Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A comparative evaluation of Bayesian inference, fuzzy time series, and adaptive neuro-fuzzy inference systems for predicting the intensity of concurrent drought-pluvial events

Nicholas J. Sokol

nsokol@utk.edu

University of Tennessee Knoxville

Research Article

Keywords: Concurrent drought-pluvial events, Standardized Precipitation-Evapotranspiration Index (SPEI), Bayesian inference, Fuzzy time series, Adaptive neuro-fuzzy inference system (ANFIS), Model evaluation, Cross-validation, Reanalysis

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10680513/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Concurrent drought-pluvial events (CDPEs) are the simultaneous occurrence, in a single month, of anomalously dry conditions in one region of the contiguous United States and anomalously wet conditions in another, forming a spatial dry-wet dipole. This study evaluates three modeling methods that make few assumptions about data structure — Bayesian inference, fuzzy time series (FTS), and an adaptive neuro-fuzzy inference system (ANFIS) — for their ability to reproduce the intensity of 52 CDPEs, defined as the amplitude of the Standardized Precipitation-Evapotranspiration Index (SPEI) dipole. Each model is trained on seven synoptic predictors drawn independently from two reanalysis products, the North American Regional Reanalysis and the ERA-Interim reanalysis, each at three atmospheric levels (1000, 850, and 750 mb). Model skill is assessed with mean absolute error under event-grouped five-fold cross-validation and a held-out test split, against a climatological mean baseline and a persistence baseline, with pairwise Wilcoxon signed-rank tests and false-discovery-rate correction. No model reliably outperforms the baselines: of sixty pairwise comparisons, thirteen reach nominal significance but none survives correction, and persistence is the strongest predictor in every configuration. Predicted distributions explain this: Bayesian and ANFIS predictions collapse toward the SPEI mean rather than reproducing the bimodal dry/wet structure, while FTS reproduces the spread at the cost of accuracy. Variable-contribution analysis attributes most fitted signal to the meridional wind and moisture-flux terms, consistent with north-south moisture transport, though this does not yield skill above baseline. CDPE intensity is only weakly predictable from contemporaneous synoptic state at monthly resolution.

1 Introduction

Statistical and computational models are used routinely to analyze and predict drought and pluvial events, and the skill of any such model is shaped jointly by the volume and veracity of its input data (Essou et al. 2017; Schnase et al. 2017) and by the computational technique applied to those data (Choubin et al. 2016; Scully 2010). Assessing the respective contributions of the data source and the modeling method is therefore a prerequisite to understanding model performance and improving predictive capability. This study addresses both by evaluating three predictive modeling methods against explicit baselines, using two reanalysis products at three atmospheric levels, to reproduce the intensity of concurrent drought-pluvial events. The three methods — Bayesian inference, fuzzy time series (FTS), and an adaptive neuro-fuzzy inference system (ANFIS) — make few explicit assumptions about the underlying data distribution, instead learning empirical relationships between predictor and response variables (Cooley et al. 2007; Bacanli et al. 2009; Lukoševičius and Jaeger 2009). That property makes them attractive for a phenomenon as irregular as a CDPE.

A concurrent drought-pluvial event is the simultaneous occurrence, in a single month, of a severe drought in one region of the contiguous United States (CONUS) and a severe pluvial in another, forming a spatial dry-wet dipole. Such events arise from a combination of oceanic forcing (van der Schrier and Barkmeijer 2007), internal atmospheric variability (Cook et al. 2011), synoptic feedbacks (Fink et al. 2004), and land-atmosphere interactions (Dai 2011), and their intensity provides a compact measure of

the strength of the dry-wet dipole. Whether that intensity can be predicted from the contemporaneous synoptic state of the atmosphere, and whether the choice of model or reanalysis product changes the answer, is the question this study addresses.

1.1 Bayesian inference for hydroclimatic prediction

Bayesian inference estimates the probability of an event given its prior occurrence and the influence of indicator variables, where the indicators are the independent variables believed to influence a given event. Several studies have used Bayesian inference to analyze and model severe drought and precipitation, most often to estimate return intervals (Wu et al. 2014; Kwon et al. 2016; Sikorska and Seibert 2018). Applied to the atmospheric drivers of drought, Bayesian approaches have identified the variables that contribute most to increasing event frequency and intensity, frequently implicating soil-moisture modifications that enhance drought (Wu et al. 2014; Kwon et al. 2016). A recognized strength of the method is that it returns an explicit measure of uncertainty, which makes it informative for spatially diverse hydrological extremes where the contribution of individual variables across a distribution is of interest (Cooley et al. 2007; Zhang et al. 2015). Because it can estimate the probability of variables influencing events that are difficult to model through traditional means (Sikorska and Seibert 2018), Bayesian inference is a natural candidate for investigating a complex spatial phenomenon such as a CDPE.

1.2 Fuzzy time series

Fuzzy logic is a form of probabilistic modeling that assigns degrees of membership based on the observed range of a variable, an idea introduced by Zadeh (1965), who showed that verbal descriptions of ranges within a dataset could be represented by a strength of likeness to a given linguistic label (e.g., cold, cool, warm, hot). Fuzzy time series extends this idea by using a universe of discourse to build fuzzy relationship sets capable of predicting future values (Song and Chissom 1993). The earliest applications predicted university enrollment from prior-year values (Song and Chissom 1991) and, subsequently, temperature (Chen and Hwang 2000; Song and Chissom 1993). Later refinements addressed the scaling of time intervals (Huarng 2001) and the construction of multivariate fuzzy relations (Tsaur et al. 2005). Fuzzy approaches are often regarded as well suited to hydrological modeling because their loosely bounded classes mimic the graded transitions of natural systems (Hundecha et al. 2001; Bogardi et al. 2003; Oladokun et al. 2017). The capacity of fuzzy time series to represent complex, noisy data through fuzzy relationships makes it a reasonable candidate for CDPE prediction, given the chaotic nature of the data underlying these events.

1.3 Adaptive neuro-fuzzy inference systems

The adaptive neuro-fuzzy inference system (ANFIS) is a fuzzy method that removes the subjectivity of defining membership classes by learning them with a neural network. Although relatively new to climatology and weather modeling, ANFIS has been applied widely (Chaudhuri and Chattopadhyay 2013; Polykretis et al. 2019). Araya-Muñoz et al. (2017) used spatial weights to define membership functions in

an ANFIS study of natural-hazard uncertainty in Concepción, Chile, reporting strong agreement between training and test partitions. Bacanli et al. (2009) combined station precipitation with the Standardized Precipitation Index (SPI) to build an ANFIS drought forecast and reported high accuracy relative to competing methods, and Mekanik et al. (2016) found that ANFIS outperformed purely traditional models for seasonal precipitation on root-mean-square error and mean absolute error (MAE). Choubin et al. (2016) reached the same conclusion in a comparison against multiple linear regression and a multilayer perceptron. Other applications have reported large accuracy gains but also limitations in topographically complex terrain and for events dominated by internal atmospheric variability (Akrami et al. 2013; Nguyen et al. 2015; Gocić et al. 2015), a limitation common to forecasting models more broadly (Hoerling et al. 2012; Cook et al. 2016; Deser et al. 2018).

