hydrology

MDPI

Review

Modeling Groundwater Resources in Data-Scarce Regions for Sustainable Management: Methodologies and Limits

Iolanda Borzì

Department of Mathematics, Computer Sciences, Physics and Earth Sciences, The University of Messina, 98166 Messina, Italy; iolanda.borzi@unime.it

Abstract: Groundwater modeling in data-scarce regions faces significant challenges due to the lack of comprehensive, high-quality data, impacting model accuracy. This systematic review of Scopus-indexed papers identifies various approaches to address these challenges, including coupled hydrological-groundwater models, machine learning techniques, distributed hydrological models, water balance models, 3D groundwater flow modeling, geostatistical techniques, remote sensing-based approaches, isotope-based methods, global model downscaling, and integrated modeling approaches. Each methodology offers unique advantages for groundwater assessment and management in data-poor environments, often combining multiple data sources and modeling techniques to overcome limitations. However, these approaches face common challenges related to data quality, scale transferability, and the representation of complex hydrogeological processes. This review emphasizes the importance of adapting methodologies to specific regional contexts and data availability. It underscores the value of combining multiple data sources and modeling techniques to provide robust estimates for sustainable groundwater management. The choice of method ultimately depends on the specific objectives, scale of the study, and available data in the region of interest. Future research should focus on improving the integration of diverse data sources, enhancing the representation of complex hydrogeological processes in simplified models, and developing robust uncertainty quantification methods tailored for data-scarce conditions.

Keywords: data-scarce aquifer; ungauged aquifer; groundwater modeling; sustainable groundwater resource management; uncertainty analysis; groundwater data lack

check for updates

Academic Editor: Peiyue Li

Received: 3 December 2024
Revised: 3 January 2025
Accepted: 8 January 2025
Published: 9 January 2025

Citation: Borzì, I. Modeling Groundwater Resources in Data-Scarce Regions for Sustainable Management: Methodologies and Limits. *Hydrology* **2025**, *12*, 11. https://doi.org/10.3390/hydrology12010011

1. Introduction

Groundwater is a critical resource that plays an essential role in supporting human life and the environment. It serves as a primary source of drinking water for approximately 50% of the global population and accounts for about 40% of the water used for irrigation, making it indispensable for food production and agricultural sustainability [1]. In regions where surface water is scarce, groundwater often represents the only reliable source of water, particularly in arid and semi-arid areas [2]. As the Earth's population continues to grow, projected to reach nearly 11 billion by 2100, the demand for food and water will increase significantly, highlighting the need for the sustainable management of groundwater resources [1]. The proper management of this limited resource is crucial to support the growing demand and prevent depletion, which can have severe environmental and socioeconomic consequences [2–5]. For the above-mentioned reasons, groundwater modeling is crucial for sustainable water resource management due to its ability to provide insights into the behavior of groundwater systems under various conditions [6–9].

The quantitative modeling of groundwater systems represents an indispensable tool for achieving sustainable water resource management [10]. In the recent literature, Secci et al. [10] compared different modeling approaches (process-based models, data-driven models, and system-dynamics models) for groundwater sustainable management, highlighting their characteristics and advantages. They found that process-based models provide a theoretical framework for understanding groundwater dynamics but require extensive parameterization, which can be challenging in heterogeneous aquifer systems. Conversely, data-driven models leverage available data to predict groundwater behavior but depend heavily on the quality and quantity of the input data, making them less effective when data are limited. Then, system dynamics modeling offers a holistic view by incorporating socio-economic factors into groundwater management strategies, facilitating stakeholder engagement and policy development. For example, widely used numerical models such as MODFLOW [11] and FEFLOW [12] offer reliable simulations results but needs robust datasets for proper calibration processes [13].

All these models provide critical insights into aquifer dynamics, enabling the prediction of groundwater behavior under various water stress scenarios, including increased withdrawals, changing precipitation patterns, and land-use modifications. This predictive capability is vital for developing management strategies that balance human needs with ecological sustainability [10]. Moreover, groundwater models assist in optimizing groundwater protection efforts. They can identify vulnerable areas and assess the potential risks associated with over-extraction or contamination, enabling planners to implement effective conservation measures. Furthermore, groundwater modeling plays a pivotal role in fostering integrated water resource management (IWRM) by explicitly incorporating socio-economic factors into the decision-making process [9,14]. By coupling hydrological simulations with economic considerations and social impact assessments, these models facilitate scenario analysis that considers the needs of diverse stakeholders. The latter can include agricultural communities, urban water utilities, environmental protection agencies, and indigenous populations. This holistic approach empowers the development of adaptive management strategies that can evolve and respond to dynamic conditions, such as population growth, climate change impacts, and evolving water demands [9,10,15].

Hydrogeological studies and aquifer modeling are often hindered by the absence of long-term water table records and the difficulties in accessing relevant locations, making it challenging to manage local resources effectively [4]. Therefore, a proper modeling of groundwater resources can be challenging in data-scarce regions due to the lack of sufficient hydrogeological data and piezometric level time-series. As a result, alternative approaches and techniques must be explored to assess and manage groundwater resources in such situations.

This review paper aims to explore various modeling techniques that have been employed to assess and manage groundwater resources in data-scarce regions. The focus will be on the applicability, limitations, and potential for further advancements of these techniques. The review will provide a comprehensive overview of the approaches used to tackle the challenges of groundwater management in areas with limited data availability, with the goal of informing future research and practical applications in this field.

The primary objective of this study is to address three key research questions regarding the following:

- Data Scarcity and Groundwater Modeling: how do scientists in the recent literature address the challenges posed by data scarcity or lack when developing groundwater models?
- Methodological Evaluation: what are the strengths and weaknesses of various methodologies employed for groundwater modeling in data-scarce environments?

- Future Research Priorities: what are the promising avenues for future research in the field of ungauged aquifer modeling to ensure sustainable groundwater management?

This review article is organized as follows: Section 2 details the systematic review protocol employed to select relevant research papers for inclusion in this literature review. The protocol's application to the Scopus database is explained, ensuring a comprehensive and unbiased selection of studies. Section 3 provides an in-depth analysis of the diverse methodological approaches identified during the literature review process. These approaches are categorized and described, highlighting their key features and applications in data-scarce regions. Section 4 offers a critical discussion of the methodological characteristics, advantages, limitations, and uncertainties associated with each approach. This section synthesizes findings across studies, comparing different methods to provide a comprehensive overview of their strengths and weaknesses. Section 5 identifies research priorities and future directions in the field of groundwater modeling in data-scarce regions. This section draws on the gaps and challenges identified in the reviewed literature to suggest areas for further investigation and methodological development. Section 6 concludes the review by summarizing key findings, emphasizing the importance of adapting methodologies to specific regional contexts, and highlighting the potential of integrated approaches for sustainable groundwater management in data-limited environments.

2. Materials and Methods

In this study, a systematic literature review has been conducted to identify, select, screen, and filter scientific articles from the Scopus database.

2.1. Literature Review Protocol

The search terms and keywords used are "Groundwater", "Aquifer", "Modeling", "Modelling", "Data-Scarce", "Data Scarce", "Ungauged", "Sustainable", "Management". The search has been conducted both in the "KEY" field and in the "TITLE-ABS-KEY" field. The KEY fields means that the search for a specific term will be conducted in the keywords of the document. Keywords included terms selected by both authors and publishers. The TITLE-ABS-KEY field means that a search for the terms will be conducted in the title, in the abstract, and in the keywords of each document. This ensures that the search is broad enough to consider enough information and documents. The above-mentioned words have then been combined with "AND" and/or "OR" logical operators in order to take into account synonymous, alternative, or equivalent terms and to perform a proper search of the database. In particular, the following text combinations were used for searching in Scopus database: "groundwater OR aquifer AND modelling OR modeling AND data-scarce OR ungauged AND sustainable AND management"; "groundwater OR aquifer AND modelling OR modeling AND data-scarce OR ungauged AND aquifer AND management"; "groundwater AND modelling AND data-scarce AND aquifer AND management"; "groundwater AND modelling AND ungauged AND aquifer AND management"; "groundwater AND modelling AND ungauged AND aquifer AND sustainable AND management". For all the searches, the time range was set "between 2010 and 2024" to consider only recent and updated references.

First, all duplicates from different searches were removed. Then, the search results were processed following the PRISMA protocol [16] represented as flowchart in Figure 1. A number of 35 records were identified through Scopus database searching. Cross-referencing and manual search on the topic produced 28 extra records. From the combination of both search sources, a total of 63 records were collected. Of those 63 records, 5 were excluded, as they were not peer-reviewed scientific articles, and the remaining 58 were then screened. Of the 58 screened articles, 13 of them were removed from the database due to lack of relevance, after reading the abstract and contents, and 7 were removed because full paper was not found or accessible. At the end of the screening protocol, a total of 38 scientific papers were collected and studied in depth for the purpose of this review.

