A STUDY ON THE INTEGRATION OF GIS AND EPIC MODEL: METHODOLOGY AND APPLICATION

GIS を利用した統合 EPIC モデルに関する研究：その方法論および応用

Guoxin TAN*, Ryosuke SHIBASAKI*

Abstract: Erosion Productivity Impact Calculator (EPIC) has been widely used to analyze the relationship between soil erosion and agricultural productivity. It has the ability to simulate yields for many different crops in field, and each crop has unique values for the model parameters. In this paper, a methodology to integrate Geographic Information System (GIS) with EPIC model is proposed. This integration applies a loose coupling approach that involves a commercial GIS package and EPIC program. Data are exchanged using either ASCII or binary data format between two software packages without a common user interface. A case of crop yield simulation is tested with GIS-based EPIC model in China. The work of data handling and result presentation can be rent to GIS. Examinations show that result simulated with GIS-based EPIC model is acceptable, and thus it can serve as an important tool for decision-maker dealing with land resource management at global level.

1. INTRODUCTION

The world is rapidly changing. Growing population densities lead to scarcity of land and widespread changes in land use. In order to support a sound natural resource management, it is necessary to understand the productivity of land resources at global level.

One well-known method to assess food production on global scale is FAO-based model (FAO 1978-80). It is a simple model, and is just used to predict potential land productivity. Erosion Productivity Impact Calculator (EPIC) is another very popularly used model to simulate crop yields. It was developed by the USDA to analyze the relationship between soil erosion and agricultural productivity at field level. The model integrates the major processes that occur in the soil-crop-atmosphere-management system, including hydrology, weather, erosion, nutrients, plant growth, soil temperature, tillage, plant environmental control and economics (Williams et al, 1990; Williams, 1995). Many test research about EPIC model have been performed using different data and parameters (Roloff et al, 1998; Bryant et al, 1992; Edwards et al, 1994). It seems that EPIC is well suited for relative comparisons of soils, crops and management scenarios (Bouzaher et al, 1993).

In order to improve simulating accuracy, actual site information is necessary. There are some
approaches used to estimate air temperature and precipitation at actual sites, such as climate-model simulations of temperature variability (Washington and Meehl, 1986), extraction from satellite observations (Spencer and Christy, 1990), and analyses based on station-network records. Interpolation approach based on station records is regarded as the most accurate or precise one and has been investigated extensively (Karl et al., 1989). This method uses a lot of correction factors and large amount of data in interpolating air temperature and precipitation at unsampled points in space from meteorological observation. Geographic Information System (GIS) have emerged as powerful tools in the management and analysis of large amount of basic data and information. It is being applied increasingly to a variety of situations, including the simulation of crop productivity and the evaluation of land resources (FAO/IIASA, 1991). The objective of this paper is to integrate GIS with EPIC model, so we can lend data handling to the application of GIS and enable EPIC simulation more efficiently on global scale.

This paper is organized into five sections. After a brief background introduction in Section 1, the methodology of integrating GIS with EPIC model and some interpolation methods are discussed in Section 2; then the material and data required in our research are introduced in Section 3; A test of rice productivity in China and result analysis will be given in Section 4, followed by conclusion in Section 5.

2. METHODOLOGY

The methodology for integration of EPIC model and GIS technique is briefly illustrated in Figure 1. Firstly we define Agro-ecological cells (AECs), which divide an area of land into smaller units with a unique combination of climate, soil and other related land attributes (FAO/IIASA, 1991). Secondly we create GIS database and provide EPIC model with the data in grid format. After operation variables such as planting and harvesting date are determined, crop yields can be simulated with EPIC model in each grid cell. There are three main components in this flowchart: EPIC model, the approach to integrate GIS with EPIC model, and data handling.

2.1 EPIC Model

EPIC uses a daily time step to simulate weather,
hydrology, soil erosion by wind and water, nutrient cycling, tillage, crop management and growth, and field-scale costs and returns. It uses a general plant growth model with crop specific parameters to simulate the growth of rice, wheat, maize, grain, sorghum, soybean and so on (Figure 2). In this model, interception of solar radiation is estimated as a function of the crop's leaf area index. The leaf area index is simulated with equations dependent upon heat units, the maximum leaf area index for the crop, a crop parameter that initiates leaf area index decline, and five stress factors.

