

# **Modeling land use changes using Cellular Automata. An example from the North China Plain**

Bode, Frank<sup>1)\*</sup>; Doluschitz, Reiner<sup>1)\*</sup>

\*) – International Research Training Group “Modeling Material Flows and Production Systems for Sustainable Resource Use in Intensified Crop Production in the North China Plain”, University of Hohenheim 769, Stuttgart, Germany; <http://rtgchina.uni-hohenheim.de/>

1) - Computer Applications and Business Management in Agriculture, University of Hohenheim (410C), 70593 Stuttgart, Germany

## **Abstract**

Within the context of the multidisciplinary, collaborative International Research Training Group project “Modeling Material Flows and Production Systems for Sustainable Resource Use in Intensified Crop Production in the North China Plain,” land use modeling is one task. Therefore it is necessary to get an idea of the regional extent of specific land use systems, their development and possible changes in the future.

While typical GIS approaches in land use modelling is widespread in the meanwhile, another concept of regional representation is tested. Describing the land use with Cellular Automata offers the opportunity to recognise the dependencies between neighbouring spatial entities more accurate. Cellular Automata are one approach to model changes in land use. In contrast to a raster GIS Cellular Automata are able to simulate land cover changes in a more dynamic way. Each cell can interact with neighbouring cells to calculate next stages in cell status. One chance in working with Cellular Automata is to include analysis made in the different subprojects within the whole framework of the Research Training Group to create transition rules for the cell stages. The possibility to create cells in hierarchical order opens up the opportunity to regionalise data. One base -besides others- for the regionalization is data gained from field experimental sites in the North China Plain.

## **Key words:**

Agriculture, Cellular Automata, Environment, GIS, Land use changes, North China Plain

## **1. Introduction**

A growing population and a rising standard of living necessitate an increasing food production in the Peoples Republic of China. As arable land availability is limited and decreasing with proceeding urbanisation food production can only be increased by augmenting productivity of the land (higher production intensities, fertilizer-input, irrigation, improved crop rotations and higher yields). As it is also true in other intensive agricultural systems, serious environmental problems (water availability and pollution, air pollution, soil contamination and erosion) are raising. On the other hand, the North China Plain is one of the most important regions of cereal crop production in China. Wheat and maize rotations are common. In addition, vegetables play an important role in the daily Chinese diet. Vegetables are grown both in the field and in greenhouses, and there is a long tradition in greenhouse vegetable production in China. Due to the increasing demand of vegetables also during the off-season the percentage of greenhouse vegetable production has increased significantly, especially in the vicinity of big cities, and this trend is still going on.

The collaborative project “International Research Training Group (IRTG): Modeling Material Flows and Production Systems for Sustainable Resource Use in Intensified Crop Production in the North China Plain”, by the University of Hohenheim, Germany and the China Agricultural University, Beijing, China (<http://rtgchina.uni-hohenheim.de>), tend to develop cropping systems and management practices for the North China Plain, that guaranty high intensities and are, at the same time, environmentally, economically and socially sound. Main hypotheses of the overall project are

- (1) Adjustments in cropping systems (wheat/maize, vegetables) and management practices provide potential for sustainable resource protection (soil, water, air, nutrients) on a high yield level.
- (2) Adjusted cropping systems and management practices meet farm level (economic and organisational) and sectoral (demand) requirements, provide regional benefits and can be linked to the sectoral (political, institutional) frameworks.
- (3) Modeling approaches at all research levels are available and suitable to the conditions in the North China Plain to transfer spot results to a broader surrounding (regional, sectoral).

## **2. General remarks**

### **2.1 Project presentation**

Within the framework of the IRTG, the subproject with the duty of regionalization aims to develop a regionalized approach to understanding the impact and expansion of specific land uses. It pursues the primary goal of studying the regional consequences of locally investigated agricultural purposes. The central tool is the development and maintenance of a GIS based integrated Agro-Environmental Information System (AEIS). To evaluate the changes in land use in the North China Plain remote sensing data were analyzed. Therefore a repeated investigation and classification of land use types using satellite images is the central part of the work. The generation of a meaningful Land-Use Information System was the primary focus of the work at the University of Hohenheim during the first year of research in 2004/2005.

The basic structure and concept of the Agro-Environmental Information System was previously developed by Bareth et al. (2002). In the first phase of the IRTG preparatory work was done. Subsequently an Agro-Environmental Information System was established. This Information System includes different sectoral Information Systems. Components of these systems which are backed by a Geographical Information System (GIS) are:

- Base Geo Data Information System (BGDIS)
- Soil Information System (SIS)
- Climate Information System (CIS)
- Land Use Information System (LUIS)
- Agricultural Management Information System (AMIS)

In addition, geographical base information was gathered to put together the necessary information and to have a common geographical basis. The geo database was established within the software package ArcGIS. Based on the Geographical Information System a regional process-based agro-ecological model was developed and applied (Liu et al. 2006, Liu

et al. 2007). The results show regionally differentiated changes in the content of Soil Organic Carbon depending on land use systems and farm management.

The content of the information system is made up of land use information gathered from satellite images on the one hand, and maps and other geographical data sources as well as data from empirical surveys (e.g. concerning management practices) on the other hand. Furthermore, official statistics, like the Agricultural Yearbooks, are used to complete the information base. The results of other subprojects within the IRTG were made accessible by including them in the central meta-database of the entire research project. From there on, the current work is to develop and later to implement a modeling concept to regionalize the data to get to well based conclusions about the environmental and agro-structural consequences of recent land use systems changes and intensity shifts in the North China Plain.

