Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
36th NOAA Annual Climate Diagnostics and Prediction Workshop
Fort Worth, TX, 3-6 October 2011

Seasonal Forecasting Using the Climate Predictability Tool (CPT)

S. J. Mason

International Research Institute for Climate and Society, The Earth Institute of Columbia University

The CPT is a software package developed by the International Research Institute for Climate and Society (IRI) designed for making seasonal climate forecasts. The CPT is an easy-to-use tool that runs on Windows 95+, and a source-code version that can be compiled under other types of operating systems is available for running batch jobs. It is available free of charge from the IRI's web page: http://iri.columbia.edu/climate/tools/cpt/. At the time of writing, the latest version of CPT (11.10) consists of about 50,000 lines of Fortran 95 code, and uses compiler extensions for the user-interface and graphics, and so is not portable to other platforms. The source-code version consists of about 33,000 lines of Fortran 90 code, and provides no graphics functionality. This version has been tested with gfortran, ifort, nagfor, pgf90, and pgf95, but up-to-date versions of the compilers are required to avoid issues with compiler bugs.

The underlying goal in developing the CPT has been to promote the widespread creation and communication of quality-controlled seasonal climate forecasts that address specific needs of different user groups. The software was initially developed to enable forecasters at National Meteorological Services (NMSs) in Africa to produce updated seasonal forecasts for their country, and to provide greater consistency in inputs to the Regional Climate Outlook Forums (RCOFs) to facilitate consensus building, but the CPT has been used widely beyond the RCOFs (Fig. 1).

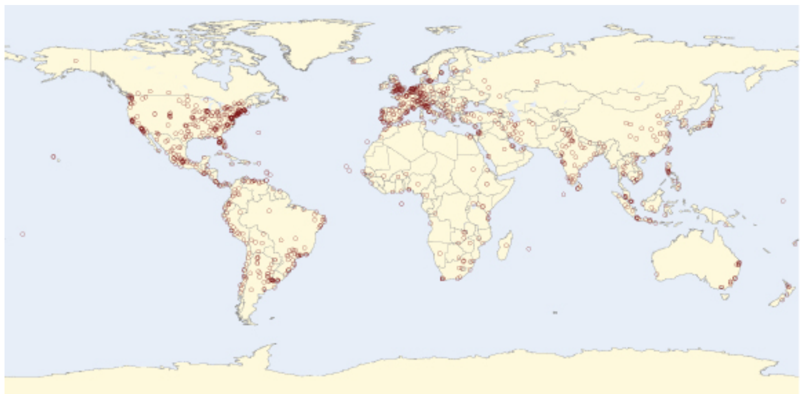

Fig. 1 Locations of downloads of the CPT software, 2004 – 2011.

There are two main approaches to generating seasonal forecasts: using large-scale models of the global atmosphere, known as general circulation models (GCMs), or using a statistical approach to relate seasonal climate to changes in sea-surface temperatures, such as those associated with El Niño, or to other predictors. In the former case, predictions are made for large-areas, and are often not very relevant for specific locations. In addition, because of the coarse scale at which the GCMs operate, the geography in the models is often distorted, and so geographical locations can be displaced. These GCM outputs therefore need to be adjusted so that they can be applied at the local level. The CPT tool is designed to perform both forms of prediction, namely downscaling of GCM output, and purely statistical predictions.

The statistical approach to making seasonal forecasts from sea-surface temperatures has been used for a number of years at many National Meteorological Services. Since the late 1990s, these statistical forecasts have been combined to produce a consensus forecast, representing a patchwork of nationally-based forecasts for subcontinental areas, in Regional Climate Outlook Forums (RCOFs; Ogallo *et al.* 2008). While such forums have been very successful in building the capacity to produce seasonal climate forecasts, a number of problems have emerged, and some systematic errors in the forecasts have been identified (Mason and Chidzambwa 2009). The CPT tool was designed in response to these problems, and, specifically, was developed to address the following issues:

Correspondence to: S. J. Mason, International Research Institute for Climate and Society, The Earth Institute of Columbia University, Palisades, NY; E-mail: simon@iri.columbia.edu.

a. *Slow production time*: the time taken to construct the statistical forecasts at the RCOFs was requiring long and expensive pre-forum workshops. By using CPT, forecasts can be produced in just a few hours or less. The quick production time makes it possible to hold shorter, and thus cheaper, pre-forum workshops, and provides more time at these workshops for advanced training. It also makes it viable for forecasters to produce updated forecasts on a monthly basis back in their home countries.

b. *Artificial skill*: an important step in producing a reliable seasonal forecast is to obtain a realistic estimate of how good the model predictions are. CPT performs rigorous cross-validation and retroactive tests for estimating skill levels, and adjusts the forecasts accordingly. Extensive diagnostic statistics are provided, including most of the scores and procedures recommended by the WMO CBS SVSLRF (including the calculation of significance levels and error bars).

