Automated analysis of temperature variance profiles to determine inundation state of wetlands

Thomas L. Anderson, Jennifer L. Heemeyer, William E. Peterman, Michael J. Everson, Brittany

H. Ousterhout, Dana L. Drake, Raymond D. Semlitsch

Electronic Supplementary Material

Appendix A

We have provided annotated R code and example data files from iButton output to be used for example analyses at https://mospace.umsystem.edu/xmlui/ At the main interface of this website, search for the authors' last names or the title of the manuscript that leads to a file named "iButton_ESM.zip" which includes the following:

DESCRIPTION OF FILES INCLUDED:
1) iButton_SourceCode.R: This file contains the code and functions necessary to process raw iButton .txt files. **YOU SHOULD NOT HAVE TO MODIFY THIS FILE**

2) Execute_iButton.R: This file contains the lines of code that must be modified to process raw iButton .csv files.

3) iButton_Optimization_SourceCode.R: This file contains the code and functions necessary to process raw iButton .csv files. **YOU SHOULD NOT HAVE TO MODIFY THIS FILE**

4) Execute_iButtonOptimization.R: This file contains the lines of code that must be modified to process raw iButton .csv files.

5) Validation.txt: This is an example .csv showing the necessary format field validation must be in order to run the optimization code. Three columns are needed: 1) Site name, which must **exactly** match the iButton .csv filename for the corresponding site; 2) Date, written as dd/mm/yyy; 3) Status; either "Wet" or "Dry" (case sensitive).

6) Raw iButton.txt files. We provide 5 raw files so users can run the code (numeric files names)

OVERVIEW:
The code and files are provided as is, and the authors have limited ability to troubleshoot issues with running this code. If you have question, please address them to Bill Peterman (Bill.Peterman@gmail.com).

Further files will be updated at https://github.com/wpeterman/iBdry

We highly recommend that you install and/or update R and all R packages to the most recent versions. This code was last run/checked using R version 3.0.2. Additionally, if users are not familiar with R, we highly recommend downloading and using Rstudio (http://www.rstudio.com/). NOTE: in practice, raw iButton data generally is in .csv format, and the

ANALYZING iBUTTON DATA
1) After downloading all of your iButton data to .txt files (we use the provided OneWire software), save all files together in a unique folder
2) Open the "Execute_iButton.R" file in R
3) Modify each line of code as needed.
   - Note that only text to the right of the "<-" symbol should be modified.
   - All dates, text, and directories should be enclosed in quotes (e.g. "01/01/2001")
   - When giving the path to a directory (i.e. folder), make sure to end with a forward slash (/)

4) After modifying each line, simply run the last line of code: source(iButton.Source) to execute the function and process your raw iButton files

WHAT THE CODE DOES
This code will first import and process each iButton.csv file in turn. After reading each file and trimming it to the specified dates, the variance in temperature is calculated for each day of deployment. Then, a moving window slides over the data and calculates either the mean or variance of the daily variance measures as the window moves over the data and determines whether or not the calculated value in the window is above or below the specified variance threshold. Finally, a second moving window (consecutive days window) slides over the data to determine the number of consecutive days that the variance threshold was crossed.

Following processing, four output files will be generated:
1) A .csv file containing the raw iButton data for each pond will be exported. We have found that working with the date-time field can be challenging in Excel; this file has columns formatted to alleviate these issues.

2) A pdf file containing two plots (Fig. 1). The top plot shows the daily variance measures; the bottom plot shows the variance measure (mean or variance of the variance) calculated within the moving window. On both plots, the variance threshold is drawn as a red dashed horizontal line, and detected fill-dry events are marked by vertical black lines.

3) A "Dry_Fill_Events.csv" file will be exported. Each row in this table represents a site, and columns are the dates of dry-fill events. Note that because of the moving window analysis, the exact start/end of a dry-fill event may not align perfectly with these events in the field. The initial state of each site is indicated, and each subsequent event date indicates a switch. For example, if initial state is "Filled", the date in the Event1 column indicates the date of drying.

4) An "All_iButtons_Combine.csv" containing all of the data from each iButton file.

