How to determine the threshold

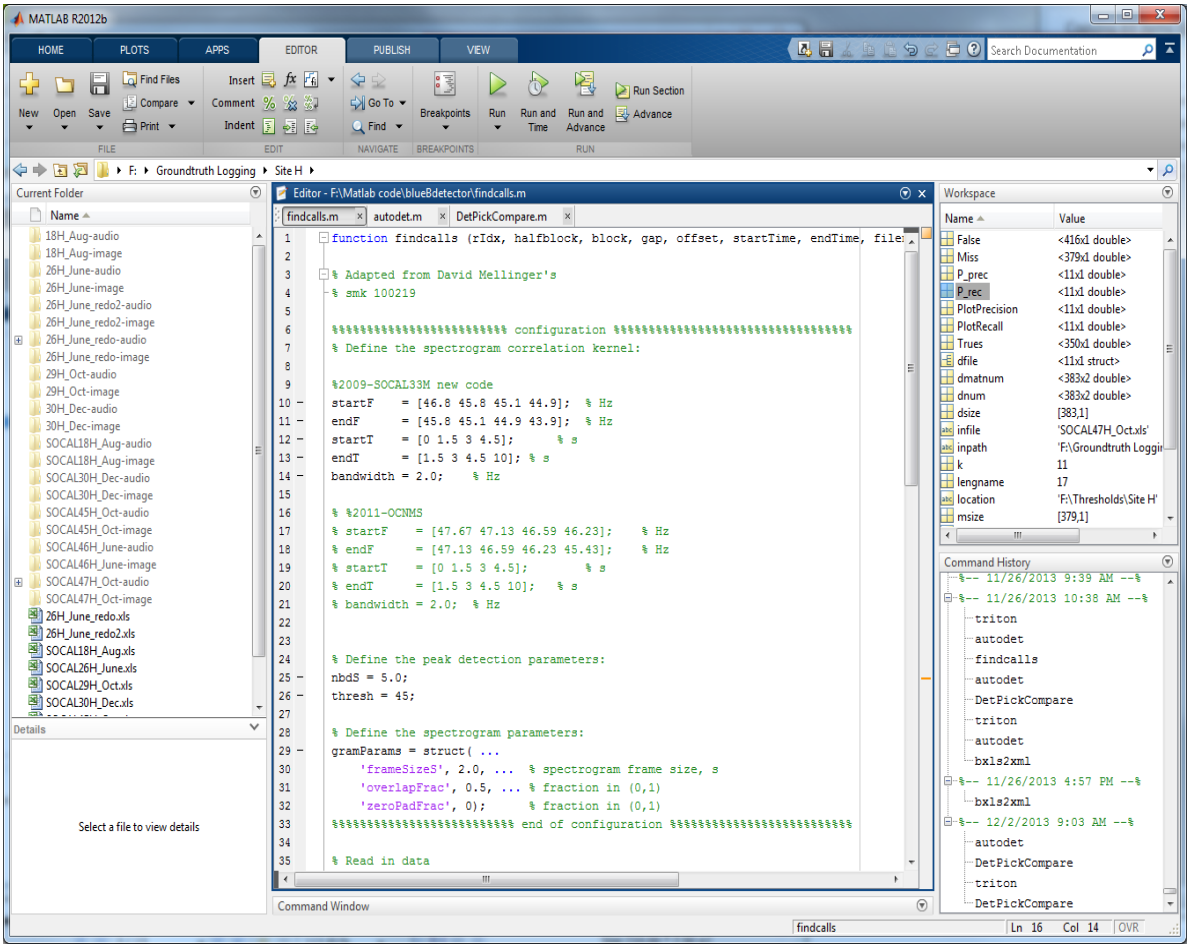
for automated blue whale B call detection

After generating a kernel and before running the detector on a dataset, you must determine the optimal threshold for that dataset. To do this you must run the detector on one xwav from the dataset at various thresholds, compare the detector's output at the various thresholds to your groundtruth (manual picks) and then choose the threshold that has the highest precision and recall.

