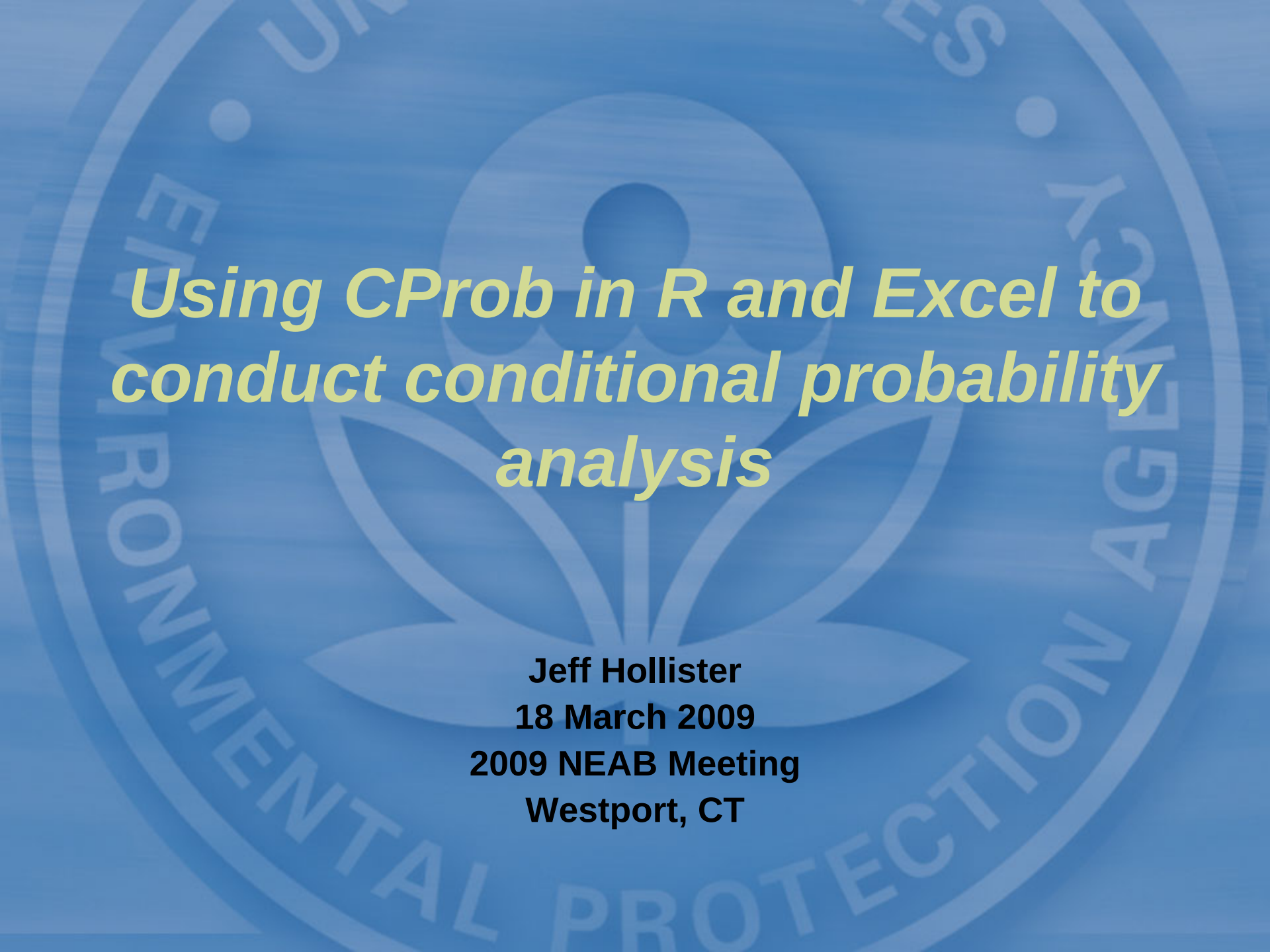


US EPA ARCHIVE DOCUMENT



Using CProb in R and Excel to conduct conditional probability analysis

**Jeff Hollister
18 March 2009
2009 NEAB Meeting
Westport, CT**

Goals of this Presentation

Goal 1: Introduce Conditional Probability

Goal 2: Introduce CProb as a tool for
conducting conditional probability analysis

Goal 3: Install and Use CProb in Excel

Goal 4: Use CProb in R

What is Conditional Probability?

