Multistep optimization of HyPix model for flexible vertical scaling of soil hydraulic parameters

J.A.P. Pollacco [a], J. Fernández-Gálvez [b,a,*], C. Rajanayaka [c], S.C. Zammit [c], P. Ackerer [d], B. Belfort [d], L. Lassabatere [e], R. Angulo-Jaramillo [e], L. Lilburne [a], S. Carrick [a], D.A. Peltzer [a]

[a] *Manaaki Whenua – Landcare Research, Lincoln, New Zealand*
[b] *Department of Regional Geographic Analysis and Physical Geography, University of Granada, Spain*
[c] *National Institute of Water and Atmospheric Research, Christchurch, New Zealand*
[d] *Université de Strasbourg, CNRS/EOST, ITES UMR 7063, Institut Terre et Environnement de Strasbourg, Strasbourg, France*
[e] *Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR5023 LEHNA, Vaulx en Velin, Lyon, France*

ARTICLE INFO

ABSTRACT

Efficient simulation of water-flow processes in the vadose zone is crucial to increase agricultural productivity within environmental limits. This requires deriving detailed soil hydraulic parameters of the soil profile that is highly challenging, particularly for heterogeneous soils. We therefore developed an alternative indirect methodology to calibrate the hydraulic parameters from soil water content time series measured at multiple depths by using the new physically based hydrological model HyPix.

We propose a novel, efficient, multistep optimization algorithm for layered soils that derives an optimal set of hydraulic parameters for a desired number of soil layers. For each selected soil layer, HyPix derives five physical, bimodal, Kosugi hydraulic parameters that describe the soil water retention and hydraulic conductivity by using a novel algorithm that reduces the degree of sensitivity and freedom of the parameters. The optimization algorithm upscales the soil hydraulic parameters by gradually incorporating the soil heterogeneity. This method overcomes the problems associated with optimization of the hydraulic parameters of each layer individually, which leads to poor results because it does not represent the cohesive soil water dynamics across the unsaturated zone.

We tested the method using soil water content measurements at different depths at five heterogeneous experimental sites in New Zealand. We show how the accuracy of the simulated water balance components increases with the number of soil layers. The multistep optimization upscales a detailed, layered profile of soil hydraulic parameters into a model with fewer layers. The methodology developed provides an estimate of the uncertainty of using a reduced number of soil layers. We also show that a pedological description can provide an indication of the minimum soil layers of vertical discretization required to accurately compute the soil water balance components.

1. Introduction

The digital soil database, S-map (https://smap.landcareresearch.co.nz; Lilburne et al., 2004; McNeill et al., 2018), with a coverage of 37% of New Zealand in 2021, provides the Kosugi (1994, 1996) soil water retention $\theta(\psi)$, and the unsaturated hydraulic conductivity $K(\theta)$ (Pollacco et al., 2013b, 2017), for up to six functional horizons to a maximum depth of 1 m, on a 150 m × 150 m spatial grid. This soil hydrological property database is termed 'Smap-Hydro'. However, most hydrological models are not designed to accommodate multiple layers and therefore require soil hydrological characterization at a coarser vertical resolution (e.g., one, two or three layers). Here we determine how to upscale hydraulic parameters to multiple soil layers, replacing a heterogeneous domain by a less heterogenous domain (Vereecken et al., 2007) for which both parameterizations produce similar hydrological responses (*drainage, evapotranspiration, root zone soil water content,* etc.) under certain scaled boundary conditions.

Using inverse modelling to derive soil hydraulic properties has

---

become increasingly popular in the last few decades (e.g., Graham et al., 2018; Wöhling and Vrugt, 2011), making it possible to obtain more representative estimates of soil hydraulic parameters compared with laboratory methods and estimates from pedotransfer methods (e.g., Al-Ashwal et al., 2021). Hydraulic parameters derived from laboratory methods are not always representative of field conditions, and direct methods for determining soil hydraulic parameters require steady-state conditions as well as restrictive initial and boundary conditions. Alternatively, inverse methods combine forward soil water flow models with appropriate optimization algorithms to find the optimal parameter set that minimizes an objective function. Advanced inverse methods combine a physically based numerical model with an algorithm for automatic parameter estimation (e.g., Pollacco, 2005; Pollacco et al., 2013a; Pollacco and Mohanty, 2012) but this becomes more challenging for heterogeneous soils.

Deriving hydraulic parameters of highly heterogeneous soils by inverting solely time series of *soil water content*, θ [L^3 L^{-3}], by using the numerical solution of the Richardson–Richard's equation (RRE, Richardson, 1922; Richards, 1931) as a forward model was attempted by Ritter et al. (2003) to optimize three layers, but they found that to be a challenging, ill-posedness problem (Pollacco et al., 2008b). Ill-posedness problems can be overcome by measuring θ and soil water pressure, ψ [L], simultaneously from multiple depths (Schelle et al., 2012). Scharnagl et al. (2011) derived the hydraulic parameters by inverting θ and using the ROSETTA pedotransfer function (Schaap et al., 2001) to provide additional information about the correlation structure of the predicted parameters that was found to be essential for the effectiveness and robustness of the methodology. Ines and Droogers (2002) successfully inverted the hydraulic parameters by inverting soil water content from two layers and *evapotranspiration* derived from remote sensing. Rezaei et al. (2016) optimized the hydraulic parameters using θ monitoring together with a crop growth model and a soil hydrological model on a 2-layer soil. Qanza et al. (2019) also successfully used inverse estimation of soil hydraulic parameters for four layers but considering null residual θ, θ_r [L^3 L^{-3}]. Over et al. (2015) introduced a hierarchical simulation and modelling framework that allows for inference and validation of the likelihood function in Bayesian inversion of vadose zone hydraulic properties: they inverted multilayer soil hydrological properties using θ observations collected in the uppermost four layers. The main drawback of this work, however, is the increased computational expense of the inversion. To the best of our knowledge, no optimization has been performed considering up to five soil layers. The only work optimizing 25 hydraulic parameters simultaneously is that of Wöhling and Vrugt (2011), but this combines observed θ and ψ from multiple depths.

A variety of upscaling procedures are available in the literature that allow migration of point-scale variables into coarser models (e.g., Abbaspour et al., 1997; Ward et al., 2006; Zhang et al., 2004). A detailed review of the most commonly used upscaling methods can be found in Vereecken et al. (2007). As illustrated in previous methods, it is important to apply an approach that is suitable for the specific needs. The widespread adoption of soil water content sensors in agricultural regions in New Zealand, the availability of site-specific time series data on soil wetting and drying patterns, and evapotranspiration processes provide an opportunity to develop a method for upscaling high-resolution multilayer Smap-Hydro parameters to assist with better parameterisation of coarser models.

To address the challenge of simultaneously optimizing many hydraulic parameters in highly heterogeneous soils, or in the presence of macroporosity, we develop a novel multistep optimization algorithm, which first models the overall flow in the profile (assuming a homogeneous soil) and then gradually introduces heterogeneities to the level required by the hydrological model. This approach is needed because optimizing the hydraulic parameters of each layer separately (for example, by matching time series θ, individually) produces poor results (Kamali and Zand-Parsa, 2016) since it does not represent the overall

soil water dynamics across the unsaturated zone (Pollacco, 2005).

Optimization is generally performed so that the simulated output from a hydrological model correspond to the observed time series θ at different depths (e.g., from time or frequency domain reflectometer). Hence, the soil hydraulic parameters are derived by minimizing an objective function formed from the observed and simulated data (Pollacco, 2005; Pollacco et al., 2008b, 2013a; Pollacco and Mohanty, 2012). Experiments have shown that the objective function has a unique solution but suffers from excessive sensitivity to some parameters (Pollacco et al., 2008a, 2008b; Pollacco and Angulo-Jaramillo, 2009). The methodology proposed by Fernández-Gálvez et al. (2021) and further validated in Vogeler et al. (2021) reduces the sensitivity of the bimodal soil Kosugi hydraulic parameters by preventing non-physical combinations of hydraulic parameters (Pollacco et al., 2008b). The Kosugi (1994, 1996) hydraulic functions are used because their parameters are physically related to the pore size distribution of the soil.

The multistep optimization method is implemented into the new performant **Hy**drological **Pix**el (HyPix) model. HyPix (Pollacco et al., 2022) is a physically-based hydrological model and solves the mixed form of the Richardson–Richard's equation, REE, by using the Newton–Raphson method. The non-linear RRE is solved with an efficient heuristic and physical time-stepping strategy using a reduced number of control parameters. HyPix also incorporates a novel algorithm to avoid 'overshooting' by controlling the Newton–Raphson step. HyPix can **(a)** process a large number of soil layers, **(b)** simulate Kosugi unimodal and bimodal hydraulic parameters for each soil layer, **(c)** simulate realistic water *ponding* at the soil surface by using a novel approach for the computation of *sorptivity* (Lassabatere et al., 2021, 2022), **(d)** compute rainfall *interception* from leaf area index, **(e)** derive *transpiration* from root water uptake, with a compensation mechanism for deeper layers where root density is limited, **(f)** compute *evaporation*, and **(g)** compute *drainage* through the bottom of the soil profile under different boundary conditions.

This paper is organized as follows: section 2 presents the theory describing the bimodal Kosugi hydraulic functions, an overview of the HyPix model, and the novel vertical multistep optimization scheme used to upscale the hydraulic parameters; section 3 describes the experimental data and methodology used to illustrate and validate the proposed scaling method; section 4 shows the results of the derived soil hydraulic and vegetation parameters from inverse modelling; section 5 illustrates some of the direct applications in future research; and section 6 summarizes the main conclusions.

