns, 1 layer, and 25 stress periods. The first stress period is 10,000 days in length and represents a pre-development condition—there is not groundwater extraction during the first stress period. The subsequent 24 stress periods are discretized into monthly periods and include time-varying boundary conditions representing cyclical recharge and pumping over the two-year period. Observations for assimilation are available during the first year (e.g., stress periods 2–13) and include two deep and one shallow groundwater level observation locations and one surface water gage location (Fig. 1). The last 12 stress periods represent an unmeasured, forecast period. Groundwater extraction occurs in six wells scattered throughout the domain for the entire simulation period, and a seventh extraction well ("new well" in Fig. 1) is added in stress period 11 to impart a nonstationary trend in the forecast period.

The higher-resolution, "complex" model mimics the lower-resolution model in properties and boundary conditions, differing only in spatial and temporal resolution. It has 120 rows, 60 columns, and 3 layers—three times as many of each as the lower-resolution model—and, with the exception of the same initial 10,000 day pre-development stress period as in the simple model, the complex model simulates daily stress periods over the two-year simulation length for a total of 731 stress periods. The location of groundwater extraction wells, SFR reaches and state observations were made to correspond between the higher- and lower-resolution models.

Four simulated quantities from the complex truth model were assimilated during the first 12 months of the 2-year simulation, yielding 48 historic state observations for conditioning:

1. One shallow (i.e., layer 1 in the complex model) groundwater location (GW_1 in Fig. 1).
2. Two deep (i.e., layer 3 in the complex model) groundwater locations (GW_2 and GW_3 in Fig. 1).
3. One surface water location at the stream outlet (SW_1 in Fig. 1).

Three simulated quantities of interest were defined, shown in Fig. 1:

1. An unmeasured groundwater level location in the upgradient hydrologic divide area of the model domain (GW_forecast in Fig. 1);

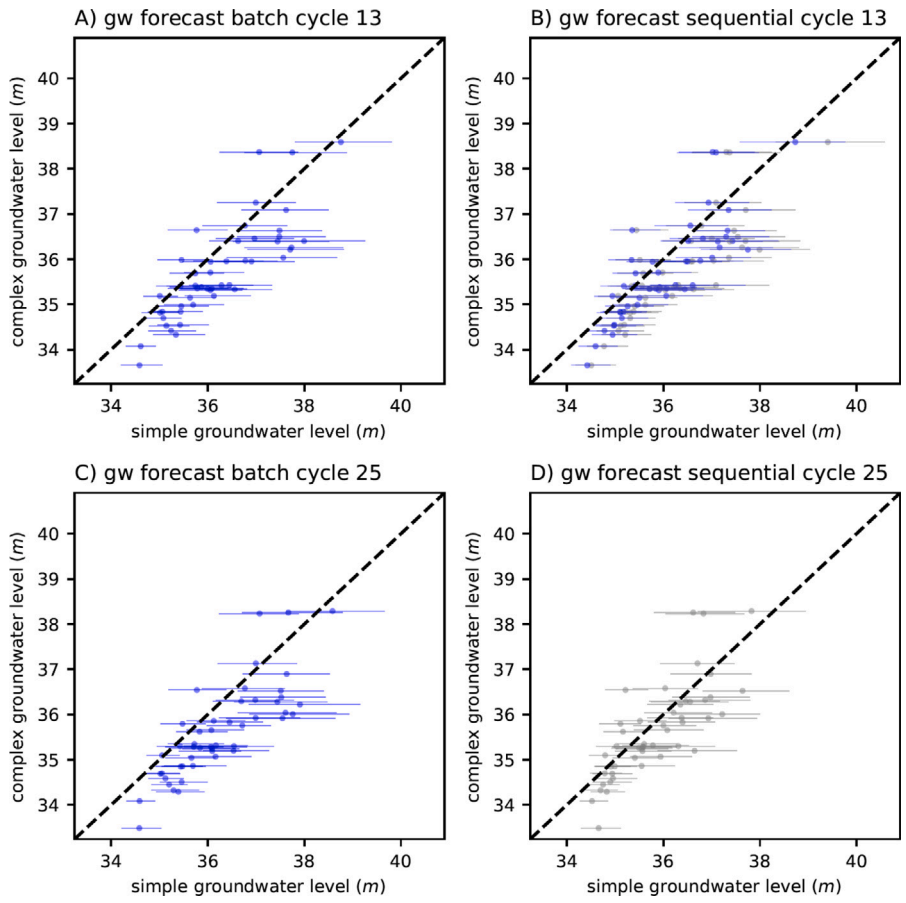


Fig. 4. Posterior (blue) and open-loop (gray) simple model outputs for the GW_forecast location using batch (left panel) and sequential (right panel) DA for assimilation cycle 13 (top panel) and open-loop cycle 25 (bottom panel).

2. Aggregated surface-water/groundwater exchange for SFR headwater reaches in rows 1–20 (green cells in Fig. 1) (reaches 1–60 in the higher-resolution model);
3. Aggregated surface-water/groundwater exchange for SFR tailwater reaches in rows 21–40 (teal cells in Fig. 1) (reaches 61–120 in the higher resolution model).

Simulated outputs of these quantities were selected to represent common quantities of interest for decision making in applied groundwater modeling. These quantities also represent a range of dependence on the historic state observations used for assimilation, or, put another way, these quantities have varying null-space dependence (Moore and Doherty, 2005). This variability will in test different aspects of the sequential and batch DA approaches to cope with model error and resulting parameter compensation and potential predictive bias.

Two parameterization scenarios were developed to evaluate the predictive power of sequential versus batch DA. The first parameterization scenario assumes all static properties as well as all boundary conditions are uncertain. This scenario, herein referred to as the "coarse scenario" makes use of the greatest number and variation of possible receptacles to receive information during the DA process, and allows for the isolation of model resolution error in comparing the two DA approaches. We note the complex model, used to generate replicates of observations in the paired-model analysis, also treated all of these quantities as uncertain, representing applied groundwater modeling settings.

The second parameterization scenario, inspired by a common approach among practitioners, assumes that all static properties and boundary conditions are uncertain, except for historic groundwater extraction. This scenario, herein referred to as the "fixed pumping scenario" allows for evaluating the effect of model process error along with resolution error, as it collapses the uncertainty of groundwater extraction estimates to zero, potentially transferring the uncertainty in historic groundwater extraction to the other uncertain model inputs. In both scenarios, static properties, such as horizontal and vertical hydraulic conductivity and confined and unconfined storage were parameterized with a multiscale approach (e.g., McKenna et al. (2019) and White et al. (2020a)) using both pilot point (e.g.,Doherty (2003) and grid-scale parameters. The recharge and groundwater extraction boundary conditions were parameterized to express both spatial and temporal uncertainty in these model inputs. Property grid-based, pilot-point, and recharge grid-based parameters were assigned distance-based prior correlations with an exponential variogram with range 500, 2000, and 1000 m, respectively. The sill was set proportional to the variance of the parameter type. The prior parameter ensemble used in both the sequential and batch DA analyses was generated in pyEMU (White et al., 2016, 2021) using spectral simulation, see White et al. (2020b) for more details. Importantly, both the sequential and batch DA approaches used the same prior parameter ensemble.

To undertake sequential DA with the above-described Freyberg model, we constructed a single-stress-period transient MODFLOW-6 model and treated the initial groundwater level for each active model cell for each assimilation cycle as dynamic states that are estimated along with the other uncertain parameters. PESTPP-DA was used to advance the sequential assimilation process across the 25 cycles—one cycle for each stress period in the batch DA approach of the simple model. Note that only cycles 2–13 have four state observations each to assimilate, while cycles 14–25 are "open-loop" cycles, where the prior ensemble is evaluated forward in time.

