Landscape Ecology (2006) 21:213–224
DOI 10.1007/s10980-005-0827-0

© Springer 2006

Research article

A method to efficiently apply a biogeochemical model to a landscape

Robert E. Kennedy[1,3,*], David P. Turner[1], Warren B. Cohen[2] and Michael Guzy[1,4]
*[1]Department of Forest Science, Oregon State University, 321 Richardson Hall, Corvallis, OR 97331, USA; [2]USDA Forest Service, PNW Research Station, 3200 SW Jefferson Way, Corvallis, OR, 97331, USA; [3]Present Address: USDA Forest Service, PNW Research Station, 3200 SW Jefferson Way, Corvallis, OR, 97331, USA; [4]Present Address: Department of Bioresource Engineering, Oregon State University, 116 Gilmore Hall, Corvallis, OR 97331, USA; *Author for correspondence (e-mail: robert.kennedy@fs.fed.us)*

Received 19 December 2004; accepted in revised form 8 July 2005

Key words: Biome-BGC, Carbon modeling, Interpolation, Mapping, Net ecosystem production, Net primary production, Oregon

Abstract

Biogeochemical models offer an important means of understanding carbon dynamics, but the computational complexity of many models means that modeling all grid cells on a large landscape is computationally burdensome. Because most biogeochemical models ignore adjacency effects between cells, however, a more efficient approach is possible. Recognizing that spatial variation in model outputs is solely a function of spatial variation in input driver variables such as climate, we developed a method to sample the model outputs in input variable space rather than geographic space, and to then use simple interpolation in input variable space to estimate values for the remainder of the landscape. We tested the method in a 100 km × 260 km area of western Oregon, U.S.A. , comparing interpolated maps of net primary production (NPP) and net ecosystem production (NEP) with maps from an exhaustive, wall-to-wall run of the model. The interpolation method can match spatial patterns of model behavior well (correlations > 0.8) using samples of only 5 t o 15% of the landscape. Compression of temporal variation in input drivers is a key step in the process, with choice of input variables for compression largely determining the upper bounds on the degree of match between interpolated and original maps. The method is applicable to any model that does not consider adjacency effects, and could free up computational expense for a variety of other computational burdens, including spatial sensitivity analyses, alternative scenario testing, or finer grain-size mapping.

Introduction

A key challenge to carbon cycle science is characterizing current and future carbon dynamics over large spatial domains. Biogeochemical models offer one means of carbon modeling, and have been applied for this purpose at regional to continental scales (Aber and Federer 1992; VEMAP Members 1995; Thornton 1998).

In a standard approach to spatial modeling, the biogeochemical model is run separately at a large number of contiguous grid cells on the landscape (e.g. VEMAP Members 1995; Ollinger et al. 1998; Law et al. 2004). Each cell contains the necessary driving variables (climate, soils, etc.) to run the model. For many biogeochemical models, computational burden at each cell can be great (Thornton 1998; Law et al. 2004), reducing the

number of model runs possible over large areas. To explore fine grain effects over large areas, to characterize model sensitivity, or to investigate alternative climate change scenarios, more efficient approaches to spatial extrapolation are desirable.

One approach is to reduce the computational burden of the model by constructing a simplified metamodel (Friedman 1996; Urban et al. 1999) that reproduces the bulk properties of a detailed model over a defined range of input (driver) variables (e.g. Williams et al. 1997, 2001; Urban et al. 1999; Acevedo et al. 2001; Alexandrov et al. 2002; Garman 2004). The structure of the metamodel is developed through hypothesis building and expert understanding of the detailed model (Williams et al. 1997; Garman 2004), and then parameterized by running the detailed model across the full range of input variables. Because of its lower computational burden, this simplified model can be applied to larger landscapes more readily than its more detailed counterpart. Although an elegant and powerful approach, development of the structure for the metamodel can be challenging, especially when multiple feedbacks in the detailed model prevent easy hypothesis building.

A simpler strategy for spatial extrapolation of a computationally burdensome model is to use lookup tables. Here, the *n*-dimensional input variable space observed on a given landscape is partitioned into discrete regions, the model is run only once in each input variable region, and the model outputs are assigned to the landscape according to the region of input space into which each point on the landscape falls (Band et al. 1991; Burke et al. 1991; Franklin 2001). Although conceptually appealing, this strategy relies on an *a priori* stratification whose definition may be arbitrary and whose discrete outputs may not be appropriate for some applications.

Nevertheless, the lookup-table approach emphasizes that most biogeochemical model outputs are entirely a function of the input variables controlling them, not a function of neighboring conditions (e.g. Aber and Federer 1992; Running and Hunt 1993; VEMAP Members 1995; Ollinger et al. 1998; Thornton 1998; Coops and Waring 2001). For any combination of input variables, the model output will be the same regardless where on the landscape it falls. Thus, the models are not truly spatial (Peters et al. 2004). Rather, spatial variation in the model outputs simply reflects the underlying geographic variability in input variables.

In the lookup-table approach, those input variables are stratified into discrete regions, but we envision an extension to the continuous-variable case (Figure 1). Rather than the standard approach of running the model in every cell on the landscape (Figure 1, black arrow), model outputs can be determined at a sample of points in the multidimensional space of the input variables, interpolated to the remaining cells in the input variable space, and mapped back into geographic space (Figure 1, open arrows). We refer to the standard path as "wall-to-wall" modeling and to the alternative path as "input-space interpolation" modeling.

