Journal of Hydrology: Regional Studies 47 (2023) 101399

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Evaluating recharge estimates based on groundwater head from different lumped models in Europe

I.K. Seidenfaden[a,*], M. Mansour[b], Hélène Bessiere[d], David Pulido-Velazquez[h], A. Højberg[a], K. Atanaskovic Samolov[i], L. Baena-Ruiz[h], H. Bishop[k], B. Dessì[e], K. Hinsby[a], N.H. Hunter Williams[j], O. Larva[f], L. Martarelli[e], R. Mowbray[k], A.J. Nielsen[a], J. Öhman[g], T. Petrovic Pantic[i], A. Stroj[f], P. van der Keur[a], W. J. Zaadnoordijk[c,l]

[a] *Geological Survey of Denmark and Greenland, GEUS – Øster Voldgade 10, 1350 Copenhagen K, Denmark*
[b] *British Geological Survey, BGS – Keyworth, Nottingham NG12 5GG, United Kingdom*
[c] *Geological Survey of the Netherlands, TNO – Princetonlaan 6, 3584 CB Utrecht, the Netherlands*
[d] *óFrench Geological Survey, BGRM - 3 avenue Claude-Guillemin, BP 36009, 45060 CEDEX 02 Orléans, France*
[e] *Geological Survey of Italy, ISPRA - Via Vitaliano Brancati 48, 00144 Rome, Italy*
[f] *Croatian Geological Survey, HGI-CGS - Milana Sachsa 2, 10 000 Zagreb, Croatia*
[g] *Geosigma, S:t Persgatan 6, 753 20 Uppsala, Sweden*
[h] *Spanish Geological Survey, IGME-CSIC – Urb. Alcázar del Genil, 4-Edif. Zulema, Bajo. 18006, Granada, Spain*
[i] *Geological Survey of Serbia, Rovinjska 12, 11000 Belgrade, Serbia*
[j] *Geological Survey Ireland, Dept. of the Environment, Climate and Communications, Booterstown Hall, Blackrock, Co Dublin, A94 N2R6, Ireland*
[k] *Geological Survey Ireland and GDG Ltd, Rathfarnham, Dublin D14 X627, Ireland*
[l] *Delft University of Technology, Faculty of Civil Engineering and Geosciences, Water Resources Section, Stevinweg 1, 2628 CN Delft, the Netherlands*

A R T I C L E I N F O

Keywords:
Groundwater recharge
Lumped modelling
Time series analysis
Europe

A B S T R A C T

Study region: The study uses 78 groundwater head time series across 10 European countries with various geological and hydrological settings.
Study focus: The estimation of groundwater recharge using time series analysis and lumped modelling based on groundwater head time series is a low-cost and practical method. However, lumped recharge estimation models based on groundwater level variations are uncertain, and successful applications are known to depend on both climate and hydrogeological setting. Here, we assess the suitability of three different models to estimate recharge (Metran - Transfer Function-Noise model, AquiMod - groundwater level driven hydrological model, and GARDÉNIA - lumped catchment model).
New hydrological insights: Results showed that while all three models generally did well during the modelling of groundwater heads, the resulting recharge estimations from the models were different. The analysis showed that the transfer-noise modelling of groundwater heads with recharge and evapotranspiration in Metran is not generally applicable for recharge estimation. The addition of physical information in AquiMod improved the recharge estimations, but the reliability was still limited without control of the water balance due to non-uniqueness. By adding discharge information to the modelling, GARDÉNIA can provide more reliable recharge values. Thus, recharge estimation from groundwater head time series without water balance information

* Corresponding author.
E-mail address: ika@geus.dk (I.K. Seidenfaden).

https://doi.org/10.1016/j.ejrh.2023.101399
Received 12 November 2022; Received in revised form 18 April 2023; Accepted 24 April 2023

must be considered uncertain with low precision, but applicability can be improved when including knowledge of the local system.

1. Introduction

Knowledge about groundwater recharge and potential changes in a future climate is important for water resource management and the assessment of a sustainable use of groundwater (e.g., Alley et al., 2002; Margat and van der Gun, 2013; Pulido-Velazquez et al., 2020). This includes assessments of the water balance, water availability, contaminant transport, nutrient loadings, and groundwater quantitative and chemical status according to European directives and guidance (European Commission, 2000, 2006, 2009). These aspects are all important to the implementation of the UN sustainable development goals and to improve our understanding of the limits to the human use of groundwater without causing ecological damage (Steffen et al., 2015; Gleeson et al., 2020).

Measuring and estimating groundwater recharge is not straightforward (Alley et al., 2002). A large variety of methods exists for recharge assessment (Custodio, 2019), e.g., in situ lysimeters, water budget calculations (including streamflow separation techniques), analysis of groundwater head time series (such as water table fluctuation methods, lumped models or transfer noise modelling), unsaturated zone modelling, and integrated hydrological modelling (Scanlon et al., 2002; Healy, 2010; Sibanda et al., 2009).

Because groundwater head time series are widely available, analysis of groundwater head time series for recharge estimation has been applied extensively in the literature (e.g. Healy and Cooke, 2002). The methods often required very little data and can thus be setup in countries and locations where data coverages are lacking or insufficient for more data intense methods (e.g., integrated modelling approaches), while being very fast to create and deploy. A fairly simple method is the water table fluctuation method (e.g., Crosbie et al., 2005), which requires information on the specific yield, which may not be known accurately. Other methods using groundwater heads can take into account boundary conditions and other processes (e.g., abstraction) that may affect the groundwater levels (e.g. Peterson and Fulton, 2019). Uncertainties in recharge estimations based on heads are, however, often substantial (Delin et al., 2007; Nimmo et al., 2015) if head are influenced by other factors.

While recharge is highly dependent on climate, it is also dependent on surface and sub-surface conditions. Thus, the recharge estimate will depend not only on the input data, but also on the recharge conditions and mechanisms, and whether the selected model can represent these mechanisms satisfactorily (Lafare et al., 2021). de Vries and Simmers (2002) recommend that water resources practitioners identify important features that influence recharge for a given location first and then apply a recharge calculation model that is suitable to actual field conditions. It is possible that some model types may not be applicable under certain conditions or settings, and therefore both governing features for recharge creation and appropriate model type may be difficult to immediately identify. Therefore, (de Vries and Simmers, 2002) recommend applying and comparing multiple independent approaches to the same case study.

The focus of this paper is to investigate the potential use of recharge estimates from three structurally different groundwater head-driven models across different climatic, hydrological, and geological settings in Europe. Two of these models have been chosen as they have been used for seasonal groundwater head forecasts (e.g. Prudhomme et al., 2017) and can be potentially used, therefore, to undertake recharge forecasts. These models are the groundwater level-driven hydrological model, AquiMod (Mackay et al., 2014a), and the lumped catchment model, GARDÉNIA (Thiéry, 2015a). In order to apply a third different model type, it was chosen to test the recharge produced by a Transfer Noise model, Metran (Zaadnoordijk et al., 2019), because this type of model is widely applied and it has been suggested that it can provide useful recharge estimates (Collenteur et al., 2021; Obergfell et al., 2019). The objective of this paper is to investigate if the successful model calibration and simulation of observed groundwater head fluctuations are sufficient indication of reasonable groundwater recharge estimates, and to explore possible limitations of their use. We focus on diffuse recharge from precipitation and present a taxonomy based on literature nomenclature to classify the calculated values to enable a proper comparison. Due to lack of in-situ measurements of recharge at the sites to verify or reject the recharge estimations, the head-driven model estimates are compared to recharge generated by other methods, e.g., local and national hydrological or integrated models and a machine learning-based European Recharge map (Martinsen et al., 2022).

The head-driven models employ contrasting approaches to assess recharge using groundwater heads, precipitation, and potential evapotranspiration as input and optionally surface water discharge and groundwater abstractions. The concept of recharge considered by the various tools and approaches is not always the same, making a comparison of results problematic (Flint et al., 2002; Barthel et al., 2021). Therefore, this study also presents a taxonomy based on literature nomenclature for classifying the recharge estimates and identifies the different recharge types estimated by the applied tools. It should be noted that the models estimate diffuse recharge, occurring from infiltrating excess precipitation through the soil and unsaturated zone, thus this recharge type is the focus here, and return irrigation and repellency of soils on flat surfaces are not considered.

The paper is structured as follows; first is the method Section (2.), including an overall description of the study sites across Europe (2.1), followed by a description of the recharge taxonomy used in the paper (2.2), and the description of the models applied (2.3). The next section describes the calibration of and the experiences from four selected case studies (3.1), followed by the overall results of all the case studies investigated in the paper (3.2 and 3.3). The final sections of the paper include the discussion (4.) and conclusions (5.). The discussion and conclusions will be focused on first, a comparison of the model's recharge estimations and performance; and secondly, identifying limitations and recommendations for models and approaches. This work has been conducted by a group of European Geological Surveys during the GeoERA TACTIC project (TACTIC, 2021b).

2. Materials and methods

2.1. Study sites

The study sites are distributed across Europe (Fig. 1), where time series of groundwater heads in 56 monitoring wells are included. Nine of these are multi-screen wells and have groundwater head time series from multiple depths, yielding a total of 78 time series for the analysis. The sites have very different characteristics in terms of climate, geology, and hydrology; an overview of the overall characteristics can be seen in Table S1 in the electronic supplementary material (ESM). The groundwater head time series have been chosen based on the availability of adequate data in terms of resolution, length, data gaps, as well as representativity of the regional aquifer system (instead of e.g. local pumping or surface water) and availability of precipitation and evapotranspiration data. No actual evapotranspiration was available, and it was assumed that differences in determination of evapotranspiration values are accounted for by the models in the translation to actual evapotranspiration and do not influence the estimated recharge.

The study sites span a range of climatological settings. Each site has been assigned to a climatic zone and together they cover five out of the six climatic zones proposed by Schneider et al. (2013). The Mediterranean group (Climatic zone - M, Table S1) is characterized by high summer temperatures and mild winters; often, precipitation is predominantly during winter. The four study sites in this group have a very high potential evapotranspiration (>1000 mm/y), with examples of both high (Italy) and low precipitation areas (Spain). The next group is the temperate oceanic (Climatic zone – TO). Here, the proximity to the coast means the precipitation is high for most of the year, with mild winters and warm summers. This group consists of the largest set of sites, 47 in total. The sites generally have precipitation between 800 and 900 mm/y and experience potential evapotranspiration in the order of 600 mm/y. The temperate zone situated away from the oceans is the temperate continental zone (Climatic zone – TC), which generally has a larger contrast between winter and summer temperatures and precipitation occurring mainly during wintertime. The four study sites located here

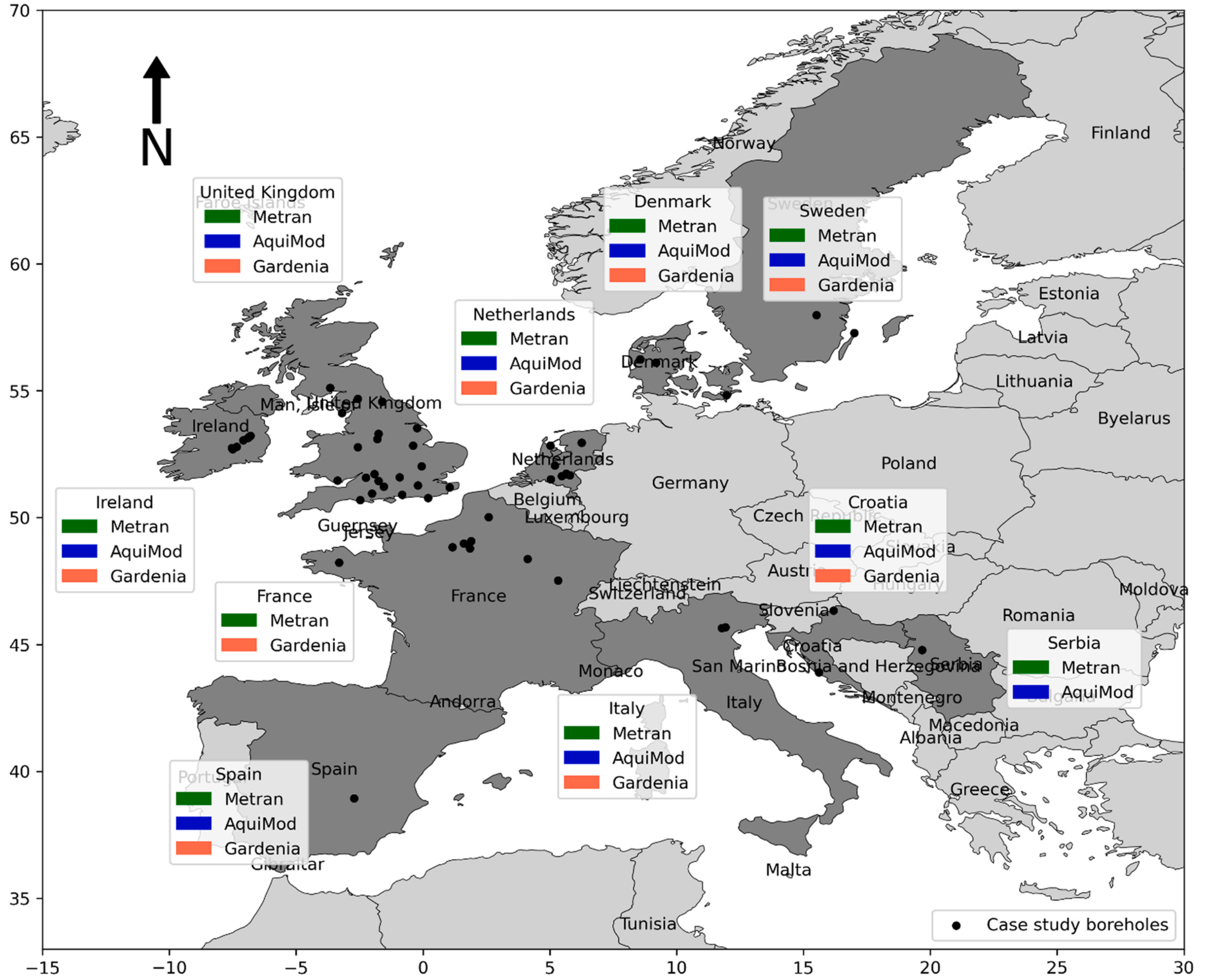

Fig. 1. Overview of study site locations, boxes indicate models used at least at one borehole at the location. Black dots indicate borehole location. UK borehole locations: British Geological Survey © UKRI.

have lower precipitation than in the oceanic area, with precipitation around 600–700 mm/y with one outlier of 850 mm/y (Drava Mura). This group generally experiences potential evapotranspiration in the order of 700 mm/y, with over 900 mm/y for Serbia. One study site lies in between these two temperate zones, in the temperate transitional zone (Climatic zone – TT), and this zone is a mix of the preceding two zones – i.e., with high seasonal contrast but precipitation occurring during all seasons.

