

## **MODEL GOODNESS OF FIT: A MULTIPLE RESOLUTION PROCEDURE**

ROBERT COSTANZA

*Coastal and Environmental Policy Program, Chesapeake Biological Laboratory,  
University of Maryland, Solomons, MD 20688-0038 (U.S.A.)*

(Accepted 7 March 1989)

### **ABSTRACT**

Costanza, R., 1989. Model goodness of fit: a multiple resolution procedure. *Ecol. Modelling*, 47: 199–215.

Quantitatively evaluating the goodness of fit of ecological simulation models is difficult, and no generally agreed upon method has evolved. This paper presents a method for quantifying the goodness of fit of spatial and/or time series data and models based on measuring the similarity of the patterns, and the idea that measurement at one resolution is not sufficient to describe complex patterns. The method yields indices that summarize the way the fit changes as the resolution of measurement changes. An expanding 'window' is used to gradually degrade the resolution of the comparison. Lack of fit can be partitioned into 'registration', 'resolution' and residual components. This allows a better understanding of the underlying patterns and type of correspondence. Multiple resolution methods yield additional information not contained in single resolution methods that is necessary to adequately evaluate the performance of complex ecological models.

### **INTRODUCTION**

No theory or model fits reality perfectly. As Albert Einstein once said: "The laws of mathematics, as far as they refer to reality, are not certain, and as far as they are certain, do not refer to reality." What scientists really need to know is exactly how well (or how poorly) their models perform over a broad range of conditions and criteria. Because there are no generally agreed upon procedures for measuring the degree of fit between simulation models and reality, this testing and evaluation process is not often given the importance it deserves. Lack of consensus results from the fact that many levels of disaggregation are used and the range of spatial and temporal resolutions over which models must operate and be tested is very large. Standard statistical tests of goodness of fit are generally applied only at a single resolution of measurement. The thesis of this paper is that there is no

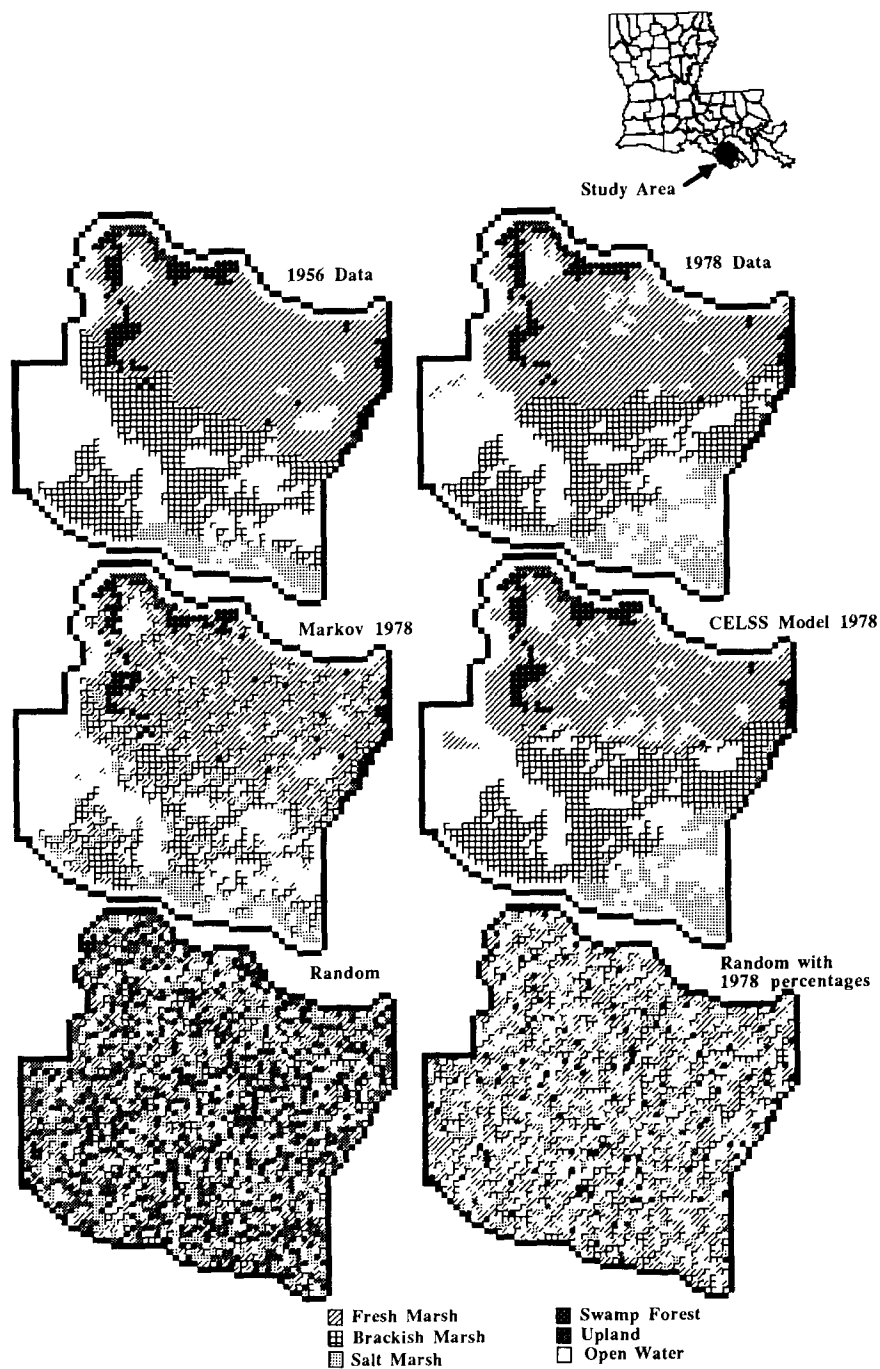


Fig. 1. Habitat distribution data and various model predictions for the Atchafalaya/Terrebonne area of coastal Louisiana. Each of the 2479 grid cells represents 1 km<sup>2</sup>.

one 'proper' resolution, but rather a range of resolutions is necessary to adequately describe the fit of models with reality.

