MONASH University

MONASH BUSINESS SCHOOL

High-dimensional time series analysis

Rob J Hyndman

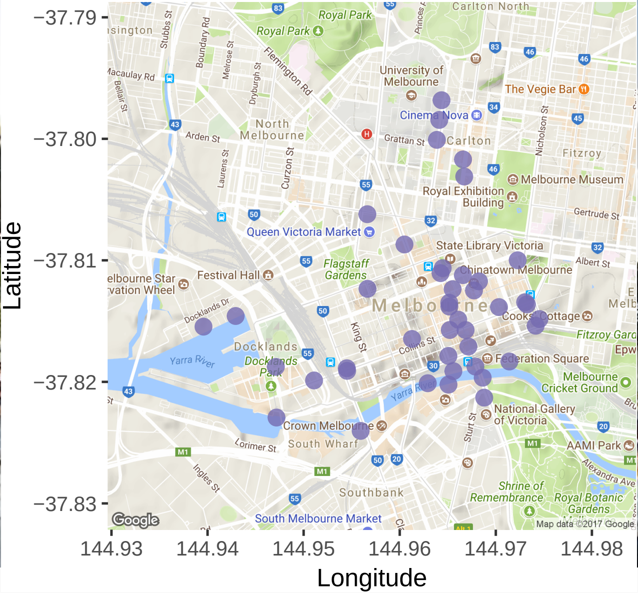
1 November 2017

ACEMS 2017 RETREAT

Outline

Pedestrian counts

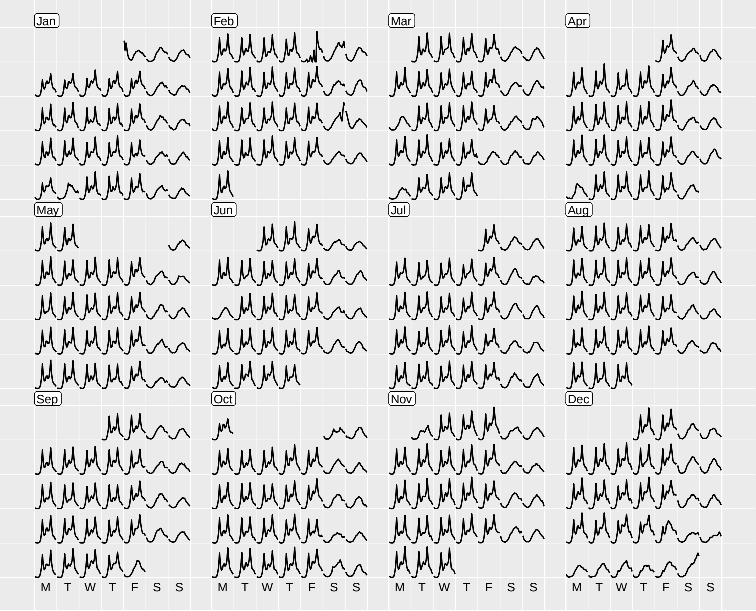
Pedestrian counts



Sub-daily time series analysis

- How to visualize many series of sub-daily data over several years?
- How to identify unusual patterns/incidents?
- How to forecast sub-daily data taking account of public holidays and special events?

Sub-daily time series analysis



data

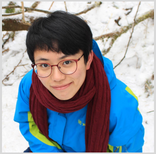
s?

unt

Sub-daily time series analysis

- How to visualize many series of sub-daily data over several years?
- How to identify unusual patterns/incidents?
- How to forecast sub-daily data taking account of public holidays and special events?

Di Cook



Earo Wang

Mitchell O’Hara-Wild

Outline

Walmart weekly sales data

Time series feature analysis

- Can we use time series features for fast identification of forecasting models?
- How to generate new time series with specified feature vectors?
- What can we say about the feature space of time series?

Time series feature analysis

- Can we use time series features for fast identification of forecasting models?
- How to generate new time series with specified feature vectors?
- What can we say about the feature space of time series?

Kate Smith-Miles

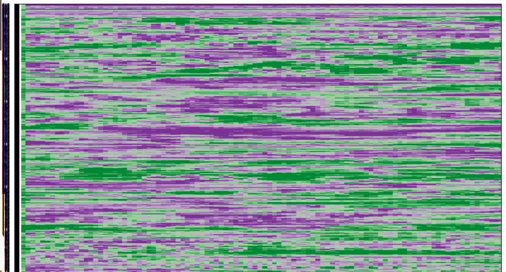
George Athanasopoulos

Thiyanga Talagala

Outline

Security monitoring

Security monitoring



Time series anomaly detection

- How to identify anomalous behaviour within streaming data?
- How to define an anomaly in a large multivariate data set?

Time series anomaly detection

- How to identify anomalous behaviour within streaming data?
- How to define an anomaly in a large multivariate data set?

Kate Smith-Miles

Mario Andrés Muñoz Acosta

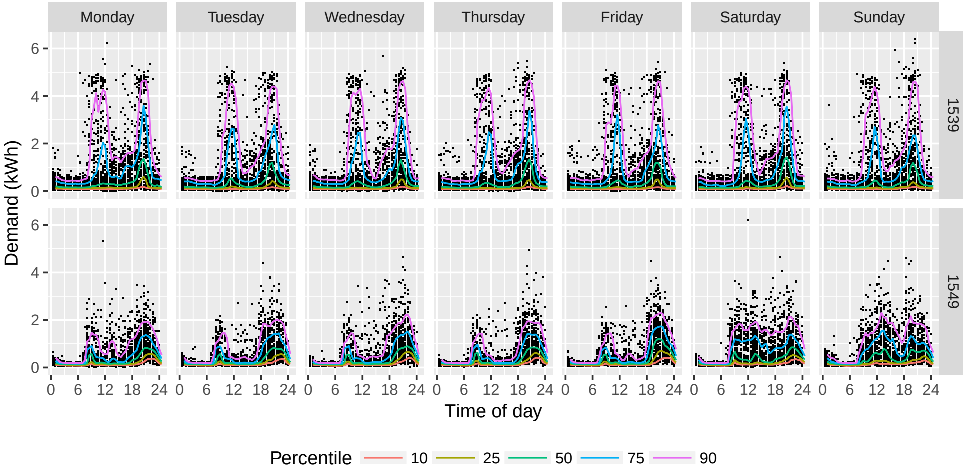
Sevvandi Kandanaarachchi

Dilini Talagala

Outline

Electricity demand

Electricity demand



Electricity demand

- How to forecast future demand by household?
- How to reconcile household demand forecasts with state and national demand forecasts?
- How to identify unusual demand patterns?
- How to measure forecast accuracy when forecasts are probability distributions within a hierarchy?

Electricity demand

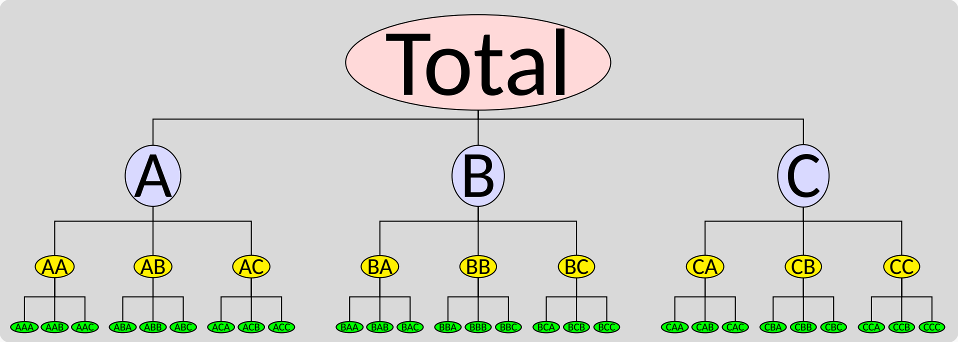
- How to forecast future demand by household?
- How to reconcile household demand forecasts with state and national demand forecasts?
- How to identify unusual demand patterns?
- How to measure forecast accuracy when forecasts are probability distributions within a hierarchy?

Souhaib Ben Taieb

Cameron Roach

Outline

Forecast reconciliation



- Huawei sales by division, group, sub-group, etc.
- Australian tourism demand by state, region, zone.

Forecast reconciliation

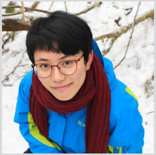
- Forecasts at all nodes must be coherent
- Bottom level typically has thousands or millions of time series
- How to define coherence probabilistically?
- How to visualize so many time series?

