

Water Resources Research

RESEARCH ARTICLE

10.1029/2021WR031241

Exploring Predictive Uncertainty at a Double-Source Managed Aquifer Recharge Site via Stochastic Modeling



Key Points:

- Predictive modeling uncertainty of conservative transport is evaluated via stochastic models of a real double-source managed aquifer recharge (MAR) site
- Predictive uncertainty due to unknown future hydrometeorological conditions is larger than that caused by unknown subsurface structure
- If the larger source in a double-source MAR site is uncertain then mixing in the aquifer is uncertain

Correspondence to:

G. Rudnik,
germanr@volcani.agri.gov.il

Citation:

Rudnik, G., Rabinovich, A., Siebner, H., Katz, Y., & Kurtzman, D. (2022). Exploring predictive uncertainty at a double-source managed aquifer recharge site via stochastic modeling. *Water Resources Research*, 58, e2021WR031241. <https://doi.org/10.1029/2021WR031241>

Received 19 SEP 2021
Accepted 8 APR 2022

G. Rudnik^{1,2} , A. Rabinovich², H. Siebner³, Y. Katz⁴, and D. Kurtzman¹ 

¹Institute of Soil, Water and Environmental Sciences, Agriculture Research Organization—Volcani Institute, Rishon LeZion, Israel, ²Faculty of Engineering, School of Mechanical Engineering, Tel Aviv University, Tel Aviv, Israel, ³Department of Environmental Hydrology and Microbiology, Zuckerberg Institute for Water Research, Jacob Blaustein Institute for Desert Research, Ben Gurion University of the Negev, Be'er Sheva, Israel, ⁴Mekorot Water Company Ltd., Tel Aviv, Israel

Abstract In Menashe Streams managed aquifer recharge (MAR) site (Israel) desalinated seawater (DSW) is recharged since 2015, alongside ephemeral stream flows. This study quantifies the uncertainty of predictions of the mixing of these two water sources in the aquifer. Mixing estimations are based on the significant difference in the content of stable water isotopes between the water sources. Uncertainties driven from aquifer heterogeneity and climate variability are compared. We use realistic flow and isotope-transport models in a multiple-realization stochastic approach considering space and time for the two drivers of uncertainty. Predictive uncertainty is evaluated via comparison of the temporal coefficient of variation of four realization ensembles. Results show that the impact of subsurface structure uncertainty on the predictions is small, compared to the uncertainty resulting from variability related to future hydrometeorological conditions. A generalized conclusion from this result is the difficulty to make long-term predictions of the mixing ratios, regardless of the level of certainty in the subsurface structure interpretation, when one source of recharge water has significant annual fluctuations. Implication on prediction of magnesium concentration is demonstrated as an example for prediction uncertainty of concentration of a solute of interest in one of the MAR sources. Furthermore, we show that considering a single source MAR site, uncertainty in mixing of the MAR water and native groundwater in wells upgradient of the recharge location is higher than prediction uncertainty in wells located downgradient. Insights are also drawn regarding the change in uncertainty with distance from the recharge pond of the DSW.

1. Introduction

In the last decades, managed aquifer recharge (MAR) is a continuously growing practice, as part of the efforts to tackle water scarcity in various regions of the world. In these operations, water is recharged artificially, via designated ponds, canals, or wells, into the aquifer, which serves as a subsurface reservoir that has evident economic advantages over an above-surface alternative. The source of the recharged water is project-specific and depends on the purpose of the MAR operation, and can include, among others, river water, municipal wastewater, and rainfall runoff (Dillon et al., 2020). At times, water from an atypical source is recharged, such as excess volumes of desalinated seawater (DSW), the production of which has become abundant in the last decade, particularly in arid and semiarid regions (Jones et al., 2019).

Since the 1960s, Menashe Streams MAR site in Israel provides a cost-efficient subsurface storage solution for runoff water that otherwise would be lost to sea. In 2015, it became a double-source MAR site, since posttreated reverse-osmosis DSW is recharged, in addition to ephemeral stream flows.

The introduction of DSW as source water for MAR is not unique to the Menashe Streams MAR site ("Menashe"). In Gulf countries, where climate is arid and DSW serves as the main source of water supply, this practice is common—with field experiments in Kuwait, Qatar, and Abu-Dhabi taking place in 1960s, 1980s, and early 2000s, respectively (Abdulrazzak, 1997; Hutchinson, 1998; Stuyfzand et al., 2017). Parimalarenganayaki (2021) concludes that soil aquifer treatment (SAT) with DSW is the most feasible of eight investigated MAR technics applied in Gulf countries. Zekri et al. (2021), using optimization techniques, conclude that 84% of the annual surplus of produced DSW in the metropolitan area of Muscat, Oman can be reused via storage and recovery, instead of returned to the sea, thus providing net benefit value of around \$16 million per year. The Israeli Mediterranean case is conceptually different than that of most Gulf countries because DSW is injected to a naturally

© 2022. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

potable-water aquifer rather than a saline aquifer; hence, regional spreading and mixing of the DSW in the aquifer is desired (e.g., for remineralization).

Whether an MAR site is initially designed, or redesignated to operate as a double source, later, during operation years, there is a need to understand the affect it will have on the system in terms of mixing of the two types of water in the aquifer. In the case of Menashe, this need is emphasized by public health concerns caused by the lack of Mg^{2+} in recharged DSW (Avni et al., 2013; Rosen et al., 2018) and its potential remineralization due to mixing with naturally fresh mineral rich groundwater (Ganot et al., 2017; Ganot, Holtzman, Weisbrod, Russak, et al., 2018; Ronen-Eliraz et al., 2017).

Scenarios, such as this, should be thoroughly examined (e.g., via reactive transport models) to support decision making at the site. In particular, stochastic models should be developed to evaluate the uncertainty in predictions. In applications of remediation techniques, seawater intrusion, aquifer depletion, and others, the advantages of stochastic modeling are widely discussed (Carle & Fogg, 2020; Fiori et al., 2016; Gotovac et al., 2009; Rajaram, 2016; Sanchez-Vila & Fernandez-Garcia, 2016) and applied (He et al., 2013, 2014, 2015; Moreno & Paster, 2018; Pool et al., 2015; Teramoto et al., 2021), often as a means to account for the modeling uncertainty which can be caused by various factors, e.g., subsurface heterogeneity, geologic interpretation, and model parameterization. An extensive literature collection deals with deterministic modeling of processes in MAR sites, as reviewed by Ringleb et al. (2016). However, in contrast to the above mentioned subbranches of hydrogeology, only a handful of authors apply stochastic or probabilistic approaches and account for the modeling uncertainty at MAR sites (Maples et al., 2019, 2020; Rodriguez-Escales et al., 2017, 2018; Sasidharan et al., 2019; Vanderzalm et al., 2013), and even then, mainly synthetic or hypothetical MAR sites are used for the investigations.

Maliva (2015) suggests that it would be beneficial to apply stochastic methods when modeling MAR systems, in order to adequately represent spatial heterogeneity in aquifer properties. However, another major source of uncertainty is introduced by the variability of climate, which can affect operational decision making at an MAR site. In MAR systems, dramatic hydrodynamic conditions in the subsurface are common, where very high recharge rates during relatively short periods of time occur. For example, in the Menashe Streams MAR site, a column of >20 m of water has been recharged at a 10.7 ha infiltration pond within a period of a few weeks (Ganot et al., 2017), compared to $\sim 20 \text{ cm yr}^{-1}$ of natural recharge, in which only a fraction of the direct rainfall reaches the water table.

In this work, we perform numerical experiments to enhance our understanding of modeling uncertainty in an MAR site. The experiments are conducted using a new predictive model of the Menashe Streams MAR site, built in a stochastic Monte Carlo framework incorporating hydrologic and water-isotope field data, and insight from recorded runoff recharge data from over 50 years of MAR operation.

The goal of this work is to incorporate the two main sources of predictive uncertainty in such an MAR site: aquifer heterogeneity and the variability of climatic conditions, in a stochastic approach, and to assess and compare their individual influence on predictive uncertainty of flow and transport modeling in (a) various locations in the MAR site (e.g., with growing distance from the infiltration ponds) and (b) different scenarios of MAR operation (e.g., climate-dependent operation with runoff recharge versus stable operation with DSW only).

We address these issues of two-source mixing within the aquifer, keeping in mind that mixing of native groundwater with recharged DSW can represent a general case in which water from a new (additional) source is introduced into an existing single source MAR site. Similar two-source mixing cases are becoming more common due to the worldwide expansion of MAR and integrated water resources management (e.g., Zhang & Wang, 2021). To the best of our knowledge, the work presented here is the first to deal with issues of predictive uncertainty pertaining to a case study of a large-scale operating MAR site.

The paper is organized as follows. Section 2 contains a description of Menashe MAR site and of the hydrological and isotope data that supports the study, followed by a comprehensive description of the numerical experiment development process. Section 3 presents the obtained results which are discussed within the context of various topics of interest (e.g., varying distances, types of uncertainty, etc.), as well as a discussion regarding the applicability and the generalization of the presented approach. Finally, conclusions are drawn in Section 4.

2. Materials and Methods

2.1. Site Description

The Menashe Streams MAR site, located above the northern part of the unconfined, sandy Israeli (Mediterranean) coastal aquifer, was designed and constructed in the 1960s as part of an effort to maintain water security in a semiarid Mediterranean climate. Winter ephemeral streams from a 110 km² Menashe Hills catchment, where runoff is high and infiltration is low, are diverted via a 16-km long channel, into a settling pond and from there, depending on water level and turbidity, flow is channeled into three infiltration ponds (referred to in this work as ponds #1, #2, and #3; Figure 1) via a system of hydraulic control structures. The site is situated on a sandy dune landscape at a mean elevation of 30 m above sea level and is dominated by a Mediterranean climate, with average temperatures of 20.2 °C and a mean precipitation of 556 mm yr⁻¹. The aquifer, stretching over an area of 2,000 km² along the coast, has thickness varying between 200 m on the coastline (western boundary) and several meters at the eastern boundary. It is composed of Pleistocene calcareous sandstone, sand and interleaved with continuous marine, and continental silt and clay lenses, overlying a highly impermeable Saqiye clay beds of the Yafo Formation (Kurtzman et al., 2012; Figure 1c). The regional mean groundwater level fluctuates between 0 and 8 m above sea level due to intense winter recharge of runoff water (historic annual mean of 10 million cubic meters (Mm³)) and summer production fluxes driven by dozens of designated wells (Figure 1). Since 2015, annual mean of ~2.1 Mm³ reverse-osmosis DSW from the Hadera Desalination Plant (total annual production ~130 Mm³ yr⁻¹) constitutes an additional source of water for the MAR site, usually, when maintenance is conducted at the National Water Carrier (days-weeks/year). Therefore, a certain annual volume is bypassed to pond #3, due to its geographical proximity to the DSW pipeline (Figure 1), which de-facto turns Menashe MAR site into a double-source MAR site.

2.2. Data

2.2.1. Sources and Sinks

Annual recharge data from the two sources (DSW and rainfall runoff, Table 1) was provided by Mekorot (the national water company). Monthly production data collected at 48 public and private wells were provided by the Israeli Water Authority and Mekorot (Table 1; per well production rates are not presented herein due to confidentiality), along with head data from several selected wells (Section 2.3.4). Rainfall data from a rain gauge located in proximity to the site (at Gan Shmuel) was acquired from the Israeli Meteorological Service website. Table 1 includes 3 years of published data (Ganot, Holtzman, Weisbrod, Bernstein, et al., 2018) and 2 years of newly acquired data (2018–2019).