For relatively simple purposes there are well-established statistical measures of goodness-of-fit (i.e.  $R^2$ ,  $\chi^2$ ). However, most ecological modeling involves quantifying the degree of matching or similarity between complex spatial and temporal patterns. Spatial pattern matching is not a straightforward statistical procedure. Even quantifying the degree to which nonspatial ecosystem modeling time series results match real patterns of ecosystem behavior is difficult and there is no universally agreed upon procedure (Gardner et al., 1980; Jørgensen, 1982). In spatial ecological analysis the literature has concentrated on determining if non-random patterns exist in point data (Ripley, 1981; Getis and Franklin, 1987), determining the optimum quadrat size based on an analysis of variance (Moellering and Tobler, 1972; Grieg-Smith, 1983), and determining whether or not two maps are statistically "different" (Pielou, 1977). But most statistical tests tend to miss or mask the details of spatial and temporal *patterns* in the data, concentrating instead on deviations from standard, random distributions. For example, a standard  $\chi^2$  test can be used to measure the fit between two land use maps, but it ignores the spatial pattern of the land use variables. It would yield information only about whether the total number of pixels in each land use category were significantly different between two maps, but not whether there was a significant difference in the pattern of two maps. The maps may both have the same total number of pixels in each land use category but be radically different in the arrangement of the pixels. The problem is more akin to the problem of matching amino acid sequences in large protein sequence data bases, which had been successfully attacked by developing "similarity scores" (Lipman and Pearson, 1985). The approach described in this paper extends the similarity scoring concept to include the possibility of performing the scoring over multiple resolutions.

The practical problems which led to the development of the procedures in this paper arose from work on spatial simulation modeling of long-term habitat succession in the Louisiana coastal zone (Sklar et al., 1985; Costanza et al., 1986, 1988). Some sample model results are shown in Fig. 1, which includes measured habitat distribution maps for 1956 and 1978 and various model predictions for 1978. We needed to gauge the models' performance in predicting the 1978 conditions from the 1956 conditions in a quantitative way that would allow parameter optimization, and we also needed an indication of the relative effectiveness of the various models. The simplest pattern sensitive goodness of fit metric under these circumstances is the percent of the cells that are correct (i.e., are the same type in both the model and data in a cell by cell comparison). The problem with this method was that it ignored 'near misses', and two maps with the same percent correct

could exhibit very different patterns in their residuals (misses). One might have random misses while the other might have systematic misses, for example at the boundaries between habitat types. We wanted a method that would give some weight to near misses as well as direct hits and tell us whether the 'pattern' was being relatively well matched. Lipman and Pearson's (1985) method of handling the 'near miss' problem was to weight comparisons based on the evolutionary likelihood of one category (amino acid) changing into another. This requires a priori information on which categories are 'most similar' to which other categories. In our case, it would have required us to determine, for example, that salt marsh is more similar to fresh marsh than it is to open-water habitat. To avoid having to make these somewhat arbitrary judgments, we developed an algorithm that uses spatial similarity in the pattern, rather than similarity between categories that must be supplied from other sources.

The method applies standard fit procedures over a number of temporal and/or spatial resolutions and uses the way the fit changes with resolution to interpret the model's performance. The procedure is first developed for categorical spatial data (the most difficult case) and then extended to interval spatial data, and to non-spatial time series data. It also borrows from ideas about edge detection as applied to problems of emulation of human vision (Rosenfeld and Kak, 1978; Marr and Hildreth, 1980) based on using moving two-dimensional 'windows' to filter the data (Spacek, 1986), but extends this idea to multiple resolutions.

Conceptually, this multiple resolution approach is similar to a "fractal dimension" (Mandelbrot, 1977, 1983) which measures the way the length of a coastline or other boundary changes as the resolution of measurement changes. For example, if we wished to determine the length of the coastline of Britain, the answer would depend on what resolution map we used (or the size of the fundamental unit of measurement). The smaller the fundamental unit of measurement (or the higher the resolution of the map) the longer the coastline measurement would be. Rather than speak of the 'length' of the coastline (which is resolution-dependent) we could alternatively speak of the way length depends on resolution. We can then develop resolution independent measures of the relationship between resolution and length. The fractal dimension is an elegantly simple example. It can be interpreted as a measure of the 'complexity' of boundaries.

The fundamental idea underlying fractals can be generalized and stated simply as: measurement at one particular resolution is always insufficient to describe complex natural phenomenon. It is necessary to measure phenomenon at several resolutions and interpret the way the results change with changing resolution to arrive at a meaningful description. The algorithm described in this paper implements this idea for model testing.

## COMPONENTS OF FIT

Lack of fit between a model and data (or between any two data series) can be partitioned among three components: registration, resolution, and residual. Lack of fit due to registration has to do with the failure of the model and data to 'line up' even though their patterns may be similar. For example, two identical pictures may show very little pixel by pixel correspondence if they are slightly rotated, if one is moved slightly horizontally or vertically, or if one is a different size. By rotating, sliding, expanding or contracting the pictures (or time series), the fit may be dramatically increased. Lack of fit due to misregistration can be removed by performing the above manipulations on the data until the fit is maximized. In analyzing two data maps, spatial misregistration can be an important problem, especially when the goal is to detect small changes between the two maps. In spatial modeling work, spatial misregistration is seldom a problem since the model is designed to spatially register with the data. Temporal misregistration can be an important problem in measuring fit, however, as we discuss further on.

