

Article

Flow Regime-Dependent, Discharge Uncertainty Envelope for Uncertainty Analysis with Ensemble Methods

Nick Martin ^{1,*}  and Jeremy White ² ¹ Southwest Research Institute, San Antonio, TX 78238, USA² INTERA Geosciences Pty Ltd., Perth, WA 6000, Australia

* Correspondence: nick.martin@alumni.stanford.edu

Abstract: A discharge uncertainty envelope is presented that provides an observation error model for data assimilation (DA) using discharge observations derived from measurement of stage using a rating curve. It uniquely represents the rating curve representation error, which is due to scale and process incompatibility between the rating curve hydrodynamic model and “true” discharge, within the observation error model. Ensemble methods, specifically, the iterative ensemble smoother (IES) algorithms in PEST++, provide the DA framework for this observation error model. The purpose of the uncertainty envelope is to describe prior observation uncertainty for ensemble methods of DA. Envelope implementation goals are (1) limiting the spread of the envelope to avoid conditioning to extreme parameter values and producing posterior parameter distributions with increased variance, and (2) incorporating a representative degree of observation uncertainty to avoid overfitting, which will introduce bias into posterior parameter estimates and predicted model outcomes. The expected uncertainty envelope is flow regime dependent and is delineated using stochastic, statistical methods before undertaking history matching with IES. Analysis of the goodness-of-fit between stochastically estimated “true” discharge and observed discharge provides criteria for the selection of best-fit parameter ensembles from IES results.

Keywords: data assimilation; ensemble methods; uncertainty analysis; PEST++; bias–variance trade-off; observation error model



Citation: Martin, N.; White, J. Flow Regime-Dependent, Discharge Uncertainty Envelope for Uncertainty Analysis with Ensemble Methods. *Water* **2023**, *15*, 1133. <https://doi.org/10.3390/w15061133>

Academic Editors: Achim A. Beylich and Aizhong Ye

Received: 4 February 2023

Revised: 2 March 2023

Accepted: 13 March 2023

Published: 15 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Environmental models are constructed to test hypotheses and make predictions in support of resource management. Simulation results are typically accompanied by significant uncertainty that must be accounted for during decision making and resource allocation. This inherent uncertainty has many sources, but herein we focus on two widely recognized sources of uncertainty: parameter variability and historical observations of system state.

Parameters are model input quantities, which are specified or predefined during model development. Parameter valuation uncertainty is addressed through the description of parameter value likelihood with probability distributions. Observations are measured quantities. From the environmental model construction perspective, observations provide “target” values used to assess model skill and capability with the goal of producing a model that simulates or predicts values similar to observed values, i.e., history matching, when the model is driven by parameter values that describe the current system state. The inherent underlying assumption promoting environmental model construction is that a model, with demonstrated history matching skill, will provide predictive capability when “new” forcing is applied that represents an unobserved scenario used to guide resource management decision making.

Data assimilation (DA) is a collection of methods and tools for the optimal combination of information from numerical model simulations with observations to obtain the “best” description of a dynamical system and the uncertainty contained within the optimal

system description. It inherently accounts for parameter and observation uncertainty. In DA, the numerical model provides a forecast which is adjusted to account for or better represent observations.

DA, as an umbrella categorization, covers a variety of methods and techniques that are derived from Bayes' theorem [1]. Contained under the DA umbrella are inverse approaches to environmental model calibration, which seek to vary input parameter values to provide the best fit between simulated and observed values, including "calibration-constrained uncertainty analyses" provided by the PEST suite of utilities [2–4].

Bayes' theorem, see Equation (1), provides the fundamental starting point shared by all DA methodologies [1]. Equation (1) quantifies the model parameter uncertainty, where k represents the model parameters, observations or targets used in calibration are h , $P()$ signifies a probability distribution, $P(k)$ is the prior parameter probability distribution, $P(h|k)$ is the likelihood function, and $P(k|h)$ is the posterior parameter probability distribution. The posterior parameter probability distribution $P(k|h)$ is the probability distribution of model parameters conditioned on observations [3]:

$$P(k|h) = P(h|k)P(k) \quad (1)$$

Modern Bayesian inverse techniques for hydrologic model calibration are an example of the use of DA methods in hydrology. The goal of the inversion approach is the selection of parameter values to generate the best fit between simulated values and observations, where the best-fit is quantitatively evaluated using goodness-of-fit metrics. In the inverse problem context, the observed values are "target" values. Assuming that the initial ranges of parameter values in the inverse problem are constrained by professional knowledge, the inverse problem approach to environmental model formulation seeks to adjust parameters in ways that are in harmony with expert judgment and improve history matching. This yields an approximation to the posterior parameter ensemble in a Bayesian sense, $P(k|h)$ in Equation (1), and explains why "calibration-constrained uncertainty analysis [4]" is a form of DA.

In DA implementation, observations, k in Equation (1), include an observation error model that contains "measurement error", or "observation error" that is propagated through the assimilation to the posterior parameter probability distribution, $P(k|h)$. "Observation error" always includes instrument error and may include representation error, which accounts for different representations by the measurements and the numerical model [1]. Representation errors in numerical weather prediction and oceanographic implementations are typically errors due to scales and physical processes that are unresolved by either the numerical model or the observations [1,5]. Hereafter, representation error attributed to scale or process incompatibility between the numerical forecast model and observations is labeled "numerical representation error".

This paper presents a discharge uncertainty envelope designed to provide an observation error model in DA implementations that use river discharge observations calculated via a rating curve from an observed stage. For this type of observation, the rating curve provides an additional source of representation error from scale and process incompatibility between the rating curve hydrodynamic model and "true" discharge. Hereafter, this form of representation error is termed "rating curve representation error". The goal of the discharge uncertainty envelope is to optimize the bias–variance trade-off impacts to the posterior parameter probability distribution, $P(k|h)$ in Equation (1). Figure 1 graphically explains the bias–variance trade-off and the contribution of the discharge uncertainty envelope to a robust DA implementation.

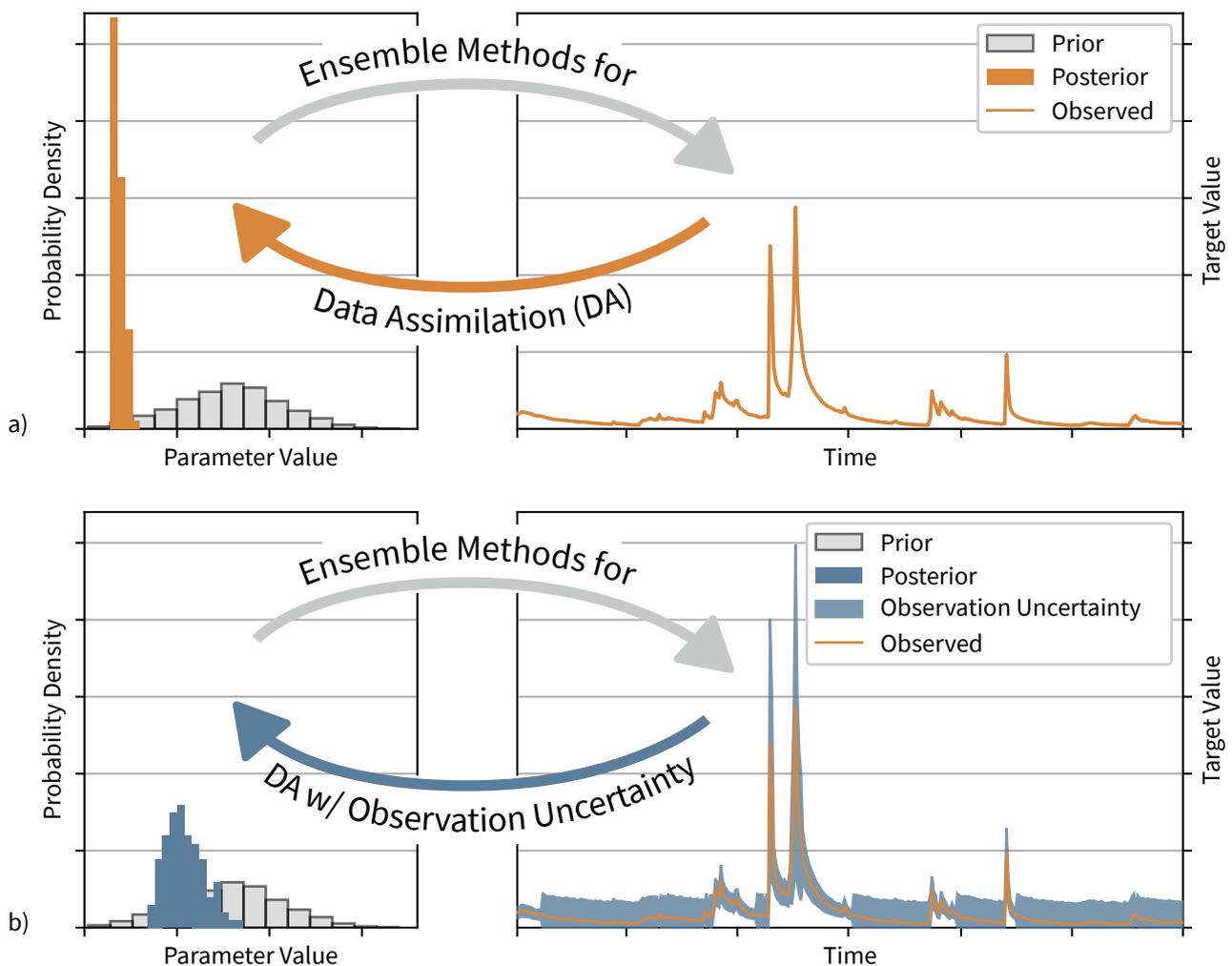


Figure 1. Conceptual schematic of observation error models in data assimilation (DA) and optimizing the bias–variance trade-off. The prior parameter distribution, “Prior”, is estimated using professional judgment with the goal of uncovering a posterior parameter distribution, “Posterior”, that produces numerical model forecasts, which are congruent with observations and that has minimum variance and bias. Panel (a) represents the case of no observation error model in DA, where bias is a concern. Here, the range of target values for each history matching interval is narrower than the expected measurement and representation error, causing the inverse solution to overfit to a narrow range of relatively unlikely parameter values as shown on the left-side of panel (a). Variance denotes the spread of parameter values in the posterior. The goal of DA is to use observations to constrain, or narrow, the posterior relative to the prior. In panel (b), the introduction of observation uncertainty and the use of an observation error model promote a posterior with an expected value within the range of relatively likely prior values from professional judgment and with reduced variance, and thus reduced uncertainty, relative to the prior. Note that a single parameter and target are shown for illustrative purposes. Ensemble methods for DA work with hundreds to millions of parameters and hundreds to thousands of targets simultaneously.

