Forecasting for lead-time period by temporal aggregation: Whether to combine and how

Thanos Goltsos

The PARC Assistant Professor of Management Science

PARC Institute, Cardiff Business School

in collaboration with Bahman Rostami-Tabar and Shixuan Wang

ISF2022

Oxford

Presentation overview

- Introduction and background information
  - Temporal aggregation
  - Forecast combinations
- Methodology
  - Forecasting methods
  - Temporal aggregation and combination approaches
  - Performance measurement
  - Data and simulation
- Results
  - Percentage best of aggregation approaches
  - Performance of combinations

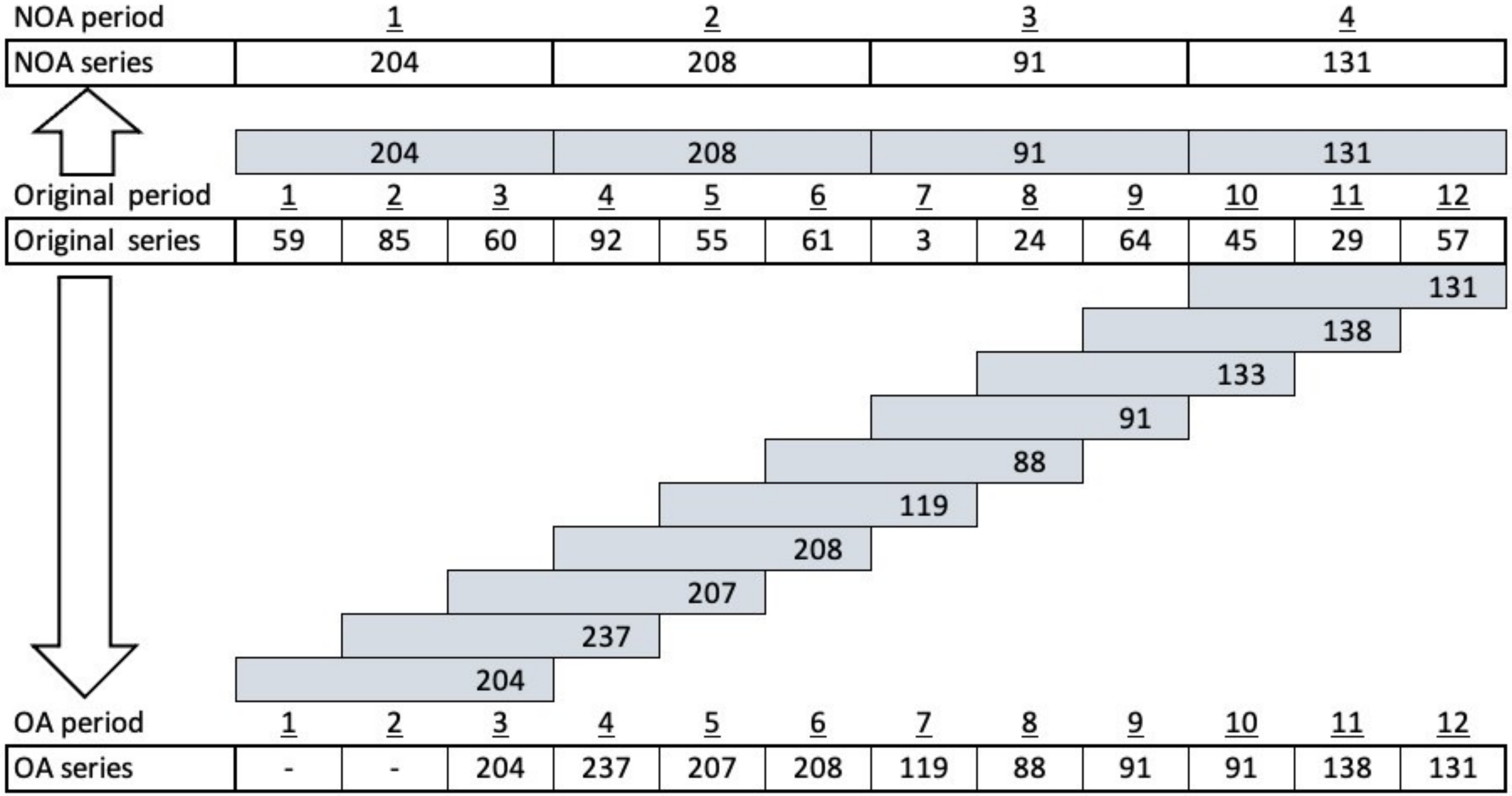
Introduction and background information

- Temporal aggregation rules/approaches
  - Aggregate forecasts (Bottom-Up approach – BU)
  - Overlapping Aggregation (OA)
  - Non-Overlapping Aggregation (NOA)
- Forecast combinations
  - Straight averages
  - Polynomial potential aggregation rule (Machine Learning Polynomial – MLP)

Background - Temporal aggregation rules

- Aggregate Forecasts (Bottom-Up approach – BU)
  - Aggregate the forecasts across a lead time
- Overlapping temporal Aggregation (OA)
  - Aggregate time-series through a sliding window approach
  - Each data point used m times
- Non-Overlapping temporal Aggregation (NOA)
  - Aggregate time-series through a moving window approach
  - Each data point used once
  - The non-overlapping temporal aggregation approach has been the main focus of the literature.