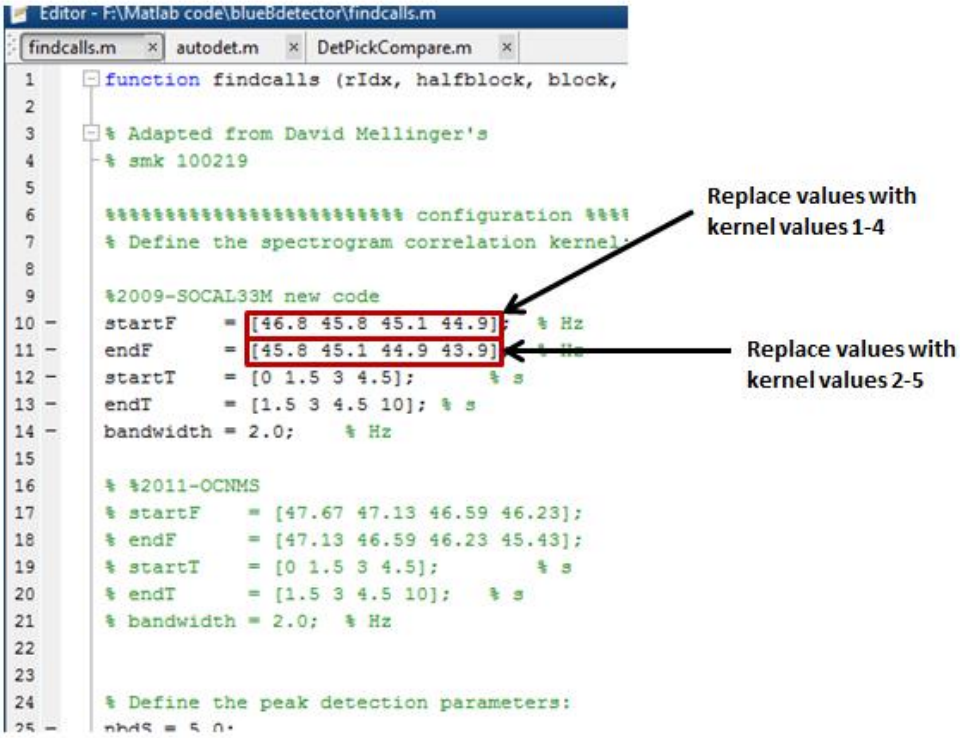
What you need:

- Matlab R2012b
- findcalls.m
- autodet.m
- DetPickCompare.m

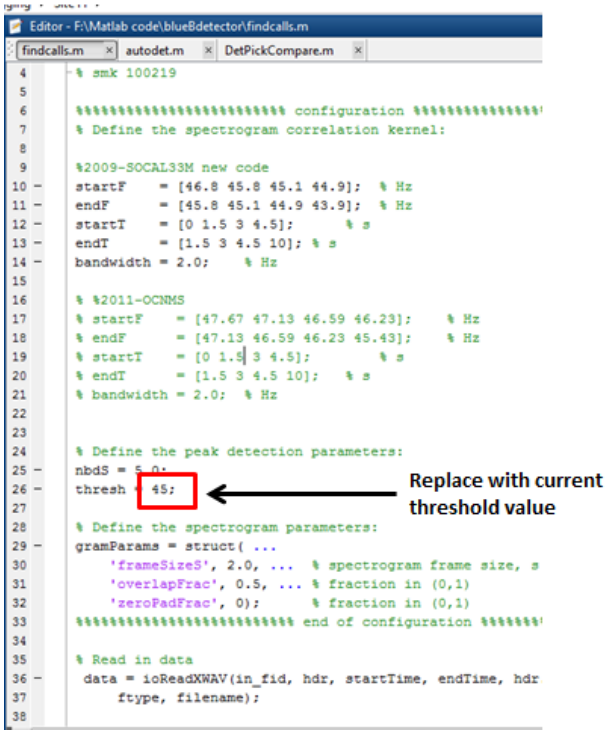
1. Open findcalls.m, autodet.m, and DetPickCompare.m in Matlab editor.



