ENSEMBLES FOR TIME SERIES FORECASTING

Mariana Oliveira & Luís Torgo

Outline

- Introduction
  - Delay-coordinate embedding
- Bagging for Time Series Forecasting
  - Bagging Variants
- Experimental Evaluation
  - Time series
  - Results
- Conclusion
- Future work

Introduction

- **Ensembles** are among the most **competitive** forms of solving predictive tasks;
- **Diversity** among ensemble members is essential;
- We aim at improving the predictive performance of ensembles in **time series forecasting**.

Delay-coordinate embedding

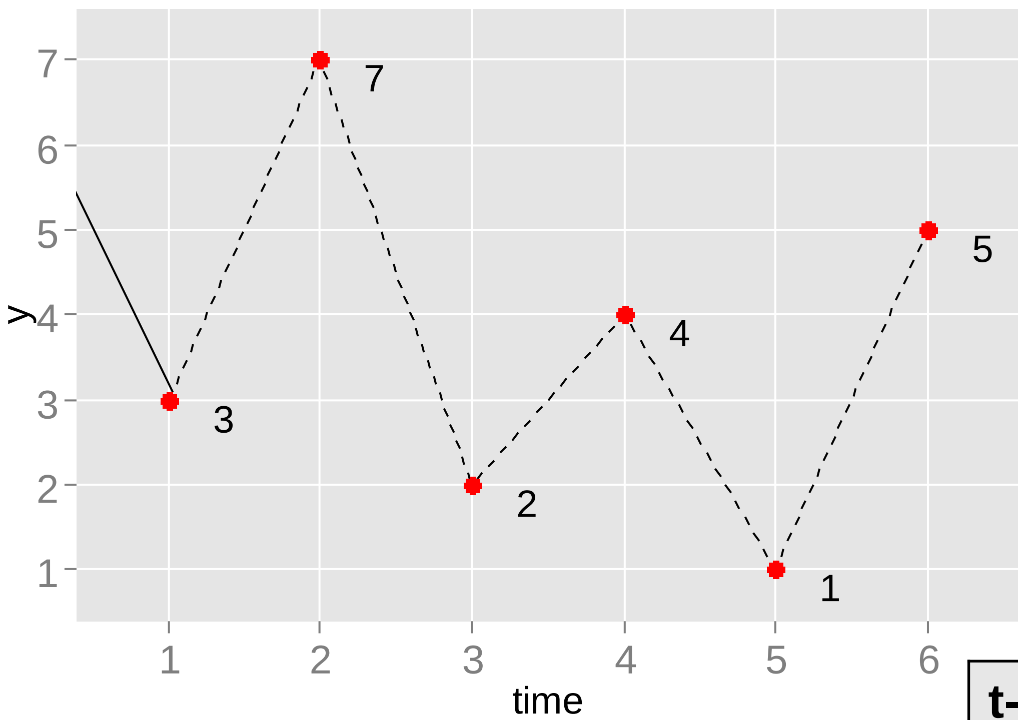
- **Delay-coordinate embedding** assumes that future values of the series are only dependent on a limited number of previous values;

- Any regression tool can then be used to obtain a model of the form

$$Y_{t+h} = f(< Y_{t-k}, \dots, Y_{t-1}, Y_t >).$$

- This requires setting the **embed size** (k) and most times there may not exist **one** single correct answer.

