Towards Responsible Time Series Forecasting

KÄLIN Remo[1] and PORTMANN Edy[2]

[1] *Lucerne University of Applied Sciences and Arts, Institute of Communication and Marketing — University of Fribourg, HUMAN-IST Institute*
Zentralstrasse 9, CH-6002 Lucerne
Switzerland
E-mail: remo.kaelin@hslu.ch

[2] *University of Fribourg, HUMAN-IST Institute*
Boulevard de Pérolles 90, CH-1700 Fribourg
Switzerland
E-mail: edy.portmann@unifr.ch

1 Background

Ensemble techniques frequently outperform individual forecasting models, yet many common approaches remain too naive to support responsible use. Simple averaging or static weighting schemes do not account for how model performance varies across operating conditions and they also provide little insight into why a particular model should dominate the aggregated prediction at a given moment. This lack of context poses challenges in high-stakes settings such as national electricity forecasting, where both trust and accountability are essential [1]. We propose an interval Type-2 fuzzy ensembling framework that uses interpretable context descriptors to determine how forecasting models are weighted. Instead of relying on equal weighting or cumulative error metrics, the method characterizes each model's historical behavior through a collection of interval Type-2 TSK rules [2]. This produces an ensemble that not only improves accuracy but also enables a transparent connection between current operating conditions and the weights assigned to individual models.

2 Methodology

We study 24-hour-ahead electricity load forecasting. At each forecast origin, nine base models produce point forecasts. The ensemble assigns weights via fuzzy evaluation of context variables derived from the realized load.

2.1 Context Variables

Five scalar descriptors are used to characterize the system context for each day. Exogenous uncertainty is measured by the day-ahead temperature forecast error, regime change is captured by a smoothed first derivative of the load series and short-term variability is quantified through a seven-day realized volatility measure. Anomaly state is determined by a back fitting ARIMA-residual detector that flags large deviations from expected behavior, while linearity reflects the midday deviation index (MDI), describing how strongly the daily load profile strays from a a linear relationship.

Acknowledgement This research is funded by the Lucerne University of Applied Sciences and Arts.

https://www.doi.org/10.15452/978-80-7599-515-5.2026.12

2.2 Fuzzy Characterization

Each proxy is partitioned using triangular and trapezoidal Type-1 membership functions, defined by (1) and (2) as described by Mendel [3].

$$\mu_A(x;a,b,c) = \begin{cases} \dfrac{x-a}{b-a}, & a \le x < b, \\ \dfrac{c-x}{c-b}, & b \le x \le c, \\ 0, & \text{otherwise}, \end{cases} \quad (1) \qquad \mu_A(x;a,b,c,d) = \begin{cases} \dfrac{x-a}{b-a}, & a \le x < b, \\ 1, & b \le x \le c, \\ \dfrac{d-x}{d-c}, & c < x \le d, \\ 0, \end{cases} \quad (2)$$

Membership parameters are derived from the empirical distribution using percentile-based thresholds. The anomaly indicator is represented by a two-level fuzzy set, with high membership assigned to the normal or abnormal state accordingly.

2.3 Interval Type-2 Extension via Rarity

To express epistemic uncertainty caused by atypical operating conditions, the Type-1 membership functions are extended to interval Type-2 sets. The rarity function $U(x)$ quantifies how atypical a proxy value x is with respect to its empirical training distribution and is defined as

$$U(x) = 1 - \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{3}$$

where μ and σ denote the empirical mean and standard deviation of the proxy. Given the upper membership function $\mu_U(x)$, the corresponding lower membership $\mu_L(x)$ is obtained by

$$\mu_L(x) = \mu_U(x)\,[1 - U(x)], \tag{4}$$

which expands the footprint of uncertainty (FOU) under rare forecasting conditions.

2.4 Type-2 TSK Rule Structure

Let m be the index of the forecasting model and b the index of the fuzzy base set, for example "Weather Uncertainty High" or "Volatility Calm". Each proxy is divided into several linguistic terms, yielding multiple one-dimensional rules per model. For each model m, these rules form a dedicated interval Type-2 Takagi-Sugeno-Kang (TSK) [2] system whose output is a model-specific score. Each rule is defined as

$$R_{m,b}: \text{ IF context is in } \tilde{A}_b \text{ THEN } y = f_{m,b}, \tag{5}$$

with antecedent interval Type-2 set $\tilde{A}_b$. For a day t, this rule fires with

$$h_{L,m,b}(t) = \mu_{b,L}(c_t), \qquad h_{U,m,b}(t) = \mu_{b,U}(c_t), \tag{6}$$

where c_t is the proxy value associated with base set b. Unlike classical multidimensional TSK systems, rules operate on single-proxy inputs. Consequents are Type-1 reliability scores derived from fuzzy-weighted MAE values in each context region according to

$$f_{m,b} = -\frac{\mathrm{MAE}_{m,b}}{\tau}, \tag{7}$$

with scaling parameter $\tau > 0$. Interval Type-2 behavior therefore results only from the antecedent firing intervals defined in (6).