2. Click on findcalls.m and replace the kernel values in black text with the kernel values for the dataset you are working with. Your kernel should consist of 5 values. Replace the numbers in the line with “startF” with values 1-4 of your kernel. Then do the same in the next line (beginning with “endF”) except with values 2-5 of your kernel.

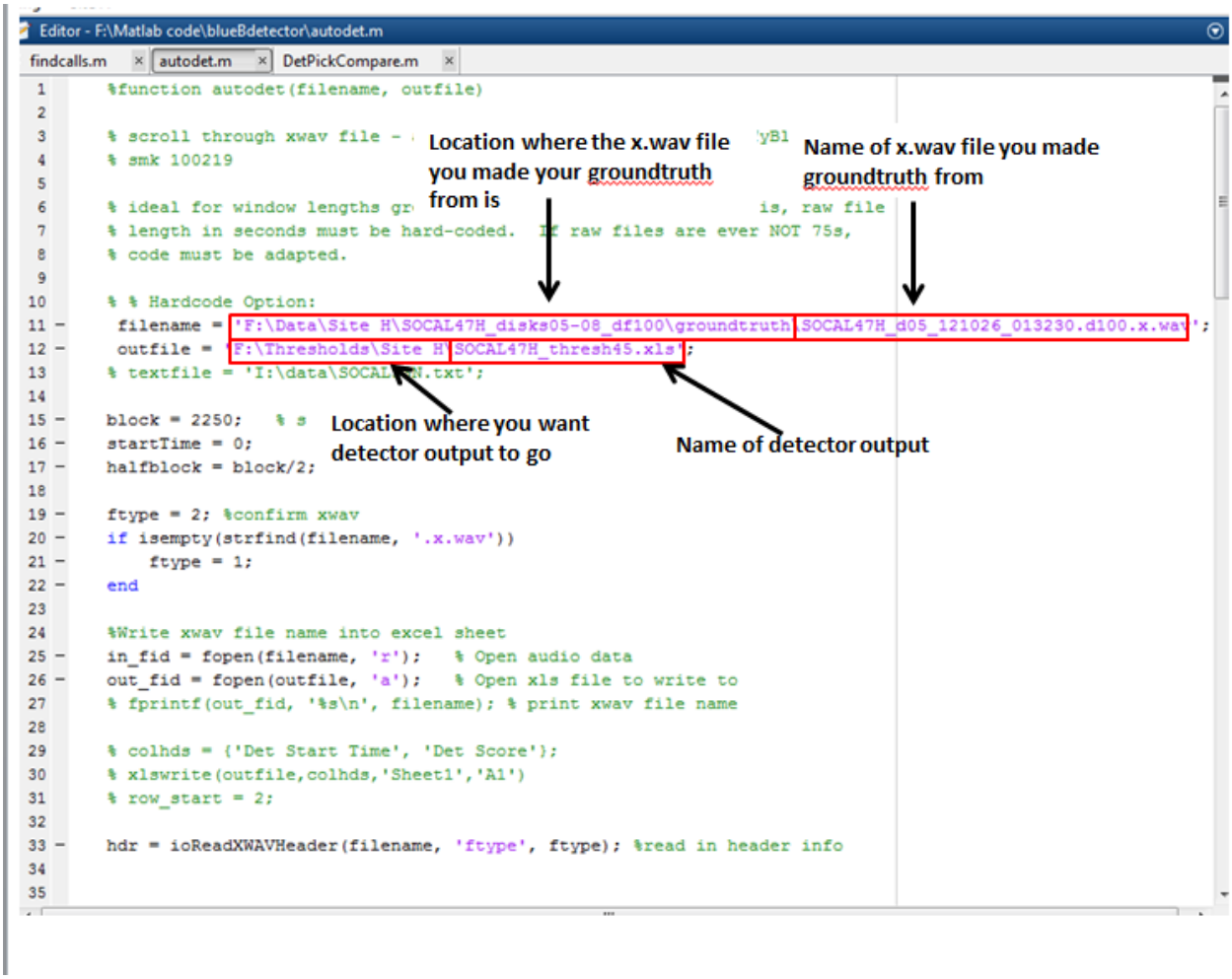


3. Scroll down a bit in findcalls.m to the line that starts “thresh =”. Replace the threshold that’s there with the threshold you are going to use (usually begin with 25).



4. **SAVE** the findcalls.m file (you will know a file has been saved if there is no asterisk after the .m – ex: findcalls.m* is unsaved, findcalls.m is saved).

