Guidance for calculating the MPCA's human disturbance score (2006 version)

Biological Monitoring Program

I. Introduction

The Minnesota Pollution Control Agency (MPCA) has developed indices of biotic integrity (IBIs) in an effort to assess the aquatic life for streams in Minnesota. In order to develop IBIs, a way to quantify human disturbance was needed to establish the IBI. To estimate human disturbance at stream sites the MPCA's Biological Monitoring Unit has designed a system to score biological sites according to the degree of cumulative human disturbance calculated from land use statistics, proximity to feedlots surface water discharges and urban land use, riparian condition, and degree of channelization. The MPCA uses the human disturbance score (HDS) to rank sites based on the degree of disturbance. The ranked sites were then used to test candidate metric response used for the statewide index of biotic integrity (IBI). Well performing IBI metrics demonstrate a response to this human disturbance gradient (e.g., percent of tolerant individual's increases with increasing disturbance).

HDS Development Dataset: A total of 1563 biological monitoring sample sites were used in the dataset to develop the HDS. Sites were used for HDS development if they met certain criteria: 1) Sites are not included in the ranking process if they are MPCA sites that were not sampleable for both invertebrates and fish, b) some non-reportable MPCA sites are included in the ranking if the reason for them being considered non-reportable was related to the stream classification and not some problem with the sampling (e.g., coldwater streams), c) Data from other sources was/is included if the sampling method used met the requirements for IBI assessment, d) streams with reach lengths less than about 80 m were automatically disqualified with the exception of MPCA sites sampled during MRAP and the Red River project because this data was collected for IBI purposes and it has already been used for assessment. d) We did not keep any Department of Natural Resources (DNR) historical data collected prior to 1986 because we surmised that anything earlier would not be representative of our earliest land use coverage.

II. Initial sites screening criteria

Sites are NOT ranked if they do not meet certain criteria. The following list outlines the site screening process used in previous HD scoring time periods

A. Sites are not included in the ranking process if they are MPCA sites that were not sampleable for both invertebrates and fish.
B. Some non-reportable MPCA sites are include in the ranking if the reason for them being considered non-reportable was related to the stream classification and not some problem with the sampling (e.g., coldwater streams)

C. Data from other sources is included if the sampling method used met the requirements for IBI assessment.
D. Streams with reach lengths less than about 80 m were automatically disqualified with the exception of MPCA sites sampled during MRAP and the Red River project because this data was collected for IBI purposes and it has already been used for assessment.
E. DNR historical data collected prior to 1986 was not used because we surmised that anything earlier would not be representative of our earliest land use coverage.

III. Criteria for scoring

The HDS system incorporates both watershed statistics and site specific criteria (Table 1). The watershed statistics are calculated using automated GIS processes; however, the site specific criteria require desk top viewing of each site. There are five criteria that require desktop viewing; guidance for scoring the site specific criteria is described below. The following table outlines the scoring system and criteria used to calculate a human disturbance score.

Table 1 Site Rank Table (updated 6/25/07 – quartiles include all data up to 2006 using 1563 total sites in dataset)

Stressor	Metric	Primary metric (10 pts max)	Secondary metric Upward Adjustment (1 point)	Secondary Metric Downward Adjustment (1 point)	Proximity Downward Adjustment (1 point)	3 x interquartile range transformed if necessary (lower, upper)	Comments
Agricultural Land use	% agriculture in watershed	Yes (cont. 0-10)				No outliers	Arcsine transformed Slight negative drainage area relationship
	% agriculture in 100 meter buffer	No	No upward adjustment	Upper Quartile >49.1		No outliers	Slight negative drainage area relationship
	Amount of agricultural landuse on 3% slope as a percentage of total watershed area of 3% slope	No	No upward adjustment	Upper Quartile >57		No outliers	
Urban land use	% impervious	Yes (cont.,0-10)			City or town at the site or immediately adjacent to site	0, 2.768	Log Transformed, Slight negative drainage area relationship
	# of road crossings per stream km	No	Lower Quartile	Upper Quartile		0, 0.986	Slight negative

