JP1.7

AUTOMATED, SUPERVISED SYNOPTIC MAP-PATTERN CLASSIFICATION USING RECURSIVE PARTITIONING TREES

Alex J. Cannon*, Paul H. Whitfield, and Edward R. Lord
Meteorological Service of Canada, Vancouver, British Columbia

1. INTRODUCTION

Synoptic map-pattern classifications attempt to cluster atmospheric circulation fields into relatively homogeneous groups. Ideally, each group should reflect a distinct synoptic situation. The resulting classification can then be used to investigate synoptic-scale meteorological controls on local environmental variables (Yarnal, 1993).

Two main criteria are used to organize synoptic map-pattern classification methods. The first criteria separates methods based on how the atmospheric circulation fields are linked with the local environment. In unsupervised approaches atmospheric circulation data are clustered without reference to local weather conditions. Alternatively, in supervised approaches joint consideration is given to both the synoptic circulation and weather observed at the surface. The second criteria separates methods based on the level of automation of the classification procedure. Automated procedures, usually computer-based, use an established algorithm to generate the classification. Manual approaches rely on human judgement and intuition to form the classification and sort maps into groups.

To date, synoptic map-pattern classifications have typically employed either an automated, unsupervised approach or a manual, supervised approach. Both methods have different strengths and weaknesses. While automation allows classifications to be built quickly and be replicated easily, the use of an unsupervised, statistically-based clustering algorithm may result in some groups that have little meteorological significance (Frakes and Yarnal, 1999). Consequently, circulation-environment links may not be well resolved using this approach. Similarly, while manual classifications are time-consuming to build and are difficult to replicate, they may be better at handling links between certain synoptic situations and surface weather conditions. As maps are grouped by a trained expert, important synoptic controls on local weather can easily be accounted for as part of the classification system.

An ideal compromise between these two approaches is one that is both automated and supervised. Hughes et al. (1993) and Zorita et al. (1995) first applied this type of approach to synoptic classification. In their work, recursive partitioning trees (Therneau, 1983; Breiman et al., 1984) were used to develop classification systems relating gridded sea-level pressure (SLP) data to rainfall conditions at stations in North America. More recently, Schnur and Lettenmaier (1998) used recursive partitioning to generate a synoptic climatology of rainfall states at stations in four regions of Australia. This method was again used by Zorita and von Storch (1999) to classify circulation-rainfall relationships on the Iberian Peninsula.

In contrast to automated, unsupervised classification systems, recursive partitioning trees are supervised and give joint consideration to both the synoptic-scale circulation and the local environment. Conceptually, this type of model is a discrete analog to continuous statistical downscaling methods (Hewitson and Crane, 1996). Similar to statistical downscaling models, inputs to recursive partitioning trees are synoptic-scale circulation fields and the output is a measure of some local environmental variable of interest. Circulation patterns relevant to the environmental variable are therefore more likely to be selected than if the classification were based only on atmospheric circulation data; patterns with little significance are more likely to be avoided. Because recursive partitioning models are also automated, improved discrimination does not come at the expense of speed or ease of replication.

Synoptic climatological applications of the recursive partitioning algorithm have tended to focus on a single local variable of interest. Hughes et al. (1993), Zorita et al. (1995), Schnur and Lettenmaier (1998), and Zorita and von Storch (1999) each related circulation conditions to regional rainfall conditions. Consequently, these classifications are of little use to researchers dealing with variables other than rainfall. Unless the target variable is a general indicator of weather conditions at the location of interest, classifications derived using recursive partitioning will be not be usable in the same manner as more general synoptic climatologies.

The primary goal of the current study is the development of an automated, supervised method for producing synoptic map-pattern classifications capable of resolving the main synoptic controls on surface weather in a given region. It is intended that classifications resulting from such a method could be used as alternatives to manual classifications (Lamb, 1972; Muller, 1977; Hess and Brezowsky, 1977) or automated, unsupervised classifications developed for a specific region (Yarnal, 1993). To achieve this goal, recursive partitioning trees are used to relate synoptic-scale atmospheric conditions to a group of surface weather-elements. Two model architectures are investigated. In the first, principal component analysis (PCA) is used to generate a one-dimensional index that summarizes the surface weather-element variables.

