Forecast reconciliation

3. Probabilistic forecast reconciliation

Rob J Hyndman

MONASH University

Photo by Edvard Alexander Rølvaag on Unsplash

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

The coherent subspace

Coherent subspace

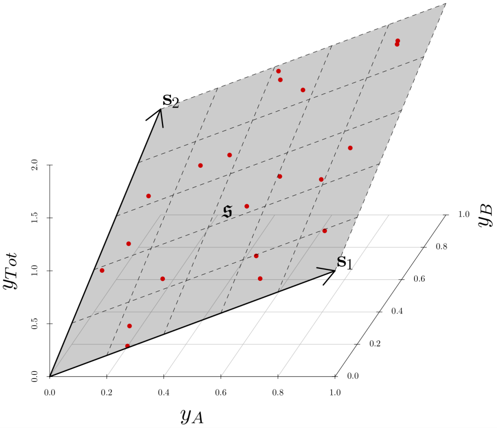
m-dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

Hierarchical time series

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

Coherent point forecasts

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



$y_{Tot} = y_A + y_B$

The coherent subspace

Coherent subspace

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

Hierarchical time series

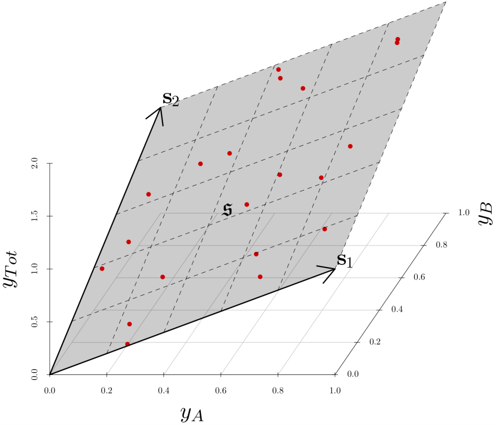
An n-dimensional multivariate time series such that $\boldsymbol{y}_t \in \mathfrak{s} \quad \forall t$.

Coherent point forecasts

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

Base forecasts

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



$y_{Tot} = y_A + y_B$

The coherent subspace

Coherent subspace

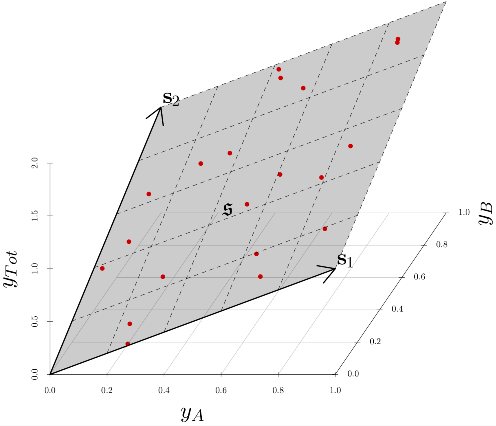
m-dimensional linear subspace $\mathfrak{s} \subset \chi^n$ for which linear constraints hold for all $\boldsymbol{y} \in \mathfrak{s}$.

Hierarchical time series

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

Coherent point forecasts

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



$y_{Tot} = y_A + y_B$

Base forecasts

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

Reconciled forecasts

Let ψ be a mapping, $\psi : \chi^n \rightarrow \mathfrak{s}$.
$\tilde{\boldsymbol{y}}_{t+h|t} = \psi(\hat{\boldsymbol{y}}_{t+h|t})$ "reconciles" $\hat{\boldsymbol{y}}_{t+h|t}$.

Coherent probabilistic forecasts

Coherent probabilistic forecasts

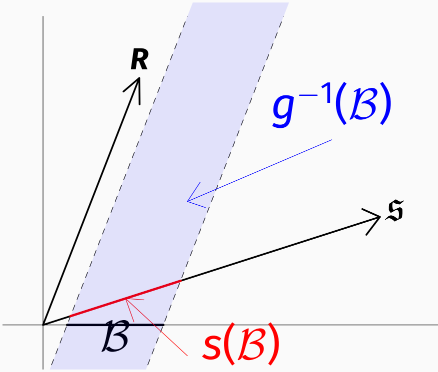
A probability triple $(\mathfrak{s}, \mathscr{F}_{\mathfrak{s}}, \breve{\nu})$ is coherent with the bottom probability triple $(\chi^m, \mathscr{F}_{\chi^m}, \nu)$, if

$$\breve{\nu}(s(\mathcal{B})) = \nu(\mathcal{B}) \quad \forall \mathcal{B} \in \mathscr{F}_{\chi^m}$$

i.e., probability of any point not on $\mathfrak{s}$ is zero.

Probabilistic forecast reconciliation

Let $(\chi^n, \mathscr{F}_{\chi^n}, \hat{\nu})$ be the base forecast. Then the reconciled probability distribution $\breve{\nu}$ is a transformation of $\hat{\nu}$ that is coherent on $\mathscr{F}_{\mathfrak{s}}$.



$\psi = s \circ g$

Panagiotelis et al. (2023) and Corani, Azzimonti, and Rubattu (2023)

Construction of reconciled distributions

Reconciled density of bottom-level

Density of bottom-level series under reconciled distribution is

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

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

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

Construction of reconciled distributions

Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_{\boldsymbol{y}}(\boldsymbol{y}) = |\boldsymbol{S}^*| \tilde{f}_{\boldsymbol{b}}(\boldsymbol{S}^-\boldsymbol{y}) \, \mathbb{1}\{\boldsymbol{y} \in \mathfrak{s}\}$$

- $\boldsymbol{S}^* = \begin{pmatrix} \boldsymbol{S}^{-\prime} & \boldsymbol{S}_\perp \end{pmatrix}^\prime$
- $\boldsymbol{S}^-$ is $m \times n$ generalised inverse of $\boldsymbol{S}$ such that $\boldsymbol{S}^-\boldsymbol{S} = \boldsymbol{I}$
- $\boldsymbol{S}_\perp$ is $n \times (n - m)$ orthogonal complement to $\boldsymbol{S}$: $\boldsymbol{S}_\perp^\prime \boldsymbol{S} = \boldsymbol{0}$

Construction of reconciled distributions

Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_{\boldsymbol{y}}(\boldsymbol{y}) = |\boldsymbol{S}^*|\, \tilde{f}_{\boldsymbol{b}}(\boldsymbol{S}^-\boldsymbol{y})\, \mathbb{1}\{\boldsymbol{y} \in \mathfrak{s}\}$$