Stressor	Metric	Primary metric (10 pts max)	Secondary metric Upward Adjustment (1 point)	Secondary Metric Downward Adjustment (1 point)	Proximity Downward Adjustment (1 point)	3 x interquartile range transformed if necessary (lower, upper)	Comments
			<0.164	>0.395			drainage area relationship
Feedlots	# of animal units per km^2	Yes (cont.0-10)			Visual evidence (from DOQ) of feedlot at site or immediately upstream of site	No outliers	Log transformed, No drainage area relationship
	# of feedlots per km^2	no	No upward adjustment	Upper Quartile >0.235		0,0.754	
Point sources	# of point sources per sqkm^2	Yes (cont.0-10)			Continuous discharge <5 stream miles into stream <50 Mi.^{2}	0, 0.444	Sqrt transformed, Negative drainage area relationship
Watershed riparian	% disturbed riparian habitat in the watershed	Yes (cont.0-10)				No outliers	Arcsine transformed No drainage area relationship
Site riparian	DOQ	Yes	Refer to table for scoring guidance				
Site channelization	Channel Condition	Yes (0 or 10)	For sites not completely channelized, add a point for each 10-% of non-channelized stream.				
Watershed channelization	% channelized stream per stream km	Yes (cont. 0-10)				No outliers	Sqroot transformed, Negative drainage

Stressor	Metric	Primary metric (10 pts max)	Secondary metric Upward Adjustment (1 point)	Secondary Metric Downward Adjustment (1 point)	Proximity Downward Adjustment (1 point)	3 x interquartile range transformed if necessary (lower, upper)	Comments
							area relationship

a) Scoring guidance for five site specific criteria (Manual Process)

There are five stressors you will need to examine, and score appropriately (Table 2). The first three are related to the proximity of stressors to the site and the last two relate to the characteristics within the sampling reach of the site. Notice that some of the scoring criteria are based on watershed size (i.e. Feedlots and Ditching). You will only need to review these criteria where the appropriate watershed criteria apply.

For metrics that require the use of digital orthophotography layers. **Use the new color orthophotography for sites sampled from 1998 to present;** for sites sampled in 1997 and earlier use the old digital orthophoto quadrangle (DOQ) layers. Previous note: In instances where the site riparian zone appears similar on both digital layers, it may be more appropriate to use the older DOQs to estimate the amount and proximity of the disturbance.

Table 2 Proximity-related stressors

Stressor	Database field	Score adjust ment	Watershed exceptions	Adjustment criteria
Urban land use	[UrbankLUProx]	-1		City or town at site or adjacent to site, or within a reasonable distance upstream to cause some measurable effect within the site reach.
Feedlots	[FeedlotProx]	-1	applies only to streams <50mi^2	Visual evidence (from orthophotography layer) of feedlot at or immediately upstream of site.
Point Sources	[PointSourceProx]	-1		**Continuous discharge <5 stream miles upstream.**
Ditching		0 to 10	applies only to streams <100 mi^2	Contrast the ditch layer % calculation with a visual estimate of ditching from the orthophotography layers (only for sites <100 mi^2). If the difference between the two layers is significant, estimate the actual % ditching and score according to Table 2.
Riparian condition		0 to 10		Average of % disturbance from 0 to 30 m and % disturbance from 0 to 15. Score according to Table 3.

b) Guidance for scoring stressors:

i. **Urban land use:** We will have to review examples with staff that applied these criteria in previous rounds in order to understand when the score adjustment is applied so that we can hopefully develop a consensus as to what is considered a "reasonable distance upstream to cause some measureable effect." Remember that this is score adjustment only applies to sites <100mi2.

ii. **Feedlots:** We will review some sites with feedlots of different sizes and proximities in order to determine where this score adjustment applies. Remember that this score adjustment only applies to sites <50 mi^2.

iii. **Point Sources**: This criterion observes whether or not there is a surface water discharge less than five miles upstream of the reach.

