Interpretable Neural Forecasting

Kin G. Olivares & Cristian Challu

Overview

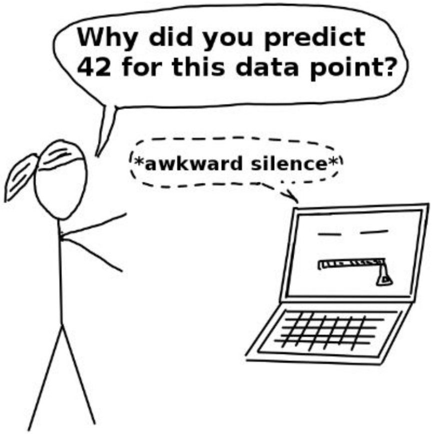
1. Preliminaries

2. NBEATSx: Neural basis expansion analysis with exogenous variables
    a. Convolutional Encoders
    b. Time Series Decomposition
    c. Electricity Price Forecasting

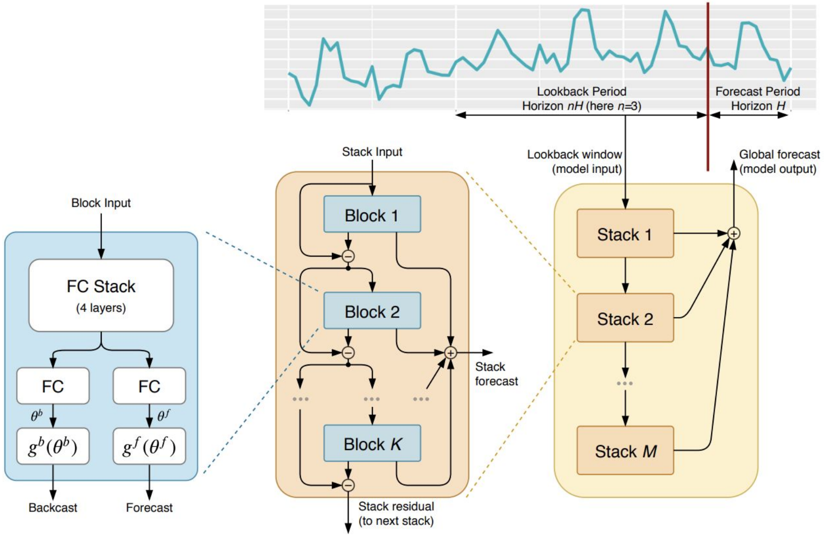
3. DeepMidas: Deep Mixed Data Sampling Regression for Long Multi-Horizon Forecasting
    a. Mixed Data Sampling
    b. Smoothness Regularization
    c. Healthcare Signals and Electricity Price Forecasting

Preliminaries

Neural forecasting has proven powerful and flexible, yet there are several situations where our understanding of the model's predictions can be as crucial as their accuracy.



NBEATS: Architecture



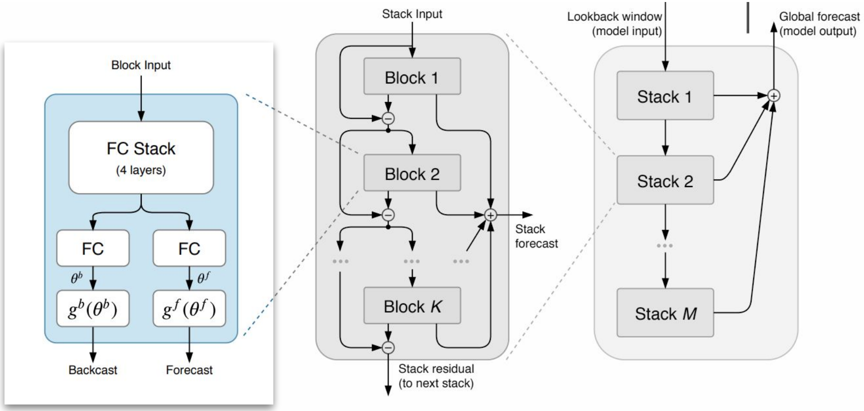
Oreshkin, B. N., Carpov, D., Chapados, N., and Bengio,Y. *N-BEATS: Neural Basis Expansion Analysis for interpretable time series forecasting*. In 8th International Conference on Learning Representations, ICLR 2020. Available at https://arxiv.org/abs/1905.10437.

