Using Temporal Hierarchies in Practice

International Symposium on Forecasting
22nd June 2015

George Athanasopoulos
Monash University

Nikolaos Kourentzes
Lancaster University

Rob J Hyndman
Monash University

Fotios Petropoulos
Cardiff University

n.kourentzes@lancaster.ac.uk

Lancaster Centre for Forecasting
www.forecasting-centre.com

Outline

1. Temporal Hierarchies
2. Do they work? Empirical results on M3
3. Significance for practice – Predicting A&E
4. Cross sectional and temporal hierarchies

Hierarchical Forecasting

In many forecasting applications there are multiple time series that are hierarchically organised:

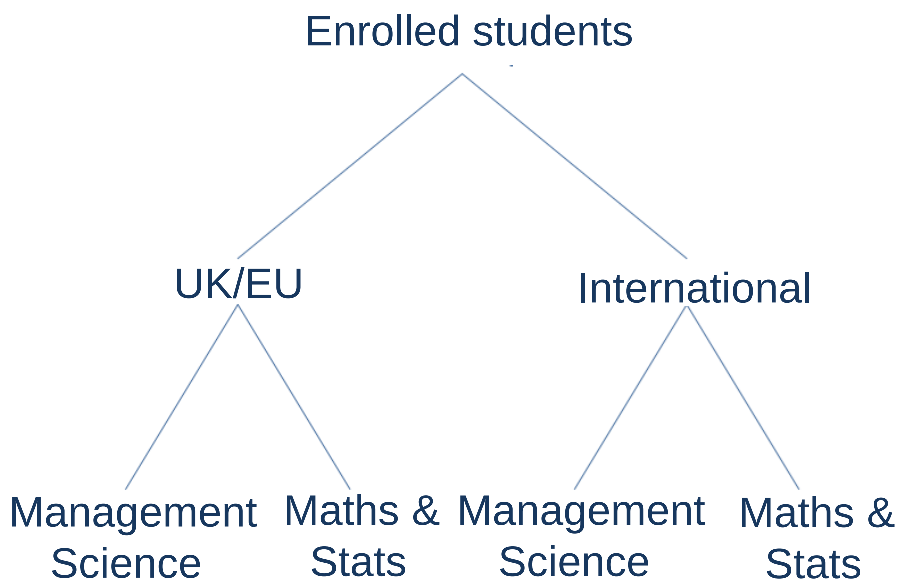
- Product categories
- Market segments
- Geographical
- Supply channels
- etc.

There are two major advantages in considering hierarchical structure in forecasting:

- Improve accuracy
- Reconcile forecasts → Decision making advantages

Hierarchical Forecasting

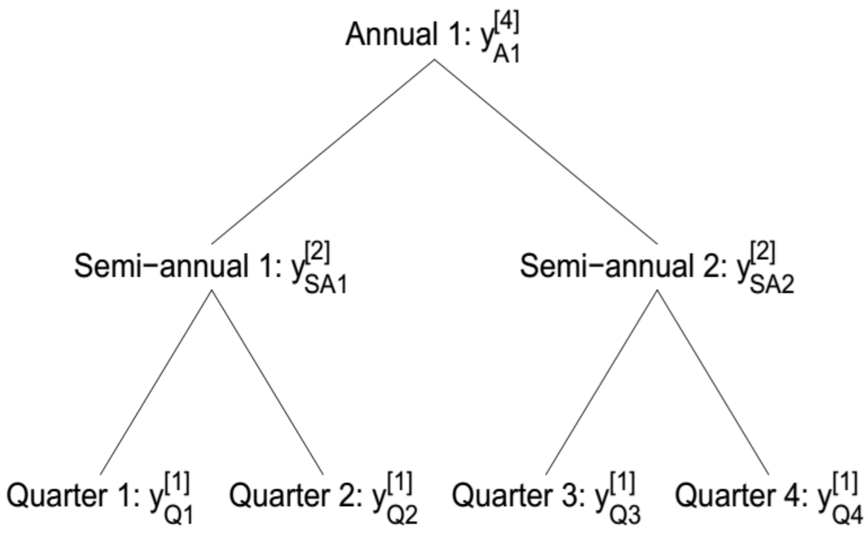
A typical example of hierarchical forecasting comes from cross-sectional applications. For example let us suppose we want to forecast student admissions:



Reconciling forecasts across levels helps improve accuracy and align decision making.

