COMBINING AND POOLING FORECASTS BASED ON SELECTION CRITERIA

DEVON BARROW | SCHOOL OF STRATEGY AND LEADERSHIP | COVENTRY UNIVERSITY

NIKOLAOS KOURENTZES, | DEPARTMENT OF MANAGEMENT SCIENCE | LANCASTER UNIVERSITY

FOTIOS PETROPOULOS | CARDIFF UNIVERSITY

OUTLINE

- M3 Competition data
- Selection and combination - The state of play
- Comparing selection methods: information criteria versus cross-validation
- Comparing combination methods based on selection approaches
- Pooling forecasts - The state of play
- Comparing methods of pooling – a new approach to pooling

Time interval between successive observations	Types of time series data						
	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57		756
Monthly	474	334	312	145	111	52	1428
Other	4			29		141	174
Total	828	519	731	308	413	204	3003

M3-COMPETITION DATASET

Frequency	# of time series	Forecast horizon	# of forecasts	Max # of observations
Yearly	645	6	8	41
Quarterly	756	8	19	64
Monthly	1428	18	19	126
Other	174	8	8	96

EXPERIMENT DESIGN

SELECTION – THE STATE OF PLAY

- Selecting forecasts for planning (inventory, scheduling, etc.):
  - *Hold-out set* - Withhold a set of observations and measure performance using MAE, MSE, MASE, MAPE, etc.. (Our preferred approach)
  - *Cross-validation* (rolling origin) – extension of the *Hold-out* method using multiple origins and lead times and measure performance using MAE, MSE, MASE, MAPE, etc.. (Our preferred approach)
  - *Information criteria*: Akaike Information Criteria, (Akakie, 1974) and Bayesian Information Criterion (Schwarz, 1978)

Agreement between AIC and BIC in selecting models

Family of Models	Yearly	Quarterly	Monthly
Exponential Smoothing	86.5%	72.7%	57.1%
ARIMA	[illegible]	[illegible]%	38.7%
No. of Series	[illegible]	[illegible]	1428

Selection uncertainty = selection anxiety

The correct identification of the ***best model*** for each series individually may lead to significant accuracy improvements (Fildes, 2001), in some cases up to 20-30% (Fildes and Petropoulos, 2014).

Spot-on selection is difficult: Model selection often does not outperform the performance of a single model (Fildes, 2001; Theta model, Assimakopoulos and Nikolopoulos, 2000; Hyndman et al., 2002).

RESULTS ON SELECTION METHODS

Mean				
	Annual	Quarterly	Monthly	Other
AIC	18.86%	9.89%	14.38%	**4.28%**
BIC	17.86%	**9.77%**	**14.29%**	4.37%
HQ	18.22%	9.84%	**14.29%**	4.32%
Validation	**17.85%**	9.95%	14.79%	4.56%

Median				
	Annual	Quarterly	Monthly	Other
AIC	9.68%	4.57%	**6.78%**	**1.91%**
BIC	**9.36%**	4.75%	6.86%	1.97%
HQ	9.47%	4.60%	**6.78%**	1.95%
Validation	9.53%	**4.39%**	6.95%	1.97%

COMBINATION – THE STATE OF PLAY

- Combining forecasts for planning (inventory, scheduling, etc.):
  - *Equal weights:* all forecasts are weighted equally
  - *Simple weights:* weights are normalised based on based on the size of the criteria used e.g. MAE, MSE, MASE, MAPE, etc.. (Our preferred approach)
  - *Information criteria:* so called Akaike Weights (Kolassa, 2011). Calculated based on AIC differences, Δ_{AIC} and likelihood relative to minimal model given as $\exp(-\frac{1}{2}\Delta_{AIC}(M))$

RESULTS ON COMBINATION METHODS

Mean				
	Annual	Quarterly	Monthly	Other
Equal weights	**16.65%**	9.64%	15.20%	4.38%
Akaike weights	17.59%	**9.59%**	**14.06%**	**4.29%**

Median				
	Annual	Quarterly	Monthly	Other
Equal weights	9.20%	4.40%	6.91%	2.03%
Akaike weights	**9.14%**	**4.37%**	**6.68%**	**1.92%**

RESULTS ON SELECTION VERSUS COMBINATION

Mean				
	Annual	Quarterly	Monthly	Other
Best selection	17.85%	9.77%	14.79%	4.28%
Best combination	**16.65%**	**9.59%**	**14.06%**	**4.29%**

