

Article

Hybrid Data-Driven Framework for Multi-Month Precipitation Forecasting over Boyacá with CHIRPS and DEM Context in Light Mode

Manuel Ricardo Pérez Reyes ^{1,3} , José Antonio Reyes Suárez ² , Marco Javier Suárez Barón ³  and Óscar Javier García Cabrejo ⁴ 

¹ Pedagogical and Technological University of Colombia (UPTC), School of Systems Engineering, Sogamoso 152210, Colombia

² Department of Bioinformatics, Faculty of Engineering, University of Talca, Talca 34600000, Chile

³ Pedagogical and Technological University of Colombia (UPTC), School of Systems Engineering, Sogamoso 152210, Colombia

⁴ Pedagogical and Technological University of Colombia (UPTC), School of Geology, Sogamoso 152210, Colombia

* Correspondence: manuelricardo.perez@uptc.edu.co

Abstract

Background: Reliable multi-month precipitation forecasts underpin flood and drought preparedness in mountainous regions. We present a hybrid data-driven framework that couples CHIRPS v2.0 precipitation with a Boyacá DEM, orchestrating feature bundles and lightweight architectures in a light-mode setting. **Methods:** Using a 60-month input window and two horizons (6 and 12 months), we implement three hybrid-inspired families (BASIC, KCE, PAFC) over a 5-by-5 cut to accelerate experimentation on a single A100 (80 GB). The framework blends data-preprocessing hybrids (temporal/DEM engineering), component-combination hybrids (attention with conv nets), and post-processing bias analysis. **Results:** BASIC yields the best balance (RMSE 63–96 mm, R² up to 0.56, near-zero bias at 12 months); KCE attention reduces bias; PAFC improves RMSE but loses explanatory power. **Conclusions:** The light-mode hybrid framework enables rapid model selection; next steps include full-domain training for Boyacá at both horizons, post-hoc bias calibration, and expanded interpretability.

Keywords: CHIRPS; DEM; Boyacá; hybrid models; data-driven framework; ConvLSTM; attention; precipitation forecasting; hydrology; GPU

Received:

Accepted:

Published:

Citation: Pérez Reyes, M.R.; Reyes Suárez, J.A.; Suárez Barón, M.J.; García Cabrejo, Ó.J. Hybrid Data-Driven Framework for Precipitation Forecasting. *Hydrology* **2025**, *1*, 0. <https://doi.org/>

Copyright: © 2025 by the authors. Submitted to *Hydrology* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multi-month precipitation forecasts underpin early warnings for floods, droughts, and reservoir operations in complex mountain terrain. In Boyacá, Colombia, sparse gauges and strong orographic gradients demand remote-sensing inputs such as CHIRPS v2.0 and topographic context from the regional DEM. We present a hybrid data-driven framework that couples these sources with engineered temporal and terrain features, enabling lightweight spatio-temporal architectures to train efficiently in a light-mode setting. Unlike purely convective baselines, the framework blends (i) data-preprocessing hybrids (seasonality encodings and DEM-derived relief), (ii) component-combination hybrids (convolutional encoders with attention blocks), and (iii) post-processing bias inspection to test how static relief, elevation clusters, and precipitation memory improve mid-range skill. Hybrid ap-

proaches have shown benefits for monthly precipitation [8], and we adopt those principles here to accelerate iteration on compact models.

Recent hydrological studies show that recurrent and attention-based models can outperform traditional learners for runoff and rainfall [2,5,7], and ensemble/stacking strategies continue to reduce bias and variance [6,8]. Building on this evidence, we target Boyacá with a 5×5 prototype grid to accelerate iteration on a single A100 GPU (80 GB) before scaling to the full departmental extent. The contribution is twofold: (i) a documented feature-engineering pipeline (BASIC, KCE, PAFC) aligned with the complete CHIRPS+DEM dataset, and (ii) a reproducible light-mode training/evaluation loop with standardized metrics that can be extended to larger domains. By keeping the workflow modular and light, the same scaffold can ingest future static covariates (soil, land cover) and climate indices while preserving reproducibility and transfer to the full Boyacá grid.