Forecasting with Temporal Hierarchies

Analogously we can construct temporal hierarchies, where now we consider for **a single time series** multiple levels of **temporal aggregation**.



The idea is that similarly to cross-sectional hierarchies, we can take advantage of the structure to **increase accuracy** at short and long term and **align forecasts of different horizons**.

Forecasting with Temporal Hierarchies

As George explained in his presentation ‘Forecasting with Temporal Hierarchies’:

$$\boldsymbol{y}_i = \left(y_i^{[m]}, \ldots, \boldsymbol{y}_i^{[k_3]\prime}, \boldsymbol{y}_i^{[k_2]\prime}, \boldsymbol{y}_i^{[1]\prime}\right)'$$

Temporally aggregated series

$$\boldsymbol{y}_i = \boldsymbol{S}\boldsymbol{y}_i^{[1]}$$

‘Summing` matrix describing the hierarchy

and for the forecasts:

Scaling matrix

Reconciled forecasts

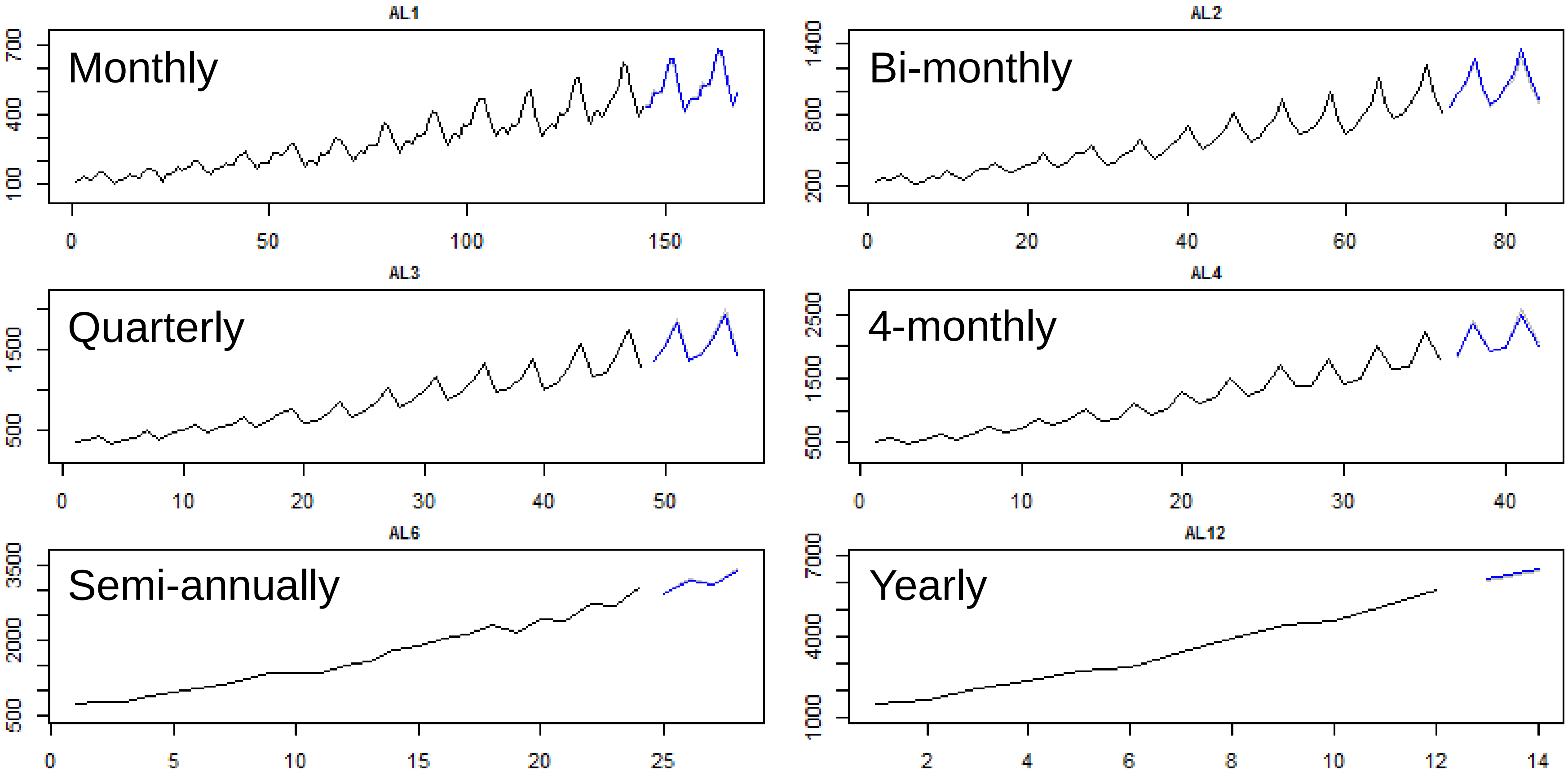
$$\tilde{\boldsymbol{y}}_h^* = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Lambda}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{y}}_h.$$