- $\boldsymbol{S}^* = \begin{pmatrix}\boldsymbol{S}^{-\prime} & \boldsymbol{S}_\perp\end{pmatrix}'$
- $\boldsymbol{S}^-$ is $m \times n$ generalised inverse of $\boldsymbol{S}$ such that $\boldsymbol{S}^-\boldsymbol{S} = \boldsymbol{I}$
- $\boldsymbol{S}_\perp$ is $n \times (n-m)$ orthogonal complement to $\boldsymbol{S}$: $\boldsymbol{S}_\perp'\boldsymbol{S} = \boldsymbol{0}$

Gaussian reconciliation

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

Construction of reconciled distributions

Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_{\boldsymbol{y}}(\boldsymbol{y}) = |\boldsymbol{S}^*|\,\tilde{f}_{\boldsymbol{b}}(\boldsymbol{S}^-\boldsymbol{y})\,\mathbb{1}\{\boldsymbol{y} \in \mathfrak{s}\}$$

- $\boldsymbol{S}^* = \begin{pmatrix}\boldsymbol{S}^{-\prime} & \boldsymbol{S}_\perp\end{pmatrix}'$
- $\boldsymbol{S}^-$ is $m \times n$ generalised inverse of $\boldsymbol{S}$ such that $\boldsymbol{S}^-\boldsymbol{S} = \boldsymbol{I}$
- $\boldsymbol{S}_\perp$ is $n \times (n-m)$ orthogonal complement to $\boldsymbol{S}$: $\boldsymbol{S}_\perp'\boldsymbol{S} = \boldsymbol{0}$

Gaussian reconciliation

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

Elliptical reconciliation

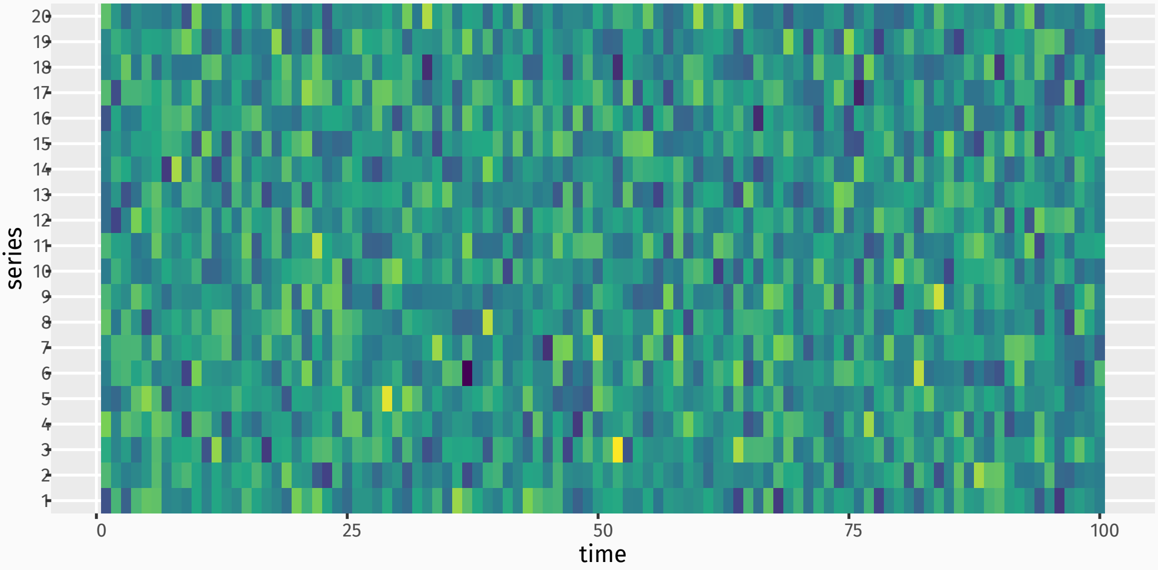
An elliptical distribution is obtained if the base forecasts are elliptical, unbiased, and $\text{rank}(\hat{\Sigma} - \Sigma) \le n_a$

Simulation from a reconciled distribution

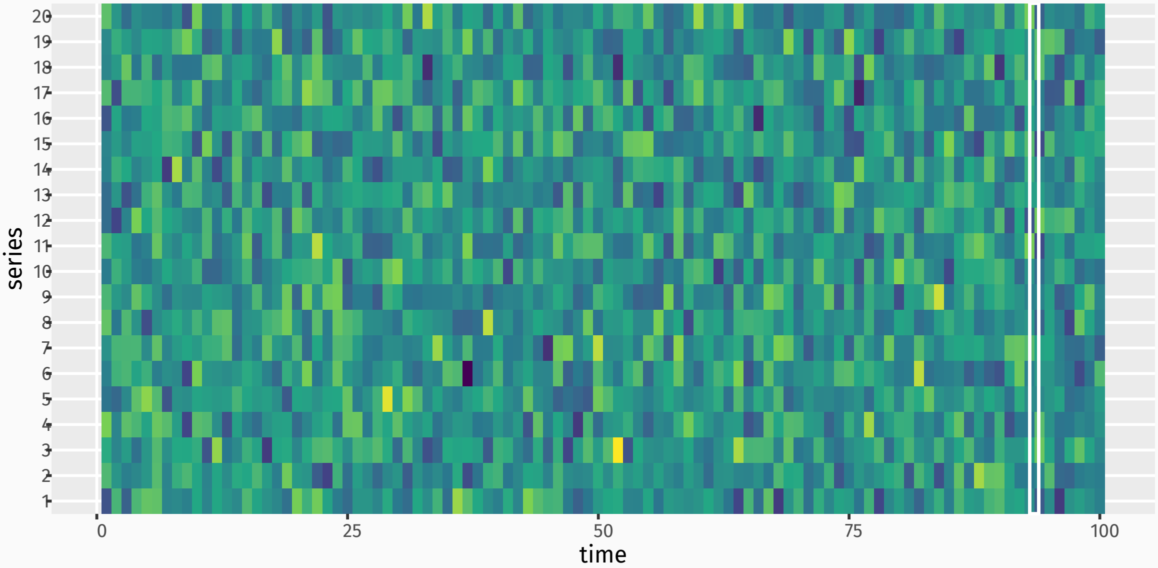
Suppose that $(\hat{\boldsymbol{y}}^{[1]}, \dots, \hat{\boldsymbol{y}}^{[L]})$ is a sample drawn from an incoherent probability measure $\hat{\nu}$. Then $(\tilde{\boldsymbol{y}}^{[1]}, \dots, \tilde{\boldsymbol{y}}^{[L]})$ where $\tilde{\boldsymbol{y}}^{[\ell]} := \psi(\hat{\boldsymbol{y}}^{[\ell]})$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

- Simulate future sample paths for each series, by simulating from each model using a multivariate bootstrap of the residuals (to preserve cross-correlations).
- Reconcile the sample paths.
- The reconciled sample paths are a sample from the reconciled distribution.

