King County Normative Flow Studies Project
Science Review Team Work Session #4
Thursday, February 26th and Friday, February 27, 2004
Mountaineers Club, Seattle

Meeting Summary
April 19, 2004
FINAL

MEETING BACKGROUND

Currently, watershed-scale regulations (e.g. retention/detention standards, zoning, tree retention, etc.) are coupled with site-specific actions (capital improvement projects, in particular) to achieve a broad range of human and ecological goals in the region's rivers and streams. These applications are crude at best and becoming more difficult to defend without a clear scientific basis. The County seeks a greater diversity and effectiveness of management actions, at both the site and watershed scales. This is an ambitious goal, given the history of such efforts. Overall, the goal of SRT participation in the Normative Flow project is to help identify constructive directions for developing these new management actions.

As with previous meetings of the SRT, topics for SRT 4 were broadly divided into two areas, small streams and large rivers (specifically, the Green River). Most of the two days were given over to the first item, whose specific meeting goal was to assess recently completed stream analyses and their results. Work on the Green River is not as advanced, so those meeting goals were more generic in scope and common to the Normative Flow project as a whole, namely:

1. To assess the value and utility of indicator-based approaches, both generally and with the specific examples at hand.
2. To give feedback on the "experimental" approach for river analysis.
3. To offer program guidance on the level of understanding needed to support regulations or other management actions.
4. To offer program guidance, particularly for the near-term work plan, on future SRT roles and (or) roles of individual SRT members.

SMALL STREAMS DISCUSSION

I. Hydrologic Data

Strong attention was focused on the quality and suitability of the simulated hydrologic data presented prior to the meeting and used for the analyses.

1. Simulation vs. gauged flows

Most of this discussion was based on presented hydrographs from Little Bear Creek that directly compared simulated with gage data on an annual scale. Although no formal analyses were made, participants noted that the simulation was slow to show the onset of autumn flows and had some apparent underestimation of the larger peaks. The model itself is understood to have good matches with the annual statistics generally used in the region to evaluate calibration quality, but there has not been a systematic evaluation of whether the chosen hydrologic metrics for this study show significant differences between simulated and gaged flows.

Other opportunities to compare simulated to gage records were noted, particularly along Bear Creek (5 gages in total). It was also noted that the simulated differences between forested and current land covers was probably *under*estimated, because recent physical changes to the channel (particularly loss of wetlands) are presumed for the forested scenario as well as the current one.

Problems with the Swamp Creek hydrologic model, particularly with the upper reaches, were also acknowledged though not described in detail. The queasiness of the SRT for using simulated flows suggests that these data should be eliminated from future analyses.

2. 15-minute vs. hourly vs. daily flows

A few graphs of daily and 15-minute data were displayed. They show apparent differences between nominally "identical" hydrographs that are as large as or greater than the comparison of gaged and simulated flows. The discussion of which was "better" turned on whether the choice of data set made any operational difference. The first step would be to look at the rank order of streams' response using both time steps, based on the selected hydrologic metrics. If those rankings change depending on the choice of time steps, then the one that more closely parallels biological response is the right one to use. The same approach can (and should) also be applied to evaluate the utility of simulated vs. gaged flows.

3. Small vs. large streams

Stream size was recognized by Konrad and Booth (2002) as a significant determinant of certain hydrologic metrics—T_{Qmean}, in particular, has a marked dependence on basin size (smaller basins are naturally flashier). The inclusion of larger watersheds was an effort to expand the number of Puget Lowland B-IBI sites, but they compromise the use of any metric that evaluates the rate of hydrograph change (e.g., fall rate). Their inclusion also raises significant biological issues (see below).

II. Biological Data

Background. At SRT 3, many of the presented "biological" indicators appeared to be more physical than biological, such as "connectivity and habitat complexity." In addition, some suggested metrics were not obviously or readily measured, such as depth to groundwater, export of particulate organic matter, or extent of hyporehic zone. Since that meeting, the County's focus for small-stream biological metrics has clearly shifted to use of the B-IBI (Karr 1998, Karr and Chu 1999, Morley and Karr 2002), with some exploration of the response of individual metrics and some additional, related biological parameters. Useful fish data have not been found.