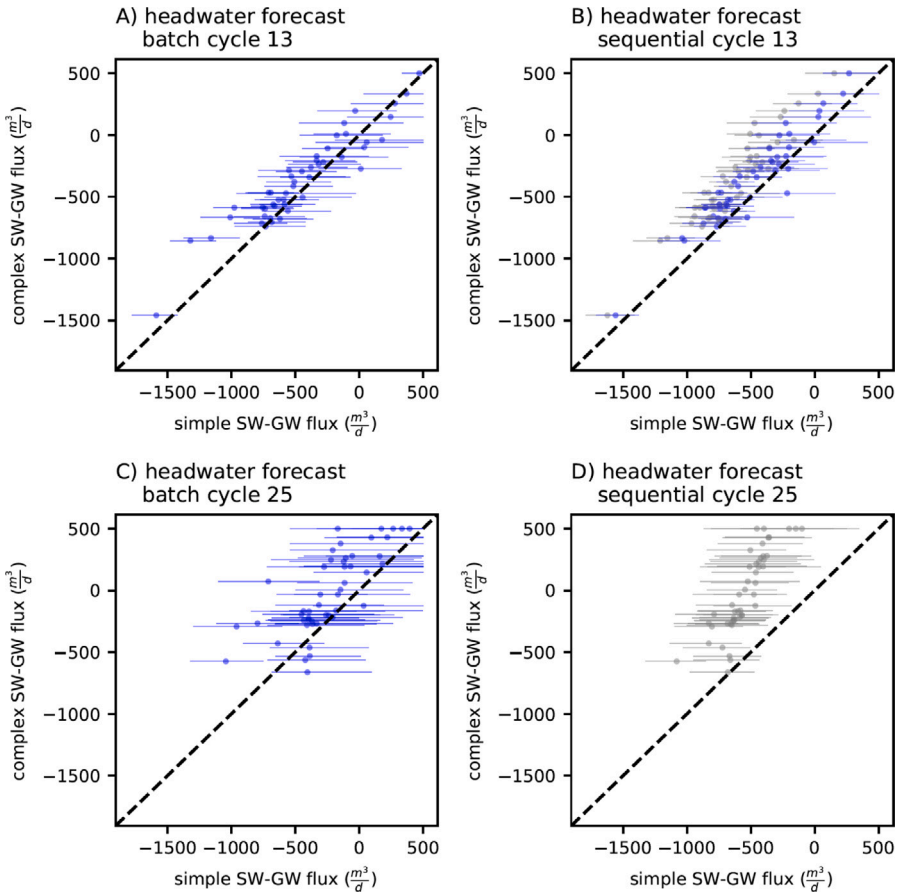


Fig. 5. Posterior (blue) and open-loop (gray) simple model outputs for the headwater SW-GW exchange flux forecast locations using batch (left panel) and sequential (right panel) DA for assimilation cycle 13 (top panel) and open-loop cycle 25 (bottom panel).

4. Paired complex–simple workflow

We extend the deterministic paired complex–simple analysis of Doherty and Christensen (2011) to statistically evaluate the performance of batch and sequential DA approaches to cope with model error. Briefly, the paired-complex–simple analysis begins by completing a Monte Carlo with the complex model using realizations drawn from the complex model prior ensemble. Each set of the realized simulation outputs from this Monte Carlo are then treated as observations for history matching in the simple model. Herein, we refer to these realized complex model simulation results as "replicates" and we generate 50 replicates for this analysis.

We extend the paired complex–simple analysis to include ensemble-based simple model results so that a more complete stochastic understanding of the simple model's behavior can be analyzed in both a prior and posterior stance. That is, for each of the 50 replicates, an ensemble of 50 simple model realizations are propagated through the batch and sequential DA algorithms, yielding a corresponding prior and posterior probability density function (PDF) for each simple model output, compared to a single value of simple model output in the original paired complex–simple analysis. By attempting to reproduce each replicate's simulated outputs at locations and times of state observations, and also having the corresponding replicate (e.g., correct) value for the quantities of interest, the empirical paired complex–simple analysis allows rigorous statistical insights into how the simple model is able to cope with model error in the context of simulating the quantities of interest. The novelty of this extended approach is the inclusion of the stochastic effect of model input uncertainty on predictions, enabling a probabilistic rather than deterministic evaluation of bias at only a modest increase in computational cost.

We present the majority of paired complex–simple analysis results in the form of "s versus $\underline{s}$" plots, where the 50 complex model replicate simulated quantities correspond to a single value (s) each on the y-axis and the simple model simulated quantities ($\underline{s}$) correspond to 50 realized (e.g., ensemble-based) x-values per complex replicate y-value. As discussed in detail in Doherty and Christensen (2011), s versus $\underline{s}$ plots (herein referred to as "s-plots") are useful model error diagnostic plots, where a bias in simple model predictions can be clearly detected as deviations from a 1:1 line. We plot the simple model simulated values as a point representing the mean and a line extending from the 5% to the 95% confidence intervals of the values. To further aid in interpretation, we show the results for cycle 13 and 25, both corresponding to the management relevant low-flow period and encompassing an assimilation cycle and open-loop cycle, respectively.

5. Results

5.1. Coarse scenario

The coarse scenario involves a structurally simple model with a high dimensional parameterization (>3500 adjustable parameters), where the source of model error arises from a coarser horizontal and vertical resolution compared to the complex model.

S-plots for the posterior assimilated GW_1 head observation show no appreciable difference in performance between batch and sequential DA in the assimilation cycle (Fig. 2A, B). The sequential DA one-step ahead forecasts for cycle 13, shown in gray in Fig. 2B, are almost indistinguishable from the posterior values, highlighting the capability of sequential DA for near-term prediction. In the open-loop cycle, the batch DA simple model first moments cluster along the 1:1 line

