Review – Yin et al, HESS-2018-601

The authors of this paper seek a method that decomposes a time series of monthly means into three components, one of which might contain a trend component, the second representative of the annual cycle, and the third representing the remaining stochastic month-to-month variability, with the decomposition being performed in such a way that the pair-wise covariance between the three components is zero. They propose an analysis of variance that achieves this objective. Competing decomposition methods are briefly assessed and found not to produce components with zero pairwise covariance (i.e., are uncorrelated), although the author's calculations suggest that the covariances obtained using these methods are small in most cases in the examples considered.

First, it is not obvious that the three methods considered correspond to three identical models for the month-to-month variation in precipitation – and indeed, it seems that the models cannot be the same. A challenge from the outset is that it is not clear what parts of model (1) are stochastic, and which are fixed, although for all three it seems that the component $P_m(t)$ represents the annual cycle, and thus has the property $P_m(t)= P_m(t+12)$. I would conclude then that $P_m(t)$ is deterministic rather than stochastic, so the notion of variance and covariance (between random variables) doesn't quite seem to fit the bill.

For the first variant of the model, discussed in 4.1, the component $P_a(t)$ is take to be a linear trend. Since the entire observational record is considered, it seems that for this method at least, $P_a(t)$ is deterministic, and thus again, the notion of variance and covariance between random variables does not apply. In this case, what is being fitted is a variant of model (1) with a deterministic linear trend, a fixed annual cycle, and residual stochastic variability. There are probably a number of ways this could be fitted other than by estimating components sequentially – and perhaps an ANOVA formulation is one of those better ways, but this should be judged in terms of the relative efficiency of parameter estimates obtained via different methods as opposed to whether the apparent covariance between components is zero.

The second variant of the model, discussed in 4.2, uses a crude low pass filter to obtain $P_a(t)$. This is rather different from a trend formulation, because the filter will pass both any deterministic change in level over time (assuming that such changes only occur on long time scales) and stochastic variability at time scales that the filter allows to pass. Note that the moving average filter does not cut off smoothly with frequency (it has messy "side lobes" that leak high frequency variance), resulting in contamination of the low frequency component by higher frequency "noise". Thus, this variant of model (1) has in mind trend plus stochastic low-frequency variability as one component, the annual cycle as a second component, and stochastic high-frequency variability as a third component. Variability and covariance of the first and third terms make sense, in so much as deterministic trend is not present in the first term. Clearly this is a different animal from that considered in the first variant of the model. Both the first term is different, and the nature of the variability that is retained in the third term is different.

The third variant of the model, discussed in Section 5, apparently uses a 2-factor ANOVA model to decompose a timeseries of monthly means into an annual effect (with a different level for each year), a month effect (with a different level for each of the 12 months of the year, thus representing the annual cycle), and a residual component. The interpretation of this type of model requires consideration of whether year and month effects are fixed or random. In this case, I would assume that year effects are random, and month effects are fixed, since they are common to all years. The partitioning of variability in an ANOVA analysis is done in such a way that, under the assumption of the Gaussian distribution and iid residual variability, the three variance components that result are statistically independent. This is all standard stuff, and I'm not sure that the long appendix is required to make essentially this point. From a climatological perspective, the interpretation of this variant of model (1) is not very different from that of the second variant of the model considered in section 4.2. The annual component presumably has deterministic trend and low frequency stochastic components, the annual cycle is deterministic, and the residual has higher frequency stochastic variability.

Again, differences between these methods, and the underlying variants of model (1) that are implicit in these methods, should be considered in terms of the different objectives of the methods (a different variant of model (1) implies a somewhat different objective), and whether one method produces better estimates (from a statistical perspective) of model parameters and properties rather than rather arbitrarily focusing on a single aspect, the covariance of component estimates.

One final note concerning a statement that appears on line 90 of the manuscript – zero covariance is synonymous with independence ONLY if the monthly time series values are Gaussian (i.e., normally) distributed. It seems very evident from the figures in the supplement that this is certainly not the case at Cobar (Lerida), e.g., see Figs S3d and S5d. While less pronounced, the similar figures for the other two locations also show some evidence of skewed residuals, and hence a departure from the Gaussian assumption.