Applied Geodata Science I

Session 10

Prof. Dr. Benjamin Stocker

Spring semester 2023

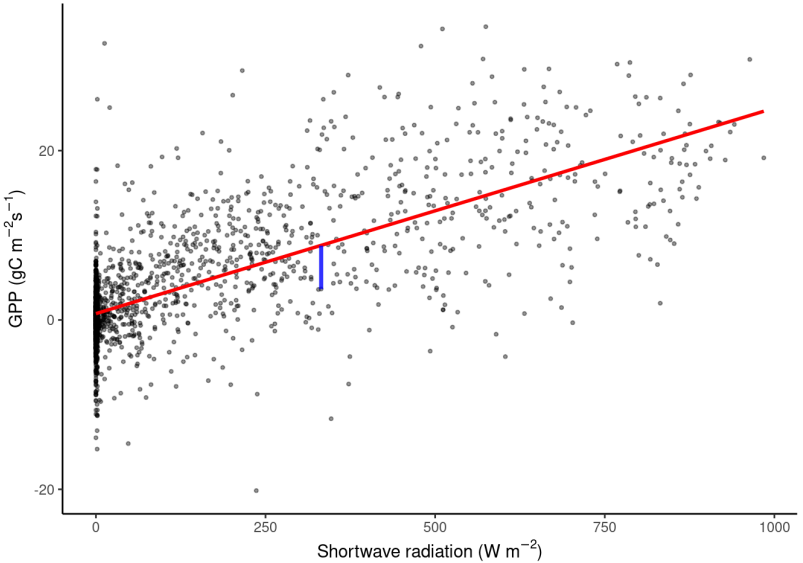
GECO

UNIVERSITÄT BERN

Recap: regression

- A regression models a **continuous** target variable as a function one or multiple continuous or categorical predictor variables.

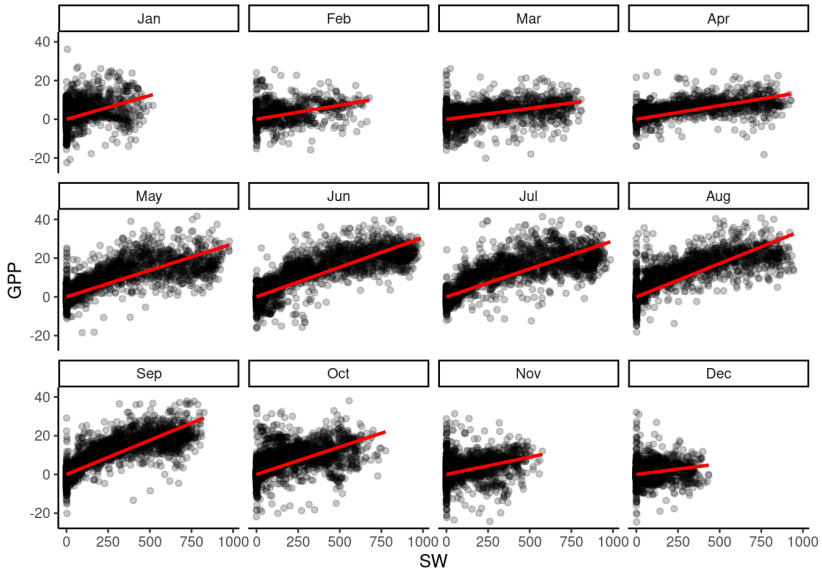
```
lm(formula = GPP_NT_VUT_REF ~ SW_IN_F, data = df)
```



Recap: regression

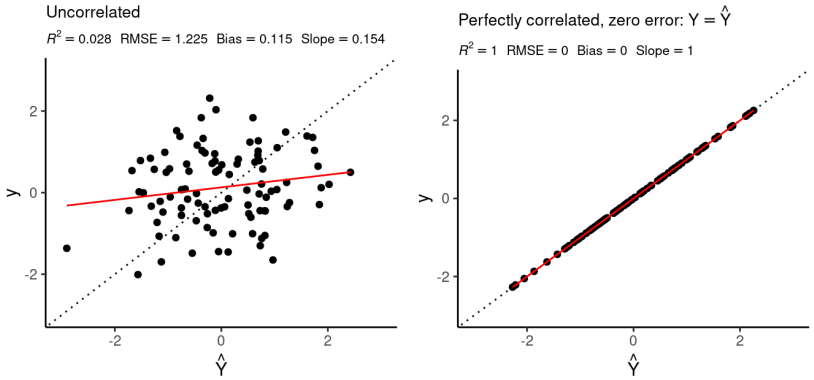
- A regression models a **continuous** target variable as a function one or multiple continuous or categorical predictor variables.

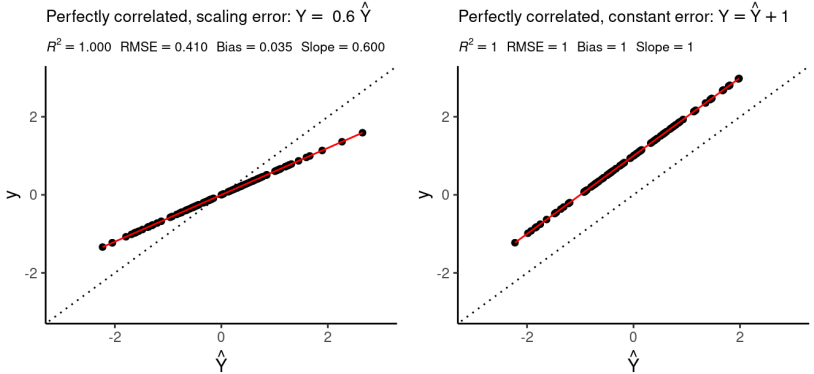
```
lm(formula = GPP_NT_VUT_REF ~ MONTH + SW_IN_F + MONTH:SW_IN_F,
    data = df_cat)
```

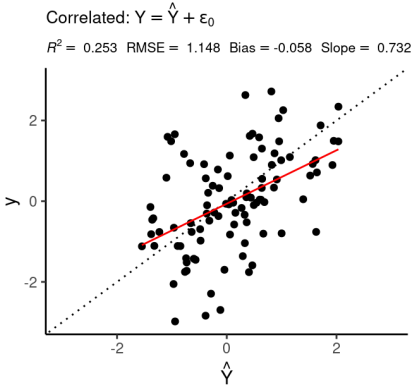


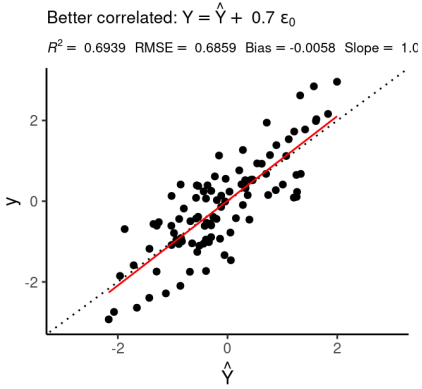
Recap: regression

- Multiple metrics measure different aspects of the model-data fit.









Recap: regression

- The R^2 always increases when increasing the number of predictors.
- However, this does not mean that the model performs better on new data (data not used for fitting the model).
- This indicates a trade-off between model complexity and generalisability.
- Metrics that penalise model complexity should be used for comparing models of different complexity (e.g., AIC).

Report Exercise: Stepwise forward regression

Classroom exercise

- Sketch pseudo-code for the stepwise forward regression
  - Define the loops.
  - At what position do you create the formula?
  - At what position do you select predictors?
  - How do you retain and complement the list of selected predictors?
  - How do you determine and update the list of remaining (“candidate”) predictors?