Virtual Beach 3.0.6 – Data Preparation for GBM model

In this module you will learn how to:

- A. Import and clean-up model-building data for your beach
- B. Process directional data (wind, currents, waves)
- C. Combine two or more predictive variables
- D. Transform the dependent bacteria variable

A. Import and clean-up model-building data for your beach

A.1. Open **Microsoft Excel** to preview the data you will be importing into **Virtual Beach 3**. Open the file "VB_Training_Data_GBM.xls".

Be sure to save your data as "*.xls" files. A plugin for **Virtual Beach** 3 is available for importing "*.xlsx" files, but there are still bugs to be worked out.

Column **B** is always the *response* variable, "ECOLI" in this example. All data to the right are potential *explanatory* variable. See the **KEY** tab of the **Excel** file for descriptions of variables used in this module. Close the **Excel** file before returning to **Virtual Beach 3**. Data cannot be imported from an open Excel file.

	<u>יאי</u> יאי			VB_Tra	aining_Data_GBM.xls	Compatibility Mo	de] - Excel		困 —		×
Fi	le Home Ins	ert F	Page i ayout	Formulas Da	ta Review Vi	ew ACROBAT	Q Tell me what	you want to do	Sign ir	P₄ Sha	re
B1		\times	fr. E	COLI							
	A	В		D	F	F	G	Н		1	
	DATETIME	_	-	_	-		_	CLOUDCOV_Qual	DOY	RRAIN6	_
2	5/21/2009 12:05		1	1	0	0	0	Sunny	141		
3	5/28/2009 12:20	5	1	1	0	0	0	Cloudy	148		
1	6/4/2009 11:55	1	1	1	0	0	0	Sunny	155		
5	6/11/2009 12:35	345	1	1	0	0	0	Mostly Sunny	162		
5	6/12/2009 14:15	18	1	1	0	0	0	Cloudy	163		
7	6/15/2009 11:25	29	1	1	0	0	0	Sunny	166		
3	6/16/2009 10:30	8	1	1	0	0	0	Cloudy	167		
)	6/17/2009 14:05	120	1	1	0	0	0	Sunny	168		
0	6/18/2009 14:05	17	1	1	0	0	0	Sunny	169		
1	6/22/2009 10:40	4	2	0	1	0	0	Sunny	173		
2	6/23/2009 11:45	76	2	0	1	0	0	Partly Sunny	174		
3	6/24/200055	15	2	0	1	0	0	Partly Sunny	175		
	► KEY	BM-M	odel-Buildin	g data_2009-1	5 GBM-Valida	tion data_2016	(+)	• •	1	: • •	

This file can be used as a template for formatting beach-specific data.

A.2. Return to Virtual Beach 3 project file created in the "Beach Orientation" module.
1. Click the Global Datasheet tab. 2. Click the Import Data icon and select the Excel file "VB_Training_Data_GBM.xls". 3. Click Open.

Free Location Global Datasheet	n Open			×
0 0 1 1 1 0 0	This PC > OS (C:) > Virtual Beach 3	o ~	Search Virtual Beach 3	Q
nport Validate Compute Manipulate Transform Go To Data Data A O Model	Organize 👻 New folder			
Add date Work with Data	- Downloads ^ Name	Date modified	Туре	Size
2	Music Britunes Fictures Videos Solutions	-8/15/2017 6:13 PM	Microsoft Excel 97	579 KB
	File name: VB_Training_Data_GBM.xls	~	V	*.xlsx;*.c ~ Cancel

A.3. In this example there is more than one worksheet in the Excel file, so you must choose which one to import. **1.** Select the worksheet **GBM-Model-Building data_2009-15. 2.** Click **OK**.

🖳 Virtua	Beach 3					
File	Location	Global	Datasheet			
0	0	34	24	3	0	Select Excel Workbook — 🗆 🗙
Import Data	Validate Data	Compute A O	Manipulate	Transform	Go To Model	Worksheets in VB_Training_Data_GBM.xls
Add	Validate	4.9	Work wit	h Data	Mouler	GBM-Model-Building data_2009-15 GBM-Validation data_2016
						KEY
						1
						2
						A
						Ok Cancel

A.4. The data table will open in **Virtual Beach 3**.



