Climate Predictability Tool (CPT)

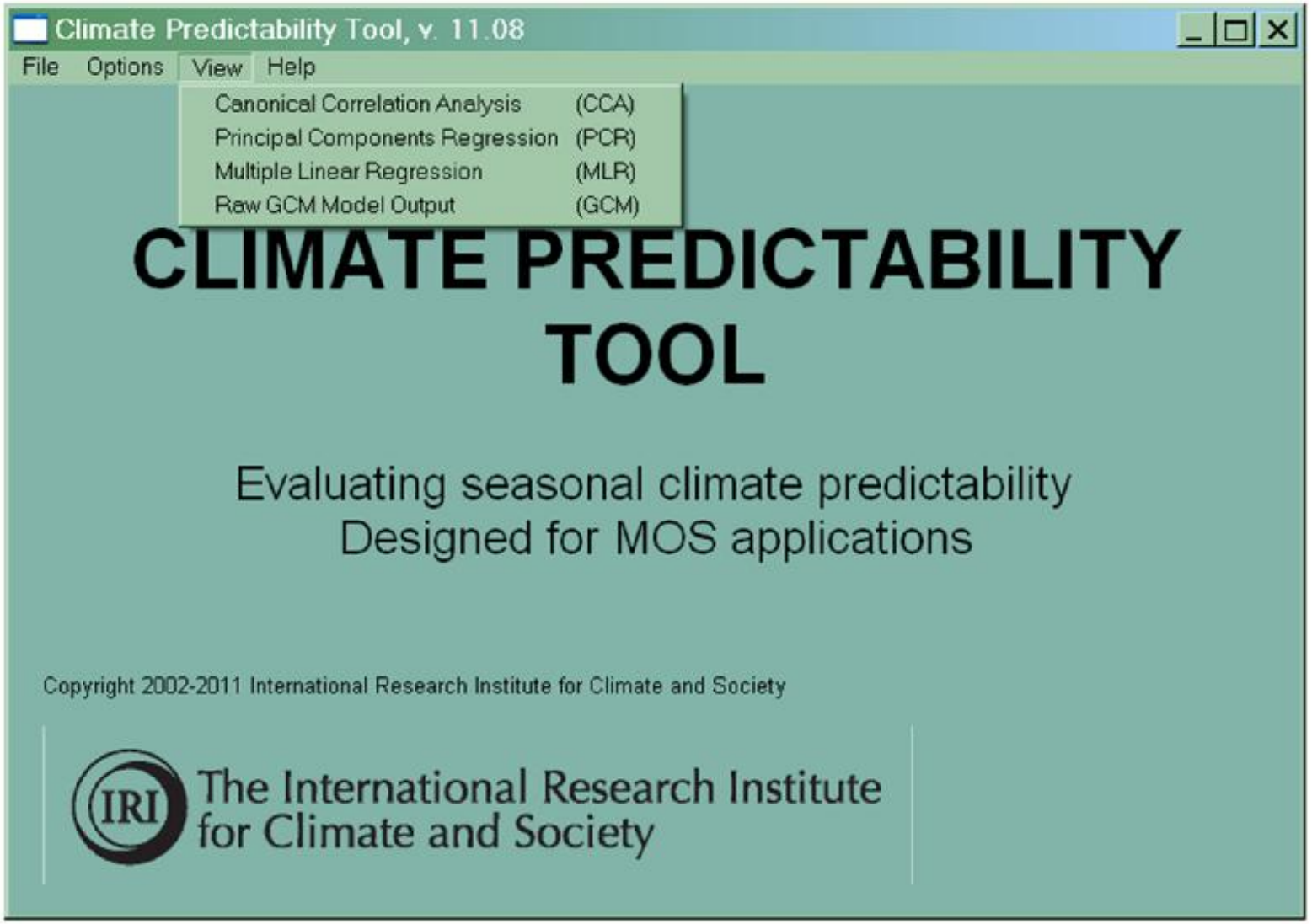
Ousmane Ndiaye and Simon J. Mason
cpt@iri.columbia.edu

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

OVERVIEW

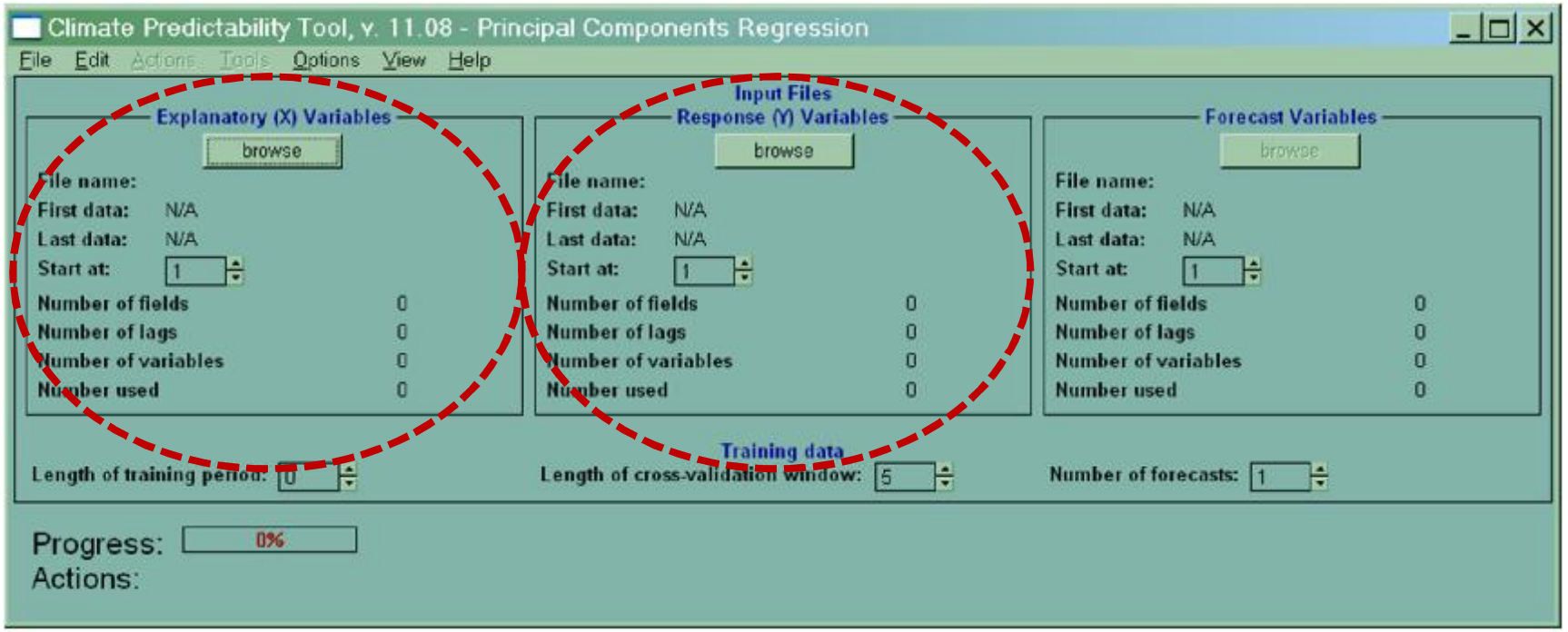
- The Climate Predictability Tool (CPT) provides a Windows package for :
  - seasonal climate forecasting
  - model validation
  - actual forecasts given updated data
- Uses ASCII input files
- Options :
  - Principal Components Regression (PCR)
  - Canonical Correlation Analysis (CCA)
  - Multi Linear Regression (MLR)
  - Global Model Output (GCM)
- Help Pages on a range of topics in HTML format
- Options to save outputs in ASCII format and graphics as JPEG
- Program source code is available for those using other systems (e.g., UNIX)

SELECTING THE ANALYSIS



Choose the analysis to perform: PCR, CCA, MLR, or GCM

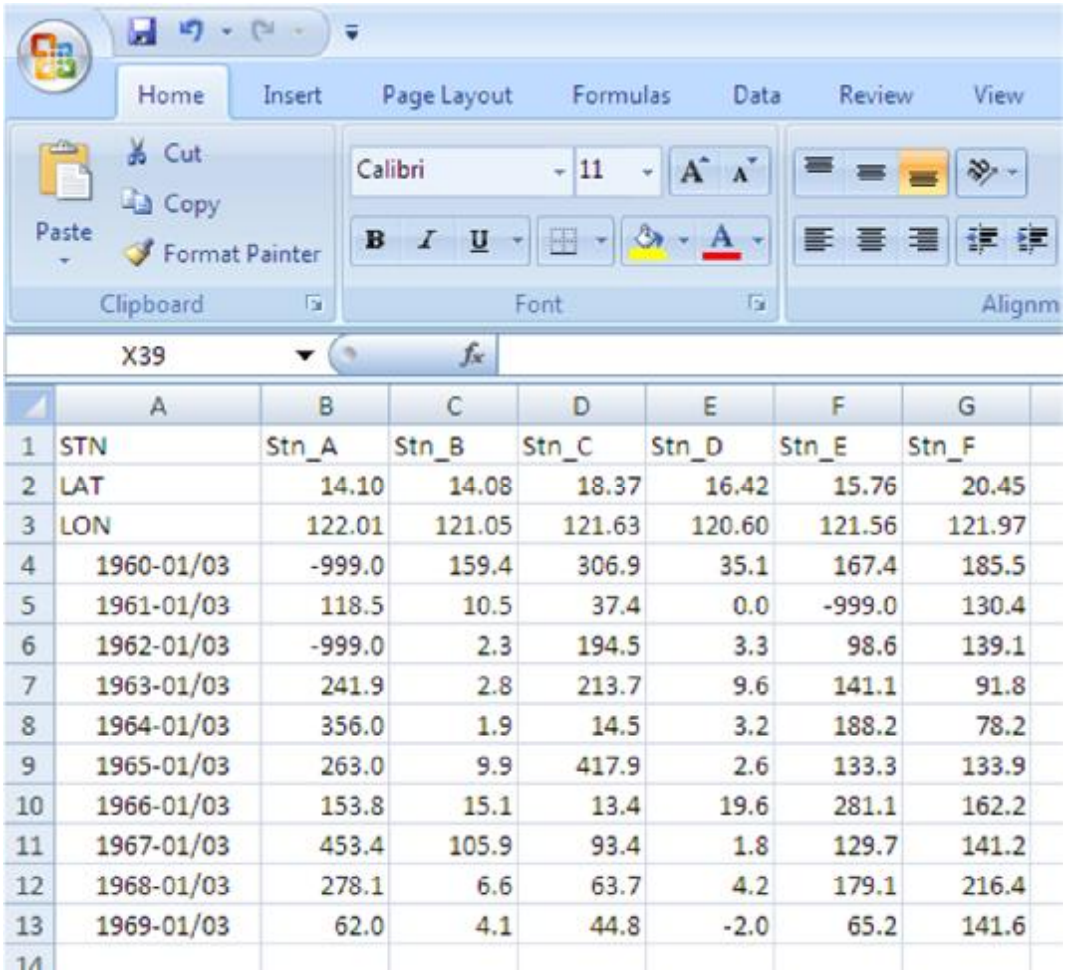
INPUT DATASETS



All analysis methods require two datasets:
"X variables" or "Predictors" dataset;
"Y variables" or "Predictands" dataset.

CPT INPUT FILE FORMATS

1. STATION files



	A	B	C	D	E	F	G
1	STN	Stn_A	Stn_B	Stn_C	Stn_D	Stn_E	Stn_F
2	LAT	14.10	14.08	18.37	16.42	15.76	20.45
3	LON	122.01	121.05	121.63	120.60	121.56	121.97
4	1960-01/03	-999.0	159.4	306.9	35.1	167.4	185.5
5	1961-01/03	118.5	10.5	37.4	0.0	-999.0	130.4
6	1962-01/03	-999.0	2.3	194.5	3.3	98.6	139.1
7	1963-01/03	241.9	2.8	213.7	9.6	141.1	91.8
8	1964-01/03	356.0	1.9	14.5	3.2	188.2	78.2
9	1965-01/03	263.0	9.9	417.9	2.6	133.3	133.9
10	1966-01/03	153.8	15.1	13.4	19.6	281.1	162.2
11	1967-01/03	453.4	105.9	93.4	1.8	129.7	141.2
12	1968-01/03	278.1	6.6	63.7	4.2	179.1	216.4
13	1969-01/03	62.0	4.1	44.8	-2.0	65.2	141.6
14							

This file-type contains :

Station_name (without spaces; ≤16 characters)

Latitude (south negative)

Longitude (west negative)

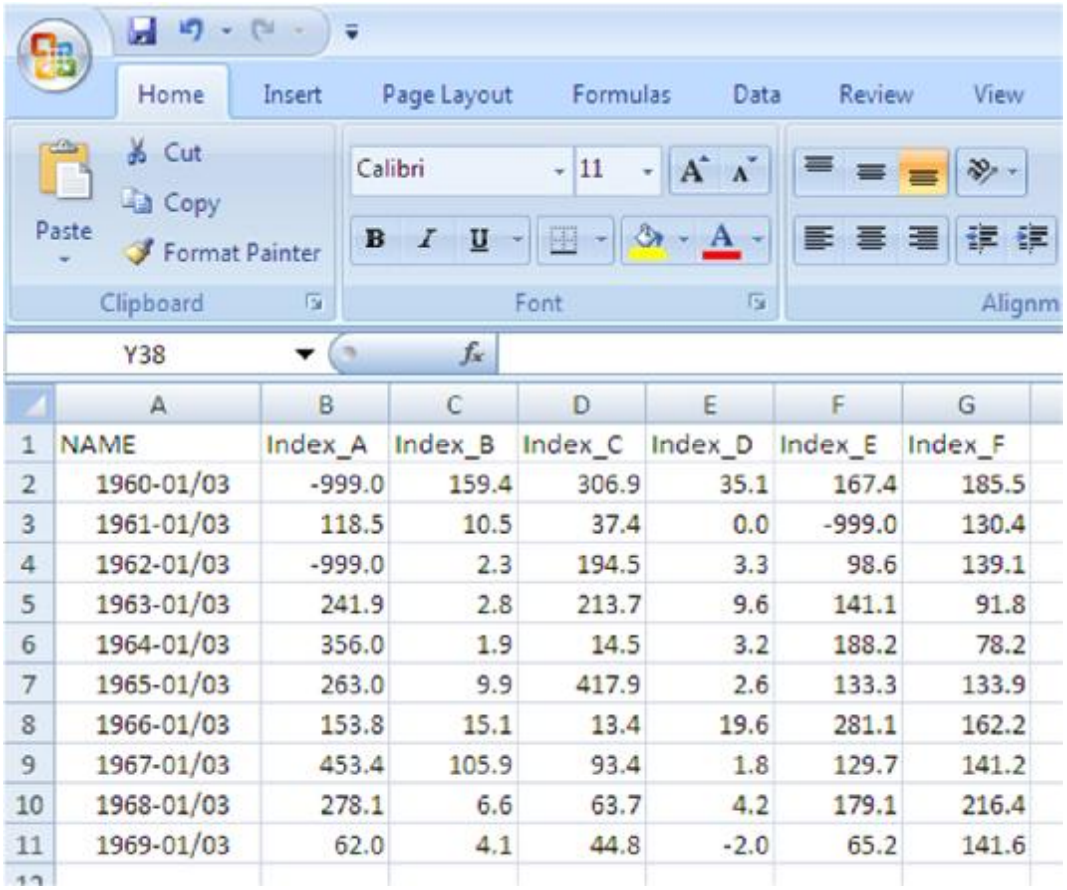
Year-season (in the first column)

Data (missing values should be filled with the same value, -999.0 for example)

Keywords:
STN, LAT, LON

CPT INPUT FILE FORMATS

2. UNREFERENCED or INDEX files



	A	B	C	D	E	F	G
1	NAME	Index_A	Index_B	Index_C	Index_D	Index_E	Index_F
2	1960-01/03	-999.0	159.4	306.9	35.1	167.4	185.5
3	1961-01/03	118.5	10.5	37.4	0.0	-999.0	130.4
4	1962-01/03	-999.0	2.3	194.5	3.3	98.6	139.1
5	1963-01/03	241.9	2.8	213.7	9.6	141.1	91.8
6	1964-01/03	356.0	1.9	14.5	3.2	188.2	78.2
7	1965-01/03	263.0	9.9	417.9	2.6	133.3	133.9
8	1966-01/03	153.8	15.1	13.4	19.6	281.1	162.2
9	1967-01/03	453.4	105.9	93.4	1.8	129.7	141.2
10	1968-01/03	278.1	6.6	63.7	4.2	179.1	216.4
11	1969-01/03	62.0	4.1	44.8	-2.0	65.2	141.6

This file-type contains :

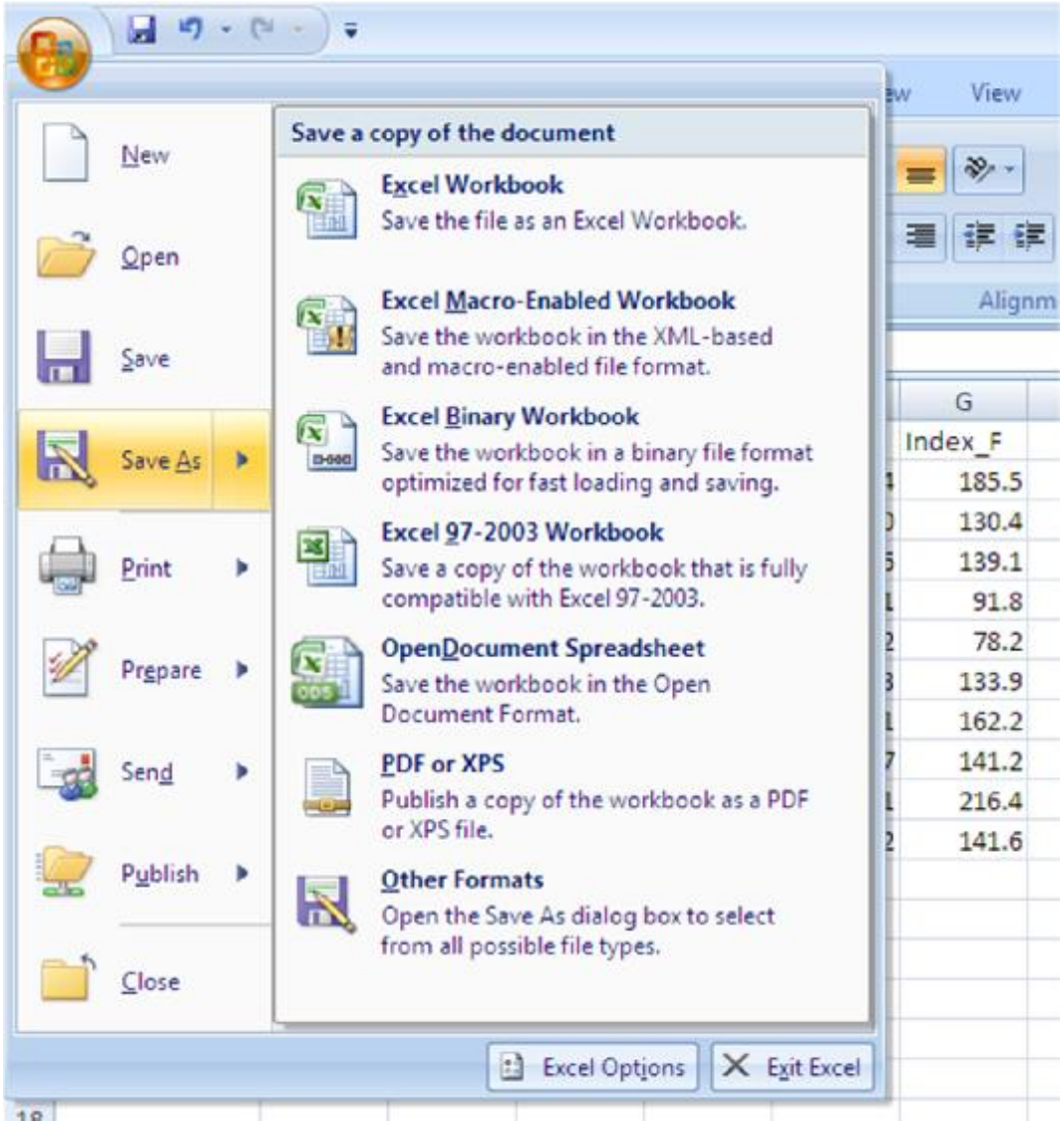
Index_name (without spaces; ≤16 characters)

Year-season (in the first column)

Data (missing values should be filled with the same value, -999.0 for example)