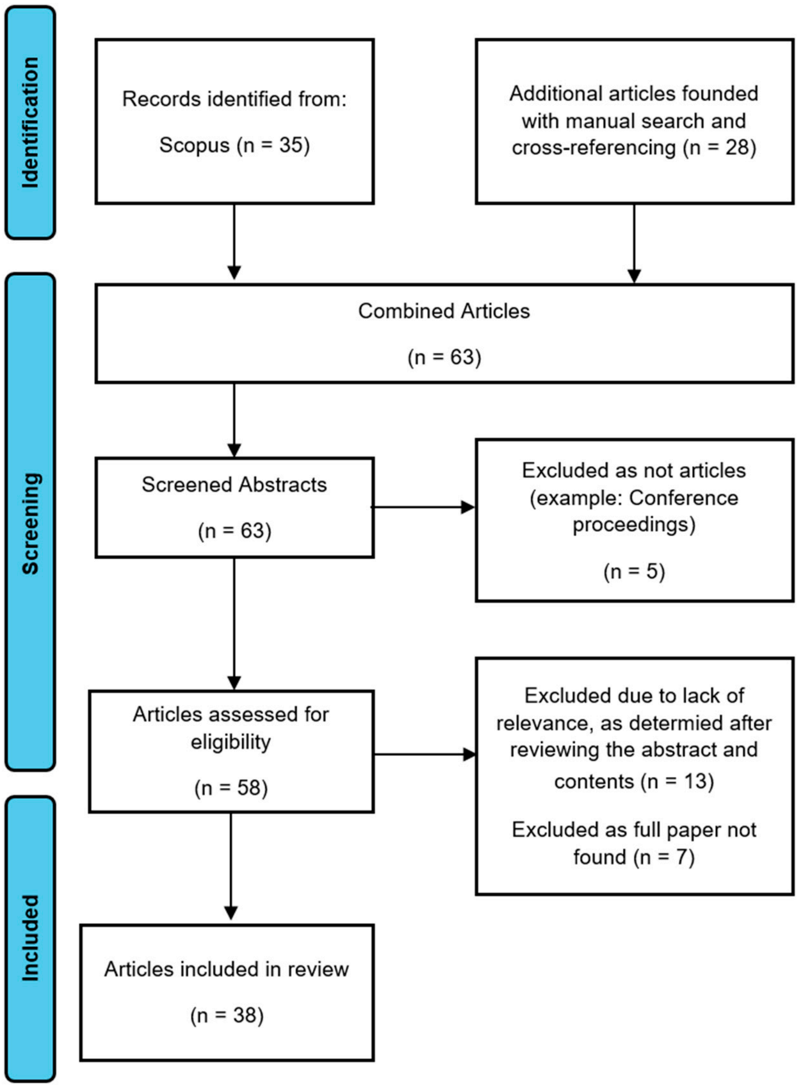

Figure 1. PRISMA [16] flowchart highlighting various stages of the literature review process.

2.2. Database Overview

As an initial overview of the article database, Figure 2 illustrates the temporal distribution of publication years for the 38 scientific papers included in this review.

Figure 2 reveals a notable increase in publications from 2011 onwards, indicating a growing interest and importance of ungauged aquifer modeling within the scientific community. This trend highlights the increasing recognition of the challenges posed by data scarcity in groundwater management and the need for innovative modeling approaches.

Figure 3 provides a visual representation of the journals where the 38 scientific articles included in this review were originally published. By analyzing the distribution of articles across different journals, we can gain insights into the specific research communities and outlets that have been actively contributing to the field of ungauged aquifer modeling. This information can help identify potential collaborations, track emerging trends, and assess the overall impact of research in this area. Given the broad interest in groundwater modeling and its implications for sustainable water resource management, understanding the diverse range of journals involved is essential for a comprehensive overview of the field.

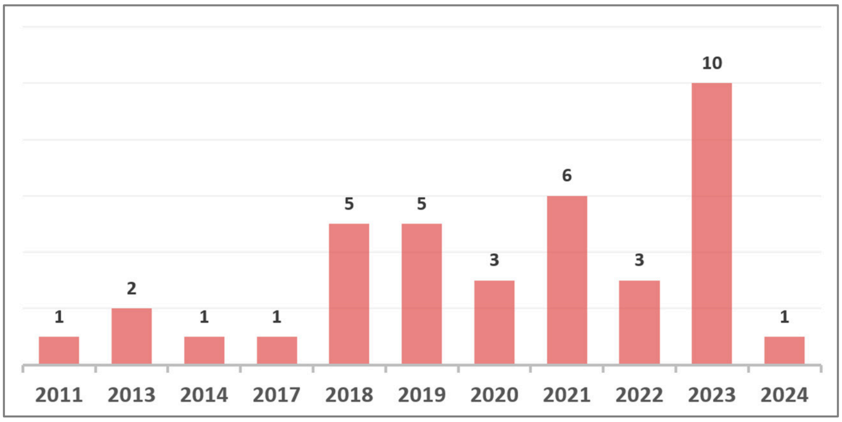

Figure 2. Temporal distribution of publication years for the 38 scientific papers included in the systematic literature review.

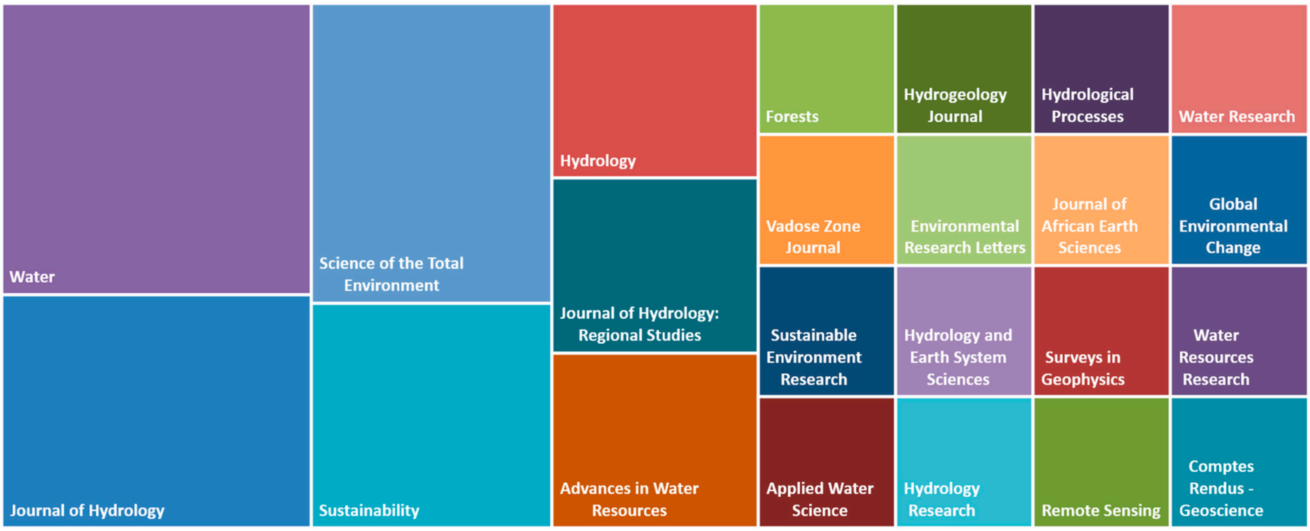

Figure 3. Distribution of selected scientific articles in the systematic literature review across peer-reviewed journals.

2.3. Case Study Countries

Figure 4 provides a global overview of the location of case studies by countries investigated in the 38 scientific papers included in this review. By mapping the geographic distribution of these case studies, we can identify regions where ungauged aquifer modeling is particularly relevant and gain insights into the diverse hydrological settings and challenges faced by researchers worldwide. A significant number of case studies are located in developing countries, particularly in Africa and Asia. This underscores the importance of groundwater resources in these regions, where water scarcity is a pressing issue, and the potential benefits of improved groundwater management through modeling. The diverse array of case study countries represented in Figure 4 underscores the global significance and broad applicability of ungauged aquifer modeling. This geographical spread, encompassing arid, semi-arid, and humid regions across various continents, highlights the universal challenges and opportunities associated with groundwater management in data-scarce environments.

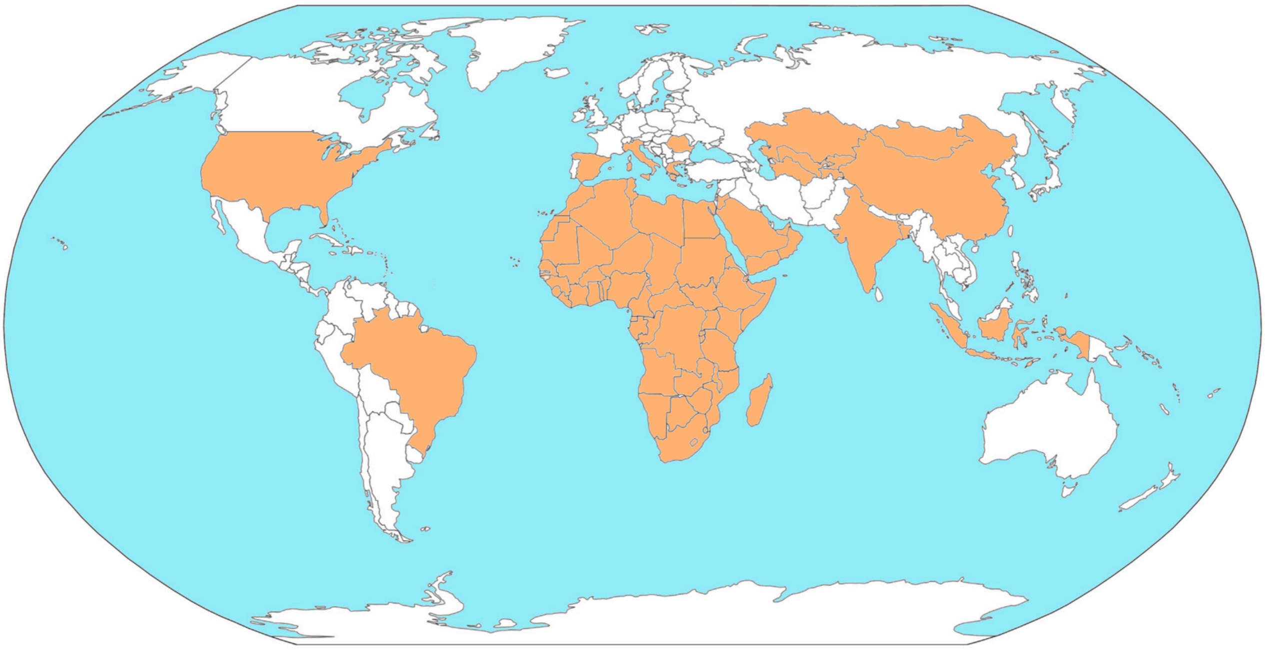

Figure 4. Global Distribution of Case Study Countries for Ungauged Aquifer Modeling within the selected papers in the literature review.

3. Methodological Approaches

The selected database includes a total of 38 scientific papers presenting a diverse array of methodologies for estimating and managing groundwater resources in data-scarce regions. All papers have been investigated and classified by methodological approaches. Table A1 in Appendix A summarizes methodological approaches highlighting advantages, limitations, input data requirements, and referencing database papers.