Once the lack of fit due to misregistration has been removed, the resolution component of fit can be analyzed. This is described by way of example in the following sections.

## SPATIAL PATTERN MATCHING USING VARIABLE RESOLUTION FITTING

Two example  $10 \times 10$  scenes with four different categories of cover (Fig. 2) are used to describe the procedure. Looking at the two scenes we can discern that the patterns are 'similar' but not identical. There is a block of category-1 cells in the upper left corner of each scene, they both have a large block of category-3 cells covering most of the right side, etc. The patterns are also certainly non-random. Humans seem to have an innate ability to recognize these kinds of pattern similarities. The problem is to systematize this ability to the point that computer algorithms can approximate it.

Pattern recognition is scale-dependent (Allen and Starr, 1985). It depends on both resolution and extent (the size of the scene). For example, Fig. 3 presents two related scenes at different scales. The enlarged view (Fig. 3a) has higher resolution, lower extent and the general pattern is not apparent. In Fig. 3b resolution has been sacrificed for increased extent and the pattern is apparent. The increased extent is not necessary to discern the pattern, however. Decreasing the resolution is often enough, as can be seen by squinting at Fig. 3a or holding it at a distance so the effective resolution is decreased.

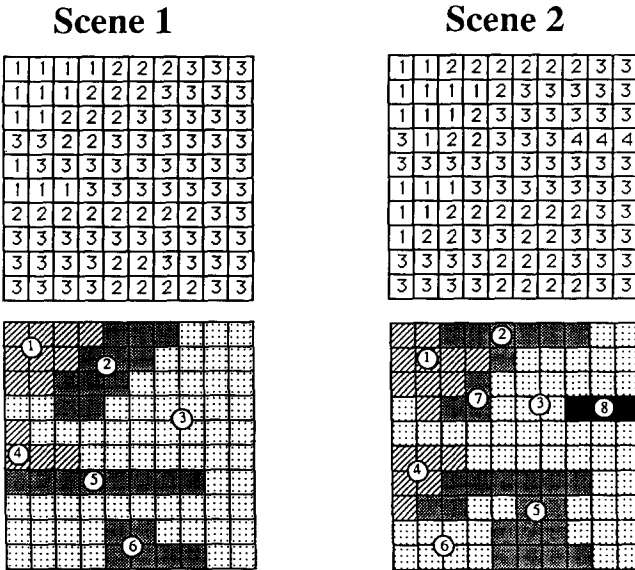


Fig. 2. Example 10×10 scenes with four categories.

The algorithm described in this paper decreases the resolution of comparisons by increasing the size of a sampling window that slides over scenes for which fit is calculated. The effect is similar to squinting to unfocus the scenes in order to perceive the larger pattern. Doing this with scene a makes the pattern apparent.

There are several alternative ways of measuring fit at a particular window size (resolution), depending on the nature of the data and other considerations. The multiple resolution approach does not depend on which measure of fit is used, as long as it is consistently applied at all resolutions. If numerical data are available, a standard coefficient of correlation ( $r^2$ ) can be used. Categorical data are more of a problem.

For a sampling window size of 1, a convenient measure of fit for categorical data is the proportion of cells accurately matched (80% in the example 10×10 scenes). As the sampling window size is increased, one would like to use an analogous measure of fit. Essentially, the data are being aggregated within each sampling window and there are several possible ways to perform this aggregation. The simplest approach is ‘proportional’ aggregation which chooses that category which is the highest proportion of the cells within the window, randomly deciding ties. This method tends to quickly eliminate rare categories, however. Most aggregation schemes replace all the cells within the sampling window with a single category, and result in a rapid loss of information about the relative proportions of the categories. For some purposes this loss of information is acceptable, but for

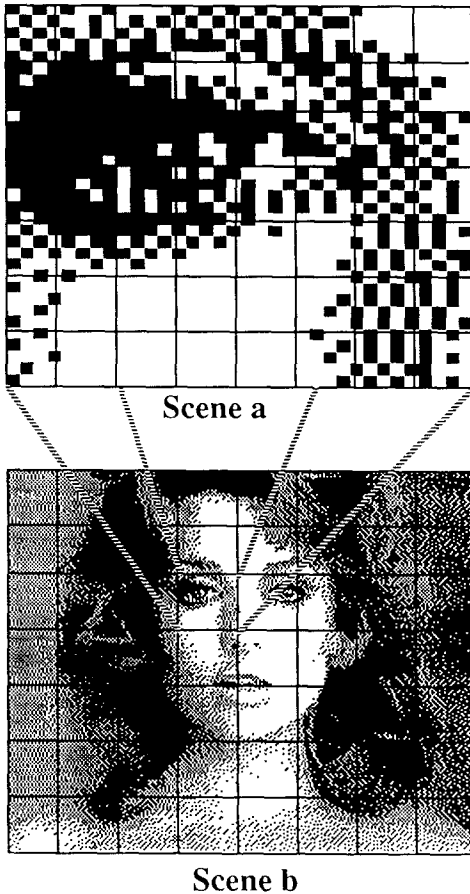


Fig. 3. Example showing the effects of resolution and extent on ability to recognize patterns. Scene a is a blowup of the right eye of what is obviously a picture of a woman in scene b, but it is only recognizable as such if we ignore some of the detail.

