



News and Load: Quantification of Economic and Social Drivers behind Regional Electricity Demand

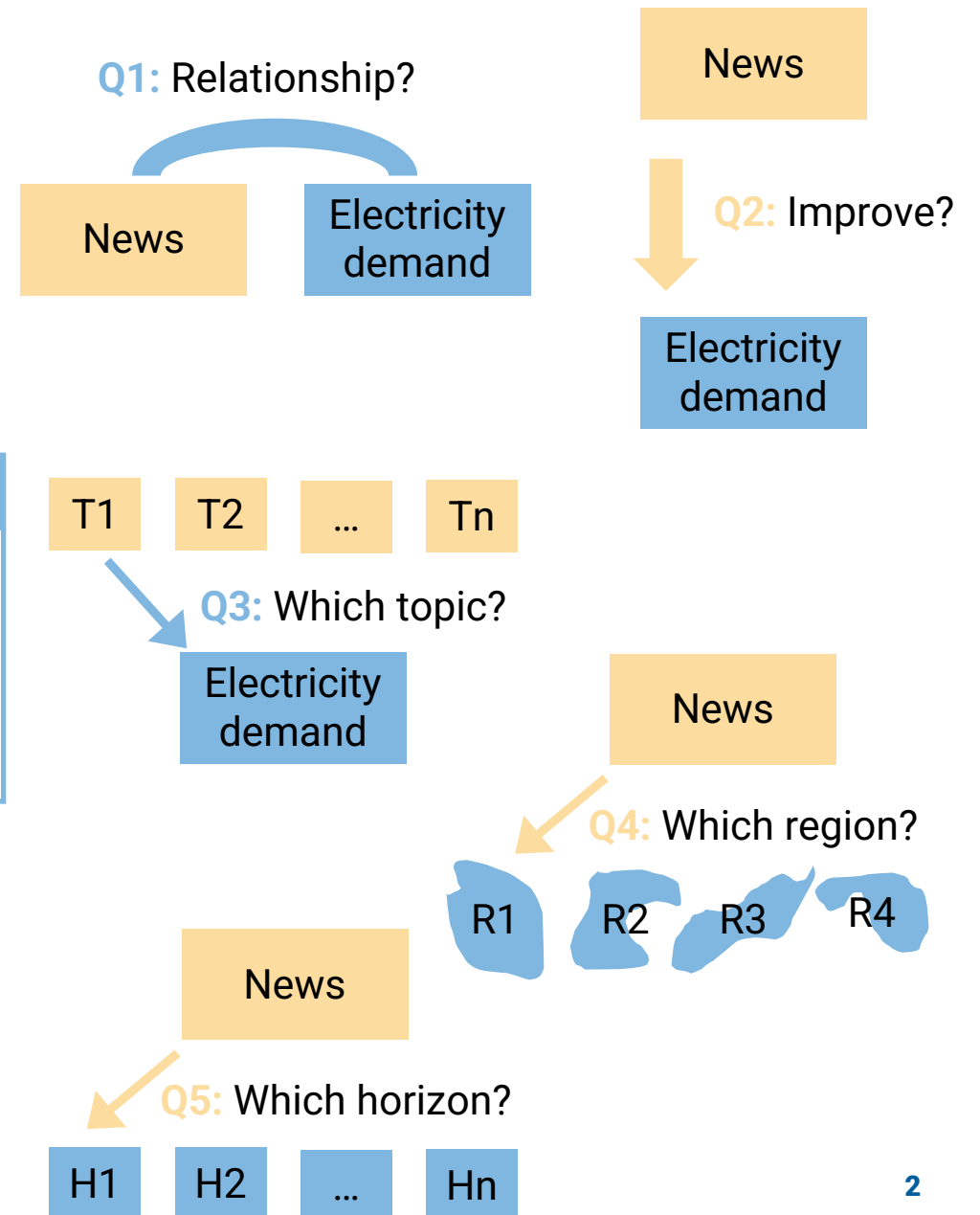
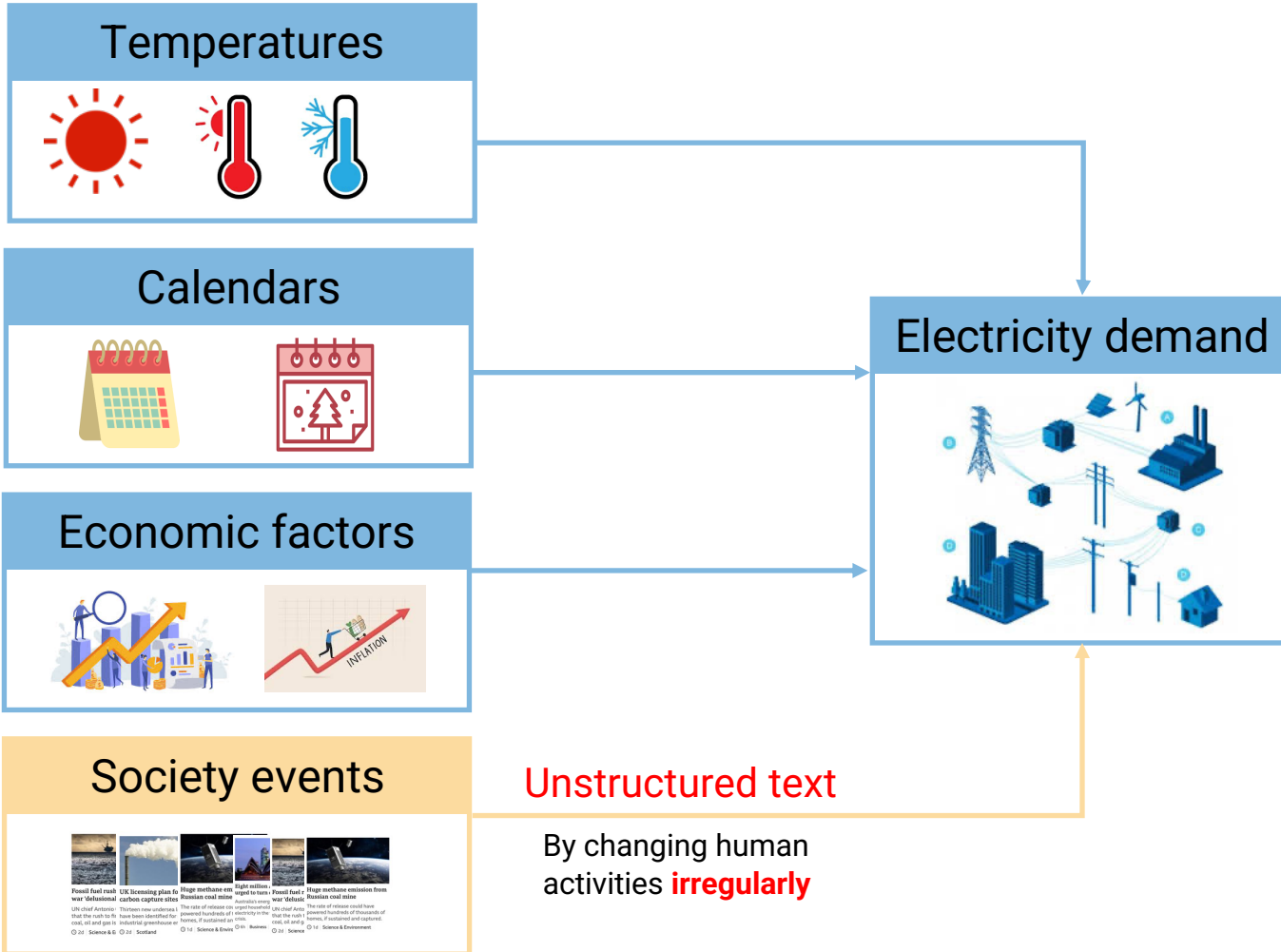
Yun BAI

