

# Forecast reconciliation

## 3. Probabilistic forecast reconciliation

Rob J Hyndman



MONASH University

Photo by Edvard Alexander Rølvaag on Unsplash

[robjhyndman.com/fr2023](http://robjhyndman.com/fr2023)

# Outline

- 1 Definition of probabilistic coherence
- 2 Evaluating probabilistic forecasts
- 3 Emergency Services Demand
- 4 Evaluating multivariate probabilistic forecasts
- 5 Example: Australian electricity generation

# Outline

- 1 Definition of probabilistic coherence
- 2 Evaluating probabilistic forecasts
- 3 Emergency Services Demand
- 4 Evaluating multivariate probabilistic forecasts
- 5 Example: Australian electricity generation

# The coherent subspace

## Coherent subspace

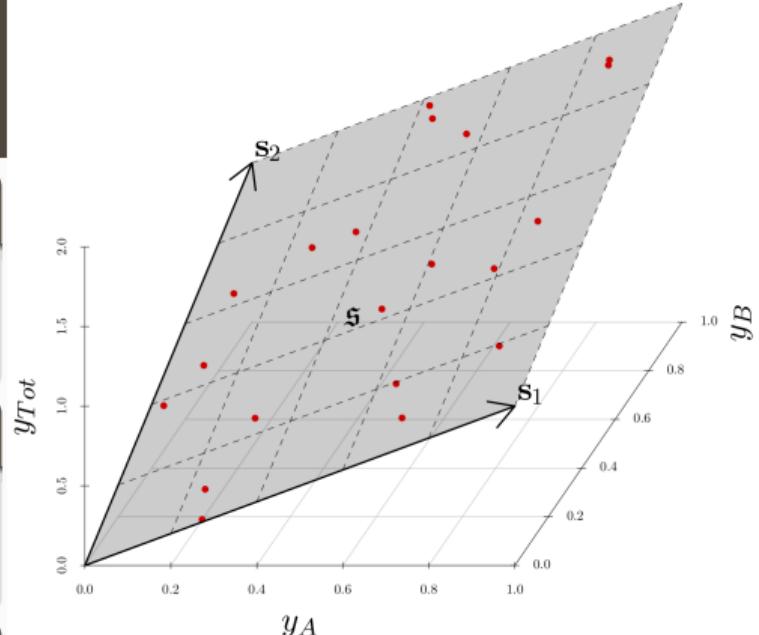
$m$ -dimensional linear subspace  $\mathfrak{s} \subset \mathbb{R}^n$  for which linear constraints hold for all  $\mathbf{y} \in \mathfrak{s}$ .

## Hierarchical time series

An  $n$ -dimensional multivariate time series such that  $\mathbf{y}_t \in \mathfrak{s} \quad \forall t$ .

## Coherent point forecasts

$\tilde{\mathbf{y}}_{t+h|t}$  is *coherent* if  $\tilde{\mathbf{y}}_{t+h|t} \in \mathfrak{s}$ .



$$y_{Tot} = y_A + y_B$$

# The coherent subspace

## Coherent subspace

$m$ -dimensional linear subspace  $\mathfrak{s} \subset \mathbb{R}^n$  for which linear constraints hold for all  $\mathbf{y} \in \mathfrak{s}$ .

## Hierarchical time series

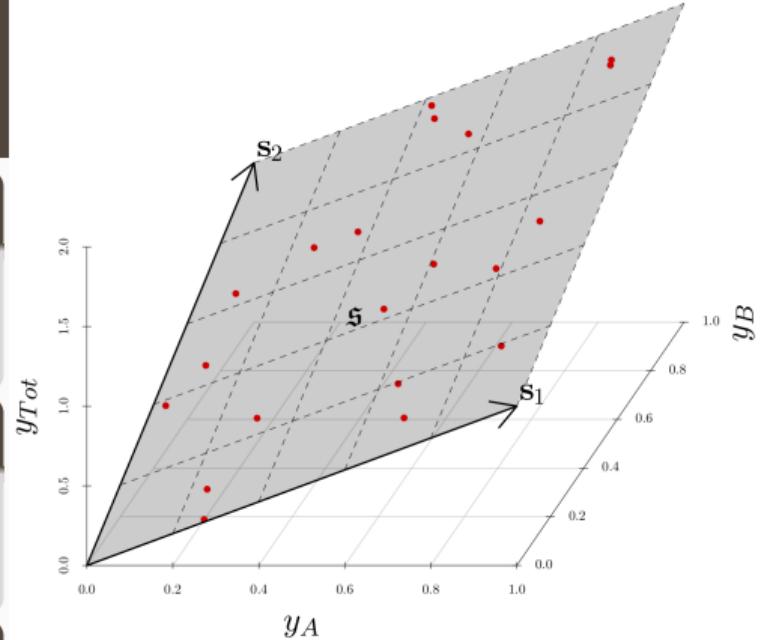
An  $n$ -dimensional multivariate time series such that  $\mathbf{y}_t \in \mathfrak{s} \quad \forall t$ .

## Coherent point forecasts

$\tilde{\mathbf{y}}_{t+h|t}$  is *coherent* if  $\tilde{\mathbf{y}}_{t+h|t} \in \mathfrak{s}$ .

## Base forecasts

Let  $\hat{\mathbf{y}}_{t+h|t}$  be vector of *incoherent* initial  $h$ -step forecasts.



$$y_{Tot} = y_A + y_B$$

# The coherent subspace

## Coherent subspace

$m$ -dimensional linear subspace  $\mathfrak{s} \subset \mathbb{R}^n$  for which linear constraints hold for all  $\mathbf{y} \in \mathfrak{s}$ .

## Hierarchical time series

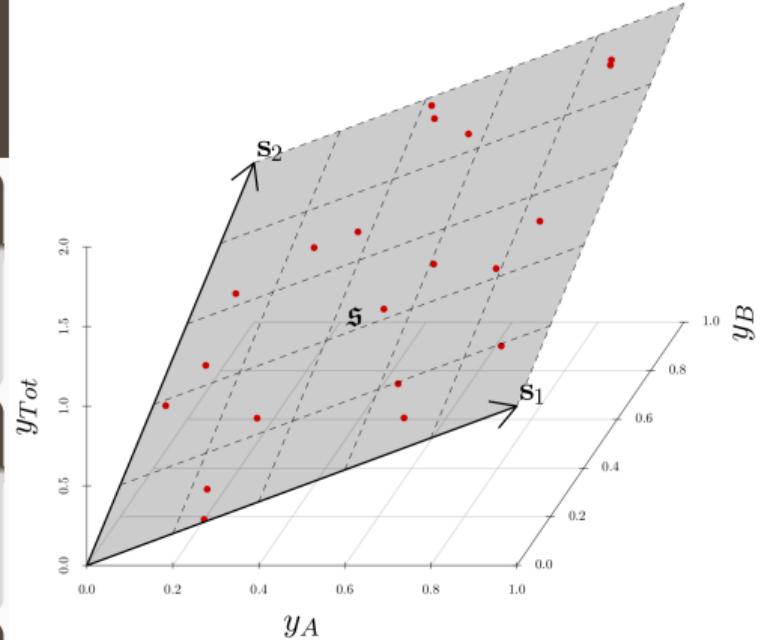
An  $n$ -dimensional multivariate time series such that  $\mathbf{y}_t \in \mathfrak{s} \quad \forall t$ .

## Coherent point forecasts

$\tilde{\mathbf{y}}_{t+h|t}$  is *coherent* if  $\tilde{\mathbf{y}}_{t+h|t} \in \mathfrak{s}$ .

## Base forecasts

Let  $\hat{\mathbf{y}}_{t+h|t}$  be vector of *incoherent* initial  $h$ -step forecasts.



$$y_{Tot} = y_A + y_B$$

## Reconciled forecasts

Let  $\psi$  be a mapping,  $\psi : \mathbb{R}^n \rightarrow \mathfrak{s}$ .  
 $\tilde{\mathbf{y}}_{t+h|t} = \psi(\hat{\mathbf{y}}_{t+h|t})$  “reconciles”  $\hat{\mathbf{y}}_{t+h|t}$ .

# Coherent probabilistic forecasts

## Coherent probabilistic forecasts

Given the triple  $(\mathbb{R}^m, \mathcal{F}_{\mathbb{R}^m}, \nu)$ , a coherent probability triple  $(\mathfrak{s}, \mathcal{F}_{\mathfrak{s}}, \check{\nu})$  is such that

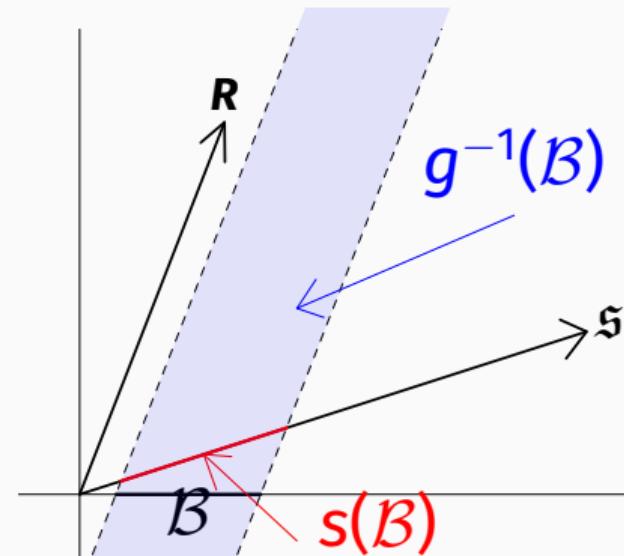
$$\check{\nu}(s(\mathcal{B})) = \nu(\mathcal{B}) \quad \forall \mathcal{B} \in \mathcal{F}_{\mathbb{R}^m}$$

## Probabilistic forecast reconciliation

The reconciled probability measure of  $\hat{\nu}$  wrt  $\psi(\cdot)$  is such that

$$\tilde{\nu}(\mathcal{B}) = \hat{\nu}(\psi^{-1}(\mathcal{B})) \quad \forall \mathcal{B} \in \mathcal{F}_{\mathfrak{s}}$$

where  $\psi^{-1}(\mathcal{B}) := \{\mathbf{y} \in \mathbb{R}^n : \psi(\mathbf{y}) \in \mathcal{B}\}$  is the pre-image of  $\mathcal{B}$ , that is the set of all points in  $\mathbb{R}^n$  that  $\psi(\cdot)$  maps to a point in  $\mathcal{B}$ .



