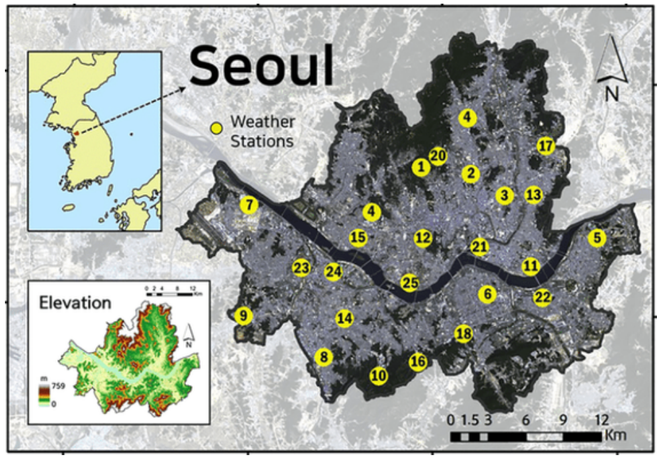
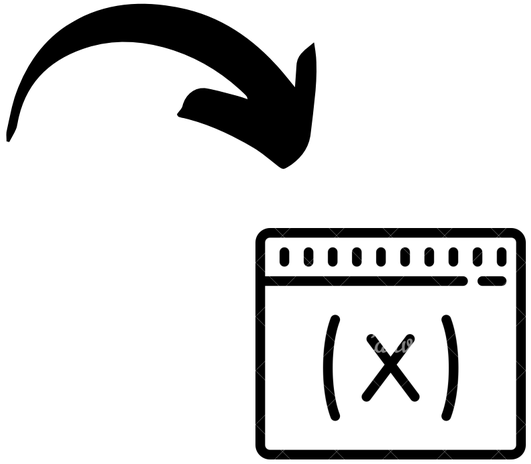
Bias correction of numerical prediction model temperature forecast Data Set

Gautier Poursin & Luca Micciche

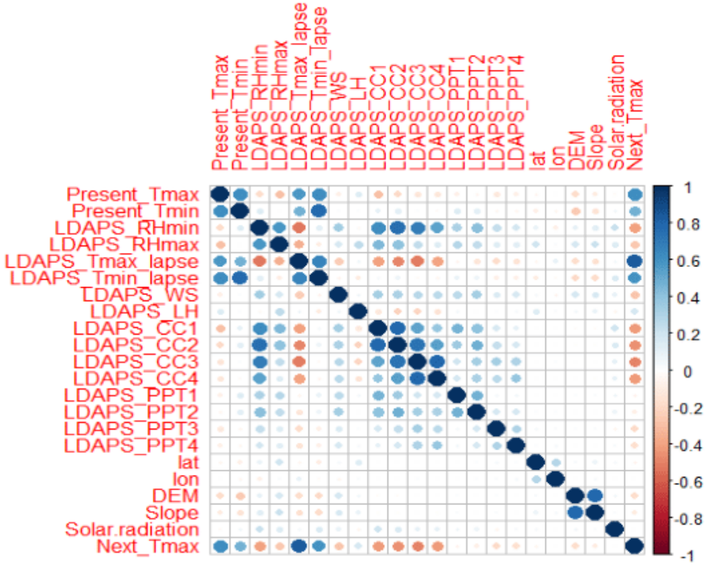




23-covariables

LDAPS

Auxiliary



Correlation for Tmax

2 TARGET

Next_Tmin

Next_Tmax

Bias correction of numerical prediction model

temperature forecast Data Set

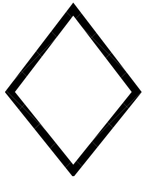
Methodology - The Methods ?