For biogeochemical models, a challenge is appropriately representing input-variable space. Climate variables that drive daily time-step models, for example, may include hundreds or thousands of dimensions of information (Thornton 1998). Fortunately, it is only the spatial variation in the temporal variables that matters. A consideration of the simplest case is instructive: If all of the cells on a landscape had identical temporal patterns of climate, then a map of model outputs for the entire landscape could be constructed by replicating the output from single model run across every cell. This is true even if that single run were extremely complicated to implement. While most landscapes show more spatial variation in the input drivers than this simplest case, spatial autocorrelation in climate and soil properties often results in spatial variation orders of magnitude lower than the temporal variation. Our method capitalizes on this fact to develop a compressed representation of spatial variation in input variables, and then applies the input-space interpolation approach to mimick spatial patterns in modeled output.

Methods

Wall-to-wall modeling

We tested our conceptual approach using the biogeochemical model BGC 4.1.1. (Thornton 1998; Thornton et al. 2002). BGC is constructed as a series of modules representing the major functions of an ecosystem, including photosynthesis,

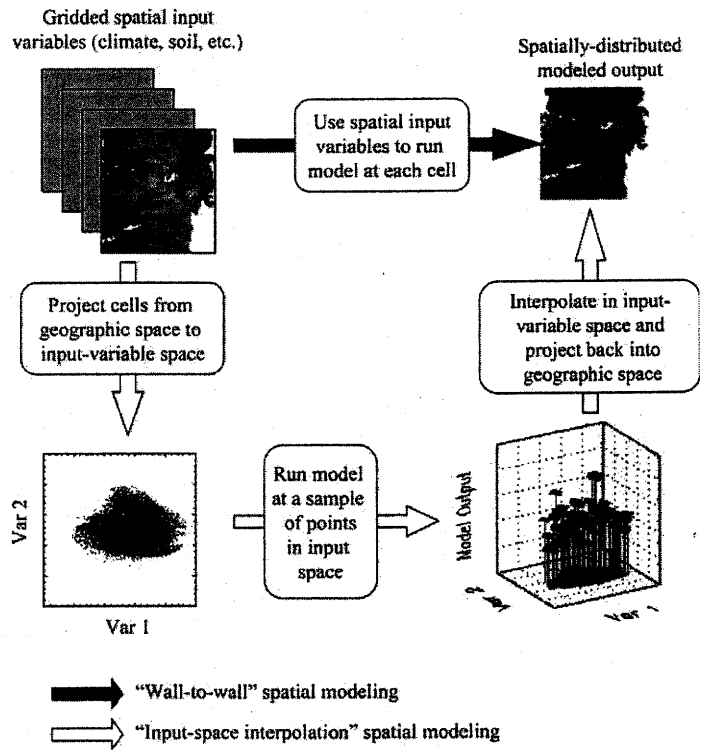

Figure 1. Two approaches for spatial modeling with ecosystem models that do not have cell to cell interactions. In "wall-to-wall" modeling (dark arrow), the model is run separately for every cell in the geographic study area. This approach is relatively inefficient, since many cells may have similar driving variables (climate, soils, etc.). The proposed alternative approach moves modeling out of geographic space into the space of the driver variables (the "input space"). Modeling is done only at a sample of grid cells, with the rest of the points being estimated by interpolation in input-space and re-projection into geographic space.

respiration, decomposition, and transpiration in plants, as well as transpiration and runoff of water. The model requires daily meteorological data, here derived from DAYMET (Thornton et al. 1997; Thornton et al. 2000) at a grain size of 1 km, which defines the grain size for BGC modeling. Climate data at each grid cell are derived from actual weather station observations, interpolated using a terrain model, and thus spatial patterns of climate data reflect both simple elevation effects as well as the long-term patterns of weather systems and air masses. At the time of this work, each 1 km cell had 18 years of daily meteorological data (365 days × 18 years for a total of 6570 input layers per meteorological variable). Soil depth and type were from a dataset described in Kern et al. (1997). The model is designed to run for an idealized vegetation type for several thousand years until several carbon and nitrogen pools reach approximate steady state. A vegetation type is defined by a suite of parameters that describe the type's ecosystem-, canopy-, organism-, and leaf-level attributes. In this study, the parameter set for evergreen needleleaf conditions supplied with the model code was adjusted slightly to better represent Douglas-fir (*Pseudotsuga mensziesii*). The model was run for a 100 km × 260 km study area in western Oregon (U.S.A.). Modeling was limited to 18,395 cells where a landcover map of the area (Law et al. 2004) suggested Douglas-fir was likely to occur (Figure 2). After spinup equilibrium was reached in each cell, we followed the approach of Law et al. (2004) in applying two post-spinup disturbances, and finally the model was allowed to run for 200 years.

The two variables tracked for this study were net primary production (NPP) and net ecosystem production (NEP), both recorded in units of

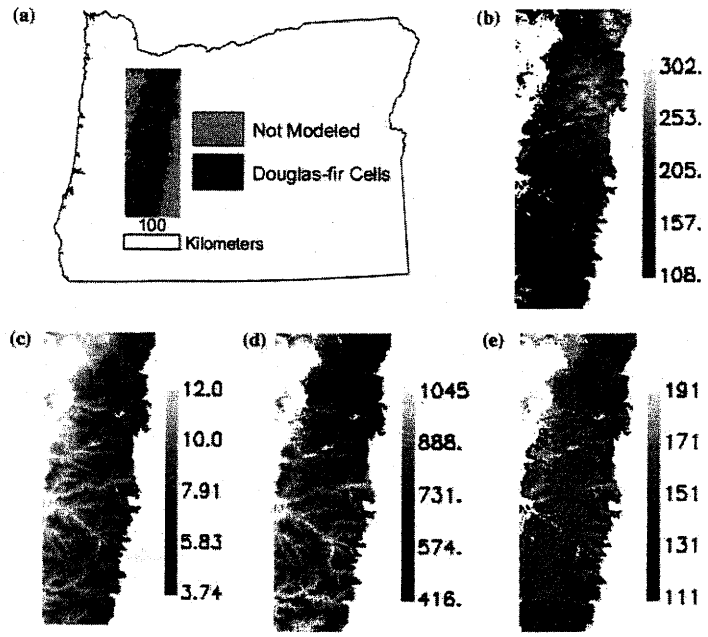

Figure 2. The study area, (a) defined as all of the cells within a 100 km × 260 km area of western Oregon's Cascade Mountains that are likely to support Douglas-fir forests. Shown are (b) average yearly precipitation (cm), (c) temperature (degrees C), and (d) vapor pressure deficit (Pa), as well as (e) estimated soil depth (cm). See Table 1 for sources of spatial data.