F Virtual Beach 3 automatically highlights the second column of the datasheet as the response variable, "ECOLI", in this example. The "Response Variable" is indicated in the left-hand panel, along with "Column Count", "Row Count" and other descriptions of the data.

	Location Glob	al Datasheet									
bort ta	Validate Data Validate	Mariipulate Tra Work with Da	S) ansform	Go Tu Model							
File		1-VB-Training-Dat	a.	DATETIME	ECOLI	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	C
Colun	n Count	96		5/21/2009 12:05.	2	1	1	0	0	0	Su
Row	ime Index DATETIME			5/28/2009 12:20.	. 5	1	1	0	0	0	Clc
	Time Index DATETIME onse Variable ECOLI			6/4/2009 11:55:	1	1	1	0	0	0	Su
nesh		e Variable ECOLI		6/11/2009 12:35.	345	1	1	0	0	0	Mc
	led Row Count	0		6/12/2009 2:15:	18	1	1	0	0	0	Ck
	led Column Count n Column Count	0		6/15/2009 11:25.	. 29	1	1	0	0	0	Su
	endent Variable Count	-		6/16/2009 10:30.	. 8	1	1	0	0	0	Ck
				6/17/2009 2:05:	120	1	1	0	0	0	Su
				6/18/2009 2:05:	17	1	1	0	0	0	Su
				6/22/2009 10:40.	. 4	2	0	1	0	0	Su
c			1	C /30 /3000 11-85	70	2	0	1	0	0	0- ~ >

A.5. Virtual Beach 3 will NOT build a model if any cells have null (missing), or nonnumeric (text) values. 1. Click the Validate Data icon to check your dataset. 2. In the pop-up window, click Scan.

Location Development Validate Data dd Validate	Data Validation Scan (Optional) Find:							
File	Action (@) Replace With: 2	ECOLI	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	1
Column Count	Delete Row	2	1	1	0	0	0	4
ow Count		5	1	1	0	0	0	
ate-Time Index Response Variable	O Delete Column	1	1	1	0	0	0	
coportos variable	Take Action Within	345	1	1	0	0	0	
Disabled Row Count	Only This Cell 📎	18	1	1	0	0	0	
isabled Column Count		29	1	1	0	0	0	
idden Column Count dependent Variable Co	Texts Addition	8	1	1	0	0	0	
acheriacur, Aguapie er		120	1	1	0	0	0	
		17	Ť	1	0	0	0	
	Identify Categorical Variables	4	2	0	1	0	0	
		70	2	n	1	0	n	
ocation Global Da	Return							_

A.6. In this example, the "CLOUDCOV_qual" column is flagged because the values are text, or non-numeric. **1.** Click the radio button next to **Delete Column**. Under **Take Action Within** make sure **Only This Column** is selected. **2.** Click **Take Action**.

Virtual Beach 3	Validation						- 0]
Location	Data Validation							
ort Validate Cor ta Data A Id Validate	Scan							
File		QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG1	CLOUDCOV_Qual	[
Column Count	Replace With:	1	1	0	0	0	Sunny	14
Row Count	O Delete Row	1	1	0	0	0	Cloudy	1
Date-Time Index Response Variable	Delete Column	1	1	0	0	0		Ē
response vanable	Take Action Witmin:	1	1	0	0	0	Mostly Sunny	1
Disabled Row Count	Only This Column 🗸	1	1	0	0	0	Cloudy	1
isabled Column Count		1	1	0	0	0	Sunny	1
idden Column Count dependent Variable Co	Take Action	1	1	0	0	0	Cloudy	1
	13	1	1	0	0	0	Sunny	1
		1	1	0	0	0	Sunny	1
	Identify Categorical Variables	2	0	1	0	0	Sunny	1
		4	n	1	n	n	Druthe Courses	1
	Betum Cancel							-
	Lancel							

A.7. Repeat step A.6 until you come to the "TRIB6" column. The variable is numeric, but some cells are empty. Do not remove the entire column. **1.** Click the radio button for **Delete Row**. **2.** Select **Entire Column** and click **Take Action**.