OPTIMIZING iBUTTON DATA

Directions are the same as listed above, and necessary notes and comments are annotated in the "Execute_iButtonOptimization.R" file. Most important is the formatting of the file used for validations. Please reference the provided example.
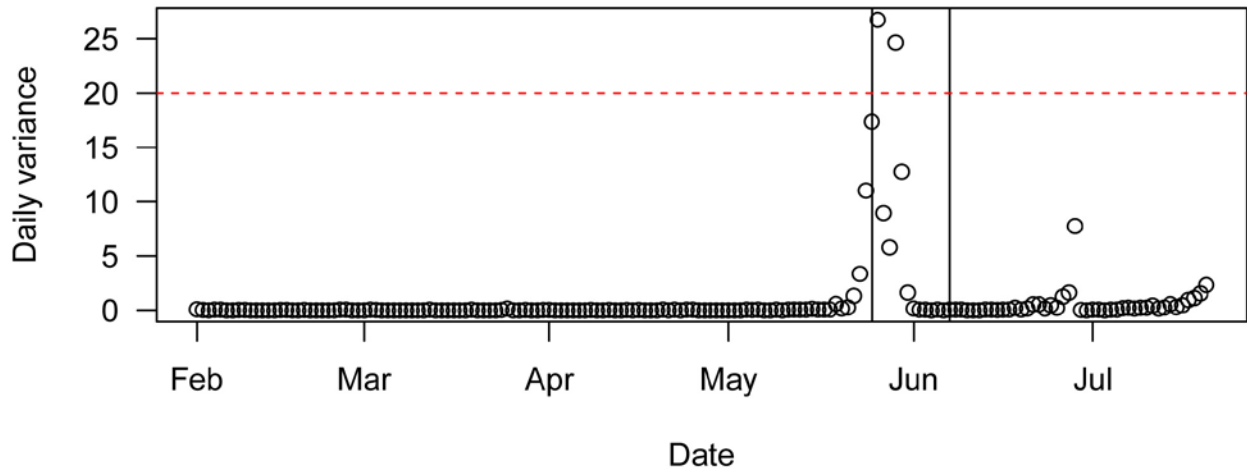
Once executed, this code will run through all combinations of the values specified for variance, moving window, and consecutive days. The code automatically tests both "mean" and "variance" methods.

This code is a little cumbersome, and the total number of parameter combinations can get large fast. Nonetheless, it is possible to search >2,000 parameter combinations on >50 iButton files in a few hours.

This code processes each file using the specified parameters, determining whether or not a pond was full or dry at each date. For each site-date combination in the validation file, it is determined whether or not the combination of parameters correctly identified the state of the pond. This assessment is done through the calculation of sensitivity (proportion predicted to be dry that were actually dry) and specificity (the proportion predicted to be wet that were actually wet). We then added sensitivity to specificity, and used this as our optimization criteria. The output file "Sensitivtiy_Specificty_Optimization.csv" contains these calculated statistics, and is sorted by rank order of best parameter combination to worst.

