RESEARCH SEMINAR

Forecasting with Deep Temporal Hierarchies

Personal Details

Filotas Theodosiou: MSc Data Science

Research Associate at VIVES University of Applied Sciences

Email: filotas.thedosiou@vives.be
LinkedIn: https://www.linkedin.com/in/filotas-theodosiou-4916881a1/

Main Resources:

1. Forecasting with Deep Temporal Hierarchies: A novel way for forecasting with temporal hierarchies based on deep learning models. - Master Thesis -

2. Theodosiou, Filotas & Kourentzes, Nikolas. (2021) Forecasting with Deep Temporal Hierarchies. SSRN Scholarly Paper ID 3918315. Social Science Research

3. Theodosiou, Filotas & Kourentzes, Nikolaos. (2021). Deep Learning Temporal Hierarchies for Interval Forecasts. International Conference of AI in Finance. Time Series Workshop

Content

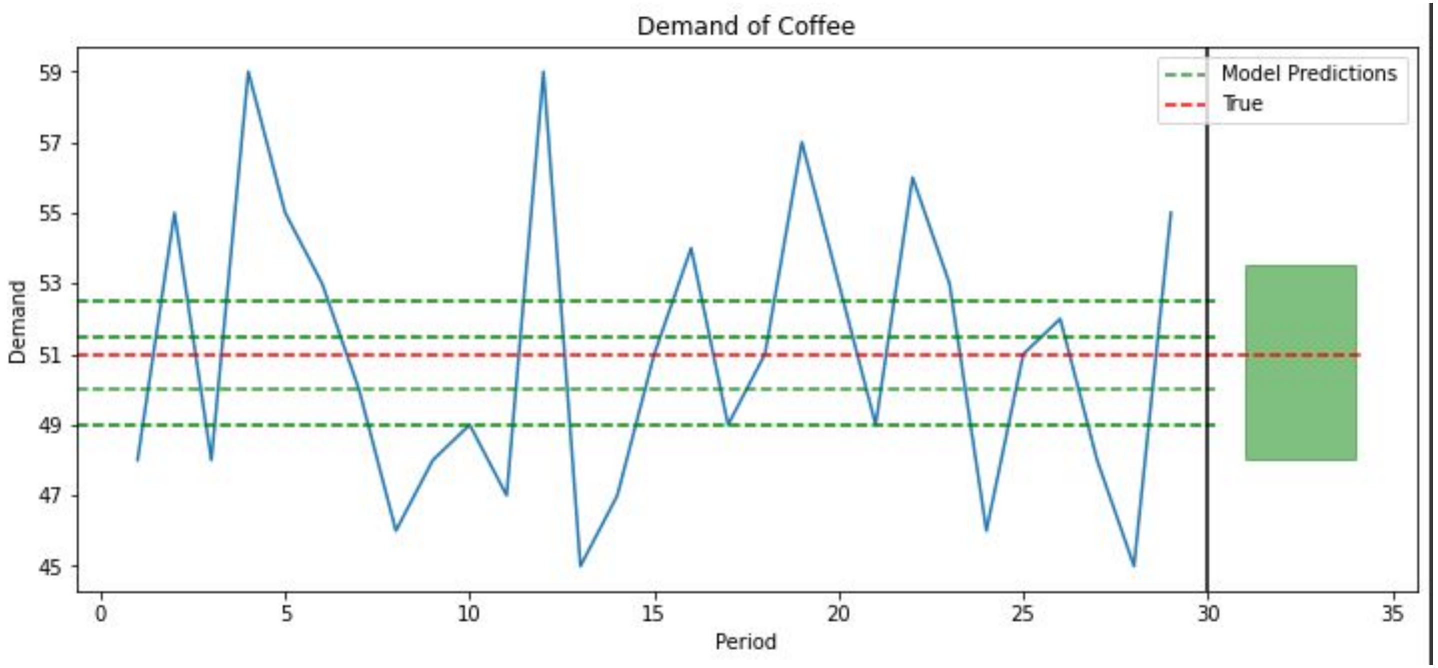
Why Temporal Hierarchies?
Model Selection Uncertainty

We forecast the demand of coffee.

- We assume demand is independent of weather/price/location
- Constant demand with mean = 51 with some noise

We generate predictions using different models
But we get high forecast uncertainty

- Model selection uncertainty
- Parametric uncertainty
- Forecast uncertainty originated from noise



Why Temporal Hierarchies?
Forecast Combinations

In every forecasting problem we look for a "good-enough" model.

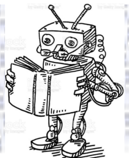
- A non trivial and usually difficult task
- Let's avoid it!

Forecast Combinations

- Instead of looking for the "best-model", combine diverse predictions
- Simple combinations show accuracy benefits

How to get diverse predictions?

Very bad forecasts will harm the combined predictions



Kourentzes, Nikolaos & Barrow, Devon & Petropoulos, Fotios. (2018). Another look at forecast selection and combination: Evidence from forecast pooling. International Journal of Production Economics. 209. 10.1016/j.ijpe.2018.05.019.

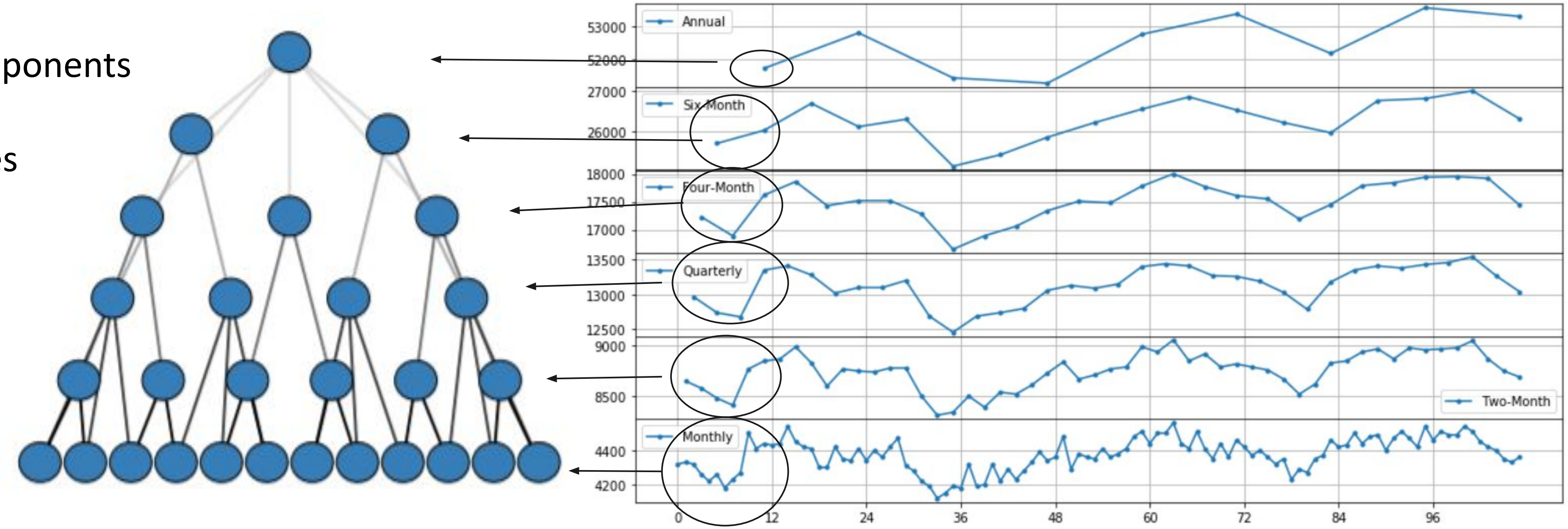
Temporal Hierarchies in a Nutshell
Introduction to Temporal Hierarchies

How to get diverse predictions?