George Athanasopoulos

Anastasios Panagiotelis

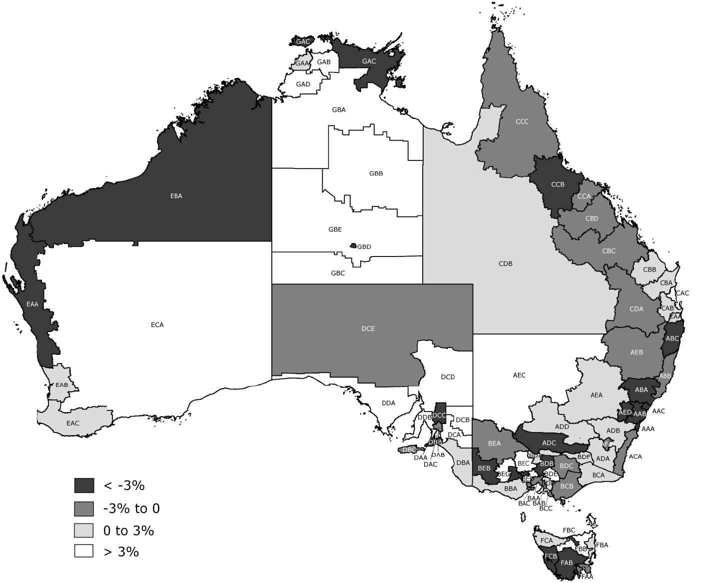
Shanika Wickramasuriya

Puwasala Gamakumara



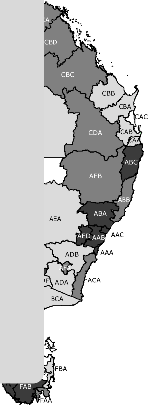
Earo Wang

Australian tourism demand

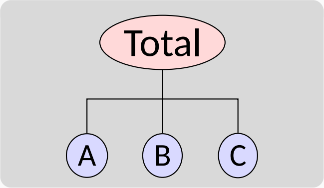


Australian tourism demand

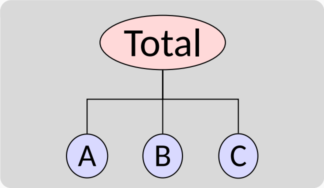
- Quarterly data on visitor night from 1998:Q1 – 2013:Q4
- From *National Visitor Survey*, based on annual interviews of 120,000 Australians aged 15+, collected by Tourism Research Australia.
- Split by 7 states, 27 zones and 76 regions (a geographical hierarchy)
- Also split by purpose of travel
  - Holiday
  - Visiting friends and relatives (VFR)
  - Business
  - Other
- 304 bottom-level series



Hierarchical time series



Hierarchical time series

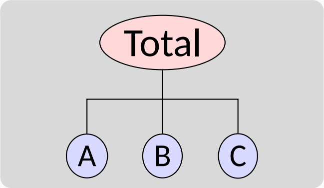


y_t : observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t.

$\boldsymbol{b}_t$: vector of all series at bottom level in time t.

Hierarchical time series



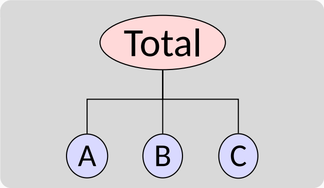
y_t : observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t.

$\boldsymbol{b}_t$: vector of all series at bottom level in time t.

$$\boldsymbol{y}_t = [y_t, y_{A,t}, y_{B,t}, y_{C,t}]' = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}$$

Hierarchical time series



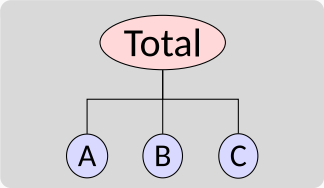
y_t : observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t.

$\boldsymbol{b}_t$: vector of all series at bottom level in time t.

$$\boldsymbol{y}_t = [y_t, y_{A,t}, y_{B,t}, y_{C,t}]' = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}$$

Hierarchical time series



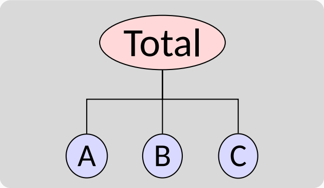
y_t : observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t.

$\boldsymbol{b}_t$: vector of all series at bottom level in time t.

$$\boldsymbol{y}_t = [y_t, y_{A,t}, y_{B,t}, y_{C,t}]' = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}}_{\boldsymbol{b}_t}$$

Hierarchical time series



y_t : observed aggregate of all series at time t.

$y_{X,t}$: observation on series X at time t.

$\boldsymbol{b}_t$: vector of all series at bottom level in time t.

$$\boldsymbol{y}_t = [y_t, y_{A,t}, y_{B,t}, y_{C,t}]' = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}}_{\boldsymbol{b}_t}$$

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$$

Disaggregated time series

Every collection of time series with aggregation constraints can be written as

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$$

where

- $\boldsymbol{y}_t$ is a vector of all series at time t
- $\boldsymbol{b}_t$ is a vector of the most disaggregated series at time t
- $\boldsymbol{S}$ is a "summing matrix" containing the aggregation constraints.

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$.

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$. (In general, they will not "add up".)

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$. (In general, they will not "add up".)

Reconciled forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

for some matrix $\boldsymbol{P}$.

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$. (In general, they will not "add up".)

Reconciled forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ projects base forecasts $\hat{\boldsymbol{y}}_n(h)$ to bottom level.

Forecasting notation

Let $\hat{\boldsymbol{y}}_n(h)$ be vector of initial h-step forecasts, made at time n, stacked in same order as $\boldsymbol{y}_t$. (In general, they will not “add up”.)

Reconciled forecasts must be of the form:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

for some matrix $\boldsymbol{P}$.

- $\boldsymbol{P}$ projects base forecasts $\hat{\boldsymbol{y}}_n(h)$ to bottom level.
- $\boldsymbol{S}$ adds them up

General properties: bias and variance

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

General properties: bias and variance

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Bias

Reconciled forecasts are unbiased iff $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$.

General properties: bias and variance

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Bias

Reconciled forecasts are unbiased iff $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$.

Variance

Let error variance of h-step base forecasts $\hat{\boldsymbol{y}}_n(h)$ be

$$\boldsymbol{W}_h = \text{Var}[\boldsymbol{y}_{n+h} - \hat{\boldsymbol{y}}_n(h) \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_n]$$

Then error variance of the reconciled forecasts is

$$\text{Var}[\boldsymbol{y}_{n+h} - \tilde{\boldsymbol{y}}_n(h) \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_n] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}'$$

Optimal forecast reconciliation

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Optimal forecast reconciliation

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Theorem: MinT Reconciliation

If $\boldsymbol{P}$ satisfies $\boldsymbol{SPS} = \boldsymbol{S}$, then

$$\min_{\boldsymbol{P}} = \text{trace}[\boldsymbol{SPW}_h\boldsymbol{P}'\boldsymbol{S}']$$

has solution $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$.

Optimal forecast reconciliation

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h)$$

Theorem: MinT Reconciliation

If $\boldsymbol{P}$ satisfies $\boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S}$, then

$$\min_{\boldsymbol{P}} = \text{trace}[\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}']$$

has solution $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$.

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_n(h)$$

Reconciled forecasts

Base forecasts

Optimal forecast reconciliation

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_n(h)$$

Reconciled forecasts

Base forecasts

Optimal forecast reconciliation

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_n(h)$$

Reconciled forecasts

Base forecasts

Assume that $\boldsymbol{W}_h = k_h \boldsymbol{W}_1$ to simplify computations.

Optimal forecast reconciliation

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_n(h)$$

Reconciled forecasts

Base forecasts

Assume that $\boldsymbol{W}_h = k_h\boldsymbol{W}_1$ to simplify computations.

WLS solution

- Approximate $\boldsymbol{W}_1$ by its diagonal.

Optimal forecast reconciliation

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_n(h)$$

Reconciled forecasts

Base forecasts

Assume that $\boldsymbol{W}_h = k_h\boldsymbol{W}_1$ to simplify computations.

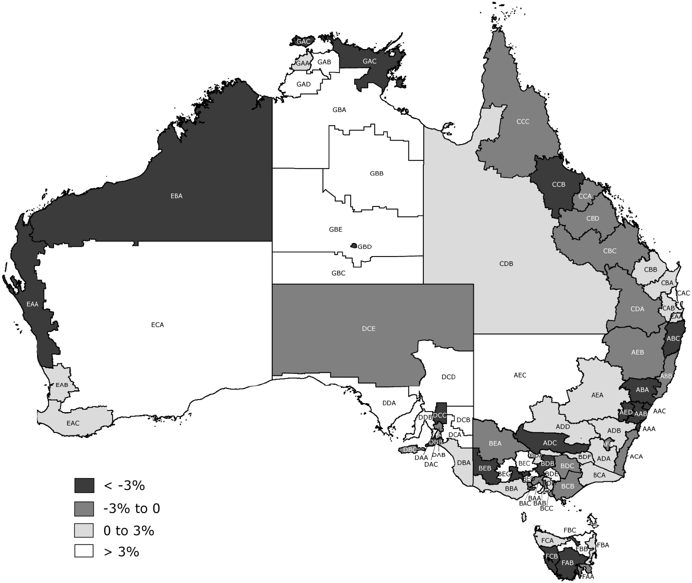
WLS solution

- Approximate $\boldsymbol{W}_1$ by its diagonal.

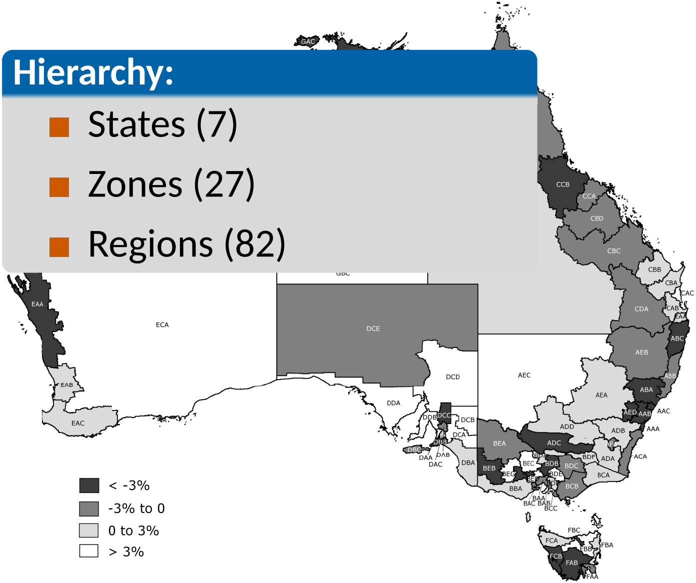
GLS solution

- Estimate $\boldsymbol{W}_1$ using shrinkage to the diagonal.