2.5 Type-Reduction and Weight Computation

Type-reduction follows Karnik and Mendel's Center-of-Sets (CoS) formulation of the Karnik-Mendel algorithm [3, 10] . For forecasting model m, the family of rules $\{R_{m,b}\}_b$ yields consequents $\{f_{m,b}\}$ and interval firing strengths $\{h_{L,m,b}(t), h_{U,m,b}(t)\}$. The CoS KM algorithm produces a type-reduced interval

$$S_{m,t} = \mathrm{KM}\{f_{m,b}, h_{L,m,b}(t), h_{U,m,b}(t)\}_b, \tag{8}$$

The model weights $w_{m,t}$, which determine the contribution of each forecasting model m at time t, are obtained from the reliability scores $S_{m,t}$ via the softmax transformation, and the final ensemble prediction $\hat{y}_t^{\mathrm{TSK}}$ is computed as the weighted sum of the individual model forecasts according to (9).

$$w_{m,t} = \frac{e^{S_{m,t}}}{\sum_j e^{S_{j,t}}}, \qquad \hat{y}_t^{\mathrm{TSK}} = \sum_m w_{m,t}\,\hat{y}_{m,t}. \tag{9}$$

3 Experimental Setup

The ensemble is evaluated on Swiss hourly load data (ENTSO-E) from 1 February 2024 to 26 September 2025 [5]. Exogenous inputs include day-ahead temperature forecasts (ECMWF IFS via Open-Meteo [4]) and national holiday indicators. Nine forecasting models are used: LightGBM [7], Holt-Winters [8], TBATS, KNN-regressor, MFLES [9], NBEATSx [6], SARIMAX, seasonal naive, and MSTL. All models produce 24-hour-ahead forecasts and are retrained weekly using an expanding window. Hyperparameters are tuned via four-fold expanding-window cross-validation. Fuzzy memberships and TSK consequents are learned from data prior to 1 July 2025, and evaluation is conducted from 1 July to 26 September 2025.

4 Results

Table 1 summarizes the out-of-sample performance. LightGBM is the strongest individual model, while uniform averaging already improves upon the single-model results. The proposed interval Type-2 TSK ensemble achieves the lowest MAE and highest R^2, demonstrating small but consistent gains from context-aware weighting.

Model / Series	n	MAE	MAPE	R^2
LightGBM	13776	499	7.81	0.596
Holt–Winters	13776	522	8.15	0.569
TBATS	13776	543	8.47	0.549
KNN-Regressor	13776	552	8.54	0.531
MFLES	13776	563	8.82	0.512
NBEATSx	13776	587	9.07	0.443
SARIMAX	13776	601	9.34	0.411
Seasonal Naive	13776	614	9.55	0.369
MSTL	13776	706	10.90	0.171
Uniform Ensemble	13776	485	7.60	0.613
TSK Ensemble	13776	**480**	**7.52**	**0.624**

Table 1: Out-of-sample accuracy over the evaluation period (1 July–26 September 2025).

5 Conclusion

We introduced an interval Type-2 fuzzy ensemble for responsible time series forecasting in critical infrastructure settings. The framework combines interpretable context descriptors with rarity-based interval Type-2 antecedents, crisp TSK consequents, Center of Sets KM type reduction, and softmax aggregation. This study constitutes an initial step within a broader, ongoing research effort on context-aware and responsible ensemble forecasting, based on practical data-driven proxies that will be further refined in future work.

Applied to Swiss day-ahead load forecasting, the proposed approach outperforms all individual models and consistently improves upon uniform averaging, while providing transparent and context-aware model weighting. Future work will focus on incorporating confidence intervals and extending both model inputs and ensemble outputs to interval-valued representations, further strengthening the role of fuzzy ensembling for trustworthy and interpretable forecasting in critical domains.

References

[1] M. T. Ribeiro, S. Singh, and C. Guestrin, Why Should I Trust You?: Explaining the Predictions of Any Classifier, *KDD*, pp. 1135–1144, 2016.

[2] M. Sugeno and G. T. Kang, Structure Identification of Fuzzy Model, *Fuzzy Sets and Systems*, vol. 28, no. 1, pp. 15–33, 1988.

[3] J. M. Mendel, *Explainable Uncertain Rule-Based Fuzzy Systems*, 3rd ed., Springer, 2024.

[4] P. Zippenfenig, Open-Meteo.com Weather API, *Software*, doi: 10.5281/zenodo.7970649, 2023.

[5] ENTSO-E, Transparency Platform: Electricity Load and Generation Data, 2025.

[6] K. G. Olivares, C. Challu, G. Marcjasz, R. Weron, and A. Dubrawski, Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx, *International Journal of Forecasting*, vol. 39, no. 2, pp. 884–900, 2023. doi: 10.1016/j.ijforecast.2022.03.001.

[7] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, LightGBM: a highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, vol. 30, pp. 3149–3157, 2017.

[8] C. Chatfield, The Holt-Winters forecasting procedure, *Journal of the Royal Statistical Society, Series C: Applied Statistics*, vol. 27, no. 3, pp. 264–279, 1978.

[9] F. Garza, M. M. Canseco, C. Challú, and K. G. Olivares, *StatsForecast: Lightning fast forecasting with statistical and econometric models*, PyCon, Salt Lake City, Utah, USA, 2022.

[10] N. N. Karnik, J. M. Mendel, and Q. Liang, “Type-2 fuzzy logic systems,” *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 6, pp. 643–658, Dec. 1999.