The fact that agricultural production systems and material flows in the North China Plain are investigated within this project by several subprojects with different approaches in spatial and temporal designs and the demand for dynamic future projections necessitates the development of a further regionalization approach. The concept of Cellular Automata will be used to attempt this. This methodology allows the creating of spatial entities which are able to interact with their relative neighborhood to determine the next step in the temporal development of each cell (Wagner 1997).

## **2.2 Research area and production systems**

The North China Plain spans across the Provinces Hebei, Henan and Shandong in the coastal lowlands of Northern China (Fig. 3.1). Additionally parts of Beijing, Tianjin, Anhui and Jiangsu Provinces are included. The area is extremely important for Chinese agricultural production. Wheat and maize rotations are common. In addition, vegetables play an important role in the daily Chinese diet. The agriculture of the region is dominated by a typical production system which tends to have two harvests in one year. In autumn a type of summer maize is harvested, after this winter wheat is sown. After wheat harvest in late spring the maize is planted. This is a fairly stable system but requires high inputs of fertilizer and water.

With 320.000 km<sup>2</sup> and a population of 200 million people the region contributes an amount of 50% of Chinese wheat and about one third of Chinese maize production, respectively (Kendy et al. 2003).

## **2.3 Research approaches**

In this paper two different but highly interdependent approaches and respective preliminary results will be presented:

- (1) GIS based regionalization strategies based on the AEIS:  
Using a Geographical Information System and foregoing analysis of remote sensing and other spatial information a regionalization of agro-environmental parameters is applied.
- (2) Dynamic future projections and scenario based simulations using Cellular Automata:  
A more complex approach is chosen to display a more dynamic type of interaction in land use systems.

### **3. Regionalization strategies using GIS**

#### **3.1 The GIS database**

The integrated Agro-Environmental Information System (AEIS) is available and provides possibilities to cope with the data collected within the entire project, and therefore it provides the basis for the ongoing strategy of the regionalization of substantial values of agricultural production. Additionally, data from third-party sources, such as the statistical data from the Research Center for Rural Economy (RCRE), which provides panel data on households in the research area from the last decades, will be used to complete the information base.

One of the hypotheses of this project is that the matching of existing data gathered within and outside this project can create new datasets with new data at plot, farm and different regional levels. Therefore, a variety of possible sources of information are surveyed and analyzed for their ability to provide valuable information for the project. The objectives are to develop, use and improve methods for a regional representation and quantification of key agro-environmental values.

A new concept of regionalization will be applied by using detailed statistical data for identifying and locating specific land use types and tracking their spatial development over time, in order to have a comprehensive view on the changes in agricultural crop cultivation area.

The investigation of the rural agrarian structure in the North China Plain, which is one of the outcomes of the statistical data analysis, will be used to gain a basis for scenarios for the future development of rural areas. Especially in competition to the growing urban areas in the North East of China, it is necessary to implement successful strategies to ensure the ability of rural people to participate with economic and social development in China (Lichtenberg and Ding 2008).

The main aim of the development of an Agro-Environmental Information System (AEIS) and its application in the modeling of sustainable resource use in the North China Plain is implemented as follows. The AEIS can be seen as a powerful tool to regionalize the results and models developed by other subprojects within the framework of the IRTG. In detail, the Agro-Environmental Information System in the North China Plain consists of the following parts:

##### **3.1.1 Base Geo Data Information System (BGDIS)**

The BGDIS is the central component of the AEIS, and is the basis for the integration of all spatial data and the spatial / regional modeling. Data contained was gained from information collected within the framework of a project that preceded the IRTG and was financially founded by the German Federal Ministry of Education and Research (BMBF). It consists of geographical key reference points related to administrative boundaries, infrastructure and topographical data, as the basis of a digital elevation model (DEM). Currently, data of Chinese county boundaries has been included in the central AEIS database at a scale of 1:50.000, as have digital elevation model data (DEM) gained from www-data at a resolution of one km<sup>2</sup>. The Universal Transverse Mercator coordinate system (UTM) is used as the grid system (UTM zone 50), based on the World Geodetic System 1984 (WGS) datum. As this projection is used for the BGDIS, all geographic data in the AEIS is likewise based on the UTM system.

A number of further datasets are available from the BMBF- project. These include railways and major roads and other infrastructural data. Additionally, available are information of lakes, rivers and topographical data, all at a high resolution. But there are difficulties in using these datasets because this data are dispersed over more than 30 separate datasets. Different origins and data histories cause numerous topological or digitizing errors.

### 3.1.2 Soil Information System (SIS)

The second major component of the AEIS is the soil information system. This system is primarily based on data from the national soil survey, which was carried out in the mid-1980s, but also on data gained within the IRTG and its predecessors, e.g. the BMBF-project located at the Dongbeiwang site near Beijing, which is also used in the current project. Data included are therefore both large scale and small scale. Large scale soil maps and soil attributes describing its chemical (e.g. pH, nutrient status) and physical (texture) properties allowing for a general characterization of the soil types within the North China Plain. Small scale, site-specific data include data regarding soil water and nutrient (N) contents and flows, at very high spatial and temporal resolutions, mainly collected at the three field experiment stations Dongbeiwang (DBW), Quzhou (QZ) and Wuqiao (WQ).

Data currently contained in the SIS is restricted to the area of the Hebei province, only. The map was digitized at a scale of 1:500,000. For some limited parts of the province, data is also available at 1:250,000. Unfortunately no larger scale of mapped information is available. This information source is assisted by county wide plot information (Liu et al. 2006)

### 3.1.3 Climate Information System (CIS)

The climate database consists of datasets from eleven meteorological stations distributed throughout the North China Plain, comprising the three field stations run by the China Agricultural University in the North China Plain investigated by most IRTG subprojects. Meteorological data is collected on a daily basis as this has proven to be necessary for the modeling approaches used by other subprojects. Although gained from single locations, the data will also be used for regionalization. For regionalization to be more meaningful, the combination of such single-location data with climatic data available for download from the “China Meteorological Administration” may prove to be suitable in the future. Other than the BGDIS and the SIS, the CIS consists primarily of historical data, and updates of the datasets will be necessary on a regular basis. So far, the data covers a period of six years.