Virtual Beach 3	Validation						- 0
Location	Data Validation						
ort Validate Cor ta Data (d Validate	(Optional) Find:						
File	Action:	RRAIN24	RRAIN48	RRAIN72	RRAIN120	RRAIN144	TRIB6
Column Count	O Replace With:	0	0	0	60.27	63.52	813 9
Row Count	Delete Row	0.88	0.88	0.88	12.76	62.9	571.7 6
Date-Time Index	O Delete Column	0	0.88	0.88	0.88	12.76	586.5 5
Response Variable	Take Action Within:	0	0	22.51	22.51	23.39	333.1 4
Disabled Row Count	Only This Row	0	0	0	22.51	22.51	193.3 2
Disabled Column Count	Only This Row	0.38	0.38	0.38	22.89	22.89	200.8 1
Hidden Column Count ndependent Variable Co	Entire Column Entire Sheet	0	0.38	0.38	0.38	22.89	206.2 2
nuependeni, valiable ce	hr	1.63	12.52	12.52	12.52	12.9	169.6 1
		0	1.63	12.52	12.52	12.52	161.4 1
	Identify Categorical Variables	2 16	16	17.63	28.52	28.52	
		_ n	10	10	C3 0C	20 23	
		-					
	Billum Cancel						

A.8. Repeat Step A.7 until a notice appears at the bottom of the pop-up window stating **No anomalous data values found**. Then click the **Return** button.

LUCAUUM	Chahal Dahashaat							
ort Validate Ga Data	Validation							
ld Validate	Scan	-						_
File	(Optional) Find:	ECOLI	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	-
Column Count	Action	345	1	1	0	0	0	
Row Count	Replace With:	. 18	1	1	0	0	0	
Date-Time Index Response Variable	(i) Delete Row	29	1	1	0	0	0	
Nesponse valiable	Delete Column	. 17	1	1	0	0	0	
Disabled Row Count	CT Delete Column	4	2	0	1	0	0	
Disabled Column Count	Take Action Within	76	2	0	1	0	0	
Hidden Column Count Independent Variable (Entire Column	15	2	0	1	0	0	
	Take Actor	3	2	0	1	0	0	
	Tank natur	11	2	0	1	0	0	
	^	5	2	0	1	0	0	
	Identify Categorical Variables	. 7	2	0	1	0	0	
¢		26	2	0	1	0	0	4
	No anomalous data values fou	ind.						-
ocation Global D	Return Cancel	-						

B. Process wind and current data

B.1. **1.** Click the **Compute A O** icon. **2.** In the pop-up window, under **Wind Data**, click the pull down arrow next to **Speed** and select WSPD. For **Direction**, select WDIR.

Location Global	Wind/Current C	official and a second	×					
Location Global	Wind Data Specify wind data	columns:						
a Data A O d Validate	Speed Direction (dea)	WVPD24 A WVPAR A WVPERP						
1/	Direction (deg)	WVPERP3 WVPAR6	QTRSEASON	PRE_JUNE21	JUNE21_JULY15	JULY16_AUG10	POST_AUG10	D ^
Column Count	Current Data	WVPAR6 WVPERP6	1	1	0	0	0	14
low Count		WVPAR12	1	1	0	0	0	14
ate-Time Index	Specify current da	WVPAR24	1	1	0	0	0	15
lesponse Variable	Speed	WVPERP24 WSPD	1	1	0	0	0	16
isabled Row Count		WDIR	1	1	0	0	0	16
lisabled Column Count	Direction (deg)	WPAR WPERP	1	1	0	0	0	16
idden Column Count		WPAR3	1	1				
dependent Variable Count		WPERP3 WPAR6	1	1	0	0	0	16
	Wave Data	WPERP6	1	1	0	0	0	16
	Specify wave data	WPAR12	1	1	0	0	0	16
	Specky wave data	WPAR24	2	0	1	0	0	17
-	Height	WPERP24 ATEMP ATEMP6	2	n	1	0	n	17 4
cation Global Datashe	Direction (deg)	ATEMP5 ATEMP12 ATEMP24 ATEMP24	-					

B.2. Repeat for **Current Data**, selecting CSPD and CDIR. Repeat for **Wave Data** selecting WVHT and WVDIR. The Beach Angle is automatically included. Click **OK**.