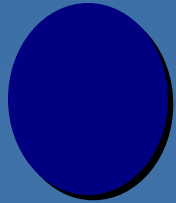
Conditional Probability is:

- Probability of one event occurring given another event has also occurred.

Environmental Context:

- Probability of an impaired environmental condition (i.e. response) given that another variable (i.e. stressor) is greater than, less than, or equal to a certain value

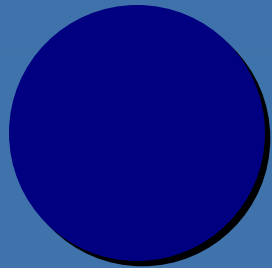
What is Conditional Probability?



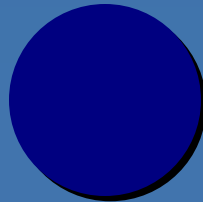
IBI = 5
TP = 0.02



IBI = 3
TP = 0.04



IBI = 1
TP = 0.8



IBI = 1
TP = 1.5



IBI = 2
TP = 0.04

Impairment = $IBI < 3$

Probability of Impairment = $3/5 = 0.6$

Conditional Probability of Impairment

<u>TP(X)</u>	<u>Impaired(Y)</u>	<u>Cond. Prob.</u>
0.02	0	$3/5 = 0.6$
0.04	1	$3/4 = 0.75$
0.04	0	$2/3 = 0.67$
0.8	1	$2/2 = 1$
1.5	1	$1/1 = 1$

CProb 1.0

Why was CProb developed?

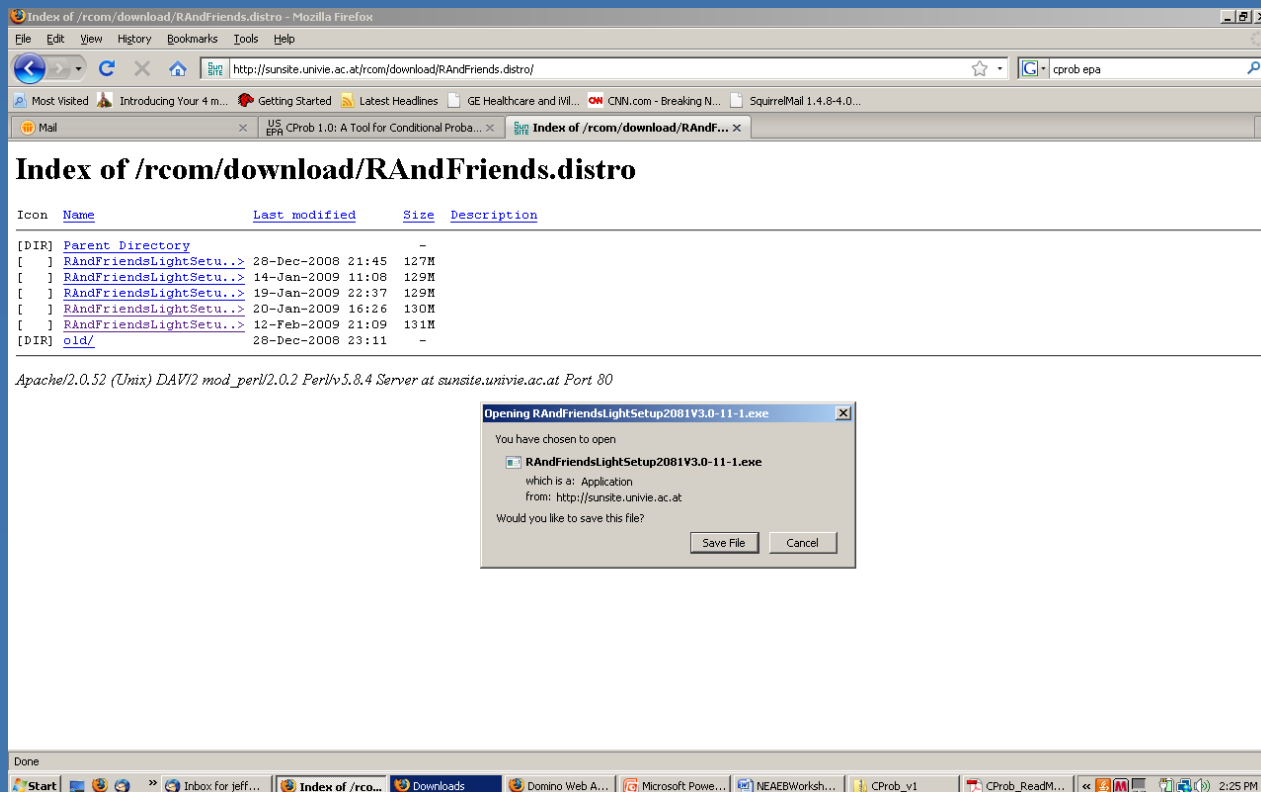
- Ease Calculations
- Make calculations reproducible with script
- Provide familiar user interface

