-Uncertainty Analysis Of An Inverse Hydrogeological Model In The North Of Bucaramanga.

Óscar Fabián Sánchez Ortiz

Degree Project To opt to the tittle of Master In Geophysics

Advisor

Wilfredo del Toro Rodríguez

Civil Eng., Msc

Co-advisor

Sandra Rocío Villamizar Amaya

Civil Eng., Msc, Phd

Universidad Industrial de Santander

Faculty of Science

Physics School

Master In Geophysics

Bucaramanga

2023

Table of Contents

Page.

List of Figures

Page.

List of Tables

Page.

List of Appendices

Page.

Resumen

Titulo: análisis De Incertidumbre De Un Modelo hidrogeológico Inverso En El Norte De Bucaramanga [*].

Autor: Óscar Fabián Sánchez Ortiz[**]

Palabras Clave: Incertidumbre, Modelado Hidrogeológico, MODFLOW 6, IES, Galería

Descripción:

El modelado hidrogeológico para el diseño de infraestructura requiere una cuantificación sólida de las incertidumbres de pronósticos para reducir los costos y optimizar la inversión. Construimos un modelo numérico de aguas subterráneas para pronosticar los caudales producidos por la construcción de una galería de drenaje. La galería se encuentra sobre un acuífero aluvial no confinado en Bucaramanga, Colombia, donde el agua subterránea está causando inestabilidad en el suelo. Se implementaron enfoques emergentes para el modelado, como el "modelado programático", el "pronóstico primero" y la "cuantificación temprana de la incertidumbre" (EUQ). El software utilizado incluyó MODFLOW 6 y PEST++, con los paquetes Flopy y Pyemu como interfaces de Python. El enfoque programático combinado con Git y GitHub para la documentación del modelo permite una alta transparencia y reproducibilidad para cualquier persona que desee auditar el modelado y el pronóstico de incertidumbre. La cuantificación de la incertidumbre se realizó utilizando el método de conjuntos iterativos suavizado (IES), el complemento de Shur (lineal) y Monte Carlo de espacio nulo (NSMC). La evaluación del modelo a través de EUQ probó los efectos de la parametrización previa del refinamiento de la cuadrícula en el pronóstico de flujo, lo que ayudó a corregir fallas en la conceptualización del modelo. Además, el IES demostró ser más liviano computacionalmente y suficientemente robusto para manejar la no linealidad e inestabilidades del modelo, pero la asimilación de datos fue deficiente en el modelo. A pesar de eso, IES fue más eficiente computacionalmente que los otros métodos. Debido a la falta de linealidad y las inestabilidades del modelo, la matriz jacobiana y la optimización por Gauss-Levenberg-Marquardt tuvieron problemas que se heredaron al NSMC y el método lineal. Sugerimos usar IES con EUQ cuando el objetivo principal es la cuantificación de la incertidumbre de un pronóstico que depende del nivel freático.

[*] Proyecto de grado

[**] Facultad de Ciencias, Escuela de Fisica. Maestria en Geofisica. Director. Wilfredo del Toro Rodríguez Civil Eng., Msc Co-director Sandra Rocío Villamizar Amaya Civil Eng., Msc, Phd

Abstract

Title: Uncertainty Analysis Of An Inverse Hydrogeological Model In The North Of Bucaramanga[*].

Author: Óscar Fabián Sánchez Ortiz[**]

Keywords: Uncertainty, Hydrogeological Modelling, MODFLOW 6, IES, Gallery

Description:

Groundwater modeling for infrastructure design requires a robust quantification of forecast uncertainties associated with porous media and other hydrological variables. Understanding, quantifying, and minimizing the uncertainty associated with numerical models reduces costs and optimizes investment; therefore, a reliable, open, and replicable model is appropriate. We built a groundwater numerical model to forecast flows yielded by a drainage gallery construction. The gallery is over an unconfined alluvial aquifer in Bucaramanga, Colombia, where groundwater is causing ground instability in an urbanized area greater than 1 km^2. The model implemented emerging approaches for workflow modeling such as programmatic modeling, forecast first, and early uncertainty quantification (EUQ). The software used included MODFLOW 6 and PEST++ for forward and inverse modeling, respectively, with Flopy and Pyemu packages as Python interfaces, all of them open source. Python's programmatic approach using open-source software combined with Git and GitHub for model documentation allows high transparency and reproducibility for anyone who wants to audit the model construction process and forecast uncertainty analysis. Quantification of forecast flow uncertainty was performed using iterative ensemble smoother (IES), Shur's complement (Linear), and Null Space Monte Carlo (NSMC). The assessment of model structure through EUQ tested the effects of prior parameterization of the grid refinement in the flow forecast helping to find and fix several flaws in the model conceptualization. Besides, the IES demonstrated to be lighter in computational burden and robust enough to handle non-linearity and instabilities of the model, but data assimilation was poor in the model, showing null improvement in the forecast after the first iteration. Despite that, IES was more computationally efficient than NSMC and Shur's complement and much more reliable than the linear method. Due to model non-linearity and instabilities, the Jacobian and the Gauss-Levenberg-Marquardt optimization had problems that were carried over to the NSMC and the linear method. We suggest that IES with EUQ should be used when the main goal of the modeling is the uncertainty quantification of a forecast that depends on the water table.

---

[*] Project of grade

[**] Facultad de Ciencias, Escuela de Fisica. Maestria en Geofisica. Director. Wilfredo del Toro Rodríguez Civil Eng., Msc Co-director Sandra Rocío Villamizar Amaya Civil Eng., Msc, Phd

Introduction

Groundwater models are usually built to forecast the results of a proposed action related to the management of water resources (Anderson et al., 2015; White, 2017). However, other purposes of groundwater modeling are hindcasting (to recreate past conditions) and screening (to develop a better understanding of the system) (Anderson et al., 2015). In any case, model calibration (also known as inverse problem or history matching) has been considered essential for groundwater modeling because it allows to decrease the uncertainty of model parameters (Zhou et al., 2014) as well as to get the most probable model for the given data (Anderson et al., 2015); therefore, decreasing the uncertainty of model predictions. In forecast modeling of groundwater systems, future conditions are usually determined using calibrated models where uncertainty limits are reported, but the uncertainty quantification is highly challenging because it is an attempt to quantify what is unknown (Linde et al., 2017)

Subsurface construction of infrastructure below the water table generates risks associated to groundwater drainage which demands critical decision making. This requires a robust quantification of uncertainty prior construction because groundwater systems cannot be exhaustively investigated (Sundell et al., 2019; Zhou et al., 2014), and decisions must be taken under uncertainty (Sundell et al., 2019). Tunnel construction, a specific case of subsurface infrastructure, requires detailed modeling due to the potential significant effects on the environment during and after their placement. One of those effects is the drainage of groundwater and the inevitable reduction of surface stream flows during construction. Furthermore, knowing the order of magnitude of the drainage flow produced by its presence would allow to more

accurately determine the initial investment and operating maintenance costs, both of which could make feasible or unfeasible a project (Suescún, 2016). Although there are examples of uncertainty quantification of tunnels in groundwater systems (Benedek & Dankó, 2009; Gokdemir et al., 2022; Suescún, 2016), only a few publications are available describing the modeling of these structures using a three-dimensional tool such as MODFLOW (Zaidel et al., 2010).

When a decision must be taken, the likelihood of undesired effects is assessed alongside relevant management decision thresholds which are defined on the basis of ecological, economic and/or cultural objectives (Hemmings et al., 2020). However, using any method to make decisions through history matching and uncertainty quantification is still an expensive process, and history matching does not always help to reduce the uncertainty of a forecast. Moreover, history matching has the potential to corrupt the quantification of predictive uncertainty hence, it is not always recommended to carry out a complete history matching process (Hemmings et al., 2020). Furthermore, ill effects arise in the history matching when there is not enough parameterization detail (coarse discretization) producing parameter compensation, bias, or underestimation of the uncertainty, even in relatively complex parameterization schemes (>2000 adjustable parameters). However, it was found that reduced parameterization schemes are enough for some types of forecasting and decision-making, especially when uncertainty is calculated without or before history matching (Knowling et al., 2019; White, Knowling, et al., 2020).

Based on the dangers around simplified coarse models, modelers may be tempted to “err on the side of caution” using a fine discretization. Nonetheless, this may introduce complications related to fine grids such as numerical instabilities, excessive computational burden, or increase of input data needs, among others (Knowling et al., 2019; White, Knowling, et al., 2020). To address and overcome the dangers related with model design, John Doherty & Moore, 2020; White, (2017)

criticized the usual workflow in which forecasts are done in groundwater models, proposing a framework oriented to prioritize and monitor the forecast throughout the model construction and uncertainty quantification. John Doherty & Moore, (2020) synthetized a framework to design groundwater models through strategic abstraction/simplification of parameters and processes to optimize the model building for the making of a particular prediction by the assimilation (or not) of pertinent information. In the same line, Hemmings et al., (2020) proposed a workflow to address decision support modeling under uncertainty where the model structure is tested saving time and computational resources. It starts with a prior uncertainty quantification that accomplishes two objectives: the identification of prior data conflicts (inability of the prior ensemble to match observations) and assessing the proximity of the forecast to the decision threshold. The First objective indicates that the model must be reformulated to represent the reality and avoid bias through history matching, and the latter is important because if the prior forecast uncertainty is far from the decision threshold, it allows to discard (or accept) the possibility of something "bad" happening. Therefore, once the objective of the modeling is fulfilled, the modeling process stops (Hemmings et al., 2020).

However, if data assimilation is needed to reduce the uncertainty, the workflow moves towards an abridge history matching and a preliminary approximation of the posterior uncertainty. This allows to assess whether data assimilation can reduce forecast uncertainty. If it does, either the modeling process is stopped because the uncertainty reduction is enough to address the management decision, or history matching is continued because it is helping to the uncertainty analysis (Hemmings et al., 2020). In case that history matching does not inform the forecast and the forecast uncertainty is not satisfactory for decision support purposes, the modeling process must be stopped and the workflow can be restarted recasting forecasts, model design,

parameterization etc., to improve the flow of information from the observations to the model parameters and forecasts (Hemmings et al., 2020).

Most of the uncertainty algorithms are computationally restricted by the number of adjustable parameters because they require running many direct models to assimilate information (i.e., by building a Jacobian matrix)(Li et al., 2018; White, 2018). This may lead to simplifications that can result in bias in the predictions (White, 2018). Uncertainty analysis in a groundwater model can be done through basic or advanced approaches (the latter with a higher computational burden), and their results are usually shown either as standard deviations or confidence intervals. The basic approach may consist of a linear uncertainty analysis based on the Jacobian matrix and a Bayesian approach, which is also known as Shur's complement (Anderson et al., 2015; Fienen et al., 2010). Error variance is another method commonly used for linear uncertainty analysis (Moore & Doherty, 2005; White et al., 2014, 2016). While uncertainty quantification by linear methods is usually good enough for basic engineering cases with the advantage of minimal computational burden, most of the groundwater modeling problems are non-linear (Anderson et al., 2015) requiring advanced approaches.

The Monte Carlo (MC) is the most common method for advanced uncertainty analysis. It makes a large number of simulations to explore model uncertainty using probability density functions to define parameters without assuming any linearity, and filtering the predictions to allow only models with target objective functions slightly above the minimum value obtained by the inverse modeling (Anderson et al., 2015). To overcome the computational burden of the MC method, there are more efficient variations such as the Monte Carlo Markov Chain (MCMC) method, which favors combinations of parameters with a good fit of the target objective function (also known as phi) (Anderson et al., 2015). MCMC is a recommended method to avoid the

underestimation of the uncertainty in many problems, as are the hydro geophysical inverse models (Linde et al., 2017).

The Null Space Monte Carlo (NSMC) is another modified MC version, widely used because it reduces the burden of calibration-constrained MC in highly parameterized models. NSMC pre-conditions the ensemble of exploring parameters (or combination of parameters) that are uninformed by the calibration dataset, yielding parameter sets that should be close to an acceptable value of the calibration objective function (Anderson et al., 2015; M. Tonkin & Doherty, 2009; White et al., 2016).

Stochastically calibrated methods, such as the Ensemble Kalman filter (EnKF), emerged as an alternative to uncertainty methods that rely on a uniquely calibrated model. These methods generate an ensemble of multiple realizations that assimilate data meanwhile representing multiple combinations of parameters with realistic patterns of spatial variability (Zhou et al., 2014). The EnKF is computationally efficient compared to classical gradient-based methods(such as Gauss-Levenberg-Marquardt) and it is becoming increasingly popular for inverse modeling in hydrogeology (Li et al., 2018). However, more recently, the iterative ensemble smoother (IES) emerged achieving results comparable to EnKF with much less computation cost (Li et al., 2018). It has been demonstrated that IES can be used to easily address history matching and uncertainty quantification of highly parameterized models, enabling low-cost access to decision support analyses (White et al., 2021).

To the best of our knowledge there are no groundwater models of tunnels in the literature applying the principles of strategic abstraction, forecast first and early uncertainty quantification. Neither we found application of IES to groundwater models of drifts, galleries, or tunnels. The purpose of this research was to apply these recent workflow trends to address uncertainty of flow

yield in an underground gallery structure. Moreover, we assessed the suitability of the different types of uncertainty quantification to get insights of the convenience or inconvenience of such methods to assess flow forecast.

Our case of study is based on the work of (UIS & AMB, (2019) where they built a groundwater numerical model to forecast flows yielded by a projected drainage gallery. The gallery would rest over an unconfined alluvial aquifer in Bucaramanga, Colombia, where groundwater is causing ground instability in an urbanized area greater than 1 km^2 (UIS & AMB, 2018, 2019). Many groundwater models were elaborated and calibrated along the different stages of the gallery design to forecast its drawdown and flow. The study obtained a wide range of results in the estimation of hydraulic conductivities and flows making evident that high uncertainty was present in the modeling process (UIS & AMB, 2018, 2019). The models in the 2019 study (UIS & AMB, 2019) were calibrated in steady state (but uncertainty was not assessed) by using a scarcely parameterized approach, using zonation and Tikhonov as regularization devices to address the ill-conditioning typically of inverse groundwater models (Anderson et al., 2015).

Under these circumstances, we saw the necessity of performing the uncertainty analysis for a numerical groundwater model of the drainage gallery to identify the appropriate flow range for the gallery design using few available data in a highly parameterized model. Furthermore, we aimed at transparency and reproducibility through a programmatic approach (Bakker et al., 2016; White, Foster, et al., 2020) using python scripts and relevant packages such as Flopy (Bakker et al., 2016) and Pyemu (White et al., 2016). This is not a minor issue because there are strong calls from the scientific community for more clarity and reproducibility in geoscience and modelling papers. (Gil et al., 2016; White et al., 2021)

1. Theoretical framework

1.1 Groundwater modeling

The groundwater modeling process has two main components: the conceptual model and the numerical model (Servicio de Evaluación Ambiental, 2012). The conceptual model relies on the experience and knowledge of the study site by the researchers; the numerical model is typically solved based on finite-difference or finite-element methods to simulate the groundwater flow. The 3D groundwater flow is represented in transient state with a complex network of sources an sinks; MODFLOW is one of the well-known solvers of the groundwater flow equation (Equation 1) (Anderson et al., 2015):

$$\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) = S_s \frac{\partial h}{\partial t} - W^*$$

Equation 1. Transient groundwater flow in 3D

where,

h is the variable of interest representing the hydraulic head, which is the sum of pressure and elevation heads

x, y, z, t are the independent spatial and time variables

$K_x,\ K_y, K_z$ are the principal components of the hydraulic conductivity tensor, according to the Darcy equation

S_s is the Storage parameter

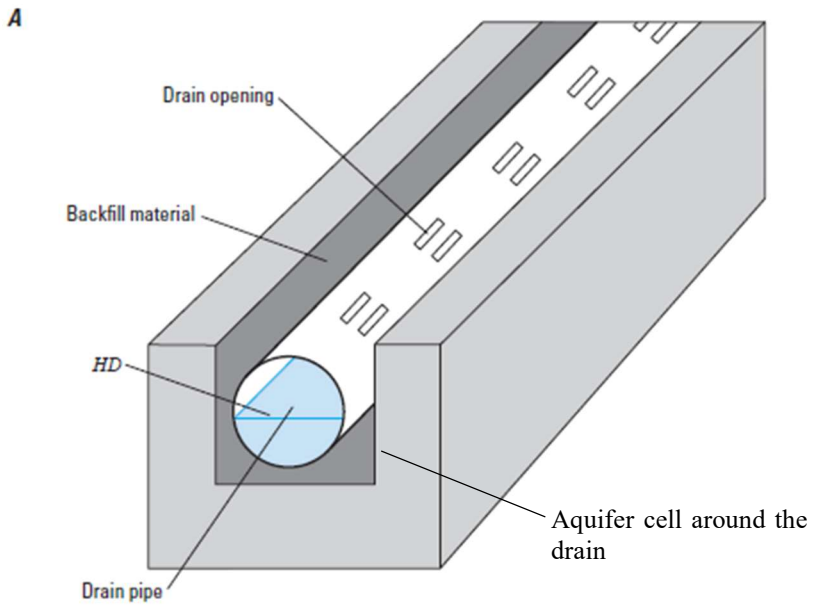
W^* are the water sources and sinks.

The MODFLOW 6 code solves the groundwater equations in steady and transient state. MODFLOW 6 is based on the control-volume finite-difference approach, where each cell can be connected to any number of surrounding cells allowing pinch out layers and a better fit of geological conditions (Langevin et al., 2017).

Galleries and tunnels in MODFLOW are usually represented with the drain (DRN) boundary condition (Suescún, 2016; Zaidel et al., 2010). The DRN package simulates features that remove water from the groundwater system at a rate proportional to the difference between the head in the aquifer cell and a fixed drain elevation when the hydraulic head is above the drain elevation. Many physical conceptualizations can be approximated by the DRN package; for instance, Figure 1 shows a three-dimensional view of a buried drain tile as used for agriculture. The slope of the pipe is assumed to be large enough that once water enters to the pipe, it is carried away without filling it. Head losses occur through three processes: around the drain inside the cell, in the immediate vicinity where the hydraulic conductivity differs from the average value used in the cell (backfill material), and through the wall of the drainpipe. Those head losses may be taken proportional to the discharge and the proportionality constant is called the drain conductance (Langevin et al., 2017).

Figure 1.

Drain conceptualization with factors affecting head loss around a buried drainpipe in a backfilled ditch (Langevin et al., 2017)



1.2 Inverse model or model calibration

In inverse modeling, comparison between simulated and observed values is carried out using a target function. Observed and simulated values are dependent variables of the system (i.e., hydraulic head). The parameter values that produce the "best fit" are defined as those producing the smallest value in the objective function (Hill & Tiedeman, 2007). Given the great amount of parameters involved in the calibration of most groundwater models, it is advisable to use all the available data without saving part of it for verification/validation (Anderson et al., 2015; J.e. Doherty & Hunt, 2010; Linde et al., 2017).

PEST and PEST++ (Model-Independent Parameter Estimation and Uncertainty Analysis) (J. E. Doherty et al., 2010; Welter et al., 2015; White, Hunt, et al., 2020)are the most used tools for parameter estimation in applied groundwater modeling. Among their notable features are (1) it has inversion methods to optimize the objective function (Anderson et al., 2015), and (2) PEST/PEST++ implements the subspace Gauss-Levenberg-Marquardt (GLM) algorithm for least-squares parameter estimation (White, Hunt, et al., 2020). GLM is also known as maximum likelihood method, and it is frequently used with zonification to regularize the solution (assuming areas with homogeneous parameters) (Zhou et al., 2014). The equation of the GLM method according to Hill & Tiedeman, (2007) is shown in Equation 2

$$({C^T X_r}^T \omega X_r \mathrm{C} + \mathrm{Im_r})\mathrm{C}^{-1}\mathrm{d_r} = C^T {X_r}^T \omega\big(y - y'(b_r)\big)$$

$$b_{r+1} = \rho_r d_r + b_r$$

Equation 2. GLM

where,

r is the number of the iteration

X_r is the sensitivity matrix of the parameters in b_r (also known as the Jacobian matrix)

ω is the weight matrix

$({X_r}^T \omega X_r)$ is the symmetrical matrix of NPxNP dimensions (Number of estimated parameters)

$\mathrm{d_r}$ is the vector of parameter change

b_r is the vector of estimated parameters at the beginning of iteration r

C is the scaling parameter.

$\mathrm{m_r}$ is the Marquardt parameter to correct vector $\mathrm{d_r}$ (also known as λ)

ρ_r is the dampening parameter to avoid *overshooting* in the $\mathrm{d_r}$ vector

When the Marquardt (λ) parameter is zero, the upgrade vector direction behaves like the Gauss-Newton (guided by the curvature/Hessian of the parameter space); as λ increases, the upgrade vector changes towards the solution of the gradient descent. PEST++ tests different values of λ to explore a larger portion of the parameter space (Welter et al., 2015).

Calibration of highly parameterized models commonly requires regularization methods to achieve a more stable and smooth target objective function (Anderson et al., 2015). PEST++ uses Tikhonov regularization and truncated singular value decomposition as regularization devices (Welter et al., 2015).

Singular value decomposition (SVD) is a method to solve ill-conditioned (i.e., non-unique solution) least square problems by factoring a G matrix (the Jacobian in groundwater calibration) (from the inverse problem Gm=d) into orthogonal matrices (Equation 3). Where S_p is a diagonal matrix of positive values listed from highest to lowest (Aster et al., 2019b)

$$Gm = d$$

$$G = USV^T$$

$$S = \begin{bmatrix} S_p & 0 \\ 0 & 0 \end{bmatrix}$$

Equation 3 Singular value decomposition

where,

$Gm = d$ is the mathematical model

G is a function matrix

m is the vector of model parameters, to be estimated in inverse problems

d is a vector of discrete observations

U and V are orthogonal matrices corresponding to the data and model space, respectively

S is a diagonal matrix with diagonal elements called singular values.

In ill posed problems, the lowest values of matrix S are close to zero, indicating that there are projections of the parameters in the null space. To avoid the instability of the inverse model, those low values are truncated (TSVD). This generates a worse fit in data but avoids overfitting, especially where the noise is present (Aster et al., 2019a). SVD is usually used with Tikhonov regularization to avoid unreasonable values and overfitting (M. J. Tonkin & Doherty, 2005).The Tikhonov regularization incorporates the soft knowledge (i.e., prior knowledge of the solution) by augmenting the measurement objective function with a second regularization objective function (Equation 4) (Anderson et al., 2015).

$$\Phi_{total} = \sum_{i=1}^{n} (\mathrm{w_i r_i})^2 + \sum_{j=1}^{q} \big(\mathrm{f_i(p)}\big)$$

Equation 4 Target objective function with Tikhonov regularization

where,

The first term is the sum of squared weighted (w) residuals (r)

The second is the sum of functions that penalizes deviation of the parameters (p) from the soft knowledge.

Tikhonov regularization have different parameter-dependent functions to minimize (also known as different orders of Tikhonov regularization). Those regularization measures could be a function of preferred values (zeroth order), or values of the first or the second derivatives of the parameter matrix (m), reflecting preference for flat (first order) or smooth (second order) models (Aster et al., 2019c).

Highly parameterized models have practical limits in the number of parameters for inverse modeling due to the computational burden of estimating a great quantity of parameters. This can be handled by using pilot points, where parameters are estimated in discrete locations distributed

over the model domain, and the rest of the node/cells are assigned by spatial interpolation. The geological and hydrogeological phenomena, although unknown in every point in space, usually shows some continuity behavior which can be expressed statistically by variograms or covariance matrices (J. E. Doherty et al., 2010). The pilot point method was developed to introduce spatial variability by kriging geostatistical interpolation (Zhou et al., 2014)

1.3 Stochastic modeling and uncertainty analysis

The maximum likelihood method yields a unique representation of the aquifer, heterogeneous enough to reproduce the observed data, but it does not necessarily represent the real spatial variability, failing in forecasting phenomena that depend on that variability (Zhou et al., 2014). Stochastic modelling may be a better strategy because the calibration (or inverse model) can be described from a probabilistic approach as the method that selects the most probable model (and forecast) given the existing knowledge.

The Bayes equation (Equation 5) describes the probability of A given information B. The posterior belief ($P(A|B)$) is a function of the prior belief $P(A)$ and the likelihood $P(B|A)$(Anderson et al., 2015).

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Equation 5 Bayes model

The uncertainty analysis may be used to report an estimate of uncertainty surrounding a forecast to address a regulatory requirement or guide management decisions. To reach that objective there are different types of methods with different requirement of computational resources. Linear uncertainty analyses usually require only the sensitivity matrix, then little effort

is required after calibration (Anderson et al., 2015). Two of the most common linear uncertainty methods are Shur's complement and Error Variance Analysis (White et al., 2016). The first one is described in section 0 (Linear analysis within the Methodology). The second one is outlined in Equation 6 (Anderson et al., 2015).

$$\sigma_s^2 = y^t C(p) y - y^t C(p) X^t [X C(p) X^t + C(\varepsilon)]^{-1} X C(p) y$$

Equation 6 Error variance analysis in Bayes formulation

σ_s^2 is the variance of the forecast uncertainty

y is the parameter sensitivity vector of the forecast S

$C(p)$ is the parameter covariance matrix representing expert knowledge

X is the Jacobian Matrix

$C(\varepsilon)$ is the covariance matrix of measurement error of targets.

There are more complex methods to assess uncertainty based on Monte Carlo with no assumptions of linearity (Anderson et al., 2015). The more relevant to our work is Null Space Monte Carlo (M. Tonkin & Doherty, 2009). The method is described in section 0(Post calibration Monte Carlo analysis within the Methodology) but here it is outlined in eight steps.

1. From the base model calculate the sensitivity matrix and construct and calibrate super parameters (using SVD)
2. Calculate the Jacobian (X) with the calibrated parameters
3. Undertake SVD decomposition of $X^T Q X$ to get V1,V2, E1, E2
4. Determine the number of columns in V1 and V2 (dimensionality of the solution space)
5. Generate parameters stochastically using the covariance of parameters
6. Estimate $V_2 V_2^T$ and project differences between stochastic parameters and the calibrated parameters onto the null space.

7. Add the projected values to calibrated parameters values
8. Recalibrate the super parameters in case of misfit.

The IES algorithm is a useful tool to address uncertainty; it uses GLM but the Jacobian is calculated stochastically. It uses an ensemble of random models to develop history matching of all the models and simultaneously quantify their uncertainty. The approximated Jacobian matrix used in IES has a lower computational burden because the number of parameters is not linked to the number of runs necessary to get a Jacobian, allowing the history matching and uncertainty quantification of highly parameterized models (White, 2018). The mathematical implementation of IES is presented in section 0(Iterative Ensemble Smoother within the Methodology).

2. Objectives

2.1 Research question

How to perform the best flow forecast uncertainty analysis of a drainage gallery located in the north of Bucaramanga using few data in a highly parameterized numerical model?

2.2 General objective

Perform the uncertainty analysis for a numerical groundwater model of a drainage gallery to identify the appropriate flow range for the gallery design.

2.3 Specific objectives

Calculate the initial forecast flow by building a prior groundwater model of the north of Bucaramanga in transient state.

Estimate the most probable flow at the gallery through inverse numerical modeling

Analyze the uncertainty of flow forecast for the gallery using ensemble, Monte Carlo, and linear methods to identify the most suitable method for the site.

3. Study Site

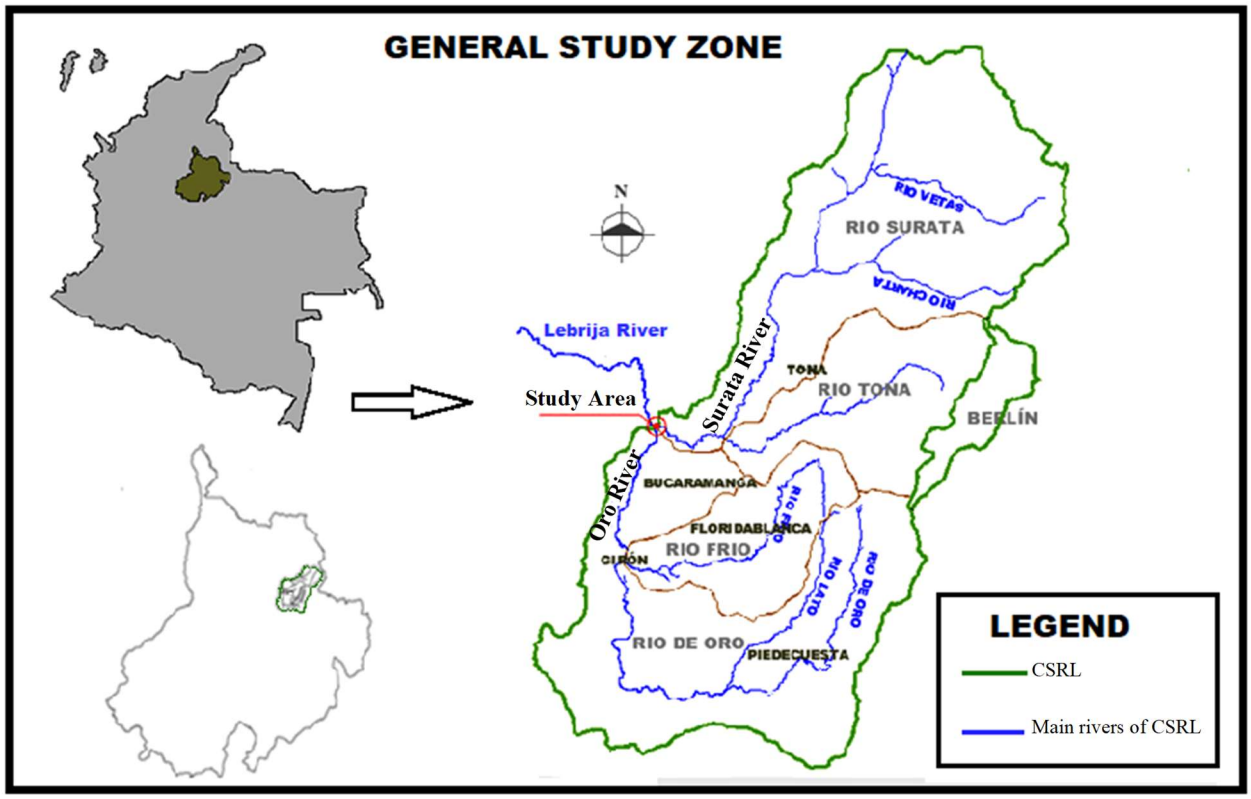
3.1 General information of the study site

The city of Bucaramanga is at the lowest point of the upper basin of the Lebrija River (CSRL), a 1,280 km^2 watershed located in the northwestern foothills of the eastern Andean range of Colombia (Figure 2). A rough topography extends throughout 80% of its area, with slopes between 30 and 200%, in elevations that range between 1,000 and 3,500 m.a.s.l. The remaining 20% of the area corresponds to rolling to flat terrain, located at the south-western side of the watershed with elevations between 640 and 1000 m a.s.l. The main rivers that drain along the basin (Suratá and Oro) are born in the Santander Massif (north-east extent of the basin) but their longitudinal profile softens substantially when they reach the Bucaramanga area (Gomez et al., 2015). Annual precipitation in the area presents a bimodal trend marking two rainy seasons and

two dry (less rainy) seasons. In the mountainous Central area (elevations up to 1,800-2,000 m.a.s.l.) the maximum rainfall can reach 2,000 millimeters per year. In the lowest (west) and highest (east) part of the basin, rainfall is between 700 and 900 mm per year. Isotopic an geochemistry data show that the alluvial aquifers in the lowest part of the basin receive their recharge from the upper part of the basin (Gomez et al., 2015).

Figure 2.

General study area



Note. Adapted by the author from Gomez et al., 2015

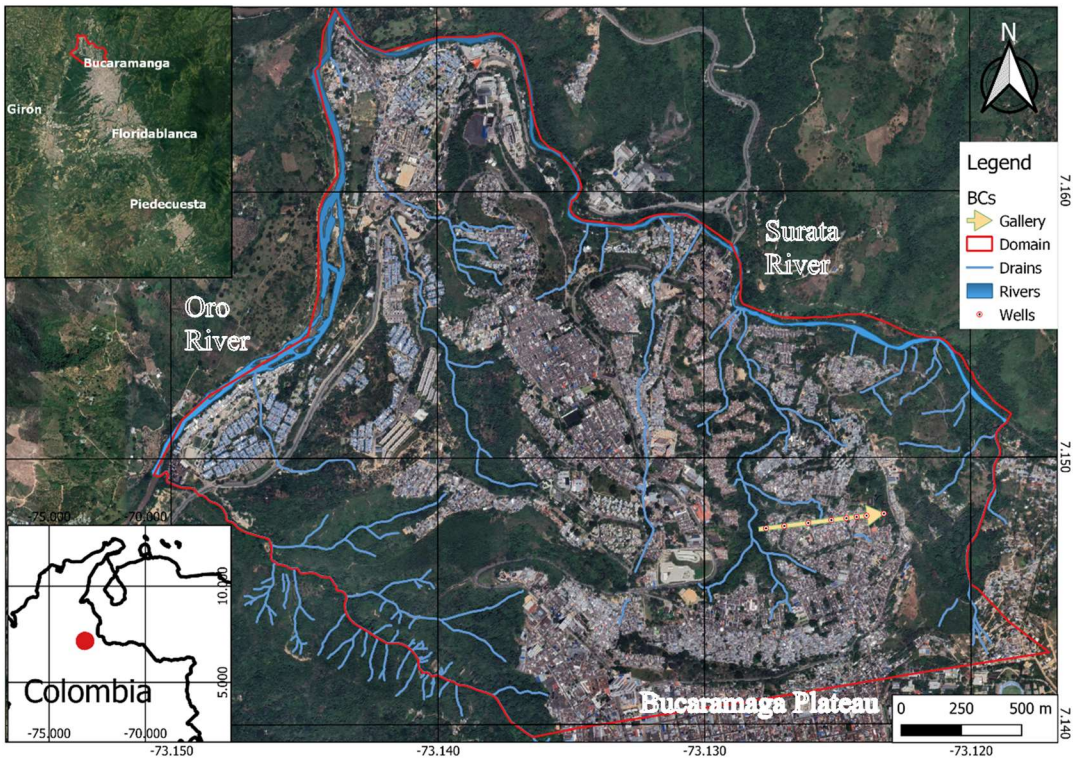
The study site is in the city of Bucaramanga, specifically, on the northern side of the Bucaramanga's plateau, in an area of 7 hectares (

Figure ***3***). The area is limited by The Suratá river on the north-east side, the Oro River on the north-west side, the Picha Creek on the south-west side, and the UIS Creek on the south-east side. The two main rivers that drain the CSRL (Suratá and Oro) are perennial in the study site. This area is of economic and social interest because there have been mass movements that have affected more than 5000 inhabitants and 2000 buildings in the last 40 years. A previous study proposed that to increase the safety factors in the area, it is necessary to decrease pore pressures by lowering the water table (UIS & AMB, 2018). To do this, a drainage gallery was proposed (location shown by the yellow arrow in

*Figure **3***). However, before constructing the gallery it is necessary to understand the behavior of the hydrogeological system; this would help to predict the geotechnical and environmental impact of this gallery.