SRT 4 discussion on biological data. Although there was general agreement with the strategy of using B-IBI as the primary/only biological metric in small streams, some potential problems with the current data set were discussed.

Many of the sites have data collected by multiple people in multiple years, and not always at exactly the same location (sometimes intentionally, sometimes not). Some hydrologic "points of interest" (POI) have several distinct biological sampling sites associated with them, including some that are already known to have B-IBI scores that are clearly not a result of differences in hydrology. For example, one POI in Little Bear shows a 19 to 32 B-IBI range along a reach that includes an intentionally wide variety in local land cover. Discussion during this part of the meeting focused on whether the various scores were of equivalent quality and represented a homogeneous data set, with the issues grouped into four main categories:

1. Different data sources

Cursory review of individual graphs of total B-IBI values, and the consultants' report that individual metrics from all sources showed no apparent differences with different observers, provided a tentative conclusion that there were no *systematic* differences between different sets of data as a result of different field teams, agencies, or counting labs. This discussion did not fully resolve, however, whether levels of effort in the field were applied consistently in all cases; information is insufficient to determine whether protocols, site selection, and time of year and flow were truly equivalent. Such data may or may not be available.

2. Small total sample populations

A number of sites and scores were included where the total number of individuals counted was less than 500. The rationale offered was the desire to have some high-urban sites represented in the data set (many of which have low total population counts), but the recommendation from the SRT was clear—scores with total populations <500 simply cannot be used, unless small replicate samples can be combined to get fewer aggregated samples, each above (or at least close to) 500 individuals.

3. Systematically "low" years

Year 2000 was noted by all as a particularly low-scoring year for virtually all sites and all observers. There was general agreement that the current set of analyses should proceed with those points excluded, but there was also strong encouragement to determine the likely cause(s) for this Year-2000 anomaly. For example, were invertebrate densities lower across all taxa or were just the B-IBIs lower? What configuration of IBI metrics yielded those lower values? Hydrologically, do rainfall patterns and (or) stream discharges offer an explanation of any Year-2000 anomaly (noting that streams evaluated need not be restricted to those with available IBI data)? Was it a particularly dry year, or, perhaps, an unusually wet summer?

4. Different stream types (gradient, elevation, size)

A number of relative high-elevation, steep-channel sites were included in the original data set, primarily tributaries to the Issaquah Creek system that drain Cougar and Tiger mountains. In addition, several large systems (notably, the lower mainstem of Issaquah and Bear creeks) were included, raising questions about their biological equivalency that paralleled the discussion of these sites' hydrologic equivalency. It was noted that multiple populations of streams were being included, and the final recommendation was to exclude both the high-elevation and large trunk streams from further analyses. They should not be included in the ongoing analyses until their respective levels of biological difference are evaluated (and, if they are distinct, then until appropriate reference conditions, sampling protocols, configuration of metrics, and scoring thresholds are defined).

III. Choice of Hydrologic Metrics

Background. At SRT 3 (March 2003), extensive discussion of hydrologic data pointed to the potential use by King County of the Indicators of Hydrologic Alteration (IHA), T_{Qmean} (Konrad and Booth, 2002), seasonal stability of flows, and seasonal transitions. Discussion with the SRT at that time focused on the particular "families" of indicators chosen—i.e. monthly flows, extreme flows, frequency/duration of extremes, rates of change. There was a recognized need to define a shorter, targeted list of hydrologic indicators. Towards the end of SRT 3, the following set of hydrologic indicators for small streams was identified as likely prospects:

- High-flow pulses, especially during normally low-flow or stable flow periods
- Peak flows during emergence time for fish, and/or during spawning.
- Days above bed-sediment transport threshold.
- Timing and duration of first flow increase in fall
- Stream power, particularly during times of salmon egg incubation
- Timing and duration of e.g. 7-day low flow
- Frequency-duration relationships, such as flows under a certain threshold metric, or duration of low flows below critical threshold for a given stream
- Magnitude of 1, 7, 21-day low flows
- Baseflow stability
- Full-year and/or seasonal T_{Qmean} (e.g., late summer).