Current Data 3 0.1261 -0.01659 0.1717 -0.03284 0.1625 -0.00388 Specify current data columns: -0.0162 0.0718 -0.2873 0.09819 -0.437 0.1281 Specify current data columns: -0.0162 0.03684 -0.1361 0.03951 -0.1784 0.04436 Direction (deg) CDIR -0.0051 0.04907 -0.1737 0.04087 -0.1929 0.02522 13 -0.131 0.04138 -0.146 0.04777 -0.1977 0.06138 09 0.06425 -0.009418 0.09981 0.004245 0.1182 -0.01211 187 -0.1005 -0.01455 -0.08743 -0.01437 -0.1853 0.03237 99 -0.209 0.05336 -0.1797 0.04137 -0.1853 0.03237 187 -0.1005 -0.0133 -0.04925 -0.01564 -0.05114 0.004655 187 -0.02085 0.01331 -0.04001 0.02431 -0.01950 0.00567 11 <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>Wind Data</th>									Wind Data
Direction (deg) WDIR P3 CPAR6 CPERP6 CPAR12 CPERP12 CPAR24 CPERP24 3 0.1261 -0.01659 0.1717 -0.03284 0.1625 -0.03281 Current Data								columns:	Specify wind data
P3 CPAR6 CPERF6 CPAR12 CPERP12 CPAR24 CPERP24 Current Data 3 0.1261 -0.01659 0.1717 -0.03284 0.1625 -0.03688 Specify current data columns: 8 -0.1912 0.0718 -0.2873 0.09819 -0.437 0.1281 Specify current data columns: -0.102 0.03684 -0.1361 0.03951 -0.1784 0.04436 Direction (deg) CDIR -0 0.06425 -0.09418 0.04907 -0.1289 0.02522 13 -0.1331 0.04138 -0.146 0.04777 -0.1977 0.06125 -0.01211 187 -0.1055 -0.0145 -0.08743 -0.004245 -0.01349 -0.01155 09 -0.2094 0.05336 -0.1797 0.04137 -0.1833 0.03237 Wave Data Specify wave data columns: - - - - 0.00536 -0.1797 0.04137 -0.1833 0.03237 F22 -0.004242 -0.01033	 						-	WSPD ~	Speed
Current Data 8 0.1912 0.0718 0.2873 0.09819 0.437 0.1281 Specify current data columna: 5 0.102 0.03684 0.1361 0.03951 0.1784 0.0436 Direction (deg) CDIR 2 0.04087 0.1281 0.04087 0.1281 0.04087 0.1281 0.04436 Wave Data CDIR 2 0.06425 0.009418 0.09981 0.04087 0.1182 0.01151 Wave Data Specify wave data columns: 2 0.0005 0.01331 0.04001 0.02431 0.01839 0.02536 Wave Data Specify wave data columns: 2 0.0004242 0.01331 0.04001 0.02431 0.07802 0.02366 127 0.0004242 0.01033 0.04255 0.01564 0.05114 0.004655 126 0.02745 0.00895 0.1101 0.00833 0.1167 0.02096 48 0.0004213 0.01828 0.03929 0.04252 0.04437 0.008455	CPERP24	CPAR24	CPERP12	CPAR12	CPERP6	CPAR6	RP3	WDIR ~	Direction (deg)
Specify current data columna: 5 0.102 0.03684 0.1361 0.03951 0.1784 0.0436 Specify current data columna: 0 0.2051 0.04907 0.1737 0.04087 0.1289 0.02522 Direction (deg) CDIR 0 0.06425 0.00918 0.09981 0.04077 0.1977 0.01182 0.01182 Wave Data 0 0.00536 0.01045 0.08743 0.004234 0.01349 0.01211 187 0.1005 0.01045 0.08743 0.004734 0.01349 0.01329 9 0.209 0.05336 0.1797 0.04137 0.1853 0.03237 57 0.02085 0.00131 0.04001 0.02431 0.07802 0.02666 722 0.002422 0.0133 0.04925 0.01564 0.05114 0.004655 266 0.02442 0.007306 0.04628 0.004371 0.01095 0.00697 01 0.07245 0.00895 0.1101 0.00833 0.1157	-0.03638	0.1625	-0.03284	0.1717	-0.01659	0.1261	3		
Specify current data columns. 0.2051 0.04907 0.1737 0.04087 0.1289 0.02522 Direction (deg) CDIR 0 0.0311 0.04138 0.146 0.04777 0.1977 0.06138 Wave Data 0 0.06425 0.009418 0.09811 0.004245 0.1182 -0.01211 187 0.1005 -0.01045 -0.08743 -0.004734 -0.01349 -0.0155 09 -0.209 0.05336 -0.1797 0.04137 -0.1853 0.03237 57 -0.02085 0.00131 -0.04001 0.02431 -0.07802 0.02666 722 -0.004242 -0.01033 -0.04925 -0.01564 -0.05114 0.004655 226 0.02442 0.007306 0.04628 0.004371 0.01095 0.00697 01 0.07245 -0.00895 0.1101 -0.008333 0.1157 -0.02096 48 -0.0024213 0.01828 0.03929 0.004252 0.04437 0.008455	 0.1281	-0.437	0.09819	-0.2873	0.0718	-0.1912	8		Current Data