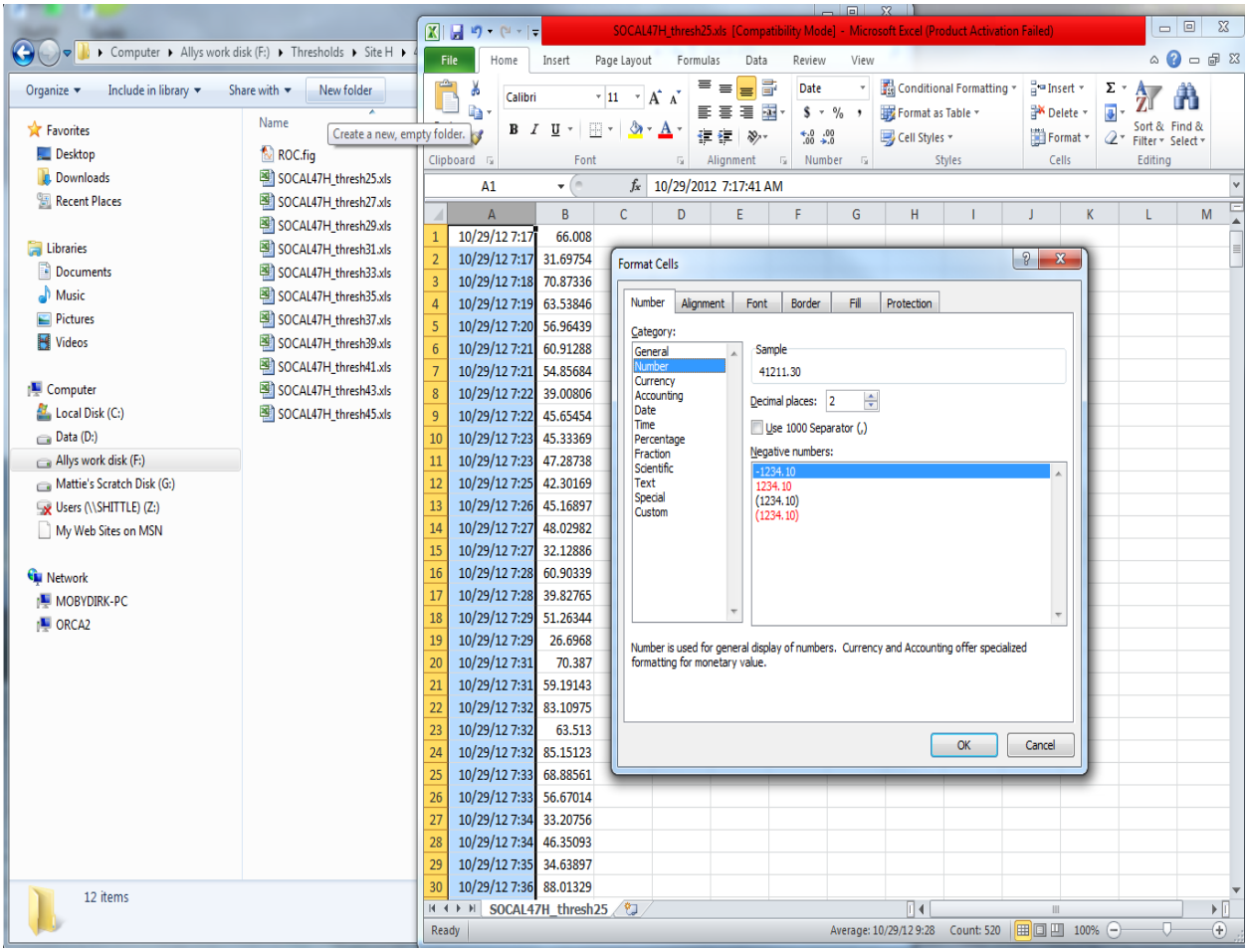
5. Click on autodet.m. Replace the purple text after "filename" with the location of the x.wav you made your groundtruth from, followed by the specific name of that x.wav file. Replace the purple text after "outfile" with the location where you would like Matlab to put the output from the detector (this will be easiest if you have a folder specifically for thresholds) and the name you would like that file to be called (typically something like SOCAL47H_thresh25.xls where you have your deployment info as well as the threshold you are using). **NOTE:** when entering locations for where Matlab should look for files and place them make sure you don't delete the ' at either end of the location)



