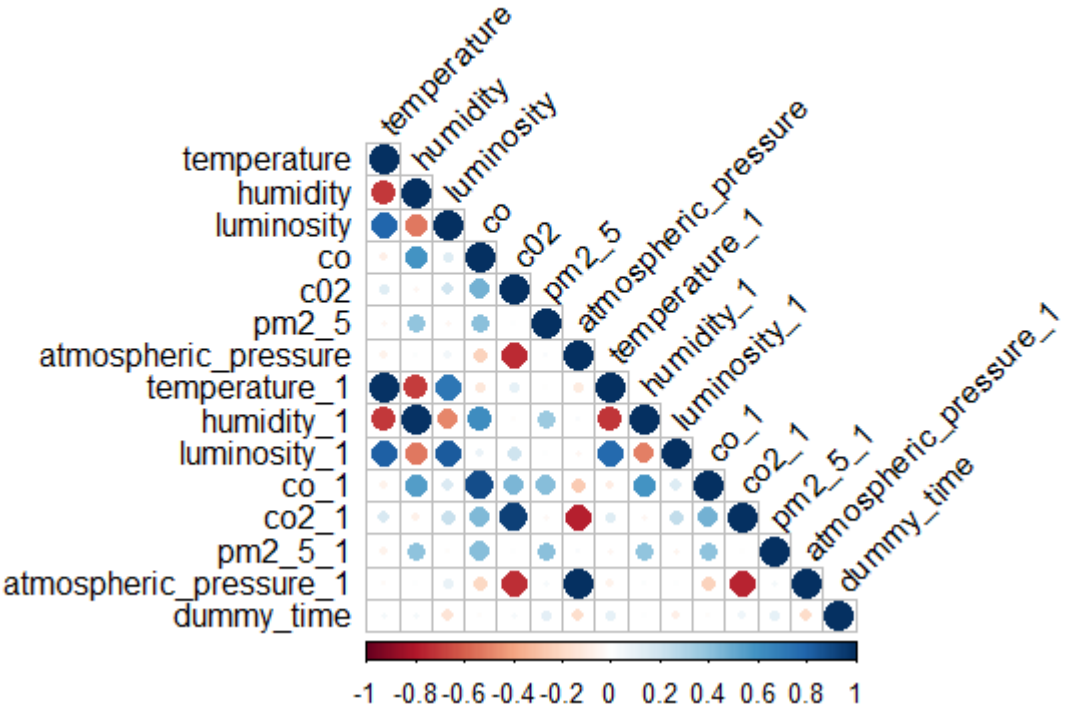
Ever Weather

Βελτίωση μοντέλου με τις μεθόδους PCR & PLSR

Περιεχόμενα

- Training dataset summary
- PCR theory
- PLS theory
- Scree Plot
- Cross Validation (Temperature, Humidity, Luminosity, Atm_pressure)
- Results Table
- Predictions Plots/ Weather Forecast

Corr Plot


temperature
humidity
luminosity
co
c02
pm2_5
atmospheric_pressure
temperature_1
humidity_1
luminosity_1
co_1
co2_1
pm2_5_1
atmospheric_pressure_1
dummy_time
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

PCR

- Principal components regression (**PCR**) is a regression technique based on principal component analysis (**PCA**).
- The basic idea behind **PCR** is to calculate the **principal components** and then use some of these components as predictors in a linear regression model fitted using the typical least squares procedure.
- Some of the most notable advantages of performing PCR are the following:

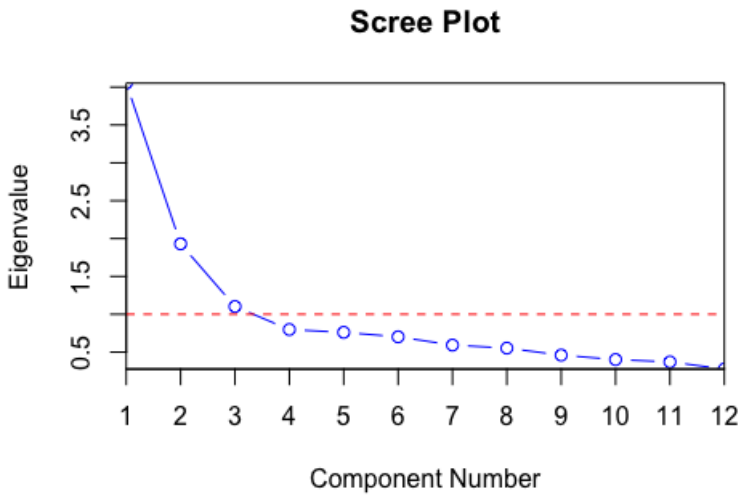
    1. Dimensionality reduction
    2. Avoidance of multicollinearity between predictors
    3. Overfitting mitigation

PLS

- The Partial Least Squares Regression procedure estimates partial least squares (PLS, also known as "projection to latent structure") regression models. PLS is a predictive technique that is an alternative to ordinary least squares (OLS) regression, canonical correlation, or structural equation modeling, and it is particularly useful when predictor variables are highly correlated or when the number of predictors exceeds the number of cases.
- PLS combines features of principal components analysis and multiple regression. It first extracts a set of latent factors that explain as much of the covariance as possible between the independent and dependent variables. Then a regression step predicts values of the dependent variables using the decomposition of the independent variables.