Simulation from a reconciled distribution

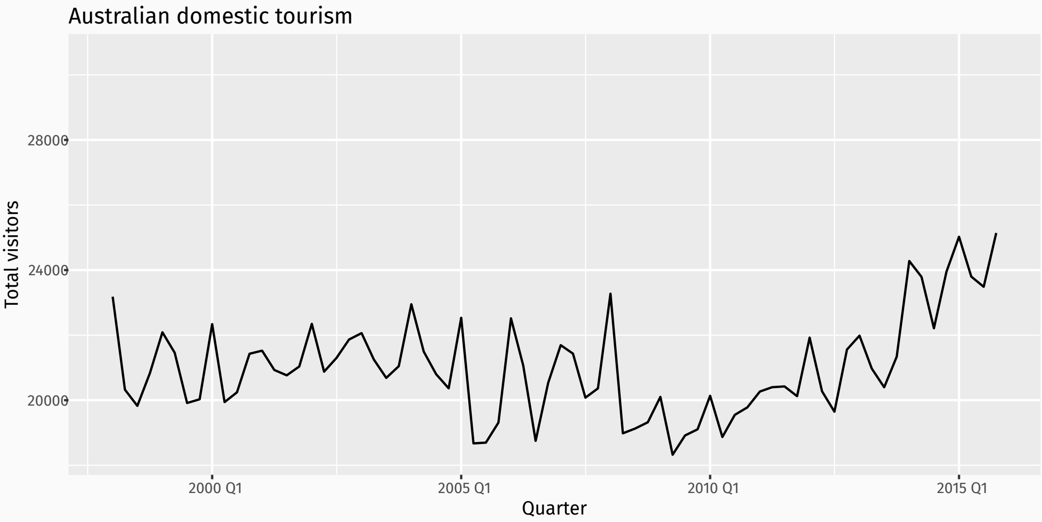


Simulation from a reconciled distribution

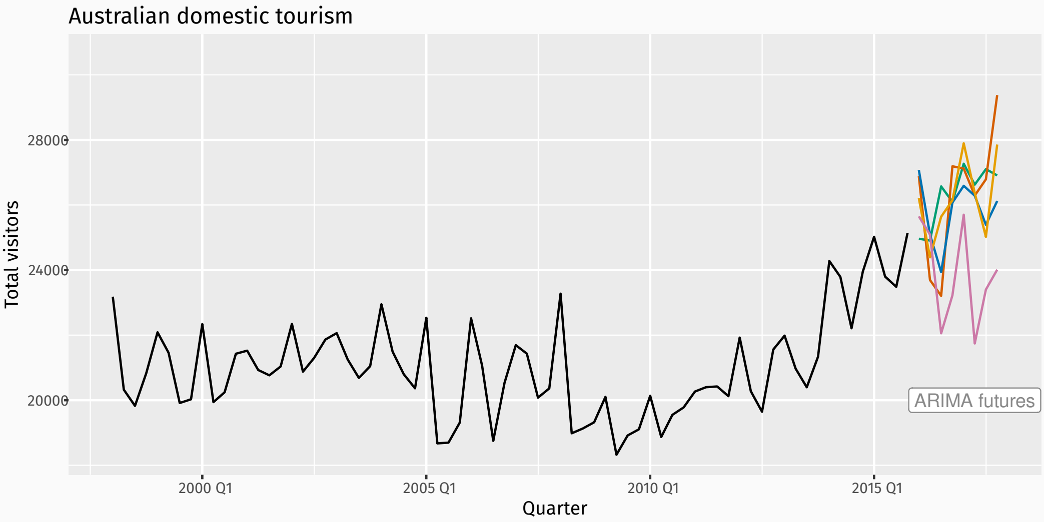


Outline

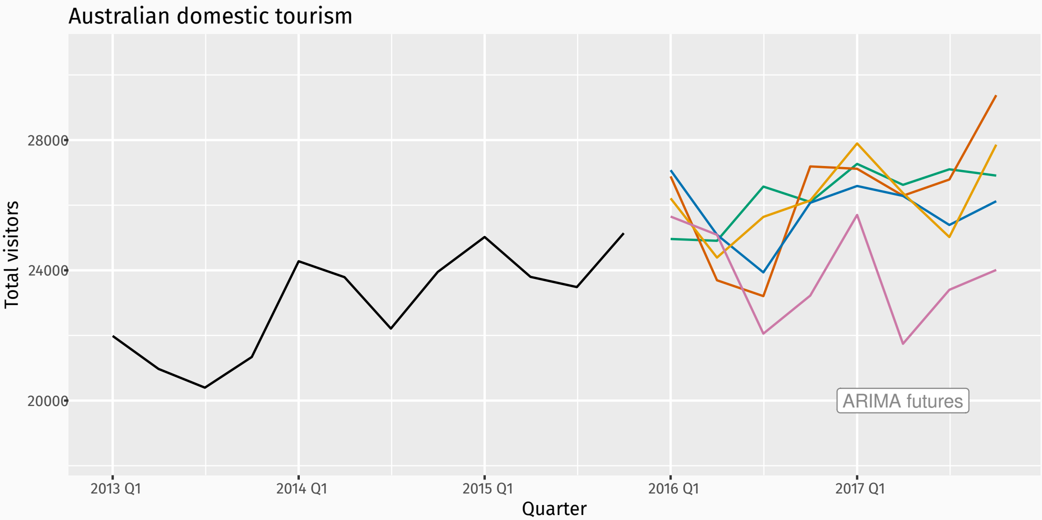
Evaluating probabilistic forecasts



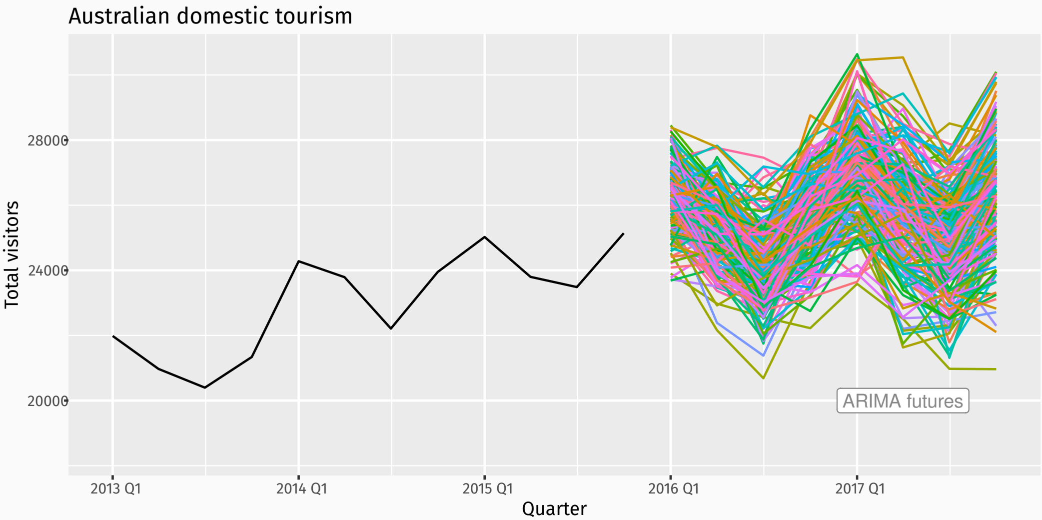
Evaluating probabilistic forecasts



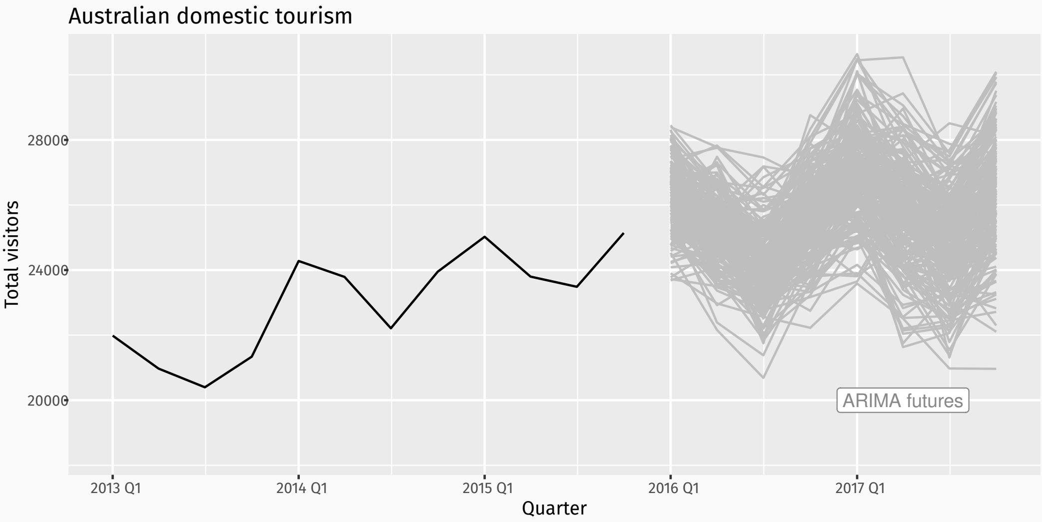
Evaluating probabilistic forecasts



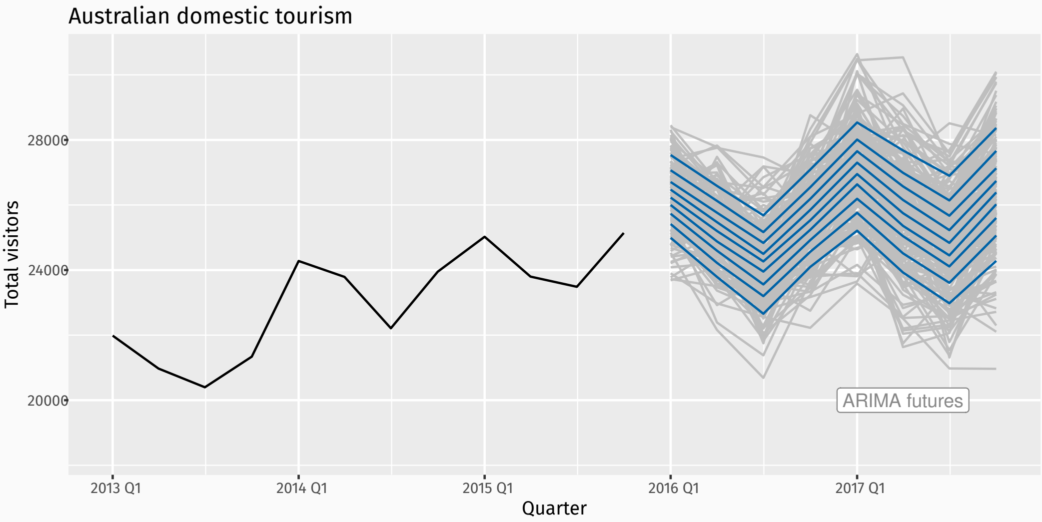
Evaluating probabilistic forecasts