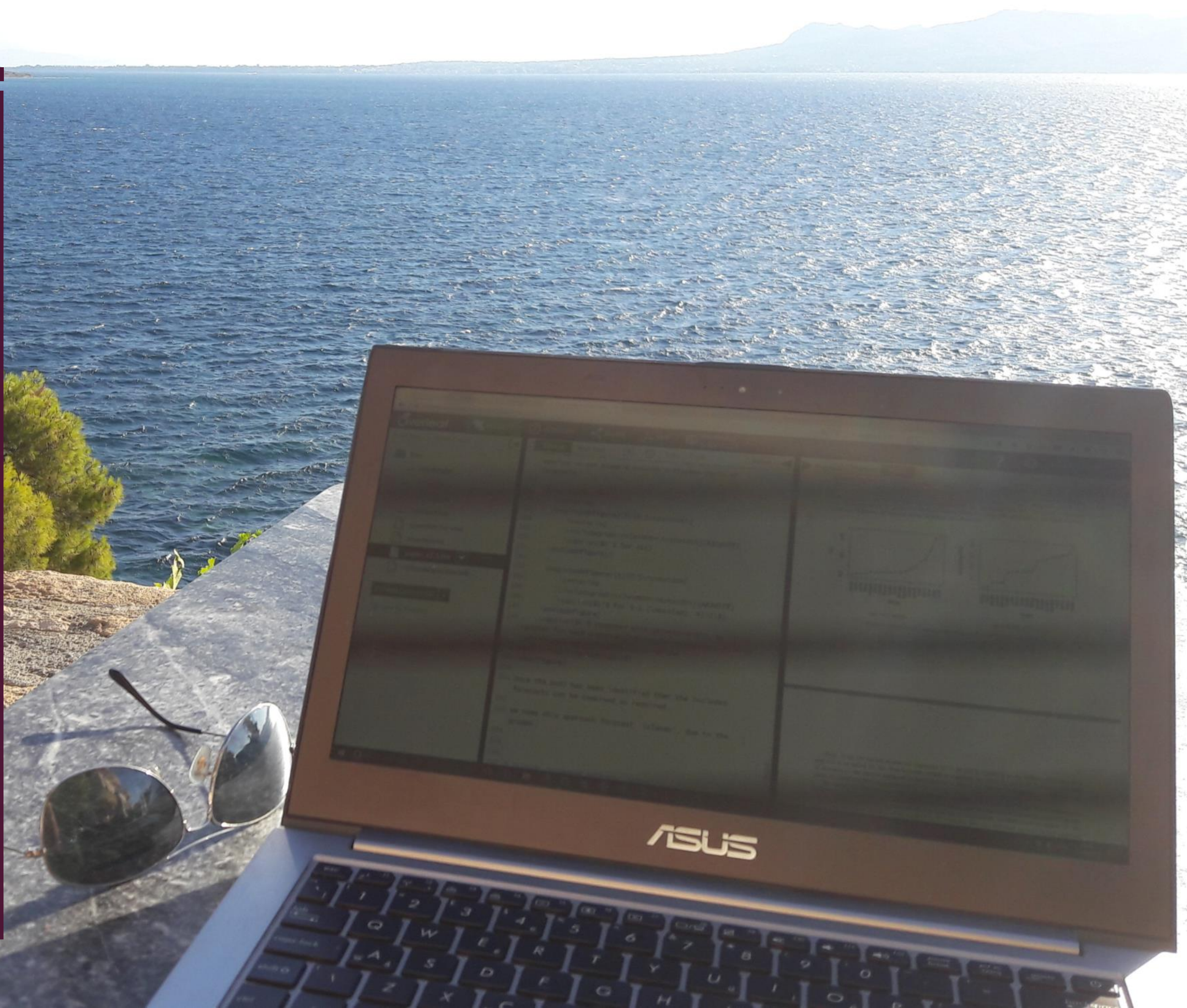
Median				
	Annual	Quarterly	Monthly	Other
Best selection	9.36%	4.39%	6.78%	**1.91%**
Best combination	**9.14%**	**4.37%**	**6.68%**	1.92%

