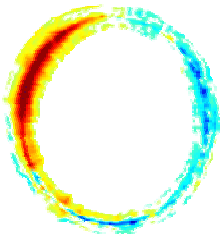
# Climate Predictability Tool (CPT)

***Ousmane Ndiaye and Simon J. Mason***

***cpt@iri.columbia.edu***



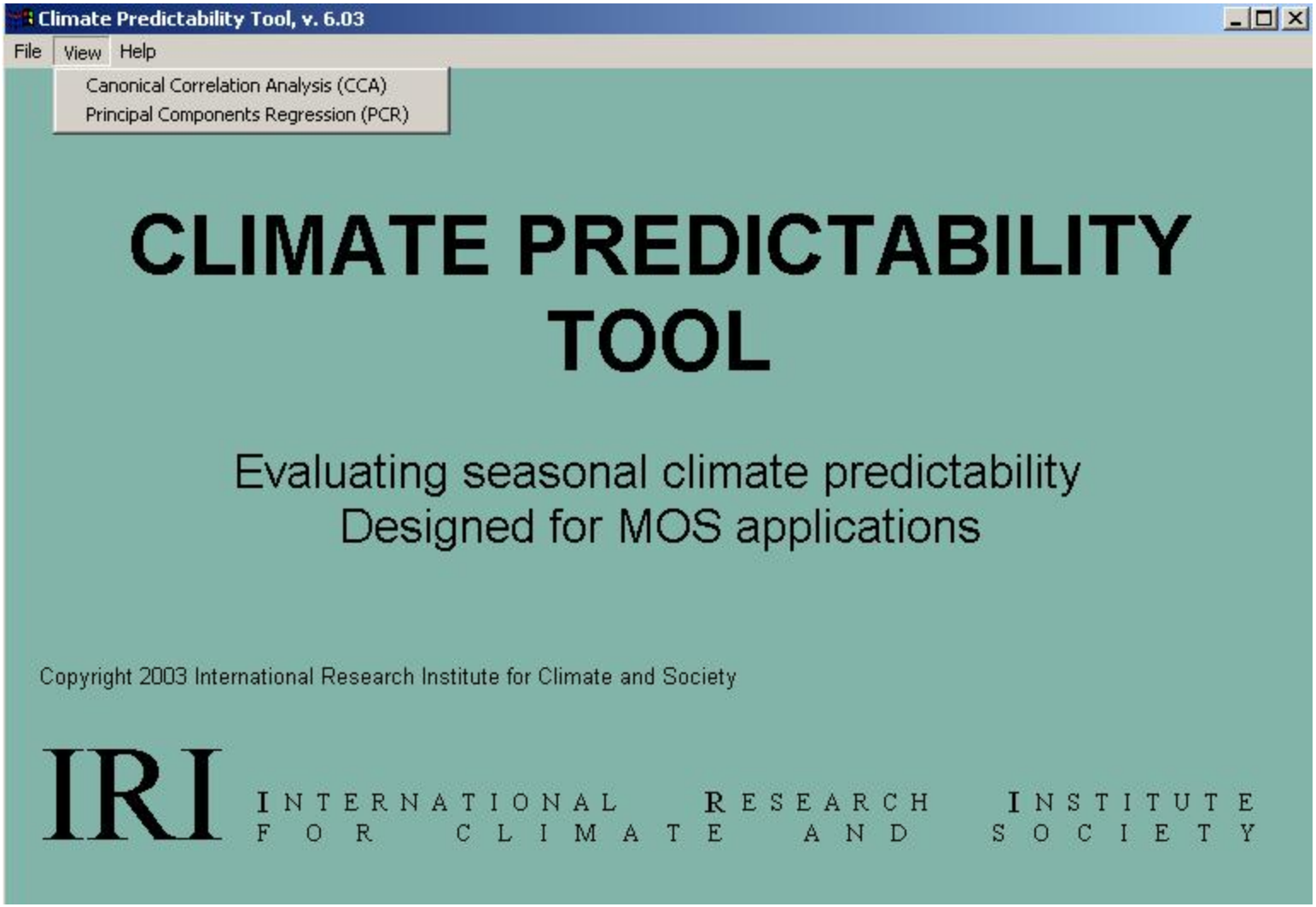
**International Research Institute for Climate and Society**

**The Earth Institute of Columbia University**

# OVERVIEW

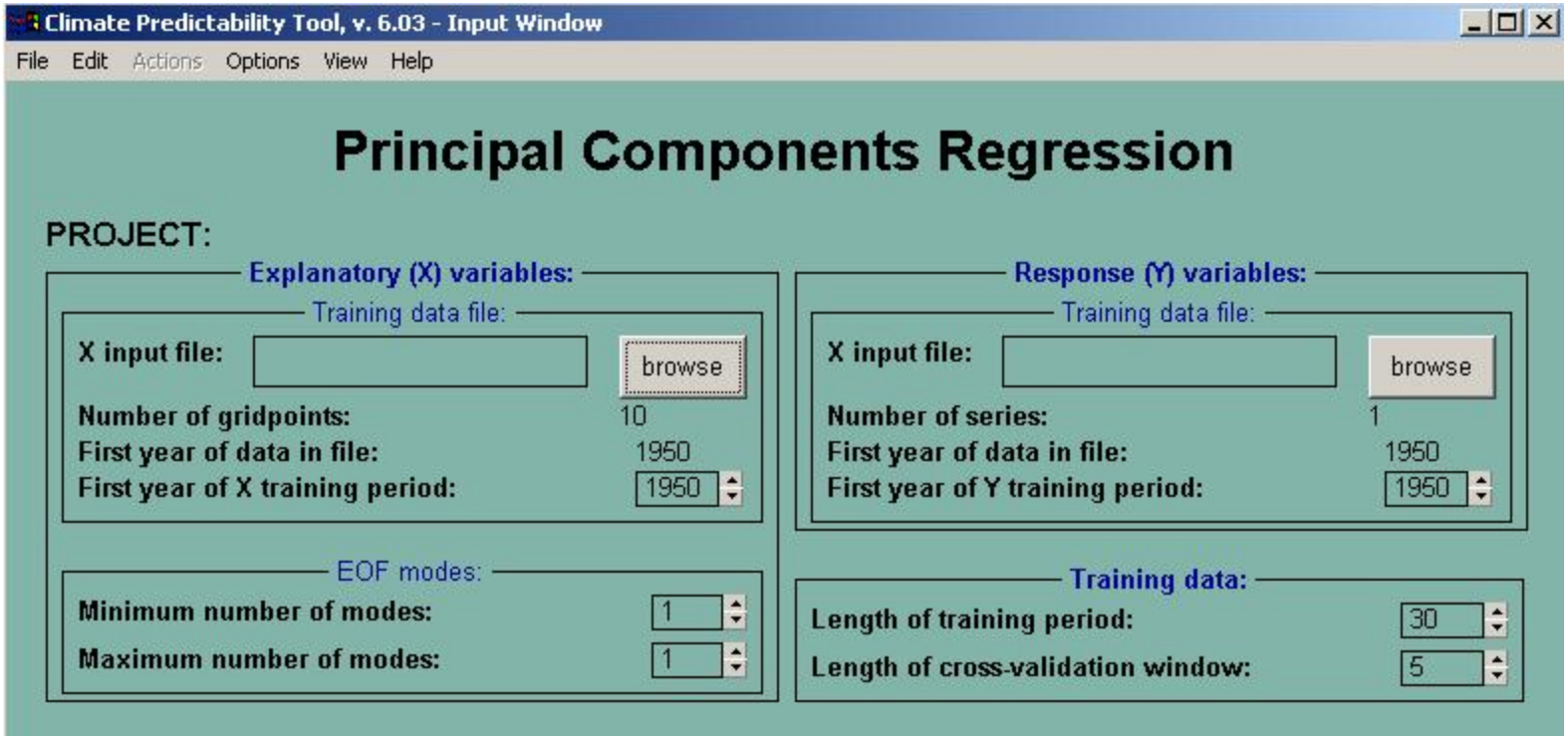
- The Climate Predictability Tool (CPT) provides a Windows package for :
  - Ø seasonal climate forecasting
  - Ø model validation
  - Ø actual forecasts given updated data
- Uses ASCII input files
- Options :
  - Ø principal components regression (PCR)
  - Ø canonical correlation analysis (CCA)
- Help Pages on a range of topics in HTML format
- Options to save outputs in ASCII format and graphics as JPEG
- Program source code is now available for those using other systems (e.g., UNIX)

# SELECTING THE ANALYSIS




Choose the analysis to perform: PCR or CCA

# INPUT DATASETS

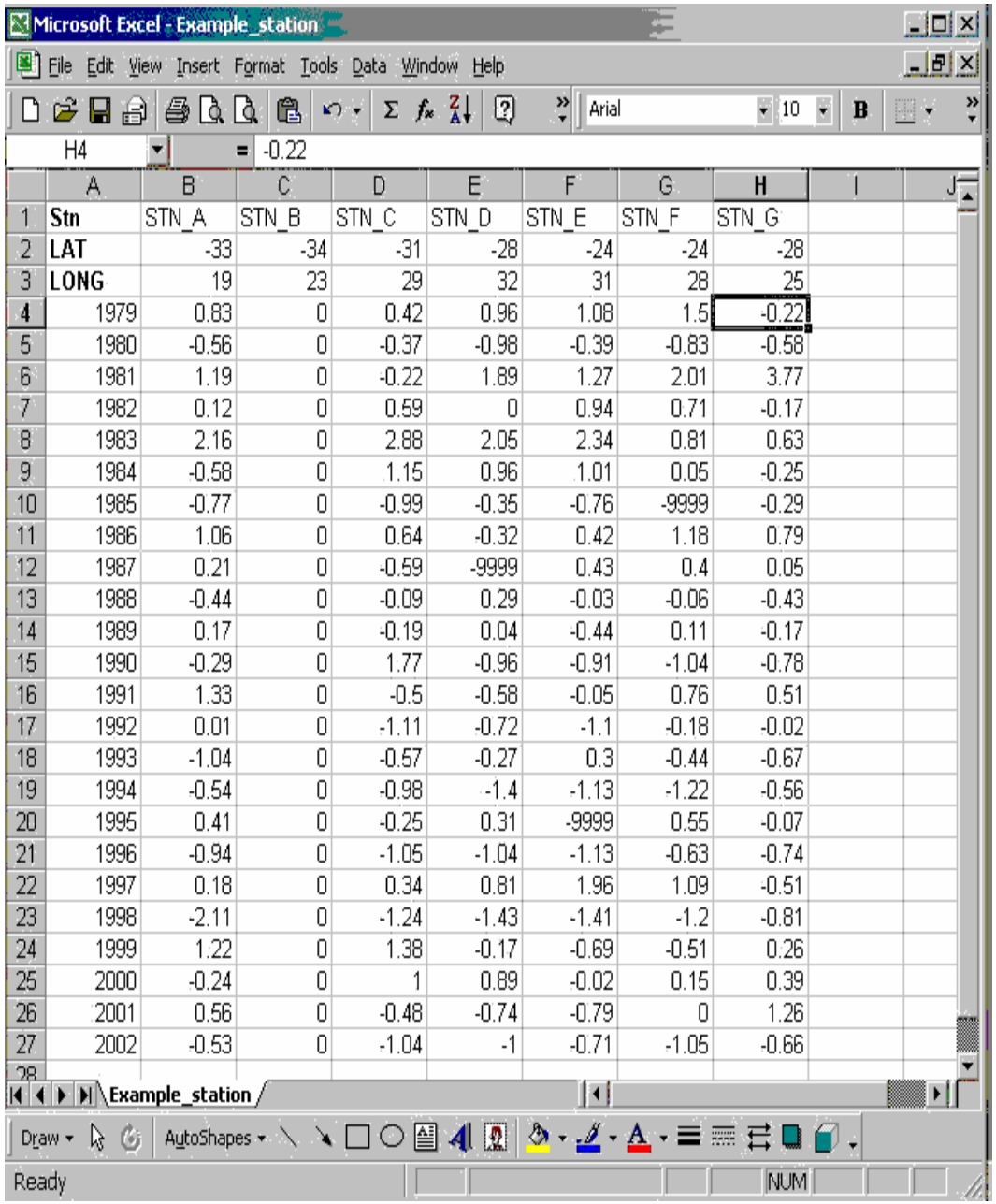



Both analysis methods require two datasets:

"X variables" or "X Predictors" dataset;

"Y variables" or "Y Predictands" dataset.

# CPT INPUT FILE FORMATS



Microsoft Excel - Example_station

File Edit View Insert Format Tools Data Window Help

Arial 10 B

H4 = -0.22

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Stn | STN_A | STN_B | STN_C | STN_D | STN_E | STN_F | STN_G |
| 2 | LAT | -33 | -34 | -31 | -28 | -24 | -24 | -28 |
| 3 | LONG | 19 | 23 | 29 | 32 | 31 | 28 | 25 |
| 4 | 1979 | 0.83 | 0 | 0.42 | 0.96 | 1.08 | 1.5 | -0.22 |
| 5 | 1980 | -0.56 | 0 | -0.37 | -0.98 | -0.39 | -0.83 | -0.58 |
| 6 | 1981 | 1.19 | 0 | -0.22 | 1.89 | 1.27 | 2.01 | 3.77 |
| 7 | 1982 | 0.12 | 0 | 0.59 | 0 | 0.94 | 0.71 | -0.17 |
| 8 | 1983 | 2.16 | 0 | 2.88 | 2.05 | 2.34 | 0.81 | 0.63 |
| 9 | 1984 | -0.58 | 0 | 1.15 | 0.96 | 1.01 | 0.05 | -0.25 |
| 10 | 1985 | -0.77 | 0 | -0.99 | -0.35 | -0.76 | -9999 | -0.29 |
| 11 | 1986 | 1.06 | 0 | 0.64 | -0.32 | 0.42 | 1.18 | 0.79 |
| 12 | 1987 | 0.21 | 0 | -0.59 | -9999 | 0.43 | 0.4 | 0.05 |
| 13 | 1988 | -0.44 | 0 | -0.09 | 0.29 | -0.03 | -0.06 | -0.43 |
| 14 | 1989 | 0.17 | 0 | -0.19 | 0.04 | -0.44 | 0.11 | -0.17 |
| 15 | 1990 | -0.29 | 0 | 1.77 | -0.96 | -0.91 | -1.04 | -0.78 |
| 16 | 1991 | 1.33 | 0 | -0.5 | -0.58 | -0.05 | 0.76 | 0.51 |
| 17 | 1992 | 0.01 | 0 | -1.11 | -0.72 | -1.1 | -0.18 | -0.02 |
| 18 | 1993 | -1.04 | 0 | -0.57 | -0.27 | 0.3 | -0.44 | -0.67 |
| 19 | 1994 | -0.54 | 0 | -0.98 | -1.4 | -1.13 | -1.22 | -0.56 |
| 20 | 1995 | 0.41 | 0 | -0.25 | 0.31 | -9999 | 0.55 | -0.07 |
| 21 | 1996 | -0.94 | 0 | -1.05 | -1.04 | -1.13 | -0.63 | -0.74 |
| 22 | 1997 | 0.18 | 0 | 0.34 | 0.81 | 1.96 | 1.09 | -0.51 |
| 23 | 1998 | -2.11 | 0 | -1.24 | -1.43 | -1.41 | -1.2 | -0.81 |
| 24 | 1999 | 1.22 | 0 | 1.38 | -0.17 | -0.69 | -0.51 | 0.26 |
| 25 | 2000 | -0.24 | 0 | 1 | 0.89 | -0.02 | 0.15 | 0.39 |
| 26 | 2001 | 0.56 | 0 | -0.48 | -0.74 | -0.79 | 0 | 1.26 |
| 27 | 2002 | -0.53 | 0 | -1.04 | -1 | -0.71 | -1.05 | -0.66 |

Example_station

Draw AutoShapes

Ready NUM

## 1. STATION files:

This file-type contains :

**Station_name** (without spaces; ≤16 characters)

**Latitude** (in signed degrees)

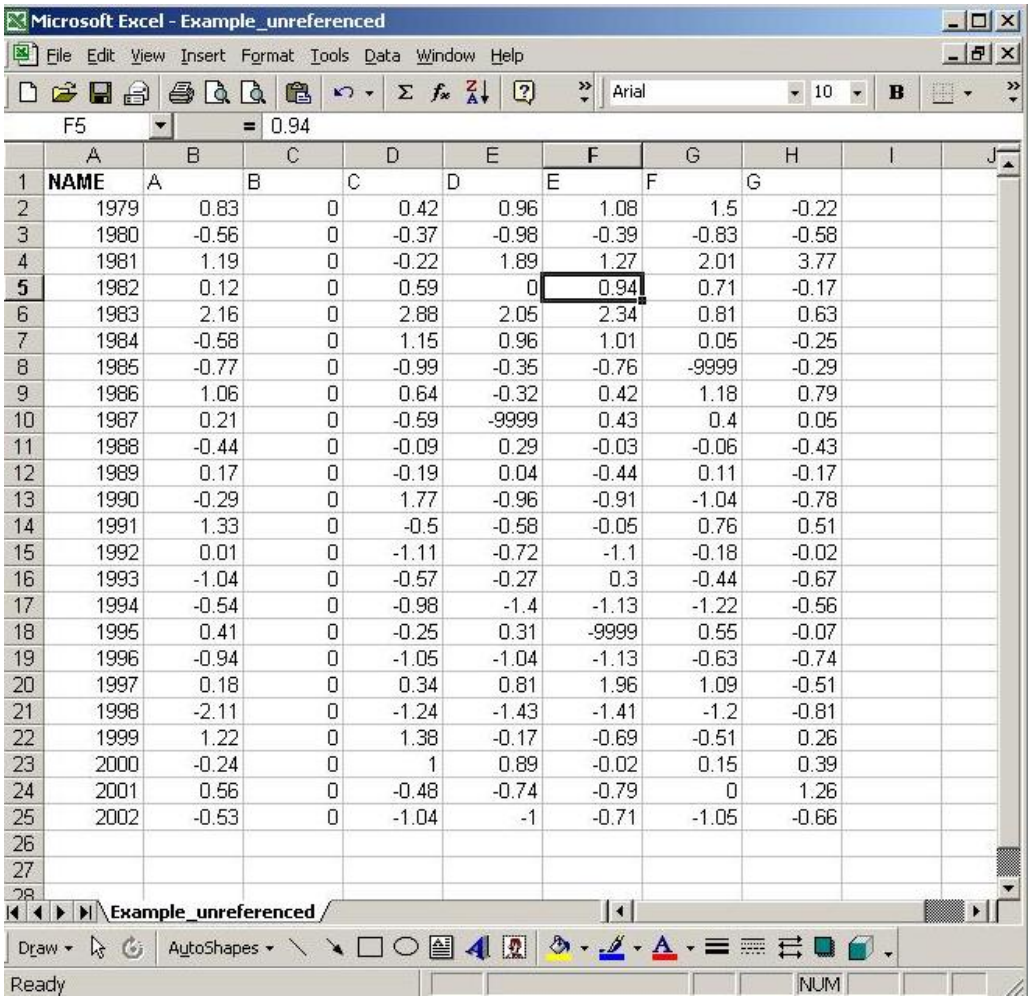
**Longitude** (signed degrees)

**Year** (in the first column)

**Data** (missing values should be filled with the same value, -9999 for example)

**Keywords:**
STN, LAT, LONG

# CPT INPUT FILE FORMATS




| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | NAME | A | B | C | D | E | F | G | | |
| 2 | 1979 | 0.83 | 0 | 0.42 | 0.96 | 1.08 | 1.5 | -0.22 | | |
| 3 | 1980 | -0.56 | 0 | -0.37 | -0.98 | -0.39 | -0.83 | -0.58 | | |
| 4 | 1981 | 1.19 | 0 | -0.22 | 1.89 | 1.27 | 2.01 | 3.77 | | |
| 5 | 1982 | 0.12 | 0 | 0.59 | 0 | 0.94 | 0.71 | -0.17 | | |
| 6 | 1983 | 2.16 | 0 | 2.88 | 2.05 | 2.34 | 0.81 | 0.63 | | |
| 7 | 1984 | -0.58 | 0 | 1.15 | 0.96 | 1.01 | 0.05 | -0.25 | | |
| 8 | 1985 | -0.77 | 0 | -0.99 | -0.35 | -0.76 | -9999 | -0.29 | | |
| 9 | 1986 | 1.06 | 0 | 0.64 | -0.32 | 0.42 | 1.18 | 0.79 | | |
| 10 | 1987 | 0.21 | 0 | -0.59 | -9999 | 0.43 | 0.4 | 0.05 | | |
| 11 | 1988 | -0.44 | 0 | -0.09 | 0.29 | -0.03 | -0.06 | -0.43 | | |
| 12 | 1989 | 0.17 | 0 | -0.19 | 0.04 | -0.44 | 0.11 | -0.17 | | |
| 13 | 1990 | -0.29 | 0 | 1.77 | -0.96 | -0.91 | -1.04 | -0.78 | | |
| 14 | 1991 | 1.33 | 0 | -0.5 | -0.58 | -0.05 | 0.76 | 0.51 | | |
| 15 | 1992 | 0.01 | 0 | -1.11 | -0.72 | -1.1 | -0.18 | -0.02 | | |
| 16 | 1993 | -1.04 | 0 | -0.57 | -0.27 | 0.3 | -0.44 | -0.67 | | |
| 17 | 1994 | -0.54 | 0 | -0.98 | -1.4 | -1.13 | -1.22 | -0.56 | | |
| 18 | 1995 | 0.41 | 0 | -0.25 | 0.31 | -9999 | 0.55 | -0.07 | | |
| 19 | 1996 | -0.94 | 0 | -1.05 | -1.04 | -1.13 | -0.63 | -0.74 | | |
| 20 | 1997 | 0.18 | 0 | 0.34 | 0.81 | 1.96 | 1.09 | -0.51 | | |
| 21 | 1998 | -2.11 | 0 | -1.24 | -1.43 | -1.41 | -1.2 | -0.81 | | |
| 22 | 1999 | 1.22 | 0 | 1.38 | -0.17 | -0.69 | -0.51 | 0.26 | | |
| 23 | 2000 | -0.24 | 0 | 1 | 0.89 | -0.02 | 0.15 | 0.39 | | |
| 24 | 2001 | 0.56 | 0 | -0.48 | -0.74 | -0.79 | 0 | 1.26 | | |
| 25 | 2002 | -0.53 | 0 | -1.04 | -1 | -0.71 | -1.05 | -0.66 | | |
| 26 | | | | | | | | | | |
| 27 | | | | | | | | | | |

## 2. <u>UNREFERENCED or Indices files</u>:

The data are not referenced (no latitudes and longitudes):