Speed CSPD -0.2051 0.04907 -0.1737 0.04087 -0.1289 0.02522 13 -0.1331 0.04138 -0.146 0.04777 -0.1977 0.06138 Direction (deg) CDIR - - - - - - - - - - - - - - - - 0.04087 - - - - - - - 0.04087 - 0.1289 0.02522 - - - - - - - 0.1977 0.01977 0.01977 0.01977 0.01974 - 0.01134 - 0.01245 0.01134 - 0.01473 - 0.0183 0.03237 57 - 0.0205 0.01331 - 0.04001 0.02431 - 0.02366 722 - 0.0004242 - 0.01331 - 0.04001 0.02431 - 0.02455 - 0.01544 - 0.02146 0.004252	0.04436	-0.1784	0.03951	-0.1361	0.03684	-0.102	5	ta columos:	Specify current da
Direction (deg) CDIR 13 0.1331 0.04138 0.146 0.04777 0.09777 0.06138 Direction (deg) CDIR 09 0.06425 -0.009418 0.09981 0.004245 0.1182 -0.01211 187 -0.1005 -0.01045 -0.08743 -0.004734 -0.01349 -0.01155 09 -0.209 0.05336 -0.1797 0.04137 -0.1853 0.03237 Wave Data 57 -0.02085 0.001331 -0.04001 0.02431 -0.07802 0.02366 Specify wave data columns: - - - - - - 0.004242 - - 0.04332 - 0.01351 - 0.04371 0.01095 0.02665 126 0.02442 -0.01033 -0.04925 -0.01564 -0.05114 0.00657 01 0.07245 -0.008955 0.1101 -0.008333 0.1167 -0.02096 48 -0.0004213 0.01828 0.03929 0.004252 0.04437	 0.02522	-0.1289	0.04087	-0.1737	0.04907	-0.2051			
Wave Data -0.0005 -0.01045 -0.08743 -0.004734 -0.01349 -0.01155 Specify wave data columns: -0.000420 0.0209 0.05336 -0.1797 0.04137 -0.1853 0.03237 Height WVHT -0.0004242 -0.01033 -0.04925 -0.01564 -0.05114 0.00465 10 0.07245 -0.00895 0.1101 -0.008393 0.1167 -0.02096 48 -0.0004213 0.01828 0.03239 0.04437 -0.04437 -0.008355	0.06138	-0.1977	0.04777	-0.146	0.04138	-0.1331	13	CSPD ~	Speed
Wave Data -0.005 -0.0145 -0.08743 -0.01479 -0.01349 -0.01155 Dy ave Data 0.0209 0.05336 -0.1797 0.04137 -0.1853 0.03237 Specify wave data columns:	-0.01211	0.1182	0.004245	0.09981	-0.009418	0.06425	09	1	
Wave Data 57 -0.02085 0.001311 -0.04001 0.02431 -0.07802 0.02366 Specify wave data columns: Height -0.0004242 -0.01033 -0.04925 -0.01564 -0.05114 0.00452 Leight WVHT -0.07245 0.00895 0.1101 -0.00833 0.1167 -0.00964 Direction (dep) WVDIR -0.0004213 0.01828 0.03929 0.004252 0.04437 -0.008545	-0.01155	-0.01349	-0.004734	-0.08743	-0.01045	-0.1005	187		Direction (deg)
Specify wave data columns: 722 -0.0004242 -0.01033 -0.04925 -0.01564 -0.05114 0.00465 Height WVHT 0 0.02442 0.007306 0.04628 0.004371 0.01095 0.00697 Direction (dep) WVDIB 0 0.002421 0.01828 0.03929 0.04252 0.04437 -0.009645	0.03237	-0.1853	0.04137	-0.1797	0.05336	-0.209	09		
Understand User	0.02366	-0.07802	0.02431	-0.04001	0.001331	-0.02085	57		Wave Data
Height WVHT 226 0.02442 0.007306 0.04628 0.004371 0.01095 0.00697 Direction (dea) WVDIB 0.07245 -0.00895 0.1101 -0.008333 0.1167 -0.02966 48 -0.0004213 0.01828 0.03929 0.004252 0.04437 -0.008545	0.00465	-0.05114	-0.01564	-0.04925	-0.01033	-0.0004242	722	a columns:	Specify wave data
Direction (dep) WV/DIR 01 0.07245 -0.00855 0.1101 -0.00833 0.1167 -0.0296 WV/DIR 48 -0.0004213 0.01828 0.03929 0.004252 0.04437 -0.008545	0.00697	0.01095	0.004371	0.04628	0.007306	0.02442	226		
Direction (deg) WVDIR	-0.02096	0.1167	-0.008393	0.1101	-0.00895	0.07245	01	WVHT ~	Height
Direction (deg) WVDIR	-0.008545	0.04437	0.004252	0.03929	0.01828	-0.0004213	48	hanna	-
07 -0.0003726 0.05574 -0.03808 0.04123 -0.02744 0.04941	0.04941	-0.02744	0.04123	-0.03808	0.05574	-0.0003726	07	MADIK ~	Direction (deg)
9 -0.2809 0.1204 -0.2872 0.1212 -0.1852 0.1123	0.1123	-0.1852	0.1212	-0.2872	0.1204	-0.2809	9		
Reach Angle (deg): [21.1542739868164 02 -0.2167 0.07601 -0.1695 0.0632 -0.222 0.0754	0.0754	-0.222	0.0632	-0.1695	0.07601	-0.2167	02	Angle (deg): 21.1542739868164	Beach

