Applied Energy 406 (2026) 127268

Contents lists available at ScienceDirect

Applied Energy

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

Large-scale modeling of solar water pumps using machine learning

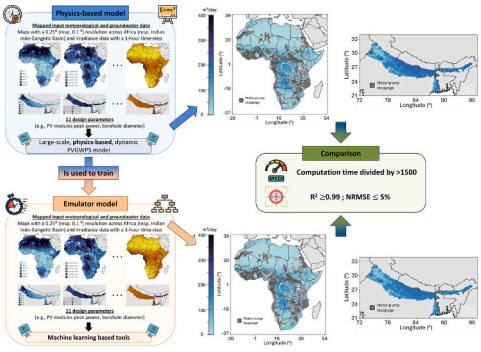
Guillaume Zuffinetti [a,b,c,d], Simon Meunier [a,b,*], Céline Hudelot [c], Donald John MacAllister [e], Gopal Krishan [f], Evelyne Lutton [g], Prosun Bhattacharya [h,i], Peter K. Kitanidis [j], Alan M. MacDonald [e]

[a] Université Paris-Saclay, CentraleSupélec, CNRS, Group of electrical engineering Paris (GeePs), 91192 Gif-sur-Yvette, France
[b] Sorbonne Université, CNRS, GeePs, 75252 Paris, France
[c] Université Paris-Saclay, CentraleSupélec, Mathematics and Computer Science Laboratory for Complexity and Systems (MICS), 91190 Gif-sur-Yvette, France
[d] Department of Sustainable Development, Environmental Science and Engineering, KTH Royal Institute of Technology, Teknikringen 10B, SE-100 44 Stockholm, Sweden
[e] British Geological Survey, The Lyell Centre, Research Avenue South, Edinburgh EH14 6AJ, Scotland, United Kingdom
[f] National Institute of Hydrology, Roorkee, Uttarakhand, India
[g] INRAE, AgroParisTech, Université Paris-Saclay, UMR MIA-PS, 22 Place de l'Agronomie, Palaiseau, France
[h] KTH-International Groundwater Arsenic Research Group, Department of Sustainable Development, Environmental Science and Engineering, KTH Royal Institute of Technology, Teknikringen 10B, SE-100 44 Stockholm, Sweden
[i] Digital Futures Faculty, KTH Royal Institute of Technology, Teknikringen 10B SE-100 44 Stockholm, Sweden
[j] Department of Civil and Environmental Engineering, Stanford University, Stanford, United States

HIGHLIGHTS

- Machine learning tools are developed to accelerate a physics-based solar pump model
- The designed emulators are applied to Africa and the Indian Indo Gangetic Basin
- These emulators reduce the computation time of the physics-based model by >1500
- These emulators achieved high accuracy: $R^2 \geq 0.99$, NRMSE $\leq 5\%$.
- They open up the way for optimizing the large-scale deployment of solar pumps

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Solar energy
Water pumping
Africa
Indo-Gangetic Basin
Machine learning
Modeling
Emulators

ABSTRACT

Photovoltaic Groundwater Pumping Systems (PVGWPSs) have experienced growing interest, particularly in two key regions. In Africa, they offer a means to improve water availability for millions. In northern India, they could help decarbonize the agricultural sector. However, large-scale deployment must be approached carefully to avoid risks such as groundwater overextraction or widespread unmet irrigation demand. To support informed deployment, a large-scale, physics-based, dynamic PVGWPS model is introduced, that simulates pumping capacities of PVGWPS. Given the computational intensity of this model, machine learning-based emulators are explored to replicate its results more efficiently without significant loss in accuracy. The emulator operates in two stages. First, it predicts whether the motor-pump will stop due to water level dropping below the operational

* Corresponding author at: Université Paris-Saclay, CentraleSupélec, CNRS, Group of electrical engineering Paris (GeePs), 91192 Gif-sur-Yvette, France.
E-mail address: simon.meunier@centralesupelec.fr (S. Meunier).

https://doi.org/10.1016/j.apenergy.2025.127268
Received 29 September 2025; Received in revised form 27 November 2025; Accepted 13 December 2025
Available online 30 December 2025
0306-2619/Crown Copyright © 2025 Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

threshold. Among the models tested, the Gradient Boosting Classifier model performed best. Second, when no stoppage is predicted, the emulator estimates the pumping capacity of the PVGWPS. Among the models tested for this second task, the Random Forest Regressor gave the most accurate results. Applied to datasets from Africa and the Indo-Gangetic Basin within India, the emulator achieved high accuracy ($R^2 \geq 0.99$, NRMSE ≤ 5 %) while reducing computation time by more than a factor of 1500. The emulators thus offer high computational speed and sufficient accuracy to open the way to addressing large-scale dispatch problems, such as the optimal positioning and pre-sizing of PVGWPSs at regional, national, or even continental scales while considering a large number of possible climate scenarios. Coupled with sustainability analyses (not explored in this study), they could serve as powerful upstream decision-support tools for PVGWPSs planning, complementing more detailed, site-specific analyses.

1. Introduction

1.1. Literature review

Photovoltaic groundwater pumping systems (PVGWPSs) have garnered significant attention over the past decade, with over 1.3 million PVGWPSs installed globally by 2023 [1]. There has been particular interest in these systems in Africa and in South Asia.

In Africa, an estimated 400 million people still lack access to basic drinking water service [2], and less than 10 % of the continent's cultivated land is irrigated [3]. Although surface water is often shallower and more economical to extract, groundwater represents the largest source of freshwater on the continent, albeit variably distributed [4]. Moreover, unlike surface water, groundwater often requires no treatment [4,5] and its slower response to meteorological changes makes it a natural buffer against climate variability [6–8]. While careful management of natural resources is essential [3], groundwater can help meet the increasing demand for domestic and agricultural water across the continent [9]. In this context, PVGWPSs, which harness affordable and low carbon photovoltaic energy, offer a promising solution to enhance water access in off-grid areas. These systems are already economically viable in various contexts [10], with technological advancements enhancing their durability [11], and local case studies demonstrated encouraging outcomes [12,13]. Consequently, the development of this technology is actively promoted across the continent by governments and international institutions such as the World Bank or UNICEF [14,15].

In South Asia, the Indo-Gangetic Basin (IGB), home to approximately 15 million groundwater pumping systems (GWPSs) for irrigation [16], is one of the most intensively cultivated regions globally [17,18]. With annual groundwater extraction for irrigation surpassing 200 km^3 [19], the region accounts for about 20 % of global groundwater withdrawal [20]. This extraction is particularly concentrated in Punjab and Haryana where, despite a century of groundwater accumulation [21], groundwater is currently over exploited [19]. The region's GWPSs are primarily powered by diesel or connected to a carbon-based grid [16], making them substantial contributors to greenhouse gas emissions [22]. For instance, the 20 million grid-connected pumps and 10 million diesel pumps in India are estimated to emit over 200 million tons of CO_2 annually, representing more than 5 % of the country's total CO_2 emissions [23]. Given the imperative of SDG 13, which calls for the decarbonization of all economic sectors [24], a low-carbon alternative to these traditional pumping systems is required. In response, the Indian Ministry of New and Renewable Energy (MNRE) initiated a solar pumping program in 1992 to convert existing GWPSs into PVGWPSs [10]. By 2023, India has emerged as a global leader in PVGWPS installation, with over one million systems deployed for agricultural use [1]. However, despite this progress, the adoption rate within the Indian part of the IGB region remains modest, with less than 5 % of GWPSs converted to PVGWPSs [25].

Therefore, PVGWPSs hold significant potential for advancing sustainable development in both Africa and South Asia. However, the higher initial capital cost of PVGWPSs compared to that of conventional GWPSs [16,18] slow their widespread adoption, despite their lower lifecycle costs [22]. Consequently, widespread subsidies are often required to facilitate the large-scale deployment of PVGWPSs [26,27]. Nevertheless, by facilitating access to free solar energy, these subsidies could also exacerbate groundwater over-exploitation [3,28]. Thus, to maximize the impact of such subsidies, it is essential to deploy PVGWPSs in regions where their pumping capacities can align with local groundwater demands, while also safeguarding against over-exploitation. To support this effort, it is important to develop large-scale PVGWPS models that can quantitatively assess their pumping performance and identify where they are most effective.

Several PVGWPS models have been proposed to investigate the potential of PVGWPSs over different large-scale geographical areas: in Ethiopia [29], Ghana [30], Egypt [31,32], Algeria [33], China [34–36], Spain and Morocco [37], in the Sahel [38], in sub-Saharan Africa [39,40] or even for the whole African continent [41]. However, they remain limited in several ways. Many of these studies [29–32,36,38] do not model the operation of PVGWPSs, which prevents them from quantitatively accounting for the solar and hydrogeological resources. Campana et al. [34], Rubio-Aliaga et al. [37], Falchetta et al. [39], and Xie et al. [40] have quantified the pumping capacity of PVGWPSs by modeling the energy system using physics-based PVGWPS models. Nevertheless, they opt for monthly average irradiance figures rather than hourly or sub-hourly time series data, which reduces the model's ability to accurately estimate the abstractable groundwater volume of PVGWPSs [41,42]. Analyses from [34,37,39] also only consider the depth to the water table without modeling the drawdown at the pump, despite its significant influence on pumping capacities [41]. While simulating the drawdown, Xie et al. nevertheless do not consider the saturated thickness of the aquifer and the depth of the motor-pump [40], although these impose limits on the maximum possible drawdown and consequently the pumping flow rate [41]. Another study uses sub-hourly average irradiance values and considers the saturated thickness and the motor-pump depth [41]. It also highlights that, in certain locations (notably where the transmissivity is low), the PVGWPS regularly stops due to the water level in the borehole reaching the motor-pump. However, it uses a steady-state model that does not account for the specific yield and the time-dependent dynamics of pumping [43]. Moreover, it takes 10 h to [41] to compute the pumping capacities of PVGWPSs at 62500 locations across Africa. Such a model is thus considered too time-consuming to be used for large-scale decision-making processes, such as the optimal positioning and pre-sizing of PVGWPSs at regional, country or even continental scale, especially if multiple climate scenarios and their associated uncertainties are to be considered. These analyses could nonetheless support large-scale investments in PVGWPS projects, making them valuable to governments and funding organizations. Machine learning tools could help address this computing time challenge. However, to our knowledge, only Haddad et al. [44] have used machine learning (specifically, Regression Neural Networks) to predict the pumping capacity of a PVGWPS. Nevertheless, their study did not consider hydrogeological factors and focused on a single location, preventing the generalization of their findings to other locations.

1.2. Research gap and question, contributions and method overview

Review of the literature therefore reveals that there is no PVGWPS

model at large-scale (e.g., regional or continental) that is both computationally efficient and accurate. Our hypothesis is that this research gap can be addressed by developing machine learning-based emulators trained on the outputs of detailed physics-based dynamic models. Consequently, our research question is: can the combination of physics-based modeling and machine learning allow to simulate PVGWPS pumping capacities at large-scale in a significantly reduced computing time while maintaining high accuracy?

To answer this question, a two-step method is proposed in this article. First, a large-scale, physics-based, dynamic PVGWPS model to simulate the pumping capacities of PVGWPSs is developed. Second, machine learning-based emulators are designed to reproduce the results of the physics-based model at a much faster rate. Publicly accessible meteorological and groundwater input data are employed to apply the models across Africa and the Indian IGB. A visual overview of the approach is shown in the graphical abstract of this article.

Therefore, the first novel contribution of this work lies in the detailed design of a large-scale, physics-based and dynamic PVGWPS model which considers the specific yield, the saturated thickness, and hourly irradiance values. The second and main contribution of this work is the development of large-scale emulator models based on machine learning tools to improve the calculation time of the physics-based model without significantly reducing the quality of its estimation. Together, these contributions directly address the main identified research gap by providing a computationally efficient and accurate framework for the large-scale simulation of PVGWPS performance. Finally, the last contribution is the presentation of results spanning the Indian IGB, where quantitative large-scale PVGWPS models had never been applied.