kg C/m^2 yr. NPP is defined as total autotrophic carbon fixation, minus respiration losses due to maintenance and growth within the plant. NEP is the total system balance of carbon, here defined as NPP minus heterotrophic respiration and losses to fire. NEP and NPP vary significantly over the 200-year model run. Directly after disturbance, NEP is initially negative (carbon flowing out of the terrestrial system) because of decomposition of residual biomass, and then it slowly becomes positive as NPP, which has quickly returned to positive flux (into the terrestrial system), accumulates and as material from the prior stand available for decomposition diminishes (see Law et al. (2004), Figure 8 for more detail of typical model output in this ecosystem). Because of the cyclic repetition of the DAYMET data, an appropriate temporal grain for analysis of these temporal trends is 18 years. Therefore, all NPP and NEP data were aggregated into non-overlapping 18-year bins, referred to as age classes, using simple averaging.

This wall-to-wall modeling represents the "truth" set. While the model itself does not truly describe the system, the wall-to-wall modeling is a faithful representation of the behavior of the model in every 1 km cell across the study area. The goal of the input-space interpolation modeling approach is to efficiently match the behavior of the model over space.

Input-space interpolation modeling

Because the model ignores adjacency effects among cells, spatial patterns of modeled NPP and NEP are determined solely by spatial patterns in meteorological and soils data (Figure 2). As noted above, the high dimensionality of the climate data in a given cell is irrelevant for spatial mapping. The first step in input-space interpolation, then, is to develop a simplified representation of spatial variation in input variables. This was achieved through several compression steps. First, the yearly mean or total (as appropriate for each variable) was calculated for each meteorological variable listed in Table 1 for each of the 18 years of record. For precipitation only, spring and summer averages were considered separately to

Table 1. Summaries of spatially-distributed input variables for the study area shown in Figure 2.

Variable	Description	Units	Minimum[a]	Maximum[a]	Median[a]	Source
Pr	Average total precipitation (Jan. 1 to Dec. 31)	cm	108.9	302.1	175.5	DAYMET[b]
PrSp	Average spring precipitation (April 1 to June 30)	cm	19.2	60.1	34.8	DAYMET[b]
PrSu	Average summer precipitation (July 1 to Sep. 30)	cm	6.4	22.2	11.6	DAYMET[b]
T	Average daily temperature	°C	3.7	12.1	8.8	DAYMET[b]
Tmn	Average daily minimum temperature	°C	−1.8	6.6	3	DAYMET[b]
Sw	Average daily shortwave radiation flux	W/m^2	244.8	330.7	289	DAYMET[b]
VPD	Average daily vapor-pressure deficit	Pa	417	1045.6	730.1	DAYMET[b]
SDpth	Soil depth	cm	111	192	163	Kern et al. 1997

[a]Minimum, maximum, and median values are for the population of grid cells in the study area; see Figure 2.
[b]DAYMET citations: (Thornton et al. 1997; Thornton et al. 2000).

reflect the dominant seasonal climate of the region (Waring and Franklin 1979). This temporal aggregation was carried out for each cell on the landscape separately.

Next, variation across cells was characterized using principal-component analysis (PCA). PCA is a tool that compresses an *n*-dimensional space into a reduced set of transformed, orthogonal axes that more efficiently capture the bulk of the variation in the original space (Jongman et al. 1995). The more highly correlated the dimensions of the original space, the greater the compression. Before PCA was applied, all meteorological data were confirmed to be normally distributed, and then were standardized to across-year mean and unit standard deviation. This preserved interannual variation within a meteorological variable while providing equal weighting across meteorological variables with different units of measurement. In PCA, each cell was assigned principal component scores that described the cell's position along new axes through the multivariate meteorological data space. Because a first principal component axis captures the greatest variance in a dataset, the map of first principal component axis scores (PC Image #1) represented the dominant spatial pattern in the 18-year record of each meteorological variable, with successively higher-order PC images representing diminishing sources of spatial variation. Soil depth was standardized to unit standard deviation and used as a single-layer spatial input variable. (Note: Soil texture was not considered here because its spatial patterns were identical to those of soil depth, but with much less variability in range). The PC images and single soil depth image will be referred to collectively as "spatial input variables."

Taken together, these spatial input variables define the input variable space of the model, and define the bounds of the response surface of the model as applied to the particular landscape of study. However, spatial variation in the different meteorological indices is highly correlated across indices. If all meteorological indices were perfectly correlated, for example, then any one of them could be used alone to define the meaningful spatial variation of model outputs. Parsimony argues for use of a combination of spatial input variables that most efficiently captures spatial pattern in model output. Twelve combinations of spatial input variable were tested, with combinations chosen to be illustrative of expected meteorological and edaphic controls on the model (Table 2). For each combination, two cases were considered: one using only the first PC image (the dominant spatial pattern) of each spatial input variable, and one using both the first and second PC images of each

Table 2. Summary of spatial input variable combinations tested for input-space interpolation.

