# The Risks of Imperfect Knowledge: Reliability Trade-Offs in Physics-Aware Neural Networks for Climate Projections

Mengxue Zhang[1], Kai-Hendrik Cohrs[1], Miguel-Ángel Fernández-Torres[2], Esther Rodrigo-Bonet[1], and Gustau Camps-Valls[1]

[1] Image Processing Laboratory (IPL), Universitat de València, Paterna, València, 46980, Spain
[2] Department of Signal Theory and Communications, Universidad Carlos III de Madrid (UC3M), Madrid, 28911, Spain

**Abstract.** Hybrid models combining physical knowledge with machine learning in Earth system science face two fundamental challenges: equifinality, where distinct parameterizations yield equivalent training performance, and uncertainty arising from internal climate and scenario variability. To investigate the relationship between these challenges, this study evaluates how physical correctness affects predictive stability by categorizing physics-aware neural networks for evapotranspiration estimation into under-constrained, imperfect, and perfect knowledge subtypes. These models are tested using IPSL-CM5A-LR climate data across multiple realizations under historical and RCP scenarios (2.6, 4.5, and 8.5). Results show that while physical constraints mitigate equifinality via model regularization, their effectiveness depends strictly on the accuracy of the embedded knowledge. Imperfect constraints degrade performance, particularly under high-emission scenarios, whereas low-forcing regimes amplify model sensitivity to internal variability. Consequently, physical constraints alone do not guarantee reliability in climate projections; robust uncertainty quantification and regularization regarding climate change remain essential for developing trustworthy neural Earth system models.

**Keywords:** Hybrid Model · Physical Constraint · Model Extrapolation · Evapotranspiration Estimation · Climate Forcing.

## 1 Introduction

The integration of process-based physical knowledge with data-driven machine learning has led to the emergence of hybrid modeling frameworks, which have recently gained significant attention in Earth system science [22,5]. In a typical hybrid framework, deep neural components are seamlessly coupled with physical equations to combine their respective advantages: model flexibility, physical interpretability, and consistency with scientific knowledge [10,24]. In particular, due to the physical structure, these models promise trustworthy extrapolation, an issue that plain data-driven models famously struggle with [25].

Nonetheless, hybrid modeling is not without limitations. A main hurdle to the adaptability and reliability of hybrid models for extrapolation is equifinality [3,13], where many combinations of physics and machine learning may lead to similar performance on the data in-distribution, but strongly differ in internal representations. The main reason for this phenomenon is that the latent components of the hybrid model may not be well-constrained, leading to similar untrustworthy extrapolations with large uncertainty even when the physics are correct. Identifying and addressing this issue is a key step toward developing trustworthy neural Earth system models [16,32].

A solution is the introduction of further prior knowledge on the neural networks within the hybrid modeling framework. Physics-aware (also known as knowledge-guided) neural networks, for instance, incorporate additional knowledge through varying strategies, such as pretraining, semi-parametric structures or physics-guided losses [6,12]. In Earth-system modeling, physics-aware neural networks (NNs) have been used for estimating evapotranspiration (ET) [9,27]. As a key component modulating Earth's energy, water, and carbon cycles, accurate ET estimation is critical for addressing global hydrological challenges under climate change [19]. Numerous applications have demonstrated the superiority of these physics-aware NNs over purely mechanistic or strictly statistical models [33,15,8].

One issue with the introduction of auxiliary information, such as data or equations, in the Earth system sciences is that data can be highly noisy, and physical equations are often only approximate representations of reality. Karpatne et al. [11] suggest classifying the introduced knowledge into *perfect* and *imperfect knowledge*, where mostly all equations of Earth system science fall into. This raises a relevant question: *is knowledge useful even if it is imperfect?*

In this study, we systematically explore the relationship between varying qualitative degrees of physics constraints in hybrid models and their extrapolation capacities in climate projections. We do this by categorizing physics-aware NNs into distinct subtypes based on the correctness of their embedded physical constraints: under-constrained, imperfect, and perfect.
Our contributions are as follows:

1. We establish a novel analytical framework that isolates the effects of varying degrees of physical correctness on predictive stability.
2. We demonstrate that while physical regularization mitigates equifinality, integrating imperfect physical knowledge can severely degrade future predictions, sometimes performing worse than under-constrained models.

3. We reveal that the effectiveness of physical constraints is highly sensitive to climate forcing: extreme thermodynamic shifts (e.g., RCP 8.5) amplify parametric errors, whereas low-forcing regimes (e.g., RCP 2.6) amplify the model's sensitivity to internal variability.

