Environmental Modelling and Software 109 (2018) 191–201

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions

Jeremy T. White

GNSScience, Wairakei, New Zealand

ARTICLE INFO

Keywords:
Modeling
Uncertainty
Iterative ensemble smoother
Gauss-Levenberg-Marquardt

ABSTRACT

An open-source, scalable and model-independent (non-intrusive) implementation of an iterative ensemble smoother has been developed to alleviate the computational burden associated with history-matching and uncertainty quantification of real-world-scale environmental models that have very high dimensional parameter spaces. The tool, named pestpp-ies, implements the ensemble-smoother form of the popular Gauss-Levenberg-Marquardt algorithm, uses the pest model-interface protocols and includes a built-in parallel run manager, multiple lambda testing and model run failure tolerance. As a demonstration of its capabilities, pestpp-ies is applied to a synthetic groundwater model with thousands of parameters and to a real-world groundwater flow and transport model with tens of thousands of parameters. pestpp-ies is shown to efficiently and effectively condition parameters in both cases and can provide means to estimate posterior forecast uncertainty when the forecasts depend on large numbers of parameters.

Code and data availability

Statically-linked pestpp-ies binaries for PC and Mac OSX operating systems are included in the code repository, https://github.com/dwelter/pestpp, along with a Microsoft Visual Studio solution and makefiles for both Mac OSX and Linux operating systems. The repository includes several fully-worked pestpp-ies examples ranging from the 1-parameter analytical verification test from Chen and Oliver (2013) to a 1.1-million parameter synthetic model used to test the scalability of the implementation. The Freyberg example (Freyberg, 1988) presented below is also included in the benchmark suit. The datasets used for the Hauraki example application are available upon request from Waikato Regional Council.

To increase transparency and reproducibility of the examples presented herein, the model input datasets for both of these example models were constructed with flopy (Bakker et al., 2016) and the pest model-interfaces were constructed with pyemu (White et al., 2016); the driver python scripts used to construct the model interfaces for Freyberg example are included in the code repository.

1. Introduction

For an environmental model to serve effectively in a decision-support role, it must include estimates of the reliability of the important simulated outcomes (e.g., the forecasts of interest). This reliability analysis typically relies on some form of uncertainty quantification, where model input uncertainty is discretized into parameters and the model is used to propagate the uncertainty in these parameters to uncertainty in the forecasts. Given the central role of parameterization in uncertainty analysis, it is critically important to use enough parameters to appropriately (and robustly) represent model input uncertainty (e.g. Moore and Doherty 2005; Dausman et al., 2010; Knowling and Werner 2016; White et al., 2017). Unfortunately, most algorithms for real-world environmental model parameter estimation (PE) and uncertainty quantification (UQ) are computationally constrained by number of adjustable parameter—the “curse of dimensionality”. Because of this constraint, assumptions must be employed to reduce the number of parameters, a form of model simplification (Doherty and Christensen, 2011; Cooley and Christensen, 2006). This form of simplification is not without consequence: it can lead to model error phenomena such as parameter compensation and (undetectable) forecast bias (e.g., Doherty and Christensen, 2011; White et al., 2014).

Recently, iterative ensemble smoothers (iES) have emerged as a class of algorithms for PE and UQ that relax or eliminate the computational bound induced by the number of parameters (Chen and Oliver, 2012); these algorithms can be formulated as model independent (non-intrusive), making them an attractive option for application to existing (or legacy) environmental modeling analyses. While many successful applications of iES are presented in the literature (e.g, Bocquet and Sakov (2013); Emerick and Reynolds (2013); Bocquet and Sakov

E-mail address: j.white@gns.cri.nz.

https://doi.org/10.1016/j.envsoft.2018.06.009
Received 21 March 2018; Received in revised form 24 May 2018; Accepted 1 June 2018
Available online 21 June 2018

(2014); Bailey and BaÃ[1] (2010); Crestani et al. (2013) among others), to my knowledge, there is not a generally-applicable, model-independent, scalable implementation of iES that can be readily employed by practitioners to solve a wide-range of real-world environmental modeling problems. To fill this need, herein, I present pestpp-ies, a model-independent form of iES based on the GLM formulation of Chen and Oliver (2013).

The remainder of this manuscript presents a brief overview of GLM and iES theory, describes the implementation of the pestpp-ies and presents application of pestpp-ies to two example problems. The Supplementary data includes detailed input instructions and algorithmic workflows for pestpp-ies, as well as additional information from the second example problem.

2. Background and theory

The ensemble smoother (ES) was first proposed by Van Leeuwen and Evensen (1996) as a "batch" update alternative to the sequential Ensemble Kalman filter (EnKF) (Evensen, 1994) where all past states (and parameters) are estimated in single update step. While the ES enjoyed some success, it did not perform as well as the ensemble Kalman filter (Van Leeuwen and Evensen, 1996). To improve the performance of ES, Chen and Oliver (2013) wrapped an ES approximation to the tangent linear operator (Jacobian matrix) within the iterative framework of the widely-used Gauss-Levenberg-Marquardt (GLM) algorithm (Oliver et al., 2008; Doherty, 2015). The iterative nature of the iES greatly improved the algorithms ability to minimize the sum-of-squared residuals ($\mathcal{L}_2$ norm) objective function for nonlinear problems, while the ensemble approximation to the Jacobian matrix of the GLM algorithm greatly reduced the computational constraint induced by using large numbers of parameters.

Briefly, the standard, regularized (maximum a posteriori) form Gauss-Levenberg-Marquardt algorithm (GLM) (Hanke, 1997) is:

$$\delta_\theta = -((\mathbf{J}^T\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{J}) + (1+\lambda)\boldsymbol{\Sigma}_\theta^{-1})^{-1}(\boldsymbol{\Sigma}_\theta^{-1}(\theta - \theta_0) + \mathbf{J}^T(\mathbf{d}_{\text{sim}} - \mathbf{d}_{\text{obs}})) \quad (1)$$

where $\mathbf{J}$ is the tangent linear operator known as the Jacobian matrix, λ is the Marquardt dampening parameter, $\boldsymbol{\Sigma}_\varepsilon$ is the measurement noise covariance matrix, $\boldsymbol{\Sigma}_\theta$ is the prior parameter covariance matrix, $\mathbf{d}_{\text{sim}} - \mathbf{d}_{\text{obs}}$ is the residual vector, θ is the current parameter vector, θ_0 is the initial parameter vector, and δ_θ is the parameter upgrade (change) vector. Equation (1) differs from the maximum likelihood form of the GLM by inclusion of $\boldsymbol{\Sigma}_\theta$, which enforces regularization on δ_θ.

Many modern, model-independent implementations (e.g., pest (Doherty, 2015), pestpp (Welter et al., 2015), ucode (Poeter et al., 2014)) of the GLM algorithm rely on filling the Jacobian matrix of Equation (1) with finite-difference approximations to the partial first derivatives:

$$\mathbf{J}[\text{sim}_i, \text{par}_j] = \frac{\partial \text{sim}_i}{\partial \text{par}_j} \approx \frac{\Delta \text{sim}_i}{\Delta \text{par}_j} \quad (2)$$

where 'sim_i' is the simulated equivalent to observation 'obs_i'.

Using the finite-difference approximation is advantageous because it allows implementations of GLM to be parallelized and model-independent (non-intrusive). The model-independence trait is important because it allows any input to any forward model to be treated as a parameter and any output from any forward model—including processed and derived outputs—to be treated as observations. It also facilitates flexibility in forming the composite objective function involving several types of observations—an important consideration when dealing with imperfect, real-world models (e.g., Doherty and Welter, 2010; White et al., 2014).

