Ecological Modelling, 62 (1992) 275–293
Elsevier Science Publishers B.V., Amsterdam

Comparing global vegetation maps with the Kappa statistic

Robert A. Monserud [a] and Rik Leemans [b]

[a] *Intermountain Research Station, Forest Service, U.S. Department of Agriculture, 1221 S. Main St., Moscow, ID 83843, USA*

[b] *Global Change Department, National Institute of Public Health and Environmental Protection, P.O. Box 1, 3720 BA Bilthoven, Netherlands*

(Accepted 16 December 1991)

ABSTRACT

Monserud, R.A. and Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. *Ecol. Modelling*, 62: 275–293.

The Kappa statistic is presented as an objective tool for comparing global vegetation maps. Such maps can result from either compilations of observed spatial patterns or from simulations from models that are global in scope. The method is illustrated by comparing global maps resulting from applying a modified Holdridge Life Zone Classification to current climate and several climate change scenarios (CO_2 doubling). These scenarios were based on the results of several different general circulation models (GCMs). The direction of change in simulated vegetation patterns between different GCMs was found to be quite similar for all future projections. Although there were differences in magnitude and extent, all simulations indicate potential for enormous ecological change. The Kappa statistic proved to be a useful and straightforward measure of agreement between the different global vegetation maps. Furthermore, Kappa statistics for individual vegetation zones clearly indicated differences and similarities between those maps. The Kappa statistic was found to be most useful for rank ordering of agreement, both across a series of maps and across the various vegetation zones within a map.

INTRODUCTION

Predicting ecosystem response to climate change is currently one of the major issues in ecological research (Walker and Greatz, 1989; Houghton et al., 1990). Current approaches for assessing such response can be charac-

Correspondence to: R.A. Monserud, Intermountain Research Station, Forest Service, U.S. Department of Agriculture, 1221 S. Main St., Moscow, ID 83843, USA.

terized as either static or dynamic modelling. Static modelling assumes equilibrium conditions. Essentially, it is a modern implementation of classic vegetation–climate classifications (e.g., Emanuel et al., 1985a, b; Guetter and Kutzbach, 1990). The dynamic models, which are based on species characteristics and individual plant responses, try to capture the transient response of vegetation to a changing climate (Shugart, 1990). The static models are used on a global scale, presenting large-scale vegetation distributions, while the dynamic approach is used to assess the transient response on smaller scales. A promising approach for assessing vegetation response to a changing climate is to combine both types of models. Static models determine shifts in the large-scale vegetation patterns, while dynamics models are used to determine local vegetation response in the shifting regions. The static models are thus spatially explicit, while the dynamic models are needed for temporal realism.

Recent research by the Biosphere Dynamics Project at the International Institute of Applied Systems Analysis (IIASA) has produced maps of global vegetation patterns (Leemans, 1989; Prentice et al., 1989; Solomon and Leemans, 1990). These maps are primarily based on static models. They have been implemented using gridded global data bases with relevant climatic and physical data for the land surface of the earth. Because of the complexity of the patterns displayed on those maps, it is difficult to objectively compare any two such maps. The focus of this paper is on the development of objective statistics for summarizing differences and similarities between global vegetation maps.

GLOBAL VEGETATION MAPS

Climate and vegetation are closely related, especially on a global scale (Budyko, 1986). This observation allows us to define global vegetation patterns in a climatic parameter space. Global vegetation maps are obtained by first stating a hypothesis relating climatic factors to vegetation and then solving and plotting the outcome of that hypothesis for a network of points that represent the climate of the earth's land surface. This can be done in a straightforward manner for many different climates, including reconstructed past climates, observed current climates, and projected future climates. Absolutely necessary for this type of analysis is a series of data bases that associate the relevant climatic factors with all locations in the network of points or cells. We use a dedicated Geographic Information System (GIS) that provides the storage and linkage structure of all data bases and enables analysis, overlaying and comparisons of maps. Additional tasks of this GIS are map construction, plotting and data base development.

Geographic climate data base

Leemans and Cramer (1991) describe the development of a gridded global data base for terrestrial climate. The data base contains average monthly values for temperature (°C), precipitation (mm) and percentage cloudiness and is obtained by interpolation of a selected series of ca. 7500 long-term weather records worldwide. The resulting temperature values are corrected for topography using a lapse-rate correction scheme (Strahler and Strahler, 1987). The resolution of this data base is 0.5° latitude by 0.5° longitude (cell size of approximately 55 × 55 km at the equator). The total network contains 62483 land pixels. The additional cells represent oceans, large water bodies and Antarctica and are ignored because they are "structural zeros" (Bishop et al., 1975) in any analysis comparing change in terrestrial vegetation. The data base is mainly used as a basis for displaying global patterns of climatic indices using GIS techniques.

Holdridge life zone classification

We use the Holdridge (1947, 1967) Life Zone Classification to illustrate the procedures developed in this paper. Holdridge held that the natural vegetation in an area could be determined objectively by local climate. He defined 39 life zones using three climatic parameters: biotemperature, mean annual precipitation, and potential evapotranspiration (PET) ratio (Fig. 1A). Biotemperature is defined as the mean of monthly (or daily) positive temperatures. Evapotranspiration is here simply a linear function of biotemperature. PET ratio is the ratio of evapotranspiration to mean annual precipitation. Holdridge used biotemperature to delineate latitudinal and elevational zones and the PET ratio to differentiate humidity provinces. Finally, a strong geometric structure was imposed on the three indices, so that the life zones could be displayed as hexagons of constant size in a two-dimensional triangular space (Fig. 1A).

