The Arizona Water Quality Index: A Communication Tool for Water Quality Summaries

Doug McCarty
Associate Hydrogeologist
April 2018

OFR-18-01

Contents

1.0 Introduction and Background

The Arizona Water Quality Index (AWQI) is a tool developed to better communicate water quality information in a concise and understandable way to the general public, water quality professionals, and decision-makers. The WQI operates on a set of disparate water quality data with water quality standards that differ by constituent from locations having differing designated uses. It generates a single standardized number reported on a scale from 0 to 100, with 100 representing the best water quality. The criterion for a top score is the uniform attainment of water quality standards applying to the reach or site identified. Indices can be tracked over time to demonstrate improving or degrading water quality conditions. A sub-index generated from the body of data can report on the severity of single impairment analytes on the same scale. The general index considers the percentage of distinct chemical parameters exhibiting exceedances relative to the population of distinct chemical parameters, the percentage of water quality exceedances relative to the total population of individual water quality results, and the magnitude of excursions over the most restrictive water quality standard. A combination of the method (Section 3.0), data set of core parameters and any impairment analytes (Section 4.0), and water quality standards serving as the criteria by which these data are evaluated (Section 5.0) comprise the three essential elements that must be in place for an index number to be generated.

2.0 Index Approaches, Uses, and History

2.1 Summary of Methods

A number of approaches towards the development and implementation of water quality indices can be considered. A detailed exposition of these is beyond the scope of this paper, but a brief outline of the most commonly employed methods will be given. Harbans Lai of *Water Efficiency: The Journal for Water Resource Management* has summarized several of these approachs online (Lai, 2011). Mathematical bases are outlined in Appendix A. According to Lai, the various methods include:

> **a) Weighted arithmetic mean** *(Cude 2001) - In this model, different water quality components are multiplied by a weighting factor and are then aggregated using simple arithmetic mean.*
>
> **b) Weighted geometric mean** *(McClelland 1974)- Similar to arithmetic weighted mean, each water quality component is weighted by a power factor, and then WQI is calculated using the geometric mean procedure.*
>
> **c) Un-weighted harmonic square mean** *(Dojlido et al. 1994 cited by Cude 2001) - This model is considered an improvement over the weighted arithmetic mean and the weighted geometric mean. This allows the most impaired variable to impart the greatest influence on the water quality index and acknowledges that different water quality variables will pose differing significance to overall water quality at different times and locations.*
>
> **d) Using the fuzzy logic model** *(Lermontov et al. 2009 and Nasiri et al. 2007) - This model employs artificial intelligence (AI) concept and helps capture uncertainties and inaccuracies in knowledge data. It can represent qualitative knowledge and human inference process—quite common in expressing water quality parameters—without a precise quantitative analysis.*

e) Baseline comparative model *(UNEP 2007) - This model compares water quality observations to benchmark values of different parameters instead of normalizing observed values to subjective rating curves. The benchmark values may be derived from national, state, or local water quality standards, or site-specific background values. The Canadian Council of Ministers of the Environment (CCME) used this approach for their model known as Canadian Water Quality Index (CWQI). The Global Environmental Monitoring System (GEMS)/Water Program of the United Nations Environment Program (UNEP) adopted and used the CWQI model for evaluating the quality of drinking water around the globe (UNEP 2007).*

(Lai, 2011)

Advantages of the weighted approaches include the relative transparency and straight-forward nature of the calculations. These were among the first indices to be developed and used. EPA uses components of Brown's (1970) original index formulation in various studies to the present day. The arithmetic mean approach is still used today in countries including Argentina and Turkey. McClellan (1974), who co-authored Brown's original paper, was concerned about a lack of sensitivity to low-value parameters of the weighted arithmetic mean approach, a phenomenon later termed "eclipsing" (Walsh, 2012). This was a prime motivator in his development and presentation of the geometric mean index approach. The disadvantage of both weighted methods (a., b. above) is the necessity of deriving a system of weights, which requires a subjective assessment as to which parameters should be assigned the higher weights and the relative ordering of parameters in the scheme. The subjective nature of this can introduce a bias into the index calculation.

The unweighted harmonic mean approach, developed by Cude for widespread use in Oregon, claims the advantage of not relying upon arbitrary weights. Doljido (1994) determined that this approach retained great sensitivity to the most impaired analyte while still allowing for other variables to influence the index. However, one drawback of this approach was found to be the so-called "ambiguity" effect, where all sub-indicators may indicate good water quality, but the overall indicator does not (Walsh, 2012). EPA has adopted and used the general harmonic mean method for specific rules promulgations across the U.S.

The fuzzy logic approach is based upon a method that recognizes and incorporates measures that are more qualitative than quantitative in nature. It is intended to characterize conditions where more subjective assessments relying upon linguistic and anecdotal data are of prime importance or are the most predominant data available. According to Chang (2001),

> Fuzzy set theory has been developed and extensively applied since 1965 (Zadeh, 1965). It was designed to supplement the interpretation of linguistic or measured uncertainties for real-world random phenomena. These uncertainties could originate with non-statistical characteristics in nature that refer to the absence of sharp boundaries in information. However, the main source of uncertainties involving in a large-scale complex decision-making process may be properly described via fuzzy membership functions.

The strength of the fuzzy logic approach in appropriate employment is also the primary drawback of the method. Fuzzy logic purports to accommodate subjective uncertainties in reporting and decision-making and in doing so makes subjective assessments about the fuzzy-logic weighting of different classes of data. Thus, it can “fuzzify” data that is capable of speaking for itself with more traditional statistical measures of uncertainty, such as variances, standard deviations, and errors of the mean. In other words, fuzzy logic approaches can introduce an unnecessary level of subjectivity to the determination of a water quality index. Where quantitative data exists and is sufficient to accurately characterize water quality conditions, adopting the premises of a fuzzy logic approach can dilute the significance of an index determination. If concerns about uncertainty exist for quantitative measures, in almost all cases these can be better addressed through incorporating more data and expanding set size requirements, thereby reducing tolerance and confidence limits around the index calculation.

The baseline comparative model, represented by the Canadian WQI and discussed in depth in Sections 2.6 and 3.0, has a number of advantages associated with it. There are no subjective rating curves employed in index calculations. Likewise, there are no subjectively-weighted classes or fuzzy-logic categories necessary to implement. Water quality standards or guidelines, which typically serve as the baseline for comparison, are established through an independent process which takes into account the toxicity or deleterious effects upon water quality of each water quality variable considered, based upon each variable’s unique characteristics. Therefore, additional weighting is unnecessary, since the guidelines themselves are based upon scientifically-established thresholds of water quality health. Variables are considered upon an equally-weighted basis, thus avoiding the possibility of eclipsing. The method has considerable flexibility in addressing a wide variety of water quality variables, and with modifications, it is scalable in addressing complete data sets, consistent limited data sets, individual designated uses, or individual analytes. Arizona has chosen to adopt the framework of the baseline comparative model, both for the several advantages it confers and for its compatibility and adaptability in several respects with Arizona’s methods of assessing water quality data.

2.2 National Sanitation Foundation WQI

The first practical use of a water quality index in the United States was by the National Sanitation Foundation in 1970 (Brown, 1970). A group of water quality scientists and officials were asked to select the most important water quality variables from a list of 35 possible parameters. The resulting nine top choices (dissolved oxygen, fecal coliform, pH, 5-day BOD, temperature change, total phosphate, nitrate, turbidity, and total solids) were combined into an index. The group was then asked to graph the water quality on a scale of 0-100 for typical observed ranges in each of the variables. Their results were tabulated and used to establish consensus rating curves for each of the variables. Water quality results were converted to “Q-values” on a scale of 0-100 by comparing with rating curves and then multiplied by weighting coefficients based on a test’s importance to overall water quality. Results for the nine variables were then summed to give a final index value.

The NSF WQI followed work by Horton (1965), who developed an initial index of ten variables and used arithmetic aggregation for the ten variables. The weighted sum of the ten variables was then multiplied by temperature and "obvious pollution" (Walsh, 2012). Some of Horton's development methods were later deemed somewhat arbitrary, and NSF thus convened and used the input of the afore-mentioned expert panel and dropped the multiplicative terms from the index calculation. Versions of this basic approach are still in use today by EPA when evaluating some Regulatory Impact Analyses; the weights used in EPA's six–parameter WQI are drawn directly from the weights originally used in Brown (1970).