Geologically, the aquifers at the sites are also very diverse (Table S1). Most of the aquifers in this study are sedimentary aquifers comprising of porous media like silt, sand and gravel or fissured sedimentary aquifers. This accounts for 25 of the 56 sites. Four sites are alluvial aquifers with sand or gravel media, and eight represent karstic environments with limestone. Observations from both unconfined and confined groundwater settings are present among the 78 groundwater head time series, with a prevalence of unconfined times series (48 out of 78). The hydrology of the sites is generally not very affected by surface runoff, tile drainage or rivers (~20% of the sites are affected by one or more of the three). For the sites where stream discharge data are available, it was included in the modelling for ten site and not used for two sites, where the applied tool does not use discharge or it is not relevant to include. Groundwater abstractions are present at only 6 sites and are always included in the modelling.

2.2. Definition of recharge types

In a broad sense, groundwater recharge is water that is added to the saturated zone (Meinzer, 1923). More specifically, recharge can be defined as the downward flow of water reaching the water table adding to groundwater storage as suggested by e.g., Healy (2010a, 2010b). More loosely, recharge is also sometimes defined as the downward flux of water towards the water table (e.g. Healy, 2010a, 2010b and Fitts, 2013). It is often divided into diffuse recharge from precipitation or irrigation and localized recharge from surface waters and wadis (Scanlon et al., 2002). We focus on diffuse recharge from precipitation.

Throughout the vertical migration of water in the soil column, multiple subsurface processes influence the downward flux of water, e.g., by storage change or evapotranspiration, and change the amount of water accessible for further infiltration. Thus, the recharge will differ depending on where in the subsurface the infiltrating water is considered as recharge, and this taxonomy may therefore be unclear or imprecise (Staudinger et al., 2019). None of these definitions include flow between groundwater systems, where one aquifer may provide recharge to another.

To facilitate the interpretation and comparison of the recharge results from the tools used in this study and other sources, this paper distinguishes three types of recharge based on the specific location in the hydrogeological system, and linked to the water flow process involved before or at that location: potential recharge, recharge (based on the definition by Healy, 2010a, 2010b) and inter-aquifer leakage. The three types are illustrated in Fig. 2 with the associated fluxes in the subsurface. The vertical (green) arrows are the different recharge types, decreasing if storage increases and water moves away with lateral flow.

The red box symbolizes losses near the surface due to processes such as surface runoff, evapotranspiration (including root uptake, interception, and soil evaporation) affecting the soil moisture deficit. Together with precipitation, these losses determine the potential recharge leaving the root zone. Potential recharge is equal to the water surplus arising from precipitation and irrigation after actual evapotranspiration and reducing the soil moisture deficit and surface runoff to surface water systems. Potential recharge has many terms in the literature and is also called percolation, net infiltration or drainage (see e.g., Hendrickx and Walker, 1997).

Some of this potential recharge will be captured by the unsaturated zone storage, restricting soil layers and interflow. The remainder of the potential recharge reaches the groundwater table and becomes part of the saturated zone. This recharge type is simply referred to as recharge. The recharge is thus the gross amount of water that recharges the upper phreatic aquifer through the water table.

In a groundwater aquifer system with one or more confining layers, there may be recharge to confined aquifer due to downward flow through an overlying confining aquitard. This recharge type is termed inter-aquifer leakage, and will have a different hydraulic

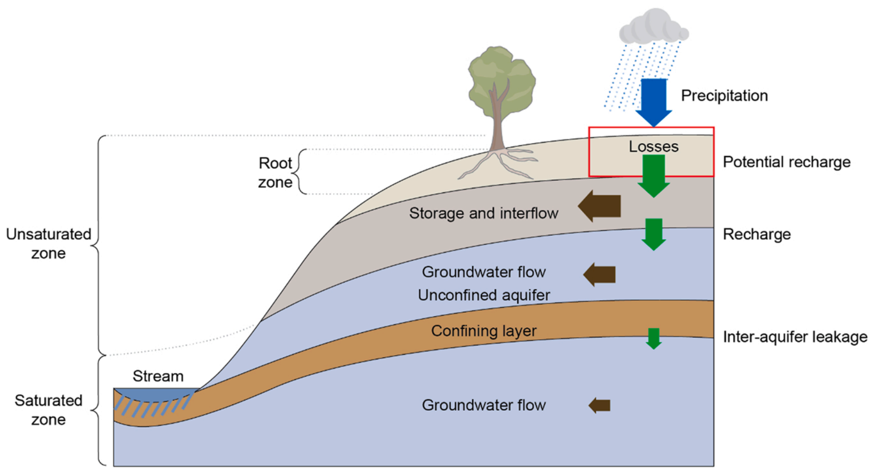

Fig. 2. Definition of the three recharge types, (potential recharge, recharge and inter-aquifer leakage) used in this study.

head response than the recharge type defined above. This type covers all aquifers in the system that are underneath an aquitard and are not phreatic.

In a multi-aquifer system, there may be multiple recharge locations (in case of perched conditions) and multiple inter-aquifer leakages (in the case of multiple confined aquifers)In that case it is necessary to specify the depth the recharge is associated with.

Depending on the spatial resolution, the schematisation of Fig. 2 does not capture all small-scale processes, especially small surface water bodies (e.g. streams or gullies) or tile drainage. In that case, infiltrating water may be drained to surface water. Depending on the method of evaluation it may then be impossible to separate the related fluxes from the surface runoff (for potential recharge) or the interflow (for recharge).

2.3. Recharge estimation models

Three models are applied in this study to estimate the recharges: Metran, AquiMod and GARDÉNIA. In Table 1 a general overview is shown for the three models; these time series or zero-dimensional models are also introduced in subsections 2.3.1, 2.3.2, and 2.3.3, respectively. Table 2 provides a summary of the conceptual representation of processes included in these models together with the number of parameters to be identified during calibration. Recharge estimates from these models are compared to results of measurements or models already present at the sites; therefore, a brief overview of the methods is included in Section 2.3.4.

2.3.1. Transfer noise (Metran)

Metran is a software program for modelling groundwater head time series using a Transfer Function-Noise approach with precipitation and reference evapotranspiration data (Zaadnoordijk et al., 2019). The precipitation transfer function is a three-parameter function based on the gamma distribution, which is a unimodal function with a wide range of possible shapes (see the parameters in Table 2). The transfer function for the reference evapotranspiration is the same, except for a multiplication factor, referred to as the evapotranspiration factor (f_e). The noise-model is an autoregressive model with a single parameter. The five resulting parameters of the applied Metran models (total size of precipitation response, two shape parameters of the precipitation response, evapotranspiration factor and decay parameter of the noise model) have no direct physical meaning that can be used to evaluate the calibration.

Metran runs on a webserver to provide models explaining the variation of groundwater heads in the Dutch national database (http://www.grondwaterstandeninbeeld.nl). It can also be used offline and with different explanatory variables and time series in various formats.

Obergfell et al. (2019) proposed a recharge (R) estimate based on such a time series model which is equal to the rainfall (N) minus the reference evapotranspiration (E) multiplied by the evapotranspiration factor (f_e):

$$R = N - (f_e * E) \quad (1)$$

This was found to give a reasonable estimate of the long-term average recharge in an area with a phreatic aquifer without surface runoff and interflow (Obergfell et al., 2019). In that case, the potential recharge, recharge, and inter-aquifer leakage recharge all have the same value. The evapotranspiration factor f_e was smaller than one, indicating that the actual evapotranspiration was smaller than the reference evapotranspiration.

Usually, it is safe to assume that the actual evapotranspiration is not greater than the reference evapotranspiration. This leads to the interpretation that a value of the evapotranspiration factor greater than one indicates loss of water e.g. due to drainage, surface runoff, or interflow to surface waters. In this situation, the factor f_e is interpreted as the reduction of the precipitation contribution to recharge, while the actual evapotranspiration is assumed to be equal to the reference evapotranspiration. The latter assumption seems reasonable if the groundwater table is so close to the surface that drainage and surface runoff are important. This gives the following approximation for recharge:

Table 1
Overview of components included and data inputs and outputs from the three models. x – Included in the model; a - please see footnote.

TYPE	DATA	Metran	AquiMod	GARDÉNIA
INPUT DATA	Precipitation	x	x	x
	Temperature	* *		x
	Potential evapotranspiration	x	x	x
	Groundwater abstraction	x	x	x
	Groundwater levels	x	x	x
	Stream Discharge			x
SURFACE OUTPUT DATA	Surface runoff		x	x
	Actual evapotranspiration		x	x
	Snowmelt			x
	Stream discharge		*	x
SUB-SURFACE OUTPUT DATA	Groundwater levels	x	x	x
	Potential recharge			
	Recharge	x	x	x
	Inter-aquifer leakage		x	

* * Metran can be used with temperature instead of evapotranspiration, but this requires special adaptation in the calculation of recharge.
[a] AquiMod model produces aquifer discharges. Under certain conceptualisation, these discharges can be considered as stream discharges.

Table 2
Overview of parameters and conceptualisation of the three models.

Aquimod		GARDENIA	METRAN
Conceptualisation	• Uses three modules (soil zone, unsaturated zone, saturated zone) • Soil zone module: Account for soil moisture changes based on the use of the root constant, soil moisture at full saturation and soil moisture at wilting parameters. • Unsaturated zone: Applies the Weibull function to smooth and delay the water leaving the soil zone • Saturated zone: can be made of one or multiple compartments based on the geological complexity of the aquifer	• Three modules (surface reservoir, intermediate reservoir, and groundwater reservoir) • Surface reservoir: accounts for the evapotranspiration and snow fall • Intermediate reservoir: Takes the effective precipitation calculated in the surface reservoir and accounts for fast flow (interflow) • Groundwater reservoir: Takes percolation or recharge from the intermediate reservoir and calculate groundwater slow flow. It may have multiple orifices to account for groundwater flows with different velocity.	• Based on a Transfer Function-Noise approach use with here only precipitation and evapotranspiration as input. • Precipitation and evapotranspiration transfer function based on the gamma distribution • The transfer function for the reference evapotranspiration is equal to the precipitation function except for a multiplication factor • AR(1) noise modelRecharge is estimated by deducting the evapotranspiration multiplied by the evapotranspiration factor from rainfall
Number of parameters to calibrate	• Soil zone module (three parameters): root constant, depletion factor, soil moisture at full saturation and soil moisture at wilting parameters. • Unsaturated zone (two parameters): k and lambda parameters that are defining the shape of the Weibull distribution • Saturated zone (for each compartment): base elevation, hydraulic conductivity (k) and storage coefficient (S).	• Surface reservoir: A parameter that defines the shape of the curve controlling the calculation of both the effective precipitation and the evapotranspiration • Intermediate reservoir: A parameter used in a non-linear equation. • Groundwater reservoir: One parameter controlling the exponential equation used to empty the saturated volume. If this compartment is emptied using two routes, two parameters will be used.	• Three parameters for the precipitation transfer function. • One parameter for the evapotranspiration factor. • One noise decay parameter in the noise model. • Total of five parameters .

$$R = N - (f_e * E)\ (f_e \leq 1) \tag{2}$$

$$R = (N/f_e) - E\ (f_e > 1) \tag{3}$$

2.3.2. Groundwater level based hydrological model (AquiMod)

AquiMod is a lumped numerical model that was originally developed to investigate the behaviour of a groundwater system. The aim is to predict missing historical groundwater levels and calculate future groundwater levels (Mackay et al., 2014a). The model captures the conceptual understanding of a groundwater system using three numerical modules. These represent the soil zone, the unsaturated zone, and the saturated zone.

Water infiltration into the soil zone is calculated using the FAO Drainage and Irrigation Paper 56 (FAO, 1998) approach as modified by Griffiths et al. (2006). The first module accounts for actual evapotranspiration from crops and the soil moisture deficit to calculate excess water from rainfall. The excess water is split into infiltration and runoff. The delay of infiltrated water movement within the unsaturated zone is controlled by the second module using a two-parameter Weibull probability density function. Finally, the fluctuation of groundwater levels is simulated by the third module representing the saturated zone using an explicit numerical form of the mass balance equation. The saturated zone can be represented by several vertically arranged aquifers to capture the conceptual understanding of the site. A full description of the mathematical equations used to simulate the movement of water in these three zones is detailed by Mackay et al. (2014a) and Mackay et al. (2014b). This study focuses on the amount of water infiltration into the aquifer, consequently, the recharge estimated from AquiMod is considered recharge or inter-aquifer leakage.

