Water Resources Research

RESEARCH ARTICLE

10.1029/2020WR029519

Key Points:

- We built one spatial and two spatiotemporal models using Bayesian Maximum Entropy to estimate Depth to Groundwater (DTG) in a rural region
- The addition of probabilistic data, based on local knowledge collected through Community Science Research, improved the space-time model
- The watershed's portion more likely to have shallow DTG increased from 13% in an average year to 56% in a La Niña year

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. M. Gómez,
amgomez@live.unc.edu

Citation:

Gómez, A. M., Serre, M., Wise, E., & Pavelsky, T. (2021). Integrating community science research and space-time mapping to determine depth to groundwater in a remote rural region. *Water Resources Research*, *57*, e2020WR029519. https://doi.org/10.1029/2020WR029519

Received 23 DEC 2020
Accepted 27 MAY 2021

Integrating Community Science Research and Space-Time Mapping to Determine Depth to Groundwater in a Remote Rural Region

A. M. Gómez[1,2], M. Serre[3], E. Wise[2], and T. Pavelsky[4]

[1]Escuela Ambiental, Facultad de Ingeniería, Universidad de Antioquia, Medellín, Colombia, [2]Department of Geography, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [3]Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, NC, USA, [4]Department of Geological Sciences, University of North Carolina, Chapel Hill, NC, USA

Abstract Continuous depth to groundwater (DTG) data collection is challenging in remote regions. Community participation offers a way to both increase data collection and involves the local community in scientific projects. Local knowledge, which is often descriptive, can be difficult to include in quantitative analysis; however, it can increase scientists' ability to formulate hypotheses or identify relevant environmental processes. We show how Community Science Research can add useful descriptive information for a study based in rural Colombia. To estimate the spatiotemporal distribution of DTG, the community collected water level measurements during a wet (La Niña) year and an average year. We built one spatial and two spatiotemporal models (with and without probabilistic data) using Bayesian Maximum Entropy. Due to the inclusion of local knowledge, the spatiotemporal model with probabilistic data reduced its mean square error by a factor of 15 compared to the spatial model. Using this model, we found that 13% of the study area has a high probability of very shallow DTG (<0.1 m) during an average year, whereas during La Niña, this area increases to 56%. The difference in shallow DTG between the average and wet year implies that after reaching a precipitation threshold, the study region may lose its flow regulation capacity, contributing to flooding during extreme precipitation events. Our approach presents a method to incorporate local knowledge in data-driven models by combining qualitative and quantitative information.

Plain Language Summary Groundwater is a key source of water supply in many regions, supporting crop yields and maintaining water levels in rivers and wetlands. In unconfined aquifers, groundwater may reach the surface during wet periods, contributing to overland flow and intensifying erosion. Identifying groundwater level changes helps to establish water and land management activities. However, continuous depth to groundwater (DTG) data collection, essential for identifying groundwater level changes, is challenging in remote rural areas. We show how Community Science Research, an approach involving active community participation, added crucial information to a statistical model to represent shallow aquifer's groundwater levels in Colombia. The community collected DTG during an extreme wet year and an average year in a middle-low-elevation watershed. We created DTG maps using three statistical models. DTG is better represented by the model that combines descriptive observations with DTG measurements. We also created a map with the probability that the groundwater is near the surface and showed that the area was much larger in the wet year than during the average year. This difference implies that after the watershed receives a lot of precipitation, its flow regulation capacity decreases, which is threatened by land-use activities.

1. Introduction

Depth to groundwater (DTG), the depth measured from the terrain surface to the groundwater table, is essential to identify groundwater availability and groundwater-surface water interactions (Fan et al., 2013). Groundwater can influence the landscape, acting as a discharge system feeding the tributaries that support streams (Margat & van der Gun, 2013), creating water-logged soil conditions that define wetland ecosystems, and supplying water to the root zone to maintain plant photosynthetic activity (Lewandowski et al., 2019). The landscape also influences groundwater as groundwater flow is often related to topography.

Surface waters and areas where deep infiltration occurs can recharge aquifers (Sophocleous, 2002; Winter et al., 1998). In addition to its ecosystem function, groundwater is the primary water supply in many regions, sometimes acting as the only source (UNESCO & UN-Water, 2017). In the absence of regulation, land use activities may substantially change shallow groundwater systems' capacity to regulate flow or attenuate groundwater pollution (Gleeson et al., 2016). In precipitation-driven groundwater systems, extreme rainfall changes can cause groundwater table subsidence or flooding (Marchetti & Carrillo-Rivera, 2014), influencing erosion. These effects can subsequently impact ecosystem function and services. Although continuous data collection is challenging, these data are needed to understand the links of shallow groundwater to the landscape, estimate groundwater storage, and establish limits of extraction. Large, long-term DTG datasets are mostly limited to developed countries (Fan et al., 2013), constraining the understanding of groundwater function and services and limiting land and water management decision-making in these regions.

Community participation in scientific projects provides a way to address environmental questions with a meaningful social impact and to reduce information gaps (Arias et al., 2016; Sandoval, 2004; Wiggins & Crowston, 2011). Approaches to community participation in scientific projects can be defined as a function of the type of relations, the strategies implemented, and the level of engagement developed between the public and the researchers (e.g., citizen science, participatory action, community-based, social monitoring, etc.). In this study, we use the term Community Science Research (Cooper et al., 2007) to describe research projects in which the public participates in significant ways. Significant participation can take place at different stages where the participants engage in activities that may last beyond the projects (Wijnen et al., 2012), influence local governance decision-making processes (Arias et al., 2016), and help solve scientific questions that increase the system's knowledge (Baldwin et al., 2012; Le Coz et al., 2016).

Community science projects have a long history in data collection in several biogeoscience disciplines. In hydrology, the participation of the public and explicit design of community science projects has been increasing (Buytaert et al., 2014). There are several existing examples of community participation in water level and precipitation data collection (Piragua project, Colombia, https://www.piraguacorantioquia.com.co/piragua/; Pluviometros Cuidadanos, Chile, http://milluvia.dga.cl/index.php; CoCoRaHS, USA, https://www.cocorahs.org/"; K. E. Little et al., 2016; Lowry et al., 2019; Weeser et al., 2018); its use in modeling river discharge (Starkey et al., 2017); and on identifying lake water storage changes (S. Little et al., 2021). Most projects concentrate on data collection, leaving data analysis and modeling to the scientist (Assumpção et al., 2018; Njue et al., 2019). Beyond data collected by the local population, their knowledge of the landscape also encompasses local environmental conditions that help to define the logistics to access or install measurement devices and provide a qualitative understanding of hydrologic systems. This type of information is not always formally collected or included in the analysis and model implementation process. Therefore, high levels of active participant engagement and valuable qualitative descriptions are left out of the quantitative data analysis. The qualitative nature and the potential high levels of uncertainty of the descriptions made by local inhabitants make it challenging to incorporate into the model construction. Overcoming this challenge will be key in integrating community knowledge into models and data analysis.

One potential solution to add local knowledge to models is through the creation of a probabilistic representation using Bayesian Maximum Entropy (BME). BME incorporates general information about the salient variable (spatial dependencies, conceptual assumptions, etc.) by maximizing an entropy function (Christakos, 1990; He & Kolovos, 2018; Serre & Christakos, 1999). BME is an extension of linear geostatistical approaches (kriging-based methods). It can combine data that carry higher levels of uncertainty, known as soft data, with spatiotemporal measurements that have lower uncertainty, known as hard data. In groundwater hydrology, BME has been used to effectively map groundwater flow direction using water table data (Serre & Christakos, 1999). In this study, each measurement had a source of error reported during the monitoring campaigns. Those water table measurements with an identified source of randomized errors or errors manually flagged by experts were defined as soft data. Hard data were defined as water table measurements with no recorded random errors and no errors manually flagged by experts after data collection. Another example of applied BME estimates aquifer hydraulic conductivity (Serre et al., 2003), soft data correspond to hydraulic conductivity derived from porous media analysis at different sites and hard data are hydraulic conductivity measurements performed at specific locations. These combinations of soft and hard data minimized estimation errors of the hydraulic subsurface properties in saturated media.

BME has also been applied to assess water quality by combining space-time variability of the water quality within the river network (Akita et al., 2007). In this study, soft data included detected nitrate concentrations in water, using concentrations under the detection limit, while hard data included nitrate concentrations over the detection limit (Messier et al., 2014). Also, BME served to combine different temporal scales of arsenic concentrations, with soft data defined as arsenic concentrations at coarser temporal scales and private wells, and hard data as the concentrations officially provided by the local authorities (Sanders et al., 2012). In all these cases, the use of BME approaches led to an improvement in the detection of water quality levels. BME has also been used to design monitoring networks (Hosseini & Kerachian, 2017) by incorporating the uncertainty of having areas with no monitoring stations and different start monitoring dates in the modeling design. These examples show the potential of BME to include ancillary information with high levels of uncertainty in data-driven models.

To our knowledge, BME has not previously been used to model DTG. In the absence of continuous measurements, the shallow groundwater table is often assumed to be a function of the topography and simulated using classic geostatistical interpolation methods, which require high resolution elevation maps. However, topographic information is often only available at coarse resolution in remote rural regions. Therefore, modeling DTG as a function of the topography may hide shallow DTG in lowlands, close to depressions, that may impact identification and mapping of near-saturated areas that affect discharge (Snyder, 2008).

The central research questions of this study are: How can descriptive information be incorporated into DTG mapping? Does the inclusion of this information improve model performance? We show the potential for combining DTG quantitative and qualitative data collected in a community science project in the 481 km^2 Man River middle-low elevation watershed (Bajo Cauca region, Colombia). The data collected by the community previously helped to generate monthly groundwater table maps using classic kriging interpolation (Palacio, 2014). This interpolation used the elevation from the Shuttle Radar Topography Mission data set (http://srtm.csi.cgiar.org/), with a 30 m spatial resolution, SRTM-30. The resulting maps were useful to identify potential flow direction in the shallow aquifer. However, monthly average groundwater table estimates may hide peaks, missing potential high and low values and making it difficult to identify rapid water table responses to precipitation. We present a different approach to obtain weekly DTG. The regular flooding in the region during extreme wet seasons (Betancur-Vargas et al., 2017) motivated us to apply our results to address how extreme precipitation influences changes in DTG.

