Virtual Beach 3.0.6 – Building a GBM Model

Building, Evaluating and Validating 'Anytime Nowcast' Models

In this module you will learn how to:

A. Build and evaluate an anytime 'GBM' model
B. Optimize a GBM model by removing variables
C. View a GBM model within the Prediction tab
D. Import a historical data table to validate your models
E. Save a GBM model as a new project file

☛ **Virtual Beach 3** project files (.vb3p) allow users to save their work at any stage of the model building, evaluation, or refinement process, and to share their work with other users. Imported data are saved within the project, so the file is *stand-alone*, and can be shared without any other files.

☛ When re-opening a **Virtual Beach 3** project file, ("filename.vb3p"), the beach location and orientation are saved, but that the default base map is displayed. All data, manipulations, and transformations are saved.

A. Build and evaluate an 'Anytime' GBM model

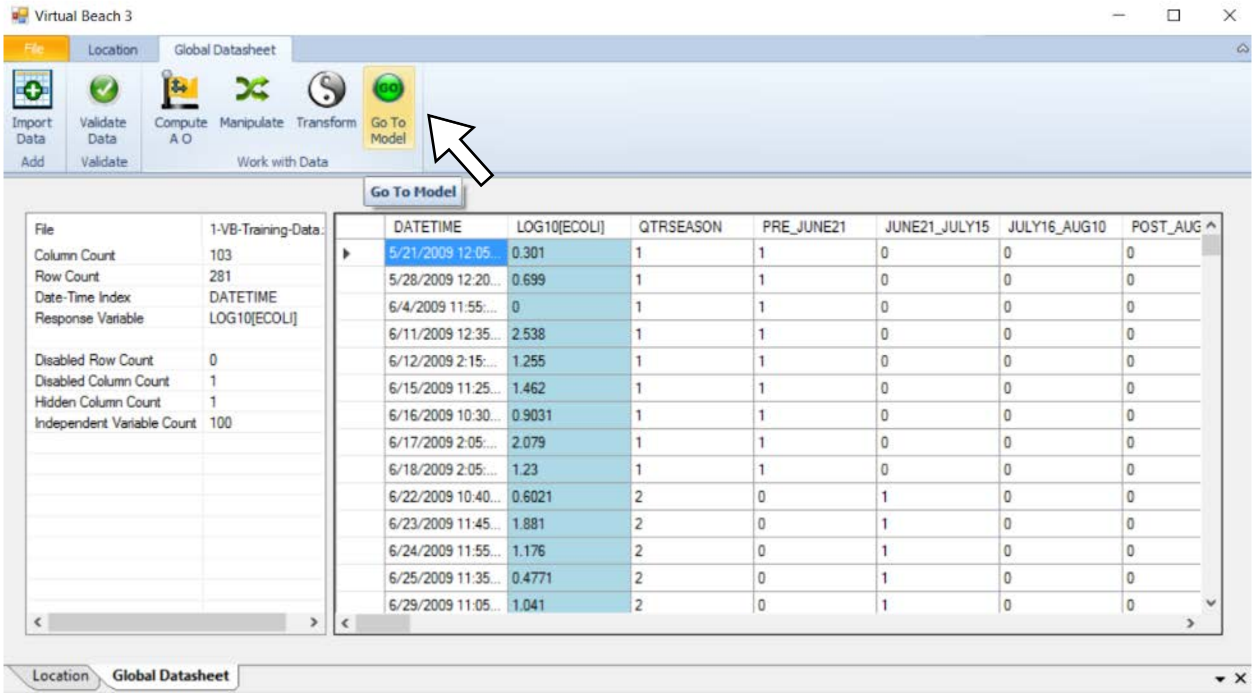
Many nowcast models operate using a combination of field-collected and web-based data. Two recent advances allow efficient development and cost-effective operation of nowcast models using *only* web-based data. Having a separate model that relies only on web-based data allows users to operate the model on days when field-collected data may not be available.

The first is the Environmental Data Discovery and Transformation (EnDDaT) Web data portal (https://cida.usgs.gov/enddat/dataDiscovery.jsp). Developed by the USGS Office of Water Information, EnDDaT enables Virtual Beach users to download location-specific, pre-processed data on any number of hydro-meteorological variables (e.g., radar-estimated rainfall, currents, waves, river discharge, etc.) from NOAA and USGS, anywhere in the Great Lakes and eventually other parts of the country.

The second is **Virtual Beach 3**'s incorporation of the Gradient-Boosting Machine (GBM) as an alternative method of developing and operating nowcast models. GBM is a "machine learning" technique for the rapid development of complex decision-tree models. Unlike the more traditional Multiple-Linear Regression (MLR) model, the GBM model is not limited in the number of explanatory variables that can be used. This allows users to create models with a much larger array of online data.

GBM does not require the variables to be independent of one another nor have a linear relationship to the response variable. This method performs best with data sets >100 observations.

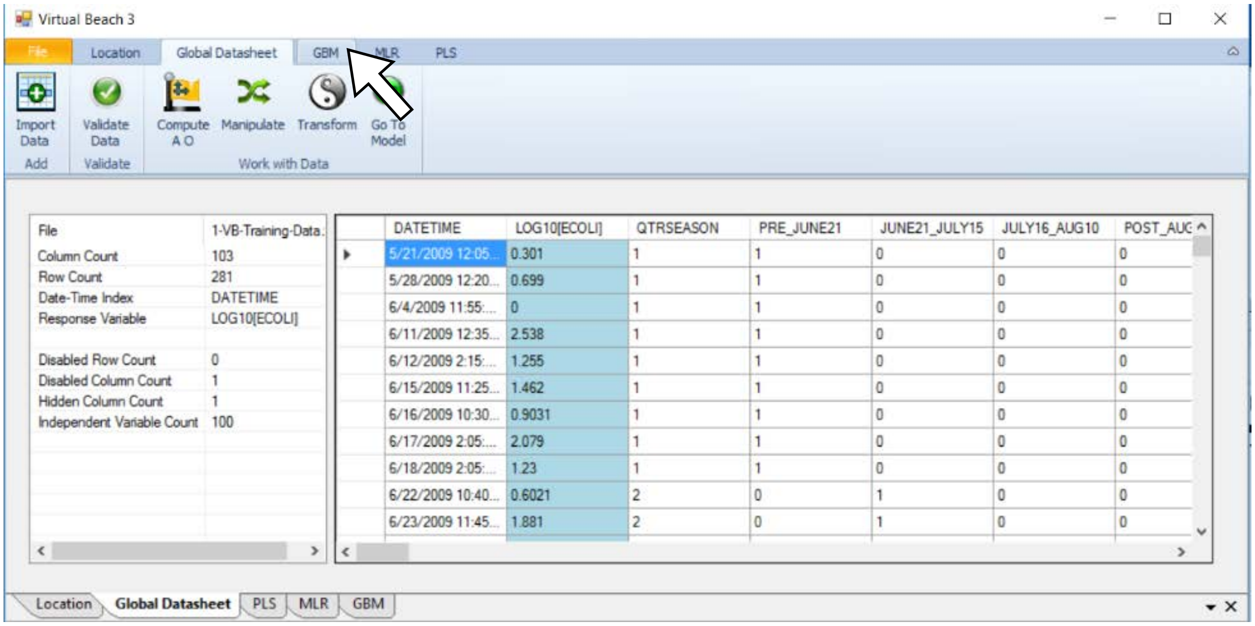
A.1. Open the file saved at the end of the **Virtual Beach 3.0.6 Data Prep-GBM** module. In the Global Datasheet tab, click the **Go to Model** icon.