Supervised by: Andrea MICHIORRI and Simon CAMAL

yun.bai@minesparis.psl.eu

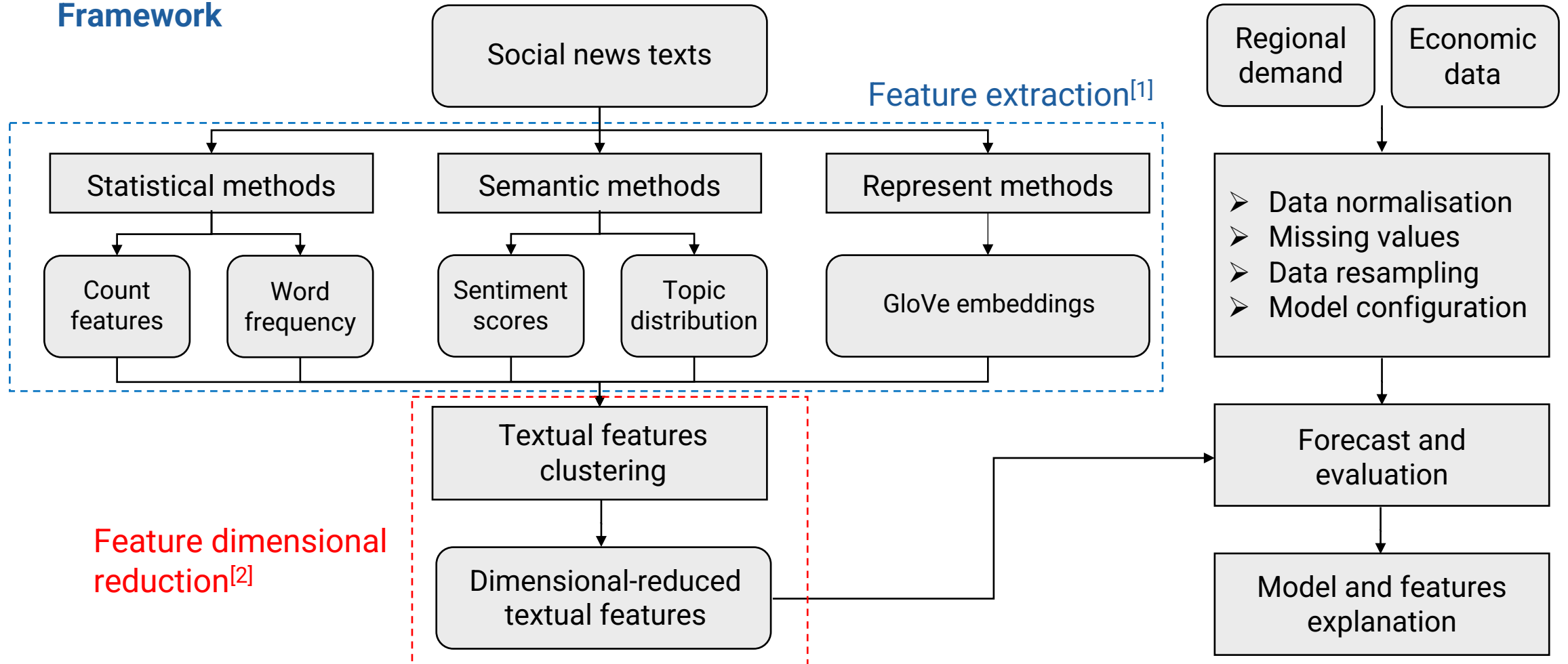
2024.07.01

1 CONTEXT



2 METHODOLOGY

Framework



[1] Bai, Yun, Simon Camal, and Andrea Michiorri. "News and load: A quantitative exploration of natural language processing applications for forecasting day-ahead electricity system demand." *IEEE Transactions on Power Systems* (2024). Doi: [10.1109/TPWRS.2024.3361074](https://doi.org/10.1109/TPWRS.2024.3361074)

[2] Bai, Yun, Simon Camal, and Andrea Michiorri. "Investigation of the Impact of Economic and Social Factors on Energy Demand through Natural Language Processing." *arXiv preprint arXiv:2406.06641* (2024).

2 METHODOLOGY

Textual features

- ❖ **Statistical features:**
count features, word frequency
- ❖ **Semantic features:**
Sentiment scores, topic distributions
- ❖ **Represent features:**
GloVe embeddings

Count features


- Count **numbers** of words, sentences, pieces of news, unique words, non-stopping words, electric-related words.
- Count average **length** of sentences and news (with unit words)
- Section **type** (Asia, Europe, Scotland,...)

