A note on presentation of environmental data

By Jaroslav Mohapl

Abstract

What to keep in mind when presenting environmental data.

Key Phrases Scientific data, environmental monitoring.
Key Words Data, science, environment, monitoring.

Introduction

Numerous manuscripts sent for a review to a scientific journal fail to provide vital information about the data serving as the empirical evidence supporting the conclusions of the study. The following recalls the elementary information authors want include should their work meet the basic standards.

Time Matters

Compared to data originating in designed experiments, environmental measurements, such as temperature, precipitation, wind speed, concentrations of particles etc. are observational and seasonal in nature. Hence, the time the observation period started and ended usually matters and should be quoted. A reasonably precise description of the sampling location(s), say using coordinates or a map, is also desirable. In long term monitoring, specimens are collected regularly on daily, weekly or monthly basis and the sampling frequency is certainly of interest to the reader. It is not unusual for air quality monitoring, due to resource limits, to sample irregularly, in which case, a more detailed description of the sampling scheme is highly appreciated. What authors provide very rarely is the motivation for the choice of the particular time frame and sampling locations. Of course, data are pressures valuables and citing them as collected under special circumstances

might make them useless in other studies. On the other hand it is the careful design of the sampling scheme that yields information of highest value. Good reasons for the choice of the sampling scheme are thus very powerful arguments for acceptance of the paper.

Provide Sample Size

The most common ways of presenting data are summary statistics and time series plots. The first essential statistics the reader wants to know is the sample size. Sometimes, data are missing, because everything in this world malfunctions occasionally. The number and percentage of missing observations have to be available, in particular if the author intends to use the data as regular time series. The impact of missing observations under such treatment should also be discussed. The basic descriptive statistics are average, standard deviation, sample minimum and maximum. While computation of the average is straight forward, computation of the deviation always deserves an explanatory note. It is usually the square root of the sum of squares divided sometimes by the sample size N, sometimes by N-1. There is no reason to prefer one over the other without an underlying model for the observations, but the information is always good to know, in particular for small sample sizes. Many authors describe variability of the data using the average of absolute deviations from the average. This is not wrong, but with the ordinary deviation, using sum of squares, we are always ready to calculate a t-test, which may come handy. There is a long list of other statistics, such as median, quartile, geometric mean and geometric standard deviation authors can include. As to time series plots, do not forget to tell the reader what is in the plot, particularly if you display other than raw data.

State Precision

One detail authors always neglect mentioning is the precision of the observations. Precision of the data determines precision of the summary statistics.

2

Collecting integer data and presenting their average with six decimals makes no sense. Precision comes to mind particularly in long term studies. For example, accuracy of data collected by a modern digital thermometer capable of depositing analog temperatures directly onto a hard drive is higher than that of measurements obtained 150 years ago by a mercury tube. The first device functions with precision 0.01℃ in the range from -50℃ to 50℃, the second goes hardly beyond plus-minus 1℃ or 2℃ and its performance in extreme temperatures is questionable.

Summary

The minimum information one should present about the data is: when, how long and how often did we collect, how many data do we have and how many of them are useless. Then add precision of the measurements and the basic sample characteristics, such as the sample mean and sample deviation. Add a nice time series plot if appropriate and you are all set to discuss your project!