The model is calibrated by changing the hydraulic parameter values included in the three modules, an overview of the parameters can be seen in Table 2. In AquiMod all the hydraulic parameters have a physical meaning such as the root constant, soil moisture at wilting point and at saturation, hydraulic conductivity and storage coefficient values of each compartment representing the saturated part of the aquifer. Only the flow with the unsaturated part of the aquifer is simulated using a mathematical equation which replicates the impact of the unsaturated zone on the groundwater flow. The calibration is done using a Monte Carlo approach that produces several simulations that are equally acceptable. These acceptable simulations are all used to calculate uncertainty bounds around the estimated recharge values.

2.3.3. Lumped catchment model (GARDÉNIA)

The computer code GARDÉNIA (modèle Global A Réservoirs pour la simulation des Débits et des Niveaux Aquifères) is a lumped hydrological model for the simulation of relationships between time series of stream or spring flow data at the outlet of a watershed, and/or groundwater level data at an observation well situated in the underlying aquifer (unconfined aquifer) and the rainfall received over the corresponding catchment. Groundwater abstractions can be considered if necessary.

The behaviour of an aquifer system is represented by a series of interconnected tanks. Non-linear transfer functions improve the capability of this representation to simulate a complex system. GARDÉNIA simulates the water cycle through a series of 3 or 4 connected tanks that represent respectively:

1. The top few decimetres of soil that are subjected to the influence of evapotranspiration (root zone);
2. An intermediate zone of rapid flow or quick flow;
3. One or two aquifer zones of delayed, slow flow.

The outflow from one reservoir to another is controlled by simple laws, specific to each reservoir; these laws are governed by the model parameters (active storage, duration of outflow, overflow threshold, etc.).

GARDÉNIA determines the hydrological balance for the basin: actual evapotranspiration, runoff, infiltration, and recharge. The hydrological balance can be used for the evaluation of groundwater recharge to the groundwater table (recharge). GARDÉNIA also provides river flow, groundwater level or recharge data over the period for which precipitation and potential evapotranspiration data are known.

One or two series of observations can be considered: river flows at the basin outlet and/or representative groundwater levels at an observation well located in the basin, also taking impacts of groundwater or river abstraction into account.

Calibration is done in a semi-automatic way. The user provides an initial set of parameters and indicates which parameters should be optimised (see an overview of parameters in Table 2). Gardenia's hydraulic parameters have also a physical meaning, similarly to AquiMod but there are also linear equations that are used to control the movement of water in the subsurface and the parameters of which are optimised during calibration. The model uses a non-linear optimisation algorithm for calibration adopted from the Rosenbrock method (Rosenbrock, 1963). The model varies the chosen parameters (within a range of values defined by the user) and searches for the set which gives the best fit between observed and simulated values.

2.3.4. Performance evaluation

As the three models are different, so are the ways their performance is evaluated. The performance of a Metran model is expressed by three scores of 0, 1 and 2 based on the model evaluation criteria given by Zaadnoordijk et al. (2019). A score of 0 indicates the model is not acceptable, 1 that it is acceptable (potentially useful), and 2 that it fulfils additional criteria (good). The criteria are related to the explained variance, independence of the noise part from the transfer part, and the chance of overfitting. The inability to fit an acceptable model can be due to any or all of the following conditions: 1. A variable with an important influence on the groundwater heads was not included, e.g., abstraction; 2. Non-representative input data e.g., inadequate precipitation or evapotranspiration data; 3. Insufficient information in the calibration data or errors in groundwater heads; 4. a linear response with the gamma response function cannot provide an adequate approximation of the groundwater system. However, to aid the comparison of the three models, the Nash-Sutcliffe Efficiency (NSE; see Nash and Sutcliffe, 1970) of Metran's groundwater level simulation have also been calculated.

The performance of AquiMod is routinely evaluated using standard performance criteria such as the Root Square Mean Error (RMSE) and the Nash-Sutcliffe Efficiency (NSE; see Nash and Sutcliffe, 1970) performance measures, while NSE is used in GARDÉNIA to evaluate the performance of both groundwater heads and river discharge.

While many studies set criteria that relate the goodness of calibration to the NSE values (Moriasi et al., 2007; Kalin et al., 2010; Kouchi et al., 2017), there exist no standard values of NSE used for the acceptance/failing of AquiMod simulations. An NSE value of 1.0 indicates a perfect match between the observed and simulated data. In this study we consider AquiMod simulations that achieve an

Table 3
Additional methods/models from which results are used for comparing with the recharge estimations.

TYPE	MODEL	RESOLUTION	REFERENCE	COUNTRY (PILOT/-S)	RECHARGE TYPE
MEASUREMENT	Lysimeter	Point	Patrčević (1995)	Croatia (36)	2
SOIL STORAGE MODEL	ZOODRM	National	Mansour et al. (2018)	UK (all UK)	1
HYDROLOGICAL MODEL	NAM	Local	Nielsen and Hansen (1973)	Denmark (3–5)	2
	HYPE	National	Lindström et al. (2010)	Sweden (1,2)	2
	SACRAMENTO	Local	(Pérez-Sánchez et al. 2019) Burnash et al. (1973)	Spain (40)	2
INTEGRATED HYDROLOGICAL MODEL	MARTHE	Local	Amraoui et al., (2004, 2019)	Somme (34)	1,2,3
	MIKE SHE	National	Højberg et al. (2013)	Denmark (3–5)	1,2,3
	Refined NHI-LHM / iMOD	Local	TACTIC (2021a)	Netherlands (21)	1,2,3
	NHI-LHM / iMOD	National	De Lange et al. (2014)	Netherlands (21)	1,2,3
MACHINE LEARNING	TACTIC EU Recharge map	Europe	Martinsen et al. (2022)	Europe (all)	1

NSE score higher than 0.7 as good simulations that can capture the observed data behaviour. Those with NSE score of 0.5 are still included as they simulated groundwater level time series show acceptable reflection of the observed ones. For the GARDÉNIA model, previous work (Bouanani et al., 2012; Jeannin et al., 2021) and experience the performance of GARDÉNIA have showed that NSE, is deemed acceptable when it is above 0.7, good when above 0.8 and very good for NSE values above 0.9 This criterion is also implemented here and GARDÉNIA simulations that produce NSE values greater than 0.7 are accepted for analysis.

2.3.5. Other results used for comparison

Recharge results from models and measurements from other studies were also collected and used in this study for comparing and assessing recharge estimates. The models are mostly from national modelling studies, but some site-specific model results are also available. The model types are both hydrological and integrated hydrological models. An overview of the additional results used is shown in Table 3. The models will be described briefly in the following section, but for more detailed information the reader is referred to the references given in Table 3.

In Croatia recharge was estimated from lysimeter measurements (Patrčević, 1995) and water balance analyses (Urumović et al., 1981; Brkić, 1999). This recharge estimation has also been used as input in a MODFLOW (Harbaugh et al., 2000) model (Brkić et al., 2021). MODFLOW is a groundwater flow simulation program and needs groundwater recharge as input either as a calibration parameter or an input value/ time series, instead of climate input (e.g., precipitation, pot. evapotranspiration).

Three additional hydrological models are used: NAM, SACRAMENTO, and HYPE. The NAM model (Nielsen and Hansen, 1973) is a lumped conceptual catchment model based on four water storage compartments where water is moved through the system using nine empirical parameters. Recharge is calculated as the water moving from the root zone storage to the groundwater storage and only

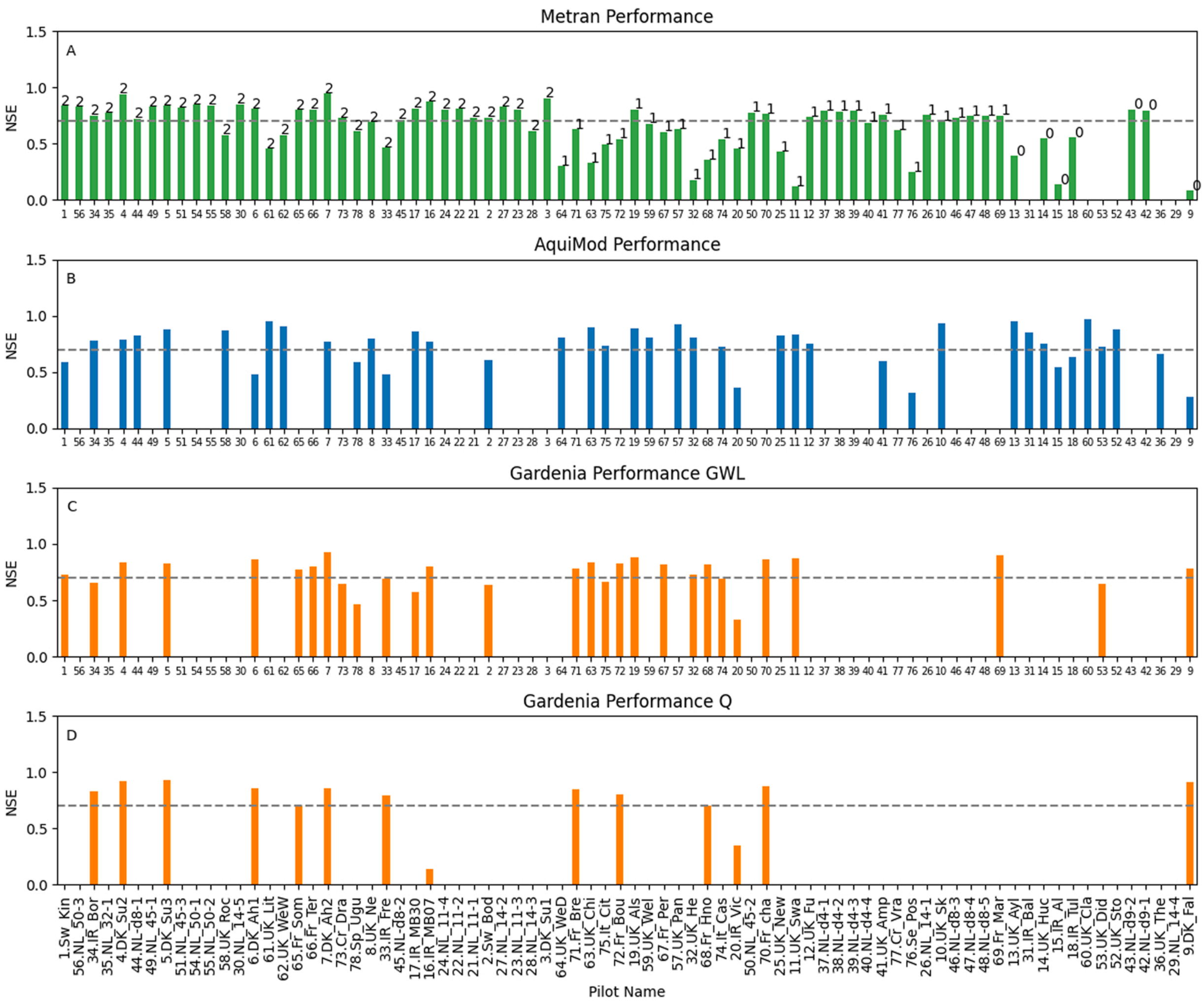

Fig. 3. Overview of the performance of the different models after calibration. A: Metran performance shown for NSE, the model is however evaluated by a score of either Best= 2, OK= 1 and not accepted= 0; B: AquiMod model performance evaluated on NSE from groundwater levels. C and D: Performance of GARDÉNIA simulations on groundwater level (C) and stream discharge (D) evaluated on NSE. Models are sorted based on Metran performance. The dotted lines indicate the acceptable performance level. Contains numerical data: British Geological Survey © UKRI 2023.

represents recharge to the saturated zone (TYPE2). The model was set up and calibrated for three catchments containing the Danish sites within the TACTIC project (TACTIC, 2021a). HYPE is a semi-distributed conceptual hydrological model, which was developed in Sweden, and is now used as the national Swedish hydrological model for discharge and nutrient loads, S-HYPE (Strömqvist et al., 2012). The S-HYPE model is used extensively in Sweden and multiple uses have been well documented. In this study, the recharge in HYPE reflects percolation to the groundwater table (TYPE2), where the calculation of net precipitation employs a generalised empirical relationship for evapotranspiration that has been calibrated for the national model (involving a water balance of c. 400 catchments). The SACRAMENTO model was used in Spain, where it provided the input of groundwater recharge (TYPE2) to the local MODFLOW model setup for the Upper Guadiana Basin (Collados-Lara et al., 2021; Surge, 2018).

The national recharge estimation tool the Zooming Object Oriented Distributed Recharge Model (ZOODRM) is used in United Kingdom (Mansour et al., 2018). The model calculates recharge using characteristics of plant growth and water uptake, as well as information on soil characteristics and climatic conditions. The model operates on grid basis and covers the entire British Mainland.

The hydrodynamic model of the chalk aquifer of the Somme basin uses a finite difference groundwater modelling approach implemented in BRGM's MARTHE code (Thiéry, 2015b; Thiéry et al., 2018). Groundwater is the primary water resource for water supply in the French Somme Department. Hydrodynamic modelling of the Somme basin (Amraoui et al., 2004) is used to study the impacts of climate change on groundwater recharge to the chalky aquifer, piezometric level fluctuations and streamflow to the drainage network (Amraoui, 2019). Like MODFLOW, MARTHE is a groundwater flow simulation program that needs groundwater recharge as input. Integrated hydrological models also form the basis of the national water resources model of Denmark (based on MIKE SHE) and the Netherlands (based on MODFLOW for groundwater (de Lange et al., 2014)). Common to these models is the integrated calculation of unsaturated zone, surface water and groundwater fluxes, and they can therefore estimate all three recharge types. Water fluxes in the unsaturated zone are calculated based on e.g., Richard's equation (MARTHE) or a 2-layer solver (MIKE SHE). The saturated zone fluxes are calculated by finite-difference solving Darcy equations for 3D groundwater flow. The national application of the models is well documented in previous publications for the Netherlands (de Lange et al., 2014), and Denmark (Højberg et al., 2013; Stisen et al., 2019). Also included is a local high-resolution version of the Dutch national model run for the catchment of DeRaam (TACTIC, 2021a).