What does it do?

- Conditional Probability
- Bootstraps Confidence Intervals
- Cumulative Distribution Functions
- Plots - Scatter Plots, CDFs, and Conditional Probability

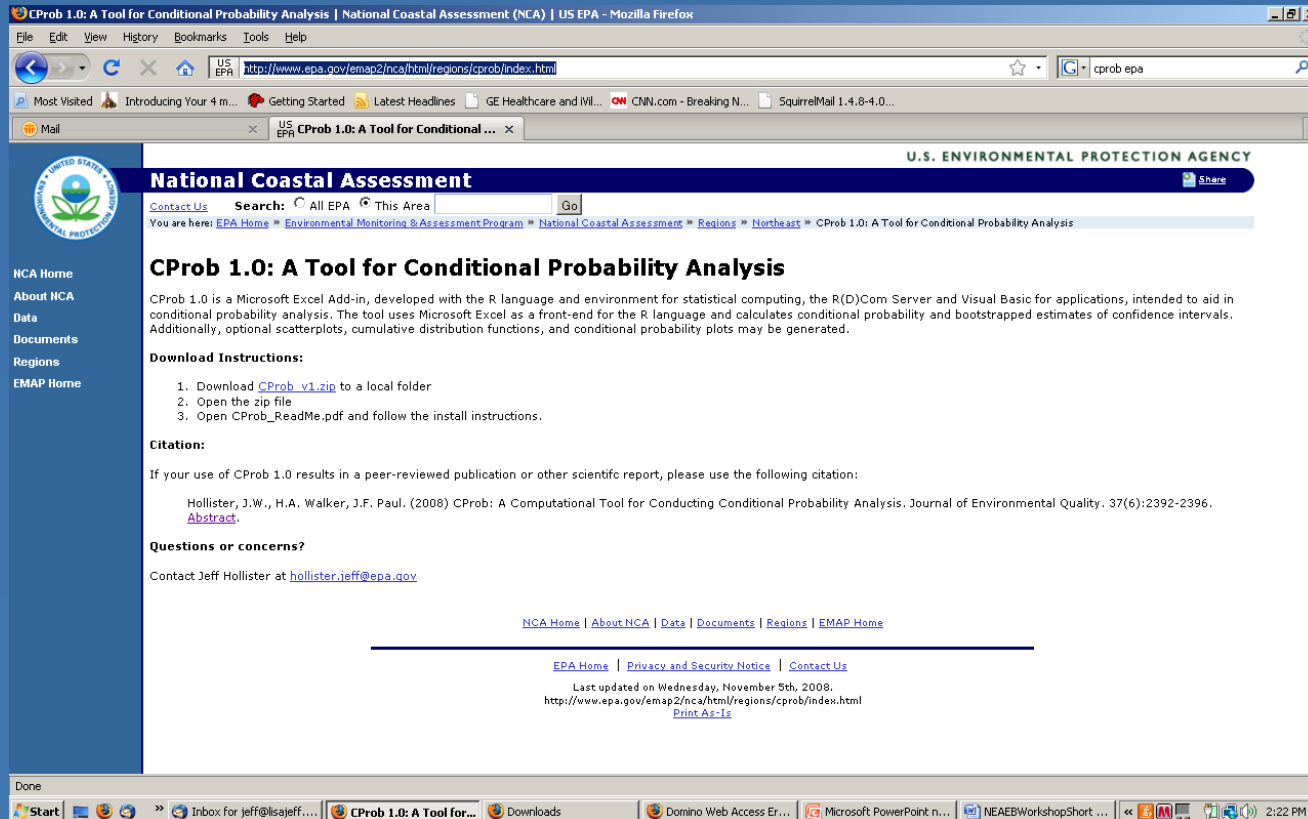
Using CProb 1.0 in Excel: Download

- Download R and R(D)Com Server:
 - Use most recent R and Friends:
<http://sunsite.univie.ac.at/rcom/download/RAndFriends.distro/>



Using CProb 1.0 in Excel: Download

- Download CProb 1.0
 - <http://www.epa.gov/emap2/nca/html/regions/cprob/index.html>



The screenshot shows a Mozilla Firefox browser window displaying the EPA's National Coastal Assessment (NCA) website. The address bar shows the URL <http://www.epa.gov/emap2/nca/html/regions/cprob/index.html>. The page title is "CProb 1.0: A Tool for Conditional Probability Analysis | National Coastal Assessment (NCA) | US EPA - Mozilla Firefox". The page content includes the EPA logo, a search bar, and a navigation menu. The main heading is "CProb 1.0: A Tool for Conditional Probability Analysis". Below this, a paragraph describes the tool as a Microsoft Excel Add-in developed with the R language and environment for statistical computing. A "Download Instructions" section lists three steps: 1. Download [CProb_v1.zip](#) to a local folder, 2. Open the zip file, and 3. Open [CProb_ReadMe.pdf](#) and follow the install instructions. A "Citation:" section provides a citation for Hollister, J.W., H.A. Walker, J.F. Paul. (2008) CProb: A Computational Tool for Conducting Conditional Probability Analysis. Journal of Environmental Quality. 