## 2 Related Work

*Equifinality and Physical Regularization* The issue of model equifinality, or non-uniqueness, arises when different parameterizations or model structures yield equivalent system representations. In data-driven settings, this corresponds to model multiplicity (the Rashomon effect), where distinct architectures or weight configurations achieve similar predictive performance [4,23]. In Earth system modeling, equifinality has a physical interpretation: for instance, different combinations of surface ($r_s$) and aerodynamic ($r_a$) resistance can produce identical evapotranspiration (ET) estimates despite representing different underlying processes [8]. A common mitigation strategy is the use of physics-aware neural networks, which constrain the hypothesis space through physical regularization [6]. Embedding physical constraints in the loss function helps filter implausible solutions [8,32], while alternative approaches introduce causal structure [7]. However, most work assumes that the imposed physical priors are correct—an assumption that may fail under extrapolation to unseen climate regimes.

*Uncertainty in Hybrid Climate Projections* Uncertainty in climate modeling stems from the mismatch between simplified models and complex real-world dynamics. While well studied in purely physical and purely data-driven models, uncertainty in hybrid models is inherently multifaceted, arising from data-driven components, physical components, and their interactions [30,32]. Physics-aware neural networks propagate data uncertainty through differentiable modules, while physical components introduce additional uncertainty from internal climate variability and emission scenarios [18,24]. A key gap is the interaction between physical constraints and future uncertainty. Relationships that appear stable during training may break down under novel climate conditions. When constraints are imperfect, this can amplify uncertainty and lead to unreliable projections. This study addresses this gap by examining how the correctness of physical constraints affects model stability under future climate variability.

## 3 Data and Hybrid Modeling Framework

### 3.1 Climate Data

We used outputs from the IPSL-CM5A-LR[3] climate model, which participated in CMIP5, covering the historical period (1861–2005) and three emission scenarios (RCP 2.6, RCP 4.5, and RCP 8.5) for the period 2006–2100. IPSL-CM5 is a

[3] https://cmc.ipsl.fr/ipsl-climate-models/ipsl-cm5/

full Earth system model that, in addition to physical atmosphere–land–ocean–sea ice components, includes representations of the carbon cycle, stratospheric chemistry, and tropospheric chemistry. IPSL-CM5A-LR is the low-resolution version, with a resolution of $1.9^\circ \times 3.75^\circ$. To account for the internal variability of the chaotic climate system, four realizations (R1-R4) were run in parallel. These RCP scenarios, with different pathways and realizations under varied initial conditions, collectively yield an ensemble of future climate forcing, as a benchmark for ET extrapolations.

### 3.2 Evapotranspiration and Physical Formulation

The IPSL-CM5A-LR model provides physically consistent estimates of ET for historical periods and future scenarios. However, these estimates are computed using sophisticated sub-grid biological processes. To isolate key experimental factors, potential ET is approximated with the Penman-Monteith (PM) [21,20] equation, which incorporates two critical biophysical parameters: aerodynamic resistance ($r_a$, measured in s m$^{-1}$) and surface resistance ($r_s$, measured in s m$^{-1}$):

$$\lambda ET = \frac{sR_n^* + \rho_a C_p D / r_a}{s + \gamma(1 + r_s/r_a)}, \tag{1}$$

where $\lambda ET$ is the latent heat flux (MJ m$^{-2}$); $\lambda$ is the latent heat of vaporization (J kg$^{-1}$), calculated as a function of air temperature; $s$ is the slope of the saturation vapor pressure curve (Pa K$^{-1}$); $R_n^*$ is the net radiation minus ground heat flux (MJ m$^{-2}$); $\rho_a$ is the air density (kg m$^{-3}$); $C_p$ is the specific heat of air at constant pressure (J kg$^{-1}$ K$^{-1}$); $D$ is the vapor pressure deficit at 2m height, i.e., the difference between saturation vapor pressure and actual vapor pressure (KPa); ; and $\gamma$ is the psychrometric constant (Pa K$^{-1}$) [21,20]. These intermediate variables are derived from the monthly climate data.