Change your viewpoint through Temporal Aggregations

Temporal Aggregations:

- Filter high frequency components (seasonality)
- Boost high frequency ones (trend)
- Reveal hidden structural information



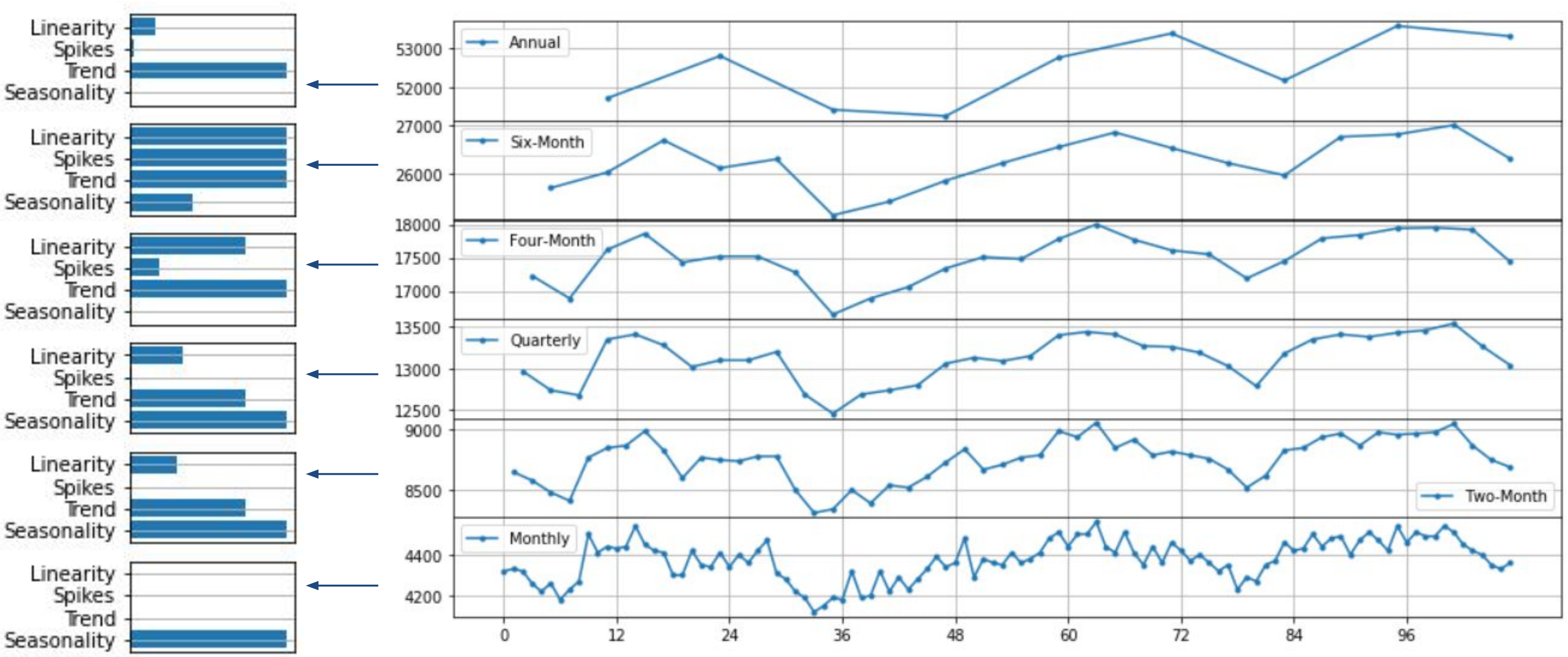
Temporal Hierarchies in a Nutshell
Introduction to Temporal Hierarchies

How to get diverse predictions?
Change your viewpoint through Temporal Aggregations

The Diversity Among Levels

Features are strengthen/weaken at each level

What if we combine information from each level into a single forecast?

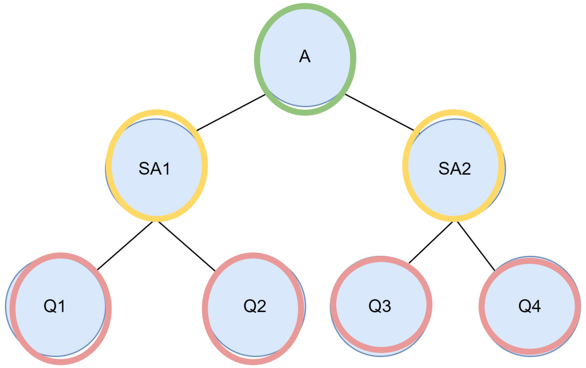


Features/components estimated using tsfeatures in R
- Spikes measures spikness. Measured as the variance of the leave-one-out variances of the remainder component after STL
- Linearity is measured based on the coefficients of an orthogonal quadratic regression

Temporal Hierarchies in a Nutshell
How to Forecast with Temporal Hierarchies (THieF)

Hierarchical Forecasting.

- Bottom up
- Top down -> Biased



b = (Q1, Q2, Q3, Q4)
y = (A, SA1, SA2, **b**)
y = **Sb**

$$\boldsymbol{S} = \begin{array}{cl} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & \boldsymbol{I}_m & & \end{bmatrix} & \begin{array}{l} \text{Top level} \\ \text{Middle level(s)} \\ \\ \text{Bottom level} \end{array} \end{array}$$

Temporal Hierarchy Forecasts (THieF)

1. Generate Base Forecasts $\hat{\boldsymbol{y}}_h$
2. Reconcile (Just a fancy word for combine) into the bottom level
3. Reconstruct the entire hierarchy

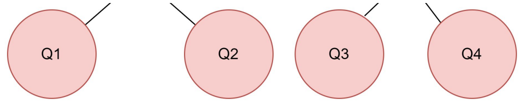
Reconciliation:
Multiply with **G**

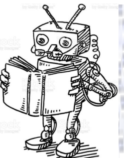
???