POOLING FORECASTS – THE STATE OF PLAY

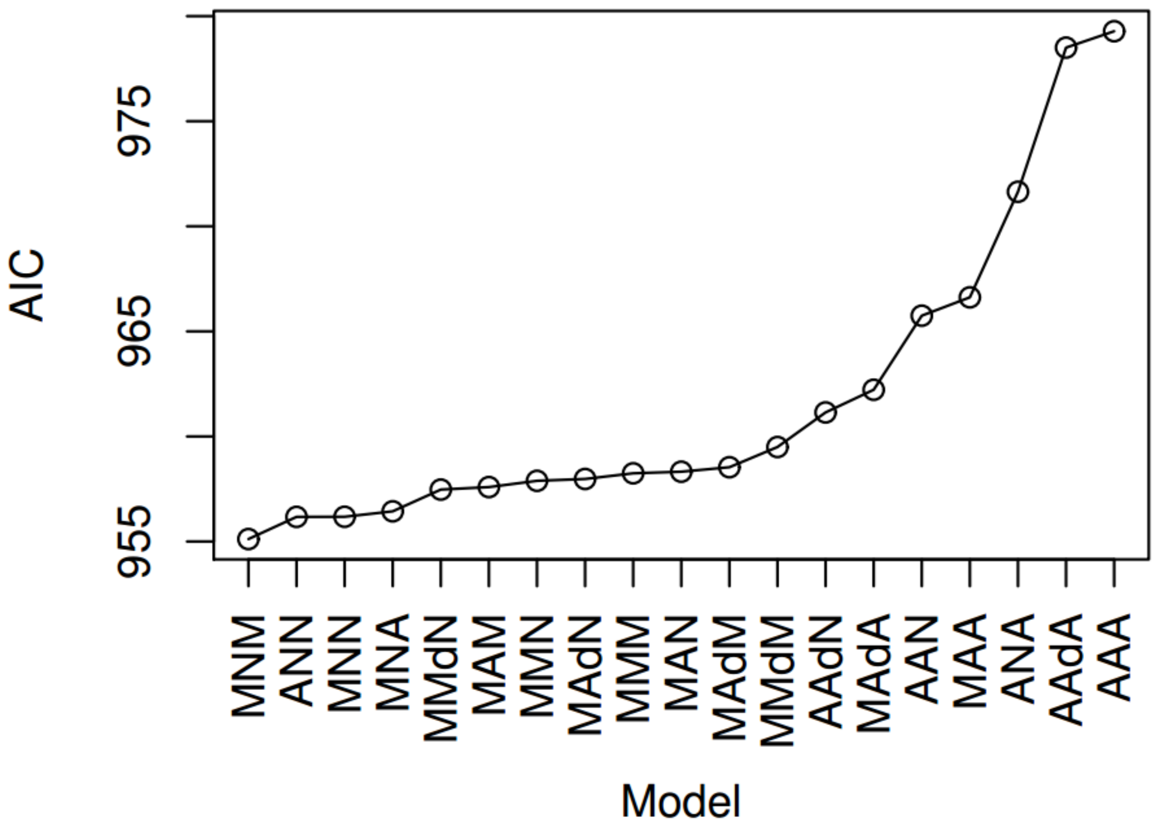
- Pooling forecasts for planning (inventory, scheduling, etc.):
  - Top 2 and Top 3: selecting the second and third best forecasting models according to some criteria.
  - Trimming: discard top best and worst forecasts (10%, 20%, 30%, ….)
  - Quartile pooling: assign each forecast model to a quartile and combination 1st quartile (Aiolfi and Timmermann, 2004) which first.
  - Islands? **Proposed Method**