However, using finite-difference approximated derivatives has a downside: the computational burden of filling the Jacobian matrix is directly related to the number of parameters. As the number of parameters increases, so too does the number of model runs required to fill the Jacobian matrix during each iteration of GLM. To overcome this computational constraint, Chen and Oliver (2013) reformulated the successful GLM algorithm to use a Jacobian matrix derived empirically from an ensemble of random parameter values. The resulting GLM formulation is (following Chen and Oliver (2013):

$$\boldsymbol{\Delta}_\theta = -((\mathbf{J}_{emp}^T\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{J}_{emp}) + (1+\lambda)\boldsymbol{\Sigma}_\theta^{-1}))^{-1}(\boldsymbol{\Sigma}_\theta^{-1}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0) + \mathbf{J}_{emp}^T(\mathbf{D}_{\text{sim}} - \mathbf{D}_{\text{obs}})) \quad (3)$$

where $\mathbf{D}_{\text{obs}}$ and $\mathbf{D}_{\text{sim}}$ are the observation and simulated-equivalent ensembles ($N_e \times N_{obs}$), respectively, $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}_0$ are the current and initial parameter ensembles ($N_e \times N_{par}$), and $\boldsymbol{\Delta}_\theta$ is the parameter upgrade matrix ($N_e \times N_{par}$). Using this formulation, $\mathbf{J}_{emp}$, the empirical Jacobian, is calculated as:

$$\mathbf{J}_{emp} \approx \boldsymbol{\Sigma}_\varepsilon^{\frac{1}{2}}\boldsymbol{\Delta}_{\text{sim}}\boldsymbol{\Delta}_{\text{par}}^{-1}\boldsymbol{\Sigma}_\theta^{-\frac{1}{2}} \quad (4)$$

where

$$\boldsymbol{\Delta}_{\text{sim}} = \boldsymbol{\Sigma}_\varepsilon^{-\frac{1}{2}}(\mathbf{D}_{\text{sim}} \ominus \overline{\mathbf{D}}_{\text{sim}})/\sqrt{N_e - 1} \quad (5)$$

$$\boldsymbol{\Delta}_{\text{par}} = \boldsymbol{\Sigma}_\theta^{-\frac{1}{2}}\left(\boldsymbol{\Theta} \ominus \overline{\boldsymbol{\Theta}}\right)/\sqrt{N_e - 1} \quad (6)$$

where $\overline{\mathbf{D}}_{\text{sim}}$ and $\overline{\boldsymbol{\Theta}}$ are the mean values of the simulated equivalents to observations and parameters across their respective ensembles and N_e is the number of realizations ("$\ominus$" denotes a broadcast subtraction operation). In this formulation, to fill an approximate Jacobian matrix, the model needs to only be run once for each member of the ensemble (i.e. realization) rather than once for each parameter, effectively eliminating the computational burden induced by using a large number of parameters. This formulation maintains the ability to be model independent and includes the flexibility of standard, deterministic implementations of the GLM (e.g. Doherty 2015; Welter et al., 2015; Poeter et al., 2014). Additionally, since an ensemble of parameter realizations (typically drawn from the prior parameter covariance matrix,) is propagated through the algorithm until an acceptable fit with observations is found, this algorithm also yields an estimate of the posterior parameter distribution that can be used to quantify uncertainty in forecasts of interest. Interested readers are referred to Chen and Oliver (2013) and the references cited therein for more detailed and rigorous explanation of this algorithm.

3. Implementation and workflow

3.1. General

A form of equation (3) has been implemented in pestpp software suite (Welter et al., 2015) and is named pestpp-ies; the implementation is based on the "LM-EnRML" algorithm of Chen and Oliver (2013). The code is written mostly in C++11 and makes heavy use of the existing pestpp code base and the Eigen numerical linear algebra template library (Guennebaud et al., 2010). The pestpp-ies iES implementation achieves model independence through the use of the popular pest model-interface protocols (Doherty, 2015). Some key features of the implementation include:

- serial or parallel ensemble evaluation, the latter of which completed via a built-in TCP/IP run manager. See Welter et al. (2015) for parallel run manager usage details.
- multiple (and option parallel) λ testing. Experience in environmental modeling suggests evaluating different values of λ can greatly increase the efficiency of finding an optimal solution (Doherty, 2015).
- subset ensemble evaluation. Although evaluating multiple lambdas is an important consideration, in the iES framework of Chen and Oliver (2013), each lambda evaluation requires propagating the

entire ensemble forward (i.e. running the model once for each realization). pestpp-ies allows users to evaluate only a subset of the whole ensemble for each lambda value, thereby saving many model evaluations while still identifying which value of lambda is likely optimal in reducing the objective function. Once the optimal lambda value is found for a given iteration, the remaining realizations in the corresponding ensemble are evaluated (in parallel) before the next iteration is started.

- fault tolerance: failed run and "bad" run handling. If a particular realization of parameter values causes the forward run to fail (not run to completion) or yields an objective function that is beyond a user-specified threshold of "acceptable", then that realization is simply dropped from the analysis.
- support for drawing (multivariate) Gaussian realizations.
- support for inequality constraints. Observations assigned an observation group name beginning with "less_"/"l_" (less than) or "greater_"/"g_" (greater than) will be treated as inequality constraints. If the corresponding simulated value does not violate the inequality, its residual (e.g., location in $\mathbf{D}_{sim} - \mathbf{D}_{obs}$) is assigned zero—it does not contribute to the objective function or the upgrade calculations. See White et al. (2018) for more information about implementing these types of constraints.

3.2. Input

pestpp-ies can construct all of the necessary elements needed to apply Equation (3) "on-the-fly" from the information in tbhe pest control file, allowing the numerous existing and current modeling analyses using pest and pestpp to apply this tool without any modification. However, detailed algorithmic control is exposed through optional keyword arguments added to the end of the pest control file. For example, existing parameter and/or observation ensembles can be loaded from CSV files as can a "restart" observation ensemble. This functionality allows practitioners to restart a pestpp-ies analysis from a previous analysis, use parameter realizations generated from geostatistical simulations, or to pre-filter a parameter ensemble by eliminating realizations that do not suite some external criteria. A complete listing of the optional inputs for pestpp-ies, along with algorithmic flowcharts, are included in the Supplementary data.

3.3. Output

pestpp-ies writes a run record file (".rec") and a performance log file (".pfm") during each application. Furthermore, the current parameter ensemble and corresponding simulated ensemble are written to comma-separated value (CSV) files at the end of each iteration. pestpp-ies also writes several iteration-summary CSV files corresponding to the measurement, regularization, composite, and actual objective function values for active realizations. Through the use of the + +ies_verbose_level (3) optional argument (see Supplementary data), users can also have all temporary matrices used in various calculations dumped to ASCII files.

3.4. Other considerations in applying pestpp-ies

The prior parameter distribution, $\mathcal{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$ plays a central role in Equation (3) and is critical to a successful application of pestpp-ies. It, therefore, deserves special consideration. The mean vector of the prior parameter distribution ($\boldsymbol{\mu}_\theta$) is taken from the initial parameter estimates in the pest control file. The prior parameter covariance matrix $\boldsymbol{\Sigma}_\theta$ can be formed in several ways. If correlation is not recognized in prior parameter distribution, then $\boldsymbol{\Sigma}_\theta$ can be constructed "on-the-fly" and without user intervention from the parameter bounds in the pest control file—this is the default case if no additional options are specified. By default, this approach assumes the parameter bounds represent a 4 standard deviation envelop ($\pm 2\sigma$ around the initial parameter values), which is approximately 95% confidence interval, although users can specify a different number of standard deviations that the parameter bounds represent (see Supplementary data). If correlation is recognized in the prior parameter distribution, then users may specify an optional parameter covariance matrix file. This file follows the pest file formats and may be ASCII or binary format. In many cases, the correlation in the prior parameter distribution may be implied by a geostatistical structure (e.g., variogram), in which case, practitioners may use the support utilities in pyemu (White et al., 2016) to help construct this matrix.

The algorithm encoded in pestpp-ies is significantly different from standard, deterministic implementation of GLM. As such, there are some important considerations that can greatly improve the efficiency of applying pestpp-ies to complex, numerically "delicate", real-world environmental models. For example, in standard GLM, practitioners must take care to use very small solver closure tolerances in the forward model to ensure the fidelity of the finite-difference partial first derivatives in the **J** matrix. However, pestpp-ies functions with a much coarser **J**; high fidelity model outputs are therefore not as important for successful history matching, meaning larger closure tolerances may be used, which, in some settings, may decrease the computational burden of each model forward run.

In sampling from the prior parameter distribution, parameter realizations may be drawn that induce instability in the forward model. This instability may manifest as either a failure of the forward model run, corrupted/null model outputs, or as an extreme increase model runtime. pestpp-ies offers several ways to cope with these problems. Firstly, pestpp-ies tolerates failures of the forward run seamlessly—realizations that cause the model forward run to fail (detected by the run manager) are removed from the active ensembles. Secondly, through the "+ +ies_bad_phi" argument (see Supplementary data), users can identify forward runs with corrupt and/or null output, which are also removed from the active ensembles. To cope with the issue of an extreme increased forward run time, the parallel run manager built into pestpp-ies accepts an additional argument ("+ +overdue_giveup_fac()") which results in forward runs that have been running for greater than *overdue_giveup_fac*mean*(*runtime*) being marked as model run "failures".

4. Example applications

Two example applications are presented to demonstrate the applicability of pestpp-ies to efficiently history match environmental models with high-dimensional input spaces. First, a synthetic groundwater model based on the model of Freyberg (1988) is considered. The Freyberg application is used to validate the performance of pestpp-ies. Then, pestpp-ies is applied to a real-world groundwater flow and transport model of the Hauraki Plains, New Zealand, to demonstrate the capability of pestpp-ies to history match and quantify uncertainty in a high-dimensional, real-world setting.