2.3 Oregon WQI

The state of Oregon was one of the first states to develop and use a Water Quality Index in 1980. Oregon index developers eventually migrated to the unweighted harmonic square mean approach outlined above (Approach c.), incorporating eight water quality variables oriented towards protecting recreational uses. Cude (2001a) summarizes Oregon's approach:

> *The OWQI analyzes a defined set of water quality variables and produces a score describing general water quality for Oregon's rivers and streams. The water quality variables included in the OWQI are temperature, dissolved oxygen (percent saturation and concentration), biochemical oxygen demand, pH, total solids, ammonia and nitrate nitrogens, total phosphorus, and fecal coliforms. Raw data for each variable are transformed into unitless sub index values, with values of 10 being worst case and 100 being ideal. ... The OWQI was designed to permit comparison of water quality among different stretches of the same river or between different watersheds…The OWQI calculation formula, an unweighted harmonic square mean function, accounts for the variability of factors limiting water quality in different watersheds (Dojlido et al., 1994). This formula allows the most impaired variable to impart the greatest influence on the OWQI.*
>
> *It acknowledges that different water quality variables will pose differing significance to overall water quality at different times and locations…The OWQI aids in the assessment of water quality for general recreational uses (i.e., fishing and swimming). The OWQI cannot determine the quality of water for specific uses, nor can it be used to provide definitive information about water quality without considering all appropriate chemical, biological, and physical data.*

Oregon has a long history of well over 30 years with its index, with watershed reports generated on a regular basis comparing index values state-wide and regular trend analyses conducted at the same sites in the state-wide network of monitoring stations. Of note, Oregon trends are calculated over rolling ten year periods with statistical evaluation of trend conducted where data sets are large enough ($n \geq 30$) to establish significance.

2.4 State of Washington WQI

The state of Washington evaluates a set of eight variables, including temperature, dissolved oxygen, pH, fecal coliform, total nitrogen, total phosphorus, total suspended sediment, and turbidity (Hallock, 2002). Washington uses a quadratic equation in the form of

$$\mathbf{WQI = a + b_1\,(Constituent) + b_2\,(Constituent)^2}$$

Equation 1

to convert each individual water quality result to an index value between 0 and 100. Coefficients vary depending on the class of waters, the ecoregion, the constituent considered, and sometimes seasonal differences. Results are then aggregated, weighted, and processed in various ways to generate an overall number. Geomeans were used for the fecal coliform sub-index. The harmonic mean was employed to combine highly-correlated TSS and turbidity values and incorporate the lowest-scoring variable of the two. Results are evaluated relative to water quality standards to maintain designated uses for variables where numeric criteria exist; for other variables (sediment and nutrients), results are compared to regional expectations. Waters scoring above 80 are considered to "meet expectations." Waters scoring between 40 and 80 are of moderate concern, while waters below 40 do not meet expectations and are considered "highest concern." Washington's index developed from the conceptual approach behind the NSF WQI; as such, it shares some of the advantages and some of the disadvantages of the weighted arithmetic mean approach.

Washington summarizes the benefits and drawbacks of their methodology below:

> *Our WQI indicates whether water quality was either poorer than expected or poorer than necessary to support beneficial uses at a particular location. There are disadvantages, however, to comparing results to expectations or beneficial use requirements. For one thing, this approach requires subjective determinations of the beneficial uses that a particular stream segment should support, the level of water quality required to support those uses, and how critical a variation from that level of quality is. For several key parameters, the first two of these determinations are already codified in Washington's Administrative Code (WAC 173-201A). Another disadvantage is that, by design, the WQI indicates how well water quality at a station meets expectations but not how good the absolute quality is. Comparing WQIs for different stations does not indicate which station has the better absolute water quality unless expectations for both stations were the same.*
>
> *The methodology used to determine WQI scores was originally developed by the Environmental Protection Agency (EPA), Region 10. Initial development was documented only in the "gray" literature, but the methodology is similar to and perhaps based on the well-known National Sanitation Foundation index. This index uses curves to relate concentrations or measurements of various parameters to index scores and then aggregates scores to a single number. The EPA curves were "a synthesis of national criteria, state standards, information in the technical literature, and professional judgment" (Peterson and Bogue, 1989). Washington's index is based largely on these curves, adjusted to reflect local water quality standards criteria.*

(Washington, 2009)

2.5 California Central Coast Index

The state of California also uses water quality indices. The Central Coast Regional Water Quality Control Board (CCRWQCB), one of nine such boards in the state, has led the way in developing an ambitious program to generate index numbers for central California watersheds. CCRWQCB has compiled thresholds from many different sources and instead of measuring baseline acceptability of water quality, seeks to measure the true health of water bodies and watersheds in the region. Uses are split generally into human health index and an aquatic life index. Within each of these categories, broad suites of metrics are measured and evaluated, such as salts, metals, pathogens (bacteria), pesticides, and nitrogen species. As an example, toxicity, conventional analytes, ammonia, and biostimulation are some of the suites considered for aquatic life. Canadian-type scoring (outlined subsequently) is employed, but each suite is considered independently within the larger category of its orientation and so the scope sub-index of the Canadian calculation is eliminated. Excursions beyond thresholds are quantified through a "Magnitude of Exceedance Quotient" (MEQ). In each suite of possible stressors, either a harmonic mean or a worst-case individual score is taken. From the suites comprising the larger use/index, an overall harmonic mean (aquatic life) or harmonic mean/worst-case value (human health) index is determined. Provision is also made for incorporation of watershed health indicators, such as Habitat Indices, Stream Condition Indices, and Percent Natural Cover in the Watershed. This approach allows for the grading of watersheds on a health score basis by suite of interest. Examples can be seen on their website at http://www.ccamp.info/ca/view_data.php?org_id=rb3#pagetop.

2.6 Canadian/British Columbia Water Quality Index

The foundation of the Arizona Water Quality Index is provided by the Canadian Water Quality Index developed in 2001. Prior to its adoption nationwide by the Canadian Council of Ministers of the Environment (CCME), the index was used by the province of British Columbia in a slightly altered form. Other provinces developed variations of an index with the same purpose, used indices that ultimately were incorporated as one component of the final index, or used British Columbia's index formulation with modifications. However, the British Columbia water quality index (BCWQI) had the broadest and most comprehensive applicability of the province indices that CCME considered for adoption, and it also featured the longest and most extensive development and application period. Like Arizona, British Columbia had different designated uses and water quality objectives for its waters where the index was to be applied, including drinking water, recreation, irrigation, livestock watering, aquatic life and wildlife. Like Arizona, British Columbia had water quality guidelines that monitoring results were evaluated against. In researching possible existing indices, it was these similarities that eventually led to the adoption of the BC/Canadian approach for Arizona's needs. The structural similarities between BC's water quality evaluations and Arizona's water quality evaluations indicated a high degree of compatibility between the methods and provides for a relatively seamless adoption process.

3.0 Method

The CCME published a technical report in 2001 (CCME, 2001a) that outlines the components and calculations of the index. It is reproduced here as the most succinct explanation of the index composition[1] :

<u>CCME Water Quality Index Formulation</u>

The index consists of three factors:

Factor 1: Scope

*F1 (**Scope**) represents the extent of water quality guideline non-compliance over the time period of interest. It has been adopted directly from the British Columbia Index:*

$$F_1 = \left(\frac{Number\ of\ failed\ variables}{Total\ number\ of\ variables}\right) \times 100$$

*Where **variables** indicates those water quality variables with objectives which were tested during the time period for the index calculation[2].*

Factor 2: Frequency

*F2 (**Frequency**) represents the percentage of individual tests[3] that do not meet objectives ("failed tests"):*

$$F_2 = \left(\frac{Number\ of\ failed\ tests}{Total\ number\ of\ tests}\right) \times 100$$

The formulation of this factor is drawn directly from the British Columbia Water Quality Index.

Factor 3: Amplitude

*F3 (**Amplitude**) represents the amount by which failed test values do not meet their objectives. F3 is calculated in three steps. The formulation of the third factor is drawn from work done under the auspices of the Alberta Agriculture, Food and Rural Development.*

---

[1] Equations in CCME excerpt not included in document equation numbering.

[2] The index definition of a ***variable*** is a set of results for the same measured chemical or physical property. Examples of variables include dissolved oxygen, total arsenic, pH, lead, and dissolved copper.

[3] The index definition of a ***test*** is an individual sample result on a specific day and time for a specific variable. An example of a hypothetical test for a site might be the concentration of dissolved copper on January 1, 2000. An example of a ***failed test*** would be a dissolved copper exceedance of the water quality standard for a designated use at the site on Jan. 1, 2000.

(i) The number of times by which an individual concentration is greater than (or less than, when the objective is a minimum) the objective is termed an "excursion" and is expressed as follows.

