Development of Climate Change Indicators and Metrics

STEP 6 ASSESSMENT NOTES

October 4, 2017

Introduction

As the Chesapeake Bay climate change indicators project has evolved and unfolded, it has become apparent to ERG that the scoring approach we originally proposed in our Work Plan (Task 2) and the criteria we laid out in our Frameworks Assessment (Task 3) may not be sufficient to prioritize indicators for workgroup consideration and ultimately for inclusion in the implementation plan. Thus, we have developed an adjusted, enhanced approach for scoring and prioritizing candidate indicators. This document reiterates the original scoring criteria, identifies strengths and limitations of these criteria, provides more detail on how we will apply the criteria—including a standardized rubric for assigning high-moderate-low scoring—and identifies supplementary considerations and how we can include them in the scoring process.

Original Scoring Criteria

During the framework development stage of this project, ERG and CBP agreed to the following initial critera:

- Every indicator must meet the following criteria:

    1. **Topical relevance:** The indicator provides information about physical climate trends, ecological or societal reponse, or programmatic progress toward resilience. The connection to climate change is documented or can be explained easily.
    2. **Spatial coverage:** The indicator provides information that is specific to the Chesapeake Bay, the Chesapeake Bay watershed, or geographic sub-units within the watershed.
    3. **Temporal coverage:** Multiple years of data are available to describe changes or trends, and the latest available data are timely.
    4. **Actual observations:** The indicator is based on observed data. Modeling and statistical inference (if any) is limited to spatial interpolation between data points, such as the process used to generate a gridded map.
    5. **Credible methods:** The indicator is based on sound data collection and analytical methods that reflect the state of the science.
    6. **Data quality and integrity:** The data provider uses quality assurance procedures to ensure data quality and management systems to protect the integrity of the data.
    7. **Objectivity:** The indicator is developed and presented in a clear, complete, and unbiased manner that accurately represents the underlying trends in physical conditions.
    8. **Uncertainty:** Sources of uncertainty are known and understood.
    9. **Transparency and reproducibility:** The specific data used and the specific assumptions, analytical methods, and statistical procedures employed are clearly stated. Documentation is sufficient to allow the indicator to be reproduced independently.
    10. **Feasibility:** The indicator is feasible to construct, and a program is in place to continue to collect data, thereby allowing the indicator to be updated in the future.

- Certain indicators must meet the following criterion:

    1. **Peer-review validation:** If an indicator is based on physical measurements of environmental conditions, it must use data from a peer-reviewed publication, a program that uses peer-reviewed methods to collect and analyze data, and/or a program whose data have been used and validated in peer-reviewed publications. This criterion will likely apply to all indicators in the *physical climate trends* bin and certain indicators in the other two bins (for example, a measure of benthic community condition). For indicators that are not based on physical measurements, peer review is ideal but not required.

- Every indicator will ideally meet the following additional criteria, which can be used to prioritize indicators for development:

    1. **Relationship to other indicators:** The ideal indicator will complement other indicators rather than duplicating them. It fills a vital role in the organizational framework. Where possible, an ideal indicator will have established causal relationships with other indicators, which can be evaluated.

    2. **Spatial coverage:** The ideal indicator will use data collected throughout the Bay and its major tributaries or throughout the watershed, as opposed to indicators that are only measured at a few locations.

    3. **Spatial resolution:** The ideal indicator will provide at least a total or an average for the Bay, the watershed, or the individual states that are part of the watershed. Where possible, the ideal indicator will support local-scale analysis by providing data that are downscaled further—for example, data for individual sampling sites, sub-watersheds (e.g., HUC-12), NOAA climate divisions (up to 10 per state), or a gridded map.

    4. **Temporal coverage:** The ideal indicator will have many years of data available. The best indicators will have at least 30 years of data, which is a common threshold for climatological analysis. The ideal indicator will also have a defined baseline, particularly if it is used to assess progress toward resilience.

    5. **Temporal resolution:** The ideal indicator will have data with at least annual frequency, with sub-annual frequency if appropriate (e.g., where seasonal variations are important to consider).

    6. **Consistency of methods:** The ideal indicator will be based on data collection and analytical methods that are comparable across time and space. In some cases, it may be appropriate to use data that were collected or analyzed using multiple methods—for example, supplementing short-term records with longer-term records from a different source. In such cases, the data visualization should distinguish between the different sources, such as by inserting a discontinuity in a time series or plotting multiple lines on a graph. The CBNERR indicators by UMCES and Chesapeake Data provide a good example of this approach.