Word frequency


- Count the number of documents that a **certain word** appear.
- Select words appear in more than 500 titles, 1000 descriptions and 10000 text bodies, ensuring final number of words is balanced.

Sentiment scores

Polarity

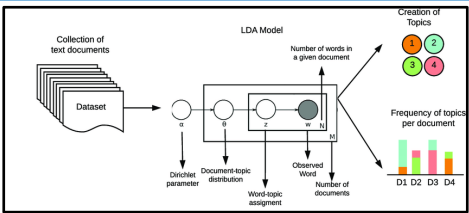


Subjectivity



Compute the polarity and subjectivity scores of a document with TextBlob dictionary.

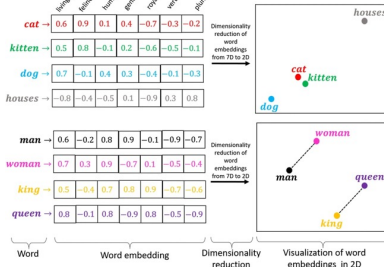
Topic distributions



Obtain the latent topic probabilistic from unlabelled news big data

GloVe embeddings

Word	house	man	woman	king	queen		
cat	0.6	0.9	0.1	0.4	-0.7	-0.3	-0.2
kitten	0.5	0.8	-0.1	0.2	-0.6	-0.5	-0.1
dog	0.7	-0.1	0.4	0.3	-0.4	-0.1	-0.3
houses	-0.8	-0.4	-0.5	0.1	-0.9	0.3	0.8
man	0.6	-0.2	0.8	0.9	-0.1	-0.9	-0.7
woman	0.7	0.3	0.9	-0.7	0.1	-0.5	-0.4
king	0.5	-0.4	0.7	0.8	0.9	-0.7	-0.6
queen	0.8	-0.1	0.8	-0.9	0.8	-0.5	-0.9



Represent each word with a 100d GloVe embedding pre-trained model.

2 METHODOLOGY

Social textual features

	Textual features type	Num of features
News titles	Count features	27
	Word frequency	456
	Sentiment scores	18
	Topic distributions	87
	GloVe embeddings	100
News descriptions	Count features	27
	Word frequency	329
	Sentiment scores	18
	Topic distributions	100
	GloVe embeddings	100
News bodies	Count features	27
	Word frequency	550
	Sentiment scores	18
	Topic distributions	69
	GloVe embeddings	100

Date

February-March 2021

February 8-21, 2021

February-March 2021

February 5-March 19, 2022

February 4-20, 2022

June 27-July 10, 2022

July 6-31, 2022

July 28-August 8, 2022

November 20-December 18, 2022

Sport Events Name

Six Nations Championship

Australian Open

India vs. England Test Series

Six Nations Championship

Winter Olympics

Wimbledon Championships

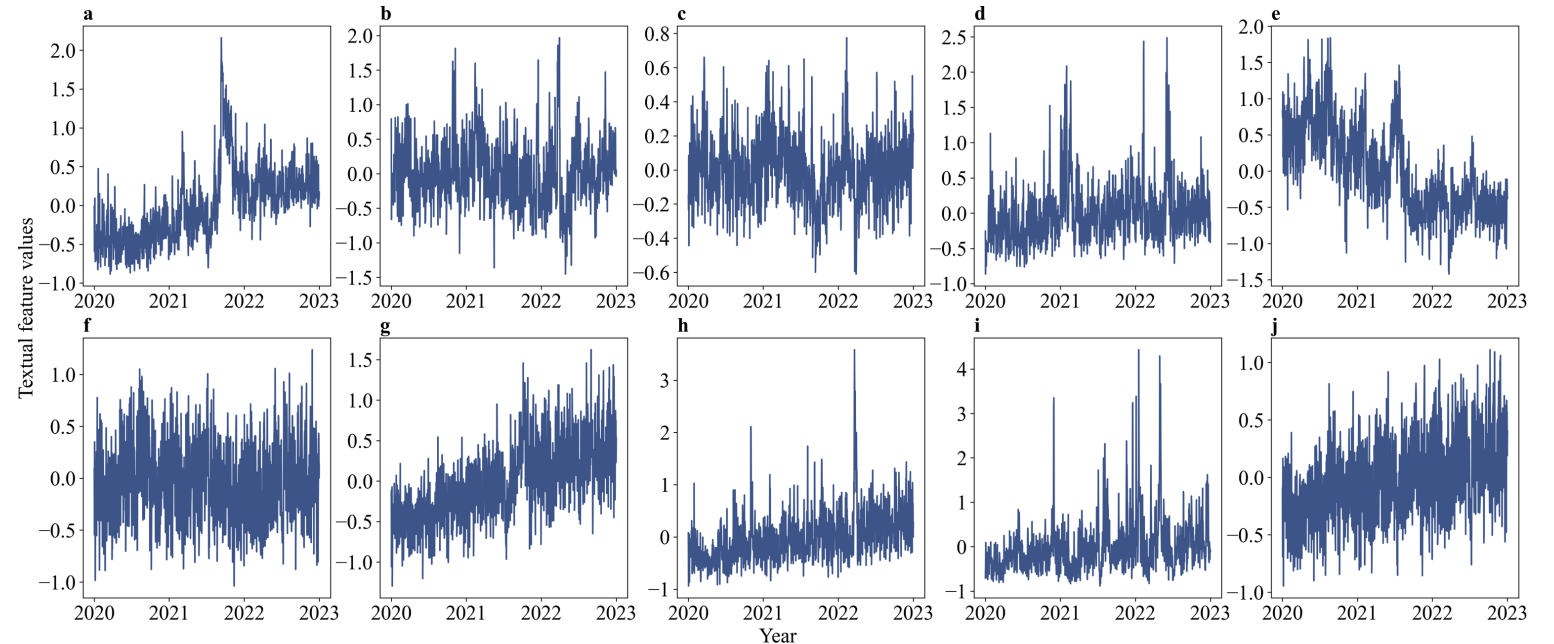
UEFA Women's Euro

Commonwealth Games

FIFA World Cup



military conflicts transportation travel and leisure sports events* pandemic control



regional economics

strikes

family life

election

energy markets

2 METHODOLOGY

Models

❖ Naïve models

- Persistence
- Smart Climatology
- Linear combination of persistence and smart climatology