4.1. Freyberg

The first example application is a synthetic MODFLOW-2005 (Harbaugh, 2005) model due to Freyberg (1988). This example application was used to evaluate the performance and outcomes of pestpp-ies. The groundwater flow model has 1 layer, 40 rows and 20 columns of 250 m finite-difference model cells. The model has three stress periods:

1. History-matching: 1-day, steady-state
2. Forecast: 1825-day, transient
3. Forecast: 1-day, steady-state (using the same forcing as stress period II).

Conceptually, water enters the model domain via rainfall recharge and exists the domain via abstraction wells, surface-water discharge (MODFLOW RIV-type boundary) and basin outflow (specified head

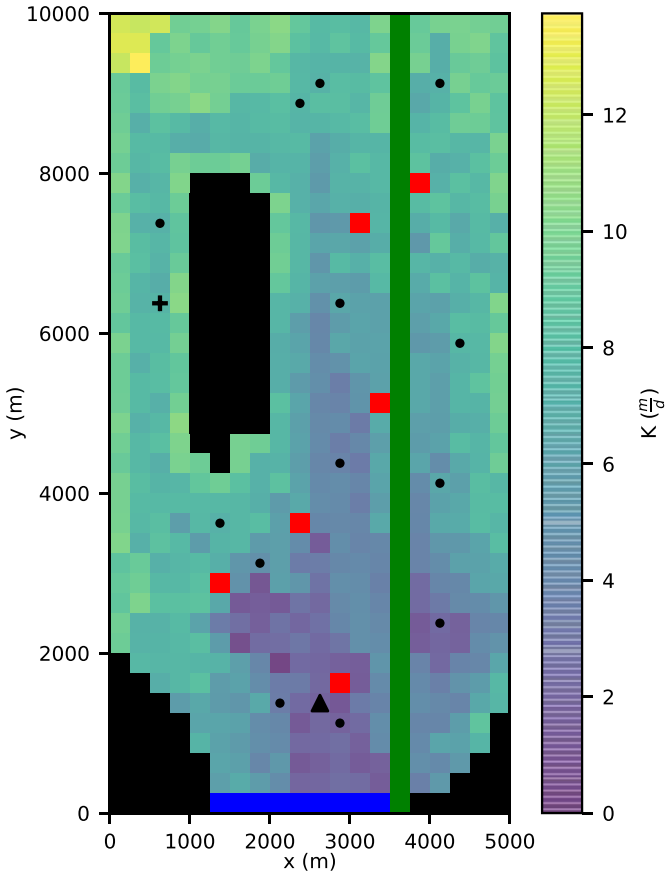

Fig. 1. Schematic of the Freyberg model domain and "true" hydraulic conductivity field. Inactive areas are show in black; Boundary conditions include abstraction wells(red), head-dependent flux (RIV-type) (green), and specified head (blue). Black dots show groundwater level observation locations, the particle starting location is shown as a black cross, the forecast groundwater level location is shown as a black triangle.

boundary) (Fig. 1). Stress periods II and III receive less recharge and have increased groundwater abstraction.

For this example application, three outputs from the model were treated as forecasts:

- groundwater level at an unmeasured location at the end of the stress period II
- surface-water/groundwater exchange rate for all RIV boundaries during stress period II
- particle travel time during the during the forecast stress periods (II and III).

The particle travel-time forecast was implemented using a single particle released near the divide (Fig. 1) and tracked forward using MODPATH6 (Pollock, 2012) until reaching a boundary (stress period III was added to ensure that the particle reaches a boundary). These forecasts represent a wide range of model outputs and also represent outputs that are influenced by different parameters and parameter components (e.g. different degrees of null space dependence (Moore and Doherty, 2006)), which provides an important check on the algorithm encoded in pestpp-ies.

Forecast quantities of interest were added to the pest model interface as zero-weighted observations to facilitate monitoring of these important quantities throughout testing and application of pestpp-ies (e.g., White (2017)). By including the forecast in the pest model interface, the final observation ensemble ($\mathbf{D}_{sim}$) includes estimates of the forecast posterior distribution.

Table 1
Prior parameter uncertainties specified for the Freyberg example problem.

Type	count	transform	standard deviation
hydraulic conductivity	800	log	0.25
historic recharge	800	log	10%
future recharge	800	log	10%
historic abstraction rate multiplier	6	log	37.5%
future abstraction rate multiplier	6	log	37.5%

The "true" model inputs (hydraulic conductivity (Fig. 1), historic and forecast recharge, historic and forecast groundwater abstraction) were run forward with MODFLOW-2005 and MODPATH6 to generate the "true" value for the 13 observation locations (Fig. 1) as well as for the forecasts of interest. Normally-distributed stochastic noise with a mean of 0.0 and standard deviation of 2.0 m was added to the 13 "true" groundwater levels to form the history-matching dataset.

4.1.1. Parameterization

A total of 2412 parameters were specified to represent model input uncertainty; the parameterization includes many model inputs that are commonly recognized as uncertain in real-world groundwater models (Table 1).

The prior parameter covariance matrix ($\mathbf{\Sigma}_\theta$) was specified as a block diagonal matrix; spatially-distributed parameters were assigned distance-based covariates using an exponential variogram with a range of 2500 m and sill of 1.0. The resulting covariance matrix for each spatially-distributed parameter type was scaled by the standard deviations shown in Table 1.

4.1.2. Application

Three different pestpp-ies analyses were undertaken using 30, 50, and 100 realizations, respectively; 10 iterations of Equation (3) were completed for each of the analyses. Other optional arguments used for this analysis include:

- ies_lambda_mults(0.1,1.0,10,0) - evaluate three values for λ each iteration
- ies_subset_size(10) - run the first ten realizations for each of the three $\mathbf{\Theta}_\lambda$ ensembles to identify which is optimal
- ies_use_prior_scaling(false)
- parcov_filename(prior.cov) - use a $\mathbf{\Sigma}_\theta$ stored in ASCII format in the file "prior.cov"

As a verification of the uncertainty estimates yielded by pestpp-ies, a Monte Carlo rejection sampling analysis (e.g., Tarantola, 2005) was completed by evaluating 100,000 parameter realizations generated from the initial parameter estimates and prior parameter covariance matrix—samples from the prior distribution. Once the 100,000 model runs were complete, they were post-processed; realizations yielding an objective function less than or equal to 13.0 (the theoretical, error-based-weighting minimum) were used to collectively form the posterior distribution for the three forecasts. The posterior forecast distributions yielded by the rejection sampling were compared to the posterior distributions yielded by pestpp-ies. Note the Monte Carlo analysis was also completed using pestpp-ies by specifying + + ies_num_reals(100000) and setting the pest control variable noptmax to −1.

4.1.3. Results

Each of the 30-, 50- and 100-realization cases matched the 13 groundwater level observations to near (or below) the theoretical minimum objective function value of 13.0 in a very limited number of model runs (Fig. 2)—a single iteration of standard, deterministic GLM requires 2412 model runs. For comparative purposes, pestpp Welter et al. (2015) was applied to this problem using regularized Tikhonov (Tikhonov and Arsenin, 1977) inversion and "on-the-fly" subspace