Delay-coordinate embedding



k=3

t-3	**t-2**	**t-1**	**t**
2	4	1	5
7	2	4	1
3	7	2	4

Bagging for Time Series Forecasting

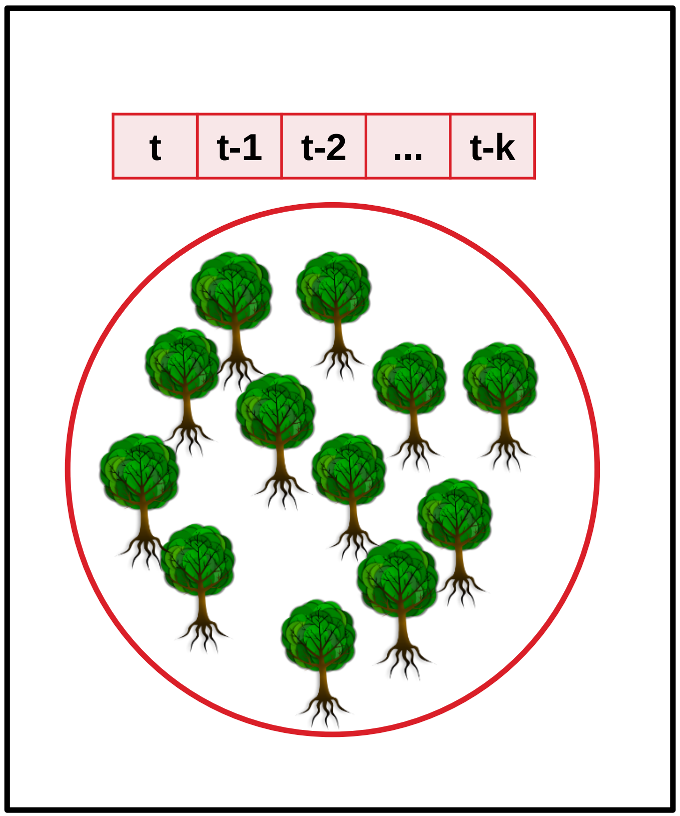
- We propose variants of **bagging** of regression trees;
- Diversity generation of our variants explores **specific properties** of time series prediction tasks;
- We will compare the performance of our proposals against that of standard bagging, our baseline.

Bagging for Time Series Forecasting

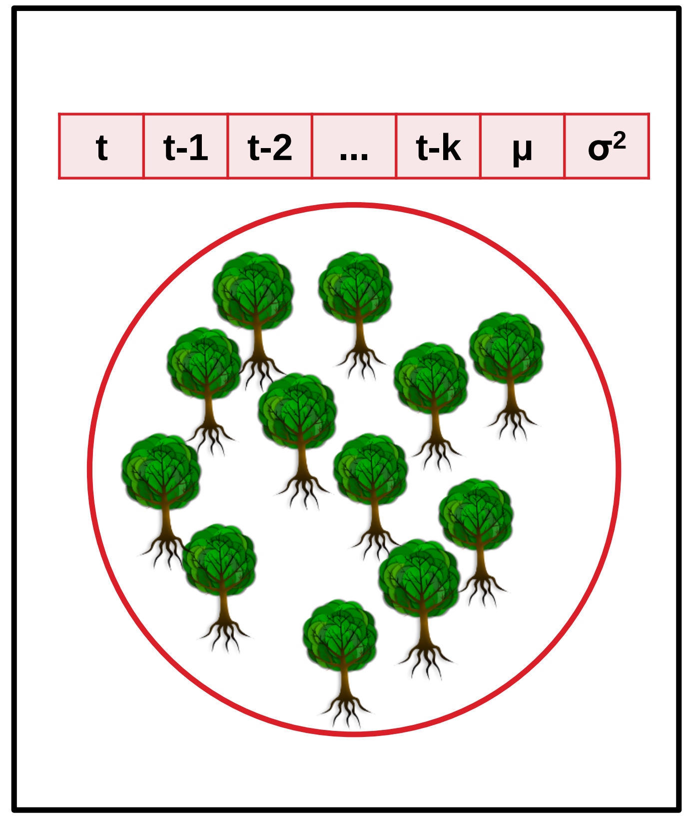
- There are many possible ways of describing the recent dynamics of a time series through a set of predictors;
- Our initial set of proposed bagging variants use
  - **different embed sizes** given a maximum embed size k_{max} ;
  - summary statistics of recent values as **additional predictors**.