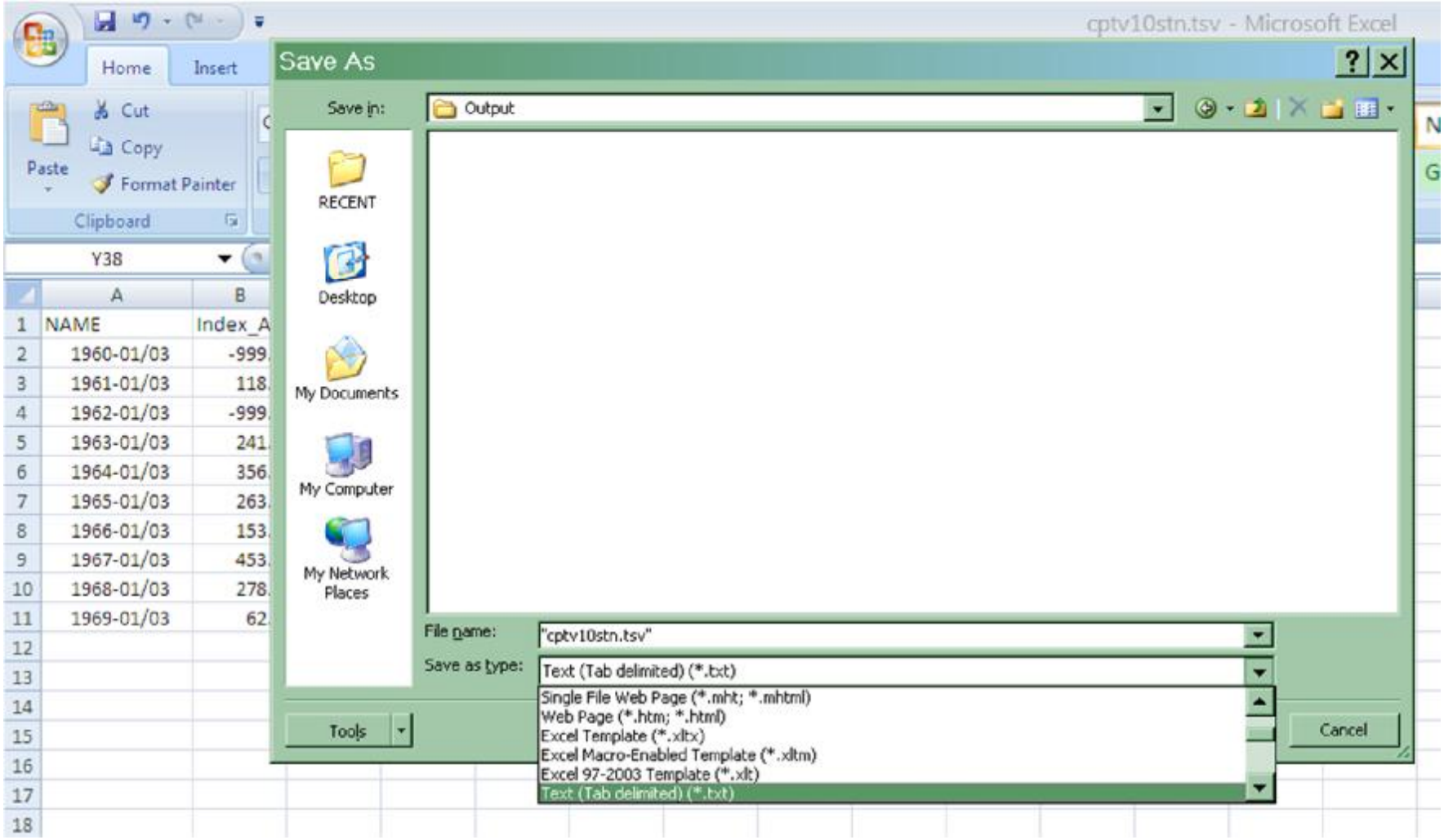
Keywords:
NAME or YEAR

CPT INPUT FILE FORMATS



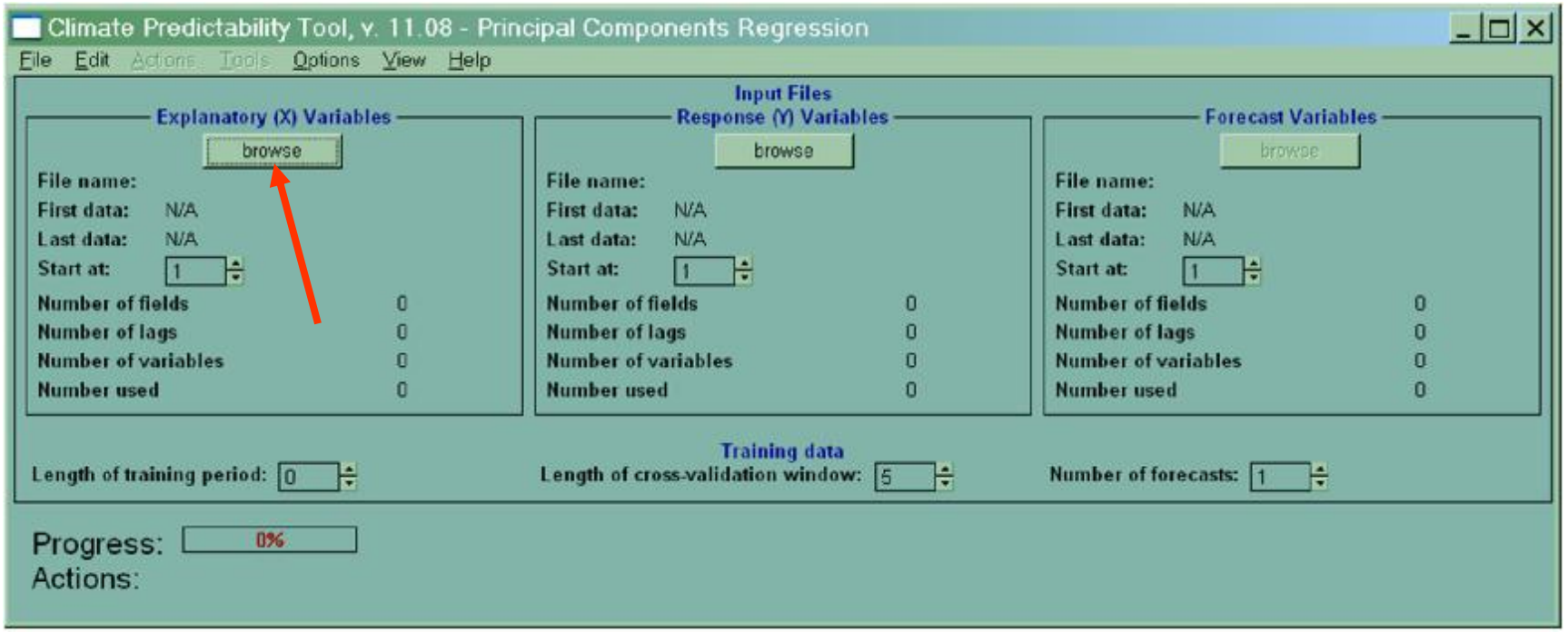
The input files could be easily made using a spreadsheet such as Excel

CPT INPUT FILE FORMATS



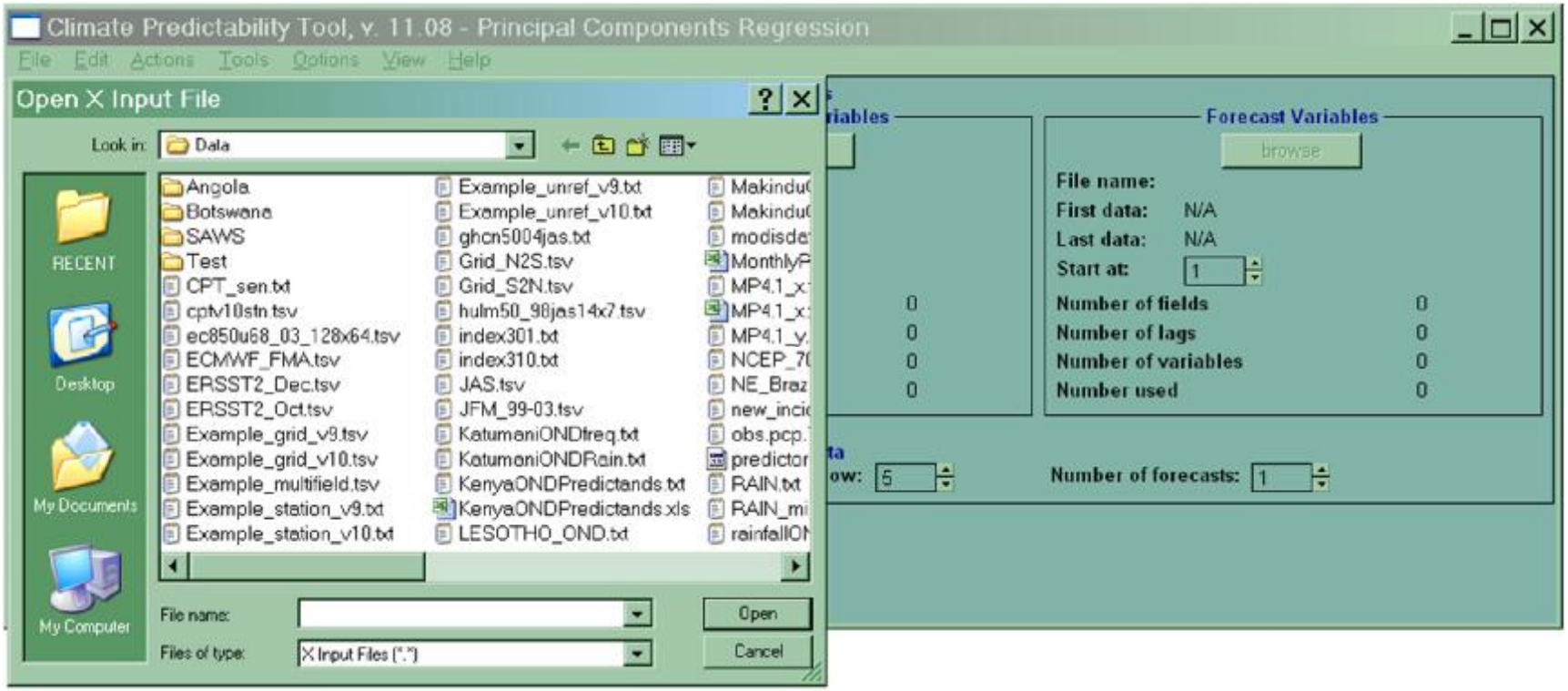
In Excel the file should be saved as:
"Text (Tab delimited) (*.txt)"

SELECTING INPUT FILES



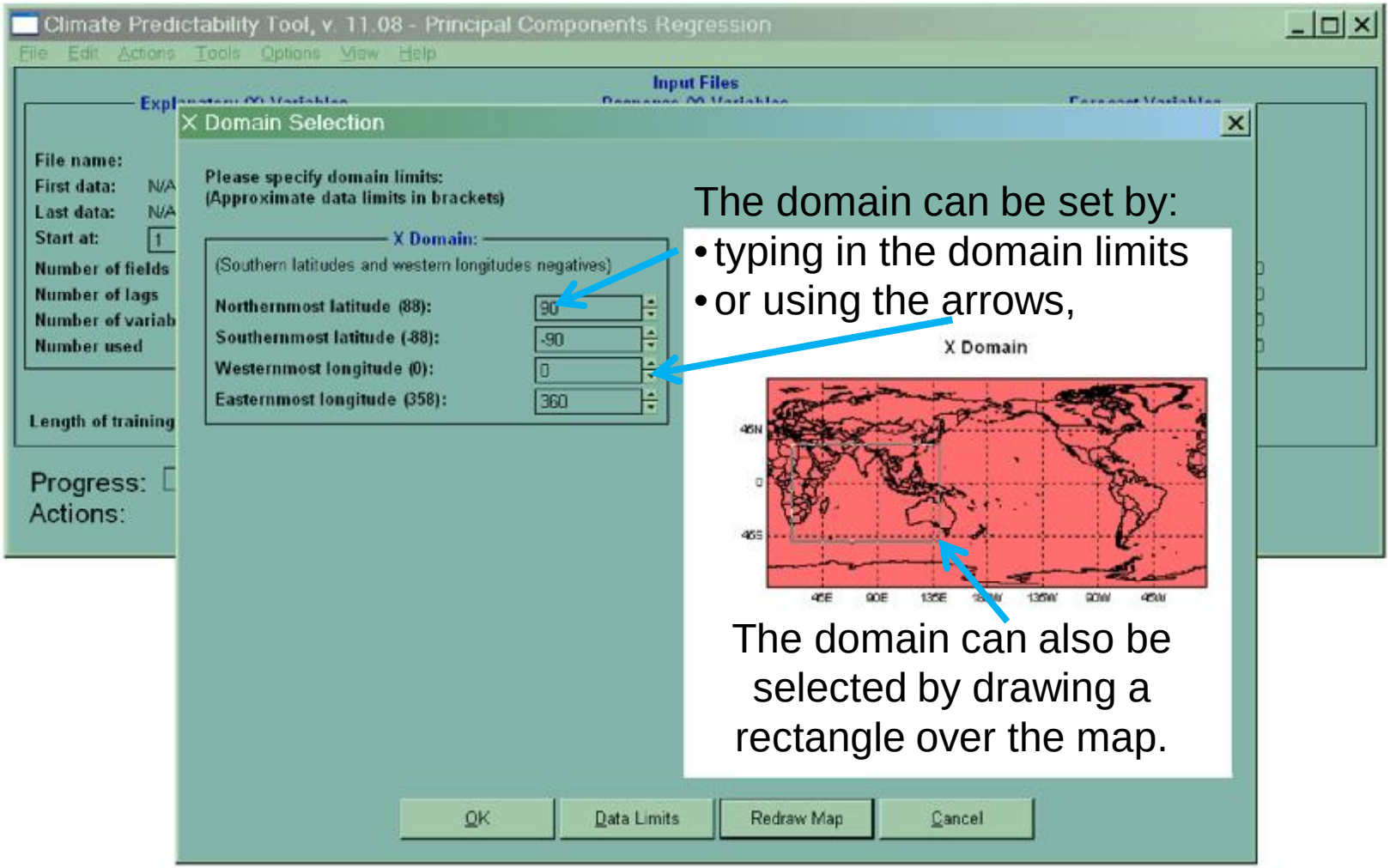
To select input files just click on browse.

SELECTING INPUT FILES



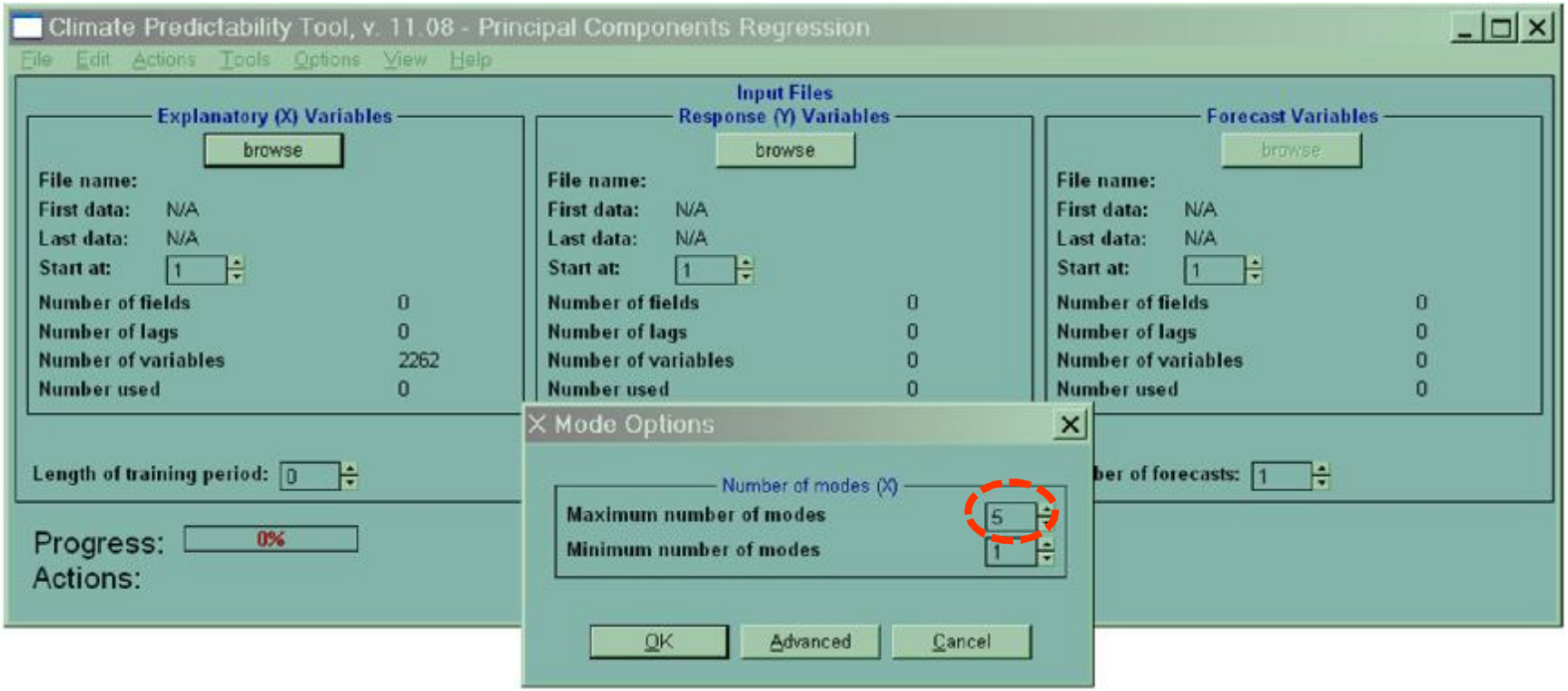
CPT opens a browser, which by default looks for data in:
C:\Documents and Settings\user\Application Data\CPT\Data\
or the directory specified during installation.
You can search for data from any other directory.

SELECTING INPUT FILES



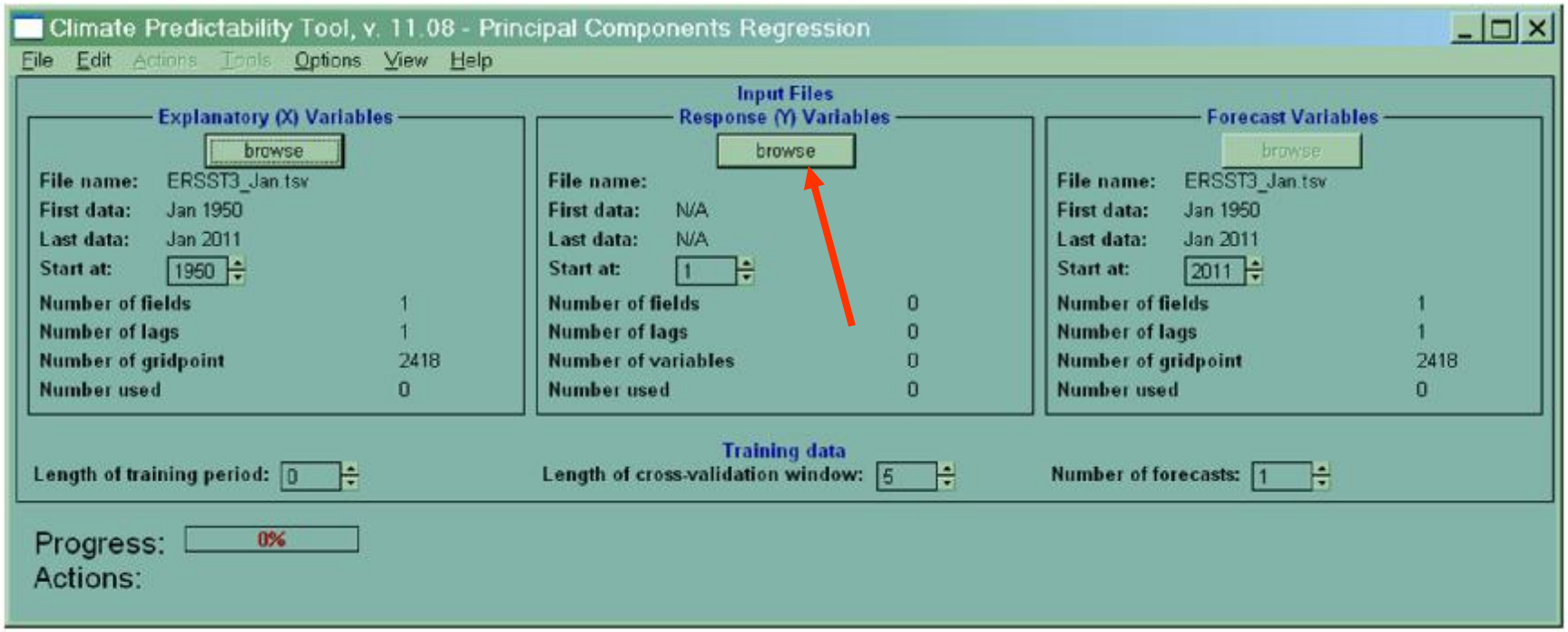
For gridded and station datasets, CPT lets you choose the spatial domain over which you want to perform your analysis. For the X file you should choose a domain from where the predictors are known to affect climate over the region to be predicted.

