Multi-Scale Characterization of Land Use Pattern and its Changes in China

Youqi Chen, Peter H. Verburg* and Qingbo Zhou

Abstract

This paper analyzes the pattern of land use in China by GIS to reveal the factors that determine the distribution of the different land use types. Firstly correlation and regression analyses were used to identify the major explanatory variables from a large set of candidate determining factors. We found that the distribution of all the land use types in China was best described by a combination of different biophysical and socio-economic factors. The distribution of different crop types was closely related to climatic differences over the country. Emphasis was placed on the influence of the scale of analysis on the results of the study. On the basis of all these correlation and regression analyses, a multi-scale spatial analysis by GIS modeling on the distribution of different land use types has been carried out. Both resolution of the data and the extent of the study area influenced the revealed relations. The systematic and quantitative characterization of the land use distribution presented in this paper can be used in spatially explicit land use models.

Introduction

The natural land cover is heterogeneous at all levels of observation, due to the heterogeneity of the environmental conditions, such as temperature and moisture conditions (Woodward, 1987). The actual distribution, however, only occasionally results from physical limitations. More usually the ecologically allowable is a subset inside what is physically possible. This is mainly due to competition between individual plants or between ecosystems as a whole (Meisel and Turner, 1998). Human use of land has altered the structure and functioning of ecosystems (Vitousek *et al.*, 1997). Human activities override natural changes of ecosystems. Agriculture, forestry, and other land-management practices have modified entire landscapes and altered plant and animal communities of many ecosystems throughout the world (Ojima *et al.*, 1994). The most spatially and economically important human uses of land globally include cultivation in various forms, livestock grazing, settlement and construction, reserves and protected lands, and timber extraction (Turner II *et al.*, 1994). A better understanding of the determining factors of land use change is of crucial importance to the study of global environmental change (Turner II *et al.* 1994).

The study of land use in China is relevant as China has a long history of land use under ever increasing population pressure in a very diverse natural environment. In recent years, the high population pressure, in combination with economic development, has resulted in the conversion of arable land area into non-agricultural uses. Together with the loss of arable land through degradation this situation is undermining China's food production capacity (Huang and Rozelle, 1995).

The different processes determining the land use pattern have each their own optimal scale level at which the process can be studied (Fresco, 1995). For the system as a whole, there is not an optimal scale level (Levin, 1993). Therefore, a multi-scale approach, including the study of the system at different levels of integration,

Institute of Natural Resources and Regional Planning, CAAS, Baishiqiao Road 30, Haidian, Beijing 100081, P.R.China;
* Wageningen Agricultural University, The Netherlands

enables to analyze the system and variety in processes more comprehensively. The main objective of this paper is to provide a spatially explicit characterization of land use in China by the identification of the major determining factors and to observe the spatial scale dependency of the relations found.

Multi-scale issues

Land use patterns result from many processes within the landscape that act over a large number of scales and are linked together in hierarchies. Phenomena occurring at any level are affected by mechanisms occurring at levels below and above (Gibson *et al.*, 1998). Ecologists have stated that the issues of scaling are a fundamental problem in ecology, if not in all of science (Levin, 1993).

Scale refers to the spatial or temporal dimension used to measure and study a phenomenon. Each analysis of spatial pattern should incorporate explicitly or implicitly scale into the process of identifying research objects: the very act of identifying a particular pattern means that scale, extent, and resolution have been employed. These choices over scale, extent, and resolution critically affect the type of pattern that will be observed: patterns that appear at one level may be lost at lower or higher levels; patterns that occur over one extent of a dimension may disappear if the extent is increased or decreased (Allen and Hoekstra, 1991). Consequently, changes in scale often imply a change in correlation structure. Occurrences that are positively correlated in a large-scale universe can become negatively correlated in a smaller universe. The most obvious example is the effect of urban extension on the utilization of arable land. On a large scale, because of the great potential market, more advanced technologies and the big financial pool provided by the development of city, extensity and efficiency of all the agricultural activities will be promoted. Therefore, agriculture in the suburban areas is more developed and modernized. However, on a lower scale, due to the urban anticipation caused by urban sprawl, the farmers in the nearby area of a city are expecting that their land will be occupied for urban uses for a higher renting price. Therefore, the most popular way to handle their land at present is to use it extensively, and even for some time to leave it fallow. All these imply that the effect of urban growth on arable land use varies with the scale. Thus, observations and theories derived on one scale may not apply on another.

Although the concepts of scale dependence are most often applied to natural vegetation on scales ranging from a couple of centimeters to a few kilometers, they are also highly relevant to the study of agro-ecosystems (Fresco and Kroonenberg, 1992). Studies by Veldkamp and Fresco (1997) have shown that the analysis of land use patterns is subject to scale dependency also at grain sizes ranging from a few kilometers to the national level. Therefore, the scale dependency of land use patterns in China can be studied in a systematic way by varying the resolution and extent of analysis.

Data and methods

1 Data

Land cover and land use are constrained by environmental factors such as soil characteristics, climate and topography (Turner II *et al.*, 1994) while human factors determine where and to what extent land cover is modified at a certain location. Based on a review of the literature dealing with the factors determining land use, a large set of candidate determining factors was selected for our analysis. And for these parameters spatially explicit information was collected. Because detailed information for all factors is not always available, variables were included which can be regarded as proxies for the determining factors. Illiteracy can, for example, be considered as an indicator of “access to information and means”, while “distance to urban centers” determines the access to markets, both for selling and access to means of production. A list of all variables contained in the spatially explicit database is presented in the Appendix. Data based on statistics are valid for 1991.

All the data have to be converted to a regular grid to match the representation of the different data sources and ease the analysis. The basic grid size, to which all the data are converted, is 32x32 km (~1000 km^2), which is equivalent to the average county size (the basic administrative units for all countrywide statistical data) in the eastern part of China. The average county size in the western part of China is much larger, due to the presence of large uninhabited areas (*i.e.* deserts and mountains). All the statistics were recalculated excluding the deserts as designated in a vegetation map (CAS, 1979/1996). All the other land use types in the statistics are assumed to be found in the area outside the deserts while population density in the deserts is assumed to be zero.

The dependent variables in our analysis, the distribution of the different land use types, are represented by designating the relative cover of the land use types in each grid cell. This data representation differs from the traditional way of representing data, where for each grid the dominant land use type is identified. The data representation is a direct result of the data contained in the census and other statistical surveys, which indicate for each administrative unit the area occupied by the different land use types.