2. Materials and Methods

2.1. Data access and preprocessing pipeline

CHIRPS v2.0 monthly precipitation (0.05° , $\sim 5.4\text{--}5.5$ km) is downloaded from the Climate Hazards Center (<https://data.chc.ucsb.edu/products/CHIRPS-2.0/>). Daily CHIRPS is aggregated to monthly totals by summing daily accumulations before constructing windows. The Boyacá DEM comes from NASA SRTM (native 90 m, <https://srtm.csi.cgiar.org/>); it is bilinearly resampled upstream to 0.05° to co-register with CHIRPS, yielding a common grid of 59×61 cells for the departmental extent. Static derivatives (slope, aspect, relief) are computed after resampling. The engineered NetCDF file (complete_dataset_with_features_with_clusters_elevation_windows_imfs_with_onehot_elevation_clean.nc) stores 518 monthly timesteps, 6165 grid points, and 15 variables; it underpins both full-grid and light-cut experiments.

2.2. Study area and spatial extent

Boyacá spans Andean valleys and high plateaus, with sharp orographic gradients that modulate convection. CHIRPS at 0.05° yields 59 cells north–south and 61 west–east covering the department. For rapid iteration, a 5×5 light cut (indices lat [28:33], lon [30:35]) samples representative relief and climatic regimes while guaranteeing that the A100 fits all families in memory. The full-grid pipeline remains ready for scaling once feature choices are frozen.

2.3. Compute environment

Experiments ran on Google Colab Pro+ with one NVIDIA A100 (80 GB HBM), 167 GB system RAM, and 235 GB storage. The light cut guarantees all models fit in GPU memory and allows fast epoch turnaround for grid searches.

2.4. Feature bundles and preprocessing

Sliding windows of 60 months are built with stride 1; targets are the next H months. Missing values are forward-filled per pixel; static layers (DEM, relief) are broadcast across time. Three bundles feed the families:

- **Basic (BASIC):** Temporal encodings (year, month, month_sin, month_cos, doy_sin, doy_cos); daily precipitation summaries from CHIRPS (max_daily_precipitation, min_daily_precipitation, daily_precipitation_std); static terrain (elevation, slope, aspect). All channels are z-normalized with training statistics.
- **KCE:** BASIC plus elevation clusters encoded as one-hot (elev_high, elev_med, elev_low) to guide meteorological attention; inputs are z-normalized and clipped at 3σ .

- **PAFC:** KCE plus precipitation memory lags (`total_precipitation_lag1`, `lag2`, `lag12`); static DEM/relief retained as embeddings; per-channel standardization using training statistics.

Reverse-engineering the analysis notebooks shows that KCE clusters derive from elevation quantiles computed on the full Boyacá DEM, followed by one-hot encoding to stabilize attention heads. PAFC inherits KCE and adds lags of monthly totals to capture persistence; the lags are capped to avoid exploding gradients and standardized with training statistics. All bundles draw from the same engineered file; the light-mode 5×5 cut slices the cube without altering normalization.

2.5. Model families and architectures

- **BASIC:** ConvLSTM/ConvRNN baselines with residual and bidirectional variants; a light Transformer baseline for temporal attention. These act as hybrid component-combination models by mixing convolutional encoders and attention heads. Residual skip paths stabilize deep temporal stacks, and bidirectional gates preserve seasonality within the 60-month window.
- **KCE:** ConvLSTM variants with meteorological attention aimed at reducing spatially coherent bias. Attention heads attend to elevation clusters to redistribute weights across valleys and ridges, exemplifying hybrid spatial reweighting of convolutional features.
- **PAFC:** ConvLSTM architectures with parameter-efficient attention, emphasizing compactness; static DEM/relief channels stabilize training. Depthwise separable kernels and squeeze-excite style attention reduce memory footprint, yielding a lightweight hybrid between conv encoders and efficient attention.

ConvGRU2D was unavailable; ConvGRU models were excluded after unit tests validated custom losses and layers.