Background - Temporal aggregation rules



Background – Forecast combinations

- Combining has long been widely considered to be beneficial for forecasting in various fields (Clemen, 1989)
  - Combination of different forecasting methods (statistically or otherwise derived)
  - Combination of forecasts calculated from different temporally aggregated frequencies (mostly NOA; e.g., Athanasopoulos et al., 2017; Kourentzes 2014)
  - In the current work we explore the combination of forecasts calculated from different aggregation rules (BU, OA, NOA)
    - This can be used in parallel with all the other methods
    - To the best of our knowledge the first attempt at this

Background – Forecast combinations: forecasting methods

- He and Xu (2005) proposed a self-organising forecast combination method and showed it outperformed linear and neural network combination approaches
- Kolassa (2011) propose the use of Aikake weights on exponential smoothing forecasts and show that it consistently outperformed the use of single ’best’ forecasts (when those were selected by information criteria)
- Simple combination approaches seem to perform reasonably well compared to more complex ones (Clemen, 1989; Hibon and Evgeniou, 2005; Jose and Winkler, 2008)

Background – Forecast combinations: temporal aggregation

- Combinations can lead to forecast accuracy improvements (Andrawis et al., 2011)
- Kourentzes et al. (2014) recommended using multiple levels of TA and combining the separate forecasts (MAPA)
  - Benefits from managing the modelling risk, utilises the established gains of forecast combination (Barrow and Kourentzes, 2016; Blanc and Setzer, 2016)
- Since, modelling with multiple TA levels has been used successfully to intermittent demand, promotional modelling and inventory management (Petropoulos and Kourentzes, 2014; Kourentzes and Petropoulos, 2016; Barrow and Kourentzes, 2016)

Methodology

- Forecasting methods
- Temporal aggregation and combination approaches
- Performance measurement
- Data and simulation

Methodology – Forecasting methods

- Exponential Smoothing (ES) state space family of models: ETS (Error, Trend, Seasonality) (see Athanasopoulos 2021, or Hyndman et al., 2008)
  - The trend and seasonality components can be none (N), Additive (A) or multiplicative (M)
  - The trend can additionally be damped or not
  - The error term can also be additive (A) or multiplicative (M)
- AutoRegressive Integrated Moving Average (ARIMA) models
- We use automatic ETS and ARIMA from the fable package in R

Methodology – Aggregation and combination approaches

- The three aforementioned aggregation approaches independently
- Straight averages
  - The average approach is an equal weight combination of Bottom-Up (BU), Non-overlapping (NOA) and Overlapping (OA) temporal aggregation approaches.

$$\hat{y}_{T,m}^{Average} = \frac{\hat{y}_{T,m}^{BU} + \hat{y}_{T,m}^{NOA} + \hat{y}_{T,m}^{OA}}{3}$$

- Polynomial potential aggregation rule (Machine Learning Polynomial – MLP) (learning mechanism)

$$\hat{y}_{T,m}^{MLP} = p_t^{BU}\hat{y}_t^{BU} + p_t^{NOA}\hat{y}_t^{NOA} + p_t^{OA}\hat{y}_t^{OA}$$

Methodology – Performance measurement

- Mean Absolute Scaled Error (MASE) (Hyndman and Koehler, 2006)
  - Absolute errors of out of sample errors are scaled by the mean absolute error of the one step ahead, in sample naïve forecasts
- Mean Absolute Percentage Errors (MAPE) were also calculated but omitted

