Abstract  Groundwater recharge plays an important role in water resource management. In arid and semi-arid areas, the groundwater recharge under agricultural fields constitutes a large portion of the total recharge. Estimating the agricultural recharge is important for groundwater management and assessing contamination risk. However, estimations of the recharge fluxes are difficult due to the soil heterogeneity and the complex nature of the flow processes in the soil. Here, we suggest a new method for estimating the recharge flux probability distribution that accounts for the soil heterogeneity. The method is demonstrated using a unique data set of long-term continuous measurements of water content under an agricultural field. The measurements are used to assign weights to a large ensemble of hydraulic parameter sets. The weights are then used to derive the probability distribution of flux under different agricultural practices. We found that the distribution of the recharge fluxes varies considerably between one annual crop and a cycle of three crops over 2 years. However, the distribution of the fraction of the input water that infiltrates does not vary much between the two practices. Under natural conditions, with no irrigation, the recharge flux is much smaller, and its distribution is narrower. Our approach provides the full probability distribution of fluxes rather than a single expected value of the flux or its mean.

1. Introduction

Groundwater recharge is an essential component of the water cycle. The total recharge is often divided into several components, including diffuse recharge, focus recharge, and agricultural recharge (Le Coz et al., 2013; Meixner et al., 2016; Wang et al., 2021). Agricultural recharge is important not only for its effect on the amount of groundwater (Bühlke, 2002; Ficklin et al., 2010) but also because it affects the contamination risk and solute leaching into the groundwater (Bühlke, 2002; Foster et al., 2018; Kurtzman et al., 2013; Scanlon et al., 2002; Walker et al., 2020). Taking water flux measurements in the unsaturated zone is challenging. Experiments that are based on lysimeters can provide direct and accurate measurements of input and output fluxes (Duncan et al., 2016; Scanlon et al., 2002). Up-scaling the results of a lysimeter experiment to the field scale, however, is difficult due to differences in the boundary conditions (e.g., wind, precipitation, and the container’s physical boundaries) and the internal soil structure (often the container is filled with perturbed soil).

Estimation of groundwater recharge requires a knowledge of the output fluxes. However, except for lysimeter experiments, these quantities cannot be directly measured. The flux may be estimated using groundwater age, a water table fluctuation analysis (Green et al., 2008), and a mass balance of tracers (Scanlon et al., 2002, 2006). A common method for estimating fluxes is using field measurements in the vadose zone to calibrate a model describing the flow and the flux (and groundwater recharge) (Turkeltaub et al., 2015). The advantage of using a model is that once calibrated, it is robust and can be used with any boundary and initial conditions. The disadvantages are the difficulty in assessing the range and values of the hydraulic parameters to represent the soil properties, the difficulty in describing the boundary conditions, and the difficulty in choosing the optimal model for balancing complexity and accuracy.

In order to assess the uncertainties associated with groundwater recharge and other hydraulic quantities, the probability distribution of the flux (or of other hydraulic quantities) is required. There are various methods for estimating the probability distribution of these quantities (Wu & Zeng, 2013), including considering specific forms of the probability distribution of model hydraulic parameters (e.g., Chen et al., 2018; Kosugi, 1996; Sánchez-Vila et al., 2009; Zhai & Benson, 2006) and the Bayesian approach (e.g., Abbaspour et al., 1999; Feyen et al., 2003;
Gao et al., 2019; Guber et al., 2009; Moreira et al., 2016; Schmidt-Walter et al., 2020; Scholer et al., 2012; Tsai & Li, 2008; Xu & Valocchi, 2015; Yones et al., 2019). Another approach combines methods for parameter selection and the Bayesian approach (Vesselinov & Harp, 2012), with a Monte Carlo-based estimation of the probability distributions (Mattis et al., 2015). Assessing the uncertainty in the values of the hydraulic parameters can also be done using an ensemble Kalman filter (Brandhorst et al., 2017), or the generalized likelihood uncertainty estimation (GLUE) method (Beven & Binley, 1992; Christiaens & Feyen, 2002; Kumar et al., 2010; Minasny & Field, 2005), or by combining several different pedotransfer functions for the same data set (Baroni et al., 2010; Loosvelt et al., 2011). These approaches rely on data collected from the field for the likelihood assessment.

Measurements in the unsaturated zone can provide information about the water content (WC) and solute concentrations. Common methods include collecting soil cores (Avishai et al., 2016; Russo & Bouton, 1992; Russo et al., 1997; Staver & Brinsfield, 1998), using natural or applied tracers (Allison et al., 1994; Flury et al., 1994; Morris et al., 2008) and detecting them in the soil cores or in the aquifer boreholes, and installing moisture sensors (Oehsner et al., 2003) and suction cups (Patterson et al., 2000) in vertical boreholes or excavated channels. All these methods involve disturbance of the soil structure, and some of them provide data for a limited time scale. Regardless of the measurement technique, the soil heterogeneity imposes a challenge on adequately sampling the soil hydraulic characteristics and conditions.

At the field scale, the heterogeneity plays an important role (Botros et al., 2009; De Vries & Simmers, 2002; Hartmann et al., 2017). However, most models describe a specific soil type with specified hydraulic parameters. The hydraulic parameters are not always directly measurable and are not homogeneous in space. Moreover, heterogeneity in flow characteristics may occur even in a seemingly homogeneous domain (Rimon et al., 2007).