2 Aggregation levels and regional stratification

To test the hypothesis that relationships between determining factors and the land use distribution will change with the scale of analysis, we used artificial, grid-based data sets that differ in grid cell size. These artificial aggregation levels were obtained through aggregation of the dependent and independent variables by averaging the data of 2 by 2 cells (64x64 km), 3 by 3 cells (96x96 km), 4 by 4 cells (128x128 km), 5 by 5 cells (160x160 km) and 6 by 6 cells (192x192 km). The six aggregation levels that were obtained this way are schematically presented in Fig. 1. Aggregation of data by a grid-based plurality procedure, where the most

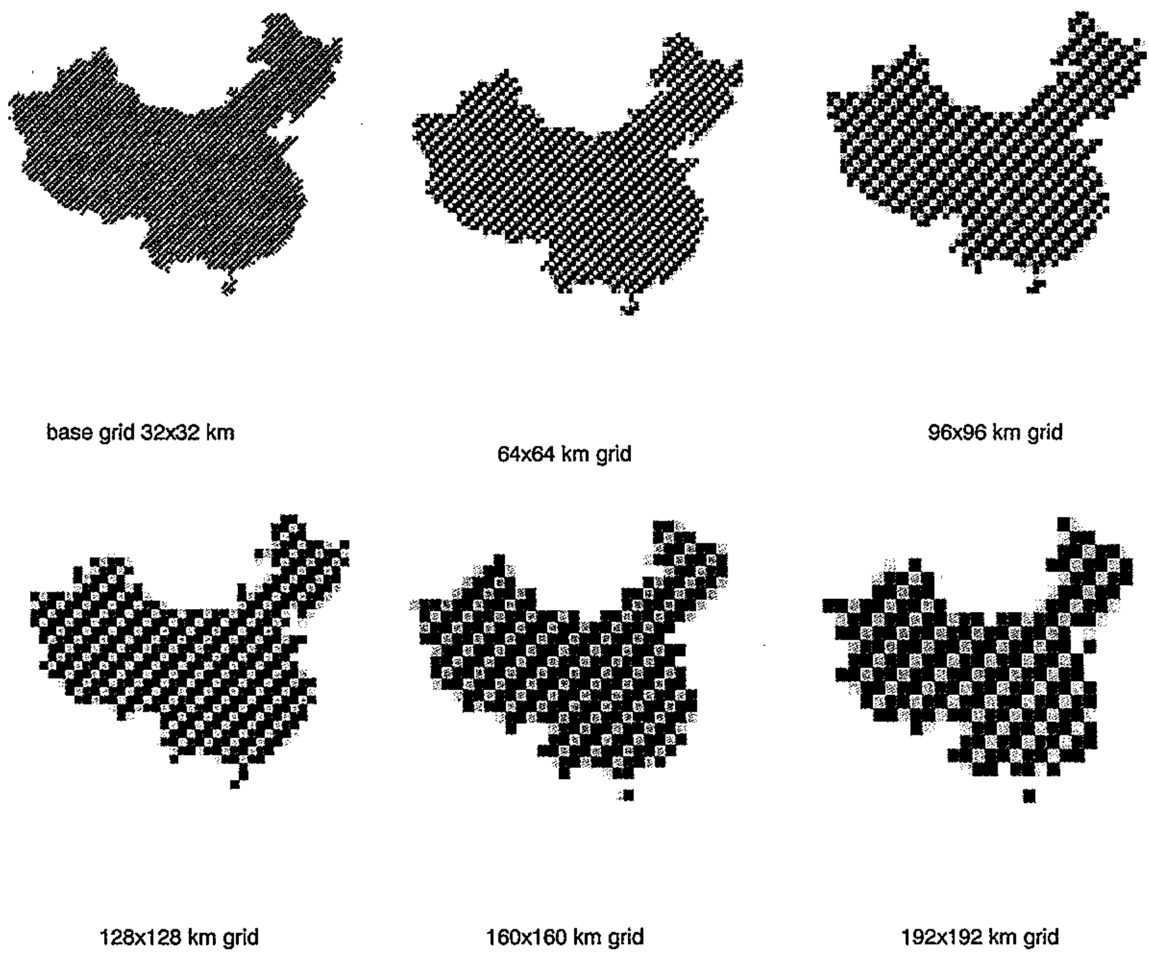

Fig. 1 Schematic representation of aggregation levels used in the analysis

frequently occurring sub-grid cover type is used to code each grid cell at each degraded resolution, introduces bias, as some class proportions will decrease and others will increase with scale depending on the spatial and probability distribution of the cover types. However, all the information is retained during our aggregation procedure because of the sub-pixel information data structure applied in this study.

Next to the grain size, it is hypothesized that also the extent of the analysis will influence the results. Therefore all the analyses were performed for the country as a whole as well as for eight individual regions. The eight regions have been defined based on geographical/natural conditions, demographic and economic features, and province-level administrative subdivisions (Fig. 2). This stratification differs slightly from the frequently used regionalization of the United States Department of Agriculture (Crook, 1993b) which includes the densely populated Sichuan province in the same region as the sparsely populated, remote Tibetan plateau. In our subdivision two separate regions have been created.

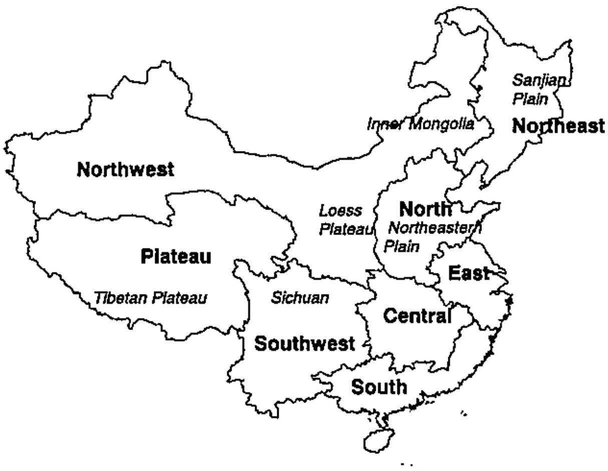

Fig. 2 Regional subdivision of China

3 Statistical methods

Statistical techniques are well suited to achieve the investigative objectives of this study. Simple correlation analysis was used to quantify the relation between individual candidate determining factors (independent variables) and the land use distribution (dependent variables). If no common pattern exists for a candidate determining factor, the correlation will likely reflect this fact in nonsignificant results. It is essential to emphasize that an explanatory variable found to be significantly related to the land use distribution, does not necessarily imply that it is a direct cause of the land use pattern; correlation does not necessarily imply causality.

