

CProb Version 1.0 ReadMe

CProb.xla version 1.0 is a Microsoft Excel Add-in developed with Microsoft Excel Visual Basic for Applications, R-Excel, R(D)Com Server and R. The Add-in was developed with Microsoft Excel 2003 as a front end interface and conducts a conditional probability analysis using R version 2.4.0 and R(D)Com 2.0 /RExcel 1.5 as the statistical processor/R connector. CProb is freely available from <http://www.epa.gov/emap/nca/html/regions/cprob/index.html>.

To Install:

The CProb.xla Add-in requires Microsoft Excel, R and R(D)Com server. The installation steps assume that no prior installations of R, R(D)Com server, or CProb exist. If any of these have been previously installed, the safest option is to remove them prior to the new installation. The instructions are as follows:

1. If not already installed, install Microsoft Excel.
2. Install R, the R(D)Com server, and RExcel with the most current R and Friends installer, accepting all defaults. The current R and Friends installer is available from <http://sunsite.univie.ac.at/rcom/download/RAndFriends.distro/RAndFriendsLightSetup2081V3.0-10-1.exe>.
3. After installing R and Friends, activate RExcel through Start/Programs/R/RExcel/Activate RExcel Add-in.
4. Save CProb.xla on your local machine. Open Microsoft Excel, click on Tools:Add-Ins and Browse to CProb.xla. Make sure it is checked and click ok. Another message may appear asking if you'd like to save the add-in. You may choose either yes or no. A new menu should now be added to your Microsoft Excel Menu Bar entitled "Conditional Probability".

(Note: CProb.xla borrows extensively from RExcel. As such, ensure RExcel is installed correctly and working prior to adding the CProb Add-in. If RExcel is not installed and running, a Microsoft Visual Basic warning will likely result: “Can’t find project or library”)

To Calculate Conditional Probability:

A demonstration dataset is included with this documentation (DummyData.xls). After following the installation procedures above, open this spreadsheet and if necessary, add in the CProb.xla Add-in by checking (or browsing to) the CProb Add-in check box in Tools:Add-ins.

Details of the Conditional Probability Menu:

The menu contains three items: "About", which provides information about development environments and contacts; "View R Scripts", which provides access to the R code used to generate plots, calculate conditional probability, and bootstrap confidence intervals; and "Run Conditional Probability", which opens the primary interface for CProb. This interface has two tabs “Conditional Probability Analysis” and “Conditional Probability Plotting” (Figures 1 & 2).

The “Conditional Probability Analysis” tab (Figure 1) accepts all input variables to run the CPA, calculate confidence intervals by bootstrap resampling, and include probability survey inclusion probabilities. It has three sections: “Data”, “Required Arguments”, and “Optional Arguments”. The “Data” section accepts a worksheet range or single column for the stressor and response variables. Check the “With Column Name” checkbox if the first row of data contains a column name. The “Required Arguments” specify the details of the response variable threshold, with “Impairment Cutoff Direction” and “Cutoff Value”, and the probability direction of the calculation with “Probability Direction”. Often, the “Probability Direction” will correspond with the direction of the stressor in which impairment is expected. For instance, biological impairment might be expected with higher percentages of fine material in bedded sediment and a “Probability

Direction” of “Greater Than” could be used. However, the direction of the expected impairment is not always known and in those instances, CPA would be more exploratory and both directions might be used.

The "Optional Argument" section allows for inclusion probabilities with the "Weights" argument. Confidence interval's alpha (eg., 0.05 for 95% confidence intervals) and the bootstrap iterations can also be adjusted. Lastly, the “Conditional Probability Plotting” tab (Figure 2) provides controls for generating scatterplots, cumulative distribution functions, and conditional probability plots. Checkboxes control which plots are included. Axis and plot titles may be changed via the text boxes.

