2 avoid common statistical problems

an explorative analysis of soil moisture data

2015-11-16 21:00

Introduction

Alain Zuur, Ieno Elena and Chris Elphick published **A protocol for data exploration to avoid comon statistical problems** in 2010 (Methods in Ecology and Evolution). They proposed 7 steps in order to prepare a data set towards the use of statistical methods in order to avoid type I or type II errors, thereby reducing the chance of making wrong ecological conclusions or recommendations.

The recomended protocol includes the following questions:

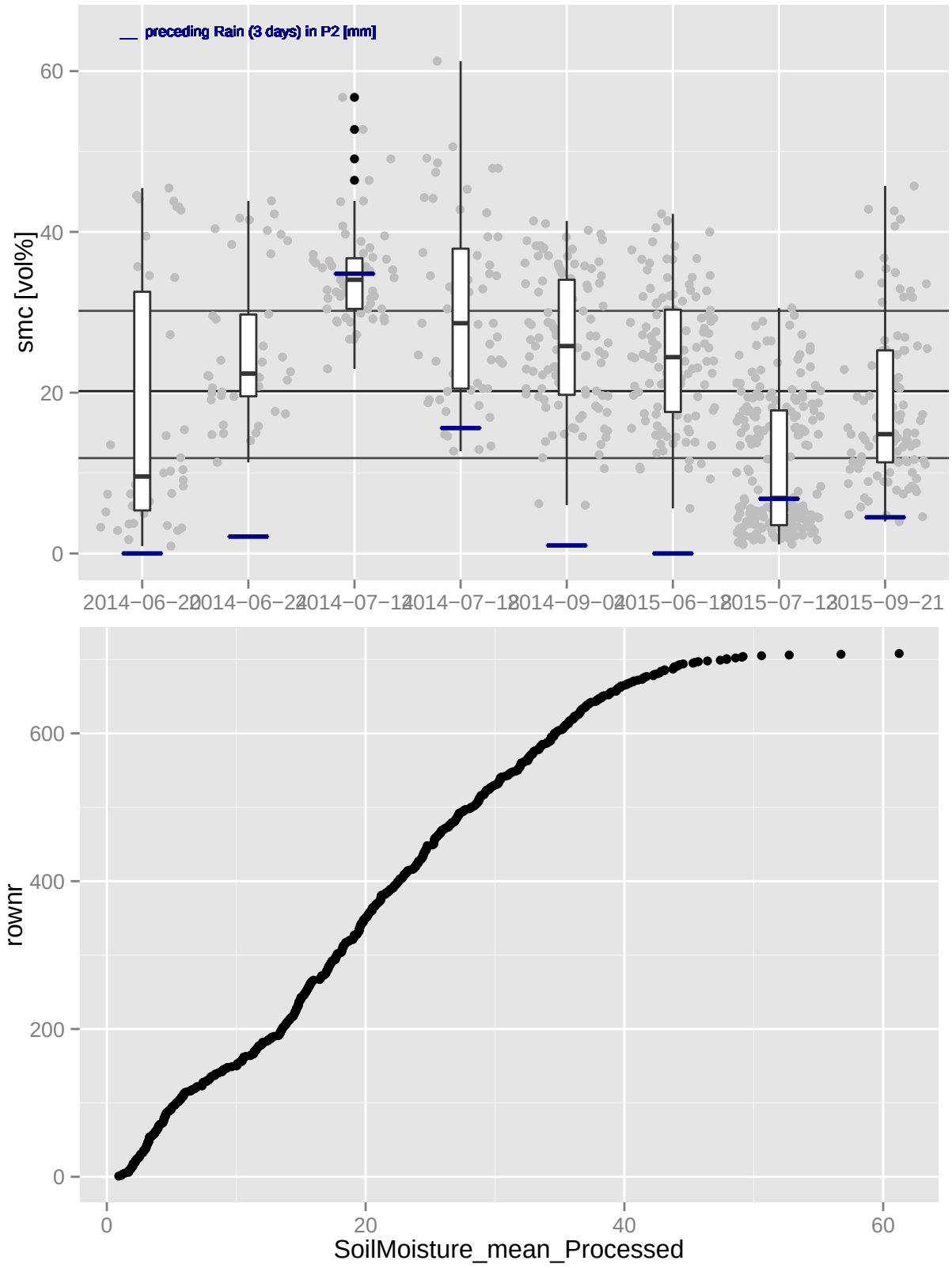
1. Are there outlier in x and y?
2. Do we have homogeneity of variance?
3. Are the data normally distributed?
4. Are there a lot of zeros in the data?
5. Is there collinarity among the covariates?
6. What are the relationships between y and x variables?
7. Should we consider interaction?
8. Are observations of the response variable independent?

In this document we will answer them by means of a visual data exploration.

The data set we are working on is explain in detail within the R-packages SoilMoisturePattern. For more information install the package and read the data set documentation:

ARE THERE OUTLIER IN X AND Y?

Outlier can be detected with boxplots or Cleveland dotplots (row number vs. observation).



Zuur Alain F., Ieno Elena N., Elphick Chris E. (2010): A protocol for data exploration to avoid common statical problems, Methods in Ecology and Evolution, 1, 3-14.