Forecasts at each level (produced by any means)

The scaling matrix Λ can be approximated using:

- Temporal hierarchy structure
- Variance of forecasts at each temporal level

An example



The forecast (blue line) is the result of reconciling the forecasts across all temporal aggregation levels. Here ETS was used, but any model can be used and it does not have to be the same across levels.

Evaluation on the M3 dataset

Monthly dataset

% change over base (MAPE rows) · relative error over base (GRMSE rows)

Aggregation level	h	ETS Base	ETS BU	ETS WLS_s	ETS WLS_v	ARIMA Base	ARIMA BU	ARIMA WLS_s	ARIMA WLS_v
MAPE									
Annual	1	9.4	-7.7	**-17.2**	-14.7	13.6	-30.0	**-38.3**	-36.9
Semi-annual	3	11.4	2.4	**-7.9**	-5.1	12.8	-1.5	**-13.2**	-9.7
Four-monthly	4	11.5	3.6	**-5.0**	-3.0	12.8	-0.1	**-9.0**	-6.8
Quarterly	6	12.7	2.4	**-6.3**	-3.9	14.3	-2.0	**-11.8**	-9.0
Bi-monthly	9	14.2	3.2	**-4.6**	-2.5	16.1	-3.2	**-11.7**	-9.3
Monthly	18	20.0	0.0	**-1.4**	-0.6	22.1	0.0	**-12.6**	-10.1
Average			*0.6*	***-7.1***	*-5.0*		*-6.2*	***-16.1***	*-13.6*
GRMSE									
Annual	1	1.00	0.81	**0.75**	0.78	1.00	0.76	**0.69**	0.71
Semi-annual	3	1.00	1.01	**0.95**	0.97	1.00	0.98	**0.90**	0.93
Four-monthly	4	1.00	1.02	**0.97**	0.99	1.00	0.99	**0.93**	0.95
Quarterly	6	1.00	1.02	**0.97**	0.98	1.00	0.97	**0.91**	0.93
Bi-monthly	9	1.00	1.01	**0.96**	0.97	1.00	0.98	**0.93**	0.95
Monthly	18	1.00	1.00	**0.96**	0.97	1.00	1.00	**0.95**	0.97
Average			*0.98*	***0.93***	*0.94*		*0.95*	***0.89***	*0.91*

BU: Bottom-Up
WLS_s: Structural scaling
WLS_v: Variance scaling

To compare with M3 results on sMAPE the results are:

WLS_s: ETS: 13.69%

WLS_v: ETS: 13.87%

ETS: 14.45% (Hyndman et al., 2002)

Theta: 13.85% (best original performance)