For implementation, the uncertainty envelope requires a DA framework that provides for the propagation of prior observation uncertainty through the analysis to posterior parameter uncertainty and for the adjustment of constraint on possible posterior parameter values to reflect the prior observation uncertainty. Ensemble methods, which are a type of DA implementation, provide an explicit and natural approach to accommodate a wide range of observation uncertainty representations. Consequently, the discharge uncertainty envelope is specifically designed for use with the iterative ensemble smoother

(iES) algorithm [6,7] within the PEST++ suite of DA tools, hereafter PESTPP-IES, as the data assimilation engine because PESTPP-IES prior-data conflict capabilities address numerical representation error [3,7]. The discharge uncertainty envelope addresses measurement error and rating curve representation error, and the combination of PESTPP-IES and the uncertainty envelope generates a DA implementation with a complete observation error model accounting for measurement, numerical representation, and rating curve representation error.

2. Data and Methods

The primary data sets for this study are observations of river discharge. These data are analyzed using flow duration curves (FDCs). A stochastic simulation approach is then employed to generate the expected uncertainty envelope.

2.1. Discharge Observations

Data sets for this study are continuous, daily discharge observations from a collection of stream gauging stations in central Texas (TX). Figure 2 displays study site location and gauging station configuration across the study area. Table 1 provides descriptive information for the 16 gauging stations shown on Figure 2. These data sets are publicly available online [8,9].

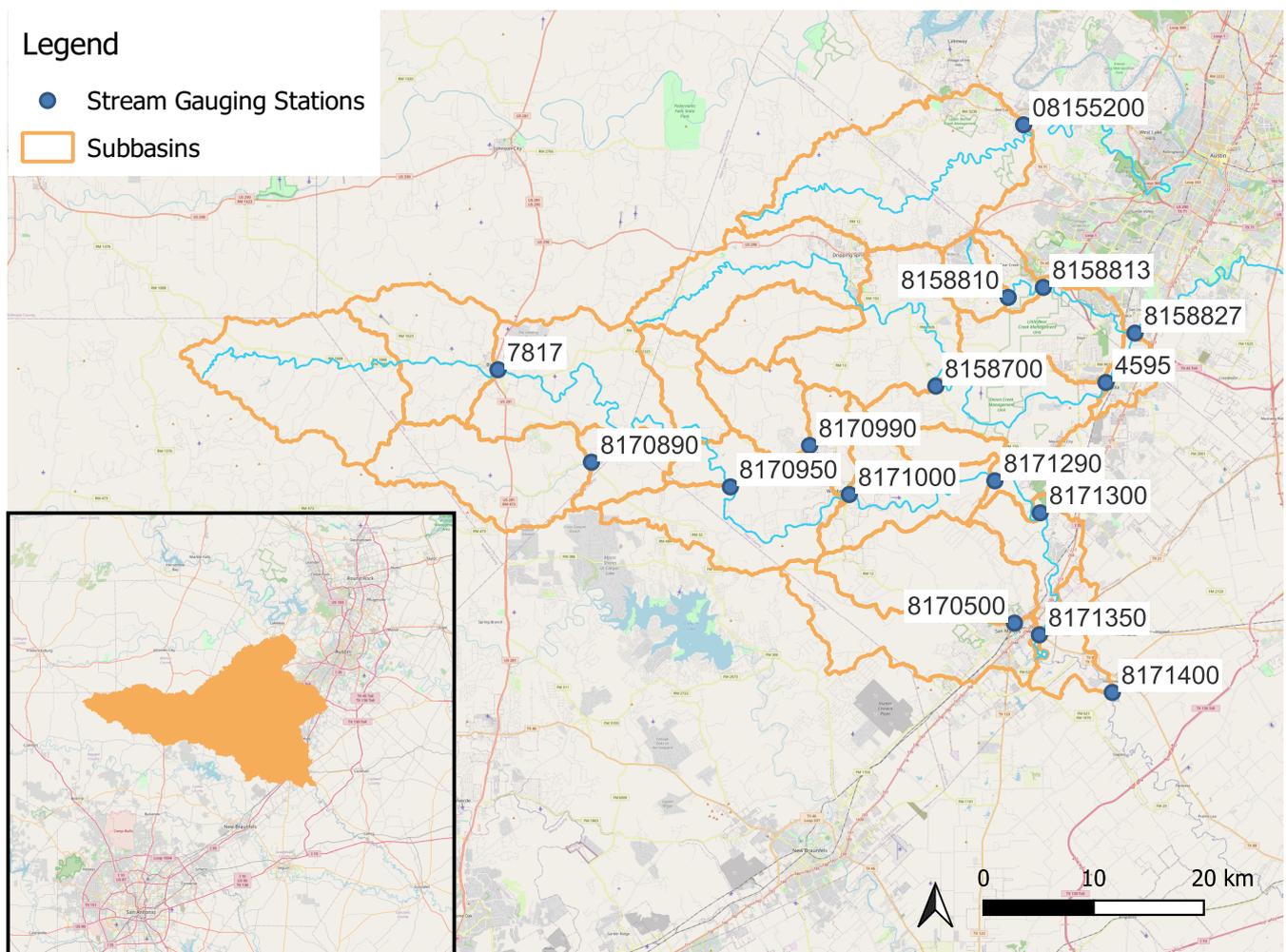


Figure 2. Study site location, stream gauging station configuration, and sub-basins. The study site is between San Antonio and Austin in south-central Texas (TX). 16 stream gauging stations are shown on the Blanco River, San Marcos River, Onion Creek, Bear Creek, and Barton Creek.

Data sets were acquired as daily observations. Daily discharge units are listed as cubic meters per second, days ($\text{m}^3/\text{s-days}$). In some situations in this analysis, the monthly averaged discharge is required. The monthly averaged discharge is the average across the calendar month of the daily discharge observations and is presented with units of cubic meters per second, months ($\text{m}^3/\text{s-months}$).

Several of these stream-gauging stations provide unusual data sets whose records show a sharp decrease in discharge moving downstream for medium to low flows and identify that there are strong, spatially localized “losing” conditions. Discharge decreases in these locations because of communication with the Balcones Fault Zone (BFZ) Edwards Aquifer, which provides a significant source of water for the city of San Antonio and the San Antonio–Austin corridor. A subset of the stream gauging stations in this study is used to estimate recharge for the BFZ Edwards Aquifer [10,11].

All of the stations in the study use a water stage recorder method of discharge calculation. Stage is the normal depth of the water at the measurement location. A stage–discharge rating curve is developed for the site and used to calculate or estimate discharge given the observed stage. The rating curve is traditionally developed using discrete, concurrent measurements of stage and discharge [12]. A combination of a water stage measurement with a rating curve is often used to estimate discharge because discharge is difficult to measure continuously while multiple methods for continuous measurement of stage are available [13].

Multiple sources of uncertainty affect the rating curve-based stream discharge calculated from a stage measurement. However, there are two primary uncertainty components: (1) measurement error in the concurrent stage and discharge measurements employed to derive the rating curve and (2) model error from the limitations of the stage to discharge transformation provided by the rating curve. Model error includes the necessity of extrapolation to stages and discharges that are both higher and lower than those available in the underlying collection of concurrent stage and discharge measurements and includes theoretical, physics-based limitations on the correlation between normal depth and discharge [13,14].

For analysis of discharge data sets in this study, three flow regime-based discharge uncertainty estimates are used that are provided by [14] as typical values: (1) ± 50 –100% for low flows, (2) ± 10 –20% for medium or high (in-bank) flows, and (3) ± 40 % for out of bank flows. These three flow regimes explicitly represent the flow regime-dependent uncertainty associated with the stage–discharge estimation approaches.

2.2. Estimation of Flow Regime from Flow Duration Curves (FDCs)

A FDC depicts the relationship between magnitude and frequency of discharge for a particular river basin and provides an estimate of the amount of time that a given discharge has been matched or exceeded during the historical record. The FDC is the complement of the cumulative distribution function (CDF) of stream discharge for the selected analysis interval (e.g., daily, monthly, or annual). It provides a comprehensive, graphical description of historical variability in the stream discharge [15].

In this study, daily FDCs are used as the basis for the identification of the flow regime. Ref. [16] provides a review of the low flow hydrology, summarizes low flow measures and indices, and describes methods for index estimation or calculation from discharge time series. Discharges within the range of 70–99% time exceedance, as extracted from the FDC, are typically used as design low flows; these flows are important for in-stream ecological support and are of primary interest for water managers. The mean annual runoff (MAR) provides an upper bound or threshold to the low flow regime; MAR is the mean of the observed annual total discharge. The long-term mean daily discharge (MDF) is obtained by dividing MAR by the number of seconds in a year [16]. For this study site, MDF is used to delineate the low flow boundary because many of the discharge data sets have zero discharge at the 70% time exceedance threshold in the daily FDC.

Table 1. Stream gauging station metadata.

Gauge Id	Agency	River/Stream	Description	Watershed Area (km ²)	Gauge Type	WY ⁴ 2021 Data Quality
8155200	USGS ¹	Barton Creek	Barton Ck at SH 71 nr Oak Hill, TX	232.3	water stage recorder	good
8158700	USGS ¹	Onion Creek	Onion Ck nr Driftwood, TX	321.2	water stage recorder	fair
8158810	USGS ¹	Bear Creek	Bear Creek below Farm to Market Road 1826 near Driftwood, TX	31.6	water stage recorder	fair
8158813	USGS ¹	Bear Creek	Bear Creek at Spillar Ranch Road nr Manchaca, TX	52.8	water stage recorder	fair
8158827	USGS ¹	Onion Creek	Onion Creek at Twin Creeks Road near Manchaca, TX	468.8	water stage recorder	fair
8170500	USGS ¹	San Marcos River	San Marcos Rv at San Marcos, TX (San Marcos Springs)	126.7	water stage and water velocity recorders	good
8170890	USGS ¹	Little Blanco River	Little Blanco Rv at FM 32 nr Fischer, TX	130.0	water stage recorder	fair
8170950	USGS ¹	Blanco River	Blanco Rv at Fischer Store Rd nr Fischer, TX	696.7	water stage recorder	fair
8170990	USGS ¹	Cypress Creek	Jacobs Well Spg nr Wimberley, TX	NA	water stage and water velocity recorders	poor
8171000	USGS ¹	Blanco River	Blanco Rv at Wimberley, TX	919.4	water stage recorder and crest stage gauge	fair
8171290	USGS ¹	Blanco River	Blanco Rv at Halifax Rch nr Kyle, TX	1012.7	water stage recorder	fair
8171300	USGS ¹	Blanco River	Blanco Rv nr Kyle, TX	1067.1	water stage recorder	fair
8171350	USGS ¹	Blanco River	Blanco Rv at San Marcos, TX	1129.2	water stage recorder	fair
8171400	USGS ¹	San Marcos River	San Marcos Rv nr Martindale, TX	1416.7	water stage recorder	fair
7817	LCRA ²	Blanco River	Blanco River at Blanco, BRBT2	283.6	water stage recorder	poor ³
4595	LCRA ²	Onion Creek	Onion Creek at Buda, BDUT2	431.8	water stage recorder	poor ³

¹ USGS is the United States Geological Survey [9]. <https://waterdata.usgs.gov/tx/nwis/rt> (accessed on 13 September 2022), ² LCRA is the Lower Colorado River Authority [8]. <https://hydromet.lcra.org/> (accessed on 13 September 2022), ³ LCRA station data are demarcated as poor because they do not publish water year summary reports, ⁴ WY stands for water year.