NBEATS: Architecture

Expansion coefficients

$$\mathbf{h}_l = \mathbf{FCNN}_l\left(\mathbf{y}_{l-1}^{back}, \mathbf{X}_{l-1}\right)$$

$$\boldsymbol{\theta}_l^{for} = \mathbf{LINEAR}^{for}\left(\mathbf{h}_l\right) \qquad \boldsymbol{\theta}_l^{back} = \mathbf{LINEAR}^{back}\left(\mathbf{h}_l\right)$$

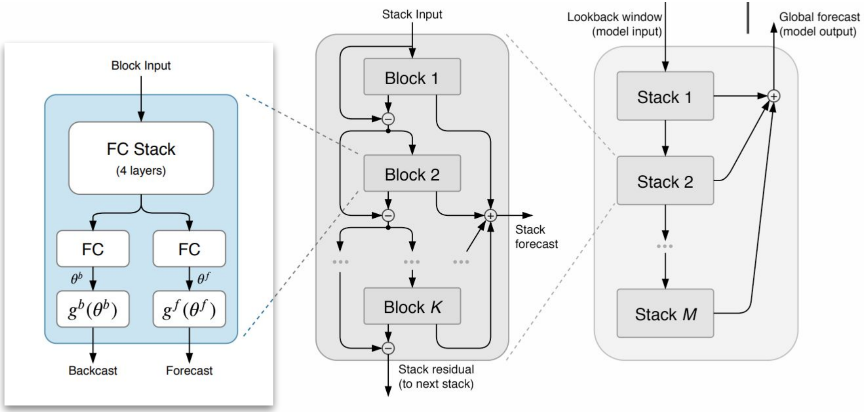


Oreshkin, B. N., Carpov, D., Chapados, N., and Bengio,Y. *N-BEATS: Neural Basis Expansion Analysis for interpretable time series forecasting*. In 8th International Conference on Learning Representations, ICLR 2020. Available at https://arxiv.org/abs/1905.10437.

NBEATS: Architecture

Basis expansion

$$\hat{\mathbf{y}}_l^{back} = \sum_{i=1}^{|\boldsymbol{\theta}_l^{back}|} \theta_{l,i}^{back}\, \mathbf{v}_{l,i}^{back} \equiv \boldsymbol{\theta}_l^{back}\, \mathbf{V}_l^{back} \qquad \hat{\mathbf{y}}_l^{for} = \sum_{i=1}^{|\boldsymbol{\theta}_l^{for}|} \theta_{l,i}^{for} \mathbf{v}_{l,i}^{for} \equiv \boldsymbol{\theta}_l^{for}\, \mathbf{V}_l^{for}$$

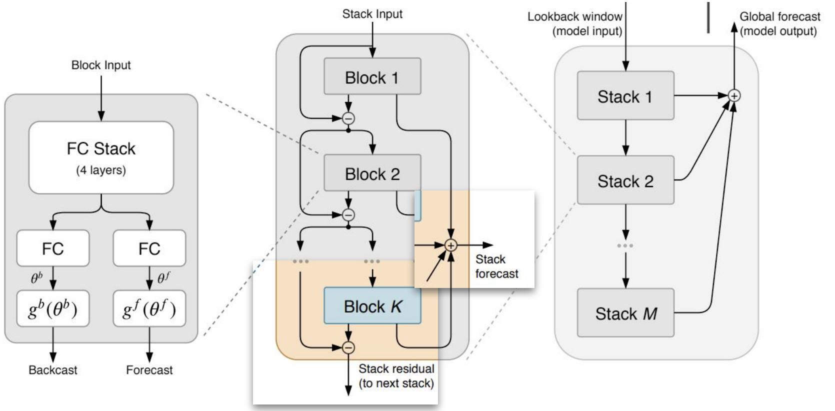


Oreshkin, B. N., Carpov, D., Chapados, N., and Bengio,Y. *N-BEATS: Neural Basis Expansion Analysis for interpretable time series forecasting*. In 8th International Conference on Learning Representations, ICLR 2020. Available at https://arxiv.org/abs/1905.10437.

NBEATS: Architecture

Doubly residual stacking

$$\mathbf{y}_l^{back} = \mathbf{y}_{l-1}^{back} - \hat{\mathbf{y}}_{l-1}^{back} \qquad \hat{\mathbf{y}}^{for} = \sum_{l=1}^{S\times B} \hat{\mathbf{y}}_l^{for}$$



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. Available at https://arxiv.org/abs/1905.10437.

NBEATSx: Neural basis expansion analysis with exogenous variables

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 https://arxiv.org/abs/2104.05522.

NBEATSx: Motivation and Contributions

Interpretable neural forecasting is largely unexplored, the recent well performing model *Neural Basis Expansion* (NBEATS) lacked exogenous variable inputs. Yet there are several applications where exogenous variables are fundamental.

