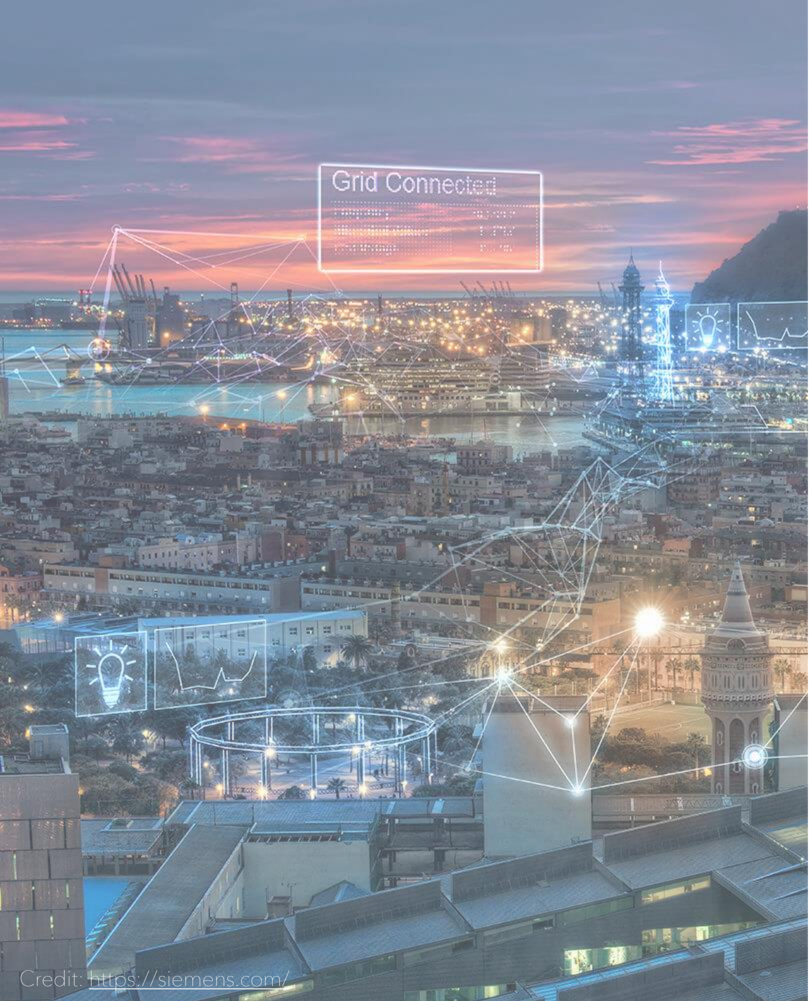
Forecasting hierarchies with coherency-learning

Julien Leprince, Jan Kloppenborg Møller, Waqas Khan, Henrik Madsen, Wim Zeiler

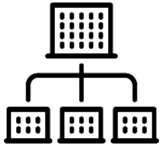
Sep 7th 2023

Building Services



Credit: https://siemens.com/

Building to grid energy flexibility



Energy **balancing**

Energy **flexibility identification**

BEEBOP

Need for information coherency between aggregation layers

Forecasting building loads

Situation

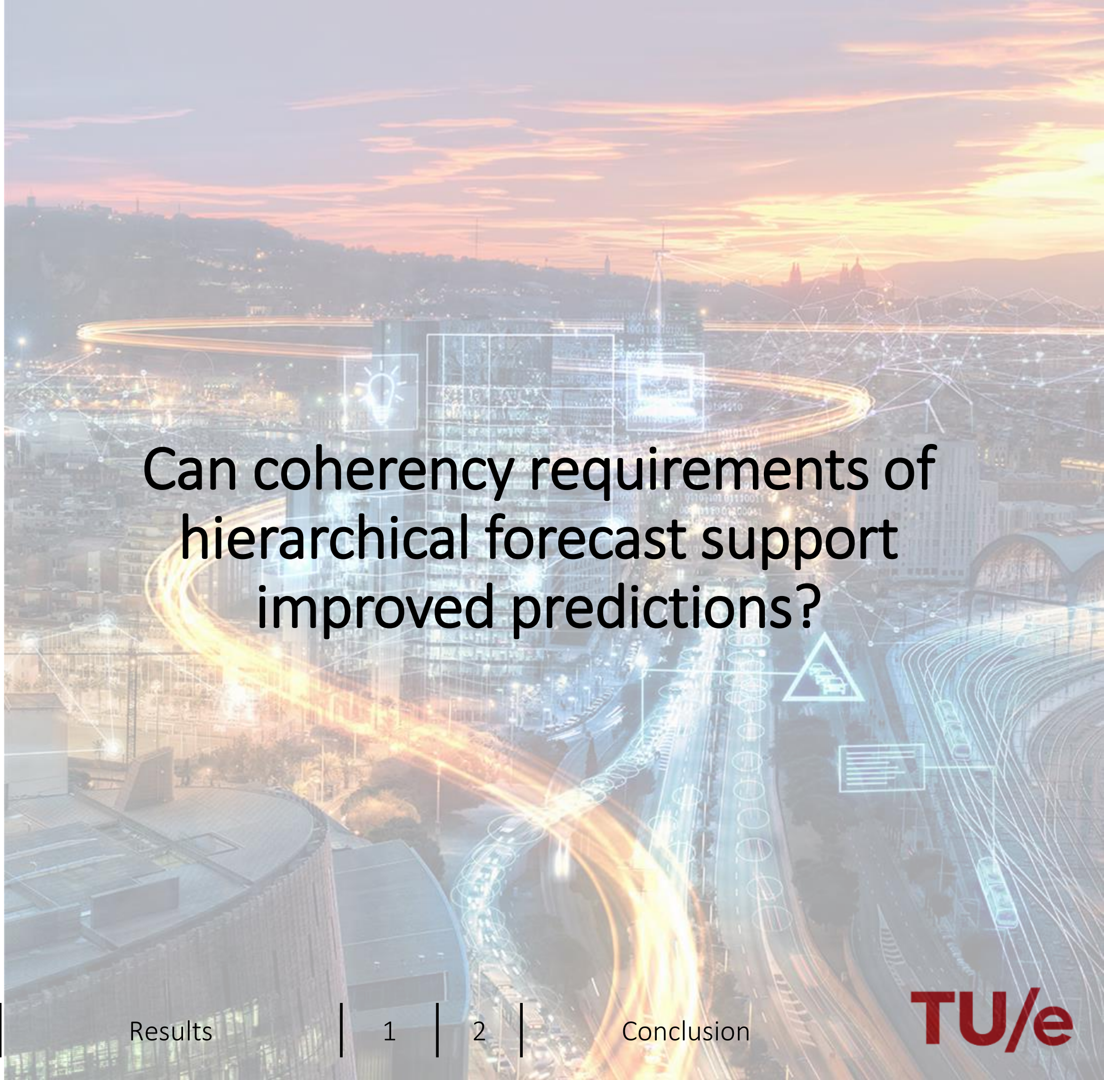
Multiple **disconnected** system **scales**
Buildings – Districts – Energy grid

Problem

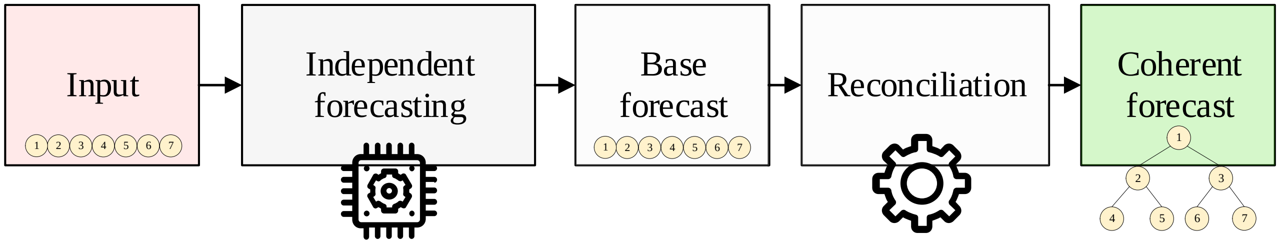
Incoherent forecasts due to independent data and predictions