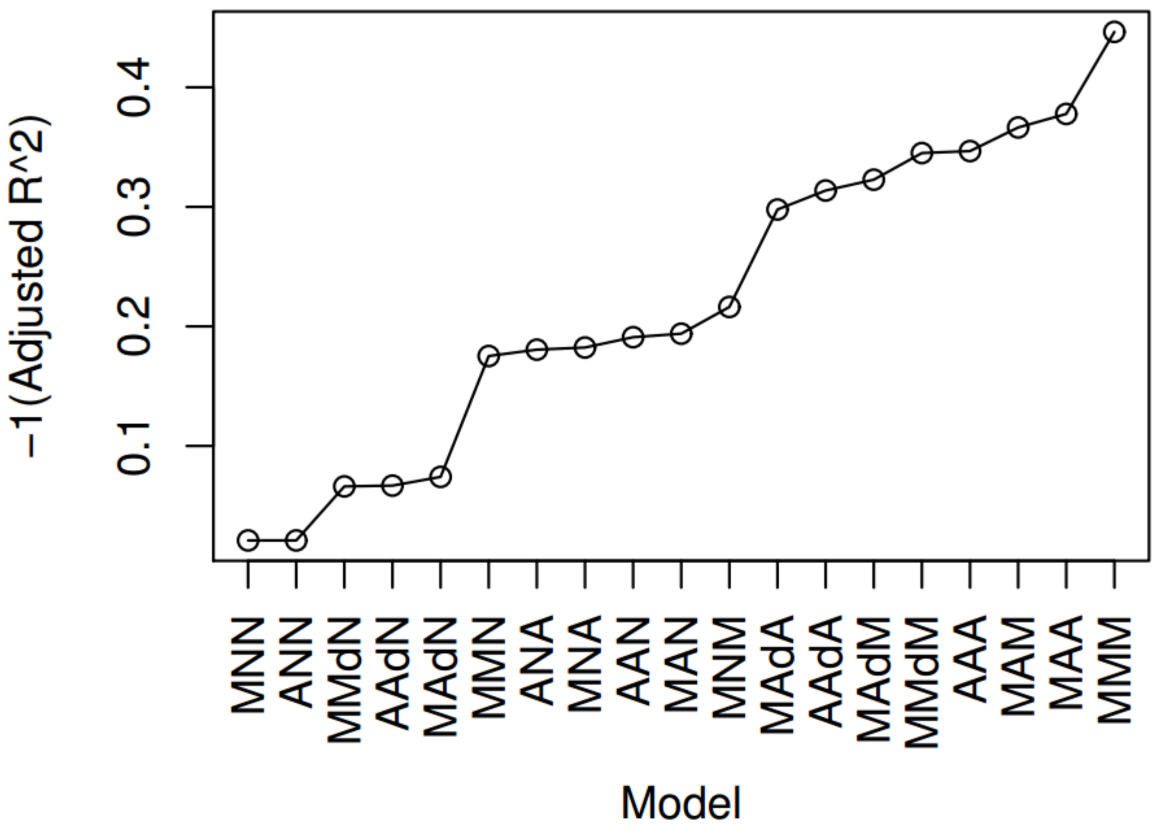
FORECAST ISLANDS

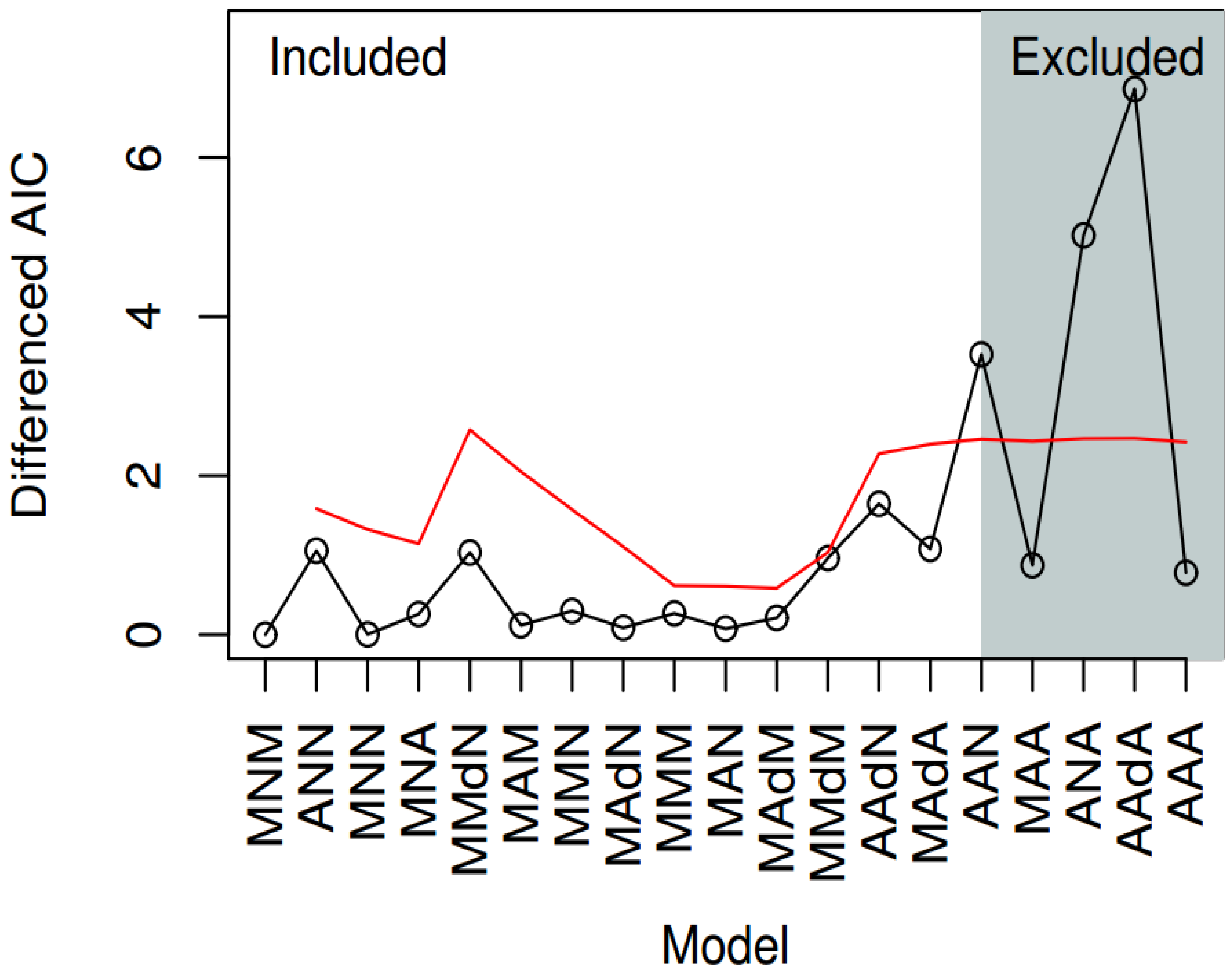
- A heuristic to form forecast pools
- Model and method independent
- Use of any criteria: information criterion like AIC, a Cross-Validation statistic, or adjusted R2



