High-Resolution Peak Demand Estimation Using Generalized Additive Models and Deep Neural Networks

Jonathan Berrisch, Michał Narajewski, Florian Ziel

University of Duisburg-Essen
2023-09-13

Motivation

High Resolution Peak Demand Estimation Challenge

- Organized by Western Power Distribution and Catapult Energy Systems
- Does limited high-resolution monitoring help estimate future high-resolution peak loads?

The Objective:

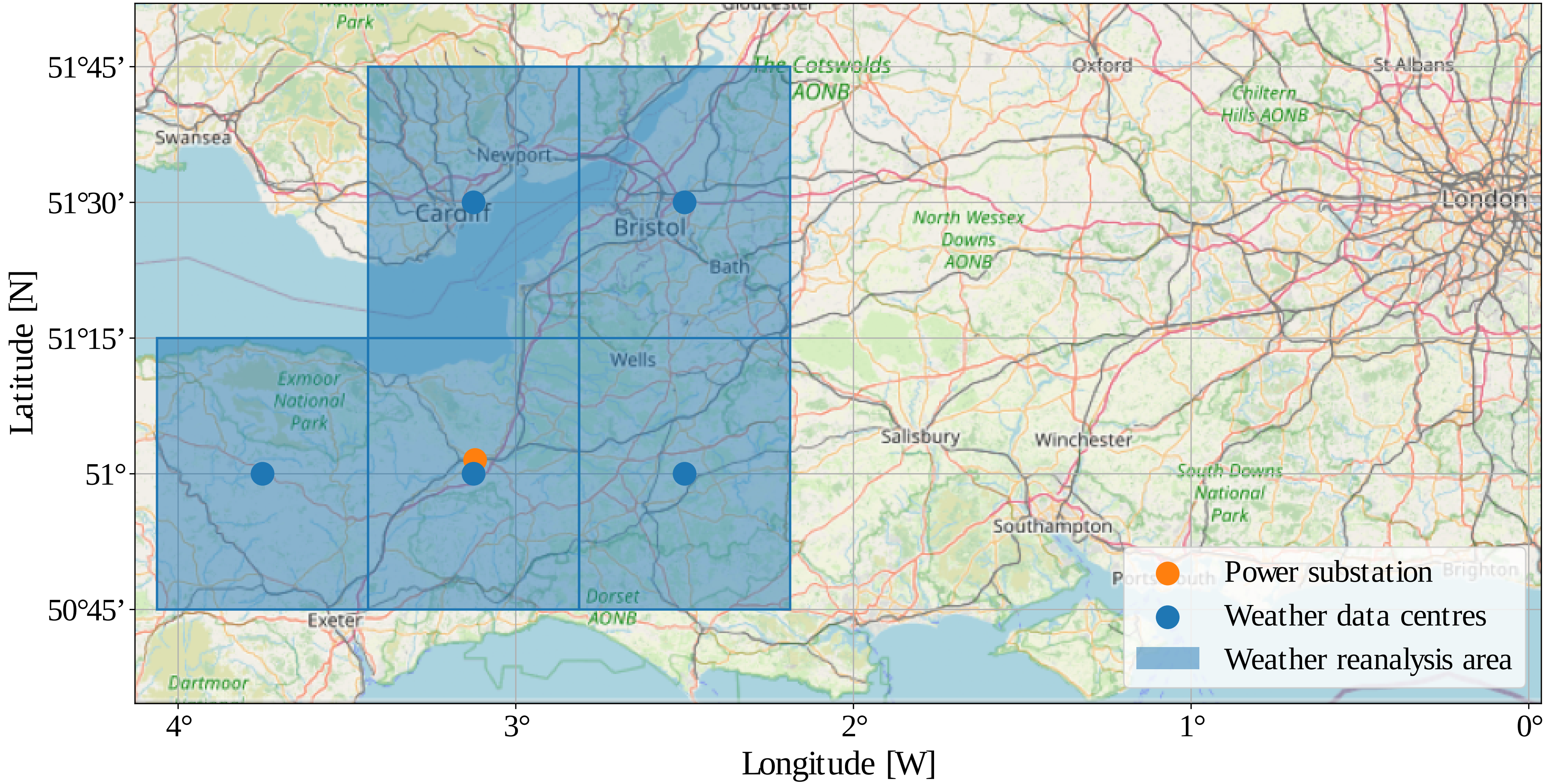
- Estimate minimum and maximum electricity load values (one-minute resolution)
- Given data with only a 30-minute resolution
- One single substation, every half-hour of September 2021