# Construction of reconciled distributions

## Reconciled density of bottom-level

Density of bottom-level series under reconciled distribution is

$$\tilde{f}_b(\mathbf{b}) = |\mathbf{G}^*| \int \hat{f}(\mathbf{G}^- \mathbf{b} + \mathbf{G}_\perp \mathbf{a}) d\mathbf{a}$$

- $\hat{f}$  is density of incoherent base probabilistic forecasts
- $\mathbf{G}^-$  is  $n \times m$  generalised inverse of  $\mathbf{G}$  st  $\mathbf{G}\mathbf{G}^- = \mathbf{I}$
- $\mathbf{G}_\perp$  is  $n \times (n - m)$  orthogonal complement to  $\mathbf{G}$  st  $\mathbf{G}\mathbf{G}_\perp = \mathbf{0}$
- $\mathbf{G}^* = (\mathbf{G}^- : \mathbf{G}_\perp)$ , and  $\mathbf{b}$  and  $\mathbf{a}$  are obtained via

the change of variables  $\mathbf{y} = \mathbf{G}^* \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \end{pmatrix}$

# Construction of reconciled distributions

## Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_y(\mathbf{y}) = |\mathbf{S}^*| \tilde{f}_b(\mathbf{S}^- \mathbf{y}) \mathbb{1}\{\mathbf{y} \in \mathfrak{s}\}$$

- $\mathbf{S}^* = (\mathbf{S}^{-'} \quad \mathbf{S}_\perp)'$
- $\mathbf{S}^-$  is  $m \times n$  generalised inverse of  $\mathbf{S}$  such that  $\mathbf{S}^- \mathbf{S} = \mathbf{I}$
- $\mathbf{S}_\perp$  is  $n \times (n - m)$  orthogonal complement to  $\mathbf{S}$  such that  $\mathbf{S}'_\perp \mathbf{S} = \mathbf{0}$

# Construction of reconciled distributions

## Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_y(\mathbf{y}) = |\mathbf{S}^*| \tilde{f}_b(\mathbf{S}^- \mathbf{y}) \mathbb{1}\{\mathbf{y} \in \mathfrak{s}\}$$

- $\mathbf{S}^* = (\mathbf{S}^{-'} \quad \mathbf{S}_\perp)'$
- $\mathbf{S}^-$  is  $m \times n$  generalised inverse of  $\mathbf{S}$  such that  $\mathbf{S}^- \mathbf{S} = \mathbf{I}$
- $\mathbf{S}_\perp$  is  $n \times (n - m)$  orthogonal complement to  $\mathbf{S}$  such that  $\mathbf{S}'_\perp \mathbf{S} = \mathbf{0}$

## Gaussian reconciliation

If the incoherent base forecasts are  $N(\hat{\mu}, \hat{\Sigma})$ , then the reconciled density is  $N(\mathbf{S}\mathbf{G}\hat{\mu}, \mathbf{S}\mathbf{G}\hat{\Sigma}\mathbf{G}'\mathbf{S}')$

# Construction of reconciled distributions

## Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_y(\mathbf{y}) = |\mathbf{S}^*| \tilde{f}_b(\mathbf{S}^- \mathbf{y}) \mathbb{1}\{\mathbf{y} \in \mathfrak{s}\}$$

- $\mathbf{S}^* = (\mathbf{S}^{-'} \quad \mathbf{S}_\perp)'$
- $\mathbf{S}^-$  is  $m \times n$  generalised inverse of  $\mathbf{S}$  such that  $\mathbf{S}^- \mathbf{S} = \mathbf{I}$
- $\mathbf{S}_\perp$  is  $n \times (n - m)$  orthogonal complement to  $\mathbf{S}$  such that  $\mathbf{S}'_\perp \mathbf{S} = \mathbf{0}$

## Gaussian reconciliation

If the incoherent base forecasts are  $N(\hat{\mu}, \hat{\Sigma})$ , then the reconciled density is  $N(\mathbf{S}\mathbf{G}\hat{\mu}, \mathbf{S}\mathbf{G}\hat{\Sigma}\mathbf{G}'\mathbf{S}')$

## Bootstrap reconciliation

Reconciling sample paths from incoherent distributions works.

# Simulation from a reconciled distribution

Suppose that  $(\hat{\mathbf{y}}^{[1]}, \dots, \hat{\mathbf{y}}^{[L]})$  is a sample drawn from an incoherent probability measure  $\hat{\nu}$ . Then  $(\tilde{\mathbf{y}}^{[1]}, \dots, \tilde{\mathbf{y}}^{[L]})$  where  $\tilde{\mathbf{y}}^{[\ell]} := \psi(\hat{\mathbf{y}}^{[\ell]})$  for  $\ell = 1, \dots, L$ , is a sample drawn from the reconciled probability measure  $\tilde{\nu}$ .

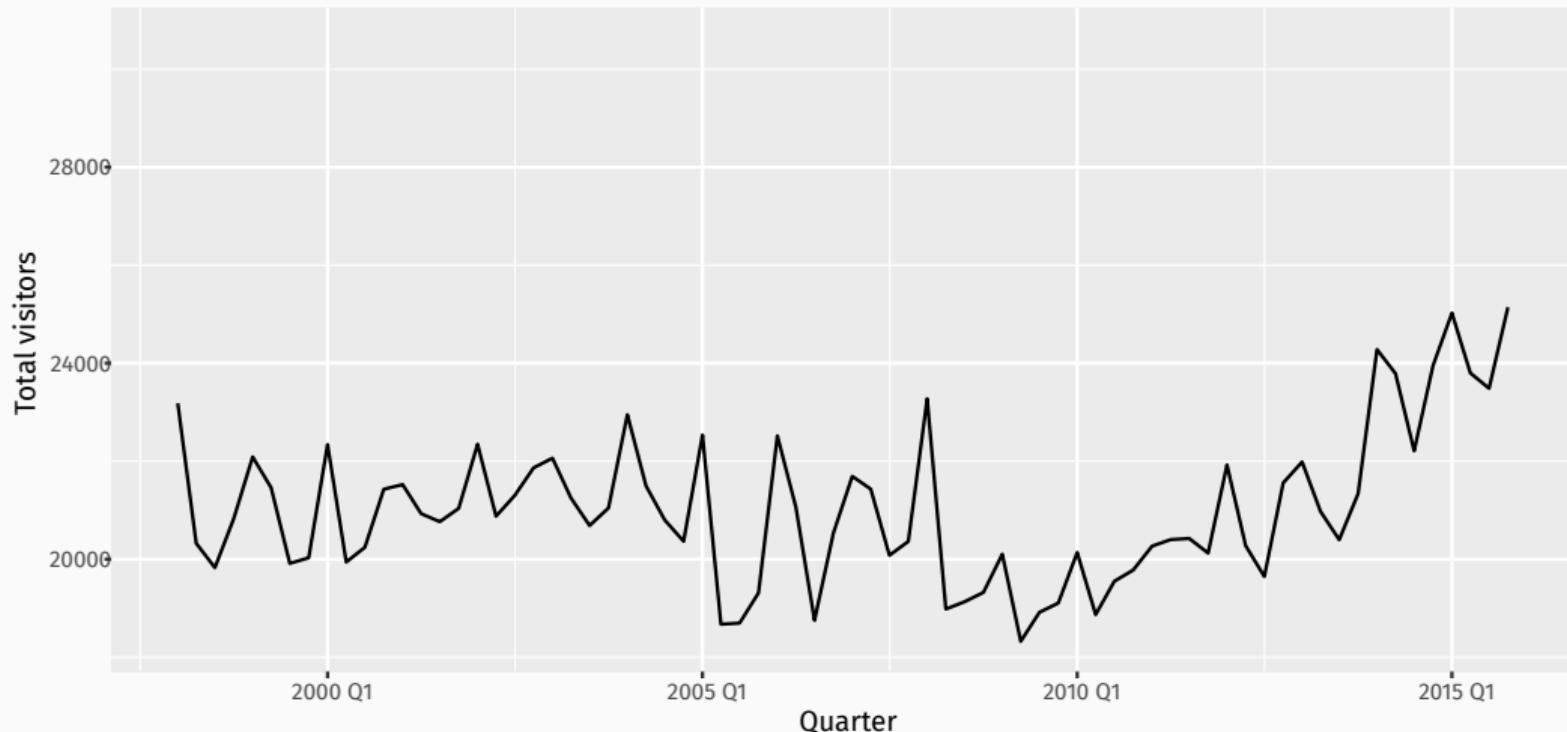
- Simulate future sample paths for each series using a multivariate bootstrap of the residuals from each model.
- Reconcile the sample paths.
- The reconciled sample paths are a sample from the reconciled distribution.

# Outline

- 1 Definition of probabilistic coherence
- 2 Evaluating probabilistic forecasts
- 3 Emergency Services Demand
- 4 Evaluating multivariate probabilistic forecasts
- 5 Example: Australian electricity generation