37(6):2392-2396. A "Questions or concerns?" section provides contact information for Jeff Hollister at hollister.jeff@epa.gov. The footer includes links to "NCA Home", "About NCA", "Data", "Documents", "Regions", and "EMAP Home", as well as "EPA Home", "Privacy and Security Notice", and "Contact Us". It also states the page was last updated on Wednesday, November 5th, 2008, and provides the URL <http://www.epa.gov/emap2/nca/html/regions/cprob/index.html> and a "Print As-Is" link.

CProb 1.0: A Tool for Conditional Probability Analysis

CProb 1.0 is a Microsoft Excel Add-in, developed with the R language and environment for statistical computing, the R(D)Com Server and Visual Basic for applications, intended to aid in conditional probability analysis. The tool uses Microsoft Excel as a front-end for the R language and calculates conditional probability and bootstrapped estimates of confidence intervals. Additionally, optional scatterplots, cumulative distribution functions, and conditional probability plots may be generated.

Download Instructions:

1. Download [CProb_v1.zip](#) to a local folder
2. Open the zip file
3. Open [CProb_ReadMe.pdf](#) and follow the install instructions.

Citation:

If your use of CProb 1.0 results in a peer-reviewed publication or other scientific report, please use the following citation:

Hollister, J.W., H.A. Walker, J.F. Paul. (2008) CProb: A Computational Tool for Conducting Conditional Probability Analysis. Journal of Environmental Quality. 37(6):2392-2396. [Abstract.](#)

Questions or concerns?