Even though the Holdridge Life Zone hypothesis is simplistic, it has nevertheless proven useful in elucidating both the importance and limitations of climate as a determinant of vegetation (Prentice, 1990). This simplicity is probably the main attraction of the hypothesis, for it requires only data that is generally available. Several different global implementations of the Holdridge Life Zone Classification exist (e.g., Emanuel et al., 1985a, b; Henderson-Sellers, 1991; Leemans, 1989; Prentice, 1990; Prentice and Fung, 1990).

Visually, only partial agreement is observed between the Holdridge Life Zone Classification and the vegetation map of Olson and Watts (1982). To improve this agreement the original 39 Life Zones of Holdridge were aggregated into larger units (Table 1 and Fig. 1A). The resulting 14 vegetation zones more closely represented the major biomes and were therefore used as the biogeographical units in this study.

Climate change

The Holdridge Life Zone Classification is intrinsically static. Vegetation is viewed as responding immediately to a change in climate. With such a naive viewpoint, vegetation is not seen to have any transient response, any feedback or delay, any dynamics. In spite of these limitations, the approach nevertheless has some utility in the absence of a workable dynamic alternative for the global scale. Because the Holdridge classification system uses only basic climatic variables that are generally available (temperature, precipitation), it is straightforward to predict a vegetation response to any climate scenario that can be expressed with those variables.

General circulation models (GCMs) of the atmosphere have become increasingly popular tools for predicting the climatic response to a variety of global atmospheric disturbances. GCMs attempt to numerically simulate the dynamics of the atmosphere, coupled with the surface water and energy balances (Harrison, 1990). After dividing the earth's surface and atmosphere vertically into strata and horizontally into grid cells and then specifying initial conditions, the equations of state are simultaneously solved for all cells in all strata, while constraining for conservation of energy and momentum (Hansen et al., 1983). Basically this amounts to solving the Navier–Stokes equations for the movement of a fluid around a sphere.

In this study, we relied primarily on GCM predictions from the GFDL model (developed at the Geophysical Fluid Dynamics Laboratory of NOAA at Princeton) to estimate the global climatic response to a doubling of CO_2 (Wetherald and Manabe, 1986; Manabe and Wetherald, 1987).

Construction of Holdridge Life Zone maps

Climatic output from the GFDL general circulation model was used to create Holdridge Life Zone maps for doubled CO_2 climate. Several steps are needed to create the global vegetation-change scenarios:

(1) A complete global implementation of the Holdridge Life Zone Classification is created using the current climate data base (Leemans and Cramer, 1991). The resulting 39 classes are then aggregated iteratively into 14 classes (Table 1) by trying to most closely resemble the patterns shown on the Olson et al. (1983) map of current global vegetation (Anonymous, 1990). The result is the current climate vegetation map used in subsequent map comparisons (Fig. 1B).

