

---

# Forecasting with Randomized Loss Backpropagation

---

**Olivier Sprangers**

AIRLab, NL  
University of Amsterdam, NL  
o.r.sprangers@uva.nl

**Sebastian Schelter**

University of Amsterdam, NL  
s.schelter@uva.nl

**Maarten de Rijke**

University of Amsterdam, NL  
m.derijke@uva.nl

## Abstract

Large-scale forecasting for complex systems is usually done using ML using scalar losses. In this work, we demonstrate that using randomized aggregated losses improves forecasting performance. More specifically, study the setting where commonly scalar loss functions are used. In our work, we demonstrate that using aggregate losses to optimize individual forecasts improves forecasting performance of an arbitrary collection of forecasts for these time series. Such a collection may include a hierarchical setting, where forecasting practitioners are interested in obtaining coherent forecasts for a set of hierarchical aggregations of time series.

## 1 Introduction

Contemporary large-scale forecasting applications require forecasting many time series concurrently Böse et al. [2017]. Often, there are dependencies between these individual time series. For example, there may be a hierarchical dependency, where individual time series may be aggregated into increasingly larger aggregations that are meaningful from the perspective of the forecasting application, such as individual product demand that may be aggregated into department demand, to store demand, to regional demand. Another example is dependencies between products themselves, where demand for product A (e.g. ) is correlated with the demand of product B. Ideally, when creating a forecast model for such a setting, we would like to take into account all these dependencies. This is a hard problem because of two reasons. First, taking into account correlations between individual products suffers from the curse of dimensionality; in the full 'dense' setting for every additional individual time series we need to compute its correlation with all other time series. Second, forecasting practitioners are commonly interested in aggregations too, for example in the case of hierarchical forecasting Hyndman et al. [2011]. Traditional methods commonly solve the first issue by using some approximation of the correlation matrix (e.g. a sparse matrix - add refs). The second issue is commonly solved by adding an explicit forecast reconciliation step when creating hierarchical forecasts. In this work, we show that we can solve both issues using randomized aggregated loss backpropagation. First, we formalize the notion of a randomized aggregated loss in time series problems. Then, we theoretically show that using randomized aggregated loss reduces bias and variance in the recursive forecasting setting. We empirically verify our theoretical result on a set of common hierarchical forecasting benchmarks and show that our solution improves forecasting performance up to [x%]. Finally, we show that randomized aggregated losses can yield forecasts that are better coherent across arbitrary aggregations, even when comparing against methods that specifically aim to optimize for specific aggregations, such as in hierarchical forecasting. The main contributions of this paper are:

1. We formalize the notion of randomized aggregate loss for large-scale time series problems
2. We theoretically show that using a randomized aggregated loss function reduces bias and variance in both the direct and recursive forecasting setting
3. We empirically demonstrate that a randomized aggregated loss improves hierarchical forecasting performance up to [x%] compared to using a scalar loss function

4. We provide a pip-installable Python package that allows practitioners to easily use our aggregated loss framework.

## 2 Related work

**Hierarchical Forecasting** Aggregated forecasts are commonly considered in the context of hierarchical forecasting Hyndman et al. [2011]. In this setting, there is a fixed hierarchical organization between time series, and one is interested in finding an optimal reconciliation between forecasts of different levels within the hierarchy. Such reconciliation steps are often performed as post-processing steps, i.e. a forecast is made for every time series and its hierarchical aggregations, and after training these forecasts are reconciled to satisfy the required hierarchical structure Hyndman et al. [2011, 2016], Taieb et al. [2017], Ben Taieb and Koo [2019], Wickramasuriya et al. [2019]. Limitations of these approaches is that they require a post-processing step and they consider a fixed aggregation structure. In our work, we demonstrate how using dynamic aggregations during training can improve forecasts at various aggregation levels. More recently and most closely related to our work, Han et al. [2021] introduced SHARQ, a method that reconciles probabilistic hierarchical forecasts during training by employing a regularized loss function that enforces hierarchical consistency of bottom-up forecasts. However, the hierarchical structure of the time series problem is considered fixed in SHARQ whereas we consider a structure that changes dynamically during training, and we will show that it is the dynamic changing of the structure that provides the additional benefits in terms of metrics improvement over other methods.

The benefit of forecasting with temporal hierarchies has been studied by Athanasopoulos et al. [2017]. However, this work considers only fixed temporal hierarchies and aims to produce forecasts on multiple hierarchies to improve individual forecasts. In our work, we do not create forecasts at higher aggregations but only use higher aggregation gradient information during the training of our model.