Figure 1. Example of figure produced after processing raw iButton data. The top plot shows the daily variance measures; the bottom plot shows the variance measure (mean or variance of the variance) calculated within the moving window. On both plots, the variance threshold is drawn as a red dashed horizontal line, and detected fill-dry events are marked by vertical black lines.
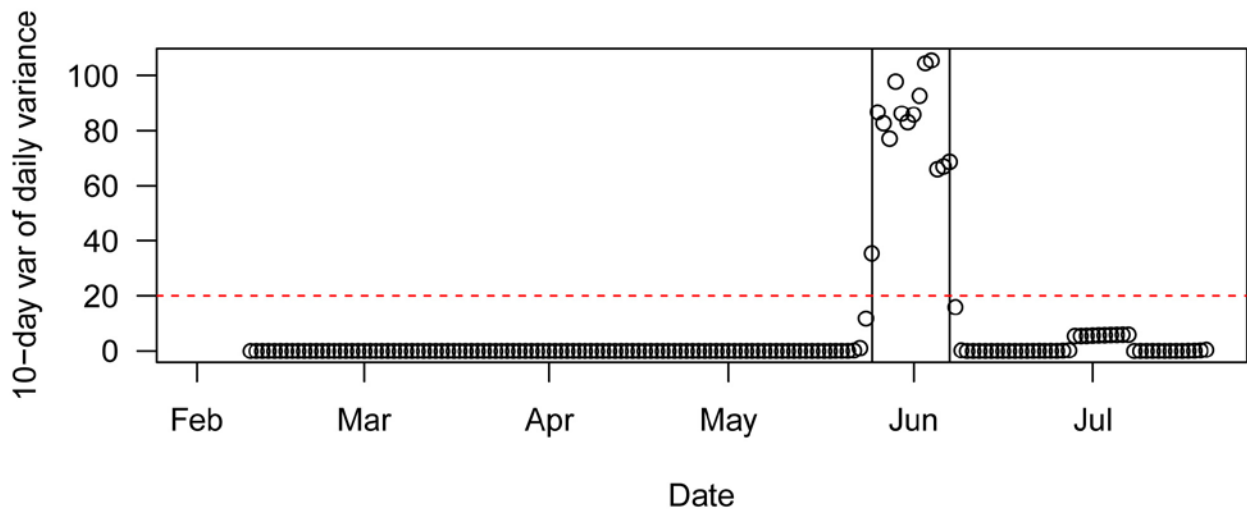
Table 1: Example summary output from automated R code (see ESM) that analyzed temperature variance profiles of iButtons for a subset of ponds. 'Initial State' is whether the pond had water at the time of deployment, 'Days Deployed' is the length of time that the iButtons were in a pond, 'Days Wet' is the number of days the pond had water (i.e. hydroperiod), 'Days Dry' is the number of days the pond was considered dry, and the 'Number of Drying/Filling Events' is the number of times variance increased above/below the specified threshold (see ESM). Parameters were set to a window size of 11 days, a threshold of 22, and a 2-day consecutive threshold that had to be exceeded for a switch in inundation state (see ESM). Ponds 333 and 120 correspond to Figure 3a, Ponds 110 and 237 correspond to Figure 3b, Ponds 120 and 278 correspond to Figure 3c, and Ponds 57, 122, 242, 255 correspond to Figure 3e.

Pond	Start Date	Initial State	Days Deployed	Days Wet	Days Dry	Number of Drying Events	Number of Filling events
110	7/5/2012	Dry	150	84	66	0	1
120	7/5/2012	Dry	150	0	150	0	0
122	7/5/2012	Filled	150	150	0	0	0
237	7/5/2012	Dry	150	98	52	0	1
242	7/5/2012	Filled	150	141	9	1	1
255	7/5/2012	Filled	150	150	0	0	0
278	7/5/2012	Dry	150	74	76	2	3
333	7/5/2012	Dry	150	14	136	2	2
57	7/5/2012	Filled	150	150	0	0	0

Appendix B: Pictures of Experiment 2 Set-Up

Automated analysis of temperature variance profiles to determine inundation state of wetlands

Thomas L. Anderson, Jennifer L. Heemeyer, William E. Peterman, Michael J. Everson, Brittany

H. Ousterhout, Dana L. Drake, Raymond D. Semlitsch

Design of Experiment 2:

Figure B1: Experiment 2 set up and design. The top two panels are the open (top left) and shaded (top right) treatments, and the bottom two panels show iButton attached and set up.

1   **Appendix C: Initial Analysis and Results of Temperature Variance**

Automated analysis of temperature variance profiles to determine inundation state of wetlands

Thomas L. Anderson, Jennifer L. Heemeyer, William E. Peterman, Michael Everson, Brittany H. Ousterhout, Dana L. Drake, Raymond D. Semlitsch

2

3

4   <u>Methods and Initial Analysis of Temperature Variance</u>

5   Our primary metric to determine inundation state of each pond were daily variances in

6   water temperature. We expected that high daily variance would be associated with dry ponds,

7   and low variance with wet ponds, due to the buffering capacity of water to fluctuations in

8   ambient air temperatures (Sowder and Steel 2012). We tested the accuracy of prediction for the

9   inundation state of ponds (wet or dry) in the field study by performing three single-day visits to

10   ponds in the fall of 2012 (n = 27 ponds on 30 August and all ponds with iButtons twice during

11   October). We initially tested the accuracy of six haphazardly selected variance thresholds in

12   August (s^2 = 10, 13, 14, 15, 20, and 25; i.e. ponds were predicted to be dry when daily exceeded

13   each of these values), and seven thresholds for both October visits (s^2 = 5, 10, 12, 14, 15, 20, 25).