When the test value must not exceed the objective:

$$excursion_i = \left(\frac{Failed\ Test\ Value_i}{Objective_j}\right) - 1$$

For the cases in which the test value must not fall below the objective:

$$excursion_i = \left(\frac{Objective_j}{Failed\ Test\ Value_i}\right) - 1$$

ii) The collective amount by which individual tests are out of compliance is calculated by summing the excursions of individual tests from their objectives and dividing by the total number of tests (both those meeting objectives and those not meeting objectives). This variable, referred to as the normalized sum of excursions, or nse, is calculated as:

$$nse = \frac{\sum_{i=1}^{n} excursion_i}{Total\ number\ of\ tests}$$

iii) F3 is then calculated by an asymptotic function that scales the normalized sum of the excursions from objectives (nse) to yield a range between 0 and 100.

$$F_3 = \left(\frac{nse}{0.01 * nse + 0.01}\right)$$

The CCME WQI is then calculated as:

$$CCMEWQI = 100 - \left(\frac{\sqrt{{F_1}^2 + {F_2}^2 + {F_3}^2}}{1.732}\right)$$

The factor of 1.732 arises because each of the three individual index factors can range as high as 100. This means that the vector length can reach

$\sqrt{100^2 + 100^2 + 100^2}$ = $\sqrt{30{,}000}$ = *173.2* *as a maximum. Division by 1.732 brings the vector length down to 100 as a maximum.*

Discussion

…Of the many problems inherent in the development of this index, two required most attention: the varying scale of measurements, and the range of exceedance. Differing scales of measurement are characteristic of water quality analyses. Some parameters, such as pesticides, may be environmentally significant at $ng \cdot L^{-1}$ ranges, while others are significant at the $mg \cdot L^{-1}$ range. Adopting the objective-oriented approach developed in British Columbia allows these types of data to be assembled in the same multivariate index formulation, since the metric of interest is the comparison of the measured data relative to its objective.

The objective-oriented BC WQI also avoided the problem of weighting parameters. There were discussions of how to deal with objective exceedance for a parameter such as phosphorus relative to a parameter with more toxic associations such as PCBs. The Subcommittee felt that since the relative toxicities of different chemicals were addressed during the development of water quality objectives, further weighting was not warranted.

Another problem frequently encountered in reporting on water quality data is results below the analytical detection limit. It is problematic to deal with these numbers statistically, but the problem is avoided with the CCME WQI approach. 'Less than' values are used in the index as observations which are within the objectives, so the results are counted, but all the statistical problems associated with how to deal with them are circumvented.

Applying the Index

Applying this index to water quality data sets must be done with due regard to how the index is formulated. Experience with the British Columbia index has shown that misapplication, or use of the index for purposes for which it was not designed, can lead to erroneous conclusions. There are several rules for application that should be taken into consideration:

a) ***Index comparisons should only be made when the same sets of objectives are being applied.***
The CCME WQI allows the index user to select the objective set on which to compare measured water quality. This is a design feature that increases the versatility of the index considerably but allows for misuse. Different jurisdictions in Canada use different objectives for water quality, and there are usually different objectives for different water uses. Objectives designed for the protection of water used for irrigation or livestock watering will be different from those designed to protect sensitive aquatic life. If an index value is calculated on one set of objectives and compared to an index value based on a completely different set of objectives, any conclusions drawn will be wrong.

b) ***Index comparisons should only be made using the same sets of parameters****.*
This is common sense "apples to apples" reasoning. Comparing a site where most of the measured parameters are pesticides to a site where most of the measured parameters are metals will yield information of limited value. It is possible to obtain index values under these conditions, but comparison of these types of sites will only tell the user how each site is doing relative to those objectives. There is no way the index can replace a detailed site assessment of different types of pollutants. Similarly, if a trend through- time index series is calculated for a

specific site and the number and type of water quality parameters change significantly during the course of the time series, meaningless conclusions may be drawn.

c) ***Care should be taken with older data.***
Many data sets can go back to times when the sensitivity of analytical methodology was considerably less than with more modern methods. This is of particular concern in cases where there are older results that appear to be just above the detection limit. For example, metals data generated in the 1970's may have been obtained using colorimetric methods with detection limits significantly above current water quality objectives. All analytical methods are capable of producing 'false positive' results and incorporation of these into the index can provide misleading conclusions. For example, if older cadmium data was derived from a method with a detection limit of 0.01 $mg \cdot L^{-1}$ *there will probably be results at (or slightly above) the detection limit. These may or may not be valid. If these data are run in the index against an objective of 0.0002* $mg \cdot L^{-1}$ *false positives will represent very large excursions over the objective and questionable index values will result.*

d) ***The index should be run on parameter sets relevant to the water body being tested.***
Several jurisdictions, including Ontario and Québec have older data sets where large suites of parameters were tested. The CCME WQI should only include 'relevant' parameters in the calculation. Because of the way the index is calculated, the inclusion of many parameters (for example, all pesticides in a 'scan') may result in unrealistically low index values. For example, gas chromatographic – mass spectrometer scans will often provide large amounts of data on many chemicals simultaneously. Including all of these data in index calculations will artificially depress the index value. This will be of particular concern in trend-through-time index evaluations when the number of tested parameters varies significantly, or in situations where comparisons between sites are desired.

e) ***Minimal data sets should not be used.***
The CCME WQI was not designed to replace proper evaluation of water quality conditions through thorough assessment of water quality chemicals of concern. The CCME WQI should not be run with less than four parameters and four sampling visits per year.

Despite these restrictions on its use, the CCME WQI has been successfully applied in several Canadian jurisdictions and has produced values that contain valuable information with regard to trends through time and spatial discrimination of impacted and non-impacted sites. The committee feels that it has application as a management and communication tool if applied appropriately.

(CCME, 2001a)

4.0 Data

As a measure originating with the Impaired Waters Identification Rule in 2002 (ADEQ, 2002), the core parameter data set addressing each designated use was developed to ensure that data sets considered for possible impairment listings met minimal requirements for numbers of samples, seasonal distribution of those samples, and temporal and spatial independence of the data considered. Each designated use for an assessment unit (stream reach or lake) had certain water quality variables designated as core parameters. These variables were chosen to ensure that the most important elements for the meeting of the designated use reflect satisfactory conditions and thereby allow the use to be fulfilled. Each designated use had its own set of water quality variables selected, with those selected reflecting best professional judgment as to which variables were the most critical to consider when evaluating whether uses were hampered. The following discussion is presented in ADEQ's Methods and Technical Guidance document (2014):

> ***Core Parameters and Seasonal Distribution -*** *… Monitoring data are collected at sites and during conditions selected to be representative of the varying conditions. Since a water quality standard might be more likely to be exceeded during critical conditions such as high or low flows or during seasonal conditions when recreation is more active, samples should be collected under different conditions to determine whether the surface water is really "attaining" its designated uses (seasonal distribution).*
>
> *Although all parameters with numeric standards are used for assessment, ADEQ has chosen a set of indicators, called "core parameters," necessary to assess whether each designated use is attaining standards. Arizona's core parameters are shown in the table below.*
>
> *Core parameters were selected based on EPA's CALM (Consolidated Assessment and Listing Methodology) guidance (2002), although they are limited due to the lack of narrative standards implementation procedures. CALM guidance places strong emphasis on narrative water quality standards, suggesting that core indicators should include bioassessments, habitat assessments, ambient toxicity testing, contaminated sediment, health of individual organisms, nuisance plant growth, algae, sediments, and even odor and taste.*

Core Parameters

DESIGNATED USE	CORE PARAMETERS
Aquatic and Wildlife	Dissolved oxygen (not required if ephemeral) Stream flow (if a stream) Sample depth (if a lake) pH Total nitrogen (if nutrient standards established) Total phosphorus (if nutrient standards established) Dissolved cadmium, copper, and zinc and hardness
Fish Consumption	Total mercury
Full Body or Partial Body Contact	*Escherichia coli* (not required if ephemeral) pH
Domestic Water Source	Nitrate/nitrite or nitrate pH Fluoride Total arsenic, chromium or chromium VI, and lead
Agricultural Irrigation	pH Total boron and manganese
Agricultural Livestock Watering	pH Total copper and lead

However, Arizona is currently limited to physical-chemical parameters. Arizona's choice of core parameters will change in future assessments as new numeric and narrative standards, criteria, and assessment tools are developed.

Core parameters were chosen using the following criteria:

- *Frequently exceeded standards in past assessments;*
- *Routinely included in ambient monitoring suites;*
- *Lab reporting limits routinely below applicable surface water criteria;*
- *Critical toxicity recognized; and*
- *Standards and implementation procedures support application of the criteria.*

For example, dissolved metals exceedances and low pH measurements are often found in historic mining areas. E. coli bacteria and nitrate were chosen because they can cause serious human illness or death if standards are exceeded, and they are important in determining support of Body Contact and Domestic Water Source designated uses.

Core parameters must be sampled at least three times and samples must be reasonably distributed at different times of the year to reflect seasonal changes (seasonally distributed). For assessment purposes, it is ensured that at least one sample is collected in each of the four seasons: winter (December – February), spring (March – May), summer (June – August), and fall (September – November). If this does not occur, and the designated use is not "impaired," then the designated use is assessed as "inconclusive."…

To assess a designed use, all core parameters must be represented seasonally. For example, although numerous E. coli bacteria samples were collected, the assessment unit is assessed as attaining Full Body Contact only if pH was also collected with seasonal distribution…

(ADEQ, 2014)

The core parameter set provides a standardized scaffold by which water quality index determinations can be calculated consistently from one determination to the next. Retrieval of all core parameters for all of the designated uses of a site or reach ensures that important variables for the most fundamental and essential assessment of the condition of water quality are considered. ADEQ samplers collecting field data and taking samples are tasked with ensuring that core parameter coverage is obtained during the course of their sampling or investigations.