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

$$q_j = \frac{y_j - \hat{y}_j}{\frac{1}{n-1}\sum_{t=2}^{n} |y_t - y_{t-1}|}$$

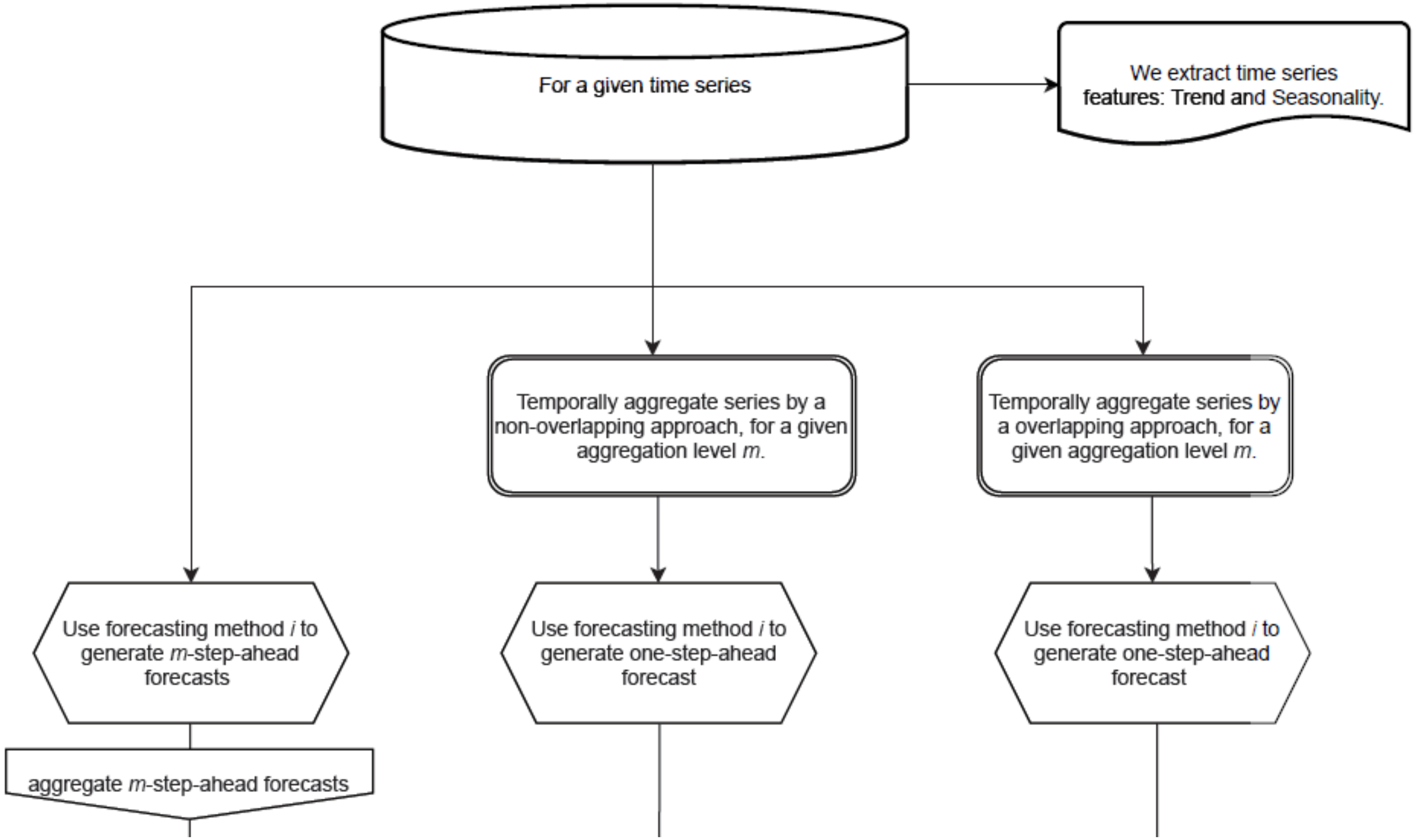
Methodology - Data

- Quarterly, monthly and daily subsets of M4 forecasting competition (Makridakis et al., 2018) datasets to evaluate empirically the forecast accuracy of five approaches for a given forecasting method
- M4 datasets include time series from various sectors such as demographic, industry, finance, economics, and others.

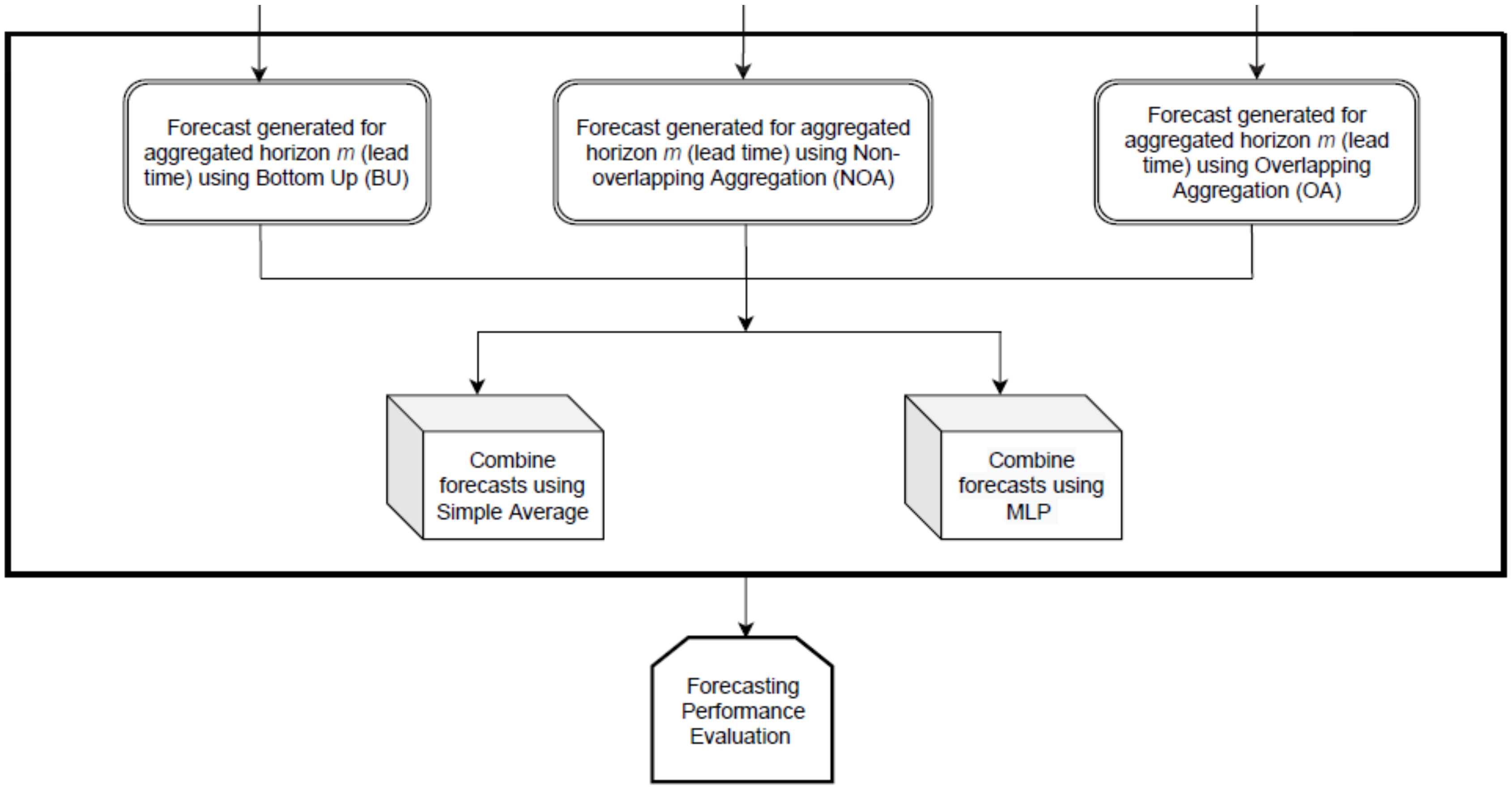
Table 1: The number of time series in M4 competition data

M4 series	Total	Component			
		(N,N)	(N,S)	(T,N)	(T,S)
Quarterly	24,000	5,869	2,959	9,325	5,847
Monthly	48,000	11,412	8,752	11,841	15,995
Daily	4,227	3,335	43	819	30

