Mixture model-based atmospheric air mass classification: a probabilistic view of thermodynamic profiles

Jérôme Pernin (1), Mathieu Vrac (2), Cyril Crevoisier (1), and Alain Chédin (1)

(1) LMD, IPSL, CNRS, Ecole polytechnique, Université Paris-Saclay, 91128 Palaiseau, France
(jerome.pernin@lmd.polytechnique.fr), (2) LSCE, IPSL, CEA, CNRS, UVSQ, Centre d'Études de Saclay, Orme des
Merisiers, Bât. 701, 91191 Gif-sur-Yvette, France

Air mass classification has become an important area in synoptic climatology, simplifying the complexity of the atmosphere by dividing the atmosphere into discrete similar thermodynamic patterns. However, the constant growth of atmospheric databases in both size and complexity implies the need to develop new adaptive classifications. Here, we propose a robust unsupervised and supervised classification methodology of a large thermodynamic dataset, on a global scale and over several years, into discrete air mass groups homogeneous in both temperature and humidity that also provides underlying probability laws. Temperature and humidity at different pressure levels are aggregated into a set of cumulative distribution function (CDF) values instead of classical ones. The method is based on a Gaussian mixture model and uses the expectation–maximization (EM) algorithm to estimate the parameters of the mixture. Spatially gridded thermodynamic profiles come from ECMWF reanalyses spanning the period 2000–2009. Different aspects are investigated, such as the sensitivity of the classification process to both temporal and spatial samplings of the training dataset. Comparisons of the classifications made either by the EM algorithm or by the widely used k-means algorithm show that the former can be viewed as a generalization of the latter. Moreover, the EM algorithm delivers, for each observation, the probabilities of belonging to each class, as well as the associated uncertainty. Finally, a decision tree is proposed as a tool for interpreting the different classes, highlighting the relative importance of temperature and humidity in the classification process.