Idea

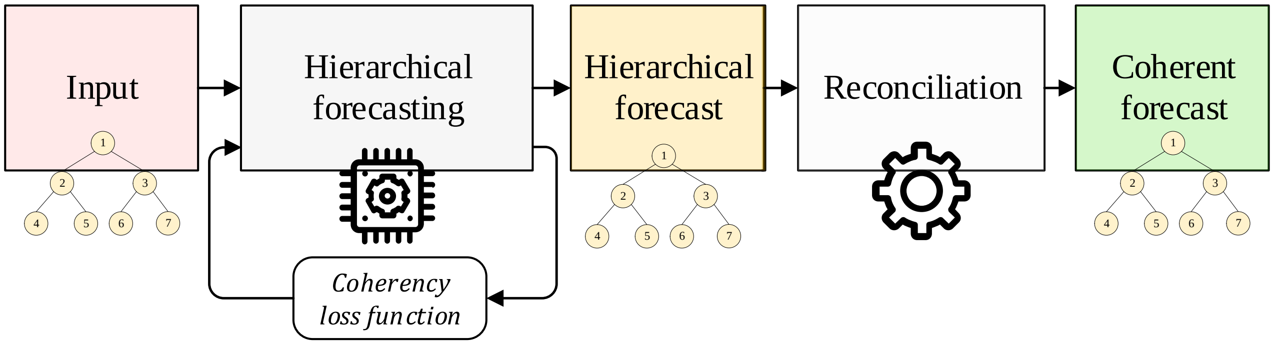
Exploit methods from hierarchical forecasting to improve prediction performances



Approach



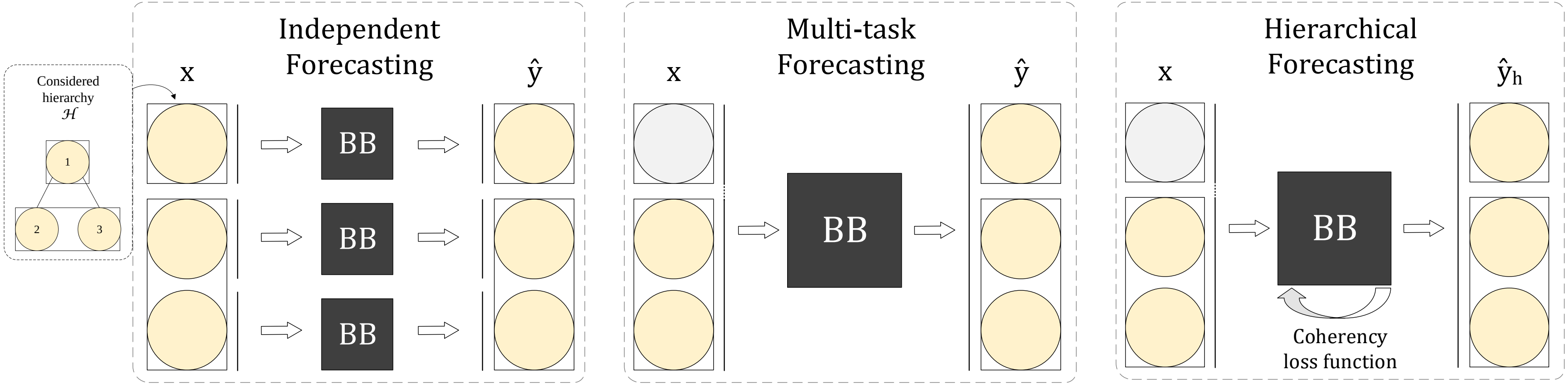
$$\tilde{y} = S\,(S^T\Sigma^{-1}S)^{-1}S^T\Sigma^{-1}\hat{y}$$



$$\hat{y}_h - S\,(S^T\Sigma^{-1}S)^{-1}S^T\Sigma^{-1}\hat{y}_h$$

Towards hierarchical learning

$$\mathcal{L}^c(\mathcal{Y},\hat{\mathcal{Y}}|\Theta) = \frac{1}{h}\sum_{t=1}^{h}\left(\hat{\boldsymbol{y}}_t - S\left(S^T\Sigma^{-1}S\right)^{-1}S^T\Sigma^{-1}\hat{\boldsymbol{y}}_t\right)^2$$



$$\mathcal{L}^b(\mathcal{Y},\hat{\mathcal{Y}}|\Theta) = \frac{1}{h}\sum_{t=1}^{h}\left(y_t - \hat{y}_t\right)^2$$

$$\mathcal{L}^h(\mathcal{Y},\hat{\mathcal{Y}}|\Theta) = \frac{1}{h}\sum_{t=1}^{h}\left(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t\right)^2$$

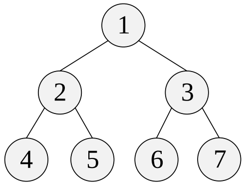
$$\mathcal{L}^c(\mathcal{Y},\hat{\mathcal{Y}}|\Theta) = \frac{1}{h}\sum_{t=1}^{h}\left(\boldsymbol{y}_t - \tilde{\boldsymbol{y}}_t\right)^2$$

$$\mathcal{L}^{hc}(\mathcal{Y},\hat{\mathcal{Y}}|\Theta) = \alpha\mathcal{L}_t^h + (1-\alpha)\mathcal{L}_t^c$$

Covariance matrices

Σ

Considered hierarchy $\mathcal{H}$



Summation matrix and $\boldsymbol{y}$ vector

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{S} \quad \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix}}_{\boldsymbol{y}}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

identity - *id*

$$\sigma_K^2 \begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

structural - *str*

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_7^2 \end{bmatrix}$$

variance - hvar

$$\begin{bmatrix} \sigma_{k1}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{k2}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{k2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{k3}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{k3}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{k3}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{k3}^2 \end{bmatrix}$$

variance - svar

$$\Lambda_{hvar}^{1/2} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \rho_{2,3} & 0 & 0 & 0 & 0 \\ 0 & \rho_{2,3} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \rho_{4,5} & \rho_{4,6} & \rho_{4,7} \\ 0 & 0 & 0 & \rho_{4,5} & 1 & \rho_{5,6} & \rho_{5,7} \\ 0 & 0 & 0 & \rho_{4,6} & \rho_{5,6} & 1 & \rho_{6,7} \\ 0 & 0 & 0 & \rho_{4,7} & \rho_{5,7} & \rho_{6,7} & 1 \end{bmatrix} \Lambda_{hvar}^{1/2}$$

covariance - *kcov*

$$\Lambda_{hvar}^{1/2} \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \rho_{1,4} & \rho_{1,5} & \rho_{1,6} & \rho_{1,7} \\ \rho_{1,2} & 1 & \rho_{2,3} & \rho_{2,4} & \rho_{2,5} & \rho_{2,6} & \rho_{2,7} \\ \rho_{1,3} & \rho_{2,3} & 1 & \rho_{3,4} & \rho_{3,5} & \rho_{3,6} & \rho_{3,7} \\ \rho_{1,4} & \rho_{2,4} & \rho_{3,4} & 1 & \rho_{4,5} & \rho_{4,6} & \rho_{4,7} \\ \rho_{1,5} & \rho_{2,5} & \rho_{3,5} & \rho_{4,5} & 1 & \rho_{5,6} & \rho_{5,7} \\ \rho_{1,6} & \rho_{2,6} & \rho_{3,6} & \rho_{4,6} & \rho_{5,6} & 1 & \rho_{6,7} \\ \rho_{1,7} & \rho_{2,7} & \rho_{3,7} & \rho_{4,7} & \rho_{5,7} & \rho_{6,7} & 1 \end{bmatrix} \Lambda_{hvar}^{1/2}$$