About the Algorithm

Linear/Stepwise

Lasso

K-Nearest Neighboor



Ridge

About the models

All the data

Per years

Per stations

With clusters

Statistical indicators

$$BIAS = \frac{\sum_{i=1}^{n}(\widehat{yi} - yi)}{n}$$

$$MAE = \frac{\sum_{i=1}^{n}|\widehat{yi} - yi|}{n}$$

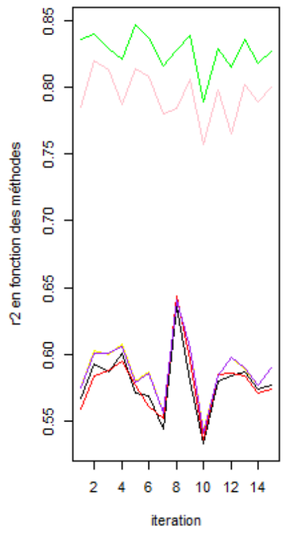
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \widehat{yi})^2}{\sum_{i=1}^{n}(yi - \bar{yi})^2}$$

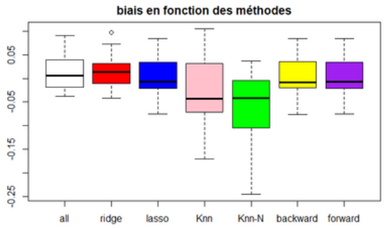
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{yi} - yi)^2}{n}}$$

Bias correction of numerical prediction model temperature forecast Data Set

Models without clusters

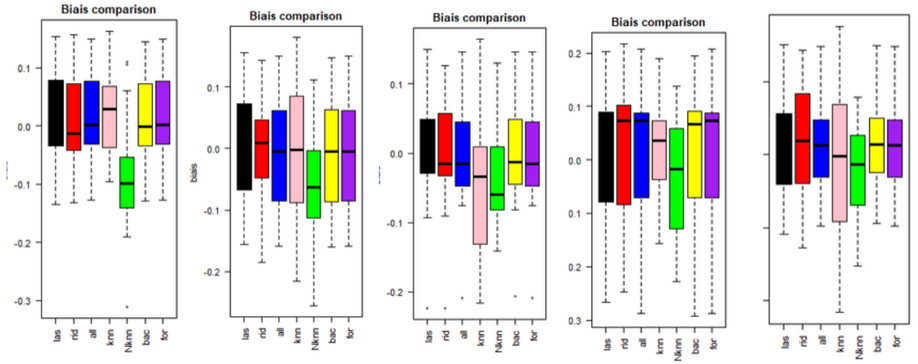
With all the data





- Bias and RMSE are very low for all
- give a first impression of the results

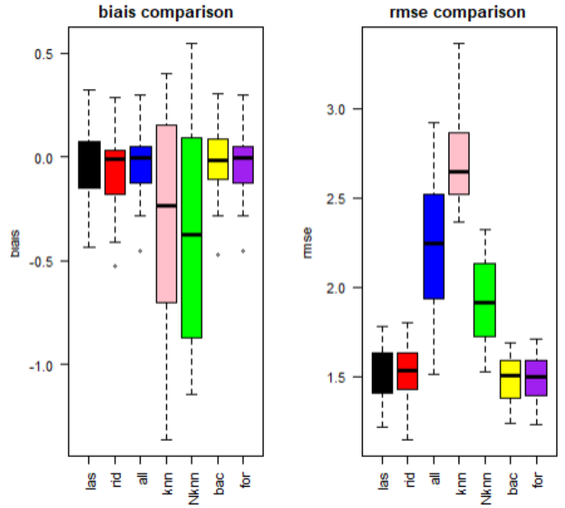
Per year



2013 - 2014 - 2015 - 2016 - 2017

- lot of disparity on the bias
- reducing the generality of the model

Per stations



- the bias is very scattered
- there is a lost of informations

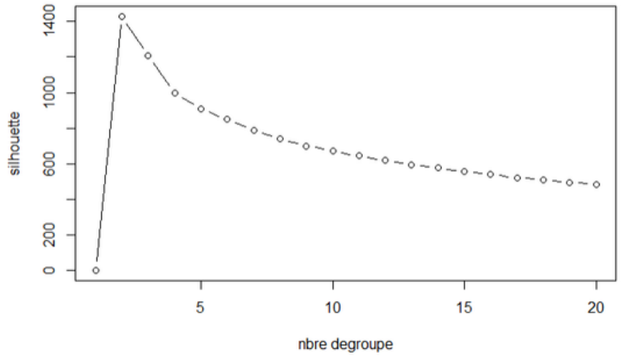
Results are not relevant: new idea => **computing clusters**