No data set for the determination of WQIs can be considered complete if it omits data for water quality variables previously determined as impaired for the reach or water body of concern. Misleading pictures of the overall health of the reach would result from indices calculated from core parameters alone when the reach has been identified as impaired, with multiple exceedances of a water quality standard that is not covered by the core parameter data set. Consequently, it is strongly recommended that any determination of a general index for a given time frame and a site or reach/water body should include all data within the time frame for impairment analytes. An Excel-based VBA program designed to calculate indices allows for the entry of up to two water quality variable codes for additional impairment analytes beyond the core parameter data set.

Beyond the compilation of the core parameter set and specific impairment analytes, it serves little purpose, and in fact can prove to be detrimental, to expand the data retrievals to all sampled parameters in a time frame. The addition of extra water quality variables and additional test results can have the effect of “watering down” the index and result in the reporting of a higher value than is warranted. Furthermore, sites frequently have different objectives in their establishment and sometimes have differing suites of chemicals sampled in the execution of program sampling plans. This variation, when left unconstrained, can result in the indices that appear the same but are calculated on very different sets of data. The restriction of index runs to the core parameter data set and impairment analytes standardizes comparisons to the extent reasonably possible while also ensuring that all designated uses are considered in the general index number.

5.0 Water Quality Standards as Criteria

Arizona water quality standards generally serve as the basis for the criteria considered in the index calculations. An exceedance of the water quality standard is the yardstick by which the index is decremented from a top score of 100; an exceedance will simultaneously count as one of *n* water quality variables showing adverse water quality impacts in the scope term (F_1) of the calculation (Equation 2), one of the total number of results in the data set showing adverse impacts in the frequency term (F_2) of the calculation, and the magnitude of the exceedance is tabulated relative to the standard for the amplitude term (F_3) of the index calculation. However, while water quality standards serve as the basis for the calculation, this is not to suggest that the index does not deviate from strict 305(b) water quality assessment methodology. The index follows a modified and streamlined methodology developed specifically for it and thus may occasionally generate results that do not necessarily accord with the assessment status of the reach or water body.

$$WQI = 100 - \left(\frac{\sqrt{{F_1}^2 + {F_2}^2 + {F_3}^2}}{1.732} \right)$$

Equation 2

Results are considered only once relative to the most stringent standard. A result that exceeds the most stringent standard associated with a designated use is not further evaluated against less stringent standards of the other designated uses that apply. Likewise, a result that does not exceed the most stringent standard is not further counted as not exceeding less stringent standards for the other designated uses that apply. Such consideration would artificially inflate or suppress an index value through multiple evaluations of a single result. Visit-specific hardness levels are considered for dissolved metals calculations, with each result compared to the standard applicable for the reported hardness value for that sample. Chronic standards are used for comparison for aquatic and wildlife uses of cold, warm, and effluent-dependent water (edw) streams or sites. Acute dissolved standards are used for comparison for the ephemeral A&W designated use.

A number of Arizona water quality standards rely on a metric other than a comparison of a simple single-valued result to a numeric standard. Annual means, geometric means within a 30 day period, 90^{th} percentile values, and median concentrations are a few of these other metrics. Generally, the WQI does not attempt to address these more specialized forms of standards expressions. For *E. coli* evaluations, only the single sample maximum is considered, consistent with Arizona assessment methodology of *E. coli* results. Where suspended sediment concentration constitutes an impairment analyte, individual results are measured against the SSC median standard for the reach. No attempt is made in index calculation to address the SSC standard requirement of a minimum of four samples with each taken seven days apart, nor is an attempt made to screen data according to the 48 hour window exclusion after storm events. The method of calculating Arizona nutrient annual means is unwieldy and impractical for incorporation in the index, and 90^{th} percentile requirements for nutrient evaluations are disregarded. Instead, individual nutrient results are compared to site-specific single sample maximum values (after the addition of major nitrogen species in the case of N), and an unstratified-time period mean is taken of all nutrient values in the data set for comparison against the mean standard for both nitrogen and phosphorus where applicable.

It is emphasized here that the occasional disagreement of an index value with assessment status is a feature, not a bug. It is a different perspective or window on the data, and as such may serve as a corrective or qualifying adjunct to the reach's formal impairment status. Users of the index may initially be concerned that the WQI disagrees with assessment status, or that it may not show a problem as severe when the reach is formally considered as impaired. It would be a mistake to attempt to make the index conform to the assessment status; it may well be that the assessment status is questionable, that the data used for listing was suspect or non-representative, or the water body is otherwise incorrectly assessed. The index, if left in its objective and independent form, may be the only tool available that can demonstrate that. Or, it may be that the different methods of calculating and assessing water quality simply point to differing conclusions about the state of water quality in the reach. In the same way that scientists are encouraged not to throw out outliers simply because they don't conform to expectations, users of the water quality index who are cognizant of the assessment status of the reach are encouraged to keep an open mind towards the possibility of disagreement and allow that condition to exist if necessary. Such disagreements may in fact eventually shed light on water quality standards that are not properly or adequately researched before implementation, or an assessment methodology that may be deficient in certain respects.

6.0 Analyte-specific Indices

The scope, frequency, and amplitude components of the WQI can be modified to yield an index number specific to a given analyte. This approach is valuable in assessing the severity of individual impairments on the same scale of 0 to 100 as the general index. In essence, the scope term of index calculation drops out, and the index is determined with the frequency and amplitude components alone. For the sake of continuity and clarity, the sub-terms are not re-numbered in Equation 3 below. With preparatory calculations the same as for the general index for the retained sub-indices, the Analyte WQI is then calculated as:

$$Analyte\ WQI = 100 - \left(\frac{\sqrt{{F_2}^2 + {F_3}^2}}{1.414}\right)$$

Equation 3

Where F_2 is restricted to the number of exceedances and results for the individual analyte alone and F_3, likewise, is calculated on the basis of only the individual analyte being considered.

The factor of 1.414 derives from the elimination of the scope term, resulting in a maximum possible numerator value of 141.4.

$$\sqrt{100^2 + 100^2} = \sqrt{20{,}000} = 141.4$$

Division by 1.414 allows for reporting on the 0-100 scale with 100 as a maximum.

7.0 Categorization of Results

Canada has developed a spectrum of classification associated with its index as outlined below:

> *The assignment of CCME WQI values to categories of water quality is termed "categorization" and represents a critical but somewhat subjective process. Categorization should be based on the best available information, expert judgement, and the general public's expectations of water quality. The categorization presented here is preliminary and will no doubt be modified as the index is tested further. Because of the nature of the index, it is impossible to determine from an index range whether the ranking is due to extreme excursions in one variable, or frequent small excursions in one or more variables. … Once the CCME WQI value has been determined, water quality can be ranked by relating it to one of the following categories:*
>
> ***Excellent:*** *(CCME WQI Value 95-100) – water quality is protected with a virtual absence of threat or impairment; conditions very close to natural or pristine levels. These index values can only be obtained if all measurements are within objectives virtually all of the time.*
>
> ***Good:*** *(CCME WQI Value 80-94) – water quality is protected with only a minor degree of threat or impairment; conditions rarely depart from natural or desirable levels.*
>
> ***Fair:*** *(CCME WQI Value 65-79) – water quality is usually protected but occasionally threatened or impaired; conditions sometimes depart from natural or desirable levels.*

Marginal: *(CCME WQI Value 45-64) – water quality is frequently threatened or impaired; conditions often depart from natural or desirable levels.*

Poor: *(CCME WQI Value 0-44) – water quality is almost always threatened or impaired; conditions usually depart from natural or desirable levels.*

(CCME, 2001a)

As mentioned in the narrative, the act of classification into broad categories is a subjective act, and as such, the boundaries between classes are, at least initially, somewhat arbitrarily established, and subject to further refinement as data accrues. The spectrum of water quality outlined above (Excellent -- > Poor) is generally endorsed for Arizona's WQI. Considered at a more general level, "Excellent " and "Good" categories can be considered "Acceptable" or "Generally Satisfactory" water quality; "Fair" and "Marginal" classes can be considered as "At Risk" or "Of Concern" water quality; and the "Poor" class rates as "Unacceptable" or "Unsatisfactory" water quality. Early indications are that general WQIs can differ considerably in the categorization from analyte-specific WQIs for the same site or reach. Broad-scale WQIs frequently receive a higher categorization than the indices for impairment analytes; this is because the presence of core parameters not deemed as impaired in the reach often serves to attenuate the general index and lift it to a higher classification while impaired analyte calculations do not enjoy the same advantage. In some cases however, where there are a number of minor (in degree or amplitude) exceedances scattered across several variables, the converse may be true; the general number may be suppressed to a lower categorical description than any analyte-specific score. It should be noted as well that the scale breaks for categorization may ultimately differ between general indices and specific indices. These distinctions in class determinations will take time to discern.