2. Materials and Methods

2.1. Study Area

The 481 km^2 Man River middle-low watershed drains the foothills between the Western and the Central Andes Cordillera in Colombia (Figure 1). The relief is low with a landscape characterized by extensive valleys and rolling hills. The middle-low watershed's elevation ranges between 12 and 148 m.a.s.l.. The Man River, originating at 1,050 m.a.s.l., is a tributary of the Cauca River, a main river in Colombia. Its most important tributaries are the Quebradona and the Ciénaga Colombia creeks (Figure 1a). The Ciénaga Colombia, a wetland to the north-east, regulates flow during the wet season and provides ecological and cultural services to the region's socio-economic development (Santa-Arango et al., 2010). The climate is humid (average relative humidity 78%) and warm (average temperatures between 25 and 30 °C). The average precipitation is 2,800 mm/year, with a dry period between December and March and a wet period between April and November. The region's geology is primarily composed of Tertiary clastic sedimentary rocks of continental origin, represented by the Cerrito Formation (Ngc), which occupy 404.3 km^2 of the watershed; Sincelejo Group sedimentary rocks (NgQs), 17.5 km^2; and recent Quaternary alluvial deposits (Qal), 57.4 km^2 (Figure 1c). These features form three hydrogeological units: a shallow aquifer, an aquitard, and a confined aquifer (Betancur et al., 2012). Our work centers on the shallow aquifer, formed by Man River alluvial deposits and non-consolidated saprolite from the Cerrito Formation, with depths ranging between 13 and 65 m (Palacio et al., 2013). Previous studies in this region identified groundwater and surface water connections (Santa-Arango et al., 2010), and groundwater recharge occurring directly from precipitation (Palacio & Betancur, 2007).

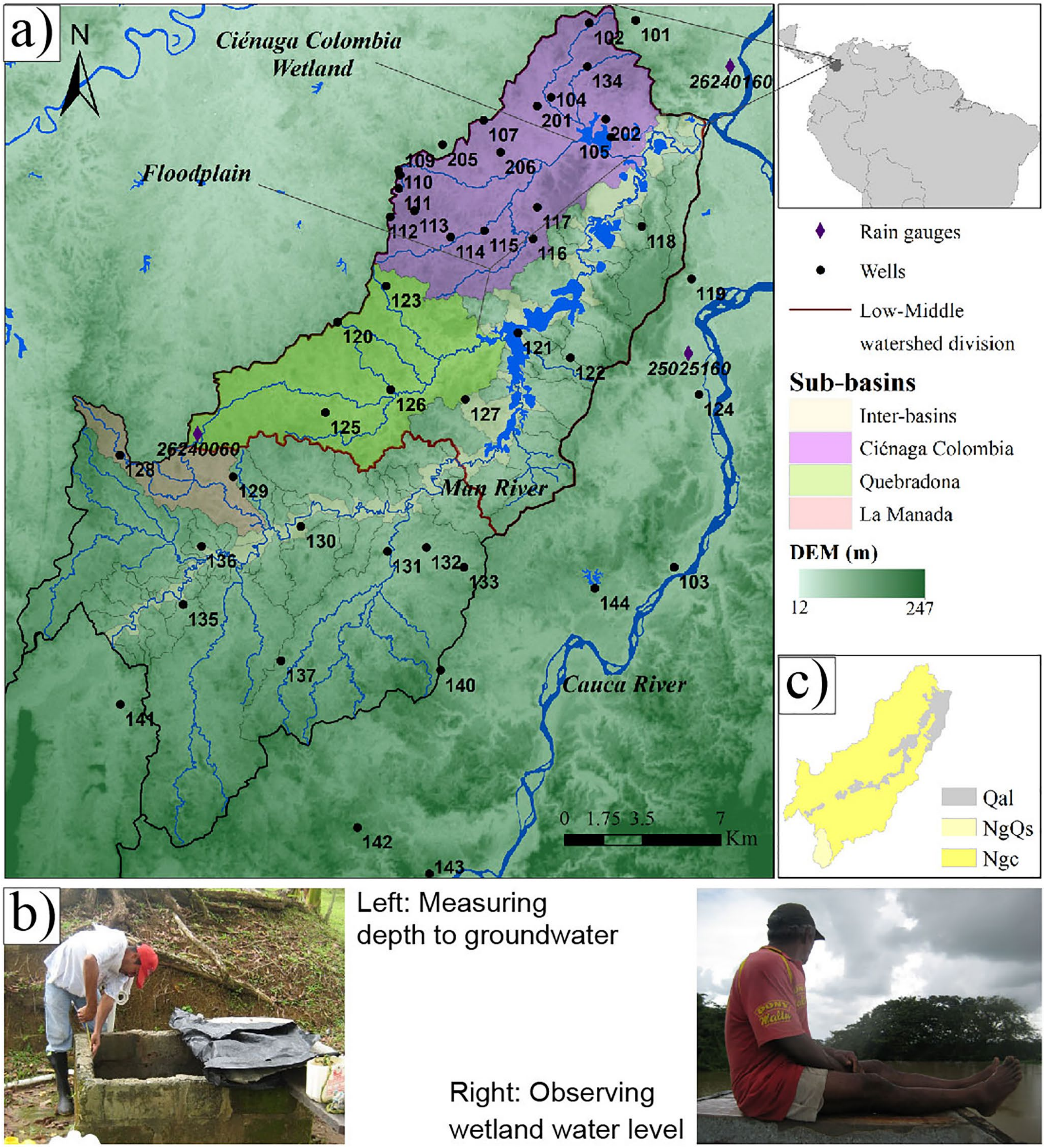

Figure 1. Study area and community science activities. (a) Wells used as monitoring stations. (b) Activities developed by the community (source: Grupo GIGA, Universidad de Antioquia). (c) Geology of the region describing the shallow aquifer: Qal, recent Quaternary alluvial deposits (12% of the watershed); Ngc, Cerrito Formation (84%); and NgQs, Sincelejo Group (3%).

A 2003 water well inventory indicated 147 dug and five drilled wells in this area; 86% are for domestic use (CORANTIOQUIA & Universidad de Antioquia, 2003). At the time of this project, no water supply treatment system was operational in the area. In 2011, 66.8% of the population from the municipalities where the watershed is located lived in poverty (Departamento Administrativo Nacional de Estadística, 2012).

A few people engage in small-scale agriculture and fishing, but the main economic activity is cattle raising (Instituto Geográfico Agustín Codazzi, 2007a). Cattle ranching occupies around 80% of the watershed (Instituto Geográfico Agustín Codazzi, 2007b) and along with strip mining, land conflicts, and illicit crops, leads to population displacement (Cuartas et al., 2000).

2.2. Community Science Data Collection

The DTG monitoring network started with the analysis of a conceptual hydrogeological model for the Bajo Cauca region (Betancur, 2008), that estimates groundwater flow directions at a regional scale and identifies aquifer units and their thickness. The potential wells and associated collaborators were obtained from the 2003 well inventory (CORANTIOQUIA & Universidad de Antioquia, 2003) based on three criteria: (1) spatial distribution across the watershed, (2) accessibility, and (3) well depth. The wells' spatial distribution was selected based on areas where groundwater-surface water interactions were and were not expected. These decisions were based on expert knowledge of the region and the conceptual hydrogeological model. The accessibility conditions aimed to facilitate visits to the sites and access to a good cellphone signal to receive data from the collaborators. Finally, the wells' depth helped to identify the hydrogeological unit from which water is extracted. With this selection, six field campaigns were conducted. In the first reconnaissance campaign, each household's well location and contact information registered in the well inventory was confirmed with the collaborators at each house and farm where the wells are located. The rest of the sampling campaigns were designed to reinforce data collection procedures, collect hydrogeochemical data, share guidance, and follow up on the maintenance of each well's sanitary condition.

The campaigns were designed to facilitate data collection with limited resources. Potential collaborators were introduced to the project, and characteristics of water wells, pumps, and pumping times were identified. Collaborators were trained in water depth measurements at each site and provided with a tape measure and forms to register measurements. Each collaborator was trained individually. The distance was measured from the top of the well to the water table weekly before pumping. The height from the top of the well to the ground was subtracted later to obtain the DTG (Figure 1b). Weekly phone calls gathered information and verified anomalous values, representing extremely low or high DTG. The verification followed a 2-step process. The first step was to verify by phone the time of the day and how the collaborator was measuring. The second step was during the sampling campaign. In most cases, phone call verification was enough to improve data collection. Some collaborators moved during the study, which is common in this area; data were either no longer collected or the collaborator trained a new person to monitor their well. This practice of knowledge transfer was identified in highly engaged participants. In addition to obtaining DTG, collaborators located close to the wetland area or the Man river described weather conditions and observed surface water levels. These descriptions were not systematically collected among collaborators but were mostly provided during the phone calls to collect the data or during the sampling campaigns. After the third campaign, the monitoring network was updated to include additional wells, following the same procedure established with the first group of wells. Ultimately, we managed to collect a total of 2,397 high-quality noncontinuous data at 44 wells between 2008 and 2009. These data would have been challenging and expensive to acquire with a traditional approach.

Although the project did not provide monetary compensation to the collaborators, hydrogeochemical data collected to identify groundwater sources and surface-groundwater interactions (Santa-Arango et al., 2010) were reinterpreted in terms of their sanitary meaning, and the results were provided to the collaborators. Those data were for the benefit of the collaborators and are not part of this research study. In addition, after finding poor sanitary conditions at some of the wells, recommendations were provided, such as covering wells to prevent animal and debris access and simple disinfection techniques (Organization World Health, 1999). The data collection team followed up with each household regarding these recommendations. Some households shared advice with other relatives or friends not included in the monitoring network. We did not measure or collect information on how much of this information transfer occurred among households.