### 3.1.4 Land-Use Information System (LUIS)

The LUIS also comprises historical datasets, as well as current information. It depicts current and historical land use patterns in the North China Plain, which will be based on a combination of (a) statistical data e.g. from statistical yearbooks, (b) questionnaires, which were performed in the past, but are also being undertaken by subprojects within the framework of the IRTG, and (c) remote sensing techniques. The development of the LUIS is therefore the most laborious part of setting up the AEIS, despite the large amount of time also spent on digitalizing data for the SIS. The development of techniques to detect different crops in satellite images is a major focus.

The general idea of the initial approach is the identification of the potential winter wheat/summer maize cropping area, as this crop rotation system is the main focus of the IRTG. The respective area is characterized by rather dense vegetation cover of young winter

wheat plants in spring, while most other fields and orchards are only sparsely vegetated. In autumn, the fields with winter wheat / summer maize rotation will again be densely vegetated. But during this time of the year, the same holds true for other crops, orchards etc. Nevertheless, a detection of areas being densely vegetated both in spring and autumn will be a first clue about the extent of the area potentially cultivated by winter wheat and summer maize. Once this potential area is identified, it will be used in the future for more detailed analysis, in order to exclude areas that are also densely vegetated in both spring and autumn but are not winter wheat / summer maize rotations. These other combinations could be winter wheat with other crops, or multi-cropping of vegetables.

For the identification of areas with dense vegetation cover, both the Normalized Difference Vegetation Index (NDVI) and a principle component analysis were carried out. The NDVI is based on the fact that green leaves reflect a great deal of the incoming radiation in the near-infrared (NIR) spectrum (0.7 – 1.1  $\mu\text{m}$ ), but only a small portion of the incoming red (R) radiation ( $\sim 0.65 \mu\text{m}$ ). This is a rather unique characteristic and the relationship between the reflected R and NIR radiation values and therefore enables the identification of vegetated areas. On the other hand, most satellite images contain spectral bands covering both the R and NIR parts of the spectrum. For Landsat ETM+ images, band three covers radiation of 0.63-0.69  $\mu\text{m}$ , whereas band four shows reflected radiation of 0.75-0.90  $\mu\text{m}$ . Hence, these two bands were used to derive the NDVI for selected parts of the North China Plain situated around the field stations.

Once the areas covered in dense vegetation according to the NDVI images were identified, a classification was attempted using large square areas showing a high NDVI as training areas, because winter wheat fields are usually of square shape. The classification itself was then based on the results of a principle component analysis of seven bands.

This not only highlights the central position of this subproject within the IRTG-framework, but also the need for the integration of data gained from other subprojects, in the setup of this GIS. Furthermore, the GIS will form the basis of the integrative approach that couples a GIS-model. The coupling will be performed in order to depict the current situation of sustainable management within the North China Plain, but also to enable the development and evaluation of the sustainability of future scenarios of changes in land use and agricultural management practices. The AEIS itself is GIS-based and incorporates a subset of information systems belonging to a wide range of disciplines from physical geography and socio-economics.

### 3.1.5 Agricultural Management Information System (AMIS)

This information system is based mainly on information obtained from questionnaires, but to some extent also from official statistical sources. It includes databases that are only partially linked to geographical information, because information with a spatial dimension is not always available. Similar to the LUIS, the data included in the AMIS also covers both the past and current situation, which again leads to the necessity of regular updates of the data to reflect the current agricultural management practices in the different regions of the study area.

## 3.2 Results up to now

The results for the areas of the experiment stations of the China Agricultural University are presented below. The station Dongbeiwang is located at the northern border of Beijing City, The stations Wuqiao and Quzhou are situated in the more rural areas in the south of Hebei Province (Fig. 3.1).

### 3.2.1 Dongbeiwang site

The field station at Dongbeiwang is special, because it is located in the vicinity of Beijing, hence the heavy air pollution originating from China's capital becomes clearly visible on most satellite images at Dongbeiwang. Furthermore, Beijing is also affected by dust storms originating from Inner Mongolia, which regularly obscure the sky. The satellite images were taken on May 2003, when most of the city, Dongbeiwang village and the surrounding area were covered in clouds, which may be attributable to one of the two above-mentioned reasons. Nevertheless, it is hardly possible to calculate any NDVI based on this image. Generally, the variation of the land use pattern in this region remains extremely high, as urban and suburban areas are interlinked with numerous fields, and also with plantations, orchards and semi-natural vegetation in the neighboring mountains. Hence, any classification of this area is extremely difficult due to both the mostly very poor quality of satellite images and the great heterogeneity of land use patterns. Therefore, the evaluation of the satellite data for this region was postponed to later date.

### 3.2.2 Wuqiao site

The area around Wuqiao is characterized by its predominant use for agricultural purposes, which is interrupted by numerous small villages. In the images from May 2001 and 2003, the edges of the image are covered in dense vegetation, whereas the central and northern parts of the image are free of vegetation. A comparison of the classification results based on these two images shows that the overall, the densely vegetated area did not change much to a large extent. In autumn, nearly almost the entire area is covered in dense vegetation, with roads, villages and cities, as well as rivers, becoming detectable as non-vegetated areas. An overlay of the autumn image from 2000 and the image of the following spring helps to detect the potential winter wheat / summer maize growing area, which amounts to 20 – 30% of the area covered by the satellite images.