The numeric breaks between classes that Canada has adopted are likely based upon their evaluation of a national data set over several years of application; they are not recommended for uncritical acceptance for the Arizona WQI. Canada's description of each category prominently refers to the time element ("frequently", "almost always", "rarely", etc.) in its evaluation of index numbers. While some consideration of time is implicit in the compilation of index data sets, it is not clear that time is the most important or even the most prevalent element for consideration in the Arizona WQI. Canada has abundant water resources and true perennial waters with robust sampling programs that routinely visit sites multiple times in any given time frame. Arizona differs significantly from Canada in these regards. The difference in flow regimes and the importance of differing time considerations is particularly true for data sets of ephemeral reaches, which by definition flow only for limited periods of time. Consideration must be given to the relatively infrequent sampling visits to each site or reach (usually on a quarterly basis, but sometimes monthly) and the episodic and rotating visitation schedule; generally, most Arizona sampling does not follow one site or reach continuously for a focused long-term trend analysis. Rather, the focus for Arizona is more oriented towards determining what is exceeding water quality standards, and what proportion the exceedances/impairments occupy relative to the entire data set. For these reasons and others, it is recommended that Arizona begin to establish an empirical data base of its own for evaluation as to where class breaks lie. Compiling such a data base will take some time and a substantial body of data where indices have been evaluated. These evaluations should be broken out between general WQIs and analyte-specific WQIs.

Provisional class boundaries for descriptive categorization based on a limited data set thus far suggest the following:

- Excellent/Good – 100 : No WQI standard exceedances
- Fair - 80-99 : Occasional exceedances of mild to moderate magnitude
- At-Risk - 60-79 : Exceedances of mild to moderate magnitude common, or occasional exceedances of high magnitude
- Poor - Below 60: Frequent and persistent exceedances of moderate to high magnitude

8.0 Examples

Early work with the index has been applied to varying types of waters in Arizona for various time periods to assess how the index responds to real-world data and to gauge the difference in responses between broad-scale general index numbers and analyte-specific index values. The index has been run on individual sampling sites as well as whole reaches. The data set has been sufficiently varied such that the entire range of index values has been explored; some high-quality waters in the state have logged scores of 100 for the time periods evaluated, while several severe impairments were characterized in the 10-30 range. One individual analyte score of 0 was also recorded. Common time periods examined included pre-defined water quality assessment periods, historic al pre-TMDL periods, five year planning periods, TMDL project sampling periods, and post-TMDL implementation periods. Types of sites or reaches generally included impaired waters, identified “high quality waters,” long-term spatial aggregation sites, and sites/locations with completed TMDLs. See Tables 1 and 2 for a summary of index values.

It is not unusual or particularly difficult for an Arizona water body to achieve a score of 100; all that is necessary is that no water quality exceedances were logged for the core parameter data set, and the water body is unimpaired for any other constituent. Unlike California’s index, where a high score indicates pristine water conditions of the highest caliber, the objective for use of the Arizona index is to evaluate results according to water quality standards that form the basis of the biennial water quality assessment. The introduction of any exceedance in the data set decrements the index from 100. Therefore, a top score for the Arizona WQI indicates simply that the water is not considered water quality-limited for the constituents characterizing the most fundamental level of designated use fulfillment. An Arizona index value can be considered a baseline water quality evaluation against the minimum thresholds for water use set forth by the Clean Water Act.

One advantage of using a data set paired down to core parameters and impairment analytes is that the index becomes more sensitive in reflecting changes in water quality. With the use of complete data sets, the signal of water quality degradation could frequently be lost in the sheer number of results returned if degradation happened selectively and was not broadcast equally across all constituents sampled. For example, in paired instances in the test collection, one site was evaluated for the five years prior to a large wildfire and the five year period after the same wildfire. The general WQI dropped from 80 to 40 from one period to the next as a consequence of the fire. The same location was tested for a planning period baseline after the post-wildfire period, and it was noted the general WQI had rebounded to 77 for this period. The range of response in this instance is desirable, given the magnitude of effects noted on water quality from this major fire. As water quality subsequently recovered to near its former level, the index bounced back readily, showing elasticity with no “hang-over” effect. This illustrates another feature of the index; the index is evaluated solely on the numbers reflected in the data set for the sampling visits. There is no carry-over effect from prior degradation, as long as the data set no longer contains the numbers causing a previous degradation. We can see as a result that the length of the time frame evaluated, the choice of starting dates, and the consistency between compared sets are all-important in generating index values that can be interpreted with confidence.

The index responds slowly when water quality exceedances exceed the standard by more than an order of magnitude. This is largely attributable to the calculations of the F3 sub index, which generally assume linear space as the domain of water quality responses to stressors. For all but a couple of constituents, the F3 term evaluates individual results by simple linear multiples of factors beyond the standard. Exponential responses are generally not assumed. Additionally, in the normalization of the sum of excursions back to a 0-100 scale, the entire range of excursions must be expressed on the 100 point sub-scale. Thus, less than a one order of magnitude difference between standard and existing concentrations would show a near-linear response between 10 and 100; a second order of magnitude difference would report between 1 and 10; and a third order of magnitude difference differs only from 0.1 to 1. Consequently, the score of 0 from an analyte-specific index previously mentioned has recovered only to 1 thus far in the implementation phase, though concentrations of the impairment analyte have dropped substantially; existing concentrations are still well above one order of magnitude different from the standard.

Reach	Description	Type Site/Reach	Interval Descriptor	General WQI
15060103-015	Cherry Creek	High-quality Waters	12/14 Assessment period	100
15060203-026B	West Clear Creek	High-quality Waters	12/14 Assessment period	100
15020001-027	South Fork LCR	High-quality Waters	12/14 Assessment period	94
15020001-013B	West Fork LCR	High-quality Waters	12/14 Assessment period	93
15040005-025	Eagle Creek	High-quality Waters	12/14 Assessment period	91
15040004-025B	Blue River	High-quality Waters	06/08 Assessment period	86
15060105-013A	Tonto Creek Reach 13A	Impaired reach	Post TMDL two year period	91
15060103-887	Gibson Mine Tributary	Listing Period - Impaired water	2006/08 Listing Assessment Period	24
15040002-004	Gila River - Bitter Creek to NM State line	Listing Period - Impaired water	2008 Listing Assessment Period	--
15040005-022	Gila River - Yuma Wash to Bonita Creek	Listing Period - Impaired water	2006/08 Listing Assessment Period	--
15050203-001	San Pedro River - Aravaipa Creek to Gila River	Listing Period - Impaired water	2004 Listing Assessment Period	84
15070201-003	Colorado River - Morelos Dam site 09522000	Long-term aggregation site	Planning Period Baseline	95
14070006-001	Colorado River - Lee's Ferry 09380000	Long-term aggregation site	Planning Period Baseline	91
15040005-022	Upper Gila UGGLR448.61 09448500	Long-term aggregation site	Planning Period Baseline	83
15060103-004	Salt River - Roosevelt Lake to Pinal Creek	Long-term aggregation site	Pre- R-C fire , 5 years frame	80
15040005-022	Upper Gila UGGLR448.61 09448500	Long-term aggregation site	Post-TMDL data"	79
15060103-004	Salt River SRSLR107.43 09498500	Long-term aggregation site	Planning Period Baseline	77
15040005-022	Upper Gila UGGLR448.61 09448500	Long-term aggregation site	Pre-TMDL data*	50
15060103-004	Salt River - Roosevelt Lake to Pinal Creek	Long-term aggregation site	Post- R-C fire, 5 year frame	40
15060105-013A	SRTON056.59	Reach aggregator site	Post TMDL two year period	96
15070201-003	Gila River - Coyote Wash to Fortuna Wash	TMDL completed - 4A	TMDL Project Sampling	83
15040002-004	Gila River - Bitter Creek to NM State line	TMDL completed - 4A	TMDL Project Sampling	80
15040005-022	Gila River - Yuma Wash to Bonita Creek	TMDL completed - 4A	TMDL Project Sampling	80
15070101-008	Gila River - Gillespie Dam to Centennial Wash	TMDL completed - 4A	TMDL Project Sampling	71
15050203-001	San Pedro River - Aravaipa Creek to Gila River	TMDL completed - 4A	TMDL Project Sampling	40
15020002-004	LCR - Silver Creek to Carr Lake Draw	TMDL completed - 4A	TMDL Project Sampling	34
15060105-353	Christopher Creek	Impaired reach	Historic, Pre-TMDL	88