Figure 3.

Plane view of study area showing boundary conditions and location of the drainage gallery, northern Bucaramanga, the arrow in the gallery represents the direction of the construction process.



Note. Source: author, using a" Google Hybrid" base map.

3.2 Geological frame of the study zone

In the area of Bucaramanga (Santander-Colombia), geology is complex with the upper part of the Lebrija Basin composed of weathered and fractured igneous or metamorphic rocks (Santander Mountain) with high slopes and the lower part (Bucaramanga plateau) composed of alluvial formation (Gomez et al., 2015). The Bucaramanga – Santa Marta Fault divides the two geologic systems (Gomez et al., 2015).

The geological frame of the study area is the Bucaramanga Formation, a quaternary deposit with composition from base to top: Organos Member (Qbo), Finos Member (Qbf), Gravoso Member (Qbg) and Limos Rojos member (Qblr) (Franco & Ramírez, 2018). The Organos member of the Bucaramanga Formation covers the area of interest (see Figure *4*); it is a deposit whose lithological characteristics allowed inferring that it was formed under two sedimentary environments: alluvial processes and fluvial transport (Lima & Medina, 2015).

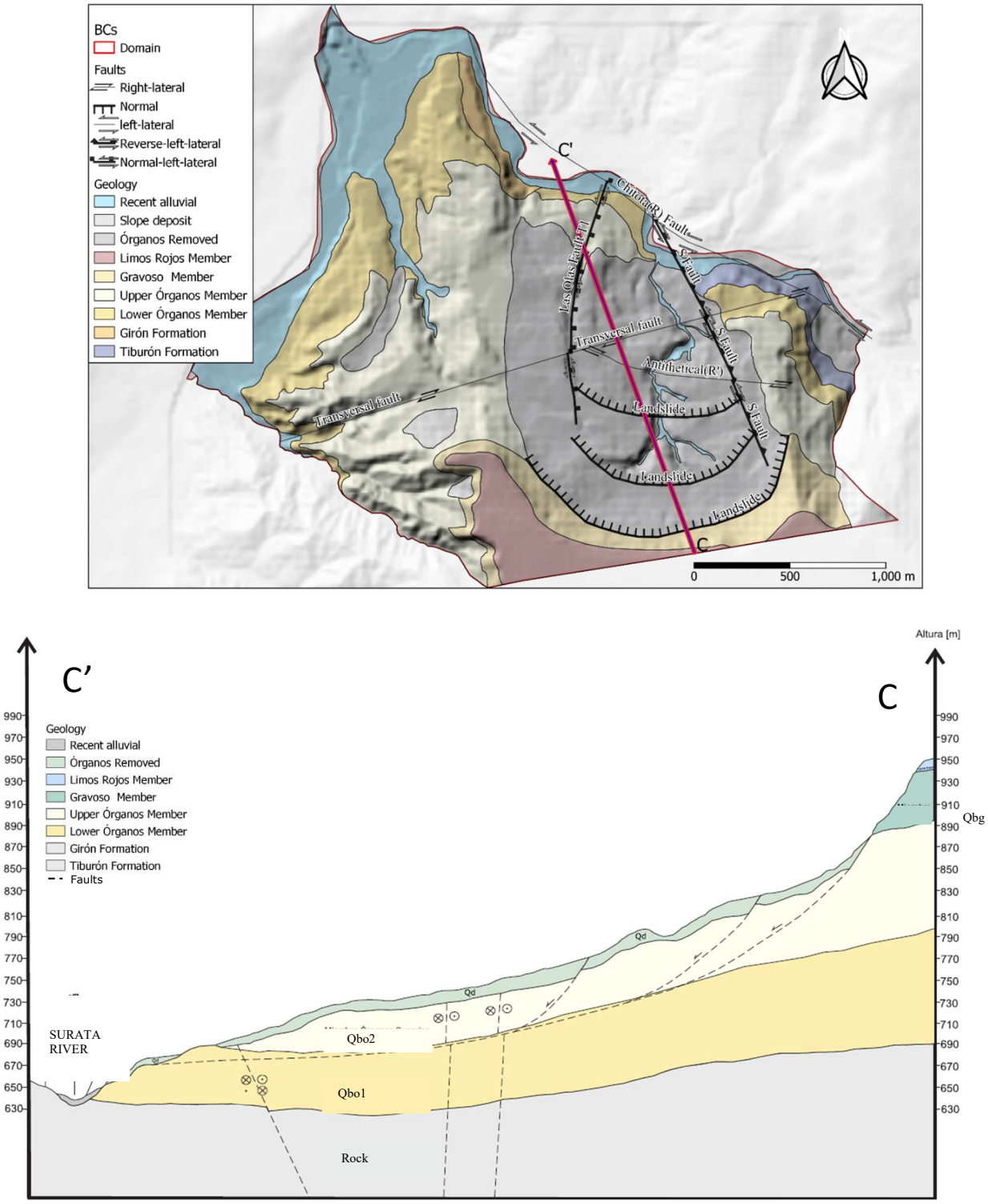
From the granulometry, geometry, fabric, composition, and presence of organic matter within the Organos Member, it is possible to distinguish -in their order- 7 main sedimentary facies.

- Facies Gm: Clast-supported gravels arranged chaotically with slight stratification
- Facies P: Pedogenic carbonate
- Facies Fr: Varicolored clays with the presence of organic matter.
- Facies Sh: Massive sands with scattered pebbles and crude laminar stratification.
- Facies Gms: Sand-supported gravels with slight imbrications.
- Facies Gcm: Gravel supported.
- Facies Pf: Ferricrete.

The structural model that governs the study area is the Bucaramanga Fault system (Velandia & Bermúdez, 2018), which is presented as individual traces, with the complexity of ridges and elongated valleys controlled by synthetic riedel traces I, antithetical (R') and some post-Riedel strokes (P). the geology is partially distorted by the presence of rotational slip faults besides the faults of the structural system (see C-C' profile of Figure *4*, bottom). Those discontinuities in the geological system in the Organos member add complexity and uncertainty to the geological and hydrogeological knowledge of the region.

Figure 4.

Geologic map of the study area. Top: plan view; Bottom: geologic profile for the C-C' section



Note. Adapted by the Author from Gómez Arroyo & Gómez Palencia, 2021 and UIS & AMB, 2019

3.3 Drainage gallery

Drainage galleries to improve slope stability are very common in big and deep landslides (Suarez, 2008). UIS & AMB, 2018 proposed a gallery to lower pore pressures in a pre-feasibility study, and UIS & AMB, (2019) upgraded the studies to detailed engineering. The gallery is to be located in the southeast zone of the study zone, starting its construction at the main creek of the area with a positive slope in the eastward direction, allowing the gallery´s discharge to flow west toward the creek. The location of the gallery and associated wells are shown in

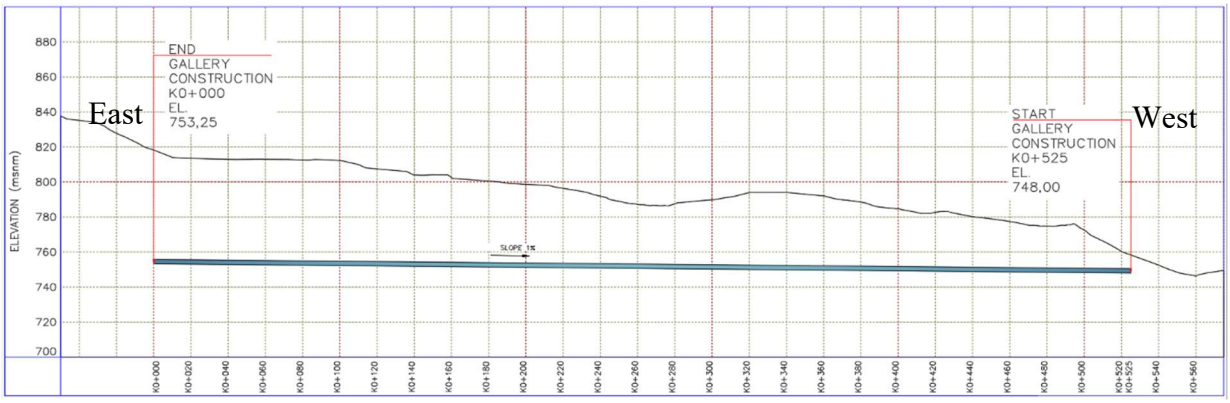
Figure ***3***.

The alignment of the gallery has a straight central axis with a total length of 525 m with a longitudinal slope of 1% (Figure *5*). Eight gallery wells are proposed to help drain the aquifer in case low conductivity layers prevent it (Figure *6*). The cross-section of the gallery has a horseshoe shape, covered with shotcrete, with a maximum height of 3.8m. Seven radial drains every 10 meters allow to drawdown the water table (

Figure 7). The gallery construction rate is established at 2.2 m/d.

Figure 5.

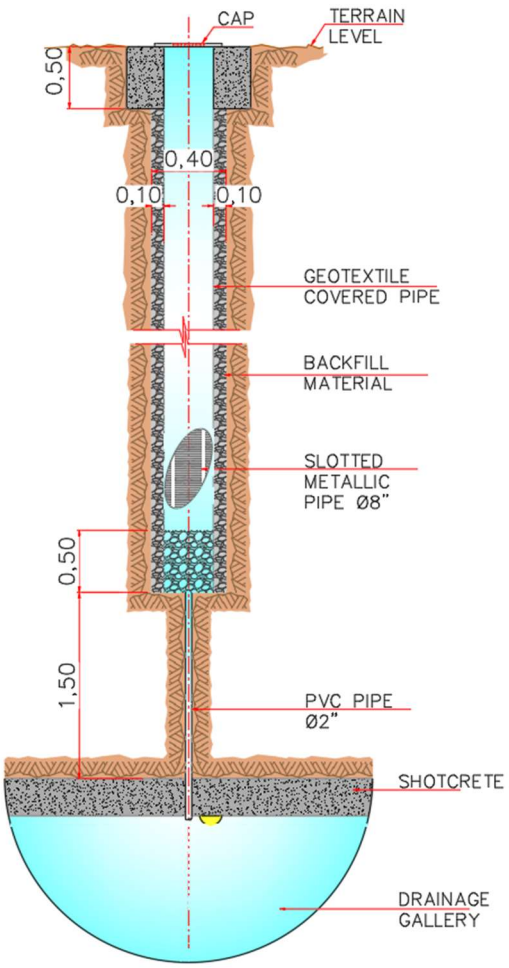
Gallery Profile



Note. Adapted by the author from UIS & AMB, 2019

Figure 6.

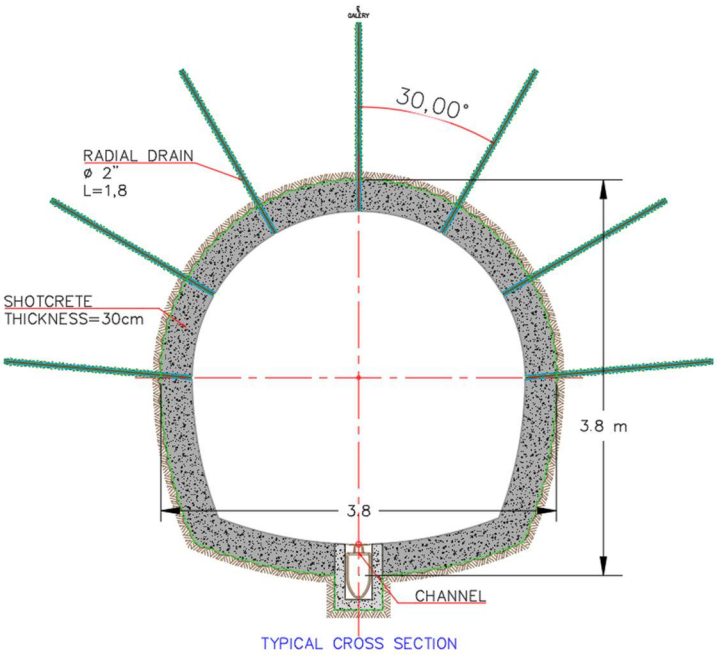
Gallery wells scheme



Note. Adapted by the author from UIS & AMB, 2019

Figure 7.

Gallery design scheme



Note. Adapted by the author from UIS & AMB, 2019

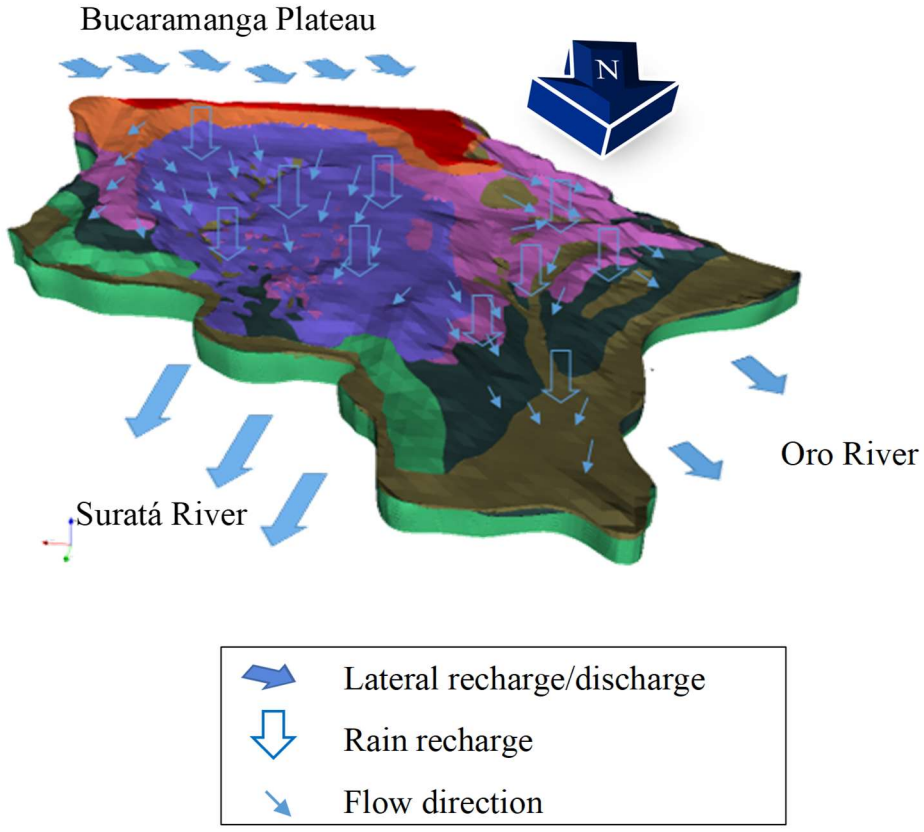
3.4 Hydrogeological conceptual model for the study zone

A recent review by Enemark et al., 2019 states that "hydrogeological conceptual models are collections of hypotheses describing the understanding of groundwater systems". For the case of our study zone, there are three studies that contribute to the general understanding of the system. Antolínez Quijano, 2014 proposed a general conceptual model of the entire Bucaramanga formation. UIS & AMB, 2019 proposed the conceptual model of the groundwater system used in this work, and Medina Baez, 2021 complemented those findings with hydrochemistry analyses. The main processes described in the conceptual model used in this work are flow from south to north coming from a lateral recharge from the Bucaramanga's plateau; local recharge by rain, and

discharge to drains and rivers, especially the Suratá River (Figure *8*). Despite the general understanding that these conceptual models are one of the major sources of uncertainty in groundwater flow and transport modeling (Enemark et al., 2019; Refsgaard et al., 2012), they provide a starting point for specific site analyses. For the case of this study site, the hypotheses are supported by previous studies (Antolínez Quijano, 2014; Gómez Arroyo & Gómez Palencia, 2021; Gomez et al., 2015; Medina Baez, 2021; UIS & AMB, 2018, 2019). More information about the conceptual model can be found in Appendix A.

Figure 8.

Conceptual model scheme.



Note. Adapted by the author from UIS & AMB, 2019

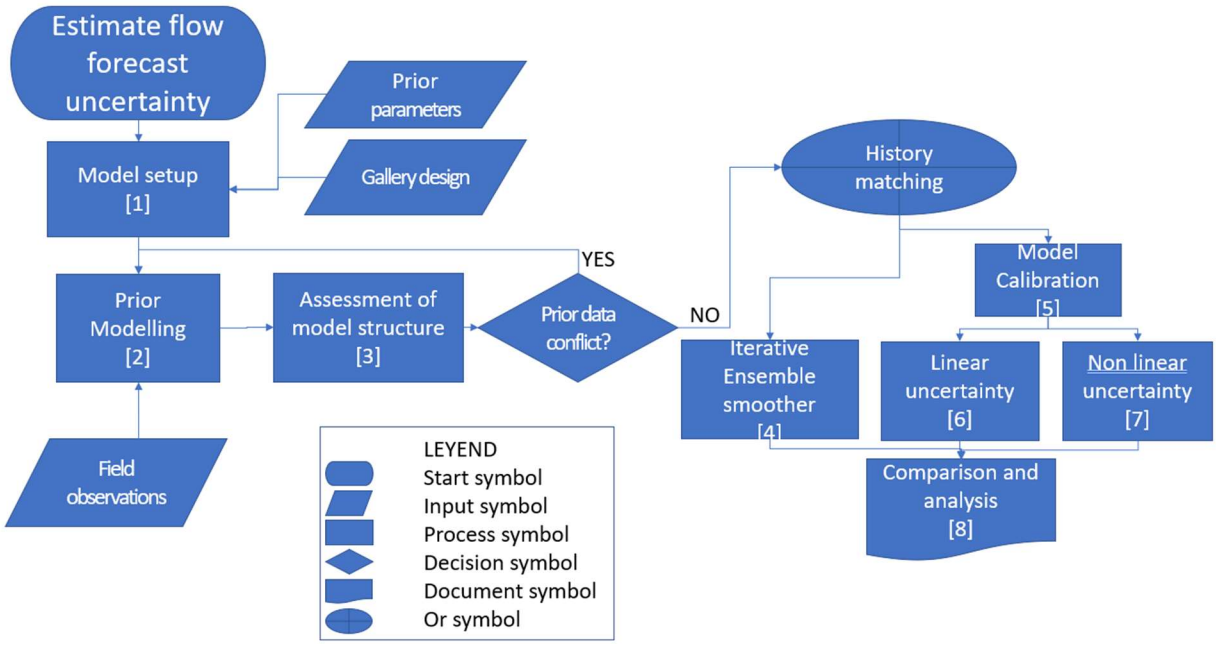
4. Methodology

This study develops an uncertainty assessment of the forecasted flows yielded by a drainage gallery to be built over an unconfined aquifer. To obtain a better assessment of the posterior uncertainty of the forecasted variables (i.e., to better assess the uncertainty of the modeled flows), we applied relevant approaches such as forecast first - FF (White, 2017), strategic abstraction - SA (John Doherty & Moore, 2020), early uncertainty quantification – EUQ (Hemmings et al., 2020) and programmatic approach – PA (Bakker et al., 2016; White et al., 2016, 2021; White, Foster, et al., 2020) throughout the different stages (represented by rectangles) of the workflow (

Figure ***9***). FF and SA were applied from model setup through the assessment of the model structure (1 - 3), EUQ was applied from model setup through the assessment of posterior uncertainty (1 - 4), and PA was used in almost the entire workflow (1 - 7). As the arrows show, some processes are interdependent along the workflow, while others are not (i.e., iterative ensemble smoother - IES (4) vs. linear uncertainty analysis (6)).

Figure 9.

Graphical representation of the workflow



4.1 Model Setup [1]

We used the MODFLOW 6 code to solve groundwater equations in steady and transient state. MODFLOW 6 is based on the control-volume finite-difference approach, where each cell can be connected to any number of surrounding cells allowing pinch out layers and a better fit of geological conditions. The Groundwater flow model in MODFLOW 6 (GWF) is divided into "packages" where a package is the part of the model that deals with a single aspect of a simulation. Packages included within the GWF model include those related to internal calculations of groundwater flow (discretization, initial conditions, hydraulic conductance, and storage), stress packages (constant heads, wells, recharge, rivers, general head boundaries, drains, and evapotranspiration), and advanced stress packages (streamflow routing, lakes, multi-aquifer wells,

and unsaturated zone flow) (Langevin et al., 2017). Conceptual model inputs were taken mainly from UIS & AMB, 2019 where boundary conditions were slightly modified, but raster geological inputs are exactly the same. Details of the model setup are shown in the following subsections.

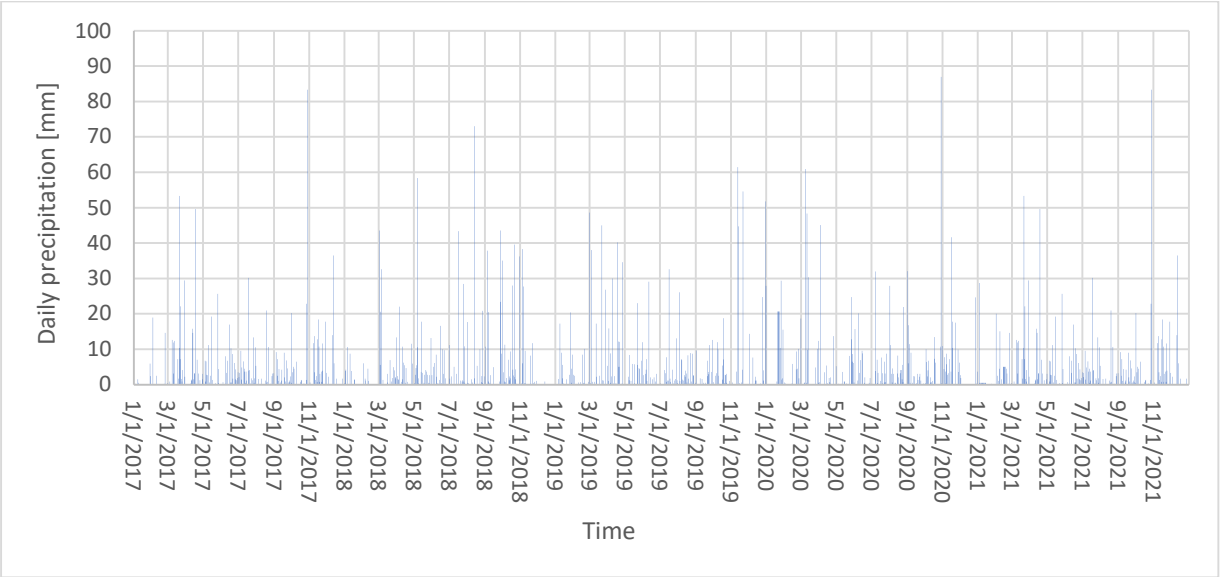
4.1.1 Boundary conditions (BCs).

The addition or removal of water from a groundwater flow model is represented either as a perimeter or internal boundary condition (BC) and are frequently used to represent streams, lakes, springs, wetlands and other more sophisticated features (Anderson et al., 2015). In this work, BCs that enter water into the model (lateral and local recharge) change accordingly to precipitation patterns which are taken from the weather and climate Colombian authority (IDEAM). BCs for the study zone can be seen in Figure 12.

Local recharge is defined as a Recharge BC (Rech) (Neuman-type BC); it changes daily as a percent of the total rain in the area (Figure *10*) (10%, according to the water budget presented in UIS & AMB, 2018). Creeks are defined as drain BCs (DRN) (Cauchy-type BC); Oro and Suratá rivers (to the north) are defined as constant head BCs (CHD) (Dirichlet-type BC). Hydraulic heads in CHD and DRN BC are set 1 m above the ground surface.

Figure 10.

Daily precipitation in the study area where the 10% is used as local recharge (Bucaramanga IDEAM Gauge). Data after February 2021 was not available, then data of 2017 was used to fill the gap in the future time.

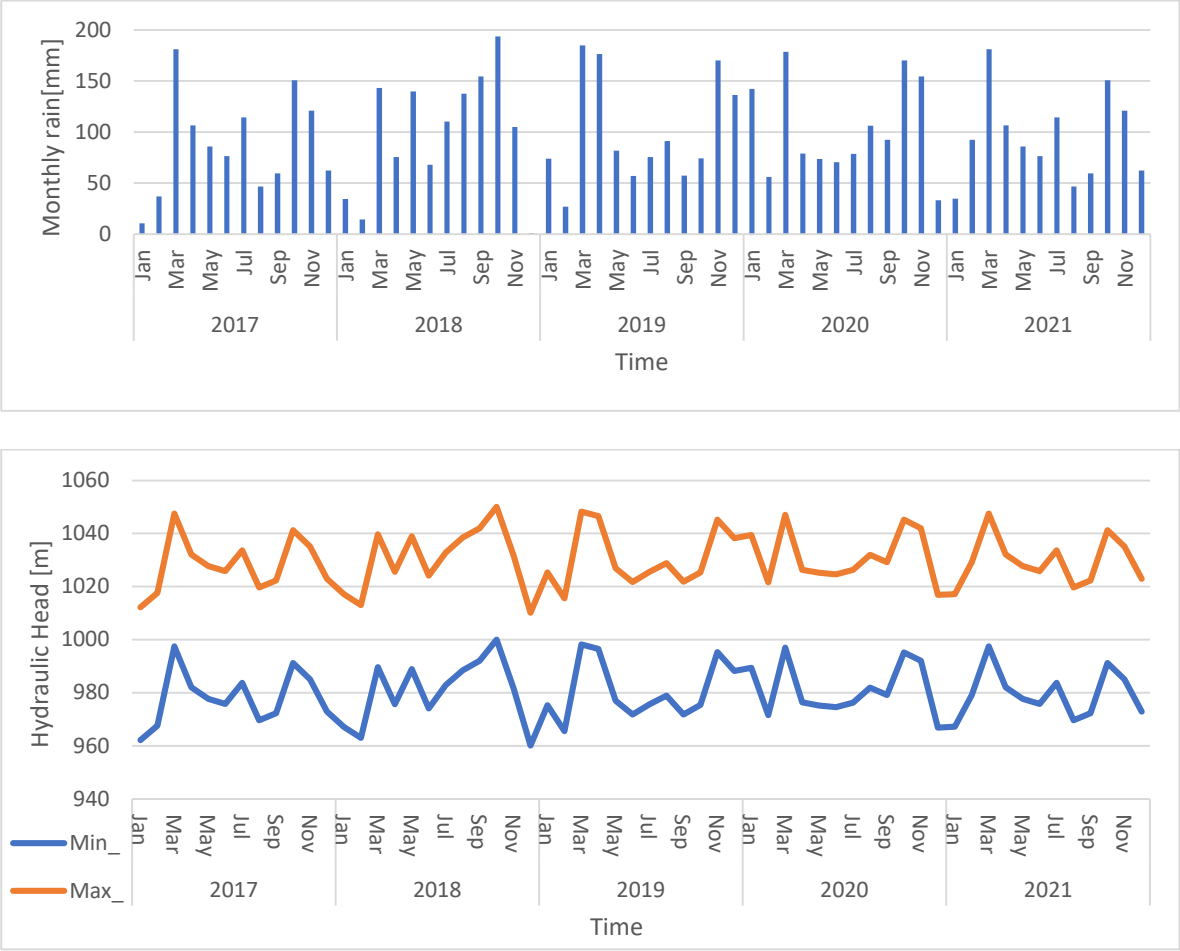


Lateral recharge from the Bucaramanga plateau is introduced by a general head BC (GHB) (Cauchy-type BC) that is typically used to simulate flow from a distant boundary located outside of the model domain (Anderson et al., 2015). This is done applying a conductance estimated by taking distance (approximately 2000 m from the original water source used for Antolínez Quijano, 2014), the model layer thickness, the cell size, and permeability from the initial value set in the model (Table *1*). Lateral recharge varies monthly, and proportional to monthly rain. Variability of head values in the source is around 1020 and 1040 (Antolínez Quijano, 2014) and were made proportional the maximum and minimum variation in monthly rain. The piezometric surface predicted by previous models around our proposed GHB shows a gradient correlated with the

gradient of the ground surface (Antolínez Quijano, 2014); because of this, the top surface of the model was used to correct the hydraulic head in the GHB toward lower values (Figure *11*).

Figure 11.

Range of monthly variations of hydraulic heads[m] in GHB proportional to rain. Top: monthly variations of rain[mm], bottom monthly variations of hydraulic heads in GHB



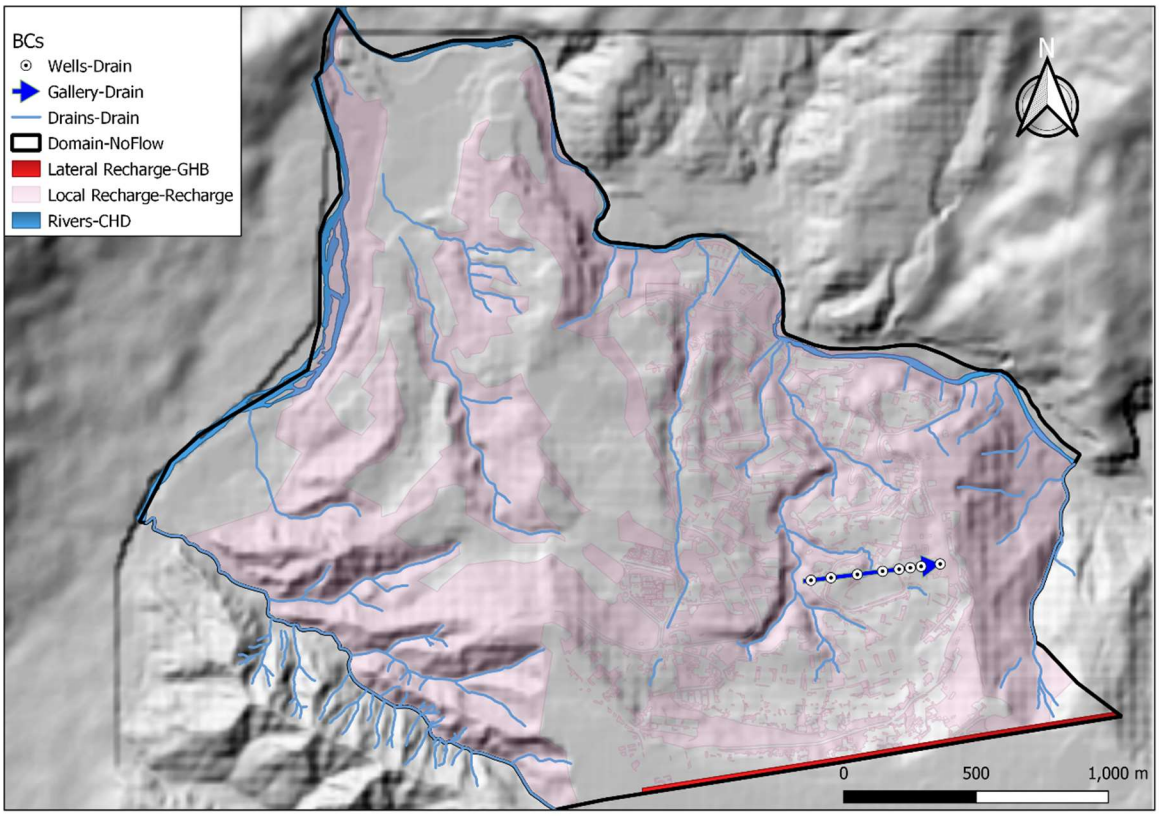
The gallery is modelled as a drain BC just like the gallery wells, using the detailed design schedule and drawings of the previous study (UIS & AMB, 2019). The typical section (

Figure 7) shows the geometry and number of drains that will control the conductance value. The drain condition is activated when the gallery in each cell is completely built, this is checked using the gallery construction rate (2.2 meters of gallery per day). Wells are also activated when they are completely built in the scheduled construction and are modelled as drain BCs; the bottom of the drain is set to be 2.5 meter above gallery top.

The gallery construction process must be modelled in transient state because the maximum flow happens when the gallery is completely built or before that. Besides, activating all the drain BCs of the gallery suddenly, would result in an overestimation of the maximum flows caused by the abrupt gradient change. Modelling the gallery only in steady state would show the flow in the long term, which is less than the maximum value.

Figure 12.

Boundary conditions



4.1.2 Horizontal grid discretization.

The model is built using the MODFLOW 6´s DIS package to configure an irregularly structured grid with a coarse discretization in areas away from the gallery, and a more refined grid at and around the gallery zone. The spacing transition between the refined and coarse grid was done using a smooth factor of 1.5 to avoid convergence problems. Schemes of model discretization are shown in Figure 13 and Figure *14*

4.1.3 Vertical discretization and hydraulic properties

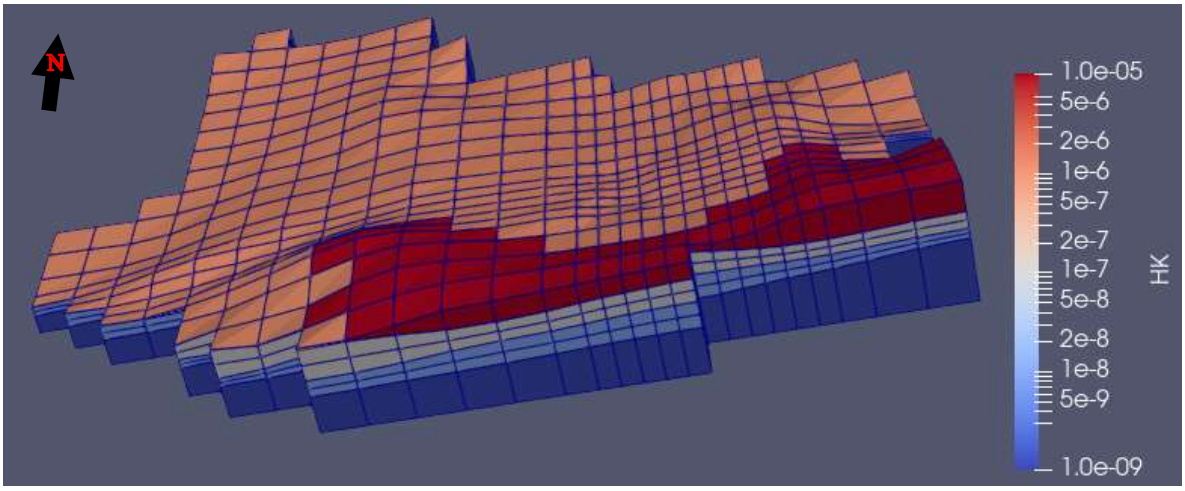
Model layers were defined according to the geological layers, but vertical gradients seen through piezometer data suggested that more vertical discretization was needed.; thus, geological

layers with many model layers were defined, using the same hydraulic conductivity which was assigned by employing the same external file. Additionally, a minimum thickness was introduced to avoid model errors by pinch outs. When a geological layer disappears, those zero-thickness cells are assigned as vertical through cells in the idomain package.