Bagging variants

E

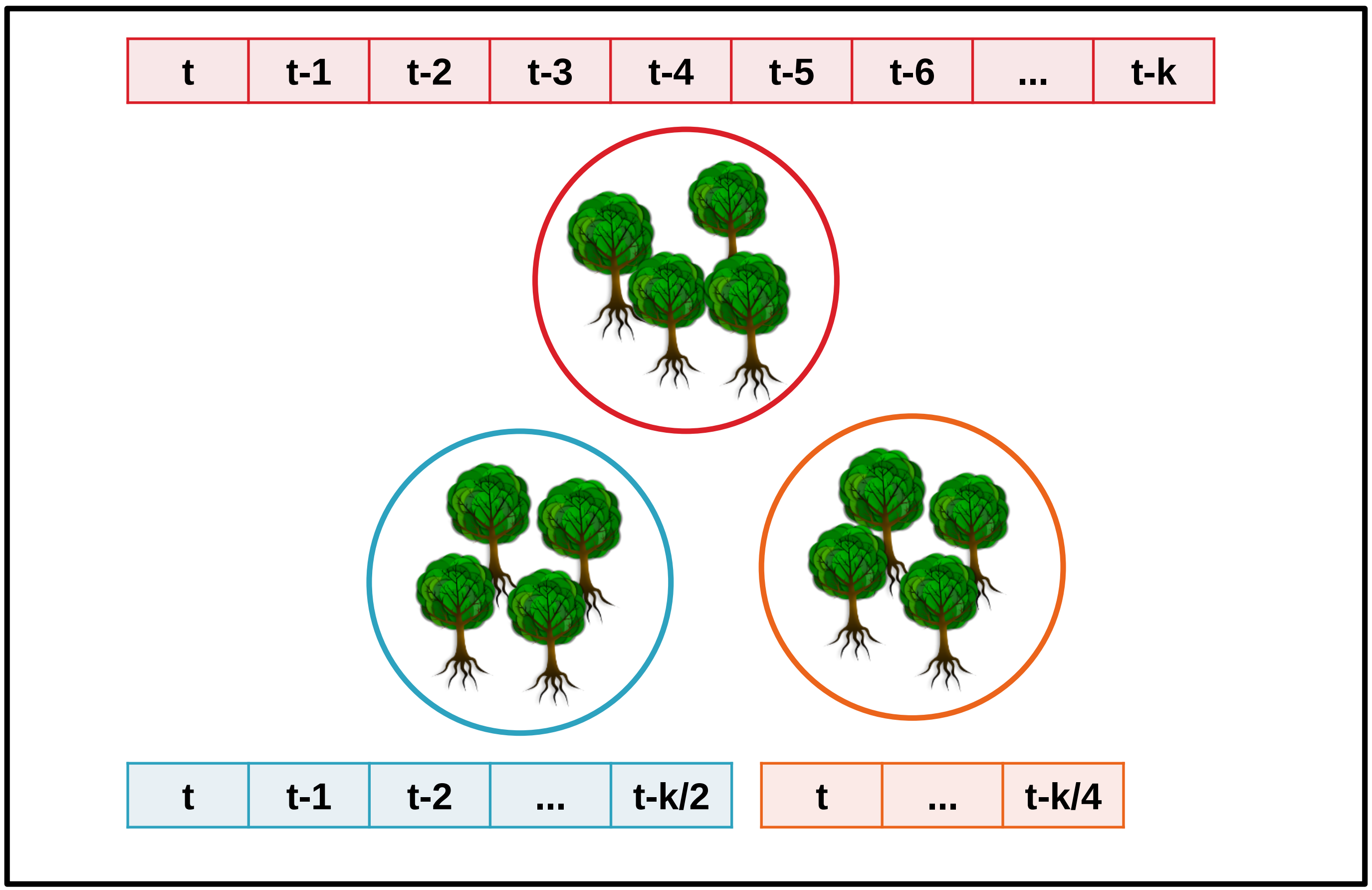


E+S



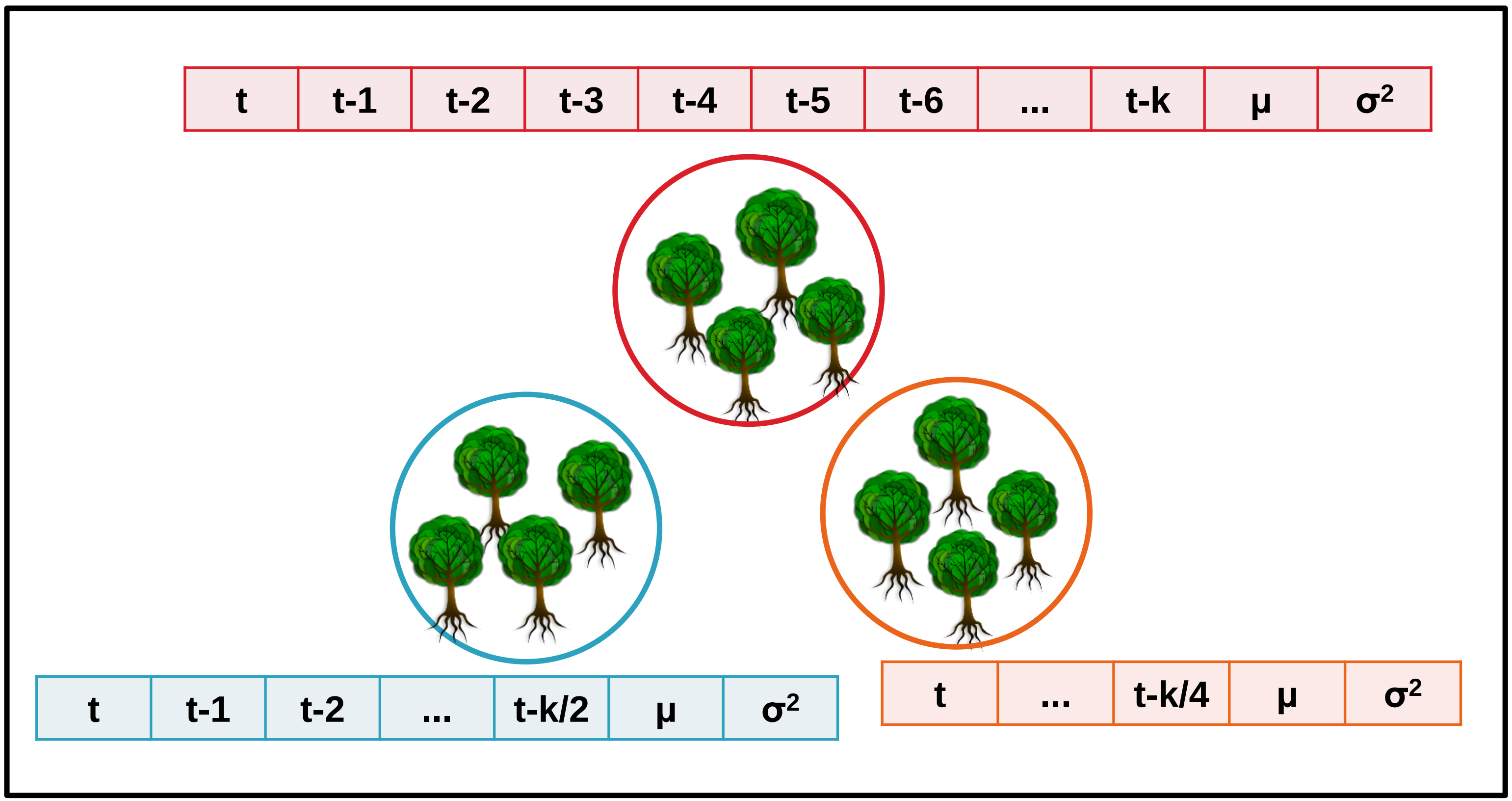
Bagging variants

DE



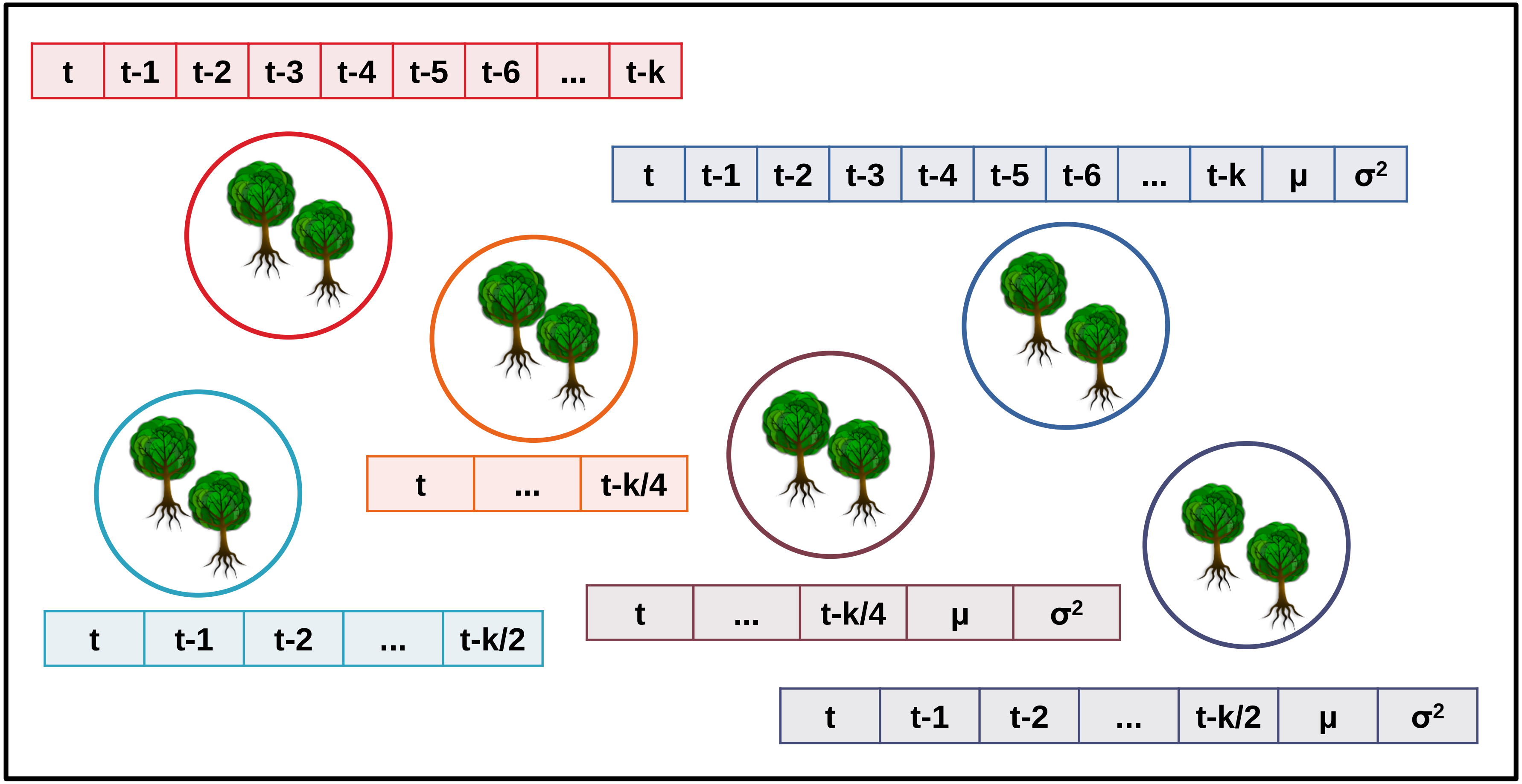
Bagging variants

DE+S



Bagging variants

DE±S



Experimental Evaluation

- **Data:** 14 real world time series;
- **Metric:** Standard Mean Squared Error (MSE);
- **Experimental procedure:** Monte Carlo simulations
  - randomly selected 10 points in time
  - training on the previous 50% observations
  - testing on the following 25%;
- **Statistical Significance:**
  - Wilcoxon signed rank tests with p-value < 0:05;
- **Tested setups:**
  - Different number of models in the ensemble (*M*);
  - Difference value of the maximum embed used (k_{max});

Time series

We use the **series of the differences** between successive values of each original time series;

Each series was treated **separately** from the others in their respective data source.