6. Click Run.

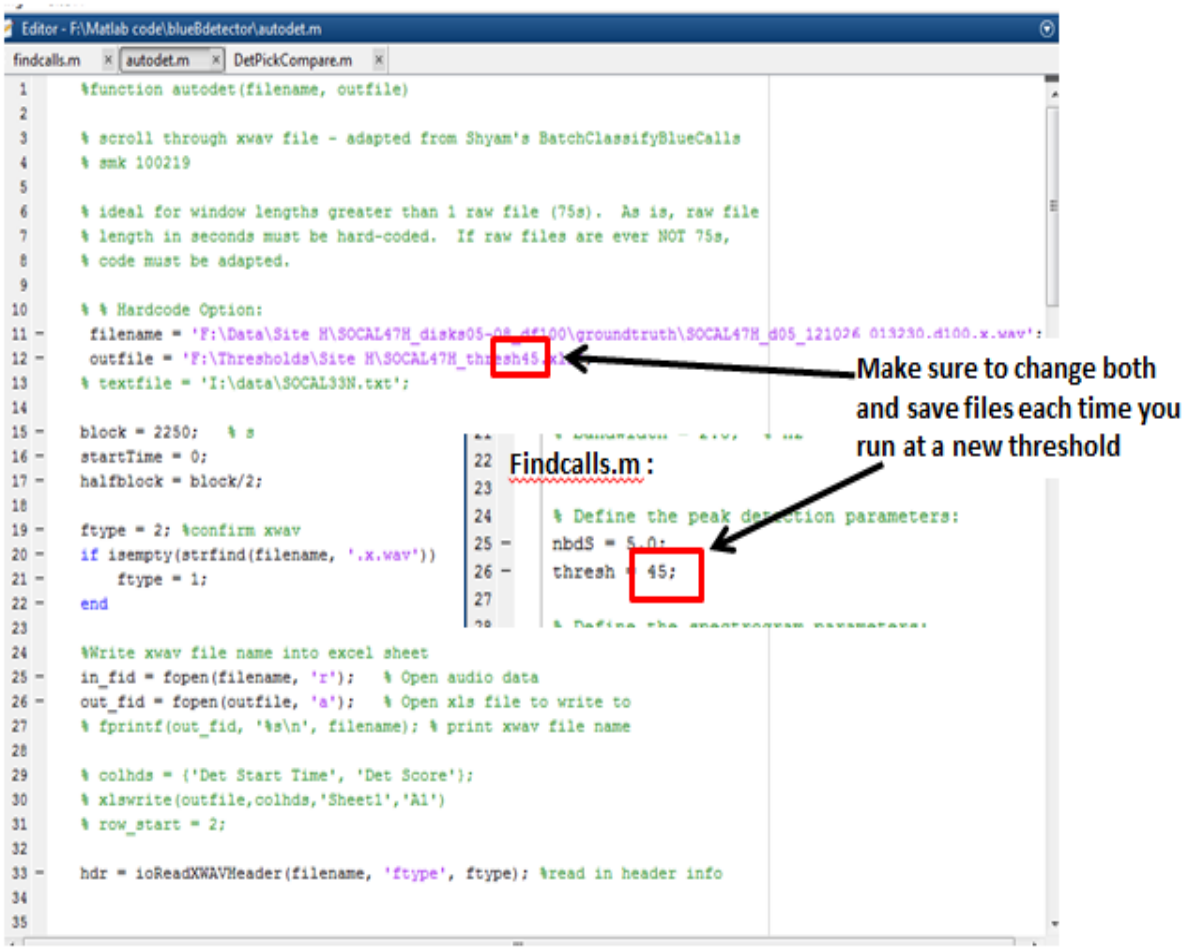
7. Matlab should start analyzing the x.wav. You will see its progress in the command window as it moves through sections of the x.wav - this will take a few minutes.