Plant growth is constrained by water, nutrient, and temperature stresses. The potential biomass is adjusted daily if one of the plant stress factors is less than 1.0 using the product of the minimum stress factor. The water stress factor is computed by considering supply and demand. The temperature stress factor is computed by using a function dependent upon the daily average temperature, the optimal temperature, and the base temperature for the crop. The N and P stress factors are based on the ratio of accumulated plant N and P to the optimal values. The aeration stress factor is estimated as a function of soil water relative to porosity in the root zone (Williams, 1990).

Perennial crops maintain root systems through frost–induced dormancy and start regrowth when average daily air temperature exceeds the base temperature specified for the plant. The crop yield is estimated using the harvest index concept. Harvest index increases as a nonlinear function of heat units from zero at the planting stage to the optimal value at maturity. The harvest index may be reduced by high temperature, low solar radiation, or water stress during critical crop stages.

Two kinds of standard data sets are developed as EPIC input files. One is basic input file, which includes miscellaneous field information such as climatic data, soil data, and management information. The other consists of parameter files such as crop parameter file, tillage parameter file, pesticide parameter file, and fertilizer parameter file. These parameters for most of the major crops have been established and do not need to be modified if there is no specific knowledge or specific application.

2.2 Method for Integrating GIS with EPIC

In order to facilitate the storage, manipulation and handling of complex EPIC spatial information, it is very necessary to input all raw spatial data into a geographical information system. The nature of data handling and analysis, which involves data editing, conversion, interpolation and overlay, can lend to the application of GIS. With the aid of GIS, it is possible for the EPIC model to simulate crop yields efficiently on global scale, and to allow a flexible presentation of results according to user needs.

There are several different approaches to integrate GIS with other modeling, such as embedding method, loose coupling and tight coupling method (Sui, 1998). In our research, the loose coupling approach is used to integrate GIS with the EPIC model. This approach uses two different packages directly. One is a standard GIS package (Arc/Info) and another is EPIC program (EPIC version 5300). They are integrated by combining various data layers on the physical aspects of agricultural envi-

![Figure 2 EPIC Model](image)
ronments such as soil, landform and climate, via data exchange using either ASCII or binary data format between these two packages, which do not have a common user interface. The advantage of this approach is that redundant programming can be avoided.

Map input, data handling, spatial analysis, and map output capabilities of GIS are used for preparation of land resource database required by the EPIC model. Data transferring programs are developed independently. So data can flow from GIS-created database into the EPIC model and modeling results can be transferred to GIS for further processing and presentation.

One more important point to be noted about the EPIC model is that it was not designed for the application of GIS. In fact, some EPIC parameters such as crop management practice data are very difficult or impossible to derive from GIS. Crop management descriptions must specify the timing of individual operations either by the date or by the fraction of growth period. Nonetheless, a compromise approach has been developed. In this approach, automatic processes provided by the EPIC model such as automatic irrigation and fertilization are adopted. The planting and harvesting date of each crop type each grid cell are determined automatically by comprehensive consideration of weather and crop condition. A modified EPIC program is developed to fit into these needs. This modified program can access field data directly from multiple input layers in GIS grid format.

2.3 Data Handling Method

The data required by EPIC model is firstly entered into GIS. They have been obtained from a multitude of sources, originally having many different formats and scales. For example, weather station records are stored in tabular format; soil, land use and DEM data have different map scales; planting and harvesting dates are unavailable, and thus need to be predicted or estimated. One of the important tasks of GIS data handling is to derive real site (grid) spatial data, and to output these data in grid format.

Weather station network with long-term records and good spatial coverage is uncommon (Willmott et al., 1991). Two problems arise from this paucity of well-conditioned observational networks. Estimating a time-averaged weather or climate variable at unsampled location by spatial interpolation is relatively unreliable, and in turn, areal averages made from the network observations can be biased. Topographically informed interpolation is one way to enhance the traditional interpolation of air temperature, which exploits the relationship between air temperature and elevation. Elevation effects on air temperature have been statistically included by several authors (Ishida and Kawashima, 1993).