Multiple regressions are used to derive comprehensive models that describe the pattern of land use as a function of a combination of determining factors. Multi-collinearity is very common in studies of complex systems, as many candidate determining factors are closely related (*e.g.* geomorphology and soil). A step-wise regression procedure was used to select the determining factors that yield a significant contribution to the regression model while multicollinearity is at acceptable low levels. For each regression equation, the adjusted coefficient of determination (R^2) is a measure for variation in the relative cover of the specific land use type that can be explained by the model variables. The standardized betas indicate the number of standard deviation changes in the dependent variable associated with a standard deviation change in the independent variable if all other variables are held constant. They therefore indicate the relative importance of a variable for land use change in a given regression equation.

For reasons of simplicity, we have only evaluated a linear regression. This approach will reduce the explaining power of the regression models, as some determining factors will be related to land use in a nonlinear way. On the other hand, relationships that are nonlinear on fine scales tend to be flatter (*i.e.* more linear) on coarser scales (for more details see the discussion by Rastetter *et al.*, 1992).

Geographical patterns nearly always a exhibit a positive spatial autocorrelation. Stronger spatial autocorrelation indicates a stronger tendency for like values to cluster. Therefore, there is generally a lack of independence among observations in spatial data sets. Part of this autocorrelation is induced by the use of the arbitrary delineation of spatial units of observation (*e.g.* census tracts, county boundaries) and problems of spatial aggregation. In addition, and separately from this measurement issue, the inherent spatial organization and spatial structure of phenomena will tend to generate complex patterns of interaction and dependencies which are of interest by themselves. The existence of autocorrelation among regression residuals contradicts the assumption, critical for the validity of the statistical inferences associated with regression, that errors are independently distributed. When this condition is not satisfied, t statistics do not give reliable indications of the significance of the variables. Correcting the statistical procedures for autocorrelation is not feasible as no satisfactory methods are yet available to deal with spatial autocorrelation in spatial methods. Therefore spatial autocorrelation was not corrected for in the models.

It is also important to note that sample sizes varied between scales and regions. This makes the comparison of correlations and regression coefficients by means of the standard errors, t-values, and P-values problematic. These statistics should therefore be interpreted as general indicators.

Space limitations allow us to present only representative results. All the presented regressions and correlations are statistically significant ($P < 0.001$) unless otherwise indicated. Presented R^2 values are adjusted R^2. However, due to the large sample size, reported and adjusted R^2 were nearly identical.

Results and interpretations

1 Land use patterns

The results of a correlation analysis of the distribution of the main land use types is presented in Table 1. The analysis was performed for the country as a whole at the finest data resolution (32x32 km). For the different land use types, the 10 variables that show the highest correlation with the distribution of land use are listed.

The distribution of cultivated land is strongly correlated with the distribution of population, especially with the distribution of the agricultural population. This relation shows the rural character of China, where population and agriculture are strongly clustered. Other important factors explaining the distribution of cultivated land are the suitability of the soil for irrigated rice cultivation, elevation, temperature and some hydrological conditions. This means that cultivated land is also strongly related to the suitability of the soil for agriculture. The distribution of horticulture is determined by a similar set of variables. However, here we find that temperature,

Table 1 Pearson correlation coefficients* for the determining factors most related to the distribution of the different land cover types (n=9204)

Cultivated land		Horticulture		Grassland		Forest		Unused land		Built-up land	
PAG91	0.81	PTOT91	0.45	TMP_10C	-0.52	PRC_50M	0.67	PRC_RNG	-0.64	PTOT91	0.82
PAGLF91	0.79	PAG91	0.44	TMP_AVG	-0.51	PRC_TOT	0.66	PRC_50M	-0.58	PAG91	0.80
PTOT91	0.79	PRURLF91	0.44	TMP_MAX	-0.51	PRC_RNG	0.66	PRC_TOT	-0.52	PRURLF91	0.77
PRURLF91	0.77	PRC_TOT	0.42	MEANELEV	0.47	SUN_TOT	-0.56	SUN_TOT	0.42	PAGLF91	0.75
S1IRRPAD	0.60	PAGLF91	0.42	ILLIT	0.44	GEOMOR1	0.53	GEOMOR3	0.39	PURB91	0.58
NSIRRPAD	-0.59	TMP_MIN	0.41	TMP_MIN	-0.40	GEOMOR5	-0.46	PAG91	-0.36	S1IRRPAD	0.58
MEANELEV	-0.54	PRC_RNG	0.41	PAG91	-0.39	TEXT1	-0.42	PTOT91	-0.36	NSIRRPAD	-0.55
SMAXHIGH	0.50	TMP_AVG	0.40	PAGLF91	-0.38	GOODDRAI	0.41	PAGLF91	-0.36	BADDRAIN	0.50
TMP_MAX	0.49	TMP_10C	0.38	PRURLF91	-0.38	TMP_10C	0.40	PRLPER91	-0.35	MEANELEV	-0.49
BADDRAIN	0.49	PRC_50M	0.37	PTOT91	-0.38	TMP_MIN	0.39	PRURLF91	-0.35	PRC_RNG	0.47

*all significant at $P<0.0001$

especially the minimum temperature and the number of months with temperatures above 10℃ are important determinants. Horticultural crops often need specific environmental conditions. Especially frost limits the distribution of most horticultural crops. The distribution of grasslands, forests and unused land is very much determined by environmental conditions. From the correlations it is clear that these land use types are mainly found in areas unsuitable for agriculture far away from concentrations of population. As the largest areas of grassland in China are found in the highly elevated Plateau region and in Inner Mongolia, grasslands are correlated with low temperatures and high elevations. The high and positive correlation with the percentage of the population that is illiterate can be explained by the remoteness of this area. Most of the forests are found in the South of China and in the southwestern mountain fringes. This area is also characterized by the highest amounts of precipitation in China and relatively high temperatures. The positive correlation with mountainous geomorphology and negative correlation with plain land shows the location of forest on sloping land, unsuitable for agriculture. The category of unused land is dominated by the large desert areas, leading to high correlations with low precipitation and many sunshine hours. Finally, the distribution of built-up land is correlated with the same factors as the distribution of cultivated land. Obviously, the strongest correlation is found with the total population density followed by biophysical variables indicating the suitability of the land. These correlations show that the expansion of built-up land leads to a loss of mainly high-quality, primary farmland.