Australian tourism



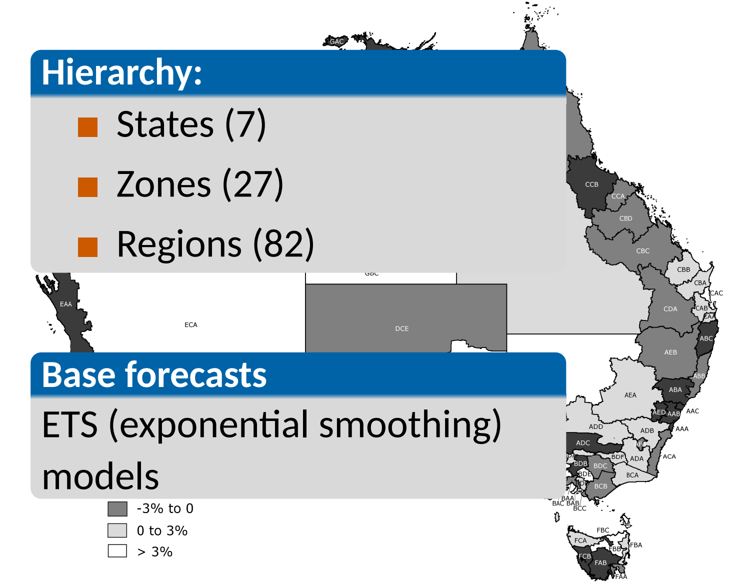
Australian tourism



Hierarchy:

- States (7)
- Zones (27)
- Regions (82)

Australian tourism



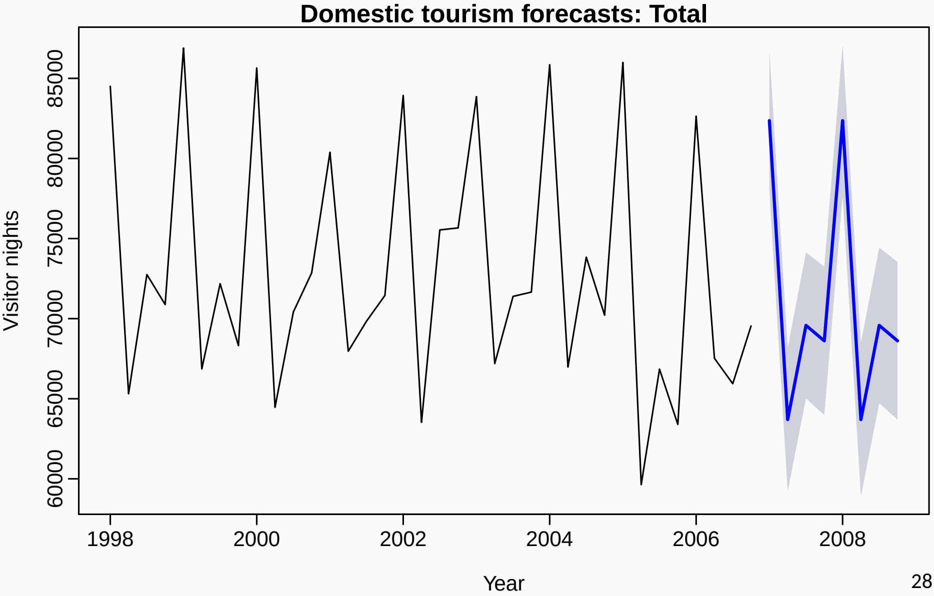
Hierarchy:

- States (7)
- Zones (27)
- Regions (82)