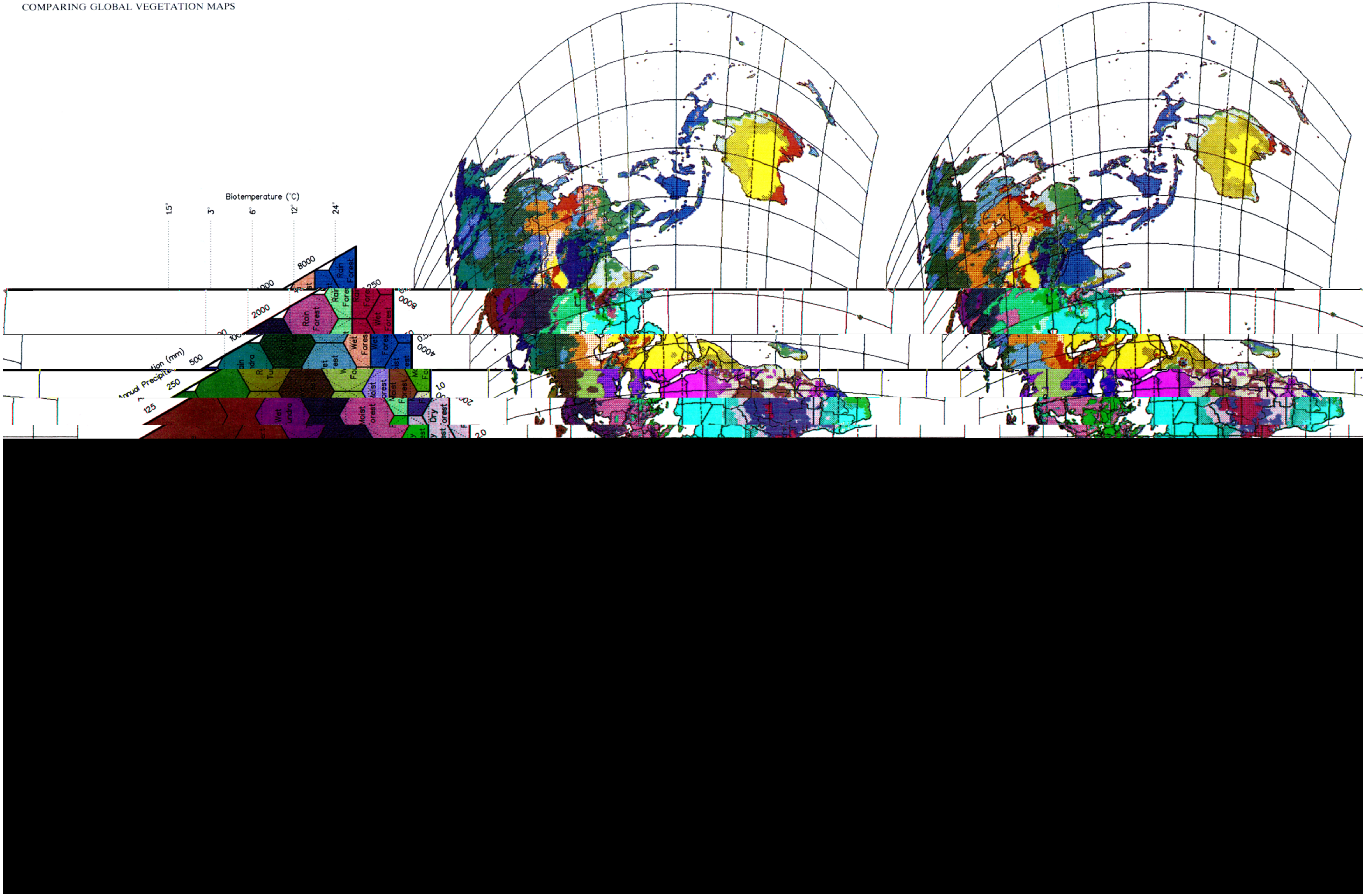

Fig. 1. (A) The Holdridge Life Zone Classification. The color indicates the clustering used in the Holdridge Life Zone world map. (B) The Holdridge Life Zone world map. This vegetation map is derived from current climate. (C) The Holdridge Life Zone map for climate change. This world map is derived from a double CO_2 climate scenario as predicted by the GFDL general circulation model.

TABLE 1

The scheme for aggregating the Holdridge Life Zones into the vegetation zones used in this study. Areas are given for the simulation using current climate

Holdridge Life Zone	Vegetation zone	Area (1 000 km^2)
Polar Ice		
Polar Desert	Tundra	1 037
Subpolar Moist Tundra		
Subpolar Wet Tundra		
Subpolar Rain Tundra	Forest Tundra	886
Subpolar Dry Tundra		
Boreal Desert		
Boreal Dry Scrub	Cold Parklands	281
Boreal Moist Forest		
Boreal Wet Forest		
Boreal Rain Forest	Boreal Forest	1 512
Cool Temperate Moist Forest		
Cool Temperate Wet Forest		
Cool Temperate Rain Forest	Temperate Forest	998
Cool Temperate Steppe	Steppe	741
Cool Temperate Desert		
Cool Temperate Desert Scrub	Cool Desert	402
Warm Temperate Moist Forest		
Warm Temperate Wet Forest		
Warm Temperate Rain Forest	Warm Temperate Forest	322
Warm Temperate Thorn Steppe		
Warm Temperate Dry Forest	Chapparal	563
Warm Temperate Desert		
Warm Temperate Desert Scrub		
Subtropical Desert		
Subtropical Desert Scrub		
Tropical Desert		
Tropical Desert Scrub	Hot Desert	2 085
Subtropical Thorn Woodland		
Tropical Thorn Woodland		
Tropical Very Dry Forest	Tropical Semi-Arid	953
Subtropical Dry Forest		
Tropical Dry Forest	Tropical Dry Forest	1 485
Subtropical Moist Forest	Tropical Seasonal Forest	1 508
Subtropical Wet Forest		
Subtropical Rain Forest		
Tropical Moist Forest		
Tropical Wet Forest		
Tropical Rain Forest	Tropical Rain Forest	846

(2) Using the GFDL general circulation model (as well as other GCMs), the differences between the control run and the doubled CO_2 run are determined. These differences are then interpolated from the original GCM grid to the somewhat finer 0.5° grid used in the IIASA studies. The absolute value of the temperature differences and the ratio of the precipitation estimates are calculated for each pixel.

(3) The changed climate data base is created by adding the temperature differences to the current climate data base and by multiplying precipitation by the precipitation ratio predicted by the GCM runs.

(4) A new world map with an aggregated Holdridge Life Zone Classification (14 classes) is determined for the GCM scenario by using the changed climate data base (Fig. 1C).

COMPARING GLOBAL VEGETATION MAPS

When the resulting maps are plotted in color, it is possible to visually examine and compare the maps for differences (cf. Figs. 1B and 1C; see also Anonymous, 1990). This is nearly impossible in black and white (cf. Leemans, 1989). Comparison in color is nevertheless difficult and tedious, for the maps are quite complex, even when the number of vegetation zones is as few as 14. An additional problem becomes apparent: because there are so many pixels (62483), the map examiner usually compares only small subsets of points, points subjectively chosen because they represent regions with which the examiner is familiar. The result of such a comparison is often a subjective judgment based on incomplete information. The need for an objective measure of agreement between two given maps is obviously great.

Statistical considerations

Costanza (1989) and Turner et al. (1989) have recently developed new approaches for comparing patterns in spatial ecosystems models. Their main goal was to develop quantitative methods for comparing spatial patterns in order to evaluate the performance of such spatial models. Although their methods are directly applicable to global map comparisons, there is a major problem. Their derived indices are all defined for a continuous grid. World maps contain both land and ocean grid cells, whereby the latter are treated as missing values. The stable border between continents and oceans heavily influences the value of their proposed indices and make them of little value for comparing global terrestrial patterns.