File	1-VB-Training-Data.
Column Count	103
Row Count	281
Date-Time Index	DATETIME
Response Variable	LOG10[ECOLI]
Disabled Row Count	0
Disabled Column Count	1
Hidden Column Count	1
Independent Variable Count	100

DATETIME	LOG10[ECOLI]	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG
5/21/2009 12:05...	0.301	1	1	0	0	0
5/28/2009 12:20...	0.699	1	1	0	0	0
6/4/2009 11:55:...	0	1	1	0	0	0
6/11/2009 12:35...	2.538	1	1	0	0	0
6/12/2009 2:15:...	1.255	1	1	0	0	0
6/15/2009 11:25...	1.462	1	1	0	0	0
6/16/2009 10:30...	0.9031	1	1	0	0	0
6/17/2009 2:05:...	2.079	1	1	0	0	0
6/18/2009 2:05:...	1.23	1	1	0	0	0
6/22/2009 10:40...	0.6021	2	0	1	0	0
6/23/2009 11:45...	1.881	2	0	1	0	0
6/24/2009 11:55...	1.176	2	0	1	0	0
6/25/2009 11:35...	0.4771	2	0	1	0	0
6/29/2009 11:05...	1.041	2	0	1	0	0

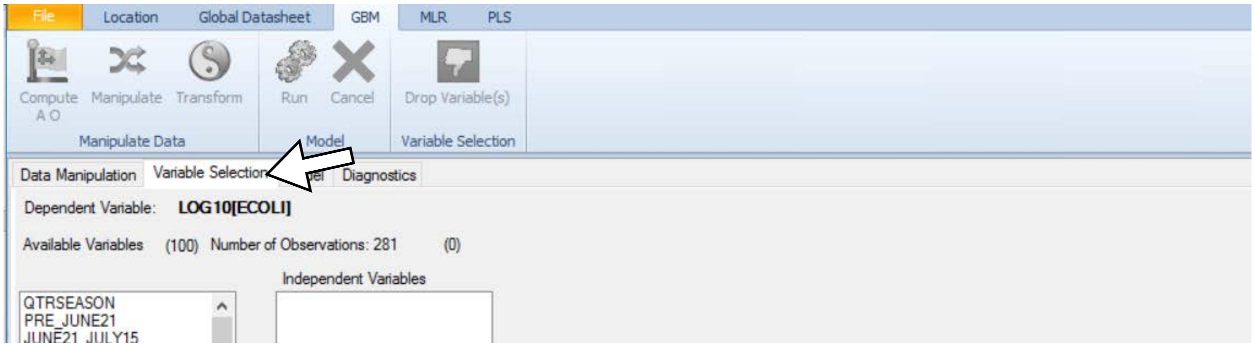
A.2. Now three tabs corresponding to each model **Virtual Beach 3** can create are available. Click on the **GBM** tab. A copy of the main data table labeled **Data Manipulation** will open. You can manipulate the data specifically for the GBM model in this tab without effecting the global data sheet.



File	1-VB-Training-Data.
Column Count	103
Row Count	281
Date-Time Index	DATETIME
Response Variable	LOG10[ECOLI]
Disabled Row Count	0
Disabled Column Count	1
Hidden Column Count	1
Independent Variable Count	100

DATETIME	LOG10[ECOLI]	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG
5/21/2009 12:05...	0.301	1	1	0	0	0
5/28/2009 12:20...	0.699	1	1	0	0	0
6/4/2009 11:55:...	0	1	1	0	0	0
6/11/2009 12:35...	2.538	1	1	0	0	0
6/12/2009 2:15:...	1.255	1	1	0	0	0
6/15/2009 11:25...	1.462	1	1	0	0	0
6/16/2009 10:30...	0.9031	1	1	0	0	0
6/17/2009 2:05:...	2.079	1	1	0	0	0
6/18/2009 2:05:...	1.23	1	1	0	0	0
6/22/2009 10:40...	0.6021	2	0	1	0	0
6/23/2009 11:45...	1.881	2	0	1	0	0

A.3. Click the the **Variable Selection** sub-tab.



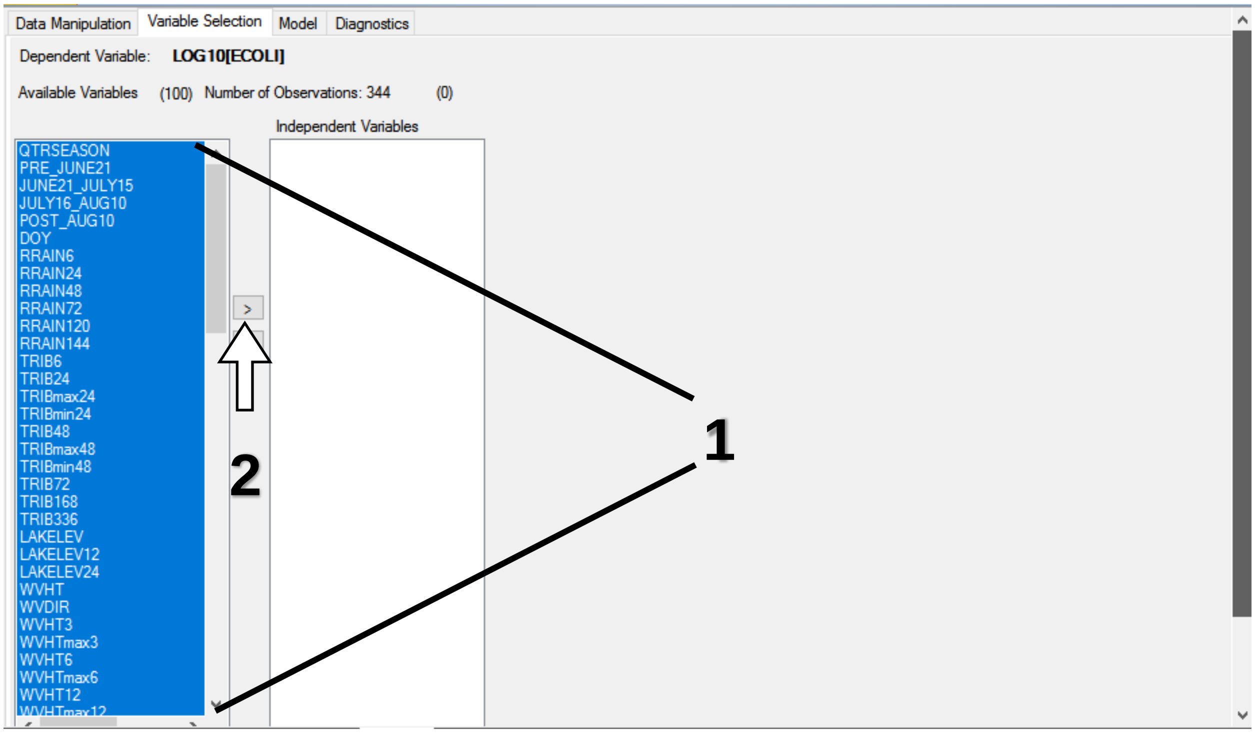
Dependent Variable: LOG10[ECOLI]

Available Variables (100) Number of Observations: 281 (0)

Independent Variables

QTRSEASON
PRE_JUNE21
JUNE21_JULY15

A.4. **1.** Under **Available Variables** in the left panel, select all independent variables by clicking the first variable QTRSEASON, holding down the shift key, scrolling to the bottom of the list and then clicking the last variable DIFF[TRIBmax24,TRIBmin24]. **2.** Click the right-arrow **›** button to move the selected variables to the right-hand panel. In this example, you should end-up with **100** Independent Variables.



Dependent Variable: LOG10[ECOLI]

Available Variables (100) Number of Observations: 344 (0)