14   These values were chosen as they represented actual daily temperature values recorded by

15   iButtons verified to be underwater when ponds were visited in October. We visually confirmed

16   and compared the observed inundation state at these three visits to the calculated daily variance

17   in the previous 24 hrs from the iButton data, and totaled the number of correct predictions at

18   each variance threshold. We then examined three pond characteristics of the incorrectly

19   predicted ponds (pond area, depth and canopy closure) in the two October visits where we had

20   the largest sample sizes, which we expected to influence the accuracy of our methodology.

21   Measurements of pond area, depth and canopy closure are described in Peterman et al. (2014).

22   We tested whether these three factors influence the agreement of the predicted inundation state

23   with the observed state using a generalized linear model with a binomial error structure (i.e.

24   logistic regression with a binary response, "Agree" or "Disgree"). We tested for the effects of

25   depth, area and canopy closure for all variance thresholds combined, as well as each individual

26   variance threshold.

27

28   <u>Results of initial analysis for field-deployed iButtons</u>

29   Of the 159 iButtons that were deployed from July–December 2012, 130 were recovered

30   from the field, and data were successfully downloaded from 95 of them (27% failure rate of

31   iButtons that were recovered). Of the 27 ponds visited on 30 August 2012, 25 of the ponds with

32   iButtons were dry and two were inundated; the mean variance for the iButtons in the dry ponds

33   was 103.8 ± 74.59 SD, and 10.0 and 1.4 for the two inundated ponds, respectively. We found

34   that the greatest number of correct predictions was produced with variance thresholds of 13 to 15

35   (93% accuracy) in comparison to thresholds of 10, 20 and 25 which resulted in 88% accuracy

36   (Figure A1). For thresholds 13-25, all of the incorrect assignments of inundation were predicted

37   as wet when the ponds were dry (Figure A1). Using a variance threshold of 10, one pond with a

38   variance of 10.04 was incorrectly predicted to be dry.

39   Combining the two visits in October (n = 169 observations), 139 ponds were inundated

40   with water and 30 were dry. The inundation state was correctly predicted 80% (n=135) of the

41   time for all variance thresholds except for $s^2 = 5$ (69% correct; A1). The proportion of wet and

42 dry incorrect predictions changed based on the variance threshold that was tested, where higher

43 variance thresholds resulted in ponds being incorrectly predicted to be wet and lower variance

44 thresholds incorrectly predicted ponds to be dry (Figure A1). Four of the incorrect dry

45 predictions were due to the iButton placement; they were placed along the shoreline which dried,

46 but water remained in deeper portions of the pond. Therefore, while we categorized these

47 situations as incorrect in our assessment, the iButtons effectively captured a mimicked drying

48 event as the shoreline receded and refilled. For the other incorrect ponds, they generally had

49 very small pond areas (median area: 63m, Figure A2), shallow water depths (< 0.3 m, Figure

50 A3), or had either high or low canopy closure (Figure A4). Across all variance thresholds,

51 canopy closure, area and depth significantly predicted whether ponds agreed with the true

52 inundation state (depth: $p < 0.001$; canopy: $p < 0.001$; area: $p = 0.05$). Individual thresholds were

53 more variable, however, where canopy was only significant for thresholds greater than 12 (15: p

54 $= 0.01$; 20: $p = 0.005$; 25: $p = 0.008$), depth significant for a variance threshold of 5 ($p = 0.01$),

55 and area was not significant for any individual thresholds.