Common approaches for assessing heterogeneity include: coring followed by direct measurements of saturated WC (θs), residual WC (θr), and saturated hydraulic conductivity (Ks) (Russo & Bouton, 1992; Russo et al., 1997; Sudicky et al., 2010); cores can also be analyzed for particle size distribution, which can be used in a soil parameter model such as ROSETTA (Schaap et al., 2001); assigned spatial distribution of the hydraulic parameters corresponding to known soil classes can also be applied (Carsel & Parrish, 1988; Tóth et al., 2017).

The assessment of soil heterogeneity using soil texture and particle size distribution alone does not guarantee an adequate description of the hydraulic properties of the heterogeneous soil (Hopmans et al., 2002). These properties include preferential flow paths and structural effects that lie beyond the local information. Moreover, in practice, the measurements required to assess the local hydraulic parameters are destructive, and therefore, characterizing the heterogeneity with a high enough resolution is not possible for a soil that will later be used for experiments/measurements.

Here, we use direct hydraulic measurements in order to assess the hydraulic properties of a heterogeneous vadose zone under an agricultural field. The database used here is taken from a field study that included continuous measurements of WC in unsaturated and undisturbed soil profiles (Weissman et al., 2020).

Our goal is to present a new method that combines measurements across deep sections of a heterogeneous unsaturated zone with modeling, to predict the entire probability density function of fluxes. The uniqueness of the method is twofold: (a) it relies on long-term continuous hydraulic measurements in assessing the relevance of different hydraulic parameters; and (b) a wide range of hydraulic parameter values is considered, and each set of parameter values is weighted according to the goodness-of-fit of the simulated WC to multiple field scale measurements. It is well known that the actual groundwater recharge is determined by the integrated sum of fluxes from all flow paths (De Vries & Simmers, 2002). The flux probability distribution function (PDF) incorporates the weighted relative fluxes occurring through the various flow paths that cross the natural heterogeneous unsaturated zone, while the common modeling approach provides a prediction of the most probable value of the flux or its mean. Therefore, predicting the recharge flux probability distribution is essential for contamination risk assessment and sustainable groundwater management.

2. Materials and Methods

Our approach for estimating the probability distribution of hydraulic quantities (e.g., WC at a specific depth or the flux at a given depth) is based on multiple measurements under the same field. Different sensors detect different WC curves (see Figure 2) representing different flow paths dictated by the conditions above and in the vicinity of...
the sensors. In order to derive the probability distribution of quantities that are not directly measurable, one must use a flow model. Every model requires parameters to characterize the soil, but the different measured curves may suggest different parameters. We use the multiple measurements in order to estimate the entire probability distribution of the quantities of interest rather than suggesting a single expected value. We consider a wide range of hydraulic parameter values and assess the goodness-of-fit of the simulated WC curves (using the different sets) to the measured WC curves. The goodness-of-fit is used to assign a weight to each set, assuming that the relative goodness-of-fit represents the set's relative importance in describing the field's hydraulic properties. Once the sets are weighted, we run simulations under every boundary condition of interest, and the quantity of interest that is simulated by each set is recorded. The different simulated values are then weighted according to the weights of the sets that were used for simulating them, and the entire probability distribution is estimated. Note that in this approach, each set contributes (unless its weight is zero) to the PDF. In what follows, we provide more details about the data, the model we used to simulate the water flow, the parameter sets, the weighting process, the validation of the predicted probability distributions, and the boundary condition scenarios for which we provide predictions of the recharge flux probability distributions.

### 2.1. Field Experiment

The data used in this study were collected as part of a field experiment in a semi-arid area in Israel. The experiment tested the effects of irrigation water's different salinity and fertilizer levels on water percolation and solute migration across deep unsaturated soil (≈4.5 m) under an agricultural field. A full description of the experimental setup and a detailed discussion of the measured results are provided in Weissman et al. (2020), while a brief description of the relevant parts of the setup is presented here. During a period of 2 years, three irrigated crops were grown in the field: potato, corn, and cotton. Four different treatments, in terms of salinity and fertilizer levels, were applied to different plots in the field using separated irrigation systems. Eight field plots were monitored using vadose zone monitoring systems (VMSs) (Dahan et al., 2009; Rimon et al., 2011; Turkeltaub et al., 2015, 2016). The VMS included 24 flexible time domain reflectometry (FTDR) sensors, based on a modification of Acclima time domain transmissometer (TDT) SDI-12 moisture sensors. The sensors were mounted on flexible sleeves, which were installed in slanted boreholes. The soil moisture sensors were placed at three depth levels (one sensor at each depth level in each plot): 150–190 cm, 250–290 cm, and 360–400 cm. This setup enabled monitoring the water flow in unsaturated and undisturbed deep soil. Volumetric WC measurements were taken in 30-min intervals, and were later averaged to daily values. A data set of over 900 days of measured WC values is used here. These data can be found in Weissman et al. (2022).

