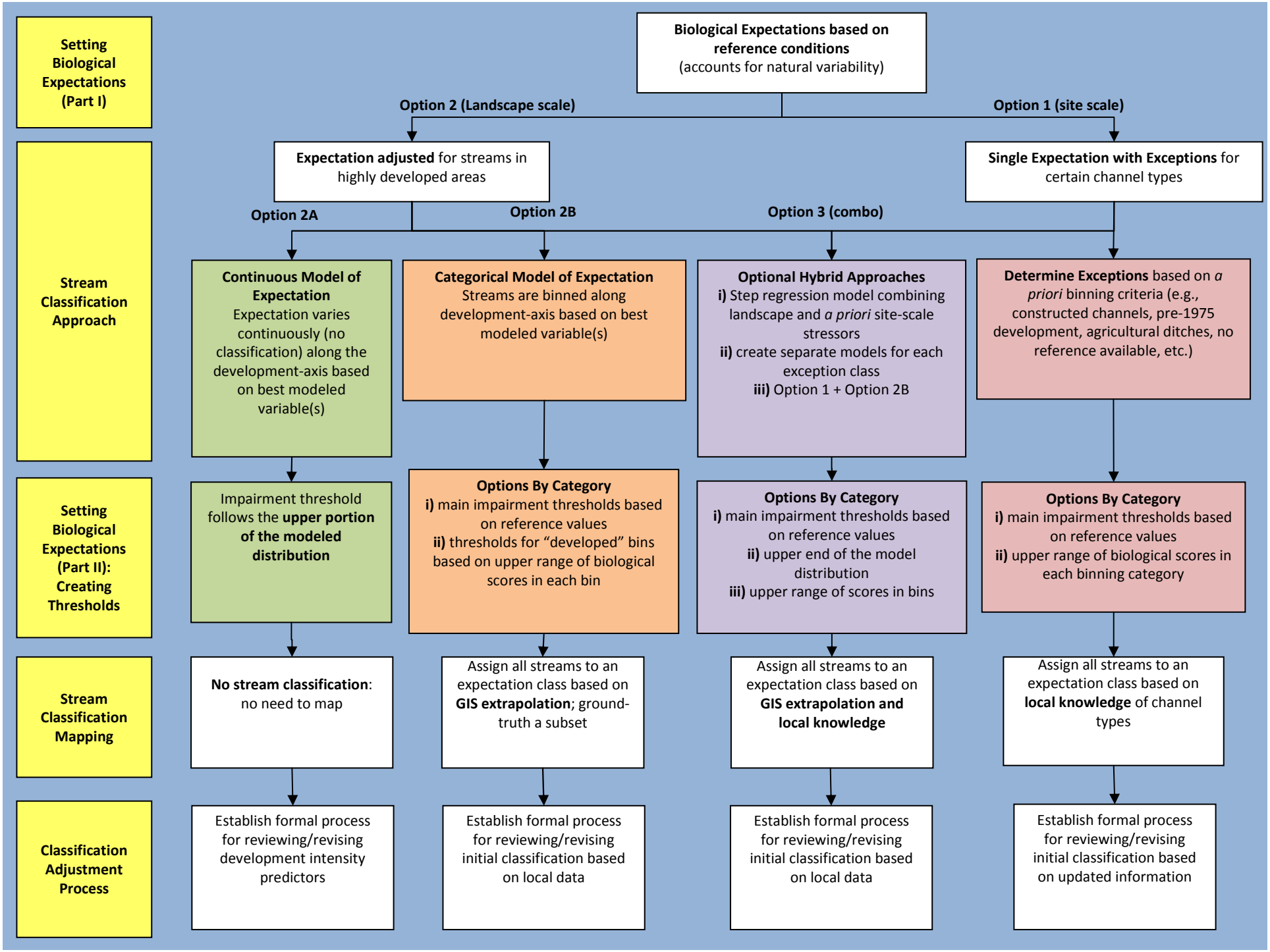
Modeling the alternative approaches to stream classification

Jason May, Larry Brown, and Ian Waite

USGS

4/20/11


Setting Biological Expectations (Part I)
Biological Expectations based on reference conditions (accounts for natural variability)
Option 2 (Landscape scale)
Option 1 (site scale)
Stream Classification Approach
Expectation adjusted for streams in highly developed areas
Single Expectation with Exceptions for certain channel types
Option 2A
Option 2B
Option 3 (combo)
Continuous Model of Expectation
Expectation varies continuously (no classification) along the development-axis based on best modeled variable(s)
Categorical Model of Expectation
Streams are binned along development-axis based on best modeled variable(s)
Optional Hybrid Approaches
i) Step regression model combining landscape and a priori site-scale stressors
ii) create separate models for each exception class
iii) Option 1 + Option 2B
Determine Exceptions based on a priori binning criteria (e.g., constructed channels, pre-1975 development, agricultural ditches, no reference available, etc.)
Setting Biological Expectations (Part II): Creating Thresholds
Impairment threshold follows the upper portion of the modeled distribution
Options By Category
i) main impairment thresholds based on reference values
ii) thresholds for "developed" bins based on upper range of biological scores in each bin
Options By Category
i) main impairment thresholds based on reference values
ii) upper end of the model distribution
iii) upper range of scores in bins
Options By Category
i) main impairment thresholds based on reference values
ii) upper range of biological scores in each binning category
Stream Classification Mapping
No stream classification: no need to map
Assign all streams to an expectation class based on GIS extrapolation; ground-truth a subset
Assign all streams to an expectation class based on GIS extrapolation and local knowledge
Assign all streams to an expectation class based on local knowledge of channel types
Classification Adjustment Process
Establish formal process for reviewing/revising development intensity predictors
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on updated information

Approach:

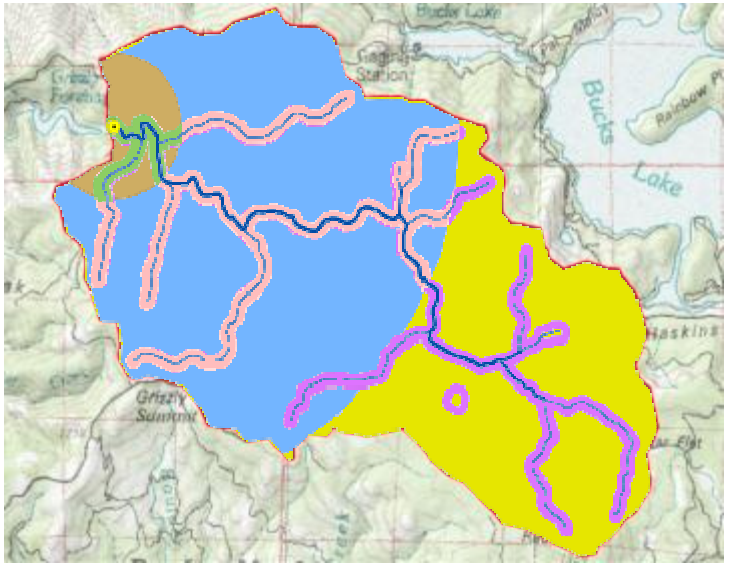
- Focus on SMC-Xeric region
  - Based on strong gradient
  - data density (allows for subsetting)
- Stepwise multiple linear modeling
- Response variable(s):
  - So. Cal. IBI, O/E, EPT richness
- "Best model" determined from
  - Adjusted R^2
  - AIC: Akaike Information Criterion
    - Models with AIC within 2 points are equally plausible

Data Reduction

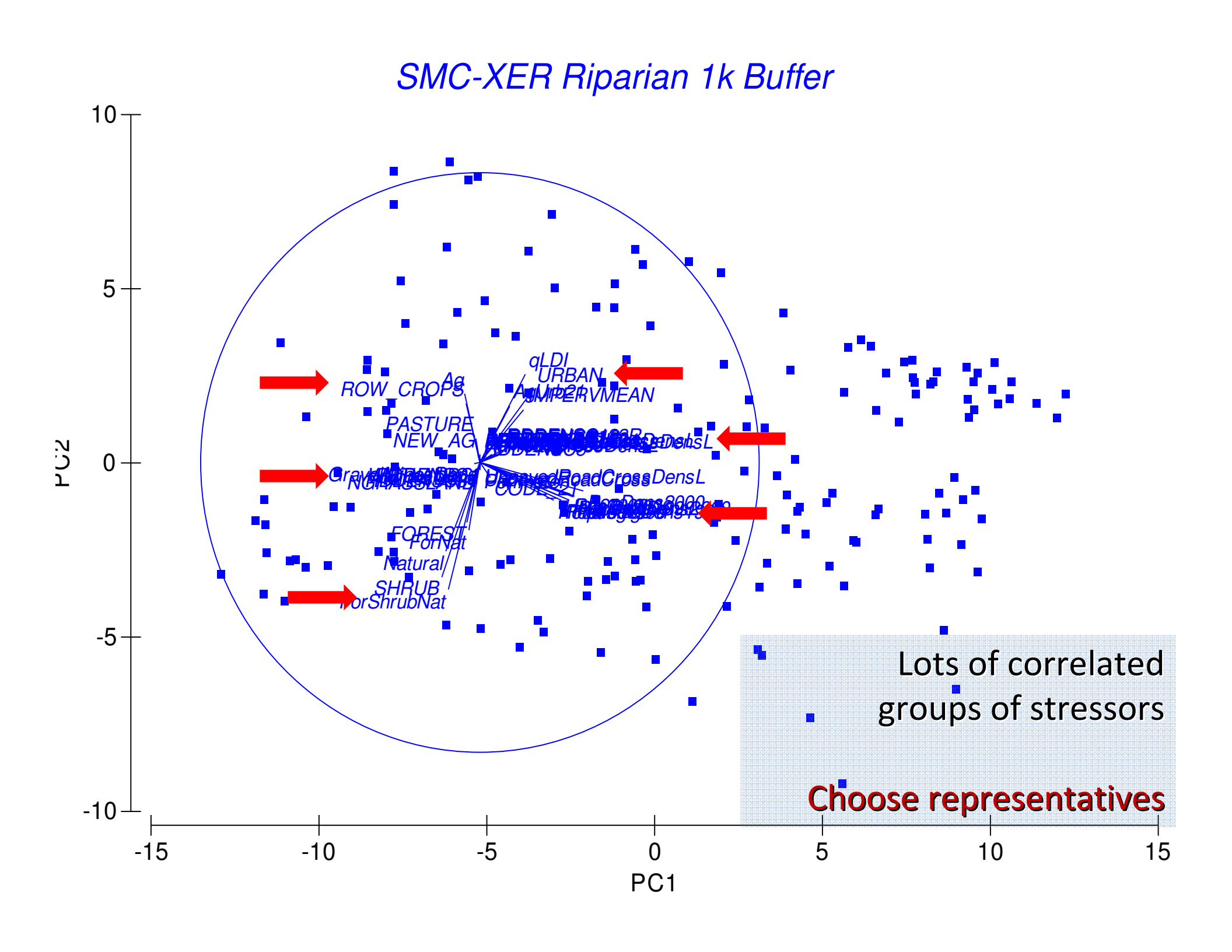
- Methods
  - Correlation, PCA, Random Forest
- GIS scale selection
- Winnowing process
  - Original 167 variables reduced to 44 (~25% left)
    - Reach
    - riparian 1k
    - watershed
  - Secondary data reduction with Random Forest on 44 variable set