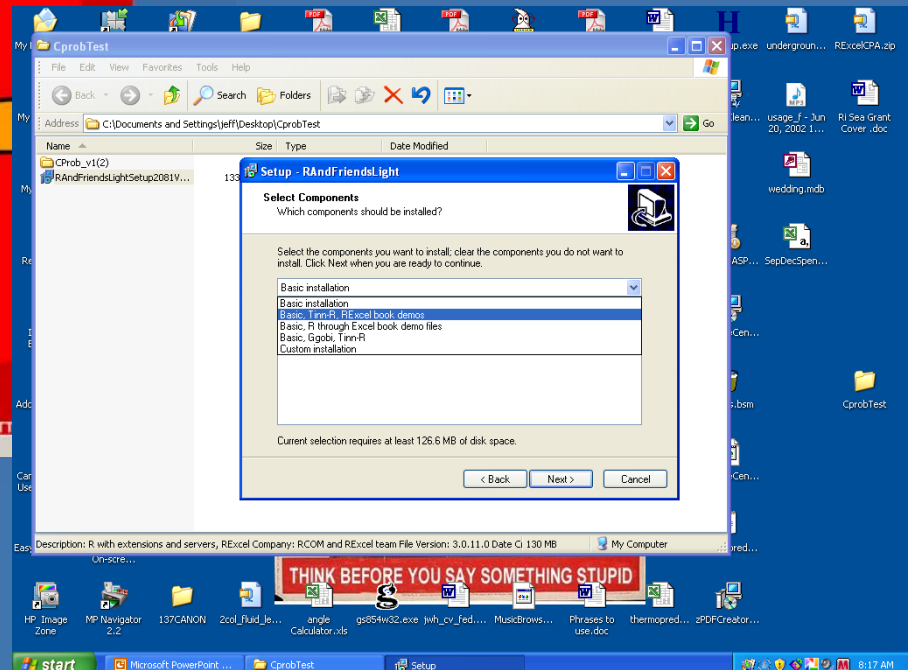
Contact Jeff Hollister at hollister.jeff@epa.gov

[NCA Home](#) | [About NCA](#) | [Data](#) | [Documents](#) | [Regions](#) | [EMAP Home](#)

[EPA Home](#) | [Privacy and Security Notice](#) | [Contact Us](#)

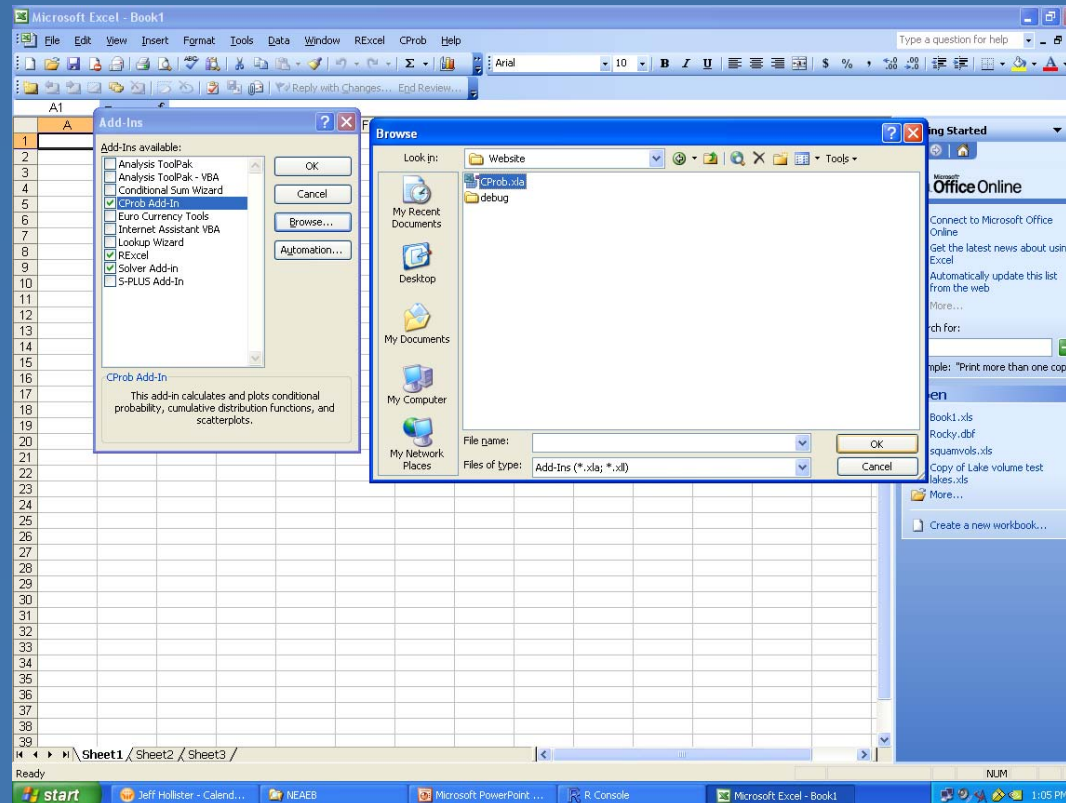
Last updated on Wednesday, November 5th, 2008.
<http://www.epa.gov/emap2/nca/html/regions/cprob/index.html>
[Print As-Is](#)

