Multivariate ensemble post-processing using generative machine learning methods

Jieyu Chen, Tim Janke, Sebastian Lerch

jieyu.chen@kit.edu

Karlsruhe Institute of Technology
Chair of Statistics and Econometrics

Multivariate post-processing techniques are developed to retain the spatial, temporal, or inter-variable dependencies that univariate post-processing fails to preserve.

- Forecast data from ECMWF, with 50 ensemble members; The same dataset as has been used in Rasp and Lerch (2018)[1].
- 10 years of data available, splitting into training (2007–2014), validation (2015), and test (2016) sets.
- Focusing on preserving the **spatial** dependency (multiple locations) of **temperature** forecast in multivariate post-processing.
- Auxiliary weather variables: 'd2m', 'cape', 'q_pl850', 'sp', 'tcc', 'u_pl500', 'u_pl850', 'v_pl500', 'v_pl850', 'gh_pl500', 'u10', 'ssr', 'v10', 'str', 'sshf', 'slhf'.

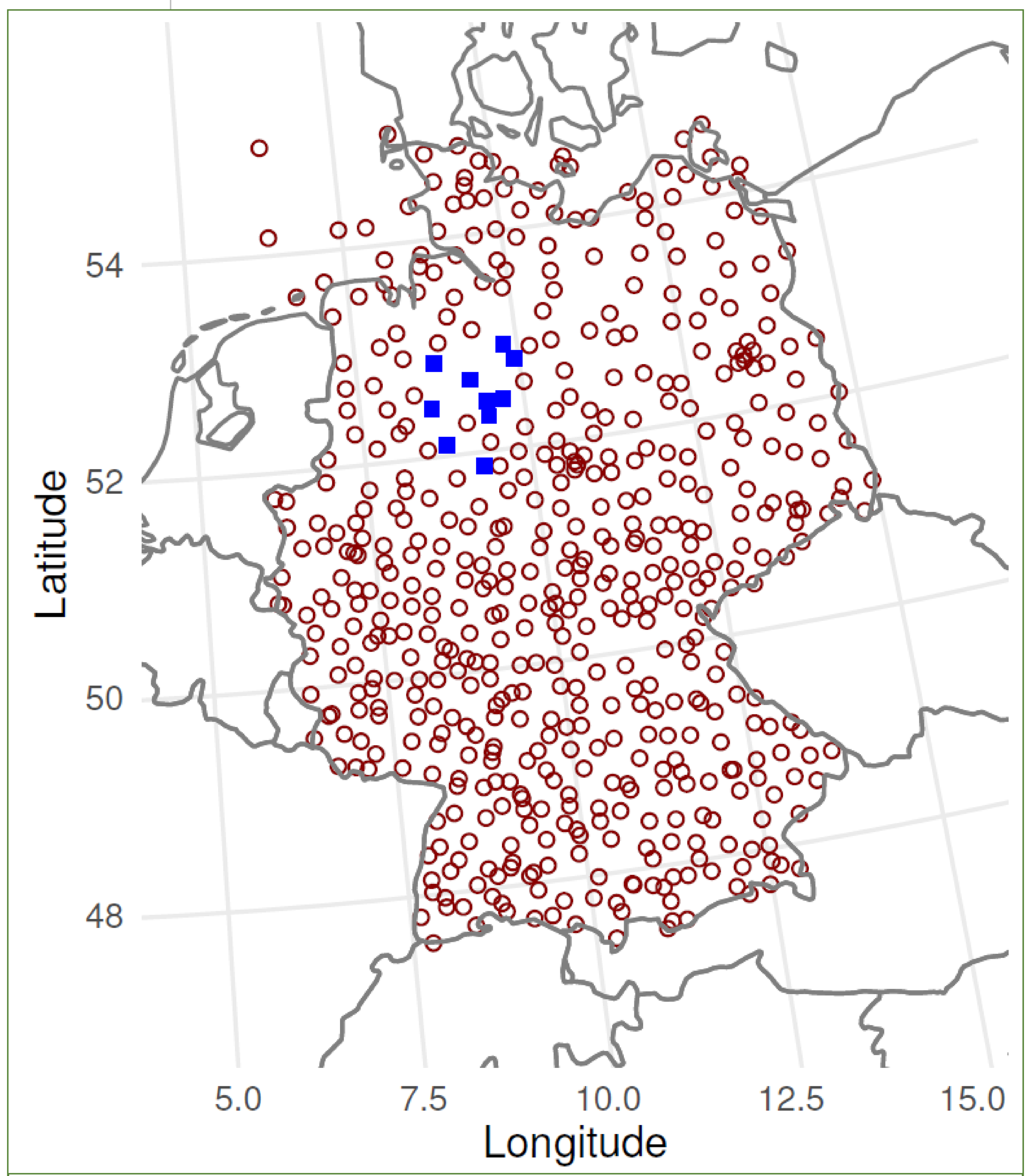

Fig. 1: Locations of weather stations in the dataset; Spatial dependence between 10 nearby stations of temperature forecast is considered in this study.