### 2.2. Modeling Setup

Continuous measurements of the temporal variations in the sediment WC at multiple depths enabled the estimation of the soil hydraulic properties' distribution, as it is reflected in the hydraulic heterogeneity. This, in turn, enables the use of numerical models in order to predict the PDF of water and solute fluxes.

We assume that the multiple measurements adequately represent the soil heterogeneity in the field. Therefore, a comparison of the simulation results, based on different hydraulic parameters, with the different measurements can help in assessing the relative importance of each parameter set.

Our main interest is the water flux in the deep unsaturated zone, and therefore, we simulate the water flow using a 1D spatial model. It was found that the 1D model may underestimate solute fluxes and groundwater contamination hazards (Russo et al., 2013). However, our modeling approach is not based on a single 1D soil profile but rather on a large number of soil profiles, each represented by a different set of soil hydraulic parameters and weighted according to the similarity to the different measurements. It is important to clarify that the use of a simple 1D model does not imply that the soil is indeed a simple homogeneous column. The actual representation of the soil is a weighted combination of all the 1D profiles.

Our simulations are based on the HYDRUS 1D program (Šimůnek et al., 2013), which solves the Richards equation numerically. We assume the 1D approximation to be sufficient, as we do not expect any lateral flows in our field. We used the simplest closure model, the van Genuchten-Mualem single porosity model (Van Genuchten, 1980) with no hysteresis. The vertical depth of the profile is 500 cm with a single textural layer (a uniform soil column). Model complexities, such as heterogeneous 1D columns (two or more textural layers), 2D/3D simulations, hysteresis, or dual porosity closure models, can be considered with this approach if one finds any such complexities to be essen-
that the soil heterogeneity was well captured. We tested our results using even wider ranges and finer resolutions. The different parameter sets that we used spanned a wide range of values with a fine enough resolution to ensure realistic range (e.g., very high saturated WC, meaning (such as residual WC, etc.). Therefore, some of the parameter values that we use here are outside the realistic range (e.g., very high saturated WC, meaning (such as residual WC, etc.). Therefore, some of the parameter values that we use here are outside the realistic range (e.g., very high saturated WC, meaning (such as residual WC, etc.). Therefore, some of the parameter values that we use here are outside the

In order to model the important effects of the crops on the soil water dynamics, the root water uptake was included in the dynamics. The root uptake model was defined as the default Feddes transpiration reduction function, Feddes et al. (1978), with specific root water uptake parameters for each of the three crops. Despite its simplicity, it was shown that this representation of the root water uptake performs well in simulating soil water dynamics (de Melo & de Jong van Lier, 2021). The parameters for potato and corn were taken from the built-in database in HYDRUS, and for cotton, we used the parameters that were suggested by Taylor and Ashcroft (1972) and were used by Forkutsa et al. (2009). The root growth data were defined according to the beginning and ending times of each crop in the field experiment.

The water flow boundary conditions were an atmospheric boundary condition with surface runoff and free drainage. The field experiment included eight individual field plots, and each of them was irrigated with slightly different water amounts and timing. For this reason, our simulation setup included eight independent simulations with eight sets of irrigation inputs (corresponding to the irrigation in the field plots). The precipitation, potential evapotranspiration (PET), and leaf area index (LAI) were identical for all eight simulations. The precipitation for the time period of 1996–2020 and part of the PET data for that time were measured at a meteorological station that is located 800 m from the field experiment. The meteorological data were retrieved through Agrimeteo (2021), Israeli Meteorological Service (2021) and from D. Bonfil (2021, personal communication, 10 August). Most of the missing PET data were calculated using the Penman-Monteith equation based on the method by Allen et al. (1998) and after Turkeltaub et al. (2018), using daily values of maximum temperature, minimum temperature, relative humidity, radiation, and wind speed that were measured in the same meteorological station and retrieved from the Israeli Meteorological Service (2021). The remaining missing data were interpolated (52 out of 9,132 daily PET values) by creating a normal distribution for each day of the year, based on the existing measurements, and taking a random value within this distribution. The LAI data for potato, corn and cotton were interpolated from 6, 12, and 9 data points, respectively, which were graphically generated based on previous studies (Bufon et al., 2011; Liu et al., 2002; Monteiro et al., 2006; Villa et al., 2017). The interpolated LAI data were then adjusted (rescaled) according to local studies (Gitelson et al., 2003; Kaplan & Rozenstein, 2021; Nguy-Robertson et al., 2014). All of the irrigation, LAI, and meteorological data that we used are available in the Data Set S1 in Supporting Information S1 and in Weissman et al. (2022).

The depths of the observation points were specified for each of the eight setups (corresponding to the eight plots) according to the depths of the WC sensors that were used in each plot in the experiment. The initial condition for all the simulations was a WC of 0.2, throughout the soil profile. However, all the simulations were initiated at a time corresponding to 3 years before the experimental period (namely, a 3-year spin-up), with atmospheric boundary conditions and no irrigation or crops, reflecting the actual situation in the field. This spin-up was added in order to attain soil conditions that resemble those in the field before the experiment. Obviously, the simulated WC during the spin-up period was not used in our analysis.