iv. **Ditching:** The ditch coverage underestimates the actual percent ditching in a watershed. For large watersheds we might assume that the percent ditching at any given location is underestimated in roughly the same proportions, however for small watersheds the estimates can be very different relative to each other. To attempt to correct for this problem we are going to contrast the ditch layer with a visual estimate of ditching from the orthophotography layers for all sites <100 square miles drainage area (DA). If the difference is significant, estimate the actual percent ditching and score according to Table 3 below.

v. **Riparian conditions:** View the riparian zone from 0 to 15 m and 0 to 30 m. For each distance, visually calculate the percent of disturbance observed (row crop, lawn or golf courses, roads, etc). Take the average percent disturbance of these two calculations and scores according to Table 4.

Table 3
Ditch Watershed scoring
Visual estimate of the **% ditching** in the watershed (sites<100 square miles DA)

% ditch	New scoring (from Scott's notes)
0	10
10	7
20	5.5
30	5
40	3.5
50	3
60	2
70	1
80	0.5
90	0.25
100	0

Table 4

Score	RiparianCondition Average of % disturbance from 0-30 m. and % disturbance from 0-15 m.
10	Site has fully intact riparian zone (i.e. natural veg.) within approximately 30 m of stream banks (both banks, full length)
9	Almost all (>90%) riparian zone is natural with less than 10% disturbance (i.e. row crop agriculture, pasture, urban, or residential disturbance)
8	80 to 90 % natural
7	70 to 80 % natural
6	60 to 70 % natural
5	50 to 60 % natural
4	40 to 50 % natural
3	30 to 40 % natural
2	20 to 30 % natural
1	10 to 20 % natural
0	<10 % natural

vi. **For watershed statistics:**

First, check to be sure drainage delineations look appropriate. Obvious errors should be noted and not used to score until a correction is made.

At least 40% of the watershed area is required to rank a site. The 40% cutoff was chosen because it would allow us to rank the lower mainstem sites on the St. Croix, Red, and Minnesota Rivers. Statistics from sites that have between 40 and 100% of their watershed area in Minnesota have been adjusted by reducing the stream miles in the watershed by the percentage of watershed area outside of the state.

For automated scoring of watershed characteristics, outliers were removed by determining the interquartile range of the dataset and setting the top score as the value that equal three times the interquartile range. Sites scoring less than 0 were adjusted to 0.

Transformations were done on all continuously scored metrics to reduce the Skewness. In all cases, the transformation decreased the skewness appreciably but not all metrics had a skewness/SE skewness <2. No metrics are normally distributed.

We will wait to assign any kind of weighting to each metric until after the scoring process is complete.

Most metrics have a slightly negative relationship with drainage area. For primary metrics this means that a greater proportion of small streams will have the lowest scores. For secondary metrics it means that corrections that tend to lower the score will be conducted on a greater proportion of small streams.

Table 5 Ditch Watershed scoring
Visual estimate of the % ditching in the watershed (sites mi^2 DA).

% ditch	Old scoring	New
0	10	10
10	8	7
20	7	5.5
30	6.3	5
40	5.5	3.5
50	5	3
60	4.4	2
70	3.7	1
80	3	0.5
90	2	0.25
100	0	0

c) Automated scoring process: Overall land use analysis

i. Data sources:

a) NLCD 2001 (V: \land_cover\nlcd_2001\mn_nlcd)

b) Drainage area polygons clipped to MN boundary

ii. The land use composition analysis was run using the tabulate areas Python script for overlapping polygons. This script has the unfortunate pattern of running fine through several (between 2 and 300) drainage polygons, then bailing with no explanation. The solution used was to break the drainage area polygons into separate feature classes of < 500 polygons each, running the script concurrently on two machines on the subsets, restarting as needed, until all polygons had been analyzed.

a) Script: X:\Databases\Water_Quality\Biological_Monitoring\GIS\Scripts

d) Land use analysis for areas with > 3% slope

i. Data sources:

b) NLCD 2001 (V:\land_cover\nlcd_2001\mn_nlcd)

c) NED 30 m DEM (V:\elevation\national_elevation_dataset\ned30\mn_ned)

d) Drainage area polygons clipped to MN boundary

ii. The first step is to create the slope surface, which is done by running the Spatial Analyst Slope tool, setting the output measurement to PERCENT_RISE.

