Project: The Role Of Holiday In Probabilistic Load Forecasting: A Spanish Case Study

Authors: Öcal Kaptan & Ly Le Thi

Supervisors: Prof. Dr. Florian Ziel

Essen, July 18th, 2024

1- Introduction

- Problem:

➢ Probabilistic load forecasting for Spanish electricity demand

➢ Horizon: 168 hours ahead

➢ Focus: Holiday effects

- Methods: Generalized Additive Models for Location Scale and Shape (GAMLSS), Distributional Neural Networks (DDNN) and AutoRegressive with Exogenous Inputs (ARX) as baseline model

2. Data – Regressors selection

- 3 key variables for load forecasting: calendar features, weather features, and socio-economic features
- Utilized data in this project:

- Day-ahead Load of Spain. Unit: Gigawatt (GW).
- Calendar data: days of the week and hours of the day.
- Holiday data: 35 public and regional holidays in Spain
- Weather data: forecast and actual temperature (Unit: Degrees Celsius (°C))

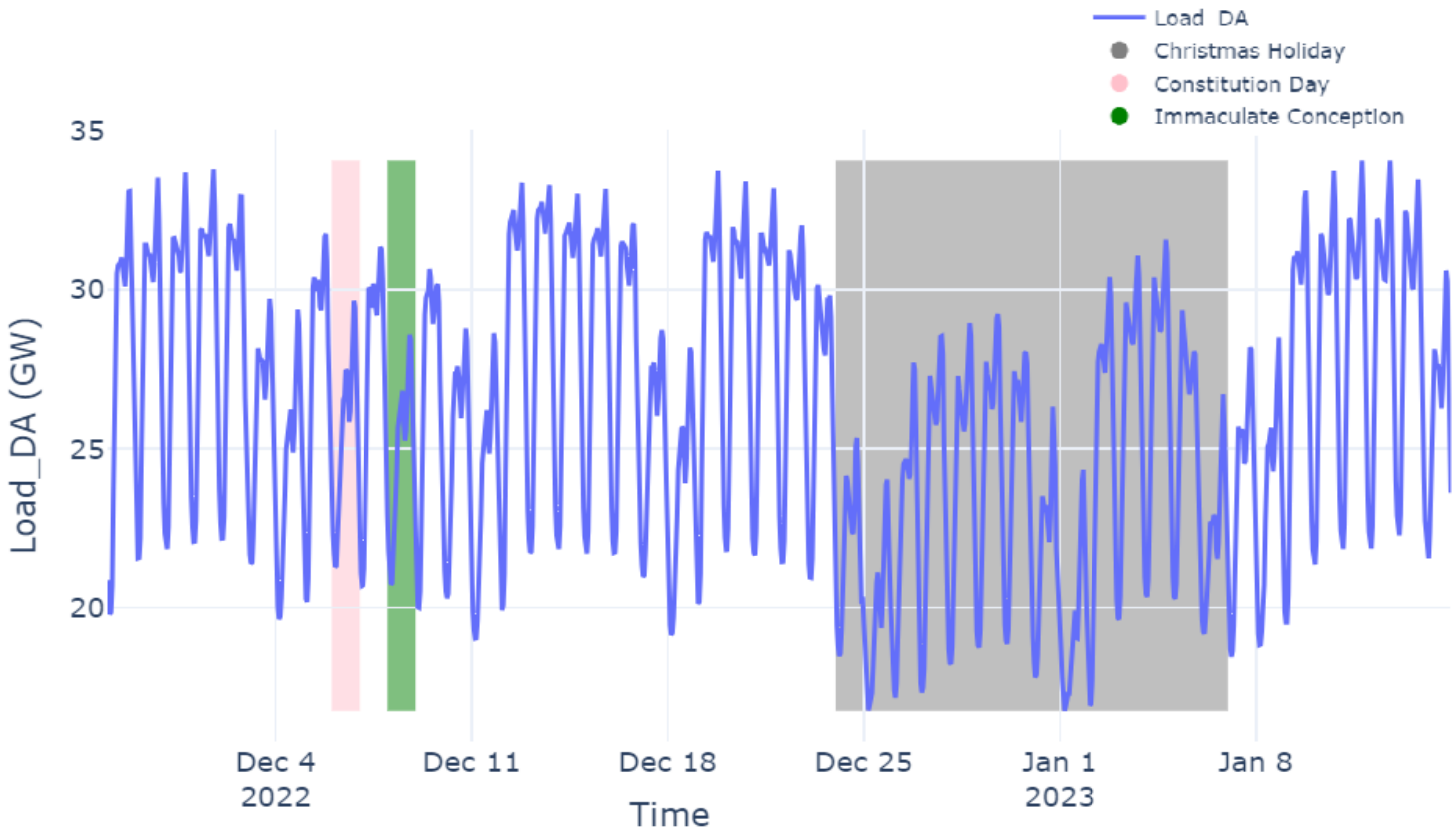
- Data range: 2018-01-01 at 00:00:00
  to 2024-06-01 at 21:00:00

Training: 2018-01-01 00:00:00 to 2022-12-01 08:00:00
Validation: 2022-12-01 09:00:00 to 2023-12-01 08:00:00
Testing: 2023-12-01 at 09:00:00 to 2024-06-01 at 08:00:00

2. Data – Holiday affect

UNIVERSITÄT
DUISBURG
ESSEN
Open-Minded

- Public Holidays: nationwide, affecting all counties
  - ➢ Flexible holidays: fall on the same weekday each year: Easter Monday, Maundy Thursday,...
  - ➢ Fixed holidays: Fall on the same days of the year regardless of the day of the week, e.g. Christmas Day (25/12), Labour Day (01/05),...
- Regional Holidays: specific to certain states or regions, e.g. Easter Monday, Day of the Valencian Community,...
- Christmas holiday period: 24/12 – 06/01



