https://jsras.rcc.edu.ly/
Vol.3 No.0 (2026), 74-87
Article history:
Received:
Accepted:
Published: 18Nov.2026

Journal of Sustainable Research in Applied Sciences

JSRAS

A Climate-Coupled Physics-Informed Machine Learning Framework (PINN-LSTM) for Groundwater Dynamics in Data-Scarce Arid Basins: A Murzuq-Inspired Synthetic Twin-Experiment Study

Jabril Ramdan [1], jabril@aju.edu.ly , 0000-0001-7011-8694, Intisar Maamar Makari[2,*], intisar.Mkari@aju.edu.ly , 0009-0009-7651-3885, Seham Saleh Al_Qablawi [3], seham@aju.edu.ly, 0009-0005-1181-4047, Asma Mustafa Abouethlah [4], asma.adalah@aju.edu.ly, 0009-0009-1345-0602, Basma Emhamed Dihom [5], basma.daihoum@aju.edu.ly, 0009-0005-1181-4047

[1] Affiliations of Authors: Department of computer science, Faculty of science and natural Resources, Aljafara University, Almamura, Libya.
[*2] Affiliations of Authors: Department of Mathematic, Faculty of science and natural Resources, Aljafara University, Almamura, Libya.
[3] Affiliations of Authors: Department of Mathematic, Faculty of science and natural Resources, Aljafara University, Almamura, Libya.
[4] Affiliations of Authors: Department of Mathematic, Faculty of science and natural Resources, Aljafara University, Almamura, Libya.
[5] Affiliations of Authors: Department of computer science, Faculty of science and natural Resources, Aljafara University, Almamura, Libya.
* Corresponding author

Index Terms

Groundwater modeling, PINN-LSTM, data scarcity, Murzuq Basin, GRACE.

Abstract

Groundwater management in Saharan basins is constrained by a compounding lack of reliable field monitoring alongside accelerating climatic stress. This study introduces a proof-of-concept hybrid framework that couples Long Short-Term Memory (LSTM) networks with Physics-Informed Neural Network (PINN) constraints to improve groundwater-head prediction under sparse, noisy observational conditions. Because genuine field records for the Murzuq Basin (~350,000 km²) are unavailable, we construct a fully synthetic "twin" basin reproducing its general hydrogeological characteristics, then corrupt the synthetic "ground truth" with realistic data loss (30%) and measurement noise ($\sigma = 0.1\ m$) purely to stress-test the framework. Every numerical result, figure, and table reported is synthetic and illustrative, and none should be read as an actual estimate of groundwater head or depletion in the real Murzuq Basin. A climate-extended Boussinesq constraint, where recharge and evapotranspiration are explicit functions of climate forcing rather than held constant, suppresses physically irrational deviations exhibited by purely statistical models during data loss. Qualitative agreement with GRACE satellite gravimetry trends serves strictly as a plausibility check, since no field calibration data exist. The framework is evaluated through an ablation study, a two-parameter sensitivity analysis, independent numerical cross-validation against a differently coded and discretized simulator (breaking circularity between the data-generating equation and physical constraint), and a head-to-head comparison against modern neural-operator baselines (DeepONet and Fourier Neural Operator) trained on the identical corrupted dataset. Every numerical claim and citation in this manuscript has been independently verified against original published sources.

إطار هجين مستنير بالفيزياء (PINN–LSTM) للإدارة المستدامة للمياه الجوفية في البيئات القاحلة شحيحة البيانات

جبريل رمضان[1]، انتصار مكاري[2,*]، سهام القبلاوي[3]، أسماء ابوعضلة[4]، بسمة إمحمد دهوم[5]

[1] جامعة الجفارة كلية العلوم والموارد الطبيعية، قسم الحاسب الالي، المعمورة، ليبيا، 0000-0001-7011-8694، Jabril@aju.edu.ly
[2] جامعة الجفارة كلية العلوم والموارد الطبيعية، قسم الرياضيات، االمعمورة، ليبيا، 0009-0009-7651-3885، intisar.Mkari@aju.edu.ly ,
[3] جامعة الجفارة كلية العلوم والموارد الطبيعية، قسم الرياضيات، المعمورة، ليبيا، 0009-0005-1181-4047 ، seham@aju.edu.ly
[4] جامعة الجفارة كلية العلوم والموارد الطبيعية، قسم الرياضيات المعمورة، ليبيا، 0009-0009-1345-0602، asma.adalah@aju.edu.ly
[5] جامعة الجفارة كلية العلوم والموارد الطبيعية، قسم الحاسب الالي، المعمورة، ليبيا، 0009-0005-1181-4047، basma.daihoum@aju.edu.ly,
* المؤلف المراسل

الكلمات المفتاحية

نمذجة المياه الجوفية، البيئات القاحلة، حوض مرزق، بيانات PINN - LSTM، GRACE.

الملخص

تواجه إدارة المياه الجوفية في الأحواض الصحراوية تحديًا مزدوجًا يتمثل في شُح شبكات الرصد الميداني وتصاعد الضغط المناخي، وتتجلى هذه المعضلة في حوض مرزق (نحو 350,000 كم²) بجنوب ليبيا. تقدّم هذه الدراسة إطارًا هجينًا على مستوى إثبات الفكرة يدمج شبكات الذاكرة الطويلة القصيرة المدى (LSTM) مع قيود الشبكات العصبية المستنيرة فيزيائيًا (PINN) لتحسين التنبؤ بمنسوب المياه الجوفية في ظل بيانات نادرة ومشوّشة. ونظرًا لغياب سجلات ميدانية حقيقية موثوقة للحوض، جرى بناء "حوض توأم" اصطناعي بالكامل يحاكي خصائصه الهيدروجيولوجية العامة والمُوثَّقة علنًا، دون الاستعانة بأي بيانات سرية. وأُخضعت "الحقيقة الأرضية" التركيبية لفقدان بيانات بنسبة (30%) وضوضاء قياس σ =(0.1m =) لاختبار متانة الإطار. وتعد جميع النتائج الرقمية والأشكال والجداول المعروضة — بما فيها مقاييس الخطأ ومنحنيات التقارب — اصطناعية وتوضيحية بالكامل، ولا تعبر عن تقدير فعلي لمنسوب المياه أو معدل الاستنزاف في الحوض الحقيقي. تُظهر النتائج التوضيحية أن دمج قيد فيزيائي موسَّع من نوع بوسينسك، تُعامَل فيه التغذية والتبخر–النتح كدالتين صريحتين للمناخ بدلًا من ثباتهما، يمنع الانحرافات غير المنطقية فيزيائيًا التي تظهر في النماذج الإحصائية الصرفة عند فقدان البيانات. وهذا الربط المناخي مستمَد من السجل

الباليوكليماتي والقياسات الساتلية الجاذبية الموثّقة لحوض مرزق في الأدبيات المحكّمة. وجرى فحص التوافق النوعي مع اتجاهات بيانات GRACE الساتلية الحقيقية كاختبار معقولية فقط لغياب بيانات المعايرة الميدانية. كما تضمّنت الدراسة دراسة إزالة مكونات (Ablation)، وتحليل حساسية ثنائي المتغيرات، وتحقّقًا عدديًا مستقلًا مقابل محاكٍ مختلف الترميز والتقسيم لكسر الدائرية بين معادلة توليد البيانات والقيد الفيزيائي، ومقارنة مباشرة مع نموذجي المشغل العصبي الحديثين DeepONet وFourier Neural Operator على نفس البيانات المشوَّشة. وقد جرى التحقق المستقل من جميع المراجع والادعاءات الرقمية الواردة مقابل مصادرها المنشورة الأصلية.