iii. Next, the Spatial Analyst Reclassify tool is used to set all slope values greater than 3 to 1, and all values less than 3 to <no_data>

iv. The resultant layer (with all cells having a value of 1) is added to the NLCD 2001 raster using the raster calculator, subtracting 1 from the output values to yield a layer with correct land use class values.

v. The land use composition analysis was run using the tabulate areas Python script for overlapping polygons. This script has the unfortunate pattern of running fine through several (between 2 and 300) drainage polygons, then bailing with no explanation. The solution used was to break the drainage area polygons into separate feature classes of < 500 polygons each, running the script concurrently on two machines on the subsets, restarting as needed, until all polygons had been analyzed.

- Script X:\Databases\Water_Quality\Biological_Monitoring\GIS\Scripts
- Land use raster for land with > 3% slope: X:\Databases\Water_Quality\Biological_Monitoring\Streams\Landuse\nlcd01lu_3pct

e) Riparian land use data development and analysis

i. Data sources:

a) NLCD 2001 (V:\land_cover\nlcd_2001\mn_nlcd)

b) NHD 1:24k flowlines, areas, and waterbodies

c) Drainage area polygons clipped to MN boundary

ii. Buffered data – all buffers were produced by basin

a) 100m buffer polygons were generated from 1:24k flowlines, areas, and waterbodies.

b) The erase tool was then used, first with the area buffer erasing the flowline buffer, then with the waterbody buffer erasing the flowline buffer, then the area buffer erasing the waterbody buffer. The yields three polygon data sets that buffer all flowlines, areas, and waterbodies by 100m, but do not overlap. These steps were linked into a model, then run using batch processing for each basin. The individual basin feature classes for each polygon type were then loaded into a common feature class for each polygon type using ArcCatalog.

c) Buffered waterbodies with type = wetland were extracted from the original (non-erased) waterbody buffer feature class. This subset was used in a location selection to select those wetlands that overlapped either the original (non-erased) flowline or area polygons. The selection was then switched, yielding a selected set that included only swamp/marsh polygons that were not within 100 m of a flowline or area polygon. These polygons were not included in the analysis.

d) The buffered polygons were converted to separate raster layers (by basin), which were then mosaicked together. The resulting raster had a homogenous cell value (1). The Spatial Analyst raster calculator was used to add the riparian grid to the NLCD 2001 grid, yielding a data set that had zero values outside of the riparian areas, and land use coded values within the riparian areas.

iii. The last step of this data development was to subtract a value of 1 from the riparian land use grid, as the cell value (1) from the riparian cells had been added to the NLCD values.

iv. The land use composition analysis was then run using the tabulate areas script for overlapping polygons.

a) Script: X:\Databases\Water_Quality\Biological_Monitoring\GIS\Scripts

b) Land use within 100 m riparian buffer: X:\Databases\Water_Quality\Biological_Monitoring\Streams\Landuse\Bufflu2001

f) Impervious surface watershed analysis

i. Data sources:

a) The U of M Remote Sensing Lab's impervious surface grid from 2000 (V:\land_cover\um_rsl_2000\impervious\impervious_2000\impervious_statewide_2000_final.img)

b) drainage area polygons clipped to MN boundary

c) intermediate data sets as required by the Python script

d) the analysis uses a Python script that runs the Spatial Analyst zonal statistics process on overlapping polygons. The mean value of the impervious cell value describes the overall impervious percentage for the drainage area.

e) The script can be found here: X:\Databases\Water_Quality\Biological_Monitoring\GIS\Scripts

g) Animal units and feedlots per square kilometer

i. Data sources:

a) DELTA.V_FEEDLOTS_FC (SDE layer) Feedlot points, with total animal units as an attribute.

ii. Drainage area polygons, split into separate feature classes by basin, and clipped to MN boundary.

a) The steps in this analysis have been compiled into an ArcGIS ModelBuilder model.