2. Data – Holidays calculation

Public Holidays

- Flexible holidays: same weekday each year, disrupts the usual weekly load.For D ∊ Flex, 6 +24 +6 around public holidays are set to 1, starting from 18:00.

$$B_k^D(t) = \begin{cases} 1 & \text{if } k \leq \text{HoD} \\ 0, & \text{otherwise} \end{cases}$$

- Fixed holidays: same day every year, varies depending on the day of the week. For F ∊ Fix, If holidays are on:

- Sunday: 0 (low-level load day), -Tue, Wed, Thu: C(t) is 1 (high-level load day)
- Mon, Sat, Sun: C(t) is between 0 and 1

HoD(t) is the hours of the day counting from 1, 2, . . . , 36 at time point t, starting from 18:00

$$C(t) = \max\left(1 - \frac{\text{high-level load target}(t) - \text{actual load target}(t)}{\text{high-level load target}(t) - \text{low-level load target}(t)}, 1\right)$$

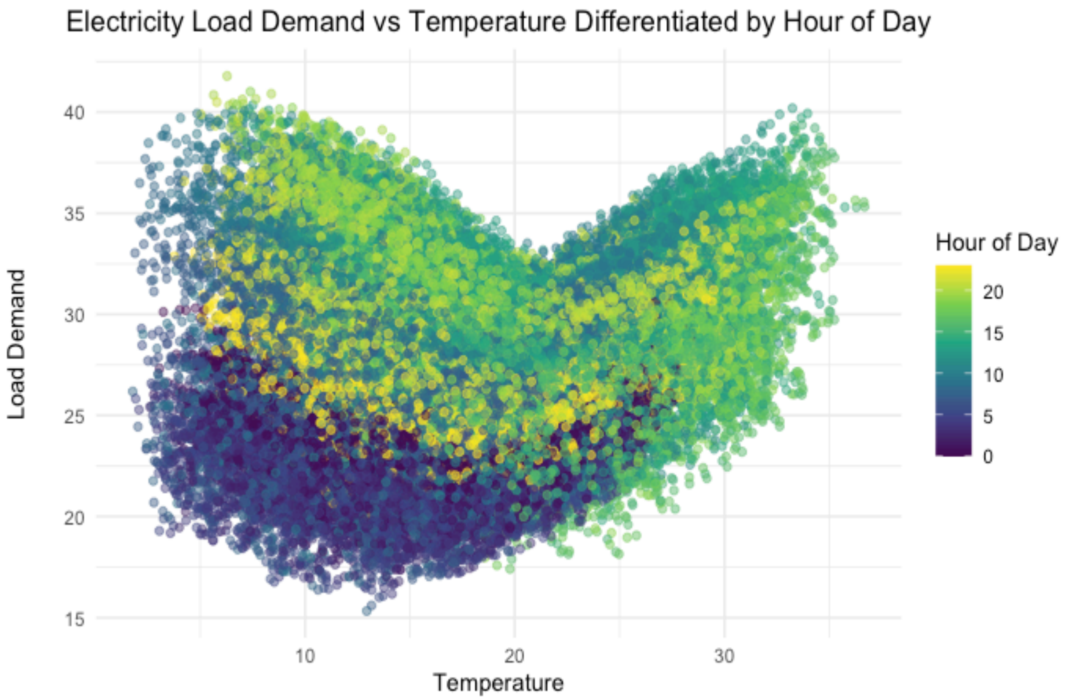
$$B_k^F(t) = \begin{cases} C(t) & \text{if } k \leq \text{HoD} \\ 0, & \text{otherwise} \end{cases}$$

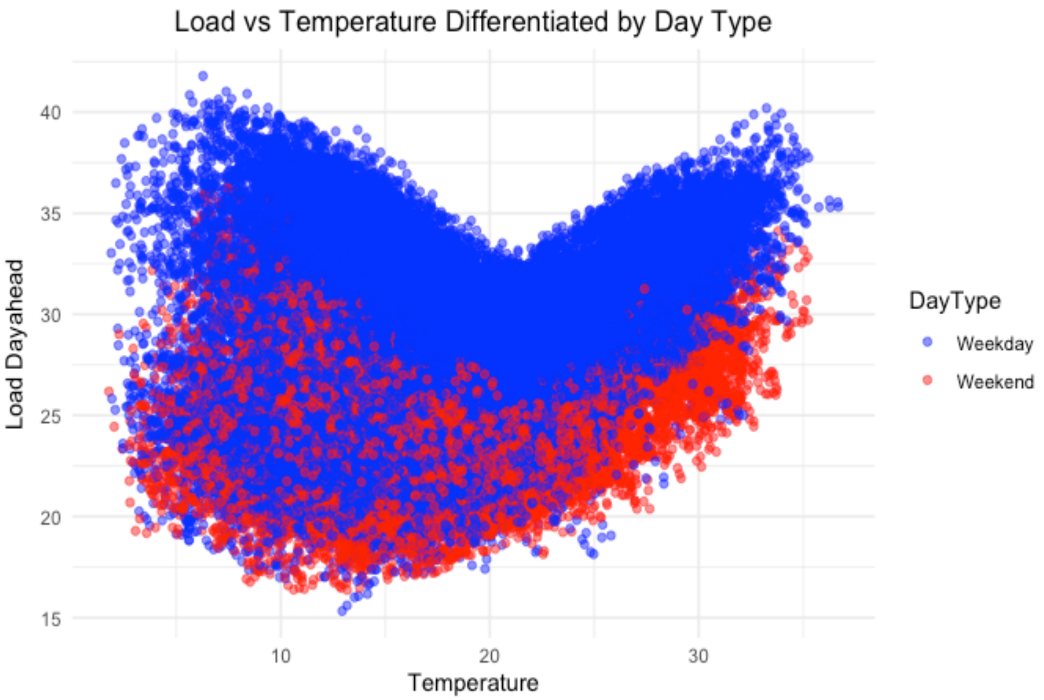
Regional Holidays: If more than 10% of the national population celebrated that holiday, then the day of the holiday is assigned a value of 1 otherwise 0.