The numerical values of $r_a$ are estimated using the Big Leaf formulation described in [14], which computes $r_a$ as the sum of the aerodynamic resistance for momentum ($r_{\mathrm{am}} = \mathrm{WS}/\mathrm{U}^{*2}$) and the canopy boundary layer resistance for heat ($r_{\mathrm{bh}} = 6.2\mathrm{U}^{*-0.667}$), where WS is the wind speed (m s$^{-1}$) and U$^*$ is the friction velocity (m s$^{-1}$), derived from turbulent surface stress.

The ground-truth values of $r_s$ are obtained by inverting the PM equation using the IPSL-CM5A-LR outputs of actual ET and other meteorological forcing variables:

$$r_s = \frac{\rho_a C_p D + r_a s(R_n^* - \lambda ET)}{\gamma \lambda ET} - r_a. \tag{2}$$

To exclude the influence of water stress and frozen water, we follow the approach in [29]. For each model year, the annual mean $r_s$ is calculated by averaging the monthly $r_s$ values over all grid cells and months that are not water-limited. This simplification retains the influence of climate change on ET while reducing the number of intermediate factors, thereby enabling tractable evaluation of physical constraints under climate forcing.

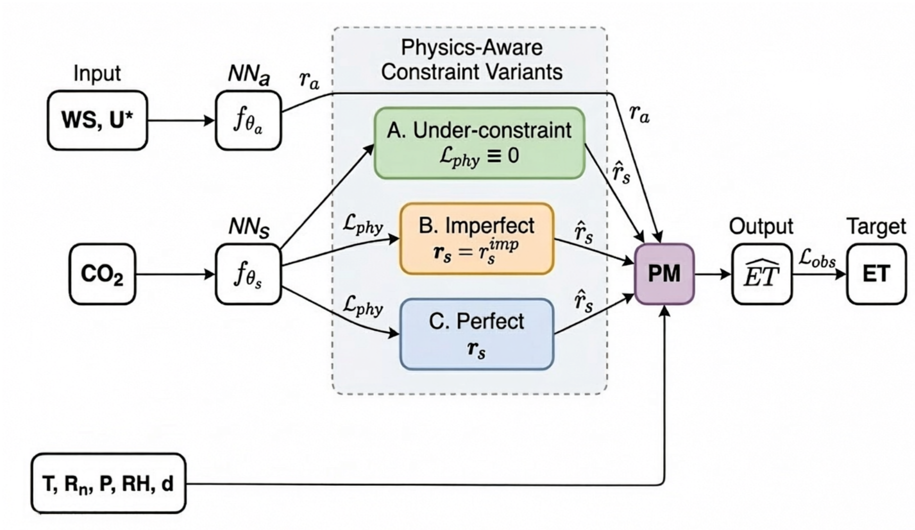



Fig. 1: Overview of the proposed hybrid models and the physical constraint variants. The constraint block illustrates how the predicted $r_s$ is constrained by different physical knowledge, utilizing either zero penalty (A), imperfect target $r_s^{imp}$ (B), or a perfect ground-truth replacement $r_s^{true}$ (C), where the neural network is completely bypassed. The physical penalty ($\mathcal{L}_{phy}$) acts on the intermediate $r_s$ representations, while the data-driven observation loss ($\mathcal{L}_{obs}$) minimizes the errors between the prediction ($\widehat{\text{ET}}$) and the target (ET).

### 3.3 Proposed Hybrid Model and Physical Constraints

Since potential ET simulations depend jointly on $r_a$ and $r_s$, our hybrid models are designed to learn these two biophysical variables, from which ET is then derived using the PM equation. While existing studies have demonstrated the effectiveness of hybrid models by imposing prior constraints on $r_a$ [8], we introduce a constraint on the loss minimizing surface resistance $r_s$. This simplification is motivated by the fact that surface resistance is more sensitive to global warming than aerodynamic resistance. Empirically, rising $CO_2$ and increased vapor pressure deficit can drive physiological stomatal closure, resulting in a substantial increase in surface resistance. Elevated $CO_2$ directly induces stomatal closure and reduces stomatal density, thereby increasing resistance to water vapor flux [1,17]. In contrast, aerodynamic resistance exhibits less change under these climatic shifts [29,28]. However, aerodynamic resistance is also undergoing indirect climate-driven changes, and alters the land-atmosphere energy balance [31]. Additionally, soil dryness and leaf area index also create a non-linear response in $r_s$ that is difficult to simplify.

