EMS Annual Meeting Abstracts
Vol. 12, EMS2015-573-3, 2015
15th EMS / 12th ECAM

When can we classify a model as good or bad? - Quality assurance of mesocale predictions in high complex terrain

Ronny Petrik (1), K. Heinke Schlünzen (1), and Dietmar Öttl (2)

(1) University of Hamburg, Meteorological Institute, Germany (ronny.petrik@zmaw.de), (2) Government of Styria, Section Air Quality Control, Austria

Simulations of mesoscale models are used to be evaluated in order to ensure that the forecasts are physically consistent and reproducing the measurements and the underlying atmospheric processes. For instance, operationally used NWP models are repeatedly quality checked applying specific evaluation methods and statistics. Having determined the evaluation measures like mean errors, skill scores or hit rates, two main problems arise with the interpretation. First, the meaning of the absolute error is difficult to assess because nobody really knows, which portion of the error can be attributed to problems inherent in the 'model vs. observation' comparison (consider the measurement accuracy and the spatial representativeness) and which portion of the error can be really attributed to shortcomings in the model dynamics or physics. Second, it is hard to conclude just from the error, whether a model can be classified as good or not.
Indeed there is the possibility to compare the evaluation results of the own model with other models. Nevertheless, as soon as we are considering those model applications used to prepare certificates / assessments for future planned industrial sites or factories, a final decision has to be taken with respect to the suitability of a specific model. In particular in regions with complex terrain high quality and high-resolution simulations have to be performed to guarantee a realistic dispersion of industrial pollutants. Therefore, we aim to examine and derive evaluation criteria for mesoscale atmospheric models. Our motivation also arises from the fact that in Mid Europe there are no well-defined quality criteria for consultant-applied models - thus, we are supported by the German Federal Environment Agency and by the German Association of Engineers.
We will simply start from the basic assumption that our derived evaluation criteria shall separate physically consistent and proper results of prognostic mesoscale models from the results of simplier diagnostic and other 'low-order' models. To find some candidates suitable for a full-scale realistic evaluation experiment, much time has been spent on the selection of reference (observational) data sets and on the validation of individual measurement sites. Three different data sets remained after a critical assessment: a postfrontal flow around an isolated hill in stable atmospheric conditions (Sophienhöhe), the establishment of a counter-flow in the Graz basin in winter during inversion weather conditions, and observed drainage flows in the Stuttgart basin in spring time. In our presentation we propose a general method to derive evaluation critera and thresholds. A kind of boot-strap analysis is addressed, which allow for a seperation beween 'good' models and 'bad' models, i. e. the discrimination of 'low-order' models as well as inproper mesoscale simulations.