The results are provided for five PVGWPS sizes, with a particular focus on PVGWPSs of 3000 W_p, which represents a typical PVGWPS size for irrigation [10,27,45]. Presenting quantitative results for large-scale areas facilitates the comparison of different regions and helps identify zones with the greatest potential for PVGWPSs. Moreover, the ability offered by the developed emulators to simulate the pumping performance of PVGWPSs at large-scale in a strongly reduced computing time, that must be combined with sustainability analyses (not covered in this article, see note[1]), helps pave the way for the development of upstream strategies to optimally position and pre-size PVGWPSs at wide (e.g., regional, country) scale ahead of local-scale implementation approaches. The method is detailed in Section 2. The results are shown in Section 3 and are discussed in Section 4.

2. Materials and methods

In this section, PVGWPSs are first described in Section 2.1. Secondly, the input location-dependent data required for this study are presented in Section 2.2. Then, a physics-based model to estimate the groundwater volume abstractable by a PVGWPS across Africa and the Indian IGB is proposed in Section 2.3. Finally, in Section 2.4, machine learning-based emulators are developed.

2.1. Photovoltaic groundwater pumping systems

In rural areas, PVGWPSs have emerged as a low-carbon and cost-effective solution for enhancing water access both for domestic and irrigation use [49]. The simplest architecture for a PVGWPSs includes photovoltaic (PV) modules, a controller, and a motor-pump. Some configurations include storage options like batteries or water tanks to reserve energy or water for periods of low sunlight, albeit at the cost of increased system complexity and expense [50,51]. The selection of the motor-pump type depends on the water source [52]: surface pumps are commonly used for extracting water from streams, rivers, or shallow groundwater (less than 7 m deep), whereas submersible pumps access deeper groundwater, which is more resilient to climate variability and less susceptible to surface contamination [6,8].

The off-grid submersible PVGWPS architecture considered for this study is presented in Fig. 1a. This architecture is common for groundwater abstraction with PV energy [12,53]. The motor and the pump are built-in together [10] and the motor-pump set is submersed in the borehole [54]. Control equipment is also installed between the PV modules and the motor-pump and/or directly integrated to the motor-pump set [10,55]. This equipment allows the motor-pump to stop and also to operate the motor-pump and the PV modules at their best operating points [10]. In the case of a surface pump (see Fig. 1b), the only difference is the motor-pump being placed above ground, lifting water by suction. This method prevents the motor-pump from lifting water beyond 7 m. Therefore, in this article, surface pumps and submersible pumps are modelled by the same equations, and will not be studied separately, the only difference being the pumping depth H_p which cannot be higher than 7 m in the case of a surface pump. In addition, in the following, the term "pumping depth" H_p is used for both submersible and surface pumps and represents the maximum depth at which water can be pumped. In this article, generic PVGWPSs are considered, with the size of the motor-pump proportional to the peak power of the photovoltaic modules. The PV modules peak power is therefore used as a proxy for the size of the PVGWPSs.

Two operating modes can be distinguished:

- The first operating mode is when the water level in the borehole H_b never reaches the pumping depth H_p, so it does not cause the motor-pump to stop. This is the most common operation for a PVGWPS, and is referred to, in this study, as 'no motor-pump stoppage'.
- The second operating mode is when the water level in the borehole H_b reaches the pumping depth H_p at least once during the simulation period. When it occurs, it typically happens around midday on sunny days, as this is when the electrical power from the PV modules, the pumping flow rate and, therefore, the drawdown are at their highest [41]. Additionally, this situation is generally encountered in locations with low transmissivity, which leads to increased drawdown and a higher risk of the water level falling below the motor-pump or pump intake [41,56]. This is referred to as 'motor-pump stoppage' in this study.

2.2. Input location-dependent data

To fulfill these objectives, input location-dependent data for Africa and the Indian IGB are used and summarized in Table 1. This table notably shows that the input datasets have different spatial resolutions. For the rest of the article, for each region, the resolution of the irradiance maps, 0.25° for Africa and 0.1° for the Indian IGB, is used. They are considered sufficient for the large-scale analysis carried out in this study. These resolutions of 0.25° and 0.1° were applied to all input datasets for their respective regions using nearest neighbor interpolation via the Rasterio library in Python [57]. Datasets originally in shapefile format were rasterized at the same resolutions also using Rasterio.

For the African datasets, for the static water depth $H_{b,s}$, the transmissivity T, and the saturated thickness H_{st}, the original source provides only value ranges, not exact values, for each location. The midpoint of each range is generally used, except in the following cases: if $H_{b,s}$ exceeds 250 m, 300 m is considered (same for H_{st}); if $H_{b,s}$ falls between

[1] Studying the sustainability of PVGWPS development, particularly across the Indian IGB which already counts numerous over-exploited locations, is crucial. However, the complexity of the mechanisms involved, such as the high seasonality of groundwater recharge and levels [46], spatial heterogeneity of the resource [47], interactions with major rivers [48], and the distinction between natural and artificial recharge [19], makes it challenging to also address sustainability in this single study. Consequently, this study focuses on the pumping capacities of PVGWPSs, while future studies will concentrate on studying the sustainability of the large-scale deployment of PVGWPSs.

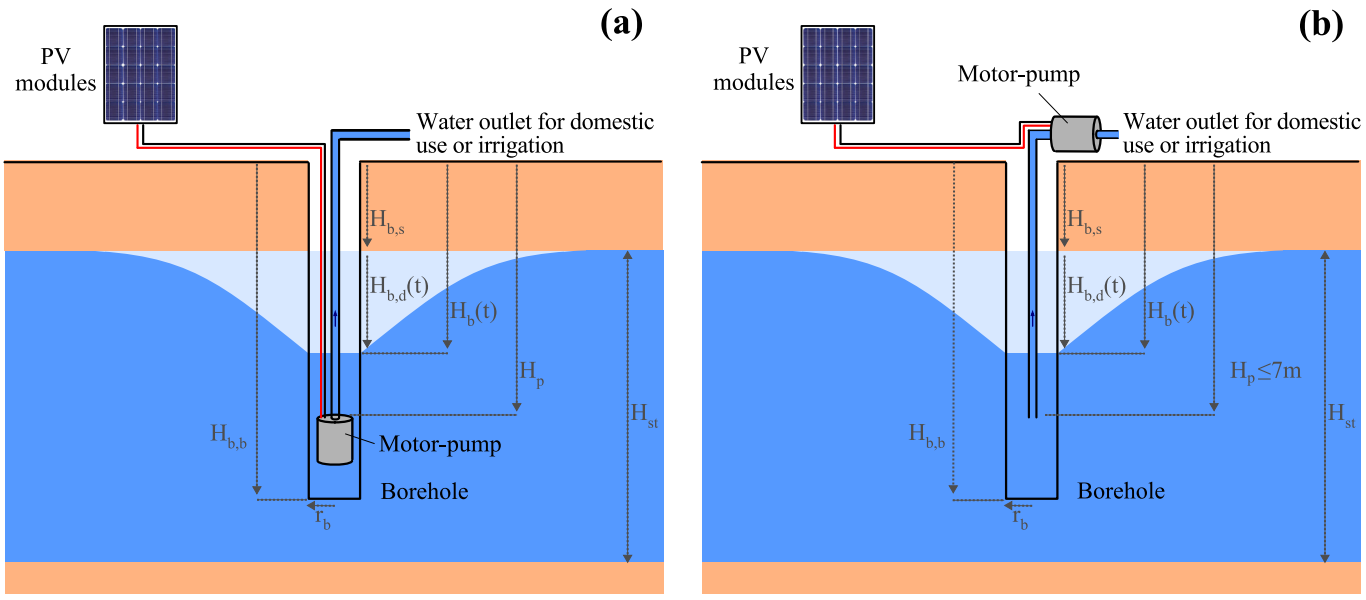


Fig. 1. Architectures for (a) a submersible and (b) a surface photovoltaic groundwater pumping system (PVGWPS). Adapted from [41]. Abbreviations: r_b: radius of the borehole; $H_{b,s}$: static water depth (corresponding to the water depth in the borehole when there is no pumping); $H_{b,d}$: drawdown; H_b: water depth in the borehole; H_{st}: saturated thickness of the aquifer; $H_{b,b}$: borehole depth; and H_p: pumping depth. All the lengths are defined as positive.

Table 1
Input location-dependent data for Africa/Indian IGB.

Data	Symbol	Description	Spatial resolution	Temporal resolution and coverage	Year of release	Provider
Annual mean static water level (m)	$H_{b,s}$	Depth of water in the borehole when there is no pumping.	0.05° / 0.04°	One value for each location	2012/ 2025	British Geological Survey [56] / India-WRIS (Water Resource Information System) [59]
Aquifer saturated thickness (m)	H_{st}	Vertical thickness of the hydrogeologically defined aquifer unit in which the pore spaces are saturated with water	0.01° / shapefile		2012/ 2016	British Geological Survey [4,56] / British Geological Survey [19,58]
Aquifer transmissivity (m^2/s)	T	Rate at which groundwater flows horizontally through an aquifer	0.05° / shapefile			
Specific yield (–)	S_y	Equal to the effective porosity, which is the porosity of a rock available to contribute to fluid flow through the rock	0.25°/ shapefile			
2 m temperature (°C)	T_a	Ambient temperature 2 m above the ground level	0.25° / 0.1°	One temporal vector for each location.	2024	ERA5 [60] / ERA5-Land [61]
Surface solar radiation downwards (J/m^2)	SSRD	Amount of solar radiation reaching the surface of the Earth.	0.25° / 0.1°	Data for 2024 with a time step of 1 h		

Note: When the cell format is *a / b*, *a* refers to the African datasets, and *b* refers to the Indian IGB datasets.
Note: Meteorological data from 2024 are used in this study; however, multi-year datasets are available to support multi-year planning.

0 and 7 m, 7 m is used [41]. For the aquifer transmissivity across the Indian IGB, only ranges of aquifer conductivity across 7 groundwater typologies were available [58]. For each typology, the midrange value is chosen and is multiplied by the aquifer thickness to obtain the aquifer transmissivity at each location across the Indian IGB [4]. For the static water level across the Indian IGB, 6 classifications were provided [59]. The midrange value is chosen for each classification, except for the highest one (“>40 m”) which is taken equal to 50 m.

Due to the lack of available information, the input groundwater data provided in Table 1 is considered to remain constant over time. In Supplementary fig. 1 (see Appendix A), the annual mean static water level $H_{b,s}$, the saturated thickness H_{st}, the aquifer transmissivity T, the specific yield S_y, the annual mean of the ambient temperature at 2 m T_a, and the annual mean of the irradiance on the plane of the PV modules G_{PV} are plotted for each region (see Section 2.3 for information on the calculation of the irradiance on the plane of the PV modules G_{PV} based on the surface solar radiation downwards SSRD).

2.3. Physics-based model

In this section, the methodology for one pixel is presented. Note that the term “pixel” is used interchangeably with the term “location” in this study. The methodology is then the same for every pixel.