FORECAST ISLANDS






Included
Excluded
Differenced AIC
0
2
4
6
MNM
ANN
MNN
MNA
MMdN
MAM
MMN
MAdN
MMM
MAN
MAdM
MMdM
AAdN
MAdA
AAN
MAA
ANA
AAdA
AAA
Model

FORECAST ISLANDS: THE ALGORITHM

1. Let $\mathrm{C} = \{\mathrm{c_i}\}$ be the values of an appropriate criterion to assess the forecasts for $i = 1, \ldots, k$ forecasts
2. Transform the criterion to ensure that a smaller value is best
3. Order the forecasts from best to worst.
4. From the sorted metric construct $C' = \{0, \Delta C\}$, where Δ is the differencing operator
5. Island Threshold $\mathrm{T} = \mathrm{Q3} + 1.5\mathrm{IQR}$, where $\mathrm{Q3}$ is the 3rd quartile and IQR is the inter-quartile range.
6. Include all forecasts in the pool up until $\mathrm{C}' \geq \mathrm{T}$.

RESULTS ON POOLING: *INFORMATION CRITERIA*

	Mean				Median			
	Annual	Quarterly	Monthly	Other	Annual	Quarterly	Monthly	Other
Islands, equal weights	16.86%	**9.42%**	**14.13%**	**4.29%**	**9.15%**	**4.22%**	**6.55%**	**1.94%**
1st Quartile, equal weights	**16.42%**	9.93%	15.15%	4.97%	9.18%	4.67%	7.15%	2.77%

	Mean				Median			
	Annual	Quarterly	Monthly	Other	Annual	Quarterly	Monthly	Other
Islands, Akaike weights	17.62%	**9.59%**	**14.07%**	4.29%	9.13%	**4.37%**	**6.68%**	1.93%
1st Quartile, Akaike weights	**16.47%**	9.82%	14.37%	**4.27%**	**8.96%**	4.56%	6.79%	**1.83%**