Table 1. General WQIs for selected sites and reaches

Reach	Description	Impairment	Imp. 1 An	Impairment	Imp. 2 An	Impairm	ıp. 3 An	Im	ıp. 3 Ana	Time Frame
15060103-015	Cherry Creek	**N.A.**								7/1/06-6/30/11
15060203-026B	West Clear Creek	**N.A.**								7/1/06-6/30/11
15020001-027	South Fork LCR	**N.A.**								7/1/06-6/30/11
15020001-013B	West Fork LCR	**N.A.**								7/1/2006-6/30/2012
15040005-025	Eagle Creek	**N.A.**								7/1/06-6/30/11
15040004-025B	Blue River	**N.A.**								1/1/2000-12/31/2005
15060105-013A	Tonto Creek Reach 13A	**94**	D.O.	**83**	E. coli					1/1/2013 - 12/31/2014
15060103-887	Gibson Mine Tributary	**0**	Cu, Diss	**36**	pH					1/1/2000 - 12/31/2004
15040002-004	Gila River - Bitter Creek to NM State line	**31**	SSC							10/30/2002-12/31/2006
15040005-022	Gila River - Yuma Wash to Bonita Creek	**56**	SSC							1/1/2000-12/31/2006
15050203-001	San Pedro River - Aravaipa Creek to Gila River	**61**	E.coli							1/1/1998-12/31/2002
15070201-003	Colorado River - Morelos Dam site 09522000	**96**	D.O.	**100**	Se					7/1/2008 - 6/30/2013
14070006-001	Colorado River - Lee's Ferry 09380000	**N.A.**								7/1/2008 - 6/30/2013
15040005-022	Upper Gila UGGLR448.61 09448500	**100**	SSC	**76**	E. coli					7/1/2008-6/30/2013
15060103-004	Salt River - Roosevelt Lake to Pinal Creek									6/1/1997 - 6/30/2002
15040005-022	Upper Gila UGGLR448.61 09448500	**73**	SSC	**73**	E. coli	**48**	Pb			9/1/2010 - 12/31/2013
15060103-004	Salt River SRSLR107.43 09498500	**100**	SSC	**96**	E. coli	**90**	Total N	**88**	Total Phos	7/1/2008-6/30/2013
15040005-022	Upper Gila UGGLR448.61 09448500	**58**	SSC	**63**	E. coli	**63**	Pb			1/1/1976 - 12/31/2008
15060103-004	Salt River - Roosevelt Lake to Pinal Creek									7/1/2002 - 6/30-2007
15060105-013A	SRTON056.59	**73**	E. coli							1/1/2013 - 12/31/2014
15070201-003	Gila River - Coyote Wash to Fortuna Wash	**91**	B	**100**	Se*					6/1/2006-3/31/2008
15040002-004	Gila River - Bitter Creek to NM State line	**100**	SSC*	**47**	E. coli	**75**	Pb			6/1/2006-3/31/2008
15040005-022	Gila River - Yuma Wash to Bonita Creek	**86**	SSC	**51**	E. coli					6/1/2006-3/31/2008
15070101-008	Gila River - Gillespie Dam to Centennial Wash	**20**	B	**12**	Se					6/1/2006-3/31/2008
15050203-001	San Pedro River - Aravaipa Creek to Gila River	**26**	E.coli							5/1/2007-4/30/2012
15020002-004	LCR - Silver Creek to Carr Lake Draw	**31**	SSC	**37**	E.coli					11/1/2006-9/30/2010
15060105-353	Christopher Creek	**90**	E.coli	**99**	Phosphorus					1994-2003

Table 2. Impairment Analyte WQIs for sites and reaches of Table 1

9.0 Sensitivity Analysis and Data Adequacy QA

9.1 Constituents with Exponential Responses

Selected parameters were considered in a sensitivity analysis to test index responses and compare assumptions of linear responses to assumptions of exponential responses. *E. coli*, as a variable that tend to respond in exponential fashion in stormflow conditions, was tested to determine if the default linear calculation for the magnitude of excursions led to disproportionately suppressed index scores. Hypothetical data sets of sizes n = 4, n = 10, and n = 25 were constructed, with single and multiple (2) exceedances at both just over the standard and at ten times the standard. Five different calculation protocols were tested – the default linear multiple calculation, the default formula substituting log 10 values for linear values, the default formula substituting natural log values for linear values, and a simple difference of both log 10 values and natural logarithm values. Results are displayed in Figure 1.

1 exceedance at 10x standard

n =	Linear Calc	Log10 space	LN space	Log10 Diff	LN Diff
4	40	64	64	62	56
10	66	92	92	90	85
25	80	94	94	94	92

1 exceedance at just over standard

n =	Linear Calc	Log10 space	LN space	Log10 Diff	LN Diff
4	64	65	65	65	64
10	93	93	93	93	93
25	97	97	97	97	97

2 exceedances at 10x standard

n =	Linear Calc	Log10 space	LN space	Log10 Diff	LN Diff
4	24	48	48	45	37
10	52	85	85	82	74
25	70	94	94	92	88

2 exceedances at just over standard

n =	Linear Calc	Log10 space	LN space	Log10 Diff	LN Diff
4	49	50	50	50	49
10	85	86	86	85	85
25	94	94	94	94	94

Figure 1. Exceedance simulations on index responses

Reviewing results with an eye towards an assessment protocol criterion of a 10% exceedance rate indicating impairment and an index threshold of 80 or lower raising concern, the default linear calculation (yellow-shaded columns) responded too drastically to marginal exceedances and indications of impairment. Other methods responded less sensitively, in some cases not resulting in a large-enough movement of the needle. Ultimately, the log 10 difference method (orange-shaded columns) occupied the happy medium, maintaining acceptably satisfactory scores for marginal exceedances, and showing a decline for more serious impairments that would draw attention while not unduly dragging down the overall score. This sensitivity exercise led to adoption of the log 10 difference as the F3 calculation protocol for constituents expected to respond in storm flows in an exponential fashion – namely, *E. coli* and suspended sediment concentration. These have since been modified through use and subsequent review to a natural log difference.

9.2 Set Size as an Index Determinant

The exercise also highlights the importance of set sizes as a major determinant of index responses. It is clear to see in the yellow-shaded columns of Figure 1 that the same raw number and magnitude of exceedances can result in vastly different index readings, ranging from index values of 64 to 97 and 49 to 94 for one and two exceedances respectively just over the standard, and from 40 to 80 and 24 to 70 for one and two exceedances respectively at values of 10 times the standard. These variations are entirely attributable to the number of samples under consideration in each set; a fixed number of exceedances in a set can constitute a widely-varying percentage of a set size without floors. This percentage is reflected in the F2 sub-index calculation. Large sets are not a concern; the index response as set size grows larger decreases and stabilizes as it approaches its asymptotic limit, and the confidence interval around a large-set index narrows as the set grows, assuring a more reliable number. However, small sets are a great concern, as index values are highly unstable and subject to great changes simply due to the expansion of set size. This variability is completely independent of the addition of any other exceedances or the magnitude of any additional exceedances.

Consequently, use of the index on small sample sets is not recommended. Water quality sampling is an exercise in statistical characterization, with random variations possible for any given sample. The first sample of a set of visits may turn up an exceedance, or the 25th sample might. A larger set of values collected leads to more certainty about measures of central tendency, values and ranges of outlier events, and the percentage of visits that can be expected to result in exceedances of water quality standards. But little can be said with confidence about a single value, or a low number of values, since variation in a population is distributed randomly and an exceedance of standards may occur in any given sampling event, whether the first or the tenth. Index numbers generated from only a few samples have a much higher degree of uncertainty associated with them, and a much lower level of reliability and confidence that can ascribed to the accurate characterization of the water body. As a provisional rule-of-thumb, set sizes of less than three visits should not be evaluated, and set sizes of less than 10 should be reported as provisional indices. When an adequate number of indices have been generated so that set sizes and index variation can be statistically evaluated against one another, it will be possible to state with more confidence where an absolute cut-off would be. Ultimately, values should be dictated by the characteristics of the data and evaluations themselves.

Recommendations for minimal data set sizes necessarily imply recommendations on time frames for which the index is employed. ADEQ considers data collected at least seven days apart as the minimum time to establish temporal independence of samples. This rule is used to aggregate samples for consideration in the water quality assessment. In absolute terms, one month is the minimum to achieve a four sample set, but logistical considerations for sampling routinely prevent this from occurring. This is illustrated well by the fact that the *E. coli* water quality standard, which incorporates geomean criteria requiring four samples in 30 days, is almost never assessed on this criteria, as such frequent data collection is not feasible in almost all instances. Practically speaking, achieving a minimum data set of four for even a qualified index value requires a minimum of two months of collection time, since field visits are only rarely scheduled on a persistent bi-weekly basis. More typically, visits are made on a monthly or quarterly basis: monthly visits on average might be expected in monsoon season when conditions can change rapidly, while quarterly visits are usual to characterize water quality when conditions are stable and not expected to change much. Seasonal or semi-annual time frames for index determinations are the minimum that make practical sense for generation of indices, with much longer time frames (annual, 5 year, historical) more suitable yet. The water quality index is not a tool that is designed, nor is it suitable, for snapshot evaluations and cursory judgments about water quality conditions.

9.3 Criteria for Data-Adequate Indices

The following criteria are adopted to provide the minimum assurance that data sets with a general water quality index score can be considered data-adequate and thus sufficiently reliable to attain at least provisional status.

