EXPLORING A TYPOLOGY FOR MODELS IN GEOGRAPHY

KATRINA RICHARDS

Dr Katrina Richards, Department of Geography, University of Otago, Dunedin, New Zealand

INTRODUCTION

Models can be grouped and subdivided in a myriad of ways, and numerous typologies exist to classify scientific models. The most widely known for models in geography is that by Chorley (1964; 1967), which is visualized as a flow chart encompassing hardware, mathematical, experimental, natural analogue, and system approaches. Chorley's over-arching classification of model 'types' was innovative in its time and demonstrated the "extraordinary breadth, flexibility, and potential, of the model-concept" (Harvey 1969, p157). Physical geography has embraced the use of models and especially mathematical models. These are often classed according to their scale (e.g. long- or short-term, meso- or micro-scale), mathematical approach (e.g. deterministic *vs* stochastic), or expected outcome, e.g. rainfall data *vs* weather forecasts. In theory, geographical models may be also classified using combinations of grain (resolution), spatial area and temporal duration (Pereira 2002), or using the configuration of temporal and spatial lags between predictor and predicted data sets (Elhorst 2001). Despite this abundance of classification schemes, it remains intrinsically useful to explore new typologies, because these may suggest relationships not previously considered, and can lead to a deeper understanding of models, model building and theory. A typology of models is in effect itself a model, i.e. a model of model types. Thus, to paraphrase Harvey (1969), it may enable one to better visualize relationships, make broad comparisons between models, organize a collection of models, formulate explanations, or help to extend existing theory (Chorley 1964; 1967; Harvey 1969).

This paper presents a typology for predictive mathematical models in geography. It does not seek to replace classification schemes suggested by previous authors such as Chorley (1964; 1967). Rather it dissects and explores a specific sub-set of models (mathematical, predictive) that are strongly represented in the modern geographical literature. The scheme is explored, and first-order verification undertaken, using (1) hypothetical case studies, and (2) selected examples from the climatological literature. The typology was deduced from general principles, and is in theory transferable to other disciplines, e.g. hydrology, geomorphology, or wherever predictive models feature.

There is little consensus in the scientific literature regarding the language of prediction, many terms are used, e.g. hindcast, forecast, backcast, retro-prediction and post-prediction, and their precise meaning may vary between publications. In this paper prediction is used in a general sense to mean a predictive statement about an event. 'Hindcast' is used to mean a prediction concerning events that have already happened, and the term 'forecast' is reserved for predictions concerning truly future events, i.e. which had not yet happened at the time when the prediction was made.

THEORY

The proposed typology is based on six descriptors, grouped into three mutually exclusive and all-inclusive pairs, as shown in Figure 1. In the visualization data sets (depicted as squares) have a temporal location on the x-axis, and occupy information space on the y-axis. Here, 'information' is used rather loosely to mean data that are coherent and organized in some manner. In each case the independent (predictor) data set is shaded, and the dependent (predicted) data set is white. Time runs from past (left hand side, lhs) to future (right hand side, rhs), and (where shown) a black circle represents the present.

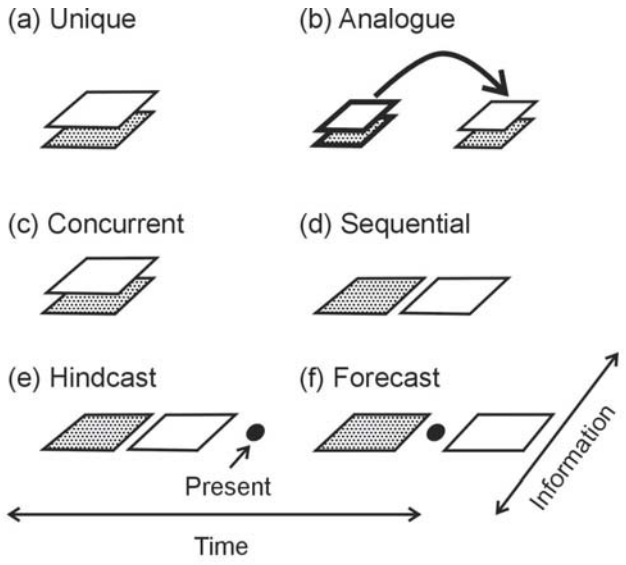

Figure 1: Six descriptors for predictive models in geography, arranged in three mutually exclusive and all-inclusive pairs, showing independent (shaded squares), dependent (white) and analogue (wide border) data sets, analogue links (arrows), and the location of the present (black circle).