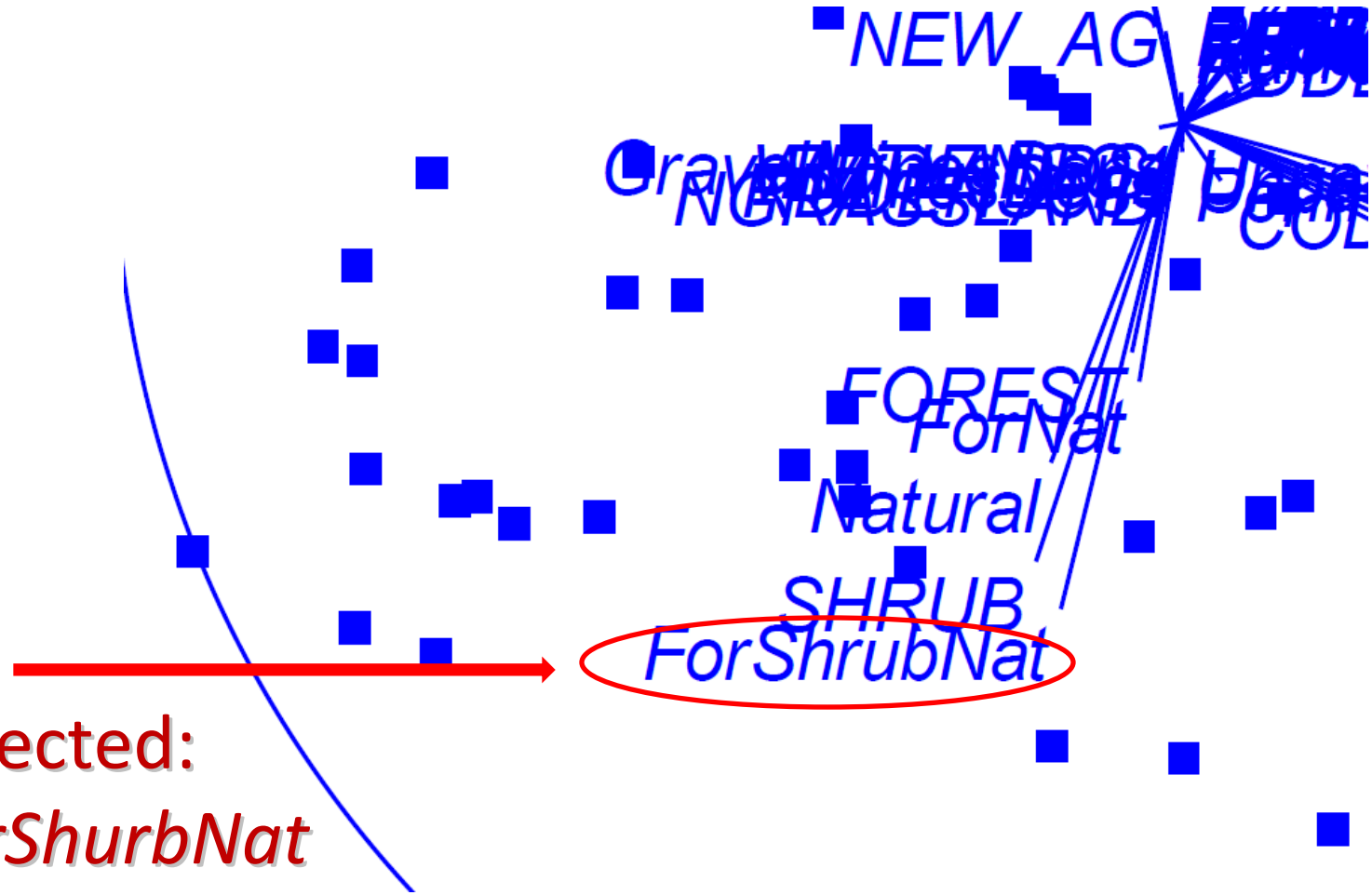
Deciding which GIS scales to model

- Initial decision to use Watershed scale info (WS)
- Want the scale most different from WS
  - Used Primer routine-RELATE to test for similarity between WS and other scales of GIS information



Resemblance Matrix A	Resemblance Matrix B	Rho	scale
WS	1k	0.323	1k radial clip
WS	5k	0.541	5k radial clip
WS	r1k	**0.270**	1k riparian buffer
WS	r5k	0.461	5k riparian buffer
WS	rWS	0.852	WS riparian buffer


SMC-XER Riparian 1k Buffer
10
5
0
-5
-10
PC2
-15
-10
-5
0
5
10
15
PC1
qLDI
URBAN
Ag
ROW_CROPS
PASTURE
NEW_AG
FOREST
ForNat
Natural
SHRUB
NonShrubNat
Lots of correlated
groups of stressors
Choose representatives



Selected:
ForShurbNat

Reach level 'Surrogates'

Model Variable	Surrogates	
COND	*ALK*	*CL*
NTL	*CL*	
Elevation	*XSLOPE*	
P_SAFN	*Log Rel. Bed Stab.*	
W1_HALL	*general disturbance*	
Precipitation(PPT)	*PRISM-general climate*	
Temperature (TEMP)	*PRISM-general climate*	

SMC-XER:

Correlation of Biotic Indicators (Spearman's)

	O_E_0	O_E_05	IBI_Score	EPT
O_E_0	1			
O_E_05	0.85	1		
IBI_Score	0.79	0.63	1	
EPT	0.81	0.69	0.72	1

- Responses are relatively highly correlated
- Perform differently in the models

SMC-XER- Variables for Modeling Effort

Bio-Indicators	Reach	Riparian 1k	Watershed
O_E_0	COND	r1k_PC1	ws_PC1
O_E_0.5	NTL	r1k_Ag	ws_Ag
So. Cal. IBI Score	Elevation	r1k_CODE_21	ws_CODE_21
EPT Richness	P_SAFN	r1k_URBAN	ws_URBAN
	W1_HALL	r1k_qLDI*	ws_qLDI*
	PPT	r1k_AgUrb21	ws_AgUrb21
	TEMP	r1k_ForShrubNat	ws_CanalPipe24kPer
		r1k_HousingDens2000	ws_DamDensArea
		r1k_IMPERVMEAN	ws_ForShrubNat
		r1k_PASTURE	ws_GravelMinesDens
		r1k_PavedRoadCross	ws_GRAZING
		r1k_PopDens2000	ws_HousingDens2000
		r1k_RDDENSC1234	ws_IMPERVMEAN
		r1k_RoadRailroadCross	ws_LengthNoPipe24k
		r1k_ROW_CROPS	ws_MinesDens
		r1k_WETLANDS	ws_PASTURE
			ws_Pipe24k
			ws_PopDens2000
			ws_RDDENSC1234
			ws_ROW_CROPS
			ws_WETLANDS

**** qLDI = ∑[(Code21 *2)+(AG*6)+(URBAN*8.5)]***
-Coefficient values modified from Brown and Vivas (2005)

Secondary data reduction phase

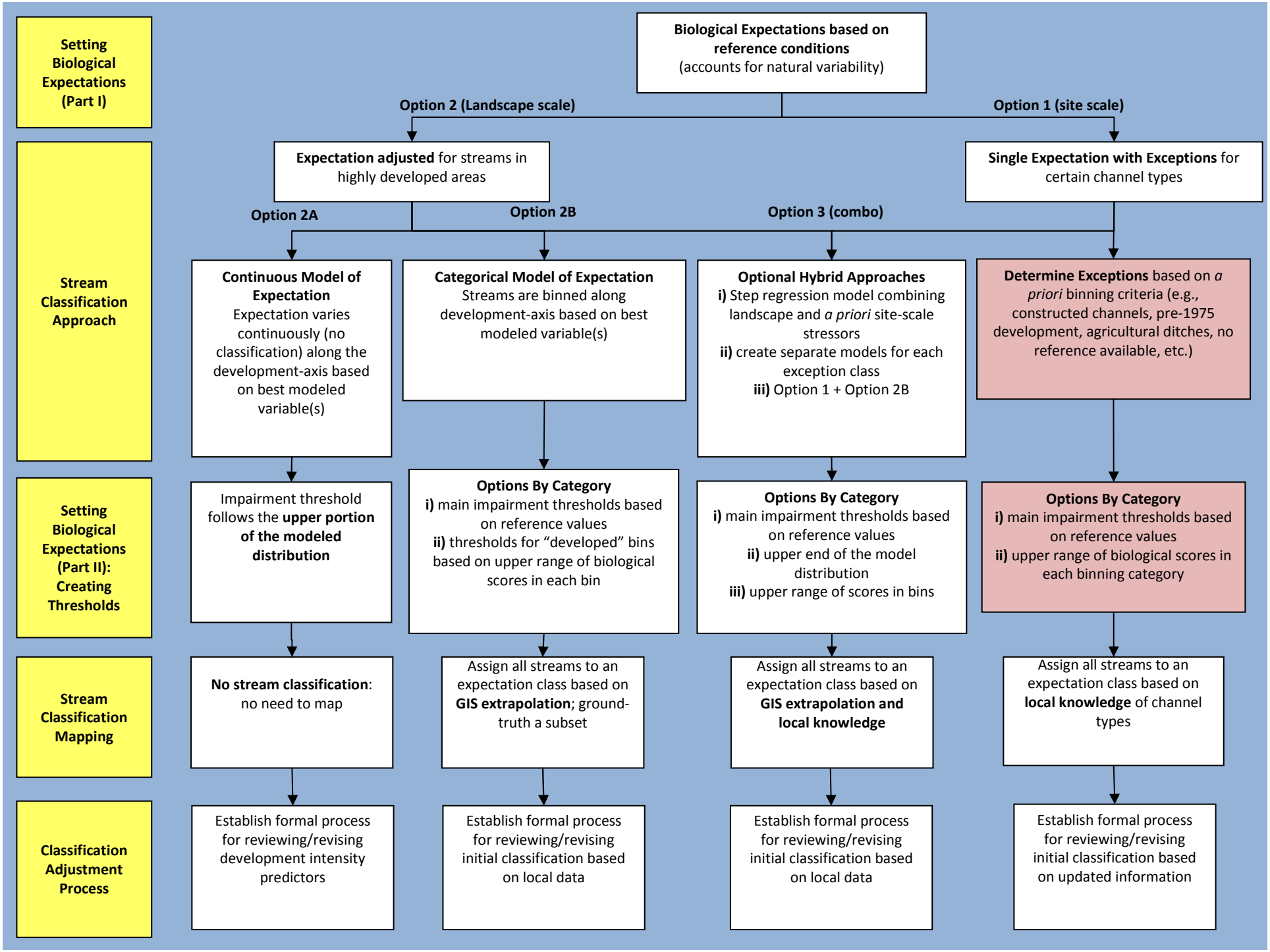
IBI, OE, EPT,& OE_0.5

Random Forest

Used the top 20 variables to start the modeling process

So. Cal. IBI Score **(56% var exp)**		**O/E (46% var exp)**	
r1k_AgUrb21	17.84	P_SAFN	19.23
ws_IMPERVMEAN	17.45	r1k_URBAN	17.83
ws_RDDENSC1234	17.12	r1k_AgUrb21	17.45
ws_URBAN	17.04	r1k_qLDI	16.59
r1k_qLDI	16.78	r1k_IMPERVMEAN	16.38
ws_AgUrb21	16.75	ws_URBAN	15.54
ws_qLDI	16.35	ws_qLDI	14.72
ws_HousingDens2000	14.98	ws_IMPERVMEAN	14.69
r1k_IMPERVMEAN	14.91	NTL	13.33
PPT	14.18	ws_ForShrubNat	12.99
ws_ForShrubNat	13.88	ws_AgUrb21	12.35
ws_PopDens2000	13.49	ws_RDDENSC1234	11.57
ws_CODE_21	13.45	COND	10.72
COND	12.84	TEMP	10.64
r1k_URBAN	12.66	PPT	9.93
Elevation	11.75	r1k_ForShrubNat	9.89
W1_HALL	11.48	r1k_PC1	9.70
r1k_ForShrubNat	9.96	ws_HousingDens2000	9.61
r1k_PopDens2000	9.82	ws_PopDens2000	8.91
NTL	9.19	ws_DamDensArea	8.88
r1k_RDDENSC1234	8.61	ws_GRAZING	8.50
r1k_HousingDens2000	8.32	r1k_HousingDens2000	8.14
r1k_PC1	8.18	ws_LengthNoPipe24k	7.87