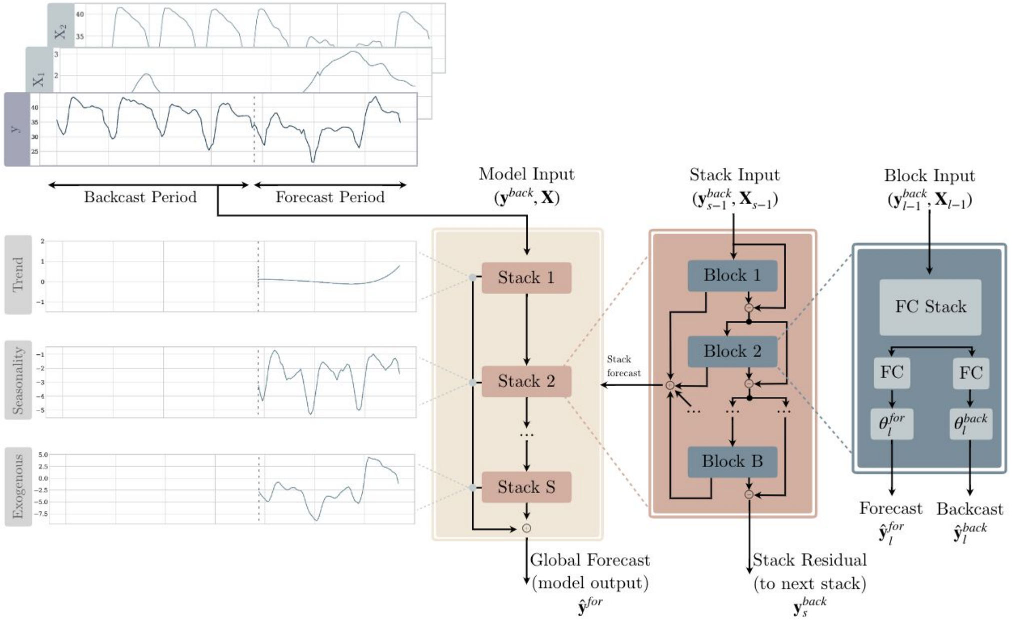
The NBEATSx architecture our contributions include:

1. Convolutional Encoders for the Exogenous Variables.
2. Novel Time Series Decomposition method.
3. State-of-the-Art performance on EPF and healthcare data.

NBEATSx: Architecture

Building blocks of the NBEATSx are structured as a system of multilayer fully connected networks with ReLU based nonlinearities.

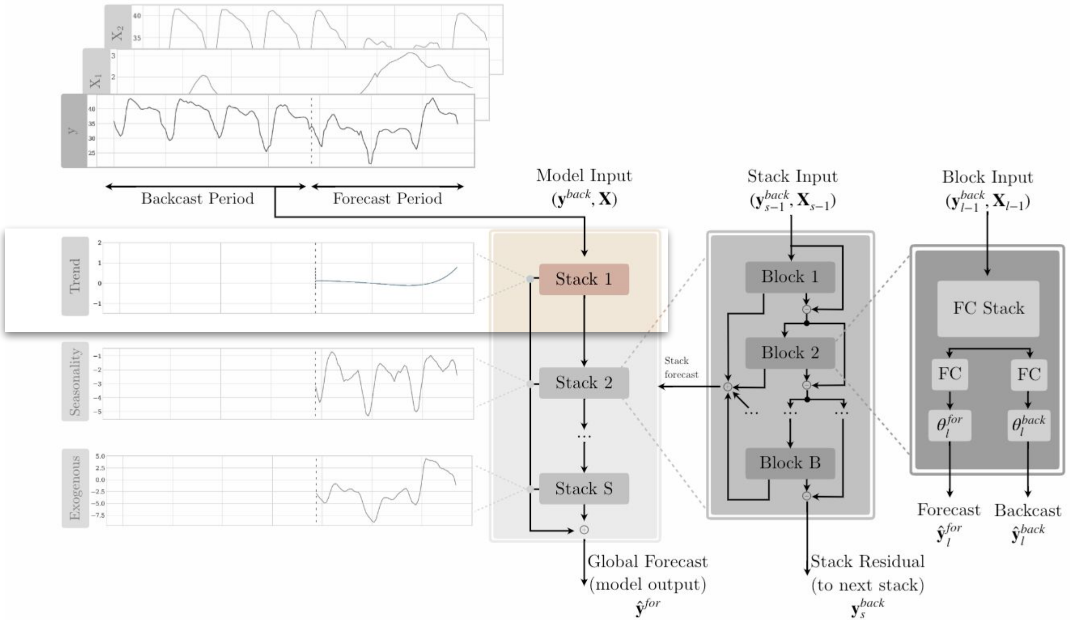
Blocks overlap using the *doubly residual stacking* principle for the backcast and forecast outputs of the l-th block, and the blocks within a stack may share weights.