There is a large and growing literature for analyzing spatial patterns and spatial processes (e.g., Pielou, 1977; Ripley, 1981; Cliff and Ord, 1981;

Gaile and Willmott, 1984). Much of this literature concentrates on the deviations from standard — mostly random — patterns and is devoted almost exclusively to answering the following question: "What underlying process could have produced this map or spatial pattern?" Surprisingly, the answer is trivial when doing spatial modeling within the framework of a GIS: the underlying process is known exactly. It is the interaction of the spatial hypothesis and the relevant features of the terrain programmed into the GIS. The map is a picture of this interaction.

Thus, the real question of interest involves the comparison of two maps generated by known processes or map series obtained from repeated measurements. This is properly a question of agreement — pixel by pixel agreement. Bishop et al. (1975, p. 394) explain that the distinction between agreement and association for nominal data is that for two responses to agree they must fall into the identical category, while for two responses to be perfectly associated it is only necessary to be able to predict the category of one response from the category of the other. A table displaying paired responses may exhibit high association along with either high or low agreement.

Although much work has been done on various measures of association, the literature on judging agreement is quite small. In their seminal work on measures of association, Goodman and Kruskal (1954) considered agreement to be a special case. Given a table displaying the results of two observers assigning each of N items into one of c categories, the categories for rows in a table of agreement must appear in exactly the same order as the categories for columns. This simple restriction gives meaning to the main diagonal of any agreement table (Bishop et al., 1975).

The Kappa statistic

The contributions of Goodman and Kruskal (1954) notwithstanding, the seminal work on agreement is Cohen (1960). Consider the following table

The main diagonal contains the proportions of observed agreement between the two maps for each category. Their sum is the overall proportion of observed agreement:

$$p_0 = \sum_{i=1}^{c} p_{ii}$$

Although p_0 is the simplest and most frequently used index of agreement (it is often called an *intraclass correlation coefficient*), it is not without problems (Fleiss, 1981). It is reasonable to expect that some degree of agreement will occur by chance alone. Cohen (1960) discovered a natural means for correcting for chance. Observing that the marginal totals contain information about the magnitude of chance agreement, Cohen calculated the overall proportion of chance-expected agreement:

$$p_e = \sum_{i=1}^{c} p_{i.}p_{.i}$$

that occurs if the rows are independent of the columns. Although the difference $p_0 - p_e$ is a useful measure of agreement, Cohen (1960) improved it by normalizing by the largest possible value for the given marginal totals (namely, $1 - p_e$). The resulting statistic is called *Kappa*:

$$\hat{\kappa} = \frac{p_0 - p_e}{1 - p_e}$$

Kappa has desirable properties. It takes on a value of 1 with perfect agreement ($p_0 = 1$). It has a value close to zero when the observed agreement is approximately the same as would be expected by chance ($p_0 \approx p_e$). In addition, the Kappa statistic does not assume that the marginal probabilities are equal for the two observers or maps.

An individual $\hat{\kappa}_i$ can also be calculated for each of the c categories. A straightforward way to think about calculating individual Kappas is to partition the overall matrix of proportions into the following 2 by 2 matrix with only two categories: i and not-i (it is easiest to think of category i as the first one, especially since their ordering is arbitrary).

$$\begin{array}{c|ccc}
p_{11} & p_{12} & \cdots & p_{1c} \\
\hline
p_{21} & p_{22} & \cdots & p_{2c} \\
\vdots & \vdots & \ddots & \vdots \\
p_{c1} & p_{c2} & \cdots & p_{cc}
\end{array}$$

This matrix reduces to the following 2 by 2 table for measuring agreement on a single category:

Map A	Map B		Total
	Category i	All others	
Category i	p_{ii}	$p_{i.}-p_{ii}$	$p_{i.}$
All others	$p_{.i}-p_{ii}$	d	$1-p_{i.}$
Total	$p_{.i}$	$1-p_{.i}$	1

where $d = 1 - p_{i.} - p_{.i} + p_{ii}$. The main attraction of this formulation is that it quickly produces d (the large number of proportions that both maps categorize as other than i) by subtraction, since the column and row totals ($p_{.i}$ and $p_{i.}$) are easy to calculate. The following formula (see p. 217 in Fleiss, 1981, but with a different notation) can then be used to calculate the Kappa statistic for category i:

$$\hat{\kappa}_i = \frac{2[p_{ii}d - (p_{i.}-p_{ii})(p_{.i}-p_{ii})]}{p_{i.}(1-p_{.i}) + p_{.i}(1-p_{i.})}$$

Fleiss' formula can be reduced to the following more intuitive shorter form:

$$\hat{\kappa}_i = \frac{p_{ii} - p_{i.}p_{.i}}{(p_{i.}+p_{.i})/2 - p_{i.}p_{.i}}$$

An additional desirable property is that the overall value of $\hat{\kappa}$ is also equal to a weighted average of the individual $\hat{\kappa}_i$'s. Dividing the sum (over all categories) of the numerators in the preceding formula by the sum of the denominators yields the overall Kappa statistic:

$$\hat{\kappa} = \frac{\sum_{i=1}^{c} (p_{ii} - p_{i.}p_{.i})}{\sum_{i=1}^{c} [(p_{i.}+p_{.i})/2 - p_{i.}p_{.i}]}$$

$$= \frac{\sum_{i=1}^{c} p_{ii} - \sum_{i=1}^{c} p_{i.}p_{.i}}{\sum_{i=1}^{c} (p_{i.}+p_{.i})/2 - \sum_{i=1}^{c} p_{i.}p_{.i}}$$

$$= \frac{p_0 - p_e}{1 - p_e}$$

Because the asymptotic sample variance of $\hat{\kappa}$ has been derived, it is straightforward to do hypothesis testing with Kappa (see excellent summary by Fleiss, 1981, Chapter 13). This is rarely an interesting or useful way to compare two maps, however, because of the rather large sample sizes involved. (With $N = 62\,483$ almost *any* two global maps will be significantly different.)

