DMIDAS: Deep Mixed Data Sampling Regression for Long Multi-Horizon Time Series Forecasting

Cristian Challu¹ Kin G. Olivares¹ Gus Welter¹ Artur Dubrawski¹

Abstract

Neural forecasting has shown significant improvements in the accuracy of large-scale systems, yet predicting extremely long horizons remains a challenging task. Two common problems are the volatility of the predictions and their computational complexity; we addressed them by incorporating smoothness regularization and mixed data sampling techniques to a well-performing multilayer perceptron based architecture (NBEATS). We validate our proposed, DMIDAS, on high-frequency healthcare and electricity price data with long forecasting horizon (∼ 1000 timestamps) where we improve the prediction accuracy by 5% over state-of-the-art models, reducing the number of parameters of NBEATS by nearly 70%.

1. Introduction

Recently neural forecasting has shown great success on improving the accuracy of forecasting systems. Long-horizon forecasting remains a challenging for neural networks as often times their expressiveness translates into excessive computational complexity and volatility. We address these limitations, improving on well performing multi-horizon models with temporal mixed data sampling techniques and smoothness regularization. Our contributions include:

(i) Mixed Data Sampling: We incorporate sub-sampling layers before fully-connected networks, and observe that this technique significantly reduces the memory footprint and the amount of computation, while maintaining the effective memory of the model.

(ii) Smoothness Regularization: We induce smoothness of the multi-horizon model’s predictions by reducing the dimension of its outputs and matching the forecast with the original frequency through interpolation. We add L1 regularization to shrink the weights towards a sparse representation, which robustifies the model against the data’s higher frequency noise.

(iii) DMIDAS architecture: The two regularization techniques naturally motivate the Deep Mixed Data Sampling regression (DMIDAS), that improves the forecast decomposition capabilities of NBEATS by specializing the blocks of the architecture on different frequencies of the data, reducing its volatility, and computational complexity, while maintaining its predictive power.

We compare DMIDAS on long horizon forecasting tasks against well-established benchmarks: a parsimonious fully-connected network (MLP; Lago et al. 2021), the Dilated Recurrent Neural Network (DiRNN; Chang et al. 2017), the Exponential Smoothing Recurrent Neural Network (ES-RNN; Smyl 2019) and the Neural Basis Expansion Analysis (NBEATS; Oreshkin et al. 2020). DMIDAS reduced RMSE by 3% and MAE by 5% against 2nd best model. We improve the average RMSE across datasets on 6%, 11%, 10% over NBEATS-G, NBEATS-I and ESRNN respectively.

Figure 1. Arterial pressure (ART) and French European Power Exchange electricity price, along with the long-multi-horizon forecasts of DMIDAS. High-Frequency data often poses challenging forecasting tasks when series display heterogeneous behavior across series and signals have non-stationary dynamics.
The remainder of the work is structured as follows. Section 2 reviews relevant literature, Section 3 introduces notation and describes the methodology, Section 4 contains the empirical findings. Finally in Section 5 we wrap up and conclude.

2. Literature Review

Neural forecasting methods have become an increasingly active area of research in recent years, as these models have been successfully adopted in multiple domains such as demand forecasting (Wen et al., 2017; Salinas et al., 2020), weather prediction (Nascimento et al., 2021) energy markets (Olivares et al., 2021; Gasparin et al., 2019) and excelled at several forecasting competitions (Smyl, 2019; Oreshkin et al., 2020; Lim et al., 2020). For a comprehensive survey of neural forecasting see (Benidis et al., 2020).

Regarding our approach to tackle the long horizon forecasting task we found the research over multi-step-ahead forecasting strategies and mixed data sampling regressions to be the most relevant, we summarize below.

Multi-step-ahead forecasting strategies. Comprehensive investigations on the bias and variance behavior of multi-step-ahead forecasting strategies observed that the direct strategy has a low bias and high variance, with the benefit of avoiding error accumulation across the horizon exhibit by classic recursive strategies. Finally, variants of the joint or multi-horizon strategy allow for the best trade off of the variance and bias effects because they maintain great expressiveness while still sharing parameters (Bao et al., 2014; A tiya & Taieb, 2016; Wen et al., 2017).

Mixed data sampling regression. Previous forecasting literature recognized challenges of extremely long horizon predictions, and proposed mixed data sampling regression (MIDAS) to ameliorate the problem of parameter proliferation while preserving high frequency temporal information (Ghysels et al., 2007; Armesto et al., 2010). MIDAS regressions maintained the classic recursive forecasting strategy of linear auto-regressive models, but defined a parsimonious fashion of feeding the inputs to the model.