* *Corresponding author address:* Alex J. Cannon, Science Division, Meteorological Service of Canada, 700-1200 W. 73rd Ave, Vancouver, BC V6P 6H9; e-mail: alex.cannon@ec.gc.ca.

This index is then used as a target in a standard univariate regression tree built using the recursive partitioning algorithm. In the second, atmospheric circulation variables are related to the weather-elements directly using a multivariate extension to the univariate regression tree approach. The latter approach avoids the loss of information resulting from the compression of the weather-elements data down to a single variable. The ability of each model to predict variations in the weather-element data is evaluated and compared using a cross-validation procedure. Results are measured against those from an unsupervised map-pattern classification based on the k-means clustering algorithm.

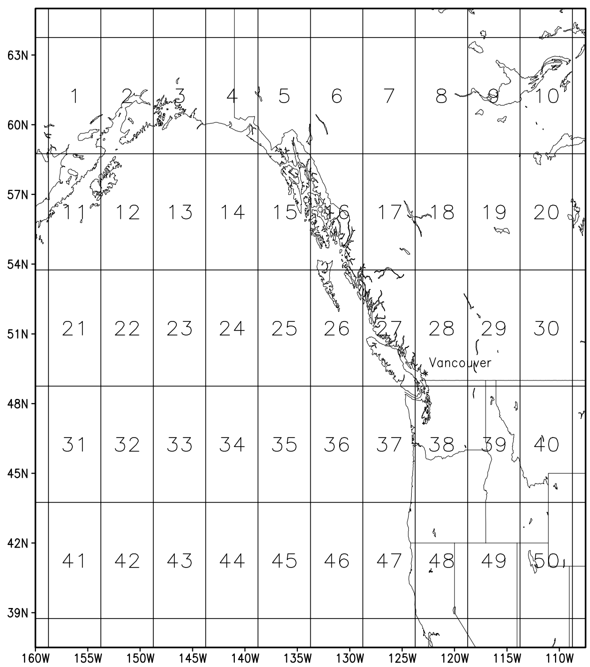

Figure 1. Map showing the grid-points used for the atmospheric circulation data and the location of the surface observation station.

2. DATA

2.1 *Atmospheric circulation*

Gridded SLP and 500-hPa geopotential height data from the NCEP/NCAR model reanalysis project (Kalnay et al., 1996) were used as inputs to the synoptic map-pattern classifications. Daily averages from 1953 to 1998 were obtained for a region covering western North America and the north Pacific Ocean (40°N-62.5°N; 157.5°W-110°W). Data were first smoothed spatially by averaging the 2.5°×2.5° resolution grids (10×20) to 5°×5° grids (5×10). The spatial domain is shown in Figure 1.

To reduce the impact of seasonal variability in the magnitude of circulation data on the classifications, moving average filters were applied prior to identification of the map-patterns (Hewitson and Crane, 1992; Yarnal, 1993). For each day in the analysis, grid-point values were expressed as deviations from a mean value calculated using all data from the 13-days centered on the day of interest. This form of filter preserves spatial patterns in the data but removes variations in average magnitude occurring on time-scales longer than 13-days. The 13-day window was selected following Hewitson and Crane (1992) and is based on power spectra and correlograms of the daily SLP and 500-hPa geopotential height time-series. The filter therefore prevents classifications from being overwhelmed by seasonal variations in average magnitude while still capturing variability occurring on time-scales less than the life-span of typical synoptic-scale systems (Hewitson and Crane, 1992; Yarnal, 1993). Map-patterns that reflect only gross seasonal features of the atmospheric circulation are thereby avoided.