Christmas holiday period: Holidays period 24/12 – 06/01 is set to 1, otherwise 0-

- 21/35 holidays are chosen for modelling

2. Data – Calendar and weather effect



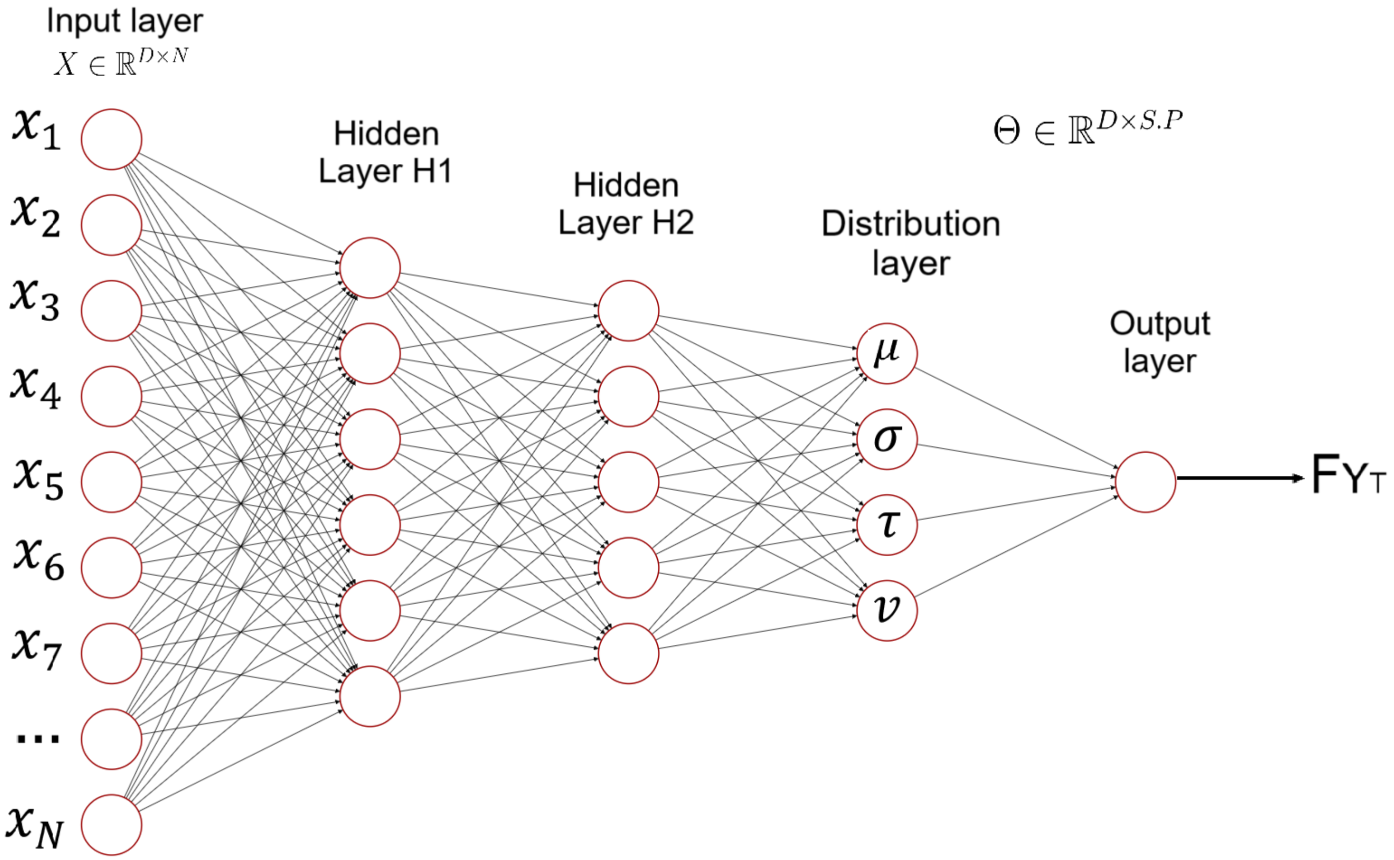


- Calendar effect: Hour of the Day, Hour of the week (or Day of the week), Hour of the year
- Weather effect: Historical temperature for training and forecasted temperature for forecasting

  Lagged temperature intervals—1 hour, 2 hours, 3 hours, 24 hours (1 day), and 48 hours (2 days), 3-hour mean temperature, minimum and maximum temperatures over the past 24 hours, and trend.

  9 weather features are chosen for modelling
- Historical Load: lag values of 1, 2, 6 and 7 days of load day ahead

3. Methods – Distributional Deep Neural Networks (DDNN)



- Input layer, 2 hidden layers, output layer with parameter layer Θ consists of P distribution parameters for each of the S output features with S = 168 representing the 168 hourly load predictions for the next 7 days for Spain.

3. Methods – Distributional Deep Neural Networks (DDNN)

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Optimization problem: Minimizes negative log-likelihood of the observed data given the predicted distribution parameters

$$\hat{\Theta} = \arg\min_{\Theta} \left(-\sum_{i=1}^{n} \log f(L_i; \Theta, x_i) \right)$$

To mitigate the overfitting risk, regularization is applied for the loss function

$$\mathscr{L}_{\text{reg}}(L; F(\Theta; x)) = \mathscr{L}(L; F(\Theta; x)) + \sum_{i=0}^{I-1} \lambda_{1,i} \|H_i\|_k + \sum_{i=0}^{I-1} \lambda_{2,i} \|W_{i+1}\|_k + \sum_{i=0}^{I-1} \lambda_{3,i} \|b_{i+1}\|_k$$

$$+ \sum_{p=0}^{P} (\lambda_{1,I,p} \|H_I\|_k + \lambda_{2,I,p} \|W_{I+1}\|_k + \lambda_{3,I,p} \|b_{I+1}\|_k)$$

3. Methods – Distributional Deep Neural Networks (DDNN)

- Hyperparamter tuning
- Not consider weather data in DDNN.
- Fixed: all features, 2 hidden layers, 100 epochs, batch size 32.
- Number of neurons in each hidden layers: 16-128.