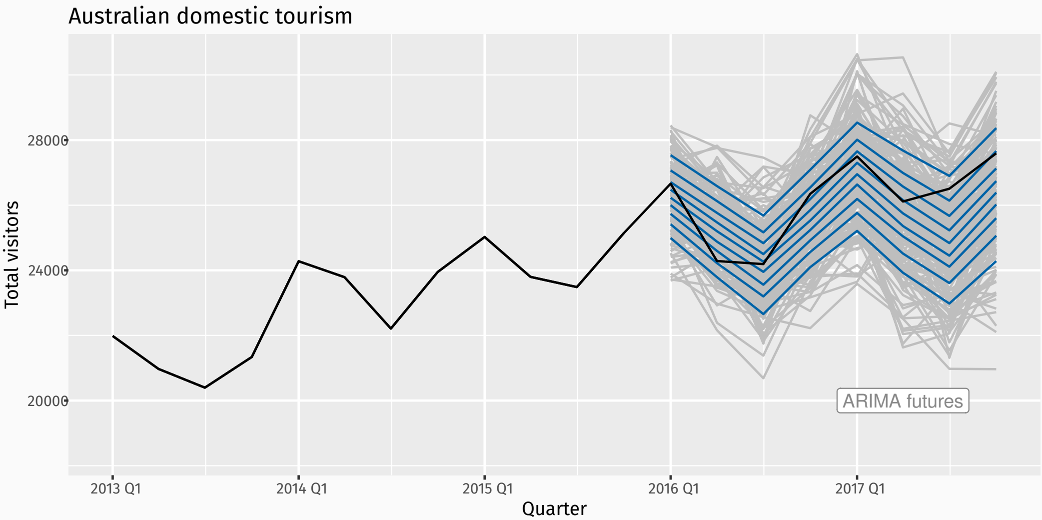
Evaluating probabilistic forecasts



Evaluating probabilistic forecasts



Evaluating probabilistic forecasts



Evaluating probabilistic forecasts

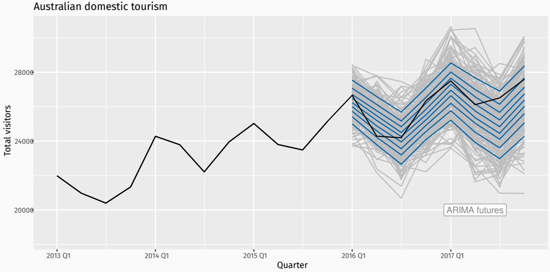
y_t = observation at time t

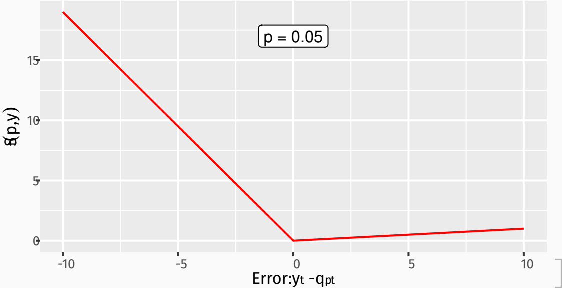
$q_{p,t}$ = quantile forecast: prob. p, time t

Quantile score

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

- 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)