2. Theory

2.1. Soil hydraulic functions

2.1.1. Bimodal Kosugi soil hydraulic functions

HyPix uses the bimodal Kosugi (1994, 1996) soil hydraulic functions. The choice of the Kosugi soil hydraulic functions is based on the physical interpretation of the parameters in relation to the soil pore size distribution and the fact that these parameters can be constrained by exploiting the relationship between them (Fernández-Gálvez et al., 2021; Pollacco et al., 2013b). Moreover, the selection of bimodal functions is based on the prevalence of soils with a bimodal pore system (e.g., Jarvis, 2007; McLeod et al., 2008), where macropores and micropores lead to a two-stage drainage. Fast flow (macropore flow) can occur when the water pressure head exceeds the threshold needed to activate the macropore network, adding to the matrix flow. Below this threshold, only the matrix participates in the flow (Fernández-Gálvez et al., 2021). The representation of the soil water retention curve $\theta(\psi)$ [L^3 L^{-3}] and the unsaturated hydraulic conductivity $K(\psi)$ [L T^{-1}] functions is based on the dual porosity model of Pollacco et al. (2017):

$$
\begin{cases}
\theta(\psi) = \theta_{\mathrm{Mat}}(\psi) + \theta_{\mathrm{Mac}}(\psi) \\[2mm]
\theta_{\mathrm{Mat}}(\psi) = \frac{1}{2}[\theta_{s\mathrm{MacMat}} - \theta_r]\,\mathrm{erfc}\left[\frac{\ln \psi/\psi_m}{\sqrt{2}\,\sigma}\right] + \theta_r \\[3mm]
\theta_{\mathrm{Mac}}(\psi) = \frac{1}{2}[\theta_s - \theta_{s\mathrm{MacMat}}]\,\mathrm{erfc}\left[\frac{\ln \psi/\psi_{m\mathrm{Mac}}}{\sqrt{2}\,\sigma_{\mathrm{Mac}}}\right]
\end{cases}
\tag{1}
$$

(2021) and further validated in Vogeler et al. (2021), which reduces the number of parameters to be optimized without compromising the fit of the hydraulic functions, while the estimated hydraulic parameters still have physical meaning.

This set of constraints can be summarized as follows: θ_r is derived from σ, $\psi_{m\mathrm{Mac}}$ and σ_{Mac}, which are considered constant for a fixed value of ψ_{MacMat} equal to 100 mm, and ψ_m and σ are dynamically constrained, based on the assumption that $\theta(\psi)$ and $K(\theta)$ are lognormally distributed. Therefore, the number of hydraulic parameters to be optimized is

$$
\begin{cases}
K(\psi) = K_{\mathrm{Mat}}(S_e(\psi)) + K_{\mathrm{Mac}}(S_e(\psi)) \\[3mm]
S_e(\psi) = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \frac{1}{2}\left[\frac{\theta_{s\mathrm{MacMat}} - \theta_r}{\theta_s - \theta_r}\,\mathrm{erfc}\left[\frac{\ln \psi/\psi_m}{\sqrt{2}\sigma}\right] + \frac{\theta_s - \theta_{s\mathrm{MacMat}}}{\theta_s - \theta_r}\,\mathrm{erfc}\left[\frac{\ln \psi/\psi_{m\mathrm{Mac}}}{\sqrt{2}\sigma_{\mathrm{Mac}}}\right]\right] \\[3mm]
K_{\mathrm{Mat}}(S_e(\psi)) = K_s \frac{\theta_{s\mathrm{MacMat}} - \theta_r}{\theta_s - \theta_r}\,\sqrt{S_e(\psi)}\left[\frac{1}{2}\,\mathrm{erfc}\left[\frac{\ln \psi/\psi_m}{\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}}\right]\right]^2 \\[3mm]
K_{\mathrm{Mac}}(S_e(\psi)) = K_s \frac{\theta_s - \theta_{s\mathrm{MacMat}}}{\theta_s - \theta_r}\,\sqrt{S_e(\psi)}\left[\frac{1}{2}\,\mathrm{erfc}\left[\frac{\ln \psi/\psi_{m\mathrm{Mac}}}{\sqrt{2}\sigma_{\mathrm{Mac}}} + \frac{\sigma_{\mathrm{Mac}}}{\sqrt{2}}\right]\right]^2
\end{cases}
\tag{2}
$$

where erfc is the complementary error function; θ [L^3 L^{-3}] represents the *volumetric soil water content* and ψ [L] the *soil water pressure*, considering $\psi > 0$ for unsaturated soils (i.e., *matrix suction*); θ_s [L^3 L^{-3}] and θ_r [L^3 L^{-3}] are the *saturated* and *residual volumetric soil water content*, respectively; ln ψ_m and σ [-] denote the mean and standard deviation of ln ψ, respectively, in the soil matrix domain; ln $\psi_{m\mathrm{Mac}}$ and σ_{Mac} [-] denote the mean and standard deviation of ln ψ, respectively, in the macropore soil domain (with the argument of ln in units of length, i.e., ψ_m, ψ, and $\psi_{m\mathrm{Mac}}$ in [L]); $\theta_{s\mathrm{MacMat}}$ [L^3 L^{-3}] is the volumetric saturated water content that theoretically differentiates inter-aggregate pores (structural macropores) and matrix domains (intra-aggregate micropores), defining the corresponding soil water pressure threshold between macropore and matrix ψ_{MacMat} [L]; $S_e(\psi)$ [-]denotes the *effective saturation* as a function of ψ with values between 0 and 1; K_s [L T^{-1}] is the *saturated hydraulic conductivity*; and $K(S_e(\psi))$ [L T^{-1}] refers to the *unsaturated hydraulic conductivity*, written as a function of $S_e(\psi)$. For the case where $\theta_{s\mathrm{MacMat}} = \theta_s$, Eq. (1) and Eq. (2) reduce to the unimodal Kosugi soil hydraulic functions.

2.1.2. Constraining bimodal Kosugi soil hydraulic parameters

Estimating the bimodal Kosugi soil hydraulic parameters from observed θ requires the simultaneous estimation of eight parameters (θ_s, θ_r, σ, ψ_m, K_s, $\theta_{s\mathrm{MacMat}}$, σ_{Mac}, and $\psi_{m\mathrm{Mac}}$) using limited measurement data. Here, we use the term 'optimize' in the estimation of the parameters since the task is to produce a set of values for the parameters that optimizes (typically, minimizes) an objective function. Therefore, optimization is used in the sense of parameter estimation.

In certain cases full inversion of the data to produce parametric estimates of the bimodal Kosugi soil hydraulic model is not possible, for example, where measurements have a restricted range (Pollacco et al., 2008b), leading to highly sensitive parameters (Pollacco et al., 2008b, 2013a; Pollacco and Mohanty, 2012). Reducing the sensitivity of the parameters requires either adding novel independent measurements or incorporating constraints to reduce the effective complexity of the model. Since gathering new measurements is, in this case, not feasible, here we use the set of constraints proposed by Fernández-Gálvez et al.

reduced from eight to five using the principles of soil physics. Although θ_s can be derived from total porosity when soil bulk density and particle density data are available, in this case it is an optimized parameter with a feasible range determined from the maximum observed θ in the corresponding layer. The soil hydraulic parameters are optimized using the bimodal $\theta(\psi)$ and $K(\theta)$ model described in section 2.1.1. This is performed by matching simulated θ time series derived from the HyPix model with observed θ time series at the corresponding depth.

The dynamic, physically feasible range of the five optimized bimodal Kosugi soil hydraulic parameters using the full set of constraints derived from Fernández-Gálvez et al. (2021) is indicated in Table 1. Note that ψ_{MacMat} is a constant, with a value of 100 mm and ln ψ_{MacMat} is higher than ln $\psi_{m\mathrm{Mac}}$ by three times σ_{Mac} ($P_\sigma = 3$); Eq. (3) in Fernández-Gálvez et al. (2021). Here, θ_r is derived from σ in Eq. (3) and Table 1, and is given by:

$$
\begin{cases}
\sigma^* = \frac{\sigma - \sigma_{\min}}{\sigma_{\max} - \sigma_{\min}} \\[3mm]
\theta_r(\sigma) = \theta_{r\mathrm{Max}}\frac{1 - e^{-\alpha_1 \cdot \sigma^{*\alpha_2}}}{1 - e^{-\alpha_1}}
\end{cases}
\tag{3}
$$

where $\theta_{r\mathrm{Max}}$ is set at 0.2, the maximum value for θ_r that was found to be satisfactory (Pollacco et al., 2013b); $\alpha_1 = 15$ and $\alpha_2 = 4$ are two optimized empirical parameters; σ^* [-] is the normalized σ; and $\sigma_{\min}$ [-] and $\sigma_{\max}$ [-] are set at 0.75 and 4 from Table 1. Therefore, with all these simplifications and additional constraints, we define a model that requires only five parameters: θ_s, σ, ψ_m, K_s, and $\theta_{s\mathrm{MacMat}}$.

2.2. Summary of HyPix model

Modelling unsaturated flow in highly heterogeneous soils can accurately be performed by solving the RRE, which is commonly adopted by hydrological and soil vegetation atmosphere transfer models. The HyPix model implements improvements to solve the RRE by including a dynamic physical smoothing criterion for controlling the Newton–Raphson step, and a novel time-stepping management scheme based on ψ without introducing further parameters (Pollacco et al., 2022). The solution of

Table 1

Feasible dynamic range of the optimized bimodal soil Kosugi hydraulic parameters from observed θ using the relationship between σ and ψ_m (Fernández-Gálvez et al., 2021) that dynamically constrained the set of hydraulic parameters. The feasible range of K_s is derived from Carsel and Parrish (1988), $\psi_{MacMat} = 100$ mm, $P_\sigma = 3$, and $\theta_r(\sigma)$ is described in Eq. (3).

