Atmos. Chem. Phys. Discuss., 14, C3934–C3936, 2014
www.atmos-chem-phys-discuss.net/14/C3934/2014/
© Author(s) 2014. This work is distributed under
the Creative Commons Attribute 3.0 License.

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Spatial variations and development of land use regression models of levoglucosan in four European study areas" *by* A. Jedynska et al.

Anonymous Referee #2

Received and published: 20 June 2014

The paper describes the measurements of Levoglucosan in four European regions at different seasons (summer, winter and an intermediate season, either spring or autumn). Based on the measurements, land use regression (LUR) models are developed for an assessment of the relevance of Levoglucosan as a marker for wood burning. The relevance is quantified by the correlation of Levoglucosan with other air pollution indicators and its variance "explained" by the developed LUR. While the addressed questions might be of scientific relevance the paper suffers from basic technical failures and vague descriptions of the applied statistical methods and developed LUR models.

According to Stephen F. Gull (Cambridge, 1994) data analysis as simply being 'a dialogue with the data'. But this statement does not mean that data analysis is simply applying test statistics to the data and interpreting the computed numbers. Each statistical method is based on fundamental assumptions on the data. If these assumptions are violated, any conclusion drawn from this statistics might be correct or wrong.

The authors applied students t-tests " to calculate the difference (and significance) between site types and between seasons. " This sentence is a good representative example of the weakness of the paper. 1) The sentence is imprecise as the authors do not compute difference between site types and between seasons. I guess the authors mean that they compute the difference of Levoglucosan measured at different site types and at different times. 2) Statistical tests do not allow to compute anything but are designed for falsification of hypothesis. 3) Student t-tests is the term for a bulk of statistical tests based on the corresponding probability distribution. A detailed description of the applied statistical tests would it allow for other researcher reproducing the results in this paper. 4) Student t-tests require the underlying distributions to have the same variance. A glimpse at the box-plot in Figure 2 would have revealed that this requirement is obviously violated. The Welch's t test allows the distributions to differ in their variances. 5) Student t-tests require the underlying distributions to be Gaussian. The measured Levoglucosan concentrations are obviously not Gaussian distributed. Thanks to the central limit theorem non Gaussian distributed data could also be treated, however require a sufficient number of samples. But as the rule of thumb of N=30 samples is violated, the significance of the applied student t-tests are questionable.

One of the main approaches, the development of LUR models is described rather short making it impossible reproducing the approach. Again a crucial assumption of the involved statistical tool has not been examined or discussed. The coefficient of determination Rˆ2 as a measure of how much of the variance of the data is explained by the model requires a linear relationship between dependent and independent variables.

The authors did not describe any diagnostic tools investigating the linear assumption making any conclusion valueless.

Linear regression models imply real valued dependent variables. Levoglucosan concentrations obviously cannot be modeled by real numbers but they are tied to be non-negative. The authors should discuss this misfit, e.g. how they prevent the LUR models from predicting 'negative' concentrations.

The authors introduce several threshold values without any explanation, e.g. different p-values for different statistical tests. In Section 2.5 two different threshold values for outlier detection have been defined. What is the reason for these obviously arbitrary chosen values? How sensitive is the scientific conclusions of the paper to these thresholds? High sensitiveness of the scientific conclusions on these threshold values renders any conclusions valueless as there is no reason not to change the threshold .

All in all, the paper lacks in precision of the method descriptions, in the examination of the requirements of used statistical methods and in the application of a non suitable Student t-test.

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 13491, 2014.