MODELING THE ALTERNATIVE APPROACHES

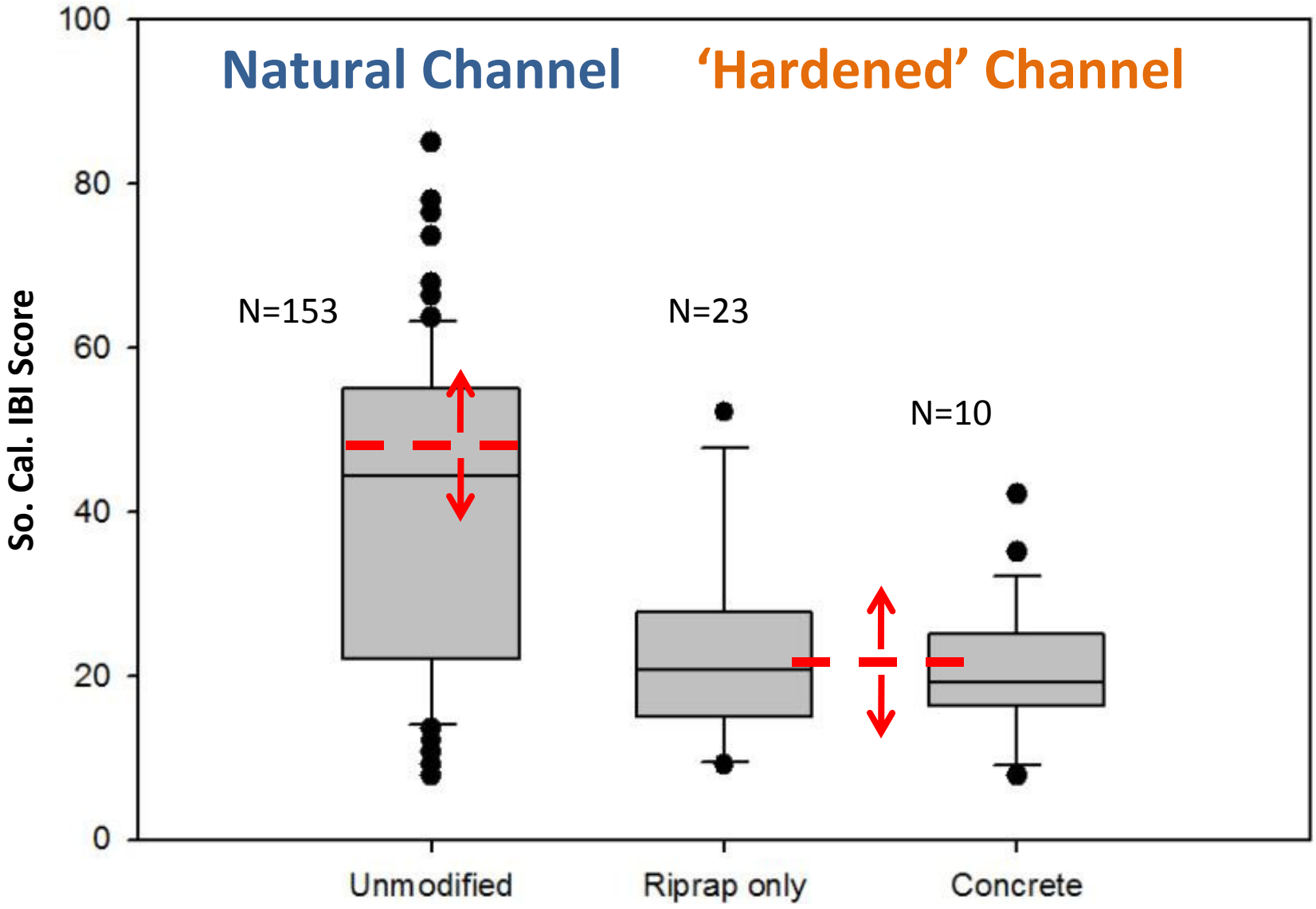

Setting Biological Expectations (Part I)
Biological Expectations based on reference conditions (accounts for natural variability)
Option 2 (Landscape scale)
Option 1 (site scale)
Stream Classification Approach
Expectation adjusted for streams in highly developed areas
Single Expectation with Exceptions for certain channel types
Option 2A
Option 2B
Option 3 (combo)
Continuous Model of Expectation
Expectation varies continuously (no classification) along the development-axis based on best modeled variable(s)
Categorical Model of Expectation
Streams are binned along development-axis based on best modeled variable(s)
Optional Hybrid Approaches
i) Step regression model combining landscape and a priori site-scale stressors
ii) create separate models for each exception class
iii) Option 1 + Option 2B
Determine Exceptions based on a priori binning criteria (e.g., constructed channels, pre-1975 development, agricultural ditches, no reference available, etc.)
Setting Biological Expectations (Part II): Creating Thresholds
Impairment threshold follows the upper portion of the modeled distribution
Options By Category
i) main impairment thresholds based on reference values
ii) thresholds for "developed" bins based on upper range of biological scores in each bin
Options By Category
i) main impairment thresholds based on reference values
ii) upper end of the model distribution
iii) upper range of scores in bins
Options By Category
i) main impairment thresholds based on reference values
ii) upper range of biological scores in each binning category
Stream Classification Mapping
No stream classification: no need to map
Assign all streams to an expectation class based on GIS extrapolation; ground-truth a subset
Assign all streams to an expectation class based on GIS extrapolation and local knowledge
Assign all streams to an expectation class based on local knowledge of channel types
Classification Adjustment Process
Establish formal process for reviewing/revising development intensity predictors
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on updated information

Option-1 Site scale

- Single expectation with exceptions for certain channel types

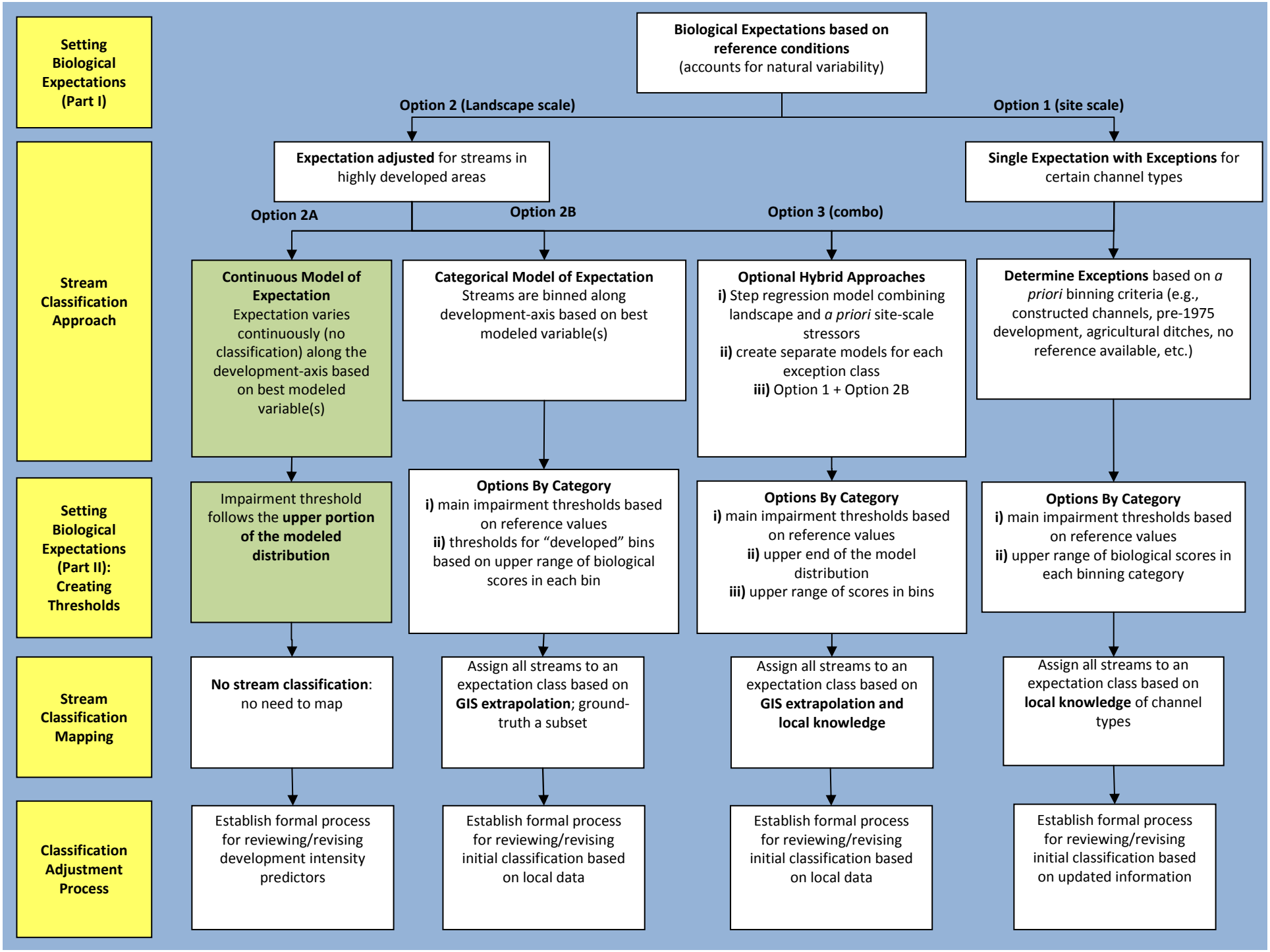
- Exceptions are based on *a priori ,* likely non-technical decisions
  - Concrete lined, Ag ditches, etc.

Example Site Scale Approach using *a priori* Stream Classification from So. Cal. (SMC)



Lesson learned from option 1

- Simple: If we the had data in hand we could do it tomorrow
- We do not really have this data
  - Categories like this are not available in data or GIS coverages
  - This “estimate” is used later in the modeling because it is the best we have
- Where is the threshold?