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_h$$

Reconciled coherent predictions

Base forecasts





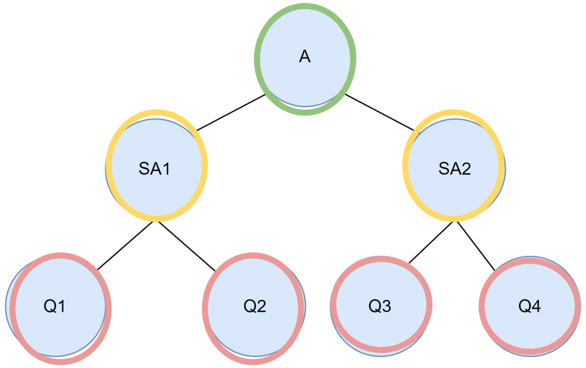
Panagiotelis, A., Athanasopoulos, G., Gamakumara, P.,& Hyndman, R.J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. International Journal of Forecasting, 37, 343–359.

Athanasopoulos, George & Hyndman, Rob & Kourentzes, Nikolaos & Petropoulos, Fotios. (2017). Forecasting with Temporal Hierarchies. European Journal of Operational Research. 262. 10.1016/j.ejor.2017.02.046.

Temporal Hierarchies in a Nutshell
How to Forecast with Temporal Hierarchies (THieF)

Hierarchical Forecasting.

- Bottom up
- Top down -> Biased



b = (Q1, Q2, Q3, Q4)
y = (A, SA1, SA2, **b**)
y = **Sb**

$$S = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & I_m & & \end{bmatrix}$$

Top level
Middle level(s)
Bottom level

Temporal Hierarchy Forecasts (THieF)

1. Generate Base Forecasts $\hat{y}_h$
2. Reconcile (Just a fancy word for combine) into the bottom level
3. Reconstruct the entire hierarchy

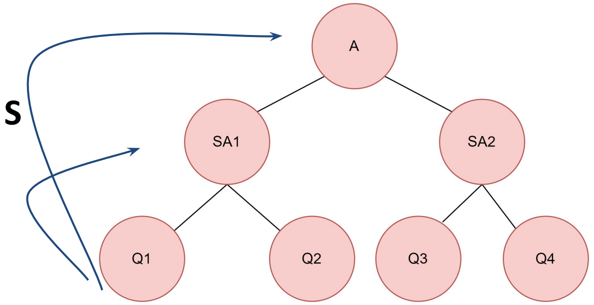
???

$$\tilde{y}_h = S\color{red}{G}\hat{y}_h$$

Reconciled coherent predictions

Base forecasts

Reconstruct
Multiply with **S**



Panagiotelis, A., Athanasopoulos, G., Gamakumara, P.,& Hyndman, R.J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. International Journal of Forecasting, 37, 343–359.

Athanasopoulos, George & Hyndman, Rob & Kourentzes, Nikolaos & Petropoulos, Fotios. (2017). Forecasting with Temporal Hierarchies. European Journal of Operational Research. 262. 10.1016/j.ejor.2017.02.046.

Temporal Hierarchies in a Nutshell
Estimate G

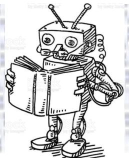
Optimal Reconciliation Matrix G: $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$
(in terms of least squares)

$\boldsymbol{W}_h$ is the positive definite covariance matrix of base forecast errors.

Problem: Estimating $\boldsymbol{W}_h$ in practise is very challenging.
Solution: Again, avoid the problem and rely on approximations

Two diagonal estimators which will be useful later:
1. Structural Scaling: A diagonal matrix containing the number of errors contributing to each level.
2. Variance Scaling: A diagonal matrix with the in-sample one-step-ahead error variance of each level.

Possible Limitations:
1. Strict Linearity → limited capacity
2. Partial Covariance Estimation → partial consideration of relationships between levels
3. Assumptions → each estimators relies on a number of assumptions



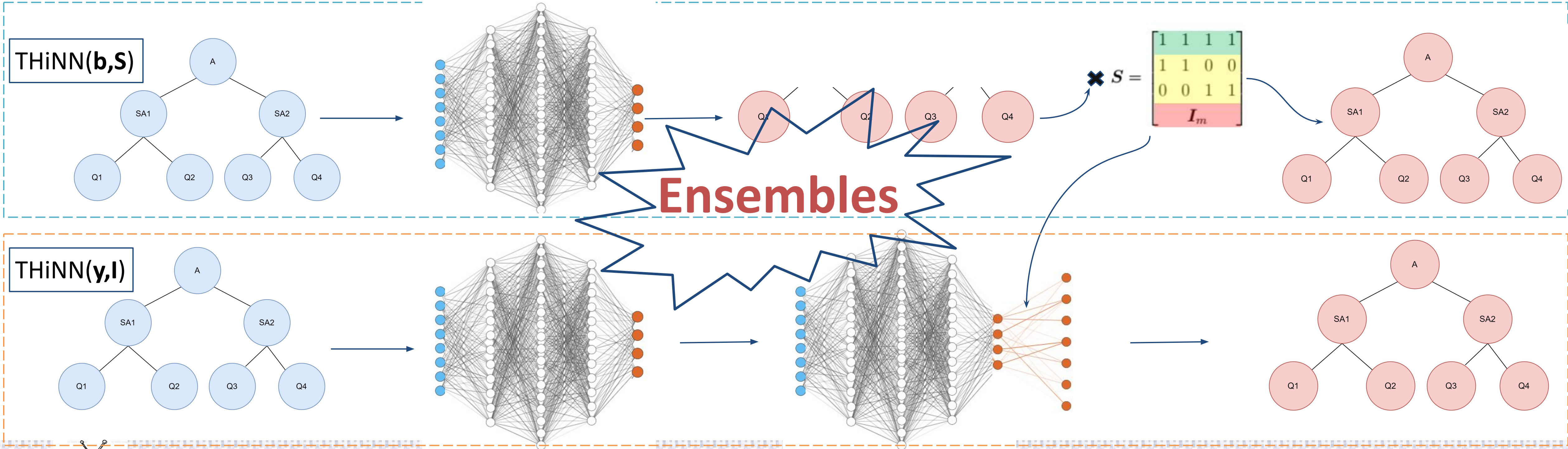
Shanika L. Wickramasuriya, George Athanasopoulos & Rob J. Hyndman (2019) Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization, Journal of the American Statistical Association, 114:526, 804-819

Athanasopoulos, George & Hyndman, Rob & Kourentzes, Nikolaos & Petropoulos, Fotios (2017). Forecasting with Temporal Hierarchies. European Journal of Operational Research. 262. 10.1016/j.ejor.2017.02.046

Temporal Hierarchy Neural Network (THiNN)
A Non-Linear Data-Driven Reconciler

Approximate G with Temporal Hierarchy Neural Network (THiNN)

- NNs: Non-Linear function approximators with success in numerous tasks
- Relax classical restrictions
- Preference over classical ML (Gradient Boosted Decision Trees) due to architectural freedom.