covariance - *cov*

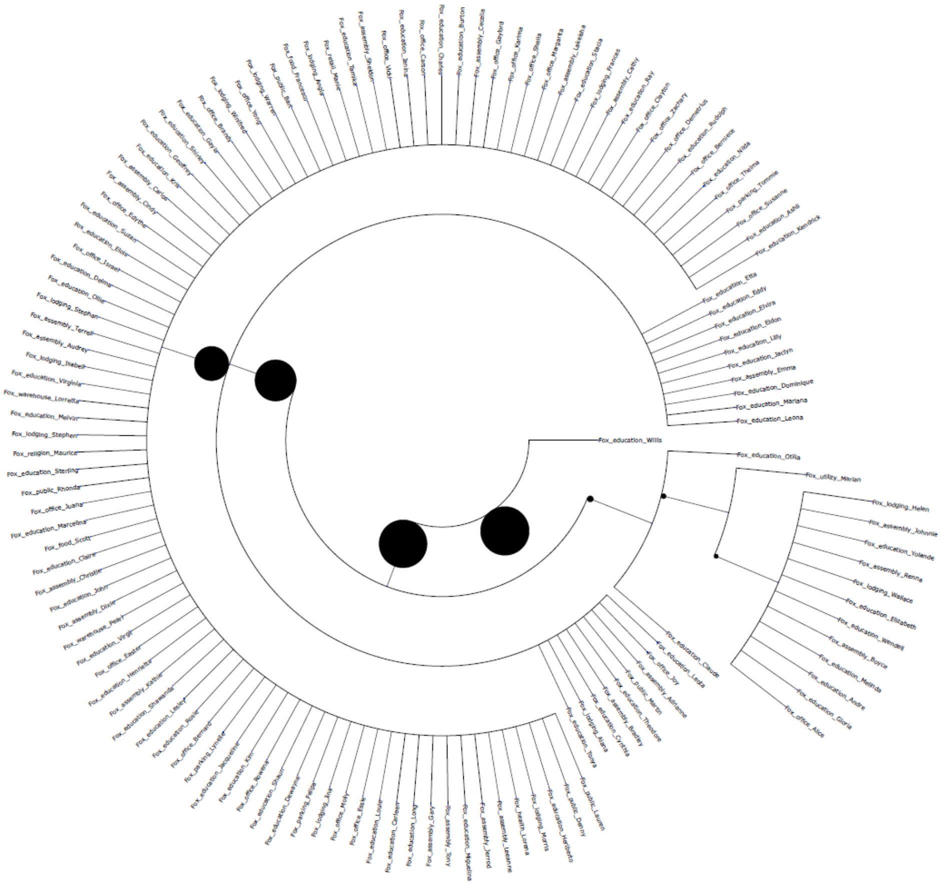
Case studies

Eneco

41 residential households
The Netherlands
Horizon: 3 years, 2019-2022

BUILDING DATA GENOME PROJECT 2

133 electric-meter measurements
United States of America
Horizon: 2 full years, 2016-2017

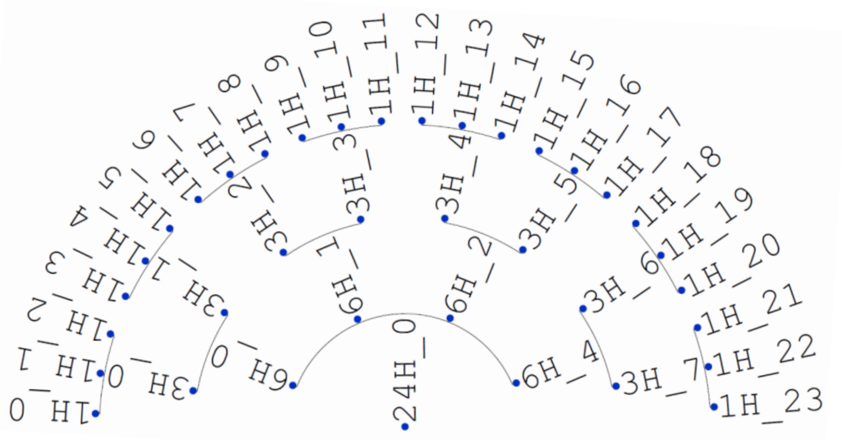


Credit: https://freepik.com/

Case studies

41 residential households
The Netherlands
Horizon: 3 years, 2019-2022

133 electric-meter measurements
United States of America
Horizon: 2 full years, 2016-2017

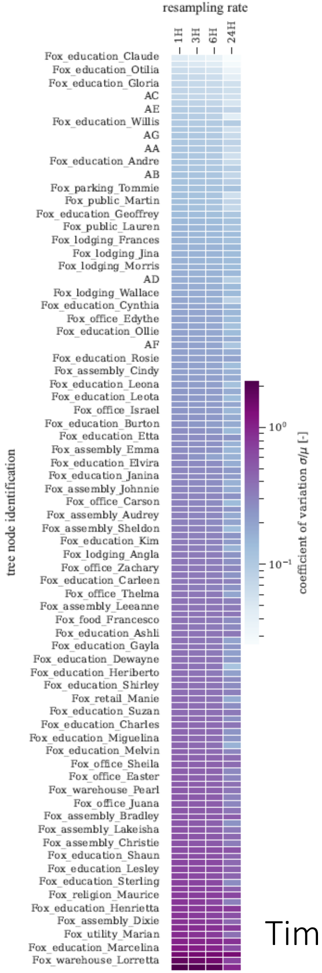


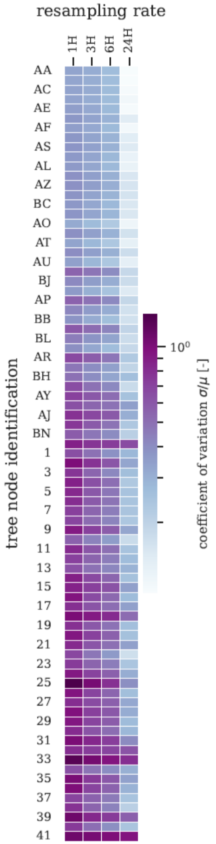
Credit: https://freepik.com/

Case studies

41 residential households
The Netherlands
Horizon: 3 years, 2019-2022

133 electric-meter measurements
United States of America
Horizon: 2 full years, 2016-2017