2.2.2. Water Isotopes

In a biannual sampling campaign (2018–2019), water samples from five active production wells located in the vicinity of the infiltration ponds (M7, M21, M2, M9, and M6; Figure 1) were collected in sealed polyethylene tubes and kept on ice until reaching the lab. Care was taken to avoid headspace and prevent evaporation. Samples were analyzed for stable water isotopes by a cavity ring-down spectroscopy (CRDS) analyzer (L2130-i, Picarro) in the Zuckerberg Institute for Water Research (ZIWR) at Ben Gurion University and results are presented in Table 2. The isotopic values are expressed in δ²H (‰), and calculated relative to a standard as follows:

$$\delta_{\text{sample}}(\text{in } \text{‰}) = \left(\frac{R_{\text{sample}}}{R_{\text{standard}}} - 1 \right) * 1,000 \quad (1)$$

where $R = {}^2\text{H}/{}^1\text{H}$ is the ratio between heavy and light isotopes. Calibration was conducted using the international VSMOW-SLAP scale with an accuracy of ±0.5‰. The desalinated water carries the relatively enriched (heavy) signature of the east Mediterranean seawater and remains high during the reverse-osmosis desalination process. Therefore, δ²H can be used to trace the spreading and mixing of the DSW plume within the aquifer, following a binary system assumption, which was proven to be suitable for the Menashe MAR site by Ganot, Holtzman, Weisbrod, Bernstein et al. (2018). According to this approach, a significant contrast exists between the enriched (heavy) DSW (δ²H = 11.3‰) and the depleted (light) natural water (i.e., fresh groundwater sampled before

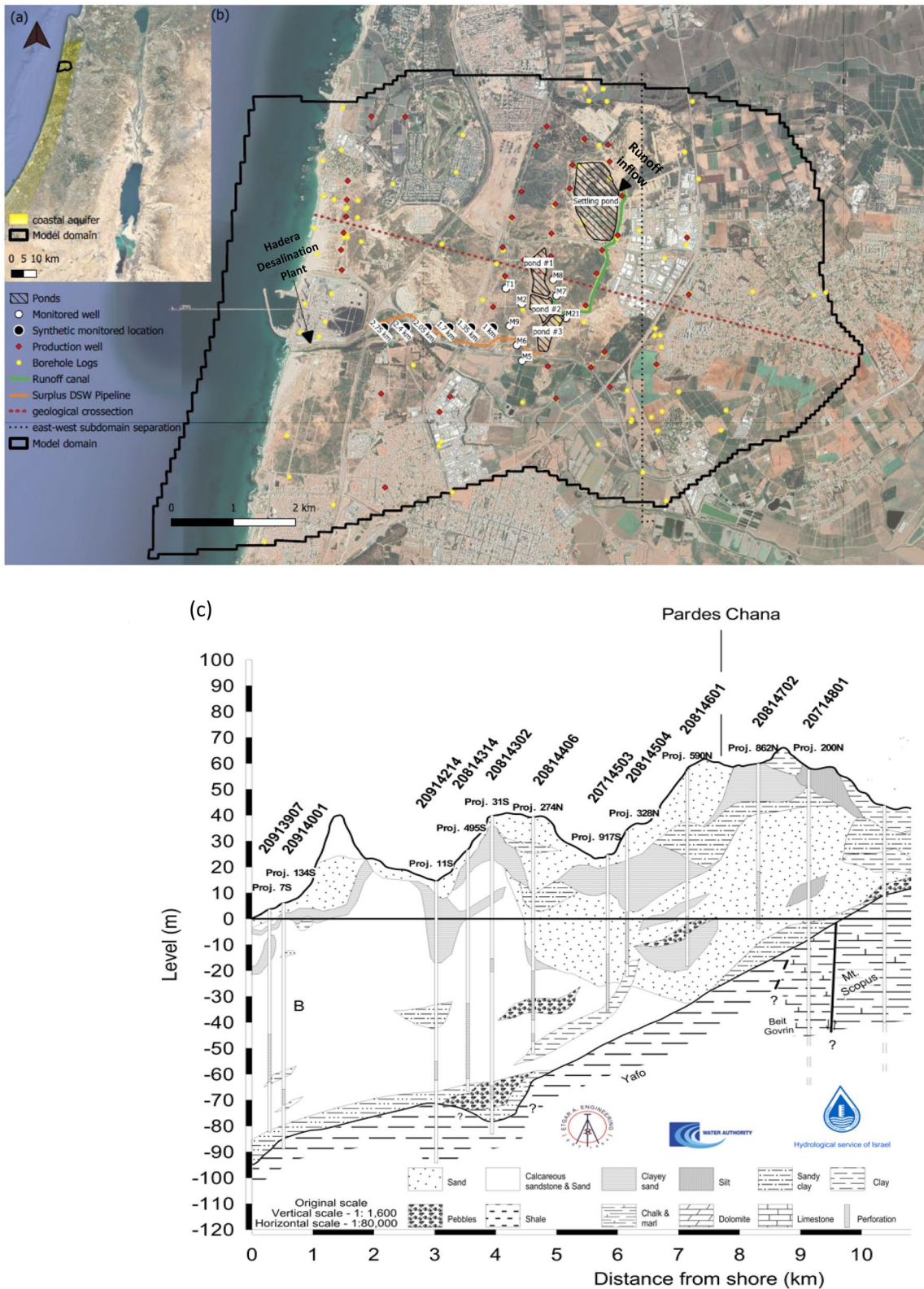


Figure 1. Map of the study region. (a) Location of the Israeli Coastal Aquifer (yellow) and model domain (black polygon). (b) The Menashe Streams managed aquifer recharge (MAR) site. Rainfall runoff from an upstream north-eastern catchment is diverted into the settling and infiltration ponds, and surplus desalinated seawater (DSW) is diverted to pond #3 from the Hadera Desalination Plant (Background aerial: © Google Earth 2021). (c) Representative geological cross-section (source: Hydrological Service of Israel). Cross-section location marked with dashed red line at (b).

Table 1
Annual Recharge Volumes in Menashe MAR Site in 2015–2019, From Two Sources, and Annual Production Volumes

Year	Annual recharge volumes in Mm ³		Production in Mm ³ from public and private wells (48 in total)
	Rainfall runoff	DSW	
2015	3.2	2.6	13.9
2016	0	1.3	15.4
2017	1.7	0.6	14.4
2018	7.8	2.2	15.6
2019	13.1	3.7	17.1

2015, runoff and rain; $\delta^2\text{H} = -22.7\text{‰}$). The $\delta^2\text{H}$ value of each water sample containing a mixture between the two endmembers is then used to establish the normalized concentration of DSW (“ C_{DSW} ,” can also be referred to as “DSW fraction” or “mixing ratio”) in the sample (Equation 2 and Table 2). The data presented in Table 2 supplement the existing published data (Ganot, Holtzman, Weisbrod, Bernstein, et al., 2018) from the years 2015–2017

$$C_{\text{DSW}} = \frac{\delta^2\text{H}_{\text{sample}} - \delta^2\text{H}_{\text{freshwater}}}{\delta^2\text{H}_{\text{DSW}} - \delta^2\text{H}_{\text{freshwater}}} * 100 \quad (2)$$

2.3. Model Development

First the model domain and the stochastic hydrogeological framework is presented (Figure 2; Section 2.3.1), followed by a description of the model's predictive phase, which builds on randomized time series of runoff recharge (Section 2.3.2). A numerical experiment consisting of four scenarios is described in Section 2.3.3 (Figure 3). The predictive stage in all scenarios of the numerical experiment follows a 5-year historic stage (2015–2019), which is described in Section 2.3.4 and was used to calibrate and validate the properties of the four categorical variables that were used in all the realizations of the aquifer. Additionally, this single calibrated and validated realization (referred to as the “validated realization”) is applied as the single aquifer realization in the scenario accounting for the climatic uncertainty (the blue scenario, Section 2.3.3).

2.3.1. Model Domain and Stochastic Subsurface Model

The entire model domain covers $\sim 70\text{ km}^2$ (see domain boundaries in Figure 1) with a $70 \times 70\text{ m}$ horizontal cell size. In the vertical direction, the aquifer thickness varies between 50 and 100 m, which are divided into 24 layers with a z -direction cell length of maximum 5 m. These cell dimensions were chosen to be able to represent aquifer heterogeneity with sufficient accuracy while keeping a balanced trade-off with computational costs. In total, the computational 3D grid consists of 470,000 cells.

The subsurface is represented by four categorical variables (i.e., hydrogeological units, hydrofacies, Table 3), as interpreted from over 100 borehole logs drilled during numerous geological studies conducted in the region over the decades (Figure 1b). The Transition Probability (T-ProGS) method, which is a Markov chain approach to geostatistical analysis and simulation of spatial distributions of categorical variables (Carle, 1999), was applied to generate the hydrogeological model of the subsurface in a single (Ganot, Holtzman, Weisbrod, Bernstein, et al., 2018) and multirealization approach, all conditioned to the log data. In the single realization approach subsurface model, constructed in a previous study (Ganot, Holtzman, Weisbrod, Bernstein, et al., 2018; Section 2.3.4), the model domain includes two subdomains ($\sim 2/3$ western and $1/3$ eastern, location of separation plane is marked with a dashed black line; Figure 1b) in order to increase hydrogeological realism, in which an evident difference exists in characteristics between two statistical subregions, as reflected in well logs' data (i.e.,

Table 2
 $\delta^2\text{H}$ Values at Five Sampled Wells (Years 2018–2019) and the Respective DSW Concentrations

Date	Well M7		Well M21		Well M2		Well M9		Well M6	
	$\delta^2\text{H}$ (‰)	C_{DSW} (%)	$\delta^2\text{H}$ (‰)	C_{DSW} (%)	$\delta^2\text{H}$ (‰)	C_{DSW} (%)	$\delta^2\text{H}$ (‰)	C_{DSW} (%)	$\delta^2\text{H}$ (‰)	C_{DSW} (%)
01 May 2018	-17.7	14.7	-17.8	14.3	-9.2	39.5	-10.3	36.4	-10.0	37.2
14 October 2018	-16.1	19.3	-16.9	17.1	-6.7	46.9	-11.9	31.8	-11.2	33.7
14 May 2019	-12.1	31.1	-12.1	31.1	-11.0	34.2	-5.5	50.6	-7.0	46.0
31 December 2019	-16.3	18.8	-12.8	29.0	–	–	-7.5	44.8	-7.4	44.8

Note. Well M2 malfunctioned in the second campaign of 2019.

Table 3
Hydrofacies Characteristics and Their Validated Model Parameters

Hydrogeological unit	1	2	3	4
Rock type	Gravel, beach rock, calcareous sandstone with shells and gravel	Calcareous sandstone, sand	Loam, sandy loam, loamy sand, marine silty sand	Clay and silt of marine or terrestrial origin
Unit proportions in eastern subdomain (%)	7.7	35.7	37.0	19.6
Unit proportions in western subdomain (%)	2.5	71.2	17.3	9.0
Unit proportions in entire domain (%)	4	59	23.5	13.5
Vertical mean lens length ^a (m)	5.45	14.6	6.87	4.63
Horizontal/vertical lens ratio ^a (–)	10.0	19.4	10.0	70.0
Horizontal K (K interval ^b) (m d ⁻¹)	50 (25, 75)	12 (6, 18)	6 (3, 9)	0.01 (0.005, 0.015)
Vertical K (m d ⁻¹)	12.5	3	1.5	0.01
Specific storage (m ⁻¹)	0.002	0.0015	0.001	0.001
Specific yield	0.35	0.12	0.12	0.1
Longitudinal dispersivity (m)	20	20	20	20
Porosity	0.35	0.19	0.17	0.1

^aWestern subdomain. ^bSee Section 2.3.3.

higher % of clay in the east; Figure 1c and Table 3). The western and eastern subdomains were generated separately, as following: (a) geostatistical properties of the categorical variables of the eastern and the western subdomains were calculated using log data collected from boreholes to the east and west of the subdomain separation plane, respectively (Figure 1b). (b) Two aquifer realizations were generated, conditioned to logs of the eastern and western subdomains, respectively. (c) The two realizations were superposed in a way that to the east of the separation plane, the eastern realization prevails and to the west of it—the western realization prevails. Thus, a single realization was created, which is conditioned to the entire set of log data. The significant differences in volumetric proportions of the four hydrofacies between the two subdomains (Table 3) are indicative of the statistical difference between the subdomains, which is accounted by the above-described superposition method. The calibrated hydrogeologic properties of the single subsurface model are presented below (Table 3, Section 2.3.4).