Smoothness Regularization. Interpolation techniques to augment the resolution of modeled signals has a very long tradition (Meijering, 2002), with applications in many fields like signal and image processing. In time series forecasting it has many applications, from completing unevenly sampled data and noise filters, to temporal hierarchical forecasting (Chow & Ioh Lin, 1971; Fernandez, 1981). For deep learning, interpolation has been seen in computer vision applications (Noh et al., 2015; Ye et al., 2018).

To our knowledge, temporal interpolation has not been explicitly leveraged in neural forecasting model’s architectures to induce smoothness of its predictions.

3. Methodology

3.1. Neural Basis Expansion Analysis

The Neural Basis Expansion Analysis (NBEATS) is a state-of-the-art deep learning univariate model. The key idea of the model is to perform local nonlinear projections onto basis functions across multiple blocks. Each block consists of a fully-connected neural network (MLP) which learns coefficients for the backcast and forecast outputs on a predefined basis. The backcast output is used to clean the inputs of subsequent blocks, while the forecasts are summed to compose the final prediction. The blocks are grouped in stacks, each specialized in learning a different characteristic of the data using different basis functions.

The NBEATS architecture is composed of \( S \) stacks with \( B \) blocks each. The input \( y^b \) of the first block consists of \( L \) lags of the target time-series \( y \), while the inputs of the following blocks include residual connections with the backcast output of the previous block. Within each \( l \)-th block, the first component consists of a MLP that learns hidden vector \( h_l \), which is then passed to a linear layer to produce \( \theta^l_h \) and backcast \( \theta^l_b \) expansion coefficients. After the MLP, each \( l \)-th block includes a basis expansion operation between the coefficients learnt and the block’s basis function. This transformation results in the backcast \( \hat{y}^l_b \) and forecast \( \hat{y}^l_f \) outputs of the block. NBEATS’ block operations are:

\[
\begin{align}
    h_l &= \text{MLP}_l (y^b_{l-1}) \\
    \theta^l_h &= \text{LINEAR}^l (h_l) \\
    \theta^l_b &= \text{LINEAR}^b (h_l) \\
    \hat{y}^l_f &= \theta^l_h V^l_f \\
    \hat{y}^l_b &= \theta^l_b V^l_b
\end{align}
\]

\( V^l_f = I_{H \times H} \) is used in the generic NBEATS-G version.

3.2. Deep Mixed Data Sampling Regression

3.2.1. Mixed Data Sampling

In order to capture long time dynamics while not over-parametrizing the model we added pooling layers\(^1\). Different kernel sizes induce mixed frequencies of the inputs on the backcast window. These sub-sampling layers effectively reduce the number of parameters, limiting the memory footprint and the amount of computation, while maintaining the original receptive field. This technique was previously explored in the MQ-CNN architecture (Wen et al., 2017).

\(^1\)The pooling layers can be average pooling, max pooling or simple stride down sampling.
We define the coefficients \( \theta \). We design the architecture so that each block specializes on while the final predictions are constructed using interpolation. The expressiveness of the output dimension of each stack guides the specialization of the additive predictions on different frequencies, while the final predictions are constructed using interpolation.

3.2.2. SMOOTHNESS REGULARIZATION

For most multi-horizon forecasting models, and in particular NBEATS as specified in Equation (1), the outputs’ cardinality corresponds to the horizon’s dimension, \(| \theta_f^l | = H \). As the dimension of the forecast horizon grows, so the model prediction’s volatility. To solve this issue, DMIDAS defines the dimensions of its forecast coefficients in terms of the expressivity ratio \( r_l \) that controls the number of parameters per unit of time, now \(| \theta_f^l | = [r_l H] \). To recover the target horizon \( H \) on the original frequency DMIDAS uses temporal interpolation:

\[
y^*_f l t = \left( \theta_{t_1}^l + \left( \frac{\theta_{t_2}^l - \theta_{t_1}^l}{H} \right) (t - t_1) \right)
\]

(2)

In Equation (2), \( t_2 \) denotes the closest available time index in the future of \( t \) where the coefficients \( \theta_f^l \) have an associated value, analogous to \( t_1 \) for the past. A consequence of the temporal interpolation, is that the predictions are continuous and smooth between each coefficient \( \theta_{t_1}^l \) and \( \theta_{t_2}^l \).

3.2.3. DMIDAS ARCHITECTURE

The ideas from multi-horizon forecasting, mixed data sampling regressions and smoothness regularization through interpolation naturally converge into DMIDAS approach. We design the architecture so that each block specializes on different sampling frequencies of the inputs and outputs of the model simultaneously, by selecting the expressivity ratio. We define the coefficients \( \theta_f^l \) and \( \theta_r^l \) for each \( l \)-th block to be uniformly spaced, as shown in Figure 2.

We use exponentially increasing expressivity ratios through the depth of the architecture blocks that allows to model complex dependencies, while controlling the number of parameters used on each output layer. If the expressivity ratio is defined as \( r_l = r^l \) then the space complexity of DMIDAS scales geometrically \( O((H(1-r^l))/(1-r)) \), while the NBEATS-G scales linearly \( O(HB) \).