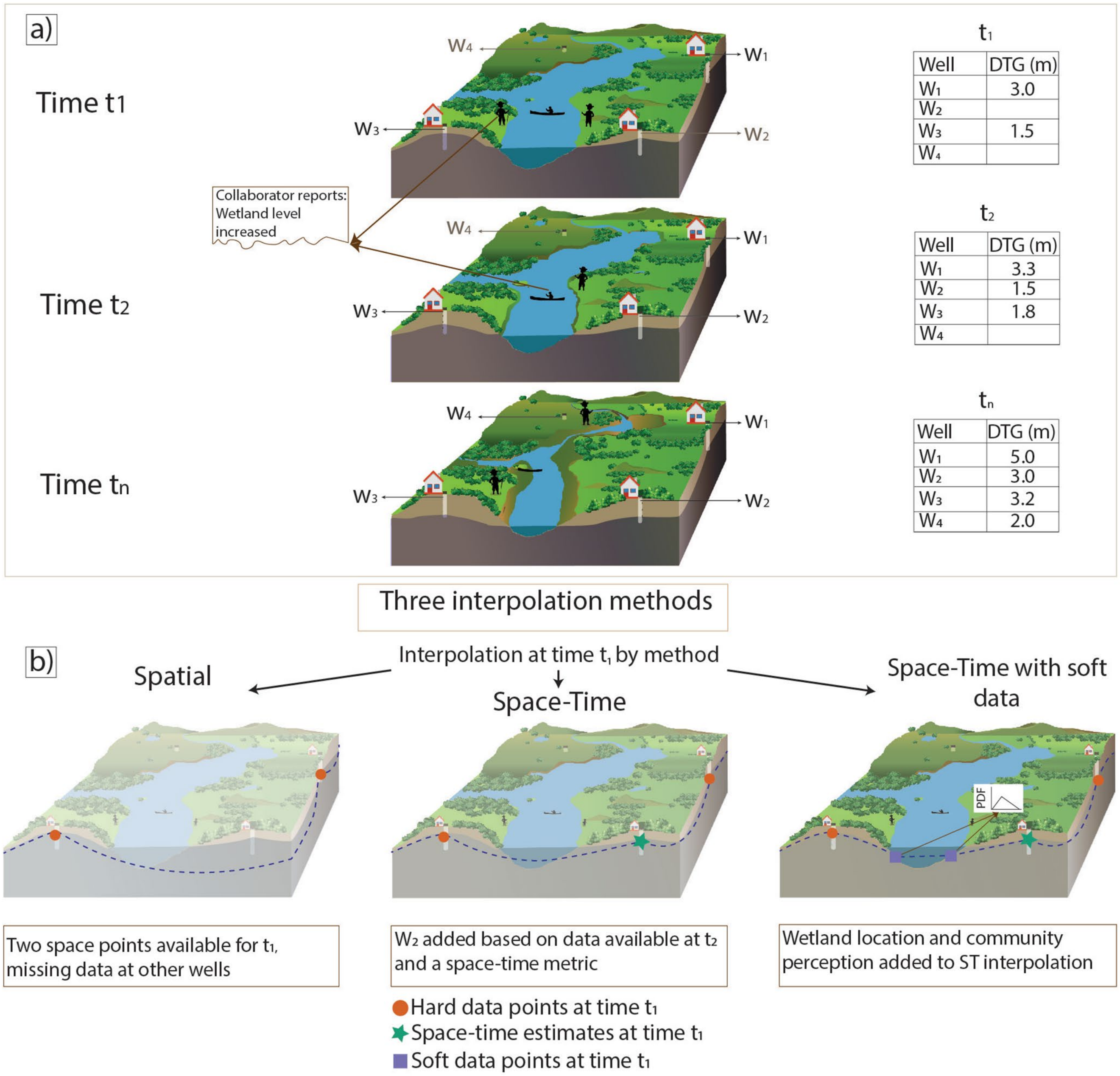

Figure 2. Schematic differences in the interpolation methods. (a) Shows a watershed conceptual block-diagram closed to the wetland location at three different times: t_1 and t_2 correspond to the wet season and t_n to the dry season. Tables t_1, t_2, and t_n show examples of the data collected. Collaborators qualitatively reported the wetland stage, mainly during the wet season. (b) Depth to groundwater (DTG) interpolation example at time t_1(dashed line). Spatial interpolation (S) interpolates only data collected at t_1; space-time interpolation (ST) incorporates correlations in time of DTG to improve interpolation results; in the space-time with soft data (STSD) interpolation, descriptions from collaborators are incorporated to further improve results.

2.3. Model Framework and Implementation

In this study, we built three model combinations using the BME framework: spatial interpolation with only observed DTG at each monitoring wells, that is, hard data (S), space-time with only hard data (ST), and space-time with probabilistic or soft data, that is, data corresponding to expected low DTG at the wetland locations (STSD). In the spatial interpolation, S, we interpolated each week independently, assuming no correlation in time, a common approach in classic geostatistical analysis. For the space-time interpolations (ST and STSD) we hypothesized high correlations of DTG in time would improve DTG estimates (Figure 2).

The summary of the BME framework provided below is adapted from previous work; the reader is referred to these papers for more details on the mathematical underpinnings of the BME framework (Christakos et al., 2001; Serre & Christakos, 1999). The BME framework is a nonlinear extension of the classical kriging methods of linear geostatistics. BME has the capacity to use only the knowledge base kriging can process. In that case, the BME equation reduces to a linear kriging estimator with a Gaussian posterior probability distribution function, PDF. However, using information processing principles such as information entropy maximization and epistemic Bayesian principles, the knowledge base that BME considers can extend beyond those used in kriging.

Let $X(p)$ be a space-time random field (S/TRF) representing the variation of DTG across space-time coordinate $\boldsymbol{p}$=($\boldsymbol{s}$,t), where $\boldsymbol{s}$ is a geographical location, and t is time. We denote as $G = \{m_X(p), c_X(p,p')\}$ the general knowledge characterizing $X(p)$, where $m_X(p)$ and $c_X(p,p')$ are the mean and covariance of $X(p)$, respectively. We denote $S = \{x_h, f_S(x_s)\}$ the site-specific knowledge characterizing the data at hand, where x_h are the space-time hard data DTG values observed at each monitoring wells, x_s are space-time soft data DTG values corresponding to where we know the DTG is low, and $f_S(x_s)$ are the density functions characterizing the shallow DTG and its uncertainty.

BME can be summarized in three stages: Prior, Posterior, and Interpretive. At the prior stage, we use the *Maximum Entropy* principle of information theory to create a prior PDF $f_G(x_{map})$ that integrates the general knowledge G, where $x_{map} = (x_k, x_h, x_s)$ is the value of $X(p)$ at points $p_{map} = (p_k, p_h, p_s)$ and where p_k is an estimation point of interest. The PDF statistical properties are consistent with G, and maximize the amount of choice in the value DTG can take. G consists of the knowledge of the mean and covariance of $X(p)$, which are statistical moments up to order two only. As a result, f_G is the multivariable Gaussian PDF with means and covariance specified in G.

At the posterior stage, we integrate the site-specific knowledge S using an epistemic *Bayesian* conditionalization Equation 1.

$$f_K(x_k) = A^{-1} \int dx_s f_S(x_s) f_G(x_k, x_h, x_s) \tag{1}$$

In Equation 1, A is a normalization constant. This equation creates the BME posterior PDF, $f_K(x_k)$, providing a full stochastic representation of $X(p_k)$, the DTG at the estimation point p_k.When we restrict the analysis by not using the soft data, Equation 1 reduces to Equation 2:

$$f_K(x_k) = f_G(x_k, x_h) / \int dx_k f_G(x_k, x_h) = f_G(x_k \mid x_h) \tag{2}$$

which is the conditional PDF x_k gives x_h under the general knowledge base G. Since $f_G(x_k, x_h)$ is Gaussian, the conditional PDF in Equation 2 is also Gaussian, which means the PDF is a linear combination of observed values that correspond to the kriging approach. Since the mean is assumed constant and calculated within a local estimation neighborhood, BME reduces to moving window ordinary kriging when only household well observation data are used. This is the case of the S and ST approaches. This kriging limiting case makes BME attractive since it means that BME reduces to a linear kriging estimator whenever the analysis is restricted to hard data, but it extends to a nonlinear and non-Gaussian estimator when soft data is used. Finally, at the interpretive stage, we calculate the BME mean $\tilde{x}_k$ and corresponding posterior BME variance $\tilde{v}_k$ of the BME posterior PDF at estimation points p_k on an estimation grid to obtain BME estimation maps and the corresponding uncertainty of DTG.

Prior to model implementation, we examined the general spatial and temporal distribution of the raw data. Because of its high right skewness (most of DTG are close to 0), we normalized the data by transforming DTG using natural logarithm, and used it in all the interpolation methods expressed in ln(depth [m]) units, here denoted as ln-depth. This decision was made to have comparable results across the models since BME reduces to a linear Gaussian estimation when only wells measurements are included. We defined x_h as all the 2,395 space-time hard data DTG values observed at the 44 household monitoring wells, x_s as the 2,600 space-time soft data DTG values corresponding to where we know the DTG is low (i.e., these points are geographically located at 50 nodes uniformly distributed across the wetland area and during the wet season, April-November), and $f_S(x_s)$ were the corresponding 2,600 triangular probability density functions characterizing the shallow DTG and its uncertainty.

For STSD, the triangular PDF was based on precipitation records combined with community members' descriptions of the water levels in the wetland and floodplains during wet months (end of April to November). The collaborators reported increasing wetland surface levels and shallow DTG in nearby wells. These observations, while not quantitative, gave us tools to hypothesize the aquifer connections to the wetland. Using collaborators' qualitative statements, we defined a probability function of shallow DTG in the wetland during the wet season. First, we used the piezometers 201 and 206 installed in the Ciénaga Colombia sub-catchment (Figure 1a) to identify the relation between the intra-annual precipitation pattern and groundwater level change. These piezometers have complete DTG records and are at 500 and 860 m distance to the nearer stream. We built a Cumulative Deviation of the Mean, CDM, to identify rainfall changes that reflect DTG changes. The shift from negative to positive slope in the CDM curve marked the beginning of the wet conditions (Custodio & Llamas, 1996). Second, we defined uniformly and randomly distributed points in the wetland as representative wetland data locations, i.e., soft data, using the wetland and inundated areas data set (Lasso et al., 2014). For each, we defined a triangular PDF with the lower, middle, and upper bound parameters obtained from well 121, the only one in the Man River floodplains. The lower bound was set to 0 m (minimum depth expected), the middle to the minimum of non-zero (0.01 m), and the upper to the maximum (0.8 m) DTG during the wet season. Based on the qualitative descriptions made by the community about the low topographic locations close to the wetland, we expect DTG at those locations will follow a similar pattern to that observed in well 121 during the wet season. However, heterogeneity is expected to alter the parameters of the PDF function. To limit the assumption of equal behavior, we limited the PDF function to the wet months in the wetland area.

In this study, the BME framework was implemented using MATLAB functions from BMElib version 2.0c (available at https://mserre.sph.unc.edu/BMElib_web/). The BME implementation parameters chosen for this study varied according to the model approach. In the spatial approach, S, we used six hard data points in the estimation neighborhood, with a maximum 1,000 km spatial radius and a maximum temporal radius of zero weeks. In S and ST approaches, we considered 300 hard data points in the estimation neighborhood, with the same spatial radius as in the spatial approach and four weeks as the maximum temporal radius. For space-time with soft data approach (STSD), we used two soft data points as the maximum soft data estimation neighborhood. We used an estimation grid of 250 by 200 estimation points and included the hard data points and Voronoi points. The mean within the estimation neighborhood was assumed to be constant and equal to the mean of the observational data within that neighborhood (i.e., the average of the closest hard data points). The estimation grid was expanded outside the middle-low watershed domain to avoid edge effects.