Herein, for the multirealization approach, 30 equally probable subsurface realizations are generated via T-ProGS (Figure 2). Thirty was found to be a sufficient number of subsurface realizations to reach statistical convergence as shown in Appendix A. The above described east-west separation was kept. We note that the eastern part of the flow domain has very little effect on the stochastic model results at locations of interest near the infiltration ponds and westward. The most eastern well analyzed is M21 which is 17 grid cells far from the east-west separation line (Figures 1 and 2). Thus, the eastern part of the single validated realization from Ganot, Holtzman, Weisbrod, Bernstein et al. (2018) was inserted in all realizations to the east of the separation plane, replacing the coinciding generated subset of cells (Figure 2). Furthermore, the calibrated values of the hydrogeologic properties were adopted from the single realization model (Ganot, Holtzman, Weisbrod, Bernstein, et al., 2018) to the multirealization model.

Table 3 presents a summary of the types and proportions of the four hydrofacies used to represent the subsurface, their main geostatistical properties (vertical mean lens length, i.e., the T-ProGS equivalent to the classic correlation length), and other properties calibrated in Ganot, Holtzman, Weisbrod, Bernstein et al. (2018, see Section 2.3.4). Finally, we note that model boundaries are deliberately located far away from the subdomain of interest (proximity of the infiltration ponds), thus model results are not affected by issues of seawater intrusion and tides.

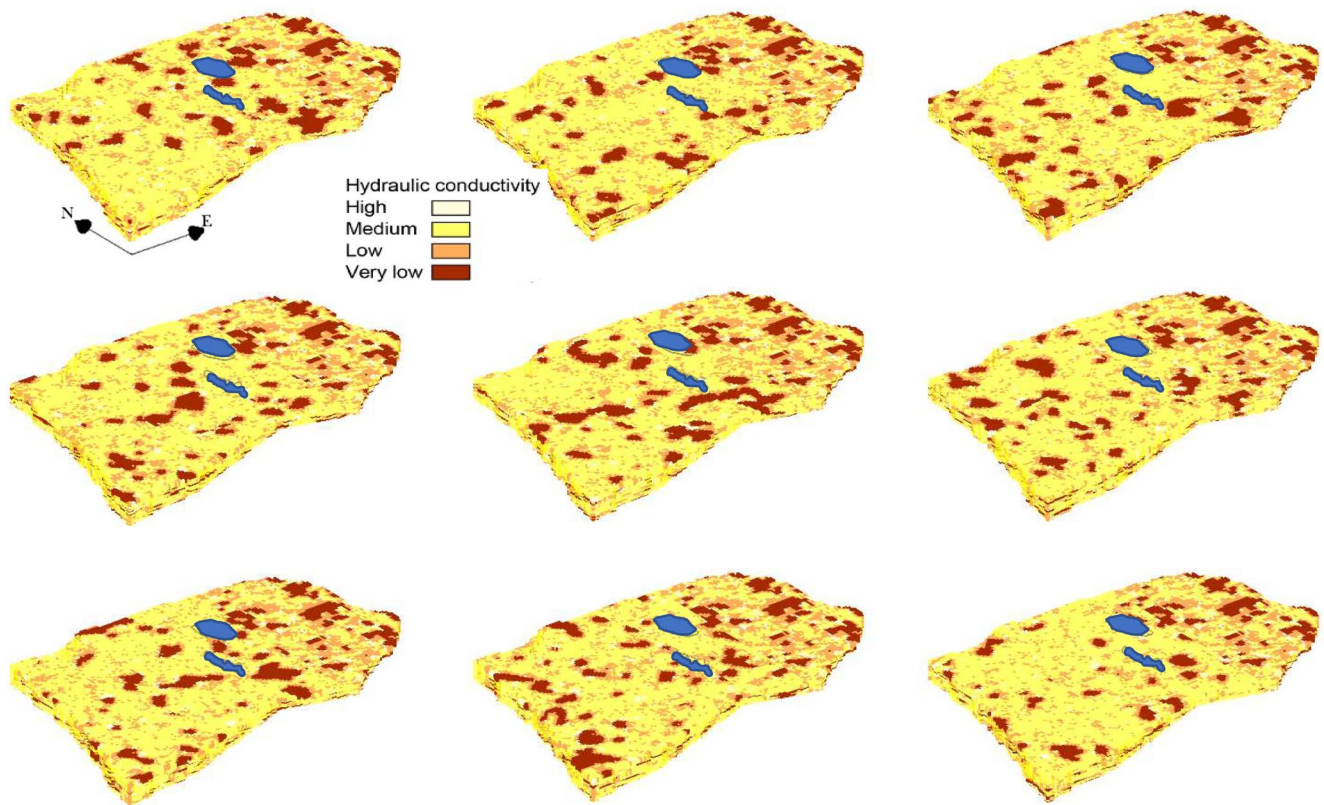


Figure 2. Nine random subsurface realizations (out of 30) used for scenarios with a variable subsurface (5× vertical magnification). The settling and infiltration ponds of Menashe managed aquifer recharge (MAR) site are marked in blue for spatial reference (see Figure 1). The four colored facies' hydraulic conductivities correspond with the hydrological units in Table 3.

2.3.2. Uncertainty of Runoff Recharge and the Transient Model Stages

All predictive simulations in this study consist of two subsequent stages: (a) a 5-year historic stage (2015–2019, “stage I,” see Section 2.3.4) and (b) a 25-year long projection phase (2020–2044, “stage II,” described herein). Stage II incorporates the following assumptions and properties: (a) Specified time-varying heads at boundary wells (i.e., wells located on the bounding polygon of the model and provide the temporal data for the specified head boundary condition) are set to reoccur annually according to the monthly observed heads of the latest available data (year 2019). In preliminary simulations (not described here), the model was found to be not sensitive to other possible settings of well boundary conditions (e.g., constant head values which are set as the means of available data at each well), due to the remoteness of the boundaries of the model from the subdomain of interest. (b) Production rates are annually reoccurring, with the same spatial and temporal distribution as those of 2019, i.e., 17.1 Mm³ total annual production. (c) The managed recharge of DSW is reoccurring annually, exclusively into pond #3 and with a constant volume which is the annual mean of 2015–2019, i.e., 2.14 Mm³. (d) In scenarios with double-source recharge during stage II, annual runoff recharge is predetermined as follows: (I) A gamma distribution (applicable to distributions of annual river flow and precipitation; Markovic, 1965) is fitted to the histogram of 50 years of historic recharge data from the Menashe MAR site (Kurtzman & Guttman, 2020) resulting in shape and rate parameters of 1.552 and 0.156, respectively (Figure 4a). (II) A time series of 25 years of runoff recharge volume is randomized to fit the gamma distribution (Figure 4). For the stochastic-climatic ensemble, 30 such equally probable time series were generated (Figure 4b). It should be noted that 30 was found to be a sufficient amount of realizations for our purpose as shown in Appendix A. (III) Annual runoff recharge volumes of each year are distributed among the settling ponds, pond #1 and pond #2 according to Equation 3, where $V_{T,i}$, $V_{SP,i}$, $V_{1,i}$, and $V_{2,i}$ are recharge volumes (Mm³) in year i at the entire site, the settling pond, infiltration pond #1 and infiltration pond #2, respectively (Figure 1). In short, runoff is distributed so that the first 3 Mm³ are assigned to the settling pond while any volumes above 3 Mm³ are divided in between the two infiltration

Table 4
Simulated Ensembles and Their Characteristics

Ensemble (color)	Subsurface structure	Aquifer parameters	Stage II runoff recharge	Stage II DSW recharge
Single Sub-Multiple RR (blue)	Validated realization (Ganot, Holtzman, Weisbrod, Bernstein, et al., 2018)	Validated, constant	Varying randomized time series at each simulation (Figure 4b)	Constant annual
Multiple Sub-Single RR (green)	Stochastic realizations	Validated, constant	Single randomized time series at each simulation	Constant annual
Multiple Sub-0 RR (yellow)	Stochastic realizations	Validated, constant	No runoff recharge	Constant annual
Multiple Sub Rand K-0 RR (magenta)	Stochastic realizations	Randomized inner-facies K	No runoff recharge	Constant annual

ponds. (IV) Annual volumes of runoff and DSW recharge were temporally distributed over each single year with reoccurring typical tendencies of recharge fluxes (according to the infiltration-recharge calibrated model of Ganot et al. (2017)). (e) Direct precipitation is applied uniformly on the entire model domain with reoccurring monthly rates (i.e., means of historic monthly data with a 0.4 recharge coefficient, which is a realistic coefficient for sparsely vegetated, sandy soil in the Mediterranean climate, used in similar regions to the studied aquifer)

$$\left\{ \begin{array}{l} V_{sp,i} = V_{T,i} \\ V_{1,i} = V_{2,i} = 0 \\ V_{sp,i} = 3 \\ V_{1,i} = 0.6(V_{T,i} - 3) \\ V_{2,i} = 0.4(V_{T,i} - 3) \end{array} \right\} \begin{array}{l} V_{T,i} < 3 \\ \\ V_{T,i} \geq 3 \end{array} \quad (3)$$

2.3.3. Simulated Ensembles

In order to address the issues of interest, numerical experiments were conducted, consisting of four simulation ensembles, as following (summarized in Table 4 and Figure 3): (a) Single Sub-Multiple RR (blue ensemble in following plots) consists of 30 simulations with the same reoccurring (validated) subsurface realization and a different randomized runoff recharge scenario at stage II for each simulation (Figure 4b). Mean annual runoff recharge of the different realizations varies between 7.7 and 12.9 Mm³. (b) Multiple Sub-Single RR (green ensemble): 30 stochastic subsurface realizations (Figure 2) and a single, reoccurring, scenario of randomized runoff recharge at stage II (see Figure 4c). Runoff recharge realization #3 was chosen since its mean annual runoff recharge of 9.9 Mm³ (st. dev. = 5.5), is close to the historic mean value. Ensembles (a) and (b) are designed in a way that allows to account for an individual source of uncertainty by “canceling” the other in a given ensemble. However, the decision of using a single reoccurring realization of the subsurface (in (a)) and of runoff recharge time series (in (b)), is accounted for by means of a sensitivity analysis of the ensemble to a chosen deterministic “reality.” In both ensembles (a) and (b), the same analysis was performed with a different, arbitrary deterministic single realization which did not affect the statistics of the ensemble and main conclusions of this study. (c) Multiple Sub-0 RR (yellow ensemble): 30 stochastic subsurface realizations and a single source of recharge (DSW) at stage II. (d) Multiple Sub Rand K-0 RR (magenta ensemble): 30 stochastic subsurface realizations with hydraulic conductivity randomized within the category (hydrofacies) and a single source of recharge (DSW) at stage II. Parameter values used for the ensembles (a), (b), and (c), as shown in Table 3, are as calibrated in Ganot, Holtzman, Weisbrod, Bernstein et al. (2018) and validated here (Section 2.3.4). Ensemble (d) was constructed in order to account for the uncertainty presented by well-log geological interpretation when using discrete categorical variables, and for the fact that values which are calibrated for a certain subsurface realization are applied (in ensembles (b) and (c)) with other equally probable realizations, without recalibration. In ensemble (d), at each realization, each grid cell is assigned a value of hydraulic conductivity, which is randomized from a uniform distribution according to its preassigned facies. The minimum and maximum bounds of the distribution for each facies were set as 50% and 150% of the