calculating fit, it is to be avoided if possible. A simple formula was devised that retains more of the information about the relative proportions of categories within the sample window in estimating fit while decreasing the resolution. The fit for each sampling window is estimated as 1 minus the proportion of cells that would have to be changed to make the sampling windows each have the same number of cells in each category, regardless of their spatial arrangement. For example, if a particular  $2 \times 2$  window had two cells of forest and two of marsh in both scenes, the fit would be 100% regardless of how the cells were arranged within the windows. If one sampling window had one forest and three marsh, while the other had two of each category, the fit would be 75% (since one cell out of four would have to be changed to make the fit 100%). The fit for the whole scene for a

particular sampling window size is the average fit over all the sampling windows of that size. The sampling window is moved through the scene one cell at a time until the entire image is covered.

The following formula for the fit at a particular sampling window size ( $F_w$ ) implements this idea for aggregation. It reduces to the percent correct if the sampling window size is 1 and retains information about relative category proportions as the window size is increased:

$$F_w = \frac{\sum_{s=1}^{t_w} \left[ 1 - \frac{\sum_{i=1}^p |a_{1i} - a_{2i}|}{2w^2} \right]}{t_w} \quad (1)$$

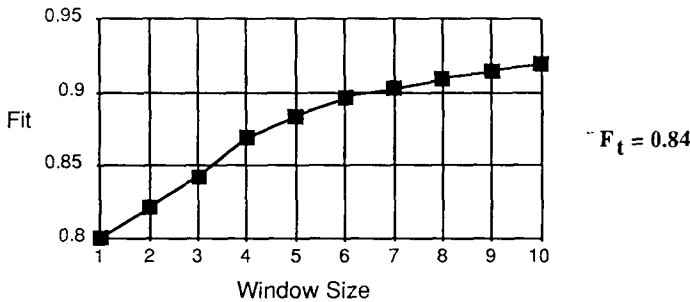
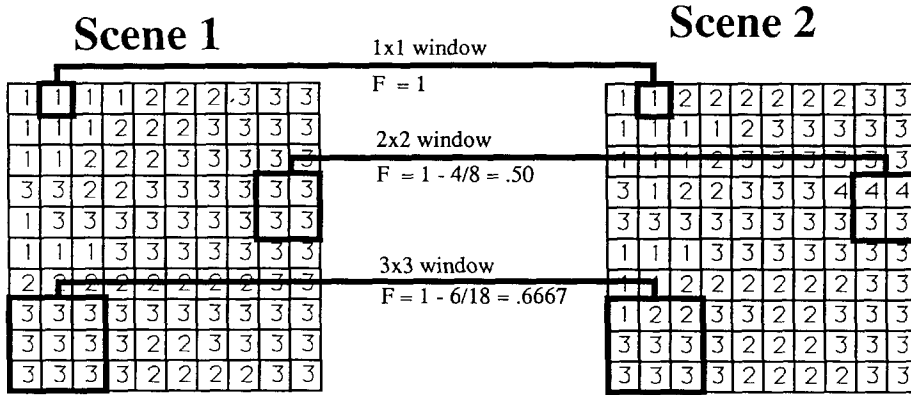


Fig. 4. Example of the multiple resolution fitting procedure for two sample 10×10 scenes with four categories. The plot shows that the two scenes match moderately well, as there is significant improvement in the fit with increasing window size. The total fit ( $F_t$ ) is an exponentially weighted average over all the window sizes with  $k = 0.1$ . See text for further explanation.



where  $F_w$  is the fit for sampling window size  $w$ ,  $w$  the dimension of one side of the (square) sampling window,  $a_{ki}$  the number of cells of category  $i$  in scene  $k$  in the sampling window,  $p$  the number of different categories (e.g., habitat types) in the sampling windows,  $s$  the sampling window of dimension  $w$  by  $w$  which slides through the scene one cell at a time, and  $t_w$  the total number of sampling windows in the scene for window size  $w$ .

One can then plot the fit between the scenes ( $F_w$ ) vs. the size of the sampling window ( $w$ ) as in Fig. 4. Using other aggregation methods or fit estimates would yield similar (but not identical) results, but the fundamental idea would still apply. If the plot behaves as it does for the two example scenes (it increases rapidly at first) then the pattern between the two scenes is very well matched even though the initial fit at window size 1 is relatively low. This would occur if the patterns between the scenes were similar, but the precise boundaries in the maximum resolution scenes were slightly off. Conversely, if the plot starts at a relatively low fit and is flat, then the spatial pattern is not well matched even though the initial fit might be higher than that for the two sample scenes. If the scenes were exactly identical, a perfect match over all window sizes would result. If two randomly generated scenes were compared, the expected fit would start at  $1/p$  and rapidly increase to 1 as the expanding sampling window made the statistical similarity between any two random scenes apparent.