The overall loss function is formulated as:

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{obs}}(y, \widehat{y}) + \lambda_{\text{phy}}\mathcal{L}_{\text{phy}}(r_s, \widehat{r_s}), \quad (3)$$

where $\mathcal{L}_{\text{obs}}(y, \widehat{y})$ denotes the data-driven loss computed from observations, and $\mathcal{L}_{\text{phy}}(r_s, \widehat{r_s})$ represents the physics-based constraint. The coefficient $\lambda_{\text{phy}} \geq 0$ controls the relative weight of the physical regularization during training.

To systematically evaluate the effects of constraints, we categorize the physical constraints into three subtypes by degree of correctness: under-constraint, imperfect, and perfect. The subtypes of hybrid models are defined by different assumptions regarding the target $r_s$, as illustrated in Fig 1:

A. **Under-constraint ($\mathcal{L}_{\mathbf{phy}} \equiv 0$)**: The hybrid model is under-constrained, as both intermediate variables $(r_a, r_s)$ are simultaneously estimated from only the ET target. No physical prior is thus applied.
B. **Imperfect ($r_s = r_s^{imp}$):** The incorporated physics shares the same structural form as the true physics but contains parametric errors or uncertainty. The NNs jointly estimate $r_a$ and $r_s$, penalized by an imperfect knowledge or empirical approximation of surface resistance $r_s^{imp}$ (e.g., Stockle [26], Yang [29], or the historical ground truth). These constraints are frequently imperfect, as observations may not accurately proxy future climate regimes, and incorrect proxies may lead to even worse approximations.
C. **Perfect (Replacement):** The assumed physics perfectly matches the true system dynamics. In this special case, the NNs estimate only $r_a$, and the prediction of $r_s$ is entirely bypassed; the PM equation strictly receives the ground truth $r_s$ during both training and inference.

Note that our focus is on hybrid models that embed physical constraints directly into the loss function. Alternative strategies, such as physical feature engineering, multi-task learning, multiphysics-informed, and causal hybrid models, are left for further discussion.

## 4 Experimental Design

In this Section, we first describe the employed climate data (Section 4.1), then propose the comparative baseline models (Section 4.2) and finally describe the employed training setting and selected evaluation metrics (Section 4.3).

### 4.1 Study Domain and Data Preparation

We chose a global study region spanning from 90°N to 70°S in latitude and covering all longitudes from 180°W to 180°E, with a grid size of $85 \times 96$. The study period extends from 1966 to 2100. This period was divided into non-overlapping subsets: the training data cover Jan. 1966 to Dec. 1995, the validation data cover Jan. 1996 to Dec. 2005 for model selection, and the testing data cover Jan. 2006 to Dec. 2100 for evaluation purposes.

To represent atmospheric and surface states, our models utilize seven core input variables: mean temperature at 2 m, wind speed at 10 m, net radiation, surface pressure, relative humidity, day of the month, and friction velocity. Atmospheric $CO_2$ is additionally included to capture the impact of anthropogenic climate forcings. To preserve the underlying dynamics, all variables are mapped to consistent physical units without other preprocessings.

### 4.2 Baseline Models for Evapotranspiration Estimation

To comprehensively evaluate our proposed framework, we compare it against five baseline models for ET estimation:

- Data-driven models: Random Forest (RF) and Deep Neural Network (DNN).
- Physical methods: The reference crop Penman–Monteith model (PM-RC) [2]; PM-Stockle, which adjusts $r_s$ for C3/C4 crops, as suggested by [26]; and PM-Yang, which links $r_s$ changes to atmospheric $CO_2$ concentration [29]. An optimized version, PM-Yang-Cali, is also included, in which parameters are calibrated using a least-squares Trust-Region Reflective algorithm.

All baseline models receive the full suite of input variables. Conversely, to ensure physical consistency, the hybrid models first predict the latent variables $r_a$ and $r_s$ to estimate ET based on the PM equation, partitioning the inputs: the variables for predicting $r_a$ are restricted to wind speed and friction velocity, while the variable for predicting $r_s$ is only the atmospheric $CO_2$.

### 4.3 Model Training and Evaluation Metrics

To enhance robustness against outliers, both the data-driven loss ($\mathcal{L}_{obs}$) and the physical penalty ($\mathcal{L}_{phy}$) are formulated using the Mean Absolute Error (MAE) objective:

$$\min_{\theta_{r_a},\theta_{r_s}} \frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{ET}_i - \widehat{\mathrm{ET}}_i\right| + \lambda_p\left|r^*_{s,i} - \hat{r}_{s,i}\right|, \tag{4}$$

where $N$ is the batch size, $\mathrm{ET}_i$ and $\widehat{\mathrm{ET}}_i$ are the true and predicted evapotranspiration, respectively, and $r^*_{s,i}$ denotes the constraint target for the specified model category.

