

Data Preparation for B-IBI recalibration

12/10/12 by Jo Wilhelm, King County

DATA CORRECTIONS

B-IBI scores and associated metrics were originally downloaded from the Puget Sound Stream Benthos (PSSB, www.pugetsoundstreambenthos.org) on 11/28/12. The data were screened for typos and potential mistakes and corrections were made as feasible:

- **Geography:**
 - Latitude and Longitude were mapped to make sure data appeared correct, and in all cases site coordinates projected within the state of Washington (previous data entry errors placed some sites in the middle of the Pacific Ocean).
 - WRIA classification was also checked. One location, Pierce County site BIBI-042 - Lacamas Creek had erroneously been grouped in WRIA 26 (Cowlitz) because there are at least two Lacamas Creeks in Washington State. A second Lacamas Creek was added to the PSSB so that there is now a Lacamas Creek (Cowlitz) and a Lacamas Creek (Nisqually) and the WRIA classification for BIBI-042 – Lacamas Creek was corrected to be in WRIA 11 (Nisqually).
- **Date:** Three sites were identified where the sampling month did not match other sampling time frames for the given project.
 - Pierce County site BiBi-010 - Canyon Falls originally stated that it was sampled on 12/8/11. Carla Vincent checked the original data sheet and determined it should have been August 12, 2011 (8/12/11, not 12/8/11). This was corrected in the PSSB.
 - Kitsap County site KCSSWM-042 on Wilson Creek was sampled 12/06/06. The [original data](#) were consulted, and it was confirmed that the data sheet says December (and it spells it out).
 - King County site 09NEW2076 on Newaukum Creek was originally entered as March 31, 2005. Based on the [original data](#), this was corrected to 8/31/05 in the PSSB.
- **Surface area:** Data were screened for surface areas other than 3, 8, or 9 square feet total. A lot of data were entered or sample metadata was corrected to fix several errors (see summary table).

Sq Ft	# Visits	Corrected	Not Corrected	Notes
1	52	52	0	51 were changed from 1 ft ² to 3 ft ² because of data entry error; 1 had data for two missing replicates entered
2	12	5	7	5 samples were updated, corrected, or missing surface area information was corrected; After consulting original data, 7 samples were confirmed to only have 2 ft ² for one reason or another.
5	1	1	0	
6	6	0	6	All were part of 9 ft ² protocol, but one rep not collected or reported so only 6 ft ² . Checked original data.
13.5	14	0	14	NOAA slack sampler method; they did collect 13.5 ft ²
16	9	5	4	The 5 fixed were all Kitsap-EPA side by side sites; fixed by making EPA site a QC rep; the other 4 were Ecology sites, only 2 in Puget Sound and Karen Adams confirmed those were mesocosm sites and should be omitted b/c of methodology differences.

DATA DOWNLOAD

Once all the above corrections were made to the data in the PSSB, data were ready to be downloaded again on 12/5/12 for all stream visits. Below are the PSSB metadata that describe the criteria used for this download:

All Site visits

Title: Scores

Date Filter: All events at each site

Replicate Handling: Sum replicates' quantities, then calculate scores

Taxa Attributes: Fore, Wisseman (2012) Draft 4

Taxa Resolution: Resolution used by lab

Custom Taxa Metrics: Chironomidae, Oligochaeta, Acari

Organisms per Visit: At most 500, subsampled when over

Sorted by: Location, Agency, Project, Site Code, Date

Generation Time: Wednesday, December 05, 2012 11:40 AM

http://pugetsoundstreambenthos.org/Download/Scores.ashx?R=4&TA=2012&adv=1&TM=Chironomidae%2C%20Oligochaeta%2C%20Acari&MinO=-1&MaxO=500&oo=7&d=4

Internal King County [link](#) to data spreadsheet.

DATA CLEANUP

There are four Snohomish County sites whose site codes always download as dates. These were edited to text.

PSSBCode	DownloadError	SiteID
5-54	May-54	983
5-42	May-42	952
5-37	May-37	951
7-72	Jul-72	958

The data download generated 3,827 site visits which were reduced down for the following reasons:

- 235 site visits not in the Puget Sound basin
- 96 site visits > 500 m elevation
- 118 site visits sampled outside of July – October (many of these are from projects on ephemeral streams)
- 2 site visits for methodology differences (e.g., mesocosm recolonization)
- 23 site visits for surface area (not 3, 8, or 9 sq ft)
- 241 site visits do not have land use/land cover data (LULC data was generated for 1,132 sites, but 58 of these do not have invertebrate data and some PSSB sites have been added since this analysis was done in early 2011).