3.1. Coupled Hydrological-Groundwater Models

Coupled hydrological-groundwater models [17–22] offer a comprehensive approach to simulating the complex interactions between surface water and groundwater systems in data-scarce regions. These models typically integrate surface hydrological processes with groundwater dynamics, allowing for a more holistic representation of the water cycle. For instance, Ref. [17] employs a one-way coupling of the CREST hydrological model with MODFLOW-NWT, enabling fine-resolution (500 m) modeling at regional scales. Another approach [18] sees the introduction of a modified IHACRES rainfall-runoff model adapted for river basins connected to deep groundwater aquifers, using a two-store routing module to simulate groundwater and spring discharge. Ebrahim et al. [19] presents an integrated approach using MODFLOW-OWHM, which combines remote sensing, rainfall-runoff modeling, and 3D dynamic modeling to assess groundwater resources in hard-rock semi-arid terrain. These coupled models offer several advantages, including the ability to represent surface–groundwater interactions, incorporate remote sensing data, and simulate complex hydrological processes. However, they also face challenges such as high computational demands, potential error propagation between model components, and the need for extensive parameterization. Despite these limitations, coupled hydrological-groundwater models prove valuable for understanding water resource dynamics and informing sustainable management practices in data-scarce environments.

3.2. Machine Learning Techniques

Machine learning techniques [23–25] offer promising approaches for groundwater modeling in data-scarce regions. Gaffoor et al. [23] compares ensemble and deep learning algorithms, specifically Gradient Boosting Decision Trees (GBDT) and Long Short-Term Memory Neural Networks (LSTM-NN), for modeling groundwater levels. These techniques show strength in working with limited historical data, with LSTM capturing temporal dependencies and GBDT performing well with small sample sizes. Rafik et al. [24] presents a novel combination of SWAT hydrological modeling, downscaled GRACE satellite data, and machine learning techniques (Random Forest and Support Vector Regression) for groundwater level forecasting. This integrated approach leverages multiple data sources to overcome data scarcity, providing both spatial and temporal groundwater level estimates. Fletcher et al. [25] introduces an adaptive management framework that integrates Bayesian learning with stochastic dynamic programming to assess opportunities for learning about groundwater availability over time. This approach allows for flexible decision-making under uncertainty, particularly valuable in data-scarce regions where initial uncertainty is high but can be reduced through ongoing monitoring. While these machine learning techniques offer significant advantages in predicting groundwater levels and supporting adaptive management, they also face challenges related to data quality, model complexity, and the need for some ground-truthing or validation data.

3.3. Distributed Hydrological Models

Distributed hydrological models, particularly the Soil and Water Assessment Tool (SWAT), are widely used for simulating hydrological processes in data-scarce regions [26–28]. SWAT is a semi-distributed, physically based model that operates on a daily time step and divides watersheds into sub-basins and hydrologic response units (HRUs). It offers a comprehensive representation of hydrological processes, making it suitable for assessing the impacts of land use changes, climate variations, and management practices on water resources. In the Kilombero floodplain of Tanzania [27], SWAT was applied using a combination of local precipitation data and satellite-based rainfall estimates to overcome data scarcity, achieving fair model performance with Nash–Sutcliffe Efficiency (NSE) values of 0.43 and 0.23 for calibration and validation periods, respectively. Similarly, in the upper Mara River Basin in Kenya [28], SWAT was employed to evaluate the impacts of land use and climate change on hydrology, utilizing coarse resolution datasets and satellite-derived rainfall estimates. While these models offer detailed process representation and the ability to simulate complex interactions between land use, climate, and hydrology, they face challenges in data-scarce environments. These include extensive parameterization requirements, which can be difficult to fulfill with limited data, and the need for careful calibration and validation. Despite these limitations, distributed models like SWAT prove valuable for understanding water resource dynamics and informing sustainable management practices in data-poor regions.

3.4. Water Balance Models

Water balance models [29–32] offer a valuable approach for estimating groundwater recharge and resources in data-scarce regions. These models typically utilize readily available spatial data such as land use/land cover maps, soil maps, and digital elevation models to provide spatially explicit recharge estimates. For instance, Ref. [32] employs the WetSpass spatially distributed water balance model integrated with GIS and remote sensing for groundwater recharge estimation in the Upper Gelana Watershed, Ethiopia. This method is particularly suitable for data-scarce regions as it incorporates land use and soil data. An inverse hydrogeological balance method has been applied in [30,31] at annual and monthly scales, respectively, highlighting the role of evapotranspiration assessment in groundwater resource estimation. While these water balance models offer the advantage of large-scale applicability with limited data, they may oversimplify complex hydrological processes and often rely on empirical relationships. The temporal resolution is typically limited, with many models providing only yearly averages. Despite these limitations, water balance models remain a valuable tool for initial groundwater resource assessment in data-poor environments, offering a balance between data requirements and spatial coverage.

3.5. 3D Groundwater Flow Modeling

Three-dimensional groundwater flow modeling, particularly using MODFLOW, is a powerful approach for simulating complex groundwater systems in data-scarce regions. Refs. [20,33]'s method allows for the detailed representation of aquifer geometry and properties, enabling the quantification of inter-basin groundwater flow and complex boundary conditions. In [33], MODFLOW-2005 was used to develop a steady-state groundwater flow model for the Gallocanta Lake watershed in Spain, successfully quantifying inter-basin groundwater flow exchanges with adjoining basins. The model achieved good calibration, with a Nash–Sutcliffe Efficiency Index of 0.8, despite data scarcity. Rödiger et al. [20] employed MODFLOW as part of the MODFLOW-OWHM (One-Water Hydrologic Flow Model) to assess the impact of groundwater development and climate change on Jordan's multi-aquifer system. This integrated approach combines surface and groundwater processes, accounting for climate change scenarios on a national scale. However, these models face challenges in data-scarce environments, as they require significant geological and hydrogeological data for parameterization. The models may also simplify complex aquifer interactions due to data limitations. Despite these challenges, 3D groundwater flow modeling with MODFLOW proves to be a valuable tool for understanding groundwater dynamics and informing sustainable management practices in data-limited regions.

3.6. Geostatistical and Geophysical Techniques

Geostatistical techniques, particularly ordinary kriging [34], offer valuable approaches for analyzing groundwater table variability and trends in data-scarce regions. The study in Sylhet, Bangladesh, applied ordinary kriging to interpolate groundwater levels from 46 observation wells, providing insights into spatial and temporal variations over a 15-year period. This method allows for the spatial interpolation of groundwater levels, the quantification of spatial autocorrelation, and an estimation of uncertainty. The key advantage of kriging is its ability to provide predictions for ungauged areas, which is crucial for identifying vulnerable zones and informing groundwater management decisions. Varouchakis et al. [35] further enhanced this approach by incorporating physical laws and local geographic features into residual kriging, improving prediction accuracy in data-scarce conditions. However, these techniques face limitations, such as the assumption of spatial stationarity and a sensitivity to data quality and distribution. The accuracy of kriging predictions heavily depends on the spatial distribution and quality of available data points. Additionally, these methods may not fully capture complex hydrogeological processes. Notably, Mohamed et al. [36] employ hydro-geophysical monitoring using gravity data to assess the North-Western Sahara Aquifer System's groundwater resources, showcasing the potential of geophysical methods in areas with limited traditional data. Despite challenges, geostatistical techniques prove to be valuable tools for understanding groundwater dynamics and supporting sustainable management practices in data-limited environments.

3.7. Remote Sensing-Based Approaches

Remote sensing-based approaches [32,37–44] offer valuable solutions for estimating groundwater resources in data-scarce regions. Siavashani [37] explored the use of satellite-based climate data from the CHADFDM (Climate Hazards Group InfraRed Precipitation with Stations Data Fusion for Drought Monitoring) platform for groundwater recharge estimation in arid and semi-arid areas. This approach provides continuous spatial and temporal coverage of precipitation and temperature data, which is crucial for recharge estimation in regions with limited ground-based observations. Demisse et al. [32] utilized the WetSpass model integrated with GIS and remote sensing tools to estimate groundwater recharge in the Upper Gelana watershed, Ethiopia. This method leverages readily available spatial data such as land use/land cover maps, soil maps, and digital elevation models to provide spatially explicit recharge estimates. Jòdar et al. [38] combined lumped hydrological models with remote sensing data, particularly MODIS snow cover and albedo products, to evaluate water resources in semi-arid, high-altitude ungauged watersheds.

Sun et al. [39] demonstrates the use of remote sensing and GIS techniques to map prospective water resources and monitor land use/land cover changes in arid regions, providing valuable insights into spatially explicit recharge estimates.

Central Asia was investigated by [40,41], respectively. [40] investigated the impact of land-cover change on groundwater levels in the Tarim Basin, using remote sensing and hydrological modeling to understand the dynamics of groundwater depletion in desert regions, and [41] mapped groundwater-dependent ecosystems in arid Central Asia, using remote sensing and GIS to identify areas vulnerable to land degradation due to groundwater extraction.

Springer et al. [42] emphasized the role of space-based observations for groundwater resource monitoring over Africa, highlighting the use of satellite data to fill gaps in ground-based observations and provide continuous spatial and temporal coverage. Hasan et al. [43] assessed physical water scarcity in Africa using GRACE and TRMM satellite data, demonstrating the utility of satellite-based observations in quantifying groundwater depletion and surface water changes. Feng et al. [44] evaluated groundwater depletion in North China using GRACE data and ground-based measurements, highlighting the importance of integrating satellite and in situ data for accurate assessments.

These approaches demonstrate the potential of remote sensing data to overcome limitations in ground-based observations and provide continuous spatial and temporal coverage of key hydrological variables. However, they also face challenges such as the need for validation with ground-based data and potential inaccuracies in capturing local-scale heterogeneities. Despite these limitations, remote sensing-based approaches prove to be valuable tools for understanding groundwater dynamics and supporting sustainable management practices in data-limited environments.