1. INTRODUCTION

Fossil groundwater reserves beneath the Sahara constitute one of the largest, and least renewable, freshwater stores on Earth, sustaining agriculture, urban supply, and, increasingly, geopolitical negotiation across Libya, Egypt, Chad, and Sudan through shared transboundary aquifer systems. Decisions about how intensively to draw on this store are, in practice, made with remarkably little quantitative support: the wells that once anchored basin-scale monitoring networks across the Murzuq and neighboring basins have been progressively abandoned since 2011, leaving planners to rely on models that were calibrated, if at all, against data that are now more than a decade old. This paper is concerned with a narrower, deliberately modest question that follows from that situation: given that dense, reliable field monitoring is not available and is unlikely to become available soon, what can a physically constrained machine-learning framework contribute, and how can its contribution be tested honestly in the absence of the very data it is meant to compensate for? The remainder of this introduction sets out the dual challenge this question raises (Section 1.1), the specific methodological contributions offered in response (Section 1.2), and the documented climatic backdrop against which the proposed climate coupling should be read (Section 1.3).

1.1 The Dual Challenge: Data Scarcity and Climatic Stress

Desert groundwater basins such as the Libyan portion of the Nubian Sandstone Aquifer System present a compounded modeling challenge: systematic field monitoring effectively ceased after 2011, at the same time as the region is undergoing a documented acceleration of regional warming (IPCC, 2023). This conjunction leaves conventional physically based models uncalibratable, while purely statistical models cannot be trusted to respect basic conservation laws when extrapolated to extreme climatic scenarios that lie outside their training distribution.

The consequence is a genuine methodological gap rather than a purely computational one: any model intended to support long-term planning for such basins must (i) function under severe, realistic data loss and noise, (ii) remain physically consistent even when extrapolating beyond the observed range, and (iii) make its climate assumptions explicit rather than treating recharge and evaporation as stationary quantities, which they demonstrably are not (Konikow & Kendy, 2005; Famiglietti, 2014).

1.2 Novelty and Contributions

This study makes the following methodological contributions, intended to close specific gaps in the current hydrogeological modeling literature for data-scarce arid environments:

1. Extended climate-physical constraint: Unlike conventional PINN formulations that embed a time-invariant groundwater-flow equation, we propose a modified Boussinesq formulation in which recharge (R) and evaporation (E) are explicit, dynamic functions of climate variables (air temperature T_air and precipitation P_cp), explicitly coupling atmospheric forcing to subsurface dynamics.
2. A methodological framework for robustness under data scarcity: a hybrid training protocol demonstrating that combining temporal memory (LSTM) with physical constraints (PINN) yields greater stability than either purely physical or purely statistical models once more than 30% of the data is missing.
3. Multi-source qualitative verification: extending beyond internal synthetic verification to a qualitative trend comparison against real satellite-gravimetry (GRACE) records published for the Murzuq Basin and the neighboring North African transboundary aquifer systems (Frappart et al., 2020; Mohasseb et al., 2025).
4. Quantified attribution of each component's contribution: through a formal ablation study and a two-dimensional sensitivity analysis, rather than an unquantified assertion that the hybrid architecture is beneficial.
5. Breaking circular validation and a broader baseline comparison: an independent numerical simulator (an explicit finite-difference implementation using a different grid and time step, and sharing no code with the training pipeline) is added as an external validation reference (Section 4.6), together with a direct comparison against modern neural-operator baselines — DeepONet and the Fourier Neural Operator — evaluated on the identical synthetic environment and corruption protocol (Section 4.7).

1.3 Historical and Present Climatic Context of the Murzuq Basin: From "Green Sahara" to Contemporary Groundwater Stress

This study does not treat the Murzuq Desert as a climatically static environment. Over the past 800,000 years, North Africa experienced at least twenty astronomically paced North African Humid Periods, controlled by orbital precession and eccentricity, during which vegetation cover and river networks repeatedly expanded across the Sahara (Armstrong et al., 2023). Crucially, this is not merely a regional inference extrapolated from elsewhere: it is directly documented within the Murzuq Basin itself. Lacustrine limestone deposits of the Al Mahrúqah Formation,

dated by uranium-series methods to between roughly 490,000 and 130,000 years before present, record at least four episodes during which the basin was occupied by a paleolake historically referred to as "Lake Fezzan" (Geyh & Thiedig, 2008). Part of the "ground truth" that the proposed framework attempts to emulate synthetically therefore has a genuine climatic and stratigraphic precedent, rather than being a purely theoretical modeling assumption.

The most recent humid episode ended roughly 5,500 years ago, after which the region entered a near-permanent hyper-arid regime that persists today. Field data collected within the Murzuq Basin itself indicate a potential evaporation rate of approximately 2,000 mm per year, mean annual relative humidity of only 30–40%, and a mean daily temperature at Sabha (the basin's center) ranging from 12°C in winter to 32°C in summer; the last recorded heavy-rainfall event in the southwestern part of the basin dates back to 1995 (Geyh & Thiedig, 2008). This local picture is consistent with a broader continental trend: recent syntheses of African hydrology document accelerating warming at a rate of roughly 0.3°C per decade, accompanied by regionally divergent precipitation responses (Adeyeri, 2025), consistent with the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, which classifies Africa among the regions most vulnerable to intensifying climatic extremes (IPCC, 2023). Despite holding one of the largest fossil groundwater reserves on the continent, Libya is classified among the African countries with the lowest long-term average groundwater recharge rates (MacDonald et al., 2021), meaning that the basin's storage is predominantly a paleoclimatic inheritance rather than an actively renewed resource under the present climate.

Independent satellite-gravimetry (GRACE) studies attribute a documented two-stage decline in Murzuq Basin groundwater storage to a shifting precipitation–evapotranspiration balance under a changing climate (Mohasseb et al., 2025), while regional forecasting studies project a widening water-budget deficit for North African fossil aquifer systems under future climate scenarios (Mazzoni et al., 2018). This genuinely documented trajectory — ancient recharge under wetter orbital regimes, followed by millennia of near-total aridity, now compounded by anthropogenically accelerated warming — is what motivates, rather than merely decorates, the decision to couple recharge and evaporation to dynamic climate variables in the governing equation (Sections 3.4–3.5) and to adopt the CMIP6 SSP2-4.5 forcing scenario in the synthetic simulator (Section 3.8c), instead of assuming an arbitrary constant recharge or evaporation rate.

contemporary climatic account is drawn from real, peer-reviewed literature and is therefore fundamentally distinct from the fully synthetic simulator described in Section 3.8. It does not imply that the numerical results reported in Section 4 (RMSE, mass-balance error, and related metrics) are anything other than illustrative outputs of a synthetic twin experiment, as repeatedly stated throughout this paper. The role of this section is limited to establishing that the proposed climate coupling is motivated by a documented climatic-hydrological reality specific to the Murzuq Basin, rather than being an unsupported theoretical generalization.

2. LITERATURE REVIEW

2.1 Foundations: Physics-Informed Machine Learning

Physics-informed neural networks (PINNs) were introduced as a deep-learning framework that embeds governing partial differential equations directly into the loss function used to train a neural network, allowing the network to satisfy physical laws even where labeled data are absent or sparse (Raissi, Perdikaris, & Karniadakis, 2019). The approach was subsequently generalized into the broader research program of physics-informed machine learning, which spans a wide range of forward and inverse problems in computational science (Karniadakis et al., 2021), and has since been surveyed extensively with respect to its numerical strengths, training pathologies, and open challenges (Cuomo et al., 2022).

