Technical Report Number 06-6
Department of Statistics
Virginia Polytechnic Institute and State University

Clustering Monitoring Stations Based on Two Rank-Based Criteria of Similarity of Temporal Profiles

David Farrar and Eric Smith

September, 2006

Abstract. To support evaluation of water quality trends, a water quality variable may be measured at a series of points in time, at multiple stations. Summarization of such data and detection of spatiotemporal patterns may benefit from the application of multivariate methods. We propose hierarchical cluster analysis methods that group stations according to similarities among temporal profiles, relying on standard clustering algorithms combined with two proposed, rank-based criteria of similarity. An approach complementary to standard environmental trend evaluation relies on the incremental sum of squares clustering algorithm and a criterion of similarity related to a standard test for trend heterogeneity. Relevance to the context of trend evaluation is enhanced by transforming dendrogram edge lengths to reflect cluster homogeneity according to a standard test. However, the standard homogeneity criterion may not be sensitive to patterns with possible practical significance, such as region-specific reversal in the sign of a trend. We introduce a second criterion, which is based on concordance of changes in the water quality variable between pairs of stations from one measurement time to the next, that may be sensitive to a wider range of patterns. Our suggested criteria are illustrated and compared based on application to measurements of dissolved oxygen in the James River of Virginia, USA. Results have limited similarity between the two methods, but agree in identifying a cluster associated with a locality that is characterized by pronounced negative trends at multiple stations.

Keywords. Cluster analysis, environmental monitoring, environmental trends, James River, nonparametric procedure, Ward's method.

Acknowledgements. Funding for this research was received from the U.S. EPA via a Science to Achieve Results (STAR) Grant RD 83136801-0. The research has not been subjected to any USEPA review and so no USEPA endorsement should be inferred. We gratefully acknowledge Carl Zipper and Jason Hill for assistance in accessing and interpreting water quality data.

CONTENTS

Introduction 4
1. Cluster Analysis Methodology 5
1.1 Hierarchical Clustering of Monitoring Stations Based on Similarities in Temporal Profiles 5
1.2 A Criterion of Trend Similarity Related to a Standard Trend Homogeneity Test 7
1.3 Clustering by Incremental Sum of Squares with Transformed Plotting Heights 9
1.4 An Approach Based on Concordant and Discordant Changes over Time 10
2. An Empirical Comparison of the Two Criteria 12
2.1 Measurements of Dissolved Oxygen in the James River of Virginia 12
2.2 Results 13
3. Discussion 16
References 19
FIGURES 20
Appendix 1. Station codes and geographic coordinates for 71 stations used in the analysis. 34
Appendix 2. Updated R function library 36

INTRODUCTION

An important use of water quality data is to evaluate trends over time (Gilbert, 1987; Helsel and Hirsch, 1992; Millard and Neerchal, 2000). Information on temporal trends may be useful in formulating management strategies to maintain and improve water quality. In practice, statistical trend evaluation may involve separate analyses for individual monitoring stations. However, interpreting the volume of statistical results from a station-by-station analysis with multiple water quality endpoints may be challenging. Multivariate and graphical techniques may help us to recognize patterns in data representing multiple stations. Regional patterns in particular may be useful in formulation of region-specific water quality management strategies.

Here, we will explore use of hierarchical cluster analysis (HCA), briefly reviewed in Section 1.1, to identify groups of stations with similar temporal profiles, for a water quality endpoint measured at repeated points over time. In choosing a specific HCA procedure, important decisions include the criterion of similarity among the temporal profiles, and an algorithm that can use the chosen criterion to identify groups of stations. (We will sometimes use the term *similarity* for brevity, while recognizing that some calculations actually require a measure of *dissimilarity*.) We focus primarily on the problem of quantifying similarity among temporal profiles, relying on standard hierarchical cluster analysis software to perform clustering. Different clustering algorithms have been used, depending on the criterion of similarity among temporal profiles.

We compare two rank-based criteria of similarity. The first, described in Section 1.2, is related to a standard test for trend homogeneity (van Belle and Hughes, 1984). Data for a station are reduced to a scalar summary statistic, and clustering may be based on differences in values of the statistic between pairs of stations. The effect is that stations are grouped according to the strength and sign of trend. HCA based on this criterion may complement standard trend evaluations.

However, reduction of the data for a station to a scalar summary may result in lower sensitivity to patterns with some practical importance, such as a reversal in trend in a particular a region. Therefore in Section 1.3 we propose an alternative criterion based on concordance of changes in the water quality endpoint between pairs of stations, from one measurement time to the next.

The volume of cluster analysis literature is large and reflects applications in diverse disciplines. Summaries are provided by general texts such as Gordon (1999), Romesburg (1984), Seber (1984), and Venables and Ripley (1994). To avoid repetitious citation, we will simply note that where we do not give a specific reference for some aspect of cluster analysis, more extensive exposition is provided by Seber (1984) and other general treatments.

1. METHODS

1.1 Hierarchical Clustering of Monitoring Stations Based on Similarities in Temporal Profiles

Our objective will be to group monitoring stations based on similarities in temporal profiles, when a water quality endpoint has been measured at a series of points in time at each station. Based on data comprising *n* measurement stations, each measured at *T* times, we may let x_{it} denote the *t*th measurement for the *i*th station. In practice -- as in applications presented here -- measurements may be missing for some measurement times, for some stations.

Several multivariate techniques are available for evaluation of information in the form of an index of pairwise similarity or dissimilarity among objects (Seber, 1984). Similarity or dissimilarity values comparing *n* objects are conveniently arrayed in an $n \times n$ matrix with d_{ij} in the *i*th row and *j*th column. Reasonable requirements for dissimilarity criteria are $d_{ij} \geq 0$, $d_{ii} = 0$, and $d_{ij} = d_{ji}$ (Seber, 1984; Venables and Ripley, 1998).

For our purposes it is useful to evaluate dissimilarity information using what may be the most transparent approach, namely agglomerative HCA. The results from such a procedure are displayed graphically as a *dendrogram*, such as those displayed in many of our figures. The agglomerative approach is iterative, starting with each object viewed as a separate cluster, and at each iteration joining two clusters, carried forward from previous iterations, until objects are joined in a single cluster. In the first iteration a cluster is formed by joining two objects that are most similar according to the chosen criterion. Some subsequent iterations require choice of a *linkage criterion*, which can be used to quantify the similarity of clusters that may contain multiple objects. A typical hierarchical cluster analysis program will provide multiple options. For example, the R library function `hclust` Version 2.3, used for our examples, provides seven linkage criteria. For our examples we have relied on the *average linkage approach* and *Wards method*, with the choice depending on the criterion of temporal profile similarity.

Once a dendrogram has been created using an HCA procedure, clusters can be identified based on a desired number of clusters, or based on a specified dissimilarity. An important issue is specification of the number of clusters supported by the data. Clusters will tend to be more homogeneous internally if more are extracted, but extracting more clusters will not necessarily result in a more useful classification. As an alternative to specifying the number of clusters one may seek to identify individual clusters that are in some sense well supported, without necessarily committing to an exhaustive classification.

The procedures that we discuss have been programmed in R (R Core Development Team, 2005). In addition to using `hclust` to generate dendrograms, we use `cutree` to extract clusters from a dendrogram object. A library of R functions is included as an appendix.

1.2 A Criterion of Trend Similarity Related to a Standard Trend Homogeneity Test

Van Belle and Hughes (1984) introduced a chi-square statistic for testing homogeneity of environmental trends. Their objective of trend homogeneity testing is evidently closely related to our objective of using cluster analysis to find homogeneous groups of stations. A rank-based statistic Z is computed for each station. When the objective is to test the trend for one station, a *p*-value is computed by referring Z to a standard normal distribution. The homogeneity of trend test combines the Z statistics representing n stations, $Z_1, \ldots, Z_n$ say.

It is useful to summarize computation of the Z statistic for a single station. The statistic is related to Kendall's τ (Kendall, 1962) as used for relating water quality endpoint to time, involving consideration of each pair of measurement times. We suppose that the water quality endpoint is measured at each of T times for a particular station. From the T measurements we compute $Z = (S + \delta) / \sqrt{\mathrm{Var}(S)}$. Here S is computed by considering each of $T(T-1)/2$ distinct pairs of measurement times. Each of these pairs is scored as "tied," "concordant," or "discordant," according as the water quality endpoint is equal in value for both measurement times, increases in value from the first to the second for both, or decreases from the first to the second for both. (The terminology reflects whether or not the sign of change for the water quality endpoint agrees with the sign of change of the time variable.) Ties are possible because of limited measurement precision, or because non-detects are present in the data and counted as ties. S is computed by subtracting the number of discordant pairs from the number of concordant pairs. The value of δ is –1, 0, or 1 according as S is positive, equal to 0, or negative. $\mathrm{Var}(S)$ is an estimate of sampling variance under an assumption of no trend, in our applications incorporating a standard adjustment for tied measurements (Gilbert, 1987, Expression 17.2).

The statistics for *n* stations, $Z_1, \ldots, Z_n$ say, may be used in a test of homogeneity of trend. We compute test statistic $\chi^2_H = \sum_{i=1}^{n} \left(Z_i - \bar{Z} \right)^2$ where $\bar{Z}$ is the average of $Z_1, \ldots, Z_n$. To compute a *p*-value, for testing a null hypothesis of no trend heterogeneity, χ^2_H is referred to a chi-square distribution with $n - 1$ degrees of freedom. A statistically significant result is taken as evidence for trends differing in sign or magnitude among stations. For our purposes, it is good to observe that χ^2_H is a sum of squared deviations from a mean -- a corrected sum of squares or SS_c -- considering that some cluster analysis procedures are designed to minimize SS_c.