Install R and Friends:

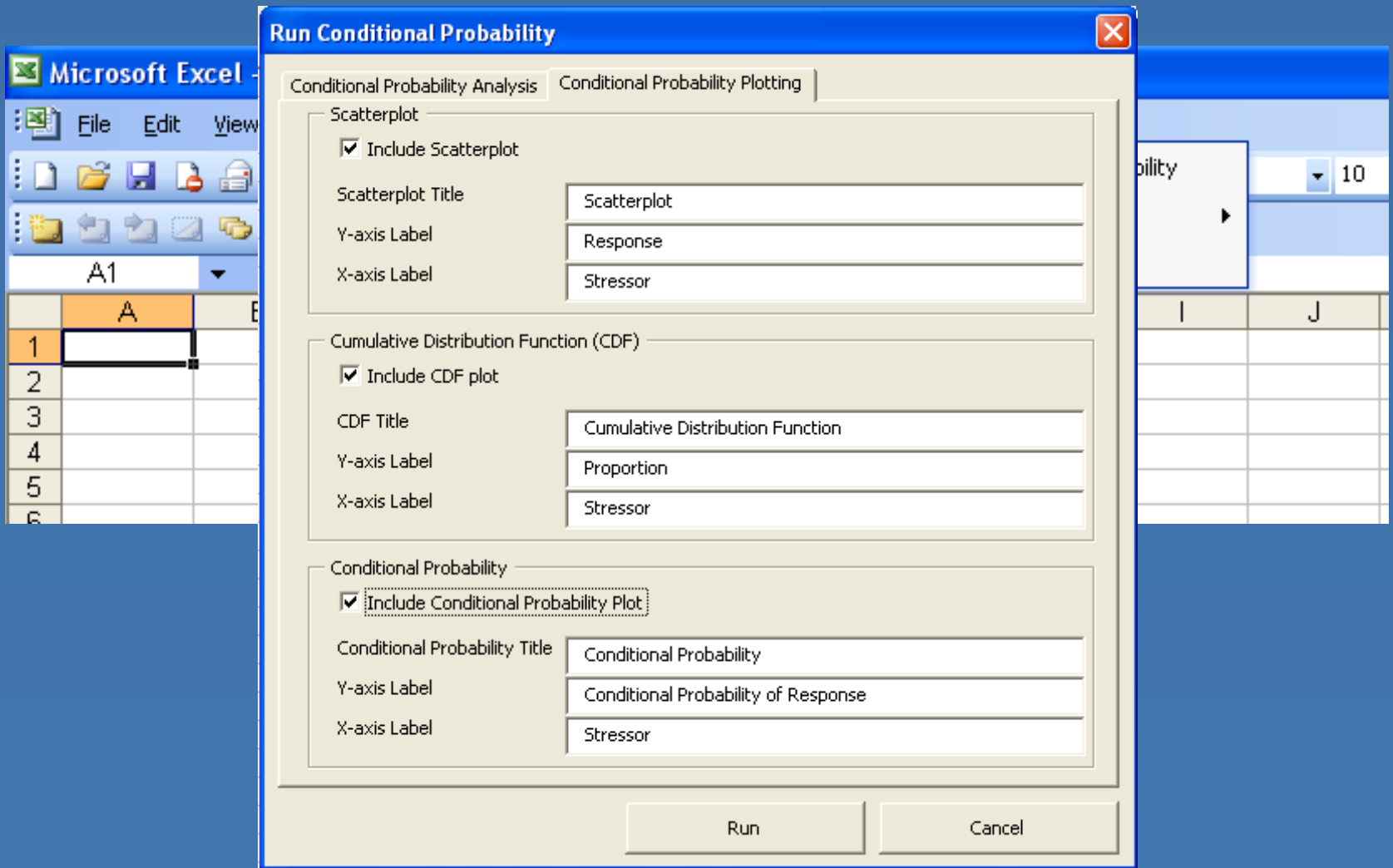


Using CProb 1.0 in Excel: Installation

Add CProb Add-in (Currently does NOT work with Office 2007):



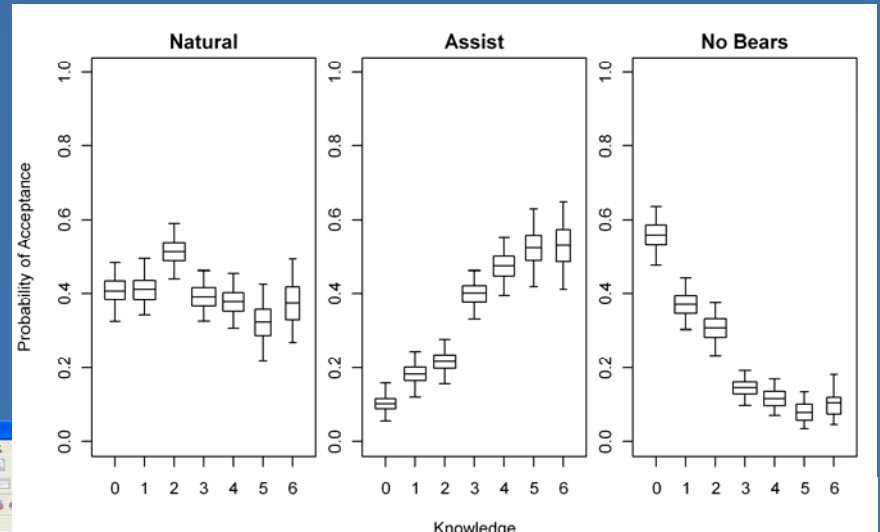
Using CProb 1.0 in Excel: User Interface



***Using CProb 1.0 in Excel:
EMAP Demo***

Using CProb 1.0 in R: Why?