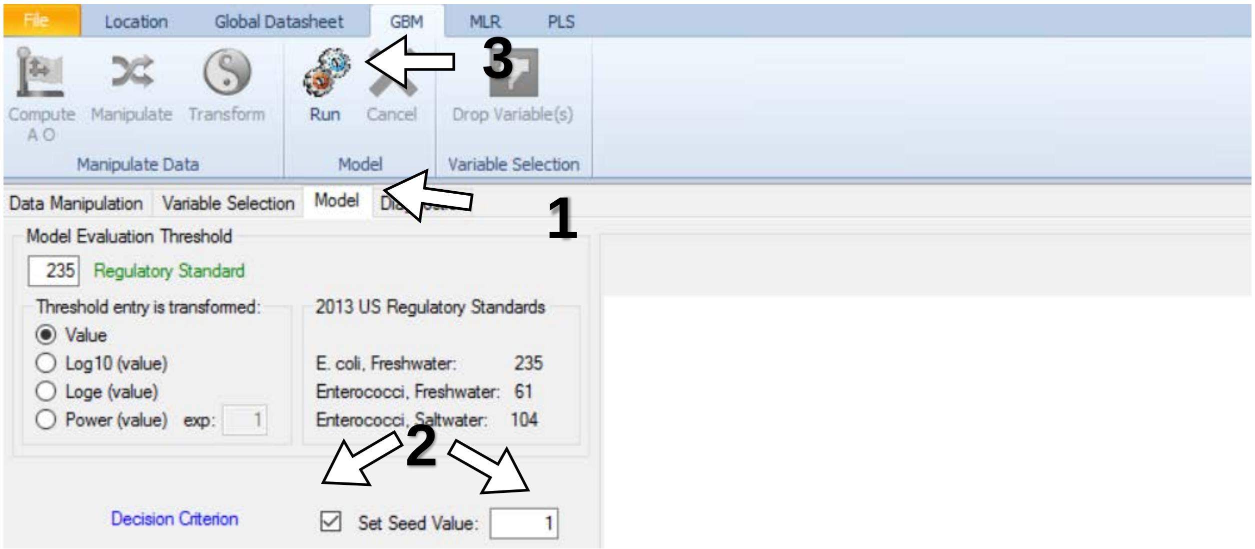
Independent Variables

QTRSEASON
PRE_JUNE21
JUNE21_JULY15
JULY16_AUG10
POST_AUG10
DOY
RRAIN6
RRAIN24
RRAIN48
RRAIN72
RRAIN120
RRAIN144
TRIB6
TRIB24
TRIBmax24
TRIBmin24
TRIB48
TRIBmax48
TRIBmin48
TRIB72
TRIB168
TRIB336
LAKELEV
LAKELEV12
LAKELEV24
WVHT
WVDIR
WVHT3
WVHTmax3
WVHT6
WVHTmax6
WVHT12
WVHTmax12

GBM is a nonlinear technique – it is a decision tree that makes binary decisions so even if variable are closely correlated, the GBM model will not be overly influenced by the interaction. Additionally, if transformations or interactions are important, GBM will find them and use them for you.

A.5. **1.** To run your GBM model, open the **Model** tab. **2.** For everyone in the training to have exactly the same model, check the **Select Seed Value** box. Enter the number 1 in the box. **Virtual Beach 3** uses a random number generator to produce numbers for creating models. By setting the seed value to the same number, the analysis can be reproduced by others. **3.** Click the "Run" button. The process should not take more than 2-3 minutes.

Under **Model Evaluation Threshold**, 235 is automatically entered for the Regulatory Standard. This represents the common US EPA standard of 235 CFU of *E. coli* for issuing a swim advisory.



Model Evaluation Threshold

235 Regulatory Standard

Threshold entry is transformed:

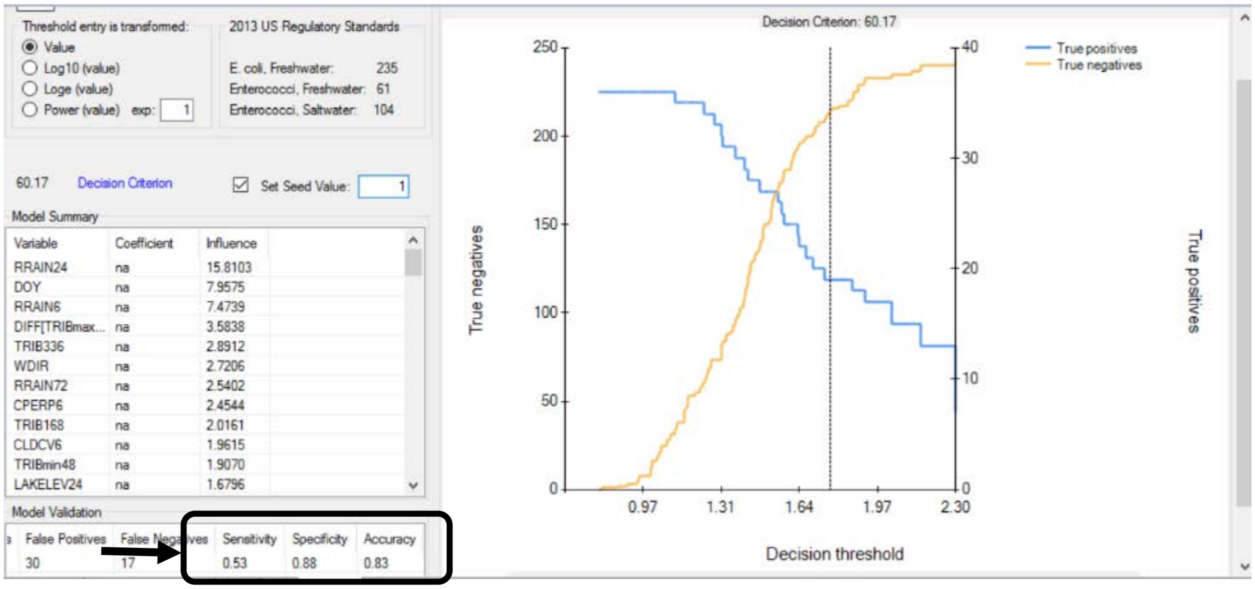
● Value
○ Log10 (value)
○ Loge (value)
○ Power (value) exp: 1

2013 US Regulatory Standards

E. coli, Freshwater:	235
Enterococci, Freshwater:	61
Enterococci, Saltwater:	104

☑ Set Seed Value: 1

A.6. A line graph shows **True Positives** (correctly predicted ***exceedances***) in blue and **True Negatives** (correctly predicted ***non-exceedances***) in yellow.



Threshold entry is transformed:

● Value
○ Log10 (value)
○ Loge (value)
○ Power (value) exp: 1

2013 US Regulatory Standards

E. coli, Freshwater:	235
Enterococci, Freshwater:	61
Enterococci, Saltwater:	104

60.17 Decision Criterion ☑ Set Seed Value: 1

Model Summary

Variable	Coefficient	Influence
RRAIN24	na	15.8103
DOY	na	7.9575
RRAIN6	na	7.4739
DIFF[TRIBmax...	na	3.5838
TRIB336	na	2.8912
WDIR	na	2.7206
RRAIN72	na	2.5402
CPERP6	na	2.4544
TRIB168	na	2.0161
CLDCV6	na	1.9615
TRIBmin48	na	1.9070
LAKELEV24	na	1.6796

Model Validation

False Positives	False Negatives	Sensitivity	Specificity	Accuracy
30	17	0.53	0.88	0.83

In the bottom-left panel, scroll all the way to the right, and note the values (percentages) reported for "**Sensitivity**" (true positives / (true positives + false negatives)), "**Specificity**" (true negatives / (true negatives + false positives)) and "**Accuracy**" ((true positives + true negatives) / number of total observations).

The important metrics of model performance are not the common statistical measures of 'fit', like R-square, nor are they measures of 'precision', like mean absolute error. Rather, they are **sensitivity** and **specificity**. These key measures, in turn, are related to the model-specific, adjustable, **decision criteria**.

KEY TERMS

Sensitivity: The percentage of correctly predicted water-quality exceedances (true positives) out of all measured, or observed, exceedances. Over 0.50 [50%] is considered good and the rule of thumb for this model.

Model example: 19 / (19 +17) = 0.53 [53%]

Specificity: The percentage of correctly predicted non-exceedances (true negatives) out of all measured, or observed, non-exceedances. Over 0.90 [90%] is considered good and the rule of thumb for this model.

Model example: 215/ (215 + 30) = 0.88 [88%]

Accuracy: The percentage of correctly predicted exceedances and non-exceedances out of all results. Do *not* use accuracy as the sole basis for setting Decision Criteria. Often the Decision Criterion corresponding to highest Accuracy has an unacceptably low Sensitivity. The goal is not to maximize accuracy, but to find an optimal balance of Sensitivity and Specificity, using the 50% - 90% rule-of-thumb.

Model example: (19+215) / (19+215+30+17) = 0.83 [83%]

	TRUE (RIGHT Prediction)	**FALSE** (WRONG Prediction)	
POSITIVES (As predicted by model)	**19** (points really OVER standard)	**30** (points really under standard)	**SENSITIVITY**
NEGATIVES (As predicted by model)	**215** (points really under standard)	**17** (points really OVER standard)	**SPECIFICITY**

Decision Criteria: The prediction thresholds that determine whether an actual exceedance of a regulatory standard. In GBM, when **Virtual Beach 3** has finished developing a model, it automatically recommends a **decision criterion**.