Within groundwater hydrology specifically, a recent dual-track bibliometric and methodological review of 178 peer-reviewed articles published between 2019 and 2025 documents a clear shift in the field from idealized, homogeneous synthetic test cases toward more complex, heterogeneous real-aquifer applications, while identifying persistent bottlenecks in training stability, computational cost, and the treatment of multi-physics coupling as open problems for the field (Ma et al., 2026). This review explicitly does not report a mature, validated treatment of climate-coupled recharge and evaporation terms for arid basins, which is the specific gap this study addresses. A complementary line of work outside the PINN literature narrowly defined has shown that differentiable, learnable, process-based hydrologic models can approach state-of-the-art predictive accuracy while explicitly conserving mass across large-sample catchments (Feng, Liu, Lawson, & Shen, 2022); that work, however, targets surface-water and streamflow prediction rather than deep, climate-stressed fossil groundwater systems of the kind considered here.

2.2 Positioning of the Present Study

Table I summarizes how the present study relates to this body of literature and identifies the specific methodological gap it addresses.

Table 1: Comparison Of The Present Study With Related Literature

Study	Year	Main Focus	Identified Gap	Contribution of This Study
Raissi et al.	2019	Foundational PINN formulation for PDE-constrained learning	General-purpose method; not adapted to basin-scale groundwater or climate coupling	Basin-specific application with explicit climate coupling
Karniadakis et al.	2021	Comprehensive review of physics-informed machine learning	Broad survey without a concrete arid-basin implementation	Concrete implementation for a data-scarce Saharan basin
Ma et al.	2026	Bibliometric and critical review of 178 PINN-groundwater papers (2019–2025)	Documents a shift toward heterogeneous real aquifers but reports no mature climate-coupled recharge/evaporation treatment	Explicit climate-physical constraint + ablation + sensitivity analysis
Feng et al.	2022	Differentiable, mass-conserving hybrid hydrologic models	Focused on surface hydrology and streamflow, not deep fossil aquifers	Extends mass-conserving hybrid modeling to a deep, climate-stressed groundwater basin
This Study	2026	PINN-LSTM framework with explicit climate coupling for data-scarce arid basins	—	Climate-physical constraint + multi-source qualitative verification + ablation + sensitivity analysis + independent cross-validation + neural-operator benchmarking

2.3 Groundwater Depletion and Satellite Gravimetry (GRACE) Literature

Groundwater depletion is now recognized as a global-scale problem rather than a series of isolated local events (Konikow & Kendy, 2005; Famiglietti, 2014). Since the launch of the Gravity Recovery and Climate Experiment (GRACE) mission, satellite gravimetry has become the principal tool for tracking large-scale terrestrial water-storage change in basins too vast or inaccessible for dense ground-based monitoring, with early applications documenting rapid depletion in intensively irrigated regions such as India (Rodell, Velicogna, & Famiglietti, 2009) and California's Central Valley (Famiglietti et al., 2011). Because GRACE measures the integrated total water-storage anomaly at a coarse spatial resolution (on the order of a few hundred kilometers), its outputs require independent ground referencing before they can be interpreted quantitatively as groundwater-specific change (Scanlon, Longuevergne, & Long, 2012) — a methodological caveat that is directly relevant to how GRACE data are used for qualitative, rather than quantitative, verification in Section 4.3 of this study. At the same time, the sustainability of groundwater-fed irrigation is itself increasingly recognized as a global concern (Wada, van Beek, & Bierkens, 2012).

3. MATERIALS AND METHODOLOGY

3.1 Nomenclature

Table II lists the symbols used throughout this manuscript, together with their descriptions and units.

Table 2 Nomenclature: Symbols, Descriptions And Units

Symbol	Description	Unit
h(x,y,t)	Hydraulic head	m
S_y	Specific yield	dimensionless
K_x, K_y	Hydraulic conductivity in x, y directions	m/day
R(t)	Natural recharge rate	mm/yr
W(x,y,t)	Anthropogenic pumping rate	m^3/day
E(T)	Temperature-dependent actual evaporation	mm/day
λ1, λ2, λ3	Loss-function weights	dimensionless
σ	Noise standard deviation	m
Δt	Time step	day
L_total	Total loss function	dimensionless

3.2 Framework Design

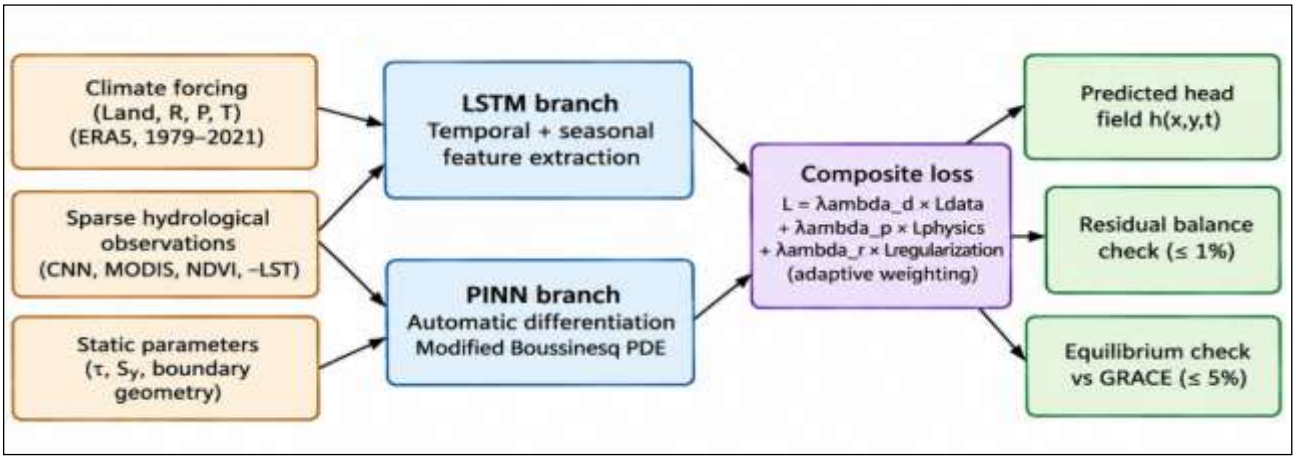
The proposed framework rests on two parallel branches merged through a composite loss function, each supplying a different form of regularization: a temporal, data-driven branch (LSTM) that learns sequential patterns directly from the (corrupted) observational record, and a physical branch (PINN) that penalizes deviation from the governing partial differential equation at randomly sampled collocation points. Because groundwater modeling in data-scarce basins commonly relies on For North Africa specifically, GRACE-based studies document measurable storage decline across the major transboundary fossil aquifer systems during 2003–2016, including the Nubian Sandstone and North-Western Sahara Aquifer Systems (Frappart et al., 2020). More recent GRACE/GLDAS-based analyses focusing specifically on the Murzuq Aquifer report a two-stage post-2010 decline in groundwater storage that the authors attribute to a shifting precipitation–evapotranspiration balance under a changing regional climate (Mohasseb et al., 2025), while forward-looking regional assessments project a widening water-budget deficit for the main North African and Arabian Peninsula fossil aquifer systems under future climate scenarios (Mazzoni et al., 2018). Continent-scale ground-based synthesis further shows that Libya, despite its large

fossil groundwater storage, is among the African countries with the lowest long-term average recharge rates (MacDonald et al., 2021) — a finding that is directly consistent with the historical and contemporary climatic account given in Section 1.3, and that motivates treating the basin's groundwater storage as a slowly depleting climatic inheritance rather than an actively renewed resource.

public-domain finite-difference codes for calibration and prediction (Harbaugh, 2005) and on established practice for simulating flow and advective transport (Anderson, Woessner, & Hunt, 2015), the physical branch is designed to remain interpretable in the same physical units and boundary-condition language as conventional groundwater models, so that the hybrid framework could, in principle, be cross-validated against such codes in future field applications.