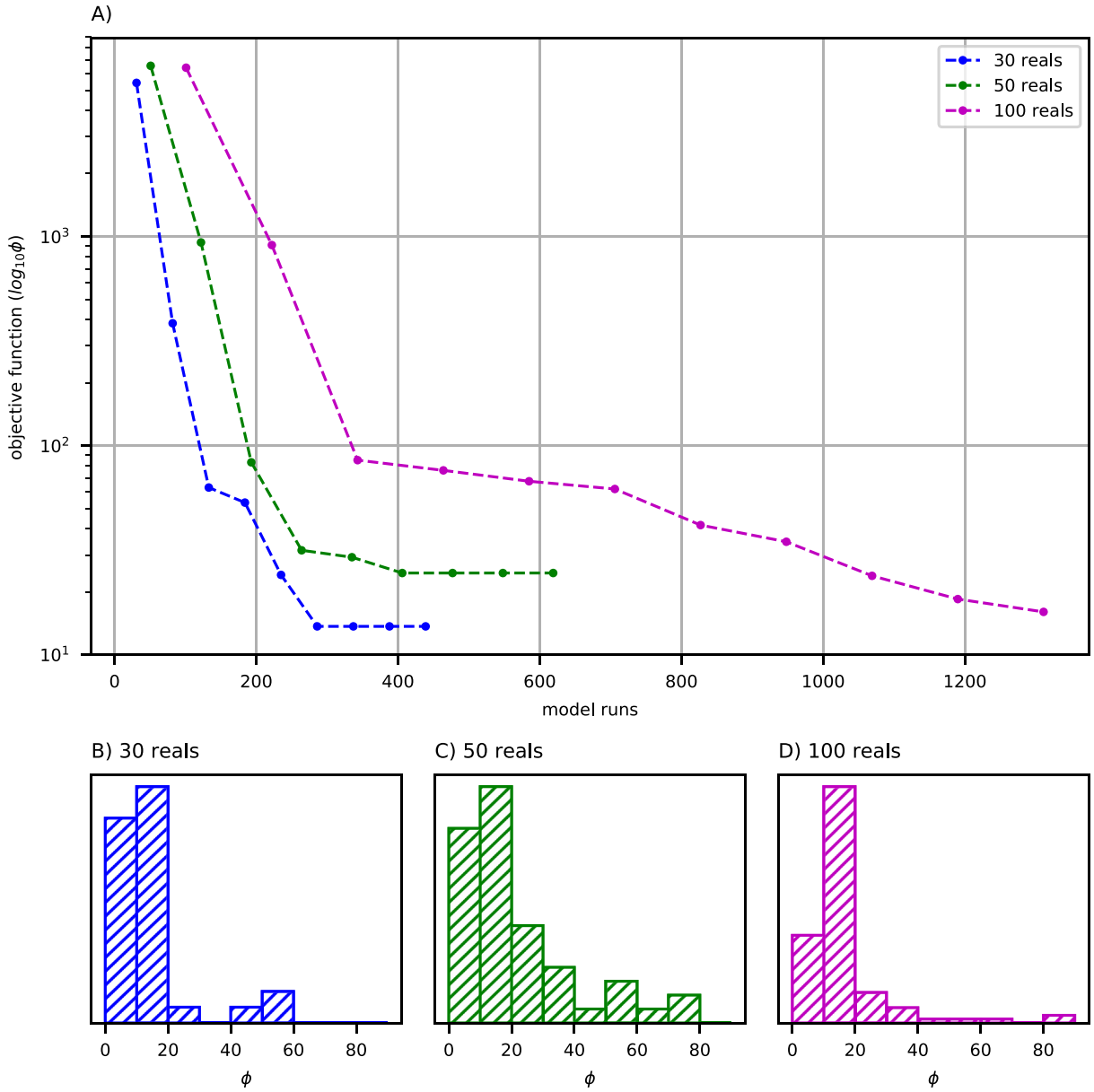

Fig. 2. A) Summary of the mean objective function versus number of model runs for each of the three cases and the final phi distribution for the B) 30-realization C) 50-realization and D) 100-realization cases. All cases are able to match the observations to the level of measurement noise.

reparameterization (e.g. the "SVD-Assist" methodology of Tonkin and Doherty (2005)). pestpp achieved an objective function minimum of 13.0 after required 2450 model runs (more than 10 times the average number of pestpp-ies model runs). More importantly, pestpp yields a single set of history-matched parameters, whereas pestpp-ies yields an ensemble of history-matched parameters.

A single hydraulic conductivity realization from the 10th iteration of each of the three cases is shown on Fig. 3, along with the minimum error variance hydraulic conductivity fields from the 10th iteration (the minimum error variance parameters are included in the pestpp-ies analysis by adding the initial parameter values in the control file as an additional realization—a default behavior). These plots reveal that the conditioned stochastic realizations maintain their general stochastic character while the minimum error variance field shows considerably less variability compared to the stochastic fields—similar to the solution yielded by application of Equation (1) (e.g., deterministic GLM). This is an encouraging outcome since, in certain modeling settings, a single, deterministic parameter set may be of use—the minimum error variance parameters are the best candidates in this situation (see Fig. 4).

The Monte Carlo rejection sampling identified 275 realizations that yield an objective function value less than or equal to 13.0 (an acceptance rate of approximately 0.275%). These 275 realizations were collectively treated as the "correct" posterior distribution for each of the three forecasts. Comparing the posterior distributions yielded by pestpp-ies to the "correct" posterior distributions indicates that pestpp-ies is appropriately estimating both the central tendency and the distribution of the forecast posterior distributions for the three forecasts used in this test problem.

4.2. Hauraki Plains

pestpp-ies was applied to a real-world groundwater flow and transport model constructed for the Hauraki Plains on the North Island of New Zealand (Fig. 5). Briefly, the model is a finite difference MODFLOW-NWT (Niswonger et al., 2011) model with 7 layers, 124 rows, and 70 columns with a uniform grid spacing of 1000 m. Model layer thicknesses were assigned to coincide with production and/or monitoring zones (layer thicknesses are shown in the Supplementary data).

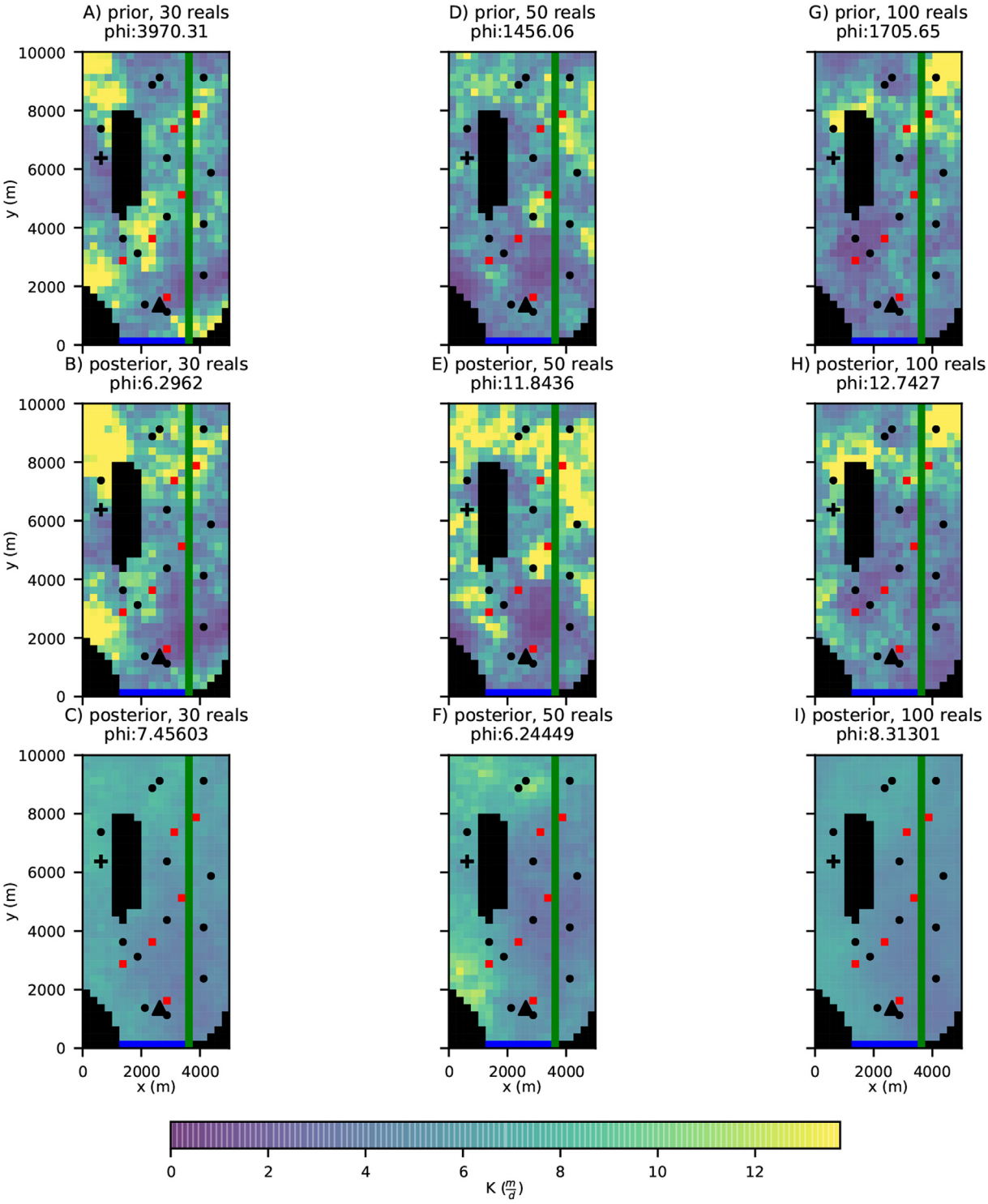

Fig. 3. The prior and posterior hydraulic conductivity field for the first realization in the ensemble at 10th iteration. The colormap of each field uses range of the "true" hydraulic conductivity field for reference. The final minimum error variance realizations (bottom row) show little heterogeneity, Note the initial minimum error variance field was a uniform 5 $\frac{m}{d}$.