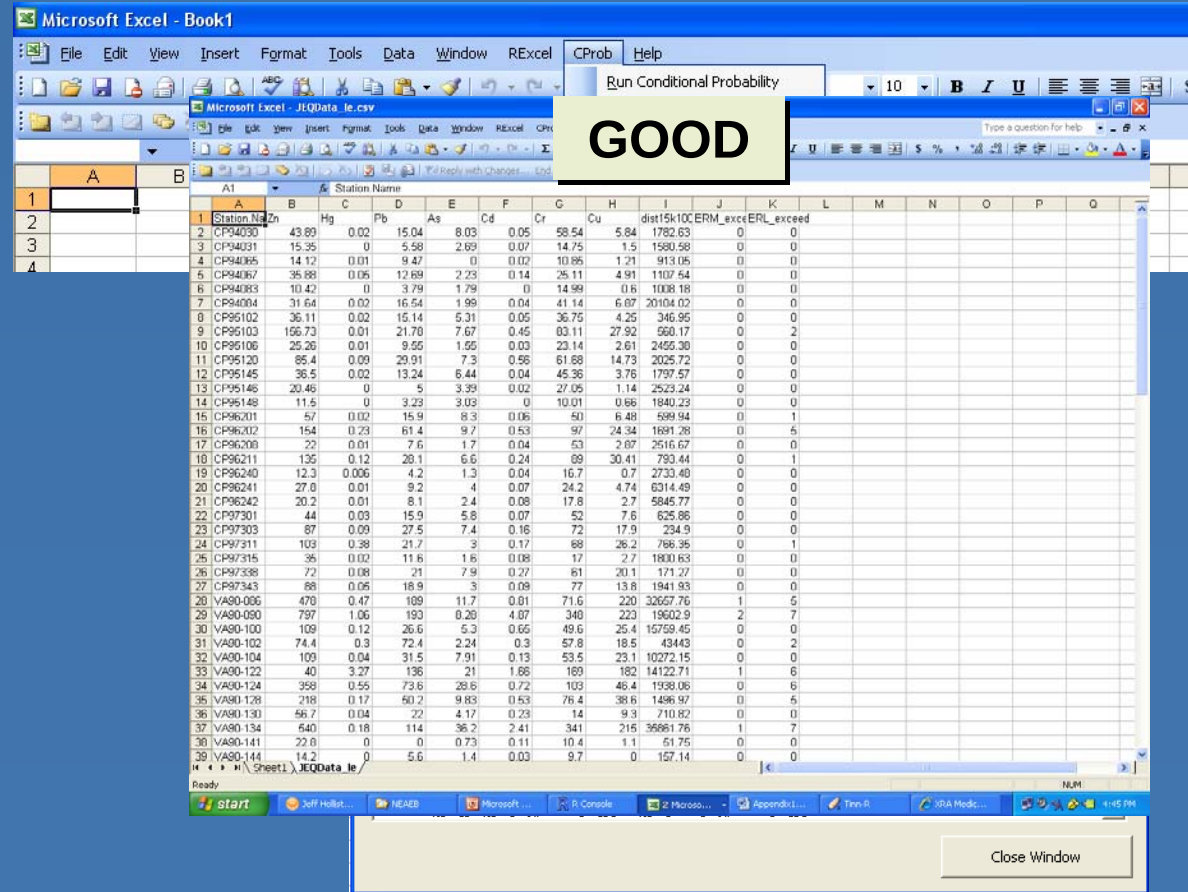
- Batch Processing
- Repeatability
- Access to R



```
Tim R: [Untitled1*]  
File Project Edit Format Search Options Tools  
R  
BeakAnalysis2R [Untitled1*]  
# R function to generate bootstrapped conditional probabilities and  
# plot boxplots for each distribution.  
# Generates figures in Appendix 1 and Table 4 in Morzillo et al.  
# Coded by J.W. Hollister, Nov-Dec 2008  
# Input data requires three input vectors with binary responses for  
# each recovery strategy (x1,x2, and x3) and an input vector with the  
# different levels of each conditioning factor (y).  
probplot3panel<-function(x1,x2,x3,y,filename,xlabel,ylabel,mtitle1,  
mtitle2,mtitle3,ci=0.05){  
# Function to calculate bootstrapped probabilities -- B=1000  
bootpc<-function(xdf){  
xbootpc<-data.frame()  
for(i in 1:1000){  
testc<-by(xdf[,2],function(x){sample(x[,1],replace=T)})  
