

Interactive comment on “Chemical composition of rainwater at Maldives Climate Observatory at Hanimaadhoo (MCOH)” by R. Das et al.

Anonymous Referee #2

Received and published: 6 October 2010

My first concerns about the paper are over data quality, and until these issues—and the ones raised below—are dealt with satisfactorily, I'd argue that there is really no point in commenting on the later sections of the paper.

First, on P. 17574 the authors write: “At least once a week and when found visibly contaminated the wet-only collector was cleaned as part of quality control scheme using a plastic brush and de-ionized water (18MΩcm resistivity) from a spray bottle in different combinations.”

What sorts of contaminants were seen on the funnels? How often were the samplers found to be contaminated? Was one of the samplers contaminated more often than the other? This is a serious concern because of the very high concentrations of nssCa⁺⁺ found in some samples. Unless contamination can be discounted as a source for

C8323

nssCa⁺⁺, I'd argue that any detailed discussion of its sources would be overly speculative, and the data for the other ions may be compromised as well. {As an aside, I would be interested in seeing the high nssCa⁺⁺ samples plotted with a different symbol in Fig. 2—how were the ion balances for those samples? Also, it would be interesting to see the samples from the different trajectory groups plotted with different symbols in Fig. 2 and Fig. 3}

Beyond the visible contamination, I'm concerned about contamination that wasn't visible, and along these lines, I think some additional information on the sampling protocol needs to be included. What concerns me here is that there is no indication that the collector was cleaned after each sample was collected. What I'm wondering is whether all of the rain that falls on the funnel actually ends up in the collection bottle. If not, I'd be concerned that any droplets that remained on the funnel would evaporate and the ions that were in solution would carry over and be collected in the subsequent sample. And of course, any droplets left on the funnel also would be prime sites for reactions with gaseous substances. So if the samplers were not cleaned between samples, I'd be concerned about sampling artifacts and their impact on data quality.

Along these lines, I didn't see any mention of blanks. These should be routinely collected for QA/QC and if any were collected, the data for them should be summarized, at least briefly—as supplemental material if nothing else. If blanks were not collected, I would consider this a serious flaw in the experimental design, possibly a fatal flaw, depending upon the authors' response to the possible lack of cleaning issue raised above.

I may have missed it, but I didn't see some information that I think should be included in the paper. First are at least summary statistics for the numbers of samples and the volumes of the samples. This information should be included in Table 1 and it should be given for each trajectory group.

I'm concerned about the volumes of the samples because of possible volume effects

C8324

on concentrations—it appears (Fig. 5) that there are higher concentrations of some the analytes, such as NO₃⁻, in the lower volume (more concentrated?) samples. Are the volumes of the samples comparable for all of the trajectory stratified groups?

[Perhaps a bit tangential, but I would also be interested in knowing what time of day the rain events most often occurred for the various groups, i.e., are the “photochemical environments” similar among groups?]

On P. 17576, the quality of the NH₄⁺ and nssK⁺ data is discussed, and it raises some flags about the validity of the results. I am particularly concerned about the nssK⁺ data because in two of the trajectory-defined groups (marine and mixed), nssK⁺ is a very small percentage of the total K⁺. More to the point, the concern is that for those groups nssK⁺ is a small difference between two large numbers, and so very noisy and likely with little meaning. I’m particularly concerned about the inclusion of this variable in the PCA because of the large uncertainties in its concentrations (I have somewhat lesser concerns over using the NH₄⁺ data in the PCA for the same reason). Although I admittedly have limited experience with the technique, I have found that the inclusion or exclusion of a single variable can greatly influence the outcome of PCA. Even if not the results are not presented in the final version, I’d suggest that authors try several runs of the PCA with and without nss Ca⁺⁺, nssK⁺ and NH₄⁺ to see how robust the results are.

As noted in the review RC C7496 already posted, PCA requires a large number of samples, and this is another reason I’m concerned about the number of samples. The following quote is from SAS support:

“Principal component analysis is a large-sample procedure. To obtain reliable results, the minimal number of subjects providing usable data for the analysis should be the larger of 100 subjects or five times the number of variables being analyzed.” (<http://support.sas.com/publishing/pubcat/chaps/55129.pdf>)

Further, I am curious about the way in which the relationships among the ions were

C8325

evaluated. The relationships are presented in several tables of correlations, but there is no evidence that the variables are in fact linearly related. It’s generally known that a relatively small number of extreme points can exert a disproportionate influence on correlations, i.e., leading to “significant” correlations when the data are not really linearly related. Some x-y plots might be added, even if as supplemental material.

[I don’t actually know how the data should be handled for these analyses, but are the correlation calculations weighted by volume?]

Back to PCA. One of the assumptions of PCA is this:

Linearity. The relationship between all observed variables should be linear. (same SAS reference as above)

Thus I think it would be important to address the linearity issue in some detail.

My next concern is over the use of trajectories in interpreting the precipitation data. First, the trajectories are calculated for one point in time (5 pm) each day. I’m wondering how representative these will be relative to the actual time the rain event occurred. For example, the 5 pm trajectory could be more than 12 hrs after the event occurred if the rain fell shortly after midnight. Second, and more problematic in my mind: how are the heights of the trajectories matched to the rain event? My understanding is that in-cloud processes are much more important determinants of what gets into the rain than below-cloud scavenging. So, are the trajectories matched to the altitude where the precipitation formed, and if so how? In the same vein, how do the trajectories deal with convective precipitation? Wouldn’t the strong vertical motions in convective storms make the trajectories almost meaningless? Wouldn’t convective events upwind of the sampling site similarly limit the usefulness of trajectories? Even for stratiform precipitation, I’d think the trajectories would need to be carefully matched to the heights of the air masses where the droplets formed—as the term “stratiform” would imply.

C8326

I think the authors also need to provide some more detailed information on how the trajectory groups were defined. Was a clustering algorithm used to objectively establish groups, or was this done by eye? I suspect the latter, and if so, was information on the vertical dimension of the trajectories taken into account? Frankly, to my eye, at least some of the mixed trajectories look quite similar to those in the marine group. At the very least, Fig. 4 should include plots showing the vertical characteristics of the trajectory groups, and this dimension should be given considerably more attention. In particular, the heights of the trajectories as they passed over potential source regions could well determine whether pollutants and other continental substances were entrained into the air that eventually was sampled at MCOH.

As the stratification of the data by trajectories is the foundation for much of the interpretation of the results, I think the authors need to establish that their approach is robust.

Again, until these questions about data quality and the approach used to interpret the data are addressed, I think further comments on the results would be unwarranted.

Interactive comment on Atmos. Chem. Phys. Discuss., 10, 17569, 2010.

C8327