The first pair of descriptors is Unique-Analogue. This pair defines the primary subdivision in the typology, and describes a profound difference in model structure. In a Unique model (Figure 1a) the event of interest is viewed as singular, e.g. flood data (white square) may be predicted based on rainfall data (shaded) for the same day. In an Analogue model (Figure 1b), an inference is made that if two events agree in certain respects, they will agree also in other aspects. Here, an event consists of a pair of associated data sets, one independent and one dependent, and an arrow links the analogue event (lhs) to the event of interest (rhs). To illustrate this convention, if an analogy is made between rainfall in 1992 (shaded, lhs) and 2002 (shaded, rhs), this may allow flood data from 1992 (white, lhs) to be used to predict flooding in 2002 (white, rhs).

The second division is based on the pair Concurrent-Sequential. In a Concurrent model (Figure 1c) independent and dependent sets are coincident, e.g. rain and flooding data for the same day. In a Sequential model (Figure 1d) sets occur at different times and are

separated on the x-axis. For example, snowfall in winter (shaded) might be used to predict flooding due to snow melt in the following spring (white).

Hindcast-Forecast is the third pair. It is assigned the lowest rank in the typology because it represents a contrast solely in the timing of events relative to the present. That is, a distinction is made between a predictive model run and tested using historical data (Hindcast; Figure 1e), and a true forecast, i.e. tested against future data, which become available with the passage of time. Note that in Figure 1e all sets lie to the lhs (past) of present (black circle), whereas in Figure 1f, the dependent (predicted) set is located in the future (rhs).

The three pairs can be separately applied, but their real power comes when they are combined, as shown in the schematic diagram in Figure 2. In the visualization events occupy two planes in time-information space. The upper (Unique) and lower (Analogue) planes represent the primary division in model type. Each plane is further subdivided according to the descriptors Concurrent (up-page), Sequential (down-page), Hindcast (lhs) and Forecast (rhs), and symbols and conventions follow those described for Figure 1. In combination, the six descriptors specify eight types of predictive model. These are labeled (a) to (h) in Figure 2.

