

Multi-Source Uncertainty Aware Fusion for Soil Moisture Estimation

Eleftherios Polychronakis^{†,‡}, Archana Kannan^{*}, James D. Campbell^{*}, Mahta Moghaddam^{*},
Panagiotis Tsakalides^{†,‡}, Grigorios Tsagkatakis^{†,‡}

[†]Institute of Computer Science, Foundation for Research and Technology - Hellas, Greece.

[‡]Computer Science Department, University of Crete, Greece.

^{*} Department of Electrical and Computer Engineering, University of Southern California

Abstract—Soil moisture (SM) estimation from fusion of heterogeneous sources requires quantifying prediction uncertainties, not just point estimates, to account for different error sources. In this work, we propose an uncertainty-aware multi-modal fusion framework that combines predictive distributions from data-assimilation (SMAP L4) soil-moisture time series and satellite (CYGNSS delay-Doppler maps) to produce daily SM estimates. Performance is evaluated using collocated in-situ measurements from SoilSCAPE sites. The time-series leg produces a forecast distribution from a model with exogenous inputs, while the CYGNSS leg uses probabilistic regression on DDM-derived features to produce an independent retrieval distribution. We fuse the two distributions using the one-dimensional 2-Wasserstein (W2) barycenter in quantile space, and we include an L_2 barycenter baseline based on linear pooling of the predictive CDFs for comparison. A central focus of the paper is weight selection: we compare several practical ways to choose scalar fusion weights from a validation set (including equal weighting, RMSE-based tuning, quantile-space error, and CRPS-based tuning) and analyze how weight choice changes both accuracy and probabilistic skill.

Index Terms—Data assimilation, uncertainty quantification, machine learning, soil moisture, GNSS-R, CYGNSS

I. INTRODUCTION

Soil moisture (SM) estimation from heterogeneous observations is naturally probabilistic, since different paradigms introduce distinct uncertainty sources: forecasting uncertainty arises from prediction errors and their temporal accumulation, whereas retrieval uncertainty reflects measurement noise and components of SM variability that cannot be resolved from the available observations

In-situ soil-moisture measurements are highly informative and are often treated as ground reference, but they are inherently local and spatially sparse. To obtain spatially continuous estimates, data-assimilation products such as SMAP L4 provide consistent soil-moisture fields by combining models and observations, although they typically arrive with latency [1]. Complementarily, satellite observations including GNSS-Reflectometry from platforms like CYGNSS offer low-latency broad coverage through indirect measurements whose

This work has been supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “3rd Call for H.F.R.I.’s Research Projects to Support Faculty Members & Researchers” (H.F.R.I. Project Number: 26302) and by the TITAN ERA Chair project (contract no. 101086741) within the Horizon Europe Framework Program of the European Commission.

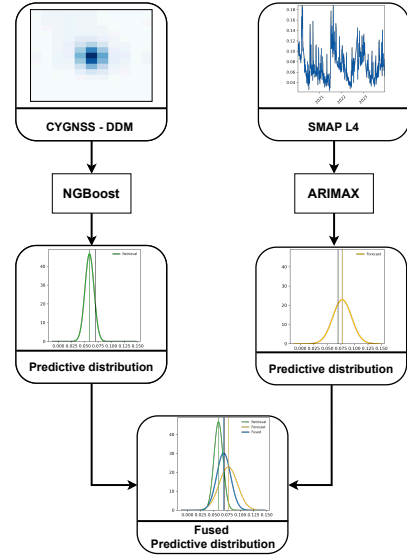


Fig. 1. Overview of the proposed multi-source probabilistic fusion framework: CYGNSS DDM observations are used for soil-moisture retrieval, while SMAP L4 data-assimilation time series are used to generate forecasts. The resulting predictive distributions are fused via 1D Wasserstein-2 (W2) barycenter quantile averaging to produce a calibrated soil-moisture estimate with associated uncertainty.

information content varies with surface conditions and topography [2–4]. This motivates a forecast–retrieval formulation, where a forecasting leg propagates the most recent SMAP L4 state forward in time, and a retrieval leg provides observational updates from CYGNSS when available.