To optimize hyperparameters, we performed a standard grid search, selecting the values that yielded the lowest validation RMSE. For the RF, the maximum tree depth was set to 15, and the ensemble consisted of 10 trees. All NNs were trained for 100 epochs using the Adam optimizer with a batch size of 1024, an initial learning rate of 0.001, 128 hidden units with ReLU activations, and 3 hidden MLP layers for DNN, 2 for predicting $r_a$, and 4 for predicting $r_s$, as in [8]. Early stopping with a patience of 10 epochs was applied. For the hybrid models, the hyperparameter $\lambda_p$ was empirically set to $10^{-2}$ to give prior knowledge a moderate influence in the overall loss function.

The model achieving the lowest validation RMSE during training was selected for the experimental results. All reported results are averaged over ten independent runs. Quantitative metrics are computed for a comprehensive comparison. Specifically, we use Root Mean Squared Error (RMSE) to estimate absolute predictive errors, Percent Bias (PBIAS) to assess systematic bias, and Root Mean Squared Calibration Error (RMSCE) to assess calibration reliability, which measures the area between the calibration curve and the diagonal representing perfect calibration. The uncertainty stems from the ensembles of future climate (different realizations and RCP scenarios), and the lower RMSCE values indicate better reliability.

## 5 Results and Discussion

In this Section, we first compare our method with other baseline models (Section 5.1), and then perform an ablation study (Section 5.2).

### 5.1 Comparison with Baseline Models

In our first experiment, we compared the performance of the baselines against our proposed Physics-aware NN (specifically, the Imperfect variant using PM-Yang). Quantitative metrics in terms of RMSE, PBIAS, and RMSCE are reported in Fig. 2a. The experimental results demonstrate the superiority of the Physics-aware NN across almost all metrics compared to purely data-driven models (RF, DNN) and strictly physical models (PM-RC, PM-Stockle, PM-Yang).

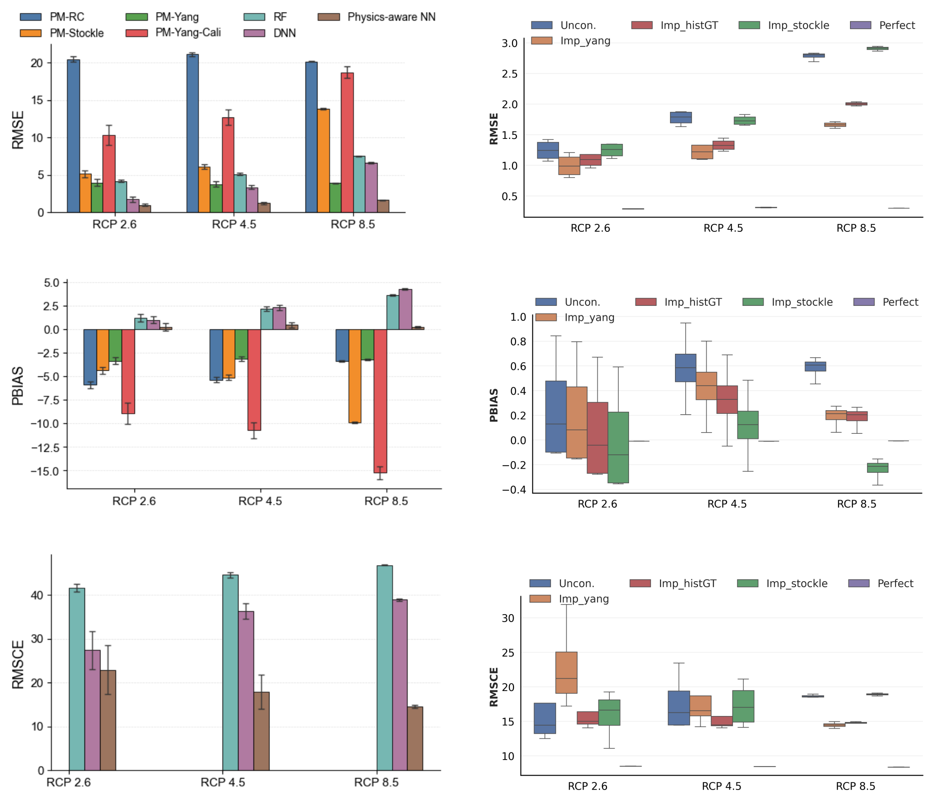
However, the performance of the Physics-aware NN is distinctly affected by the different climate forcings. First, as climate signals shift from RCP 2.6 to RCP 8.5, predictive errors increase, indicating that extreme extrapolation conditions impose significant challenges on the Physics-aware NN. This degradation suggests that imperfect physical constraints may fail to capture the complex dynamics under strong forcing scenarios. Second, a notable limitation is observed in RCP 2.6, where the RMSCE variance is significantly higher than that of RCP 8.5. This divergence suggests that a Physics-aware NN may be sensitive to the internal stochastic variability. Intuitively, the reason may be that the misalignment between the rigid constraints of equations and the stochastic nature of data leads to over-constrained predictions that fail to capture low-signal-to-noise regimes.

