Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Economia e Statistica Agroalimentare

Ciclo XXIV

Settore Concorsuale di afferenza: 13/A2- POLITICA ECONOMICA

Settore Scientifico disciplinare: SECS-P/02 POLITICA ECONOMICA

Titolo Tesi

An Analysis of Living Conditions in Rural Villages in China. A New Geography of the Country.

Presentata da: Nica Claudia Calò

Coordinatore Dottorato

Prof. Roberto Fanfani

Relatore

Prof. Roberto Fanfani

Esame finale anno 2012

Acknowledgements

I would like to thank my supervisor, Prof. Roberto Fanfani, and the researcher Tommaso Pirotti for their useful comments and recommendations. Acknowledgments are due to all the other Professors for their suggestions and criticism. Many thanks to my dear friends and PhD colleagues for the kind support provided, I am glad to have met you.

Special thanks are owed to Sara, for having given me a place to stay and finish the thesis.

Finally I would like to say thanks to my father, my mother and my sister who have never let me get down and to Francesco who is always by my side with love, reminding me to trust in myself.

CONTENTS

LIST OF TABLES AND FIGURES

LIST OF ABBREVIATIONS

Compute MBF: Compute Membership Function (Fuzzification Method)

CoM: Center of Maximum (Defuzzification Method)

BSUM: Bounded Sum Fuzzy Operator for Result Aggregation

MIN: Fuzzy Operator for AND Aggregation

MAX: Fuzzy Operator for OR Aggregation

GAMMA: Compensatory Operator for Aggregation

PROD: Fuzzy Operator for Composition

LV: Linguistic Variable

MBF: Membership Function

RB: Rule Block

Introduction

The role played by Chinese rural areas is still fundamental for the general development of the country from a political, economic and social point of view. In particular, the way in which the rural areas have influenced the social stability of the whole country has been widely discussed due to their strict relationship with the urban areas where most people from the countryside emigrate searching for a job and a better life.

However, in recent years many studies have mostly focused on the urbanization phenomenon with little interest in the living conditions in rural areas and in the deep changes which have occurred in some, mainly agricultural provinces.

In 1978, one of the most important reforms in the history of China was introduced, well-known as the *Household Responsibility System* (or HRS), followed by many other reforms that significantly changed the structure of the country. One main consequence, of course, has been increasing differences, especially between the rural and urban areas, which are as strictly related to each other as they are distant in terms of growth rate and living conditions.

An analysis of the level of infrastructure is one of the main aspects which highlights the principal differences in terms of living conditions between rural and urban areas. Since the 1980s, several studies have been conducted to show the relationship between agricultural development and infrastructure development (Antle 1983). It is generally recognised that not very much attention has been paid to the role of infrastructure in the literature, due to a lack of reliable data on various infrastructure indicators. However, by using a traditional source accounting approach, the Agricultural Census and other official sources, it is possible to identify the specific role of rural infrastructure and other public capital in explaining productivity difference among regions. This throws new light on how to allocate limited public resources for both growth and regional equity purposes (Fan, Zhang 2004). This thesis will focus on these available detailed data on rural infrastructure from the Agricultural Census of 2006, in order to underline the main characteristics of Chinese rural areas in terms of living conditions, thus stressing regional disparities.

One fundamental consequence of these regional disparities is the difference in productivity within Chinese provinces. Therefore, in order to reduce the estimation bias some recent studies add new variables in the production function estimation, such as government investment in roads, electrification, education and other public investment in rural areas (Fan, Zhang 2004).

The main aim of this work is to describe and analyze the social and living conditions in Chinese rural villages in terms of infrastructure and service development, utilizing data calculated at the village unit, in order to show the main disparities which are still present among different provinces and municipalities in China.

In fact, the Chinese government has invested a lot in poverty alleviation, especially in the post-reform years, by adopting several anti-poverty policies and strategies that can be divided into 4 phases: 1) 1978-1985 when poverty alleviation was mainly due to general economic growth; 2) 1986-1993 with programs mostly targeted at poor areas and focused on the improvement of infrastructure, considered as fundamental for enhancing the capacity of poor areas and rural households; 3) 1994-2000 with the launch of the Poverty Alleviation Plan to eliminate absolute poverty through the tax favorite policy, financial support and a social-economic development program, together with support to labor migration; 4) 2001-2010 with the promotion of the Development Oriented Poverty Alleviation Program in Rural China whose aim was threefold, i.e. guarantee sustainable subsistence to poor people by increasing income and living conditions, focus on integrated development means to improve infrastructure, technology, education, public health care and cultural development and finally, support farmers' participation in poverty alleviation planning and implementation[1]. All these investments and programs have, of course, enhanced the level of development and reduced poverty in the poorest regions but it is still not enough for some provinces. The differences are not only between rural and urban zones but also between the richer metropolitan areas and coastal provinces of the East and the Central and Western provinces which are more peripheral and marginal.

In this thesis, I first carried out the analysis through the multivariate statistics approach in order to define the new map of rural areas, and then the results have been compared with an index elaborated through the Fuzzy Expert/Inference System.

The nature of the data has placed several limits on this analysis. First, they refer to only one year (2006) and this means that we simply get a snapshot of the country's situation. Moreover, it was not possible to get data at a more disaggregated level, namely at county level instead of provincial level.

[1] In October 2004, the Rural Survey Organization of the Chinese National Bureau of Statistics presented the paper "Poverty Statistics in China" at the "International Conference on Official Poverty Statistics: Methodology and Comparability", held in Manila, Philippines.
http://www.nscb.gov.ph/poverty/conference/papers/4_poverty statistics in china.pdf

1. Chinese Economic Development in Recent Decades and deep changes in rural areas[2]

Analysts, politicians and international players from all over the world look at China as one of the most powerful countries on the international scenario, and as a country whose economic development can significantly impact on the economies of the rest of the world. In fact, recent news about the lower growth rate of China has immediately resulted in negative reactions from the most important stock exchanges.

Undoubtedly, the changes in the last few decades have been enormous and this has attracted even more attention from other countries. There are, however, many aspects of these changes that are less well known. This is not only because of the sheer size of China - with a population of more than 1.3 billion – but also because of the lack of information on the enormously large and varied rural areas, where more than 50% of the Chinese population still live.

The rapid development of the Chinese economy in recent decades is the result of the combined effect of the reforms began in 1978. In particular, the contribution of decollectivization, price adjustments and institutional reforms has been deeply analysed, demonstrating how these reforms have positively impacted on agricultural growth (Lin 1992).

Hovever still some answers are missing about how much these reforms in agriculture influecend the recent changes occurred in the country in particular in terms of great reduction in hunger and malnutrition and then in terms of reduction of poverty, thus increasing the social stability of the whole country.

The opening of the Chinese market and the development of the Chinese economy has increasingly highlighted the country's importance as a both a producer and consumer of agricultural and food products, as well as its influence on international markets and on the negotiations taking place at present under the aegis of the WTO (Lohmar *et al.* 2009).

In fact today, the sum of imports and exports of China amounts to almost 70% of GDP, and in the period from when the reforms began in 1978 to today, China has reached the level of Germany and the USA in world merchandise trade in terms of exports and imports. In 2009, China overtook Germany as the lead exporter of merchandise while becoming the second largest importer after the USA (WTO 2010).

[2] The first paragraphs of the thesis are partially based on the results reported in a joint publication with my supervisor, where we have presented the results obtained from a common analysis of the data of the Chinese agricultural censuses.

Table 1 - Total Value of Imports and Exports in China – 100 ml USD (1978-2010)

Year	**Total Imports and Exports**	**Total Exports**	**Total Imports**	**Balance**
1978	206.4	97.5	108.9	-11.4
1980	381.4	181.2	200.2	-19
1985	696	273.5	422.5	-149
1990	1154.4	620.9	533.5	87.4
1995	2808.6	1487.8	1320.8	167
2000	4742.9	2492	2250.9	241.1
2005	14219.1	7619.5	6599.5	1020
2010	29739.9	15777.5	13962.4	1815.1

Source: China Statistics Yearbook, 2011

According to data from OECD since 2001, when China became member of WTO, the balance of payments of the country increased dramatically.[3] In 2001 it was $17.4 billion while in 2007 it reached $371 billion. During the last years, because of the international crisis, it has slightly decreased to 297 $billion in 2009 while in the second quarterly of 2010 it was $126.5 billion.

After China joined the WTO in 2001 there have been significant changes largely due to the reduction in tariffs on some products, such as soya and cotton. For this reason China has moved from being mainly self-sufficient in food, and thus an exporter, to being a net importer of foodstuffs and agricultural products. About cereals (rice, maize and wheat) China is still self-sufficient but this can change, especcially in the case of maize, which can be used as animal feed and consumption can be expected to increase because of increased meat consumption in the next few years. The increase in imports of agricultural products has only been partly compensated for by an increase in exports of processed foodstuffs and fresh vegetables. However, this must be contrasted, as we have said above, with the unprecedented growth in the positive balance of payments for manufactured products.

Table 2 - Foreign Trade (USD billion)

Item	**2003**	**2004**	**2005**	**2006**	**2007**
Total Value of Imports and Exports	**851**	**1,155**	**1,422**	**1,760**	**2,174**
Total Exports	438	593	762	969	1,218
Primary Goods	**35**	**41**	**49**	**53**	**62**
Manufactured Goods	403	553	713	916	1,156
Total Imports	413	561	660	791	956
Primary Goods	**73**	**117**	**148**	**187**	**243**
Manufactured Goods	340	444	512	604	713
Balance	**25**	**32**	**102**	**177**	**262**

[3]This does not include the balance with Hong Kong, which is then re-exported, mainly to the USA. Estimates, which more than double this surplus, are obtained by calculating the deficits which the main trading nations have with China, using data from the IMF.

Source: China Statistics Yearbook, 2011

In order to assure the availability of important products, China is establishing new bilateral trade agreements with major soya and cotton producers such as Brazil, with regards to the utilization of resources and agricultural production. In certain cases, such as Africa, these agreements are becoming increasingly precise, with more or less direct Chinese management of large areas of cultivated land in developing countries. In 2007 alone, China signed more than thirty "agricultural cooperation" agreements with African countries which involved more than two million hectares of land. The Chinese have obtained the right to cultivate palm-oil for bio-fuel on 2.8 million hectares of land in the Congo, which will thus become the largest palm-oil plantation in the world.[4] In a continent such as Africa, obtaining credits for agricultural production and innovation is very difficult or impossible (sub-Saharan Africa spends much less than India on R&D). China, however, has established eleven research centres in Africa for improving cereal production and this investment will also have positive effects for local producers.[5]

Besides the well recognised role of agriculture with regards to food security and, thus, reduction of rural poverty it is important to analyse the vast areas of rural China to better understand the role of agriculture even in terms of the relationships between rural workforce and the developments of other productive activities more centered around industrialized and urbanized provinces on the Eastern coast.

Based on the census data for agriculture and rural areas (1996 and 2006), an analysis of the changes registered in this decade will allow us to find answer about crucial questions on future development in China, from both an economic and social point of view. One crucial issue lies in the regional differences and the support given by rural areas to urban areas either for the development of industry then for satisfying the food requirements.

Even a brief analysis of the policies which have been adopted recently in favor of agriculture and rural areas provides important information which can help us to understand whether the development in China in recent years will continue as it has in the past or whether a shortage of working people will occur. With reference to this argument, many scholars have provided interesting analyses, often conflicting with one another. While some of them have concluded that China has reached the Lewis turning point, approaching a new era of workforce shortages

[4] The Chinese authorities are also negotiating with Zambia for the use of two million hectares to be used for oil-palms for bio-fuels. This has provoked protests from the Zambian opposition. Chinese companies in Zambia already produce 25% of the eggs sold in the capital, Lusaka. Finally, it is estimated that about one million Chinese farmers worked in Africa this year, which is an enormous number.

[5] China and Brazil will diverge more and more because of their different endowment of resources (land and water in particular) and China will probably become the major importer of food and non food products (soya beans, cotton) while Brazil will become the major exporter among the emerging countries.

from rural areas (Zhang, Yang, Wang 2010), some other Chinese scholars have conducted analyses showing that in China a phase characterized by a shortage of working people can cohesist with higher salaries (Zhang, 2011).

It is important to say that the reforms in China have produced some significant effects on poverty reduction. According to the WB data, in 1978 Chinese per-capita income was about $1,000 in PPP (prices from 2000), while in India and in Sub-Saharan African countries it was slightly higher (respectively $1,250 and $1,700). The reform of 1978 had a significant impact on poverty in China that, in less than ten years, decreased from 33% to 15% of the population. Even the per capita income increased reaching, in 2006, around $2,000 in current prices and more that $7,700 in PPP. The international scenario changed with China in a better position compared to other emerging countries in particular India where other importan reforms in rural areas were adopted but with a different approach (Gulati Fan 2007).

In the following years, poverty reduction progressed but at lower levels and, according to some scholars, this is due to the relationship between poverty and income growth and inequality. In particular, there has been an analysis of the negative impact of rising income inequality on poverty reduction, above all in rural areas (Yao 2000). Nowadays, according to the WB, poverty in rural areas is around 10%, while the Chinese statistics register poverty at around 5% of the population.

The aim of the following paragraphs is to underline the importance of those factors, such as changes in productive structure and the consumption model, and their consequences on internal problems in China and at an international level.

The data in table 3 show the main information about Chinese economy which significantly developed in the last thirty years reshaping the world economic geography. In fact in less than two decade China increased its GDP reaching the level of the largest economies in the world.6 Chinese economic growth has been far superior to that of India and Brazil, the other two large emerging countries. Even during the global financial and economic crisis of 2007 to 2009, the Chinese growth rate compared favorably with that of the developed economies, and recent data for 2010 confirmed that the average annual growth rate had returned to 10%.

[6] According to World Bank estimates, (World Development Indicators) China had an annual growth rate of about 10% between 1990 and 2006. During the same period, Indian GDP grew by 150% and Brazilian GDP by 50%.

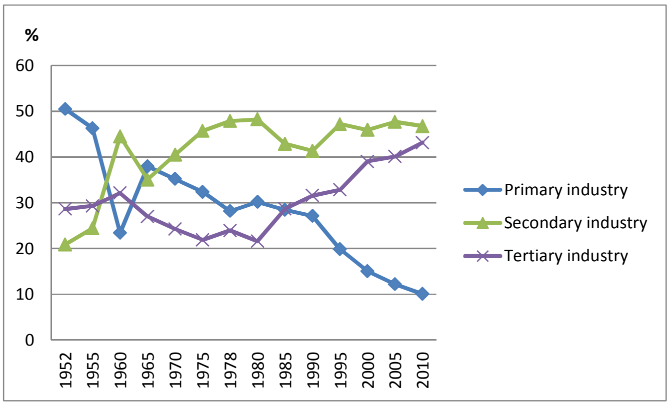
Table 3 - Principal Aggregate Indicators of National Economic and Social Development and Growth Rates in China (1978-2010)

Item	Aggregate Data				Indices and Growth Rates (%)		
					Average Annual Growth Rate		
	1978	1990	2000	2010	1979-2010	1991-2010	2001-2010
Population (10,000 people)							
Population at Year-end	96259	114333	126743	134091	1.0	0.8	0.6
Urban	17245	30195	45906	66978	4.3	4.1	3.8
Rural	79014	84138	80837	67113	-0.5	-1.1	-1.8
Employment (10,000 people)							
Employment	40152	64749	72085	76105	2.0	0.8	0.5
Registered Unemployment in Urban Areas	530	383	595	908	1.7	4.4	4.3
National Accounting (100 million yuan)							
Gross National Income	3645.2	18718.3	98000.5	403260.0	9.9	10.5	10.7
Gross Domestic Product	3645.2	18667.8	99214.6	401202.0	9.9	10.5	10.5
Primary Industry	1027.5	5062.0	14944.7	40533.6	4.6	4.0	4.2
Secondary Industry	1745.2	7717.4	45555.9	187581.4	11.4	12.5	11.5
Tertiary Industry	872.5	5888.4	38714.0	173087.0	10.9	10.7	11.2
Gross Domestic Product by Expenditure Approach	3605.6	19347.8	98749.0	394307.6			
Final Consumption Expenditure	2239.1	12090.5	61516.0	186905.3			
Household Consumption Expenditures	1759.1	9450.9	45854.6	133290.9			
Government Consumption Expenditure	480.0	2639.6	15661.4	53614.4			
Net Export of Goods and Services	-11.4	510.3	2390.2	15711.5			
Investment in Fixed Assets (100 million yuan)							
Total Investment in Fixed Assets		4517.0	32917.7	278121.9		22.6	23.0
Urban		3274.4	26221.8	241430.9		23.8	24.2
Rural		1242.6	6695.9	36691.0		18.3	17.3
Foreign Trade							
Total Value of Imports and Exports (USD 100 million)	206.4	1154.4	4742.9	29740.0	16.8	17.6	20.2
Exports	97.5	620.9	2492.0	15777.5	17.2	17.6	20.3
Imports	108.9	533.5	2250.9	13962.4	16.4	17.7	20.0

Source: China Statistics Yearbook, 2011

This strong economic development has underlined the great differences in China. About productive sectors the contribution of agriculture to GDP had fallen from 30% in 1978-1980 to a little more than 10% in 2010 while services had grown from 22% to more than 40% and industry remained important and contributed more than 46% of GDP in 2010.

Figure 1 - Percentage Composition of Chinese GDP by industry sectors from 1952 to 2010



* The data from 1952 to 1977 come from a different source

Source: China Statistics Yearbook, 2011

Table 4 – Percentage Composition of Main Indicators of National Economic and Social Development in China (1978-2010)

Item	1978	1990	2000	2010
Population and Employment				
Population				
Urban	17.9	26.4	36.2	49.9
Rural	82.1	73.6	63.8	50.1
Employment				
Primary Industry	70.5	60.1	50.0	36.7
Secondary Industry	17.3	21.4	22.5	28.7
Tertiary Industry	12.2	18.5	27.5	34.6
National Accounting				
Primary Industry	28.2	27.1	15.1	10.1
Secondary Industry	47.9	41.3	45.9	46.8
Tertiary Industry	23.9	31.6	39.0	43.1
Imports and Exports of Goods				
Composition of Exports				
Primary Goods		25.6	10.2	5.2
Manufactured Goods		74.4	89.8	94.8
Composition of Imports				
Primary Goods		18.5	20.8	31.1
Manufactured Goods		81.5	79.2	68.9

Source: China Statistics Yearbook, 2011

Knowledge of rural China, and, in particular, of Chinese agriculture is still incomplete and fragmentary. This is because of the vast size of the territory and the vastly different situations, as well as the enormous changes which have taken place in recent decades.

Since 1978 Chinese agriculture has been deeply changed from a system based on cooperatives and collective management to a new one with land management and cultivation being entrusted to rural households, while the land itself has remained the property of the village, the city or the state.

1.1. The role of rural areas in Chinese economic development: migration and workforce surplus

One of the most important contributions of agriculture to economic development has always been to provide a workforce to the other two sectors of the country's economy, i.e. industry and services. The transition from an economy mainly based on agriculture, characterized by underemployment and low productivity, to one with high employment in services and in industry or construction, is a significant contributor to increased productivity, economic growth and the development of urban and industrialized areas. Despite the improvement of living conditions in rural areas, it is the disparities with urban areas which are often emphasized in this development process, together with new problems in terms of income distribution and welfare.

An analysis of the Chinese case is particularly interesting because of the huge amount of farm workers who have left the countryside in the past three decades searching for a more or less precarious job in sectors of industry, in particular construction, and in services, often in provinces far away from those of their residence. The strong economic growth over the past two decades has led to record rates of increases in GDP of 10% per year, concentrated in coastal provinces and large municipalities in eastern China. These also experienced the benefit of the other major policy, initiated in 1978, to open trade with the rest of the world, culminating with China's entry into the WTO in 2001. The central and western provinces are different, with agriculture still playing an important role, and employment in this sector is significantly high in rural areas.

The results of two national censuses of agriculture and rural areas in 1996 and 2006 provide important information specifically about the role of rural economic development in China. At the same time, they allow us to quantify the reduction in agricultural employment and how it has particularily affected in diverse ways the different Chinese provinces and rural areas. The surveys also allow us to answer several questions about the effects of this huge flow of labor. In particular, it is important to know if, despite this reduction in agricultural labor, employment in small and very small family farms (over 200 million, as we saw earlier) is still characterized by a strong presence of underemployment and precarious labor and what is the possibility of new labor supplies for the future development of extracurricular activities in agricultural and urban areas. In addition, the movement of labor, especially for men and the younger age classes, influences the structure of family farms and the role that women play in conducting business? The presence of more than 130 million migrants moving from rural areas to work in extracurricular activities, often far from their place of residence, causes huge problems both in areas of immigration and of origin. This enormous mass of migrant workers, most of whom are linked to family farms, can change the occupational structure of rural areas and family farms, since in this context family businesses, though small and very small, play a role related to social safety nets for the kind of precarious and seasonal occupations involving the vast majority of immigrants. An analysis of these and other important issues can be useful, not only for understanding current reality better, but also for having more guidance on possible future developments in the vast rural areas and in the whole of China itself.

The main characteristics of migrant workers are very interesting. Almost two thirds are male and one third female. Two thirds are between 20 and 40 years of age, and most of these are between 20 and 30 years of age.

In terms of education, migrants present very specific characteristics with over 70% having a high school education level and another 19% a primary education level. The percentage of migrants with no education is very low (1.2%) and much lower than that of workers in rural areas (6.8%). Lack of education means less possibility to migrate from rural areas. Even the presence of higher education degrees among migrants is slightly lower than that of other workers (8.7% vs. 9.8%), but in this case this may be due to greater ability to find work in the area of origin. Indeed, this is particularly true for migrants who have a higher level of education (college or higher), whose percentage is similar to those of other workers in rural areas (around 1.2%).

Table 5 - Main characteristics of migrant workers from rural areas (2006)

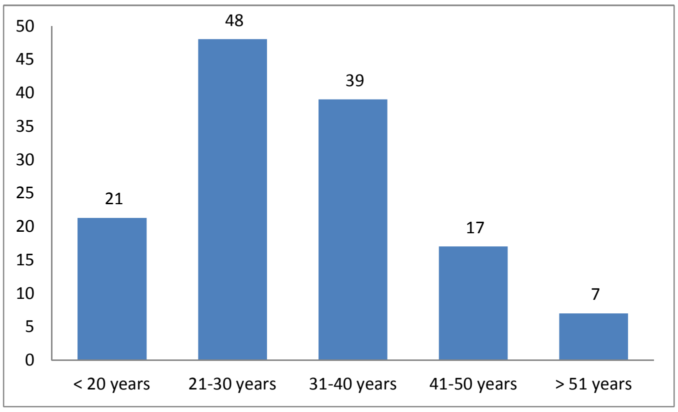
Percentage by gender and education level	National total	East	Central	West	Northeast
Male	64.0	65.8	62.8	63.1	70.2
Female	36.0	34.2	37.2	36.9	29.8
Age Classes					
< 20 years	16.1	14.2	17.6	16.1	16.7
21-30 years	36.5	36.1	36.6	36.7	35.4
31-40	29.5	27.3	29.3	32.2	25.4
41-50	12.8	15.4	11.9	11.1	15.3
> 51 years	5.1	7.0	4.6	3.9	7.2
Education level (%)					
Never went to school	1.2	0.9	1.1	1.7	0.5
Primary School	18.7	15.0	16.5	24.9	20.1
Middle School	70.1	70.9	73.0	65.5	71.8
Secondary school	8.7	11.4	8.4	6.9	5.9
College (higher level)	1.3	1.8	1.0	1.0	1.7

Source: Second National Census of Agriculture and Rural Areas (2006)

If we compare the migrant workers with all other workers in rural areas, they differ from them in several aspects, particularly from those employed in agriculture. Of the nearly 132 million migrants, the vast majority are male (64%), while in general the workers in rural areas are male and female in equal proportion. The situation is completely different, however, if we refer to those employed in agriculture where, because of the migration of many men, the percentage of women is slightly higher - more than 53% of the total.

With regards to structure by age, this is much more homogeneous for migrants and mainly concentrated in the ages between 20 and 40 years (66% of the total), with a high prevalence in the class between 20-30 years (over 36%) . The percentage of migrants younger than 20 years and those over 40 years is about 16-17%. Migrants older than 51 years represent only 5% of the total. On the other hand, there is the structure by age of other workers in rural areas, where most belong to higher age classes: workers between 20 and 40 slightly exceed 41% of the total, while 45% are workers over 41 years old and 25% are over 51.

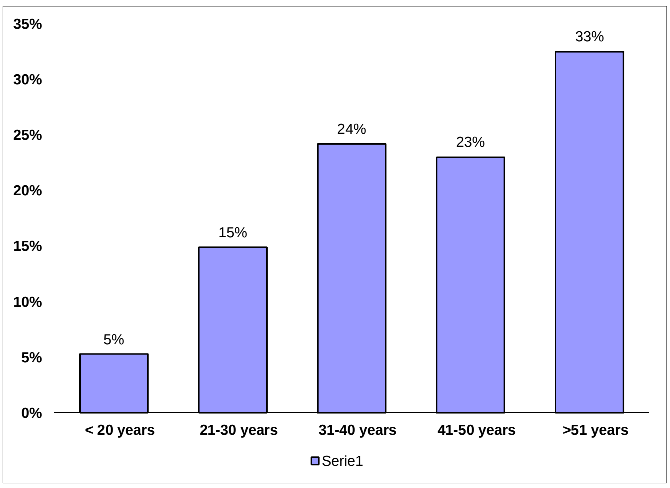
Figure 2 - Millions of migrants by age classes (2006)



Source: Second National Census of Agriculture and Rural Areas (2006)

These differences are even more evident when a comparison is made with farm workers in family households, from which most of the migrants come. In the family households, more than 55% of agricultural workers are over 41 years old, and nearly one third of the total are over 51.

Figure 3 Percentage of agricultural labor by age classes (2006)

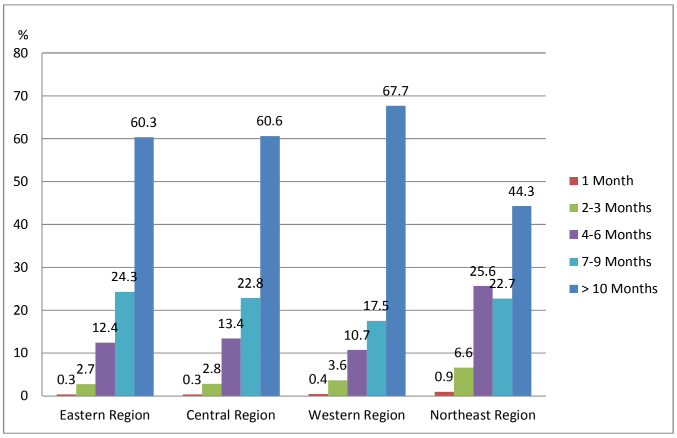


Source: Second National Census of Agriculture and Rural Areas (2006)

Migration distances are also noteworthy, with almost 50% of migrants working in a different province from where they are resident. In fact, while the other characteristics of migrants from rural areas are quite uniform, the differences are very remarkable with regards to migration distance, largely influenced by the degree of development of the province and the possibility of finding work in non-agricultural sectors near their place of residence. In 2006, almost half of migrants in rural areas moved within their province of residence, while the other half moved outside of the province, often very far away. This is due to the size of China and transportation difficulties over long distances and between rural and urban areas. In particular, it is observed that in the rich eastern provinces, more interested in the processes of economic development and urbanization, less than 20% of migrants are working outside their native province while in the central provinces, from where a substantial proportion of migrants come, and in the western provinces, they constitute over two thirds (67.6% and 60% respectively). This means that over 33 million people from central provinces, and almost 25 million from those in the west, move to the coastal provinces.

The length of time that migrants work away from home is also very long, often between seven and ten months a year. Around 110 million (85% of migrants) move for over 7 months and almost 82 million (60%) for over 10 months per year. Temporary migration is not significant, with a very low percentage of people migrating for less than three months (3-4%) in all the provinces of China.