Ref. [17] presents a process-based diagnostic approach to model evaluation. As part of signature diagnostic measures related to vertical soil moisture redistribution, they partition an FDC into three different segments: (1) high-flow segment demarcated by 0–2% flow exceedance probabilities on the FDC; (2) mid-segment with 2–70% flow exceedance probabilities; and (3) low-flow segment with the typical 70–100% flow exceedance probabilities used for design bases. In this study, we assume that 0.02 (or 2%) flow exceedance probability provides the threshold between out-of-bank high flows and in-bank high flows.

2.3. Stochastic Demarcation of Discharge Uncertainty Envelope

A Monte Carlo model is employed to generate realizations of “synthetic” discharge from the time series provided for each gauging station. The synthetic time series incorporates flow regime-dependent uncertainty to provide realizations of what the actual stream discharge could be given expectations for error in the calculated discharge record from the gauging station. Flow regime dependent uncertainty of: (1) ± 50 –100% for low flows, (2) ± 10 –20% for medium or high (in-bank) flows, and (3) ± 40 % for out of bank flows [14] is used to derive synthetic discharge. To implement the flow regime delineation, two discharge thresholds are required: (1) out-of-bank high flow and (2) low flow. The 2% exceedance threshold on the FDC is used for the out-of-bank threshold, and the MDF is used for the low flow threshold as discussed in Section 2.2.

Following [18], the synthetic discharge time series, x_i , is divided into a deterministic part, d_i , and a random component, e_i , of the generating scheme shown in Equation (2), where the subscript i denotes the interval of the time series:

$$x_i = d_i + e_i \quad (2)$$

The deterministic portions, d_i , are the discharge observations from the gauging station. The random components, e_i , are estimated stochastically from the flow regime for each d_i and expected error percentage corresponding to the flow regime, P_f , as shown in Equation (3). P_f values are selected from ± 50 –100% for low flows, ± 10 –20% for medium or high (in-bank) flows, and ± 40 % for out-of-bank flows [14]. The largest value in the error range is used for P_f ; consequently, 100% is used for low flows and 20% for medium flows. d_i values below the low flow threshold were also adjusted to the low flow threshold discharge for use in Equation (3). Unaltered d_i values are used in Equation (3) for medium and high flow regimes:

$$e_i = \zeta_i \times (P_f d_i) \quad (3)$$

The stochastic part of the implementation comes from the random sampling variate, ζ_i , which is sampled for each time interval from either a standard normal distribution ($\mu = 0.0$ and $\sigma = 1.0$) representing unbiased measurement error or a shifted normal distribution ($\mu = 1.0$ and $\sigma = \sqrt{2}$) representing measurement and rating curve representation error. μ is the population mean, and σ is the population standard deviation. Rating curve representation error denotes the difference between the discharge estimated from the rating-curve model and “true” discharge.

In the application of Equation (3), the form of the normal distribution for the random sampling variate is selected based on the water year (WY) 2021 data quality assessment, which is listed in Table 1. If the gauging station was assigned a data quality assessment of “fair” or “poor”, the biased measurement and representation error descriptor is used for that station. The unbiased measurement error descriptor is used for stations with data quality assessed as “good”.

One thousand synthetic discharge time series are calculated with Equation (2) for each gauge. Each of these time series is a Monte Carlo realization derived from the application of Equations (2) and (3). The expected uncertainty envelope for each time series interval, i , and each gauging station is estimated using the interval root mean square error (RMSE) between synthetic realizations, $x_{i,r}$, and the actual gauge data set, d_i , using Equation (4). r

represents the realization index, which goes from $1, \dots, R$ where $R = 1000$ or the number of realizations. $RMSE_i$ provides a standard error estimate for each history matching time, i , in the dynamic DA analysis:

$$RMSE_i = \sqrt{\frac{\sum_{r=1}^R (x_{i,r} - d_i)^2}{R}} \quad (4)$$

Goodness-of-Fit Metrics

Three goodness-of-fit metrics are used to compare the synthetic time series to the gauge record: (1) Nash–Sutcliffe efficiency (NSE) [19], (2) Kling–Gupta efficiency (KGE) [20], and (3) sum of NSE and KGE (Σ_{NK}). Metrics are calculated for each synthetic time-series realization, r . The NSE, defined in Equation (5), overcomes two common issues with correlation-based, goodness-of-fit metrics, which are sensitivity to extreme values (outliers) and lack of sensitivity to additive and proportional differences between model predictions and observations [19]:

$$NSE_r = 1.0 - \frac{\sum_{i=1}^N (x_{i,r} - d_i)^2}{\sum_{i=1}^N (x_{i,r} - \bar{x}_r)^2} \quad (5)$$

The KGE metric, see Equation (6), was developed through the decomposition of the NSE into the linear correlation coefficient between observed and simulated values, ρ , a measure of relative variability in the simulated and observed values, α , and a bias component, β [20]. Both NSE and KGE range from $-\infty$ to 1.0 with 1.0 representing the “perfect” match:

$$KGE_r = 1 - \sqrt{(\rho_r - 1)^2 + (\alpha_r - 1)^2 + (\beta_r - 1)^2} \quad (6)$$

$$\rho_r = \frac{\sum_{i=1}^N ((x_{i,r} - \bar{x}_r)(d_i - \bar{d})) / N}{\sigma_d \sigma_{x_r}} \quad (7)$$

$$\alpha_r = \frac{\sigma_d}{\sigma_{x_r}} \quad (8)$$

$$\beta_r = \frac{\mu_d}{\mu_{x_r}} \quad (9)$$

The sum of NSE and KGE (Σ_{NK}), see Equation (10), ranges from $-\infty$ to 2.0 with 2.0 representing the “perfect” match. Σ_{NK} provides a simple means to leverage the value of NSE and KGE in a single metric:

$$\Sigma_{NK,r} = NSE_r + KGE_r \quad (10)$$

2.4. Uncertainty Analysis with Ensemble Methods

An expected uncertainty envelope for stream gauge discharge data sets is used to derive an observation error model that portrays measurement error and rating curve representation error. This enhanced observation error model is only useful in history matching (i.e., DA) approaches that propagate observation uncertainty into the posterior parameter probability distribution, $P(k|h)$ in Equation (1).

The approach described in this paper is designed specifically for use with the PEST++ Iterative Ensemble Smoother (IES) toolset, PESTPP-IES [3,7,21]. PESTPP-IES simultaneously adjusts an ensemble of parameter realizations and uses an ensemble of residuals to seek an approximate posterior parameter ensemble. Residuals are the differences between observations, or targets, and simulated values. PESTPP-IES was selected for the DA framework because it provides for explicit incorporation of observation error models into the calibration through specification of a standard error estimate for each history-matching

time, it makes no assumptions regarding model linearity in the sampling of the posterior parameter probability distribution and is thus better suited to DA using watershed and surface water numerical models than the other tools in the PEST++ suite, and it (and ensemble methods in general) is relatively efficient, computationally, with large numbers of input parameters [7].

PEST++ tools, including PESTPP-IES, seek to minimize an objective function (Φ) to identify the the posterior parameter distribution that produces best fit, history matching. Φ is the sum of squared, weighted residuals. Because PESTPP-IES works with ensembles of parameters, it produces ensembles of simulated values and residuals. The user must then select the posterior parameter probability distributions, $P(k|h)$ in Equation (1), by choosing one or more parameter ensembles that provide best-fit or equally best-fit simulated values. A single ensemble of parameters could be selected that produces the minimum Φ value. However, parameter and observation uncertainty is typically significant, and a collection of parameter ensembles will often provide the appropriate description of posterior parameter uncertainty. When selecting a collection of best-fit parameter ensembles, a Φ threshold can be used to identify collection members; in this case, parameter ensembles that produce Φ values less than or equal to the threshold are the collection members [3,7,21].

PESTPP-IES provides a “base realization” for conceptual checking and evaluation of the calibration process. The “base realization” uses parameters directly from the input specifications, and not subspace parameters obtained via singular value decomposition (SVD), and pairs these parameters with unadjusted observed values for targets. The “base realization” can also be employed to facilitate the selection of a Φ threshold [3,7,21].

Variability between parameter fields that sample the posterior parameter probability distribution, $P(k|h)$ in Equation (1), comes from three sources. The first is prior parameter uncertainty, described with $P(k)$. The second is insufficient information within the observed data set of important outcomes to significantly reduce parameter uncertainty. In other words, available data may be insufficient to constrain the range of feasible parameter values. The third source comes from observation error [3].

The concern with discharge data sets derived from stage observations is that these calculated values are contaminated by measurement error, derived from imperfect stage–discharge estimation, and also representation error, derived from using a rating curve which provides an imperfect hydrodynamics model. Rating curve representation error is an extra representation error component for calibration within a DA framework in addition to that expected from numerical representation error.

PESTPP-IES employs the realizations of random parameter fields and also realizations of additive observation uncertainty (i.e., added to observed values) to explore posterior parameter uncertainty and thus the predictive uncertainty of the model, arising from both parameter uncertainty and observation uncertainty. Using observation uncertainty (or the observation error model), PESTPP-IES creates a slightly different “target data set” for use in each adjustment of each random parameter field during model calibration. The altered target data set is created through the addition of a realization of observation uncertainties to observed values [3]. The generation of the targets plus additive uncertainties realizations is simplified if the shape of the error distribution that generates the uncertainties is known; consequently, most DA implementations use Gaussian additive observation model errors [1,3], which results in a formulation similar to Equation (3), except with a standard deviation value times the standard normal variate. In this study, the $RMSE_i$, from Equation (4) in Section 2.3, is provided to PESTPP-IES as this standard deviation value for each history matching time, i .