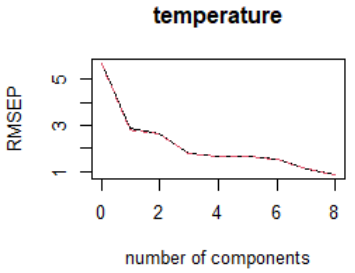
Scree Plot

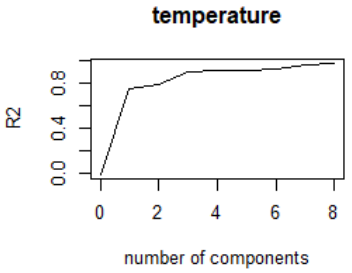
- The scree plot, which displays the variance explained by each principal component, helps you decide how many components to retain by identifying the "elbow" of the plot.
- The "elbow" is located at the point where the slope of the curve changes most abruptly from steep to flat data's variability.
- The components before the elbow are those that capture most of the important variability in the data.
- Components after the elbow are likely capturing noise and redundant information.
- The idea is to choose a number of components that provide a good balance between simplicity (fewer components) and retaining most of the important information in the data.

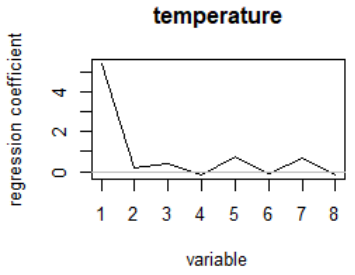


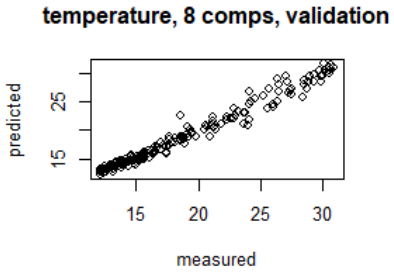
Temperature Cross Validation

PCR

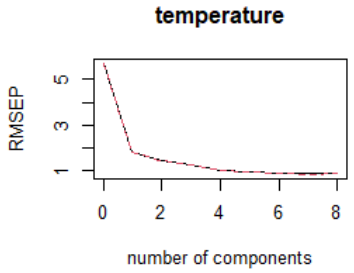


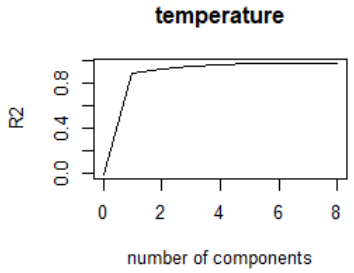


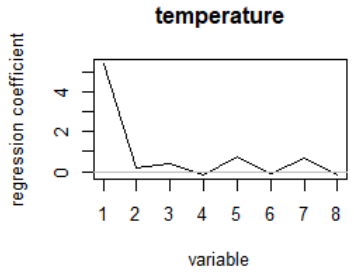