c. *Dependence on one model*: it has been demonstrated extensively that the best seasonal forecasts are produced by generating a number of predictions and then combining these, perhaps by simply taking the average of the predictions. In the past, however, there has been a strong reliance on the prediction from a single model, simply because the effort invested in constructing the statistical model tended to encourage an over-confidence in the prediction from this model. Inputs from other sources (most notably the GCM predictions) were largely down-played because of the lack of a sense of ownership of these additional products. By making it much easier to generate a set of predictions by using CPT, the official forecasts from the National Meteorological Services and from the RCOFs now consider a broader range of inputs than was possible in the early years of seasonal climate forecasting efforts.

d. *Unreliable probabilities*: seasonal climate forecasts are expressed probabilistically because of the large uncertainties involved in forecasting the next few months. However, most statistical forecasting methodologies do not explicitly indicate the uncertainty in the prediction, and so this uncertainty has to be estimated. Some simple and intuitive ways of estimating forecast uncertainty are unfortunately not very reliable (Mason and Mimmack 2002). A more reliable system has been implemented in CPT based upon the error variance of the cross-validated forecasts over the training period.

e. *Forecast format*: seasonal forecasts are typically presented as the probability that the seasonal rainfall total, for example, will fall within pre-defined ranges. These ranges are most commonly set as the upper and lower terciles of the historical rainfall totals. This format tends to be unpopular with users of the forecasts, partly because the forecast is too abstract (it is not clear how much rainfall is meant by the upper and lower terciles), and partly because the forecast is too unspecific (the upper and lower terciles are often not very interesting thresholds). Within CPT there are options that provide considerable flexibility to tailor forecasts for specific user requirements, including options to redefine the ranges. Forecasts can be produced in a variety of formats, and detailed information is provided so that the forecast can be communicated to the end users in easy-to-understand terms.

Although the software is designed to implement best-practices, a second underlying principle has been to respond to user requests for improved functionality and ease of use. Major enhancements to the CPT have been released approximately once per year. The major new versions are indicated in Table 1.

The software has been introduced to most of the RCOFs and is now used fairly extensively: there are currently over 650 registered users of the Windows versions from all over the globe. Visitors to the CPC Africa Desk are provided extensive training in the software. Numerous training workshops in the use of the software have already been held, including under the auspices of the WMO CLIPS activities.

References

Mason, S.J., and S. Chidzambwa, 2009: Verification of RCOF Forecasts. *WMO RCOF Review 2008 Position Paper*, 26 pp.

Mason, S.J., and G.M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *Journal of Climate*, **15**, 8–29.

Ogallo L.J., P. Bessemoulin, J.P. Ceron, S.J. Mason, and S.J. Connor, 2008: Adapting to climate variability and change: the Climate Outlook Forum process. *J. World Meteor. Org.*, **57**, 93–102.

CPT Versions		
Version	**Date**	**New features**
CPT 0		MATLAB code for performing canonical correlation analysis
CPT 1	Dec 2002	Translation into about 650 lines of Fortran 77 as interface to LAPACK SVD routines (requires recompilation each time) Principal component regression
CPT 2	Aug 2003	Conversion to Fortran 95 Graphical user interface Validation statistics Option to calculate a forecast using updated predictors
CPT 3	Feb 2004	Mapping of station data Missing value estimation
CPT 4	Feb 2005	Improved graphics Bootstrap confidence intervals and p-values
CPT 5	Aug 2005	Forecast uncertainty estimates based on prediction intervals WMO SVSLRF verification procedures Optional transformation of predictand to Gaussian distribution Options to define categories using climatological probabilities other than 33%, or using absolute values
CPT 6	Nov 2005	Multiple concurrent users Exceedance probability curves Options to calculate forecasts as, and to define categories in terms of, anomalies or standardized anomalies
CPT 7	Aug 2006	Option to set model update interval in retroactive procedure Option to set a zero-bound (useful for precipitation forecasts) Option to calculate forecasts as, and to define categories in terms of, % of average
CPT 8	May 2007	Retroactive forecast probabilities and verification including attributes diagrams Multiple linear regression option Permitted input of daily data
CPT 9	Mar 2008	Break up of executable to reduce executable size by using DLLs to simplify download of updates Major internal restructuring of code for closer parallel between Windows and Linux versions Additional verification procedures
CPT 10	Oct 2009	Multiple predictor fields, including extended EOFs New user-interface "Ensemble" forecasts based on error variance Additional verification procedures New input data formats for improved identification of forecast lags Forecasts as odds Option to define categories in terms of analogue years Calculation of p-values for skill maps Improved graphics functions
CPT 11	Jan 2011	Multi-lingual user-interface GCM verification option Probabilistic verification scores for retroactive forecasts Category thresholds, averages, and correlation maps Improved missing value estimation Improved handling of daily data Prevention of over-fitting of CCA models Decreased sensitivity to different compilers

Table 1 History of CPT versions indicating the major new features.