DMIDAS enjoys several advantages. First, it reduces the volatility of the model’s outputs and improves the overall computational efficiency. Second, the additive decomposition capabilities inherited from NBEATS improves by making each block specialize on different frequencies. Furthermore, the expressivity ratio \( r_l \) can be specified using the generating process domain expertise to improve forecasting performance. Third, L1 regularization can continue to shrink the weights towards a sparse representation, robustifying the model to higher frequency noise.

The additive forecast decomposition of DMIDAS provides valuable information beyond the trend-seasonality decomposition. Figure 3 shows a forecast of DMIDAS and its corresponding decomposition for ART and compares it with the non-interpretable decomposition of the NBEATS-G.

4. Experiments

Vital Signs Dataset. This dataset consists of de-identified high-frequency vital signs collected from the intensive-care at the University of Pittsburgh Medical Center Hospitals (UPMC) over three years. It contains arterial blood pressure and heart rate. The data was collected from Phillips Data Warehouse Connect, and it was prepared and de-identified under Institutional Review Board review and approval.
pressure (ART) and Pulse Oximetry Photoplethysmogram (PLETH) waveforms for 98 patients. Each patient has 90-minute data, the last five minutes comprise the test set.

**Electricity Price Dataset.** We consider electricity price forecasting (EPF) datasets of five major power markets ⁴, namely Nord Pool, the Pennsylvania-New Jersey-Maryland market, and the European Power Exchange for Belgium, France and Germany. Each market contains six years of history, we keep the last three months of each as test set.

### 4.1. Training Methodology and Evaluation

For ART and PLETH we train global models. For EPF we train separate models for each market. A simple mean ensemble of four random initializations is used to forecast. Model selection is done with a bayesian optimization technique which explores the hyperparameter space using tree-structured Parzen estimators (HYPEROPT; Bergstra et al. 2011). The configurations that reach the lowest validation mean absolute error (MAE) are evaluated on the test data.

We evaluate the accuracy with the mean absolute error (MAE) and root mean squared error (RMSE). DMIDAS consistently achieved the best performance on long horizons, with a monotonic increasing relative improvement vs horizon. It improves the RMSE 3% on average and MAE on 5% against the 2nd best model, including several models and forecasting strategies, such as the recursive forecast of the DilRNN, while remaining interpretable. DMIDAS improve the average RMSE across datasets on 6%, 11%, 10% over NBEATS-G, NBEATS-I and ESRNN respectively.

Table 1. Forecast accuracy measures for long-horizon tasks. The reported metrics are mean absolute error (MAE) and root mean squared error (RMSE). Smallest errors are highlighted in bold.

| Data | H | Metric | MLP | DilRNN | ESRNN | NBEATS-I | NBEATS-G | DMIDAS |
|------|---|--------|-----|--------|-------|----------|----------|--------|
|      |   |        |     |        |       |          |          |        |
| ART  | 120| RMSE   | 6.51| 6.62   | 5.89  | 5.25     | 5.30     | 5.44   |
|      |    | MAE    | 0.046| 0.073  | 0.079 | 0.081    | 0.082    | 0.073  |
|      | 480| RMSE   | 9.21| 9.12   | 8.73  | 7.80     | 7.68     | 7.92   |
|      |    | MAE    | 0.067| 0.071  | 0.065 | 0.063    | 0.061    | 0.062  |
|      | 960| RMSE   | 12.23| 11.84  | 14.17 | 9.87     | 9.87     | 9.92   |
|      |    | MAE    | 0.094| 0.095  | 0.094 | 0.095    | 0.095    | 0.095  |
| PLETH| 120| RMSE   | 6.31| 6.42   | 5.89  | 5.25     | 5.30     | 5.44   |
|      |    | MAE    | 0.045| 0.046  | 0.043 | 0.043    | 0.043    | 0.044  |
|      | 480| RMSE   | 9.21| 9.12   | 8.73  | 7.80     | 7.68     | 7.92   |
|      |    | MAE    | 0.067| 0.071  | 0.065 | 0.063    | 0.061    | 0.062  |
|      | 960| RMSE   | 12.23| 11.84  | 14.17 | 9.87     | 9.87     | 9.92   |
|      |    | MAE    | 0.094| 0.095  | 0.094 | 0.095    | 0.095    | 0.095  |
| EPF  | 24 | RMSE   | 9.84| 9.66   | 9.55  | 9.49     | 9.34     | 9.84   |
|      |    | MAE    | 0.055| 0.055  | 0.059 | 0.049    | 0.049    | 0.054  |
|      | 336| RMSE   | 12.94| 12.99  | 12.84 | 13.29    | 13.00    | 12.84  |
|      |    | MAE    | 0.080| 0.080  | 0.080 | 0.075    | 0.075    | 0.078  |
|      | 672| RMSE   | 18.03| 17.45  | 17.33 | 16.92    | 18.11    | 15.88  |
|      |    | MAE    | 0.038| 0.038  | 0.038 | 0.038    | 0.038    | 0.038  |