Most prior work on these and related events has relied on traditional model-comparison or regression techniques to relate event occurrence to specific variables. The flexible computational methods described above may improve both predictive accuracy and understanding of which variables most influence these events. Accordingly, this study has three objectives: (1) to quantify how well Bayesian inference, fuzzy time series, and ANFIS reproduce the intensity of CDPEs, evaluated against a climatological baseline and a persistence baseline; (2) to characterize the spread in error and uncertainty among the three models and across data configurations; and (3) to determine which predictor variables contribute most to each model's output. Section 2 describes the data, Sect. 3 the methods, Sect. 4 the results and their interpretation, and Sect. 5 the conclusions.

2 Data

The 1-month Standardized Precipitation-Evapotranspiration Index (SPEI) is used to identify drought and pluvial events and to define event intensity. SPEI was developed to improve upon SPI by incorporating a measure of potential evapotranspiration (Vicente-Serrano et al. 2010), which yields a more complete representation of drought severity under warming than a precipitation-only index (Beguería et al. 2014). SPEI is drawn from the Global SPEI Database (SPEIbase; Beguería et al. 2014) over the CONUS. For each identified CDPE, intensity is sampled at fixed positions across the dry and wet zones — the most extreme dry pixel, a moderate dry pixel, a moderate wet pixel, and the most extreme wet pixel — together with an event-mean value, giving a consistent set of sampling points per event across a total of 52 events.

The North American Regional Reanalysis (NARR; Mesinger et al. 2006) is an assimilated gridded product providing 32-km resolution over North America from the surface (1000 mb) to the top of the atmosphere (10 mb), assimilating radiosonde, dropsonde, aircraft, satellite, and surface observations. Monthly means are used, matching the monthly time scale of the SPEI-based events. NARR has been used extensively to study drought (Barandiaran et al. 2013; Mo et al. 2011) and precipitation (Ruiz-Barradas and Nigam 2006; Bukovsky and Karoly 2007). It is valued for its resolution but criticized for over- and under-estimation of precipitation and moisture over heterogeneous terrain (Bukovsky and Karoly 2007); for the synoptic variables of interest here it has been shown to have high veracity for event observation (Trubilowicz et al. 2016).

The ERA-Interim reanalysis (ERAI; Dee et al. 2011) is a global gridded product spanning 1979 onward at approximately 0.75° (~ 80–130 km) resolution and six-hourly output, from 1000 to 100 mb, constructed from land, radiosonde, aircraft, and satellite observations together with sea-surface temperatures (Simmons et al. 2007). ERAI is distinguished from NARR by its four-dimensional variational assimilation, and although its spatial resolution is coarser it has repeatedly been shown to represent synoptic-scale patterns with high fidelity (Decker et al. 2012), while performing less well over complex terrain and for features with short periodicities such as the low-level jet (Berg et al. 2015). Feeding each product independently to each model, at three atmospheric levels (1000, 850, and 750 mb), quantifies the respective roles of data source and modeling method. These levels bracket the surface and lower-tropospheric flow most relevant to moisture transport.

3 Methods

3.1 Models and predictors

Bayesian inference, fuzzy time series, and ANFIS are each used to hindcast CDPE intensity, defined as the SPEI amplitude of the dry-wet dipole sampled at the fixed within-event positions described in Sect. 2. Each model receives the same seven predictors — air temperature, specific humidity, zonal wind, meridional wind, zonal moisture flux, meridional moisture flux, and pressure — chosen because they have been identified as most influential in the formation and maintenance of drought and pluvial conditions (Seager and Hoerling 2014). Each model is run separately for every combination of reanalysis product (NARR, ERAI) and atmospheric level (1000, 850, 750 mb), giving six data configurations per model and eighteen model runs in total.

Bayesian inference estimates the probability of a CDPE intensity value conditional on the predictor magnitudes. It requires a prior distribution, which may be based on prior knowledge, on previous analysis, or on Jeffreys' prior (Firth 1993), which is constructed from the statistical variance of the observed values without requiring an externally specified prior. The prior used here is formed from the average of Jeffreys' prior and the empirical probability of occurrence derived from the identified events. The model is expressed as

$$P(\theta \mid D) = [\, P(D \mid \theta) \times P(\theta) \,] \,/\, P(D), \quad (1)$$

where P(θ) is the prior, P(D | θ) is the likelihood of the observed data given the prior distribution, and P(D) is the evidence — the probability of the data summed across all values of θ, weighted by the subjective belief in those values. Each predictor is treated as a probability density over the frequency or magnitude observed in conjunction with a given CDPE intensity, and the model outputs intensity values at each within-event sampling position.

Fuzzy time series builds fuzzy relations among values in a series to predict the next observation (Song and Chissom 1991, 1993). A fuzzy set Y(t) describes a subset of the universe of discourse — the total range of values — for a given variable, and fuzzy relations are formed through the min-max composition

$f_i(t) = f_i(t-1) \circ R_{ij}(t, t-1)$, (2)

where ∘ denotes the min-max composition and R is the relation between elements of the set (Song and Chissom 1993). As values in one list co-occur more frequently with values in another, the relation between them strengthens, and these relations are combined with the fuzzy sets to predict the value at the next step. Predictors are treated as degrees of truth — a measure of how strongly a value falls within a range that contributes to the observed response — and the model reproduces SPEI intensity for the events.

ANFIS combines fuzzy logic with a neural network that learns the membership functions, removing the researcher's subjectivity in specifying membership weights (Azeem et al. 2000). The Takagi-Sugeno formulation is used, in which fuzzy rules take the form

if x is A and y is B, then $f_1 = p_1x + q_1y + r_1$, (3)

with the network arranged in layers that assign weights, classify inputs through membership functions, and produce an output (Wang and Lee 2002). Predictors are not transformed but assigned a weighting scheme reflecting their observed influence on the response, and the model is trained over a fixed number of epochs.