Four model layers were defined initially, but the first one is divided in two different geological units that are not continuous: Qd (orange color), a local colluvial layer, and Graboso member Qbg (red color) pinching out in the north of the study zone (Figure 13). Qbg can be important because it has an aquifer known as Superior Aquifer of Bucaramanga. The layers following vertically down are Qbo2 (upper Organos member of Bucaramanga Formation, grey color), Qbo1 (Lower Organos member of Bucaramanga Formation, light blue color), and rock (dark blue color).

Figure 13.

Simplified 3D model representing the initial model discretization and hydraulic conductivity assignment



Vertical permeability was defined as anisotropy Kv/Kh. Specific storage (Ss) and specific yield (Sy) were defined in the Storage package (STO) in the same way as hydraulic conductivity. Table *1* shows the hydraulic properties set at the beginning of the modelling process. These values were proposed by the author based on UIS & AMB, 2019. An example of the model setup is shown in Figure *14* with profiles and boundary conditions.

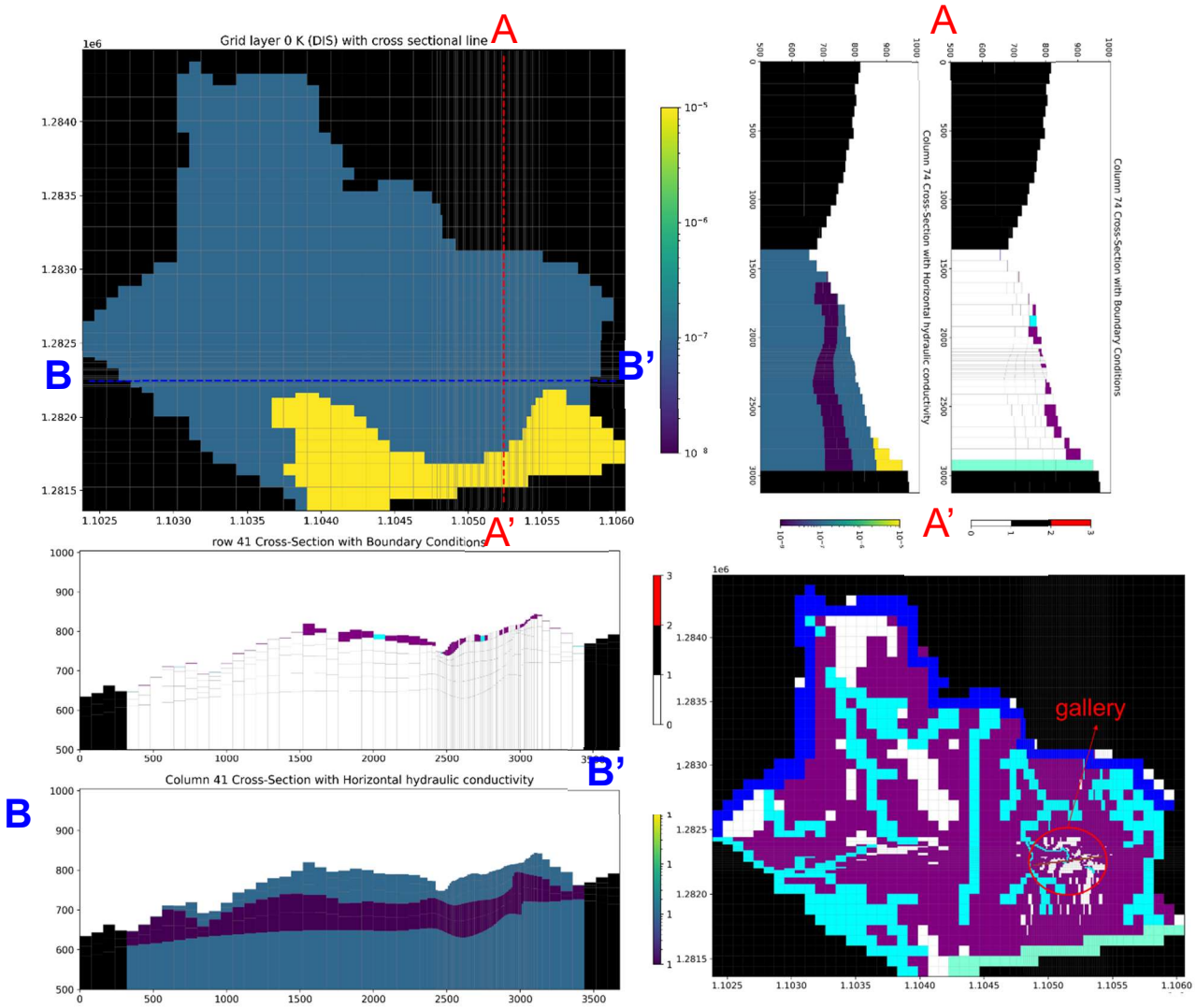
Table 1.

Initial hydraulic properties used for the geological layers of the model

Formation	k[m/s]	kv(anisotropy)	Ss	Sy
Qd	1.00E-07	1.00E-01	1.00E-04	1.00E-01
Qbg	1.00E-05	1.00E-01	1.00E-04	1.00E-01
Qbo2	1.00E-07	1.00E-01	1.00E-04	1.00E-01
Qbo1	1.00E-08	1.00E-01	1.00E-04	1.00E-01
rock	1.00E-09	1.00E-01	1.00E-05	1.00E-01

Figure 14.

Model structure scheme of the numerical groundwater model. Top-left plane view: hydraulic conductivities of the first model layer; bottom-right plane view: boundary conditions for the first model layer (recharge in purple, drains in cyan, GHB in aquamarine); top-right and bottom-left show model cross-sections A-A' and B-B', respectively.


Grid layer 0 K (DIS) with cross sectional line
A
A'
B
B'
row 41 Cross-Section with Boundary Conditions
Column 41 Cross-Section with Horizontal hydraulic conductivity
Column 74 Cross-Section with Horizontal hydraulic conductivity
Column 74 Cross-Section with Boundary Conditions
gallery
1e6

4.1.4 Time window and temporal discretization

The time window used for the modeling process is four and a half years, using daily stress periods, starting on the first day of 2017 and finishing in mid-2021. The first stress period was set as steady and de the following as transient (in the STO package) to use the steady-state model solution as initial condition for the transient time steps. Daily time steps of 86400 seconds (seconds in one day) were set for each transient stress period because the model is defined in international units (seconds). The gallery starts its construction in the third year and 3 months (starting on day 1185). The first 3 years of modelling are used to "heat up" an calibrate the model. Then recharge values used at the beginning are recycled and used again for the following time steps. Gallery construction takes place until day 1424. From there, their BCs will be active until the end of the simulation (finishing on day 1642).

4.1.5 Transparency and reproducibility

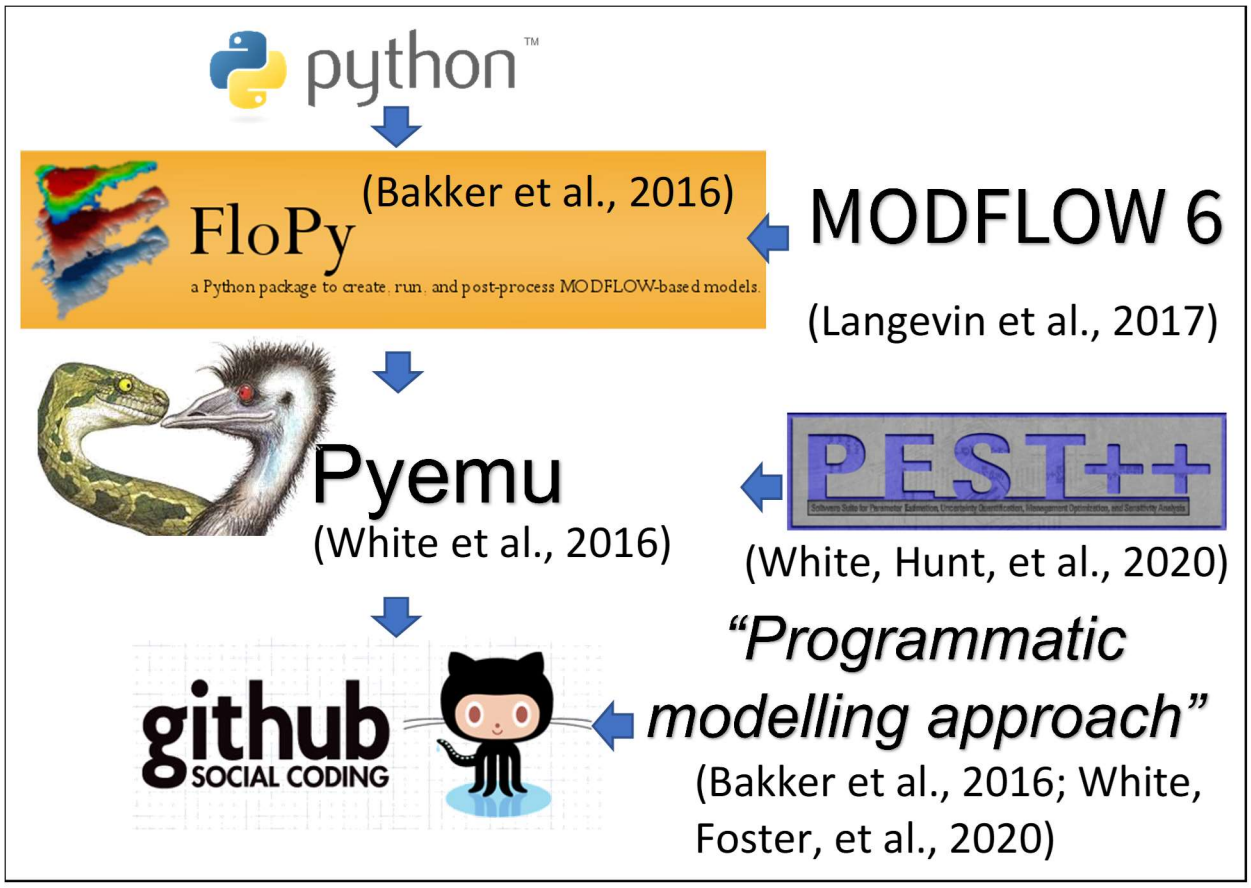
Reproducibility and transparency were intended throughout the process; this implied an important effort in the programmatic approach. We used Python with its libraries to build up the model and Git/Git Hub to document it, as outlined in Figure *15*. Open-source software combined with **Git** and **GitHub** for model documentation allows high transparency and reproducibility for anyone who wants to audit the model construction process and forecast uncertainty analysis. Python libraries are the interface for direct and inverse software modelling. Flopy (Bakker et al., 2016) handles MODFLOW 6 which solves the groundwater equations in steady and transient state.

Pyemu (White et al., 2016) is the python framework that interacts with Pest++, a software used for uncertainty analysis (White, Hunt, et al., 2020).

The philosophy of the model's construction follows an automatic approach to allow a constant review in model design. Python's Flopy library builds inputs for model runs taking GIS inputs to define extension, domain, boundary conditions and model layers. Cell sizes and the number of layers supplied in the code determine thicknesses and boundary condition (BC) cells by taking data from shapefiles and raster files and adapting them to the model estructure.

Figure 15.

Tool workflow used to guarantee reproducibility and transparency.



4.2 Prior numerical modelling

4.2.1 Prior uncertainty quantification [2]

Following the methodology of Hemmings et al., (2020) we developed a prior modelling to identify a suitable setting for the model to reproduce field observations and assess its effect in model forecast. Many parameters settings were tested to identify its effects on model outputs, especially flow forecast (White, 2017).

We defined parameters to be statistically modelled using Pyemu and Pest++. The Pestpp-ies executable was used with zero iteration (noptmax parameter in Pest control file) to develop the prior modelling; then, every input used here can be used as input in the iterative ensemble smoother (ies) process (process [4] in Figure 6).

The parameters selected to be statistically varied are horizontal hydraulic conductivities, vertical conductivities, specific storage, specific yield, general head conductance and recharge. To allow heterogeneity in realizations we used pilot points every 2 cells in hydraulic properties and 1 conductance per layer in GHB. A variability was assigned to those parameters to define a probability density function which determined the variability of the prior parameters.

4.2.2 Assessment of model structure [3]

We modified the model in search of a balance between simplicity and accuracy (John Doherty & Moore, 2020), avoiding a complex setting to attain data assimilation and evade prior data conflicts. After setting model complexity as “complex” in the MODFLOW 6 solver and the

statistical variability of parameters (Table *2*) in a pest++ file, we ran many Monte Carlo (MC) models with gaussian distributions of parameters and assessed the optimum configuration of the model.

Table 2.

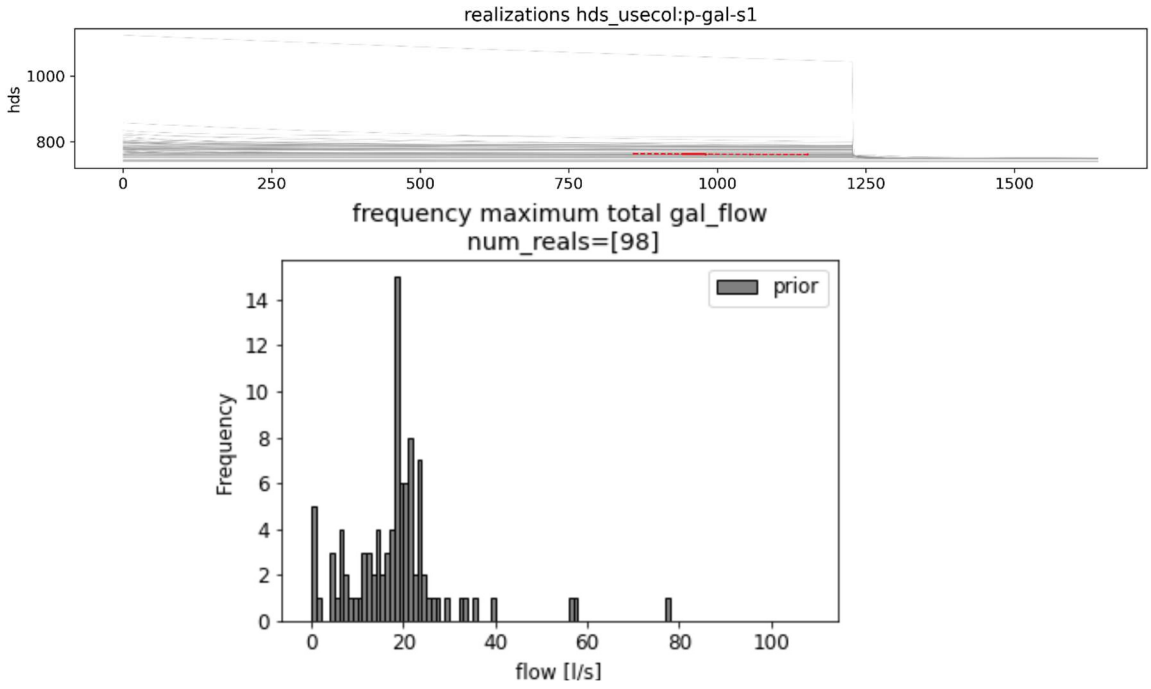
Variability of parameters. Initial high and low multipliers assigned to groups of parameters represent ±2 standard deviations of the initial values in a Gaussian distribution. The maximum and minimum bound restrict the values of possible parameters in the history matching.

Parameter group	Low bound multipliers	high bound multipliers	Minimum bound	Maximum bound
Horizontal hydraulic conductivity (HK)	0.1	10	1e-10	1E-4[m/s]
Vertical hydraulic conductivity (Kv/Kh) (VK)	0.1	10	100	0.01
Specific Storage (SS)	0.1	10	1e-6	1e-2
Specific yield (SY)	0.3	3	0.01	0.5
General head conductance (Cond)	0.1	10	0	100(virtually unlimited)
Recharge (Rech)	0.3	3	0	1100 [mm/yr]

We analyzed the results of prior MC modelling using mainly two types of graphics. The first type shows all the prior realizations of one observation of the model (and sometimes compared with a measured field value) (Figure *16*, top). The second shows the frequency of maximum flows yielded by the gallery (Figure *16*, bottom). When measured field values are outside of the ensemble of prior modelled values (prior data conflict), the structure of the prior model must be reviewed to avoid biases generated by history matching compensation. This prior data conflict can be neglected if the misfit is in non-important observations, which are those deemed uncorrelated with the model forecast (i.e., observations far from the gallery or very deep).

Figure 16.

Example of output graphs generated by the prior Monte Carlo analysis used to assess model structure. Top: ensemble of results of one model observation (grey) compared with the correspondent field observation (red). Bottom: histogram showing the distribution of maximum flows in the stochastic realizations.



4.3 History Matching and uncertainty quantification.

We compared model observations with available piezometer data at points where this information was reported from previous studies (Medina Baez, 2021; UIS & AMB, 2019) (Figure *17*), vibrating wire piezometers have more than one observation as is shown in Table *4*. These points are mainly located at or around the gallery zone. Model observations were placed also in the gallery and its wells to monitor the behavior of flows in those BCs. Although field observations

were used in the prior modelling, they were used only to identify prior data conflicts. In this phase those observations were used for data assimilation, which allowed updating the initial parameters.

The history matching process requires a model observation for each field observation, except for the boundary conditions where observations are used to monitor the flows. Also, for each field observation we set up a measurement error (Table *3*) to indicate measurement precision (reliability) and weights (relative importance) in the history matching process.

Table 3.

Measurement error of every observation expressed as a standard deviation

Observation	desvstd	Observation	desvstd
P-Gal-S1	0.2	P-RN-S7_1	1
P-Gal-S2	0.2	P-RN-S7_2	0.7
P-Gal-S3	0.2	P-RN-S7_3	0.5
P-Gal-S4	0.2	P-RN-S7_4	0.5
P-Gal-S5	0.2	P-E3-S6_1	0.7
P-Gal-S6A	0.2	P-E3-S6_2	0.5
P-LZ-4	0.2	P-E3-S6_3	0.5
P-M-6	0.2	P-E3-S6_4	0.5
P-MS-1	0.2	P-E3-S6_5	0.5
P-MS-2	0.2	P-VR-S_1	0.5
P-VR-1	0.2	P-VR-S_2	0.5
		P-VR-S_3	0.5

The parameters modified in the history matching process are the same as above (section 0); the parameterization structure is shown in Table *5*. To allow model heterogeneity but reducing the number of adjustable parameters, pilot points were configured in some parameters using Kriging and semi variograms to express covariance between close spatial parameters. The semi variogram used was exponential with a range of 3 times the maximum separation of the pilot points (6 times the separation of the outer refinement). The drain conductance was not set as an adjustable

parameter; instead, it was set as a high value, making the hydraulic conductivity in the area control the flow to the BC (Anderson et al., 2015).

Figure 17.

Model observations used for history matching and BCs

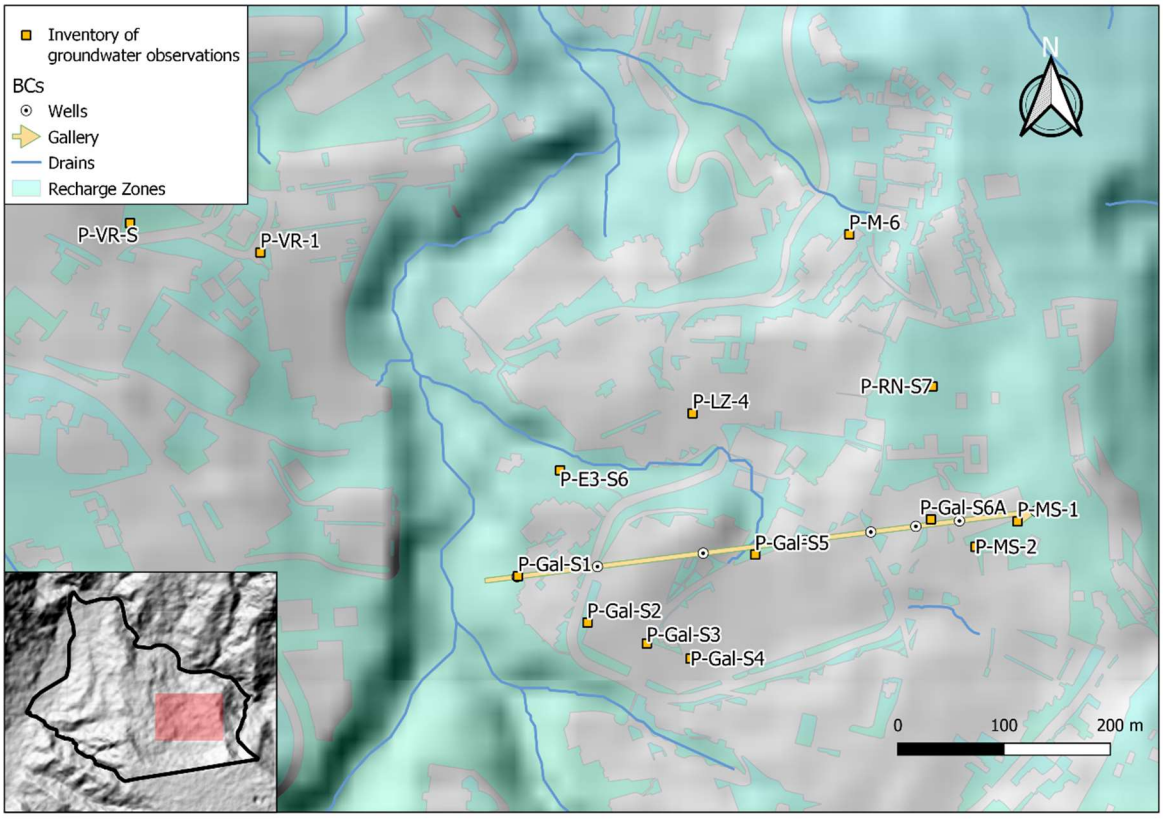


Table 4.

Depth of observations (m) in vibrating wire piezometers

Piezometer	Sensor Depth (m)
P_RN_S7_4	10.5
P_RN_S7_3	40.5
P_RN_S7_2	91.5
P_RN_S7_1	121.5
P-VR-S_3	26.5

Piezometer	Sensor Depth (m)
P-VR-S_2	40.5
P-VR-S_1	77.5
P-E3_S6_5	9
P-E3_S6_4	23.5
P-E3_S6_3	35.5
P-E3_S6_5	50.1
P-E3_S6_1	141.5

Table 5.

*Settings for the adjustable parameters in history matching. *Constant parameters are those that don't have horizontal heterogeneity.*

Parameter Name	Parameter type	External files to modify
Horizontal hydraulic conductivity (HK)	Pilot Point	Geological layers - HK
Vertical hydraulic conductivity (VK)	Pilot Point	Geological layers - VK
Specific Storage (SS)	Pilot Point	Geological layers - SS
Specific yield (SY)	Pilot Point	Geological layers - SY
General head conductance (Cond)	Constant*	Model layers - Cond
Recharge (Rech)	Constant*	One Median Recharge value - Rech

4.3.1 Iterative ensemble smoother (IES) [4]

Once a suitable model configuration was defined, we proceeded to execute data assimilation by iterative ensemble smoother (IES) and assess forecast uncertainty at the same time. We did it using Pestpp-ies which implements a reformulated Gauss-Levenberg-Marquardt algorithm. The latter uses a Jacobian matrix derived empirically from an ensemble of random parameter values (White, 2018). The empirical Jacobian matrix then is used to update the

parameters of the ensemble of realizations. The mathematical formulation is described in Equation *7*.

$$\Delta_\theta = -\left(\left(J_{emp}^T \Sigma_\varepsilon^{-1} J_{emp}\right) + (1+\lambda)\Sigma_\theta^{-1}\right)^{-1} \left(\Sigma_\theta^{-1}(\Theta - \Theta_0) + J_{emp}^T (D_{sim} - D_{obs})\right)$$

$$J_{emp} \approx \Sigma_\varepsilon^{\frac{1}{2}} \Delta_{sim} \Delta_{par}^{-1} \Sigma_\theta^{-\frac{1}{2}}$$

$$\Delta_{sim} = \Sigma_\varepsilon^{-\frac{1}{2}} (D_{sim} \ominus \bar{D}_{sim}) / \sqrt{N_e - 1}$$

$$\Delta_{par} = \frac{\Sigma_\theta^{-\frac{1}{2}} (\Theta \ominus \bar{\Theta})}{\sqrt{N_e - 1}}$$

Equation 7. GLM and IES formulation for ensembles

Where:

D_{obs} , D_{sim} are the observation and simulated-equivalent ensembles N_exN_{obs}

Θ, Θ_0 are the current and initial parameter ensembles N_exN_{par}

Δ_θ is the parameter update matrix

J_{emp} is the empirical Jacobian

$\bar{D}_{sim}$, $\bar{\Theta}$ are the mean values of the simulated equivalents to observations and parameter across their respective ensembles

N_e is the number of realizations

$\ominus$ denotes a broadcast subtraction operation

λ The Marquardt dampening parameter

Σ_θ is the prior parameter covariance matrix

Σ_ε is the measurement noise covariance matrix

Δ_{sim} is the change in simulated values

Δ_{par}^{-1} is the change in the parameters

In accordance with Hemmings et al., (2020), an abridge history matching is desirable before a complete data assimilation because sometimes there is not much improvement in the model forecast, or just because the modelling objective could be already achieved with one iteration. Then we tested if the field data contains valuable information for the model forecast with one iteration of IES and checked results.

Most of the parameters and settings used for IES (Equation *7*) are the same as those used in the prior analysis of chapter 0 since the pest++ executable is the same, and parameters are almost the same for both analyses. The covariance matrix of observations was defined automatically by PEST++ using measurement errors. The number of realizations was initially set to 100 but was updated according to our findings. "Autodataloc" was used as true to use the automatic functionality of pest to define the localization matrix. To deal with the problem of excessive run times caused by random parameters, if an overdue simulation has already taken more than twice the average run time, it is declared failed by default (White, 2018). Because of this, in every iteration there will be a different number of successful models, different to the initial number of realizations, because there are failed and delayed models.

Finally, we developed a complete history matching coupled with uncertainty quantification with many iterations where we sampled a posterior distribution of parameters and the flow forecast.

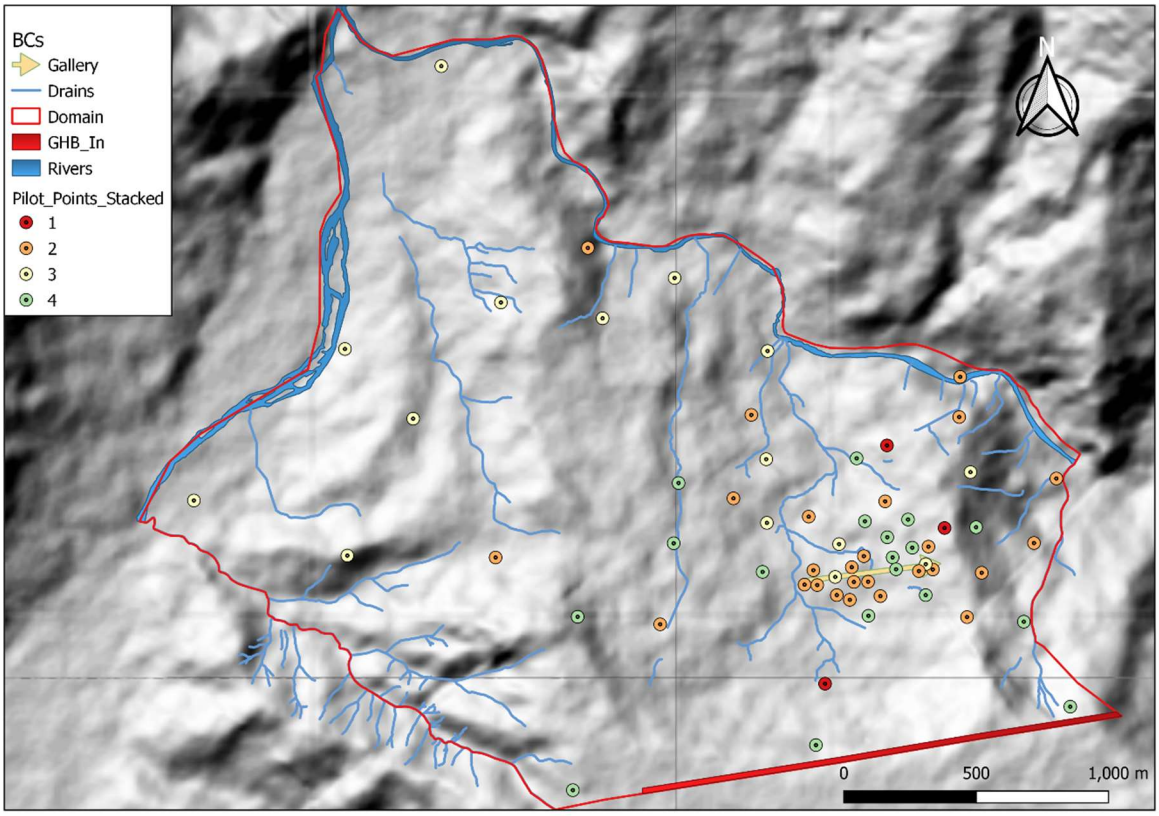
4.3.2 Model Calibration [5].

We optimized the model parameters to get the "most probable" unique model with the minimum heterogeneity necessary to match field observations. The optimization is carried out with Pestpp-GLM which implements a subspace Gauss-Levenberg-Marquardt(GLM) algorithm (White, Hunt, et al., 2020). Pestpp-GLM uses a hybrid regularization approach joining the truncated singular value decomposition (TSVD) and the Tikhonov regularization (M. J. Tonkin & Doherty, 2005).Also, to perform model calibration, the number of parameters must be lower compared with those used for IES, given that Levenberg Marquardt depends on a Jacobian matrix whose size its proportional to the number of parameters and it is more computationally constrained (White, 2018).

The pilot point automatic separation feature included in Pyemu was not used because the irregular cell width generated big holes without pilot points and the range of the semi variogram was changed to 8 times the maximum cell distance (960 m). Pilot points were manually put in the positions shown in Figure 18 but not equally located in every layer because each layer has different extension, observations, and importance due to its interaction with the forecast. Regularization conditions are imposed over all the parameters and set as a *zero order Tikhonov regularization*; that is, initial values are the preferred values. The *Sy* in the 2 deepest geological layers and the *Ss* parameter were not defined with pilot points, but as a constant parameter per geological layer to reduce the number of parameters and the computational burden of model calibration. This decision helped to decrease the parameterization because it was mandatory reducing the run time and they were considered less important than others, especially less important than Sy in the first two geological layers (most of the flow of the gallery is considered to come from its pore dewatering).

Figure 18.

Pilot point locations in the calibration process. The color of the pilot points indicates if they are located in one (red), two (orange), three (yellow) or four geological layers (green).



Most of the settings of the previous analyses are used to set up a *Pestpp-glm* run but in GLM it is necessary to ensure the high fidelity of the finite difference partial derivatives (Jacobian matrix) (White, 2018). To do so, it was necessary to take care of the solver criteria and change the model structure to avoid low quality sensitivities and model failures. The essential change of the base model was the Newton Raphson approach, but this more rigorous approach exposed the numerical weaknesses of the model and led to its restructuring as shown below.

The important changes in the solver and model structure were:

- Newton Raphson approach

- This method has a much better handling of dry cells and avoids big values yielded in observations located on dry cells.

- Adaptative time stepping
  - This allows transient stress periods to change their time steps if the model does not converge. The adaptive time stepping can improve dramatically stress period convergence, but it usually implies longer run times.

- Solver settings
  - Change from high complexity to a more customized approach that fits the model characteristics and ensures model convergence with acceptable run times.

- Lower conductance of drains
  - Conductance of drains is a relatively unknown value. DRN BCs are designed to have lower permeability than its surrounding cells, and high values can cause model instabilities.

- No thin cells (less than 5 meters in thickness)
  - Cells that change from fully saturated to dry and vice versa easily can cause convergence problems.
  - High slopes combined with thin cells cause adjacent cells not touching each other, thus violating control volume conditions. No thin cells alleviates stability problems in the solution.

- Smoother topographic raster.
  - A smoother topographic raster avoids large changes in model top elevations causing adjacent cells not touching each other, thus violating control volume conditions.

- Refinement of coarse areas:
  - A more refined surface in coarse areas avoids large jumps in zones with steep topography.

- Change in the initial heads
  - To make convergence easier, it is better to start close to the solution. Then a python function was included in every run to replace the starting heads with the solution of the last successful model.

The changes mentioned above were defined following an empirical approach, that is, testing the convergence of many models and analyzing their behavior. The models were selected by running *PEST++* with random parameters and selecting failed models (usually 10) to test settings to make them converge. Although some settings could help to achieve convergence in more models, they could lead to high running times, even in models where convergence was easily achieved. Therefore, the final settings were defined to succeed in around 70% of the stochastic models in reasonable times.

The number of total iterations was set to 4 (NOPTMAX). To accept the FOSM-realization if the *phi* is lower than the lambda-testing phi, *Glm_accept_mc_phi* was defined as TRUE (White, Hunt, et al., 2020). This hybrid approach helps to avoid local minima and speed up the model inversion.

4.3.3 Linear analysis [6]

Once each iteration of model calibration is finished, Pestpp-glm uses the Jacobian matrix and the covariance to continue with an uncertainty quantification of all the observations, including observations to assess the gallery behavior.