Data:

- From Nov. 2019 to Sept. 2021 (30-minute resolution)
- MERRA-2 weather reanalysis data from five locations close to the substation

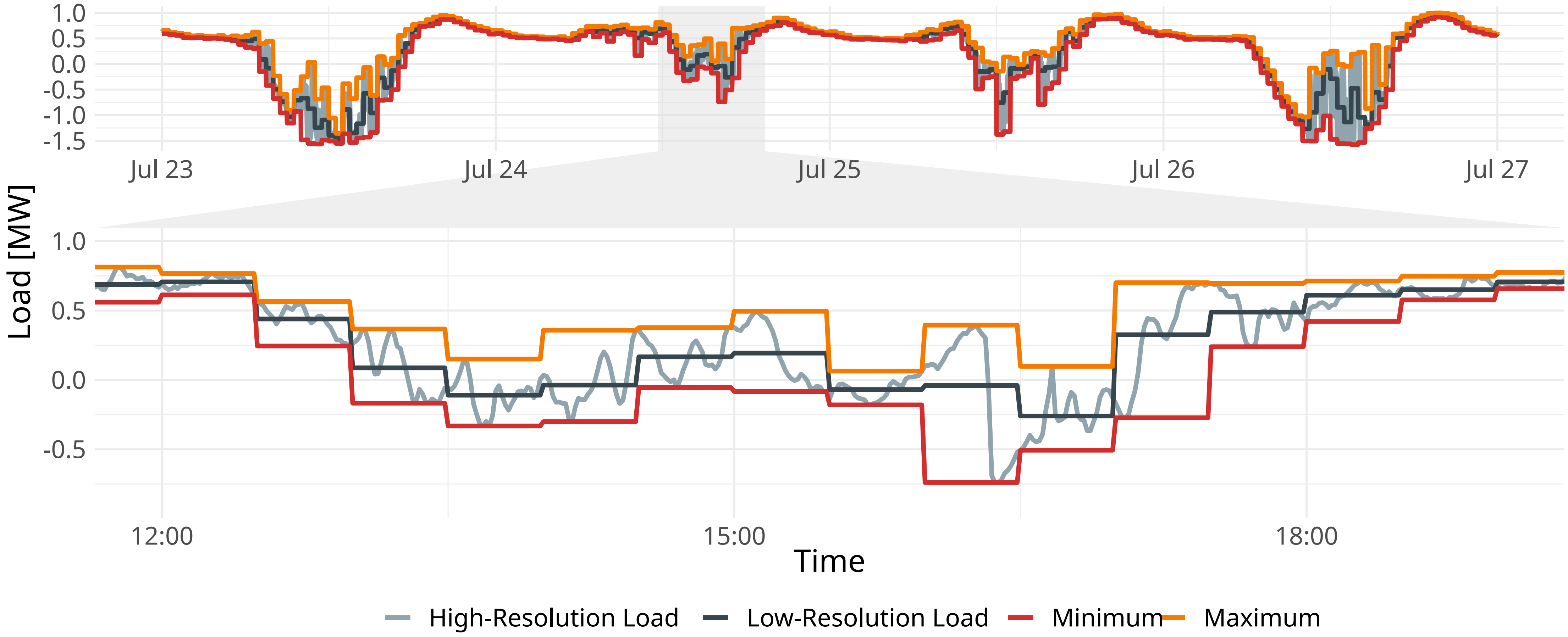
Motivation

Location Overview



Motivation

Data Overview



Data

Targets: $\Delta_t^{\min}$ and $\Delta_t^{\max}$

Possible explanatory variables:

The *half-hourly* load: L_t

Discrete second order central difference (DSOCD):

$L_t'' = L_{t-1} - 2L_t + L_{t+1}$

Deterministic components (to capture potential seasonal characteristics)

- Daily D_t number of hours in a day
- Weekly W_t number of hours in a week
- Annual A_t number of hours in a meteorological year with 365.24 days

Weather Inputs: Temperature, Windspeed (North / East), Solar, Humidity

The Figure (next slide) shows how these variables correlate.