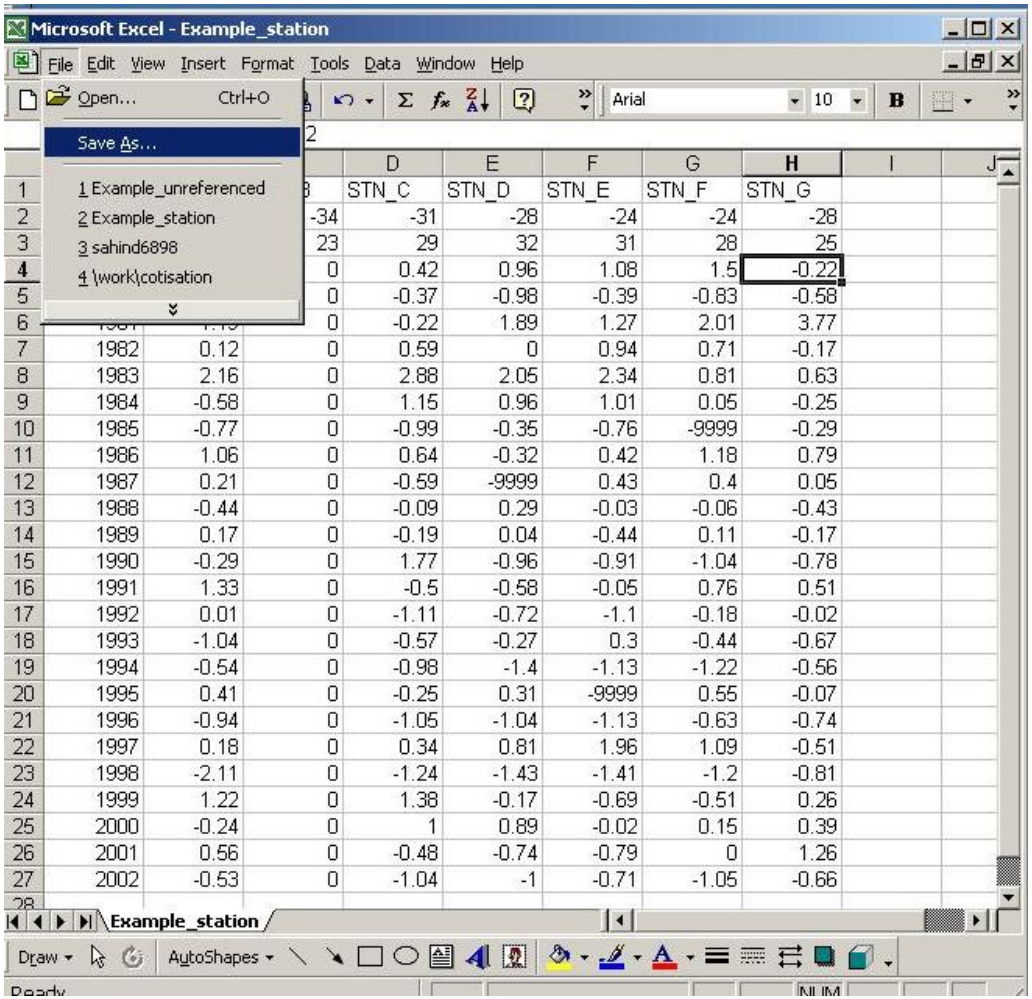
**Index_name** (without spaces; ≤16 characters)

**Year** (in the first column)

**Data** (with missing data)

**<u>Keywords:</u>**
NAME or YEAR

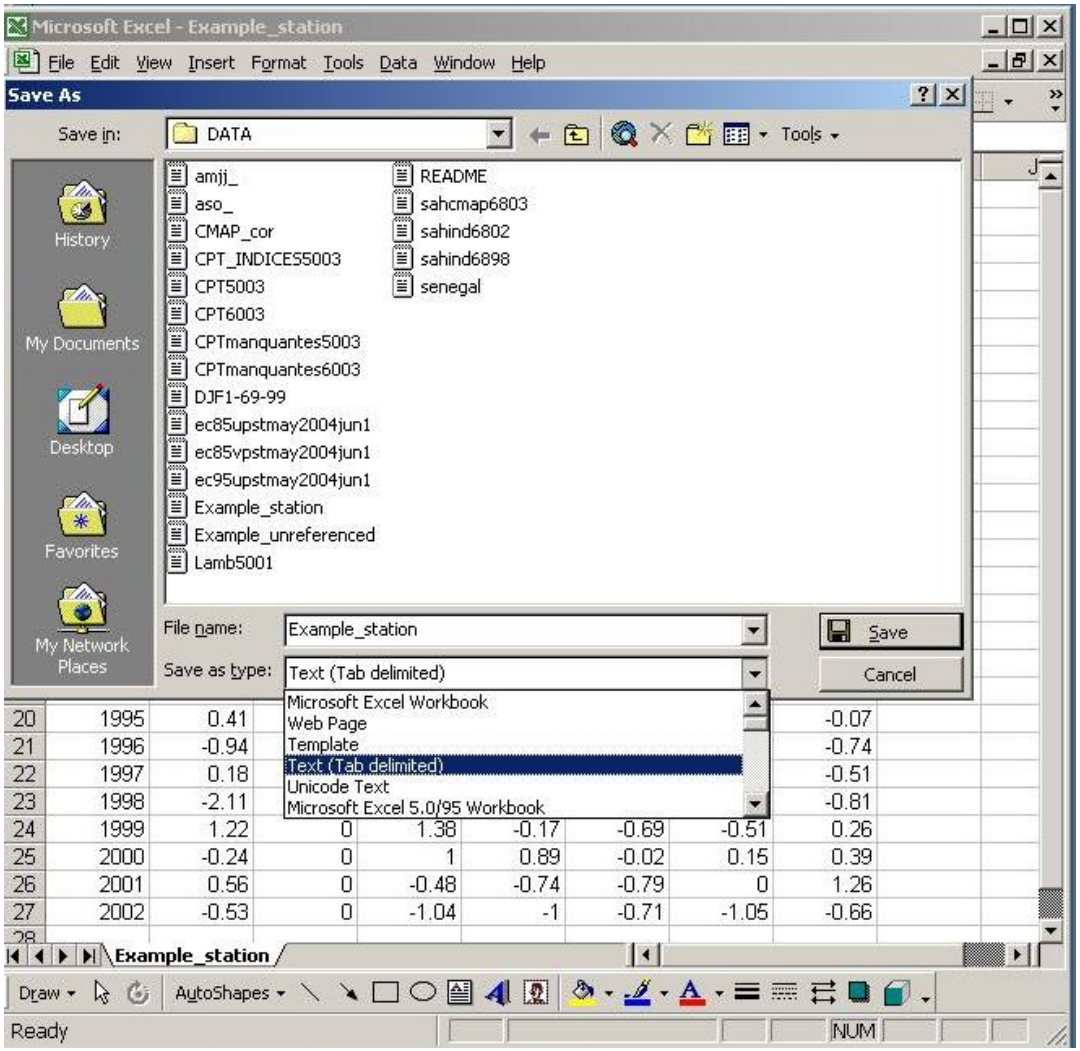
# CPT INPUT FILE FORMATS




| | | | | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | STN_C | STN_D | STN_E | STN_F | STN_G |
| 2 | | | -34 | -31 | -28 | -24 | -24 | -28 |
| 3 | | | 23 | 29 | 32 | 31 | 28 | 25 |
| 4 | | | 0 | 0.42 | 0.96 | 1.08 | 1.5 | -0.22 |
| 5 | | | 0 | -0.37 | -0.98 | -0.39 | -0.83 | -0.58 |
| 6 | | | 0 | -0.22 | 1.89 | 1.27 | 2.01 | 3.77 |
| 7 | 1982 | 0.12 | 0 | 0.59 | 0 | 0.94 | 0.71 | -0.17 |
| 8 | 1983 | 2.16 | 0 | 2.88 | 2.05 | 2.34 | 0.81 | 0.63 |
| 9 | 1984 | -0.58 | 0 | 1.15 | 0.96 | 1.01 | 0.05 | -0.25 |
| 10 | 1985 | -0.77 | 0 | -0.99 | -0.35 | -0.76 | -9999 | -0.29 |
| 11 | 1986 | 1.06 | 0 | 0.64 | -0.32 | 0.42 | 1.18 | 0.79 |
| 12 | 1987 | 0.21 | 0 | -0.59 | -9999 | 0.43 | 0.4 | 0.05 |
| 13 | 1988 | -0.44 | 0 | -0.09 | 0.29 | -0.03 | -0.06 | -0.43 |
| 14 | 1989 | 0.17 | 0 | -0.19 | 0.04 | -0.44 | 0.11 | -0.17 |
| 15 | 1990 | -0.29 | 0 | 1.77 | -0.96 | -0.91 | -1.04 | -0.78 |
| 16 | 1991 | 1.33 | 0 | -0.5 | -0.58 | -0.05 | 0.76 | 0.51 |
| 17 | 1992 | 0.01 | 0 | -1.11 | -0.72 | -1.1 | -0.18 | -0.02 |
| 18 | 1993 | -1.04 | 0 | -0.57 | -0.27 | 0.3 | -0.44 | -0.67 |
| 19 | 1994 | -0.54 | 0 | -0.98 | -1.4 | -1.13 | -1.22 | -0.56 |
| 20 | 1995 | 0.41 | 0 | -0.25 | 0.31 | -9999 | 0.55 | -0.07 |
| 21 | 1996 | -0.94 | 0 | -1.05 | -1.04 | -1.13 | -0.63 | -0.74 |
| 22 | 1997 | 0.18 | 0 | 0.34 | 0.81 | 1.96 | 1.09 | -0.51 |
| 23 | 1998 | -2.11 | 0 | -1.24 | -1.43 | -1.41 | -1.2 | -0.81 |
| 24 | 1999 | 1.22 | 0 | 1.38 | -0.17 | -0.69 | -0.51 | 0.26 |
| 25 | 2000 | -0.24 | 0 | 1 | 0.89 | -0.02 | 0.15 | 0.39 |
| 26 | 2001 | 0.56 | 0 | -0.48 | -0.74 | -0.79 | 0 | 1.26 |
| 27 | 2002 | -0.53 | 0 | -1.04 | -1 | -0.71 | -1.05 | -0.66 |

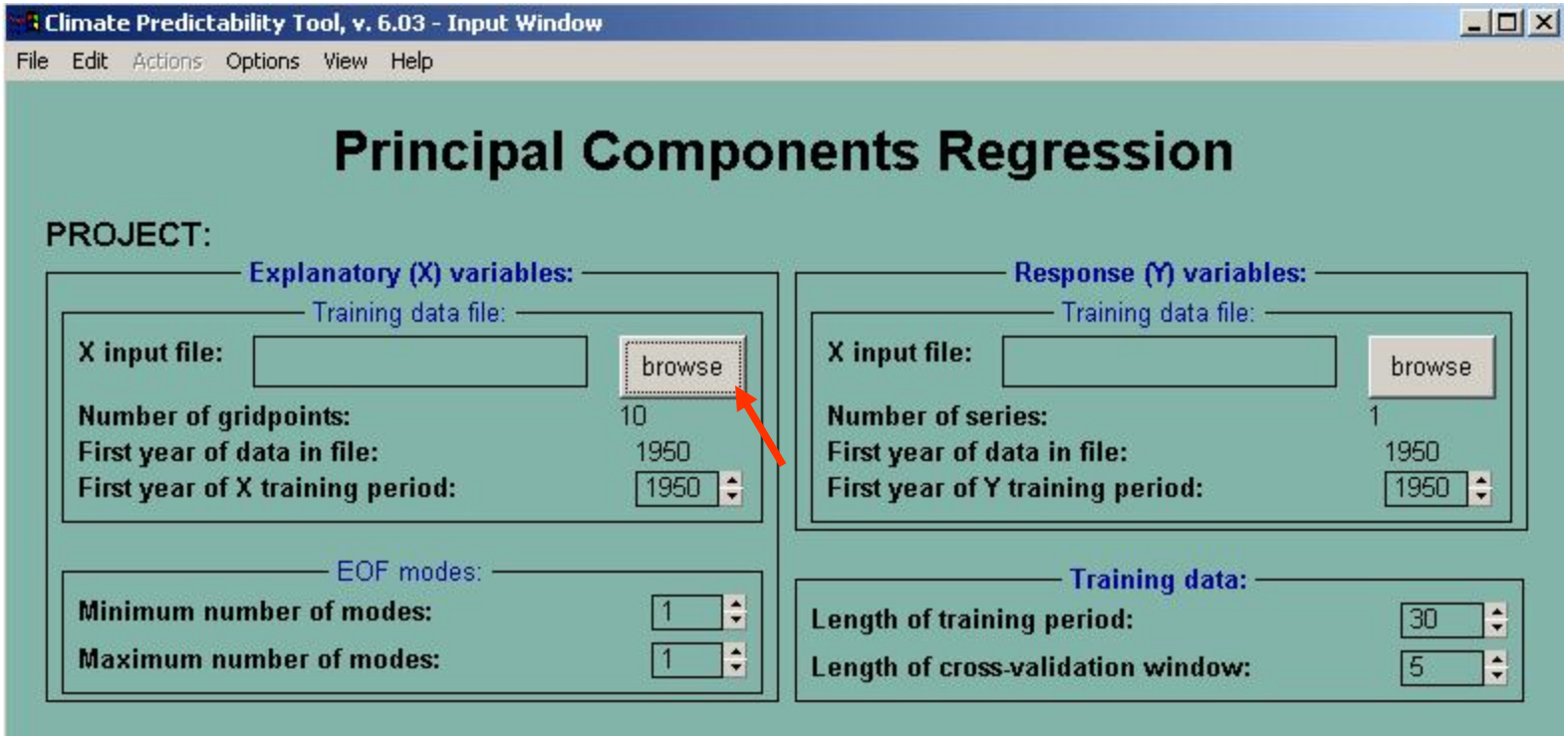
The input files could be easily made using a spreadsheet such as Excel

# CPT INPUT FILE FORMATS



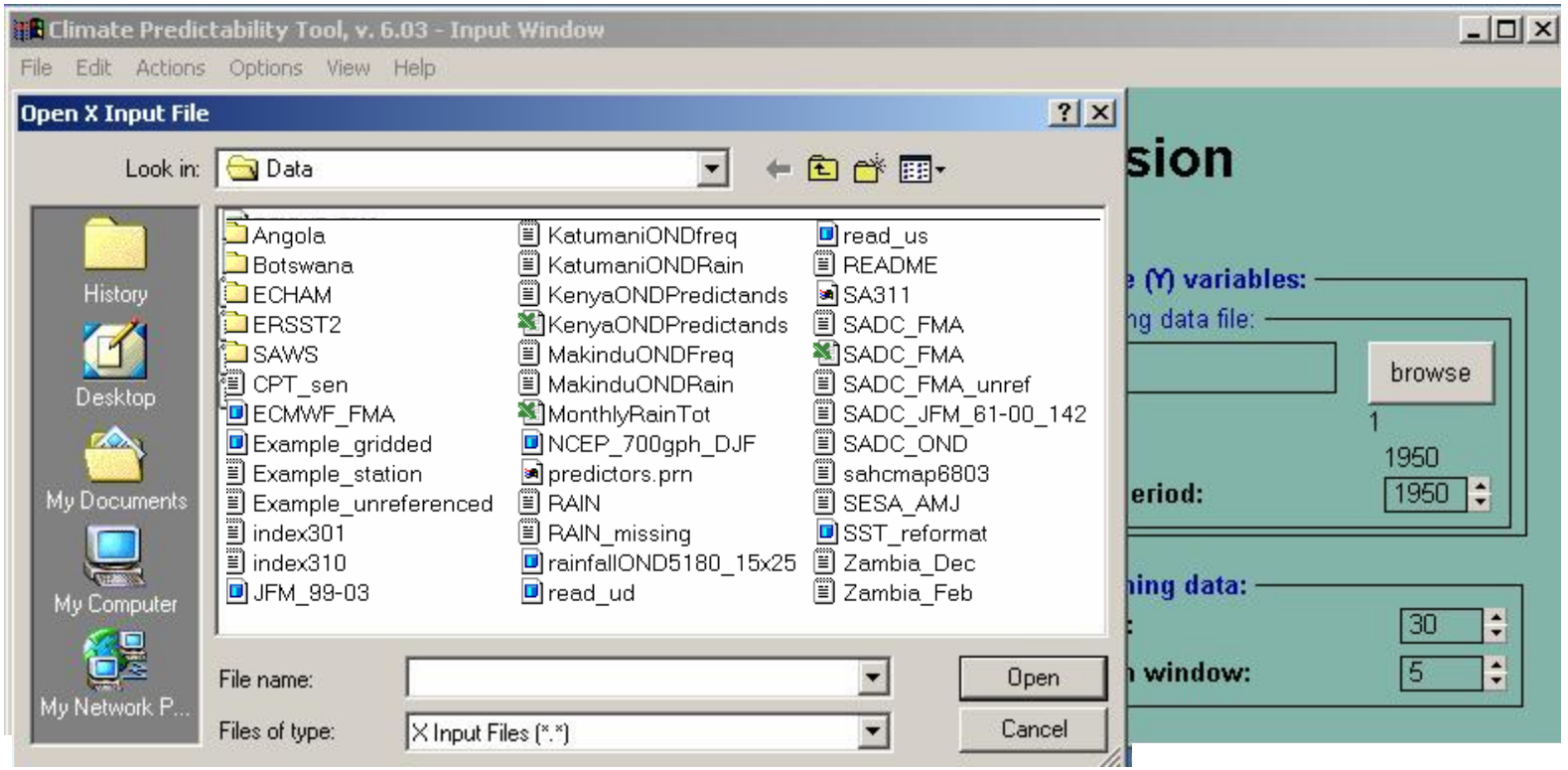

In Excel the file should be saved as:
“Text, tab delimited”

# SELECTING INPUT FILES



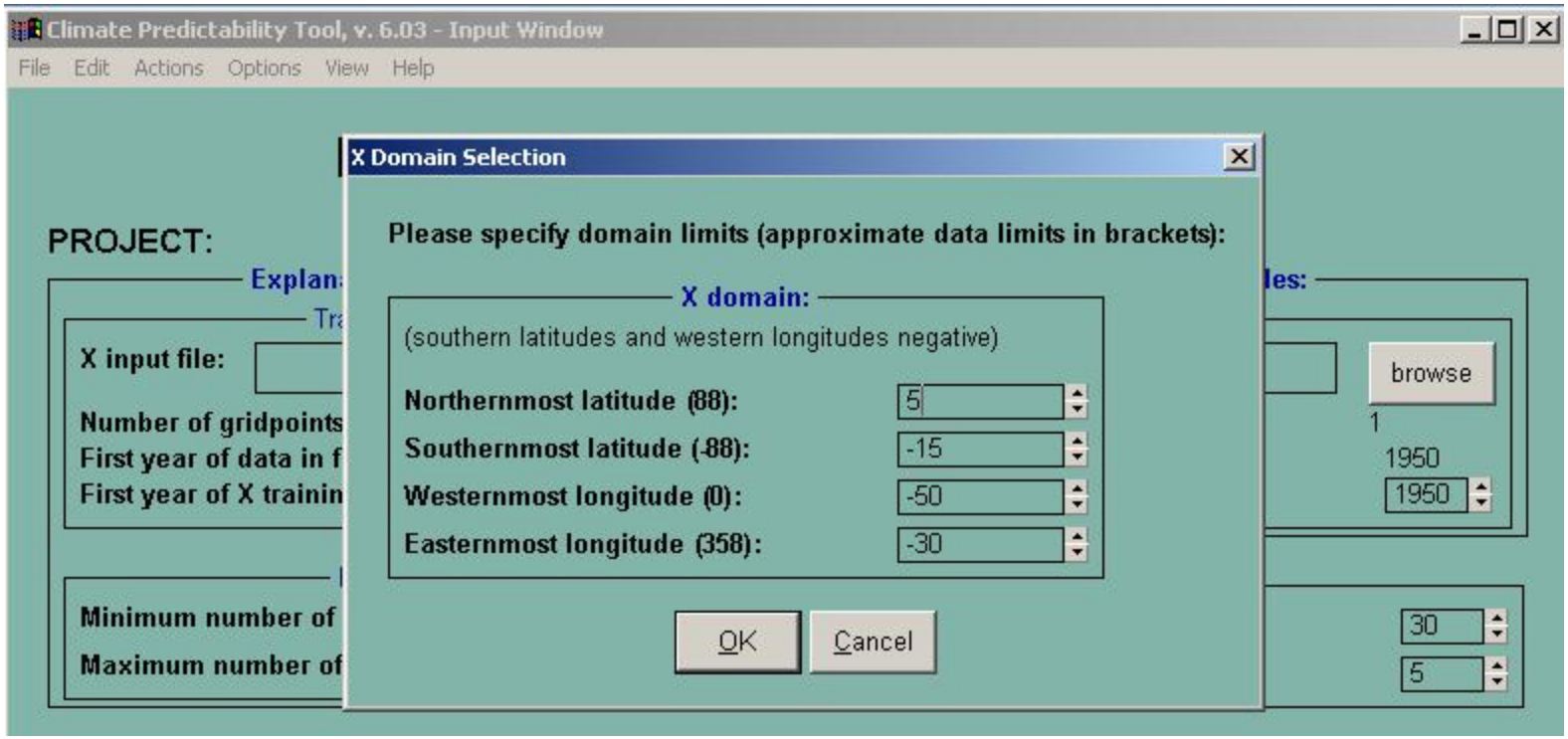

To select input files just click on browse.

# SELECTING INPUT FILES



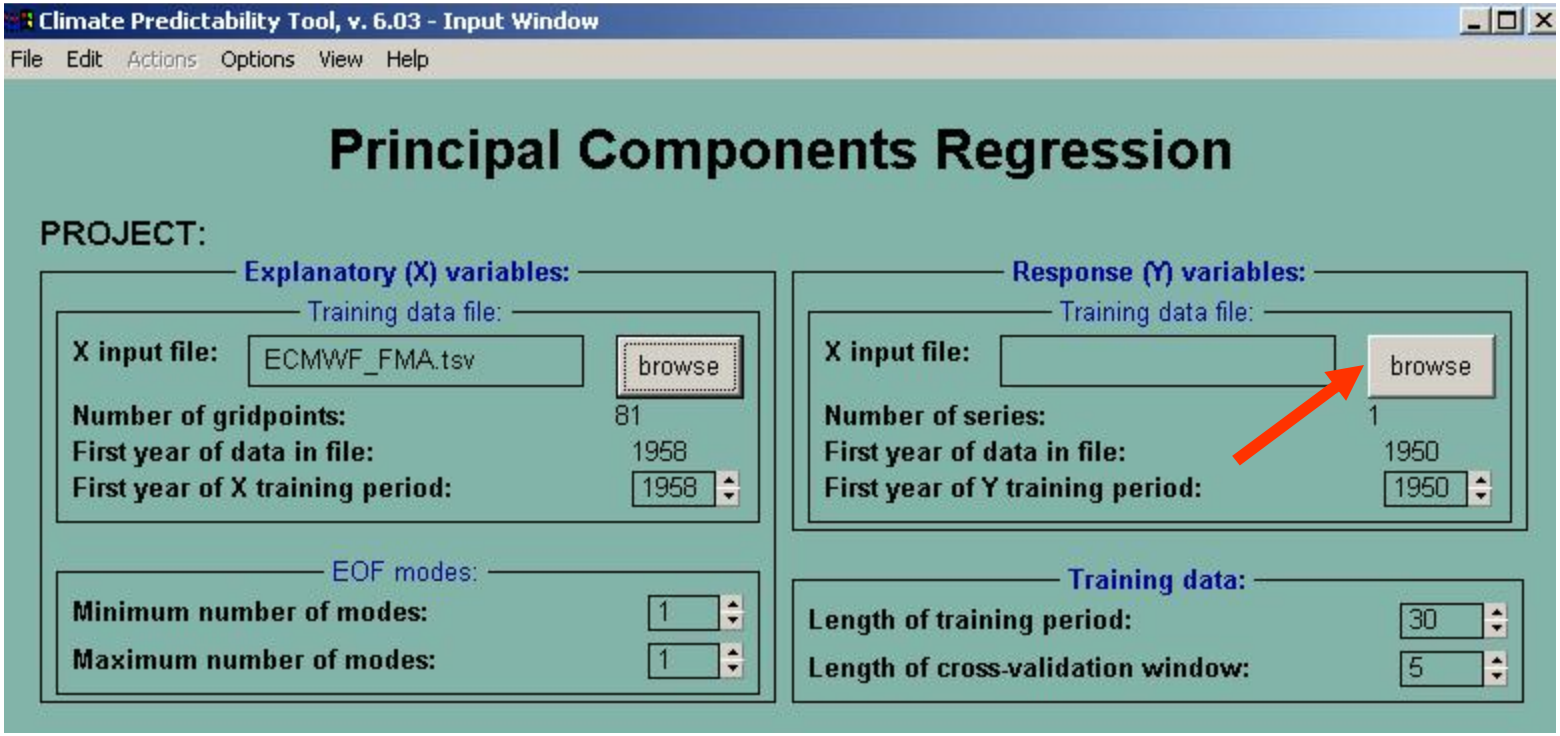

CPT opens a browser, which by default looks for data in:
C:\Documents and Settings*user*\Application Data\CPT\DATA\
You can search for data from any other directory.