56

57

58

59

60

61

62

63

64
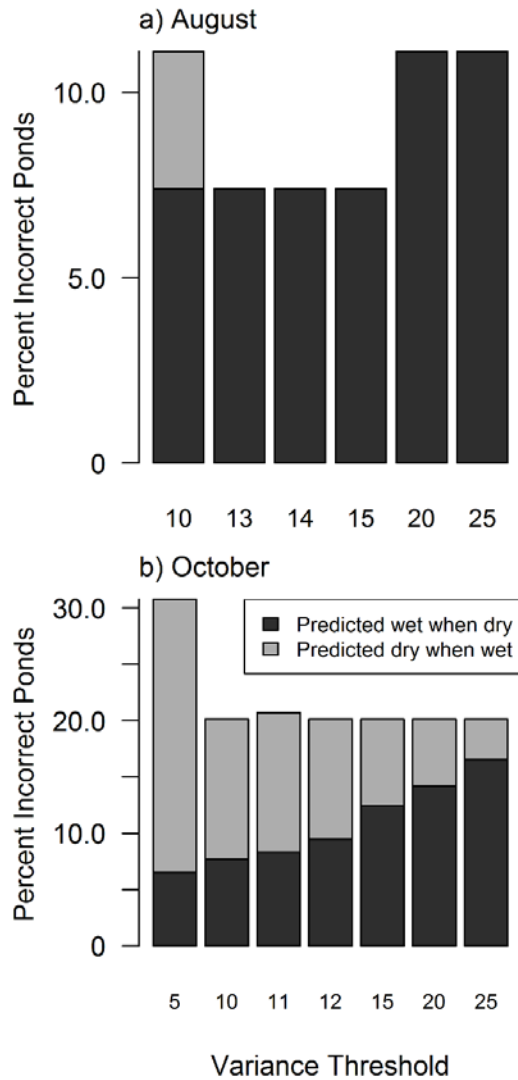
Fig. A1 (a) The percentage of predicted inundation states that were incorrect when compared with field observation on 30 August 2012 at variances of 10, 13, 14, 15, 20, and 25. (b) The percentage of predicted inundation states that were incorrect when compared with field observation conducted at two separate time points in October 2012 at variance thresholds of 5, 10, 11, 12, 15, 20, and 25. Dark gray bars represent the number of ponds predicted as wet when the ponds were dry, and light gray bars represent ponds predicted as dry when the ponds were wet