3.8. Isotope-Based Methods

Isotope-based methods [45,46] offer valuable insights into groundwater recharge processes and aquifer interactions in data-scarce regions. Mattei et al. [45] presents an innovative approach using pore water isotope fingerprints to understand spatiotemporal groundwater recharge variability in ungauged watersheds. This method extends 1D unsaturated zone flow modeling from the profile scale to the watershed scale using GIS-based index methods. The approach allows for the estimation of soil hydraulic parameters and recharge rates based on a single field campaign, without requiring long-term monitoring. It successfully captures both the dynamics and quantity of recharge at different scales, with results comparing well to those from spatial water balance models calibrated using long-term discharge data. Rusli et al. [46] combined environmental water tracer (EWT) data analysis, including stable isotopes, with numerical groundwater flow modeling to quantify aquifer interactions. This integrated approach provides both qualitative and quantitative insights into groundwater dynamics, helping to validate model performance in data-scarce areas. However, these methods face some limitations, such as the need for specialized isotope analysis, potential uncertainties in age estimates, and challenges in representing complex karst or fractured systems. Despite these constraints, isotope-based methods prove to be powerful tools for understanding recharge processes and groundwater dynamics in data-limited environments, offering unique insights that may not be obtainable through other techniques.

3.9. Global Model Downscaling

Global model downscaling, Res. [38,47] offers a valuable approach for improving groundwater resource estimation in data-scarce regions. Ben-Salem et al. [47] focused on mapping steady-state groundwater levels in the Mediterranean region, specifically the Iberian Peninsula, by combining global groundwater models with geostatistical downscaling techniques and in situ observations. This method leverages the wide spatial coverage of global models while enhancing spatial resolution through downscaling. The study found that conditioning the average simulated water table depth with at least 50% of available observations (approximately three wells per 1000 km^2) resulted in a well-reproduced spatial groundwater pattern (R^2 = 0.65). Therefore, Ref. [38] combined a lumped hydrological model (HBV) with remote sensing data to evaluate water resources in semi-arid, high-altitude ungauged watersheds. This innovative approach allows for the assessment of water resources in challenging environments where traditional data collection is difficult. Both studies highlight the potential of integrating global models with local data and remote sensing products to overcome limitations in data-scarce regions. However, these approaches still face challenges, such as the reliance on the accuracy of global models and the need for some in situ data for validation, which can be limited in truly data-scarce areas.

3.10. Integrated Modeling Approaches

Integrated modeling approaches [19,24,25,48,49] offer comprehensive solutions for groundwater assessment and management in data-scarce regions. Ref. [19] presented an integrated hydrogeological modeling approach using MODFLOW-OWHM, which combines remote sensing, rainfall-runoff modeling, and 3D dynamic modeling to assess groundwater resources in hard-rock semi-arid terrain. This approach allows for a detailed representation of surface–groundwater interactions and irrigation abstraction. Rafik et al. [24] showcased a novel combination of SWAT hydrological modeling, downscaled GRACE satellite data, and machine learning techniques for groundwater level forecasting, integrating multiple data sources to overcome data scarcity. Klaas et al. [49] introduced the Head-Guided Zonation method combined with particle-tracking simulation for developing groundwater vulnerability zones in karst areas, allowing for the spatial variation of aquifer properties with minimal input data. Fletcher et al. [25] presented an adaptive management framework that integrates Bayesian learning with stochastic dynamic programming, enabling flexible decision-making under uncertainty. Mazzoni et al. [50] focused on forecasting water budget deficits and groundwater depletion in major fossil aquifer systems in North Africa and the Arabian Peninsula, utilizing long-term climate and hydrological data to predict future water scarcity. Pan et al. [51] presented an integrated modeling approach to assess the impact of climate change on groundwater and surface water in the South Aral Sea area, combining climate models with hydrological and groundwater flow models to predict future water resource availability. Alcalà et al. [52] proposed a feasible methodology for groundwater resource modeling in sparse-data drylands, applying it to the Amtoudi Oasis in the northern Sahara. This approach combines limited field data with remote sensing and modeling techniques to support sustainable groundwater use. Again, in the Sahara region, Ref. [53] addressed the issue of hydrologic data scarcity in the Tindouf basin, presenting strategies for dealing with limited data through the use of remote sensing, modeling, and adaptive management approaches.

These papers collectively underscore the importance of leveraging multiple data sources and advanced technologies to overcome data scarcity, providing robust estimates and sustainable management strategies for groundwater resources in arid and semi-arid regions. Integrated approaches offer the advantage of combining multiple data sources and modeling techniques to provide comprehensive water resource assessments. However, they often face challenges related to data quality, scale transferability, and the representation of complex hydrogeological processes. The choice of integrated modeling approach ultimately depends on the specific context, available data, and management objectives of the study area.

4. Discussion

4.1. Methodologies Comparison

The methodologies presented in the selected papers showcase a wide range of approaches for estimating and managing groundwater resources in data-scarce regions. Coupled hydrological-groundwater models [17–19] offer detailed process representation but require significant data inputs compared with simpler water balance models [29,31,32] which can provide useful insights with limited data. Ref. [29] used readily available physiographic data to estimate runoff coefficients and natural aquifer recharge. This method is particularly suitable for large-scale recharge estimation in data-scarce regions, although they may oversimplify complex hydrological processes.

Comprehensive hydrological models like SWAT [26–28] provide detailed process representation but face parameterization challenges in data-poor environments. Physical-based groundwater models like MODFLOW, FEFLOW, and SWAT are powerful tools for simulating complex hydrogeological systems, but their application in data-scarce regions presents significant challenges. Modelers have developed various strategies to overcome the issue of data scarcity when using these models. A common approach is the integration of remote sensing data to supplement limited ground-based observations. For instance, in [17], Khadim et al. coupled the CREST hydrological model with MODFLOW-NWT, leveraging remote sensing data to overcome data scarcity and provide a comprehensive view of the aquifer system at a fine resolution (500 m). Similarly, Ref. [27] demonstrated the use of satellite-based rainfall estimates (RFE) in combination with local precipitation data to drive a SWAT model in the Kilombero floodplain, Tanzania. Another strategy involves the use of global datasets and downscaling techniques; [47] showcased an approach that combines global groundwater models with geostatistical downscaling and limited in situ observations to map steady-state groundwater levels in the Mediterranean region. This method allows modelers to leverage large-scale datasets while improving spatial resolution through downscaling techniques.

Modelers also employ innovative parameterization methods to address data limitations. In [21], Griffiths et al. describe the parameterization of a national-scale groundwater model (TopNet-GW) for New Zealand using a priori parameter sets derived from national hydrogeological datasets. This approach enables groundwater modeling in ungauged catchments at various scales, demonstrating how readily available datasets can be utilized to overcome local data scarcity. The integration of multiple data sources and modeling techniques is another key strategy. Ref. [24] presents a novel combination of SWAT hydrological modeling, downscaled GRACE satellite data, and machine learning techniques for groundwater level forecasting in Morocco region. This integrated approach helps fill gaps in ground-based observations and provides both spatial and temporal groundwater level estimates. Researchers also adapt model structures to suit data-scarce conditions. In [22], Sahoo et al. introduced an enhanced hillslope-storage Boussinesq model that incorporates surface ponding, unsaturated zone processes, bedrock leakage, and root-zone water balance, making it suitable for data-scarce regions. Such adaptations allow for the representation of key hydrological processes with minimal data requirements. In karst aquifer systems, where data scarcity is often compounded by complex hydrogeology, approaches like the Head-Guided Zonation (HGZ) method [49] allow for spatial variation of aquifer properties with minimal input data. This method divides the model domain into zones of piecewise constant hydraulic properties based on available groundwater level data, providing a way to represent spatial variability in aquifer properties despite limited data. Modelers are also increasingly adopting adaptive management frameworks, as demonstrated in [25], which integrates Bayesian learning on groundwater observations with stochastic dynamic programming. This approach allows for the quantification of learning potential in environmental modeling and supports decision-making under uncertainty, particularly valuable in data-scarce regions where initial uncertainty is high but can be reduced through ongoing monitoring. While these strategies help overcome data scarcity to some extent, it is important to note that they often involve trade-offs between model complexity, data requirements, and uncertainty.

All the above-mentioned models provide detailed insights into system dynamics but often require extensive parameterization and may struggle with feedback complexities in data-limited environments. In contrast, data-driven models, particularly machine learning techniques like GBDT and LSTM-NN [23], can work effectively with limited historical data, capturing temporal dependencies and performing well with small sample sizes. However, they may not capture underlying physical processes not represented in the training data. When employing machine learning (ML) models [23,24] to predict groundwater levels in data-scarce regions, ensuring model performance despite limited data is a critical challenge. The researchers addressed this issue through several strategies to guarantee the robustness of their models. Firstly, the use of ensemble and deep learning algorithms, specifically Gradient Boosting Decision Trees (GBDT) and Long Short-Term Memory Neural Networks (LSTM-NN), was instrumental [23]. These algorithms are designed to work effectively with small sample sizes and limited historical data. For instance, LSTM-NN captures temporal dependencies in the data, which is crucial for predicting groundwater levels over time, while GBDT performs well with small sample sizes and can handle missing data points efficiently. To further enhance model performance, in [23], Gaffoor et al. focused on careful data preprocessing and feature engineering. This involved selecting relevant input features that are most correlated with groundwater levels and transforming the data to meet the assumptions of the models. For example, handling missing data points and normalizing the input data helped in improving the model's stability and accuracy. Additionally, they employed cross-validation techniques to evaluate the model's performance on unseen data. This approach helps in preventing overfitting and provides a more realistic estimate of the model's predictive power in real-world scenarios.

The integration of remote sensing data [24,32,37–44] provides broad spatial coverage, addressing data scarcity issues, but may sacrifice local-scale accuracy. For example, Refs. [32,37] highlight the use of satellite-based datasets like CHADFDM and MODIS to provide continuous spatial and temporal coverage of precipitation, evapotranspiration, and other hydrological variables. This helps in overcoming the limitations of sparse ground-based observations and provides essential input data for models like MODFLOW and SWAT. Geostatistical methods [34,35] excel in spatial interpolation but may not capture underlying physical processes. Ref. [34] demonstrates the use of kriging to analyze groundwater table variability and trends, providing insights into areas with limited data points. This approach helps in quantifying spatial autocorrelation and providing uncertainty estimates, which are crucial in data-scarce environments.