Bias correction of numerical prediction model

temperature forecast Data Set

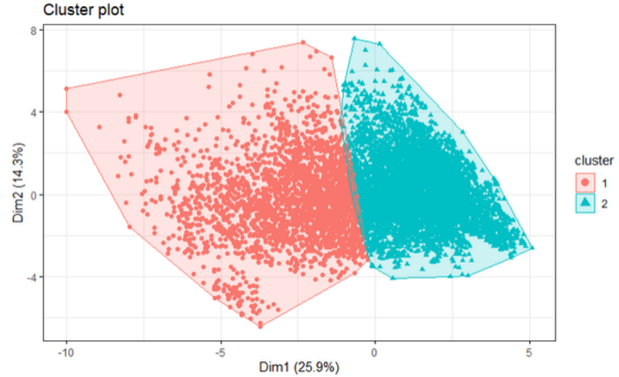
The chosen methods - Cluster

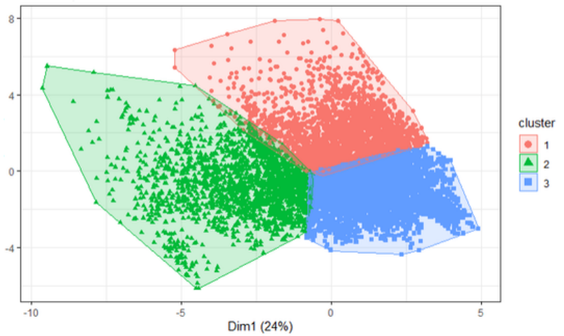
Silhouette

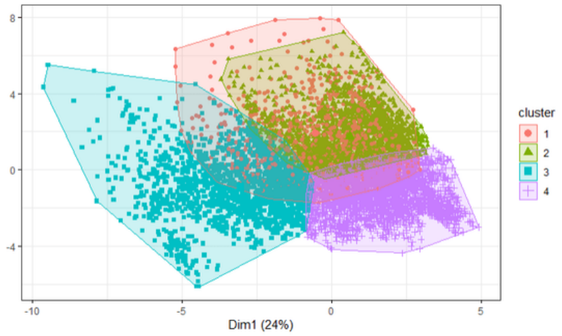


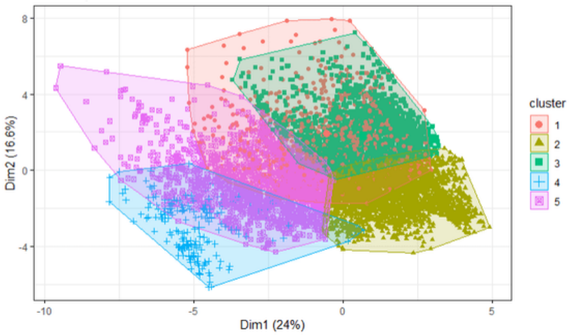
K < 4

K > 3









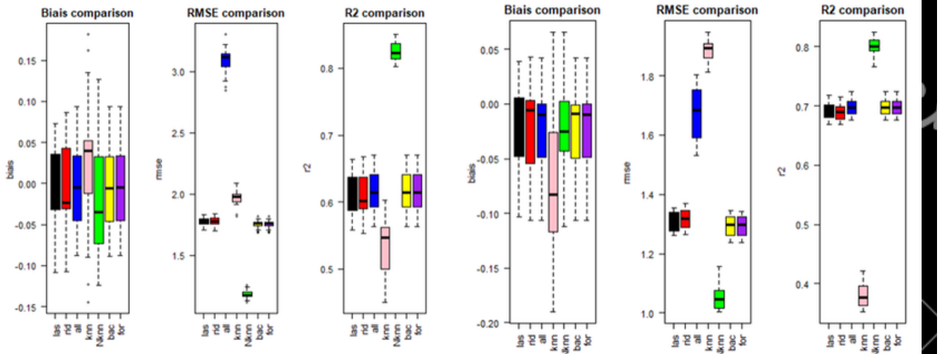
- 2/3 several clusters
- sharing small borders