Setting Biological Expectations (Part I)
Biological Expectations based on reference conditions (accounts for natural variability)
Option 2 (Landscape scale)
Option 1 (site scale)
Stream Classification Approach
Expectation adjusted for streams in highly developed areas
Single Expectation with Exceptions for certain channel types
Option 2A
Option 2B
Option 3 (combo)
Continuous Model of Expectation
Expectation varies continuously (no classification) along the development-axis based on best modeled variable(s)
Categorical Model of Expectation
Streams are binned along development-axis based on best modeled variable(s)
Optional Hybrid Approaches
i) Step regression model combining landscape and a priori site-scale stressors
ii) create separate models for each exception class
iii) Option 1 + Option 2B
Determine Exceptions based on a priori binning criteria (e.g., constructed channels, pre-1975 development, agricultural ditches, no reference available, etc.)
Setting Biological Expectations (Part II): Creating Thresholds
Impairment threshold follows the upper portion of the modeled distribution
Options By Category
i) main impairment thresholds based on reference values
ii) thresholds for "developed" bins based on upper range of biological scores in each bin
Options By Category
i) main impairment thresholds based on reference values
ii) upper end of the model distribution
iii) upper range of scores in bins
Options By Category
i) main impairment thresholds based on reference values
ii) upper range of biological scores in each binning category
Stream Classification Mapping
No stream classification: no need to map
Assign all streams to an expectation class based on GIS extrapolation; ground-truth a subset
Assign all streams to an expectation class based on GIS extrapolation and local knowledge
Assign all streams to an expectation class based on local knowledge of channel types
Classification Adjustment Process
Establish formal process for reviewing/revising development intensity predictors
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on updated information

Continuous Model of Expectation 2A

- Expectation varies continuously (no classification) along the development-axis based on best modeled variable(s)

- Impairment threshold follows the upper portion of the modeled distribution

MODELING OUTPUTS FOR BEST MODELS USING ALL VARIABLES

So. Cal. IBI Score &
EPT Richness

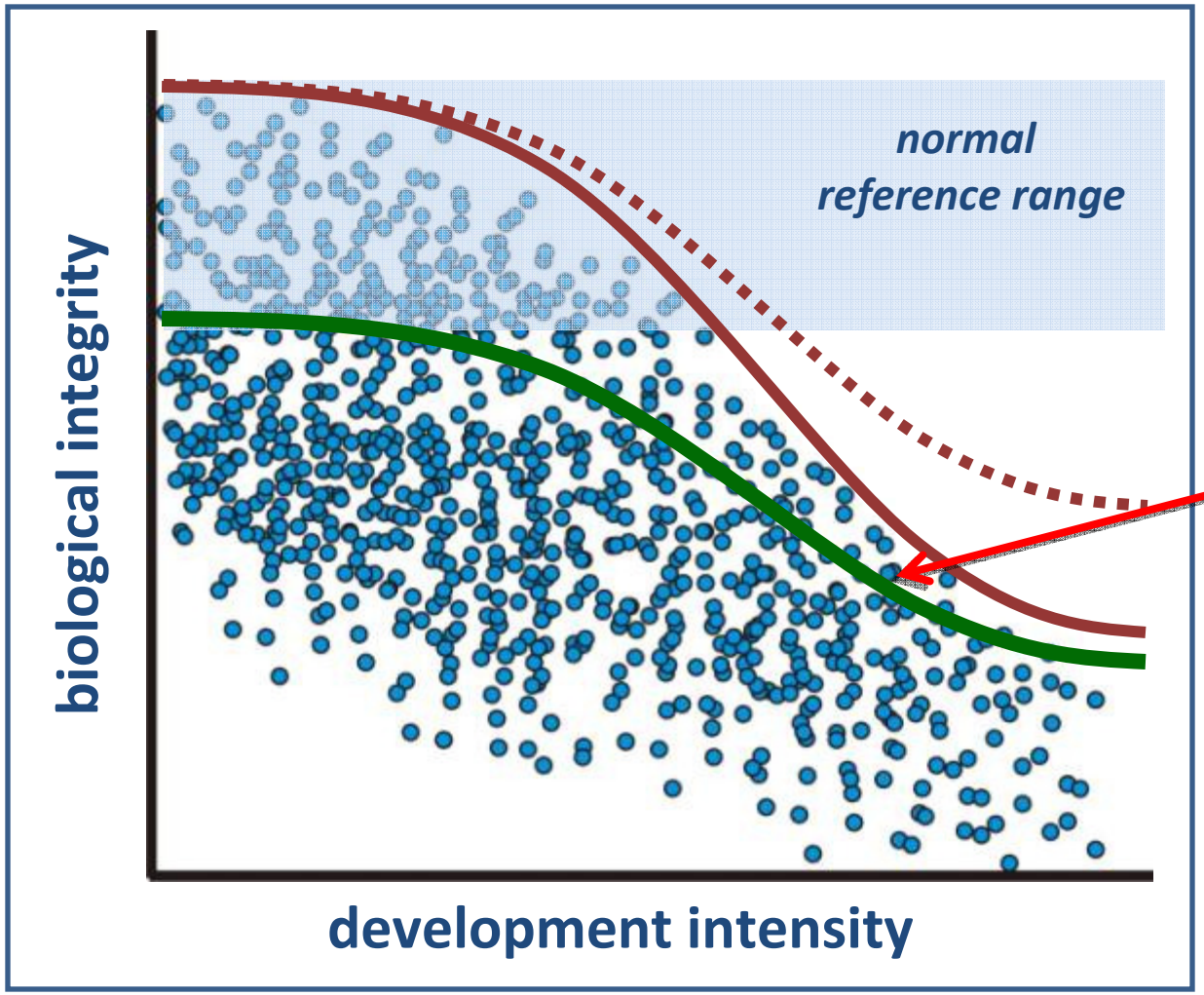
SMC-XER EPT Richness

Variable	adj-R2	AIC	reach	1k riparian	watershed
r1k_qLDI	0.3729	370.4		x	
r1k_IMPERVMEAN	0.3684	371.73		x	
r1k_URBAN	0.3628	373.38		x	
r1k_AgUrb21	0.3348	381.35		x	
ws_URBAN	0.2999	390.88			x
r1k_qLDI + ws_URBAN + NTL	**0.426**	**355.91**	**x**	**x**	**x**
ws_qLDI + r1k_IMPERVMEAN + ws_RDDENSC1234	0.4249	356.25		x	x
r1k_qLDI + r1k_URBAN + ws_qLDI	0.4226	356.99		x	x
r1k_qLDI + ws_URBAN + Elevation	0.4221	357.17	x	x	x
r1k_qLDI + ws_URBAN + ws_IMPERVMEAN	0.4218	357.27		x	x
r1k_IMPERVMEAN + ws_qLDI + NTL + ws_RDDENSC1234	**0.4346**	**354.07**	**x**	**x**	**x**
r1k_qLDI + ws_URBAN + NTL + Elevation	0.4343	354.18	x	x	x
r1k_qLDI + r1k_URBAN + ws_qLDI + NTL	0.4327	354.69	x	x	x
r1k_qLDI + r1k_URBAN + ws_URBAN + NTL	0.4325	354.76	x	x	x
r1k_qLDI + ws_URBAN + ws_IMPERVMEAN + NTL	0.4318	354.99	x	x	x

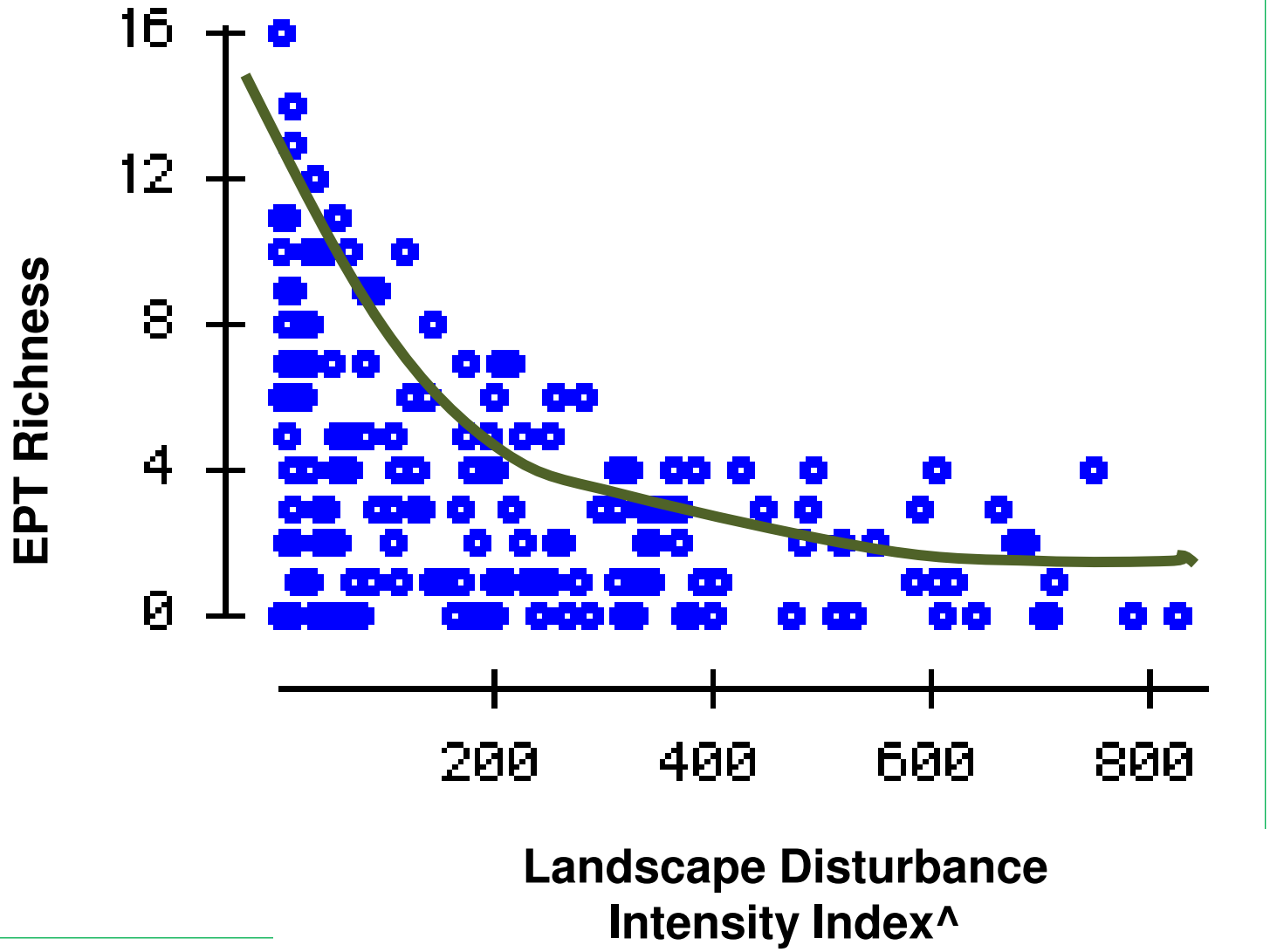
SMC-XER SO.CAL. IBI Score

Variable	adj-R2	AIC
ws_URBAN	0.4635	962.82
ws_qLDI	0.4581	964.68
ws_IMPERVMEAN	0.4266	975.21
r1k_qLDI	0.4129	979.58
ws_AgUrb21	0.4106	980.33
r1k_AgUrb21+ws_IMPERVMEAN+ws_URBAN	0.5519	931.31
ws_IMPERVMEAN+ws_URBAN+r1k_qLDI	0.5516	931.43
r1k_AgUrb21+ws_URBAN+PPT	0.5506	931.84
ws_URBAN+r1k_qLDI+COND	0.5502	932.03
ws_URBAN+ r1k_qLDI+PPT	0.5497	932.23
r1k_AgUrb21+ws_IMPERVMEAN+ws_URBAN+PPT	**0.566**	**926.35**
r1k_AgUrb21+ws_IMPERVMEAN+ws_URBAN+Elevation	0.5652	926.67
ws_IMPERVMEAN+ws_URBAN+r1k_qLDI+PPT	0.5644	927.04
ws_IMPERVMEAN+ws_URBAN+r1k_qLDI+Elevation	0.5631	927.57
ws_IMPERVMEAN+ws_URBAN+r1k_qLDI+COND	0.5617	928.17