Figure 4 Percentage of migrants by region and by length of working time



Source: Second National Census of Agriculture and Rural Areas (2006)

One final remark needs to be made about the type of activities in which migrants are employed. Approximately 57% of them work in industry (including construction), and 40% in services, while less than 3% work in agriculture. The main sector of employment is the manufacturing sector (almost 57%), but even the construction sector is very important with over the 38% of the total. Most migrants in this sector are male (70%), while the percentage of female workers rises from 30% in construction to reach about 43% in the service industries.

An analysis of the characteristics of migrants allow us to evaluate how they can impact on the structural transformations in the economic sectors of rural areas in particular, due to their continued links with part-time farms and their villages of origin. These relationships are an element of social safety against the precariousness of seasonal and migrant labor, contributing considerably to the "social stability" of the entire country by providing an element of social and economic security and a sort of “social cushioning” effect. This became especially evident in 2007 and 2008 when the loss of jobs caused by the recession resulted in millions of workers - some 28 million according to some estimates - returning to their farms, which in the meantime had been run by the women of the household.

The displacement of this large number of workers from rural to urban and industrialized areas in 2006, put more emphasis on the great structural change that has seen a sharp reduction in employment, especially in agriculture; over 82 million workers in the decade 1996-2006. This sharp reduction and the large number of migrants are closely linked with the process of urbanization that resulted in a tremendous increase of the Chinese population in urban areas, of about 400 million people.

Even more significant change will occur in the structure by age of those employed in agriculture in rural China when migration affects not only individuals but entire rural families. This will depend, in particular, on the new policies concerning rural development, especially with regards to land tenure. If they allow households to maintain the rights of land management, then households will be able to rent the land. If this happens, it will significantly increase the potential for the migration of entire families, with a significant impact on both the rural areas of origin and the urban areas of settlement. China's future development will depend on these aspects, not only in rural but also in urban areas.

1.2. Regional inequalities and Living conditions in modern China

Chinese economic development has, however, been accompanied by other changes which have resulted in great regional differences, as well as enormous demographic changes both in rural and urban areas. There has been an increase in the population, albeit at a rate less than in the past, and also an increase in urbanization. The ageing population has also increased the differences between urban and rural areas and between the coastal and metropolitan areas, on the one hand, and the interior and western provinces, on the other. A brief description of these important processes may help us to better understand the ongoing transformation in China.

In 2010, the Chinese population was more than 1.3 billion, compared to just over 962 million in 1978.[7] This population increase was, however, concentrated almost exclusively in the urban areas. Indeed, the rural population fell from 790 million, or 82% of the population, in 1978 to almost 670 million, or 50% of the population, in 2010. This phenomenon was particularly evident after the 1980's. During the same period the urban population grew from slightly more than 172 million to more than 660 million, with an average rate of increase of 4.4% a year. The population increase of more than 400 million was almost exclusively concentrated in the urban areas. Today, almost 50% of the Chinese population live in urban areas, compared with just over 19% thirty years ago[8].

To this growth of the urban population, one must also add the enormous number of migrant workers, as already discussed in the previous paragraph. In fact, the concentration of the population in urban areas and the enormous internal migration will also characterize the changes in progress in the next few years, and will have an enormous impact on the economic and social structure of both the areas where the immigrants come from and those where they go to work. Urbanization plays an important role, not only in increasing inequalities in the provinces themselves and between the rural areas and the cities, but also, and above all, at family level. Data on net disposable household income shows this very clearly. In 2010, urban households had an income of some 19,000 Yuan (almost $3,000) while rural households had an income of only 5,900 Yuan (just under $940). The increase in this difference can be seen from consumption patterns. While in the middle 1980's urban households consumed 2½ times as much as rural households, today they consume 3½ times as much.

The combination of urbanization and the increase in unequal consumption has resulted in a situation where in 2010 consumption by urban households amounted to three quarters of total consumption (76%), while that of rural households had fallen from 60% in 1978 to a little less than a quarter (23%). This also helps us to understand the immensity of the changes which have taken place in recent decades.

[7] The rate of population increase has been more than 1.1% per year for the last thirty years, although there has been a marked reduction due to birth control policies. Thus, it fell from 1.5% in the 1980's to just over 0.6% in the new millenium (2000-2007).

[8] OECD estimates say, however, that the Chinese urban population will only become greater than the rural one in 2016, while in many developing countries this threshold has already been passed.

Table 6 - Average Household Consumption Expenditure Per capita in China (1978-2010) – yuan at current prices

Year	Rural Household	Urban Household	Urban/Rural Ratio (Rural Household= 1)*
1978	138	405	2,9
1980	178	489	2,7
1985	349	765	2,2
1990	560	1596	2,9
1995	1313	4931	3,8
2000	1860	6850	3,7
2005	2579	9644	3,7
2010	4455	15907	3.6

*The effect of price differentials between urban and rural areas has not been removed in the calculation of the urban/rural consumption ratio.

Source: China Statistics Yearbook, 2011

This disparity between the provinces and the municipalities has been increasing in favor of Peking, Shanghai and the East coast provinces during the same period, and particularly in the last twenty years. Slightly less than 40% of the Chinese population live in Peking, Shanghai and the Eastern provinces[9], yet they produce more than 57% of the GDP and are responsible for almost 90% of Chinese trade with the rest of the world. Development and per capita income are much lower in the vast central provinces and particularly in the mountainous Western frontier provinces. To illustrate this, per capita income in China in 2010 was on average over 19,000 Yuan (about 3,000 $), but was over 25,000 Yuan in the Eastern provinces, significantly higher than that of the central and peripheral mountainous Western provinces (17,000 Yuan, almost 2,700 $).

This disparity is accentuated and more evident when one considers the different structural

[9] According to the official classification we consider as Eastern provinces: Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Fujian, Shandong, Guangdong, Hainan, Zhejiang; Central provinces: Shanxi, Anhui, Jiangxi, Henan, Hubei, Hunan; North-Eastern provinces: Jilin, Heilongjiang, Liaoning; asWestern provinces: Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang, Inner Mongolia.

realities. Looking at the data for 2010, with respect to the percentage of the population involved in agriculture as an indicator of the level of economic development, and per capita GDP as an indicator of wealth, we find that there is a clear negative correlation between these variables. In most Chinese provinces the agricultural workforce makes up between 30% and 60% of the total. The extreme cases are, on the one side, Beijing and Shanghai with 4.9% and 3.9%, and on the other side, Tibet with 53%. The per capita GRP in the Western provinces is between 13,000 and 27,000 Yuan, while a growing number of richer Eastern provinces are reducing the agricultural workforce and thus increasing by a still greater extent the per capita income of the population. In Beijing and Shanghai the agricultural workforce makes up about 4% of the population and the per capita income is around 74,000 Yuan which is almost four times higher than that of most of the other Chinese provinces.

This great structural, area and zonal difference has also drawn attention to the problem of assured food supplies and the contribution that agriculture and the rural zones make to general economic development in China. The great increases in both food production and consumption, especially during the 1990's, have made China one of the world's major producers and consumers of agricultural products. Today, China is the greatest producer and consumer of pork, poultry and eggs, as well as of agricultural commodities such as rice and wheat. Most of these commodities are for the internal market even though their value in terms of exports is constantly increasing. There has also been strong growth in the production and consumption of fruit and vegetables - including potatoes, tomatoes, asparagus and garlic – and China is the world leader in this area.

Table 7 - Main exports commodities in terms of volume and value (10,000 USD)

Item		2009		2010	
		Volume	Value	Volume	Value
Live Poultry	(10,000 heads)	696	2,608	696	2,610
Frozen, Fresh Beef	(10,000 tons)	1	6,121	2	10,909
Frozen, Fresh Pork	(10,000 tons)	9	26,272	11	33,201
Frozen Chicken	(10,000 tons)	7	13,627	10	20,828
Aquatic and Seawater Products	(10,000 tons)	209	680,851	243	880,218
Fresh Eggs	(million units)	1,111	7,889	1,298	9,789
Rice	(10,000 tons)	79	52,506	62	41,868
Maize	(10,000 tons)	13	3,171	13	3,335
Vegetables	(10,000 tons)	636	499,576	655	798,093
Fresh Vegetables	(10,000 tons)	424	218,760	421	377,180
Mandarins and Oranges	(ton)	985,127	50,640	814,322	52,098

Apples	(ton)	1,171,805	71,213	1,122,953	83,163
Pine Nut Kernels	(ton)	7,862	14297	7,127	15,928
Soybean	(10,000 tons)	35	23,714	16	11,825
Peanuts	(10,000 tons)	24	21,823	19	24,176
Edible Vegetable Oil	(ton)	114,019	15,149	92,461	12,262
Sugar	(ton)	63,886	3,365	94,348	6,386
Natural Honey	(ton)	71,831	12,570	101,138	18,251
Tea	(ton)	302,952	70,495	302,525	78,412
Dried Capsicum	(ton)	91,025	14,262	44,535	11,500
Canned Pork	(ton)	36,133	9,765	41,932	10,604
Canned Mushroom	(ton)	285,791	36,014	329,621	46,185
Beer	(10 000 liters)	21,030	12,269	19,410	11,234
Casings	(ton)	69,219	78,798	74,961	82,930
Cotton (Cotton Wool)	(ton)	8,249	1812	6,453	921

Source: China Statistics Yearbook, 2011

2. Data source: the Second National Agricultural Census of China (2006)

Every kind of research and study, in every type of field, needs a significant amount of data to be considered a valid study. Unfortunately, when we refer to China, and to many other developing or emerging economies, this statement needs to be contextualised for the differences in definition, methodologies and availability of data, and with regards to the quality of data, which is often not very reliable for several reasons. However tremendous the development of China in the last few years, the lack of statistical data remains a constant limit for many researchers.

Many sources are not accessible to foreigners and in many cases the data do not exist at all. In fact in many countries the statistica bureau of government do not release household-level data and therefore census data are the only vailable. In this context the Census data are very useful and powerful but still very limited due to the fact that the Census only covers one year, every ten years. Fundamentally this means loss in accuracy and in precision. However some studies investigate this aspect by calculating the error due to the use of aggregate census data especially for drawing high resolution maps of spatial patterns in poverty. The results show that errors associated is relatively low and the analysis can be considered relatively accurate (Minot, Baulch 2004).

In the following paragraph I will report the main differences between the two censuses in order to throws some light on the fundamental changes of a vast area of rural China.

2.1. The two National Censuses of Agriculture and rural China (1996 and 2006)

When in 1996 the first national agricultural census in China has been carried out the lack of information about Chinese rural areas was impressive. This was probably one of the most important limit to the development of national strategies aimed at reducing the rate of poor people in the country. Ten years after the first census, in 2009, the Chinese National Bureau of Statistics published the results of the second national census started in December 2006. The comparison between the two censuses is a fundamental source of information about the huge transformation happened in the rural areas from several points of view. The censuses covered not only agriculture and farm typologies (by size, labor, activities, households and non-households) but also private and public non-agricultural companies. They also provided information on the living conditions, education and mobility of a large part of the Chinese population. The surveys concern rural households, villages and towns and the data refers to the different types of households (rural and agriculture households), the rural conditions of

agricultural production and land use, rural labor and employment, irrigation and mechanization, presence and characteristics of migrants, rural livelihoods and infrastructure services.
The results show a significant changes in many relevant aspects from the number and average size of household and non-households farms to the reduction of rural population and workforce. With regards to the number of farms it is worthy of note that the household increased every where but the increase was greatest in the Eastern coastal provinces (+7%), less in the Central provinces (+2.3%) and least in the Western mountainous border provinces (+1%).

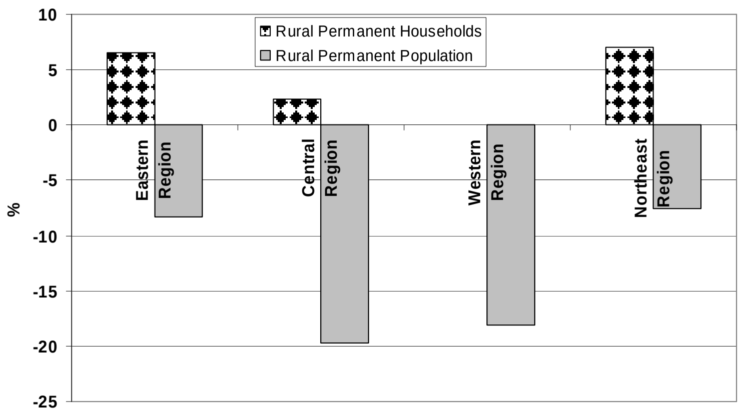
Table 8 - Number of Agricultural Holdings by Region in 1996 and 2006

Region	Household Agricultural Holdings (household)		Non-Household Agricultural Holdings(unit)		Household Agricultural 2006/1996		Non-Household Agricultural 2006/1996	
	1996	**2006**	**1996**	**2006**	**number.**	**%**	**number.**	**%**
National total	193,088,158	200,159,115	357,736	395,180	7,070,957	3.7	37,444	10.5
Eastern Region	60,940,891	65,500,377	145,693	193,181	4,559,486	7.5	47,488	32.6
Central Region	59,250,561	60,599,525	92,431	90,077	1,348,964	2.3	-2,354	-2.5
Western Region	60,613,762	61,280,654	91,403	86,921	666,892	1.1	-4,482	-4.9
Northeast Region	12,282,944	12,778,559	28,209	25,001	495,615	4.0	-3,208	-11.4

Source: Second national agricultural census of 2006, China Statistics Press, Beijing 2009

With regards to non-family run farms, mostly belonging to villages or cities, the most of changes were in the richer Eastern costal provinces (+32%) almost exclusively in Fujian and Zhejiang.
One important change taking place in Chinese agriculture is, however, the increasingly important role being played by these farms. While the overall number of pigs and poultry raised on such farms is not of great importance, they are on average large concerns, with an average number of more than 5,300 heads on poultry farms. Indeed, in recent years developments in animal husbandry production, and in particular in pig and poultry production, have been concentrated in large intensive farms run collectively by villages, cities and cooperatives. The presence and widespread use of animal husbandry confirms that the agricultural reforms have played a key role in reducing famine and malnutrition, and thus, also poverty, in rural areas.
These large non-family concerns are of particular importance in fish-farming, as fish becomes an increasingly important part of the Chinese diet. They are mainly collectives, and are aimed at serving the growing demand for fish in China, in particular in the urban areas.
One important change regards the reduction in rural workforce and rural population which for more than three quarters took place in the Central and Western provinces.

Figure 5 China % change in rural household and in rural population 1996-2006



Source: Second national agricultural census of 2006, China Statistics Press, Beijing 2009

Although the agricultural workforce is falling it is still a very important segment of total employment in rural zones. Indeed, according to the 2006 census, 71% of workers work on household farms in rural areas.

A brief mention should be made of developments in the workforce in non-household farms. These increased in number by more than 10%, from 358,000 in 1996 to more than 395,000 in 2006. There was a similar reduction in the workforce on these farms (-19.8%) as in household farms, but it was much less in the more developed Eastern provinces (-2.2%) and much greater in the Central (-44%) and Western provinces (-33%), while in the North Eastern provinces the workforce actually grew by 12%.

Finally some remarks have to be done with regards to the average size of rural household.

Overall, household farms manage some 122 million hectares of arable land. Thus, they are on average very small, with on average less than half a hectare of land each, or about 8 mu in Chinese measurements. The increase in the number of farms was, however, accompanied by a great reduction in the rural population. The number of permanent residents (resident for more than six months a year) fell from 874 million in 1996 to a little more than 745 million in 2006. This is a reduction of 128 million, or about 15% of the total.

The average number of permanent workers per farm went from 2.2 in 1996 to 1.7 in 2006. There was also a reduction in non-household farms where the number of permanent workers fell from almost 24 per farm to slightly more than 17 per farm in the same period. The average number of permanent workers on household farms is much lower in the coastal provinces (1.4 per farm in 2006), higher in the central provinces (1.7 per farm in 2006) and highest in the Western provinces (more than 2 per farm in 2006). To be more precise, there is an inverse relationship between the number of workers employed on household farms and the importance of employment in agriculture as a percentage of total employment in the individual provinces. This inverse relationship became stronger between 1996 and 2006. For example, in 1996 there were more than 3 workers per farm in the Westernmost provinces of Tibet and Qinghai, while at the other extreme, in the richest coastal provinces such as Peking and Shangai, the number of workers per farm was less than half that figure (1.5 workers per farm). In 2006, in Tibet and Xinjiang there were between 2.5 and 3 workers per farm while in Shanghai and Zhejiang there was less than one (0.6 and 0.8 respectively).

2.2. Some explanatory notes about data from the census

In the previous paragraph, I mentioned the main characteristics of the data used in my analysis. These data, collected in the Second National Agricultural Census, refer to specific conditions related to rural areas in China.

Given that an analysis of living conditions in the rural villages is the main objective of my study, I selected all variables from the Second National Census of Agriculture and rural areas. In particular, all variables related to villages and collected in the 4th section of the Census.

I then chose for my analysis those variables referring to the main aspects of living conditions in rural areas, namely in the villages, where living conditions present a lower level of development when compared to urban areas. Therefore, the selected variables focus mainly on the level of development of infrastructure and facilities. Some elements are particularly interesting, such as those related to environmental issues.

The selected variables are: Highways; Buses; Road Material; Electricity; Telephones; Kindergartens; Libraries; Cable TV: Primary Schools; Middle Schools; Drinking Water; Garbage Disposal; Biogas Pits; Improved Toilets; Clinics; Qualified Doctors; Hospitals; Supermarkets; Motor Wells; Water Ponds. For each of these there are some important explanatory notes from the census that will help in understanding the importance of these data and in the analysis of the results.

The 31 variables selected and used in our analysis relate to the main aspects of rural life: 1) Transportation facilities, 2) Electrical power and communication, 3) Culture and education, 4) Environmental health, 5) Sanitary facilities, 6) Medical Care, 7) Market development, 8) Drinking water. Secondary data taken from the National Agricultural Census (2006) are used and calculated, as appropriate. The complete list of the variables is reported below in table 9.

Table 9 - Variables groups of living conditions in rural China

Group	Variables
Transportation facilities	Villages with Highway*, Closest Bus Station/Port to Village Commune 0km, Road Material Cement & Tar (in village),
Electric power and communication	Villages with Electricity, Villages with Tel Network
Culture and education	Villages with Kindergarten(s), Villages with Library(ies), Villages with Cable TV, Closest Primary School to Village Commune 0km, Closest Middle School to Village Commune 0km,
Environmental health	Villages with Purified Drinking Water, Villages with Garbage Disposal Sites, Villages with Bio gas Pit,
Sanitary facilities	Villages with Improved Toilets
Medical care	Villages with Clinics, Villages with Qualified Doctors, Closest Clinic/Hospital to Village Commune 0km,
Market development	Villages with shops/supermarkets >50m2
Drinking water	Villages with wells with motor pumps Villages with water Ponds/Reservoirs.

*The word "highway" refers to well-built public roads, as defined in the Census book.

Source: National Agricultural Census Office of China (2009), *Abstract of the Second National Agricultural Census in China*, China Statistics Press.

The first variable refers to the number of villages with roads reaching the village from the outside. The word "highway" refers to well-built public roads, as defined in the Census book.

The second variable "buses" means the closest bus station or port and refers to the distance from the village to the closest bus station or port. Bus station means a location where there are regular or frequent buses passing by and one can get on the bus with a beckon.

The variable "road material" refers to the type of materials paving major roads from outside to inside the village. Road pavement includes cement, asphalt, gravel, brick and stone plate and other pavements. If there is more than one road from the outside to the village, the type of pavement of the highest grade of road is selected.

The variables "electricity" and "telephone" refer to the number of natural villages with access to electricity for undertaking normal production and living activities and to the number of natural villages with access to a fixed telephone or mobile phone for contact with the outside.

"Kindergartens" means the actual number of kindergartens and child care centers in the village at the end of 2006 that are established by collective, individual or other agencies and registered by the education administration at and above county level. It includes pre-school as well as unofficial kindergartens and child care centers with a certain scale (more than 10 children) without the approval of the education administration. The information reported in this variable are clearly interesting given the fact that the presence of this kind of service is not so usual in the rural villages.

In the variable "libraries" is the number of libraries or cultural centers, i.e. all specialized institutions engaged in loaning reading materials and mass cultural works in the village, established by village collective, individual or other agencies for the purpose of serving the public.

While "cable TV" refers to the number of natural villages that have installed a cable TV reception device and can receive TV programs normally.

Two very relevant variables are those referring to the closest primary and middle school, which means the distance from the village committee to the closest primary and middle school, no matter whether the primary school belongs to the village or to the township.

The variable "drinking water" is even more relevant, since it means whether drinking water of households at the site of the village committee has been centrally treated through purification and disinfection at the years-end. Drinking water from running a water plant is seen as being centrally treated through purification.

The variable "garbage disposal" means whether there is garbage treatment facility within the village to centrally treat garbage, or when there is no garbage treatment facility, the garbage is cleaned, transported and managed in a unified manner. While "Improved Toilets" means that most or all residents use toilets with digestion tanks or biogas digesters or three-separation tanks, some residents use public toilets or there are designated areas in other villages for dumping faeces. They have basically developed new systems compared to those of open faeces' tanks, faeces' pits, dry toilets and simple toilets which were more used in the past.

There are three variables which all relate to health care services. First is "Clinics" that refers to the clinics established by various economic organizations or individuals in the village that are licensed by public health administration at or above county level. Clinics need to have a fixed location for operations and mainly engage in medical treatment activities. It excludes specialized dental clinics and other establishments specialized in the sale of medicines. Secondly, "Qualified doctors" which refers to the physicians living the village who have obtained a license for practicing medicine from the public health administration at or above county level and who are practicing medicine. And finally, "Hospitals" refering to the distance from the domicile of the village committee to the closest hospital.

“Supermarkets” is, in fact, the number of shops and supermarkets with more than $50m^2$ of floor space. In particular, it refers to shops or supermarkets in the village that are engaged in the wholesale or retail of commodities with an operating area greater than 50 m^2.
Finally “Motorized wells” refers to wells with pumps driven by diesel engine, electric motor or other power machines for the irrigation of farmland, excluding wells to be installed with pumps. And “Water ponds” is the number of natural or dug ponds and reservoirs for irrigation solely owned by the village or jointly owned by multiple villages at the end of 2006.
“Roads” refers to pathways that automobile and tractors can go along.
These variables will be analyzed later when I comment on the results of the clustering process. Here, let just summarize that the basic infrastructure (such as transportation, communication and electricity supply) are quite well developed in rural China, with most of the villages having indicators of over 70%. However, while there is a large diffusion of primary schools in rural China, high schools have indicators that do not exceed 50%. Health facilities and tourism are very under-developed, mostly at around 10%-20%, and even lower in many villages. The difference between provinces, as we will see, are often relevant for many of the variables analysed.

3. Analysis of the data: the methodology

I apply two statistical methodologies in the first part of my analysis: the Principal Component Analysis and the Cluster Analysis. The first one has been applied in order to reduce the number of original variables selected from the Chinese Second Agricultural Census of 2006. In fact, I choose 20 original variables for my analysis, as has already been explained in the previous section about data. Thanks to the PCA, I reduced the variables into 4 Principal Components (PC) which were then utilised in the Cluster Analysis (CA).

3.1. Principal component analysis (PCA)

In social sciences researchers often deal with large dataset including several highly correlated variables which therefore put some constraints on the analysis, both in terms of simplification and synthesis. For this reason, researchers are usually interested in reducing the number of variables. Principal component analysis (PCA) is a statistical technique aiming at a linear transformation of an original set of variables into a smaller set of uncorrelated variables retaining the main information of the original ones (Dunteman 1989). In particular, it is a data reduction method, which helps to summarize and order the information in a large data set and, hence, to avoid double counting. Originally conceived by Pearson (1901), it has been developed by Hotelling (1993) who introduced the most famous formulation of this technique.
Since the objective of PCA is to maximize the variability explained by the components, the total variability of the *p*extracted components equals the total variability of the *k*original variables. It permits easier selection of a sub-set of components (Mario Mazzocchi, 2008).
The original variables are linearly transformed through the equation of PCA that can be expressed as follows:

$$C = XA' = \begin{cases} c_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p \\ c_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2p}x_p \\ \qquad\qquad \ldots\ldots \\ c_p = a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pp}x_p \end{cases} \qquad (1)$$

Where C is the n x p matrix of principle component scores,X is the data matrix, and A is the p x p matrix of component loadings. Once the matrix A has been computed, the component scores (i.e., $a_1, a_2, \ldots a_p$) can be mathematically calculated to maximize the variation of the linear composite or, equivalently, to maximize the sum of the squared correlations of the principal component with the original variables.

$$\hat{a}_{ij} = \frac{a_{ij}}{\sqrt{\lambda_j}} \qquad (2)$$

where λ_j is eigenvalues.

Geometrically, each principal component is the line of closest fit to the *n* observations in the *k* dimensional variable space. It minimizes the sum of the squared distances of the n observations from the lines in the variables space representing all the principal components.

In this analysis, the components are computed on a covariance matrix since the original variables selected are measured in the same units, namely in % terms. According to the covariance matrix computed here, the variables used in the analysis are highly correlated, so it is reasonable to apply the PCA to extract useful indicators. The method used to determine the number of components is based on the previously explained level of variance[10] and Guttman-Kaiser's criterion[11]. I extracted four components with eigenvalues greater than 1, which were able to explain nearly 80% of the initial variances (see 3.1).

The data in the table 10 are the same as those reflected in the figure of the screeplot. The curve tends to be smooth after the 4th component. This confirms that choosing the first 4 components is justifiable.

Table 10 - Total Variance Explained

	Initial Eigenvalues (λ)		
Components	Total	% of Variance	% Cumulative
1	.263	40.852	40.852
2	.129	20.001	60.853
3	.104	16.188	77.040
4	.046	7.086	84.127
5	.027	4.195	88.322
6	.019	2.967	91.289
7	.015	2.353	93.641
8	.012	1.883	95.524
9	.008	1.226	96.750
10	.007	1.063	97.814

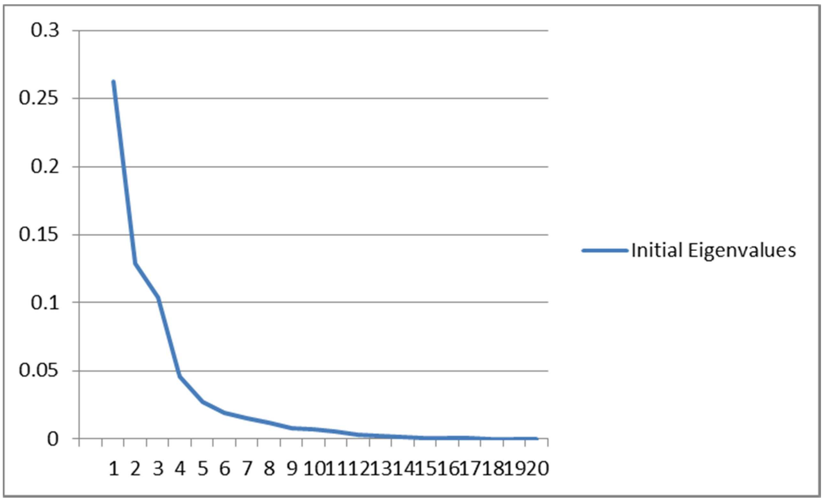
Source: author processing on Second National Agricultural Census in China (2009)

The value of initial eigenvalues (λ) presented in the table are the variance of each components and it allows us to calculate how much of the total variance of the variables each component can explain.

[10] The components should be able to explain 70% - 80% of the total variability.

[11] *Guttman-Kaiser's criterion:* considering the components with eigenvalues greater than or equal to 1. A lower level would explain less than the standardized variance (=1).

Figure 6 - The screeplot diagram: number of components and their initial eigenvalue



Source: author elaboration on data from Chinese Second National Agricultural Census 2006

Applying the Kaiser rule and on the base of the screeplot diagram, 4 components are selected. Table 11, here below, shows the unrotated component matrix.