Input Variables[a]
Pr + Sw + T
Pr + T + Tmn
Pr + T + VPD
PrSu + PrSp + Sw + T
PrSu + PrSp + T + Tmn
PrSu + PrSp + T + VPD
SDpth + Pr + T
SDpth + PrSu + PrSp + T
SDpth + PrSu + PrSp + Sw + T
SDpth + PrSu + PrSp + T + Tmn
SDpth + PrSu + PrSp + T + VPD
SDpth + Pr + T + Tmn + VPD

[a]Variables defined in Table 1

spatial input variable (except for soil depth, which has only one layer). This resulted in a total of 24 combinations in spatial input variables, the simplest having 3 dimensions (the first PC image from three spatial input variables) and the most complex having 9 dimensions (two PC images from four spatial input variables, plus standardized soil depth). A final PCA was then applied to each of these combinations, and the first three axes of this final PCA used to describe the spatial variation in the spatial input variables.

Although a variety of space-filling approaches to response-surface sampling are available (Box and Draper 1987; Myers and Montgomery 2002), a simple regular grid was chosen here. A lattice was constructed through the three-dimensional space in each of the 24 cases, with equal spacing between lattice intersections. The actual grid cells closest (using Euclidean distance) to lattice intersections were used as sample cells; lattice intersections greater than one cubic hypotenuse length from an actual cell were dropped. At each sampled cell, the model value from the wall-to-wall approach was extracted. (In an actual implementation of the approach, no such wall-to-wall output would be available, and the model would be run in each sample cell.)

Non-sampled cells were estimated by interpolation in input variable space. For each cell, the NEP or NPP value was estimated by linear inverse-distance weighted averaging of the nearest 8 sampled lattice-intersection cells (in three-dimensional input variable space). Once NEP or NPP values were assigned in input-variable space, the cells were mapped back into geographic space to produce a map of estimated model outputs. If the lattice of sample points was sparse, interpolation distance would be large and the smoothing properties of interpolation exacerbated. To investigate this effect, tests were repeated at five grid densities, ranging from $<1\%$ to 15% of the landscape for NEP outputs only.

Comparisons

Modeled outputs from the input-space interpolation approach were compared on a cell-by-cell basis with all cells in the truth datasets derived using the wall-to-wall approach. Two metrics of comparison were calculated across all cells for each of the 11 age classes and each of the five grid densities: a simple correlation and the square root of mean-square error (RMSE). These two metrics were also calculated for a randomly-selected subset of validation cells, and compared to the metrics derived from the wall-to-wall values to provide an estimate of the proportion of independent cells needed to faithfully capture error in the interpolation method. Tested validation cell proportions ranged from 0.5% to 10% of the landscape.

Multiple regression in input variable space was investigated as a possible alternative to interpolation. Sampling of the input variable space was identical to that used for interpolation. A multiple linear regression of NEP or NPP values from all sampled points on the three axes of the input space was conducted. The resultant model was then applied to the non-sampled points to produce estimates of NEP or NPP in the input-variable space. As with interpolation, those estimated values were then projected into geographic space to produce a map of modeled outputs.

Results

Figure 3 shows a comparison of interpolated and wall-to-wall NEP maps for one age class at two lattice-point sampling densities using one combination of spatial input variables (SDpth + Pr + T + Tmn + VPD). It provides a visual reference for the results presented subsequently.

As expected, the interpolation method works better when more points are sampled for the interpolation. When only 0.4% of the cells are sampled in the input-variable space, the interpolation approach produces a map that captures only the approximate patterns of the wall-to-wall map (Figure 3a). The image to image histogram plot shows scatter about the 1:1 line (Figure 3b). When the input-space lattice samples 12% of the cells, the maps match well visually (Figure 3c), and the image-to-image histogram plot shows strong attraction to the 1:1 line with relatively little scatter.

The two summary metrics capture these visual patterns. The correlation between the maps provides a quantitative measure of the fit, and is independent from the units of NEP or NPP being estimated. As expected, the poorly-matched map

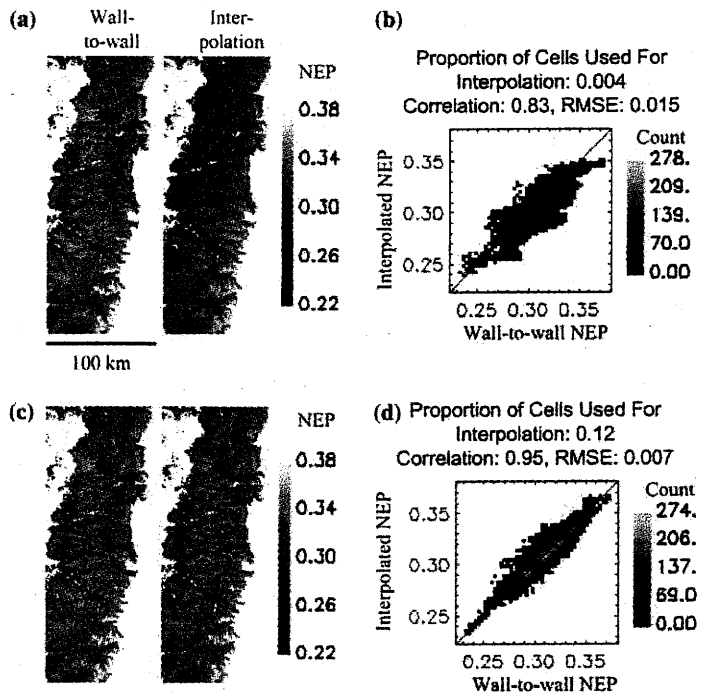

Figure 3. Comparing wall-to-wall NEP maps (left-hand maps) with input-space interpolation NEP maps (right-hand maps) for interpolation age class 73–90 years, for sampling proportions of (a) 0.004 and (c) 0.12. NEP units are kg C/m^{2}year, with positive values representing uptake by the terrestrial system. Spatial input variables for were soil depth, yearly precipitation, average temperature, average minimum temperature, and average vapor pressure deficit. Plots (b) and (d) show two-dimensional histograms of the interpolated map against the wall-to-wall map, and report the correlation and RMSE of the plots. Higher sampling intensity for interpolation resulted in a better match.

had a lower correlation than a well-fit map (compare Figure 3b and d). The RMSE provides a measure of the potential error in units of NEP or NPP, useful for interpreting the magnitude of the error directly.