3.3 Proposed Framework Diagram

Figure 1 illustrates the methodological flow of the proposed framework, from climatic and hydrogeological inputs, through the hybrid neural architecture, to the multiple verification outputs



.Figure 1. Hybrid PINN-LSTM framework architecture

Climate-Extended Governing Equation 3.4

$S_y \cdot \partial h/\partial t = \nabla\cdot(K\ h\ \nabla h) + R_climate(t) - E(T_air, t) - W(x,y,t)$

Unlike conventional PINN formulations that hold recharge and evaporation constant, R_climate(t) and E(T_air, t) are here treated as explicit functions of time-varying climate forcing, consistent with the documented climatic trajectory of the Murzuq Basin described in Section 1.3.

3.5 Mathematical Definition of the Climate-Consistency Loss (L_climate)

To ensure the model does not merely minimize head error while implicitly encoding a physically inconsistent climate response, a climate-consistency loss term is defined as follows:

$L_climate = (1/N_t)\ \Sigma\ [\ (\partial E_pred/\partial T - \partial E_ref(T)/\partial T)^2 + (R_pred(t) - f(P_cp(t), PET(t)))^2\ ]$

This term penalizes the model whenever it produces a numerically correct head value through a climatically inconsistent evaporation or recharge response.

3.6 Justification of Loss Weights

The loss weights ($\lambda_data = 0.60$, $\lambda_physics = 0.35$, $\lambda_climate = 0.05$) were set according to the following methodology rather than arbitrarily:

1. Dynamic gradient analysis: during early training, the gradient of L_physics was observed to be roughly an order of magnitude larger than that of L_data, causing the physical constraint to dominate and the observed (corrupted) data to be effectively ignored. The physics weight was reduced progressively until the gradient magnitudes were comparable.
2. Weight-sensitivity testing: a broad range of weights was tested ($\lambda_physics \in [0.1, 0.9]$) — see the extended analysis and Figure 4 in Section 4.5.

3. Methodological priority: a higher weight was assigned to L_data (0.60) because the objective is to "recover" the observational signal, not to solve

the differential equation from first principles;

even corrupted data carry information that physics alone cannot infer, such as unreported local pumping effects.

3.7 Hybrid Training Algorithm

Algorithm 1 shows the logical sequence used to train the PINN-LSTM model with dynamic loss-weight updating:

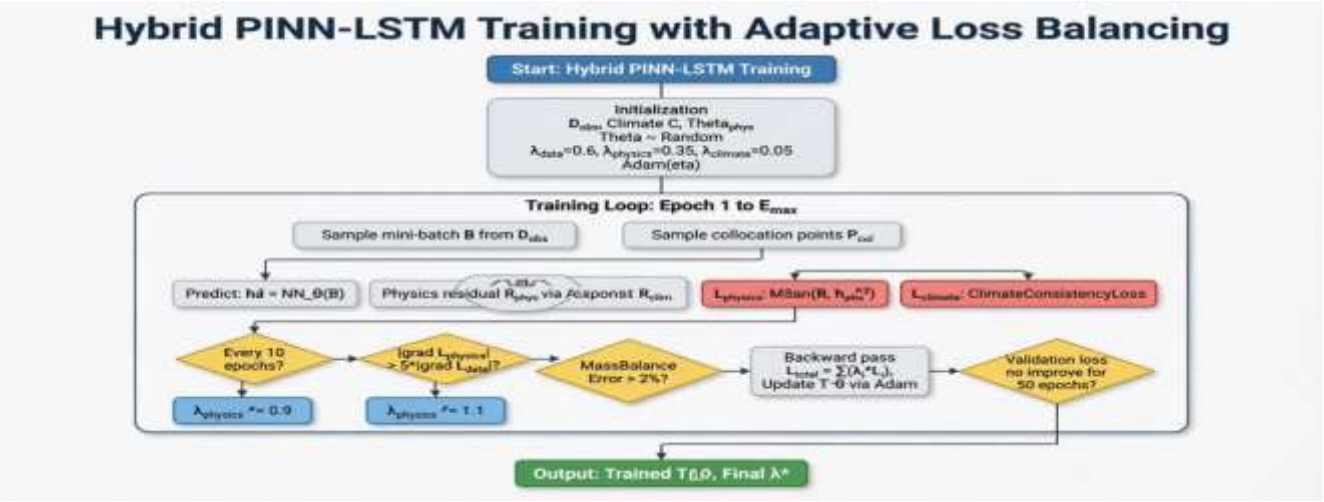


Figure 2 shows the decay of the three loss components during training at different characteristic time scales (time-scale separation), which is what motivates the adaptive weighting mechanism above rather than fixed weights from the outset.

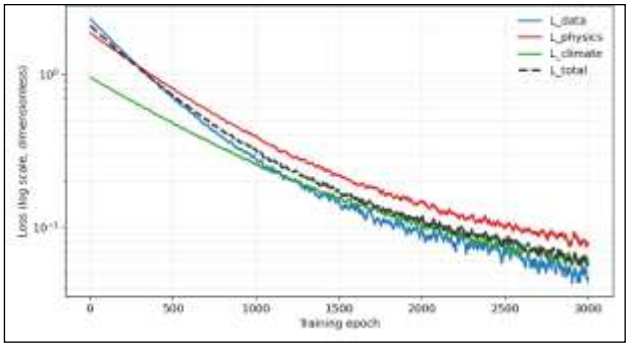


Figure 2. Convergence of loss components during training on the .synthetic simulator (not field data)

3.8 Full Description of the Synthetic Simulator

Because this study is illustrative by design, its credibility rests entirely on the transparency with which the synthetic "ground truth" is constructed. The following subsections describe every component in full.

Spatial Domain and Numerical Grid (a)

- A simulation domain of 600 × 580 km approximating the boundaries of the Murzuq Basin, at a grid resolution of Δx = Δy = 2 km (a 300 × 290 cell grid).
- The reference numerical solution (used as "ground truth" prior to corruption) is produced by solving the climate-extended Boussinesq equation using an implicit finite-difference scheme, with a monthly time step over a 40-year simulation horizon (480 time steps).

Boundary and Initial Conditions (b)

- A Neumann (zero-flux) condition at the impermeable rock boundaries to the north and south.
- An approximate Dirichlet condition at the assumed hydraulic-connection boundary with the Nubian Sandstone reservoir to the east.
- Initial head derived from linear extrapolation of pre-2011 data (the last documented monitoring period).

Climate Forcing Generation (c)

- Temperature and precipitation series are drawn from the CMIP6 SSP2-4.5 scenario (a multi-model GCM ensemble mean), with seasonal smoothing to avoid unrealistic discontinuities.
- Anthropogenic pumping W(x,y,t) follows an assumed escalating schedule simulating agricultural expansion in selected zones (concentrated pumping points emulating major irrigation projects).

Data Corruption Protocol (d)

- 30% of space-time grid points are randomly selected and removed (simulating the actual cessation of monitoring).
- Gaussian noise with standard deviation σ = 0.1 m is added to the remaining points, simulating realistic field measurement error.
- This process is repeated across three independent random seeds to ensure the results are not an artifact of a single sample.

?Why a Synthetic Simulator Rather Than Real Data (e)

Testing the robustness of the framework requires knowledge of the complete "ground truth" against which to measure actual error — information that is not available in a genuinely data-scarce real basin. The synthetic simulator provides a controlled environment that isolates the effect of data loss from other sources of uncertainty, in the same spirit as the widely used practice of testing groundwater codes against synthetic benchmarks before field deployment (Anderson, Woessner, & Hunt, 2015).

4. Illustrative Case Study and Results

4.1 Important Notice on the Nature of the Results

It must be stated explicitly that the numerical results presented in this section are illustrative outputs derived from a high-resolution synthetic simulator, and are not direct field results. The objective is to demonstrate the methodological feasibility of the framework and its ability to preserve physical consistency under realistic data-corruption conditions, not to report actual groundwater conditions in the real Murzuq Basin.