The last tool used for comparison of recharge is a machine learning aided approach developed in the TACTIC project for the calculation of potential recharge (TYPE1) at a pan-European scale (Martinsen et al., 2022). The machine-learning algorithm uses gridded national model estimates of infiltration coefficients as training data, applying a range of explanatory variables to establish the governing variables for infiltration across Europe. The resulting infiltration coefficient map is then multiplied onto a European net precipitation map generated from the E-Obs precipitation dataset at 0.1° resolution and satellite data for evapotranspiration (Cornes et al., 2018). In this study, potential recharge is obtained for the grid location of the study sites. More information on the method can be acquired from TACTIC (2021b).

3. Results

3.1. Calibration results

The performance of the models after calibration from the different study sites is presented in Fig. 3. As not all study sites have all three tools, some areas are left blank on the graphs.

The Metran model is applied for all 78 time series. As seen in Fig. 3A, 35 out of 78 time series have the best performing model fit (score = 2), while 31 are acceptable (score = 1). For a small group of 12, it was not possible to calibrate a usable Metran model for the site (score = 0). For comparative reasons NSE is shown by the bar height in the fig.

ure. As can be seen a high NSE is not always a guarantee for a good score and vice versa, due to the two-stage nature of the standard performance evaluation. Overall, 31 of the Metran models have NSE values above 0.75. AquiMod is applied to 40 time series in 37 study sites (Fig. 3B). Most of the model runs (28 time series) produce very high NSE values (>0.70) for the groundwater level, while 7 sites produce NSE values between 0.5 and 0.7. Poorer performance was found for five sites in Serbia, Denmark and Ireland. GARDÉNIA is used for 30 groundwater head time series (Fig. 3C). Sixteen sites of these include stream discharge (Fig. 3D). Generally, performances are good, with 19 time series for groundwater heads and ten time series for stream discharges having NSE values larger than 0.70 and eight time series for groundwater heads and two time series for stream discharges with NSE above 0.50 but below 0.70. Performance deteriorates for four sites in Ireland and in Spain.

The calibration of the different models shows that there is not always a correspondence between when the models succeed or fail. For time series where all three models are applied (22 time series), 32% of the time series show that all succeed in making a reasonable calibration target. For 10 time series, one model fails to produce a reliable result, while four time series only have one model succeeding out of three. For the time series where two models are applied (29 time series), both models are successful in 62% of the cases. One of the two models fails in 11 of the cases. There is, thus, seldom a correlation between the poor performance of the models. There is also no clear tendency that it is always the same model that fails, generally, Metran and AquiMod have a fail rate of 15% and 10%, respectively, while GARDÉNIA fails in 37% of all applications.

3.2. Selected case studies

To illustrate the modelling of recharge using the three models, four case study sites are described in more details. Each case study illustrates and highlights different structural differences or limitations of one or all three models. The case studies have been chosen both to illustrate important challenges with applying the models and to cover a range of geologies and climatologies across Europe. It should be noted that both AquiMod and GARDÉNIA provide recharge time series, while Metran provides mean annual values of recharge. A total of four contrasting cases are included here: Spain, Serbia, Ireland and Sweden. These will be briefly introduced in each section.

3.2.1. South - Upper Guadiana, Spain (Time series 78)

The Campo de Montiel aquifer covers an area of nearly 2200 km^2 located within a semiarid region in the Upper Guadiana Basin (Spain). It is mainly composed of limestones and Jurassic carbonate formations. Its altitude varies between 800 and 1000 m asl. The mean annual temperature is about 14 °C. Annual precipitation varies between 400 and 500 mm. The predominantly dry climate and the prevalence of irrigated agriculture mainly supplied with groundwater resources has led to a high and permanent conflict between agriculture and ecosystem conservation, with significant impacts on the wetlands (protected RAMSAR sites) and aquifers (quantity and quality) in the past (Baena-Ruiz et al., 2021; Baena-Ruiz and Pulido-Velazquez, 2020).

GARDÉNIA, AquiMod and Metran modelling tools were employed to estimate groundwater recharge based on more than 35 years of available data (from April 1978 to – October 2015) in one piezometer. The groundwater head time series (Fig. 4A) are from an unconfined aquifer and are related to recharge. The mean recharge values (Fig. 5A) obtained with GARDÉNIA, AquiMod and Metran are 69.4 mm/year, 69.8 mm/year, and 421 mm/year, respectively. The differences between the mean recharge estimated by applying GARDÉNIA or AquiMod with respect to site-specific model (66.2 mm/year) (Table 1) are not significant (maximum of 5.6%), meanwhile the mean recharge obtained with Metran is significantly greater. Metran was run for 10 time series in the basin but only succeeded in producing a model for one. Even though this model passed the statistical test, it has been also dismissed as it has almost no

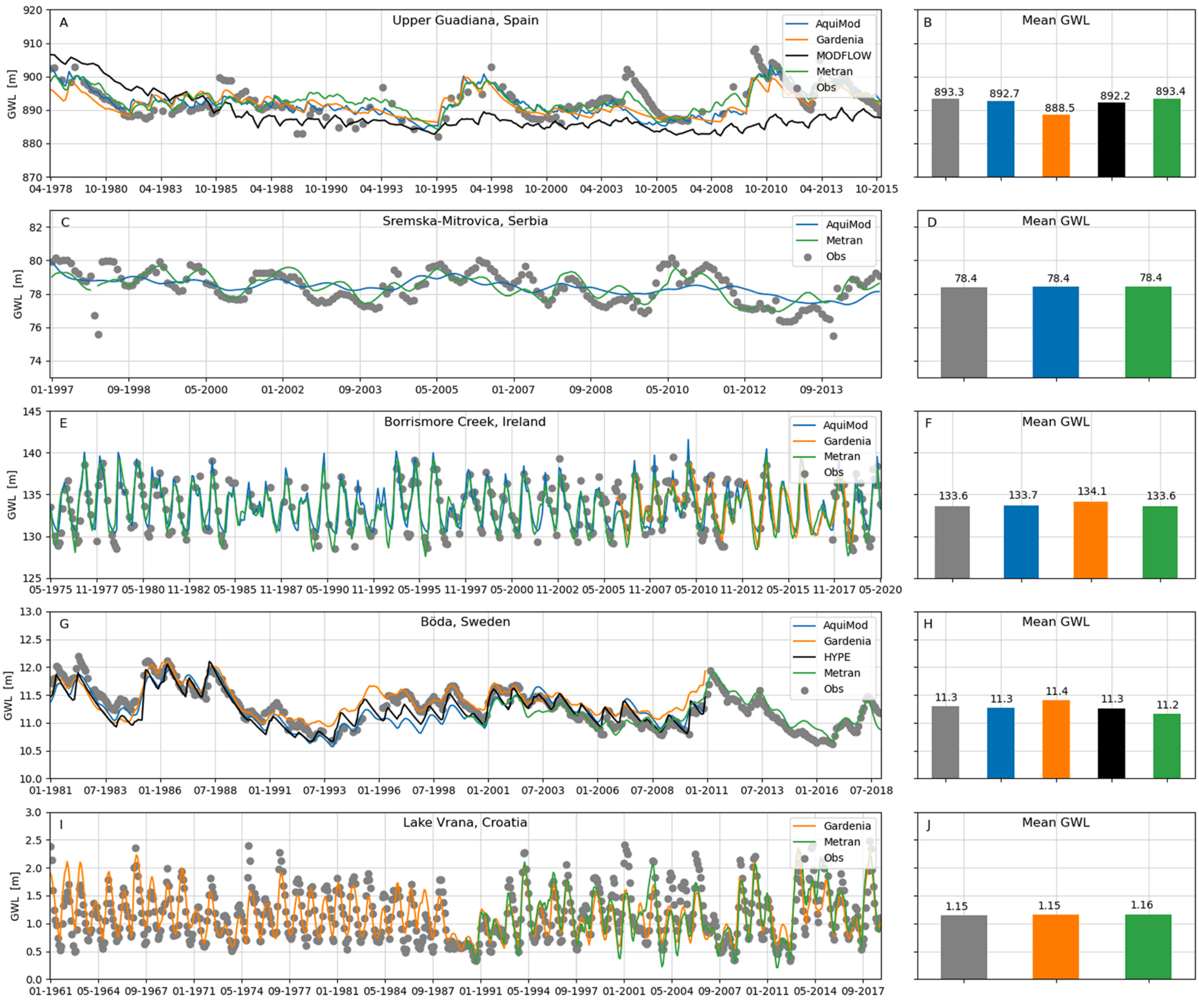

Fig. 4. Observed and simulated groundwater levels [m] with GARDÉNIA, AquiMod, Metran and any local model present at the case study (Left), mean values are shown to the right (green bar). Selected results are shown for the four case studies: Spain (A), Serbia (C), Ireland (E) and Sweden (G).

Fig. 5. Simulated recharge values (mm/month) with GARDÉNIA, AquiMod and any local model present in the case study (left), with mean annual recharge values for the same simulations (right) with the addition of Metran (green bar) that only provides results on annual basis. Selected results are shown for the four selected case studies: Spain (A), Serbia (C), Ireland (E), and Sweden (G).

evapotranspiration response indicating that groundwater fluctuations are independent of evapotranspiration.

The long time series make it possible to perform a sensitivity analysis of the results for the period of data employed to calibrate and validate the GARDÉNIA and AquiMod models. Three calibration scenarios were carried out using the entire dataset and 70% and 50% of the period, while the rest of the dataset was reserved for validation. Table 3 shows the mean recharge in the period 1978–2015 and the goodness of fit in the simulated groundwater levels obtained for different calibration scenarios. Both models provide similar mean recharge values and acceptable results in groundwater level forecast. In this case study, most AquiMod models provide a slightly better simulation of groundwater level than GARDÉNIA, although the estimated recharge provided by GARDÉNIA is more similar to the site-specific model. Table 4.

Fig. S1 of the electronic supplementary material (ESM) shows the observed and simulated groundwater levels obtained with the different calibrated models in the period 1978–2015. It shows that both models, GARDÉNIA and AquiMod, have the ability to reproduce groundwater levels within the different validation periods. Therefore, in addition to assessing groundwater recharge, they can be employed as tools for short-term groundwater level forecasting.

Table 4
Mean recharge (mm/year) and goodness of fit of groundwater level simulations with GARDÉNIA and AquiMod (in parentheses the percentage of period simulated) in Campo de Montiel (Spain). Cali = Calibration period; Vali= Validation period. R2 = r-squared, NSE=Nash Sutcliffe coefficient, RMSE= Root Mean Square Error, ME= Mean Error.

	GARDÉNIA (50%)		GARDÉNIA (70%)		GARDÉNIA (100%)	
MEAN RECHARGE	69.1 mm/year		69.4 mm/year		66.9 mm/year	
PREFORMANCE (GWL)	CALI.	VALI.	CALI.	VALI.	CALI.	VALI.
R2	0.62	0.67	0.54	0.77	0.76	-
NSE	0.57	0.07	0.46	0.50	0.57	-
RMSE	3.18	5.17	3.30	4.17	3.39	-
ME	0.00	-3.40	0.08	-2.54	0.25	-
	AQUIMOD (50%)		AQUIMOD (70%)		AQUIMOD (100%)	
MEAN RECHARGE	68.9 mm/year		69.8 mm/year		69.8 mm/year	
PREFORMANCE (GWL)	CALI.	VALI.	CALI.	VALI.	CALI.	VALI.
R2	0.62	0.88	0.62	0.72	0.78	-
NSE	0.43	0.59	0.59	0.37	0.55	-
RMSE	3.67	3.43	2.90	4.66	3.45	-
ME	1.28	-1.43	0.52	-1.80	-0.06	-

3.2.2. *East - Posavina, Serbia (Time series 76): Borehole affected by river stage*

The pilot area (Posavina) is situated in the Sava River Basin in Serbia and is mostly covered with Quaternary sediments. This area stretches from the border with Croatia on the West to Belgrade in the East, and from Fruška Gora Mountain in the North to Cer Mountain in the South. The territory covers an area of about 5250 km^2, and the Sava River catchment size is 9057 km^2. The major groundwater reserves are accumulated in thick Quaternary sediments where high yield intergranular aquifers are formed. The aquifer is recharged from precipitation and groundwater inflow. However, the main inflows to this aquifer are from surface waters. The analysis of the groundwater regime indicates that the influence of the river on the groundwater table weakens with distance from Sava River. To simulate groundwater recharge and prediction of future groundwater levels, AquiMod and Metran models were applied.

AquiMod represents the groundwater system as a closed homogeneous medium, with no impact from any outer boundary or feature, whether physical or hydrological. With the significant influence of river stages on groundwater heads, it was therefore not possible to obtain a good AquiMod model (Fig. 4C). Metran, on the other hand, succeeded in capturing the groundwater dynamics (Fig. 4C) but only when the river water levels were added as an additional input signal.