The groundwater volume V_p abstractable by a single off-grid PVGWPS during a certain period can be estimated thanks to the flow rate profile at the pump during this period. Thus, V_p is given by:

$$V_p = \int_{\text{period}} Q(t)dt \quad (1)$$

where Q is the pumping flow rate during the period. The pumping flow rate of the PVGWPS at time t can be determined by solving the following equation [41,62]:

$$\forall t, \quad (\beta + \nu \cdot L_p + K) \cdot Q(t)^3 + H_{b,s} \cdot Q(t) + H^a_{b,d}(t) \cdot Q(t) - \frac{P(t) \cdot \eta_{mp}}{\rho \cdot g} = 0 \qquad (2)$$

where β is the borehole losses coefficient, L_p the pipe lengths (taken equal to the pumping depth in this article), ν the linear pipe losses coefficient, K the junction losses coefficient, $H^a_{b,d}$ is the head loss due to aquifer losses, η_{mp} the efficiency of the motor-pump, ρ the water density (equal to 1000 kg/m^3) and g the gravitational acceleration (equal to 9.81 m/s^2). Typical ranges for the parameters are provided in Table 2. The power supplied to the pump changing every hour (due to hourly irradiance values), the head loss due to aquifer losses $H^a_{b,d}$ can be expressed as [63]:

$$H^a_{b,d}(t) = \frac{1}{4\pi T} \int_0^t Q(\tau) \cdot \frac{e^{-\frac{S_y \cdot r_b^2}{4T(t-\tau)}}}{t - \tau} d\tau \qquad (3)$$

where r_b is the borehole radius. $H^a_{b,d}$ is the drawdown solution from the Theis equation considering an unsteady flow rate [63]. This solution implies confined aquifer conditions, which might not always represent the actual field conditions. In unconfined aquifer conditions, similar equations could be implemented [64]. Nevertheless, the findings from [65] highlight that the results in terms of abstractable volume assuming unconfined conditions would be very similar (in [65] the abstractable volume differs on average by less than 5 % between confined and unconfined conditions). In addition, to our best knowledge, there are no maps indicating whether aquifers are predominantly confined or unconfined across the different parts of the considered areas, which prevents a consistent spatial application of different formulations. Thus we found sufficient to use only Eq. (3) for this large-scale study. Considering that the maximum power point tracking of the PV modules is properly achieved, the following model is used to calculate the power P produced by the modules [34,66]:

$$P(t) = \frac{G_{PV}(t)}{G_0} \cdot P_P \cdot \left(1 + \gamma \cdot \left(T_a(t) + \frac{NOCT - 20}{800} \cdot G_{pv}(t) - 25\right)\right) \qquad (4)$$

where G_0 is the reference irradiance (1000 W/m^2), P_P the peak power of the PV modules, γ the loss coefficient related to PV modules temperature, T_a the ambient temperature, and NOCT the nominal operating cell temperature. G_{PV} is the irradiance on the plane of the PV modules, which is maximized by appropriately setting the azimuth and tilt angles of the PV modules according to the location. To this end, the azimuth angle of the PV modules α is set to 180° for locations in the northern hemisphere and 0° for those in the southern hemisphere [67]. The tilt angle of the PV modules is set to [67]:

$$\theta = \begin{cases} \max(10, 1.3793 + (1.2011 + (-0.014404 + 0.000080509\phi)\phi)\phi) & \text{if } \phi > 0 \\ \min(-10, -0.41657 + (1.4216 + (0.024051 + 0.00021828\phi)\phi)\phi) & \text{if } \phi < 0 \end{cases} \qquad (5)$$

where ϕ is the latitude of the location. Eq. (5) also ensures that the absolute tilt angle of the PV modules is always greater than 10°, providing sufficient inclination for effective cleaning by rainfall [41]. Given these angles, and after converting SSRD data into Global Horizontal Irradiance (GHI), Direct Horizontal Irradiance (DHI), and Direct Normal Irradiance (DNI) thanks to the PVlib python library, PVlib then estimates the irradiance on the plane of the PV modules G_{PV} [68]. An albedo of 0.2 is assumed, representative of cropland, a typical land cover in the rural areas under study [69]. A map of the annual mean of G_{PV} for each region is shown on Supplementary fig. 1f and ll (see Appendix A).

To implement Eq. (3) numerically, it is discretized by considering that the flow rate varies at each one-hour time-step (i.e., the temporal resolution of irradiance data). By integrating Eq. (3) and (4) into Eq. (2), one can see that, after the discretization of the integral, Eq. (2) is a polynomial equation at each time-step *t*. When solving Eq. (2), the only physically feasible solution of the equation is taken.

The proposed model considers the possible stops of the motor-pump when the water depth in the borehole H_b is found to be deeper than the pumping depth H_p (see Section 2.1). It also considers that the motor-pump starts only when the power supplied by the PV modules is higher than a power $P_{mp,0}$, particularly when the sun has risen sufficiently in the morning and the irradiance has surpassed a certain threshold.

Finally, after calculating the pumping flow rate for every time step, the groundwater volume abstractable by an off-grid PVGWPS V_p of installed power P_p can be calculated for each desired period (see Eq. (1)).

It takes the physics-based model ~9 h to estimate the pumping capacities of a given PVGWPS for the 40,100 locations across Africa and ~ 1.5 h for the 5500 locations across the Indian IGB for a given set of input design parameters (e.g., the PV modules peak power P_p, the borehole radius r_b, see the list of all design parameters in Supplementary Table 1, in Appendix A). Note that the computation times in this study are obtained with the following computer server: Intel(R) Xeon(R) W-2245, 3.90 GHz, 8 cores.

Table 2
Ranges considered for each input location-dependent data and design parameters.

Input Data	Range	References
Annual mean static water level (m) $H_{b,s}$	0 – 300	[56]; [19]
Saturated thickness (m) H_{st}	10 – 500	[56]
Aquifer transmissivity (m^2/s) T	10^{-6} – $5{\cdot}10^{-1}$	[56]; [58]
Specific yield (–) S_y	10^{-5} – 0.5	[88]; [89]; [90]
Pumping depth (m) H_p	$H_{b,s}$ + [1–100]	[41]; [91]
Borehole radius (m) r_b	0.05 – 0.5	[92]
Borehole losses (s^2 m^{-5}) β	10^3 – 10^6	[66]; [93]; [94]; [95]; [96]
Linear pipe losses (s^2 m^{-6}) ν	0 – 10^4	[62]; [97]; [98]
Junction losses (s^2 m^{-5}) K	0 – 10^5	[62]; [97]; [98]
Motor-pump efficiency (–) η_{mp}	0.1 – 0.8	[99]; [100]; [101]
Peak power of the PV modules (W$_p$) P_p	100 – $10{\cdot}10^3$	[41]; [27]
Loss coefficient related to PV modules temperature (%/°C) γ	−0.45 - -0.30	[102]; [103]
Nominal operating cell temperature (°C) NOCT	41 – 46	[104]; [103]
Starting power of the motor-pump (W) P_{mp0}	[10 % - 30 %]·P_p	[101]

2.4. Emulator models

2.4.1. Overview and models selection

The computing times (reported in the previous section) required to evaluate the dynamic PVGWPS model for Africa/the Indian IGB are too high for large-scale decision-making processes, such as optimally positioning and pre-sizing PVGWPSs at extensive scale (e.g., a country, a continent), particularly when several climate scenarios and their uncertainties need to be accounted for. Indeed, a robust optimization taking into consideration uncertainties would require numerous model evaluations. For instance, using a genetic optimization algorithm with a population size of 100 and 100 generations (which are common default hyper parameter values for genetic optimization algorithms [70,71]) would take the algorithm 10,000 evaluations of the model to run. This would then take ~10 years to optimally size the PVGWPSs at every location across Africa using the above physics-based model, and ~ 2 years for the Indian IGB.

To address this challenge, emulator models are proposed in this section. The concept of an emulator involves replacing the time-consuming physics-based model with a machine-learning-based alternative relying on classification or regression tools (e.g., neural networks, random forests) to replicate physics-based model results at a faster rate [72]. Therefore, this study explores emulator models designed to predict the abstractable groundwater volume of a PVGWPS at a specific location, based on input location-dependent data, particularly irradiance time-series, and design parameters of the PVGWPS, such as the peak power of the PV modules and the borehole radius.

A critical aspect of the problem addressed in this study is the consideration of pumping stoppages when the water table at the borehole drops below the pumping depth, caused by the drawdown induced by pumping (see Section 2.1). Indeed, when this situation does not occur, the relationship between the power supplied to the motor-pump by the PV modules and the abstractable groundwater flow rate is found to be a power law relationship, though dependent on location and design parameters (see Supplementary note 1, in Appendix A). This power law relationship between the two time-series has the potential to substantially reduce the complexity of the emulator models.

As a result, the problem has been divided into two steps:

- Step 1: A first machine-learning based model (called 'motor-pump stoppage forecasting model') predicts whether the PVGWPS is likely to stop due to the water level declining to the point of reaching the pumping depth when pumping.
- Step 2:
  - o If the first model does not predict that the PVGWPS is likely to stop, a second machine-learning based model (called 'Prediction model for abstractable groundwater volume') can be used to forecast the abstractable groundwater volume.
  - o If the first model forecasts that the PVGWPS is likely to stop, the physics-based model can be employed to compute the abstractable volume. Indeed, training machine learning models in this case is more challenging as no pattern between the power supply and the abstractable flow rate has been found. In any case, practitioners try to avoid this situation in the field as much as possible, as it may damage the motor pump.

Motor-pump stoppage forecasting model: The task is a classification problem, i.e., the output of the model is a binary number: 1 if the motor-pump is likely to stop during the simulation period due to the drawdown (and thus the water level) reaching the pumping depth when pumping, or 0 if it is not. The drawdown is influenced by the power supplied to the pump from the PV modules and by hydrogeological factors (particularly the transmissivity) [41,56]. As the maximum drawdown is expected to occur for the maximum power supply (reached at the maximum irradiance)[2] [73], it is assumed that if the motor-pump can operate without stopping at the maximum power supply, it will continue functioning under normal conditions. This assumption simplifies the problem by eliminating the need for a full time-series analysis of power supply. Instead, only the maximum power supply over the year is considered. Consequently, the problem is reduced to a classification task based on structured (i.e., time-independent) data enabling the use of relatively simple machine learning models.

Prediction model for abstractable groundwater volume: Given the fact that it is a time-series problem which can be reduced to a power law problem (see Supplementary note 1, in Appendix A), the model will then predict the power law regression coefficients of the power law relationship between power supply and the abstractable groundwater flow rate, using only structured (i.e., time-independent) data. This includes input location-dependent data and design parameters of the PVGWPS. That allows for the use of relatively straightforward machine learning models. Finally, the abstractable groundwater flow rate time-series is obtained by applying the regression coefficients to the power supply time-series. The volume is then obtained by integrating the flow rates.

Given the nature of the problems, the following models are tested in this study:

- **Random Forest (RF):** This method builds multiple independent decision trees in parallel and aggregates their outputs to provide a final prediction. A decision tree is a tree-like structure where each node represents a decision based on a feature, and branches represent the outcome of the decision. A **Random Forest Classifier (RFC)** is used for the motor-pump stoppage forecasting model and a **Random Forest Regressor (RFR)** for the prediction model for the abstractable groundwater volume.
- **Gradient Boosting (GB)**: Unlike RF, this other tree-based method builds trees sequentially rather than in parallel. Each new tree focuses on correcting the mistakes made by the previous ones. In GB, the decision trees are typically shallow, with only a few levels. A **Gradient Boosting Classifier (GBC)** is used for the motor-pump stoppage forecasting model and a **Gradient Boosting Regressor (GBR)** for the prediction model for the abstractable groundwater volume.
- **Multi-layer perceptron (MLP)**: This is a neural network composed of three main components: an input layer, one or more hidden layers, and an output layer. Each layer is composed of neurons, and every neuron from one layer is connected to every neuron of the next one. Between each layer, the data is transferred to the next one through a linear transformation, introducing weights and biases, and at each neuron, the data is transformed once more through an activation function. One MLP is developed for **Classification** for the motor-pump stoppage forecasting model and one for **Regression** for the prediction model for the abstractable groundwater volume.

Fig. 2 provides a block diagram illustrating the overall workflow of the emulator models. The input and output data for each model are shown. It is worth noting that, to reduce the problems' complexity, some input parameters are aggregated into "macro-parameters," such as the quadratic losses term λ (which adds together the borehole losses, the linear pipe losses, and the junction losses), and the well function

[2] This assumption is supported by an analysis conducted on 300 randomly selected locations where the physics-based model returns a stoppage. Indeed, the motor-pump was observed to stop at the maximum solar power for all of these 300 locations.