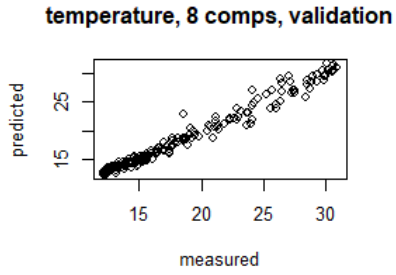


PLSR



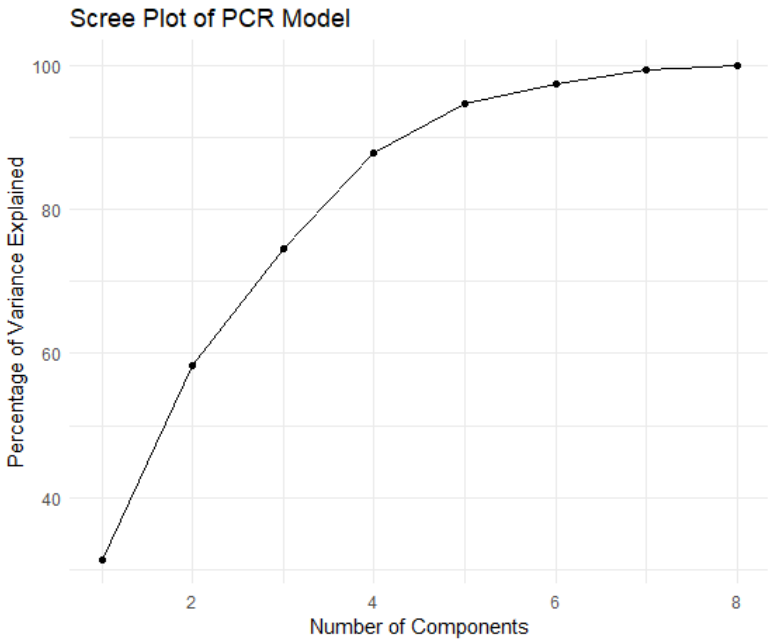




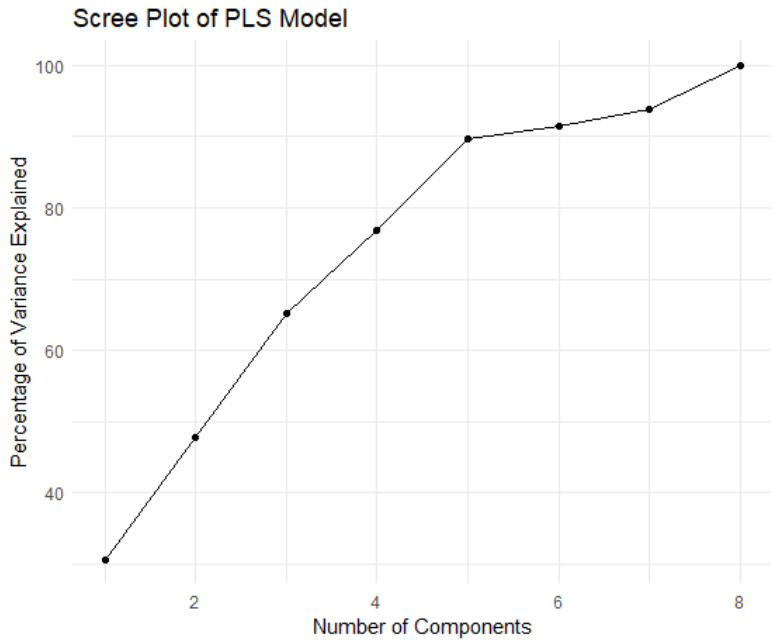


Temperature scree plots

PCR

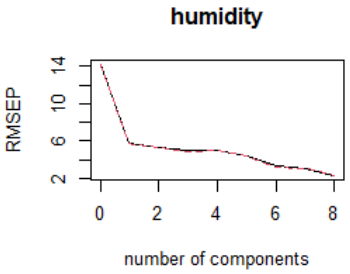


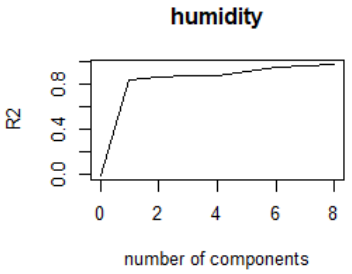
PLS

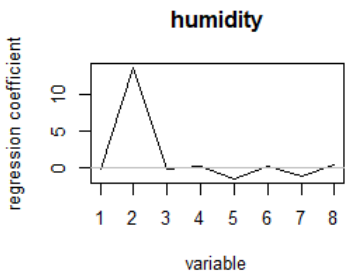


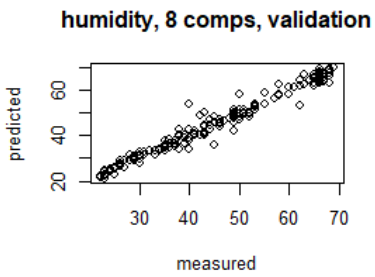
Humidity Cross Validation

PCR

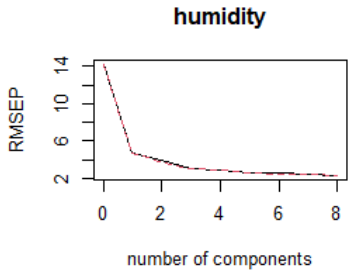


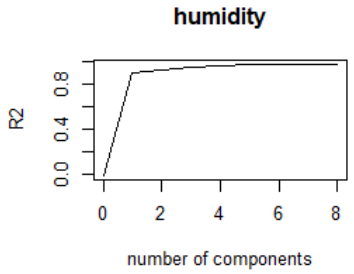


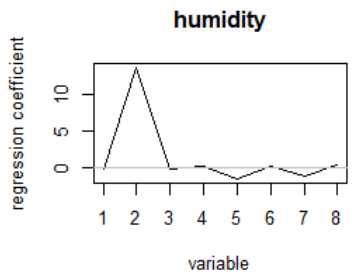


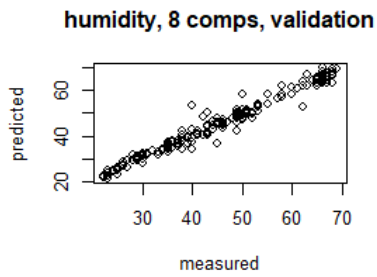