Evaluation on the M3 dataset

Quarterly dataset

Aggregation level	h	ETS Base	ETS BU (% change over base)	ETS WLS_s	ETS WLS_v	ARIMA Base	ARIMA BU (% change over base)	ARIMA WLS_s	ARIMA WLS_v
MAPE									
Annual	2	7.8	-26.1	-26.0	**-26.6**	7.8	-21.6	-20.7	**-25.0**
Semi-annual	4	6.8	-7.4	-6.7	**-7.8**	6.9	-2.9	-1.5	**-6.1**
Quarterly	8	7.2	0.0	2.0	**-0.1**	7.6	0.0	1.7	**-2.5**
Average			*-11.2*	*-10.2*	***-11.5***		*-8.2*	*-6.9*	***-11.2***
GRMSE (relative error over base)									
Annual	2	1.00	0.67	0.71	**0.68**	1.00	0.76	0.79	**0.75**
Semi-annual	4	1.00	**0.93**	0.96	**0.93**	1.00	0.94	0.96	**0.93**
Quarterly	8	**1.00**	**1.00**	1.05	**1.00**	1.00	1.00	1.03	**0.99**
Average			***0.87***	*0.91*	***0.87***		*0.90*	*0.93*	***0.89***

BU: Bottom-Up
WLS_s: Structural scaling
WLS_v: Variance scaling

To compare with M3 results on sMAPE the results are:

WLS_s: ETS: 9.93%

WLS_v: ETS: 9.80%

ETS: 9.94% (Hyndman et al., 2002)

Theta: 8.96% (best original performance)

An application: Predicting A&E admissions

A major challenge in running A&E wards is predicting the staffing requirements for:

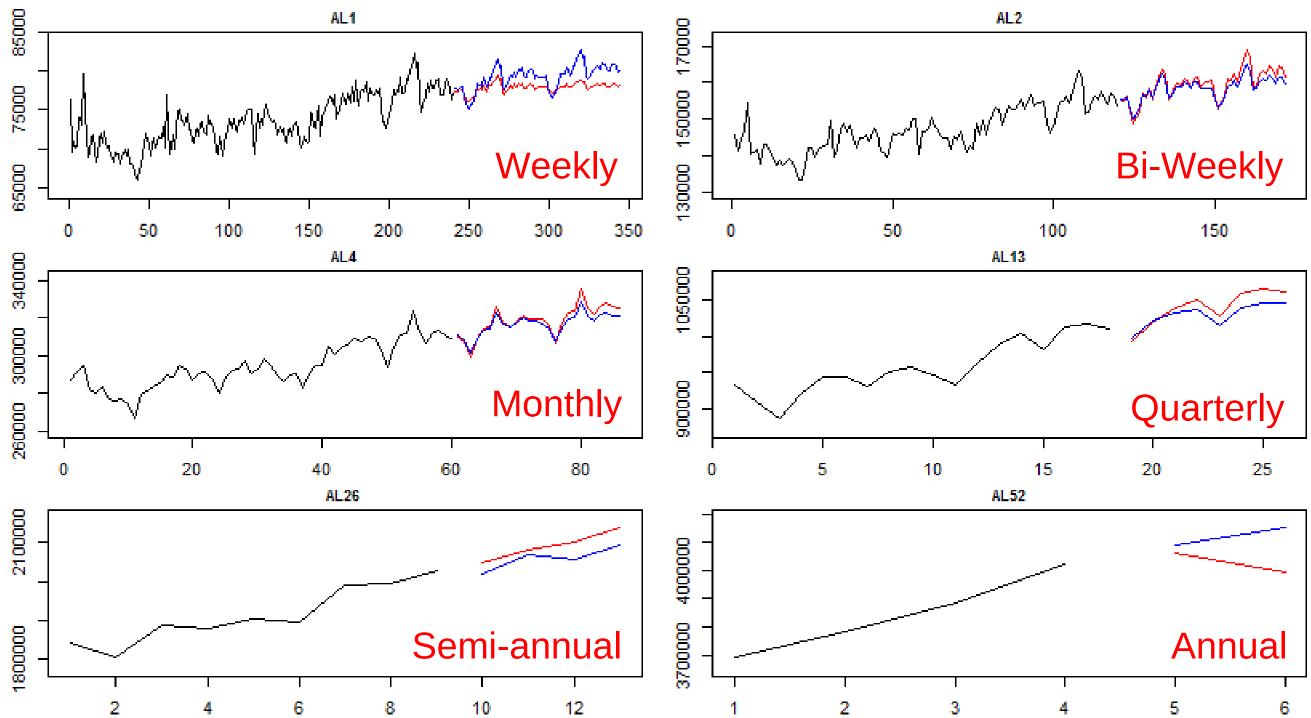
- short term: to ensure adequate response to emergencies
- long term: to ensure appropriate staffing levels, training, etc.