In the example from Figure 3, the match with the original wall-to-wall map was poorer when regression rather than interpolation was used (Figure 4a and b). This held true across nearly all NEP tests conducted for this study (Figure 4c). From here forward, results will be limited to maps built using interpolation.

The correlation of maps reached an asymptotic value as the proportion of training cells increased. The level of this asymptote varied depending on the spatial input-variables used to build the input-variable space, and the age class being interpolated (Figure 5). For any given age class, the spatial input-variable combination with highest maximum correlation tended to reach its asymptote more quickly than those combinations with lower maximum correlation.

Maximum correlation varied across age classes for a given combination of spatial input variables, and different combinations showed different temporal trajectories (Figure 6). Three distinct groups of spatial input-variable combinations emerged (separated as Figure 6a, 6b, and 6c). Groups 1 and 2 (Figure 6a and b) performed well or moderately well across age classes, but had opposing patterns with age. Group 3 performed poorly for older age classes (Figure 6c). Notable is the inclusion of soil depth as a spatial input variable in Groups 1 and 2, and its omission from Group 3.

NPP results were similar to NEP results. For brevity, only one example test from each of the Groups identified in Figure 6 is shown in Figure 7. The maximum correlations for NPP in Group 1 were slightly lower than for the NEP maps, but were more consistent across age classes. Groups 2

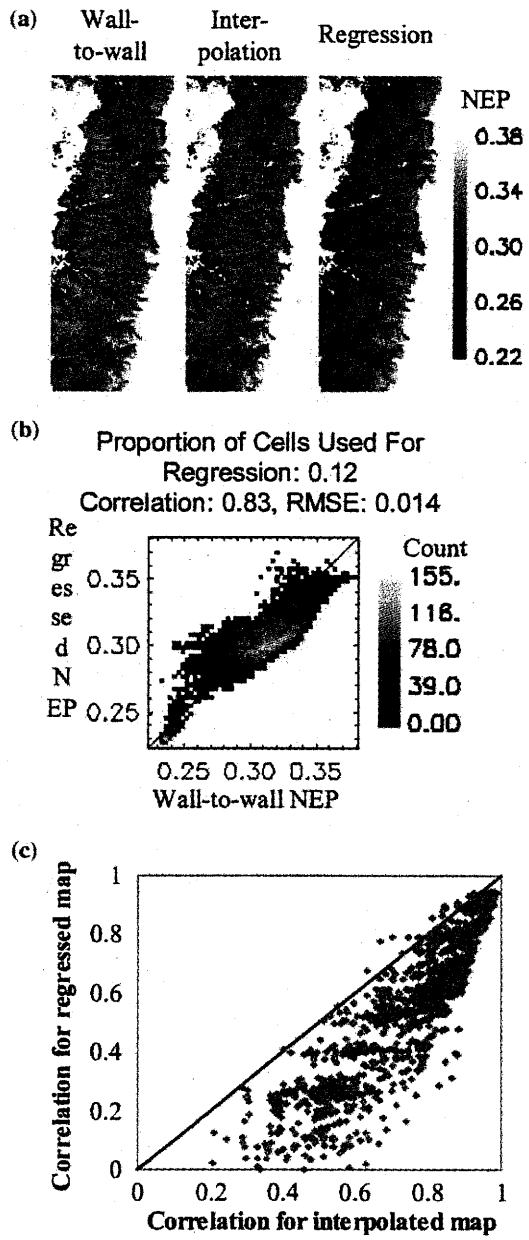

Figure 4. Comparing maps derived using interpolation versus regression. (a) Wall-to-wall, interpolated, and regressed NEP maps for ages 73–90 years using a sampling proportion of 0.12. Input variables used to define the input space are the same as in Figure 3. (b) Two-dimensional histogram for the regressed map against the wall-to-wall map. Compare to the interpolated version in Figure 3d. (c) Plots of the correlations for regressed maps against correlations for interpolated maps for 1144 NEP maps spanning a range of input variable combinations and sampling proportions.

and 3 were also more stable for NPP than for NEP maps, and generally had higher correlations than for NEP.

A summary of the use of validation points is shown in Figure 8. Across all NEP tests, correlations of interpolated and wall-to-wall maps were well-characterized using a sample of validation points. Even at relatively low sampling density (3% of cells), the correlation as estimated by validation points strongly tracked the true correlation (Figure 8a). As the proportion of cells used for validation points increased, the match improved slightly (Figure 8b). A regression line anchored at zero had a slope near 1 (ideal) for all validation proportions greater than 3%.

Discussion

When a model lacks cell-to-cell interactions, its behavior across geographic space depends entirely on the spatial patterns of the variables that drive it. Insofar as the driver variables are redundant across space, wall-to-wall modeling of each cell is inefficient. Here, we described and tested the input-space interpolation approach to more efficiently extrapolate spatial patterns in the biogeochemical model BGC (Thornton 1998).

In essence, the input-space interpolation approach develops a site- and input-variable-specific metamodel of the spatial behavior of BGC. This metamodel is not a simplified representation of the functional behavior of the model at a given point in space, and as such lacks the conceptual elegance of other metamodeling efforts (e.g. Williams et al. 1997; Urban et al. 1999, Garman 2004). Rather, it is a simplified representation of the *spatial behavior* of the model. The advantage of this approach is that the entire functional complexity of the underlying model is retained, and only the relatively simpler spatial variation is approximated.