Table 11 - Components Matrix

Original Variables	Principal Components			
	1	2	3	4
Highways	.034	.050	.015	.009
Buses	.042	.003	.053	-.019
Road_Material	.219	.006	-.004	.046
Electricity	.035	.069	.038	.023
Telephone	.038	.062	.035	.012
Kindergartens	.049	.088	.030	-.041
Libraries	.094	-.028	-.015	.026
Cable_TV	.173	.118	-.020	-.081
Primary_Schools	-.060	.067	.051	.071
Middle_Schools	-.002	.009	.005	.004
Drinking_Water	.232	-.039	-.052	.020
Garbage_Disposal	.216	-.032	-.059	.032
Biogas_Pit	-.117	.152	-.016	.113
Improved_ Toilets	.150	.025	-.090	.043
Clinics	.051	.121	.059	.037
Qualified_Doctor	.061	.129	.077	.013

Hospitals	-.003	.007	.001	-.001
Supermarkets	.111	.027	.071	-.036
Motor_Wells	.077	-.057	.240	-.017
Water_Ponds	-.050	.170	-.086	-.096

Source: data from the Second National Agricultural Census, author processing with software SPSS

Table 12 shows the rotated solution which can be easier to interpret. In fact, even though the retained principal components are interpretable, some researchers prefer to rotate the principal components, as is typically done in factor analysis. An orthogonal rotation is just a shift to a new set of coordinate axes in the same subspace spanned by the principal components. As we have seen, each principal component will find the two coordinate axes that define the plane which is the closest fit to the essentially two dimensional swarm of points. However, any two other perpendicular coordinate axes lying in the same plane can also help in describing the observations, maintaining all the information. Like the principal components, the new coordinate axes are also defined by their correlations (loadings) with the original variables but their pattern will be more conceptually appealing, thus allowing for a simpler interpretation.
There are several rotation procedures but the Varimax method with Kaiser normalization is considered the most popular. In this rotation the axes are rotated so as to maximize what is called the varimax criterion. This criterion results in a new set of orthogonal coordinate axes where each new coordinate axis has either large or small loadings of the variables on it.

Table 12 - Rotated Component Matrix

Variables	Components			
	1	2	3	4
Highways	.005	.033	.001	.004
Buses	-.013	.002	.061	.018
Road_Material	.245	.149	-.030	-.150
Electricity	.006	.089	.004	-.016
Telephone	.002	.061	.013	.001
Kindergartens	-.026	.029	.070	.132
Libraries	.066	.011	-.020	-.058
Cable_TV	.041	-.037	.153	.473
Primary_Schools	-.011	.198	-.063	-.152
Middle_Schools	.000	.002	.000	-.001
Drinking_Water	.283	-.031	-.043	-.063
Garbage_Disposal	.267	.005	-.088	-.093

Biogas_Pit	.025	.483	-.363	-.275
Improved _Toilets	.219	.073	-.189	-.055
Clinics	.012	.227	.004	-.032
Qualified_Doctor	-.014	.219	.068	.035
Hospitals	.000	.001	.000	.001
Supermarkets	-.009	.011	.150	.085
Motor_Wells	-.163	.105	.605	-.173
Water_Ponds	-.111	-.076	-.034	.590

Source: data from the Second National Agricultural Census, author processing with software SPSS

From the PCA result matrix, we can interpret the four principle components as follows:

The PC1 gives higher values to the variables related to road materials, purified water, garbage disposal sites, improved toilets, those items strictily related to basic needs for a decent living standard.

The PC2 gives higher values to variables such as availability of electricity, telephone networks, clinics and qualified doctors which were defined as public services.

The PC3 gives higher values to variables related to the distance from bus stations or ports, kindergartens and supermarkets which were defined as the availability of, or proximity to, facilities and services.

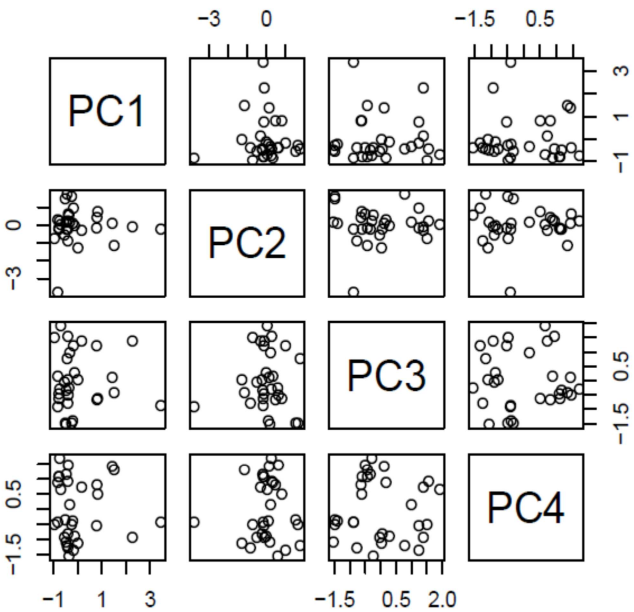
The PC4 does not provide much information since it seems to give higher values only to water ponds and cable TV.

Table 13 - Original variables spread out by components

Components	**Variables**
C 1	Villages with Purified Drinking Water; Villages with Garbage Disposal Sites; Villages with Improved Toilets
C 2	Villages with Electricity; Villages with Tel Network; Closest Primary School to Village Commune 0km; Closest Middle School to Village Commune 0km
C 3	Closest Bus Station/Port to Village Commune 0km Villages with kindergartens Villages with shops/supermarkets >50m2;
C 4	Villages with water Ponds/Reservoirs Villages with Cable TV

The first four components, which explain about 80% of the initial variance among the clusters, are related to living conditions, ICT technologies and proximity to different facilities. Both the local social conditions and the degree of openness of villages are relevant when characterizing the reality of rural villages in China. The other components are less capable of explaining the main differences between rural areas, and they suggest that there is more homogeneity between the villages, even when the level and presence of facilities and services is not high. One should note that Tibet is in a particular position, as it is the outlier for most of the variables considered. Before proceeding with the next step of the analysis, which consists of the cluster analysis, it can be useful to graphically analyse the distribution of the four principal components resulting from the PCA run in the previous paragraph. In fact, the simple scatterplot matrix here below would be a graphical support to revealing or not revealing a clear pattern of clusters in the data.

Figure 7 - The simple scatterplot matrix



Source: data from the Second National Agricultural Census, author processing with statistical system R

More specifically, this diagram shows the pairwise relationship between the 4 Principal Components detected from the PCA. In fact, the advantage of employing principal component analysis in cluster analysis is to be able to plot the component scores and visually search for clusters of observations as we have done above. Even though the four principal components are by definition uncorrelated to each other, this diagram can help answer the question as to whether there is clustering by groups, and also provide some initial information about the distribution of data and the possible ways to cluster the Chinese provinces.

3.2. Cluster analysis

The cluster analysis can be considered one of the most powerful methodologies applied within the domain of multivariate data in order to synthesize information coming from dataset of different sizes. Indeed, the tecniques of cluster analysis are adopted in several research fields, including the social sciences. This kind of technique is particularly useful for classifying objects on the basis of similarities with respect to the clustering variables, rather than on predetermined threshold values (Yang, Hu 2008).

Many scholars within the domain of statistical analysis do not consider this methodology as a consistent one, given the fact that it is strictly dependent on the nature of the data and strongly influenced by the point of view of the researcher (Steinbach, Kumar 2003). However, international organizations and public institutions always set criteria for classification that can be considered as relatively subjective as the techniques of cluster analysis. Geographic distribution facilitates in formulation of new policies, especially those designed to policy reduction. Moreover this knowledge helps in evaluating progress and detecting th impact of other factors (Minot, Baulch 2004).

With cluster analysis, researchers can group data objects based on one specific piece of information found in the data. The goal is to obtain groups whose objects are the most similar to each other, and the most different compared to the objects belonging to other groups. It is broadly recognised that there is not just one definition of a cluster and that it depends significantly, as already said, on the nature of the data and on the personal discretion of the researcher; that is, the results that he/she would like to obtain from the analysis (Tan, Steinbach, Kumar 2006). This is one evident argument against the cluster analysis.

To run a cluster analysis it is fundamental to follow some basic steps starting from the measure of distance for individual observations and proceeding with the choice of clustering algorithm, the definition of distance between the clusters, the determination of the number of clusters and finally, the validation of the results (Mazzocchi 2008).

The effect of the different measures would significantly influence the results only in the presence of outliers (Mazzocchi 2008). Therefore, I finally used the Euclidean distance which is the most common measure.

At the same time, I focused most of my study on choosing the clustering algorithm. In fact, there are several possible algorithms, and each of them can be chosen or excluded for different reasons. These I will introduce shortly before explaining the one I used in my analysis.

3.2.1.Hierarchical and Partitional (or No-hierarchical) Algorithms

A first important distinction, among all the possible clustering methods, is with regards to the reference vectors and whether they are predefined or not. When we have *a priori* knowledge of the data patterns the data will be classified with supervised clustering. On the other hand, with unsupervised clustering, which refers to all the algorithms that do not imply any *a priori* knowledge, the classification is based on the structure that is observed within the data itself (Kaufman, Rousseeuw 1990).

The types of unsupervised clustering algorithms can be divided into two groups: hierarchical and the non-hierarchical or partitioning methods. The main difference lies in the fact that with the non-hierarchical methods the number of clusters is defined by the researcher, while with the hierarchical method it is the result of the algorithm. Therefore, it is evident that the hierarchical method can be considered more objective from this point of view. Clearly, each method presents weaknesses and strengths.

The hierarchical method does not produce a particular number of clusters but the result is more a set of nested clusters starting from two possible situations: one where every cluster contains only one single observation - that is, the number of clusters equals the number of observations. – and the other which starts with one cluster containing all the units. The first method is called *agglomerative*, since it proceeds by a series of successive steps, and in each step two clusters are merged until the last step when only one cluster is left. The second method is called *divisive*, since all the *n* individuals are separated successively into groups. The Hierarchical method is, compared to the Partitional one, less reliable, since with this method it is not possible to repair what was done in the previosu step. This rigidity has been therefore considered the key to both the success and the disadvantage of the hierarchical method (Kaufman, Rousseeuw 1990).

The Partitional method should compensate for the rigidity of the hierarchical method. This category of cluster analysis methods includes algorithms such as K-means and Partitioning Around Medoids PAM. However a significant difference between K-means and PAM algorithm lies in the higher sensitivity of the first one to the outliers. Indeed PAM is less sensitive since it uses most centrally located objects (Park, Lee, Jun 2008). In general PAM is known as the most powerful but it is does not work efficiently with large dataset (Han et al, 2001).Because my dataset is not so large, I will focus on the PAM algorithm, defined as a more robust version of the K-means algorithm.

The first step of PAM consists of computing *k* representative objects, called *medoids*, which can be defined as the objects of a cluster which minimize its average dissimilarity to all the objects in the same cluster. The detection of these representative objects is meaningful since they can represent various aspects in the structure of the investigated dataset for our study (Kaufman, Rousseeuw 1990).

After finding the set of *k medoids*, i.e. the representative object of each cluster, each object of the data set is assigned to the nearest medoid. In fact, as far as all the other algorithms are concerned, the aim is to cluster *p* elements x_j, $j \in \{1, \ldots, p\}$ and this means minimizing the dissimilarities among all the elements within each cluster, while maximizing the dissimilarities between each cluster. The most used measure of dissimilarity is the same as for the hierarchical method: the Euclidean distance. The *k* representative objects should minimize the objective function, which is the sum of the dissimilarities of all the objects to their nearest medoid. It can be run in two steps: *BUILD-step* that sequentially selects *k* "centrally located" objects, to be used as initial medoids and *SWAP-step* that is carried out if the objective function can be reduced by interchanging (swapping) a selected object with an unselected object and ends only when the objective function can no longer be decreased.

There is additional information on the diameters and separation of clusters. Even more important and useful is the graphical display called *silhouette*, that provides information about the goodness of the clusters obtained. The silhouette plot shows the silhouettes of all clusters next to each other and it is quite useful for deciding the number of clusters. Through the silhouette plot it is possible to check whether the objects are well represented or not by the cluster to which they have been assigned.

It is normally a good rule to run PAM several times, each time for different values of *k*, and then compare the resulting silhouette plots. The average silhouette width, *s(i)*, can be used to select the 'best' number of clusters, by choosing that *k* which yields the highest silhouette width.

After having calculated all the *s(i)* for all the objects included in each cluster they can then be plotted in the diagram in order to show the average dissimilarity of *I*, *a (i)*, to all other objects belonging to the same cluster A. This average dissimilarity will then be calculated for all the other objects not belong to cluster A but to any other clusters that can be defined in the data *b (i)*. This measure of the average dissimilarity needs to evaluate whether each object has been assigned to the right cluster, i.e. to the one that minimizes its dissimilarity with the other objects in the same cluster. In particular, we have to take into consideration the value of s(i) that is calculated by combining the single average dissimilarity related to each cluster. Here it is the formula:

$$s\,(i) = \frac{b\,(i) - a\,(i)}{\max\{a\,(i), b(i)\}}$$

Based on the average dissimilarities (*a (i)*, *b (i)*, etc..) it is then be possible to calculate the average silhouette width both for each cluster $\overline{s_k}$, and for the entire dataset $\bar{s}$.
The aim is to find the *k* that maximizes the silhouette coefficient (SC).

$$SC = \max \bar{s}(k)$$

Here below is a table (14) with a range of possible values of the Silhouette Coefficient (SC) and possible ways to interprete them, as introduced by Kaufam and Rousseeuw (Kaufman, Rousseeuw 1990)

Table 14 - The range of Silhouette Coefficient and its interpretation

OF SC	*INTERPRETATION*
0.71-1.0	A strong structure has been found
0.51-0.70	A reasonable structure has been found
0.26-0.50	The structure is weak and could be artificial. Try additional methods of data analysis.
< 0.25	No substantial structure has been found

Source: Kaufman, Rousseeuw 1990

Concerning the choice of the type of clustering algorithm to use, I stated at the very beginning of the paragraph that it is influenced both by the nature of the data and the purpose of the analysis. It is therefore strongly recommended to run more than one analysis through different methods and then compare the results (Kaufman, Rousseeuw 1990). I also put the stress on the characteristics of the data analysed here and how they limited the choice from several points of view. Considering one particular aspect, that is the size of the sample used in this analysis, and considering that some methods of cluster analysis can be adopted mainly with a large sample, I finally opted for using more than one algorithm in order to compare the final results obtained from each of them. Indeed, I first ran the analysis on the basis of the hierarchical method, measuring the distance between the observations in terms of Euclidean distance. Thanks to some basic rules that can be adopted, I calculated the most reasonable number of clusters. In particular, I will integrate the analysis with some graphical tools that provide useful information in order to define the best number of clusters. Two of those tools are: the screeplot and the silhouettes. With both of them showing the graphical distribution of the data it is possibile to make the best choice with regards to the number of clusters.
In the following paragraphs, I will present the results from the application of the different methods: the first one is the hierarchical and then the Partinioning Around Medoids (PAM).

The analysis has been run with the help of the software SPSS and R, a system for statistical computation.

3.3. A new geography of rural China: the results of cluster analysis

According to the official geographic classification of China, there are four main groups of provinces: Eastern coastal area, Central area, Western area and North-East. Interestingly, these rank from developed to undeveloped in socio-economic terms as one moves from the Eastern to the Western part of China.

Figure 8 - The official geographical classification of Chinese provinces

Source:Chinese yearbook

There is also a more specific and detailed classification which divides China into seven zones: Northern, Central, Southern, Eastern, Northeast, Southwest and Northwest. Though these classifications are based on the geographical location, the districts in the these main geographical zones still have a lot in common in socio-economic terms.

Figure 9 - Geographic Regions of China (First Agricultural Census 1996)

Source: First National Agricultural Census, China 1996

Figure 10 - Geography of Economic Regions of China

Source: First National Agricultural Census, China 1996

Based on the results gathered from the cluster analysis ran in this thesis, I have been able to draw up a map of Chinese rural villages in terms of the level of development of infrastructure, facilities and services .

In particular, with the support of the ArcGIS (version 9.2) software, it was possible to observe the geographic distribution of the results obtained with the two methods of cluster analysis.

In the next paragraphs I will report the results of my analysis both those of the hierarchical and the PAM algorithm. In paragraph 4 I will go further with the analysis by adopting the fuzzy logic approach in order to build a new index about living conditions in the rural villages of

China. This will allow me to enlaring the analysis with additional results that it will be possible to compare with those coming from the clustering.
Then in the last paragraph I will present the maps reporting all the results of the analysis carried out in this thesis. Through the maps it would be easier making a comparison and underline the main trends in the geographical distribution of living conditions in the rural villages of China. This new geography of rural China could add some information with regards to the official classification previously introduced.

3.3.1. Results of the Hierarchical algorithm

The clustering process in this paragraph will use a hierarchical method to analyze the 31 provinces and municipalities in China (excluding Hong Kong, Macau, the two special administrative regions and Taiwan), based on the four components extracted by PCA from the initial 21 variables and which are able to explain around 80% of the total variability.
Here, Ward's hierarchical method is used[12] and the Euclidean measure of distance[13]. For a generic number of co-ordinates n, the equation between two observations i and j is as follows:

$$D_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ki} - x_{kj})^2}$$

where X_{ki} is the measurement of the K^{-th} variable on the i^{-th} observation. All of the data are processed by the statistic software SPSS.
In this sector, the Q-mode cluster, which refers to the sample, is used.
Here below (table 15) the agglomeration process obtained selecting the Euclidean quadratic distance as a measure of distance for the individual objects and the Ward method for the distance between clusters.

12 In *Ward's method*, the sum of squared distances is firstly computed within each of the clusters, then the aggregation between two clusters with the smallest increase in the total sum of squared distances is chosen.
13 *Euclidean distance*, the best-known measure of distance, is the length of a line segment connecting two points.

Table 15 - The Agglomeration process of Hierarchical Method

Stage	Number of Clusters	Combined Cluster		Nesting Distance	Difference in nesting distance
		Cluster 1	Cluster 2		
1	30	18	23	0.016	
2	29	20	25	0.068	0.052
3	28	27	28	0.128	0.06
4	27	17	18	0.216	0.088
5	26	13	19	0.323	0.107
6	25	6	8	0.505	0.182
7	24	14	22	0.688	0.183
8	23	4	27	0.902	0.214
9	22	21	24	1.179	0.277
10	21	3	16	1.624	0.445
11	20	7	15	2.105	0.481
12	19	29	31	2.627	0.522
13	18	12	17	3.168	0.541
14	17	6	7	3.852	0.684
15	16	12	14	4.559	0.707
16	15	10	11	5.531	0.972
17	14	4	30	6.529	0.998
18	13	1	2	7.715	1.186
19	12	4	29	9.359	1.644
20	11	5	6	11.135	1.776
21	10	20	21	13.311	2.176
22	9	10	13	15.782	2.471
23	8	1	9	21.513	5.731
24	7	3	4	28.135	6.622
25	6	10	12	36.076	7.941
26	5	3	20	45.821	9.745
27	4	5	10	59.807	13.986
28	3	3	26	76.029	16.222
29	2	1	3	94.263	18.234
30	1	1	5	120	25.737

Source: data from the Second National Agricultural Census.

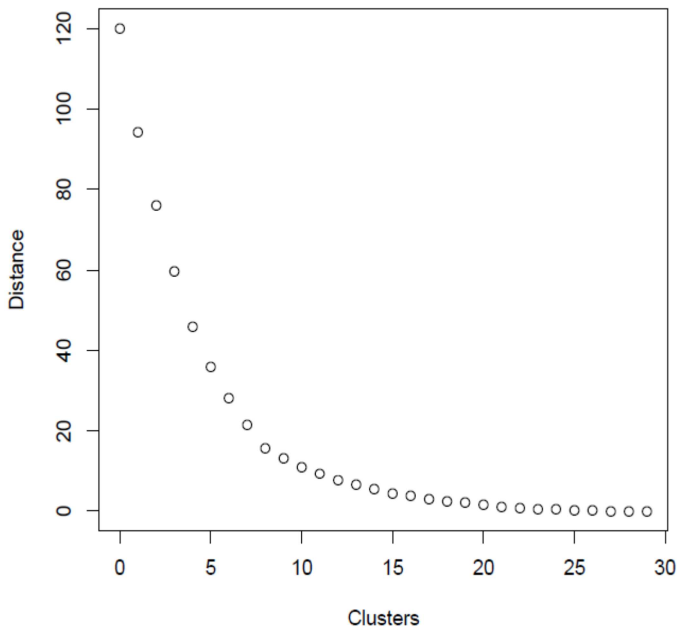
Author process with SPSS software

The agglomeration process resulting from the hierarchical method is a table with all the steps of the process, and for each step it shows the nesting distance between the two clusters that have been merged.

One rule for the determination of the ideal number of clusters tells us that the number of clusters can be defined by looking at the nesting distance difference between the maximum possible number of clusters, that is the number of observations (N), and the stage (first column of table 15) before the observed large increase in the nesting distance.

It follows that the ideal number of clusters in my analysis could be more than one, since there are several stages with difference in nesting distance. Therefore, it can be either: 31-27=4, since we observed a first significant difference in nesting distance at the 27th stage of the agglomeration process; or 31-23/24=8/7. Based on this nesting distance, it is possible to draw the screeplot diagram in which the difference in nesting distance is on the y axis while the number of clusters is on the x. The ideal number of clusters will correspond to the one after which the slope of the curve seems to be zero. Then the curve is almost horizontal. In the diagram here below, I have reported the data from my analysis and it emerges that the ideal number is between 7 and 8.

Figure 11 - The screeplot diagram of number of clusters and distance



Source: data from the Second National Agricultural Census, author's processing with R software

I will compare this result with those from the rest of my analysis.

In particular, since it looks as if the number of clusters is still not definitive, I will try to find the best solution by using some other graphical tools that are normally associated with the other algorithm that I applied in my analysis.

3.3.2. Results of the PAM algorithm

On the basis of the results of the hierarchical method, one could be convinced to choose the number of clusters between 7 and 8, but even 4 was suggested as a possible good solution. These are the results gathered by the screeplot diagram and rule based on the variation in the nesting distance.

However, as I said at the beginning of this paragraph, the nature of data placed several limits on my analysis and forced me to choose to run the analysis several times in order to have more results to compare.

In this paragraph, I will show the results from the PAM algorithm which, as a more robust method, shows more clearly the characteristics of the data.

As for the hierarchical method, the main objective of the PAM algorithm is to find clusters that can better express the differences among the objects belonging to different clusters and the similarities within the cluster itself. This algorithm aims at defining the *k* representative objects among the objects of the data set, and these K objects have to represent various information about the structure of the data. For each different algorithm included in the group of non-hierarchical (or partitioning) methods, the representative objects are called and defined in different ways. Often they are generically called centrotypes and for the PAM algorithm they are called medoids. The clusters will be identified around these medoids, by calculating the distance/dissimilarity of each observation from/with the choosen medoid. Of course, the more a medoid can reduce the average dissimilarity within each cluster the better the partition is.

I ran my analysis with the support of the statistical software R.

The first time I ran the PAM algorithm, choosing 4 as the number of clusters and then I ran the same algorithm again but with a different number of clusters, starting with 7. The kind of data that I entered is not a matrix of dissimilarity and the measure of distance is the Euclidean distance. It is possible to choose the medoids but I did no do so, leaving the programme to choose the best ones.

```
pam (PCdata, 4, diss= FALSE, metric="euclidean", medoids=NULL,
stand=FALSE,
cluster.only=FALSE, do.swap=TRUE, keep.data=FALSE, trace.lev=0)
```

The algorithm gives the following output:

For 4 clusters

```
Medoids:
          ID        PC1       PC2       PC3       PC4
Jilin      7 -0.326153  0.212808  0.997920  0.168447

Gansu     28 -0.553535 -0.516698 -0.040648 -0.902270

Guangdong 19  0.802654  0.457477 -0.643697  0.813746

Sichuan   23 -0.776989 -0.132240 -0.435825  1.083758
```

```
Clustering vector:
Beijing  Tianjin    Hebei      Shanxi         Inner Mongolia
1        1          1            2              1
Liaoning  Jilin    Heilongjiang     Shanghai      Jiangsu
1         1              1              3          3
Zhejiang  Anhui    Fujian      Jiangxi        Shandong
3         4         3           4               1
Henan    Hubei    Hunan        Guangdong      Guangxi
1         4      4              3              3
Hainan   Chongqing   Sichuan       Guizhou        Yunnan
2        4             4             2               3
Tibet    Shaanxi   Gansu         Qinghai        Ningxia
2        2          2            2              2
Xinjiang
2
```

```
Objective function:
   build     swap
1.218008 1.202461
```

For 7 clusters

```
Medoids:
          ID        PC1       PC2       PC3       PC4
Shanghai   9  3.436582 -0.183800 -0.882612 -0.402841

Jilin      7 -0.326153  0.212808  0.997920  0.168447
```

```
Shaanxi   27 -0.410504 -0.239082  0.049437 -0.781954

Guangdong 19  0.802654  0.457477 -0.643697  0.813746

Sichuan   23 -0.776989 -0.132240 -0.435825  1.083758

Guangxi   20 -0.510834  1.520676 -1.488208 -0.319545

Tibet     26 -0.833299 -3.777897 -0.888873 -0.406099
```

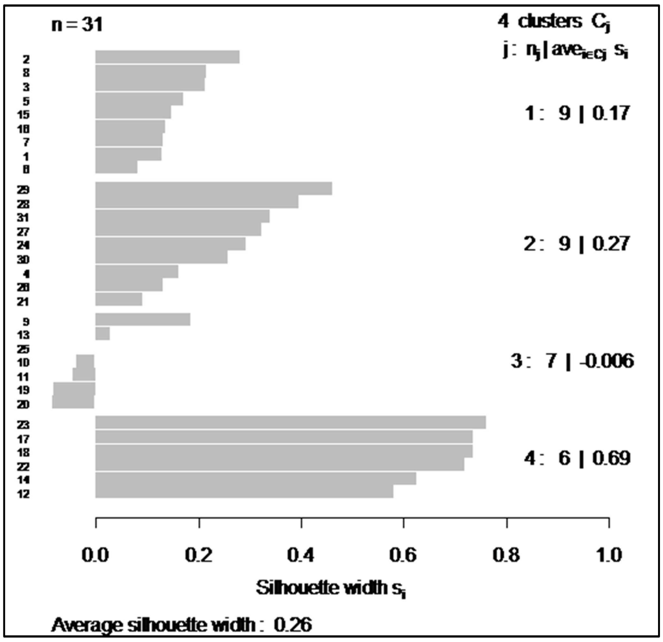
```
Clustering vector:
Beijing   Tianjin    Hebei      Shanxi         Inner Mongolia
1          2          2          3              2
Liaoning   Jilin   Heilongjiang    Shanghai        Jiangsu
2          2       2               1              4
Zhejiang   Anhui   Fujian     Jiangxi       Shandong
4          5       4          5             2
Henan     Hubei      Hunan     Guangdong      Guangxi
2         5          5          4              6
Hainan  Chongqing    Sichuan    Guizhou        Yunnan
6        5           5           6              6
Tibet   Shaanxi     Gansu     Qinghai        Ningxia
7       3           3         3              3
Xinjiang
3
```

```
Objective function:
    build      swap
0.8551363 0.8469568
```

I ran the PAM cluster analysis based on the PAM algorithm several times, chosing a different number of clusters everytime with a different output and a graphical display.