3.2 Baselines

Because a flexible model can appear skillful simply by reproducing the central tendency of a target, all three models are evaluated against two baselines. The climatological baseline predicts every held-out event at the mean SPEI intensity of the training set, capturing no event-to-event variation. The persistence baseline predicts each event from the most recent prior observation, capturing whatever short-memory structure exists in the intensity series without using any synoptic predictor. A model is considered skillful only if it improves on both references: a model that does not beat the mean baseline has not learned to discriminate strong events from weak ones, and a model that does not beat persistence has not extracted information from the synoptic predictors beyond what the previous value supplies.

3.3 Cross-validation and held-out testing

Two evaluation protocols are used. First, grouped five-fold cross-validation constrains all sampling points belonging to a single CDPE to the same fold. This grouping is necessary because multiple sampling points are drawn from each event; a random point-level split would place points from the same event in both the training and validation partitions, leaking event-specific information across the split and inflating apparent skill. Grouping by event yields an out-of-sample error estimate, reported as the mean and standard deviation of MAE across folds. Second, a temporally disjoint test half holds out the later events, so that no test event informs training. Error is measured with MAE, which is directly interpretable in SPEI units and less sensitive to a small number of large residuals than the mean squared error (Willmott and Matsuura 2005); both are reported.

3.4 Significance testing

For each data configuration, the absolute-error distributions of every pair of predictors — each model against each baseline, and the models against one another — are compared using the Wilcoxon signed-rank test (Wilcoxon 1945), a paired, non-parametric test appropriate for error distributions that are not normally distributed and that are paired by held-out event. Because this produces sixty pairwise tests across the six configurations, the false-discovery rate across the full family of tests is controlled using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995), and both raw and FDR-adjusted p-values are reported. Only FDR-adjusted significance is treated as evidence of a real skill difference, since a count of raw-significant results at $p < 0.05$ would be expected to include several false positives given the number of comparisons.

3.5 Variable contribution

For the Bayesian and ANFIS models, each predictor's contribution to the fitted output is quantified as the share of predicted variance attributable to that predictor within a given configuration. Fuzzy time series is omitted from this analysis because its rule-based formulation does not yield a comparable per-variable importance metric. Contributions are reported for the NARR 1000-mb configuration, with other configurations summarized in the text where they differ.

3.6 Spatial reconstruction

To assess whether the predicted intensities are sufficient to recover the spatial structure of an event, the predicted values are mapped back to their original sampling locations for a representative event (March 1980) and interpolated by nearest neighbor to reconstruct the dry-wet dipole. This reconstruction uses only the sampled points, not a pixelwise prediction over the full grid, and serves as a qualitative check on whether a model reproduces the sign and rough placement of the dipole.

3.7 Implementation

All models, baselines, cross-validation, and statistical tests were implemented in Python using standard scientific-computing libraries.

4 Results and discussion

4.1 Model performance against baselines

Held-out test MAE values are given in Table 1 and Fig. 1b. The climatological mean baseline has an MAE of 1.72 SPEI units in every configuration, by construction, and the persistence baseline improves on it to 1.38 units. The best learned model in each configuration ranges from 1.38 to 1.87 units: FTS reaches 1.41 at NARR 1000 mb and matches persistence at 1.38 at ERAI 750 mb, and the Bayesian model reaches 1.48 at NARR 750 mb, but in no configuration does any model fall clearly below the persistence

value. ANFIS produces held-out MAE between 1.64 and 1.75 units, at or above the mean baseline in several configurations and never below persistence.

The grouped cross-validation results (Fig. 1a) are consistent, with an additional contrast in variance. The Bayesian and ANFIS models have low fold-to-fold standard deviations (typically 0.02–0.16 units), indicating stable performance, whereas the FTS standard deviation reaches 0.30 units, indicating accuracy that depends strongly on which events fall in the training partition. The lowest cross-validation means belong to the Bayesian model at ERAI 750 mb (1.29 units) and NARR 750 mb (1.39 units) and to FTS at NARR 1000 mb (1.37 units), but the FTS values carry wide error bars that overlap the other models. In no configuration does a learned model's cross-validation MAE fall far enough below the 1.38-unit persistence value to indicate skill once its variability is accounted for.

Table 1
Held-out test mean absolute error (MAE, SPEI units) for the two baselines and the three models, by reanalysis product and atmospheric level (n = 52 held-out events per configuration)

Data / level	Baseline (mean)	Persistence	Bayesian	FTS	ANFIS
NARR 1000 mb	1.72	1.38	1.64	1.41	1.70
NARR 850 mb	1.72	1.38	1.63	1.51	1.66
NARR 750 mb	1.72	1.38	1.48	1.80	1.64
ERAI 1000 mb	1.72	1.38	1.72	1.87	1.75
ERAI 850 mb	1.72	1.38	1.69	1.62	1.66
ERAI 750 mb	1.72	1.38	1.51	1.38	1.65

MAE values are rounded to two decimal places

The pairwise significance results (Fig. 1d) show no reliable skill difference among the learned models or between any model and the mean baseline. Of the sixty Wilcoxon tests, thirteen reach nominal significance at $p < 0.05$, dominated by two effects: persistence versus the mean baseline (raw $p = 0.008$ in all six configurations) and several ANFIS-versus-persistence and Bayesian-versus-persistence comparisons in which persistence is the better predictor. After Benjamini-Hochberg correction across the full family of tests, no comparison remains significant at the 0.05 level; the smallest FDR-adjusted p-value is 0.073, for the persistence-versus-baseline contrasts. The strongest signal in the comparison is therefore the advantage of persistence over the climatological mean.

4.2 Predicted intensity distributions

The error results follow from the shape of the predicted distributions (Fig. 2). The observed SPEI intensities are strongly bimodal, clustering near − 1.7 for the dry poles and near + 1.6 for the wet poles, with little mass near zero, a consequence of the CDPE definition, which selects extreme dry and wet cells. The Bayesian model predicts almost every event near the SPEI mean, forming a single tall spike

close to zero, which is why its MAE hovers near the mean baseline. ANFIS behaves similarly, concentrating its predictions in a narrow band around zero with only slight spread toward the observed modes. Fuzzy time series is the only model that reproduces a spread of predicted values across the dry and wet ranges, and at NARR and ERAI 750 mb it partially recovers the two observed modes, but with enough scatter into the intermediate range that its overall error is no lower than that of the mean-reverting models. The models thus face a trade-off between matching the mean, which produces low bias against a bimodal target but no discrimination, and matching the spread, which produces some discrimination but high variance; none resolves the trade-off in a way that beats persistence.