Methodology - Simulation



Methodology - Simulation



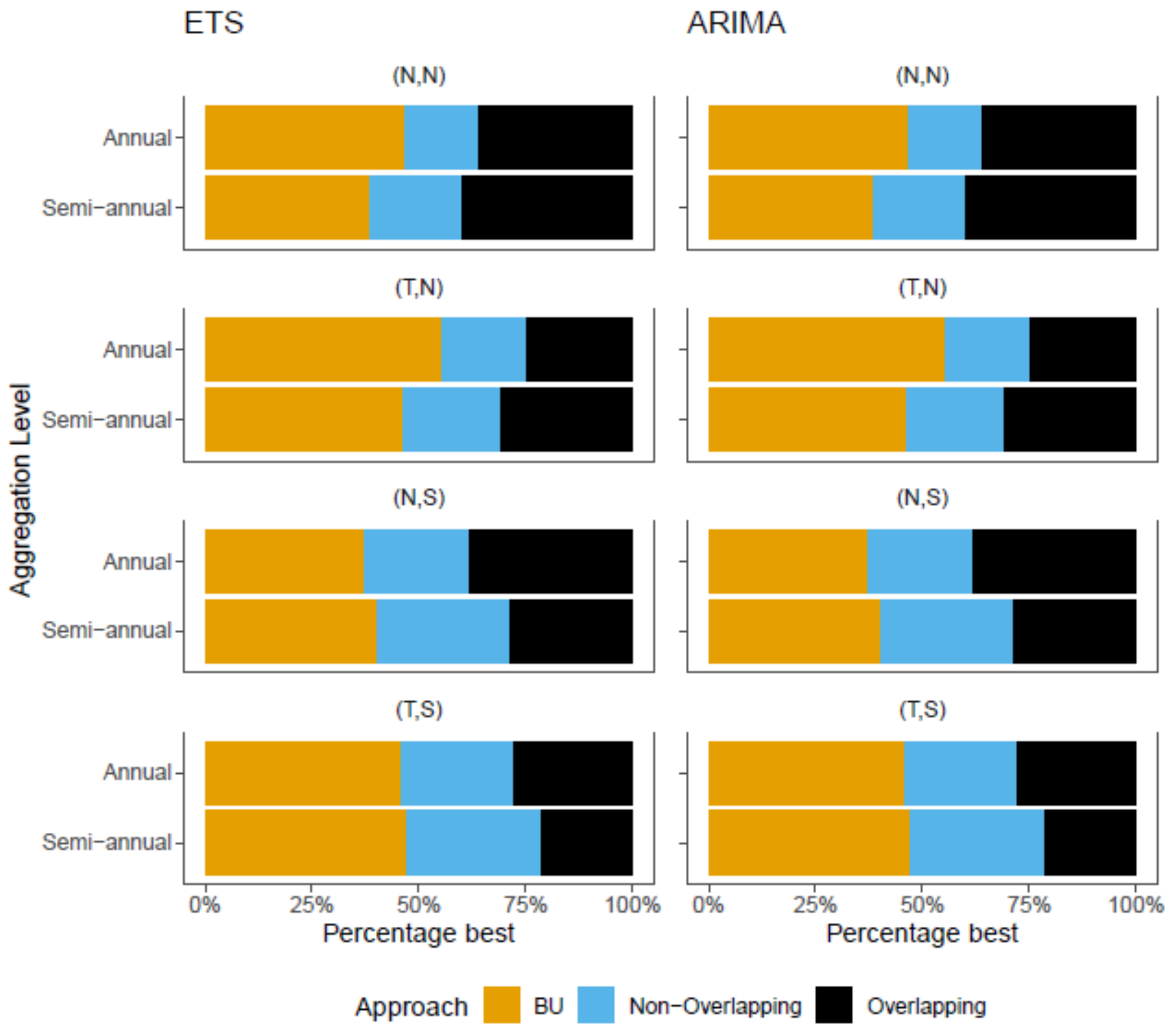
Methodology - Simulation

- Lead-time forecasts equal to the aggregation levels
  - Quarterly time series, m = 2, 4 (annual and semi-annual)
  - Monthly time series, m = 2, 3, 4, 6, 12 (bi-monthly to annual)
  - Daily time series, m = 2, 3, 4, 5, 6, 7 (2 days to 1 week)(7 days)
- Rolling origin forecast evaluation to determine the forecast accuracy of each approach, for a given forecasting method and aggregation level.
  - We use the training set to generate the forecast for the first given lead-time in the out-of-sample, followed by computing the error metric.
  - Then, we include one new observation in the training set and continue the process until the number of observations left in out-of-sample equals the aggregation level. This will be the last generated forecast.
- Supercomputing facilities were used to run the experiment. The computational time was 9 weeks.