Fig. 3.1: Locations of the Experiment Stations Dongbeiwang, Wuqiao and Quzhou of the China Agricultural University in the North China Plain.

### 3.2.3 Quzhou site

NDVI images from Quzhou taken in May 2000 and May 2003 show a clear separation of this area into a western portion that is densely vegetated in the spring of both years and an eastern portion that exhibits a sparse vegetation cover, interrupted by some small "green" vegetation islands. Nevertheless, an overlay of the classification results of these two images shows a strong decrease of more than 20% in the densely vegetated area between 2000 and 2003. As a result, primarily the upper NW corner of the image remains as a potential area of winter wheat / summer maize cropping, together with the area surrounding the administrative centre of the county.

Similar to the conditions in Wuqiao, nearly the entire area around the Quzhou field station is covered in dense vegetation in autumn 2001. Only in some areas in the NE corner of the

image differences can be observed. Here larger areas appear to have a very sparse vegetation cover both in spring and autumn. An overlay of the spring and autumn images again highlights the probability of a strong shift in the cropping pattern in this region between 2000 and 2003. While 52 to 61% of the area was potentially covered in winter wheat / summer maize in 2000, in 2003 only 31 to 42.5% fell into this cropping category. Unfortunately, only one autumn image of 2001 could be taken as a reference, due to the persistent cloud cover in autumn in this region in most years.

#### 3.2.4 Modeling results

Despite the above mentioned data gaps still associated with some of the information systems that make up the AEIS, small-scale AEIS have been established, primarily by Chinese members of this research project. These two small AEIS cover the area of Huimin County in Shandong Province and the area of Quzhou County. Especially the first modeling trials were run using the DNDC model (Liu et al. 2006). Modeling the soil organic matter (SOM) content, Liu et al. (2006) managed to show that the SOM is currently very low, but bound to increase in the coming years if current management practices continue and if rates of fertilization are strongly reduced. Highest gains in SOM contents may be obtained by a larger amount of crop residues being returned to the fields, according to the scenarios included in the trial runs.

## 4. Scenario approach using Cellular Automata (CA)

### 4.1 The course of model development

The course of model development mainly includes two major steps, the

- (1) development of GIS-based modeling approaches and the
- (2) development of modeling scenarios.

#### (1) Development of GIS-based modeling approaches:

As a step following the set-up of the AEIS data gained from the field studies and first modeling approaches, developed and tested by other subprojects within the IRTG will be linked to the AEIS-databases. Existing models will be regionalized based on the AEIS database and optimized to match the status quo. The major goal of this exercise is the development of models that effectively simulate the status quo; i.e. the effect of current land-use patterns and management practices, in combination with prevailing climatic conditions and soils occurring in the NCP, on the agricultural environment and the sustainability of resource utilization. Models will especially have to be optimised with respect to water-use efficiency and environmental pollution resulting from over-fertilization.

#### (2) Development of modeling scenarios:

The further step is the development of future scenarios, taking into account changes in agricultural management and land-use patterns as well as economic and political changes, and investigating strategies for a reduction in the input of resources, while maintaining stable high production and reducing negative environmental effects. Furthermore, these scenarios will incorporate envisaged future plant breeding successes, with data again derived from subprojects within the IRTG framework. The major goal is to reduce the negative environmental impact of the farming systems, again mainly focusing on pollution and water-use efficiency, while maintaining high yields. Future scenarios will also be developed with policy analysis and results from statistics. This will show the influences of decisions made on

different policy levels on production success and the environmental factors at the farm, village and regional levels.

## **4.2 General concept of Cellular Automata**

The general concept of Cellular Automata is known since the late 1960's (Sarkar 2000). The application of the "Game of Life" is one of the simplest and best known examples. The early and very simple models developed to two- dimensional representations of space. In a perfect environment these models are running well and give easily to interpret results.

Each cell is established as a unit, which in dependency of the own state and the interaction with neighboring cells is able to determine next steps of a temporal development.

Cellular Automata consist of four main features: geometry, state, time and transition rules.

### **4.2.1 The geometry**

For each cell it is important to define a set of geometrical conditions. Every individual cell needs to have a spatial dimension. And this dimension has a determined form. It is possible for cells to have any possible irregular shape. But in many cases the shape is a rectangular form. So the base geometrical situation is similar to a grid based raster - GIS.

Additionally to the shape a boundary has to be defined. Strongly connected with this topic is the definition of the neighborhood relations. The neighborhood is considerably essential to establish the true advantages of a Cellular Automata in relation to a raster GIS. Characteristic for Cellular Automata is the ability to take actions and states of neighboring cells in account, when calculation the next step of cell development. The actions may differ between just "recognizing" the neighboring cell and interacting reciprocally. The initially set up geometry can not be changed during the simulation, it is not allowed to divide cells to smaller units.

### **4.2.2 The state**

Each of the cells representing a real area has a certain state at a certain time. The state is defined from a group of finite number of discrete possible spatial variables, for example a land use type. The cell is autonomic in deciding which state it will capture in the next time step. The calculation will be done with respect and dependency to its current state and the states of neighboring cells and defined neighborhood function.

### **4.2.3 The Time**

In the Cellular Automata concept there is a global time device for all cells. So, every cell is forced to calculate the next state at a synchronized time. The calculation based on transition rules.

### **4.2.4 The transition rules**

Transition rules are besides the geometry the core of the Automata. The transition rule defines what action is going on when time for calculating the next step comes. The states of each cell changes according to the defined rules. A cell takes input from its on state history as well as from its neighborhood. This is responsible for the dynamic behavior of Cellular Automata.

