Mathematical Statistics
Stockholm University
Research Report **2017:14**,
http://www.math.su.se

Towards a flexible statistical modelling by latent factors for evaluation of simulated responses to climate forcings: Part III

Ekaterina Fetisova* Anders Moberg† Gudrun Brattström‡

October 2017

Abstract

Evaluation of climate model simulations is a crucial task in climate research. In a work consisting of three parts, we propose a new statistical framework for evaluation of simulated responses to climate forcings, based on the concept of latent (unobservable) variables. In Part I, several latent factor models were suggested for evaluation of temperature data from climate model simulations, forced by a varying number of forcings, against climate proxy data from the last millennium. In Part II, focusing on climatological characteristics of forcings, we deepen the discussion by suggesting two alternative latent variable models that can be used for evaluation of temperature simulations forced by *five specific* forcings of natural and anthropogenic origin. The first statistical model is formulated in line with *confirmatory factor analysis* (CFA), accompanied by a more detailed discussion about the interpretation of latent temperature responses and their mutual relationships. Introducing further *causal links* between some latent variables, the CFA model is extended to a *structural equation model* (SEM), which allows us to reflect more complicated climatological relationsnhips with respect to all SEM's variables. Each statistical model is developed for use with data from a single region, which can be of any size. Here, in Part III, the performance of both these statistical models and some models suggested in Part I is evaluated and compared in a pseudo-proxy experiment, in which the true unobservable temperature is replaced by temperature data from a selected climate model simulation. The present analysis involves seven regional data sets. Focusing first on the ability of the models to provide an adequate and climatologically defensible description of the unknown underlying structure, we may conclude that given the climate model under consideration, the SEM model in general performed best. As for the factor model, its assumptions turned out to be too restrictive to describe the observed relationships in all but one region. The performance of another factor model, reflecting the assumptions typically made in many D&A studies, can be characterised as unacceptable due to its high sensitivity to insignificant coefficient estimates. Regarding the fourth statistical model analysed - a factor model with two indicators and one latent factor - it can be recommended to apply it with caution due to its sensitivity to departures from the independence assumptions among the model variables, which can make the interpretation of the latent factor unclear. The conclusions above have been confirmed in some form of a cross-validation study, presuming the availability of several data sets within each region of interest. Importantly, the present pseudo-proxy experiment is performed only for zero noise level, implying that the five SEM models and one factor model await further investigation to fully test their performance for realistic levels of added noise.

Keywords: Confirmatory Factor Analysis, Structural Equation models, Measurement Error

*Department of Mathematics, Stockholm University; katarina@math.su.se
†Department of Physical Geography, Stockholm university, Sweden; anders.moberg@natgeo.su.se
‡Department of Mathematics, Stockholm University; gudrun@math.su.se

models, Climate model simulations, Climate forcings, Climate proxy data, Detection and Attribution