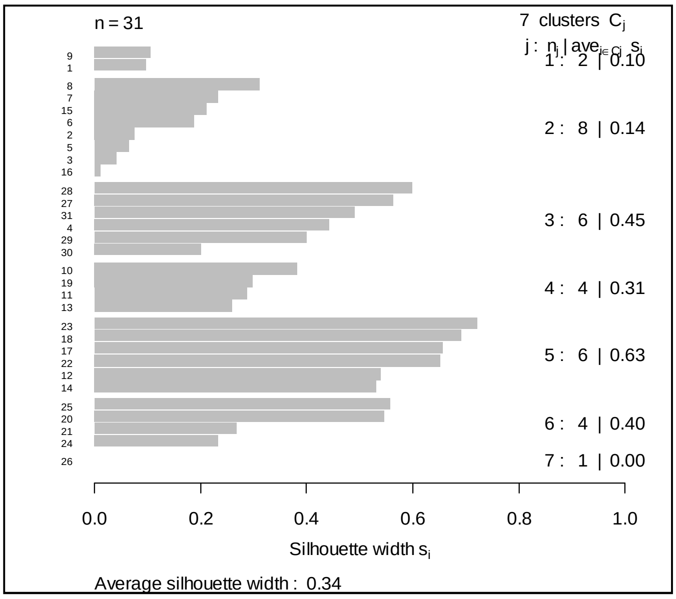
Figure 16 shows the silhouette for the first option with the number of clusters equal to 4

Figure 12 - The silhouette plot of PAM, 4 clusters



Source: Our processing with the help of R software

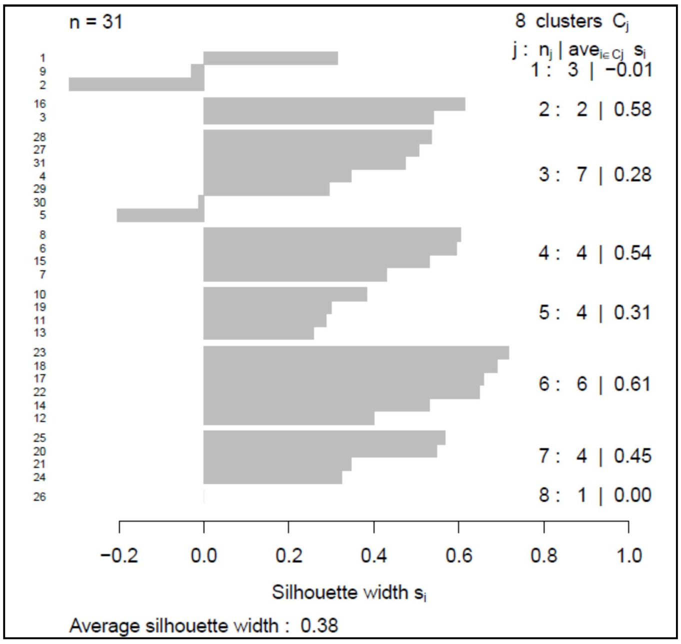
Figure 13 - The silhouette plot of PAM, 7 clusters



Source: Our processing with the help of R software

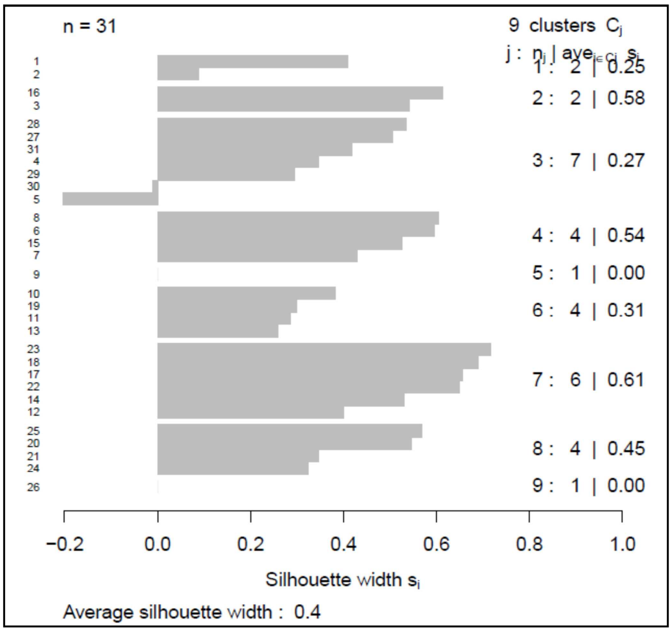
The main important information is the average silhouette width that, as explained, is the avarege of the s (i) for all objects belonging to a certain cluster. It helps to identify clear-cut clusters against weak clusters: the ones with larger average silhouette width are more pronounced (Kaufman, Rousseeuw 1990).

Figure 14 - The silhouette plot of PAM, 8 clusters



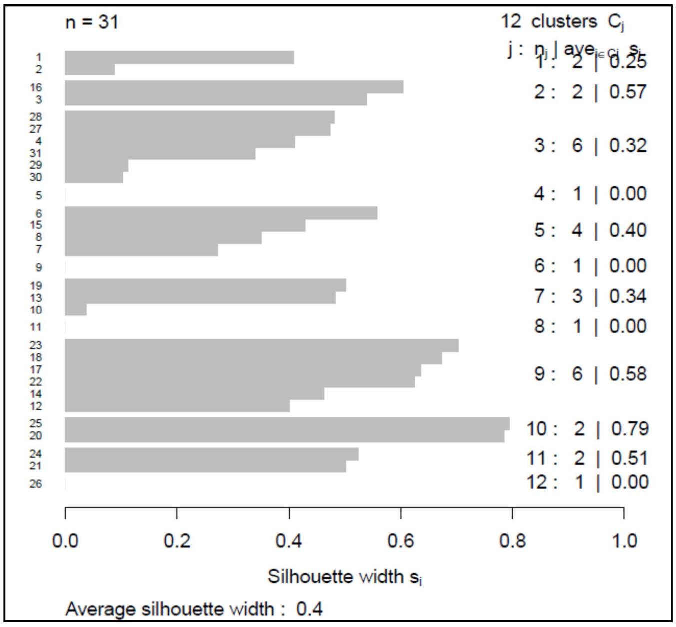
Source: Our processing with the help of R software

Figure 15 - The silhouette plot of PAM, 9 clusters



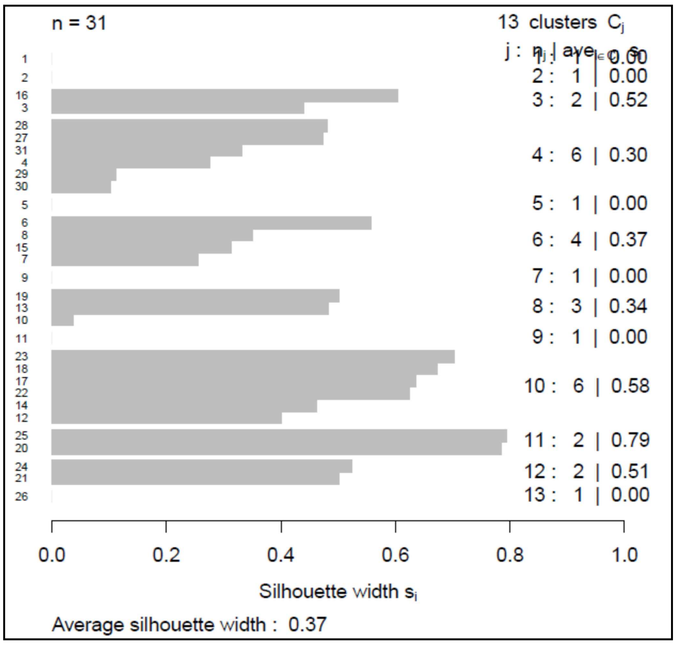
Source: Our processing with the help of R software

Figure 16 - The silhouette plot of PAM, 12 clusters



Source: Our processing with the help of R software

Figure 17 - The silhouette plot of PAM, 13 clusters



Source: Our processing with the help of R software

The results that I obtained show quite clearly that the most reasonable number of clusters goes from 9 to 12.

For a number of clusters lower than 9 or higher than 12, we have an SC lower than 0.4, which the SC observed for a number of clusters between 9 and 12. In a previous paragraph, I in fact explained that a weak SC means that the number of clusters is not sufficiently fair.

However, even for 9 or 12 clusters the results can not be easily interpreted and the best option seems to be choosing additional methods of data analysis.

The results so far reported show, in fact, that it is not possible to group Chinese provinces on the basis of the differences in the level of infrastructure and services accessible to rural villages. This can be interpreted as a confirmation that there are some significant differences among Chinese provinces for some variables but not for all of them.

However, generally speaking, it can be interesting to analyse what the level of development in terms of infrastructure and facilities in the villages of Chinese rural areas is. The further analysis carried out in the following paragraph of this thesis will add more information about that.

In particular, it is important to evaluate whether the level of development of rural villages in China is good enough or not.

3.4. Fuzzy Logic applied to the analysis of living conditions in the rural villages of China

In the previous paragraphs, I presented the first results about the analysis of living conditions in the rural villages of China based on the two methodologies of Principal Component Analysis and Cluster analysis. In particular, I was intent on showing the distribution of infrastructure in the rural villages by taking into consideration the fact that several policies have been implemented in the last few years in order to support the economic development of this part of China.

Here below, I will explain in more detail the additional analysis that I carried out for describing the availability of services and the level of infrastructure development in rural areas, and then the disparities within Chinese provinces. However, the argument that I want to analyse presents some common elements of those contexts, usually characterized by a much deeper complexity and a high correlation among all the selected variables. This makes it difficult to create a model describing the context (Zadeh 1968).

Indeed, the analysis of the level of development in terms of infrastructure and facilities in the rural villages of Chinese provinces can be considered as an example of a measurement problem with such an intrinsic complexity that is not easy to preserve it with other techniques.

The aim of this part of the thesis is, therefore, to approach this argument through Fuzzy Logic, considered increasingly to be a valid alternative to traditional techniques of analysis for constructing non linear models based on heuristic information (Passino, Yurkovich 1997), combined with the Fuzzy Experts System.

Many models assume that the phenomenon we want to study is linear. We know that this is a strong assumption with several limitations and that it is reasonable to consider this assumption unrealistic and wrong.

The Fuzzy Logic technique can make the relationship among the selected variables explicit through a set of rules, while keeping the complexity of the measuring problem. It is a mathematical approach more similar to human logic, since it does not provide any cut-offs that strictly classify the observations in well defined values. As for Fuzzy clustering, Fuzzy Logic does not assume that there are distinct classes but allows each observation to be linked to its probability of belonging to a class (Oksanen 2012). In the meantime, this technique provides a crisp indicator that can be disaggregated in relevant intermediate variables (Addabbo, Facchinetti, Pirotti 2011).

Moreover, the advantage of the Fuzzy Logic approach is that by using it, it is possible to capture all the range of possible values that could be observed or included.

In the classical approach, every observation is included in one specific category and it is not possible to put the stress on the nuances. This means that in the classical approach every observation can be accepted as a member of one class or another, and it is not possibile to explain the cases when the same observation is classified in more than one class. This is exactly what the Fuzzy Logic approach enables us to do. Through the concept of the membership function, it is possible to grade the membership of one observation into several different classes. classical

The fundamental difference between Fuzzy Logic and the classical logic approach is the definition of the membership functions in those sets. In fact, in Fuzzy Logic this function can assume more values than in the classic approach where the codomain is represented only by two values, 0 and 1, while in Fuzzy Logic those values are the extremes of an interval [0,1].

This concept can be formalized in mathematical terms:

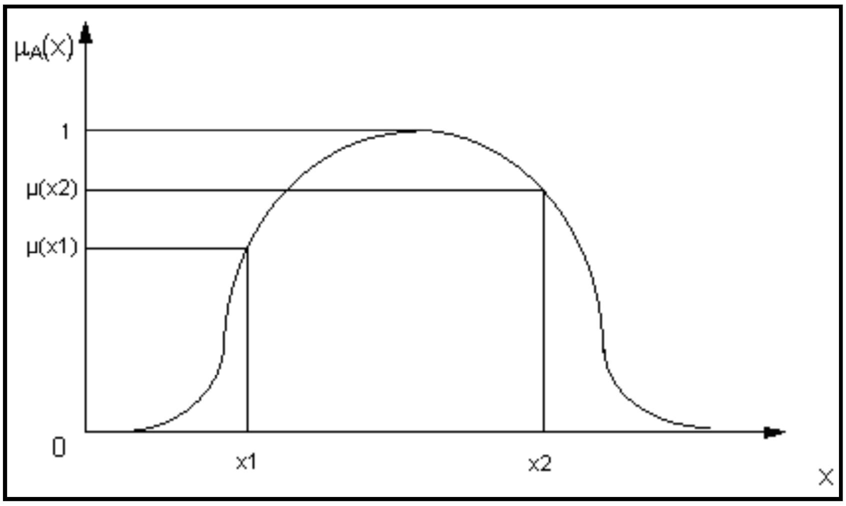
$$A = \{(x, \mu_A(x)) : x \in A, \mu_A(x) \in [0,1]\}$$

Where $\mu_A(x)$ is the membership function of the set called A. The membership function indicates how much x is member of the set A. Therefore:

- If $\mu_A(x) = 0$ then X does not belong to set A;
- If $\mu_A(x) = 1$ then X is completely in set A;
- If $\mu_A(x) = y$ then X is member of set A inasmuch as the value of "y"

In the figure below is the function of membership of a generic fuzzy set.

Figure 18 - The function of membership



The inferential process, in fact, consists of several phases, starting with the fuzzification of the inputs; that is, the transformation of the original variables into fuzzy numbers. This transformation is based on the definition of the rules and application of logic operators, which is the second step of the analysis. It follows the significance from the “antecedent” to the “consequent” and then the aggregation of the output. In fact, the fuzzification of original variables, which means the creation of the membership function, can be done following more that one technique. A first method is called total fuzzy and relative, and it strictly depends on the distribution of the variable in the reference population (Cheli, Lemmi 1995). In the methodology choosen here, the identification of membership functions is based on experience and existing literature, thus it is called the Fuzzy Experts System (FES). The rules are dictated by “experts” responsible for aggregating the original variables into intermediate ones and then to an output variable.The aggregation process is, therefore, based only on linguistic rules and attributes and, above all, is not at all data dependent since there is no knowledge of the data in advance. The Fuzzy Experts System is defined as a “function approximator”, able to explain the relationship between the input and output variables (Addabbo, Facchinetti, Pirotti 2011).

The final phase is the defuzzification of the results. In the next paragraphs, each single step of the analysis will be described and the final results obtained. These I will finally present based on a geographical perspective in order to highlight the notable disparities throughout the Chinese territory.

This will allow me to compare these results with the results from the first part of my analysis.

3.4.1. The input variables and the decision tree

In this paragraph, I will introduce the orginal variables selected for the analysis. As explained in the previous part of the study, I have selected some variables from the Second National Agricultural Census that refers to infrastructure in order to better describe the level of development of the availability of infrastructure and services.

With regards to the limits of the data, I cross-refer to the paragraph dealing with the data used in this study.

Here, I will just present the original variables that were selected for this part of the study in more detail. However, they are the same as in the previous parts of the analysis.

In this paragraph, I added some more considerations and more information to make the analysis included in this part of the thesis clearer.

Firstly, I listed in the table here below the names of the variables, together with the labels, in order to make clearer what each variable refers to. From here on the variables will be mentioned in the analysis only through their names.

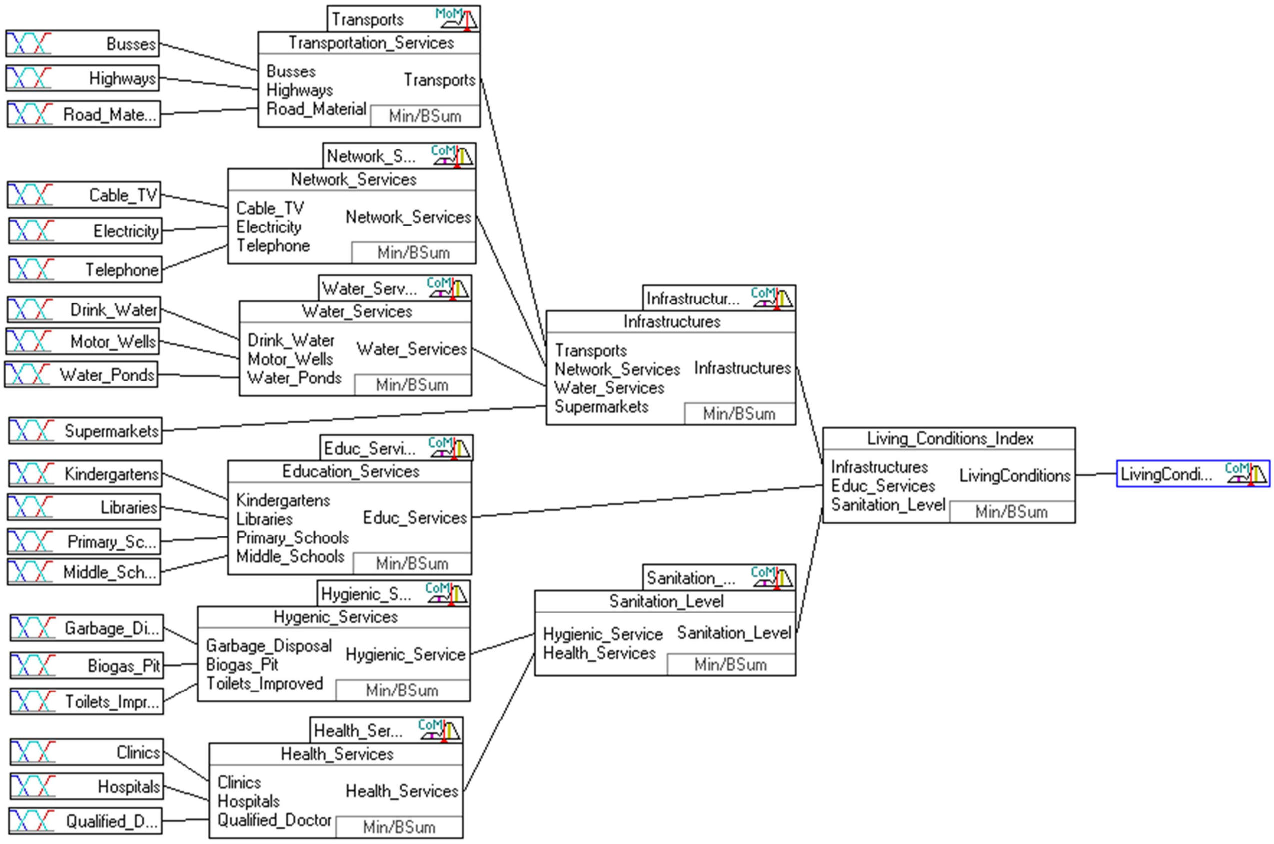
Table 16 - Labels and names of variables used in the analysis

Variable labels	***Variable name***
Villages with Highway	Highways
Closest Bus Stat/Port to Vil Comm 0km	Buses
Road Material Cement&Tar (in village)	Road_Material
Villages with Electricity	Electricity
Villages with Tel Network	Telephone
Villages with Kindergarten(s)	Kindergartens
Villages with Library(ies)	Libraries
Villages with Cable TV	Cable_TV
Closest Primary School to Vil Comm 0km	Primary_Schools
Closest Mid School to Vil Comm 0km	Middle_Schools
Villages with Purified Drinking-Water	Drinking_Water
Villages with Garbage Disposal Sites	Garbage_Disposal
Villages with Bio gas Pit	Bio gas_Pit
Villages with Improved Toilets	Improved_Toilets
Villages with Clinics	Clinics
Villages with Qualified Doctors	Qualified_Doctor
Closest Clinic/Hosp to Vil Comm 0km	Hospitals
Villages with shops/supermarkets >50m2	Supermarkets
with Motored Wells	Motor_Wells
with water Ponds/Reservior	Water_Ponds

In this study, both inputs and outputs of the fuzzy system are real numbers whose relations pass through the definition of fuzzy sets. The original inputs have to be fuzzified; that is, they are transformed into fuzzy sets, which are elaborated on the basis of the fuzzy rules that have been previously outlined, in order to produce fuzzy conclusions. Finally, through defuzzification, those fuzzy conclusions are converted into new real numbers which are the outputs of the fuzzy system.

In the Fuzzy Logic approach the first element is to design a “decision tree” consisting of the Fuzzy Expert System which better represents the argument of our analysis. The last element, on the right side of the tree, is the final output and the first elements on the left side are the inputs of the FES. The branches between the inputs and the ouput are the intermediate outputs of the exert system.

Figure 19 - Structure of the Fuzzy Logic System



Source: our processing

The final output in this decision tree is the index that I have named Living Conditions Index (LCI). This is the main result of this part of my analysis based on the FES. The LCI is connected to the inputs of the system trhough the intermediate outputs and the branches of the tree, and they are fuzzy variables. Those variables are as important as the final output, since they are indexes themselves that add useful information to the analysis. In fact, by defuzzifying those intermediate indexes, I obtained new values which are real numbers, from 0 to 1. Those indexes always refer to some main aspect of living conditions, such as the level of education, hygienic, health, water and network services, and the condition of the transport system. They are a first result of the aggregation of the original variables but still not as synthetic as the final output, the LCI.

3.4.2.The fuzzyfication

This paragraph introduces all the linguistic variables, their definition and all the membership functions related to them. Here, it will be explained how the input variables have been transformed into “fuzzy sets”. Previously, I said that fuzzyfication consists of translating the original variables into groups of sets defined by linguistic variables, which are not real numbers but linguistic terms used to translate the original real values into linguistic values. Each group has a specific membership function which has already explained in the/a previous paragraph. It goes from 0, which means no membership, to 1, meaning the maximum level of membership, and it can assume all the values included in this range from 0 to 1.

More precicely, I will now present how the 21 original variables will be progressively transformed and finally reduced to 9 output variables. This transformation will be executed following the rule blocks that have been outlined in order to control this part of the analysis. Besides those rule blocks, it has been necessary to define the relative membership functions, of which there are 113, in total. The system structure identifies the fuzzy logic inference flow from the input variables to the output variables. The fuzzification in the input interfaces translates analogue inputs into fuzzy values. The fuzzy inference takes place in rule blocks which contain the linguistic control rules. The output of these rule blocks are linguistic variables. I will talk more in depth about this in following paragraph concerning the inference process.

The defuzzification in the output interfaces translates those linguistic variables into analogue variables.

The following tables list all the system variables as well as the respective fuzzification or defuzzification method. The properties of all the base variables and the term names are also listed.

In table 17 is the list of input variables with an explanation of the term names referring to all of them and their minimum and maximum level.

Table 17 - Variables of Group "Inputs"

Variable Name	Unit	Min	Max	Term Names
Biogas_Pit	Percentage	0	1	low medium high
Buses	Percentage	0	1	low medium high
Cable_TV	Percentage	0	1	low medium high
Clinics	Percentage	0	1	low medium high
Drinking_Water	Percentage	0	1	low medium high
Electricity	Percentage	0	1	low medium high
Garbage_Disposal	Percentage	0	1	low medium high
Highways	Percentage	0	1	low medium high
Hospitals	Percentage	0	1	low medium high
Kindergartens	Percentage	0	1	low medium high
Libraries	Percentage	0	1	low medium high

Variable Name	Unit	Min	Max	Term Names
Middle_Schools	Percentage	0	1	low medium high
Motor_Wells	Percentage	0	1	low medium high
Primary_Schools	Percentage	0	1	low medium high
Province_N	Units	1	31	
Qualified_Doctor	Percentage	0	1	low medium high
Road_Material	Percentage	0	1	low medium high
Supermarkets	Percentage	0	1	low medium high
Telephone	Percentage	0	1	low medium high
Improved _Toilets	Percentage	0	1	Low medium high
Water_Ponds	Percentage	0	1	low medium high

Source: table resulting from the fuzzy logic analysis run with the support of fuzzyTECH software.

Table 18 - Variables of Group "Outputs"

#	Variable Name	Unit	Min	Max	Term Names
22	Educ_Services	Units	0	1	very_low low medium high very_high
23	Health_Services	Units	0	1	very_low low medium high very_high
24	Hygienic_Services	Units	0	1	very_low low medium high very_high
25	Infrastructure	Units	0	1	very_low low medium_low medium medium_high high very_high
26	LivingConditions	Units	0	100	term1 term2 term3 term4 term5 term6 term7 term8 term9
27	Network_Services	Units	0	1	very_low low medium

#	Variable Name	Unit	Min	Max	Term Names
					high very_high
28	Sanitation_Level	Units	0	1	very_low low medium_low medium medium_high high very_high
29	Transport	Units	0	1	very_low low medium high very_high
30	Water_Services	Units	0	1	very_low low medium high very_high

Source: table resulting from the fuzzy logic analysis run with the support of fuzzyTECH software

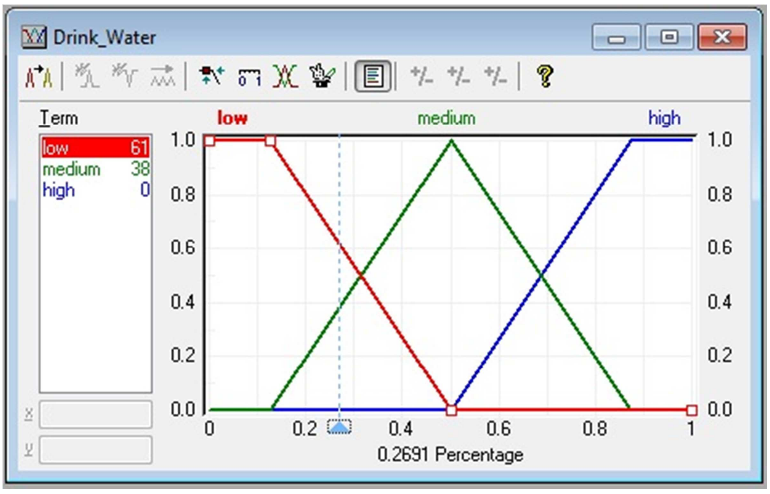
3.4.3. The inference: the rule blocks

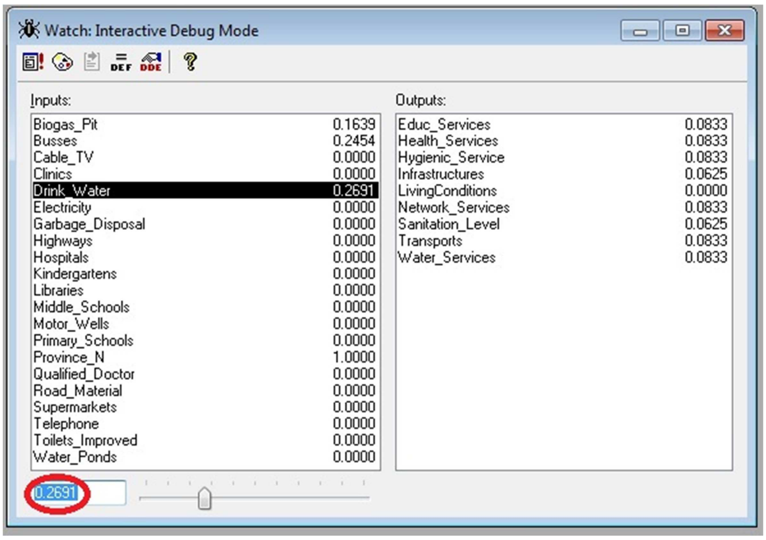
One of the previous paragraphs explained that fuzzy inference takes place in rule blocks which contain the linguistic control rules. The output of these rule blocks are linguistic variables. The defuzzification in the output interfaces translates them into analogue variables.
Each term is defined by a membership function (MBF). Each membership function defines the associated degree of membership of the linguistic term for any value of the input variable. The membership functions of all the terms of one linguistic variable are normally displayed in one graph.
Linguistic variables have to be defined for all inputs, outputs and intermediate variables. The membership functions are identified by the definition points that are the extreme values through which the linguistic terms can move. The figure here below gives one example of the graph which shows those definition points. In this graph, referring to the input variable "drinking water", there are three linguistic terms that define the variables: low; medium and high, and each of them can assume values from 0 to 1, depending on the rule of block. On the left side we can read the level of the term.

