ITAT: Cluster Analysis Code Development

1. Brief description of project, including key tasks
    a. The Integrated Trends Analysis Team (ITAT) requires assistance to refine code used in running the cluster analysis of our tidal trends water quality trends. Cluster analysis is a procedure that groups a set of items according to some measure of similarity between the items. In recent years the Chesapeake Bay Program has developed the *baytrends* package for assessing trends at stations in the fixed station network. This package uses generalized additive models (GAMs) to estimate curvilinear trend lines for a water quality time series (Figure 1.). See Murphy et al. 2019 for details on the GAM methods used in the *baytrends* package. Curvilinear trend estimation can reveal new and interesting information by capturing short term trend events due to meteorology, BMP implementation or other forcing functions, but this new information comes at a cost. Because the curvilinear trend lines are a more complex description of the trend, it is more difficult to discern which stations are exhibiting similar trends. Grouping stations with similar trends is important for assessing whether BMPs are influential over broad geographical areas. It is also important for differentiating regions where progress is satisfactory from regions where more attention to reverse degrading trends is needed. To address this important task of grouping trends, CBP developed methods based on cluster analysis. Base code has been developed for the cluster analysis, but it is in a format that only the contractor (Elgin Perry) can use. Elgin's time is focused this coming year and next mainly on the 4D interpolator and may retire prior to completing the code, so this need is pressing to continue generating this informative analysis.

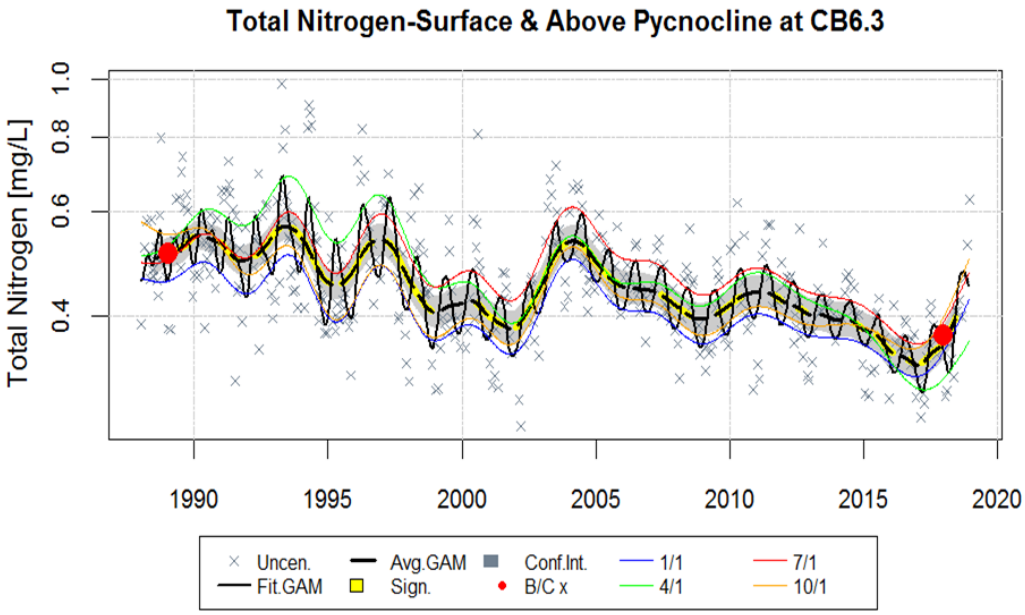

Figure 1. Illustration of curvilinear trend line fitted by baytrends.

2. Targeted audience / user base
    a. The need is to develop the code in a format that a technical analyst like Breck Sullivan, Kaylyn Gootman, or Rebecca Murphy may use. The results from the code are included in ITAT's tributary summaries which are continually updated with the most up to date water quality data.
3. GIT priorities that will be addressed through project funding and implementation
    a. One of STAR's functions is to explain ecosystem condition and change by enhancing the understanding of spatial and temporal patterns. Cluster analysis is one approach supported by researchers and analysists from various governmental, academic, non-profit, and private organizations to conduct analysis and synthesis of the science on changes in water quality. The results of cluster analysis are included in communication materials such as the ITAT Tributary Summaries that are updated frequently and are used to show change over time in a specific tributary and the factors influencing those changes (Figure 2. and Figure 3.). The Tributary Summaries support the progress of reporting annual results as part of our Water Quality Standards Attainment and Monitoring (WQSAM) Outcome, "Continually improve our capacity to monitor and assess the effects of the management actions being taken to implement the Chesapeake Bay Total Maximum Daily Load (Bay TMDL) and improve water quality. Use monitoring results to report annual progress being made in attaining water quality standards and trends in reducing nutrient and sediment in the watershed." Project funding will also support a CBP science need under the WQSAM Outcome with information on it available here.