Another approach is to make use of a second air temperature or precipitation field, one observed over a different time period but on a much higher spatial resolution station network (Willmott and Robeson 1995)—termed Climatologically Aided Interpolation (CAI). This approach is done by interpolating climatologically averaged station air temperatures or precipitation observed on the higher-resolution station network to the locations of the stations in the lower-resolution network, but at unsampled location.

Interpolation methods by distance weighting, such as applications of spatial regression, thin-plate splines, kriging and inverse-distance weighting are popular because of its simplicity and relative accuracy (Willmott, 1995). Each performs somewhat differently, but all estimate climate with the accuracies on the same order (Robeson, 1994). In order to reduce the uncertainty in monthly climatic interpolation due to factors such as inadequate of sample points or ill-distribution of sample points in external interpolation, a quality control model is proposed here. The formula is
\[ \Delta T = \frac{1}{n} \sum_{i=1}^{n} \Delta T_i \]
\[ \sigma = \sqrt{\frac{\sum_{i=1}^{n} (\Delta T_i - \Delta \bar{T})^2}{(n-1)}} \]

Where \( \Delta T_i \) is the difference between the lower-resolution station of interest \( T_i \) and the climatological \( \bar{T}_i \) in high-resolution, \( n \) is the number of lower-resolution stations, \( \Delta \bar{T} \) is average of difference, and \( \sigma \) is standard deviation.

Suppose \( \Delta \bar{T}_j \) is the interpolated value at grid \( j \). If there are not enough stations in the fix distance nearby grid \( j \), then we use \( \Delta \bar{T} \) instead of \( \Delta \bar{T}_j \). Consulting the table of standardized normal distribution, we know the probability that a certain residual is larger than triple standard deviation \( |\Delta \bar{T}_j| > 3\sigma \) to be 0.003. So if \( \Delta \bar{T}_j \) is more than \( \Delta \bar{T} + 3\sigma \), we think interpolation in grid \( j \) is ill-conditioned and make \( \Delta \bar{T}_j = \Delta \bar{T} + 3\sigma \). Otherwise if \( \Delta \bar{T}_j < \Delta \bar{T} - 3\sigma \), then we use \( \Delta \bar{T}_j = \Delta \bar{T} - 3\sigma \).

3. MATERIALS AND DATA REQUIREMENTS

EPIC is a sophistical model. It requires a very large amount of input data. Only important input data used in our examination, such as weather, soil and management data, are introduced here.

3.1 Weather Data

EPIC model uses a stochastic weather generator to generate daily weather from monthly climatic parameters. The basic data set needed for each site is a record of monthly maximum and minimum temperature, precipitation, STD of maximum daily air temperature, STD of minimum daily air temperature, STD of daily precipitation, skew coefficient for daily precipitation, probability of wet day after dry day, and probability of wet day after wet day.

The weather data we use in our test is from Global Daily Summary produced by National Climatic Data Center with time series of daily maximum and minimum temperature, daily precipitation, and elevation for 256 available terrestrial stations in China over the period from 1980 to 1990. Kriging method aided by climatologically and topographically interpolation with quality control model is applied. A monthly global long-term average data interpolated from nearly 18000 terrestrial stations recorded from 1890 to 1989 by Legates and Willmott (1990) serves as high-resolution climatology. A very high-resolution DEM at a spatial resolution of 5' is also used to obtain the georeferenced heights for grid networks.

3.2 Soil Data

EPIC can accept up to 20 parameters for 10 soil layers. However, only a minimum of 7 parameters are required: depth, percent sand, percent silt, bulk density, PH value, percent organic carbon, and percent calcium carbonate. Other soil parameters can be estimated by EPIC itself. In addition, specifying data for 3 to 4 layers should be adequate for most application.

The Spatial Database of Soil Properties (test version, 1999) is an accessible soil data set on pedosphere properties on a global scale. This soil database is provided by The Global Soil Task cooperated by the Data and Information System (DIS) framework activity of the International Geosphere–Biosphere Programme (IGBP) (Scholes et al., 1995). The highest resolution of this database is 5'. EPIC soil parameters, including percent sand, percent silt, bulk density, PH, percent organic carbon, and percent calcium carbonate, can be generated from Spatial Database of Soil Properties directly. In our test, four soil layers are applied. The soil-depth intervals are 0–10cm, 10–30cm, 30–50cm and 50–80cm respectively.