# SELECTING INPUT FILES



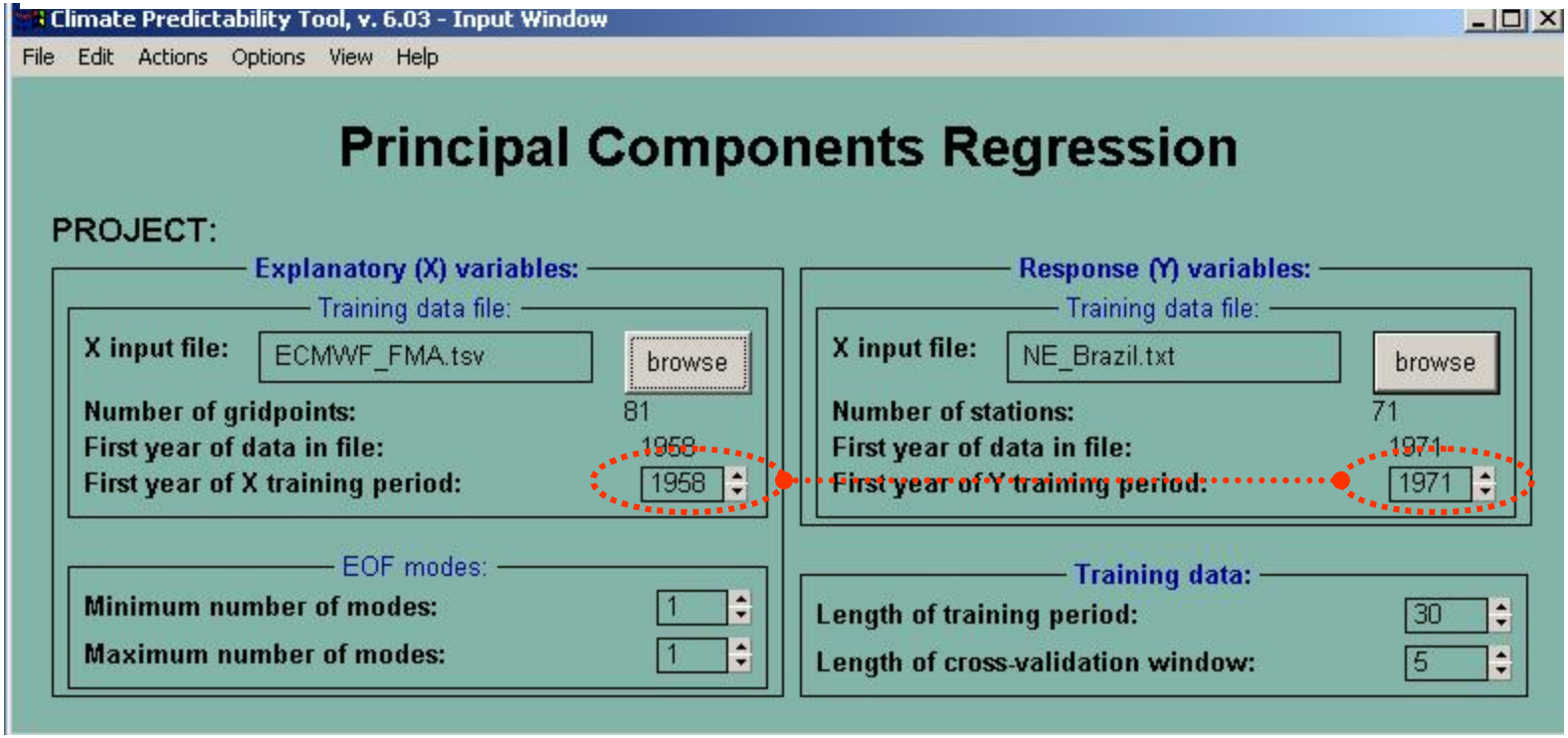

For gridded and station datasets, CPT lets you choose the spatial domain over which you want to perform your EOF or CCA analysis. In general the domain is known in advance through experience.

# SELECTING INPUT FILES




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

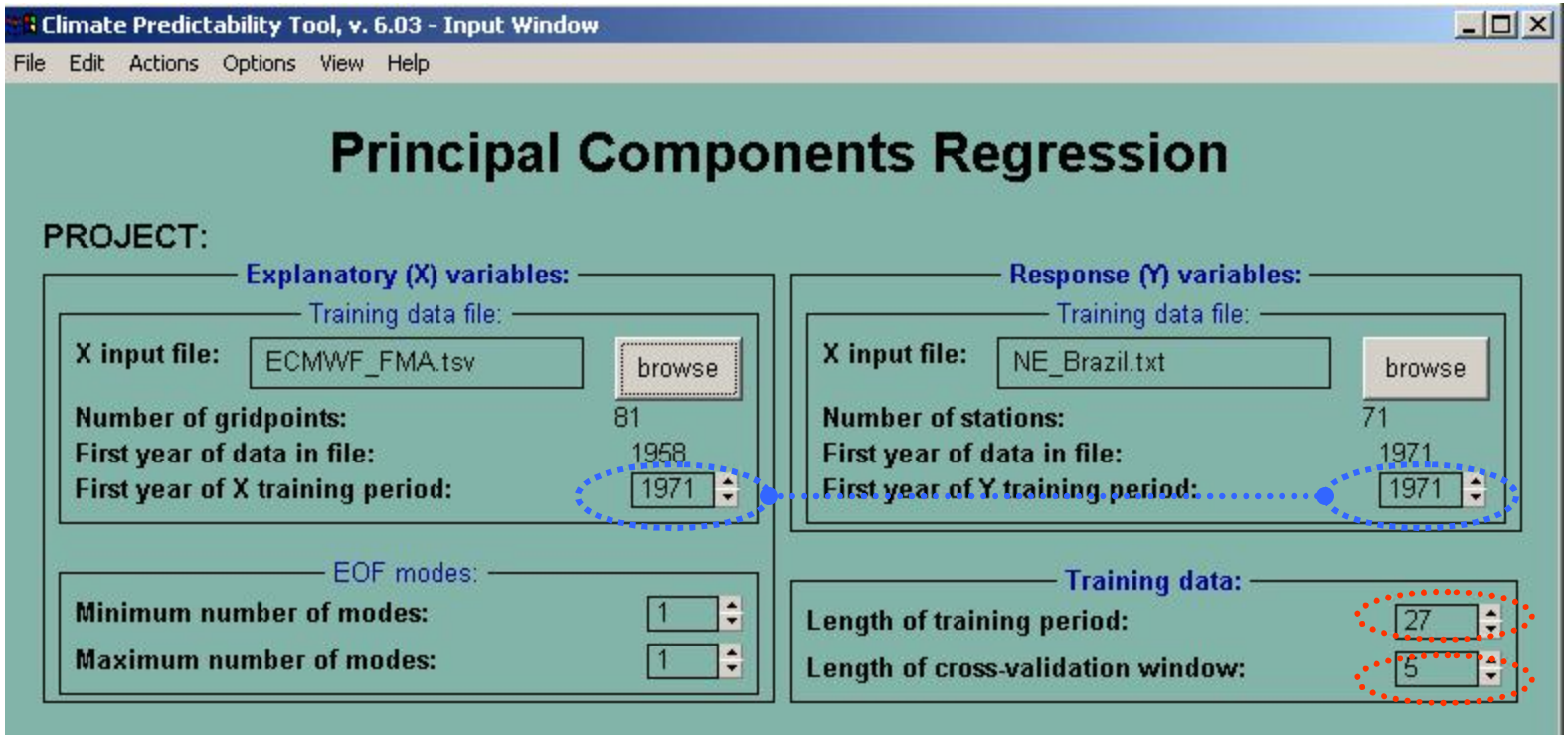
# SETTING THE TRAINING PERIOD




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 year in the two files.

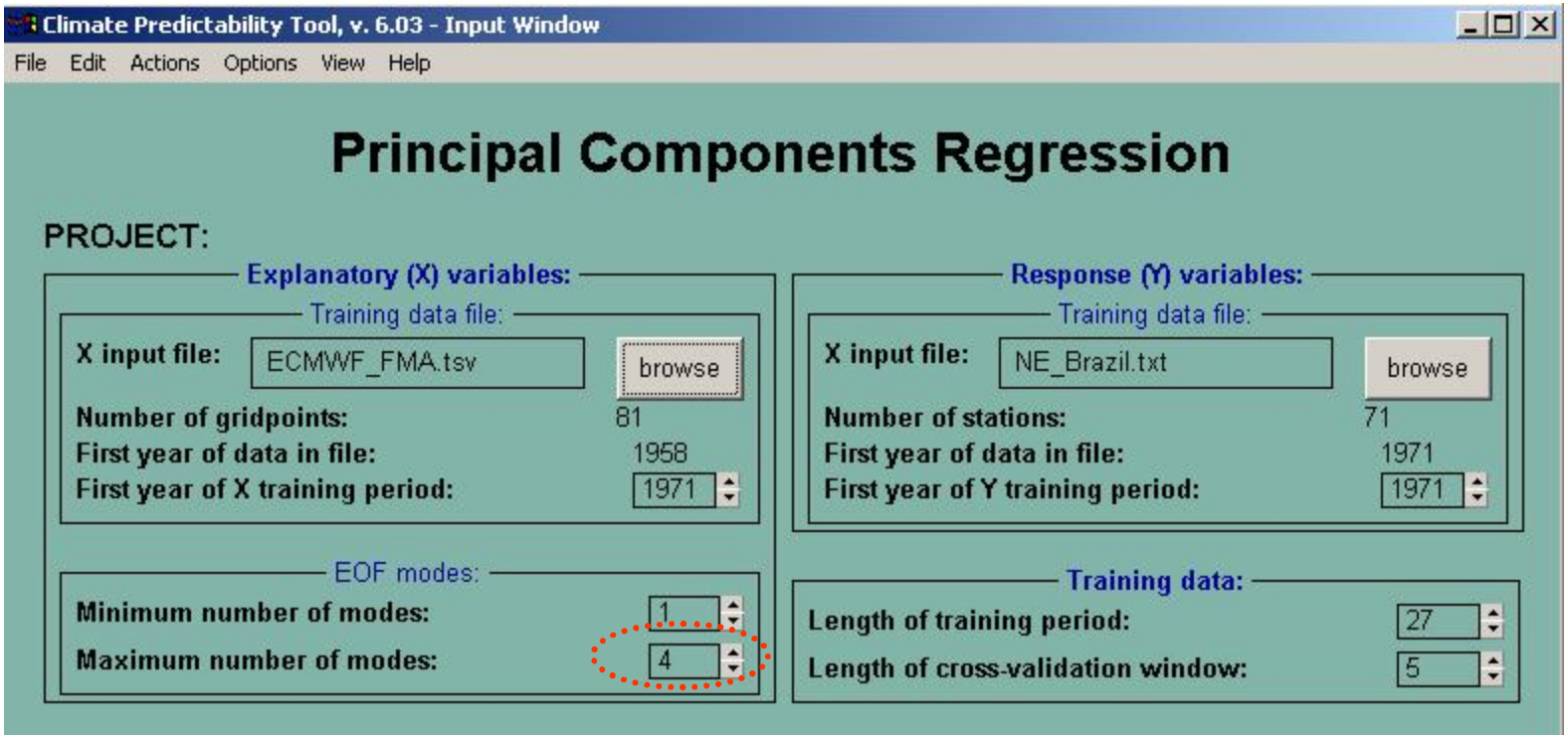
You should make sure the lag is correct if you cross the calendar year while using the DJF or JFM season, for example. In this case the starting year for file X may need to be one year earlier than for file Y.

# SETTING THE TRAINING PERIOD



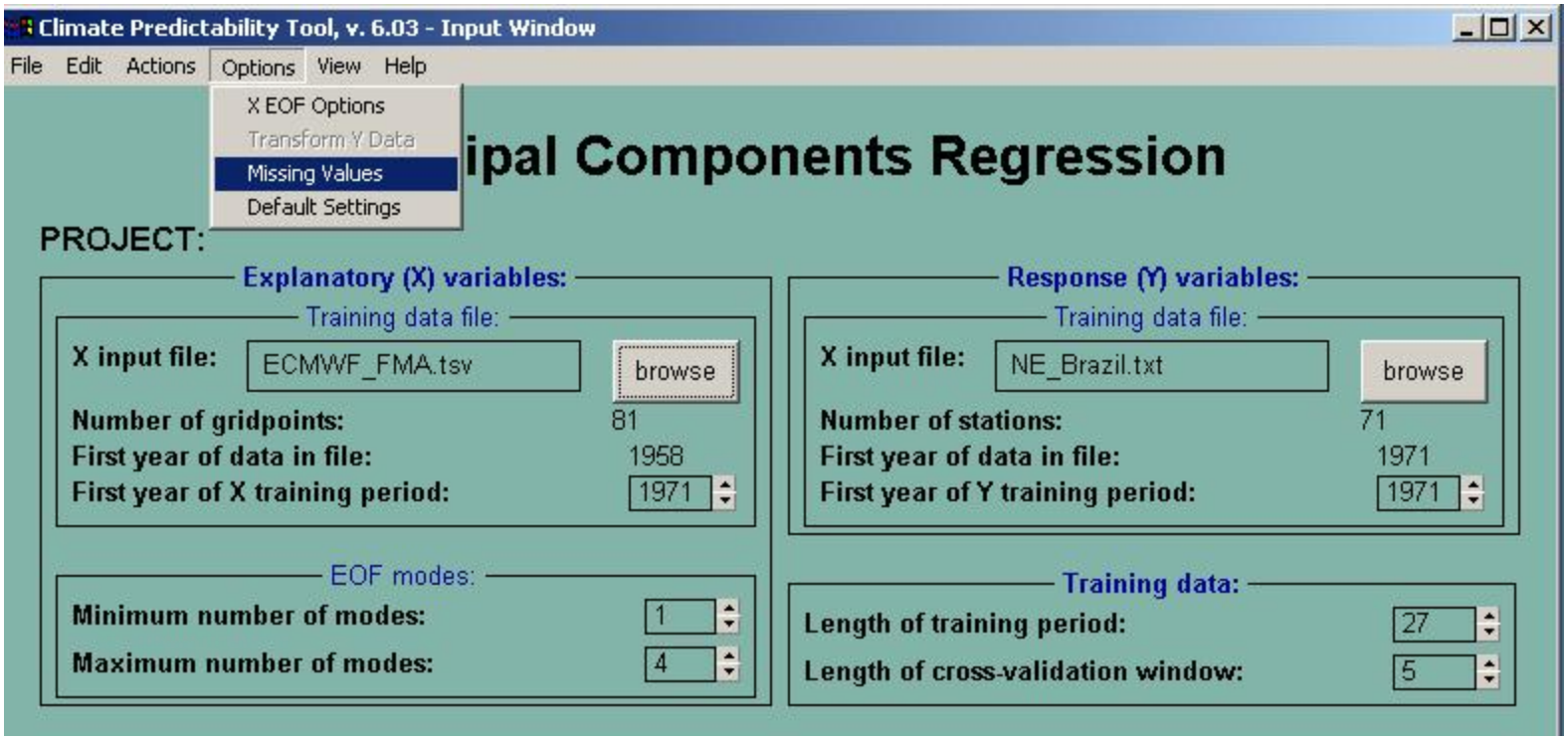

You have to specify the length of the training period as well as the length of the cross-validation window.

# SETTING ANALYSIS OPTIONS



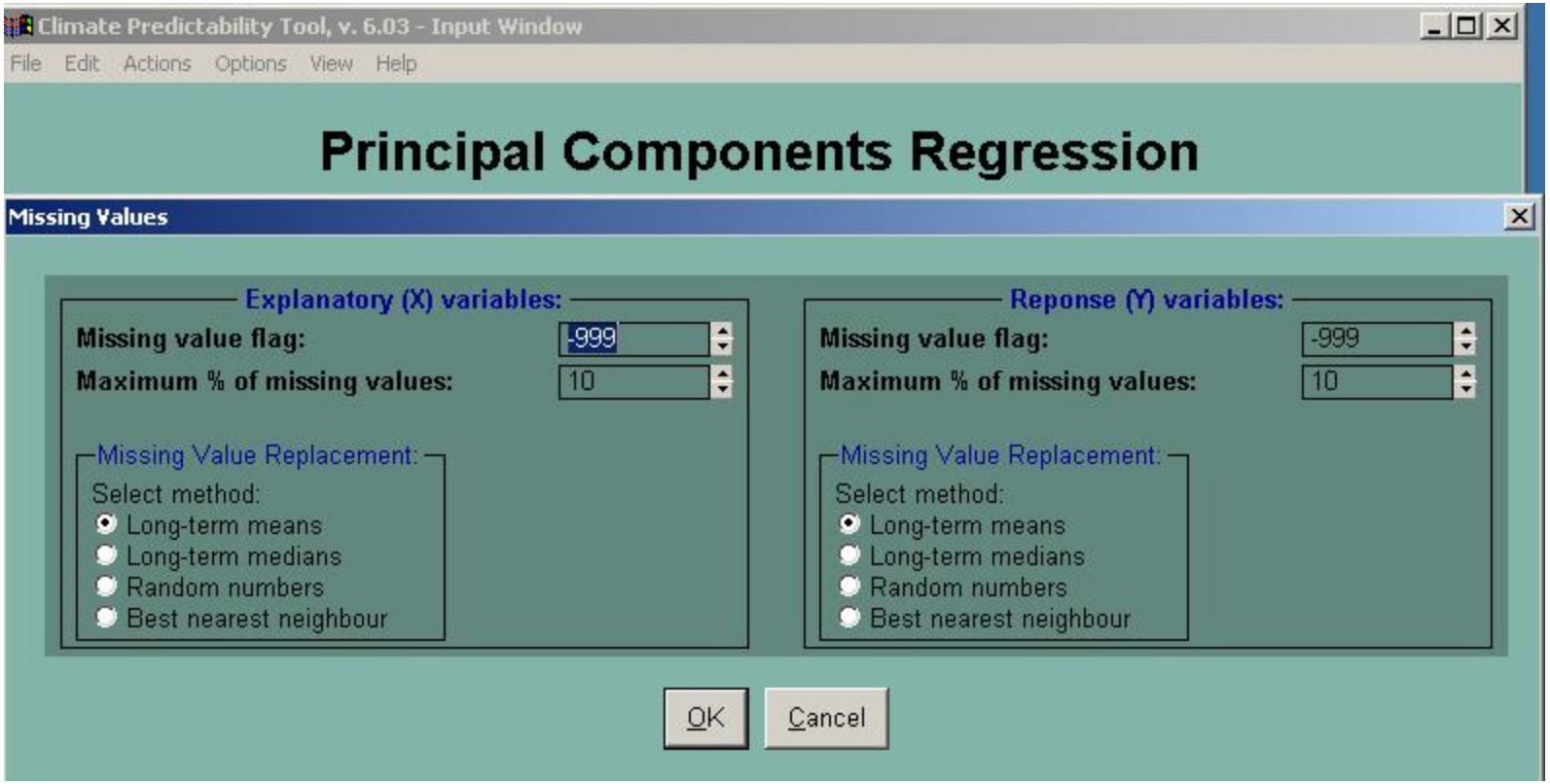

You have to choose the number of EOFs for the predictor fields used to fit the model. If you set the minimum to be less than the maximum, CPT will find the optimum number of modes between the two numbers. However, if you set the minimum equal to the maximum, then CPT will use that number of modes.

# MISSING VALUES



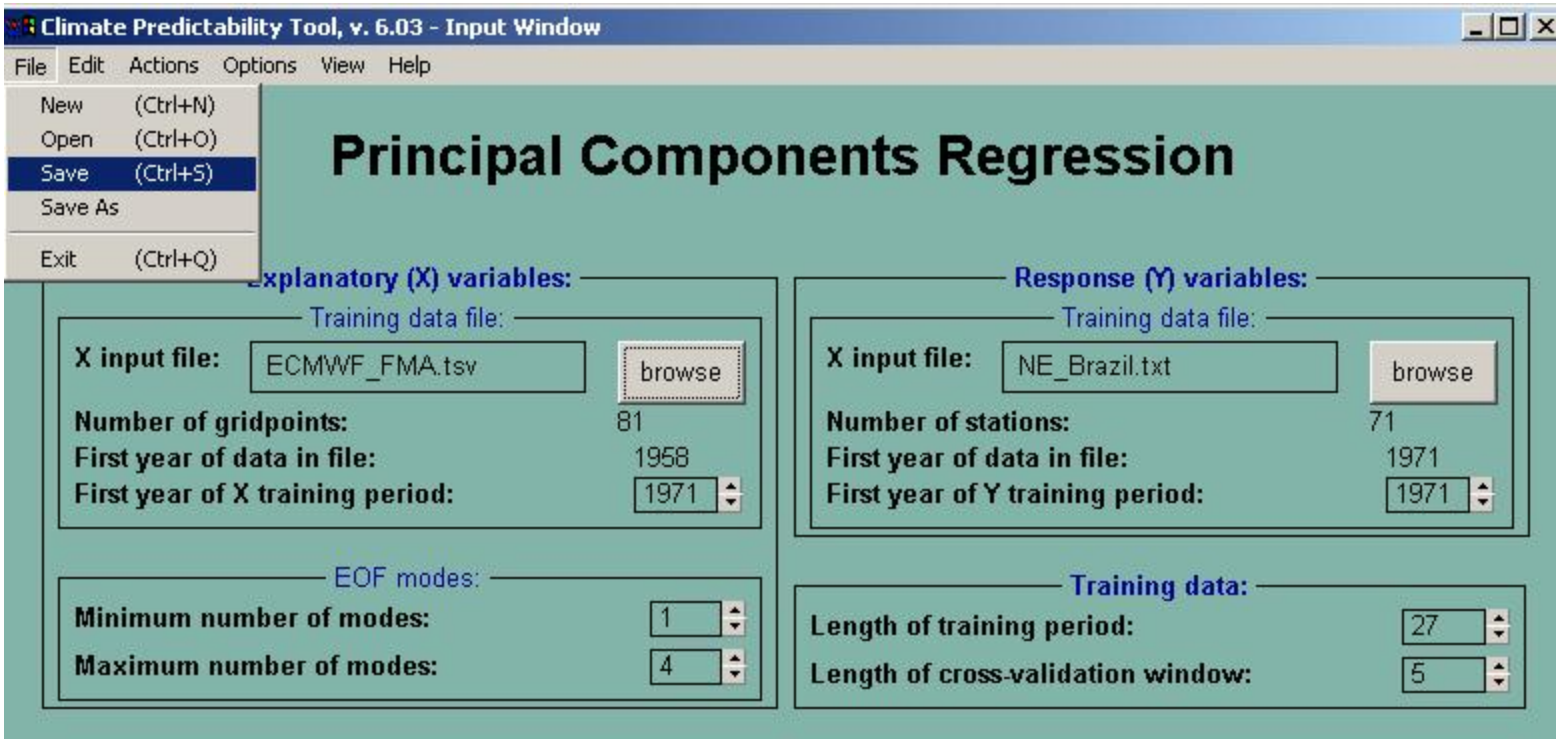

If you have missing values in your dataset, you need to specify what you want CPT to do with them.