We only considered the mean and covariance of DTG observations at the prior stage for all the modeling approaches. As a result, the prior stage is multivariate Gaussian. Although, higher statistical moments up to an even order can also be considered, though we did not attempt that. In the STSD approach, we considered both hard and soft data at the posterior stage, which leads to a non-Gaussian posterior pdf and a nonlinear combination of the observations. This BME posterior PDF is particularly adept at integrating the knowledge of the triangular PDFs at the wetland soft data points. The integration allows us to incorporate soft data derived from knowing the location of wetlands and from household knowledge of wetland flooding during the wet season. This unique feature of BME makes it an ideal framework to process data from a community science project.

2.4. Model Evaluation

To quantify each model performance, we used leave-one-out cross-validation over the entire space-time domain, using 4,397 validation values in total, consisting of 2,397 well observations and 2,600 wetland values obtained by taking the expected value of the probability density function at each soft data point. We quantified the total error (estimate minus validation value) in terms of systematic error (i.e., error that is consistent and can be removed through bias correction), and random error (i.e., residual error after systematic errors are removed), which indicate the degree of precision (Reyes et al., 2017). For the statistical metrics, we used the following notation. The first letter designates the statistical operator: M = mean, V = variance, S = Standard Deviation, MS = Mean Square, Cov = Covariance. The last letter represents the values to which each statistic is applied. Z = ln-depth estimations, O = ln-depth validation value observed at the well or wetland points, and E = Z-O = ln-depth errors. Let $e_i = z_i - o_i$ be the error at the space-time i, and n be the number of space-time validation values. Positive errors mean overestimation and negative errors mean underestimation.

The Mean Error, ME; (Equation 3) and the Variance of Error, VE, (Equation 4) quantify the systematic and random errors, respectively. The systematic errors explain the bias of the method, while the random error describes the precision. These two errors are quantified in the total error or Mean Square Error (Equation 5) (Reyes et al., 2017).

$$\mathrm{ME} = \sum_{i=1}^{n} e_i \;/\; n \tag{3}$$

$$\mathrm{VE} = \sum_{i=1}^{n} \left(e_i - ME\right)^2 \;/\; n \tag{4}$$

$$\mathrm{MSE} = \sum_{i=1}^{n} {e_i}^2 \;/\; n = ME^2 + VE \tag{5}$$

Finally, we used the Pearson coefficient of determination R^2 (Equation 6) to quantify the fraction of the variance explained by the estimates.

$$R = cov\left(O,Z\right) / \left(SO \times SZ\right) \tag{6}$$

where cov(O, Z) is the covariance between O and Z. Knowing that VE = V(O,Z), VE can be written as in Equation 7.

$$\mathrm{VE} = \mathrm{SO}^2 + \mathrm{SZ}^2 - 2 \times \mathrm{cov}\left(\mathrm{O,Z}\right) \tag{7}$$

Combining Equations 6 and 7, R^2 can also be expressed as:

$$\mathrm{R}^2 = \left(\mathrm{VZ} + \mathrm{VO} - \mathrm{VE}\right)^2 / \left(4 \times \mathrm{VZ} \times \mathrm{VO}\right) \tag{8}$$

which allows interpreting the influence of the random errors over the variance of the estimates.

2.5. Areas Impacted by Shallow Groundwater

Mapping DTG helps to reveal groundwater-surface water connectivity and quantify seasonal fluctuation in groundwater level and flooding (Fan et al., 2013). With the STSD model, we established the areas more susceptible to shallow groundwater in response to precipitation, given that 2008 was a wet year in which the middle-low watershed experienced flooding. To accomplish this goal, we used BME mean and variance and calculated the probability that DTG is less than a cutoff value of interest at each estimation point, that is, Prob[DTG < 0.1 m]. Values smaller than the cutoff value of 0.1 m are interpreted as shallow DTG which are more likely to maintain flow in surface streams, connect to the wetland, and increase the probability of flooding during extreme precipitation periods. We used the BME posterior Probabilistic Distribution Function (PDF) to calculate for each space-time estimation point the probability that DTG is less than 0.1 m, Prob[depth < 0.1]. The probabilities were classified as high = [0.8–1.0], moderate = [0.6–0.8), low = [0.4–0.6), and very low = [0.0–0.4).

With the probability maps, we accounted for areas of a high probability of shallow groundwater. We built a time series of the area and compared it with the cumulative average weekly precipitation. The precipitation records were obtained from the three closest weather stations managed by the National Institute of Hydrology, Meteorology and Environmental Studies of Colombia, IDEAM, (http://dhime.ideam.gov.co/atencionciudadano/). Two stations (i.e., Guarumo-La Lucha, National ID: 25025160 and La Coquera, National ID:

26240160) are located outside the middle-low watershed, around 3 km from the catchment to the Northeast. The third station is situated at the upper side of the Quebradona sub-catchment (i.e., Manizales, National ID: 26240060) (Figure 1). We used this information to analyze the influence of intensity, amount of precipitation per week, and frequency, days of continuous precipitation, in the shallow DTG area.

3. Results and Discussion

3.1. Monitoring Network Results

The monitoring period overlapped with the 2008 La Niña year, with 3,375 mm of precipitation (700 mm above average); 2009 was relatively dry and received 1,119 mm less rainfall than 2008 (2,255 mm) (data obtained from the IDEAM). The decrease in precipitation in 2009 compared to 2008 was reflected in the DTG. During the wet season, average DTG ranged between 0.9 and 1.3 m in 2009, whereas in 2008, it ranged between 0.5 and 1.1 m (Figure 3). During the 2009 dry season, average DTG was 2.4 m, while in 2008, DTG reached 3.3 m in March (Figure 3b). DTG ranged from 0 to 20 m; for 75% of the data points DTG was less than 6 m (Figure 3b). Wells with DTG greater than 6 m were located to the North (wells 120 and 123), and to the South, outside the middle-low watershed (well 140).

The closest monitoring stations to Ciénaga Colombia wetland were piezometer 202 and well 105 (Figures 1 and 3). During the 2008 wet season, DTG in piezometer 202 reached the surface in June, while for well 105 the minimum DTG was 1 m. These differences are explained by the fact that well 105 is topographically higher (55 m.a.s.l) than piezometer 202 (47 m.a.s.l). Well 121 was the only one located in the Man River floodplain (Figure 3). In this well, groundwater reached the surface during the two wet seasons, fluctuating from 0 to 0.8 m, and had a maximum DTG (3.29 m) during the dry period of 2008 (Figure 3).

DTG varied with the season and well's location (Figures 1 and 3). For wells located at a high topographic location, at the catchment boundary or between two tributaries (e.g., wells 120, 127, 128, and 140), the water table was deeper during the entire period compared to wells located in topographic depressions. Consistently shallower depths were found close to the springs of the Ciénaga Colombia creek sub-catchment, especially during the wet season, at Quebradona Creek (well 126), and at La Manada Creek (well 129). Difficulty accessing wells located close to the wetland impeded measurement of DTG in these areas.

3.2. Cross-Validation and Method Selection

The cross-validation MSE metric, calculated by combining well and wetland validation data, showed a reduction in the overall error (MSE) when using space-time models compared to spatial interpolation (Table 1). The MSE dropped from 9.41 (ln-depth)^2 in the S model to 7.21 (ln-depth)^2 in the ST model. This reduction is explained by an improvement in precision (VE), but not in bias (ME), as depths to groundwater are consistently overestimated due to lack of hard data collected near the wetland. Adding soft data generated by knowledge of low DTG at the wetland locations and wet weeks (i.e., weeks in which the water table rose to the ground in the vicinity of wetland areas) resulted in further reduction of the MSE from 7.21 (ln-depth)^2 in ST to 0.61 (ln-depth)^2 in STSD. This change resulted from the addition of soft data, which caused a decrease in VE, from 3.81 to 0.41 (ln-depth)^2, and ME, from 1.84 to 0.44 (ln-depth). Therefore, the STSD model reduced the overestimation of DTG, and the total error by a factor of 15 (MSE_S/MSE_{STSD}) when S is compared to STSD. Consistent DTG overestimation could be due to the lack of well observations in the wetland. Although further reductions could be reached with more well observations in the wetland area, our wetland space-time data contribute to a better representation of the DTG in the floodplain. Cross-validation results using only the well and wetland validation data, data agree with these results and are included in the supporting information (Table S1).

The Pearson regression coefficient, R^2, increased from 0.004 for the S model to 0.268 for the ST model. Since VO is the same for S and ST, an increase in R^2 may be due to a decrease in VE or an increase in VZ. Here, the increase in R^2 is due to an effect on both VE and VZ. VE decreased from 5.97 (ln-depth)^2 for the S model to 3.81 (ln-depth)^2 in the ST model (i.e., the ST model produces smaller error estimates), while VZ simultaneously increased from 1.23 (ln-depth)^2 in S to 2.23 (ln-depth)^2 (i.e., the ST model provides a stronger contrast between small vs. large DTG values). These results reveal the benefits of building a space-time interpolation

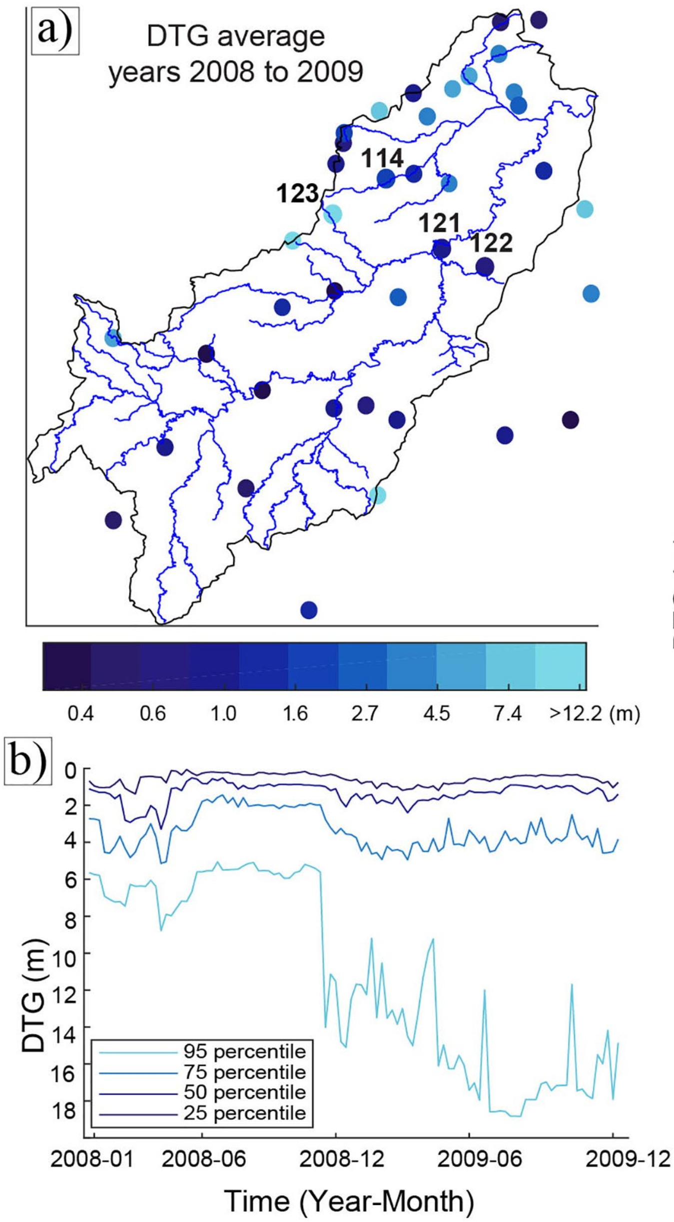

Figure 3. Summary of data collected in the watershed. (a) annual average distribution of depth to groundwater (DTG) during 2008 and 2009. (b) weekly DTG percentiles in all wells. (c) examples of DTG time series chosen based on depth range.