The model is steady-state and uses long-term average forcing conditions and long-term average groundwater levels and surface-water fluxes for conditioning. Conceptually, water enters the aquifer system through recharge and discharges to the local surface-water system, which flows north to the Firth of Thames.

The initial estimates of horizontal hydraulic conductivity were derived from a geologic model of the Hauraki Plains (White et al., 2015); the geologic model was spatially intersected with the model layers, forming hydraulic conductivity zones (the layer-based zonation is show in the Supplementary data).

The fate and transport of nitrate is an important issue in the Hauraki Plains. As such, mt3d-usgs (Bedekar et al., 2016) was used to simulate the movement of nitrate within the aquifer and surface-water system (surface water transport was implemented using the SFT process available in mt3d-usgs) for a period of 50 years of steady-state hydrologic conditions and nitrate loading. Nitrate loading was simulated as a

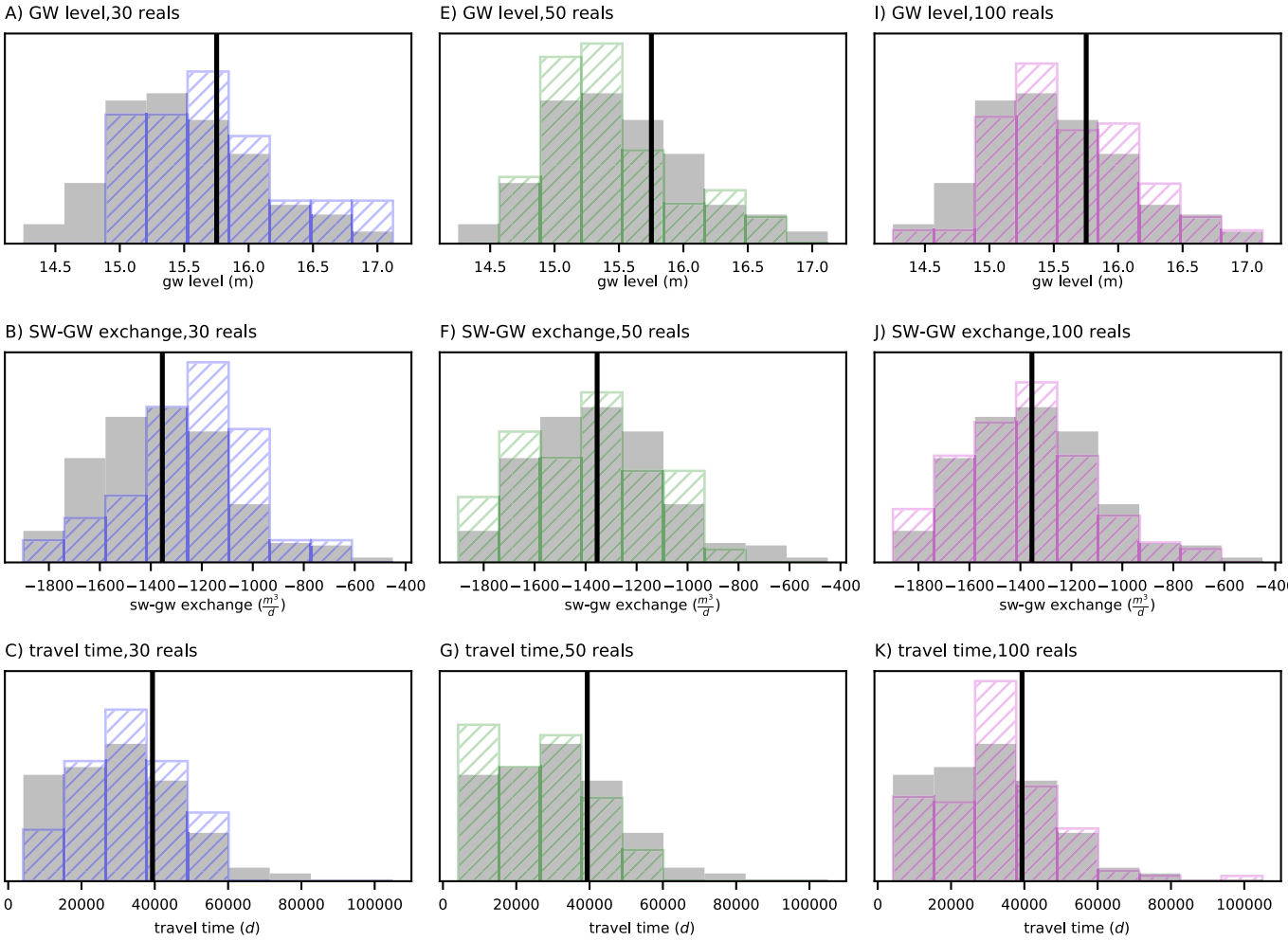

Fig. 4. A comparison of the posterior ensembles found by Monte Carlo rejection sampling (grey) and pestpp-ies (hatched colors) for the groundwater level forecast (top row), surface-water/groundwater exchange forecast (middle row) and travel time forecast (bottom row). The Monte Carlo rejection sampling yielded 275 realizations with an objective function less 13.0. The vertical, solid black line marks the "true" value of each forecast, which is bracketed by significant probability mass in all cases. In general, the pestpp-ies posterior compares well with the rejection-sampling posterior.

mass loading boundary condition in model layer 1 based on the long-term loading estimates for the Hauraki Plains derived from overseer modeling (Ledgard et al., 1999). De-nitrification was simulated using a first-order decay rate; initial estimates for the spatial distribution of decay rates were derived from Wilson et al. (2018) (the initial first-order decay rate estimates are shown in the Supplementary data).

Other relevant model features include:

- long-term average groundwater abstraction derived from available records. Groundwater abstraction is simulated with the MODFLOW WEL package
- long-term average recharge estimates were used from the New Zealand TopNET model (McBride et al., 2016) and simulated with the MODFLOW RCH package
- Surface water network simulated using the MODFLOW SFR package (Niswonger and Prudic, 2005) (Note: long-term average runoff and direct inflows to reaches derived from the TopNET model simulations)

For the purposes of this example analysis, the forecasts of interest are:

- long-term average groundwater levels at all active model cells in layer 1 (3106 forecast values)
- long-term average surface-water/groundwater exchange for all SFR segments with in the model domain (Fig. 5) (183 forecast values)
- long-term average groundwater nitrate concentration at all active model cells in layer 1 (3106 forecast values).

Similar to the Freyberg example application, the forecast quantities of interest are listed as zero-weight observations in the pest control file, resulting in the simulated forecast values being recorded after every model run. In this way, the final simulated ensemble ($\mathbf{D}_{sim}$) contains samples from the posterior forecast distributions.

4.2.1. Parameterization

Model input uncertainty was discretized into 83,905 parameters. All spatially-distributed input uncertainty was conceptualized as occurring both at the scale of the geologic model (large-scale zones) and also at the model-cell scale (grid scale). To accomodate these two scales of recognized model input uncertainty, the parameterization of spatially-distributed hydraulic properties used multiplier arrays of both geologic model scale (zones) and arrays of individual model cell multipliers. In this way, the zone-based parameters estimate the broad scale "mean" property value within a given zone while the grid-scale parameters account for intra-zone heterogeneity. This combined multiplier array parameterization was applied to horizontal and vertical hydraulic conductivity, porosity and first-order reaction coefficients in all model layers and to recharge. Vertical hydraulic conductivity was specified as the ratio of horizontal to vertical hydraulic conductivity. Additional

parameters were specified for hydraulic conductance, runoff and inflow for each of the 183 SFR segments (runoff and inflow multiplier parameters were only specified for segments with non-zero runoff or inflow, respectively). The nitrate loading rate each active cell in model layer 1 was assigned a multiplier parameter as well. Given the expected relation between the groundwater abstraction rates and the history matching observations (and forecasts of interest), and the presence of uncertainty in the abstraction rate estimates, a multiplier parameter was added to each of 617 abstraction wells.

The prior parameter covariance matrix ($\boldsymbol{\Sigma}_\theta$) used in the application of pestpp-ies to the Hauraki Plains model was conceptualized as a block-diagonal matrix, where off-diagonal elements (i.e. covariates) were specified for the grid-scale parameters using an exponential variogram with a range of 10,000 m and a sill of 1.0. The resulting covariance matrices associated with grid-scale parameters were multiplied by the variances implied by the associated prior parameter confidence limits (shown in the Supplementary data). The remaining parameters were treated as independent in the prior parameter covariance matrix. Both the prior parameter covariance matrix and the initial parameter ensemble were generated using pyemu (White et al., 2016).