SRT discussion of hydrologic metrics. It was noted from a cursory review of simulated hydrographs from Little Bear Creek, which compared forested and current land uses, that the major changes on an annual time scale were changes in the onset of high flows in the fall and mid-winter (i.e. earlier in the autumn) and bigger individual storm peaks. The timing of flows in the late winter, spring, and summer is virtually unchanged.

Subsequent discussion of how to identify a suitable (and tractable) set of hydrologic metrics relied on plots provided for Little Bear Creek that displayed the year-to-year variability for a large number of candidate metrics based on simulated flows. The only seasonal simulation data made available was for springtime, which in the judgment of SRT 3 was not the most critical season and did not show significant relationships in the current data set.

There was discussion of how to recognize useful metrics of "general hydrologic conditions." A suggested starting point, based on the Little Bear Creek plots, was to apply the simple criterion that the metric should show a systematic difference over all years of the simulation between forested and current conditions (i.e. strong signal relative to interannual variability). There was also some reluctance to use a metric that depended on knowledge of absolute flow magnitudes, because comparison between basins would require some type of (not yet explored) scaling factor (most likely by use of the mean or median flow).

The metrics that best met these criteria for the Little Bear Creek simulation data are:

1. High pulse count
2. Fall rate (cfs/day)
3. Fall count (0.1 rule)

Others that worked reasonably well under these criteria were:

4. Rise rate (cfs/day)
5. Rise count (0.1 rule)
6. High pulse start
7. High pulse end

Each of these metrics followed the relationship between urbanization and flow anticipated by many decades of urban hydrology, and so there was general agreement that they are characterizing the broad response of these watersheds even if the absolute magnitude of that response, and the relative ranking between sites and years, was still somewhat uncertain in light of the data questions raised above.

Other candidate metrics that showed modest systematic differences between current and forested land-use simulation scenarios included one- and multi-day annual minima (although HSPF does not include any effects from irrigation or septic-tank recharge) and T_{Qmean}. Tested metrics that did not show strong differences included monthly and annual average discharge (particularly spring through mid-autumn), Julian day of annual minimum and maximum flow and of the beginning and end of low flow, and the total duration of low and high pulses. Flow maxima, particular those for intervals of about a month or less, show very strong systematic increases from forested to current conditions, but the connection of single-storm peak magnitudes to biological response is not well established and so these metrics were not pursued.

Other considerations included the importance of avoiding strong cross-correlations between chosen metrics, particularly if they are basically measuring the same thing. In the B-IBI metrics, for example, mayfly taxa richness and mayfly relative abundance are not both included because they measure more or less the same aspect of biology. Mayfly taxa richness and stonefly taxa richness, however, are both included because they measure different things (even though they are statistically correlated). The need was reiterated to have a clear, plausible connection between the hydrologic attribute being characterized by the metric and an ecological response. The hydrologic variables, in total, should capture ecologically relevant components of the flow regime – timing, magnitude, frequency, duration, rate-of-change – although many will display intercorrelations. Finally, the potential for antecedent conditions (e.g., previous-year flows) to influence current-year biota was noted from an ecological perspective and suggesting a need to explore these data for any relationships.

This discussion provided the basis for a second iteration of hydrology-biology relationships, generated and plotted between the two days of the SRT meeting. Four hydrologic metrics were displayed in that analysis: the three showing the strongest land-use dependencies from the Little Bear Creek hydrologic simulation (high pulse count, fall rate, and fall count) and T_{Qmean}, in recognition of that metric's prior success in discriminating biological response in the UW data set.

IV. Observed Hydrologic-Biological Relationships

Regressions and correlations

The overnight runs of the four hydrologic metrics on all (and various subsets) of the biological data yielded several clear results:

- Several of the hydrologic metrics are well correlated. One or two could be omitted without much loss of statistical power, although all metrics that are judged to have direct relevance to biology should be incorporated in any final set.
- Hydrologic variables correlated better with the B-IBI than with any of the component B-IBI metrics. This result was not surprising, since each metric has high variance but the individual variations tend to average out when combined, and because nearly all of the biological metrics that were considered are part of the B-IBI.
- None of the univariate plots suggested any improvement to hydrologic-biological relationships by considering time lags of one to several years.