As an extension of typical Cellular Automata Sonnenschein and Vogel (2003) describe the possibility to vary the shape of cells to get a closer relation to real live application. In some applications the typical grid cell type may not be suitable for a correct description of the underlying processes. Additionally, they introduce a defined number of higher levels of cells, which may control cells of lower level in hierarchical order. All levels cover the same extent and the time step in every level is synchronic, but can differ between the levels. So a higher order of cells may act only every five time step of a lower one. The neighborhood functions also may be differed in a way that not only adjacent cells can influence each other. A software tool “EcoScape” developed by the University of Oldenburg is a realization of this concept (Sonnenschein and Vogel 2003).

### **4.3 Applications of Cellular Automata**

The history of Cellular Automata shows, that they were used in chemical science as well as in physics. Atmospheric pollution and dispersion processes were successfully modeled with Cellular Automata (Ganguly et al. 2003). Many studies tended to simulate more biological features. Cellular Automata are able to describe and display processes ruling biological invasive events as Hogeweg (1988) described. Furthermore, diseases dispersion can be simulated by models using Cellular Automata (Tischendorf et al. 1998). Here a rabies simulation model for Eastern Germany is presented. This type of spatial modeling is well known to be a suitable solution for land use change models, especially in more urbanized areas (O’Sullivan and Torrens 2001, Yeh and Li 2003). Furthermore, Cellular Automata have been previously applied to ecological topics. The use of this approach for agricultural purposes is not as widespread, but the advantages and limits of its use are outlined by Parker et al. (2003). Cellular Automata as a tool have some advantages because with it, it is possible to build simple models, which are nevertheless capable of representing fairly complex systems. The combination and further development of the approaches of land use system modeling and advanced Cellular Automata promises to have good results in land use system modeling. Especially because of their ability to represent spatial dimensions in contrast to approaches like multi agent systems, Cellular Automata can be used to for a regionalized agricultural modeling (Parker et al. 2003). The base information for modeling with Cellular Automata can be transferred from the GIS system.

Within this project, the Chinese counterparts have experience with some Cellular Automata software solutions, especially the development of the CLUE model (Verburg et al. 2004). The combination of data mining and Cellular Automata is a powerful tool for modeling complex spatial systems (Li and Yeh 2004).

The adoption of this concept for this specific agricultural question is based on the assumption emerged from previous results that there are some preliminary restrictions for farmers that determine their choice of crop. Such restrictions include the location, the time of sowing and harvest, and the possibility of changing management strategies. The changes of the spatial extent and the distribution of land use are two of the main targets of this subproject. Within a favorable area for decision making, a possible change in the land use system will be located. Successful farmers and villages will be an example for others and successful farming strategies will be adopted by closely related production communities.

### **4.4 The importance of transition rules**

Transition rules and neighboring rules are central elements of the Cellular Automata. On the one hand the general structure has to be build. Here the type and the quality of connection

between cells are formulated. The definition which cells influence each other which and how this influence is realized is essential for the results. The construction is mostly based on expert knowledge. The problematic issue concerning expert knowledge is often the way to gather it. Experiences from building knowledge bases for expert system approaches can be used and assist this difficult task. Communication with experts has to be direct and supposed to be subjective. There may be also some problems with transferability. For a general understanding the rules may be structured this way, but quantification is hardly possible like this.

An analysis of recent developments in different agricultural parameters shows the structures and spatial extent of agrarian changes (LUIS). These analyses can be used to parameterize the rules, which are used to define the transition in one cell state from one time period to the next. To create the rules through analysis is one important precondition to modeling the spatial consequences of different land-use practices. With these results, a model that considers the special dependencies and developments in land use change can be built with using the information of the AEIS.

The main point is to build up a data warehouse and to use different methods to extract interesting information out of it. The data warehouse contains all different kinds of heterogeneous data from multiple information sources. For further use the multivariate natured data are integrated and standardized. Afterwards the data warehouse is used for data mining and Online Analytical Processing (OLAP) tools are applied. The data mining technique that was applied by Abdullah et al. (2003) is called Recursive Noise Removal (RNR). This is an unsupervised clustering method that can run on any database table containing alphanumeric values. With this method the authors were able to find two clusters as a starting point to establish certain rules and to deduce threshold values for different events. Furthermore, using OLAP and statistical analysis operations Abdullah et al. (2003) could locate recurrent patterns in the data. This information will be used to quantify rules. There are different suitable methods to do this. There are a couple of statistical (empirical) procedures and mathematical equations. One difficulty is that statistical procedures strongly recommend a clear data structure, strong assumptions due to the distribution and the level of measurement. Additionally, in the forerun of an analysis the dependency must be clearly understood. However, most of the transition rules built like this are not unambiguous because they use mathematical equations instead of using explicit transition rules. With respect to data quality and availability methods of multivariate statistical analysis can be applied. Regression models seem to be one suitable means of analysis if accuracy and preciseness of the data are appropriate. This will be decided based on a detailed analysis of the data in question.

Data mining seems to be an appropriate solution and is used additionally to regression models to analyze the complex structure of the datasets gained in the IRTG. The statistical assumptions made by data mining, in contrast to typical statistical approaches, are lower, which means that data from different origins and in different resolutions and formats can be used (Duda et al. 2001). To create the transition rules for Cellular Automata due to the amount and heterogeneity of available data sources from within the project the method of data mining was chosen (Witten and Eibe 2005). Due to the vast amount of available data it is necessary to automate the procedure of rule definition as much as possible.

Therefore, a machine learning system such as data mining has strong advantages. In consequence an evaluation of suitable software tools to achieve this was done. Finally the "Java" stand-alone tool "Weka" will be eventually chosen for initial analysis (Witten and Eibe 2005). Besides the traditional analysis with statistical software packages, further open source

software was tested. Some problems must still be solved in the fields of data collection and clearing, data format and software applications, especially because of inconsistent import- and export formats, respectively.