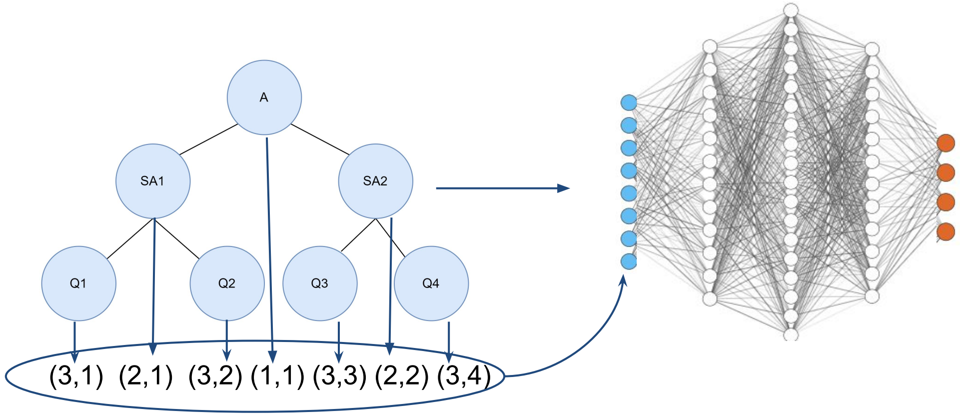


Theodosiou, Filotas & Kourentzes, Nikolas. (2021) Forecasting with Deep Temporal Hierarchies. SSRN Scholarly Paper ID 3918315. Social Science Research

Temporal Hierarchy Neural Network (THiNN)
Positional Information

Limitation of Feed Forward NNs:

- Input is random flatten values
- No positional information
- No level-specific information
- Hard time understanding values



Positional Encoding :

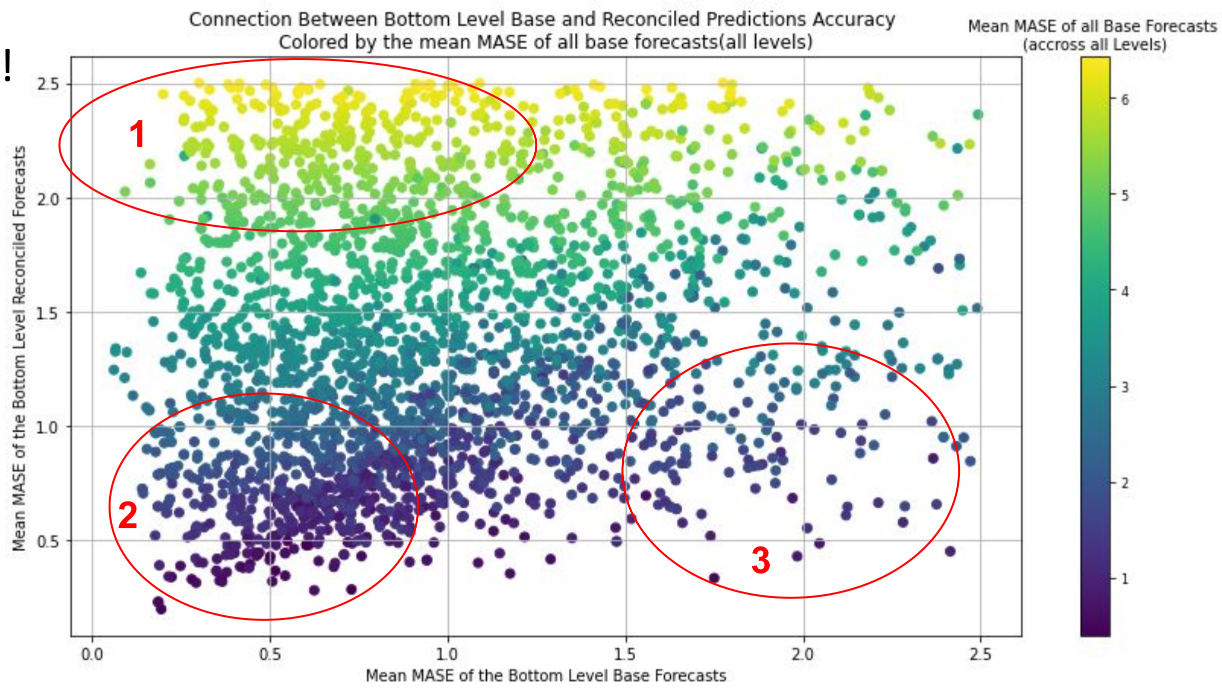
- Inspired by Transformers
- Described the location (level) and the position (time-step) of the forecast.
- Pair of type (level, time-step) -> fails on large hierarchies
- Pairs of (cos,sin)
- One-hot-encoded values

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. arXiv:1706.03762 [cs].

Temporal Hierarchy Neural Network (THiNN)
Connection between Base and Reconciled Forecasts

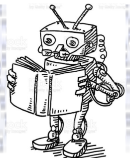
Inaccurate initial forecasts can lead to bad forecast combinations!
So, Bad base forecasts can negatively affect reconciliation?

- Measure MASE of bottom level (MASE_b) base forecasts (ETS)
- Measure MASE of all (MASE_tot) base forecasts (ETS)
- Measure MASE of bottom level (MASE_rec) reconciled
  - Use Structural Scaling (no parametric uncertainty)
  - Accurate reconciled bottom level => Accurate reconciled hierarchy (coherence)



Results:

1. Low MASE_B & High MASE_tot => High MASE_rec
2. Low MASE_B & Low MASE_tot = > Low MASE_rec
3. High MASE_B & Low MASE_tot => Low MASE_rec



Pritularga, K.F., Svetunkov, I., Kourentzes, N., 2021. Stochastic coherency in forecast reconciliation. International Journal of Production Economics 240, 108221. https://doi.org/10.1016/j.ijpe.2021.108221

Kourentzes, Nikolaos & Barrow, Devon & Petropoulos, Fotios. (2018). Another look at forecast selection and combination: Evidence from forecast pooling. International Journal of Production Economics. 209. 10.1016/j.ijpe.2018.05.019.