SETTING ANALYSIS OPTIONS



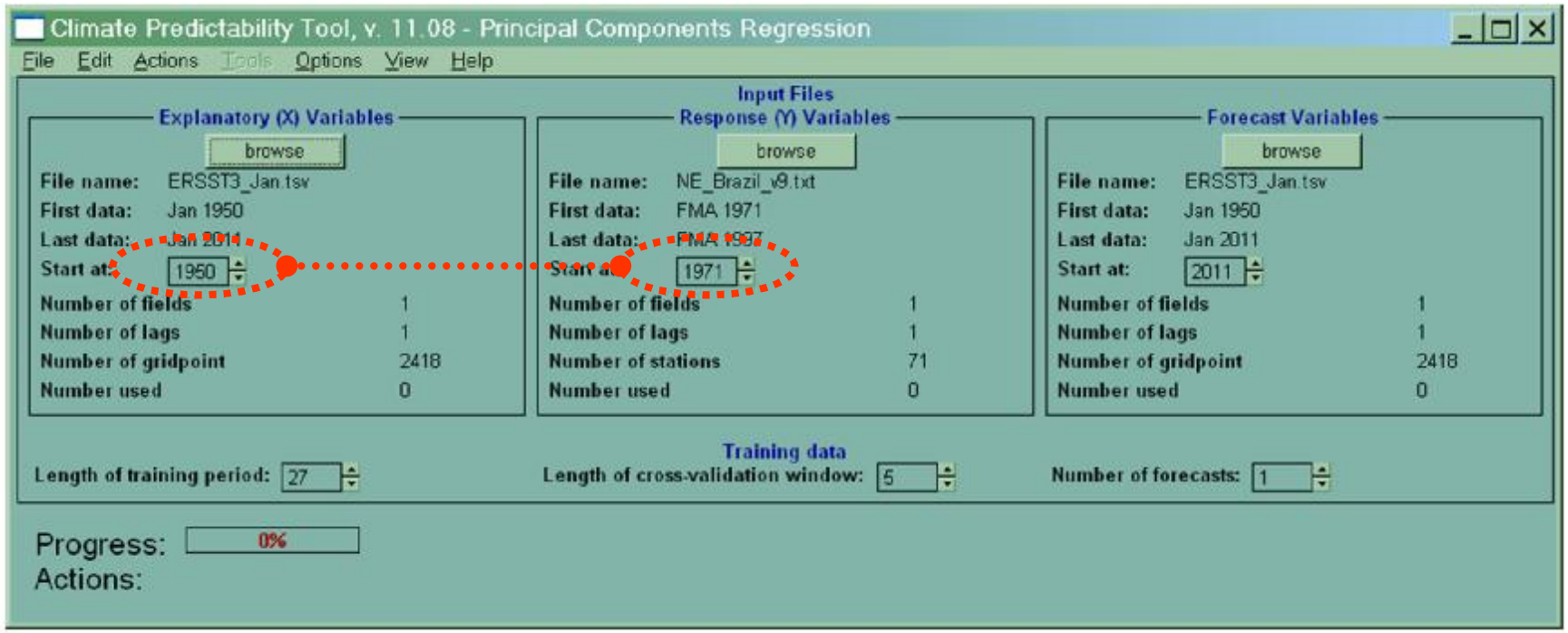
For PCR and CCA, you have to choose the number of EOFs for the predictor fields used to fit the model. If you set the minimum to be less than the maximum, CPT will find the optimum number of modes between the two numbers. If you set the minimum equal to the maximum, then CPT will use that number of modes.

SELECTING INPUT FILES



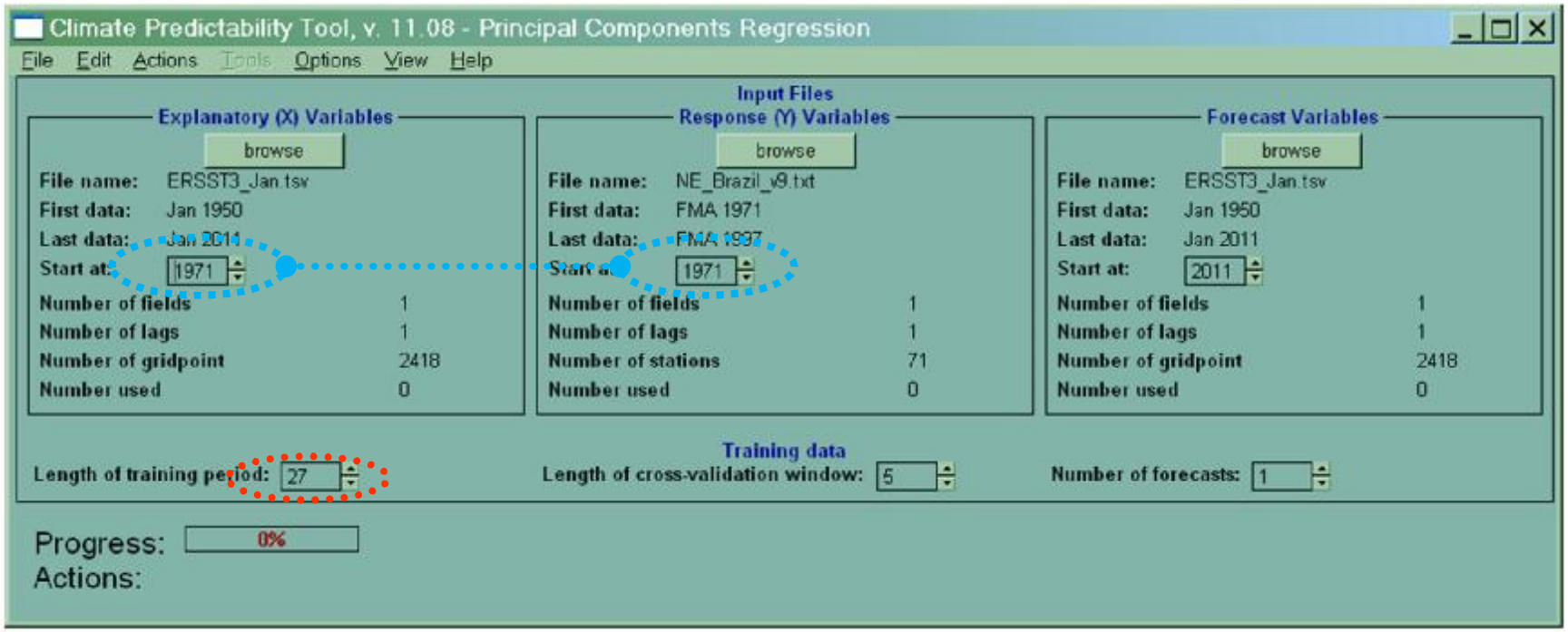
Proceed in the same way to select your file containing the Y variables (predictands).

SETTING THE TRAINING PERIOD



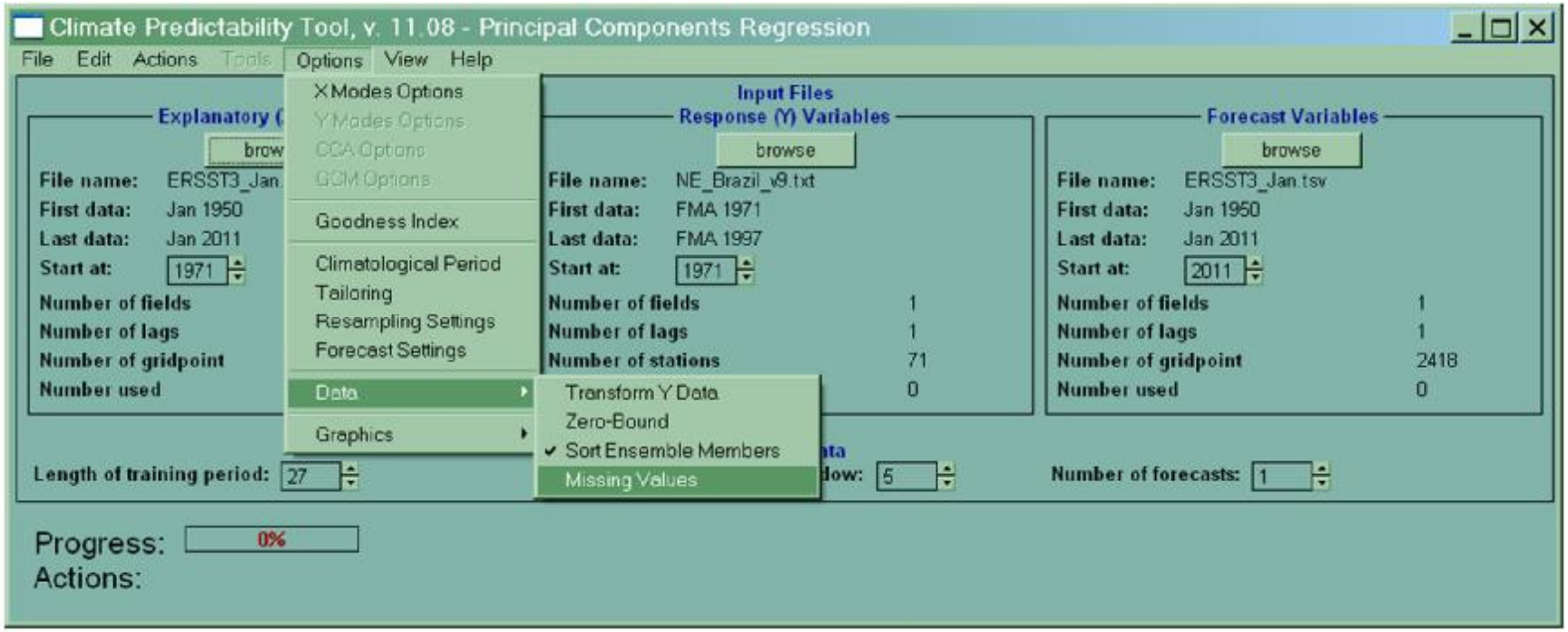
By default CPT usually starts the analysis from the first years in the X and Y files; note that these years could be different. You would normally set them equal to the latest of the two first years in the files. (In the example, the start date for the X file would normally be set to 1971.)
If you cross the calendar year while using December predictors, for example, the starting year for the X file will need to be one year earlier than for the Y file. (In this case, 1970.)
If you use a NDJ or DJF season, the year is for the first month.

SETTING THE TRAINING PERIOD



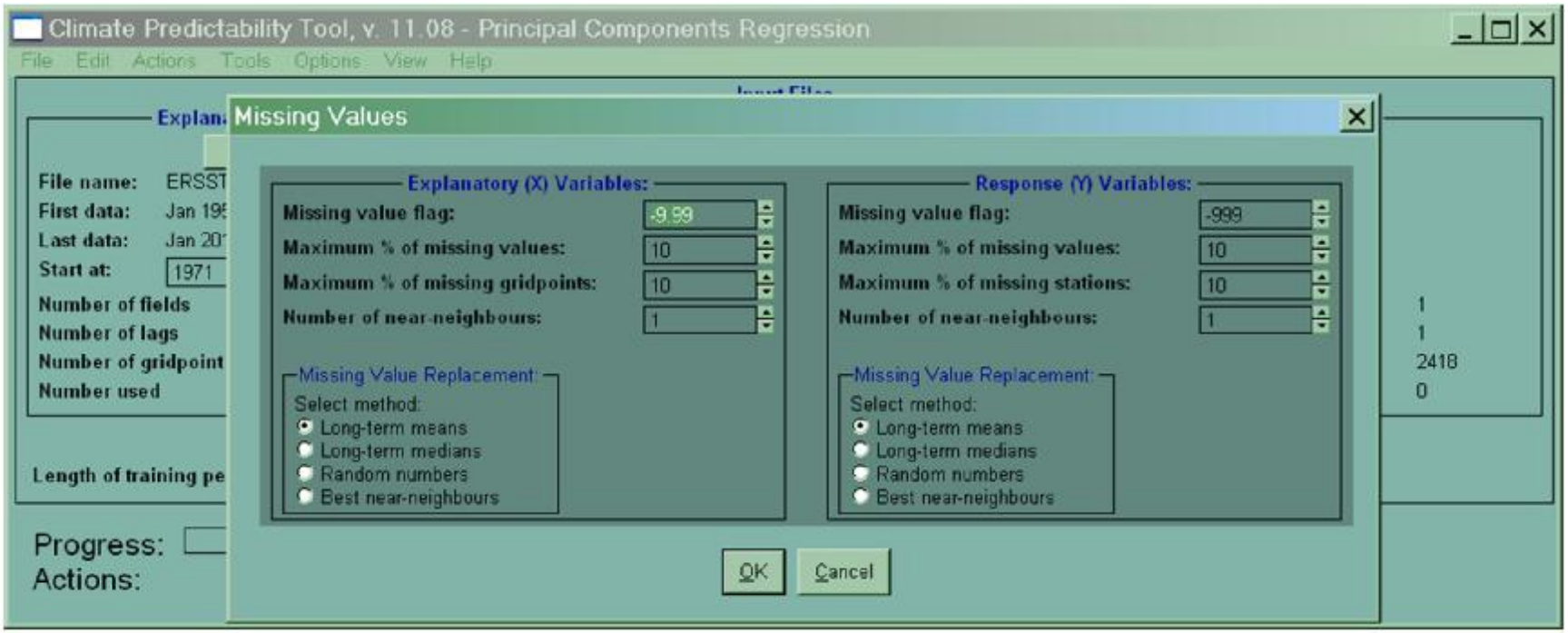
You have to specify the length of the training period.
By default, CPT will try to use as many years as are available.

MISSING VALUES



If you have missing values in your dataset, you need to specify what you want CPT to do with them.

MISSING VALUES



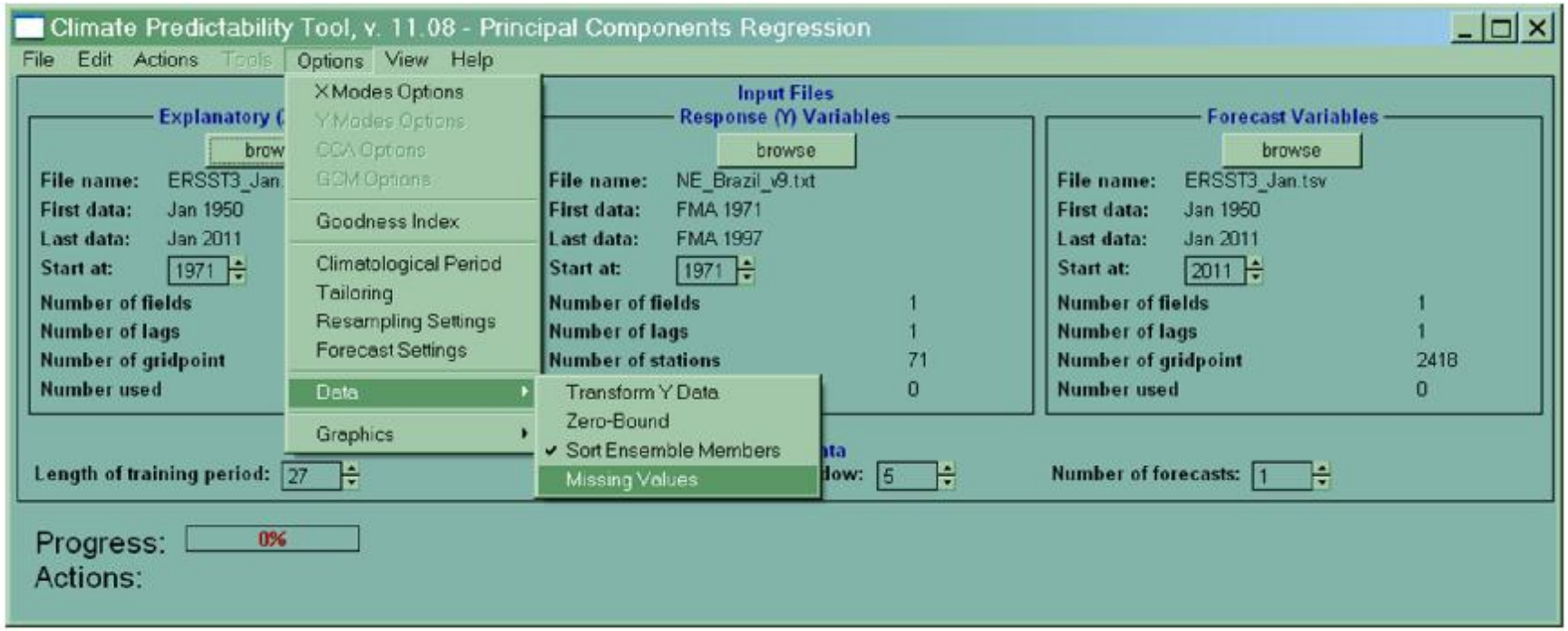
Next to the Missing value flag box, you need to specify the number in your dataset that represents a missing value.

You can choose the Maximum % of missing values. If a station has more than that percentage of missing values, CPT will not use that station in its model.

You can choose the Maximum % of missing stations. If a year has more than that percentage of missing values, CPT will not use that year in its model.

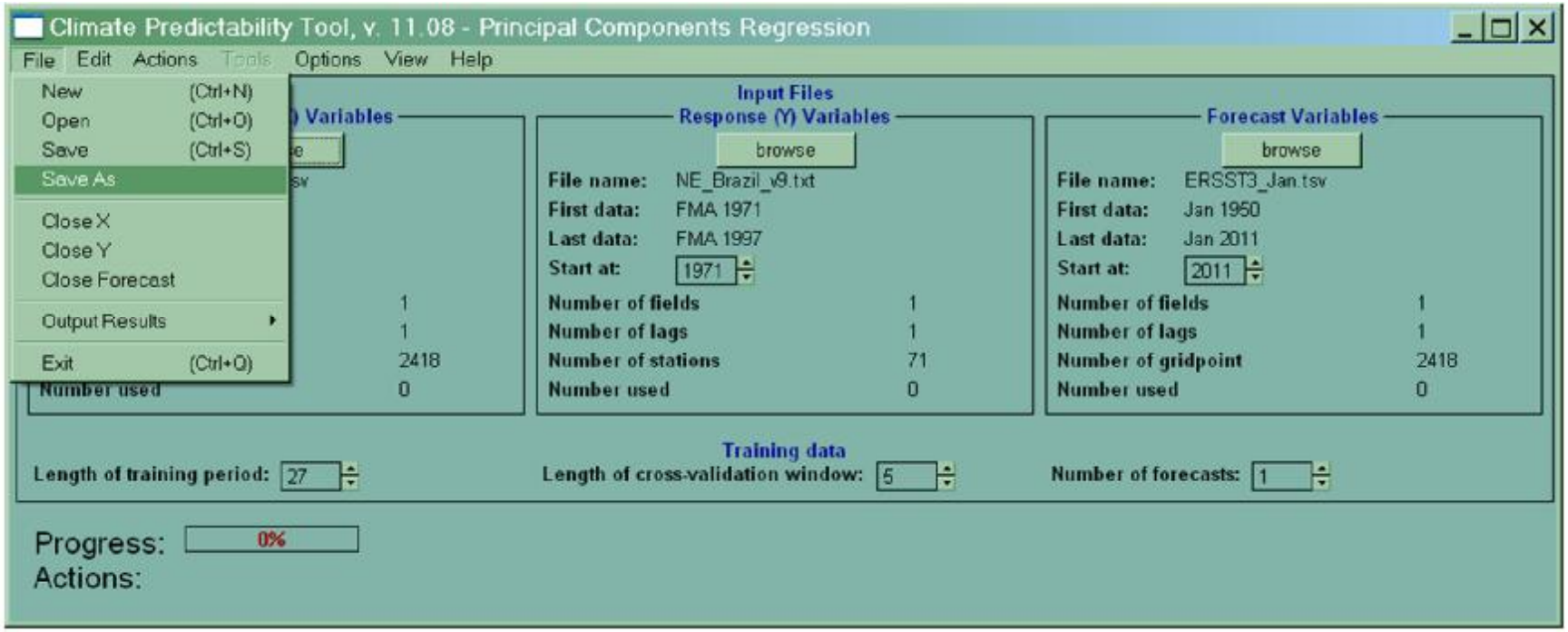
You can also choose which method you want CPT to use to replace the values. If you choose Best nearest neighbours then CPT will use the Number of near-neighbours that you specify.

RAINFALL



If you are predicting rainfall, you may want to switch on the Zero-Bound, which will force CPT never to predict negative values. If the rainfall data are positively skewed (have occasional very large values), you may also want to switch on Transform Y Data, which will help to prevent lowest forecast probabilities on "normal".