	θ_s [m^3 m^{-3}]	θ_r [m^3 m^{-3}]	σ [-]	ψ_m [mm]	K_s [cm h^{-1}]	$\theta_{sMacMat}$ [m^3 m^{-3}]	ψ_{mMac} [mm]	σ_{Mac} [-]
Min	Max(θ)	$\theta_r(\sigma)$	0.75	$\sqrt{\psi_{MacMat}}\,e^{\sigma P_\sigma}$	0.02	$0.75\,\theta_s$	$\sqrt{\psi_{MacMat}}$	$\dfrac{\ln \psi_{MacMat}}{2P_\sigma}$
Max	0.65		4.00	$\psi_{MacMat}\,e^{\sigma P_\sigma}$	30.00	θ_s		

the RRE is based on Maina and Ackerer (2017), for which the RRE partial differential equation is solved using a *cell-centred finite-volume* (implicit finite differences) scheme for the spatial discretization, with an *implicit Euler scheme* for the temporal discretization by using the *weighted average* inter-cell hydraulic conductivity. Assuming a rigid solid matrix, the mixed form of the RRE is written as:

$$\frac{\theta_i(\psi_i^t) - \theta_i(\psi_i^{t-1})}{\Delta T^t} - S_o \frac{\theta_i(\psi_i^t)}{\theta_{s_i}} \frac{\psi_i^{t-1} - \psi_i^t}{\Delta T^t} = \frac{Q_{i-\frac{1}{2}}^t - Q_{i+\frac{1}{2}}^t}{\Delta Z_i} - Sink_i(\psi_i^{t-1}) \quad (4)$$

where ΔT^t [T] is the *time-step* at time t; ΔZ_i [L] is the *mesh size* of cell i, with the vertical coordinate positive downwards; θ_i [L^3 L^{-3}] is the *volumetric soil water content* of cell i; θ_{s_i} [L^3 L^{-3}] is the *saturated volumetric soil water content* of cell i; S_0 [L^{-1}] is a parameter that accounts for *fluid compressibility*, which is assumed to be constant with depth; ψ_i [L] is the *soil water pressure* of cell i, considering $\psi < 0$ for unsaturated soils; Q [L T^{-1}] is the *soil water flux* based on the extended Darcy–Buckingham's law, which is positive downward and negative when water moves upwards; $Q_{i-\frac{1}{2}}^t$ [L T^{-1}] is the flux entering cell i and $Q_{i+\frac{1}{2}}^t$ [L T^{-1}] is the flux exiting cell i; and $Sink_i$ [L^3 L^{-3} T^{-1}], taken as positive, is the *sink* term defined as the volume of water per unit time removed from cell i by soil *evaporation* and *root water uptake*. Additional details of the HyPix model can be found in Pollacco et al. (2022).

2.3. Novel vertical multistep optimization

The aim of this study is to derive a strategy to optimize the five hydraulic parameters (θ_s, σ, ψ_m, K_s, $\theta_{sMacMat}$) of the parsimonious bimodal $\theta(\psi)$ and $K(\theta)$ functions described in section 2.1.1 for each layer of the soil profile by using the HyPix model. Each layer of the profile corresponds to the depths where θ is experimentally measured. For example, in a site where θ values are measured at five depths, 25 (5 × 5) hydraulic parameters need to be optimized.

It is challenging to optimize 25 parameters simultaneously.

However, Pollacco (2005) found that optimizing each layer separately produces poor results, particularly for a highly heterogeneous soil profile in the presence of lenses of clay, pebbles, or macropore flow. This is because an optimization method that isolates each layer without considering the overall water flow in the soil profile causes unrepresentative parameters, and thus poor representation of the soil water fluxes (Kamali and Zand-Parsa, 2016). Therefore, we present an inverse modelling algorithm for layered soils.

When optimizing the hydraulic parameters, the soil profile is initially considered homogeneous. Then, a stepwise grouping of local layers (zones defined by the end-user) allows heterogeneous patterns to be addressed. The optimization of the different layers in a specified order and pattern is presented in Table 2 (Table 2a for odd and Table 2b for even number of layers), where 0 or 1 indicates which soil layers are optimized simultaneously at each specific step. For example, layers containing the number 1 in Table 2 show the grouping of different layers

Table 3

Relative errors, ζ, computed between *reference* (greatest number of layers) and *simulated* (reduced number of layers): *drainage at the bottom of the soil profile*, *evapotranspiration*, and *soil water content of the root zone* (top 600 mm).

	Relative error [%]	
Drainage	$\zeta_Q = \dfrac{\left\| \sum_{t=1}^{t=N_t} \Delta T^t Q_{sim_{N_z+\frac{1}{2}}}^t - \sum_{t=1}^{t=N_t} \Delta T^t Q_{ref_{N_z}}^t \right\|}{\sum_{t=1}^{t=N} \Delta T^t Q_{ref_{N_z}}^t}$	Eq. (5)
Evapotranspiration	$\begin{cases} \Delta Et = \Delta Evap + \Delta Rwu \\ \zeta_{et} = \dfrac{\left\| \sum_{t=1}^{t=N_t} \Delta Et_{sim}^t - \sum_{t=1}^{t=N_t} \Delta Et_{ref}^t \right\|}{\sum_{t=1}^{t=N_t} \Delta Et_{ref}^t} \end{cases}$	Eq. (6)
Root zone θ	$\zeta_{swc} = \dfrac{\left\| \sum_{t=1}^{N_t} \sum_{i=1}^{Z_{Nroot}} \theta_{sim_i}^t - \sum_{t=1}^{N_t} \sum_{i=1}^{Z_{Nroot}} \theta_{ref_i}^t \right\|}{\sum_{t=1}^{N_t} \sum_{i=1}^{Z_{Nroot}} \theta_{ref_i}^t}$	Eq. (7)

Table 2

The vertical multistep optimization method for an even number of measurement depths is described, where the order in which the optimization is performed for the different groups of layers, iL: Table 2a for five and Table 2b for four measurement depths. The layers that correspond to 0 are the cells that keep their values from the previous optimization steps. The *N layer splitting* 'mimics' the number of measurement depths.

Table 2a. Odd measurement depths: 5

Layers	Opt_1	Opt_2	Opt_3	Opt_4	Opt_5	Opt_6	Opt_7	Opt_8	Opt_9
iL	1	2a	2b	3a	3b	4a	4b	5a	5b
1	1	1	0	1	0	0	0	0	0
2	1	1	0	0	1	1	0	0	0
3	1	1	0	0	1	0	1	0	0
4	1	0	1	0	0	0	0	1	0
5	1	0	1	0	0	0	0	0	1

Table 2b. Even measurement depths: 4

Layers	Opt_1	Opt_2	Opt_3	Opt_4	Opt_5	Opt_6	Opt_7
iL	1	2a	2b	3a	3b	4a	4b
1	1	1	0	1	0	0	0
2	1	1	0	0	1	0	0
3	1	0	1	0	0	1	0
4	1	0	1	0	0	0	1

(zones) in which the soil hydraulic parameters have the same optimal value (i.e., homogeneous layer).

In the first step (*Opt_1*) it is assumed that the soil is homogeneous, and therefore the whole profile is modelled with five optimized soil hydraulic parameters, and the same values were given for each parameter in the different layers of the soil profile. The derived effective optimal hydraulic parameters will be used for hydrological models requiring only one homogeneous layer. In the second step (*Opt_2*) (*optimal hydraulic parameters for models requiring two layers*), only the parameters of the upper half of the profile are optimized, maintaining the bottom half (below the root zone) with the value of optimized parameters derived from the previous step; and then the third step (*Opt_3*) (*optimal hydraulic parameters for models requiring three layers, etc.*) operates on the deeper zone, keeping the upper half of the profile with the values of the parameters with values of the previously optimized parameters.

Optimizing from top to bottom was observed to produce better results than from bottom to top, because water percolates downwards, so a change of the hydraulic parameters of the top layer will affect the lower layer. Note that the top layer is also the layer where the water content has larger variations with time. Each zone of the profile is successively split into two zones from top to bottom, and the optimization is repeated by copying the values of the optimal parameters from the previous optimization step. The vertical multistep optimization assumes that the number of soil layers, *iL*, corresponds to the number of θ observations. The centre of each soil layer is also given by the corresponding θ observation.

2.3.1. Evaluation of improvements in accuracy by increasing the number of layers

The procedure for identifying the relative errors between optimization performed with the greatest number of layers and simulated with a reduced number of layers is presented in Table 3. It is assumed that outputs from the HyPix model computed with the greatest number of layers, *iL*, is the reference (free of error), *ref*, and simulations computed with a reduced number of layers are simulated, *sim*. The reduction of error in different steps of computation (a) *drainage* [Eq. (5)], (b) *evapotranspiration* [Eq. (6)] and (c) *root zone soil water content* (top 600 mm) [Eq. (7)] is possible by increasing the number of layers, *iL*, as presented in Table 2.

2.3.2. Weighted objective functions of the multistep optimization

The global optimizer searches for the optimal hydraulic parameters by minimizing a *weighted objective function* (*WOF*). We selected the robust global optimizer BlackBoxOptim v0.6.1 (https://github.com/robertfeldt/BlackBoxOptim.jl) written in Julia (Bezanson et al., 2017) and selected the *adaptive_de_rand_1_bin_radiuslimited* method. For every optimization step, the maximum evaluation functions that are allowed to be optimized are controlled by *MaxFuncEvals*, which was set to 100.