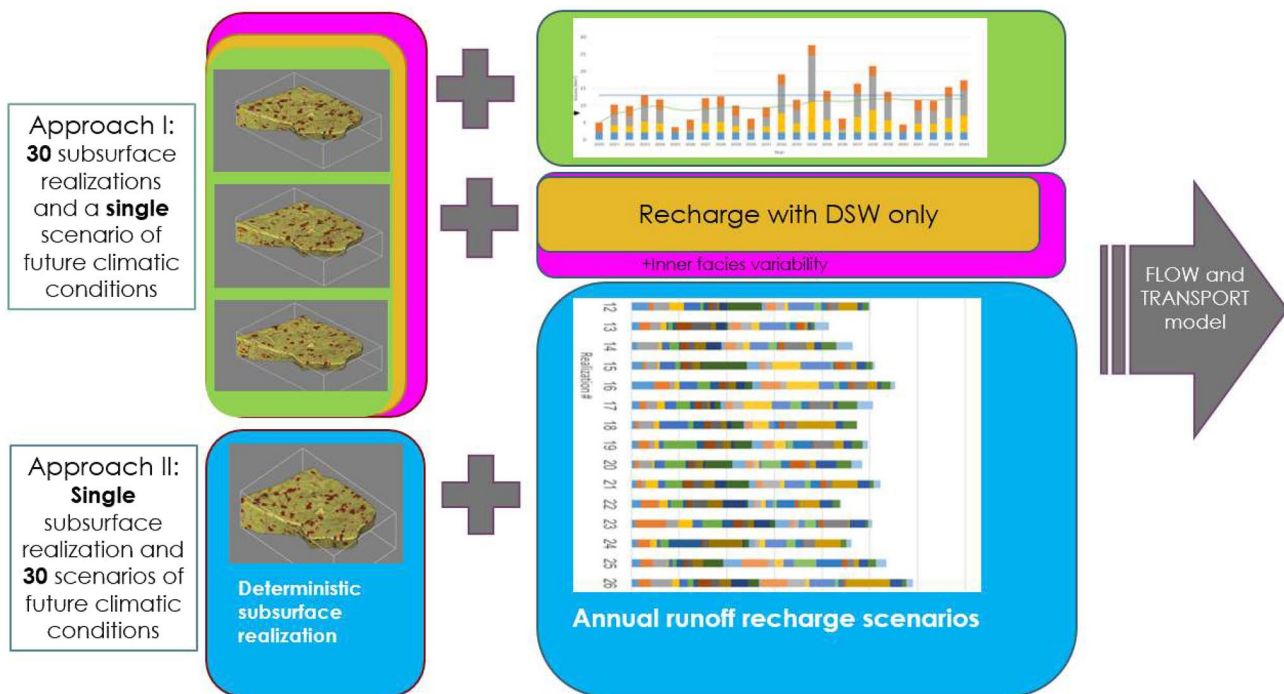


Figure 3. Schematic of the conducted numerical experiment with four ensembles.

validated values of horizontal hydraulic conductivity (Table 3), to create ranges of values that would remain realistic.

In total, 120 simulations were conducted using the Python scripting package FloPy (Bakker et al., 2016), incorporating MODFLOW-2005 (Harbaugh et al., 2017), and MT3DMS (Zheng & Wang, 1999) for flow and transport, respectively. Each simulation consists of 2,125 stress periods: 1,825 daily stress periods at stage I followed by 300 monthly stress periods at stage II.

2.3.4. The Calibrated/Validated Realization

The hydraulic and geostatistical properties of the generated subsurface realizations (Figure 2 and Table 3) were adopted (Section 2.3.1) from a single realization that was calibrated to 2015–2017 data (Ganot, Holtzman, Weisbrod, Bernstein, et al., 2018) and validated herein with the new 2018–2019 data. The single realization flow and transport model were run with MODFLOW (Harbaugh et al., 2017) and MT3DMS (Zheng & Wang, 1999) codes, respectively, with daily stress periods and the following boundary conditions: specified hydraulic head of 0 was set at the aquifer western boundary at sea level (due to the nonsensitivity of the model to western head boundary conditions, tidal conditions were not implemented) and a time-varying head was specified at northern, southern, and eastern aquifer boundaries according to head observations at wells located in the vicinity of the boundaries. The same type of boundary conditions was used for the stochastic realization runs of stage II (Section 2.3.2). Inputs of managed recharge for the calibrated realization model were distributed between ponds #1 and #2 as described in Equation 3. Infiltration pond #3 is strictly used for DSW recharge. Inputs of pumping rates and operation times of production wells distributed throughout the region were taken according to data provided by the Water Authority and Mekorot for the calibration and validation period (2015–2019). In the water isotope-transport model, following Section 2.2.2, mass influx boundary conditions of 100% were set at the sea boundary and for pond #3 recharge (which receives only DSW), and 0% elsewhere (ambient natural groundwater, runoff recharge, and direct rainfall-recharge). These percentages indicate the concentration of DSW in the water. Figures 5a and 5b show simulated head results and available observed values at wells M8 and M5 (Figure 1). To account for missing values in these two wells and provide further validation assurance, an additional observation well was used, well T1, which was not used during the calibration process of the model. Results shown in

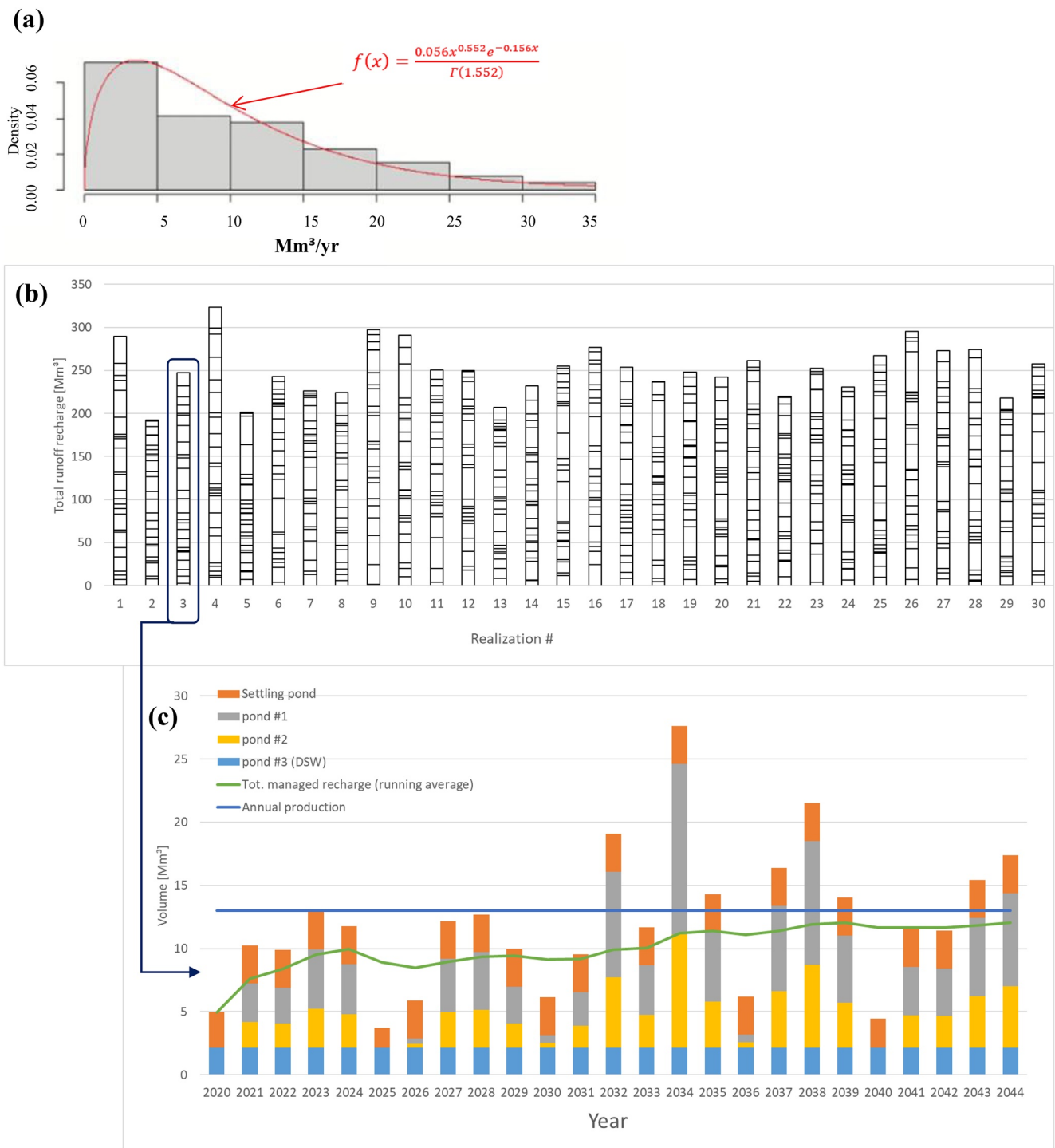


Figure 4. (a) Gamma distribution PDF (red) fitted to the histogram of 50 years of runoff recharge (gray bins). (b) Thirty synthetic time series of annual runoff recharge volumes, randomized from gamma distribution fitted to historic data. Each stacked bar represents the total annual runoff recharge value for the entire site, used as input for ensemble Single Sub-Multiple RR (blue ensemble). Each of the 25 columns within a single bar represent a year of stage II (2020–2044 from the bottom up). (c) A single synthetic time series of annual recharge values, reoccurring at stage II of each simulation of the Multiple Sub-Single RR (green) ensemble. Spatial distribution (by the single pond) is assigned according to Equation 3 (Figure 1).

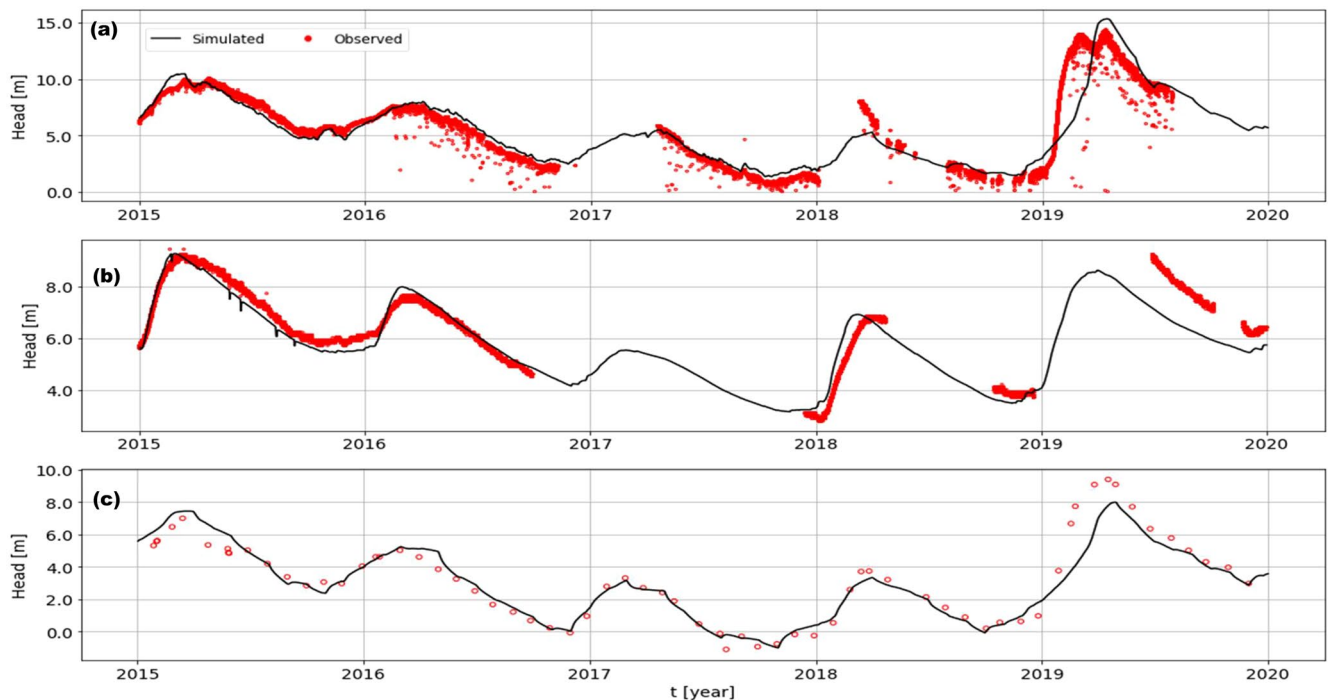


Figure 5. (a–c) Observed and simulated head values at wells M8, M5, and T1, respectively. Wells M8 and M5 were used for calibration (2015–2017) and validation (2018–2019). Well T1 (sampled monthly) was not used for calibration, thus the whole 5 years serve as a validation period. Highly scattered values and missing data at wells M8 and M5 are due to periodic pumping and malfunction.

Figure 5c present a good fit between observed and simulated heads as indicated by the correlation coefficients between observed and simulated heads: 0.98, 0.95, and 0.94 for the calibration ($n = 42$), validation ($n = 25$), and entire period ($n = 67$), respectively. Larger error observed in 2019 is on account of the dramatic hydrologic conditions of this year, compared to calibration years in which there was much smaller runoff recharge (Table 1).