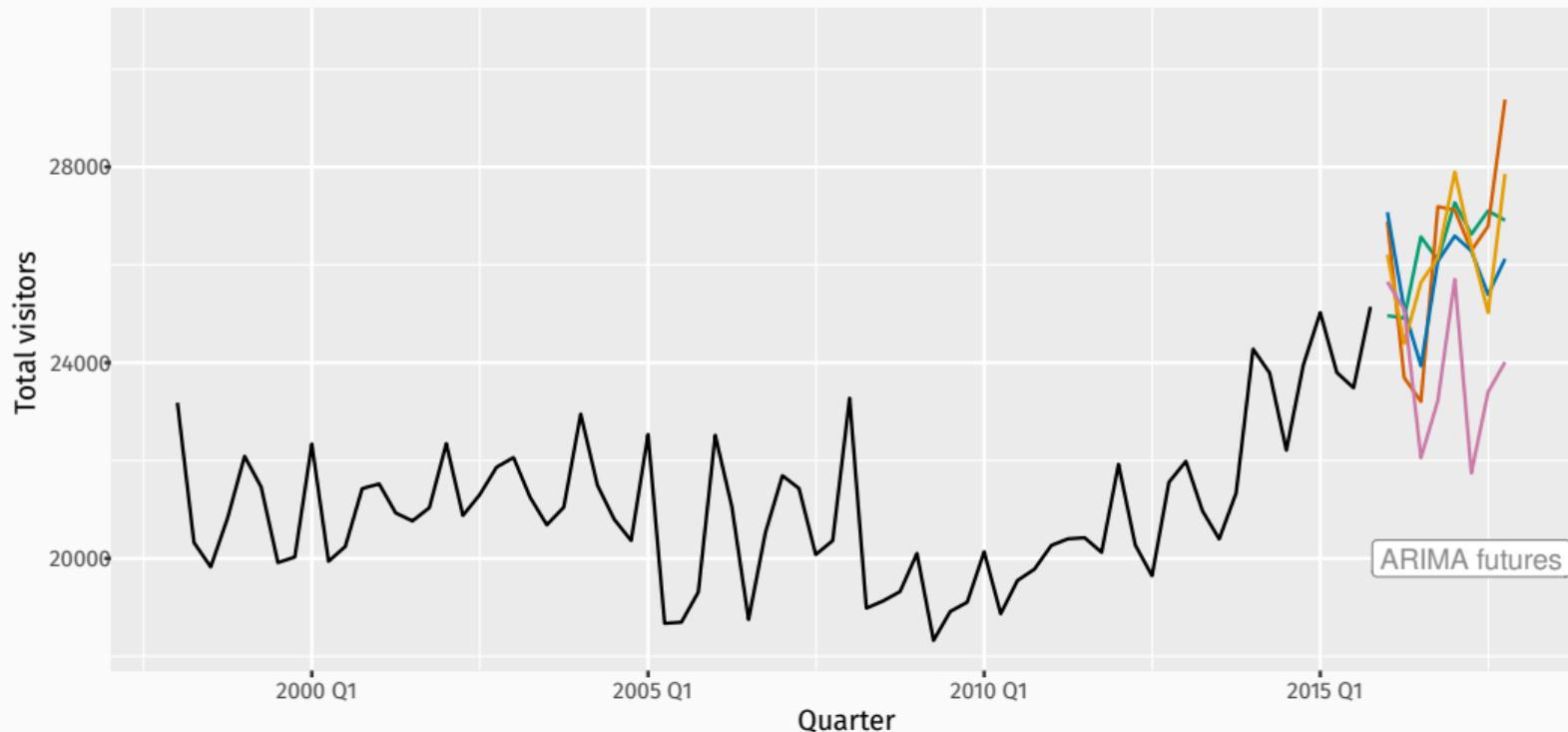
# Evaluating probabilistic forecasts

Australian domestic tourism



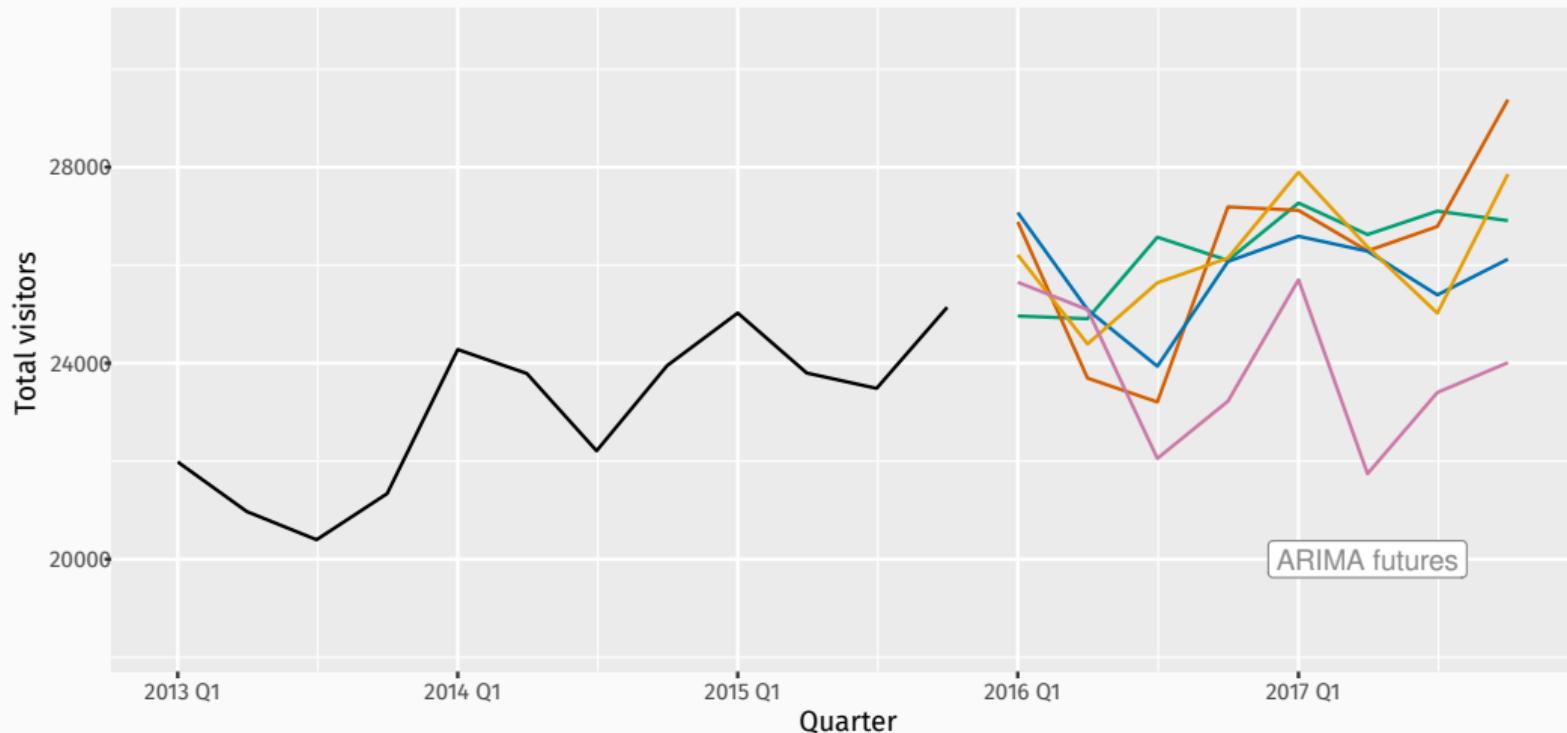
# Evaluating probabilistic forecasts

Australian domestic tourism



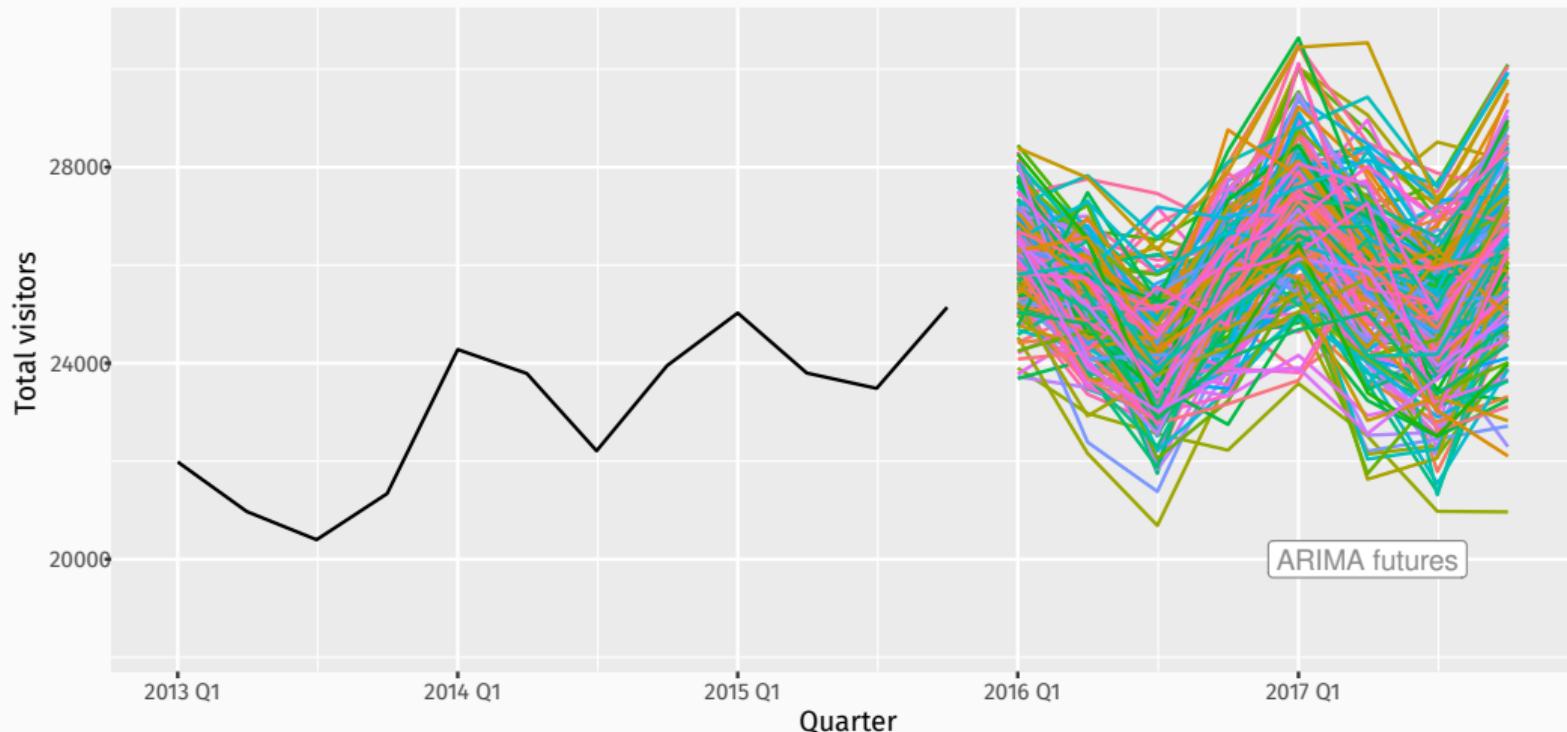
# Evaluating probabilistic forecasts

Australian domestic tourism



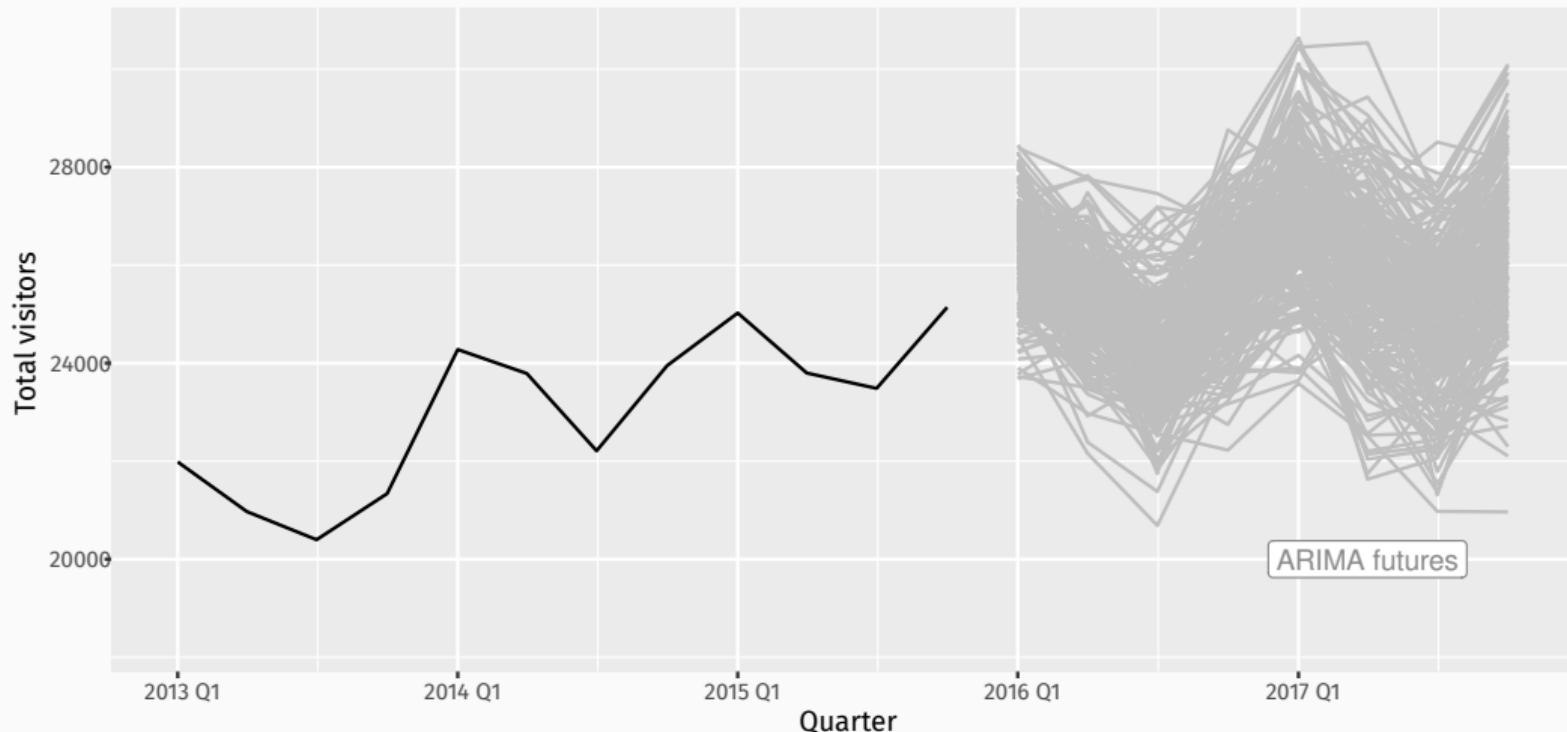
# Evaluating probabilistic forecasts

Australian domestic tourism



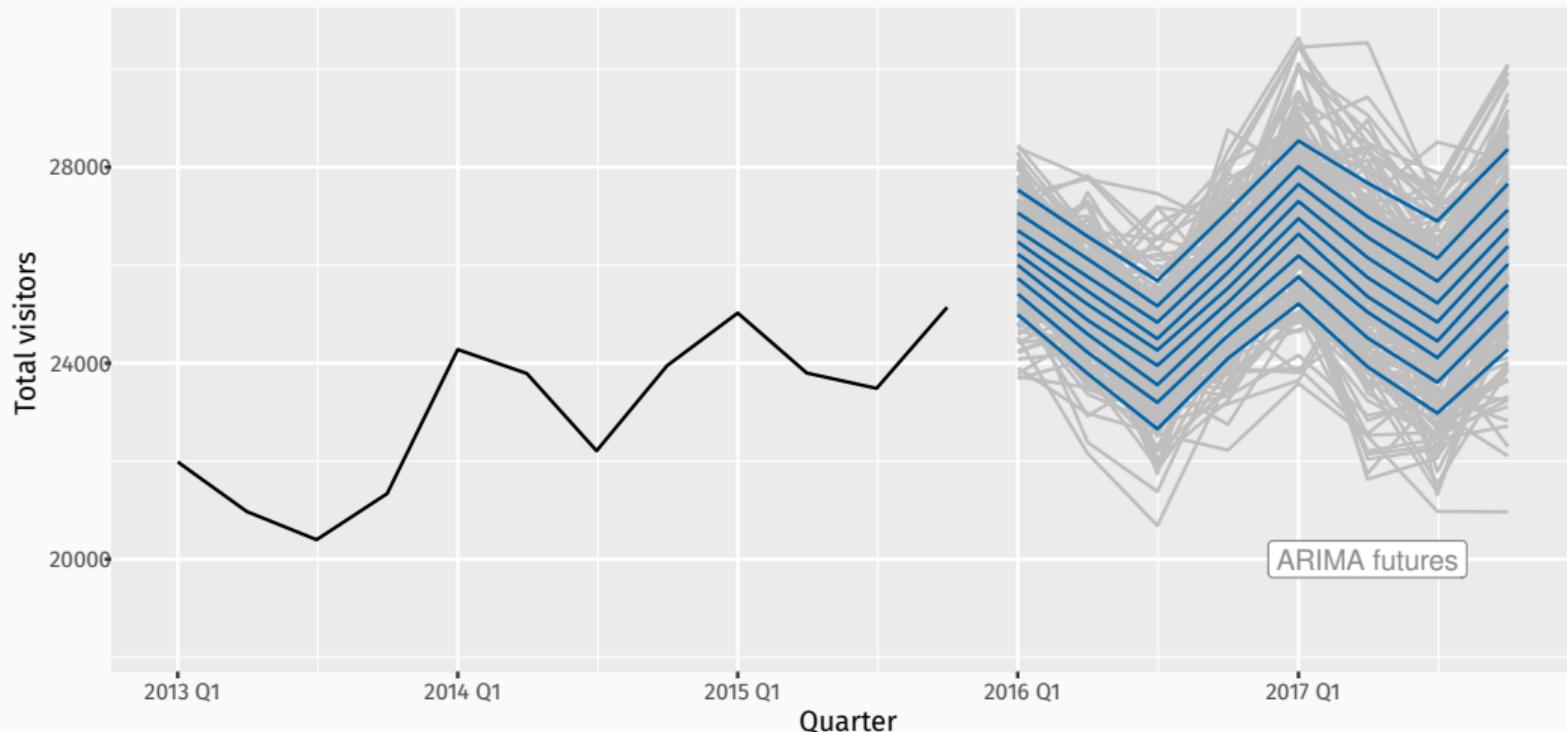
# Evaluating probabilistic forecasts

Australian domestic tourism



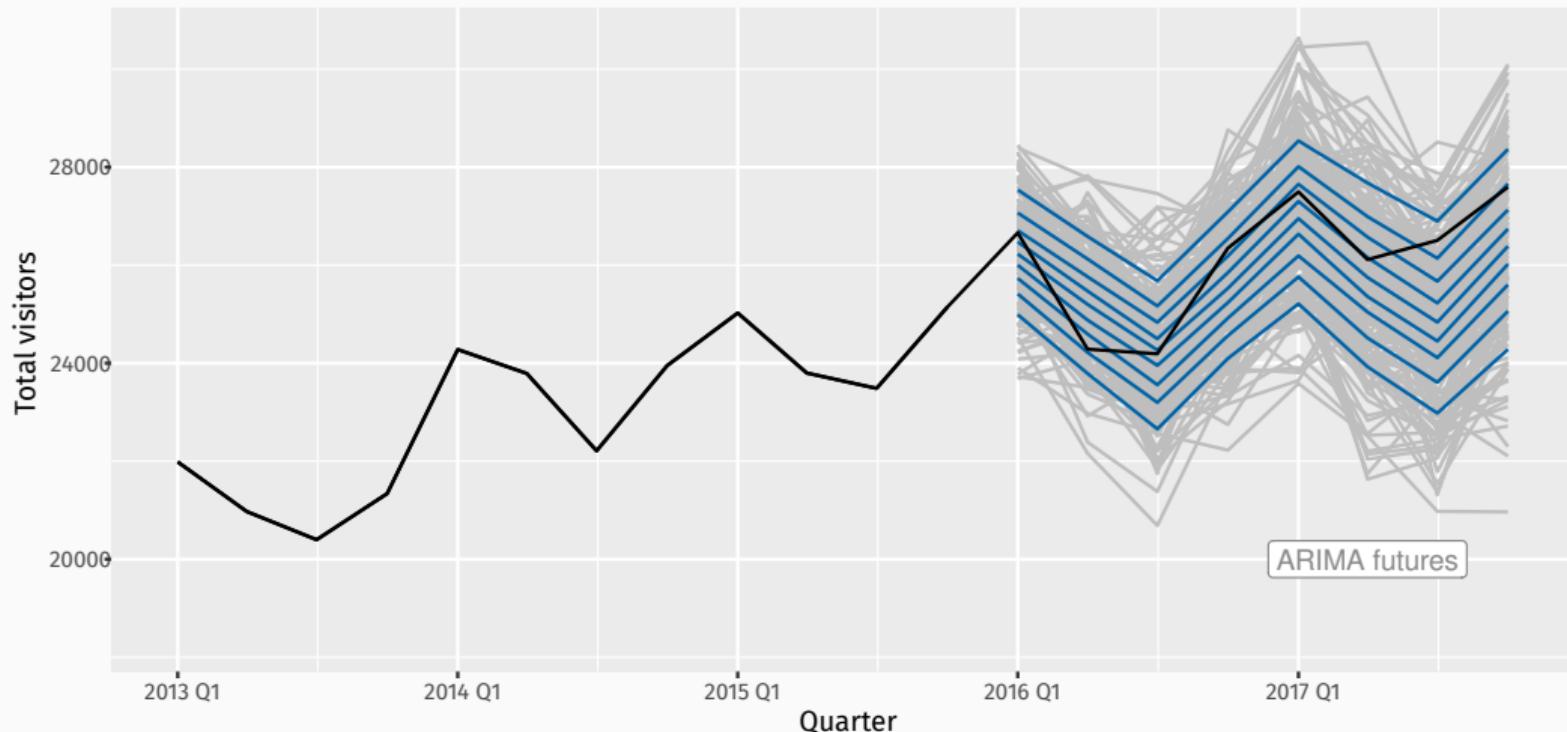
# Evaluating probabilistic forecasts

Australian domestic tourism



# Evaluating probabilistic forecasts

Australian domestic tourism



# Evaluating probabilistic forecasts

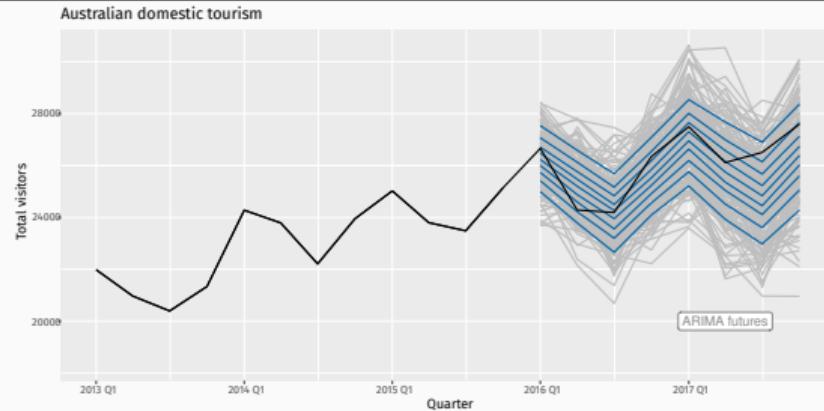
$y_t$  = observation at time  $t$

$q_{p,t}$  = quantile forecast: prob.  $p$ , time  $t$