CProb Results and Output:

CProb returns two new worksheets and optional output graphs (Figures 3-5). The worksheets, entitled "Conditional Probability Output" and "Cumul. Dist. Function Output" are returned to Microsoft Excel®. The selected optional plots are generated by R (Figure 5 and figures 3a-c in Hollister et al. In Review). The five columns in the "Conditional Probability Output" worksheet are: "Sressor", "Raw.Data.Probability", "Bootstrap.Probability", "Lower.CI", and "Upper.CI." "Sressor" is the sorted Criterion Metric (i.e. X). "Raw.Data.Probability" contains the conditional probabilities calculated from the raw data. These are the plotted conditional probabilities (Figure 4 and 3c in Hollister et al. In Review). "Bootstrap.Probability" contains the mean bootstrapped probabilities and are not plotted. "Lower.CI" and "Upper.CI" are the upper and lower limits of the confidence interval determined by the "Confidence Interval Alpha" argument. In our example these are the lower 2.5% and upper 97.5% (i.e. 95% confidence interval) of the bootstrapped estimates. The “Cumul. Dist. Function Output” worksheet contains the results of three cumulative distribution functions; one for all samples which are in the first two columns, one for all samples that are greater than the response threshold and one for all samples less than the

threshold which are in the last 4 columns. The good and bad designations are dependent upon the response cut off direction. If weights are included they are also used in generating the conditional probabilities, confidence intervals, and the cumulative distribution functions. These resultant output worksheets are included for users wishing to generate custom graphs and figures different from the default CProb graphs.

To Run CProb with DummyData.xls:

- 1.) Select Conditional Probability: Run Conditional Probability
- 2.) Add in worksheet range for the criteria metric (i.e. the metric which impairment probability is conditioned on) and the Impairment Metric (i.e. the metric defining impaired biological condition).
- 3.) Choose the Impairment Cutoff Direction and Value (i.e. EPT Taxa less than 9)
- 4.) Choose the probability direction. Easiest way to think of this is as an exceedance. For instance, if you expect higher probability of impairment for higher values of the Criteria Metric, it is an exceedance and choose “greater than”
- 5.) Optional arguments may be left unchanged. If you choose to alter some and not others, you may leave the other optional arguments unchanged if you wish.
- 6.) To create plots, select the “Conditional Probability Plotting” tab, otherwise skip to step 8.
- 7.) Select which of the plots you’d like to included in the plot. If you choose, edit the title and axis labels.
- 8.) Click Run

On the following pages are a screenshot of the Conditional Probability GUI and the resultant output and graphs using the DummyData.xls dataset. Two additional dataset, JEQData_wq.csv and JEQData_le.csv are also included in the .zip file. These datasets were used in Hollister et al. (In Review). These include the water quality data from Paul et al (2005) and the landscape data

which has been used in part of numerous studies (e.g. Hollister et al 2004, Hollister et al 2008a, and Hollister et al 2008b).

Suggested Citation

If your use of CProb results in a publication or technical report, please cite the manuscript published in the Journal of Environmental Quality:

Hollister, J. W., H. A. Walker, J. F. Paul. 2008. CProb: A Computational Tool for Conducting Conditional Probability Analysis. Journal of Environmental Quality. 37:2392-2396.

References:

Hollister, J.W., M.L. Gonzalez, J.F. Paul, P.V. August, and J.L. Copeland. 2004. Assessing the accuracy of the national land cover dataset area estimates at multiple spatial extents. Photogrammetric Engineering and Remote Sensing 70:405-414.