Validation of the transport model was performed against isotope data (Table 2) from five wells collected during four additional field campaigns, performed in May 2018, October 2018, May 2019, and December 2019. In the model, tracer concentrations are normalized according to Equation 2. A reasonably good fit between observed and simulated concentrations is presented in Figure 6 with correlation coefficient values of 0.75, 0.81, and 0.88 in the calibration ($n = 21$), validation ($n = 19$), and total period ($n = 40$), respectively. The DSW breakthrough curves of the 5-year phase of the calibration and consecutive validation years (2015–2019; “stage I”) are presented in Section 3.

3. Results and Discussion

In the following sections and figures, resulting heads and mixing ratios are in units of meters and $\%_{\text{DSW}}$ in a sample, respectively. The predictive uncertainty of heads and mixing ratios is presented here via the transient coefficient of variation (CV, i.e., standard deviation of a population normalized by its mean) of mixing ratios (“ $\text{CV}_{\text{conc. DSW}}$ ”) at several selected locations during stage II.

3.1. Distance From Recharge Pond

We investigate groundwater heads and DSW to runoff water mixing ratios as well as the uncertainty of their predictions at varying distances from the infiltration ponds. Synthetic Monitored Locations (SML) are defined (Figure 1). These SML are six model cells located within a distance between 1.0 and 2.75 km downgradient (west) from an assumed central point in pond #3 (where the DSW is recharged), with 0.35 km separating any two

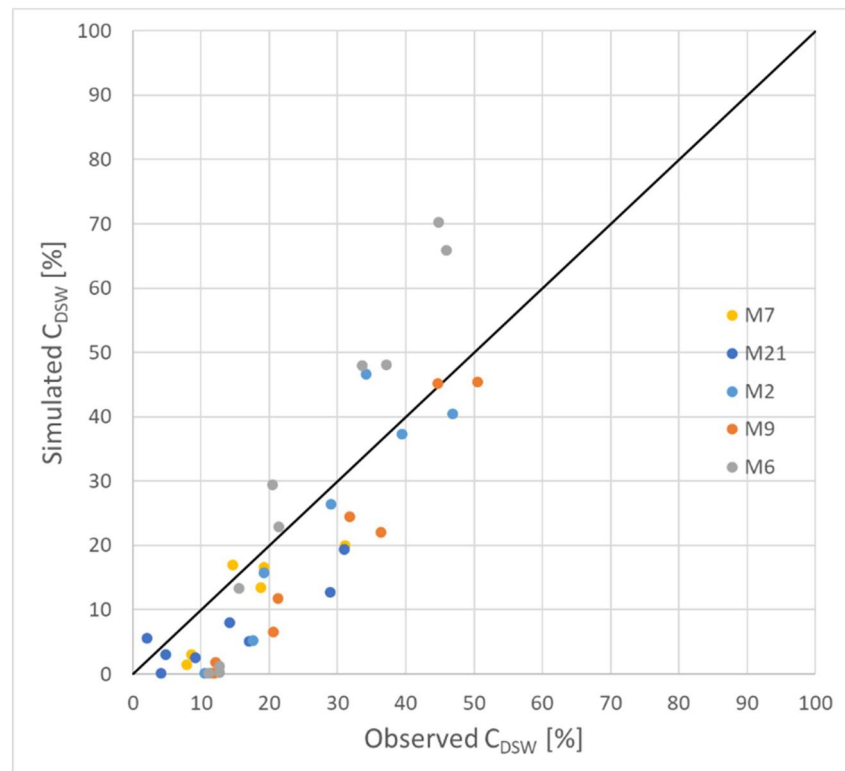


Figure 6. Comparison of observed and simulated desalinated seawater (DSW) concentration values at five wells used for transport model validation.

consecutive SML cell centers. Vertically, the SMLs are located at a mid-domain layer (layer 12 out of 24) and were chosen at locations without pumping wells in their proximity, in order to depict the state of the aquifer in regions not directly affected by pumping.

In Figure 7, heads and mixing ratios (proportional to $\delta^2\text{H}$ values; Table 2, Equation 2) at the six locations are presented. The closest location (SML #1) was chosen at 1 km, which is the minimum distance to have 0 DSW concentration at the beginning of stage II. It is clearly seen in Figures 7a–7f that predicted head results are prominently more variable for the “Single Sub-Multiple RR” (blue) ensemble compared to other ensembles (regardless of distance). From Figures 7g–7l, it is evident that the first arrival of the DSW plume progresses $\sim 70 \text{ m yr}^{-1}$ in all simulated ensembles (typical order of magnitude of velocity in this aquifer $\sim 100 \text{ m yr}^{-1}$ far from pumping wells, e.g., Kurtzman et al., 2012).

Figure 8 presents the CV of DSW concentrations as a function of time for the four different simulation ensembles. Before the first arrival of DSW to the SML, there is no uncertainty associated with the mixing ratio, i.e., CV is 0 initially. Then, CV rises to a peak a few years after the first arrival.

A peak CV value of 2.0 is found for the Single Sub-Multiple RR (blue) ensemble (Figure 8a), at a distance of 1 km, ~ 2 years into the predictive stage, where it is larger than other ensembles' CV by up to a factor of 2.6, at this distance. The blue peak drops significantly with distance (by $\sim 100\%$ within 1 km; Figure 8a) indicating the decreased impact of climatic variability (represented by runoff recharge) on the uncertainty at larger distances from the recharge source. However, this is not the case in the Multiple Sub-Single RR (green) ensemble, where maximum peak CV values drop after the first SML but then rise back up and are generally between 0.6 and 0.9, with no evident trend which proves that uncertainty of predictions is not necessarily proportional to distance, when the aquifer heterogeneity is the main source of uncertainty.

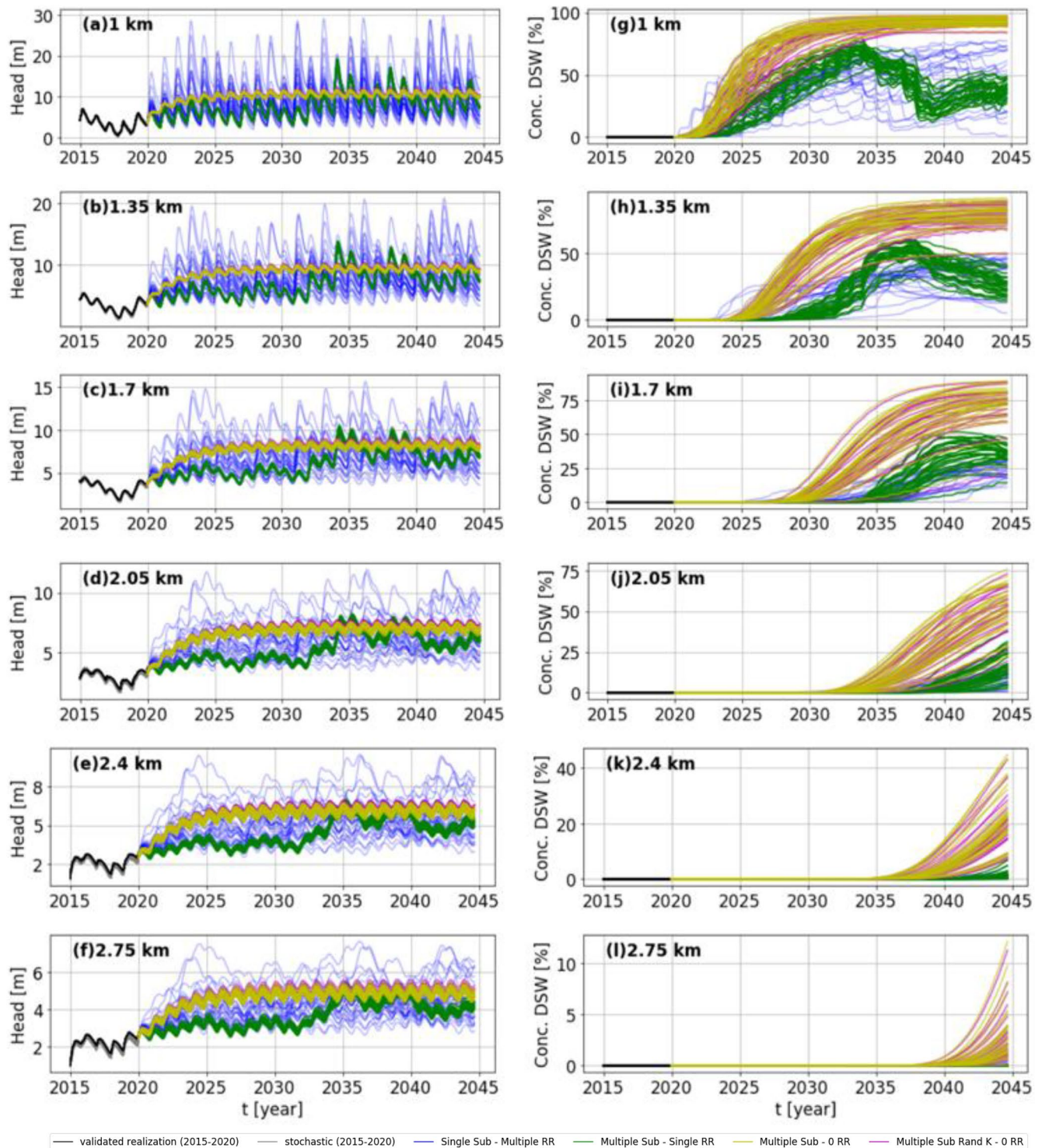


Figure 7. Transient groundwater heads (a–f) and fractions of desalinated seawater (DSW) values (g–l and Equation 2) at Synthetic Monitored Locations (SMLs) 1–6, respectively, as a function of time, for the four simulation ensembles. The cells are located at increasing distances from pond #3 (DSW infiltration pond), from 1 to 2.75 km at SMLs 1 and 6, respectively (350 m between two consecutive cells, see Figure 1).