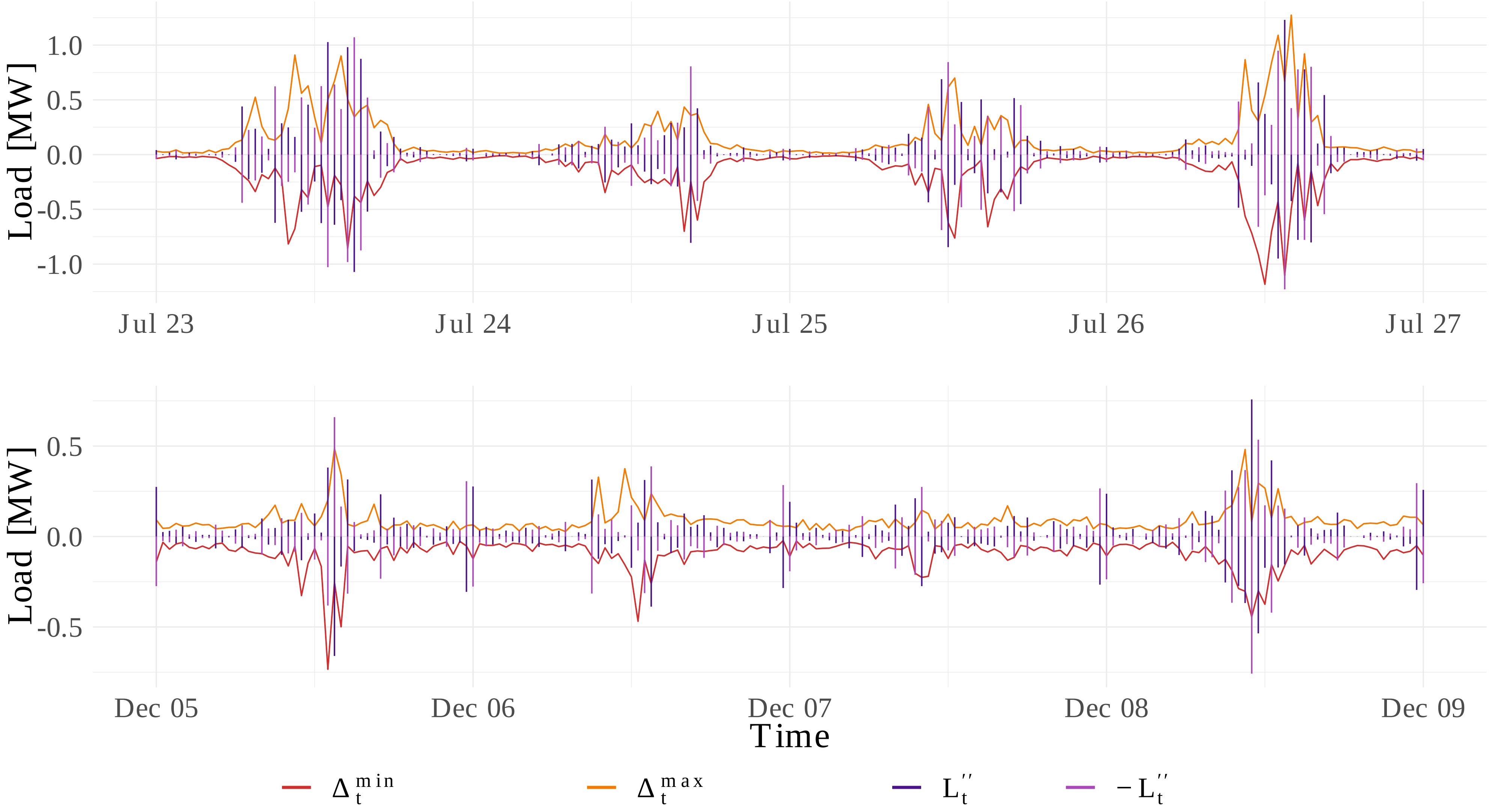
- Lower triangle:
  - Pearson's correlation
- Upper triangle:
  - Distance correlation

Distance correlation: non-linear dependency measure that takes values in $[0, 1]$ and characterizes stochastic independence Székely, Rizzo, and Bakirov (2007).