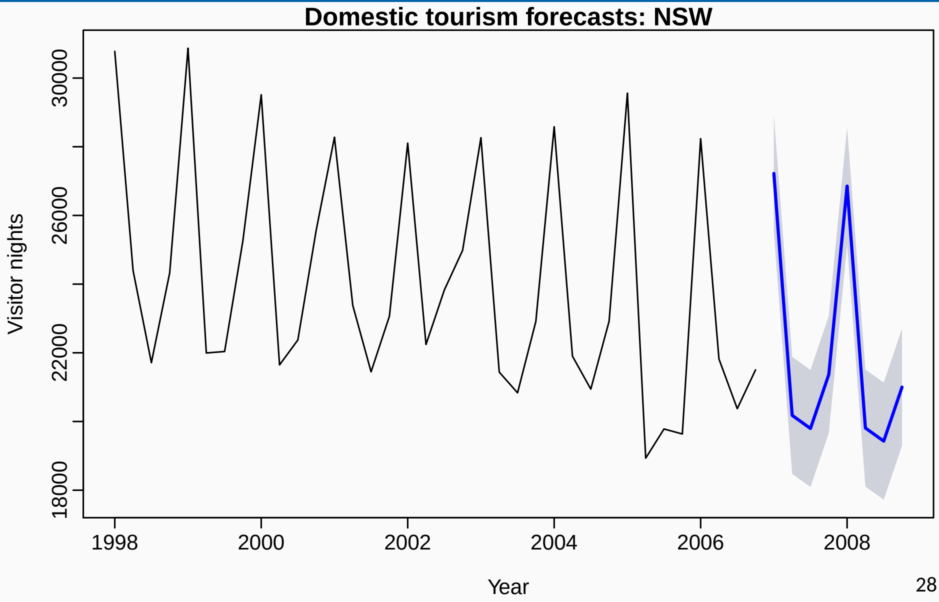
Base forecasts

ETS (exponential smoothing) models

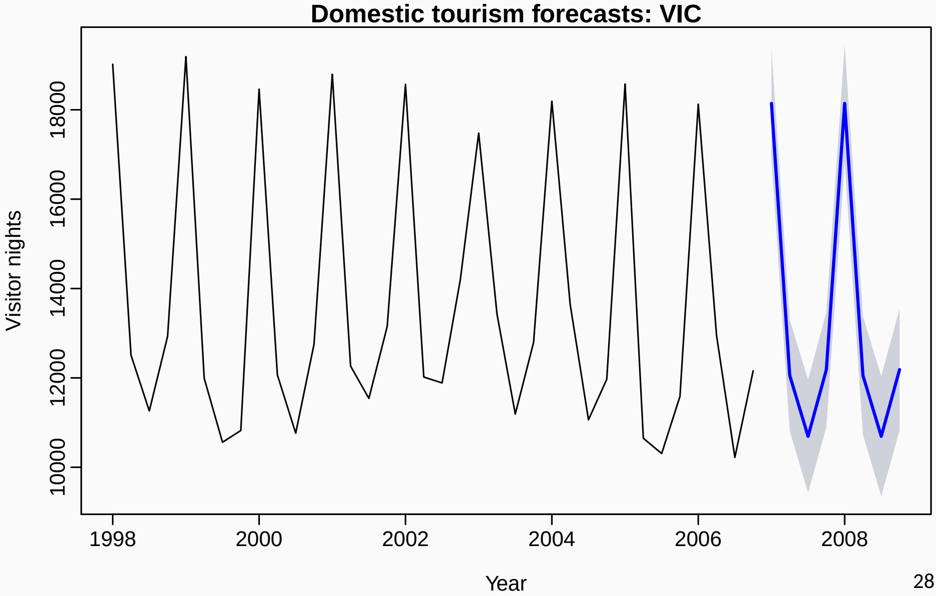
Base forecasts



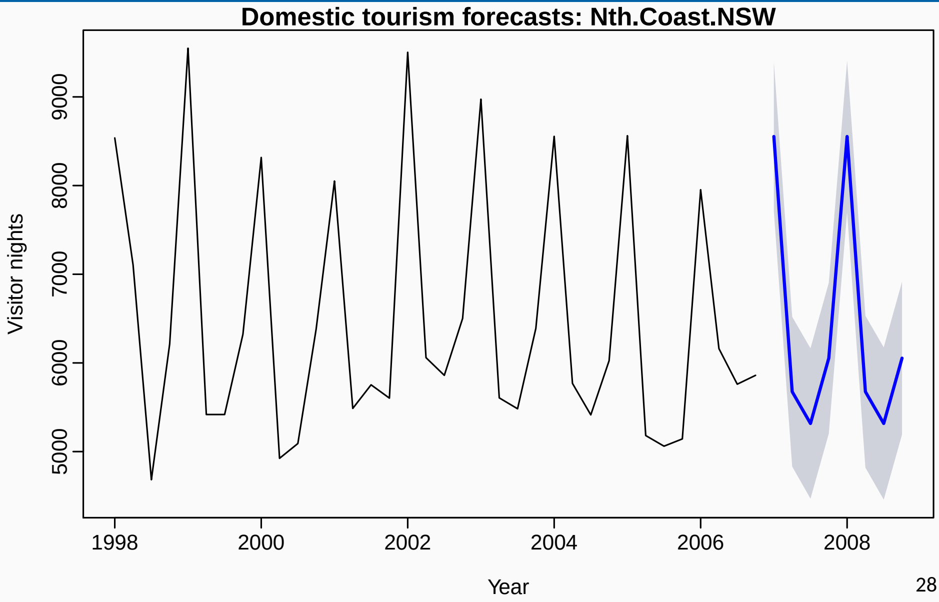
Base forecasts



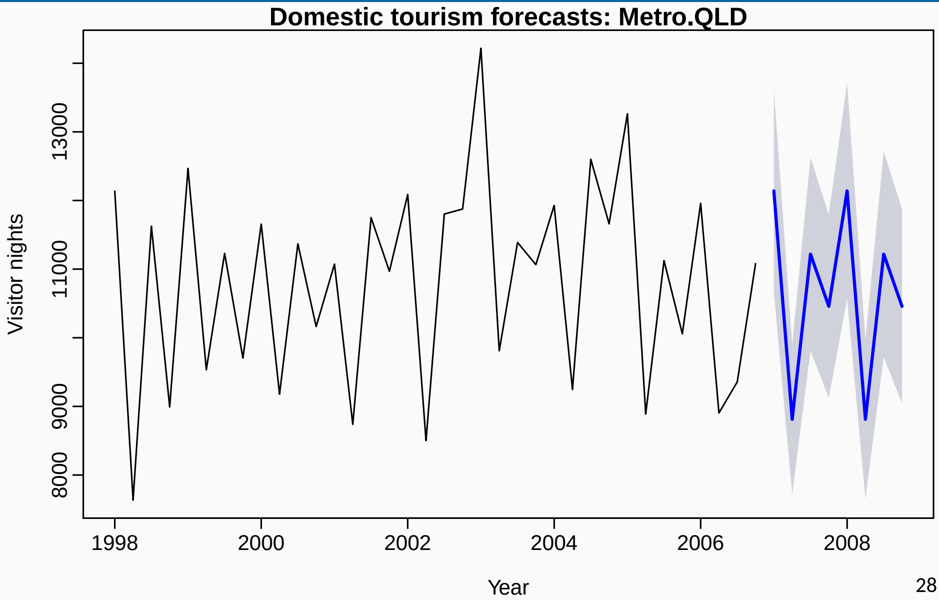
Base forecasts



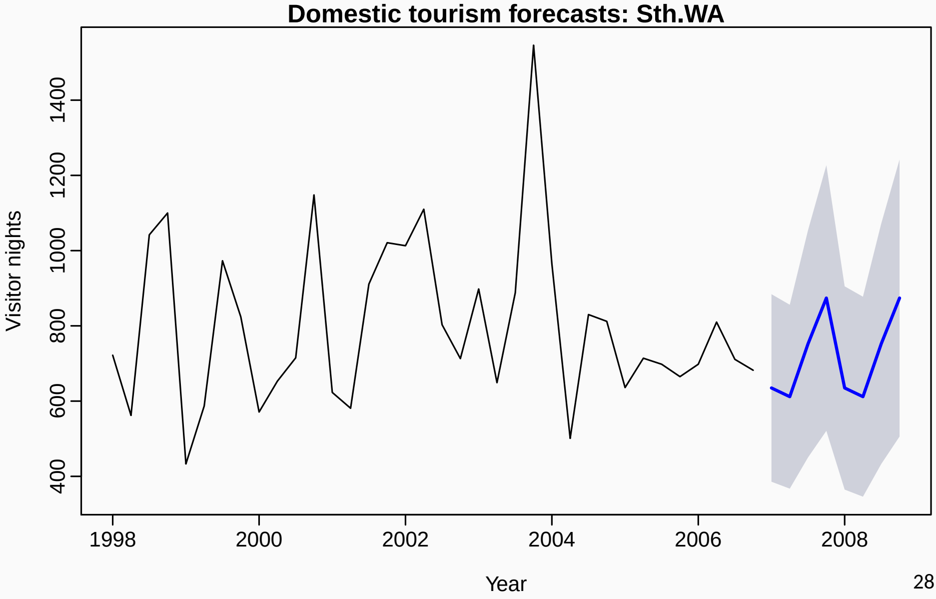
Base forecasts



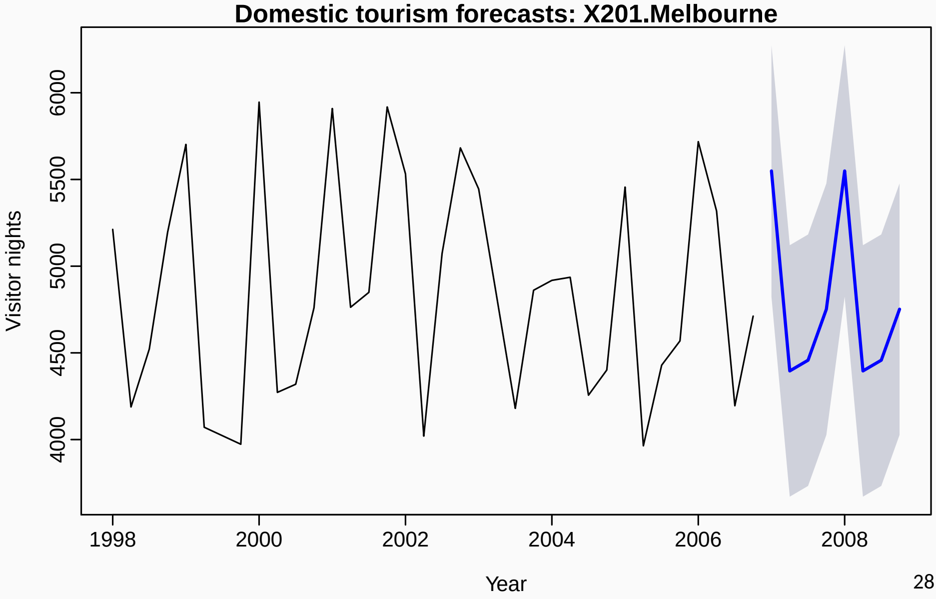
Base forecasts



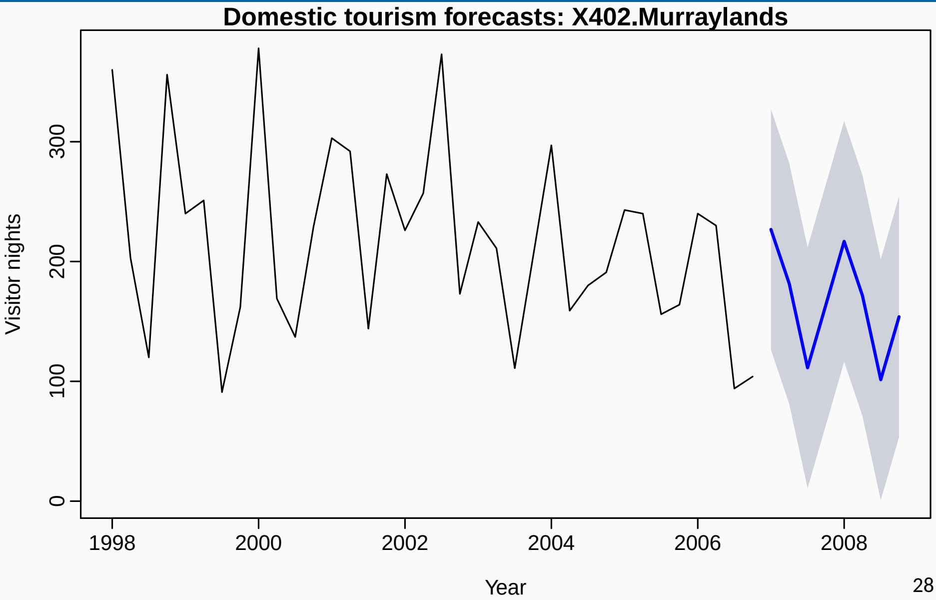
Base forecasts



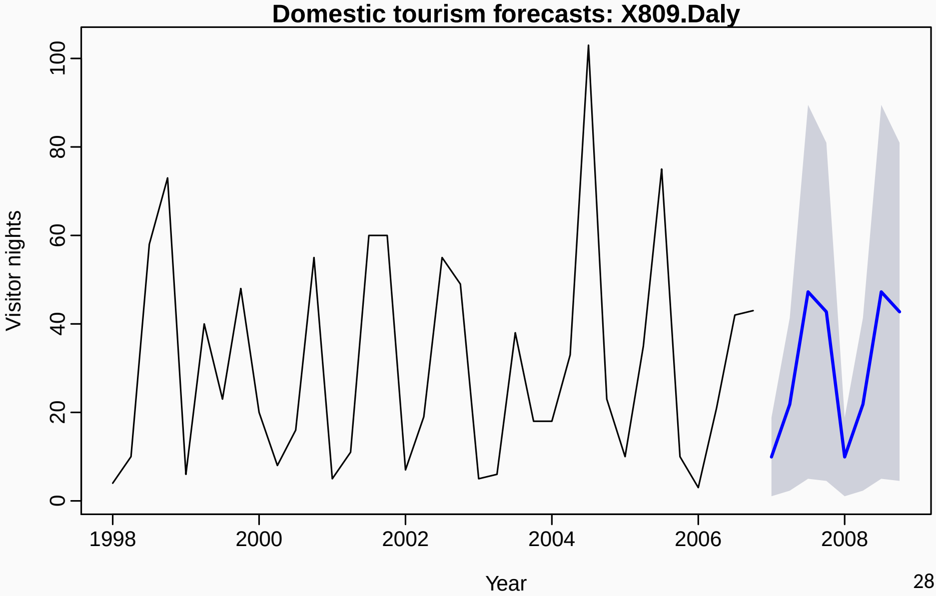
Base forecasts



Base forecasts

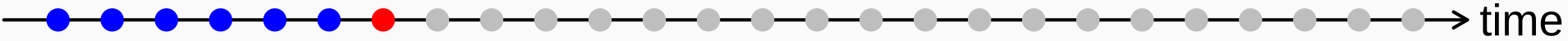


Base forecasts



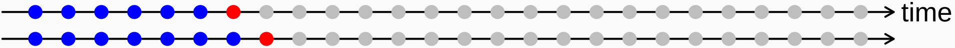
Forecast evaluation

Training sets **Test sets $h = 1$**



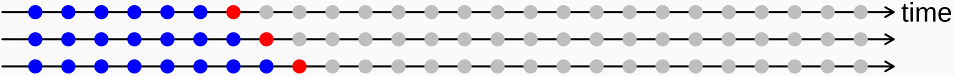
Forecast evaluation

Training sets **Test sets $h = 1$**



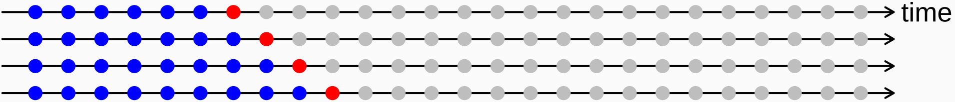
Forecast evaluation

Training sets **Test sets $h = 1$**



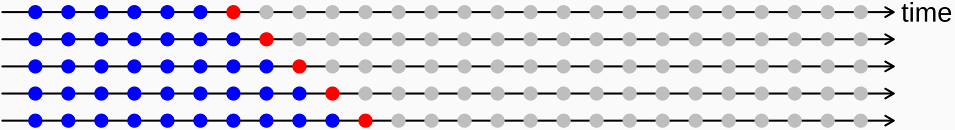
Forecast evaluation

Training sets **Test sets $h = 1$**



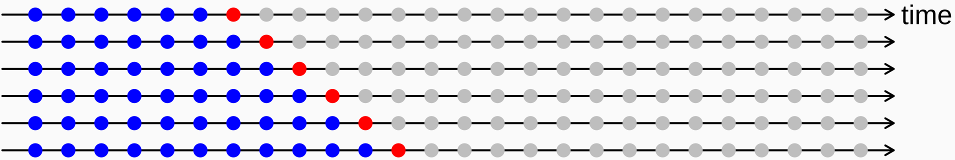
Forecast evaluation

Training sets **Test sets $h = 1$**



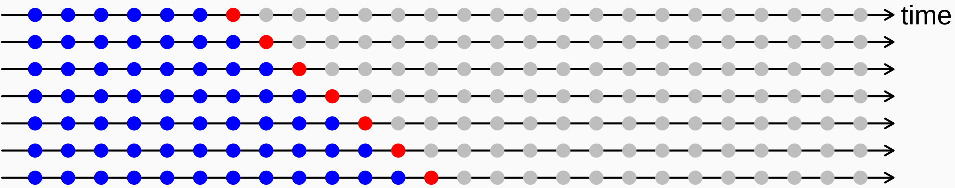
Forecast evaluation

Training sets **Test sets $h = 1$**



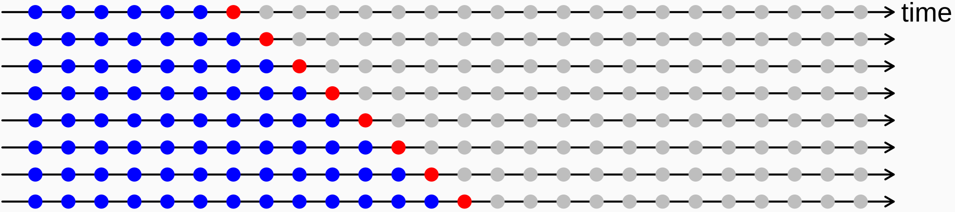
Forecast evaluation