Evidently, the approach may be extended to test homogeneity of groups of stations, allowing heterogeneity among groups (*e.g.*, Table 1). A chi-square value is obtained by computing χ^2_H for each group, as for testing homogeneity of the group, and summing across groups. The *p*-value is computed by referring the statistic to a chi-square distribution with $n - k$ degrees of freedom, the sum of degrees of freedom over groups.

Here our objective is to use Z statistics to group monitoring stations. χ^2_H provides a criterion that can be used to identify homogeneous groups. The dissimilarity of the *i*th and *j*th stations can be taken to be the squared difference of Z scores $\left(Z_i - Z_j \right)^2$. We will denote this dissimilarity criterion ZDiffs^2 (for Z differences squared). Some justification is provided in the next section. A familiar statistical manipulation relates the sum of ZDiffs^2, over pairs of stations, to the test statistic χ^2_H:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(Z_i - Z_j \right)^2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(Z_i - \bar{Z} + \bar{Z} - Z_j \right)^2 = n \times \chi^2_H .$$

ZDiffs^2 can be computed for a pair of stations with no measurement times in common. In case of missing measurements, it can happen that there are few measurement times in common for some stations. It seems sensible to incorporate, among criteria for selection of stations, some criterion based on a minimum number of common measurement times.

1.3 Clustering by Incremental Sum of Squares with Transformed Plotting Heights

To perform clustering based on the ZDiffs^2 criterion, we use the incremental sum of squares method (also known as Ward's method) because of a straightforward relationship to our objective of finding groups of stations homogeneous according to χ^2_H. Let $\chi^2_H\ (C)$ denote the test statistic for trend homogeneity, computed using stations in a set C, but with a value of zero if C contains only one station. The increment from merging two clusters C_i and C_j is

$$\Delta\chi^2_{\text{H}}\left(C_i, C_j\right) = \chi^2_{\text{H}}\left(C_i \cup C_j\right) - \chi^2_{\text{H}}\left(C_i\right) - \chi^2_{\text{H}}\left(C_j\right).$$

According to Ward's method, as expressed for our context, the pair of clusters merged at a given step is such as to minimize such an increment. It is clear that the effect at a given iteration is also to minimize the chi-square statistic, described in the previous section, for a test of simultaneous trend homogeneity.

In a default dendrogram based on Ward's method (*e.g.*, Figure 3), dendrogram branch-points are plotted at twice the χ^2_H increment (Seber, 1984). In the present context the effect of the factor of 2 is that when a cluster includes only two stations, the height plotted equals ZDiffs^2. We find it useful to modify the dendrogram by substituting plotting heights more closely related to standard trend homogeneity computations (e.g., Figure 4). We plot Cluster *C* at height $\sqrt{\chi^2_{\text{H}}\ (C) \,/\, \left(n_C - 1\right)}$ where $\chi^2_{\text{H}}\ (C)$ is the chi-square statistic and n_C the number of stations for *C*. We observe that for the test of homogeneity of C, the expected value of the chi-square distribution equals its degree of freedom $n_C - 1$. Therefore values of our modified plotting height greater than 1 are larger than expected under an assumption of homogeneity. (Of course, such a distributional assumption does not take into account that our groups are defined so as to minimize chi-square increments.) We apply the square-root transformation in view of the skewness of the ZDiffs^2 distribution over pairs of stations, which without our transformation results in very short relative heights for the first clusters joined.

In programming such a modification of a dendrogram, it is convenient to work in a programming environment where a dendrogram object can be generated using a library function and then modified, and the modified dendrogram plotted or otherwise evaluated. This allows modification of selected components of the dendrogram specification, such as plotting heights in our approach, while other components are handled by library routines. For some manipulations of dendrograms, it is convenient to rely on recursive functions. For example, the count of stations for a cluster is the sum of counts for member clusters. Similarly, cluster χ^2_H values are amenable to recursive computation.

1.4 An Approach Based on Concordant and Discordant Changes over Time

For a second criterion of similarity among temporal trends, which we think may be sensitive to a wider array of temporal profile similarities, we propose the Rank Temporal Profile Similarity Index (RTepsi). For a given pair of stations, the value of the index is based on concordance between stations, in temporal changes in the water quality endpoint.

It is helpful in this context to adopt a modified notation. Suppose that the comparison of stations *i* and *j* is based on T_{ij} measurement times with values for both stations, $x_{i1}, \ldots,$ $x_{iT_{ij}}$ for station *i*, and $x_{j1}, \ldots, x_{jT_{ij}}$ for station *j*. In the sequel, for brevity, we will let T denote the number of time-points, with the understanding that in case of missing measurements at some times, this may actually depend on the stations compared.

We will say that water quality endpoint changes for the two stations are "concordant" for a pair of measurement times when the measured value increases from the first time to the second for both stations, decreases for both, or is unchanged for both. A pair of measurement times will be called "discordant" for two stations if not concordant. Ties in particular are counted as concordances. Then our measure of similarity for a pair of

stations is the fraction of pairs concordant, out of the $T\ (T\ -\ 1)\ /\ 2$ pairs with measurements for the two stations. Our index can be expressed formally as

$$\text{RTepsi}_{ij}\ =\ \frac{1}{T\ (T\ -\ 1)\ /\ 2}\ \sum_{k=1}^{T-1}\ \sum_{l=k+1}^{T}\ \Big[\ I\ \Big(\big(x_{il}\ -\ x_{ik}\big)\ \big(x_{jl}\ -\ x_{jk}\big)\ >\ 0\Big)\ +\ I\ \big(x_{il}\ =\ x_{ik}\big)\ I\ \big(x_{jl}\ =\ x_{jk}\big)\Big]$$

where the indicator function $I\ (c)$ equals 1 or 0 according as condition *c* does or does not hold.

Where a criterion of dissimilarity is required -- rather than a criterion of similarity -- we subtract RTepsi values from one. The result can be described as the fraction of pairs of measurement times that are discordant, relative to the number of pairs with data available.

In our implementation, computation of RTepsi for a pair of stations is subject to a minimum count of years with measurements for both stations. To impose this minimum count we first form a matrix with RTepsi values, including all stations with some measurements available. For any pair of stations with too few times measured for both, the missing value code is entered in the appropriate cell of the matrix. Stations are then deleted one at a time until the matrix contains no missing values, at each step deleting the station associated with the largest number of values missing.

To perform hierarchical cluster analysis, we use the average linkage criterion, which is relatively conventional and easily explained. According to that approach, the dissimilarity for a pair of clusters is the average of dissimilarities, over pairs of objects with one member of the pair belonging to each of the clusters compared.

2. AN EMPIRICAL COMPARISON OF THE TWO CRITERIA

2.1 Measurements of Dissolved Oxygen in the James River of Virginia

The Virginia Department of Environmental Quality (VADEQ) samples water quality from streams and rivers in Virginia (USA), for evaluation of water quality status and trends. We report comparisons of the two criteria based on measurements of dissolved oxygen (DO, $mg \bullet L^{-1}$) from 71 stations on the James River and its tributaries, that met criteria for computation of both measures of similarity. The locations of the 71 stations are displayed in Figure 1.

Most stations had multiple measurements in some years. Before computing similarity criteria, the data were reduced to annual median values, so that each combination of station year is represented by a single value at most. The data used are limited to the 15-year series ending in 2004. Stations were included such that there were 5 or more years with measurements for both stations in any pair included, within the 15-year series.

Station labels used in our graphical displays (e.g., Figure 1) incorporate a 3-letter stream code from the VADEQ data, and a numeric index representing order of distance from the stream mouth. One stream (JMS) represents stations actually located on the James River, accounting for 20 stations. Other streams correspond to tributaries, which are represented very unequally. There are 11 stations for the Pagan River (PGN), 2-6 for 5 other tributaries, and a single station each for 22 tributaries.

Appendix 1 lists the 71 stations contributing data to our analysis, according to station codes used by VADEQ. The appendix shows the correspondence between our station codes and the longer codes used by VADEQ, which incorporate distance in miles from stream mouth. The appendix allows retrieval of detailed information for any station of interest from the VADEQ (2006) web site.

2.2 Results

An exploration of regional patterns in temporal profile information may naturally include separate graphical analyses for pre-defined subsets of the data, such as river basins or physiographic provinces. Figure 2 displays temporal profiles for annual medians for the Pagan River and James River, which together account for almost half of the stations used in our analyses. These results may be compared to subsequent graphs for clusters based on our procedures. A feature that is important in subsequent analyses is an apparent negative trend for stations associated with the Pagan River.

Our initial analysis involved clustering the set of 71 stations. As a result of observing interesting patterns involving the Pagan River and Jones Creek stations, additional analyses focused on stations for those streams.

Dendrograms based on our two criteria are displayed in Figures 3 and 4. In Figure 4 cluster indices have been added to station labels, for 4 clusters identified from the dendrograms, using the `cutree` function. Cluster indices appearing on the ZDiff^2 dendrogram are default indices generated by `cutree`. Cluster indices appearing on the RTepsi dendrogram were chosen so as to maximize the number of stations with the same cluster index in both dendrograms, so that where possible similar clusters receive the same index in both dendrograms.

A feature in common for the two dendrograms is a cluster with a relatively large number of stations from the Pagan River. Apart from such a cluster the dendrograms do not display conspicuous similarities. The two criteria are not strongly correlated across pairs of stations (Figure 5). Table 1 displays a cross-classification of stations according to the two criteria, after extracting 4 clusters.