In this example, the **decision criterion** has automatically been set to approximately 63 CFU/100 mL, colony-forming units (CFU). This is a value of the original (un-transformed) response variable, *E. coli*. The regulatory standard is 235 CFU. We will adjust this threshold value later on in step B.3.

In GBM models, the optimal **decision criterion**, as suggested by the 50% - 90% rule of thumb, is typically much lower than 235 CFU, often, it will be lower than 100 CFU.

Particularly on those days with very high levels of *E. coli* at the beach, model-predicted concentrations will typically be lower than the actual values. In effect, most nowcast models are "muted." That is, the predicted extremes are not as high as the actual extremes. This is normal.

While the concept of using decision criteria that are different from 235 CFU may seem confusing at first, it is critical that you not simply insert 235 or some other common threshold in place of the optimal threshold as identified through the process highlighted above. Using a sub-optimal threshold for simplicity sake will result in increased decision errors meaning more missed advisories or unnecessary advisories.

As will be highlighted in the EnDDaT module, regardless of the numeric value automatically selected for the decision criterion, it will correspond to an exceedance probability of 50%. Even though the exact number for the optimal decision criterion may not match the 235 regulatory standard, it will correspond to a 50% probability of exceeding the standard.

B. Optimize a GBM model by removing variables

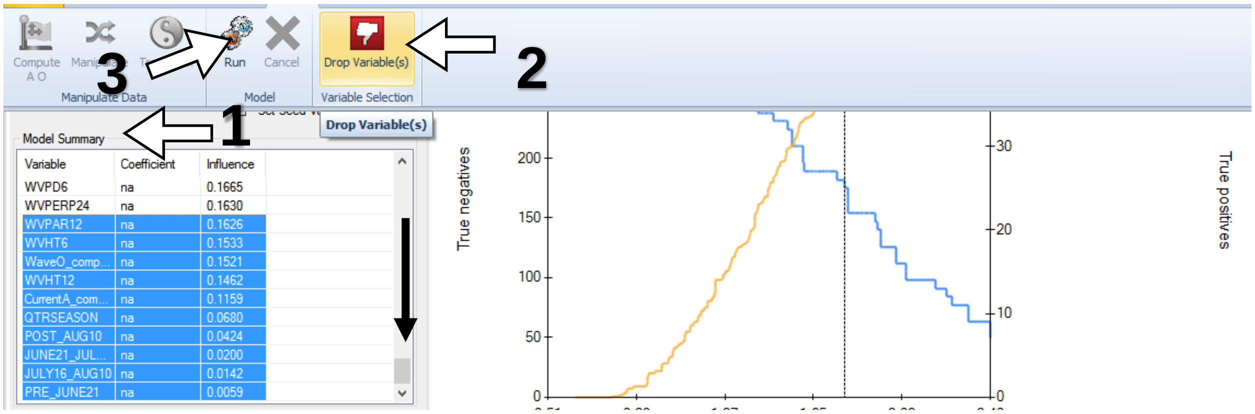
By combining the less limiting modeling approach of GBM with the large number of data available through the EnDDaT Web data portal, it is possible to develop Nowcast models with 100 or more independent variables for each of the observed, or sampled, bacteria samples collected. HOWEVER, just because gathering many independent variables is possible, it is not necessarily a good idea.

Recent tests have shown that the addition of more and more variables to a GBM model only improves its predictive model to a certain extent. At some point, the addition of more independent variables that are less related to water quality will cause a model's predictive power to actually decline.

In addition, the more online data services a model depends on, the more inefficient and unreliable its daily operation will be. Asking a model to look for more data will make it run slower since data calls will take longer to complete or be unavailable if there are service outages or other technical issues.

In this section, we will optimize the GBM quickly and effectively to either increase or maintain the model's predictive power by reducing the number of variables included.

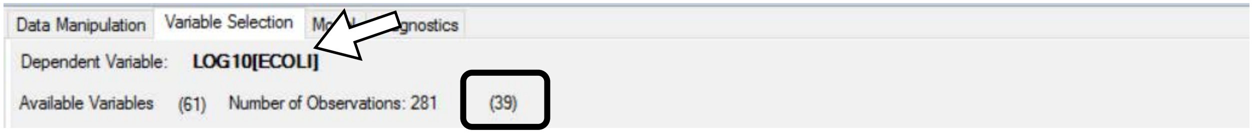
B.1. **1.** Under the **Model Summary** sub-tab, scroll to the bottom of **Variables** list. This list is ordered by these variables' relative Influence on response variable, *E. coli*. Select the bottom 10%, or 10, variables. **2.** Click the **Drop Variables** icon. **3.** Click the **Run** icon.



Model Summary

Variable	Coefficient	Influence
WVPD6	na	0.1665
WVPERP24	na	0.1630
WVPAR12	na	0.1626
WVHT6	na	0.1533
WaveO_comp...	na	0.1521
WVHT12	na	0.1462
CurrentA_com...	na	0.1159
QTRSEASON	na	0.0680
POST_AUG10	na	0.0424
JUNE21_JUL...	na	0.0200
JULY16_AUG10	na	0.0142
PRE_JUNE21	na	0.0059

B.2. Repeat Step B.1 removing the next 10% least influential variables to generate new models. In this example, you will run the model next with 81 variables, then 73, 66, 59, 53, 48, 43, and finally 39 variables. Click the **Variable Selection** sub-tab if you need to see how many variables were used in the previous model before clicking the **Drop Variable(s)** icon. .



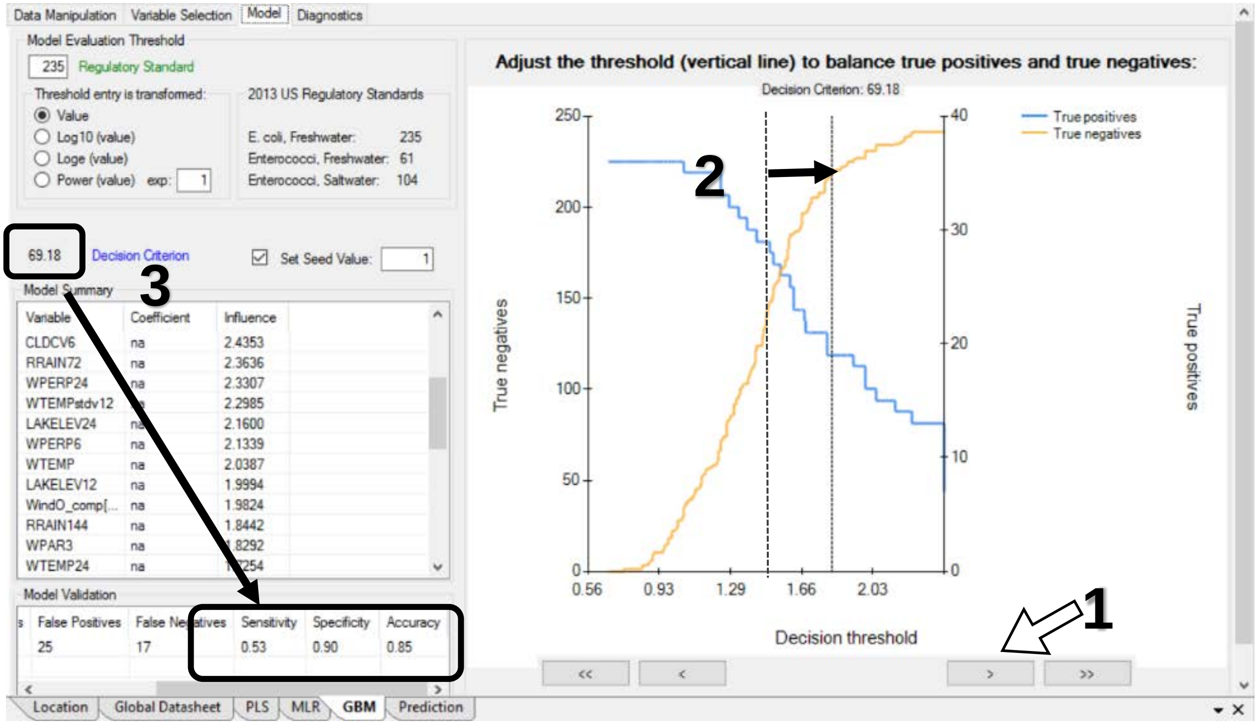
Dependent Variable: LOG10[ECOLI]

Available Variables (61) Number of Observations: 281 (39)

B.3. There is no perfect number of independent variables when creating a model, however a reduction in variables would increase the operational efficiency of running the model on a daily basis. Generally, reducing the variables to less than 45 will make the **Virtual Beach 3** program run more quickly. Keeping more than 30 independent variables ensures the model incorporates multiple factors for predicting bacteria levels.