Temporal Hierarchy Neural Network (THiNN)
Base Forecasts Accuracy "Hints"

Insights:

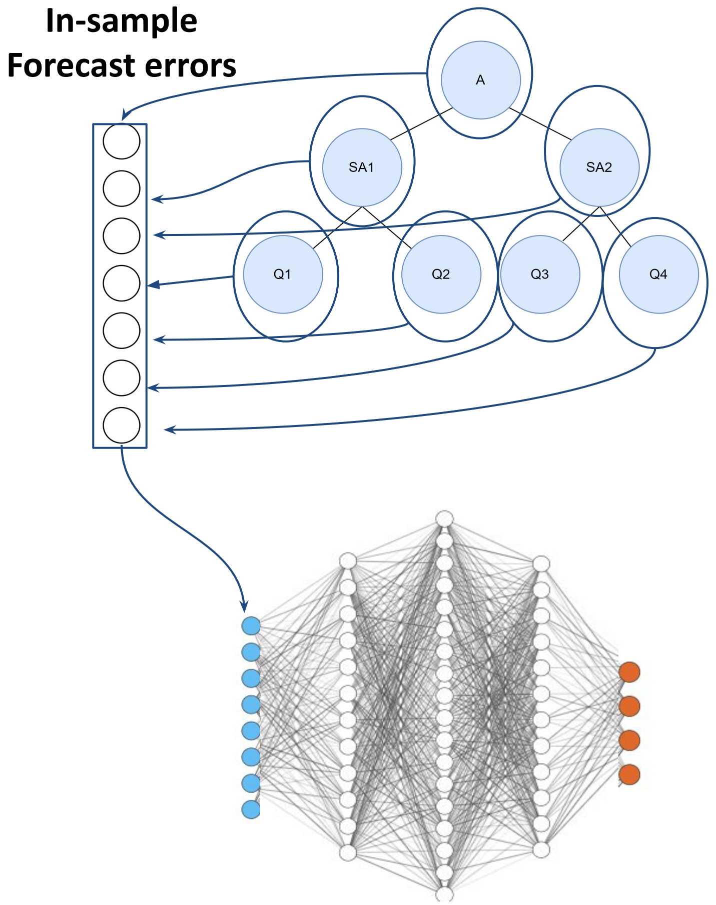
- Having one/two accurate base forecasts does not guarantee good reconciliation
- Bad forecasts drag reconciled predictions outside the accurate subspace
- Problems with model misspecifications on base forecasts
  (eg ETS for all base forecasts on "difficult" frequencies?)

Ideally THiNN should focus on more accurate base forecasts.
Not possible.

Assumption: in-sample performance gives an adequate picture for out-of-sample

Include mean in-sample accuracy of each base forecast

- THiNN corrects base forecasts based on their in-sample accuracy
- THiNN is not overly affected by inaccurate base forecasts
- Inaccurate base forecasts will be dragged "more" into the accurate subspace



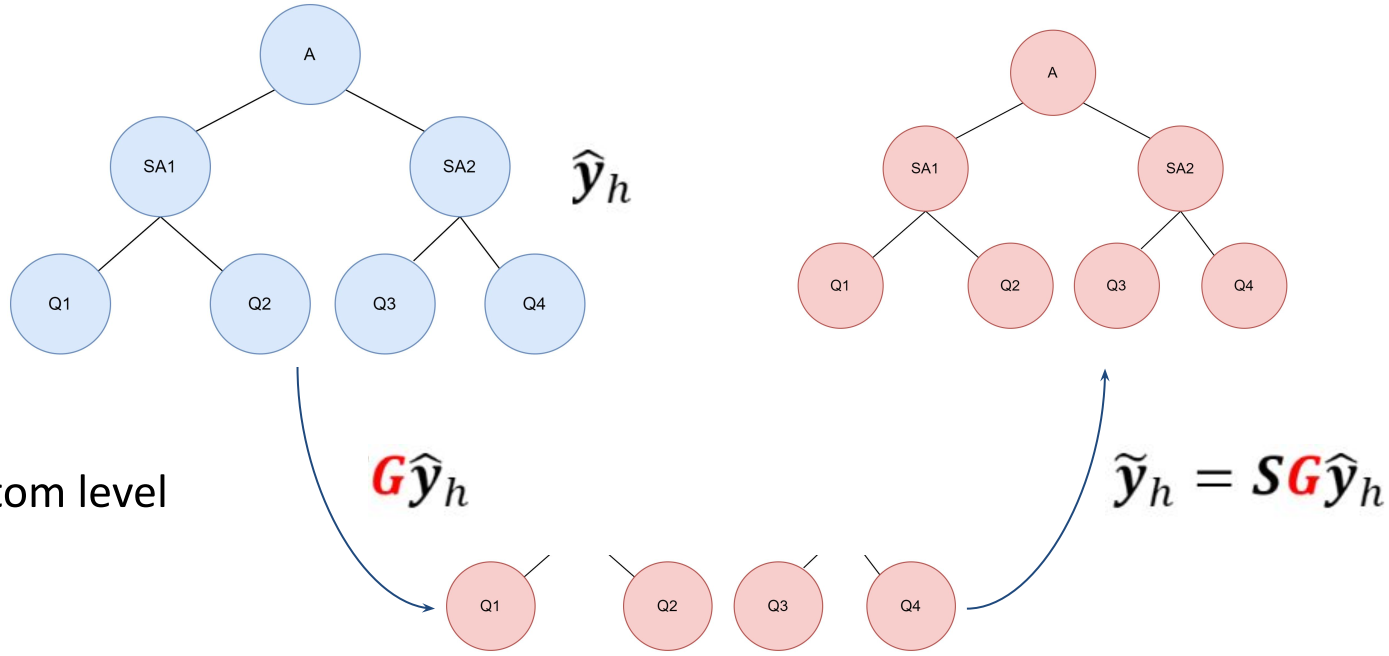
Temporal Hierarchy Neural Network (THiNN)
Training the models

Let's review THieF:

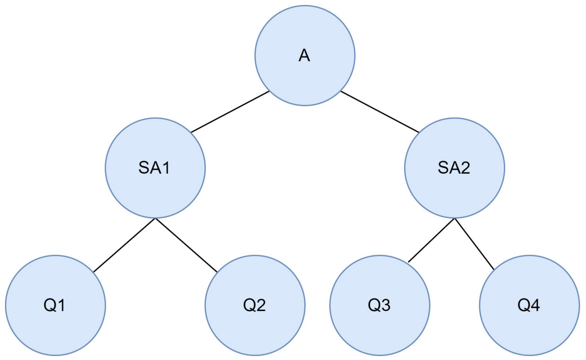
1. Get base forecasts for each level
2. Reconcile at the bottom level
3. Reconstruct the hierarchy with bottom-up

Bottom-up:

- Unbiased
- Ensures coherence
- Good hierarchical results given accurate bottom level