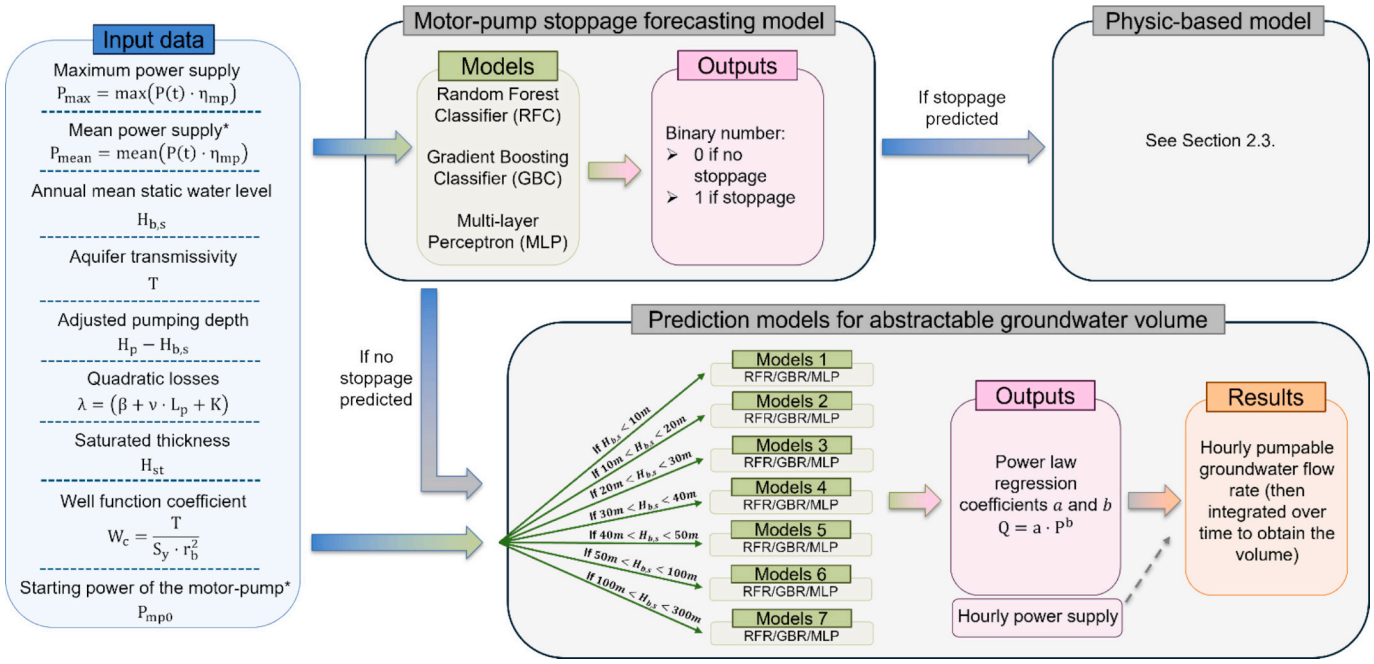


Fig. 2. Block diagram of the emulator models to simulate the pumping capacities of a photovoltaic groundwater pumping system. *: Input data only used for the prediction model for abstractable groundwater volume. The other input data are used for both models.

coefficient W_c. The expressions of the macro-parameters are provided in Fig. 2, in the "Input data" panel. Also, from Eq. (2), it can be observed that, as the static water level $H_{b,s}$ increases, the power law relationship approaches linearity, regardless of other input parameters. This occurs because the non-linear effects associated with drawdown and losses become negligible when the static water level increases. Conversely, when $H_{b,s}$ is lower, the deviation from linearity increases, though this effect may be moderated by a high transmissivity. Thus, to simplify the prediction model for abstractable groundwater volume, the model is divided into seven sub-models based on static water level values: from 0 to 10 m, 10 to 20 m, 20 to 30 m, 30 to 40 m, 40 to 50 m, 50 to 100 m, and 100 to 300 m (see Fig. 2). This approach prevents each sub-model from being overly complex and allows more focus on either the linear or non-linear aspects of the problem. Consequently, stating that an "RFR" (resp. GBR, resp. MLP) is applied implies that each of the seven sub-models corresponds to an RFR (resp. GBR, resp. MLP).

The models have been selected as they are considered state-of-the-art in the field of machine learning and have been used for energy systems and in hydrology. For instance, MLP can be used for solar production forecasting [74,75], and for studying variations in groundwater levels in India [76,77]. Decision tree-based algorithms (e.g., Random Forest and Gradient Boosting) have also been utilized for predicting groundwater levels [78,79], but also for estimating groundwater recharge [80,81], as well as for forecasting streamflow [82,83]. They can also be used for fault detection in PV systems [84], or detection of cleaning intervention on PV modules [85]. Each of these models requires tuning several hyperparameters, including the optimization algorithms, the loss functions, the number of layers and neurons and the activation functions specifically for the MLPs, and the number of decision trees specifically for the RFs and GBs. In this study, the PyTorch library is used to implement the MLPs [86], and the scikit library [87] is used to implement the RFs and the GBs within a Python environment. Hyperparameters have been determined through fine-tuning by hand and are listed in Supplementary note 2 (see Appendix A). In summary, the MLPs comprised 3–4 hidden layers with 50 neurons per layer and ReLU activation functions (Sigmoid or Quantile output), and were trained over 1000–1500 epochs using the Adam optimizer. The RF models were configured with 20–30 trees, and the GB models with 200 trees, the latter using binary cross-entropy or quantile loss functions depending on the target variable.

2.4.2. *Training and validation data*

To incorporate the outputs of the physics-based model into the emulator development process, the physics-based model is first run using random values for the input location-dependent data (except for the irradiance on the plane of the PV modules G_{PV} and the 2 m temperature T_a) and for the design parameters, uniformly selected from literature-based ranges deemed realistic. The outputs of the physics-based model are then used to train and validate the emulators, with the aim of reproducing the model behavior for the same inputs. Note that the selection of each input location-dependent data and design parameters is made to ensure a consistent set of parameters. For instance, even though the pumping depth is randomly selected from a wide range (see Table 2), it is always between the static water level $H_{b,s}$ and the static water level plus the saturated thickness $H_{b,s} + H_{st}$. For the irradiance on the plane of the PV modules G_{PV} and the 2 m temperature, to ensure large and realistic datasets, the values covering the African continent and the ones covering the IGB (see Table 1) were used. Thus, the physics-based model is run using real irradiance and temperature values along with random input location-dependent data and design parameters, producing hourly flow rate data. These datasets provide information for 45,600 locations with hourly data throughout an entire year. The literature-based ranges considered for each input location-dependent data and design parameters are listed in Table 2. Training and validating the emulator models across such broad ranges aims to enable the development of emulators that are not only suitable for our datasets but can also be applied in other contexts. In addition, although we use a single year of data, the spatial coverage for the training and validation includes a large number of pixels spanning diverse meteorological conditions, thus capturing a range of meteorological conditions that one might encounter at a single site over multiple years. This supports the robustness and generalizability of the emulator for multi-year planning.

It is worth noting that, since 7 ranges of static water level $H_{b,s}$ are considered for the prediction model for abstractable groundwater volume (see Fig. 2), the physics-based model is run 8 times for the 45,600

instances. Indeed, the first run covers static water levels from 0 to 300 m to generate training and validation datasets for the motor-pump stoppage forecasting model. It is then run seven more times, each time for a specific static water level range (e.g., 0–10 m, 50–100 m), while keeping all other parameter ranges unchanged. These additional runs generate training and validation datasets for each sub-model of the prediction model for abstractable groundwater volume.

When predicting motor-pump stoppage, the initial dataset presented an imbalance, as the motor-pump stopped in only ~25 % of instances. To address this imbalance and avoid introducing bias into the model training, the dataset when predicting motor-pump stoppage was reduced to make it more balanced, such that motor-pump stoppages occur in ~50 % of the instances. In contrast, when predicting the abstractable groundwater volume when the motor-pump does not stop, the dataset was treated differently. In this case, all instances where the motor-pump stops were completely removed. Finally, the models are trained with 70 % of the instances (chosen randomly) and validated on the remaining 30 %. This hold-out validation ensures that the validation is performed on data not seen during training. It applies to all types of data considered, including the meteorological data.

3. Results

3.1. Emulator models statistics

In this section, results are given for the validation datasets (see Section 2.4.2).

Motor-pump stoppage forecasting model: The confusion matrixes for each fine-tuned classification model (see the value of each hyper-parameter in Supplementary note 2, in Appendix A) are shown in Fig. 3. While the MLP is worse than the other 2, the confusion matrixes of the RFC and the GBC show similar results with ~96 % of instances correctly classified. Nevertheless, it is worth noting that in the context of efficiently sizing a PVGWPS, it is preferrable to predict that the motor-pump stops when it actually does not, than predicting that the motor-pump does not stop while it actually does. In this context, the GBC model appears to be the most appropriate model to detect motor-pump stoppage, as it has the lowest occurrence of the "predicted no-stop/true stop" error (93 occurrences over 6572 simulated cases). Note that it takes each model less than 0.02 s to classify the 6572 simulated cases.

Even though the training process of the GBC may appear as a black-box, a feature from the scikit-learn library allows to calculate the "importance" of each input variable on the predicted results, i.e., how much the GBC model relies on each input variable [105]. Importance is calculated as a numerical value, normalized between 0 and 1. The closer it is to 1, the more important it is. Transmissivity is found to be the most important parameter (importance of 0.65), which is logical as it is one of the main factors influencing the drawdown [56]. The adjusted pumping depth H_p-$H_{b,s}$ also proves to be relatively important (importance of 0.2), which is expected as it represents the physical limit for the drawdown. Finally, while the static water level, as well as the maximum power supply, have non-negligible importance (both importances around 0.07), the other parameters contribute insignificantly. Therefore, these findings help in building confidence in the GBC model's training process.

Prediction model for abstractable groundwater volume: To verify the prediction results, the abstractable groundwater flow rates estimated by the emulators were compared with the ones estimated by the physics-based model using two complementary indicators: the coefficient of determination (R^2) and the normalized root mean square error (NRMSE). R^2 measures how well the emulator reproduces the variations of the physics-based model's outputs, by comparing the prediction errors with the natural variability of the physics-based results. The NRMSE evaluates the root mean square of the differences between the results of the physics-based model and the emulators, normalized by the mean value of the non-zero abstractable groundwater flow rates estimated by the physics-based model. Together, these metrics indicate how accurate the emulator's predictions are (with values of R^2 closer to 1 and lower NRMSE values indicating better performance). The computation time to predict the final abstractable groundwater flow rate time series is also provided as the main goal of the emulator models is to reduce the computation time of the physics-based model. It is worth noting that the results for the three models (RFR, GBR, and MLP), which predict the power law regression coefficients of the relationship between power supply and the abstractable groundwater flow rate time-series, are presented after the final calculation of the abstractable groundwater flow rate time-series. This means that the R^2 and NRMSE metrics are calculated for the abstractable groundwater flow rate time-series, not for the power law regression coefficients. Additionally, the given computation times account for the time needed to apply the power law to the power supply time-series.

The results of each fine-tuned model (see the value of each hyper-parameter in Supplementary note 2, in Appendix A) are shown in Table 3. For each model, the reported values represent the mean values obtained across the seven sub-models (see Fig. 2). However, the conclusions drawn in this section remain applicable to each individual sub-model, as their results are consistent with the mean values.

The RFR stands out as the best model to predict the groundwater volume abstractable by a PVGWPS with the best R^2 and NRMSE values. Although the MLP is the fastest model, the difference in absolute computation time compared to the RFR is relatively small (only a 0.3 s difference compared to a physics-based model which originally takes hours). Therefore, the RFR has been chosen as the best overall model to forecast the groundwater volume abstractable by a PVGWPS.

Once again, a feature from the scikit-learn library enables the calculation of each input variable's "importance" in the predicted

Table 3
Quantitative results of each emulator regression model for predicting the groundwater volume abstractable by a photovoltaic groundwater pumping system.

	R^2	NRMSE	Computation time (s)
RFR	0.997	3.46 %	2.27
GBR	0.983	8.52 %	2.27
MLP	0.992	5.79 %	1.97

RFC	Predicted no-stop	Predicted stop
True no-stop	3049	200
True stop	158	3165

GBC	Predicted no-stop	Predicted stop
True no-stop	3022	196
True stop	93	3261

MLP	Predicted no-stop	Predicted stop
True no-stop	2681	568
True stop	161	3162

Fig. 3. Confusion matrices of the RFC (Random Forest Classifier), the GBC (Gradient Boosting Classifier), and the MLP (Multi-Layer Perceptron).

results, indicating the extent to which the RFR models depend on each variable. The mean and maximum power supply, along with the static water depth, transmissivity, and quadratic losses, logically emerge as the most influential parameters in predicting the groundwater volume abstractable by a PVGWPS. Nevertheless, their relative importances are varying with the static water level range considered (see Fig. 2). The other parameters have negligible impact.