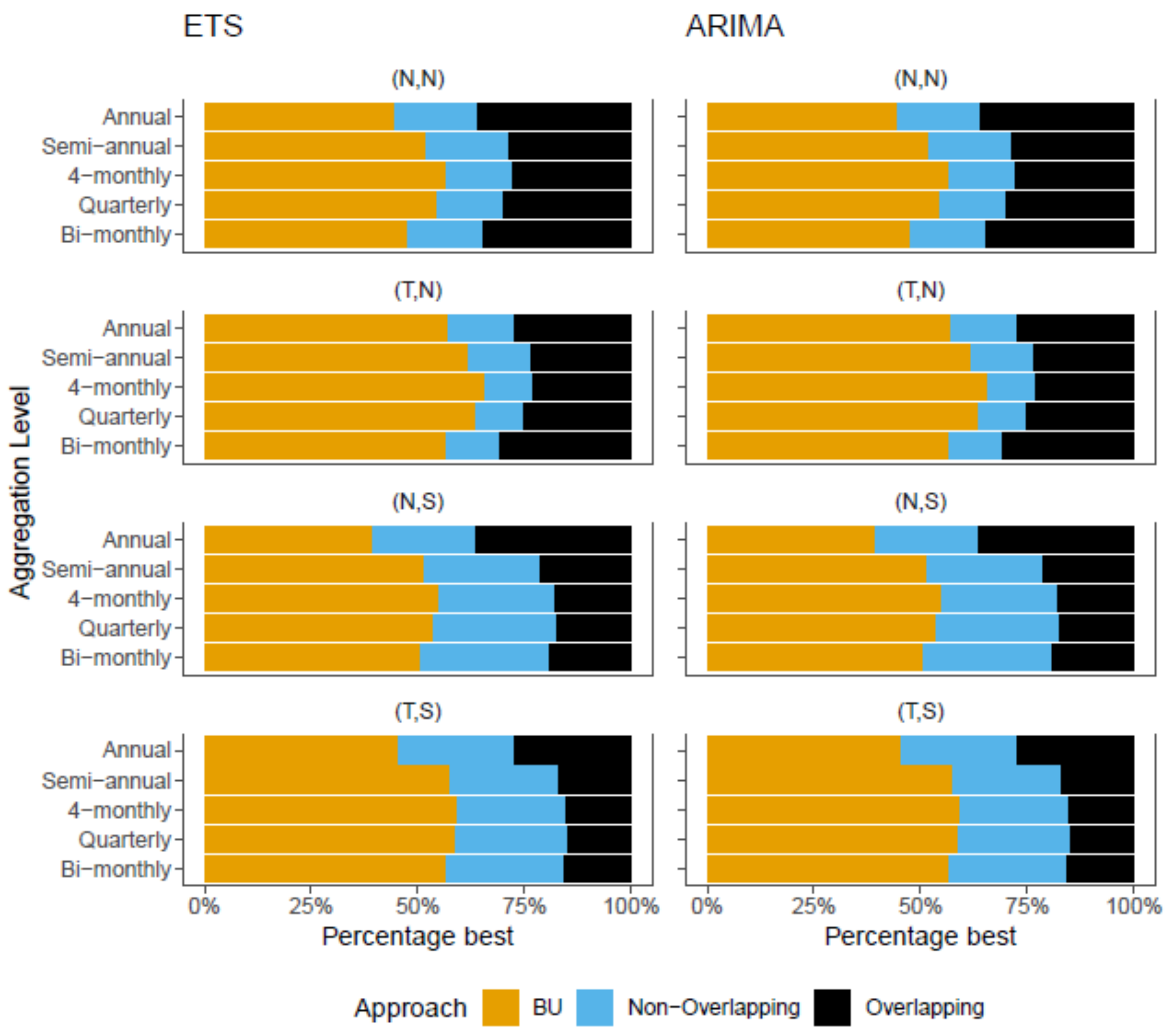
Results

- Percentage best of aggregation approaches
  - Simply how many times approach A outperformed approach B
- Performance of combinations
  - As measured by mean absolute scaled error (MASE)
- All results are statistically significant according to the Multiple Comparison with the Best (MCB) method (Demšar, 2006)

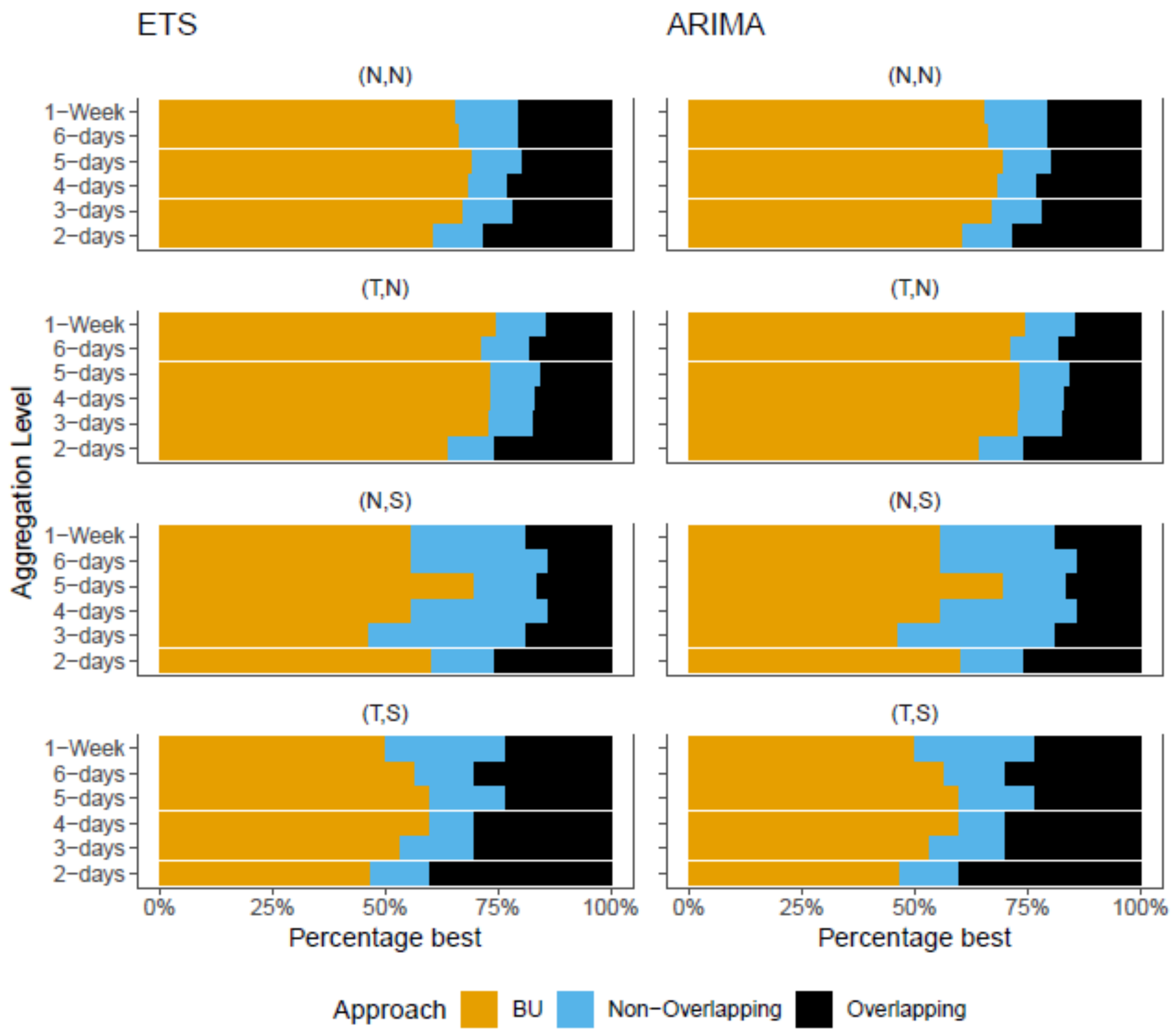
Results - Percentage best TA (Quarterly)