We designed the novel *WOF* to address the issue that observed θ, θ_{obs} [L^3 L^{-3}], in deeper layers (below the root zone), contains greater uncertainties compared to measurements in the root zone. It is also important that simulated θ, θ_{sim} [L^3 L^{-3}], gives more accurate predictions in the root zone than below the root zone. This is because most of the dynamism of soil evaporation and root water uptake takes place in the root zone. Prioritizing the root zone by optimizing from top to bottom is performed in the algorithm used by the vertical multistep optimization (section 2.3). This is performed by introducing a weighting, *W*, in the *WOF* by assuming that the measurements placed at different depths are equally spaced. The *WOF* is computed as follows:

$$\begin{cases} W_{iL} = 2\dfrac{N_{iL} + 1 - iL}{N_{iL}(N_{iL} + 1)} \\[2mm] \sum_{iL=1}^{N_{iL}} W_{iL} = 1 \\[2mm] WOF_\theta = \sqrt{\dfrac{\sum_t^{N_t} \sum_{iL}^{N_{iL}} W_{iL} \left| \theta_{obs_{iL}}^t - \theta_{sim_{iL}}^t \right|^2}{N_{iL} N_t}} \end{cases} \tag{8}$$

where N_{iL} is the total number of layers, *iL*, where θ is measured; N_t is the total number of time-steps; and θ_{obs} and θ_{sim} are observed and simulated θ, respectively.

3. Material and methods

3.1. Data

3.1.1. Monitoring periods for calibration and validation

The monitoring periods of the θ at each depth for the five experimental sites were split in two, using one period for model calibration and another for model validation. The dates of calibration and validation for each site were selected to consider the different installation periods, sensor stabilization after installation, erroneous data, data gaps, and the occasional presence of the groundwater table in deeper layers at some sites. The calibration and validation periods were chosen to include a wide variety of dry and wet periods, and simulations were always preceded by a 2-month warm-up period. There were not enough data for validation for one of the sites (*Waihou*) due to malfunctioning of the sensors.


Fig. 1. Location of the experimental sites in the North Island of New Zealand.

3.1.2. Climate

Daily values of *precipitation*, ΔPr [L], were measured adjacent to the soil water content probes using a tipping bucket electronic rain gauge, recording at 0.2 mm resolution. *Potential evapotranspiration*, ΔPet [L], was derived from the New Zealand Virtual Climate Stations network (VCS) (Tait and Woods, 2007) with estimates based on the spatial interpolation of actual data observations made at climate stations located around the country (Tait, 2010). For a given variable X, we used the symbol ΔX to indicate the change in the variable X in a given time step as follows, $\sum_{t=1}^{T} X^t - \sum_{t=1}^{T-1} X^t$, e.g., ΔPr is the cumulative precipitation between two-time steps. The yearly average *precipitation* and *potential evapotranspiration* using representative wet and dry periods are shown in the Results section (Table 5) by using HyPix model.

3.1.3. Soils

The sites used in this study are located in the Waikato region of New Zealand (Fig. 1). All soils have formed from airfall volcanic tephra, but vary in their soil physical properties and heterogeneity, particularly their texture and profile drainage characteristics. A brief description of the soils as follows:

Well-drained soils

The *Taupō* soil is a sandy-textured soil formed from volcanic airfall pumice material (New Zealand classification: *Podzolic Orthic Pumice Soil;* USDA Soil Taxonomy classification: Orthod (Hewitt, 2010)). The *Otorohanga* and *Waihou* soils are also formed from airfall volcanic material, but with finer tephra material compared with *Taupō*, resulting in silty loam topsoil textures grading to silty clay in the subsoil. They are classified in New Zealand as Typic Orthic Allophanic Soils, and in Soil Taxonomy as a Haplohumult (Hewitt, 2010). All soils are characterized by well-drained morphology, having yellow-brown coloured subsoils with no redox mottles, indicating that the frequency and duration of internal waterlogging is minimal and oxidation processes predominate. Based on the soil morphology, the *Otorohanga* and *Waihou* soils would be considered to have the least heterogeneous soil profiles of those used in this study, since they have a reduced number of distinct layers. The *Taupō* soil, although free draining, is expected to show more heterogeneity in water movement due to the stone content (Cichota et al., 2016).

Imperfectly and poorly drained soils

The *Hamilton* soil has a silt loam topsoil overlying clayey, textured subsoils, having formed into strongly weathered volcanic tephra. It is classified in New Zealand as a Typic Orthic Granular Soil, and in Soil Taxonomy as a Haplohumult (Hewitt, 2010). The drainage class indicates a degree of subsoil drainage impediment that may result in short-term waterlogging during wet periods, although oxidation processes predominate most of the time. The *Waitoa* soil is a silty-textured soil. It is classified in New Zealand as a Typic Orthic Gley Soil, and in Soil Taxonomy as a Haplohumult (Hewitt, 2010). The drainage class indicates a slow-draining subsoil, with subsoil saturation occurring most years in winter and early spring.

The soil morphology of the *Hamilton* and *Waitoa* shows the greatest heterogeneity of the soils in this study, as reflected in previous studies of soils with impeded drainage features (McLeod et al., 2008; Vogeler et al., 2019).

3.1.3.1. Time series of soil water content: θ. θ measurements were recorded at 15 min intervals with *Decagon* 5TM *frequency domain reflectometers* located at up to five depths for each of the experimental sites. The monitoring depths for the different sites differ corresponding to the soil morphological features at each individual site. Both the rain gauges and the θ sensors were calibrated following the New Zealand standard (Duncan et al., 2013). In particular, the θ were calibrated against neutron probe measurements taken at particular times throughout the monitoring period. The neutron probe data were calibrated with θ measurements from sample weighting *in situ* and after oven drying for each depth and site.

3.1.4. Vegetation

Experimental sites are five mixed, non-irrigated pasture grass sites. The trapezoidal water stress response function (Feddes et al., 1978) was used with parameters for mixed pasture grass in New Zealand derived from Van Housen (2015). For all sites we used default values at four soil water pressures: $\psi_{Feddes1} = 100$ mm, $\psi_{Feddes2} = 250$ mm, $\psi_{Feddes3} = 5000$ mm, and $\psi_{Feddes4} = 80,000$ mm. These parameters are not as sensitive as the crop coefficient, K_C, which we optimized (Pollacco et al., 2008a).

The K_C, and leaf area index, LAI, vary throughout the growing season. Due to the lack of K_C and LAI data it was assumed that there is a positive linear relationship between LAI and pasture growth, as already suggested by many authors (e.g., Kaur and Singh, 2013; Punalekar et al., 2018), and a positive linear relationship between K_C and pasture growth. Allen and Pereira (2009) showed that K_C is a function of the *fraction of ground cover* and *crop height*, which is also related to pasture growth. Pasture growth (kg DryMatter/ha/day) data were provided by DairyNZ for the Waikato region (https://www.dairynz.co.nz/media/5793237/average-pasture-growth-data-waikato-2020-v1.pdf) with the following normalized monthly values [0–1] from January to December: 0.48, 0.24, 0.15, 0.23, 0.34, 0.14, 0.00, 0.23, 0.65, 1.00, 0.79, and 0.72. The range of K_C, 0.8 to 0.95, was taken from rotated grazing pasture according to FAO irrigation paper 56 (Allen et al., 1998), and the range of the LAI, 0.19 to 5.10, was taken from Van Housen (2015). Normalized LAI and K_C are calculated as:

$$\text{normalised} value = \frac{value - value_{\min}}{value_{\max} - value_{\min}} \tag{9}$$

where the corresponding value of the vegetation parameters, $LAI_{\min}$, $LAI_{\max}$, $K_{C\min}$, and $K_{C\max}$, as well as the maximum saturated storage capacity of a wet canopy ($Sint_{sat}$ [L]), are optimized after multistep *Opt_1* and reoptimized after multistep *Opt_9* or *Opt_7*, depending on whether the number of soil layers is five or four. Then the multistep optimization is rerun with the updated vegetation parameters. For all sites, maximum root depth, Z_{Nroot}, was averaged to 800 mm (Vogeler and Cichota, 2019), and the percentage of roots in the top 300 mm, Rdf_{top}, was 90% (Evans, 1978).

3.2. HyPix model parameters

The vertical discretization of the soil profile has 37 cells for soils with five layers and 25 cells for soils with four layers. Depth between 0 and 500 mm is discretized with cells of mesh size $\Delta Z = 20$ mm, and depth between 500 and 1200 mm is discretized with cells of mesh size $\Delta Z = 50$ mm. The vertical discretization is the same as the one used in Pollacco et al. (2022). Simulations use the standard values of the HyPix model parameters, as shown in Table 4.

Table 4
Standard values for the HyPix model parameters.

$\Delta\psi_{active}$ [mm]	$\Delta T_{\min}$ [s]	$\Delta T_{\max}$ [s]	$P_{\Delta T\_Rerun}$ [-]	N_k [-]	$WB_{residual}$ [s^{-1}]	$\Delta\theta_{\max}$ [L^3 L^{-3}]	$\Omega_{\min}$ [-]
1	30	3600	1.5	70	10^{-10}	$6. \, 10^{-3}$	0.2

3.3. Goodness of fit

The goodness of fit between the model outputs (θ_{sim}) and the corresponding observations (θ_{obs}) was assessed using the best concordance correlation coefficient, *CCC*, and the refined index of agreement, *dr*, proposed by Willmott et al. (2012).

The *CCC*, is defined as follows:

$$CCC = \frac{2\rho\sigma_{\theta_{obs}}\sigma_{\theta_{sim}}}{\sigma^2_{\theta_{obs}} + \sigma^2_{\theta_{sim}} + \left(\mu_{\theta_{obs}} - \mu_{\theta_{sim}}\right)^2} \tag{10}$$

where $\mu_{\theta_{obs}}$ and $\mu_{\theta_{sim}}$ are the observed and simulated means of θ at each depth where θ is measured, $\sigma^2_{\theta_{obs}}$ and $\sigma^2_{\theta_{sim}}$ are the observed and simulated variances of θ at each depth where θ is measured, and ρ is the Pearson correlation coefficient between observed and simulated values. *CCC* is equal to 1 for a perfect fit.