Decision makers need accurate short term forecasts (weekly) as well as annual (and longer) forecasts to build on capacity and capabilities at the A&E wards, as well as for budgeting.

Collect weekly data for UK A&E wards.
13 time series: covering different types of emergencies and different severities (measured as time to treatment)
Span from week 45 2010 (7th Nov 2010) to week 24 2015 (7th June 2015)
Series are at England level (not local authorities).

An application: Predicting A&E admissions

Total Emergency Admissions via A&E



Red is the prediction of the base model (ARIMA)
Blue is the temporal hierarchy reconciled forecasts (based on ARIMA)

Observe how information is `borrowed' between temporal levels. Base models for instance provide very poor weekly and annual forecasts

An application: Predicting A&E admissions

Objective: Accurately predict to support staffing and training decisions. Note that aligning the short and long term forecasts is important for consistency of planning and budgeting.

Test set: 52 weeks.
Rolling origin evaluation.
Forecast horizons of interest: t+1, t+4, t+52 (1 week, 1 month, 1 year).
Evaluation GMRAE (relative to base model)
As a base model auto.arima (forecast package R) is used.

Hierarchical level	Horizon	GMRAE
Annual level	t+1	0.590
Weekly level	t+1	1.031
Weekly level	t+4	0.894
Weekly level	t+1 - t+52	0.840

Where do the improvements come from?

- Temporally hierarchical view of the data & reconciliation
- `Long-term' view of data
- Base model uncertainty

Decision making and hierarchical forecasting

Hierarchical (or grouped) forecasting can improve accuracy, but their true strength lies in the reconciliation of the forecasts → aligning forecasts is crucial for decision making.

Is the reconciliation achieved useful for decision making?

Cross-sectional

- Reconcile across different items.
- Units may change at different levels of hierarchy.
- Suppose an electricity demand hierarchy: lower and higher levels have same units. All levels relevant for decision making.
- Suppose a supply chain hierarchy. Weekly sales of SKU are useful. Weekly sales of organisation are not! Needed at different time scale.

Temporal

- Reconcile across time units/horizons.
- Units of items do not change.
- Consider our application. NHS admissions short and long term are useful for decision making.
- Suppose a supply chain hierarchy. Weekly sales of SKU is useful for operations. Yearly sales of a single SKU may be useful, but often not!
- Operational → Tactical → Strategic forecasts.