Table 19 - Definition points of Membership Function of input variables "drinking water"

Term Name	**Shape/Par.**	**Definition Points (x, y)**		
Low	linear	(0, 1) (1, 0)	(0.1275, 1)	(0.5, 0)
Medium	linear	(0, 0) (0.8725, 0)	(0.1275, 0) (1, 0)	(0.5, 1)
High	linear	(0, 0) (1, 1)	(0.5, 0)	(0.8725, 1)

Figure 20 - Graph with the definition points of linguistic terms (picture of the output fuzzyTECH software)

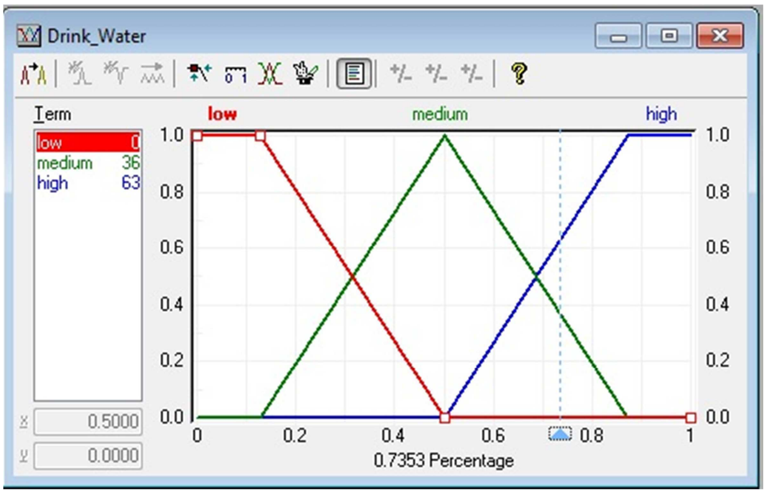


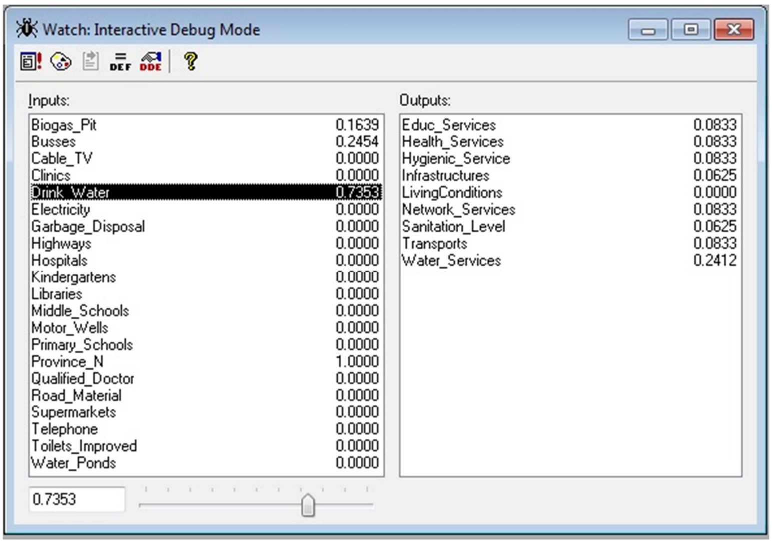


The figure below shows what happens if the value of the input variable is changed. As we can see, there is a change in the rules that will be activated.

In the first example, with a value of 0.27, the rules activated are low (0.61) and medium (0.38) while high is not activated.

Figure 21 - Graph with the definition points of linguistic terms (picture of the output fuzzyTECH software)





In the second example, here above, with a different value for the input variable (0.74), the rules activated are medium (0.36) and high (0.63).

Once the fuzzification of the original variables has been done, it is important to outline the rule blocks containing the control strategy of a fuzzy logic system. They are sets of rules “if-then” where the “if” describes the situation while the “then” part describes the response of the fuzzy system in this situation.

The two methods of aggregation are specified in the rule block: the first will be used for the aggregation of the “if” part, i.e. the antecedent; the second method for the aggregation of the “then”, the consequent.

The mathematical operator in the first part of the block is the “AND”, which consists of choosing the Minimum among all the values.

In the second part, I chose to use the BSUM (Bounded Sum) which is a sum of all the values up to 1; that is, the maximum value that can be reached.
The table with all the rule blocks is in the appendix of this thesis, here below is just an extract of it as an example.
In the rule block each variable can be considered more or less relevant in terms of the effect it can have on the output. In this study, given the characteristics of the selected variables, I proceeded by assigning the same relevance to all the variables and then by considering their effect to be always positive and equal to +2. In this scheme, in fact, it could be possible to assign a sort of weight that could be "1", "2" in both the two directions (positive or negative) or "0".

Electricity +2, Telephone +2, Cable TV +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	3
Number of Outputs:	1
Number of Rules:	27

Table 20 - Rules of the Rule Block "Network_Services"

IF			**THEN**
Cable_TV	**Electricity**	**Telephone**	**Network_Services**
Low	Low	low	very_low
Low	Low	medium	very_low
Low	Low	high	low
Low	Medium	low	very_low
Low	Medium	medium	low
Low	Medium	high	medium
Low	High	low	low
Low	High	medium	medium
Low	High	high	high
Medium	Low	low	very_low

IF			THEN
Medium	Low	medium	low
Medium	Low	high	medium
Medium	Medium	low	low
Medium	Medium	medium	medium
Medium	Medium	high	high
Medium	High	low	medium
Medium	High	medium	high
Medium	High	high	very_high
High	Low	low	low
High	Low	medium	medium
High	Low	high	high
High	Medium	low	medium
High	Medium	medium	high
High	Medium	high	very_high
High	High	low	high
High	High	medium	very_high
High	High	high	very_high

Each rule block is characterized by a logic relationship, expressed through a mathematical operator of aggregation.

3.4.4. The defuzzification

The results of the rule of block are fuzzy numbers that are not real values but linguistic values, numbers defined by terms and their level of activation.

For this reason, the fuzzy logic analysis ends with the defuzzification which consists of once again transforming the fuzzy values into "crisp" number; that is, into real values.

Several Defuzzification Methods can be adopted: Center of Maximum (CoM); Mean of Maximum (MoM); Center of Area (CoA); Hyper CoM; Fuzzy Output Force; Categorical MoM.

The one that I used in my analysis is the Center of Maximum (CoM).

Those different methods can result either in to the "most plausible result" or the "best compromise". The "best compromise" is produced by the methods: CoM (Center of

Maximum), CoA (Center of Area) and the CoA BSUM. In the first method (CoM) the typical numerical value for each scaled membership function is firstly determined; that is, the mean of the numerical values corresponding to the degree of membership at which the membership function was scaled.

$$X_{final} = \frac{(x_1\mu_1 + x_2\mu_2 + .. x_n\mu_n)}{(\mu_1 + \mu_2 + .. \mu_n)}$$

where x_n is the typical numerical value for the scaled membership function n, and μ_n is the degree of membership at which membership function n was scaled.

The CoA firstly reduces the domain of each linguistic term through Alpha-cut where $\alpha=\mu_i$ (x) and then calculates the integral of the areas under the Alpha-cuts. The "most plausible" result is produced by the methods: MoM (Mean of Maximum), Categorical MoM (Categorical Mean of Maximum), MoM BSUM, a version especially for efficient VLSI implementations. The first method selects the typical value of the most valid output linguistic term.

The defuzzification method used in this analysis is the CoM since this method provides results that are more robust and balanced than for the other methods. It is the balanced average of the values related to the maximum values of each term and its activation value.

3.4.5. The Living Condition Index (LCI) in rural villages of China

On the basis of the results of this analysis important considerations can be made. First, it can be observed that for most of the Chinese provinces the level of the main services and transport in rural villages is low as highlighted by the data in table here below (table 21). In many provinces the level of the main services and transport does not reach values that can be considered positive. For instance, with regards to Education services, four provinces (Guizhou, Tibet, Qinghai and Ningxia) registered an index equal to "0". For Hygienic services, this figure increases to eight provinces (Inner Mongolia Jilin Heilongjiang Tibet Gansu Qinghai Ningxia Xinjiang), often provinces with a significantly high level of urbanization.

Table 21 - Defuzzification of intermediate indexes

Province	**Education Services**	**Hygienic Services**	**Health Services**	**Water Services**	**Network Services**	**Transport**
Beijing	0.189	0.400	0.521	0.696	1	1
Tianjin	0.041	0.339	0.520	0.387	1	0.5
Hebei	0.124	0.062	0.654	0.326	0.942	0.75
Shanxi	0.169	0.031	0.458	0.016	0.980	0.5
Inner Mongolia	0.021	0	0.527	0.127	0.920	0.25
Liaoning	0.143	0.008	0.75	0.185	1	0.5
Jilin	0.077	0	0.632	0.137	1	0.75
Heilongjiang	0.132	0	0.720	0.288	1	0.5
Shanghai	0.001	0.75	0.745	0.25	1	0.75
Jiangsu	0.159	0.153	0.75	0.429	1	0.5
Zhejiang	0.078	0.352	0.252	0.265	1	0.75
Anhui	0.103	0.015	0.75	0.245	0.934	0.25
Fujian	0.197	0.195	0.688	0.095	1	0.75
Jiangxi	0.171	0.102	0.710	0.142	1	0.5
Shandong	0.030	0.009	0.535	0.532	1	0.75
Henan	0.164	0.110	0.75	0.236	0.951	0.5
Hubei	0.026	0.021	0.633	0.078	1	0.5
Hunan	0.043	0.005	0.517	0.144	0.980	0.25
Guangdong	0.138	0.260	0.682	0.154	1	0.5
Guangxi	0.155	0.329	0.649	0	1	0.25
Hainan	0.116	0.273	0.304	0.099	0.915	0.25
Chongqing	0.065	0.033	0.656	0.191	1	0.25
Sichuan	0.047	0.06	0.510	0.193	1	0.25
Guizhou	0	0.101	0.373	0	0.839	0.25
Yunnan	0.155	0.271	0.75	0.013	0.985	0.25
Tibet	0	0	0	0	0.181	0

Shaanxi	0.045	0.031	0.481	0.000	0.924	0.25
Gansu	0.045	0	0.509	0.017	0.839	0.25
Qinghai	0	0	0.460	0	0.657	0.25
Ningxia	0	0	0.710	0.039	0.786	0.5
Xinjiang	0.052	0	0.253	0.144	0.860	0.25

Source: Our processing

The table here below shows the intermediate indeces which are the last step of aggregation in the decision tree before calculating the final Living Conditions Index (see figure 19).

The information gathered from those indeces are useful as the previous ones (table 21). In particular Tibet is confirmed to be the province with the worse condition both in terms of infrastructure and sanitation level. For some provinces it is interesting to note how the two indexes reach level significantly different between each other. For instance in the province of Shandong we observe the infrastructure index seems to be much higher (0.67) than for sanitation services (0.20).

In the second to last step before the final output it was possible to aggregate these intermediate indexes in order to obtain two more intermediate indexes, more synthetic than the previous one but still not as synthetic as the final LCI.

Table 22 - Defuzzification of intermediate indeces (last step before final output)

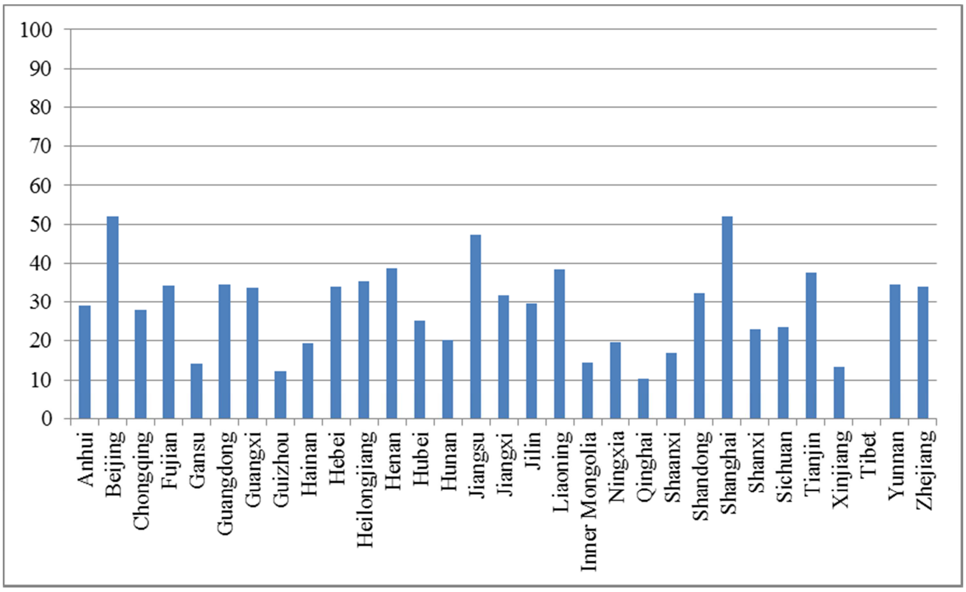
Province	**Infrastructure**	**Sanitation Level**
Beijing	0.839	0.437
Tianjin	0.628	0.392
Hebei	0.583	0.316
Shanxi	0.426	0.177
Inner Mongolia	0.390	0.185
Liaoning	0.656	0.339
Jilin	0.559	0.255
Heilongjiang	0.633	0.314
Shanghai	0.693	0.830
Jiangsu	0.742	0.435
Zhejiang	0.591	0.247
Anhui	0.473	0.343
Fujian	0.532	0.428
Jiangxi	0.442	0.370
Shandong	0.674	0.199

Henan	0.580	0.407
Hubei	0.429	0.270
Hunan	0.427	0.183
Guangdong	0.510	0.452
Guangxi	0.379	0.470
Hainan	0.338	0.230
Chongqing	0.432	0.292
Sichuan	0.432	0.225
Guizhou	0.274	0.177
Yunnan	0.416	0.514
Tibet	0.005	0
Shaanxi	0.384	0.188
Gansu	0.355	0.173
Qinghai	0.238	0.140
Ningxia	0.383	0.307
Xinjiang	0.380	0.026

Source: Our processing

The figure here below shows the results by province of the fuzzy analysis with the level of living conditions expressed as a new index which takes values between “0” and “100”. It is evident that the level of the index is higher only in few provinces but in general it can be stated that the conditions of rural villages are far from an optimal level.

Figure 22 - The Living Conditions Index (LCI) by provinces



Source: Our processing

3.4.6. Comparison between official geographical classification and the results of my analysis

The results of the cluster analysis show that the choice of group for the Chinese provinces is slightly artificial, and for this reason I further analysed the data through the mathematical approach of fuzzy logic in order to extrapolate as much information as possible from my data.

In fact, the conclusions derived from the cluster analysis do not answer the main research question behind this thesis: to what extent can we consider the level of infrastructure and services in the rural villages of China sufficiently developed?

The results from the fuzzy logic approach provide very useful information through the final synthetic index, which I have called the Living Conditions Index (LCI).

However, despite the fact there are not substantial differences between the Chinese provinces, this index allows us to detect a significant shortage of infrastructure and services in many rural Chinese villages.

If we consider that this rural context still plays a fundamental role in terms of social stability, this suggests the reason why it is very important to place the stress on these areas. Most Chinese policies focus on the development of rural areas, above all in those areas affected by a phenomenon of migration inflows after long periods of outflows.

Here are the results obtained from the fuzzy logic analysis with the LCI and the other indexes. These indexes underline the aspects that make the differences among the Chinese provinces more significant. In addition, it is interesting to compare these results with the results of the cluster analysis as detected with the PAM algorithm.

Table 23 - Indices from the fuzzy logic, spread out by cluster (4)

Cluster	Province	Education Services	Hygienic Services	Health Services	Water Services	Network Services	Transport
1	*Beijing*	0.189	0.400	0.521	0.696	1	1
	Tianjin	0.041	0.339	0.520	0.387	1	0.5
	Hebei	0.124	0.062	0.654	0.326	0.942	0.75
	Henan	0.164	0.110	0.75	0.236	0.951	0.5
	Inner Mong	0.021	0	0.527	0.127	0.920	0.25
	Heilongjiar	0.132	0	0.720	0.288	1	0.5
	Jilin	0.077	0	0.632	0.137	1	0.75
	Liaoning	0.143	0.008	0.75	0.185	1	0.5
	Shandong	0.030	0.009	0.535	0.532	1	0.75
Cluster Avarage 1		**0.102**	**0.103**	**0.623**	**0.324**	**0.979**	**0.611**
2	*Gansu*	0.045	0	0.509	0.017	0.839	0.25
	Ningxia	0	0	0.710	0.039	0.786	0.5
	Qinghai	0	0	0.460	0	0.657	0.25
	Shaanxi	0.045	0.031	0.481	0.000	0.924	0.25
	Shanxi	0.169	0.031	0.458	0.016	0.980	0.5
	Xinjiang	0.052	0	0.253	0.144	0.860	0.25
	Guizhou	0	0.101	0.373	0	0.839	0.25
	Hainan	0.116	0.273	0.304	0.099	0.915	0.25
	Tibet	0	0	0	0	0.181	0
Cluster Avarage 2		**0.047**	**0.048**	**0.394**	**0.035**	**0.776**	**0.278**
3	*Shanghai*	0.001	0.75	0.745	0.25	1	0.75
	Fujian	0.197	0.195	0.688	0.095	1	0.75
	Guangdong	0.138	0.260	0.682	0.154	1	0.5
	Jiangsu	0.159	0.153	0.75	0.429	1	0.5
	Zhejiang	0.078	0.352	0.252	0.265	1	0.75
	Guangxi	0.155	0.329	0.649	0	1	0.25
	Yunnan	0.155	0.271	0.75	0.013	0.985	0.25
Cluster Avarage 3		0.126	0.330	0.645	0.172	0.998	0.536
4	*Anhui*	0.103	0.015	0.75	0.245	0.934	0.25
	Chongqing	0.065	0.033	0.656	0.191	1	0.25
	Hubei	0.026	0.021	0.633	0.078	1	0.5
	Hunan	0.043	0.005	0.517	0.144	0.980	0.25
	Jiangxi	0.171	0.102	0.710	0.142	1	0.5
	Sichuan	0.047	0.06	0.510	0.193	1	0.25
Cluster Avarage 4		**0.076**	**0.039**	**0.629**	**0.166**	**0.986**	**0.333**

Source: our processing

Figure 23 - Indices from the fuzzy logic, spread out by cluster (7)

Cluster	Province	Education Services	Hygienic Services	Health Services	Water Services	Network Services	Transport
1	*Beijing*	0.189	0.400	0.521	0.696	1	1
	Shanghai	0.001	0.75	0.745	0.25	1	0.75
Cluster Avarage		**0.095**	**0.575**	**0.633**	**0.473**	**1.000**	**0.875**
2	*Tianjin*	0.041	0.339	0.520	0.387	1	0.5
	Hebei	0.124	0.062	0.654	0.326	0.942	0.75
	Henan	0.164	0.110	0.75	0.236	0.951	0.5
	Inner Mongolia	0.021	0	0.527	0.127	0.920	0.25
	Heilongjiang	0.132	0	0.720	0.288	1	0.5
	Jilin	0.077	0	0.632	0.137	1	0.75
	Liaoning	0.143	0.008	0.75	0.185	1	0.5
	Shandong	0.030	0.009	0.535	0.532	1	0.75
Cluster Avarage		**0.091**	**0.066**	**0.636**	**0.277**	**0.977**	**0.563**
3	*Gansu*	0.045	0	0.509	0.017	0.839	0.25
	Ningxia	0	0	0.710	0.039	0.786	0.5
	Qinghai	0	0	0.460	0	0.657	0.25
	Shaanxi	0.045	0.031	0.481	0.000	0.924	0.25
	Shanxi	0.169	0.031	0.458	0.016	0.980	0.5
	Xinjiang	0.052	0	0.253	0.144	0.860	0.25
Cluster Avarage		**0.052**	**0.010**	**0.479**	**0.036**	**0.841**	**0.333**
4	*Fujian*	0.197	0.195	0.688	0.095	1	0.75
	Guangdong	0.138	0.260	0.682	0.154	1	0.5
	Jiangsu	0.159	0.153	0.75	0.429	1	0.5
	Zhejiang	0.078	0.352	0.252	0.265	1	0.75
Cluster Avarage		**0.143**	**0.240**	**0.593**	**0.236**	**1.000**	**0.625**
5	*Anhui*	0.103	0.015	0.75	0.245	0.934	0.25
	Chongqing	0.065	0.033	0.656	0.191	1	0.25
	Hubei	0.026	0.021	0.633	0.078	1	0.5
	Hunan	0.043	0.005	0.517	0.144	0.980	0.25
	Jiangxi	0.171	0.102	0.710	0.142	1	0.5
	Sichuan	0.047	0.06	0.510	0.193	1	0.25
Cluster Avarage		**0.076**	**0.039**	**0.629**	**0.166**	**0.986**	**0.333**
6	*Guangxi*	0.155	0.329	0.649	0	1	0.25
	Guizhou	0	0.101	0.373	0	0.839	0.25
	Hainan	0.116	0.273	0.304	0.099	0.915	0.25
	Yunnan	0.155	0.271	0.75	0.013	0.985	0.25
Cluster Avarage		**0.106**	**0.243**	**0.519**	**0.028**	**0.935**	**0.250**
7	*Tibet*	0	0	0	0	0.181	0
Cluster Avarage		**0.000**	**0.000**	**0.000**	**0.000**	**0.181**	**0.000**

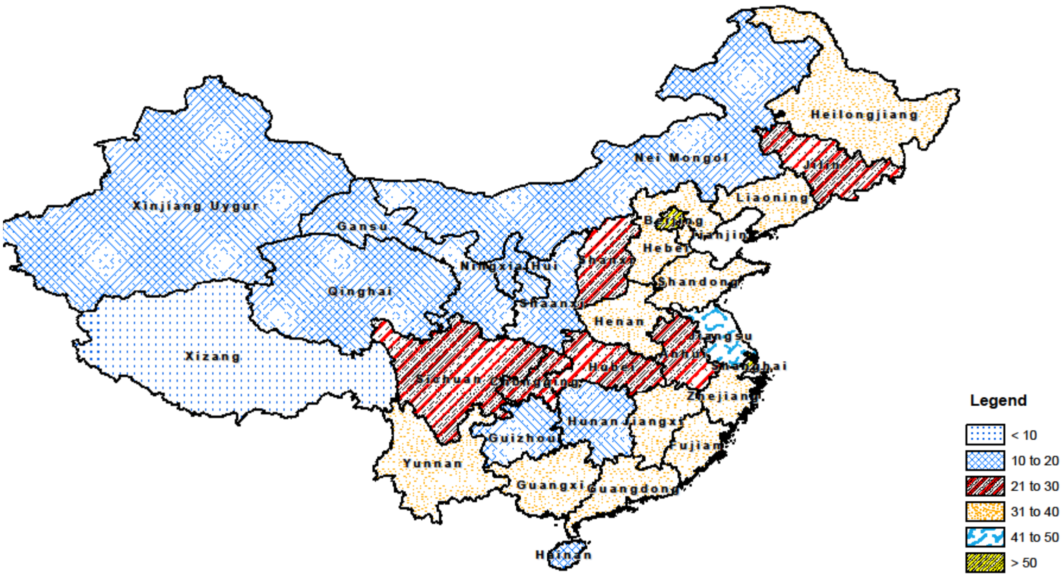
Source: our processing

The three maps here below show a similar condition, that is, a similar geographical distribution of living conditions in the rural villages of China. This is a confirmation of the results detected from the analysis: firstly, through the methodology of cluster analysis and secondly, through the fuzzy logic approach. However, these maps do not explicitly say anything about the level of development, except for the one showing the results of the fuzzy logic. They offer important information about the differences and the similarities within the Chinese provinces.

The first map here below is the geographical distribution of the LCI, the second is the map of clusters of rural villages based on the PAM algorithm with 4 clusters, while the third one is the map of results based on the PAM method with 7 clusters.[14]

The map reporting the results of the LCI shows to which class each province belongs. In fact, the first class, with the lowest level of LCI, includes only Tibet, while the last class, with the highest LCI, includes only Beijing and Shanghai. In fact, this classification reflects the general differences in the Chinese provinces, as already mentioned in previous paragraphs. Except for some provinces, this different distribution is reflected in the map based on the results of cluster analysis.

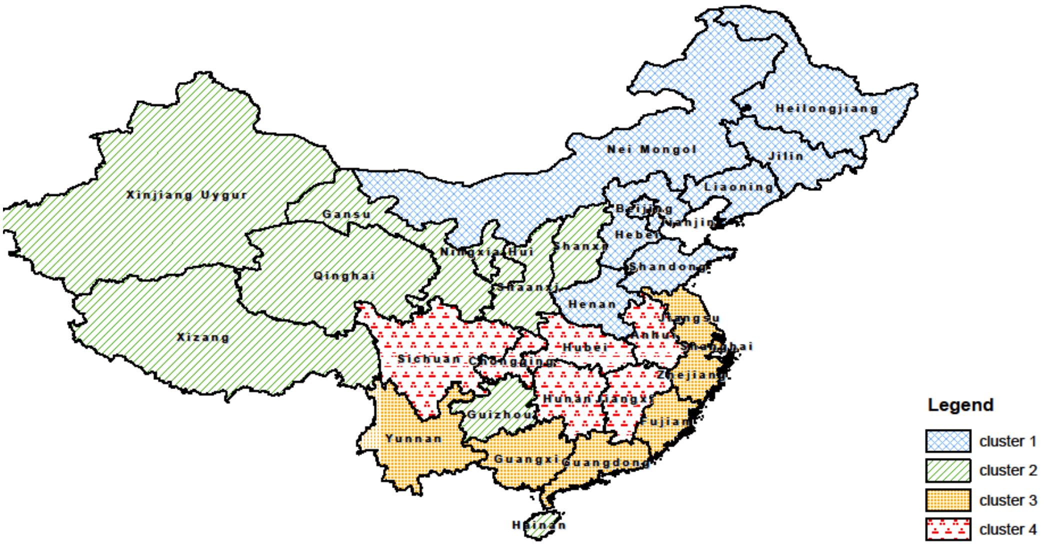
Figure 24 - The map of rural villages based on the LCI classes



Source: Author's configuration

[14] For other maps with 6 and 9 clusters see the appendix 1.

Figure 25 - The map of rural villages based on the results of the PAM algorithm (4 clusters)



Source: Author’s configuration

Figure 26 - The map of rural villages based on the results of the PAM algorithm (7 clusters)

Conclusion

This thesis was intended to put the stress on the fundamental role played by the rural infrastructure development in rural areas of China, with main focus on the variables who are related to the living conditions. This has to be considered a central issue to take into consideration not only for the effects that this can have on the agricultural production technical efficiency but even more for planning a more balanced growth among provinces and between rural and urban areas, able to assure social stability in China.

The analysis of living conditions in the rural villages of China, specifically in terms of access to services and facilities, reported in this thesis does not pretend to be an exhaustive one. Nonetheless the results described in previous paragraph offer several issues which are useful to define and improve the present policies of rural development and which can be further analysed.

First the clustering of living conditions and the Living Condition Index (LCI) proposed in this thesis allow us to conclude by saying that the geography of the social and living conditions in China is significantly different from the traditional and official classifications from the Statistical Yearbook and also from that of the two national agricultural censuses (1996 and 2006). From one side the results of clustering show there are at the same time some general and homogenous standards and very deep differences in many relevant aspects of the quality of life in the villages. While on the other side the LCI based on the fuzzy expert approach provide us with a clear snapshot of the rural China where still a lot of basic needs have to be improved. This is a fundamental aspect which has to be considered in designing future policies and actions in the country.

There is the need of different policies to avoid the permanence of differences within the rural areas and in particular compared to the fast growing of Eastern provinces and urban areas. In the next few years both phenomenon of urbanization and massive internal migration, characterizing the recent Chinese economic development, will have a decisive impact on the whole country and in particular on the small Chinese household farms in the rural areas. One of the major challenges for the future economic and social development of China will be the trade-off between the support to the economic development in urban areas and the need of specific policies of rural development in rural areas. Those areas will be characterized at the same time by a further reduction in the agricultural workforce and on the other side by a counter flow of migration. Given the role played by the small Chinese household in terms of food safety and in general of social stability it can not be denied how important will be to assure better living conditions even in the rural areas. The social cohesion of China will be possible only if the Chinese government will manage to face these challenges and if it will be reduced the tremendous disparities between sectors, regions and families moving towards a more balanced development. Thus, the interest in new policies addressed to rural areas is due to the fact that the development of these areas is essential for social stability and for a sustainable growth in the future.