2.6. Training protocol and metrics

All models train with Adam (10^{-3}), batch size 2, 150 epochs, patience 50 on validation MAE, and checkpointing of the best epoch. Metrics: RMSE, MAE, R^2 , and mean bias on the held-out validation set. Cross-horizon deltas (ΔRMSE , ΔR^2) quantify degradation/improvement when extending H from 6 to 12 months. Comparisons are family-wise and cross-family. The metrics are defined as: $\text{RMSE} = \sqrt{\frac{1}{N} \sum_i (y_i - \hat{y}_i)^2}$, $\text{MAE} = \frac{1}{N} \sum_i |y_i - \hat{y}_i|$, $R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$, and $\text{Bias} = \frac{1}{N} \sum_i (\hat{y}_i - y_i)$. RMSE penalizes peaks, MAE is robust to outliers, R^2 reports explained variance, and bias captures systematic over- or under-estimation.

3. Results

3.1. Global performance

Table 1 and Table 2 summarize averages by family. BASIC is the most balanced (bias near zero at $H = 12$). KCE reduces bias compared to PAFC via attention, while PAFC lowers RMSE yet sacrifices R^2 . Best-per-family models are shown in Tables 3 and 4. Figures 3 and 4 gather evolution and per-model comparisons, and Figure 5 covers bias and horizon deltas.

Table 1. Average metrics by experiment for $H = 6$ (`metrics_summary.csv`).

Experiment	RMSE (mm)	R^2	Bias (mm)
BASIC	96.0	0.45	-17.6
KCE	164.9	-2.07	-44.8
PAFC	116.4	0.15	-15.4

Table 2. Average metrics by experiment for $H = 12$ (`metrics_summary.csv`).

Experiment	RMSE (mm)	R^2	Bias (mm)
BASIC	63.4	0.56	-5.9
KCE	96.5	-0.36	-32.1
PAFC	115.2	-1.47	-53.8

Table 3. Best models per experiment for $H = 6$.

Experiment	Model	RMSE (mm)	R^2
BASIC	ConvLSTM_Bidirectional	81.9	0.61
KCE	ConvLSTM_Residual	92.6	0.50
PAFC	ConvLSTM_EfficientBidir	92.1	0.51

Table 4. Best models per experiment for $H = 12$.

Experiment	Model	Horizon	RMSE (mm)	R^2
BASIC	ConvRNN	1	51.63	0.71
KCE	ConvLSTM_MeteoAttention	9	58.87	0.62
PAFC	ConvLSTM_Residual	3	59.17	0.63

3.2. Architecture overview

Figure 2 depicts the end-to-end flow: CHIRPS and DEM feed preprocessing, bundle selection, family-specific encoders, and the decoder head evaluated with hydrological metrics.



(a) Colombia with Boyacá in red (QGIS-derived).

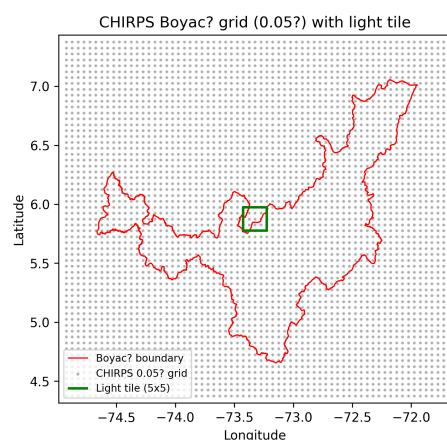
(b) CHIRPS grid (0.05°) with Boyacá boundary and light tile.

Figure 1. Spatial context. Left: QGIS map of Colombia with Boyacá highlighted in red. Right: CHIRPS-aligned Boyacá domain at 0.05° with the departmental boundary; the light-mode tile (5×5) used for rapid experiments is shaded.