    7. **Uncertainty:** The ideal indicator will have low uncertainty—for example, small error bars or narrow confidence intervals.

    8. **Other limitations:** The ideal indicator will have few counfounding factors or other limitations that make it difficult to interpret the data or draw conclusions.

    9. **Understandability:** The ideal indicator will provide a clear depiction of observations that can be understood by both technical and non-technical users.

Strengths and Limitations of the Original Scoring Criteria

The original criteria are intended for assessing the quality of an existing indicator or data source. Yet it has become increasingly clear that some of the most desirable topics do not yet have data collected—and we want those topics to remain on our final list. We don't want to rule out an indicator that *could* meet the criteria someday, just because it does not meet the criteria right now.

Although they are meant to be objective, the criteria as written still require some judgment on the part of the scorer, which could pose a particular problem if scoring is performed by a team of people who interpret the wording of the criteria differently. Thus, one way to enhance the criteria is to lay out specific standards for what constitutes a high, moderate, or low score, etc. This document does so below.

A few of the criteria cannot be assessed at this early stage in the indicator development process:

- "Objectivity: The indicator is developed and presented in a clear, complete, and unbiased manner that accurately represents the underlying trends in physical conditions." *Assessing this criterion requires a review of how the indicator is compiled, presented, and documented. It can only be fully assessed after the indicator development stage.*
- "Transparency and reproducibility: The specific data used and the specific assumptions, analytical methods, and statistical procedures employed are clearly stated. Documentation is sufficient to allow the indicator to be reproduced independently." *Assessing this criterion requires a review of how the indicator is compiled, presented, and documented. It can only be fully assessed after the indicator development stage.*
- "Relationship to other indicators: The ideal indicator will complement other indicators rather than duplicating them. It fills a vital role in the organizational framework. Where possible, an ideal indicator will have established causal relationships with other indicators, which can be evaluated." *This criterion cannot be scored for each indicator in a vacuum. It can only be assessed when we see the narrowed list of high-priority indicators that begins to take shape, because it requires that each indicator be assessed in relation to the other indicators that are part of the recommended suite. By necessity, this criterion must be assessed later.*

Criteria that focus on data quality will contribute to overall scoring, but they are unlikely to help us identify indicators that *add the most value* because of their topical relevance, their uniqueness, or their connections to management actions. These other types of considerations are embodied in the voting and scoring activities that workgroup members and other stakeholders conducted in the spring and summer, where a vote for a topic generally reflected an assessment of *value* more than an assessment of data quality. It has become increasingly clear to us that we will need to include measures of value and give them substantial weight in the calculus that leads us to our narrowed list of indicators to propose for implementation. Otherwise, a calculation based purely or largely on data quality runs the risk of selecting indicators because they are easy, not because they are useful.

The original scoring criteria could score most of the indicators in one bin (say, Physical Climate Trends) higher than the indicators in another bin (say, Resilience), simply due to differences inherent in these types of indicators. We could envision using the criteria scores to help us select the best indicators within each bin, but not for comparisons across bins. In fact, the best use of the original scoring criteria might be to distinguish between different data options for the same topic, rather than for comparisons between indicators. For example, if three data sources are available to describe trends in sea level, our scoring criteria could provide an objective means to identify which source offers the highest quality indicator. Thus, if an indicator has more than one potential data source, we should separately score each one.

Enhanced Scoring Approach

1. **Parse the table out to facilitate scoring by data source.** Within each indicator topic, if we have identified more than one viable metric or data source, we will split each metric or source on a separate row in the matrix.

2. **Score each row based on the required criteria, reflecting the quality of the currently available data.** Each criterion will be applied on a yes/no basis, excluding #7 and #9, which are designed to be applied after indicators have been developed. An indicator or data source either meets each requirement or it does not. A final column will tally whether each data source meets all of the applicable required criteria. A perfect score is either 8 or 9, depending on whether the indicator is derived from a physical (not administrative) measurement that requires peer-reviewed data collection and analytical methods. ERG will have multiple people score the indicators, and we will combine their scores.

3. **Identify how many of the required criteria could conceivably be met in the future.** This step requires consideration of data collection programs or analyses that are still under development. The resulting score will indicate whether we think each indicator is a viable candidate for future development if sufficient data do not exist already.