2.2 *Surface weather-elements*

Five variables were selected to describe surface weather conditions in the vicinity of Vancouver, British Columbia (Figure 1). Hourly observations of surface temperature, dew-point temperature, percent cloud opacity, wind speed, and wind direction were obtained from Environment Canada for the monitoring station located at Vancouver International Airport. These five variables were chosen to reflect standard observations reported by Environment Canada's hourly airport observing stations. Variables selected were similar to those used in other studies that defined synoptic types based on local airmass characteristics (Yarnal, 1993; McGregor and Bamzelis, 1995; Kalkstein et al., 1996; Greene et al., 1999). Local pressures at Vancouver were excluded because gridded SLP data were a part of the atmospheric circulation data set. This ensured that the two sets of data were independent. For the remaining variables, daily mean values for the period 1953-1998 were calculated from the hourly observations. Mean wind speeds and circular mean wind directions were converted into u and v wind components (east-west and north-south respectively). Similar to the filtering applied to the circulation data, effects of seasonality were removed from the surface weather-element data by expressing variables as deviations from centered 13-day moving averages (Yarnal, 1993).

3. METHOD

3.1 *Recursive partitioning trees*

Recursive partitioning trees are data-driven statistical models capable of representing nonlinear and interactive relationships between input variables and one or more output variables. For a complete description of the recursive partitioning algorithm the reader is referred to Therneau (1983), Breiman et al. (1984), and Therneau and Atkinson (1997). In this study, two specific forms of recursive partitioning models, univariate regression trees and multivariate regression trees, are described. Burrows et al. (1995) provide an excellent description of univariate regression trees- recursive partitioning models with a single continuous output variable- from a meteorological forecasting perspective. Multivariate regression trees, discussed by Yu and Lambert (1999), are a recent extension to univariate regression trees that allow multiple continuous output variables.

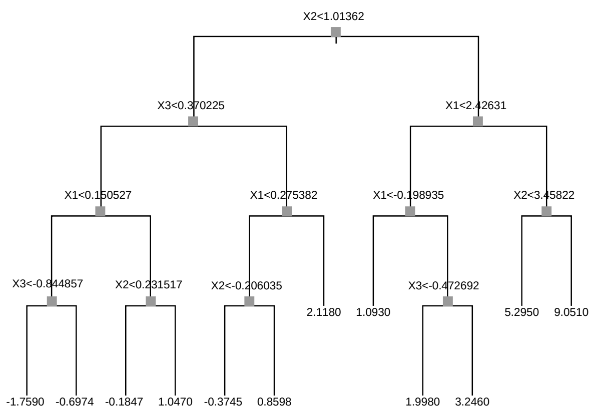

Figure 2. Example of a univariate regression tree with three inputs (*X1*, *X2*, and *X3*). Decision nodes are marked with squares. If the decision rule at a given node is true, cases follow the left branch; if false, cases follow the right branch. The number below each terminal node is the average of the output variable values (*Y*) for cases assigned to that node.

The goal of the recursive partitioning algorithm is to separate the input space in such a way that output variable cases are placed into groups that are as homogenous as possible. As shown in Figure 2, the partition is represented using a tree-like structure. Inputs to the model are presented at the top of the tree and criteria determining which branch each case proceeds to are made at decision nodes. Depending on whether the criterion at the decision node is met or not, cases either follow the left branch or the right branch down the tree. Cases are assigned to classes based on the terminal nodes they reach in the tree; each terminal node defines a potential class in the synoptic climatology.

Predicted output values for classes are also assigned by the recursive partitioning model. Output values for cases in terminal nodes are averaged; these mean values are used as predictions for the classes. In the univariate regression tree shown in Figure 2, for example, a case with inputs $X1 = 1$, $X2 = 2$, and $X3 = 3$ would yield a predicted output $Y = 3.246$, the mean value of all cases assigned to that terminal node during training. Counting from left to right, this case would be a member of class 10. Residual errors associated with the tree are given by the difference between the observed outputs and the corresponding predicted value. If the observed value were equal to 3, the residual error for this example would be -0.246. In multivariate regression trees, terminal nodes yield predicted values for each of the output variables.