Training sets **Test sets $h = 1$**



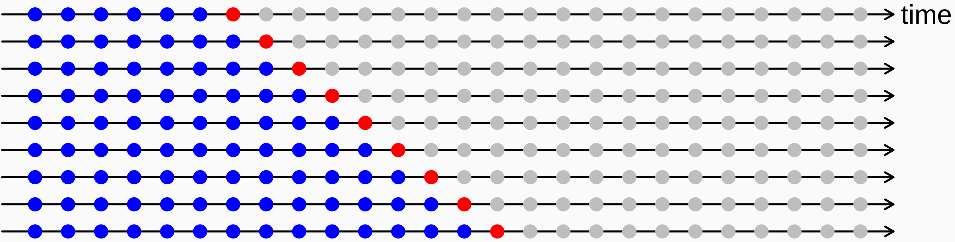
Forecast evaluation

Training sets **Test sets $h = 1$**



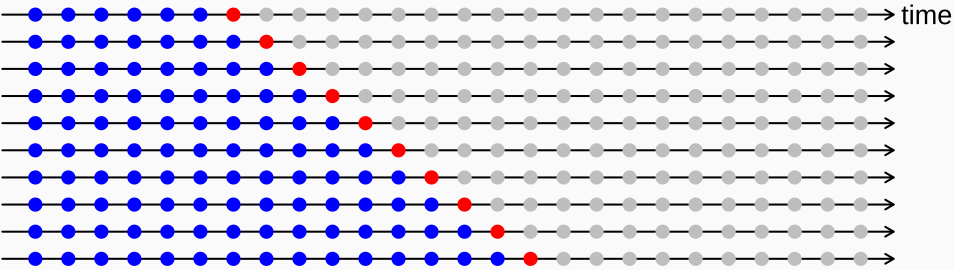
Forecast evaluation

Training sets **Test sets $h = 1$**



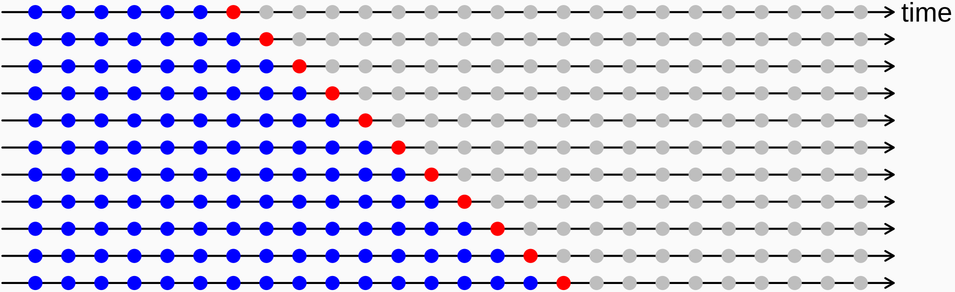
Forecast evaluation

Training sets **Test sets $h = 1$**



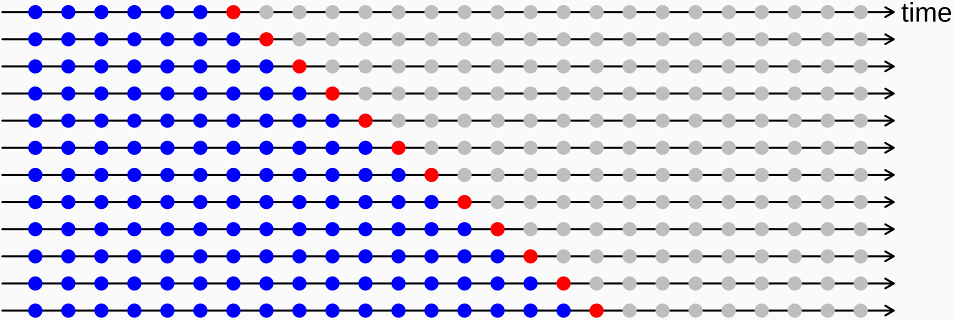
Forecast evaluation

Training sets **Test sets $h = 1$**



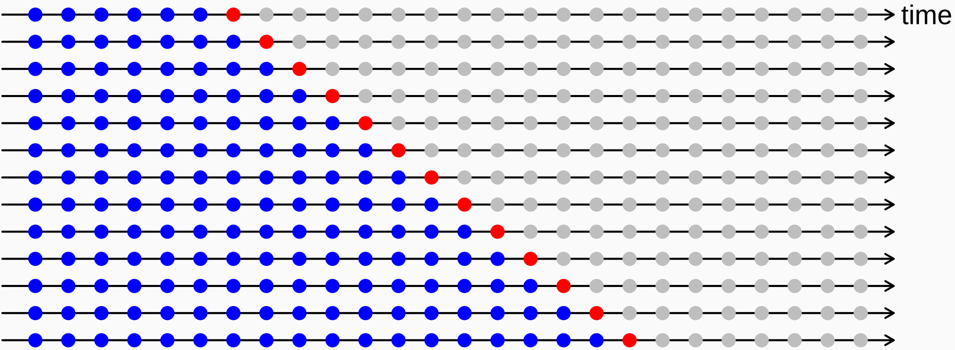
Forecast evaluation

Training sets **Test sets $h = 1$**



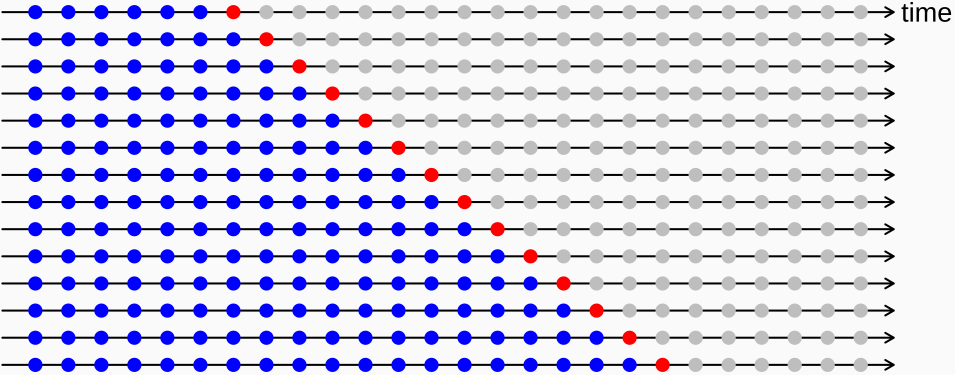
Forecast evaluation

Training sets **Test sets $h = 1$**



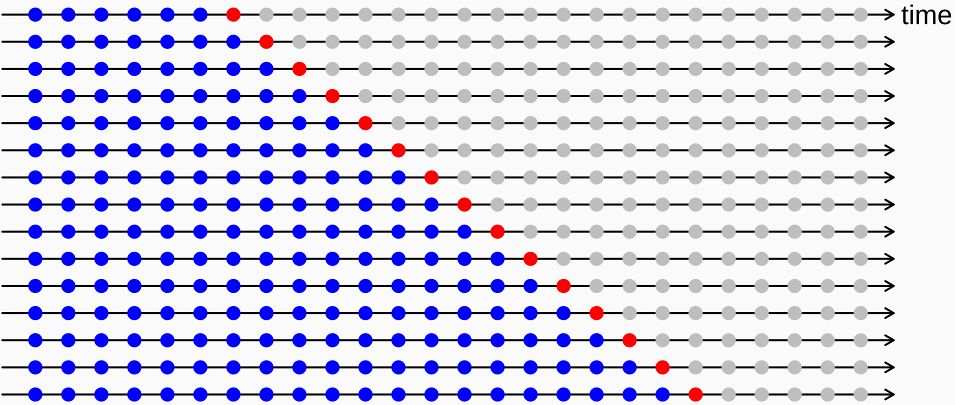
Forecast evaluation

Training sets **Test sets $h = 1$**



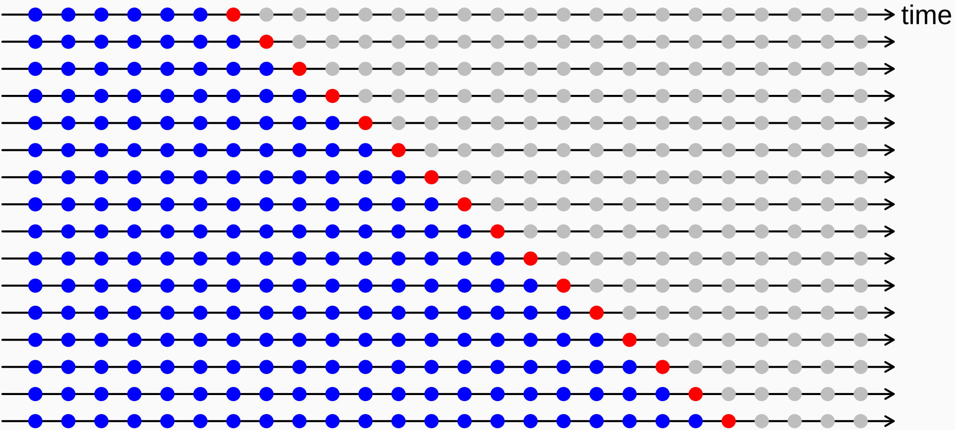
Forecast evaluation

Training sets **Test sets $h = 1$**



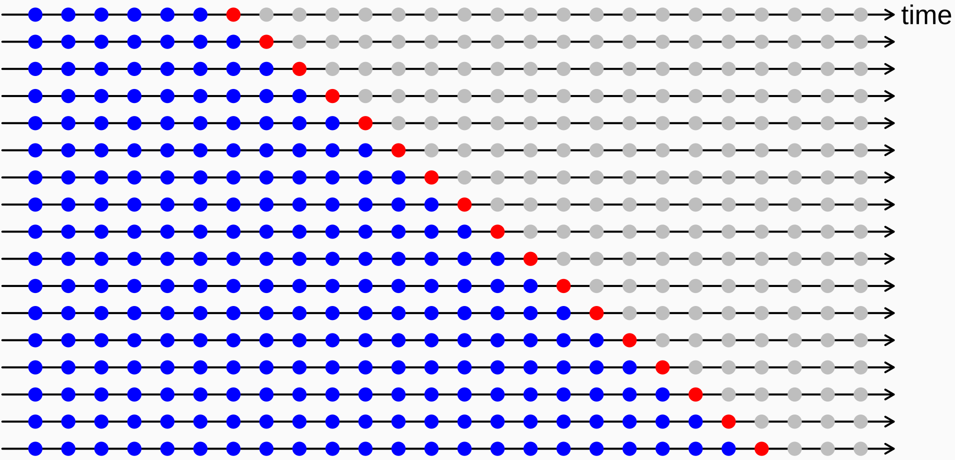
Forecast evaluation

Training sets **Test sets $h = 1$**



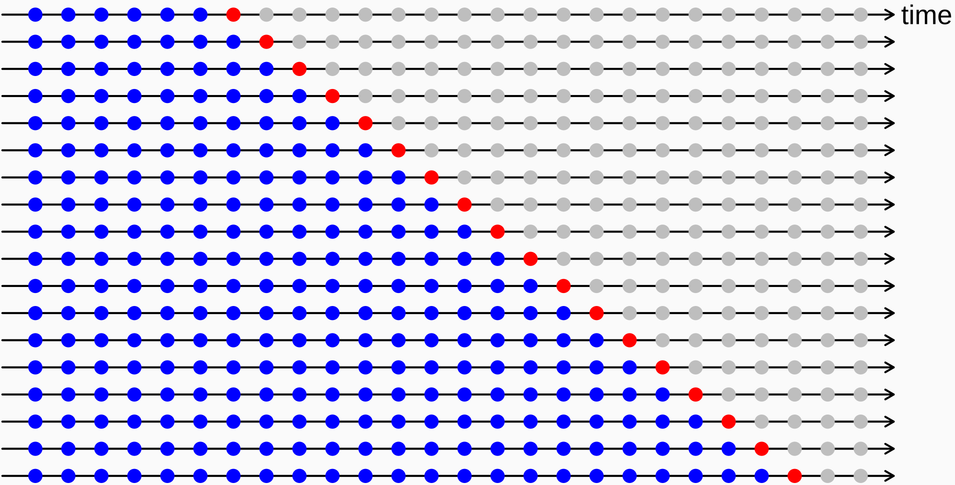
Forecast evaluation

Training sets **Test sets $h = 1$**



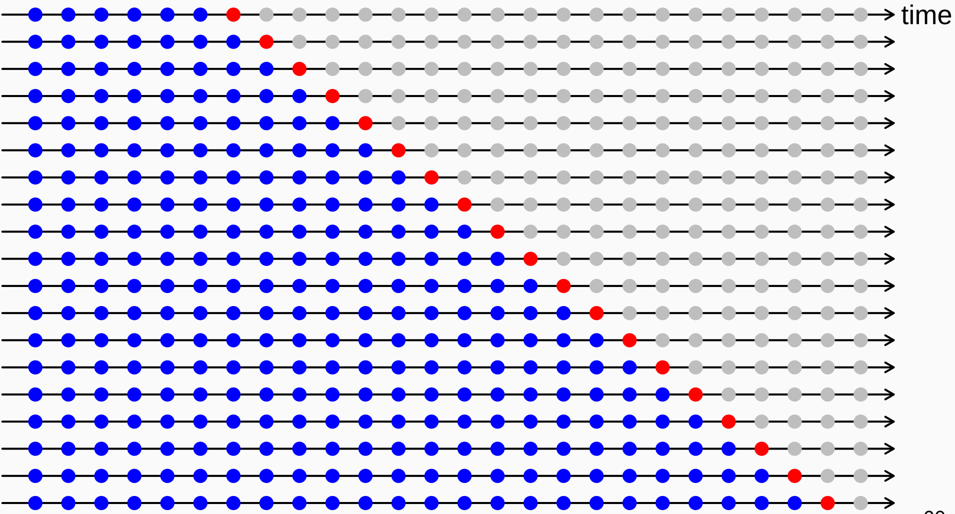
Forecast evaluation

Training sets **Test sets $h = 1$**



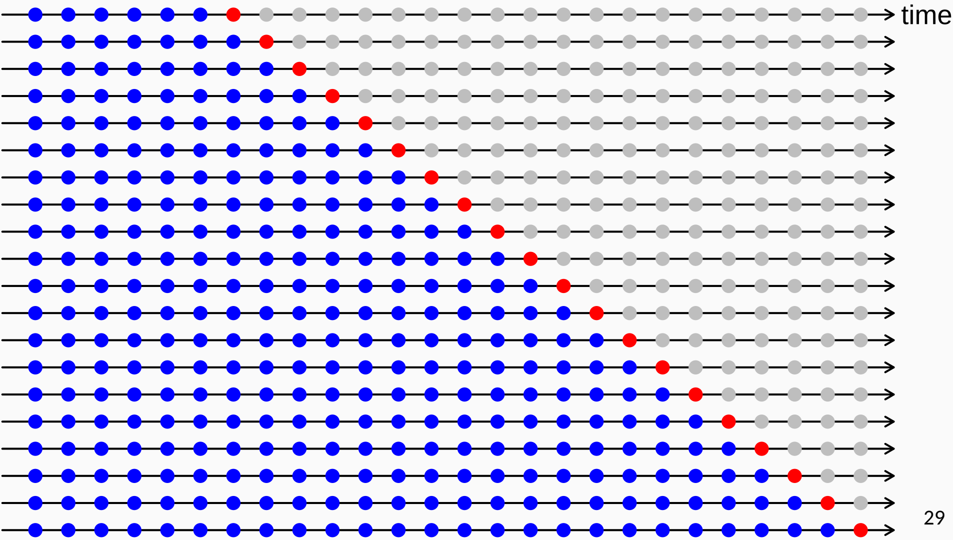
Forecast evaluation

Training sets **Test sets $h = 1$**



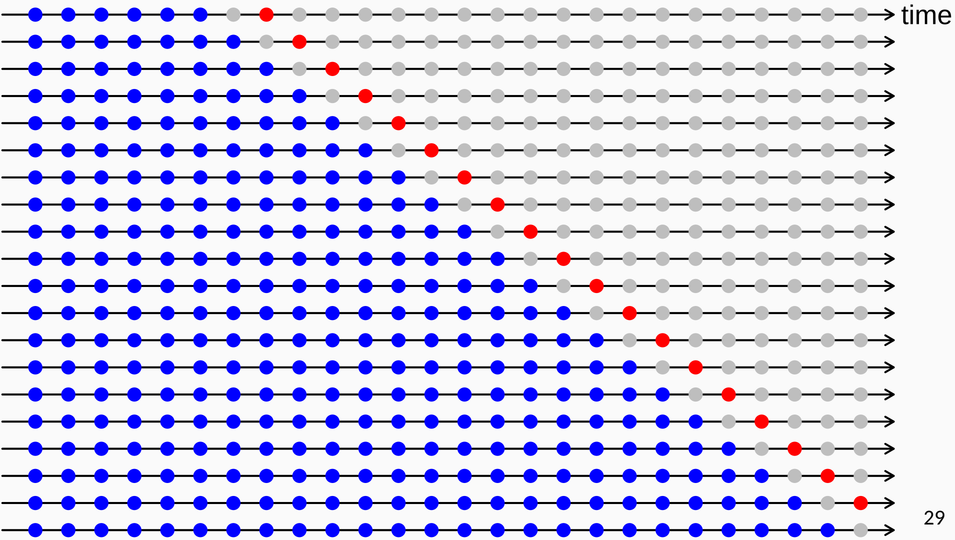