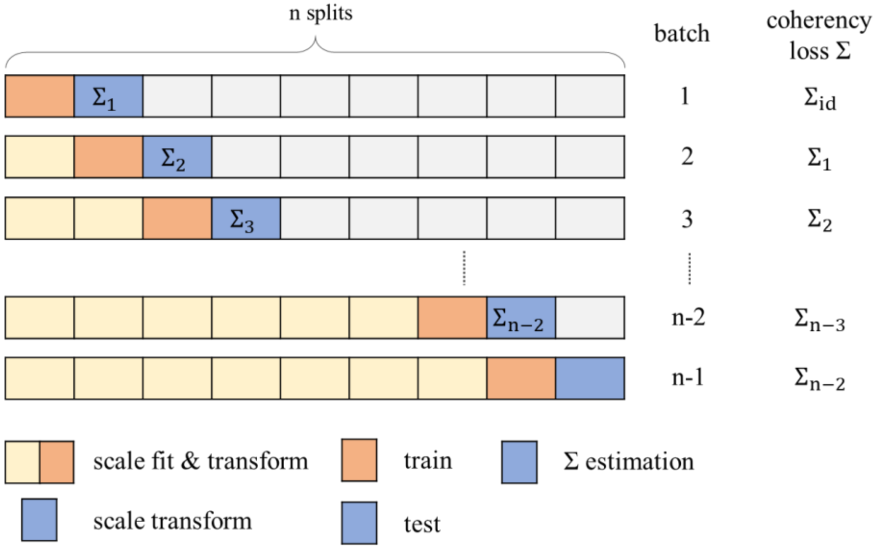


Time series variability

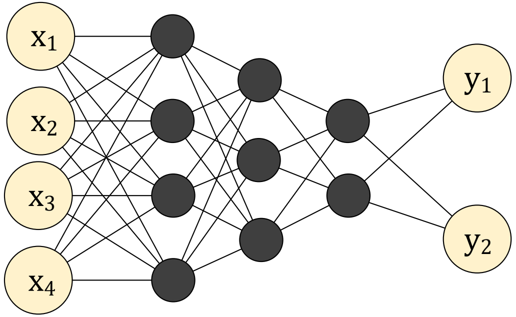
Credit: https://freepik.com/

Training setup

Data partitioning & transformation



Designing hierarchical regressors



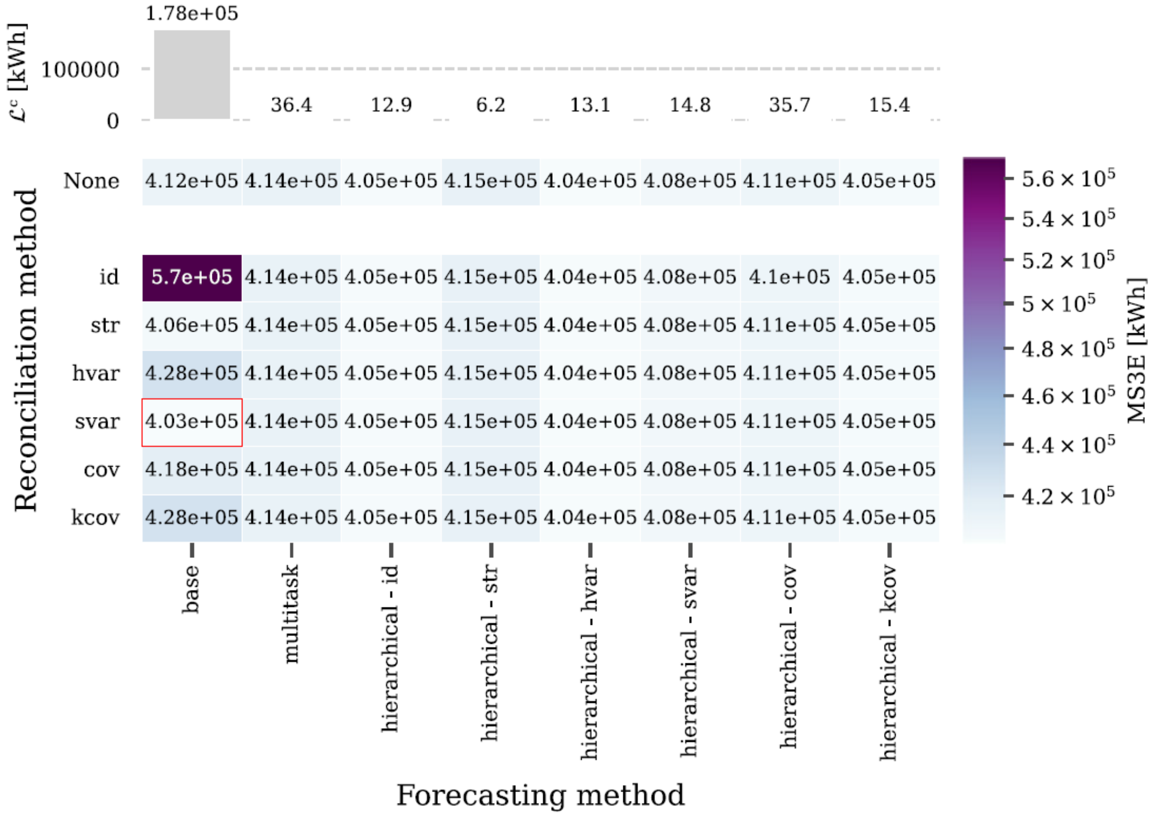
Activation function
sigmoid

Dropout ratio
0.2

Results

Eneco dataset
41 residential households