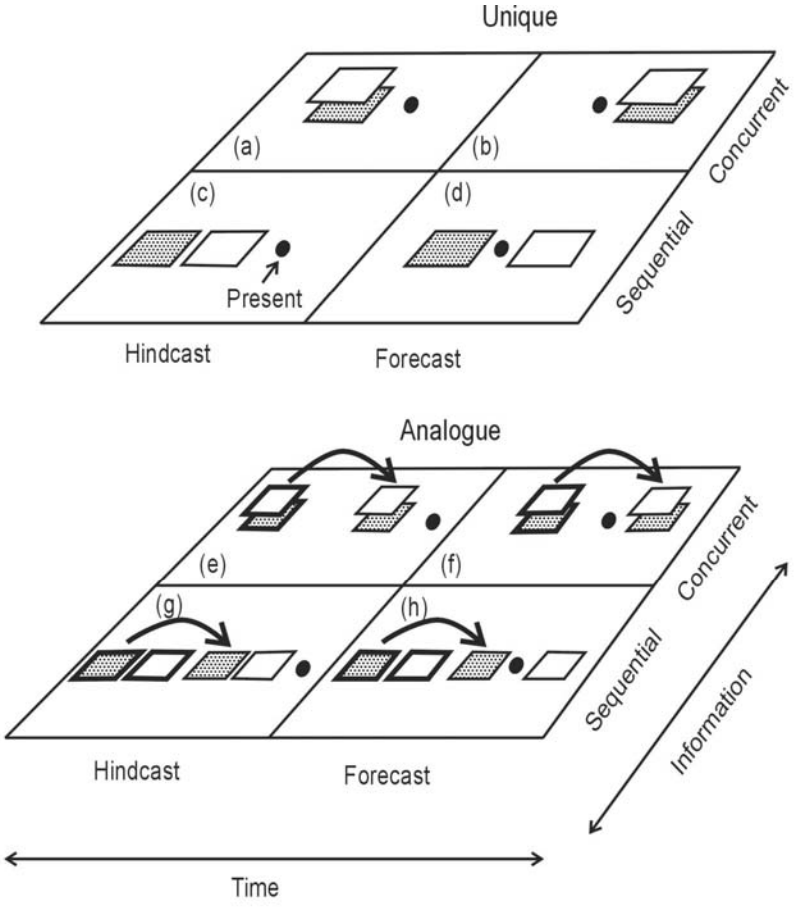

Figure 2: Visualization of the scheme, with symbols as for Figure 1. The eight model types are labeled (a) to (h) for discussion purposes.

EXPLORING THE TYPOLOGY

Hypothetical Case Studies

The typology is first explored using hypothetical cases studies. They deal with dust soiling of orchard fruit; however, another topic could have been used. This was chosen because (1) dust prediction potentially involves interesting combinations of antecedent and subsequent conditions, (2) dust soiling has economic importance because it reduces fruit worth, and (3) dustiness in general is a familiar phenomenon, so the illustrative examples should be reasonably easy to follow.

Unique-Concurrent-Hindcast: Daily dust soiling is predicted for 1992, based on weather conditions, and a numerical model that links the weather observed on a particular day (e.g. windiness) to dust conditions on the same day. The event of interest is viewed as unique (no analogy is made), weather conditions on the day of interest are used to inform the prediction for that day, and the prediction is a hindcast—run and tested using historical data. Thus, the model plots entirely to the lhs (past) of present, as shown in Figure 2a.

Unique-Concurrent-Forecast: Dust soiling for tomorrow is predicted based on the weather forecast for that day, and an understanding of how weather and dust conditions are causally linked. The model is essentially similar to that shown in Figure 2a, except both data sets are located in the future. That is, a true forecast is made and the model plots entirely to the rhs of present in Figure 2b.

Unique-Sequential-Hindcast: Using historical data, dust soiling on picked fruit is predicted based on wind speeds observed one day previous to the harvest date. The sets are treated as unique, the prediction for a particular day is based on data observed during the previous day, and the model is run and tested against existing data. Hence, it plots to the lhs of present, with a temporal separation between sets (Figure 2c).

Unique-Sequential-Forecast: An orchardist, hearing rain on the roof in the night, predicts that the following day will have minimal dust soiling, because fruit will be washed clean. Again, the sets are treated as unique and are sequential, but here the predicted event lies in the future. Thus, the independent data (rainfall) lie to the lhs, and the dependent data (soiling) to the rhs of present (Figure 2d).

Analogue-Concurrent-Hindcast: Dust soiling is predicted based on historical dust data, and a model linking soiling intensity to wind and soil moisture conditions. To develop a prediction, the day of interest is first paired with the day having the most similar wind and soil moisture conditions. Dust conditions on the analogue day are then used to predict dust conditions for the day of interest. Here, an analogue is explicitly chosen, the relevant independent and dependent sets are concurrent for both the analogue event and event of interest, and all sets consist of past data. Four sets are present (Figure 2e): observed environmental (shaded) and soiling (white) data for the analogue (bold) day, and observed environmental (shaded) and predicted soiling (white) data for the day of interest. As in Figure 1b, an arrow indicates the analogue linkage between events.

Analogue-Concurrent-Forecast: Severe dust soiling is forecast for next September, based on the forecast that there will be at least three weeks of road construction work during that month, and using an analogy made with a previous September when road construction occurred. A historical analogue is used to make a true forecast, and the relevant independent and dependent sets are concurrent for each event. The four sets are: past road (shaded) and dust (white) data for the analogue (lhs), and forecast road (shaded) and dust (white) data for the event of interest (rhs) (Figure 2f).