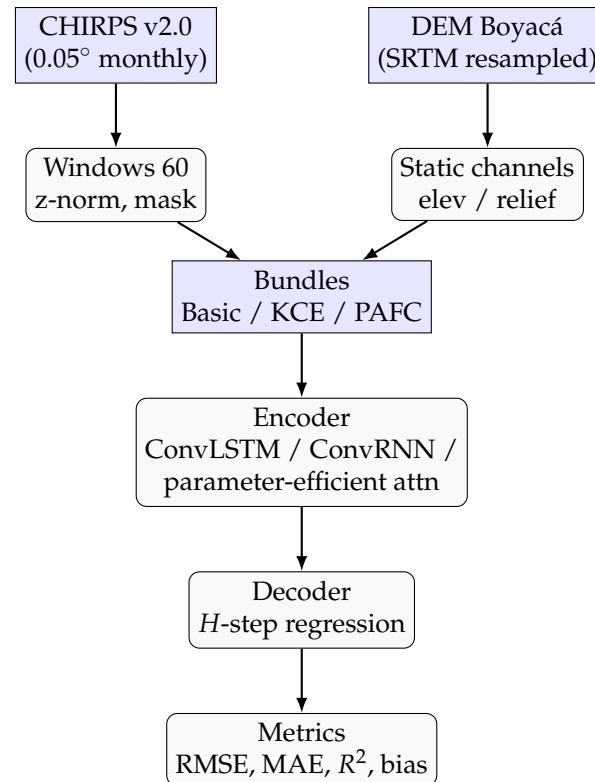


Figure 2. High-level data and model flow. Basic uses temporal CHIRPS + seasonality; KCE adds meteorological context for attention; PAFC adds static DEM-derived relief with parameter-efficient attention.

3.3. Figures from experiments

119

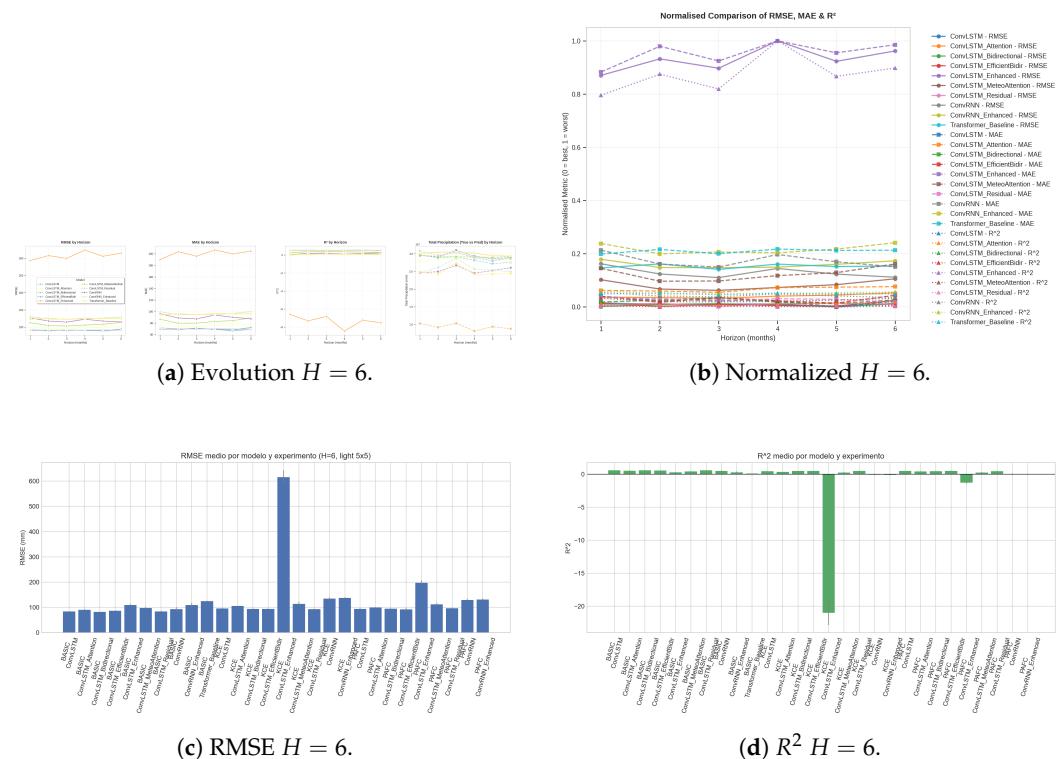


Figure 3. Training and per-model comparisons for $H=6$. BASIC converges fastest; KCE trims bias; PAFC improves RMSE but loses explanatory power.