NBEATSx: Architecture

Trend block

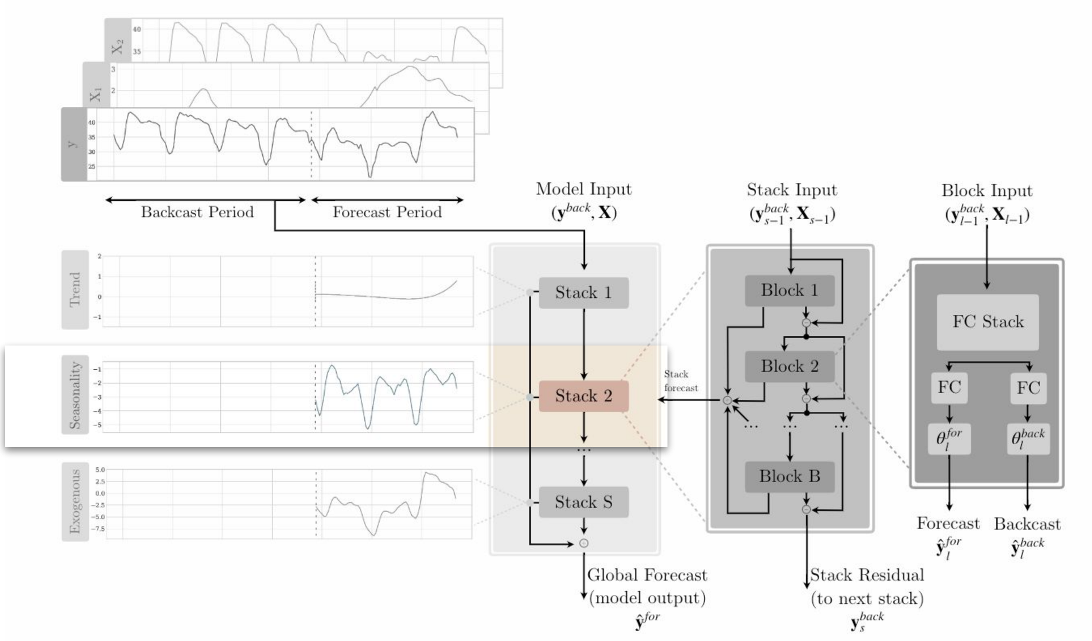
$$\hat{\mathbf{y}}_l^{trend} = \sum_{i=0}^{p} \theta_{l,i}^{trend}\, \mathbf{t}^i \equiv \boldsymbol{\theta}_l^{trend}\, \mathbf{T}$$



NBEATSx: Architecture

Seasonality block

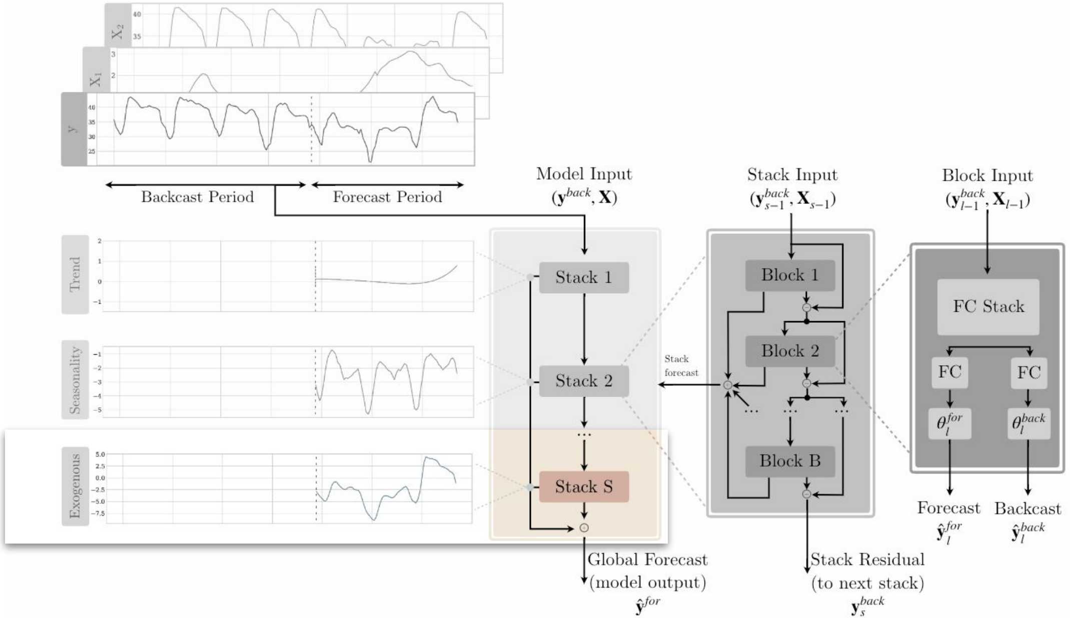
$$\hat{\mathbf{y}}_l^{seas} = \sum_{i=0}^{\lfloor H/2-1 \rfloor} \theta_{l,i}^{seas} \cos(2\pi i \mathbf{t}) + \theta_{l,i+\lfloor H/2 \rfloor}^{seas} \sin(2\pi i \mathbf{t}) \equiv \boldsymbol{\theta}_l^{seas} \, \mathbf{S}$$