This is done using the first order second moment method called *Shur's complement* (Fienen et al., 2010; White, Hunt, et al., 2020). Shur's complement implemented in Pestpp-GLM assumes

that the standard deviation of measurement noise associated with each observation is proportional to the current observation residual. Then it calculates a posterior parameter covariance matrix and the posterior uncertainty estimates for forecasts (Equation *8*).

$$\bar{\Sigma}_\theta = \Sigma_\theta - \Sigma_\theta J^T [J \Sigma_\theta J^T]^{-1} J \Sigma_\theta$$

$$y = \frac{\partial s}{\partial \theta}$$

$$\sigma_s^2 = y^T \Sigma_\theta y$$

$$\bar{\sigma}_s^2 = y^T \bar{\Sigma}_\theta y$$

Equation 8. Shur's complement equations

where,

$\Sigma_\theta, \bar{\Sigma}_\theta$ Prior and posterior covariance matrix

J Jacobian Matrix

σ_s^2, $\bar{\sigma}_s^2$ Prior and posterior uncertainty estimates for forecast **s**

y vector of forecast sensitivity to each parameter.

4.3.4 Post calibration Monte Carlo analysis [7]

Pestpp-GLM uses in every iteration the Null Space Monte Carlo (NSMC) method (by using "glm_iter_mc = true") to identify model forecast variations strongly related with a lack of field data informing model parameters and forecasts.

The process is described in detail in M. Tonkin & Doherty, (2009) and it is summarized below:

- Calculate Base parameter sensitivities (Jacobian),

- Undertake SVD decomposition of the Jacobian matrix to identify solution and null space,
- Produce multiple stochastic realizations of model parameters,
- Project differences between the realization and the base model (Calibrated parameters until that iteration) onto the null space,
- Add projected differences to the base (or calibrated) parameter field,
- Use each parameter set to run the model and assess uncertainty.

In case the objective function of any realization is lower than that achieved by GLM, that realization is taken as the new updated model instead of the one obtained with the optimization by the gradient of the target function.

4.4 Comparisons and Analysis [8]

We compared the different methods based on their ease of application, types of results, and computational burden. The latter is especially important because this research is carried out with the limited resources of a desktop personal computer. Also, we assessed the complementarity of methods and analyzed advantages of the implementation of recent approaches as Early Uncertainty Quantification and Forecast First. Furthermore we compared uncertaity results with the decision threshold of 30 l/s which is the lowest flow used for the hydraulic design of elements in the gallery (UIS & AMB, 2019).

5. Results

5.1 Model repository and reproducibility of results.

Several setups were tried according to the model forecast results. All the code and history tracking are stored in a public GitHub repository. Access to the information is possible through the link **https://github.com/oscarfasanchez/Thesis_mf6_pest.** The person interested in revising the repository should have basic knowledge of Python, Git, Modflow and Pest++. Running the code for reproducibility purposes requires that the computer has python packages installed and pest++ executables. The versions used for the model runs are Python 3.8.5, Pyemu 1.2.0, Flopy 3.3.6, MODFLOW 6.3.0, Pest++ 5.1.23. Detailed results can be found in Appendix B

The GitHub repository contains the python files in the 09_Python folder; the two main files are "modelo.py" which builds the base model, and "Pest_inv_unc.py" which runs history matching and uncertainty quantification. Binaries and other types of files necessary to run the scripts, are available through links in the README.md file (a git practice is only track plain text files, not binaries).

5.2 Assessment of model structure

The assessment of model structure through prior uncertainty quantification tested the refinement of a general coarse grid, especially, in areas close to the gallery. We used three different grid structures, namely, a coarse, a medium, and a fine resolution (Table 6), finding many problems

in the coarse resolution in terms of scarce match with field observations which led to the medium refinement. The medium refinement overcame most of those problems, but the overestimation of maximum flows led to the fine grid.

Table 6.

*Grid refinement of the model structure and history matching parameterization * the fine model with the Newton-Raphson formulation is not used until the model calibration in the GLM approach*

Grid Refinement	Total Parameters	#pp lay0 qd_qbg	#pp lay2 qbo2	#pp lay3 qbo1	#pp lay4 rock	Outer grid spacing [m]	Gallery Grid spacing[m]	Run time in base model
Coarse	1544	96	96	96	96	180	80 m	45 seconds
Medium	17160	1072	1072	1072	1072	80	10 m	8 minutes
Fine	89416	5588	5588	5588	5588	80	3 m	40 minutes
Fine NR*	461	54	49	33	36	40	3	72 minutes

The prior Monte Carlo modeling with the coarse grid produced results where field observations fell inside the ensemble of prior realizations (Figure 19 and Figure *21*) and outside of it (Figure 20 and Figure *22*). The comparison between head field measurements (red) and heads of prior realizations (gray) in these figures helped identifying structural problems of the model.

Figure 19.

Example of field observations inside prior head realization with coarse grid, using the initial parameter range

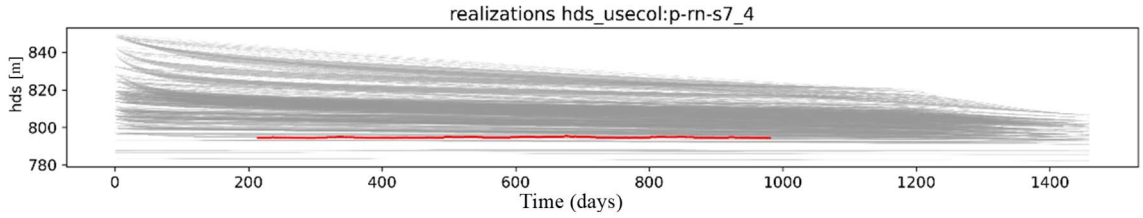
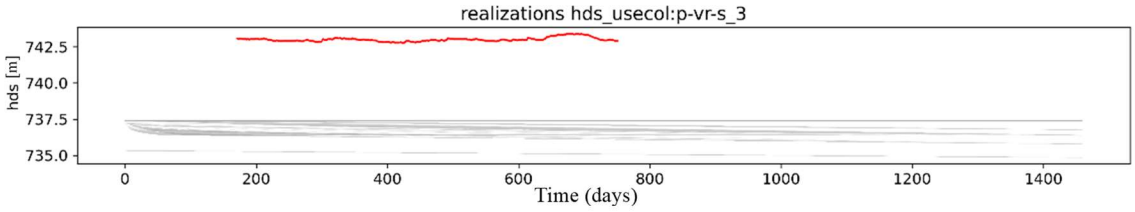


Figure 20.

Example of field observations outside prior head realizations with the coarse grid, using the initial parameter range



The offset between the observed heads and the ensemble of prior realizations in Figure 20 suggested that we needed to update the recharge boundary conditions (

Table **7**) and hydraulic conductivity bounds (Table 8) aiming to loosen the range of possible inflows and conductivities in the model. Both tables show the initial and updated multipliers.

Table 7.

Upper and lower bound multipliers of recharge boundary conditions after and before the update of the coarse model

	ghb_bound	rch_bound
initial	[0.1,10]	[0.3,3]
updated	[0.01,100]	[0.25,4]

Table 8.

Upper and lower bound multipliers of hydraulic conductivities and mean hydraulic conductivity of rock (bottom layer) in the coarse model

	hk_bound	K_roc
initial	[0.1,10]	1e-09
updated	[0.005,500]	1e-7

These changes were successful in removing the offset in many cases (Figure *21*) but, for some important observations (those deemed correlated with the model forecast), the ensemble of models still could not represent them (Figure *22*). The gallery effect in p-gal-s1 observation point can be seen in time ~1250 (Figure *21*).

Figure 21.

Example of field observations inside prior head realizations for the coarse grid, using the updated parameter ranges

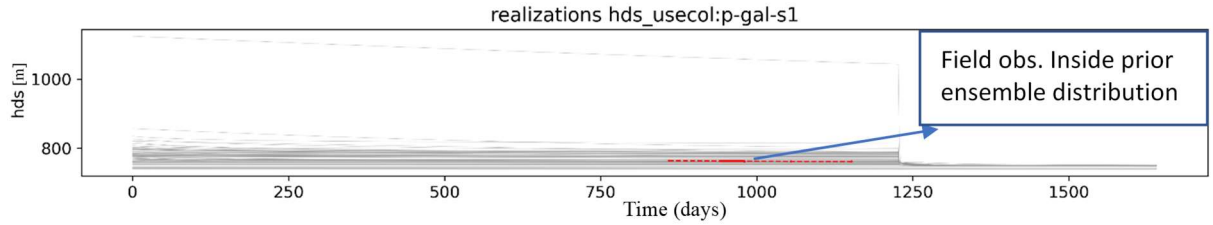
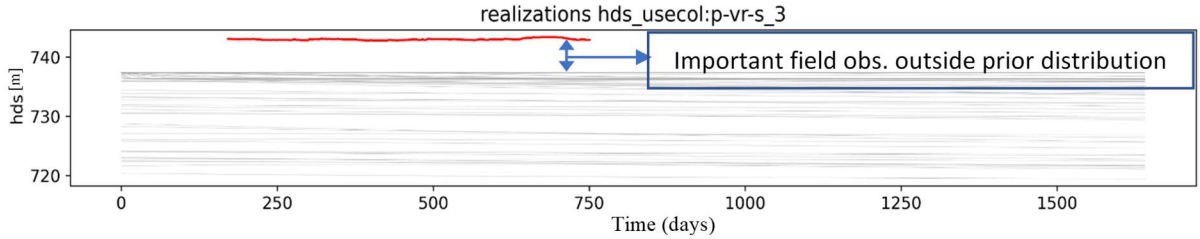


Figure 22.

Example of important field observations outside prior head realizations for the coarse grid, using the updated parameter ranges



Finally, using the medium refinement solved the disability to match important observations (Figure 23), although some non-important are still outside the ensemble of realizations (

Figure ***24***).

Figure 23.

Example of field observations inside the prior ensemble due to parameter update and change to a medium grid

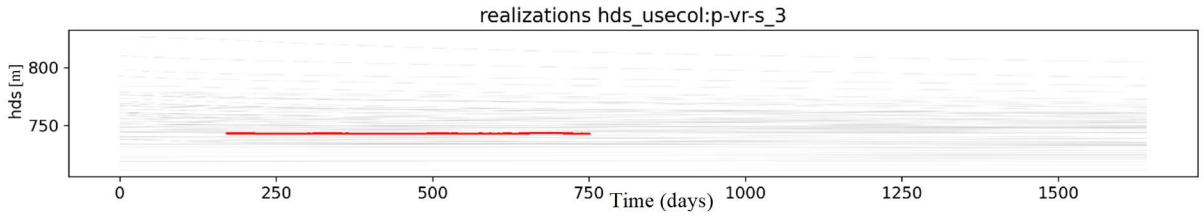
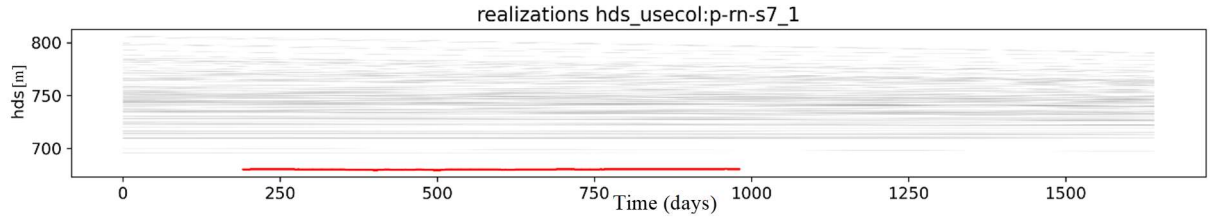


Figure 24.

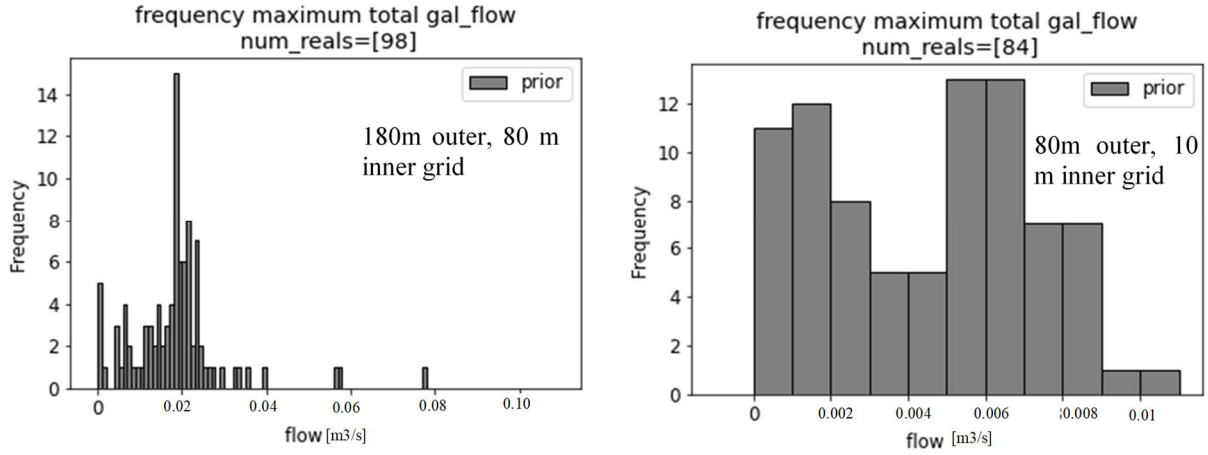
Non-important (very deep) field observations not matching ensembles in the medium grid model



The distribution of gallery's maximum flows with the medium grid is much narrower than in the coarse (updated) model (Figure *25*).

Figure 25.

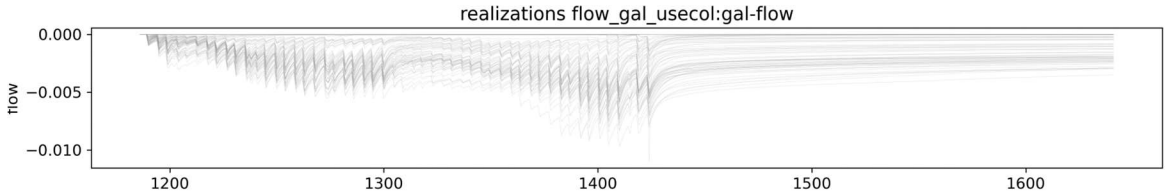
Frequency distribution of maximum flows for the coarse (left) and medium (right) grid resolution. running initially 100 realizations



Prior realizations of gallery flow for the medium grid show unnatural strong peaks related to the activation of the boundary condition in every cell (Figure 26).

Figure 26.

Realizations of gallery flow in the medium grid model.

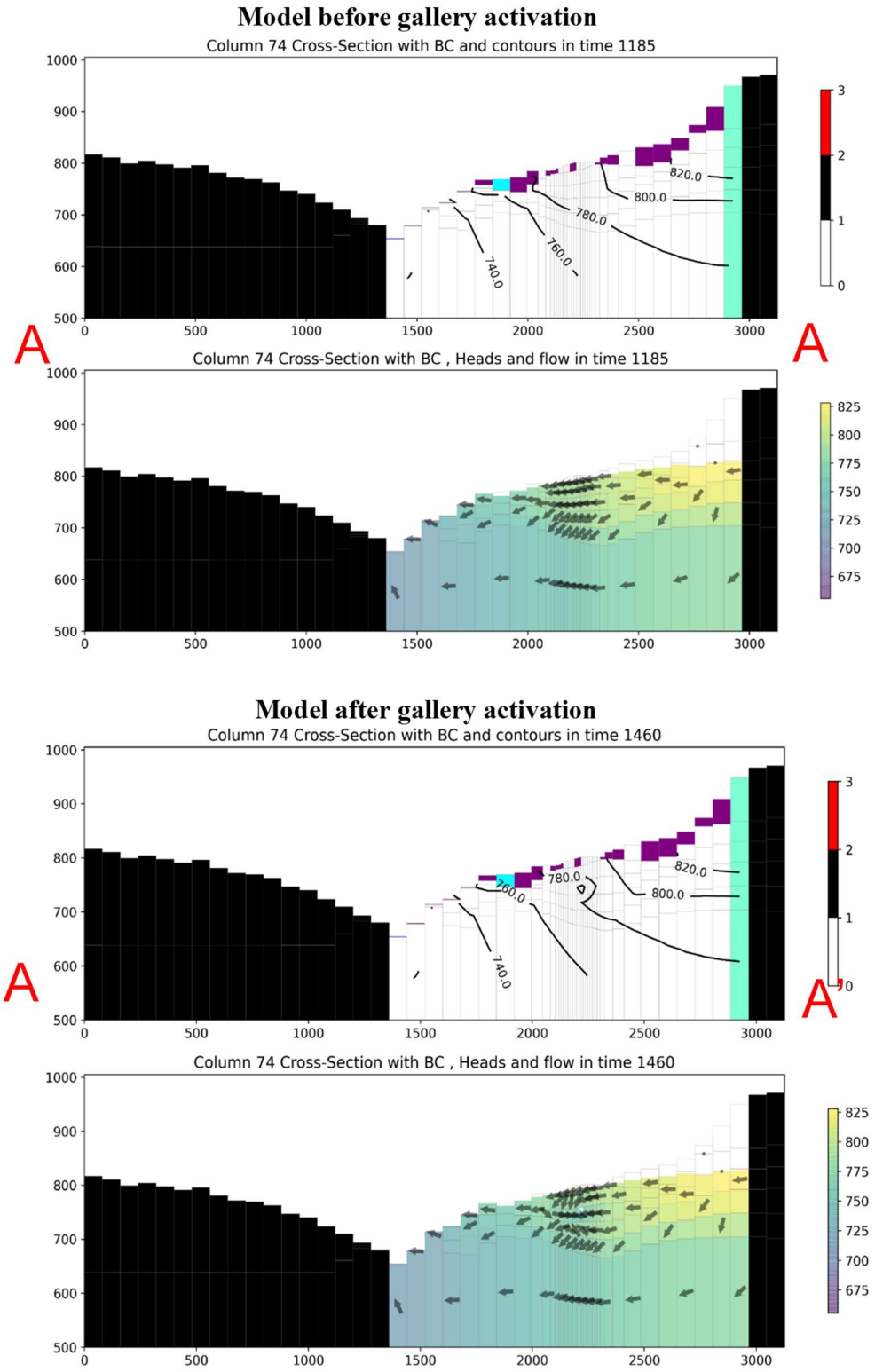


The effect of the gallery construction on the groundwater system at cross section A-A', using the medium grid base model is shown in

Figure 27. Arrows indicate flow direction and colors show hydraulic heads. The difference between the upper and lower panels relates to the effect in flow and head after gallery's activation where flow goes toward the gallery and drawdown occurs close the gallery.

Figure 27.

Groundwater heads and flow before (top panel) and after (bottom panel) gallery activation.

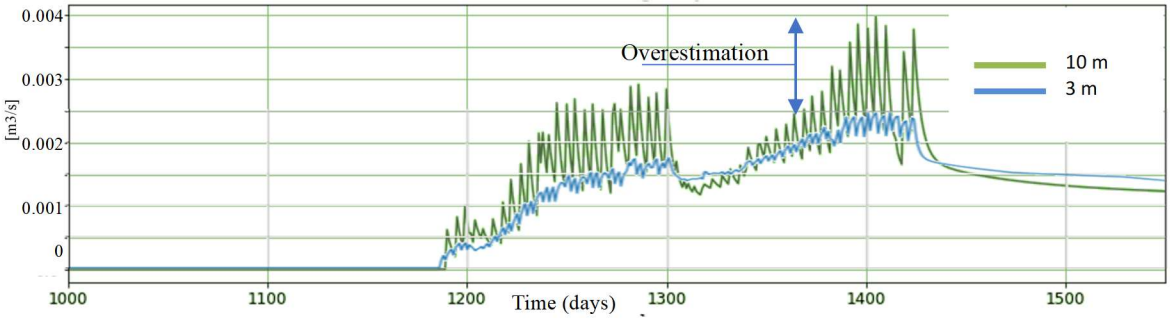


In the bottom figures of each panel, arrows indicate flow direction and colors show hydraulic heads. In the upper figure of each panel color indicate BCs: recharge in purple, drains in cyan, GHB in aquamarine.

Comparison of the base realization of the coarse (green) vs. fine (blue) grid, confirms that the medium mesh is overestimating maximum flows (Figure 28). The coarse (10 m) grid model has strong peaks due to lack of continuity in the activation of BC (DRN cell is activated every 3-4 days).

Figure 28.

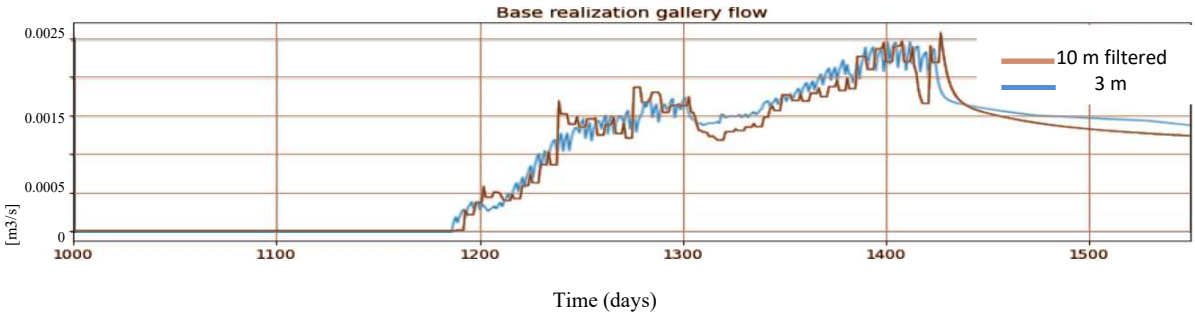
Comparison of flows between medium and fine base models



However, the performance of the medium grid model (brown) gets closer to that of the fine grid model (blue) if we change the flow of each day for the minimum flow of the previous 4 days (Figure 29).

Figure 29.

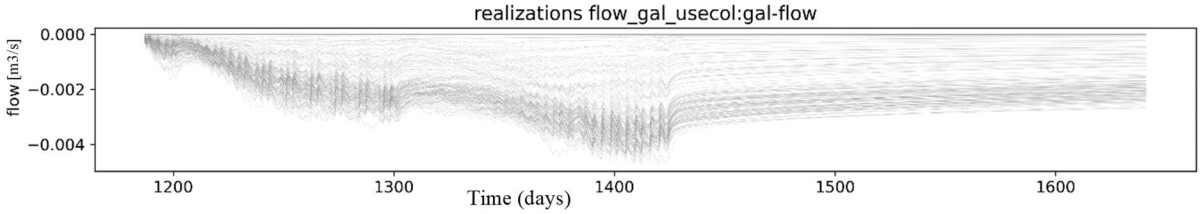
Comparison of flows between medium (brown) and fine (blue) base models, using the minimum value of the previous 4 days in the medium discretized model



In the ensemble of flows (m^3/s) for the fine grid model (Figure 30) it can be seen that all the flow lines concentrate in the higher values, and the maximum flows do it above 2.5 l/s in time around 1400 days. Also, strong peaks have been softened compared with those of the medium model.

Figure 30.

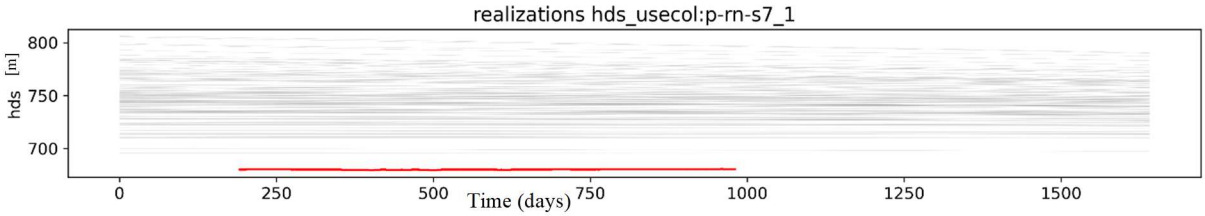
Prior realizations of gallery flow (m^3/s) in the fine grid model, negative values denote groundwater outflow in the MODFLOW convention.



As in the previous model there are some non-important observations than our model cannot simulate (Figure 31) and are the same observations that the previous model was not able to reproduce.

Figure 31.

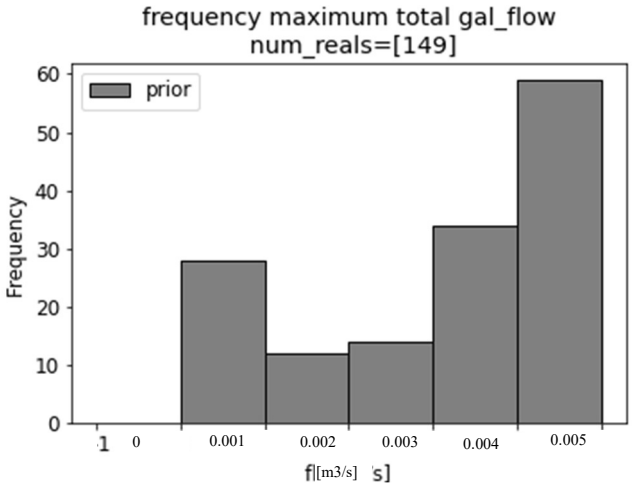
Example of non-important (very deep) field observation not matching ensembles in fine model



The histogram distribution of maximum flow (Figure 32) consolidates as a Non-gaussian distribution in the fine model, with peaks in very low flows (less than 1 l/s) and around 4-5 l/s.

Figure 32.

Frequency distribution of prior maximum flows for the fine grid model, running initially 150 realizations



5.3 Iterative Ensemble Smoother (IES)

Prior distributions are the base for one iteration of IES; nevertheless, we generated new prior ensembles for the analysis of one or more iterations of IES. This was done for the medium and fine models. In general, although there is evidence of data assimilation in the ensemble of models, the forecast is mildly affected by the history matching (i.e. the calibration improve but the flow forecast doesn't change), especially after the first iteration of IES.

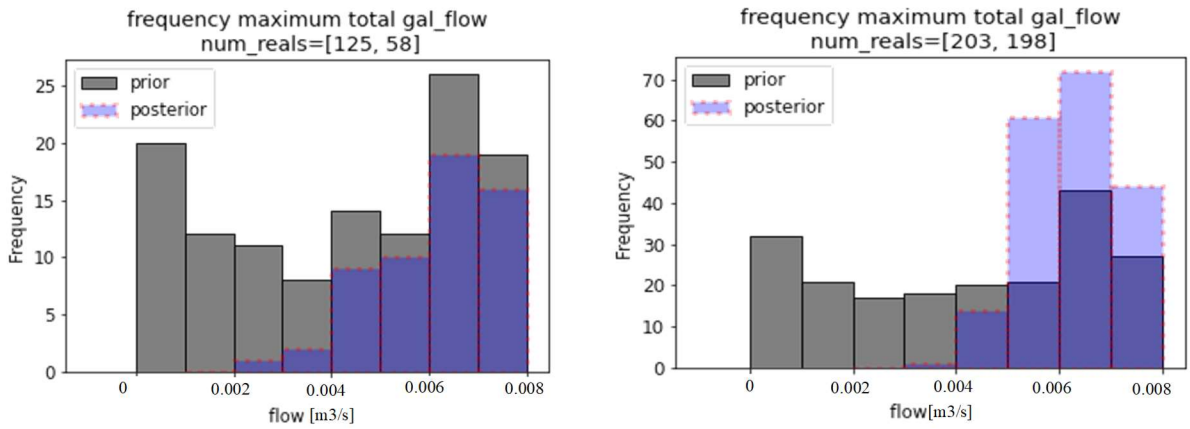
For the case of the medium grid model, IES helps to assimilate data, but we started to see symptoms of instabilities in the model. This is evidenced by the loss of more than half posterior realizations due to model failures [125 vs 58] in the prior and first iteration, respectively (

Figure ***33***, left). Although a normalization could be performed to mitigate the loss of information, it could lead to bias. Thus, we decided to undertake a new attempt with more realizations to mitigate model failures. This led us to repeat the prior calculation with more realizations (250 instead of 150). In the new attempt we saw fewer losses in the first iteration with respect to the prior and a convergence trend around 6-7 l/s (

Figure ***33***, right).

Figure 33.

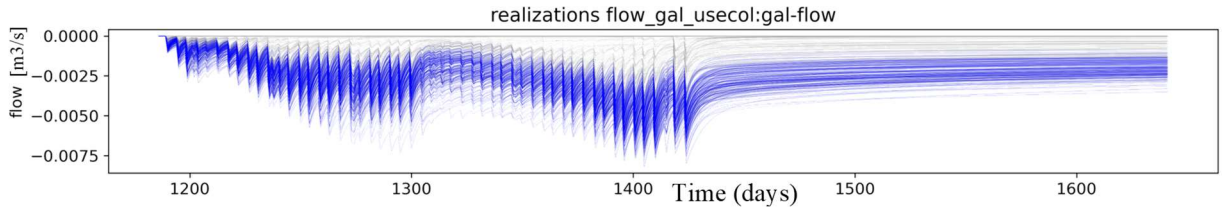
Frequency distribution of maximum flows for the first iteration of the medium grid model, running initially 150 realizations(left), and 250 realizations(right). In brackets is shown the number of successful realizations in the prior and the first iteration



The posterior distribution of flows (blue) falls inside the prior distribution (gray) with a trend of the posterior distribution of flows to be narrower than the prior and concentrated towards higher flow values (Figure 34).

Figure 34.

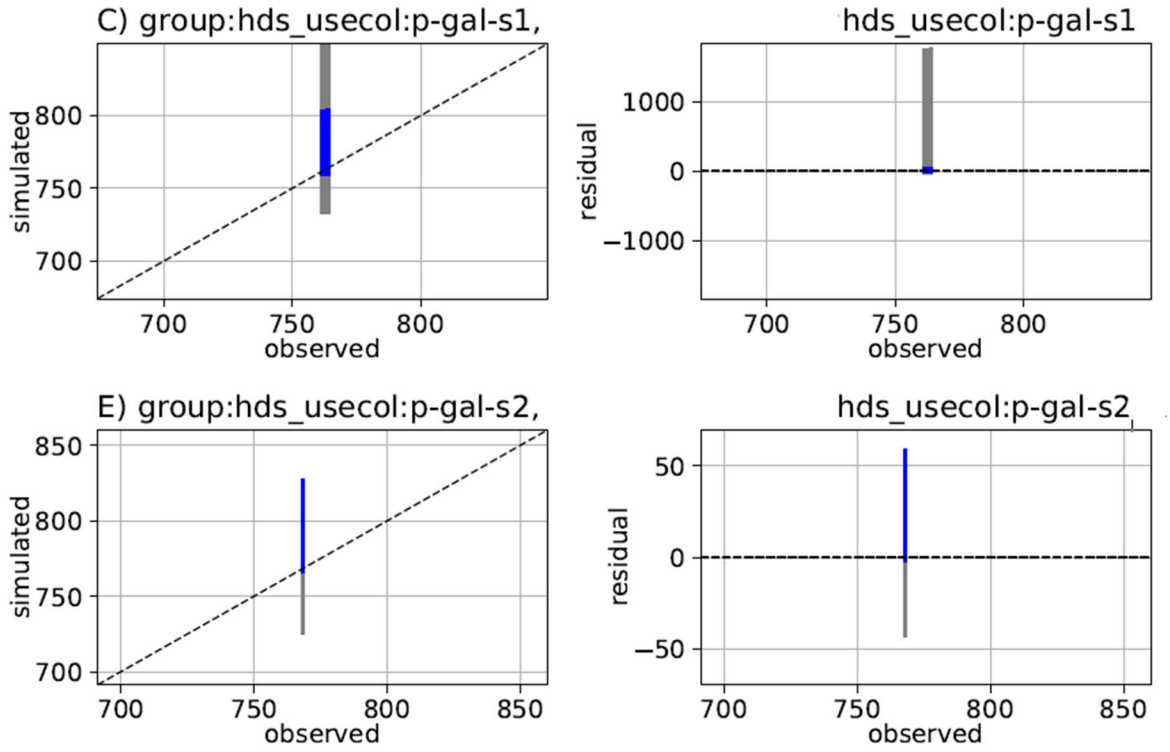
Prior (gray) and posterior (blue) realizations of gallery flow(m^3/s) in the medium grid model, negative values denote groundwater outflow in MODFLOW convention.



The posterior distribution of hydraulic heads shows an improvement of fit compared with the prior(closeness to 1-1 diagonal in Figure *35*, left), represented by lower residuals(simulated minus observed values)(Figure *35*, right graph), also can be noticed that posterior values are concentrated toward higher values(blue values above dashed black lines).

Figure 35.

Example of model fit of heads in the medium grid model. Left: simulated vs measured head values in two observations. Right: residual vs observed. Prior and posterior (first iteration) realizations of the medium model are shown in gray and blue, respectively.



The evolution of the objective function happens in the first and the second iteration, and after that it stalls, or even, it gets worse (Table *9*). Furthermore, we could only find improvement

of the objective forecast in the first iteration because after this first iteration the IES stalls and lose realizations because of model failures (

*Figure **36***).

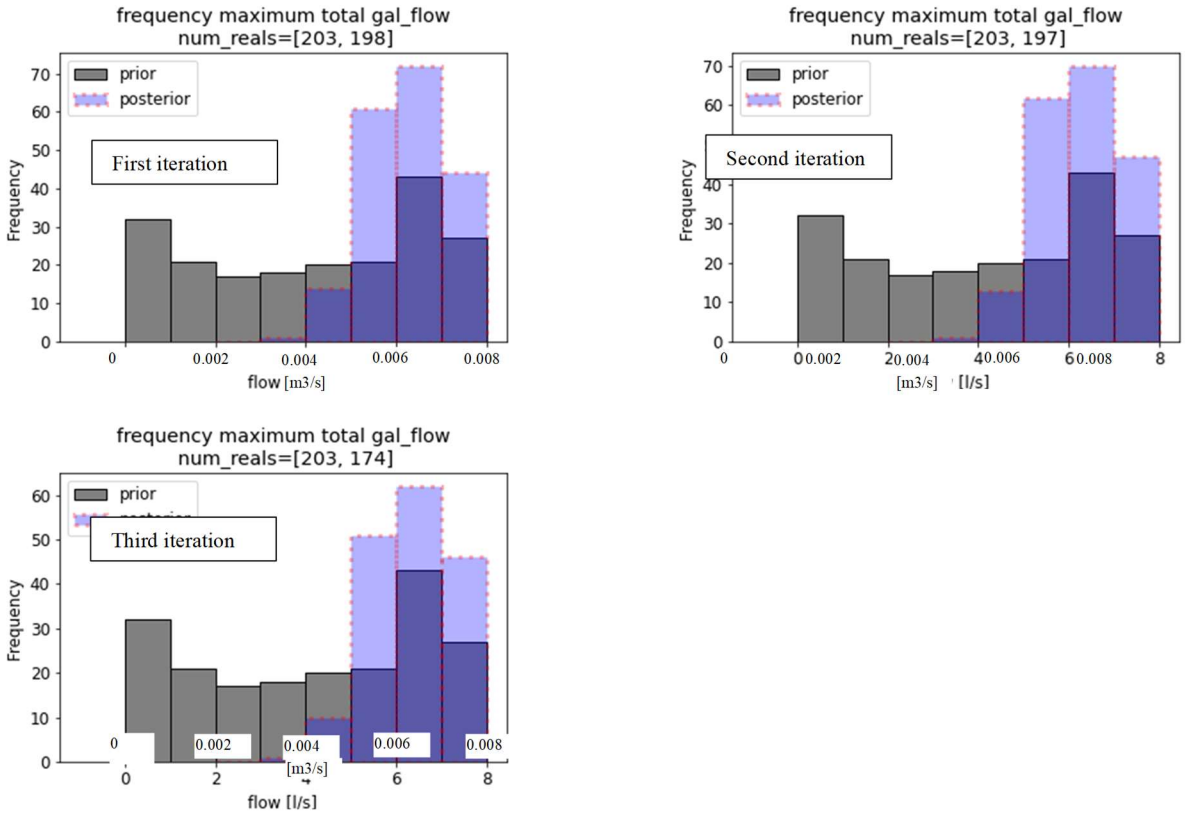
Table 9.