RESULTS ON POOLING: *CROSS-VALIDATION*

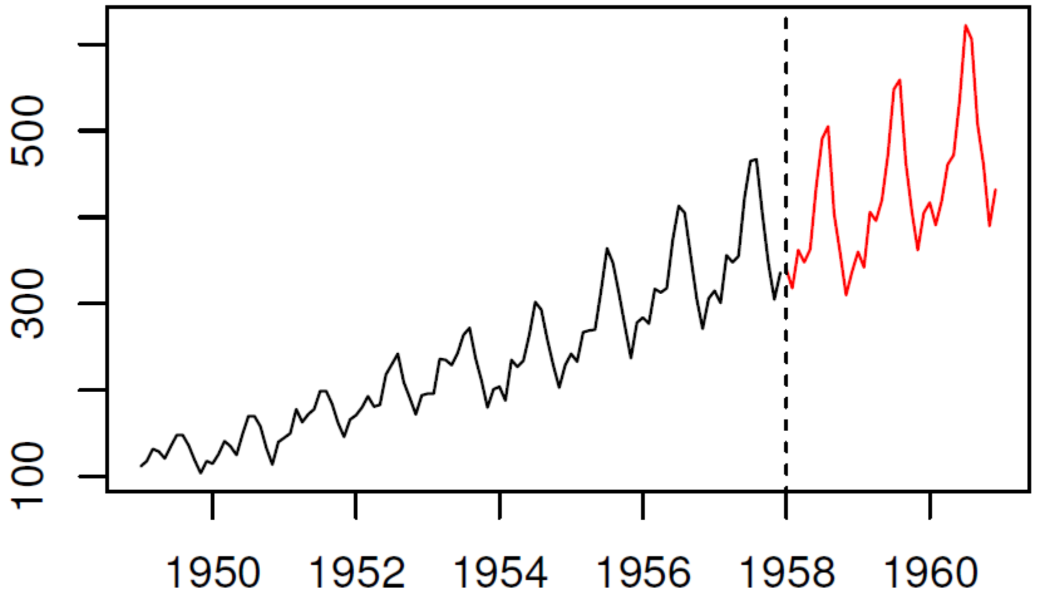
	Mean				Median			
	Annual	Quarterly	Monthly	Other	Annual	Quarterly	Monthly	Other
Islands, equal weights	16.71%	**9.53%**	**14.44%**	**4.30%**	**9.03%**	**4.15%**	**6.69%**	**1.98%**
1st Quartile, equal weights	**16.42%**	9.93%	15.15%	4.96%	9.25%	4.67%	7.15%	2.76%

	Mean				Median			
	Annual	Quarterly	Monthly	Other	Annual	Quarterly	Monthly	Other
Islands, CV weights	16.77%	**9.48%**	**14.29%**	4.39%	**9.14%**	**4.22%**	**6.66%**	1.98%
1st Quartile, CV weights	**16.68%**	9.67%	14.45%	**4.33%**	9.51%	4.36%	6.82%	**1.88%**

CONCLUSIONS

- Little difference in AIC and BIC (agrees with Billah et al., 2006) **and now cross-validation** (extension to Billah et al., 2006)
- The **combine all** approach is always better
- **Pooling improves on the combine all approach**, results consistent across combination methods
- Performance benefits of pooling improves with number of models combined