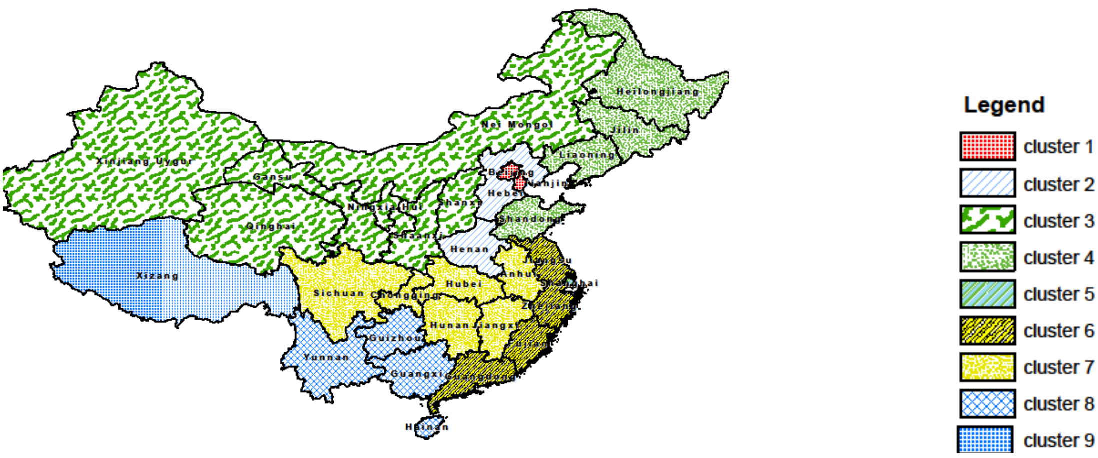
Even the problem of assured food security remains a challange for this country where the food consumption of households is constantly increasing and changing. The food security will depend not only on the increase in consumption and the productive capacity of Chinese agriculture but also on the structural changes and on the difficulties and lack of organization of the whole food production chain. Beside the aggregation of the production of millions of small farms, it will needed a more efficient organization of logistics and distribution instead of the current fragmentation with many different public and private operators, and different stages in the food chain. Even the limits emerged from the analysis are important for the ideas they offer for further in depth research. The main limit is the nature of data, as largely explained in the first paragraphs. These limits could be overtake with more disaggregated data or with more information about the distribution of the main variables utilized in the analysis of living conditions in rural China.

Appendix 1

The map of rural villages based on the results of the PAM algorithm (6 clusters)

Source: our processing

The map of rural villages based on the results of the PAM algorithm (9 clusters)



Source: our processing

Appendix 2

Input Variable "Biogas_Pit"

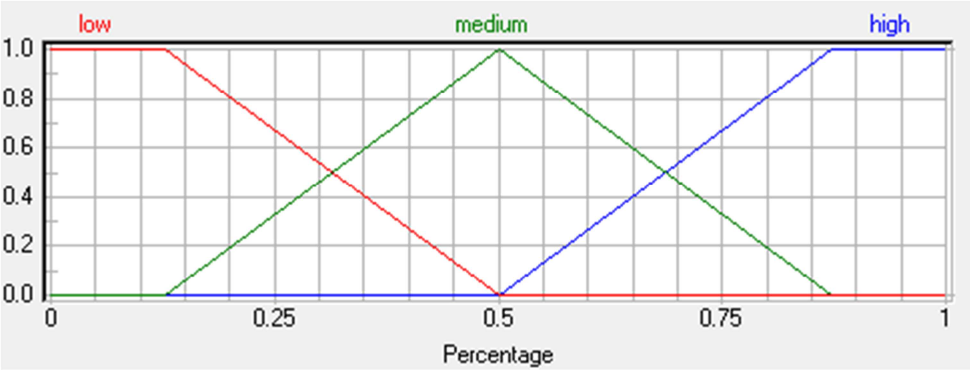


Figure 1: MBF of "Biogas_Pit"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 1: Definition Points of MBF "Biogas_Pit"

Input Variable "Buses"

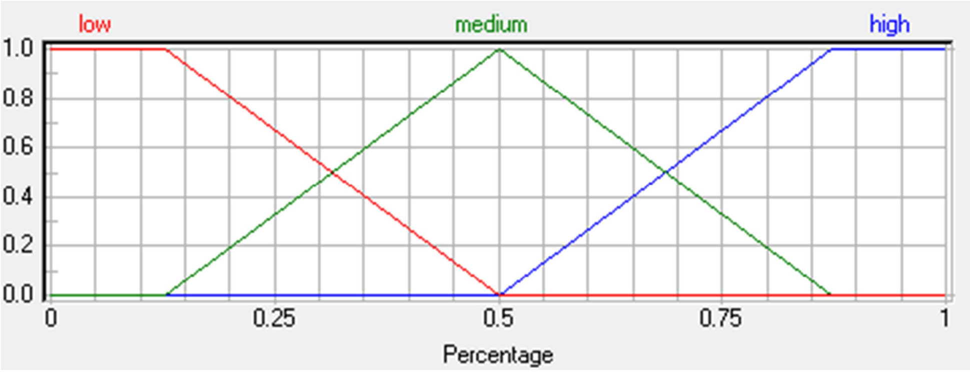


Figure 2: MBF of "Busses"

Term Name	**Shape/Par.**	**Definition Points (x, y)**		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 2: Definition Points of MBF "Buses"

Input Variable "Cable_TV"

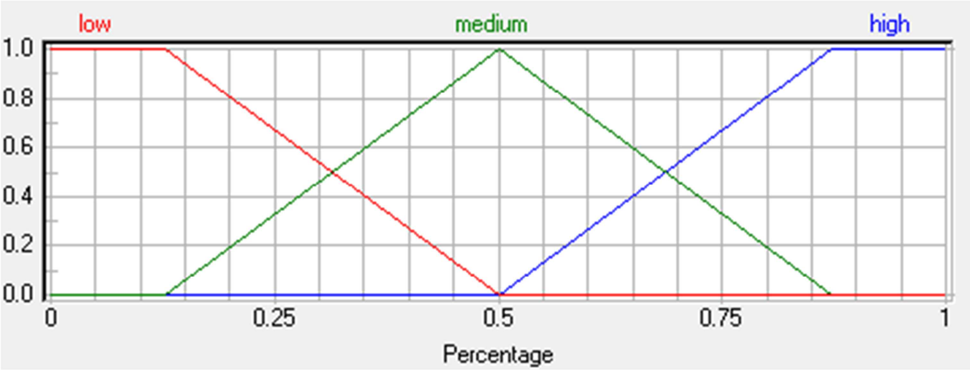


Figure 3: MBF of "Cable_TV"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 3: Definition Points of MBF "Cable_TV"

Input Variable "Clinics"

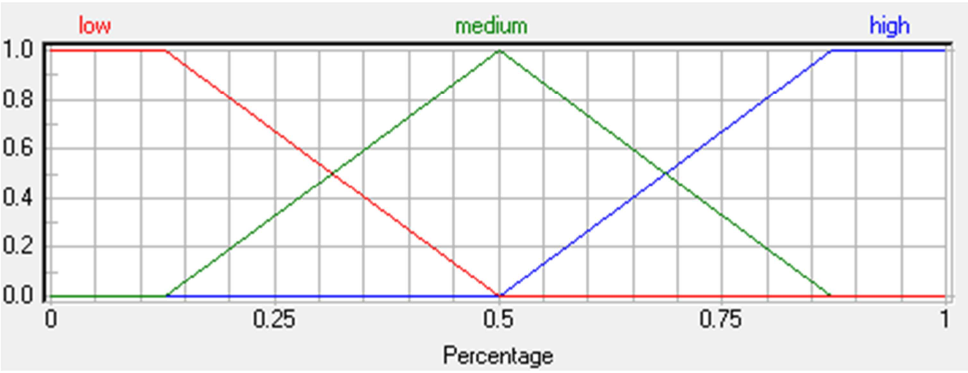


Figure 4: MBF of "Clinics"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)

Term Name	Shape/Par.	Definition Points (x, y)		
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 4: Definition Points of MBF "Clinics"

Input Variable "Drinking_Water"

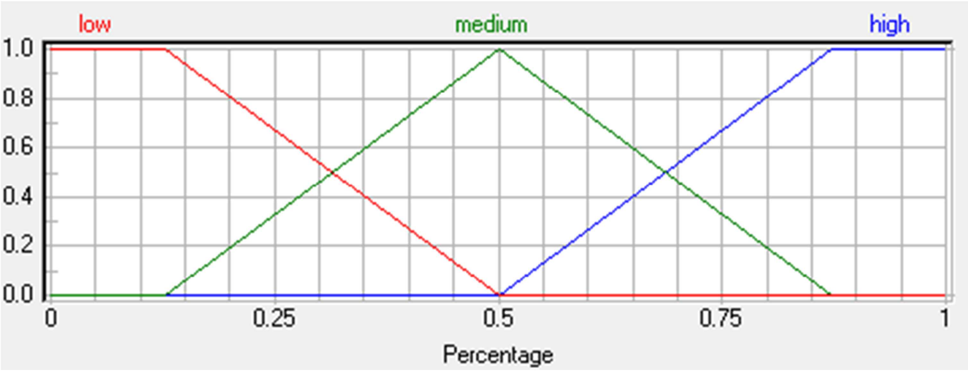


Figure 5: MBF of "Drinking_Water"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 5: Definition Points of MBF "Drinking_Water"

Input Variable "Electricity"

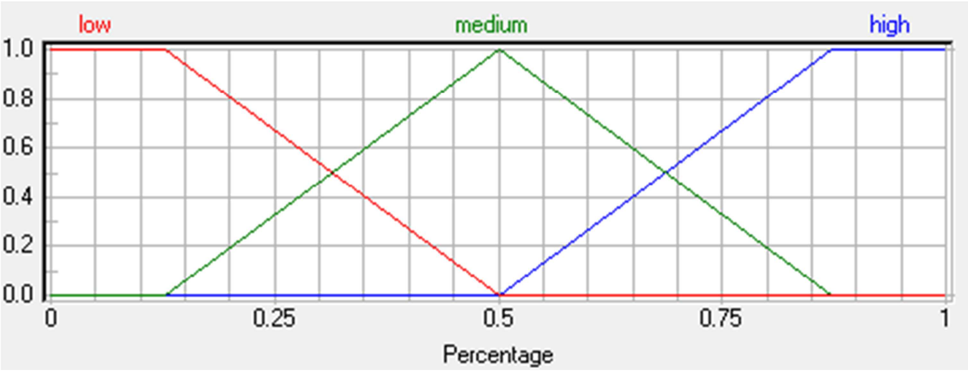


Figure 6: MBF of "Electricity"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 6: Definition Points of MBF "Electricity"

Input Variable "Garbage_Disposal"

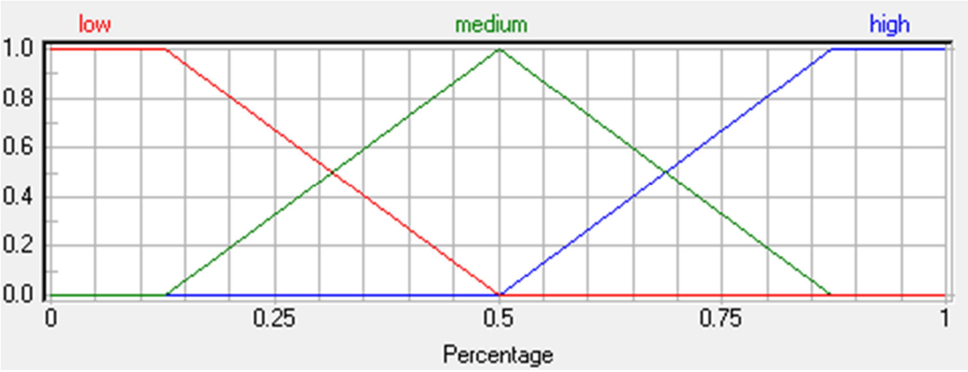


Figure 7: MBF of "Garbage_Disposal"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 7: Definition Points of MBF "Garbage_Disposal"

Input Variable "Highways"

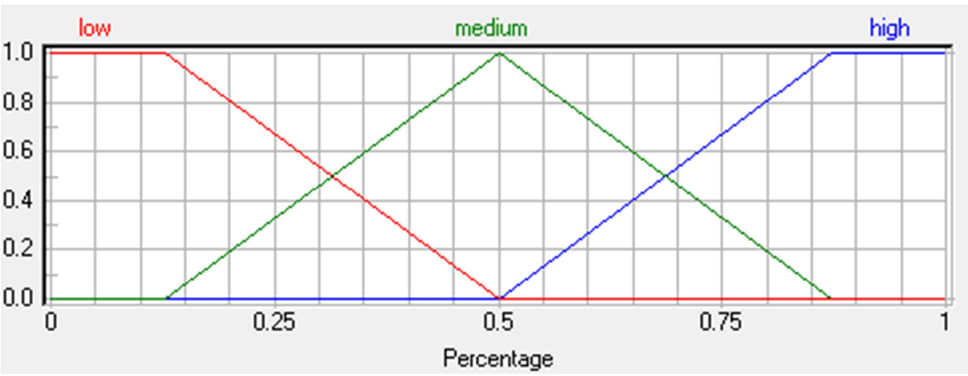


Figure 8: MBF of "Highways"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 8: Definition Points of MBF "Highways"

Input Variable "Hospitals"

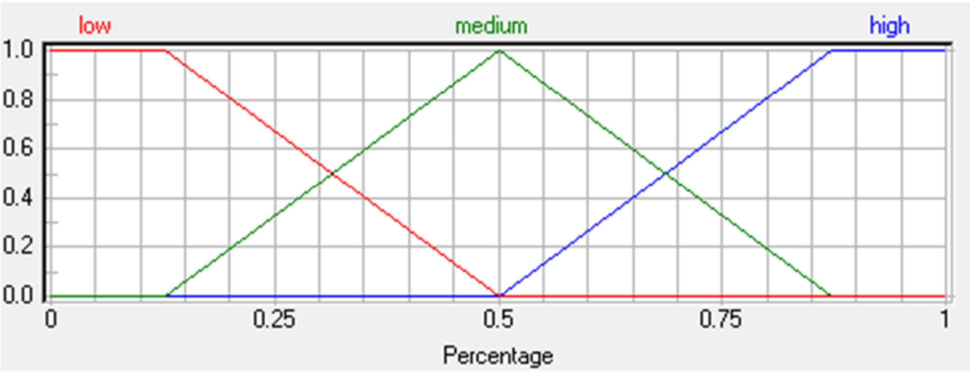


Figure 9: MBF of "Hospitals"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)

Term Name	Shape/Par.	Definition Points (x, y)		
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 9: Definition Points of MBF "Hospitals"

Input Variable "Kindergartens"

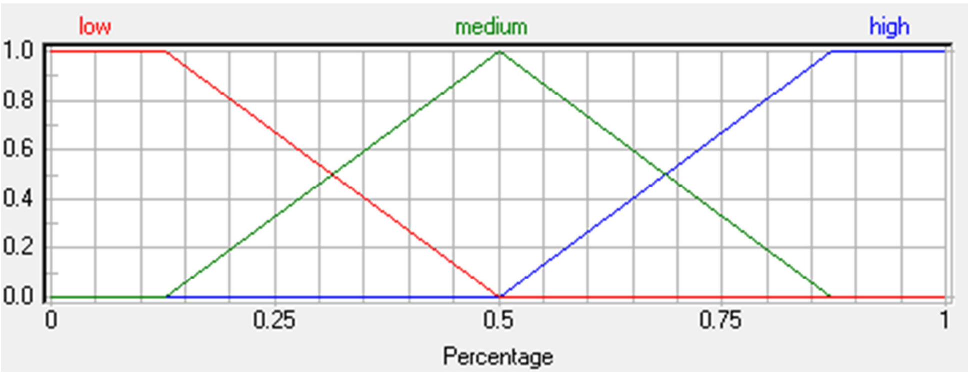


Figure 10: MBF of "Kindergartens"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 10: Definition Points of MBF "Kindergartens"

Input Variable "Libraries"

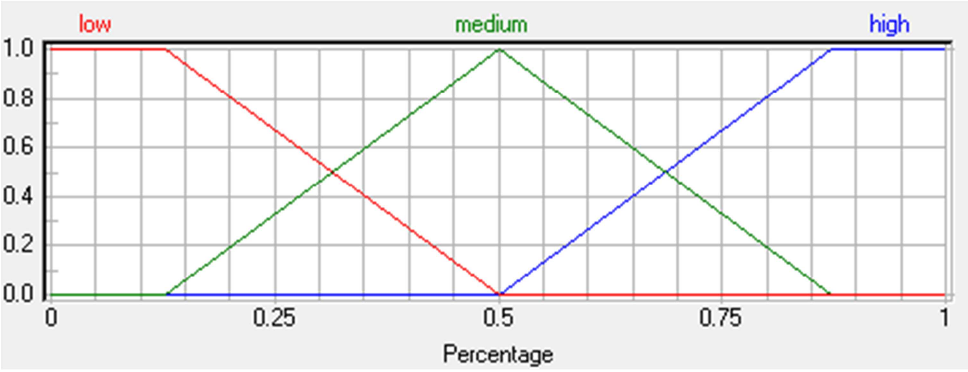


Figure 11: MBF of "Libraries"

Term Name	**Shape/Par.**	**Definition Points (x, y)**		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 11: Definition Points of MBF "Libraries"

Input Variable "Middle_Schools"

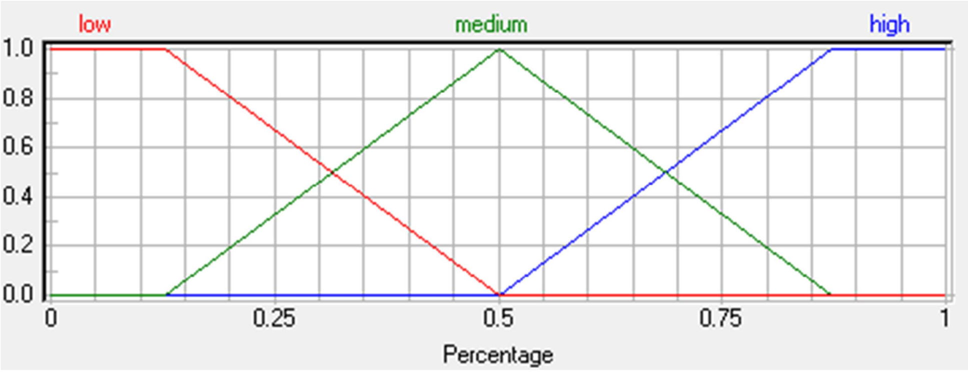


Figure 12: MBF of "Middle_Schools"

Term Name	**Shape/Par.**	**Definition Points (x, y)**		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 12: Definition Points of MBF "Middle_Schools"

Input Variable "Motor_Wells"

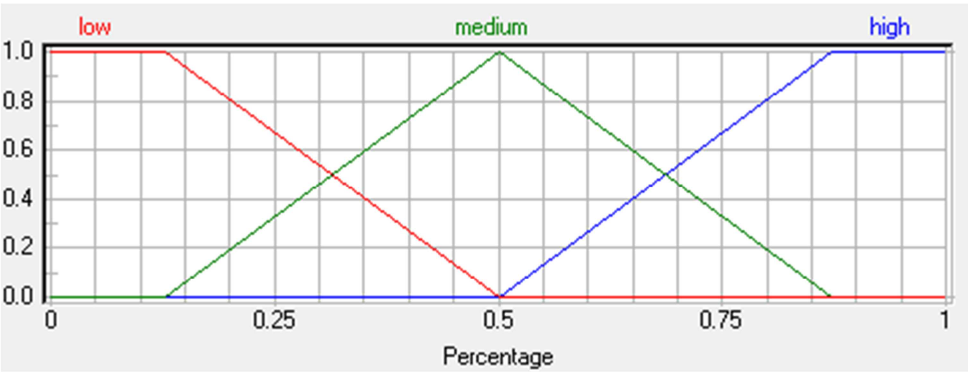


Figure 13: MBF of "Motor_Wells"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 13: Definition Points of MBF "Motor_Wells"

Input Variable "Primary_Schools"

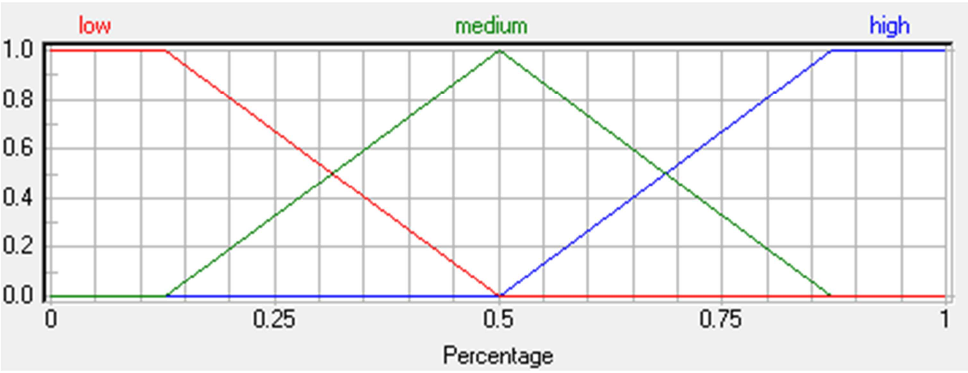


Figure 14: MBF of "Primary_Schools"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)

Term Name	Shape/Par.	Definition Points (x, y)		
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 14: Definition Points of MBF "Primary_Schools"

Input Variable "Province_N"

Input Variable "Qualified_Doctor"

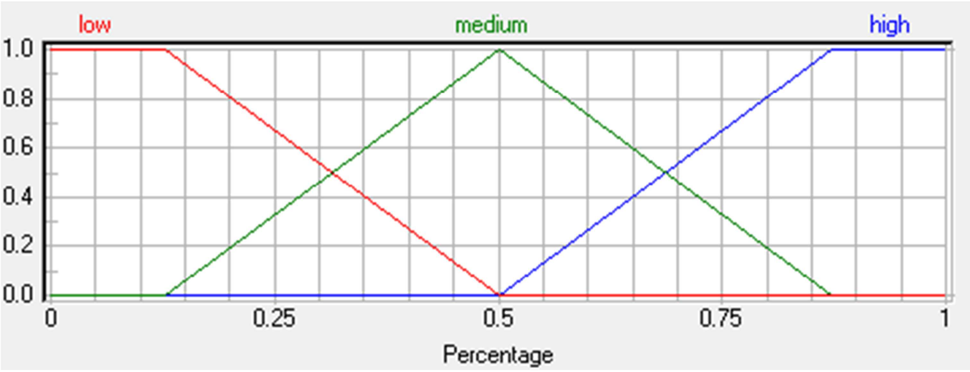


Figure 15: MBF of "Qualified_Doctor"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 15: Definition Points of MBF "Qualified_Doctor"

Input Variable "Road_Material"

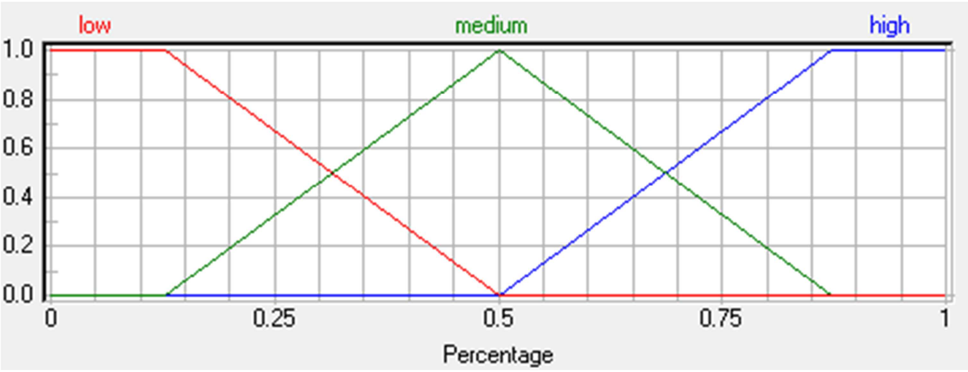


Figure 16: MBF of "Road_Material"

Term Name	**Shape/Par.**	**Definition Points (x, y)**		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 16: Definition Points of MBF "Road_Material"

Input Variable "Supermarkets"

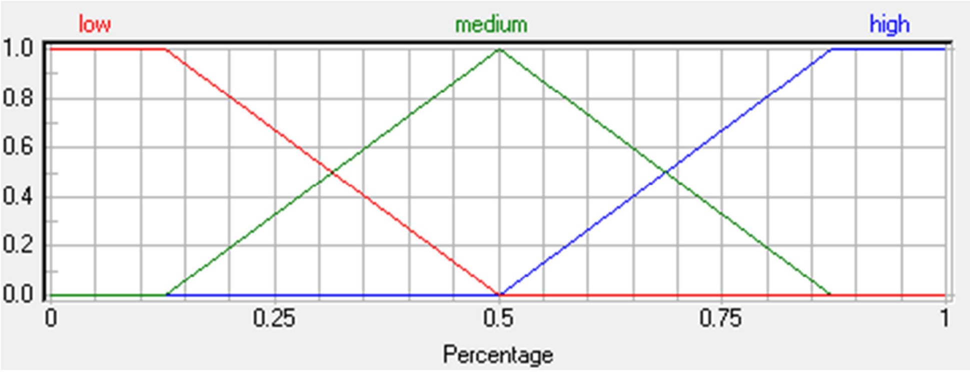


Figure 17: MBF of "Supermarkets"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 17: Definition Points of MBF "Supermarkets"

Input Variable "Telephone"

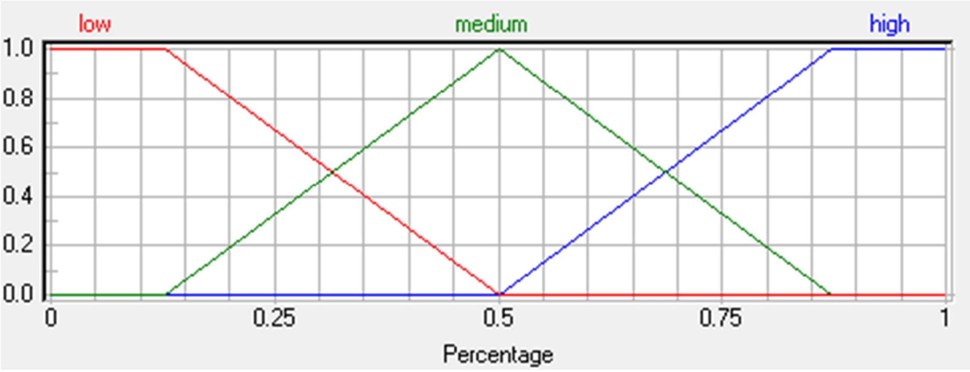


Figure 18: MBF of "Telephone"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 18: Definition Points of MBF "Telephone"

Input Variable " Improved_ Toilets"

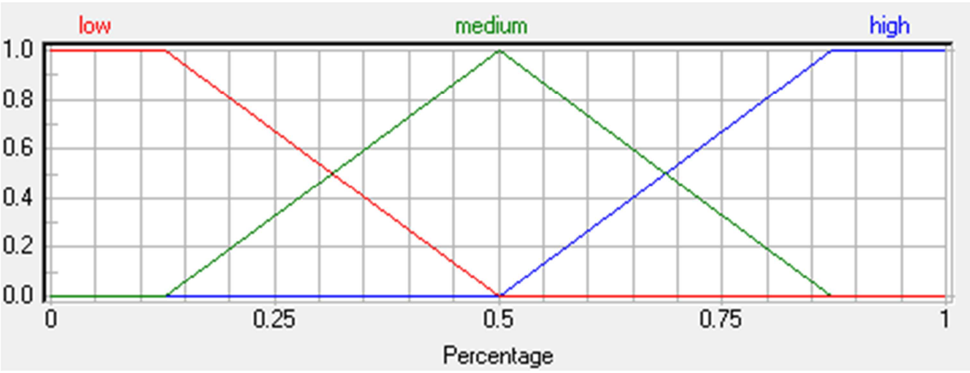


Figure 19: MBF of " Improved_ Toilets"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)

Term Name	Shape/Par.	Definition Points (x, y)		
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 19: Definition Points of MBF " Improved_ Toilets "

Input Variable "Water_Ponds"