Temporal Hierarchy Neural Network (THiNN)
Training the models



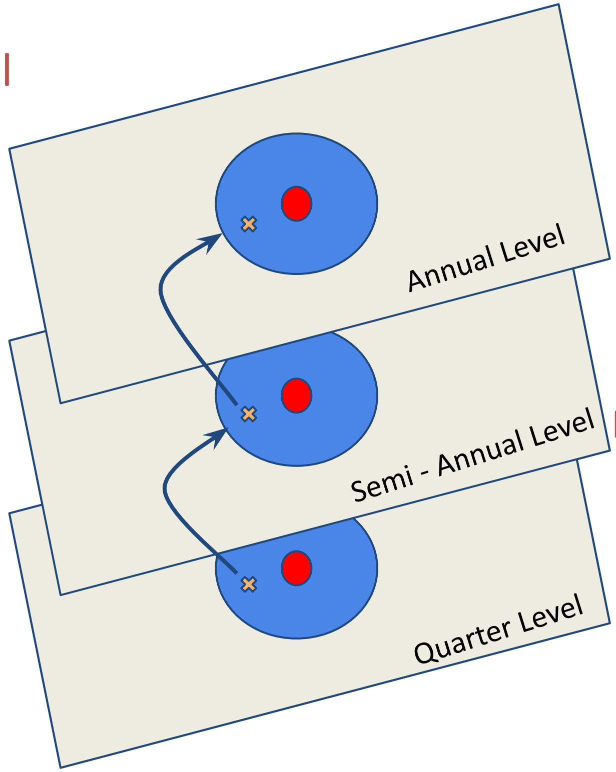
Bottom-up:

- Unbiased
- Ensures coherence
- Good hierarchical results given accurate bottom level

Basic THiNN:

- Gets a set of base forecasts
- Tunes weights to reconcile on bottom level (THieF style)
- Through coherence constraints builds the complete hierarchy

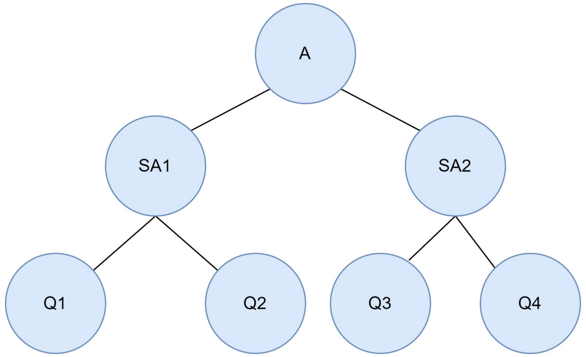
If bottom level is accurate = whole hierarchy is accurate



Properties

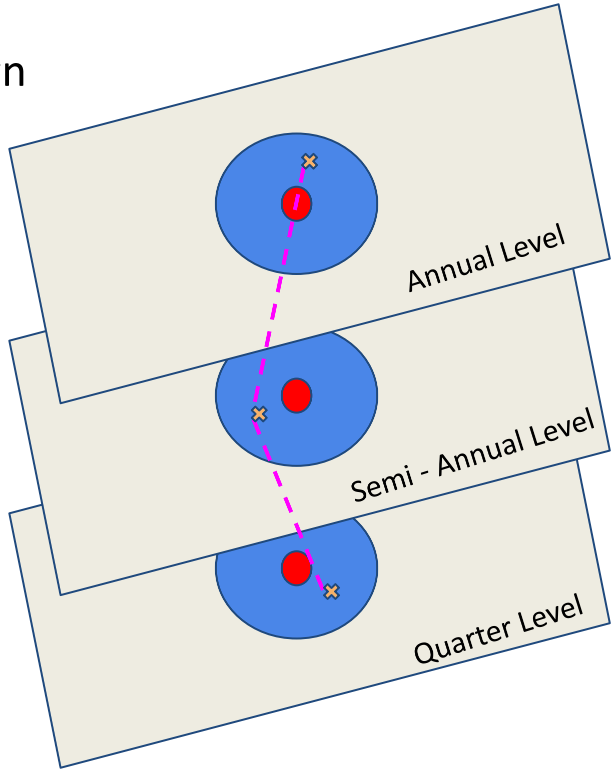
- Focus is on bottom level
- We rely for our reconciliation on bottom level (1-step)
- If it does not work we lose the reconciliation

Temporal Hierarchy Neural Network (THiNN)
Training the models



Multi - Level Reconciliation

- Instead of tuning weights to only reconcile on bottom level
- Simultaneously optimize on every level and then align
- Objectives of reconciling at each level are identical.
- Shared parameters drags all predictions

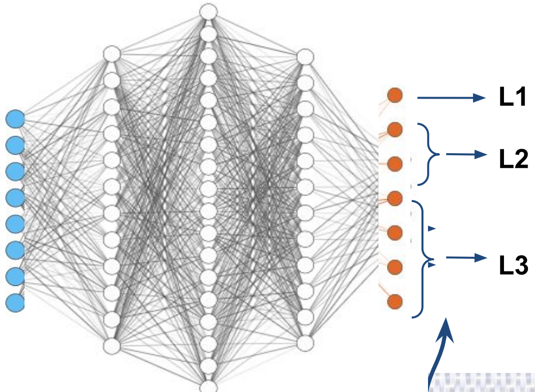


Glossary:

- Base forecasts
- Reconciled Forecast
- Shared Parameters

Reconciliation Level: Annual
Steps: i steps to the right

Multi - Task Learning

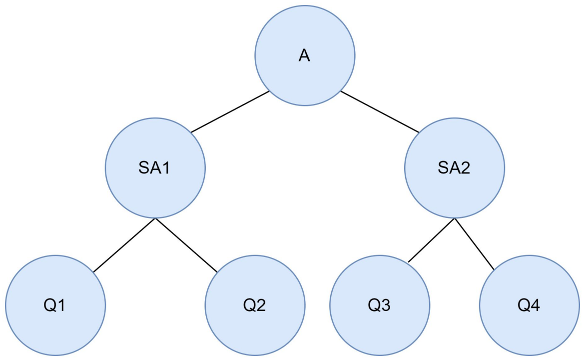


Every Level has its own loss function

Limitation:

Predictions are not coherent.

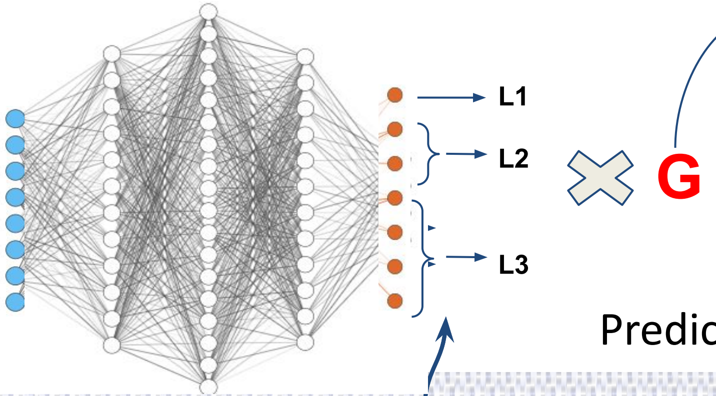
Temporal Hierarchy Neural Network (THiNN)
Training the models



Multi - Level Reconciliation

- Instead of tuning weights to only reconcile on bottom level
- Simultaneously optimize on every level and then align
- Objectives of reconciling at each level are identical.
- Shared parameters drags all predictions