These target realizations, observed discharge values combined with additive observation uncertainty, allow PESTPP-IES to identify and examine prior-data conflict [22–24]. Prior-data conflict occurs where simulated targets, produced from the prior parameter ensemble, do not agree in a statistical sense with the observed values plus observation uncertainty. Agreement is measured, within PESTPP-IES, by the statistical distance between the ensemble of simulated outputs and the ensemble of target realizations. Lack of

agreement implies that extreme parameter values or extreme parameter combinations will be needed to reproduce the conflicted observation values. In this case, extreme parameter estimates can also be referred to as “biased” estimates; the continuation of parameter adjustments in the presence of prior-data conflict will generate parameter bias and forecast bias [3]. The identification and elimination of prior-data conflict provide adjustments that address representation error related to the numerical model’s ability to simulate the observations. When prior-data conflicts are removed, numerical representation error is not propagated into posterior parameter uncertainty.

2.5. Numerical Model Employed for Data Assimilation (DA)

DA seeks to optimally combine information from numerical models with observations. The expected discharge uncertainty envelope is designed to be used as the observation error model within the PESTPP-IES, DA framework. The hydrological model used with PESTPP-IES in the case study implementation, illustrating the use and advantages of the observation uncertainty envelope, is the Hydrological Simulation Program FORTRAN (HSPF) [25,26].

HSPF is a collection of routines that can simulate the hydrologic, and associated water quality, processes on pervious and impervious land surfaces and in streams and well-mixed impoundments. It provides a comprehensive package for the simulation of watershed hydrology and surface water-related considerations at the watershed scale. HSPF was originally developed as the Stanford Watershed Model in the 1960s [25].

Conceptually, HSPF utilizes the constructs of hydrologic response units (HRUs) and well-mixed reservoirs to create a network of ordinary differential equations (ODEs) and empirical relationships. This network is an example of hydrologic routing [27] and can be solved in watershed routing order, i.e., upstream to downstream because the required discharge inflow information is obtained from only the upstream routing unit. Physical process representation in HSPF is much less advanced and accurate numerically relative to the models typically used in weather and oceanographic DA implementations. Consequently, the numerical representation error for HSPF to simulate stream discharge is expected to be more significant than the numerical representation error components typically encountered in weather and oceanographic DA implementations.

HSPF is applied to represent watershed processes across the study site, identified in Figure 2. In the PESTPP-IES analysis, 948 parameters are varied, which characterize pervious and impervious watershed properties and discharge from stream segments (i.e., FTABLE parameterization). In total, 2411 targets are used in the objective function, Φ ; of these, 791 targets are monthly averaged discharge observed at the stream gauging stations listed on Table 1. The $RMSE_i$, generated as described in Section 2.3, provides the specified standard error for each of these 791 discharge targets. HSPF simulates 1 January 2015 through 31 December 2019 with a daily time step. The first six simulation months are used as a “burn-in” period and do not contribute to the objective function calculation.

3. Results

Two related sets of results were obtained: (1) observation uncertainty for discharge targets and (2) parameter uncertainty analysis accounting for parameter and observation uncertainty. Expected uncertainty envelopes provide the discharge observation error model and are developed independently of PESTPP-IES using the methods described in Section 2.3. Observation uncertainty is provided to PESTPP-IES as a specification, and PESTPP-IES considers observation uncertainty in conjunction with parameter uncertainty to constrain the range of optimal parameter values, as discussed in Section 2.4.

3.1. Observation Uncertainties Estimated with Expected Uncertainty Envelopes

FDCs provide the means to delineate flow regime for discharge observation uncertainty analysis. A FDC was created for each gauging station listed in Table 1 using daily

averaged discharge observations. Flow regime threshold results and summary statistics of the discharge data sets are presented in Table 2.

Figure 3 displays FDCs from two Blanco River gauging stations, 8171000 and 8171300. 8171000 is on the upstream side of the BFZ Edwards Aquifer Recharge Zone, and 8171300 is just downstream of the Recharge Zone. Although the watershed area of 8171300 is about 148 km² larger than that of 8171000, the 70th percentile discharge, on Figure 3 and in Table 2, for 8171300 is 0.03 m³/s-days, while the corresponding discharge for 8171000 is about 0.85 m³/s-days. Mid-range high to high flows (i.e., discharges greater than the 20th percentile discharge) are approximately equal for 8171000 and 8171300.

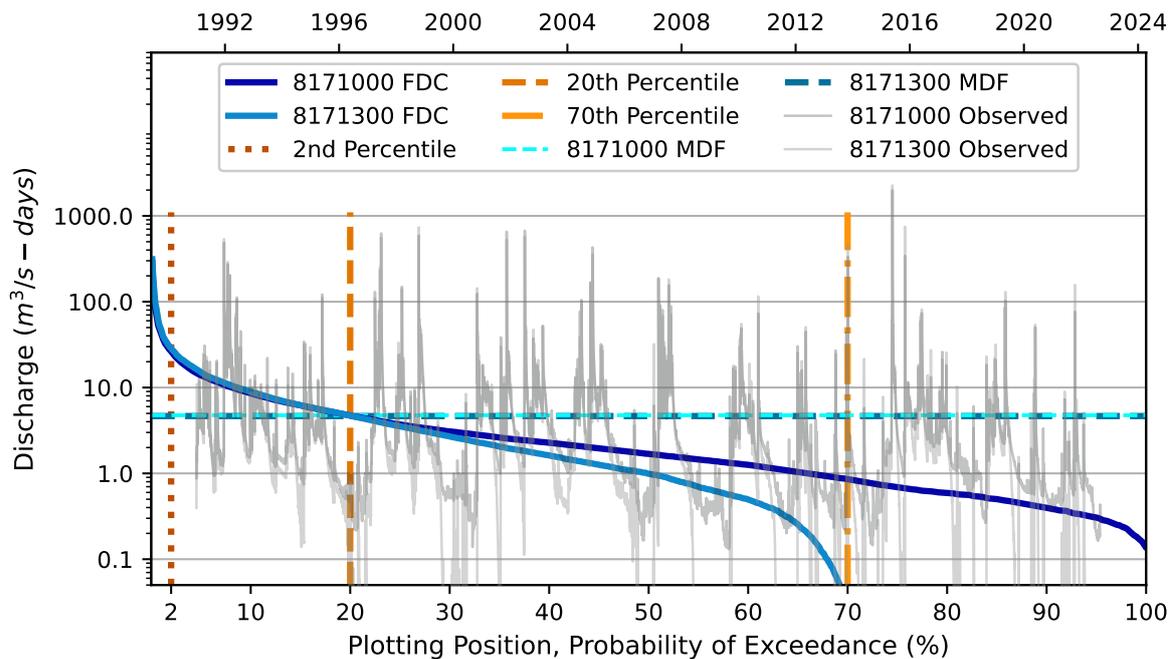


Figure 3. Flow duration curves (FDCs) for two Blanco River gauging stations on opposite ends of the BFZ Edwards Aquifer Recharge Zone: 8171300 is at the downstream edge of the Recharge Zone, and 8171000 is at the upstream edge. The low flow threshold for this study is the mean daily flow (MDF) in Table 2, and MDF for these two gauging stations is approximately equal. Note that discharge at 8171300 is generally less than discharge at 8171000 for values less than the MDF.

Figure 4 shows FDCs from two Onion Creek gauging stations, 8158700 and 4595. These two stations are on opposite sides of the BFZ Edwards Aquifer Recharge Zone. For 4595, it has an approximately 111 km² greater contributing area than 8158700, but the 70th percentile discharge, in Figure 4 and Table 2, for 4595 is zero and that for 8158700 is about 0.05 m³/s-days. The daily FDCs for 8158700 and 4595 only have similar shapes above the 2nd percentile threshold for out-of-bank high flows.

The MDF value in Table 2 provides the low flow threshold for each gauging station, and the 2nd percentile provides the out-of-bank high flow threshold. 1000 synthetic discharge time series were created from the analysis of the gauging station record using Equation (2). The $RMSE_i$ calculated from the comparison of synthetic discharge series to observed series from the gauging station provides the expected uncertainty for each time interval. The series of expected uncertainties creates the expected uncertainty envelope for each gauge that spans all time intervals, and the expected uncertainty envelope provides an estimate of dynamic observation uncertainty.

For NSE, Equation (5), KGE, Equation (6), and Σ_{NK} , Equation (10), goodness-of-fit metrics were calculated from the comparison of each synthetic discharge time series to the gauged time series. Tables 3 and 4 provide a summary of the 1000 calculated metric values (i.e., a value is calculated for each realization) for daily time series. Tables 5 and 6 provide

summary metrics calculated from comparison of monthly series. As expected, estimation of the synthetic discharge series using the biased variate produces relatively lower NSE, KGE, and Σ_{NK} values.

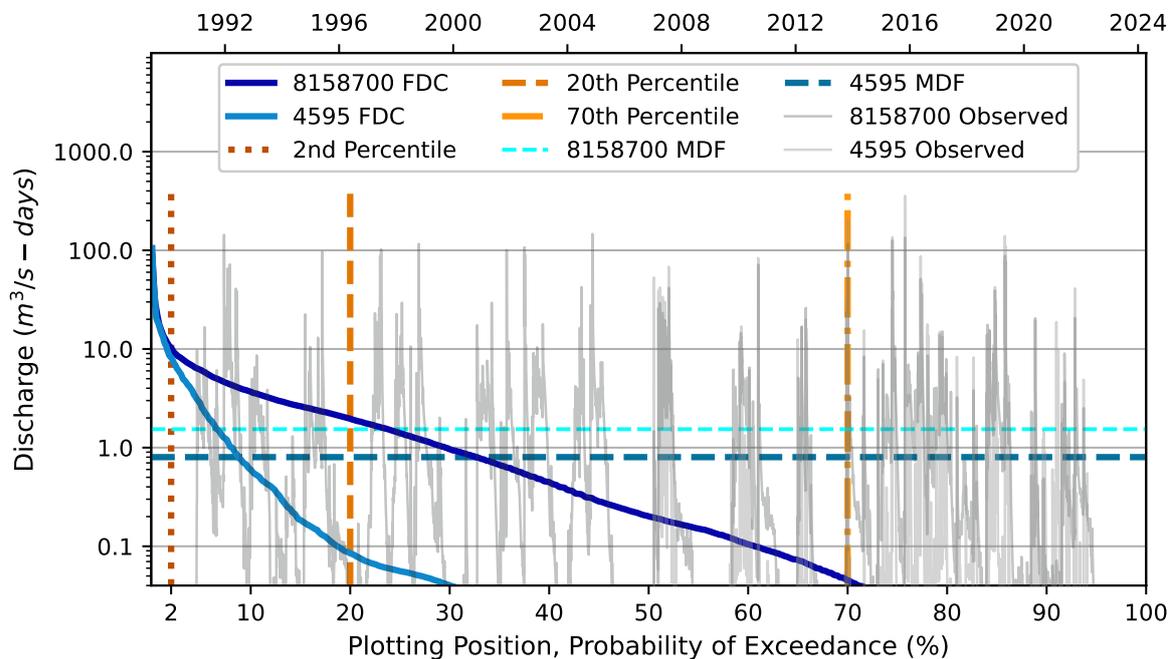


Figure 4. Flow duration curves (FDCs) for two Onion Creek gauging stations on opposite ends of the BFZ Edwards Aquifer Recharge Zone. The contributing area for 4595 is about 110 km² larger than that for 8158700. For all discharges with a greater than 2% probability of exceedance, the discharge observed at 8158700 is significantly larger than the discharge at 4595.