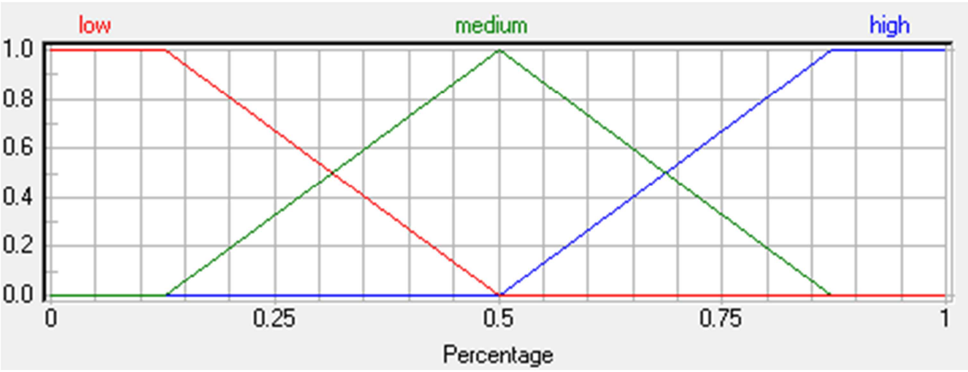


Figure 20: MBF of "Water_Ponds"

Term Name	Shape/Par.	Definition Points (x, y)		
low	linear	(0, 1)	(0.1275, 1)	(0.5, 0)
		(1, 0)		
medium	linear	(0, 0)	(0.1275, 0)	(0.5, 1)
		(0.8725, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.8725, 1)
		(1, 1)		

Table 20: Definition Points of MBF "Water_Ponds"

Output Variable "Educ_Services"

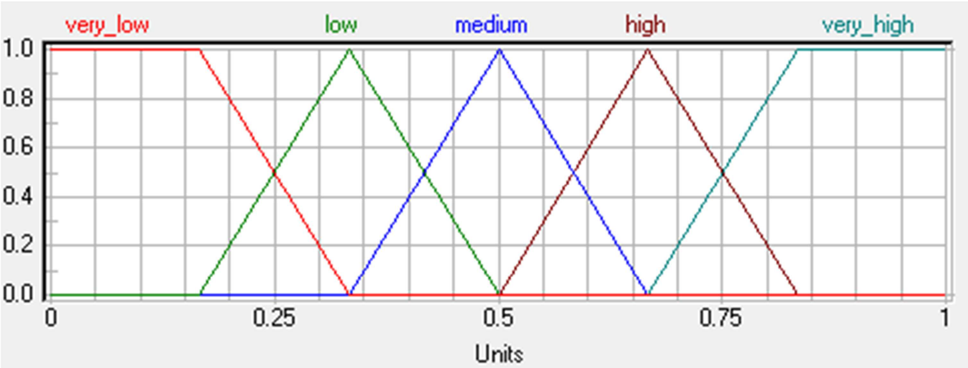


Figure 21: MBF of "Educ_Services"

Term Name	Shape/Par.	Definition Points (x, y)		
very_low	linear	(0, 1) (1, 0)	(0.16666, 1)	(0.33334, 0)
low	linear	(0, 0) (0.5, 0)	(0.16666, 0) (1, 0)	(0.33334, 1)
medium	linear	(0, 0) (0.66666, 0)	(0.33334, 0) (1, 0)	(0.5, 1)
high	linear	(0, 0) (0.83334, 0)	(0.5, 0) (1, 0)	(0.66666, 1)
very_high	linear	(0, 0) (1, 1)	(0.66666, 0)	(0.83334, 1)

Table 21: Definition Points of MBF "Educ_Services"

Output Variable "Health_Services"

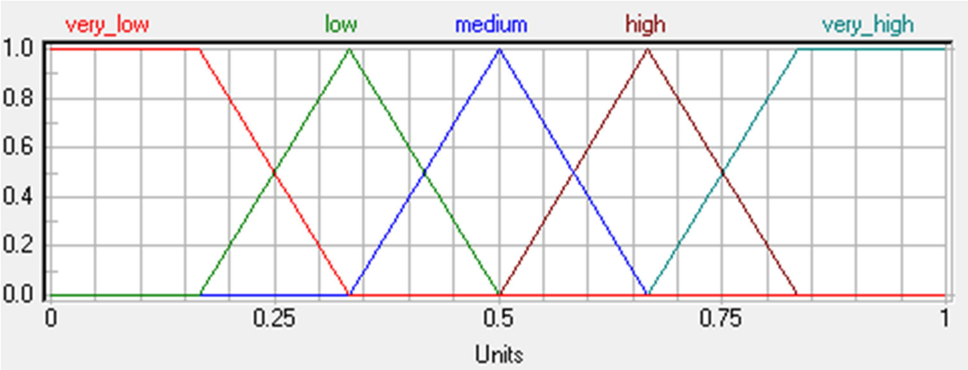


Figure 22: MBF of "Health_Services"

Term Name	Shape/Par.	Definition Points (x, y)		
very_low	linear	(0, 1)	(0.16666, 1)	(0.33334, 0)
		(1, 0)		
low	linear	(0, 0)	(0.16666, 0)	(0.33334, 1)
		(0.5, 0)	(1, 0)	
medium	linear	(0, 0)	(0.33334, 0)	(0.5, 1)
		(0.66666, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.66666, 1)
		(0.83334, 0)	(1, 0)	
very_high	linear	(0, 0)	(0.66666, 0)	(0.83334, 1)
		(1, 1)		

Table 22: Definition Points of MBF "Health_Services"

Output Variable "Hygienic_Service"

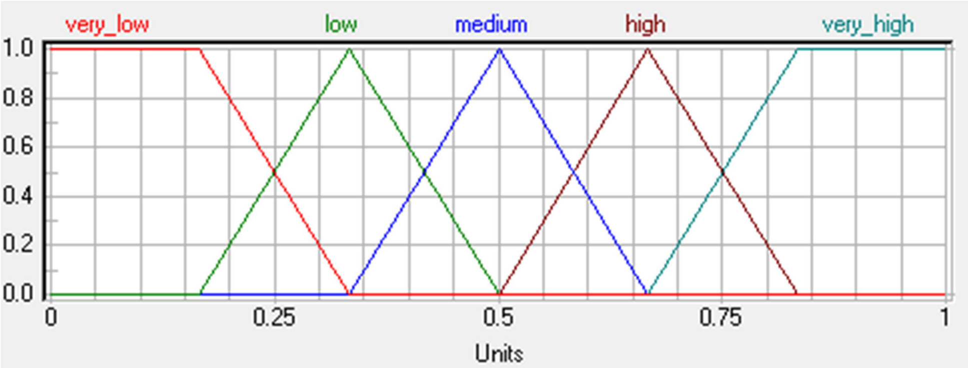


Figure 23: MBF of "Hygienic_Service"

Term Name	Shape/Par.	Definition Points (x, y)		
very_low	linear	(0, 1)	(0.16666, 1)	(0.33334, 0)
		(1, 0)		
low	linear	(0, 0)	(0.16666, 0)	(0.33334, 1)
		(0.5, 0)	(1, 0)	
medium	linear	(0, 0)	(0.33334, 0)	(0.5, 1)
		(0.66666, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.66666, 1)
		(0.83334, 0)	(1, 0)	
very_high	linear	(0, 0)	(0.66666, 0)	(0.83334, 1)
		(1, 1)		

Table 23: Definition Points of MBF "Hygienic_Service"

Output Variable "Infrastructures"

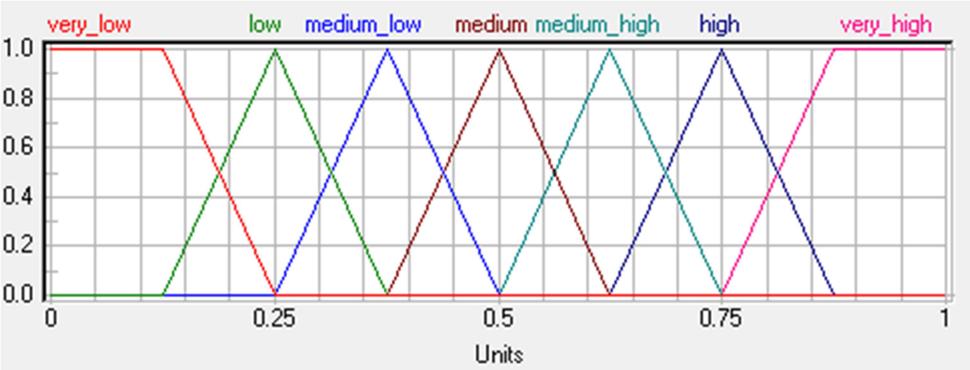


Figure 24: MBF of "Infrastructure"

Term Name	Shape/Par.	Definition Points (x, y)		
very_low	linear	(0, 1)	(0.125, 1)	(0.25, 0)
		(1, 0)		
low	linear	(0, 0)	(0.125, 0)	(0.25, 1)
		(0.375, 0)	(1, 0)	
medium_low	linear	(0, 0)	(0.25, 0)	(0.375, 1)
		(0.5, 0)	(1, 0)	
medium	linear	(0, 0)	(0.375, 0)	(0.5, 1)
		(0.625, 0)	(1, 0)	
medium_high	linear	(0, 0)	(0.5, 0)	(0.625, 1)
		(0.75, 0)	(1, 0)	
high	linear	(0, 0)	(0.625, 0)	(0.75, 1)
		(0.875, 0)	(1, 0)	
very_high	linear	(0, 0)	(0.75, 0)	(0.875, 1)
		(1, 1)		

Table 24: Definition Points of MBF "Infrastructure"

Output Variable "LivingConditions"

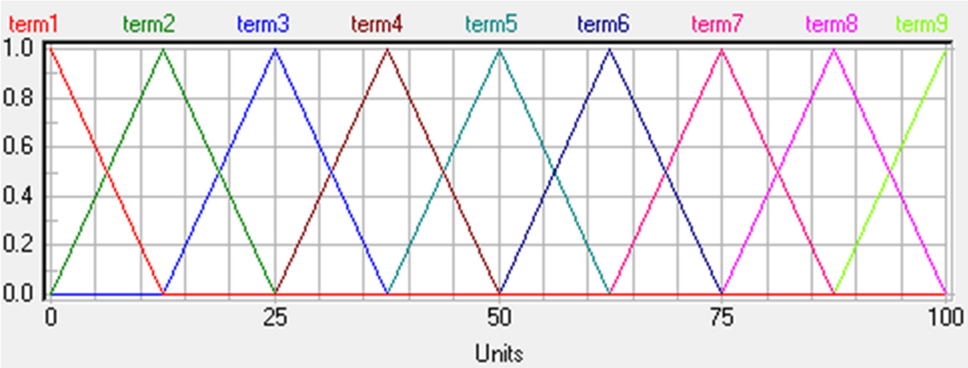


Figure 25: MBF of "LivingConditions"

Term Name	Shape/Par.	Definition Points (x, y)		
term1	linear	(0, 1)	(12.5, 0)	(100, 0)
term2	linear	(0, 0)	(12.5, 1)	(25, 0)
		(100, 0)		
term3	linear	(0, 0)	(12.5, 0)	(25, 1)
		(37.5, 0)	(100, 0)	
term4	linear	(0, 0)	(25, 0)	(37.5, 1)
		(50, 0)	(100, 0)	
term5	linear	(0, 0)	(37.5, 0)	(50, 1)
		(62.5, 0)	(100, 0)	
term6	linear	(0, 0)	(50, 0)	(62.5, 1)
		(75, 0)	(100, 0)	
term7	linear	(0, 0)	(62.5, 0)	(75, 1)
		(87.5, 0)	(100, 0)	
term8	linear	(0, 0)	(75, 0)	(87.5, 1)
		(100, 0)		
term9	linear	(0, 0)	(87.5, 0)	(100, 1)

Table 25: Definition Points of MBF "LivingConditions"

Output Variable "Network_Services"

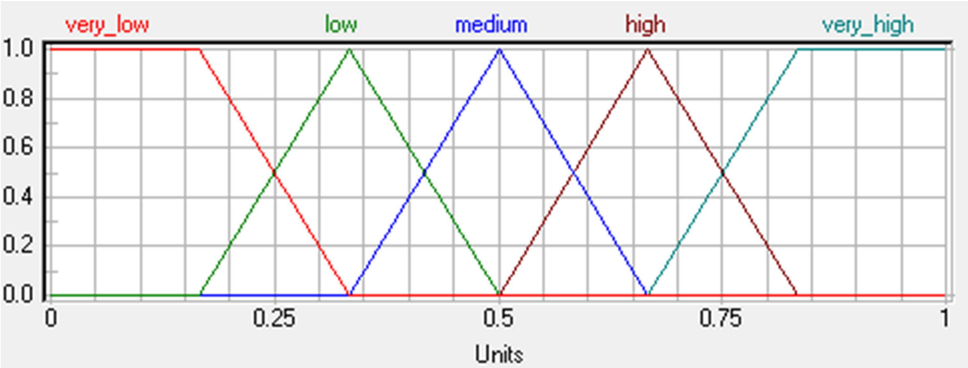


Figure 26: MBF of "Network_Services"

Term Name	Shape/Par.	Definition Points (x, y)		
very_low	linear	(0, 1)	(0.16666, 1)	(0.33334, 0)
		(1, 0)		
low	linear	(0, 0)	(0.16666, 0)	(0.33334, 1)
		(0.5, 0)	(1, 0)	
medium	linear	(0, 0)	(0.33334, 0)	(0.5, 1)
		(0.66666, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.66666, 1)
		(0.83334, 0)	(1, 0)	
very_high	linear	(0, 0)	(0.66666, 0)	(0.83334, 1)
		(1, 1)		

Table 26: Definition Points of MBF "Network_Services"

Output Variable "Sanitation_Level"

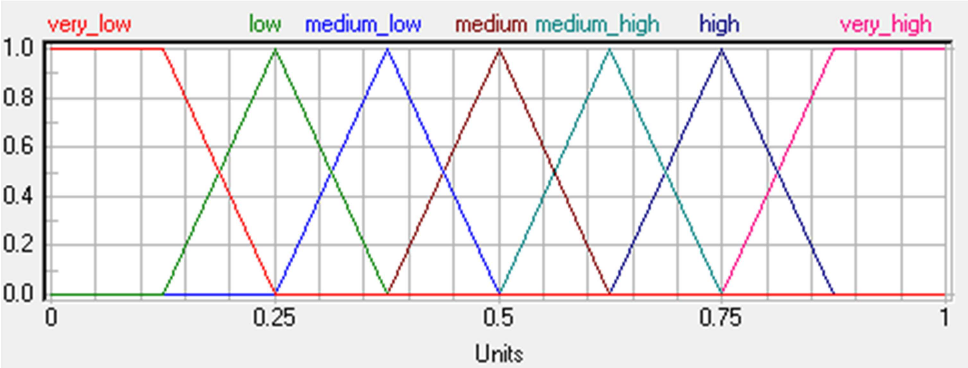


Figure 27: MBF of "Sanitation_Level"

Term Name	Shape/Par.	Definition Points (x, y)		
very_low	linear	(0, 1)	(0.125, 1)	(0.25, 0)
		(1, 0)		
low	linear	(0, 0)	(0.125, 0)	(0.25, 1)
		(0.375, 0)	(1, 0)	
medium_low	linear	(0, 0)	(0.25, 0)	(0.375, 1)
		(0.5, 0)	(1, 0)	
medium	linear	(0, 0)	(0.375, 0)	(0.5, 1)
		(0.625, 0)	(1, 0)	
medium_high	linear	(0, 0)	(0.5, 0)	(0.625, 1)
		(0.75, 0)	(1, 0)	
high	linear	(0, 0)	(0.625, 0)	(0.75, 1)
		(0.875, 0)	(1, 0)	
very_high	linear	(0, 0)	(0.75, 0)	(0.875, 1)
		(1, 1)		

Table 27: Definition Points of MBF "Sanitation_Level"

Output Variable "Transport"

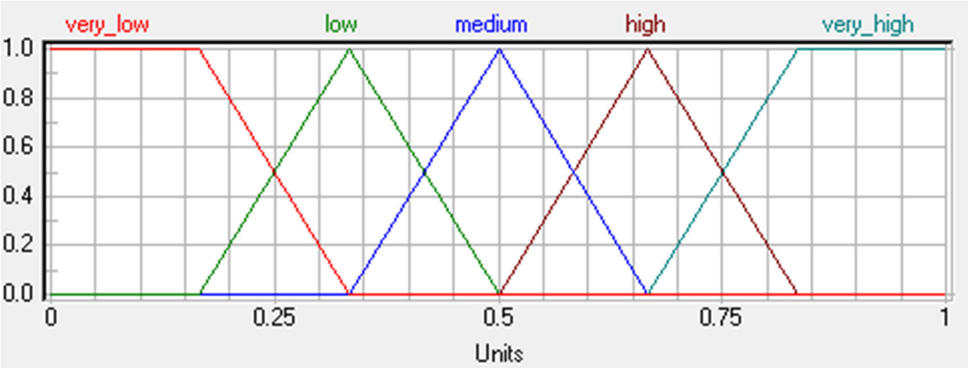


Figure 28: MBF of "Transport"

Term Name	Shape/Par.	Definition Points (x, y)		
very_low	linear	(0, 1)	(0.16666, 1)	(0.33334, 0)
		(1, 0)		
low	linear	(0, 0)	(0.16666, 0)	(0.33334, 1)
		(0.5, 0)	(1, 0)	
medium	linear	(0, 0)	(0.33334, 0)	(0.5, 1)
		(0.66666, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.66666, 1)
		(0.83334, 0)	(1, 0)	
very_high	linear	(0, 0)	(0.66666, 0)	(0.83334, 1)
		(1, 1)		

Table 28: Definition Points of MBF "Transport"

Output Variable "Water_Services"

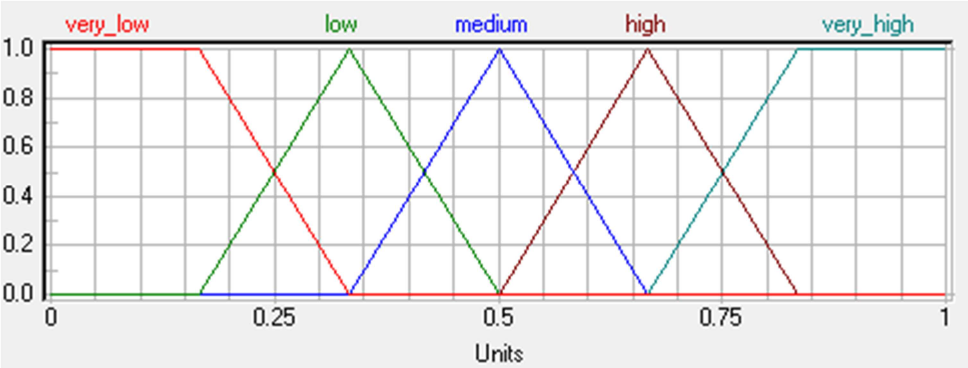


Figure 29: MBF of "Water_Services"

Term Name	**Shape/Par.**	**Definition Points (x, y)**		
very_low	linear	(0, 1)	(0.16666, 1)	(0.33334, 0)
		(1, 0)		
low	linear	(0, 0)	(0.16666, 0)	(0.33334, 1)
		(0.5, 0)	(1, 0)	
medium	linear	(0, 0)	(0.33334, 0)	(0.5, 1)
		(0.66666, 0)	(1, 0)	
high	linear	(0, 0)	(0.5, 0)	(0.66666, 1)
		(0.83334, 0)	(1, 0)	
very_high	linear	(0, 0)	(0.66666, 0)	(0.83334, 1)
		(1, 1)		

Table 29: Definition Points of MBF "Water_Services"

The Rules Block

Rule Block "Education_Services"

Kindergartens +2

Libraries +2

Primary Schools +2

Middle Schools +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	4
Number of Outputs:	1
Number of Rules:	81

IF				**THEN**	
Kindergartens	**Libraries**	**Primary_Schools**	**Middle_Schools**	**DoS**	**Educ_Services**
low	Low	low	low	1.00	very_low
low	Low	low	medium	1.00	very_low
low	Low	low	high	1.00	low
low	Low	medium	low	1.00	very_low
low	Low	medium	medium	1.00	low
low	Low	medium	high	1.00	low
low	Low	high	low	1.00	low
low	Low	high	medium	1.00	low
low	Low	high	high	1.00	medium
low	Medium	low	low	1.00	very_low
low	Medium	low	medium	1.00	low
low	Medium	low	high	1.00	low
low	Medium	medium	low	1.00	low
low	Medium	medium	medium	1.00	low

IF				THEN	
low	Medium	medium	high	1.00	medium
low	Medium	high	low	1.00	low
low	Medium	high	medium	1.00	medium
low	Medium	high	high	1.00	high
low	High	low	low	1.00	low
low	High	low	medium	1.00	low
low	High	low	high	1.00	medium
low	High	medium	low	1.00	low
low	High	medium	medium	1.00	medium
low	High	medium	high	1.00	high
low	High	high	low	1.00	medium
low	High	high	medium	1.00	high
low	High	high	high	1.00	high
medium	Low	low	low	1.00	very_low
medium	Low	low	medium	1.00	low
medium	Low	low	high	1.00	low
medium	Low	medium	low	1.00	low
medium	Low	medium	medium	1.00	low
medium	Low	medium	high	1.00	medium
medium	Low	high	low	1.00	low
medium	Low	high	medium	1.00	medium
medium	Low	high	high	1.00	high
medium	Medium	low	low	1.00	low
medium	Medium	low	medium	1.00	low
medium	Medium	low	high	1.00	medium
medium	Medium	medium	low	1.00	low
medium	Medium	medium	medium	1.00	medium
medium	Medium	medium	high	1.00	high
medium	Medium	high	low	1.00	medium
medium	Medium	high	medium	1.00	high
medium	Medium	high	high	1.00	high
medium	High	low	low	1.00	low
medium	High	low	medium	1.00	medium
medium	High	low	high	1.00	high

IF				**THEN**	
medium	High	medium	low	1.00	medium
medium	High	medium	medium	1.00	high
medium	High	medium	high	1.00	high
medium	High	high	low	1.00	high
medium	High	high	medium	1.00	high
medium	High	high	high	1.00	very_high
high	Low	low	low	1.00	low
high	Low	low	medium	1.00	low
high	Low	low	high	1.00	medium
high	Low	medium	low	1.00	low
high	Low	medium	medium	1.00	medium
high	Low	medium	high	1.00	high
high	Low	high	low	1.00	medium
high	Low	high	medium	1.00	high
high	Low	high	high	1.00	high
high	Medium	low	low	1.00	low
high	Medium	low	medium	1.00	medium
high	Medium	low	high	1.00	high
high	Medium	medium	low	1.00	medium
high	Medium	medium	medium	1.00	high
high	Medium	medium	high	1.00	high
high	Medium	high	low	1.00	high
high	Medium	high	medium	1.00	high
high	Medium	high	high	1.00	very_high
high	High	low	low	1.00	medium
high	High	low	medium	1.00	high
high	High	low	high	1.00	high
high	High	medium	low	1.00	high
high	High	medium	medium	1.00	high
high	High	medium	high	1.00	very_high
high	High	high	low	1.00	high
high	High	high	medium	1.00	very_high
high	High	high	high	1.00	very_high

Table 30: Rules of the Rule Block "Education_Services"

Rule Block "Health_Services"

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	3
Number of Outputs:	1
Number of Rules:	27

IF			THEN	
Clinics	**Hospitals**	**Qualified_Doctor**	**DoS**	**Health_Services**
low	Low	low	1.00	very_low
low	Low	medium	1.00	very_low
low	Low	high	1.00	low
low	Medium	low	1.00	very_low
low	Medium	medium	1.00	low
low	Medium	high	1.00	medium
low	High	low	1.00	low
low	High	medium	1.00	medium
low	High	high	1.00	high
medium	Low	low	1.00	very_low
medium	Low	medium	1.00	low
medium	Low	high	1.00	medium
medium	Medium	low	1.00	low
medium	Medium	medium	1.00	medium
medium	Medium	high	1.00	high
medium	High	low	1.00	medium

IF			THEN	
medium	High	medium	1.00	high
medium	High	high	1.00	very_high
high	Low	low	1.00	low
high	Low	medium	1.00	medium
high	Low	high	1.00	high
high	Medium	low	1.00	medium
high	Medium	medium	1.00	high
high	Medium	high	1.00	very_high
high	High	low	1.00	high
high	High	medium	1.00	very_high
high	High	high	1.00	very_high

Table 31: Rules of the Rule Block "Health_Services"

Rule Block "Hygenic_Services"

Garbage Disposal +2

Biogas Pit +2

Toilets Improved +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	3
Number of Outputs:	1
Number of Rules:	27

IF	THEN

IF			THEN	
Garbage_Disposal	**Biogas_Pit**	**Toilets_Improved**	**DoS**	**Hygienic_Service**
low	Low	low	1.00	very_low
low	Low	medium	1.00	very_low
low	Low	high	1.00	low
low	Medium	low	1.00	very_low
low	Medium	medium	1.00	low
low	Medium	high	1.00	medium
low	High	low	1.00	low
low	High	medium	1.00	medium
low	High	high	1.00	high
medium	Low	low	1.00	very_low
medium	Low	medium	1.00	low
medium	Low	high	1.00	medium
medium	Medium	low	1.00	low
medium	Medium	medium	1.00	medium
medium	Medium	high	1.00	high
medium	High	low	1.00	medium
medium	High	medium	1.00	high
medium	High	high	1.00	very_high
high	Low	low	1.00	low
high	Low	medium	1.00	medium
high	Low	high	1.00	high
high	Medium	low	1.00	medium
high	Medium	medium	1.00	high
high	Medium	high	1.00	very_high
high	High	low	1.00	high
high	High	medium	1.00	very_high
high	High	high	1.00	very_high

Table 32: Rules of the Rule Block "Hygenic_Services"

Rule Block "Infrastructures"

Trsanportation +2

Network +2

Water +2

Supermarkets +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	4
Number of Outputs:	1
Number of Rules:	375

IF				THEN	
Transport	**Network_Services**	**Water_Services**	**Supermarkets**	**DoS**	**Infrastructure**
very_low	very_low	very_low	low	1.00	very_low
very_low	very_low	very_low	medium	1.00	very_low
very_low	very_low	very_low	high	1.00	low
very_low	very_low	low	low	1.00	very_low
very_low	very_low	low	medium	1.00	low
very_low	very_low	low	high	1.00	medium_low
very_low	very_low	medium	low	1.00	very_low
very_low	very_low	medium	medium	1.00	low
very_low	very_low	medium	high	1.00	medium_low
very_low	very_low	high	low	1.00	low
very_low	very_low	high	medium	1.00	medium_low
very_low	very_low	high	high	1.00	medium
very_low	very_low	very_high	low	1.00	low
very_low	very_low	very_high	medium	1.00	medium_low
very_low	very_low	very_high	high	1.00	medium
very_low	low	very_low	low	1.00	very_low
very_low	low	very_low	medium	1.00	low

IF				THEN	
very_low	low	very_low	high	1.00	medium_low
very_low	low	low	low	1.00	very_low
very_low	low	low	medium	1.00	low
very_low	low	low	high	1.00	medium_low
very_low	low	medium	low	1.00	low
very_low	low	medium	medium	1.00	medium_low
very_low	low	medium	high	1.00	medium
very_low	low	high	low	1.00	low
very_low	low	high	medium	1.00	medium_low
very_low	low	high	high	1.00	medium
very_low	low	very_high	low	1.00	medium_low
very_low	low	very_high	medium	1.00	medium
very_low	low	very_high	high	1.00	medium
very_low	medium	very_low	low	1.00	very_low
very_low	medium	very_low	medium	1.00	low
very_low	medium	very_low	high	1.00	medium_low
very_low	medium	low	low	1.00	low
very_low	medium	low	medium	1.00	medium_low
very_low	medium	low	high	1.00	medium
very_low	medium	medium	low	1.00	low
very_low	medium	medium	medium	1.00	medium_low
very_low	medium	medium	high	1.00	medium
very_low	medium	high	low	1.00	medium_low
very_low	medium	high	medium	1.00	medium
very_low	medium	high	high	1.00	medium
very_low	medium	very_high	low	1.00	medium_low
very_low	medium	very_high	medium	1.00	medium
very_low	medium	very_high	high	1.00	medium_high
very_low	high	very_low	low	1.00	low
very_low	high	very_low	medium	1.00	medium_low
very_low	high	very_low	high	1.00	medium
very_low	high	low	low	1.00	low
very_low	high	low	medium	1.00	medium_low
very_low	high	low	high	1.00	medium