Forecast evaluation

Training sets **Test sets $h = 1$**



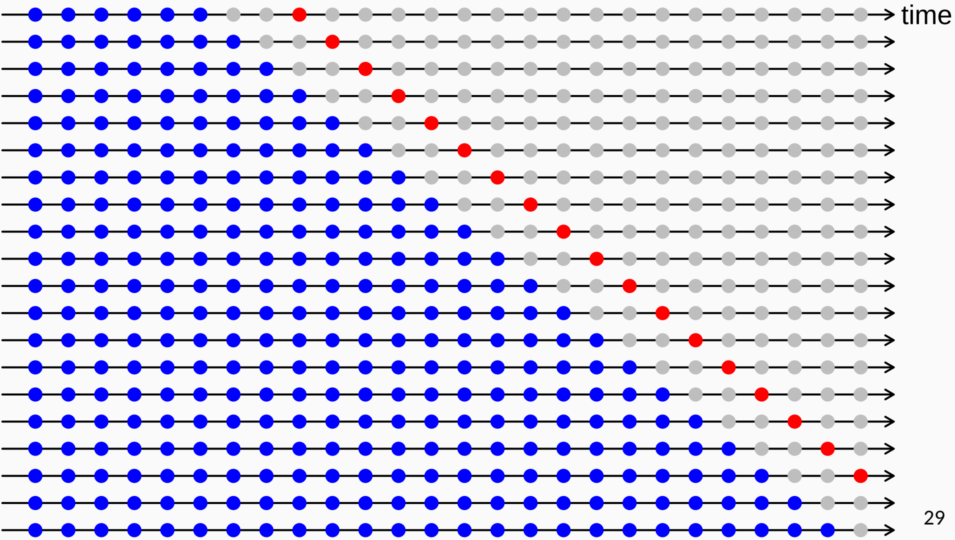
Forecast evaluation

Training sets **Test sets $h = 2$**



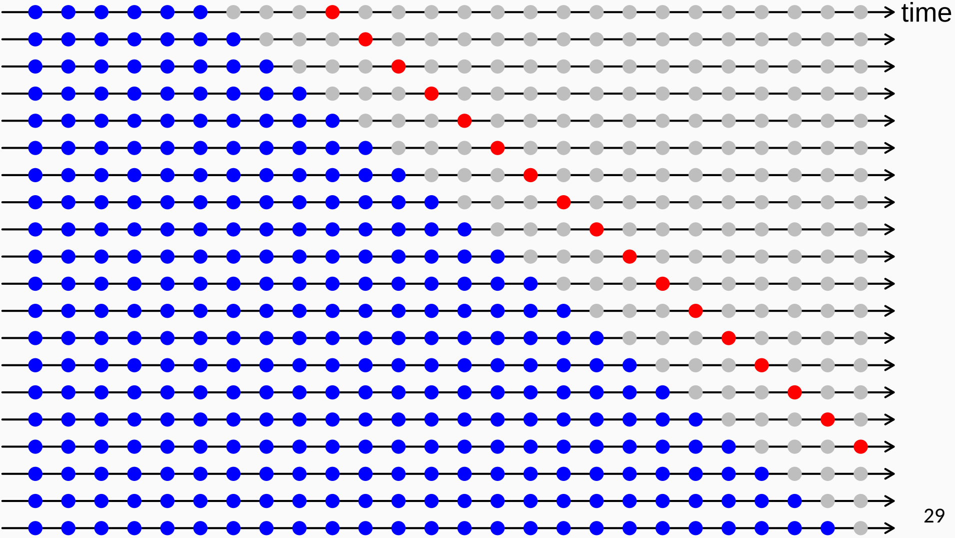
Forecast evaluation

Training sets **Test sets $h = 3$**



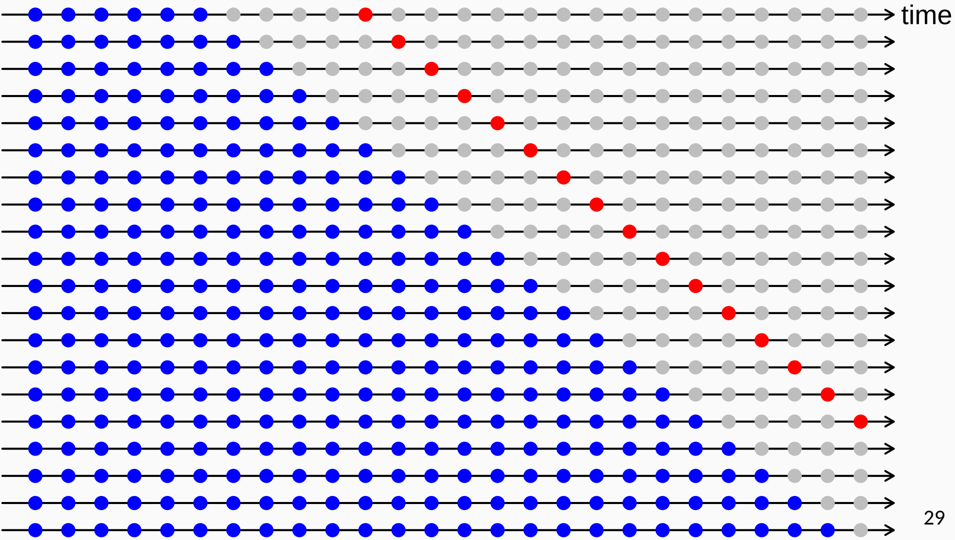
Forecast evaluation

Training sets **Test sets $h = 4$**



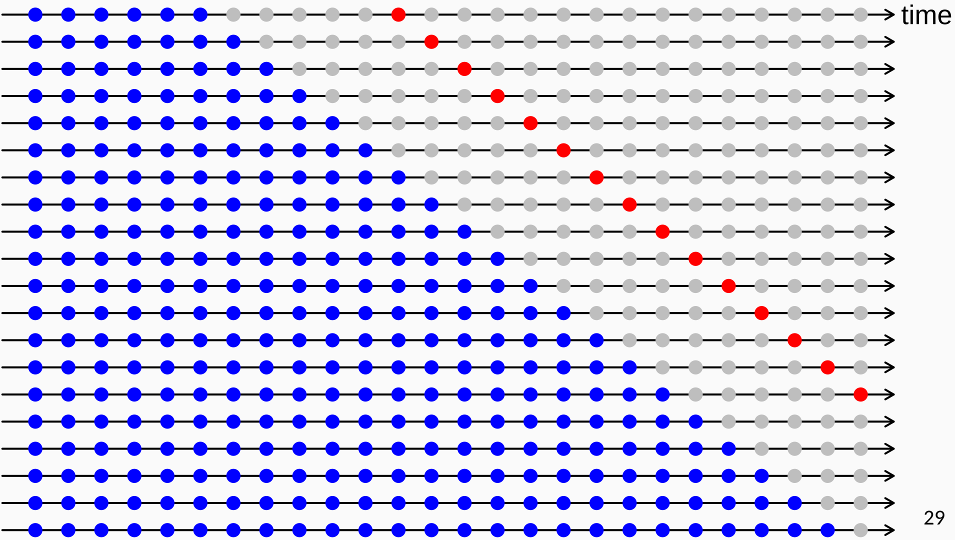
Forecast evaluation

Training sets **Test sets $h = 5$**



Forecast evaluation

Training sets **Test sets $h = 6$**

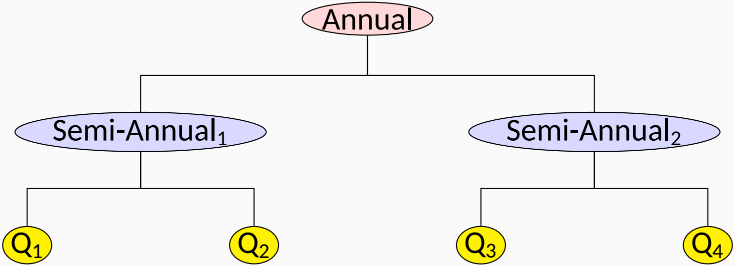


Hierarchy: states, zones, regions

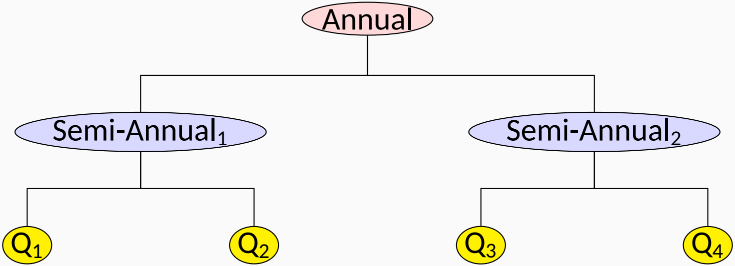
RMSE	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	Ave
			Forecast horizon				
			Australia				
Base	1762.04	1770.29	1766.02	1818.82	1705.35	1721.17	**1757.28**
Bottom	1736.92	1742.69	1722.79	1752.74	1666.73	1687.43	**1718.22**
WLS	1705.21	1715.87	1703.75	1729.56	1627.79	1661.24	**1690.57**
GLS	1704.64	1715.60	1705.31	1729.04	1626.36	1661.64	**1690.43**
			States				
Base	399.77	404.16	401.92	407.26	395.38	401.17	**401.61**
Bottom	404.29	406.95	404.96	409.02	399.80	401.55	**404.43**
WLS	398.84	402.12	400.71	405.03	394.76	398.23	**399.95**
GLS	398.84	402.16	400.86	405.03	394.59	398.22	**399.95**