Presence(?) of a factor ceiling

Although some graphs showed moderate linear relationships between biological and hydrologic metrics, the overarching impression was one of scattered data with at best a loose fit to a central line. Discussion began with whether multiple biological data from different locations on the stream, but associated with the same hydrologic data point ("point of interest," or POI), should be combined into a single value to minimize the scatter.

Suggested alternatives were to average all data for a POI (all sites, all observers, all years), focus on the highest score (which presumably represent the best conditions possible for the site, given its "systemic" [e.g. hydrologic] constraints), or to cull the data to use only a limited set of the highest quality information. These are not all mutually exclusive—even a culled data set would show residual scatter, because we do not expect hydrology to explain all variation in B-IBI, and single values of B-IBI for a given POI (whether maximum or averaged) would also scatter with respect to values from other POI. Consensus of the SRT was, most importantly, to explore more closely the reasons for the scatter of data. In particular:

- Use only truly comparable biological data, eliminating those scores where sample size, sampling protocol, type of human influence, or stream type is clearly different from the standard applications of B-IBI; and
- Do *not* average scores from multiple sites with the same POI, or from the same site in multiple years.

Then distribution of data points was discussed in greater detail. Many of the plots suggested a "factor ceiling," a pattern with an upper (sloping) limit to the data and lots of scatter below. The discussion first questioned whether the apparent presence of factor ceilings within many of these graphs lent support to applying the analytical framework of a factor ceiling (as opposed to a linear regression) for these data. The general consensus was that they did, and so much of the remaining discussion focused on whether hydrology was the "defining" factor for the ceiling, what other factors might be causing the scatter beneath it, and how these various hypotheses might be tested.

Historically, the factor-ceiling approach was first applied using the term "maximum species richness lines," rather than factor ceiling distribution, in an analysis of fish data from watersheds throughout the Midwest (Michigan, Illinois, Kentucky, Nebraska, North and South Dakota) twenty years ago (Fausch et al. 1984). More recently, this pattern was tied (Karr and Chu 1999, page 52-53) to an analogous concept termed a "factor ceiling distribution" in physiology and ecology (Thomson et al. 1996, Blackburn et al. 1992, Scharf et al. 1998).) This approach is also similar to that expressed in quantile regression, where a regression line is fit through the upper bound of the "ceiling" of points (Cade and Cuo 2000, Cade and Noon 2003). Implicit in all these approaches is the assumption that the upper ceiling represents a "limit" imposed by the predictor variable (here, hydrologic variables plotted on the X-axis.)

Factor ceiling(s) and the distribution of data

The relationships between various hydrologic metrics and B-IBI expressed by these graphs suggest, but do not require, that the presence and location of the upper envelope is determined by hydrologic conditions. An alternative explanation would be that some *other* factors, moderately (or well) correlated with hydrologic metrics, are the actual controlling variable(s). For example, scatter plots of total impervious area vs. B-IBI show a very similar pattern. (Note also that other graphs with other [non-hydrologic] independent variables might *also* show a factor-ceiling distribution. That outcome is neither proven nor precluded by these results.)

If hydrology-biology data can be described as a factor-ceiling distribution, then variability in other presumably relevant parameters (e.g., water quality or physical habitat) should describe much of the scatter below the factor ceiling—the "best" sites plotting closest to the envelope, poorer sites plotting progressively lower. This is a testable hypothesis (and it should be tested).

Other prior data addressing this issue offer some guidance. For example, several data sets here and elsewhere (most recently, Morley and Karr 2002) strongly suggest that intact riparian zones (or their absence) directly influence B-IBI over distances as short as a few 100 m. Other such factors of likely relevance were suggested in the discussion here and at SRT 3. These include:

- Bed stability
- Channel width and depth relative to flow
- Sediment channel geometry, including a measure of sediment mobility
- Bank stability
- Canopy closure
- Heterogeneity of channel bed topography
- Channel confinement
- Chemical outflows and other water-quality parameters

- Sediment sources due to erosion

This list provides an obvious starting point for exploring potential "secondary" factors that might explain the scatter of the hypothesized factor-ceiling distribution.