The *dr* index, with values between -1.0 and 1.0, indicates the sum of the magnitudes of the differences between the model-predicted and observed deviations about the observed mean relative to the sum of the magnitudes of the perfect-model ($\theta_{obs} = \theta_{sim}$ for all i) and observed deviations about the observed mean. This index is used because, in general, it is more rationally related to model accuracy than other existing indices (Willmott et al., 2012). The *dr* index can be written as follows with $c = 2$.

$$dr = \begin{cases} 1 - \dfrac{\sum_{i=1}^{N_t}|\theta_{sim_i} - \theta_{obs_i}|}{c\sum_{i=1}^{N_t}|\theta_{obs_i} - \overline{\theta_{obs}}|}, & \text{when} \quad \sum_{i=1}^{N_t}|\theta_{sim_i} - \theta_{obs_i}| \leqslant c\sum_{i=1}^{N_t}|\theta_{obs_i} - \overline{\theta_{obs}}| \\[2em] \dfrac{c\sum_{i=1}^{N_t}|\theta_{obs_i} - \overline{\theta_{obs}}|}{\sum_{i=1}^{N_t}|\theta_{sim_i} - \theta_{obs_i}|} - 1, & \text{when} \quad \sum_{i=1}^{N_t}|\theta_{sim_i} - \theta_{obs_i}| > c\sum_{i=1}^{N_t}|\theta_{obs_i} - \overline{\theta_{obs}}| \end{cases} \tag{11}$$

A value for *dr* of 0.5, for example, indicates that the sum of the error magnitudes is one-half of the sum of the perfect-model-deviation and observed-deviation magnitudes. When $dr = 0.0$, it signifies that the sum of the magnitudes of the errors and the sum of the perfect model-deviation and observed-deviation magnitudes are equivalent. When $dr = -0.5$, it indicates that the sum of the error magnitudes is twice the sum of the perfect-model-deviation and observed-deviation magnitudes. Values of *dr* near -1.0 can mean that the model-estimated deviations about $\overline{\theta_{obs}}$ are poor estimates of the observed deviations; but they can also mean that the observed variability is insignificant. When approaching the lower limit of *dr*, interpretation should be made cautiously.

4. Results

4.1. Hydraulic and vegetation parameters derived from inverting observed θ

4.1.1. Water balance components

The rainfall interception algorithm in HyPix (Pollacco et al., 2022) uses vegetation parameters described in section 3.1.4. The *interception loss* is computed for the five pasture grass sites in Table 5, ranging from 8 to 20%, with annual *throughfall precipitation* values corresponding to 2298 and 823 mm, respectively. As expected, the *interception loss* is greater for sites with smaller annual rainfall (*Waitoa* and *Waihou*) compared to sites with a larger annual rainfall (*Otorohanga* and *Taupō*), because the saturated storage capacity, $Sint_{sat}$, of the interception model has greater impact on smaller rainfall events than on large ones.

Simulated yearly drainage values are lower at sites with lower precipitation, while simulated yearly evaporation values show a distinctive behaviour at each site. The closure of the simulated water balance is computed by the normalized soil water balance to the infiltration (Error *WB** in Table 5), with very small discrepancies between $4.4 \cdot 10^{-6}$ and $0.2 \cdot 10^{-6}\%$. The results of interception loss between 8 and 20% are similar to values derived by Thurow et al. (1987) for *curly mesquite* (short grass) and *sideoats* (mid-grass) dominated sites, with 10.8 and 18.1% interception loss, respectively. Note that interception varies monthly with *LAI* and K_C (section 3.1.4). The goodness of fit between θ_{obs} and θ_{sim} at all depths for each site is represented by the WOF_θ values in Table 5, with satisfactory results between 0.009 and 0.014.

Table 6 shows the optimal values of vegetation parameters described in HyPix (Pollacco et al., 2022) for the five experimental sites. Due to the limited vegetation information available it was not possible to fully validate the rainfall interception model and its vegetation parameters, as an excellent fit between observed and simulated θ does not guarantee that the vegetation parameters be physical plausible. Pollacco et al. (2008a) showed that when the vegetation parameters are optimized against observed θ, assuming the soil hydraulic parameters are known (which is not the case here), the outputs of the *drainage*, *transpiration*, and *interception* are particularly sensitive to small uncertainties of the θ data, and therefore the optimized vegetation parameters suffer from correlated parameters. The reason for having correlated parameters (Pollacco et al., 2008a, 2008b; Pollacco and Angulo-Jaramillo, 2009) is that the *rainfall interception* and the *evapotranspiration* parameters are linearly linked, which means that *interception* compensates for over/under-predicting *evapotranspiration*, and vice versa. If the water balance between *interception* and *evapotranspiration* is inadequate, then the *drainage* will compensate without affecting the global water balance.

4.1.2. Model calibration and validation

A monitoring period, long enough to include a variety of wet and dry

Table 6
Optimal values of vegetation parameters for the five experimental sites. Minimum and maximum values of leaf area index (LAI_{min} and LAI_{max}) and crop coefficient (K_{Cmin} and K_{Cmax}) as well as the maximum saturated storage capacity of a wet canopy ($Sint_{sat}$).

Site	LAI_{min} [-]	LAI_{max} [-]	K_{Cmin} [-]	K_{Cmax} [-]	$Sint_{sat}$ [mm]
Waitoa	0.369	3.440	0.755	1.077	1.978
Waihou	3.000	4.547	0.788	0.965	1.781
Taupō	2.108	3.906	0.760	0.805	1.288
Otorohanga	0.153	3.405	0.785	1.198	1.747
Hamilton	2.918	4.949	0.754	0.910	1.358

Table 5
Yearly average values of precipitation (ΔPr) and potential evapotranspiration (ΔPet) together with water balance components from HyPix model, using the calibration periods with representative wet and dry periods, of simulated throughfall precipitation ($\Delta Pr_{through}$), computed percentage interception loss, simulated drainage ($\Delta Drainage$), and simulated evapotranspiration ($\Delta Sink$) for the five experimental sites. The last two columns list the computed relative error for the normalized soil water balance to the infiltration (Error *WB**), and goodness of fit between θ_{obs} and θ_{sim} at all depths for each site (WOF_θ).

Site	ΔPr [mm]	ΔPet [mm]	$\Delta Pr_{through}$ [mm]	*Interception loss* [%]	$\Delta Drainage$ [mm]	$\Delta Sink$ [mm]	Error *WB** [%]	WOF_θ [mm³mm⁻³]
Waitoa	1028	910	823	20	436	430	$4.4 \cdot 10^{-6}$	0.011
Waihou	1217	761	1099	10	628	433	$0.2 \cdot 10^{-6}$	0.009
Taupō	1646	773	1489	10	1015	423	$0.7 \cdot 10^{-6}$	0.009
Otorohanga	1818	875	1581	14	1162	430	$2.4 \cdot 10^{-6}$	0.009
Hamilton	2487	413	2298	8	2020	100	$0.2 \cdot 10^{-6}$	0.014

Table 7

Optimal bimodal Kosugi soil hydraulic parameters described in Section 2.1, for each layer derived by inverting observed θ using the multistep optimization algorithm (Section 2.3) at the five experimental sites. The optimization methodology uses the relationship between σ and ψ_m that dynamically constrained the set of hydraulic parameters (Fernández-Gálvez et al., 2021), $\psi_{m\text{Mac}}$ and σ_{Mac} are derived from $\psi_{\text{MacMat}} = 100$ mm and considering $P_\sigma = 3$ as shown in Table 1, and θ_r is derived from σ as described in Eq. (3). Therefore, only five hydraulic parameters (θ_s, σ, ψ_m, K_s, $\theta_{s\text{MacMat}}$) of the bimodal Kosugi hydraulic functions are optimized, while the other three hydraulic parameters (θ_r, σ_{Mac}, and $\psi_{m\text{Mac}}$) are derived from alternative data.

Site	Layer	θ_s [m^3 m^{-3}]	σ [-]	ψ_m [mm]	K_s [cm h^{-1}]	$\theta_{s\text{MacMat}}$ [m^3 m^{-3}]	θ_r [m^3 m^{-3}]	σ_{Mac} [-]	$\psi_{m\text{Mac}}$ [mm]
Waitoa	1	0.551	2.706	14982.671	21.776	0.444	0.180	0.768	10
	2	0.464	2.528	33521.144	7.387	0.438	0.158	0.768	10
	3	0.544	3.521	17626.676	26.432	0.489	0.200	0.768	10
	4	0.531	1.404	59135.804	4.878	0.427	0.006	0.768	10
	5	0.526	2.773	20451.083	1.567	0.459	0.186	0.768	10
Waihou	1	0.476	2.656	13528.521	15.174	0.440	0.175	0.768	10
	2	0.534	2.074	61953.129	17.031	0.400	0.077	0.768	10
	3	0.534	2.074	61953.129	17.031	0.400	0.077	0.768	10
	4	0.508	3.456	55820.530	26.147	0.419	0.200	0.768	10
	5	0.477	3.356	9262.989	11.005	0.419	0.200	0.768	10
Taupō	1	0.560	2.626	10086.399	7.867	0.465	0.171	0.768	10
	2	0.560	2.626	10086.399	7.867	0.465	0.171	0.768	10
	3	0.510	2.443	8113.736	7.255	0.505	0.145	0.768	10
	4	0.510	2.443	8113.736	7.255	0.505	0.145	0.768	10
Otorohanga	1	0.564	2.287	40152.950	13.797	0.464	0.117	0.768	10
	2	0.528	1.465	47219.519	19.946	0.430	0.008	0.768	10
	3	0.493	2.063	49484.511	24.435	0.445	0.075	0.768	10
	4	0.521	1.485	39769.927	24.344	0.434	0.009	0.768	10
	5	0.518	2.352	13620.086	11.866	0.513	0.129	0.768	10
Hamilton	1	0.460	3.823	268865.502	1.647	0.423	0.200	0.768	10
	2	0.613	3.137	1923546.614	23.685	0.494	0.199	0.768	10
	3	0.663	1.151	2662049.694	13.860	0.539	0.001	0.768	10
	4	0.634	2.608	4510797.576	10.480	0.580	0.169	0.768	10

periods at each site, was used for calibrating the soil hydraulic properties for each layer derived from inverse modelling by using the multistep optimization algorithm implemented in HyPix (section 2.3). The optimal bimodal Kosugi soil hydraulic parameters from inverse modelling by using the multistep optimization algorithm with observed θ at the five experimental sites are shown in Table 7. The selection of the calibration period is crucial to provide an accurate estimate of the hydraulic parameters, and, due to malfunctioning of the *Waihou* site, there were not enough data left for validating this site. After successfully automatically calibrating the hydraulic and vegetation parameters at each site, HyPix was run for a validation period using the optimal hydraulic and vegetation parameters obtained from calibration.