1. Increase the Decision Criterion by clicking on the right-arrow **›** button under the graph. **2.** This will move the horizontal line to the right, which in turn will affect the model's Sensitivity and Specificity. **3.** Keep clicking **›** until you have achieved a Sensitivity/ Specificity balance as close to 50%/ 90% as possible.

By adjusting the Decision Criterion, using the **‹** or **›** buttons) until you have achieved an optimal balance of Sensitivity/ Specificity. In this example, a criterion value of approx. 69 (CFU) results in a Sensitivity of 0.53, Specificity of 0.90, and an overall Accuracy of roughly 0.85.



Model Evaluation Threshold

235 Regulatory Standard

Threshold entry is transformed:

● Value
○ Log10 (value)
○ Loge (value)
○ Power (value) exp: 1

2013 US Regulatory Standards

E. coli, Freshwater:	235
Enterococci, Freshwater:	61
Enterococci, Saltwater:	104

69.18 Decision Criterion

☑ Set Seed Value: 1

Model Summary

Variable	Coefficient	Influence
CLDCV6	na	2.4353
RRAIN72	na	2.3636
WPERP24	na	2.3307
WTEMPstdv12	na	2.2985
LAKELEV24	na	2.1600
WPERP6	na	2.1339
WTEMP	na	2.0387
LAKELEV12	na	1.9994
WindO_comp[...	na	1.9824
RRAIN144	na	1.8442
WPAR3	na	1.8292
WTEMP24	na	1.7254

Model Validation

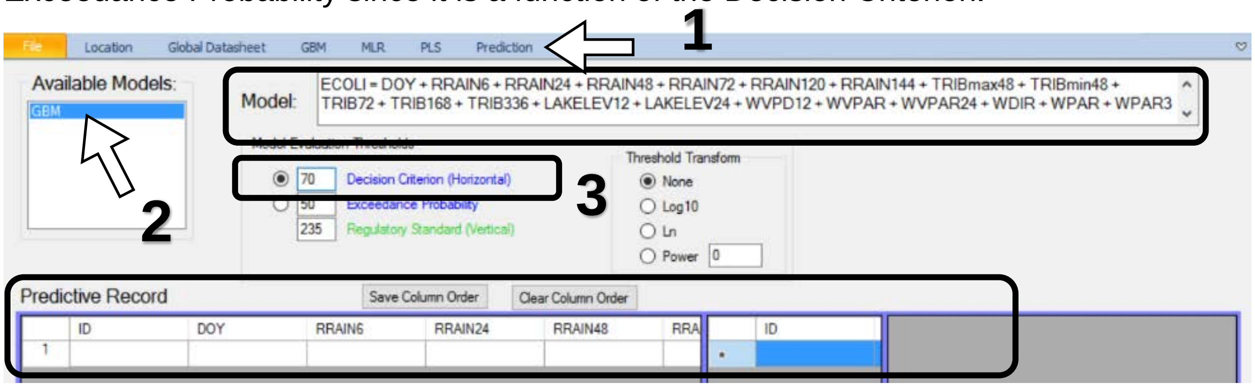
s	False Positives	False Negatives	Sensitivity	Specificity	Accuracy
	25	17	0.53	0.90	0.85

C. View a GBM model within the Prediction tab

The **Prediction** tab shows a model in the format that the eventual Nowcast operator will use to make routine water-quality predictions. It is here that the daily observations of explanatory variables like antecedent rainfall, wave height, and gull counts will be manually entered if collected during a sanitary survey at the beach or downloaded via online services such as EnDDaT. The model created in this GBM module only used independent variable data gathered from EnDDaT. This data gathering process will be explained in more detail in the "EnDDaT Module".

Up to 3 different models can be accommodated in a single Prediction tab: 1 GBM model, 1 PLS model, and 1 traditional MLR model.

C.1. **1.** Click on the **Prediction** tab. **2.** Under **Available Models** click **GBM**. The model equation is displayed and a row of blank cells appears under **Predictive Record**. **3.** The Decision Criterion set during the GBM model-building and optimization is not useful for beach decisions since it isn't a whole number. For easier interpretation, change the value to the closest multiple of 5 or 10. In this example, 70 is the decision criteria that produced the best sensitivity and specificity. You do not need to change the Exceedance Probability since it is a function of the Decision Criterion.



Model: ECOLI = DOY + RRAIN6 + RRAIN24 + RRAIN48 + RRAIN72 + RRAIN120 + RRAIN144 + TRIBmax48 + TRIBmin48 + TRIB72 + TRIB168 + TRIB336 + LAKELEV12 + LAKELEV24 + WVPD12 + WVPAR + WVPAR24 + WDIR + WPAR + WPAR3

● 70 Decision Criterion (Horizontal)
○ 50 Exceedance Probability
235 Regulatory Standard (Vertical)

Threshold Transform: ● None ○ Log10 ○ Ln ○ Power 0

Predictive Record

	ID	DOY	RRAIN6	RRAIN24	RRAIN48	RRA
1						

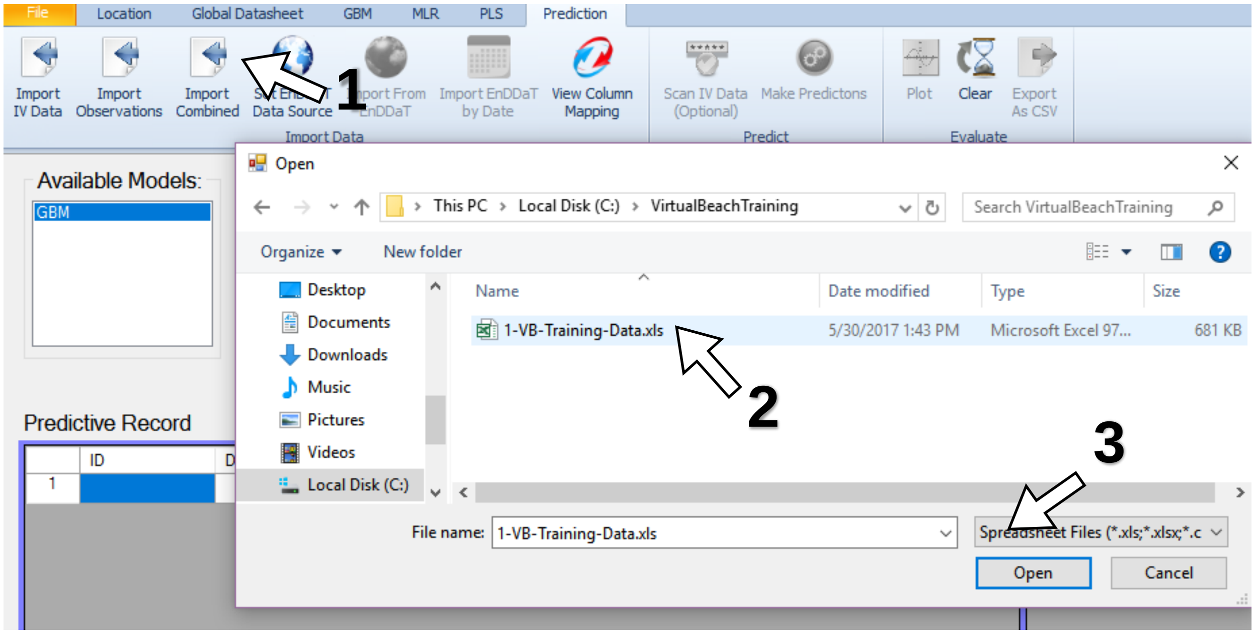
	ID
*	

Model Equation: The text box at the top-center of the **Prediction** tab contains the mathematical expression of the selected model. In the case of MLR and PLS models, this equation will include numeric coefficients that define the independent relationship with each explanatory variable and the response variable; e.g., 'ECOLI'. In the case of GBM models, there are no coefficients.