			Regions				
Base	93.15	93.38	93.45	93.79	93.50	93.56	**93.47**
Bottom	93.15	93.38	93.45	93.79	93.50	93.56	**93.47**
WLS	93.02	93.32	93.38	93.72	93.39	93.53	**93.39**
GLS	92.98	93.27	93.34	93.66	93.34	93.46	**93.34**

Outline

Temporal hierarchies



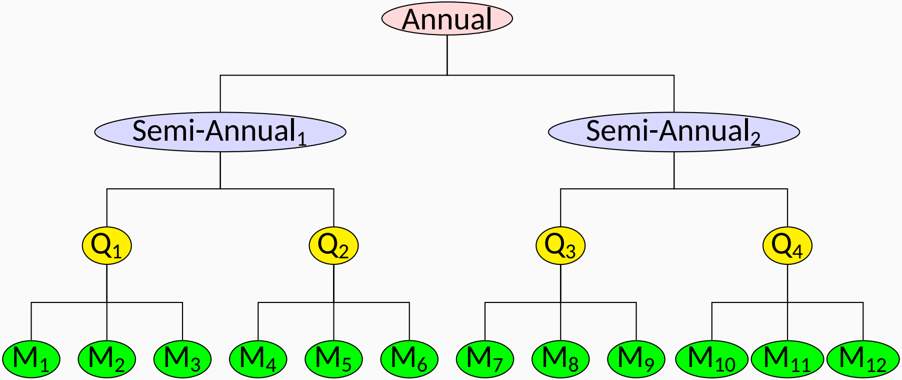
Temporal hierarchies



Basic idea:

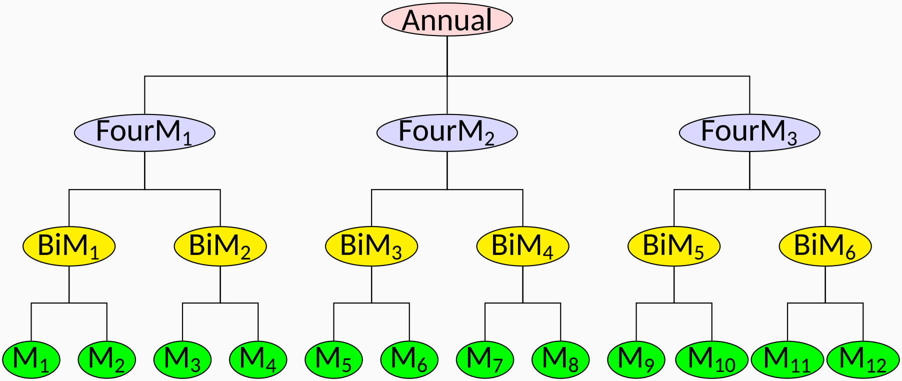
- Forecast series at each available frequency.
- Optimally reconcile forecasts within the same year.

Monthly series



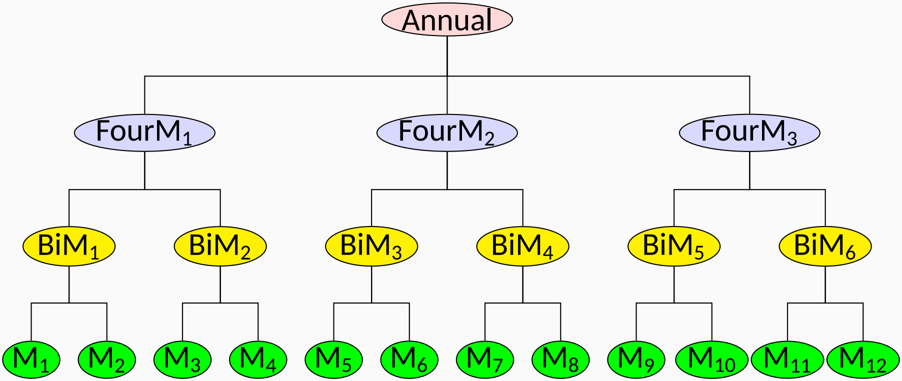
- $k = 2, 4, 12$ nodes

Monthly series



- $k = 2, 4, 12$ nodes
- $k = 3, 6, 12$ nodes

Monthly series



- $k = 2, 4, 12$ nodes
- $k = 3, 6, 12$ nodes
- Why not $k = 2, 3, 4, 6, 12$ nodes?

Monthly data

$$
\underbrace{\begin{pmatrix} A \\ SemiA_1 \\ SemiA_2 \\ FourM_1 \\ FourM_2 \\ FourM_3 \\ Q_1 \\ \vdots \\ Q_4 \\ BiM_1 \\ \vdots \\ BiM_6 \\ M_1 \\ \vdots \\ M_{12} \end{pmatrix}}_{(28\times 1)}
=
\underbrace{\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & \vdots & & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & & & \vdots & & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
 & & & & & & \boldsymbol{I}_{12} & & & & &
\end{pmatrix}}_{\boldsymbol{S}}