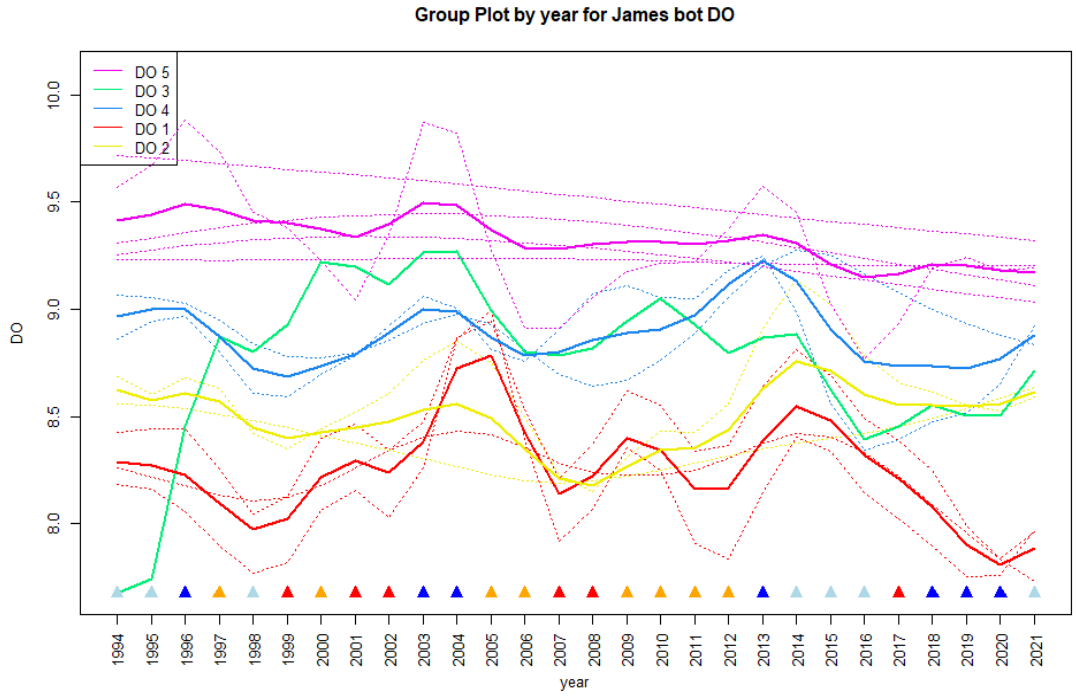

Figure 2. Year means plotted with station groups segregated by color. Multiple dash line traces within group show variability among station within groups.

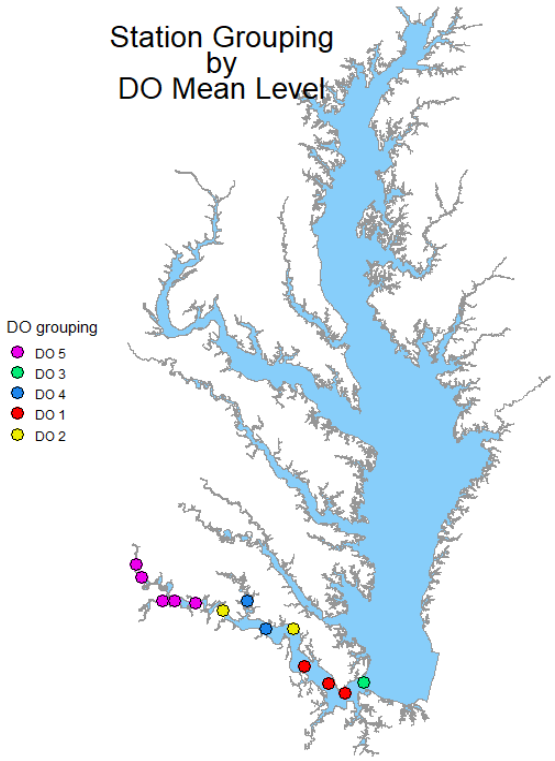

Figure 2: Cluster Map showing station groups by DO levels for James.

4. Identification of (any) cross-GIT application(s) and/or support
    a. The code development would not have any cross-GIT applications, but the product the results are included in, the Tributary Summaries, is a cross-GIT product by including findings from the Land Use Workgroup and Submerged Aquatic Vegetation Workgroup.
5. Intended results
    a. Completed code for ITAT members to use for producing results for communication products. The final product will be similar to the R package developed for baytrends which allows the user to input their own data and produce results and graphs. The initial results will be for current ITAT members to use, but future funded work could lead to an interactive website for users to view results based on the Chesapeake Bay tidal trends similar to baytrendsmap.
6. Projected budget
    a. $25,000 - $30,000
7. What other funding sources have been pursued for the project
    a. Jon Harcum and Erik Leppo were working to complete this, but their time and funding has now been dedicated to the 4D Interpolator work. They would need dedicated funding strictly for refining and completing the code for the cluster analysis, or we need other contractor support to complete it.