PLSR



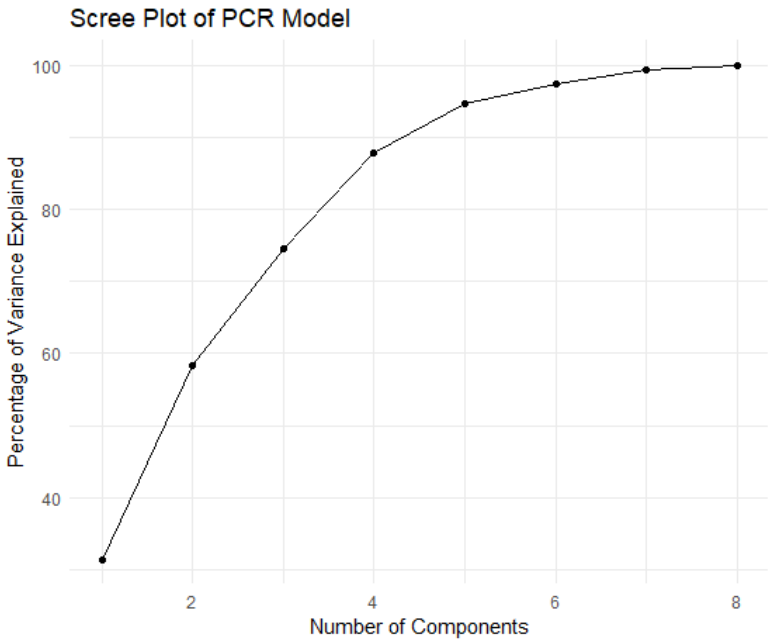




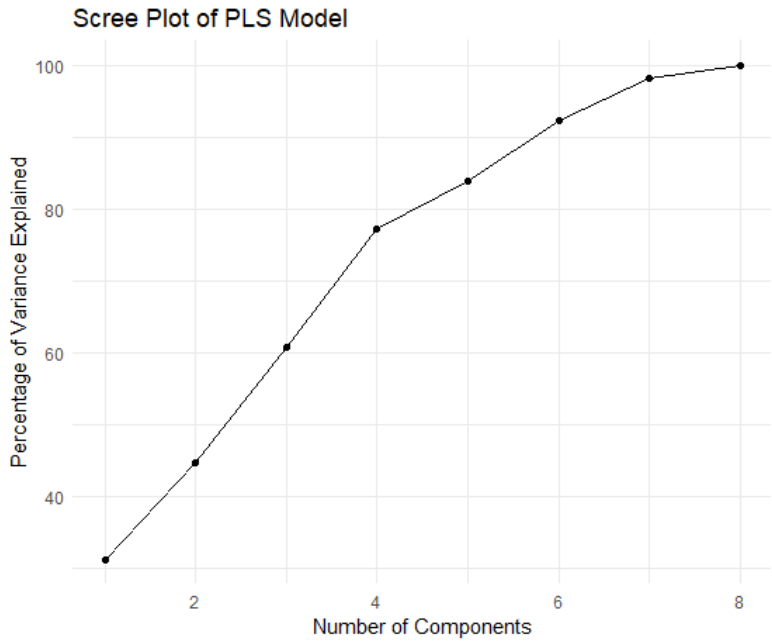


Humidity scree plots

PCR

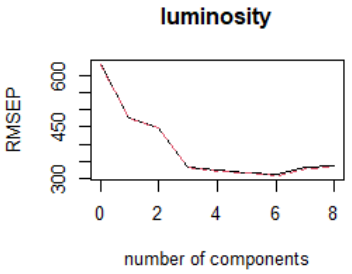


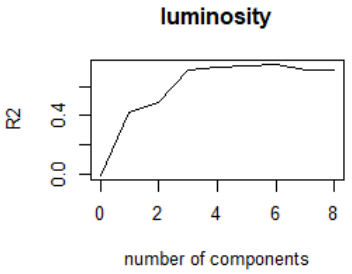
PLSR

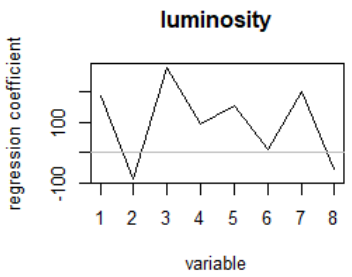


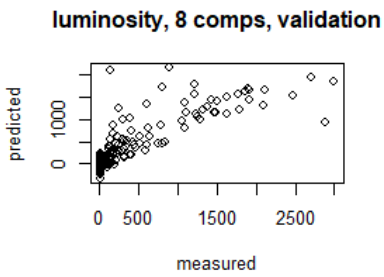
Luminosity Cross Validation

PCR

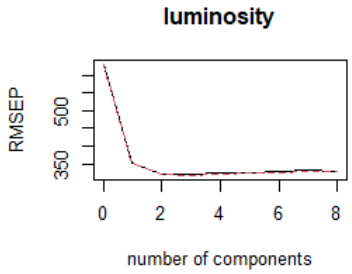


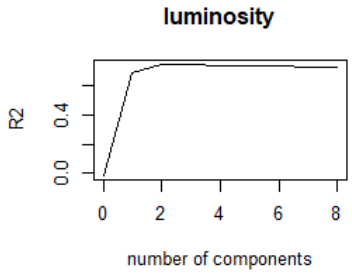


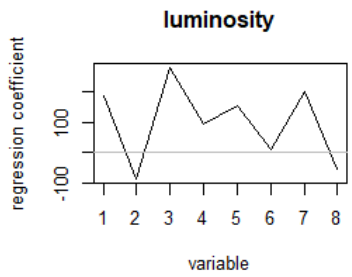


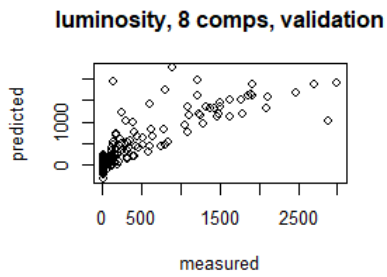