NBEATSx: Architecture

Exogenous block,
Convolutional encoders

$$\mathbf{C}_l = \mathrm{TCN}(\mathbf{X}_l) \qquad \hat{\mathbf{y}}_l^{exog} = \sum_{i=1}^{N_c} \theta_{l,i}^{f} C_{l,i} \equiv \boldsymbol{\theta}_l^{f} \mathbf{C}_l$$



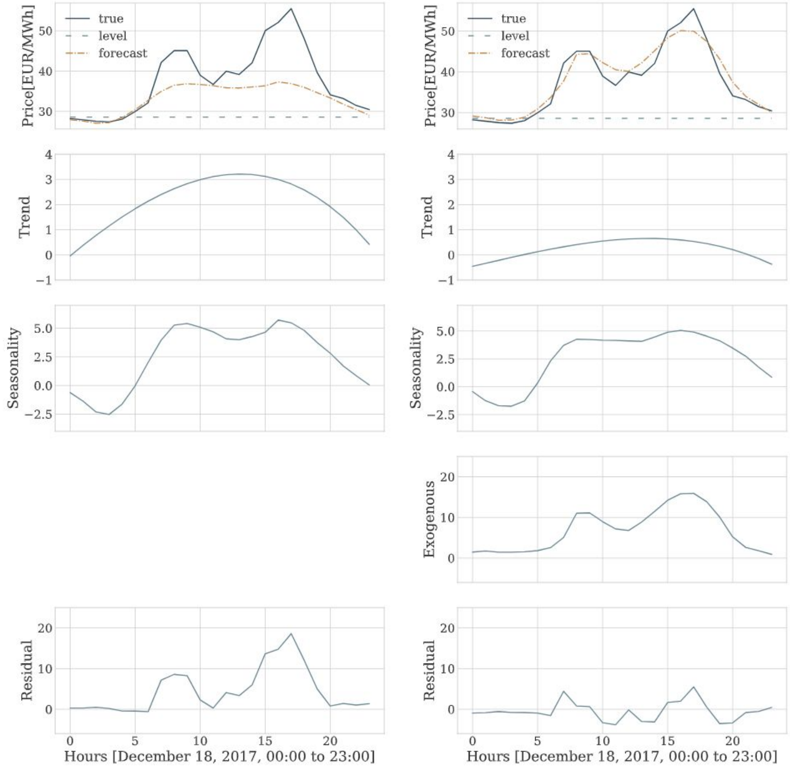
NBEATSx: Decomposition

NP electricity price day-ahead forecasts decomposed using interpretable variants of NBEATS and NBEATSx.

The **top row** of graphs shows the original signal and the level, the latter is defined as the last available observation before the forecast.

The **second row** shows the polynomial trend components, the third and fourth rows display the complex seasonality modeled by nonlinear Fourier projections and the exogenous effects of the electricity load on the price, respectively.

The **bottom row** graphs show the unexplained variation of the signal. The use of electricity load and production forecasts turns out to be fundamental for accurate price forecasting.



NBEATSx: Empirical Results

Table 3: Forecast accuracy measures for day-ahead electricity price predictions of *ensembled models*. The ESRNN and NBEATS do not include time dependent covariates. The reported metrics are *mean absolute error* (MAE), *relative mean absolute error* (rMAE), *symmetric mean absolute percentage error* (sMAPE) and *root mean squared error* (RMSE). The smallest errors in each row are highlighted in bold.

		AR1	ESRNN	NBEATS	ARx1	LEARx	DNN	NBEATSx-G	NBEATSx-I
NP	MAE	2.26	2.09	2.08	2.01	1.74	1.68	**1.58**	1.62
	rMAE	0.71	0.66	0.66	0.63	0.55	0.53	**0.50**	0.51
	sMAPE	6.47	6.04	5.96	5.84	5.01	4.88	**4.63**	4.70
	RMSE	4.08	3.89	3.94	3.71	3.36	3.32	**3.16**	3.27
PJM	MAE	3.83	3.59	3.49	3.53	3.01	**2.86**	2.91	2.90
	rMAE	0.79	0.74	0.72	0.73	0.62	**0.59**	0.60	0.60
	sMAPE	14.5	14.12	13.57	13.64	11.98	**11.33**	11.54	11.61
	RMSE	6.24	5.83	5.64	5.74	5.13	5.04	5.02	**4.84**
EPEX-BE	MAE	7.2	6.96	6.84	7.19	6.14	**5.87**	5.95	6.11
	rMAE	0.88	0.85	0.83	0.88	0.75	**0.72**	0.73	0.75
	sMAPE	16.26	15.84	15.80	16.11	14.55	**13.45**	13.86	14.02
	RMSE	18.62	16.84	17.13	18.07	15.97	15.97	**15.76**	15.80
EPEX-FR	MAE	4.65	4.65	4.74	4.56	3.98	3.87	3.81	**3.79**
	rMAE	0.78	0.78	0.80	0.76	0.67	0.65	**0.64**	**0.64**
	sMAPE	13.03	13.22	13.30	12.7	11.57	10.81	**10.59**	**10.69**
	RMSE	13.89	11.83	12.01	12.94	10.68	11.87	11.50	**11.25**
EPEX-DE	MAE	5.74	5.60	5.31	4.36	3.96	3.41	3.31	**3.29**
	rMAE	0.71	0.70	0.66	0.54	0.49	0.42	**0.41**	**0.41**
	sMAPE	21.37	20.97	19.61	17.73	15.75	14.08	**13.99**	**13.99**
	RMSE	9.63	9.09	8.99	7.38	7.08	5.93	5.72	**5.65**