Table 1
Cross-Validation Model Results for the Hard and Soft Data Points

Model	MSE (ln-depth)^2	VE (ln-depth)^2	ME (ln-depth)	R^2 unitless	VZ (ln-depth)
S	9.41	5.97	1.86	0.004	1.23
ST	7.21	3.81	1.84	0.268	2.23
STSD	0.61	0.41	0.44	0.930	3.71

Note. For each cross-validation evaluation, MO is −1.93 (ln-depth), and VO is 5.06 (ln-depth)^2.

over a pure spatial analysis. For the ST model, high temporal correlation of DTG informed the model in areas where DTG were missing for a particular date. R^2 increased further, to 0.930, in the STSD model due to an increase in VZ from 2.23 (ln-depth)^2 in ST to 3.71 (ln-depth)^2 in STSD, and a decrease in VE from 3.81 (ln-depth)^2 to 0.41 (ln-depth)^2. The dual-action on VE and VZ reflects the positive impact of adding soft data describing the expected low DTG in the wetland and floodplains during wet periods.

Inclusion of wetland space-time data improved DTG estimates (Figures S3–S4 in the supporting information). The models showed shallow DTG in the northwest of the middle-low watershed, close to springs and depressions, especially in the wet season. However, models without soft data (S and ST) failed to represent the expected small depths near wetland and floodplains (DTG estimation resulted in larger values) due to the lack of observations in the topographic low-elevation sectors of the catchment.

The collaborators reported increases in flooded areas and wetland open water during the wet season. Our results can spatially represent these findings and observations at a weekly temporal scale. Moreover, our findings are consistent with previous studies in the region that used hydrogeochemistry and isotopic techniques to identify groundwater recharge sources (Palacio & Betancur, 2007), and surface-groundwater interactions in the Man River middle-low watershed (Palacio et al., 2013) and the Ciénaga Colombia wetland (Santa-Arango et al., 2010). According to these studies, in the Quebradona and Ciénaga Colombia sub-catchment, evaporation from the aquifer occurs in the upper left side of each of the sub-catchments (Palacio et al., 2013). Moreover, direct recharge takes place across the watershed. This pattern was identified by detecting similar isotopic composition in the rainwater and the shallow groundwater, which suggested recent groundwater in the aquifer with potential short resident time (Palacio & Betancur, 2007). These water composition similarities were consistent during the wet and dry seasons, implying a high dependency of the aquifer on precipitation, which also may explain some of the shallow depths obtained during the wet year.

BME total error reduction is consistent with BME applications in water quality (Akita et al., 2007; Messier et al., 2014), highlighting that accounting for temporal correlation results in significant decrease in the total error. Our approach allowed us to partition the total error into systematic and random error to analyze them separately. We found that the main contribution in the STSD model is that it increases the precision (i.e., reduce the random errors) more than it reduces the bias, although the bias is also reduced (i.e., systematic error decrease). All methods overestimated DTG. This may be explained by the lack of hard DTG data in the low-elevation locations. Nevertheless, STSD overestimated DTG the least compared with the S and ST model. This consistent overestimation implies more contrast between deep and shallow DTG is expected in the area and predictions can be improved by adding new hard data measurements at the depression locations.

3.3. Areas Impacted by Shallow Groundwater

The probability of shallow groundwater (DTG < 0.1 m) increased as precipitation increased. The area of high probability of shallow DTG expanded during April–November, and decreased between December and March (Figure 4a). After the middle-low watershed received six or more continuous days of precipitation in a week, the area with shallow DTG increased, reaching its maximum (56%) the second week of May 2008 (Figure 4a). We saw a bigger increase in the area in 2008, the wetter year, suggesting a relationship between days with continuous precipitation, rainfall intensity, and shallow DTG (Figure 5). Conversely, shallow DTG area decreased two weeks after reaching its peak when precipitation also decreased, suggesting that soil moisture's antecedent and posterior conditions influence shallow DTG area changes. Our results suggest a nonlinear relationship may be occurring between precipitation intensity and frequency, and the fluctuation of shallow DTG areas (Figure 5). We found a relationship between the aquifer response to precipitation and the intensity and duration of the rain. Longer DTG records are necessary to define the threshold in precipitation after which groundwater contributes to overland flow and subsequent flooding, and the nature of the nonlinear relationship defining this threshold.

The distribution of the high probability of shallow DTG areas indicates the aquifer's potential exchange of groundwater and surface water. Zones with consistently shallow DTG included the Ciénaga Colombia wetland and the Man River floodplains in the low watershed, i.e., topographic depressions, and the up-stream La Manada and Quebradona sub-catchments, where the headwaters are located (Figures 6 and 7). Shallow DTG areas up-stream of the main sub-catchments may explain the permanent streamflow of the Man tributaries throughout the year. Low probabilities of shallow DTG, situated to the North in the watershed's upper margin (Figure 4b), are consistent throughout the two-year period (Figures 6 and 7).

Previous studies suggested a dominant contribution of the aquifer to the surface water, and evaporation from the upper La Manada and Quebradona sub-catchments (Betancur, 2008; Palacio et al., 2013). Evaporation may explain the rapid decrease in the shallow DTG area after it reaches its peak (Figure 4a). Another explanation of this rapid decline of shallow DTG area could be associated with the soil characteristics and the sparse vegetation. Eighty percent of the watershed is associated with sandy clay and loam soils derived

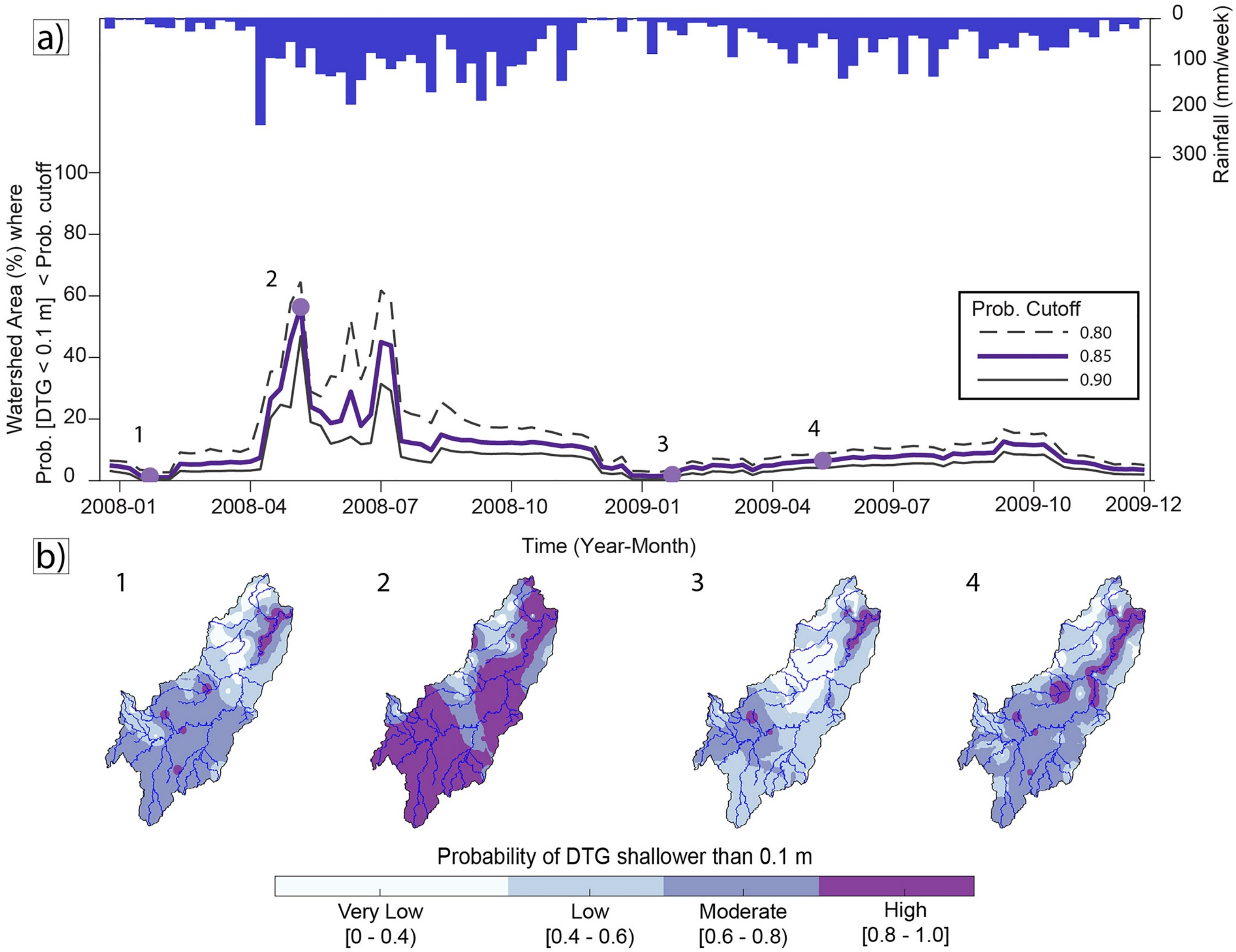

Figure 4. Temporal and spatial distribution of Prob[DTG < 0.1 m]. (a) Weekly precipitation compared to weekly area of high Prob[DTG < 0.1 m] expected at 0.8, 0.85, and 0.9 cutoffs. Values at each probability cutoff are expressed as a percentage of the total watershed area (481.3 km^2), including wetlands (11.1 km^2). Precipitation is the average of the three closest national weather stations. (b) Example weekly maps of the spatial distribution of Prob [DTG < 0.1 m] for time points 1, 2, 3, and 4 (see a), illustrating dry and wet months.

from poorly consolidate sandstones with a hydraulic conductivity of about 0.2 cm/h (Instituto Geográfico Agustín Codazzi, 2007a). These soil characteristics allow the soils to reach saturation quickly. Therefore, DTG increases during continuous and intense rainy days.