Regression trees are built using an algorithm that selects and creates decision nodes so that output variable cases are placed into increasingly similar groups. The decision rule at each new node is chosen by iteratively searching through the input variables to find the split that maximizes a measure of node homogeneity. The splitting criterion *SC* is given by

$$SC = SS_T - (SS_L + SS_R) \tag{1}$$

where SS_T is the sums-of-squares for the node (equal to the summed residual error defined above), and SS_L and SS_R are the sums-of-squares for the left and right branches respectively. Choosing the highest value of *SC* leads to the split that maximizes the sums-of-squares between the new branches. In univariate trees, errors are summed over cases of the output variable. In multivariate regression trees, errors are summed over cases of each output variable and then combined.

The algorithm that creates nodes is controlled by two parameters: N_s, the minimum number of cases in a node required to attempt creating a split, and N_t, the minimum number of cases in a terminal node. By default, the recursive partitioning algorithm sets N_s to 20 and N_t to $N_s/3$. In synoptic climatological applications where sample sizes are generally quite large, final tree-structure is not very sensitive to these parameters; changes are instead reflected in computation time.

New nodes are created until each terminal node contains a minimum number of cases or no further splits can be made because the splitting criterion has converged. By default, tree size is not limited during the initial fitting process; tree building continues until terminal nodes have reached the minimum size defined by N_t or *SC* has been maximized and no further splits are possible. Since tree size is not limited, models may overfit the training data used to grow the tree and may not reflect the underlying relationships between inputs and outputs. In the most extreme case, when N_t is set to one and each case is allowed to reach its own terminal node, the residual error of the tree on training data will be zero. The model will have memorized both the structure underlying the data and also noise; performance of the model on data not used in the building process may be poor as a result. As a remedy, overfit trees are pruned by removing unnecessary branches from the model, thereby reducing the number of terminal nodes. The pruning step is very important in synoptic climatological applications as the number of terminal nodes determines the number of map-patterns.

The amount of pruning is determined by inspecting out-of-sample estimates of model performance calculated using cross-validation (Weiss and Kulikowski, 1991). In *N*-fold cross-validation, the available data are first split into *N* equal size bins. Full trees are then built using data from *N*-1 groups and residual errors for pruned sub-trees are calculated using the data remaining in the left-out bin. This procedure is repeated *N* times for each sub-tree, rotating the training and left-out bins at each fold of the cross-validation. The *N* error estimates from cross-validation are then collected and their mean and standard error *SE* values are computed and stored. Following cross-validation, the cross-validated errors for the pruned trees are plotted against the number of terminal nodes. An example is given in Figure 3; in this case, the error values plotted are proportions of unexplained variance for the pruned trees. As the number of terminal nodes initially increases, cross-validated error typically decreases quite rapidly. This is followed by a relatively flat plateau, and, as overfitting of the training data occurs, a gradual increase in cross-validated error. Trees that are close to minimizing the cross-validated error (i.e. those along the plateau in the plot) are likely to perform well on true out-of-sample data. In practice, performance of these models will be very similar; as a result, the simplest is chosen for sake of parsimony.

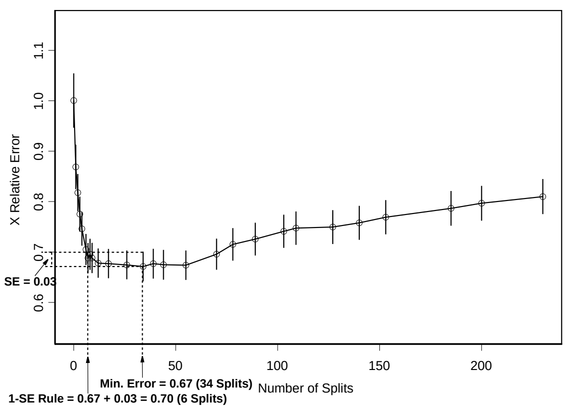

Figure 3. Example cross-validation error plot obtained during the recursive partitioning procedure. Error bars show +/- 1-SE. The minimum cross-validated error and the error associated with the 1-SE rule are marked.