Figure 6 displays hierarchical cluster analysis results obtained using only the Pagan River and Jones Creek stations. Special mention seems appropriate for results for two particular stations. The single station for James Creek (JOG-1 or 2-JOG000.62), located at the

confluence with the Pagan River, clusters relatively closely with Pagan River stations. Among Pagan River stations, the most distinctive profile is apparently associated with the station furthest upstream (PGN-11 or 2-PGN010.07). Some subsequent figures focus on these particular stations.

Figure 7 compares the profiles for two stations of particular interest (JOG-1, PGN-11) to the combined set of Pagan River and Jones Creek profiles. The profile for PGN-11 is seen to be relatively flat, compared to other Pagan River profiles. An apparent difference between dendrograms generated according to the two criteria is that using the RTepsi criterion there is a greater tendency for adjacent stations to cluster together. (Again, note that stations are numbered from the mouth towards headwaters.)

In selecting the number of clusters supported by the data, one may consider the results from the pooled homogeneity test, with pooling over the clusters. In Table 2 we display some results from the pooled test, with different numbers of clusters. The conventional test of van Belle and Hughes, with 70 degrees of freedom, corresponds to the case of a single cluster $(k = 1)$. The statistically significant result ($p \leq 0.01$) provides support for an effort to identify patterns. Statistical significance disappears with two or more clusters based on the ZDiff^2 criterion, while with the RTepsi criterion the test is significant with up to 3 clusters (suggesting at least 4 clusters). We expect that, with $k \geq 2$, true false positive rate of the test will be lower than the nominal rate particularly with the ZDiffs^2 criterion, which identifies clusters so as to minimize the test statistic.

Figures 8 and 9 display temporal profiles for clusters extracted from each dendrogram, somewhat arbitrarily assuming 4 clusters. Figure 10 displays profiles for the Pagan River and Jones Creek along with a cluster identified using each method, with a majority of stations from the Pagan River. Coordinates of stations assigned to the 4 clusters are mapped in Figure 11. The clusters appear to overlap broadly, displaying little indication of regional patterns.

Table 1. Cross-classification of stations according to cluster under two criteria.

		RTepsi Cluster			
		1	2	3	4
$ZDiff^2$ Cluster	1	18	11	2	0
	2	3	8	3	1
	3	1	5	2	8
	4	0	0	0	9

Table 2. Homogeneity of trend tests with pooling over clusters. (Clusters have been identified based on two different hierarchical cluster analysis procedures.)

K	degrees of freedom	$ZDiff^2$ Criterion		RTepsi Criterion	
		pooled chi-square statistic	*p*-value	pooled chi-square statistic	*p*-value
1	70	182.2	<0.01	182.2	<0.01
2	69	45.1	0.99	130.6	<0.01
3	68	21.9	1.00	95.3	0.02
4	67	13.8	1.00	73.5	0.27
5	66	6.0	1.00	64.0	0.55

3. DISCUSSION

The most pronounced pattern that we have detected, reflecting data for multiple stations, is a negative trend of DO for stations associated with the Pagan River and Jones Creek (PGN/JOG). The region that includes these stations has been sampled relatively intensively because of specific water quality concerns. Therefore it can be argued that a specific cluster for these streams is an artifact of high sampling density. (If some other region were sampled with similar intensity, the result could be a cluster representing that region.) Nevertheless, we think the results support that the methods provide efficient recognition of the most important patterns.

Given that the Pagan River and Jones Creek empty into the James River estuary and nearby points, with neither emptying into the other, the correlation between the two streams may reflect the action of tides.

Low DO can be associated with ecological degradation. However, the results for PGN/JOG do not establish an ongoing pattern of degradation. The pattern observed might reflect efforts to control nutrient enrichment, which often has the role of an ecological stressor. For practical reasons, measurements are taken during daytime. In presence of high solar irradiation, nutrient enrichment may enhance photosynthetic activity and lead to high DO. Such an increase in DO may be transient particularly in a warm stream. The negative trends for PGN/JOG may be partly due to early spikes in DO, and the profiles may now be stable (Jason Hill, personal communication). Information on diurnal variation may help to evaluate this interpretation.

Our limited empirical comparison does not seem to support a strong preference between the two criteria of temporal pattern similarity. However, in practice the relatively limited specificity of the $ZDiffs^2$ criterion could be important in some situations. Figure 12 displays a hypothetical example where three stations have identical *Z* statistics, although the temporal profiles differ in ways that could be important, if such a pattern is encountered in practice. In practice, a change in the sign of trend, as for Station A, might

reflect a change in land use or introduction or removal of a source of pollution. In general, we expect that *Z* statistics are not reliable for capturing regional patterns in non-monotone profiles. Non-monotone, regional profiles potentially include local effects of climatic fluctuations, as well as changes in the sign and magnitude of trends. Results from separate analysis with Pagan River and Jones Creek data are consistent with a conclusion that the RTepsi criterion may capture more information in the temporal profiles, relevant for grouping stations.

As is often the case in applications of hierarchical cluster analysis, inspection of the dendrograms in our case does not lead to optimism in the possibility of finding a simple, automated procedure for determining the number of clusters. While a cluster with a large number of Pagan River stations seems well supported, an automated procedure that would identify such a cluster might identify additional clusters that are not as well supported.

Figure 13 suggests a plausible approach for selecting the number of clusters when relying on hierarchical cluster analysis with the RTepsi criterion. The average within-cluster value, viewed as a measure of within-cluster homogeneity, is plotted against the number of clusters for 1-10 clusters. Our average is computed in two stages, first averaging over pairs of stations within each cluster, then averaging the results from the first step, over clusters. Formally, for *k* clusters $C_1, \ldots, C_k$, we compute

$$\frac{1}{k}\sum_{l=1}^{k}\left(\frac{1}{N_{C_l}}\sum_{(i,j)\in C_l}\text{RTepsi}_{ij}\right)$$

where N_{C_l} is the number of stations in C_l. It is no surprise that homogeneity as measured increases with the number of clusters. However, if the curve had showed evidence of a plateau the graph might have been taken to suggest a number of clusters, considering the use of analogous plots in multivariate analysis and statistical modeling. Unfortunately, the approach does not seem to suggest a definite number of clusters in our case. A possible improvement might incorporate a penalty for increasing the number of clusters, perhaps based on an expected increase in averaged similarity.

We have focused on patterns in the data for a single water quality endpoint measured at multiple points in time. Trend evaluation is typically required for multiple measured variables as well as for multiple stations. Therefore we think it is desirable to explore multivariate techniques designed to simultaneously evaluate multiple measurements, particularly rank-based procedures (Lettenmeier, 1976; Rheem, 1992).

Parametric and semiparametric alternatives to a rank-based approach may have the effect of clustering based on estimated profiles that smoothed, relative to the profiles of actual measurements. Such alternatives may include functional data analysis procedures (Henderson, 2006) or incorporation of spatial or temporal autocorrelations.

REFERENCES

Gilbert, R.O. (1987). Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Gordon, A.D. (1999). Classification. (2nd edition) Chapman & Hall/CRC, New York.

Helsel, D.R., and Hirsch, R.M. (1992). Statistical Methods in Water Resources. Elsevier, New York.

Henderson, B. (2006). Exploring between site differences in water quality trends: a functional data analysis approach. Environmetrics 17:65-80.

Kendall, M.G. (1962). Rank Correlation Methods. (3rd edition) Hafner Publishing, New York.

Lettenmaier, D.P. (1976). Detection of trends in water quality data from records with dependent observations. Water Quality Research 12:1037-1046.

Millard, S.P., and Neerchal, N.K. (2000). Environmental Statistics with S-Plus. CRC Press, New York.

R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org`.

Romesburg, H.C. (1984). Cluster Analysis for Researchers. Wadsworth Inc., London.

Seber, G.A.F. (1984). Multivariate Observations. Wiley, New York.

Smith, E.P., Rheem, S., and Holtzman, G.I. (1993). Multivariate assessment of trend in environmental variables. pp. 489-508 in G.P. Patil, C.R. Rao, and N.P. Ross (eds.), Multivariate Environmental Statistics. Elsevier Science, New York.

van Belle, G., and Hughes, J.P. (1984). Nonparametric tests for trend in water quality. Water Quality Research 20:127-136.

Venables, W.N., and Ripley, B.D. (1998). Modern Applied Statistics with S-Plus. (2nd edition). Springer, New York.

Virginia Department of Environmental Quality (VADEQ). (2006). [Web site providing detailed information on monitoring stations.] VA DEQ, `https://www.deq.state.va.us/webapp/`

FIGURES

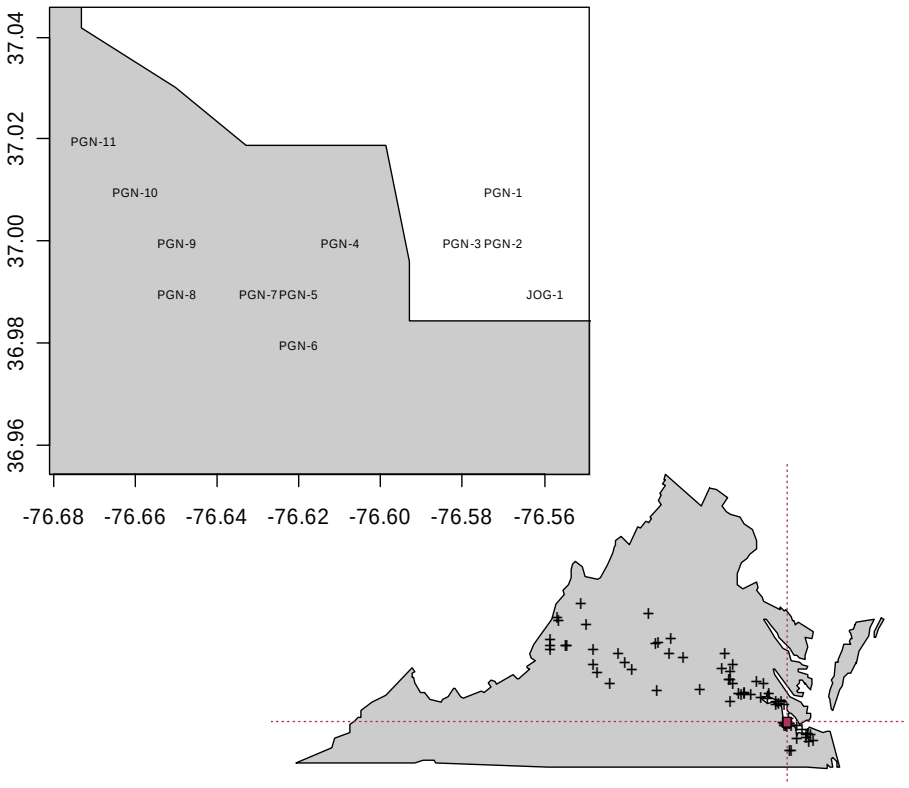

Figure 1 Monitoring stations for the James River basin. A map of Virginia shows the locations of 71 monitoring stations. A rectangular region that appears at the juncture of dotted lines in the Virginia map is enlarged in the upper figure, with axes representing degrees longitude and latitude.