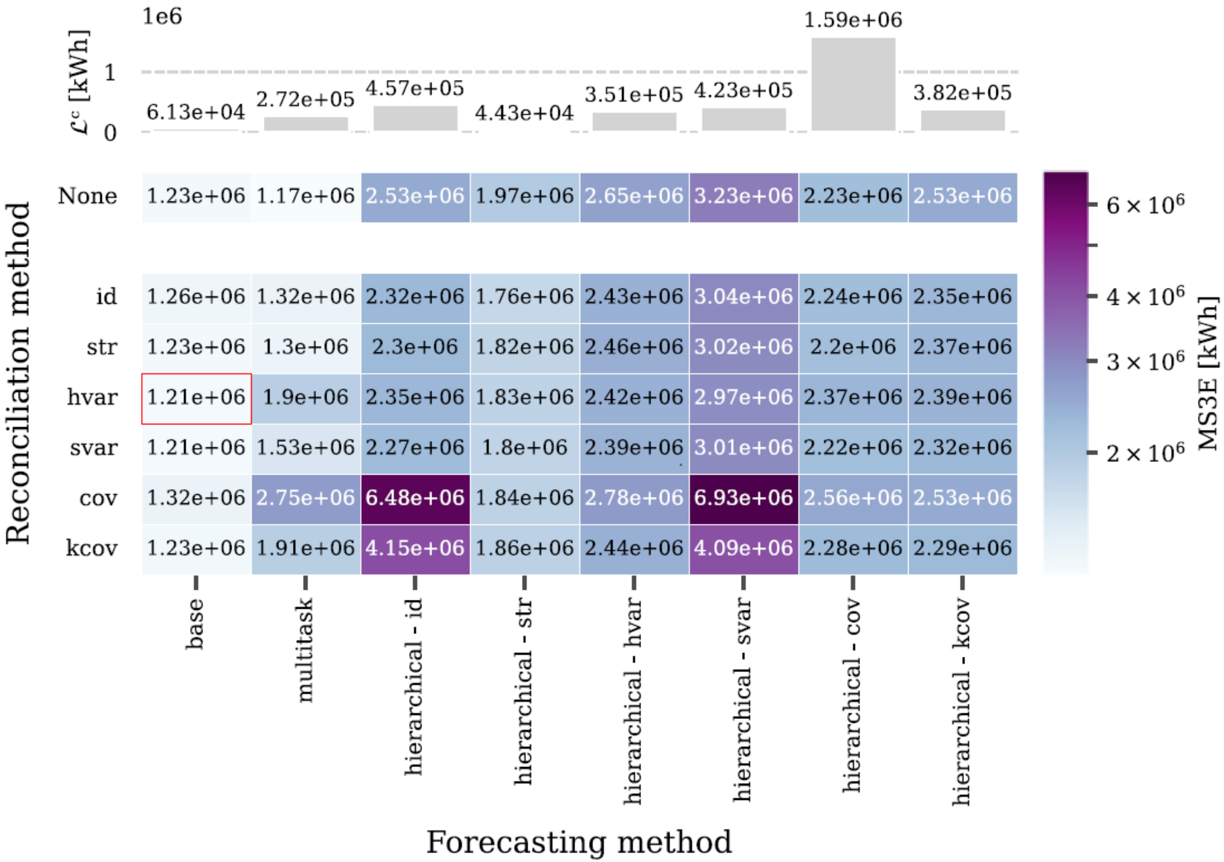
Spatial hierarchy
Hour-ahead forecast



Results

Eneco dataset
41 residential households

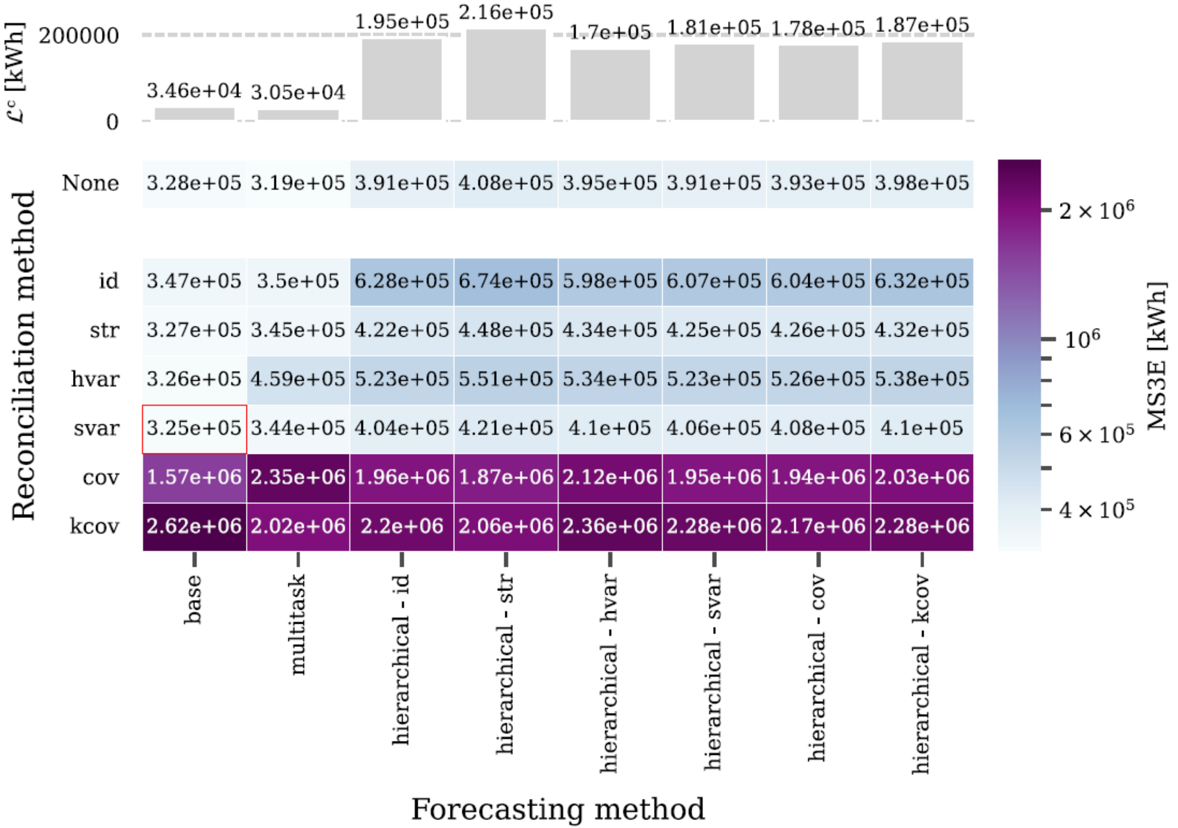
Temporal hierarchy
Day-ahead forecast



Results

Eneco dataset
41 residential households

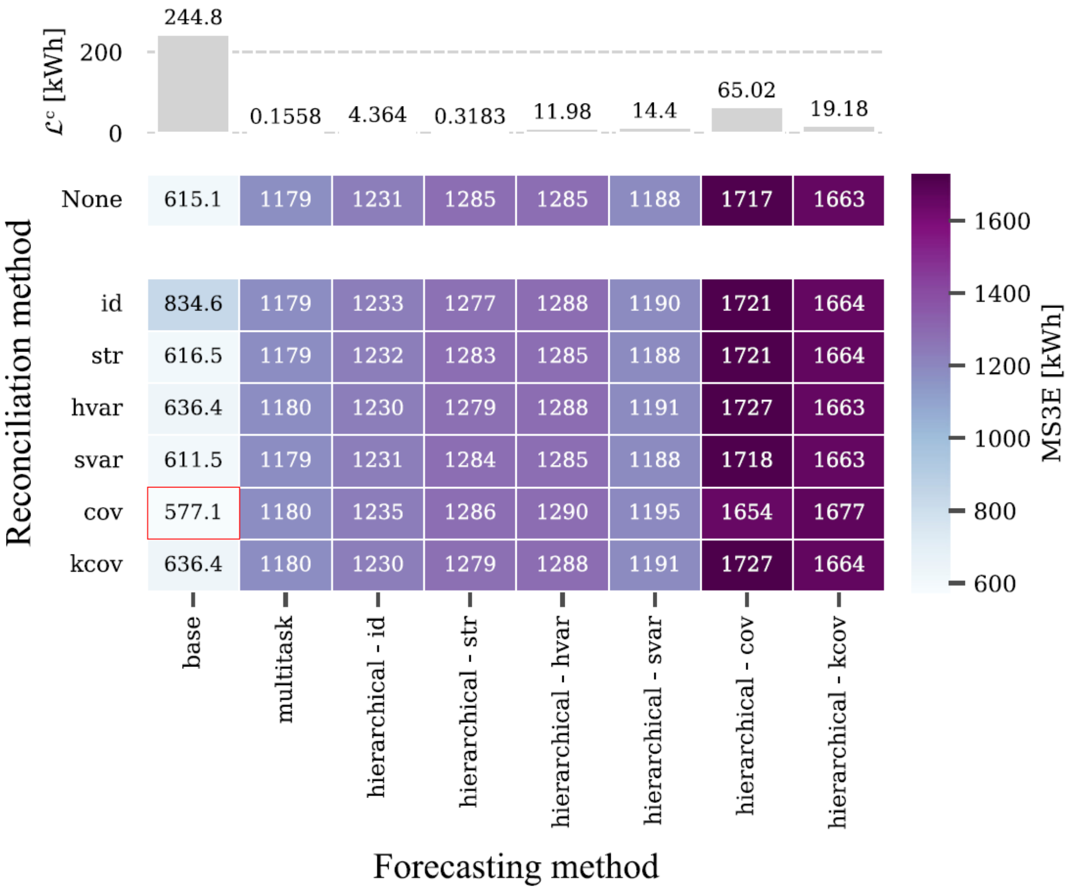
Spatiotemporal hierarchy
Day-ahead forecast