Evolution of the target objective function in the ensemble of medium models. Bolded values show where there is a noticeable improvement in the statistics

iteration	Mean	standard_deviation	min	max
0	2.57E+64	4.75E+64	4.86E+07	1.99E+65
1	**2.74E+62**	1**.15E+63**	**3.27E+07**	**1.42E+64**
2	**1.19E+08**	**7.96E+07**	3.27E+07	**5.43E+08**
3	1.21E+08	8.28E+07	3.27E+07	5.43E+08
4	1.22E+08	8.08E+07	4.29E+07	5.43E+08
5	1.23E+08	8.54E+07	5.19E+07	5.43E+08

Figure 36.

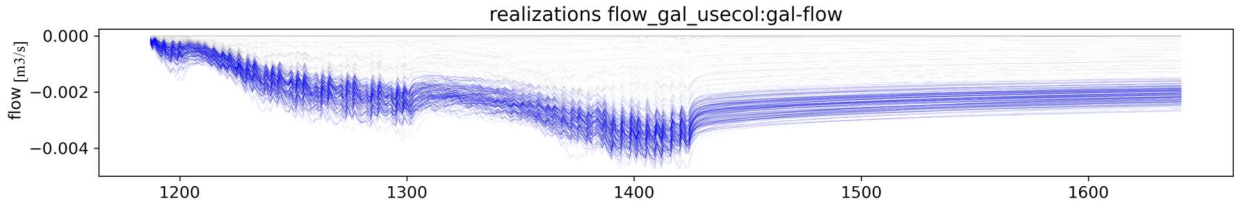
Evolution of maximum flows distribution in the ensemble of medium models(250 realizations)



Results of the fine grid model are, in general, similar to those of the medium model with the same behavior of improvement, but lower values of flows. flow yield by the gallery is shown in Figure 37, the posterior realizations tend to higher values, but without the peaks of the medium model.

Figure 37.

Prior (gray) and posterior (blue) realizations of gallery flow (m^3/s) in the fine grid model, negative values denote groundwater outflow in MODFLOW convention.

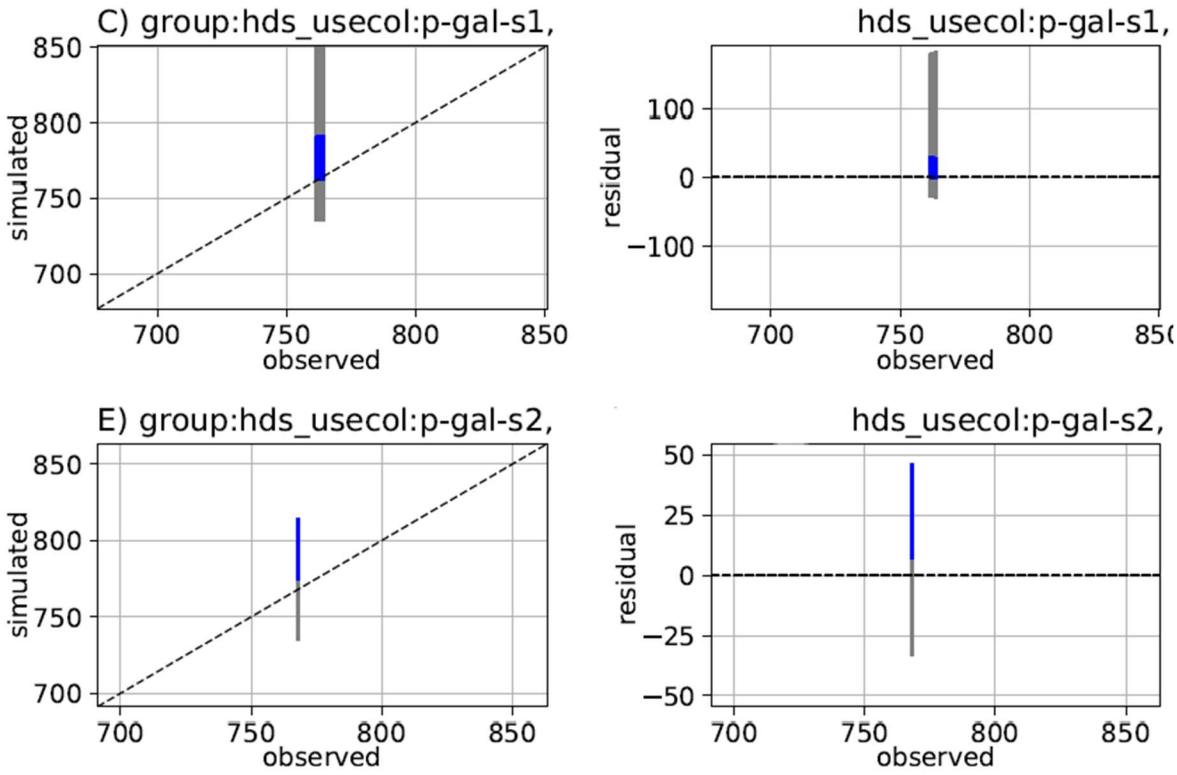


The fit in the fine grid model tends to degenerate in some observations (

Figure ***38***, lower graph), where the ensemble of models don't contain the observed values, suggesting some problems in the upgrade of the parameters.

Figure 38.

Example of model fit of heads in the fine model. Left: simulated vs measured values of two observations. Right: residual vs observed. Prior and posterior (first iteration) realizations of the fine model are shown in gray and blue, respectively



In the same way as the medium grid model, the fine grid model does not find more improvement in the objective function beyond the second iteration (Table 10), and it stalls in the forecast after the first iteration (Figure *39*). The final distribution of flows is asymmetrical with a peak in the right side around 4-5 l/s

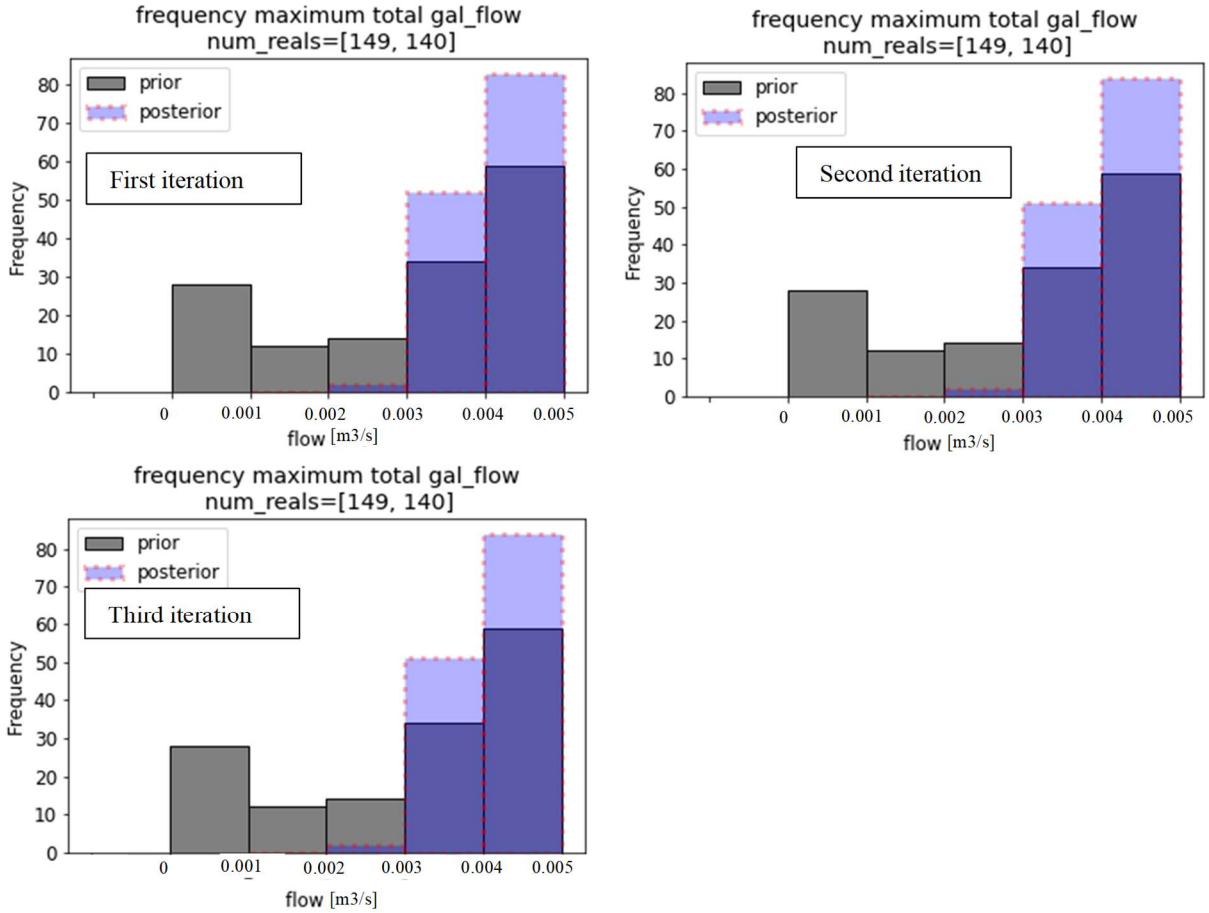
Table 10.

Evolution of target objective function in the ensemble of fine models. Bolded values show where there is a noticeable improvement in the statistics

iteration	mean	standard_deviation	min	max
0	2.02E+64	4.61E+64	4.06E+07	1.99E+65
1	**1.21E+62**	**4.62E+62**	**3.64E+07**	**1.88E+63**
2	**1.10E+08**	**6.36E+07**	3.64E+07	**3.84E+08**
3	1.10E+08	6.36E+07	3.64E+07	3.84E+08
4	1.10E+08	6.36E+07	3.64E+07	3.84E+08

Figure 39.

Evolution of maximum flow distribution in the ensemble of fine grid models (250 realizations)



5.4 Model calibration

The GLM process in model calibration had low success updating parameters and reducing the uncertainty, and the NSMC had a better performance reducing the target objective function than GLM. Moreover, uncertainty results estimated by methods relying on the Jacobian matrix were dubious.

The first parameter optimization by GLM after calculating the Jacobian matrix estimated the λ parameter to try different directions of the upgrade vector and explore widely the parameter space (Welter et al., 2015) (Table *11*). All λ parameters resulted in a *phi* (value of the target objective function) bigger than the initial one. Since a *phi* value is generated by every NSMC realization, the values of *phi* generated by NSMC can be compared with lowest got by GLM. In case *phi* by NSMC is lower than that found by GLM, it can be used as a new starting point for the next iteration. In the fine-NR model NSMC yielded *phi* values lower than 70% of the initial *phi* which is a better performance than that of the GLM algorithm (Figure *40*); this means that λ optimizations of GLM are discarded to use the Monte Carlo realization with the lowest phi in the next iteration of GLM.

Table 11.

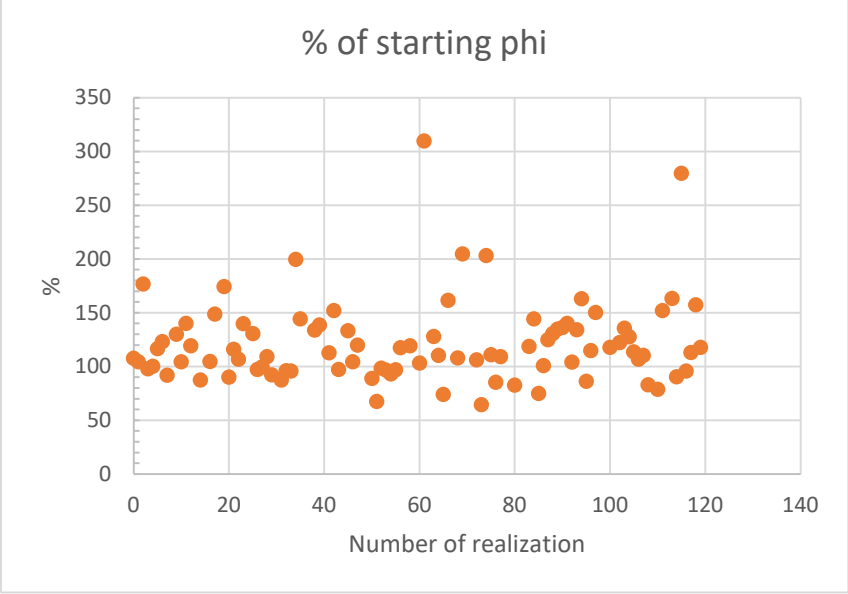
Summary of GLM lambda upgrade runs in the first iteration. Type indicates if the vector upgrade is being scaled, Length is the length of the upgrade vector, and Phi-lambda is the value of the target objective function achieved with that vector

Iteration	Lambda (λ)	Type	Length	Phi-lambda	% of starting phi
1	0.1	normal	9.00E+00	2.89E+07	100.24
1	0.1	scale(0.75)	4.62E+00	2.89E+07	100.17
1	1	normal	9.00E+00	2.89E+07	100.24

Iteration	Lambda (λ)	Type	Length	Phi-lambda	% of starting phi
1	1	scale(0.75)	4.62E+00	2.89E+07	100.18
1	10	normal	9.17E+00	2.95E+07	102.15
1	10	scale(0.75)	4.77E+00	2.93E+07	101.58
1	100	normal	9.24E+00	3.41E+07	118.07
1	100	scale(0.75)	4.84E+00	3.24E+07	112.3
1	1000	normal	1.10E+01	3.66E+07	126.81
1	1000	scale(0.75)	5.99E+00	3.35E+07	116.25

Figure 40.

Monte Carlo phi summary of the first iteration.



The model presented more instability in the second iteration of GLM, having several model failures computing Jacobian sensibilities. The number of failed models were 28 of 461 parameters, approximately 6% of the total sensitivity values in Jacobian. In this second iteration, λ upgrade runs had a very poor performance and one of the trials of λ failed showing that the model is suffering of an important stability problem that could affect the capacity of GLM to update parameters (Table *12*).

Table 12.

Summary of GLM lambda upgrade runs in the second iteration. Type indicates if the vector upgrade is being scaled, Length is the length of the upgrade vector, and Phi-lambda is the value of the target objective function achieved with that vector

Iteration	Lambda(λ)	Type	Length	Phi-lambda	% of starting phi
2	0.1	normal	5.83E+00	1.91E+07	100.04
2	0.1	scale(0.75)	3.00E+00	1.92E+07	100.26
2	1	normal	5.83E+00	1.91E+07	100.06
2	1	scale(0.75)	NA	run failed	run failed
2	10	normal	5.83E+00	1.91E+07	100.04
2	10	scale(0.75)	3.00E+00	1.92E+07	100.26
2	100	normal	5.83E+00	1.91E+07	100.01
2	100	scale(0.75)	3.00E+00	1.91E+07	100.21
2	1000	normal	5.84E+00	1.90E+07	99.58
2	1000	scale(0.75)	3.00E+00	1.91E+07	99.93

The *phi* estimated by NSMC was again lower than that found by GLM (Figure *41*), thus the next step would be again determined by the stochastic method (NSMC). But at this point the optimization was stopped because every GLM iteration is an expensive process, and not significant improvement was achieved by it, instead the NSMC was updating the parameters. Although resulting in a poor fit of the measured values (Figure *42*) the computational burden of more iterations would imply a waste of computational resources and time.

Figure 41.

Monte Carlo phi summary of the second iteration

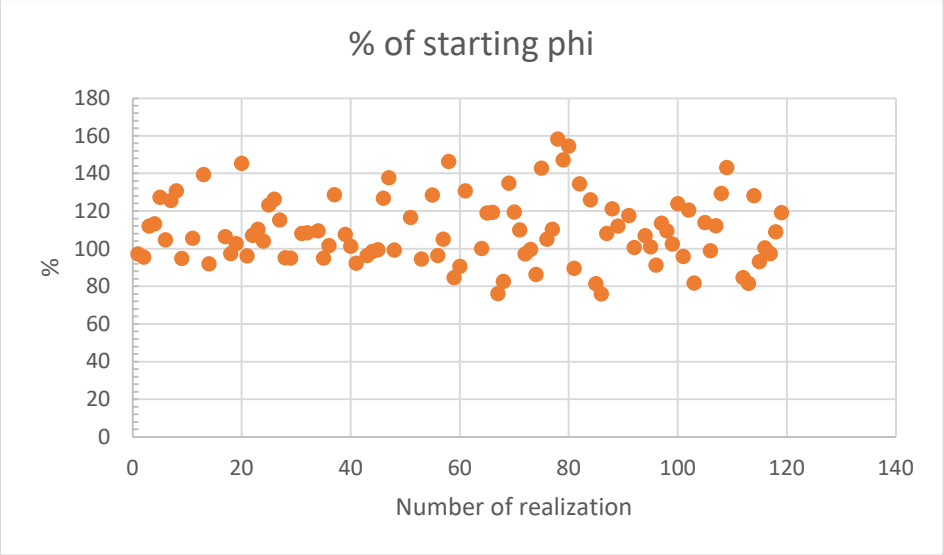
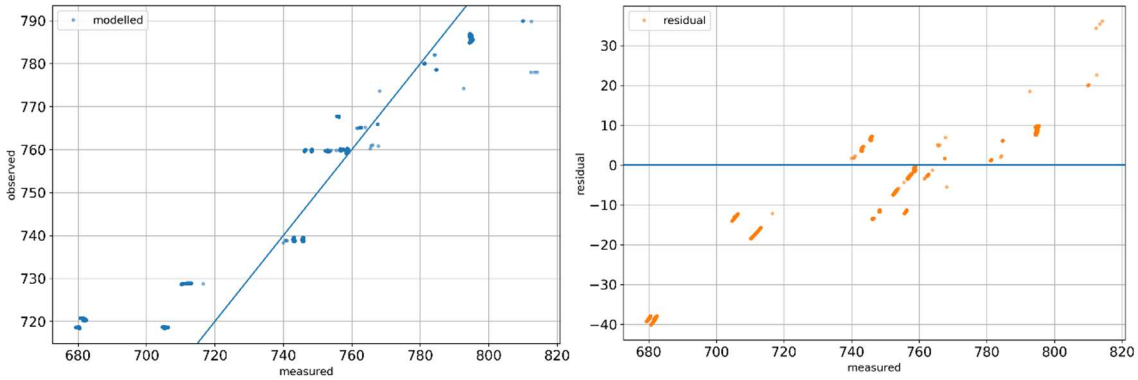


Figure 42.

Model fit after the first iteration of GLM. observed vs measured values at left and residual vs measured a right. The values are hydraulic head in meters.

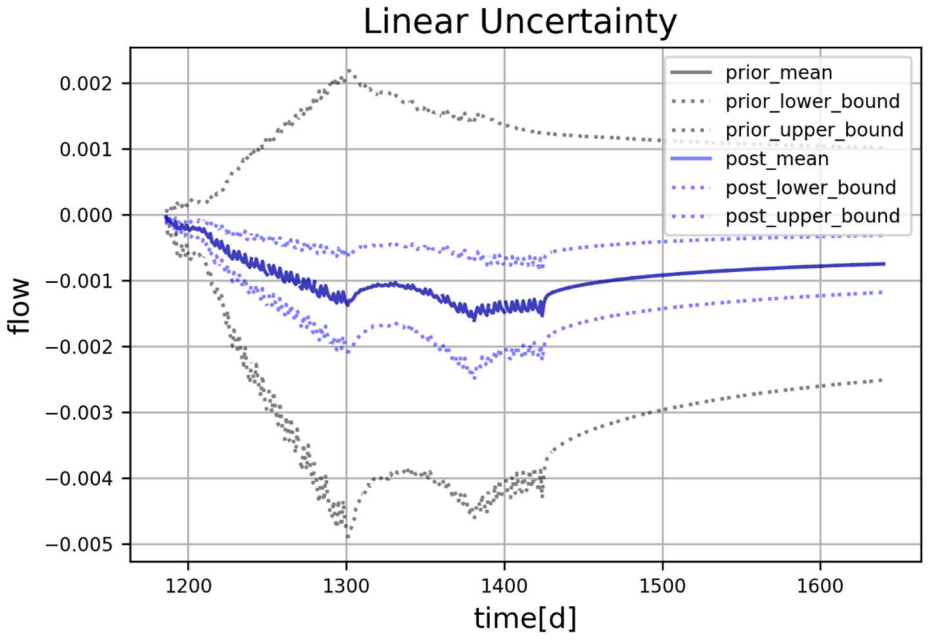


5.4.1 Shur's complement for linear uncertainty analysis

The first calculation of linear uncertainty after computing the Jacobian matrix shows flows within a range of 2 standard deviations (close to 95% of confidence intervals). Prior flows have values up to 5 l/s in day 1300, in a different day where the maximum flow of the mean is located (day 1380 approximately) (Figure 43). The linear uncertainty method yields symmetric results around the mean, showing positive flow values inside the uncertainty limits. This should be ignored because positive flow values are not possible (the gallery injecting water to the groundwater system).

Figure 43.

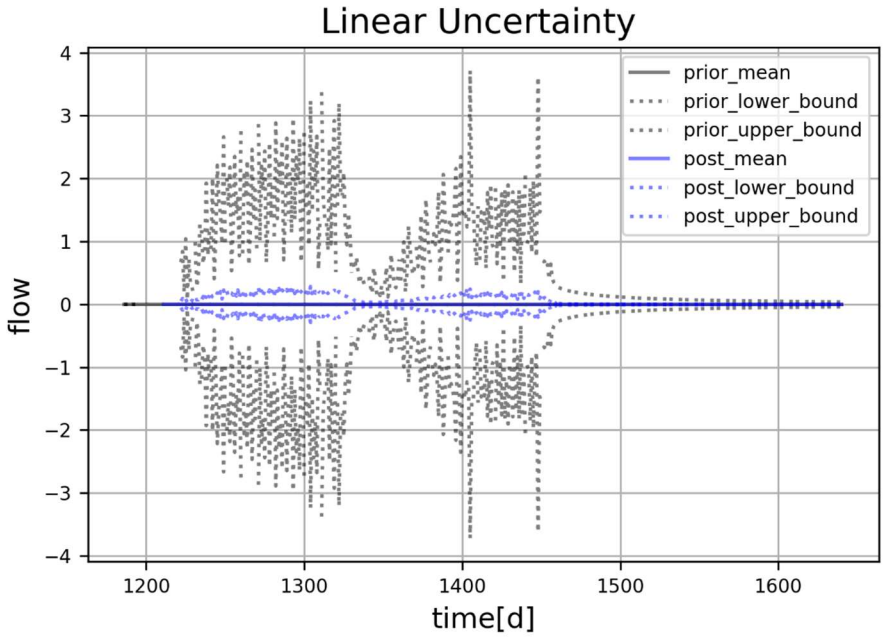
Prior and posterior linear uncertainty of flow (m^3/s) in the initial model. The prior mean and the post mean are superimposed; negative values denote groundwater outflow in the MODFLOW convention.



The linear uncertainty computed after the first iteration shows extremely high values for the prior and an erratic behavior with high changes in uncertainty along time (Figure 44). Although the posterior values are much lower than the prior, they still are big values that should be taken with caution.

Figure 44.

Prior and posterior linear uncertainty of flow (m^3/s) after the first iteration using NSMC. Negative values denote groundwater outflow in the MODFLOW convention.



The results of prior and posterior NSMC (Figure 45) show values up to 2.5 l/s in the prior and up to 4.5 l/s in the posterior, with posterior realizations showing a significant shift towards higher values (Figure 46).

Figure 45.

Prior (gray) and posterior (blue) realizations by NSMC

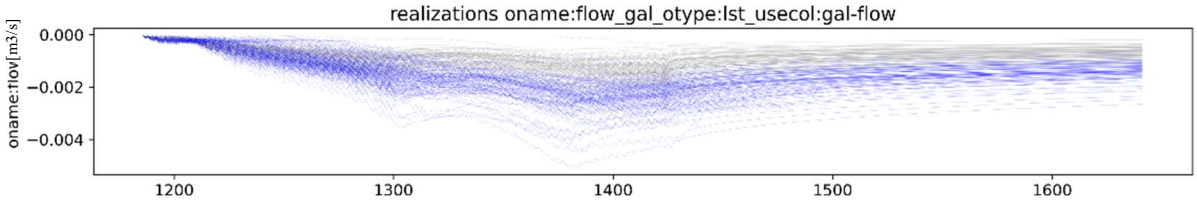
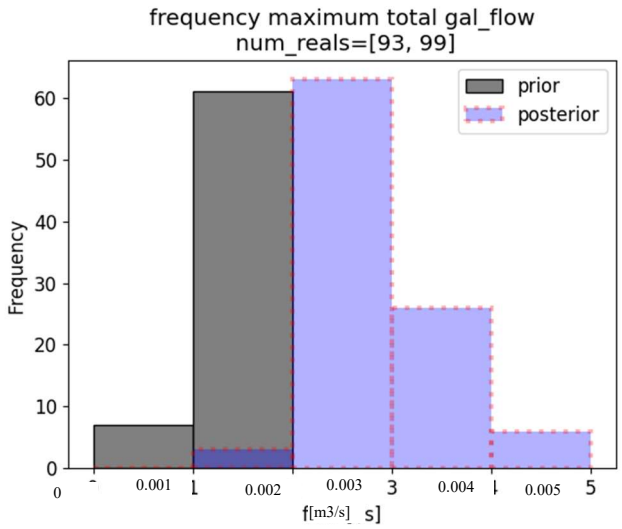


Figure 46.

Frequency distribution of maximum flows generated by NSMC



5.5 Time and computational burden

The run time spent in every model is similar in IES and GLM despite changes made to the fine model to use the Newton Raphson (NR) solver and different parameterization of the history matching. The important difference between the processes in computational burden is related to the number of runs, which for GLM is linked to the number of parameters. A summary of the average times used for the different processes used to estimate the uncertainty is shown in (Table

13). The table does not show total time because different number of threads (The computer used have 6 cores and 12 threads) were used for each process (8 in the fine vs 11 in the Fine-NR). It is important to notice that average run time is different to the run time of the base model (shown in Table *6*) (i.e., base run time of the Fine-NR model is 72 minutes). Also, all these times should be seen in a relative way and will be different in every computer.

Table 13.

Number of model runs and average time in every iteration for the fine models used in GLM and IES. Note: the linear uncertainty analysis has a long run time because the uncertainty of all the observations was assessed; the time required to only assess the uncertainty of gallery flow would be around 6 hours.

Model	Process	Avg time 1 it	#Runs 1it	Avg time 2 it	#Runs 2 it	Avg time 3 it	#Runs 3 it	Avg time 4 it	#Runs 4 it
fine	IES-Jacobian /uncertainty	128 min	149	135 min	139	147 min	136	138 min	136
fine	IES-Lambda	108 min	20	129 min	29	122 min	36	114 min	32
Fine-NR	GLM-Jacobian	138 min	461	169 min	433	-	-	-	-
Fine-NR	GLM-Lambda	138 min	10	158 min	9	-	-	-	-
Fine-NR	GLM-Linear uncertainty	32 h*	1	31 h*	1	-	-	-	-
Fine-NR	GLM-NSMC	132 min	93	151 min	99	-	-	-	-

6. Discussion

The main goal of this work was to perform the best flow forecast uncertainty analysis for the drainage gallery using a few data in a highly parameterized numerical model. To find the best way to assess the uncertainty of the flow yielded by the gallery, we applied three types of analyses: IES (White, 2018), Shur's complement (Fienen et al., 2010; White et al., 2016) and NSMC (M. Tonkin & Doherty, 2009; White et al., 2016). The last two are commonly used and are done after model calibration, a time expensive process. Therefore, we included in our analysis an ensemble method like the Iterative Ensemble Smoother. IES was successfully used here to address uncertainty despite its low success in history matching. Furthermore, to be time-efficient (computationally), we used elements from the Early Uncertainty Quantification proposed by Hemmings et al., 2020. Although not all the results are conclusive, the overall result is that a lighter method like IES could be more appropriate to assess uncertainty with less risk of wasting time and computational effort, at least if the uncertainty of a forecast is the main goal instead of history matching.

6.1 Assessment of model structure

The early quantification of the uncertainty in the coarse grid model showed a structural problem because the model produced very low hydraulic heads (Figure 20). A higher inflow in recharge added to wider parameter bounds was implemented to help the ensemble match field measurements. Also, the rock layer was set with a higher hydraulic conductivity because it is

highly fractured in some places. While some hydraulic head model observations, showed low values, others, especially deep observations, presented higher values than expected. This was, clearly, a prior data conflict as stated by Hemmings et al., 2020 because the lack of refinement and a narrow biased prior distribution of conductivity was not able to reproduce some processes (Knowling et al., 2019; White, Knowling, et al., 2020). This could be confirmed when most of the observations were fitted after refining the mesh (Figure 23) and updating the prior distributions (

Table **7** and Table 8).

Regarding the non-important observations that were not matched (

Figure ***24***), those are either horizontally or vertically away from the gallery or observations of low (i.e. isolated measurements of old studies building issues in piezometers). Improving the model to match the deep observations would require more vertical discretization, but it would generate a more complex model without clearly helping our forecast which is about dewatering close to the topographic surface.

Regarding model refinement, it is clear that mesh size is critical (Knowling et al., 2019). The histograms of prior flows indicate a reduction of modeled flows with refinement, with values up to 80 l/s in the coarse model (Figure *25*) down to maximum flows of 5 l/s in the fine one (Figure 32). Also, the histogram distributions show a non-Gaussian trend, suggesting that this prediction is not suitable for linear methods.

There are two main reasons to explain changes in flow with model discretization. The first one is that theoretically, the smaller the cell, the more accurate the solution, and, in the same way, a nodal spacing similar to the dimension of the well is needed to accurately represent heads (Anderson et al., 2015); here, a cell spacing similar to the dimensions of the gallery is necessary to represent flows accurately. Nevertheless, numerical experiments carried out in finite-difference grids showed that when the cell size is similar to the gallery dimensions, the steady flow is underestimated by around 15%, and this numerical error is minimized either in very fine grids or in cell sizes three times larger than the size of tunnel/gallery (Zaidel et al., 2010).

The second reason that explains changes in maximum flows with model refinement is due to large cells suddenly activating a boundary condition generating great peaks in flow that stabilize quickly to a lower value (Figure 26 and Figure 28) and imprinting this effect in the histogram of maximum flow. This is supported when comparing the fine model with the medium using the minimum value of the previous 4 days (a gallery BC is activated every 4-5 days accordingly to

digging rate times); in both cases (Figure 29), the flow is very similar, and peaks are eliminated. On the other hand, we see that flow in the long term does not converge to the same value with both models, probably because of the effect of accuracy mentioned above.

Finally, all the workflow approaches used to keep the model as simple as possible (John Doherty & Moore, 2020) detected prior data conflicts (Hemmings et al., 2020) and watched the parameterization effect in the model forecast (Knowling et al., 2019; White, 2017; White, Knowling, et al., 2020). These two actions led us to select the fine model and their ensemble of forecasts, which predicted a maximum flow of 5 l/s in a bimodal distribution.

In the detailed design proposed in UIS & AMB, (2019), the lower design flow corresponded to the external drain channel which was designed to support 30 l/s. If we were to follow the recommendations proposed by Hemmings et al., (2020) strictly, we could stop the modeling process at this point because the probability of exceeding the threshold of this forecast is practically null; but, we kept working on the assessment of the uncertainty by different methods for the sake of making a complete analysis.

6.2 Iterative Ensemble Smoother (IES)

IES showed a good performance reducing the uncertainty of the forecast in the first iteration but null improvement in the following iterations, indicating that hydraulic head observations have little data about the forecast. However, in the updating process, we found symptoms of model instability that affected the results.

We had to estimate the prior uncertainty twice in the medium model because even with the medium refinement the model was a little unstable and many realizations failed (

Figure ***33***) (many stable realizations were automatically killed if they presented long-run times). Also, we can observe the posterior distribution of maximum flows that seems to take a gaussian shape. Furthermore, we can see in

Figure ***33*** and Figure 34 that one iteration of data assimilation circumscribed the forecast to a lower extent, meaning that there is some information in the field observations that informs the prediction.

We can see that differences between simulated and observed values got lower in the medium model (Figure *35*) but even though there is an improvement in the fitting until the second iteration (Table *9*), the improvement was not significant in the forecast after the first iteration (

*Figure **36***). The aforementioned happens probably because we only have head observations which only help us to discard very unlikely scenarios generated by the big uncertainties in the prior, but it is incapable of further improvement because there is very little information of the forecast inside the observations (John Doherty & Moore, 2020; Hemmings et al., 2020).

The calculation of the fine model´s prior and posterior uncertainty found that the reduction of the flow forecast uncertainty in the fine model was stagnant after the first iteration (Figure *39*), however showing an improvement in the ensemble of target objective functions until the second iteration (Table 10); this is the same behavior as with the medium model. These results support the point that there is not much information in the field observations of the flow forecast. The reason why IES is incapable of optimizing more the target objective function is probably that the number of realizations is much less than the number of parameters and it usually generates errors in the estimated cross-covariance between parameters and observations clouding the optimization(White, Hunt, et al., 2020).

Running the IES for the fine grid resulted in the failure of more models; this is ~200 in the medium (

*Figure **36***) vs ~150 in the fine grid model (Figure *39*) (both cases started with 250 realizations). The worsening of stability conditions can explain why the posterior simulated observation cannot fit as well as the medium model (

Figure ***38***).

