Geophysical Research Abstracts
Vol. 18, EGU2016-15689, 2016
EGU General Assembly 2016

Artificial bias typically neglected in comparisons of uncertain atmospheric data

Mikko R.A. Pitkänen (1,2), Santtu Mikkonen (2), Kari E.J. Lehtinen (2,1), and Antti Arola (1)

(1) Finnish Meteorological Institute, Atmospheric Research Centre of Eastern Finland, Kuopio, Finland (mikko.pitkanen@fmi.fi), (2) University of Eastern Finland, Department of Applied Physics, Kuopio, Finland

Researchers in atmospheric sciences frequently disregard data uncertainty in their choice of methods for data analysis and visualisation. Such methods include the widely used standard least squares line fitting in combination with some variations of scatter plots when comparing two different data sets of the same physical quantity. When using these methods, random data uncertainty (eg. measurement uncertainty) causes artificial systematic bias in the comparison between the extreme values of the data sets, which is then often interpreted falsely as a consequence of some true physical phenomenon or instrument misbehavior. This artificial bias is recognized as regression to the mean (RTM), that is a known effect in the field of statistics, but mostly disregarded in atmospheric sciences and not acknowledged at all in the vast majority of publications in our field. All kinds of data comparisons are subject to the bias, as long as uncertainty is present in the data.

This work introduces the concept of RTM bias and demonstrates the necessity of considering the RTM effect in comparisons of data with uncertainties. We not only visualize the RTM effect with synthetic data but also use simulations based on real atmospheric data to estimate the magnitude of RTM bias in data comparisons common in our field. Typically, RTM bias is greater when the reference data (often on the x-axis) has greater uncertainty. For example, mid-visible aerosol optical thickness determined using a sun photometer may have a fairly low uncertainty of +-0.01 and, thus the RTM effect is small when using it as reference data. On the other hand UV index measurements with a broadband instrument may have an uncertainty of 10 % and higher, and the bias caused by RTM becomes larger. The bias caused by RTM is typically greatest for the extreme values of the data sets, emphasizing the need to account for RTM bias when comparing and interpreting these cases.