PLS



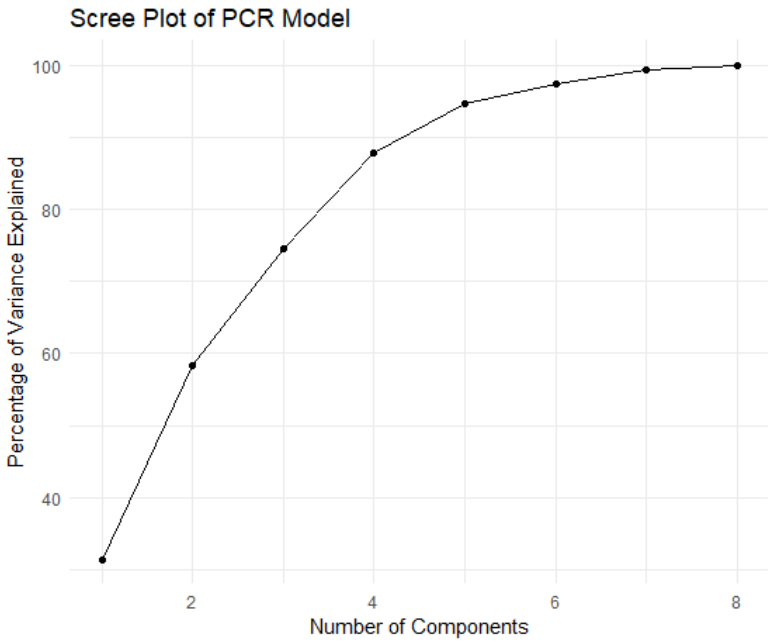




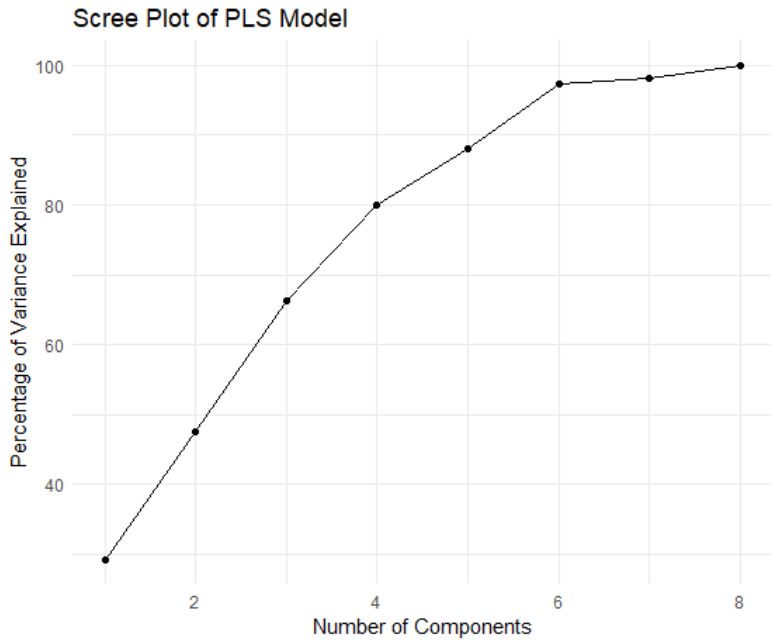


Luminosity Scree plot

PCR

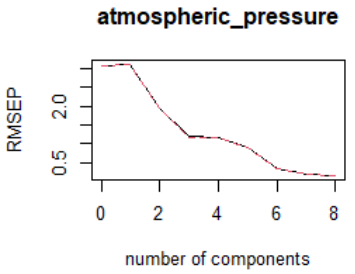


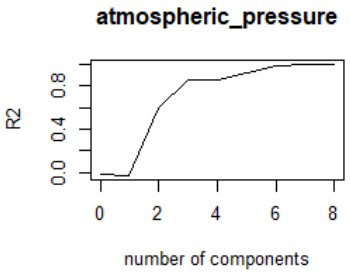
PLSR

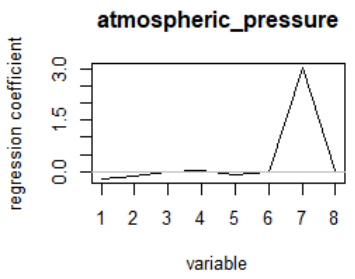


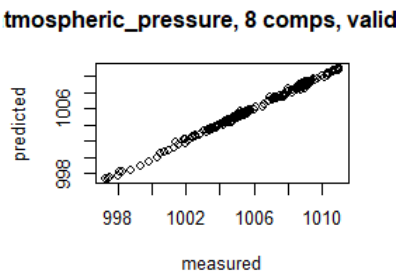
Atm_pressure Cross Validation

PCR

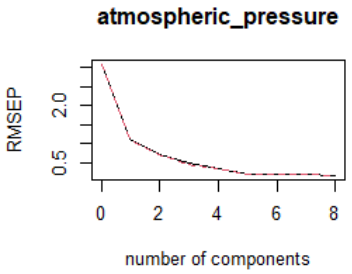


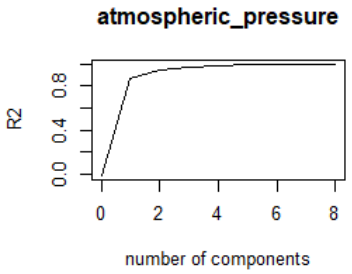


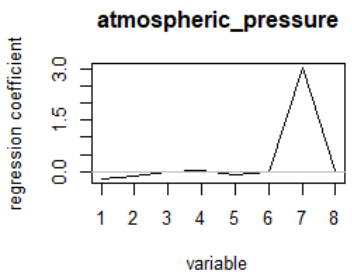


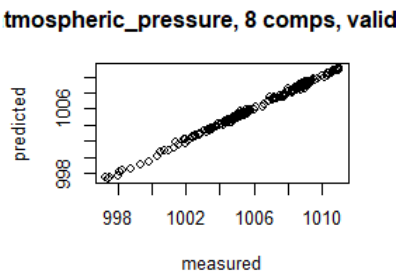