2.3. Parameter Setup

The model includes five varying parameters: the residual WC, $\theta_r$, the saturated WC, $\theta_s$, the saturated hydraulic conductivity, $K_s$, the pore size distribution parameter, $n$ and the fitting parameter, $\alpha$. In what follows, we refer to each combination of these five parameters’ values as a parameter set. In this work, we use a weighted combination of different parameter sets to represent the hydraulic behavior of the entire field. We do not consider these parameters as directly measured quantities representing their physical meaning (such as residual WC, etc.). Therefore, some of the parameter values that we use here are outside the realistic range (e.g., very high saturated WC, $\theta_s$ values). However, all our parameter values are mathematically allowed, and the sets obey the expected relations between the parameters (e.g., $0 < \theta_r < \theta_s < 1$).

The different parameter sets that we used spanned a wide range of values with a fine enough resolution to ensure that the soil heterogeneity was well captured. We tested our results using even wider ranges and finer resolutions...
to verify that our results are not sensitive to the specific choice of parameter range and resolution. The parameter values that were used to generate the parameter sets are specified in Table 1. Our setup consisted of 9,900 different parameter sets, each of which was used to simulate the water dynamics in the field’s eight plots.

### 2.4. Weighting Setup

Each set is weighted according to the goodness-of-fit of the simulated WC (using this set of parameters) to the measured WC in each of the sensors within the depth level of interest. It is important to note that despite the fact that we used depth homogeneous hydraulic parameters (i.e., the hydraulic parameters are identical throughout the simulated soil layer), we assign the weights at each depth level independently of the other levels.

We use the root mean square error (RMSE) as a measure of the goodness-of-fit. The RMSE is defined as:

\[
RMSE = \sqrt{\frac{1}{N_t - 1} \sum_{t_{\text{start}}}^{t_{\text{end}}} (W_{Cm}(t) - W_{C}(t))^2},
\]

where \(W_{Cm}(t)\) is the measured daily averaged WC on day \(t\), \(W_{C}(t)\) is the simulated daily averaged WC on day \(t\), and \(N_t\) is the number of days between the beginning of the weighting period, \(t_{\text{start}}\), and its end, \(t_{\text{end}}\). The RMSE value is calculated for each parameter set, with respect to each measured curve (individual sensor), separately. The weighting period can be different for the different purposes of validation or prediction. The weight of each parameter set, \(W_i\), is set according to the following function:

\[
W_i = \left( \frac{1}{RMSE_i} \sum_{i} \frac{1}{RMSE_i} \right)^{-1}.
\]

Here, \(i\) indicates the parameter set, and \(x\) is a power value that can be used to tune the weight differences between the parameter sets. Note that in general, it is possible to use other weighting functions that monotonically decrease with the RMSE. In addition, in our analysis, we did not have any set that resulted in a vanishing RMSE, so there was no divergence of the expressions in Equation 2.

### 2.5. Validation Setup

The method proposed here relies on two assumptions: (i) that the different sensors in the field provide adequate sampling of the heterogeneity of the soil hydraulic parameters; and (ii) that the simplistic model we use to simulate the water flow is good enough. The latter assumption is motivated by: (a) the fact that we are interested in the hydraulic quantities in the deep vadose zone; and (b) several tests (the results of which are not shown here) that suggested that including several soil layers (each with a different set of hydraulic parameters) or hysteresis does not considerably affect the results. In order to test

| Parameter | Values |
|-----------|--------|
| \(\theta_r\) (cm\(^3\) cm\(^{-3}\)) | 3 values: 0.01, 0.075, 0.14 |
| \(\theta_s\) (cm\(^3\) cm\(^{-3}\)) | 11 values: \(0.225 \pm \frac{0.075-0.225}{10}, 0 \leq i \leq 10; i \in \mathbb{N}\) |
| \(\alpha\) (cm\(^{-1}\)) | 10 values: \(0.005 + \frac{0.075-0.005}{4}, 0 \leq i \leq 4; 0.0875 + \frac{1.44-0.075}{5}, 1 \leq i \leq 5; i_1, i_2 \in \mathbb{N}\) |
| \(n\) | 5 values: 1.2, 1.325, 1.45, 1.575, 1.7 |
| \(K_s\) (cm day\(^{-1}\)) | 6 values: 20, 316, 612, 908, 1,204, 1,500 |

Note. Graphic presentation of the values in this table is shown in Figure 1.

---

**Figure 1.** Ranges (bars) and specific values (circles) that were used to define the parameter sets. The values are also presented in Table 1.
these assumptions (i and ii above), we perform a test to validate the method. The validation is done by using the WC data from four field plots as “training data” and the data from the remaining four field plots as “validation data.” The training data were taken from the measurements in field plots 1–4 and from the simulations of the same plots (with the corresponding irrigation boundary conditions). These plots were chosen for the training data since the measured WC values in these plots span the widest WC range at all three depth levels. The training data are used to assign the weights to the different parameter sets. The weighted ensemble of simulated WC curves is used to estimate the probability distribution of the WC. Finally, the assessment of the validation results is done by comparing the predicted PDF with the observed PDF.