❖ Linear models

- Least Absolute Shrinkage and Selection Operator (LASSO) regression
- Generalised Additive Models (GAM)

❖ Machine learning models

- Support Vector Regression (SVR)
- ExtraTrees regressor
- Light Gradient Boosting Machine (LightGBM)
- Probabilistic Gradient Boosting Machines (PGBM)

Error metrics

$$\text{RMSE} = \frac{1}{30D} \sum_{H=1}^{30} \sum_{d=1}^D \sqrt{\frac{1}{24} \sum_{h=1}^{24} (y_{h,d,H} - \hat{y}_{h,d,H})^2}$$

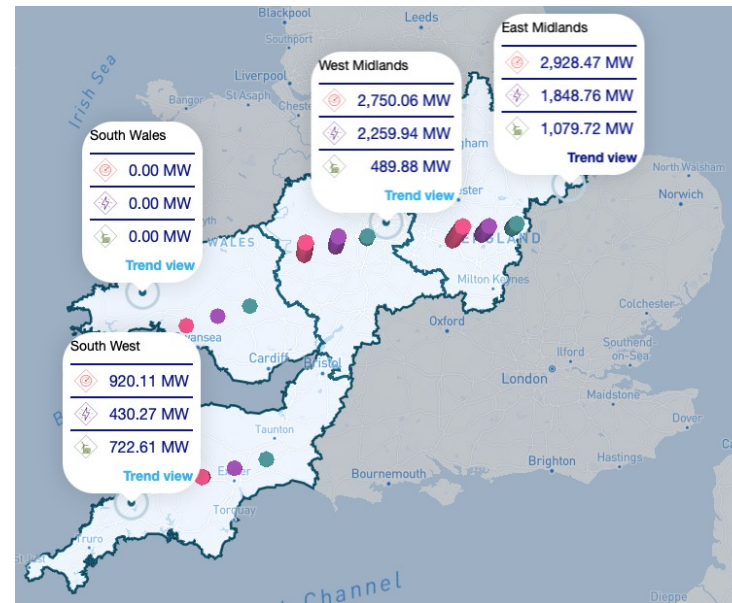
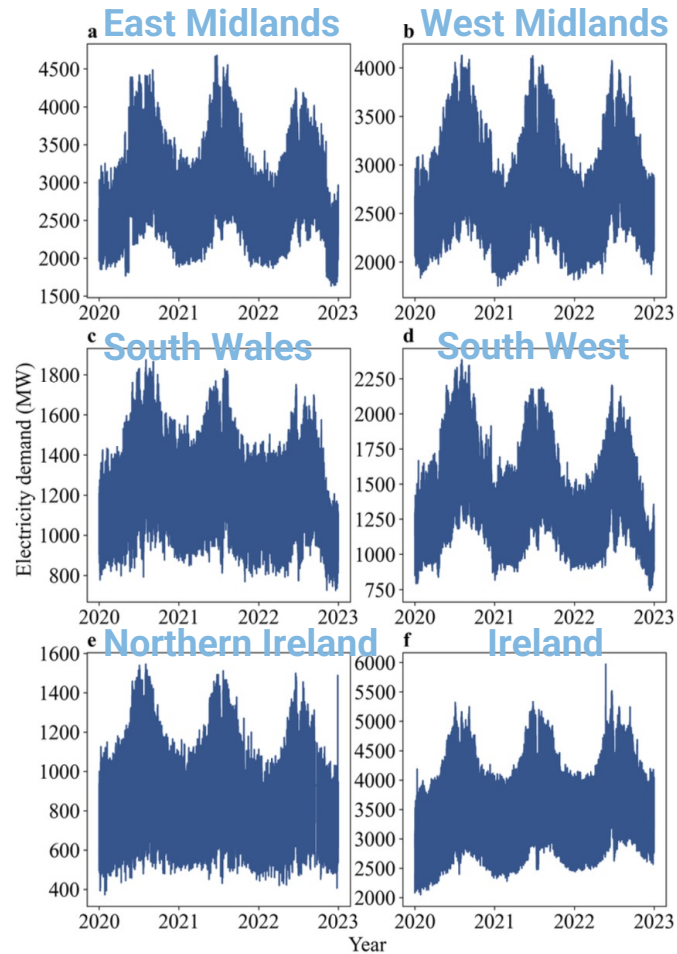
$$\text{MAPE} = \frac{100\%}{30 \times D \times 24} \sum_{H=1}^{30} \sum_{d=1}^D \sum_{h=1}^{24} \left| \frac{y_{h,d,H} - \hat{y}_{h,d,H}}{y_{h,d,H}} \right|$$

$$\text{CRPS} = \frac{100\%}{30 \times D \times 24} \sum_{H=1}^{30} \sum_{d=1}^D \sum_{h=1}^{24} \int [F(\hat{y}_{h,d,H}) - \mathbf{1}_{\{\hat{y}_{h,d,H} \geq y_{h,d,H}\}}]^2 d\hat{y}$$