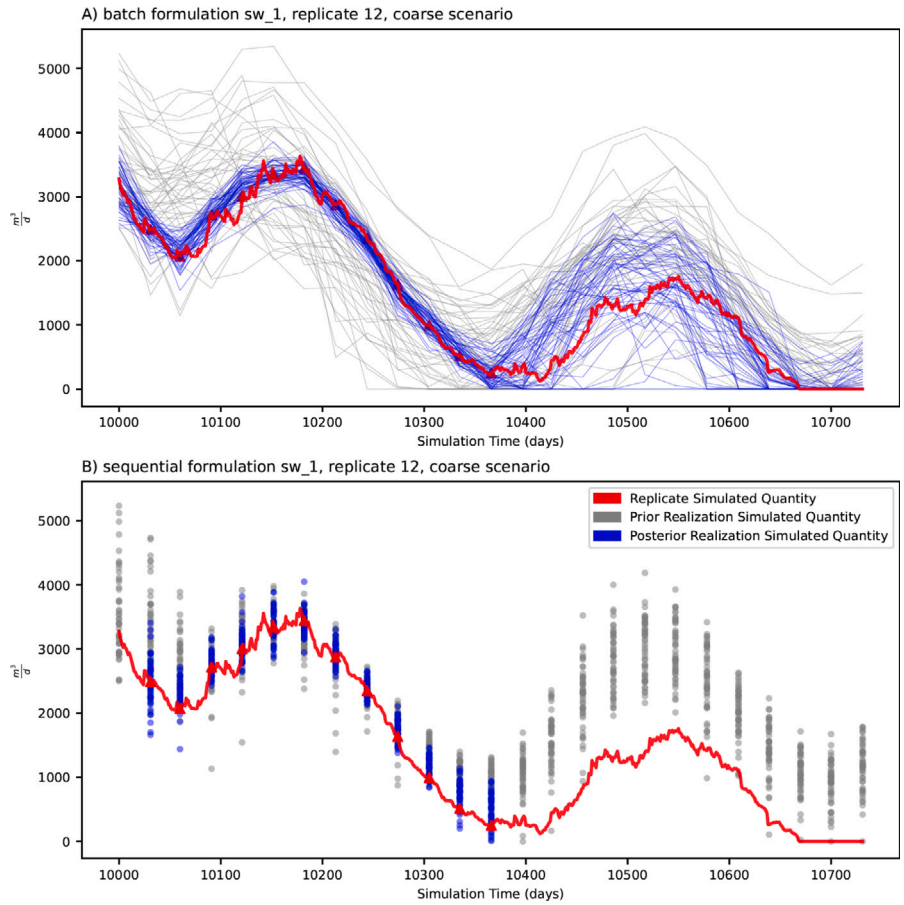


Fig. 6. Posterior (blue) and open-loop (gray) simple model outputs and complex model replicate outputs (red) for the SW_1 observation location using batch (top panel) and sequential (bottom panel) approaches for the coarse scenario.

(Fig. 2C) while the sequential DA first moments show a trend with higher simple model head predictions at lower complex replicate values (Fig. 2D). This suggests that the sequential algorithm struggled to assimilate observations from simple model replicates with more pronounced declining trends, a symptom that becomes exacerbated over time in the open-loop cycles.

The two DA approaches perform similarly in their ability to reproduce replicate surface water (SW_1) flux predictions during the assimilation period, though in cycle 13 the batch DA simple model first moments cluster along the 1:1 line and the sequential DA posterior first moments are slightly shifted to the right (Fig. 3A, B). The one-step ahead forecast predicts the general slope but overpredicts the magnitude of posterior first moments. As with the groundwater assimilation location (Fig. 2D), this overprediction is even more pronounced in open-loop cycle 25 (Fig. 3D). The introduction of a new pumping well in cycle 11 produces a declining trend that slightly degrades the one-step ahead predictive ability, which is ameliorated by history matching and allowing for model head states to be adjusted (Fig. 3B). As the declining trend continues throughout the forecast period, so does the predictive performance of sequential DA (Fig. 3D). With the batch DA approach, the simple model first moments in the forecast period plot along a moderately steeper slope than the 1:1 line (Fig. 3C), however the complex replicate value is almost always contained within the simple model ensemble values.

Predictions for the forecast locations shed light on whether the assimilation process led to erroneous parameter conditioning in space during the assimilation period and space and time during the forecast period. For the groundwater forecast location, batch and sequential DA show similar performance during both the assimilation and open-loop cycles (Fig. 4). This result is surprising given the biased open-loop predictions made with the sequential approach at the GW_1 assimilation location (Fig. 2D). For the headwater SW-GW exchange flux forecast location, the simple model first moments of both DA approaches closely predict the complex replicate value during the assimilation period, clustering along the 1:1 line (Fig. 5A, B). Similar to the GW_1 and SW_1 assimilation locations, however, the sequential approach systematically overpredicts the flux from GW to SW compared to the complex replicate values in the forecast cycle (Fig. 5D, note that more negative indicates a higher flux from GW to SW).

Plots of observed versus simulated quantities can shed further light on the history-matching and forecast performance of the two DA approaches in the coarse scenario. We feature the outputs from one replicate (Fig. 6) and refer readers to the supporting information for the observed versus simulated plots for additional replicates and all assimilation and forecast locations. Fig. 6 reinforces the similar performance of batch and sequential DA during the history matching period, where the posterior ensemble (blue lines and dots) encompass the complex replicate streamflow at SW_1. In the forecast period (second half of the simulation), the batch DA posterior ensemble encompasses the truth model output, while the sequential DA prior ensemble generally overpredicts the complex replicate flows compared to batch DA.

Overall, the coarse scenario results reveal minimal difference in the ability of batch and sequential DA to reproduce historical observations and modest differences in long-term predictive outcomes, particularly with surface water and SW-GW exchange fluxes, in the presence of model resolution error. With respect to short-term predictive performance, the sequential DA one-step ahead forecasts closely predict the posterior simulated head quantities at both assimilation and forecast locations, with slightly worse predictive performance for streamflow and SW-GW exchange quantities.

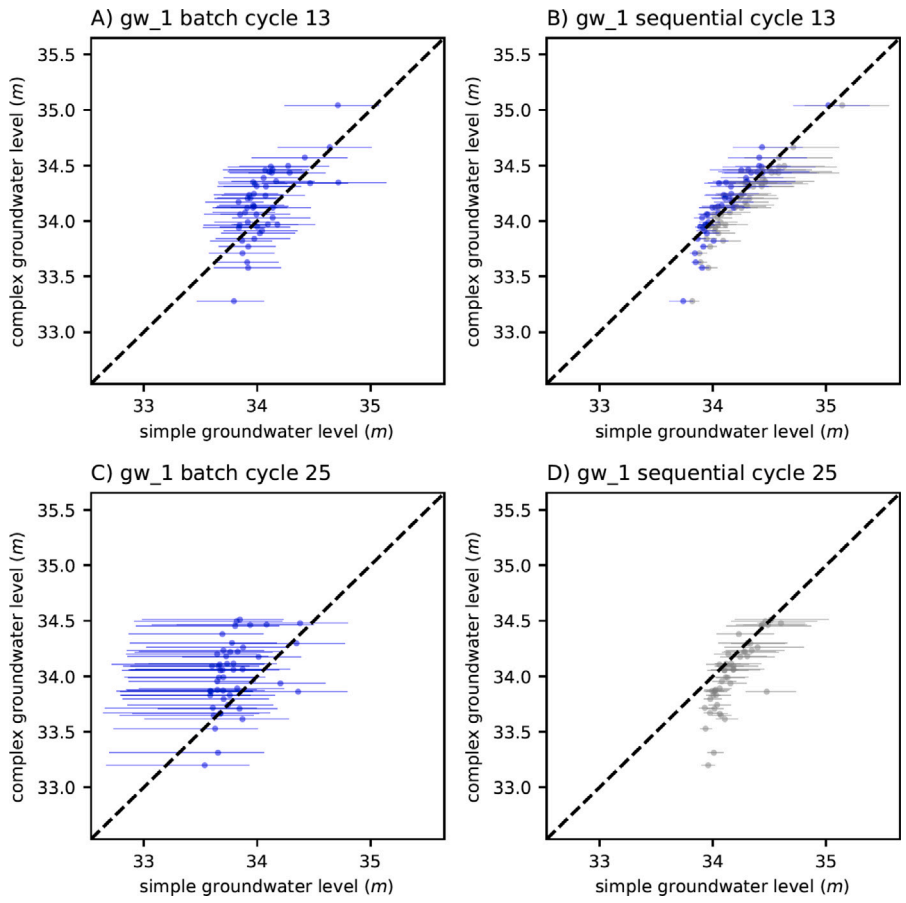


Fig. 7. Posterior simple model outputs for the GW_1 observation locations using batch (left panel) and sequential (right panel) DA for assimilation cycle 13 (top panel) and open-loop cycle 25 (bottom panel).