To ensure that we use data showing responses to the input during the experiment rather than data reflecting pre-experimental conditions, we did not use the initial part of the data. The responses at different depth levels occurred at different times, and therefore, the used period for different depth levels is different. We refer to the day on which the experiment started as day 1. The beginning of the period of used data was set as \( t_{\text{start}} = 122, 327, \text{and } 513 \text{ days, for depth levels 1–3, respectively (shown in Figure 2).} \) The end of the period of used data was set to \( t_{\text{end}} = 1,000 \).

Our goal is to provide a reliable prediction of the flux probability distribution, rather than the common goal of predicting the most probable value of the flux or its mean. Therefore, validation of our method entails comparing the distribution of measured WC values with the predicted distribution, using the validation data (of field plots 5–8). Obviously, this comparison omits the precise synchronization between the predicted values and the measured ones. However, due to the limited data, we do not compare daily distributions for WC values. The observed distribution is derived from a histogram of the measured values or, equivalently, from the empirical cumulative distribution function. The predicted distributions are generated by dividing the entire range of predicted values into bins and then assigning, for each bin, a weight corresponding to the sum of weights of all the sets that simulated predicted values within the range of that bin. This method was used for both the predicted WC distributions and the predicted flux distributions that are shown in the following sections. Note that the WC distributions, which are based on predictions for multiple times, were normalized accordingly.

One way to compare measured and predicted distributions is the reliability diagram. These diagrams show the observed quantile versus the predicted quantile. We created each reliability diagram by searching through the predicted PDF for the quantiles that are closest to the 0.1 interval quantiles (0.9, 0.8…), with the limitations that are dictated by the structure of the analyzed PDF (bin sizes). Each resulting reliability curve had a different number of points of slightly different quantiles. The diagonal of the reliability diagram represents a perfect prediction of the probability distribution, and the deviation of the curve from the diagonal measures the similarity between the observed and predicted distributions.

### 2.6. Weight Function Optimization

At this stage, we used the reliability diagrams in order to optimize the power value, \( x \) (Equation 2). It is important to note that as opposed to the weights, which are based on an assessment of the goodness-of-fit using the RMSE, the final product, the flux PDF, is assessed using the reliability diagram. The optimum is found by varying the width of the weight function (by varying \( x \)) in order to minimize the deviation of the reliability curve from the diagonal. This deviation, \( M \), is defined as

\[
M = \frac{1}{N_q} \sum_{i} |p_i - o_i|
\]
where \( p_i \) and \( o_i \) are the predicted and observed quantiles in the reliability diagram of the PDF for which the weight function is being optimized, and \( N_q \) is the number of quantiles (points along the curve) that are considered in the reliability diagram.

The optimization was done in two stages: first by testing power values in fixed intervals that were relative to the tested range, and then by optimizing around the best \( x \) value (of the initially tested values). Optimization in the second stage was done using the `fminbnd` function in MATLAB (MATLAB Fminbnd, 2022), which is based on a golden section search and parabolic interpolation. The optimization was done for each depth level separately (namely, for each depth level, we used a different weighting function).

### 2.7. Flux Prediction

#### 2.7.1. Short-Term Predictions

Short-period predictions of water flux were analyzed in order to demonstrate the use of the weighting function (see Section 3.2 below). These predictions were generated using the same simulation setup that was used for the validation (Section 2.5), but with two repetitions of the experimental period assuming that the conditions of the first 1,000 days were repeated in the following 1,000 days. The simulation time that was analyzed for fluxes spanned 2,000 days.

The weights for this setup were calculated with respect to all the sensors. The weight of each parameter set, for each plot, is the average of the weights assigned to the set with respect to the three sensors (at three different depths) in this plot. The different plots have slightly different boundary conditions, and therefore, there is no average over the plots. The final flux PDF is based on the 9,900 × 8 different flux values, each with its corresponding weight.

#### 2.7.2. Long-Term Predictions for Different Scenarios

Long-period predictions of water flux distributions under the experimental agricultural field are created for three scenarios. The scenarios are simulated using the full parameter setup (see Section 2.3). The first scenario corresponds to repeating crop cycles that were actually practiced in the field during the experiment (three crops in 2 years: potato, corn, and cotton). The second scenario corresponds to repeating cycles of one crop a year (corn). The third scenario corresponds to no cultivation of the field, that is, resembling natural conditions with no irrigation. The irrigation amounts and timing for the first two scenarios were set to be identical to the irrigation provided to the field during the corresponding measurement periods. The LAI data were identical to the data that were used in the validation setup. For all scenarios, we simulate 25 years using the measured atmospheric conditions from 1996 to 2020 (see Section 2.2). The first 3 years are identical in all scenarios with no crops and atmospheric conditions only (spin-up), and the flux during the remaining 22 years is analyzed. The weights for the flux analysis are based on the optimized power value of depth level three only (3.6–4 m), which is the closest to the reference depth for the flux (3.5 m). For the two scenarios with crop cycles, the flux PDF is based on simulations spanning not just the different hydraulic parameters but also the slightly different boundary conditions as were measured in the field experiment. For the natural conditions, there is no difference between the boundary conditions of the different field plots. The results of the analyses include a flux probability distribution for each scenario.