4.2.2. Observations

A total of 495 long-term average groundwater levels and 33 surface-water fluxes were used for history matching (locations are shown in the Supplementary data). The long-term average groundwater level observations were assigned a measurement standard deviation based on the number of measurements available for averaging at a given location:

- number of measurements > 10: 1.0 m
- 10 < number of measurements > 5: 4.0 m
- 5 < number of measurements: 10 m

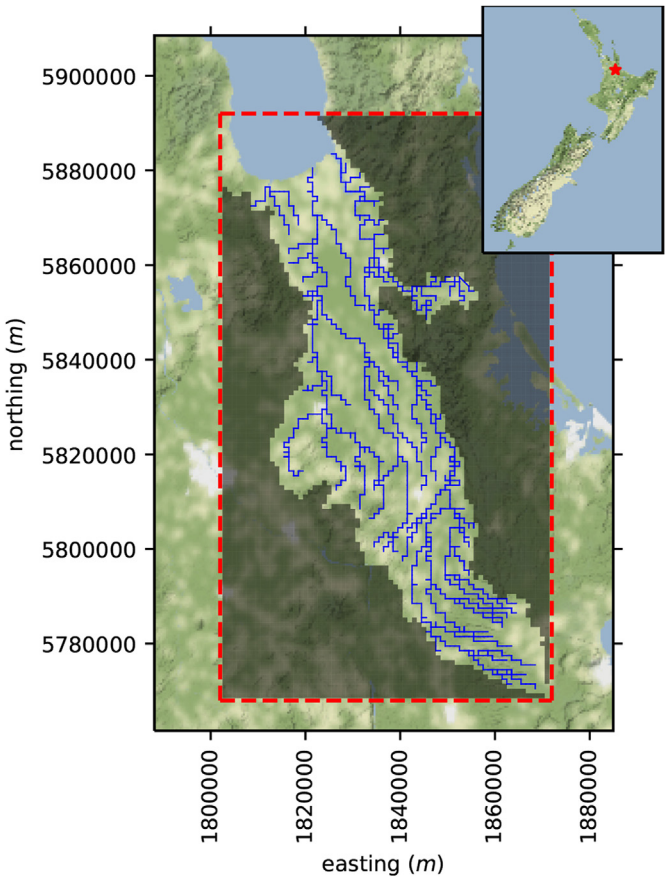

Fig. 5. Hauraki Plains model extent (red dashed line), layer 1 inactive area (shaded) and surface water network (blue lines).

The long-term average surface-water flux observations were assigned measurement standard deviation of 10% of the long-term average flux, reflecting the expected heteroscedastic character of the measurement errors.

Given the lack of observations of water table elevation (simulated groundwater level in model layer 1; see Supplementary data for groundwater level observation locations), it is important to provide some constraints to enforce the physical condition that the simulated water table be at (near) or below land surface. This preferred condition was implemented using "less-than" inequality constraints: the simulated groundwater levels in every active model cell in model layer 1 assigned an "observed value" equal to land surface (the MODFLOW discretization top input); these "observations" were assigned to an observation group "less_k0_hds", which identifies them as less-than inequality constraints. Each inequality constraint was assigned a weight of 0.1.

4.2.3. Application

An ensemble of 100 realizations was used for the Hauraki Plains analysis. Since the observations used for history matching do not contain information pertaining to porosity, first-order reaction rates or nitrate loading rates, these parameters were not subjected to estimation, rather the values for these parameters were not adjusted from the stochastic values in initial parameter ensemble ($\boldsymbol{\Theta}^0$). This can be thought of as a form of localization/regularization (Chen and Oliver, 2017) to prevent unwarranted adjustment of these parameters. Future work will focus on the inclusion of transport-specific observations and conditioning these parameters. However, given the dependence of simulated groundwater and surface-water nitrate concentrations on these parameters, it is important to express their uncertainty in the analysis.

The following optional arguments were used in the application of pestpp-ies to the Hauraki Plains model:

- + + ies_lambda_mults(0.1,1.0,10)
- + + ies_subset_size(10)
- + + ies_parameter_ensemble(hauraki.par.csv)
- + + ies_bad_phi(50000)

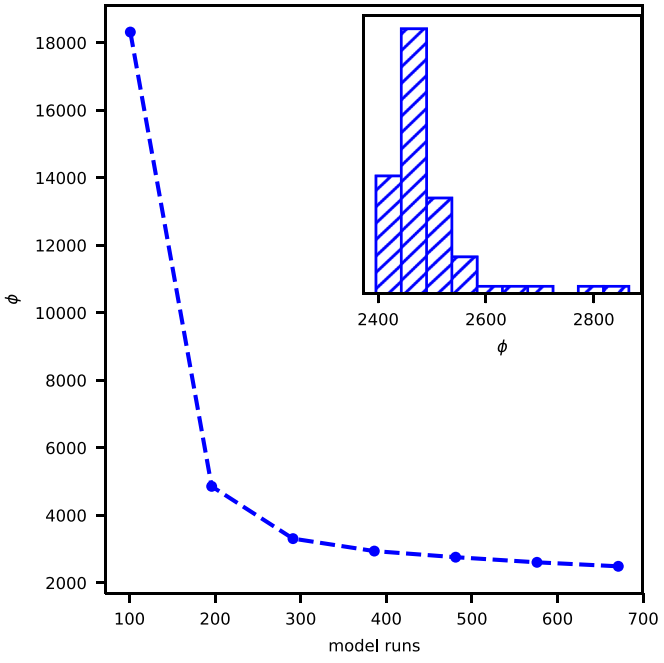

Fig. 6. Summary of mean objective function vs number of model runs. The inset shows final objective frequency distribution.

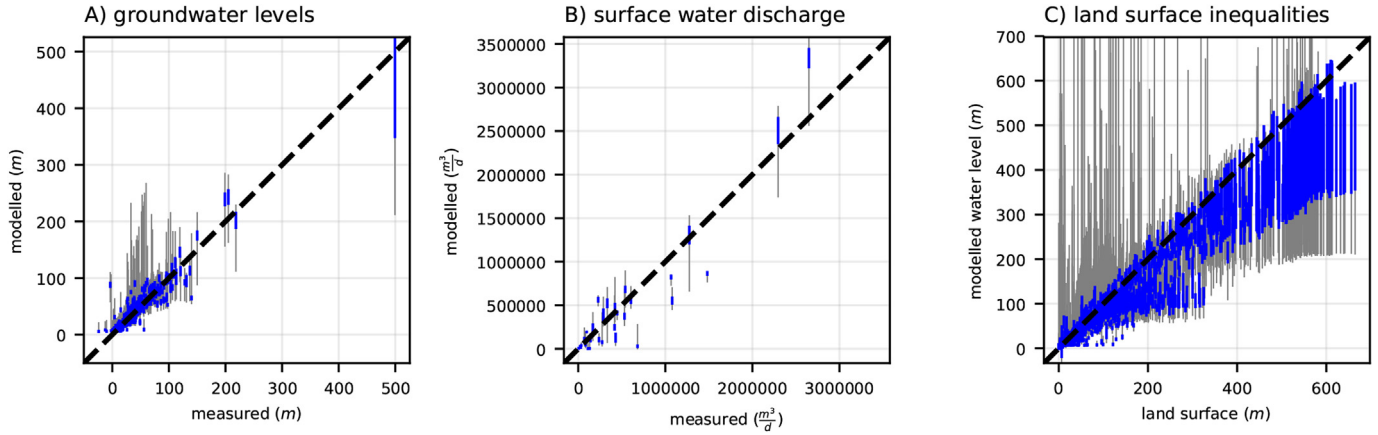

Fig. 7. Plot of observed to simulated A) groundwater levels, B) surface water fluxes and C) water-table inequality constraints. The vertical gray bars show the prior range; the vertical blue lines show the posterior range. The effect of the inequality constraints is seen as limiting the posterior ranges to below land surface.

4.2.4. Results

pestpp-ies was run for 6 iterations for a total of 827 model runs (Fig. 6); the match to both the long-term average groundwater levels and surface-water fluxes were improved significantly after 3 iterations (Figs. 6 and 7).

In general, the posterior parameter realizations retain the stochastic character of the prior realizations, only they have been adjusted via Equation (3) to better reproduce the observation data. For example, the prior and posterior horizontal hydraulic conductivity arrays for the first realization in the ensemble are shown on Fig. 8 (additional realizations are shown in the Supplementary data).