8. Once Matlab has finished running you can open the xls containing the automatic detections. Convert the cells containing detections (first column) to a time format. Using the first and last detection you noted from your groundtruth, delete all of the detections that come before and after your manual picks. Then, convert the detections to a number format.



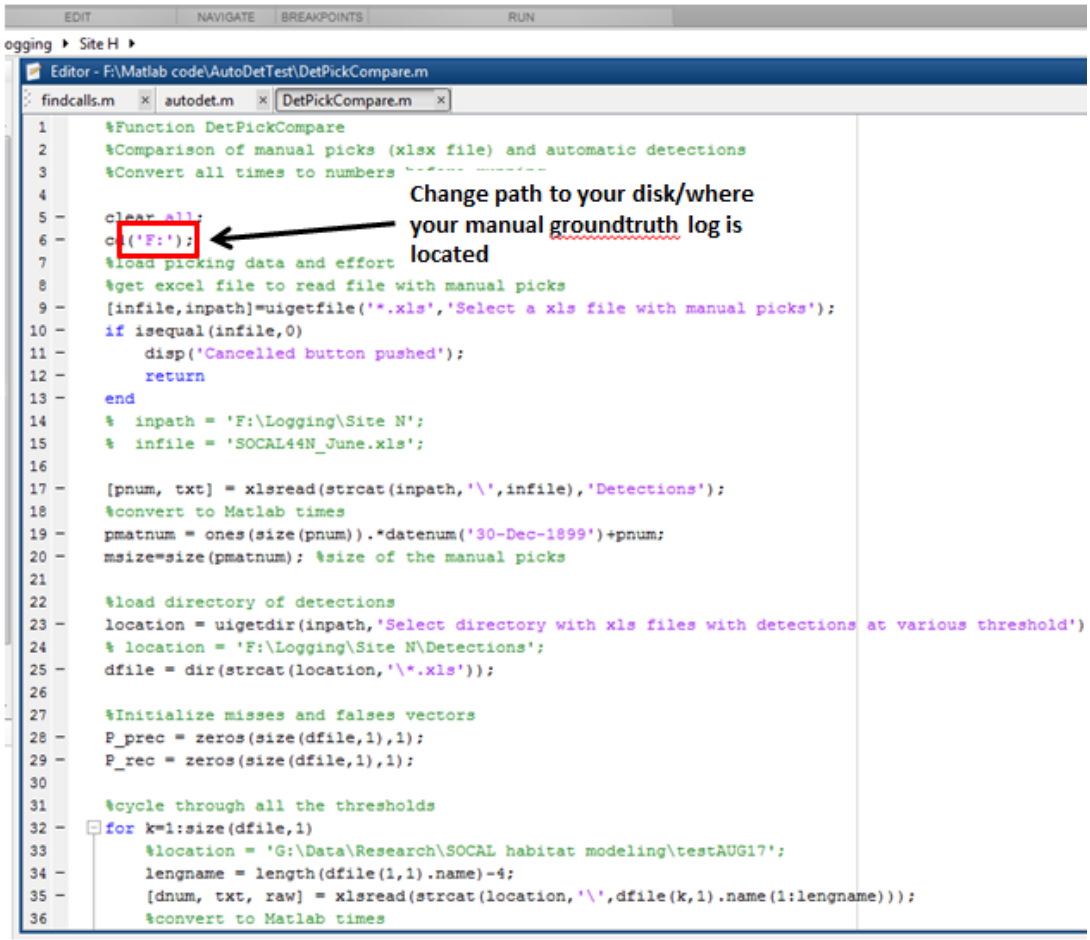
9. **SAVE** the excel sheet to replace the unedited detector output.