Finally, the distribution of maximum flows with the fine grid model is still bimodal in the prior but evolves toward an asymmetric distribution with a maximum value slightly lower than in the medium model (5 l/s), but still suggesting a non-gaussian behavior (Figure *39*).

6.3 Model calibration

The GLM process showed to be insufficient (or at least inefficient) to reduce the uncertainty of the model despite its high computational burden, providing support to the hypothesis of the system´s highly nonlinear behavior. Because of this, the process of updating parameters was carried out by the NSMC instead of GLM (we stopped the optimization after two iterations of GLM).

The first iteration of the GLM optimization found bad results, yielding in every case a target objective function worse than in the base model (Table *11*); this could indicate that the "multidimensional surface" to optimize has a highly nonlinear behavior. On the other hand, when we use the NSMC, which is a method that moves over the null space of the Jacobian matrix to avoid affecting the target objective function by changing insensitive parameters, we could appreciate two important things. The first one is that NSMC has a much better performance finding a lower phi than the GLM method (64% of starting *phi*), and the second, is that 70% of realizations have a higher phi than the base model (Figure *40*). The NSMC behavior supports the hypothesis of high nonlinearity in the observation and suggests that the model could be in a local minimum of the objective function.

The lower phi in NSMC realizations was used to follow the next iteration of GLM which behaved better than in the first iteration but still performed poorly, finding values between 99.5% and 100.26 % of the previous phi (Table *12*). Additionally, one lambda iteration and 6% of the runs in the Jacobian failed. This suggests instability in the models and their derivatives, which can cast doubt on the uncertainty analysis that relies on the Jacobian.

In the NSMC realizations of the second iteration, we got 65% of models with a higher phi than the base model, with a minimum value of 76% (Figure *41*). This is, again, a better improvement of the objective function than with GLM, and it is, again, a warning of the nonlinearity of the phi surface.

The model was going to keep estimating derivatives starting from the best set of parameters found by NSMC, nevertheless, the GLM process was not helping to update parameters due to the "shape" of the objective function, and it was very expensive in terms of the computational burden. Because of this, we decided to stop the model because it made no sense to keep calculating derivatives that were useless for the model calibration.

If the purpose of history matching is calibrating a model by the stochastic sounding of the model space, then, other algorithms such as particle swarm optimization (Eberhart & Kennedy, 1995) or even IES are more efficient to do this. However, IES is not necessarily better than GLM to update parameters as can be seen in Table *11* and Table *12*, where every phi of GLM is lower than the minimum phi in IES (Table *9* and Table 10) (although adjustable parameterization of both models is different).

When we analyze the spread of observed vs. simulated values, we can easily find that the model is far from being fitted, where lower values of the hydraulic head are overestimated by the

model and the higher values are underestimated (Figure *42*). This is probably because the hydraulic conductivity is too low.

6.4 Comparison of uncertainties

The IES algorithm has the best performance among the methods used here because Shur's complement and the NSMC inherit the problems of the GLM method related to the computational burden, and non-linearity of the solution (especially for Shur's complement). Nevertheless, the NSMC showed acceptable behavior and made important contributions to the history-matching process.

Results after calculating the uncertainty by Shur's complement (Figure 43) show a similar maximum value than in the IES, a bit more than 4 l/s. The main difference between the two methods is in time; the maximum value in the linear estimation (using two standard deviations) would be 5 l/s on day 1300 while for IES it happens after day 1400 (Figure 37). The posterior estimation of uncertainty by Shur's complement, however, predicts a smaller value of about 2.5 L/s in a time similar to the IES realizations (Figure 43). This could suggest that the linear method could be as useful as IES to estimate uncertainty, obviously dismissing the upper values of the graph (positive values) which would imply a gallery injecting flow to the ground.

Nevertheless, these results should be taken carefully, because the behavior of GLM and IES analyzed before indicates a non-gaussian distribution and a very nonlinear objective function that would invalidate the results of a method that relies on the linearity of the Jacobian to estimate uncertainty and the Gaussianity of the uncertainty. Furthermore, when we used Shur's complement in the next iteration, we obtained very erratic results with extremely high values in the "prior" and

"posterior" (Figure 44). Values Greater than 1000 l/s cannot be considered and are the result of a corrupted Jacobian matrix which leads to unreliable results.

On the other hand, NSMC does not rely completely on the linearity of the Jacobian and shows a more consistent behavior in both iterations (Figure 45 and Figure 46). It shows a trend in flows toward higher values, similar to those obtained through IES, with the peak moving from smaller to values bigger than 2 l/s. However, it needs to be noted that NSMC is a method designed to use when the model has achieved the minimum error variance (by minimizing *phi*) (John Doherty & Moore, 2020) and the posterior uncertainty relies on the fact that the major source of uncertainty is those values which cannot be calibrated. Since the model is not fitted (Figure *42*) (but it's in a local minimum), and the actual values are not "close" to the real values, this may be an underestimated or biased uncertainty.

Given that data assimilation by GLM was not successful and uncertainty always must be overestimated to avoid model failure (Knowling et al., 2019), NSMC is not a reliable method for the uncertainty of flow. Furthermore, the linear method which relies strongly on the consistency and linearity of the Jacobian matrix added to the Gaussianity of the uncertainty is not suitable for the model built in this study. Based on this, we conclude that the best option to assess the uncertainty of the present (or coming models) is the IES method (Figure *39*) given that it explores larger portions of the parameter space hence, it is better prepared to overcome local minima and highly nonlinear models (Hunt et al., 2021).

In UIS & AMB, 2019, three models were tested and calibrated (in steady state), the second model having more observations to calibrate than the first one, and the third one more than the second one. Only the second model was used for the predictions despite the third one having better statistics of calibration than the second one. This was mainly (among other reasons) because the

calibrated model with all the observations (the third one) presented a lower hydraulic conductivity; therefore, a forecast of flow with the third model would yield a lower flow, having more risk of underestimating the prediction. then, to avoid the underestimation of the forecast of the previous study, it was overestimated the flow yielded by the gallery, and also shows why the uncertainty always must be estimated. In our study site, maximum flows are likely to be between 3 and 5 l/s which is a safe flow for the design proposed in UIS & AMB, 2019 (much less than 30 l/s). If we compare with the maximum prediction used in UIS & AMB, 2019, they forecast a maximum flow of about 90 l/s.

6.5 Limitations and future work

Our analyses suggest that EUQ with IES is the best strategy to address the uncertainty of our model and coming models, or even of similar models where the main objective is the uncertainty quantification and not the model parameters. Nonetheless, we acknowledge that there are other possible methods, not used here, to assess uncertainty such as the Ensemble Kalman filter which is a relatively light method in terms of computational burden (Li et al., 2018) that has benefits in terms of handling model error (Alzraiee et al., 2022; Markovich et al., 2022), among other advantages.

Although the IES uncertainty was done with the fine model before using the newton Raphson formulation and some underestimation of the flow could exist due to the less rigorous approach of the solution, it is very unlikely that it could significantly affect our results for three reasons. The first one is that IES does not ask for a rigorous approach to succeed. After all, it uses a coarse Jacobian where the errors of some models are diluted among hundreds of models (White,

2018). The second one is because, in the standard formulation (where cells get dry), observations in dry cells yield a huge value, and therefore a big value in the objective function (Table 10), then the IES algorithm pushes the models toward high values of hydraulic heads to avoid dry cells and it did not affect the forecast as it can be seen in the change between the first and the second iteration (Table 10 and Figure *39*). The third reason that makes the absence of the Newton Raphson formulation a despicable matter is that the forecast is far away from the decision threshold of 30 l/s.

The computer threads used in the fine model vs the fine-NR are different, then the run times in both models are not directly comparable, rather, both times are very similar despite the differences in model settings, therefore the importance of focusing on the number of runs instead of run time. Given that the models used in IES and GLM have different discretization (80m vs 40m), different solver settings (standard vs Newton Raphson), and different history matching parameterization (density of pilot points or constant layer parameters), there is not a real comparison about the results of both mathematical methods. Therefore, the utility of this work is not the pure contrasts in times and results but the empirical comparison in the implementation and convenience of such methods using a computer with limited capabilities.

We are aware that there are different ways to build a model, and some of them could lead to better data assimilation and are worthy of being tested in future work. The first one could be a less rigorous way to represent the geological layers and faults giving more responsibility to the data assimilation to represent geological features (John Doherty & Moore, 2020) but as it was mentioned before, our observations do not have much information of the forecast to implement this approach (John Doherty & Moore, 2020). Another option could be not adapting the model layers to the geological layers but instead maintaining (or almost) horizontal model layers where

the geology is only represented by hydraulic properties, but this approach would need more vertical discretization that was worthy to avoid given our computational limitations. Another aspect that could be improved in the parameterization involves greater rigor in the geostatistical variation of parameters because in this work wasn't available such data and may have an impact on the results.

Despite our computational limitations, we acknowledge that more vertical discretization could be desirable to represent better vertical flow and fit deep observations. Moreover, a more complex mesh could avoid the ill effects caused by calibrating a model with not enough flexibility (White, Knowling, et al., 2020); this could have caused the unfitting of some observations after one iteration of IES (

Figure ***38***). Nevertheless, the loss of some realizations could also explain why there is no more match between the posterior and some observations, and repeating the exercise as was done before with the medium model (

Figure ***33***) could be the solution, though it does not seem that the distribution of maximum flows could change much. Another possible approach could be dropping those observations to avoid the possibility of corrupting the model. moreover could be important to check the effect of those observations considered " non-important" to confirm its dispensability. In the same way, it may be important to test the relevance of full heterogeneity of the water storage parameters for the flow forecast because is known that oversimplification can lead to the underestimation of the uncertainty (Knowling et al., 2019), which is a frequent problem in GLM(White, 2018).

Another change in the base model could be how the gallery BC is implemented; for instance, smoothing the activation of the boundary condition to eliminate the peaks of the medium model (Figure 28) could be useful to try with a less refined model and decrease the computational burden, although as it was mentioned before, this would not delete all the side effects of the medium mesh nor significantly change the results.

Approaches that can be implemented in further studies to help with the computational burden are related to software and hardware. Using a more powerful computer, a network of computers or even cloud computing could overcome most of the computing limitations. On the other hand, using unstructured grids (i.e., DISU Package), could help to refine only the zones of interest of the model and avoid unnecessary calculations. Here it was not used because unstructured grids have very limited support in Pyemu and would need the implementation of other not foreseen software different to the PEST++ suite and Pyemu to handle it (i.e., executables of PEST utilities).

Finally, our study was only focused on the flow forecast. It might be important to include the drawdown forecast as well or pay more attention to the final distribution of parameters in

further studies. In fact, the inclusion of drawdown uncertainty would enable the assessment of the effectiveness of the gallery to increase the safety factor in the zone, which is, its final purpose.

7. Conclusions

We developed a workflow to assess the forecast uncertainty of a drainage gallery flow (extensible to other types of infrastructure) where we used a non-linear, highly parameterized model, having relatively few monitoring data for the history matching relevant to the forecast. We found in the process that the forecast first philosophy (White, 2017) combined with Early uncertainty quantification (EUQ) is highly valuable for model design and almost essential to get competent uncertainty forecasts.

Although the EUQ methodology is essentially agnostic concerning the methodologies adopted for the uncertainty quantification (Hemmings et al., 2020), we found IES desirable over other alternatives that rely on a rigorous Jacobian as NSMC which can sub-estimate the uncertainty in a non-completely calibrated model stuck in local minima or linear methods not always reliable in complex models (usually non-linear because they model the water table).

Even though two iterations of GLM were not successful in our case, probably after more iterations the algorithm would improve, nevertheless, it is still a waste of resources. Instead, IES after one iteration using an overestimation of uncertainty seems to limit the breadth of the forecast even where there is very little information of the forecast in the observations.

Our findings suggest when parameter calibration is an important objective of the modeling (i.e., to know more of the subsurface structure); other approaches could be useful to address non-linearity, like undertaking IES at the beginning of the process (or other non-linear optimization algorithms as PSO (Eberhart & Kennedy, 1995) and finish with the implementation with a rigorous gradient method (Serrano et al., 2017).

Finally, in future work we recommend doing the data assimilation of the new information using IES, furthermore, IES could be used to assess the uncertainty of the gallery's effectiveness (in terms of factor of safety), This is due to its flexibility to couple the geotechnical calculations of safety factors to the groundwater model and instantly get the factor of safety uncertainty.

Bibliographic References

Alzraiee, A. H., White, J. T., Knowling, M. J., Hunt, R. J., & Fienen, M. N. (2022). A scalable model-independent iterative data assimilation tool for sequential and batch estimation of high dimensional model parameters and states. *Environmental Modelling and Software*, *150*(December 2021), 105284. https://doi.org/10.1016/j.envsoft.2021.105284

Anderson, M. P., Woessner, W. W., & Hunt, R. J. (2015). Applied Groundwater Modeling. In *Applied Groundwater Modeling* (2nd ed.). https://doi.org/10.1016/c2009-0-21562-5

Antolínez Quijano, W. L. (2014). *Modelo numérico del flujo de agua subterranea en las formaciones acuíferas de bucaramanga*. Universidad Industrial de Santander.

Aster, R. C., Borchers, B., & Thurber, C. H. (2019a). Rank Deficiency and Ill-Conditioning. In *Parameter Estimation and Inverse Problems*. https://doi.org/10.1016/b978-0-12-804651-7.00008-0

Aster, R. C., Borchers, B., & Thurber, C. H. (2019b). Review of Probability and Statistics. *Parameter Estimation and Inverse Problems*, 341–362. https://doi.org/10.1016/b978-0-12-804651-7.00019-5

Aster, R. C., Borchers, B., & Thurber, C. H. (2019c). Tikhonov Regularization. In *Parameter Estimation and Inverse Problems*. https://doi.org/10.1016/b978-0-12-804651-7.00009-2

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., & Fienen, M. N. (2016). Scripting MODFLOW Model Development Using Python and FloPy. *Groundwater*, *54*(5), 733–739. https://doi.org/10.1111/gwat.12413

Benedek, K., & Dankó, G. (2009). Stochastic hydrogeological modelling of fractured rocks: A generic case study in the Mórágy Granite Formation (South Hungary). *Geologica Carpathica*, *60*(4), 271–281. https://doi.org/10.2478/v10096-009-0019-y

Caballero Caceres, S. A., & Gomez Navas, E. (2011). *Pruebas De Bombeo En Pozo Profundo Y Pozos De Observacion En La Ciudad De Bucaramanga. Sergio*.

Doherty, J. E., Fienen, M. N., & Hunt, R. J. (2010). Approaches to Highly Parameterized Inversion : Pilot-Point Theory , Guidelines , and Research Directions. In *U.S. Geological Survey*. https://doi.org/10.3133/sir20105168

Doherty, J.e., & Hunt, R. j. (2010). Approaches to highly parameterized inversion: a guide to using PEST for groundwater-model calibration. *U. S. Geological Survey Scientific Investigations Report 2010-5169*, 70. https://doi.org/2010-5211

Doherty, John, & Moore, C. (2020). Decision Support Modeling: Data Assimilation, Uncertainty Quantification, and Strategic Abstraction. *Groundwater*, *58*(3), 327–337. https://doi.org/10.1111/gwat.12969

Eberhart, R., & Kennedy, J. (1995). A new optimizer using particle swarm theory. *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, 39–43.

Enemark, T., Peeters, L. J. M., Mallants, D., & Batelaan, O. (2019). Hydrogeological conceptual model building and testing: A review. *Journal of Hydrology*, *569*(December 2018), 310–329. https://doi.org/10.1016/j.jhydrol.2018.12.007

Fienen, M. N., Doherty, J. E., Hunt, R. J., & Reeves, H. W. (2010). Using Prediction Uncertainty Analysis to Design Hydrologic Monitoring Networks: Example Applications from the

Great Lakes Water Availability Pilot Project. *U.S. Geological Survey*, *Scientific Investigations Report 2010 – 5159*, 44.

Franco, M., & Ramírez, A. (2018). *Reconstrucción paleoambiental de la fm. bucaramanga* [Universidad Industrial De Santander]. http://tangara.uis.edu.co/biblioweb/tesis/2018/171985.pdf

Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., Karlstrom, L., Lee, H., Mills, H. J., Oh, J., & Pierce, S. A. (2016). Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science*, *3*(10), 388–415. https://doi.org/10.1002/2015EA000136

Gokdemir, C., Li, Y., Rubin, Y., & Li, X. (2022). Stochastic modeling of groundwater drawdown response induced by tunnel drainage. *Engineering Geology*, *297*(September 2021), 106529. https://doi.org/10.1016/j.enggeo.2022.106529

Gómez-Arroyo, M. A., & Sánchez-Ortíz, O. F. (2020). Geometric modeling of hydrofacies applied to the northern sector of Bucaramanga, Santander, Colombia. *Tecnologia y Ciencias Del Agua*, *11*(4). https://doi.org/10.24850/j-tyca-2020-04-04

Gómez Arroyo, M. A., & Gómez Palencia, M. F. (2021). *Aproximacion al modelo de hidrofacies del sistema acuifero del sector norte de Bucaramanga, Colombia* [Universidad Industrial de Santander]. http://tangara.uis.edu.co/biblioweb/tesis/2021/180839.pdf

Gomez, S., Taupin, J. D., & Rueda, A. (2015). Estudio hidrodinámico, geoquímico e isotópico de las formaciones acuíferas de la región de Bucaramanga (Colombia) Hydrodynamic, geochemical and isotopic study of formations aquifers in the region of Bucaramanga (Colombia). *Revista Peruana Geo-Atmosferica RPGA*, *61*(4), 44–61.

Hemmings, B., Knowling, M. J., & Moore, C. R. (2020). Early Uncertainty Quantification for an Improved Decision Support Modeling Workflow: A Streamflow Reliability and Water Quality Example. *Frontiers in Earth Science*, *8*(November), 1–22. https://doi.org/10.3389/feart.2020.565613

Hill, M. C., & Tiedeman, C. R. (2007). Effective Groundwater Model Calibration. In *Effective Groundwater Model Calibration* (1st ed.). John Wiley & Sons, 2006. https://doi.org/10.1002/0470041080

Hunt, R. J., White, J. T., Duncan, L. L., Haugh, C. J., & Doherty, J. (2021). Evaluating Lower Computational Burden Approaches for Calibration of Large Environmental Models. *Groundwater*, *59*(6), 788–798. https://doi.org/10.1111/gwat.13106

Knowling, M. J., White, J. T., & Moore, C. R. (2019). Role of model parameterization in risk-based decision support: An empirical exploration. *Advances in Water Resources*, *128*(February), 59–73. https://doi.org/10.1016/j.advwatres.2019.04.010

Langevin, C. D., Hughes, J. D., Banta, E. R., Niswonger, R. G., Panday, S., & Provost, A. M. (2017). Documentation for the MODFLOW 6 Groundwater Flow Model: U.S. Geological Survey Techniques and Methods. *USGS: Techniques and Methods 6-A55*, 197. https://doi.org/https://doi.org/10.3133/tm6A55

Li, L., Puzel, R., & Davis, A. (2018). Data assimilation in groundwater modelling: ensemble Kalman filter versus ensemble smoothers. *Hydrological Processes*, *32*(13), 2020–2029. https://doi.org/10.1002/hyp.13127

Lima, M., & Medina, J. (2015). *Evolución climática y estratigrafía del miembro Órganos de la formación Bucaramanga* [UNIVERSIDAD INDUSTRIAL DE SANTANDER]. http://tangara.uis.edu.co/biblioweb/tesis/2015/160331.pdf

Linde, N., Ginsbourger, D., Irving, J., Nobile, F., & Doucet, A. (2017). On uncertainty quantification in hydrogeology and hydrogeophysics. *Advances in Water Resources*, *110*(October), 166–181. https://doi.org/10.1016/j.advwatres.2017.10.014

Markovich, K. H., White, J. T., & Knowling, M. J. (2022). Sequential and batch data assimilation approaches to cope with groundwater model error: An empirical evaluation. *Environmental Modelling & Software*, *156*(August), 105498. https://doi.org/10.1016/j.envsoft.2022.105498

Medina Baez, M. P. (2021). *Caracterización hidrogeoquímica de la contaminación del recurso hídrico subterráneo en la zona norte de Bucaramanga*. Universidad Industrial de Santander.

Moore, C., & Doherty, J. (2005). Role of the calibration process in reducing model predictive error. *Water Resources Research*, *41*(5), 1–14. https://doi.org/10.1029/2004WR003501

Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., & Troldborg, L. (2012). Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. *Advances in Water Resources*, *36*, 36–50. https://doi.org/10.1016/j.advwatres.2011.04.006

Serrano, J. O., Abreo, S. A., Ramirez, A. B., & Pestana, R. C. (2017). A cycle-skipping analysis in transformed domains for full waveform inversion using particle swarm optimization (PSO). *15th International Congress of the Brazilian Geophysical Society \& EXPOGEF, Rio de Janeiro, Brazil, 31 July-3 August 2017*, 1787–1792.

Servicio de Evaluación Ambiental, S. (2012). *Guía para el Uso de Modelos de Aguas Subterráneas en el SEIA*.

https://www.sea.gob.cl/sites/default/files/migration_files/guias/Guia_uso_modelo_aguas_subterraneas_seia.pdf

Suarez, J. (2008). DESLIZAMIENTOS: TECNICAS DE REMEDIACION. Obras de drenaje y subdrenaje. In *Deslizamientos Tecnicas de Remediación* (Vol. 2, pp. 47–106).

Suescún, L. C. (2016). Modelación analítica y numérica para predicción y calibración de caudales de infiltración en obras subterráneas: Colombia, Francia/Italia y España. *Universidad Nacional de Colombia*, 200. http://www.bdigital.unal.edu.co/55471/1/1014200617.2016.pdf

Sundell, J., Norberg, T., Haaf, E., & Rosén, L. (2019). Economic valuation of hydrogeological information when managing groundwater drawdown. *Hydrogeology Journal*, *27*(4), 1111–1130. https://doi.org/10.1007/s10040-018-1906-z

Tonkin, M., & Doherty, J. (2009). Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. *Water Resources Research*, *45*(12). https://doi.org/10.1029/2007WR006678

Tonkin, M. J., & Doherty, J. (2005). A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research*, *41*(10), 1–16. https://doi.org/10.1029/2005WR003995

UIS, G. de I. G., & AMB. (2018). *ESTUDIO DE AMENAZA, VULNERABILIDAD Y RIESGO POR MOVIMIENTOS EN MASA DEL SECTOR NORTE DE BUCARAMANGA (BARRIOS LIZCANO, ESPERANZA III, MIRADOR, JOSÉ MARÍA CÓRDOBA, VILLA HELENA, VILLA ROSA, VILLA MARÍA); INFORME DE DISEÑOS A NIVEL DE PRE FACTIBILIDAD DE LAS*.

UIS, G. de I. G., & AMB. (2019). *ESTUDIO DE FACTIBILIDAD Y LOS DISEÑOS DE INGENIERÍA DE DETALLE, PARA LA CONSTRUCCIÓN DE UNA GALERÍA DE DRENAJE EN EL BARRIO LA ESPERANZA II DEL MUNICIPIO DE BUCARAMANGA, COMO MEDIDA PARA LA REDUCCIÓN DEL RIESGO POR MOVIMIENTO EN MASA PROFUNDOS* (Vol. 1, Issue 1).

Velandia, F., & Bermúdez, M. A. (2018). The transpressive southern termination of the Bucaramanga fault (Colombia): Insights from geological mapping, stress tensors, and fractal analysis. *Journal of Structural Geology*, *115*(July), 190–207. https://doi.org/10.1016/j.jsg.2018.07.020

Ward, D., Goldsmith, R., Andres, J., Cruz, J., Hernán, R., & Eduardo, G. (2013). *Geología de la plancha 109 y 120*. 1977. http://srvags.sgc.gov.co/Flexviewer/Estado_Cartografia_Geologica/

Welter, D. E., White, J. T., Hunt, R. J., & Doherty, J. E. (2015). Approaches in Highly Parameterized Inversion : PEST ++ Version 3, A Parameter ESTimation and Uncertainty Analysis Software Suite Optimized for Large Environmental Models. *U.S. Geological Survey Techniques and Methods 7-C12*, *Techniques and Methods 7-C12*, 54. http://dx.doi.org/10.3133/tm7C12.%0Ahttp://dx.doi.org/10.3133/tm7C12

White, J. T. (2017). Forecast First: An Argument for Groundwater Modeling in Reverse. *Groundwater*, *55*(5), 660–664. https://doi.org/10.1111/gwat.12558

White, J. T. (2018). A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environmental Modelling and Software*, *109*(June), 191–201. https://doi.org/10.1016/j.envsoft.2018.06.009

White, J. T., Doherty, J. E., & Hughes, J. D. (2014). Quantifying the predictive consequences of model error with linear subspace analysis. *Water Resources Research*, *50*(2), 1152–1173. https://doi.org/10.1002/2013WR014767

White, J. T., Fienen, M. N., & Doherty, J. E. (2016). A python framework for environmental model uncertainty analysis. *Environmental Modelling and Software*, *85*, 217–228. https://doi.org/10.1016/j.envsoft.2016.08.017

White, J. T., Foster, L. K., Fienen, M. N., Knowling, M. J., Hemmings, B., & Winterle, J. R. (2020). Toward Reproducible Environmental Modeling for Decision Support: A Worked Example. *Frontiers in Earth Science*, *8*(February). https://doi.org/10.3389/feart.2020.00050

White, J. T., Hemmings, B., Fienen, M. N., & Knowling, M. J. (2021). Towards improved environmental modeling outcomes: Enabling low-cost access to high-dimensional, geostatistical-based decision-support analyses. *Environmental Modelling and Software*, *139*(February), 105022. https://doi.org/10.1016/j.envsoft.2021.105022

White, J. T., Hunt, R. J., Fienen, M. N., Doherty, J. E., & Survey, U. S. G. (2020). Approaches to highly parameterized inversion: PEST++ Version 5, a software suite for parameter estimation, uncertainty analysis, management optimization and sensitivity analysis. *Techniques and Methods*, 64. http://pubs.er.usgs.gov/publication/tm7C26

White, J. T., Knowling, M. J., & Moore, C. R. (2020). Consequences of Groundwater-Model Vertical Discretization in Risk-Based Decision-Making. *Groundwater*, *58*(5), 695–709. https://doi.org/10.1111/gwat.12957

Zaidel, J., Markham, B., & Bleiker, D. (2010). Simulating seepage into mine shafts and tunnels with MODFLOW. *Ground Water*, *48*(3), 390–400. https://doi.org/10.1111/j.1745-6584.2009.00659.x

Zhou, H., Gómez-Hernández, J. J., & Li, L. (2014). Inverse methods in hydrogeology: Evolution and recent trends. *Advances in Water Resources*, *63*, 22–37. https://doi.org/10.1016/j.advwatres.2013.10.014

Appendices

Appendix A. Conceptual Model complement

Limits of the area and geology

The Area of study is in the Bucaramanga Formation (Qtf) which is a geological unit at the southwest of the mountainous region (Figure *47*) composed of weathered and fractured igneous rocks(Gomez et al., 2015).The Bucaramanga formation has 2 main hydrogeological units, Qbg and Qbo members(Antolínez Quijano, 2014; Gomez et al., 2015), but only the last one is present in the area of the gallery. A first choice for the study area would be the complete geological unit as done by Antolínez Quijano, (2014), but is too big for the purpose of the gallery analysis, the a smaller area chosen by UIS & AMB, (2019) is used . The limits of the study zone are defined at the north by the two most important rivers of the CSRL (Oro and Surata). UIS and Picha creeks at the east and west are deemed to be hydraulic features that separate the hydrogeological system of the area from other parts of the same hydrogeological unit (Bucaramanga formation).To the south our study area is fully connected to rest of the aquifer of the Organos member of the Bucaramanga formation as is shown by the isopotentials of the system (Figure 46). Then to define a limit without a physical or hydraulic boundary close to the area of interest an arbitrary limit is selected where the geomorphology change. In this limit, a GHB condition can be used later to simulate the lateral inflow, either extracting local BC from a bigger model or simulating a distant boundary condition.

Figure 47.

Geological map of Bucaramanga 1:100.000 (Ward et al., 2013)

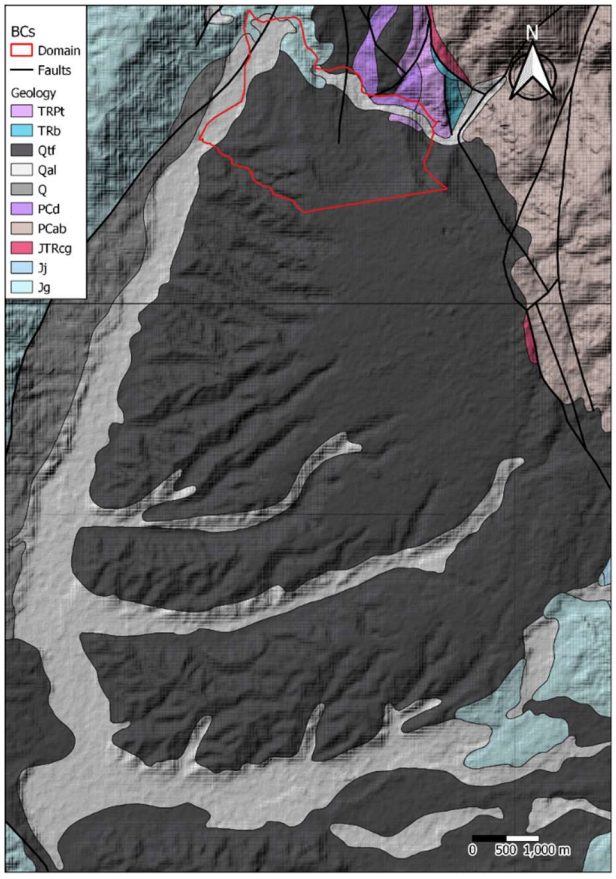
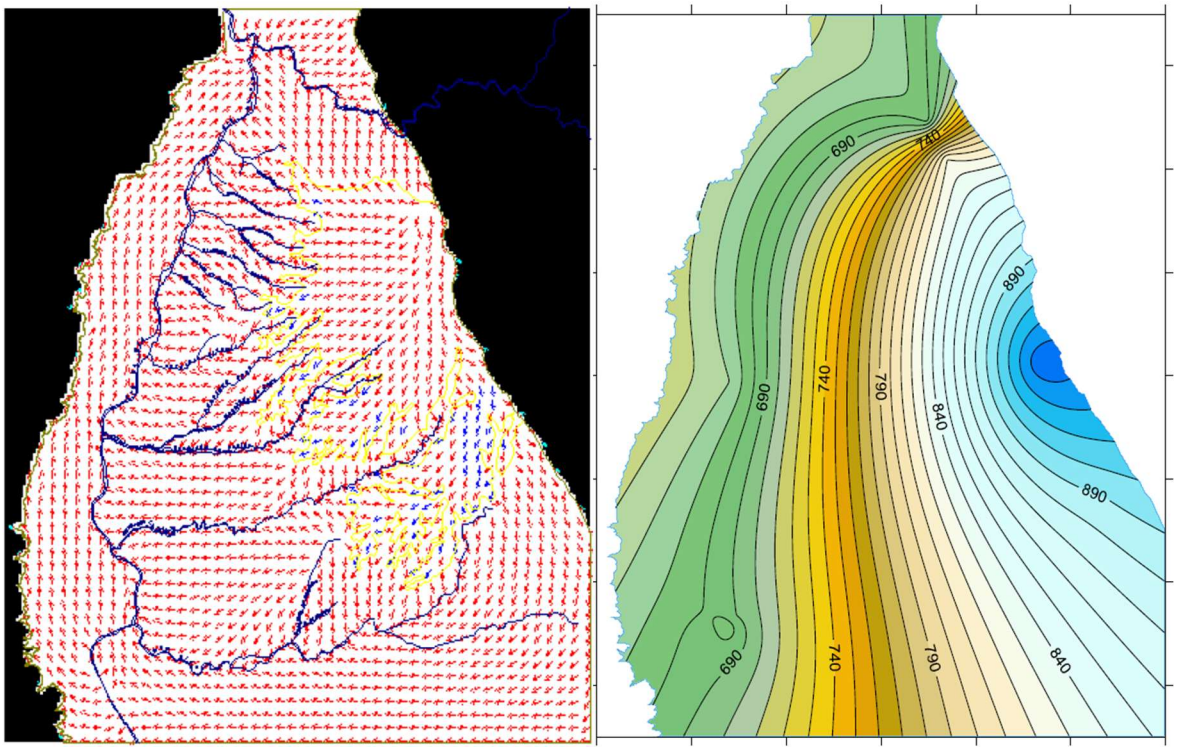


Figure 48.

hydraulic behavior of the Qbo member, at left directions of Flow in a numerical model, at right isopotentials measured from wells.(Antolínez Quijano, 2014)



Hydrostratigraphy

There are 5 hydrogeological units: Rock, Qbo1, Qbo2, Qbg, Qd. the first one is composed of all the sedimentary consolidated geological units. Qbo1, Qbo2 and Qbg are members of the Bucaramanga formation. Qd is a colluvial deposit formed of landslides of the Bucaramanga formation. The geometric distribution of the layers in the study area is showed in **Error! Reference source not found.** to **Error! Reference source not found.**. The profiles shown were used to build the 3D geological model (Figure 57) that will be used for the groundwater numerical model.

Figure 49.