A much more useful way to use Kappa for map comparison is provided by Landis and Koch (1977). They have characterized different ranges of $\hat{\kappa}$ based on the degree of agreement that they suggest. Values greater than approximately 0.75 indicate very good to excellent agreement (1.0 is perfect agreement), values between 0.4 and 0.75 indicate fair to good agreement, and values of 0.4 or less indicate poor agreement. Values close to 0.0 mean that the agreement is no better than would be expected by chance. Although it is possible to have a minimum value that is negative, a negative $\hat{\kappa}$ indicates exceedingly poor agreement. Threshold values used in the current paper for separating the different degrees of agreement for the Kappa statistic are listed in the following table:

Lower bound	Degree of agreement	Upper bound
< 0.05	No	0.05
0.05	Very poor	0.20
0.20	Poor	0.40
0.40	Fair	0.55
0.55	Good	0.70
0.70	Very good	0.85
0.85	Excellent	0.99
0.99	Perfect	1.00

Although the Kappa statistic appears well suited to judging agreement between maps, few applications could be found in the ecological literature. Congalton et al. (1983) is a notable exception (note that they term the statistic "KHAT").

COMPARISONS OF THE HOLDRIDGE LIFE ZONE MAPS

GFDL comparisons

A natural way to compare different vegetation maps is in terms of a change in area for each vegetation zone. Table 2 compares the change in area between the GFDL climate change projection (Wetherald and Manabe, 1986; Manabe and Wetherald, 1987) and the Holdridge Life Zone map for current climate. First, the total area of each vegetation zone is displayed for each map. Next, the size of the stable area is presented (i.e., the area categorized identically in both maps), along with the percentage of

TABLE 2

Change in area between the GFDL climate change map and the current climate vegetation map. The Kappa statistic for assessing agreement between maps is 0.43. This indicates only *fair agreement* between the current climate and climate change vegetation maps

Vegetation zone	Area comparison (units = 1000 km^2)				
	Current climate map	Climate change map	Stable area	% Stable area	Kappa statistic
Tundra	1036.89	429.29	429.29	41.4%	0.62
Forest Tundra	885.71	394.17	15.94	1.8%	−0.04
Cold Parklands	280.99	284.47	89.04	31.7%	0.32
Boreal Forest	1512.04	961.49	144.23	9.5%	−0.00
Temperate Forest	997.77	1185.85	407.69	40.9%	0.30
Steppe	741.13	1158.90	400.81	54.1%	0.35
Cool Desert	401.76	304.43	138.91	34.6%	0.37
Warm Temperate Forest	321.67	195.58	34.75	10.8%	0.11
Chapparal	562.86	740.61	40.50	7.2%	0.02
Hot Desert	2085.22	2065.00	1830.82	87.8%	0.86
Tropical Semi-Arid	953.36	1398.89	814.77	85.5%	0.66
Tropical Dry Forest	1485.48	1956.96	1216.70	81.9%	0.66
Tropical Seasonal Forest	1507.83	1002.29	689.08	45.7%	0.52
Tropical Rain Forest	845.77	1540.65	843.04	99.7%	0.68
Totals	13618.49	13618.58	7095.57	52.1%	0.43

the current climate vegetation map that remained stable. These area statistics allow one to determine how much a vegetation zone is shrinking, expanding, stable, or shifting (cf. Figs. 1B and 1C). The Kappa statistic is then calculated for each vegetation zone and for the entire map (Table 2). Note that the Kappa statistic is often close to the proportion of stable area, but only when a given zone has approximately the same area in both maps (e.g., Cold Parklands, Cool Desert, Hot Desert). This similarity breaks down if a zone is expanding or shrinking from one map to another. For example, over 99% of the original area in Tropical Rain Forest is stable in the GFDL projection. However, the amount of Tropical Rain Forest predicted by GFDL has doubled. Thus, the Kappa statistic for Tropical Rain Forest is 0.68 instead of being near 0.99.

The overall value of Kappa for the GFDL climate change comparison is 0.43. This indicates *fair agreement* with the current climate vegetation map. Furthermore, only 52% of the area has remained stable. Zones that expanded greatly are Steppe, Tropical Semi-Arid, Tropical Dry Forest, and Tropical Rain Forest. Zones undergoing considerable shrinkage are Tundra, Cold Parklands, Boreal Forest, and Tropical Seasonal Forest. Judging

TABLE 3

Kappa statistic and the corresponding qualitative degree of agreement between all possible pairs of Holdridge maps examined. The map labeled "Holdridge" is derived from current climate and all others are derived from the CO_2 doubling climate change scenario

	Kappa statistic and degree of agreement between maps				
	Holdridge	GFDL	GISS	OSU	UKMO
Holdridge	1.00	0.43	0.51	0.57	0.35
GFDL	Fair	1.00	0.67	0.65	0.71
GISS	Fair	Good	1.00	0.75	0.62
OSU	Good	Good	V. good	1.00	0.56
UKMO	Poor	V. good	Good	Good	1.00

by the Kappa statistic, the only vegetation zones that show at least good agreement with the current climate vegetation map are Tundra, Hot Desert, and Tropical zones of Semi-Arid, Dry Forest, and Rain Forest. Hot Desert, of course, is the most stable zone. The locations of the Forest Tundra, Boreal Forest, Warm Temperate Forest, and Chapparal zones have almost completely changed.