3.2. Detailed results for one location

Based on the results from the previous Section, the combination of the GBC (for predicting motor-pump stops) and of the RFR (for predicting abstractable volume) is considered in the rest of the article. This combination is now applied to the real input location-dependent data and fixed design parameters (detailed in Supplementary table 1, in Appendix A) to predict the groundwater volume abstractable by a PVGWPS. Before showing averaged results at large-scale (see Section 3.3), this section describes results for one randomly selected location. An interesting aspect of working with random forest is the ability to use the central limit theorem to also obtain the confidence intervals of the model [106]. The central limit theorem states that, assuming the individual decision trees within the Random Forest are independent and that the number of trees is sufficiently large (20 trees are used here, see Supplementary note 2, in Appendix A), the average prediction from the RFR model follows a normal distribution. This means that confidence intervals can be computed around the predicted values. In practice, this allows estimates of the uncertainty of the model's output, providing a more robust interpretation of the results.

Therefore, Fig. 4 illustrates the time-series of the groundwater flow rate abstractable by a PVGWPS of 3000 W_p throughout the first day of the year at location: latitude: −29°, longitude: 28.28° (Lesotho). Note that no motor-pump stop was predicted beforehand by the GBC for this location. This location has been selected as both its transmissivity and static water level are low ($3.5{\cdot}10^{-5}$ m^2/s and 7 m respectively), making it suitable for evaluating the emulator's robustness under conditions that deviate from the central ranges. In Fig. 4, the blue line represents the abstractable groundwater flow rate estimated by the physics-based model, the orange-dotted line represents the abstractable groundwater flow rate predicted by the emulator, and the grey zone is the confidence interval of the emulator at 95 %. Although the emulator does not perfectly match the physics-based solution, the latter remains within or close to the confidence interval. An additional analysis is provided in Supplementary note 3 (see Appendix A), where we compare both the large-scale physics-based and emulator models to field data from an installed PVGWPS in Burkina Faso.

3.3. Large-scale results

Fig. 5 illustrates the annual mean daily groundwater volume abstractable by a PVGWPS of 3000 W_p across Africa and the Indian IGB. Fig. 5a and d show the results of the physics-based model across Africa and the Indian IGB respectively. The chosen design parameters to obtain these maps are listed in Supplementary Table 1 (see Appendix A). Figs. 5b and e show the results of the emulator across Africa and the Indian IGB respectively. Finally, Figs. 5c and f show the absolute relative difference between the physics-based model and the emulator across Africa and the Indian IGB respectively. It is important to note that the figures relative to the physics-based model do not show results when the motor-pump is predicted to stop to facilitate the comparisons. Nevertheless, the physics-based model is able to calculate the groundwater abstractable volume even though the motor-pump does stop (as done for instance to compute the emulator model statistics, see Fig. 3 in Section 3.1).

With regard to the motor-pump stoppage forecasting model, the comparison of Fig. 5a and b and the one of Fig. 5d and e underscore the efficacy of the emulator in predicting motor-pump stoppage due to drawdown reaching the pumping depth. Across Africa, the forecasting model misclassifies fewer than 1 % of locations. At locations where the transmissivity is low ($3.5{\cdot}10^{-5}$ m^2/s), the forecasting model misclassifies only 0.68 % of locations. This highlights the strong performance of the GBC, even where the physics-based model deviates from linearity. Across the Indian IGB, the forecasting model does not misclassify any locations.

With regard to the abstractable groundwater volume estimated by the physics-based model, the comparison of Fig. 5a and d highlights the relatively higher pumping capacity in the Indian IGB compared to Africa. This difference can primarily be attributed to the comparatively lower static water levels in the Indian IGB, as well as to the higher transmissivity values observed in this region. Thus, according to the physics-based model, the median daily groundwater volume abstractable by a 3000 W_p across Africa is 28 m^3/day, ranging from 6 m^3/day to 198 m^3/day, whereas the corresponding value for the Indian IGB is 158 m^3/day, with a range of 32 m^3/day to 236 m^3/day.

With regard to the abstractable groundwater volume estimated by the emulator, comparisons with the physics-based model highlight the high performance of the emulator. For the African dataset, the NRMSE between the physics-based model and the emulator is 5.06 %, with an R^2 value of 0.996.[3] The absolute relative difference between the physics-based and emulator models for the groundwater volume abstractable by a 3000 W_p across Africa ranges from 10^{-4} % to 42 %, with 95 % of locations falling between 0.11 % and 8.9 %. At locations with a transmissivity of $3.5{\cdot}10^{-5}$ m^2/s, the median absolute relative difference is 5.1 %, showing that the emulator remains accurate even where the physics-based model is less linear. Finally, the emulator estimates the pumping capacity of the 3000 W_p PVGWPS for the 40,100 locations across Africa in just 17 s (including less than 0.2 s for the GBC which predicts motor-pump stops) achieving a ∼ 1900-fold reduction in computation time compared to the physics-based model.

For the Indian IGB dataset, the NRMSE between the physics-based model and the emulator is 3.69 %, with an R^2 value of 0.995.[4] The absolute relative difference between the physics-based and emulator models for the estimated groundwater volume across the Indian IGB ranges from 0.02 % to 10.7 %, with 95 % of locations falling between 1.1 % and 9.1 %. At locations with a low static water level (1 m), the median absolute relative difference is 4.7 %, further confirming the strong performance of the emulator even where the physics-based model exhibits more pronounced nonlinearity. Finally, the emulator estimates the pumping capacity of the 3000 W_p PVGWPS for the 5500 locations across the Indian IGB in just 3 s (including less than 0.2 s for the GBC), again demonstrating a computation time reduction (by a factor of ∼1600).

The difference between the computation-time reductions obtained for Africa (40,100 locations, reduction of ∼1900) and for the Indian IGB (5500 locations, reduction of ∼1600) highlights that the computation-time reduction factor depends on the number of studied locations. To characterize this dependency more comprehensively, Supplementary fig. 4 in Appendix A provides a runtime scaling curve showing how the computation time reduction varies with the number of studied locations.

Similar results as Fig. 5 are shown for 4 other PVGWPS sizes: 100, 1000, 5000, and 10,000 W_p in Appendix A (supplementary figs. 5 to 8). In addition, Supplementary fig. 9 shows the lower and upper bounds of the confidence interval of the daily water volume abstractable by a 3000 W_p PVGWPS estimated by the emulator. The physics-based solution falls within the 95 % confidence interval for 77 % and 91 % of locations of the

[3] For comparison, assuming a single global regressor model, rather than the seven regressor sub-models used in this study (see Figure 2) results in an NRMSE of 14.1 % and an R^2 of 0.970 for the African dataset.

[4] For comparison, assuming a single global regressor model, rather than the seven regressor sub-models used in this study (see Figure 2) results in an NRMSE of 13.0 % and an R^2 of 0.940 for the Indian IGB dataset.

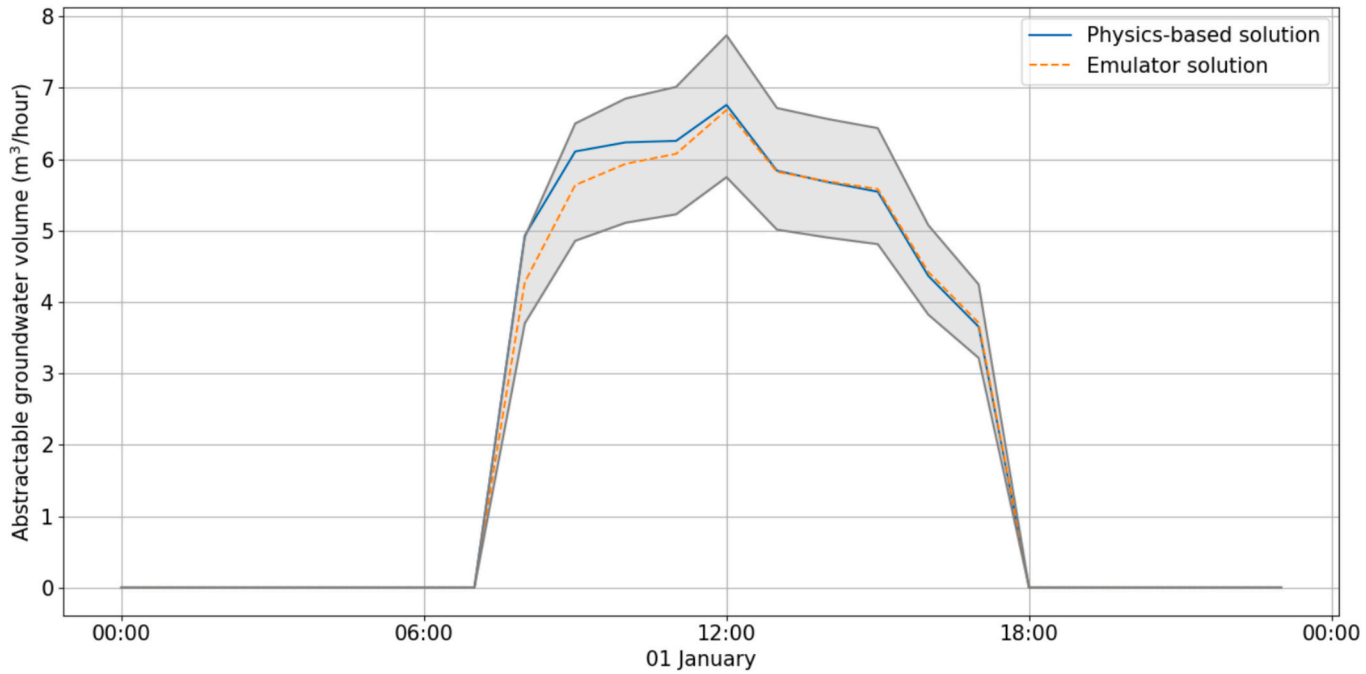


Fig. 4. Time-series of the groundwater flow rate abstractable by a photovoltaic groundwater pumping system of 3000 W_p throughout the first day of the year for the location (latitude: -29°; longitude: 28.28°). The blue line represents the abstractable groundwater flow rate estimated by the physics-based model, the orange-dotted line represents the abstractable groundwater flow rate predicted by the emulator, and the grey zone is the confidence interval of the emulator at 95 %. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

African and Indian IGB datasets respectively. At locations where it lies outside, the median absolute relative difference between the physics-based model and the nearest bound is 1.57 % and 1.67 % for the African and Indian IGB datasets respectively. Thus, even at locations where the physics-based results are not within the confidence interval, the deviation from it remains small. Finally, Supplementary note 4 in Appendix A presents a sensitivity analysis that quantifies how selecting mid-range values for hydrogeological parameters (see Section 2.2) influences the pumpable groundwater volume estimated by the large-scale physics-based model.

4. Discussion

4.1. Influence of groundwater levels

The results of this study underscore the linkage between the groundwater volume abstractable by an off-grid PVGWPS and the static water table. To estimate the abstractable groundwater volume, the static water table is assumed to remain constant during the whole year in this study, due to the lack of data. However, seasonal fluctuations in the water table are expected, notably across the Indian IGB, due to its alluvial geology and the pronounced differences between the dry and monsoon seasons [47,107,108]. Such seasonality can impact the estimations of the abstractable groundwater volume. With static water levels typically lower during the dry season and higher during the wet season [108], the evaluation of the physics-based model in this article tends to overestimate the PVGWPS potential for the dry season and underestimate it for the wet season. It should be noted, however, that if the data are available, there is no technical difficulty in evaluating the model using time series of static water depth as input. The emulator could for instance be applied iteratively for different static water levels. The long-term viability of PVGWPSs could also be explored by analyzing future groundwater conditions and demands in the context of climate change [109]. Thanks to their strongly reduced computation time, the proposed emulator models can facilitate these long-term analyses of PVGWPS over their lifespan (~20 years [66]) and for a large number of scenarios.