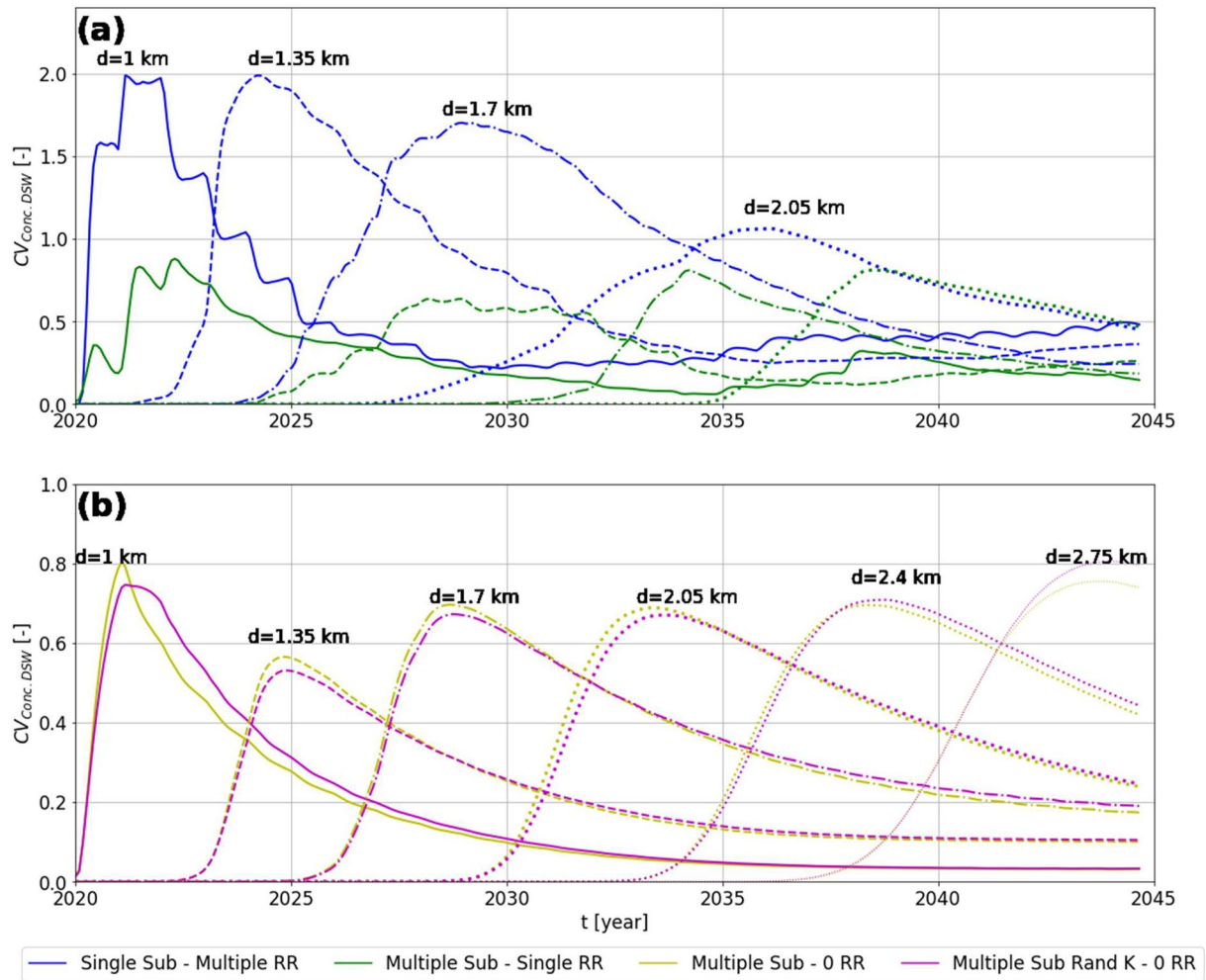


Figure 8. Coefficient of variation (CV) of desalinated seawater (DSW) concentrations as a function of time at increasing distances from pond #3 for the predictive stage (2020–2044), considering the four simulation ensembles: (a) double-source managed aquifer recharge (MAR) ensembles (i.e., with DSW and runoff recharge) and (b) single constant source MAR ensembles (i.e., DSW only). Note difference in vertical axis scales.

Peak CV values of all three ensembles with deterministic recharge time series are stable over distance (green scenario in Figure 8a and both scenarios in Figure 8b), while the stochastic recharge scenario is not (blue scenario; Figure 8a), due to high variability in annual recharge volumes.

3.2. Climate Driven Versus Subsurface Driven Uncertainty

Two main drivers of uncertainty that can be discussed in the context of predictive uncertainty of subsurface mixing processes at an MAR site are uncertainty attributed to the heterogeneity of the subsurface and the uncertainty arising from climatic variability (which affects recharge rates). In order to discuss the separate contribution of each driving source, we compare the results of two ensembles: Single Sub-Multiple RR (blue ensemble) and Multiple Sub-Single RR (green ensemble).

Figures 9a–9e present simulated and observed DSW fraction (2015–2019) and predictions (2020–2044) in the water produced (simulated) at downgradient production wells M2, M9, and M6 as well as upgradient wells M7 and M21 (Figure 1). A measure for the degree of uncertainty is the envelope of probability, i.e., the envelope surrounding the clustered curves in these figures, for each ensemble. A much narrower envelope of probability is seen for the Multiple Sub-Single RR (green) ensemble, which leads to lower CV values in Figures 9f–9j, in comparison to the Single Sub-Multiple RR (blue) ensemble in all wells throughout the entire stage II simulation.

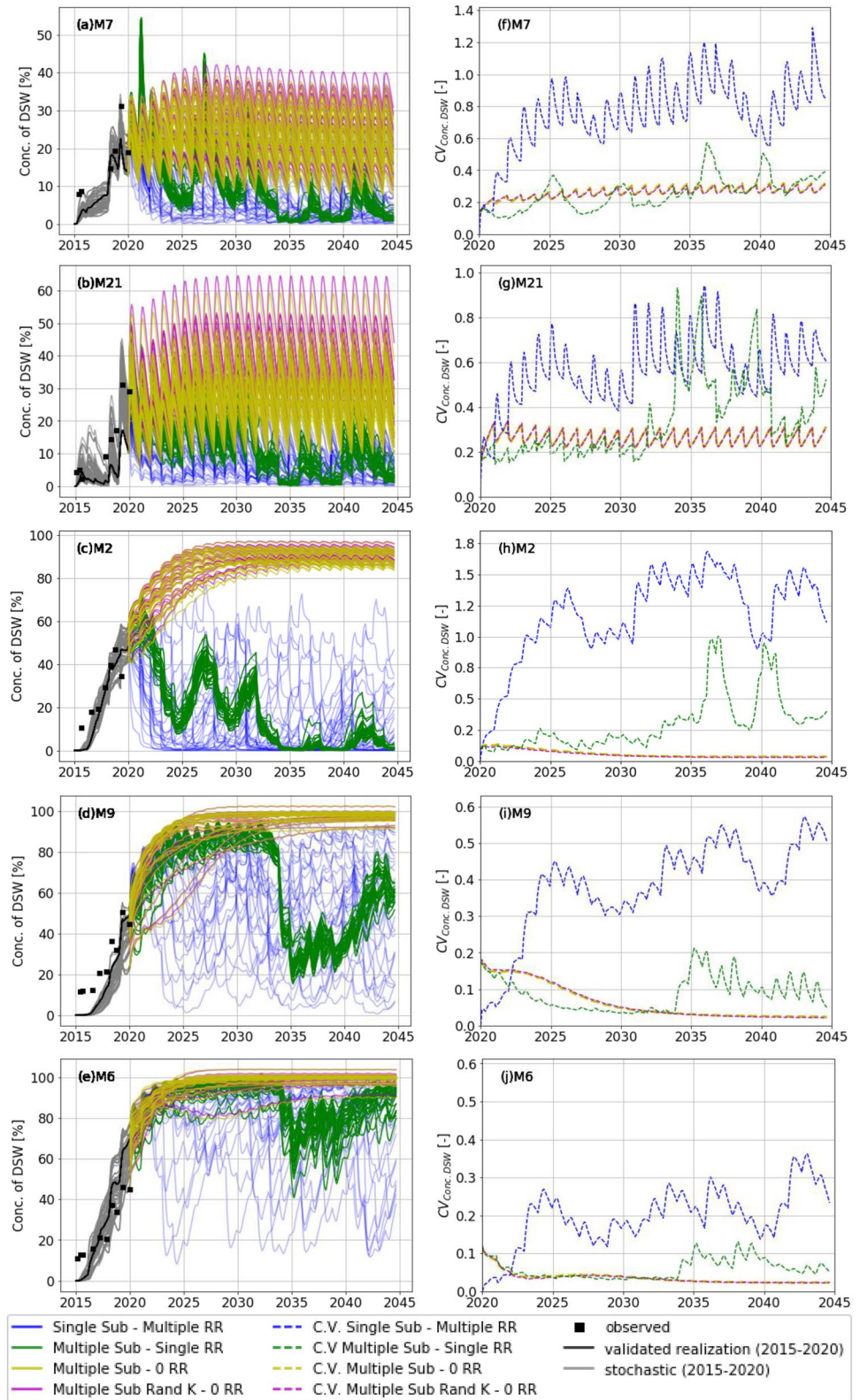


Figure 9. (a–e) Resulting concentrations of desalinated seawater (DSW) as simulated in wells M7, M21, M2, M9, and M6, respectively, for the four simulated ensembles. (f–j) Coefficient of variation of concentrations for the five wells at the predictive stage.

An exception is seen in the two first years of the predictive stage at wells M9 and M6 (Figures 9i–9j). In the green ensemble, the maximum CV peaks ($CV = 1.0$) can be observed in well M2 (Figure 9h). In all other locations, and apart from these peaks, the green ensemble presents relatively low predictive uncertainty, which leads to a conclusion that well-scale mixing predictions can be made confidently at a double-source MAR site, when recharge volumes from both sources are predefined, regardless of structural uncertainty. The significant uncertainty caused by the randomization of future annual recharge of runoff volumes at each realization can be further noticed by the sudden rise of CV in all downgradient wells around the beginning of the second half of stage II in the Multiple Sub-Single RR (green) ensemble. This coincides with the increase in mean annual runoff recharge volume (Figure 4c), reoccurring at each simulation at this ensemble, from 7.8 Mm³ at the first half of stage II (2020–2031) to 12.2 Mm³ at the second half (2032–2044).

3.3. MAR Source Type

To demonstrate the importance and usefulness of the new model, we present an implementation of an additional scenario represented by the Multiple Sub-0 RR (yellow) ensemble, as follows. We assume that due to a hypothetical reason (e.g., land use policies) it is decided that only pond #3 remains operative and DSW is the sole source of recharge with a constant annual flux. However, production rates at wells remain the same. In this ensemble, as seen in Figure 9, DSW fractions at downgradient wells (M2, M6, and M9) stabilize at a value above 80% after ~10 years into stage II, maintaining a sustainable recharge and recovery operation with easily controlled concentrations and low predictive uncertainty. These high fractions of DSW in downgradient wells are close to the deterministic prediction for these wells predicted by Ganot, Holtzman, Weisbrod, Bernstein et al. (2018). The similarity is due to the fact that in Ganot, Holtzman, Weisbrod, Bernstein et al. (2018) the deterministic prediction was based on reoccurring 2015 MAR operation, in which runoff recharge was small and was all in the settling pond whereas DSW recharge was relatively high (Table 1), not very far from the MAR scenario of the yellow ensemble here. On the other hand, at upgradient wells (M7 and M21), DSW fraction of produced water is fluctuating annually between 10% and 45% due to incoming fluxes of natural groundwater from the east and the predictive uncertainty is higher (Figure 9, yellow and magenta). Moreover, comparing CV results of Multiple Sub-Single RR (green) and Multiple Sub-0 RR (yellow) ensembles, in which the representation of structural uncertainty is equivalent, but recharge sources and volumes differ, suggests that structural uncertainty has a larger effect on predictive uncertainty in a double-source MAR site in comparison to a single source MAR site, mainly in the second part of stage II, where runoff recharge volumes are higher. This is inconclusive, however, due to similar CV values at the first part of stage II, when runoff recharge is moderate.

3.4. Hydraulic Conductivity Uncertainty

The role of hydraulic conductivity uncertainty in predictions is investigated via the Multiple Sub Rand K-0 RR (magenta) ensemble in Figure 9. The similarity between CV results of this ensemble and the Multiple Sub-0 RR (yellow) ensemble at all locations and times (Figures 9f–9j) are indicative of the fact that representation of the additional structural uncertainty caused by inner facies variability of hydraulic conductivity, in most cases, does not impact the predictive uncertainty. The inner facies variability causes only slight increase in uncertainty at well M9 close to the breakthrough of the new source of recharge water (Figures 9d and 9i). Furthermore, in the four closest SMLs to pond #3 (e.g., within 2.05 km from the pond), a slight decrease occurs in the CV peaks (Figure 8b).

Another conclusion arising from the comparison of Multiple Sub Rand K-0 RR (magenta) ensemble and the Multiple Sub-0 RR (yellow) ensemble regards the calibration procedure. Their similarity suggests that (at least in the case of our model) hydraulic conductivity values of a hydrofacies subsurface model, which are calibrated and validated against transient data for a certain aquifer realization, can be confidently applied for other, equally probable realizations generated with the same geostatistical parameters, without the need to recalibrate for each realization. This is a result of the finding that heterogeneity is more dominantly represented by the four facies distribution in space, than the inner facies heterogeneity.

Table 5
Predicted Probabilities (P) to Meet the Recommended Minimum Concentration for [Mg²⁺] in Produced Water in Five Different Wells and Four Simulated Ensembles (P = # of Realizations [Mg²⁺] > 10.0 mg L⁻¹/# of Realizations in Ensemble)

	M7	M21	M2	M9	M6
Average [Mg ²⁺] before DSW recharge (mg L ⁻¹)	14	11.2	11	9.7	11.9
<i>P</i> ([Mg ²⁺] > 10 mg L ⁻¹) in January 2030					
Single Sub-Multiple RR (blue)	1	0.70	0.56	0	0
Multiple Sub-Single RR (green)	1	0	0.06	0	0
Multiple Sub-0 RR (yellow)	1	0	0	0	0
Multiple Sup Rand K-0 RR (magenta)	0.96	0	0	0	0

Note. See Figure 1 for wells and DSW infiltration pond location.