Since many variables indicated in Table 1 are highly correlated, some variables yield non-significant contributions in regression models. Multiple regression models, as derived for the different land use types, indicate that the most comprehensive description of the distribution of land use is obtained when at least one variable is included concerning the population distribution, one variable indicating the soil quality, one geomorphological variable and one climatic variable. Explanations range between 0.3 for horticulture and 0.8 for cultivated land.

2 Effects of spatial scale

The influence of spatial scale on the relations found above was studied by comparing the effects of both grain size and the spatial extent of the study.

The effect of grain size was studied by comparing the results of a correlation and regression analysis at six different artificial aggregation levels. Fig. 3 displays the correlation coefficient of a selection of variables for

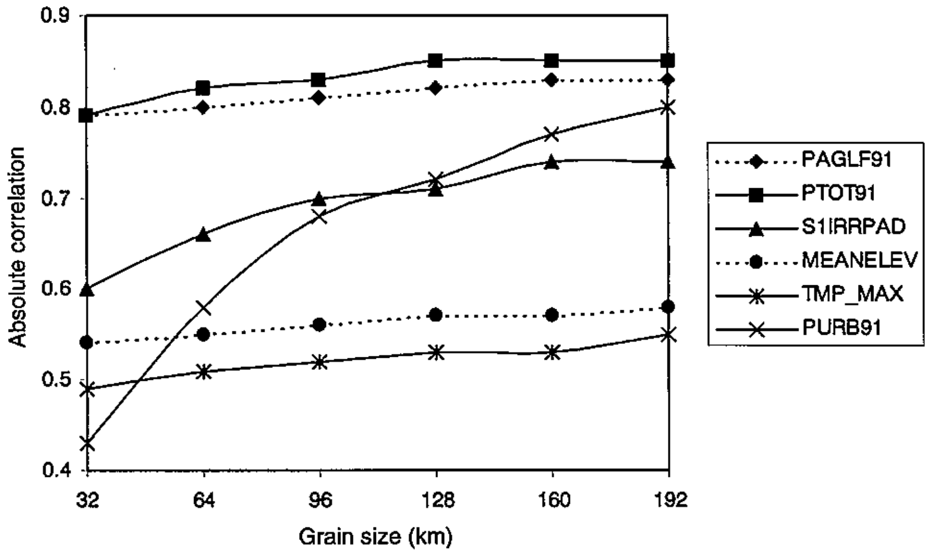

Fig. 3 Absolute correlation between the relative cover with cultivated land and a number of determining factors at different aggregation levels (all correlations significant at P<0.0001)

cultivated land at the six aggregation levels. For all the variables the correlation coefficient increased with grain size. This increase in correlation coefficient may have been caused by areal aggregation, which reduced the variability. However, the correlation coefficients were not inflated consistently with successive aggregation and for the different variables. Whereas the correlation coefficients of maximum temperature (TMP_MAX) and mean elevation (MEANELEV) remained approximately constant over the range of grain sizes, the coefficients for soil suitability (S1IRRPAD) and urban population density (PURB91) increased strongly with grain size. The change in the correlation structure is very much related to the spatial variability of the variable and the distance over which the parameter affects land use. The strong increase in correlation between urban population and cultivated land can be explained by the influence a city has on land use in the surrounding area. With the increase in grain size, an increasing part of the cultivated land around a city falls within the same grid-cell as the urban population, yielding high correlations.

Changes in the correlation structure also result in different multiple regression models at different aggregation levels. Table 1 presents the multiple regression models derived for the distribution of horticultural land at different aggregation levels. The overall explanation of horticultural land remained low at all aggregation levels, presumably due to the different environmental requirements of all crops lumped under horticultural crops (*e.g.* tea, mulberry plantations and orchards). Although the main variables that explain the pattern of horticultural land were similar over the range of scales, there were differences in the relative importance of the different variables and different additional variables were added that mediate the relations. Similar to the results for cultivated land, the urban population became a more important variable at higher aggregation levels.

The influence of the extent of analysis was tested by correlation and regression analysis after stratification of the base grid data according to regions (Fig. 3). The most important results for a number of regions are given below. Numbers between brackets indicate the correlation coefficients.

1) In the Northeast, the land use distribution is very much dependent on geomorphology and temperature. Cultivated land and orchards are all situated in the plain areas (northeastern plain and Sanjiang plain), which have comparatively favorable temperatures and deep (DEEP: 0.67), fertile soils (FERT3: 0.34), suitable for agriculture. The mountainous areas surrounding the plain have a harsh climate only suitable for forests and grasslands. Forests and grasslands are mainly distributed according to precipitation. Areas with higher precipitation are afforested (PRC_TOT: 0.41) while the drier areas are mainly covered by grassland (PRC_TOT: -0.47).
2) In the Northwest, climatic parameters limit the distribution of land use types. In this dry region, cultivated land is only found in the wet and relatively warm areas while forests and grasslands are found in the relatively wet, but cooler areas. All the other parts of the region are too dry for plant growth (PRC_TOT: -0.71) and are therefore classified as unused land, *i.e.* desert. Typical for the region is the positive correlation between cultivated land and the area with loess deposits (GEOMOR2: 0.61). The correlation with the loess deposits also explains the lack of clearly negative correlations with slope and other variables representing mountainous land forms, as they are found in other regions. The intensive cultivation of the large loess-covered region explains these observations. The Loess Plateau, a very extensive area (530,000 km^2 - larger than Spain), has been used for agriculture since ancient times. In those times, the conditions were very favorable for agriculture: the land had a suitable humidity and forest-steppe or grassy vegetation, it possessed fertile and loose soil developed from loess, it was a flat land, which was easily reclaimed as farmland and was convenient for travel. However, the climatic conditions, with highly variable rainfall which falls in high-intensity rainstorms, together with the soil characteristics made the area very susceptible to soil erosion. Nowadays, the loess plateau is one of the most severely eroded landscapes in the world, which is steeply dissected by deep gullies and valleys. Therefore, the once prosperous agricultural region has turned into one of the most backward regions, which has lost most of its favorable characteristics for agriculture. With respect to the correlation results, this fact means that in earlier times the distribution was more

correlated with flat and fertile land, similar to the other regions. Nowadays the distribution can only be explained by the occurrence of high population densities and loess deposits.