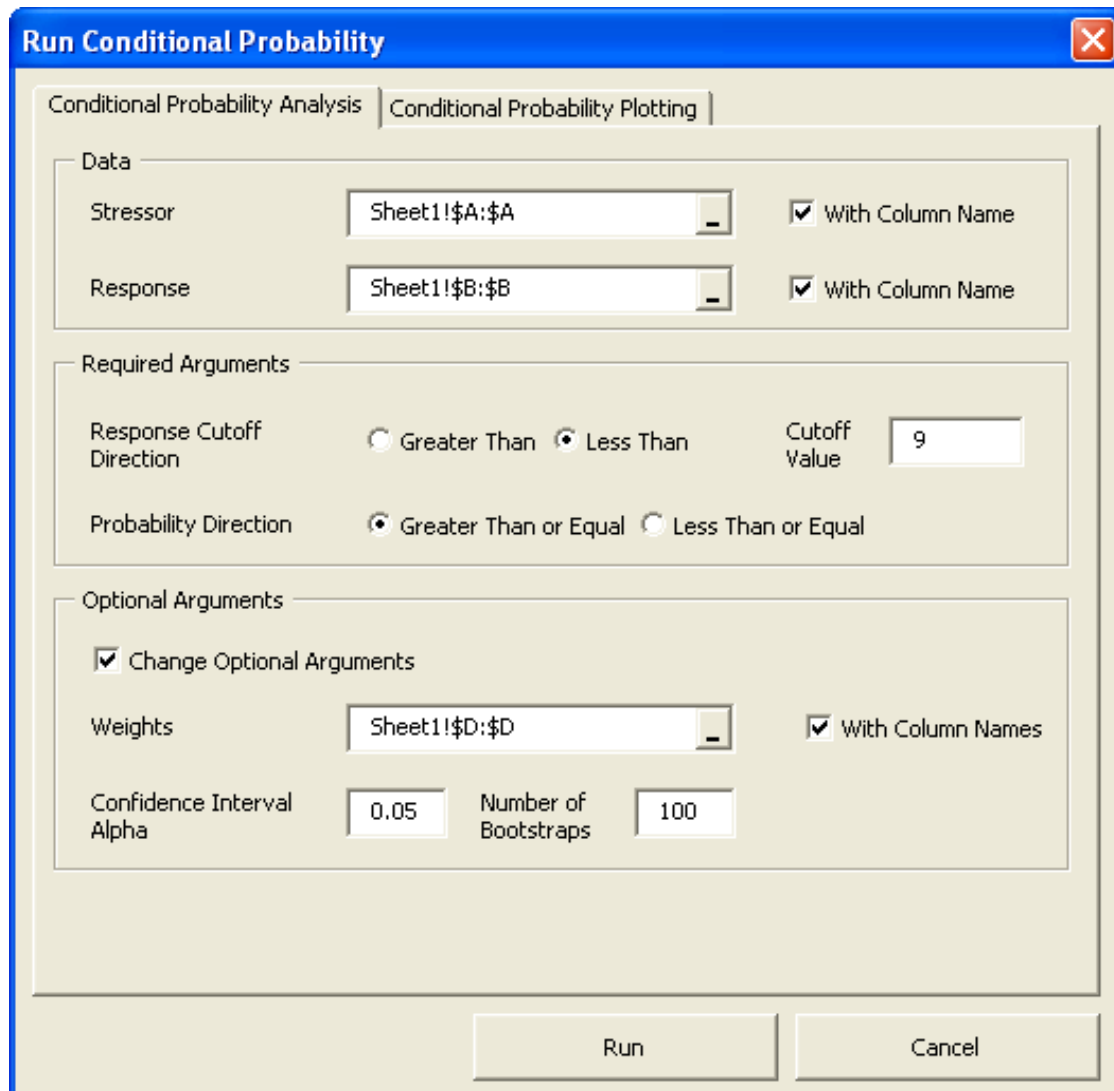
Hollister, J.W., H.A. Walker, and J.F. Paul. In Press. CProb: A Computational Tool for Conducting Conditional Probability Analysis. Journal of Environmental Quality.

Hollister, J.W., P.V. August, J.F. Paul, and H.A. Walker. 2008a. Predicting estuarine sediment metal concentrations and inferred ecological conditions: An information theoretic approach. Journal of Environmental Quality 37:234-244.

Hollister, J.W., P.V. August, and J.F. Paul. 2008b. Effects of spatial extent on landscape structure and sediment metal concentration relationships in small estuarine systems of the united states' mid-atlantic coast. Landscape Ecology 23:91-106.

Paul, J.F., and M.E. McDonald. 2005. Development of empirical, geographically specific water quality criteria: A conditional probability analysis approach. *Journal of the American Water Resources Association* 41:1211-1223.

Figure 1. Conditional Probability Analysis Tab Screen Capture



The image shows a software dialog box titled "Run Conditional Probability". It has a blue title bar with a close button in the top right corner. The dialog is divided into two tabs: "Conditional Probability Analysis" (selected) and "Conditional Probability Plotting".

Data Section:

- Stressor:** A text box containing "Sheet1!\$A:\$A" with a dropdown arrow to its right. A checkbox labeled "With Column Name" is checked.
- Response:** A text box containing "Sheet1!\$B:\$B" with a dropdown arrow to its right. A checkbox labeled "With Column Name" is checked.

Required Arguments Section:

- Response Cutoff Direction:** Two radio buttons: "Greater Than" (unselected) and "Less Than" (selected). To the right is a text box labeled "Cutoff Value" containing the number "9".
- Probability Direction:** Two radio buttons: "Greater Than or Equal" (selected) and "Less Than or Equal" (unselected).