Correlation

	$\Delta_t^{\min}$	$\Delta_t^{\max}$	L_t	L''_t	$\widetilde{L}''_t$	Temp	Solar	WindN	WindE	Press	Humid	D_t	W_t	A_t
$\Delta_t^{\min}$		0.83	0.54	0.46	0.48	0.29	0.65	0.05	0.19	0.05	0.11	0.30	0.05	0.11
$\Delta_t^{\max}$	-0.70		0.64	0.47	0.49	0.29	0.68	0.08	0.16	0.05	0.09	0.31	0.05	0.12
L_t	0.43	-0.59		0.37	0.39	0.53	0.82	0.13	0.07	0.21	0.35	0.33	0.05	0.30
L''_t	0.04	0.21	-0.22		0.97	0.16	0.38	0.03	0.11	0.03	0.05	0.20	0.03	0.06
$\widetilde{L}''_t$	0.04	0.22	-0.21	0.99		0.18	0.42	0.04	0.12	0.03	0.06	0.21	0.03	0.06
Temp	-0.26	0.27	-0.55	0.01	0.01		0.51	0.16	0.13	0.20	0.88	0.20	0.03	0.46
Solar	-0.57	0.64	-0.84	0.04	0.04	0.53		0.10	0.09	0.17	0.22	0.37	0.05	0.20
WindN	0.02	-0.07	0.13	0.00	0.00	0.10	-0.10		0.26	0.32	0.21	0.04	0.04	0.10
WindE	-0.17	0.12	0.06	0.00	0.00	0.09	-0.02	0.26		0.32	0.15	0.08	0.05	0.11
Press	0.01	0.02	-0.19	0.00	0.00	0.05	0.16	-0.28	-0.32		0.19	0.01	0.04	0.22
Humid	-0.08	0.04	-0.30	0.00	0.00	0.89	0.23	0.18	0.09	-0.05		0.05	0.03	0.48
D_t	-0.04	0.05	0.13	0.05	0.00	0.14	-0.01	0.01	0.03	0.01	0.05		0.14	0.00
W_t	-0.01	0.00	0.03	0.01	0.00	0.02	-0.01	0.03	-0.04	-0.03	0.02	0.14		0.01
A_t	0.03	-0.04	0.00	0.00	0.00	0.26	-0.03	-0.01	0.02	-0.21	0.31	0.00	-0.01	

Data



Load [MW]
1.0
0.5
0.0
-0.5
-1.0
Jul 23
Jul 24
Jul 25
Jul 26
Jul 27
Load [MW]
0.5
0.0
-0.5
Dec 05
Dec 06
Dec 07
Dec 08
Dec 09
Time
Δ_t^{min}
Δ_t^{max}
L_t''
$-L_t''$

Feature Set and Models

Selected features based on preliminary analysis:

Variable type	Included feature	Number
Lagged load	L_{t-1}, L_t, L_{t+1}	3
Lagged DSOCD load	$\widetilde{L}''_{t-4}, \ldots, \widetilde{L}''_{t+4}$	9
Weather inputs	Temp, Solar, WindN, WindE, Press, Humid	6
Seasonal inputs	D_t, W_t, A_t	3

Considered Models:

- Generalized Additive Model (GAM)
- Multilayer Perceptron Network (MLP)
- Combinations of the above

Competition Benchmark:

Naive

$$L_t^{\max} = L_t$$

$$L_t^{\min} = L_t$$

Modelling Approach: GAM

Generalized Additive Model (GAM)

$$\Delta_t^m = \sum_{i=1}^{L} f_i(X_{t,1}, \ldots, X_{t,N}) + \epsilon_t \quad (1)$$

where $m \in \{\min, \max\}$

Traditional framework using cubic B-splines

GAM.full Specification (with all 2-way interactions):

$$\Delta_t^m = \sum_{i=1}^{N} b_{k_0}(X_{t,i}) + \sum_{i=1}^{N} \sum_{j=1, j>i}^{N} b_{k_1,k_2}(X_{t,i}, X_{t,j}) + \epsilon_t \quad (2)$$

b_{k_0} and b_{k_1,k_2} denote univariate and bivariate splines with k_0, and (k_1, k_2) knots. Tensor interaction splines -> only capture the joint effects.

We set $k_0 = 27$ and $k_1 = k_2 = 9$. Thus, linear terms are specified by 27 parameters and bivariate terms by 81 parameters.