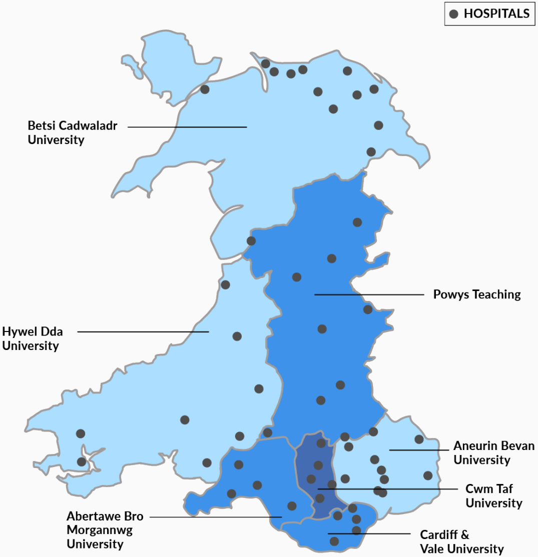




Outline

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

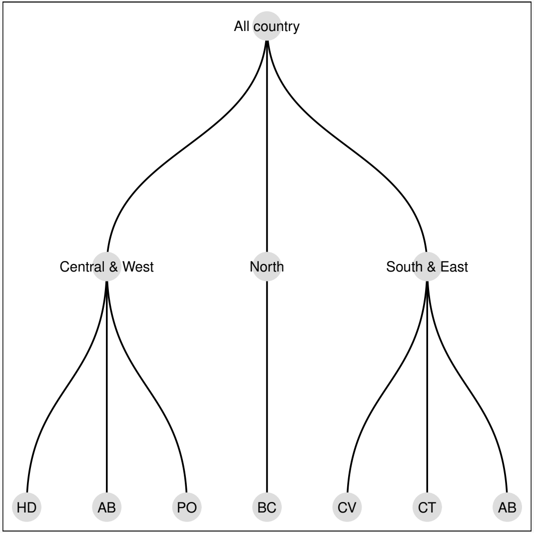
Wales Health Board Areas



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
Red
Amber
Green

*

Nature of incident
Chest pain
Stroke
Breathing problem
...
Abdominal pain

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

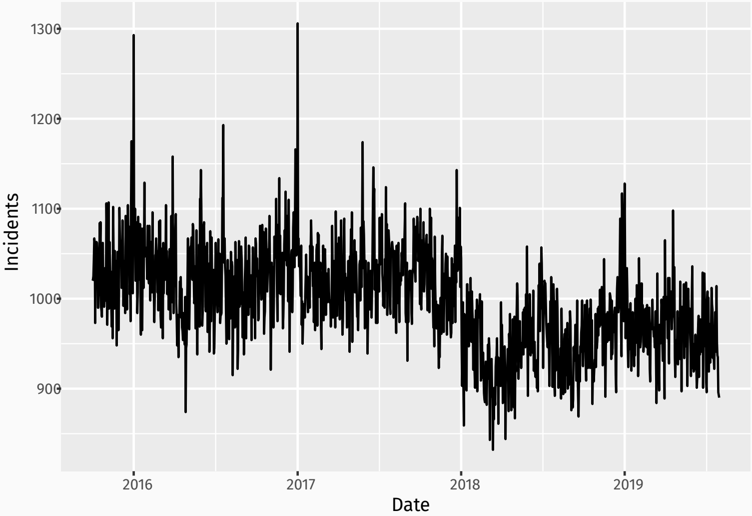
Data

```
# 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

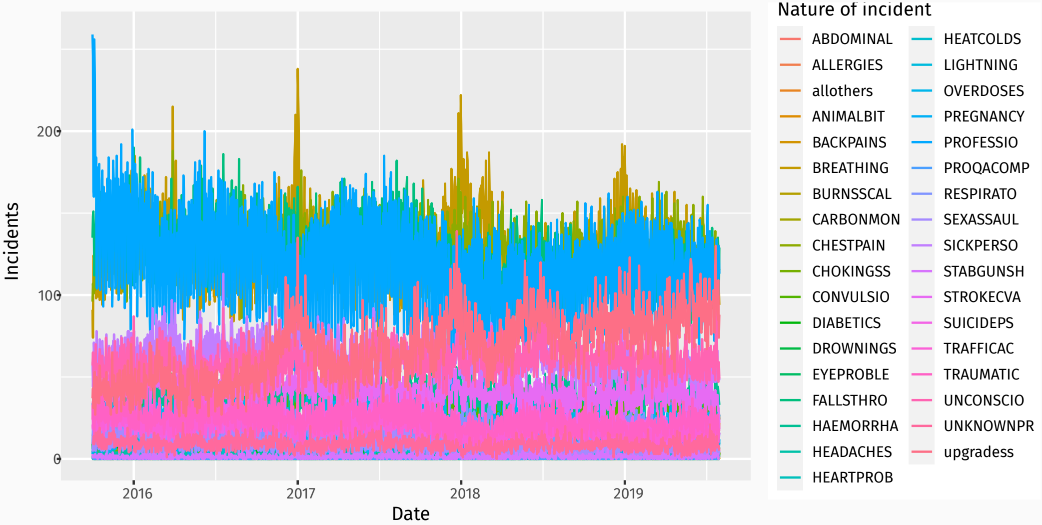


Daily incidents by control area

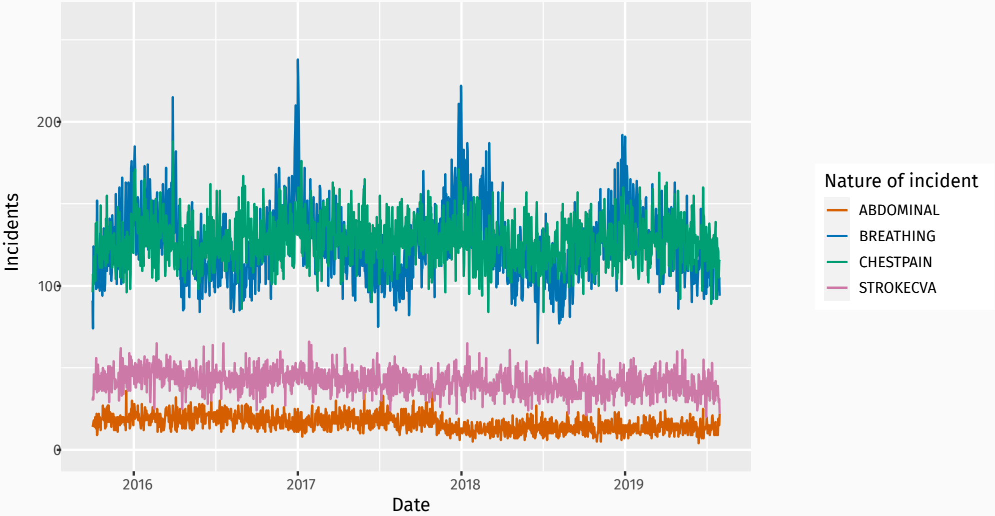
Data incidents by health board

Data incidents by priority

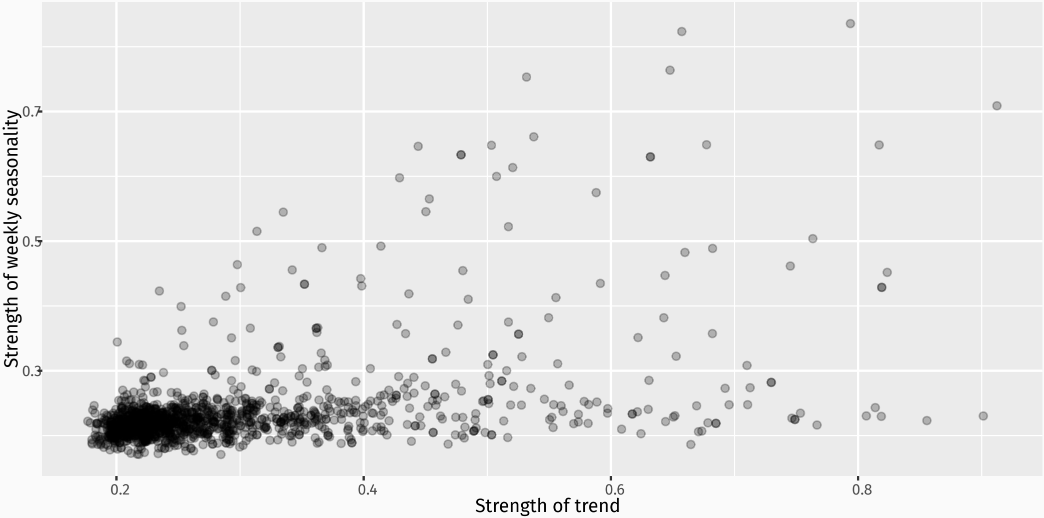
Data incidents by nature of incident



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.