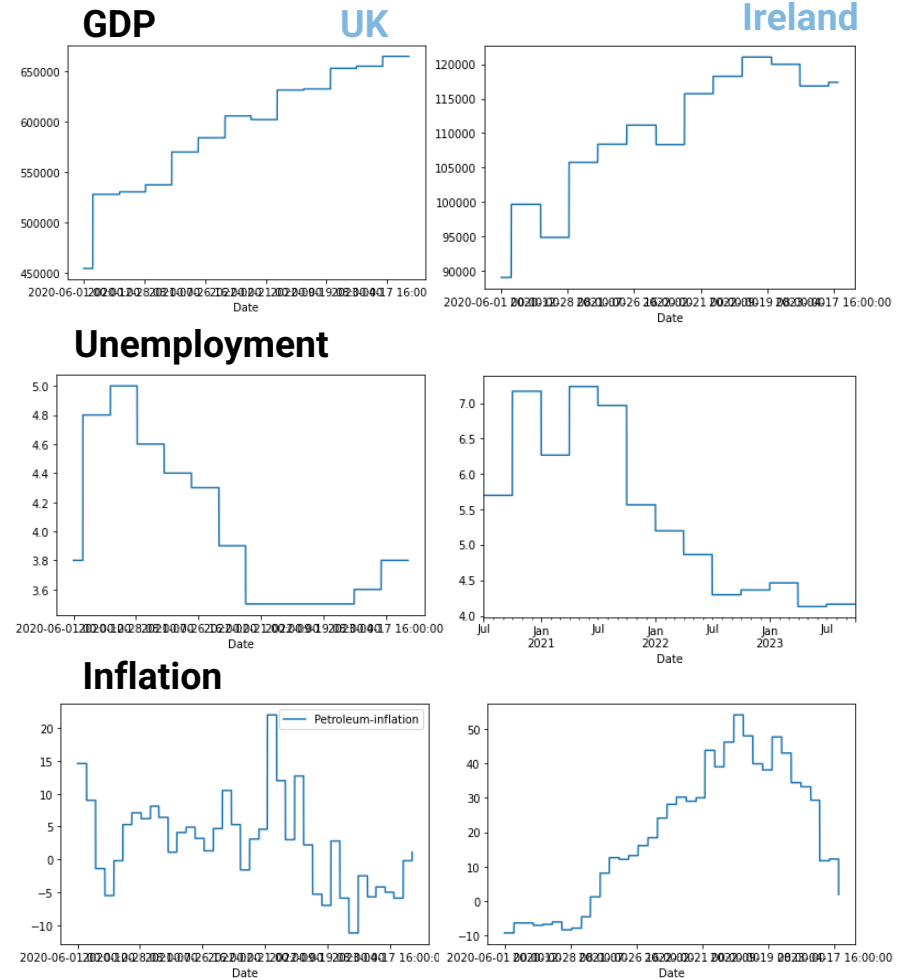
3 EXPERIMENTAL RESULTS

Datasets

Regional electricity demand data

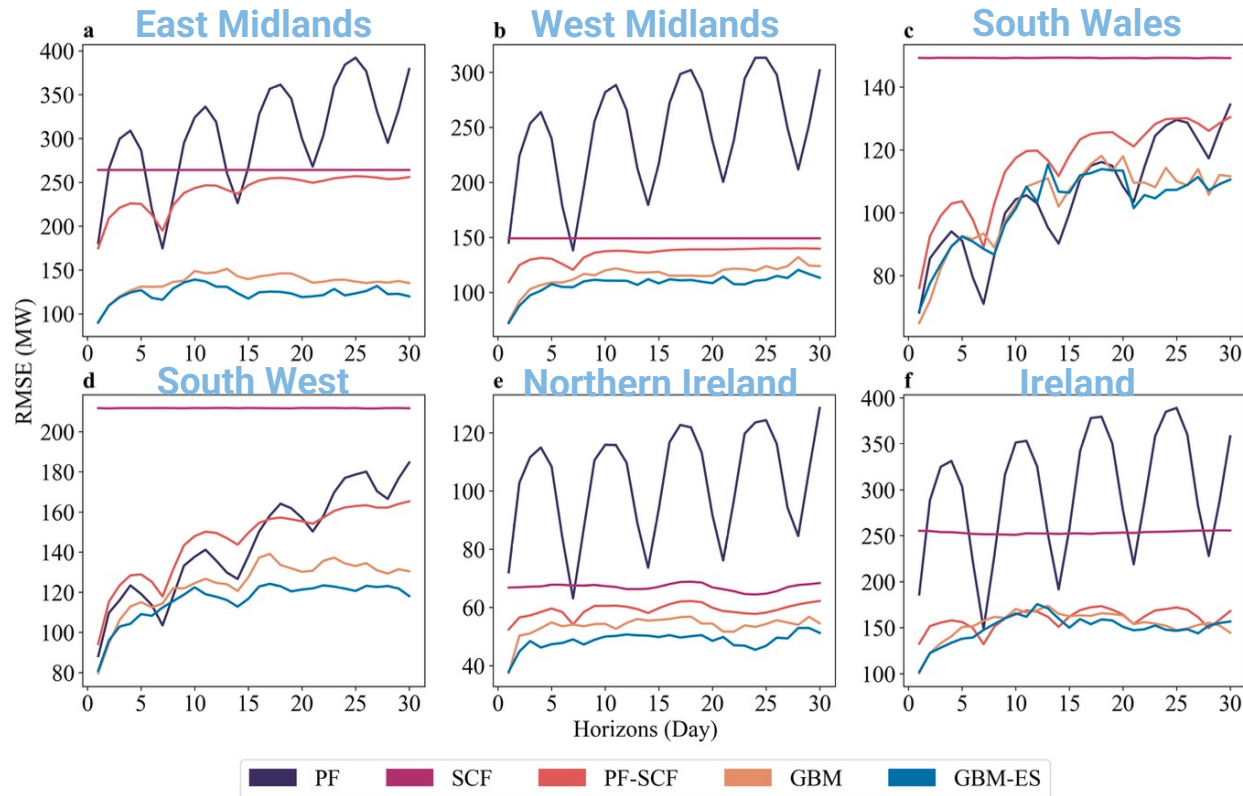


National economic data



3 EXPERIMENTAL RESULTS

Deterministic forecasting results

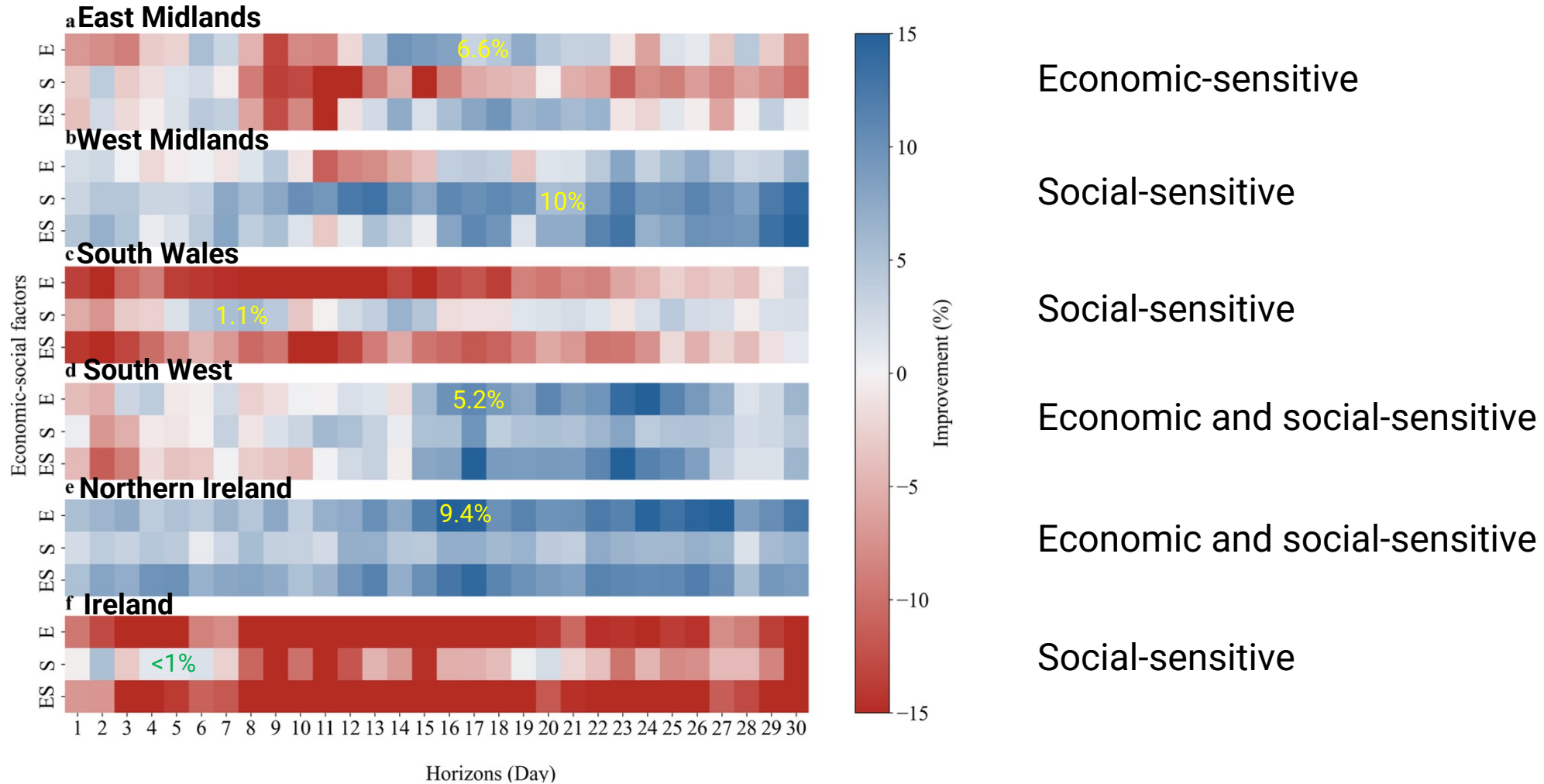


PF: persistence
 SCF: smart climatology
 PF-SCF: linear combination of persistence and smart climatology
 GBM: LightGBM
 GBM-ES: LightGBM model with economic or social features