Simulations corresponding to the calibration and validation periods are shown in Fig. 2, where time series of precipitation ΔPr, throughfall precipitation $\Delta PrThrough$, ponding $\Delta Hpond$, drainage ΔQ, potential evapotranspiration ΔPet, sink term $\Delta Sink$, and evaporation $\Delta Evap$ are presented above the θ values at depths corresponding to measurements for the five experimental sites. Note that the *intercepted precipitation* varies monthly with *LAI* and K_C (section 3.1.4). Simulated values of θ at different measuring depths were found to be close to the observations and follow the general trend according to the forcing input data, with larger changes in θ near the surface and less responsive values deeper in the soil profile. *Rainfall* events are followed by a sudden increase in θ that gets reflected progressively down into the soil profile. After these episodes, θ decreases progressively, with an approximately simultaneous increase in both simulated *drainage* and *evapotranspiration*. A sudden increase in *evapotranspiration* values occurs because after each *rainfall* event the amount of water available next to the soil surface is greater. *Drainage* does not take place immediately after a *rainfall* event: there is a delay in the *drainage* pulse, with differences related to the amount of water reaching the surface during each *rainfall* event and the hydraulic characteristics of each experimental site.

The goodness of fit between measured and simulated θ for the calibration and validation periods at each of the experimental sites using the dynamically constrained set of hydraulic parameters is shown in Table 8. A better fit is obtained for the first four sites (*Waitoa, Waihou, Taupō,* and *Otorohanga*), with higher $CCC \geq 0.743$ and $dr \geq 0.666$ values compared to the ones for the *Hamilton* site with $CCC = 0.596$ and $dr >$

0.533. The goodness of fit for validation is only marginally lower than for calibration, providing confidence in the retrieved hydraulic and vegetation parameters implemented in HyPix for each site. Only for the *Hamilton* site, the *dr* index is reduced by half for the validation period due to both the higher soil heterogeneity and the shorter monitoring period with available data. In fact, the *Hamilton* site, with its slow-draining subsoil and a high soil heterogeneity, is the most challenging site studied which also be due to errors in the climate data.

One of the challenges addressed by the automatic multistep optimization algorithm (section 2.3) and the weighted objective function [Eq. (8)] is to predict the root zone θ more accurately and ensure the predictions below the root zone remain acceptable. This is because the water balance is mainly determined in the root zone where the sink term is active. This is challenging, since greater uncertainties are found in measuring θ in the root zone due to presence of roots and macropores, which can be seen in the *Otorohanga* site (Fig. 2) where θ_{sim} modelled at 100 mm is more accurate during the validation period compared with the calibration period. This may be because the time required for the top sensor to settle down is greater than for the other sensors (contact between the sensor and the surrounding soil).

4.1.3. Multistep optimization for vertical scaling

The improvement of the fit between observed and simulated θ with increasing splitting of the soil profile (layers), as described in Table 2a (five depths of measuring θ) and Table 2b (four depths of measuring θ), depends on the:

a) heterogeneity of the soil profile;
b) aptitude of the model to represent the hydrological processes, which is quantified by the success of the RRE to reproduce θ_{obs};
c) accuracy of θ_{obs}.

The maximum number of splits of the soil profile required by a hydrological model to obtain the desired accuracy in a particular output is difficult to establish, since it does not depend solely on the severity of heterogeneity in the soil, but also on the required accuracy of the outputs of the model. Table 9 shows the minimum number of layers required to achieve a particular accuracy in *drainage, evapotranspiration,*

Calibration

Validation
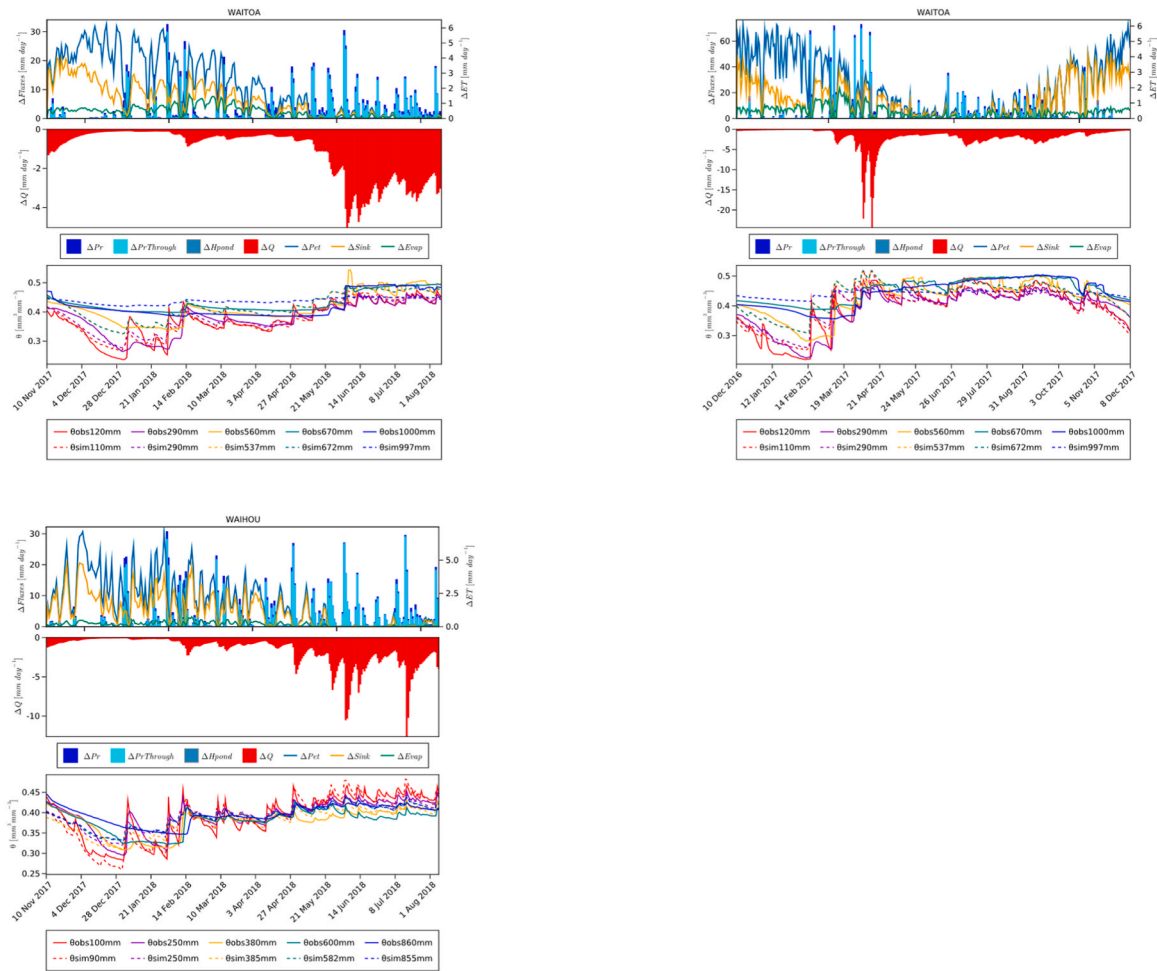
Fig. 2. Simulated water balance components with HyPix for the experimental sites and simulated and measured θ during the calibration and validation periods for the experimental sites. As explained in the text, there was no sufficient data to validate Waihou site. Plots include ΔPr precipitation, throughfall precipitation $\Delta PrThrough$, ponding $\Delta Hpond$, drainage ΔQ, potential evapotranspiration ΔPet, sink term $\Delta Sink$, evaporation $\Delta Evap$, and measured (continuous line) and simulated (dashed line) θ at depths corresponding to measurements for each site.

and *soil water content of the root zone* for each of the experimental sites by assuming that outputs from the HyPix model computed with the greatest number of layers, *iL*, is the reference (free of error), as indicated in Table 3. The selected accuracy indicated in Table 9 for each model output is chosen as an example, and the measure of the accuracy is evaluated by the relative errors described in Table 3.

Fig. 3a shows the decrease of goodness of fit between observed and simulated θ, represented by the WOF_θ (Eq. (8)), with the increasing amount of splitting of the soil profile using the multistep optimization scheme described in section 2.3 for each site. As can be observed, the decrease is less pronounced for less heterogenous soils, with the higher impact for the most heterogeneous site. Fig. 3b–d illustrate the results corresponding to the algorithms, which compute the percentage errors (Table 3) as a function of the amount of splitting for (a) *drainage*, ζ_Q, (b) *evapotranspiration*, ζ_{et}, and (c) *soil water content of the root zone* (top 600 mm), ζ_{SWC}. In the evaluation of the errors, the discretization with the greatest amount of splitting of the profile (greatest number of layers) was considered as the reference.