An overview of the sequences of work steps to receive transition rules is shown in Fig. 4.1.

Fig. 4.1: Flowchart for exemplary reconstruction of the transition rules for Cellular Automata (from Li and Yeh 2004).

Within the data mining process there are a several procedures available. The advantage in using decision trees lies in their ability to represent the rules with easy to understand parameters, like Li and Yeh (2004) showed.

#### **4.5 Research questions aimed at**

A preliminary set of research questions which shall be analyzed in the medium range include:

- detection of changes in the spatial extent of certain land use and
- distribution of land use,
- investigation of expansion / adoption dynamics of successful management practices,
- spatially explicit modeling of environmental consequences of specific land use types.

#### **4.6 Project data sources, availability and outlook**

To enhance the base data, further remote sensing is necessary to discover recent development in land use change and to ensure the accuracy and the timeliness of the results. Some parts of the necessary image classification in GIS will be carried out in the future, as satellite images will be taken after the harvest in autumn. Additionally, the preparation and analysis of satellite images and the GIS- digitalization of analog maps will be important and time consuming. Due to the fact that some spatial information is not available or affordable in digital format, a digitalization of larger amounts of geo-information will be necessary. To get access to Chinese statistics and other Chinese-language data sources.

The working program of the project in the further phase includes the maintenance of the datasets of the Agro-Environmental Information System, as well as further steps for setting up transition rules for Cellular Automata gained from statistical data analysis will be done. The ongoing work will include the provision of new relevant spatial data, such as updated satellite images or more detailed maps of Wuqiao County and Quzhou County in southern Hebei province. The necessity of updated spatial information is obvious due to the topic of researching land use changes in the last years, in order to be able to make better suggestions for future scenarios. Maps will include detailed topographical information on a larger scale. Due to the availability of different land use data, one of the needs will be to verify and standardize the data to be incorporated into the information system. Parts of this information will be sampled in the other subprojects of the Research Training Group. Subprojects concerning plant production and cropping systems will provide information about local land use practices and the possible resulting changes in yield, while projects engaged in rural

development and farm management will deliver data about the resulting farm-income. Additionally, further resources like data from empirical and statistical surveys and remote sensing data will be made accessible. The new datasets will have to be evaluated and reorganized and searched for suitable new datasets which will be included in the existing system. Recent satellite images will be analyzed to carry on the recording of land use changes in selected counties in the research area of the North China Plain. Further topographical, and other geographical analog map information, will be gathered and digitized. In the near future these projects will also include a set of modeling approaches which can be used for regionalization. In addition and by using the available results of the GIS approach and the collected and analyzed data a combined modeling approach will be applied. In the long run, a set of decision support rules according to the results of modeling will be built up. The upscaling modeling and simulation approach within this subproject will be set up with an attempt to run Cellular Automata. Different Working Packages can be identified. The parameters are known as production relevant factors of natural production functions as well as parameters of socio-economic conditions. Some of these will be received by analyzing the data sources from within the subproject, while others will be gained from other subprojects. Preliminary to the modeling and simulation, several statistical analyses will have to be carried out in order to identify structures and rules for setting up a system and its possible interactions. Data mining techniques will be used to find structures in the data set that need not be linear like in regression analysis. The data used has a complex structure and will be delivered by different surveys and subprojects. The available datasets include household surveys gathered in recent years, as well as data from our Chinese partners preliminary to this project. Also, official statistical datasets like the panel data from the Research Center for Rural Economy (RCRE).

In a first step, an approach will build a set of rules to couple regionalized datasets. In a test phase, a simplified model with a less developed cell type will be created.

In a bottom-up approach, the concept of representation of space will be realized by building cells. Each cell can represent different spatial extents of the real world, for example one plot can represent one field or one farm. At different regional levels, different parameters must be taken into account. Each cell can be connected with each of its neighboring cells and can interact with one or more of them according to special rules.

This approach will be improved by setting up an advanced type of asymmetrical and hierarchical Automata. This will increase the complexity in regard to spatial, technical and questions of content. The results will be used to calibrate the further development of the regional model.

The necessary information for manipulating agricultural production is available at the local or regional levels. Especially the cooperation with subprojects dealing with the socioeconomic influences on farmers and those effects on land use will be a step forward to get an evident impression of conditions in North China agricultural economic situation. The conditions different management practices apply on or resulting from material flows will be carried out by subprojects working on plant and soil material flows. The results will give hints about which factors have the most impact and are known to be relevant for regionalization. A scenario-driven approach for modeling effects of different land use systems is a result that can be expected from applying cellular construction.

## **5. Conclusions**

The North China Plain is one of the most important agricultural production areas in China. Environmental problems and the need to keep up high yields lead to seek for alternative and sustainable agricultural productions systems and resource use. A Sino- German cooperation research project aims to do so, with a set of different approaches. To achieve a meaningful set of data, locally gathered data must be transferred to a higher scale level. Besides a regionalization approach using GIS methodology, Cellular Automata as an advanced method is presented. The concept and background of Cellular Automata and their ability to represent land use systems and their changes are shown. In respect to the agricultural production systems and the availability of data Cellular Automata are able to model these systems more dynamically. The concept is presented and the further steps within the project are outlined.

## **6. Acknowledgements**

This paper is based on the first Sino-German International Research Training Group “Modeling Material Flows and Production Systems for Sustainable Resource Use in Intensified Crop Production in the North China Plain” jointly supported by the German Research Council (DFG) and the Chinese Ministry of Education (MoE).