Results

BDG2 dataset
133 electric-meters

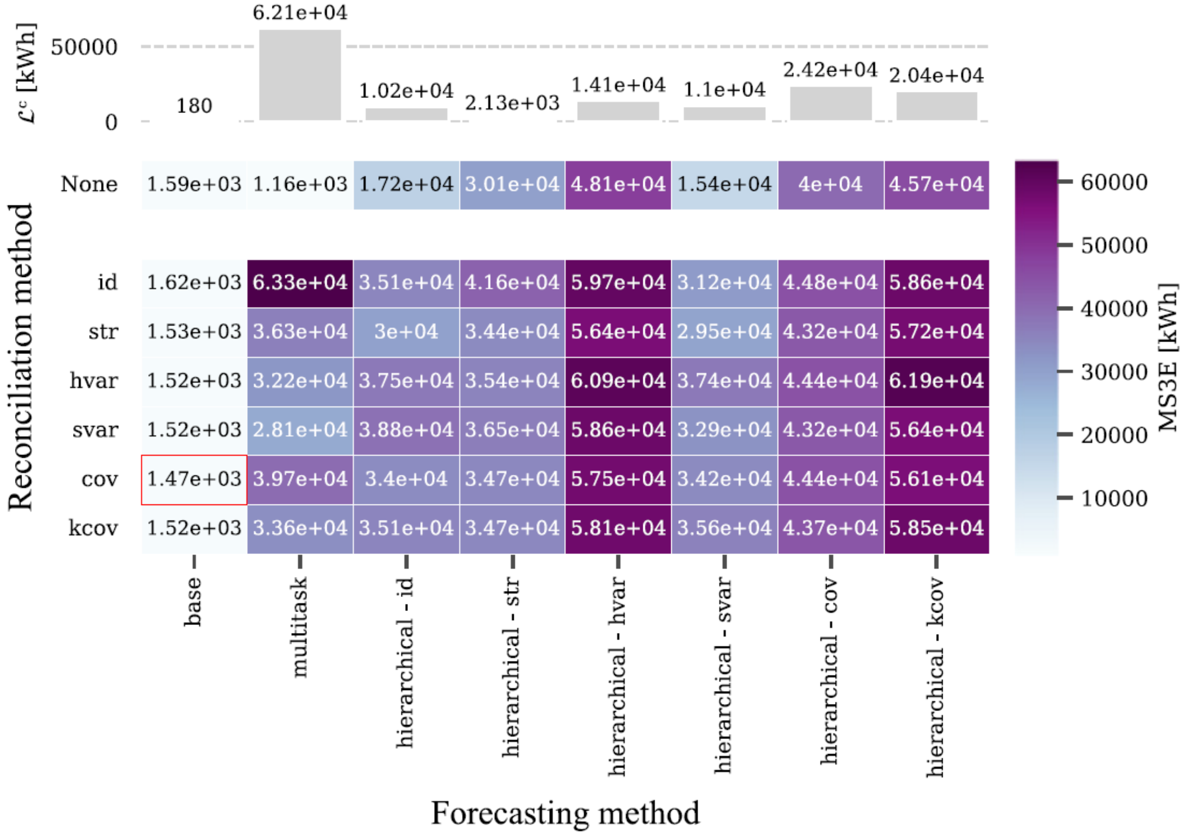
Spatial hierarchy
Hour-ahead forecast



Results

BDG2 dataset
133 electric-meters

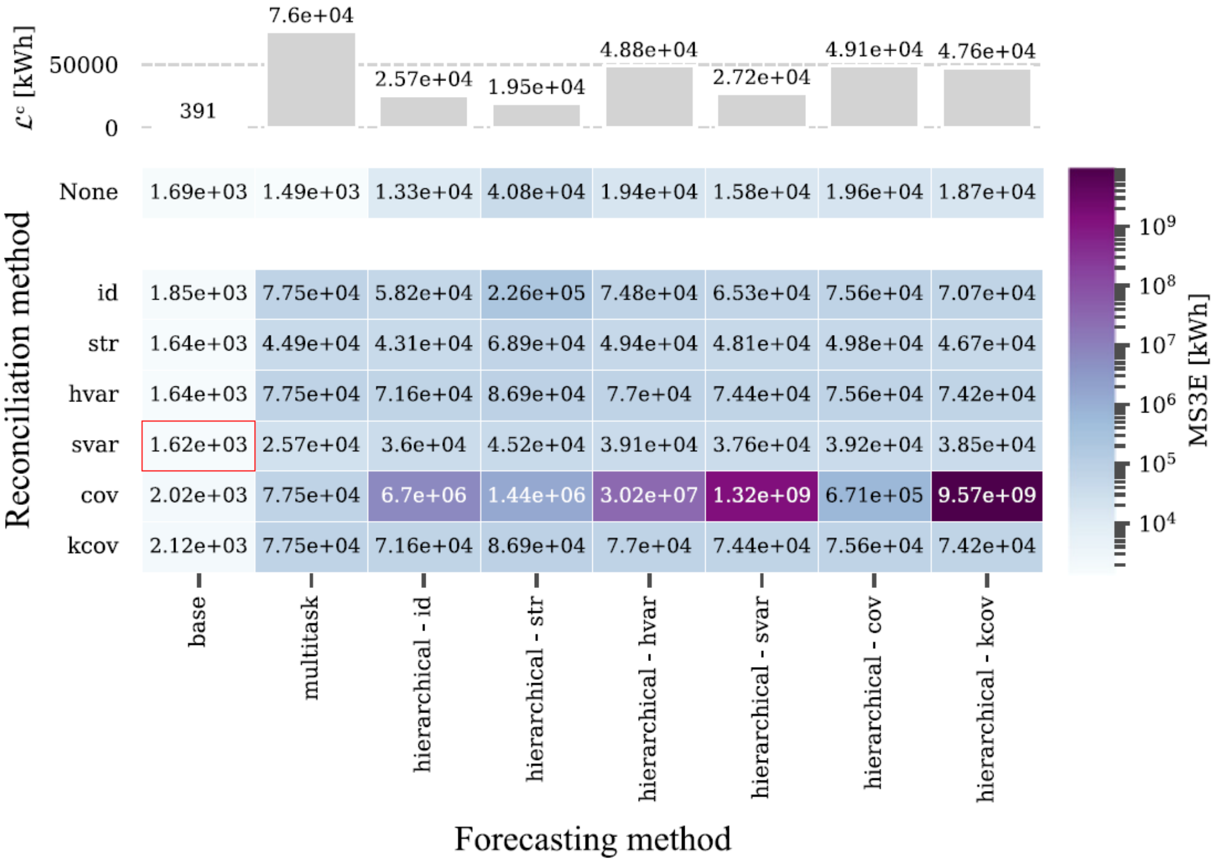
Temporal hierarchy
Day-ahead forecast



Results

BDG2 dataset
133 electric-meters

Spatiotemporal hierarchy
Day-ahead forecast



Findings & method adjustments

Arduous learning

Many weights to update with little data

- ➢ Explore architectures with fewer weights

Induced coherency over accuracy

Temporal hierarchy - periodicity falling on prediction horizon & fewer data points

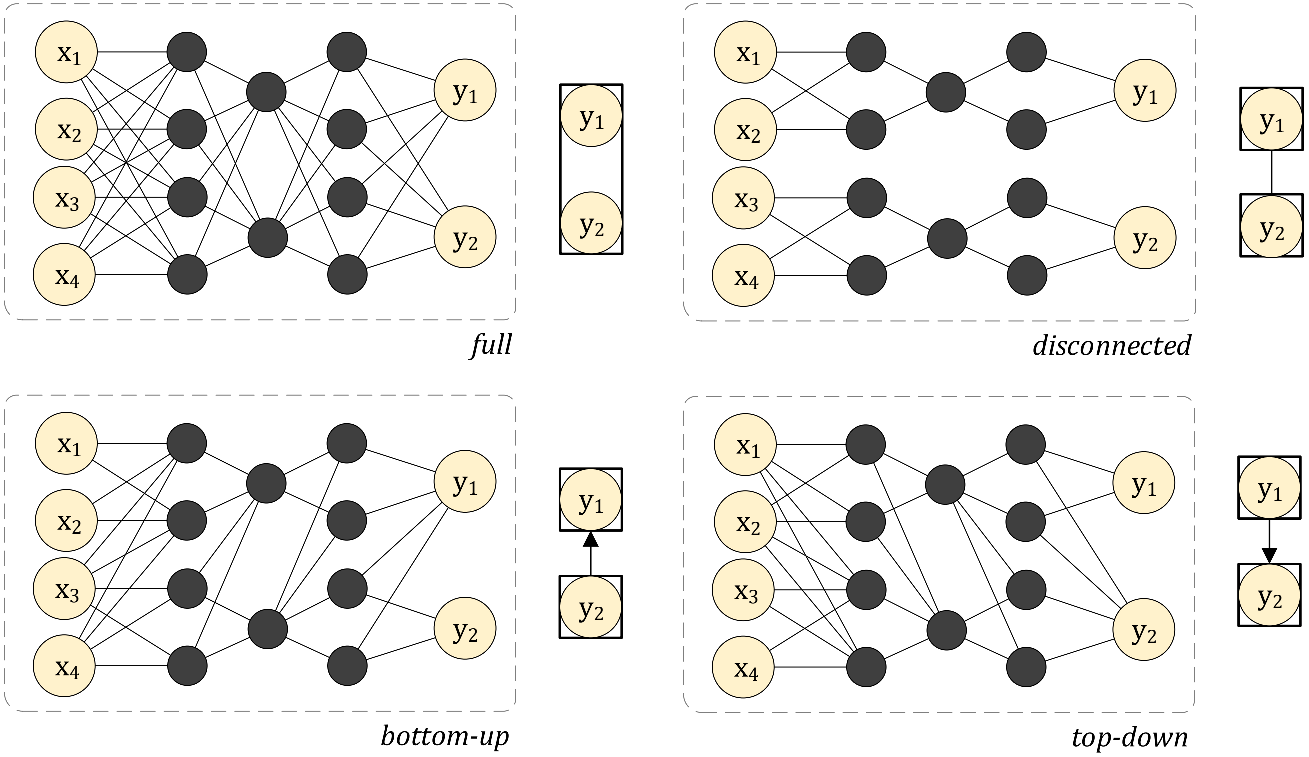
Faulty coherency learning

Normalization tempers the coherency learning

- ➢ Explore methods robust to varying ranges of target values
- ✓ Batch normalization

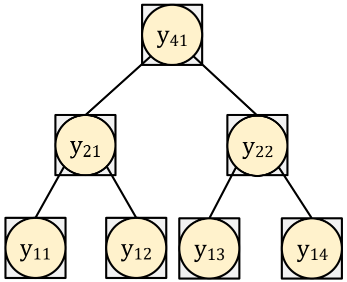
Method extension

Tailored designs of neural networks for efficient learning

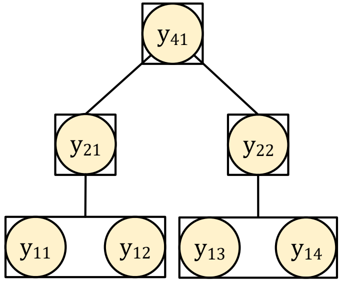


Method extension

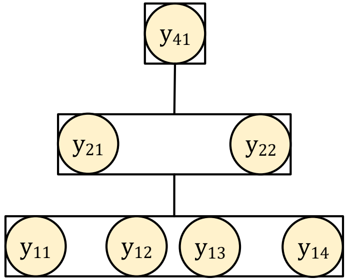
Tailored designs of neural networks for efficient learning