## Quantile score

$$S_t(p, y) = \begin{cases} 2(1 - p)|y_t - q_{p,t}|, & \text{if } y_t < q_{p,t} \\ 2p|y_t - q_{p,t}|, & \text{if } y_t \geq q_{p,t} \end{cases}$$

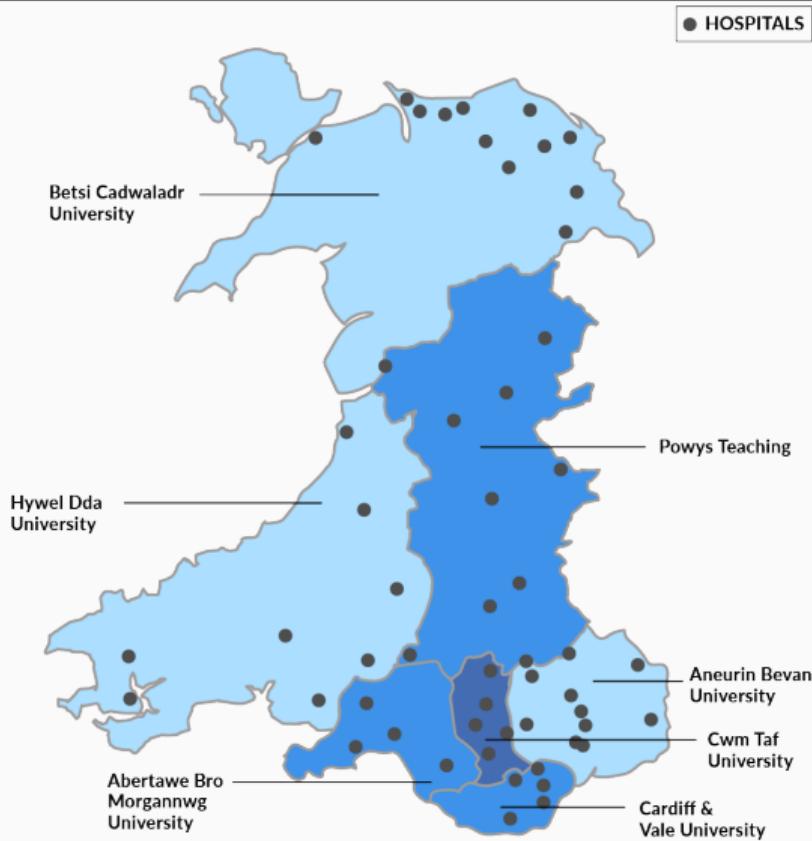
- Low  $S_t$  is good
- Multiplier of 2 often omitted, but useful for interpretation
- $S_t$  like absolute error, weighted to account for likely exceedance
- Average  $S_t(p, y)$  over  $p$  = CRPS (Continuous Rank Probability Score)



# Outline

- 1 Definition of probabilistic coherence
- 2 Evaluating probabilistic forecasts
- 3 Emergency Services Demand
- 4 Evaluating multivariate probabilistic forecasts
- 5 Example: Australian electricity generation

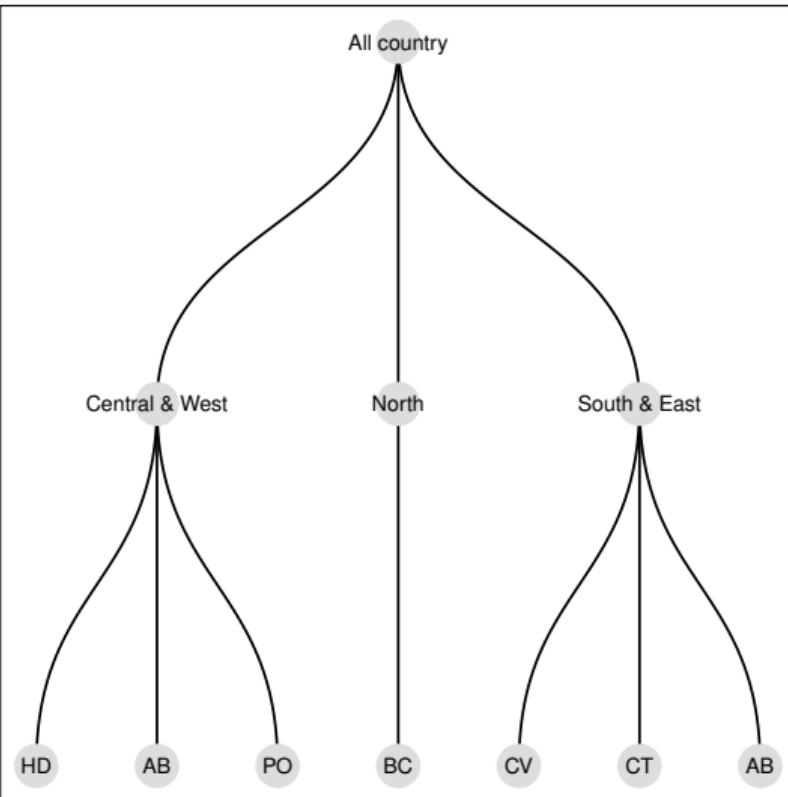
# Wales Health Board Areas



# Data

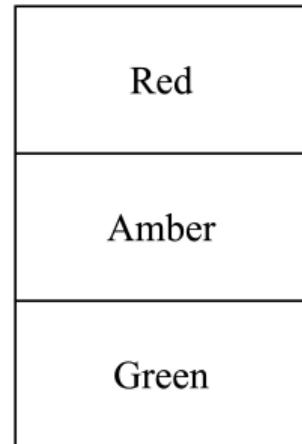
- Daily number of attended incidents:  
1 October 2015 – 31 July 2019
- Disaggregated by:
  - ▶ control area
  - ▶ health board
  - ▶ priority
  - ▶ nature of incidents
- 2,142,000 rows observations from 1,530 time series.