4.3 Predicted and observed intensity series

Tracing predicted and observed intensity event by event across all six configurations (Fig. 3) reinforces the distributional result. The observed series (black) oscillates strongly between the dry and wet extremes, while the Bayesian (red) and ANFIS (blue) predictions remain compressed near the centre of the range, tracking the sign of the larger swings only weakly and rarely reaching the observed amplitude. The fuzzy time series predictions (green) swing more widely and occasionally overshoot the observed values, attempting the full range of the target without tracking individual events accurately. No model follows the observed series closely across timesteps, and the pattern is consistent across reanalysis products and atmospheric levels.

4.4 Variable contribution

Within the fitted models, the predictor contributions are dominated by the meridional terms (Fig. 1c; Table 2). For the Bayesian model at NARR 1000 mb, meridional moisture flux (27.1%) and meridional wind (25.4%) together account for more than half of the explained variance, followed by specific humidity (15.6%), air temperature (12.8%), and zonal moisture flux (10.3%), with pressure (8.7%) and the zonal terms contributing little. ANFIS concentrates its weight even more sharply on the same two meridional predictors, assigning roughly 61% to meridional moisture flux and 39% to meridional wind and effectively zero to the remaining five variables. The prominence of the north-south moisture-transport terms is physically coherent: the dry-wet dipoles that define CDPEs are organized by meridional advection of moisture from a pluvial region toward or away from a drought region. This coherent attribution nonetheless coexists with an absence of skill above baseline; the models identify physically reasonable predictors but cannot convert them into accurate event-level intensity predictions at this resolution.

Table 2
Percentage contribution of each predictor to model output for the NARR 1000-mb configuration. Fuzzy time series is omitted because its rule-based structure yields no comparable per-variable importance metric

Predictor	Bayesian (%)	ANFIS (%)
Air temperature (T)	12.8	0.0
Specific humidity (q)	15.6	0.0
Zonal wind (u)	0.0	0.0
Meridional wind (v)	25.4	39.2
Zonal moisture flux (uq)	10.3	0.0
Meridional moisture flux (vq)	27.1	60.8
Pressure (P)	8.7	0.0

Values are the percentage of fitted variance attributable to each predictor within the configuration; columns need not sum to exactly 100 due to rounding

4.5 Spatial reconstruction

Mapping the predicted intensities back to their sampling locations for the March 1980 event (Fig. 4) shows the practical consequence of the distributional differences. The observed field is a clear dipole, with a broad wet anomaly across the southern and central United States and drier conditions to the north. Fuzzy time series produces the closest reconstruction, recovering both a wet region in the south-central United States and drier conditions elsewhere, though with weaker amplitudes than observed. ANFIS recovers the sign of the wet anomaly but with strongly muted intensity, consistent with its mean-reverting predictions. The Bayesian reconstruction renders both poles as near-neutral and does not express the dipole. No model reproduces the observed amplitude, consistent with the distributional results; the reconstruction confirms that FTS and, to a lesser extent, ANFIS capture dipole placement, while intensity itself is not well predicted.

4.6 Interpretation and limitations

Taken together, the results indicate that CDPE intensity is only weakly predictable from the contemporaneous synoptic state at the monthly resolution and event-buffer scale examined here. The mechanism is visible in the predicted distributions: because the CDPE definition selects extreme dry and wet cells, the target intensity distribution is strongly bimodal with no central mass, and a model minimizing mean error against such a target is pulled toward predicting the mean, incurring a moderate error on every event rather than a large error on the events it misclassifies. The Bayesian and ANFIS models follow this route, while fuzzy time series attempts the full range and converts that ambition into scatter; the two strategies reach similar overall error, and neither beats persistence.

The variable-contribution analysis is physically coherent even though it does not produce skill. Both the Bayesian and ANFIS models place the majority of their weight on the meridional wind and meridional moisture-flux terms, and ANFIS does so almost exclusively, consistent with the role of north-south moisture transport in organizing the dry and wet poles: anomalous meridional flux carries moisture toward the pluvial region while its suppression starves the drought region. That the models converge on physically reasonable predictors yet cannot predict intensity accurately suggests that the limiting factor is the information content of the monthly, contemporaneous synoptic state, rather than the identity of the predictors.

Several limitations bear on interpretation and indicate directions for improvement. First, seasonality is not controlled: the most heavily weighted predictors carry a strong seasonal cycle, and pooling events across all seasons forces each model to reconcile winter and summer regimes within a single set of relations or membership functions, so fitting season-specific models is a natural next step. Second, the models use only the CDPE month; a CDPE reflects moisture anomalies that rarely develop within a single month, so predictors aggregated over the preceding months of circulation may carry information that the contemporaneous state does not. Third, the sample is small (52 events), which limits model complexity and widens the confidence intervals on every skill estimate, an intrinsic consequence of the rarity of the phenomenon. Fourth, the reconstruction is limited to sampled points rather than a pixelwise spatial prediction, so it confirms dipole placement but not full spatial structure. Finally, only the linear and rule-based relationships intrinsic to these three methods were tested; the results do not exclude the possibility that another model could predict CDPE intensity, only that these three, as configured, do not beat persistence at monthly resolution.

5 Conclusion

Three low-assumption computational models — Bayesian inference, fuzzy time series, and ANFIS — were evaluated for their ability to reproduce the intensity of 52 concurrent drought-pluvial events over the contiguous United States, using two reanalysis products at three atmospheric levels and comparing every model against a climatological mean baseline and a persistence baseline under event-grouped cross-validation and a held-out test split. No model reliably outperforms the baselines: across sixty pairwise Wilcoxon comparisons none survives false-discovery-rate correction, and persistence is the strongest predictor in every configuration. The Bayesian and ANFIS models collapse their predictions toward the SPEI mean rather than reproducing the bimodal dry/wet structure of the observations, while fuzzy time series reproduces the spread at the cost of accuracy.

The models converge on a physically coherent set of predictors, weighting the meridional wind and meridional moisture-flux terms most heavily, consistent with the role of north-south moisture transport in organizing the dry-wet dipole. That this attribution does not translate into predictive skill points to the information content of the monthly, contemporaneous synoptic state as the limiting factor rather than the choice of predictors or model. The most promising routes to improvement are to control for seasonality by fitting season-specific models, to incorporate multi-month antecedent circulation rather

than the event month alone, and to enlarge the sample where possible. CDPE intensity, as represented by these data and methods, is only weakly predictable from the contemporaneous synoptic state at monthly resolution.

Declarations

Funding

None.

Competing interests

None.