Results - Percentage best TA (Monthly)



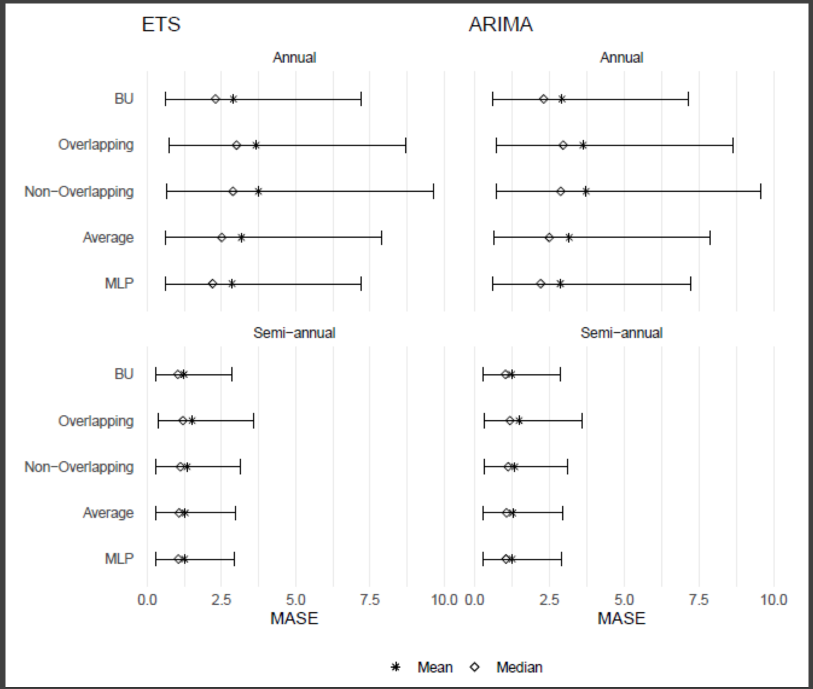
Results - Percentage best TA (daily)



Results summary – Percentage best

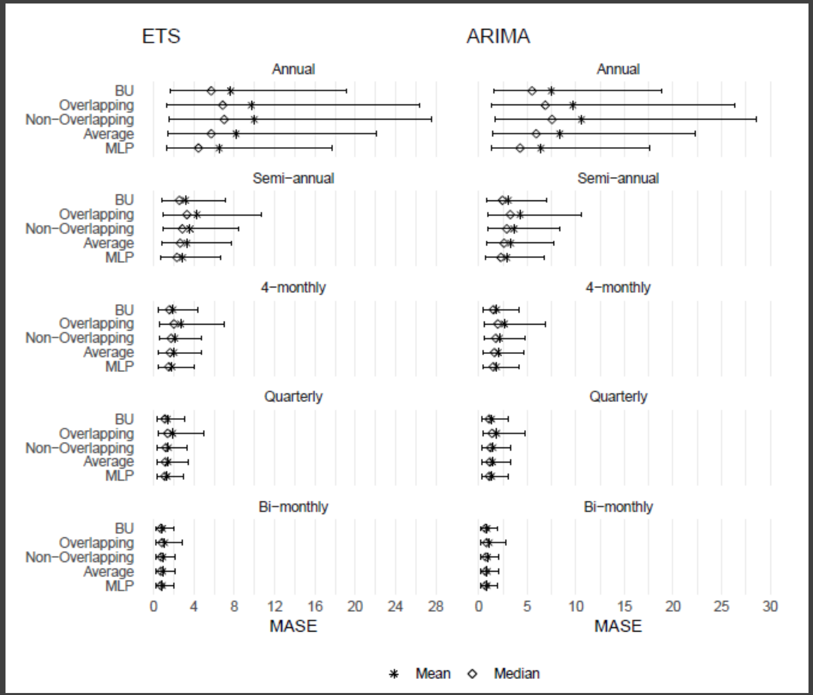
- When forecast over lead-time is required, TA approaches might not always provide more accurate forecasts.
- BU is a reliable competitor (almost 50% percentage best – 75% for daily) regardless of whether there is any pattern such a trend or seasonality in the time series
- Temporal aggregation approaches for time series with seasonality and trend are slightly better compared to no trend and seasonality
- BU, overlapping and Non-overlapping temporal aggregation may have their own merits and areas of comparative overperformance which is part of the motivation

Results - Performance of combinations (Quarterly, ETS)