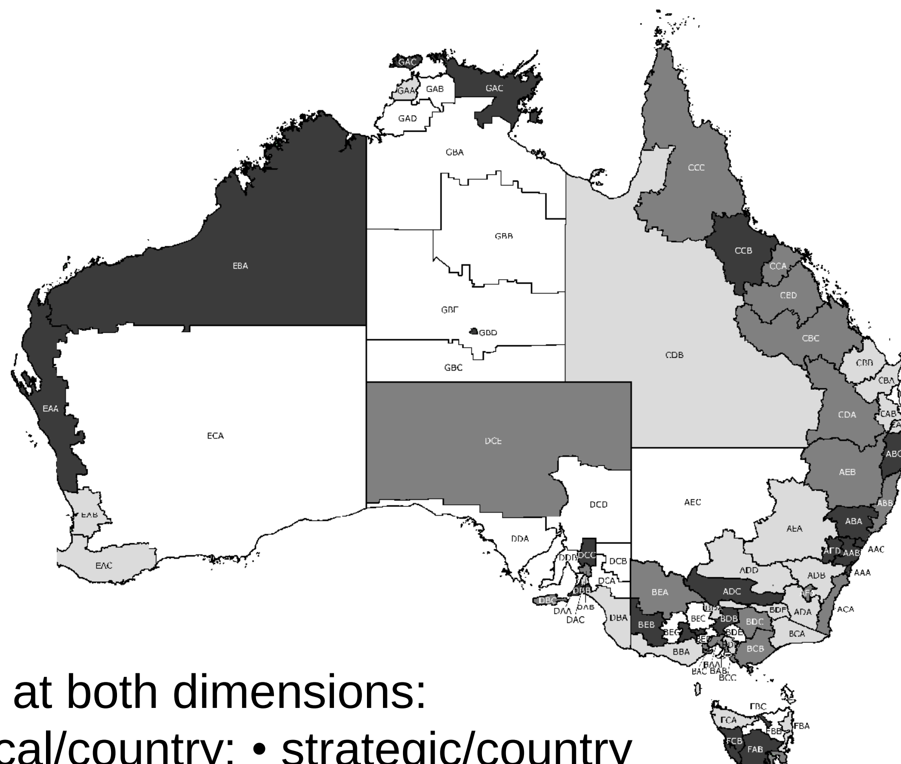
Cross-temporal hierarchies

Temporal hierarchies permit aligning operational, tactical and strategic planning, while offering accuracy gains → useful for decision making

BUT there can be cases that strategic level forecasts are not required for each item, but at an aggregate level.

Let us consider tourism demand for Australia as an example. Local authorities can make use of detailed forecasts (temporal/spatial) but at a country level weekly forecasts are of limited use.

- Temporal: tactical → strategic
- Cross-sectional: local → country



Cross temporal can support decisions at both dimensions:
• Tactical/local; • strategic/local; • tactical/country; • strategic/country

Cross-temporal hierarchies: Tourism demand

56 (bottom level) quarterly tourism demand series → Athanasopoulos et al. (2011)

- 6 years in-sample
- 3 years out-of-sample horizon: up to 2 years
- rolling origin evaluation

Cross-temporal hierarchical forecasts:

- Most accurate
- Most complete reconciliation (one number forecast)
- Flexible decision making support

MAPE %

Level	No. of series	ETS	Theta
		Base forecasts per series	
Overall	89	32.26	28.74
Top	1	5.61	5.96
Level 1	4	9.08	9.05
Level 2	28	28.39	24.68
Bottom	56	36.32	32.58
		Temporally reconciled	
Overall	89	30.46	28.19
Top	1	5.75	6.13
Level 1	4	9.29	9.04
Level 2	28	27.18	24.21
Bottom	56	34.06	31.95
		Cross-temporally reconciled	
Overall	89	30.26	**28.04**
Top	1	6.02	**5.88**
Level 1	4	9.11	**8.70**
Level 2	28	25.91	**23.87**
Bottom	56	34.39	**31.90**

Conclusions

- Temporal hierarchies provide a new class of hierarchical forecasts that can be produced for any time series.
- Applicable to forecasts produced by any means → theoretically elegant hierarchical combination of forecasts.
- Joins operational, tactical and strategic decision making by reconciling forecasts → satisfies a business need that has remained unmet
- Potential to increase forecasting accuracy and mitigate modelling uncertainty
- Combining cross-sectional and temporal hierarchies: forecasts reconciled across conventional hierarchy and forecast horizons → one number forecast → superior decision making.

Thank you for your attention!
Questions?

Nikolaos Kourentzes
Lancaster University Management School
Lancaster Centre for Forecasting - Lancaster, LA1 4YX
email: n.kourentzes@lancaster.ac.uk
Forecasting blog: http://nikolaos.kourentzes.com

www.forecasting-centre.com/

Lancaster Centre for Forecasting
www.forecasting-centre.com