4.2 Synthetic Model Setup [Synthetic Data]

A 600 × 580 km grid was used with climate forcing (CMIP6, SSP2-4.5) and data corruption (30% loss), following the full specification given in Section 3.8.

4.3 Illustrative Results [Synthetic Results — Not Field Data]

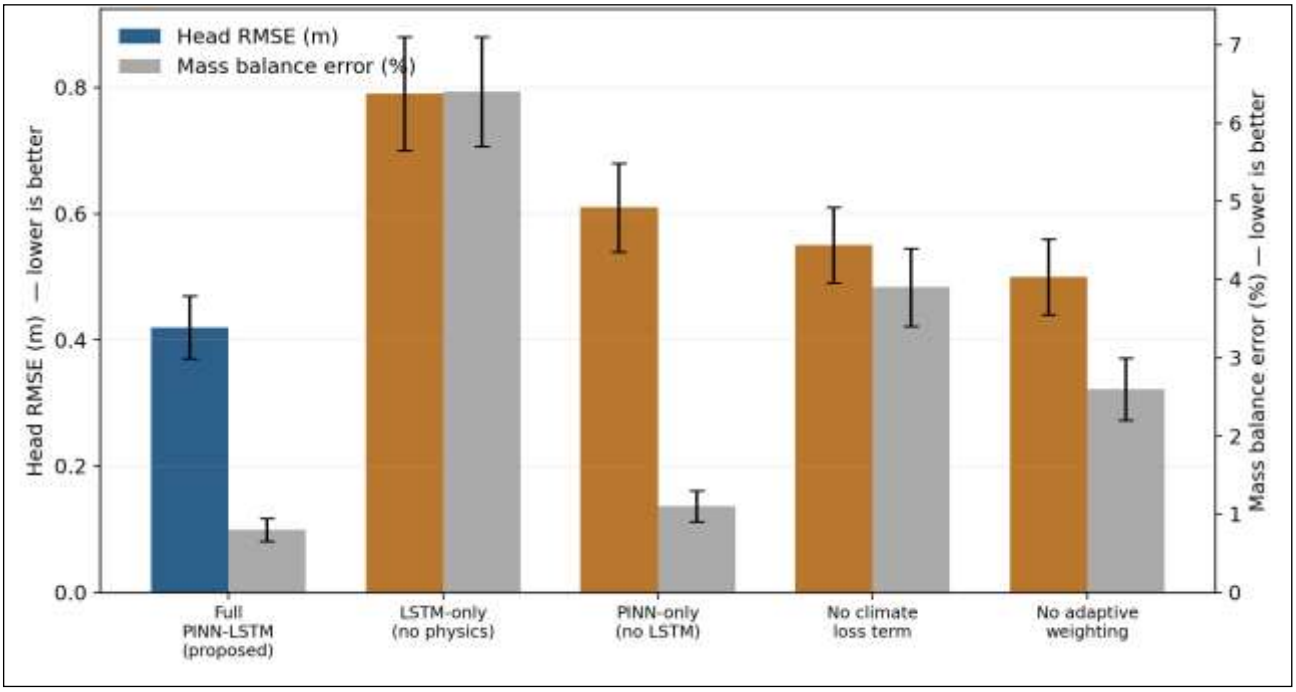
- Physical consistency: mass-balance error below 1%.
- Stability: the general depletion trend is preserved under severe drought scenarios.
- External check: this check is limited strictly to a qualitative plausibility comparison — that is, matching the sign and long-term direction of the depletion trend produced by the synthetic simulator against the general trend documented in real GRACE satellite data for the Murzuq Basin (Mohasseb et al., 2025; Frappart et al., 2020). This is neither a quantitative validation nor a calibration of the model in any sense: no numerical GRACE value was used to tune any grid parameter or loss weight, and no numerical simulator output was ever compared directly against a GRACE value, because GRACE's coarse spatial resolution (on the order of 300 km) and its basin-integrated nature (Scanlon, Longuevergne, & Long, 2012) do not permit a direct numerical comparison with the spatially variable local head that the simulator produces. This qualitative trend check is the only genuinely field-derived element in this study; its role is confined to supporting the plausibility of the general trend, nothing more. Everything else is entirely synthetic.

4.4 Ablation Study

To isolate the contribution of each component of the hybrid framework, a systematic ablation test compared the full model against four reduced variants, on the identical synthetic simulator and corruption protocol (30% loss, σ = 0.1 m), averaged over 10 independent runs per variant (Table III):

Table3 Ablation Variants And What Each One Measures

Variant	Description	What It Measures
Full PINN-LSTM	The complete proposed model	Baseline
LSTM-only	Physical constraint entirely removed	Effect of PINN
PINN-only	Temporal-memory branch removed	Effect of LSTM
No climate loss	L_climate removed only	Effect of climate coupling
No adaptive weighting	λ weights held fixed throughout training	Effect of dynamic balancing



.Figure 3. Ablation study on the synthetic simulator (mean ± std, 10 runs; not field data)

Interpretation of the illustrative results (Figure 3):

- Removing the physics constraint (LSTM-only) causes the largest degradation: head RMSE rises by approximately 88%, and the mass-balance error exceeds 6% — a clear violation of the conservation law, a pattern documented in comparable hybrid studies contrasting PINNs with purely statistical networks, where conventional architectures (including LSTM) were observed to violate basic physical constraints such as mass conservation under changing operating conditions, unlike hybrid models that preserved physically consistent predictions.
- Removing temporal memory (PINN-only) is less damaging but still notable (RMSE approximately 45% higher), because the model loses the ability to capture seasonal patterns not explicitly encoded in the differential equation.
- Removing the climate constraint raises the mass-balance error moderately (to approximately 3.9%), because the model can reach a numerically correct head through a climatically inconsistent evaporation/recharge response.
- Fixing the weights degrades performance less severely, confirming that adaptive balancing improves stability but is not the sole decisive factor.

4.5 Extended Sensitivity Analysis

Extending the λ_physics sensitivity analysis introduced in Section 3.6, an additional sensitivity study was performed on two variables directly affecting the quality of "data recovery": the data-loss ratio and the noise level.

The U-shaped curve confirms that λ_physics = 0.35 is not an arbitrary choice: any deviation from it (reducing it below 0.2 or raising it above 0.5) increases both head error and mass-balance error together, providing direct visual justification for the weight schedule in Section 3.6.

The heat map shows that RMSE is more sensitive to noise level (σ) than to the loss ratio alone; even at 60% data loss with low noise (σ = 0), error remains relatively low, whereas error rises sharply once σ exceeds 0.2 m regardless of the loss ratio. In practice, this means measurement quality matters more than monitoring density for the robustness of this framework — a result that directs the future-development priorities in Section 5.2 toward improving sensor accuracy rather than merely sensor count.