Optional Arguments Section:

- A checkbox labeled "Change Optional Arguments" is checked.
- Weights:** A text box containing "Sheet1!\$D:\$D" with a dropdown arrow to its right. A checkbox labeled "With Column Names" is checked.
- Confidence Interval Alpha:** A text box containing "0.05".
- Number of Bootstraps:** A text box containing "100".

Buttons: At the bottom right, there are two buttons: "Run" and "Cancel".

Figure 2. Conditional Probability Plotting Tab Screen Capture

The screenshot shows a software dialog box titled "Run Conditional Probability" with a close button (X) in the top right corner. The dialog has two tabs: "Conditional Probability Analysis" and "Conditional Probability Plotting", with the latter being the active tab. The dialog is organized into three main sections, each with a title and a set of options and input fields.

Scatterplot

- ☒ Include Scatterplot
- Scatterplot Title: Scatterplot
- Y-axis Label: EPT Taxa
- X-axis Label: Percent Fines

Cumulative Distribution Function (CDF)

- ☒ Include CDF plot
- CDF Title: Cumulative Distribution Function
- Y-axis Label: Proportion of Samples
- X-axis Label: Percent Fines

Conditional Probability

- ☒ Include Conditional Probability Plot
- Conditional Probability Title: Conditional Probability
- Y-axis Label: Probability of EPT tax < 9 if $X \geq X_c$
- X-axis Label: X_c , Percent Fines

At the bottom of the dialog are two buttons: "Run" and "Cancel".

Figure 3. Output Worksheet Screen Capture

Microsoft Excel - DummyData.xls

File Edit View Insert Format Tools Data Window REExcel CProb Help

Arial 10

Reply with Changes... End Review...

	A	B	C	D	E	F	G	H	I	J
1	Stressor	Raw.Data.	Bootstrap.	Lower.CI	Upper.CI					
2	2	0.524905	0.48205	0.331526	0.619965					
3	10	0.53537	0.49166	0.338136	0.632325					
4	11	0.584564	0.536838	0.369206	0.690429					
5	12	0.598746	0.549862	0.378164	0.707179					
6	14	0.610964	0.561082	0.385881	0.72161					
7	15	0.629778	0.57836	0.397763	0.74383					
8	18	0.660893	0.606934	0.417415	0.78058					
9	22	0.684889	0.628971	0.432571	0.808922					
10	25	0.714361	0.656037	0.451185	0.843731					
11	27	0.730162	0.670548	0.461165	0.862394					
12	33	0.712361	0.681799	0.469826	0.858953					
13	34	0.744113	0.697247	0.469518	0.892292					
14	38	0.760647	0.717509	0.5059	0.93543					
15	41	0.769448	0.72396	0.510287	0.939912					
16	45	0.756621	0.732525	0.525903	0.943262					
17	46	0.7981	0.763425	0.506792	0.93303					
18	47	0.839135	0.792908	0.53037	0.97226					
19	50	0.863077	0.812393	0.545503	1					
20	51	0.851987	0.832834	0.589683	1					
21	53	0.834346	0.846283	0.610392	1					
22	55	0.827039	0.844796	0.606167	1					
23	56	0.808013	0.833223	0.562844	1					
24	57	0.77877	0.813912	0.496257	1					
25	60	0.758495	0.796859	0.450093	1					
26	61	0.869573	0.84619	0.48513	1					
27	63	0.839039	0.864026	0.571866	1					
28	67	0.82746	0.861438	0.544664	1					
29	72	0.825727	0.860548	0.540591	1					
30	75	0.817512	0.854917	0.520418	1					
31	77	0.785435	0.832931	0.43612	1					
32	78	0.748567	0.80766	0.33923	1					
33	80	0.746679	0.806291	0.334269	1					
34	83	1	0.93056	0.447674	1					
35	88	1	0.93673	0.462971	1					
36	90	1	0.954232	0.605802	1					
37	91	1	0.97	1	1					
38	99	1	0.97	1	1					
39										

Sheet1 Sheet2 Sheet3 Conditional Probability Output Cumul. Dist. Function Output

Ready

Figure 5. Screen capture of graphic output from the Conditional probability Add-in.