➢ Results: (ND) is (71, 53) and JSU is (75, 55).

- Learning Rate: 10^{-5} to 10^{-1} on a logarithmic scale.

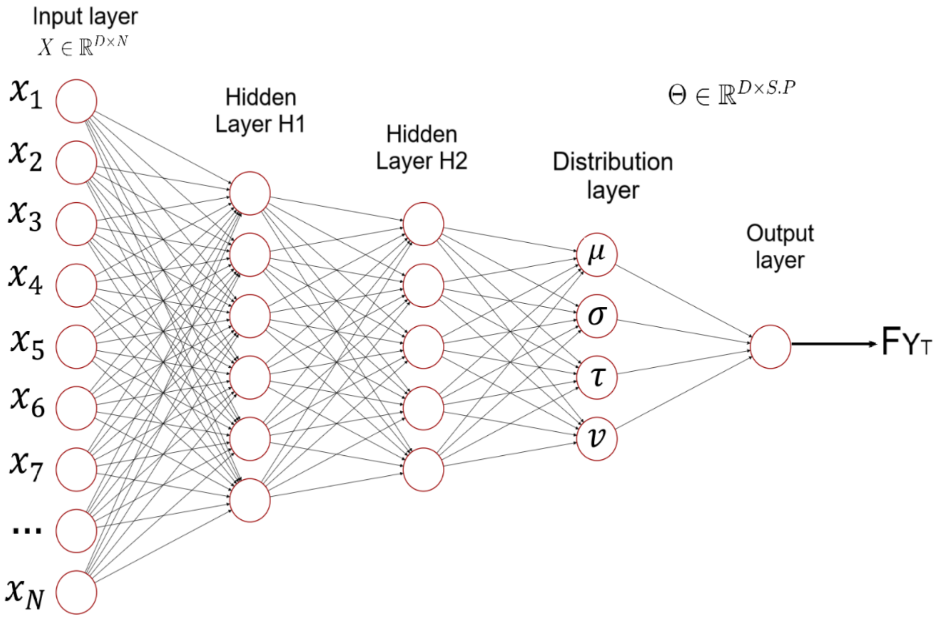
➢ Results ND is 0.091 and for JSU is0.0026.

- Dropout Rate: between 0 and 1.

➢ Only JSU need drop out layer with rate of 0.0055.

- Regularization rate: 10^{-5} to 10^{-1} on a logarithmic scale.
- Activation Functions: tunning using ReLU, ELU, Sigmoid, Tanh, Softplus, and Softmax.

➢ Results: JSU, the ReLU is utilized and for ND the Softplus .



3. Methods– Generalized Additive Models for Location Scale and Shape (GAMLSS)

GAMLSS MODEL: $$g_k(\theta_k) = X_k\beta_k + \sum_{j=1}^{J_k} s_{kj}(x_{kj})$$

where $g_k(\,.\,)$ link function which is determined by the distribution parameter θ_k, X_k is design matrix, β_k is the parameter vector and s_{jk} is smooth functions of an explanatory variable x_{jk}. Unknown parameters are estimated by maximising the penalized log-likelihood given by

$$\ell_p = \sum_{i=1}^{n} \log f(y_i \,|\, \theta_i) - \frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{J_k} \lambda_{jk}\gamma_{jk}^{\top} G_{jk}\gamma_{jk}$$

where G_{jk} is a symmetric matrix that depends on a vector of smoothing parameters λ_{jk} .

Adaptive LASSO

$$\hat{\beta}_{\lambda}^{lasso} = \arg\min_{\beta} \left\| \mathbf{y} - \sum_{j=1}^{p} \mathbf{x}_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j| \, .$$

where w is the weights to different coefficients. By using adaptive weights, it reduces the bias in the coefficient estimates that is common with standard Lasso. Adaptive Lasso method is implemented in gamlss.lasso package in R.

3. Methods – Benchmark and evaluation

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Auto-Regressive with Exogenous Input (ARX)

- Simple time series model as a benchmark model
- Incorporates both past values itself (auto-regressive components L(t-h)) and past values of external (exogenous – weekend dummy (0 for weekdays and 1 for weekends)

$$L_t = \phi_0 + \phi_1 L_{t-24} + \phi_2 L_{t-48} + \phi_3 L_{t-144} + \phi_4 L_{t-168} + \phi_5 D_t + \varepsilon_t,$$