5.2. Fixed pumping scenario

In the fixed pumping scenario, model error arises from both a coarser resolution in the simple compared to the complex model as well as misspecified pumping rates, as the complex prior Monte Carlo varies pumping rates for each replicate and the simple model treats the pumping as a fixed quantity. This scenario was designed to reflect a common practice in applied groundwater modeling where pumping is treated as "known" and not adjustable during data assimilation, introducing model process error.

The two DA approaches diverge in their ability to reproduce historical observations in the fixed well scenario. For the shallow groundwater observation location, a noticeable deviation in slope from the 1:1 line can be seen for batch DA (Fig. 7A). Here, the simple model batch DA results overestimate lower heads from the complex replicates and underestimate higher heads from the complex replicates, reflecting an inability of batch DA to reproduce the range in complex replicate head observations. For this same observation location, sequential DA shows a similar bias in the lower replicate head values, but is otherwise able to reproduce the majority of replicate head values without bias (Fig. 7B). A similar, albeit more pronounced pattern can be seen in the open-loop cycle (Fig. 7C, D).

Notably, the batch DA simple model results show much larger variance than the sequential DA results and variance in the sequential DA simple model results follows a decreasing trend with decreasing complex replicate value (Fig. 7C, D). The larger variance in the batch results occurs because the complex replicate values are in conflict with (i.e., outside the range of) the simple model prior ensemble predictions, encouraging the batch algorithm to explore more extreme parameter values and combinations to match the observations. The sequential approach, on the other hand, is able to overcome the prior-data conflict through updating the head states, but is potentially over-conditioning the parameters that influence simulated head at GW_1 leading to smaller variances in both the assimilation (Fig. 7B) and open-loop cycles (Fig. 7D). This over-conditioning is not consistent across the replicates, but increases with decreasing complex replicate value. Since the lower replicate GW_1 head values are likely driven by higher pumping rates in the complex model, these replicate values may reflect a higher degree of model process error between the complex and simple models and, as a consequence, a higher degree of over-conditioning by the sequential algorithm.

The bias in reproducing historical observations amidst model resolution error and fixed well pumping is more pronounced for the streamflow assimilation location (Fig. 8). The posterior first moments of the simple model realizations differ strongly between the two DA approaches, with batch DA plotting along a steeper regression line, driven by an inability to reproduce the upper and lower range of complex replicate streamflow values (Fig. 8A). Meanwhile, sequential DA shows an impressive ability to reproduce the complex replicate streamflow quantity (Fig. 8B), with first moments falling along the 1:1 line. Here, uncertain head states in the sequential DA approach subsume the noise introduced by variable pumping rates in the complex model replicates relative to the fixed pumping amounts in the simple model realizations, minimizing bias that would affect the head-dependent streamflow fluxes. Similar to the coarse scenario, the sequential approach completely overpredicts streamflow in the open-loop cycle (Fig. 8D) by not assimilating the long-term trend from introducing a new well in cycle 11. The batch DA results, despite showing strong bias in the slope of the simple model first moments, are better able to reproduce the long-term trend evidenced by the complex replicate value falling within the wide 95% confidence intervals of the simple model ensembles (Fig. 8D).

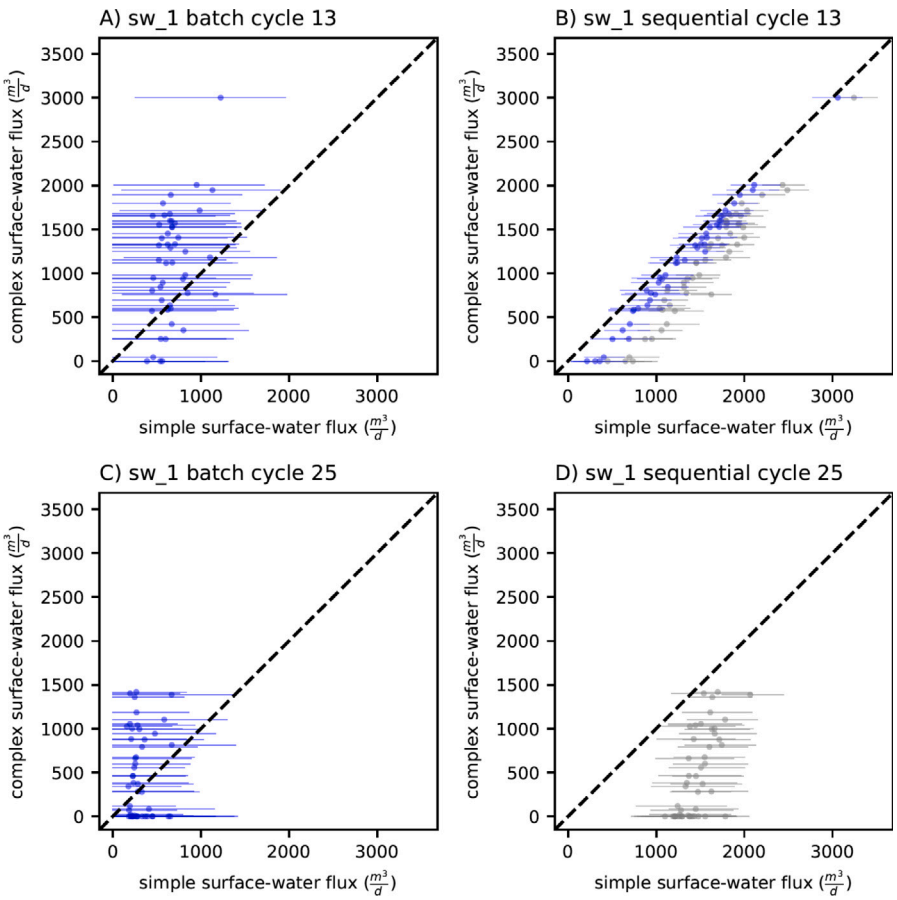


Fig. 8. Posterior (blue) and open-loop simple model outputs for the SW_1 streamflow observation location using batch (left panel) and sequential (right panel) DA for assimilation cycle 13 (top panel) and open-loop cycle 25 (bottom panel).

Despite the fixed well scenario containing process error along with model resolution error, the two approaches again perform equally well for the GW_forecast location in both the assimilation and open-loop cycles (Fig. 9). One explanation for this is an implicit localizing of parameter compensation through a high degree of spatial parameter resolution such as with grid- or pilot point-based parameterization. Here, the use of grid-scale and pilot point adjustable parameters in the simple models means that parameter compensation from assimilating observations that contain structural noise in one model location does not necessarily permeate the entire domain. In other words, data assimilation with observations that contained structural noise arising from missing parameters (i.e., variable pumping in the complex model) at GW_1, GW_2, and GW_3 did not negatively impact predictions at GW_forecast because no data was assimilated in that region of the model (Fig. 1). An alternative yet not mutually exclusive explanation is that the groundwater forecast location is not impacted by the variable pumping in the complex replicates owing to its distant location from the extraction wells, making the replicate GW_forecast simulated value less marred by the model process error.

Batch DA shows substantial predictive bias at the SW-GW exchange forecast locations (Fig. 10A, C). That is, the complex replicate simulated value varies widely, but the simple model realizations show only small changes in the ensemble first moments, leading to steeper slopes in the s-plots. This is not surprising given that batch DA struggled to reproduce historical head (Fig. 7) and SW flux 8 observations in the fixed well scenario as the SW-GW exchanges are head-dependent fluxes. Sequential DA, on the other hand, matches the 1:1 slope with both the one-step ahead and posterior simulated quantities during the assimilation period, despite no SW-GW flux observations being assimilated (Fig. 10B). This enhanced forecast performance begins to break down in the first open loop cycle, and remains that way for the remainder of the simulation (Fig. 10D). This is also not surprising given that the sequential DA approach updates head predictions with each assimilation cycle. Since the SW-GW exchange flux is a head-dependent process, close reproduction of heads in the assimilation locations results in close reproduction of the SW-GW exchange fluxes, but only during the assimilation cycles when head states are being estimated. In essence, the head assimilation keeps the sequential DA process "on track" in the presence of otherwise corrupting model error.