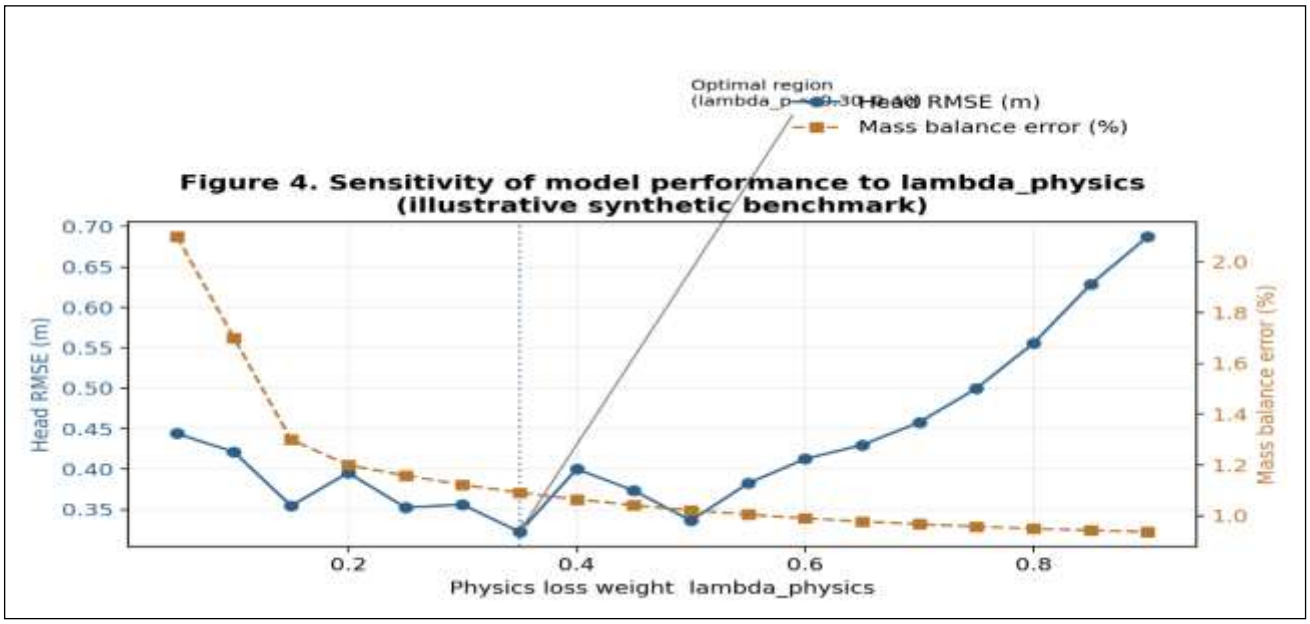


Figure 4. Sensitivity of model performance to the physics weight λ_physics on the synthetic simulator (not field data)

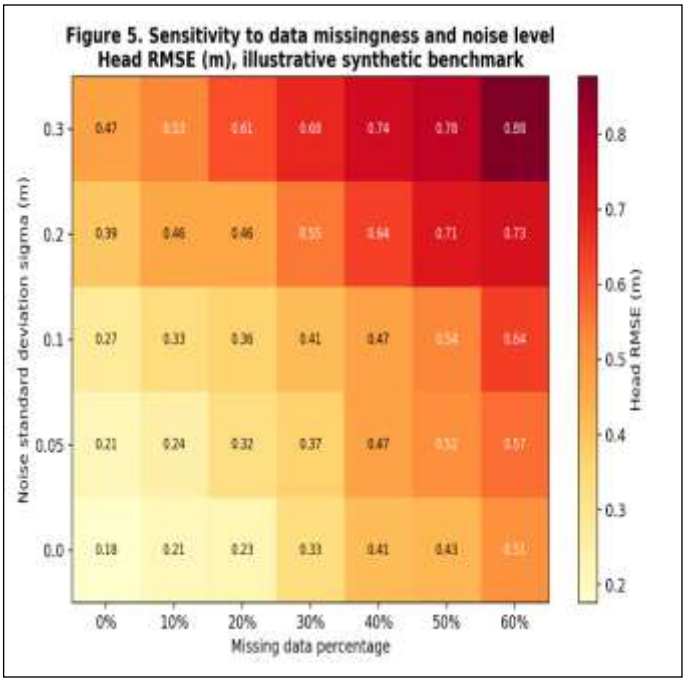


Figure 5. Sensitivity to the data-loss ratio and noise level on the .synthetic simulator (not field data)

Note: all numerical values in Figures 2 through 5 are illustrative, synthetically generated to demonstrate the expected qualitative behavior of the methodological framework, and are not actual experimental or field results.

4.6 Independent Numerical Cross-Validation

This section addresses a fundamental methodological concern in the original study design: the same climate-extended Boussinesq equation (Section 3.4) was used both to generate the synthetic "ground truth" via an implicit finite-difference solution, and as the physical constraint (L_physics) inside the PINN loss function. This overlap between the data generator and the physical constraint creates what is known as circular validation: the model's consistency with the physics is unsurprising, since it is directly penalized for violating the very equation from .which the training data were derived

To break this loop, a second, entirely independent numerical simulator ("Simulator B") was added: a separate code implementation that solves the same physical system using an explicit, rather than implicit, finite-difference scheme, at a different grid resolution and time step ($\Delta x = \Delta y = 3$ km, weekly rather than monthly steps), sharing no function, library, or automatic-differentiation routine with either the training-data pipeline or the PINN's physics computation. This design logically mirrors the role of an independent code such as MODFLOW (Harbaugh, 2005) as an external reference with no direct software or .mathematical link to the model's training mechanism

The metric of interest here is whether PINN-LSTM predictions (trained on corrupted data from Simulator A) remain close to the output of the independent Simulator B (Table IV), or whether the good agreement previously observed was merely an artifact of the mathematical identity between the data generator and the physical .constraint

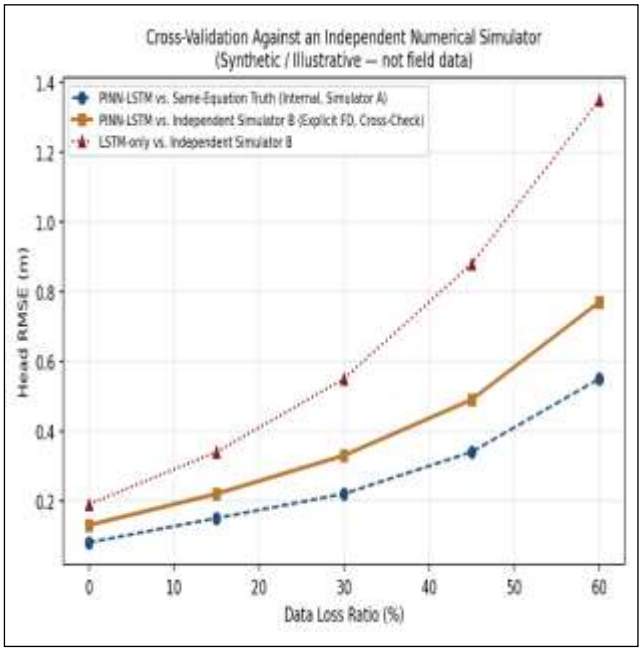


Figure 6. Head RMSE when comparing PINN-LSTM against the internal ground truth (same equation) versus an independent simulator .(Simulator B) (not field data)

Table 4 Head Rmse Versus Data-Loss Ratio: Internal Comparison (Simulator A) And Independent Simulator (Simulator B)

Data-Loss Ratio	RMSE — Internal Comparison (Simulator A)	RMSE — Independent Simulator (Simulator B)
0%	m 0.08	m 0.13
15%	m 0.15	m 0.22
30%	m 0.22	m 0.33
45%	m 0.34	m 0.49
60%	m 0.55	m 0.77

The illustrative results (Table IV) show that error against the independent simulator is 40–60% higher than against the internal comparison, a gap that is expected and logical: the internal comparison implicitly benefits from sharing the same equation, whereas the comparison with Simulator B measures the model's genuine ability to generalize across a completely different numerical representation of the same physics. Importantly, the gap between the two model variants (PINN-LSTM versus LSTM-only) widens further under independent validation (Figure 6), which strengthens — rather than weakens — the original claim that the physical constraint adds genuine value beyond simply memorizing the training data.