## **7. References**

- Abdullah, A., Brobst, S., Pervaiz, I., Umer, M. & Nisar, A. (2003). “Agri Data Mining/Warehousing: Innovative Tools for Analysis of Integrated Agricultural and Meteorological Data”, In proceedings of International Workshop on Frontiers of Information Technology, COMSATS (CIIT), Islamabad, Dec. 2003.
- Bareth, G., Yu, Z. & Doluschitz, R. (2002). Set up of a Regional Agricultural Environmental Information System (RAEIS) in China. Environmental Informatics 2002, Vienna.
- Duda, R. O., Hart, P. E. & Stork, D. G. (2001). Pattern Classification, Second Edition, 680 P.
- Editorial Committee of China's Agricultural Yearbooks. China's Agricultural Yearbooks (various issues). Beijing: China's Agricultural Press.

Ganguly, N., Sikdar, B. K., Deutsch, A., Canright, G. & Chaudhuri, P. P. (2003). A survey on cellular automata. Technical report, Centre for High Performance Computing, Dresden University of Technology, December 2003.

Hogeweg, P. (1988). Cellular Automata as paradigm for ecological modelling. *Applied mathematics and computation* 27, 81-100.

Kendy, E., Molden, D. J., Steenhuis, T. S., Liu C., & Wang J. (2003). Policies Drain the North China Plain - Agricultural Policy and Groundwater Depletion in Luancheng County, 1949-2000. International Water Management Institute, Colombo.

Li, X. & Yeh, A.G. (2003). Simulation of Development Alternatives using Neural Networks, Cellular Automata, and GIS for Urban planning. *Photogrammetric Engineering & Remote Sensing* 69/9, 1043-1052.

Li, X. & Yeh, A.G. (2004). Data mining of Cellular Automata's transition rules. *Int. J. Geographical Information Science* 18/ 8, 723-744.

Lichtenberg, E. & Ding, C. (2008). Assessing farmland protection policy in China. *Land Use Policy*, 25, 59–68

Liu, Y., Bode, F., Yu, Z. & Doluschitz, R. (2007, September). Land use simulation in the North China Plain under different scenarios - An integrated GIS approach. (Paper presented at the Conference on the Science and Education of Land Use: A transatlantic, multidisciplinary and comparative approach, Washington D.C.).

Liu, Y., Yu, Z., Chen, J., Zhang, F., Doluschitz, R., & Axmacher, J.C. (2006). Changes of Soil Organic Carbon in an intensively cultivated agricultural region: A Denitrification-Decomposition (DNDC) modelling approach. *Science of the total environment* 372/1, 203-214.

O'Sullivan, D. & Torrens, P. (2001). Cellular models of urban systems. In Bandini, S. & T. Worsch (Eds) 2001, *Theoretical and Practical Issues on Cellular Automata*, Proceedings of the Fourth International Conference on Cellular Automata for Research and Industry (ACRI 2000), 108–116.

Parker, D.C., Manson, S.M., Janssen, M.A., Hoffmann, M.J. & Deadman, P. (2003). Multi-Agent Systems for Simulation of Land-use and Land-cover Change: A Review. *Annals of the Association of American Geographers*, 93 (2), 2003, 314-337.

Sarkar, P. (2000). A Brief History of Cellular Automata, *ACM Computing Surveys*, Vol. 32 (1), 80-107.

Sonnenschein, M. & Vogel, U. (2003). Hierarchische asymmetrische Zellulare Automaten zur Modellierung ökologischer Systeme auf mehreren Skalen. 5. Workshop "Theorie und Modellierung von Ökosystemen", Kölpinsee 2001, 37–50.

Tischendorf, L., Thulke, H.-H., Staubach, C., Müller, M. S., Jeltsch, F., Goretzki, J., Selhorst, T., Müller, T., Schlüter, H. & Wissel, C. (1998). Chance and risk of controlling rabies in large-scale and long-term immunized fox populations. *Proceedings of the Royal Society - B*. Volume 265, Number 1399, 839-939.

Verburg, P.H., Veldkamp, A., Willemsen, L., Overmars, K.P. & Castella, J.-C. (2004).  
Landscape Level Analysis of the Spatial and Temporal Complexity of Land-Use Change. In:  
DeFries, R., G. Asner and R. Houghton (eds.): Ecosystems and Land Use Change, 153,  
Wageningen, 217-230.

Wagner, D. F. (1997). Cellular automata and geographic information systems. *Environment  
and Planning B*, 24, 219-234.

Witten, I. H. & Eibe, F. (2005): *Data Mining: practical machine learning tools and techniques*.  
2<sup>nd</sup> ed.