SAVING PROGRAM SETTINGS



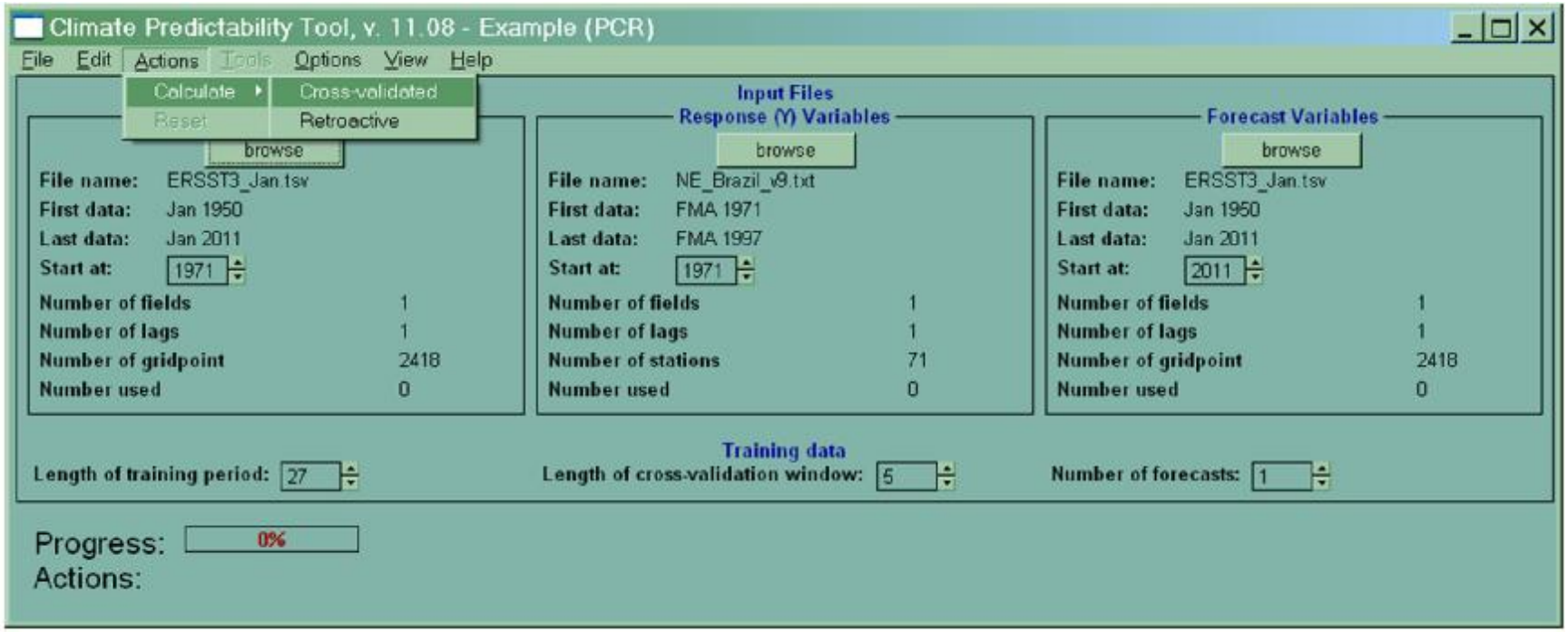
Once you have selected the input files and your settings it is a good idea to save these settings in a project file to recall them later:

File ~ Save

By default, CPT saves all the project files in the subdirectory

C:\Documents and settings\user\Application Data\CPT\Projects\

RUNNING CPT



Then you can run the analysis:
Actions ~ Calculate ~ Cross-validated

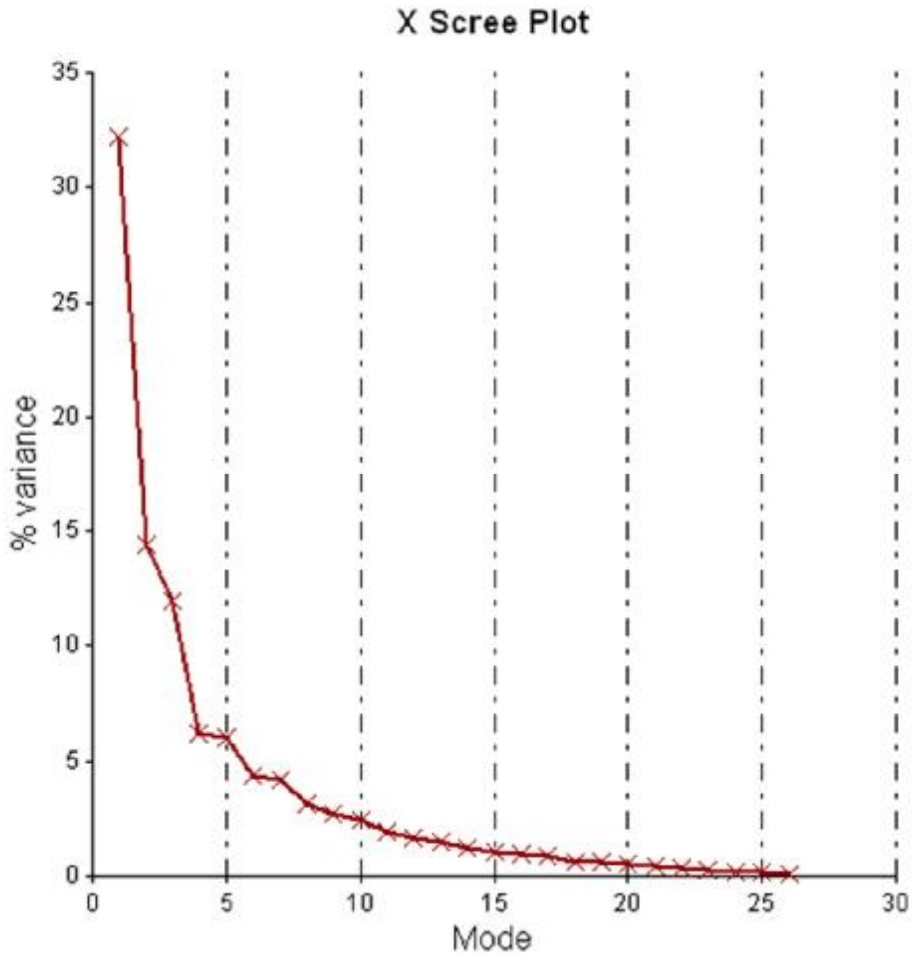
DATA ANALYSIS

CURRENT		OPTIMUM	
Number of Modes	Goodness Index	Number of Modes	Goodness Index
1	-0.105	1	-0.105
2	0.031	2	0.031
3	0.214	3	0.214
4	0.144	3	0.214
5	0.112	3	0.214

Optimizing the number of EOF modes:

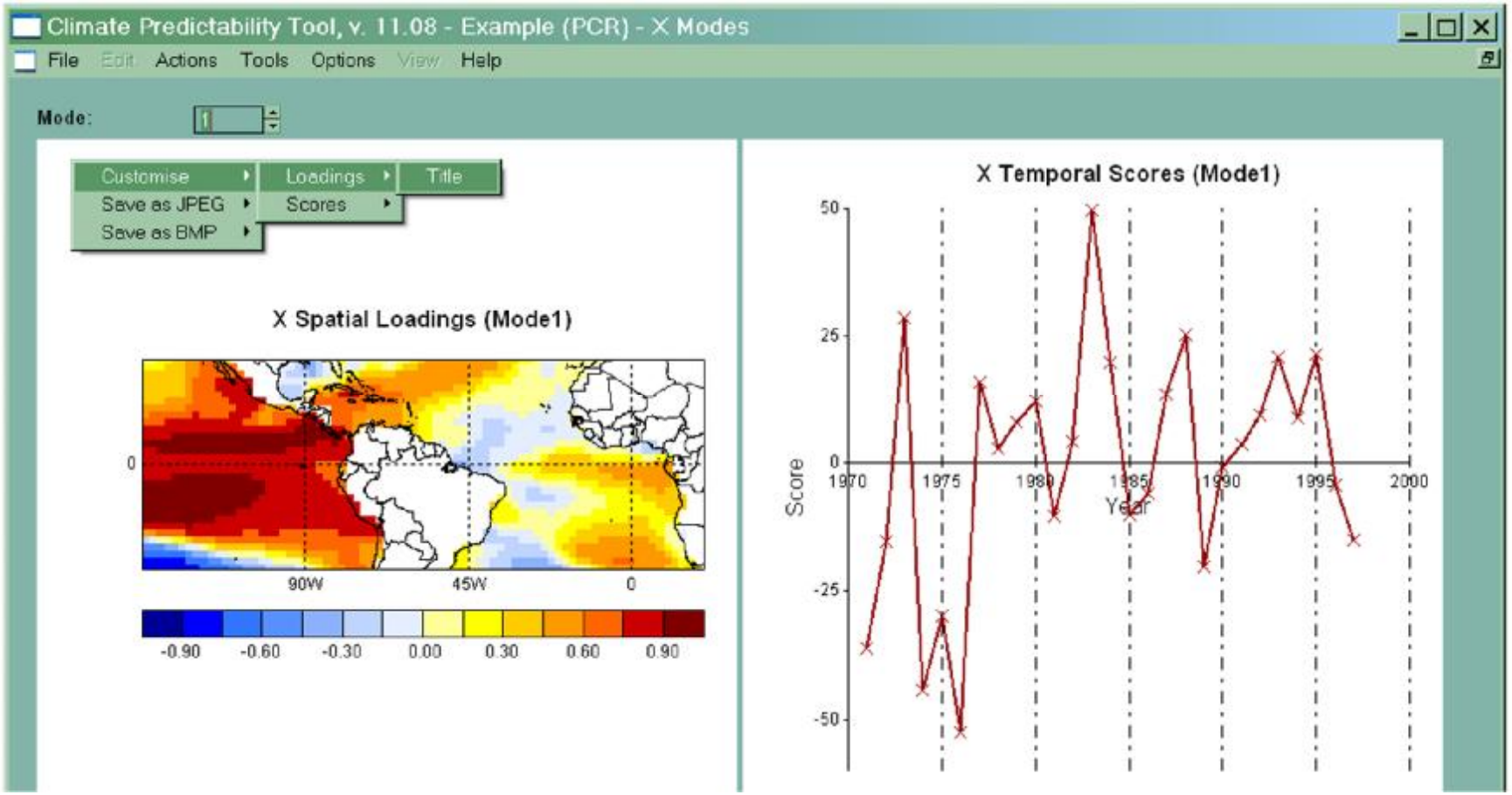
1. CPT uses EOF #1 to make cross-validated forecasts then calculates a "goodness index" summarizing how good all the forecasts are (the closer to 1.0 the better). Then CPT uses EOF #1 and #2 to remake cross-validated forecasts and calculates a new goodness index for these, and so on until all five EOFs have been used.
2. At each step CPT compares the goodness indices and retains under the column "OPTIMUM" the highest goodness index and the corresponding number of EOFs (in the example above, 3).
3. CPT uses this number of EOFs (i.e., 3) to build the model.

RESULTS – GRAPHICS



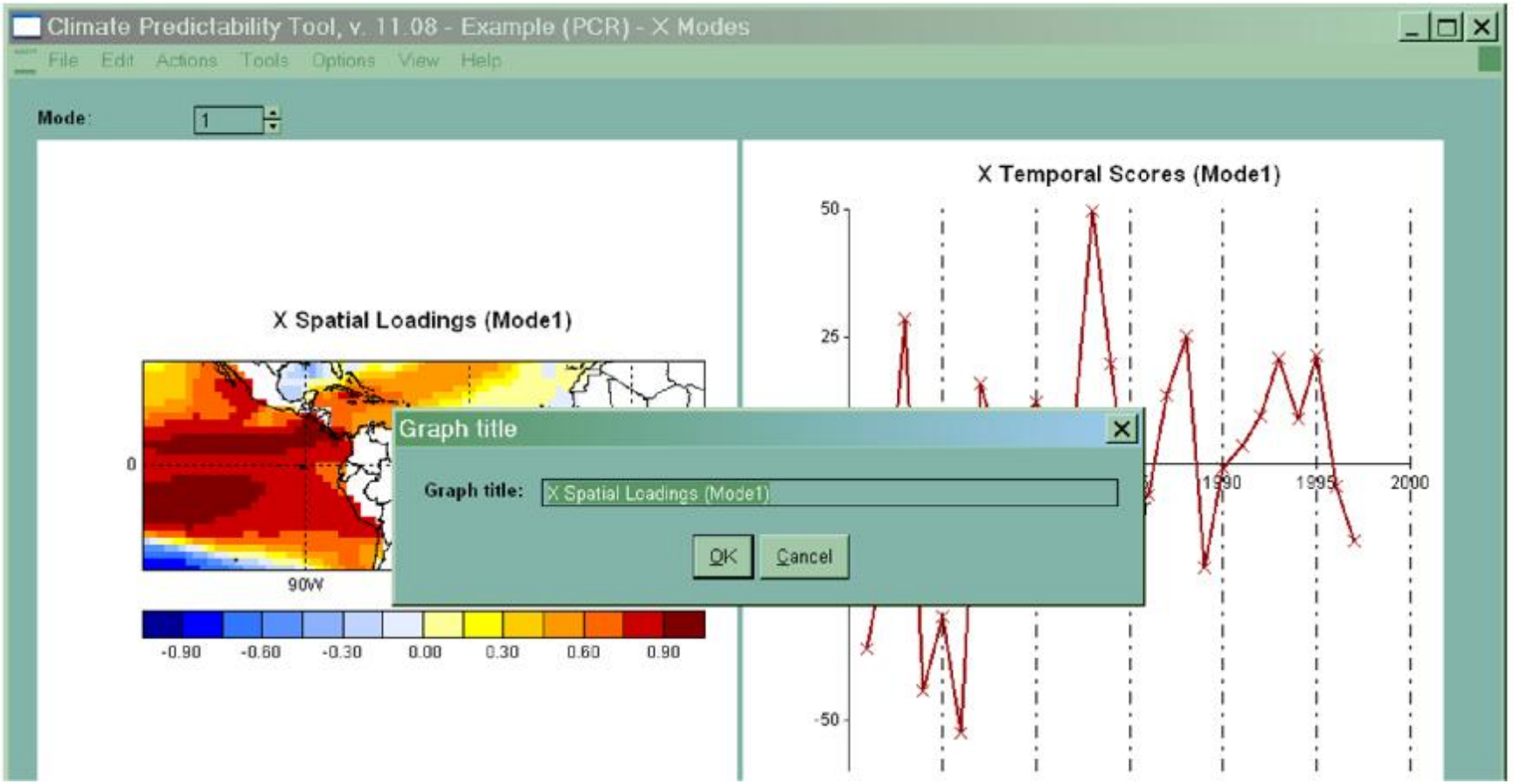
The menu Tools ~ Modes ~ Scree plots displays the percentage of variance associated with each EOF plotted.

RESULTS – GRAPHICS



1. The menu Tools ~ Modes ~ X EOF loadings and scores displays the loading pattern of each EOF and the temporal series.
2. CPT allows you to customize and save each graphic by:
   right-clicking on the mouse
   selecting the graphic to customize / save

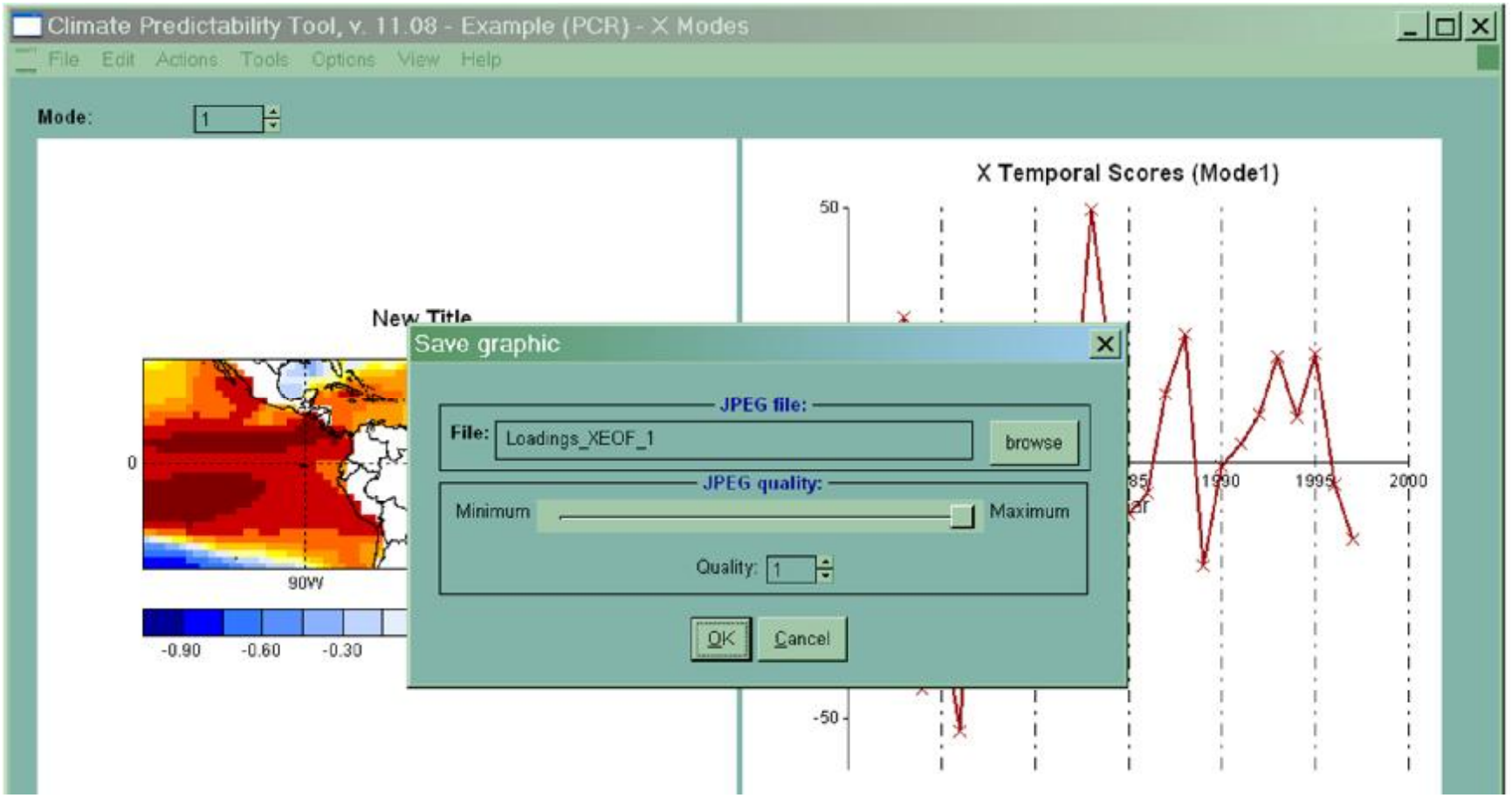
CHANGING THE TITLE



To change the title of the map

1. right-click the mouse
2. go to EOF Loadings
3. click on Title

SAVING GRAPHICS

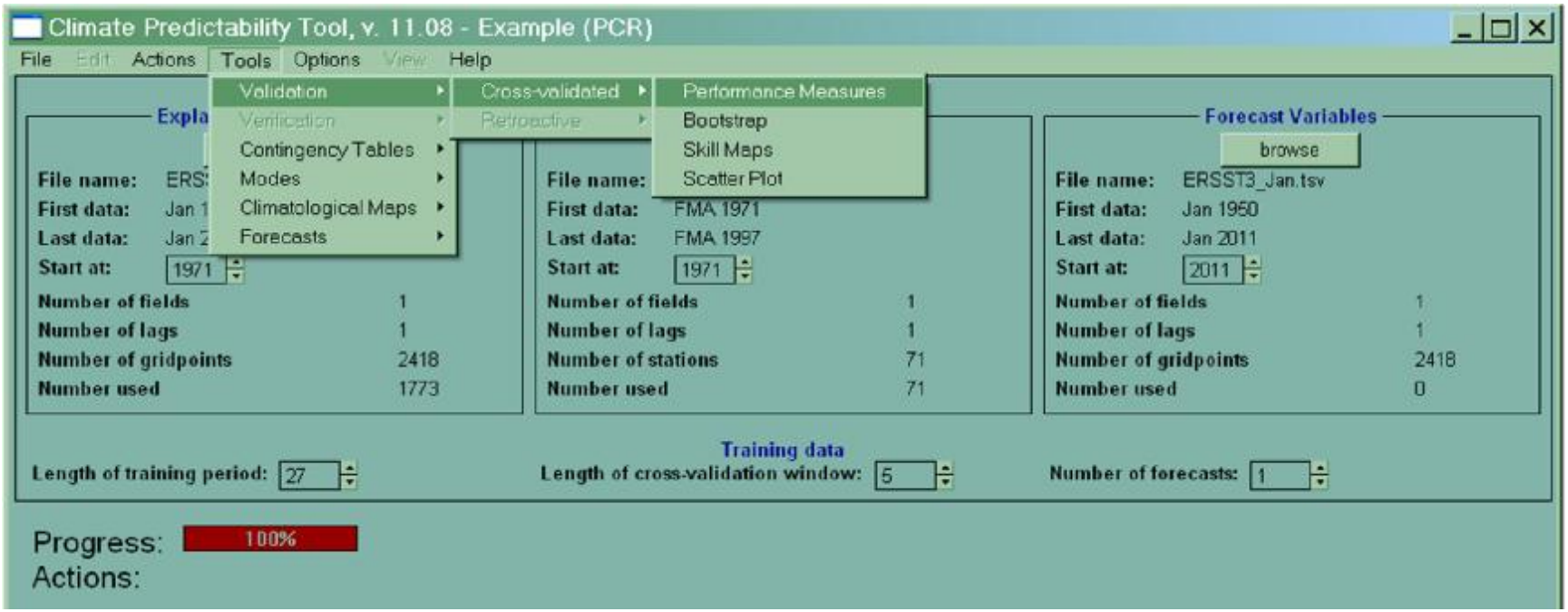


You can choose the name of the graphic output file by clicking on browse. You can adjust the quality of the JPEG graphic as well.