Multi - Task Learning



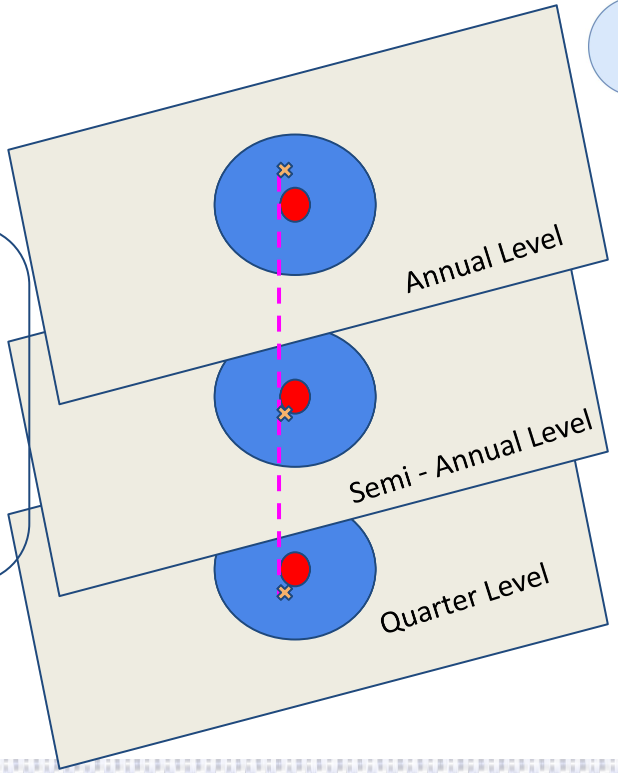
Every Level has its own loss function

G on Structural Scaling:

- Fixed
- Equal to errors on each level
- Benefits accurate base forecasts

Limitation:

Predictions are not coherent.



Glossary:

- Base forecasts
- Reconciled Forecast
- Shared Parameters

Reconciliation Level: Structural Scaling
Steps: ALIGNED

Temporal Hierarchy Neural Network (THiNN)
A Quick Experiment

Accuracy benefits from each addition

- Data: 10.000 monthly time series from M4 (random selection)
  Last 12 months kept as test set
- Accuracy Metric: Mean Absolute Scaled Error (MASE)
- Base Forecasts: ETS
- Benchmarks: Structural Scaling

Results:

- NN reconciles outperform benchmarks in general
- THiN(y,I) is not performing well
- Positional information not helpful → misused??
- In-Sample errors show benefits
- THiNN - MT and THiNN(y,S) perform the best

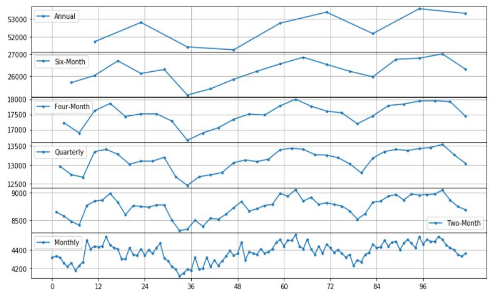
Model	Bottom Level MASE	Complete Hierarchy MASE
Base Forecasts	1.1861	0.9314
Structural Scalling	1,1772	0.91487
THIN(b,S)	1.1744	0.86714
THiN(y,I)	1.3524	0.9974
THiN - MT	1.1752	0.8599
THiN(b,S) + Pos	1.7952	1.2145
THiN(y,I) + Pos	1.6222	1.594
THiN - MT + Pos	1.6612	1.2138
THiN(b,S) + Errors	1.1651	0.8554
THiN(y,I) + Errors	1.5021	1.1102
THiN - MT + Errors	**1.1616**	**0.8334**

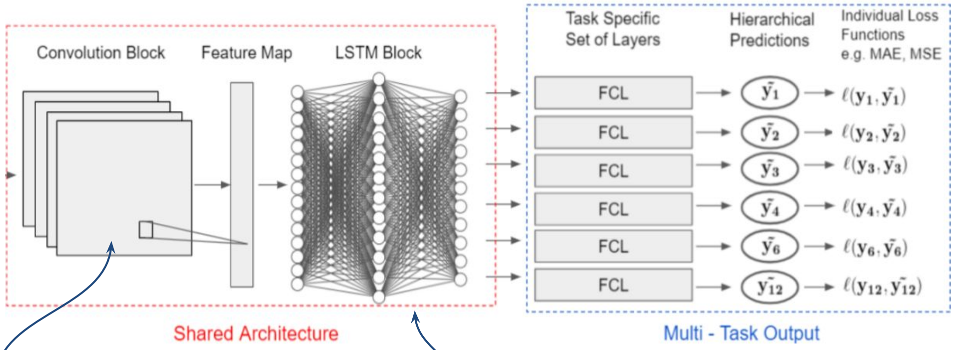
End-to-End Architectures
DeepTHieF

Deep Temporal Hierarchies Forecasting (DeepTHieF)

- End-to-end learning
  - Single-Model → Base Forecasts + Reconcilers is computationally heavy
  - Shared parameters → End-to-end architectures perform better in such occasions
- All about representation manipulation
- Novel, informative way to represent a single time series

➢ Higher capacity models with
Better understanding of time series features



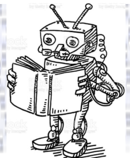


Multi - Head Output
Multi-Task module optimized each for each level and then reconciles

Input Hierarchy
Unlike THieF has a not restricted number of temporal levels

CNN Block
Extract Local Relationships among levels

LSTM Block
Models sequential part of time series



Theodosiou, Filotas & Kourentzes, Nikolaos. (2021). Deep Learning Temporal Hierarchies for Interval Forecasts. International Conference of AI in Finance. Time Series Workshop

Theodosiou, Filotas & Kourentzes, Nikolas. (2021) Forecasting with Deep Temporal Hierarchies. SSRN Scholarly Paper ID 3918315. Social Science Research

End-to-End Architectures
Another Quick Experiment

Experimental Setup:

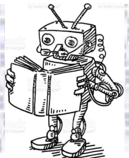
- 12.000 time series from M4
- 12-fold Cross-Validation
- Evaluate Bottom Level & Complete Hierarchy
- Metric MASE

Findings:

1. DeepTHieF outperforms all methods
2. THiNN reconcilers perform better than benchmarks
3. THiNN(y,I) performs the best among reconcilers