# 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 also choose which method you want CPT to use to replace the values.

# SAVING PROGRAM SETTINGS

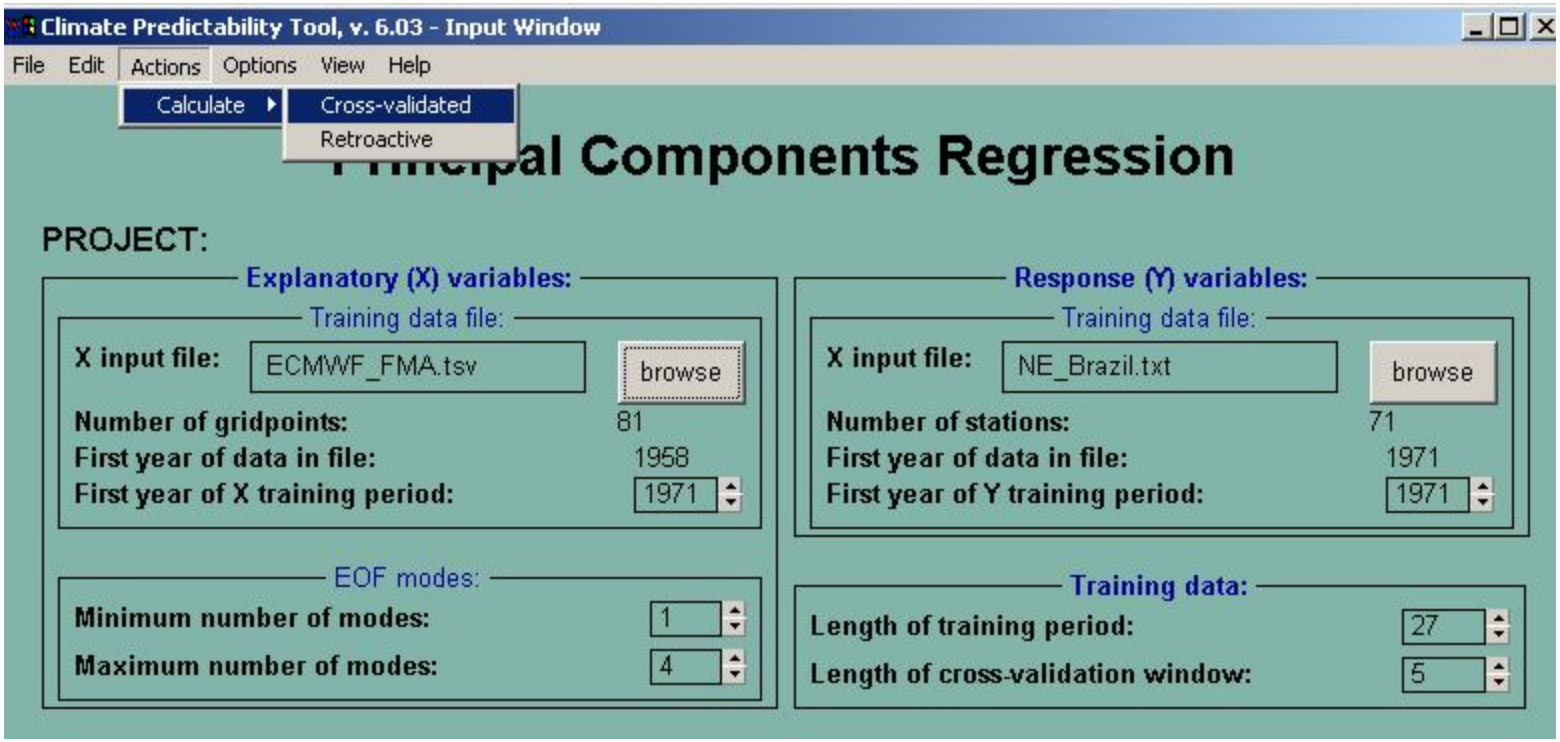



Once you have selected the input files and your settings it is a good idea to save these settings in a project file to recall them later:

File => Save

By default, CPT saves all the project files in the subdirectory C:\Documents and Settings*user*\Application Data\CPT\Projects\

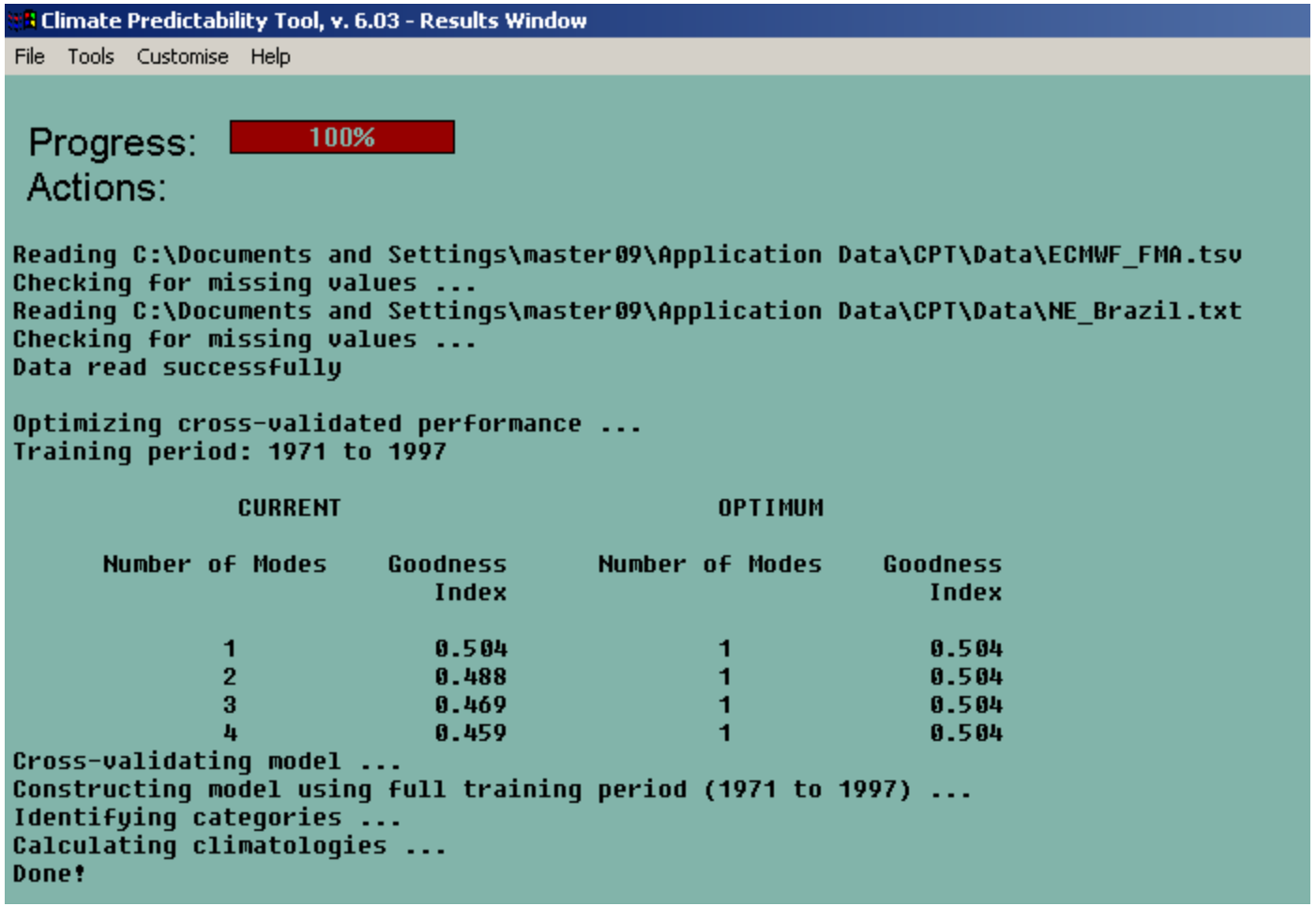
# RUNNING CPT




Then you can run the analysis:

Actions => Calculate => Cross-validated

# DATA ANALYSIS




CPT begins the specified analysis in a new "Results Window". Here you can see the steps of the analysis and of the optimization procedure.

# DATA ANALYSIS

| CURRENT | | OPTIMUM | |
|---|---|---|---|
| Number of Modes | Goodness Index | Number of Modes | Goodness Index |
| 1 | 0.504 | 1 | 0.504 |
| 2 | 0.488 | 1 | 0.504 |
| 3 | 0.469 | 1 | 0.504 |
| 4 | 0.459 | 1 | 0.504 |

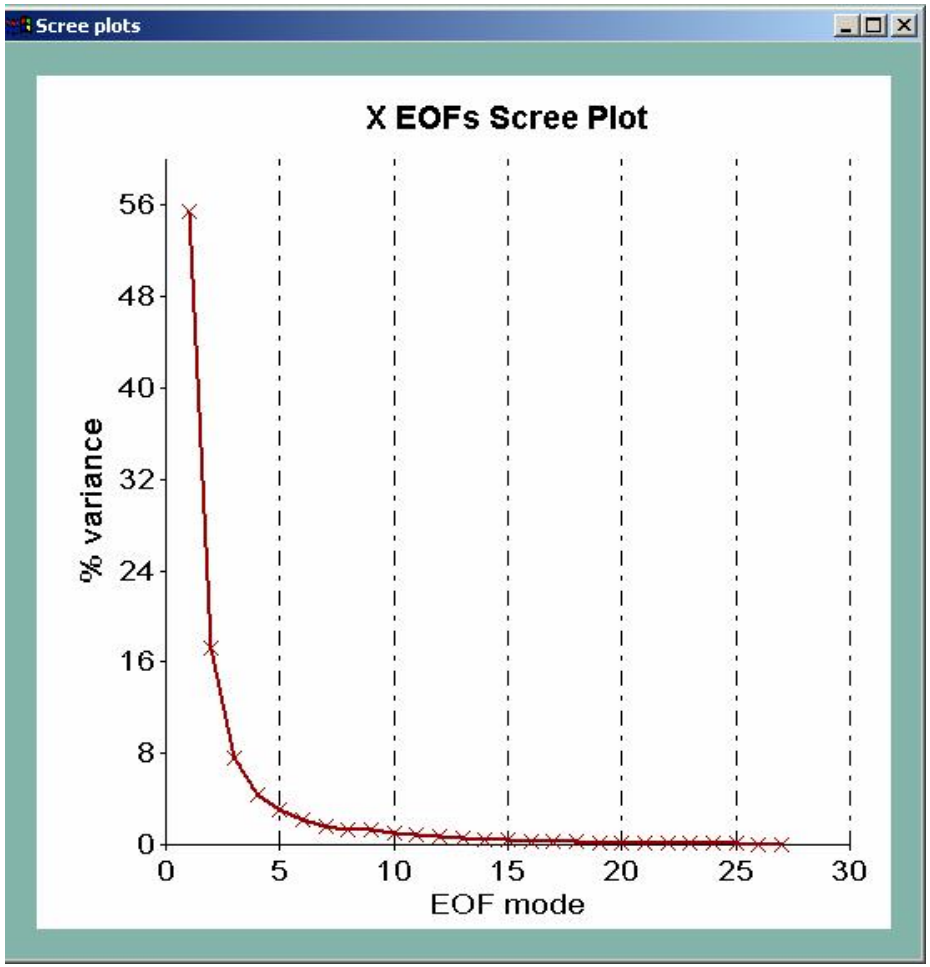
**Optimizing the number of EOF modes:**

**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 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 using all four EOFs.

**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, 1).

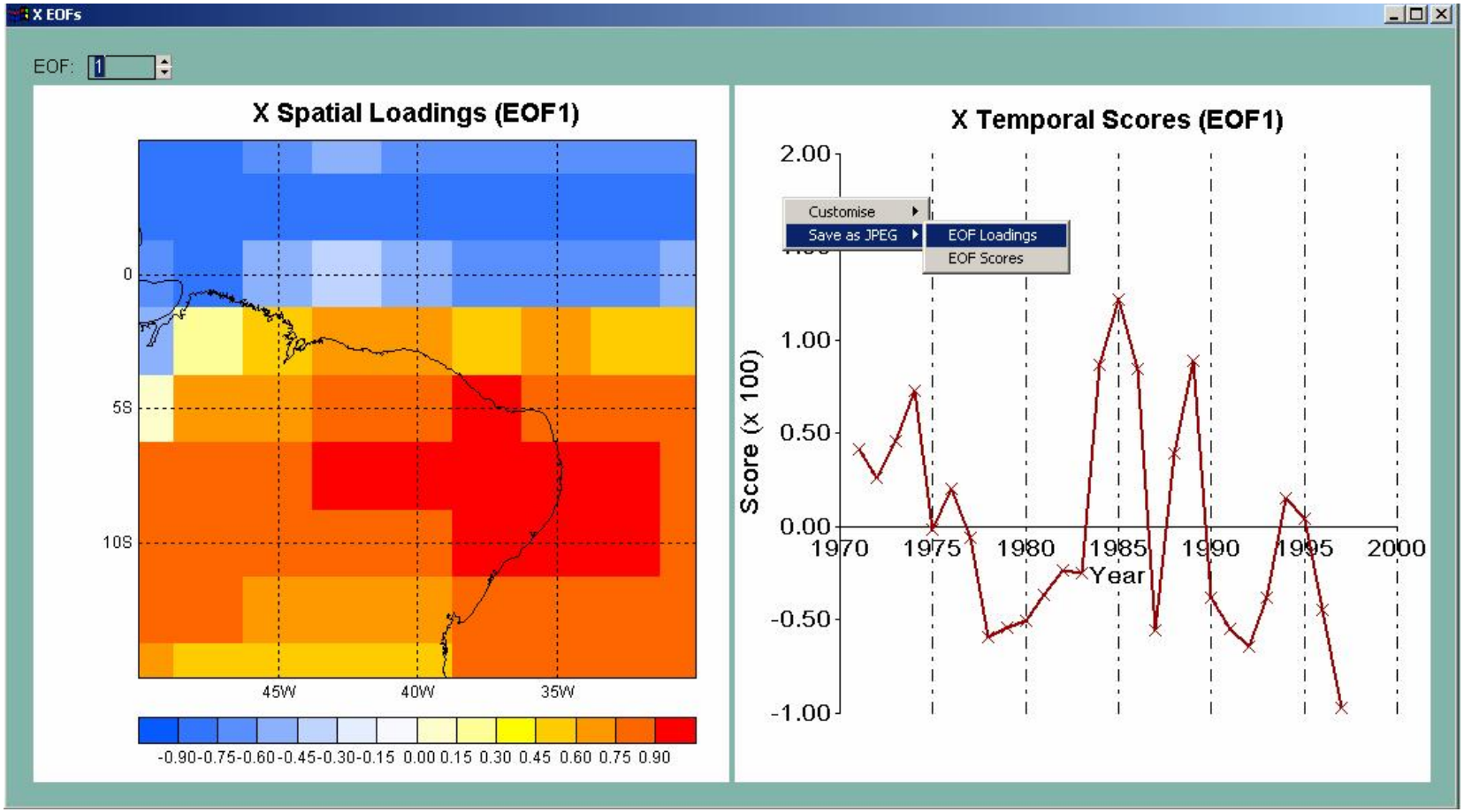
**3.** CPT uses this number of EOFs (i.e. 1) to build the model.

# RESULTS : graphics




The menu Tools => Graphics => Scree plots displays the percentage of variance associated with each EOF plotted.

# RESULTS : graphics



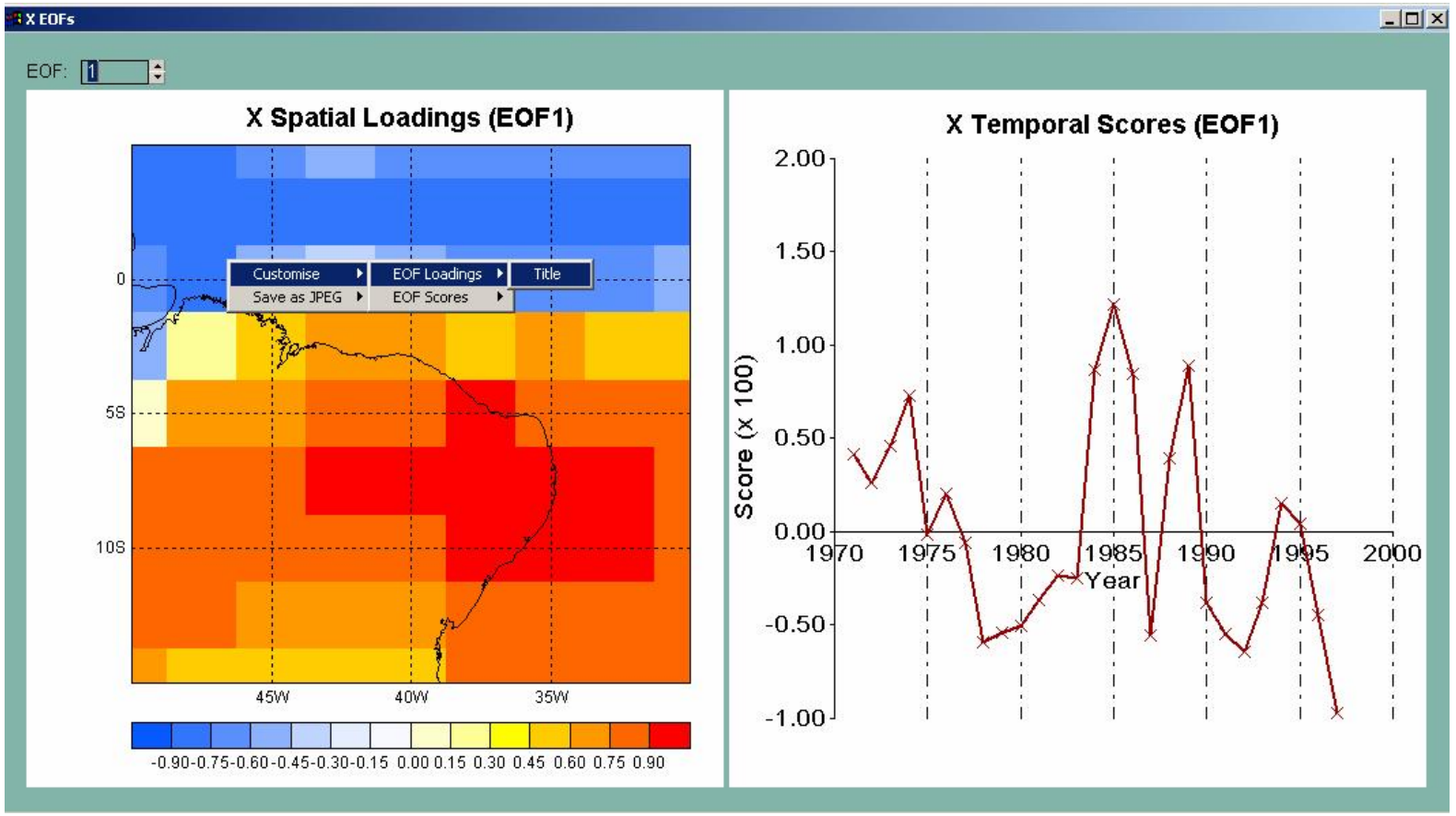

**1.** The menu Tools => Graphics => X EOF loadings and scores displays the loading pattern of each EOF and the temporal series.

