



Once more regression to the mean

¹Junghans, M., ¹Leich, A., ¹Nippold, R., ^{1,2}*Wagner, P.

*lead presenter

¹peter.wagner@dlr.de, DLR Institute for Transport Systems, Germany

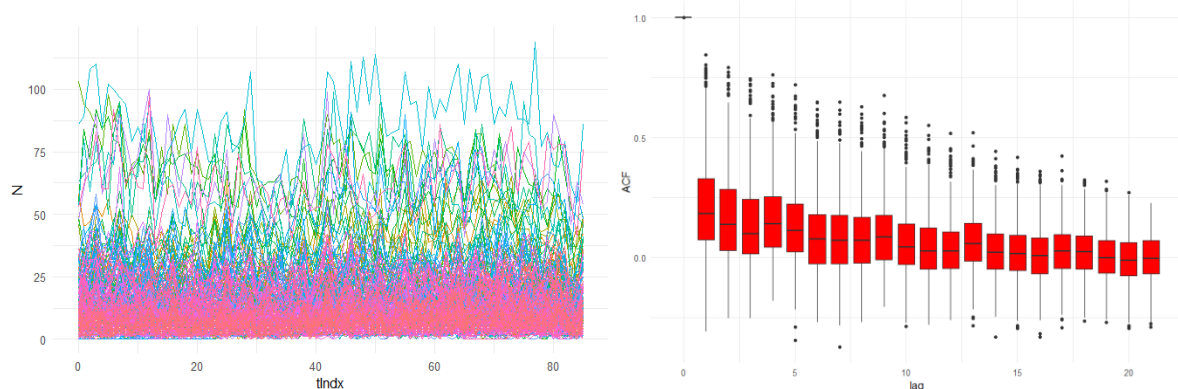
² Institute of Land and Sea Transportation, TU Berlin, Germany

Keywords: regression to the mean, auto-correlation, traffic safety

There is a beautiful paper by Ezra Hauer [1] demonstrating the regression to the mean (R2M) effect in traffic safety. In addition, there are a lot of papers detailing this more, but we feel that there is a lack in a clear assessment of this effect. Most papers even do not acknowledge it, as is true for practioneers as well – at least here in Germany, where dark spots are recognized and subsequently improved upon based on surprisingly small crash numbers. Of course, the same happens (if at all, but this is yet another story) with the observed improvements in traffic safety after certain measures have been introduced: they seem to be often exaggerated and claimed, without considering R2M at all.

By analyzing a large crash data-base, we can characterize R2M with a lot of time-series and see how strong this effect is in reality. Furthermore, for such a data-set one can watch for real changes in the sense that a real jump is visible in the data (the left part of the Figure below shows the time-series of the Top 500 intersections with a time aggregation of 12 weeks). It turns out, that these time-series display an astonishing amount of auto-correlation (right part of the Figure, which is an aggregated auto-correlation function), which in a certain sense counteracts R2M. The presence of auto-correlations is already known, but it seems that this effect is not well quantified so far.

The most important conclusion from this research is that the R2M effect is weaker than what is to be expected from a simple draw of independent crash-numbers, such as it has been done in [1]. Clearly, this is due to the auto-correlation in the data from a single site, and the auto-correlation itself is most likely from the fact that the traffic flows are very reproducible, and are one of the main determinants of crashes: as can be seen from count data, the daily and weekly traffic pattern of demand do reproduce very well, therefore introducing a strong auto-correlation in the counts, which are transferred to the crash-numbers.





International Co-operation on Theories and
Concepts in Traffic Safety

In addition to this, these time-series can be characterized further by the relation between their variance v and their mean μ : some of them displaying a Poissonian behavior where $v = \mu$, and some of them displaying over-dispersion $v > \mu$, which is a strong sign of the negative binomial distribution. The final version of the paper will try to quantify the effect of the auto-correlation on the size of the R2M, thereby giving some credit to the practioneers.

[1] Ezra Hauer (1986), *On the estimation of the expected number of accidents*, Accident Analysis & Prevention, **18**(1), 1-12, [https://doi.org/10.1016/0001-4575\(86\)90031-X](https://doi.org/10.1016/0001-4575(86)90031-X).