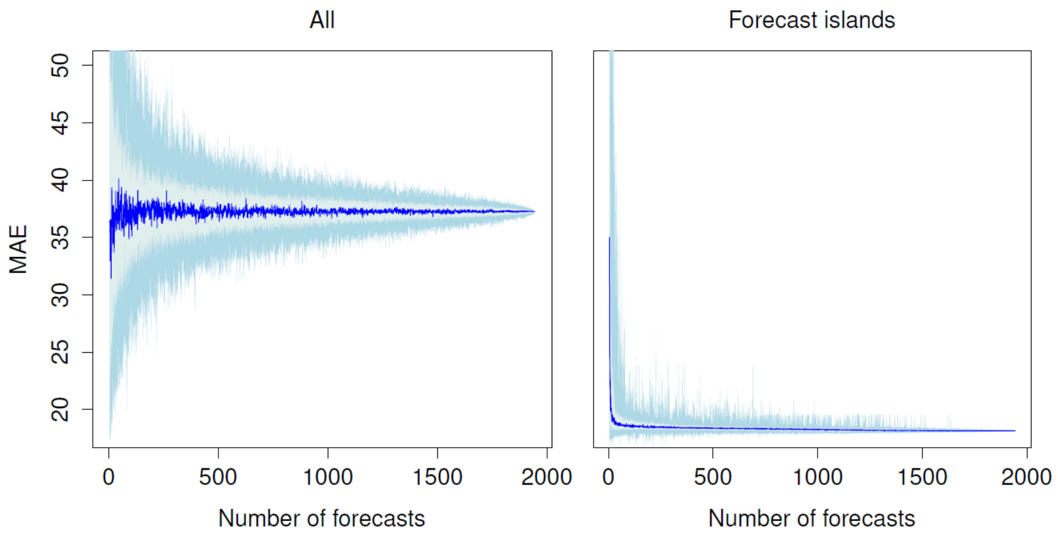
SIMULATION SETUP



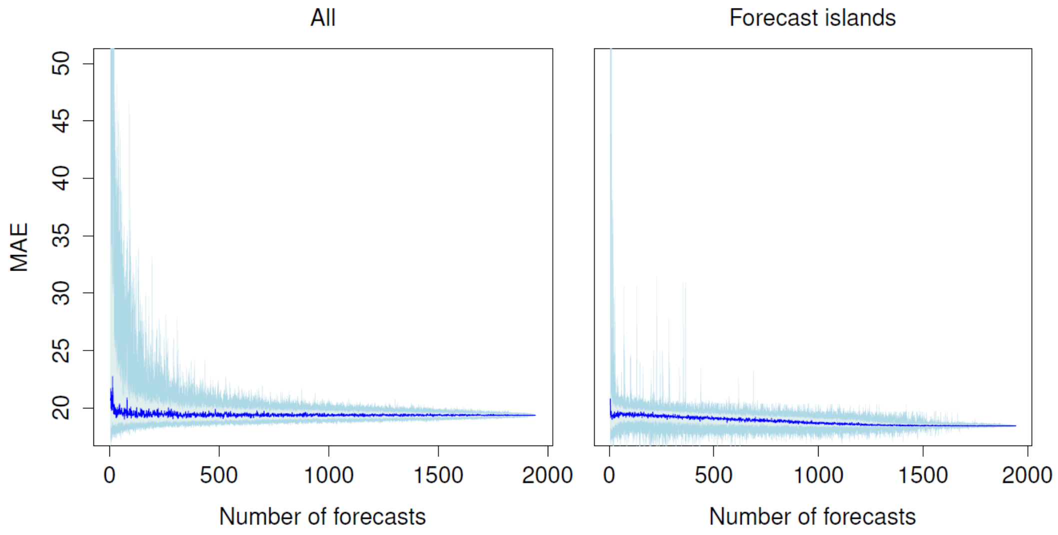
Monthly airline passenger time series.

- Time series: Airline Passenger Data
- Fitting sample: 108 observations
- Fitted Models: SARIMA(p,d,q)(P,D,Q) models with p = q = (0; 1; 2; 3; 4; 5), d = (0; 1; 2), P = Q = (0; 1; 2) and D = (0; 1)
- Combining forecasts k=2:1944, k = 1 is model selection
- Minimum of 50 different randomly selected sets of models,
- Distribution of MAE,: min, 20%, median, 80% and max

RESULTS OF A SIMULATION

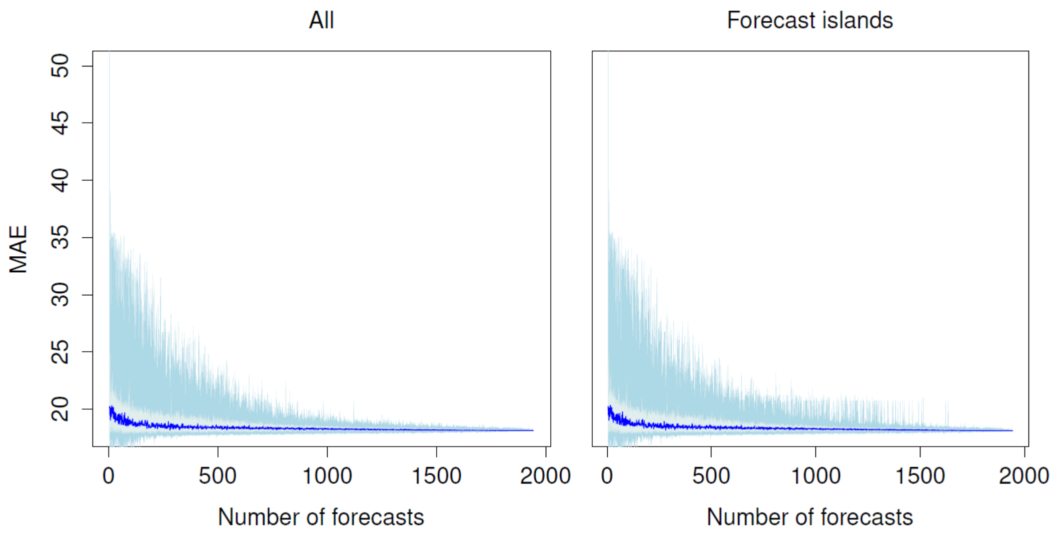


MAE for different pool sizes with **mean** combination operator.