B.3. Scroll to the far-right end of the table. <u>Six</u> new columns have been added to the end of the global data sheet and that the unprocessed wind, current, and wave data columns are now inactive (red text):

Wind A_comp: along-shore wind speed Wind O_comp: toward shore wind speed Current A_comp: along-shore current speed Current O_comp: toward shore current speed Wave_A_comp: along-shore wave height Wave_A_comp: on-shore wave height

Location Glob	al Datasheet										
ort Validate Data A O	Manipulate Work with	Transfor Data	m Go To Model								
File	1-VB-Training-E	Data:	6	CPERP24	WindA_comp[WDII	WindO_comp[WDI	CurrentA_comp[CD	CurrentO_comp[CD	WaveA_comp[WVI	WaveO_comp[[W. ^
Column Count	101			-0 03638	5.224	1.006	0.09794	0.01167	0.2734	0.2603	
Row Count	281			0.1281		0.00	-0 1678	.0.0771		-0.1104	
Date-Time Index	DATETIME	_		0.04436	-0.8499	-2.613	-0.07477	-0.03144	-0.07845	-0.06703	
Response Variable	ECOLI	- 11		0.02522	-8.08	-3.52	-0.2327	-0.06925	-0.6112	0.01645	_
Disabled Row Count	0			0.06138	-4.677	-1.274	-0.1248	-0.04007	-0.3093	0.0552	-
Disabled Column Count	6			-0.01211	-1.941	0.4918	0.0493	0.02722	-0.01497	0.01758	-
Hidden Column Count	0			-0.01155	0.04079	-2.718	-0.0708	-0.01735	-0.1998	-0.002227	-
Independent Variable Count	93			0.03237		-2 762		-0.04655		0 06824	-
<		>	c			7 110	171.11	11 10 16 6.7	-0-3000		>

C. Combine two or more predictive variables

Interaction Terms: In situations where two predictive variables are themselves correlated, meaning they interact with one another in terms of how they influence water quality, it may be beneficial to combine them into a single interaction term by **multiplying** them together. Combined the two variables may be better predictors of water quality than if included individually.