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

Forecasting methods

3 **GLM**: Poission Regression

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

and $\boldsymbol{x}_{T+h}$ 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|)
(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 ***
---
```

Significance

***	$p < 0.001$
**	$p < 0.01$
*	$p < 0.05$
.	$p < 0.1$

Forecasting methods

4 **TSGLM**: Poisson Regression

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

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

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

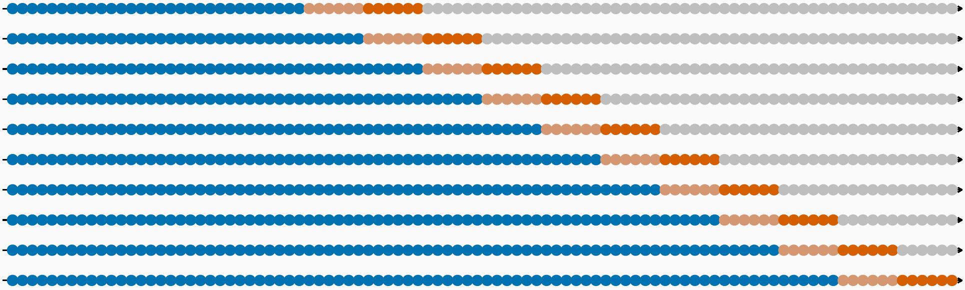
- spline trend
- day of the week
- annual Fourier seasonality
- public holidays
- school holidays
- Christmas Day
- New Year's Day