- no separate clusters
- sharing borders

Bias correction of numerical prediction model temperature forecast Data Set

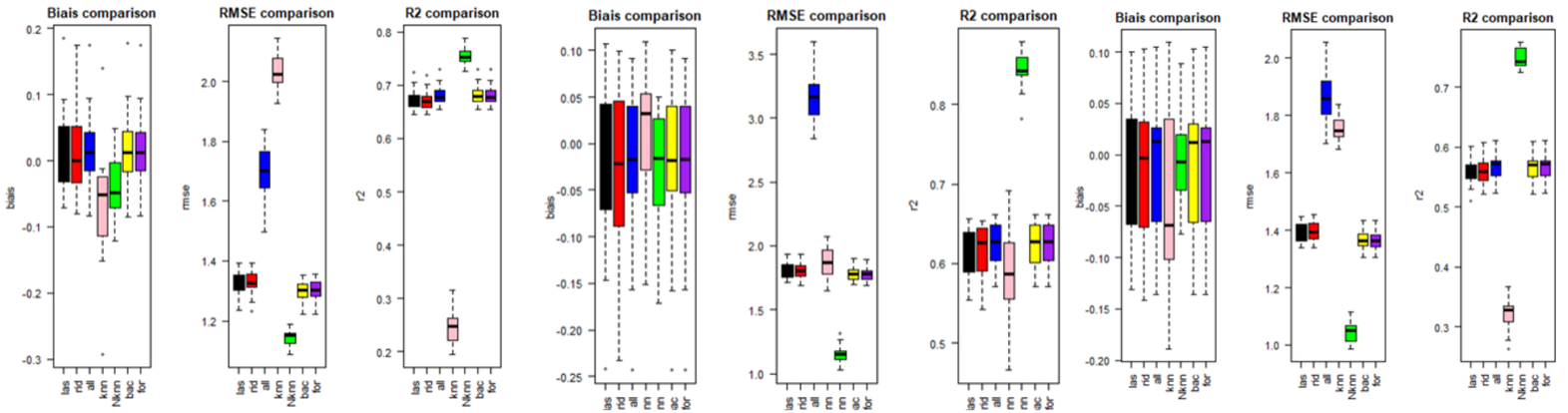
How many clusters ?