Combined Categories: Some variables are either yes or no. The 0 is "no" and the 1 is "yes". In situations where binary variables represent successive categories of some qualitative variable, like visually-observed water clarity, it may be useful to combine them into a single binary variable by **summing** them. The resulting variable will have a value of 1 when *either* of the two conditions is present. This can be especially helpful when there is little functional distinction between the categories or few cases in which one of the conditions is ever observed. In this example, the difference between TURBID and OPAQUE water is not very distinct; if the water is turbid, it was probably also opaque.

Change-in-Flow Variables. In situations where continuous stream flow data are preprocessed over different timeframes, **subtracting** one temporal snapshot from another can create proxy variables for *changes* in flow. The difference between 24-hour maximum and minimum flow rates indicates whether recent tributary discharge has been consistent or very different after a flash flood event.

C.1. **1.** Click the **Manipulate** icon. **2.** In the pop-up window, ctrl-select TRIBmax24 and TRIBmin24. **3.** Click the right-arrow ">" button.

rt Validate Comput a Data A D d Validate	e Manipulate Tra Work with Da	QTRSEASON PRE_JUNE21 JUNE21_JULY15	^ >	Variables in Expression	
ile	1-VB-Training-Data	JULY16_AUG10 POST_AUG10 DOY RRAIN6 RRAIN24	<	● Sum ○ Diff ○ Max ○ 1	Min () Mean () Product
Row Count	281	RRAIN48 RRAIN72		T	
ate-Time Index	DATETIME	RRAIN120			
Response Variable	ECOLI	RRAIN144		Add Remove	2nd Order Interactions
Disabled Row Count	0	TRIB24 TRIBmax24			
isabled Column Count	6	TRIBmin24			
Hidden Column Count	0	TRIBmax48			
ndependent Variable Count	93	TRIBmin48	2 -		

C.2. 1. Click the radio button next to **Difference**. 2. Click the **Add** button.

ie.	Location	Globa	Datasheet	🖳 Manipulate				7
Dort	Validate	Compute	Manipulate Tra	Build Expression Independent Variables			Variables in Expression	
ta Id	Data Validate	AO	Work with Da	QTRSEASON PRE_JUNE21 JUNE21_JULY15 JULY16_AUG10 POST_AUG10	^	>	TRIBmax24 TRIBmin24	
File Colum	in Count		1-VB-Training-Data	DOY RRAIN6 RRAIN24 RRAIN48			O Sum Diff O Max O Min O Mean O Produc	t
Row C			281	RRAIN72			DIFEITRIBmax24,TRIBmin24]	
	Time Index		DATETIME	RRAIN120				_
Respo	onse Variable		ECOLI	RRAIN144 TRIB6			Add Jamove 2nd Order Interaction	ns
Disabl	led Row Cour	nt	0	TRIB24 TRIB48			2	
Disabl	led Column C	ount	6	TRIBmax48			-	
Hidde	n Column Cou	unt	0	TRIBmin48 TRIB72				
Indepe	endent Variat	ble Count	93	TRIB168	*			
<			2		-	ОК	Cancel	

C.3. The change-in-flow variable **DIFF[TRIBmax24,TRIBmin24]** has been added. This approximates whether and to what extent the previous 24 hours of tributary discharge has been constant or varied a lot.

	al Datasheet	🖳 Manipulate	
Validate Compute ta Data A O	Manipulate Tra Work with Da	Build Expression Independent Variables QTRSEASON PRE_JUNE21 JUNE21_JULY15 JULY16_AUG10	Variables in Expression TRIBmax24 TRIBmin24
Column Count Row Count Date-Time Index Response Variable Disabled Row Count Disabled Column Count Hidden Column Count	1-VB-Training-Data 101 281 DATETIME ECOLI 0 6 0 93	POST_AUG10 DOY RRAIN6 RRAIN24 RRAIN24 RRAIN72 RRAIN120 RRAIN120 RRAIN144 TRIB6 TRIB24 TRIB24 TRIB48 TRIBmax48 TRIBmax48 TRIBmi48 TRIBmi48 TRIB522 TRIB168	Sum Diff Max Min Mean Product DIFF[TRIBmax24,TRIBmin24] Add Remove 2nd Order Interaction DIFF[TRIBmax24,TRIBmin24]