Continuous Threshold



Real data example with one variable

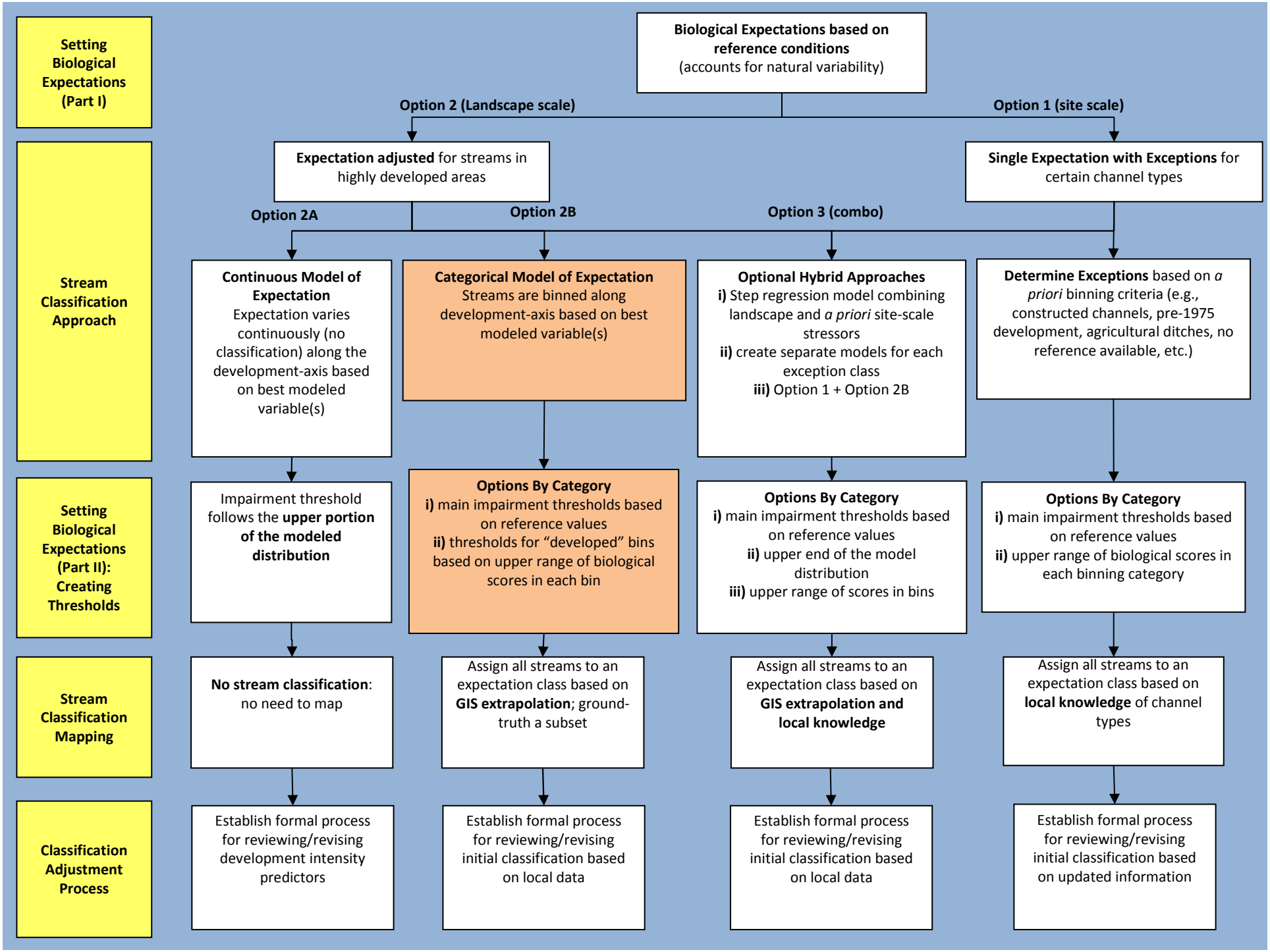


*qLDI = ∑[(Code21 *2)+(AG*6)+(URBAN*8.5)]*

- Coefficients values modified from Brown and Vivas 2005

Lessons learned from Option 2A

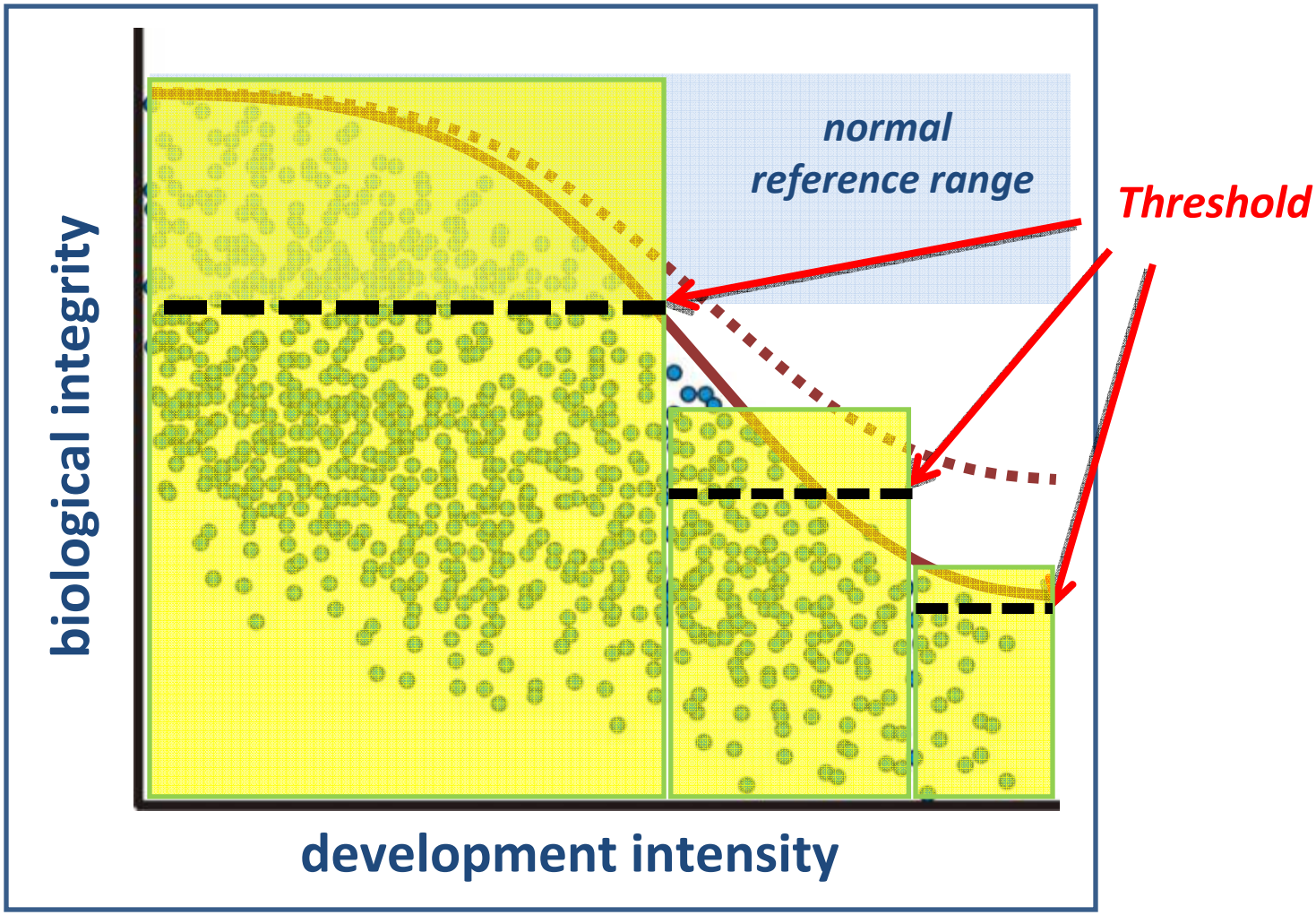
- Modeling continuous data is feasible with the data available
- Model selected the best fit parameters
  - Urban and Ag land uses, imperviousness
- Better to include more or fewer variables in the models?
  - Tradeoffs between complexity and precision
- Identifying thresholds using the upper end of the biological distribution appears achievable


Setting Biological Expectations (Part I)
Biological Expectations based on reference conditions (accounts for natural variability)
Option 2 (Landscape scale)
Option 1 (site scale)
Stream Classification Approach
Expectation adjusted for streams in highly developed areas
Single Expectation with Exceptions for certain channel types
Option 2A
Option 2B
Option 3 (combo)
Continuous Model of Expectation
Expectation varies continuously (no classification) along the development-axis based on best modeled variable(s)
Categorical Model of Expectation
Streams are binned along development-axis based on best modeled variable(s)
Optional Hybrid Approaches
i) Step regression model combining landscape and a priori site-scale stressors
ii) create separate models for each exception class
iii) Option 1 + Option 2B
Determine Exceptions based on a priori binning criteria (e.g., constructed channels, pre-1975 development, agricultural ditches, no reference available, etc.)
Setting Biological Expectations (Part II): Creating Thresholds
Impairment threshold follows the upper portion of the modeled distribution
Options By Category
i) main impairment thresholds based on reference values
ii) thresholds for "developed" bins based on upper range of biological scores in each bin
Options By Category
i) main impairment thresholds based on reference values
ii) upper end of the model distribution
iii) upper range of scores in bins
Options By Category
i) main impairment thresholds based on reference values
ii) upper range of biological scores in each binning category
Stream Classification Mapping
No stream classification: no need to map
Assign all streams to an expectation class based on GIS extrapolation; ground-truth a subset
Assign all streams to an expectation class based on GIS extrapolation and local knowledge
Assign all streams to an expectation class based on local knowledge of channel types
Classification Adjustment Process
Establish formal process for reviewing/revising development intensity predictors
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on updated information