The input-space interpolation method appears to work well. By sampling only 10 to 15% of the cells in our study area, the approach captured the bulk of the patterns of modeled NEP and NPP values across space (Figure 3) and time (Figures 6a, 6b, and 7). More importantly, the error of prediction of model output was generally quite small. Figure 9 shows mean RMSE by age class of NEP predictions for all of the runs in Groups 1

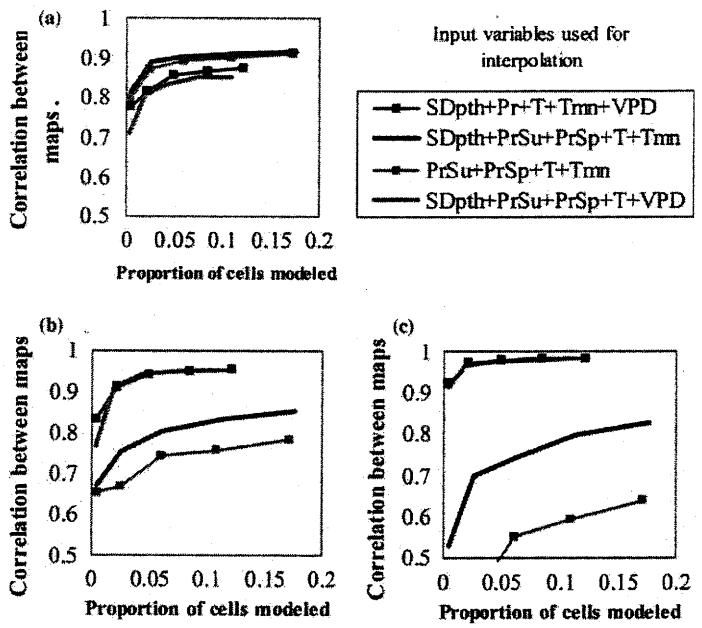

Figure 5. Examples of the correlation between wall-to-wall and interpolated maps of NEP for modeled stands at (a) 19–36 years, (b) 73–90 years, and (c) 163–180 years, for four different combinations of input-space drivers for interpolation. Definitions of input-space drivers are given in Table 1. The relative performance of different input variable combinations varied across age classes.

and 2 (of Figure 6). Except for the youngest age class, mean error was just over 3%. Error in estimates was smaller using interpolation than regression, underscoring that local fitting in input variable space was more appropriate than a global fit. Note, however, that other forms of regression may work better than the linear approach used here, especially those that use local fits, non-linear structures, or arbitrarily complex variable combinations. While reducing error relative to the regression approach tested here, such complexity would appear unwarranted given the success of the simple interpolation approach.

The compression of the input-variable space appears to set the upper bounds on the accuracy of the interpolated maps. Asymptotic levels of correlation were reached at fairly low sampling intensities (Figure 5), suggesting that further increases in sampling density or better input variable space sampling strategies would do little to improve correlation. Rather, correlation was only improved when a better set of spatial input variables was used to describe the input variable space (Figures 6 and 7). It is critical to emphasize that correlation and accuracy of the maps refer only to how well the interpolation approach mimicks spatial patterns of modeled NEP and NPP values, not how well those maps represents *actual* NEP or NPP values in the natural world.

The chief drawback of this approach is its sensitivity to the choice of spatial input-variables. Had only input-variable combinations from Group 3 been tested, the results would have been poor. Thus, a certain degree of testing of different variables is necessary, using an independent set of validation points for establishing error rates. An understanding of the mechanistic controls within the model should lead to judicious choice of input variables for interpolation.

Interpreting relative importance of different spatial input variables sheds light on model behavior. The success of interpolation varied over the time course of model runs, suggesting that controls on spatial patterns of model outputs varied over time. Soil depth was a consistently important control on modeled carbon dynamics (both NEP and NPP), especially at older age classes. All of the input variables in Groups 1 and 2 (the better-fitting Groups) included soil depth as an input variable, and the only difference between

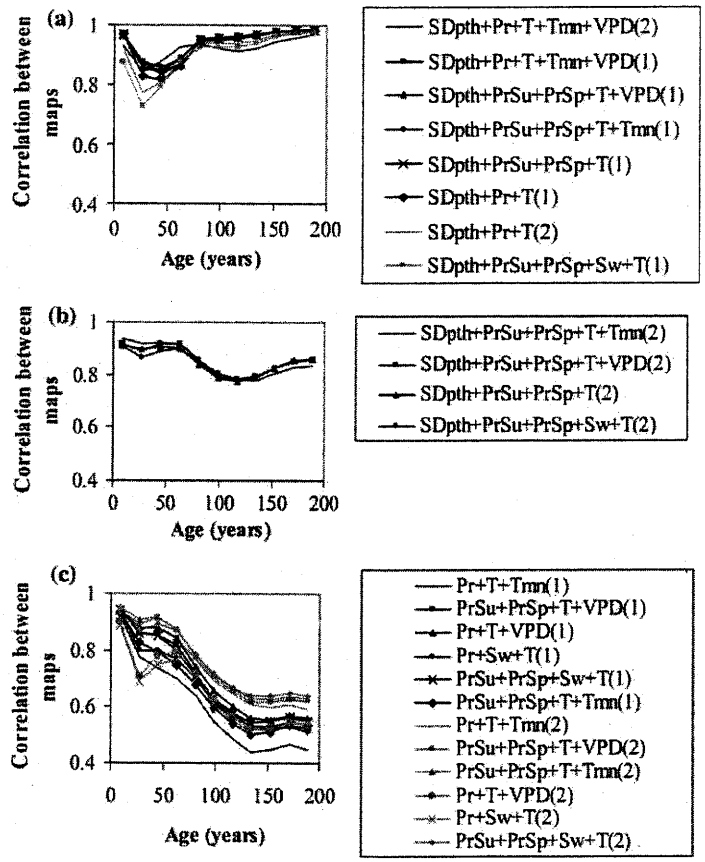

Figure 6. Maximum correlation between interpolated maps and wall-to-wall maps of NEP plotted against age class. Definitions of variables are given in Table 1. Numbers in parentheses are the number of PC layers from each meteorological variable used to build the input-variable space. (a) Group 1: Input variables that result in better fits at older ages than younger ages. (b) Group 2: Input variables that result in better fits at young ages and moderate fits at old ages. (c) Group 3: Input variables that fit well at young ages but very poorly at older ages.