3.2.3. *West – Ireland (Time series 15, 16, 17, 18, 20, 31, 33 & 34): Aquifer complexity*

The boreholes are placed in the southeast part of Ireland and where the river basins of the Suir, Nore, Barrow and Slaney rivers, as well as a number of smaller coastal river catchments are located. The main aquifers of interest are diffusely karstified limestone aquifers and extensive glacio-fluvial sand and gravel aquifers. Groundwater from these regionally important aquifers provides up to 90% of the public water supply in some counties, is important for agricultural and industrial uses, and supports many small private supplies. Groundwater-surface water interaction is significant, and baseflows are high where rivers cross the principal aquifers. The karst aquifers occupy the valleys and lowlands and are generally unconfined. Permeability varies with depth due to fracturing and karstification. Sands and gravels occur mainly along the major rivers, but also form extensive 'domes'. Groundwater is generally unconfined, and flow pathways are simple. The remainder of the area is underlain by poorly productive fissured bedrock aquifers. The karstic and fissured bedrock aquifers have low effective porosity and very limited storage, which renders them vulnerable to altered winter recharge patterns and extended dry periods. Effective rainfall is lowest in southeast Ireland and there are currently concerns about water resources availability and demand, and it is anticipated that this area will become progressively water stressed under future climate change scenarios.

AquiMod and Metran were applied to all eight groundwater level monitoring records (Fig. 3), while six were simulated by GARDÉNIA. Detailed examples are shown on Fig. 4E/Figure F and Fig. S2B & C/S3B &C (ESM). The 30-year average (1981–2010) groundwater recharge for catchments to the eight boreholes were also estimated using the updated Irish National Recharge map (Hunter Williams et al., 2021). NSEs for AquiMod historical simulations range from 0.36 to 0.86, and − 0.24–0.8 for GARDÉNIA. A lower limit of 0.5 was chosen for acceptance. Six of eight AquiMod simulations, and four of six GARDÉNIA exceed that lower limit. Four of eight Metran models have the highest acceptance, one was acceptable, and three were insufficient.

Metran recharge estimates are consistently higher than AquiMod, GARDÉNIA and the national model. AquiMod estimates tend to be lower or the same as the national model, with no systematic pattern for GARDÉNIA results. The reasons for differences or difficulties in calibrating are likely to be multiple, and different depending on the site. Metran seems to be most applicable to gravel aquifers and more straightforward karst groundwater systems. At two sites, riverine influence on groundwater level is suspected, Metran does not calibrate without the river stage as explanatory variable. Using GARDÉNIA requires river flow data for the most robust simulations. Due to the lateral heterogeneity within Irish catchments, it is possible that the groundwater level variations are not representative of the nearest gauging stations, which encompass a much larger catchment. GARDÉNIA also requires a full calendar year of data and does not allow missing head data while the other models do.

Overall, AquiMod seems to be most suited to simulation of groundwater levels in the Irish aquifers included. However, it should be noted that the more complicated karst and fractured bedrock aquifers required two-layer models at least to better represent the groundwater flow systems, and more remains to be done in this regard. Clearly, a good historical model fit will not guarantee predictive power in future climate change scenarios, since different recharge dynamics, timings and groundwater stages are likely to interact non-linearly with the heterogeneous aquifers.

3.2.4. North - Böda & Kinda, Sweden (Time series 1 & 2): Snow cover projections

In rural areas of Sweden, the supply of drinking water typically relies on private wells drilled in crystalline bedrock, which drain the storage of the overlying soil layers (typically, up to a few meters of till). In many areas, groundwater storage is limited by low effective porosity and compartmentalised (fragmented) soil coverage. This limited storage implies that the drinking-water consumption must be

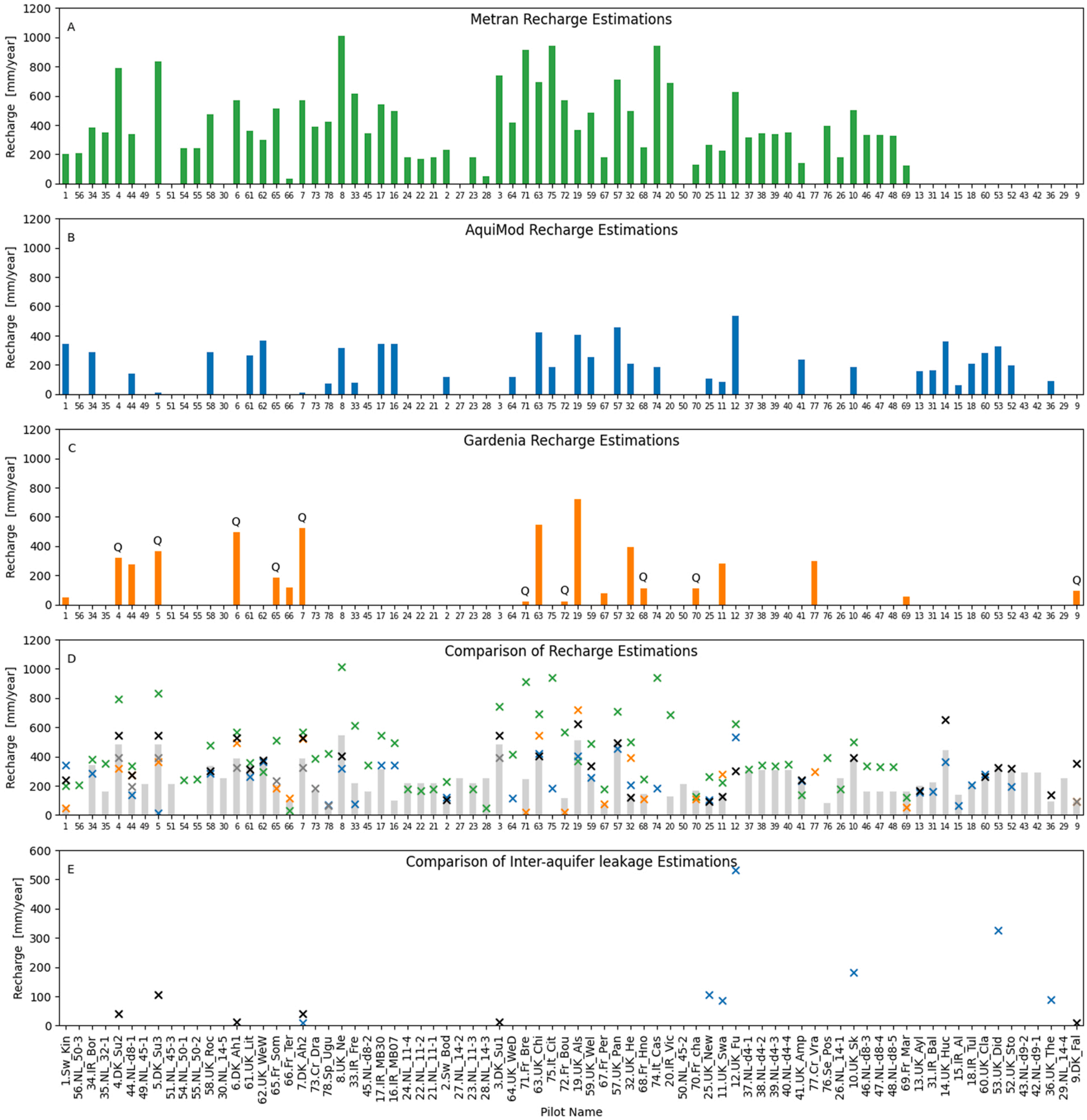

Fig. 6. Recharge estimations by the three models: A – Metran (green), B – AquiMod (blue), and C – GARDÉNIA (red). Pilots are sorted by Metran performance like shown in Fig. 5. Q (plot C) indicates that the GARDÉNIA model was calibrated with both groundwater level and stream discharge. D - Comparison of recharge estimations by the three models (Metran, AquiMod and GARDÉNIA) with results from already established models at the pilot either national (black) or local (grey); and potential recharge estimated by the European recharge map (grey bars, Martinsen et al., 2022). E - Comparison of inter-aquifer leakage estimations by the AquiMod model with results from already established models at the pilot either national (black) or local (grey). Contains numerical data: British Geological Survey © UKRI 2023.

in close balance with the recharge, not only in terms of the long-term means but – more importantly – with seasonal drought patterns, otherwise the groundwater aquifers are in risk of depletions. Seasonal droughts are periods when the recharge is negligible, such that any water consumption relies on drainage of groundwater storage. The groundwater storage determines, therefore, the upper bound for consumption during the drought period and the vulnerability to the length of the drought period; in turn, this implies that the main concern of the ongoing climate changes is the signs of longer seasonal droughts.

At present, the consumption from private wells is adapted to two types of seasonal droughts: 1) winter drought, when precipitation is stored as snow (dominates in the north), and 2) summer drought when interception and evapotranspiration exceed precipitation (dominates in the south). In areas of substantial snow storage, compared to soil storage, the snowmelt has the important role of re-filling the storages prior to the onset of the summer drought.

Climate change is expected to shift the balance of seasonal patterns from winter drought towards longer summer drought. Moreover, the role of snowmelt for re-filling the soil storage will weaken, particularly in the south of Sweden, as the snow storage will diminish (more precipitation will fall as rain during winter) and the snowmelt will advance to an earlier date (thereby elongating the summer drought). On the brighter side, climate change may shorten the winter drought in northern Sweden. Simulations for the Swedish pilots demonstrate the need to account for a retreating snow cover in climate change projections on future recharge and water supply.

Three tools were tested and compared: AquiMod (which does not account for snow storage) and GARDÉNIA and HYPE (which does account for snow storage). All tools could be reasonably well calibrated for the present-day situation (Figs. 4G and 5G; Fig. S2A and S3A of the ESM); the fit for AquiMod was NSE = 0.59 and 0.61 for two pilots, whereas it was NSE = 0.73/0.64 and NSE = 0.81/ 0.67 for GARDÉNIA and HYPE, respectively. Obviously, this signifies that data calibration is not undisputed evidence of model validity as the conceptual model also needs to capture the key processes involved. A classical dilemma of models is that of equifinality (Beven, 1993), where the model produces the "right output", but for the "wrong reasons". Thus, if the models are to be used for the projection of future conditions, it is vital that the models are able to correctly represent the processes that are expected to change in the future, e.g., in this case, that the projection model includes a snow storage module.

3.3. Recharge estimations

The recharge estimation for all the sites is shown in Fig. 6 by each of the models Metran (A), AquiMod (B) and GARDÉNIA (C), again as not all study sites have applied all three models, some areas are left blank on the graphs. Recharge results for models with inferior performance are not shown, this means that Metran results are only shown for a performance of 1 or 2. Negative recharge estimations, which may be due to local discharge, are left out of the plot but are found by Metran for the boreholes NL_14 and NL_45 (time series 26–30 & 49–51). One well is in an area that can be connected with an area with artificially managed water levels, so here there may be upwelling discharge, however this is not the case for the other well, there is therefore no physical explanation to this value and the estimate is therefore discarded. As mentioned earlier, a poor score by Metran may indicate that the input time series does not hold enough information to describe the groundwater level fluctuations in the observed time series. Thus, in these cases it is potentially not possible to estimate recharge using the groundwater level data (with the available climate and abstraction data), and therefore any modelling efforts by AquiMod and GARDÉNIA could be disregarded. However, in this case we have chosen to only disregard model results due to the model's own performance. Therefore GARDÉNIA results are shown for simulations with NSE> 0.7 for both groundwater levels and discharge, and AquiMod results are shown for all models except for sites where the model performance was

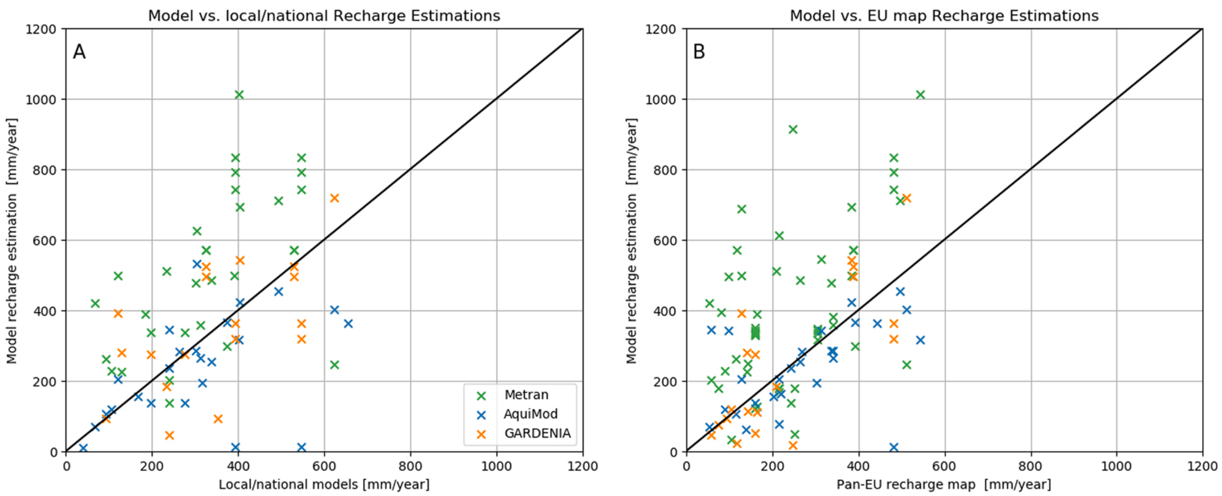

Fig. 7. Recharge estimations by the three models Metran (green), AquiMod (blue), and GARDÉNIA (red) plotted against recharge estimated by results from already established tools at the pilot (A) and the European recharge map (B). Contains numerical data: British Geological Survey © UKRI 2023.

deemed unacceptable as mentioned in the section above (4, 6, 9, 20 and 76). As the models are sorted by Metran performance (the same sorting as Fig. 5), the AquiMod and GARDÉNIA models with failed Metran score= 0 are easily identified as the models with missing columns to the right on Fig. 6.