### 5. Conclusion

We identified two challenges in long-horizon forecasting tasks, namely the volatility of multi-horizon model’s predictions and their computational complexity. We proposed the **Deep Mixed Data Sampling regression** (DMIDAS), that incorporates smoothness regularization through interpolation and sub-sampling techniques to NBEATS. The resultant parsimonious model outperforms state-of-the-art benchmarks on long-horizon forecasting tasks, improving the MAE 5% on average and the RMSE 3%, while reducing the numbers of parameters of the NBEATS model by nearly 70%. Additionally, DMIDAS has an interpretable forecast decomposition that provides valuable information beyond the classic trends and seasons.

---

⁴Available at the EPFtoolbox library (Lago et al., 2021).
References

Armesto, M. T., Engemann, K. M., and Owyang, M. T. Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, 92:521–536, 2010. URL https://research.stlouisfed.org/publications/review/2010/11/01/forecasting-with-mixed-frequencies.

Atiya, A. and Taieb, B. A bias and variance analysis for multistep-ahead time series forecasting. *IEEE transactions on neural networks and learning systems*, 27(1): 2162–2388, 2016. URL https://pubmed.ncbi.nlm.nih.gov/25807572/.

Bao, Y., Xiong, T., and Hu, Z. Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing*, 129:482–493, 2014. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2013.09.010. URL https://www.sciencedirect.com/science/article/pii/S092523121300917X.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. Algorithms for hyper-parameter optimization. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 24, pp. 2546–2554. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cf1d2577bc2619bc635690-Paper.pdf.

Chang, S., Zhang, Y., Han, W., Yu, M., Guo, X., Tan, W., Cui, X., Witbrock, M., Hasegawa-Johnson, M. A., and Huang, T. S. Dilated recurrent neural networks. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/32bb90e8976a9ab5298d5d410fe66f21d-Paper.pdf.

Chow, G. C. and loh Lin, A. Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4):372–375, 1971. ISSN 00346535, 15309142. URL http://www.jstor.org/stable/1928739.

Fernandez, R. B. A Methodological Note on the Estimation of Time Series. *The Review of Economics and Statistics*, 63(3):471–476, August 1981. URL https://ideas.repec.org/a/taf/restat/v63y1981i3p471-76.html.

Gasparin, A., Lukovic, S., and Alippi, C. Deep learning for time series forecasting: The electric load case. *CoRR*, abs/1907.09207, 2019. URL http://arxiv.org/abs/1907.09207.

Ghysels, E., Sinko, A., and Valkanov, R. Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90, 2007. doi: 10.1080/07474930600972467. URL https://doi.org/10.1080/07474930600972467.

Lago, J., Marcjasz, G., De Schutter, B., and Weron, R. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983, 2021. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2021.116983. URL https://www.sciencedirect.com/science/article/pii/S0306261921004529.

Nascimento, R. C., Souto, Y. M., Ogasawara, E. S., Porto, F., and Bezerra, E. STconvS2S: Spatiotemporal convolutional sequence to sequence network for weather forecasting. *Neurocomputing*, 426:285–298, 2021. URL http://arxiv.org/abs/1912.09363.

Noh, H., Hong, S., and Han, B. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

Olivares, K. G., Challu, C., Marcjasz, G., Weron, R., and Dubrawski, A. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx. *International Journal of Forecasting, submitted*, Working Paper version available at arXiv:2104.05522, 2021.
Oreshkin, B. N., Carpov, D., Chapados, N., and Bengio, Y. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. In 8th International Conference on Learning Representations, ICLR 2020, 2020. URL https://openreview.net/forum?id=r1ecqn4YwB.

Salinas, D., Flunkert, V., Gasthaus, J., and Januschowski, T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, 36(3):1181–1191, 2020. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2019.07.001. URL https://www.sciencedirect.com/science/article/pii/S0169207019301888.

Smyl, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. International Journal of Forecasting, 07 2019. doi: 10.1016/j.ijforecast.2019.03.017.

Wen, R., Torkkola, K., Narayanaswamy, B., and Madeka, D. A Multi-horizon Quantile Recurrent Forecaster. In 31st Conference on Neural Information Processing Systems NIPS 2017, Time Series Workshop, 2017. URL https://arxiv.org/abs/1711.11053.

Ye, J. C., Han, Y., and Cha, E. Deep convolutional framelets: A general deep learning framework for inverse problems. SIAM Journal on Imaging Sciences, 11(2):991–1048, 2018.