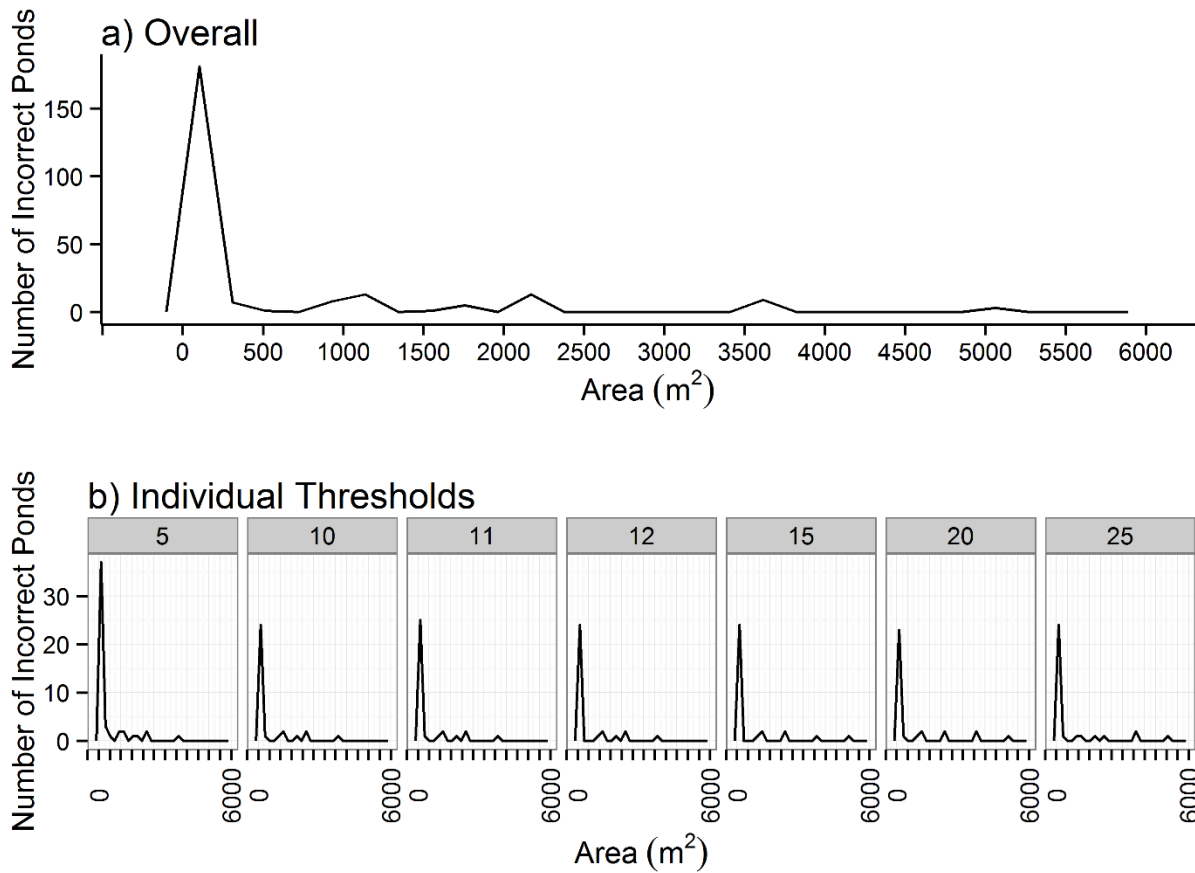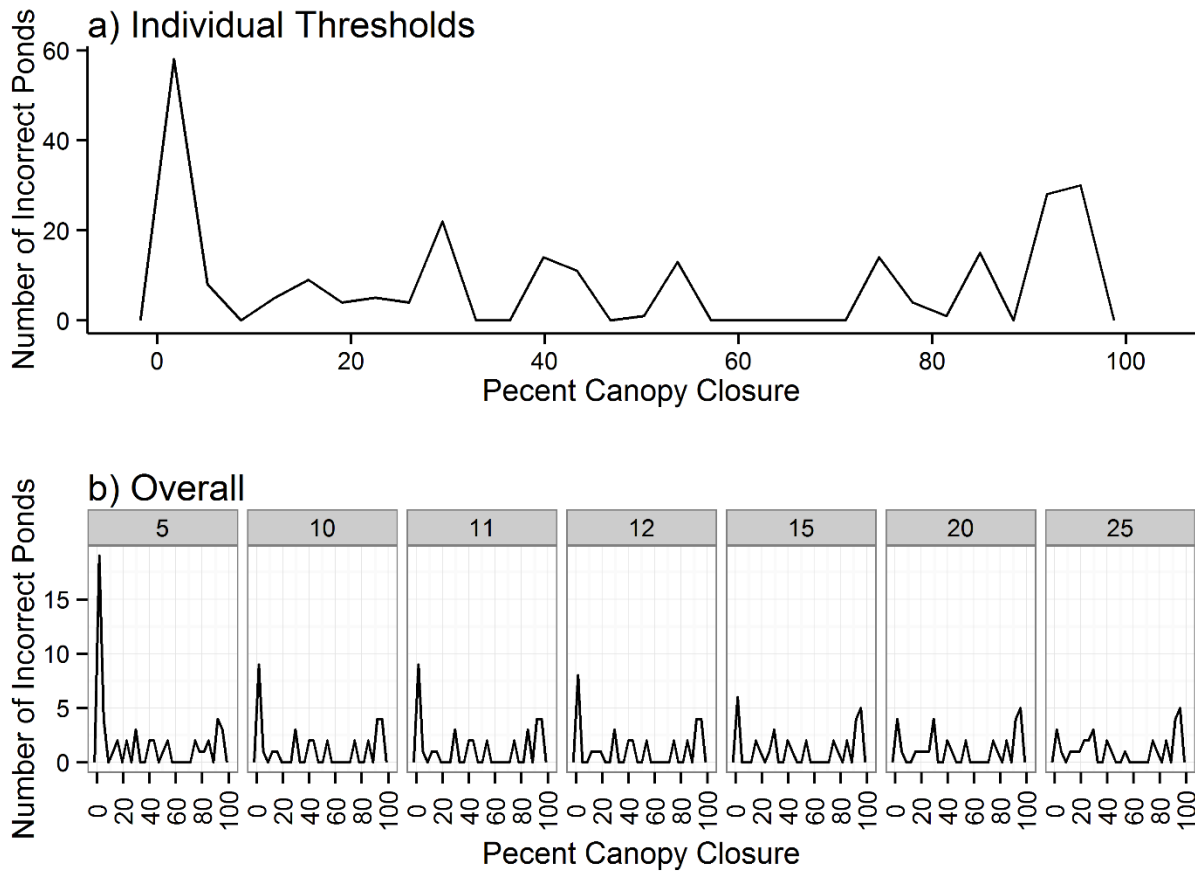
Fig. A2 The number of ponds with incorrectly predicted inundation states in October 2012 based on pond area. The top panel shows overall the number of ponds (i.e. count total across all variance thresholds), and the bottom six panels indicate the number of incorrect ponds grouped by the variance threshold that was tested, indicated by the gray number above each panel. One outlier was omitted from the graph that had an area of 12,000 m^2. $N = 257$ total incorrect predictions