Environmental tracer methods [45,46] offer unique insights into groundwater dynamics but are often limited in spatial and temporal resolution. The Head-Guided Zonation method [49] offers a tailored approach for karst aquifers but may oversimplify complex heterogeneities. Adaptive management approaches [25] introduce flexibility in decision-making but require long-term commitment to data collection. In [25], Fletcher et al. integrated Bayesian learning with stochastic dynamic programming to assess opportunities for learning about groundwater availability over time. This approach allows for flexible decision-making under uncertainty, which is particularly valuable in data-scarce regions where initial uncertainty is high but can be reduced through ongoing monitoring.

The choice between these approaches ultimately depends on the specific context, available data, and management objectives, with many studies (e.g., [19,24,48]) integrating approaches and combining multiple methods to provide the most robust assessments in data-scarce regions.

The complementarity between these approaches is evident in integrated modeling strategies, such as those presented in [24,46], which combine process-based models with data-driven techniques and remote sensing data. This integration leverages the strengths of each method, providing more robust estimates in data-scarce conditions. For instance, the combination of SWAT modeling with machine learning and GRACE satellite data [24] addresses both the need for physical process representation and the ability to work with limited ground-based observations.

In practical applications, the choice between process-based and data-driven models often depends on the specific context and available data. Process-based models are particularly valuable when understanding the underlying physical mechanisms is crucial, such as in climate change impact assessments [20]. Data-driven models, on the other hand, excel in scenarios where rapid predictions are needed and historical data, though limited, is available. The adaptive management framework presented in [25] demonstrates how these approaches can be combined over time, using data-driven models for initial estimates and gradually incorporating more process-based elements as data become available through ongoing monitoring.

Ultimately, the complementarity of these methods suggests that a hybrid approach, tailored to the specific challenges and data availability of each region, may provide the most comprehensive and reliable groundwater resource assessments in data-scarce environments.

4.2. Advantages and Benefits

The diverse methodologies for estimating and managing groundwater resources in data-scarce regions offer several advantages and benefits. Coupled hydrological-groundwater models [17–19] provide comprehensive representations of surface–groundwater interactions, allowing for fine-resolution modeling at regional scales. Spatially distributed water balance models [29–32] offer the ability to estimate recharge over large areas with limited data, incorporating land use and soil information. Machine learning techniques [23,24] can work effectively with limited historical data, capturing temporal dependencies and performing well with small sample sizes. The integration of remote sensing data [24,32,37–44] helps overcome limitations in ground-based observations, providing continuous spatial and temporal coverage of key variables like precipitation and evapotranspiration. In [39], Sun et al. demonstrate the use of remote sensing and GIS techniques for mapping prospective water resources and monitoring land use/land cover changes in arid regions. This approach provides spatially explicit recharge estimates and is particularly valuable in areas with limited ground-based data.

Environmental tracer methods [45,46] offer insights into recharge processes and aquifer interactions without requiring long-term monitoring, providing both qualitative and quantitative information. Adaptive management approaches [25] allow for flexible decision-making under uncertainty, incorporating learning from new data over time. The use of global models with downscaling techniques [47] leverages large-scale datasets for regional applications, improving spatial resolution through geostatistical methods. In [36], Mohamed et al. employed hydro-geophysical monitoring using gravity data to assess groundwater resources, showcasing the potential of geophysical methods in areas with limited traditional hydrological data. By combining GRACE satellite data with GLDAS land surface models, the approach provides a comprehensive view of groundwater storage changes at a regional scale. This integration allows for continuous spatial and temporal coverage, overcoming limitations of sparse ground-based observations in data-scarce regions. The method enables the estimation of long-term average recharge and the analysis of groundwater mass balance components.

Three-dimensional groundwater flow modeling with MODFLOW [30,33] allows for the detailed representation of aquifer geometry and properties, enabling the quantification of inter-basin groundwater flow.

The combination of multiple data sources and modeling techniques [19,24,48] provides more robust estimates in data-scarce environments. Refs. [50,51] present integrated modeling approaches that combine multiple data sources and modeling techniques to provide comprehensive water resource assessments under various climate change and socioeconomic scenarios. These methods allow for the evaluation of future water scarcity and groundwater depletion rates, which is crucial for long-term water resource planning. In [40], Wang et al. utilize remote sensing and hydrological modeling to understand the impact of land-cover change on groundwater levels, demonstrating the value of integrating multiple data sources for an improved understanding of groundwater dynamics. Moreover, in [44], the effectiveness of the integration of GRACE satellite data with ground-based measurements for evaluating groundwater depletion is demonstrated, highlighting the synergistic benefits of combining multiple data sources.

Enhanced hillslope-storage Boussinesq models [22] account for multiple hydrological processes and incorporate surface and subsurface interactions. Head-Guided Zonation with particle-tracking simulation [49] allows for the spatial variation of aquifer properties with minimal input data. Overall, each of these approaches demonstrates the potential to overcome data limitations and provide valuable insights for sustainable groundwater management in challenging environments.

4.3. Limits and Challenges

Despite their advantages, the methodologies presented in the literature review face several limitations and challenges in data-scarce environments. Coupled hydrological-groundwater models [17–19] may struggle with feedback complexities and error propagation between components. Spatially distributed models [29,32] often rely on empirical relationships and may oversimplify complex hydrological processes, particularly in karst or fractured rock systems. Machine learning approaches [23,24] are heavily dependent on data quality and quantity, and may not capture underlying physical processes not represented in historical data. Remote sensing-based methods [24,37–44] require validation with ground-based data and may not capture local-scale heterogeneities, with accuracy varying by region and sensor type. For instance, the knowledge-driven Analytical Hierarchy Process (AHP) technique used in [39] relies heavily on expert judgment for weighting criteria, which can introduce subjectivity and potential bias.

Environmental tracer methods [45,46] can be limited by uncertainties in age estimates and coarse spatial resolution. Comprehensive models like SWAT [26–28] require extensive parameterization, which is challenging in data-scarce conditions and may lead to equifinality issues. Geostatistical methods [34,35] assume the stationarity of spatial correlation, which may not hold in complex hydrogeological settings. Many approaches struggle with representing karst systems [48,49] due to their inherent complexity and heterogeneity. Global groundwater models [20,47] face challenges in representing local-scale processes and heterogeneities due to their coarse resolution. Papers focusing on data from GRACE satellites [40,42–44] face limitations in spatial and temporal resolution, with uncertainties increasing for smaller study areas. The coarse resolution of GRACE data (typically around 300 km) makes it challenging to apply these methods at local scales relevant for groundwater management. Additionally, separating groundwater signals from other water storage components requires auxiliary datasets, which may introduce additional uncertainties. The use of steady-state assumptions in some models [21,33] may not capture temporal variability in groundwater-surface water exchanges. Moreover, the integration of multiple data sources in [36,50] helps to overcome some limitations of individual datasets but also introduces challenges in data harmonization and error propagation.

Integrated modeling approaches [19,48] often require extensive data from multiple sources and may simplify complex processes due to data limitations. These modeling approaches also face challenges related to parameter estimation and model structure in data-scarce environments [50,51]. The lack of comprehensive ground-based observations for calibration and validation can lead to significant uncertainties in model outputs. Furthermore, representing complex groundwater–surface water interactions and human interventions in these models remains a challenge.

Adaptive management frameworks [25] require long-term commitment to data collection and model updating, which can be challenging in resource-constrained environments. Papers focusing on specific regions or aquifer systems [41,52,53] highlight the difficulties in transferring methodologies between different hydrogeological settings. The unique characteristics of each study area, such as the complex geology of karst systems or the deep aquifers in arid regions, require tailored approaches that may not be easily generalizable.

A common challenge across many of these studies is the validation of results in data-scarce regions. The lack of comprehensive ground-based monitoring networks makes it difficult to assess the accuracy of remotely sensed or modeled groundwater estimates. Additionally, many of these methods struggle to capture the dynamic nature of groundwater systems, particularly in regions experiencing rapid changes due to climate variability or human activities.

Overall, many of these methods face challenges related to data quality, scale transferability, and the representation of complex hydrogeological processes in data-scarce regions. Finally, while these advanced techniques offer valuable insights into groundwater resources, there remains a gap between scientific findings and practical application for groundwater management in many African [36,39,42,43,52,53] and Middle Eastern countries [40,41,44,51]. Translating complex methodologies into actionable information for local water resource managers and policymakers remains a significant challenge.

4.4. Input Data Requirement

The input data requirements for groundwater modeling in data-scarce regions vary across the methodologies explored, but generally encompass a range of spatial and temporal datasets. For coupled hydrological-groundwater models [17–22], key inputs include digital elevation models, land use/land cover data, soil data, meteorological data (precipitation, temperature), geological data, and limited groundwater level observations for calibration [17,19], or streamflow time series [18]. Spatially distributed water balance models like WetSpass [32] require land use/land cover maps, soil maps, digital elevation models, meteorological data, and limited groundwater level data for validation. Inverse hydrogeological models [30,31] require precipitation, temperature and/or aridity index information, digital elevation models, and land cover data. Machine learning approaches [23,24] typically need historical groundwater level data, meteorological data, and potentially remote sensing data like GRACE. Refs. [39,40] utilize remote sensing and GIS data, including digital elevation models, land use/land cover maps, geological information, and climate data such as rainfall. Refs. [50,51] integrate multiple data sources, including climate data, hydrological observations, and socioeconomic information. Integrated modeling approaches, such as [53], address hydrologic data scarcity by leveraging a combination of limited ground observations, remote sensing data, and modeling outputs. Comprehensive hydrological models such as SWAT [26–28] demand extensive parameterization, including digital elevation models, land use/land cover data, soil data, meteorological data, and streamflow data for calibration. Three-dimensional groundwater flow modeling with MODFLOW [20,33] requires detailed geological data, aquifer properties, recharge estimates, boundary conditions, groundwater level observations, and streamflow data when available. Geostatistical methods [34,35] primarily need groundwater level measurements and well location coordinates. Satellite-based approaches [37] rely heavily on remote sensing datasets, often complemented by ground-based observations for validation, and [36] relies on gravity data from satellite observations, specifically GRACE data, along with land surface model outputs. Environmental tracer methods [45,46] require specialized isotope analysis data along with hydrogeological information. The national-scale groundwater parameterization approach [21] utilizes extensive national hydrogeological datasets, digital elevation models, river network data, and geological maps.