4.2. Considerations about sustainability

The primary emphasis in this article has been on estimating the pumping capacity of PVGWPSs. However, even though African and Indian IGB's high photovoltaic and groundwater potentials make PVGWPSs attractive for meeting both current and future water demands in each region, the uncontrolled deployment of these systems can also have negative impacts on groundwater resources that extend beyond their potential benefits [110,111]. While they can reduce greenhouse gas emissions, when they are improperly implemented and sized, they may indeed exacerbate groundwater depletion by impeding optimal water utilization practices [112–114]. This, in turn, could notably worsen an already preoccupying groundwater situation in the western part of the Indian IGB [109]. Thus, further research should also focus on studying the sustainability of such systems with regards to groundwater resources. Future works could for instance quantify the renewable groundwater resources across the studied regions, particularly considering the impact of climate change. This research should take into consideration the role of groundwater resources at sustaining ecosystemic services [115]. It should also account for the linkage between groundwater recharge and groundwater abstraction [116], which will likely evolve as the climate changes. Incorporating these processes into the sizing of PVGWPSs would help ensure the long-term sustainability of the systems throughout their life cycle.

The sustainability of PVGWPSs is not only determined by technical and environmental considerations, but also by the social and institutional contexts within which they are deployed. Governance structures, regulatory mechanisms, incentives and social conditions all influence how technology is adopted and groundwater managed [117]. For instance, a case study from rural Mali highlighted that solar water kiosks could improve both water access and revenue generation without increasing groundwater abstraction [118]. Conversely, subsidy of solar-powered irrigation in Nepal led to increased groundwater use [3]. To identify and ultimately manage and mitigate such risks, the adoption of

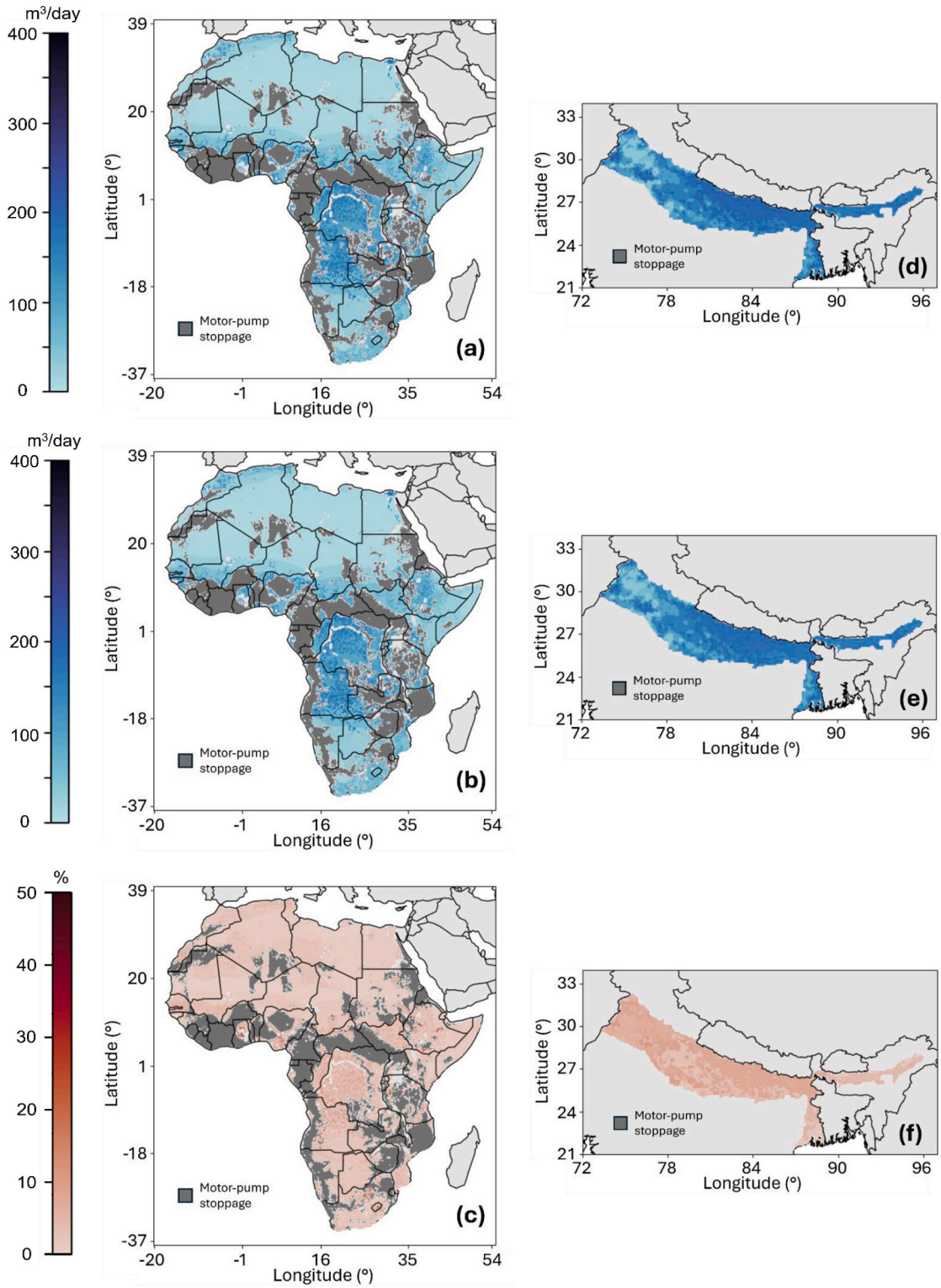


Fig. 5. Estimated daily water volume abstractable by a 3000 W_p photovoltaic groundwater pumping system (PVGWPS) (m^3/day), based on the physics-based model, across Africa (a) and the Indian IGB (d). Estimated daily water volume abstractable by a 3000 W_p PVGWPS (m^3/day), based on the emulator, across Africa (b) and the Indian IGB (e). Absolute relative difference between the physics-based and emulator models for the water volume abstractable by a 3000 W_p PVGWPS (%) across Africa (c), and the Indian IGB (f). Grey pixels correspond to locations where the motor-pump is predicted to stop due to water level reaching the pumping depth.

PVGWPSs must be accompanied by policy safeguards and widespread monitoring. In India, for example, new PVGWPS installations are prohibited in areas where groundwater resources are already overexploited [26]. Moreover, PVGWPS deployment can be coupled with remote monitoring tools to support real-time groundwater management [26].

4.3. Modeling limitations

The principal limitation of this study stems from the inherent challenges of evaluating a model at an extensive spatial scale. Both large-scale groundwater and irradiance input data are estimations, lacking

specificity for local resources. This limitation is particularly relevant for data related to static water depth, transmissivity, and saturated thickness, where significant variations may occur over short distances [119]. Consequently, the outcomes of this large-scale analysis should be considered approximate, providing an overview and serving as a complement to more precise local analyses. Given the potential strong local variations in hydrogeological parameters impacting abstractable volume [119], a coordinated and detailed local investigation and monitoring of groundwater resources is an essential component of the deployment of PVGWPSs [120,121].

Another limitation of the study is the analysis of model uncertainties, especially when dealing with machine-learning tools. Although machine learning can predict complex phenomena with very good results, the models often operate as "black boxes", making it difficult to understand how uncertainties propagate through the model. While this question of model uncertainty has been briefly addressed via the introduction of confidence intervals thanks to the specific architecture of Random Forests, there is a whole strand of machine learning research aimed at making machine learning models interpretable. One example is the rise of physic-informed neural networks, which aim to incorporate physical principles into learning models, and thus could help to better understand how uncertainties propagate into the models [122].

5. Conclusion

A large-scale, physics-based, dynamic model that uses publicly and freely available groundwater and irradiance data to estimate the pumping capacities of PVGWPSs across Africa and the Indian IGB is proposed. The model, which accounts for the different components of the PVGWPS, simulates the evolution of the pumping flow rate with a one-hour time step during a year. Using this model, the abstractable volume by a PVGWPS of 100, 1000, 3000, 5000 and 10,000 W_p is estimated for all locations across the 2 regions, with the results for 3000 W_p particularly described. For a given PVGWPS size, results indicate higher abstractable volumes in the Indian IGB than across Africa, notably due to lower static water depths and higher transmissivity values across the Indian IGB. Therefore, the average annual daily abstractable volume per PVGWPS of 3000 W_p across Africa is estimated to 28 m^3/day, ranging from 6 m^3/day to 198 m^3/day, whereas the corresponding value for the Indian IGB is 158 m^3/day, with a range of 32 m^3/day to 236 m^3/day.

A drawback of this physics-based model is its lengthy computation time: it requires ~9 h to estimate the pumping capacities of PVGWPSs across the 40,100 locations of the African dataset, and ~ 1.5 h for the 5500 locations of the Indian IGB dataset. This notably renders the model unsuitable for solving large-scale optimal dispatch problems accounting for uncertainty, which requires thousands of model evaluations. Such tasks indeed require several thousand iterations of the model. Therefore, machine-learning based emulator models are developed to accelerate the physics-based model. The emulator models are built to ensure their applicability in other contexts. Two models are developed: one model to predict whether the PVGWPS is likely to stop due to the water level declining below the operational threshold when pumping; and one model to predict the groundwater volume abstractable when the PVGWPS is not likely to stop. The models were trained and validated through data augmentation.

Compared to a Random Forest Classifier and a Multi-Layer perceptron, the Gradient Boosting Classifier is identified as the most accurate model for predicting motor-pump stoppages, with only 4 % of errors. It also demonstrates the lowest occurrence of the "predicted no-stop/true stop" error, the most critical error in the study's context. The model mainly focuses on the transmissivity and the pumping depth values, which literature also identifies as important determinants of pump stoppages. When applied to the African dataset, the model misclassifies less than 1 % of locations, and misclassifies no locations when applied to the Indian IGB dataset.

For the prediction of abstractable groundwater volumes, the Random Forest Regressor is found to be the best model when compared to a Gradient Boosting Regressor, and a Multi-Layer perceptron. Applied to the African (resp. Indian IGB) dataset, the combination of the GBC for classification and RFR for regression demonstrates high performances with an R^2 of 0.996 (resp. 0.995), and a NRMSE of 5.06 % (resp. 3.69 %), while reducing the computation time by a factor 1900 (resp. 1600) compared to the physics-based model.

Despite its limitations, this study can help identify regions where PVGWPSs are most promising. In particular, the study provides valuable information about the PVGWPS performance in terms of abstractable groundwater volumes at different locations. Furthermore, it shows how machine-learning based models can accelerate physics-based models for PVGWPS without significantly reducing the quality of estimations. This makes them better suited to address certain complex challenges, such as the optimal positioning and pre-sizing of PVGWPSs at large (e.g., regional, country) scale, while accounting for several possible climate scenarios and their associated uncertainties. Still thanks to the reduced computation time, such machine-learning tools could also be used to study the potential impacts of a large scale deployment of PVGWPSs on the groundwater resources. Combined with sustainability analyses (not examined in this study), these insights could support decisions for large-scale investments in PVGWPS projects, which could be of interest to governments and funding organizations. Additionally, the proposed models can be used as initial screening tools to estimate the pumping performance of PVGWPSs for water sector practitioners, such as local companies or NGOs.

CRediT authorship contribution statement

Guillaume Zuffinetti: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Simon Meunier:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Céline Hudelot:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Donald John MacAllister:** Writing – review & editing, Validation, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Gopal Krishan:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Evelyne Lutton:** Writing – review & editing, Visualization, Validation, Methodology, Conceptualization. **Prosun Bhattacharya:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Conceptualization. **Peter K. Kitanidis:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Conceptualization. **Alan M. MacDonald:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization.

Code availability

The code developed for this study is publicly available at the following link: https://doi.org/10.5281/zenodo.17961240. The code allows to reproduce the results of the article, with both the physcis-based and the machine-learning-based models. For any questions about the code, please contact simon.meunier@centralesupelec.fr.

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.