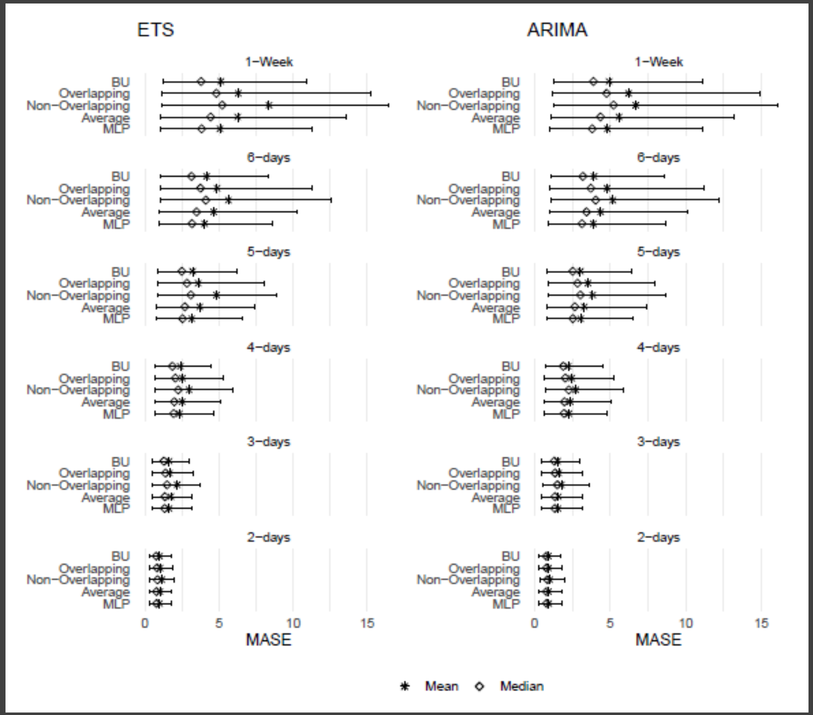
Aggregation level	Pattern	Approach				
		MLP	Average	Overlapping	Non-overlapping	BU
Annual	(N, N)	2.880 (**2.298**)	3.048 (2.439)	3.088 (2.458)	3.610 (2.843)	**2.845** (2.307)
	(T, N)	**2.820** (**2.096**)	3.294 (2.575)	4.085 (3.423)	4.008 (3.083)	2.868 (2.194)
	(N, S)	**3.086** (**2.473**)	3.178 (2.544)	3.244 (2.586)	3.573 (2.760)	3.159 (2.567)
	(T, S)	**2.769** (**2.126**)	3.115 (2.467)	3.791 (3.184)	3.555 (2.679)	2.894 (2.303)
Semi-annual	(N, N)	1.225 (1.044)	1.226 (1.048)	1.260 (1.068)	1.315 (1.117)	**1.206** (**1.026**)
	(T, N)	1.115 (0.877)	1.132 (0.893)	1.232 (1.007)	1.251 (0.991)	**1.104** (**0.872**)
	(N, S)	1.587 (1.392)	1.614 (1.396)	2.006 (1.647)	1.617 (1.395)	**1.566** (**1.345**)
	(T, S)	1.351 (1.170)	1.406 (1.214)	1.959 (1.558)	1.384 (1.172)	**1.324** (**1.124**)

Results - Performance of combinations (Monthly, ETS)



Aggregation level	Pattern	Approach				
		MLP	Average	Overlapping	Non-overlapping	BU
Annual	(N, N)	**6.675 (4.456)**	7.958 (5.38)	8.491 (5.723)	9.822 (6.939)	7.673 (5.590)
	(T, N)	**6.483 (4.299)**	8.785 (5.834)	11.161 (7.643)	11.472 (7.66)	7.603 (5.495)
	(N, S)	**6.826 (4.760)**	7.883 (5.579)	8.288 (5.857)	9.191 (6.585)	8.119 (6.252)
	(T, S)	**6.215 (4.394)**	8.064 (5.957)	10.23 (7.96)	9.417 (6.820)	7.396 (5.671)
Semi-annual	(N, N)	**3.023 (2.500)**	3.236 (2.635)	3.398 (2.800)	3.642 (2.871)	3.170 (2.637)
	(T, N)	**2.440 (1.809)**	2.893 (2.153)	3.502 (2.640)	3.511 (2.607)	2.714 (2.045)
	(N, S)	**3.421 (2.803)**	3.875 (3.167)	5.161 (3.992)	4.025 (3.315)	3.82 (3.156)
	(T, S)	**2.835 (2.290)**	3.366 (2.695)	5.105 (3.833)	3.442 (2.768)	3.156 (2.548)
4-monthly	(N, N)	**1.886 (1.620)**	1.951 (1.668)	2.041 (1.751)	2.155 (1.819)	1.916 (1.635)
	(T, N)	**1.430 (1.063)**	1.574 (1.181)	1.845 (1.421)	1.837 (1.420)	1.499 (1.115))
	(N, S)	**2.288 (1.937)**	2.538 (2.115)	3.501 (2.752)	2.551 (2.146)	2.457 (2.056)
	(T, S)	**1.824 (1.497)**	2.121 (1.693)	3.330 (2.354)	2.057 (1.691)	1.948 (1.578)
Quarterly	(N, N)	**1.358** (1.179)	1.382 (1.198)	1.446 (1.259)	1.496 (1.292)	1.366 (**1.167**)
	(T, N)	**0.998 (0.731)**	1.054 (0.772)	1.198 (0.909)	1.195 (0.917)	1.012 (0.733)
	(N, S)	**1.722 (1.471)**	1.862 (1.567)	2.515 (1.986)	1.854 (1.553)	1.807 (1.529)
	(T, S)	**1.338 (1.093)**	1.524 (1.198)	2.372 (1.644)	1.448 (1.181)	1.394 (1.121)
Bi-monthly	(N, N)	0.881 (0.758)	0.886 (0.764)	0.931 (0.799)	0.926 (0.799)	**0.877 (0.745)**
	(T, N)	0.615 (0.423)	0.627 (0.429)	0.684 (0.470)	0.679 (0.493)	**0.608 (0.412)**
	(N, S)	**1.191 (1.030)**	1.218 (1.050)	1.504 (1.241)	1.223 (1.043)	1.208 (1.031)
	(T, S)	**0.895** (0.717)	0.951 (0.745)	1.329 (0.951)	0.919 (0.737)	0.900 (**0.713**)

Results - Performance of combinations (Daily, ETS)