Objectively, the smallest tree that is within one *SE* of the minimum is usually selected as the optimum model (Therneau and Atkinson, 1997). The original tree built using the full dataset is then pruned using this criterion, commonly referred to as the 1-*SE* rule. In Figure 3, for example, the minimum cross-validated error is 0.67 with a *SE* of 0.03; the 1-*SE* rule would select the tree with an error equal to 0.67+0.03=0.70, in this case one with only 6 terminal nodes. In a synoptic climatological context, however, the 1-*SE* rule may select a tree with more nodes than can be easily interpreted. Instead, the user can choose to prune the tree to a smaller size, sacrificing model performance for a more compact classification.

3.2 *Synoptic map-pattern classifications*

To generate the synoptic map-pattern classifications, filtered SLP and 500-hPa geopotential height data were used as inputs to univariate and multivariate regression tree models. Following the recommendation of McKendry (1994), data from both atmospheric levels were considered simultaneously in the classifications. Unlike previous studies which used circulation PCs (Zorita et al., 1995; Zorita and von Storch, 1999; Schnur and Lettenmaier, 1998), grid-point data were used as model inputs in this study. As recursive partitioning algorithms are able to handle large data sets and are not sensitive to correlations between inputs (Burrows et al., 1995), data compression and decorrelation of inputs using PCA is not strictly required prior to classification. In addition, McKendry et al. (1995) suggested that map-pattern classifications based on grid-point data can be applied to GCM scenarios with greater ease than methods requiring a PCA pre-processing step. Brinkmann (1999) found that information necessary for adequate discrimination between synoptic classes may be contained in higher-order PCs not retained in the PCA. As a result, classifications using original grid-point data may be better than those using data reduction methods on the circulation fields.

PCA was used to define an index summarizing surface weather conditions in Vancouver. Scores on the leading PC of the five filtered weather-element variables were used as outputs in the univariate regression tree model. The leading PC was calculated using a P-mode PCA on the correlation matrix of the variables (Yarnal, 1993). Variance explained by the PC over the five weather-elements was 42%, ranging from a minimum of 22% for the v wind component to a maximum of 68% for dew-point temperature. The five filtered weather-elements were used directly as outputs in the multivariate regression tree. This avoided loss of information due to reduction of the weather-element data down to a single dimension.

Following preparation of the circulation data and the weather-element index, univariate and multivariate regression trees were used to generate the synoptic map-pattern classifications. Default values of N_s and N_t (equal to 20 and 7 respectively) were used in the current study. Lowering values of the control parameters did not affect the structure of the final pruned trees.

For comparison with the recursive partitioning models, unsupervised synoptic map-pattern classifications based on the k-means clustering algorithm were also produced. Standardized values of the filtered SLP and 500-hPa geopotential height data were clustered using a batch k-means algorithm (Anderberg, 1973). Cluster centers in the k-means algorithm were initialized using the maximum-norm procedure described by Katsavounidis et al. (1994).

To facilitate comparisons between the recursive partitioning and benchmark models, the number of classes was held constant in the current study. The number of classes for all models was selected by inspecting cross-validation error plots for pruned recursive partitioning trees. In the following discussion, univariate regression tree models are referenced as URT, multivariate regression trees as MRT, and k-means clustering models as KMC. Plots for the recursive partitioning models are presented in the top and bottom panels of Figure 4 respectively. For URT and MRT models, the 1-*SE* rule selected 53 classes and 65 classes with cross-validated values of unexplained variance equal to 0.585 and 0.728 respectively. Note that values of unexplained variance reported in Figure 4 for URT and MRT models are not comparable. Values for URT models are referenced against the leading weather-element PC, not the full set of variables.