### 3. Simulation Results and Analysis

In what follows, we present the measurement and analysis results. We start (in subsection 3.1) by presenting some of the measured and simulated WC curves. The measurements from the multiple sensors in the same field are the basis for the hydraulic heterogeneity assessment in our work. The apparent differences between the measurements of the different sensors under the same field (see also Figures S1–S7 in Supporting Information S1) demonstrate the necessity of a probability distribution estimation rather than the estimation of a single (most probable) value. Next, we demonstrate the effects of the weighting process on the predicted distribution of the water flux at a certain depth (Section 3.2). The validation of our method is presented in Section 3.3, where we present the reliability diagrams of our predicted WC PDFs. Finally, we apply our method for the prediction of the water flux PDFs under the experimental agricultural field with different crop cycles (and, therefore, different irrigation regimes; see Section 3.4).
3.1. Water Content Data

Examples of measured and simulated WC curves at the three depth levels of plot 2 (see Section 2 for the description of the field) are depicted in Figure 2. The dotted vertical lines denote the beginning of the period whose WC data were used for weighting the parameter sets and for validation. This analyzed period extends to day 1,000. The light gray shaded area represents the entire prediction range (from all the parameter sets considered here) for each day. The dark gray shaded area represents the 90% confidence interval of the predictions (i.e., not including the highest and lowest 5% of predicted WC).

The solid line represents the measurements in the plot, and the dashed lines represent the simulated WC using the 10 sets with the highest weight as calculated using the data in the analyzed period at each depth level. Similar figures of the other seven field plots are shown in Figures S1–S7 in Supporting Information S1. The highest weight sets in Figure 2 fit well with the arrival time of the main measured wetting front at depth level three and with the high WC values that follow this arrival. The predicted curves by the highest weight sets at depth two in field plot 2, and in additional cases in other field plots, captured the approximate arrival time of the main wetting front but failed to predict the constant high values after this arrival or the following drainage.

3.2. Water Flux Distributions

In order to demonstrate the effect of the weights on the predicted flux distributions, we calculated the PDF of fluxes at a depth of 3.5 m (the simulated results were obtained using an observation node in HYDRUS). The weights were analyzed using power values of \( x = 0, 10, 100, \) and 1,000 (panels a–d of Figure 3, respectively). Using \( x = 0 \) corresponds to no weighting, that is, each parameter set has an equal weight regardless of the goodness-of-fit of the WC curve simulated using that parameter set. A very large value of \( x \) corresponds to assigning all the weight to the best performing sets only, similar to the approach used in inverse modeling. When no weights were used (\( x = 0 \)), the flux distribution is bi-modal, and the most probable flux is from 25 to 50 mm year\(^{-1}\), but with a probability of only \( \approx 0.06 \). The mean of this distribution, however, is 303 mm year\(^{-1}\) with a standard deviation (STD) of 213 (Table 2). These results show that a single value of the most probable flux cannot represent the wide range of possible flux values through the heterogeneous soil. Note that the bi-modality stems from the wide range of hydraulic parameter values that are possible for the soil type in the field (see the Methods section for more details). Increasing the power value \( (x) \), namely, increasing the weights of the sets that better fit the measurements, increases the probability of smaller fluxes (and, obviously, reduces the probability of large fluxes). However, the peak of the distribution shifts to the range of 125–150 mm year\(^{-1}\) for \( x = 100 \). For this power \( (x = 100) \), the distribution becomes narrower with an STD of 183 mm year\(^{-1}\). In Table 2, the mean and STD for different powers of the weighting function are listed. It is apparent that even for a weighting function that is narrow, we still obtain a wide range (the coefficient of variation is not small) due to the fact that different sets with different predictions have similar RMSE values and, therefore, similar weights.

### Table 2

| Panel | Power value \( (x) \) | Mean (mm year\(^{-1}\)) | STD (mm year\(^{-1}\)) |
|-------|----------------------|-------------------------|------------------------|
| a     | 0                    | 303                     | 213                    |
| b     | 10                   | 298                     | 221                    |
| c     | 100                  | 228                     | 183                    |
| d     | 1,000                | 180                     | 116                    |

3.3. Validation

Our validation is based on comparisons between predicted and observed PDFs of WC. The validation results are summarized in Figure 4 in which the predicted and estimated WC probability distributions are shown (left column of Figure 4), together with corresponding reliability diagrams (middle column of Figure 4). The different rows correspond to three depth levels.

The weights were calculated using an optimized power value, \( x \) (Equation 2). The optimization was tested in the ranges of \( 0 < x < 100 \) and \( 0 < x < 200 \),
where the scan intervals were relative to the tested range (see Section 2.6). In all cases, the optimal values were below 100, and therefore, they were determined using the results of the optimization in the 0 < x < 100 range. The optimization curves are shown in the right column of Figure 4, and the objective function values and optimal power values are summarized in Table 3. The predicted flux distributions using the weighting function with the optimal values of x were more reliable than the probability distributions based on equally weighted parameter sets. The optimization resulted in the minimized mean distances in the reliability diagrams of 0.059, 0.052, and 0.107, at depth levels 1, 2, and 3, respectively. These values (and the PDFs and reliability diagrams in Figure 4) show that the predicted distributions captured well the observed distributions. The predicted distributions at depth levels 1 and 3 show some overestimation of the high WC values. This overestimation is the result of the high WC values that were observed in the field plots that were used for weighting (1–4). These high WC values were not observed in the field plots that were used for validation (5–8). The small optimal value of x = 1.47 at depth level 1 indicates that despite the large ensemble of parameter sets, there are no large differences between the weights assigned to each set (a result opposing the underlying assumption of inverse modeling in which a single set of parameters is obtained and used).