- ❖ Using the last observed value as forecast will show weekly patterns
- ❖ Combination of persistence and smart climatology is a strong benchmark
- ❖ Improvements can be observed by adding social or economic features into LightGBM models.

3 EXPERIMENTAL RESULTS

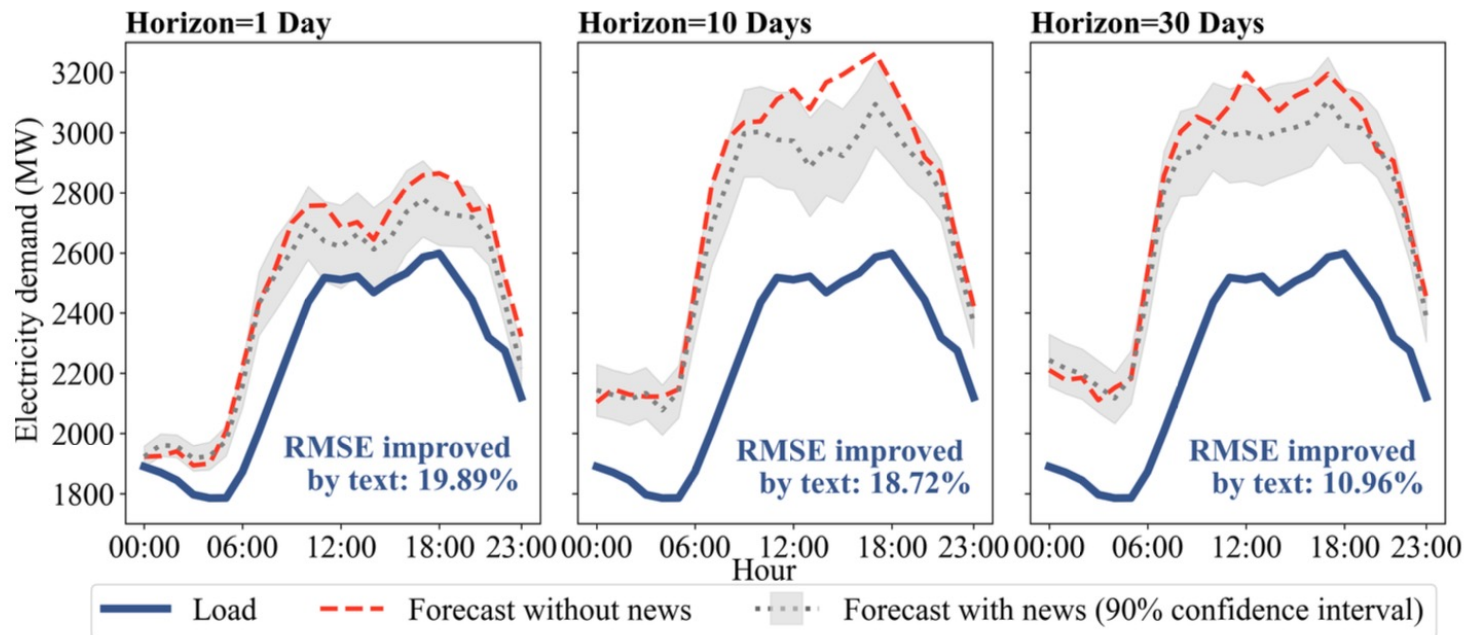
Probabilistic forecasting results



3 EXPERIMENTAL RESULTS

An example for a special day

Truth and forecasting demand on 2023-05-08 (Monday)