The minimum splitting needed to achieve a particular goal is described in Table 9. As previously indicated, Table 9 shows that the minimum number of required layers depends on the output of interest.

Clearly, accurate computation of *evapotranspiration* requires the greatest number of layers, compared to *drainage* and *soil water content in the root zone*. This is explained by the fact that *evapotranspiration* is influenced by the movements of water from deeper layers to the surface and therefore requires accurate layering along the profile. The dynamics of *evapotranspiration* and *soil water content* occur only in the root zone, and therefore the impact of having accurate hydraulic information at greater depths is less important. The fact that small errors in θ cause large errors in *drainage* and *evapotranspiration* was already reported by Pollacco et al. (2008a) and Pollacco and Mohanty (2012), which is why in Table 9 the threshold of ζ_{SWC} is 2.5% and not 5%. Whilst these results are indicative of the effects of reducing the number of layers, providing guidance on the minimum number of layers to be used for modelling at the catchment or region level would require a larger number of experimental sites.

4.1.3.1. Time of execution of multistep optimization. The yearly average execution time for each site of performing the multistep optimization by running HyPix model is described in Table 10. HyPix is run with Julia 1.7.3 on a laptop equipped with 64 GB RAM and *Intel® Xeon® E-2186M*
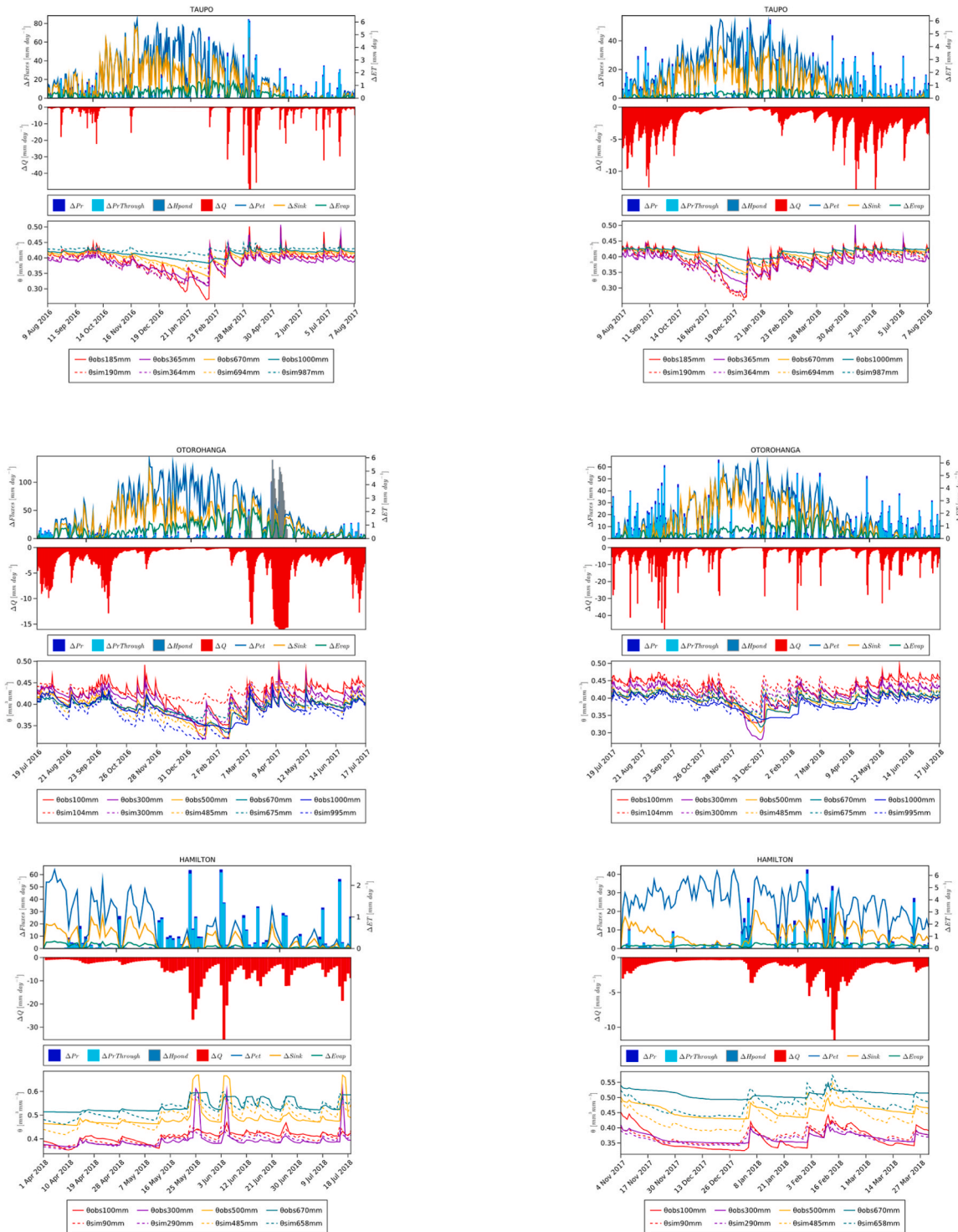
Fig. 2. (*continued*).

CPU @ 2.9 GHz processor. To facilitate inter comparison the simulations are scaled to a period of 1 year. Table 10 shows that there are differences in execution time which are largely due to the heterogeneity of the soil layers (Pollacco et al., 2022) which influences the efforts of the RRE to converge.

4.1.4. Modelled heterogeneity compared to pedological description
   Table 11 classifies the result reported in Table 9 and shows the

minimum number of layers to accurately model each site. The classification derived from inverse modelling is compared to the soil characteristics derived from the soil database S-map. Despite the limited number of experimental sites included in this study, Table 11 shows there is a clear relationship between the number of layers required for adequate representation of the soil hydraulic parameters and the heterogeneous soil description derived from S-map. Both the soil heterogeneity and the drainage class from S-map can provide an indication of

Table 8

Goodness of fit for the profile soil water content evaluated by the best concordance correlation coefficient, *CCC* [Eq. (10)], and the refined index of agreement, *dr* [Eq. (11)], between model outputs and observations for the calibration and validation periods for each of the experimental sites.

Site	Calibration		Validation	
	CCC	*dr*	*CCC*	*dr*
Waitoa	0.753	0.705	0.699	0.673
Waihou	0.743	0.670	–	–
Taupō	0.819	0.666	0.791	0.594
Otorohanga	0.785	0.691	0.605	0.649
Hamilton	0.596	0.533	0.505	0.262

Table 9

Minimum number of layers for each site required to achieve a certain accuracy in (a) *drainage*, ζ_Q, (b) *evapotranspiration*, ζ_{et}, and (c) *soil water content of the root zone*, by assuming that outputs from the HyPix model computed with the greatest number of layers, *iL*, is the reference (free of error) as indicated in Table 3. ζ are described in Table 3.

Site	$\zeta_Q < 5\%$	$\zeta_{et} < 5\%$	$\zeta_{SWC} < 2.5\%$
Waitoa	4	9	6
Waihou	4	4	1
Taupō	2	6	1
Otorohanga	3	9	3
Hamilton	3	7	5

the vertical discretization required to accurately compute the soil water balance components. For example, well-drained soils can generally be adequately described by a relatively small number of layers. By contrast, poor or *imperfectly drained* soils require more complex descriptions with more layers.

5. Future research

Next steps in implementing this approach to scaling Smap-Hydro information include a sensitivity analysis of deriving hydraulic parameters by using the vertical multistep optimization assuming fixed vegetation parameters, assessing the value of remotely sensed sources of surface soil water content and transpiration, and testing the scaled hydrological parameterizations in the various hydrological models used in New Zealand.

Another application of the multistep optimization is that observed θ derived at multiple depths (e.g., from time or frequency domain reflectometers) is becoming readily available in precision agriculture in New Zealand to control irrigation rates (Drewry et al., 2019; Ekanayake and Hedley, 2018; El-Naggar et al., 2020). Therefore, the proposed methodology of deriving unique sets of hydraulic parameters from

Table 10

Yearly execution time: **(a)** for single HyPix run, **(b)** per step described in Tables 2a and b, and **(c)** for the full multistep optimization.