Acknowledgments

This research was supported by DATAIA convergence institute as part of the"Programme d'Investissement d'Avenir", (ANR- 17-CONV-0003) operated by CentraleSupélec. AMM and DJM publish with the permission of the Director of the British Geological Survey (BGS). This research is partly supported by the BGS International NC program 'Geoscience to tackle Global Environmental Challenges' (NERC reference NE/X006255/1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2025.127268.

Data availability

The data used in this study (along with a read-me file) are publicly available at the following link: https://doi.org/10.5281/zenodo.17961240. The read-me file notably provides the sources and the links to all the raw data.

References

[1] IRENA. Off-grid renewable energy statistics 2023 [online]. 2023, https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2023/Dec/Off-grid_Renewable_Energy_Statistics_2023.pdf [Accessed 26 July 2025].
[2] UNICEF. Progress on drinking water, sanitation and hygiene in Africa (2000–2020) [online]. 2022, https://www.unicef.org/wca/reports/progress-drinking-water-sanitation-and-hygiene-africa-2000-2020 [Accessed 15 June 2025].
[3] World Bank. The hidden wealth of nations: Groundwater in times of climate change [online]. https://www.worldbank.org/en/topic/water/publication/the-hidden-wealth-of-nations-groundwater-in-times-of-climate-change; 2023 [Accessed 19 August 2025].
[4] MacDonald AM, Bonsor HC, Dochartaigh BÉÓ, Taylor RG. Quantitative maps of groundwater resources in Africa. Environ Res Lett 2012;7(2):024009.
[5] Lapworth DJ, Boving TB, Kreamer DK, Kebede S, Smedley PL. Groundwater quality: global threats, opportunities and realising the potential of groundwater. Sci Total Environ 2022;811:152471.
[6] Calow RC, MacDonald AM, Nicol AL, Robins NS. Ground water security and drought in Africa: linking availability, access, and demand. Groundwater 2010;48(2):246–56.
[7] MacAllister DJ, MacDonald AM, Kebede S, Godfrey S, Calow R. Comparative performance of rural water supplies during drought. Nat Commun 2020;11(1):1099.
[8] Taylor RG, Scanlon B, Döll P, Rodell M, Van Beek R, Wada Y, et al. Groundwater and climate change. Nat Clim Chang 2013;3:322–9.
[9] Taylor RG, Koussis AD, Tindimugaya C. Groundwater and climate in Africa—a review. Hydrol Sci J 2009;54(4):655–64.
[10] Chandel SS, Naik MN, Chandel R. Review of solar photovoltaic water pumping system technology for irrigation and community drinking water supplies. Renew Sustain Energy Rev 2015;49:1084–99.
[11] Mapurunga Caracas JV, De Carvalho Farias G, Moreira Teixeira LF, De Souza Ribeiro LA. Implementation of a high-efficiency, high-lifetime, and low-cost converter for an autonomous photovoltaic water pumping system. IEEE Trans Ind Appl 2014;50(1):631–41.
[12] Burney J, Woltering L, Burke M, Naylor R, Pasternak D. Solar-powered drip irrigation enhances food security in the Sudano–Sahel. Proc Natl Acad Sci U S A 2010;107(5):1848–53.
[13] Meunier S, Quéval L, Darga A, Dessante P, Marchand C, Heinrich M, et al. Sensitivity analysis of photovoltaic pumping systems for domestic water supply. IEEE Trans Ind Appl 2020;56(6):6734–43.
[14] World Bank. Development projects: Accelerating solar water pumping via innovative financing [online]. https://projects.worldbank.org/en/projects-operations/project-procurement/P161757; 2022 [Accessed 01 August 2025].
[15] UNICEF. Solar-powered water systems [online]. 2025, https://www.unicef.org/wash/solar-powered-water-systems [Accessed 01 April 2025].
[16] Shah T, Rajan A, Rai GP, Verma S, Durga N. Solar pumps and South Asia's energy-groundwater nexus: exploring implications and reimagining its future. Environ Res Lett 2018;13(11):115003.
[17] Balasubramanian V, Adhya TP, Ladha JK, Hershey C, Neate P. Enhancing eco-efficiency in the intensive cereal-based systems of the indo-Gangetic Plains. In: Eco-efficiency: From vision to reality. Cali, Colombia: CIAT; 2013. p. 99–115.
[18] Jain M, Fishman R, Mondal P, Galford GL, Bhattarai N, Naeem S, et al. Groundwater depletion will reduce cropping intensity in India. Sci Adv 2021;7(9):eabd2849.
[19] MacDonald AM, Bonsor HC, Ahmed KM, Burgess WG, Basharat M, Calow RC, et al. Groundwater quality and depletion in the indo-Gangetic Basin mapped from in situ observations. Nat Geosci 2016;9(10):762–76.
[20] UNESCO. The United Nations world water development report 2022: Groundwater: Making the invisible visible [online]. https://unesdoc.unesco.org/ark:/48223/pf0000380721; 2022 [Accessed 06 August 2025].
[21] MacAllister DJ, Krishan G, Basharat M, Cuba D, MacDonald AM. A century of groundwater accumulation in Pakistan and Northwest India. Nat Geosci 2022;15(5):390–6.
[22] Agrawal S, Jain A. Sustainability of solar-based irrigation in India. New Delhi: CEEW; 2016.
[23] Rathore PKS, Das SS, Chauhan DS. Perspectives of solar photovoltaic water pumping for irrigation in India. Energ Strat Rev 2018;22:385–95.
[24] United Nations. Take urgent action to combat climate change and its impacts [online]. https://sdgs.un.org/goals/goal13; 2025 [Accessed 08 August 2025].
[25] Ministry of Jal Shakti. State-wise report of 6th Census of Minor Irrigation Schemes (Volume-2) [online]. 2023, https://jalshakti-dowr.gov.in/document/state-wise-report-of-6th-census-of-minor-irrigation-schemes-volume-2/ [Accessed 02 September 2025].
[26] PM-KUSUM. National Portal PM-KUSUM [online]. 2025, https://pmkusum.mnre.gov.in/#/landing [Accessed 08 August 2025].
[27] Verma S, Kashyap D, Shah T, Crettaz M, Sikka A. Solar: Solar irrigation for agriculture resilience—A new SDC-IWMI regional partnership. Colombo, Sri Lanka: IWMI; 2018.
[28] Jamil MK, Smolenaars WJ, Ahmad B, Dhaubanjar S, Immerzeel WW, Lutz A, et al. The effect of transitioning from diesel to solar photovoltaic energy for irrigation tube wells on annual groundwater extraction in the lower Indus Basin, Pakistan. J Agric Food Res 2025;20:101799.
[29] Schmitter P, Kibret KS, Lefore N, Barron J. Suitability mapping framework for solar photovoltaic pumps for smallholder farmers in sub-Saharan Africa. Appl Geogr 2018;94:41–57.
[30] Gebrezgabher S, Leh M, Merrey DJ, Kodua TT, Schmitter P. Solar photovoltaic technology for small-scale irrigation in Ghana: Suitability mapping and business models. In: Agricultural water management – Making a business case for smallholders. vol. 178. Colombo, Sri Lanka: International Water Management Institute (IWMI); 2021.
[31] Salim MG. Selection of groundwater sites in Egypt, using geographic information systems, for desalination by solar energy in order to reduce greenhouse gases. J Adv Res 2012;3(1):11–9.
[32] Sayed E, Riad P, Elbeih S, Hagras M, Hassan AA. Multi criteria analysis for groundwater management using solar energy in Moghra oasis. Egypt Egypt J Remote Sens Space Sci 2019;22(3):227–35.
[33] Ammar H, Boukebbous SE, Benbaha N. Photovoltaic water pumping system site suitability analysis using AHP GIS method in southern Algeria. In: In: 2018 4th international conference on renewable energies for developing countries (REDEC), 1–5. IEEE; 2018.
[34] Campana PE, Leduc S, Kim M, Olsson A, Zhang J, Liu J, et al. Suitable and optimal locations for implementing photovoltaic water pumping systems for grassland irrigation in China. Appl Energy 2017;185:1879–89.
[35] Gao X, Liu J, Zhang J, Yan J, Bao S, Xu H, et al. Feasibility evaluation of solar photovoltaic pumping irrigation system based on analysis of dynamic variation of groundwater table. Appl Energy 2013;105:182–93.
[36] Yu Y, Liu J, Wang Y, Xiang C, Zhou J. Practicality of using solar energy for cassava irrigation in the Guangxi autonomous region. China Appl Energy 2018;230:31–41.
[37] Rubio-Aliaga Á, Garcia-Cascales MS, Sánchez-Lozano JM, Molina-Garcia A. Multidimensional analysis of groundwater pumping for irrigation purposes: economic, energy and environmental characterization for PV power plant integration. Renew Energy 2019;138:174–86.
[38] FAO. Solar irrigation potential in the Sahel [online]. 2024, https://openknowledge.fao.org/handle/20.500.14283/cd1326en [Accessed 03 September 2025].
[39] Falchetta G, Semeria F, Tuninetti M, Giordano V, Pachauri S, Byers E. Solar irrigation in sub-Saharan Africa: economic feasibility and development potential. Environ Res Lett 2023;18(9):094044.
[40] Xie H, Ringler C, Mondal MAH. Solar or diesel: a comparison of costs for groundwater-fed irrigation in sub-Saharan Africa under two energy solutions. Earth's. Future 2021;9(4):e2020EF001611.
[41] Meunier S, Kitanidis PK, Cordier A, MacDonald AM. Aquifer conditions, not irradiance determine the potential of photovoltaic energy for groundwater pumping across Africa. Commun Earth Environ 2023;4(1):52.
[42] Meunier S, Queval L, Darga A, Dessante P, Marchand C, Heinrich M, et al. Influence of the temporal resolution of the water consumption profile on photovoltaic water pumping systems modelling and sizing. 2018 7th International Conference on Renewable Energy Research and Applications (ICRERA). 2018. p. 494–9.
[43] Fraidenraich N, Vilela OC. Dynamic behavior of water wells coupled to PV pumping systems. Prog Photovolt Res Appl 2007;15(4):317–30.
[44] Haddad S, Benghanem M, Mellit A, Daffallah KO. ANNs-based modeling and prediction of hourly flow rate of a photovoltaic water pumping system: experimental validation. Renew Sustain Energy Rev 2015;43:635–43.
[45] Wazed SM, Hughes BR, O'Connor D, Calautit JK. A review of sustainable solar irrigation systems for sub-Saharan Africa. Renew Sustain Energy Rev 2018;81:1206–25.
[46] Akhtar S. Spatial-temporal trends mapping and geostatistical modelling of groundwater level depth over northern parts of indo-Gangetic Basin, India. J Geogr Environ Earth Sci Int 2023;27(10):96–112.