Overall, while the specific data requirements vary, most methods benefit from a combination of spatial datasets (e.g., topography, geology, land use), temporal datasets (e.g., meteorological data, groundwater levels), and ancillary information on aquifer properties and boundary conditions.

4.5. Outputs and Uncertainties Handling

The methodological approaches explored in the reviewed papers produce various outputs and handle uncertainties in different ways. Coupled hydrological-groundwater models [17–22] provide comprehensive spatial and temporal outputs of groundwater levels and fluxes but face uncertainties from error propagation between model components. To address this, Ref. [17] conducted error propagation analysis to quantify how bias in forcing datasets contributes to systematic and random errors, while [18] conducted a sensitivity analysis to model parameters. Physiography-based indirect methods [29] yield estimates of runoff coefficients and recharge rates, with uncertainties stemming from simplifications of complex hydrological processes.

Machine learning (ML) techniques in [23] offer predictions of groundwater levels with associated confidence intervals, handling uncertainties through ensemble approaches and probabilistic outputs. The integration of multiple data sources, such as remote sensing data and limited ground-based observations, was crucial in augmenting the available dataset. This multi-source approach helped in filling gaps in the historical data and provided a more comprehensive view of the groundwater system, thereby enhancing the overall performance and reliability of the ML models. In summary, the researchers guaranteed the performance of their ML models by leveraging ensemble and deep learning algorithms, careful data preprocessing, cross-validation techniques, and the integration of multiple data sources. These strategies collectively helped in mitigating the challenges associated with limited data, ensuring that the models could provide accurate and reliable predictions of groundwater levels in data-scarce regions.

Distributed hydrological models like SWAT [26–28] produce detailed water balance components and streamflow estimates, with uncertainties related to parameter equifinality. These studies often use sensitivity analysis and multi-objective calibration to address uncertainties. The inverse hydrogeological balance method application in [30] provides spatially distributed groundwater resource estimation but uncertainties are linked to DEM resolution and ET calculation methods. In [33], MODFLOW was used to assess inter-basin groundwater flow, providing detailed groundwater level and flow fields. It handles uncertainties through careful calibration and validation, achieving a Nash–Sutcliffe Efficiency Index of 0.8. Geostatistical techniques [34,35] produce interpolated groundwater level maps with estimation variances, explicitly quantifying spatial prediction uncertainties. Satellite-based approaches [36] offer continuous spatial and temporal coverage of recharge estimates but face uncertainties in the accuracy of remote sensing products. These studies often validate results against ground-based observations. In [36], Mohamed et al., who used hydro-geophysical monitoring with gravity data, addressed uncertainty by combining multiple data sources, including GRACE satellite data, GLDAS land surface models, and in situ observations. This integrated approach helps to reduce uncertainties associated with individual datasets. In [50], Mazzoni et al. conducted a comprehensive error analysis to assess uncertainties in water budget modeling, calculating average standard deviation errors for spatial approximations and propagating these errors through the entire simulation. Alcalà et al. [52] addressed uncertainty in sparse-data drylands by combining limited field data with remote sensing and modeling techniques. In [43,51], researchers used ensemble approaches, combining multiple climate models or scenarios to capture a wider range of potential outcomes and associated uncertainties. Ref. [53] specifically focused on dealing with hydrologic data scarcity, likely employing techniques such as data assimilation or Bayesian methods to quantify and reduce uncertainties.

Integrated modeling approaches [24,46] provide comprehensive water resource assessments, handling uncertainties through multi-model ensembles and data assimilation techniques. The adaptive management framework presented in [25] produced dynamic water resource plans with quantified learning potential, explicitly incorporating uncertainty into decision-making processes.

In addition, Ref. [54] provided a comprehensive review of uncertainty sources in groundwater recharge estimation, focusing on data-scarce tropical, arid, and semiarid regions. While not presenting a specific methodology, this review offers valuable insights into the challenges and potential solutions for recharge estimation in data-limited environments. This paper emphasizes the importance of uncertainty analysis in recharge estimation and highlights that most studies use multiple methods to provide a range of recharge estimates rather than conducting detailed uncertainty analyses for individual methods. This approach, while providing insight into the potential range of recharge, lacks the ability to identify uncertainties in individual methods or input data.

Overall, while the specific outputs and uncertainty handling methods vary, there is a growing trend towards more comprehensive uncertainty quantification and communication in groundwater modeling studies.

5. Research Priorities and Future Directions

Based on the reviewed papers, several key research priorities emerge for groundwater modeling in data-scarce regions. One promising direction is the integration of multi-source data, combining traditional ground-based measurements with remote sensing and geophysical techniques. Scientists could explore ways to merge GRACE satellite data with ground-based piezometric measurements and geophysical surveys to provide a more comprehensive understanding of groundwater dynamics at various scales. This approach could be particularly valuable in regions where traditional monitoring networks are sparse or non-existent. For instance, Ref. [24] highlights the potential of integrating SWAT hydrological modeling, downscaled GRACE satellite data, and machine learning techniques for groundwater level forecasting. Future research should explore more sophisticated ways to merge remote sensing data, global model outputs, and limited ground observations to improve model accuracy and reliability.

Another important research direction is the development of advanced machine learning and deep learning algorithms specifically tailored for groundwater modeling in data-scarce environments, as highlighted in [23], which compared ensemble and deep learning algorithms. Future studies could focus on transfer learning or federated learning approaches, which could help address the challenges of limited local data by leveraging knowledge from data-rich regions or related domains. Additionally, researchers should investigate the potential of hybrid models that combine physics-based and data-driven approaches, potentially offering more robust and interpretable results in data-limited contexts.

Improving methods for downscaling global or regional datasets to local scales is another crucial research area, as emphasized in [38,47]. This could involve developing more sophisticated statistical or machine learning techniques for disaggregating coarse resolution remote sensing data or global model outputs to provide meaningful inputs for local groundwater models. Researchers should also focus on developing robust uncertainty quantification methods that account for the multiple sources of uncertainty in data-scarce regions, including input data, model parameters, and conceptual model uncertainties.

Additionally, there is a pressing need to develop robust uncertainty quantification methods that account for the multiple sources of uncertainty in data-scarce regions, including input data, model parameters, and conceptual model uncertainties [54].

Future research should also focus on incorporating citizen science data and local knowledge into groundwater models to help fill data gaps and improve model performance. This may involve developing new data collection protocols and quality control measures for citizen-generated data, as suggested by the citizen science initiative in [17]. To effectively implement this approach, several practical steps can be taken. Firstly, developing standardized data collection protocols tailored to citizen scientists is crucial. These protocols should be simple yet rigorous, ensuring data quality while remaining accessible to non-experts. For instance, citizens could be trained to measure and report groundwater levels in private wells using low-cost water level meters, following a consistent methodology. Quality control measures could include automated data validation checks, cross-referencing with existing data, and periodic expert verification. Additionally, user-friendly mobile applications could be developed to facilitate data collection and submission, incorporating GPS location tagging and photo documentation to enhance data reliability. To integrate local knowledge, structured interviews and participatory mapping exercises could be conducted with long-term residents to capture historical trends and identify critical groundwater features. This qualitative information could then be systematically coded and incorporated into model parameterization or used to validate model outputs. Furthermore, establishing a centralized, open-access database for citizen-generated data would enable researchers and water managers to easily incorporate this information into their models. Research, then, is also needed on how to effectively combine hard data with soft information, such as expert knowledge and qualitative observations, in model development and calibration.

Developing adaptive modeling frameworks that can evolve as new data become available is another important direction. This includes creating flexible model structures that can be easily updated and methods for assimilating new data streams in real-time. Such adaptive approaches are crucial for supporting sustainable groundwater management in regions where data availability may change over time, as evidenced in [25]. To achieve these research directions, key technological innovations will be required, such as the development of low-cost, robust sensors for groundwater monitoring, improved algorithms for processing and interpreting remote sensing data, and advanced computational techniques for handling large, heterogeneous datasets in groundwater modeling applications.

Finally, there is a need for more comprehensive studies that compare and integrate different modeling approaches, as seen in [24,46]. This could involve systematic comparisons of various techniques (e.g., physical-based models, data-driven approaches, and hybrid methods) across different hydrogeological settings and data availability scenarios. Such comparative studies would help identify the most appropriate modeling strategies for different contexts and guide future research efforts in groundwater modeling for data-scarce regions.

6. Conclusions

This comprehensive literature review of 38 papers reveals a diverse array of methodologies for estimating and managing groundwater resources in data-scarce regions and for uncertainty handling. The approaches range from coupled hydrological-groundwater models and machine learning techniques to satellite-based datasets, isotope fingerprinting methods, and geophysical monitoring using gravity data. Each methodology offers unique advantages in addressing specific aspects of groundwater assessment and management in data-poor environments. The integration of remote sensing data, machine learning algorithms, and global model downscaling techniques has shown promise in overcoming data limitations. For instance, Refs. [32,37,38] demonstrate the value of satellite-based datasets like CHADFDM and MODIS in providing continuous spatial and temporal coverage of crucial hydrological variables. Machine learning approaches, such as those presented in [23,24], offer the ability to make accurate predictions with limited historical data, capturing temporal dependencies and performing well with small sample sizes. However, these methods also face challenges related to data quality, scale transferability, and the representation of complex hydrogeological processes.

This review highlights a trend towards combining multiple data sources and modeling techniques to provide more robust estimates. For instance, several studies integrate remote sensing data with traditional hydrological models or combine machine learning with physical-based approaches. This multi-faceted approach allows researchers to leverage the strengths of different methodologies while mitigating their individual weaknesses.

Innovative approaches for data-scarce regions are also emerging, showcasing the potential of hydro-geophysical monitoring using gravity data to assess groundwater resources in areas with limited traditional hydrological data [36], or proposing a feasible methodology for groundwater resource modeling in sparse-data drylands, combining limited field data with remote sensing and modeling techniques [52].