Evaluation: Pinball score and DM test

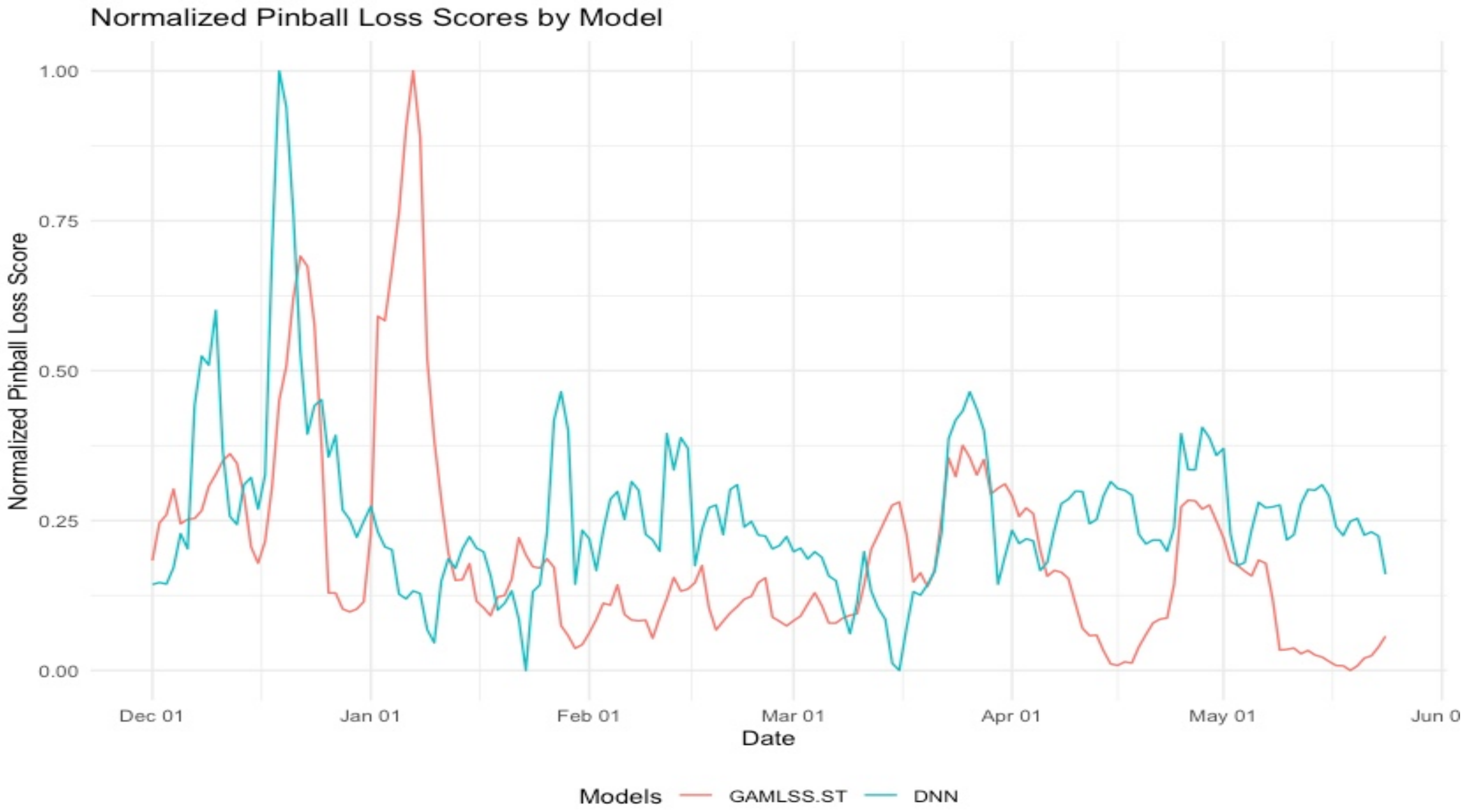
$$\text{Pinball}^{\alpha}_{(L_{d,s},\hat{L}_{d,s})} = \begin{cases} \alpha(L_{d,s} - \hat{L}_{d,s}) & \text{if } L_{d,s} \geq \hat{L}_{d,s} \\ (1-\alpha)(\hat{L}_{d,s} - L_{d,s}) & \text{if } L_{d,s} < \hat{L}_{d,s} \end{cases}$$