In contrast, limited aspects of physical habitat structure, described by easy-to-count metrics like large woody debris, have a very uneven record of predicting the biological quality of a given site. Larson et al. (2001), for example, found no change at all in B-IBI following addition of large woody debris in five urban channels, and improvement in a sixth only when compared to a significantly more urbanized upstream site. This is not because physical habitat is unimportant, but because "pieces of large wood" may capture only a small dimension of the larger context of physical habitat.

Summary of discussion and conclusions

At this stage in the project's development, the SRT reached several conclusions regarding the further exploration of the data, with an eye towards eventual implementation in a management context:

- Those sites with the highest quality of source data (i.e. most consistent and well-collected data, not necessarily the highest scores) should be evaluated across a full suite of conditions drawn from the list above. Relationships between hydrologic metrics and biological variables (i.e. B-IBI) should be evaluated for those sites that are not clearly impaired by known non-hydrologic factors (such as identifiable water-quality pollution or a degraded riparian corridor). The hypothesis—hydrologic alteration can determine a factor ceiling for the data—fails if such a pattern is not apparent.
- Even with such a pattern we would not expect to fully explain variation in B-IBI scores solely in terms of hydrologic variables, because other unmeasured factors have not been included (e.g., in-stream and riparian-zone conditions). This anticipated outcome should not inspire extensive searches for other variables to include in a regression model. It also should not invalidate the utility of demonstrating that a relationship does in fact exist.
- If hydrologic factors do explain the variation in the B-IBI scores for the reduced set of sites, a basis for regulating flow in pursuit of non-degraded biological conditions is established. However:

    1. Regulating for one factor only (e.g., hydrology) does not guarantee non-degradation, because there may be other such factors that influence conditions at other times or at only slightly lower levels of impairment.
    2. The presence of a ceiling relative to an independent variable of hydrologic alteration does not confirm hydrology as the sole determinant of best-attainable conditions.
    3. Presence of a "hydrologic" factor ceiling does not preclude other such ceilings, relative to other variable(s).

    In summary, improved hydrology may be shown as a strongly defensible management goal, but it will not guarantee improved biological conditions everywhere.

V. Next Steps

In summary, a range of tasks are needed to follow-up on the information presented and discussed at the meeting:

- Data quality and homogeneity must be addressed before any further analyses are run. As part of that effort, the hydrologic data should be tested systematically for dependence on the time step used and for differences between simulated vs. gage data. Eliminate sites on substantially larger systems and on mountain streams, or else conduct the systematic studies necessary to show that they do not have to be excluded because they are homogeneous with the other data.
- Cull the biological data to yield a homogeneous, high-quality set, using defined rules for selection that acknowledge the range of applicability for the metrics and methodology of the Puget Lowland B-IBI.
- Evaluate conditions at those sites with the least degree of recognized non-hydrologic disturbance. Focus this analysis on local riparian conditions, anticipating that grouping sites by their degree of local riparian cover will identity these least-impaired sites. Also consider chemical and other physical conditions, to the extent that data are available.

  The purpose of this additional analysis is to achieve a measure of confidence that hydrologic alteration has a recognizable effect on biological health, and thus "managing hydrology" is a defensible mandate. Appropriate hydrologic metrics for management purposes should also become evident from this effort; they will *not* be those historically used by the County and others to regulate flow. "Predicting B-IBI from hydrology," in contrast, is ***not*** the objective here (or anywhere else).
- Evaluate the intercorrelation of the various metrics, particularly hydrologic. Exclude those metrics that show high correlation with each other and that measure very similar flow attributes. The importance of an articulated mechanistic relationship to biological health has already been largely incorporated by the initial list of candidate hydrologic metrics; those that show strong correlation but lack such an explanation should be further explored.
- Describe how to operationalize the hydrologic metrics; determine whether continuous hydrologic modeling should be (or is) necessary to apply them.

If the County is planning to move forward with these recommendations, an experimental and analytical design should be presented to the SRT for review as soon as available.