### 5.2 The Impact of Physical Correctness

The ablation study in Fig. 2b highlights how the correctness of physical constraints dictates the ability of physics-aware NNs to mitigate equifinality and bias through regularizations. All constrained models present lower PBIAS than the under-constrained model. Specifically, the perfect variant represents the theoretical upper bound of this approach, effectively eliminating parameter ambiguity by anchoring latent variables to ground-truth values (yielding RMSEs $\sim 0.30$).

However, our findings reveal that the efficacy of these constraints depends on their parametric accuracy. Imperfect priors (PM-Stockle), which perform unreliably in Fig. 2a, can lose their regulatory effect on physics NN under extreme climate forcing (RCP 8.5). In this case, the integration of imperfect physical knowledge may induce untrustworthy bias, severely degrading future predictions, and being worse than under-constrained models.

Furthermore, the effectiveness of physical constraints is modulated by different climate forcings; in low-forcing regimes like RCP 2.6, all imperfect models exhibit higher internal variability (the *noise*), exacerbating equifinality, inconsistent physical states across different realizations. In contrast, in high-forcing scenarios like RCP 8.5, the intensified thermodynamic drivers reduce stochastic uncertainty while increasing predictive errors, suggesting that imperfect physical




(a) (b)

Fig. 2: Comparison of ET estimation performance across RCP scenarios over the test period (2016-2100). Left column shows all models, right column presents the ablation study of physics-aware neural networks. Each panel reports mean performance over realizations (R1–R4), with uncertainty indicating variability.

constraints may fail under extrapolation to unseen climate regimes. Therefore, while physics-aware networks offer powerful regularization, injecting imperfect physical knowledge can catastrophically degrade future projections under extreme climate shifts.

## 6 Conclusion

In this study, we assessed the reliability of physics-aware neural networks for estimating evapotranspiration under future climate change scenarios. We demonstrated that equifinality and uncertainty are inherently coupled, jointly influencing model performance.

By categorizing hybrid models into distinct subtypes based on their physical correctness, our experiments show that while physical constraints successfully mitigate equifinality through structural regularization, their effectiveness depends strictly on the correctness of the embedded physical knowledge. Crucially, imperfect constraints can severely degrade predictive performance, particularly under high-forcing scenarios (e.g, RCP 8.5), whereas low-forcing regimes (e.g., RCP 2.6) amplify the model's sensitivity to internal climate variability.

These findings highlight a key consideration for the machine-learning-based Earth science community: in hybrid modeling, physical interpretability is not a guarantee of reliability, but is rather contingent on prior errors and climate uncertainties. Future work should focus on robust uncertainty quantification and developing adaptive regularization strategies that are resilient to non-stationary climate change, ultimately enhancing the trustworthiness of neural Earth system models.

## Acknowledgements

This work was supported by the ERC Synergy Grant USMILE (GA 855187) and the EU projects XAIDA (GA 101003469), AI4PEX (GA 101137682) and ELIAS (GA 101120237). M.-Á. F.-T. acknowledges the support by the Universidad Carlos III de Madrid, Spain (project 2025/00803/001), the Regional Government of Madrid (TEC-2024/COM-322), and the computational resources and technical support provided by Artemisa, funded by the ERDF and Comunitat Valenciana, and the Instituto de Física Corpuscular, IFIC (CSIC-UV).