GAM.red:

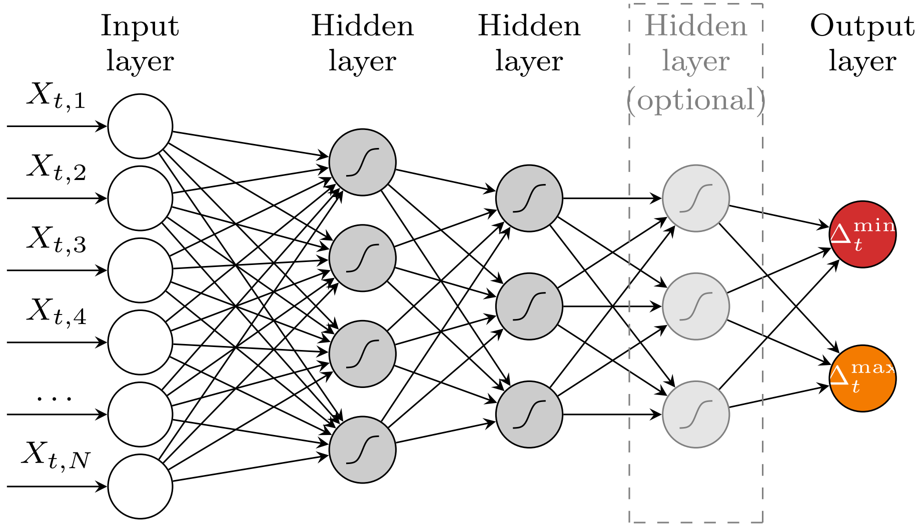
Interactions with L_t, $\widetilde{L}_t''$ and Solar_t.

GAM.simple:

$$\Delta_t^m = b_{k_0}(L_t) + b_{k_0}(\widetilde{L}_t'') + b_{k_0}(\text{Solar}_t) + \epsilon \quad (3)$$

Modelling Approach: MLP

Multilayer Perceptron Network:



Hyperparameters are tuned using OPTUNA Python package Akiba, Sano, Yanase, Ohta, and Koyama (2019)