b) The first five steps of the analysis are performed on separate data sets for each basin, which are merged before the density analysis.

iii. The process first runs an intersect between the drainage areas and the feedlots

iv. The intersected points are then run through the frequency tool, with FieldNum as the frequency field

v. Intersected points are also run through the Summary Statistics tool, with total animal units (TOTAL_AU) as the statistics field and FieldNum as the case field.

a) The resultant tables from the frequency and summary stats operations are then joined on FieldNum.

b) This joined table is subsequently joined to the drainage areas on FieldNum, yielding a table (not a feature class).

c) All of the basin's tables are merged together.

d) Fields for animal units per km2 and feedlots per km2 are added, and then calculated using data developed in preceding steps.

h) Percentage ditched per stream KM and road crossings per stream KM

i. Data sources:

a) Mn/DOT ditch layer

b) Drainage area polygons

c) NHD 1:24k flowlines

d) Intersections as points between NHD24k flowlines and Mn/DOT Roads

ii. The steps in this analysis have been compiled into an ArcGIS ModelBuilder model.

a) The first operations run the percent ditched calculation

i. Initially, the model runs parallel tasks with both the NHD flowlines and the DOT ditches.

ii. Stream lines are intersected with the drainage area polygons.

iii. Line length fields (DitchKM and FlowKM) are added to the intersected line feature classes.

iv. DitchKM and FlowKM are calculated from the feature's length field (in meters) using [Shape_length]/1000.

v. The Summary Statistics tool is run using the sum function on FlowKM and DitchKM with "Name" as the case field (this sums based on the polygon's Name field).

v. The ditch and flowlines feature classes are joined, and the PctDitched field is added, then calculated using PctDitched = ([SUM_DitchKM]/[SUM_FlowKM]) * 100.

b) The following operations run the road crossing per stream calculation.

i. Drainage polygons are intersected with the road/stream intersection points, generating multiple overlapping points in areas with overlapping polygons.

ii. The Frequency tool is run on the output layer using FieldNum as the case field.

c) This is then joined to the output of NHD output of the initial percent ditched operations – which has FlowKM.

d) Intersections per stream km are calculated using [FREQUENCY]/[SUM_FlowKM].

i.) Point sources per km2

i. Data sources:

a) DELTA.WQ_STATIONS (SDE Layer)

b) Drainage area polygons

c) The steps in this analysis have been compiled into an ArcGIS ModelBuilder model.

  i. The first step is to intersect the points with the polygons, joining all attributes. This gives the FieldNum attribute from the polygons to all points that fall within them, generating one copy of each point for each polygon where overlaps occur.

  ii. The Frequency tool is run on the FieldNum field to generate a total number of point sources per FieldNum – point sources per polygon.

  iii. The frequency table is joined to the polygons on the FieldNum field and number of point sources per km^2 is calculated and added to a new field.

Update queries for HDS scoring: Once the raw metric values are calculated using GIS analysis, queries that make the proper transformations need to be run. The ideal situation is to make a copy of the existing database (HDS_2009) and delete the existing records, thus leaving existing columns and queries. Table 6 shows the Access queries used in scoring and also discusses necessary transformations that need to be performed before running the access queries. Final HDS scores are derived by summing all metric scores and corrections.

Table 6: Metric Query Table Containing Access Queries and Necessary Transformations

Query	Metric	Scoring Access Query Code	Transformation?
	% agriculture in watershed	10-((-(0-10)/(1.30544337719725-0)*([TotalRankings2009]!PCTAGASN)))	Arcsine transformed Slight negative drainage area relationship
	% agriculture in 100 meter buffer	IIf(TotalRankings2009!PCTAGRIPARIAN>49.1,-1,0)	Slight negative drainage area relationship
	Amount of agricultural landuse on 3% slope as a percentage of total watershed area of 3% slope	IIf(TotalRankings2009!AGPCT3PCTS>57,-1,0)	
Urban land use	% impervious	IIf([TotalRankings2009]![PCTIMPLOG]>2.768,0,10-((-(0-10)/(2.768-0)*([TotalRankings2009]![PCTIMPLOG]))))	Log Transformed, Slight negative drainage area relationship
	# of road crossings per stream km	IIf([TotalRankings2009]![ROADCROS]>0.395,-1,IIf([TotalRankings2009]![ROADCROS]<0.164,1,0))	Slight negative drainage area relationship