## References

1. Adams, M.A., Buckley, T.N., Binkley, D., Neumann, M., Turnbull, T.L.: Co2, nitrogen deposition and a discontinuous climate response drive water use efficiency in global forests. Nature Communications **12**(1), 5194 (2021), https://doi.org/10.1038/s41467-021-25365-1
2. Allen, R.G., Pereira, L.S., Raes, D., Smith, M., et al.: Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. Fao, rome **300**(9), D05109 (1998), https://www.fao.org/4/x0490e/x0490e00.htm
3. Beven, K., Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. Journal of Hydrology **249**(1), 11–29 (2001). https://doi.org/https://doi.org/10.1016/S0022-1694(01)00421-8, https://www.sciencedirect.com/science/article/pii/S0022169401004218
4. Breiman, L.: Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science **16**(3), 199–231 (2001). https://doi.org/https://doi.org/10.1214/ss/1009213726
5. Camps-Valls, G., Reichstein, M., Zhu, X., et al.: Advancing deep learning for earth sciences: From hybrid modeling to interpretability. In: IGARSS 2020. pp. 3979–3982. IEEE (2020). https://doi.org/https://doi.org/10.1109/IGARSS39084.2020.9323558

6. Camps-Valls, G., Svendsen, D., Martino, L., Muñoz-Marí, J., Laparra, V., Campos-Taberner, M., Luengo, D.: Physics-aware Gaussian processes in remote sensing. Applied Soft Computing **68**, 69–82 (Jul 2018). https://doi.org/https://doi.org/10.1016/j.asoc.2018.03.021
7. Cohrs, K.H., Varando, G., Carvalhais, N., Reichstein, M., Camps-Valls, G.: Causal hybrid modeling with double machine learning—applications in carbon flux modeling. Machine Learning: Science and Technology **5**(3), 035021 (jul 2024). https://doi.org/10.1088/2632-2153/ad5a60, https://dx.doi.org/10.1088/2632-2153/ad5a60
8. ElGhawi, R., Kraft, B., Reimers, C., et al.: Hybrid modeling of evapotranspiration: inferring stomatal and aerodynamic resistances using combined physics-based and machine learning. Environ. Res. Lett. **18**(3), 034039 (2023). https://doi.org/https://doi.org/10.1088/1748-9326/acbbe0
9. Hu, X., Shi, L., Lin, G., Lin, L.: Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation. Journal of Hydrology **601**, 126592 (2021). https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126592
10. Irrgang, C., Boers, N., Sonnewald, M., et al.: Towards neural earth system modelling by integrating artificial intelligence in earth system science. Nat. Mach. Intell. **3**(8), 667–674 (2021). https://doi.org/https://doi.org/10.1038/s42256-021-00374-3
11. Karpatne, A., Jia, X., Kumar, V.: Knowledge-guided machine learning: Current trends and future prospects (2024), https://arxiv.org/abs/2403.15989
12. Karpatne, A., Kannan, R., Kumar, V.: Knowledge Guided Machine Learning: Accelerating Discovery using Scientific Knowledge and Data. Chapman and Hall/CRC, 1 edn. (2022)
13. Khatami, S., Peel, M.C., Peterson, T.J., Western, A.W.: Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty. Water Resources Research **55**(11), 8922–8941 (2019). https://doi.org/https://doi.org/10.1029/2018WR023750, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023750
14. Knauer, J., El-Madany, T.S., Zaehle, S., Migliavacca, M.: Bigleaf—an r package for the calculation of physical and physiological ecosystem properties from eddy covariance data. PloS one **13**(8), e0201114 (2018). https://doi.org/https://doi.org/10.1371/journal.pone.0201114
15. Koppa, A., Rains, D., Hulsman, P., et al.: A deep learning-based hybrid model of global terrestrial evaporation. Nat. Commun. **13**(1), 1912 (2022). https://doi.org/https://doi.org/10.1038/s41467-022-29543-7
16. Kraft, B., Jung, M., Körner, M., et al.: Towards hybrid modeling of the global hydrological cycle. Hydrol. Earth Syst. Sci. Discuss. **2021**, 1–40 (2021). https://doi.org/https://doi.org/10.5194/hess-26-1579-2022
17. Li, W., Migliavacca, M., Miralles, D.G., Reichstein, M., Anderegg, W.R., Yang, H., Orth, R.: Disentangling effects of vegetation structure and physiology on land–atmosphere coupling. Global Change Biology **31**(1), e70035 (2025). https://doi.org/https://doi.org/10.1111/gcb.70035
18. Martínez-Ferrer, L., Álvaro Moreno-Martínez, Campos-Taberner, M., García-Haro, F.J., Muñoz-Marí, J., Running, S.W., Kimball, J., Clinton, N., Camps-Valls, G.: Quantifying uncertainty in high resolution biophysical variable retrieval with machine learning. Remote Sensing of Environment **280**, 113199 (2022). https://doi.org/https://doi.org/10.1016/j.rse.2022.113199, https://www.sciencedirect.com/science/article/pii/S0034425722003091