Standard approaches for multivariate post-processing:

- Copula-based, two-step methods;
  - First post-process ensemble forecasts univariately;
  - Then reorder post-processed samples to reintroduce multi-dimensional dependencies, based on the rank order structure learned from
    - raw ensemble members (ensemble copula coupling[2], ECC),
    - or a set of historical observations (Schaake shuffle[3], SSh),
    - or a fitted Gaussian copula (Gaussian copula approach[4], GCA).

Generative machine learning models

[build upon the work of Janke and Steinke (2020)[5]]

- ✓ Conceptually simpler, new class of data-driven multivariate distributional regression models;
- ✓ Multivariate probabilistic forecasts are directly obtained as output, bypassing the copula-based two-step approaches;
- ✓ Allows generation of unlimited number of post-processed samples;
- ✓ Enables incorporating additional weather variable information into the multivariate post-processing;

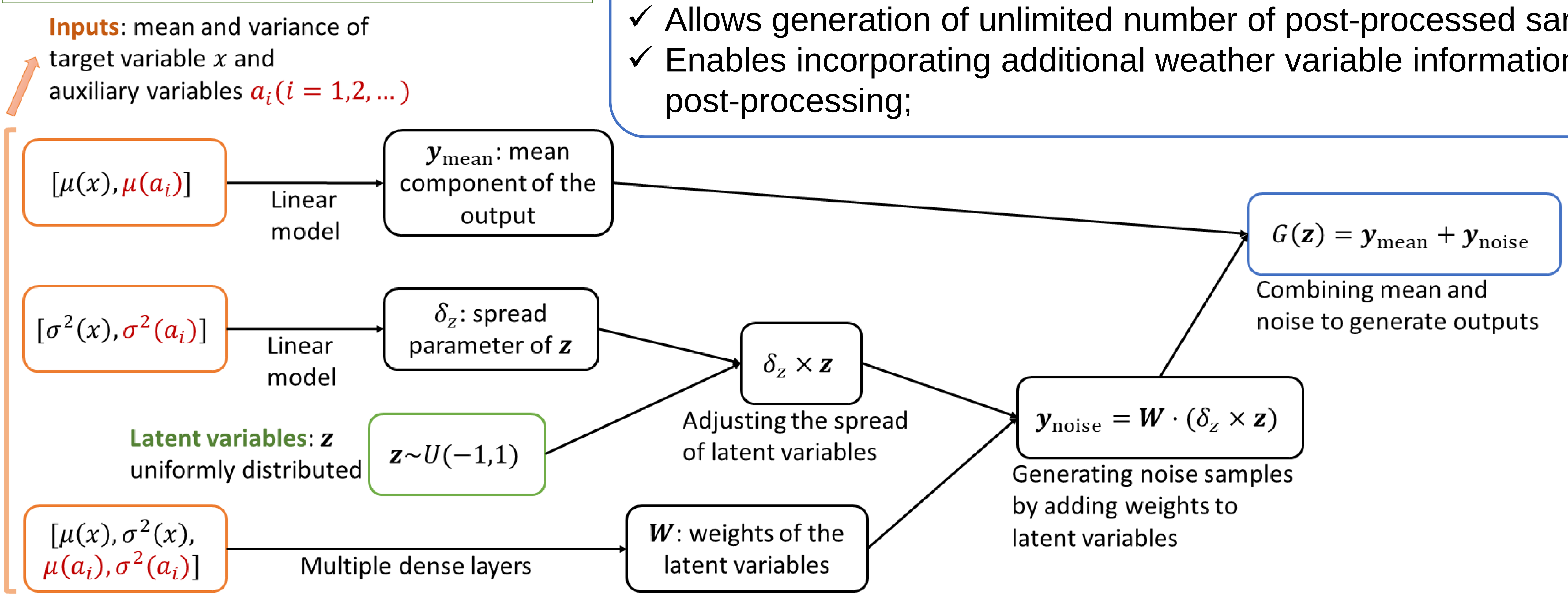

Fig. 2: Structure of the generative machine learning model for multivariate post-processing; The model is trained to minimize the energy score using stochastic gradient descent methods.

