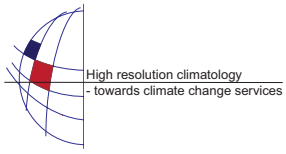
EMS Annual Meeting Abstracts
Vol. 7, EMS2010-517, 2010
10th EMS / 8th ECAC



Gaussian Mixture Models for forecasting and filling of climatological time series

D. Calandria-Hernández, J.M. Hidalgo-Muñoz, D. Argüeso, S.R. Gámiz-Fortis, M.J. Esteban-Parra, and Y. Castro-Díez

Facultad de Ciencias. Universidad de Granada, Física Aplicada, Granada, Spain (ycastro@ugr.es, +34 958243214)

In Statistics, a Gaussian Mixture Model (GMM) is a probabilistic density estimation method that consists of a linear combination of normal distributions. Model parameters, i.e. means and variances of each normal distribution, as well as the linear combination coefficients, can be estimated easily using the Expectation Maximization (EM) algorithm. Since the model expression is non-linear in its parameters (particularly regarding to the means and variances), GMMs are considered non-linear.
In this work, a methodology to forecast time series based on a GMM is presented. Such model provides several advantages over other classical forecasting models, i.e. Autoregressive with exogenous variables (ARX) model. To forecast a multimodal variable (for example, temperature), GMM offers a more intuitive representation because each normal distribution can be adjusted over each mode. Besides, GMMs can be viewed as an ensemble of ARX models. Because of each ARX model can have different variance, the ensemble provides a similar behavior to an Autoregressive Conditional Heteroskedasticity (ARCH) model, allowing a better forecasting of the error band.
The model has been applied for forecasting monthly time series of river flow in the Iberian Peninsula. The streamflow from a particular station is forecasted using the station itself (autoregression) and also the information provided by those stations with better correlation coefficients (exogenous variables). Data used in the model embraced the period between 1960-2005, using the interval 1960-1980 for calibration and 1961-2005 for validation. Results obtained highlight the usefulness of this methodology regarding to other classical forecasting models. This technique results particularly successful for filling missing data in the Iberian streamflow series.

Acknowledgements: The Spanish Ministry of Science and Innovation, with additional support from the European Community Funds (FEDER), project CGL2007-61151/CLI, and the Regional Government of Andalusia project P06-RNM-01622, have financed this study.