- The target quantile α is ranging from 0.05 to 0.95 with the step of 0.05

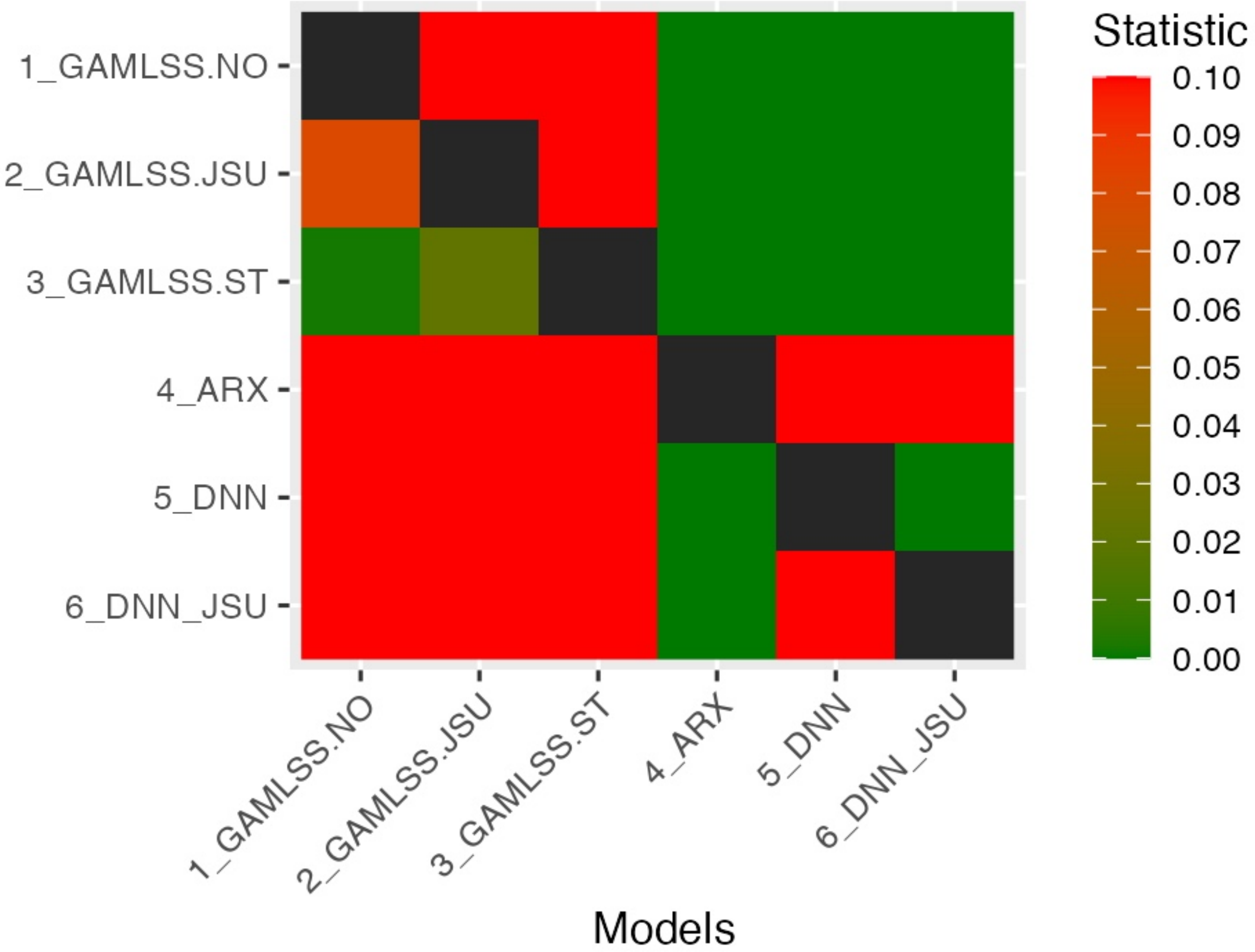
4.- Results – Pinball score

Table 1 Average Pinball Loss Score on Test Set

Model	GAMLSS:NO	GAMLSS.JSU	GAMLSS.ST	ARX	DDNN:NO	DDNN:JSU
PB SCORE	0.5426	0.5392	0.5304	2.18	1.11	1.89



4. Results - **Diebold-Mariano test**



5 - Conclusions

- By incorporating holiday effects, extra calendar data, and temperature variations, all GAMLSS models, especially GAMLSS.ST, demonstrated superior performance with the lowest pinball score
- GAMLSS.NO and GAMLSS.JSU performed well, close to GAMLSS.ST
- DDNN models with only holidays and historical load showed moderate performance and had some advantage over GAMLSS in few forecasted hours
- The ARX model yielded a significantly higher pinball score
- Future research:

➢ Deeper research on holiday and calendar effects

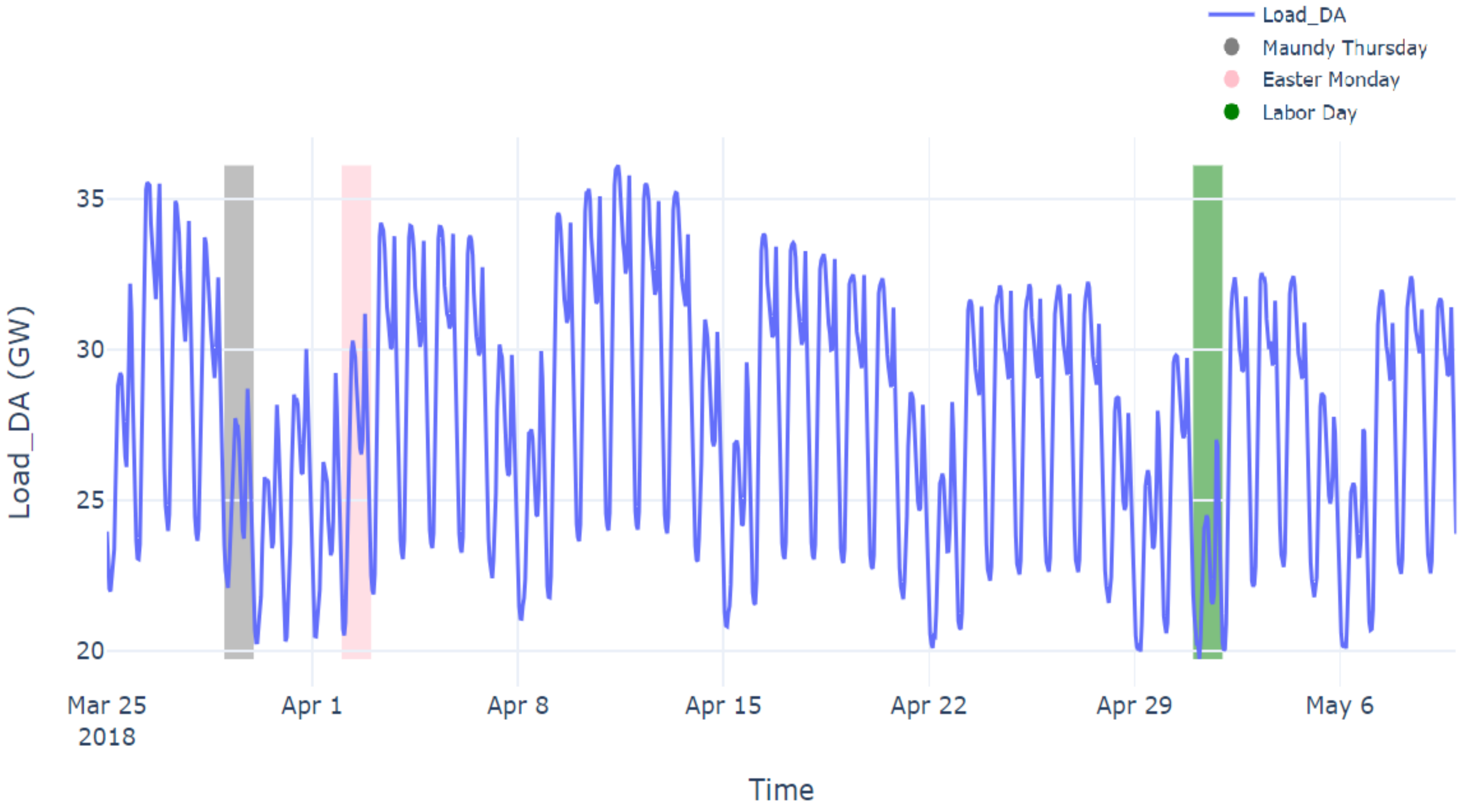
➢ Incorporate weather data into DDNN and optimize the DDNN by exploring a better set of hyperparameters to achieve better results.