xbootpc<-lapply(test,function(x){table(x)[2]/(sum(table(x)))})  
xbootpc<-data.frame(colnames=t(xbootpc$),unlist(xbootpc))  
xbootpc<-rbind(xbootpc,xbootpc)  
xbootpc  
}  
# Function to calculate by widths - widths used to show relative sample  
# size for each category in the boxplots  
bywidths<-function(xdf){  
sx<-by(xdf[,2],function(x){sample(x[,1],replace=T)})  
vec1<-0  
for(i in 1:length(sx)){  
vec1[i]<-sqrt(length(unlist(sx[i])))  
vec<-vec1/sum(vec1)  
vec  
}  
# Function to create plots  
plotit<-function(xbootit,xlabel,ylabel,mtitle,filename,wid,ci){  
bxp1<-boxplot(xbootit[,2]-xbootit[,1],xname=wid,plot=F)  
bxp1$names[1]<-by(xbootit,xbootit[,1],function(x){quantile(x)[.25]  
}
```

Using CProb 1.0 in R: Script and Data

- Script is accessible from Excel
- Read Script into R
 - Two Ways
 - `source("cprob.R")`
 - Copy and Paste
- Read Data into R
 - K.I.S.S.



***Using CProb 1.0 in R:
National Lakes Assessment
Demo (I Hope!!)***