Econometric methods in Climate Research

ECONCLIM

17 September 2003

Type of Instrument: STREP

Submission stage: OUTLINE proposal

Activity code: NEST-2003-1 ADVENTURE

Duration of project: 3 years

Proposal Abstract

Many statistical analysis methods currently used in climate research rely on the simplest of statistical assumptions – development of advanced methods that take into account such effects as serial correlation has regression results, and other highly realistic statistical problems, have not been widely appreciated in climate research. Development of advanced statistical methods in the fields of econometrics, theoretical physics and geostatistics has now surpassed the methods still used in climate research. We want to transfer knowledge of some of these advanced methods into climate science, because of the expected gains in reliability of statistical work, and the impacts of this on society and climate physics. We have three areas of statistics we wish to transfer, and three climate problems inside of which we will investigate the impact of the new methods. We choose a common model-based dataset for analysis in order to simplify the focus of the project. The proposal does not fall within any other FP6 thematic priority. In particular, it is not an ordinary climate research proposal, but a cross-disciplinary project intended to transfer knowledge so that future climate research may benefit.

Preamble: We want to bring experts in applied statistics together with experts in climate research, with the specific goal to transfer some of the very advanced methods, which have proven to be fruitful in former field, to the latter. We want to do this because statistics is widely used in climate research but nevertheless the methods used seem to be neglecting fundamental problems, which can be overcome by using methods used in other fields, such as econometrics. Therefore, we believe strongly that huge benefits for the quality of climate research, and therefore society, will result. Our project is thus a 'knowledge-transfer' project, and seeks the means to bring experts from the two fields together and test, develop and transfer powerful methods already used in other fields.

Introduction: Presently, the application of statistical methods in climate science appears somewhat stagnant and isolated, compared to fileds like econometrics and others, despite the fact that the problems studied have common characteristics. Most recent effort in climate science has been invested in climate *modelling*, but similar resources have not been invested in developing or introducing new wstatistical analysis methods. The least squares method was apparently invented by Legendre around 1800 and has become one of the most widely used techniques for estimating parameters in linear models in climate research. However, this method rests on tacit assumptions that are not explicitly tested or reviewed during use. Since the 1950's important improvements to this method have been developed and presented by econometrics, and widely used in that field, but has not spread into climate research. Appreciative improvements in estimating regression and other linear models will be available for climate research, by the adoption of these new methods, and we wish to effectuate that transfer in this project.

The main objectives of our proposed work are therefore:

- To transfer advanced statistical methods to climate science, which have been applied successfully in other fields such as econometrics, geostatistics and statistical physics, and
- assess their potential applicability in selected climate problems.

This research is very novel because transfer of knowledge has never been done from other *applied* statistical fields to the climate research community in a *systematic* way. Transfer efforts have been seen between the field of *mathematical* statistics and climate science – e.g. the Geophysical Statistics Project (GSP) at NCAR (http://www.cgd.ucar.edu/stats/), and the International meetings on Statistical Climatology (IMSC). We believe, that the transfer of well-tested and widely used advanced methods from other *applied* fields into the field of climate research is a superior approach.

The project is very ambitious because we are challenging a number of tacit assumptions commonly made in the analysis of climate data. It is commonly assumed that processes in climate have short memory, and that standard regression assumptions apply (i.e., that residuals are 'white' or independent). We are considering methods that do not assume these things and offer readily available remedies. Bringing these methods into climate science may bring into question some commonly accepted results, such as climate reconstructions, averaged instrumental period climate series, and climate change detection/attribution work.

We expect the impact of our work to be like the difference between planting a tree and planting a forest: Individual efforts to transfer econometric, or other advanced statistical methods, to climate science have been seen (e.g. some papers here Toll/Voss, Kaufman/Stern, Thomson), but these remain isolated – like trees in a field, rather than a growing forest. An example of a successful 'forest-growing' transfer is the use of principal component analysis (PCA) methods in climate science – here hundreds, if not thousands, of papers have been written on these methods and they have become standard in the community. What we want to do is transfer *other* methods that are equally important and spread the word to the climate research community. Incidentally, the PCA methods commonly rely on the very type of standard assumptions that we wish to challenge and offer better methods for, so the potential impact of our idea can be seen. Simply put, we have the ambition to change statistical practice in climate science with our proposed project.