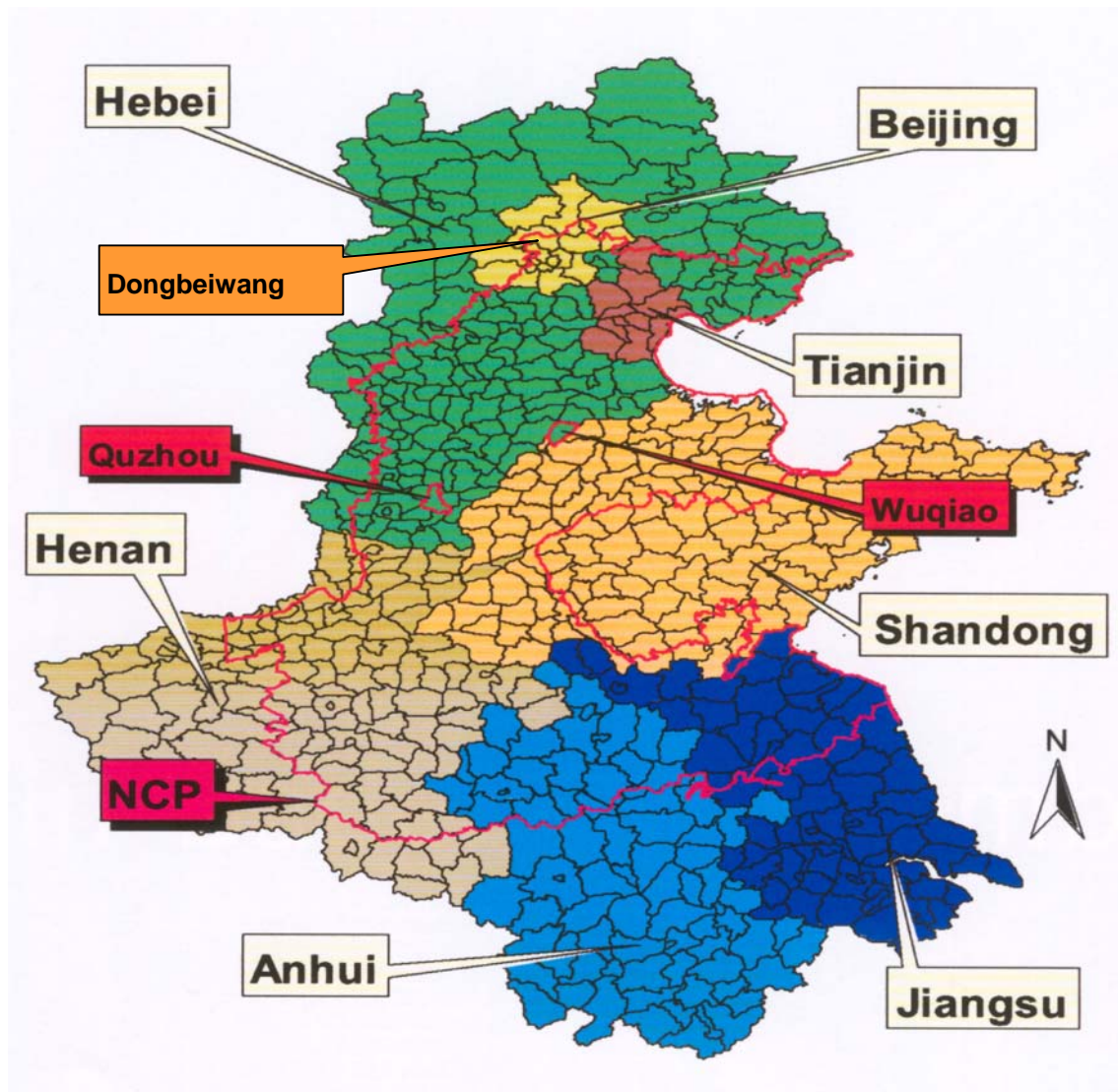


Fig. 3.1: Locations of the Experiment Stations Dongbeiwang, Wuqiao and Quzhou of the China Agricultural University in the North China Plain.

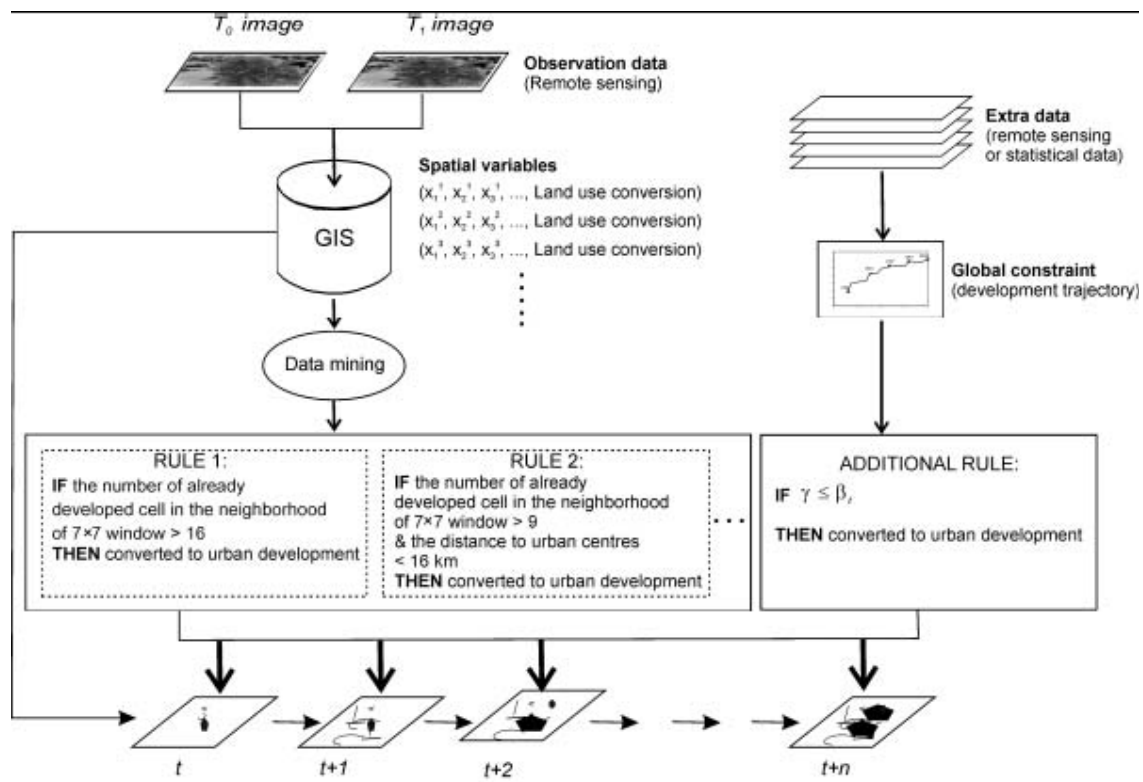


Fig. 4.1: Flowchart for exemplary reconstruction of the transition rules for Cellular Automata (from Li and Yeh 2004).