With K=2 clusters



- seems to be less scattered
- values seems to be better

With K=3 clusters

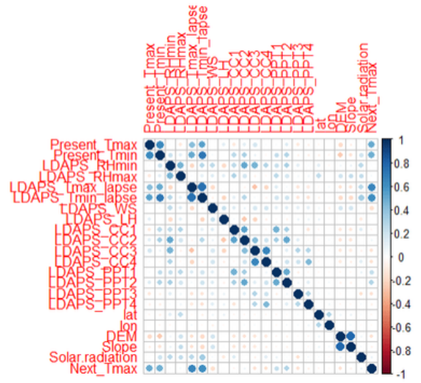


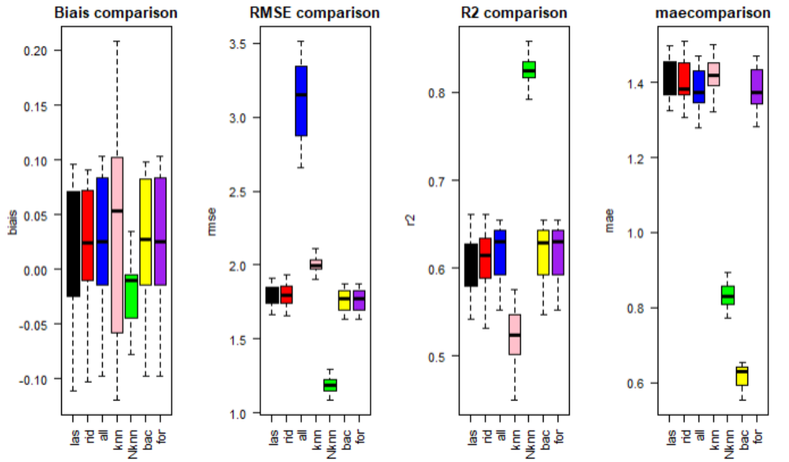
- lot of disparity

Choice : 2 clusters

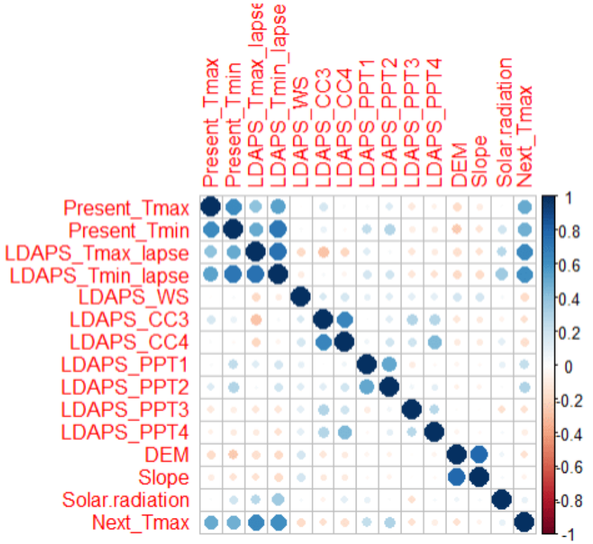
Bias correction of numerical prediction model temperature forecast Data Set

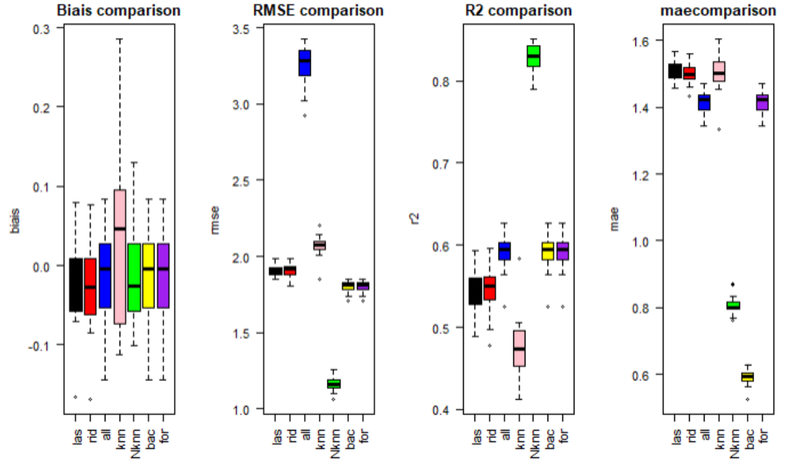
Variable selection with cluster 1





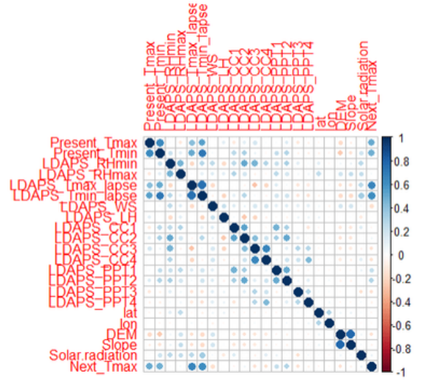
Selection of variables:

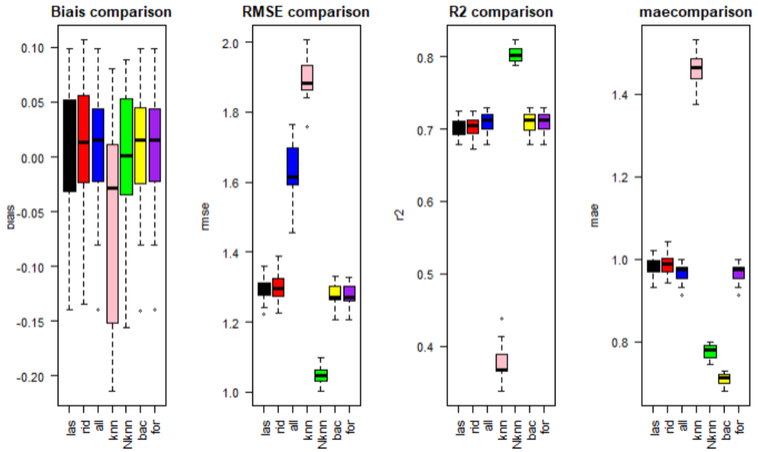




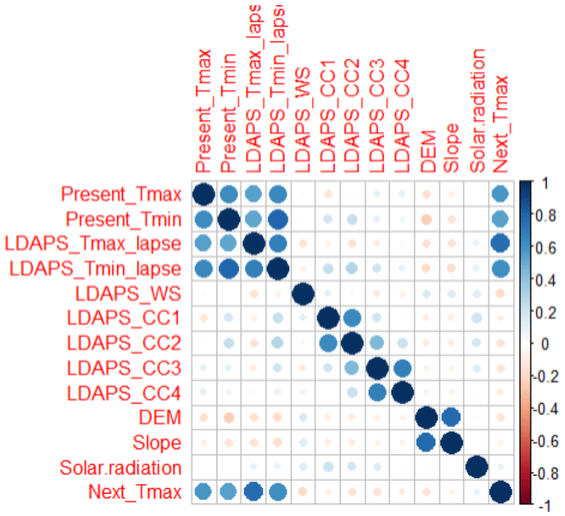
Bias correction of numerical prediction model temperature forecast Data Set

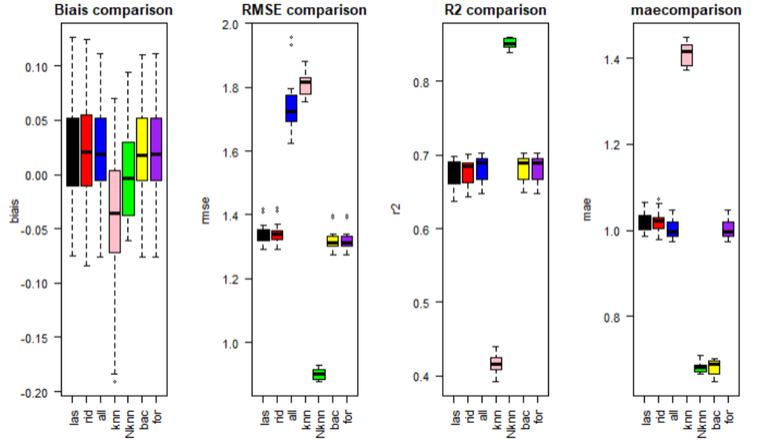
Variable selection with cluster 2





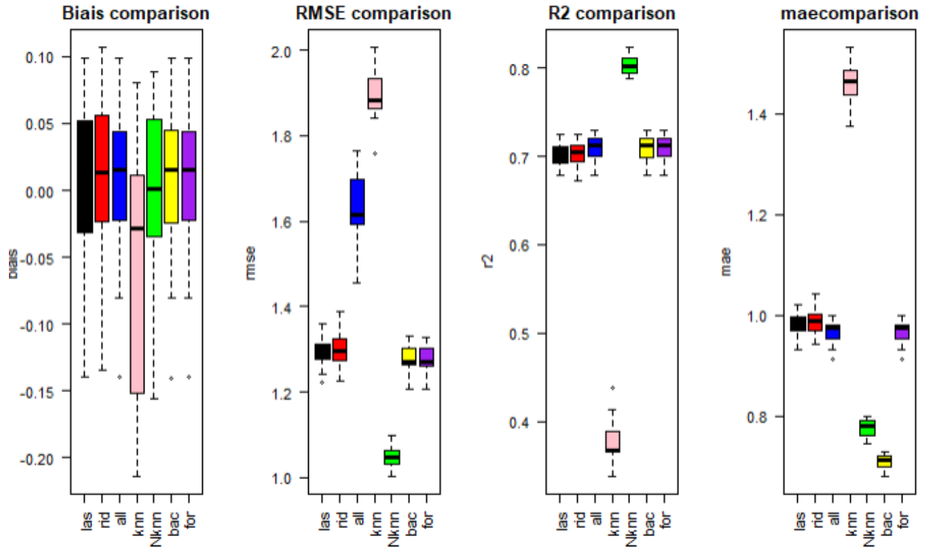
Selection of variables:



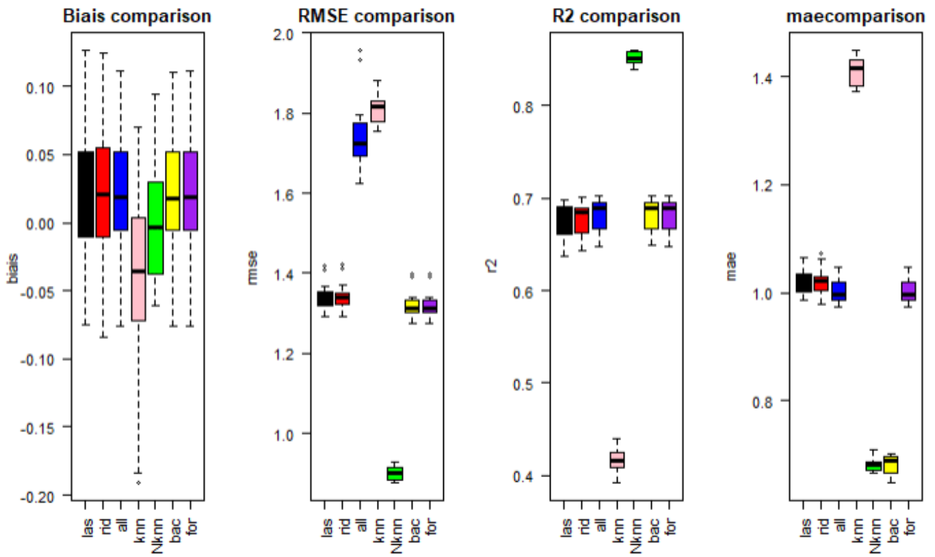


Bias correction of numerical prediction model temperature forecast Data Set

How to choose a method and how to validate a model?



Model with all variables



Selection of variables

- Similarities between Bias, RMSE, R2 and MAE
- Medians are closer

} **We can validate the model**

- KNN_Scale has the best RMSE and R2. His MAE is smaller than 1.
- All bias are close to 0

} **KNN_Scale is the best regression method**

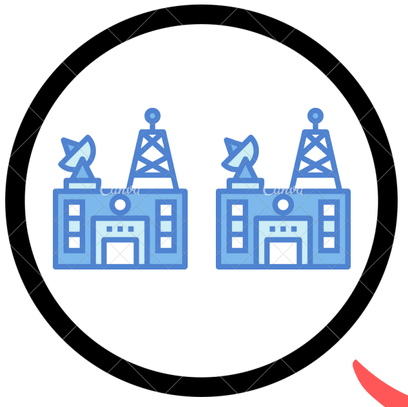
Bias correction of numerical prediction model