3) In the southwestern region, which contains the foot slopes of the Himalayan Mountains, differences in altitude are large. This explains the importance of the elevation for the distribution of population and land use types. While agriculture is found at low elevations (MEANELEV: -0.72) grasslands are found at high elevations (MEANELEV: 0.75). Forests are not strongly correlated to the mean altitude in this region (MEANELEV: 0.07) but mainly found on steep slopes (SLOPE: 0.50). Typical for this region is the negative correlation between cultivated land and high soil fertility (FERT3: -0.48) and the positive relation between soil fertility and grasslands (FERT3: 0.64). These relations do not follow the national pattern but can be explained by the occurrence of fertile soils in the higher mountain ranges and plateaus, which are too high and too cold for agricultural use. Therefore, in spite of the lower soil fertility, agricultural activities need to be concentrated in the soils at lower locations.

4) In the Plateau region, agricultural land use is concentrated around the sparsely distributed centers of population (PTOT91: 0.79, PAG91: 0.97). Due to its extremely high elevation, the relation between forests and temperature is opposite to the relation found in other regions: here forest is positively correlated with temperature (TMP_MIN: 0.75). Grasslands and unused land (bare rock, etc.) occupy the remaining area.

The results show a great deal of resemblance but also clear differences between regions. When the country is analyzed as a whole, the general relation between land use and its determining factors can be found. After stratification, region-specific conditions appear in the results. Although the explaining power (R^2) of the regression in the individual regions is generally similar or slightly lower than the R2 of the regressions for the whole country, the overall explanation of the land use distribution is much clearer after stratification. In Fig. 4, this is visualized for cultivated land by presenting the actual distribution of cultivated land (A), the distribution as modeled by the regression equation valid for the country as a whole (B) and the distribution when the regressions are made for the eight individual regions (C). The overall explanation of the distribution of cultivated land equals 0.81 for the countrywide regression. When for all the regions the region-specific regressions are used, an overall correlation of 0.92 is obtained, showing the important contribution of region-specific factors. Especially in the northeastern region a much better representation of the distribution of cultivated land is found due to the inclusion of climatic and geomorphological variables into the regression equation for this region.

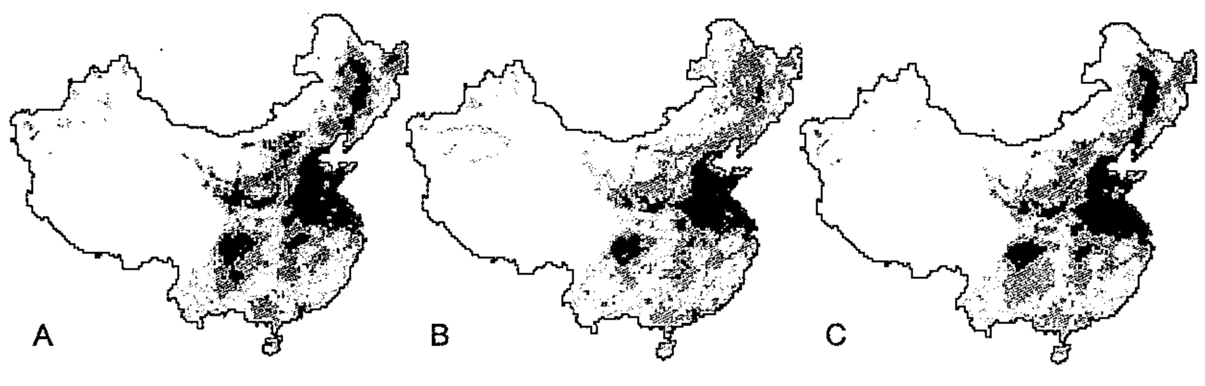

Fig. 4 A: Actual distribution of cultivated land (dark colors indicate high cover percentages)
B: Distribution of cultivated land as modeled with the country-wide regression equations
C: Distribution of cultivated land as modeled with the region-specific regression equations

3 Crop cultivation patterns

Chinese statistics traditionally distinguish grain crops, economic crops and other crops. The group of grain crops includes rice, wheat, corn, sorghum, millet, other miscellaneous grains, tubers (potatoes) and soybeans. Economic crops include oil crops (peanuts, rape-seed, sunflower, sesame, etc.), tobacco, sugar crops (sugar-beet and sugarcane), cotton, fiber crops (jute, hemp), medicinal crops and a miscellaneous category. Other crops include vegetables, melons, fodder crops and green manure. Table 2 gives correlations between a selection of different determining factors and the distribution of the different crop categories. The crop categories are represented by the percentage of the total harvested area that is devoted to the specific crop category. Additionally correlations are given for the percentage of all land in a grid-cell used for the cultivation of the crop category. From the results we can see that economic crops are more important in the cropping systems in areas that have higher average temperatures, i.e., the southern part of China. Grain crops are a more important part of the cropping systems in the drier and cooler northern part of China. In the fertile delta area of the country, farmers prefer a higher percentage of grain crops within the cropping system, while the percentage of economic crops does not show a significant relation with soil fertility. Most of the economic crops require a high input of labor for cultivation and harvesting. This can be seen from the correlation structure with the agricultural labor force density. A higher share of economic crops is found when the labor force is relatively large while lower labor force densities prevail where grains dominate the cropping system.

Table 2 Pearson correlation coefficients between crop distribution and a selection of determining factors for all grid cells containing more than 5% cultivated land (n=4643)

	% of Grain crops in all crops	% of Total land used for grain crop cultivation	% of Economic crops in all crops	% of Total land used for economic crop cultivation
TMP_AVG	0.43[a]	-0.00	0.15[a]	0.16[a]
PRC_TOT	-0.36[a]	-0.23[a]	-0.04[b]	-0.10[a]
FERT1	-0.35[a]	-0.43[a]	0.04[b]	-0.25[a]
FERT3	0.24[a]	0.38[a]	0.00	0.24[a]
S1MAIZER	0.27[a]	0.36[a]	-0.03	0.15[a]
PAGLF91	-0.20[a]	0.60[a]	0.14[a]	0.58[a]
GEOMOR5	-0.08[a]	0.47[a]	0.26[a]	0.50[a]

[a] significant at P<0.001
[b] significant at P<0.01

China has a wide variety of grain crops. However, their spatial distribution is often limited to part of the country. Table 3 shows how the ratio of the individual grain crops to all grain crops relates to a selection of determining factors. Climatic factors are important for the distribution of the different grain crops over the country.