DeepMIDAS: Deep Mixed Data Sampling Regression for Long Multi-Horizon Forecasting

Challu, C., Olivares, K. G., Welter, G., and Dubrawski, A. *DeepMIDAS: Deep Mixed Data Sampling Regression for Long Multi-Horizon Time Series Forecasting*. Submitted to the 38th International Conference on Machine Learning Time Series Workshop, PMLR 139, 2021. Working Paper version available at https://arxiv.org/abs/2106.05860.

DeepMIDAS: Motivation and Contributions

Long-horizon forecasting remains a challenging task for neural networks due to,

1. Volatility/Variance of multi-horizon models
2. Long-term dependencies
3. Computational complexity

Our work makes the following contributions:

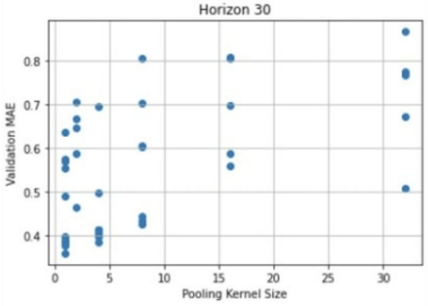
1. Mixed Data Sampling with pooling layers
2. Smoothness Regularization with temporal interpolation
3. DeepMIDAS model, up to 10% RMSE improvement with 70% less parameters.

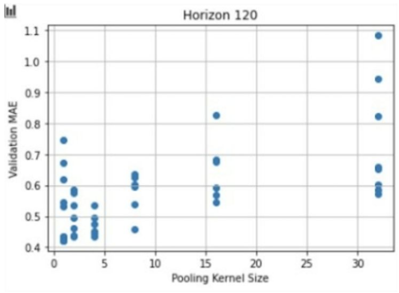
DeepMIDAS: Mixed Data Sampling

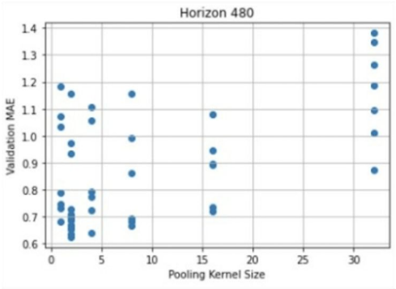
- In order to capture long time dynamics while not over-parametrizing we propose the use of max-pooling layers. These sub-sampling layers effectively reduce the number of parameters, limiting the memory footprint and the amount of computation, while maintaining the original receptive field.

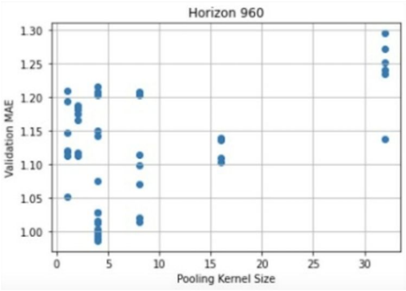
- Different kernel sizes induce mixed frequencies of the inputs on the backcast window.

- Optimal kernel size (selected during hyperparameter tuning) monotonically increases with horizon.









DeepMIDAS: Smoothness Regularization

For most joint multi-horizon forecasting models the outputs' cardinality corresponds to the horizon's dimension. We introduce the *expressivity ratio* r_l which controls the number of unconstrained outputs pero unit of time, so now,

$$|\boldsymbol{\theta}_l^f| = \lceil r_l H \rceil$$

To recover the original dimension, we rely on *temporal interpolation* (linear for now),