19. Milly, P.C., Dunne, K.A.: Potential evapotranspiration and continental drying. Nat. Clim. Change **6**(10), 946–949 (2016). https://doi.org/https://doi.org/10.1038/nclimate3046
20. Monteith, J.L.: Evaporation and environment. Symposia of the Society for Experimental Biology **19**, 205–234 (1965), https://repository.rothamsted.ac.uk/item/8v547/evaporation-and-environment
21. Penman, H.L.: Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **193**(1032), 120–145 (1948). https://doi.org/10.1098/rspa.1948.0037, https://doi.org/10.1098/rspa.1948.0037
22. Reichstein, M., Camps-Valls, G., Stevens, B., et al.: Deep learning and process understanding for data-driven earth system science. Nat. **566**(7743), 195–204 (2019). https://doi.org/https://doi.org/10.1038/s41586-019-0912-1
23. Rudin, C., Zhong, C., Semenova, L., Seltzer, M., Parr, R., Liu, J., Katta, S., Donnelly, J., Chen, H., Boner, Z.: Amazing things come from having many good models. arXiv preprint arXiv:2407.04846 (2024), https://arxiv.org/pdf/2407.04846
24. Shen, C., Appling, A.P., Gentine, P., et al.: Differentiable modelling to unify machine learning and physical models for geosciences. Nat. Rev. Earth Environ. **4**(8), 552–567 (2023). https://doi.org/https://doi.org/10.1038/s43017-023-00450-9
25. Shen, X., Meinshausen, N.: Engression: Extrapolation through the lens of distributional regression (2024), https://arxiv.org/abs/2307.00835
26. Stockle, C.O., Williams, J.R., Rosenberg, N.J., Jones, C.A.: A method for estimating the direct and climatic effects of rising atmospheric carbon dioxide on growth and yield of crops: Part i—modification of the epic model for climate change analysis. Agricultural systems **38**(3), 225–238 (1992). https://doi.org/https://doi.org/10.1016/0308-521X(92)90067-X
27. Tosan, M., Gharib, M., Attar, N., Maroosi, A.: Enhancing evapotranspiration estimation: a bibliometric and systematic review of hybrid neural networks in water resource management. Computer Modeling in Engineering & Sciences **142**(2), 1109 (2025). https://doi.org/https://doi.org/10.32604/cmes.2025.058595
28. Vahmani, P., Jones, A.D., Li, D.: Will anthropogenic warming increase evapotranspiration? examining irrigation water demand implications of climate change in california. Earth's Future **10**(1), e2021EF002221 (2022). https://doi.org/https://doi.org/10.1029/2021EF002221
29. Yang, Y., Roderick, M.L., Zhang, S., McVicar, T.R., Donohue, R.J.: Hydrologic implications of vegetation response to elevated co2 in climate projections. Nature Climate Change **9**(1), 44–48 (2019). https://doi.org/https://doi.org/10.1038/s41558-018-0361-0
30. Yip, S., Ferro, C.A., Stephenson, D.B., Hawkins, E.: A simple, coherent framework for partitioning uncertainty in climate predictions. Journal of climate **24**(17), 4634–4643 (2011). https://doi.org/https://doi.org/10.1175/2011JCLI4085.1
31. Zhang, G.: Amplified summer wind stilling and land warming compound energy risks in northern midlatitudes. Environmental Research Letters **20**(3), 034015 (2025). https://doi.org/https://10.1088/1748-9326/adb1f8
32. Zhang, M., Fernández-Torres, M.Á., Cohrs, K.H., et al.: Calibration and uncertainty quantification for deep learning-based drought detection. Int. J. Appl. Earth Obs. Geoinf. **140**, 104563 (2025). https://doi.org/https://doi.org/10.1016/j.jag.2025.104563
33. Zhao, W.L., Gentine, P., Reichstein, M., et al.: Physics-constrained machine learning of evapotranspiration. Geophys. Res. Lett. **46**(24), 14496–14507 (2019). https://doi.org/https://doi.org/10.1029/2019GL085291