Analogue-Sequential-Hindcast: Dust soiling on a particular day is predicted using rainfall data from the previous day, based on the concept that fruit will be less dusty when rain has been abundant. Specifically, the best analogue day is selected from a historical data set on the basis of rainfall records, and dust conditions for the 24-hour period following that are then used to predict soiling for the 24 hours following the day of interest. A specific analogue is chosen, the data sets are sequential, and all data are historical. As in Figure 2g, all sets—observed rainfall (shaded) and dust (white) conditions for the analogue (bold), and observed rainfall (shaded) and predicted dust (white) conditions for the event of interest—plot to the lhs of present.

Analogue-Sequential-Forecast: Dust soiling on picked fruit is predicted based on historical records of fruit quality, and rainfall data for the two weeks prior to each harvest. Each harvest is first paired with the harvest with the most similar antecedent rainfall. Soiling data for the analogue are then used to predict soiling conditions for the harvest of interest. Specifically, using up-to-date rainfall data (i.e. for the previous two weeks), soiling intensity is predicted on fruit that will be picked tomorrow. Here, an analogue is used, the sets are sequential, and the predicted data (soiling) plot to the rhs of present (Figure 2h).

A review of these hypothetical scenarios suggests that the typology may be valid and useful. It seems to describe several important characteristics of predictive models, the visual representation is useful for describing relationships between data sets, and a hypothetical case study was able to be devised for each model type.

Examples From The Climatological Literature

Hypothetical cases notwithstanding, a more rigorous test is to explore the typology using examples from the scientific literature. Examples are drawn from climatology, specifically, from the numerous papers that involve precipitation or precipitation-related topics. For simplicity, the discussion is limited to a few cases per model type (Table 1). Models that merely describe features, correlate data, or are run for calibration (i.e. training) purposes, are excluded, and spatial issues and model accuracy are not addressed.

Table 1: Selected models used to illustrate the eight model types.

Unique

(a) Concurrent-Hindcast Cherry (1991) Rowell et al. (1995) Menz (1997)	**(b) Concurrent-Forecast** Cherry (1991)
(c) Sequential-Hindcast Moss et al. (1994) Hastenrath et al. (1995) Landman and Mason (1999)	**(d) Sequential-Forecast** Hastenrath et al. (1995) Climatology Research Group, University of Witwatersrand

Analogue

(e) Concurrent-Hindcast Zorita and von Storch (1999) Davis et al. (2000)	**(f) Concurrent-Forecast** Cherry (1992) Davis et al. (2000)
(g) Sequential-Hindcast Roebber and Bosart (1998) O'Brien et al. (2000)	**(h) Sequential-Forecast** Kruizinga and Murphy (1983) Royal Netherlands Meteorological Institute

Unique-Concurrent-Hindcast: Several researchers in New Zealand have used large-scale climate and environmental indices to predict inflows for lakes from which hydroelectricity is generated. For example, Cherry (1991) used concurrent, historical values of the Southern Oscillation Index (SOI, associated with circulation patterns in the Pacific), Quasi-Biennial Oscillation (QBO, which describes equatorial wind patterns), and sun spot number (SSN) to predict the smoothed annual inflow for lakes in the Clutha catchment, New Zealand. In many tropical and monsoon areas of the world, a key research goal is the accurate prediction of seasonal rainfall patterns. For example, Rowell et al. (1995) predicted July-August-September (JAS) rainfall anomalies for tropical North Africa for the period 1949-1990, based on concurrent sea surface temperature patterns (SST, normally expressed as an anomaly from the mean). Remote sensing provides an alternative approach: Menz (1997) predicted precipitation patterns for East Africa based on concurrent satellite-derived (METEOSAT) cloud data for the region and adjacent Indian Ocean. In each of these examples, no specific analogue was chosen, and independent (climate index/SST/cloud) and dependent (inflow/rainfall) sets were concurrent and historical. Thus, the models are classified as Unique-Concurrent-Hindcast (Figure 2a).