The adaptive management frameworks and uncertainty analysis methods presented in some studies offer valuable insights into decision-making under uncertainty, which is crucial in data-scarce environments.

Despite the advancements, challenges persist. Many methods struggle with representing complex systems such as karst aquifers or capturing local-scale heterogeneities. The trade-off between model complexity and data requirements remains a significant consideration in choosing appropriate methodologies. Papers focusing on specific regions or aquifer systems (e.g., [41,52]) highlight the difficulties in transferring methodologies between different hydrogeological settings.

A common challenge across many studies is the validation of results in data-scarce regions. The lack of comprehensive ground-based monitoring networks makes it difficult to assess the accuracy of remotely sensed or modeled groundwater estimates. Additionally, many methods struggle to capture the dynamic nature of groundwater systems, particularly in regions experiencing rapid changes due to climate variability or human activities.

Future research should focus on improving the integration of diverse data sources, enhancing the representation of complex hydrogeological processes in simplified models, and developing robust uncertainty quantification methods tailored for data-scarce conditions.

Ultimately, the choice of methodology depends on the specific context, available data, and management objectives of the study area, emphasizing the need for flexible and adaptable approaches in groundwater resource estimation and management in data-scarce regions.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A

Table A1. Classification of Methodological approaches, Advantages, Limitations, Input data requirements, and References.

Methodology Type	Advantages	Limitations	Input Data Required	Reference
Coupled hydrological-groundwater models	Comprehensive representation of surface–groundwater interactions Fine-resolution modeling at regional scales Incorporates both surface water and groundwater processes Accounts for multiple hydrological processes	May not fully capture complex feedback between surface and groundwater Prone to error propagation Computationally intensive	Digital elevation model Land use/land cover data Soil data Meteorological data Geological data Limited groundwater level observations	[17–22]
Machine learning techniques	Can work with limited historical data Captures temporal dependencies Potential for accurate predictions in data-scarce environments	Performance depends on data quality and quantity May not capture complex hydrogeological processes Requires careful tuning and training	Historical groundwater level data Meteorological data Remote sensing data (e.g., GRACE) Hydrological model outputs	[23–25]
Distributed hydrological models (e.g., SWAT)	Comprehensive representation of hydrological processes Suitable for large catchments Incorporates land use and management practices	Requires extensive parameterization Computationally intensive for long-term simulations May struggle with groundwater processes	Digital elevation model Land use/land cover data Soil data Meteorological data Streamflow data for calibration	[26–28]
Water balance models	Suitable for large-scale recharge estimation Uses readily available data Provides spatially explicit recharge estimates	May oversimplify complex hydrological processes Limited temporal resolution (often annual) Relies on empirical relationships	Land use/land cover maps Soil maps Digital elevation model Meteorological data	[29–32]

Table A1. *Cont.*

Methodology Type	Advantages	Limitations	Input Data Required	Reference
3D groundwater flow modeling (e.g., MODFLOW)	Detailed representation of aquifer geometry and properties Can simulate complex boundary conditions Enables quantification of inter-basin groundwater flow	Requires extensive geological and hydrogeological data Challenging to parameterize in data-scarce conditions Computationally intensive	Detailed geological data Aquifer properties Recharge estimates Boundary conditions Groundwater level observations	[20,33]
Geostatistical and geophysical techniques	Provides spatial interpolation and uncertainty estimates Can work with limited data points Incorporates spatial autocorrelation	Assumes stationarity of spatial correlation May not capture complex hydrogeological processes Accuracy depends on spatial distribution of data points	Groundwater level measurements Well location coordinates Auxiliary variables (e.g., elevation, distance to rivers) Gravity data	[34–36]
Remote sensing-based approaches	Provides continuous spatial and temporal coverage Overcomes limitations of ground-based observations Suitable for large-scale applications	Requires validation with ground-based data May not capture local-scale heterogeneities Accuracy varies by region and sensor type	Satellite imagery (e.g., GRACE, MODIS) Digital elevation model Limited ground-based observations for validation	[32,37–44]
Isotope-based methods	Provides insights into recharge processes without long-term monitoring Captures both dynamics and quantity of recharge Useful for ungauged watersheds	Requires specialized isotope analysis Limited spatial resolution Uncertainties in age estimates	Water samples Precipitation isotope data Soil physical properties Meteorological data	[45,46]
Global model downscaling	Leverages global models for regional applications Improves spatial resolution Incorporates local observations to enhance accuracy	Relies on accuracy of global models May not capture local-scale heterogeneities Requires some in-situ data for validation	Global model outputs In-situ groundwater level observations Digital elevation model Hydrogeological data	[38,47]
Integrated modeling approaches	Combines multiple data sources and techniques Provides comprehensive water resource assessments Addresses data scarcity issues	Complex to implement and calibrate May propagate errors across multiple model components Computationally intensive	Various data types depending on integrated models Remote sensing data In-situ observations Hydrogeological data	[19,24,25,48–53]

References

1. Poeter, E.; Fan, Y.; Cherry, J.; Wood, W.; Mackay, D. *Groundwater in Our Water Cycle—Getting to Know Earth's Most Important Fresh Water Source*; The Groundwater Project: Guelph, ON, Canada, 2020.
2. Drias, T.; Khedidja, A.; Belloula, M.; Badraddine, S.; Saibi, H. Groundwater modelling of the Tebessa-Morsott alluvial aquifer (northeastern Algeria): A geostatistical approach. *Groundw. Sustain. Dev.* **2020**, *11*, 100444. [CrossRef]
3. Ren, B.; Liang, J.; Dai, J.F.; Li, X.; Sun, T.; Han, Y. Trend prediction of regional groundwater level with GISFEFLOW model in Beijing Mihuaishun plain, China. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *330*, 052037. [CrossRef]
4. García-Rodríguez, M.; Antón, L.; Martinez-Santos, P. Estimating groundwater resources in remote desert environments by coupling geographic information systems with groundwater modeling (Erg Chebbi, Morocco). *J. Arid Environ.* **2014**, *110*, 19–29. [CrossRef]
5. Roldán-Cañas, J.; Moreno-Pérez, M.F. Water and Irrigation Management in Arid and Semiarid Zones. *Water* **2021**, *13*, 2446. [CrossRef]
6. Maxwell, R.M.; Condon, L.E.; Kollet, S.J. A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geosci. Model Dev.* **2015**, *8*, 923–937. [CrossRef]
7. Bierkens, M.F.P.; Bell, V.; Burek, P.; Chaney, N.; Condon, L.E.; David, C. Hyper-resolution global hydrological modelling: What's next? *Hydrol. Process.* **2015**, *29*, 310–320. [CrossRef]
8. Clark, M.P.; Fan, Y.; Lawrence, D.M.; Adam, J.C.; Bolster, D.; Gochis, D.J. Improving the representation of hydrologic processes in Earth System Models. *Water Resour. Res.* **2015**, *51*, 5929–5956. [CrossRef]