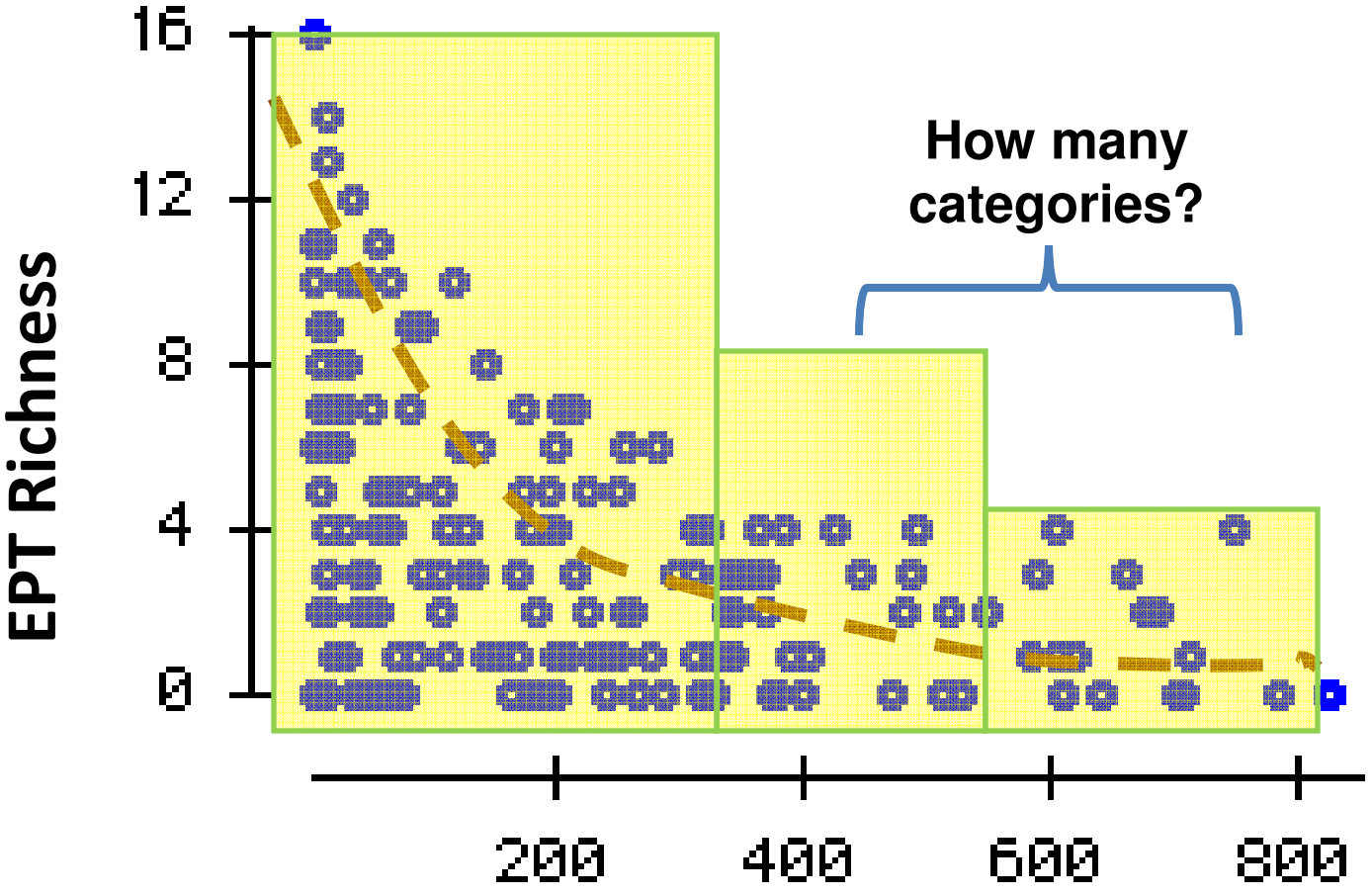
Categorical Model of Expectation-2B

- Streams are binned along development-axis based on best modeled variable(s)

- Options By Category
  - i) main impairment thresholds based on reference values
  - ii) thresholds for “developed” bins based on upper range of biological scores in each bin


normal
reference range
Threshold
biological integrity
development intensity

Real data example with one variable

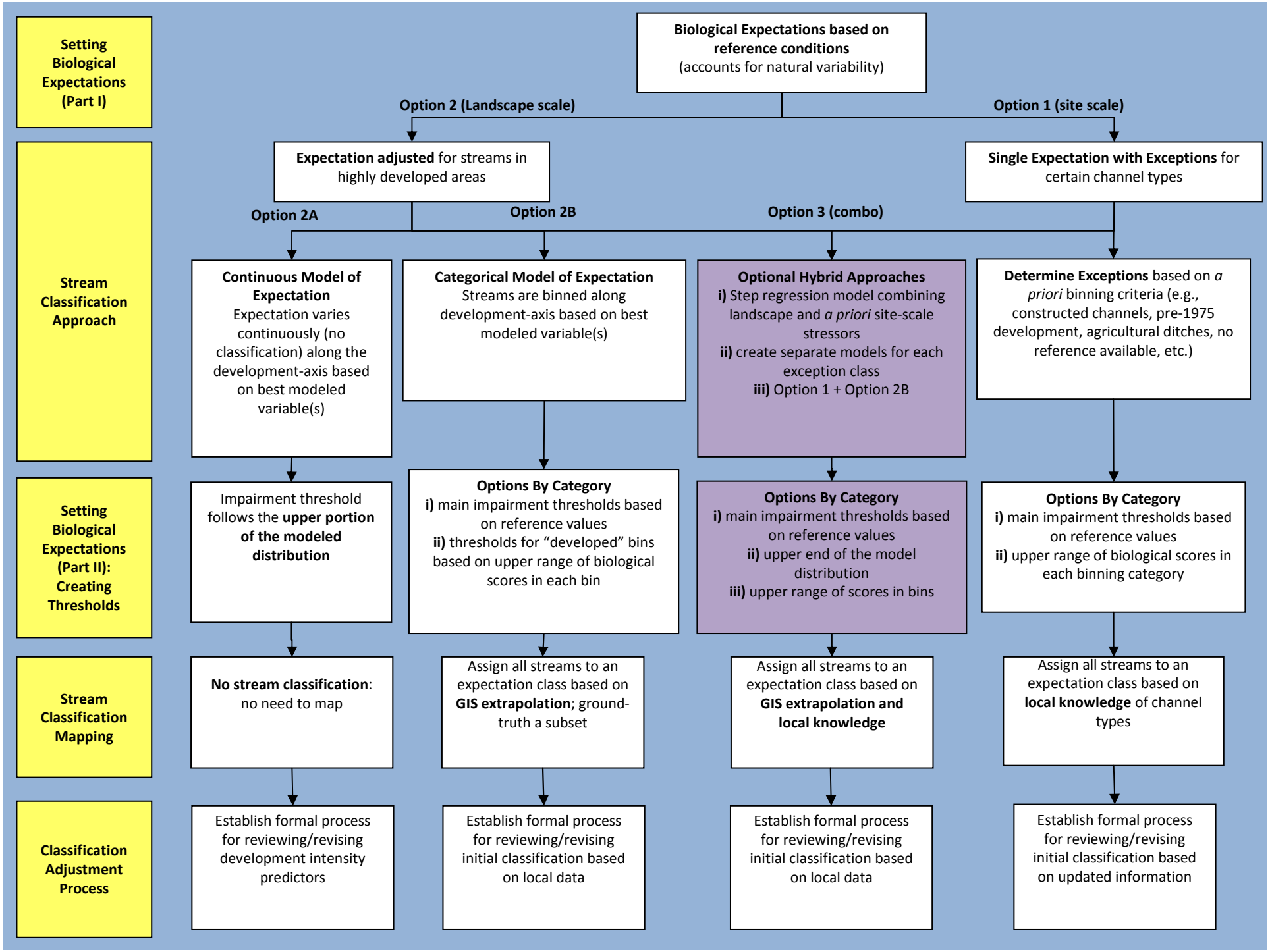


*^qLDI = ∑[(Code21 *2)+(AG*6)+(URBAN*8.5)]*

- Coefficients values modified from Brown and Vivas 2005

Lessons Learned from Option 2B

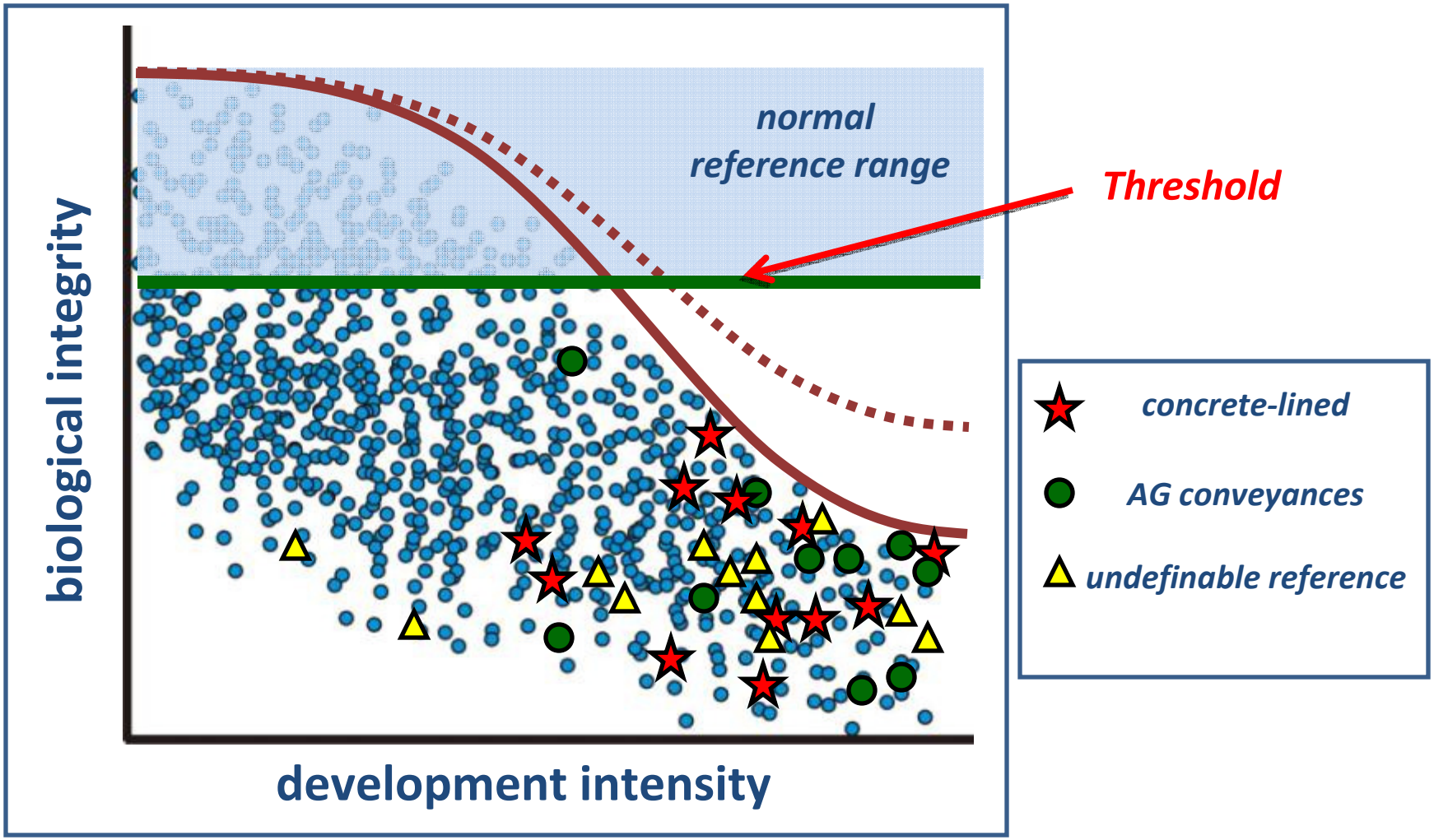
- Similar lessons from 2A
  - Expectations are all model driven
- Not clear where to divide bins
  - Likely requires consensus based exercise
- Criteria for selecting thresholds are less clear than 2A