In this work we couple two probabilistic pipelines and fuse their predictive distributions (Fig. 1). The first pipeline produces a forecast distribution from the SMAP L4 soil moisture time series collocated to each site using a model with exogenous inputs. The second maps CYGNSS delay-Doppler map (DDM) information to soil moisture via probabilistic regression [5, 6]. We combine these distributions using a one-dimensional Wasserstein-2 (W2) barycenter [7–9], which corresponds to weighted averaging of quantiles. For comparison, we also consider linear pooling, i.e., a weighted mixture of the two predictive CDFs [10–12].

The key novelties of this work are:

- Fusion of forecast and retrieval predictive distributions via a 1D Wasserstein-2 barycenter, enabling distribution-level combination rather than mixing point estimates
- Benchmarking the proposed W2 fusion against linear pooling and point baselines using RMSE and Continuous Ranked Probability Score (CRPS)
- Systematically evaluate fusion-weight selection objectives (RMSE, quantile-space MSE, CRPS), showing that quantile-space weighting provides the most stable accuracy–uncertainty trade-off across SoilSCAPE jr1–jr3 [13] and horizons.

While previous work established a probabilistic forecast–retrieval decomposition for soil moisture estimation [14], the focus of this work is to offer a stronger and more principled fusion mechanism, leading to higher performance.

II. PROPOSED FRAMEWORK

A. Problem setup and probabilistic predictors

Let y_t denote soil moisture at time index t . We treat collocated SMAP L4 surface soil moisture as the prediction target for training and evaluation. For each site we evaluate at forecast horizon H over $t \in \{\ell + H - 1, \dots, N\}$ (i.e., targets that admit an H -step-ahead forecast origin within the record). The sequence is split chronologically into training, validation, and test subsets. For each horizon H , we generate H -step-ahead predictions and score only the H -ahead target at each forecast origin.

Retrieval leg: Let $\mathbf{x}_t^{(r)}$ be features constructed from auxiliary variables (including seasonal Fourier terms) and a flattened delay Doppler map (DDM). [15].

$$p_r(y_t | \mathbf{x}_t^{(r)}) = \mathcal{N}(\mu_{r,t}, \sigma_{r,t}^2). \quad (1)$$

We use NGBoost with a Normal output distribution, trained on the training subset by minimizing the negative log-likelihood to estimate $(\mu_{r,t}, \sigma_{r,t})$.

Forecasting leg: We fit a state-space ARIMAX model on the training subset [16, 17]. On validation and test, at each forecast origin we produce an H -step-ahead predictive distribution and then update the filtered state sequentially as new observations become available (with parameters fixed). This yields Gaussian forecasts

$$p_f(y_t | \mathcal{F}_{t-H}) = \mathcal{N}(\mu_{f,t}, \sigma_{f,t}^2), \quad (2)$$

we obtain the predictive standard deviation from the model’s $100(1 - \alpha)\%$ predictive interval and convert it to σ under a normality assumption.

B. Fusion rules

1) *W2 barycenter fusion:* Let $Q_r(u, t)$ and $Q_f(u, t)$ be the quantile functions of $p_r(\cdot | t)$ and $p_f(\cdot | t)$. The fused quantile function is

$$Q_c(u, t; w) = w Q_r(u, t) + (1 - w) Q_f(u, t), \quad u \in (0, 1), \quad (3)$$

which corresponds to weighted quantile averaging in one dimension. We use $\hat{y}_t = Q_c(0.5, t; w)$ as the point prediction and obtain prediction intervals by evaluating Q_c at desired quantiles.

Gaussian closed form: For Gaussian legs, $Q(u, t) = \mu_t + \sigma_t \Phi^{-1}(u)$, hence $p_c(\cdot | t) = \mathcal{N}(\mu_{c,t}, \sigma_{c,t}^2)$ with

$$\mu_{c,t}(w) = w \mu_{r,t} + (1 - w) \mu_{f,t}, \quad (4)$$

$$\sigma_{c,t}(w) = w \sigma_{r,t} + (1 - w) \sigma_{f,t}. \quad (5)$$

We note that the Gaussian parameterization is adopted for simplicity and closed-form fusion, however, the proposed W2 quantile-space fusion is distribution-agnostic and can be applied directly to empirical/non-Gaussian predictive quantiles without modification.