Table 4: Mean and median MASE results for global training

Forecast	Original frequency		Complete Hierarchy	
	Mean	Median	Mean	Median
Base				
ETS	2.6267	1.2742	0.3485	0.2095
Reconciliation & ETS				
Structural	2.6590	1.2973	0.3223	0.1898
THiNN($\hat{\boldsymbol{b}}$,$\boldsymbol{S}$)	2.6374	1.2823	0.3066	0.1791
THiNN($\hat{\boldsymbol{y}}$,$\boldsymbol{I}$)	2.6246	1.2663	0.3040	0.1766
THiNN($\hat{\boldsymbol{b}}$,$\hat{\boldsymbol{w}}$)	2.6294	1.2649	0.3091	0.1845
THiNN($\hat{\boldsymbol{y}}$,$\hat{\boldsymbol{w}}$)	2.7352	1.3361	0.3173	0.1848
DeepTHieF	**1.5479**	**1.0616**	0.2077	0.1659



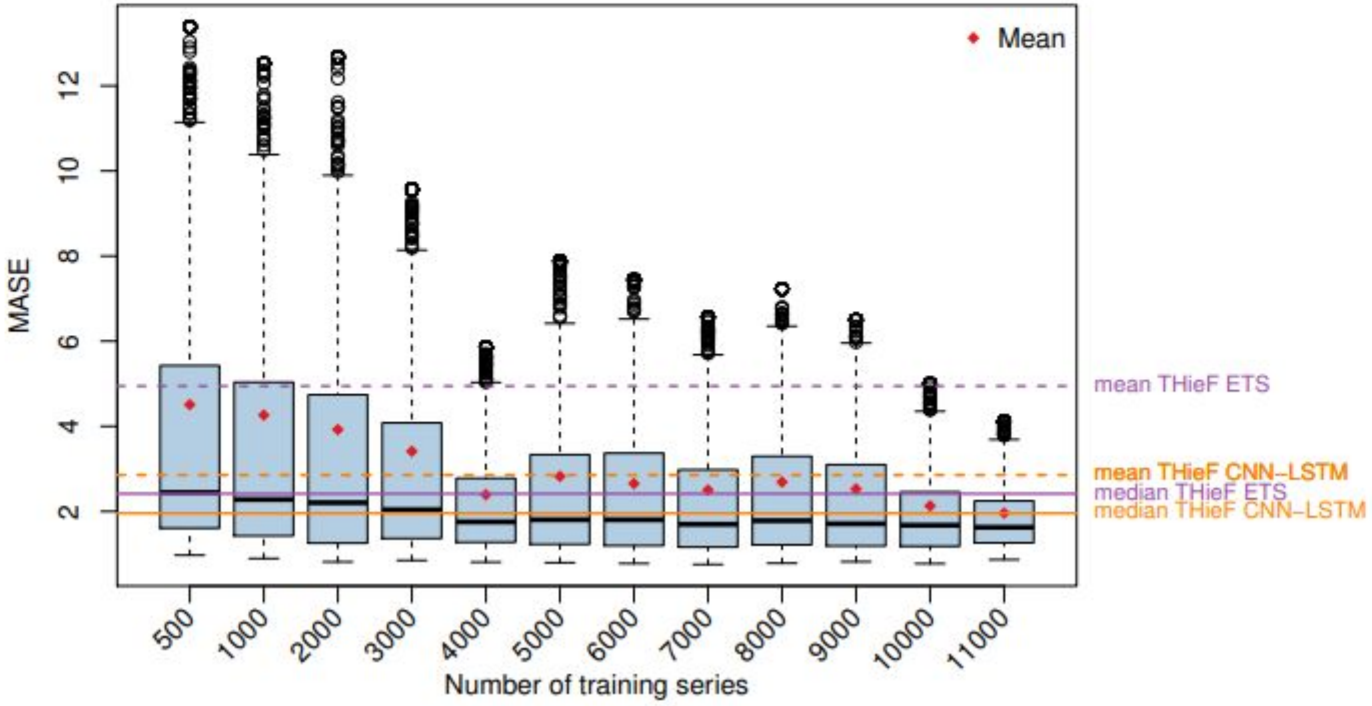
Theodosiou, Filotas & Kourentzes, Nikolaos. (2021). Deep Learning Temporal Hierarchies for Interval Forecasts. International Conference of AI in Finance. Time Series Workshop

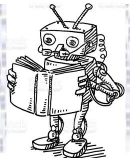
Theodosiou, Filotas & Kourentzes, Nikolas. (2021) Forecasting with Deep Temporal Hierarchies. SSRN Scholarly Paper ID 3918315. Social Science Research

End-to-End Architectures
DeepTHieF Data Requirements

Requirements for Global Training:

- Global DL Forecasting models performance increases linearly to data size
- Still DeepTHieF shows a robust performance on a reduced number of time series
- THieF has an advantage on a small number of time series.





Theodosiou, Filotas & Kourentzes, Nikolaos. (2021). Deep Learning Temporal Hierarchies for Interval Forecasts. International Conference of AI in Finance. Time Series Workshop

Theodosiou, Filotas & Kourentzes, Nikolas. (2021) Forecasting with Deep Temporal Hierarchies. SSRN Scholarly Paper ID 3918315. Social Science Research

End-to-End Architectures
Prediction Intervals (PIs)

THieF by construction is optimized for Point Forecast Reconciliation.

Approaches in Literature for Reconciling PIs

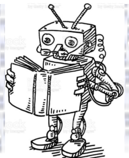
- Naive approach: Reconcile lower and upper prediction interval
- Amazon: Sample from a distribution and map to the coherent subspace
- Empirical Methods

The Flexibility of NNs

- NNs can be optimized on producing PIs
- Appropriate Loss Function
- THiNN can also output quantiles

Examples include but not limited

1. Interval score
   a. Output Upper & Lower Interval
   b. Can be combined with MAE -> similar scales
2. Probabilistic
   a. Assume a distribution
   b. Predict the parameters (eg μ, σ on Gaussian)



Theodosiou, Filotas & Kourentzes, Nikolaos. (2021). Deep Learning Temporal Hierarchies for Interval Forecasts. International Conference of AI in Finance. Time Series Workshop

Syama Sundar Rangapuram, Lucien D Werner, Konstantinos Benidis, Pedro Mercado, Jan Gasthaus, and Tim Januschowski. 2021. End-to-End Learning of Coherent Probabilistic Forecasts for Hierarchical Time Series. In Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)

Conclusions & Future of NNs and Temporal Hierarchies

- THieF's success is based on combining diverse information
- Different viewpoints produce different features
- THieF struggles when base forecasts are very inaccurate.

We proposed THiNN -> A Data-Driven Non-Linear Reconciler based on THieF

- Positional Information
- Base forecasts accuracy
- Multi-task Learning

DeepTHieF & End-to-End Architectures

- Novel Input representation based on Temporal Hierarchies
- Single NN does not work at each task (not in forecasting)
- Change how global NNs view time series

MEER INFO

Research Associate
Filotas, Theodosiou
filotas.theodosiou@vives.be
+306975721445

Campus Brugge
Xaverianenstraat
Xaverianenstraat 12
050 30 51 00

Campus Kortrijk
Doorniksesteenweg 145
056 26 41 60

facebook.com/viveshogeschool

#veryvives