Fig. 6D and E show a comparison between the results from the three models. The time series may both be from unconfined and confined aquifers, and thus for AquiMod the recharge estimation is sometimes not recharge (Fig. 6D), but inter-aquifer leakage (Fig. 6E). Looking across the results where all three models are used, and performance was accepted (22 time series), there are very little general agreements between the three models, in relation to recharge estimates. For some timeseries, recharge estimations differ in the order of 700 mm/y, while much better agreement is reached in others. Often Metran recharge values are higher than the other two tools, but not always. For the unconfined aquifers (11 time series) Metran estimates recharge with a mean of 530 mm/y. For the same boreholes, AquiMod estimates a mean of 220 mm/y, while GARDÉNIA estimates a mean of 290 mm/y. Thus, the uncertainty of recharge estimations is generally high, and while multiple methods for assessing recharge exist, as these three tools applied here, their assumptions and results often deviate substantially, as also found by e.g., dos Santos et al. (2021) and Walker et al. (2018). This is supported by a study by Walker et al. (2018), applying nine different recharge methods on a catchment in Ethiopia; they found that the range of recharge estimations could be reduced substantially when correctly identifying the recharge type (minimum, potential, actual), the appropriate scale of the estimation and accounting for the uncertainty of the method.

3.4. *Comparing recharge estimations to local and national tools*

The true values of recharge for the sites are unknown, therefore in order to try to identify reliable recharge estimates, the recharge estimated by the three tools is compared with local and national model results whenever possible (Fig. 6D and E). For 32 time series, there are recharge estimations from other tools, a table of the annual recharge estimates for the 32 sites are available in the ESM, Table S2.

A comparison between the three models and the already established local and national results can also be seen in Fig. 7A. From the scatter plot in Fig. 7A, there is an inclination for Metran recharge to be higher than the local and national tool estimates (above the 1:1-line). For AquiMod and GARDÉNIA points are on both sides of the 1:1-line. Thus, for most of these sites, AquiMod and GARDÉNIA estimates were fairly close to the local and national model results, with a mean absolute deviation of 38% and 46%, respectively, for the GARDÉNIA run where discharge is used in the objective function deviation is 29%. Metran recharge estimations generally were higher, with mean absolute deviation of 90%. Taking out all runs but AquiMod and GARDÉNIA, where Metran failed gave similar deviations of 41% and 48%, respectively (and for GARDÉNIA with discharge 27%). Thus, the bias of Metran is quite higher than for the other models. An investigation of the correlation of the estimates show that the models have similar correlation. Metran has a correlation (Pearson) of 0.60 to recharge estimates by other tools, while AquiMod and GARDÉNIA correlation are 0.55 and 0.6

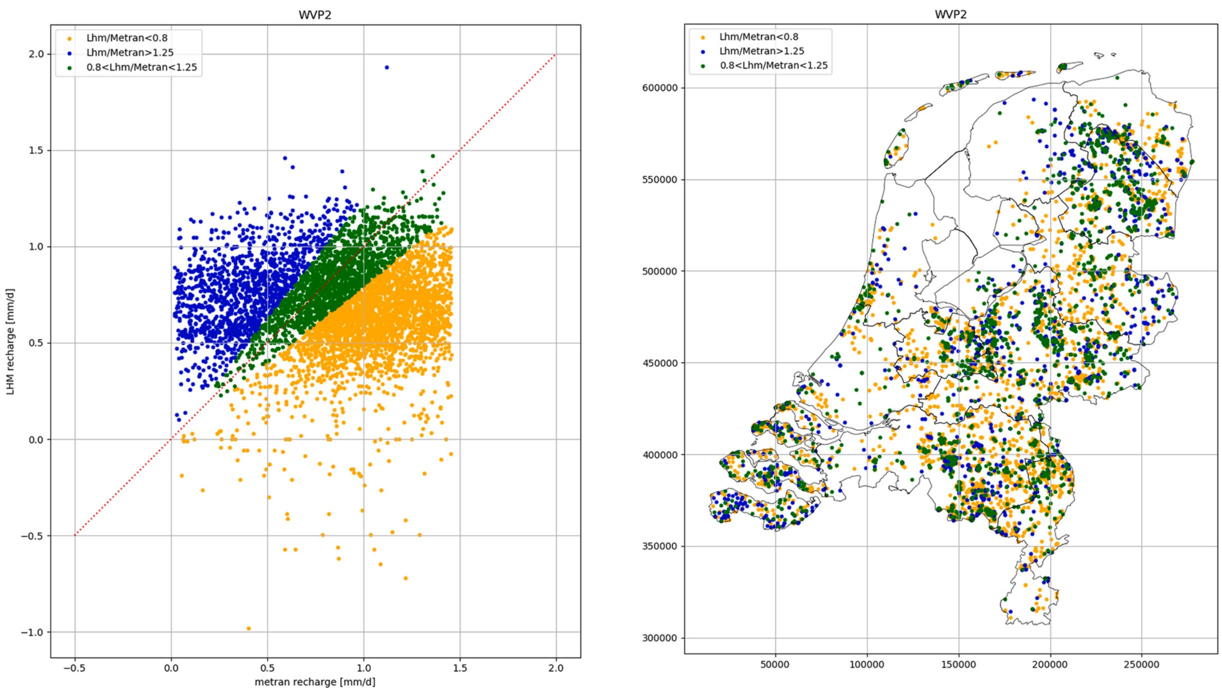

Fig. 8. Left - Correlation of recharge estimations by Metran and the Dutch National Hydrological model in the same boreholes. Right - Spatial distribution of the boreholes. Yellow indicates that Metran is underestimation recharge, green are boreholes with similar values to the national model, while blue indicates overestimation by Metran.

respectively.

To investigate this high bias further, recharge values estimated using a larger number of time series in the Netherlands by Metran are compared to the outputs from the integrated Dutch National Hydrological model (http://nhi.nu/, De Lange et al., 2014).

The comparison shows that Metran estimates do not compare well with the results from the integrated model (Fig. 8), and there is no visible pattern in the locations with reasonable performance (green), underestimation (yellow), or overestimation (blue), Fig. 8b. Metran only uses precipitation and evapotranspiration; however, its application does not give a reliable estimate of recharge values (either net precipitation or net precipitation minus runoff and local drainage). This suggests that special conditions or additional information are required to obtain reliable recharge estimates from groundwater head time series in combination with precipitation and evapotranspiration. Obergfell et al. (2019) applied a similar method with the additional constraint that the seasonal harmonic of the observed head matches the sum of the transformed seasonal harmonics of the input variables in the time series model for a specific area, to obtain results that compared well with independent estimates. However, the validity and consequences of this constraint are unknown.

This study does, thus, point to AquiMod and GARDÉNIA being appropriate methods for estimation of recharge, when considering the uncertainties inevitably attached to the estimation of groundwater recharge from hydraulic head measurements. However, it is only possible to validate with other (local and national) modelling results also prone to uncertainties, making this deduction only indicative. Furthermore, this study focuses on diffuse recharge, while indirect recharge from streams, lakes, and rivers are not considered; however, indirect recharge type are dominant in arid or semidry (semi-arid) areas and if omitted, the recharge to the entire aquifer may thus be underestimated for these climatologies.

3.5. Comparing recharge estimations to Pan-European estimate

Fig. 6D also includes potential recharge values obtained from the Pan-Eu recharge map (Martinsen et al., 2022). Two limitations should be noted in this comparison. First, the Pan-Eu recharge map is created using the national hydrological model's infiltration coefficient combined with other spatial variables in the machine learning algorithm. These are the same national models (applied in Spain, France, Ireland, Denmark and the Netherlands) for which a subset of estimated recharge values was used for comparison to the recharge estimate in the previous section. Thus, these two datasets for comparisons are not entirely independent. Secondly, the values from the recharge map are of potential recharge, i.e., infiltration recharge leaving the soil zone. They are not, therefore, directly comparable to the previous recharge and inter-aquifer leakage estimations. However, this means that Pan-EU recharge estimation should be equal to or larger than the other estimations, as interflow, drainage and runoff to surface water have not been removed. Furthermore, the map is generated at EU scale, and must therefore be assumed to be less accurate than local models, and models like GARDÉNIA where the calibration data are used to complete the hydrological balance. Nonetheless, some tendencies in the recharge distribution at some of the sites are replicated, e.g., lower potential recharge in Sweden and Ireland than in Denmark, high variability in potential recharge estimates in the UK, and lower potential recharge values in France and Spain. From the scatter plot of the three models and the Pan-EU recharge map in Fig. 7B, it can be clearly observed that there is a tendency for Metran recharge to be above the 1:1-line, meaning the recharge values are generally overestimated in relation to the Pan-EU map. For AquiMod and GARDÉNIA the points are more scattered, but with most points being below the 1:1-line as expected noting that the type of recharge calculated by these models is either recharge or inter-aquifer leakage rather than the potential recharge provided by the Pan-Eu map. The statistics for mean absolute deviation and correlation (given in parenthesis) are here for AquiMod and GARDÉNIA deviation of 50% (0.41) and 51% (0.76), respectively, with GARDÉNIA with discharge 22% (0.68) and Metran 112% (0.72).

4. Discussion

4.1. Applicability and limitations of the time series models

Recharge estimation based on groundwater head time series are valuable tools as they require very little data and can be easily and widely applied. Time series models are a subset of these tools, and they, however, cover a range of different approaches and their applicability, limitations and precision are not always well known. For some recharge estimation is a by-product of the tool, and the credibility of this recharge estimation is often uncertainty.

Time series models like Metran use precipitation and evapotranspiration (combined with abstraction) as input for calculating recharge. For Metran, it was shown that there are several problems in using this tool to assess recharge. The simulated groundwater head levels may be reasonable or good, and close to the measurements for a site. The average long-term recharge from Metran was shown to have a high bias but to correlate fairly well with local and national recharge estimates for the study sites used in this study, but on further investigation concerningly low correlation, and lack of systematics in the estimations was found when comparing to results from the distributed National model in the Netherlands. That implies that a reliable estimation of recharge is not possible using Metran, and this possibly also applies to similar transfer-noise models. These shortcomings will increase the uncertainty even further if the model is used to predict climate change impact.

In contrast, Metran was shown to effectively model groundwater head levels and is therefore a powerful and fast tool for assessing groundwater level. The Metran model was also successfully used to identify periods of time within a time series where the noise ratio was low, thereby assisting in the setup of other recharge estimation models as in boreholes in Italy. The flexible setup of the model also made it possible to adjust the model to sites where special conditions were governing the hydrology, e.g., at the Serbian site near the Sava River.

A more physical description of the recharge process combined with a subsurface representation as in AquiMod can help to improve the recharge estimates (Collenteur et al., 2021; Peterson and Fulton, 2019). Depending on the available information, the choice of representation of the saturated zone may be arbitrary and parameter values may be hard to check. For the AquiMod model it was found that for several time series it was difficult or not possible to obtain a calibrated model (where the other models succeeded). In some cases, this was related to special conditions at the site that could not be simulated by the model, e.g., river influence in Serbia. In other cases, the model seems to be very sensitive to gaps in the dataset, and the quality of the input/calibration data. This meant in some cases that obtaining a calibrated model was only possible when using part of the dataset, e.g., in Italy. Obviously, the lack of a snow module means that the model is less appropriate to apply in snow dominated areas, especially for projections into future climate. It is the only model of the suite tested that allows the modeller to directly incorporate (hydro)geological understanding of the area, and to investigate inter-aquifer leakage to deeper aquifers. The model was also found to have good predictive capability regarding groundwater levels, as shown in the case of Spain.

The deviations and correlations to local and national models were acceptable for AquiMod, however despite the inclusion of a physical description of the subsurface, AquiMod did not always assure the simulation of meaningful recharge estimates in spite of good calibration performance for groundwater levels. This is due to the non-uniqueness of the problem, where equally good groundwater level performance can be obtained by scaling the recharge estimate and/or aquifer storage, as the water balance is not contained in the modelling framework. This issue may to some degree be addressed by investigating all successful AquiMod runs during a calibration, providing a variation in the recharge estimate that may be further constrained by prior or expert knowledge on recharge magnitudes (other tools, simple groundwater flow (Darcy) evaluation, field measurements). It is therefore found that it is preferable to have an additional tool for comparison, when applying AquiMod for recharge estimation. When these limitations are kept in mind, AquiMod may be used as an informative tool for recharge estimation, where percentiles (25th percentile, 50th percentile, 75th percentile) of the estimated recharge values can be calculated based on the accepted model runs.

The non-uniqueness can only be eliminated if the water balance is directly accounted for in the model. This is possible in GARDÉNIA that can use surface water discharge observations to constrain the recharge, depending on knowledge of the surface area of the catchment and the other terms of the water balance. Determining the true magnitude of the water balance terms for a lumped catchment may be challenging. This is especially the case in lowland areas where groundwater and surface water catchment areas may not coincide. Also, in wetlands where there may be multiple surface water connections with varying flow directions. For well-defined groundwater systems, boundary fluxes have little influence if the area is large enough, but the relative importance of groundwater abstractions depends on the local situation. For karstic system, GARDÉNIA offers different options depending on the behaviour of the basin. There is the possibility of adjusting the area of the watershed or introducing external flows inside the hydrological system or even introducing an external exchange factor. For this, a robust conceptual model is needed. The size of the contribution area of groundwater recharge is generally a very sensitive parameter in GARDÉNIA calculations. For the different modelled sites in this study, only one rain gauged station was used. But GARDÉNIA does offer the possibility of adjusting the climatic input parameters (rain and PET) to correct the deviation especially for large surface basins.

Generally, the GARDÉNIA had relatively low bias and fairly good correlation to the national and local models, however like for AquiMod, application of GARDÉNIA is problematic without river discharge information to guide the storage term. This was observed at various sites, where low recharge, to large surface runoff estimates, was compensated by a change in the storage coefficient. The joint use of groundwater level and discharge data allows a more robust model less subject to the problem of equifinality. And investigating the performance of the simulations where discharge was included in the objective function compared to the national and local model showed a similar correlation as for all simulations but a clear improvement in bias. Thus, GARDÉNIA is a robust and easy-to-use model, especially with a good knowledge of the hydrological system and reliable data, and calibration is quite straightforward with GARDÉNIA because of the automatic calibration.