To use these measures to determine an overall degree of fit between two maps the information in the plot of window sizes vs. fit must be summarized. A weighted average of the fits at different window sizes is one possible way of summarizing the overall fit that allows more weight to be given to smaller window sizes while not totally ignoring the large window sizes. For this purpose one can use the following formula:

$$F_t = \frac{\sum_{w=1}^n F_w e^{-k(w-1)}}{\sum_{w=1}^n e^{-k(w-1)}} \quad (2)$$

where  $F_t$  is a weighted average of the fits over all window sizes,  $F_w$  the fit for sampling windows of linear dimension  $w$ ,  $k$  a constant, and  $w$  linear dimension of a sampling window. This formula gives exponentially less weight to the fit at lower resolution. The value of  $k$  determines how much weight is to be given to small vs. large sampling windows. If  $k = 0$ , all window sizes are given the same weight. At  $k = 1$ , only the first few window sizes will be important. The relative importance of matching the patterns precisely vs. crudely must be answered in the context of the model's objectives and the quality of the data. From previous studies on aggregation

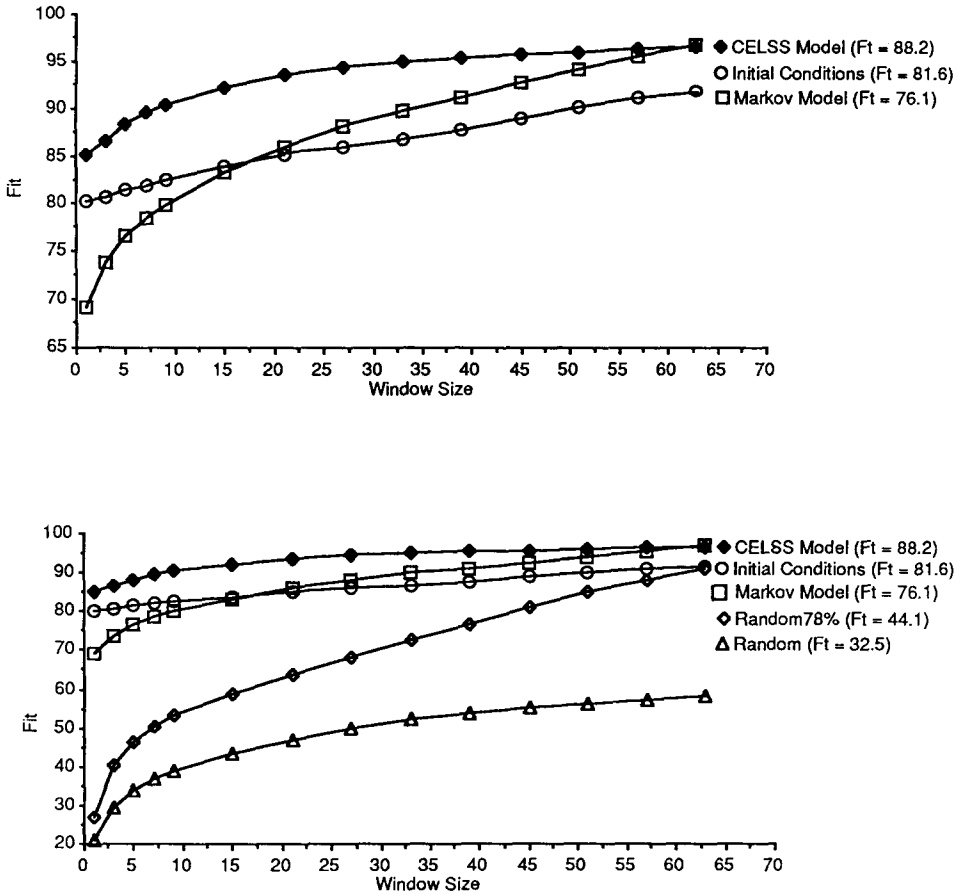


Fig. 5. Example plots of fit vs. window size for the various models in Fig. 1 compared to the 1978 data. Upper graph shows the CELSS model, the Markov model, and the initial (1956) conditions on an expanded  $y$  axis.

error in data and models (e.g., Gardner et al., 1982), we know that simple models may always be somewhat incorrect, but they may do a good job of matching the general patterns of the data, especially when there is high variance and uncertainty in the data. For the purposes of matching spatial patterns of land use, we have found that a value of  $k \approx 0.1$  gives an 'adequate' amount of weight to the larger window sizes. For the example in Fig. 2,  $k = 0.1$  yields  $F_t = 0.84$ .

The power of the technique in interpreting the patterns of correspondence between data and various models can be seen in Fig. 5, which shows the variable resolution procedure applied to the spatial modeling results summarized in Fig. 1. Fit vs. window size is plotted for the 1978 data compared with the 1978 CELSS simulation model prediction (labeled 'CELSS model'),

and several 'null models'. The most simple-minded null model is a completely random distribution of habitats with all of the categories being equiprobable (labeled 'random equiprobable' in the figure). A slightly more reasonable null model is a random distribution but with the same overall land cover probabilities (frequencies) as the 1978 data (labeled 'random with 1978 frequencies' in the figure). A simple Markov chain model is a slightly more sophisticated null model. It incorporates the statistical trend of habitat changes in an area in the form of transition probabilities (or frequencies). In our case, we use the frequencies of transition of each habitat type into each other type over the 1956–1978 interval. The fit over various window sizes for the Markov model is labeled 'Markov model' in the figure. Finally, the 1956 initial conditions represent a null model that predicts no change (labeled 'initial conditions' on the figure).

The CELSS simulation model performs better than any of the null models, but its performance can best be judged *in relation* to the null models. The simulation model fits the 1978 data better than 'random equiprobable' ( $F_t = 88.2\%$  vs.  $32.5\%$ ) and 'random with 1978 frequencies' ( $F_t = 88.2\%$  vs.  $50\%$ ). The fit for 'random with 1978 frequencies' approaches one as window size increases indicating that the overall percentages of habitat types are the same but the pattern is not well matched.