**2.** CPT allows you to customize and save each graphic by:

right-clicking on the mouse

selecting the graphic to customize / save

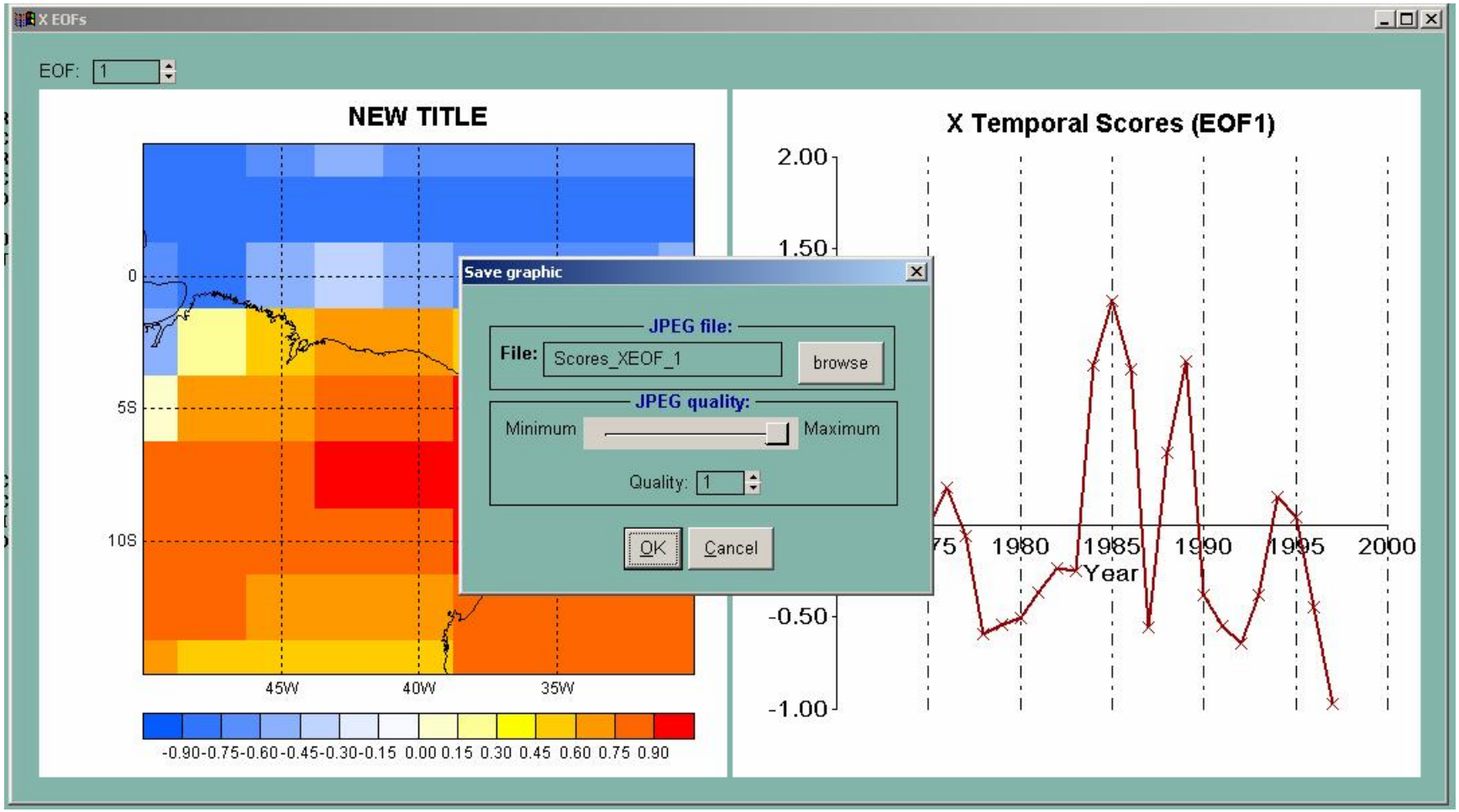
# CHANGING THE TITLE




To change the title of the graph

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

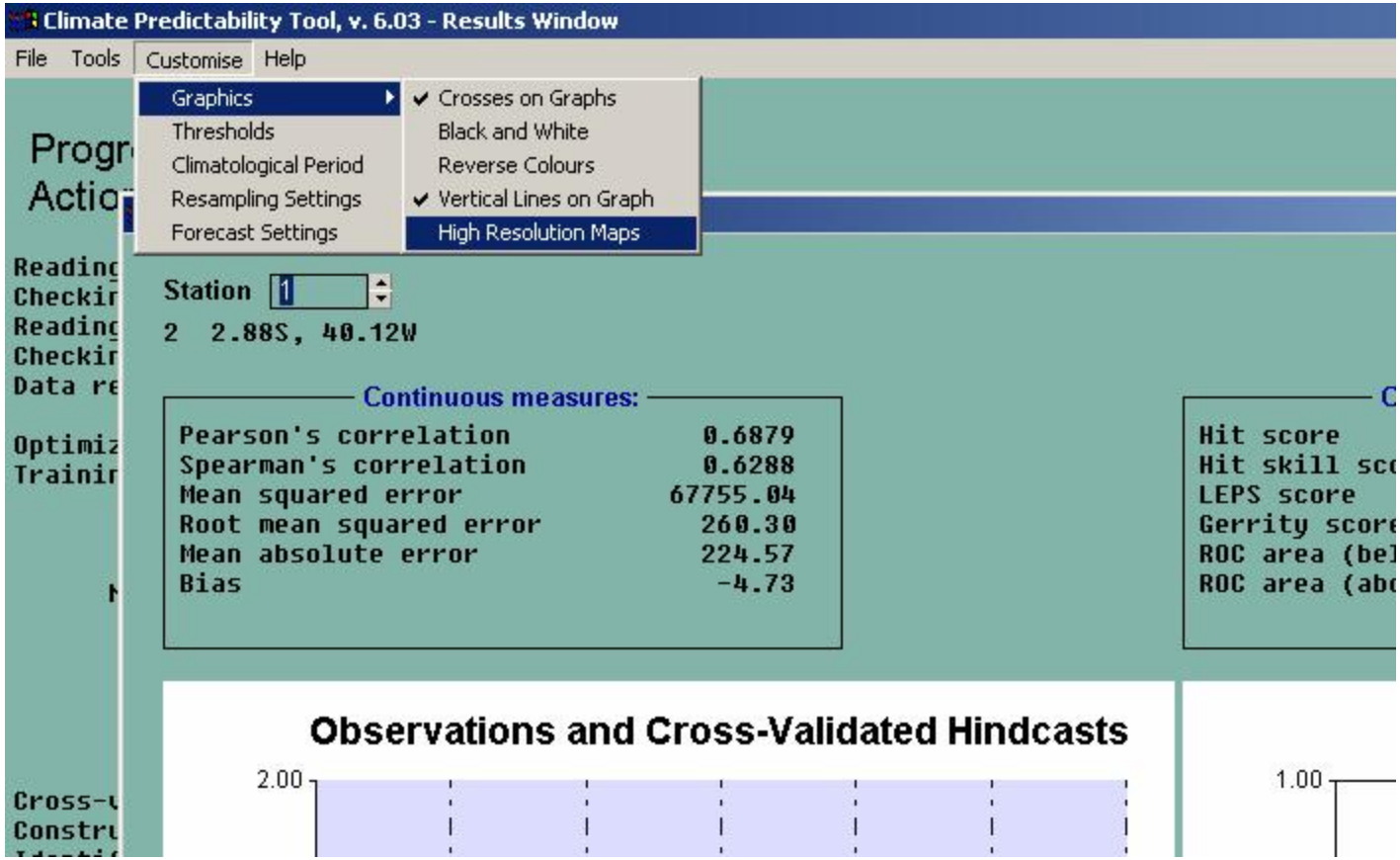
# SAVING GRAPHICS




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:\CPT\Output\

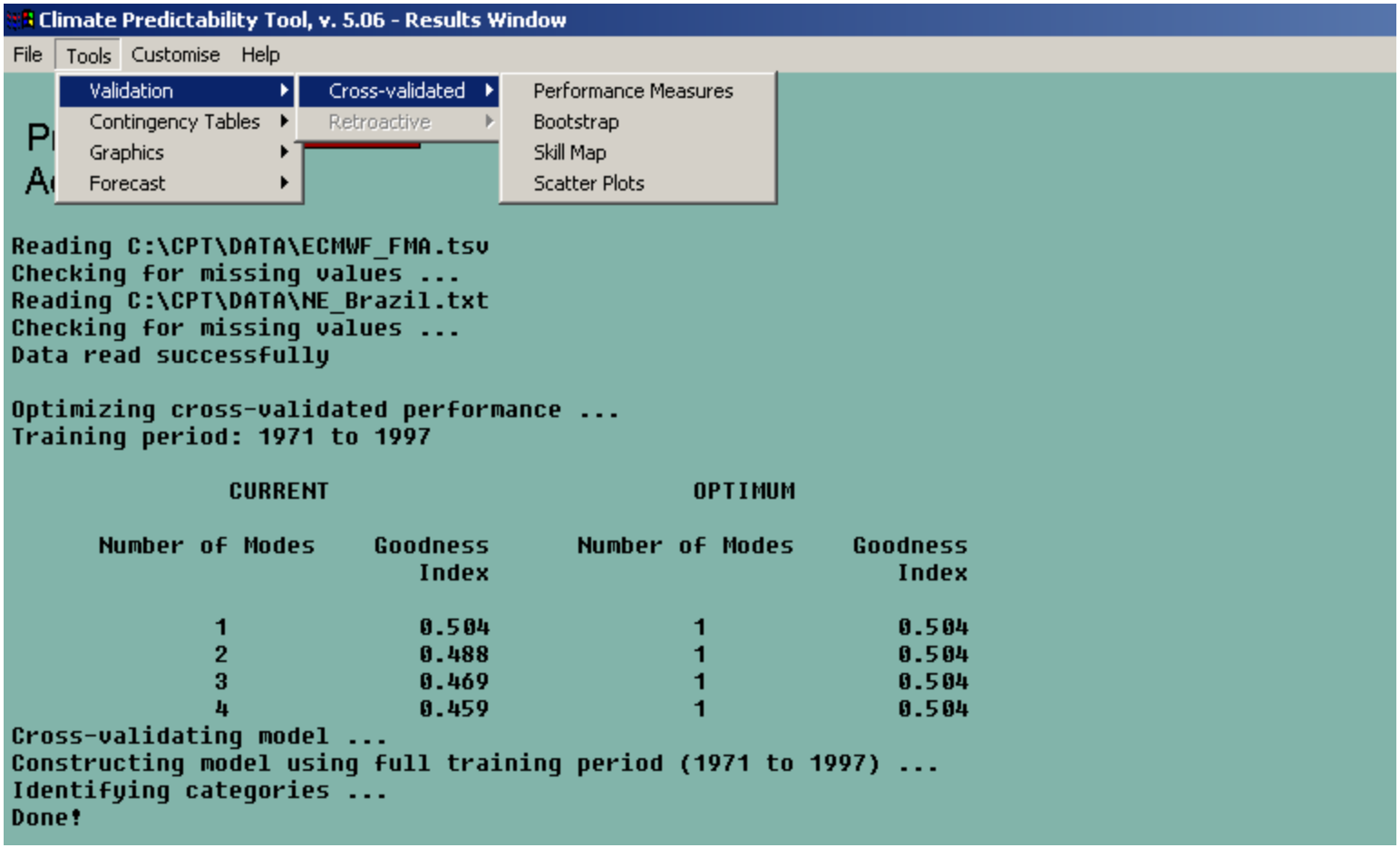
# SHOWING HIGH RESOLUTION MAPS




If you want to get a better quality map, you can change the setting to high resolution.

Customize => Graphics => High Resolution Map

# RESULTS



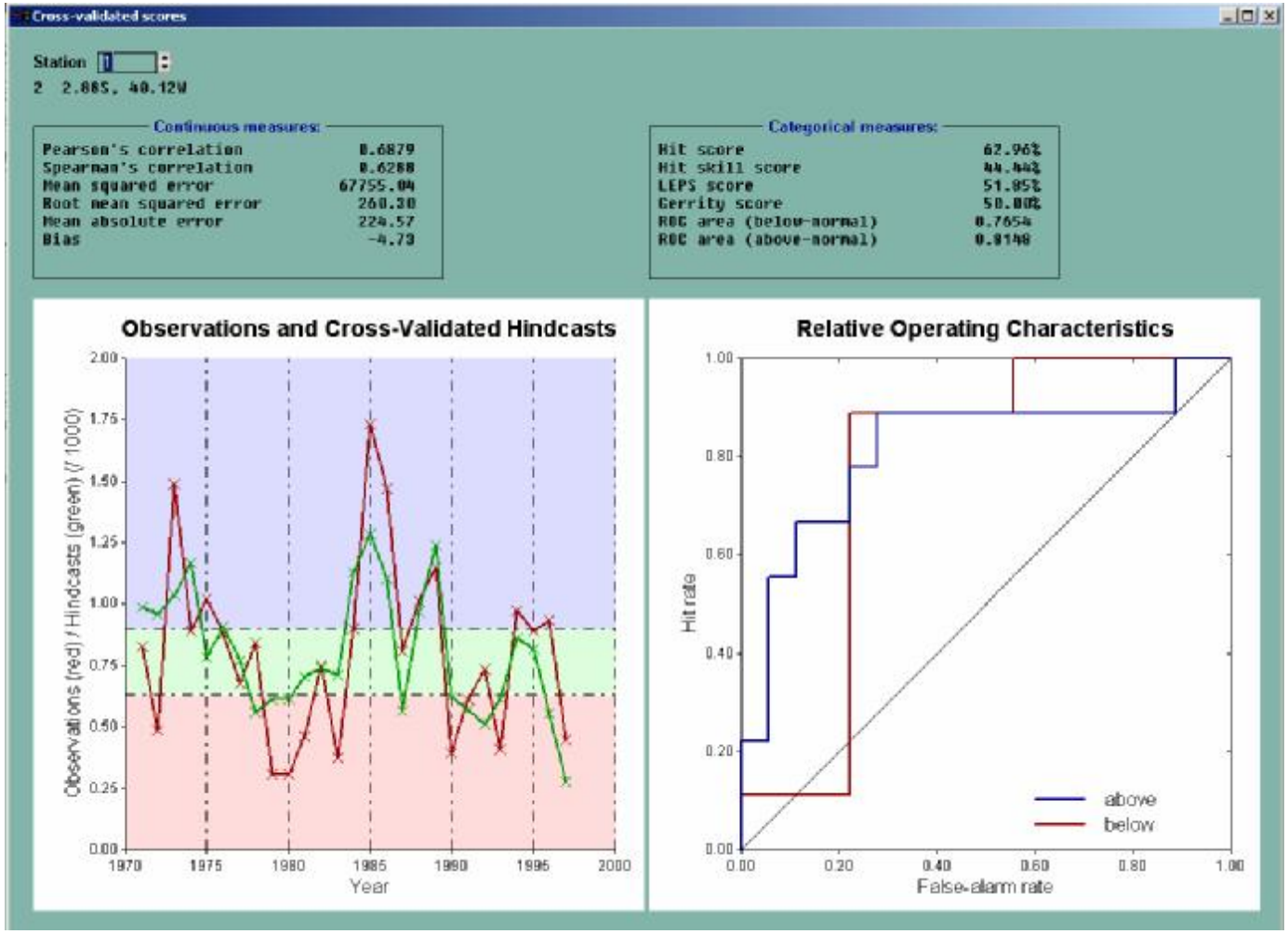

To see the results go to the menu “Tools”:

Validation : shows skill, hindcasts and observed series

Contingency Tables : shows contingency tables

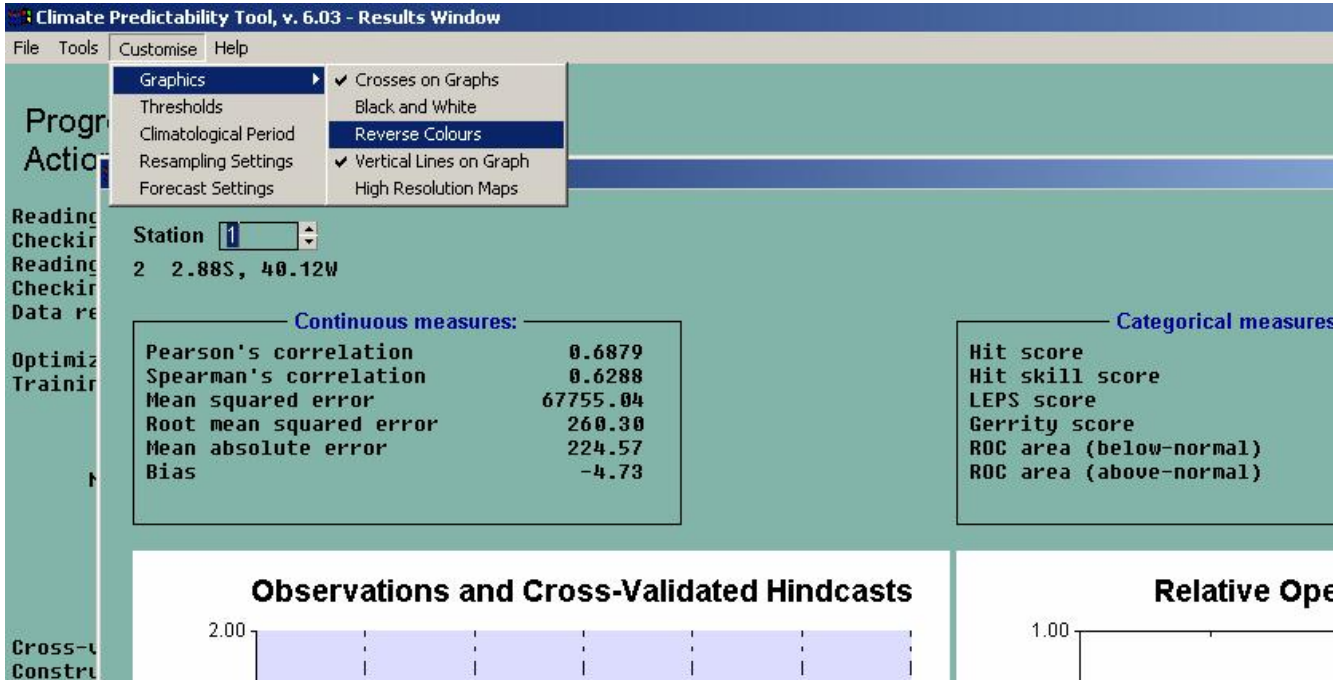
Graphics : shows the EOFs time series, loading patterns and scree plot

# RESULTS




To see the series forecasted and observed at each station/grid go to:
Tools => Validation => Cross-Validated => Performance Measures
This menu displays some statistics of the forecast, such as correlation coefficient, RMSE, ROC etc (for more details refer to the help page).

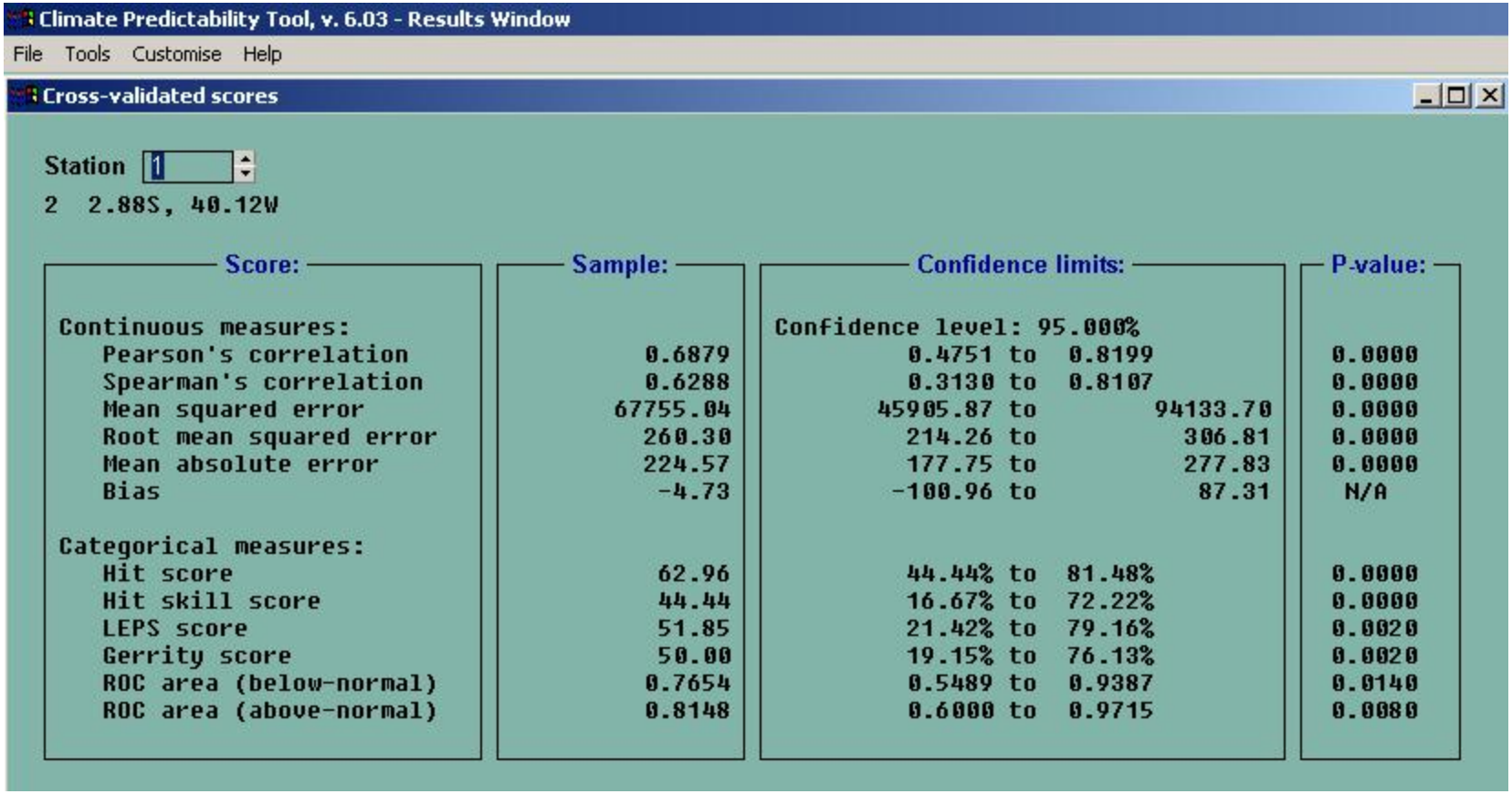
# REVERSING THE COLORS




Customise => Graphics => Reverse Colors

If you are forecasting temperature instead of precipitation, then it would be more intuitive to have red (hot/above) and blue (cold/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 UNCERTAINTY

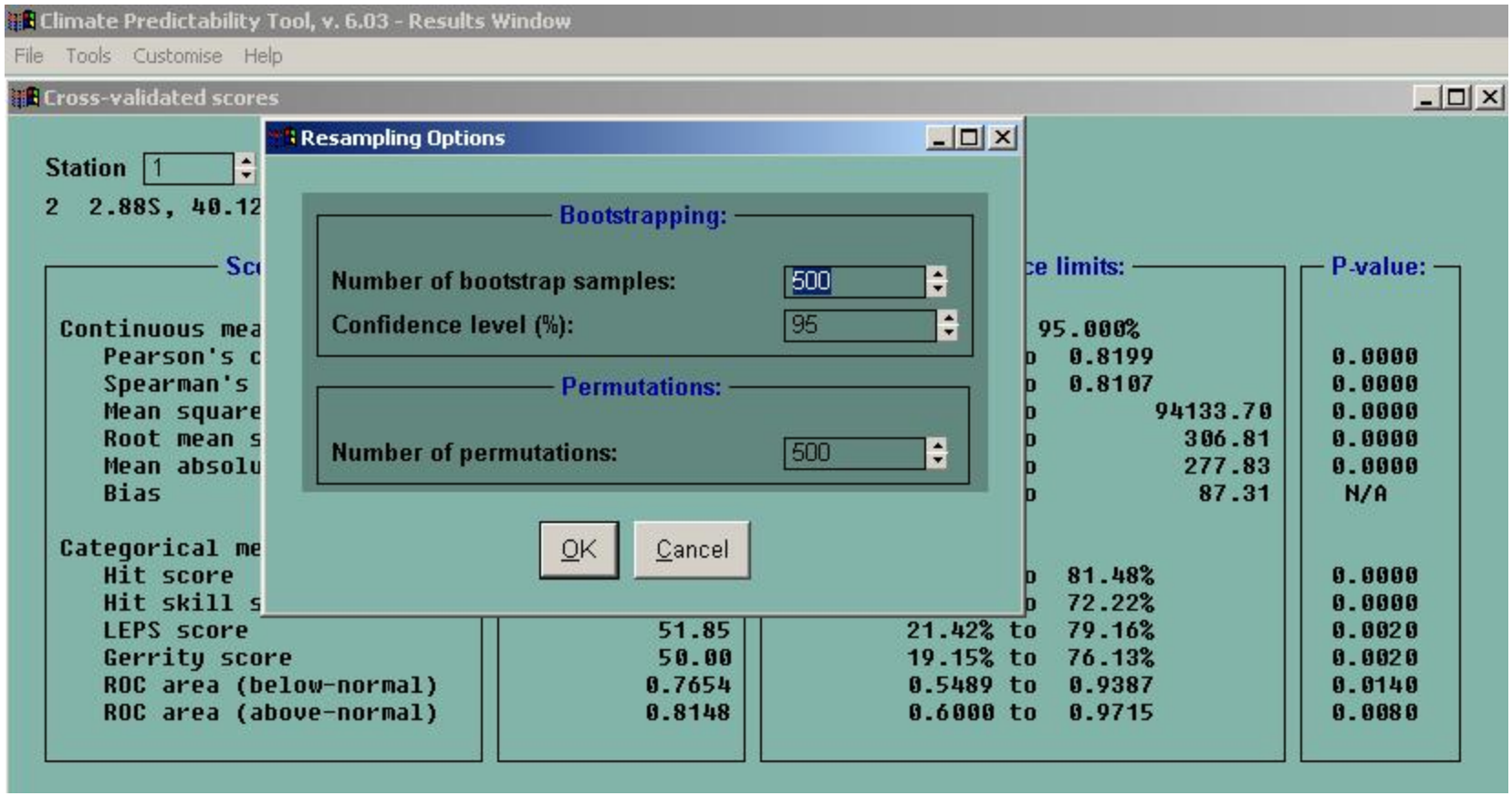



| Score: | Sample: | Confidence limits: (Confidence level: 95.000%) | P-value: |
|---|---|---|---|
| **Continuous measures:** | | | |
| Pearson's correlation | 0.6879 | 0.4751 to 0.8199 | 0.0000 |
| Spearman's correlation | 0.6288 | 0.3130 to 0.8107 | 0.0000 |
| Mean squared error | 67755.04 | 45905.87 to 94133.70 | 0.0000 |
| Root mean squared error | 260.30 | 214.26 to 306.81 | 0.0000 |
| Mean absolute error | 224.57 | 177.75 to 277.83 | 0.0000 |
| Bias | -4.73 | -100.96 to 87.31 | N/A |
| **Categorical measures:** | | | |
| Hit score | 62.96 | 44.44% to 81.48% | 0.0000 |
| Hit skill score | 44.44 | 16.67% to 72.22% | 0.0000 |
| LEPS score | 51.85 | 21.42% to 79.16% | 0.0020 |
| Gerrity score | 50.00 | 19.15% to 76.13% | 0.0020 |
| ROC area (below-normal) | 0.7654 | 0.5489 to 0.9387 | 0.0140 |
| ROC area (above-normal) | 0.8148 | 0.6000 to 0.9715 | 0.0080 |