The study showed very varying estimates of recharge at the different sites, and as the true value is unknown, it is not possible to firmly identify the model giving the correct recharge. Nevertheless, GARDÉNIA was closest to the available national local estimates for the sites especially when discharge was included in the calibration. Thus, the dataset is too limited to draw any firm conclusions and the study should be regarded as the initial steps in identifying the usefulness and precision of these tools for different settings. Additionally, it is found that using long time series (>20 years) is preferable to improve the confidence in the model results.

Generally, results points to that the time series models are especially useful for locations were other factors like abstraction or surface waters influence the water table observations, but that very complex and heterogeneous study sites such as karstic areas are problematic/difficult to simulate with all models, and here more detailed integrated modelling is needed.

4.2. Can adequate prediction of groundwater level guarantee correct recharge estimation?

The study shows very varying estimates of recharge at the different sites for models that all have very good performance during calibration. Thus, it is possible to reproduce the observed behaviour of groundwater levels but to perform poorly in the calculation of recharge. This may be due to failure of the predictive power or non-uniqueness of the model as mentioned above, or due to application in hydrological or geological settings that they cannot appropriately represent (limitations in model structure). The model structure deficiency may get compensated by hydraulic parameter values that are not representative of the hydrogeological system characteristics. Consequently, the user knowledge of the complexity of the hydrogeological system can therefore limit the robustness of the model output, and errors are easily made and difficult to identify if only one model is applied. Therefore, applying more than one model on a case is recommend as also mentioned by e.g., de Vries and Simmers (2002).

This study was limited to three, conceptually very different, groundwater level models used to estimate groundwater recharge.

However, while several similar and different approaches exist, results from this study may not always be directly transferable. Although the study covers a large and diverse selection of cases in Europe, it was not possible to cover all climate types and geological settings. A valuable next step in this process would therefore be to expand this study by including a large set of tools and sites, and thereby, developing a formulation of a rigorous scheme for the selection of appropriate recharge estimation tools based on overall study site/data characteristics. This would be especially useful if the different models and approaches can be integrated and benchmarked against each other and additional independent methods to reduce the overall uncertainty in recharge estimation procedures, as demonstrated by Crosbie and Rachakonda (2021), where recharge from chloride mass balance estimates were constrained using remote sensing.

The calibration performance of the AquiMod and GARDÉNIA models, and whether the models are deemed acceptable/adequate or not, are in this case based on expert judgement and best practice, even though the applied methodology relies on an automated approach to optimise the solution. For example, the election of a stricter or relaxed performing measure threshold may lead to different model results being used. The evaluation of the model is, therefore, also a contributor to the uncertainty of the results. It is possible that if water balance could somehow be constrained in the AquiMod model more runs would succeed, as the optimization results could be a different parameter solution.

In this study, the reason for a failure of any of the models to provide reasonable results is sought in specific conditions/quality/setting of the data and study site. However, as is true for all modelling work, uncertainties are also depended on the choices, approach, and experience of the modeller. Thus, combining the scheme for the selection of appropriate tools with a more elaborate framework for evaluating whether or not when a model's performance is acceptable or not, is a necessary next step when moving forward.

4.3. Recharge taxonomy

In this study, we have tried to divide recharge into different recharge types in order to compare the same type of recharge with each other. The taxonomy adopted here is coarse and may not always be applicable. As mentioned in Section 4.1, due to small-scale processes it may not always be possible to make a clear distinction between potential recharge and recharge. Also, in multiple aquifer systems recharge and inter-aquifer recharge may occur at multiple depths, and the depth of the aquifer is, in that case, important.

4.4. Limitations in the comparison to other methods

A major limitation in the study is the fact that no "true" values for recharge are known at any of the sites. Recharge estimations from the three models can therefore only be compared to other modelling (numerical or machine learning) results, which may not represent the correct recharge at the site and are therefore uncertain. Additional conditioning of the modelling results could be undertaken by obtaining recharge from independent estimations by in situ measurements, like chloride mass balance based on deposition from the atmosphere (Custodio, 2010) where a long term recharge can be established or age tracers. These measurements were unfortunately not available at the study sites but could potentially be used for evaluating model performance when compared to model estimates. Other methods like lysimeters also allow direct measurement of recharge. However, depending on the location of the lysimeter and the variability of the landscape around it, the transferability of the measured values to regional water balance or groundwater model will vary because of the typical variability of soil cover and land use (Fitts, 2013; Witte et al., 2019). The representativity of the measured recharge also varies because of the impact of some land uses and the associated soil treatments on lysimeter functioning (Voortman et al., 2015).

Furthermore, modelling results from local or national models are not available at all sites. Studies like the investigation done in the Netherlands for the Metran model utilizing a very high number of sites, could be a potential way to improve and test the reliability of the models, even though such a study will be extremely time consuming for models like GARDÉNIA and AquiMod, where knowledge of the geology and hydrology are essential to build into the setup.

5. Conclusions

This study compared and analysed results and limitations of three conceptually different groundwater time series models for recharge estimation (Metran, AquiMod and GARDÉNIA) explored across different climatic, hydrological, and geological settings. To facilitate this comparison a recharge taxonomy was developed, where recharge is defined in three types: 1. Potential recharge; 2. Recharge (reaching the phreatic aquifer); 3. Inter-aquifer leakage.

In lack of true measurements of recharge, the groundwater level recharge models are compared to estimation by local and national models for the different sites. Here it was shown that AquiMod and GARDÉNIA show deviations of 38% and 46%, while Metran was somewhat higher 90%, the correlation coefficient were similar for the three models, but highest for GARDÉNIA, and here the bias greatly improved when incorporating discharge in the simulation. The transfer noise model (Metran), based on groundwater head measurements driven by precipitation and evapotranspiration, was shown to be a powerful model for assessing groundwater levels, data noise and for incorporating site-specific conditions. But due to the high bias in combination with a poor correlation seen for a larger sample of Dutch boreholes, it was found that such Transfer Function-Noise models are potentially inaccurate recharge estimation tools and should be used with care.

The groundwater level driven hydrological model (AquiMod) and the lumped catchment models (GARDÉNIA), both based on climate input, and calibration of governing parameters to hydraulic head measurements, were generally shown to be successful at simulating groundwater levels. However, it was shown that special care must be taken, when estimating recharge with a model setup without water balance constraints (discharge data), as in AquiMod. The non-uniqueness means that the best procedure when using these types of models would be to use an ensemble of successful model runs and thereby include an estimation of uncertainty of the recharge estimate and/or confine results with alternative information on recharge magnitudes.

It was found that when water balance is accounted groundwater levels time series models can be very relevant tools for relatively simple study sites where boundary conditions (abstraction, river) affect head fluctuations, but for complex and heterogeneous sites more detailed integrated modelling is needed. Furthermore, in light of the uncertainties of the methods, applying multiple methods on a site is recommended.

Overall, as there is generally a lack of on-site recharge measurements for validation of these tools, a multiple tools approach is recommended to capture uncertainty ranges, when estimating recharge from groundwater heads. The large uncertainty of the recharge estimated by any of the tools must, thus, be considered when applying any of these models for recharge estimation or future prediction.

Funding Information and Conflicts of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest. Mansour publishes with the permission of the Executive Director of the British Geological Survey (UKRI). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 731166, the participating geological survey organisations and Innovation Fund Denmark under agreement no. 8055–00073B.

CRediT authorship contribution statement

I.K. Seidenfaden: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **M. Mansour, H. Bessiere, D.P. Pulido-Velázquez, W.J. Zaadnoordijk**: Conceptualization, Writing – original draft, Writing – review & editing, Methodology. **K. Atanskovic Samolov, L. Baena-Ruiz, H. Bishop, B. Dessì, K. Hinsby, N.H. Hunter Williams, O. Larva, L. Martarelli, R. Mowbray, A.J. Nielsen, J. Öhman, T. Petrovic Pantic, A. Stroj, P. van der Keur**: Writing – review & editing, Methodology. **P. van der Keur**: Writing – review & editing, Project administration. **A. Højberg**: Conceptualization, Writing – review & editing, Methodology, Project administration.

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.

Data availability

Data will be made available on request.

Acknowledgements

The authors wish to thank the European Union's Horizon 2020 for providing funding for the project. A special thanks to Grith Martinsen and Simon Stisen for supplying data from the Pan European Recharge map.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2023.101399.

References

Alley, W.M., Healy, R.W., LaBaugh, J.W., Reilly, T.E., 2002. Flow and storage in groundwater systems. Science (296), 1985–1990. https://doi.org/10.1126/science.1067123.

Amraoui, N., Sbai, M.A., Stollsteiner, P., 2019. Assessment of Climate Change Impacts on Water Resources in the Somme River Basin (France). Water Resour. Manag. 33, 2073–2092. https://doi.org/10.1007/s11269-019-02230-x.

Amraoui N., Wuilleumier A., Thiery D., Caous J.Y., Noyer M.-L., Gaudfefroy M.J., 2004, Mise à jour du modèle des hautes eaux de la Somme (Update of the high water model of the Somme, in French). Report BRGM/RP-53211-FR.

Baena-Ruiz, L., Pulido-Velazquez, D., 2020. A novel approach to harmonize vulnerability assessment in carbonate and detrital aquifers at Basin Scale. Water 12 (11), 2971. https://doi.org/10.3390/w12112971.

Baena-Ruiz, L., Pulido-Velazquez, D., Collados-Lara, A.J., de Dios Gómez, J., 2021. A preliminary lumped assessment of pollution risk at aquifer scale by using the mean residence time. analyses of potential climate change impacts. Water 13 (7), 943. https://doi.org/10.3390/w13070943.

Barthel, R., Stangefelta, M., Giese, M., Nygren, M., Seftigen, K., Chen, D., 2021. Current understanding of groundwater recharge and groundwater drought in Sweden compared to countries with similar geology and climate. Geogr. Ann.: Ser. A, Phys. Geogr. 1–23. https://doi.org/10.1080/04353676.2021.1969130.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Adv. Water Resour. 16 (1), 41–51. https://doi.org/10.1016/0309-1708(93)90028-E.