Compared to the 'initial conditions' the CELSS model might not appear to be significantly better if only the fit at window size 1 is considered ( $85\%$  vs.  $81\%$ ). The initial conditions model appears to fit fairly well because only about 20% of the cells actually changed type between 1956 and 1978. But the pattern of fit is revealed by looking at the plots of fit vs. window size. The simulation model plot increases rapidly as window size increases indicating that the 'pattern' between the model and data is well matched. The initial conditions model exhibits a flatter plot indicating the pattern is not well matched. Total weighted fit ( $F_t$ ) is significantly higher for the CELSS model than the initial conditions null model ( $88.2\%$  vs.  $81.6\%$ ).

The Markov model has a lower initial fit ( $69\%$ ) than the initial conditions model ( $81\%$ ), but the fit increases more rapidly as the window size increases. At the maximum window size it fits better than the initial conditions model since the total number of cells of each type is closer to the 1978 data. Total weighted fit for the Markov model is less than the initial conditions model ( $81.6\%$  vs.  $76.1\%$ ) since by randomly placing the transition cells the detailed spatial pattern is disrupted while increasing the fit at larger window sizes.

#### STATISTICAL SIGNIFICANCE

Since the  $F_t$  values generated by our procedure do not follow a normal distribution, one cannot calculate a statistical probability ( $P$  value). The

most satisfactory approach at present seems to be to generate  $z$  values following the method of Lipman and Pearson (1985), where:

$$z = \frac{(F_i / \text{mean of randomly permuted } F_i)}{\text{standard deviation of randomly permuted } F_i}$$

The randomly permuted  $F_i$  values are generated by randomly rearranging one of the maps while retaining the same total number of pixels in each category. At least 20 random permutations should be used and Lipman and Pearson (1985) suggest the following significance guidelines:

- $z > 3$ : possibly significant
- $z > 6$ : probably significant
- $z > 10$ : significant.

The results would be similar to comparing the CELSS model 1978 map (Fig. 1) to the Random with 1978 percentages map for several random permutations to test for significance of the fit with the 1978 data map. Because of the large number of cells involved and the highly patterned nature of the data maps, the mean of the randomly permuted  $F_i$  will generally be much lower than  $F_i$  (0.882 vs. 0.50 yielding a ratio of about 1.7) and the standard deviation of the randomly permuted  $F_i$  will be relatively low (around 0.1) yielding  $z$  scores much greater than 10 and highly significant comparisons.

#### TEMPORAL PATTERN MATCHING

The multiple resolution procedures are also applicable to time series rather than spatial data. In analysing time series data the sampling 'window' takes in progressively more time rather than progressively more space. In both cases, we are interested in summarizing the effects of 'smoothing' or averaging the data on goodness of fit. The procedures are also applicable to data that are ordinal, interval, or ratio rather than categorical (or nominal), by simply altering the formula for the fit at a particular window size (equation 1). For ratio and ordinal data the coefficient of correlation ( $r^2$ ) can be substituted for  $F_w$  in equation (1).

An example similar to one that might occur in ecological modeling work is given in Fig. 6a, which shows hypothetical 'data' and 'model' time series. In this example the data are ratio values (meaning relative to some absolute zero, for example biomass or  $K$ ) and the fit is calculated as the coefficient of determination between the data and the model ( $r^2$ ).

In this example the effects of misregistration can be clearly seen. Fig. 6b is a plot of the data compared to the model output lagged by two time steps. Fig. 7a is a plot of fit versus window size (in this case the number of intervals in a moving average of the data) for the unlagged data (Fig. 6a).

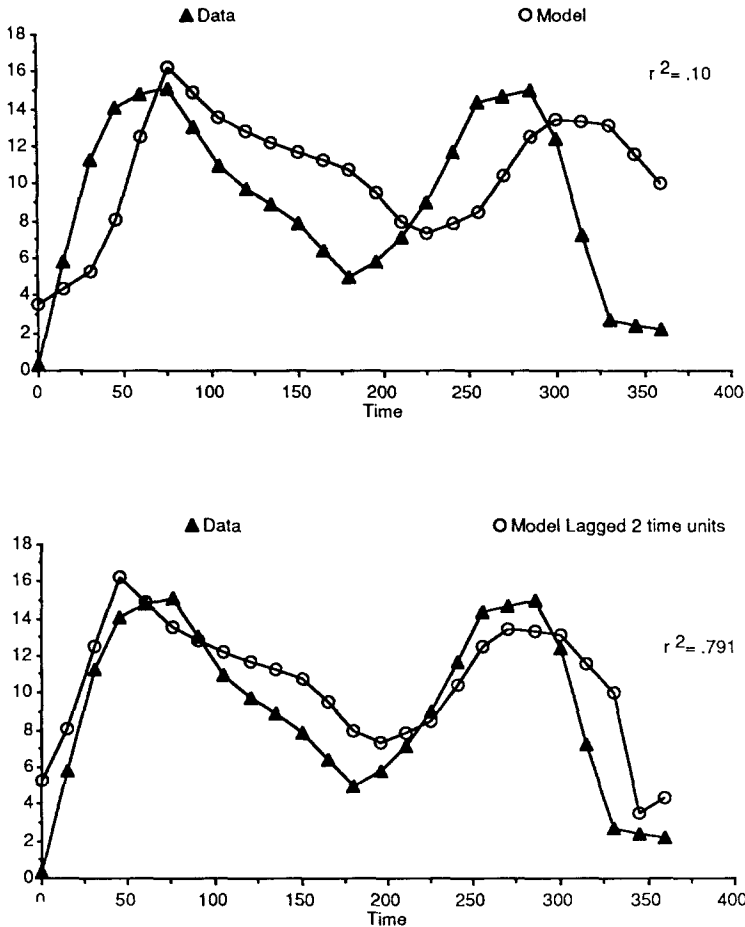


Fig. 6. Hypothetical data and model curves (top) and the same curves with the model lagged by two points.