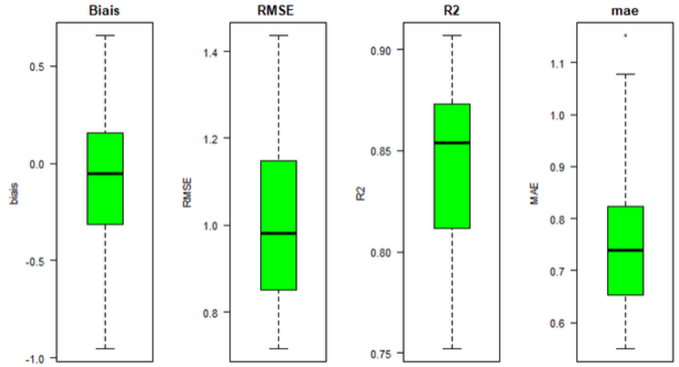
temperature forecast Data Set

Cross-Validation

20 stations to predict 5 stations x50

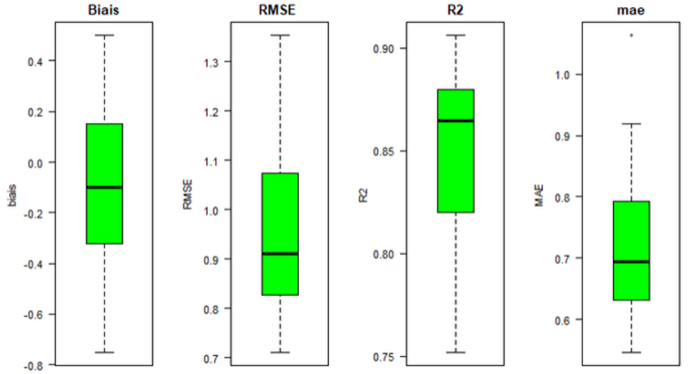
cluster 1





Bias= -0.03 RMSE=0.98 R2=0.86 MAE= 0.75

cluster 2



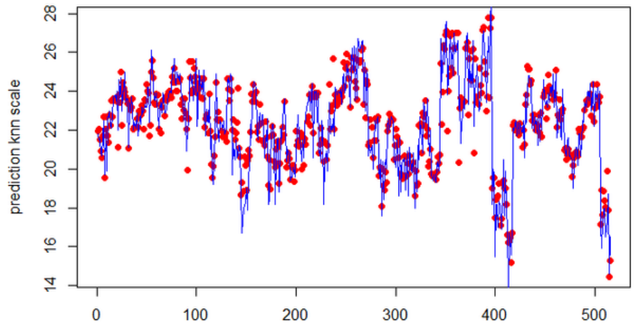
Bias= -0.09 RMSE=0.92 R2=0.87 MAE= 0.7

Bias correction of numerical prediction model temperature forecast Data Set

To summarize

Next_Tmin

Cluster 1



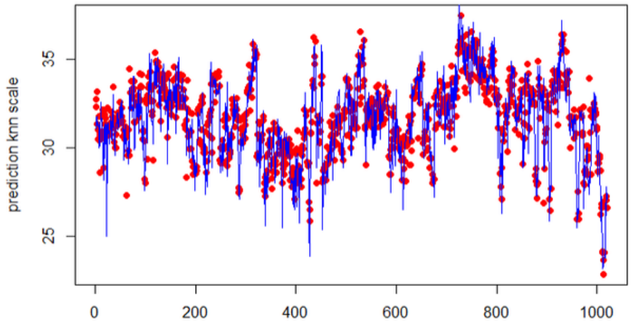
Model:

- Present_Tmin/Tmax
- LDAPS_RH_Min/Max
- LDAPS_TMin/TMax_lapse
- LDAPS_LH
- LDAPS_CC2
- LDAPS_PPT1/2/3/4
- DEM
- Slope
- Solar.Radiation

Bias= -0.09 RMSE=0.84 R2=0.86 MAE= 0.61

Next_Tmax

Cluster 1



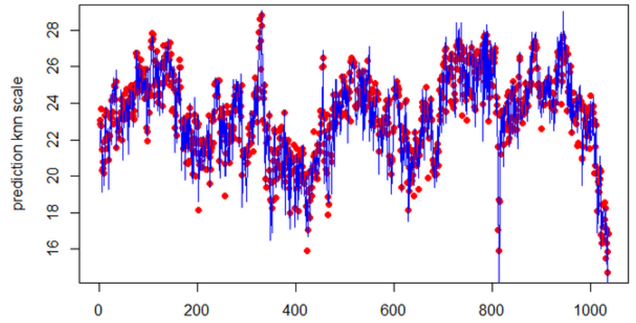
Model:

- Present_Tmin/Tmax
- LDAPS_TMin/TMax_lapse
- LDAPS_WS
- LDAPS_CC3/4
- LDAPS_PPT1/2/3/4
- DEM
- Slope
- Solar.Radiation

Bias= 0.007 RMSE=1.01 R2=0.81 MAE= 0.75

Next_Tmin

Cluster 2



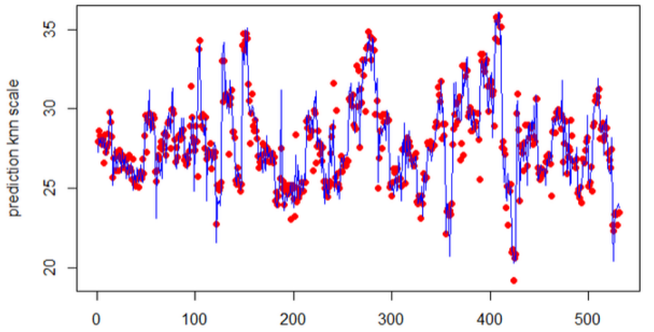
Model:

- Present_Tmin/Tmax
- LDAPS_RH_Min
- LDAPS_TMin/TMax_lapse
- LDAPS_CC1/2/3/4
- Slope
- Solar.Radiation

Bias= -0.08 RMSE=0.88 R2=0.87 MAE= 0.67

Next_Tmax

Cluster 2



Model:

- Present_Tmin/Tmax
- LDAPS_TMin/TMax_lapse
- LDAPS_WS
- LDAPS_CC3/4
- LDAPS_PPT1/2/3/4
- DEM
- Slope
- Solar.Radiation

Bias= 0.016 RMSE=1.00 R2=0.87 MAE= 0.70