The spatial distribution of the percent forecast uncertainty reduction resulting from history matching shows spatial variability as well as differences across the three forecast types (Fig. 9). Percent forecast

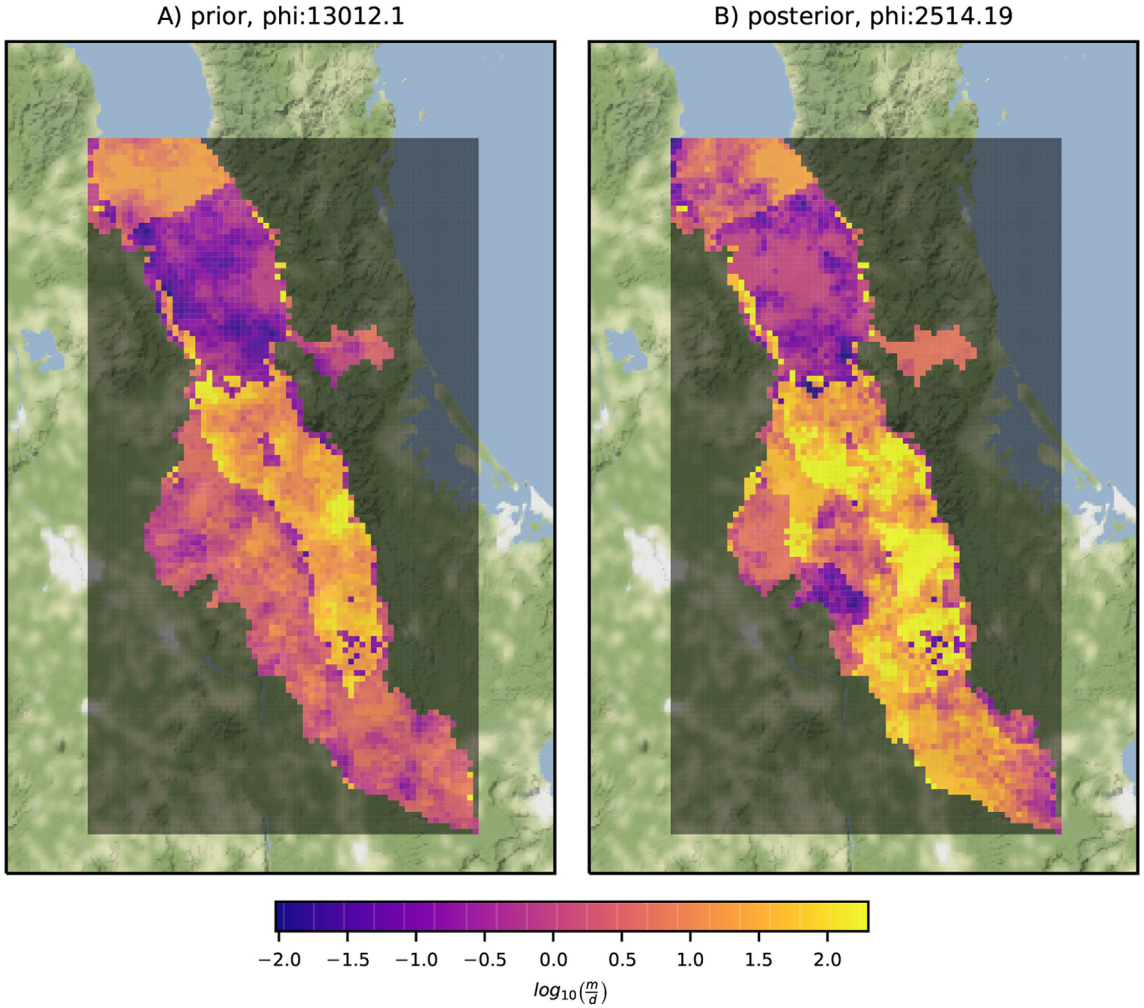

Fig. 8. The prior and posterior horizontal hydraulic conductivity arrays for the first realization in the ensemble at iteration 6. These arrays include the effects of the both the zone and grid-scale multiplier parameters.

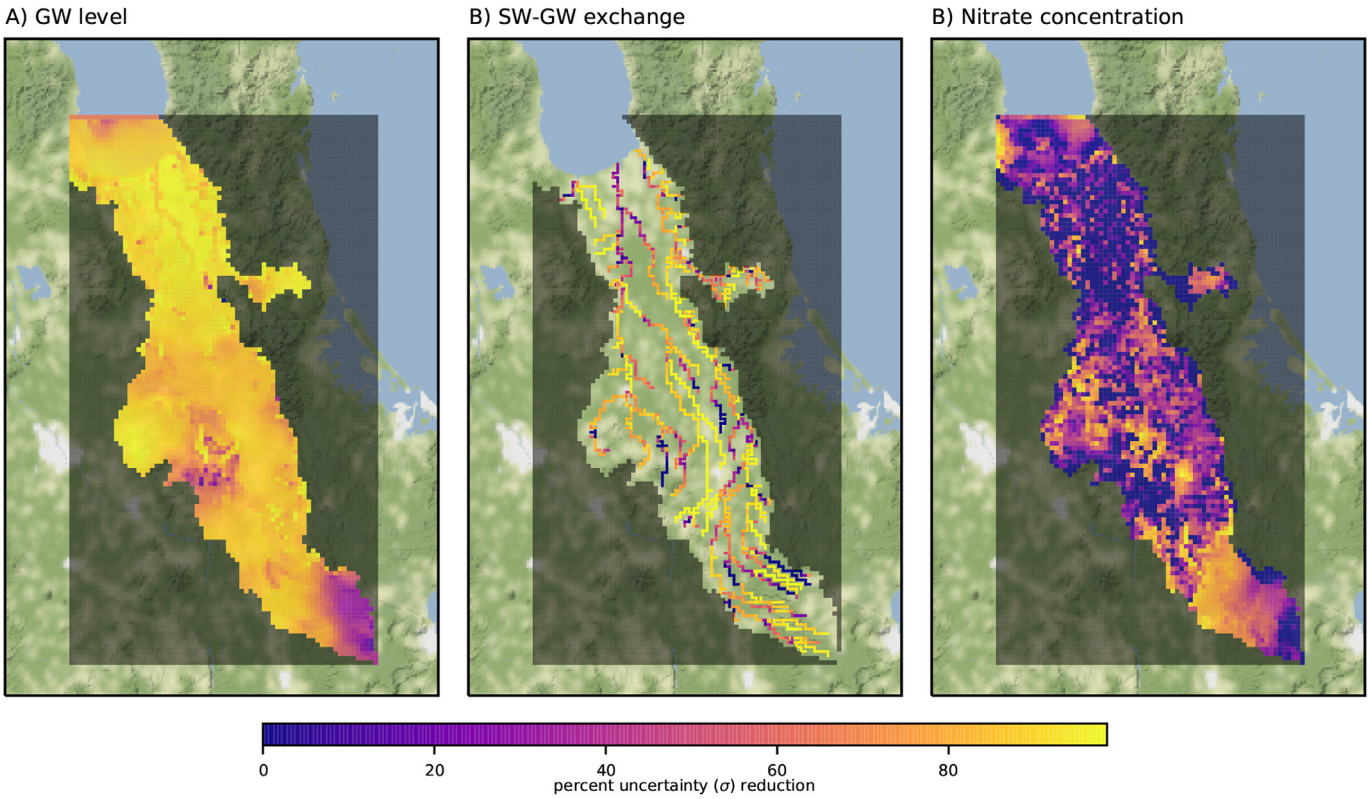

Fig. 9. Percent uncertainty reduction (posterior vs. prior standard deviation) for A)groundwater levels, B) surface-water/groundwater exchange and C) nitrate concentration forecasts. The groundwater level forecasts show considerable reduction across much of the model domain, while the surface-water/groundwater exchange forecasts show more spatial variability. In general, the nitrate concentration forecasts show less reduction in uncertainty and increased spatial complexity compared to the groundwater level forecasts.

uncertainty reduction was calculated using the initial and final simulated equivalent to observation ensembles as

$$\Delta_{sigma,s} = 100.0\left(1.0 - \frac{\overline{\sigma}_s}{\sigma_s}\right) \tag{7}$$

where σ_s and $\overline{\sigma}_s$ are the prior and posterior standard deviations, respectively, for forecast s. This calculation was repeated for each forecast (i.e. each active model cell for groundwater level and nitrate concentration forecast and each SFR reach for surface-water/groundwater exchange forecasts).

As expected, history matching to several hundred groundwater levels has reduced the uncertainty in simulated groundwater levels at many unmeasured locations. Furthermore, the uncertainty in surface-water/groundwater exchanged was also substantially reduced for several SFR segments. However, many SFR segments display no reduction in uncertainty from history matching. The spatial distribution of nitrate shows less reduction in uncertainty as a result of history matching groundwater levels and surface-water fluxes; the spatial pattern of uncertainty reduction is also more complex than the groundwater level forecasts, likely attributable to this forecast's dependence on local-scale heterogeneity. In general, the forecast uncertainty results are not unexpected and underscore the importance of uncertainty quantification to place model results in a reliability context.