Data availability

The Standardized Precipitation-Evapotranspiration Index data are available from the Global SPEI Database (SPEIbase) at https://spei.csic.es/database.html. The North American Regional Reanalysis is available from the NOAA Physical Sciences Laboratory at https://psl.noaa.gov/data/gridded/data.narr.html, and the ERA-Interim reanalysis is available from the European Centre for Medium-Range Weather Forecasts at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim. The identified event set and the model output analysed during the current study are available from the author on reasonable request.

Code availability

The custom Python code used to implement the models, baselines, cross-validation, and statistical tests is available from the author on reasonable request.

Ethics approval

Not applicable. This study did not involve human participants or animals.

References

1. Akrami SA, El-Shafie A, Jaafar O (2013) Improving rainfall forecasting efficiency using modified adaptive neuro-fuzzy inference system (MANFIS). Water Resour Manag 27(9):3507–3523. https://doi.org/10.1007/s11269-013-0361-9
2. Araya-Muñoz D, Metzger MJ, Stuart N, Wilson AMW, Alvarez L (2017) Assessing urban adaptive capacity to climate change. J Environ Manage 183:314–324. https://doi.org/10.1016/j.jenvman.2016.09.060

3. Azeem MF, Hanmandlu M, Ahmad N (2000) Generalization of adaptive neuro-fuzzy inference systems. IEEE Trans Neural Netw 11(6):1332–1346. https://doi.org/10.1109/72.883425
4. Bacanli UG, Firat M, Dikbas F (2009) Adaptive neuro-fuzzy inference system for drought forecasting. Stoch Environ Res Risk Assess 23(8):1143–1154. https://doi.org/10.1007/s00477-008-0288-5
5. Barandiaran D, Wang SY, Hilburn K (2013) Observed trends in the Great Plains low-level jet and associated precipitation changes in relation to recent droughts. Geophys Res Lett 40(23):6247–6251. https://doi.org/10.1002/2013GL058296
6. Beguería S, Vicente-Serrano SM, Reig F, Latorre B (2014) Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. Int J Climatol 34(10):3001–3023. https://doi.org/10.1002/joc.3887
7. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
8. Berg LK, Riihimaki LD, Qian Y, Yan H, Huang M (2015) The low-level jet over the southern Great Plains determined from observations and reanalyses and its impact on moisture transport. J Clim 28(17):6682–6706. https://doi.org/10.1175/JCLI-D-14-00719.1
9. Bogardi I, Bardossy A, Duckstein L, Pongracz R (2003) Fuzzy logic in hydrology and water resources. In: Demicco RV, Klir GJ (eds) Fuzzy logic in geology. Academic, Amsterdam, pp 153–190
10. Bukovsky MS, Karoly DJ (2007) A brief evaluation of precipitation from the North American Regional Reanalysis. J Hydrometeorol 8(4):837–846. https://doi.org/10.1175/JHM595.1
11. Chaudhuri S, Chattopadhyay S (2013) Neuro-computing based short range prediction of monsoon rainfall. Meteorol Atmos Phys 121(3–4):143–152. https://doi.org/10.1007/s00703-013-0261-9
12. Chen SM, Hwang JR (2000) Temperature prediction using fuzzy time series. IEEE Trans Syst Man Cybern B 30(2):263–275. https://doi.org/10.1109/3477.836375
13. Choubin B, Khalighi-Sigaroodi S, Malekian A, Kişi Ö (2016) Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. Hydrol Sci J 61(6):1001–1009. https://doi.org/10.1080/02626667.2014.966721
14. Cook BI, Cook ER, Anchukaitis KJ, Seager R, Miller RL (2011) Forced and unforced variability of twentieth century North American droughts and pluvials. Clim Dyn 37(5–6):1097–1110. https://doi.org/10.1007/s00382-010-0897-9
15. Cook BI, Mankin JS, Anchukaitis KJ (2016) Climate change and drought: from past to future. Curr Clim Change Rep 2(4):164–179. https://doi.org/10.1007/s40641-018-0093-2
16. Cooley D, Nychka D, Naveau P (2007) Bayesian spatial modeling of extreme precipitation return levels. J Am Stat Assoc 102(479):824–840. https://doi.org/10.1198/016214506000000780
17. Dai A (2011) Drought under global warming: a review. WIREs Clim Change 2(1):45–65. https://doi.org/10.1002/wcc.81

18. Decker M, Brunke MA, Wang Z, Sakaguchi K, Zeng X, Bosilovich MG (2012) Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. J Clim 25(6):1916–1944. https://doi.org/10.1175/JCLI-D-11-00004.1
19. Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597. https://doi.org/10.1002/qj.828
20. Deser C, Phillips AS, Alexander MA (2018) How well do we know ENSO's climate impacts over North America, and how do we evaluate models accordingly? J Clim 31(13):4991–5014. https://doi.org/10.1175/JCLI-D-17-0783.1
21. Essou GRC, Sabarly F, Lucas-Picher P, Brissette F, Poulin A (2017) Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling? J Hydrometeorol 17(7):1929–1950. https://doi.org/10.1175/JHM-D-15-0138.1
22. Fink AH, Brücher T, Krüger A, Leckebusch GC, Pinto JG, Ulbrich U (2004) The 2003 European summer heatwaves and drought — synoptic diagnosis and impacts. Weather 59(8):209–216. https://doi.org/10.1256/wea.73.04
23. Firth D (1993) Bias reduction of maximum likelihood estimates. Biometrika 80(1):27–38. https://doi.org/10.1093/biomet/80.1.27
24. Gocić M, Motamedi S, Shamshirband S, Petković D, Ch S, Hashim R, Arif M (2015) Soft computing approaches for forecasting reference evapotranspiration. Comput Electron Agric 113:164–173. https://doi.org/10.1016/j.compag.2015.02.010
25. Hoerling M, Eischeid J, Perlwitz J, Quan X, Zhang T, Pegion P (2012) On the increased frequency of Mediterranean drought. J Clim 25(6):2146–2161. https://doi.org/10.1175/JCLI-D-11-00296.1
26. Huarng K (2001) Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets Syst 123(3):387–394. https://doi.org/10.1016/S0165-0114(00)00057-9
27. Hundecha Y, Bardossy A, Theisen HW (2001) Development of a fuzzy logic-based rainfall-runoff model. Hydrol Sci J 46(3):363–376. https://doi.org/10.1080/02626660109492832
28. Kwon HH, Lall U, Kim SJ (2016) The unusual 2013–2015 drought in South Korea in the context of a multicentury precipitation record. Geophys Res Lett 43(15):8534–8541. https://doi.org/10.1002/2016GL070270
29. Lukoševičius M, Jaeger H (2009) Reservoir computing approaches to recurrent neural network training. Comput Sci Rev 3(3):127–149. https://doi.org/10.1016/j.cosrev.2009.03.005
30. Mekanik F, Imteaz MA, Talei A (2016) Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. Clim Dyn 46(9–10):3097–3111. https://doi.org/10.1007/s00382-015-2755-2
31. Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W et al (2006) North American Regional Reanalysis. Bull Am Meteorol Soc 87(3):343–360. https://doi.org/10.1175/BAMS-87-3-343
32. Mo KC, Chen LC, Shukla S (2011) Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction? Geophys Res Lett 39(23):L23703. https://doi.org/10.1029/2012GL053598