This leaves 3,209 site visits (and 618 that were omitted for the above reasons), which were further divided:

- There are 856 most recent site visits (test data set)
 - If a most recent site visit was omitted for one of the reasons noted above, the next most recent site visit that met all criteria was selected.
- There are 610 2nd most recent site visits (validation data set)

Summary Table comparing the Development Data Set (N=856) and Validation Data Set (N=610).

Variable	Development Data Set (N=856)				Validation Data Set (N=610)			
	Mean	Min	Max	StDev	Mean	Min	Max	StDev
Elevation (m)	86.2	0.5	647.3	94.1	79.7	0.5	618.7	87.6
Total SqFt	5.4	2.0	9.0	2.8	5.1	2.0	9.0	2.7
EPT Richness Quantity	14.0	0.0	37.0	7.0	13.5	0.0	32.0	6.9
Taxa Richness Quantity	29.1	3.0	70.0	9.7	28.5	6.0	70.0	9.1
Chironomidae Richness	3.5	0.0	28.0	4.5	3.2	0.0	26.0	4.1
Oligochaeta Richness	1.0	0.0	7.0	0.6	1.0	0.0	4.0	0.3
Acari Richness	1.0	0.0	8.0	0.8	0.9	0.0	4.0	0.4
Overall Score	30.0	10.0	50.0	9.8	29.9	10.0	50.0	9.6
P_Forest_WS	52.6	0.0	100.0	31.6	49.2	0.0	100.0	31.3
P_Wetland_WS	1.8	0.0	21.5	2.4	1.7	0.0	14.8	2.2
P_Urban_WS	27.9	0.0	97.7	30.3	31.4	0.0	97.7	30.5
P_Ag_WS	13.3	0.0	62.4	11.1	13.9	0.0	56.8	10.8
P_Forest_WS_1km	48.3	0.0	100.0	30.1	45.5	0.0	100.0	30.2
P_Wetland_WS_1km	3.3	0.0	40.8	4.9	2.8	0.0	40.8	4.4
P_Urban_WS_1km	28.1	0.0	99.0	28.1	32.0	0.0	99.0	28.8
P_Ag_WS_1km	16.2	0.0	82.6	13.9	16.4	0.0	80.7	13.9
P_Forest_BF	53.8	0.0	100.0	30.5	50.8	0.0	100.0	30.4
P_Wetland_BF	3.3	0.0	58.7	4.5	3.0	0.0	23.9	3.9
P_Urban_BF	24.5	0.0	95.0	27.8	27.9	0.0	95.0	28.5
P_Ag_BF	14.4	0.0	87.1	12.0	14.9	0.0	87.1	11.7
P_Forest_BF_1km	49.9	0.0	100.0	30.2	47.4	0.0	100.0	30.7
P_Wetland_BF_1km	5.1	0.0	58.7	7.6	4.4	0.0	57.6	6.8
P_Urban_BF_1km	24.3	0.0	97.5	26.5	27.9	0.0	97.5	27.5
P_Ag_BF_1km	16.9	0.0	87.9	15.0	17.3	0.0	87.1	15.1
Tot_Rd_Lgth_WS	61450.0	0.0	944406.5	105863.9	64190.2	0.0	944406.5	109367.6
Rd_Density_WS	4.4	0.0	18.4	4.1	5.0	0.0	18.4	4.2
RdPerStrmXperKM_WS	2.2	0.0	14.2	2.3	2.4	0.0	14.2	2.3
Tot_Rd_Lgth_WS_1km	4397.3	0.0	34317.2	4464.1	4803.3	0.0	34317.2	4566.0
Rd_Density_WS_1km	4.7	0.0	21.5	3.9	5.2	0.0	21.5	4.0
RdPerStrmXperKM_WS_1km	2.1	0.0	17.8	2.2	2.3	0.0	17.8	2.3
PopDensity_WS	538.0	0.0	3163.9	737.7	612.2	0.0	3163.9	764.7
PopDensity_WS_1km	542.0	0.0	4116.6	751.7	627.4	0.0	4116.6	797.5
WA_area_HC_WS	2417.6	9.2	142416.9	6556.0	1977.3	9.2	70894.4	4317.2
WA_area_HC_WS_1km	92.8	3.0	214.3	40.2	90.9	9.2	213.9	40.0
Pslope_mean_WS	13.9	1.5	65.9	12.2	11.8	1.5	61.7	9.2
Precip_Mean_mm_WS	1436.5	475.8	3737.4	553.4	1339.0	506.6	3335.8	394.6
Length_stream_m_WS	48082.1	0.5	2588371.3	124523.2	40025.9	0.5	1372540.1	86534.3
Density_Stream_WS	2.0	0.0	6.7	0.4	2.0	0.0	6.7	0.4

The richness of chironomids, Acari, and oligochaetes were determined for future analysis regarding taxonomic resolution and area output in the downloaded data.