Additional GCM comparisons

Monserud (1990) used the preceding methods and the Holdridge Life Zone Classificaction to compare the maps produced by additional GCM projections of a doubling of CO_2:

GISS, Goddard Institute for Space Studies, NASA, Columbia U. (Hansen et al., 1983).

OSU, Oregon State University, Corvallis (Schlesinger and Zhao, 1989).

UKMO, United Kingdom Meteorological Office (Mitchell, 1983; Wilson and Mitchell, 1987).

The Kappa statistic indicates fairly large changes between the current climate vegetation map and the climate change scenarios for these four GCMs (Table 3). The largest disparities are for the UKMO projections (poor agreement: $\hat{\kappa} = 0.35$) and the GFDL and GISS projections (both fair agreement: $\hat{\kappa} = 0.43$ and 0.51). The OSU map displays the smallest change (agreement is good: $\hat{\kappa} = 0.57$) with respect to the corresponding current climate vegetation map.

Figure 2 displays the change in area by vegetation zones between the current and changed climate scenarios for each of the four GCMs. Areas to the left indicate a reduction in the original vegetation zones, while areas to the right indicate increases (either an expansion or a shift in location) as a

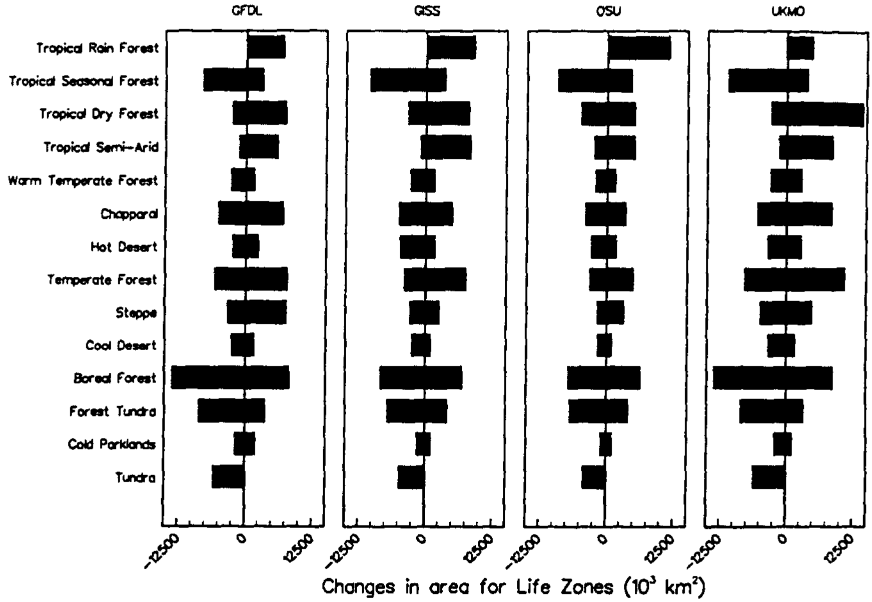

Fig. 2. Change in area by vegetation zones in the Holdridge climate change maps, for four GCM predictions. Area to the left of the zero centerline is a reduction (decrease) from the current climate vegetation map, while area to the right of the centerline is an increase in the climate change map. For example, the original area of the Tropical Seasonal Forest in the GFDL projection has shrunk from 1500 to 700×10^3 km^2 (a decrease of 800), while an additional area of 300×10^3 km^2 has been added in the GFDL climate change projection.

result of climate change. A glance at Fig. 2 indicates that the pattern of change across vegetation zones is quite similar for the four GCMs. For example, all four GCMs predict a fairly large increase in the area of the Tropical Rain Forests with climate change, without any reduction in the current land base. Conversely, all four GCMs predict a moderate decrease in the size of the Tundra with climate change, without any expansion into new locations. And all four GCMs predict large decreases in the current area of the Boreal Forest, decreases that are partly offset by a shift in the location of the Boreal Forest into new locations.

The Kappa statistic can also be used to compare any two GCM climate [illegible] (Table 3) reveals very good agreement between the OSU and GISS scenarios as well as between GFDL and UKMO. The remaining comparisons indicate good agreement. These statistics reinforce the sense of similarity apparent from examining Fig. 2.

DISCUSSION

The main objective of this research was to find an objective statistical procedure for comparing global vegetation maps. The Kappa statistic proved to be very useful for this purpose because it provided an objective measure of agreement between the whole maps and between respective vegetation zones within the comparison maps. When the Kappa statistics and the corresponding levels of agreement between these Holdridge Life Zone maps were first calculated, we found that the levels of agreement indicated by Kappa were about the same as had been judged previously by visual examination. In addition to being an objective measure of agreement, we feel that the Kappa statistic is actually measuring something about the maps that is important to humans.

Perhaps the most useful feature of the Kappa statistic (other than its objectivity) is in determining rank ordering, both across several maps and across the various categories within a given map. This rank ordering feature is quite helpful during the model development phase, for it quickly allows the model builder to find those vegetation zones that are predicted very poorly and to concentrate attention there. Furthermore, progress in modeling a particular vegetation zone can easily be monitored. Of course, no summary statistic will ever substitute for the complex visual information contained in the actual map.