Plan view of the geological profiles, Adapted by the Author from Gómez Arroyo & Gómez Palencia, 2021

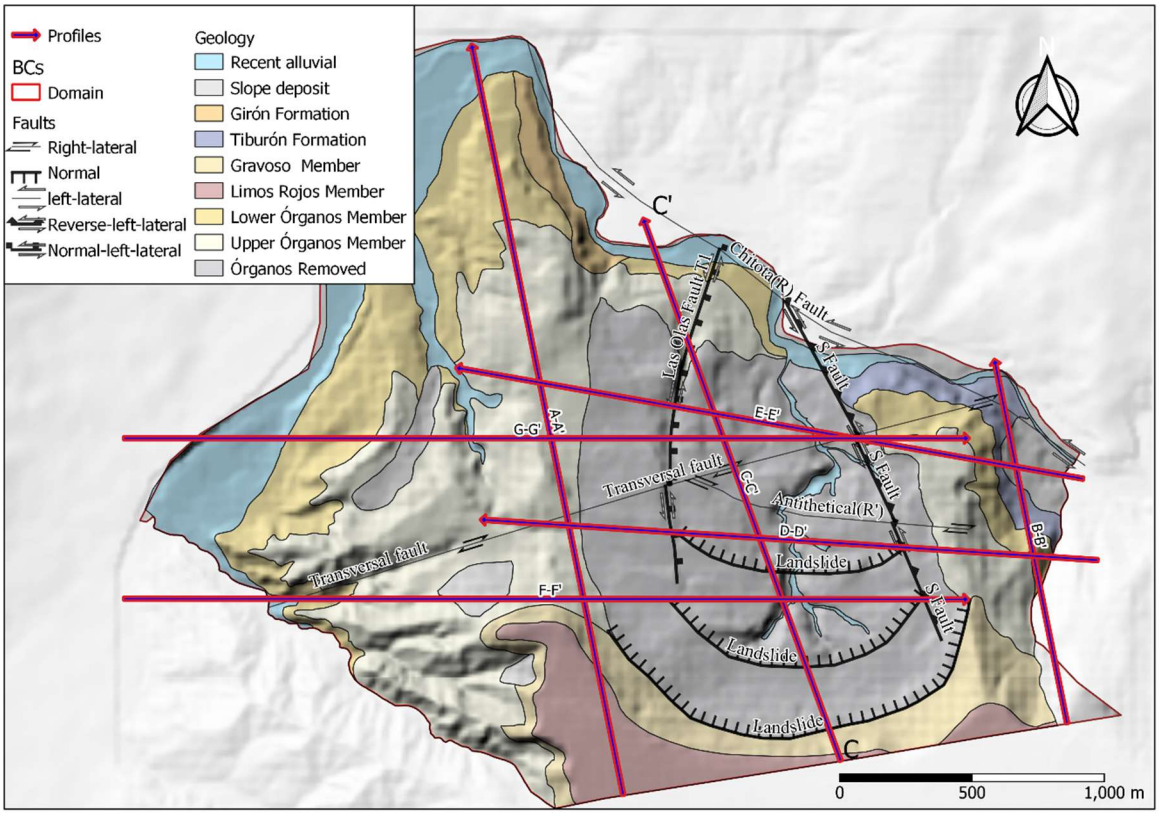


Figure 50.

Geologic profile for the A-A' section (UIS & AMB, 2019)

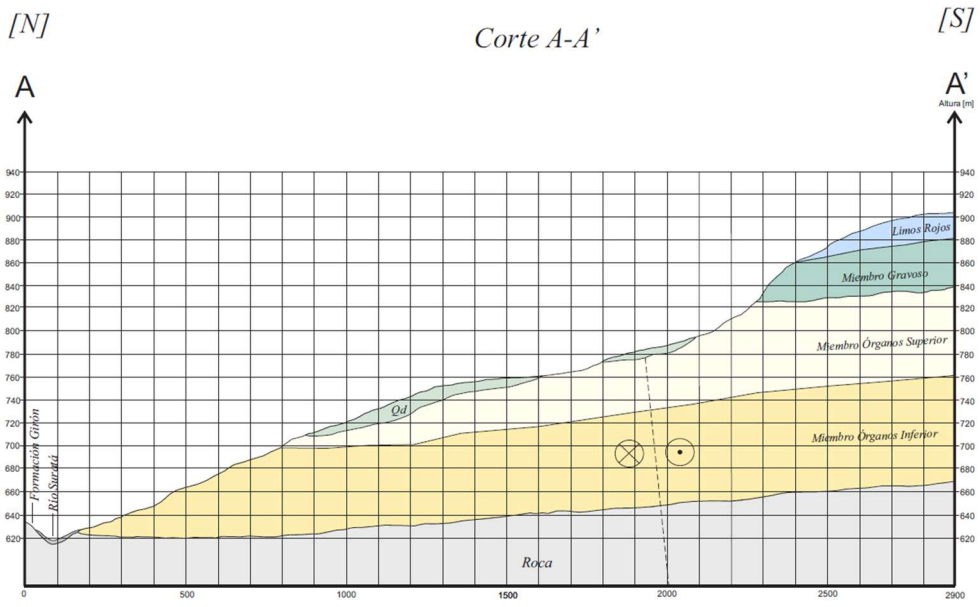


Figure 51.

Geologic profile for the B-B' section (UIS & AMB, 2019)

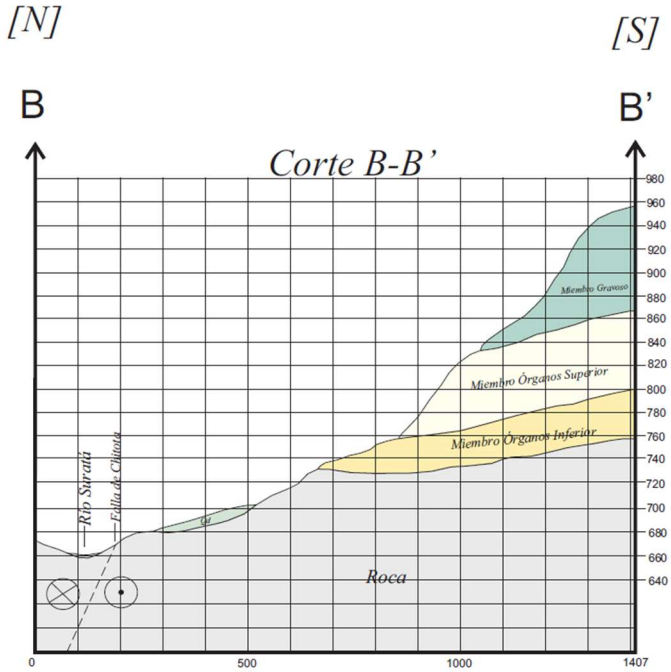


Figure 52.

Geologic profile for the C-C' section (UIS & AMB, 2019)

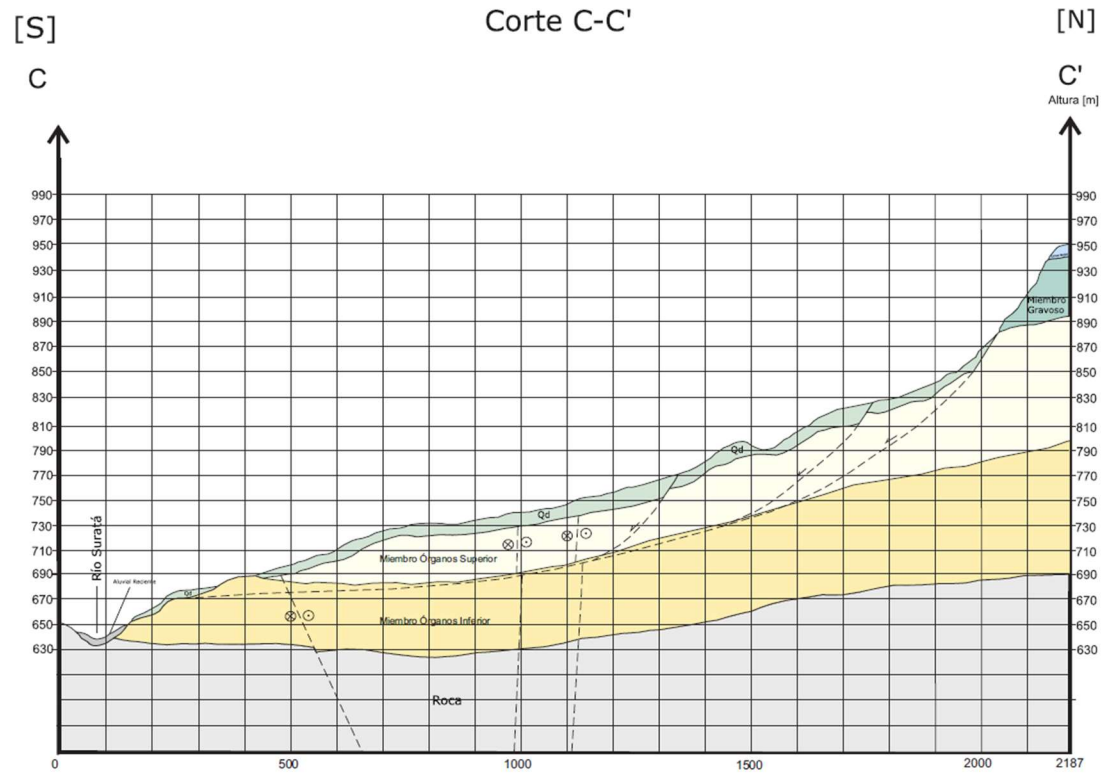


Figure 53.

Geologic profile for the D-D' section (UIS & AMB, 2019)

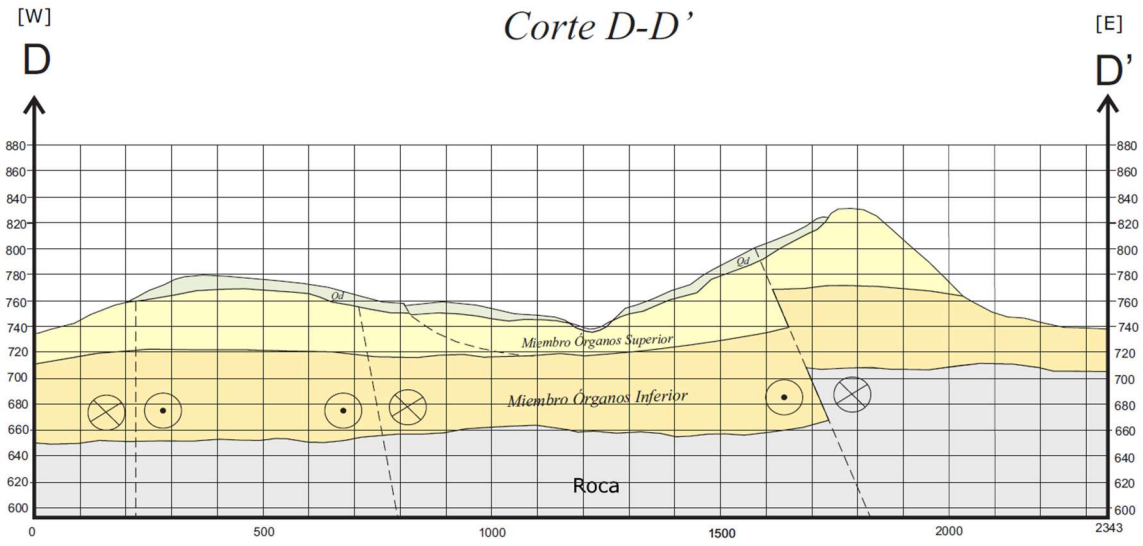


Figure 54.

Geologic profile for the E-E' section (UIS & AMB, 2019)

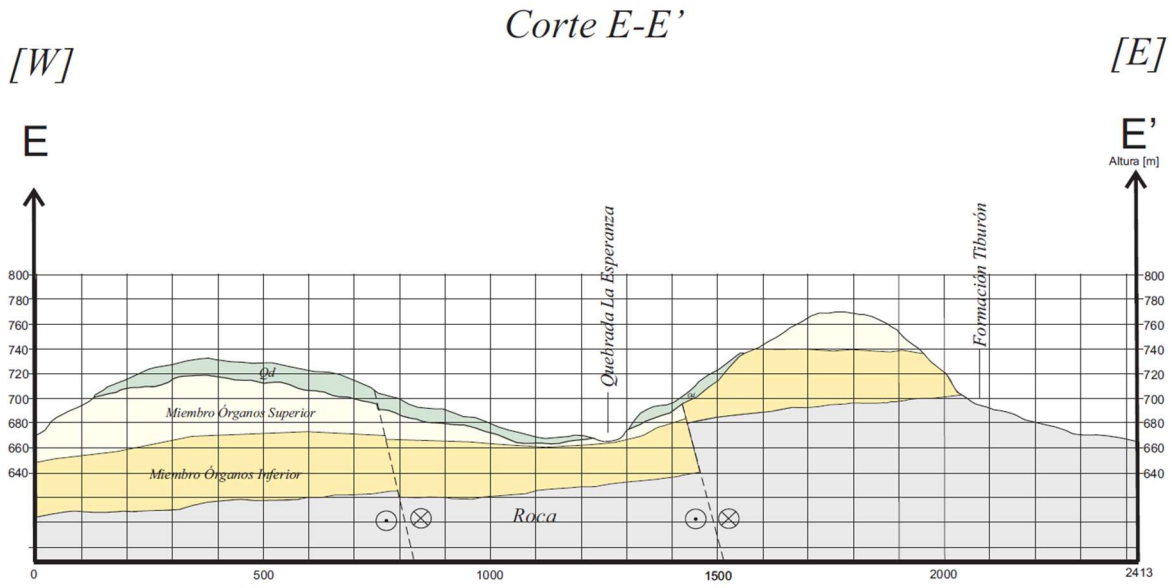


Figure 55.

Geologic profile for the F-F' section (UIS & AMB, 2019)

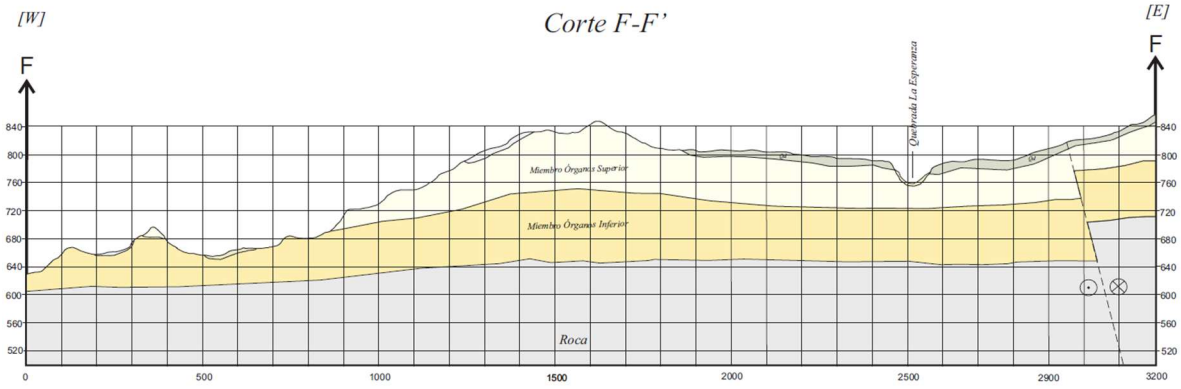


Figure 56.

Geologic profile for the G-G' section (UIS & AMB, 2019)

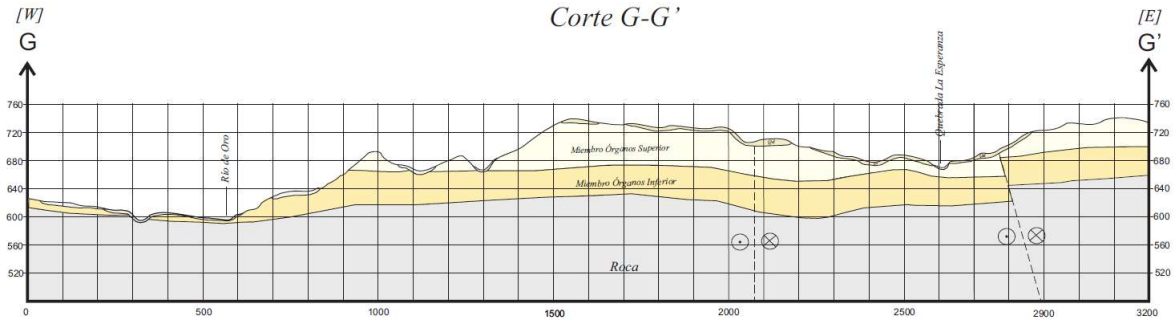
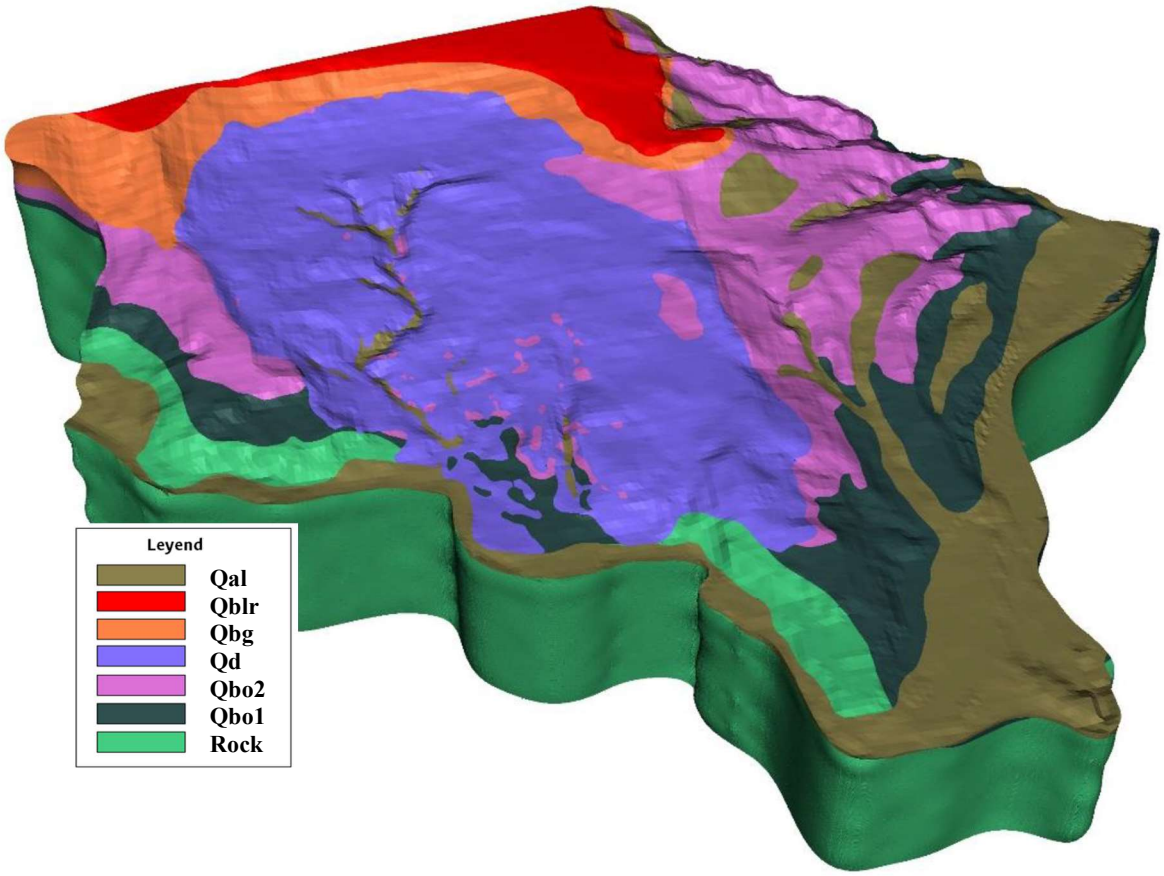


Figure 57.

3D Geological Model (UIS & AMB, 2019)



Hydrogeological Properties

The hydrogeological units of the Organos member can be divided in hydrofacies to understand better their hydraulic properties (Table *14*). In general Qboq, Qbo2 and Qd has the same hydrofacies but in different proportions(Gómez-Arroyo & Sánchez-Ortíz, 2020; Gómez Arroyo & Gómez Palencia, 2021). The hydrofacies proportion shows that the Qbo1 member has more open framework (GSM) hydrofacies, but more low potential hydrofacies(LAGr, La) than Qbo2. Also, can be seen that Qd has similar characteristics to Qbo1 but more muddy sands.

Table 14.

Hydrofacies of the Organos member(Gómez Arroyo & Gómez Palencia, 2021)

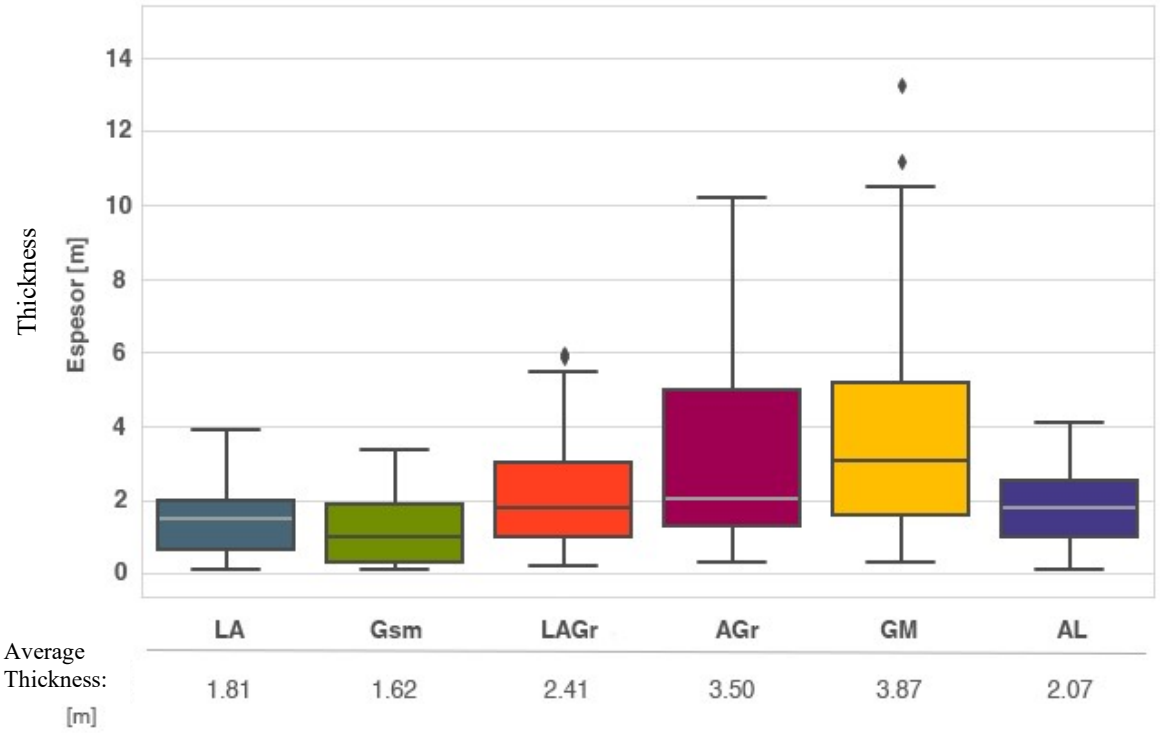
Hydrofacie	Description	Qbo1 proportion [%]	Qbo2 proportion [%]	Qd proportion [%]
Massive gravel (GM)	granular gravels, Pebbly gravels, blocky gravels, clastsupported (framework > 30%) with muddy sandy matrix.	40.48	55.44	45.98
Open-framework gravel (Gsm)	pebbly and granular gravels, locally very coarse-grained sands, with totally or partially washed matrix (little presence of matrix, mainly located on the surface of the clasts or allowing floating contact between them).	31.11	10.32	11.38
Gravelly sands (AGr)	Medium to very coarse-grained sands, granular	11.4	17.39	9.74

Hydrofacie	Description	Qbo1 proportion [%]	Qbo2 proportion [%]	Qd proportion [%]
	and pebbly gravels, with a percentage of gravel between 10% and 30%, occasional presence of clays and silts.			
Muddy sands (AL)	Very fine to fine grained sands, locally medium grained, muddy, slightly granular, with little or no occurrence of gravel (<10%). Sandy silt with a proportion of sand > 10%.	3.95	9.47	21.95
Gravelly Muds and clays (LAGr)	Silt gravelly, pebbly, gravelly sandy; pebbly and sandy clay, with a plastic matrix, with a percentage of gravel between 10% and 30%.	8.28	4.49	7.45
Muds and clays (LA)	Silts, clayey silts, silty clays, and slightly sandy silts, with plastic consistency, restricted granulometric variation, with null or atypical occurrence of non-representative larger clasts.	4.78	2.89	3.5

Most of the geological data comes from drilling cores and given that most of the area is urbanized and there is the high presence of colluvial material (without any clear organization) there is a scarcity of outcrops, then the horizontal continuity of the hydro facies is not known in the area. The vertical continuity is outlined in **Error! Reference source not found.**

Figure 58.

Distribution of the thicknesses of each hydrofacie in the area of study



The typical range of hydraulic properties in every kind of material is shown in Figure *58*. The only pumping test in the area were carried out in a closed well (in the Qbo2) between piezometers P-MS-1(depth of 20 m) and P-MS-2(Figure *17*). The well had 2 filters of 6 meters at depths of 20 m and 32 m and a deeper one of 3m to 50m. The pumping flow yielded by the well was 0.3 liters per second and the results are shown in Table 16. The Pumping test in the area showed a barrier effect, that is associated to impermeable faults. Other measurements of the hydraulic conductivity and porosity were developed in the area like permeability test and others (Table *17*). The laboratory measurements of the hydraulic conductivity tend to be biased to lower values because the drilling cores selected to test were intact due to its high cohesion. Although the

porosity is not directly a hydraulic parameter it is a proxy of the maximum value of the specific yield.

Table 15.

Literature hydraulic values of the geologic material associated to the hydrofacies (Anderson et al., 2015)

Hydrofacie	Geologic material	K (m/s)	Sy (%)	Ss (%)
Massive gravel (GM)	Silty sand/clay (muddy sandy matrix)	1E-11 - 5E-5	1 – 46	4.9E-5 – 1E-4
Open-framework gravel (Gsm)	Gravel	1E-3 – 5E-2	13 - 40	-
Gravelly sands (AGr)	Silty-Clean sand	5E-7 – 5E-3	18 - 43	4.9E-5 – 1E-4
Muddy sands (AL)	Silty sand/clay	1E-11 - 5E-5	1 - 46	9.2E-4 – 2E-2
Gravelly muds and clays (LAGr)	Silt/Clay	5e-13 - 1e-7	1 - 39	9.2E-4 – 2E-2
Muds and clays (LA)	Silt/Clay	5e-13 - 1e-7	1 - 39	9.2E-4 – 2E-2
Rock	Shale, Carbonate rocks, Sandstones; occasionally fractured	1E-5 – 10E-13	1 - 36	3.3E-6 – 6.9E-5

Table 16.

Results of the pumping test in the area (Caballero Caceres & Gomez Navas, 2011)

Puntos de Agua	Tipo de prueba	Transmisivity(m2/d)	Storage
Well	Pumping	0.6588	0.16013
Well	Recovering	0.63248	
P-MS-1	Pumping	8.1782	0.007011

Puntos de Agua	Tipo de prueba	Transmisivity(m2/d)	Storage
P-MS-1	Recovering	11.5692	

Table 17.

*Other hydraulic properties the study area. *Every consolidation test measures the porosity with little pressure (min value) and with a high-pressure equivalent to approximately 80 m of depth. ** to get porosity from the specific weight test, it was necessary to assume some parameters from the literature.*

Parameter	Geology	Min	Max	# Samples	type	Source
Kz[m/s]	Qd	6.13E-10	3.82E-09	2	Variable head	(UIS & AMB, 2018)
Kz[m/s]	Qbo2	2.89E-09	4.75E-07	6	Variable head	(UIS & AMB, 2018)
Kz[m/s]	Qbo1	1.30E-09	1.30E-09	1	Variable head	(UIS & AMB, 2018)
Kz[m/s]	Qbo2	3.34E-10	8.83E-10	5	Variable head	*(UIS & AMB, 2019)*
Kz[m/s]	Qbo1	4.11E-10	4.11E-10	1	Variable head	*(UIS & AMB, 2019)*
K[m/s]	Qbo2	4.03E-10	6.51E-08	19	Lefranc	*(UIS & AMB, 2019)*
n	Qbo2	0.179386	0.316426	6	Consolidation*	*(UIS & AMB, 2019)*
n	Qbo2	0.160863	0.223663	2	Consolidation*	*(UIS & AMB, 2019)*
n	Qd	0.097683	0.187777	4	Specific weight**	(UIS & AMB, 2018)
n	Qbo2	0.136318	0.382338	31	Specific weight**	(UIS & AMB, 2018)
n	Qbo1	0.290951	0.338994	5	Specific weight**	(UIS & AMB, 2018)

In previous studies the hydraulic conductivity of members of the Bucaramanga formation was estimated, by manual model calibration (Table *18*), and by GLM calibration using FePest

(Table *19*), both calibrated in steady state using zonation (regions of equal conductivity associated to geology). Given all this data the range of possible values of the hydraulic properties are broad, and the uncertainty of those values is high. Possibly the flow travel mainly through preferential paths of high hydraulic conductivity because the matrix shows a very low hydraulic conductivity in the laboratory test.

Table 18.

Calibrated values in steady state of a groundwater model of the Bucaramanga Formation(Antolínez Quijano, 2014), Qbo only had the 6% of the observations used in the model calibration (2 observations).

Unit	Kx-Ky min [m/s]	Kx-Ky max [m/s]	Kz min [m/s]	Kz max [m/s]
Qbo	4.2E-5	4.2E-5	1.7E-6	1.7E-6
Qbg	2.54E-5	1.12E-4	1e-5	2.12E-5

Table 19.

Calibrated values of a previous model in the study area (UIS & AMB, 2019)

Unit	Kx-Ky min[m/s]	Kx-Ky max[m/s]	Kz min[m/s]	Kz max[m/s]
Qd	1.62E-7	1.88E-6	9.54E-9	2.01E-8
Qbo2	9.32E-7	2.17E-5	1.4E-7	2.63E-6
Qbo1	9.93E-9	3.83E-8	4.29E-11	1.06E-9
Rock	1E-13	1.96E-8	1.01E-11	1.05E-11
Faults	2.85E-7	1.99E-4	5.11E-9	1.16E-4

Flow Direction, sources, and sinks

The first sketch of the directions of the flow in the Organos Member study area shows roughly a direction of flow from south to north (piezometry in

*Figure **48***), a more detailed potentiometry levels (

*Figure **60***) are seen through the inventory of groundwater points(

*Figure **59***). In the Area of study, the flow comes from the Bucaramanga plateau (main source) and goes towards(sinks) the rivers and the main creek of the area

Figure 59.

Inventory of Groundwater points in the study area

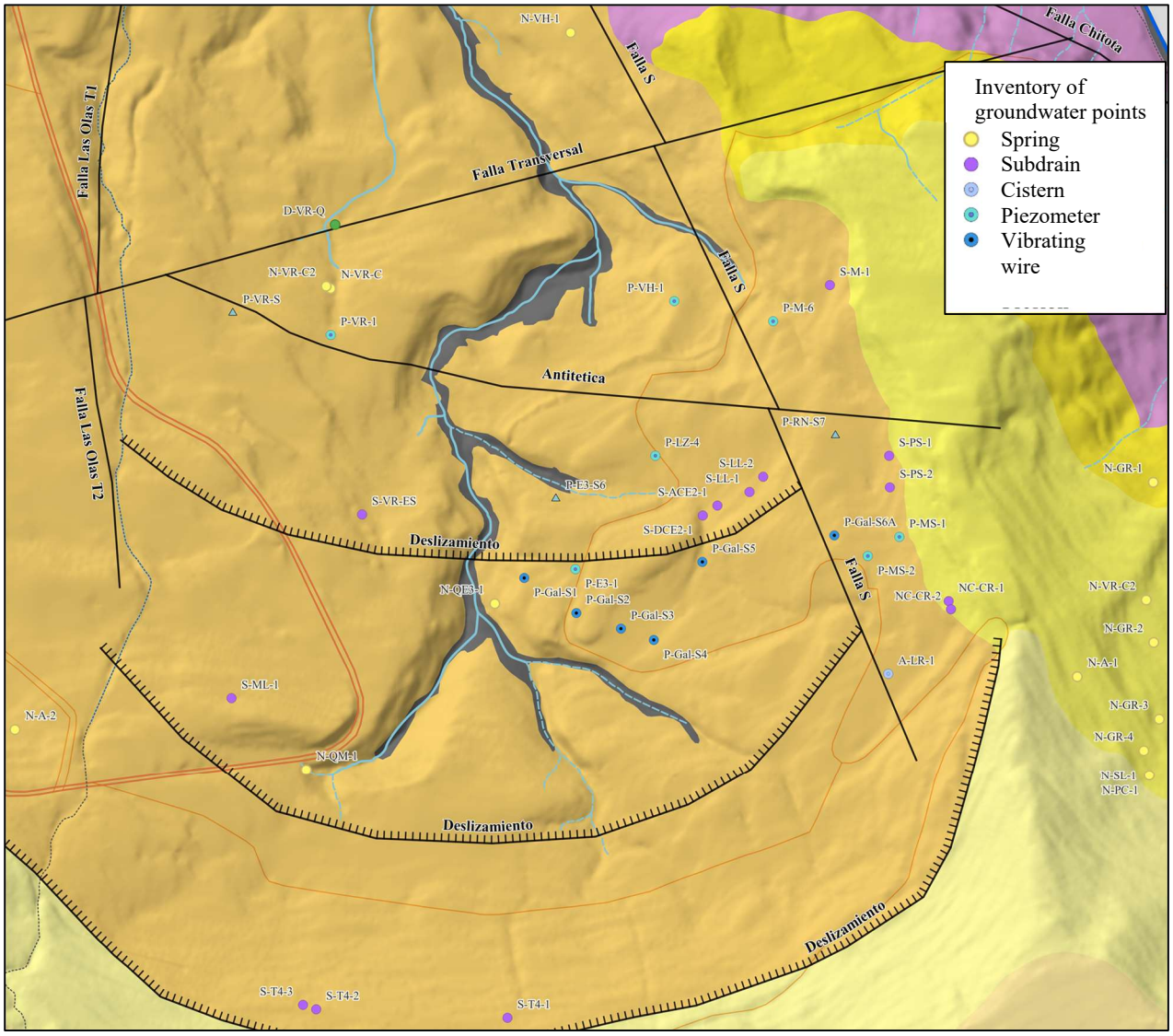
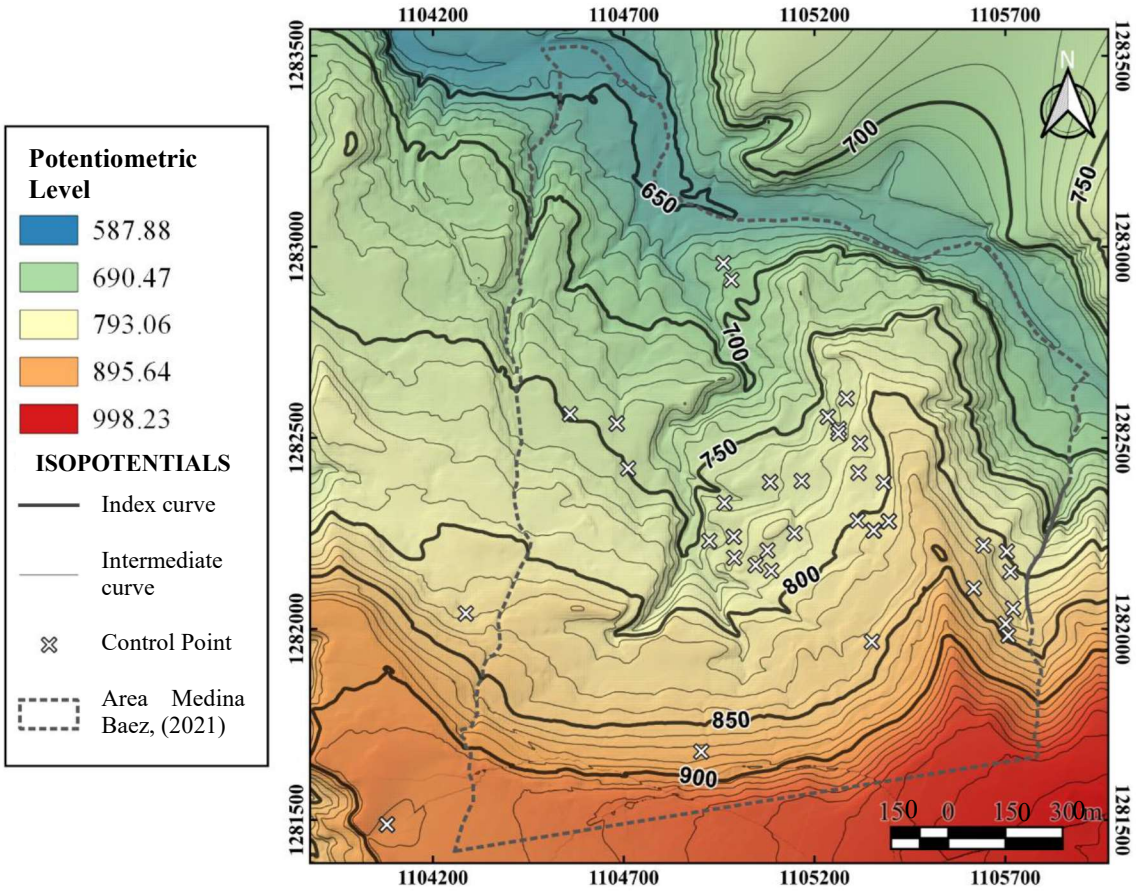


Figure 60.