Even though it is obvious (to human eyes) that the model and data curves in Fig. 6a are quite 'similar', the fit starts at 0.1 and decreases as the window size increases to a minimum of 0.02 before increasing again to 0.08 at a window size of 11. This occurs because there is a slight temporal misregistration in the two series. Fig. 7b is a plot of fit vs. lag for the two series. A lag of two time units increases the fit from 0.1 to almost 0.8. Figure 8 shows the fit vs. window size calculations performed after the model had been lagged by 2 units.  $F_t$  for this plot (with  $k = 0.1$ ) is 0.804.

We need to subtract something for the model's failure in this example to get the timing right, making it necessary to lag the model by 2 units. Again, the amount to subtract is subjective and depends on how important it is to the model's purpose to precisely match the timing of the data. Small

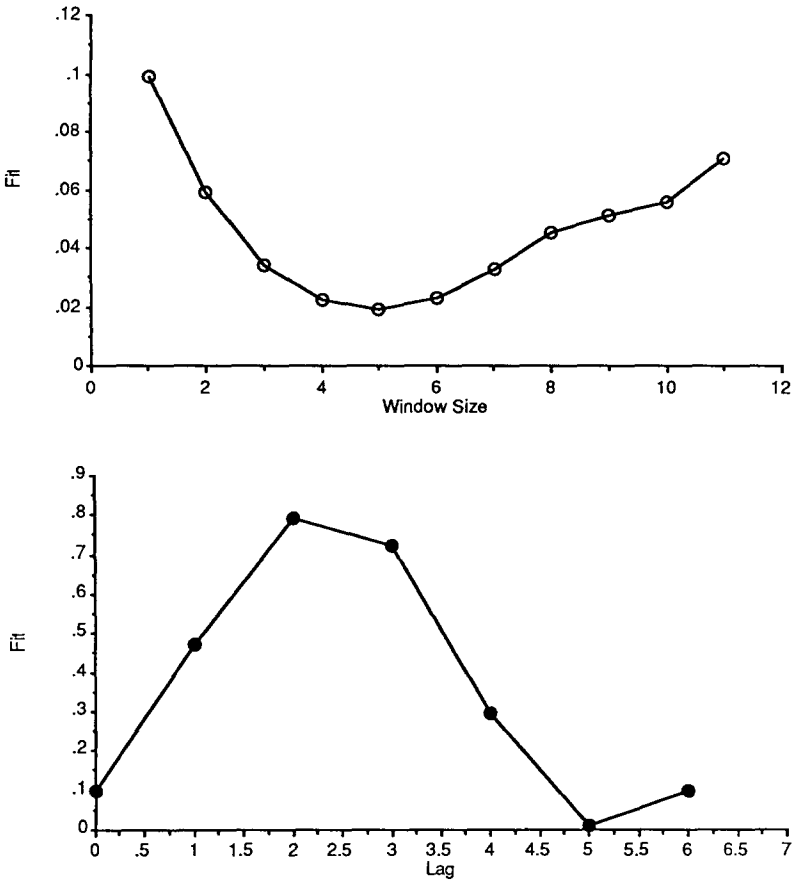


Fig. 7. Fit vs. window size of a moving average of the unlagged data and model from Fig. 6 (top) and fit vs. lag for the same data (bottom).

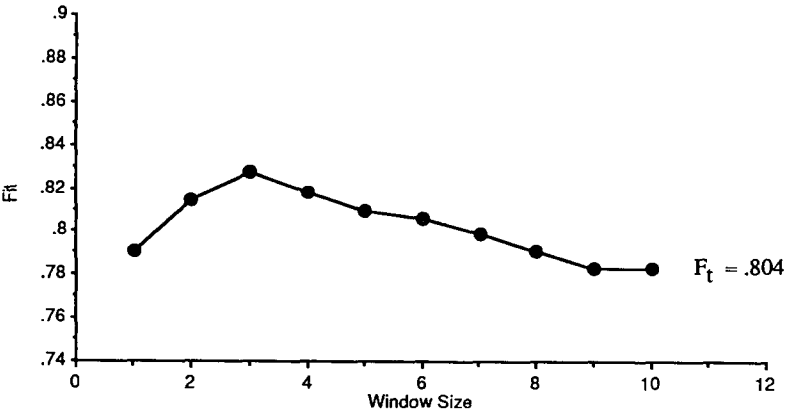


Fig. 8. Fit vs. window size for the data and model output from Fig. 6 after lagging the model by two time units.

misregistrations (low number of lags) are probably less important than large lags, so one approach would be to subtract successively more as the required lag increased. We might apply the following formula for a weighted fit over lag and window size:

$$F_{tL} = F_t e^{-k_t L} \quad (3)$$

where  $F_{tL}$  is the total weighted fit over lag and window size,  $k_t$  is the weighting factor for lag,  $L$  is the lag required to maximize the fit, for example, if  $F_t = 0.804$ ,  $L = 2$ , and  $k_t = 0.05$  then  $F_{tL} = 0.728$ .

## DISCUSSION AND CONCLUSIONS

A simple procedure for determining the overall fit between model output and data that incorporates the effects of temporal and spatial misregistration and allows a degree of pattern matching by varying the resolution of the fit has been developed. Subjective weighting factors for the importance of precise vs. general pattern matching ( $k$ ) and the importance of precise registration ( $k_L$ ) must be supplied by the user, based on the purposes of the proposed model.