MAE for different pool sizes with **median** combination operator

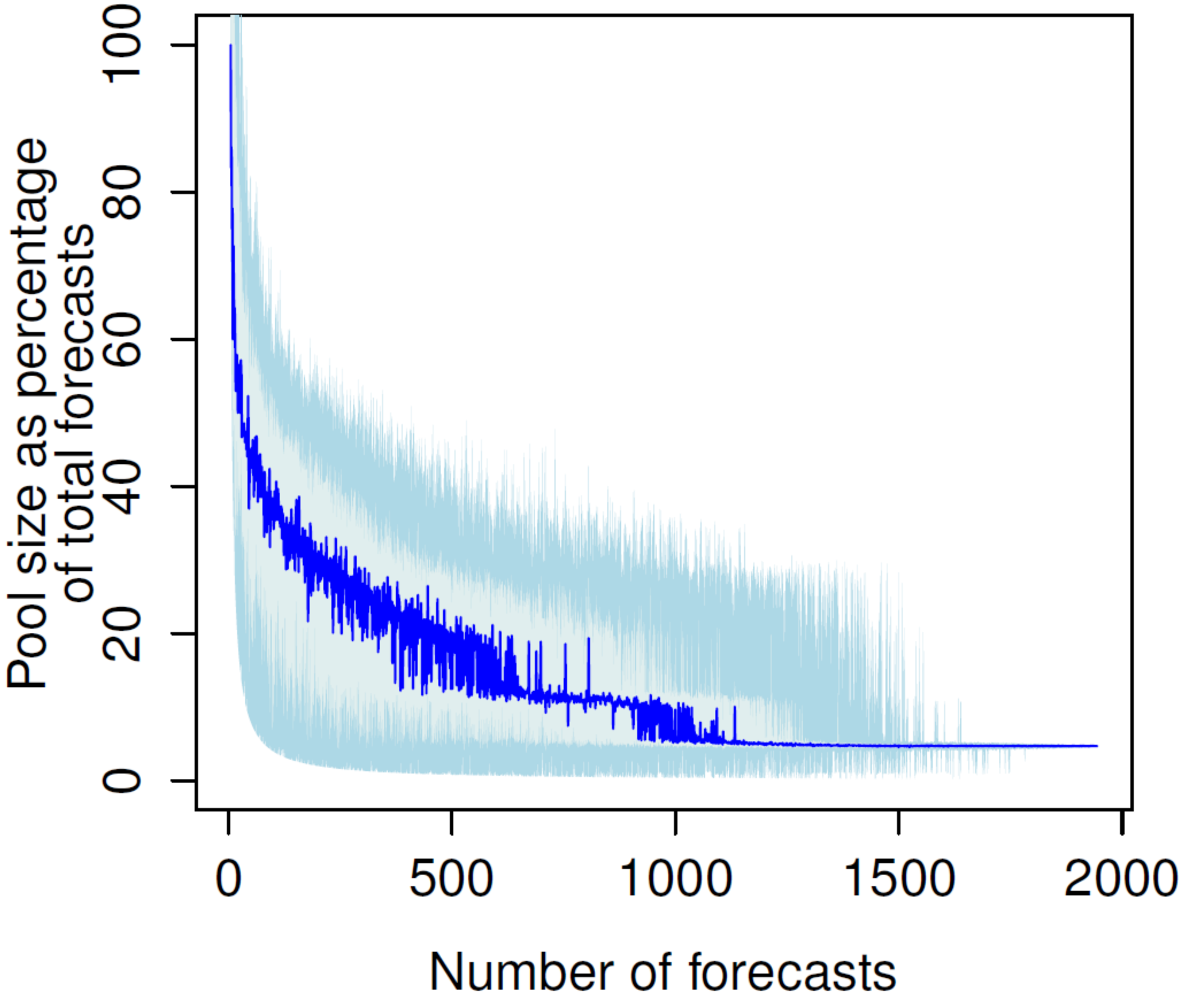
RESULTS OF A SIMULATION



MAE for different pool sizes with **BIC-weights** combination operator.

Pool	Combination method		
	Mean	Median	BIC-weights
All	37.30	19.57	18.50
Islands	**18.40**	18.81	18.52
Model selection	19.73	19.73	19.73

RESULTS OF A SIMULATION



Pool size as a percentage of all available forecasts.