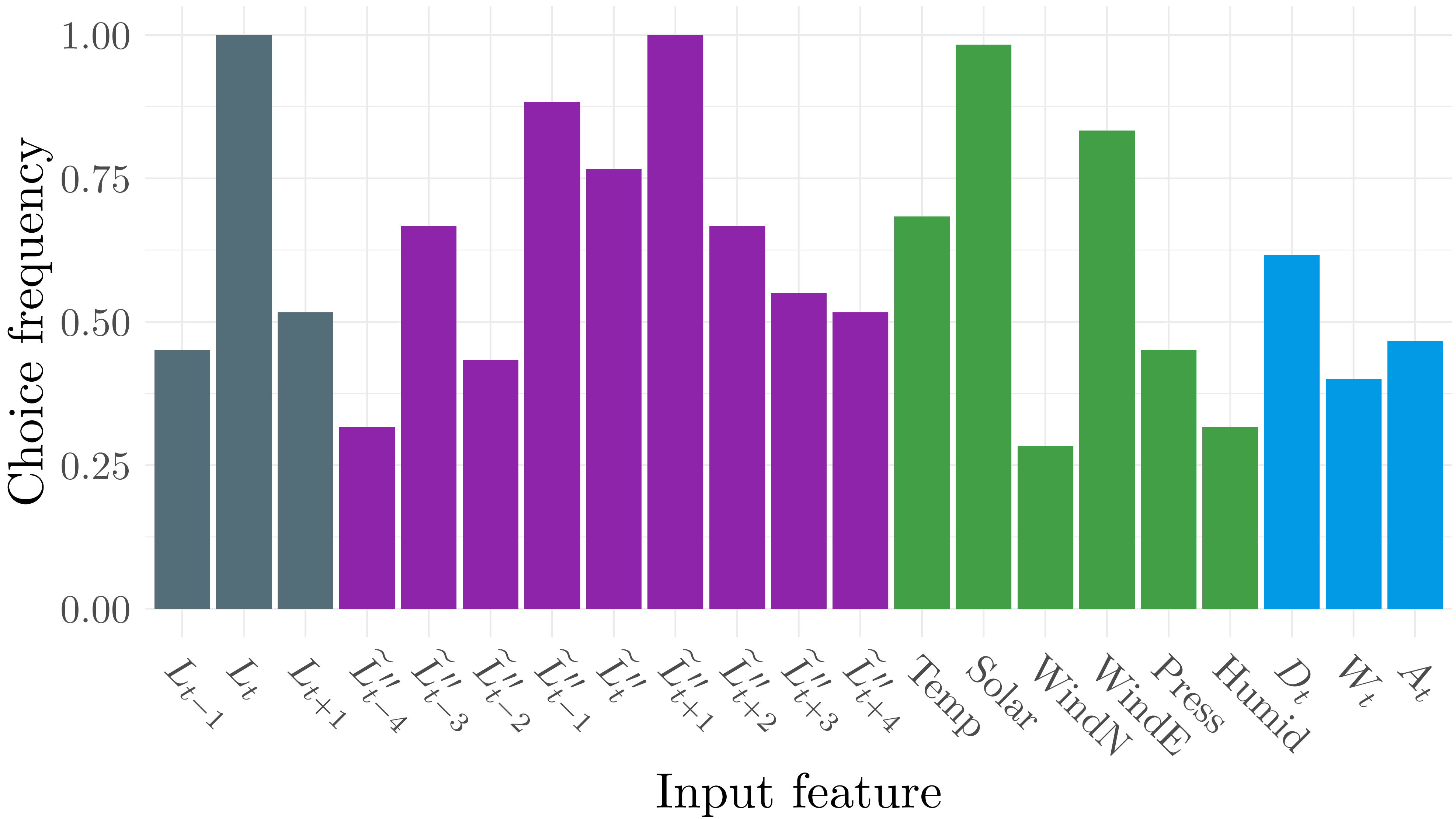
- Input feature selection
- Number of hidden layers -- either 2 or 3
- Dropout layer -- whether to use it after the input layer and, if yes, at what rate.
- Activation functions in the hidden layers: elu, relu, sigmoid, softmax, softplus, and tanh
- Number of neurons in the hidden layer drawn on an exp-scale from $[4, 128]$
- L_1 regularization -- whether to use it on the hidden layers and, if yes, at what rate.
- Learning rate for the Adam optimization algorithm drawn on an exp-scale from $(10^{-5}, 10^{-1})$ interval

GAM Parameter Significance

Min	EDF	F	Max	EDF	F
L_t	9.5	8.9	L_t	10.8	17.6
$\widetilde{L}''_{t-1}$	8.8	4.4	$\widetilde{L}''_{t-2}$	6.7	4.9
$\widetilde{L}''_{t+1}$	8.4	4.8	L_{t-1}	6.3	6.9
$\widetilde{L}''_{t}$	8.2	4.6	L_{t+1}	5.4	5.0
Humid	6.1	3.7	D_t	4.8	3.1
WindE	5.1	8.9	Temp	4.1	11.5
A_t	4.7	4.9	Solar	4.0	5.5
WindN	4.2	8.5	WindE	3.8	5.8
Temp	3.5	3.6	$\widetilde{L}''_{t+4}$	3.2	2.1
L_{t-1}	3.3	1.3	$\widetilde{L}''_{t+1}$	3.1	2.2

Min	EDF	F	Max	EDF	F
L_t, A_t	26.7	7.8	L_t, A_t	37.8	28.1
L_t, Solar	20.7	11.2	L_t, D_t	25.1	20.6
L_t, D_t	18.3	8.2	Solar, A_t	15.4	6.4
L_t, L_{t+1}	17.1	4.0	$\widetilde{L}''_t$, $\widetilde{L}''_{t-1}$	14.9	6.5
$\widetilde{L}''_t$, $\widetilde{L}''_{t-1}$	13.5	2.5	Solar, D_t	11.0	8.5
L_t, Temp	13.3	3.1	$\widetilde{L}''_t$, $\widetilde{L}''_{t+1}$	11.0	2.4
L_t, L_{t-1}	13.0	4.9	$\widetilde{L}''_t$, $\widetilde{L}''_{t+2}$	10.7	2.8
$\widetilde{L}''_t$, $\widetilde{L}''_{t+1}$	12.1	2.3	L_t, Solar	10.0	2.9
L_t, W_t	11.7	3.0	$\widetilde{L}''_t$, Temp	9.5	3.4
$\widetilde{L}''_t$, $\widetilde{L}''_{t+2}$	11.4	2.9	L_t, WindE	9.3	2.5

