Geophysical Research Abstracts
Vol. 18, EGU2016-17561, 2016
EGU General Assembly 2016

Learning Bayesian networks from big meteorological spatial datasets. An alternative to complex network analysis

Jose Manuel Gutiérrez (1), Daniel San Martín (2), Sixto Herrera (3), and Antonio Santiago Cofiño (3)

(1) Meteorology Group. Instituto de Física de Cantabria (IFCA), CSIC - Univ. of Cantabria. Santander, Spain, (2) PREDICTIA Intelligent Data Solutions. Santander, Spain, (3) Meteorology Group. Dept. of Applied Mathematics and Computer Science. Univ. of Cantabria. Santander, Spain.

The growing availability of spatial datasets (observations, reanalysis, and regional and global climate models) demands efficient multivariate spatial modeling techniques for many problems of interest (e.g. teleconnection analysis, multi-site downscaling, etc.). Complex networks have been recently applied in this context using graphs built from pairwise correlations between the different stations (or grid boxes) forming the dataset. However, this analysis does not take into account the full dependence structure underlying the data, gien by all possible marginal and conditional dependencies among the stations, and does not allow a probabilistic analysis of the dataset. In this talk we introduce Bayesian networks as an alternative multivariate analysis and modeling data-driven technique which allows building a joint probability distribution of the stations including all relevant dependencies in the dataset. Bayesian networks is a sound machine learning technique using a graph to 1) encode the main dependencies among the variables and 2) to obtain a factorization of the joint probability distribution of the stations given by a reduced number of parameters. For a particular problem, the resulting graph provides a qualitative analysis of the spatial relationships in the dataset (alternative to complex network analysis), and the resulting model allows for a probabilistic analysis of the dataset.

Bayesian networks have been widely applied in many fields, but their use in climate problems is hampered by the large number of variables (stations) involved in this field, since the complexity of the existing algorithms to learn from data the graphical structure grows nonlinearly with the number of variables. In this contribution we present a modified local learning algorithm for Bayesian networks adapted to this problem, which allows inferring the graphical structure for thousands of stations (from observations) and/or gridboxes (from model simulations) thus providing new possibilities for the multivariate analysis of these spatial datasets. An example for multi-site downscaling relating both model simulations and observations is used to illustrate the potential applications of Bayesian networks in climate applications.