Rice accounts for the largest share of all grain crops in the southern part of China, where high temperatures and high precipitation are found. However we can also see some rice far up north in the northeastern plain and in the Beijing-Tianjin area. This rice does not follow the general trend of temperature. In these areas special varieties are grown that can tolerate the cool climate, although yields are low. Reasons for still growing rice in these areas, which are more productive for other grains, are mainly historic and because of consumer preference for the high quality varieties from this area. Because of the large demand for these "good-tasting" varieties, prices are high and therefore the cultivation is profitable in spite of the low yields.

Rice distribution is negatively related to the soil suitability for rainfed agriculture, which obviously results from the poor drainage conditions needed for rice cultivation but which are unfavorable for other crops. Rice cropping is also more labor-intensive than the cultivation of other grains, as rice needs to be transplanted, which is commonly done by hand. This is reflected by the positive correlation with the percentage of the population that

Table 3 Pearson correlation coefficients between the relative distribution of different grain crops and a selection of determining factors for all grid cells containing more than 5% cultivated land (n=4643)

	% of Rice in all grain crops	% of Wheat in all grain crops	% of Potato in all grain crops	% of Corn in all grain crops	% of Soybean in all grain crops	% of Other grains in all grain crops
PRC_TOT	0.86[a]	-0.56[a]	0.25[a]	-0.42[a]	-0.26[a]	-0.43[a]
TMP_AVG	0.70[a]	-0.30[a]	0.23[a]	-0.35[a]	-0.52[a]	-0.39[a]
FERT1	0.60[a]	-0.43[a]	0.16[a]	-0.25[a]	-0.29[a]	-0.24[a]
S1MAIZER	-0.47[a]	0.32[a]	-0.11[a]	0.29[a]	0.20[a]	0.10[a]
MEANELEV	-0.36[a]	0.25[a]	0.18[a]	0.01	-0.22[a]	0.42[a]
PALPER91	0.22[a]	-0.12[a]	0.34[a]	-0.18[a]	-0.44[a]	-0.03
ILLIT	-0.14[a]	0.08[a]	0.29[a]	-0.07[a]	-0.26[a]	0.28[a]
INCOM91	0.06[a]	0.00	-0.22[a]	0.07[a]	0.08[a]	-0.15[a]

[a] significant at P<0.001

is part of the agricultural labor force.

The general pattern of potato cultivation corresponds to similar climatic conditions as found for rice. Interpretation is difficult as potato statistics consist of both sweet potato (*Ipomea batatas* (L.) Lamk) and Irish potato (*Solanum tuberosum* L.). These two types of potato have very different environmental requirements. Sweet potato grows well at high temperatures, explaining the occurrence in the southern part of China. Irish potato gives relatively low yields at high temperatures, as these are conducive to foliar development and retard tuberization. Even though the conditions are not favorable for growing potatoes (due to the low average solar radiation, high temperature at lower altitudes and high concentrated rainfall), Irish potato is an important crop in the Sichuan basin. However, in the part of the Sichuan basin where potato is grown, other staple crops will suffer even more from the constraints (Wei, 1997). The preference for higher altitudes can be seen from the positive correlation with mean elevation, which distinguishes potatoes from the conditions favorable for rice cultivation. Within the southwestern region, of which the Sichuan basin is part, the relation with temperature (TMP_AVG: -0.31) is opposite to the general trend observed for the whole country, indicating that although potatoes are found in the warmer part of the country, they are located in the coolest parts of these regions. The negative correlation with income reflects the poor conditions under which the potatoes are grown by small farmers which usually use low inputs (Wei, 1997).

Positive correlations with illiteracy and a negative correlation with income suggest the backward nature of the areas where other grain crops (*e.g.* millet and sorghum) are cultivated. This cultivation is mainly for subsistence farming.

Discussion

1 Methodology

The methodology used in this study enabled to characterize agro-ecosystems in relation to their spatial distribution. Many relations found were straightforward and corresponded to our understanding of land use patterns and crop ecology. However, the systematic analysis allowed us to define the relative importance of the different determining factors and quantify the regional differences. Although it is well understood that correlations cannot be substituted for the mechanistic understanding of relationships, correlations can play an invaluable role in suggesting mechanisms for further investigations.

In general the explaining power of the correlations and regression models is high, showing that we are able to capture the most important factors determining the land use distribution. However, we might still lack some factors that are important for the distribution of land use, *e.g.* tenure, ethnicity and tradition. However, this type of quantitative information is not available, and can therefore not be included into the analysis.

The validity of the results also depends on the temporal stability of the relations found. Temporal stability was analyzed by doing the same analysis with data for 1986. The spatial structure found was very similar to the spatial structure of 1991. This can be attributed to the small changes in land use during this short period. Unfortunately land use data for years further apart are not available. Studies in Costa Rica (Veldkamp and Fresco, 1997) and Japan (Hoshino, 1996) have shown that the land use structure in these areas is very stable, in spite of the long time span between the years of analysis and considerable changes in land use during these periods.

2 Scale issues

The results of our analysis show that the relations between land use and determining factors are influenced by the grain size and extent of analysis. The influence of the grain size on analysis can be attributed to:

- The reduction of spatial variability: coarse grain sizes obscure variability while fine grain sizes obscure the general trends. Shifts in grain size may produce more than averages or constants: they may make homogeneity out of heterogeneity and vice versa.
- Emergent properties: changes in grain size are frequently associated with new or emergent properties. In complex, constitutive hierarchies, characteristics of larger units are not simple combinations of attributes of smaller units.
- The influence some factors can have over a considerable distance: with coarse grains these factors fall within the same unit of analysis and cause therefore a change in correlation structure.
- Stronger overlap among variables: aggregation reduces intra-class variance and the size of the sample population, smoothing the distributions and reducing the number of outer values identified within each class. This can create strong overlap among variables, markedly reducing the potential value of such variables for distinguishing classes.