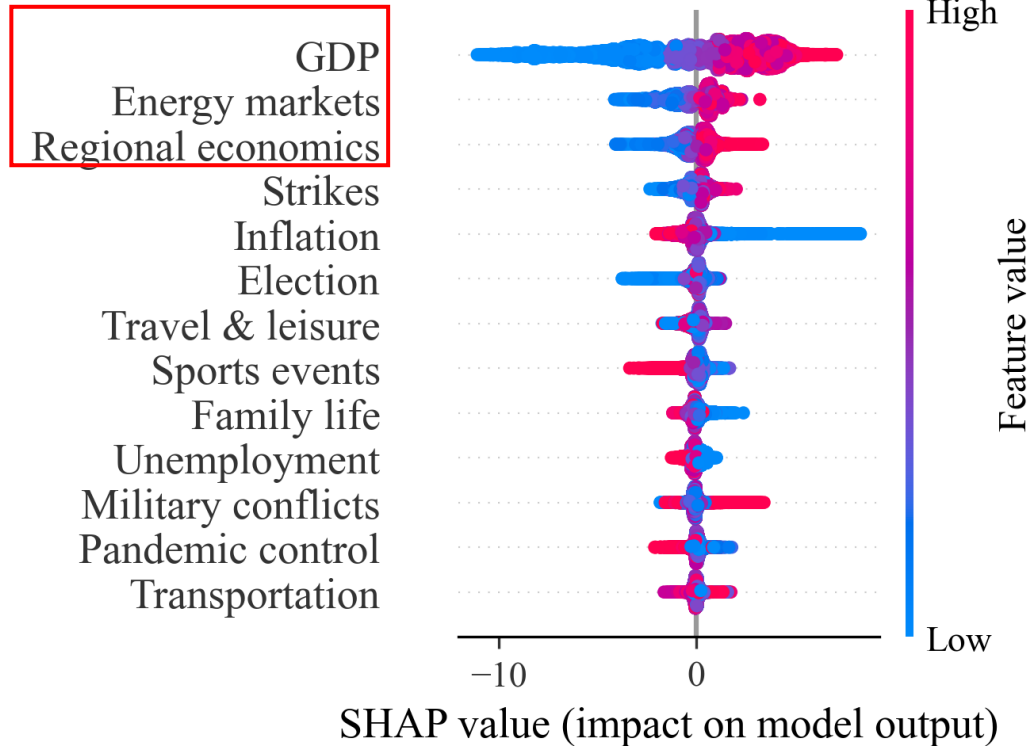
Bank holiday of the coronation of King Charles III



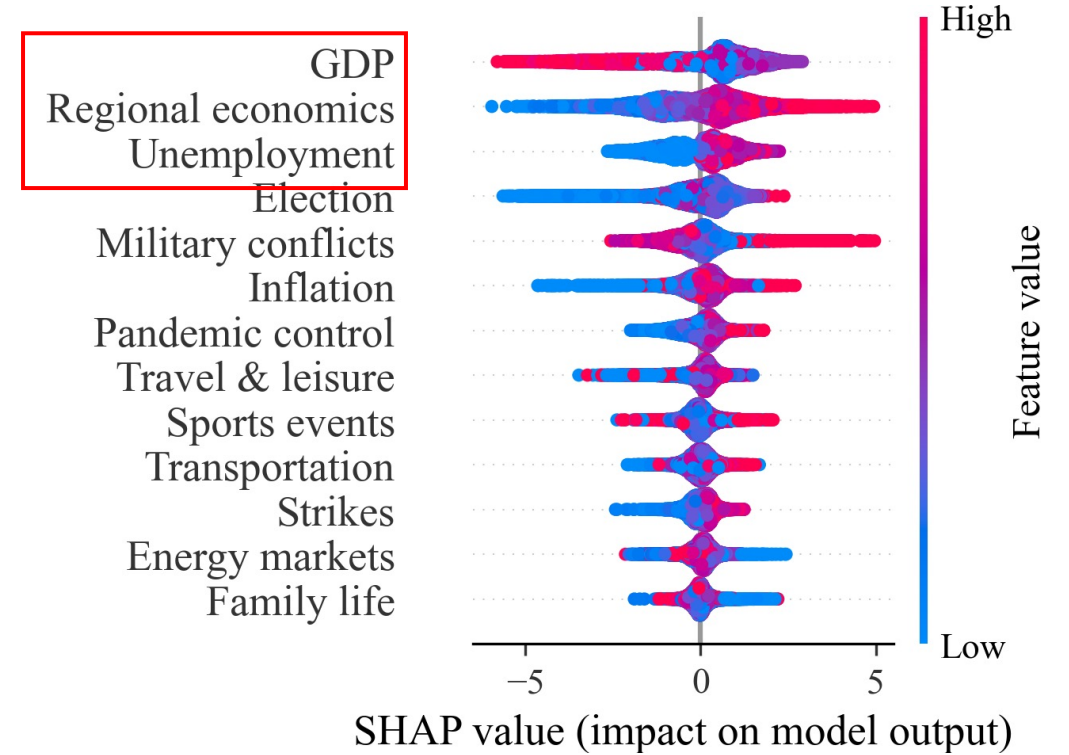
- The holiday was announced as a bank holiday on 2022-11-06
- Forecasts without news overestimate the truth demand
- This information can be captured from the news, and can pull the forecasts to truth.