3.3 Management Data

EPIC requires detailed descriptions of management practices. These descriptions must specify the timing of individual operations either by date or by fraction of the growth period (i.e., by heat units). EPIC allows the user to simulate complex crop rotations with a variety of irrigation, fertilizer,
pesticide, and tillage control options. There are two options for irrigation and fertilizer scheduled application in EPIC program: manually and automatically. Only automatic option is applied in GIS-based EPIC model, because it is impossible to obtain real fertilizer and irrigation schedules in different sites (grid cells). Automatic mode is controlled by parameters of maximum annual fertilizer and irrigation volume.

Maximum annual fertilizer volume applied for a crop is derived from county-based yearly social-economic statistical database. The statistical data included items such as population, crop yields, annual chemical fertilizer volume, and total crop area in China from year 1980 to 1991. We link this table data with China county boundary spatial data and create fertilizer application file with ARC/INFO.

Maximum annual irrigation volume is derived from land use data, due to unavailability of this kind of data. Because almost all the paddy fields in China are irrigated, we apply enough volume of irrigated water in the EPIC model processing if land use type is paddy. A list of 62 coverages taken from the 64 sheets of the 1:1,000,000 Land-use Maps of China by the Australian Center of the Asian Spatial Information and Analysis Network, Griffith University, are used.

About management parameters, commonly used practices were assumed in the GIS-based EPIC model. These practices consisted of tandem disking, listing, field cultivation prior to planting, and offset disking after harvest.

4. RESULT AND DISCUSSION

The test area covers all of China from longitude 73°50' to 135°0' east and from latitude 18°5' to 53°35' north. The grid resolution is 5' by 5'. All spatial data such as weather station data, fertilizer data, and landuse data, are converted (interpolated) to grid format. Figure 3 shows the interpolation results of 10-year mean monthly maximum air temperature in the period 1980 to 1989.

In order to check interpolation accuracy, 20 check points are selected randomly with a well distribution over China and removed from weather station network for interpolation. Interpolation errors can be estimated by these check points. We use Mean Absolute Errors (MAEs) and Standard Deviations (SDT) to evaluate the effectiveness of interpolation schemes. Table 1 shows the monthly air tempera-

![Figure 3](image-url)
Table 1  Interpolated temperature errors with or without control model (unit: Deg C)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>With</td>
<td>MAEs</td>
<td>0.40</td>
<td>0.36</td>
<td>0.34</td>
<td>0.32</td>
<td>0.32</td>
<td>0.30</td>
<td>0.33</td>
<td>0.31</td>
<td>0.30</td>
<td>0.32</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.74</td>
<td>0.68</td>
<td>0.62</td>
<td>0.62</td>
<td>0.68</td>
<td>0.75</td>
<td>0.76</td>
<td>0.74</td>
<td>0.67</td>
<td>0.66</td>
<td>0.70</td>
</tr>
<tr>
<td>Without</td>
<td>MAEs</td>
<td>0.48</td>
<td>0.41</td>
<td>0.42</td>
<td>0.39</td>
<td>0.40</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.36</td>
<td>0.39</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.92</td>
<td>0.86</td>
<td>0.73</td>
<td>0.75</td>
<td>0.90</td>
<td>0.91</td>
<td>0.91</td>
<td>0.90</td>
<td>0.84</td>
<td>0.83</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Table 2  Input parameters and selections used for the EPIC runs

<table>
<thead>
<tr>
<th>EPIC miscellaneous parameter file</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of the season when stresses start affecting harvest index</td>
<td>0.5</td>
</tr>
<tr>
<td>Water stress factor</td>
<td>1.0</td>
</tr>
<tr>
<td>Hargreaves PET equation exponent</td>
<td>0.5</td>
</tr>
<tr>
<td>-coefficient</td>
<td>0.0023</td>
</tr>
<tr>
<td>Automatic fertilizer</td>
<td></td>
</tr>
<tr>
<td>-Mineral N = NO₃⁻ - N + NH₄⁻ N (fraction)</td>
<td>0.46</td>
</tr>
<tr>
<td>-Ammonia Nit. As a fraction of mineral nitrogen</td>
<td>1.00</td>
</tr>
<tr>
<td>Potential heat units</td>
<td>1800-2000</td>
</tr>
</tbody>
</table>