We have a scientific approach that we believe is convincing because we have formed a group of leading experts from several communities intent on the knowledge-transfer. On the one hand are experts with profound experience of applying their methods to a variety of cases, such as: financial market dynamics, petroleum and mining exploration and biomedical signal analysis. On the other hand are climate scientists experienced in the statistical analysis of climate data. The different concepts and methods will be tested in a common framework of extensive datasets generated with climate models. The use of such a common dataset, and a working structure whereby statistical data analysis is performed, not in isolation but in close collaboration and conference with climate experts, ensures that such potential problems as nomenclature differences are overcome.

The concepts and methods to be transferred have all been shown to work very successfully within their field of origin – we also believe that the types of data encountered in the various fields are very similar - and this makes our proposal plausible and the effort justified. By identifying a number of 'benchmark' problems to be analysed with a selection of advanced methods we can demonstrate that the knowledge transfer will occur – intensive work on a joint dataset inside agreed-upon problem limits will give us a common understanding of what sort of problems and opportunities lie in these data and methods. We have chosen a common dataset based on highly resolved (twice daily) long (1000 and 10,000 year long) forced and unforced climate model runs.

Objectives: To transfer advanced statistical methods to climate science, which have been applied successfully in other fields such as econometrics, geostatistics and statistical physics, and assess their potential applicability in specific climate problem.

Outputs:

- Identification of possible caveats/pitfalls due to previous practise in determinations of climate reconstructions, climate change detection and attribution, downscaling or disaggregation.
- Methods to overcome these problems.
- Suggestions for changing statistical practise, the feasibility of these suggestions being demonstrated in the analysis of the model data sets.
- Software packages
- Extensive dissimination by arranging special sessions at climate meetings, writing scientific papers, using networks, such as CLIVAR and PAGES.

These objectives are realistic inside the 3 year duration of the proposed project because we are a team of participants with experience within our own fields and therefore are able to pinpoint relevant climate problems and the appropriate methodology. We also have the necessary contacts to the scientific networks in order to disseminate our results to the broad community of climate research.

Our common pool of data is already at hand from the start of the project, thus avoiding a delay from the beginning experienced within many other projects.

State of the art: No one has attempted to transfer this type of knowledge from applied statistics to climate science, on the scale we are considering. Competing research approaches (GSP, IMSC) have mainly chosen to transfer knowledge from the field of *mathematical* statistics, whereas we seek to transfer successfully tested methods that are in use in other fields. The state of the art in 'ordinary' statistics is to assume that for instance residuals in regressions are un-correlated and that the Gauss-Markov theorem underlying ordinary least squares (OLS) is therefore satisfied. However, this is often not the case, and this is often not realized, but readily available methods exist for taking residual serial correlation into account, known in econometrics. Likewise, the role of 'long term memory' or 'integration/cointegration' in time series and its impact on such mainline use as regression has not been considered in climate research. Taking the next step and assuming that time series may have serial correlation, and be describable as autoregressive series of order 1 is about as far as most applications in climate statistics have taken matters. Econometrics has a wide panoply of standard methods available for use in cases where data commonly is not well described even by an AR(1) process – e.g. co-integration methods.

Assessing technical risk elements: We have identified three main risk elements in our proposal:

1. *The new methods may not give superior results after all*
    - We do not think this a problem because the data encountered in econometrics, geoscience, biomedical applications and climate science are not dissimilar, so improved results are expected when the more sophisticated methods are transferred to climate science. We have some *indications* that new methods may give improved results, but it needs to be checked via the suggested joint analysis of model data, and checked on the basis of physical expectations about the climate system.
2. *We may not able to communicate across disciplines*
    - One aspect of this risk is *inside* the project group, but the funding we ask for will ensure the presence of the usuall tools required for avoiding this problem: <u>Frequent meetings and workshops</u> will be held in the project group and people simply talk to each other and sort out nomenclature problems and by collegial interactions pave the way for common understanding of problems and methods, <u>joint paper writing</u> is an even more intensive step in that direction, and the <u>composition of working teams</u> ensures communication across discipline borders because statisticians and climate scientists will be involved in each of the working groups we plan – so no abstracted data analysis without a foundation in reality will be possible. Another aspect is the risk of failing to communicate the availability and superiority of the 'new advanced' methods to climate scientists *outside* our project group: However, we will be active communicating the results, not just in papers but by actively taking the role of convening inter-disciplinary sessions at scientific meetings such as the annual EGS/EUG meetings, the AGU meetings, and smaller, more specific gatherings such as the GPS and ISCM. We will also spread news of what we are doing and accomplishing in such vehicles as the publications from CLIVAR and PAGES which cater to various climate science communities that rely heavily on 'standard statistical' methods, such as OLS.
    - During analysis of the model data we may come across the practical problem of not having a joint software platform to use for data analysis, but this is simply remedied by ensuring access to such a platform (e.g. the EVIEWS programme) – and we ask for funding to do this. *Outside our project group* we envisage that suitable tools should be programmed in the platforms commonly used in the climate research community so that this is no hindrance for adaptation and spread of the methods. IDL or Matlab platforms are commonly used in climate research for data analysis.

- *Methods may not be transferable*
    - It may be that method limitations make the transfer to climate science impossible. For instance, the climate system may have 'regime change behaviour' or show other nonlinear dynamics. If regime changes are frequent this problem is not inhibiting as methods can be adapted, but if climate regime changes occur once or twice in an observed interval then the methods may fail – but this will be a joint problem for all other methods, standard as well as advanced.

Methodology: Our proposed project is a knowledge-transfer project between certain fields of applied statistics and climate science – so we have tasks related to transferring knowledge and tasks related to providing for the framework inside of which we shall work – that is, the model data which is to be the central focus of our project, and upon which the various methods will be tested and modified. We have three 'methods groups' and have identified three 'climate problems' that these groups will work on. The methods are those related to 'econometrics', 'geostatistics', and 'long memory' processes, while the selected climate problems are those of reconstructions of historical climate, detection and attribution of climate change, and 'disaggregation' and spatial methods that take correlations into account.

Project structure:
Project management (WP0): Those management tasks that can be handled without reference to the scientific work will be handled by a professional management partner. Scientific management tasks will be handled by the coordinator.

Data provision (WP1): The data will be provided by a group in the project in possession of highly resolved, long (1,000 years) coupled climate model runs, and assisted by another group in possession of even longer (10,000 years) coupled model run from a climate model of 'intermediate complexity'. This task involves extracting data of a predetermined type from a large body of model outputs already in existence.

Econometric methods (WP2): The econometricists will participate in the work on all the climate problems considered in the project since ordinary least squares methods dominate as estimation tool in climate science and since improved alternatives exist in the econometricists toolbox. These alternatives range from estimation of linear models with auto-dependent residuals to estimation of models relating non-stationary timeseries and test of time series for non-stationarity. One important climate forcing (emission of greenhouse gases) is related to economic activity, which make us believe in the successful application of the econometric methods.

Geostatistics (WP3): Concepts from geostatistics are applicable in such questions as the formation of a global average temperature from unevenly distributed observing stations across the world. Commonly, series are boxed and averaged, without much attention paid to the variations in decorrelation length across the World. Improvements in this field are expected. Disaggregation methods refer to the construction of accurate point values, given local averages. This must take spatial deceleration scales into account and has applications in local climate and meteorological prediction.

Long-memory processes (WP4): Theoretical physics provides insight into time series structure. These insights can be used to diagnose time series from models and reconstructions – it is possible to seek common traits and use the presence or absence of these to understand relationships between series – for example it is possible to look at long temperature reconstructions and see if specific properties of these series are shared with observational series, and on the basis of the similarities or absence thereof make statements about how realistic the reconstruction is.

Dissemination (WP5): The work of making sure knowledge is transferred between the communities involved will take place in the work itself, as explained above, but also handled by a specific dissemination task. The leader of this task will be the project co-ordinator, who will be responsible, but all project participants will be assigned particular duties in the interest of dissemination, such as attending scientific meetings and presenting our progress, preparing materials for such assessment reporting agencies as CLIVAR and PAGES, and to ‘spread the word’ of the project goal and its potential benefits.