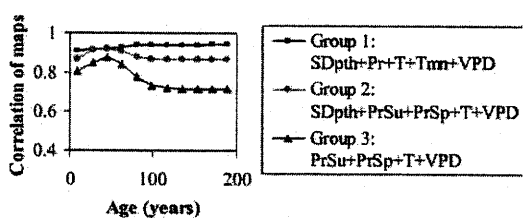

Figure 7. As in Figure 6, but for NPP rather than NEP maps, and for only one representative combination of input variables from each of the three groups shown in Figure 6.

some Group 1 and Group 2 combinations was the relative weight given to soil depth in the data compression steps (data not shown). Although not necessarily the case in natural systems, soil depth in BGC controls both the total water content that can be stored in the system and the levels of available nutrients that accumulate in the soil. Apparently, these controls have high leverage on overall model behavior, especially over long periods. As a site constant, soil depth was not considered in an extensive sensitivity analysis of BGC (White et al. 2000), which makes judging its relative importance difficult. However, it is clear that a better understanding of the effects of soil depth on the behavior of the BGC model will be useful when the model is applied over regional or continental scales.

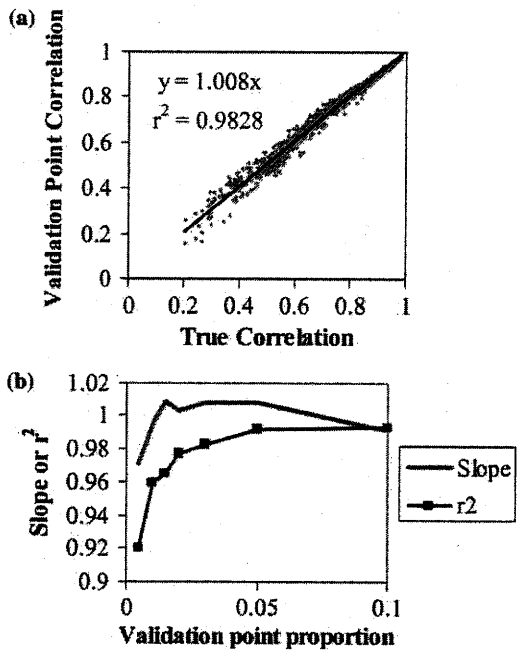

Figure 8. Summaries of tests of validation points for estimating true map correlation. (a) Plot of validation point vs. true correlation for validation point proportion of 0.03. Each point represents one map comparison. (b) Slope and r^2 of linear regressions of the type shown in (a), for a range of validation point proportions. Reasonable estimates of correlation are achieved with only 3–5% of plots used as validation points.

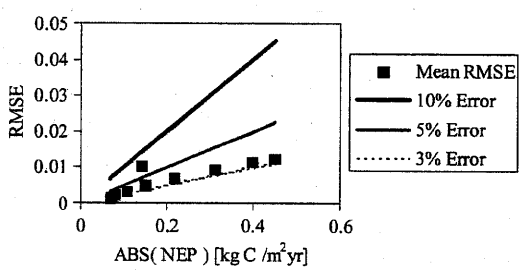

Figure 9. Mean RMSE of interpolated NEP as a function of absolute value of true NEP, for all runs in Groups 1 and 2 of Figure 6 averaged by age class, with lines indicating gradations of percent error. For most runs, error of interpolation was between 3 and 5% of the true value. The single exception was for age class 1 to 18 years, where RMSE of NEP was more than 5% of the true NEP.

Although it may aid in understanding model behavior, the input-space interpolation approach is primarily designed to improve modeling efficiency. A reasonable approximation of spatial behavior of the BGC model can be achieved using only a fraction of the points on the landscape. Indeed, although the highest-density runs were highlighted in Figures 6 and 7, the asymptotic behavior seen in Figure 5 is typical of the approach, and suggests that only 5 to 10% of cells can be used to capture the bulk of the spatial behavior of the model. The time needed for the interpolation approach itself is minimal: on a 2003-era PC computer, using the software package IDL (Research Systems Inc., Boulder CO), the input-space interpolation took only 5–50 min, a small addition to the much larger cost of running the model (12 computer days for the wall-to-wall run, for example). The order-of-magnitude savings in modeling effort can translate into greater exploration of model behavior and testing, easier development of spatial sensitivity analyses, broader testing of climate change scenarios, or modeling at finer grain sizes. The structure described here could be applied to any ecological model that does not consider cell-to-cell interactions, including other common biogeochemical models such as CENTURY (Parton et al. 1987), PnET (Aber and Federer 1992), and others (VEMAP Members 1995).

Conclusion

We have described an approach to spatial modeling that has the potential to improve modeling efficiency by nearly an order of magnitude, while retaining most of the spatial variation in model output. The approach is built on the recognition that variation in many biogeochemical models is solely a function of spatial variation in input drivers, and that modeling in input-variable space is more efficient than in geographic space. By reducing computational burden in spatial modeling, the approach can allow more time for greater exploration of model behavior under different scenarios or conditions.