Unique-Concurrent-Forecast: There is a strong potential to forecast rainfall when the predictor variables are themselves able to be forecast. Thus, a Unique-Concurrent-Hindcast approach may lead to the development of an equivalent Forecast model. For example, Cherry (1991; described above) forecast monthly smoothed inflows for Lake Roxburgh, New Zealand, for September 1991-November 1992, based on forecasts of SSN. No analogue was used, the independent (SSN) and dependent (inflow) sets were concurrent, and a true forecast was made. Thus, the model is classed as Unique-Concurrent-Forecast (Figure 2b).

Unique-Sequential-Hindcast: Climate-related phenomena are often driven by antecedent weather and environment conditions, and certain models take advantage of this temporal lag. For example, Moss et al. (1994) predicted summer flows for lakes in the Clutha catchment, New Zealand, for the period 1935-1992, using SOI values for the previous spring. Essentially similar approaches are used in Africa and India to predict seasonal rainfall. For example, Hastenrath et al. (1995) predicted rainfall conditions in the Transvaal for December-January-February (DJF) based on parameters such as SOI values for the previous JAS, and SST patterns for the previous November. In this case temporal lags (lead-times) ranged from 21 to zero months, where the latter means the predictor data were from the previous month. Similarly, Landman and Mason (1999) predicted precipitation for South Africa for October-December and January-March using SST and lead-times of five to zero months. In these examples, a specific analogue was absent, and independent (SOI/climate index/SST) and dependent (flows/rainfall) sets were sequential and historical. Thus, they are consistent with a Unique-Sequential-Hindcast classification (Figure 2c).

Unique-Sequential-Forecast: A Unique-Sequential-Hindcast approach may also lead to a forecast model, provided the predictors can be forecast. For example, Hastenrath et al. (1995; described above) made a real-time forecast in October 1994 of rainfall conditions in the Transvaal for December 1994-February 1995, based on climate-related indices for previous months. Several institutions provide operational forecasts based on Unique-Sequential-Forecast models. For example, the Climate Research Group, University of Witwatersrand, began to release Seasonal Rainfall Outlooks (i.e. rainfall forecasts) in 1992, for three- and six-month periods, based on near real-time SST data for past months (Mason 1998). In these cases, events were taken to be unique, the independent (climate index/SST) and dependent (rainfall) sets were sequential, and true forecasts were made. Thus, the models use a Unique-Sequential-Forecast approach (Figure 2d).

Analogue-Concurrent-Hindcast: General Circulation Models generate climate data at global scales. Downscaling is required to provide data at a regional scale, and analogue techniques are often used for this purpose. For example, Zorita and von Storch (1999) predicted mean DJF precipitation values for the Iberian Peninsula, Europe, for 1901-1989, by compared sea level pressure fields (SLP) in the season of interest to SLP recorded during 1951-1989. In each case, they chose the most similar season as a specific analogue, excluding the year of interest. Analogue approaches are common also in paleoclimatology, where they are used to make statements about past climates. For example, Davis et al. (2000) reconstructed Holocene (last 10 000 years) precipitation and temperature conditions for National Parks in the western Great Lakes area, USA, using fossil pollen records, and by assuming that analogous pollen assemblages were produced by similar vegetation and, thus, similar climates. In these cases, specific analogues were chosen, and independent (SLP/pollen) and dependent (rainfall/climate) sets were concurrent and historical. Thus, the models are classed as Analogue-Concurrent-Hindcast (Figure 2e).

Analogue-Concurrent-Forecast: Analogue techniques are used for forecasting in several research areas. Cherry (1992) forecast lake inflows for the Clutha catchment, New Zealand, using an analogy between the SOI and QBO values being forecast for 1992, and SOI and QBO observed during 1976-1977. Inflow values for 1976-1977 were then used to predict monthly inflow values for 1992. Historical analogues are also used, somewhat cautiously, to speculate about global warming and the impact of anthropogenic greenhouse gas emissions. For example, Davis et al. (2000; described above) identified the mid-Holocene prairie period (8000-5000 years ago) as a potential warm-dry analogue for future climate change in the western Great Lakes area, USA, and used this to made statements about ecological implications for the future. In these cases specific analogies were made, the independent (climate index/climate) and dependent (inflows/ecology) sets were concurrent, and forecasts were made. Thus, the models are classed as Analogue-Concurrent-Forecast (Figure 2f).