9. Condon, L.E.; Kollet, S.; Bierkens, M.F.P.; Fogg, G.E.; Maxwell, R.M.; Hill, M.C. Global groundwater modeling and monitoring: Opportunities and challenges. *Water Resour. Res.* **2021**, *57*, e2020WR029500. [CrossRef]
10. Secci, D.; Saysel, A.K.; Uygur, İ.; Yoloğlu, O.C.; Zanini, A.; Copty, N.K. Modeling for sustainable groundwater management: Interdependence and potential complementarity of process-based, data-driven and system dynamics approaches. *Sci. Total Environ.* **2024**, *951*, 175491. [CrossRef]
11. Harbaugh, A.W. *MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model—The Ground-Water Flow Process*; U.S. Geological Survey Techniques and Methods: Reston, VA, USA, 2005.
12. Diersch, H.-J.G. *FEFLOW—Finite Element Modeling of Flow, Mass and Heat Transport in Porous and Fractured Media*; Springer: Berlin/Heidelberg, Germany, 2013.
13. Lozano Hernández, B.L.; Marín Celestino, A.E.; Martínez Cruz, D.A.; Ramos Leal, J.A.; Hernández Pérez, E.; García Pazos, J.; Almanza Tovar, O.G. A Systematic Review of the Current State of Numerical Groundwater Modeling in American Countries: Challenges and Future Research. *Hydrology* **2024**, *11*, 179. [CrossRef]
14. Lall, U.; Josset, L.; Russo, T. A snapshot of the world's groundwater challenges. *Annu. Rev. Environ. Resour.* **2020**, *45*, 171–194. [CrossRef]
15. Gleeson, T.; Wagener, T.; Doell, P.; Bierkens, M.; Wada, Y.; Lo, M.-H. GMD Perspective: The quest to improve the evaluation of groundwater representation in continental to global scale models. *Geosci. Model Dev. Discuss.* **2021**, *14*, 7545–7551. [CrossRef]
16. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
17. Khadim, F.K.; Dokou, Z.; Lazin, R.; Moges, S.; Bagtzoglou, A.C.; Anagnostou, E. Groundwater modeling in data scarce aquifers: The case of Gilgel-Abay, Upper Blue Nile, Ethiopia. *J. Hydrol.* **2020**, *590*, 125214. [CrossRef]
18. Borzì, I.; Bonaccorso, B.; Fiori, A. A modified IHACRES rainfall-runoff model for predicting the hydrologic response of a river basin connected with a deep groundwater aquifer. *Water* **2019**, *11*, 2031. [CrossRef]
19. Ebrahim, G.Y.; Villholth, K.G.; Boulos, M. Integrated hydrogeological modelling of hard-rock semi-arid terrain: Supporting sustainable agricultural groundwater use in Hout catchment, Limpopo Province, South Africa. *Hydrogeol. J.* **2019**, *27*, 965–981. [CrossRef]
20. Rödiger, T.; Geyer, S.; Odeh, T.; Siebert, C. Data scarce modelling the impact of present and future groundwater development on Jordan multiaquifer groundwater resources. *Sci. Total Environ.* **2023**, *870*, 161729. [CrossRef] [PubMed]
21. Griffiths, J.; Yang, J.; Woods, R.; Zammit, C.; Porhemmat, R.; Shankar, U.; Rajanayaka, C.; Ren, J.; Howden, N. Parameterization of a National Groundwater Model for New Zealand. *Sustainability* **2023**, *15*, 13280. [CrossRef]
22. Sahoo, S.; Sahoo, B.; Panda, S.N. Hillslope-storage Boussinesq model for simulating subsurface water storage dynamics in scantily-gauged catchments. *Adv. Water Resour.* **2018**, *121*, 219–234. [CrossRef]
23. Gaffoor, Z.; Pietersen, K.; Jovanovic, N.; Bagula, A.; Kanyerere, T.; Ajayi, O.; Wanangwa, G. A Comparison of Ensemble and Deep Learning Algorithms to Model Groundwater Levels in a Data-Scarce Aquifer of Southern Africa. *Hydrology* **2022**, *9*, 125. [CrossRef]
24. Rafik, A.; Brahim, Y.A.; Amazirh, A.; Ouarani, M.; Bargam, B.; Ouatiki, H.; Bouslihim, Y.; Bouchaou, L.; Chehbouni, A. Groundwater level forecasting in a data-scarce region through remote sensing data downscaling, hydrological modeling, and machine learning: A case study from Morocco. *J. Hydrol. Reg. Stud.* **2023**, *50*, 101569. [CrossRef]
25. Fletcher, S.; Strzepek, K.; Alsaati, A.; Weck, O. Learning and flexibility for water supply infrastructure planning under groundwater resource uncertainty. *Environ. Res. Lett.* **2019**, *14*, 114022. [CrossRef]
26. Tudose, N.C.; Marin, M.; Cheval, S.; Ungurean, C.; Davidescu, S.O.; Tudose, O.N.; Mihalache, A.L.; Davidescu, A.A. Swat model adaptability to a small mountainous forested watershed in central Romania. *Forests* **2021**, *12*, 860. [CrossRef]
27. Näschen, K.; Diekkrüger, B.; Leemhuis, C.; Steinbach, S.; Seregina, L.S.; Thonfeld, F.; van der Linden, R. Hydrological modeling in data-scarce catchments: The Kilombero floodplain in Tanzania. *Water* **2018**, *10*, 599. [CrossRef]
28. Mango, L.M.; Melesse, A.M.; McClain, M.E.; Gann, D.; Setegn, S.G. Land use and climate change impacts on the hydrology of the upper Mara River Basin, Kenya: Results of a modeling study to support better resource management. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2245–2258. [CrossRef]
29. Ghiglieri, G.; Carletti, A.; Pittalis, D. Runoff coefficient and average yearly natural aquifer recharge assessment by physiography-based indirect methods for the island of Sardinia (Italy) and its NW area (Nurra). *J. Hydrol.* **2014**, *519*, 1779–1791. [CrossRef]
30. Borzì, I.; Bonaccorso, B.; Aronica, G.T. The role of dem resolution and evapotranspiration assessment in modeling groundwater resources estimation: A case study in Sicily. *Water* **2020**, *12*, 2980. [CrossRef]
31. Borzì, I.; Bonaccorso, B. Quantifying groundwater resources for municipal water use in a data-scarce region. *Hydrology* **2021**, *8*, 184. [CrossRef]
32. Demissie, E.S.; Gashaw, D.Y.; Altaye, A.A.; Demissie, S.S.; Ayele, G.T. Groundwater Recharge Estimation in Upper Gelana Watershed, South-Western Main Ethiopian Rift Valley. *Sustainability* **2023**, *15*, 1763. [CrossRef]

33. Arce, M.; Orellana-Macías, J.M.; Causapé, J.; Ramajo, J.; Galè, C.; Rossetto, R. Model-based assessment of interbasin groundwater flow in data scarce areas: The Gallocanta Lake endorheic watershed (Spain). *Sustain. Environ. Res.* **2023**, *33*, 32. [CrossRef]
34. Hasan, K.; Paul, S.; Chy, T.J.; Antipova, A. Analysis of groundwater table variability and trend using ordinary kriging: The case study of Sylhet, Bangladesh. *Appl. Water Sci.* **2021**, *11*, 120. [CrossRef]
35. Varouchakis, E.A.; Hristopulos, D.T. Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables. *Adv. Water Resour.* **2013**, *52*, 34–49. [CrossRef]
36. Mohamed, A.; Gonçalvès, J. Hydro-geophysical monitoring of the North Western Sahara Aquifer System's groundwater resources using gravity data. *J. Afr. Earth Sci.* **2021**, *178*, 104188. [CrossRef]
37. Siavashani, N.S.; Jimenez-Martinez, J.; Vaquero, G.; Elorza, F.J.; Sheffield, J.; Candela, L.; Serrat-Capdevila, A. Assessment of CHADFDM satellite-based input dataset for the groundwater recharge estimation in arid and data scarce regions. *Hydrol. Process.* **2021**, *35*, e14250. [CrossRef]
38. Jódar, J.; Carpintero, E.; Martos-Rosillo, S.; Ruiz-Constán, A.; Marín-Lechado, C.; Cabrera-Arrabal, J.A.; Navarrete-Mazariegos, E.; González-Ramón, A.; Lambán, L.J.; Herrera, C.; et al. Combination of lumped hydrological and remote-sensing models to evaluate water resources in a semi-arid high altitude ungauged watershed of Sierra Nevada (Southern Spain). *Sci. Total Environ.* **2018**, *625*, 285–300. [CrossRef] [PubMed]
39. Sun, T.; Cheng, W.; Abdelkareem, M.; Al-Arifi, N. Mapping Prospective Areas of Water Resources and Monitoring Land Use/Land Cover Changes in an Arid Region Using Remote Sensing and GIS Techniques. *Water* **2022**, *14*, 2435. [CrossRef]
40. Wang, W.; Chen, Y.; Wang, W.; Chen, Y.; Hou, Y. Groundwater Level Dynamic Impacted by Land-Cover Change in the Desert Regions of Tarim Basin, Central Asia. *Water* **2023**, *15*, 3601. [CrossRef]
41. Liu, C.; Liu, H.; Yu, Y.; Zhao, W.; Zhang, Z.; Guo, L.; Yetemen, O. Mapping groundwater-dependent ecosystems in arid Central Asia: Implications for controlling regional land degradation. *Sci. Total Environ.* **2021**, *797*, 149027. [CrossRef] [PubMed]
42. Springer, A.; Lopez, T.; Owor, M.; Frappart, F.; Stieglitz, T. The Role of Space-Based Observations for Groundwater Resource Monitoring over Africa. *Surv. Geophys.* **2023**, *44*, 123–172. [CrossRef]
43. Hasan, E.; Tarhule, A.; Hong, Y.; Moore, B., III. Assessment of Physical Water Scarcity in Africa Using GRACE and TRMM Satellite Data. *Remote Sens.* **2019**, *11*, 904. [CrossRef]
44. Feng, W.; Zhong, M.; Lemoine, J.M.; Biancale, R.; Hsu, H.T.; Xia, J. Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements. *Water Resour. Res.* **2013**, *49*, 2110–2118. [CrossRef]
45. Mattei, A.; Barbecot, F.; Goblet, P.; Guillon, S. Pore water isotope fingerprints to understand the spatiotemporal groundwater recharge variability in ungauged watersheds. *Vadose Zone J.* **2020**, *19*, e20066. [CrossRef]
46. Rusli, S.R.; Weerts, A.H.; Mustafa, S.M.; Irawan, D.E.; Taufiq, A.; Bense, V.F. Quantifying aquifer interaction using numerical groundwater flow model evaluated by environmental water tracer data: Application to the data-scarce area of the Bandung groundwater basin, West Java, Indonesia. *J. Hydrol. Reg. Stud.* **2023**, *50*, 2214–5818. [CrossRef]
47. Ben-Salem, N.; Reinecke, R.; Copty, N.K.; Gómez-Hernández, J.J.; Varouchakis, E.A.; Karatzas, G.P.; Rode, M.; Jomaa, S. Mapping steady-state groundwater levels in the Mediterranean region: The Iberian Peninsula as a benchmark. *J. Hydrol.* **2023**, *626*, 130207. [CrossRef]
48. de Salis, H.H.; da Costa, A.M.; Künne, A.; Fernandes, L.F.; Pacheco, F.A. Conjunctive water resources management in densely urbanized karst areas: A study in the sete Lagoas Region, state of Minas Gerais, Brazil. *Sustainability* **2019**, *11*, 3944. [CrossRef]
49. Klaas, D.K.; Imteaz, M.A.; Arulrajah, A. Development of groundwater vulnerability zones in a data-scarce eogenetic karst area using Head-Guided Zonation and particle-tracking simulation methods. *Water Res.* **2017**, *122*, 17–26. [CrossRef] [PubMed]
50. Mazzoni, A.; Heggy, E.; Scabbia, G. Forecasting water budget deficits and groundwater depletion in the main fossil aquifer systems in North Africa and the Arabian Peninsula. *Glob. Environ. Chang.* **2018**, *53*, 157–173. [CrossRef]
51. Pan, X.; Wang, W.; Liu, T.; Akmalov, S.; de Maeyer, P.; van de Voorde, T. Integrated modeling to assess the impact of climate change on the groundwater and surface water in the South Aral Sea area. *J. Hydrol.* **2022**, *614*, 128641. [CrossRef]
52. Alcalá, F.J.; Martín-Martín, M.; Guerrera, F.; Martínez-Valderrama, J.; Robles-Marín, P. A feasible methodology for groundwater resource modelling for sustainable use in sparse-data drylands: Application to the Amtoudi Oasis in the northern Sahara. *Sci. Total Environ.* **2018**, *630*, 1246–1257. [CrossRef]
53. Gonçalvès, J.; Nutz, A.; Séraphin, P.; Chekireb, A.; Kabiri, L.; Deschamps, P. Dealing with hydrologic data scarcity: The case of the Tindouf basin. *Comptes Rendus Géoscience* **2023**, *355*, 281–300. [CrossRef]
54. Beyene, T.D.; Zimale, F.A.; Gebrekristos, S.T. A review on sources of uncertainties for groundwater recharge estimates: Insight into data scarce tropical, arid, and semiarid regions. *Hydrol. Res.* **2024**, *55*, 51–66. [CrossRef]