Acknowledgments

The first author was supported under a National Aeronautics and Space Administration Earth System Science fellowship. Significant aid in model

set up was supported under by a grant from U.S. Environmental Protection Agency's Science to Achieve Results (STAR) Program (Grant # R-82830901-0). Although the research described in the study has been funded wholly or in part by the U.S. Environmental Protection Agency's STAR program, it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

References

Aber J.D. and Federer C.A. 1992. A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. Oecologia 92: 463–474.

Acevedo M.F., Pamarti S., Ablan M., Urban D. and Mikler A. 2001. Modeling forest landscapes: parameter estimation from gap models over heterogeneous terrain. Simulation 77: 53–68.

Alexandrov G.A., Oikawa T. and Yamagata Y. 2002. The scheme for globalization of a process-based model explaining gradations in terrestrial NPP and its application. Ecol. Model. 148: 293–306.

Band L.E., Peterson D.L., Running S.W., Coughlan J., Lammers R., Dungan J. and Nemani R. 1991. Forest ecosystem processes at the watershed scale: basis for distributed simulation. Ecol. Model. 56: 171–196.

Box G.E.P. and Draper N.R. 1987. Empirical Model-Building and Response Surfaces. John Wiley & Sons, New York.

Burke I.C., Kittel T.G.F., Lauenroth W.K., Snook P., Yonker C.M. and Parton W.J. 1991. Regional analysis of the central Great Plains. BioScience 41: 685–692.

Coops N.C. and Waring R.H. 2001. Estimating forest productivity in the eastern Siskiyou Mountains of southwestern Oregon using a satellite driven process model, 3-PGS. Can. J. For. Res. 31: 143–154.

Franklin S.E. 2001. Modeling forest net primary productivity with reduced uncertainty by remote sensing of cover type and leaf area index. In: Hunsaker C.T., Goodchild M.F., Friedl M. and Case T. J. (eds), Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications. Springer-Verlag, New York, NY pp.402.

Friedman L.W. 1996. The Simulation Metamodel. Kluwer Academic Publishers, Norwell, Massachusetts.

Garman S.L. 2004. Design and evaluation of a forest landscape change model for western Oregon. Ecol. Model. 175: 319–337.

Jongman R.H.G., ter Braak C.J.F. and van Tongeren O.F.R. 1995. Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge.

Kern J.S., Turner D.P. and Dodson R.F. 1997. Spatial patterns in soil organic carbon pool size in the northwestern United States. In: Lal R., Kimbal J.M., Follett R. and Stewart B.A. (eds), Soil Processes and the Carbon Cycle. CRC Press, Boca Raton, pp. 29–43.

Law B.E., Turner D.P., Lefsky M., Campbell J., Guzy M., Sun O., Van Tuyl S. and Cohen W.B. 2004. Disturbance and climate effects on carbon stocks and fluxes across Western Oregon USA. Global Change Biol. 10: 1429–1444.

Myers R.H. and Montgomery D. 2002. Response Surface Methodology: Process and Product Optimization using Designed Experiments. John Wiley & Sons, New York.

Ollinger S.V., Aber J.D. and Federer C.A. 1998. Estimating regional forest productivity and water yield using an ecosystem model linked to a GIS. Landscape Ecol. 13: 323–334.

Parton W.J., Stewart J.W.B. and Cole C.V. 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Sci. Soc. Am. J. 51: 1173–1179.

Peters D.P., Herrick J.E., Urban D.L., Gardner R.H. and Breshears D.D. 2004. Strategies for ecological extrapolation. Oikos 106: 627–636.

Running S.W. and Hunt E.R.J. 1993. Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. In: Field C.B. and Ehleringer J.R. (eds), Scaling Ecophysiological Processes: Leaf to Globe. Academic Press, San Diego, pp. 388.

Thornton P. 1998.Regional Ecosystem Simulation: Combining Surface- and Satellite-based Observations to Study Linkages between Terrestrial Energy and Mass Budgets. PhD Dissertation, University of Montana.

Thornton P., Hasenauer H. and White M.A. 2000. Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. Agric. For. Meteorol. 104: 255–271.

Thornton P., Running S.W. and White M.A. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190: 214–251.

Thornton P.E., Law B.E., Gholz H.L., Clark K.L., Falge E., Ellsworth D.S., Goldstein A.H., Monson R.K., Hollinger D., Falk M., Chen J. and Sparks J.P. 2002. Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests. Agric. For. Meteorol. 113: 185–222.

Urban D., Acevedo M.F. and Garman S.L. 1999. Scaling fine-scale processes to large-scale patterns using models derived from models: meta-models. In: Mladenoff D. and Baker W. (eds), Spatial Modeling of Forest Landscape Change: Approaches and Applications. Cambridge University Press, Cambridge, pp. 70–98.

VEMAP Members 1995. Vegetation/ecosystem modeling and analysis project: Comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate change and CO_2 doubling. Global Biogeochem. Cycles 9: 407–437.

Waring R.H. and Franklin J.F. 1979. Evergreen coniferous forests of the Pacific Northwest. Science 204: 1380–1386.

White M.A., Thornton P.E., Running S.W. and Nemani R.R. 2000. Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: net primary production controls. Earth Interact. 4–003.

Williams M., Rastetter E.B., Fernandes D.N., Goulden M.L., Shaver G.R. and Johnson L.C. 1997. Predicting gross primary productivity in terrestrial ecosystems. Ecol. Appl. 7: 882–894.

Williams M., Rastetter E.B., Shaver G.R., Hobbie J.E., Carpino E. and Kwiatkowski B.L. 2001. Primary production of an arctic watershed: an uncertainty analysis. Ecol. Appl. 11: 1800–1816.