Thank You

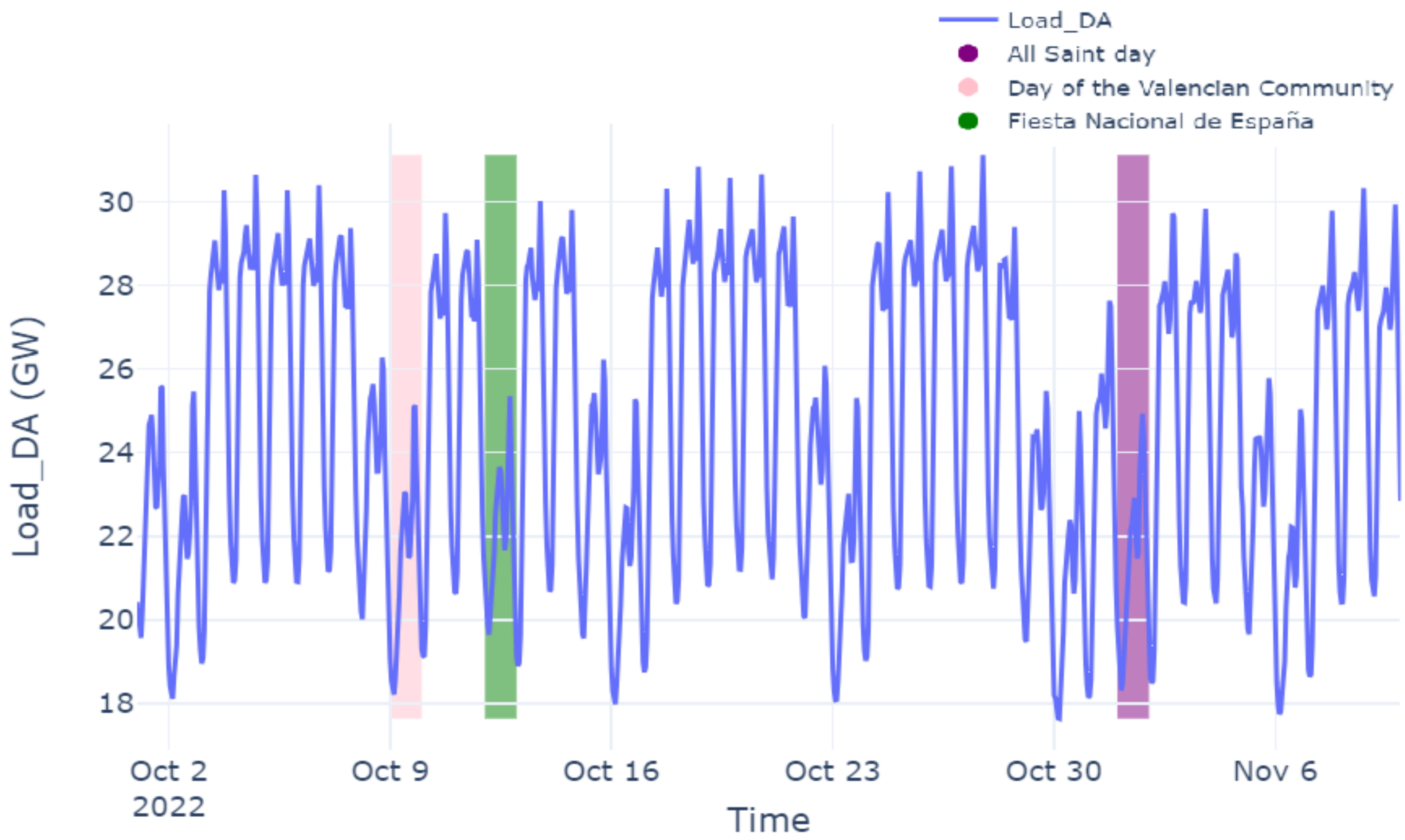
6 - Bibliography

- Ziel, F., & Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. Energy Economics, 70, 396–420.
- Ziel, F., & Liu, B. (2016). Lasso estimation for gefcom2014 probabilistic electric load forecasting. International Journal of Forecasting, 32 (3), 1029–1037.
- Ziel Florian, S. M., Muniain Peru. (2022). Gamlss.lasso: Extra lasso-type additive terms for gamlss. https://CRAN.R-project.org/package=gamlss.lasso
- Monika, Z., & Ziel, F. (2024). Efficient mid-term forecasting of hourly electricity loadusing generalized additive models. arXiv.

APPENDIX



Load_DA
Maundy Thursday
Easter Monday
Labor Day
35
30
25
20
Load_DA (GW)
Mar 25
2018
Apr 1
Apr 8
Apr 15
Apr 22
Apr 29
May 6
Time



Load_DA
All Saint day
Day of the Valencian Community
Fiesta Nacional de España
Load_DA (GW)
30
28
26
24
22
20
18
Oct 2
2022
Oct 9
Oct 16
Oct 23
Oct 30
Nov 6
Time

- GAMLSS Example Model Equation

$$g_k(\theta_k) = \beta_1 Load\_Lag_{t-168} + \beta_5 HolFlx + \beta_6 HolFix + \beta_7 XHol +$$

$$f_{1j}(Temperature) + f_{2j}(HoW) + f_{3j}(HoD) + f_{4j}(HoY)$$

- GAMLSS Penalised Log-Likelihood

$$\ell_p = \sum_{i=1}^{n} \log f(y_i \mid \theta_i) - \frac{1}{2} \sum_{k=1}^{p} \sum_{j=1}^{J_k} \lambda_{jk} \gamma_{jk}^{\top} G_{jk} \gamma_{jk}$$

Adaptive Lasso

Adaptive Lasso is an enhanced version of the standard Lasso that uses adaptive weights for penalization, improving variable selection and coefficient estimation.