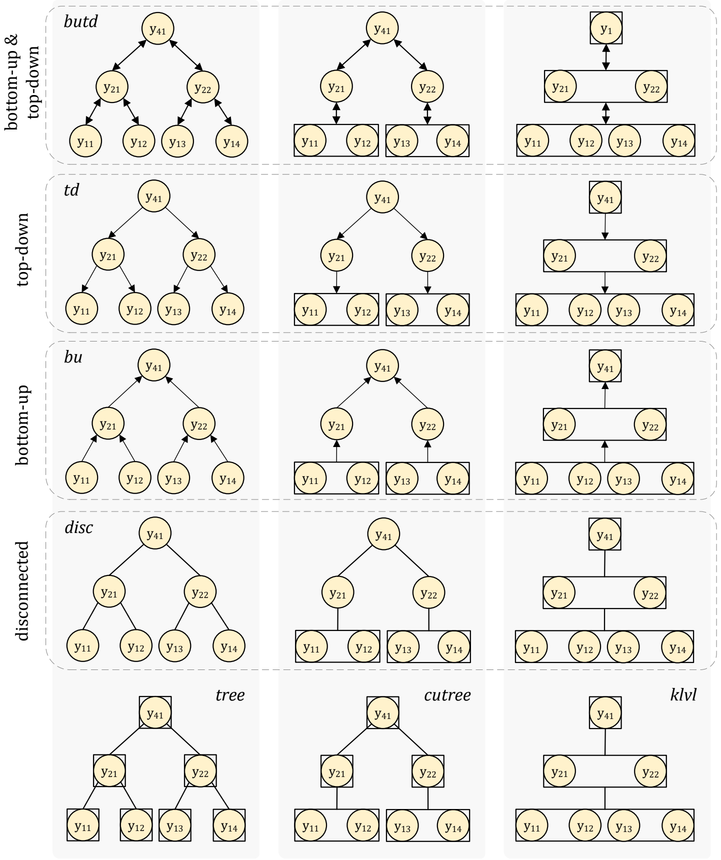
tree partitioning



cut-tree partitioning



k-level partitioning

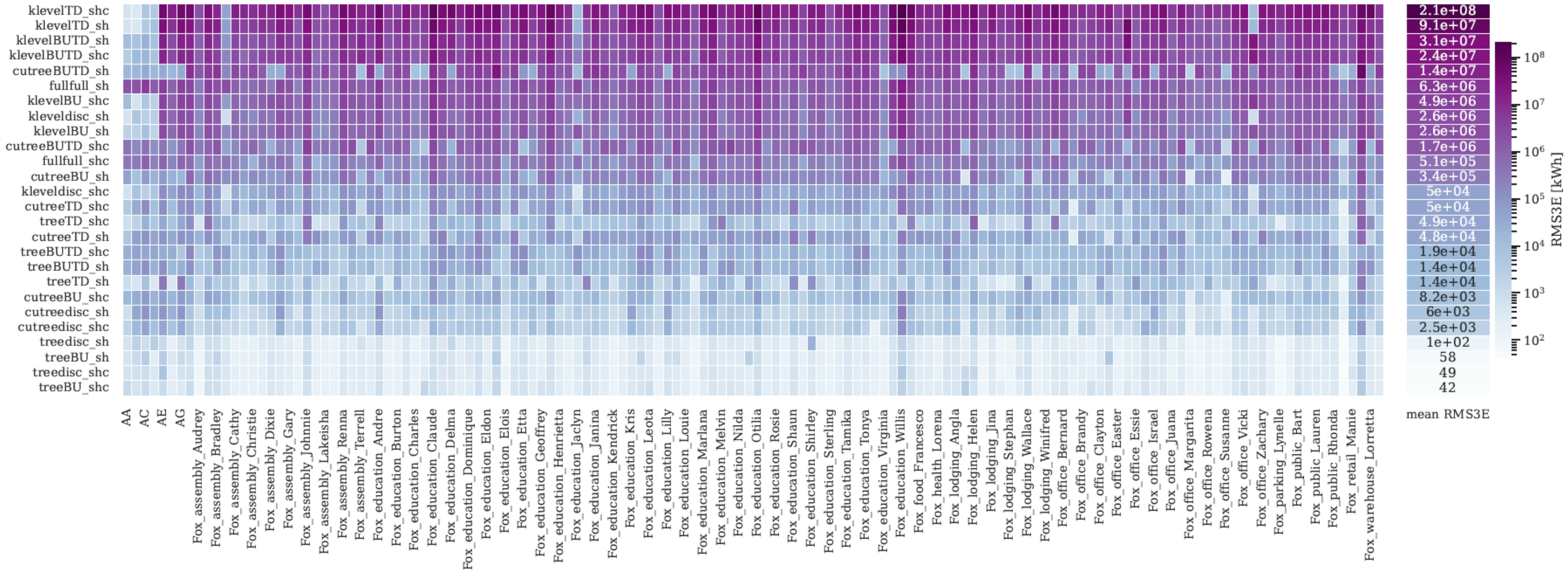


tree partitioning cut-tree partitioning k-level partitioning

Results

BDG2 dataset

Spatial tree – hour ahead



Is coherency learning worth it?

Value from connecting scales?

+ Improved forecast performances with **simpler** models

- Requires centralized data
- Complexity/value trade-off still too little

Open-source research

https://github.com/JulienLeprince

Future work

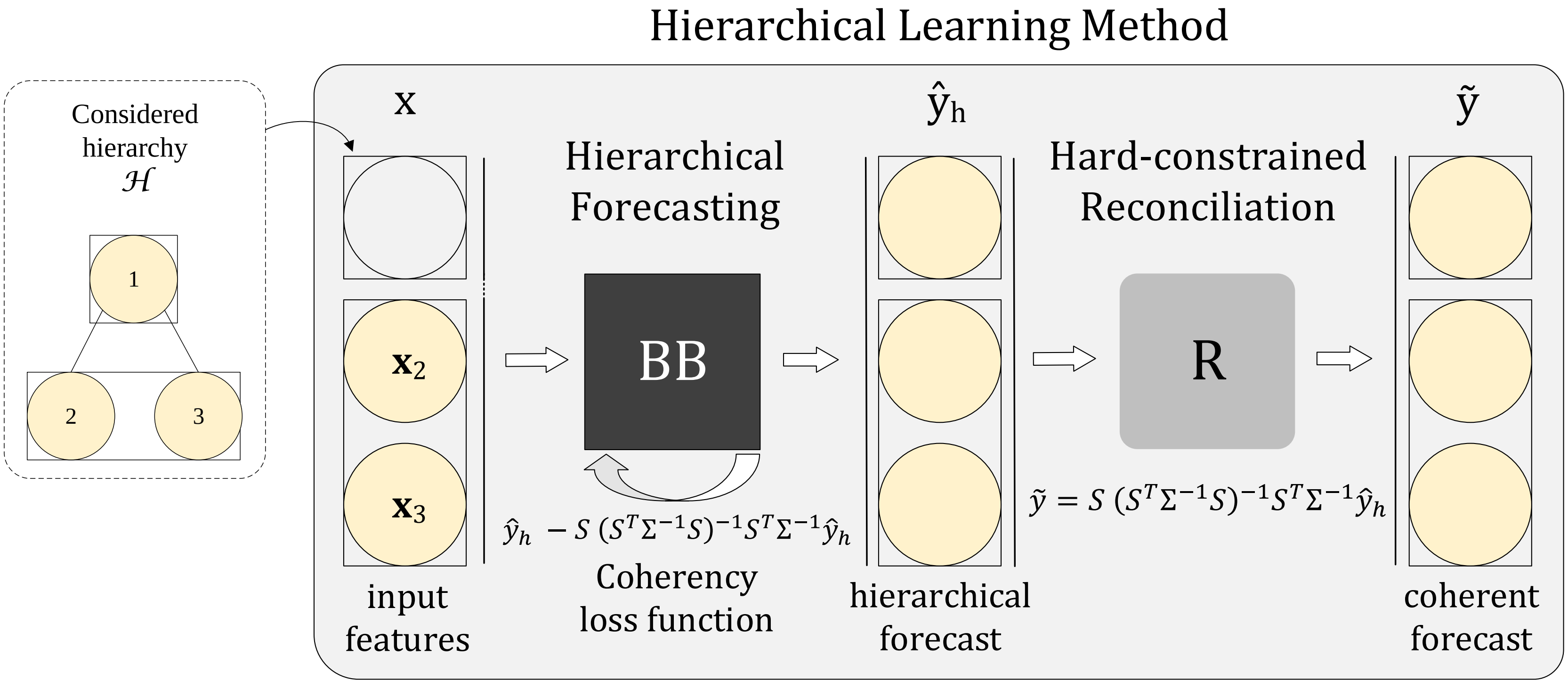
1. Other machine learning methods
2. Varying hierarchical structures
   a. Cluster homogeneity
   b. Tree depth, width etc.
3. Linking prediction performances to hierarchical time-series characteristics
4. Scalability
   a. Distributed approaches?
   b. Piecewise reconciliation