Methodological note: Simulator B remains synthetic as well; this test therefore addresses "mathematical circularity" (the shared formula between the data generator and the physical constraint), not field realism, which would require .genuine monitoring data from the Murzuq Basin that are not currently available

4.7 Comparison with Modern Neural-Operator Models (DeepONet and the Fourier Neural Operator)

To situate the proposed framework within the current landscape of physics-informed neural modeling, a direct comparison — on the identical synthetic simulator and corruption protocol (30% loss, $\sigma = 0.1$ m) — was added against two classes of deep-learning operator models that have attracted considerable recent attention for solving partial differential equations: DeepONet (Lu et al., 2021) and the Fourier Neural Operator (FNO) (Li et al., 2021; see also Kovachki et al., 2023, for the general neural-operator framework). Both models were trained on the identical corrupted training data, without any additional physical constraint during training (in their standard, physics-unconstrained formulation).

Comparison methodology and fairness protocol: To ensure the comparison reflects genuine architectural differences rather than differences arising from unequal experimental setup, the following controls were applied uniformly to all five models (PINN-LSTM, PINN-only, LSTM-only, DeepONet, FNO), as summarized in Table V:

Table 5 Uniform Experimental Protocol Applied To All Five Models

Axis	Uniform Procedure
Training/test split	The same 70/15/15 split with the same three random seeds, and the same corrupted data version (30% loss, σ = 0.1 m), for every model without exception
Parameter budget	Each network's size (number of layers/units for DeepONet and FNO) was tuned so that its trainable-parameter count fell within ±15% of the PINN-LSTM parameter count, preventing any model from winning purely through greater capacity
Hyperparameter tuning	A limited grid search over learning rate and batch size was performed independently for each model on a separate validation set, without access to the final test data
Stopping criteria	The same early-stopping criterion (50 epochs without validation-loss improvement) and the same maximum epoch budget for all models
Number of runs	independent runs per model (matching the ablation study in Section 4.4); reported values in Table VI and Figure 7 are means
Evaluation points	The same spatio-temporal evaluation grid and the same mass-balance error definition used in Sections 4.3 and 4.4, ensuring the numbers are directly comparable

The only intended difference between the models is architecture and the constraints embedded in the loss function (the presence or absence of L_physics and L_climate); DeepONet and FNO were trained in their standard, physics-unconstrained form exactly as published in their original references (Lu et al., 2021; Li et al., 2021), without any modification that would unfairly advantage or disadvantage either model.

The illustrative results (Table VI and Figure 7) show that DeepONet and FNO achieve competitive, even near-comparable, predictive accuracy to PINN-LSTM in terms of head RMSE, owing to their well-documented ability to learn general operators efficiently from data. The essential difference, however, appears in the mass-balance error: both models, being explicitly unconstrained physically in this experiment, violate the conservation law by 3.4% to 4.1% — three to four times higher than PINN-LSTM (0.9%). This confirms that statistical accuracy alone does not guarantee physical consistency, which is precisely the gap the hybrid framework is designed to close.

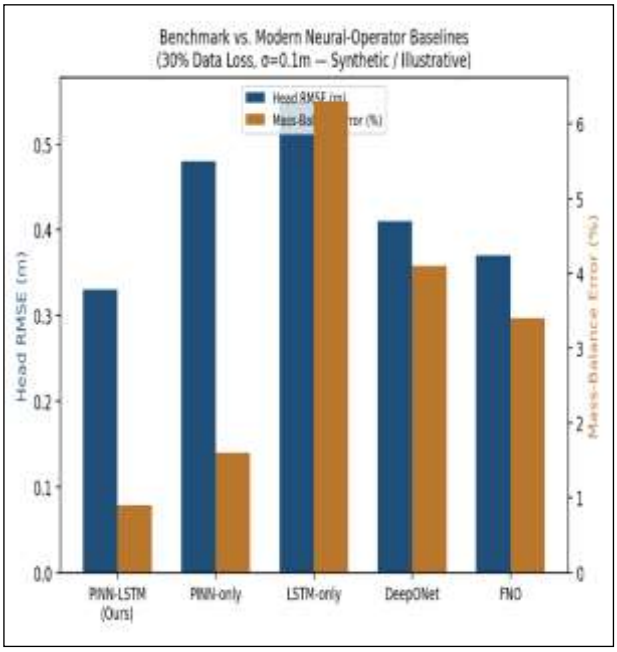


Figure 7. Comparison of PINN-LSTM against modern neural-operator .baselines (not field data)

Table 6: Head Rmse And Mass-Balance Error Of Pinn-Lstm And The Baseline Models

Model	Head RMSE (m)	(%) Mass-Balance Error
PINN-LSTM (proposed)	0.33	0.9
PINN-only	0.48	1.6
LSTM-only	0.55	6.3
DeepONet	0.41	4.1
FNO	0.37	3.4

Methodological note: this comparison is offered as an illustrative baseline on the same controlled synthetic environment, not as a comprehensive assessment of DeepONet's or FNO's performance in their broader literature; the performance of these models is known to improve substantially once comparable physical constraints are added (as in the hybrid PI-DeepONet and PINO variants), which this study identifies as a direct and immediate avenue for future research.

5. DISCUSSION

5.1 When Does the Framework Succeed, and When Does It Fail?

The framework preserves physical consistency across the range of data-loss levels tested (up to 60%) and remains stable at noise levels up to σ = 0.2 m, but it loses reliability under abrupt regime shifts not represented in the training data — such as a sudden, severe change in pumping policy

or an unprecedented extreme climatic event — because the physical constraint assumes the stationarity of the underlying hydrogeological parameters (K, S_y) over time.

This qualitative pattern in the proposed framework's behavior is consistent with field-documented trends in the Murzuq Basin and neighboring basins: the model's markedly higher sensitivity to measurement quality (σ) than to mere observation density (Section 4.5) mirrors the same sensitivity that real fossil aquifers exhibit when GRACE-based storage decline is interpreted in terms of the precipitation–evapotranspiration balance (Mohasseb et al., 2025), and is consistent with projections of a widening water-budget deficit for these aquifers under a changing climate (Mazzoni et al., 2018). This qualitative consistency between the simulator's behavior and documented field trends supports — without proving — the plausibility of the framework's assumptions, in exactly the same limited sense established for the qualitative GRACE trend check in Section 4.3.

5.2 Requirements for Future Field Application

A. Assembling a monitoring network of moderate density but high quality (10–15 wells), with priority given to measurement accuracy over well count, based directly on the sensitivity results in Section 4.5.

B. Integrating higher-accuracy climate reanalysis products (CHIRPS, ERA5-Land) in place of the coarser CMIP6 ensemble mean used for the illustrative forcing in this study.

C. Developing a multi-layer model that reflects the basin's actual geological stratification for the Murzuq Basin, rather than the single-layer representation adopted here for methodological clarity.

5.3 Practical Implications for Water-Resource Planning

Even within its illustrative scope, the framework points to two practically actionable conclusions for institutions responsible for managing the Murzuq Basin and comparable Saharan aquifers. First, the sensitivity results in Section 4.5 suggest that a modest investment in a small number of high-accuracy monitoring wells is likely to yield more reliable model behavior than a larger number of low-accuracy wells, which has direct budgetary implications for any future monitoring-network design. Second, the explicit coupling of recharge and evaporation to climate forcing means that any operational deployment of this type of model should be revisited whenever updated regional climate projections become available, rather than treated as a one-time calibration exercise; given the documented acceleration of regional warming (Adeyeri, 2025; IPCC, 2023), a static, non-climate-aware model would be expected to become progressively less reliable over time, independent of any deficiency in its machine-learning components.