\underbrace{\begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \\ M_5 \\ M_6 \\ M_7 \\ M_8 \\ M_9 \\ M_{10} \\ M_{11} \\ M_{12} \end{pmatrix}}_{\boldsymbol{b}_t}
$$

In general

For a time series $y_1, \dots, y_T$, observed at frequency m, we generate aggregate series

$$y_j^{[k]} = \sum_{t=1+(j-1)k}^{jk} y_t, \qquad \text{for } j = 1, \dots, \lfloor T/k \rfloor$$

- $k \in F(m) = \{\text{factors of } m\}$.

In general

For a time series $y_1, \dots, y_T$, observed at frequency m, we generate aggregate series

$$y_j^{[k]} = \sum_{t=1+(j-1)k}^{jk} y_t, \qquad \text{for } j = 1, \dots, \lfloor T/k \rfloor$$

- $k \in F(m) = \{\text{factors of } m\}$.
- A single unique hierarchy is only possible when there are no coprime pairs in $F(m)$.

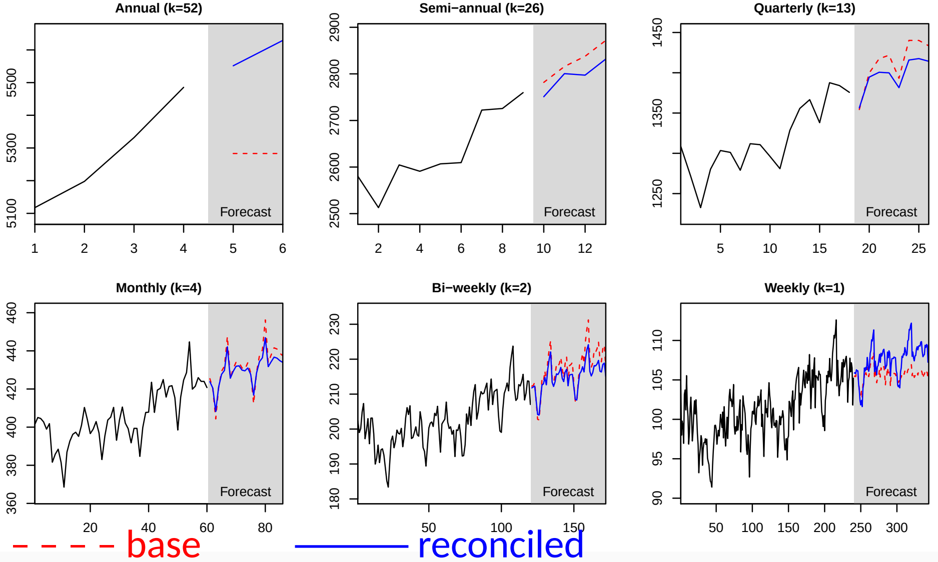
In general

For a time series $y_1, \dots, y_T$, observed at frequency m, we generate aggregate series

$$y_j^{[k]} = \sum_{t=1+(j-1)k}^{jk} y_t, \qquad \text{for } j = 1, \dots, \lfloor T/k \rfloor$$

- $k \in F(m) = \{\text{factors of } m\}$.
- A single unique hierarchy is only possible when there are no coprime pairs in $F(m)$.
- $M_k = m/k$ is seasonal period of aggregated series.

UK Accidents and Emergency Demand



Outline

R packages

Papers, packages and slides available at **robjhyndman.com**