- All parameters required for a core parameter evaluation are present in the data set.
- Each core parameter has a minimum of three values.
- Data coverage for each core parameter represents at least three of the four seasons (seasonal distribution).
- Impairment analytes, whether in the core parameter set or added to it, are not sampled disproportionately relative to the remainder of the data set. Criteria for evaluating disproportion include the following:
    - Percentages of exceedances to total samples for the analyte exceeds ten (10) percent.
    - Percentage of samples for the impaired analyte relative to the data set exceeds 2x the percentage of the parameter to total parameters in the set. Example: If one of 5 constituents in a data set (20%) is an impairment analyte, no more than 40% of individual results can be for the impairment analyte. If one of 12 parameters is an impairment analyte (8.3%), no more than 16.7% of the data set can consist of impairment sample results.

Meeting of these criteria with an additional minimum of ten samples shall be considered sufficient to attain full unqualified status for the index.

10.0 Conclusion

Arizona has adapted the approach of the Canadian Water Quality Index (CWQI) to its unique data assessment needs and the arid ecosystems and hydrologic regimes of the American Southwest. In following the Canadian model, Arizona is affirming the suitability of the UNEP baseline comparative model for its needs, wherein water quality benchmarks for variables serve as the baseline for data comparisons in index calculations. UNEP adopted the CWQI as the foundational model of its effort to develop a global water quality index. In part, it was the flexibility and wide applicability of the Canadian method that UNEP found appealing; its adoption of the CWQI in 2007 centered on a desire to apply World Health Organization drinking water guidelines to its data sets. The same flexibility is an attractive feature for Arizona as well; Arizona's water quality standards readily fulfill the role as benchmarks for data comparisons. One strong advantage of this approach is that subjective rating curves or arbitrarily-established weights are dispensed with; direct comparisons to the same standards that guide Arizona water quality assessments provide for unambiguous index calculations predicated upon the same basis that guides Arizona water quality assessments. Measurements that are meeting water quality standards do not penalize the index; where measurements exceed water quality standards, the scope, frequency, and magnitude of those exceedances is considered in the calculation. Consistency in the composition of data sets is assured by limiting index calculation to data sets comprised of impairment analytes, if any, for the water of interest and Arizona's *core parameters*. Core parameters are a set of water quality variables established to ensure that the most important variables for each designated use (i.e., those with frequently-observed exceedances, critical toxicity, and routinely sampled parameters) were considered in all assessment evaluations. Minimal thresholds for data coverage and distribution are also required for core parameters prior to assessing impairment. The adoption of the core parameter data set as its basis thus establishes a consistent framework for the employment of the AWQI. Other impairment analytes can also be considered within this framework.

The AWQI shows great promise as a metric to evaluate improvement in water quality over time. Trend analysis of indices is one of the prime reasons to pursue development and implementation of the index. Provided that care is taken to ensure data sets are consistent in composition and durations, trends with statistical significance could be determined with a good degree of reliability. The state of Oregon has been in the vanguard of uses of indices in this way, and pointers on methods of application could well be gleaned for Arizona from Oregon's experience with indices. Other features and advantages of the index are numerous; the scale is simple and easily understandable, requiring only a general orientation as to what constitutes a top score (100). The index is readily-scalable in terms of spatial application, whether in application to a single site, a stream reach, or more broadly to an entire stream when designated uses are consistent across all reaches. The index also has great flexibility when considered temporally, as it can be calculated for any time frame of interest, provided enough data is available to generate indices that satisfy data-adequacy criteria. Lastly, further development and adaptation here in Arizona of the index's conceptual basis has allowed for the creation and use of *analyte-specific indices*, with a modified mathematical basis, to assess the degree of severity of any individual variable's impairment.

In conclusion, while an index evaluation is not a substitute for the more comprehensive and rigorous analysis of water quality that accompanies state-wide assessments and other investigations, its advantage is that it provides an intuitive, easily grasped summary of important water quality analytes indicating the water quality health of Arizona streams as codified in Arizona's water quality standards. The Arizona WQI provides a simple non-technical tool by which the general public, water managers, and decision-makers can assess and generally understand the baseline water quality of any Arizona stream or sampling site in a flexible conceptual framework that can be consistently applied across the state and for any time period.

11.0 References

ADEQ, 2002. Impaired Waters Identification. Title 18. Environmental Quality. Chapter 11 Department of Environmental Quality Water Quality Standards, Article 6. Phoenix, AZ. 2002.

ADEQ, 2009. Water Quality Standards for Surface Waters. Title 18. Environmental Quality. Chapter 11 Department of Environmental Quality Water Quality Standards, Article 1. Phoenix, AZ. Jan. 2009.

ADEQ, 2014. Surface Water Assessment Methods and Technical Support. Section 3: Data Interpretation and Assessment Criteria. Unpublished. Arizona Department of Environmental Quality, Phoenix, AZ. March 2014.

Brown, R. M., N. I. McClelland, R. A. Deininger, and R. G. Tozer. 1970. A Water Quality Index—Do We Dare? Water and Sewage Works, 11: 339-343.

Canadian Council of Ministers of the Environment. 2001(a). Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, Technical Report. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Canadian Council of Ministers of the Environment. 2001(b). Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, User's Manual. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Chang, N., H. W. Chen and S. K Ning, 2001. Identification of river water quality using the Fuzzy Synthetic Evaluation approach. *Journal of Environmental Management* (2001) 63, 293–305.

Cude, C. G., 2001a. Oregon Water Quality Index: A Tool for Evaluating Water Quality Management Effectiveness. Journal of the American Water Resources Association, 37(1): 125-137.

Cude, C. G., 2001b. "The Oregon Water Quality Index (OWQI) – A Communicator of Water Quality Information", Paper presented to the American Water Resources Association Conference on Water Quality Monitoring and Modeling, San Antonio, Texas, 2001.

Dojlido, J.R., J. Raniszewski, J. Woyciechowska, 1994. Water Quality Index Applied to Rivers in the Vistula River Basin in Poland. Environmental Monitoring and Assessment 33:33-42.

EPA, 2002. *Consolidated Assessment and Listing Methodology – Toward a Compendium of Best Practices* (CALM). URL: http://www.epa.gov/owow/monitoring/calm.html .

Hallock, 2002. A Water Quality Index for Ecology's Stream Monitoring Program. Washington State Department of Ecology Publication Number 02-03-052. Seattle, WA. November, 2002. URL: https://fortress.wa.gov/ecy/publications/publications/0203052.pdf

Horton, R. K., 1965. "An Index-Number System for Rating Water Quality." Journal of the Water Pollution Control Federation **37**(3): 300-305.

Lai, 2011. The Introduction to the Water Quality Index: Expressing water quality information in a format that is simple and easily understood by common people. Water Efficiency: The Journal for Water Resource Management. September-October 2011. URL:

http://www.waterefficiency.net/WE/Articles/The_Introduction_to_the_Water_Quality_Index_15374.aspx. Accessed 2/17-19/2015.

McClelland, N. I., 1974. Water Quality Index Application in the Kansas River Basin. EPA-907/9-74-001. US EPA Region VII. Kansas City, MO.

Washington Department of Ecology, 2009. River and Stream Monitoring Water Quality Index. URL: http://www.ecy.wa.gov/programs/eap/fw_riv/docs/WQIOverview.html . Accessed April 14, 2015.

UNEP, 2007. Global Drinking Water Quality Index Development and Sensitivity Analysis Report. United Nations Environment Programme, GEM/Water Programme Office, Burlington, Ontario, Canada. URL: http://www.un.org/waterforlifedecade/pdf/global_drinking_water_quality_index.pdf

Walsh, P., Wheeler, W., 2012. Water Quality Index Aggregation and Cost Benefit Analysis. Working Paper #12-05. July, 2012. US EPA National Center for Environmental Economics. Washington, D.C. 20460.

Worcester, K., D. Paradies, J. Hunt, 2014. Central Coast Regional Board Watershed Report Cards. URL: http://www.mywaterquality.ca.gov/monitoring_council/healthy_streams/docs/healthywatersheds_krw.pdf. Accessed 2/18/2015.