**Graph Neural Networks** Our work has connections to link-prediction in Graph Neural Networks (GNNs).

**Vector Regression**

## 3 Background

We consider the problem of forecasting for a set of time series, which we can formalize as follows:

Define aggregation

## 4 Randomized Aggregated Loss

In this section, we first introduce the notion of aggregated loss functions in time series problems. Then, we derive our main theoretical result, which shows how randomized aggregated loss backpropagation can reduce bias and variance in our forecasting problem setting. Finally, we analyse the computational complexity of our proposed loss function.

### 4.1 Randomized Aggregated Loss

In forecasting problems, we typically aim to minimize a loss function  $f(\cdot)$  over a set of time series  $N$  where each time series may consist of  $T$  time steps:

$$L = \sum_{i=1}^N \sum_{t=t_0}^T f(\cdot). \quad (1)$$

For example, for the *squared error loss* this reads:

$$L = \sum_{i=1}^N \sum_{t=t_0}^T (y_{i,t} - \hat{y}_{i,t})^2, \quad (2)$$

and we can obtain the often-used *mean squared error loss* by dividing the loss by  $NT$ . We now define a general *aggregated loss* as:

$$L = \sum_{i=1}^{N+S} \sum_{t=t_0}^T w_{i,t} f(\cdot), \quad (3)$$

where  $S$  denotes the number of aggregations that we consider and  $w_i$  is a scalar weight associated with each individual time series or its aggregations. A loss function of the form (3) was used to evaluate forecasts in the recent M5 hierarchical forecasting competition Makridakis et al. [2020]. We define a *randomized aggregated loss* as a loss of the form (3) where the aggregations used to compute the loss are selected uniformly at random during training. In Figure ??, we provide a set of examples. In Figure The key insight of our paper is that we can obtain more accurate forecasts, with lower bias and variance, *by considering random aggregations during the training of our forecasting model rather than no aggregations or fixed (hierarchical) aggregations.*

## 4.2 Bias and Variance Reduction

## 4.3 Computational Complexity

# 5 Experiments

In this section we empirically verify our theoretical results.<sup>1</sup> First, we demonstrate how using our randomized aggregated loss improves forecasting performance by up to [x%]. Then, we show that using our aggregated loss results in a reduction of bias and variance, as suggested by our theoretical result.

## 5.1 Datasets

## 5.2 Results

# 6 Discussion

# 7 Conclusion

## Broader Impact

Positive: better forecasts

Negative: implicitly our system may build on dependencies between time series that may be considered biased, or could potentially be discriminatory.

## Acknowledgments and Disclosure of Funding

This research was (partially) funded by the Hybrid Intelligence Center, a 10-year program funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research.<sup>2</sup>

All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

We thank the reviewers for their constructive feedback and help on improving our work.

## References

George Athanasopoulos, Rob J. Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos. Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74, October 2017. ISSN 0377-2217. doi: 10.1016/j.ejor.2017.02.046.

<sup>1</sup>The code to reproduce our experiments can be found at LINK

<sup>2</sup><https://www.hybrid-intelligence-centre.nl/>

- Souhaib Ben Taieb and Bonsoo Koo. Regularized Regression for Hierarchical Forecasting Without Unbiasedness Conditions. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1337–1347, Anchorage AK USA, July 2019. ACM. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330976.
- Joos-Hendrik Böse, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Dustin Lange, David Salinas, Sebastian Schelter, Matthias Seeger, and Yuyang Wang. Probabilistic demand forecasting at scale. *Proceedings of the VLDB Endowment*, 10(12):1694–1705, August 2017. ISSN 21508097. doi: 10.14778/3137765.3137775.
- Xing Han, Sambarta Dasgupta, and Joydeep Ghosh. Simultaneously Reconciled Quantile Forecasting of Hierarchically Related Time Series. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pages 190–198. PMLR, March 2021.
- Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos, and Han Lin Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9): 2579–2589, September 2011. ISSN 0167-9473. doi: 10.1016/j.csda.2011.03.006.
- Rob J. Hyndman, Alan J. Lee, and Earo Wang. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*, 97:16–32, May 2016. ISSN 0167-9473. doi: 10.1016/j.csda.2015.11.007.
- Spyros Makridakis, Evangelos Spiliotis, and Vassilis Assimakopoulos. *The M5 Accuracy Competition: Results, Findings and Conclusions*. October 2020.
- Souhaib Ben Taieb, James W. Taylor, and Rob J. Hyndman. Coherent Probabilistic Forecasts for Hierarchical Time Series. In *International Conference on Machine Learning*, pages 3348–3357, July 2017.
- Shanika L. Wickramasuriya, George Athanasopoulos, and Rob J. Hyndman. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association*, 114(526):804–819, April 2019. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2018.1448825.