2) *L2 linear pooling:* We also consider the L2 linear opinion pool, defined as the mixture CDF

$$F_{L2}(y | t; w) = w F_r(y | t) + (1 - w) F_f(y | t), \quad (6)$$

where F_r and F_f are the Gaussian CDFs of the retrieval and forecasting legs. The point prediction is the mixture mean $\hat{y}_t = w \mu_{r,t} + (1 - w) \mu_{f,t}$.

C. Weight selection

For each site and horizon H , we fit a single scalar weight $w \in [0, 1]$ on the validation subset and fix it for test evaluation. Let $\mathcal{U} = \{0.05, 0.25, 0.50, 0.75, 0.95\}$.

Proposed (W2-qMSE): We select w by minimizing squared error in quantile space,

$$\hat{w}_{\text{qMSE}} = \arg \min_{w \in [0, 1]} \sum_{u \in \mathcal{U}} \sum_{t \in \text{val}} (y_t - Q_c(u, t; w))^2. \quad (7)$$

For Gaussian predictions, $Q_c(u, t; w) = \mu_{c,t}(w) + \sigma_{c,t}(w) z_u$ with $z_u = \Phi^{-1}(u)$. For symmetric $\mathcal{U}$ (so $\sum_{u \in \mathcal{U}} z_u = 0$),

$$\begin{aligned} \sum_{u \in \mathcal{U}} (y_t - Q_c(u, t; w))^2 &= |\mathcal{U}| (y_t - \mu_{c,t}(w))^2 \\ &\quad + \left(\sum_{u \in \mathcal{U}} z_u^2 \right) \sigma_{c,t}(w)^2, \end{aligned} \quad (8)$$

up to a constant factor. Thus qMSE trades off mean accuracy and dispersion of the fused predictive distribution.

Ablations (W2): We compare against (i) equal weights $w=0.5$, (ii) RMSE-tuned weights,

$$\hat{w}_{\text{RMSE}} = \arg \min_{w \in [0, 1]} \sum_{t \in \text{val}} (y_t - \mu_{c,t}(w))^2, \quad (9)$$

and (iii) CRPS-tuned weights using a one-dimensional search,

$$\hat{w}_{\text{CRPS}} = \arg \min_{w \in [0, 1]} \sum_{t \in \text{val}} \text{CRPS}(\mathcal{N}(\mu_{c,t}(w), \sigma_{c,t}^2(w)), y_t). \quad (10)$$

L2 weight : For the L2 pool (6), we select w by minimizing the mixture CRPS on validation (approximated numerically), then fix it for test. For test evaluation we refit each leg on train+val and keep the validation-selected weight fixed.

III. RESULTS

We consider CYGNSS observations after typical preprocessing, which includes: (i) converting BRCS to reflectivity, (ii) discarding DDMs flagged for low quality or with SNR $\text{SNR} < 10\text{ dB}$, and (iii) removing measurements acquired during precipitation events. Experiments follow the analysis-ready dataset construction of [14], restricted to sites jr1–jr3 because they provide substantially longer records, supporting more stable time series forecasting; the target y_t is the SMAP L4 surface soil moisture time series collocated to each site, with retrieval inputs given by collocated CYGNSS DDM observations, and auxiliary variables.

We use a Train/Validation/Test split, tuning fusion weights only on the validation set. After selecting w , we refit both component models on the Train and Validation Sets and evaluate on the held-out test set using the fixed validation-selected weight.

We evaluate lead- H performance for horizons $H \in \{2, 3, 4, 5\}$ on sites jr1–jr3, reporting RMSE of the fused median prediction, CRPS of the predictive distribution, and expected calibration error (ECE), computed as the mean absolute deviation between nominal and empirical coverage over the 50% and 90% central prediction intervals. We compare the retrieval leg, the forecasting leg, and W2 barycenter fusion under the weight selection rules discussed in Section II.

TABLE I
RESULTS FOR $H = 3$ (LOWER IS BETTER). BEST PER SITE AND METRIC IN BOLD.