Predictive Record: The bottom half of the **Prediction** tab is the **Predictive Record**. Each row represents a unique date and time for which field observations and/or remotely-measured data will be entered for each explanatory variable in the model. Then the response variable of ECOLI and the probability of exceeding the established Decision Criterion will be predicted.

Import a historical data table to validate your models

D.1. On the **Prediction** tab, click the **Import Combined** icon. **2.** Navigate to the Excel file “1-VB-Training-Data.xls”. **3.** Click **Open**.

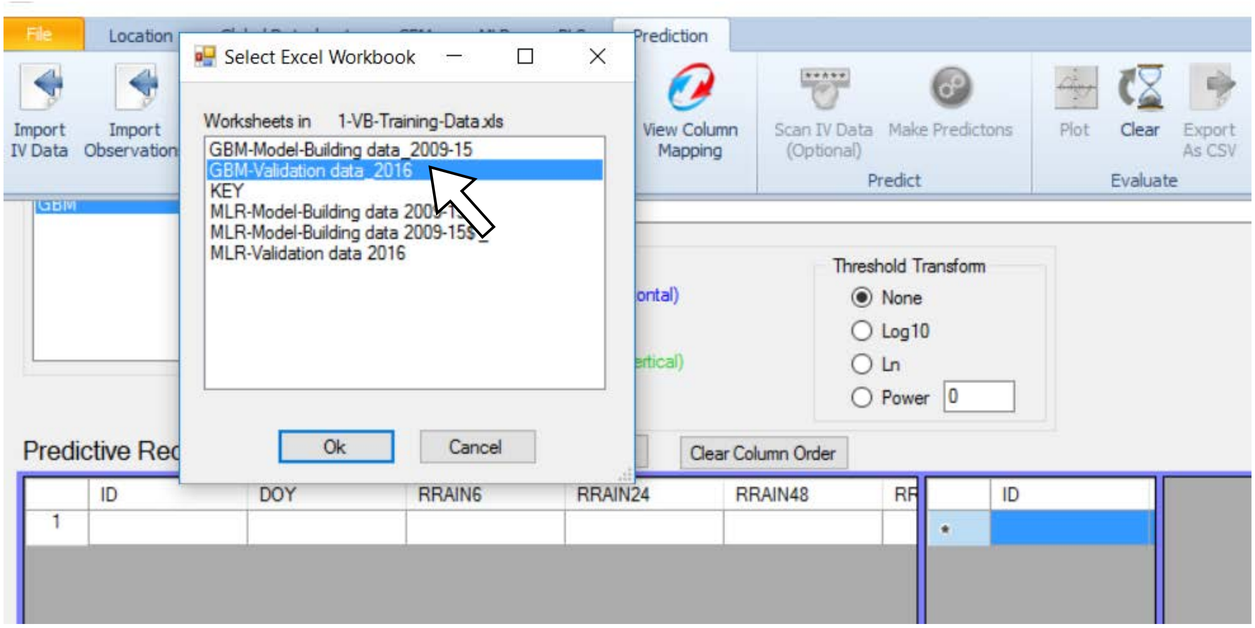


Name	Date modified	Type	Size
1-VB-Training-Data.xls	5/30/2017 1:43 PM	Microsoft Excel 97...	681 KB

File name: 1-VB-Training-Data.xls

Spreadsheet Files (*.xls;*.xlsx;*.c

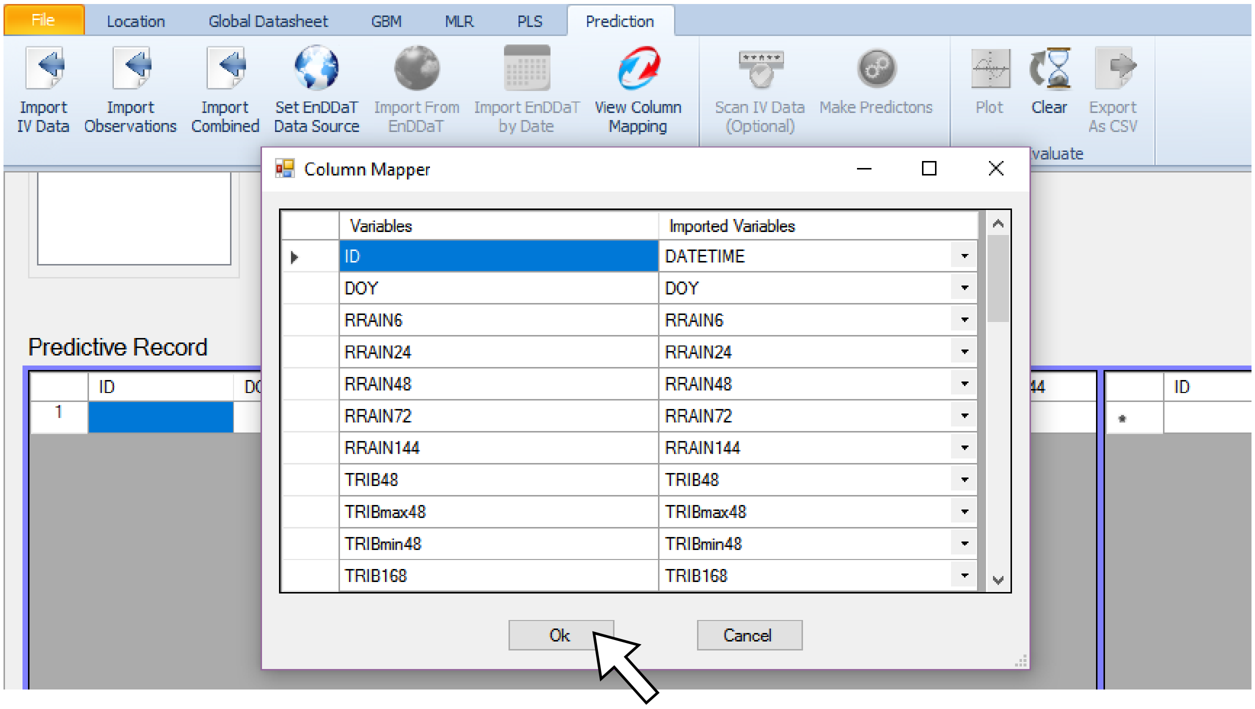
D.2. Select “GBM-Validation_data_2016” and click **OK**. Although these data are historical, they were collected after the data used to develop the model.



Worksheets in 1-VB-Training-Data.xls

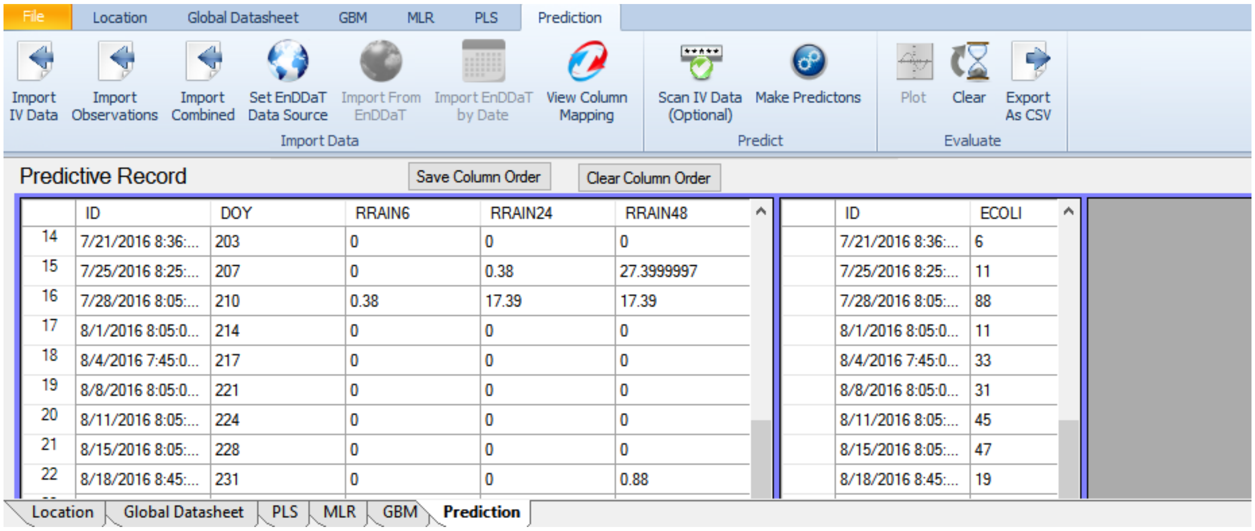
GBM-Model-Building data_2009-15
GBM-Validation data_2016
KEY
MLR-Model-Building data 2009-15
MLR-Model-Building data 2009-15$_
MLR-Validation data 2016

D.3. If this is the first time these data have been imported into the **Prediction** Tab, the **Column Mapper** window will open. **Imported Variables** must have the same names of their corresponding **Variables** in the model. You must map, or match, these variables. Since you are using data collected by the same entity, just from a later date, all variable names match. Click **OK**.