Land use activities may play an essential role in shaping the aquifer response to precipitation in the watershed. The region has alluvial mining exploitation and grazing since the 16th century (Cuartas et al., 2000). Mining activities in the Man River floodplains and other tributaries have affected the watershed's capacity to control erosion in the riverbanks. Additionally, around 80% of the middle-low watershed is used for grazing, in which a common practice is to create artificial open water ponds to provide water for the animals (Instituto Geográfico Agustín Codazzi, 2007a). These activities have reshaped creeks and channels, disrupting the surface flow (Betancur, 2008, 2014) and threatening wetland ecosystem services such as flow control, groundwater replenishment, erosion control, and food provision (Betancur-Vargas et al., 2017). This practice and the favorable atmospheric moisture conditions for evaporation may be drivers for rapid water loss after precipitation days cease. Evidence of an increase in intensity and frequency of La Niña events (Wang et al., 2019) is cause for concern, as an intensification of wet events would increase the

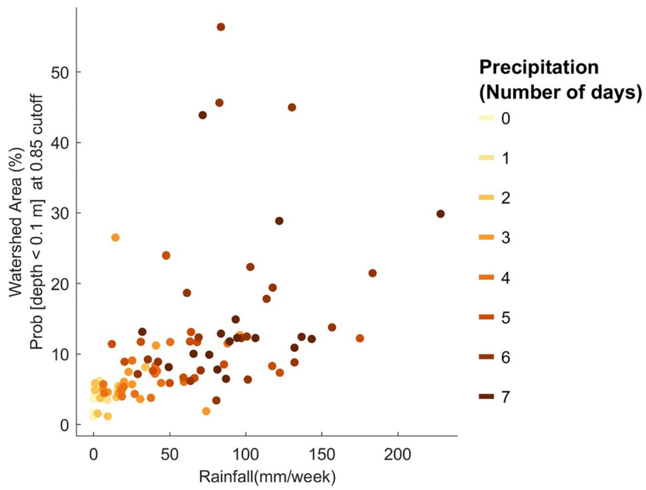

Figure 5. Relation between precipitation of the previous week and the high probability of shallow depth to groundwater (DTG) area. The combined precipitation intensity (i.e., mm per week of rainfall) and duration (i.e., number of rainy days) have an effect on the area of shallow DTG.

probability of flooding in agricultural areas, promoting sediment deposition and erosion that may cause soil compaction and reduce the infiltration capacity.

3.4. Impact of Community Knowledge on Model Implementation

Our approach allowed us to identify the combined action of intensity and frequency of precipitation over DTG response. This identification was possible due to the community knowledge that allowed to add detailed temporal data to the model. Previous studies used the classic kriging spatial interpolation to model monthly groundwater table in the River Man middle-low watershed using SRTM-30 (Palacio, 2014). While SRTM-30 was the best-known elevation model for the area, heterogeneity is missing in a relatively flat landscape (i.e., the elevation of a well located in high topography can appear lower than expected). Although we cannot compare both studies' results numerically, both studies revealed the groundwater table fluctuation in response to precipitation changes. In addition, our results provide insights about how DTG may respond to precipitation.

The fact that STSD is the best representation of DTG suggests that community knowledge added vital information to improve DTG mapping, even though incorporating this knowledge into the model implementation was not planned from the beginning of the community science project. A more systematic qualitative knowledge collection with the community (e.g., interviews or surveys, storytelling, knowledge dialog) would enhance community perceptions into the model. These systematic qualitative methods require design strategies for sustaining community engagement (Haklay, 2013). For future projects we suggest, regardless of the knowledge collection strategy, to include the community as part of the model validation. Further feedback from the community will help confirm and validate model results collectively. A technique that can be adapted for this model validation purpose is social cartography (Liebman & Paulston, 1994). This technique has proven effective in identifying relevant monitoring sites while also helps to identify environmental risk (Arias et al., 2016).

Participatory approaches are effective mechanisms to increase the knowledge about the groundwater system while involving the community in the process at different levels (Grieef & Hayashi, 2007; K. E. Little et al., 2016; Re, 2015). Our approach can be used in systems with no data for exploratory purposes (e.g., monitoring network design or seasonal groundwater-surface water connectivity detection, and aquifer characterization). Additionally, our approach is helpful for incorporating continuous groundwater monitoring because locals can be involved in different project stages, from data collection to data analysis to model validation. Building collaborative links between scientists and the community also help address research questions with a meaningful social impact (Arias et al., 2016; Haklay, 2013).

Constant communication and knowledge sharing are effective engagement mechanisms in community science projects (Assumpção et al., 2018; Cooper et al., 2007). One of the tasks that made the communication effective was to have short informal conversations with each household before discussing the data collection. Communication and trust-building are key in community science projects (Baldwin et al., 2012; K. E. Little et al., 2016). Fieldwork campaigns were also designed to build trust and establish collaborations with the households. Providing recommendations for the wells' maintenance, sharing water quality analysis results, and having a conversation with the households about activities not necessarily related to data collection were effective mechanisms to build trust, and at the same time, help to solve issues affecting water quality at the household level. These ways of communication helped to maintain the monitoring network remotely and to learn from the locals. We believe that our approach allows us to include the community knowledge in the data analysis.

10.1029/2020WR029519

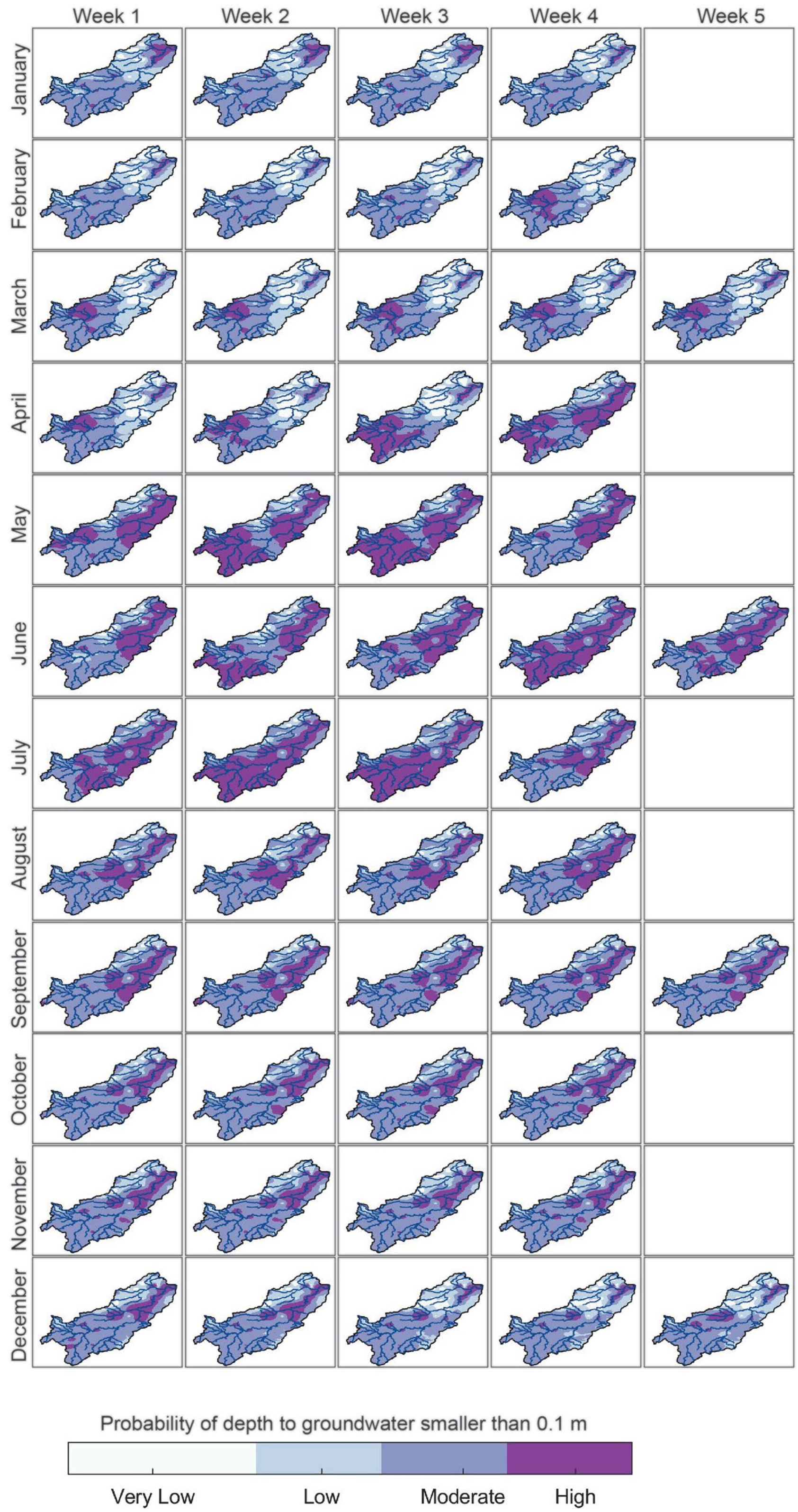

Figure 6. Spatial distribution of the probability that the weekly average of depth to groundwater (DTG) is smaller than 0.1 m (Prob. [DTG < 0.1 m]) for each week in 2008. The first week of the year is considered to start on 12/31/2007.