Site	Method	w	RMSE	CRPS	ECE
jr1	Retrieval	–	0.0280	0.0153	0.077
	Forecast	–	0.0375	0.0193	0.057
	W2 (Equal)	0.5000	0.0290	0.0153	0.054
	W2 (qMSE)	0.6384	0.0278	0.0148	0.019
	W2 (RMSE)	0.4483	0.0296	0.0155	0.054
	W2 (CRPS)	0.3975	0.0303	0.0158	0.054
	L2	0.4800	0.0296	0.0151	0.080
jr2	Inv. Var. [14]	–	0.0295	0.0159	0.152
	Retrieval	–	0.0193	0.0106	0.102
	Forecast	–	0.0244	0.0117	0.127
	W2 (Equal)	0.5000	0.0184	0.0093	0.086
	W2 (qMSE)	0.6632	0.0178	0.0093	0.030
	W2 (RMSE)	0.3720	0.0194	0.0096	0.124
	W2 (CRPS)	0.3125	0.0200	0.0098	0.126
jr3	L2	0.2800	0.0195	0.0107	0.165
	Inv. Var. [14]	–	0.0211	0.0114	0.186
	Retrieval	–	0.0243	0.0132	0.129
	Forecast	–	0.0317	0.0161	0.097
	W2 (Equal)	0.5000	0.0247	0.0125	0.059
	W2 (qMSE)	0.5949	0.0240	0.0122	0.014
	W2 (RMSE)	0.2533	0.0275	0.0139	0.075
	W2 (CRPS)	0.2625	0.0274	0.0138	0.078
	L2	0.3300	0.0261	0.0135	0.103
	Inv. Var. [14]	–	0.0261	0.0142	0.200

Table I summarizes lead- H performance at $H = 3$. At jr2, fusion improves both RMSE and CRPS relative to either individual predictor. At jr1 and jr3, retrieval remains strong in RMSE, while W2 fusion is consistently competitive in CRPS. Adding ECE shows that fusion can also improve calibration: the q-MSE weight rule yields the lowest ECE across all three sites, indicating better agreement between nominal and empirical interval coverage.

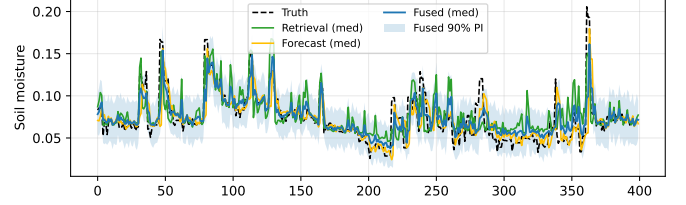
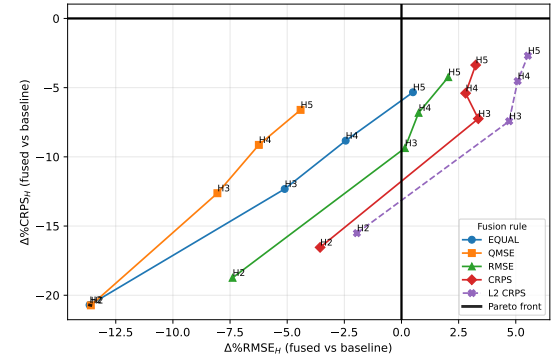
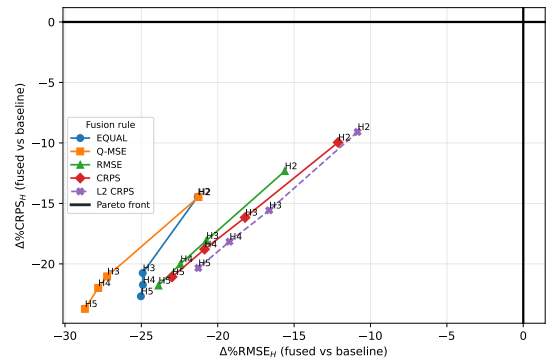


Fig. 2. JR2 example test window at lead- H showing median predictions for retrieval, forecast, and fusion, along with the fused 90% prediction interval.

Fig. 2 shows an example test window at jr2. The fused product lies between the retrieval and forecast medians and tracks the observed dynamics more stably than either leg alone. The fused prediction interval provides an interpretable uncertainty envelope around the fused trajectory.