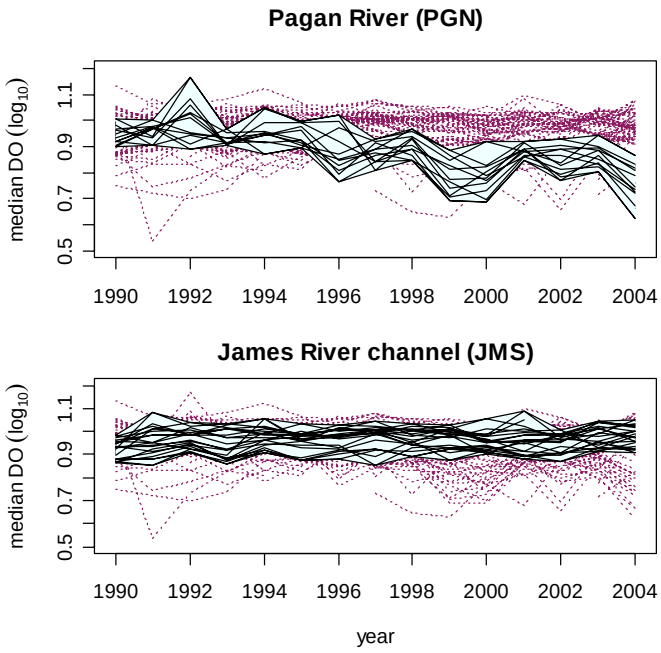

Figure 2 Time series of dissolved oxygen (mg•L-1) for stations in the Pagan River Basin (PGN) and along the James River channel. A shaded "envelope" indicates the range of concentrations for a given year. Series for other stations are displayed for comparison, using dotted lines.

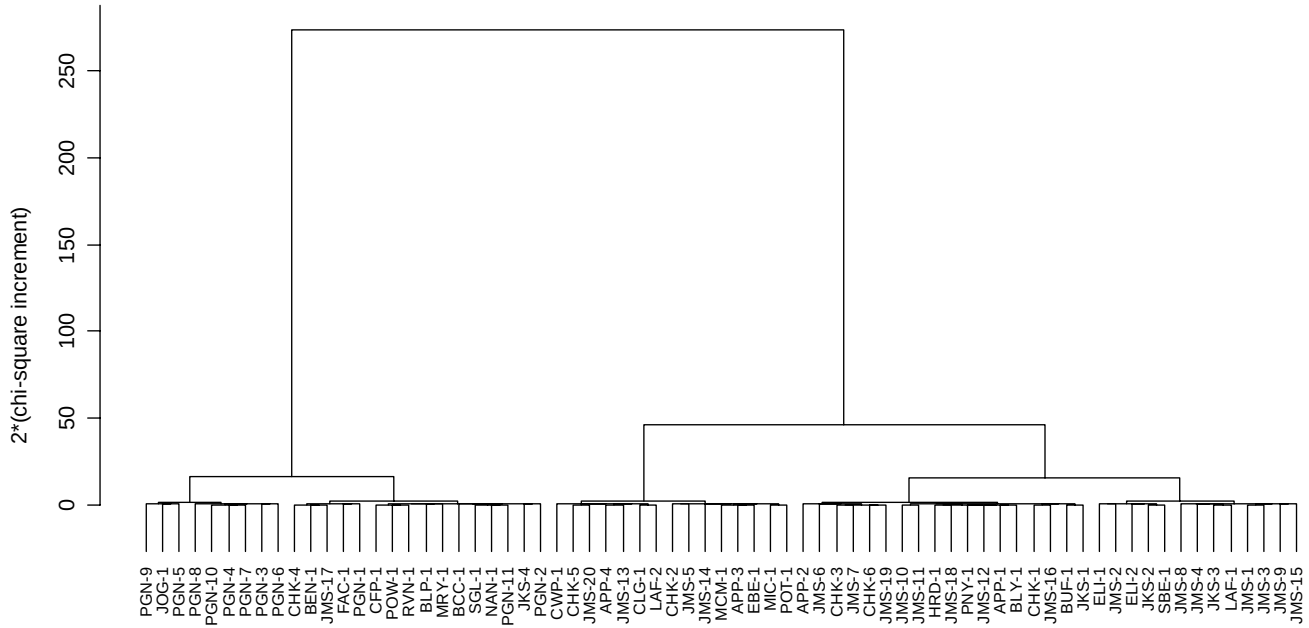

Figure 3 Dendrogram based on the ZDiff^2 dissimilarity criterion, Ward's linkage criterion, and default scaling of edge lengths.

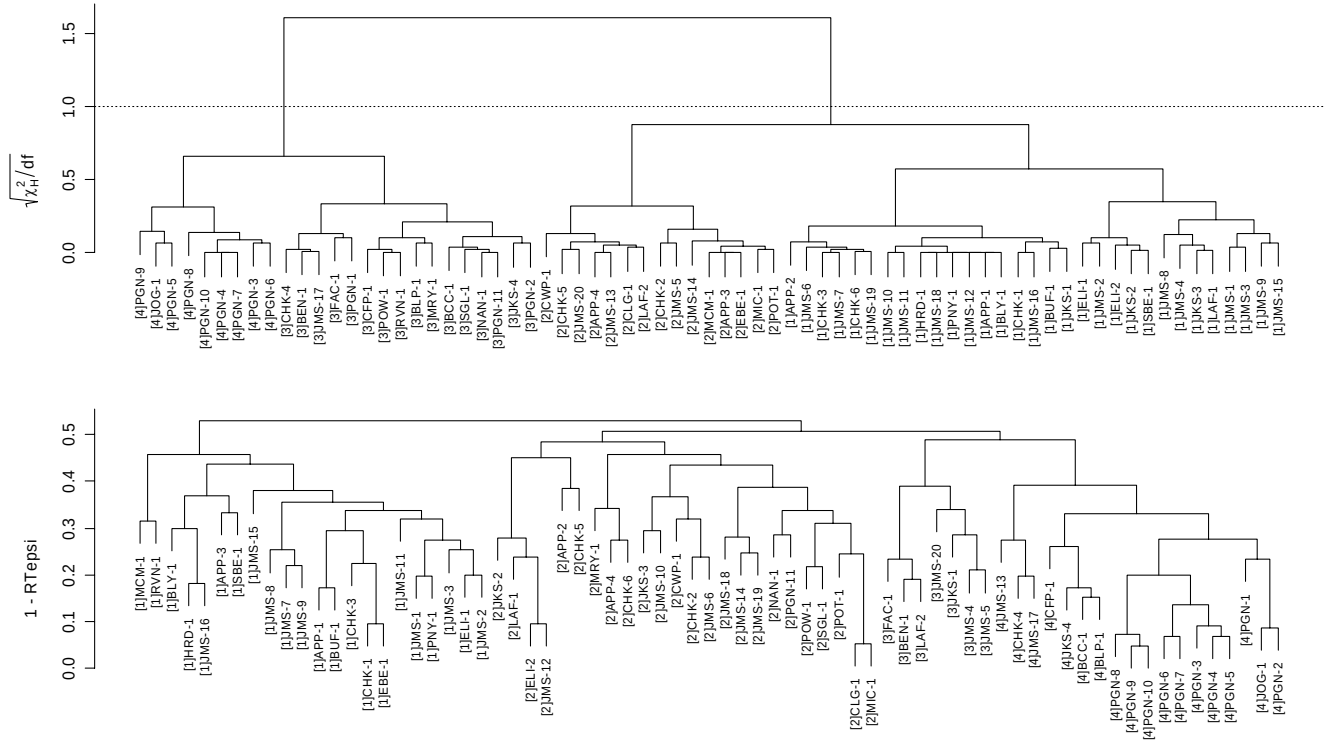

Figure 4 Dendrograms based on two criteria, ZDiff^2 (top) and RTepsi (bottom). For the upper plot the edge lengths have been transformed as described in the text. Bracketed numbers in leaf labels are cluster indices for a 4-cluster solution.