Nonparametric bootstrap reconciliation

- Fit model to all series and store the residuals as $\underset{\sim}{\varepsilon}_t$.
- These should be serially uncorrelated but cross-sectionally correlated.
- Draw iid samples from $\underset{\sim}{\varepsilon}_1, \dots, \underset{\sim}{\varepsilon}_T$ 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

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

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

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

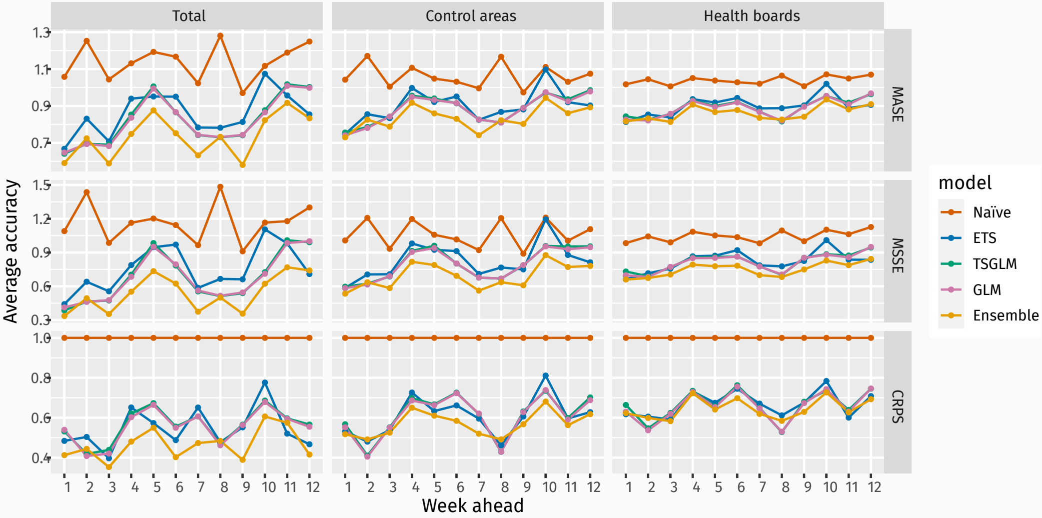
Performance evaluation

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

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

- $G_j(x)$ = 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	**0.975**
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	**0.560**	**0.706**	**0.765**	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	**1.002**
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	**0.753**	**0.844**	**0.872**	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	**5.727**	3.430	0.243
MinT	Naïve	30.368	10.902	5.498	0.313
MinT	ETS	13.515	5.967	3.547	**0.243**
MinT	GLM	13.839	5.917	3.453	0.246
MinT	TSGLM	14.000	5.947	3.455	0.248
MinT	Ensemble	**12.585**	5.728	**3.426**	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

Evaluating probabilistic forecasts

Continuous Rank Probability Score (univariate forecsts)

Forecast distribution F_t and observation y_t.

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

- 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 forecsts)

Forecast distribution F_t and observation y_t.

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

- 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)

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

Evaluating probabilistic forecasts

Continuous Rank Probability Score (univariate forecsts)

Forecast distribution F_t and observation y_t.

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

- 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)

- $\text{ES}(F_t, \boldsymbol{y}_t) = \text{E}_F||\boldsymbol{Y} - \boldsymbol{y}_t|| - \tfrac{1}{2}\text{E}_F||\boldsymbol{Y} - \boldsymbol{Y}^*||$

Log score (multivariate forecasts)

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

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	• Ordering preserved if compared using bottom-level only
Energy Score	Proper	• Full hierarchy should be used. • Rankings may change otherwise.

Score optimal reconciliation

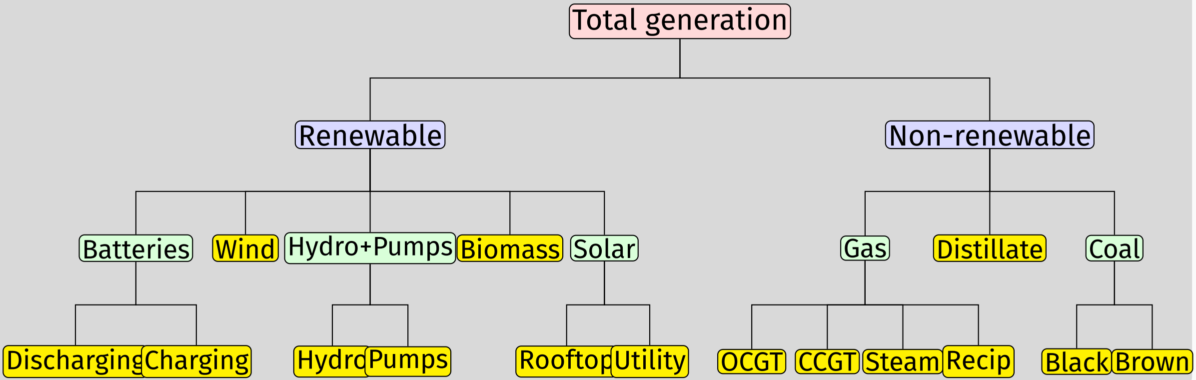
Algorithm proposed by Panagiotelis et al (2020) for optimizing $\boldsymbol{G}$ using stochastic gradient descent to optimize Energy Score.