C.4. **1.** Since the expression has been added, remove the two variables by shift-selecting TRIBmax24 and TRIBmin24. **2.** Click the left-arrow "(" button to move the variables back to the main list.

File	Location	Global Datasheet	🛃 Manipulate —		
Import Data Add	Validate Data Validate	Compute Manipulate Tra A O Work with Da	Build Expression Independent Variables Variables in Express QTRSEASON PRE_JUNE21 JUNE21 JUNE21 JULY15 JULY16,AUG10 POST_AUG10		
File Colum	nn Count	1-VB-Training-Dat 101	DOY RRAIN6 RRAIN24 O Sum O Diff O Max O Min O Mean O Product		
	Count	281	RRAIN48 RRAIN72 DIFF[TRIBmax24,TRIBmin24]	1	
	Time Index	DATETIME	RRAIN120		
Hesp	onse Variable	ECOLI	TRIB6 Add Hemove 2nd Order Interactions		
Disab	led Row Count	0	TRIB24 TRIB48 DIFF[TRIBmax24,TRIBmin24]	1	
Disab	led Column Co	unt 6	TRIBmax48		

C.5. Repeat the steps in C.2 and C.3 again to create expressions of combined variables as needed and click **OK**. Scroll to the far-right end of the table to see any new columns added through this process.

e	Location	Global	Datasheet									
ort '	Validate (Data	Compute A O	X Manipulate	S	rm Go To Model							
	Validate	AU	Work wit	h Data	Moder							
File		1	-VB-Training	-D. ^ [4	WindA_comp[WDII	WindO_comp[WDI	CurrentA_comp[CD	CurrentO_comp[CD	WaveA_comp[WVI	WaveO_comp WV	DIFF[TRIBmax24,1 ^
Column	Count		02		•	5.224	1.006	0.09794	0.01167	0.2734	0.2603	36
low Co	unt	2	81			-0.2859	-6.06	-0.1678	-0.0771	-0.3524	-0.1104	36
	ne Index		DATETIME			-0.8499	-2.613	-0.07477	-0.03144	-0.07845	-0.06703	23
Response Variable		t	COLI	- 11		-8.08	-3.52	-0.2327	-0.06925	-0.6112	0.01645	156
isabled	Row Count	0				-4.677	-1.274	-0.1248	-0.04007	-0.3093	0.0552	142
inshlar	d Column Cou	unt 6				-1.941	0.4918	0.0493	0.02722	-0.01497	0.01758	11
	Hidden Column Count					0.04079	-2.718	-0.0708	-0.01735	-0.1998	-0.002227	21

D. Transform the dependent bacteria variable

D.1. To build a usable nowcast model, bacteria counts must be transformed. Log10 is a common transformation for microbial concentrations. Right-click on the "ECOLI" column header and select Transform > Log10. Save your project file. You can now move onto the next module, "Building a GBM Model".

Fie	Location	Global Datasheet											
nport Vi Jata	alidate Data /alidate	Compute A O Work with		Go To Model									
File		1-VB-Training-D), ^	DATETIME	ECOLI	Territori		110	JULY15	JULY16_AUG10	POST_AUG10	DOY	~
File Column C	Count	1-VB-Training-D 102); ^	DATETIME 5/21/2009 12:05		Transform	5	Log10	JULY15	JULY16_AUG10	POST_AUG10	DOY 141	^
Column C Row Cou	unt	102 281); ^		2	Transform View Plots	>	Log10 Ln	JULY15	JULY16_AUG10 0	-	1	^
Column C Row Cou Date-Tim	unt	102); ^	5/21/2009 12:05	2 5	Transform	2	Log10	JULY15	JULY16_AUG10 0 0	0	141	^