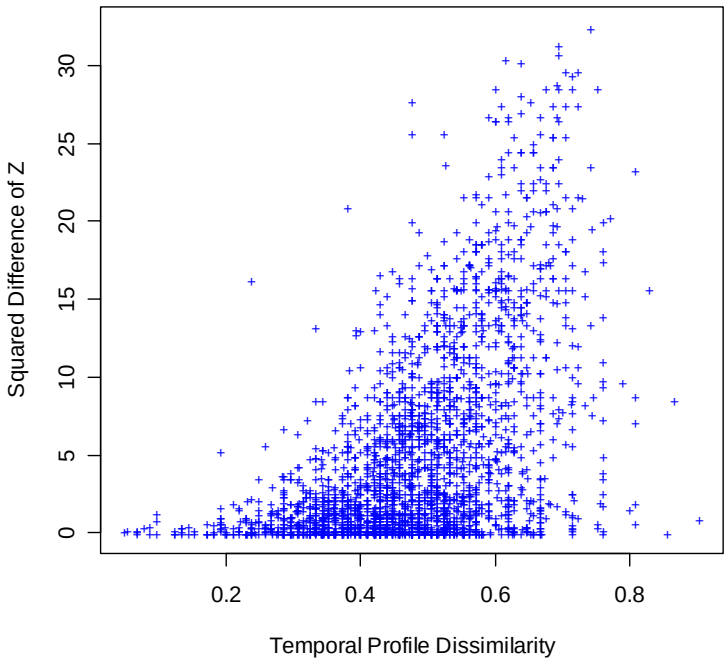

Figure 5 Scatterplot comparing two distances for the 71 stations. Each point plotted corresponds to a different pair of stations. The Spearman rank correlation is 0.50.

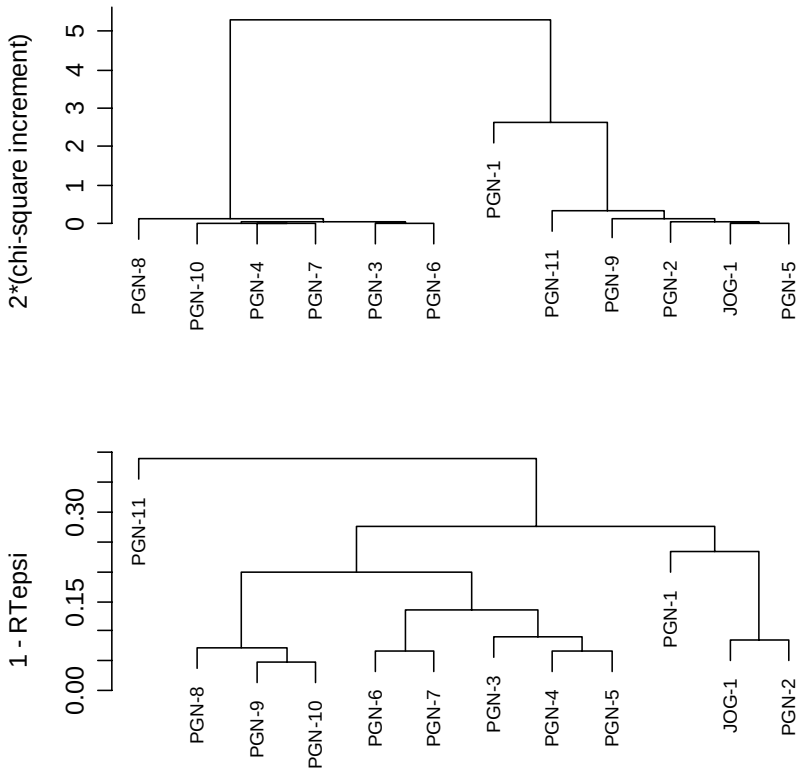

Figure 6 Dendrograms for Pagan River (PGN) and Jones Creek (JOG) stations, based on the two criteria. The upper plot relies on default edge lengths.

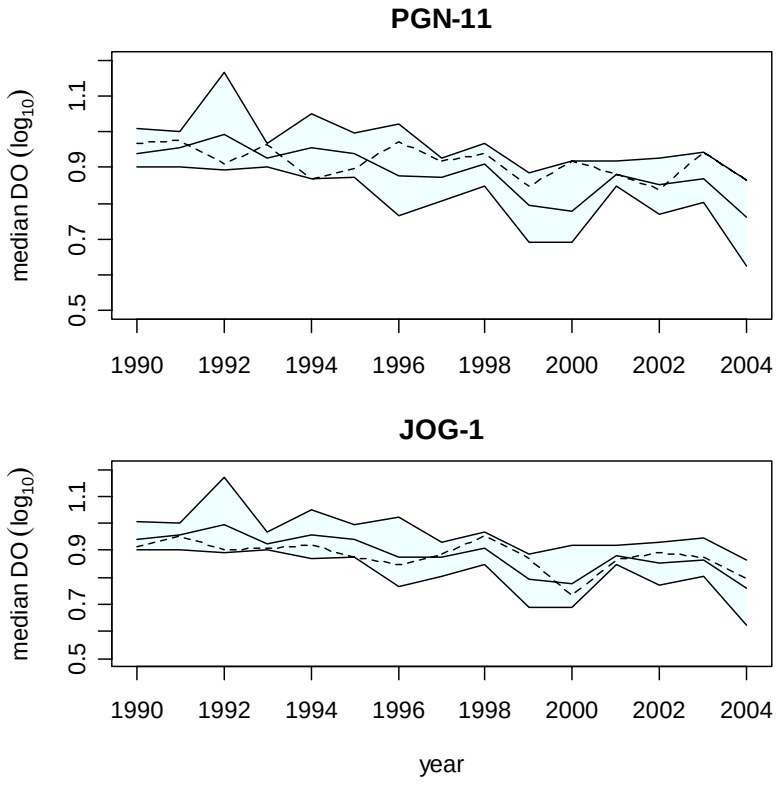

Figure 7 Profiles for stations PGN-11 (upper) and JOG-1 (lower) compared to the combined set of Pagan River and Jones Creek profiles. In each plot the profile for a station of interest is represented by a dotted line. A solid line represents the annual median values for the combined set of Pagan River and Jones Creek profiles.

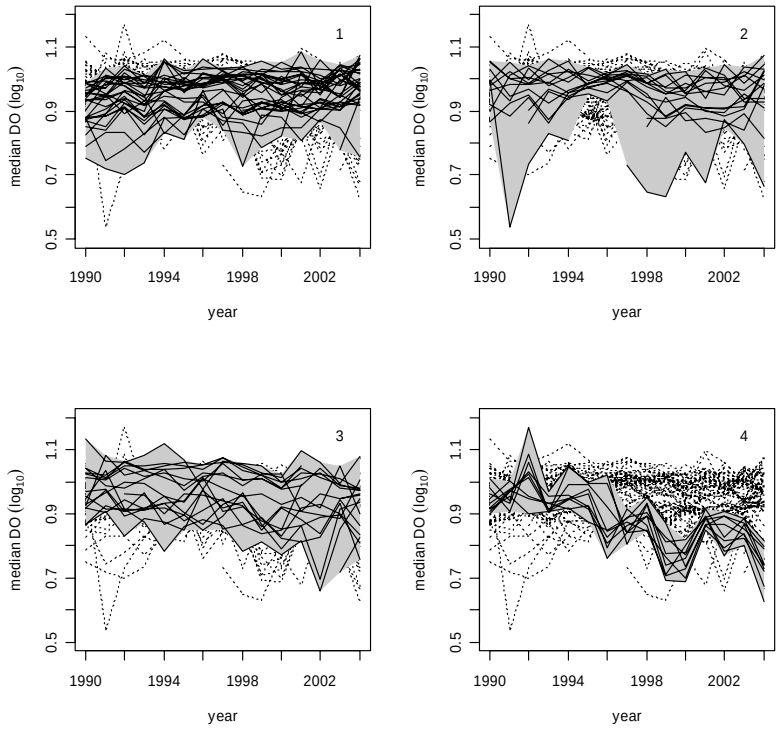

Figure 8 Temporal profiles for individual stations in four clusters identified by clustering based on the ZDiff^2 criterion.

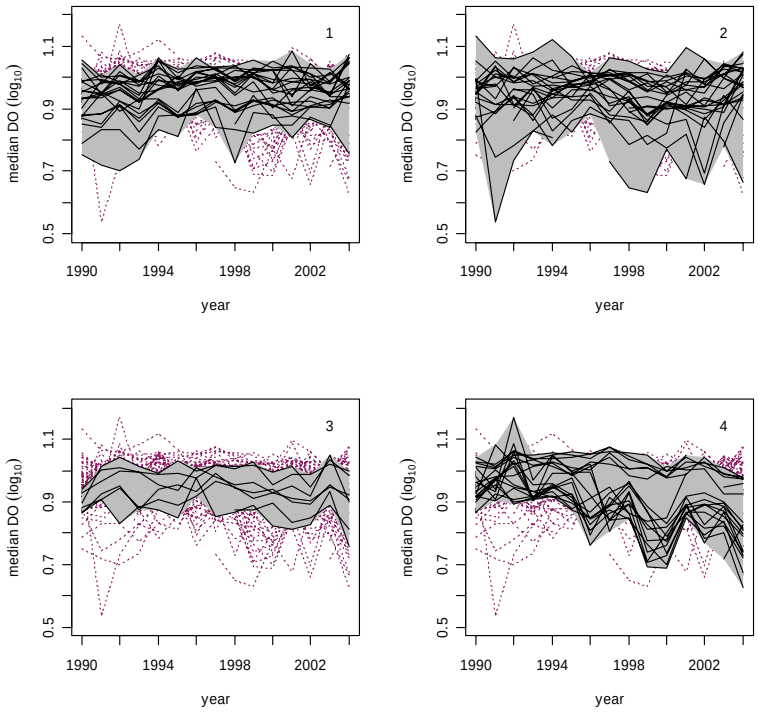

Figure 9 Temporal profiles for individual stations in four clusters identified by clustering based on the RTepsi criterion.