Appendix A: Common Index Methods
Mathematical Bases

a) Weighted arithmetic mean

Equation 4:
$$WQI = \sum_{i=1}^{n} SIi * Wi$$

Where *WQI* = Water Quality Index

SIi = Sub-index i

n= number of sub-indices

Wi = Weight given to sub-index i

b) Weighted geometric mean

Equation 5:
$$WQI = \prod_{i=1}^{n} SIi \wedge Wi$$

Where *WQI* = Water Quality Index

SIi = Sub-index i

n = number of sub-indices

Wi = Weight given to sub-index i

c) Un-weighted harmonic square mean

Equation 6:
$$WQI = \sqrt{\frac{n}{\sum_{1=1}^{n} \frac{1}{SI_i^2}}}$$

Where *WQI* = Water Quality Index

SIi = Sub-index i

n = number of sub-indices

Wi = Weight given to sub-index i

d) Using the fuzzy logic model - This approach, as demonstrated by Lermontov et al. 2009 and Nasiri et al. 2007, presents the following advantages over the conventional numerical process:

1. Ability to capture large variety of non-linear relations

2. Easily adoptable (*sic*) to local conditions

3. Could be interpreted verbally

4. Could include information that other methods cannot include such as individual knowledge and experience

5. Possibility of enhancing the results by combining qualitative information with the quantitative data that expresses the ecological status of the water body

6. Better handling of situations with missing data without affecting the results significantly

Appendix B: Core Parameter Chemical IDs and STORET Codes

DESIGNATED USE	CHEMICAL ID	CHEMICAL NAME	ANALYSIS TYPE	STORET CODE	EPA STORET NAME
A&W	3095	CADMIUM	DISSOLVED	01025	CADMIUM, DISSOLVED
A&W	3139	COPPER	DISSOLVED	01040	COPPER, DISSOLVED
A&W (where applicable)	3523	PHOSPHORUS	TOTAL	00665	PHOSPHORUS, TOTAL (MG/L AS P)
A&W	3525	ZINC	DISSOLVED	01090	ZINC, DISSOLVED
A&W	3549	DISSOLVED OXYGEN	DISSOLVED	00300	OXYGEN, DISSOLVED MG/L
A&W	3549	DISSOLVED OXYGEN	STANDARD	00301	OXYGEN, DISSOLVED, PERCENT OF SATURATION %
A&W (where applicable)	3551	KJELDAHL NITROGEN	TOTAL	00625	NITROGEN, KJELDAHL, TOTAL, (MG/L AS N)
A&W (where applicable)	3695	NITROGEN	TOTAL	00600	NITROGEN, TOTAL (MG/L AS N)
A&W (where applicable), DWS	3552	NITRATE + NITRITE	TOTAL	00630	NITRITE PLUS NITRATE, TOTAL 1 DET. (MG/L AS N)
A&W, xBC, DWS, AgI, AgL	3542	PH	TOTAL	00406	PH, FIELD
A&W, xBC, DWS, AgI, AgL	3542	PH	STANDARD	00400	PH (STANDARD UNITS)
AgI	3079	BORON (BORON AND BORATES ONLY)	TOTAL	01022	BORON, TOTAL
AgI	3319	MANGANESE	TOTAL	01055	MANGANESE, TOTAL
AgL	3139	COPPER	TOTAL	01042	COPPER, TOTAL
DWS	3042	ARSENIC, INORGANIC	TOTAL	01002	ARSENIC, TOTAL
DWS	3533	FLUORIDE	TOTAL	00951	FLUORIDE, TOTAL (MG/L AS F)
DWS	3537	CHROMIUM	TOTAL	01034	CHROMIUM, TOTAL
DWS	3537	CHROMIUM	STANDARD	01032	CHROMIUM, HEXAVALENT

DESIGNATED USE	CHEMICAL ID	CHEMICAL NAME	ANALYSIS TYPE	STORET CODE	EPA STORET NAME
DWS, AgL	3311	LEAD AND COMPOUNDS (INORGANIC)	TOTAL	01051	LEAD, TOTAL
FC	3322	MERCURY, ELEMENTAL	TOTAL	71900	MERCURY, TOTAL
FC	3322	MERCURY, ELEMENTAL	DISSOLVED	71890	MERCURY, DISSOLVED
FC	3322	MERCURY, ELEMENTAL	DISSOLVED	99920	MERCURY-DIS,FILTERED WATER,ULTRATRACE METHOD
FC	3322	MERCURY, ELEMENTAL	TOTAL	50092	MERCURY-TL,UNFILTERED WATER,ULTRATRACE METHOD
xBC	3567	E. COLI	TOTAL	99906	ESCHERICHIA COLIFORM (E.COLI) (COLILERT)
xBC	3567	E. COLI	TOTAL	31648	E. COLI - MTEC-MF N0/100ML
xBC	3567	E. COLI	TOTAL	31633	E.COLI,THERMOTOL,MF,M-TEC,IN SITU UREASE #/100ML
xBC	3567	E. COLI	TOTAL	90902	E. COLI, MODIFIED M-TEC (EPA METHOD 1603)

Appendix C: Core Parameter Standards and Formulas

STORET CODE	EPA_STORET_NAME	Standard*	Magnitude of Exceedance Formula(s)**
01025	CADMIUM, DISSOLVED (cold)	Chronic: $e^{(0.7409*LN(Hardness)-4.719)}*(1.101672-LN(Hardness)*0.04183$	R/S - 1
01025	CADMIUM, DISSOLVED (warm, edw)	Chronic: $e^{(0.7409*LN(Hardness)-4.719)}*(1.101672-LN(Hardness)*0.04183$	R/S - 1
01025	CADMIUM, DISSOLVED (ephemeral)	Acute: $e^{(1.0166*LN(Hardness)-3.924)}*(1.136672-LN(Hardness)*0.041838$	R/S - 1
01040	COPPER, DISSOLVED (cold, warm, edw)	Chronic: $e^{(0.8545*LN(Hardness)-1.702)}*(0.96)$	R/S - 1
01040	COPPER, DISSOLVED (ephemeral)	Acute: $e^{(0.9422*LN(Hardness)-1.1514)}*(0.96)$	R/S - 1
01090	ZINC, DISSOLVED (cold, warm, edw)	Chronic: $e^{(0.8473*LN(Hardness)+0.884)}*(0.978)$	R/S - 1
01090	ZINC, DISSOLVED (ephemeral)	Acute: $e^{(0.8473*LN(Hardness)+3.1342)}*(0.978)$	R/S - 1
00300	OXYGEN, DISSOLVED MG/L	7.0, 6.0, 3.0 (A&Wc, A&Ww, A&Wedw)	S/R - 1
00301	OXYGEN, DISSOLVED, PERCENT OF SATURATION %	90%	N.A.
00406	PH, FIELD	6.5 - 9.0	Lower Bound: S/R - 1; Upper Bound: R/S - 1
00400	PH (STANDARD UNITS)	6.5 - 9.0	Lower Bound: S/R - 1; Upper Bound: R/S - 1
00625	NITROGEN, KJELDAHL, TOTAL, (MG/L AS N)	Site-specific	SSM: Σ R/S - 1 Mean: I/S - 1
00630	NITRITE PLUS NITRATE, TOTAL 1 DET. (MG/L AS N)	Site-specific (10,000 ug/L for DWS)	SSM: Σ R/S - 1 Mean: I/S - 1; DWS: R/S - 1
00600	NITROGEN, TOTAL (MG/L AS N)	Site-specific	SSM: R/S - 1 Mean: I/S - 1
00665	PHOSPHORUS, TOTAL (MG/L AS P)	Site-specific	SSM: R/S - 1 Mean: I/S - 1
01022	BORON, TOTAL	1000 ug/L	R/S - 1
01055	MANGANESE, TOTAL	980 ug/L	R/S - 1
01042	COPPER, TOTAL	500 ug/L	R/S - 1
01002	ARSENIC, TOTAL	50 ug/L	R/S - 1
00951	FLUORIDE, TOTAL (MG/L AS F)	4000 ug/L	R/S - 1
01034	CHROMIUM, TOTAL	100 ug/L	R/S - 1
01032	CHROMIUM, HEXAVALENT	21 ug/L	R/S - 1
01051	LEAD, TOTAL	15 ug/L	R/S - 1
71900	MERCURY, TOTAL	0.6 ug/L	R/S - 1
71890	MERCURY, DISSOLVED	0.6 ug/L (surrogate analyte)	R/S - 1
99920	MERCURY-DIS,FILTERED WATER,ULTRATRACE METHOI	0.6 ug/L (surrogate analyte)	R/S - 1
50092	MERCURY-TL,UNFILTERED WATER,ULTRATRACE METH	0.6 ug/L	R/S - 1
99906	ESCHERICHIA COLIFORM (E.COLI) (COLILERT)	235, 576 (FBC, PBC)	LN(R) - LN(S)
31648	E. COLI - MTEC-MF N0/100ML	235, 576 (FBC, PBC)	LN(R) - LN(S)
31633	E.COLI,THERMOTOL,MF,M-TEC,IN SITU UREASE #/100ML	235, 576 (FBC, PBC)	LN(R) - LN(S)
90902	E. COLI, MODIFIED M-TEC (EPA METHOD 1603)	235, 576 (FBC, PBC)	LN(R) - LN(S)
80154#	SUSPENDED SEDIMENT CONCENTRATION	25, 80 (cold, warm)	LN(R) - LN(S)
99913#	SUSPENDED SEDIMENT CONCENTRATION, FINE FRACTI	25, 80 (cold, warm)	LN(R) - LN(S)
99914#	SUSPENDED SEDIMENT CONCENTRATION, COARSE FR/	25, 80 (cold, warm)	LN(R) - LN(S)

*Most stringent standard applied for applicable set of designated uses

** R - Result; S - Standard

- SSC not a core parameter, but special provisions made to calculate