3.5. Applicability of Modeling Approach

The stochastic experiment presented in this work can be applicable to other multisource MAR sites—existing or in the design stage. In case of a design of a new MAR site, the described approach that suggests isolating the uncertainty from different sources within different ensembles (e.g., green and blue), is intended, among others, to provide practitioners a tool, to get initial understanding of the uncertainty levels embedded within different sources. Such insight can help optimize resources allocation in the design stage of a new site. For example, when only subsurface related uncertainty is considered, one can decide that the uncertainty levels of the long-term predictions of mixing ratios of the two recharged water types are too high, and try reducing them by maximizing subsurface investigation (e.g., borehole data, geophysical campaign, etc.). However, one might realize that allocating more resources to subsurface exploration, would be wasteful, as it will not contribute to reducing the uncertainty embedded in hydroclimatic conditions of prediction years. We are aware, however, that sites with such a rich history of recorded data as Menashe are not a commodity. Where time series of runoff recharge are not available because of lack of recorded data, rainfall-runoff models can be used to generate them (considering rainfall data is much more common).

In existing single-source sites, where there is an intention to add recharge water from an additional source (e.g., DSW; Zhang & Wang, 2021), the operator should take public health issues in considerations, as mixing of the DSW with naturally fresh groundwater reduces the Mg²⁺ concentration within the produced water. In this work, this can be demonstrated via well M2, which prior to DSW recharge campaigns, produced water with 11 mg L⁻¹ [Mg²⁺] (above minimum World Health Organization (WHO) recommendation of 10 mg L⁻¹ [Mg²⁺] for public health considerations; Rosanoff, 2013), while the mean predicted Mg²⁺ concentrations for January 2030 are lower: 9.49 mg L⁻¹ (CV = 1.03) and 9.61 mg L⁻¹ (CV = 0.18), in the blue and green ensembles, respectively. These values were calculated under the assumption of conservative behavior of Mg²⁺ in the aquifer after it reaches 2.8 mg L⁻¹ in the shallow groundwater under the infiltration pond (Ganot, Holtzman, Weisbrod, Russak, et al., 2018). Probabilities for exceeding a 10 mg L⁻¹ threshold in an arbitrary future time (January 2030) in five wells for the four different ensembles are presented in Table 5. The issue of Mg²⁺ levels dropping below the minimum WHO recommendation can be clearly seen via the low probabilities in the table. Such analysis is advantageous for stake-holders and represents a more realistic picture of reality when it is presented in a probabilistic approach, considering various sources of uncertainty.

The problem of an MAR water carrying contaminants and their dilution in the aquifer is of special interest in MAR systems that uses the subsurface for SAT of low-quality water. A good example for dealing with mixing and its uncertainty of tracers of secondary effluents in wells surrounding the SAT infiltration pond of the Shafdan reclamation system (Ickson-Tal et al., 2003) is given by Gasser et al. (2010). The analysis of the dilution uncertainty of the relatively nonreactive pharmaceutical-carbamazepine is dealt in a steady-state approach. This fits the relatively steady state of the MAR-SAT source there (Tel Aviv Metropole's secondary effluents), and surrounding diluting aquifer which is relatively balanced with small fluctuations of natural recharge and abstractions. On the contrary, in the double-source MAR site analyzed here the quantity of the diluting water (runoff) is extremely variable in time and uncertain (Figure 4). Concentrating the recharge of runoff from ~100 km² watershed to ~40 ha infiltration basins amplifies the climatic uncertainty to overrule the structural uncertainty in wells close to

the infiltration basin in our case (Figure 9). Nevertheless, this climatic uncertainty declines with growing distance from the infiltration basins (Figure 8a). Hence, the modeling approach demonstrated in this study will be more advantageous in ephemeral runoff MAR sites.

4. Conclusions

The main result of this work is that in a double-source MAR site, when well-scale water mixing predictions are required, they can only be made with sufficient certainty (e.g., $CV_{\text{conc. DSW}} < 1.0$ in the Menashe MAR site, the highest CV value for the green ensemble; Figure 9h) if no climate-dependent uncertainty is involved. Subsurface heterogeneity is found to be an uncertainty driver of lesser significance than climate variability; however, the ratio between the peaks of uncertainty caused by the two drivers decreases and the similarity between them increases with growing distance from the pond in which the additional MAR source is recharged (Figure 8a).

When it is important for an MAR site operator to achieve certain mixing thresholds of the two water sources (e.g., to meet water-quality recommendations), it is beneficial if the process is simulated stochastically, via multiple aquifer realizations, in order to account and quantify hydrogeological uncertainty. This fundamental benefit of stochastic modeling is apparent in the results of Figures 9a–9e, where observed values are more likely to be represented by the envelope of the stochastic results, than by the single validated realization (deterministic) model. However, with a source of recharge water which is strongly fluctuating annually (e.g., runoff), it is very difficult to make any long-term predictions of the mixing ratios, regardless of the level of certainty in the subsurface structure interpretation. Thus, when there is a need to maintain within defined mixing thresholds, recharge volumes of the various sources should not only be planned ahead and controlled, but also, preferably, without large annual variations.

This work also shows how stochastic modeling can be used to address a pertinent issue of minimum Mg^{2+} levels recommended in drinking water. Results show that due to mixing of the low Mg^{2+} level DSW with native aquifer water, there is a high probability of overall Mg^{2+} concentrations dropping below the recommended minimum. Although for this analysis we only consider mixing, ignoring enrichment of DSW which takes place in the deep aquifer, it is a step toward further planned research, which will incorporate reactive transport in addition to mixing.

The obvious drawback of this study is its case-uniqueness; nevertheless, it can provide valuable insights for MAR site planners and operators in locations with similar climatic and hydrogeological conditions, which is probably where such double-source schemes would be considered in the first place. Moreover, the presented methodology for assessing uncertainty from different sources (i.e., structural and climatic) for a double-source MAR site can be generalized (at least) for other MAR sites, existing or designed, with two sources of recharged water, where one source is with a lesser volume (i.e., the DSW in our case), when the question of interest is either the dilution of recharged contaminated water from the secondary source, or like in our case, the remineralization (e.g., with Mg) of demineralized recharged water.

In future work, potential reactive mineral enrichment of the injected DSW will be quantified stochastically and the impact on predictive uncertainty will be investigated, a process which will be strongly supported by lessons learned from this work.

Appendix A: Statistical Conversion

The coefficient of variation of the mixing ratios “ $CV_{\text{conc. DSW}}$ ” is used in this study to quantify uncertainty of the DSW breakthrough curve. The statistically accepted large sample size 30 was chosen for the number of realizations in all ensembles. Herein, we test whether 30 is a large enough sample or does the statistics converge at this number of realizations. Figure A1 presents the $CV_{\text{conc. DSW}}$ values in the locations that produce the most uncertain results (well M2, SML1, SML2), both for the “blue” and for the “green” ensembles, where the 30 realizations represent 30 various runoff recharge scenarios, and 30 various subsurface representations, respectively. The transient CV values were calculated from varying ensemble sizes, with $n = 5, 10, 15, 20, 25,$ and 30 realizations from the respective ensemble.

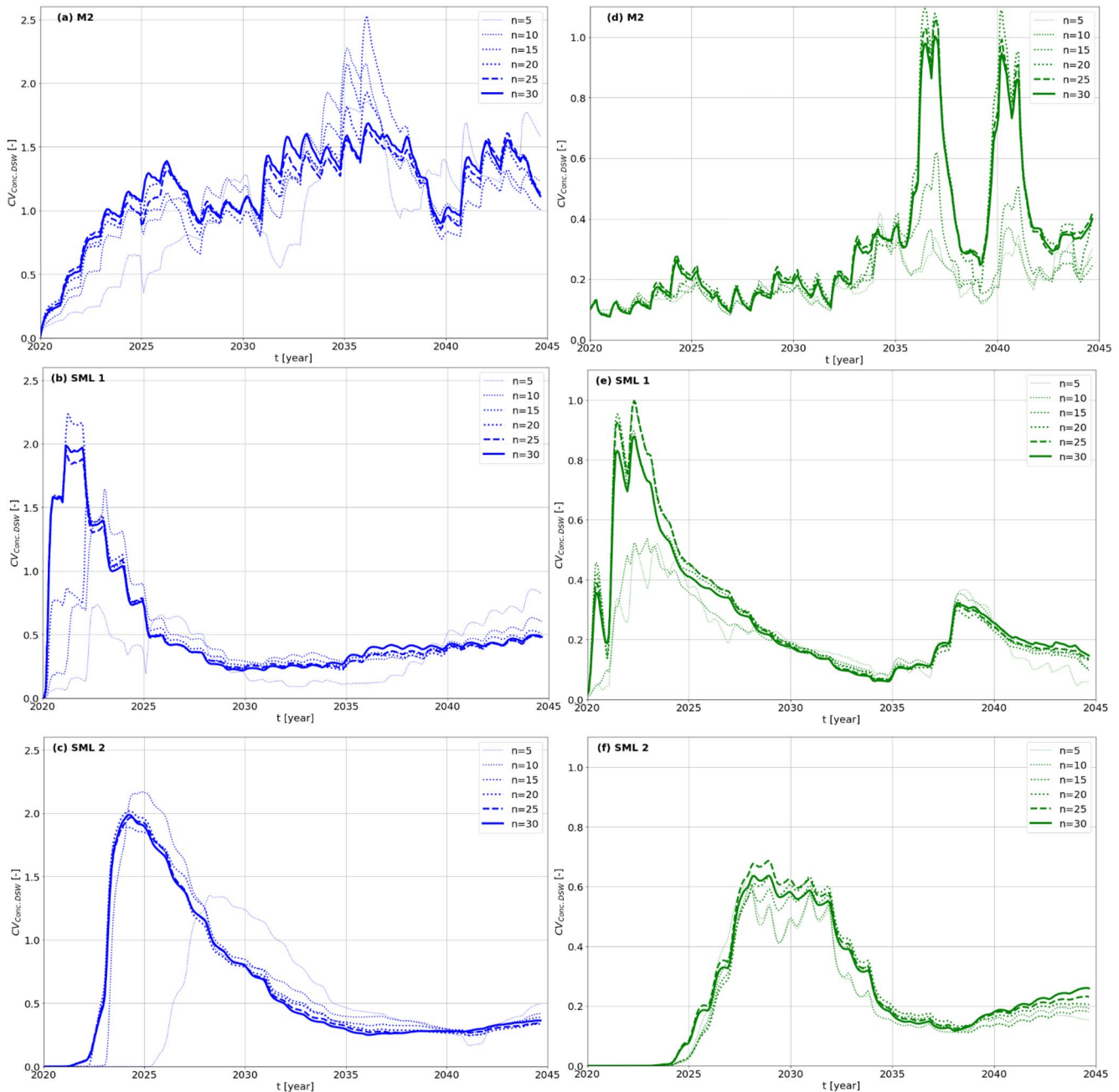


Figure A1. (a–f) Evolution of the coefficient of variation of the desalinated water breakthrough values at well M2, SML 1 and SML 2 (from top to bottom) for “blue” (left) and “green” (right) ensembles, with varying ensemble sizes (n).

As the differences between the CV values at $n = 25$ and at $n = 30$ become negligible, for both ensembles, and the CV values stabilize sufficiently for the cause of our study, it can be safely stated that $n = 30$ is a sufficient number of realizations to reach statistical convergence. It should be noted that increasing the number of realizations is always important for the sake of improving statistical convergence, however, taking in account the computational costs of the highly spatially and temporally discretized model such as here, $n = 30$ introduces balanced trade-off between accuracy (e.g., statistical convergence) and computational costs.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The underlying research data and code can be obtained from the following link: <https://doi.org/10.5281/zenodo.5512830>.

Acknowledgments

The research leading to these results received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 814066 (Managed Aquifer Recharge Solutions Training Network—MARSoLUT).