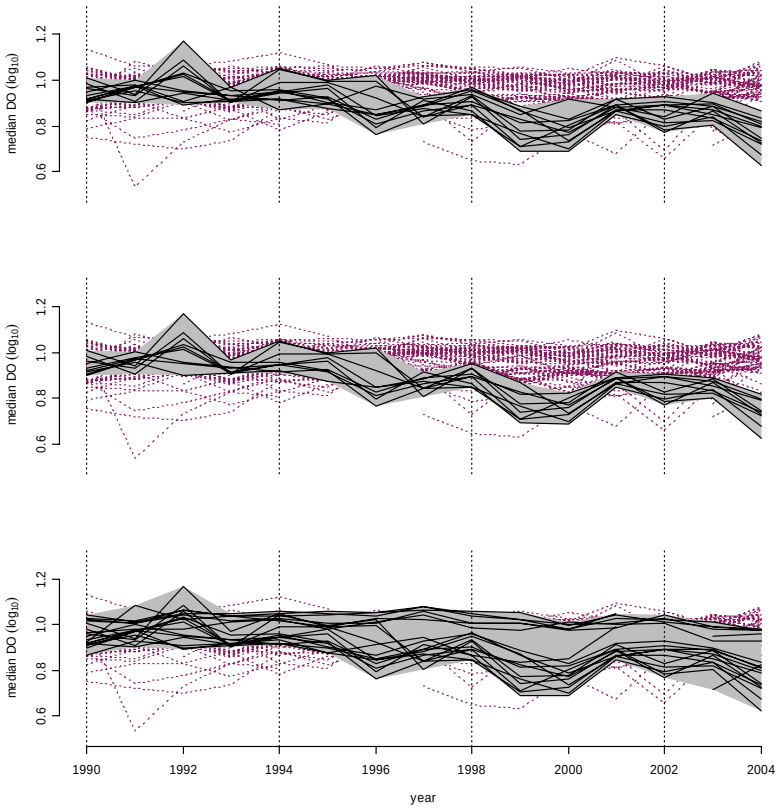

Figure 10 Profiles for the Pagan River and Jones Creek (top) compared to Cluster 4 based on use of the ZDiff^2 criterion (middle) and RTepsi criterion (bottom).

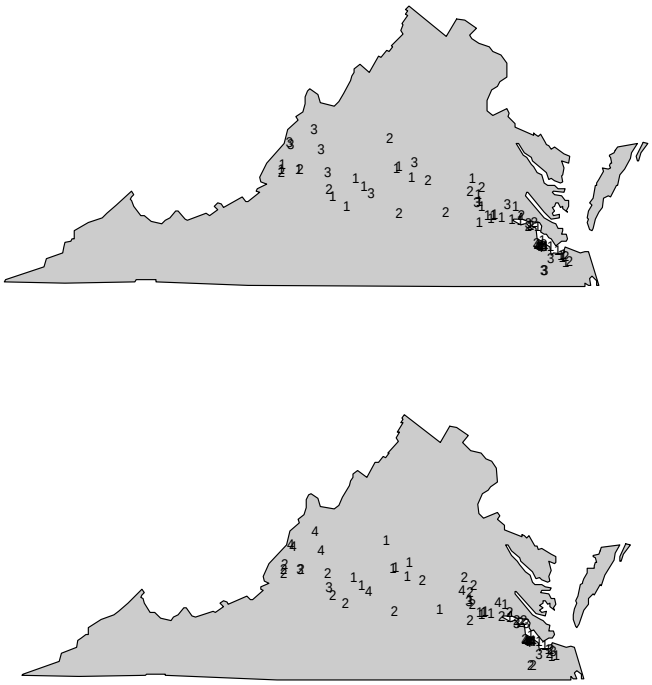

Figure 11 Coordinates of stations in 4 clusters identified by clustering based on the $ZDiff^2$ criterion (top) and the RTepsi criterion (bottom). For each solution Cluster 4 is composed primarily of Pagan River and Jones Creek stations, which are plotted separately in **Figure** 1.

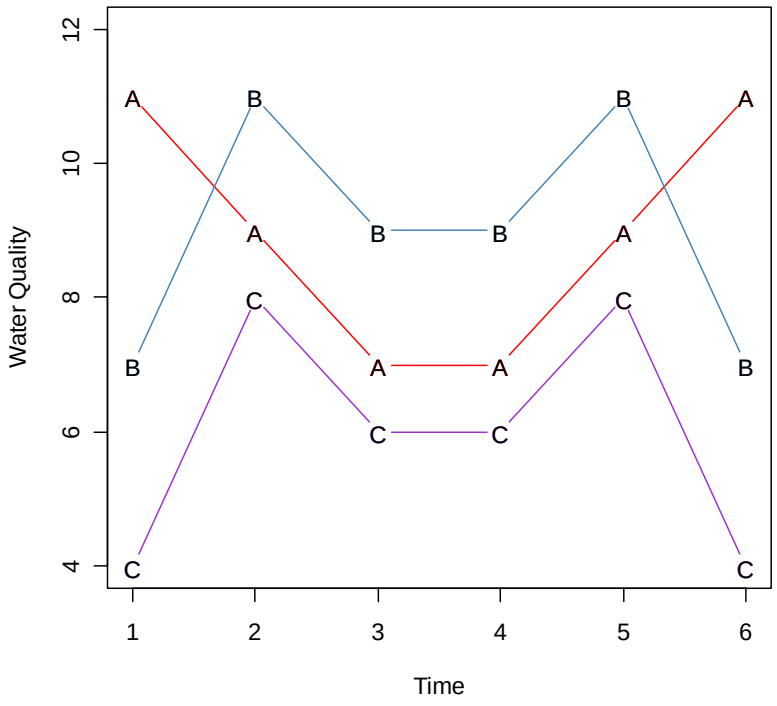

Figure 12 Hypothetical temporal profiles illustrating relative advantages between the two criteria. The profiles are indistinguishable based on *Z* statistics, which equal zero for each. However, the RTepsi is 1 comparing profiles A and C, and 0.47 for each other pair of stations.

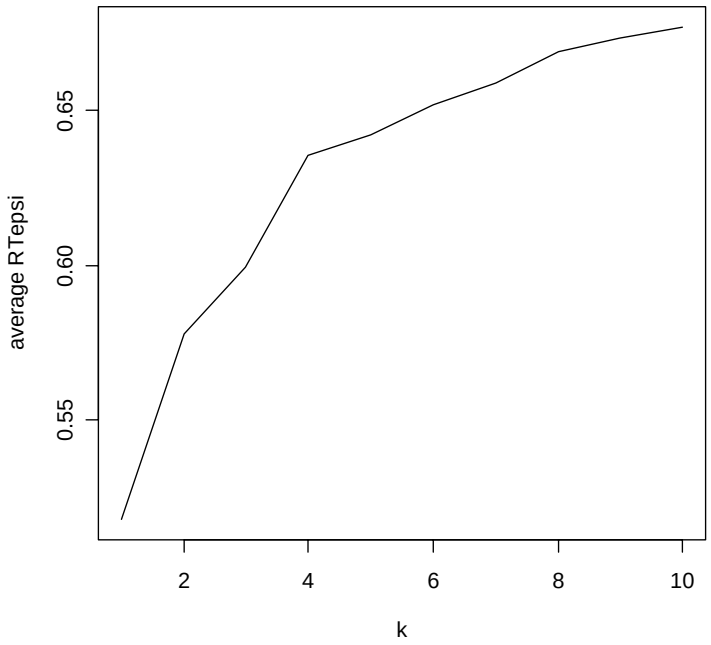

Figure 13 Relation of averaged RTepsi to number of clusters.

APPENDIX 1. STATION CODES AND GEOGRAPHIC COORDINATES FOR 71 STATIONS USED IN THE ANALYSIS.

`station` – station code as used by Virginia Department of Environmental Quality The first 2 characters are a basin code and the next the 3 a stream code. The final 5 characters give miles between the station and the mouth of the stream.

`station.1` - compact station code generated for labeling plots. The first 3 letters are the stream code extracted from `station`. The numeric code represents separate indexing of the stations included in our analyses, when sorted according to `station` (hence sorted on mileage from mouth).

latitude, longitude – coordinates of the station.

years data – number of years with one or more measurements, out of a maximum of 15.