On Tables 3–6, the mean KGE value is generally, but not always, smaller than the mean NSE value. Because the relative relationship between KGE and NSE is not entirely consistent, Σ_{NK} provides the preferred goodness-of-fit statistic for this study because it leverages the value of both statistics.

One gauging station in Tables 3–6 provides consistently lower NSE, KGE, and Σ_{NK} values; this station is 8170500, which measures the outflow from Spring Lake in San Marcos, TX. Outflow from Spring Lake provides the starting point for the San Marcos River, and the primary source of water for Spring Lake is San Marcos Springs. Figure 5 shows the FDC for 8170500. This FDC is relatively flat between the 20th and 95th percentiles and ranges from approximately 3 to 7 m³/s-days. Consequently, there are not “low flows” in this FDC, such as Figures 3 and 4, because of the significant groundwater contribution from San Marcos Springs. In this case, MDF does not provide a robust low flow threshold because base flow is unusually large relative to the other gauging stations in the study area.

For daily time series produced with the unbiased variate, the average Σ_{NK} in Table 3 is 1.53. Average Σ_{NK} across the biased variate, mean values in Table 4 is 0.89. These averages exclude gauging station 8170500. Table 5 provides unbiased variate, goodness-of-fit statistics for monthly interval time series. Here, the average Σ_{NK} is 1.73. The average biased variate Σ_{NK} is 0.92 in Table 6. These averages exclude 8170500. The monthly interval agreement always improves relative to the daily interval agreement for unbiased variate Σ_{NK} . For the biased variate, the monthly agreement generally improves relative to the daily interval agreement. Improvement for monthly, relative to daily, values is expected because the averaging process inherent in obtaining monthly averaged values from a daily series will act to counteract the noise introduced by the unbiased random variate in Equation (2).

Table 2. Flow duration curve (FDC) analysis results.

Gauge ID	Start of Record ¹	Record Length (N)	Mean Annual Runoff (MAR)	Long Term Mean Daily Flow (MDF) ²	70th Percentile Threshold ³	20th Percentile Threshold ⁴	2nd Percentile Threshold ⁵	Standard Deviation
		Days	m ³	m ³ /s-days	m ³ /s-days	m ³ /s-days	m ³ /s-days	m ³ /s-days
8155200	1 January 1991	11,578	3.901×10^7	1.24	0.03	1.48	8.40	4.64
8158700	1 January 1991	11,578	4.854×10^7	1.54	0.05	1.96	10.3	5.29
8158810	1 January 1991	11,578	7.171×10^6	0.23	0.01	0.27	1.53	0.98
8158813	30 September 2015	2540	1.352×10^7	0.43	0.04	0.44	2.00	1.02
8158827	1 July 2004	6648	3.021×10^7	0.96	0.01	0.24	9.52	8.90
8170500	1 October 1994	10,209	1.690×10^8	5.36	3.88	6.88	10.2	3.52
8170890	19 May 2016	2309	9.352×10^6	0.30	0.00	0.25	2.05	0.76
8170950	30 August 2016	2205	4.926×10^7	1.56	0.49	1.90	5.50	3.58
8170990	23 April 2005	6352	7.863×10^6	0.25	0.05	0.37	1.46	0.37
8171000	1 January 1991	11,578	1.500×10^8	4.75	0.85	4.71	26.3	25.9
8171290	1 October 2017	1808	7.088×10^7	2.25	0.73	2.64	10.7	2.96
8171300	1 January 1991	11,578	1.469×10^8	4.65	0.03	4.71	28.9	29.3
8171350	22 January 2015	2791	1.538×10^8	4.87	0.03	4.81	24.8	42.7
8171400	19 May 2011	4135	2.649×10^8	8.39	4.42	9.98	27.2	14.9
7817	17 March 2016	2371	2.365×10^7	0.75	0.34	1.03	3.77	1.66
4595	5 October 2006	5822	2.528×10^7	0.80	0.00	0.08	8.14	7.28

¹ 12 September 2022 is the end date for all records. ² Mean daily flow (MDF) provides the low flow threshold for this study and is the upper limit for low flow indices [16]. It is used here because the 70th percentile threshold is often zero or <0.1 m³/s-days. ³ The flow duration curve (FDC) 70th percentile threshold is a commonly used low flow index [16]. ⁴ The 20th percentile is a threshold for high to medium flows [17]. ⁵ The 2nd percentile is the threshold for out of bank, high flows [17] in this study.

Table 3. Daily ‘goodness-of-fit metrics’ between gauge discharge and synthetic, unbiased discharge time series.

Gauge ID	NSE			KGE			Σ_{NK}		
	Max.	Mean	Min	Max.	Mean	Min	Max.	Mean	Min
8155200	0.91	0.84	0.60	0.80	0.76	0.60	1.70	1.60	1.32
8158700	0.90	0.83	0.65	0.80	0.76	0.61	1.68	1.60	1.39
8158810	0.93	0.84	0.46	0.81	0.75	0.41	1.72	1.59	1.14
8158813	0.90	0.79	0.13	0.79	0.71	0.29	1.66	1.50	0.71
8158827	0.96	0.85	−1.04	0.82	0.72	0.25	1.74	1.57	−0.62
8170500	0.56	0.41	−0.05	0.53	0.45	0.20	0.97	0.86	0.33
8170890	0.89	0.80	0.56	0.78	0.71	0.53	1.64	1.51	1.25
8170950	0.91	0.77	−0.48	0.85	0.73	0.14	1.75	1.51	−0.02
8170990	0.77	0.74	0.69	0.80	0.77	0.73	1.56	1.51	1.46
8171000	0.96	0.83	−0.75	0.85	0.74	0.26	1.78	1.57	−0.30
8171290	0.77	0.68	0.46	0.79	0.73	0.57	1.55	1.41	1.17
8171300	0.96	0.83	−0.29	0.82	0.71	−0.14	1.76	1.54	0.20
8171350	0.99	0.77	−5.02	0.79	0.65	−0.24	1.77	1.42	−4.83
8171400	0.85	0.74	0.26	0.85	0.76	0.43	1.66	1.50	0.91
7817	0.90	0.78	−0.20	0.88	0.77	0.20	1.76	1.54	0.33
4595	0.96	0.84	−0.14	0.79	0.69	0.27	1.72	1.53	0.38

Table 4. Daily ‘goodness-of-fit metrics’ between gauge discharge and synthetic, biased discharge time series.

Gauge ID	NSE			KGE			Σ_{NK}		
	Max.	Mean	Min	Max.	Mean	Min	Max.	Mean	Min
8155200	0.80	0.72	0.61	0.43	0.27	−0.03	1.17	0.99	0.61
8158700	0.77	0.71	0.62	0.42	0.27	0.07	1.14	0.98	0.71
8158810	0.81	0.74	0.54	0.45	0.26	−0.19	1.22	1.00	0.43
8158813	0.70	0.61	0.14	0.41	0.22	−0.32	1.06	0.84	0.22
8158827	0.89	0.78	0.45	0.45	0.24	−0.35	1.31	1.02	0.26
8170500	0.12	−0.10	−0.50	−0.10	−0.25	−0.57	−0.25	−0.35	−0.66
8170890	0.70	0.63	0.41	0.39	0.23	−0.19	1.03	0.86	0.38
8170950	0.71	0.62	−0.37	0.47	0.23	−0.71	1.12	0.85	−0.20
8170990	0.39	0.35	0.28	0.37	0.32	0.24	0.73	0.67	0.60
8171000	0.88	0.75	0.19	0.47	0.24	−0.58	1.32	1.00	−0.01
8171290	0.38	0.27	−0.03	0.33	0.21	−0.08	0.60	0.48	0.23
8171300	0.88	0.77	0.11	0.46	0.25	−0.61	1.32	1.02	−0.06
8171350	0.95	0.78	−1.98	0.46	0.20	−0.90	1.40	0.98	−1.82
8171400	0.61	0.53	0.05	0.44	0.22	−0.32	0.95	0.75	0.17
7817	0.70	0.61	0.04	0.52	0.29	−0.52	1.14	0.90	−0.01
4595	0.91	0.78	0.24	0.45	0.22	−0.65	1.31	1.00	−0.09

Table 5. Monthly ‘goodness-of-fit metrics’ between gauge discharge and synthetic, unbiased discharge time series.

Gauge ID	NSE			KGE			Σ_{NK}		
	Max.	Mean	Min	Max.	Mean	Min	Max.	Mean	Min
8155200	0.97	0.95	0.90	0.81	0.79	0.77	1.77	1.74	1.68
8158700	0.96	0.95	0.91	0.81	0.79	0.76	1.77	1.74	1.69
8158810	0.96	0.94	0.85	0.81	0.79	0.73	1.76	1.73	1.61
8158813	0.94	0.89	0.76	0.81	0.77	0.71	1.74	1.66	1.48
8158827	0.97	0.93	0.64	0.79	0.74	0.53	1.74	1.67	1.34
8170500	0.87	0.80	0.64	0.91	0.86	0.78	1.78	1.66	1.42
8170890	0.98	0.96	0.80	0.83	0.78	0.67	1.81	1.74	1.52
8170950	0.98	0.95	0.65	0.90	0.83	0.63	1.87	1.78	1.30
8170990	0.96	0.95	0.93	0.84	0.82	0.80	1.80	1.77	1.73
8171000	0.98	0.94	0.65	0.85	0.81	0.67	1.82	1.75	1.34
8171290	0.97	0.94	0.86	0.89	0.84	0.77	1.86	1.77	1.63
8171300	0.98	0.94	0.72	0.82	0.78	0.54	1.79	1.73	1.33
8171350	0.98	0.91	−0.21	0.84	0.73	0.12	1.80	1.64	0.29
8171400	0.95	0.92	0.77	0.91	0.87	0.74	1.85	1.78	1.51
7817	0.98	0.94	0.58	0.90	0.83	0.61	1.87	1.77	1.23
4595	0.98	0.94	0.70	0.79	0.74	0.58	1.75	1.68	1.42

Table 6. Monthly ‘goodness-of-fit metrics’ between gauge discharge and synthetic, biased discharge time series.