References

- Abdulrazzak, M. J. (1997). Water supply augmentation through artificial groundwater recharge techniques: Water Sciences & Technology Association. In *Proceedings of the Third Gulf Water Conference, Muscat, Sultanate of Oman* (pp. 241–281).
- Avni, N., Eben-Chaime, M., & Oron, G. (2013). Optimizing desalinated seawater blending with other sources to meet magnesium requirements for potable and irrigation waters. *Water Resources*, *47*, 2164–2176. <https://doi.org/10.1016/j.watres.2013.01.018>
- 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>
- Carle, S. F. (1999). *T-PROGS: Transition Probability Geostatistical Software* (p. 84). Davis, CA: University of California. Retrieved from <http://gmsdocs.aquaveo.com/t-progs.pdf>
- Carle, S. F., & Fogg, G. E. (2020). Integration of soft data into geostatistical simulation of categorical variables. *Frontiers of Earth Science*, *8*, 565707. <https://doi.org/10.3389/feart.2020.565707>
- Dillon, P., Fernández Escalante, E., Megdal, S. B., & Massmann, G. (2020). Managed aquifer recharge for water resilience. *Water*, *12*(7), 1846. <https://doi.org/10.3390/w12071846>
- Fiori, A., Cvetkovic, V., Dagan, G., Attinger, S., Bellin, A., Dietrich, P., et al. (2016). Debates—Stochastic subsurface hydrology from theory to practice: The relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. What is characterization and stochastic theory good for? *Water Resources Research*, *52*, 9228–9234. <https://doi.org/10.1002/2015WR017525>
- Ganot, Y., Holtzman, R., Weisbrod, N., Bernstein, A., Siebner, H., Katz, Y., & Kurtzman, D. (2018). Managed aquifer recharge with reverse-osmosis desalinated seawater: Modeling the spreading in groundwater using stable water isotopes. *Hydrology and Earth System Sciences*, *22*, 6323–6333. <https://doi.org/10.5194/hess-22-6323-2018>
- Ganot, Y., Holtzman, R., Weisbrod, N., Nitzan, I., Katz, Y., & Kurtzman, D. (2017). Monitoring and modeling infiltration-recharge dynamics of managed aquifer recharge with desalinated seawater. *Hydrology and Earth System Sciences*, *21*, 4479–4493. <https://doi.org/10.5194/hess-21-4479-2017>
- Ganot, Y., Holtzman, R., Weisbrod, N., Russak, A., Katz, Y., & Kurtzman, D. (2018). Geochemical processes during managed aquifer recharge with desalinated seawater. *Water Resources Research*, *54*, 978–994. <https://doi.org/10.1002/2017WR021798>
- Gasser, G., Rona, M., Voloshenko, A., Shelkov, R., Tal, N., Pankratov, I., et al. (2010). Quantitative evaluation of tracers for quantification of wastewater contamination of potable water sources. *Environmental Science & Technology*, *44*(10), 3919–3925. <https://doi.org/10.1021/es100604c>
- Gotovac, H., Cvetkovic, V., & Andricevic, R. (2009). Flow and travel time statistics in highly heterogeneous porous media. *Water Resources Research*, *45*, W07402. <https://doi.org/10.1029/2008WR007168>
- Harbaugh, A. W., Langevin, C. D., Hughes, J. D., Niswonger, R. N., & Konikow, L. F. (2017). *MODFLOW-2005 version 1.12.00, the U.S. Geological Survey modular groundwater model*. U.S. Geological Survey Software. <https://doi.org/10.5066/F7RF5S7G>
- He, X., Højberg, A. L., Jørgensen, F., & Refsgaard, J. C. (2015). Assessing hydrological model predictive uncertainty using stochastically generated geological models: Hydrological Modelling Using Stochastic Geological Models. *Hydrological Processes*, *29*, 4293–4311. <https://doi.org/10.1002/hyp.10488>
- He, X., Koch, J., Sonnenborg, T. O., Jørgensen, F., Schamper, C., & Christian Refsgaard, J. (2014). Transition probability-based stochastic geological modeling using airborne geophysical data and borehole data. *Water Resources Research*, *50*, 3147–3169. <https://doi.org/10.1002/2013WR014593>
- He, X., Sonnenborg, T. O., Jørgensen, F., Høyer, A.-S., Møller, R. R., & Jensen, K. H. (2013). Analyzing the effects of geological and parameter uncertainty on prediction of groundwater head and travel time. *Hydrology and Earth System Sciences*, *17*, 3245–3260. <https://doi.org/10.5194/hess-17-3245-2013>
- Hutchinson, C. B. (1998). Simulation of aquifer storage recovery of excess desalinated seawater. *U.S. Geological Survey Open-File Report 98-410; USGS: Al Ain area, Abu Dhabi Emirate*. Retrieved from <https://pubs.usgs.gov/of/1998/0410/report.pdf>
- Icekson-Tal, N., Avraham, O., Sack, J., & Cikurel, H. (2003). Water reuse in Israel—the dan region project: Evaluation of water quality and reliability of plant's operation. *Water Science and Technology: Water Supply*, *3*(4), 231–237. <https://doi.org/10.2166/ws.2003.0067>
- Jones, E., Qadir, M., vanVliet, M. T. H., Smakhtin, V., & Kangmu, S. (2019). The state of desalination and brine production: A global outlook. *Science of the Total Environment*, *657*, 1343–1356. <https://doi.org/10.1016/j.scitotenv.2018.12.076>
- Kurtzman, D., & Guttman, J. (2020). Managed aquifer recharge with various water sources for irrigation and domestic use: A perspective of the Israeli experience. In A. Mukherjee, B. R. Scanlon, A. Aureli, S. Langan, H. Guo, & A. A. McKenzie (Eds.), *Global groundwater, source, scarcity, sustainability, security, and solutions*, (pp. 597–607). Amsterdam, Netherlands: Elsevier. <https://doi.org/10.1016/B978-0-12-818172-0.00044-X>
- Kurtzman, D., Netzer, L., Weisbrod, N., Nasser, A., Graber, E. R., & Ronen, D. (2012). Characterization of deep aquifer dynamics using principal component analysis of sequential multilevel data. *Hydrology and Earth System Sciences*, *16*, 761–771. <https://doi.org/10.5194/hess-16-761-2012>
- Maliva, R. G. (2015). Managed aquifer recharge: State-of-the-art and opportunities. *Water Science and Technology: Water Supply*, *15*, 578–588. <https://doi.org/10.2166/ws.2015.009>
- Maples, S. R., Fogg, G. E., & Maxwell, R. M. (2019). Modeling managed aquifer recharge processes in a highly heterogeneous, semi-confined aquifer system. *Hydrogeology Journal*, *27*, 2869–2888. <https://doi.org/10.1007/s10040-019-02033-9>

- Maples, S. R., Foglia, L., Fogg, G. E., & Maxwell, R. M. (2020). Sensitivity of hydrologic and geologic parameters on recharge processes in a highly heterogeneous, semi-confined aquifer system. *Hydrology and Earth System Sciences*, 24, 2437–2456. <https://doi.org/10.5194/hess-24-2437-2020>
- Markovic, R. D. (1965). *Probability functions of best fit to distributions of annual precipitation and runoff* (Hydrology paper, No. 8). Fort Collins, CO: Colorado State University.
- Moreno, Z., & Paster, A. (2018). Prediction of pollutant remediation in a heterogeneous aquifer in Israel: Reducing uncertainty by incorporating lithological, head and concentration data. *Journal of Hydrology*, 564, 651–666. <https://doi.org/10.1016/j.jhydrol.2018.07.012>
- Parimalarenganayaki, S. (2021). Managed aquifer recharge in the Gulf countries: A review and selection criteria. *Arabian Journal of Geosciences*, 46, 1–15. <https://doi.org/10.1007/s13369-020-05060-x>
- Pool, M., Carrera, J., Alcolea, A., & Bocanegra, E. M. (2015). A comparison of deterministic and stochastic approaches for regional scale inverse modeling on the Mar del Plata aquifer. *Journal of Hydrology*, 531, 214–229. <https://doi.org/10.1016/j.jhydrol.2015.09.064>
- Rajaram, H. (2016). Debates—Stochastic subsurface hydrology from theory to practice: Introduction. *Water Resources Research*, 52, 9215–9217. <https://doi.org/10.1002/2016WR020066>
- Ringleb, J., Sallwey, J., & Stefan, C. (2016). Assessment of managed aquifer recharge through modeling—A review. *Water*, 8, 579. <https://doi.org/10.3390/w8120579>
- Rodríguez-Escales, P., Canelles, A., Sanchez-Vila, X., Folch, A., Kurtzman, D., Rossetto, R., et al. (2018). A risk assessment methodology to evaluate the risk failure of managed aquifer recharge in the Mediterranean Basin. *Hydrology and Earth System Sciences*, 22, 3213–3227. <https://doi.org/10.5194/hess-22-3213-2018>
- Rodríguez-Escales, P., Fernández-García, D., Drechsel, J., Folch, A., & Sanchez-Vila, X. (2017). Improving degradation of emerging organic compounds by applying chaotic advection in Managed Aquifer Recharge in randomly heterogeneous porous media. *Water Resources Research*, 53, 4376–4392. <https://doi.org/10.1002/2016WR020333>
- Ronen-Eliraz, G., Russak, A., Nitzan, I., Guttman, J., & Kurtzman, D. (2017). Investigating geochemical aspects of managed aquifer recharge by column experiments with alternating desalinated water and groundwater. *Science of the Total Environment*, 574, 1174–1181. <https://doi.org/10.1016/j.scitotenv.2016.09.075>
- Rosanoff, A. (2013). The high heart health value of drinking-water magnesium. *Medical Hypotheses*, 81(6), 1063–1065. <https://doi.org/10.1016/j.mehy.2013.10.003>
- Rosen, V. V., Garber, O. G., & Chen, Y. (2018). Magnesium deficiency in tap water in Israel: The desalination era. *Desalination*, 426, 88–96. <https://doi.org/10.1016/j.desal.2017.10.027>
- Sanchez-Vila, X., & Fernandez-Garcia, D. (2016). Debates—Stochastic subsurface hydrology from theory to practice: Why stochastic modeling has not yet permeated into practitioners? *Water Resources Research*, 52, 9246–9258. <https://doi.org/10.1002/2016WR019302>
- Sasidharan, S., Bradford, S. A., Šimůnek, J., & Kraemer, S. R. (2019). Drywell infiltration and hydraulic properties in heterogeneous soil profiles. *Journal of Hydrology*, 570, 598–611. <https://doi.org/10.1016/j.jhydrol.2018.12.073>
- Stuyfzand, P., Smidt, E., Zuurbier, K., Hartog, N., & Dawoud, M. (2017). Observations and prediction of recovered quality of desalinated seawater in the strategic ASR project in Liwa, Abu Dhabi. *Water*, 9, 177. <https://doi.org/10.3390/w9030177>
- Teramoto, E. H., Engelbrecht, B. Z., Gonçalves, R. D., & Chang, H. K. (2021). Probabilistic backward location for the identification of multi-source nitrate contamination. *Stochastic Environmental Research and Risk Assessment*, 35, 941–954. <https://doi.org/10.1007/s00477-020-01966-y>
- Vanderzalm, J. L., Page, D. W., Barry, K. E., & Dillon, P. J. (2013). Application of a probabilistic modelling approach for evaluation of nitrogen, phosphorus and organic carbon removal efficiency during four successive cycles of aquifer storage and recovery (ASR) in an anoxic carbonate aquifer. *Water Resources*, 47, 2177–2189. <https://doi.org/10.1016/j.watres.2013.01.038>
- Zekri, S., Triki, C., Al-Maktoumi, A., & Bazargan-Lari, M. R. (2021). Optimization of storage and recovery of seasonal surplus desalinated water in a coastal aquifer. *Arabian Journal of Geosciences*, 14, 100. <https://doi.org/10.1007/s12517-020-06340-4>
- Zhang, Z., & Wang, W. (2021). Managing aquifer recharge with multi-source water to realize sustainable management of groundwater resources in Jinan, China. *Environmental Science and Pollution Research*, 28, 10872–10888. <https://doi.org/10.1007/s11356-020-11353-3>
- Zheng, C., & Wang, P. P. (1999). MT3DMS: A modular three-dimensional multi-species transport model for simulation of advection, dispersion, and chemical reactions of contaminants in ground-water systems. Documentation and user's guide. In *Contract Report SERDP-99-1*. US Army Engineer Research and Development.