For indications of uncertainty in the performance measures go to:

Tools => Validation => Cross-Validated => Bootstrap

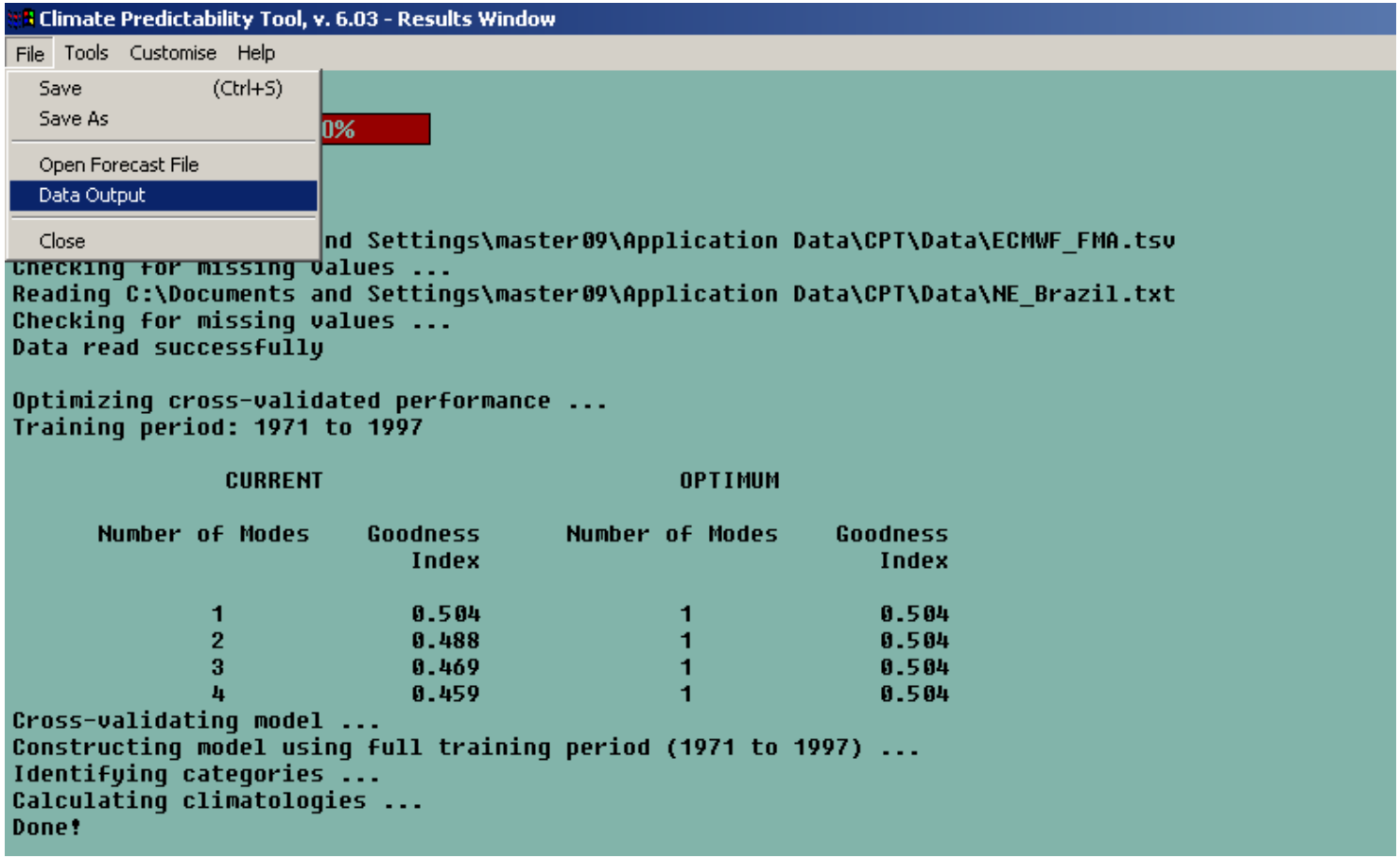
# ADJUSTING THE BOOTSTRAP SETTINGS




Customize => Resampling Settings

CPT allows you to adjust the bootstrap settings.

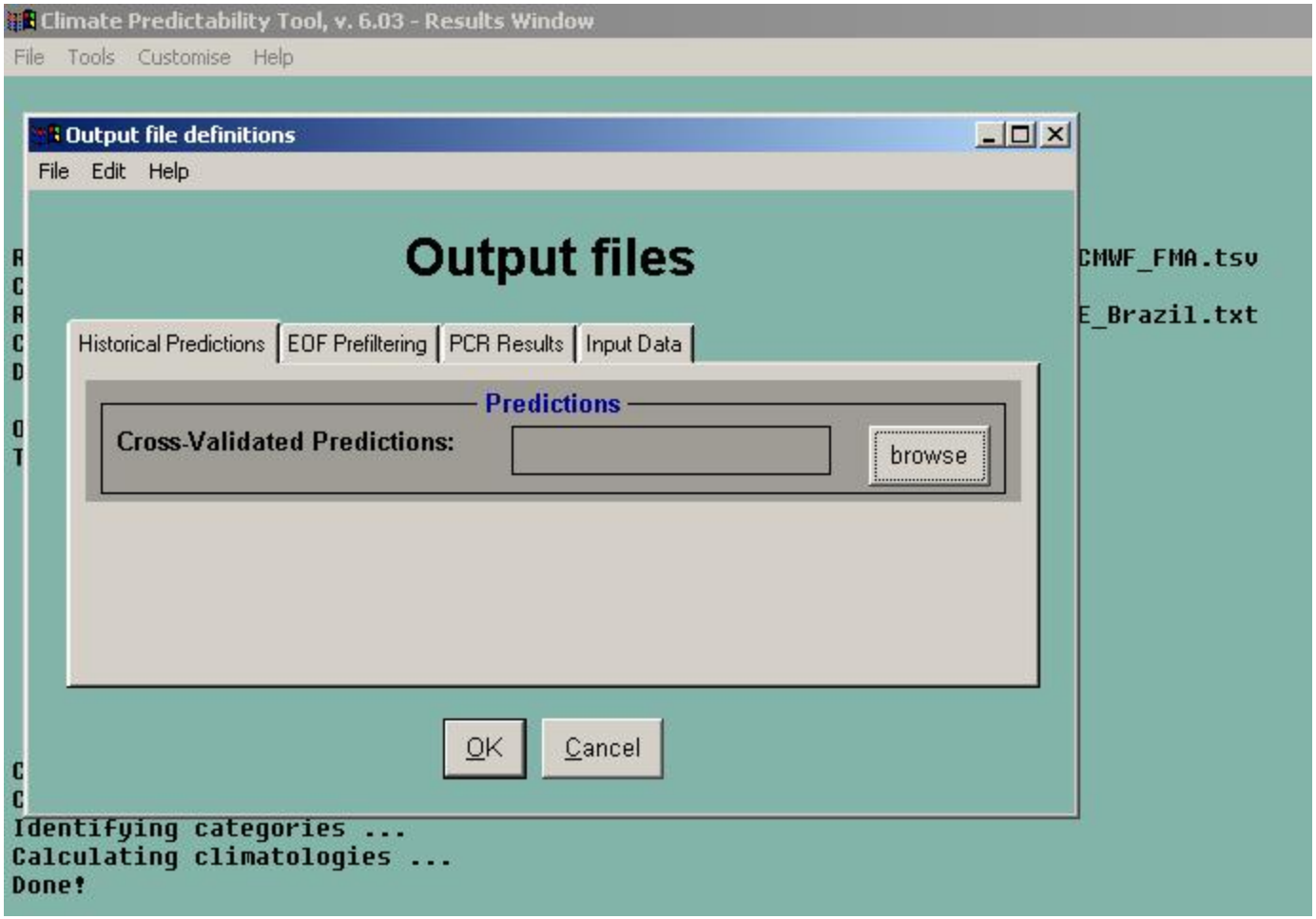
# RESULTS : data files




The menu File => Data Output allows you to save output data:

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

# SAVING OUTPUT FILES

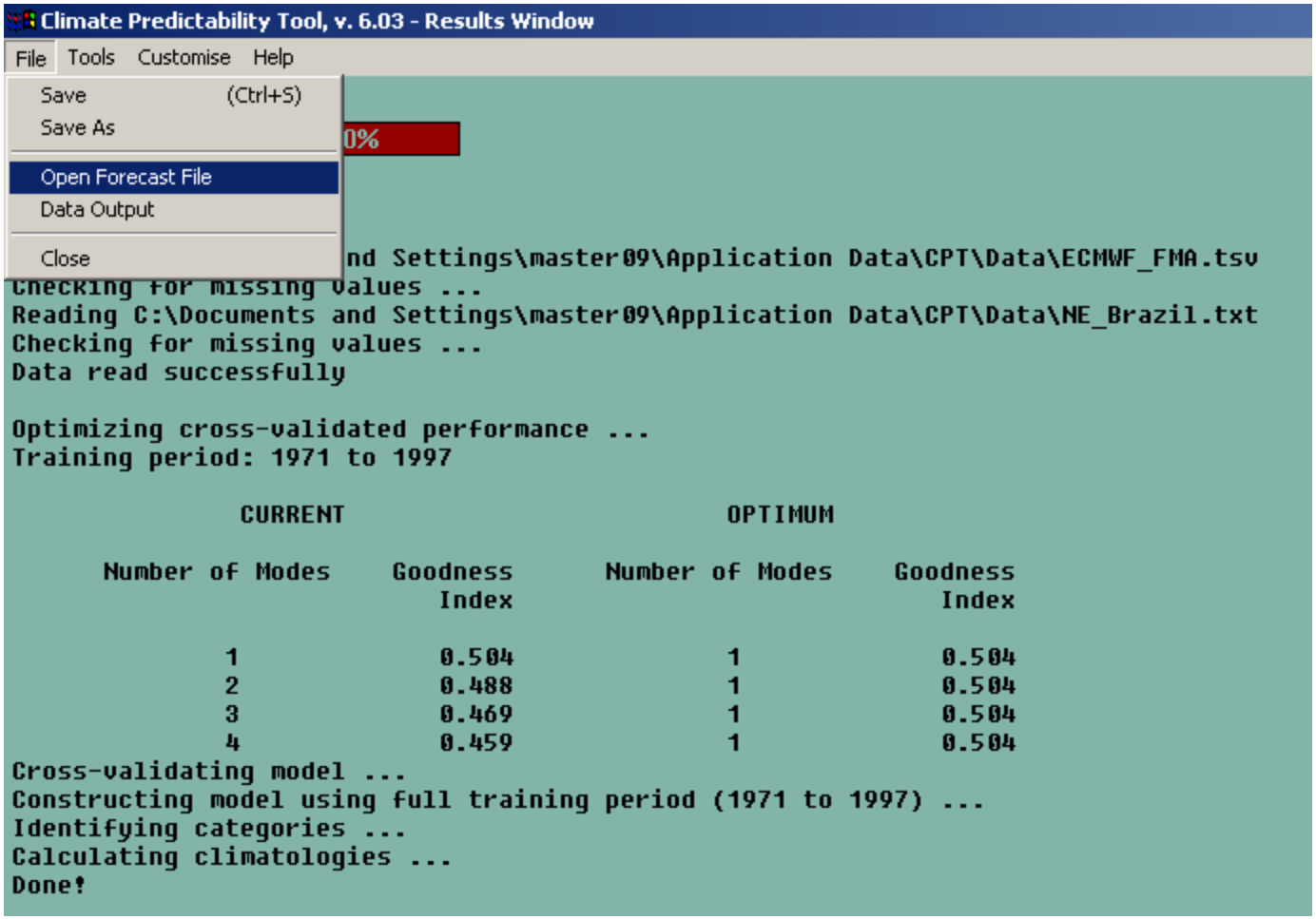



In order to save the outputs in separate files, you have to specify a file name by clicking on browse.

By default CPT saves the output files under:

C:\CPT\Output\

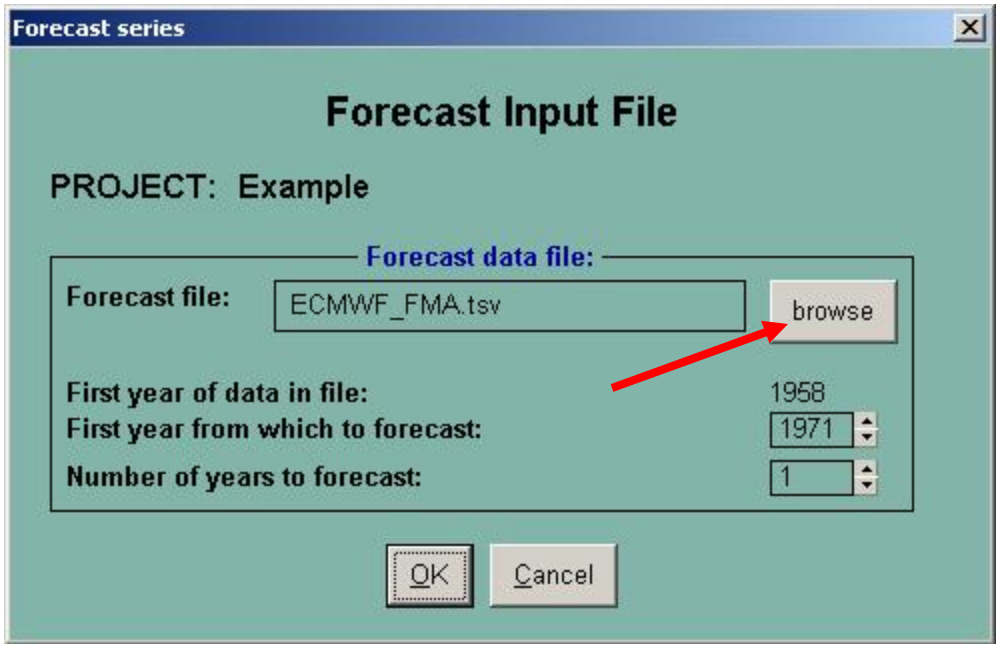
# FORECAST




Once your model is built, you can make a forecast using a forecast file with new records of the X variables:

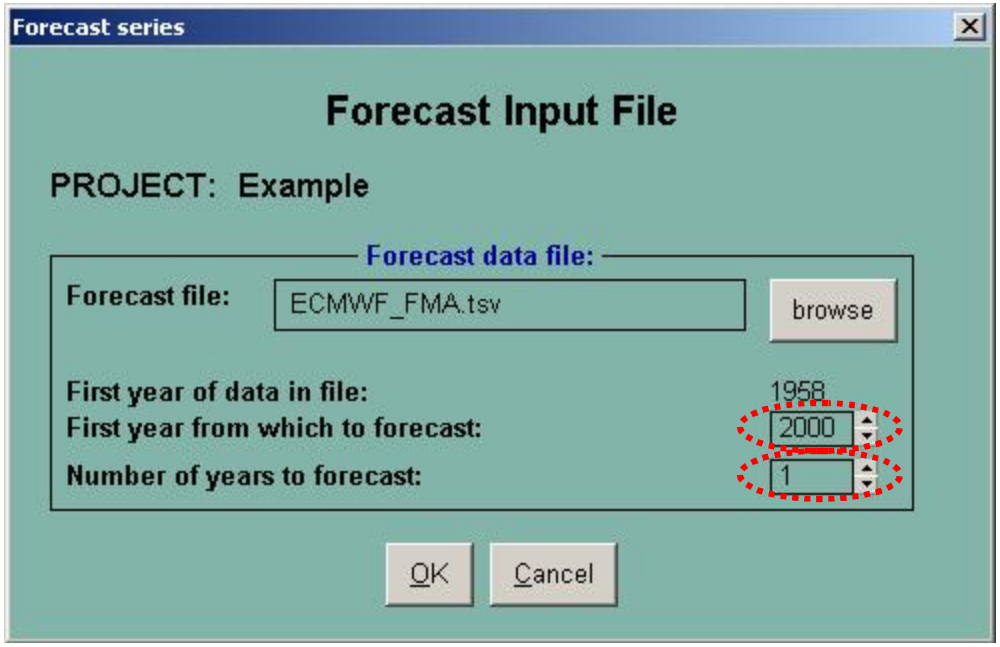
File => Open Forecast File

# FORECAST




A new window is opened.
By default CPT selects the same input predictor file.
You can change it by clicking browse.

# FORECAST

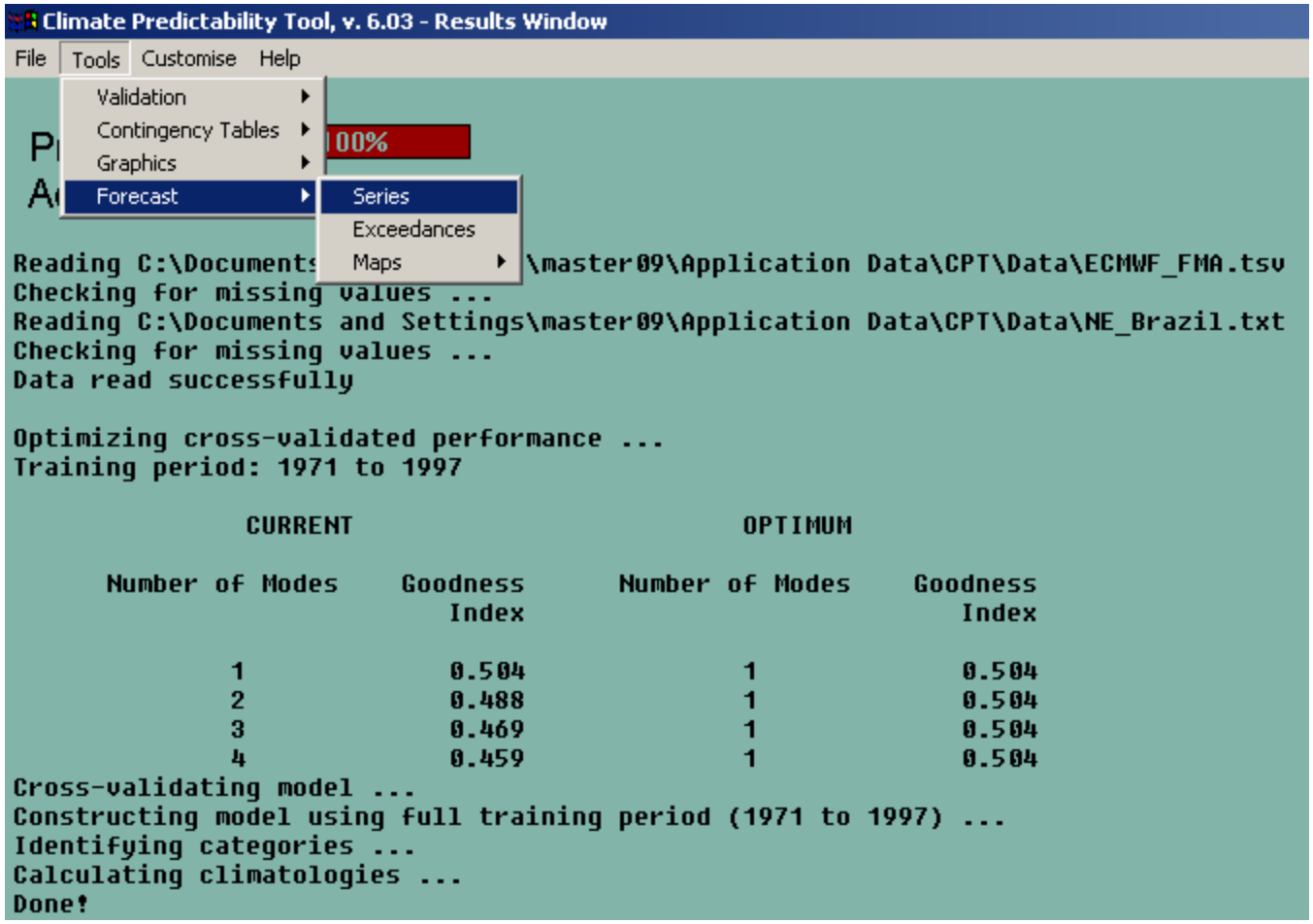



You then select:

(a) the starting year of the forecasts

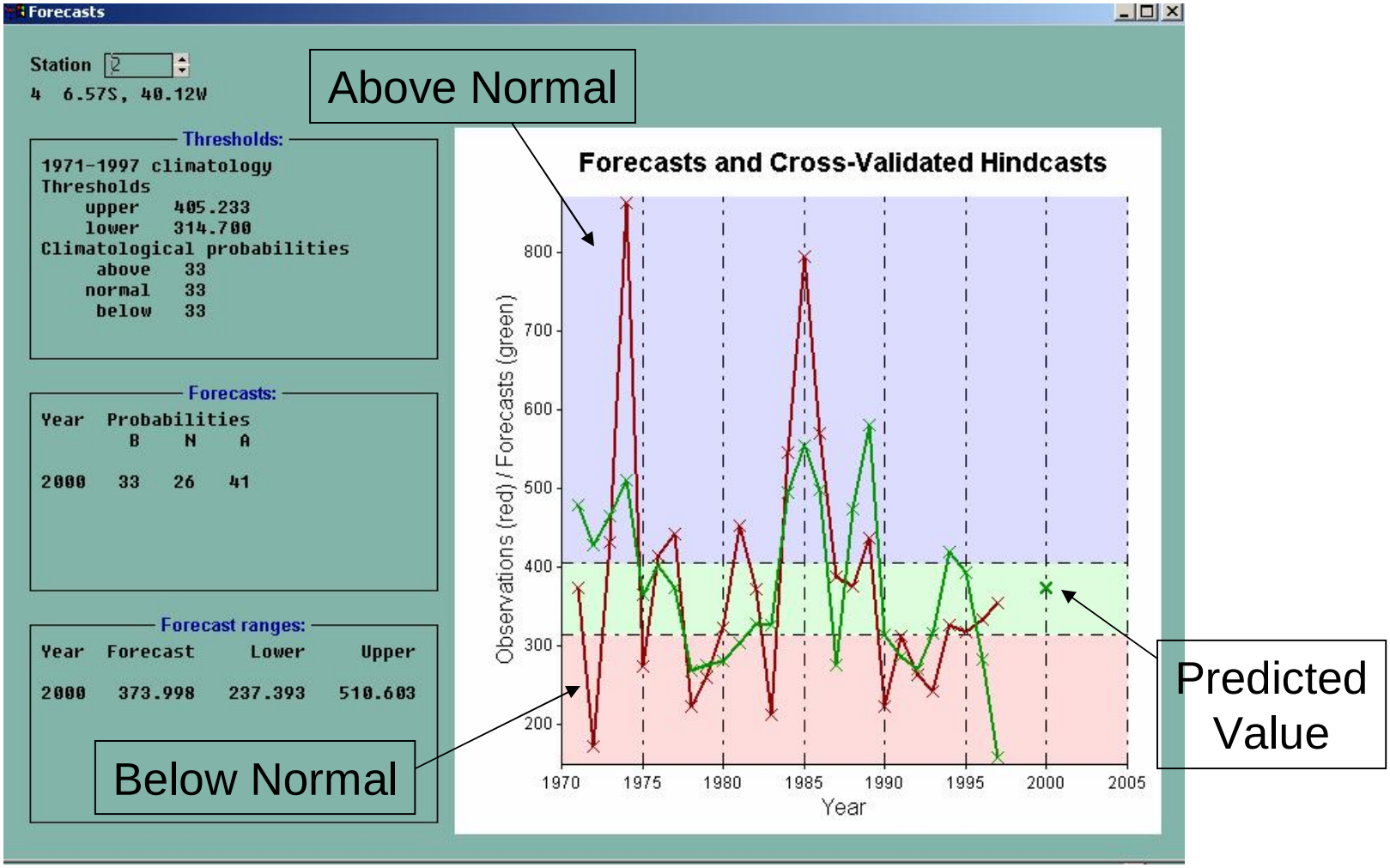
(b) the number of years to forecast

# FORECAST



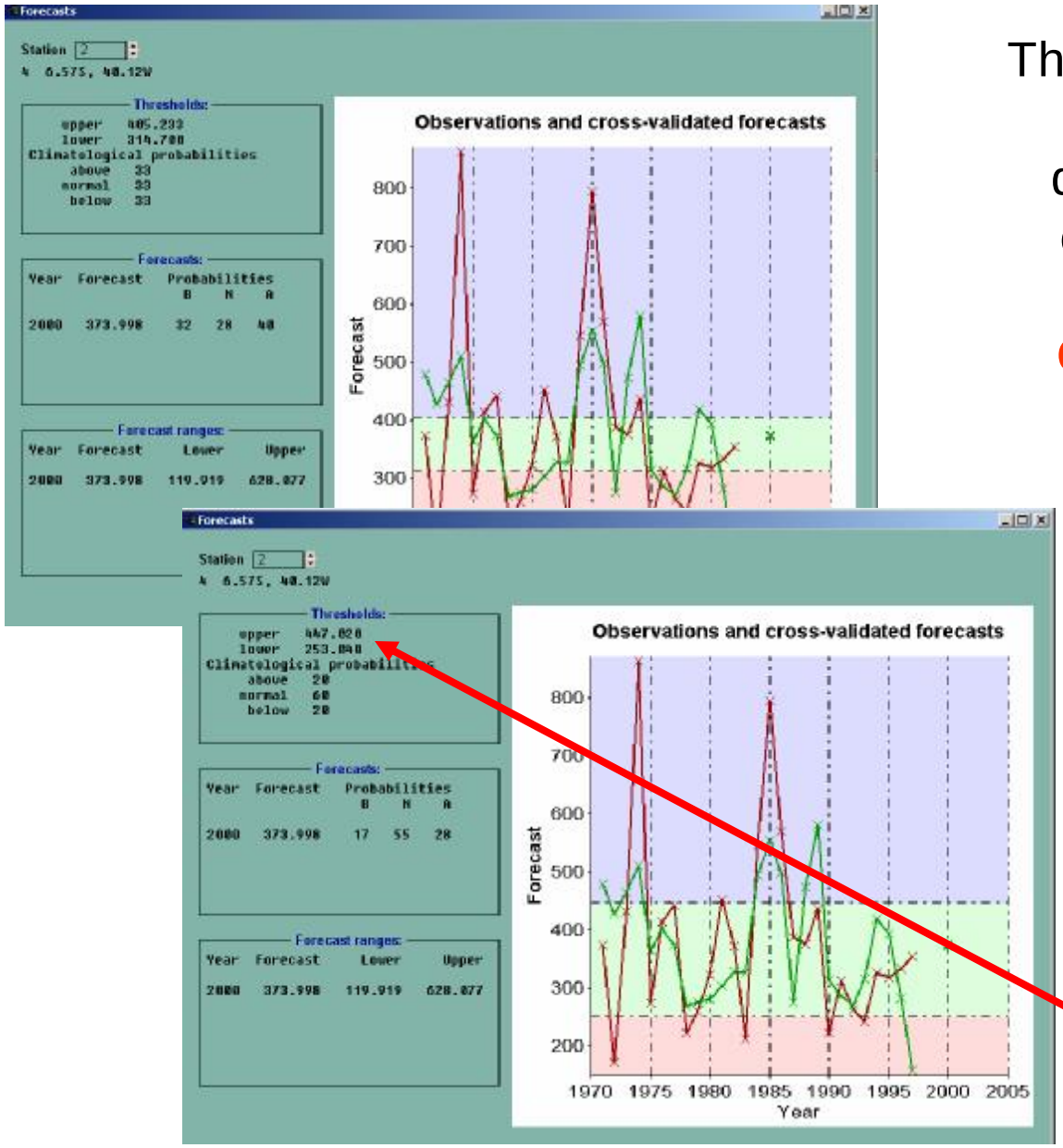

Once the file is selected and the years to forecast are chosen go to the menu Tools => Forecast => Series or Maps.

# FORECAST



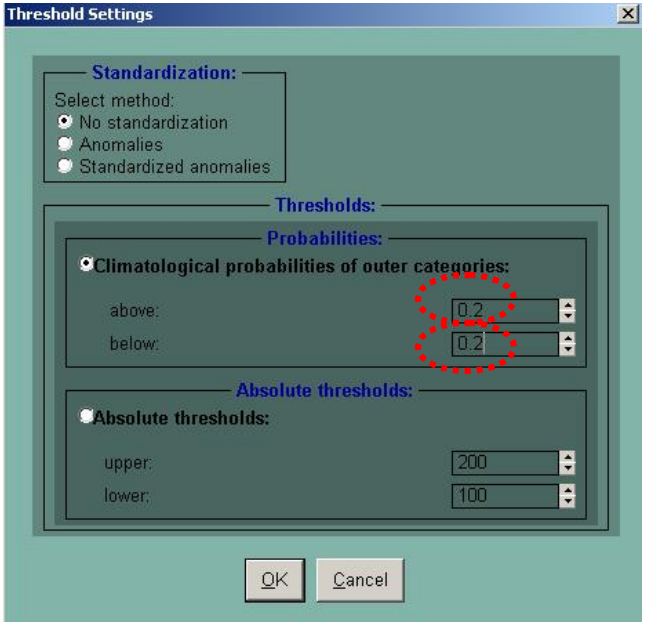

The option Series shows the predicted values (cross) for the current station as well as forecast possibilities, confidence limits for the forecast and, in the “Thresholds” box, the “category thresholds” as well as the climatological probabilities for the 3 categories.

# CHANGING CATEGORY DEFINITIONS



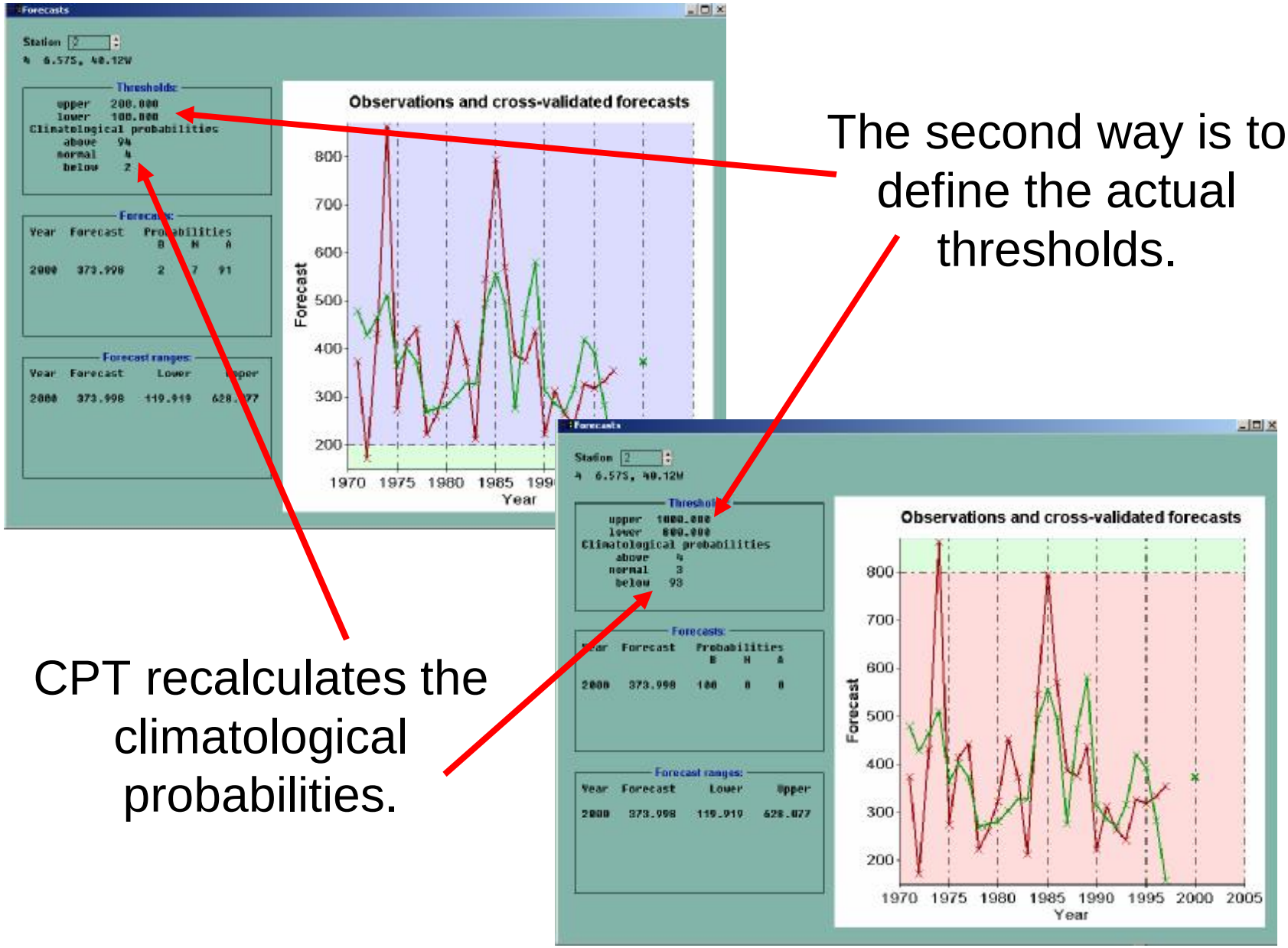

There are two ways to change how the categories are defined. The first way is to change the climatological probabilities.

Customize => Thresholds




CPT recalculates the thresholds

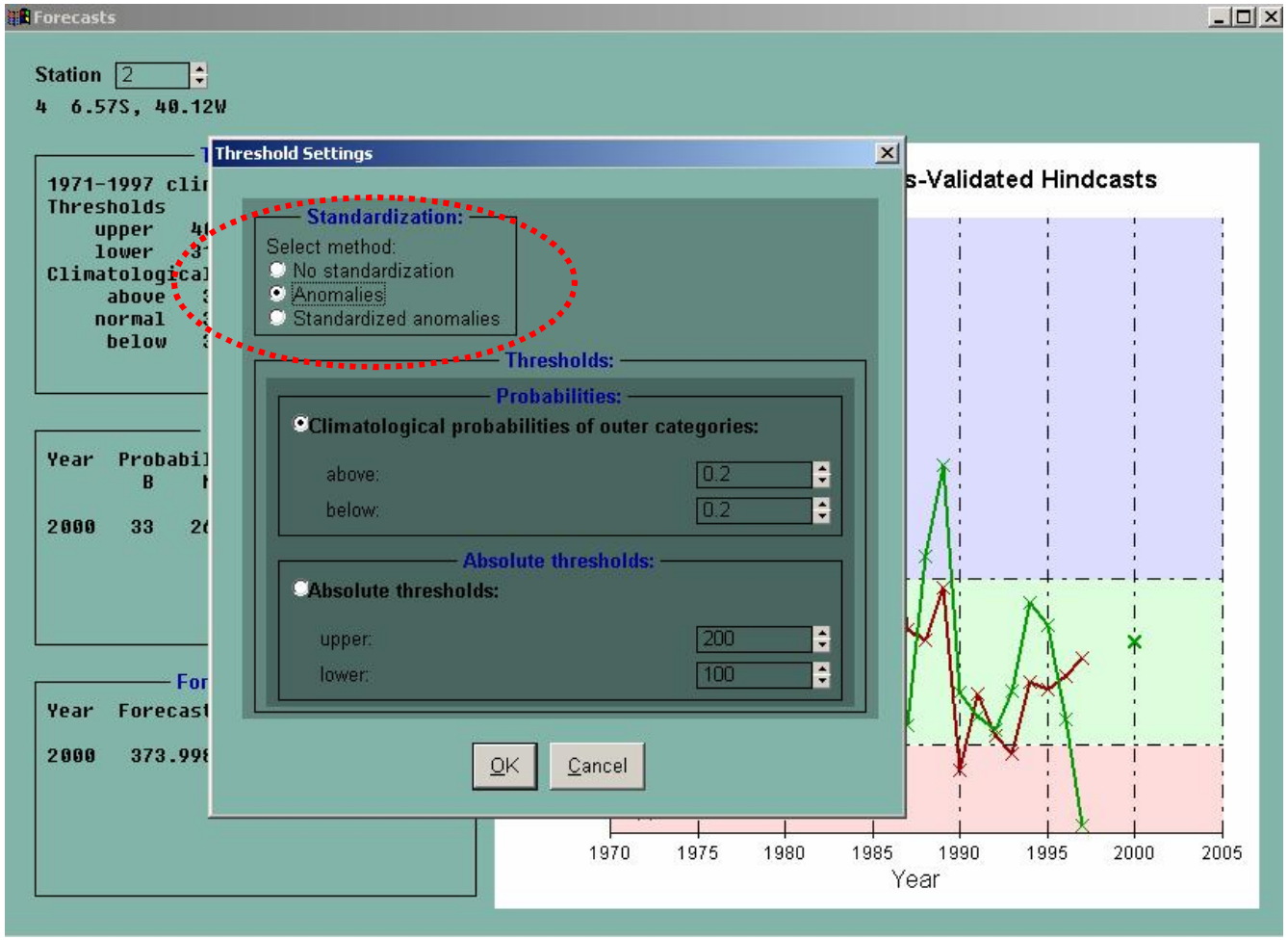
# CHANGING CATEGORY DEFINITIONS




The second way is to define the actual thresholds.

CPT recalculates the climatological probabilities.

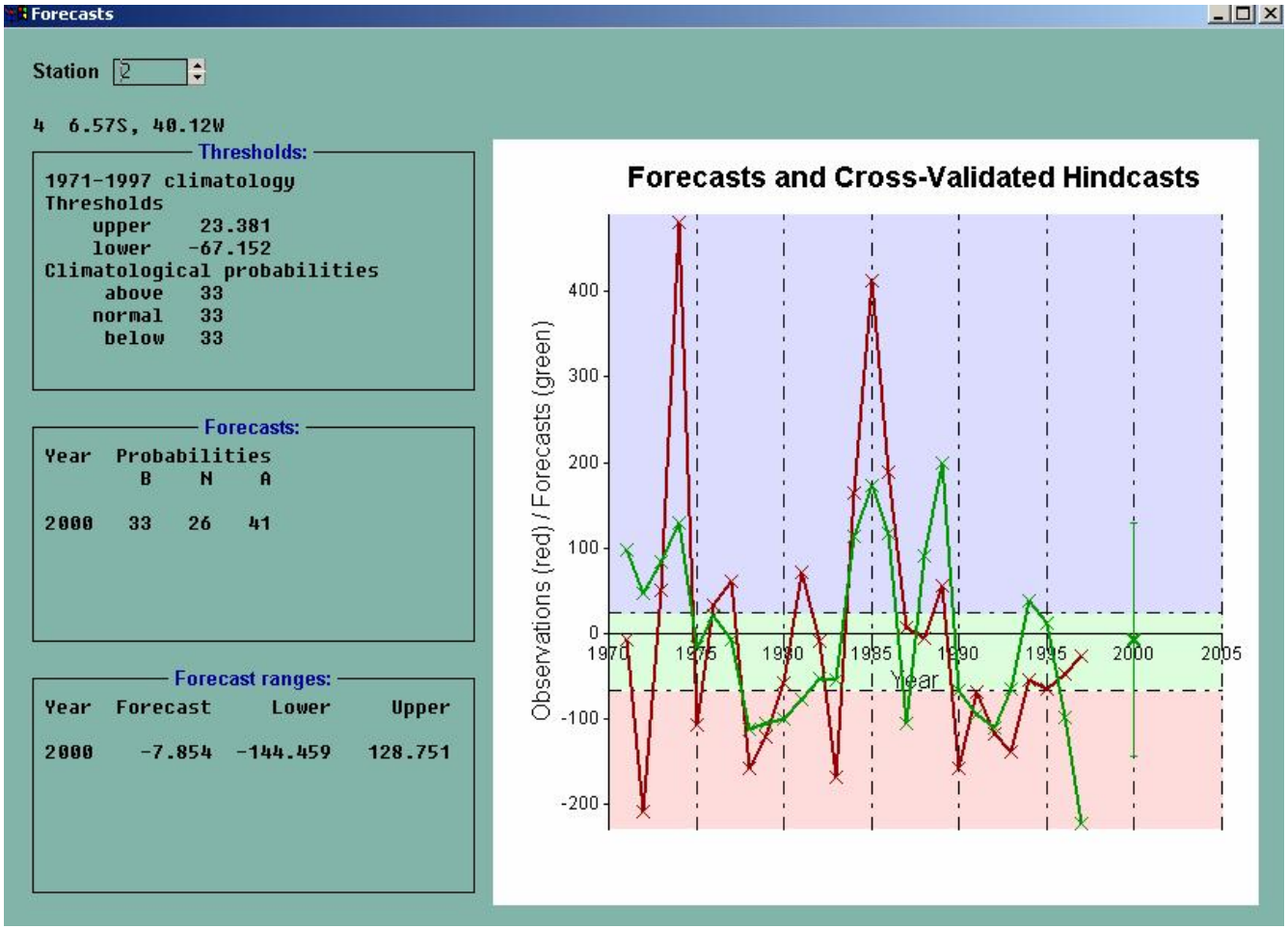
# EXPRESSING THE FORECAST AS ANOMALIES




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

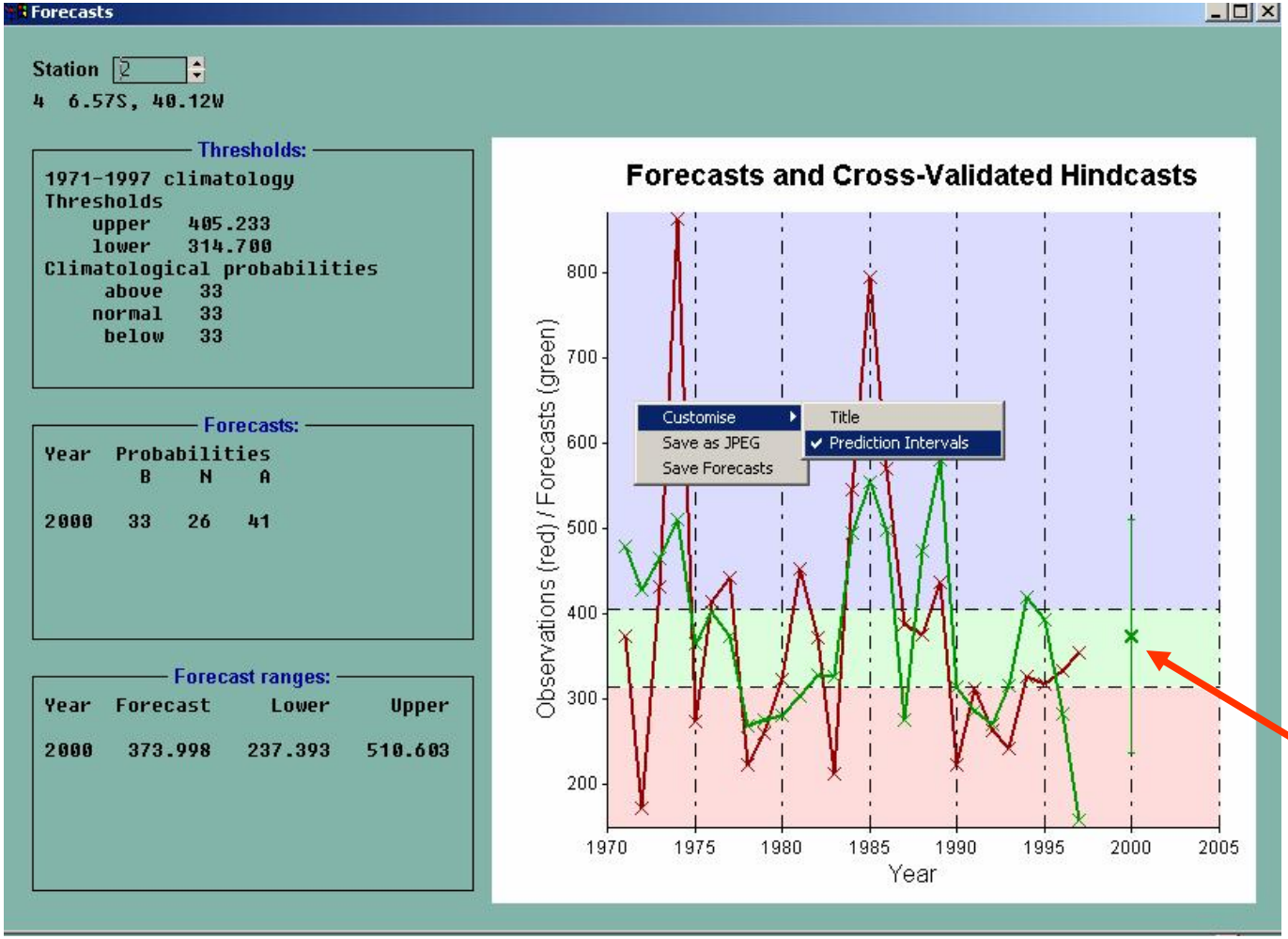
Customize => Thresholds

# EXPRESSING THE FORECAST AS ANOMALIES




The thresholds, as well as the forecast ranges, are now defined as anomalies. If absolute thresholds are set, CPT assumes that these are defined as anomalies.