Gauge ID	NSE			KGE			Σ_{NK}		
	Max.	Mean	Min	Max.	Mean	Min	Max.	Mean	Min
8155200	0.69	0.67	0.63	0.48	0.42	0.34	1.15	1.09	0.98
8158700	0.68	0.65	0.61	0.48	0.43	0.36	1.14	1.08	1.00
8158810	0.65	0.62	0.51	0.48	0.42	0.29	1.09	1.04	0.87
8158813	0.26	0.06	−0.32	0.46	0.43	0.33	0.69	0.48	0.11
8158827	0.80	0.74	0.60	0.45	0.32	−0.08	1.21	1.06	0.52
8170500	−2.64	−4.27	−6.44	0.36	0.30	0.17	−2.29	−3.97	−6.27
8170890	0.78	0.73	0.65	0.49	0.40	0.17	1.26	1.13	0.86
8170950	0.61	0.51	0.05	0.53	0.44	0.08	1.08	0.95	0.53
8170990	0.43	0.35	0.27	0.51	0.50	0.48	0.94	0.85	0.77
8171000	0.74	0.70	0.57	0.49	0.38	0.06	1.21	1.08	0.66
8171290	0.35	0.12	−0.24	0.53	0.50	0.44	0.85	0.62	0.25
8171300	0.79	0.75	0.57	0.48	0.37	0.07	1.25	1.12	0.73
8171350	0.85	0.77	0.34	0.50	0.31	−0.37	1.33	1.08	0.22
8171400	0.15	−0.16	−0.87	0.53	0.47	0.25	0.53	0.31	−0.39
7817	0.51	0.35	−0.95	0.54	0.47	0.15	0.98	0.82	−0.54
4595	0.83	0.77	0.62	0.46	0.32	−0.01	1.27	1.08	0.61

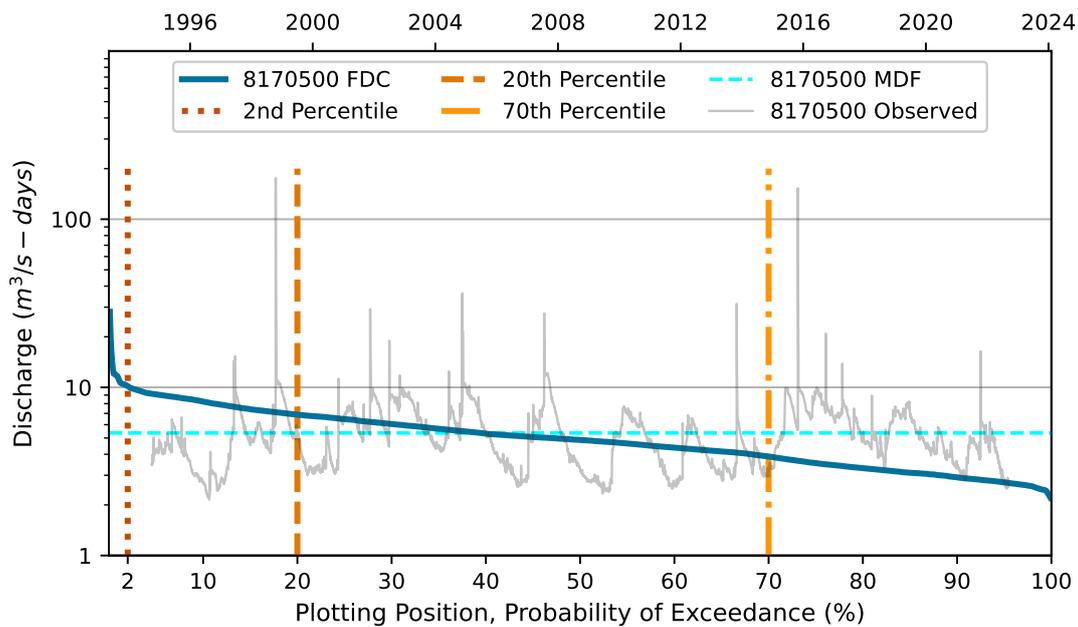


Figure 5. Daily flow duration curve (FDC) for station 8170500 on the San Marcos River. Note that all observed discharge values are greater than $2 \text{ m}^3/\text{s-days}$ denoting consistent and significant base flow from San Marcos Springs, which are located just upstream of this gauging station.

3.2. Parameter Uncertainty Analysis and Prior–Data Conflict Results

$RMSE_i$, see Equation (4), calculated for each gauging station and time interval generates the observation error model in a PESTPP-IES calibration-constrained, model predictive uncertainty analysis. These uncertainty envelopes are provided as part of target specifications using the calculated $RMSE_i$ for the “standard deviation” for each discharge target interval.

PESTPP-IES uncertainty analysis employed 3488 total HSPF model simulations, using the HSPF model described in Section 2.5, split across three iterations with 848 realizations. The “base” realization from iteration 3 had the lowest “base” realization Φ value; this Φ value was used as the Φ threshold to identify a collection of 94 best-fit parameter ensembles.

Figure 6 displays an example of the implementation of the expected uncertainty envelope as the observation error model for discharge targets in a DA implementation. This figure also shows a subset of the best-fit results from a PESTPP-IES analysis; 24 of the 94 ensembles are shown on this figure. The 24 selected ensembles are those that produce an Σ_{NK} value ≥ 1.00 , which is the minimum Σ_{NK} for 8158700 in Table 6. Results in Figure 6 are from a preliminary analysis, whose purpose is to constrain watershed parameters. The discharge targets, incorporating observation uncertainty, provide constraint on posterior parameter values for these watershed parameters.

The FDC for 8158700 is provided on Figure 4. In this implementation, monthly averaged discharges from the gauging station are the “Observed” values shown on Figure 6. The monthly averaged $RMSE_i$ was calculated by aggregating the gauge discharge time series and synthetic discharge time series to monthly averaged intervals and then using monthly, rather than daily, intervals for i in Equation (4). Shading in Figure 6 shows the $RMSE_i$ for each monthly target. A similar envelope, calculated from the standard deviation value as the monthly average gauge discharge (observed) plus the standard deviation, is labeled “ \pm Standard deviation” and shown for comparison purposes. The standard deviation of the observed discharge time series is $2.82 \text{ m}^3/\text{s-months}$.

The posterior realizations, selected as the realizations producing Φ values \leq the threshold and Σ_{NK} values ≥ 1.00 , are the ‘best-fit, simulated’ lines in Figure 6. Simulated discharge generally falls within the expected uncertainty envelope. The comparison

“±Standard deviation” envelope is typically larger than “observation uncertainty”, which is the expected uncertainty envelope.

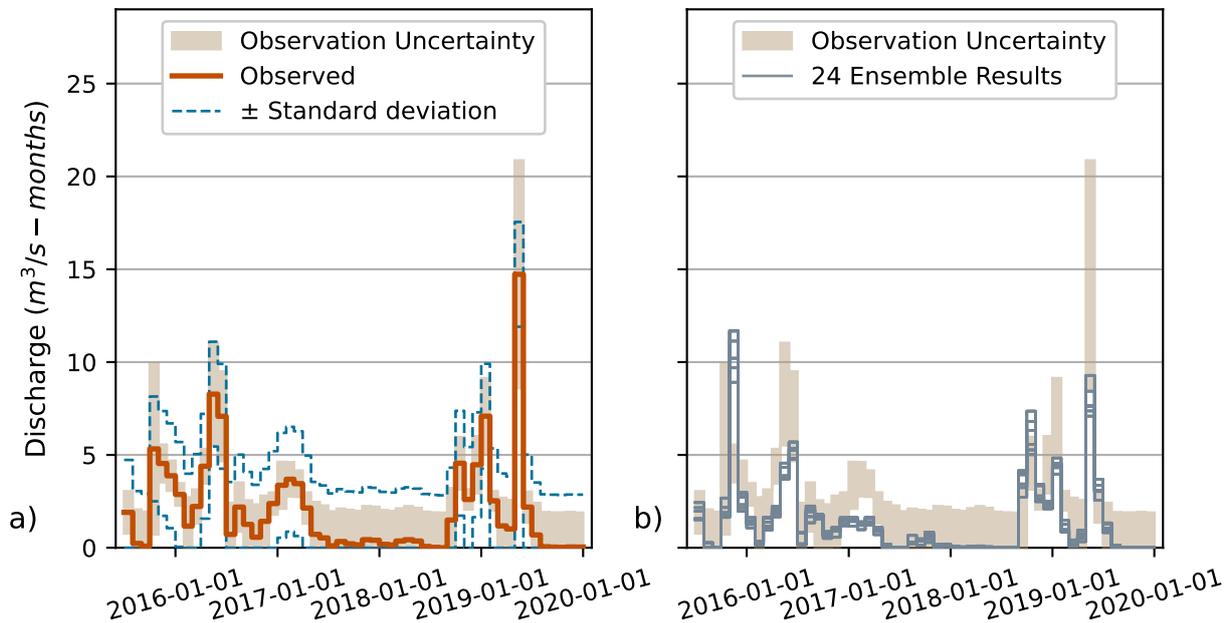


Figure 6. PESTPP-IES uncertainty analysis results for 8158700. Panel (a) displays observed discharge along with observation uncertainty descriptions. Observation uncertainty is the RMSE estimated for each month as described in Section 2.3, or the expected uncertainty envelope. The standard deviation of observed discharge is displayed for comparison with the expected uncertainty envelope, and ± standard deviation provides an independent estimate of observation uncertainty. The expected uncertainty envelope is tighter, or narrower, relative to the standard deviation envelope at low and medium discharges and is larger at high discharges. Panel (b) displays simulated discharges from 24 best-fit, parameter ensembles. These 24 best-fit ensembles achieve a Φ threshold that identifies the overall best-fit parameter ensembles and produce Σ_{NK} values ranging from 1.00 to 1.24. In Table 6, Σ_{NK} for 8158700 ranges from a minimum of 1.00 to a maximum of 1.14 with an average of 1.08.

Eight prior-data conflicts were identified at two different gauging stations during the PESTPP-IES analysis, see Table 7. There are 16 stream gauging stations that provide a total of 791 monthly averaged discharge targets; prior-data conflicts are about 1% of the discharge observations. Herein, prior-data conflicts are addressed by instructing PESTPP-IES to remove the corresponding target value from the analysis. This means that agreement between simulated and observed values for these intervals is not used to constrain optimal parameter ensembles.

Table 7. Listing of discharge target, prior-data conflicts.

Gauge ID	Month	Observed m ³ /s-months	RMSE ¹ m ³ /s-months
7817	2017-01	1.11	0.23
	2017-02	1.03	0.21
	2019-03	1.14	0.24
	2019-04	1.03	0.21
8170290	2019-02	5.37	1.12
	2019-03	3.26	0.67
	2019-04	2.74	0.70
	2019-06	4.77	0.98

¹ Root mean square error (RMSE) is calculated with Equation (4).