PLSR



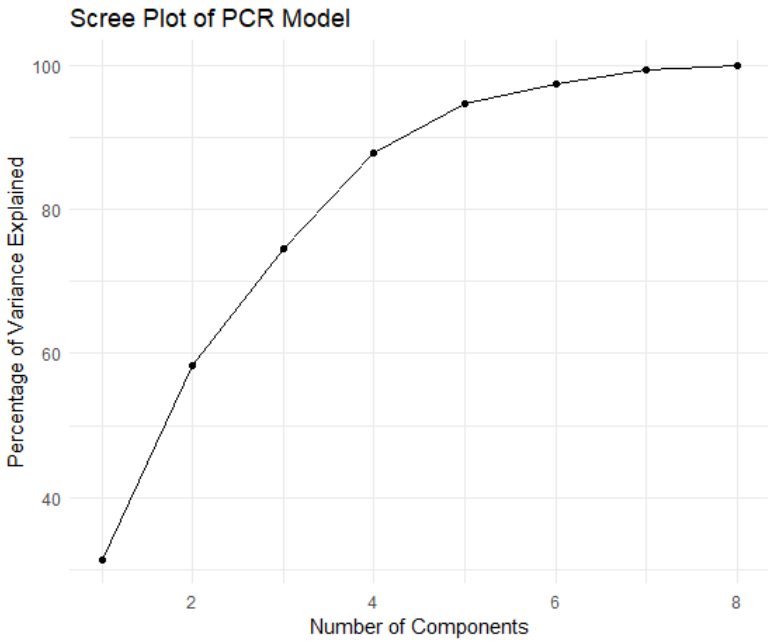




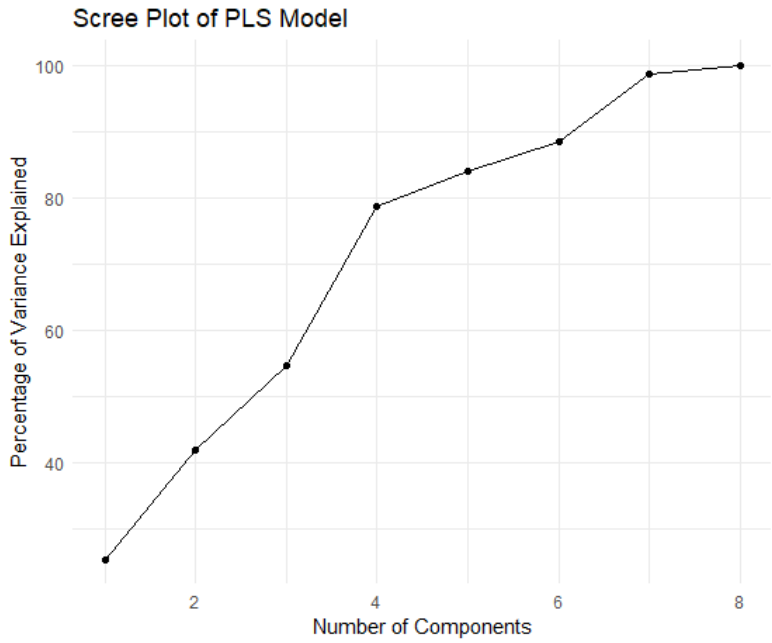


Atm_pressure scree plot

PCR



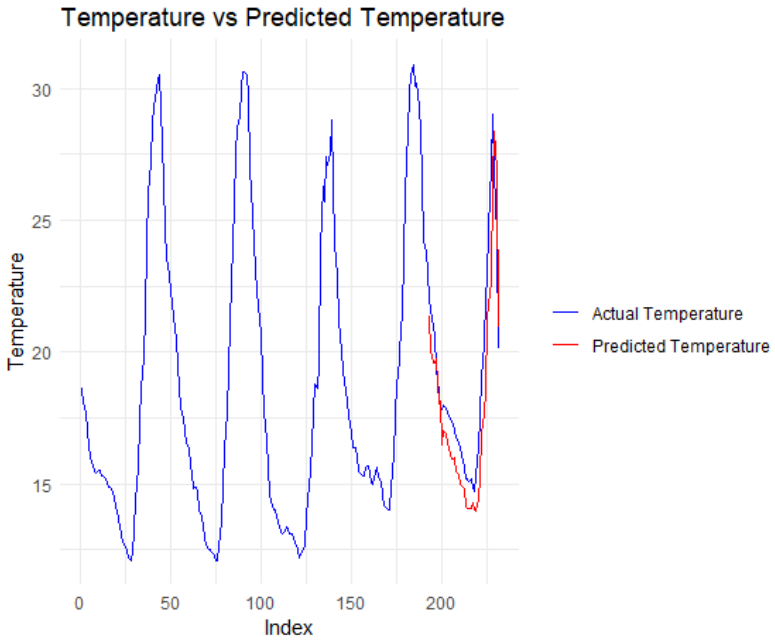
PLSR

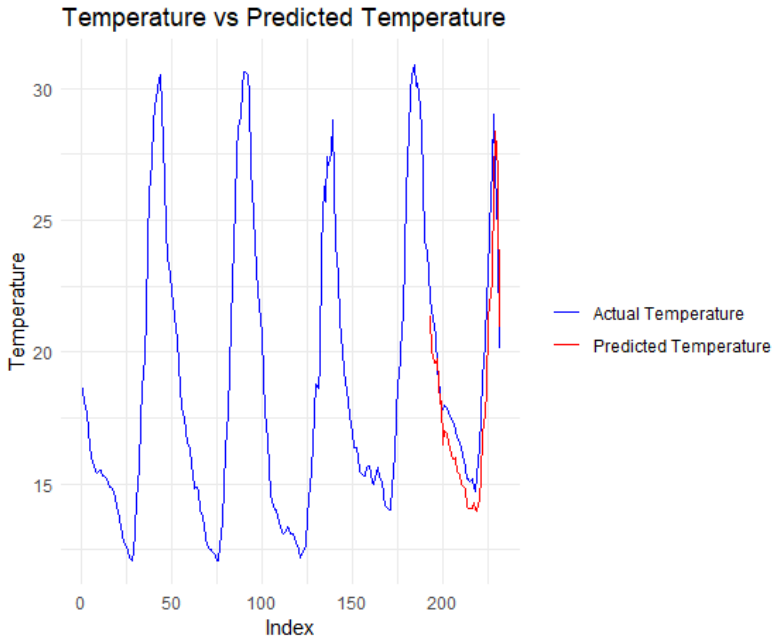


Results Table (Test)

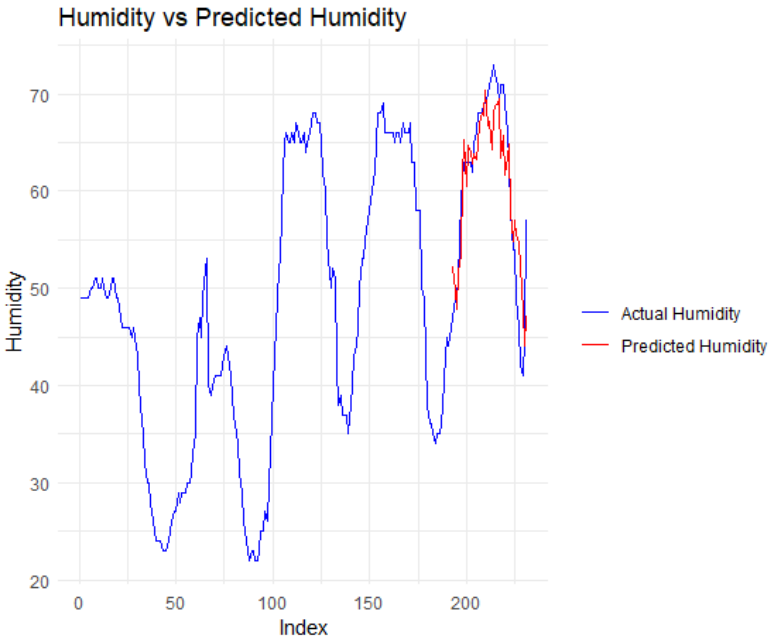
Target Variable	Comps	PCR RMSE	Comps	PLSR RMSE	Comps=8, MLR RMSE
Temperature	6	**1.70**	6	**1.34**	1.24
Humidity	6	**4.38**	6	**3.22**	2.91
Luminosity	4	**281.78**	4	**292.74**	303.62
Atm_pressure	4	**0.63**	4	**0.49**	0.37