ID	Time series	Data source	Data characteristics
1	Temperature	Bike Sharing (Fanaee-T and Gama, 2013)	Daily values from Jan. 1, 2011 to Dec. 31, 2012 (731 values)
2	Humidity		
3	Windspeed		
4	Count of total bike rentals		
5	Temperature		Hourly values from Jan. 1, 2011 to Dec. 31, 2012 (7379 values)
6	Humidity		
7	Windspeed		
8	Count of total bike rentals		
9	Flow of Vatnsdalsa river	Icelandic river (Tong et al., 1985)	Daily values from Jan. 1, 1972 to Dec. 31, 1974 (1095 values)
10	Minimum temperature	Porto weather[1]	Daily values from Jan. 1, 2010 to Dec. 28, 2013 (1457 values)
11	Maximum temperature		
12	Maximum steady wind		
13	Maximum wind gust		
14	Total precipitation		

Results

M	k_{max}	Variant	Wins/Losses
1020	20	E+S	13 (11) / 1 (1)
		DE	7 (7) / 7 (3)
		DE+S	13 (10) / 1 (0)
		DE±S	14 (12) / 0 (0)
		ARIMA	7 (3) / 7 (4)
	30	E+S	11 (9) / 3 (2)
		DE	10 (6) / 4 (3)
		DE+S	10 (5) / 4 (2)
		DE±S	10 (9) / 4 (2)
		ARIMA	6 (3) / 8 (4)

M	k_{max}	Variant	Wins/Losses
1500	20	E+S	13 (10) / 1 (1)
		DE	8 (6) / 6 (3)
		DE+S	13 (10) / 1 (0)
		DE±S	14 (12) / 0 (0)
		ARIMA	7 (3) / 7 (4)
	30	E+S	11 (9) / 3 (2)
		DE	9 (7) / 5 (3)
		DE+S	10 (7) / 4 (2)
		DE±S	10 (9) / 4 (2)
		ARIMA	6 (3) / 8 (4)

Paired comparisons:
Nr.Wins (Statistically Significant Wins)/ Nr.Losses (Statistically Significant Losses)

Results

M	k_{max}		E	E+S	DE	DE+S	DE±S	ARIMA
1020	20	mean	4.36	**2.00**	4.21	2.29	2.14	3.50
		sd	*0.84*	1.18	0.89	1.07	0.86	2.59
	30	mean	3.93	**2.29**	3.64	2.57	2.57	3.86
		sd	1.44	1.27	1.34	*1.16*	1.28	2.57
1500	20	mean	4.43	**2.00**	4.14	2.29	2.14	3.50
		sd	*0.65*	1.18	1.03	1.07	0.86	2.59
	30	mean	3.86	**2.36**	3.79	2.64	**2.36**	3.86
		sd	1.41	1.28	1.42	1.28	*1.01*	2.57