# PREDICTION INTERVALS

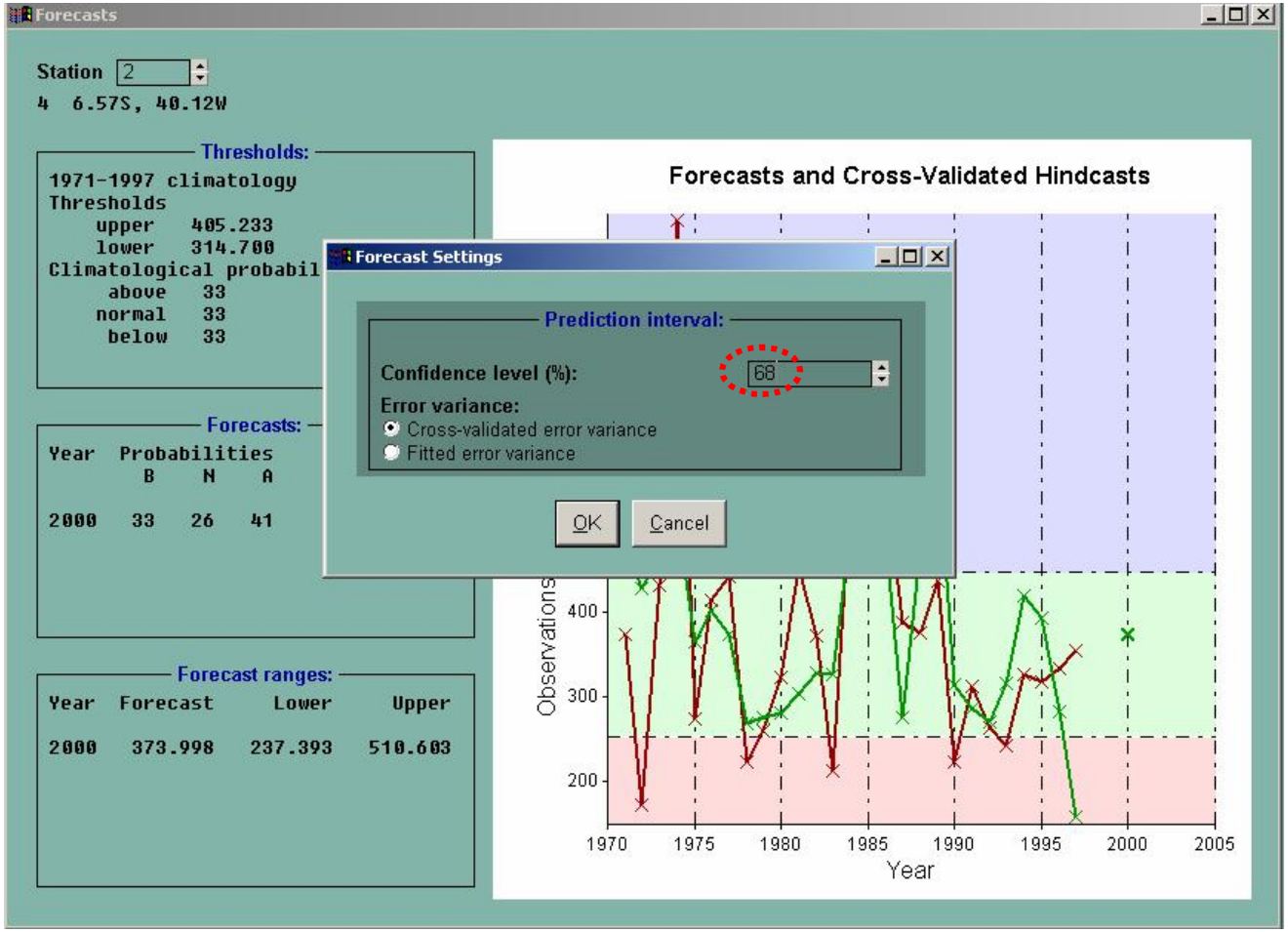



To draw error bars on the forecast, right click on the graph:

Customize => Prediction Intervals

An error bar is indicated.

# CHANGING THE PREDICTION INTERVAL

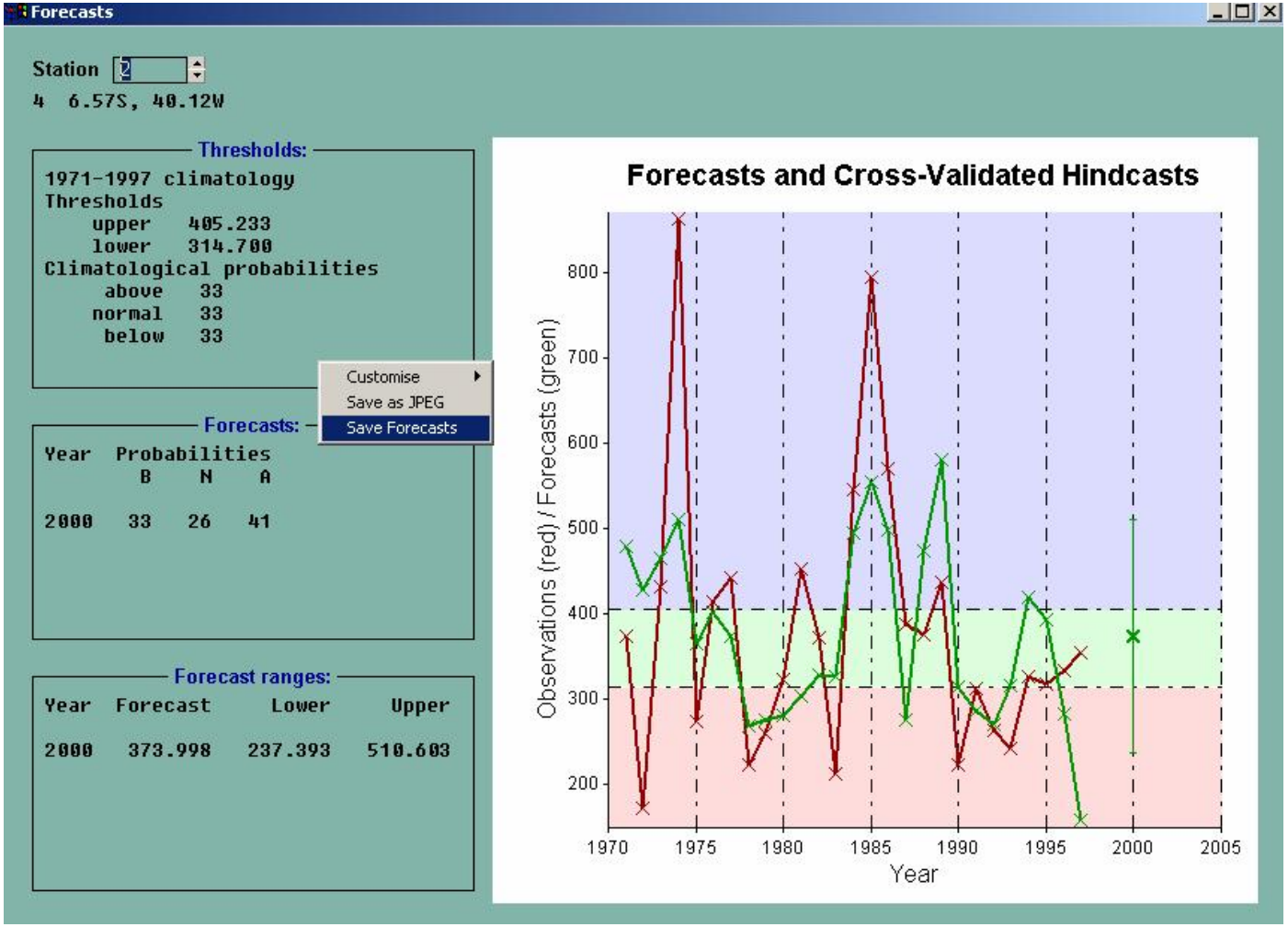



You can also change the width of the prediction interval.

Customize => Forecast Settings

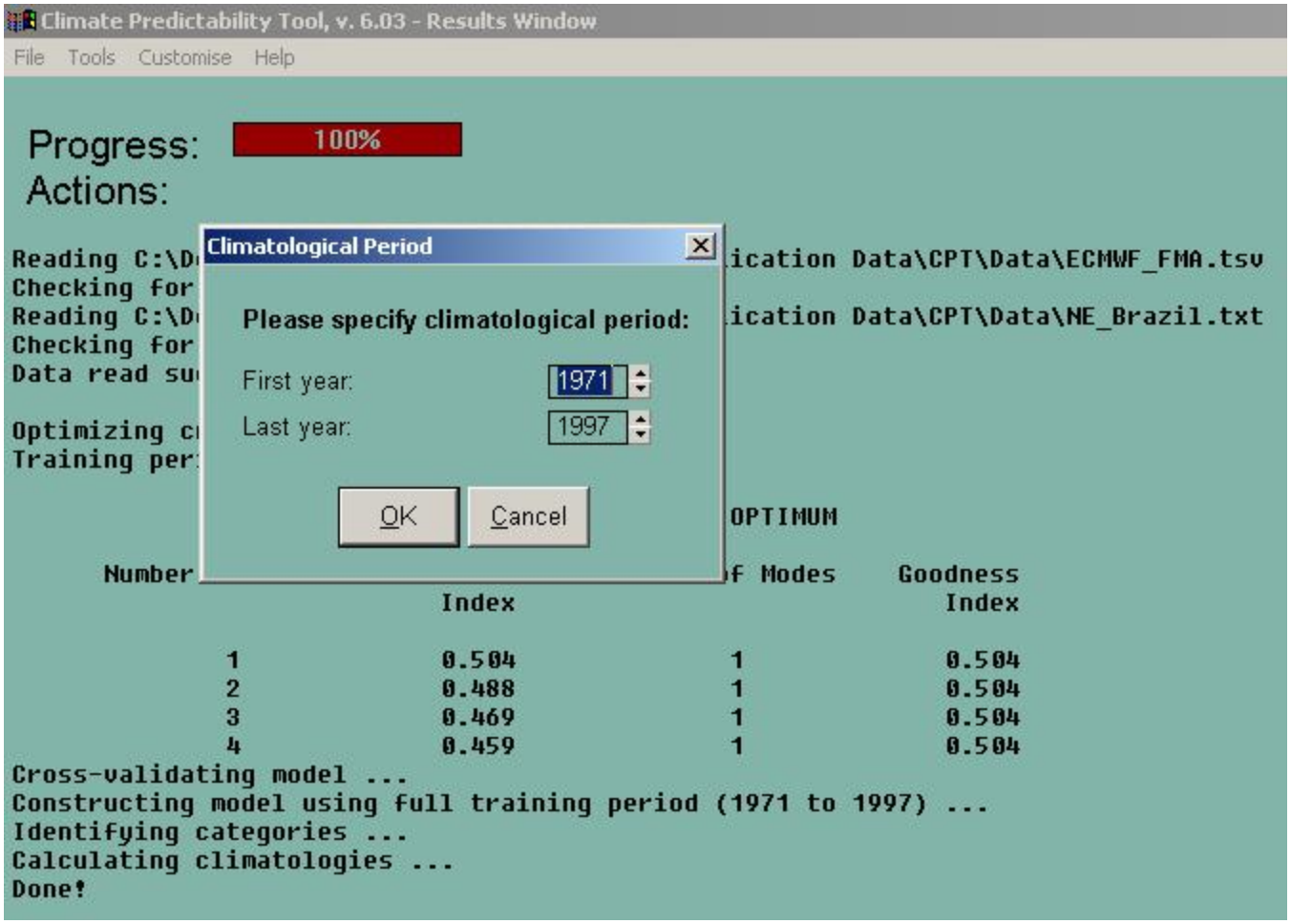
The default setting of 68.2% gives standard error bars.

# SAVING FORECASTS




To save the forecasts, right click, and specify the required output files.

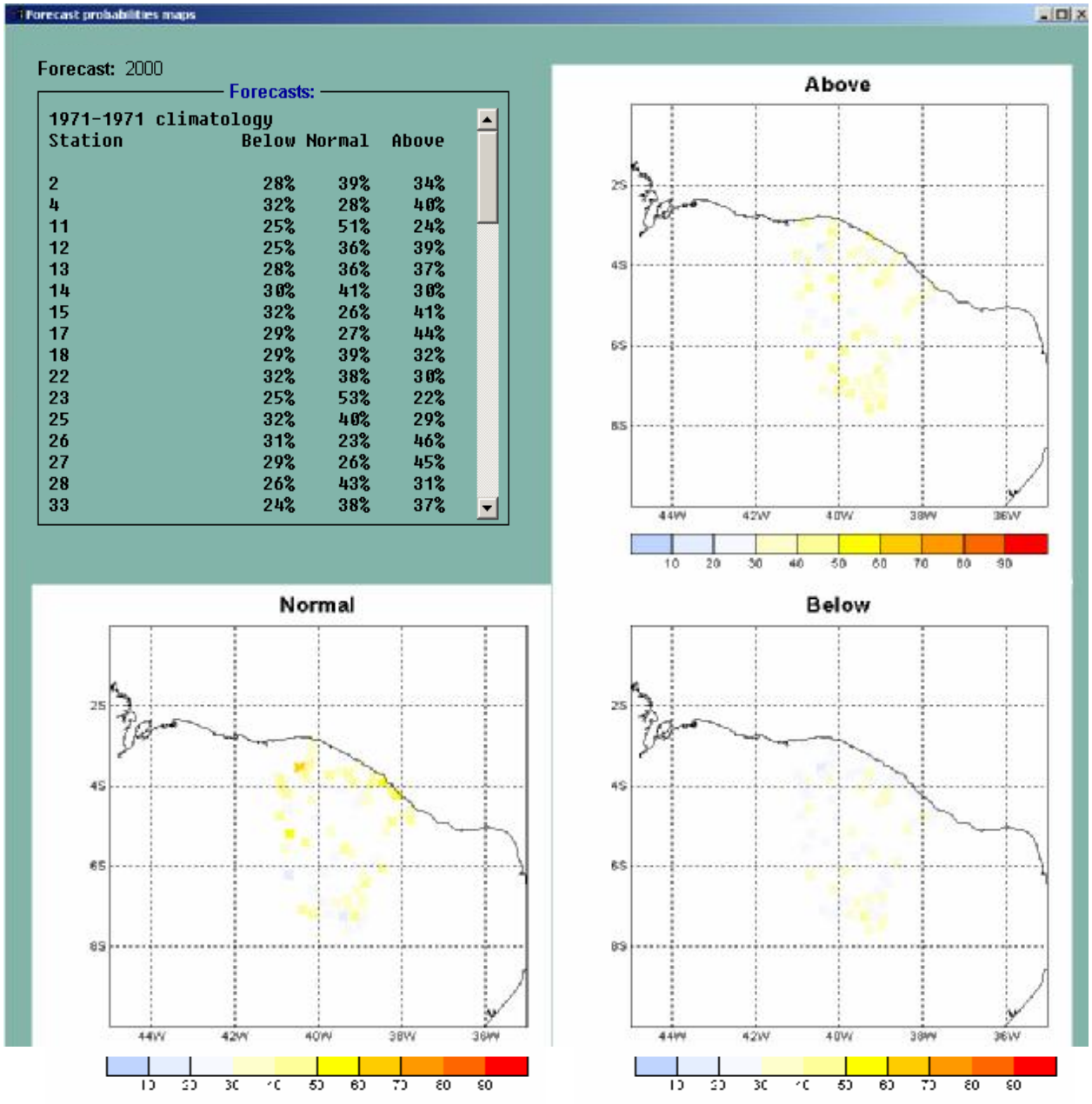
# CHANGING THE CLIMATOLOGICAL PERIOD




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:

Customise => Climatological Period

# FORECAST MAPS



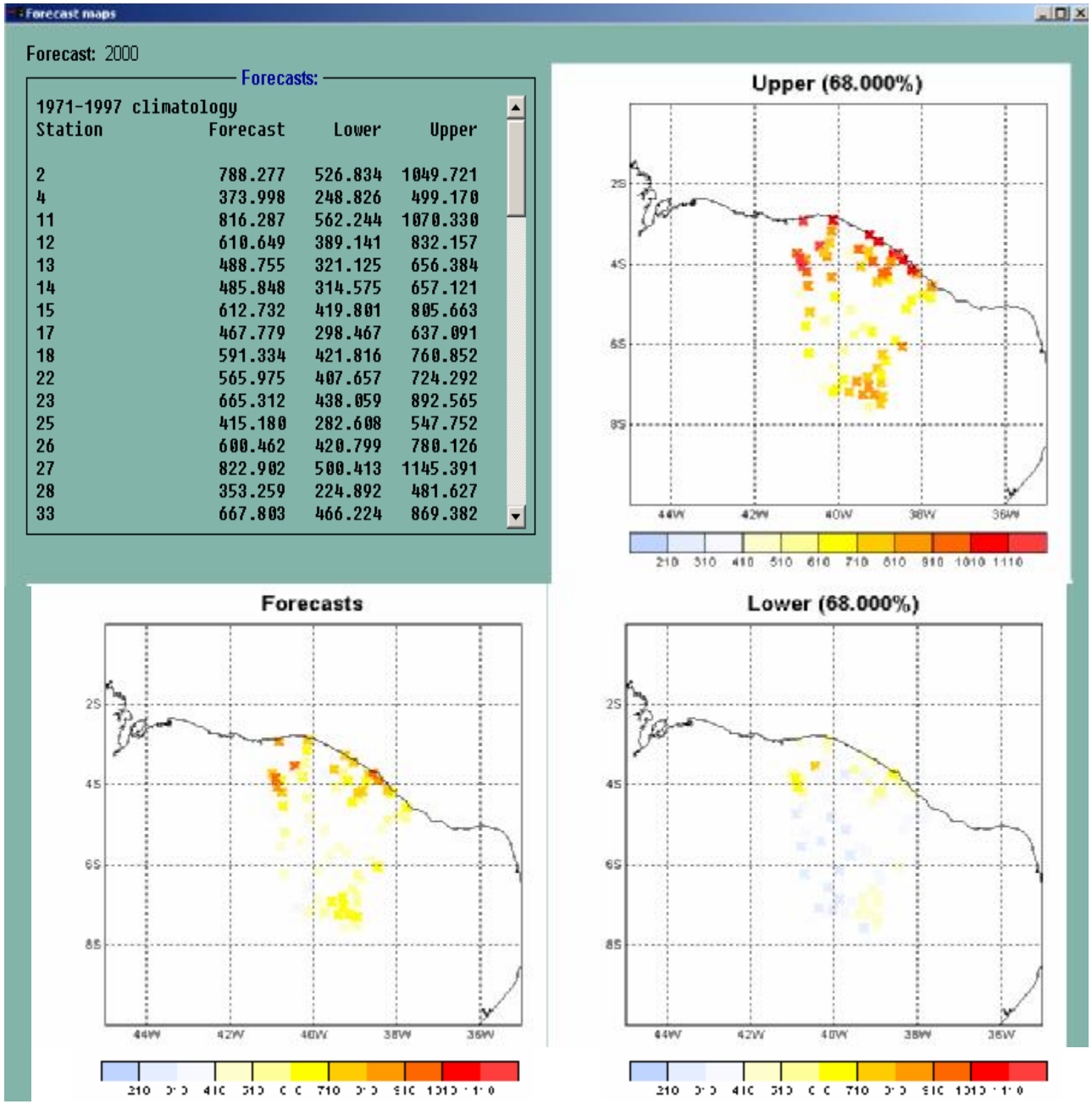

Tools => Forecast => Maps

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.

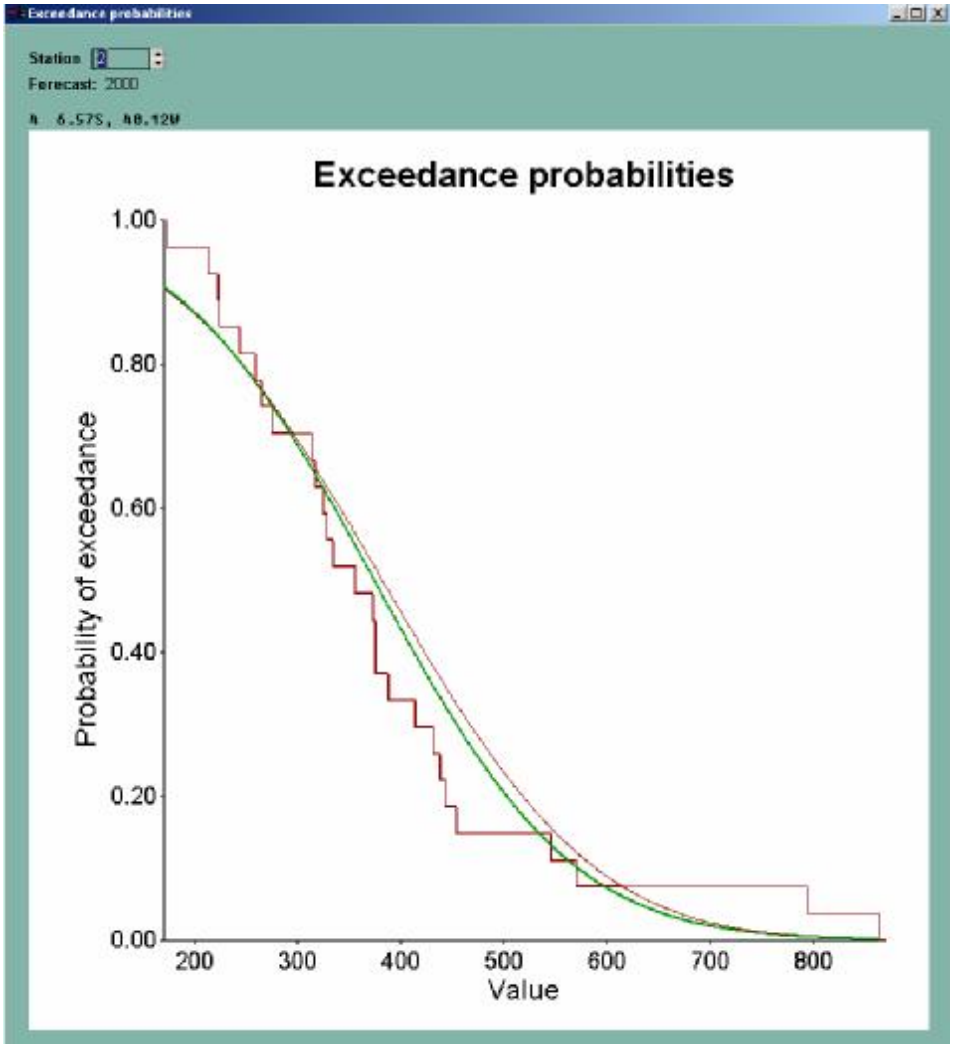
In this example it is evident that in 2000 the below-normal category has the lowest probability over most of north-east Brazil.

# 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/outreach/software/***

- 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***