Methods for climate change detection and attribution

Aurélien Ribes
CNRM - GAME, Météo France - CNRS, aurelien.ribes@meteo.fr

Abstract: Detection and attribution (D&A) have played a central role within the assessment of the human influence on climate and within IPCC's reports. Detection involves the statistical demonstration that a change has happened within climatic observations. Attribution consists in assessing the respective contributions of one or several causes to some observed change. Both require the use of climate model simulations, and are based on spatial or spatio temporal patterns of change. This paper provides a very short presentation of the classical "optimal fingerprint" method for D&A. Some recent developments, regarding the use of "error in variable" are introduced. Some of the challenging aspects of the method will be discussed too, in particular regarding the very large dimension of the typical datasets used.

Keywords: Climate change, detection, attribution, linear model, high dimension.

1 Introduction

Detection and attribution (D&A) have played a central role within the assessment of the human influence on climate and within IPCC's reports. Detection involves the statistical demonstration that a change has happened within climatic observations. Attribution consists in assessing the respective contributions of one or several causes to some observed change. Both are based on the characterisation of the spatial or spatio-temporal pattern of change corresponding to each physically plausible cause. However, specific tools from spatial statistics have been poorly used on that theme.

This paper aims primarily at giving a state of the art picture of some of the concepts, statistical tools, and current challenges in D&A analysis. The secondary attempt is to shortly discuss both difficulties and potential benefits of using spatial statistics tools.

Introduction of D&A first requires to introduce some concepts used in climate sciences. Climatologists use to first define their subject of study: the climate system. It includes the atmosphere, the ocean, and several other components (see IPCC, 2007). This system is influenced by several boundary conditions (e.g. the solar activity, the chemical composition of the atmosphere), usually referred to as *external forcings*, that may impact its state or dynamics. However, the variables used for describing the state of the system show some variability, even under fixed boundary conditions. This variability is called *internal variability*, and corresponds to the kind of variability expected while the climate is not changing.

Statistical D&A requires to have some knowledge on two parameters: first, the statistical properties or the distribution of the internal variability, and second, the expected response of the climate system to a given external forcing. Physically-based climate models are usually used for evaluating both objects instead of e.g. parametric models. Indeed, internal variability involves very specific spatial patterns and a large set of spatial scales that may hardly be accounted for in a parametric model. Instead, the use of climate model allows the evaluation from our physical understanding. D&A then requires careful comparison between observed changes and outputs from climate models.

2 Optimal fingerprint method

The more classical approach for climate change D&A is usually referred to as the *optimal fingerprint* method. This method has been gradually introduced at the end of the 90's (Hasselmann, 97, Hegerl et al., 97, Allen & Tett, 99). The latter presents this method as a linear regression of the observed climate time-series on the expected responses to the external forcings :

$$Y = \sum_{i=1}^{I} \beta_i g_i + \varepsilon, \tag{1}$$

where Y are the observations, β_i are unknown scaling factors, g_i is the expected response of the system to the i-th external forcing (as simulated by one or several climate model), and ε denotes the internal variability. In Eq. (1), Y is usually a spatio-temporal vector, Y_i typically consisting of the average of the temperature over a region, and a decade. g_i and ε have the same dimension and structure as Y.

Model (1) basically assumes that climate models have some accuracy at simulating the spatio-temporal pattern of the response to each external forcing, whereas they may fail at simulated the proper amplitude of that response. Within model (1), detection of a change associated to the forcing i corresponds to the rejection of the null hypothesis "$\beta_i = 0$". Attribution, in addition to the detection, requires to show that the observed response is consistent with the expected one, or equivalently, that the null hypothesis "$\beta_i = 1$" cannot be rejected.

Assuming that $C = \mathrm{Cov}(\varepsilon)$ is known, for example from climate models simulations, the computation of maximum likelihood estimate (MLE) for β is easy:

$$\widehat{\beta} = (G'C^{-1}G)^{-1}G'C^{-1}Y, \tag{2}$$

where $G = [g_1, \ldots, g_I]$. Under the same assumption, the distribution of the MLE is known, so as hypothesis testing on β is easy to perform.