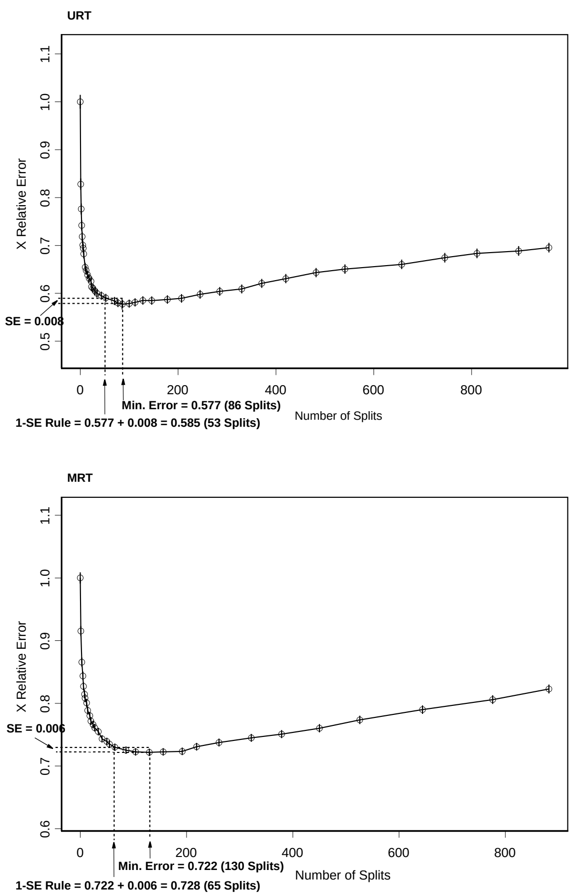

Figure 4. Cross-validated errors for recursive partitioning trees built in the current study. Results from URT models are shown in the top panel and results from MRT models are shown in the bottom panel.

As described above, the 1-*SE* rule can result in a system with too many map-patterns to be of practical use. This was true of the initial models built in the current study. Instead, a cutoff value of 25 classes was chosen for comparing the models. This provided classifications of comparable size to well-known historical synoptic climatologies (for example, 27 weather types by Lamb, 1972 and 29 classes by Hess and Brezowsky, 1977). Also, while subjective, selection of 25 classes lead to classifications that sacrificed little performance relative to those selected using the 1-*SE* rule. For example, the cross-validated error associated with MRT models using 25 classes was equal to 0.756; this represents a <3% increase in unexplained variance relative to the model recommended by the 1-*SE* rule.

4. RESULTS

Synoptic classification performance was evaluated using the method suggested by Yarnal (1993). Each synoptic classifier was evaluated in terms of its ability to predict values of the set of weather-element variables. For each weather-element considered, observed values were compared with values predicted by the synoptic classifier. Mean values variables were calculated for each of the classes in the worked synoptic climatology; days assigned to a given class then used these mean values for their predicted values.

In the current study, r^2 was used as the primary measure of classification performance. Values of the root-mean-squared error and the index of agreement (Willmott, 1981) were also calculated, but are not reported due to good agreement with r^2. Relative differences between the models were similar for the three performance measures.

To obtain unbiased estimates of model performance, 10-fold cross-validation was used to calculate average r^2 values for the period of record. Data sets were randomly split into 10 subsets of equal size. Nine sets were used to generate the synoptic classifications and the remaining set was used to test model performance on data not used in model building. Values of r^2 for the left-out set were recorded and the procedure was repeated, rotating the subsets of data used for training and testing. Reported values of r^2 are means taken over the 10 test subsets. This procedure reduces skill inflation resulting from the evaluation of performance statistics within the data set used to build the model and values are therefore lower than would be expected if cross-validation were not employed. While cross-validation estimates of classification performance were evaluated in a similar manner as those used to determine pruning of the recursive partitioning models, the two procedures were conducted separately in the analysis. Pruning cross-validation was conducted within data reserved for model building and was used exclusively to determine the appropriate number of splits (and thus classes) in the synoptic climatology.

The cross-validation procedure was used to compare the ability of each synoptic classification to predict daily values of the five filtered weather-elements over the period 1953-1998. Cross-validated values of r^2 for predictions of each weather-element are presented in Figure 5. Bars show mean values reported over the 10 cross-validation trials; lines extend from the median to show minimum and maximum values. For reference, values of explained variance for the leading weather-element PC are also reported.

Given that recursive partitioning models used weather-element data as targets, regression trees were expected to outperform the unsupervised k-means algorithm model. URT and MRT models did indeed perform better than the KMC model for all weather-elements. Averaged over the five weather-elements, a difference of 9% explained variance was noted between the KMC and URT models. In no case were cross-validated values of r^2 for the KMC model higher than those reported for the URT model. Comparing the two regression tree approaches, results for MRT models were better than those for the URT models. MRT models explained, on average, 3% more explained variance than did URT models. In this case, the range of variability over the cross-validation trials for the two models did overlap.