Analogue-Sequential-Hindcast: Certain models use an Analogue-Sequential-Hindcast approach. For example, Roebber and Bosart (1998) predicted precipitation values in the USA, based on analogues of surface and 500 hPa pressure fields. Periods 48-hours in length were chosen, and the second 24-hour period of the analogue was then used to predict precipitation patterns for the second 24 hours of the period of interest. Precipitation and cloud formation occur more easily when condensation nuclei (CN, e.g. dust or salt particles) are present in the atmosphere. O'Brien et al. (2000) predicted CN data for Mace Head, Ireland, using historical analogues chosen on the basis of a suite of variables, including precipitation, temperature, time-of-day, tide height, and wind speed and direction. In this case, a one-hour analogue was chosen, and CN conditions in the hour following that were used to predict CN values for the one-hour period following the hour of interest. These examples used specific analogues, and the independent (pressure/environmental variables) and dependent (precipitation/CN) sets were sequential and historical. Thus, the models are classed as Analogue-Sequential-Hindcast (Figure 2g).

Analogue-Sequential-Forecast: Analogue-Sequential-Forecast models are regularly used to forecast weather conditions. For example, the Royal Netherlands Meteorological Institute is routinely provided with guidance based on the forecast situation (e.g. expressed as 500 hPa pressure fields) and the thirty best historical analogues for this. Lead-times range from one to six days, and the prediction is based on weather information taken from the 24-hour period following each analogy (Kruizinga and Murphy 1983). Kruizinga and Murphy (1983) demonstrated how these 30-analogue sets may be used to produce categorical (e.g. mean and median) and probabilistic (e.g. probability distribution) forecasts. These examples used specific analogies, independent (situation) and dependent (weather) sets were sequential,

and forecasts were made. Thus, the appropriate classification is Analogue-Sequential-Forecast (Figure 2h).

The typology appears to perform in a satisfactory manner with respect to the climatological models. It seems to describe them in a useful way, and consistency exists between the general structures and relationships observed in the models and those specified in the typology. There is some evidence that the scheme groups and subdivides models realistically, because most of the studies appear only once in Table 1. Three appear twice, but solely because they present two distinct models, i.e. a Hindcast and a Forecast model, that are otherwise essentially similar in structure and approach (Cherry 1991; Hastenrath et al. 1995; Davis et al. 2000). No cases straddle the primary (Unique *vs* Analogue) and secondary (Concurrent *vs* Sequential) subdivisions in the typology. This suggests that the ranking of the pairs may be appropriate.

The visualization does not duplicate all temporal features seen in the models. For example, many involve discontinuous data sets, e.g. JAS rainfall for 1949-1990 (Rowell et al. 1995), events of different temporal length are not differentiated, and sequences and analogies are unidirectional, i.e. past to future. Some Analogue-Hindcast models (e.g. Zorita and von Storch 1999; O'Brien et al. 2000) consider both antecedent and subsequent events as potential analogues, and there is no fundamental reason why predicted data cannot precede predictor data in a Unique model. In theory, it would be relatively straightforward to add this property to the typology, i.e. by swapping shaded and white squares, bold and narrow borders, and reversing arrow directions. Analogue-Sequential models are less straightforward, however, because several sequences are possible among the four sets. Further, the relative location of the present (black circle) may vary, e.g. independent and dependent sets in a Sequential model may both be forecasts (see Figures 2d and 2h), and this would add a further degree of complexity.

On balance, it seems more valuable to retain the present level of generalization in the typology, as represented in Figure 2, and to discuss possible temporal variations in accompanying text. Nevertheless, the typology could be further refined. For example, models could potentially be compared and contrasted according to their grain (resolution) and extent (duration) if appropriate units were added to the time axis (Pereira 2002). Similarly, if appropriate units were devised for the information axis, temporal lags (Elhorst 2001) and information offsets might, perhaps, be used to identify errors in the timing and information content of predictions.