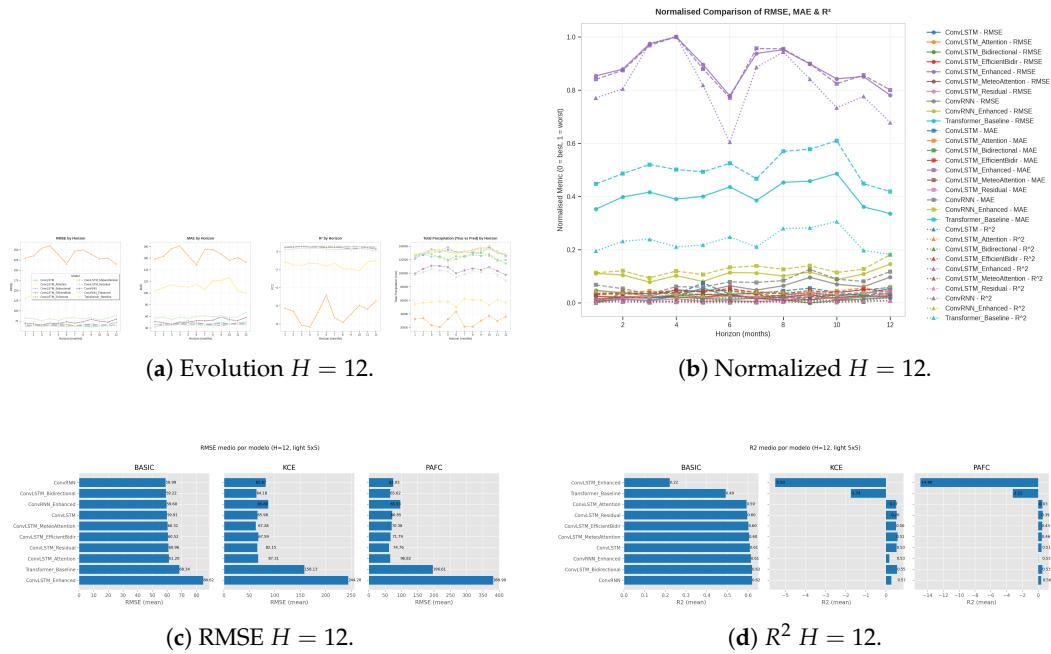


Figure 4. Training and per-model comparisons for $H=12$. BASIC remains balanced; KCE trims bias; PAFC gains RMSE at the cost of explanatory power.

The evolution panels show BASIC stabilizing early, KCE needing more epochs, and PAFC oscillating at $H = 12$. Normalized plots highlight that PAFC variance grows late in training, suggesting stronger regularization. The color palette (blue BASIC, orange KCE, green PAFC) makes these differences evident even for non-specialists. The barplots underline that bidirectional and residual ConvLSTMs dominate at $H = 6$, while compact ConvRNN variants remain competitive at $H = 12$. The R^2 drop for PAFC at $H = 12$ signals overfitting to recent lags; attention without tighter regularization hurts generalization.