In addition to the field input data discussed in Section 3, other important non-default parameters applied in our test are listed in Table 2. These parameters have been adopted from Kiniry et al. (1995), except for the potential heat units, which changed from 1800 to 2000 in different regions (Tang and Ye, 1997). For each crop type and grid cell, a program to determine possible crop combination, and the starting and ending dates of growing season is developed. This program can ensure best possible crop yields for both rain-fed and irrigated conditions, and also guarantee maximum adaptation in simulations with year-by-year historical weather conditions, or under climate distortions when analyzing different climate change scenarios. Figure 4 shows the estimated planting date for rice crop in China. Rice crop is simulated only on paddy fields. There are three sub-categories in cultivated land in China land use map: paddy field, irrigated field (except paddy), and non-irrigated field (dry land) (Figure 5).

As an example, we consider two practices for GIS-based EPIC model under different climate and fertilizer application data. One is using 10-year means in the 80s. Another is using specific year data in 1987. Figure 6 shows the mean rice yield in each paddy field in the 80s. In order to analyze simulated results, four special provinces of Liaoning, Jiangsu,
Guangdong and Sichuan, which are located in the north, east, south and west of China respectively, are selected (Figure 7). Overlaying Figure 6 with Figure 7, we can calculate the yield of rice per hectare in different provinces. Table 3 gives the simulated results and statistical values.

We can see from Table 3 that the deviations of simulated yield from statistical yield are less than 0.5 T/HA, except the yield in Jiangsu province in the year 1987. Rice yield tends to be underestimated by an average of 6% in Liaoning, Jiangsu, and Sichuan province, but over estimated by 7.6% in Guangdong province. This is because rice yields are simulated based on these two hypotheses: (i) rice crop parameters established by USDA are applicable to China; and (ii) the field management is same all over the China. In fact, since 1949 China has successfully bred more than 3400 kinds of high-quality species belonging to 40 crops, such as the breeding technology of haploid with anther (China Today Compile Committee, 1992). In addition, there are different field managements in different regions. For example, Guangdong has become the richest province since 1980. With the developing of economy, young and better-educated rural workers left rural areas in search of better paid jobs in the urban areas. So land becomes less central as a factor of production in Guangdong province. Despite these two hypotheses, the GIS based EPIC model is again able to satisfactorily estimate crop yields at regional or global level.

5. CONCLUSION

Because climate is the major determinant of crop yields (Darwin, 1998), any technique, which can improve climatic data quality, is of great benefit for the global land productivity evaluation. A topographically informed and climatologically aided interpolation method with quality control model is proposed to improve temperature and precipitation interpolation accuracy. Results show this interpolation method has a high quality and can be applied to other global interpolation of data from observational networks such as solar irradiances.

Agricultural recommendation and sustainable management of land resources need accurate crop

<table>
<thead>
<tr>
<th>Province</th>
<th>Type</th>
<th>Mean Rice Yields In 80s</th>
<th>Rice Yields In 1987</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>Estimated</td>
<td>Statistic</td>
</tr>
<tr>
<td>Liaoning</td>
<td>6.2</td>
<td>5.8</td>
<td>6.3</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>6.2</td>
<td>5.7</td>
<td>6.9</td>
</tr>
<tr>
<td>Guangdong</td>
<td>4.5</td>
<td>4.9</td>
<td>4.7</td>
</tr>
<tr>
<td>Sichuan</td>
<td>6.1</td>
<td>6.0</td>
<td>6.7</td>
</tr>
</tbody>
</table>
yield data on global or regional scale. This paper has presented a methodology for integration of GIS technique with field-based EPIC model. A case of rice simulation is tested in China. The results show that with the aid of GIS, EPIC can be applied at global level, because the simulated rice yield in the test area is acceptable. In addition, the EPIC model encompasses economics and plant environment control. So the GIS based EPIC model has a broad usage potential and will be an important support tool for the appraisal and planning of land resources. Due to the complexity and widespread practice of multiple cropping systems, the future researches are necessary on the application of GIS based EPIC model to other crops.

(受付日2000.10.31, 受理日2001.5.11)

REFERENCE


16) Sui, D.Z., 1998. GIS-based urban modeling :


