AQMEII4 Activity 2 Call March 5, 2024

Participants: Christian Hogrefe, Jesse Bash, Stefano Galmarini, Colin Lee, Olivia Clifton, Annika Vogel, Johannes Flemming, Anam Khan

At the beginning of the call, Olivia reported that she met with Anam a few weeks ago to discuss Anam's manuscript and that she expects Anam to send her a draft version of the manuscript in the near future. Olivia also shared that Anam's PhD defense has been scheduled for May.

Colin reported that he was busy with other projects and therefore has not been able to work on his AQMEII4 machine learning research lately. He hopes to be able to resume his AQMEII4 work soon and will provide an update in future meetings.

Annika summarized the exploratory "part 1" analyses she had presented at the last meeting

- Data completeness (i.e. the number of time records when observations and all models have valid data) is limited by observations and to a lesser extent CMAQ-STAGE
- There are huge differences (4-5 orders of magnitude) between state variances. The datasets can be grouped in a "high-variance" cluster (GEM-MACH*, IFS*, WR1Fchem, and MLCchem) and a "low variance cluster" for the remaining models that also includes the observations
- Some state cross-variances exceed state variances
- State correlations are typically lower for model-vs-observation pairs than model-vs-model pairs
- The pairs with the highest correlations are fairly similar across stations
- Ramat Hanadiv shows a higher range of state cross-variances and increased state correlations compared to other sites, possibly due to low data availability

During Annika's recap of her initial analyses, Johannes iterated a point from the meeting that it would be useful to stratify the analysis to separate seasonal, diurnal, and possibly synoptic scale variability.

Annika also showed a new slide depicting the monthly and hourly variations of data completeness (i.e. the number of time records for which observations and all model values are available) at each site. Results showed strong differences between sites in terms of monthly data availability fluctuations. For hourly data availability fluctuations, most sites showed higher completeness during daytime hours compared to nighttime hours.

The group discussed how these exploratory "part 1" analyses can support the formulation of concrete research questions from a deposition modeling point of view. As part of the discussion, Johannes asked whether this analysis could potentially be used to answer how models should be improved, maybe focusing on the models participating in both activity 2 and activity 1 so that deposition code improvements identified through this work could also improve grid models.

After these discussions, the group's preliminary suggestion was to formulate a research question focused on synoptic-scale variability. From a data processing point of view, this would involve deseasonalizing the data and taking out the diel cycle. Annika will initiate follow-up email discussions.

Next call: Tuesday April 2, 10:00 EDST / 14:00 GMT / 16:00 CEST