As Prentice et al. (1992) point out, the Kappa statistic is not a perfect measure of similarity between maps. Pixel-by-pixel agreement may not necessarily be the best standard for comparison. A better measure might behave like the Kappa statistic but instead use a local measure of agreement. Prentice et al. (1992) have recently developed a continuous Kappa statistic that depends on the size of a block of adjacent pixels and on a standard measure of similarity (i.e., proportion of pixels that agree) to judge agreement between blocks. When block size is one, the measure reduces to the standard Kappa statistic. As block size increases, it approaches a measure of the similarity between the overall proportions of the categories on the two maps. Ideas for the development of a generalized similarity index can also be found in Costanza (1989) and Wigley and Santer (1990). If the ocean–land border question can be resolved, then the variable resolution fitting procedure of Costanza (1989) for categorical data and the univariate and multivariate statistics of Wigley and Santer (1990) for continuous data should provide good starting points for the development of a localized similarity index.

Given the restrictions and limitations of both the Holdridge Life Zone Classification and the climatic predictions from GCMs, this analysis of the effect of climate change on vegetation zones indicates potential for enormous ecological change. Large increases in the area of tropical zones (except Seasonal Forest) are indicated, along with correspondingly large decreases in the area of polar zones (e.g., Tundra, Forest Tundra, Boreal Forest). Equally important are indications that there will be wholesale shifts in the locations of some zones (Forest Tundra, Boreal Forest, Warm Temperate Forest, Chapparal). Great stability is indicated only for Hot Desert.

Of course the limitations of the Holdridge Life Zone Classification are great. Perhaps the largest is that it is simply a static equilibrium model, assuming unrealistically that plants respond immediately to a change in climate and that problems with species migration and establishment are nonexistent. A second limitation is that the Holdridge Life Zone Classification examined here is not seasonally sensitive. Clearly, this global model is not intended to compete with a reasonable dynamic vegetation model. Perhaps it is best to view such static global models as necessary first steps in the development of an interactive, dynamic atmosphere–biosphere system (Prentice, 1990). However, whatever truth is contained in the Holdridge hypothesis casts strong warning that the potential for large ecological change is great under the projected CO_2 doubling scenarios.

CONCLUSIONS

The Kappa statistic proved to be a useful and straightforward measure of agreement between global vegetation maps. Furthermore, individual Kappa statistics for comparing a given vegetation zone between two maps clearly indicated differences and similarities between maps. The most useful feature of the Kappa statistic is in determining rank ordering, both across several maps and across the various categories within a given map. Straightforward summary statistics comparing the change in area between maps for each vegetation zone provided additional useful information for objectively determining the actual differences between maps. Despite the availability of these statistics, examining the maps themselves remained the only way to comprehend the differences in their full geographical context and detail.

After applying the Holdridge Life Zone Classification to climate change scenarios (CO_2 doubling) produced by four GCMs, the pattern of change across vegetation zones was found to be quite similar for all projections. This analysis indicates great potential for enormous ecological change. Large increases in the area of tropical zones are indicated, along with correspondingly large decreases in the area of polar zones. Only the Hot Desert was stable. Furthermore, there are indications that there will be wholesale shifts in the locations of the Forest Tundra, Boreal Forest, Warm Temperate Forest, and Chapparal zones.

ACKNOWLEDGMENTS

We are grateful to Dr. Allen M. Solomon (IIASA Biosphere Dynamics Project Leader, 1987–1990) for providing both the invitation and the funding for conducting this research at IIASA, the International Institute of Applied Systems Analysis in Laxenburg, Austria. The senior author also thanks Dr. Albert R. Stage and the Intermountain Research Station for their great flexibility and continued support. Helpful review comments from Colin Prentice, Wolfgang Cramer, Gerd Esser and Greg Biging are much appreciated.

REFERENCES

Anonymous, 1990. IIASA's climate data base. International Institute for Applied Systems Analysis, Laxenburg, Austria. IIASA's Options, March 1990, pp. 9–12.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W., 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, MA, 557 pp.

Budyko, M.I., 1986. The Evolution of the Biosphere. D. Reidel Publishing Co., Dordrecht, 423 pp.

Cliff, A.D. and Ord, J.K., 1981. Spatial Processes: Models and Applications. Pion Limited, London, 266 pp.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20: 37–46.

Congalton, R.G., Oderwald, R.G. and Mead, R.A., 1983. Assessing Landstat classification accuracy using discrete multivariate analysis techniques. Photogramm. Eng. Remote Sensing, 49(12): 1671–1678.

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

Emanuel, W.R., Shugart, H.H. and Stevenson, M.P., 1985a. Climatic change and the broad-scale distribution of terrestrial ecosystems complexes. Climatic Change, 7: 29–43.

Emanuel, W.R., Shugart, H.H. and Stevenson, M.P., 1985b. Response to comment: climatic change and the broad-scale distribution of terrestrial ecosystems complexes. Climatic Change, 7: 457–460.

Fleiss, J.L., 1981. Statistical Methods for Rates and Proportions (2nd edition). John Wiley & Sons, New York, NY, 321 pp.

Gaile, G.L. and Willmott, C.J. (Editors), 1984. Spatial Statistics and Models. D. Reidel Publishing Company, Boston, MA, 482 pp.

Goodman, L.A. and Kruskal, W.H., 1954. Measures of association for cross classifications. J. Am. Stat. Assoc., 49: 732–764.

Guetter, P.J. and Kutzbach, J.E., 1990. A modified Köppen classification applied to model simulations of glacial and interglacial climates. Climatic Change, 16: 193–215.