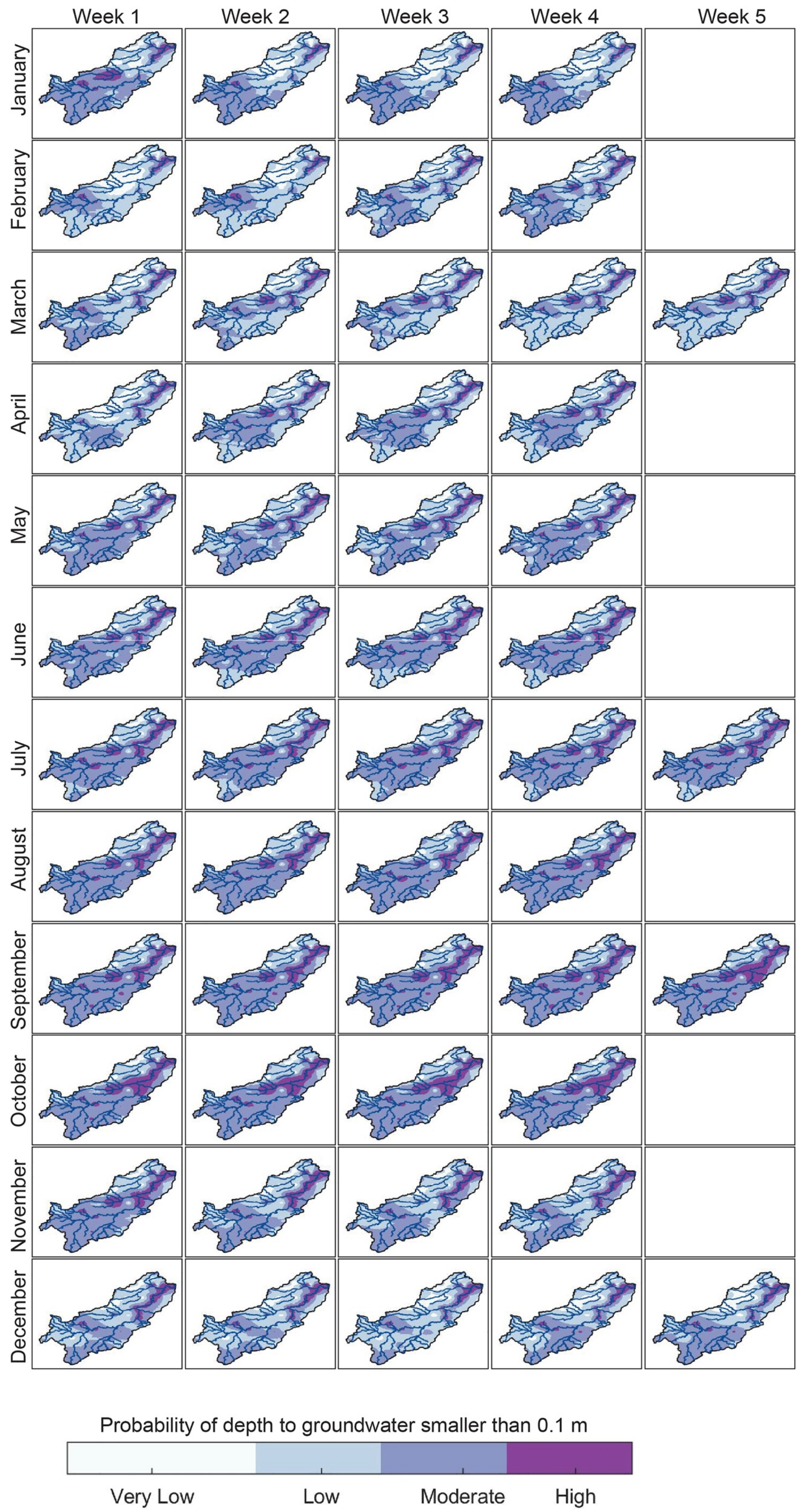

Figure 7. Spatial distribution of the probability that the weekly average of depth to groundwater (DTG) is smaller than 0.1 m (Prob. [DTG < 0.1 m]) for each week in 2009. The first week of the year is considered to start on 01/05/2009.

4. Conclusions

Community participation in scientific hydrologic studies provides benefits through both community engagement (e.g., knowledge transfer, community self-training, education outreach, and informed decision making) and the understanding of hydrologic systems. In groundwater-dependent regions with limited or no DTG information, our results suggest that the combination of Community Science Research and BME modeling can contribute to a better understanding of groundwater dynamics. We used community descriptions, locations of wetlands, and precipitation records to define a probabilistic function that informs us about shallow DTG in the floodplains and wetland areas and complements the data collected by the community. The community helped create a data set that would have been challenging to acquire with conventional data collection methods and provided local descriptive knowledge essential to improving DTG representation. BME's capabilities make it suitable to incorporate both qualitative and quantitative data into a model. Our results show how this combination contributed to the modeling effort, specifically:

1. The reduction of the total error occurred progressively from space (S), to the space-time (ST), to the space-time with soft data model (STSD). The high temporal correlation characteristic of DTG allowed improved space-time (ST) interpolations compared to spatial interpolation (S). MSE further reduced from ST to STSD by adding soft probabilistic DTG in the wetland and floodplain areas, resulting in the combined improvement in precision and a reduction in DTG overestimation and bias. Also, STSD incorporated information where DTG values were missing, increasing the variability in DTG values.
2. The spatial and weekly temporal distribution of the groundwater and surface water connections showed consistency with previous studies using hydrogeochemistry and isotopic approaches. In those studies, evaporation from the aquifer was identified to occur in areas of the main sub-catchments (Palacio & Betancur, 2007), which may explain why the area of shallow DTG decreased around two weeks after reaching its peak. Similar isotopic composition between groundwater and surface water during the wet and dry periods in topographic depressions (Palacio et al., 2013; Santa-Arango et al., 2010) explains the consistent shallow DTG in our maps. By incorporating the wetland locations and their most likely DTG during the wet season, we delineated shallow DTG in these topographic depressions.
3. Our results suggest a nonlinear relationship between precipitation intensity and frequency and the shallow groundwater area. The rapid increase in shallow DTG area in the extreme wet year compared to the average year may be related to high antecedent soil moisture conditions due to continuous rainfall, raising the groundwater table. In contrast, middle-low watershed topographic features, atmospheric conditions, and land-use practices create favorable conditions for high evaporation, contributing to decreasing the shallow DTG area over the year when precipitation ceases. Additional data will be required to investigate further specific threshold values after which shallow DTG area increases.

From a modeling perspective, one limitation of our approach is that the lack of observations in the wetland areas constrains the reduction of DTG overestimation. That said, recognizing the value of using BME in Community Science projects may contribute to the design of monitoring networks or to model use for exploration purposes. The combined use of BME and Community Science may allow for closer interactions with the collaborators, contributing to the formulation of new hypotheses and further identification of critical environmental concerns. Our results suggest that both quantitative and, crucially, qualitative information from community members can result in substantially better spatial and temporal understanding of DTG, which may be of use in many similar environments around the world. Including the community in the model validation would be a further step towards fully integrating community science projects in the broader scientific enterprise.

Data Availability Statement

All the Depth to Groundwater data used in this study are available in the repository http://doi.org/10.5281/zenodo.3923896 (License Creative Commons Attribution 4.0 International).Wetlands and inundated areas data set are available in Lasso et al. (2014). All maps of the STSD model results used in this study are available in the repository http://doi.org/10.5281/zenodo.3928587 (License Creative Commons Attribution 4.0 International).

Acknowledgments

The authors acknowledge the Grupo de Ingeniería y Gestión Ambiental, GIGA, at the Facultad de Ingeniería, Universidad de Antioquia, and the International Atomic Energy Agency (project: CRP-14031), which funded the community science research component of this study. Data Analysis was completed with support from Minciencias-Fulbright, Colombia, and Faculty for the Future fellowships awarded to Angélica M. Gómez. Special thanks to Teresita Betancur Vargas at Universidad de Antioquia for her essential contributions to the design and maintenance of the monitoring network as a director of the Community Science Research project, and her mentorship of the first author's M.S.thesis. Thanks to all the collaborators of the Man River middle-low watershed (Complete list of names in the supplemental material), and to Paola A. Palacio, Diana Montoya-Velilla, and Diana Santa-Arango. Thanks to Jhon C. Duque and Arik Tashie for their time to discuss data analysis results and to Giovanny Ruiz for his contributions to the design of Figure 2. The authors would also like to thank the editors and the anonymous reviewers for their valuable comments and feedback.

References

Akita, Y., Carter, G., & Serre, M. L. (2007). Spatiotemporal Nonattainment Assessment of Surface Water Tetrachloroethylene in New Jersey. *Journal of Environmental Quality*, *36*(2), 508–520. https://doi.org/10.2134/jeq2005.0426

Arias, P. A., Villegas, J. C., Machado, J., Serna, A. M., Vidal, L. M., Vieira, C., et al. (2016). Reducing social vulnerability to environmental change: Building trust through social collaboration on environmental monitoring. *Weather, Climate, and Society*, *8*(1), 57–66. https://doi.org/10.1175/WCAS-D-15-0049.1

Assumpção, T. H., Popescu, I., Jonoski, A., & Solomatine, D. P. (2018). Citizen observations contributing to flood modelling: Opportunities and challenges. *Hydrology and Earth System Sciences*, *22*(2), 1473–1489. https://doi.org/10.5194/hess-22-1473-2018

Baldwin, C., Tan, P.-L., White, I., Hoverman, S., & Burry, K. (2012). How scientific knowledge informs community understanding of groundwater. *Journal of Hydrology*, *474*, 74–83. https://doi.org/10.1016/j.jhydrol.2012.06.006

Betancur, T. (2008). *Una aproximación al conocimiento de un sistema acuífero tropical caso de estudio: Bajo cauca antioqueño*. (Doctoral Dissertation) (p. 221). Universidad de Antioquia, Medellín, Colombia.

Betancur, T. (2014). *Aguas subterraneas en el bajo cauca antioqueño*. Medellin: Universidad de Antioquia. ISBN 978-958-8848-84-6.

Betancur, T., Palacio, C., & Escobar, J. (2012). Conceptual models in hydrogeology, methodology and results. In *Hydrogeology. A global perspective* (pp. 203–222). https://doi.org/10.5772/28155

Betancur-Vargas, T., García-Giraldo, D. A., Vélez-Duque, A. J., Gómez, A. M., Flórez-Ayala, C., Patiño, J., & Ortiz-Tamayo, J. Á. (2017). Aguas subterráneas, humedales y servicios ecosistémicos en Colombia. *Biota Colombiana*, *18*(1), 1–28. https://doi.org/10.21068/c2017.v18n01a1

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., et al. (2014). Citizen science in hydrology and water resources: Opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*, *2*(October), 1–21. https://doi.org/10.3389/feart.2014.00026

Christakos, G. (1990). A Bayesian/maximum-entropy view to the spatial estimation problem. *Mathematical Geology*, *22*(7), 763–777. https://doi.org/10.1007/BF00890661

Christakos, G., Bogaert, P., & Serre, M. (2001). *Temporal GIS: Advanced functions for field-based applications*, Berlin, New York: Springer. https://doi.org/10.1007/978-3-642-56540-3

Cooper, C. B., Dickinson, J., Phillips, T., & Bonney, R. (2007). Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society*, *12*(2), 11. https://doi.org/10.5751/es-02197-120211

CORANTIOQUIA, & Universidad de Antioquia. (2003). *Evaluación hidrogeológica entre los municipios de Caucasia y Cáceres*. Medellin: Universidad de Antioquia.

Cuartas, M. E., Giraldo, L., Maya, M., & Bejarano, O. (2000). Bajo Cauca Desarrollo regional: Una tarea común universidad-región. In *Insitituto de Estudios regionals*, (pp. 1–120). Medellin, Colombia: Universidad de Antioquia. Retrieved from http://www.udea.edu.co/wps/wcm/connect/udea/c3957be6-8230-4d1f-9e41-845154328ce5/caracterizacion-bajo-cauca.pdf?MOD=AJPERES

Custodio, E., & Llamas, R. (1996). *Hidrología subterránea. Tomo I*. (second). Spain: Ediciones Omega.