4. **Eliminate from further consideration any row that we do not expect to at least be able to meet all of the required criteria in the future.** This step may narrow the list somewhat. We will not actually remove these rows from the matrix, but we will gray out the columns associated with the subsequent scoring steps.

5. **Score each row based on the desired criteria, reflecting the quality of the currently available data.** Each criterion will be scored on a high/moderate/low basis, with corresponding point values of 2/1/0. See the table in Attachment 1 for a detailed rubric that ERG will follow to score every indicator objectively and consistently. We will then calculate a total score for each row. As with step 2, ERG will have multiple people score the indicators, and we will combine their scores.

6. **Select the best data set for each topic.** For indicator topics with more than one source option or metric, select the row with the highest total desired criteria score. The highest-scoring row could be a hybrid indicator derived from multiple data sources—or this step could prompt us to propose a new hybrid that combines two or more rows that each scored high on their own. The highest-scoring row is the one we will carry through to the next step. Other rows can be grayed out from this point forward.

7. **Incorporate information about the value that each topic will add to a suite of indicators.** Prior to the workgroup meeting, we will insert the original vote total from the series of workgroup engagements that we conducted in the spring and summer. We hope to develop a more refined set of scores based on a more nuanced exploration of this issue at the October 16 CRWG meeting.

8. **Combine all available information to identify a draft set of approximately 20 indicators that we would propose to carry forward into an implementation plan.** This step will require us to determine relative weights for the "value" scores versus the desirable data quality criteria scores. Our initial suggestion would be to weight the value scores much more heavily—say, find the sum of the data quality score plus three times the value score. We will try this approach and see what we come up with. To cover all three bins, we can select the five highest-scoring indicators from each bin, plus five "wild card" indicators that represent the next five highest-scoring candidates from any bin. This process will provide a preliminary set of 20 indicators to at least prompt discussion.

Attachment 1: Detailed Scoring for Desirable Criteria

Criterion	High (2 points)	Medium (1 point)	Low (0 points)
1. Relationship to other indicators: The ideal indicator will complement other indicators rather than duplicating them. It fills a vital role in the organizational framework. Where possible, an ideal indicator will have established causal relationships with other indicators, which can be evaluated.	Has causal connections with four or more other proposed indicators. Addresses a unique aspect of the Chesapeake that is not covered by other indicators. *This topic will be assessed later, when the final list of recommended indicators comes more into focus.*	Has causal connections with one to three other proposed indicators. Addresses a unique aspect of the Chesapeake that is not covered or only partially covered by other indicators. *This topic will be assessed later, when the final list of recommended indicators comes more into focus.*	Has no causal connections with other indicators, and/or it overlaps with other indicators. *This topic will be assessed later, when the final list of recommended indicators comes more into focus.*
2. Spatial coverage: The ideal indicator will use data collected throughout the Bay and its major tributaries or throughout the watershed, as opposed to indicators that are only measured at a few locations.	Data are available to characterize the condition of the entire Bay or the entire watershed, etc., as appropriate for the topic in question. A site-based monitoring program can qualify for this score if the station density is sufficient to allow Bay-wide aggregation, such as with an interpolated map, or if the variable in question is one that does not vary much over space and therefore can be broadly represented by a small number of stations.	Data are available to characterize the condition of significant portions of the Bay or the watershed. A strong monitoring program that is present in Maryland but absent in Virginia (or vice-versa) could qualify for this score.	Data are available for a relatively small number of locations that are not necessarily representative of conditions across the Bay or watershed as a whole.