33. Nguyen P, Nguyen H, Vo D (2015) Drought forecasting using ANFIS — a case study in drought-prone area of Vietnam. Paddy Water Environ 13(2):145–155. https://doi.org/10.1007/s10333-013-0402-2
34. Oladokun VO, Proverbs D, Lamond J (2017) Measuring flood resilience of communities: a fuzzy logic approach. Int J Build Pathol Adapt 35(5):470–487. https://doi.org/10.1108/IJBPA-12-2016-0029
35. Polykretis C, Ferentinou M, Chalkias C (2019) A comparative study of landslide susceptibility mapping using landslide susceptibility index and artificial neural networks. Bull Eng Geol Environ 78(3):1219–1237. https://doi.org/10.1007/s10064-017-1125-1
36. Ruiz-Barradas A, Nigam S (2006) Great Plains hydroclimate variability: the view from North American Regional Reanalysis. J Clim 19(12):3004–3010. https://doi.org/10.1175/JCLI3768.1
37. Schnase JL, Carroll ML, Weber KT, Brown ME, Gill RL, Wooten M et al (2017) Toward a Monte Carlo approach to selecting climate variables in MaxEnt. PLoS ONE 12(10):e0186190. https://doi.org/10.1371/journal.pone.0186190
38. Scully ME (2010) Wind modulation of dissolved oxygen in Chesapeake Bay. Estuaries Coasts 33(5):1164–1175. https://doi.org/10.1007/s12237-010-9319-9
39. Seager R, Hoerling M (2014) Atmosphere and ocean origins of North American droughts. J Clim 27(12):4581–4606. https://doi.org/10.1175/JCLI-D-13-00329.1
40. Sikorska AE, Seibert J (2018) Value of different precipitation data for flood prediction in an Alpine catchment: a Bayesian approach. J Hydrol 556:961–971. https://doi.org/10.1016/j.jhydrol.2016.06.031
41. Simmons A, Uppala S, Dee D, Kobayashi S (2007) ERA-Interim: new ECMWF reanalysis products from 1989 onwards. ECMWF Newsl 110:25–35. https://doi.org/10.21957/pocnex23c6
42. Song Q, Chissom BS (1991) Fuzzy time series and its models. Fuzzy Sets Syst 54(3):269–277. https://doi.org/10.1016/0165-0114(93)90372-O
43. Song Q, Chissom BS (1993) Forecasting enrollments with fuzzy time series — part I. Fuzzy Sets Syst 54(1):1–9. https://doi.org/10.1016/0165-0114(93)90355-L
44. Trubilowicz JW, Shea JM, Jost G (2016) Suitability of North American Regional Reanalysis (NARR) output for hydrologic modelling and analysis in mountainous terrain. Hydrol Process 30(13):2332–2347. https://doi.org/10.1002/hyp.10795
45. Tsaur RC, Yang JCO, Wang HF (2005) Fuzzy relation analysis in fuzzy time series model. Comput Math Appl 49(4):539–548. https://doi.org/10.1016/j.camwa.2004.07.014
46. van der Schrier G, Barkmeijer J (2007) North American 1818–1824 drought and 1825–1840 pluvial and their possible relation to the atmospheric circulation. J Geophys Res Atmos 112:D13102. https://doi.org/10.1029/2007JD008429
47. Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the Standardized Precipitation Evapotranspiration Index. J Clim 23(7):1696–1718. https://doi.org/10.1175/2009JCLI2909.1

48. Wang JS, Lee CSG (2002) Self-adaptive neuro-fuzzy inference systems for classification applications. IEEE Trans Fuzzy Syst 10(6):790–802. https://doi.org/10.1109/TFUZZ.2002.805880
49. Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1(6):80–83. https://doi.org/10.2307/3001968
50. Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 30(1):79–82. https://doi.org/10.3354/cr030079
51. Wu J, Chen X, Yao H, Gao L, Chen Y, Liu M (2014) Non-linear relationship of hydrological drought responding to meteorological drought and impact of a large reservoir. J Hydrol 551:495–507. https://doi.org/10.1016/j.jhydrol.2017.06.029
52. Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353. https://doi.org/10.1016/S0019-9958(65)90241-X
53. Zhang Q, Gu X, Singh VP, Chen X (2015) Evaluation of ecological instream flow using multiple ecological indicators with consideration of hydrological alterations. J Hydrol 529:711–722. https://doi.org/10.1016/j.jhydrol.2015.08.066