The influence of the extent of analysis can be explained by the decreasing importance of local situations with an increasing extent of analysis. Our analysis for the different regions has shown that relations found at the national level are not always valid at the regional level. A smaller extent also allows the introduction of specific variables that are important for the area under analysis. Therefore, a smaller extent offers better insight into the specific situation of the region while a larger extent allows the identification of the general patterns.

These facts illustrate that decisions about the grain size and extent of the study area can greatly influence the strength and nature of the relationships observed. Investigators must be aware of the scale-related limitations of any study. The inclusion of multiple scales of observation through involvement of several grain sizes and varying extent can be critical for the understanding of land use patterns and processes.

3 Implications for land use change

The results of this study provide insights into the actual distribution of land use which results from a long period of land use changes. However, apart from giving insight, the results also have a direct relevance for land use change modeling. With the derived multiple regression models it is possible to calculate for every combination of determining factors the land use distribution that is usually found under those conditions (*i.e.* the cover calculated by the regression equations). When we want to determine what might happen to land use when a determining factor changes (*e.g.* due to urbanization), we can use the multiple regressions to calculate the land use situation normally found under the new conditions. These figures indicate in what direction the land use distribution might change. Also for the case when none of the determining factors changes, we can determine which areas will be most susceptible to land use change if the demand for a land use type changes. When it is assumed that the most important determining factors are included, the difference between actual land use and land use as calculated by the multiple regression equations has a meaning. If the actual cover with a land use type in a certain grid-cell is lower than the cover that is generally found under similar conditions, there is a potential for an increase in the cover of the considered land use type. The opposite holds true for grid-cells where the actual cover of a land use type is higher than the generally found cover. Fig. 5 illustrates this aspect for cultivated land. The dark areas in Fig. 5A indicate where actual cover with cultivated land is higher than the cover calculated by regression equations. If the demand for other competing land use types increases in these areas, it can be expected that the cover with cultivated land will decrease. Fig. 5B shows the areas where an increase in cultivated land can be expected. Based upon these principles, a dynamic simulation model was constructed to explore changes in the spatial pattern of land use in the near future (Veldkamp and Fresco, 1996). This type of model provides an essential link between small-scale mechanistic studies of land use change that offer detailed insights into specific cases that cannot necessarily be generalized (Turner II *et al.*, 1993) and integrated assessment models at the global level that often use very simplified relations between land use and its determinants, so that important driving factors are overlooked.

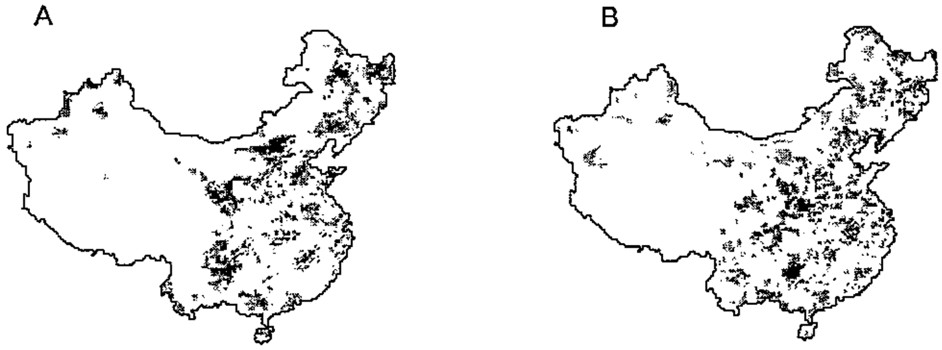

Fig. 5 Difference between actual cultivated land cover and land cover calculated with the region-specific multiple regression models (dark colors indicate large differences)
A: actual cover is larger than calculated cover; B: actual cover is smaller than calculated cover

Conclusions

From this analysis, the conclusions are as follows:

- Empirical analysis of land use distribution provides a useful methodology for an integrated characterization of agro-ecosystems on regional scales and the selection of factors that determine the land use pattern.
- The scale of analysis influences the correlation structure of the land use distribution. Thus, no description of the variability of land use makes sense without reference to the particular range of scales examined. Relations obtained on a certain scale of analysis may therefore not be directly applicable to other scales or other areas. Different research teams for land use change are, therefore, likely to reach different conclusions with respect to the determining factors of land use because of the differences in scale and extent of analysis.
- Interdisciplinary analysis is essential to understand the land use structure of China.
- The systematic, quantitative way of analysis is directly related to the modeling of land use changes in the near future.

Acknowledgments

This research was supported both by the Dutch National Programme on Global Air Pollution and Climate Change and by the project "Adaptation and Regulation Strategy of Sustaibable Development in NSTEC(North-south Transect of Eastern China)" (No.3989370) from the Natural Science Foundation of China.