[47] Bonsor HC, MacDonald AM, Ahmed KM, Burgess WG, Basharat M, Calow RC, et al. Hydrogeological typologies of the indo-Gangetic basin alluvial aquifer, South Asia. Hydrgeol J 2017;25(5):1377.
[48] Das P, Mukherjee A, Lapworth DJ, Das K, Bhaumik S, Layek MK, et al. Quantifying the dynamics of sub-daily to seasonal hydrological interactions of Ganges river with groundwater in a densely populated city: implications to vulnerability of drinking water sources. J Environ Manage 2021;288:112384.
[49] Verma S, Mishra S, Chowdhury S, Gaur A, Mohapatra S, Soni A, et al. Solar PV powered water pumping system – a review. Mater Today Proc 2021;46:5601–6.
[50] Kazem HA, Al-Waeli AH, Chaichan MT, Al-Mamari AS, Al-Kabi AH. Design, measurement and evaluation of photovoltaic pumping system for rural areas in Oman. Environ Dev Sustain 2017;19:1041–53.
[51] Meunier S, Heinrich M, Quéval L, Cherni JA, Vido L, Darga A, et al. A validated model of a photovoltaic water pumping system for off-grid rural communities. Appl Energy 2019;241:580–91.
[52] IWMI. Solar Irrigation Pump (SIP) Sizing Tool User Manual [online]. https://www.iwmi.cgiar.org/tools/sip-sizing-tool/sip_sizing_manual.pdf; 2022 [Accessed 20 August 2025].
[53] Soenen C, Reinbold V, Meunier S, Cherni JA, Darga A, Dessante P, et al. Comparison of tank and battery storages for photovoltaic water pumping. Energies 2021;14(9):2483.
[54] Sontake VC, Kalamkar VR. Solar photovoltaic water pumping system – a comprehensive review. Renew Sustain Energy Rev 2016;59:1038–67.
[55] Vezin T, Meunier S, Queval L, Cherni JA, Vido L, Darga A, et al. Borehole water level model for photovoltaic water pumping systems. Appl Energy 2020;258:114080.
[56] Bonsor HC, MacDonald AM. An initial estimate of depth to groundwater across Africa. 2011.
[57] Gillies S. Rasterio documentation [online]. https://rasterio-spestana.readthedocs.io/_/downloads/en/latest/pdf/. [Accessed 26 August 2025].
[58] MacDonald AM, Bonsor HC, Taylor R, Shamsudduha M, Burgess WG, Ahmed KM, et al. Groundwater resources in the indo-Gangetic Basin: Resilience to climate change and abstraction. 2015.
[59] India-WRIS (Water Resource Information System). Groundwater level [online]. 2025, https://indiawris.gov.in/wris/#/groundWater [Accessed 03 June 2025].
[60] Copernicus, ERA5. ERA5 hourly data on single levels from 1940 to present [online]. 2025, https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download [Accessed 24 June 2025].
[61] Copernicus, ERA5-Land. ERA5-Land hourly data on single levels from 1950 to present [online]. 2025, https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=download [Accessed 22 June 2025].
[62] Munson BR, Young DF, Okiishi TH, Huebsch WW. Fundamentals of fluid mechanics. USA: John Wiley & Sons, Inc; 2006.
[63] Bear J. Hydraulics of groundwater. Courier Corporation; 2012.
[64] Kasenow M. Applied ground-water hydrology and well hydraulics. Water Resources Publications; 2001.
[65] Zuffinetti G, Meunier S, MacAllister DJ, Kitanidis PK, MacDonald AM. A method for estimating maximum safe installable power for groundwater extraction with application to Africa. Sci Total Environ 2024;955:177062.
[66] Meunier S. Optimal design of photovoltaic water pumping systems for rural communities – A technical, economic and social approach. PhD thesis,. Université Paris-Saclay; 2019.
[67] Jacobson MZ, Jadhav V. World estimates of PV optimal tilt angles and ratios of sunlight incident upon tilted and tracked PV panels relative to horizontal panels. Sol Energy 2018;169:55–66.
[68] Holmgren FW, Hansen WC, Mikofski AM. Pvlib python: a Python package for modeling solar energy systems. J Open Source Softw 2018;3(29):884.
[69] Carrer D, Pique G, Ferlicoq M, Ceamanos X, Ceschia E. What is the potential of cropland albedo management in the fight against global warming? A case study based on the use of cover crops. Environ Res Lett 2018;13(4):044030.
[70] Elsayed SM, Sarker RA, Essam DL. A new genetic algorithm for solving optimization problems. Eng Appl Artif Intel 2014;27:57–69.
[71] Simpson AR, Priest SD. The application of genetic algorithms to optimisation problems in geotechnics. Comput Geotech 1993;15(1):1–19.
[72] Perkins WA, Brenowitz ND, Bretherton CS, Nugent JM. Emulation of cloud microphysics in a climate model. J Adv Model Earth Syst 2024;16(4):e2023MS003851.
[73] Cherni JA, Meunier S, Quéval L. Photovoltaic pumping systems for domestic sustainable water access in off-grid areas. Springer; 2024.
[74] Muhammad Ehsan R, Simon SP, Venkateswaran PR. Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. Neural Comput Appl 2017;28:3981–92.
[75] Yildirim A, Bilgili M, Ozbek A. One-hour-ahead solar radiation forecasting by MLP, LSTM, and ANFIS approaches. Meteorol Atmos Phys 2023;135(1):10.
[76] Derbela M, Nouiri I. Intelligent approach to predict future groundwater level based on artificial neural networks (ANN). Euro-Mediterr J Environ Integr 2020;5:1–11.
[77] Lohani AK, Krishan G. Groundwater level simulation using artificial neural network in Southeast Punjab, India. J Geol Geosci 2015;4(3):206.
[78] Sharafati A, Asadollah SBHS, Neshat A. A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. J Hydrol 2020;591:125468.
[79] Wang X, Liu T, Zheng X, Peng H, Xin J, Zhang B. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. Appl Water Sci 2018;8(5):125.
[80] Pazola A, Shamsudduha M, French J, MacDonald AM, Abiye T, Goni IB, et al. High-resolution long-term average groundwater recharge in Africa estimated using random forest regression and residual interpolation. Hydrol Earth Syst Sci 2024;28(13):2949–67.
[81] Sihag P, Angelaki A, Chaplot B. Estimation of the recharging rate of groundwater using random forest technique. Appl Water Sci 2020;10(7):182.
[82] Akbarian M, Saghafian B, Golian S. Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran. J Hydrol 2023;620:129480.
[83] Song Z, Xia J, Wang G, She D, Hu C, Hong S. Regionalization of hydrological model parameters using gradient boosting machine. Hydrol Earth Syst Sci 2022;26(2):505–24.
[84] Amiri AF, Oudira H, Chouder A, Kichou S. Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier. Energ Conver Manage 2024;301:118076.
[85] Heinrich M, Meunier S, Samé A, Quéval L, Darga A, Oukhellou L, et al. Detection of cleaning interventions on photovoltaic modules with machine learning. Appl Energy 2020;263:114642.
[86] Paszke A, et al. Automatic differentiation in PyTorch. In : 31st Conference on Neural Information Processing Systems, NIPS 2017:1–4.
[87] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12(Oct):2825–2830.
[88] de Graaf IE, van Beek RL, Gleeson T, Moosdorf N, Schmitz O, Sutanudjaja EH, et al. A global-scale two-layer transient groundwater model: development and application to groundwater depletion. Adv Water Resour 2017;102:53–67.
[89] Freeze RA, Cherry JA. Groundwater. New Jersey: Prentice-Hall; 1979.
[90] Hajhouji Y. Modélisation hydrologique du bassin versant de l'oued Rheraya et sa contribution à la recharge de la nappe du Haouz (bassin du Tensift, Maroc). PhD thesis, Université Paul Sabatier–Toulouse III; 2018.
[91] Ministry of Jal Shakti. Minor Irrigation Census [online]. 2014, https://mowr.nic.in/irrigationcensus/ [Accessed 12 June 2025].
[92] MacDonald AM, Barker JA, Davies J. The bailer test: a simple effective pumping test for assessing borehole success. Hydrgeol J 2008;16:1065–75.
[93] Sourcewater. Appendix G, Calculation of In-Well Losses [online]. n.d, https://www.sourcewater.ca/media/gfcdmzem/rmow-tier-3-wqra-appendix-g-h_september-2014.pdf [Accessed 05 Sept 2025].
[94] Institut de Recherche pour le Développement. Pumping test [online]. 2005, https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers16-07/010048822.pdf [Accessed 05 Sept 2025].
[95] Roman JA. Aquifer test methods to estimate transmissivity and well loss via a single pumping well. Master's thesis,. Colorado State University; 2019.
[96] JICA. The study on rural water supply in Mwanza and Mara regions in the United Republic of Tanzania [online]. n.d, https://openjicareport.jica.go.jp/pdf/11836988_09.pdf [Accessed 02 Sept 2025].
[97] Rural Water Supply Network. Handpump technologies [online]. 2025, https://www.rural-water-supply.net/en/sustainable-groundwater-management/handpump-technologies [Accessed 05 Sept 2025].
[98] altEstore. How to size pipe for solar water pumping [online]. 2025, https://www.altestore.com/pages/how-to-size-pipe-for-solar-water-pumping [Accessed 21 August 2025].
[99] Ahmed AA, Moharam BA, Rashad EE. Improving energy efficiency and economics of motor-pump-system using electric variable-speed drives for automatic transition of working points. Comput Electr Eng 2022;97:107607.
[100] Ducar I, Marinescu C. Increasing the efficiency of motor-pump systems using a vector controlled drive for PMSM application. International Symposium on Fundamentals of Electrical Engineering (ISFEE) 2014;2014:1–5.
[101] Grundfos. Submersible groundwater pumps [online]. 2025, https://product-selection.grundfos.com/us/categories/pumps/submersible-groundwater-pumps?tab=categories [Accessed 28 June 2025].
[102] Ponce-Alcántara S, Connolly JP, Sánchez G, Miguez JM, Hoffmann V, Ordás R. A statistical analysis of the temperature coefficients of industrial silicon solar cells. Energy Procedia 2014;55:578–88.
[103] Jieb YA, Hossain E. Photovoltaic systems: Fundamentals and applications. Cham: Springer; 2022.
[104] Koehl M, Heck M, Wiesmeier S, Wirth J. Modeling of the nominal operating cell temperature based on outdoor weathering. Sol Energy Mater Sol Cells 2011;95(7):1638–46.
[105] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[106] Mentch L, Hooker G. Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. J Mach Learn Res 2016;17(26):1–41.
[107] Shah T. Climate change and groundwater: India's opportunities for mitigation and adaptation. In: Water resources policies in South Asia. Routledge India; 2020. p. 213–43.
[108] Shamsudduha M, Chandler RE, Taylor RG, Ahmed KM. Recent trends in groundwater levels in a highly seasonal hydrological system: the Ganges-Brahmaputra-Meghna Delta. Hydrol Earth Syst Sci 2009;13(12):2373–85.
[109] Bhattarai N, Lobell DB, Balwinder-Singh Fishman R, Kustas WP, Pokhrel Y, Jain M. Warming temperatures exacerbate groundwater depletion rates in India. Sci Adv 2023;9(35):eadi1401.
[110] Closas A, Rap E. Solar-based groundwater pumping for irrigation: sustainability, policies, and limitations. Energy Policy 2017;104:33–7.
[111] Zuffinetti G, Meunier S. Mapping the risk posed to groundwater-dependent ecosystems by uncontrolled access to photovoltaic water pumping in Sub-Saharan Africa. Policy Research Working Paper (No. 10935). World Bank. 2024.
[112] Gupta E. Extending solar water pump subsidies: impact on water use, energy use and cropping patterns in Rajasthan: difference in differences analysis. South Asian

Network for Development and Environmental Economics Working Paper 2017: 1–99.
[113] Kishore A, Shah T. Tewari NP. Solar Irrigation Pumps: Farmers' experience and state policy in Rajasthan. Econ Polit Wkly; 2014. p. 55–62.
[114] Shah T, Verma S, Durga N. Karnataka's smart, new solar pump policy for irrigation. Econ Pol Wkly 2014;49(48):10–4.
[115] Gleeson T, Wada Y, Bierkens MF, van Beek LP. Water balance of global aquifers revealed by groundwater footprint. Nature 2012;488(7410):197–200.
[116] Cuthbert MO, Gleeson T, Bierkens MFP, Ferguson G, Taylor RG. Defining renewable groundwater use and its relevance to sustainable groundwater management. Water Resour Res 2023;59(9):e2022WR032831.
[117] Milman A, MacDonald A. Focus on interactions between science-policy in groundwater systems. Environ Res Lett 2020;15(9):090201.
[118] Wagner J, Merner S, Innocenti S, Geling A, Hope R. Can solar water kiosks generate sustainable revenue streams for rural water services? World Dev 2025; 185:106787.
[119] Owor M, Okullo J, Fallas H, MacDonald AM, Taylor R, MacAllister DJ. Permeability of the weathered bedrock aquifers in Uganda: evidence from a large pumping-test dataset and its implications for rural water supply. Hydrgeol J 2022; 30(7):2223–35.
[120] Sahuquet A, Meunier S, Cherni JA, Charpentier A, Vezin T, Darga A, et al. Photovoltaic pumping tests: a novel supervision method for photovoltaic water pumping systems. Heliyon 2024;10(21):e39718.
[121] Taylor CJ, Alley WM. Ground-water-level monitoring and the importance of long-term water-level datavol. 1217. Denver, CO, USA: US Geological Survey; 2001.
[122] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nat Rev Phys 2021;3(6):422–40.