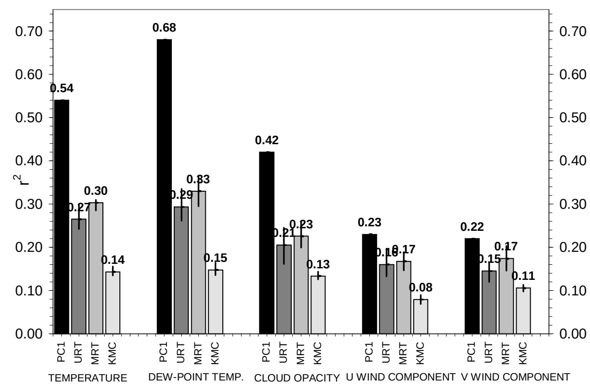

Figure 5. Mean cross-validated r^2 values for predictions of the filtered weather-element data. Error bars indicate the range of r^2 values obtained during cross-validation. Values of r^2 for the first weather-element PC are shown in black for comparison.

5. DISCUSSION AND CONCLUSION

Recursive partitioning offers a powerful method for generating synoptic map-pattern classifications. Since classifications are conditioned on weather-element data, resulting classes are more strongly associated with local weather conditions than are automated classifications based only upon synoptic-scale circulation data. In addition, the modelled tree structure and predicted values of the weather-elements can be used to help interpret relationships between the synoptic map-patterns and weather at the surface. Cross-validation error plots generated by the recursive partitioning model provide guidance for determining the appropriate number of map-patterns, a decision that is often difficult with standard unsupervised cluster analyses. While the 1-*SE* rule provides a simple, automated criterion for selecting the appropriate number of map-patterns for a given dataset, this number can be excessive for large-scale synoptic climatological analyses. Cross-validation error plots can be used to gauge the effect of selecting fewer classes than recommended by the 1-*SE* rule.

The success of map-pattern classifications developed using recursive partitioning depends on the method used to represent the weather-element data. In the current study, weather-elements were represented by the leading PC of the data in the univariate regression trees and by the full set of variables in the multivariate regression trees. As expected, results for models conditioned on the PC were slightly worse than those for models conditioned on the full set of weather-elements. Compressing the weather-element data down to a single variable resulted in a substantial loss of information. For the dataset used in the current study the leading PC explained only 42% of variance in the original weather-elements. Despite the improved performance of the MRT models, it is possible that alternative means of representing the weather-elements could lead to even better classifications using this approach. For example, Yu and Lambert (1999) found that the application of spline basis functions or PCA to their multivariate response variables improved performance of the resulting trees. Further work is required to see if the same holds true for synoptic climatological applications.

Classification results reported in the current study are strictly valid only for the local region from which the weather-elements are drawn. In this regard, the method provides results that are most similar to manual synoptic climatologies. For example, Maunder's (1968) classification of surface weather maps in the Pacific Northwest was based on links between synoptic conditions and local weather conditions over Vancouver Island. As a result, the classification was less relevant to the interior regions of British Columbia and Washington. The same is true of the classifications in the current study. The fact that the method presented is automated, however, means that it could be easily and quickly applied to a variety of specific locations. Alternatively, data from multiple stations could be combined to create a classification system valid for a larger area.

6. REFERENCES

Anderberg, M.R., 1973: *Cluster Analysis for Applications.* Academic Press, 359 pp.

Breiman, L., J.H. Friedman, R.A. Ohlsen, and J.C. Stone, 1984: *Classification and Regression Trees.* Wadsworth and Brooks, 358 pp.

Brinkmann, W.A.R., 1999: Application of non-hierarchically clustered circulation components to surface weather conditions: Lake Superior basin winter temperatures. *Theor. Appl. Climatol.*, **63**, 41-56.

Burrows, W.R., M. Benjamin, S. Beauchamp, E.R. Lord, D, McCollor, and B. Thomson, 1995: CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada. *J. App. Meteor.*, **34**, 1848-1862.