Aggregation level	Pattern	Approach				
		MLP	Average	Overlapping	Non-overlapping	BU
1-week	(N, N)	4.692 (3.749)	5.615 (4.231)	5.819 (4.59)	7.212 (4.969)	**4.604** (**3.718**)
	(T, N)	**6.303** (**2.504**)	6.032 (3.064)	4.707 (3.354)	11.095 (3.505)	8.253 (2.535)
	(N, S)	**5.619** (4.300)	7.05 (5.409)	8.237 (6.22)	8.45 (6.587)	5.622 (**4.212**)
	(T, S)	30.772 (4.732)	57.594 (5.726)	**10.305** (5.435)	132.226 (6.179)	45.217 (**4.694**)
6-days	(N, N)	3.796 (3.133)	4.292 (3.371)	4.516 (3.606)	5.218 (3.962)	**3.685** (**3.085**)
	(T, N)	4.882 (**1.94**)	4.712 (2.268)	**3.752** (2.448)	7.154 (2.997)	6.831 (2.444)
	(N, S)	4.467 (3.454)	5.368 (4.14)	6.145 (4.626)	6.421 (4.97)	**4.458** (**3.355**)
	(T, S)	17.211 (**3.913**)	23.904 (4.231)	**11.85** (4.213)	28.59 (4.968)	44.671 (3.947)
5-days	(N, N)	2.984 (2.512)	3.328 (2.616)	3.363 (2.749)	4.153 (3.012)	**2.851** (**2.455**)
	(T, N)	4.130 (**1.582**)	3.834 (1.593)	**2.838** (1.847)	5.637 (2.195)	6.205 (2.058)
	(N, S)	3.445 (2.739)	3.94 (3.082)	4.398 (3.345)	4.65 (3.631)	**3.405** (**2.642**)
	(T, S)	14.066 (**2.849**)	39.462 (3.321)	**9.237** (3.669)	85.127 (3.694)	36.409 (2.973)
4-days	(N, N)	2.233 (1.92)	2.336 (1.944)	2.375 (2.024)	2.81 (2.198)	**2.104** (**1.836**)
	(T, N)	2.290 (**1.405**)	3.6 (1.411)	**2.09** (1.531)	4.811 (1.589)	5.783 (1.574)
	(N, S)	2.566 (2.053)	2.772 (2.147)	3.003 (2.312)	3.194 (2.49)	**2.504** (**1.98**)
	(T, S)	**7.403** (2.069)	14.271 (2.484)	8.762 (2.35)	14.524 (2.915)	26.583 (**1.991**)
3-days	(N, N)	1.516 (1.323)	1.622 (1.334)	1.565 (1.369)	2.009 (1.47)	**1.436** (**1.266**)
	(T, N)	1.747 (1.126)	2.672 (1.151)	**1.661** (**0.998**)	3.659 (1.202)	4.556 (1.140)
	(N, S)	1.764 (1.432)	1.823 (1.468)	1.895 (1.557)	2.054 (1.631)	**1.709** (**1.384**)
	(T, S)	**11.671** (1.470)	14.084 (1.563)	6.718 (1.341)	22.089 (1.861)	18.328 (**1.418**)
2-days	(N, N)	0.916 (0.780)	0.919 (0.781)	0.96 (0.815)	1.027 (0.839)	**0.861** (**0.757**)
	(T, N)	1.448 (0.691)	2.033 (0.669)	2.146 (**0.664**)	**1.343** (0.72)	3.346 (0.729)
	(N, S)	**1.041** (**0.867**)	1.042 (0.863)	1.061 (0.878)	1.138 (0.935)	1.010 (0.830)
	(T, S)	**7.303** (0.900)	9.112 (0.915)	5.709 (0.837)	13.732 (0.885)	11.575 (**0.872**)

Results summary – Performance of combinations

- Quarterly:
  - MLP best for the annual lead-time, while it is the second best to forecast semi-annual lead-time, regardless of the forecasting method.
  - BU approaches is a competitive approach in both cases.
  - Using the simple average combination does not improve forecast accuracy. Both BU and MLP show less variation compared to other approaches.
  - Performance is not affected by different patterns
- Monthly:
  - MLP approach outperforms all other approaches, regardless of the forecasting method employed, gains increas with the lead-time.
  - BU approach is the second-best approach, followed by simple average, overlapping and non-overlapping temporal aggregation. BU becomes more competitive when forecasting bi-monthly lead-time
  - MLP approach shows less variation in the performance, followed by BU. Regardless of the existing pattern and the forecasting method, MLP approach is always the most accurate approach.
- Daily:
  - Both MLP and BU approaches provide accurate results
  - MLP becomes more accurate for longer lead-time, BU is more accurate for shorter lead-time.
  - Time series with trend component and no seasonality, MLP is more accurate.

Conclusions

- The superior performance of MLP is due to the online adjustment of the combination weights. MLP takes the past forecasting loss into account and penalises the poor approaches.
- Neither of the individual approaches have an overall win on forecasting accuracy.
- Performance of BU and to an extent OA are perhaps surprising compared to NOA.
- Simple average combinations of these forecasts are also underperforming.

Further research

- The proposed framework could be replicated with intermittent time series.
- Aggregating forecast and aggregating time series through temporal aggregation may lead to forecast improvement, but the conditions for this improvement remain unclear.
- More empirical investigation is required to examine the performance of temporal aggregation with weekly, daily and sub-daily time series.
- The current work can be extended to cover/combined with approaches such as MAPA and temporal hierarchies.
- While this study focused on lead-time forecasting, a further investigation is needed to evaluate the performance of proposed forecast combination approach when producing forecasts at the original, higher frequency.

Thank you very much! Happy to discuss.

Forecasting for lead-time period by temporal aggregation: Whether to combine and how

Thanos Goltsos
The PARC Assistant Professor of Management Science
PARC Institute, Cardiff Business School
in collaboration with Bahman Rostami-Tabar and Shixuan Wang

ISF2022
Oxford