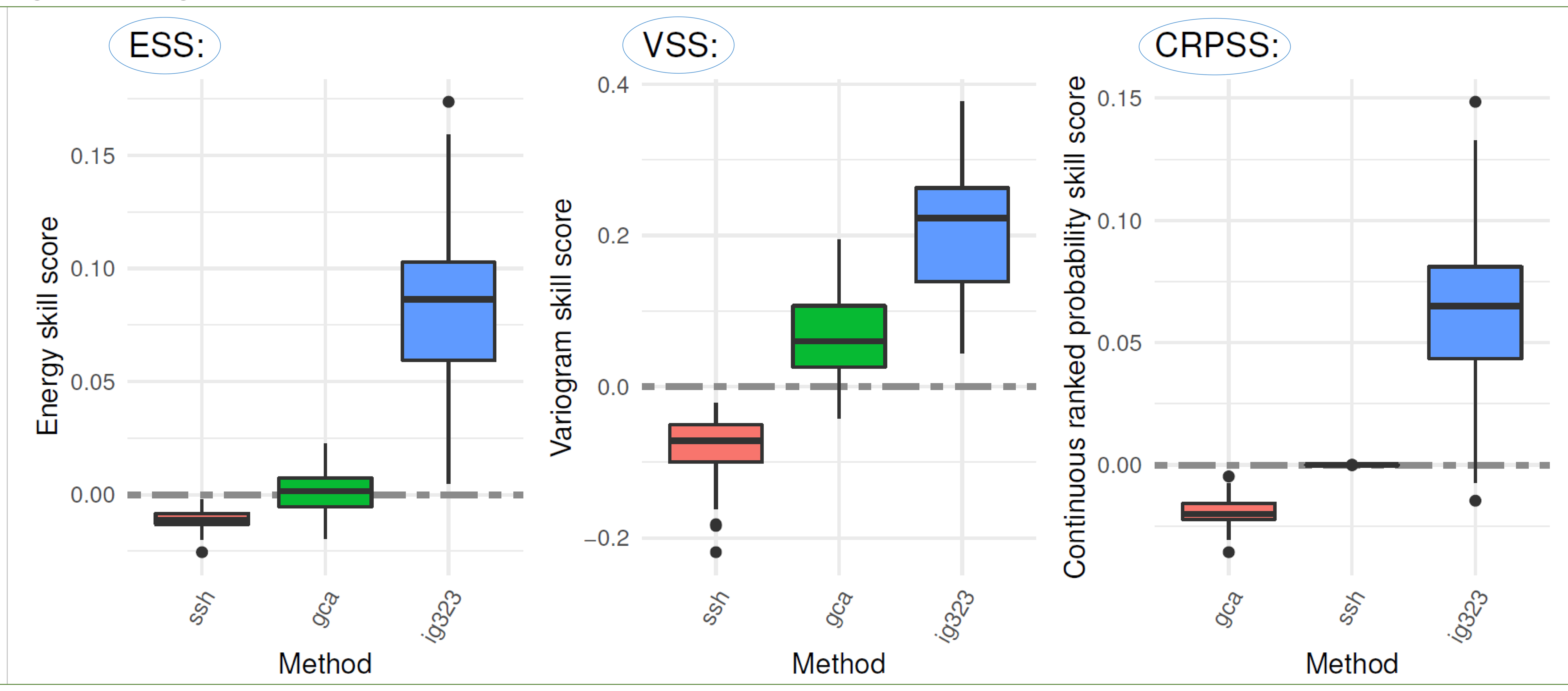

Fig. 3: Boxplots of skill scores for different multivariate post-processing approaches; Comparison between Schaake shuffle ('ssh'), Gaussian copula approach ('gca'), and the generative ML model ('ig323'); All skill scores are computed taking ECC as the reference forecast, using energy score, variogram score, and continuous ranked probability score.

References:

1) Rasp S, Lerch S. Neural networks for postprocessing ensemble weather forecasts. Monthly Weather Review. 2018 Nov 1;146(11):3885-900.
2) Schefzik R, Thorarinsdottir T L, Gneiting T. Uncertainty quantification in complex simulation models using ensemble copula coupling. Statistical science, 2013, 28(4): 616-640.
3) Clark M, Gangopadhyay S, Hay L, et al. The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. Journal of Hydrometeorology, 2004, 5(1): 243-262.
4) Möller A, Lenkoski A, Thorarinsdottir T L. Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. Quarterly Journal of the Royal Meteorological Society, 2013, 139(673): 982-991.
5) Janke T, Steinke F. Probabilistic multivariate electricity price forecasting using implicit generative ensemble post-processing. 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). IEEE, 2020.