Bouanani, R., Baba-Hamed, K., Bouanani, A., 2012. Utilisation d'un modèle global pour la modélisation pluie-débit: cas du bassin d'Oued Sikkak (NW algérien) (Use of a global model for rainfall-flow modelling: case of the Oued Sikkak basin (NW Algeria) in French). Nat. Technol. C. - Sci. De. l'Environ. 5 (2), 61–66.
Brkić, Ž., 1999, Alluvial aquifer recharge through semipermeable covering deposits in northern Croatia. PhD thesis (in Croatian), University of Zagreb, p. 157.
Brkić, Ž., Larva, O., Kuhta, M., 2021. Groundwater age as an indicator of nitrate concentration evolution in aquifers affected by agricultural activities. J. Hydrol. 602. https://doi.org/10.1016/j.jhydrol.2021.126799.
Burnash, R.J.C., Ferral, R.L., Maguire, R.A., 1973, A Generalized Streamflow Simulation System: Conceptual Models for Digital Computers; Joint Federal State River Forecast Center: Sacramento, CA, USA.
Collados-Lara, A.-J., Pardo-Igúzquiza, E., Pulido-Velazquez, D., Baena-Ruiz, L., 2021. Estimation of the monthly dynamics of surfacewater in wetlands from satellite and secondary hydro-climatological data. Remote Sens. 13 (12), 2380. https://doi.org/10.3390/rs13122380.
Collenteur, R.A., Bakker, M., Klammler, G., Birk, S., 2021. Estimation of groundwater recharge from groundwater levels using nonlinear transfer function noise models and comparison to lysimeter data. HESS 25, 2931–2949. https://doi.org/10.5194/hess-25-2931-2021.
Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M., Jones, P.D., 2018. An ensemble version of the E-OBS temperature and precipitation data sets. J. Geophys. Res. Atmos. 123, 9391–9409. https://doi.org/10.1029/2017JD028200.
Crosbie, R.S., Binning, P., Kalma, J.D., 2005. A time series approach to inferring groundwater recharge using the water table fluctuation method. Water Resour. Res. 41, W01008 https://doi.org/10.1029/2004WR003077.
Custodio, E.G., 2010. Estimation of aquifer recharge by means of atmospheric chloride deposition balance. Contrib. Sci. 6, 81–97.
Custodio E.G. Recarga natural a los acuíferos, metodología y soporte de la isotopía del agua: aplicación a la planificación hidrológica y conocimiento de las aguas subterráneas en España: informe RAEMIA. Iniciativa Digital Politècnica Oficina de Publicacions Acadèmiques Digitals de la UPC 2019.
de Lange, W.J., Prinsen, G.F., Hoogewoud, J.C., Veldhuizen, A.A., Verkaik, J., Oude Essink, G.H.P., van Walsum, P.E.V., Delsman, J.R., Hunink, J.C., Th, H., Massop, L., Kroon, T., 2014. An operational, multi-scale, multi-model system for consensus-based, integrated water management and policy analysis: The Netherlands Hydrological Instrument. Environ. Model. Softw. 59, 98–108. https://doi.org/10.1016/j.envsoft.2014.05.009.
de Vries, J.J., Simmers, I., 2002. Groundwater recharge: an overview of processes and challenges. Hydrogeol. J. 10 (1), 5–17. https://doi.org/10.1007/s10040-001-0171-7.
Delin, G.N., Healy, R.W., Lorenz, D.L., Nimmo, J.R., 2007. Comparison of local- to regional-scale estimates of ground-water recharge in Minnesota, USA. J. Hydr. 334 (1–2), 231–249. https://doi.org/10.1016/j.jhydrol.2006.10.010.
Dos Santos, R.M., Koide, S., Távora, B.E., de Araujo, D.L., 2021. Groundwater Recharge in the Cerrado Biome, Brazil—A Multi-Method Study at Experimental Watershed Scale. Water 13 (1), 20. https://doi.org/10.3390/w13010020.
European Commission, 2000, Directive 2000/60/EC.
European Commission, 2006, Directive 2006/118/EC.
European Commission, 2009, Guidance on groundwater status and trend assessment. Guidance document no. 18, Technical Report - 2009 – 026.
FAO 1998. Crop Evapotranspiration; Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56. FAO, Rome, Italy.
Fitts, C.R., 2013, Groundwater Science, Academic Press. Waltham MA, USA, ISBN: 978–9-12–384705-8.
Flint, A.L., Flint, L.E., Kwicklis, E.M., Fabryka-Martin, J.T., Bodvarsson, G.S., 2002. Estimating recharge at Yucca Mountain, Nevada, USA: Comparison of methods. Hydrogeol. J. 10 (1), 180–204. https://doi.org/10.1007/s10040-001-0169-1.
Gleeson, T., Wang-Erlandsson, L., Porkka, M., Zipper, S.C., Jaramillo, F., Gerten, D., Fetzer, I., Cornell, S.E., Piemontese, L., Gordon, L.J., Rockström, J., Oki, T., Sivapalan, M., Wada, Y., Brauman, K.A., Flörke, M., Bierkens, M.F.P., Lehner, B., Keys, P., Kummu, M., Wagener, T., Dadson, S., Troy, T.J., Steffen, W., Falkenmark, M., Famiglietti, J.S., 2020. Illuminating water cycle modifications and Earth system resilience in the Anthropocene. Water Resour. Res 56, 1–24. https://doi.org/10.1029/2019WR024957.
Griffiths, J., Young, A.R., Keller, V. 2006, Model Scheme for Representing Rainfall Interception and Soil Moisture. EnvironmentAgencyR&D Project W6–101 Continuous Estimation of River Flows (CERF), Environment Agency, UK.
Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000, MODFLOW-2000, The U.S. Geological Survey modular groundwater model – userguide to modularization concepts and the groundwater flow process. USGeol. Surv. Open-File Rep 00–92, 12.
Healy, R.W., 2010a. Estimating Groundwater Recharge. Cambridge University Press,, Cambridge. DOI:DOI: 10.1017/CBO9780511780745.
Healy, R.W., 2010b. Estimating Groundwater Recharge. Cambridge University Press,, Cambridge, UK. https://doi.org/10.1017/CBO9780511780745.
Healy, R.W., and Cooke, P.G., 2002, Using groundwater levels to estimate recharge: Hydrogeology Journal, 10, 91- 109.
Hendrickx, J.M.H., Walker, G.R., 1997, Recharge from precipitation, in: Recharge of phreatic aquifers in (semi)arid areas, edited by: Simmers, I., AA Balkema, Rotterdam, the Netherlands.
Højberg, A.L., Troldborg, L., Stisen, S., Christensen, B.B.S., Henriksen, H.J., 2013. Stakeholder driven update and improvement of a national water resources model. Environ. Model. Softw. 40, 202–213. https://doi.org/10.1016/j.envsoft.2012.09.010.
Hunter Williams, N.H., Carey, S., Werner, C., Nolan, P., 2021. Updated national groundwater recharge map. Ir. Groundw. Newsl. 59, 31–34. https://www.gsi.ie/documents/GWNewsletterNo.59.pdf.
Jeannin, P.-Y., Artigue, G., Butscher, C., Chang, Y., Charlier, J.-B., Duran, L., Gill, L., Hartmann, A., Johannet, A., Jourde, H., Kavousi, A., Liesch, T., Liu, Y., Lüthi, M., Malard, A., Mazzilli, N., Pardo-Igúzquiza, E., Thiéry, D., Reimann, T., Schuler, P., Wöhling, T., Wunsch, A., 2021. Karst modelling challenge 1: Results of hydrological modelling. J. Hydrol. 600, 126508 https://doi.org/10.1016/j.jhydrol.2021.126508.
Kalin, L., Isik, S., Schoonover, J.E., Lockaby, B.G., 2010. Predicting water quality in unmonitored watersheds using Artificial Neural Networks. J. Environ. Qual. 39 (4), 1429–1440. https://doi.org/10.2134/jeq2009.0441.
Kouchi, D.H., Esmaili, K., Faridhosseini, A., Sanaeinejad, S.H., Khalili, D., Abbaspour, K.C., 2017. Sensitivity of calibrated parameters and water resource estimates on different objective functions and optimization algorithms. Water 9 (6), 384. https://doi.org/10.3390/w9060384.
Lafare, A.E.A., Peach, D.W., Hughes, A.G., 2021. Use of point scale models to improve conceptual understanding in complex aquifers: An example from a sandstone aquifer in the Eden valley, Cumbria, UK. Hydrol. Process. 35 (5), e14143 https://doi.org/10.1002/hyp.14143.
Lindström, G., Pers, C.P., Rosberg, R., Strömqvist, J., Arheimer, B., 2010. Development and test of the HYPE (Hydrological Predictions for the Environment) model – A water quality model for different spatial scales. Hydrol. Res. 41 (3–4), 295–319. https://doi.org/10.2166/nh.2010.007.
Mackay, J.D., Jackson, C.R., Wang, L., 2014a. A lumped conceptual model to simulate groundwater level time-series. Environ. Model. Softw. 61, 229–245. https://doi.org/10.1016/j.envsoft.2014.06.003.
Mackay, J.D., Jackson, C.R., Wang, L. 2014b, AquiMod user manual (v1.0). Nottingham, UK, British Geological Survey, 34pp. (OR/14/007).
Mansour, M.M., Wang, L., Whiteman, M., Hughes, A.G., 2018. Estimation of spatially distributed groundwater potential recharge for the United Kingdom. Q. J. Eng. Geol. Hydrogeol. 51 (2), 247–263. https://doi.org/10.1144/qjegh2017-051.
Margat, J., van der Gun, J., 2013. Groundwater around the World: A Geographic Synopsis, 1st ed..,. CRC Press,, London, UK. https://doi.org/10.1201/b13977.
Martinsen, G., Bessière, H., Caballero, Y., Koch, J., Collados-Lara, J.C., Mansour, M.M., Sallasmaa, O., Pulido-Velazquez, D., Hunter Williams, N., Zaadnoordijk, W.J., Stisen, S., 2022. Developing a pan-European high-resolution groundwater recharge map – Combining satellite data and national survey data using machine learning. Sci. Total Environ. 822, 153464 https://doi.org/10.1016/j.scitotenv.2022.153464.
Meinzer, O.E., 1923. Outline of ground-water hydrology with definitions. USGS, Geol. Surv. Water-Supply Pap. 494. https://doi.org/10.3133/wsp494.
Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. Am. Soc. Agric. Eng. 50, 885–900. https://doi.org/10.13031/2013.23153.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10 (3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
Nielsen, S.A., Hansen, E., 1973. Numerical simulation of the rainfall runoff process on a daily basis. Nord. Hydrol. 4, 171–190. https://doi.org/10.2166/nh.1973.0013.

Nimmo, J.R., Horowitz, C., Mitchell, L., 2015. Discrete-Storm Water-Table Fluctuation Method to Estimate Episodic Recharge. Groundwater 53, 282–292. https://doi.org/10.1111/gwat.12177.

Obergfell, C., Bakker, M., Maas, K., 2019. Estimation of average diffuse aquifer recharge using time series modeling of groundwater heads. Water Resour. Res. 55 (3), 2194–2210. https://doi.org/10.1029/2018WR024235.

Patrčević, V., 1995, Hydrological Analysis of Vertical Water Balance of Groundwater at the Area of River Alluvium. PhD thesis (in Croatian), University of Zagreb, p. 161.

Pérez-Sánchez, J., Senent-Aparicio, J., Segura-Méndez, F., Pulido-Velazquez, D., Srinivasan, R., 2019. Evaluating Hydrological Models for Deriving Water Resources in Peninsular Spain. Sustainability 11 (10), 2872. https://doi.org/10.3390/su11102872.

Peterson, T.J., Fulton, S., 2019. Joint estimation of gross recharge, groundwater usage, and hydraulic properties within HydroSight. Groundwater 57, 860–876. https://doi.org/10.1111/gwat.12946.

Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, B., Ward, R., Jenkins, A., 2017. Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales. Hydrol. Sci. J. 62 (16) https://doi.org/10.1080/02626667.2017.1395032.

Pulido-Velazquez, D., Romero, J., Collados-Lara, A.J., Alcalá, F., Chacón, F., Baena-Ruiz, L., 2020. Using the Turnover Time Index to Identify Potential Strategic Groundwater Resources to Manage Droughts within Continental Spain. Water 12 (11), 3281. https://doi.org/10.3390/w12113281.

Rosenbrock, H.H., 1963. Some general implicit processes for the numerical solution of differential equations. Comput. J. 5 (4), 329–330. https://doi.org/10.1093/comjnl/5.4.329.

Scanlon, B.R., Healy, R.W., Cook, P.G., 2002. Choosing appropriate techniques for quantifying groundwater recharge. Hydrogeol. J. 10, 18–19. https://doi.org/10.1007/s10040-0010176-2.

Schneider, C., Laizé, C.L.R., Acreman, M.C., Flörke, M., 2013. How will climate change modify river flow regimes in Europe. Hydrol. Earth Syst. Sci. 17 (1), 325–339. https://doi.org/10.5194/hess-17-325-2013.

Sibanda, T., Nonner, J., Uhlenbrook, S., 2009. Comparison of groundwater recharge estimation methods for the semi-arid Nyamandhlovu area, Zimbabwe. Hydrogeol. J. 17, 1427–1441. https://doi.org/10.1007/s10040-009-0445-z.

Staudinger, M., Stoelzle, M., Cochand, F., Seibert, J., Weiler, M., Hunkeler, D., 2019. Your work is my boundary condition! Challenges and approaches for a closer collaboration between hydrologists and hydrogeologists. J. Hydrol. 571, 235–243. https://doi.org/10.1016/j.jhydrol.2019.01.058.

Steffen, W., Richardson, K., Rockström, J., et al., 2015. Planetary Boundaries: Guiding human development on a changing planet. Science, 1259855. https://doi.org/10.1126/science.1259855.

Stisen, S., Ondracek, M., Troldborg, L., Schneider, R.J.M., van Til, M.J., 2019, National Vandressource Model. Modelopstilling og kalibrering af DK-model 2019 (National water resources model – model construction and calibration of Denmark model 2019, in Danish). Geological Survey of Denmark and Greenland, Copenhagen, Denmark.

Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., 2012. Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale. Hydrol. Sci. J. 57 (2), 229–247.

Surge, S.L., 2018, Actualización y calibración del modelo de flujo de agua subterránea de los acuíferos del Alto Guadiana (FLUSAG) (Actualization and calibration of groundwater flow model of the aquifers of Alto Guadiana, in Spanish), Ref: TEC0004594, pp 150. Confederación Hidrográfica del Guadiana, Madrid, Spain.

TACTIC, 2021a, Pilot description and assessment of recharge and groundwater vulnerability for climate change. TACTIC Deliverable D4.2, GeoERA, ⟨https://geoera.eu/projects/tactic9/⟩.

TACTIC, 2021b, TACTIC guidelines. TACTIC Deliverable D3.5, D4.5,D5.5 & D6.5, GeoERA, ⟨https://geoera.eu/projects/tactic9/⟩.

Thiéry D. (2015a) Validation du code de calcul GARDÉNIA par modélisations physiques comparatives (validation of simulation program GARDÉNIA by comparison with physically based models, in French). Report BRGM/RP-64500-FR.

Thiéry D. (2015b) Code de calcul MARTHE – Modélisation 3D des écoulements dans les hydrosystèmes – Notice d'utilisation de la version 7.5 (simulation program MARTHE – three dimensional modelling of flow in hydrological systems – user manual version 7.5, in French). Rapport BRGM/RP-64554-FR.

Thiéry, D., Amraoui, N., NoyerML, 2018. Modelling flow and heat transfer through unsaturated chalk – validation with experimental data from the ground surface to the aquifer. J. Hydrol. 556C, 660–673.

Urumović, K., Šarin, A., Donadini, K., 1981, Groundwater Balance in the river basin district Drava-Dunav for 1979–1980 (Technical report – on Croatian). Croatian Geological Survey, Zagreb.

Voortman, B.R., Bartholomeus, R.P., van der Zee, S.E.A.T.M., Bierkens, M.F.P., Witte, J.P.M., 2015. Quantifying energy and water fluxes in dry dune ecosystems of the Netherlands. Hydrol. Earth Syst. Sci. 19 (9), 3787–3805.

Walker, David, Parkin, Geoff, Schmitter, Petra, Gowing, John, Tilahun, Seifu, Haile, A., Yimam, Abdu, 2018. Insights from a multi-method recharge estimation comparison study. Groundwater 57 (2), 245–258. https://doi.org/10.1111/gwat.12801.

Witte, J.P.M., Zaadnoordijk, W.J., Buyse, J.J., 2019. Forensic Hydrology Reveals Why Groundwater Tables in The Province of Noord Brabant (The Netherlands) Dropped More Than Expected. Water 11 (3), 478. https://doi.org/10.3390/w11030478.

Zaadnoordijk, W.J., Bus, S.A.R., Lourens, A., Berendrecht, W.L., 2019. Automated time series modeling for piezometers in the national database of the Netherlands. Groundwater volume 57 (6), 834–843. https://doi.org/10.1111/gwat.12819.