Query	Metric	Scoring Access Query Code	Transformation?
Feedlots	# of animal units per km^2	10-((-(0-10)/(8.14559283069969-0)*(TotalRankings2009!AUDENSLOG)))	Log transformed, No drainage area relationship
	# of feedlots per km^2	IIf([TotalRankings2009]![FEEDDENSITY]>0.235,-1,0)	
Point sources	# of point sources per $sqkm^2$	IIf([TotalRankings2009]![PTSQRT]>0.444,0,10-((-(0-10)/(0.444-0)*([TotalRankings2009]![PTSQRT])))))	Sqrt transformed, Negative drainage area relationship
Watershed riparian	% disturbed riparian habitat in the watershed	10-((-(0-10)/(1.52607123962617-0)*(TotalRankings2009!PCTDISTASN)))	Arcsine transformed No drainage area relationship
Site riparian	DOQ	Scored by visual Observation using GIS and Field Verification	
Site channelization	Channel Condition	Scored by visual observation and Field Verification	
Watershed channelization	% channelized stream per stream km	10-((-(0-10)/(10-0)*(TotalRankings2009!DITCHPCTSQR)))	Sqroot transformed, Negative drainage area relationship

Table 7. Total Rankings Combined_All Scores Field definitions

FIELDNUM- Field Number for sites
WBName- Water body name.
UTM83X- UTM X coordinate
UTM83Y-UTM Y coordinate
AUDENSITY- # of animal units per km^2 raw value
AUDENSLOG- Log of # of animal units per km^2 raw value
Audenscore- # of animal units per km^2 score
PCTINMN- Percent of watershed in Minnesota
PCTAGWATER- % agriculture in watershed
PCTAGASN- Arcsine transformed % agriculture in watershed
Pctagsco- % agriculture in watershed score
PCTDISTRIP-% disturbed riparian habitat in the watershed
PCTDISTASN- Arcsine transformed % disturbed riparian habitat in the watershed
Pctdistripsc- -% disturbed riparian habitat in the watershed score
POINT_DENSITY- # of point sources per $sqkm^2$
PTSQRT- Sqrt transformed # of point sources per $sqkm^2$
Ptscore- # of point sources per $sqkm^2$ score
PCTIMPER- % impervious surface in watershed
PCTIMPLOG- Log Transformed % impervious surface
Pctimpsco- Score of Log Transformed % impervious surface
PCTAGRIPARIAN- % agriculture in 100 meter buffer
Pctagripscor-score of % agriculture in 100 meter buffer
FEEDDENSITY- # of feedlots per km^2
Feedscore- score of # of feedlots per km^2
ROADCROS- # of road crossings per stream km
Roadscore- score of # of road crossings per stream km
AGPCT3PCTS- Amount of agricultural landuse on 3% slope as a percentage of total watershed area of 3% slope
ag3pctscore- score of
DITCHPCT- % channelized stream per stream km based on GIS coverage
DITCHPCTSQR- Sqroot transformed % channelized stream per stream km
DITCHPCTSCORE- score of Sqroot transformed % channelized stream per stream km
ManualDitchScore- Manually scored % channelized stream per stream km based on aerial photo for watersheds
Ditchscorefinal- Final ditch score based on manual and automated scoring
Siteriparian- Site riparian score based on visual Observation using GIS and Field Verification
Sitechannel- Site channelization score by visual observation and Field Verification
TotalRankingsCombined_AllScores Field definitions cont.
Pointsourceprox- point source proximity score correction based on GIS observation
Feedlotprox- feedlot proximity correction score based on GIS observation
Urbanluprox- Urban land use proximity score correction based on GIS observation
TotalScore_Combined- Sum of all scores
TotalScoreWOCorrections_Combined- sum of all primary scores