# Data structure



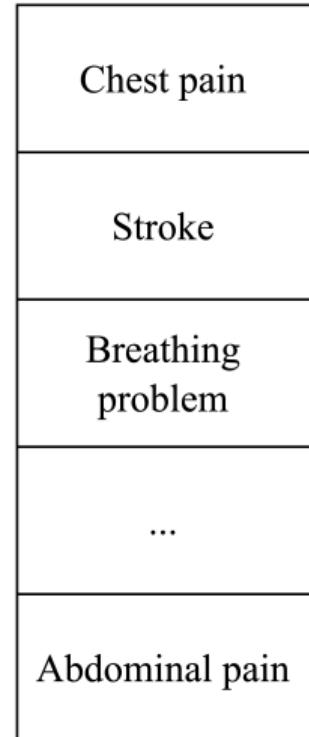
\*

## Priority



\*

## Nature of incident



# Data structure

Level	Number of series
All country	1
Control	3
Health board	7
Priority	3
Priority * Control	9
Priority * Health board	21
Nature of incident	35
Nature of incident * Control	105
Nature of incident * Health board	245
Priority * Nature of incident	104
Control * Priority * Nature of incident	306
Control * Health board * Priority * Nature of incident (Bottom level)	691
Total	1530

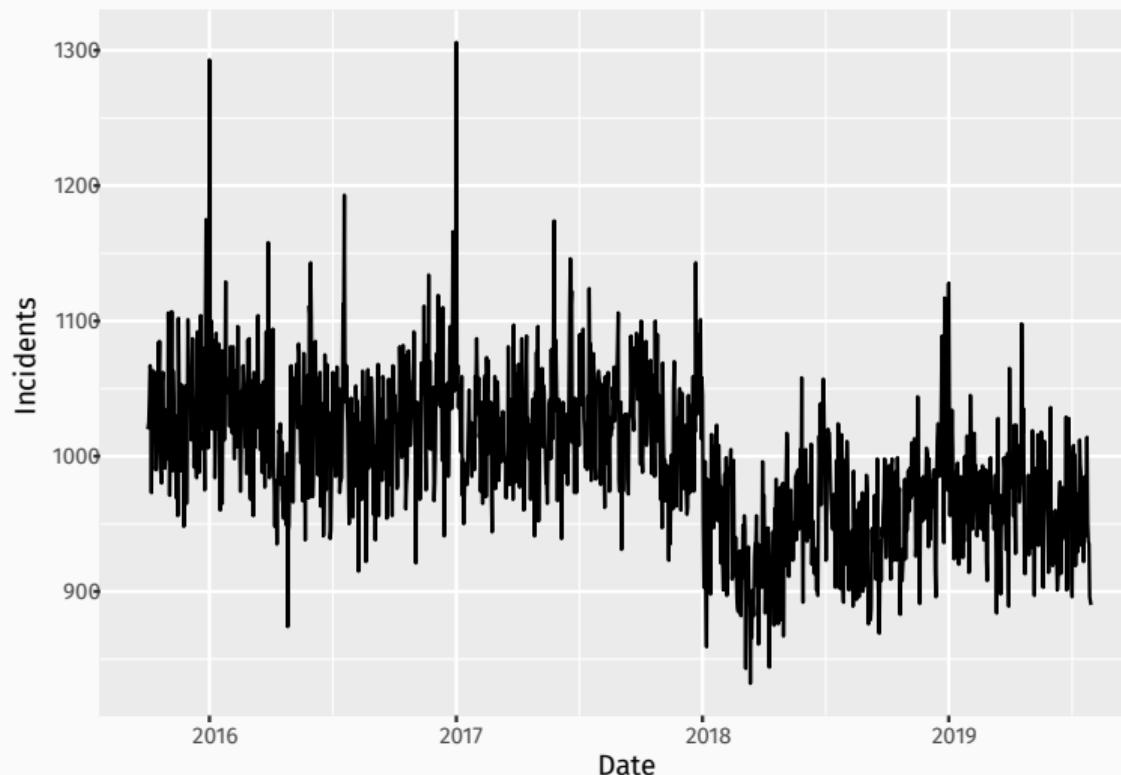
# Data

```
# A tsibble: 2,142,000 x 6 [1D]
# Key:      region, category, nature, lhb [1,530]
  date      region     category     nature      lhb       incident
  <date>    <chr*>     <chr*>     <chr*>     <chr*>     <dbl>
1 2015-10-01 <aggregated> <aggregated> <aggregated> <aggregated> 1020
2 2015-10-02 <aggregated> <aggregated> <aggregated> <aggregated> 1021
3 2015-10-03 <aggregated> <aggregated> <aggregated> <aggregated> 1025
4 2015-10-04 <aggregated> <aggregated> <aggregated> <aggregated> 1043
5 2015-10-05 <aggregated> <aggregated> <aggregated> <aggregated> 1067
6 2015-10-06 <aggregated> <aggregated> <aggregated> <aggregated> 1063
7 2015-10-07 <aggregated> <aggregated> <aggregated> <aggregated> 973
8 2015-10-08 <aggregated> <aggregated> <aggregated> <aggregated> 1057
9 2015-10-09 <aggregated> <aggregated> <aggregated> <aggregated> 1026
10 2015-10-10 <aggregated> <aggregated> <aggregated> <aggregated> 1063
# i 2,141,990 more rows
```

# Data

```
# A tsibble: 2,142,000 x 6 [1D]
# Key:      region, category, nature, lhb [1,530]
  date      region category nature     lhb      incident
  <date>    <chr*> <chr*>   <chr*>   <chr*>      <dbl>
1 2015-10-01 C      Amber    ABDOMINAL HD        0
2 2015-10-01 C      Amber    ABDOMINAL PO        0
3 2015-10-01 C      Amber    ABDOMINAL SB        0
4 2015-10-01 C      Amber    ABDOMINAL <aggregated> 0
5 2015-10-01 C      Amber    ALLERGIES HD        0
6 2015-10-01 C      Amber    ALLERGIES PO        1
7 2015-10-01 C      Amber    ALLERGIES SB        0
8 2015-10-01 C      Amber    ALLERGIES <aggregated> 1
9 2015-10-01 C      Amber    ANIMALBIT HD        0
10 2015-10-01 C     Amber    ANIMALBIT PO       0
# i 2,141,990 more rows
```