Some refinement of the method has been introduced by Allen and Stott (2003) and Huntingford (2006), in order to take into account the uncertainty at simulating the spatio temporal patterns g_i. In that case, the uncertainty may come from internal variability (within the climate model simulation), or multi-model uncertainty. The main statistical model is then slightly changed to

$$Y = \sum_{i=1}^{I} \beta_i (g_i + \nu_i) + \varepsilon, \tag{3}$$

where ν_i represents the uncertainty on g_i. Assuming, similarly to ε, that $\Sigma = \mathrm{Cov}(\nu)$ is known, the optimal estimate of β may be derived by using a *Total Least Square* (TLS) procedure instead of the *Ordinary Least Square* technique involved in the MLE mentioned before.

Such methods have led to one of the important figures of the last IPCC report that deals with the quantification of the contribution of several external forcing to the observed warming (Figure 9.6, IPCC, 2007).

3 Estimation of C and high-dimension

The method presented before assumes that C (and Σ) is known, while, in a real-life problem, it is not. Several difficulties arise from the estimation of C, that is usually done from a *control* runs (i.e. climate simulations without any change in the external forcings). We here will focus in the problem related to the high-dimension of the typical global temperature datasets.

Current datasets are providing homogenised temperatures on a 5° x 5° grid, that results in 2592 grid-points in space. D&A study typically consider a 50-yr period in time, decomposed in 5 decades. The dimension of Y is then close to 13000. Note that missing values will likely decrease this number, but won't change the typical size of, say 10^4. Consequently, C is a 10^4 x 10^4 matrix, that has to be estimated from available control runs, that are typically covering 10^4 years (when considering together control runs from various models). Classical covariance matrix estimates being very poor in such cases, the dimensionality needs to be reduced.

Two approaches have been mainly used in order to reduce this dimension while focusing on the large spatial scales. First, global temperatures have been projected onto some first spherical harmonics (e.g. Stott, 2006). Second, particularly at the regional scale (where the dimension of the dataset is smaller but remains too high), data have been projected onto the first principal components (e.g. Zwiers, 2003). In both cases, projection may reduce the accuracy of the β estimates (there are no results of optimality), and requires to choose the reduced dimension (i.e. the number of spherical harmonics or principal components), what may be sensitive.

One possible alternative consists in using a regularised estimate of the covariance matrix C, that is a linear combination of the empirical covariance matrix estimate $\widehat{C}$ and the identity (Ribes et al., 2009) :

$$\widetilde{C} = \gamma \widehat{C} + \rho I. \tag{4}$$

Such an estimate has been shown to be more accurate than $\widehat{C}$ in high-dimension (Ledoit and Wolf, 2004). To plug (4) into (2) also leads to an improved estimate of β in the context of high dimension data.

This approach may help the estimation of β, but no result of optimality has been proved. As a consequence, the problem of efficiently estimating β in the context of high dimension dataset is still open. One potentially attractive way may be to use the spatio-temporal structure of Y in order to improve the estimation of C.

4 Concluding remarks

D&A deal with one key-question regarding climate change, that is the quantification of the human contribution to the current warming. While initially based on a simple linear model, D&A involve some recent statistical tools and also provide some challenging questions, in particular related to the high dimension of the corresponding datasets.

References

Allen M. R. and Tett S. F. B. (1999) Checking for model consistency in optimal fingerprinting, *Climate Dynamics*, 15, 419-434.

Allen M. R. and Stott P. A. (2003) Estimating signal amplitudes in optimal fingerprinting, Part I: Theory, *Climate Dynamics*, 21, 477-491.

Hasselmann K. (1997) Multi-pattern fingerprint method for detection and attribution, *Climate Dynamics*, 13, 601-612.

Hegerl G. C. et al. (1997) Multi-fingerprint detection and attribution of greenhouse-gas and aerosol-forced climate change, *Climate Dynamics*, 13, 613-634.

Huntingford C. et al. (2006) Incorporating model uncertainty into attribution of observed temperature change. *Geophysical Research Letters*, 33, L05710.

IPCC (2007) Climate change 2007: the physical basis. Contribution of working group 1 to the fourth assessment report of the international panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996pp.

Ledoit O. and Wolf M. (2004) A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis*, 88, 365-411.

Ribes A., Azaïs J.-M. ans Planton S. (2009) Adaptation of the optimal fingerprint method for climate change detection using a well-conditioned covariance matrix estimate, *Climate Dynamics*, 33, 707-722.

Stott P. A. et al. (2006) Observational constraints on past attributable warming and predictions of future global warming, *Journal of Climate*, 19, 3055-3069.

Zwiers F. W. and Zhang X. (2003) Towards regional scale climate change detection, *Journal of Climate*, 16, 793-797.