All the output files are saved by default under:

C:\Documents and settings\user\Application Data\CPT\Output\

RESULTS



To see the results go to the menu “Tools”:

Validation : shows skill, hindcasts and observed series

Verification: shows probabilistic skill information for retroactive forecasts

Contingency Tables : shows contingency tables

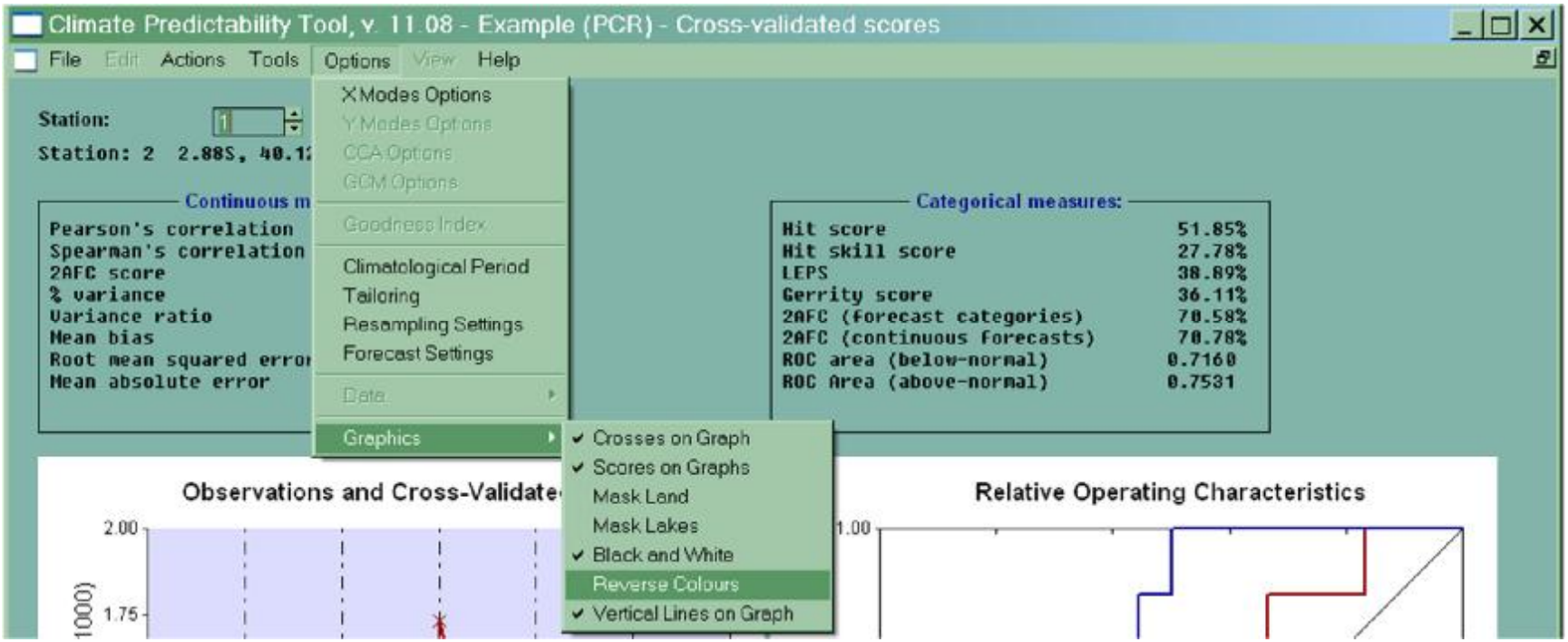
Modes : shows EOF time series, loading patterns and scree plot

Climatological Maps: shows maps of terciles and averages

RESULTS

To see the observations and cross-validated forecasts at each station go to:
Tools ~ Validation ~ Cross-Validated ~ Performance Measures

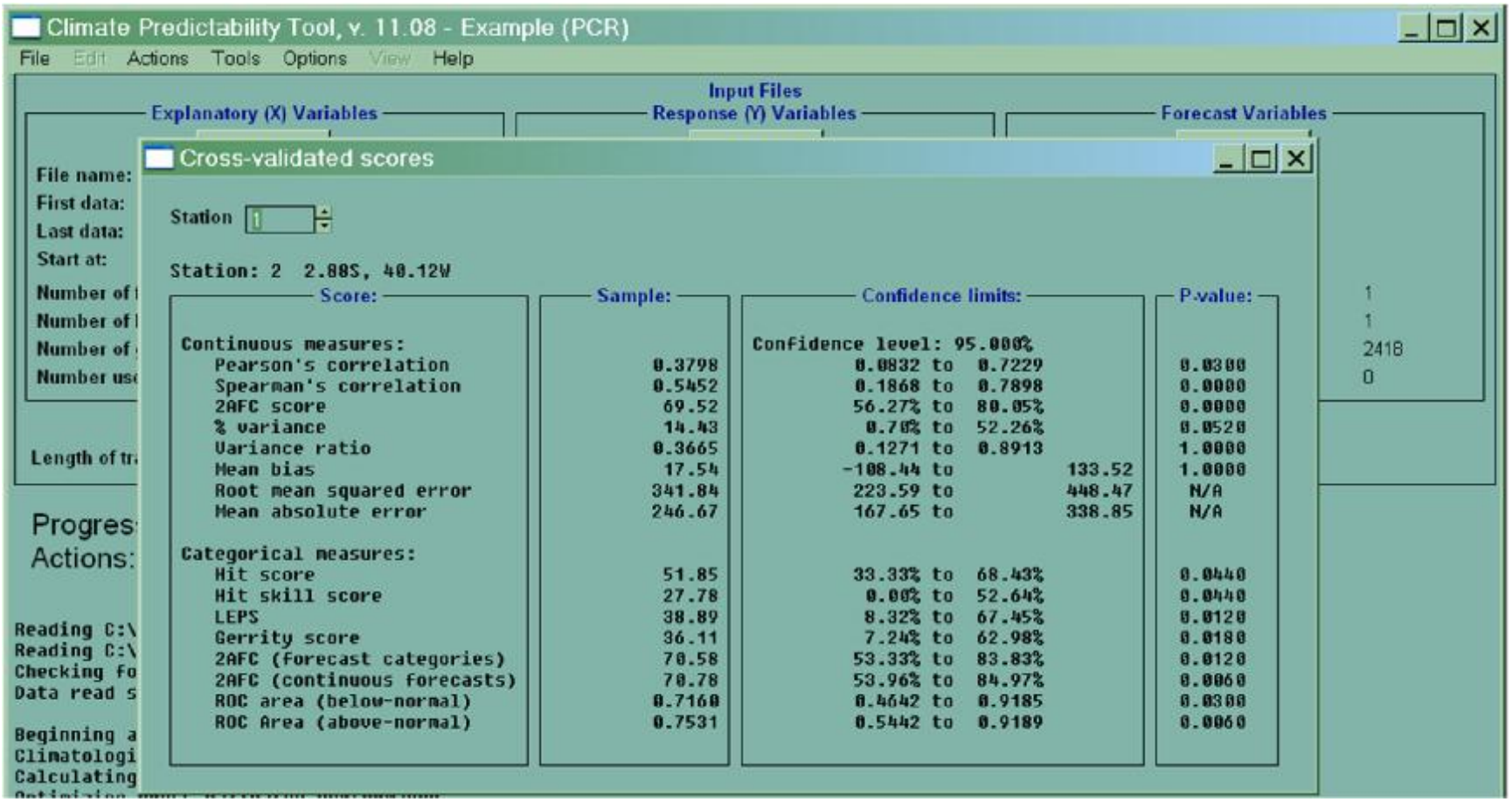
REVERSING THE COLORS



Options ~ Graphics ~ Reverse Colors

If you are forecasting temperature instead of precipitation, then it would be more intuitive to have red (hot) for above and blue (cold) for below, so you might want to invert the default colors. You might also want black and white images if they are to be included in a report or publication.

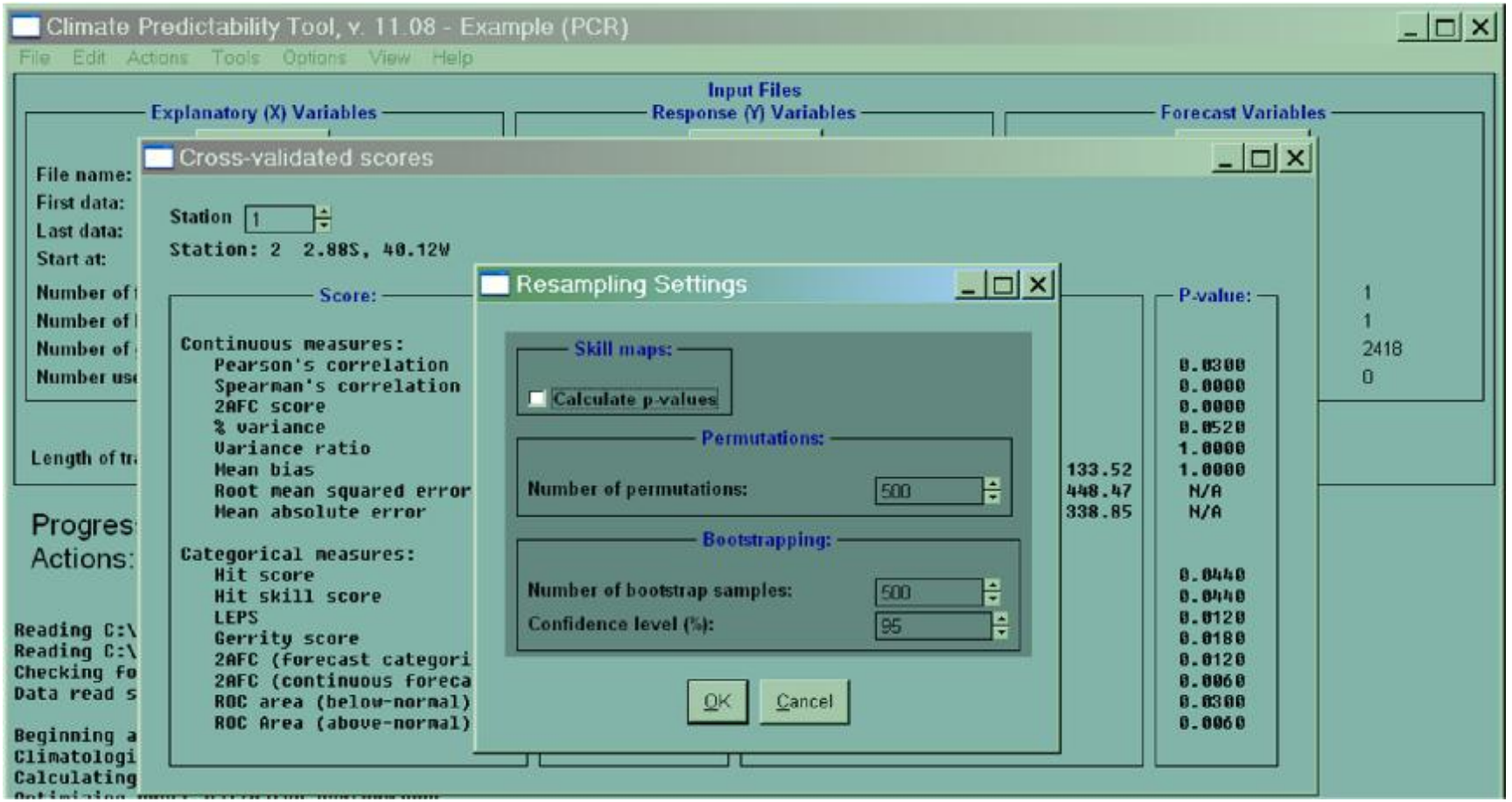
INDICATIONS OF SAMPLING ERRORS



For indications of sampling errors in the performance measures go to:

Tools ~ Validation ~ Cross-Validated ~ Bootstrap

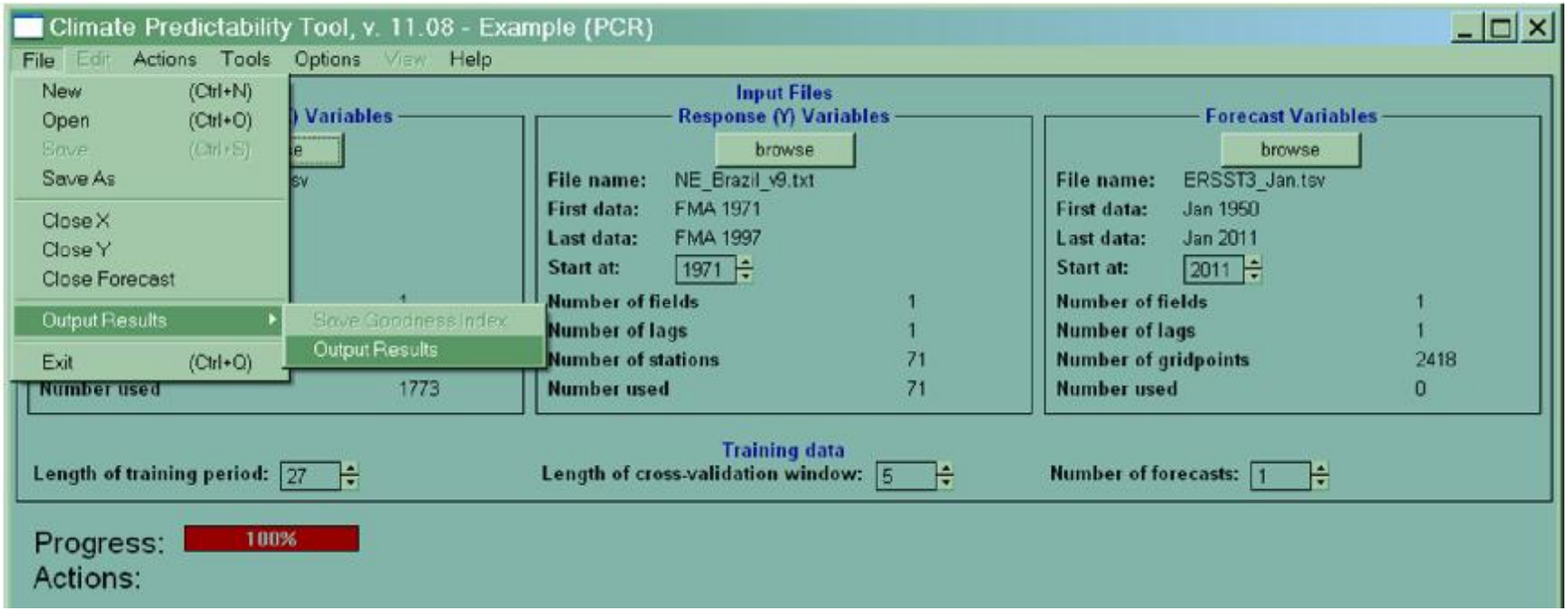
ADJUSTING THE BOOTSTRAP SETTINGS



Options ~ Resampling Settings

CPT allows you to adjust the bootstrap settings.

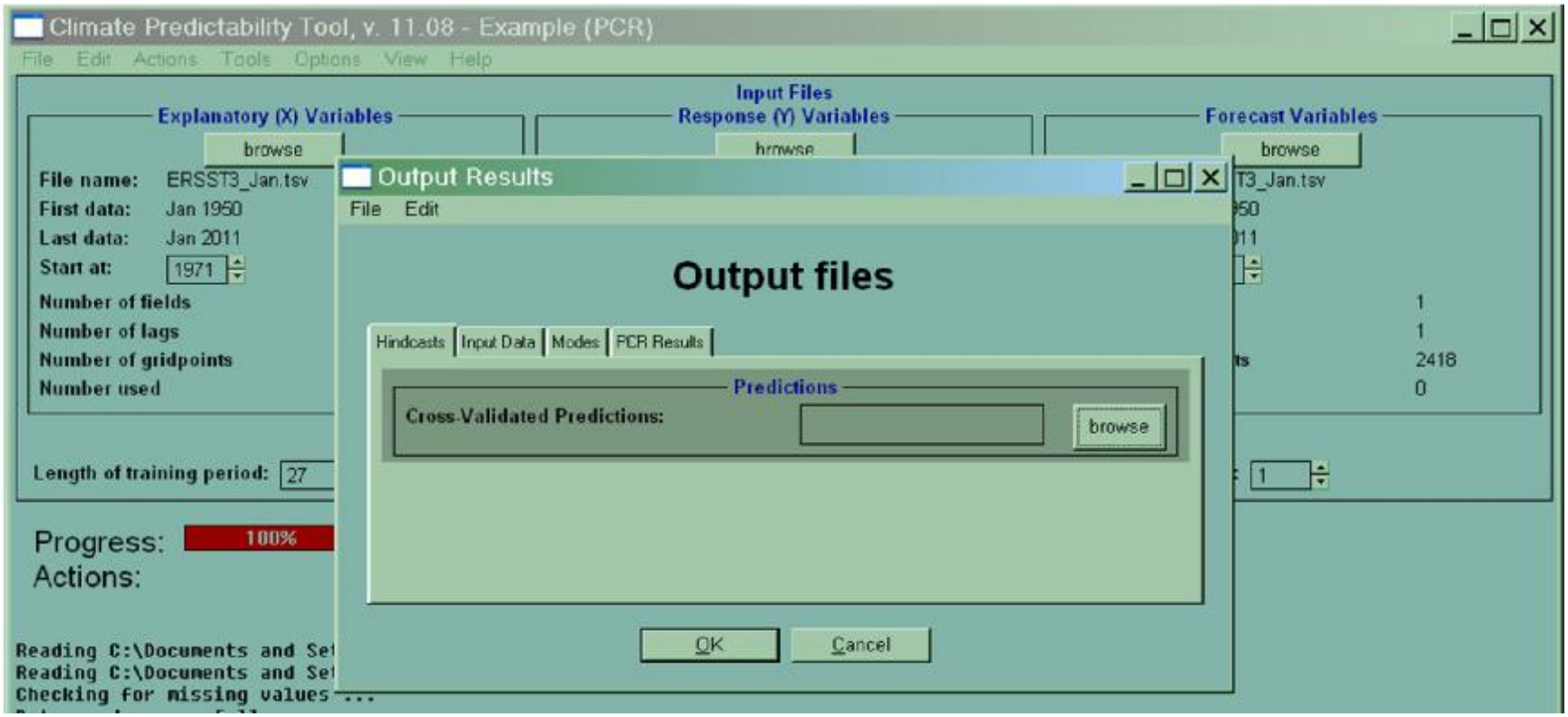
RESULTS – DATA FILES



The menu File ~ Output Results ~ Output Results allows you to save output data:

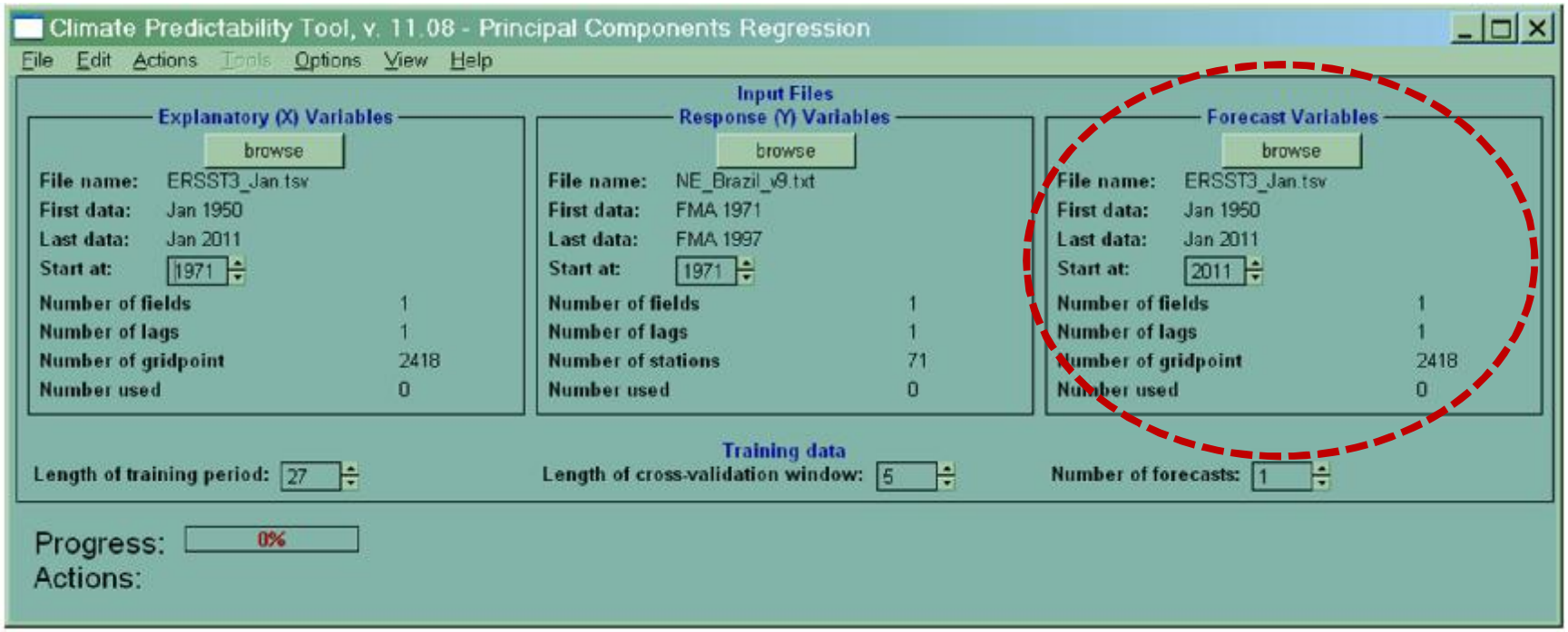
1. Cross-validated forecasts
2. The input data (with the missing values filled)
3. EOFs: time series, loading patterns, variance
4. The parameters (coefficients) of the model (example: Y=**a**x+**b**)

SAVING OUTPUT FILES



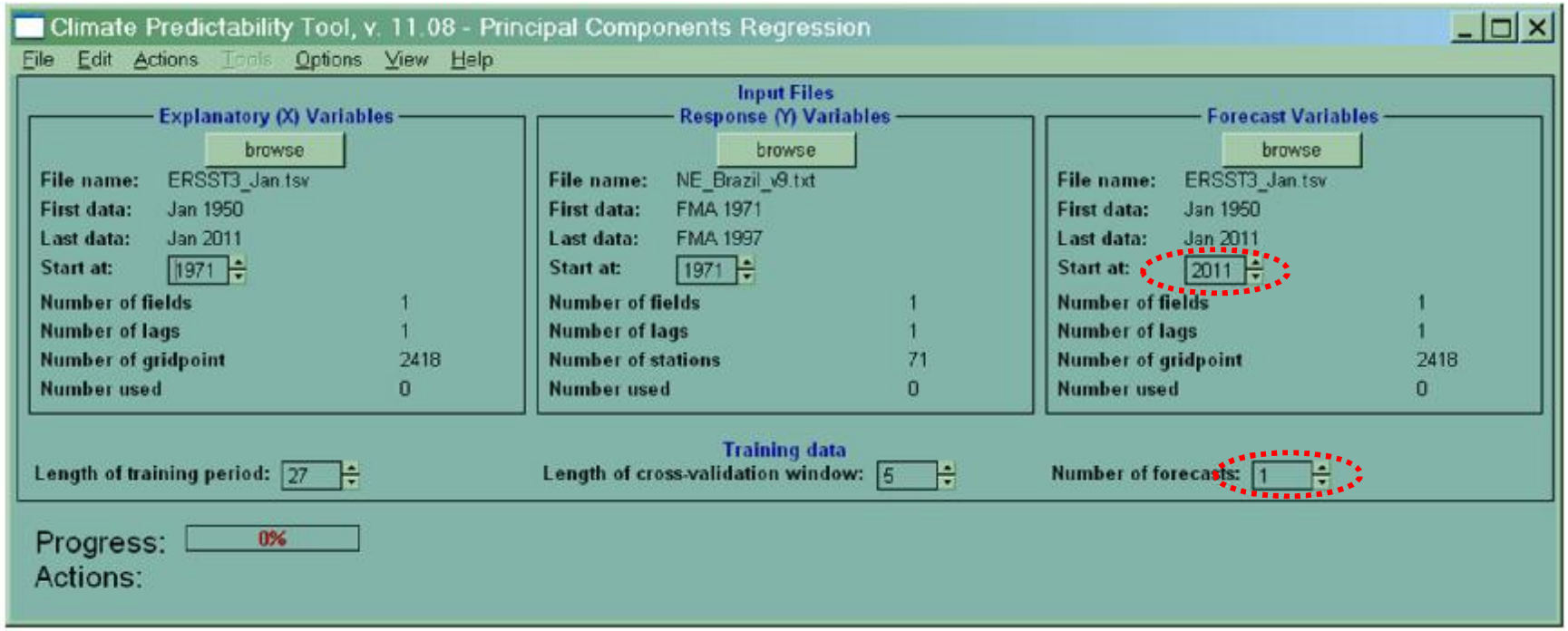
In order to save the outputs you have to specify a file name by clicking on browse. By default CPT saves the output files under:
C:\Documents and settings\user\Application Data\CPT\Output\

FORECAST



Once your model is built, you can make a forecast using a forecast file with new records of the X variables stored in a "forecast file". By default CPT selects the same input predictor file as the X file. You can change it by clicking browse.