The observed versus simulated timeseries results provide an alternate vantage point of the relative performance of batch and sequential DA approaches in the assimilation and forecast period. Using the SW_1 streamflow assimilation location and replicate 12 again as an example (see supporting information for additional observed versus simulated replicate plots), sequential DA outperforms batch DA during the assimilation period (Fig. 11B). The batch DA posterior realizations overpredict the low flows and underpredict the high flows during the assimilation period, reflecting the limitation of fixed pumping rates in the simple model. For the forecast period of the simulation, however, sequential DA completely overpredicts streamflow compared to the replicate value, missing the long-term declining trend present in the assimilation period. The batch DA approach, despite the biases in reproducing historical observations, is better able to reproduce the long-term declining trend with the posterior simple model ensemble predictions encompassing the complex replicate value (Fig. 11A). This trade-off in performance between short-term history matching of sequential DA and long-term forecasting of batch DA can be seen in other replicates with strong nonstationary trends (e.g., Figures S158, S165, and S172 in the supporting information).

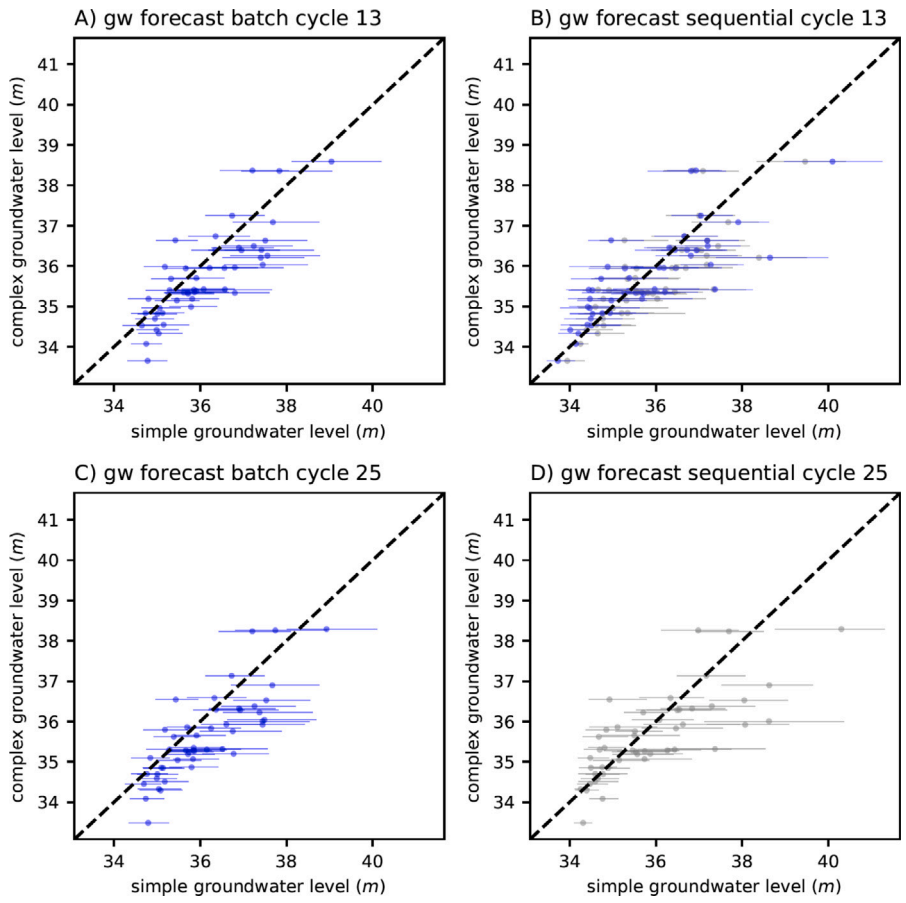


Fig. 9. Posterior (blue) and open-loop (gray) simple model outputs for the GW_forecast location using batch (left panel) and sequential (right panel) DA for assimilation cycle 13 (top panel) and open-loop cycle 25 (bottom panel).

6. Discussion

Our results demonstrate that both DA approaches can successfully reproduce historical flow model quantities in the presence of model resolution error when a high dimensional parameterization of uncertain model properties and stresses is considered. This is the case despite employing two distinct DA algorithms: batch DA, where all observations are assimilated at once and the only source of uncertainty is in the stresses and properties and sequential DA, where a subset of the observations are assimilated during discrete assimilation cycles and model states along with properties and stresses are considered uncertain and thus adjustable. This does not suggest that sequential DA does not have greater flexibility to cope with model error, but rather that the additional flexibility of state estimation may not be needed when the model error is limited to resolution error and all model properties and stresses are treated as uncertain and adjustable. This is evidenced by the infrequent occurrence of prior-data conflict in the batch DA realizations, where a high parameter dimensionality and sufficiently wide uncertainty bounds of adjustable parameters provided enough bandwidth to encompass the range of outputs from each complex replicate.

The fixed well scenario revealed important limitations to reproducing historical observations when some uncertain model inputs are treated as fixed quantities during the inversion. This amounts to model resolution error combined with process error, where the complex replicates varied pumping and the simple realizations did not. The effect is that the complex replicate simulated quantities contain information related to variable pumping rates that can only be assimilated into the "wrong" parameters in the simple models (Oliver and Alfonzo, 2018). Here, the state estimation capability of sequential DA afforded much better performance over batch DA in reproducing historical observations (see plot B in Figs. 7, 8, and 10). The lack of adjustable pumping and head states in batch DA resulted in numerous occurrences of the complex replicate values falling outside of the simple model ensemble of simulated equivalents, particularly at the upper and lower tails of the replicate values.

It is clear that for "operational" modeling settings, where a groundwater model is used to make one-step-ahead forecasts, the sequential DA approach is ideal in that it is tuned to make high-frequency forecasts where the initial states dominate predictive accuracy (see gray points in plot B of Figs. 7, 8, and 10). In this setting, the sequential DA approach affords the ability to assimilate recent observations into both states and parameters simultaneously, towards making an optimal near-term (potentially one-cycle-ahead) forecast for future system conditions. However, the state estimation capability of sequential DA does not necessarily translate to better long-term forecast performance. As demonstrated with the SW-GW exchange forecasts, sequential DA resulted in simple model predictions falling close to the 1:1 line during the assimilation cycle (Fig. 10B), but showed near vertical slopes in the open-loop cycles (Fig. 10D). The promising performance in short term, despite no SW-GW exchange fluxes being assimilated, occurs as a side effect of the state estimation, where adjusting the head states to match the complex replicate values resulted in improved head-dependent fluxes. This benefit is afforded during active assimilation cycles and degrades quickly after the first open-loop cycle.