a) Overall

b) Individual Thresholds

80
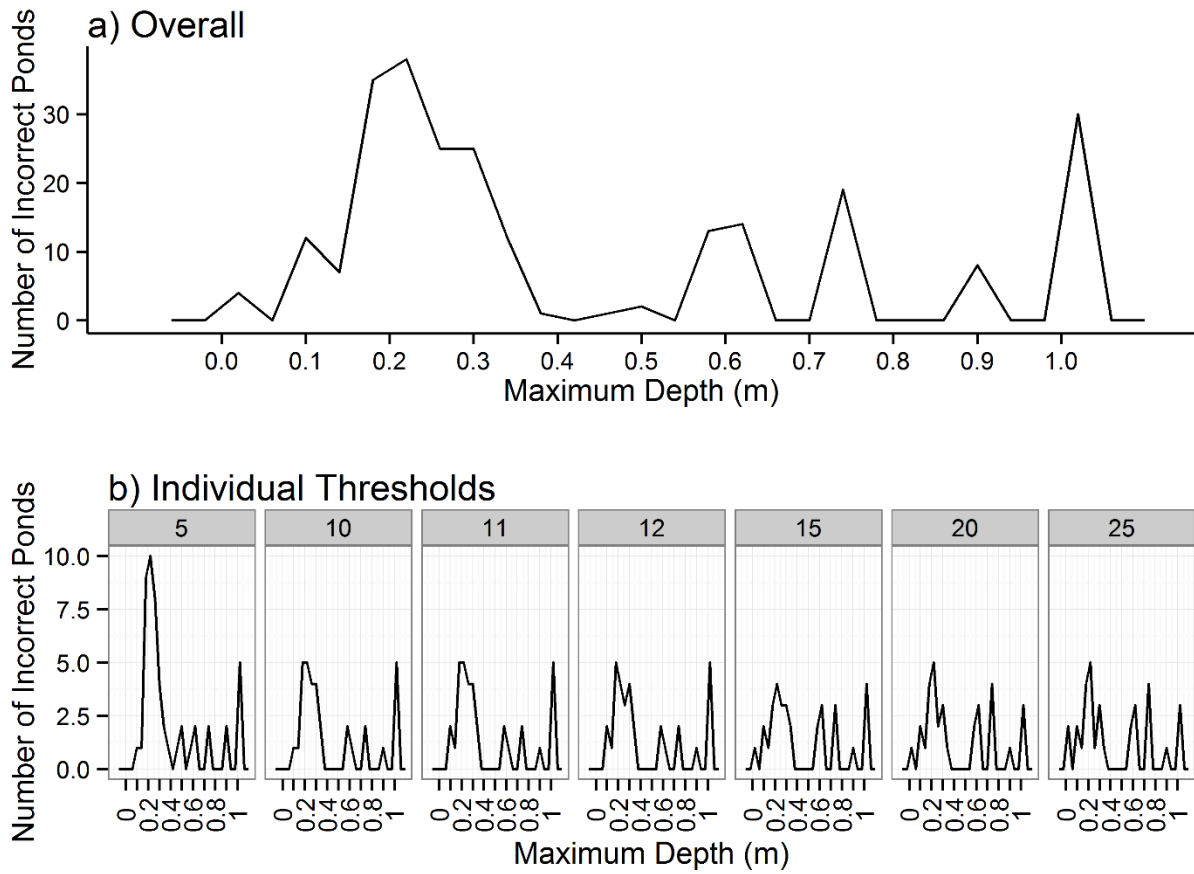
Fig. A3 The number of ponds with incorrectly predicted inundation states in October 2012 based
on maximum pond depth. The top panel shows overall the number of ponds (i.e. count total
across all variance thresholds), and the bottom six panels indicate the number of incorrect ponds
grouped by the variance threshold that was tested, indicated by the gray number above each
panel. $N = 257$ total incorrect predictions

Fig. A4 The number of ponds with incorrectly predicted inundation states in October 2012 based on percent canopy closure. The top panel shows overall the number of ponds (i.e. count total across all variance thresholds), and the bottom six panels indicate the number of incorrect ponds grouped by the variance threshold that was tested, indicated by the number within the gray box above each panel. $N = 257$ total incorrect predictions