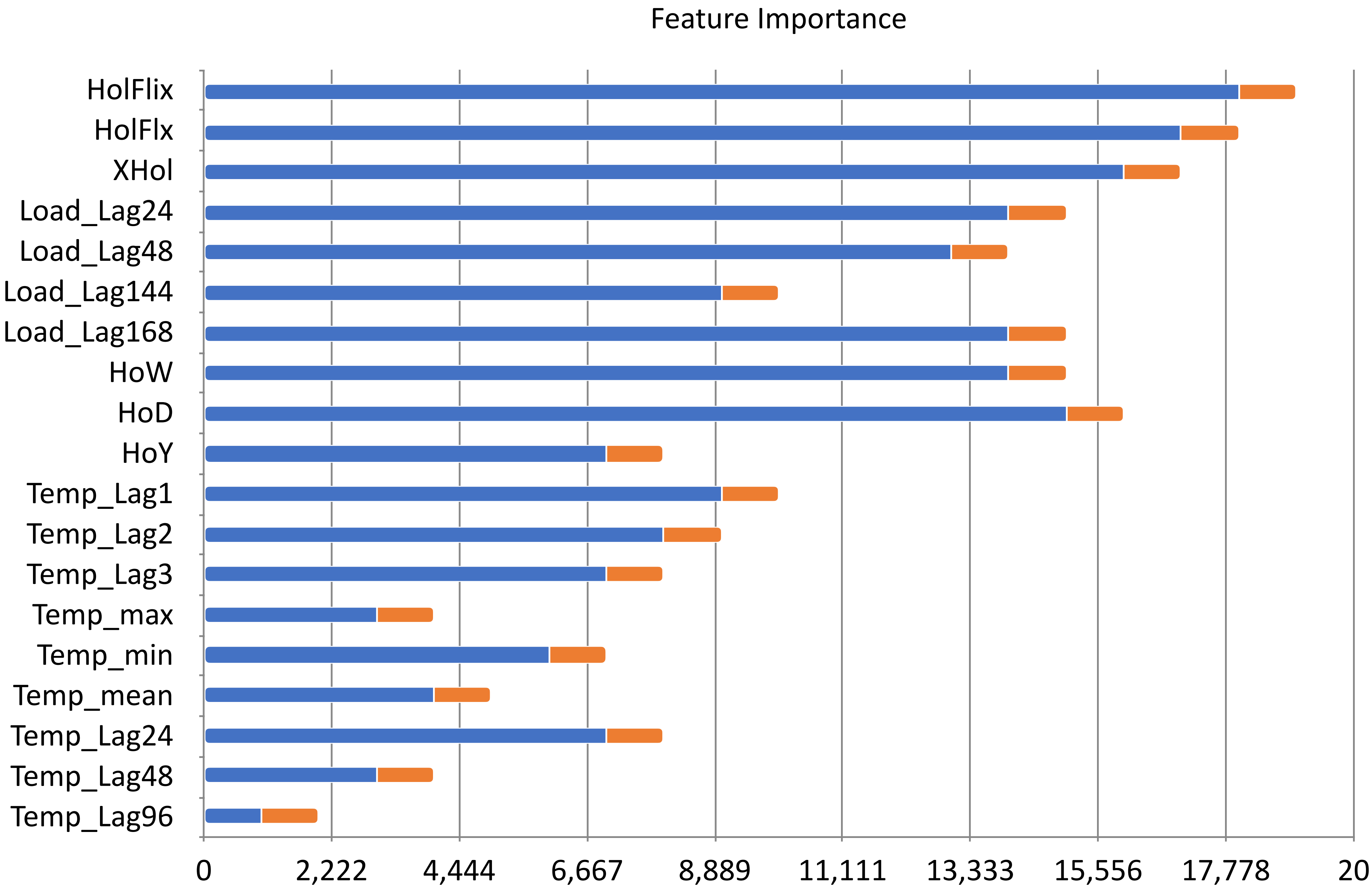
$$\min_{\beta}\left(\sum_{i=1}^{N}(y_i - X_i\beta)^2 + \lambda\sum_{j=1}^{p} w_j|\beta_j|\right)$$

1. Initial Estimation: Obtain initial estimates $\hat{\beta}_j$ using MLE or Ridge regression.

2. Adaptive Weights Calculation: Calculate adaptive weights $w_j = \frac{1}{|\hat{\beta}_j|^{\gamma}}$

3. Penalized Likelihood: Minimize the penalized likelihood function: $PL = \log(L) - \lambda\sum_{j=1}^{p} w_j|\beta_j|$ where log(L) is the log-likelihood of the GAMLSS model.

4. Feature Selection and Shrinkage: Perform feature selection and shrinkage by solving the adaptive Lasso optimization problem.


Feature Importance
HolFlix
HolFlx
XHol
Load_Lag24
Load_Lag48
Load_Lag144
Load_Lag168
HoW
HoD
HoY
Temp_Lag1
Temp_Lag2
Temp_Lag3
Temp_max
Temp_min
Temp_mean
Temp_Lag24
Temp_Lag48
Temp_Lag96
0
2,222
4,444
6,667
8,889
11,111
13,333
15,556
17,778
20

The Diebold-Mariano (DM)

We used the loss function $L_{A,B,j} = PB^{A}_{T,j} - PB^{B}_{T,j}$ of the j's forecasting window for two different forecasts A and B for DM test.

* DM-test checks if the loss $L_{A,B,j}$ is significantly different from zero.

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2 / T}}, \text{ where } d_t = L(e_{1,t}) - L(e_{2,t}) \text{ (Loss differential)}$$

Test statistics converges with increasing N to the standard normal dist.