5 FEATURE IMPORTANCE

East Midlands



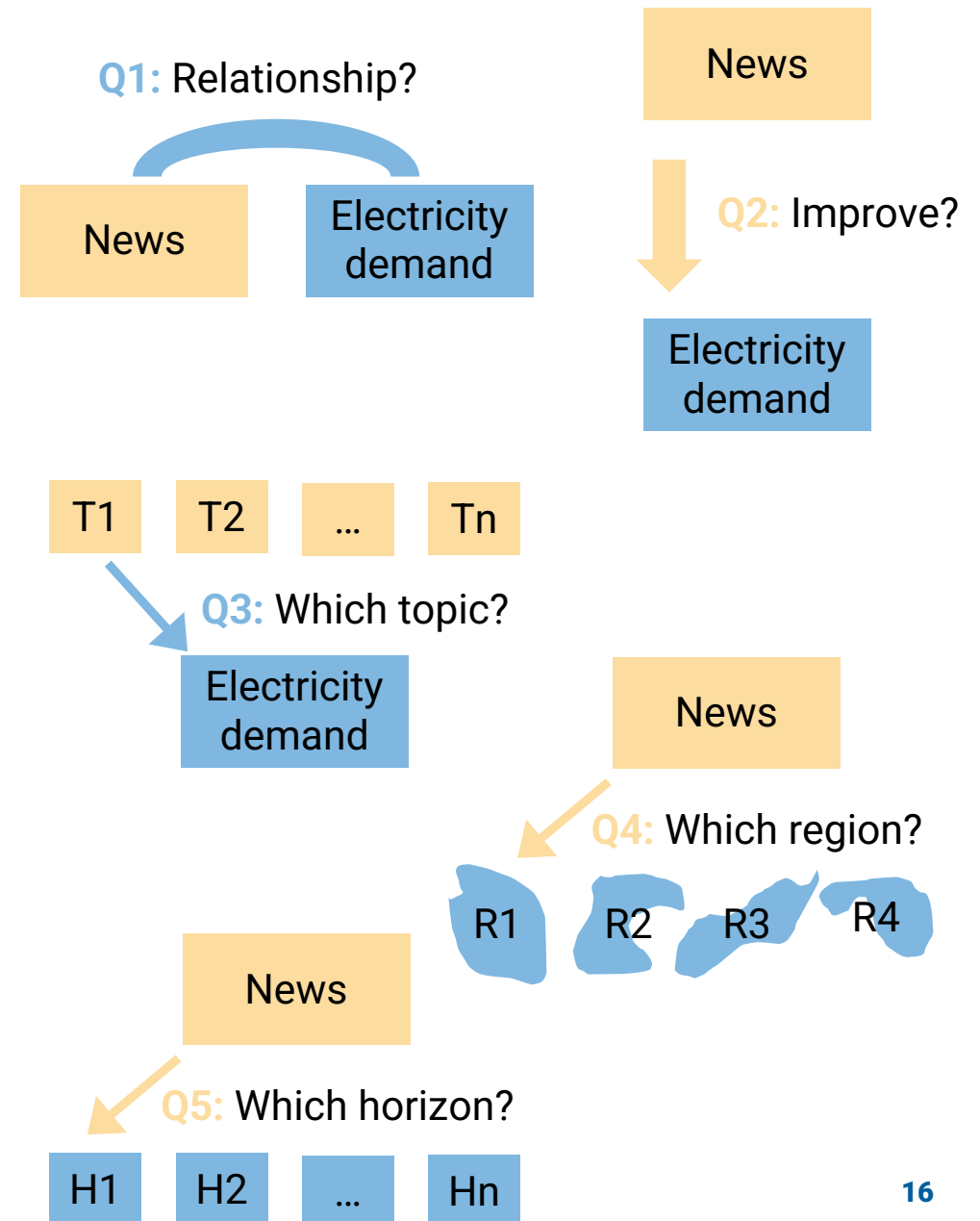
Northern Ireland



- GDP is one of the most important features in economic factors.
- Energy markets and regional economics are two significant social features.

6 CONCLUSIONS

- ✓ Social factors can result in a Granger-cause change in regional demand.
- ✓ Improvements of 3.9% and 2.3% overall for deterministic and probabilistic tasks.
- ✓ From SHAP values, energy markets, regional economics, military conflicts, strikes, and transportation
- ✓ Social: East Midlands and Northern Ireland, Economic: West Midlands and South West
- ✓ After second week.
- ❑ Future work: testing multi-language texts for demand forecasting.



7 REFERENCES

- [1] **Bai, Y.***, Michiorri, A, Camal, S. (2024). **Investigation of the Impact of Economic and Social Factors on Energy Demand through Natural Language Processing** [J]. Submitting.
- [2] **Bai, Y.***, Camal, S., Michiorri, A. (2023). A Quantitative Exploration of Natural Language Processing Applications for Electricity Demand Analysis [J]. IEEE Transactions on Power Systems. Full citations. DOI: [10.1109/TPWRS.2024.3361074](https://doi.org/10.1109/TPWRS.2024.3361074)
- [3] **Bai, Y.***, Camal, S., Michiorri, A. (2023). Electricity Demand Forecasting through Natural Language Processing with Long Short-Term Memory Networks [C]. IEEE PES ISGT Europe 2023. DOI: [10.1109/ISGTEUROPE56780.2023.10408499](https://doi.org/10.1109/ISGTEUROPE56780.2023.10408499)
- [4] **Bai, Y.***, Camal, S., Michiorri, A. (2022). Integrating text analysis in electricity load forecasting: initial findings from UK [C]. 42nd International Symposium on Forecasting, Oxford, UK. (*Oral presentation*).
- [5] **Bai, Y.¹**, Li, X.^{1*}, Yu, H., and Jia, S. (2022) Crude oil price forecasting incorporating news text [J]. International Journal of Forecasting. DOI: [10.1016/j.ijforecast.2021.06.006](https://doi.org/10.1016/j.ijforecast.2021.06.006)

Follow me:





PSL



PERSEE

Thanks for listening

**News and Load: Quantification of
Economic and Social Drivers behind
Regional Electricity Demand**

Yun BAI

Supervised by: Andrea MICHIORRI and Simon CAMAL

yun.bai@minesparis.psl.eu