Figure 7 shows an example of the uncertainty envelope implementation for Station 8171290, which has four prior-data conflicts. Station 8171290 is located between 8171000 and 8171300 (see Figures 2 and 3) on the Blanco River. Although the target values for the prior-data conflicts are removed, the simulated values for the prior-data conflict intervals are still tracked.

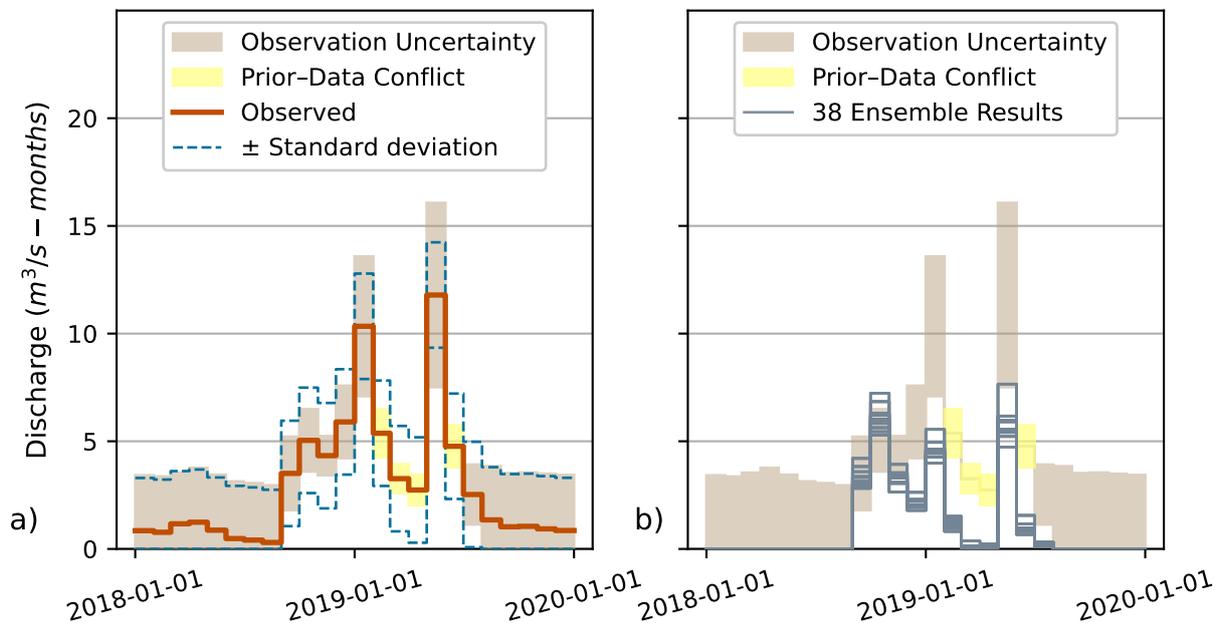


Figure 7. PESTPP-IES uncertainty analysis results for 8171290. Panel (a) displays observed discharge and observation uncertainty descriptions. “Observation uncertainty” is the RMSE, calculated as described in Section 2.3, or the expected uncertainty envelope. The standard deviation of the observed discharge time series is displayed for comparison, and “ \pm Standard deviation” provides an independent estimate of observation uncertainty. The expected uncertainty envelope is tighter, or narrower, relative to the standard deviation envelope at low and medium discharges and is larger at high discharges. Prior-data conflicts are observed values that the model cannot reproduce, when including observation uncertainty with observed values. Panel (b) shows simulated discharges from 38 best-fit, parameter ensembles. These 38 best-fit ensembles achieve a Φ threshold that identifies the overall best-fit parameter ensembles and produce a Σ_{NK} values from 0.26 to 0.71. In Table 6, Σ_{NK} for 8171290 range from a minimum of 0.25 to a maximum of 0.85 with a mean value of 0.62.

4. Discussion

In generic DA terms, the observation uncertainty envelope provides an observation error model for discharge observations that accounts for measurement error and rating curve representation error. PESTPP-IES inherently addresses numerical representation error with its prior-data conflict resolution abilities. Combining the discharge observation uncertainty envelope with PESTPP-IES produces a DA observation error model that accounts for measurement error, rating curve representation error, and numerical representation error.

The goal of the uncertainty envelope, and all DA observation error models, is to avoid bias in and minimize the variance of the posterior parameter distribution. PESTPP-IES uses the observation error model to convert targets to a range of acceptable values for each history matching time. If the range is too large, conditioning information on parameter values will be greatly reduced, and large posterior parameter uncertainties will result. This effect will produce excess variance in predictions made with the model. If the target range is too small, then too much accuracy and precision are attributed to the observed values, and the posterior parameter ensemble may be overly narrow and biased as a result of “overfitting”. This bias in estimated parameters ultimately translates to bias in the important predictive outcomes. The alternatives of encouraging overfitting, due to

artificially specifying too much target constraint, and larger spread in posterior parameter ranges, due to lack of target constraint, define a bias–variance trade-off.

Two alternative discharge observation error models, (1) no observation error model and (2) the standard deviation of the gauge record, were implicitly examined. The alternative of no observation error model is described in Figure 1; the concern with no observation error model is that the range of target values for each history matching time, i , will be too narrow and will promote overfitting and introduce bias into the posterior parameter distribution.

The other alternative discharge observation error model is the “ \pm Standard deviation” envelope shown on Figures 6 and 7. This discharge observation error model is the standard deviation of the observed monthly discharge time series added to and subtracted from the monthly averaged observed discharge. In this case, a constant standard deviation value is used for all i for targets observed at the same gauging station. Because PESTPP-IES (and DA implementations in general) uses an additive Gaussian error model for each history matching time, i , a constant value across all i will provide a poorly resolved observation error model for highly variable and dynamic observation sequences, such as those obtained from gauges in the study area and shown in Figures 3 and 4.

Examination of Figures 6 and 7 shows that “Observation Uncertainty” envelopes, which are the expected uncertainty envelope observation error model, generally have a smaller range than “ \pm Standard deviation” envelopes, which are the standard deviation observation error model, suggesting that the expected uncertainty envelope will produce relatively narrow posterior distributions, denoting a decrease in parameter variance and amelioration of the variance portion of the bias–variance trade-off.

The expected uncertainty envelope created by $RMSE_i$ produces a different additive Gaussian error model for each history-matching time, i . This permits the inclusion of different relative errors for low flows ($\pm 100\%$), high flows ($\pm 40\%$) and medium flows ($\pm 20\%$). It also allows the use of non-Gaussian random variates in Equation (3) to generate $RMSE_i$ values. River flow, and most other hydrology data sets, are often best described with non-Gaussian distributions, such as extreme value distributions.

In this study, prior-data conflicts, such as those highlighted on Figure 7, are on the falling limbs of the hydrographs. The study environment is karst terrain characterized by complex and largely unknown enhanced flow pathways in caves, conduits, and regions of enhanced secondary porosity. We hypothesize that the subsurface storm flow will be an important stream flow generation mechanism across the study area and will contribute significantly to the falling limb of discharge hydrographs. This physical process is not represented in the HSPF models used to produce Figures 6 and 7. The watershed parameter values that are the focus of the calibration-constrained, model predictive uncertainty analysis that generated these figures will not significantly impact the representation of subsurface storm flow. The exclusion of these prior-data conflicts prevents the inversion process from calibrating bias into the best-fit watershed parameter values for targets, where these parameters should not be directly relevant. This is a specific example of the inherent numerical representation error handling capabilities in PESTPP-IES.

4.1. Limitations of the Expected Discharge Uncertainty Envelope

The expected uncertainty envelope provides an observation error model accounting for measurement error and representation error between the rating-curve discharge model and observations. Consequently, it is only applicable when discharge is (1) an observed value for DA and (2) calculated or estimated rather than measured. These two limitations restrict its use case to hydrologic studies focused on regional and sub-regional water budget representation that employ hydrologic routing (i.e. using hydrologic models which simulate discharge).

Studies that use hydrodynamic models, comprised of partial differential equations solving for spatial and temporal variation in free surface elevation and conservation of momentum, to simulate river flow will not use discharge as an observation for DA. This

type of numerical model provides a more robust and accurate calculation of discharge than a rating curve and could use observed water depth from the gauging station as observations for data assimilation. When hydrodynamic models that explicitly simulate spatially variable flow, rather than discharge, are used, other types of observations that involve direct measurement of flow velocity, such as the observations obtained from an acoustic Doppler current profiler (ADCP) are often preferred [28].

For the study area, the identification of flow regime is an important component of generating the expected uncertainty envelope because low flows occur frequently and provide for the largest relative error expectation (i.e., $\pm 100\%$). Low flows are also a water resources management focus in this region. However, study areas, or sub-areas, with limited discharge variability, such as Figure 5, do not require a flow-regime dependent, uncertainty representation. With a relatively constant discharge record, the $RMSE_i$ values will likely be similar to the standard deviation of the record. In the case of limited dynamic variation in the observation record, a flow-regime dependent observation error model is not needed.

4.2. Goodness-of-Fit Metric Values

NSE and KGE, the two goodness-of-fit metrics used in this paper, have a range of $-\infty$, for no predictive skill, to 1, for perfect predictive skill. Examination of Tables 3–6 suggests that KGE is generally smaller than NSE but that there is not a completely consistent relationship between the two metrics for the study area data sets and implementation. Consequently, Σ_{NK} , which combines NSE and KGE, is used as a single goodness-of-fit metric; it has a range of $-\infty$, for no predictive skill, to 2.0, for perfect predictive skill.

Discharge observations derived from stage measurements using a rating curve are expected to generate significant observation model error components for DA because a rating curve provides an imperfect hydrodynamics model for streams and rivers. Hydrodynamics models such as [28,29] simulate spatial and temporal variation in free surface elevation and the conservation and transport of momentum, which are the primary drivers of stream and river flow at frequencies higher than daily. When constant water density is assumed, the transport of momentum can be simplified to the transport of velocity and velocity travels with the fluid, such as a scalar [29–31]. Consequently, there is not a robust, physics-based rationale to expect a strong correlation between stage and discharge for frequencies higher than the daily one, and a rating curve provides a limited and imperfect hydrodynamics model at these higher frequencies. These limitations are reflected in the mean Σ_{NK} values obtained from the comparison of stream gauge discharge data sets to the stochastically estimated discharge series, see Tables 3 and 4.

For lower-frequency discharge estimation, such as monthly, relative improvement in mean Σ_{NK} values in Tables 5 and 6 suggest that a rating curve provides a more robust hydrodynamics model for monthly discharge estimation, relative to daily discharge estimation. The unbiased variate mean Σ_{NK} , for monthly series comparison in Table 5, is always greater than 1.6, which denotes significant skill in estimation of “true” discharge values. Intuitively, it is reasonable to expect that an increase in the average stage across an interval of days to weeks correlates with an increase in the average discharge across the same interval.