The procedures have many possible applications in modeling work, and also in descriptive pattern analysis. For example, one potential use of the procedures is analysis of the degree of organization of a landscape or other pattern. Organization can be defined as the degree of departure from a random pattern. While other measures (i.e. fractal dimension) yield information about the degree of departure from a random pattern, they are not sensitive to specific spatial patterns in the data. The multiple resolution fitting procedure described above can be used to directly measure the departure from a random pattern by measuring the degree of fit between a specific landscape pattern and a random one. For example, for the scenes in Fig. 2, the weighted average fits ( $F_t$ ) were 0.445 for scene 1 and 0.517 for scene 2 when compared with a randomly generated pattern (using  $k = 0.1$ ). This indicates that scene 1 is slightly more organized (farther from random, less complex, more predictable) than scene 2. The advantage of this measure over others is that it is sensitive to aperiodic patterns or clusters and is insensitive to the complexity of the patterns. This is an advantage since it is possible to have complex, aperiodic patterns that are highly non-random. The Atchafalaya/Terrebonne marsh landscape in Fig. 1 has a total weighted fit with a random scene of 0.325, indicating it is fairly highly organized.

The multiple scale procedures allow a much more sophisticated description of patterns and of the way models (particularly spatial models) fit the data. Ecological modeling cannot progress without procedures that allow

better assessment of the relative successes of the models. The procedures outlined here are a modest step in that direction.

#### ACKNOWLEDGEMENTS

Monica G. Turner and Fred H. Sklar provided many useful ideas, criticism and support. Fred Sklar wrote the original computer code to implement the multiple resolution fitting procedure (which is available from the author on request in Basic or Fortran). Mary White and Tamer Ozmen aided in computer programming.

This research was funded in part by the U.S. Fish and Wildlife Service (a division of the U.S. Department of Interior) under Cooperative Agreement # 14-16-0009-84-921 titled: Spatial Simulation Modeling of Coastal Wetland Systems for Evaluating Management Alternatives, and by the National Science Foundation under Grant # BSR-8814272 titled: Responses of a Major Land Margin Ecosystem to Changes in Terrestrial Nutrient Inputs, Internal Nutrient Cycling, Production and Export.

Computer time on the CRAY supercomputer at the University of Illinois to run the CELSS model and the multiple resolution procedures were provided by the National Center for Supercomputer Applications and by an NSF grant on modeling landscape dynamics, Don Jameson, Principal Investigator.

#### REFERENCES

- Allen, T.F.H. and Starr, T.B., 1982. *Hierarchy: perspectives for ecological complexity*. University of Chicago Press, Chicago, IL.
- Costanza, R., Sklar, F.H. and Day, J.W., Jr., 1986. Modeling spatial and temporal succession in the Atchafalaya/Terrebonne marsh/estuarine complex in South Louisiana. In: D.A. Wolfe (Editor), *Estuarine Variability*. Academic Press, New York, pp. 387–404.
- Costanza, R., Sklar, F.H., White, M.L. and Day, J.W., Jr., 1988. A dynamic spatial simulation model of land loss and marsh succession in coastal Louisiana. In: W.J. Mitsch, M. Straškraba and S.E. Jørgensen (Editors), *Wetland Modelling. Developments in Environmental Modelling*, 12. Elsevier, Amsterdam, pp. 99–114.
- Gardner, R.H., O'Neill, R.V., Mankin, J.B. and Kumar, D., 1980. Comparative error analysis of six predator–prey models. *Ecology*, 61: 323–332.
- Gardner, R.H., Cale, W.G. and O'Neill, R.V., 1982. Robust analysis of aggregation error. *Ecology*, 63: 1771–1779.
- Getis, A. and Franklin, J., 1987. Second-order neighborhood analysis of mapped point patterns. *Ecology*, 68: 473–477.
- Grieg-Smith, P., 1983. *Quantitative Plant Ecology*. Blackwell, Oxford.
- Jørgensen, S.E., 1982. Modelling the eutrophication of shallow lakes. In: D.O. Logofet and N.K. Luckyanov (Editors), *Ecosystem Dynamics in Freshwater Wetlands and Shallow Water Bodies*, 2. UNEP/SCOPE, USSR Academy of Sciences, Moscow, pp. 125–155.



- Lipman, D.J. and Pearson, W.R., 1985. Rapid and sensitive protein similarity searches. *Science*, 227: 1435–1441.
- Mandelbrot, B.B., 1977. *Fractals. Form, Chance and Dimension*. Freeman, San Francisco, CA.
- Mandelbrot, B.B., 1983. *The Fractal Geometry of Nature*. Freeman, San Francisco, CA.
- Marr, D. and Hildreth, E., 1980. Theory of edge detection. *Proc. R. Soc. London B*, 207: 187–212.
- Moellering, H. and Tobler, W.R., 1972. Geographical variances. *Geogr. Anal.*, 4: 35–50.
- Pielou, E.C., 1977. *Mathematical Ecology*. Wiley, New York.
- Ripley, B.D., 1981. *Spatial Statistics*. Wiley, New York.
- Rosenfeld, A. and Kak, A.C., 1978. *Digital Image Processing*. Wiley, New York.
- Sklar, F.H., Costanza, R. and Day, J.W., Jr., 1985. Dynamic spatial simulation modeling of coastal wetland habitat succession. *Ecol. Modelling*, 29: 261–281.
- Spacek, L.A., 1986. Edge detection and motion detection. *Image Vision Comput.*, 4: 43–56.