Stations used in Analysis

station	station.1	latitude	longitude	years data
2-APP001.53	APP-1	37.31	-77.30	15
2-APP012.79	APP-2	37.23	-77.42	15
2-APP050.23	APP-3	37.35	-77.85	15
2-APP118.04	APP-4	37.33	-78.47	15
2-BCC004.71	BCC-1	38.07	-79.90	15
2-BEN001.42	BEN-1	36.86	-76.48	15
2-BLP000.79	BLP-1	38.20	-79.57	15
2-BLY000.65	BLY-1	37.29	-77.26	15
2-BUF002.10	BUF-1	37.61	-78.92	15
2-CFP004.67	CFP-1	37.99	-79.49	15
2-CHK002.17	CHK-1	37.26	-76.88	15
2-CHK006.14	CHK-2	37.31	-76.87	15
2-CHK023.64	CHK-3	37.40	-76.94	15
2-CHK032.77	CHK-4	37.43	-77.04	14
2-CHK062.57	CHK-5	37.60	-77.38	13
2-CHK076.59	CHK-6	37.70	-77.51	14
2-CLG000.23	CLG-1	37.23	-76.69	13
2-CWP002.58	CWP-1	37.79	-79.76	15
2-EBE002.98	EBE-1	36.84	-76.24	7
2-ELI002.00	ELI-1	36.90	-76.34	15
2-ELI004.79	ELI-2	36.87	-76.33	7
2-FAC000.85	FAC-1	37.44	-77.44	15
2-HRD011.57	HRD-1	37.81	-78.46	15
2-JKS000.38	JKS-1	37.79	-79.78	15
2-JKS023.61	JKS-2	37.79	-80.00	15
2-JKS030.65	JKS-3	37.84	-79.99	13
2-JKS058.60	JKS-4	38.04	-79.88	15
2-JMS005.72	JMS-1	36.95	-76.39	15
2-JMS013.10	JMS-2	36.99	-76.48	15
2-JMS021.04	JMS-3	37.06	-76.59	15
2-JMS032.59	JMS-4	37.20	-76.65	15
2-JMS042.92	JMS-5	37.20	-76.78	15
2-JMS055.94	JMS-6	37.27	-76.99	15
2-JMS069.08	JMS-7	37.30	-77.13	15
2-JMS074.44	JMS-8	37.32	-77.22	15
2-JMS075.04	JMS-9	37.31	-77.23	15
2-JMS099.30	JMS-10	37.40	-77.39	15
2-JMS104.16	JMS-11	37.45	-77.42	15
2-JMS110.30	JMS-12	37.53	-77.43	15
2-JMS117.35	JMS-13	37.56	-77.54	15
2-JMS157.28	JMS-14	37.67	-78.09	15
2-JMS176.63	JMS-15	37.71	-78.30	15
2-JMS189.31	JMS-16	37.80	-78.49	15
2-JMS229.14	JMS-17	37.54	-78.83	14
2-JMS258.54	JMS-18	37.41	-79.15	15
2-JMS275.75	JMS-19	37.51	-79.33	15
2-JMS282.28	JMS-20	37.59	-79.38	15
2-JOG000.62	JOG-1	36.99	-76.56	15
2-LAF001.15	LAF-1	36.91	-76.31	7
2-LAF003.83	LAF-2	36.89	-76.28	7
2-MCM005.12	MCM-1	38.10	-78.59	15
2-MIC000.03	MIC-1	37.21	-76.74	13
2-MRY014.78	MRY-1	37.75	-79.39	15
2-NAN019.14	NAN-1	36.74	-76.58	15
2-PGN000.00	PGN-1	37.01	-76.57	15
2-PGN000.80	PGN-2	37.00	-76.57	15
2-PGN001.19	PGN-3	37.00	-76.58	15
2-PGN002.58	PGN-4	37.00	-76.61	15
2-PGN003.57	PGN-5	36.99	-76.62	15
2-PGN004.57	PGN-6	36.98	-76.62	15
2-PGN005.46	PGN-7	36.99	-76.63	15
2-PGN006.65	PGN-8	36.99	-76.65	15
2-PGN007.44	PGN-9	37.00	-76.65	15
2-PGN008.42	PGN-10	37.01	-76.66	15
2-PGN010.07	PGN-11	37.02	-76.67	15
2-PNY005.29	PNY-1	37.70	-79.03	14
2-POT000.12	POT-1	37.75	-80.00	15
2-POW000.60	POW-1	37.22	-76.78	13
2-RVN015.97	RVN-1	37.86	-78.27	15
2-SBE001.53	SBE-1	36.83	-76.29	13
2-SGL001.00	SGL-1	36.74	-76.56	15