VI. What will be needed to improve County regulations and management of streamflow

Although there was general agreement that the current state of knowledge fell far short of the ideal "level of certainty" desired for regulatory action, there was also consensus that defensible understanding of urban stream systems was well within reach, and that existing regulations have been promulgated with much less scientific basis or certainty than is already in hand. Although there has never been the expectation that this (or any other) project will identify "the" flow components that determine biology, hydrologic work appears to be well on the way to characterizing the changes from normative conditions that have biological significance, and to provide strong rationale for management efforts to restore them. B-IBI does not need to be predicted, via a mechanistic or regression relationship, in order to support a management action. We only need to know that certain kinds of hydrologic changes can produce a beneficial biological response, and we need to be able to specify how to recognize and (or) measure those kinds of changes. Conversely, recognizing the existence of a factor that imposes a limit to ecological health clearly creates the rationale for regulating that factor, presuming that maintained or improved health is an articulated public good.

LARGE RIVER DISCUSSION

The stated goals for this part of the discussion was to solicit input from the SRT on how to assemble a study plan at a large-river scale that could elucidate the same kind of flow-biology relationships as are being explored for small streams. Progress in this setting will probably require an experiment/hypothesis-testing procedure (i.e., adaptive management), given the limited number of large-river examples in King County available to provide preexisting data for comparative analysis. The Green River is the County's system of choice for this effort.

Some reservations were expressed about whether current understanding the characteristics of the Green River system is adequate to begin to run "experiments." However, interest from dam managers is also high to receive specific guidance on dam operation, for example to know what discharge level makes for problems in the downstream system and under what circumstances.

Much of the remaining discussion was predicated on the assumption that a conceptual model of how the riverine ecosystem works is necessary to make initial progress. Uncertainties facing the County as they consider choosing or developing such a model include what major elements to incorporate and what interactions between different elements to emphasize. Despite some provided examples, the SRT was not able to give concrete responses in the time frame of the meeting. It was noted that a tangible example would be more likely to engage a panel discussion, particularly one whose members have limited time for the project. Also helpful would be specific articulated needs that the County wants to address by engaging the SRT (or individual scientists) in a participatory evaluation or review. There was no clear resolution to this discussion. Absent details on needs and products, SRT members offered some generic interest but no concrete commitments. It is very unlikely that the full SRT as currently constituted will remain engaged in this effort, although some members are likely to continue to participate in the discussion if the relevant agencies would find that useful.

REFERENCES

Blackburn, T. M., J. H. Lawton, and J. N. Perry. 1992. A method for estimating the slope of upper bounds of plots of body size and abundance in natural animal assemblages. Oikos 65: 107-112.

Cade, Brian S. and Qinfeng Cuo. 2000. Estimating effects of constraints on plant performance with regression quantiles. OIKOS 91:245-254.

Cade, Brian and Barry R. Noon. 2003. A gentle introduction to quantile regression for ecologists. Front Ecol Environ 2003; 1(8):412-420

Fausch, K, D., J. R. Karr and P. R. Yant. 1984. Regional application of an index of biotic integrity based on stream fish communities. Transactions American Fisheries Society 113:39-55.

Karr, J. R., 1998. Rivers as Sentinels: Using the Biology of Rivers to Guide Landscape Management. In R. J. Naiman and R. E. Bilby (Editors). River Ecology and Management: Lessons from the Pacific Coastal Ecosystems. Springer, New York, New York, pp. 502−528.

Karr, J. R., and E. W. Chu, 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington, DC.

Konrad, C. P., and Booth, D. B., 2002. Hydrologic trends resulting from urban development in western Washington streams: U.S. Geological Survey Water-Resources Investigation Report, 02-4040, 40 p.

M. L. Larson, D. B. Booth, and S. M. Morley, 2001. Effectiveness of large woody debris in stream rehabilitation projects in urban basins: Ecological Engineering 18(2), p. 211–226.

Morley, S. A., and J. R. Karr, 2002. Assessing and Restoring the Health of Urban Streams in the Puget Sound Basin. Conservation Biology 16:1498−1509.

Scharf, F. S., F. Juanes, and M. Sutherland. 1998. Inferring ecological relationships from the edges of scatter diagrams: Comparison of regression techniques. Ecology 79: 448-460.

Thomson, J. D., G. Weiblen, B. A. Thomson, S. Alfaro, and P. Legendre. 1996. Untangling multiple factors in spatial distributions: Lilies, gophers, and rocks. Ecology 77: 1698-1715.