Poor performance in long-term forecasting using sequential DA is the result of assimilating observations from one monthly cycle at a time. The adjustable head states in sequential DA treat the time-varying observations as independent between assimilation cycles, missing the variable amplitude trends present in the complex model outputs. Unsurprisingly, this forecast limitation is exacerbated in the fixed well

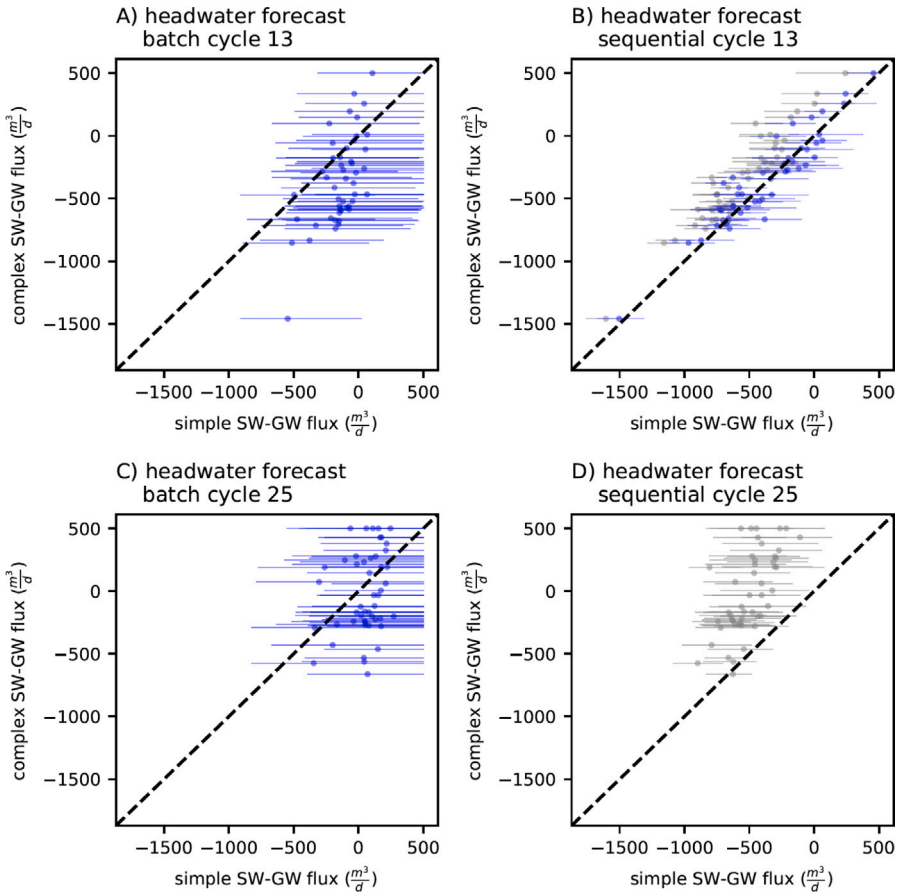


Fig. 10. Posterior (blue) and open-loop (gray) simple model outputs for the headwater SW-GW exchange flux forecast locations using batch (left panel) and sequential (right panel) DA for assimilation cycle 13 (top panel) and open-loop cycle 25 (bottom panel).

scenario. Since the pumping rates are varied in the complex model replicates, the resulting water level, streamflow, and SW-GW exchange quantities have varying amplitudes and trends. The result is that sequential DA often substantially overpredicts heads, streamflows, and SW-GW exchange fluxes in the second year (open-loop period) of the simulation in the simple model realizations (Fig. 11B). Batch DA, despite giving poor performance in reproducing historical observations in the fixed well scenario, tended to better capture the long-term trends (Fig. 11A). This long-term predictive limitation of sequential DA is less pronounced in the coarse scenario where all properties and stresses are considered uncertain (Fig. 6B). In that scenario, the sequential approach is better able to assimilate the long-term trend induced by pumping in the complex model replicate into the adjustable pumping rates in the simple model, though the simple model ensembles still overpredict the complex replicate value in the forecast period. Put simply, when the sequential algorithm is able to adjust the pumping rate parameter to reproduce history, it is better able to impart a physically based long-term trend in the forecast period. Thus, the limitation of only assimilating a subset of observations is mitigated by the parameter conditioning process when all parameters are adjustable.

7. Implications

These results have important implications for groundwater decision support modeling, given the ever present trade-off between model complexity (i.e., computationally expensive) and model error (i.e., computationally cheap) (Hunt and Zheng, 1999; Simmons and Hunt, 2012; Hill, 2006; Haitjema, 2015). Too often, discussions around groundwater model complexity are derailed by parameterization hesitancy, reducing the specter of complexity to number of free parameters as the hallmark of an overly complex model. Our results suggest the opposite to be true. Even in a relatively high dimensional free parameter example, the effect of removing parameters representing uncertainty in pumping from the inverse analysis noticeably degraded history matching and forecast performance. The adjustable head states of the sequential DA approach helped to cope with this model error during the history matching and short-term forecast period, but our results show that representing all sources of model input uncertainty, in space and time, using a batch DA approach yields superior history matching and forecast performance in both the short- and long-term. If constructing a simple decision support model for a groundwater system using a high dimensional parameterization and batch DA approach results in an unacceptable amount of model error (i.e., numerous instances of prior-data conflict), sequential DA may prove a useful tool to cope with that error in history matching and near-term prediction. Future work to this end could explore adjusting prior uncertainty bounds on the head states as a form of regularization to capture structural noise while extracting maximum information from the set of observations into model properties and stresses.

Finally, while our results point to the potential pitfalls of assimilating a subset of data in predicting nonstationary quantities, they also show strong promise for PESTPP-DA to enable one-step ahead forecasting in a decision support context. The two DA approaches in this study represent two ends of an assimilation window spectrum, the two-year period of batch DA at the wide end and the monthly stress period assimilation cycles of sequential DA at the narrow end. If the decision support objective involves forecasting near- to medium-term quantities in the presence of nonstationarity, using a wider assimilation window (e.g., seasonal) with sequential DA may provide an optimal compromise for enabling one-step ahead prediction while also assimilating low frequency (e.g., trend) information in the set of observations.