APPENDIX 2. UPDATED R FUNCTION LIBRARY

```
asClass <- function(clusterList, N=NULL)
{
  # Convert representation of a classification.
  # In input a classification is a length-K (num. clusters) list of cluster
  # objects where a cluster object is a vector of unit indices representing
  # a single cluster. In output a classification is a length-N (=num. units)
  # vector of class indices.  (D. Farrar 2006)

  K <- length(clusterList) # num. clusters
  if(is.null(N)) { N <- 0 ; for(k in 1:K) N <- N + length(clusterList[[k]]) }
  Class <- rep(NA,N);
  for(k in 1:K) Class[clusterList[[k]]] <- k
  return(Class)
} #- end def. 'asClass' ------------------------------------------------------


asClusterList<-function(
  Class,       # vector of class indices (N*1)
  K=NULL       # number of classes (if null then read from argument 1)
  )
{
  # Inverse of asClass() (D. Farrar, 2006)

  if(is.null(K)) K <- max(Class)
  N <- length(Class)
  outlist <- as.list(rep(NA,K))

  for (k in 1:K) {
    result.k <- (1:N)[Class==k]
    outlist[[k]] <- result.k
  }

  return(outlist)
} #- end def 'asClusterList' -------------------------------------------------

dendpl <- function(

  distmtx,           # symmetric matrix of distances
  method,            # 'average', 'complete', 'ward', etc.
  leaflabels=NULL,   # leaf labels
  distlabel,         # axis label depends on distance
  plot=TRUE
  )
{
  # Generation and customized plotting of dendrogram using
  # R functions hclust.
  # D. Farrar 2006

  D  <- as.dist(distmtx)              # base R function returns distance matrix
  dendro <- hclust(D, method=method) # cluster object
  dendro.relabeled <- dendro
  if(!is.null(leaflabels)) dendro.relabeled$labels <- leaflabels
  if(plot) plot(dendro.relabeled,ylab=distlabel,sub="",xlab="",main="",cex=0.75)

  return(list(dendro=dendro,
              dendro.relabeled=dendro.relabeled))

} # -- end fn defn ---------------------------------------------------------
```

```
distance.Z2 <- function(Z)
{
  # computes matrix of square differences from Z matrix
  # Input is a vector of Z statistics with names
  # D. Farrar 5/2006

  if(sum(is.na(Z))) stop("distance.Z2 - missing input not permitted.")
  statiset <- names(Z)
  distmtx  <- matrix(0,nstation,nstation,dimnames=list(statiset,statiset))
  nstation <- length(Z)
  for(i in 1:(nstation-1))
    for(j in (i+1):nstation)
      distmtx[i,j] <- distmtx[j,i] <- (Z[i] - Z[j])^2

  return(distmtx)
} #-- end fn dfn. ----------------------------------------------------------

Ktau <- function(X,minN=2) {

  # Kendall's tau and related statistics for multiple strata, e.g., stations.
  # Arguments:
  # 1) data matrix where e.g. rows correspond to stations and columns to years.
  #    Missing values are permitted.
  # 2) minimum number (non-missing) for each station
  # D. Farrar 5/2006

  numLocs <- dim(X)[1]
  Z <- S <- tau <- rep(NA,numLocs)  # intialize
  names(Z) <- rownames(X)
  for(k in 1:numLocs)             # loop over stations
  {
    kData <-X[k,]                # data for station k in kth row
    miss01<-is.na(kData)         # 1 or 0 according as missing or not
    Nk <-sum(!miss01)            # years w/ data not missing for locn k
    if(Nk >= minN)                # use locns w/at least the min num yrs
    {
      kData <- kData[!miss01]    # non-missing values for location
      S.k <- pairs.k <- 0        # S, number of non-tied pairs
      for(i in 1:(Nk-1))
      {
        for(j in (i+1):Nk)
        {
          x.i <-kData[i]
          x.j <-kData[j]
          if (x.i != x.j)           # if [not tied]
            pairs.k <- pairs.k + 1
            S.k <- S.k + ((x.i <= x.j) - (x.j <= x.i))
        }   # for(j ... [over rows]
      }     # for(i ... [over rows]
      varSk <- Nk*(Nk-1)*(2*Nk+5)/18    # variances without tie adj
      forAdj<- as.vector(table(kData))  # counts per tied group
      is.gt1<- (forAdj > 1)             # flag tied groups size gt 1
      if(any(is.gt1)) {                 # then do tie adjustment
        forAdj<- forAdj[is.gt1]
        varSk <- varSk - (1/18)*sum(forAdj*(forAdj-1)*(2*forAdj+5))
        }  # if(any( ... [tie adj]
      tau[k] <- S.k / pairs.k
      Z[k] <- (S.k - (S.k>0) + (S.k<0))/sqrt(varSk)
    }     # if(Nk>=minN) [enough data for stats ]
  }       # for k in 1...[over stations]

  return(list(Z=Z,tau=tau))

} #-- end fn. defn. -------------------------------------------------------
```

```
plotEnvelope <- function(

   x,                # (N*T) matrix where each row is a profile
   xforplot,         # year series
   forEnvelope=NULL,  # (N*T) logical - which rows to use for envelope
   asBackground=NULL, # (N*T) logical - which plot as lines in background
   asForeground=NULL, # (N*T) logical - which plot as lines atop envelope
   plotMedian=F,      # Logical - plot only medians for envelope set?
   ylim=NULL, xlab="year", ylab=NULL, main="",  # graph params
   col=c( background="maroon4", foreground="grey" )
   )
{
  # superimpose plots of multiple series, divided into 3 subsets,
  # any of which may be empty.  Optionally include median
  # of encvelope set.
  # remark:  I like to log10 the ordinate, but I do that external to
  # the function, and pass the logs to the function. (DF)(
  # D. Farrar 2006

  numprofiles <- nrow(x)

  #-- line type for foreground lines

  if(plotMedian) lty <- "dashed" else lty <- "solid"

  #-- plot axes only

  plot(xforplot,x[1,],type="n", ylim=ylim, xlab=xlab,ylab=ylab, main=main)

  #-- dotted lines for background profiles

  if(!is.null(asBackground)) for(i in 1:numprofiles) if(asBackground[i])
    points(
          xforplot,x[i,],type="l",lty="dotted",col=col[[1]]
          )

  #-- envelope for specified set


  if(!is.null(forEnvelope)) {

    forEnvM <- x[forEnvelope,]
    envL <- apply(forEnvM,2,function(x) min(x,na.rm=T))
    envU <- apply(forEnvM,2,function(x) max(x,na.rm=T))
    pgon <- rbind( cbind( xforplot, envL),
                   cbind( rev(xforplot), rev(envU))
                 )
    polygon(pgon,border=plotMedian,col=col[[2]])
    # points(xforplot,envL,type="l")
    # points(xforplot,envU,type="l")
  }

  #-- foreground profiles

  if(!is.null(asForeground)) for(i in 1:numprofiles) if(asForeground[i])
     points(
           xforplot,x[i,],type="l",lty=lty
           )

  #-- line for means

  if(plotMedian)
    points(xforplot,apply(forEnvM,2,function(x) median(x,na.rm=T)),type="l")
}
```

```
refAlign <- function(  # align classification relative to reference classfn.

  groups1,  # reference classification formatted as vector of class indices
  groups2   # classification to be aligned

  )
{
  # classifications should be in list format
  # (A cluster is a vector of unit indices.  A classification is a list
  # of clusters.)
  # D. Farrar 2006

  # function to compute indidence matrix for a classification

  N <- length(groups1)
  K <- max(groups1)

  computeZ <- function(
    classification )  # vector gives class index for each unit
  {
    Z <- matrix(0, N, K)
    for (j in 1:K) Z[classification==j,j] <- 1
    return(Z)
  }

  require(gregmisc)
  perms <- permutations(K,K)[-1,]  # enumerate permuttns neglecting identity
  numperms <- nrow(perms)

  # incidence matrices for each classification

  Z.1 <- computeZ(groups1)
  Z.2 <- computeZ(groups2)

  cat("\nreference classification with incidence matrix:\n") ;
  print(cbind(groups1,Z.1)[1:7,]);cat(" . . . \n")
  cat("\nalign:\n") ;
  print(cbind(groups2,Z.2)[1:7,]);cat(" . . . \n")

  algn.table <- crossprod(Z.1,Z.2)    # alignment matrix
  cat("\nalignment matrix:\n") ; print(algn.table)

  currtrace  <- sum(diag(algn.table)) # trace current alignment
  cat("\ntrace = ",currtrace)
  tablesum   <- sum(algn.table)       # sum for alignment matrix

  permbest <- 1:K
  ispermuted <- F

  if(K==2 & (currtrace < tablesum/2) )
  {
    ispermuted <- T
    permbest <- c(2,1)
  } else {                            # (K>2) evaluate all permutations
    traceByPerm <- rep(0,numperms)    # initialization
    for(k in 1:numperms)
      for(l in 1:K) traceByPerm[k] <- traceByPerm[k] + algn.table[perms[k,l],l]
    if(currtrace < max(traceByPerm)) {  # update incidence matrix and alignment
      ispermuted <- T
      bestperm <- perms[which.max(traceByPerm),]
    }
  } # K>2

  cat("\nbest label permutation = ", bestperm, "\n")
  cat("\nalignment matrix:\n")
  print(algn.table[,bestperm])

  groups2x <- rep(NA,N)
  for(j in 1:K) groups2x[groups2==j] <- bestperm[j]
```

```
  return(list(ispermuted=ispermuted,groups=groups2x))
} #--[ end fn. defn. ]----------------------------------------
```

```
simi.RTepsi <- function(
  dmtx,         # data matrix stations * years with row names
  minyrs=2      # required minimum years data with values in both profiles
  )
{
  # Compute a matrix of RTepsi subject to minimum count of years in common
  # with values not missing for each pair of stations.  In case of missing
  # values in the matrix, delete stations until there are none missing.
  # Calls XCorrFunc for each pair of stations.
  # D. Farrar 2006

  dmtx <- as.matrix(dmtx)
  statiset <- rownames(dmtx)
  nstation <- nrow(dmtx)
  XCorrM <- matrix(NA,nstation,nstation,dimnames=list(statiset,statiset))
  diag(XCorrM) <- 1
  for(i in 1:(nstation-1)) {
    y.i <- as.vector(dmtx[i,])
    for(j in (i+1):nstation) {
      y.j <- as.vector(dmtx[j,])
      canUse <- !(is.na(y.i)|is.na(y.j))
      if(sum(canUse) >= minyrs) {
        both <- cbind(y.i,y.j)[canUse,]
        c.ij <- XCorrFunc(both[,1],both[,2])
        XCorrM[i,j] <- XCorrM[j,i] <- c.ij
      }#if
    }# for(j ...
  }# for(i ...

  if(any(is.na(XCorrM))) {
    cat("\nDeleting stations from RTepsi matrix\n")
    while(any(is.na(XCorrM))>0) {
      n <- ncol(XCorrM)
      if(n==1) {
        stop("!! too many stations excluded")
      } else {
        numnotNA <- apply(XCorrM,1,function(x) sum(!is.na(x)))
        delstatn <- which.min(numnotNA)
        cat("\ndeleted:",colnames(XCorrM)[delstatn])
        XCorrM <- XCorrM[-delstatn,-delstatn]
      }#if
    }#while
  }#if

  return(XCorrM)

} #-- end fn dfn. ----------------------------------------------------------
```

```
vbh <- function(Z,loquacious=T) {

  # van Belle & Hughes test for heterogeneity of trend.
  # Input vector of Z's (some missing ok).  Returns p-value.
  # If second arg is TRUE then print some output.
  # D. Farrar 2006

  has.Z <- !is.na(Z)   # id which not missing
  m <- sum(has.Z)
  chiSqHom <- pHomo <- pTotal <- NA
  if(m >= 1) {
    Z <- Z[has.Z]
    chiSqTot <- sum(Z^2)      # total chi-square
    chiSqTrd <- m*mean(Z)^2   # trend chi-square
    pTotal<- pchisq(chiSqTot, m, lower.tail=F)
    if(m >= 2) {
      chiSqHom <- chiSqTot - chiSqTrd   # homogeneity chi-square
      pHomo <- pchisq(chiSqHom, m-1, lower.tail=F)
    }
  }

  if(loquacious) cat(
     "\nchi-square\nTotal\t",chiSqTot,"\nTrend\t",chiSqTrd,"\nHomog\t",
     chiSqHom,"\np=\t",pHomo )

  return(list(pHomo=pHomo, chiSqHom=chiSqHom, pTotal=pTotal))

} #-- end fn. defn. -------------------------------------------------------------
```

```
XCorrFunc <- function(v1,v2,minNotNA=2) {

  # Kendall-type cross-correlation between two series
  # 3rd arg is minimum num not missing in both vectors
  # input vectors must be equal in length.
  # See also : simi.RTepsi uses for computing distance matrix.
  # D. Farrar 2006

  nonmissing <- !(is.na(v1)|is.na(v2))
  v1 <- v1[nonmissing]
  v2 <- v2[nonmissing]
  n <- length(v1)

  if(n < minNotNA) {
    retval <- NA
    if(minNotNA==2)
      cat("\n(XCorrFunc): !!comparing series of length < 2\n")
  } else {
    npairs<-n*(n-1)/2       # pairs with neither missing
    concord <- 0            # count of concordant pairs
    for(i in 1:(n-1)) {
      for(j in (i+1):n)
        concord <- concord +
             ( (v1[j]-v1[i])*(v2[j]-v2[i]) > 0 ) +
             ( (v1[j]==v1[i])&(v2[j]==v2[i]) )
    }      # for(i ...
  retval <- concord / npairs
  }

  return(retval)            # fraction of pairs concordant

} #---[ end fn. defn. ]-------------------------------------------------------
```

```
ZDiffDendro <- function(
  D,            # matrix of distances (n*n)
  leaflabels,   # labels for dendrogram leaves
  plot=T
  )
{

  # Dendrogram for ZDiff with Ward's method and rescaled branch lengths
  # D. Farrar 2006

  #-- unscaled dendrogram

  ddobj<-dendpl(D, method="ward", leaflabels,
                distlabel=expression(Zdiff ^2),
                plot=F)$dendro.relabeled

  lKid   <- ddobj$merge[,1]
  rKid   <- ddobj$merge[,2]
  labels <- ddobj$labels
  height <- ddobj$height

  n <- length(labels)
  n.node <- chisq.node <- rep(0,n-1)      # count per internal node

  retfunc1 <- function(i) {  # recursively compute node count
    l <- lKid[i]; r <- rKid[i]
    if(l<0) lcontrib <- 1 else lcontrib <- retfunc1(l)
    if(r<0) rcontrib <- 1 else rcontrib <- retfunc1(r)
    return(lcontrib+rcontrib)
  }
  retfunc2 <- function(i) {  # recurseively compute node chi
    retval <- height[i]/2
    l <- lKid[i]; r <- rKid[i]
    if(l>0) retval <- retval + retfunc2(l)
    if(r>0) retval <- retval + retfunc2(r)
    return(retval)
  }

  for(i in 1:(n-1)) {
    n.node[i] <- retfunc1(i)
    chisq.node[i] <- retfunc2(i)
  }

  # p.node <- pchisq(chisq.node,n.node-1,lower.tail=F)

  ddobj2 <- ddobj
  ddobj2$height <- sqrt(chisq.node / (n.node-1))
  # ddobj2$labels <- station.1

  if(plot) {

    mar.0 <- par()$mar  # widen left margin to take in axis label
    par(mar=mar.0*c(1, 1.1, 1,1))

    plot(ddobj2, main="", sub="", xlab="",
      ylab=expression(sqrt(CHISQ[H] / df) ),
      cex=0.75,cex.axis=0.1)

    abline(h=1,lty="dotted")

    par(mar=mar.0)
  }
  dendro<-ddobj2
  return(dendro)
} #-------------------------------------------------------------------------
```