**Figure 4.** Summary of the validation process. Left column: predicted and observed water content probability distributions for field plots 5–8. The training data for the weights were taken from field plots 1–4. Middle column: reliability diagrams comparing the histograms on the left column. Right column: Minimization process of the mean distance in the reliability diagrams using power (x) in a range of up to 200 (orange squares) or 100 (light blue triangles). (a–c) indicate the depth levels 1–3, of 150–190 cm, 250–290 cm, and 360–400 cm, respectively. The optimal x values are indicated on the x axes of the right column. The periods from which the data were taken, and the optimal x values at each depth level are provided in Table 3.

| Panel | Depth level | Data period | f(x = 0) | f(x_{optimal}) |
|-------|-------------|-------------|----------|----------------|
| a     | 1           | Days 122–1,000 | 0.071    | f(x = 1.47) = 0.059 |
| b     | 2           | Days 327–1,000 | 0.139    | f(x = 71.33) = 0.052 |
| c     | 3           | Days 513–1,000 | 0.127    | f(x = 27.68) = 0.107 |
3.4. Long-Term Predictions

The distribution of the average water flux (over 22 simulated years) at the depth of 3.5 m was predicted for three scenarios. In order to compare between the different scenarios, we present the PDFs of the average flux and of the ratio between the cumulative flux and the total water input (irrigation and precipitation). The results are presented in Figure 5. The weights of the parameter sets were determined using the data in the analyzed period (see Section 3.1) and the optimized power ($x$) value of depth level 3 (see Section 3.3).

The two scenarios that included crops and irrigation show similar probability distributions of the ratio flux/input, with the main peak around 0.1, and wide distributions of up to $\approx 0.8$ (Figure 5). The mean and STD values of the estimated PDFs for all scenarios are summarized in Table 4. The no-crop scenario distribution has a shorter tail with smaller mean and STD, and with the main peak near zero values. The similarity of the CDFs was tested statistically using the Kolmogorov-Smirnov two-sample test. The test considered the null hypothesis that two specific samples come from the same distribution. Comparing the two irrigated scenarios resulted in a $P$ value of 0.0024. Comparing the non-irrigated scenario to each of the other two scenarios resulted in $P$ values smaller than the machine precision. The resulting $P$ values show that the two irrigated scenarios are not statistically identical, but the difference between them is much smaller than their differences from the non-irrigated scenario. These results suggest that under the climate conditions, irrigation practice, and the specific crops that occurred in the field, the water flux distribution mostly depends on the amount of water input, when it is above a certain threshold due to the irrigation, and less on the specific crop type.

The longer right tail for the irrigated scenarios (in the flux distributions and in the distributions of the ratio flux/input) indicates that an irrigated field is subject to possibly more extreme water percolation than a field with no crops. It has been shown that agricultural recharge is the main component of the

**Table 4**

| Scenario            | Mean  | STD  |
|---------------------|-------|------|
| 1                   | Three crops per 2 years | 0.317 | 0.198 |
| 2                   | One crop per year       | 0.301 | 0.201 |
| 3                   | No crop                | 0.147 | 0.142 |

**Figure 5.** Probability distributions of predicted average water flux (left) and cumulative water flux divided by the total water input (right). The probabilities were generated based on the weight distribution of the optimized setup for depth level 3 (3.6–4 m). Flux was taken at a depth of 3.5 m over 22 years, simulated in three scenarios: (a) a 2-year cycle of three different crops; (b) an annual cycle of a single crop; and (c) no crop and no irrigation. Atmospheric boundary conditions were identical in all three scenarios. Mean and standard deviation for these distributions are detailed in Table 4. The bottom panels show the cumulative distribution functions for the probability densities depicted in the top panels. The dashed vertical lines indicate the location of the quartiles.
total recharge in cultivated (and irrigated) regions (Ebrahimi et al., 2016), which is in agreement with the high values in the flux PDFs of our irrigated scenarios. Moreover, these results suggest that under agricultural fields, the risk of groundwater contamination by nitrate, for example, is not only due to excess fertilization, but also due to the increased probability of high water fluxes that trigger the nitrate leaching.

It is important to note that if a flux prediction would have been created for these scenarios using the common method of a calibrated model using an inverse solution, the resulting flux would have been a single value. In this case, the single flux value would have omitted the rest of the distribution that we show here—obviously, omitting much of the important information included in the flux PDF.

4. Conclusions

We present here a method for estimating the PDF of hydraulic quantities, and used it to derive the PDFs of WC and recharge flux. The PDF includes all the predicted values of WC or flux and their corresponding probability, instead of a single value of the most probable flux, for example, which is obtained from the common method of an inverse solution. The method is based on a comparison of simulation results, using a large ensemble of parameter sets, with multiple measurements under the same field. The heterogeneity of the soil hydraulic properties, at the field scale, is represented by the different sensors.