Forecasting hierarchies with coherency-learning

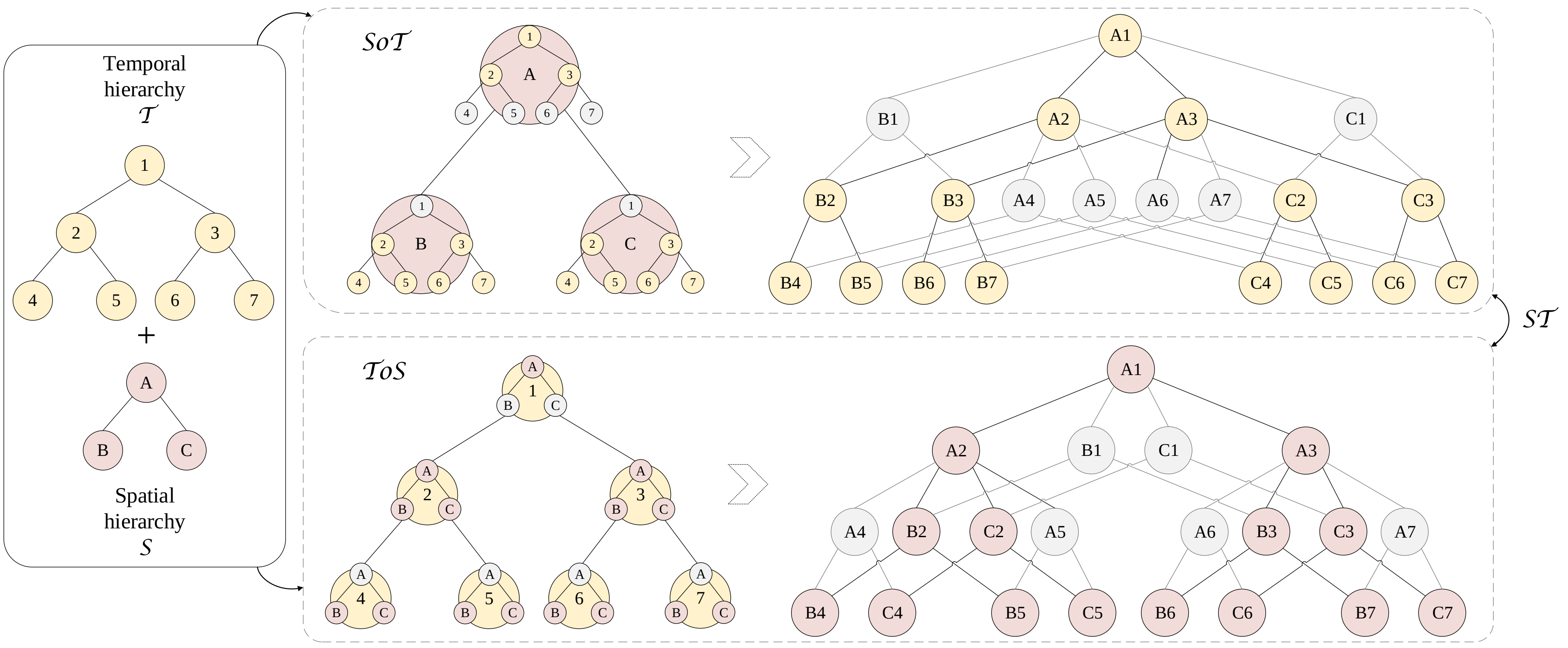
Thank you

Julien Leprince | Sep 7th 2023

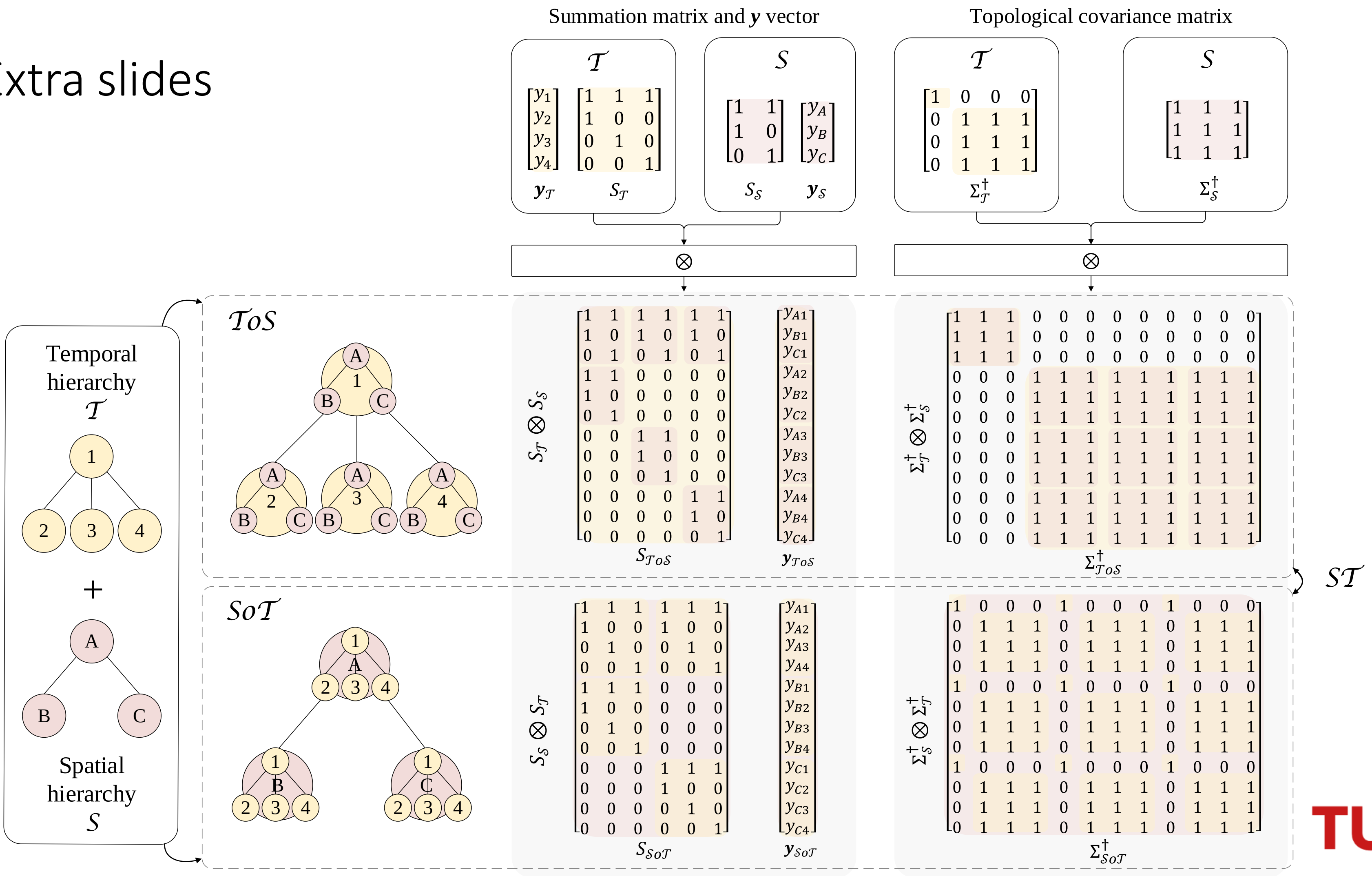
Extra slides



Extra slides

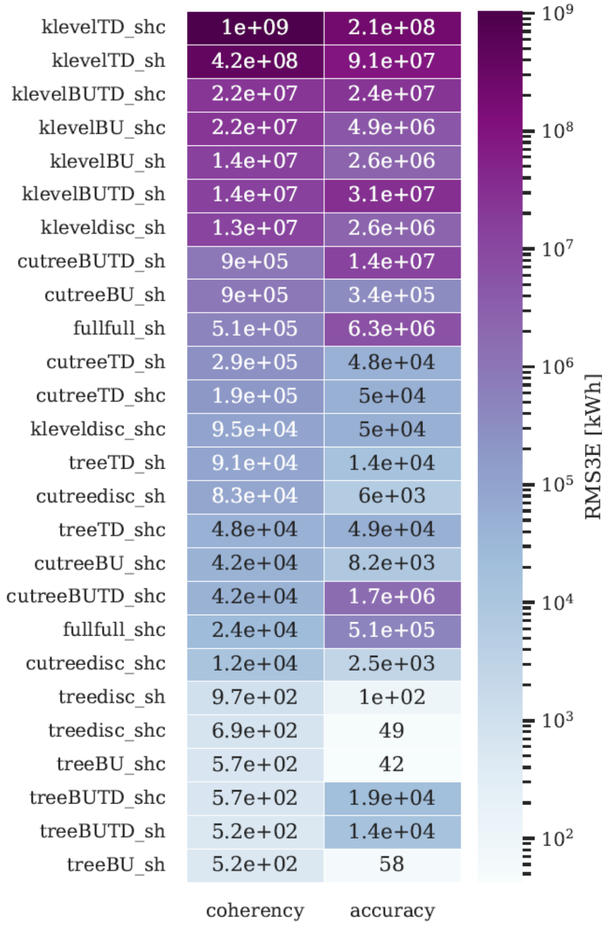


Extra slides



Extra slides

Results



Extra slides

Case studies

	Characteristics		Spatial	Temporal	Spatiotemporal
Case study 1	n	[#]	81	37	2,997
	m	[#]	41	24	984
	horizon	[h]	1	24	24
Case study 2	n	[#]	140	37	1,998
	m	[#]	133	24	1,200
	horizon	[h]	1	24	24

Credit: https://freepik.com/