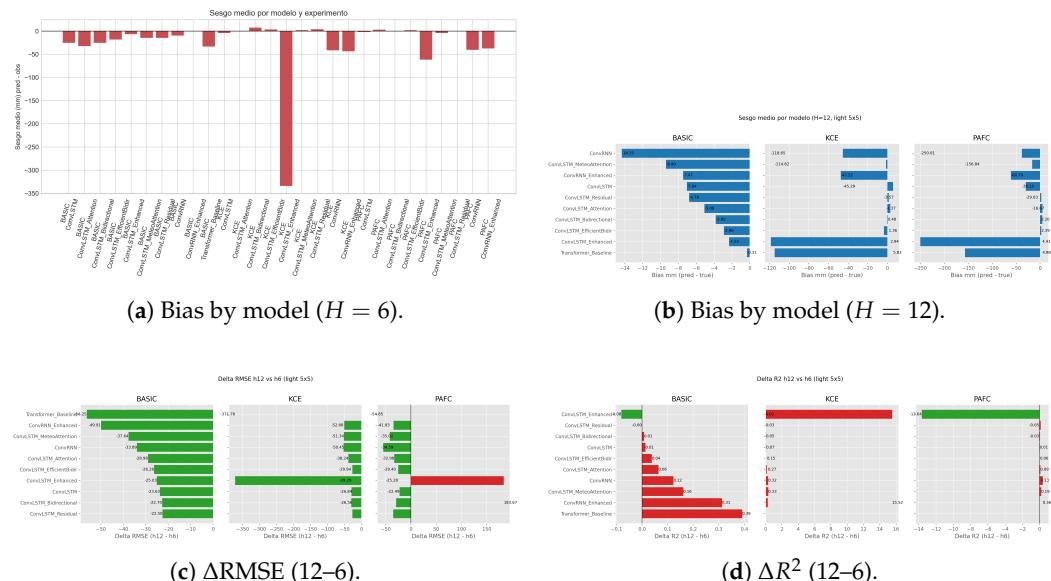


Figure 5. Bias and horizon impacts. Top: systematic biases per architecture (negative bars underestimation, positive overestimation); KCE narrows spread, BASIC centers near zero at $H=12$, PAFC underestimates more for long horizons. Bottom: extending the horizon to 12 months: green/yellow bars denote improvements; red denotes degradation. BASIC degrades least in RMSE and retains positive R^2 ; KCE gains some R^2 at mid-horizons; PAFC shows mixed changes with larger variance.

The bias panels show KCE's advantage: elevation-aware attention reins in systematic underestimation, especially at $H = 12$. BASIC remains close to zero bias after 12-month forecasting, aligning with its balanced RMSE and R^2 . Horizon deltas highlight robustness: BASIC loses the least skill when moving to 12 months, while PAFC and KCE become more sensitive to horizon length. Positive ΔR^2 bars for some KCE variants suggest attention helps mid-forecast steps but needs regularization.

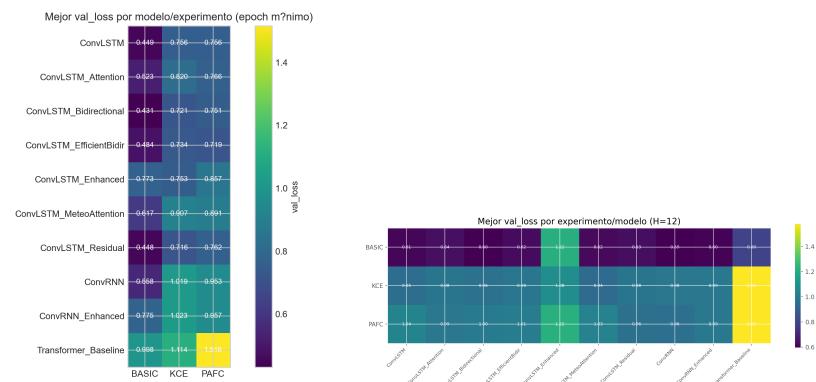
(a) Best validation loss ($H = 6$).(b) Best validation loss ($H = 12$).

Figure 6. Convergence behaviour by architecture and experiment. Cooler colors indicate lower loss. BASIC and KCE achieve the lowest validation loss at $H=6$; at $H=12$ only a subset of KCE/PAFC models remain competitive, underscoring the need for stronger regularization when adding lags and attention.

The validation-loss heatmaps confirm that architectures with residual paths and attention converge faster and deeper at $H = 6$. At $H = 12$ the spread widens, showing that only the most stable attention variants retain low loss; others require tuning of learning rate and dropout.

4. Discussion

The light-mode data-driven framework enables rapid screening of 30 architectures while preserving stable training, which is critical when iterating on feature design. With a 60-month window, the recurrent cores retain skill at $H = 12$, mirroring earlier hydrological evidence that longer contexts capture seasonality and interannual variability [2]. KCE's meteorological attention curbs systematic bias, though its gains depend on the careful scaling of elevation clusters. PAFC benefits from DEM-based relief and precipitation memory, yet the added lags can destabilize R^2 unless per-channel normalization is enforced. The compact Transformer baseline remains competitive, highlighting that lightweight attention can coexist with convolutional encoders in operational workflows.

These findings also surface practical considerations. Bias-sensitive metrics reveal that adding categorical elevation cues (KCE) or long-memory lags (PAFC) may require explicit regularization (dropout, layer normalization) to avoid overestimating extremes. Moreover, the light-mode cut is effective for hypothesis testing, but spatial generalization must be re-checked on the full Boyacá grid. Finally, the unified training recipe (Adam, early stopping on MAE, shared metrics) simplifies cross-family comparisons and provides a template to benchmark future feature bundles, including additional climate indices or soil descriptors.