Variables	Imported Variables
ID	DATETIME
DOY	DOY
RRAIN6	RRAIN6
RRAIN24	RRAIN24
RRAIN48	RRAIN48
RRAIN72	RRAIN72
RRAIN144	RRAIN144
TRIB48	TRIB48
TRIBmax48	TRIBmax48
TRIBmin48	TRIBmin48
TRIB168	TRIB168

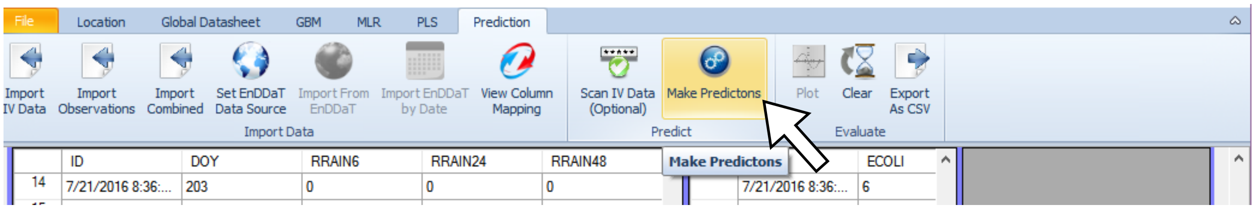
D.4. The explanatory variables, or independent variables, should now be in the left-hand panel. The *E. coli* values should be in the middle panel.



	ID	DOY	RRAIN6	RRAIN24	RRAIN48
14	7/21/2016 8:36:...	203	0	0	0
15	7/25/2016 8:25:...	207	0	0.38	27.3999997
16	7/28/2016 8:05:...	210	0.38	17.39	17.39
17	8/1/2016 8:05:0...	214	0	0	0
18	8/4/2016 7:45:0...	217	0	0	0
19	8/8/2016 8:05:0...	221	0	0	0
20	8/11/2016 8:05:...	224	0	0	0
21	8/15/2016 8:05:...	228	0	0	0
22	8/18/2016 8:45:...	231	0	0	0.88

ID	ECOLI
7/21/2016 8:36:...	6
7/25/2016 8:25:...	11
7/28/2016 8:05:...	88
8/1/2016 8:05:0...	11
8/4/2016 7:45:0...	33
8/8/2016 8:05:0...	31
8/11/2016 8:05:...	45
8/15/2016 8:05:...	47
8/18/2016 8:45:...	19

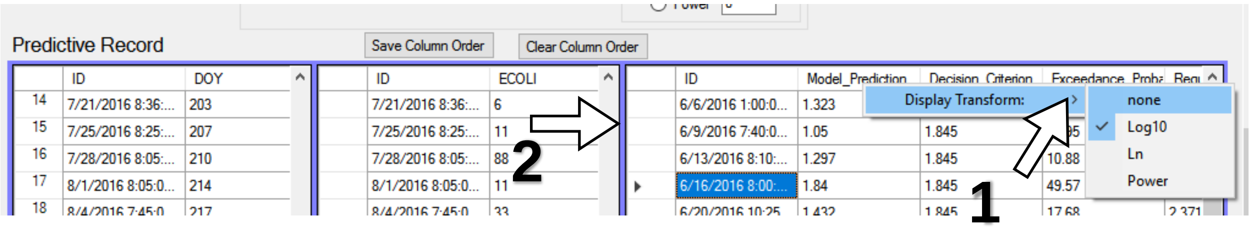
D.5. Click the **Make Predictions** icon.



	ID	DOY	RRAIN6	RRAIN24	RRAIN48
14	7/21/2016 8:36:...	203	0	0	0

	ID	ECOLI
	7/21/2016 8:36:...	6

D.6. The back-cast predictions are in Log-scale in the right panel. **1.** Right-click on the top of the **Model_Prediction** column and choose **none** to the right of **Display Transform: >** to make the *E. coli* prediction appear on a non-Log-scale. **2.** Right click and drag on the solid blue section separators to enlarge the width of the prediction panel and reduce the width of the independent variable panel.

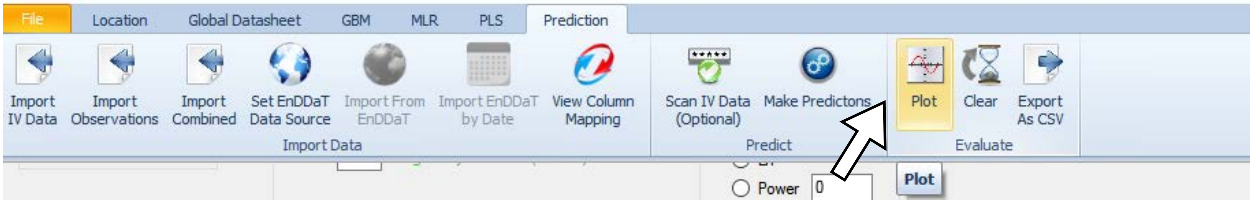


	ID	DOY
14	7/21/2016 8:36:...	203
15	7/25/2016 8:25:...	207
16	7/28/2016 8:05:...	210
17	8/1/2016 8:05:0...	214
18	8/4/2016 7:45:0	217

	ID	ECOLI
	7/21/2016 8:36:...	6
	7/25/2016 8:25:...	11
	7/28/2016 8:05:...	88
	8/1/2016 8:05:0...	11
	8/4/2016 7:45:0	33

	ID	Model_Prediction	Decision_Criterion	Exceedance_Proba	Regu
	6/6/2016 1:00:0...	1.323			
	6/9/2016 7:40:0...	1.05	1.845	95	
	6/13/2016 8:10:...	1.297	1.845	10.88	
	6/16/2016 8:00:...	1.84	1.845	49.57	
	6/20/2016 10:25	1.432	1.845	17.68	2.371

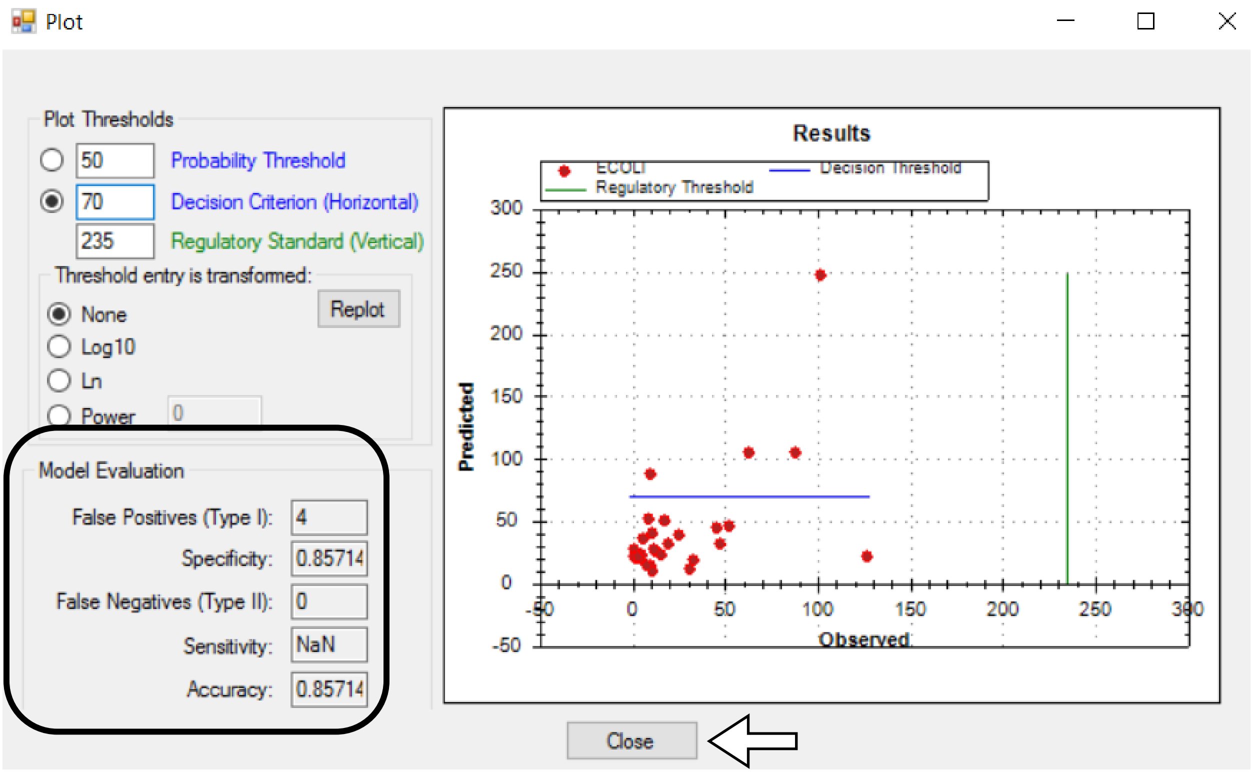
D.7. To evaluate performance visually and numerically, click the **Plot** icon.