(a) Baseline: retrieval



(b) Baseline: forecast

Fig. 4. JR2 horizon wise fusion gains in the $\Delta\%RMSE_H$ and $\Delta\%CRPS_H$ plane. Markers $H2$ – $H5$ denote forecast horizons and curves correspond to weight rules. The lower left quadrant indicates simultaneous improvement in RMSE and CRPS.

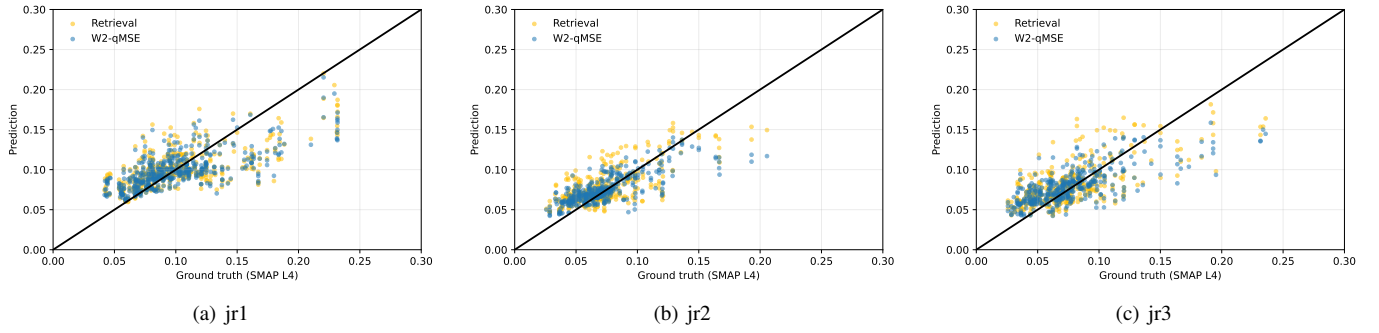


Fig. 3. Ground truth (SMAP L4) versus predictions at $H=3$ for sites jr1–jr3. Each panel shows the retrieval leg (yellow) and the W2-qMSE fused predictor (blue); the diagonal indicates perfect predictions ($\hat{y} = y$).

Fig. 4 summarizes horizon-wise fusion gains at jr2 in the $(\Delta\%RMSE_H, \Delta\%CRPS_H)$ plane, using retrieval (upper) and forecast (bottom) as separate baselines. Each point corresponds to one horizon under a weight rule. The horizontal axis is the percent change in point error relative to the baseline,

$$\Delta\%RMSE_H = 100 \frac{RMSE_{H,fused} - RMSE_{H,base}}{RMSE_{H,base}},$$

and the vertical axis is defined analogously for CRPS. Negative values indicate improvement over the baseline, so the lower-left quadrant represents simultaneous gains in both accuracy and probabilistic skill, while the upper-right quadrant indicates degradation in both. Curves connect horizons for the same weight rule to show how the RMSE–CRPS trade-off changes with lead time. The plotted Pareto front highlights the region where no other point improves both axes simultaneously. Relative to the forecast baseline, fusion points lie entirely in the improving region across horizons, whereas relative to the retrieval baseline CRPS gains remain consistent but RMSE gains are more sensitive to both horizon and weight selection.

Fig. 3 compares ground truth versus predictions at $H=3$ for the retrieval leg (green) and the W2-qMSE fused predictor (blue) across sites jr1–jr3. Points are plotted against the $y=\hat{y}$ diagonal, so perfect predictions lie on the black line; departures from the diagonal indicate error and systematic bias. Overall, the fused predictions are closer to the $y=\hat{y}$ trend than the retrieval-only predictions, with the largest deviations occurring at higher soil-moisture values where samples are sparse.

IV. CONCLUSIONS

In this work we explore the fusion of forecasting distribution from the SMAP L4 soil moisture time series with retrieval distribution from CYGNSS DDMs using a one-dimensional Wasserstein barycenter. Results indicate that fusion can improve calibration: under the qMSE rule, W2 fusion consistently reduced expected calibration error (ECE) relative to both legs, indicating better agreement between nominal and empirical interval coverage. Among the tested weight rules, the scalar qMSE choice was the most consistent, and it typically gave the lowest RMSE, CRPS and ECE, while alternatives such as

equal weights, RMSE-tuned weights, CRPS-tuned weights, L2 pooling, and inverse variance were less stable across settings. These experiments also demonstrate that calibration, sharpness, and point accuracy can move in different directions, so weight selection should match the target metric. In particular, RMSE- or CRPS-tuned weights do not necessarily yield the best calibrated predictive intervals, whereas qMSE explicitly trades off mean accuracy and dispersion and can yield more reliable uncertainty.