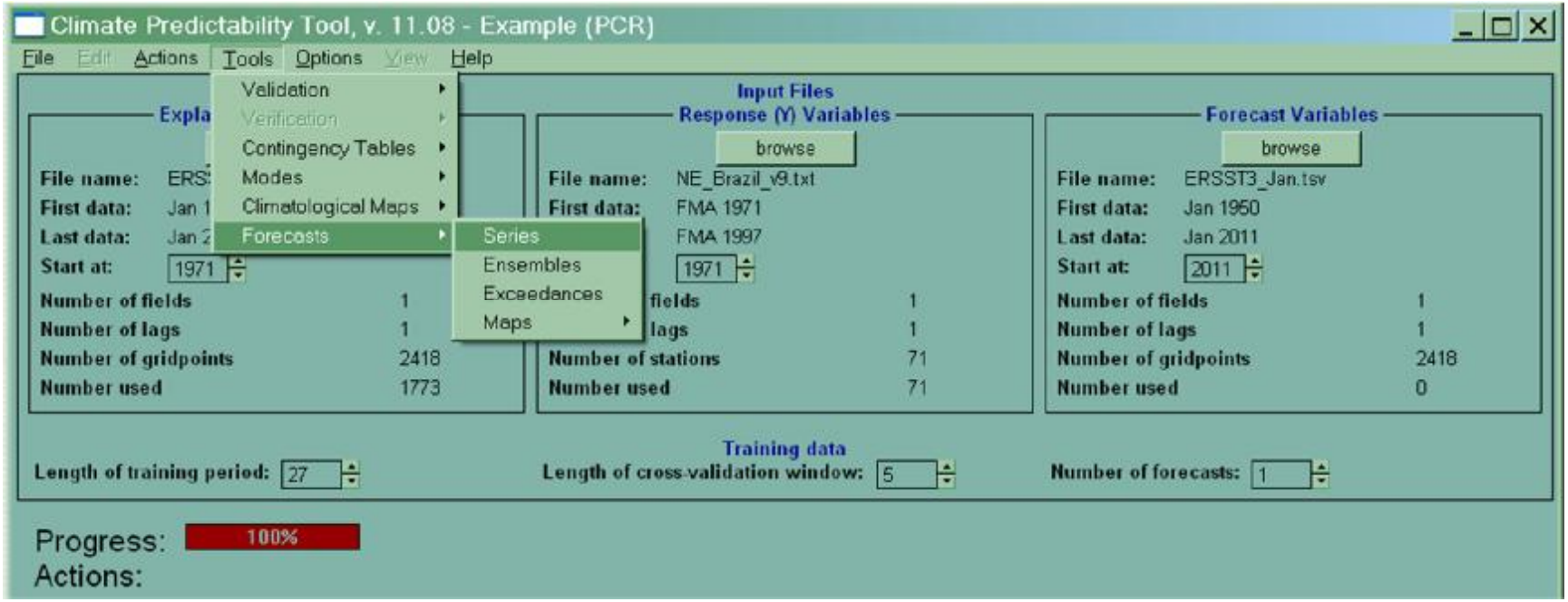
FORECAST



You then select:

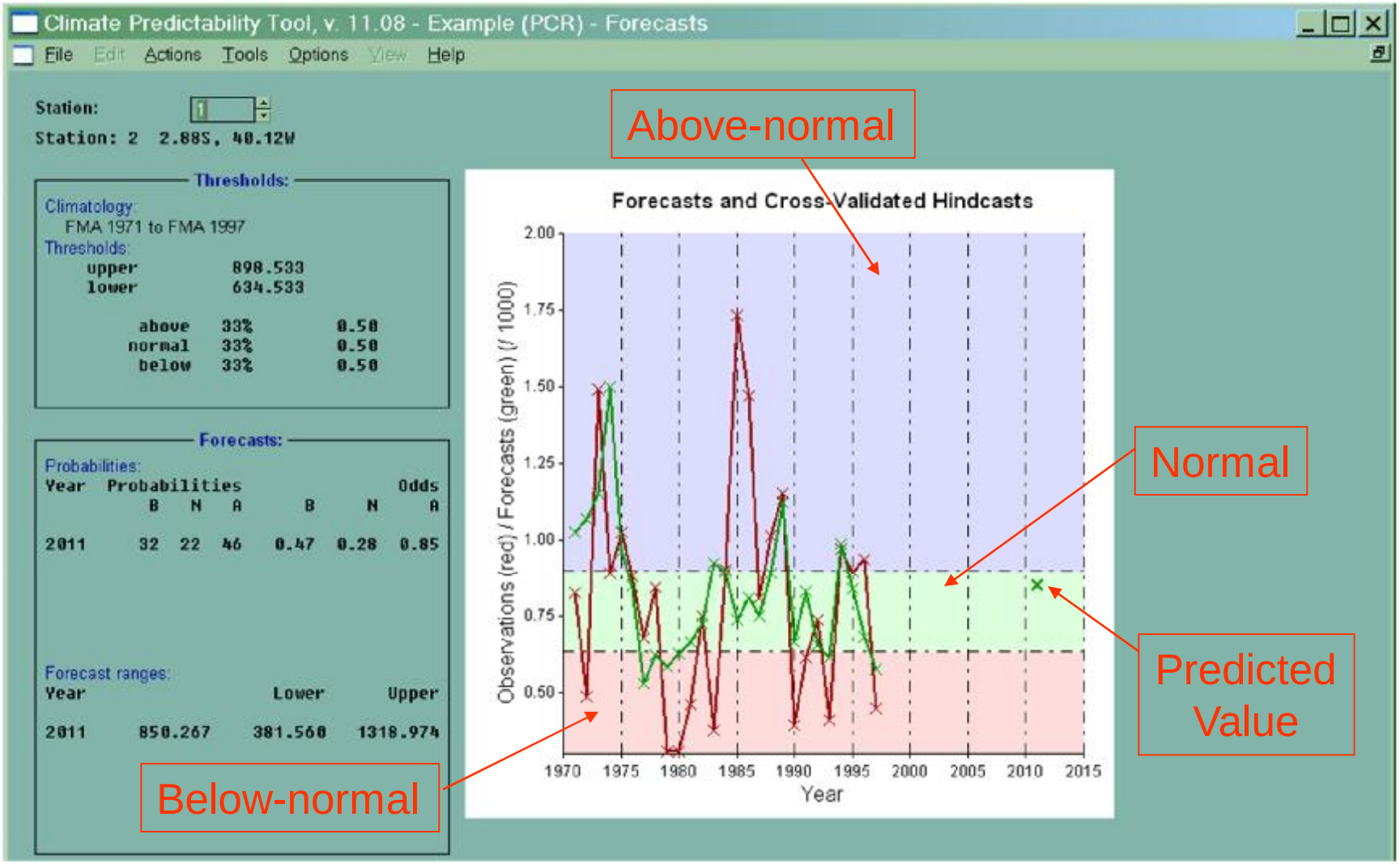
1. the starting year of the forecasts (the year is for the predictors not the predictand – for example if you are forecasting JFM 2012 from December 2011 SSTs, the year should be 2011.
2. the number of years to forecast

FORECAST



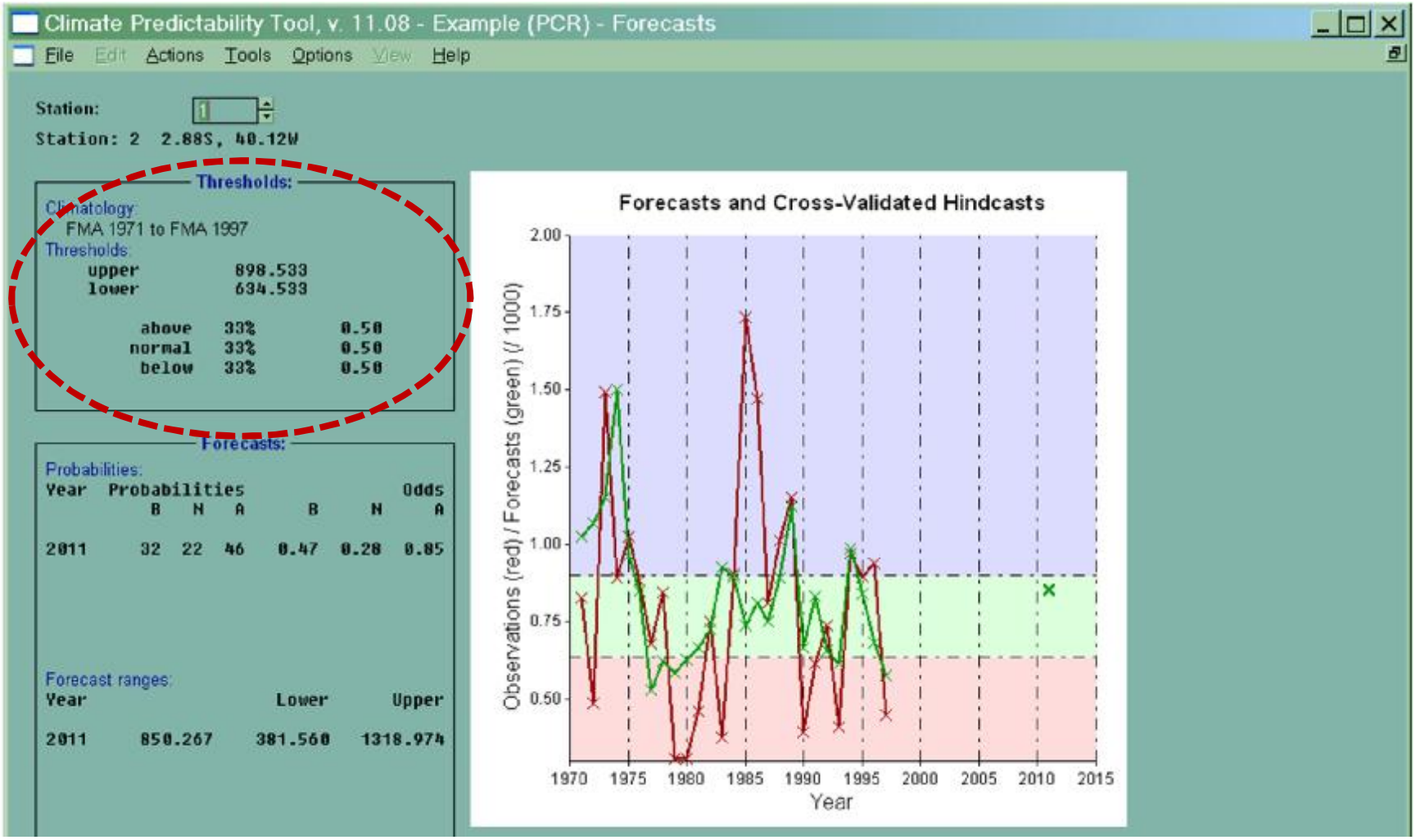
Once the file is selected and the years to forecast are chosen go to the menu Tools ~ Forecast ~ Series or Maps.

FORECAST



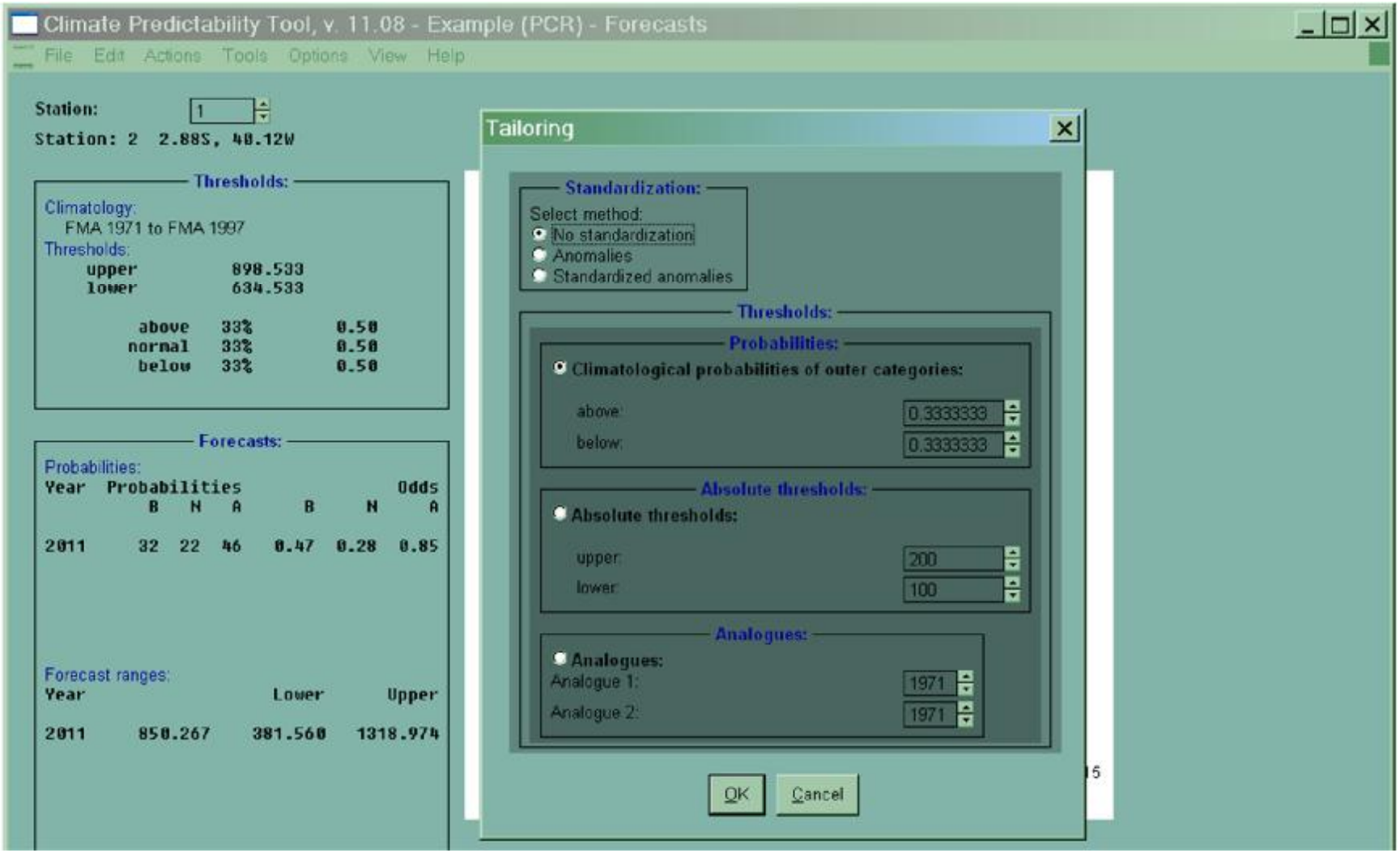
The option Series shows a graph of the cross-validated forecasts (green line) and the prediction (cross) for the current station, as well as detailed information about the forecast.

FORECAST



The “Thresholds” box indicates the definitions of below- and above-normal (less than the lower threshold and more than the upper threshold, respectively.) The climatological probabilities and odds of each of the three categories are shown.

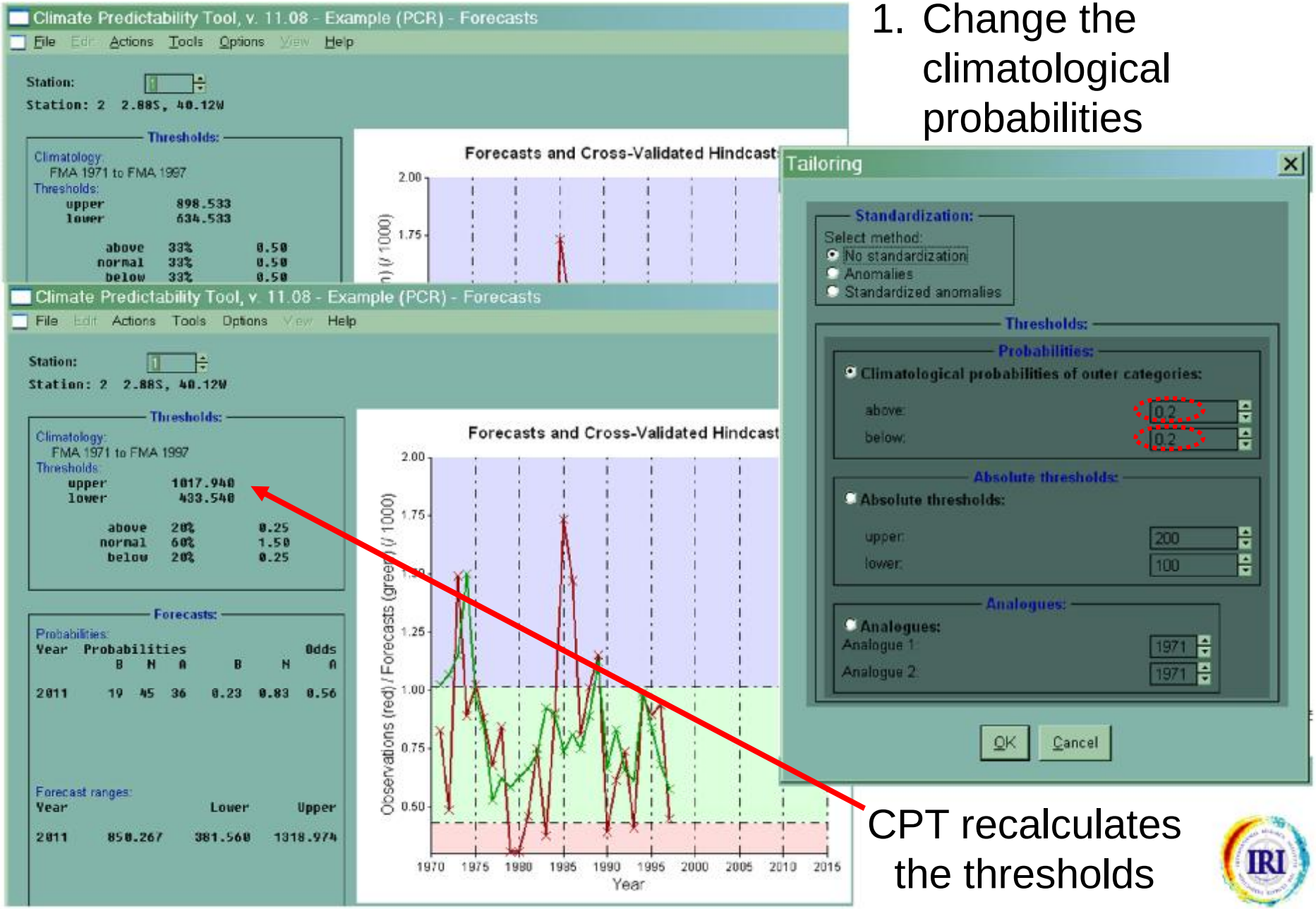
CHANGING CATEGORY DEFINITIONS



There are three ways to change how the categories are defined.