Site	Soil layers [-]	HyPix run [s]	Per Step [h]	Multistep Optimization [h]
Waitoa	5	63	1.9	17
Waihou	5	55	1.7	15
Taupō	4	31	1.1	7
Otorohanga	5	22	0.8	7
Hamilton	4	29	1.0	7
Average	–	40	1.3	11
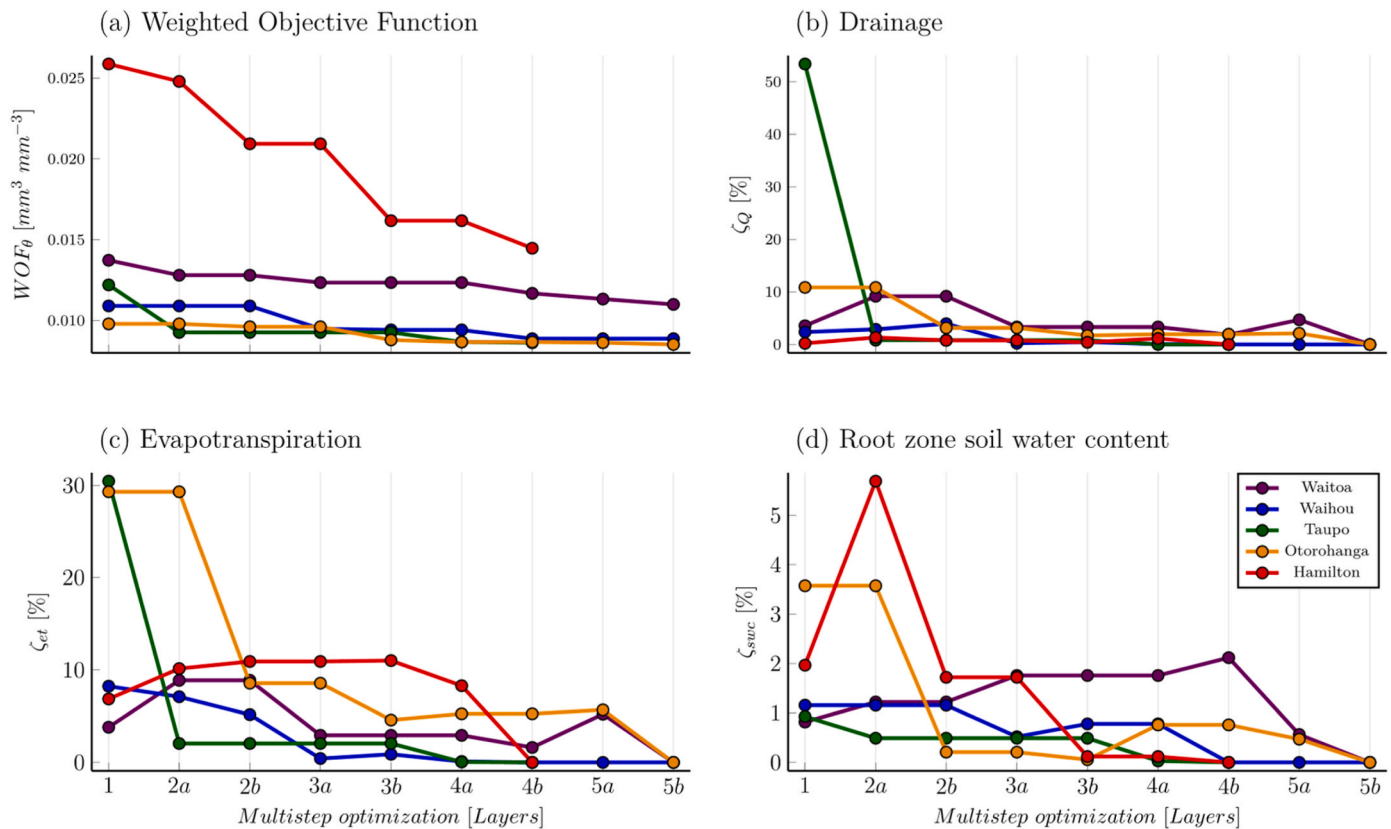
Fig. 3. Influence of splitting of the soil profile, where the multistep optimization in the X-axis is described in Tables 2a and b, in the **(a)** goodness of fit between observed and simulated θ, and percentage errors, as described in Table 3 for computed **(b)** *drainage*, **(c)** *evapotranspiration*, and **(d)** *soil water content of the root zone* (top 600 mm) using the greatest splitting of the profile (greatest number of layers, considered as the reference) and computed values using a reduced splitting (reduced number of layers from 1 to 5 (*Waitoa, Otorohanga, Waihou*) or 4 (*Taupō, Hamilton*)).

Table 11

Results of the modelling compared with the soil characteristics for each site. The *number of layers for drainage* is interpreted from (Table 8, ζ_Q), the *number of layers for evapotranspiration* is interpreted from (Table 8, ζ_{et}), and the layer heterogeneity is interpreted from (Table 8, ζ_{SWC}). The *soil heterogeneity* and *drainage class* are taken directly from the soil description in section 3.1.3.

Site	Multistep optimization			Soil description	
	Number of layers for drainage	Number of layers for evapotransp.	Layer heterogeneity	Soil heterogeneity	Drainage class
Waitoa	High	High	High	High	Poorly drained
Waihou	High	Moderate	Low	Low	Well drained
Taupō	Low	Moderate	Low	Moderate	Well drained
Otorohanga	Moderate	High	Moderate	Low	Well drained
Hamilton	Moderate	High	High	High	Imperfectly drained

observed θ time series would enable physically-based models to optimize the application of irrigation and minimize leaching through excess drainage.

6. Conclusions

We derived a novel vertical multistep algorithm that enables us to invert observed θ at multiple depths by fitting only five bimodal Kosugi hydraulic parameters to every layer and optimizing the hydraulic parameters in a particular pattern from top to bottom by successively splitting the profile. The multistep optimization algorithm puts more weight into optimizing the top layers because most of the dynamism of the soil evaporation and root water uptake takes place in the root zone. The optimization algorithm upscales the soil hydraulic parameters gradually, introducing heterogeneity, because optimizing the hydraulic parameters of each layer individually produces poor results since it does not represent the overall soil water dynamics across the unsaturated zone. It has been shown that the minimum number of layers to accurately estimate *drainage*, *evapotranspiration*, and *soil water content of the root zone* computed from the HyPix model depends on the heterogeneity of the soil and the required accuracy for each output of the model. It was found that a more detailed, layered model is required when computing *evapotranspiration* compared to *drainage*. The multistep optimization is useful to upscale a detailed layered profile of soil hydraulic parameters into a model with fewer layers and provide an estimate of the uncertainty due to the reduced number of layers. Moreover, the soil description from S-map provides an indication of the vertical discretization required to accurately compute the soil water balance components.

It was found that the Richards-Richardson equation converges faster in HyPix when the hydraulic parameters describing $\theta(\psi)$ and $K(\theta)$ are plausible. This is performed during the optimization process by dynamically constraining the relationship between ψ_m and σ. The HyPix model efficiently solves the mixed form of the Richards' equation using a *cell-centred, finite-volume* scheme for the spatial discretization, with an *implicit Euler scheme* for the temporal discretization, by using the *weighted average* inter-cell hydraulic conductivity. HyPix includes *rainfall interception*, *soil evaporation*, *root water uptake* with compensation algorithm, and *ponding* using the novel computation of sorptivity.

Software availability

The HyPix model can be downloaded from https://github.com/manaakiwhenua/SoilWater_ToolBox/tree/master/src/HyPix and is open source under the GP-3.0 License. This software is part of a set of interlinked modules implemented into the SoilWater-ToolBox ecosystem led by J.A.P Pollacco from Manaaki Whenua – Landcare Research in New-Zealand and J. Fernández-Gálvez from the University of Granada in Spain. The preliminary objectives of the SoilWater-ToolBox are to derive the soil hydraulic parameters by using a wide range of cost-effective methods. The estimated hydraulic parameters can be directly implemented into HyPix to compute the soil water budget. The SoilWater-ToolBox enables performing inter comparison and sensitivity analyses of the hydraulic parameters computed from different methods on the

soil water fluxes. The following modules are currently included in the SoilWater-ToolBox:

- **Intergranular Mixing Particle size distribution module**: derives unimodal hydraulic parameters by using particle size distribution (Pollacco et al., 2020);
- **General Beerkan Estimation of Soil Transfer parameters module**: derives the unimodal hydraulic parameters from single ring infiltration experiments (Fernández-Gálvez et al., 2019);
- **Sorptivity module**: a novel computation of sorptivity used in the General Beerkan Estimation of Soil Transfer parameters method (Lassabatere et al., 2021, 2022);
- **Saturated hydraulic conductivity module derived from unimodal and bimodal θ(ψ)** (Pollacco et al., 2013b, 2017);
- **Inverse module which inverts hydraulic parameters from θ time series**, as in this paper
- **Reduce uniqueness module of a physical bimodal soil Kosugi hydraulic parameters** from inverse modelling (Fernández-Gálvez et al., 2021) using water retention and/or unsaturated hydraulic conductivity data directly measured in the laboratory or indirectly obtained from inverting θ time series, as described in this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

Abbaspour, K.C., Van Genuchten, M.T., Schulin, R., Schläppi, E., 1997. A sequential uncertainty domain inverse procedure for estimating subsurface flow and transport parameters. Water Resour. Res. 33, 1879–1892. https://doi.org/10.1029/97WR01230.

Al-Ashwal, H.H., Alsanabani, M.M., AlMunqedhi, B.M., Gaafar, A.-R.Z., El-Sheikh, M.A., 2021. Using field experiments to estimate soil hydraulic parameters via inverse solution evaluating estimation by Hydrus-1d with other methods. Fresenius Environ. Bull. 30, 12951–12962.

Allen, R.G., Pereira, L.S., 2009. Estimating crop coefficients from fraction of ground cover and height. Irrigat. Sci. 28, 17–34. https://doi.org/10.1007/s00271-009-0182-z.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO Rome 300, D05109.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2017. Julia: a fresh approach to numerical computing. SIAM Rev. 59, 65–98. https://doi.org/10.1137/141000671.

Ward, A.L., Zhang, Z.F., Gee, G.W., 2006. Upscaling unsaturated hydraulic parameters for flow through heterogeneous anisotropic sediments. Adv. Water Resour. 29, 268–280. https://doi.org/10.1016/j.advwatres.2005.02.013.

Willmott, C.J., Robeson, S.M., Matsuura, K., 2012. A refined index of model performance. Int. J. Climatol. 32, 2088–2094. https://doi.org/10.1002/joc.2419.

Wöhling, T., Vrugt, J.A., 2011. Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data. Water Resour. Res. 47 https://doi.org/10.1029/2010WR009265.

Zhang, Z.F., Ward, A.L., Gee, G.W., 2004. A combined parameter scaling and inverse technique to upscale the unsaturated hydraulic parameters for heterogeneous soils. Water Resour. Res. 40 https://doi.org/10.1029/2003WR002925.