



## **Model selection and verification for ensemble based probabilistic forecasting of air pollution in Oslo, Norway**

Sam-Erik Walker\*

Dept. of Mathematics, Univ. of Oslo, Oslo, Norway – [samerikw@math.uio.no](mailto:samerikw@math.uio.no)

Gudmund Horn Hermansen

Dept. of Mathematics, Univ. of Oslo, Oslo, Norway – [gudmunhh@math.uio.no](mailto:gudmunhh@math.uio.no)

Nils Lid Hjort

Dept. of Mathematics, Univ. of Oslo, Oslo, Norway – [nils@math.uio.no](mailto:nils@math.uio.no)

In this paper, we discuss building time series models for forecasting of air pollution during wintertime conditions in Oslo, Norway, using ensembles of air pollution model data. Since such ensembles becomes increasingly available as part of regular air quality forecast modelling, it is important to build properly calibrated statistical models utilising such data. In particular, we focus on model selection using the Akaike and Bayesian information criteria, and verification of the forecasts using Probability Integral Transform (PIT) histograms and Brier scores. Three time series models are considered, using ensemble mean values as a primary covariate in a linear regression setting explaining observations, and modelling the residual errors as an autoregressive process, using either a constant variance; a time-varying (heteroscedastic) variance only depending on the ensemble variances; or as a combination of both. We show that for the limited, although representative, data analysed, the model incorporating both terms, seems to have an edge according to the model selection criteria and forecast verification tools used. Finally, we briefly discuss the possibility of introducing more focused model selection criteria for these types of models and data.

**Keywords:** Time series; Information Criteria; Brier score; Air quality.