10. The optimal threshold should be somewhere from 25-45. Continue to run autodet.m and increase the threshold by 2 each time you run it (ex: start at 25 and change the threshold to 27 for the next run, and continue until you've run at a threshold of 45). **Make sure you change the threshold in findcalls.m and save that as well as changing your output file name in autodet.m before you run autodet.m again.**



11. Once you have run through all the thresholds, edited each xls file so that it contains only detections within the timespan of your manual groundtruth and converted all detections into a number format, click on DetPickCompare.m.

12. At line 6 you can enter the location of your disk or the location of the folder containing your manual picks.

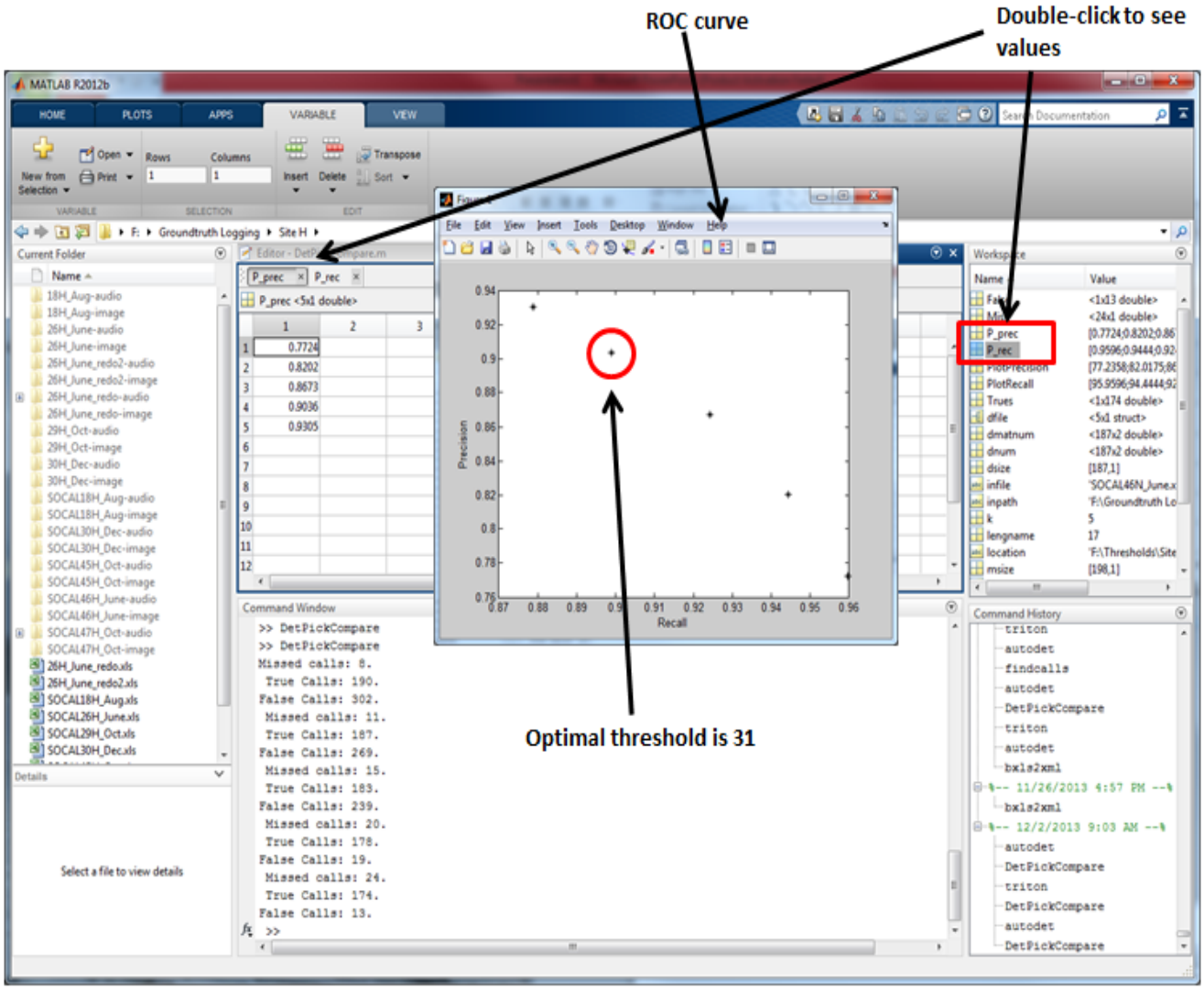



```
1   %Function DetPickCompare
2   %Comparison of manual picks (xlsx file) and automatic detections
3   %Convert all times to numbers
4
5   clear all;
6   cd('F:');
7   %load picking data and effort
8   %get excel file to read file with manual picks
9   [infile,inpath]=uigetfile('*.xls','Select a xls file with manual picks');
10  if isequal(infile,0)
11      disp('Cancelled button pushed');
12      return
13  end
14  %  inpath = 'F:\Logging\Site N';
15  %  infile = 'SOCAL44N_June.xls';
16
17  [pnum, txt] = xlsread(strcat(inpath,'\',infile),'Detections');
18  %convert to Matlab times
19  pmatnum = ones(size(pnum)).*datenum('30-Dec-1899')+pnum;
20  msize=size(pmatnum); %size of the manual picks
21
22  %load directory of detections
23  location = uigetdir(inpath,'Select directory with xls files with detections at various threshold')
24  % location = 'F:\Logging\Site N\Detections';