Criterion	High (2 points)	Medium (1 point)	Low (0 points)
3. Spatial resolution: The ideal indicator will provide at least a total or an average for the Bay, the watershed, or the individual states that are part of the watershed. Where possible, the ideal indicator will support local-scale analysis by providing data that are downscaled further.	Data are available at the scale of individual sites, sub-regions of the Bay or watershed (for example, counties), or a total or average for the Bay or the watershed. Bay-wide or watershed-wide totals or averages can be downscaled to sub-regions or to individual sites on a map to support local-scale analysis.	Data are available at the scale of a total or average for the Bay or the watershed, but no downscaling is possible to support local-scale analysis.	Data are available in the form of national, mega-regional (e.g., NCA Northeast climate region), or state totals that cannot be downscaled to just the Chesapeake Bay watershed.
4. Temporal coverage: The ideal indicator will have many years of data available. The best indicators will have at least 30 years of data, which is a common threshold for climatological analysis. The ideal indicator will also have a defined baseline, particularly if it is used to assess progress toward resilience.	Consists of 30 or more years of comparable data with minimal gaps in coverage.	Consists of 10 to 30 years of comparable data with minimal gaps in coverage.	Consists of fewer than 10 years of data, when accounting for data gaps.
5. Temporal resolution: The ideal indicator will have data with at least annual frequency, with sub-annual frequency if appropriate (e.g., where seasonal variations are important to consider).	Consists of sub-annual readings (such as continuous monitoring) for parameters that have intra-annual variation, or annual readings for parameters that do not.	Consists of at least annual readings for most of the spatial range, with generally comparable timing every year if the parameter has intra-annual variation. Or, for parameters that change slowly (such as land cover), data are available at regular intervals of five years or less.	Consists of data that are collected at infrequent or irregular intervals, or at regular intervals that are more than five years apart.

Criterion	High (2 points)	Medium (1 point)	Low (0 points)
6. Consistency of methods: The ideal indicator will be based on data collection and analytical methods that are comparable across time and space. If data were collected or analyzed using multiple methods, the data visualization should distinguish between the different sources.	Obtained with methods and equipment that are consistent across time and space, allowing for any two data points making up the metric to be completely comparable with minimal caveats.	Obtained with methods and equipment that are generally comparable over time and space, or with a limited number of different methods (e.g., two or three) that might need to be displayed with separate lines on a graph.	Obtained with inconsistent methods and equipment.
7. Peer-review validation: The ideal indicator will use data from a peer-reviewed publication, a program that uses peer-reviewed methods to collect and analyze data, and/or a program whose data have been used and validated in peer-reviewed publications.	For scientific measurements only (i.e., administrative metrics exempt): utilizes methods or approaches that have been peer-reviewed as applied in the Chesapeake Bay.	For scientific measurements only (i.e., administrative metrics exempt): based on methods or approaches that have been previously peer-reviewed in other contexts but not quite as applied to the Chesapeake; or based on methods or approaches that have not themselves been peer-reviewed, but whose results have appeared in peer-reviewed studies.	For scientific measurements only (i.e., administrative metrics exempt): Based on methods or approaches that have not previously been peer-reviewed or that have not been used to support peer-reviewed studies.
8. Uncertainty: The ideal indicator will have low uncertainty—for example, small error bars or narrow confidence intervals.	Consists of highly accurate readings with minimal measurement error, widespread data collection with a well-designed survey that minimizes sampling error, and minimal other sources of error, such as reporting error. Error bars or confidence intervals are small and the "noise" does not obscure the "signal" (if a trend is present, it is discernable).	Consists of readings of acceptable accuracy, with "noise" that would not hide a modest trend, if such a trend were present.	At least one type of error (e.g., measurement or sampling error) is sizable, and/or error bars or confidence intervals are so large that they make it impossible to discern a meaningful trend.

Criterion	High (2 points)	Medium (1 point)	Low (0 points)
9. Other limitations: The ideal indicator will have few confounding factors or other limitations that make it difficult to interpret the data or draw conclusions.	The indicator largely reflects the influence of climate change or actions taken to address climate change. Non-climatic influences are minimal by comparison. No other notable limitations are present that hamper interpretation. *Ignore limitations related to temporal or spatial considerations, uncertainty, or other attributes already captured by other criteria in this matrix.*	The indicator reflects sizable influences from climate change and actions taken to address climate change, along with sizable contributions from non-climatic influences; and/or modest limitations are present that somewhat limit interpretation. *Ignore limitations related to temporal or spatial considerations, uncertainty, or other attributes already captured by other criteria in this matrix.*	Non-climatic influences are likely larger than the influence of climate change and actions to address climate change; and/or substantial limitations are present that hamper interpretation. *Ignore limitations related to temporal or spatial considerations, uncertainty, or other attributes already captured by other criteria in this matrix.*
10. Understandability: The ideal indicator will provide a clear depiction of observations that can be understood by both technical and non-technical users.	Presents a simple concept that is easily understood by a non-scientist reader with little explanation or scientific theory needed.	Presents a concept that requires some degree of explanation, but could generally be understood (with effort) by a non-scientist reader.	Presents a complex concept that is too abstract or meaningless to a non-scientist reader without significant explanation, and that cannot be reduced to more familiar units or frames of reference without sacrificing scientific integrity.