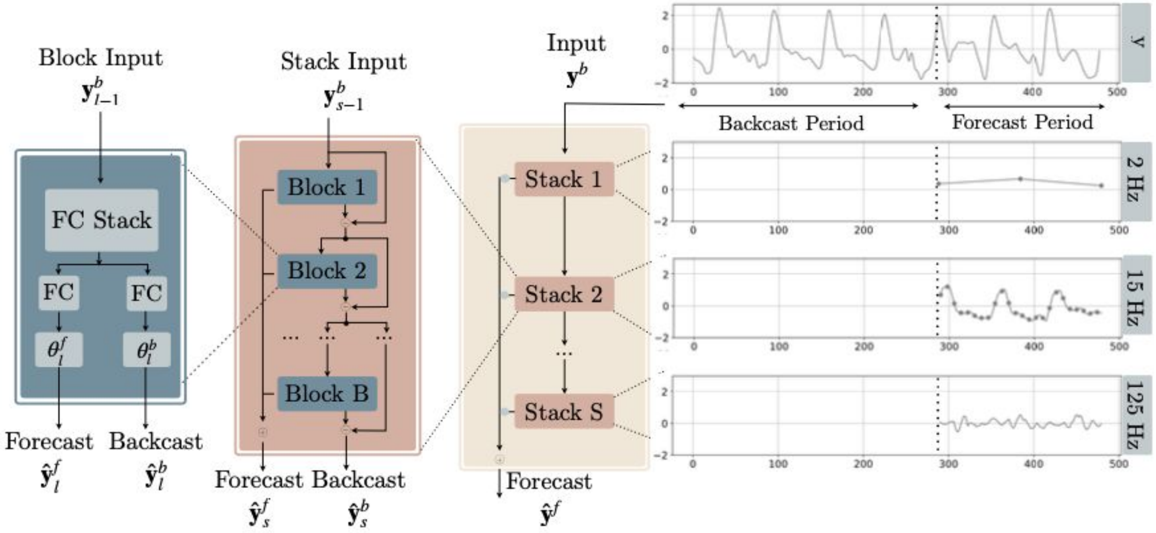
$$\hat{y}_{l,t}^f = \left(\theta_{l,t_1}^f + \left(\frac{\theta_{l,t_2}^f - \theta_{l,t_1}^f}{\frac{H}{\lceil r_l H \rceil}} \right) (t - t_1) \right)$$

We use exponentially increasing expressivity ratios through the depth of the architecture blocks to reduce the model complexity.

$$\mathcal{O}\left((H(1 - r^B)/(1 - r)) \right) \quad \text{vs} \quad \mathcal{O}\left(HB \right)$$

DeepMIDAS: Architecture

- We combine the *mixed data sampling* and *smoothness regularization* techniques with the NBEATS architecture.
- We allow for different pooling kernel sizes and expressivity ratios between *stacks*.
- Each *stack* will specialize in learning temporal dependencies on different frequencies of the data.



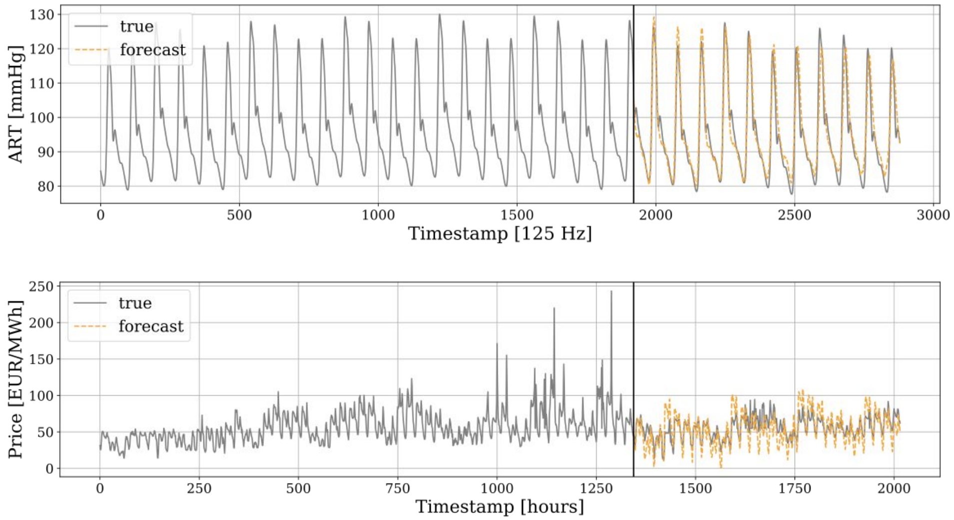
DeepMIDAS: Datasets

Healthcare

- High-frequency (125hz) vital signs collected from intensive-care units at UPMC over three years.
- Arterial Pressure (ART) and Photoplethysmogram (PLETH) vital-signs.
- 98 patients, each with a 90-minutes window.