5. Conclusions

BASIC, with engineered temporal and terrain features, offers the most reliable balance of error, explanatory power, and low bias at $H = 12$. KCE demonstrates that attention

mechanisms can further constrain bias when elevation clusters are scaled properly, suggesting a path to hybridizing BASIC with selective attention blocks. PAFC shows promise when precipitation memory is important, but it must control feature scaling to maintain R^2 at longer horizons. The framework is ready to incorporate additional climate indices and soil/static covariates, providing a reusable scaffold for data-driven precipitation forecasting in mountainous regions. Next steps include (i) retraining on the full Boyacá grid (59×61 cells) for horizons $H = 6$ and $H = 12$ using a multi-GPU cluster to preserve the full spatial context; (ii) exploring Fourier Neural Operators (FNO) [12] as a spectral alternative to ConvLSTMs, adapting the same CHIRPS/DEM preprocessing while leveraging Fourier-domain convolutions for long-range dependencies; and (iii) tightening regularization (dropout, weight decay, layer norm) for attention variants to stabilize R^2 at longer horizons. Additional preprocessing trials will include alternative window lengths and spatial smoothers to balance noise reduction with orographic detail.

Author Contributions: Conceptualization, methodology, validation, writing—original draft, review and editing: M.R.P.R.; data curation, software, investigation: J.A.R.S.; validation, visualization, review and editing: M.J.S.B.; supervision, project administration, review and editing: Ó.J.G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: CHIRPS v2.0 precipitation is publicly available at <https://data.chc.ucsb.edu/products/CHIRPS-2.0/>. The resampled DEM is derived from NASA SRTM (<https://srtm.csi.cgiar.org/>). Engineered NetCDF feature sets used in this study are available from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

1. Rivera, J.A.; Marianetti, G.; Hinrichs, S. Validation of CHIRPS precipitation dataset along the Central Andes of Argentina. *Atmospheric Research* **2018**, *213*, 437–449.
2. Gao, S.; Huang, Y.; Zhang, S.; Han, J.; Wang, G.; Zhang, M.; Lin, Q. Short-term runoff prediction with GRU and LSTM networks. *Journal of Hydrology* **2020**, *589*, 125188.
3. Hirpa, F.A.; Gebremichael, M.; Hopson, T. Evaluation of high-resolution satellite precipitation products over complex terrain in Ethiopia. *Journal of Applied Meteorology and Climatology* **2010**, *49*, 1044–1051.
4. Wani, O.A.; Mahdi, S.S.; Yeasin, M.; Kumar, S.S.; Gagnon, A.S.; Danish, F.; Al-Ansari, N.; El-Hendawy, S.; Mattar, M.A. Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas. *Scientific Reports* **2024**, *14*, 77687.
5. He, R.; Zhang, L.; Chew, A.W.Z. Data-driven multi-step prediction and analysis of monthly rainfall using explainable deep learning. *Expert Systems with Applications* **2024**, *235*, 121160.
6. El Hafyani, M.; El Himdi, K.; El Adlouni, S.E. Improving monthly precipitation prediction accuracy using machine learning models: a multi-view stacking learning technique. *Frontiers in Water* **2024**, *6*, 1378598.
7. Zhao, Y.; Luo, S.; Cai, J.; Li, Z.; Zhang, M. Monthly precipitation prediction based on the CEEMDAN-BMA model. *Water Resources Management* **2024**, *38*, 5661–5681.
8. Pérez Reyes, M.R.; Suárez Barón, M.J.; García Cabrejo, O.J. Spatiotemporal prediction of monthly precipitation: A systematic review of hybrid models. *Hydrology Research*, under review, 2025.
9. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-c. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems* **2015**, *28*, 802–810.
10. Ravuri, A.; et al. Skillful precipitation nowcasting using deep generative models of radar. *Nature* **2021**, *597*, 672–677.

11. Liu, Y.; et al. Spatiotemporal transformers for long-range precipitation forecasting. *Remote Sensing of Environment* **2023**, *288*, 113545. 208
12. Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Stuart, A.; Bhattacharya, K.; Anandkumar, A. Fourier neural operator for parametric partial differential equations. *International Conference on Learning Representations* **2021**, 1–16. 209
210
211
212

Disclaimer/Publisher's Note: The statements, opinions and data contained in this publication are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). The publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content. 213
214
215
216
217