Options ~ Tailoring

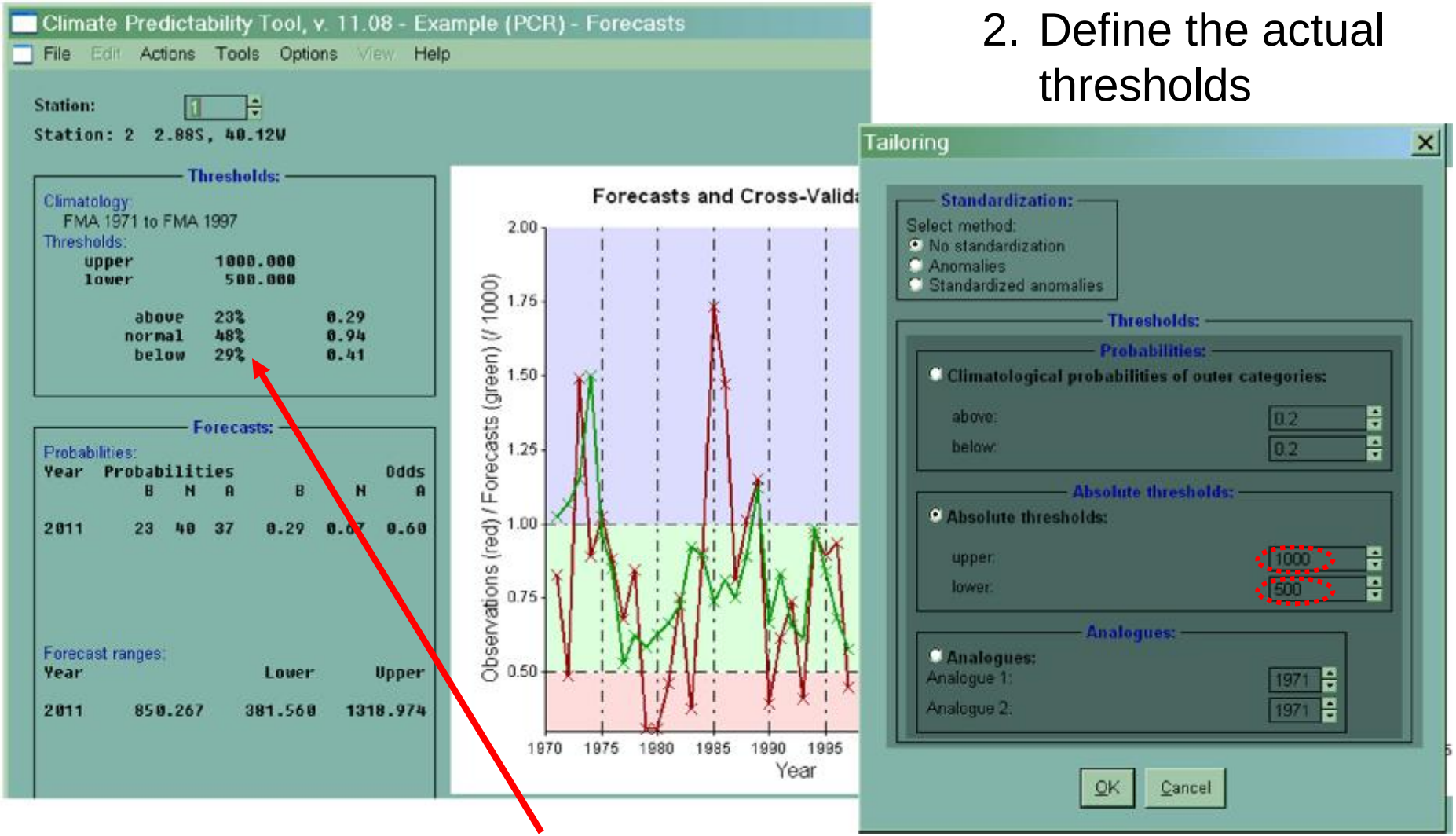
CHANGING CATEGORY DEFINITIONS



1. Change the climatological probabilities

CPT recalculates the thresholds

CHANGING CATEGORY DEFINITIONS

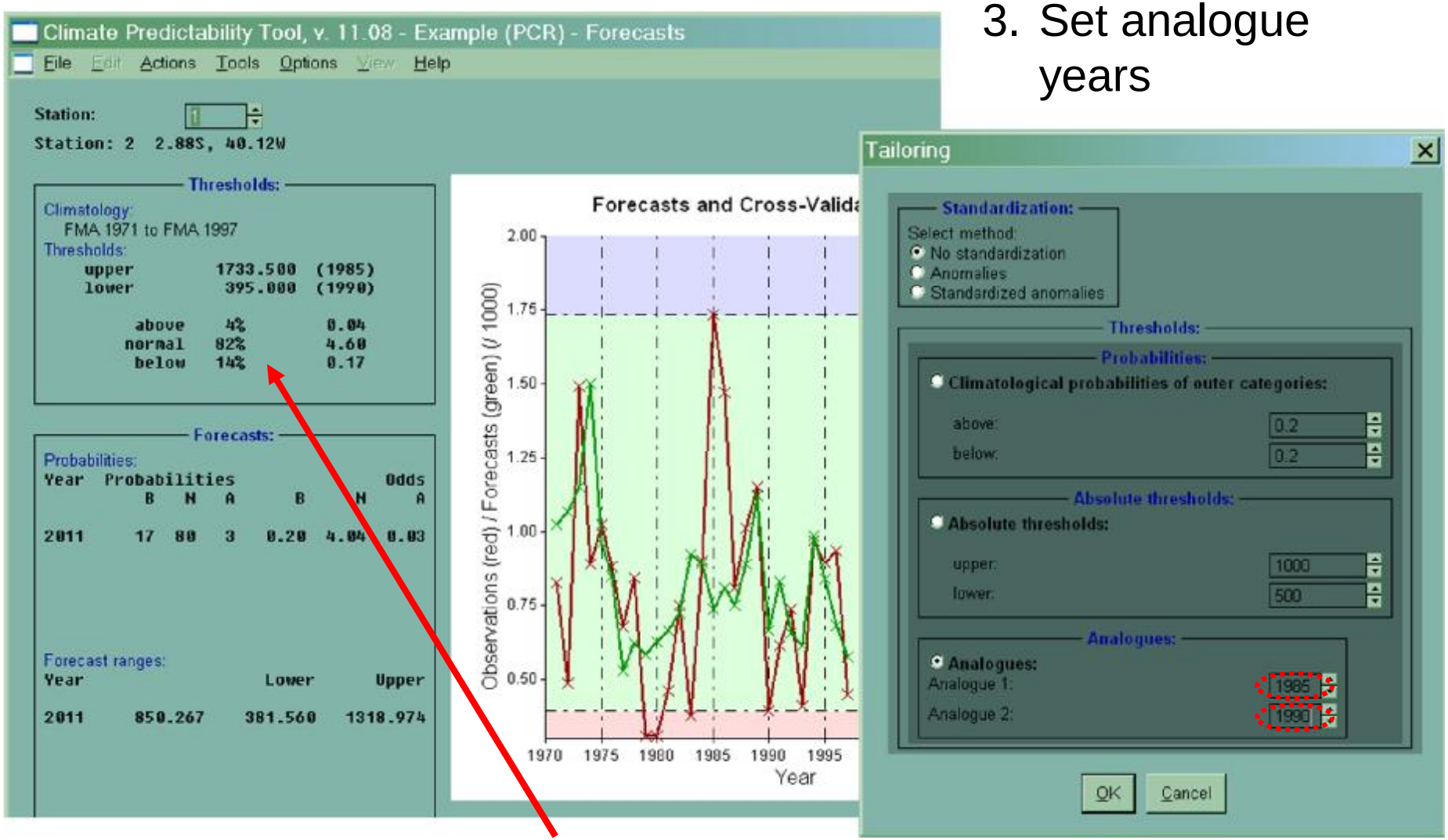


2. Define the actual thresholds

CPT recalculates the climatological probabilities

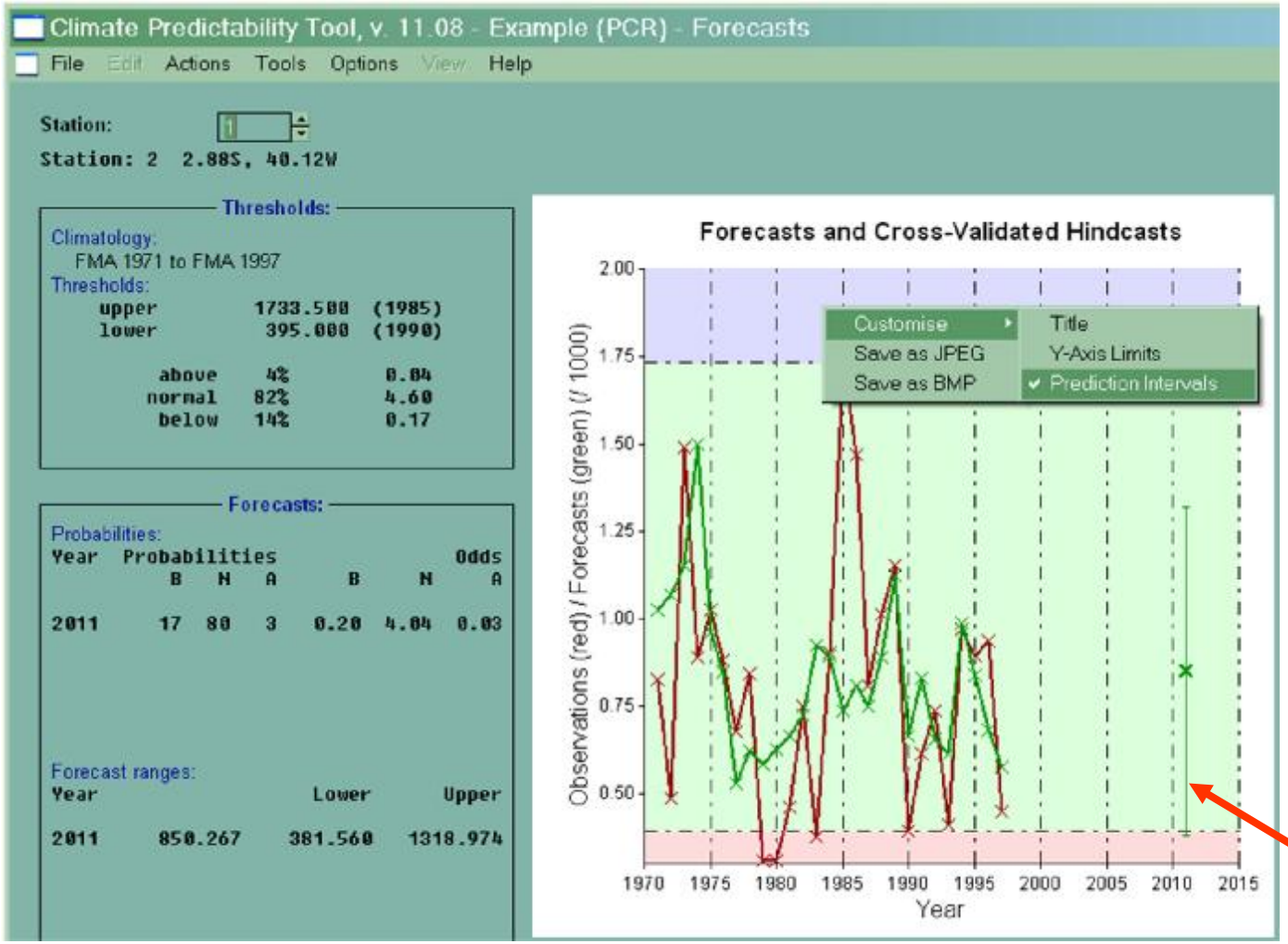
CHANGING CATEGORY DEFINITIONS

3. Set analogue years



CPT recalculates the thresholds, and the climatological probabilities and odds

PREDICTION INTERVALS

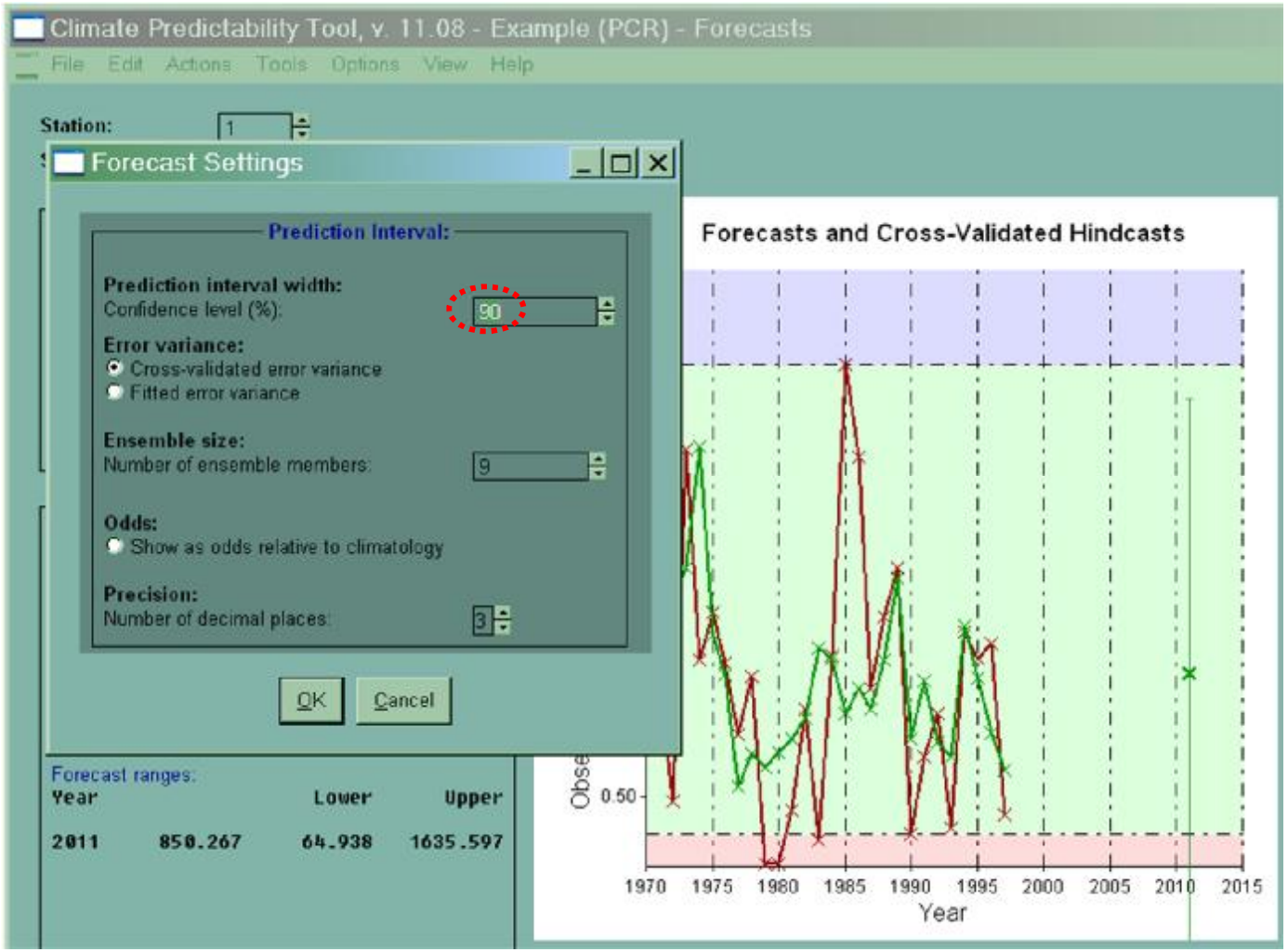


To draw error bars on the forecast, right click on the graph:

Customize ~ Prediction Intervals

An error bar is indicated.

CHANGING THE PREDICTION INTERVAL

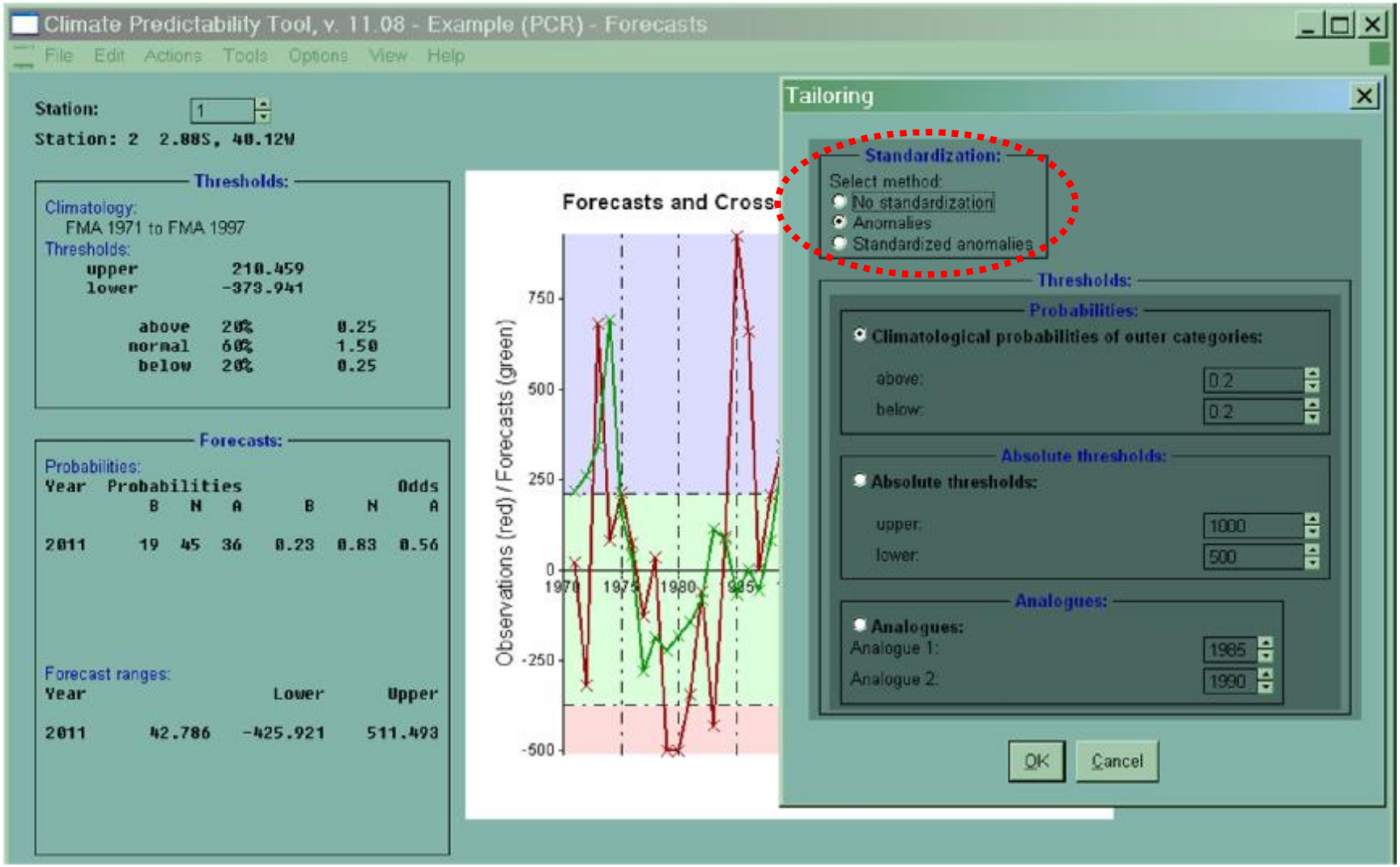


You can also change the width of the prediction interval.

Options ~ Forecast Settings ~ Prediction interval width

The default setting of 68.3% gives standard error bars.