CONCLUSION

The typology appears valid and robust. It performs in a satisfactory manner for a range of geographical examples, describes several important characteristics of predictive mathematical models, highlights similarities and differences, and presents ideas in a compact and convenient format. There are some inconsistencies, but of these are minor, and can largely be ignored because the typology is primarily designed to generalize and simplify reality. Finally, the scheme seems to possess several characteristics of a successful model (Haggett and Chorley 1967), i.e. it is simple but not too simple, provides intellectual stimulation, and suggests new questions and refinements. Given these preliminary indicators of success, further testing and/or development of the typology seems warranted.

Acknowledgements: The author would like to thank Prof Peter Holland, University of Otago, New Zealand, for his assistance with this typology and paper.

REFERENCES

Cherry, N. (1991). *Climate Indices and Clutha River Flows*. Lincoln, New Zealand: Climate Research Unit, Lincoln University.
Cherry, N. (1992). *Forecasts of the Weather Over the Next Three Seasons. Implications for Clyde Inflows*. Report for Electricorp. Lincoln, New Zealand: Climate Research Unit, Lincoln University.
Chorley, R. J. (1964). Geography and Analogue Theory. *Annals of the Association of American Geographers* 54, 127-137.
Chorley, R. J. (1967). Models in Geomorphology. In *Models in Geography*, R. J. Chorley, and P. Haggett (Eds), pp. 59-96. London: Methuen.
Davis, M., C. Douglas, R. Calcote, K. L. Cole, M. G. Winkler, and R. Flakne (2000). Holocene Climate in the Western Great Lakes National Parks and Lakeshores: Implications for Future Climate Change. *Conservation Biology* 14, 968-983.
Elhorst, J. P. (2001). Dynamic Models in Space and Time. *Geographical Analysis* 33 (2), 119-140.
Hagget, P., and R. J. Chorley (1967). Models, Paradigms and the New Geography. In *Models in Geography*, R. J. Chorley, and P. Haggett (Eds), pp. 19-41. London: Methuen.
Harvey, D. (1969). *Explanation in Geography*. Toronto: Macmillan.
Hastenrath, S., L. Greischar, and J. van Heerden (1995). Prediction of Summer Rainfall Over South Africa. *Journal of Climate* 8, 1511–1518.
Kruizinga, S., and A. H. Murphy (1983). Use of an Analogue Procedure to Formulate Objective Probabilistic Temperature Forecasts in the Netherlands. *Monthly Weather Review* 111, 2244-2254.
Landman, W. A., and S. J. Mason (1999). Operational Long-Lead Prediction of South African Rainfall Using Canonical Correlation Analysis. *International Journal of Climatology* 19, 1073-1090.
Mason, S. J. (1998). Seasonal Forecasting of South African Rainfall Using a Non-Linear Discriminant Analysis Model. *International Journal of Climatology* 18, 147-164.
Menz, G. (1997). Regionalization of Precipitation Models in East Africa Using METEOSAT Data. *International Journal of Climatology* 17, 1011-1027.
Moss, M. E., C. P. Pearson, and A. I. McKerchar (1994). The Southern Oscillation Index as a Predictor of the Probability of Low Streamflows in New Zealand. *Water Resources Research* 30, 2717-2723.
O'Brien, E. W, S. G. Jennings, M. Geever, and C. Kleefeld (2000). Relationships Between Condensation Nuclei Number Concentration, Tides, and Standard Meteorological Variables at Mace Head, Ireland. *Journal of Geophysical Research* 105 (D2), 1973-1986.
Pereira, G. M. (2002). A Typology of Spatial and Temporal Scale Relations. *Geographical Analysis* 34 (1), 21-33.
Roebber, P. J., and L. F. Bosart (1998). The Sensitivity of Precipitation to Circulation Details. Part I: An Analysis of Regional Analogs. *Monthly Weather Review* 126, 437-455.
Rowell, D. P., C. K. Folland, K. Maskell, and W. N. Ward (1995). Variability of Summer Rainfall Over Tropical North Africa (1906-92): Observations and Modelling *Quarterly Journal of the Royal Meteorological Society* 121, 669-704.
Zorita, E., and H. von Storch (1999). The Analog Method as a Simple Statistical Downscaling Technique: Comparison With More Complicated Methods. *Journal of Climate* 12, 2474-2489.