Figures

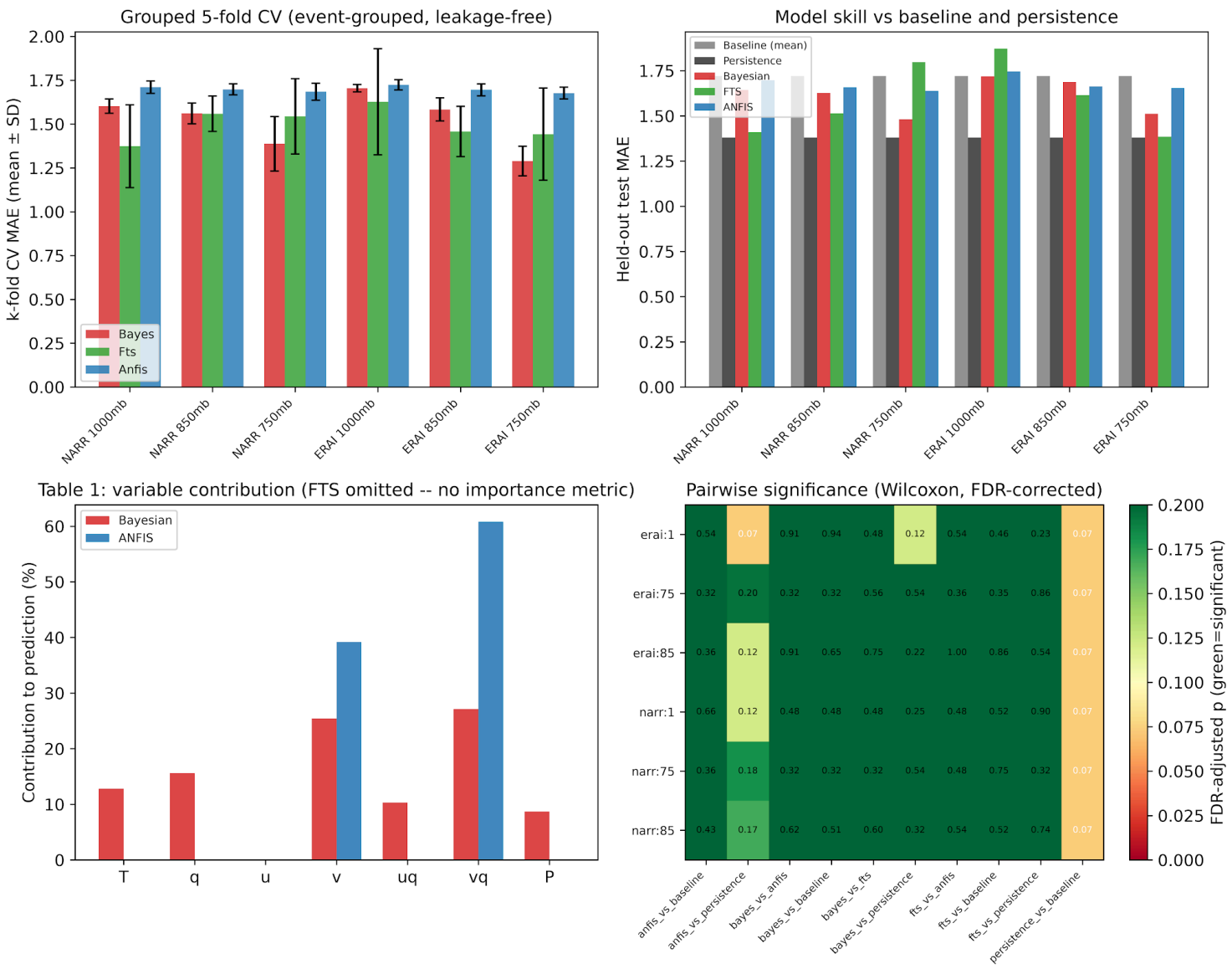


Figure 1

Model evaluation diagnostics. a Grouped five-fold cross-validation MAE (mean ± SD) for the Bayesian, FTS, and ANFIS models across the six data configurations. b Held-out test MAE for each model against the climatological mean baseline and the persistence baseline; persistence (dark grey) is the lowest bar in every configuration. c Variable contribution to prediction (%) for the Bayesian and ANFIS models at NARR 1000 mb; FTS yields no comparable importance metric and is omitted. d Pairwise Wilcoxon significance (FDR-adjusted p) for every model-versus-model and model-versus-baseline comparison; all cells exceed 0.05 after correction

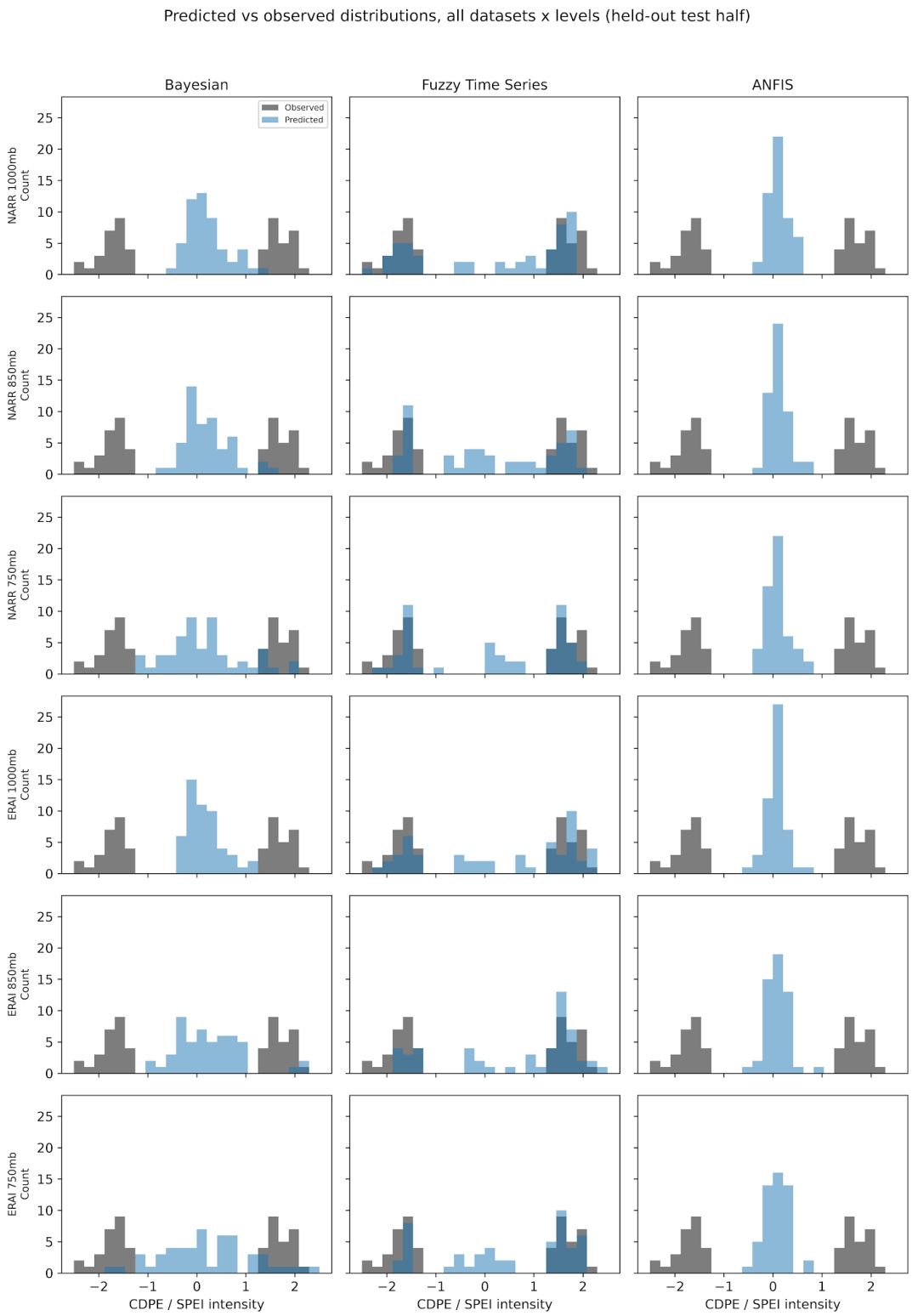


Figure 2

Predicted versus observed SPEI intensity distributions on the held-out test half, for all six data configurations (rows) and the three models (columns within each panel). Observed intensities (grey) are strongly bimodal, with dry poles near −1.7 and wet poles near +1.6. The Bayesian model (left of each panel) collapses its predictions toward the mean near zero; ANFIS (right) behaves similarly; fuzzy time

series (centre) reproduces more of the observed spread but with substantial scatter into the intermediate range