Setting Biological Expectations (Part I)
Biological Expectations based on reference conditions (accounts for natural variability)
Option 2 (Landscape scale)
Option 1 (site scale)
Stream Classification Approach
Expectation adjusted for streams in highly developed areas
Single Expectation with Exceptions for certain channel types
Option 2A
Option 2B
Option 3 (combo)
Continuous Model of Expectation
Expectation varies continuously (no classification) along the development-axis based on best modeled variable(s)
Categorical Model of Expectation
Streams are binned along development-axis based on best modeled variable(s)
Optional Hybrid Approaches
i) Step regression model combining landscape and a priori site-scale stressors
ii) create separate models for each exception class
iii) Option 1 + Option 2B
Determine Exceptions based on a priori binning criteria (e.g., constructed channels, pre-1975 development, agricultural ditches, no reference available, etc.)
Setting Biological Expectations (Part II): Creating Thresholds
Impairment threshold follows the upper portion of the modeled distribution
Options By Category
i) main impairment thresholds based on reference values
ii) thresholds for "developed" bins based on upper range of biological scores in each bin
Options By Category
i) main impairment thresholds based on reference values
ii) upper end of the model distribution
iii) upper range of scores in bins
Options By Category
i) main impairment thresholds based on reference values
ii) upper range of biological scores in each binning category
Stream Classification Mapping
No stream classification: no need to map
Assign all streams to an expectation class based on GIS extrapolation; ground-truth a subset
Assign all streams to an expectation class based on GIS extrapolation and local knowledge
Assign all streams to an expectation class based on local knowledge of channel types
Classification Adjustment Process
Establish formal process for reviewing/revising development intensity predictors
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on local data
Establish formal process for reviewing/revising initial classification based on updated information

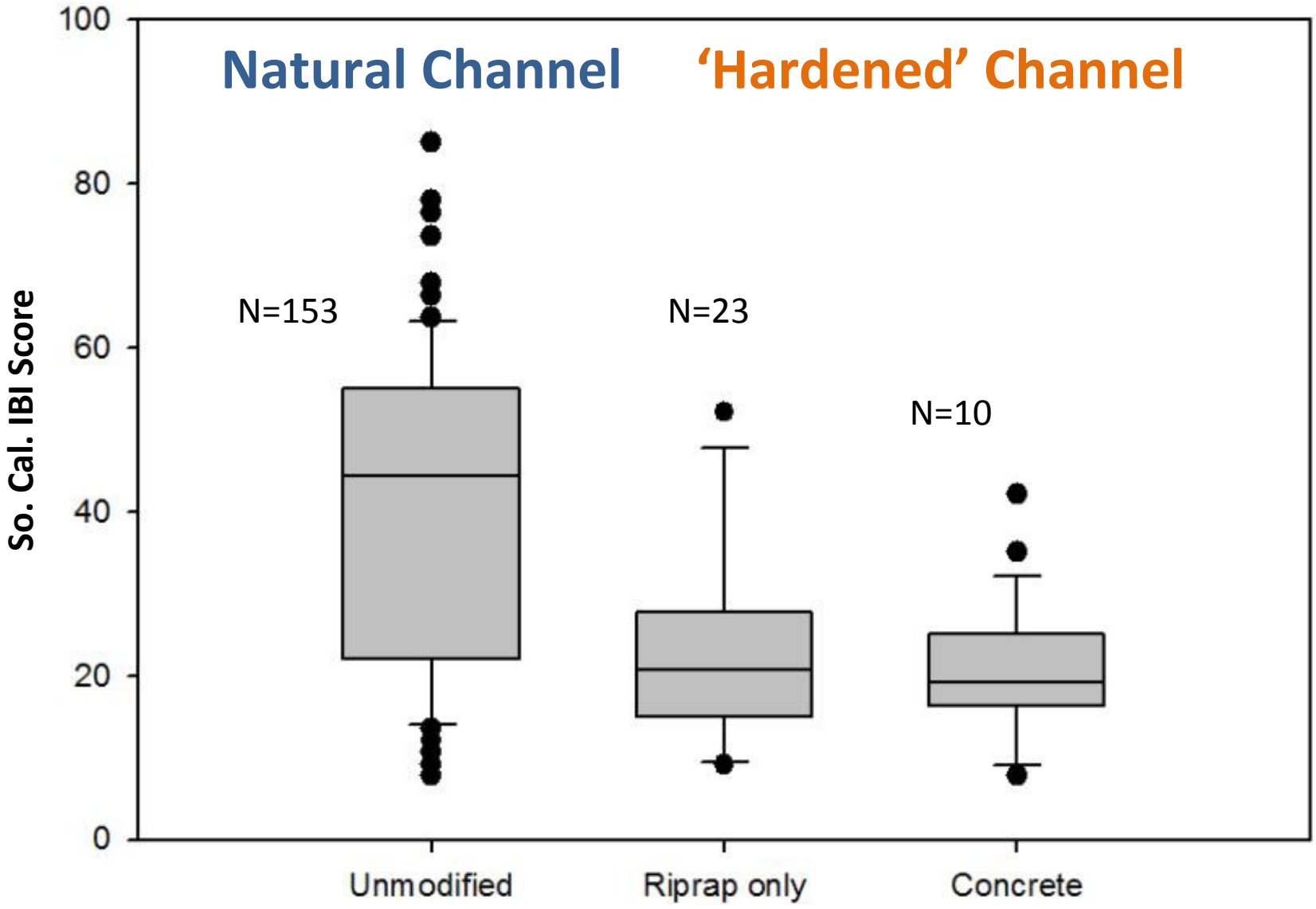
Hybrid Approaches-option 3 (combo) Requires Choices

- **i)** Create separate models for each exception class
- **ii)** Option 1 + Option 2B
- **iii)** 'Step' regression model combining landscape and *a priori* site-scale stressors
  - Initial model based on landscape variables
  - Then force reach scale habitat variables

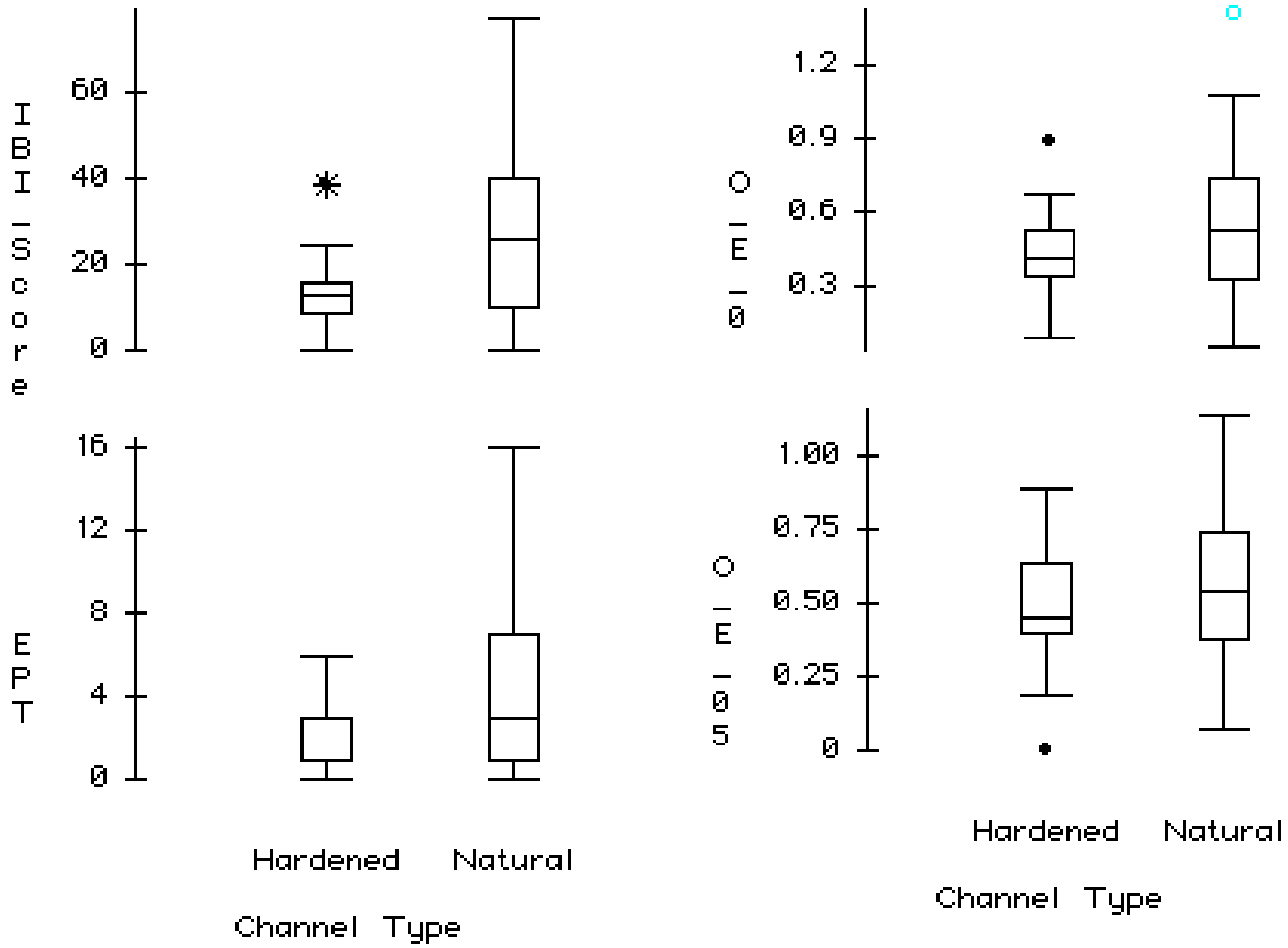
Single Standard + Exceptions



Example Site Scale Approach using *a priori* Stream Classification from So. Cal. (SMC)