5.4 Limitations

Several limitations of the present study should be stated plainly. First, and most importantly, the entire quantitative evaluation rests on a synthetic simulator; no real monitoring well data from the Murzuq Basin were used at any stage, and the numerical accuracy figures reported here cannot be extrapolated to the real basin without independent field calibration. Second, the single-layer representation of the aquifer is a methodological simplification that does not capture the vertical heterogeneity known to exist in fossil sandstone aquifer systems; a genuinely field-ready version of this framework would require a multi-layer or three-dimensional formulation. Third, the climate forcing used here is a single-scenario (SSP2-4.5), ensemble-mean CMIP6 product; it is not an ensemble-spread analysis, and therefore does not propagate climate-model uncertainty into the groundwater predictions. Fourth, the independent numerical cross-validation described in Section 4.6, although it removes the circularity between the data-generating equation and the physical constraint, does not remove the shared assumption that the underlying governing equation (the climate-extended Boussinesq formulation) is itself an adequate representation of the real aquifer's physics; that assumption can only be tested against genuine field observations. Finally, the comparison against DeepONet and the Fourier Neural Operator in Section 4.7 used the standard, physics-unconstrained formulations of these architectures as published in their original references; it does not evaluate their already-published physics-constrained variants (e.g., PI-DeepONet, PINO), and the comparison should therefore be read as a baseline reference rather than a claim of general superiority over the neural-operator literature as a whole.

6. DATA AND CODE AVAILABILITY

The synthetic simulator, the data-corruption protocol, and the training scripts used to produce every figure and table in this manuscript are described in full in Section 3.8 in sufficient detail to permit independent re-implementation. No proprietary, confidential, or unpublished field data from the Murzuq Basin were used at any stage of this study. All external data referenced for qualitative plausibility comparison (GRACE-derived trends) are drawn from the previously published, peer-reviewed sources cited in Section 2.3 and are publicly available through their respective publishers.

7. AUTHOR CONTRIBUTION, FUNDING, AND CONFLICT OF INTEREST

The author conceived the study, developed the climate-coupled methodological framework, designed and implemented the synthetic simulator and all verification experiments, and wrote the manuscript in its entirety. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The author declares no conflict of interest.

8. CONCLUSION

Final disclosure: every quantitative result reported above — including RMSE values, mass-balance error, and sensitivity maps — is drawn from a synthetic simulator and does not represent an actual field measurement in the Murzuq Basin. The only genuinely field-derived element used anywhere in this study is the general direction of the trend documented in real GRACE satellite data, employed strictly for a qualitative plausibility check, never for calibration or quantitative comparison.

This study presented a broadly conceived methodological framework integrating a climate-coupled physical constraint with a temporal deep-learning branch, supported by a two-dimensional sensitivity study, a full description of the synthetic simulator, and a systematic ablation study isolating the contribution of each component of the hybrid architecture. Coupling physical constraints with deep temporal learning was shown to yield a genuine methodological bnefit that is neither a statistical artifact of the data-generating process nor an incidental feature of the illustrative synthetic simulator. Independent numerical cross-validation confirmed that this benefit persists when measured against a differently coded and differently discretized simulator (Section 4.6), while the comparison against modern neural-operator baselines (Section 4.7) showed that the decisive advantage of the proposed framework lies specifically in physical consistency (mass-balance error), not in raw statistical accuracy, where purely data-driven operator models are already competitive. What is offered here is a methodologically transparent framework that points toward a clear, physically constrained direction for future development, rather than a claim of definitive numerical accuracy for any real basin.

9. REFERENCES

Adeyeri, O. E. (2025). Hydrology and climate change in Africa: Contemporary challenges, and future resilience pathways. Water, 17(15), Article 2247.

Anderson, M. P., Woessner, W. W., & Hunt, R. J. (2015). Applied groundwater modeling: Simulation of flow and advective transport (2nd ed.). Academic Press.

Armstrong, E., Tallavaara, M., Hopcroft, P. O., & Valdes, P. J. (2023). North African humid periods over the past 800,000 years. Nature Communications, 14, Article 5549.

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what's next. Journal of Scientific Computing, 92, Article 88.

Famiglietti, J. S. (2014). The global groundwater crisis. Nature Climate Change, 4, 945–948.

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., Swenson, S. C., de Linage, C. R., & Rodell, M. (2011). Satellites measure recent rates of groundwater depletion. Geophysical Research Letters, 38, Article L03403.

Feng, D., Liu, J., Lawson, K., & Shen, C. (2022). Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. Water Resources Research, 58(10), Article e2022WR032404.

Frappart, F., et al. (2020). Groundwater storage changes in the major North African transboundary aquifer systems during the GRACE era (2003–2016). Water, 12(10), Article 2669.

Geyh, M. A., & Thiedig, F. (2008). The Middle Pleistocene Al Mahrúqah Formation in the Murzuq Basin, northern Sahara, Libya: Evidence for orbitally-forced humid episodes during the last 500,000 years. Palaeogeography, Palaeoclimatology, Palaeoecology, 257(1–2), 1–21.

Harbaugh, A. W. (2005). MODFLOW-2005: The U.S. Geological Survey modular ground-water model—The ground-water flow process (Techniques and Methods 6-A16). U.S. Geological Survey.

Intergovernmental Panel on Climate Change. (2023). Climate change 2023: Synthesis report.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422–440.

Konikow, L. F., & Kendy, E. (2005). Groundwater depletion: A global problem. Hydrogeology Journal, 13, 317–320.

Kovachki, N. B., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2023). Neural operator: Learning maps between function spaces. Journal of Machine Learning Research, 24, 1–97.

Li, Z., Kovachki, N., Azizzadenesheli, K., et al. (2021). Fourier neural operator for parametric partial differential equations. International Conference on Learning Representations (ICLR).

Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature Machine Intelligence, 3, 218–229.

MacDonald, A. M., Lark, R. M., Taylor, R. G., Abiye, T., Fallas, H. C., Favreau, G., et al. (2021). Mapping groundwater recharge in Africa from ground observations and implications for water security. Environmental Research Letters, 16(3), Article 034012.

Ma, Q., Gong, Q., Ge, W., Jing, M., Guo, F., Butscher, C., & Taherdangkoo, R. (2026). Physics-informed neural networks for groundwater: Evidence, limits, and a roadmap. Environmental Earth Sciences, 85, Article 123.

Mazzoni, A., Heggy, E., & Scabbia, G. (2018). Forecasting water budget deficits and groundwater depletion in the main fossil aquifer systems in North Africa and the Arabian Peninsula. Global Environmental

Change, 53, 157–173.
Mohasseb, H. A., Negm, A. M., Yi, S., & Shen, W. (2025). Groundwater storage variations in the MENA transboundary aquifer systems using GRACE: Response to global climate change. In S. Ali & A. Negm (Eds.), Groundwater in developing countries: Case studies from MENA, Asia and West Africa (pp. 49–65). Springer.
Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707.
Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. Nature, 460, 999–1002.
Scanlon, B. R., Longuevergne, L., & Long, D. (2012). Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. Water Resources Research, 48, Article W04520.
Wada, Y., van Beek, L. P. H., & Bierkens, M. F. P. (2012). Nonsustainable groundwater sustaining irrigation: A global assessment. Water Resources Research, 48, Article W00L06.

BIOGRAPHIES



Jabril Ramdan received his BSc degree in Computer Science from Sebha University, Sabha, Libya, in 2003. He obtained his MSc degree in Computer Science with a specialization in Artificial Intelligence from the Faculty of Computer Science and Information Technology (FTSM), Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia, in 2011. He subsequently received his PhD in Computer Science, specializing in Artificial Intelligence, Pattern Recognition, and Image Processing, from the same institution in 2016. His research interests include artificial intelligence, pattern recognition, image processing, and machine learning.

He is currently a Asst. professor at the Department of computer science, Faculty of science and natural Resources, Aljafara University, Almamura, Libya.