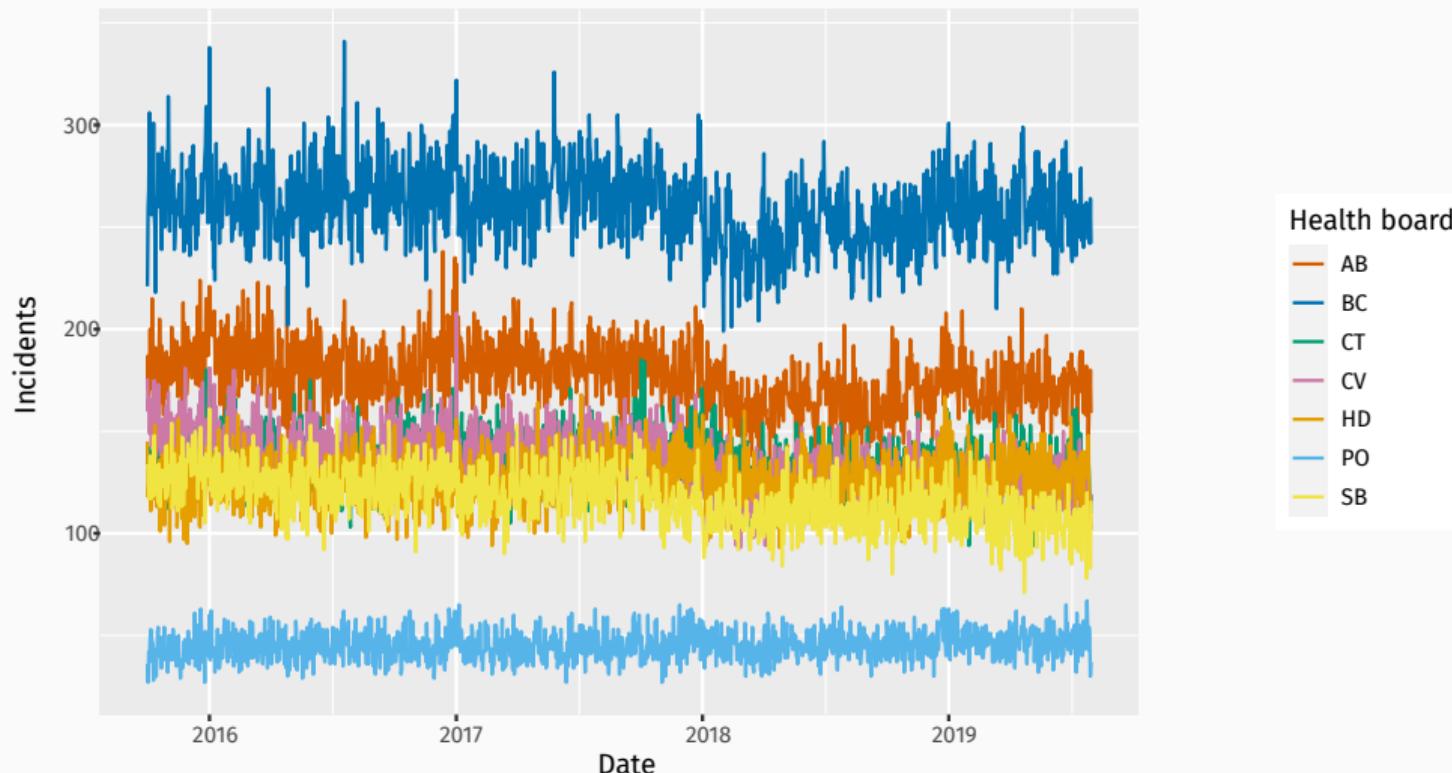
# Aggregated daily incidents



# Daily incidents by control area



# Data incidents by health board



# Data incidents by priority



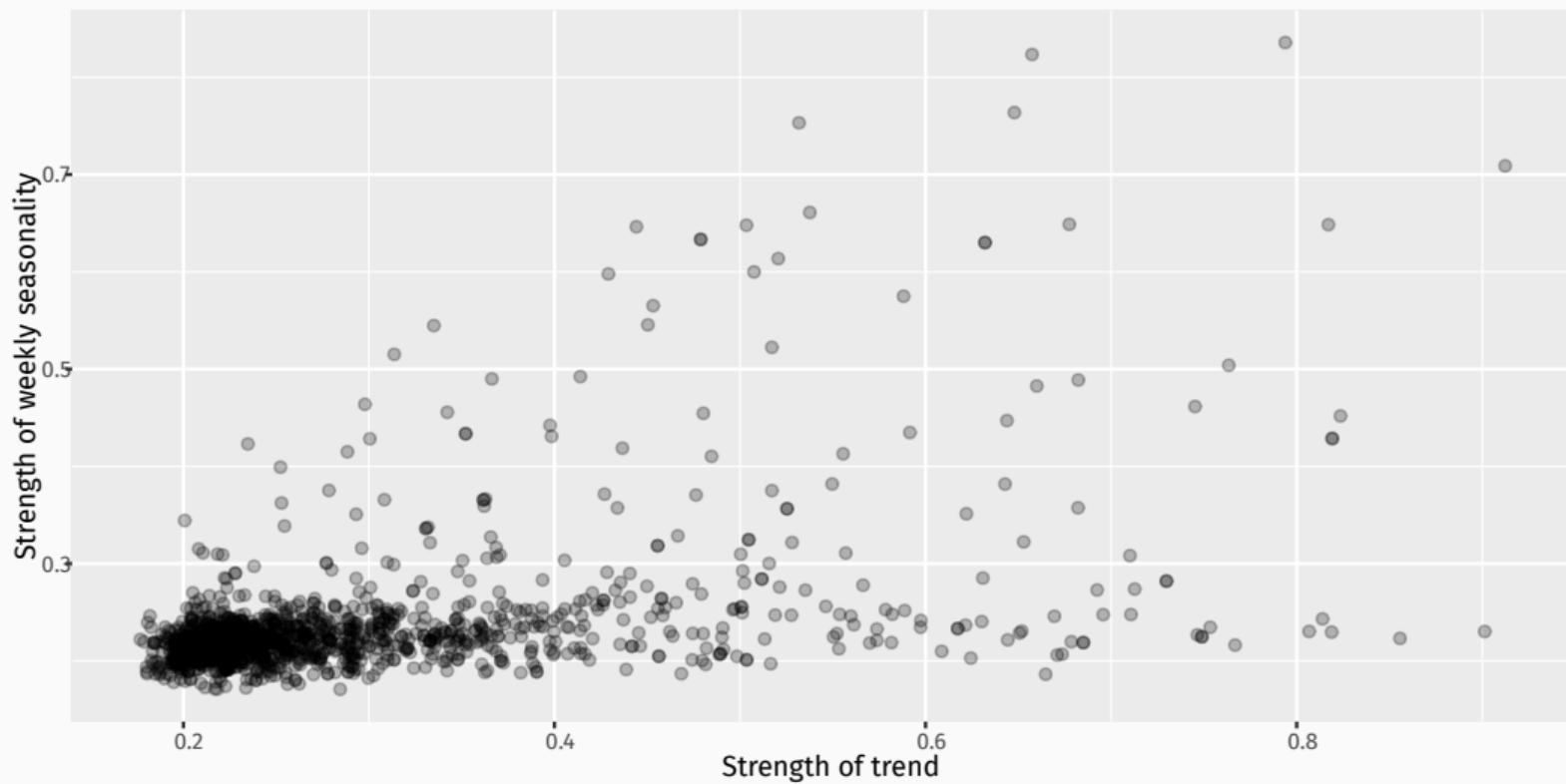
# Data incidents by nature of incident



# Data incidents by nature of incident



# Data features



# Forecasting methods

- 1 **Naïve:** Empirical distribution of past daily attended incidents.
- 2 **ETS:** Exponential Smoothing State Space models.
- 3 **GLM:** Poission Regression with spline trend, day of the week, annual Fourier seasonality, public holidays, school holidays, Christmas Day, New Year's Day.
- 4 **TSGLM:** Poisson Regression with same covariates plus three autoregressive terms.
- 5 **Ensemble:** Mixture distribution of 1-4.

# Forecasting methods

- 1 **Naïve:** Empirical distribution of past daily attended incidents.

$$y_{T+h|T} \sim \text{Empirical}(y_1, \dots, y_T)$$

# Forecasting methods

- 1 **Naïve:** Empirical distribution of past daily attended incidents.

$$y_{T+h|T} \sim \text{Empirical}(y_1, \dots, y_T)$$

- 2 **ETS:** Exponential Smoothing State Space models.

$$y_{T+h|T} \sim \text{Normal}(\hat{y}_{T+h|T}, \hat{\sigma}_{T+h|T}^2)$$

# Forecasting methods

3

## GLM: Poisson Regression

$$y_{T+h|T} \sim \text{Poisson}(\hat{y}_{T+h|T}) \quad \text{where} \quad \hat{y}_{T+h|T} = \exp(\mathbf{x}'_{T+h}\boldsymbol{\beta})$$

and  $\mathbf{x}_{T+h}$  is a vector of covariates including

- spline trend
- day of the week
- annual Fourier seasonality
- public holidays
- school holidays
- Christmas Day
- New Year's Day

# Forecasting methods

	Estimate	Std. Error	z value	Pr(> z )	Significance
(Intercept)	6.998511	0.017412	401.93	< 2e-16	***
Spline_1	0.027859	0.004740	5.88	4.2e-09	***
Spline_2	-0.088244	0.006394	-13.80	< 2e-16	***
Spline_3	-0.075036	0.004784	-15.68	< 2e-16	***
Spline_4	-0.111854	0.010202	-10.96	< 2e-16	***
Spline_5	-0.043009	0.004462	-9.64	< 2e-16	***
Monday	0.019147	0.003174	6.03	1.6e-09	***
Tuesday	-0.016414	0.003180	-5.16	2.4e-07	***
Wednesday	-0.015479	0.003184	-4.86	1.2e-06	***
Thursday	-0.006804	0.003178	-2.14	0.03230	*
Friday	0.012235	0.003156	3.88	0.00011	***
Saturday	0.005293	0.003165	1.67	0.09438	.
Fourier_S1_365	0.005365	0.001294	4.15	3.4e-05	***
Fourier_C1_365	0.008263	0.001263	6.54	6.1e-11	***
Fourier_S2_365	0.004235	0.001271	3.33	0.00086	***
Fourier_C2_365	-0.010510	0.001216	-8.64	< 2e-16	***
Fourier_S3_365	-0.000556	0.001275	-0.44	0.66303	
Fourier_C3_365	0.002650	0.001243	2.13	0.03294	*
Public_holiday	0.033278	0.005697	5.84	5.2e-09	***
School_holiday	0.004857	0.002346	2.07	0.03843	*
Xmas	-0.051902	0.016772	-3.09	0.00197	**
New years day	0.120385	0.015573	7.73	1.1e-14	***

# Forecasting methods

4

## TSGLM: Poisson Regression

$$y_{T+h|T} \sim \text{Poisson}(\hat{y}_{T+h|T})$$

where  $\hat{y}_{T+h|T} = \exp \left( \mathbf{x}'_{T+h} \boldsymbol{\beta} + \sum_{k=1}^3 \alpha_k \log(y_{T+h-k} + 1) \right)$

and  $\mathbf{x}_{T+h}$  is a vector of covariates including

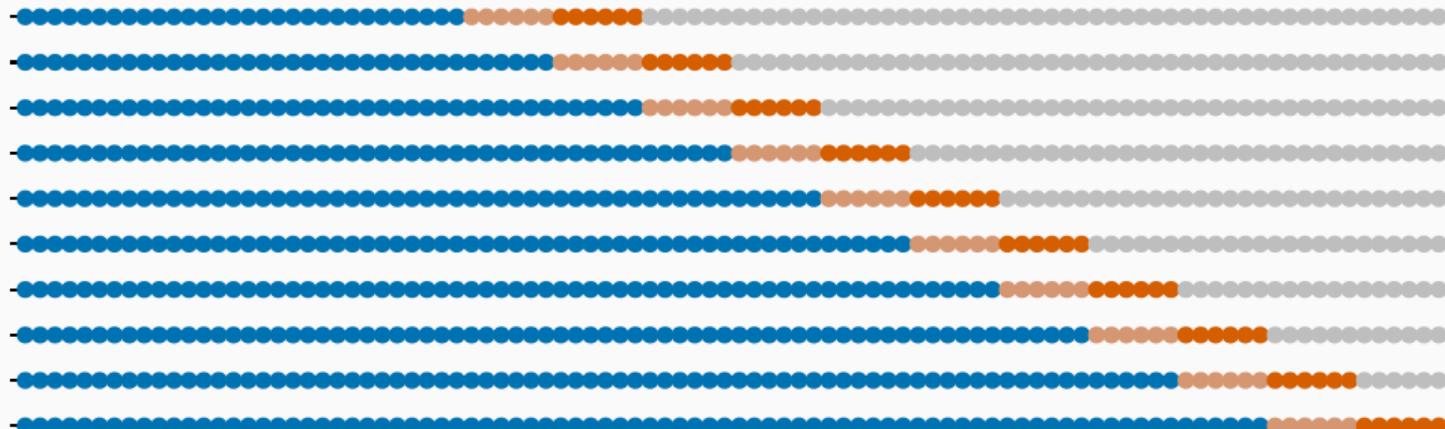
- spline trend
- day of the week
- annual Fourier seasonality
- public holidays
- school holidays
- Christmas Day
- New Year's Day

# Nonparametric bootstrap reconciliation

- Fit model to all series and store the residuals as  $\hat{\varepsilon}_t$ .
- These should be serially uncorrelated but cross-sectionally correlated.
- Draw iid samples from  $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_T$  with replacement.
- Simulate future sample paths for model using the bootstrapped residuals.
- Reconcile each sample path using MinT.
- Combine the reconciled sample paths to form a mixture distribution at each forecast horizon.

# Performance evaluation

- Ten-fold time series cross-validation
- Forecast horizon of 1–84 days
- Each training set contains an additional 42 days.
- Forecasts at 43–84 days correspond to planning horizon.