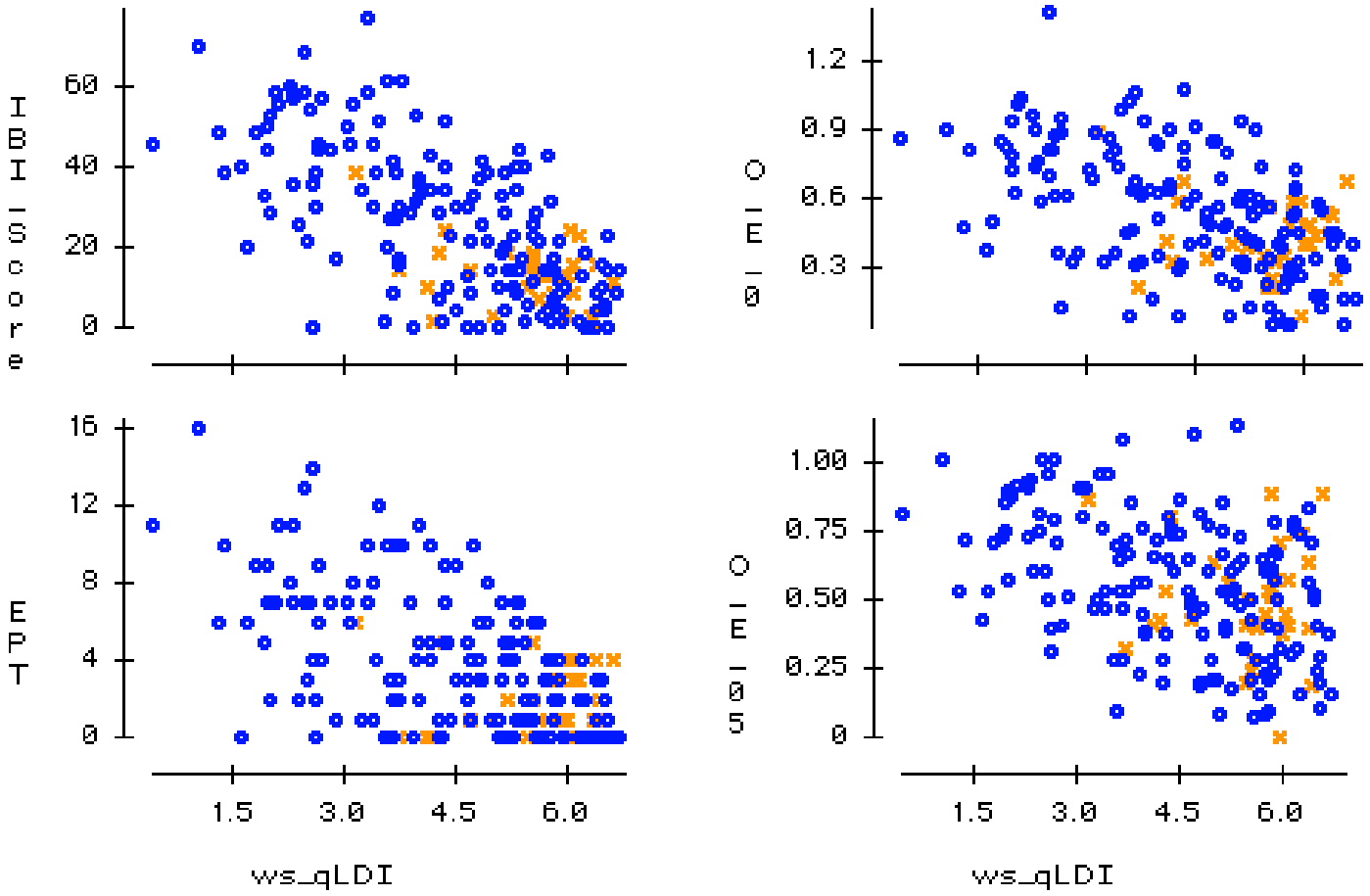


Bioindicators by Channel Type

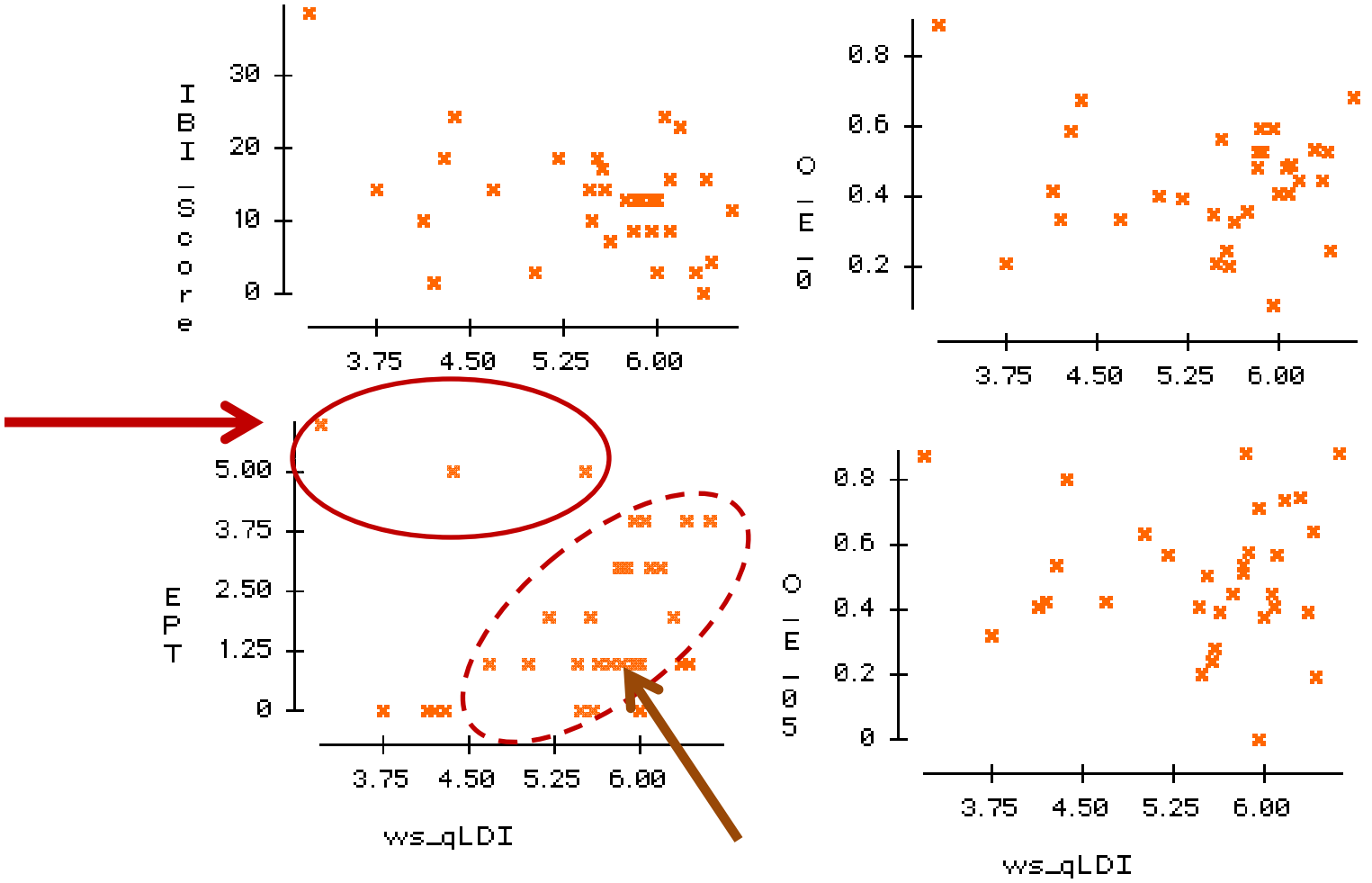


Bioindicator vs qLDI(log transformed)

(x-Hardened, o-Natural)



Hardened Channels



Separate models for Hardened Streams

Variable	adj-R2	AIC
r1k_URBAN	0.1688	131.33
r1k_qLDI	0.1259	132.99
r1k_IMPERVMEAN	0.1083	133.65
PPT + r1k_URBAN	0.2995	126.6
ws_DamDensArea + r1k_URBAN	0.2791	127.55
PPT + r1k_qLDI	0.2558	128.6
r1k_URBAN+ws_AgUrb21+ws_DamDensArea	**0.4717**	**118.18**
r1k_IMPERVMEAN+ws_AgUrb21+ws_DamDensArea	0.4272	120.84
r1k_qLDI+ws_RDDENSC1234+ws_DamDensArea	0.4196	121.28
r1k_URBAN+ws_AgUrb21+ws_DamDensArea+r1k_PavedRoadCross	**0.5109**	**116.47**
ws_PopDens2000+r1k_URBAN+ws_AgUrb21+ws_DamDensArea	**0.4821**	**118.36**
r1k_URBAN+ws_HousingDens2000+ws_AgUrb21+ws_DamDensArea	0.4788	118.57

SOCAL IBI Model w/Hardened Channels removed

Variable	adj-R2	AIC
ws_URBAN	0.4414	810.04
ws_qLDI	0.437	811.24
ws_IMPERVMEAN	0.4058	819.5
ws_URBAN+r1k_AgUrb21	0.5137	789.81
ws_URBAN+r1k_qLDI	0.5116	790.47
r1k_AgUrb21+ws_IMPERVMEAN	0.4936	796
ws_URBAN+r1k_AgUrb21+ws_IMPERVMEAN	0.5276	786.37
ws_URBAN+ws_IMPERVMEAN+r1k_qLDI	0.525	787.2
ws_URBAN+r1k_AgUrb21+PPT	0.5245	787.36
ws_URBAN+r1k_AgUrb21+ws_IMPERVMEAN+Elevation	**0.5458**	**781.3**
ws_URBAN+ws_IMPERVMEAN+r1k_qLDI+Elevation	0.5414	782.79
ws_URBAN+r1k_AgUrb21+ws_IMPERVMEAN+PPT	0.5389	783.61

Best So. Cal. IBI model with alternative scenarios

Variable	adj-R2
All data	
r1k_AgUrb21+ws_IMPERVMEAN+ws_URBAN+PPT	0.566
Hardened Channel Removed	
ws_URBAN+r1k_AgUrb21+ws_IMPERVMEAN+Elevation	0.5458
Only Hardened Channels	
r1k_URBAN+ws_AgUrb21+ws_DamDensArea	0.4717

SOCAL IBI model with added reach scale habitat variable

- **iii)** 'Step' regression model combining landscape and *a priori* site-scale stressors
- ***Added variables: W1_HALL and P_SAFN***

Variable	adj-R2	AIC
r1k_AgUrb21+ws_IMPERVMEAN+ws_URBAN+PPT	0.566	926.35
r1k_AgUrb21 + ws_IMPERVMEAN + ws_URBAN + PPT + W1_HALL	0.5721	924.65
r1k_AgUrb21 + ws_IMPERVMEAN + ws_URBAN + PPT + P_SAFN	0.5636	928.34

Lessons learned from Option 3

- Creating separate models for different classes is improbable
  - Sample size marginal even for our well sampled region
  - Results produced counter-intuitive biological responses

- Pulling out exceptional classes and assigning *a priori* expectations has potential
  - Same numerous challenges as option 1

- Landscape models with 'forced' site scale variables could yield slightly better models
  - Adds more complexity to deal with later
  - Need to agree on site-scale variables
  - Not all sites have site level data

Questions for the Panel

- Which approach/option do you think is best?
- Can you recommend improvements on the preferred option?
- What are some outcomes you would like to see at our next meeting?