References

1) Alonso, W. (1964): Location and Land Use. Toward a general theory of land rent. Harvard University Press, Cambridge, USA.
2) CAS [Chinese Academy of Sciences] (1978/1996): Soil Map of China. Edited by the Nanjing Institute of Soil Science, Chinese Academy of Sciences; published (1978) by the Map publishing house of P.R. China, Beijing. Digital version (1996) by the State Key Laboratory of Resources and Environment Information System (LREIS), Chinese Academy of Sciences, Beijing, China.
3) Crook, F. W. (1993b): China's grain production economy: a review by regions. In China, International Agriculture and Trade Reports. Situation and Outlook Series RS-93-4. United States Department of Agriculture, Economic Research Service, Washington, USA.
4) Fischer, G., Chen, Y. F. and Sun, L. X. (1998): The balance of cultivated land in China during 1988-1995. Interim report IR-98-047. International Institute for Applied Systems Analysis, Laxenburg, Austria.
5) Fresco, L. O. (1995): Agro-ecological knowledge at different scales. In Bouma, J., Kuyvenhoven, A., Bouman, B. A. M., Luyten, J. C. and H. G. Zandstra, editors. Eco-regional approaches for sustainable land use and food production. Proceedings of a symposium on eco-regional approaches in agricultural research, 12-16 December 1994, ISNAR, The Hague. Kluwer Academic Publishers, Dordrecht, The Netherlands.
6) Fresco, L. O. and Kroonenberg, S.B. (1992): Time and spatial scales in ecological sustainability. Land Use Policy 9, 155-168.
7) Gibson, C., Ostrom, E. and Ahn, T-K. (1998): Scaling Issues in the Social Sciences. IHDP Working Paper no 1. International Human Dimensions Programme on Global Environmental Change, Bonn, Germany.
8) Holling, C. S. (1992): Cross-scale morphology, geometry, and dynamics of ecosystems. Ecological Monographs, 62 (4), 447-502.
9) Hoshino, S. (1996): Statistical analysis of land-use change and driving forces in the Kansai District, Japan. Working paper WP-96-120. International Institute for Applied Systems Analysis, Austria.
10) Huang, J. and Rozelle, S. (1995): Environmental Stress and Grain Yields in China. American Journal of Agricultural Economics, 77, 853-864.
11) INRRP/CLUE [Institute for Natural Resources and Regional Planning and CLUE-group Wageningen] (1998): Statistical database of China for land cover, agriculture and population, collected by the Institute for Natural Resources and Regional Planning of the Chinese Academy of Agricultural Sciences and edited by You Qi Chen, P. H. Verburg and A. R. Bergsma. Chinese Academy of Agricultural Sciences and Wageningen Agricultural University, Wageningen, The Netherlands.
12) Levin, S. A. (1993): Concepts of Scale at the local level. In Ehleringer, J.R. and C.B. Field, editors. Scaling Physiological Processes. Leaf to Globe. Academic Press. San Diego, USA.
13) Meisel, J. E. and Turner, M. G. (1998): Scale detection in real and artificial landscapes using semivariance analysis. Landscape Ecology, 13, 347-362.
14) Ojima, D. S., Galvin, K. A. and Turner II, B. L. (1994): The global impact of land-use change. BioScience, 44(5), 300-304.
15) Rastetter, E. B., King, A. W., Cosby, B. J., Hornberger, G. M., O'Neill, R. V. and Hobbie, J. E. (1992): Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystem. Ecological applications, 2(1), 55-70.
16) Turner II, B. L., Meyer, W. B. and Skole, D. L. (1994): Global land-use/land-cover change: towards an integrated study. AMBIO, 23(1), 91-95.
17) Veldkamp, A. and Fresco, L. O. (1997): Reconstructing land use drivers and their spatial scale dependence

for Costa Rica (1973 and 1984). Agricultural systems, 55(1), 19-43.

18) Verburg, P. H., De Koning, G. H. J., Kok, K, Veldkamp, A. and Bouma, J. (1999): A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. Ecological Modelling, In press.

19) Vitousek, P. M., Mooney, H. A., Lubchenco, J. and Melillo, J. M. (1997): Human domination of Earth's Ecosystems. Science, 277, 494-499.

20) Wei, H. (1997): Agronomic and ecological studies on the potato (*Solanum tuberosum* L.) in southwest China. Seed and crop management. Thesis Wageningen Agricultural University. Wageningen, The Netherlands.

Appendix Description of land use types and candidate determining factors

Variable name	Description	Units	Source
Land use types			
CULT91	Cultivated lands	%	INRRP/CLUE, 1998
ORCH91	Horticultural lands	%	
FOREST91	Forestry lands	%	
GRASS91	Grasslands	%	
URBAN91	Lands for settlement and industry (built-up land)	%	
UNUSED91	Unused lands: deserts, glaciers, saline lands, etc.	%	
Demography			
PTOT91	Total population density	persons/km^2	INRRP/CLUE, 1998
PAG91	Agricultural population density	persons/km^2	
PURB91	Urban / non-agricultural population density	persons/km^2	
PRURLF91	Rural labor force density	persons/km^2	
PAGLF91	Agricultural labor force density	persons/km^2	
PAGPER91	% of Population belonging to agric. population	%	
PRLPER91	% of Population belonging to rural labor force	%	
PALPER91	% of Population belonging to agric. labor force	%	
PAGRUR91	% of Rural labor force belonging to agric. labor force	%	
Socio-economics			
ILLIT	Fraction of population that is illiterate (1990)	-	Skinner *et al.*, 1997
INCOM91	Net income per capita	RMB/person	INRRP/CLUE, 1998
DISTCITY	Average distance to city	Km	Tobler *et al.*, 1995
Soil-related variables			
GOODDRAI	Well-Drained soils	-	FAO, 1995
MODDRAIN	Moderately drained soils	-	
BADDRAIN	Badly drained soils	-	
SHALLOW	Shallow soils	-	
DEEP	Deep soils	-	
S1IRRPAD	Soils very suitable for irrigated rice	-	
S2IRRPAD	Soils moderately suitable for irrigated rice	-	
NSIRRPAD	Soils not suitable for irrigated rice	-	
S1MAIZER	Soils very suitable for rainfed maize	-	
S2MAIZER	Soils moderately suitable for rainfed maize	-	
NSMAIZER	Soils not suitable for rainfed maize	-	
SMAXHIGH	Soils that have high moisture storage capacity	-	
SMAXLOW	Soils that have low moisture storage capacity	-	
FERT1	Low soil fertility	-	CAS, 1978/1996
FERT2	Moderate soil fertility	-	
FERT3	High soil fertility	-	
TEXT1	Coarse soil texture	-	
TEXT2	Medium soil texture	-	
TEXT3	Fine soil texture	-	
Geomorphology			
MEANELEV	Mean elevation	m.a.s.l.	USGS, 1996
RANGEELE	Range in elevation	m	
SLOPE	Slope	degrees	
PHYSL	(fraction of the area with) level land	-	FAO, 1994
PHYSS	(fraction of the area with) sloping land	-	
PHYST	(fraction of the area with) steep sloping land	-	
PHYSC	(fraction of the area with) complex valley land forms	-	
GEOMOR1	(fraction of the area with) mountains	-	CAS, 1994/1996
GEOMOR2	(fraction of the area with) loess	-	
GEOMOR3	(fraction of the area with) eolian land forms	-	
GEOMOR4	(fraction of the area with) tableland	-	
GEOMOR5	(fraction of the area with) plain land	-	
Climate			
TMP_MIN	Temperature in coldest month	℃	Experimental dataset by W. Cramer
TMP_MAX	Temperature in warmest month	℃	
TMP_AVG	Average temperature	℃	
TMP_RNG	Difference between warmest and coldest month	℃	
TMP_10C	Number of months with temperature above 10 ℃	months	
PRC_TOT	Total yearly precipitation	mm	
PRC_RNG	Difference between wettest and driest month	mm	
PRC_50M	Number of months with precipitation above 50 mm	months	
SUN_TOT	Average percentage of sunshine	%	