# Performance evaluation

$$\text{MASE} = \text{mean}(|q_j|)$$

$$q_j = \frac{e_j}{\frac{1}{T-m} \sum_{t=m+1}^T |y_t - y_{t-m}|}$$

- $y_t$  = observation for period  $t$
- $e_j$  = forecast error for forecast horizon  $j$
- $T$  = size of training set
- $m = 7$

# Performance evaluation

$$\text{MSSE} = \text{mean}(q_j^2)$$

$$q_j^2 = \frac{e_j^2}{\frac{1}{T-m} \sum_{t=m+1}^T (y_t - y_{t-m})^2}$$

- $y_t$  = observation for period  $t$
- $e_j$  = forecast error for forecast horizon  $j$
- $T$  = size of training set
- $m = 7$

# Performance evaluation

$$\text{CRPS} = \text{mean}(p_j)$$

$$p_j = \int_{-\infty}^{\infty} (G_j(x) - F_j(x))^2 dx,$$

- $G_j(x)$  = forecast distribution for forecast horizon  $j$
- $F_j(x)$  = true distribution for same period

# Forecast accuracy



# Forecast accuracy: 43–84 days ahead

		MSSE			
Method	Model	Total	Control areas	Health boards	Bottom
Base	Naïve	1.169	1.056	1.062	1.031
Base	ETS	0.979	0.875	0.816	<b>0.975</b>
Base	GLM	0.813	0.897	0.875	1.009
Base	TSGLM	0.822	0.901	0.875	1.050
Base	Ensemble	0.599	0.729	0.774	0.993
MinT	Naïve	1.168	1.057	1.062	2.095
MinT	ETS	0.785	0.852	0.845	0.994
MinT	GLM	0.720	0.827	0.837	1.803
MinT	TSGLM	0.722	0.833	0.839	1.851
MinT	Ensemble	<b>0.560</b>	<b>0.706</b>	<b>0.765</b>	1.557

# Forecast accuracy: 43–84 days ahead

MASE						
Method	Model	Total	Control areas	Health boards	Bottom	
Base	Naïve	1.139	1.059	1.047	1.019	
Base	ETS	0.963	0.930	0.899	1.038	
Base	GLM	0.910	0.940	0.923	<b>1.002</b>	
Base	TSGLM	0.911	0.939	0.924	1.005	
Base	Ensemble	0.782	0.856	0.876	1.008	
MinT	Naïve	1.138	1.059	1.047	2.651	
MinT	ETS	0.877	0.916	0.915	1.289	
MinT	GLM	0.848	0.901	0.902	2.493	
MinT	TSGLM	0.852	0.903	0.903	2.513	
MinT	Ensemble	<b>0.753</b>	<b>0.844</b>	<b>0.872</b>	2.260	

# Forecast accuracy: 43–84 days ahead

CRPS						
Method	Model	Total	Control areas	Health boards	Bottom	
Base	Naïve	30.387	10.882	5.500	0.302	
Base	ETS	14.309	6.074	3.476	0.244	
Base	GLM	15.396	6.253	3.576	0.244	
Base	TSGLM	15.316	6.227	3.575	0.245	
Base	Ensemble	12.978	<b>5.727</b>	3.430	0.243	
MinT	Naïve	30.368	10.902	5.498	0.313	
MinT	ETS	13.515	5.967	3.547	<b>0.243</b>	
MinT	GLM	13.839	5.917	3.453	0.246	
MinT	TSGLM	14.000	5.947	3.455	0.248	
MinT	Ensemble	<b>12.585</b>	5.728	<b>3.426</b>	0.247	

# Conclusions

- Ensemble mixture distributions give better forecasts than any component methods.
- Forecast reconciliation improves forecast accuracy, even when some component methods are quite poor.
- The ensemble without the Naïve method was worse.
- Forecast reconciliation allows coordinated planning and resource allocation.

# Outline

- 1 Definition of probabilistic coherence
- 2 Evaluating probabilistic forecasts
- 3 Emergency Services Demand
- 4 Evaluating multivariate probabilistic forecasts
- 5 Example: Australian electricity generation

# Evaluating probabilistic forecasts

## Continuous Rank Probability Score (univariate forecasts)

Forecast distribution  $F_t$  and observation  $y_t$ .

$$\text{CRPS}(F_t, y_t) = \int_0^1 S_{p,t}(p, y_t) dp = E_F|Y - y_t| - \frac{1}{2}E_F|Y - Y^*|$$

- $Y$  and  $Y^*$  are iid draws from  $F_t$ .
- Optimal when  $F_t$  is truth (i.e., it is a proper score)

# Evaluating probabilistic forecasts

## Continuous Rank Probability Score (univariate forecasts)

Forecast distribution  $F_t$  and observation  $y_t$ .

$$\text{CRPS}(F_t, y_t) = \int_0^1 S_{p,t}(p, y_t) dp = E_F|Y - y_t| - \frac{1}{2}E_F|Y - Y^*|$$

- $Y$  and  $Y^*$  are iid draws from  $F_t$ .
- Optimal when  $F_t$  is truth (i.e., it is a proper score)

## Energy score (multivariate forecasts)

- $\text{ES}(F_t, \mathbf{y}_t) = E_F||\mathbf{Y} - \mathbf{y}_t|| - \frac{1}{2}E_F||\mathbf{Y} - \mathbf{Y}^*||$

# Evaluating probabilistic forecasts

## Continuous Rank Probability Score (univariate forecasts)

Forecast distribution  $F_t$  and observation  $y_t$ .

$$\text{CRPS}(F_t, y_t) = \int_0^1 S_{p,t}(p, y_t) dp = E_F|Y - y_t| - \frac{1}{2}E_F|Y - Y^*|$$

- $Y$  and  $Y^*$  are iid draws from  $F_t$ .
- Optimal when  $F_t$  is truth (i.e., it is a proper score)

## Energy score (multivariate forecasts)

$$\text{ES}(F_t, \mathbf{y}_t) = E_F||\mathbf{Y} - \mathbf{y}_t|| - \frac{1}{2}E_F||\mathbf{Y} - \mathbf{Y}^*||$$

## Log score (multivariate forecasts)

$$\text{LS}(F_t, \mathbf{y}_t) = -\log f(\mathbf{y}_t)$$

# Evaluating probabilistic forecasts

## Proper scoring rule

optimized when true forecast distribution is used.

# Evaluating probabilistic forecasts

## Proper scoring rule

optimized when true forecast distribution is used.

Scoring Rule	Coherent v Incoherent	Coherent v Coherent
Log Score	Not proper	<ul style="list-style-type: none"><li>Ordering preserved if compared using bottom-level only</li></ul>
Energy Score	Proper	<ul style="list-style-type: none"><li>Full hierarchy should be used.</li><li>Rankings may change otherwise.</li></ul>

# Score optimal reconciliation

Algorithm proposed by Panagiotelis et al (2020) for optimizing  $\mathbf{G}$  using stochastic gradient descent to optimize Energy Score.

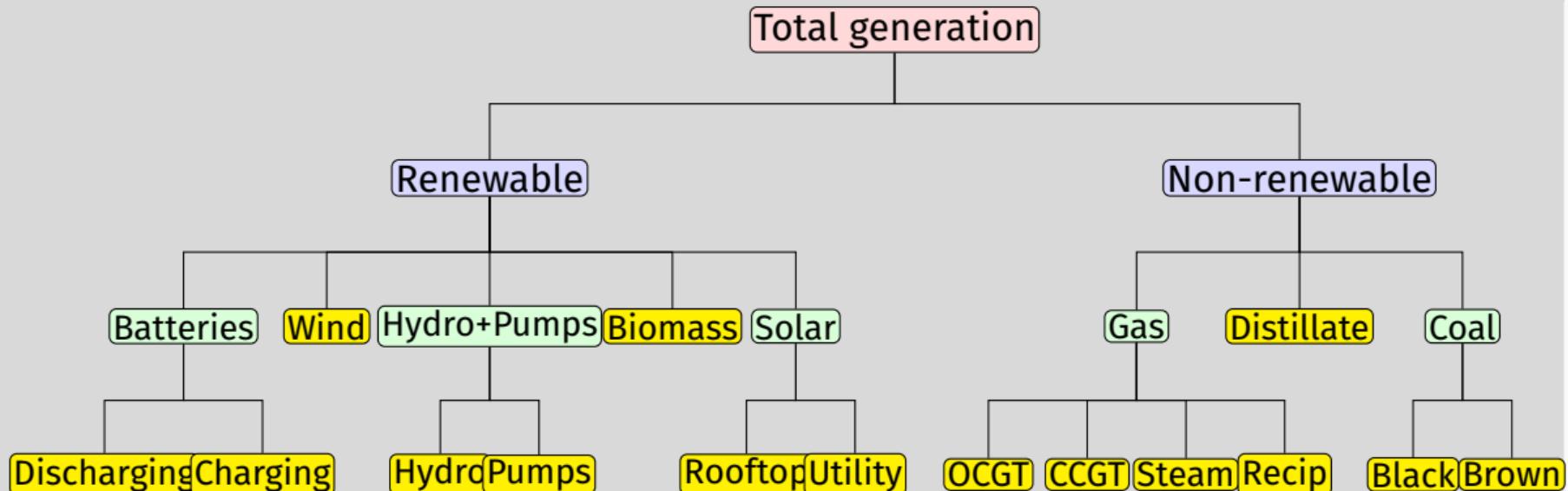
- 1 Compute base forecasts over a test set.
- 2 Compute OLS reconciliation:  $\mathbf{G} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$
- 3 Iteratively update  $\mathbf{G}$  using SGD with Adam method and ES objective over a test set

# Outline

- 1 Definition of probabilistic coherence
- 2 Evaluating probabilistic forecasts
- 3 Emergency Services Demand
- 4 Evaluating multivariate probabilistic forecasts
- 5 Example: Australian electricity generation

# Example: Australian electricity generation

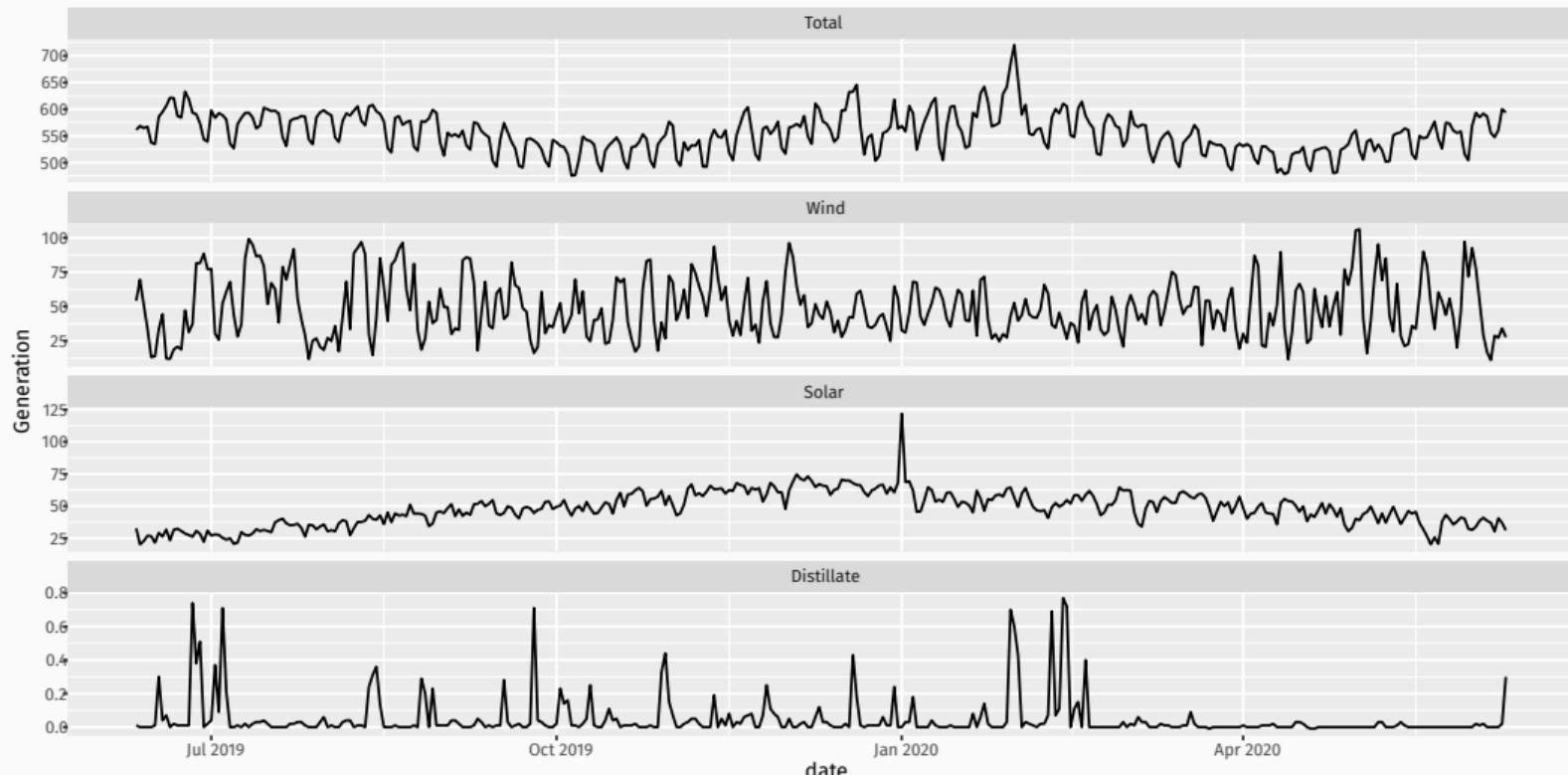
Daily time series from [opennem.org.au](https://opennem.org.au)