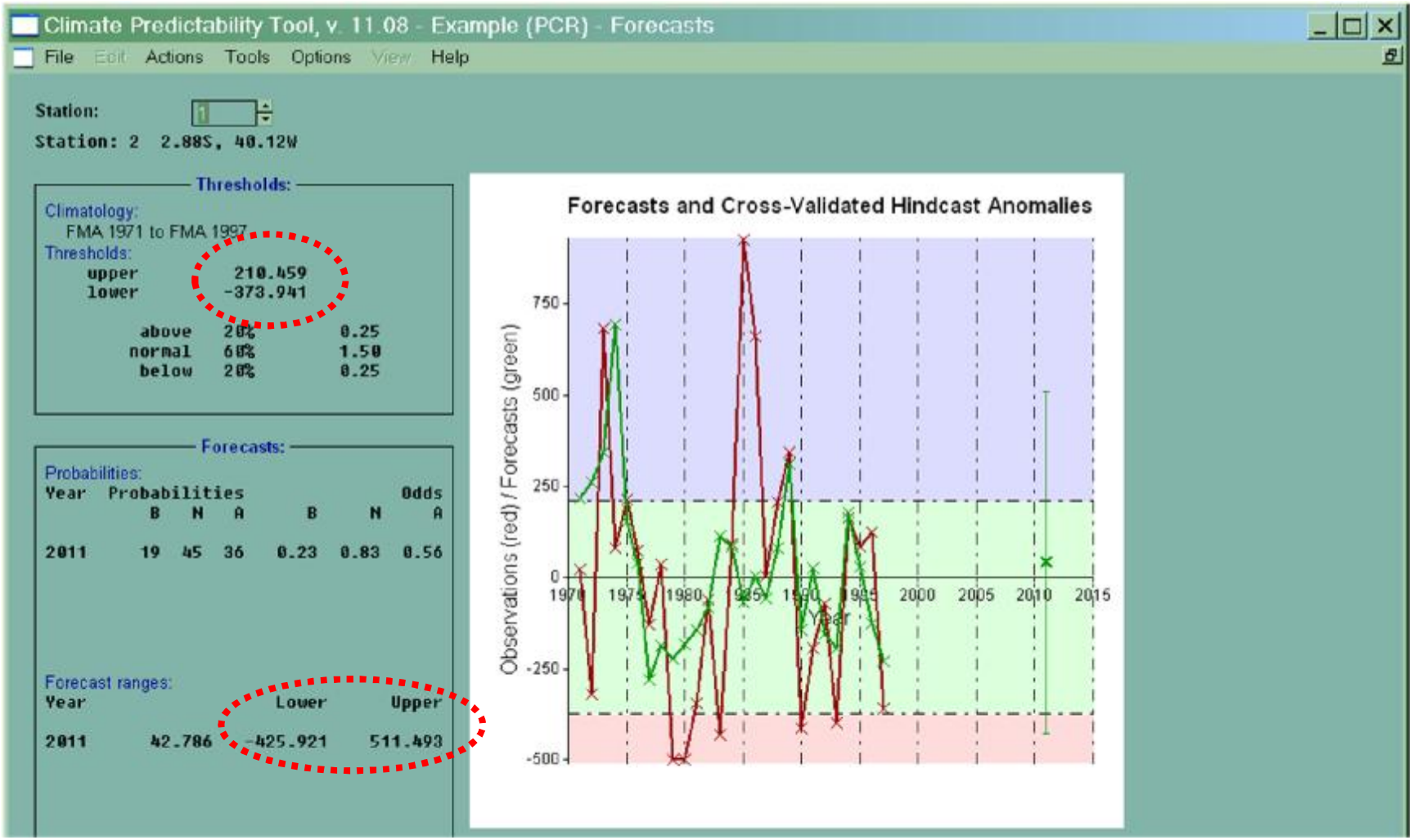
EXPRESSING THE FORECAST AS ANOMALIES



The forecast can be expressed as anomalies,
rather than absolute values:

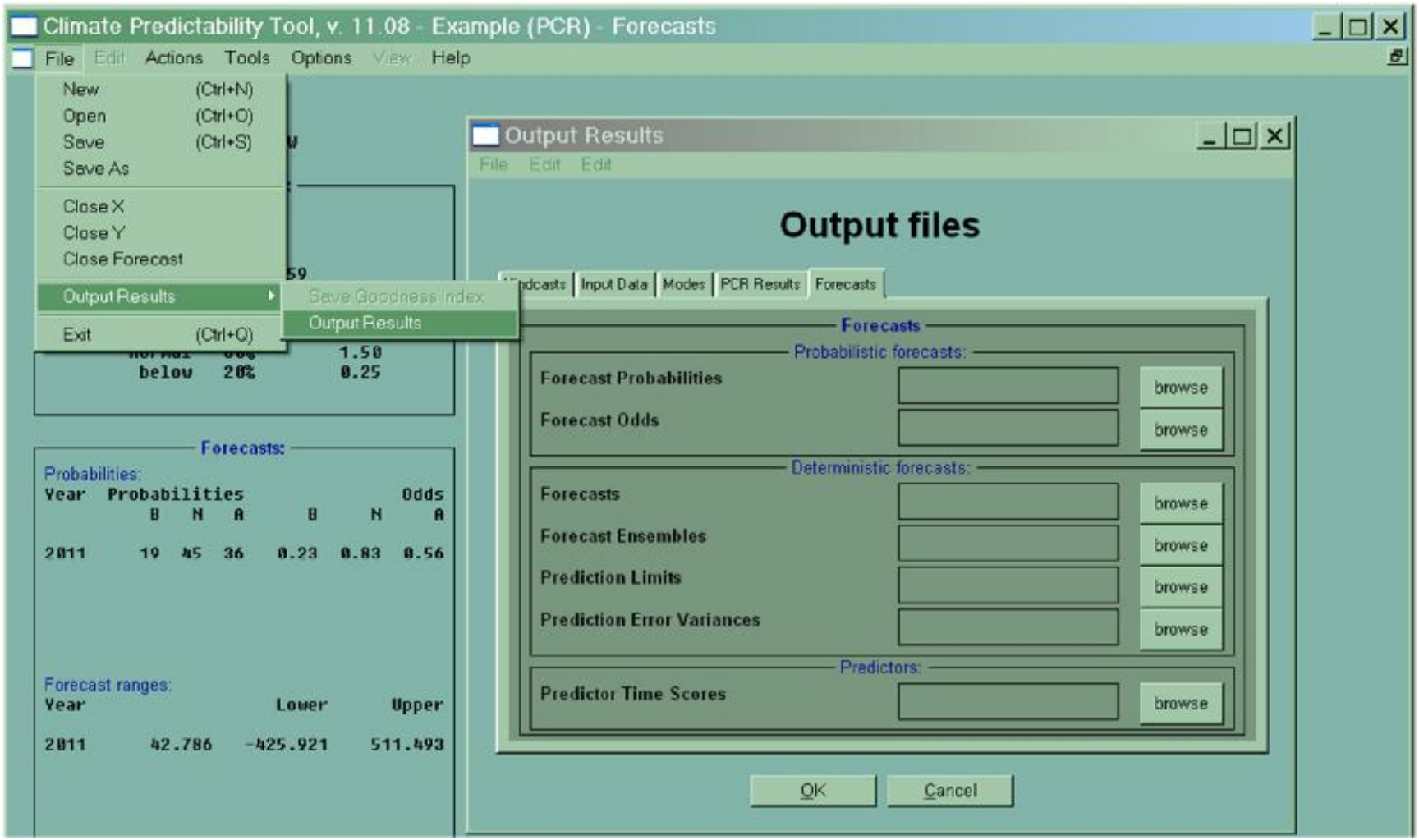
Options ~ Forecast Settings ~ Standardization ~ Anomalies

EXPRESSING THE FORECAST AS ANOMALIES



The thresholds, as well as the forecast ranges, are now defined as anomalies.

SAVING FORECASTS

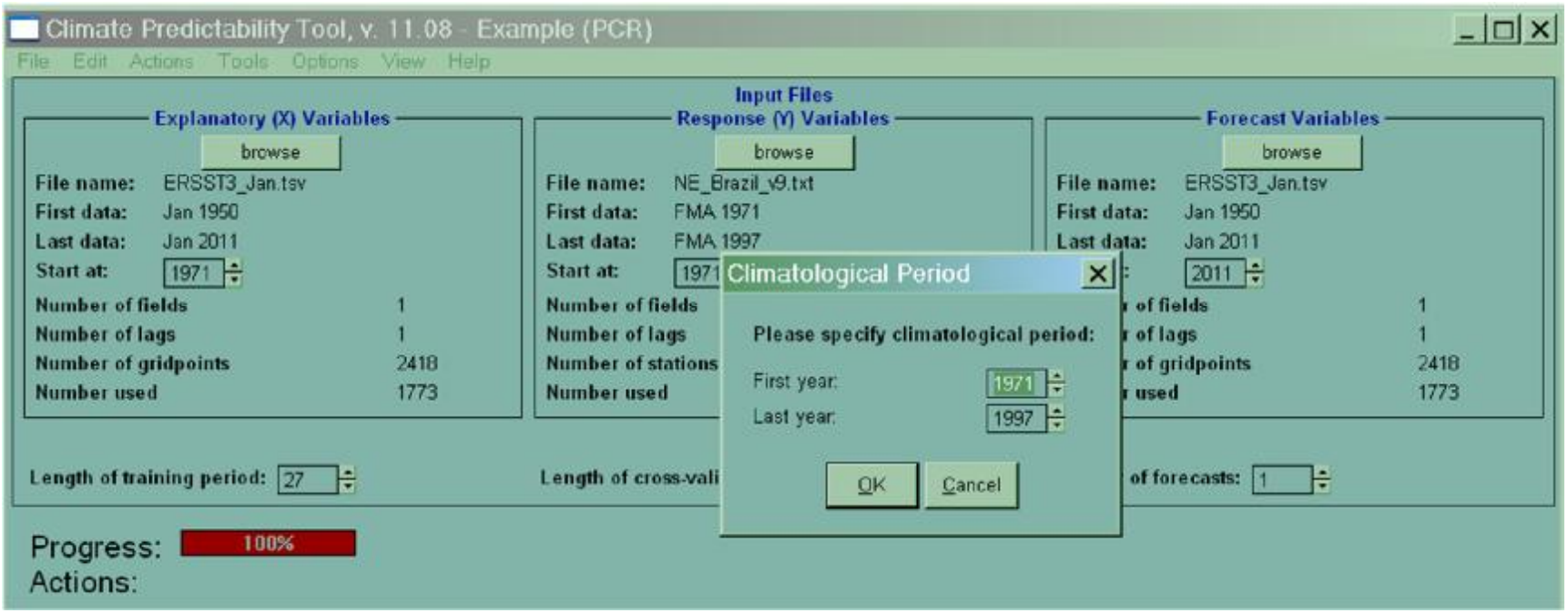


To save the forecasts, go to

File ~ Output Results ~ Output Results

and on the Forecasts tab specify the required output files.

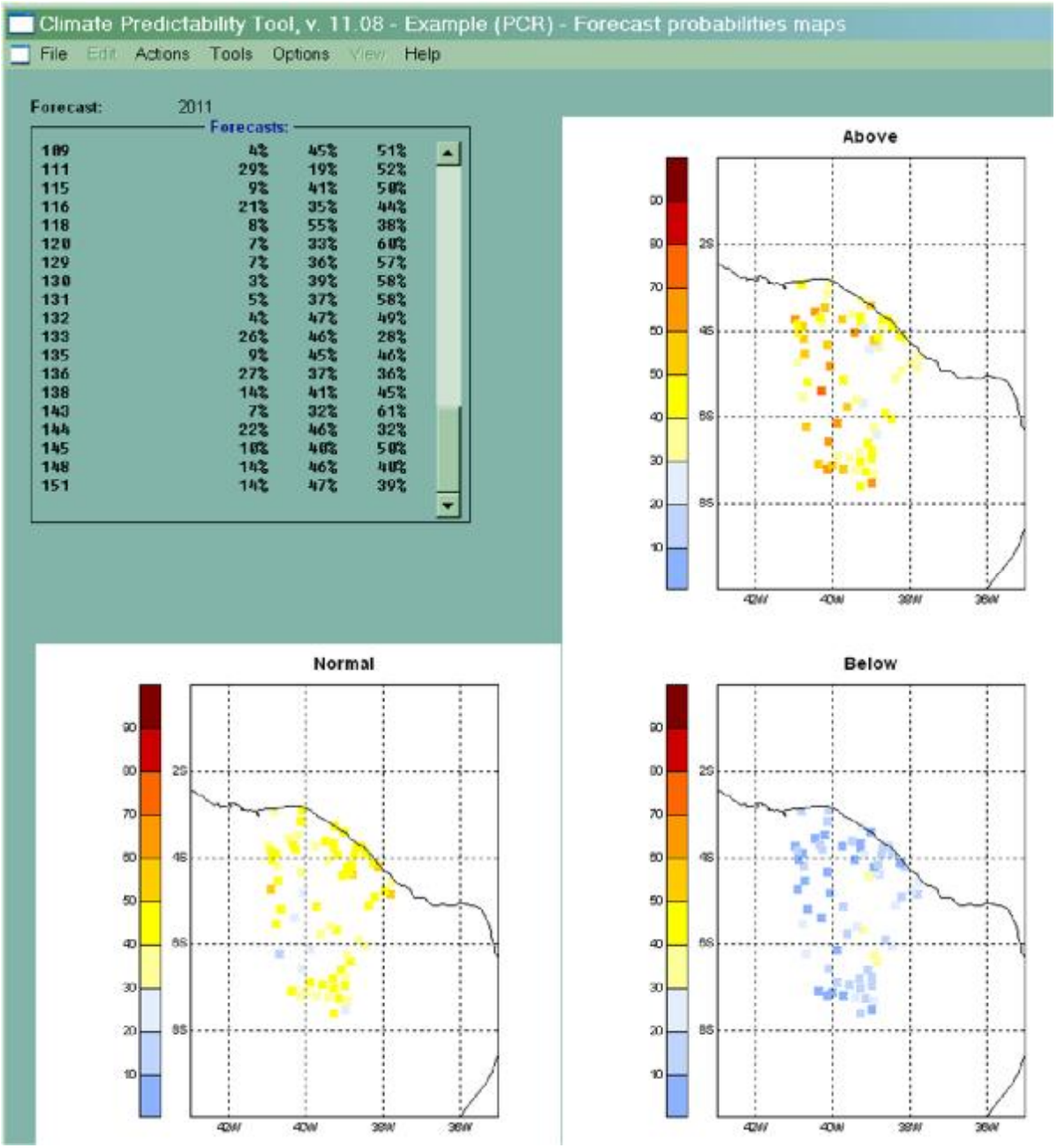
CHANGING THE CLIMATOLOGICAL PERIOD



By default, the forecast probabilities are calculated relative to a climatological period that is the same as the training period. To change the climatological period go to:

Options ~ Climatological Period

FORECAST MAPS



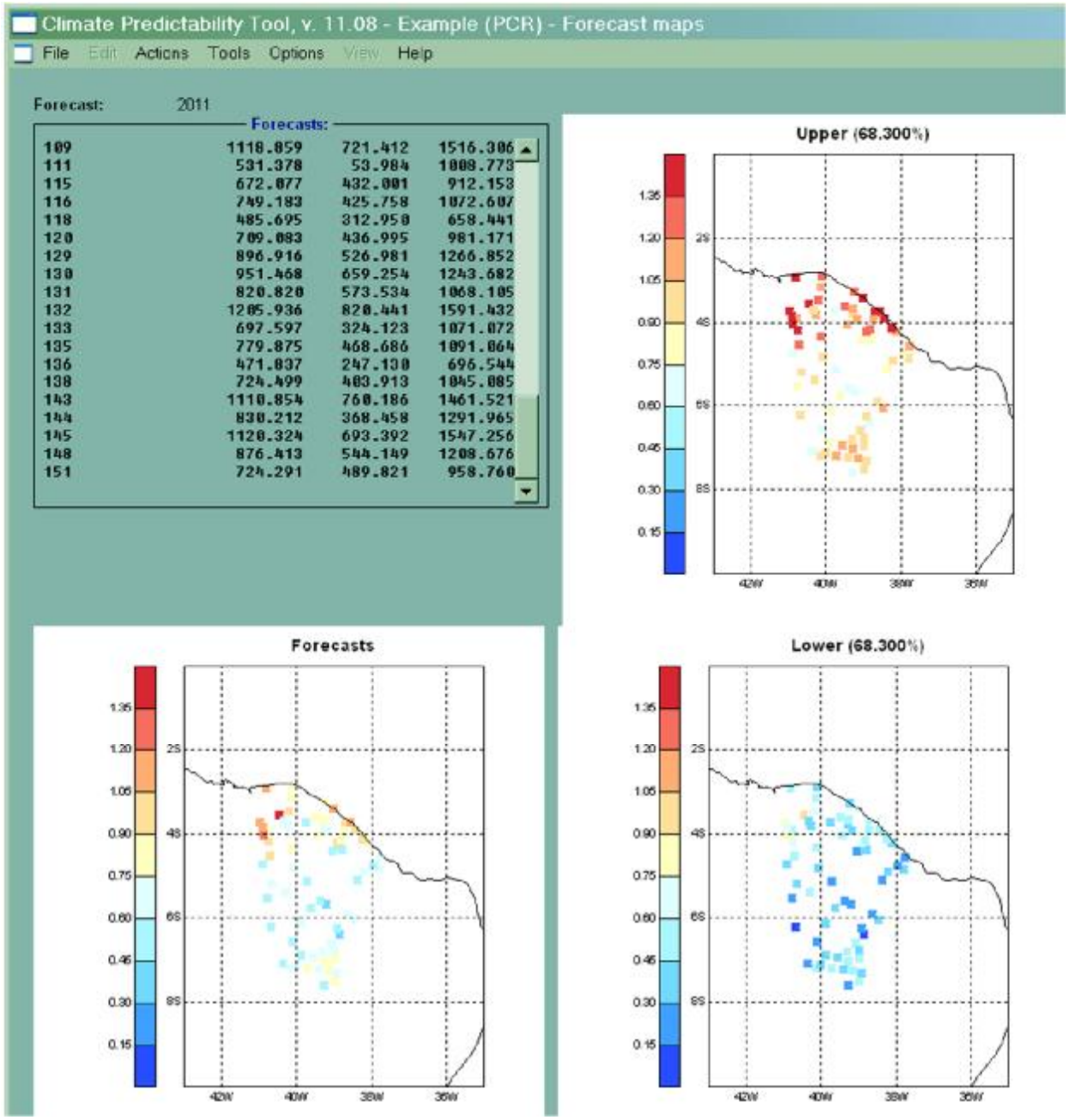
Tools ~ Forecast ~ Maps

The option Maps lets you see maps of your forecasts – either maps of the probabilities or maps of the actual forecast values.

The forecast probabilities map lists the probabilities for each category at each location as well as the spatial distribution of the probabilities.

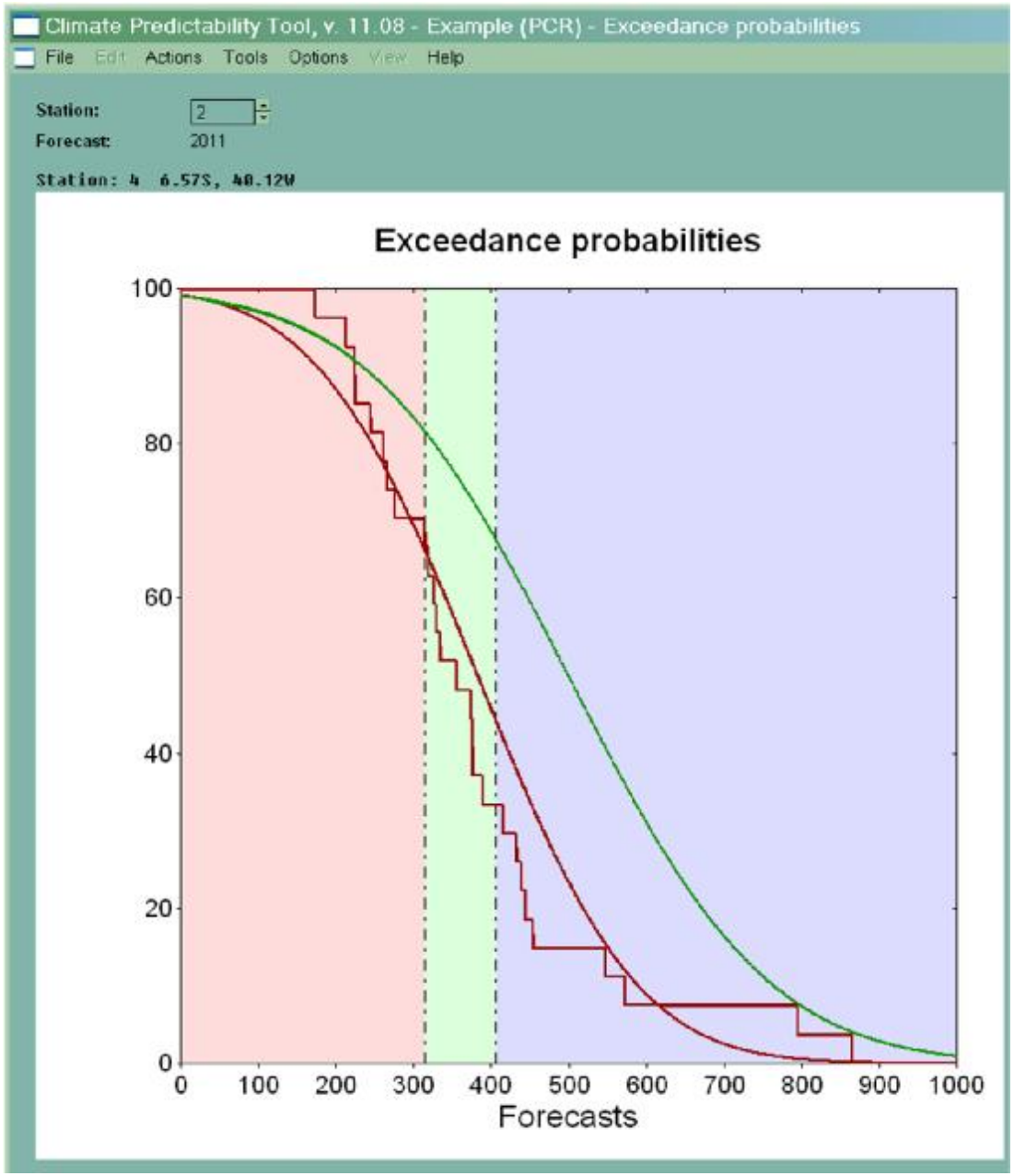
In this example, the below-normal category has the lowest probability over most of north-east Brazil.

FORECAST MAPS



The forecast values map lists the actual forecast values for each category at each location as well as the spatial distribution of the values.

EXCEEDANCE PROBABILITIES



To draw the probabilities of exceedance go to: Tools ~ Forecast ~ Exceedances

CONCLUSIONS

- For further details, read the help page of each menu and option.
- Subscribe to the user-list to be advised of updates: http://iri.columbia.edu/climate/tools/CPT/
- We want to hear from you. Your comments and questions help us to improve the CPT so do not hesitate to write to us at: cpt@iri.columbia.edu