Average isopotential levels of the groundwater system (Medina Baez, 2021), All the borders except the western border are the same than the borders used in the groundwater model. Coordinate system is in 3116, Magna Sirgas Bogotá zone



The historic behavior of the piezometric measurements is shown in Table 20. Punctual piezometric depths measured in the study area (Medina Baez, 2021) to Figure *66*. Some erratic behavior of levels (Figure 63) is later clipped in the numerical modelling to avoid noise and overfitting.

Table 20.

Punctual piezometric depths measured in the study area (Medina Baez, 2021)

date	P-Gal-S1	P-Gal-S2	P-Gal-S4	P-Gal-S5	P-Gal-S6A	P-LZ-4	P-M-6	P-MS-1	P-MS-2	P-VR-1
12/18/2018						1.90	4.20	3.92	1.20	2.70
1/28/2019								4.27	3.00	
3/2/2019						1.98	4.65	4.13	1.91	2.62
4/5/2019						2.00	2.50	1.64		2.30
5/10/2019	11.10									
11/22/2019	13.30	20.60	7.80	5.40						
2/27/2020	13.50		7.50		18.50		4.90			3.35

Figure 61.

Piezometric depths of vibrating wire RN-S7(UIS & AMB, 2019). Sensor depths are 1:121.5m, 2:91.5m, 3:40.5m, 4:10.5m

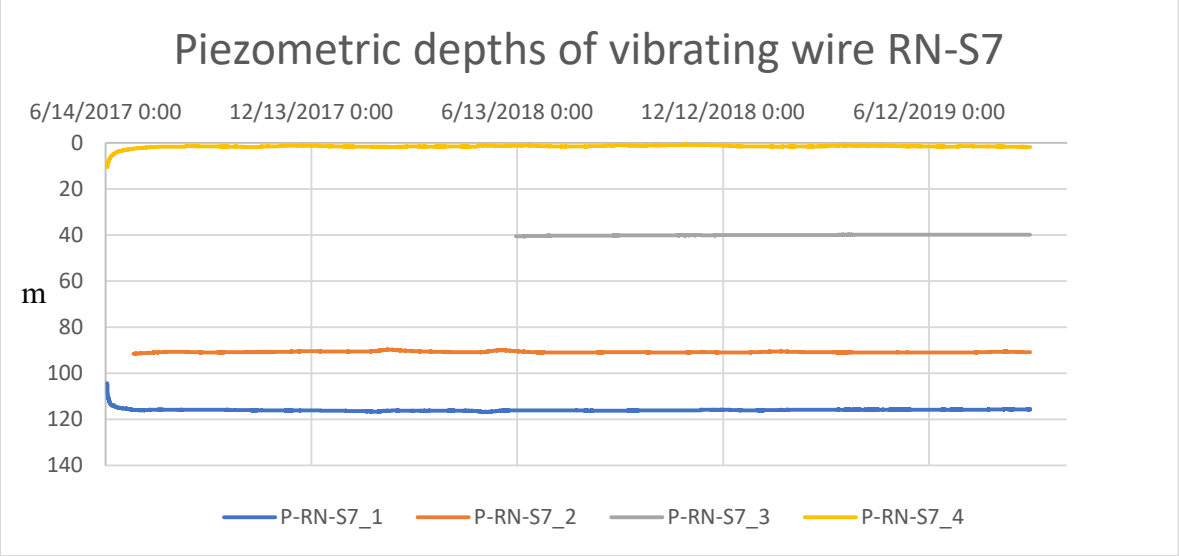


Figure 62.

Piezometric depths of vibrating wire P-E3-S6(UIS & AMB, 2019). Sensor depths are 1:141.5m, 2:50.1m, 3:35.5m, 4:23.5m, 5:9m

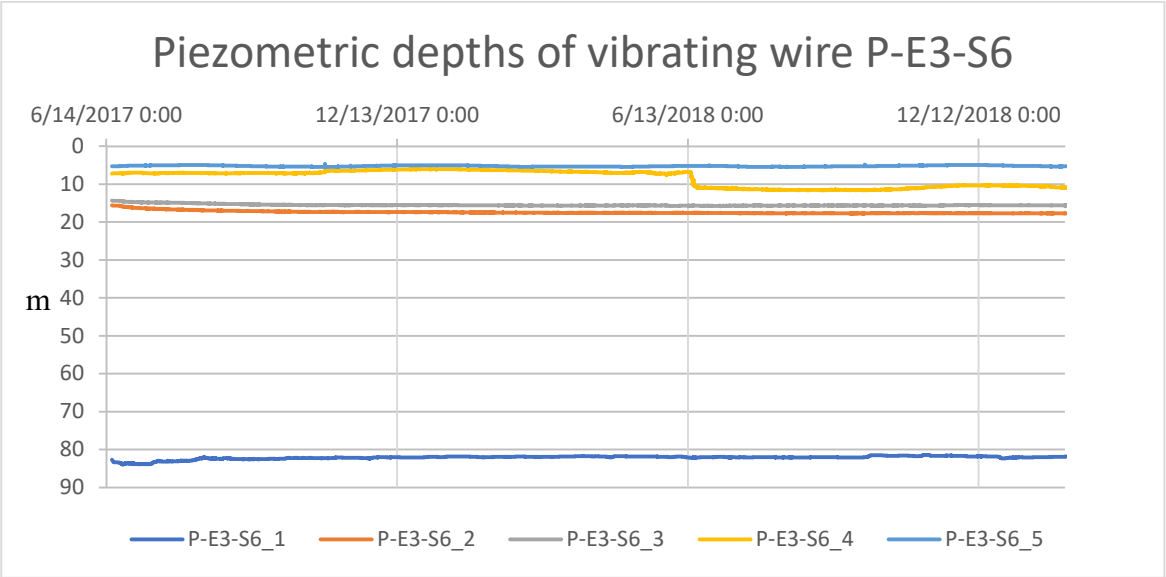


Figure 63.

Piezometric depths of vibrating wire P-VR-S(UIS & AMB, 2019). Sensor depths are 1:77.5m, 2:40.5m, 3:26.5m

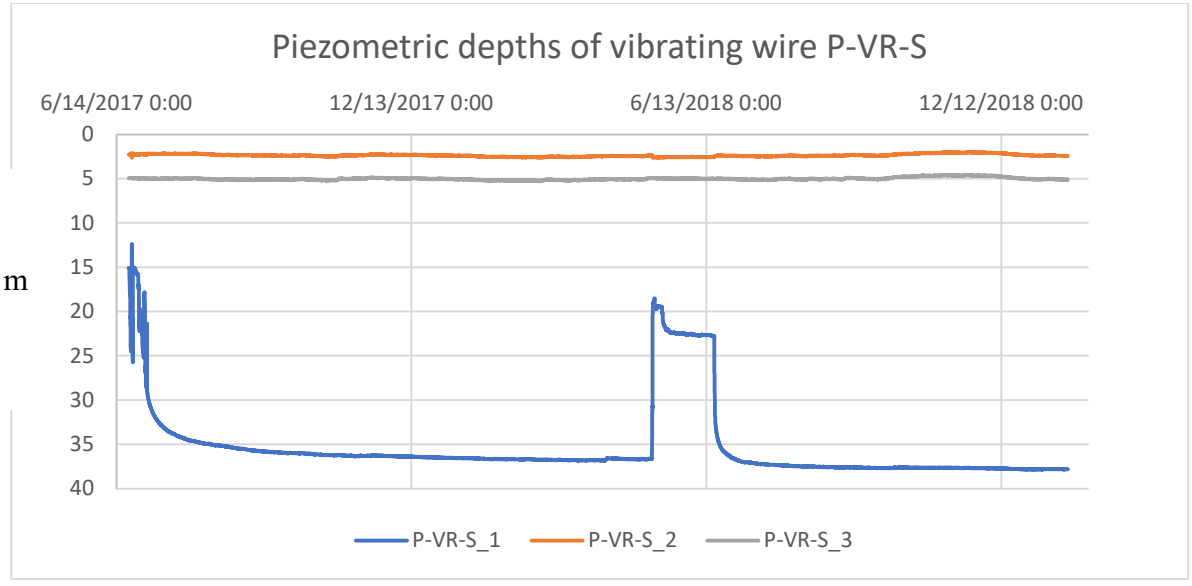


Figure 64.

Depth measurement in piezometer P-Gal-S3(Medina Baez, 2021)

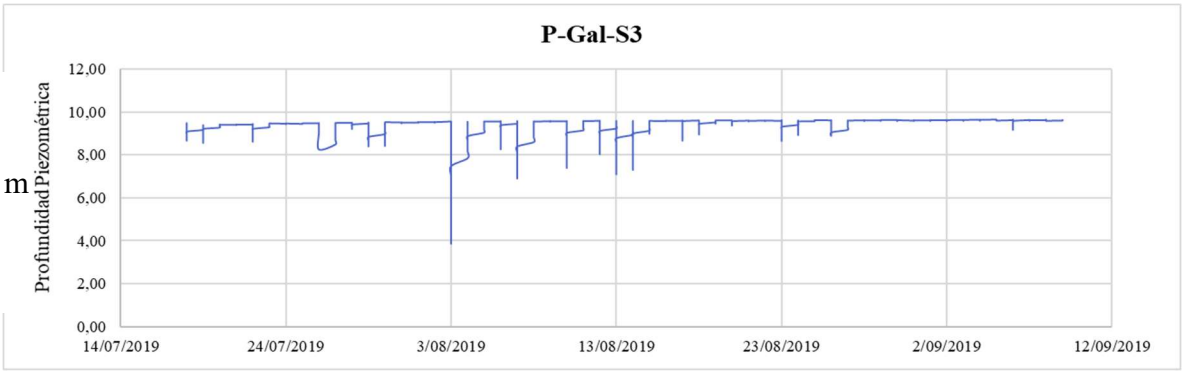


Figure 65.

Depth measurement in piezometer P-Gal-S1(Medina Baez, 2021)

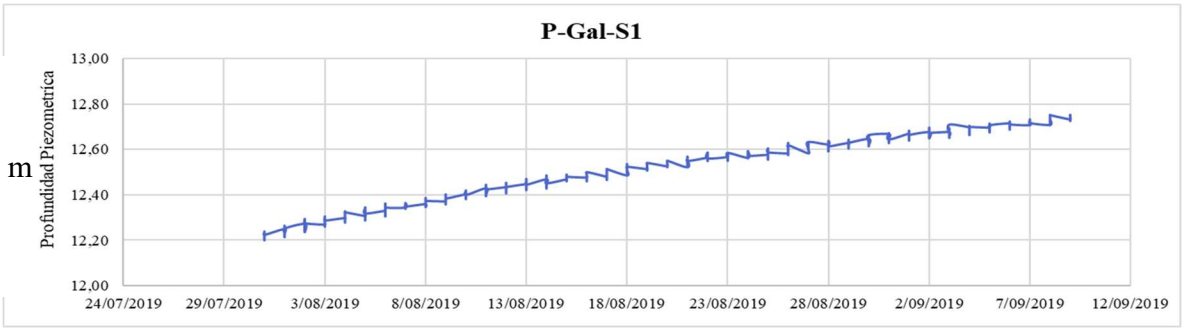
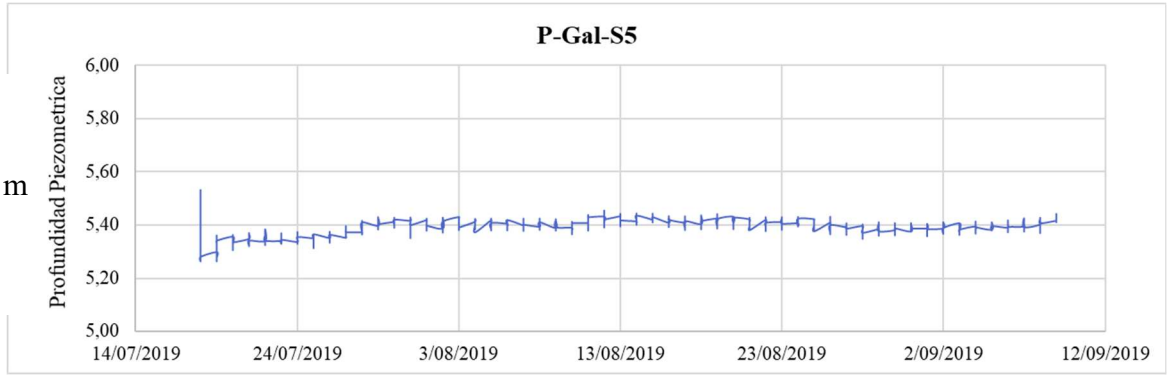


Figure 66.

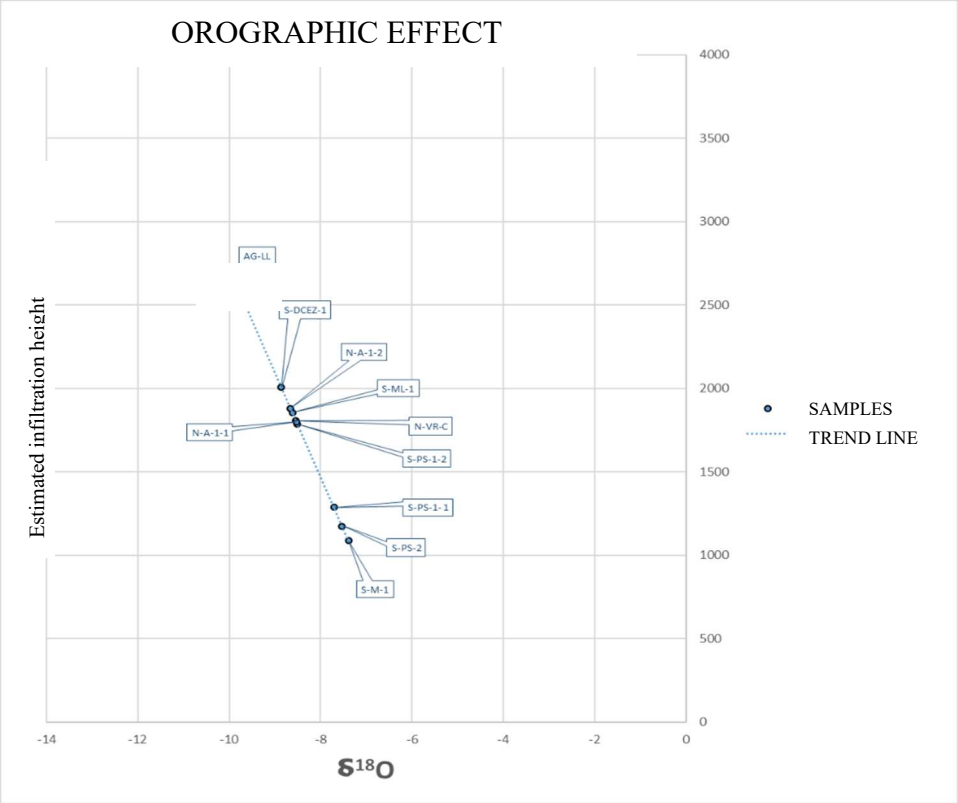
Depth measurement in piezometer P-Gal-S5(Medina Baez, 2021)



To complement the analysis of groundwater sources and recharge the water chemistry is used. The isotopic sampling of rain in the CSRL allowed to identify an orographic effect of the isotopic compositions which allows to estimate the recharge height of groundwater samples(Gomez et al., 2015). Using the orographic gradient of the stable isotopes in samples taken in the study area (Figure *67*) shows that the recharge comes from higher altitudes than the study area (Figure *52*) or at least part of the recharge. It is the same behavior of water samples of the entire Bucaramanga formation where the recharge comes from the medium and upper zones of the CSRL(Gomez et al., 2015).

Figure 67.

Estimated infiltration height of groundwater samples by the isotopic orographic effect (UIS & AMB, 2019)



The electrical conductivity shows values between 170 y 450 mS/cm which are not values of very young groundwater, the Ph is between 5.4 and 8.5 (Figure *69*, down). both the conductivity and Ph shows an anomaly at the east, probably caused by pollutants in the urban area as can be confirmed by e. coli measurements (

*Figure **68***). The ionic distribution is shown in Figure *70* where is worthy to mention that the chloride is a conservative ion which shows a trend of dilution in the direction of flow, that can be attributable to rain recharge(Medina Baez, 2021). The aforementioned is compatible with the fact that most samples showed reducing and transitional environments(Medina Baez, 2021), given that the Aquifer in the Organos member is partially isolated from the atmosphere in the Bucaramanga plateau but is exposed in the area of study to rain recharge.

Figure 68.

Microbiological measurements in the study area (Medina Baez, 2021)

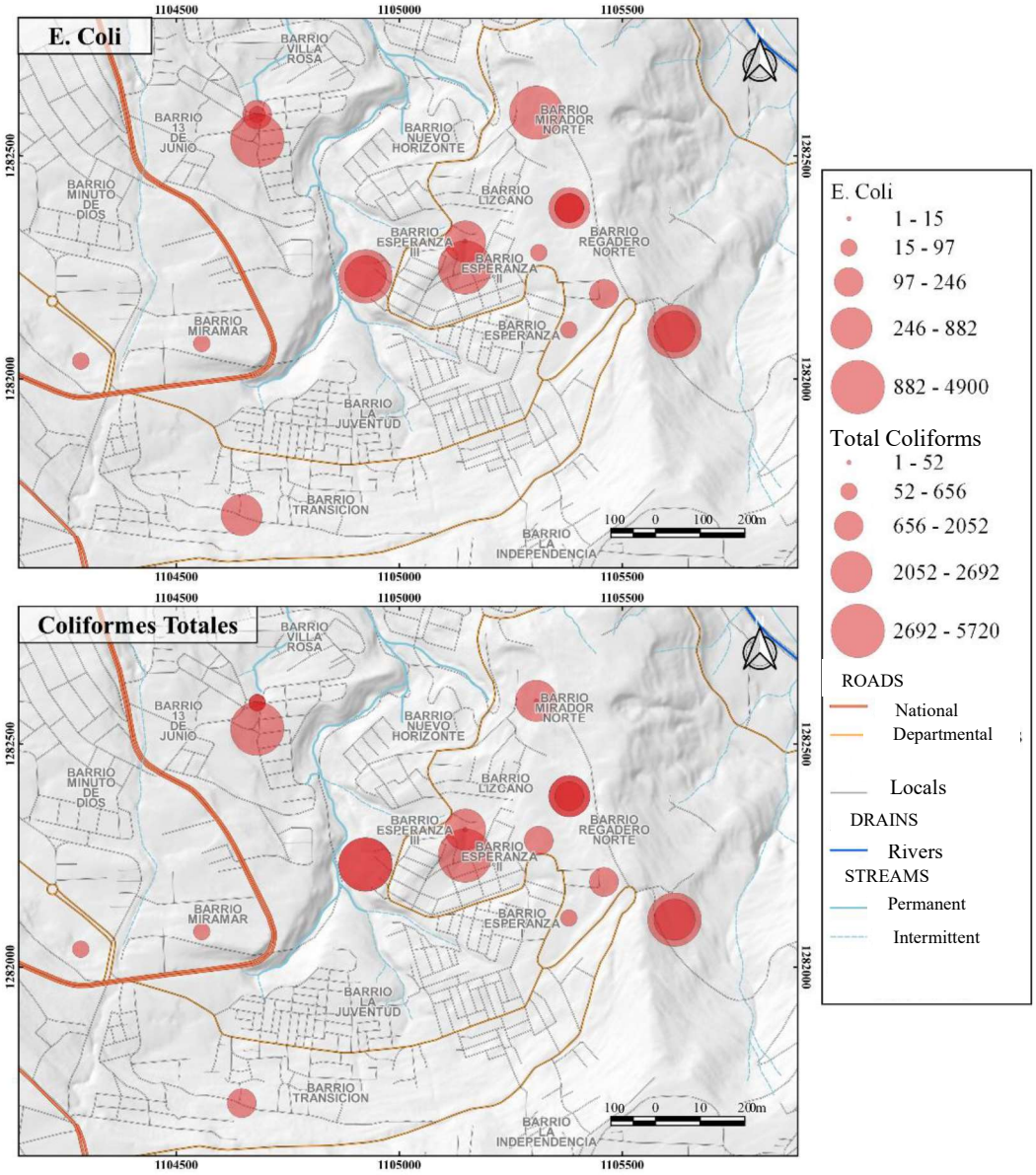


Figure 69.

Electrical conductivity [mS/cm] (up) and Ph (down) of the study area(Medina Baez, 2021)

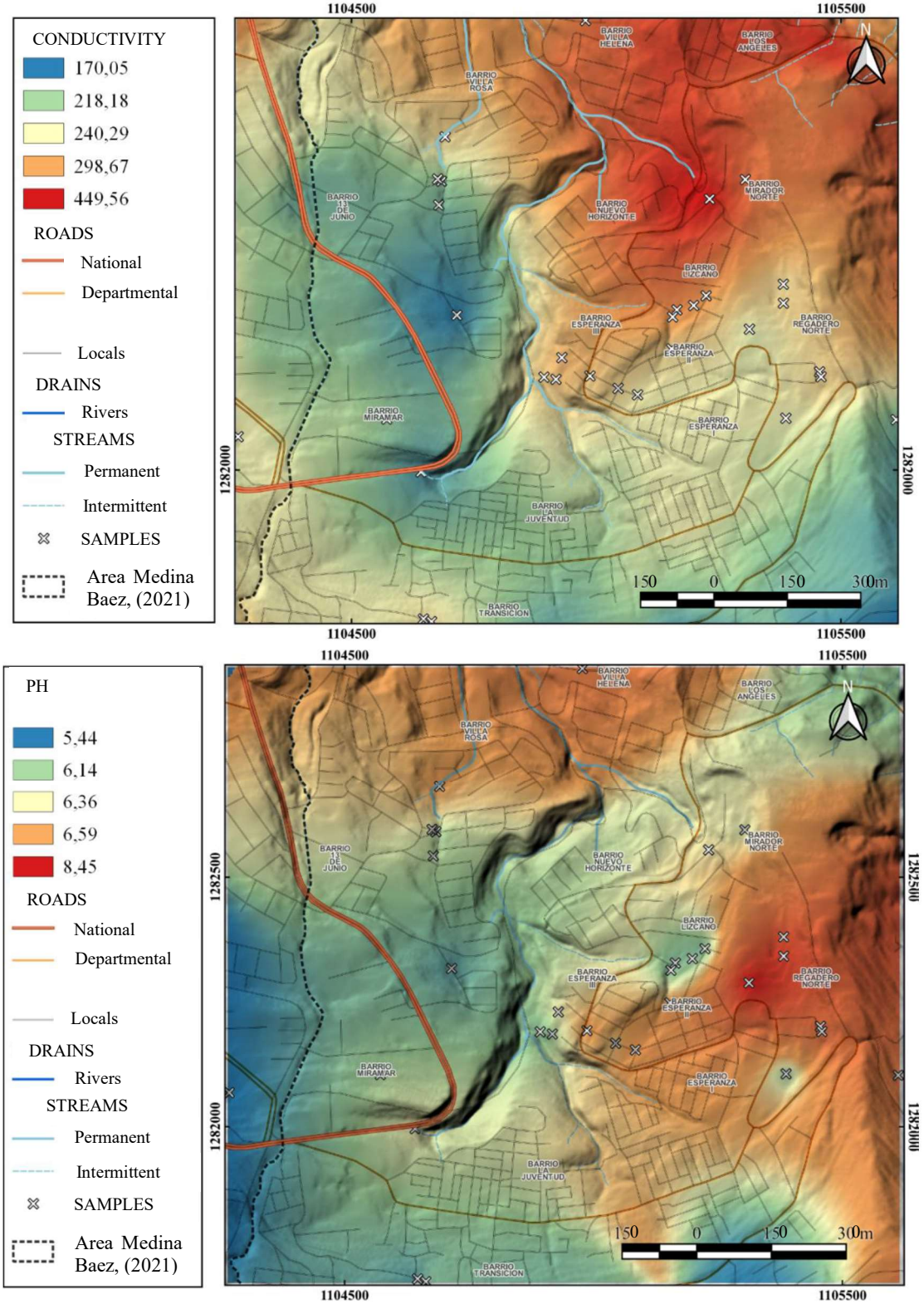
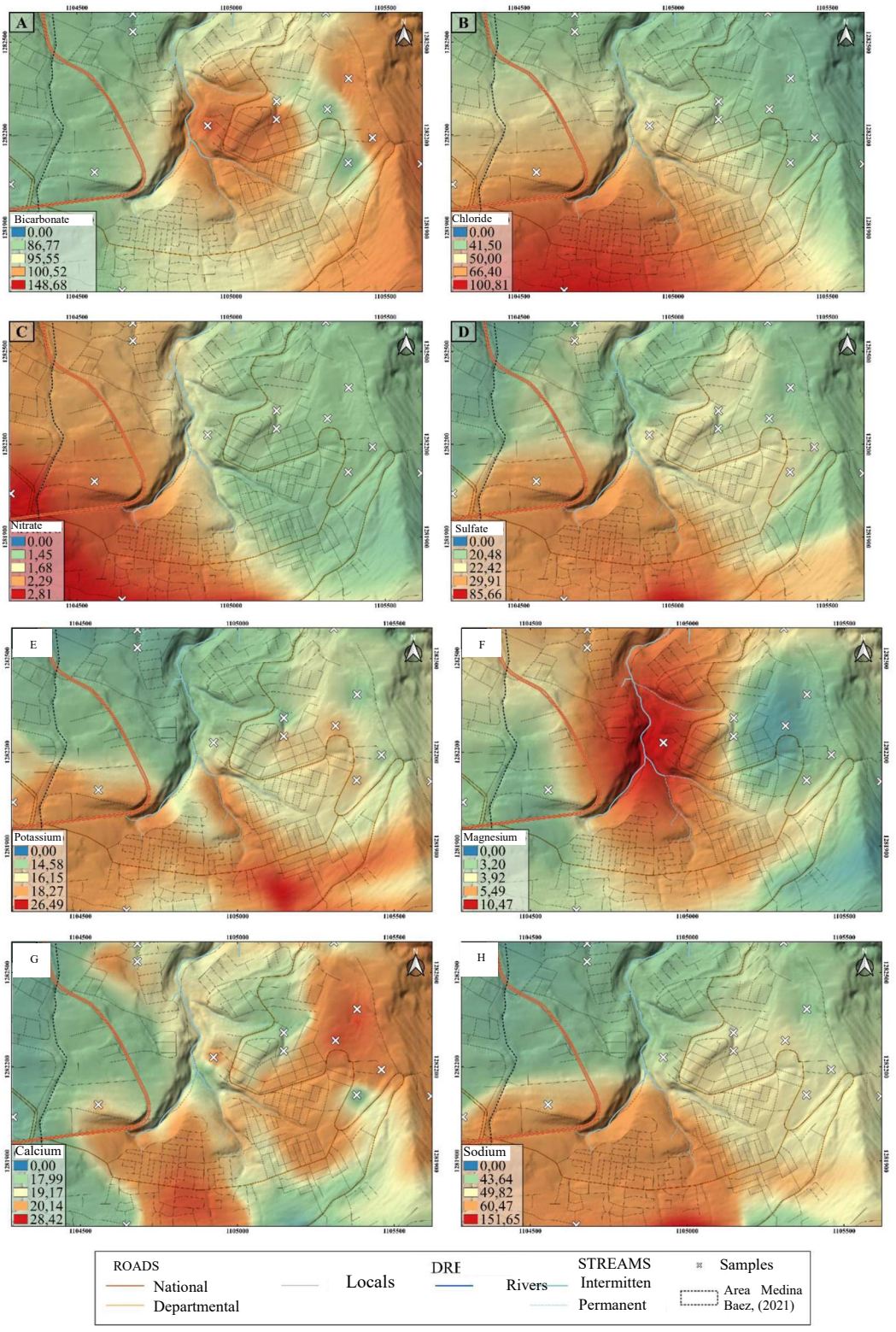


Figure 70.

Ionic concentration(mg/l) in the study area(Medina Baez, 2021),



The rain in the study area is around 1200 mm/yr (Figure *71*) with monthly variations where December, and august are usually the driest months (Figure *72*). The average temperature is 23 celcius degrees, where the estimated real Evapotranspiration is close to de potential evapotranspiration because there is usually an excess of water and is about 1070 mm/yr(UIS & AMB, 2019). Then the Evapotranspiration is approximated the 89% percent of the total precipitation, and recharge will be around 10% or less.

Figure 71.

Precipitation variations from 1982 to 2002 in UIS gauge (A gauge located right next to the south of the study area)(UIS & AMB, 2019)

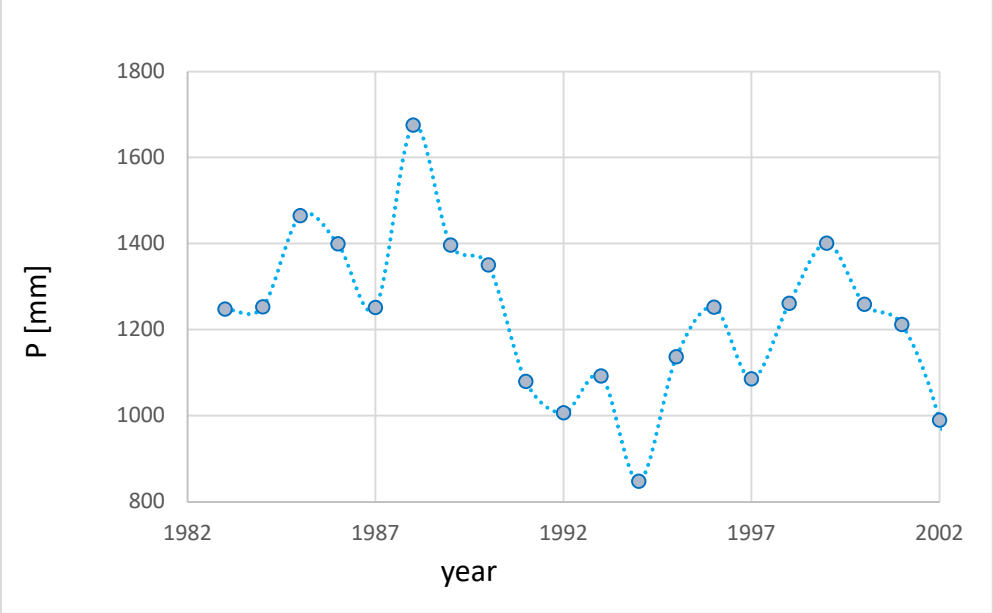
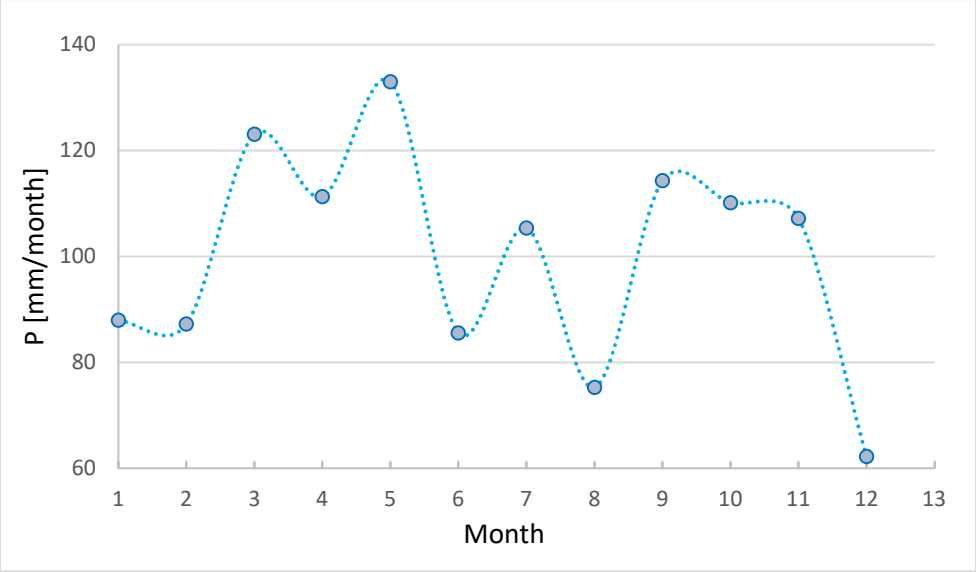


Figure 72.

Multi-year monthly mean rainfall(UIS & AMB, 2019)



Appendix B Graphs of results

See Folder attached