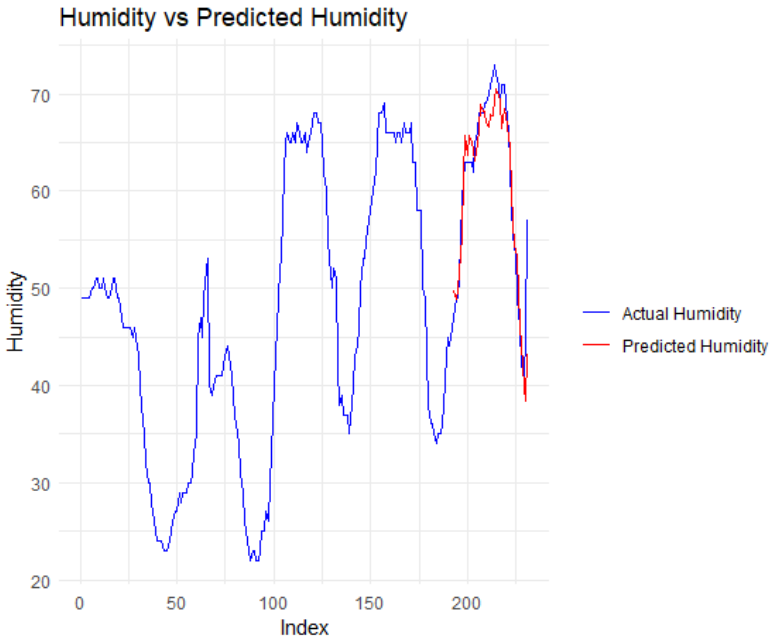
Temperature Forecasting



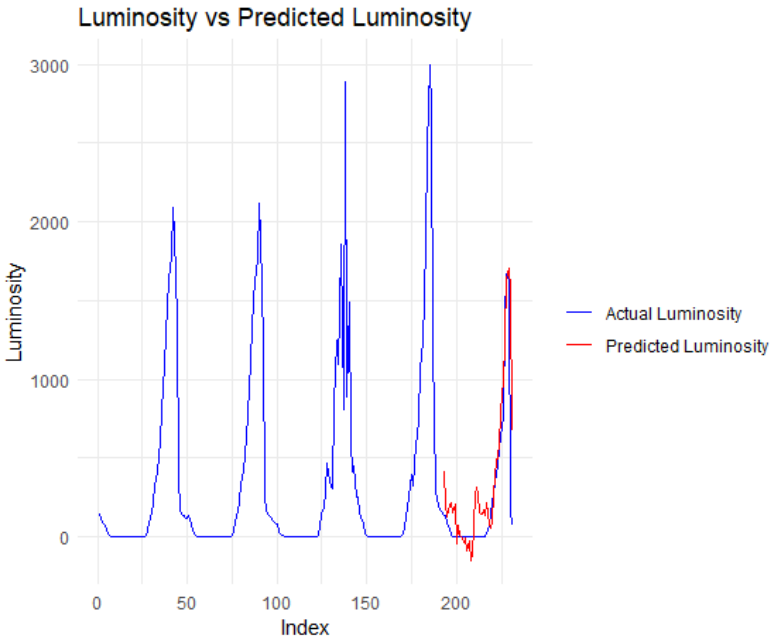


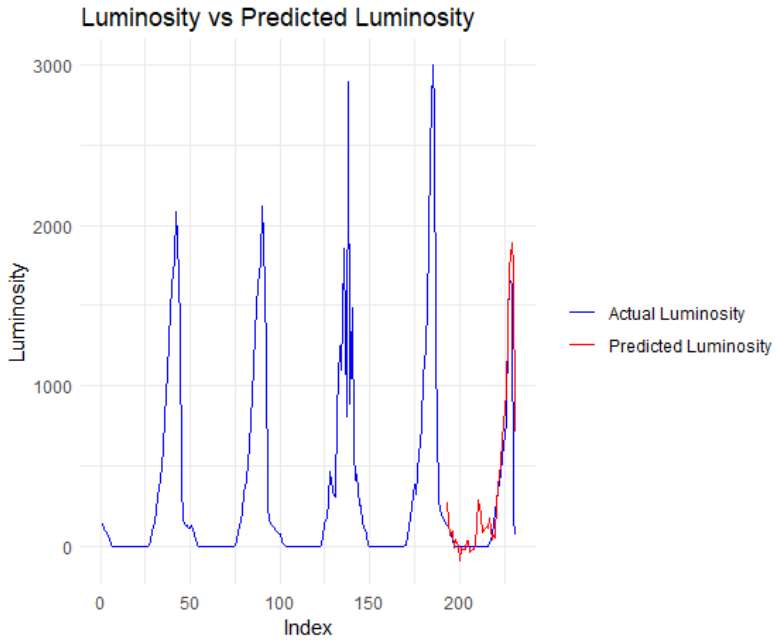
Humidity Forecasting





Luminosity Forecasting





Atm_Pressure Forecasting