PESTPP-IES produces an ensemble of parameters that provide best-fit history matching between simulated values and target values. An ensemble of parameters necessarily generates an ensemble of results. An important analysis decision when using ensemble methods is selecting which ensembles represent equally good reproduction of observed history. This selection is complicated by the fact that target data and parameterization knowledge are limited and there is uncertainty for both target and parameter values.

The NSE, KGE, and Σ_{NK} values in Tables 3–6 can be employed as thresholds, or criteria, for the selection of equally best-fit ensembles. The evaluation threshold for each gauge and each metric would be the unbiased variate mean value for gauging stations with “good” data quality during WY2021 (see Table 1) and the biased variate mean value for gauging

stations with “fair” or “poor” data quality. These criteria provide several of many criteria, assuming that other target types besides discharge are employed, by which the “final” collection of equally best-fit parameter ensembles would be selected from the PESTPP-IES ensemble results. To illustrate this, Σ_{NK} values from Table 6 are used to select the best-fit ensembles shown in Figures 6 and 7.

In conceptualization, these goodness-of-fit thresholds provide an estimate of the information content in the discharge record accounting for the observation error model. If the calibrated or trained environmental model results produce NSE, KGE, or Σ_{NK} metrics that exceed these thresholds, that is good, but it is possible that the model is being calibrated to overfit these targets, if the calibration enforces parameter selection to achieve history matching beyond the maximum thresholds.

5. Conclusions

A discharge observation uncertainty envelope is presented and developed that provides an observation error model for ensemble methods of DA. It uniquely accounts for the rating curve representation error, related to differences between the rating curve model of discharge and “true” discharge, and measurement error. In this formulation, the discharge observation uncertainty is flow regime dependent and has the largest relative uncertainty for low flows and then for high flows, with relatively reduced uncertainty for “normal” flows.

The goal of the uncertainty envelope is to avoid bias in and minimize the variance of the posterior parameter distribution. It is compared with two other observation error models: no error model and a standard deviation envelope. The discharge observation uncertainty envelope reduces bias relative to the no-error model because it provides for a range of target values for each history matching interval. It reduces variance relative to the standard deviation envelope because it generally provides a narrower range of target values for each history matching interval.

The observation uncertainty envelope is designed specifically for use with PESTPP-IES, which accounts for numerical representation error through prior-data conflict identification. The combination of the discharge observation uncertainty envelope and PESTPP-IES generates a DA observation error model that addresses the measurement error, rating curve representation error, and numerical representation error. Goodness-of-fit metric thresholds, i.e., NSE, KGE, and Σ_{NK} thresholds, are identified as part of the observation uncertainty envelope development that can be used as selection criteria for the identification of posterior parameter values from PESTPP-IES uncertainty analysis results.

Author Contributions: Conceptualization, N.M. and J.W.; methodology, N.M. and J.W.; software, J.W.; writing—original draft preparation, N.M.; writing—review and editing, N.M. and J.W.; funding acquisition, N.M. All authors have read and agreed to the published version of the manuscript.

Funding: Development of the expected uncertainty envelope approach and algorithm was funded by Southwest Research Institute Internal Research and Development Grant 15-R6209. Data set acquisition and processing were funded under Texas State University Subaward 21040-83739-1. Implementation of the expected uncertainty envelope approach for the study site was funded under Texas State University Subaward 22026-83949-1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to acknowledge the contributions of two anonymous reviewers whose comments and suggestions improved the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADCP	Acoustic Doppler Current Profiler
BFZ	Balcones Fault Zone
CDF	Cumulative distribution function
DA	Data assimilation
FDC	Flow duration curve
HSPF	Hydrological Simulation Program FORTRAN
HRU	Hydrologic response units
iES	Iterative ensemble smoother
KGE	Kling–Gupta efficiency
LCRA	Lower Colorado River Authority
MAR	Mean annual runoff
MDF	Long-term mean daily discharge
NSE	Nash–Sutcliffe efficiency
ODE	Ordinary differential equation
PEST	Parameter estimation
PDE	Partial differential equation
RMSE	Root mean square error
Std	Standard deviation
SVD	Singular value decomposition
TX	Texas
USGS	United States Geological Survey
WY	Water year

References

1. Evensen, G.; Vossepoel, F.; Jan van Leeuwen, P. *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*; Springer: Cham, Switzerland, 2022.
2. Doherty, J. PEST: Model Independent Parameter Estimation & Uncertainty Analysis. 2022. Available online: <https://pesthhomepage.org> (accessed on 27 February 2023).
3. Pest++ Development Team. PEST++: Software Suite for Parameter Estimation, Uncertainty Quantification, Management Optimization, and Sensitivity Analysis. Version 5.1.18. User Manual. 2022. Available online: <https://github.com/usgs/pestpp> (accessed on 24 October 2022).
4. Doherty, J. *Calibration and Uncertainty Analysis for Complex Environmental Models. PEST: Complete Theory and What It Means for Modelling the Real World*; Watermark Numerical Computing: Brisbane, Australia, 2015.
5. Hodyss, D.; Nichols, N. The error of representation: Basic understanding. *Tellus A Dyn. Meteorol. Oceanogr.* **2015**, *67*, 24822. [[CrossRef](#)]
6. Chen, Y.; Oliver, D.S. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* **2013**, *17*, 689–703. [[CrossRef](#)]
7. White, J.T. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environ. Model. Softw.* **2018**, *109*, 191–201. [[CrossRef](#)]
8. LCRA. LCRA HYDROMET. 2022. Available online: <https://hydromet.lcra.org> (accessed on 13 September 2022).
9. USGS Texas Water Science Center. National Water Information System: Web Interface. 2022. Available online: <https://waterdata.usgs.gov/tx/nwis/rt> (accessed on 13 September 2022).
10. Puente, C. *Method of Estimating Natural Recharge to the Edwards Aquifer in the San Antonio Area, Texas*; Water-Resources Investigation Report 78-0010; U.S. Geological Survey: Austin, TX, USA, 1978.
11. HDR Engineering, Inc.; Paul Price Associates, Inc.; LBG-Guyton Associates; Fugro-McClelland (SW), Inc. *Trans-Texas Water Program, West Central Study Area Phase II: Edwards Aquifer Recharge Analyses*; Technical Report; San Antonio River Authority and Others: San Antonio, TX, USA, 1998.
12. Turnipseed, D.; Sauer, V. *Discharge Measurements at Gaging Stations*; Techniques and Methods Book 3, Chapter A8; U.S. Geological Survey: Reston, VA, USA, 2010; ISBN 978-1-4113-2969-0.
13. Kiang, J.E.; Gazorian, C.; McMillan, H.; Coxon, G.; Le Coz, J.; Westerberg, I.K.; Belleville, A.; Sevrez, D.; Sikorska, A.E.; Petersen-Øverleir, A.; et al. A Comparison of Methods for Streamflow Uncertainty Estimation. *Water Resour. Res.* **2018**, *54*, 7149–7176. [[CrossRef](#)]
14. McMillan, H.; Krueger, T.; Freer, J. Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrol. Process.* **2012**, *26*, 4078–4111. [[CrossRef](#)]

15. Vogel, R.M.; Fennessey, N.M. Flow-Duration Curves. I: New Interpretation and Confidence Intervals. *J. Water Resour. Plan. Manag.* **1994**, *120*, 485–504. [[CrossRef](#)]
16. Smakhtin, V. Low flow hydrology: A review. *J. Hydrol.* **2001**, *240*, 147–186. [[CrossRef](#)]
17. Yilmaz, K.K.; Gupta, H.V.; Wagener, T. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resour. Res.* **2008**, *18*. [[CrossRef](#)]
18. Fiering, M.B.; Jackson, B.B. *Synthetic Streamflows*; Number 1 in Water Resource Monograph; American Geophysical Union: Washington, DC, USA, 1971. [[CrossRef](#)]
19. Legates, D.R.; McCabe, G.J. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [[CrossRef](#)]
20. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [[CrossRef](#)]
21. White, J.T.; Hunt, R.J.; Fienen, M.N.; Doherty, J. *Approaches to Highly Parameterized Inversion: PEST++ Version 5, a Software Suite for Parameter Estimation, Uncertainty Analysis, Management Optimization and Sensitivity Analysis*; Techniques and Methods 7C26; U.S. Geological Survey: Reston, VA, USA, 2020.
22. Evans, M.; Moshonov, H. Checking for prior-data conflict. *Bayesian Anal.* **2006**, *1*, 893–914. [[CrossRef](#)]
23. Alfonzo, M.; Oliver, D.S. Evaluating prior predictions of production and seismic data. *Comput. Geosci.* **2019**, *23*, 1331–1347. [[CrossRef](#)]
24. Oliver, D.S. Diagnosing reservoir model deficiency for model improvement. *J. Pet. Sci. Eng.* **2020**, *193*, 107367. [[CrossRef](#)]
25. Bicknell, B.R.; Imhoff, J.C.; Kittle, J.L., Jr.; Donigan, A.S., Jr.; Johanson, R.C.; Barnwell, T.O. *Hydrological Simulation Program—Fortran User’s Manual for Release 11*; U.S. Environmental Protection Agency: Athens, GA, USA, 1996, p. 100; ISBN EPA/600/R-97/080.
26. Donigan, A.S., Jr.; Imhoff, J.C. History and Evolution of Watershed Modeling derived from the Stanford Watershed Model. In *Watershed Models*; CRC Press: Boca Raton, FL, USA, 2006; pp. 21–45.
27. Chow, V.T.; Maidment, D.R.; Mays, L.W. *Applied Hydrology*; TATA McGRAW-HILL EDITION; McGraw-Hill Education: Gautam Buddha Nagar, India, 1988.
28. Martin, N.; Gorelick, S.M. MOD_FreeSurf2D: A MATLAB surface fluid flow model for rivers and streams. *Comput. Geosci.* **2005**, *31*, 929–946. [[CrossRef](#)]
29. Casulli, V.; Cheng, R.T. Semi-implicit finite difference methods for three-dimensional shallow water flow. *Int. J. Numer. Methods Fluids* **1992**, *15*, 629–648. [[CrossRef](#)]
30. Staniforth, A.; Côté, J. Semi-Lagrangian Integration Schemes for Atmospheric Models—A Review. *Mon. Weather Rev.* **1991**, *119*, 2206–2223. [[CrossRef](#)]
31. Martin, N.; Gorelick, S.M. Semi-analytical method for departure point determination. *Int. J. Numer. Methods Fluids* **2005**, *47*, 121–137. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.