$n = 23$  series

$m = 15$  bottom-level series

# Example: Australian electricity generation

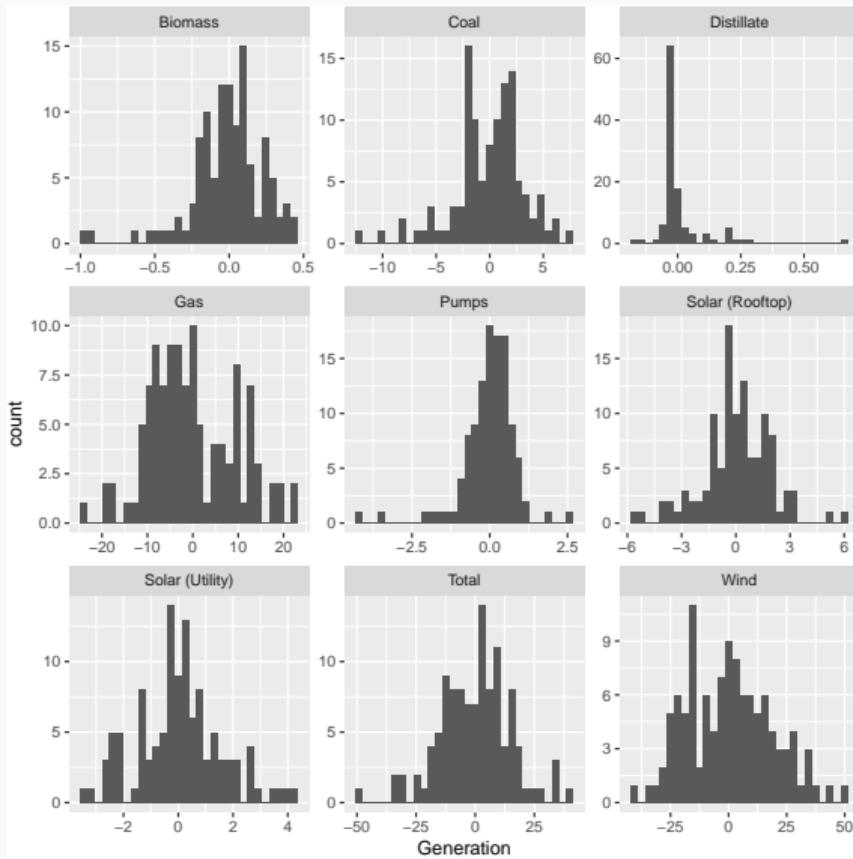


# Example: Australian electricity generation

## Forecast evaluation

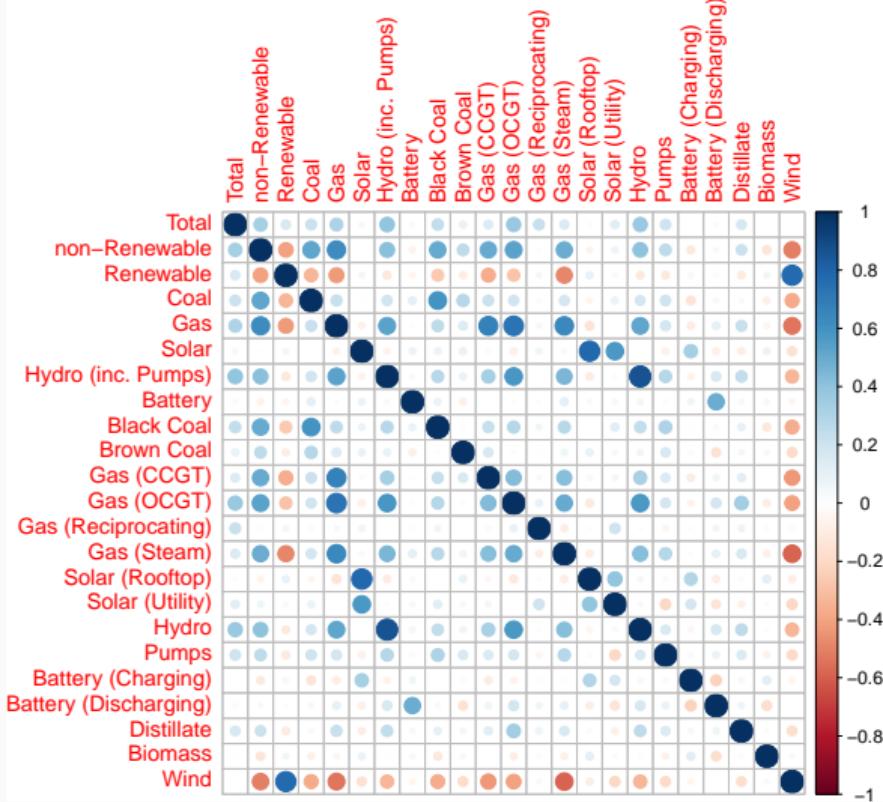
- Rolling window of 140 days training data, and one-step-forecasts for 170 days test data.
- One-layer feed-forward neural network with up to 28 lags of target variable as inputs.
- Implemented using `NNETAR()` function in `fable` package.
- Model could be improved with temperature predictor.

# Example: Australian electricity generation



**Histogram of residuals:  
2 Oct 2019 – 21 Jan 2020**  
Clearly non-Gaussian

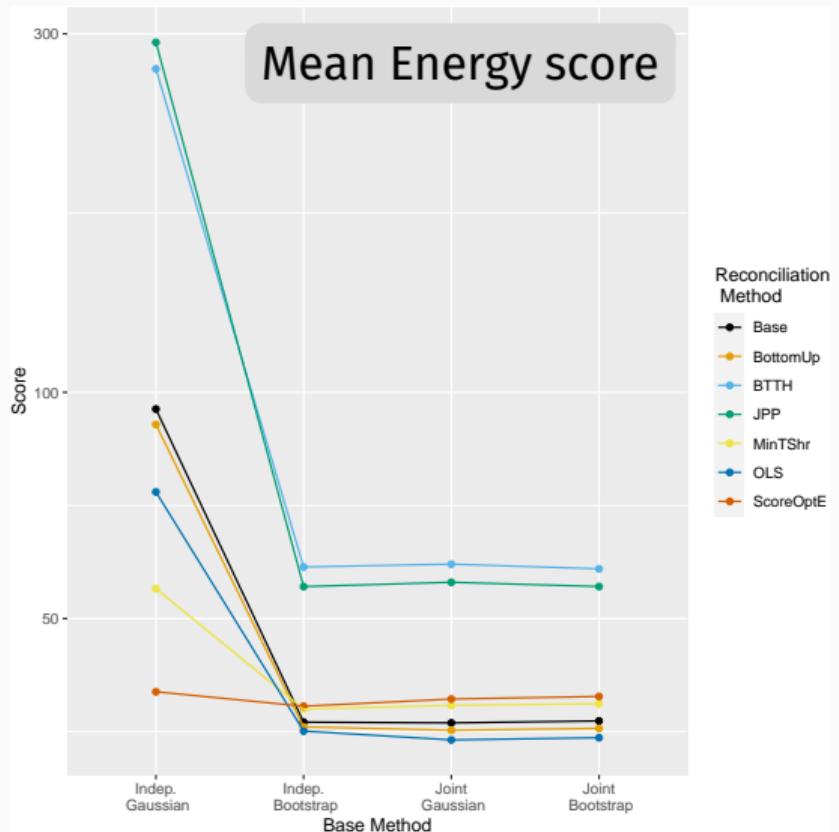
# Example: Australian electricity generation



**Correlations of residuals:**  
2 Oct 2019 – 21 Jan 2020

Blue = positive correlation.  
Red = negative correlation.  
Large = stronger correlations.

# Example: Australian electricity generation



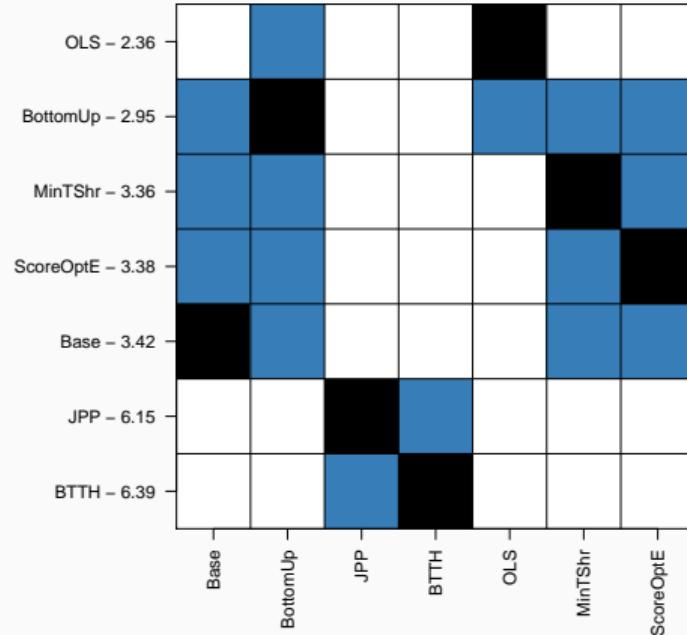
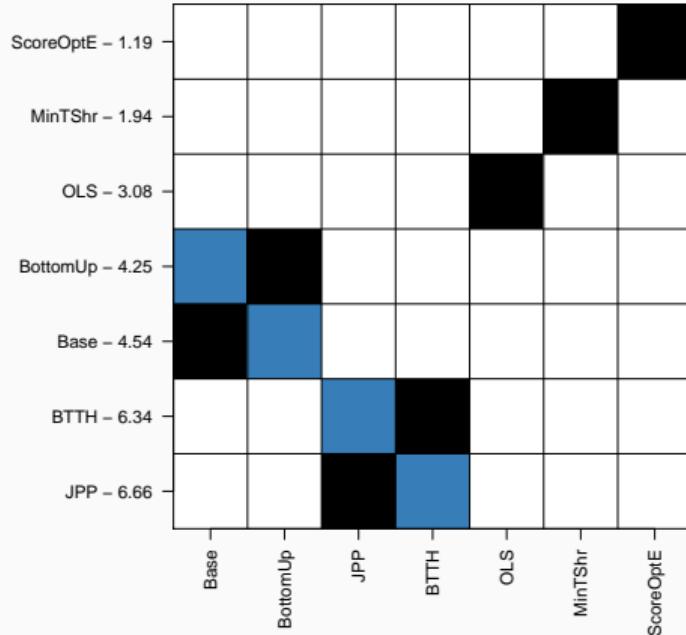
## Base residual assumptions

- Gaussian independent
- Gaussian dependent
- Non-Gaussian independent
- Non-Gaussian dependent

## Reconciliation methods

- Base
- BottomUp
- BTTH: Ben Taieb, Taylor, Hyndman
- JPP: Jeon, Panagiotelis, Petropoulos
- OLS
- MinT(Shrink)
- Score Optimal Reconciliation

# Example: Australian electricity generation



**Nemenyi test for different scores**

Base forecasts are independent and

**Nemenyi test for different scores**

Base forecasts are obtained by jointly

# References

-  Ben Taieb, S, JW Taylor, and RJ Hyndman (2021). Hierarchical Probabilistic Forecasting of Electricity Demand with Smart Meter Data. *J American Statistical Association* **116**(533), 27–43.
-  Panagiotelis, A, P Gamakumara, G Athanasopoulos, and RJ Hyndman (2023). Probabilistic forecast reconciliation: properties, evaluation and score optimisation. *European J Operational Research* **306**(2), 693–706.