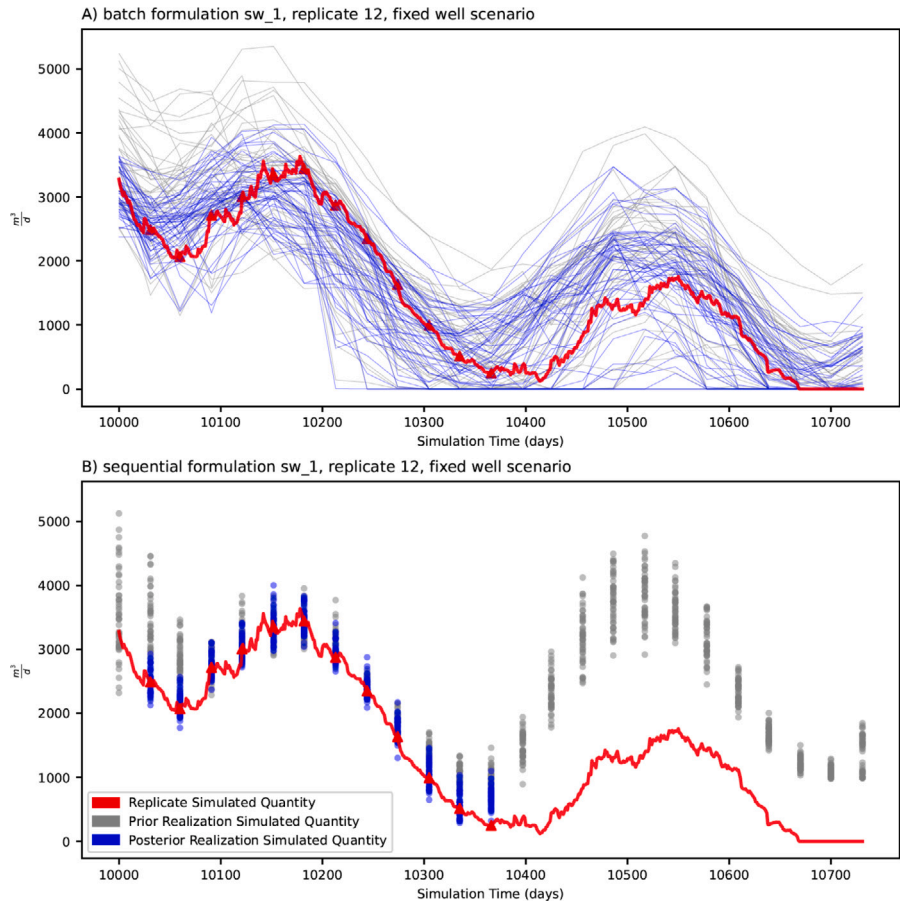


Fig. 11. Posterior (blue) and open-loop (gray) simple model outputs and complex model replicate outputs (red) for the SW_1 observation location using batch (top panel) and sequential (bottom panel) approaches for the fixed well scenario.

Future work should explore wider assimilation cycles as a strategy for improving near-term predictive performance.

8. Conclusions

This study presents an enhanced paired complex–simple approach through the use of ensembles to empirically evaluate the ability of batch and sequential DA approaches to history match and forecast in the presence of model resolution error and model resolution and process error. The empirical paired complex–simple approach and comparative evaluation comprise novel contributions to the practice of groundwater decision support modeling. Overall, we find that both approaches perform well in both history matching and forecasting when employing a high-dimensional parameterization stance, that is, treating all properties and stresses as uncertain and adjustable during the inversion process. When uncertain parameters are removed from the inversion process, the data assimilation process is degraded in different ways for batch and sequential approaches. Sequential DA is able to effectively reproduce historical observations by subsuming structural noise into the adjustable head states, but fails to reproduce the interannual, nonstationary trends in the model forecasts due to only assimilating observations one cycle (month) at a time. Batch DA is less able to match history for the groundwater and streamflow assimilation locations, which more or less propagates through to biased forecasts. Despite this, batch DA is better able to reproduce long-term trends in the assimilation and forecast quantities owing to the wider assimilation window. These results have implications for groundwater decision support modeling as they underscore the pitfalls of fixing parameters *a priori*, such as with pumping, and present a proof of concept for using adjustable model states to cope with model error in decision support modeling contexts.

9. Software and data availability

Following Donoho et al. (2008) and Stodden (2010) and in line with the FAIR principles, the authors have endeavoured to provide all relevant python code and model files for cross-platform reproducibility of the complete analysis which is available at https://github.com/kmarkovich-intera/Freyberg_DA/releases/tag/v2.0.0. The software tool applied herein is available within the open-source PEST++ code suite available at https://github.com/usgs/pestpp as a C++ code named pestpp-da. Pre-compiled, statically-linked (e.g., stand alone) binaries for Windows, Linux, and MacOS are available, as are cmake, make, and visual studio solutions. This analysis also relied heavily on python modules for Environmental Model Uncertainty analyses (pyEMU), an open-source python package available through pyPI and at https://github.com/pypest/pyemu.

CRediT authorship contribution statement

Katherine H. Markovich: Conceptualization, Methodology, Formal analysis, Writing, Visualization. **Jeremy T. White:** Conceptualization, Methodology, Formal analysis, Writing, Visualization. **Matthew J. Knowling:** Conceptualization, Methodology, Writing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Katherine Markovich reports financial support was provided by INTERA Incorporated. Jeremy White reports financial support was provided by INTERA Incorporated.

Data availability

Following Donoho et al. (2008) and Stodden (2010) and in line with the FAIR principles, the authors have endeavoured to provide all relevant python code and model files for cross-platform reproducibility of the complete analysis which is available at https://github.com/kmarkovich-intera/Freyberg_DA/releases/tag/v2.0.0. The software tool applied herein is available within the open-source PEST++ code suite available at https://github.com/usgs/pestpp as a C++ code named pestpp-da. Pre-compiled, statically-linked (e.g., stand alone) binaries for Windows, Linux, and MacOS are available, as are cmake, make, and visual studio solutions. This analysis also relied heavily on python modules for Environmental Model Uncertainty analyses (pyEMU), an open-source python package available through pyPI and at https://github.com/pypest/pyemu.

Acknowledgments

Authors KHM and JTW were funded by INTERA Incorporated.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.envsoft.2022.105498.

References

Alzraiee, A.H., White, J.T., Knowling, M.J., Hunt, R.J., Fienen, M.N., 2021. A scalable model-independent iterative data assimilation tool for sequential and batch estimation of high dimensional model parameters and states. Environ. Model. Softw. 105284. http://dx.doi.org/10.1016/j.envsoft.2021.105284, URL: https://www.sciencedirect.com/science/article/pii/S1364815221003261.

Chen, Y., Oliver, D.S., 2013. Levenberg-marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Comput. Geosci. 17 (4), 689–703. http://dx.doi.org/10.1007/s10596-013-9351-5.

Clark, M.P., Vrugt, J.A., 2006a. Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters. Geophys. Res. Lett. 33 (6), 1–5. http://dx.doi.org/10.1029/2005GL025604.

Clark, M.P., Vrugt, J.A., 2006b. Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters. Geophys. Res. Lett. 33 (6), http://dx.doi.org/10.1029/2005GL025604, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL025604. arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2005GL025604.

Doherty, J., 2003. Ground water model calibration using pilot points and regularization. Groundwater 41 (2), 170–177. http://dx.doi.org/10.1111/j.1745-6584.2003.tb02580.x, URL: https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6584.2003.tb02580.x. arXiv:https://ngwa.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-6584.2003.tb02580.x.

Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resour. Res. 47 (12), http://dx.doi.org/10.1029/2011WR010763, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR010763.

Doherty, J., Moore, C., 2020. Decision support modeling: Data assimilation, uncertainty quantification, and strategic abstraction. Groundwater 58 (3), 327–337. http://dx.doi.org/10.1111/gwat.12969, URL: https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/gwat.12969. arXiv:https://ngwa.onlinelibrary.wiley.com/doi/pdf/10.1111/gwat.12969.

Doherty, J., Welter, D., 2010. A short exploration of structural noise. Water Resour. Res. 46 (5), http://dx.doi.org/10.1029/2009WR008377, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009WR008377. arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009WR008377.

Donoho, D.L., Maleki, A., Rahman, I.U., Shahram, M., Stodden, V., 2008. Reproducible research in computational harmonic analysis. Comput. Sci. Eng. 11 (1), 8–18. http://dx.doi.org/10.1109/MCSE.2009.15.

Emerick, A.A., Reynolds, A.C., 2013. Ensemble smoother with multiple data assimilation. Comput. Geosci. 55, 3–15. http://dx.doi.org/10.1016/j.cageo.2012.03.011, URL: https://www.sciencedirect.com/science/article/pii/S0098300412000994. Ensemble Kalman filter for data assimilation.

Evans, M., Moshonov, H., 2006. Checking for prior-data conflict. Bayesian Anal. 1 (4), 893–914. http://dx.doi.org/10.1214/06-BA129.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. Oceans 99 (C5), 10143–10162. http://dx.doi.org/10.1029/94JC00572, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JC00572. arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/94JC00572.