Frakes, B., and B. Yarnal, 1997: A procedure for blending manual and correlation-based synoptic classifications. *Int. J. Climatol.*, **17**, 1381-1396.

Greene, J.S., L.S. Kalkstein, H. Ye, and K. Smoyer, 1999: Relationships between synoptic climatology and atmospheric pollution at 4 US cities. *Theor. Appl. Climatol.,* **62**, 163-174.

Hess, P., and H. Brezowsky, 1977: Katalog der Grosswetterlagen Europas (1881-1976). *Berichte des Deutschen Wetterdientes*, No. 115 Bd. 15, Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main.

Hewitson, B.C., and R.G. Crane, 1992: Regional climate in the GISS GCM: Synoptic scale circulation. *J. Climate*, **5**, 1002-1011.

Hewitson, B.C., and R.G. Crane, 1996: Climate downscaling: techniques and application. *Clim. Res.*, **7**, 85-95.

Hughes, J.P., D.P. Lettenmaier, and P. Guttorp, 1993: A stochastic approach for assessing the effect of changes in regional circulation patterns on local precipitation. *Water Resour. Res.*, **29**, 3303-3315.

Kalkstein, L.S., M.C. Nichols, and C.D. Barthel, 1996: A new spatial synoptic classification: application to air-mass analysis. *Int. J. Climatol.,* **16**, 983-1004.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.

Katsavounidis, I., C.C.J. Kuo, and Z. Zhang, 1994: A new initialization technique for generalized Lloyd iteration. *IEEE Signal Processing Letters*, **1**, 144-146.

Lamb, H.H., 1972: British Isles weather types and a register of the daily sequence of circulation patterns, 1861-1971. *Geophys. Memoirs*, No. 116, HMSO, 85 pp.

Maunder, W.J., 1968: Synoptic weather patterns in the Pacific Northwest. *Northwest Science*, **42**, 80-88.

McGregor, G.R., and D. Bamzelis, 1995: Synoptic typing and its application to the investigation of weather air pollution relationships, Birmingham, United Kingdom. *Theor. Appl. Climatol.,* **51**, 223-236.

McKendry, I.G., 1994: Synoptic circulation and summertime ground-level ozone concentrations in Vancouver, British Columbia. *J. App. Meteor.*, **33**, 627-641.

McKendry, I.G., D.G. Steyn, and G. McBean, 1995: Validation of synoptic circulation patterns simulated by the Canadian Climate Centre General Circulation Model for western North America. *Atmos.-Ocean*, **33**, 809-825.

Muller, R.A., 1977: A synoptic climatology for environmental baseline analysis: New Orleans. *J. App. Meteor.*, **16**, 20-33.

Schnur, R., and D.P. Lettenmaier, 1998: A case study of statistical downscaling in Australia using weather classification by recursive partitioning. *J. Hydrol.,* **212-213**, 362-379.

Therneau, T.M., 1983: A short introduction to recursive partitioning, *Orion Technical Report,* No. 21, Stanford University, Department of Statistics.

Therneau, T.M., and E.J. Atkinson, 1997: An introduction to recursive partitioning using the RPART routines, *Mayo Clinic Technical Report,* No. 61, Mayo Clinic, Department of Health Sciences Research, Section of Biostatistics.

Weiss, S., and C. Kulikowski, 1991: *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*. Morgan Kaufmann, 223 pp.

Willmott, C.J., 1981: On the validation of models. *Phys. Geog.*, **2**, 184-194.

Yarnal, B., 1993: *Synoptic Climatology in Environmental Analysis*. Bellhaven Press, 195 pp.

Yu, Y., and D. Lambert, 1999: Fitting trees to functional data, with an application to time-of-day patterns. *J. Comput. Graph. Stat.,* **8**, 749-762.

Zorita, E., J.P. Hughes, D.P. Lettenmaier, and H. von Storch, 1995: Stochastic characterization of regional precipitation patterns for climate model diagnosis and estimation of local precipitation. *J. Climate*, **8**, 1023-1042.

Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474-2489.