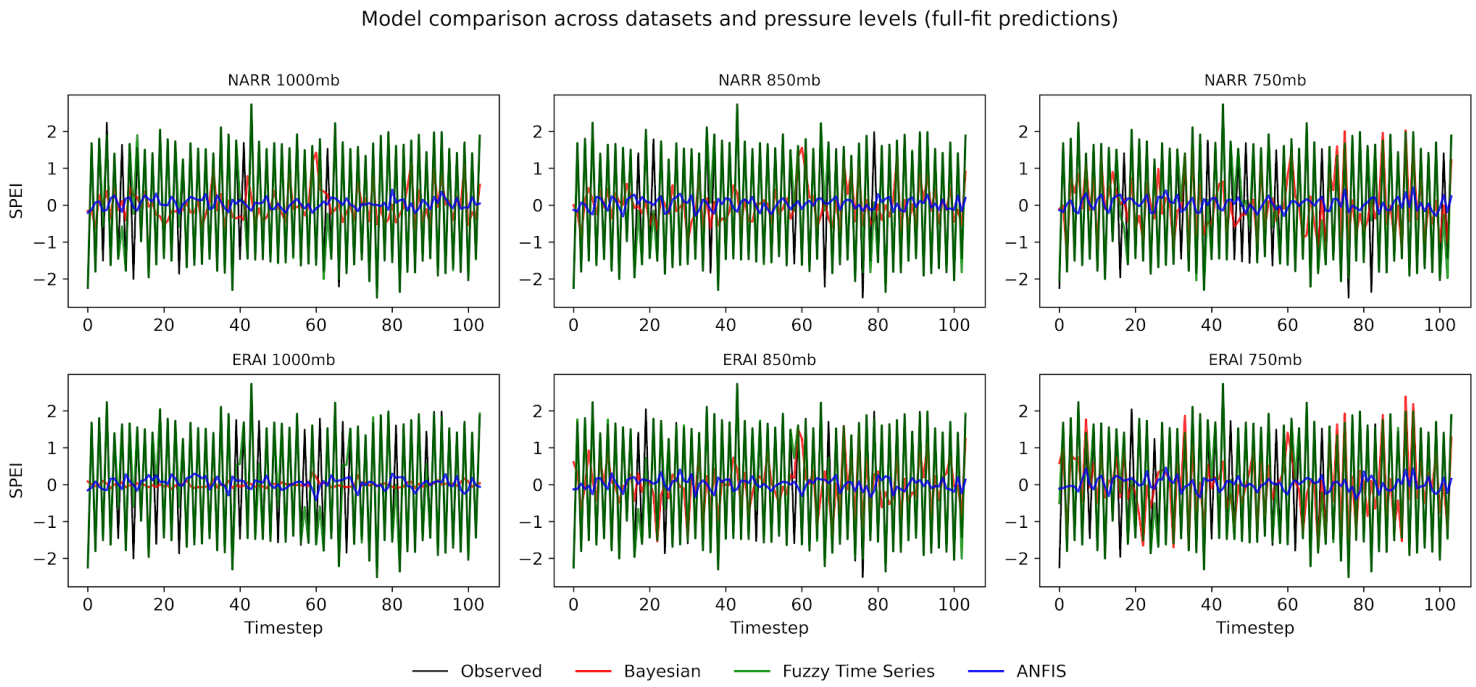


Figure 3

Observed and predicted CDPE intensity across held-out test events for the three models and all six data configurations (NARR top row, ERAI bottom row; 1000, 850, 750 mb left to right). Observed values are shown in black, Bayesian in red, fuzzy time series in green, and ANFIS in blue. The Bayesian and ANFIS predictions remain compressed near the centre of the range; the wider fuzzy time series excursions reflect attempted but inaccurate reproduction of the full intensity range

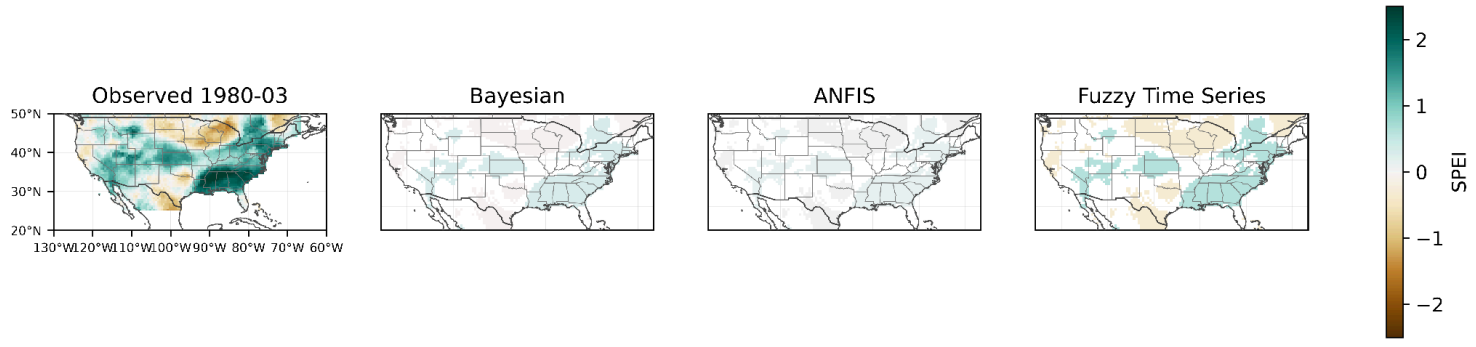


Figure 4

Spatial reconstruction of the March 1980 CDPE from predicted intensities (NARR 1000 mb), using nearest-neighbor interpolation of the sampled points. From left: observed SPEI field, Bayesian reconstruction, ANFIS reconstruction, and fuzzy time series reconstruction. Fuzzy time series recovers

the dipole placement most closely but with muted amplitude; the Bayesian reconstruction renders both poles as near-neutral