A plot of **Observed** *E. coli* values on the X-axis vs. **Predicted** *E. coli* values on the Y-axis appears. The Decision Criteria, horizontal blue line, in this example is 70 CFU and the Regulatory Standard, vertical green line, is 235 CFU.

Hover the mouse over an individual point on the scatter plot to view its associated date and time. In the screenshot below with 28 observations, there were no False Negatives (missed exceedances) resulting in a Sensitivity of "NaN" (not a number) since there were no actual exceedances in the data set from the summer of 2016. There were four False Positives (unnecessary advisories might have been issued) resulting in a Specificity of 86%. The Overall Accuracy of correct advisory and non-advisory decisions was then 86% also.

When you have completed your evaluation, click "Close" to return to the Predictive Record (table view).



Plot

Plot Thresholds

○ 50 Probability Threshold
● 70 Decision Criterion (Horizontal)
235 Regulatory Standard (Vertical)

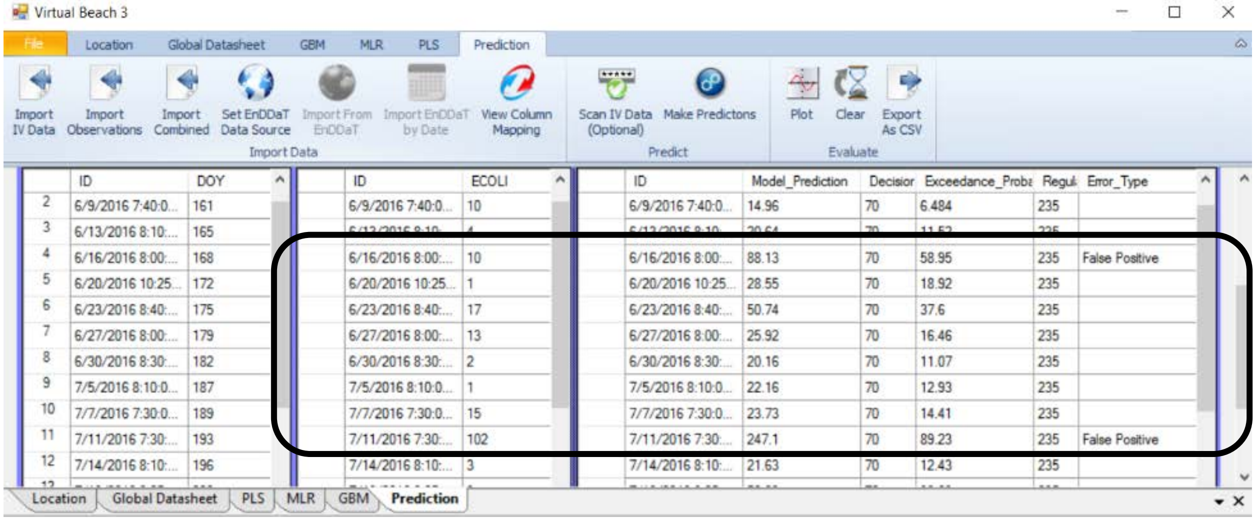
Threshold entry is transformed:
● None
○ Log10
○ Ln
○ Power 0

Replot

Model Evaluation

False Positives (Type I):	4
Specificity:	0.85714
False Negatives (Type II):	0
Sensitivity:	NaN
Accuracy:	0.85714

D.8. Adjust the width of columns in the various panels so you can more easily examine the daily results. Ideally, you will view the observed ECOLI in the middle panel, and both Exceedance_ Probability and Error_Type in the right panel. Circled below are examples of incorrectly predicted exceedances, one with a 58.95% Exceedance_ Probability and another with an 89.23 Exceedance_ Probability.



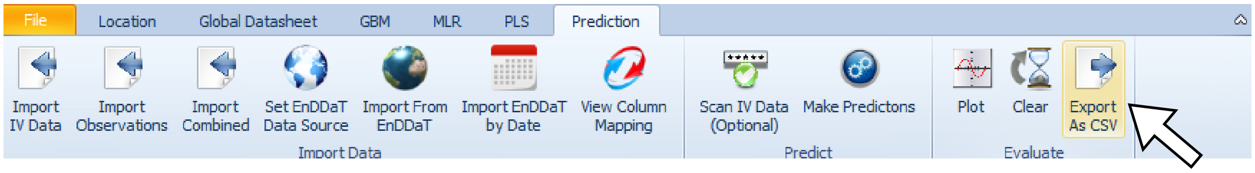
	ID	DOY
2	6/9/2016 7:40:0...	161
3	6/13/2016 8:10:...	165
4	6/16/2016 8:00:...	168
5	6/20/2016 10:25...	172
6	6/23/2016 8:40:...	175
7	6/27/2016 8:00:...	179
8	6/30/2016 8:30:...	182
9	7/5/2016 8:10:0...	187
10	7/7/2016 7:30:0...	189
11	7/11/2016 7:30:...	193
12	7/14/2016 8:10:...	196

ID	ECOLI
6/9/2016 7:40:0...	10
6/13/2016 8:10...	4
6/16/2016 8:00:...	10
6/20/2016 10:25...	1
6/23/2016 8:40:...	17
6/27/2016 8:00:...	13
6/30/2016 8:30:...	2
7/5/2016 8:10:0...	1
7/7/2016 7:30:0...	15
7/11/2016 7:30:...	102
7/14/2016 8:10:...	3

ID	Model_Prediction	Decisior	Exceedance_Proba	Regul	Error_Type
6/9/2016 7:40:0...	14.96	70	6.484	235	
6/13/2016 8:10:	20.64	70	11.52	235	
6/16/2016 8:00:...	88.13	70	58.95	235	False Positive
6/20/2016 10:25...	28.55	70	18.92	235	
6/23/2016 8:40:...	50.74	70	37.6	235	
6/27/2016 8:00:...	25.92	70	16.46	235	
6/30/2016 8:30:...	20.16	70	11.07	235	
7/5/2016 8:10:0...	22.16	70	12.93	235	
7/7/2016 7:30:0...	23.73	70	14.41	235	
7/11/2016 7:30:...	247.1	70	89.23	235	False Positive
7/14/2016 8:10:...	21.63	70	12.43	235	

D.9. To export the Predictive Record for further analysis or future reference, click the **Export As CSV** icon. Name the file something like “Validation 2016_BlueHarborBeach”, and click **OK**. CSV files can be opened directly in Excel and imported into various other analytical programs.

.



E. Save a GBM model as a new project file

E.1. On the **File** tab, select **Save As**. Navigate to the directory where you plan to keep your models and save the project as a file with “GBM” in the title. This will capture all of the work that you have completed to this point.

When you save the completed and validated nowcast predictions, all of the input, output, and validation data are saved. When saving your model file, over-writing the existing file will not change the model. Instead, the new predictive records are simply added.

Updating and re-saving model files is an effective way to track the performance of your nowcast model over the course of a beach season. If the nowcast proves to perform poorly, you can revisit steps A through D of this modules to revise the nowcast model.