Departamento Administrativo Nacional de Estadística, DANE (2012). *Reporte de indice de pobreza multidimensional–2011*. Retrieved from https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-y-desigualdad/pobreza-y-desigualdad-2011

Fan, Y., Li, H., & Miguez-Macho. (2013). Global patterns of groundwater table depth. *Sicence*, *13626*, 940–944. https://doi.org/10.1126/science.1229881

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., & Cardenas, M. B. (2016). The global volume and distribution of modern groundwater. *Nature Geoscience*, *9*(2), 161–167. https://doi.org/10.1038/ngeo2590

Grieef, L. A., & Hayashi, M. (2007). Establishing a rural groundwater monitoring network using existing wells: West nose creek pilot study, Alberta. *Canadian Water Resources Journal*, *32*(4), 303–314. https://doi.org/10.4296/cwrj3204303

Haklay, M. (2013). Citizen science and volunteered geographic information – Overview and typology of participation. In Sui, D., Elwood, S., & Goodchild, M. (Eds.), *Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice* (pp. 105–122). Dordrecht: Springer Berlin Heidelberg. https://doi.org/10.1007/978-94-007-4587-2

He, J., & Kolovos, A. (2018). Bayesian maximum entropy approach and its applications: A review. *Stochastic Environmental Research and Risk Assessment*, *32*(4), 859–877. https://doi.org/10.1007/s00477-017-1419-7

Hosseini, M., & Kerachian, R. (2017). A Bayesian maximum entropy-based methodology for optimal spatiotemporal design of groundwater monitoring networks. *Environmental Monitoring and Assessment*, *189*(9). https://doi.org/10.1007/s10661-017-6129-6

Instituto Geográfico Agustín Codazzi, & Gobernación de Antioquia. (2007a). *Estudio semidetallado de suelos de las áreaspotencialmente agricolas: Bajo Cauca departamento de Antioquia*. Medellín IGAC: Gobernación de Antioquia. ISBN 958-832-310-X.

Instituto Geográfico Agustín Codazzi, & IGAC. (2007b). *Levantamiento semidetallado de las coberturas terrestres: Departamento de Antioquia*: Bogotá, Colombia. IGAC.

Lasso, C., Guiérrez, F., & Morales-B, D. (2014). Humedales interiores de Colombia: identificación, caracterización y establecimiento de límites según criterios biológicos y ecológicos. In *X. Serie Editorial Recursos Hidrobiológicos y Pesqueros Continentales de Colombia*–255). Bogotá D.C, Colombia: Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (IAvH). Retrieved from http://hdl.handle.net/20.500.11761/9280

Le Coz, J., Patalano, A., Collins, D., Guillén, N. F., García, C. M., Smart, G. M., et al. (2016). Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand. *Journal of Hydrology*, *541*, 766–777. https://doi.org/10.1016/j.jhydrol.2016.07.036

Lewandowski, J., Arnon, S., Banks, E., Batelaan, O., Betterle, A., Broecker, T., et al. (2019). Is the hyporheic zone relevant beyond the scientific community? *Water*, *11*(11), 2230. https://doi.org/10.3390/w11112230

Liebman, M., & Paulston, R. (1994). Social cartography: A new methodology for comparative studies. *Compare: A Journal of Comparative and International Education*, *24*(3), 233–245. https://doi.org/10.1080/0305792940240304

Little, K. E., Hayashi, M., & Liang, S. (2016). Community-based groundwater monitoring network using a citizen-science approach. *Groundwater*, *54*(3), 317–324. https://doi.org/10.1111/gwat.12336

Little, S., Pavelsky, T. M., Hossain, F., Ghafoor, S., Parkins, G. M., Yelton, S. K., et al. (2021). Monitoring variations in lake water storage with satellite imagery and citizen science. *Water*, *13*(7), 949. https://doi.org/10.3390/w13070949

Lowry, C. S., Fienen, M. N., Hall, D. M., & Stepenuck, K. F. (2019). Growing pains of crowdsourced stream stage monitoring using mobile phones: The development of crowdHydrology. *Frontiers in Earth Science*, *7*. https://doi.org/10.3389/feart.2019.00128

Marchetti, Z. Y., & Carrillo-Rivera, J. J. (2014). Tracing Groundwater Discharge in the floodplain of the Parana River, Argentina: Implications for its biological communities. *River Research and Applications*, *30*(2), 166–179. https://doi.org/10.1002/rra.2629

Margat, J., & van der Gun, J. (2013). *Groundwater around the world. A greographic synopsis*–376). Boca Raton, FL: CRC Press. ISBN 9780367576509.

Messier, K. P., Kane, E., Bolich, R., & Serre, M. L. (2014). Nitrate variability in groundwater of North Carolina using monitoring and private well data models. *Environmental Science and Technology*, *48*(18), 10804–10812. https://doi.org/10.1021/es502725f

Njue, N., Stenfert Kroese, J., Gräf, J., Jacobs, S. R., Weeser, B., Breuer, L., & Rufino, M. C. (2019). Citizen science in hydrological monitoring and ecosystem services management: State of the art and future prospects. *Science of the Total Environment*, *693*, 133531. https://doi.org/10.1016/j.scitotenv.2019.07.337

Organization World Health (1999). *Guidelines for drinking-water quality: Volume 3: Drinking-water quality control in small-community supplies*. Geneva: World Health Organization. Retrieved from https://apps.who.int/iris/handle/10665/252074

Palacio, P. (2014). *Modelo Hidrológico Conceptual para la Cuenca del Río Man a partir de Técnicas Hidrológicas, Hidrogeoquímicas e Isotópicas*. (Doctoral Dissertation). (pp. 1–181). Medellín, Colombia: Universidad de Antioquia.

Palacio, P., & Betancur, T. (2007). Identificación de fuentes y zonas de recarga a partir de isotopos estables del agua (caso de estudio Bajo Cauca antioqueño). *Gestion y Ambiente*, *10*(4), 93.

Palacio, P., Dapena, C., & Betancur, T. (2013). Use of isotopic techniques for the assessment of hydrological interactions between ground and surface waters–Rio Man, Cienaga Colombia. In *Isotopes in hydrology, marine ecosystems and climate change studies*: Proceedings of the International Symposium held in Monaco, 27 March–1 April 2011, Vol. 2, Vienna: International Atomic Energy Agency (IAEA). https://www-pub.iaea.org/MTCD/Publications/PDF/SupplementaryMaterials/Pub1580_vol2_web.pdf

Re, V. (2015). Incorporating the social dimension into hydrogeochemical investigations for rural development: The Bir Al-Nas approach for socio-hydrogeology. *Hydrogeology Journal*, *23*(7), 1293–1304. https://doi.org/10.1007/s10040-015-1284-8

Reyes, J. M., Xu, Y., Vizuete, W., & Serre, M. L. (2017). Regionalized $PM_{2.5}$ community multiscale air quality model performance evaluation across a continuous spatiotemporal domain. *Atmospheric Environment*, *148*, 258–265. https://doi.org/10.1016/j.atmosenv.2016.10.048

Sanders, A. P., Messier, K. P., Shehee, M., Rudo, K., Serre, M. L., & Fry, R. C. (2012). Arsenic in North Carolina: Public health implications. *Environment International*, *38*(1), 10–16. https://doi.org/10.1016/j.envint.2011.08.005

Sandoval, R. (2004). A participatory approach to integrated aquifer management: The case of Guanajuato State, Mexico. *Hydrogeology Journal*, *12*(1), 6–13. https://doi.org/10.1007/s10040-003-0311-3

Santa-Arango, D., Martinez, C., & Betancur-Vargas, T. (2010). Uso de hidroquímica e isótopos ambientales para la evaluación de la conexión hidrológica entre el agua subterránea y el humedal ciénaga colombia. (Un Primer Acercamiento). *Gestión y Ambiente*, *11*(2).

Serre, M. L., & Christakos, G. (1999). Modern geostatistics: Computational BME analysis in the light of uncertain physical knowledge - The Equus Beds study. *Stochastic Environmental Research and Risk Assessment*, *13*(1–2), 1–26. https://doi.org/10.1007/s004770050029

Serre, M. L., Christakos, G., Li, H., & Miller, C. T. (2003). A BME solution of the inverse problem for saturated groundwater flow. *Stochastic Environmental Research and Risk Assessment*, *17*(6), 354–369. https://doi.org/10.1007/s00477-003-0156-2

Snyder, D. (2008). *Estimated depth to ground water and configuration of the water table in the Portland, Oregon area scientific investigations report 2008–5059*. Retrieved from http://pubs.usgs.gov/sir/2008/5059/

Sophocleous, M. (2002). Interactions between groundwater and surface water: The state of the science. *Hydrogeology Journal*, *10*(1), 52–67. https://doi.org/10.1007/s10040-001-0170-8

Starkey, E., Parkin, G., Birkinshaw, S., Large, A., Quinn, P., & Gibson, C. (2017). Demonstrating the value of community-based ('citizen science') observations for catchment modelling and characterisation. *Journal of Hydrology*, *548*, 801–817. https://doi.org/10.1016/j.jhydrol.2017.03.019

UNESCO, & UN-Water. (2017). Wastewater: The Untapped Resource. In UN World Water Development Report 2017. Paris, UNESCO: United Nations Educational, Scientific and Cultural Organization. http://www.unesco.org/new/en/natural-sciences/environment/water/wwap/wwdr/2017-wastewater-the-untapped-resource/

Wang, B., Luo, X., Yang, Y.-M., Sun, W., Cane, M. A., Cai, W., et al. (2019). Historical change of El Niño properties sheds light on future changes of extreme El Niño. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(45), 22512–22517. https://doi.org/10.1073/pnas.1911130116

Weeser, B., Stenfert Kroese, J., Jacobs, S. R., Njue, N., Kemboi, Z., Ran, A., et al. (2018). Citizen science pioneers in Kenya – A crowdsourced approach for hydrological monitoring. *Science of the Total Environment*, *631–632*, 1590–1599. https://doi.org/10.1016/j.scitotenv.2018.03.130

Wiggins, A., & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. In *2011 44th Hawaii International conference on system sciences* (pp. 1–10). Kauai, HI, USA: IEEE. https://doi.org/10.1109/HICSS.2011.207

Wijnen, M., Augeard, B., Hiller, B., Ward, C., & Huntjens, P. (2012). *Managing the Invisible: Understanding and improving groundwater governance*. 1 Washington, DC. Retrieved from World Bank. http://hdl.handle.net/10986/17228

Winter, T. C., Harvey, J. W., Franke, O. L., & Alley, W. M. (1998). *Groundwater and surface water: A single resource*. In Circular (Vol. 1139). U.S. Geological Survey. https://doi.org/10.3133/cir1139