5. Conclusion and future directions

pestpp-ies is an open-source, scalable, and model-independent implementation of the GLM iterative ensemble smoother algorithm of Chen and Oliver (2013). The implementation includes a built-in parallel run manager and facilities for dealing with issues that arise in real-world environmental modeling (e.g., multiple lambda testing, failed/bad run handling, minimum error variance parameter inclusion). Herein, the ability of pestpp-ies to efficiently quantify posterior forecast uncertainty for both a synthetic and real-world groundwater model has been demonstrated.

Future work will focus on tighter integration with (and better support for) geostatistical simulations, as well as research towards supporting geostatistical realizations with higher-order moments, such those arising from multi-point geostatistics. Research will also focus on enhanced support for inequality constraints and formal objective function optimization into the pestpp-ies framework as well as integration of machine learning techniques to improve the empirical Jacobian matrix estimate. Also important will be the ability to "localize" the cross-covariances between parameters and observations using expert knowledge and/or physical constraints (e.g. Chen and Oliver (2017))—this is an active research direction.

The availability of a model-independent iterative ensemble smoother provides a platform for practitioners to apply this important and emerging technology to a wide-range of environmental modeling analyses and allows practitioners to account for model input uncertainty at more realistic spatial and temporal scales. This direct and robust accounting is expected to lead to better forecast uncertainty estimates that will improve the use of environmental models in decision making.

Acknowledgments

This work was funded in part by Waikato Regional Council. I would like to acknowledge several colleagues who I have discussed iterative ensemble smoothers with, including John Doherty, Mike Fienen, Randy Hunt, and Matt Knowling. I would also like to acknowledge Brioch

Hemmings and Zara Rawlinson for pulling together the initial Hauraki Plains model input datasets, and John Hadfeld and Beavan Jenkins at Waikato Regional Council for providing several datasets used in construction and history matching of the Hauraki Plains model.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envsoft.2018.06.009.

References

Bailey, R., Ba, D., 2010. Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution. Water Resour. Res. 46 (12), W12543 n/a–n/a.
Bakker, M., Post, V., Langevin, C.D., Hughes, J.D., White, J., Starn, J., Fienen, M.N., 2016. Scripting modflow model development using python and flopy. Groundwater 54 (5), 733–739.
Bedekar, V., Morway, E., Langevin, C., Tonkin, M., 2016. MT3D-USGS version 1: a U.S. Geological Survey release of MT3DMS updated with new and expanded transport capabilities for use with MODFLOW. In: U.S. Geological Survey Techniques and Methods 6-A53, pp. 69.
Bocquet, M., Sakov, P., 2013. Joint state and parameter estimation with an iterative ensemble Kalman smoother. Nonlinear Process Geophys. 20 (5), 803–818.
Bocquet, M., Sakov, P., 2014. An iterative ensemble Kalman smoother. Q. J. R. Meteorol. Soc. 140 (682), 1521–1535.
Chen, Y., Oliver, D.S., 2012. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math. Geosci. 44 (1), 1–26.
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.
Chen, Y., Oliver, D.S., 2017. Localization and regularization for iterative ensemble smoothers. Comput. Geosci. 21 (1), 13–30.
Cooley, R.L., Christensen, S., 2006. Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media. Adv. Water Resour. 29 (5), 639–656.
Crestani, E., Camporese, M., Baú, D., Salandin, P., 2013. Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. Hydrol. Earth Syst. Sci. 17 (4), 1517–1531.
Dausman, A., Doherty, J., Langevin, C., Sukop, M., 2010. Quantifying data worth toward reducing predictive uncertainty. Ground Water 48 (5), 729–740.
Doherty, J.E., 2015. PEST and its Utility Support Software, Theory. Watermark Numerical Publishing.
Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resour. Res. 47 (12).
Doherty, J., Welter, D., 2010. A short exploration of structural noise. Water Resour. Res. 46 (W05525).
Emerick, A.A., Reynolds, A.C., 2013. Ensemble smoother with multiple data assimilation. Comput. Geosci. 55, 3–15.
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.
Freyberg, D.L., 1988. An exercise in ground-water model calibration and prediction. Ground Water 26 (3), 350–360.
Guennebaud, G., Jacob, B., et al., 2010. Eigen V3. http://eigen.tuxfamily.org.
Hanke, M., 1997. A regularizing levenberg-marquardt scheme, with applications to inverse groundwater filtration problems. Inverse Probl. 13 (1), 79.
Harbaugh, A.W., 2005. MODFLOW-2005, the U.S. Geological Survey Modular Groundwater Model – the Ground-water Flow Process, vol. 6.
Knowling, M.J., Werner, A.D., 2016. Estimability of recharge through groundwater model calibration: insights from a field-scale steady-state example. J. Hydrol. 540, 973–987.
Ledgard, S.F., Williams, P.H., Broom, F.D., Thorrold, B.S., Wheeler, D.M., Willis, V.J., 1999. Overseer Tm a Nutrient Budgeting Model for Pastoral Farming, Wheat, Potatoes, Apples and Kiwifruit. in: Best Soil Management Practices for Production. Technical report. Massey University, Palmerston North.
McBride, G., Reeve, G., Pritchard, M., Lundquist, C., Daigneault, A., Bell, R., Blackett, P., Swales, A., Wadhwa, S., Tait, A., Zammit, C., 2016. Climate Changes, Impacts and Implications for New Zealand to 2100 Synthesis Report: RA2 Coastal Case Study the Firth of Thames and Lower Waihou River.
Moore, C., Doherty, J.E., 2005. Role of the calibration process in reducing model predictive error. Water Resour. Res. 41 (5), 1–14.
Moore, C., Doherty, J.E., 2006. The cost of uniqueness in groundwater model calibration. Adv. Water Resour. 29 (4), 605–623.
Niswonger, R., Prudic, D., 2005. Documentation of the Streamflow-Routing (SFR2) Package to include unsaturated flow beneath streamsA modification to SFR1. In: U.S. Geological Survey Techniques and Methods 6-A13, pp. 50.
Niswonger, R., Panday, S., Ibaraki, M., 2011. MODFLOW-NWT, A Newton formulation for MODFLOW-2005. In: U.S. Geological Survey Techniques and Methods 6-A37, pp. 44.
Oliver, D., Reynolds, A., Liu, N., 2008. Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press.
Poeter, E., Hill, M., Lu, D., Tiedeman, C., Mehl, S., 2014. UCODE_2014, with New Capabilities ToDefine Parameters Unique to Predictions, Calculate Weights Using Simulated Values, Estimate Parameters with SVD, Evaluate Uncertainty with MCMC, and More. Number GWMI 2014-02. Integrated Groundw ater Modeling Center.
Pollock, D.W., 2012. User Guide for MODPATH Version 6A Particle-tracking Model for MODFLOW: U.S. Geological Survey Techniques and Methods. U.S. Dept. of the Interior, U.S. Geological Survey Reston, Va version 6. edition.
Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM.
Tikhonov, A.N., Arsenin, V.Y., 1977. In: Winston & Sons, V.H. (Ed.), Solutions of Ill-posed Problems. John Wiley & Sons, New York.,, , Washington, D.C..
Tonkin, M.J., Doherty, J., 2005. A hybrid regularized inversion methodology for highly parameterized environmental models. Water Resour. Res. 41 (10).
Van Leeuwen, P.J., Evensen, G., 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. Mon. Weather Rev. 124 (12), 2898–2913.
Welter, D.E., White, J.T., Doherty, J.E., Hunt, R.J., 2015. PEST + + version 3, a parameter estimation and uncertainty analysis software suite optimized for large environmental models. In: U.S. Geological Survey Techniques and Methods Report, 7–C12, pp. 54.
White, J.T., 2017. Forecast first: an argument for groundwater modeling in reverse. Groundwater 55 (5), 660–664.
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, P., Raiber, M., Della Pasqua, F., 2015. Geological Model of the Hauraki Plains: Geological Model and Groundwater Budgets (GNS Science consultancy report). .
White, J.T., Fienen, M.N., Doherty, J.E., 2016. A python framework for environmental model uncertainty analysis. Environ. Model. Software 85, 217–228.
White, J., Stengel, V., Rendon, S., Banta, J., 2017. The importance of parameterization when simulating the hydrologic response of vegetative land-cover change. Hydrol. Earth Syst. Sci. 21 (8), 3975–3989.
White, J., Fienen, M., Barlow, P., E Welter, D., 2018. A tool for efficient, model-independent management optimization under uncertainty. Environ. Model. Software 100, 213–221.
Wilson, S., Close, M., Abraham, P., 2018. Applying linear discriminant analysis to predict groundwater redox conditions conducive to denitrification. J. Hydrol. 556, 611–624.