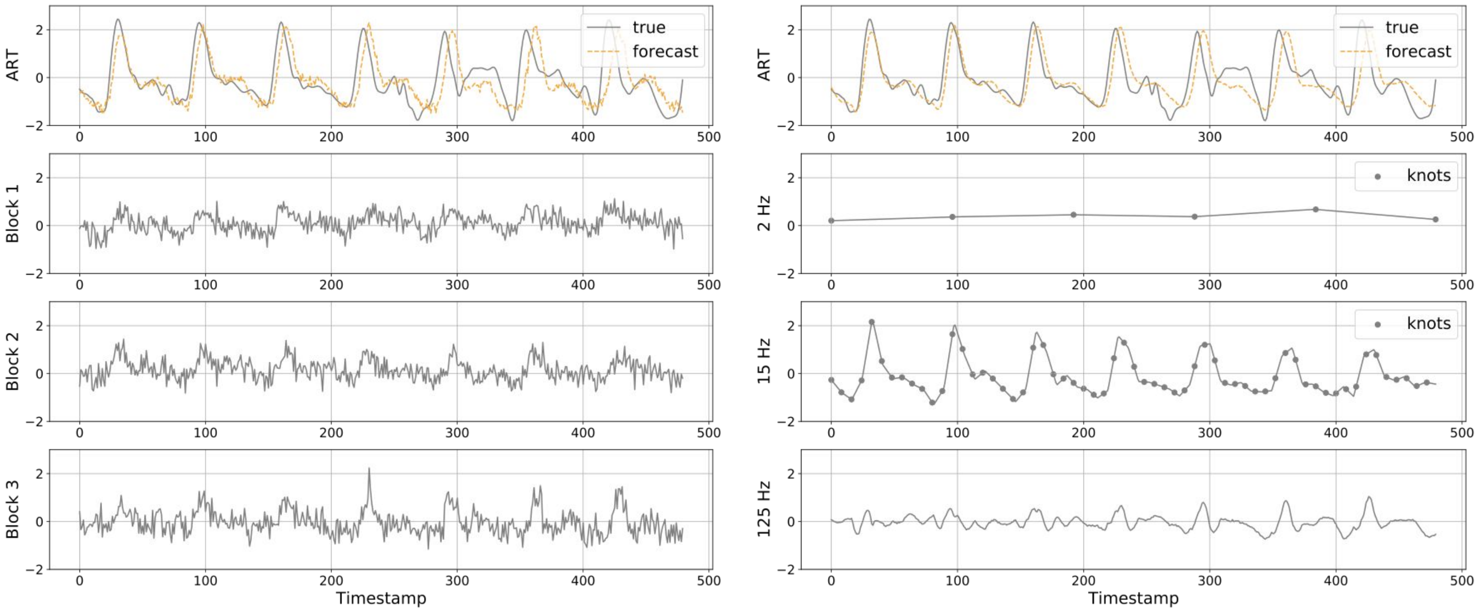
Electricity Price

- Hourly data of electricity price on 5 markets: NP, PJM, BE, FR and DE.
- Data from 01/2013 to 12/2018.



Arterial pressure (ART) and French European Power Exchange electricity price, along with the long-multi-horizon forecasts of DeepMidas. High-Frequency data often poses challenging forecasting tasks when series display heterogeneous behavior across series and signals have non-stationary dynamics.

DeepMIDAS: Interpretable Forecast



(a) NBEATS-G

(b) DeepMIDAS

Arterial Pressure (ART) and five hundred steps ahead forecasts using NBEATS-G and DeepMIDAS. The top row shows the original signal. The second, third and fourth rows show the forecast components for the each model's block, in the case of sub-Figure (b) each block specializes on different frequencies, contrary to the outputs of NBEATS-G on sub-Figure (a) that are unintelligible.

DeepMIDAS: Empirical Results

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	13.17	13.03	12.48	**11.96**	12.28	12.40
		MAE	6.33	6.62	5.89	**5.25**	5.30	5.44
	480	RMSE	18.17	17.88	22.40	17.48	16.93	**16.71**
		MAE	9.21	9.12	13.15	8.73	7.80	**7.68**
	960	RMSE	21.24	21.20	21.87	22.11	18.64	**18.01**
		MAE	12.23	11.84	12.04	14.17	9.97	**9.87**
PLETH	120	RMSE	0.054	0.056	0.060	0.051	0.050	**0.050**
		MAE	0.033	0.034	0.035	0.029	**0.028**	**0.028**
	480	RMSE	0.067	0.071	0.106	0.065	0.063	**0.061**
		MAE	0.045	0.046	0.074	0.043	**0.039**	0.041
	960	RMSE	0.079	0.081	0.078	0.082	**0.073**	**0.073**
		MAE	0.054	0.055	0.058	0.059	0.049	**0.046**
EPF	24	RMSE	9.84	9.66	9.55	9.49	9.34	**9.04**
		MAE	6.02	5.90	5.81	5.89	5.65	**5.56**
	336	RMSE	12.94	12.99	**12.84**	13.29	13.00	**12.84**
		MAE	8.80	8.80	8.75	9.05	8.76	**8.74**
	672	RMSE	18.03	17.45	17.33	16.92	18.11	**15.88**
		MAE	13.30	13.32	13.01	12.61	13.43	**11.50**

THANK YOU