Method Description

General Information

Type of Entry (*Academic, Practitioner, Researcher, Student*)	**Practitioner**
First Name	**Remo**
Last Name	**Fritschi**
Country	**Switzerland**
Type of Affiliation (*University, Company-Organization, Individual*)	**Individual**
Affiliation	**-**

Team Members (*if applicable*):

1st Member	
First Name	
Last Name	
Country	
Affiliation	
2nd Member	
First Name	
Last Name	
Country	
Affiliation	

Information about the method utilized

Name of Method	**Out-sample validated ARIMA**
Type of Method (*Statistical, Machine Learning, Combination, Other*)	**Statistical**
Short Description (up to 200 words)	**Simulation of different ARIMA models up to max ARMA(10,10) and selection using Ljungbox statistics and ACF. For the statistical top 3 models plus models up to ARMA(4,4) an out-sample simulation is calculated. The model with the smallest error (calculated with SMAPE and MASE) is selected as the final model.**

Extended Description:

Main idea:

For each time series, I create a model table: (example series W1)

AR	MA	LB_out	LB_height	ACF_in	ACF_mean	cv_mse	cv_smape	cv_mase
10	2	1	0,9780404	0,8181818	-0,0294579	0,0003584	1,7525374	1,0796457
3	9	1	0,9458985	0,8484848	-0,0285917	0,0003613	1,7512703	1,095053
2	10	1	0,9977312	0,7878788	-0,0272689	0,0003893	1,7356896	1,1063391
4	1	0,6923077	0,6223803	0,7878788	-0,0352829	0,0003539	1,8381223	1,091733
4	2	0,6923077	0,6443045	0,7575758	-0,0349566	0,0003537	1,8051984	1,0838248
4	3	0,6923077	0,6438413	0,7575758	-0,0350152	0,0003538	1,805284	1,0839271
1	4	0,5384615	0,2968972	0,7575758	-0,0407692	0,0003457	1,8573543	1,083296
2	4	0,5384615	0,3698392	0,6969697	-0,0389164	0,0003546	1,8687037	1,0924669
3	4	0,5384615	0,3902882	0,6969697	-0,0382141	0,0003545	1,8684607	1,0957512
4	4	0,6923077	0,678667	0,7575758	-0,0345563	0,0003543	1,8060596	1,0843517

The final model for the prediction in this case is the ARMA(10,2)

How to get this table:

1. Simulate models from ARMA(1,1) up to ARMA(10,10)
2. Calculate the statistical values (yellow)
    a. LB_out: Ljungbox-Values over 0.05 (share)
    b. LB_height: mean of Ljungbox-Values
    c. ACF_in: ACF-Values inside 0.05 (share)
    d. ACF_mean: calculate the mean of the ACF
3. With the values in section 2, I take to sum. This gives me the best statistical model – I believe.
4. A former teacher said, ARMA models up to 4,4 are good enough. So, I include also this models.
5. “Crossvalidation”: I do an out-sample validation. Meaning, how well would the model have performed in the past.
6. For the Error-Functions, I take your functions: smape and mase.
7. Finally, the Model with the lowest error wins.

Each time series was calculated parallel. I used the google cloud platform with an 8-core processor. Each series calculation writes a file with the prediction, a log file and and the table you have seen before.

Program-Files:

parallel.R – The actuall Run-File where you have to specify the paths and you will find a parameter-cockpit
functions.R – all the functions, which are run in parallel
createSubmission.R – takes all the output-files and creates a submission.csv