MLP Feature Importance



Study Design and Evaluation

Rolling Window Forecasting Study:

- Length: 12 Months (10/2020 - 09/2021)
- 1-Month shifts
- Evaluation by RMSE

Competition Design:

- Only evaluating 09/2021
- Rank base on Score (relative RMSE)

$$\text{Score} = \text{RMSE}(\text{Model})/\text{RMSE}(\mathbf{naive}). \quad (4)$$

Considered Models:

- **GAM.full**
- **GAM.red**
- **DNN**
- **naive**
- **Combination of GAM.full, GAM.red, and DNN**

Two additional GAM models for diagnostic purposes:

- **GAM.simple**
- **GAM.no.Weather**

Results

Overall	20/10	20/11	20/12	21/1	21/2	21/3	21/4	21/5	21/6	21/7	21/8	21/9	Avg
GAM.full	.1239 (56.3)	.0703 (58.5)	.0497 (56.6)	.0532 (57.6)	.0988 (55.4)	.1202 (57.8)	.1447 (57.2)	.1746 (56.7)	.1368 (56.3)	.1537 (53.9)	.1383 (60.2)	.1163 (56.9)	.1150 (56.9)
GAM.red	.1241 (56.2)	.0700 (58.7)	.0496 (56.6)	.0536 (57.3)	.0994 (55.2)	.1201 (57.8)	.1467 (56.6)	.1750 (56.6)	.1385 (55.8)	.1554 (53.4)	.1374 (60.5)	.1162 (56.9)	.1155 (56.7)
DNN	.1230 (56.6)	.0704 (58.4)	.0507 (55.7)	.0545 (56.6)	.1027 (53.7)	.1251 (56.1)	.1514 (55.2)	.1653 (59.0)	.1474 (52.9)	.1538 (53.9)	.1422 (59.1)	.1225 (54.6)	.1174 (56.0)
Combination	.1221 (56.9)	.0689 (59.3)	.0491 (57.1)	.0527 (58.0)	.0979 (55.8)	.1193 (58.1)	.1443 (57.3)	.1684 (58.2)	.1365 (56.4)	.1512 (54.6)	.1371 (60.6)	.1164 (56.9)	.1137 (57.4)
GAM.noWeather	.1301 (54.1)	.0707 (58.2)	.0508 (55.6)	.0533 (57.5)	.1058 (52.3)	.1253 (56.0)	.1471 (56.5)	.1822 (54.8)	.1401 (55.3)	.1584 (52.5)	.1424 (59.1)	.1249 (53.7)	.1193 (55.3)
GAM.simple	.1531 (46.0)	.0940 (44.5)	.0584 (49.0)	.0672 (46.5)	.1267 (42.9)	.1500 (47.3)	.1797 (46.8)	.2093 (48.1)	.1733 (44.7)	.1873 (43.8)	.1671 (52.0)	.1413 (47.6)	.1423 (46.7)
Naive	.2833	.1693	.1144	.1255	.2217	.2847	.3380	.4029	.3131	.3334	.3478	.2699	.2670

Wrap-Up

Estimating high-resolution elecricity peak demand using lower-resolution data:

- **GAM.full** and **GAM.red** perform similar
- **DNN** beats **GAM.full** in some Months
- **Combination of GAM.full, GAM.red, and DNN** performs best
- Weather variables improve the skill score by 1.5 percentage points on average
- **DNN** performs better at predicting maximum peak loads

Berrisch, Narajewski, and Ziel (2023)

We won the competition.

- 42.6% vs. 43.6% (second place)
- Using slightly different model

berrisch.biz/slides/23_09_stat_woche

References 1

Akiba, T., S. Sano, T. Yanase, et al. (2019). "Optuna: A next-generation hyperparameter optimization framework". In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. , pp. 2623–2631.

Berrisch, J., M. Narajewski, and F. Ziel (2023). "High-resolution peak demand estimation using generalized additive models and deep neural networks". In: Energy and AI 13, p. 100236.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). "Measuring and testing dependence by correlation of distances". In: The annals of statistics 35.6, pp. 2769–2794.