REFERENCES

- [1] R. H. Reichle, Q. Liu, R. D. Koster, W. T. Crow, G. J. M. De Lannoy *et al.*, “Version 4 of the SMAP level-4 soil moisture algorithm and data product,” 2019, reports mean latency of ~ 2.5 days; see NASA technical report / manuscript.
- [2] Z.-L. Li, P. Leng, C. Zhou, K.-S. Chen, F.-C. Zhou, and G.-F. Shang, “Soil moisture retrieval from remote sensing measurements: Current knowledge and directions for the future,” *Earth-Science Reviews*, vol. 218, p. 103673, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0012825221001744>
- [3] E. Babaeian, M. Sadeghi, S. B. Jones, C. Montzka, H. Vereecken, and M. Tuller, “Ground, proximal, and satellite remote sensing of soil moisture,” *Reviews of Geophysics*, vol. 57, no. 2, pp. 530–616, 2019. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018RG000618>
- [4] S. Yueh, R. Shah, J. Chaubell, A. Hayashi, X. Xu, and A. Colliander, “A semiempirical modeling of soil moisture, vegetation, and surface roughness impact on cygnss reflectometry data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, pp. 1–17, 11 2020.
- [5] T. Duan, A. Avati, D. Y. Ding, K. K. Thai, S. Basu, A. Y. Ng, and A. Schuler, “Ngboost: Natural gradient boosting for probabilistic prediction,” 2020. [Online]. Available: <https://arxiv.org/abs/1910.03225>
- [6] C. C. Chew and E. E. Small, “Comparison of CYGNSS reflectivity to SMAP soil moisture,” *Geophysical Research Letters*, 2018.

- [7] M. Agueh and G. Carlier, “Barycenters in the wasserstein space,” *SIAM Journal on Mathematical Analysis*, vol. 43, no. 2, pp. 904–924, 2011. [Online]. Available: <https://doi.org/10.1137/100805741>
- [8] S. S. Vallender, “Calculation of the wasserstein distance between probability distributions on the line,” *Theory of Probability and Its Applications*, 1974.
- [9] G. Peyré and M. Cuturi, “Computational optimal transport,” *Foundations and Trends in Machine Learning*, 2019.
- [10] T. Gneiting and R. Ranjan, “Combining predictive distributions,” *Electronic Journal of Statistics*, vol. 7, pp. 1747–1782, 2013.
- [11] S. G. Hall and J. Mitchell, “Combining density forecasts,” *International Journal of Forecasting*, vol. 23, no. 1, pp. 1–13, 2007.
- [12] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.
- [13] A. Melebari, R. Akbar, E. Hodges, D. Mckague, C. S. Ruf, A. R. Silva, M. Moghaddam, R. Shrestha, C. J. Lindsley, T. R. Walker, and B. E. Wilson, “Soil moisture profiles and temperature data from SoilSCAPE sites, version 2,” 2023, accessed: 2026-01-16.
- [14] G. Tsagkatakis, A. Melebari, J. D. Campbell, A. Kannan, P. Shokri, E. Hodges, M. Moghaddam, and P. Tsakalides, “Fusion of forecasting and retrieval for uncertainty aware soil moisture estimation,” in *2025 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Brisbane, Australia: IEEE, Aug. 2025.
- [15] G. Tsagkatakis, A. Melebari, R. Akbar, J. Campbell, E. Hodges, and M. Moghaddam, “Uncertainty quantification in machine learning based retrieval of soil moisture from gnss-r observations,” in *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2024, pp. 1824–1828.
- [16] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*, 2nd ed. Oxford University Press, 2012.
- [17] statsmodels developers, “SARIMAX — State Space Seasonal ARIMA with eXogenous regressors,” statsmodels documentation, 2026, accessed 2026-01-16. [Online]. Available: <https://www.statsmodels.org/stable/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html>