Hansen, J., Russell, G., Rind, D., Stone, P., Lacis, A., Lebedeff, S., Reudy, R. and Travis, L., 1983. Efficient three-dimensional global models for climate studies: models I and II. Mon. Weather Rev., III(4): 609–662.

Harrison, S.P., 1990. An introduction to general circulation modelling experiments with raised CO_2. Working Paper WP-90-27. International Institute for Applied Systems Analysis, Laxenburg, Austria, 52 pp.

Henderson-Sellers, A., 1991. Developing an interactive biosphere for global climate models. Vegetatio, 91: 149–166.

Holdridge, L.R., 1947. Determination of world plant formations from simple climatic data. Science, 105: 367–368.

Holdridge, L.R., 1967. Life Zone Ecology. Tropical Science Center, San José, Costa Rica, 206 pp.

Houghton, J.T., Jenkins, G.J. and Ephraums, J.J. (Editors), 1990. Climate Change: The IPCC Scientific Assessment. Cambridge University Press, Cambridge, UK, 365 pp.

Landis, J.R. and Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics, 33: 159–174.

Leemans, R., 1989. Possible changes in natural vegetation patterns due to a global warming. In: A. Hackl (Editor), Der Treibhauseffekt: das Problem — Mögliche Folgen — Erforderliche Massnahman. Akademie für Umwelt und Energie, Laxenburg, Austria, pp. 105–122.

Leemans, R. and Cramer, W.P., 1991. The IIASA database for mean monthly values of temperature, precipitation and cloudiness on a global terrestrial grid. Research Report RR-91-18. International Institute for Applied Systems Analysis, Laxenburg, Austria, 61 pp.

Manabe, S. and Wetherald, R.T., 1987. Large-scale changes in soil wetness induced by an increase in carbon dioxide. J. Atmos. Sci., 44: 1211–1235.

Mitchell, J.F.B., 1983. The seasonal response of a general circulation model to changes in CO_2 and sea temperature. Q. J. R. Meteorol. Soc., 109: 113–152.

Monserud, R.A., 1990. Methods for comparing global vegetation maps. Working Paper WP-90-40. International Institute for Applied Systems Analysis, Laxenburg, Austria, 31 pp.

Olson, J. and Watts, J.A., 1982. Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation (Map). NDP-017, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Olson, J., Watts, J.A. and Allison, L.J., 1983. Carbon in Live Vegetation of Major World Ecosystems. ORNL-5862, Oak Ridge National Laboratory, Oak Ridge, Tennessee, pp. 164.

Pielou, E.C., 1977. Mathematical Ecology (2nd edition), John Wiley & Sons, New York, NY, 286 pp.

Prentice, I.C., Webb, R.S., Ter-Mikhaelian, M.T., Solomon, A.M., Smith, T.M., Pitovranov, S.E., Nikolov, N.T., Minin, A.A., Leemans, R., Lavorel, S., Korzukhin, M.D., Helmisaari, H.O., Hrabovszky, J.P., Harrison, S.P., Emanuel, W.R. and Bonan, G.B., 1989. Developing a global vegetation dynamics model: results of an IIASA summer workshop. Research Report RR-89-7, International Institute for Applied System Analysis, Laxenburg, [illegible]

Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. and Solomon, A.M., 1992. A global biome model based on plant physiology and dominance, soil properties and climate. J. Biogeogr. (in press).

Prentice, K.C., 1990. Bioclimatic distribution of vegetation for general circulation model studies. J. Geophys. Res., 95 (D8): 11811–11830.

Prentice, K.C. and Fung, I.Y., 1990. The sensitivity of terrestrial carbon storage to climate change. Nature, 346: 48–51.

Ripley, B.D., 1981. Spatial Statistics. John Wiley & Sons, New York, NY, 252 pp.
Scheffé, H., 1959. The Analysis of Variance. John Wiley & Sons, New York, NY. 477 pp.
Schlesinger, M.E. and Zhao, Z-C., 1989. Seasonal climatic changes induced by doubled CO_2 as simulated by the OSU atmospheric GCM/mixed-layer ocean model. J. Climate, 2: 459–495.
Shugart, H.H., 1990. Using ecosystem models to assess potential consequences of global climatic change. Trends Ecol. Evol., 5: 303–307.
Solomon, A.M. and Leemans, R., 1990. Climatic change and landscape-ecological response: Issues and analyses. In: M.M. Boer and R.S. de Groot (Editors), Landscape Ecological Impact of Climatic Change. IOS Press, Amsterdam, pp. 293–317.
Strahler, A.N. and Strahler, A.H., 1987. Modern Physical Geography. John Wiley & Sons, New York, NY, 544 pp.
Turner, M.G., Costanza, R. and Sklar, F.H., 1989. Methods to evaluate the performance of spatial simulation models. Ecol. Modelling, 48: 1–18.
Walker, B.H. and Greatz, R.D., 1989. Effects of Atmospheric and Climate Change on Terrestrial Ecosystems. No. 5, International Geosphere-Biosphere Programme, Stockholm, pp. 61.
Wetherald, R.T. and Manabe, S., 1986. An investigation of a cloud cover change in response to thermal forcing. Climate Change, 8: 5–23.
Wigley, T.M.L. and Santer, B.D., 1990. Statistical comparison of spatial fields in model validation, perturbation, and predictability experiments. J. Geophys. Res., 95: 851–865.
Wilson, C.A. and Mitchell, J.F.B., 1987. A doubling CO_2 climate sensitivity experiment with a global climate model including a simple ocean. J. Geophys. Res., 92: 13315–13343.