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

Outline

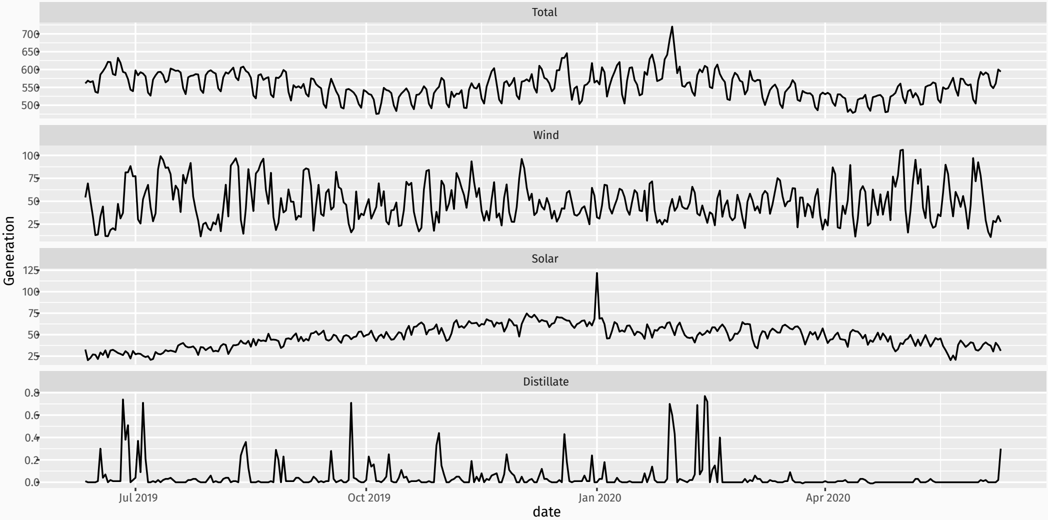
Example: Australian electricity generation

Daily time series from opennem.org.au



n = 23 series m = 15 bottom-level series

Example: Australian electricity generation

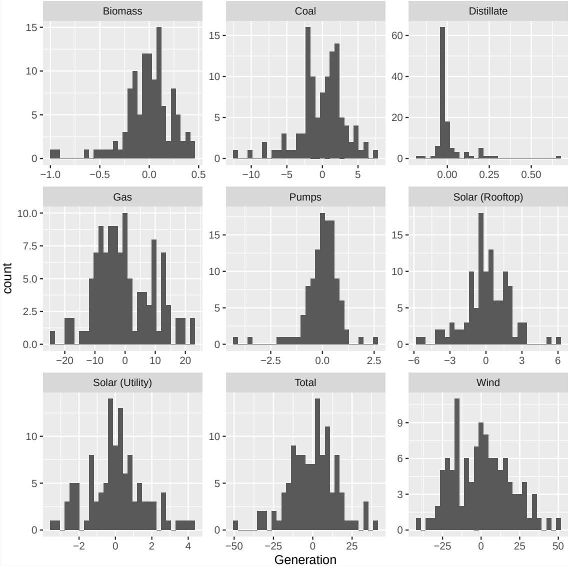


Example: Australian electricity generation

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.

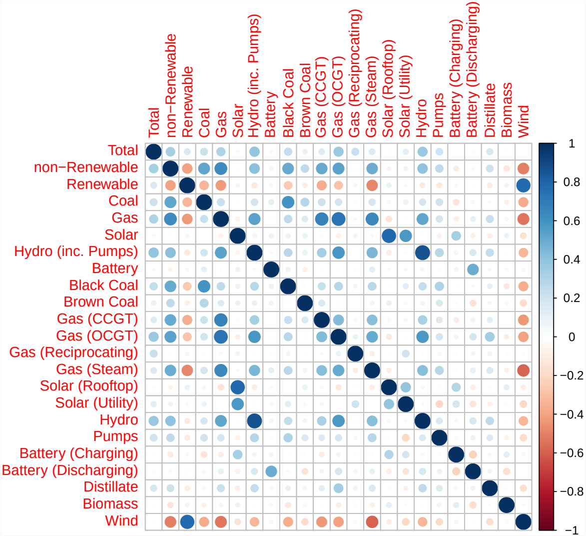
Example: Australian electricity generation



Histogram of residuals: 2 Oct 2019 – 21 Jan 2020

Clearly non-Gaussian

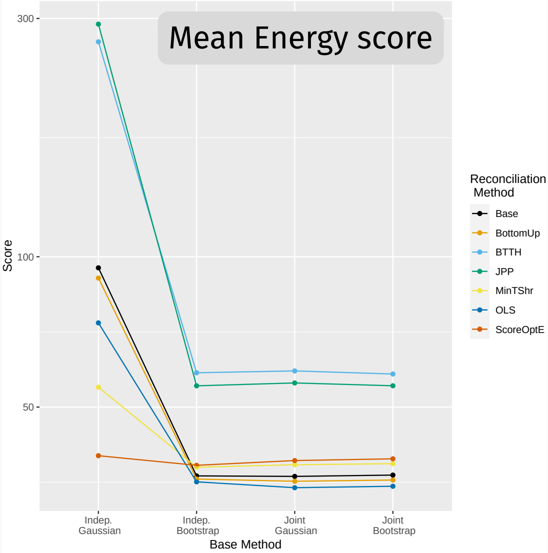
Example: Australian electricity generation



Correlations of residuals: 2 Oct 2019 – 21 Jan 2020

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

Example: Australian electricity generation



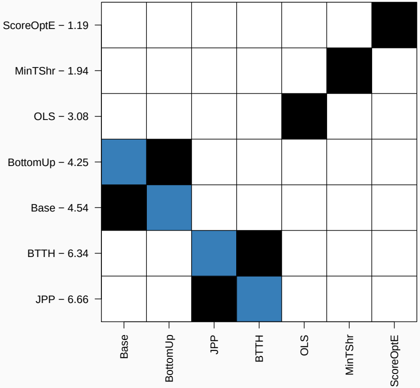
Base residual assumptions

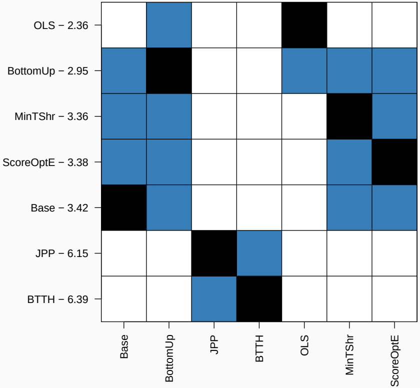
- Gaussian independent
- Gaussian dependent
- Non-Gaussian independent
- Non-Gaussian dependent

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 Gaussian.

Nemenyi test for different scores

Base forecasts are obtained by jointly bootstrapping residuals.

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.
- Corani, G, D Azzimonti, and N Rubattu (2023). Probabilistic reconciliation of count time series. *International Journal of Forecasting*. to appear.
- 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.
- Rostami-Tabar, B and RJ Hyndman (2023). "Hierarchical Time Series Forecasting in Emergency Medical Services". robjhyndman.com/publications/fem.