A unique data set of field measurements was used for implementing our method. This data set consists of measured WC in an unsaturated zone under an agricultural field. The setup included 24 moisture sensors in undisturbed profiles at depths of 1.5–4 m. The measurements were recorded continuously over 900 days of a controlled experiment, including three different irrigated crops.

Our method was validated for each of the three depth levels separately (depth levels of the WC sensors). The validation is based on the reliability diagram of the predicted PDFs for a data set that was not used in the weighting process of the different parameter sets. The reliability diagrams of all three depth levels show reasonable agreement between the predicted and observed PDFs. The weighting function that was optimized based on the reliability diagram of the third depth level was used for the long-term predictions of the flux PDFs.

The comparison of the simulated WC curves with the measured ones revealed that for some simulated curves, the timing of a considerable increase in the WC fits better to the timing of the measured increase. However, if the level of the increase is different from the measured one, the RMSE would still be high. For some applications, the correct prediction of the timing is more important than the specific level. Therefore, future studies may include the arrival time of wetting fronts as a measure of the goodness-of-fit and, thereby, as a factor dictating the weight of the parameter sets.

In the current study, we used WC measurements in order to weigh the different parameter sets. However, in studies using the VMS, there might also be information regarding the chemical composition of the water (e.g., Weissman et al., 2020). Our method may be extended to assign the weights according to the simulated solute concentrations or a combination of the solute concentrations and the WC. In other words, once the weights are obtained using WC and/or concentrations, a distribution of any hydraulic variable (e.g., WC, solute concentration, and flux) at any time and depth can be generated using numerical simulations. This may contribute to more efficient assessments of groundwater contamination risk.

In summary, our approach and results provide a new method to account for the soil heterogeneity and estimate the full PDF of groundwater recharge and its chemical fluxes. The PDF enables the design of groundwater resource management policies that rely on the likelihood of a broad range of scenarios, rather than on the most probable scenario.

Data Availability Statement

Data that was used in this study was uploaded to a Hydroshare repository and can be found in https://doi.org/10.4211/hs.be874910144b4e68b353f93991c78a0d, (Weissman et al., 2022). This repository includes boundary conditions that were used in the simulations and WC measurements.
Turkeltaub, T., Kurtzman, D., & Dahan, O. (2016). Real-time monitoring of nitrate transport in the deep vadose zone under a crop field—Implications for groundwater protection. *Hydrology and Earth System Sciences, 20*(8), 3099–3108. https://doi.org/10.5194/hess-20-3099-2016

Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal, 44*(5), 892–898. https://doi.org/10.2136/ssaj1980.03615995004400000002x

Vesselinov, V. V., & Harp, D. R. (2012). Adaptive hybrid optimization strategy for calibration and parameter estimation of physical process models. *Computers & Geosciences, 49*, 10–20. https://doi.org/10.1016/j.cageo.2012.05.027

Villa, P. M., Sarmiento, L., Rada, F. J., Machado, D., & Rodrigues, A. C. (2017). Leaf area index of potato (*Solanum tuberosum L.*) crop under three nitrogen fertilization treatments. *Agronomia Colombiana, 35*(2), 171–175. https://doi.org/10.15446/agron.colomb.v35n2.62110

Walker, G. R., Currie, D., & Smith, T. (2020). Modelling the effect of efficiency measures and increased irrigation development on groundwater recharge through a deep vadose zone. *Water, 12*(4), 936. https://doi.org/10.3390/w12040936

Wang, W., Zhao, J., & Duan, L. (2021). Simulation of irrigation-induced groundwater recharge in an arid area of China. *Hydrogeology Journal, 29*(2), 525–540. https://doi.org/10.1007/s10040-020-02270-3

Weissman, G., Bel, G., Ben-Gal, A., Yermiyahu, U., Alexandrov, B., Rasmussen, K. O., & Dahan, O. (2020). Increased irrigation water salinity enhances nitrate transport to deep unsaturated soil. *Vadose Zone Journal, 19*(1), 1–16. https://doi.org/10.1002/vzj2.20041

Weissman, G., Dahan, O., & Bel, G. (2022). BGU—VMS water content measurements in Gilat experiment 2016–2018. Hydroshare. https://doi.org/10.4211/hs.be874910144b6e68835f93991c7f8af3

Wu, J., & Zeng, X. (2013). Review of the uncertainty analysis of groundwater numerical simulation. *Chinese Science Bulletin, 58*(25), 3044–3052. https://doi.org/10.1007/s11434-013-5950-8

Xu, T., & Valocchi, A. J. (2015). A Bayesian approach to improved calibration and prediction of groundwater models with structural error. *Water Resources Research, 51*(11), 9290–9311. https://doi.org/10.1002/2015wr017912

Younes, A., Zaouali, J., Kanzari, S., Lehmann, F., & Fahs, M. (2019). Bayesian simultaneous estimation of unsaturated flow and solute transport parameters from a laboratory infiltration experiment. *Water, 11*(8), 1660. https://doi.org/10.3390/w11081660

Zhai, H., & Benson, C. H. (2006). The log-normal distribution for hydraulic conductivity of compacted clays: Two or three parameters? *Geotechnical & Geological Engineering, 24*(5), 1149–1162. https://doi.org/10.1007/s10706-005-1135-9