25  dfile = dir(strcat(location,'\*.xls'));
26
27  %Initialize misses and falses vectors
28  P_prec = zeros(size(dfile,1),1);
29  P_rec = zeros(size(dfile,1),1);
30
31  %cycle through all the thresholds
32  for k=1:size(dfile,1)
33      %location = 'G:\Data\Research\SOCAL habitat modeling\testAUG17';
34      lengname = length(dfile(1,1).name)-4;
35      [dnum, txt, raw] = xlsread(strcat(location,'\',dfile(k,1).name(1:lengname)));
36      %convert to Matlab times
```

13. Click Run. Matlab will prompt you to select an xls file with manual picks. Select the xls file that is the groundtruth for the dataset you are working with. Matlab will then prompt you to select a directory with xls files with detections at various thresholds. Select the folder where you have all of the files at different thresholds.

14. Matlab will quickly run and compare your manual picks to the detectors picks and will output the number of missed calls, true calls and false calls in the command window. Matlab will also produce a ROC curve of the different thresholds.

15. Determine which threshold is optimal using the ROC curve and the precision and recall values produced by Matlab. (Precision and Recall values will appear in the workspace window in Matlab. Double click P_prec and P_rec to see these values). **NOTE:** Ideally precision and recall will both be 0.85 or higher (try to stay close to that) – when choosing the optimal threshold keep in mind that you want to have a balance between precision and recall.



16. Save your ROC curve (either in its own folder or in the folder containing the different thresholds) and make a note somewhere of the exact precision and recall values of the threshold you are choosing.