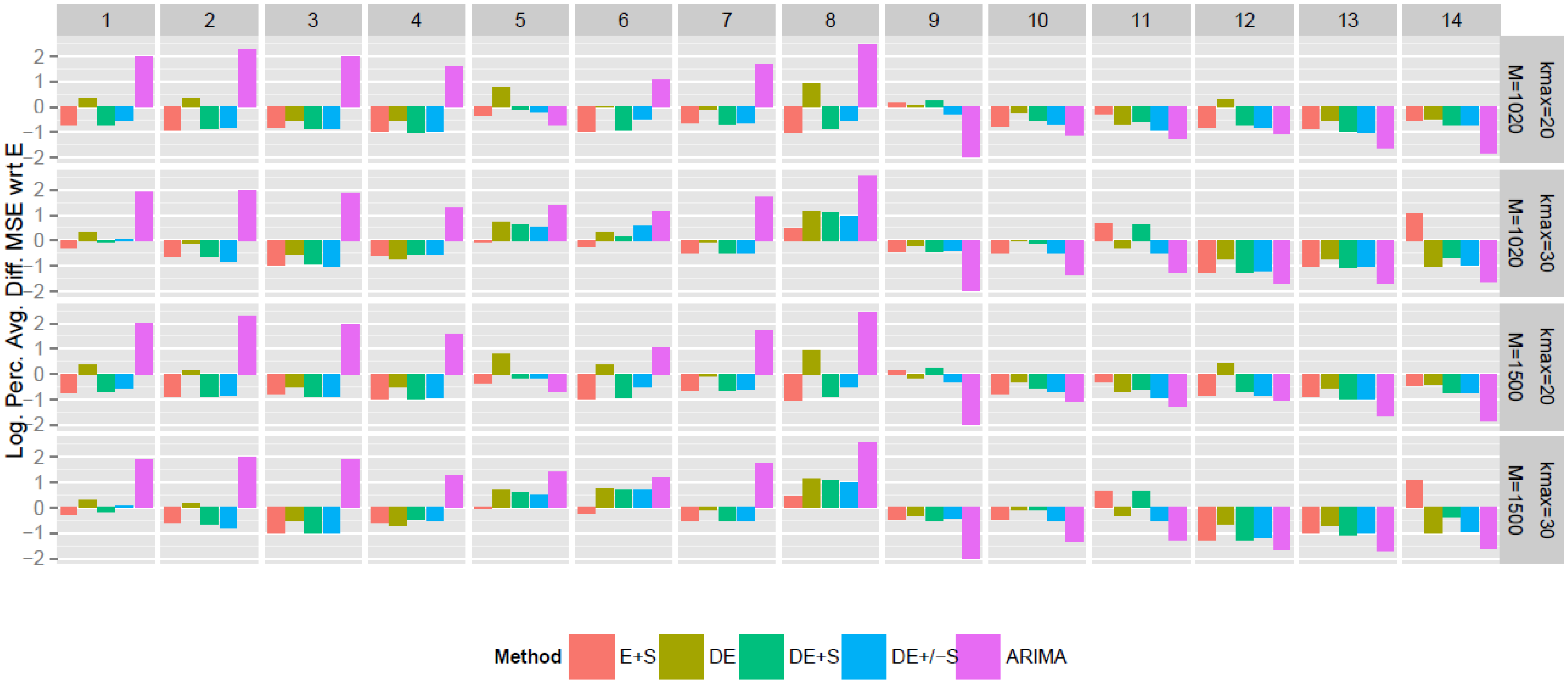
Average and standard deviation of rank for each method

Results

M	k_{max}		E+S	DE	DE+S	DE±S	ARIMA
1020	20	mean	**-4.74**	0.22	-4.54	-4.34	36.26
		sd	3.00	3.02	2.83	*2.59*	101.64
	30	mean	-2.30	-0.55	-2.23	**-3.08**	32.63
		sd	6.44	*5.20*	6.96	5.83	110.77
1500	20	mean	**-4.77**	0.27	-4.62	-4.36	36.24
		sd	2.98	3.11	2.80	*2.61*	101.57
	30	mean	-2.28	-0.15	-1.94	**-3.03**	32.76
		sd	6.39	*5.00*	7.01	5.97	110.80

Average and standard deviation of mean percentual difference wrt to the baseline

Results



$$\mathrm{sgn}(MSE_x - MSE_E).\log\left(\left|\frac{100.(MSE_x - MSE_E)}{MSE_E}\right| + 1\right)$$

Conclusion

- Proposed initial set of forms of injecting diversity into ensembles that take into account specific challenges posed by time series;
- The recent dynamics of a time series is represented using
  - different embed sizes and
  - the addition of variables summarizing the recent observed values;
- This was implemented and tested in the context of bagging regression trees, obtaining a clear advantage over standard bagging in real world data;
- Our results suggest this is a promising research direction.

Future work

- Exploring the possibility of
  - changing the amount of past data used by each model (varying training windows);
  - making the aggregation of the predictions time-dependent;
  - using other types of predictor variants.

Try it yourself:

- All code and data necessary to replicate all the results presented available at http://www.dcc.fc.up.pt/~ltorgo/ACML2014/
- All programs are written in the free and open source R software environment.

Thank you!