Freyberg, D.L., 1988. An exercise in ground-water model calibration and prediction. Groundwater 26 (3), 350–360. http://dx.doi.org/10.1111/j.1745-6584.1988.tb00399.x, URL: https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6584.1988.tb00399.x. arXiv:https://ngwa.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-6584.1988.tb00399.x.

Gupta, H.V., Clark, M.P., Vrugt, J.A., Abramowitz, G., Ye, M., 2012. Towards a comprehensive assessment of model structural adequacy. Water Resour. Res. 48 (8), http://dx.doi.org/10.1029/2011WR011044, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011044. arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011WR011044.

Haitjema, H.M., 2015. The cost of modeling. Groundwater 53 (2), 179. http://dx.doi.org/10.1111/gwat.12321.

Hendricks Franssen, H.J., Kaiser, H.P., Kuhlmann, U., Bauser, G., Stauffer, F., Mller, R., Kinzelbach, W., 2011. Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. Water Resour. Res. 47 (2), 1–20. http://dx.doi.org/10.1029/2010WR009480.

Hendricks Franssen, H.J., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res. 44 (9), 1–21. http://dx.doi.org/10.1029/2007WR006505.

Hill, M.C., 2006. The practical use of simplicity in developing ground water models. Groundwater 44 (6), 775–781. http://dx.doi.org/10.1111/j.1745-6584.2006.00227.x, URL: https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6584.2006.00227.x. arXiv:https://ngwa.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-6584.2006.00227.x.

Huang, C., Newman, A.J., Clark, M.P., Wood, A.W., Zheng, X., 2017. Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States. Hydrol. Earth Syst. Sci. 21 (1), 635–650. http://dx.doi.org/10.5194/hess-21-635-2017, URL: https://hess.copernicus.org/articles/21/635/2017/.

Hughes, J.D., Langevin, C.D., Banta, E.R., 2017. Documentation for the MODFLOW 6 Framework. Technical Report, Reston, VA, http://dx.doi.org/10.3133/tm6A57, URL: http://pubs.er.usgs.gov/publication/tm6A57.

Hunt, R.J., Doherty, J., Tonkin, M.J., 2007. Are models too simple? Arguments for increased parameterization. Groundwater 45 (3), 254–262. http://dx.doi.org/10.1111/j.1745-6584.2007.00316.x, URL: https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6584.2007.00316.x. arXiv:https://ngwa.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-6584.2007.00316.x.

Hunt, R.J., Fienen, M.N., White, J.T., 2020. Revisiting "an exercise in groundwater model calibration and prediction" after 30 years: Insights and new directions. Groundwater 58 (2), 168–182. http://dx.doi.org/10.1111/gwat.12907, URL: https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/gwat.12907. arXiv:https://ngwa.onlinelibrary.wiley.com/doi/pdf/10.1111/gwat.12907.

Hunt, R., Zheng, C., 1999. Debating complexity in modeling. Eos 80 (3), 29. http://dx.doi.org/10.1029/99EO00025.

Kim, S.S.H., Marshall, L.A., Hughes, J.D., Sharma, A., Vaze, J., 2021. Jointly calibrating hydrologic model parameters and state adjustments. Water Resour. Res. 57 (8), e2020WR028499. http://dx.doi.org/10.1029/2020WR028499, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR028499. arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020WR028499. e2020WR028499 2020WR028499.

Knowling, M.J., White, J.T., Moore, C.R., 2019. Role of model parameterization in risk-based decision support: An empirical exploration. Adv. Water Resour. 128, 59–73. http://dx.doi.org/10.1016/j.advwatres.2019.04.010, URL: https://www.sciencedirect.com/science/article/pii/S0309170819300909.

Knowling, M.J., White, J.T., Moore, C.R., Rakowski, P., Hayley, K., 2020. On the assimilation of environmental tracer observations for model-based decision support. Hydrol. Earth Syst. Sci. 24 (4), 1677–1689. http://dx.doi.org/10.5194/hess-24-1677-2020, URL: https://hess.copernicus.org/articles/24/1677/2020/.

Langevin, C.D., Hughes, J.D., Banta, E.R., Niswonger, R.G., Panday, S., Provost, A.M., 2017. Documentation for the MODFLOW 6 Groundwater Flow Model. Technical Report, Reston, VA, http://dx.doi.org/10.3133/tm6A55, URL: http://pubs.er.usgs.gov/publication/tm6A55.

McKenna, S.A., Akhriev, A., Ciaurri, D.E., Zhuk, S., 2019. Efficient uncertainty quantification of reservoir properties for parameter estimation and production forecasting. Math. Geosci. 1–19.

Moore, C., Doherty, J., 2005. Role of the calibration process in reducing model predictive error. Water Resour. Res. 41 (5).

Moore, C.R., Doherty, J., 2021. Exploring the adequacy of steady-state-only calibration. Front. Earth Sci. 9 (September), 1–17. http://dx.doi.org/10.3389/feart.2021.692671.

Oliver, D.S., Alfonzo, M., 2018. Calibration of imperfect models to biased observations. Comput. Geosci. 22 (1), 145–161. http://dx.doi.org/10.1007/s10596-017-9678-4.

Simmons, C.T., Hunt, R.J., 2012. Updating the debate on model complexity. GSA Today 22 (8), 28–29. http://dx.doi.org/10.1130/GSATG150GW.1.

Stodden, V., 2010. The scientific method in practice: Reproducibility in the computational sciences. http://dx.doi.org/10.2139/ssrn.1550193.

Watson, T.A., Doherty, J.E., Christensen, S., 2013. Parameter and predictive outcomes of model simplification. Water Resour. Res. 49 (7), 3952–3977. http://dx.doi.org/10.1002/wrcr.20145, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/wrcr.20145. arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/wrcr.20145.

White, J.T., 2018. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. Environ. Model. Softw. 109 (May), 191–201. http://dx.doi.org/10.1016/j.envsoft.2018.06.009.

White, J.T., Doherty, J.E., Hughes, J.D., 2014. Quantifying the predictive consequences of model error with linear subspace analysis. Water Resour. Res. 50 (2), 1152–1173.

White, J.T., Fienen, M.N., Doherty, J.E., 2016. A python framework for environmental model uncertainty analysis. Environ. Model. Softw. 85, 217–228. http://dx.doi.org/10.1016/j.envsoft.2016.08.017, URL: https://www.sciencedirect.com/science/article/pii/S1364815216305461.

White, J.T., Foster, L.K., Fienen, M.N., Knowling, M.J., Hemmings, B., Winterle, J.R., 2020a. Toward reproducible environmental modeling for decision support: A worked example. Front. Earth Sci. 8, 50. http://dx.doi.org/10.3389/feart.2020.00050, URL: https://www.frontiersin.org/article/10.3389/feart.2020.00050.

White, J.T., Hemmings, B., Fienen, M.N., Knowling, M.J., 2021. Towards improved environmental modeling outcomes: Enabling low-cost access to high-dimensional, geostatistical-based decision-support analyses. Environ. Model. Softw. 139 (February), 105022. http://dx.doi.org/10.1016/j.envsoft.2021.105022.

White, J.T., Hunt, R.J., Doherty, J.E., Fienen, M.N., 2020b. PEST++ Version 5, a Parameter Estimation and Uncertainty Analysis Software Suite Optimized for Large Environmental Models. U.S. Geological Survey Techniques and Methods Report 7-C26.

White, J.T., Knowling, M.J., Moore, C.R., 2020c. Consequences of groundwater-model vertical discretization in risk-based decision-making. Groundwater 58 (5), 695–709. http://dx.doi.org/10.1111/gwat.12957.