IF				THEN	
very_low	high	medium	low	1.00	medium_low
very_low	high	medium	medium	1.00	medium
very_low	high	medium	high	1.00	medium
very_low	high	high	low	1.00	medium_low
very_low	high	high	medium	1.00	medium
very_low	high	high	high	1.00	medium_high
very_low	high	very_high	low	1.00	medium
very_low	high	very_high	medium	1.00	medium
very_low	high	very_high	high	1.00	medium_high
very_low	very_high	very_low	low	1.00	low
very_low	very_high	very_low	medium	1.00	medium_low
very_low	very_high	very_low	high	1.00	medium
very_low	very_high	low	low	1.00	medium_low
very_low	very_high	low	medium	1.00	medium
very_low	very_high	low	high	1.00	medium
very_low	very_high	medium	low	1.00	medium_low
very_low	very_high	medium	medium	1.00	medium
very_low	very_high	medium	high	1.00	medium_high
very_low	very_high	high	low	1.00	medium
very_low	very_high	high	medium	1.00	medium
very_low	very_high	high	high	1.00	medium_high
very_low	very_high	very_high	low	1.00	medium
very_low	very_high	very_high	medium	1.00	medium_high
very_low	very_high	very_high	high	1.00	high
low	very_low	very_low	low	1.00	very_low
low	very_low	very_low	medium	1.00	low
low	very_low	very_low	high	1.00	medium_low
low	very_low	low	low	1.00	very_low
low	very_low	low	medium	1.00	low
low	very_low	low	high	1.00	medium_low
low	very_low	medium	low	1.00	low
low	very_low	medium	medium	1.00	medium_low
low	very_low	medium	high	1.00	medium
low	very_low	high	low	1.00	low

IF				**THEN**	
low	very_low	high	medium	1.00	medium_low
low	very_low	high	high	1.00	medium
low	very_low	very_high	low	1.00	medium_low
low	very_low	very_high	medium	1.00	medium
low	very_low	very_high	high	1.00	medium
low	low	very_low	low	1.00	very_low
low	low	very_low	medium	1.00	low
low	low	very_low	high	1.00	medium_low
low	low	low	low	1.00	low
low	low	low	medium	1.00	medium_low
low	low	low	high	1.00	medium
low	low	medium	low	1.00	low
low	low	medium	medium	1.00	medium_low
low	low	medium	high	1.00	medium
low	low	high	low	1.00	medium_low
low	low	high	medium	1.00	medium
low	low	high	high	1.00	medium
low	low	very_high	low	1.00	medium_low
low	low	very_high	medium	1.00	medium
low	low	very_high	high	1.00	medium_high
low	medium	very_low	low	1.00	low
low	medium	very_low	medium	1.00	medium_low
low	medium	very_low	high	1.00	medium
low	medium	low	low	1.00	low
low	medium	low	medium	1.00	medium_low
low	medium	low	high	1.00	medium
low	medium	medium	low	1.00	medium_low
low	medium	medium	medium	1.00	medium
low	medium	medium	high	1.00	medium
low	medium	high	low	1.00	medium_low
low	medium	high	medium	1.00	medium
low	medium	high	high	1.00	medium_high
low	medium	very_high	low	1.00	medium
low	medium	very_high	medium	1.00	medium

IF				THEN	
low	medium	very_high	high	1.00	medium_high
low	high	very_low	low	1.00	low
low	high	very_low	medium	1.00	medium_low
low	high	very_low	high	1.00	medium
low	high	low	low	1.00	medium_low
low	high	low	medium	1.00	medium
low	high	low	high	1.00	medium
low	high	medium	low	1.00	medium_low
low	high	medium	medium	1.00	medium
low	high	medium	high	1.00	medium_high
low	high	high	low	1.00	medium
low	high	high	medium	1.00	medium
low	high	high	high	1.00	medium_high
low	high	very_high	low	1.00	medium
low	high	very_high	medium	1.00	medium_high
low	high	very_high	high	1.00	high
low	very_high	very_low	low	1.00	medium_low
low	very_high	very_low	medium	1.00	medium
low	very_high	very_low	high	1.00	medium
low	very_high	low	low	1.00	medium_low
low	very_high	low	medium	1.00	medium
low	very_high	low	high	1.00	medium_high
low	very_high	medium	low	1.00	medium
low	very_high	medium	medium	1.00	medium
low	very_high	medium	high	1.00	medium_high
low	very_high	high	low	1.00	medium
low	very_high	high	medium	1.00	medium_high
low	very_high	high	high	1.00	high
low	very_high	very_high	low	1.00	medium
low	very_high	very_high	medium	1.00	medium_high
low	very_high	very_high	high	1.00	high
medium	very_low	very_low	low	1.00	very_low
medium	very_low	very_low	medium	1.00	low
medium	very_low	very_low	high	1.00	medium_low

IF				THEN	
medium	very_low	low	low	1.00	low
medium	very_low	low	medium	1.00	medium_low
medium	very_low	low	high	1.00	medium
medium	very_low	medium	low	1.00	low
medium	very_low	medium	medium	1.00	medium_low
medium	very_low	medium	high	1.00	medium
medium	very_low	high	low	1.00	medium_low
medium	very_low	high	medium	1.00	medium
medium	very_low	high	high	1.00	medium
medium	very_low	very_high	low	1.00	medium_low
medium	very_low	very_high	medium	1.00	medium
medium	very_low	very_high	high	1.00	medium_high
medium	low	very_low	low	1.00	low
medium	low	very_low	medium	1.00	medium_low
medium	low	very_low	high	1.00	medium
medium	low	low	low	1.00	low
medium	low	low	medium	1.00	medium_low
medium	low	low	high	1.00	medium
medium	low	medium	low	1.00	medium_low
medium	low	medium	medium	1.00	medium
medium	low	medium	high	1.00	medium
medium	low	high	low	1.00	medium_low
medium	low	high	medium	1.00	medium
medium	low	high	high	1.00	medium_high
medium	low	very_high	low	1.00	medium
medium	low	very_high	medium	1.00	medium
medium	low	very_high	high	1.00	medium_high
medium	medium	very_low	low	1.00	low
medium	medium	very_low	medium	1.00	medium_low
medium	medium	very_low	high	1.00	medium
medium	medium	low	low	1.00	medium_low
medium	medium	low	medium	1.00	medium
medium	medium	low	high	1.00	medium
medium	medium	medium	low	1.00	medium_low

IF				THEN	
medium	medium	medium	medium	1.00	medium
medium	medium	medium	high	1.00	medium_high
medium	medium	high	low	1.00	medium
medium	medium	high	medium	1.00	medium
medium	medium	high	high	1.00	medium_high
medium	medium	very_high	low	1.00	medium
medium	medium	very_high	medium	1.00	medium_high
medium	medium	very_high	high	1.00	high
medium	high	very_low	low	1.00	medium_low
medium	high	very_low	medium	1.00	medium
medium	high	very_low	high	1.00	medium
medium	high	low	low	1.00	medium_low
medium	high	low	medium	1.00	medium
medium	high	low	high	1.00	medium_high
medium	high	medium	low	1.00	medium
medium	high	medium	medium	1.00	medium
medium	high	medium	high	1.00	medium_high
medium	high	high	low	1.00	medium
medium	high	high	medium	1.00	medium_high
medium	high	high	high	1.00	high
medium	high	very_high	low	1.00	medium
medium	high	very_high	medium	1.00	medium_high
medium	high	very_high	high	1.00	high
medium	very_high	very_low	low	1.00	medium_low
medium	very_high	very_low	medium	1.00	medium
medium	very_high	very_low	high	1.00	medium_high
medium	very_high	low	low	1.00	medium
medium	very_high	low	medium	1.00	medium
medium	very_high	low	high	1.00	medium_high
medium	very_high	medium	low	1.00	medium
medium	very_high	medium	medium	1.00	medium_high
medium	very_high	medium	high	1.00	high
medium	very_high	high	low	1.00	medium
medium	very_high	high	medium	1.00	medium_high

IF				THEN	
medium	very_high	high	high	1.00	high
medium	very_high	very_high	low	1.00	medium_high
medium	very_high	very_high	medium	1.00	high
medium	very_high	very_high	high	1.00	very_high
high	very_low	very_low	low	1.00	low
high	very_low	very_low	medium	1.00	medium_low
high	very_low	very_low	high	1.00	medium
high	very_low	low	low	1.00	low
high	very_low	low	medium	1.00	medium_low
high	very_low	low	high	1.00	medium
high	very_low	medium	low	1.00	medium_low
high	very_low	medium	medium	1.00	medium
high	very_low	medium	high	1.00	medium
high	very_low	high	low	1.00	medium_low
high	very_low	high	medium	1.00	medium
high	very_low	high	high	1.00	medium_high
high	very_low	very_high	low	1.00	medium
high	very_low	very_high	medium	1.00	medium
high	very_low	very_high	high	1.00	medium_high
high	low	very_low	low	1.00	low
high	low	very_low	medium	1.00	medium_low
high	low	very_low	high	1.00	medium
high	low	low	low	1.00	medium_low
high	low	low	medium	1.00	medium
high	low	low	high	1.00	medium
high	low	medium	low	1.00	medium_low
high	low	medium	medium	1.00	medium
high	low	medium	high	1.00	medium_high
high	low	high	low	1.00	medium
high	low	high	medium	1.00	medium
high	low	high	high	1.00	medium_high
high	low	very_high	low	1.00	medium
high	low	very_high	medium	1.00	medium_high
high	low	very_high	high	1.00	high

IF				THEN	
high	medium	very_low	low	1.00	medium_low
high	medium	very_low	medium	1.00	medium
high	medium	very_low	high	1.00	medium
high	medium	low	low	1.00	medium_low
high	medium	low	medium	1.00	medium
high	medium	low	high	1.00	medium_high
high	medium	medium	low	1.00	medium
high	medium	medium	medium	1.00	medium
high	medium	medium	high	1.00	medium_high
high	medium	high	low	1.00	medium
high	medium	high	medium	1.00	medium_high
high	medium	high	high	1.00	high
high	medium	very_high	low	1.00	medium
high	medium	very_high	medium	1.00	medium_high
high	medium	very_high	high	1.00	high
high	high	very_low	low	1.00	medium_low
high	high	very_low	medium	1.00	medium
high	high	very_low	high	1.00	medium_high
high	high	low	low	1.00	medium
high	high	low	medium	1.00	medium
high	high	low	high	1.00	medium_high
high	high	medium	low	1.00	medium
high	high	medium	medium	1.00	medium_high
high	high	medium	high	1.00	high
high	high	high	low	1.00	medium
high	high	high	medium	1.00	medium_high
high	high	high	high	1.00	high
high	high	very_high	low	1.00	medium_high
high	high	very_high	medium	1.00	high
high	high	very_high	high	1.00	very_high
high	very_high	very_low	low	1.00	medium
high	very_high	very_low	medium	1.00	medium
high	very_high	very_low	high	1.00	medium_high
high	very_high	low	low	1.00	medium

IF				THEN	
high	very_high	low	medium	1.00	medium_high
high	very_high	low	high	1.00	high
high	very_high	medium	low	1.00	medium
high	very_high	medium	medium	1.00	medium_high
high	very_high	medium	high	1.00	high
high	very_high	high	low	1.00	medium_high
high	very_high	high	medium	1.00	high
high	very_high	high	high	1.00	very_high
high	very_high	very_high	low	1.00	medium_high
high	very_high	very_high	medium	1.00	high
high	very_high	very_high	high	1.00	very_high
very_high	very_low	very_low	low	1.00	low
very_high	very_low	very_low	medium	1.00	medium_low
very_high	very_low	very_low	high	1.00	medium
very_high	very_low	low	low	1.00	medium_low
very_high	very_low	low	medium	1.00	medium
very_high	very_low	low	high	1.00	medium
very_high	very_low	medium	low	1.00	medium_low
very_high	very_low	medium	medium	1.00	medium
very_high	very_low	medium	high	1.00	medium_high
very_high	very_low	high	low	1.00	medium
very_high	very_low	high	medium	1.00	medium
very_high	very_low	high	high	1.00	medium_high
very_high	very_low	very_high	low	1.00	medium
very_high	very_low	very_high	medium	1.00	medium_high
very_high	very_low	very_high	high	1.00	high
very_high	low	very_low	low	1.00	medium_low
very_high	low	very_low	medium	1.00	medium
very_high	low	very_low	high	1.00	medium
very_high	low	low	low	1.00	medium_low
very_high	low	low	medium	1.00	medium
very_high	low	low	high	1.00	medium_high
very_high	low	medium	low	1.00	medium
very_high	low	medium	medium	1.00	medium

IF				THEN	
very_high	low	medium	high	1.00	medium_high
very_high	low	high	low	1.00	medium
very_high	low	high	medium	1.00	medium_high
very_high	low	high	high	1.00	high
very_high	low	very_high	low	1.00	medium
very_high	low	very_high	medium	1.00	medium_high
very_high	low	very_high	high	1.00	high
very_high	medium	very_low	low	1.00	medium_low
very_high	medium	very_low	medium	1.00	medium
very_high	medium	very_low	high	1.00	medium_high
very_high	medium	low	low	1.00	medium
very_high	medium	low	medium	1.00	medium
very_high	medium	low	high	1.00	medium_high
very_high	medium	medium	low	1.00	medium
very_high	medium	medium	medium	1.00	medium_high
very_high	medium	medium	high	1.00	high
very_high	medium	high	low	1.00	medium
very_high	medium	high	medium	1.00	medium_high
very_high	medium	high	high	1.00	high
very_high	medium	very_high	low	1.00	medium_high
very_high	medium	very_high	medium	1.00	high
very_high	medium	very_high	high	1.00	very_high
very_high	high	very_low	low	1.00	medium
very_high	high	very_low	medium	1.00	medium
very_high	high	very_low	high	1.00	medium_high
very_high	high	low	low	1.00	medium
very_high	high	low	medium	1.00	medium_high
very_high	high	low	high	1.00	high
very_high	high	medium	low	1.00	medium
very_high	high	medium	medium	1.00	medium_high
very_high	high	medium	high	1.00	high
very_high	high	high	low	1.00	medium_high
very_high	high	high	medium	1.00	high
very_high	high	high	high	1.00	very_high

IF				THEN	
very_high	high	very_high	low	1.00	medium_high
very_high	high	very_high	medium	1.00	high
very_high	high	very_high	high	1.00	very_high
very_high	very_high	very_low	low	1.00	medium
very_high	very_high	very_low	medium	1.00	medium_high
very_high	very_high	very_low	high	1.00	high
very_high	very_high	low	low	1.00	medium
very_high	very_high	low	medium	1.00	medium_high
very_high	very_high	low	high	1.00	high
very_high	very_high	medium	low	1.00	medium_high
very_high	very_high	medium	medium	1.00	high
very_high	very_high	medium	high	1.00	very_high
very_high	very_high	high	low	1.00	medium_high
very_high	very_high	high	medium	1.00	high
very_high	very_high	high	high	1.00	very_high
very_high	very_high	very_high	low	1.00	high
very_high	very_high	very_high	medium	1.00	very_high
very_high	very_high	very_high	high	1.00	very_high

Table 33: Rules of the Rule Block "Infrastructure"

Rule Block "Living_Conditions_Index"

Infrastructures +2

Education +2

Sanitation +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM

Number of Inputs: 3

Number of Outputs: 1

Number of Rules: 245

IF			THEN	
Infrastructures	**Educ_Services**	**Sanitation_Level**	**DoS**	**LivingConditions**
very_low	very_low	very_low	1.00	term1
very_low	very_low	low	1.00	term1
very_low	very_low	medium_low	1.00	term1
very_low	very_low	medium	1.00	term2
very_low	very_low	medium_high	1.00	term2
very_low	very_low	high	1.00	term3
very_low	very_low	very_high	1.00	term4
very_low	low	very_low	1.00	term1
very_low	low	low	1.00	term2
very_low	low	medium_low	1.00	term2
very_low	low	medium	1.00	term3
very_low	low	medium_high	1.00	term3
very_low	low	high	1.00	term4
very_low	low	very_high	1.00	term4
very_low	medium	very_low	1.00	term2
very_low	medium	low	1.00	term2
very_low	medium	medium_low	1.00	term3
very_low	medium	medium	1.00	term4
very_low	medium	medium_high	1.00	term4
very_low	medium	high	1.00	term5
very_low	medium	very_high	1.00	term5
very_low	high	very_low	1.00	term3
very_low	high	low	1.00	term3
very_low	high	medium_low	1.00	term4
very_low	high	medium	1.00	term4
very_low	high	medium_high	1.00	term5
very_low	high	high	1.00	term5

IF			THEN	
very_low	high	very_high	1.00	term6
very_low	very_high	very_low	1.00	term4
very_low	very_high	low	1.00	term4
very_low	very_high	medium_low	1.00	term5
very_low	very_high	medium	1.00	term5
very_low	very_high	medium_high	1.00	term6
very_low	very_high	high	1.00	term6
very_low	very_high	very_high	1.00	term7
low	very_low	very_low	1.00	term1
low	very_low	low	1.00	term1
low	very_low	medium_low	1.00	term2
low	very_low	medium	1.00	term2
low	very_low	medium_high	1.00	term3
low	very_low	high	1.00	term3
low	very_low	very_high	1.00	term4
low	low	very_low	1.00	term2
low	low	low	1.00	term2
low	low	medium_low	1.00	term3
low	low	medium	1.00	term3
low	low	medium_high	1.00	term4
low	low	high	1.00	term4
low	low	very_high	1.00	term5
low	medium	very_low	1.00	term2
low	medium	low	1.00	term3
low	medium	medium_low	1.00	term3
low	medium	medium	1.00	term4
low	medium	medium_high	1.00	term4
low	medium	high	1.00	term5
low	medium	very_high	1.00	term5
low	high	very_low	1.00	term3
low	high	low	1.00	term4
low	high	medium_low	1.00	term4
low	high	medium	1.00	term5
low	high	medium_high	1.00	term5

IF			THEN	
low	high	high	1.00	term6
low	high	very_high	1.00	term6
low	very_high	very_low	1.00	term4
low	very_high	low	1.00	term4
low	very_high	medium_low	1.00	term5
low	very_high	medium	1.00	term5
low	very_high	medium_high	1.00	term6
low	very_high	high	1.00	term6
low	very_high	very_high	1.00	term7
medium_low	very_low	very_low	1.00	term1
medium_low	very_low	low	1.00	term2
medium_low	very_low	medium_low	1.00	term3
medium_low	very_low	medium	1.00	term3
medium_low	very_low	medium_high	1.00	term4
medium_low	very_low	high	1.00	term4
medium_low	very_low	very_high	1.00	term5
medium_low	low	very_low	1.00	term2
medium_low	low	low	1.00	term3
medium_low	low	medium_low	1.00	term3
medium_low	low	medium	1.00	term4
medium_low	low	medium_high	1.00	term4
medium_low	low	high	1.00	term5
medium_low	low	very_high	1.00	term5
medium_low	medium	very_low	1.00	term3
medium_low	medium	low	1.00	term4
medium_low	medium	medium_low	1.00	term4
medium_low	medium	medium	1.00	term5
medium_low	medium	medium_high	1.00	term5
medium_low	medium	high	1.00	term6
medium_low	medium	very_high	1.00	term6
medium_low	high	very_low	1.00	term4
medium_low	high	low	1.00	term4
medium_low	high	medium_low	1.00	term5
medium_low	high	medium	1.00	term5

IF			THEN	
medium_low	high	medium_high	1.00	term6
medium_low	high	high	1.00	term6
medium_low	high	very_high	1.00	term7
medium_low	very_high	very_low	1.00	term5
medium_low	very_high	low	1.00	term5
medium_low	very_high	medium_low	1.00	term6
medium_low	very_high	medium	1.00	term6
medium_low	very_high	medium_high	1.00	term7
medium_low	very_high	high	1.00	term7
medium_low	very_high	very_high	1.00	term8
medium	very_low	very_low	1.00	term2
medium	very_low	low	1.00	term3
medium	very_low	medium_low	1.00	term3
medium	very_low	medium	1.00	term4
medium	very_low	medium_high	1.00	term4
medium	very_low	high	1.00	term5
medium	very_low	very_high	1.00	term5
medium	low	very_low	1.00	term3
medium	low	low	1.00	term3
medium	low	medium_low	1.00	term4
medium	low	medium	1.00	term4
medium	low	medium_high	1.00	term5
medium	low	high	1.00	term5
medium	low	very_high	1.00	term6
medium	medium	very_low	1.00	term4
medium	medium	low	1.00	term4
medium	medium	medium_low	1.00	term5
medium	medium	medium	1.00	term5
medium	medium	medium_high	1.00	term6
medium	medium	high	1.00	term6
medium	medium	very_high	1.00	term7
medium	high	very_low	1.00	term4
medium	high	low	1.00	term5
medium	high	medium_low	1.00	term5

IF			THEN	
medium	high	medium	1.00	term6
medium	high	medium_high	1.00	term6
medium	high	high	1.00	term7
medium	high	very_high	1.00	term7
medium	very_high	very_low	1.00	term5
medium	very_high	low	1.00	term6
medium	very_high	medium_low	1.00	term6
medium	very_high	medium	1.00	term7
medium	very_high	medium_high	1.00	term7
medium	very_high	high	1.00	term8
medium	very_high	very_high	1.00	term8
medium_high	very_low	very_low	1.00	term3
medium_high	very_low	low	1.00	term3
medium_high	very_low	medium_low	1.00	term4
medium_high	very_low	medium	1.00	term4
medium_high	very_low	medium_high	1.00	term5
medium_high	very_low	high	1.00	term5
medium_high	very_low	very_high	1.00	term6
medium_high	low	very_low	1.00	term3
medium_high	low	low	1.00	term4
medium_high	low	medium_low	1.00	term4
medium_high	low	medium	1.00	term5
medium_high	low	medium_high	1.00	term5
medium_high	low	high	1.00	term6
medium_high	low	very_high	1.00	term6
medium_high	medium	very_low	1.00	term4
medium_high	medium	low	1.00	term5
medium_high	medium	medium_low	1.00	term5
medium_high	medium	medium	1.00	term6
medium_high	medium	medium_high	1.00	term6
medium_high	medium	high	1.00	term7
medium_high	medium	very_high	1.00	term7
medium_high	high	very_low	1.00	term5
medium_high	high	low	1.00	term5

IF			THEN	
medium_high	high	medium_low	1.00	term6
medium_high	high	medium	1.00	term6
medium_high	high	medium_high	1.00	term7
medium_high	high	high	1.00	term7
medium_high	high	very_high	1.00	term8
medium_high	very_high	very_low	1.00	term6
medium_high	very_high	low	1.00	term6
medium_high	very_high	medium_low	1.00	term7
medium_high	very_high	medium	1.00	term7
medium_high	very_high	medium_high	1.00	term8
medium_high	very_high	high	1.00	term8
medium_high	very_high	very_high	1.00	term9
high	very_low	very_low	1.00	term3
high	very_low	low	1.00	term4
high	very_low	medium_low	1.00	term4
high	very_low	medium	1.00	term5
high	very_low	medium_high	1.00	term5
high	very_low	high	1.00	term6
high	very_low	very_high	1.00	term6
high	low	very_low	1.00	term4
high	low	low	1.00	term4
high	low	medium_low	1.00	term5
high	low	medium	1.00	term5
high	low	medium_high	1.00	term6
high	low	high	1.00	term6
high	low	very_high	1.00	term7
high	medium	very_low	1.00	term5
high	medium	low	1.00	term5
high	medium	medium_low	1.00	term6
high	medium	medium	1.00	term6
high	medium	medium_high	1.00	term7
high	medium	high	1.00	term7
high	medium	very_high	1.00	term8
high	high	very_low	1.00	term5

IF			THEN	
high	high	low	1.00	term6
high	high	medium_low	1.00	term6
high	high	medium	1.00	term7
high	high	medium_high	1.00	term7
high	high	high	1.00	term8
high	high	very_high	1.00	term8
high	very_high	very_low	1.00	term6
high	very_high	low	1.00	term7
high	very_high	medium_low	1.00	term7
high	very_high	medium	1.00	term8
high	very_high	medium_high	1.00	term8
high	very_high	high	1.00	term9
high	very_high	very_high	1.00	term9
very_high	very_low	very_low	1.00	term4
very_high	very_low	low	1.00	term4
very_high	very_low	medium_low	1.00	term5
very_high	very_low	medium	1.00	term5
very_high	very_low	medium_high	1.00	term6
very_high	very_low	high	1.00	term6
very_high	very_low	very_high	1.00	term7
very_high	low	very_low	1.00	term4
very_high	low	low	1.00	term5
very_high	low	medium_low	1.00	term5
very_high	low	medium	1.00	term6
very_high	low	medium_high	1.00	term6
very_high	low	high	1.00	term7
very_high	low	very_high	1.00	term7
very_high	medium	very_low	1.00	term5
very_high	medium	low	1.00	term6
very_high	medium	medium_low	1.00	term6
very_high	medium	medium	1.00	term7
very_high	medium	medium_high	1.00	term7
very_high	medium	high	1.00	term8
very_high	medium	very_high	1.00	term8

IF			THEN	
very_high	high	very_low	1.00	term6
very_high	high	low	1.00	term6
very_high	high	medium_low	1.00	term7
very_high	high	medium	1.00	term7
very_high	high	medium_high	1.00	term8
very_high	high	high	1.00	term8
very_high	high	very_high	1.00	term9
very_high	very_high	very_low	1.00	term7
very_high	very_high	low	1.00	term7
very_high	very_high	medium_low	1.00	term8
very_high	very_high	medium	1.00	term8
very_high	very_high	medium_high	1.00	term9
very_high	very_high	high	1.00	term9
very_high	very_high	very_high	1.00	term9

Table 34: Rules of the Rule Block "Living_Conditions_Index"

Rule Block "Network_Services"

Electricity +2

Telephone +2

Cable TV +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	3
Number of Outputs:	1
Number of Rules:	27

IF			THEN	
Cable_TV	**Electricity**	**Telephone**	**DoS**	**Network_Services**
low	low	low	1.00	very_low
low	low	medium	1.00	very_low
low	low	high	1.00	low
low	medium	low	1.00	very_low
low	medium	medium	1.00	low
low	medium	high	1.00	medium
low	high	low	1.00	low
low	high	medium	1.00	medium
low	high	high	1.00	high
medium	low	low	1.00	very_low
medium	low	medium	1.00	low
medium	low	high	1.00	medium
medium	medium	low	1.00	low
medium	medium	medium	1.00	medium
medium	medium	high	1.00	high
medium	high	low	1.00	medium
medium	high	medium	1.00	high
medium	high	high	1.00	very_high
high	low	low	1.00	low
high	low	medium	1.00	medium
high	low	high	1.00	high
high	medium	low	1.00	medium
high	medium	medium	1.00	high
high	medium	high	1.00	very_high
high	high	low	1.00	high
high	high	medium	1.00	very_high
high	high	high	1.00	very_high

Table 35: Rules of the Rule Block "Network_Services"

Rule Block "Sanitation_Level"

Hygenic Services +2

Health Services +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	2
Number of Outputs:	1
Number of Rules:	25

IF		**THEN**	
Hygienic_Service	**Health_Services**	**DoS**	**Sanitation_Level**
very_low	very_low	1.00	very_low
low	very_low	1.00	very_low
medium	very_low	1.00	low
high	very_low	1.00	medium_low
very_high	very_low	1.00	medium
very_low	low	1.00	very_low
low	low	1.00	low
medium	low	1.00	medium_low
high	low	1.00	medium
very_high	low	1.00	medium_high
very_low	medium	1.00	low
low	medium	1.00	medium_low
medium	medium	1.00	medium
high	medium	1.00	medium_high
very_high	medium	1.00	high
very_low	high	1.00	medium_low
low	high	1.00	medium

IF		THEN	
medium	high	1.00	medium_high
high	high	1.00	high
very_high	high	1.00	very_high
very_low	very_high	1.00	medium
low	very_high	1.00	medium_high
medium	very_high	1.00	high
high	very_high	1.00	very_high
very_high	very_high	1.00	very_high

Table 36: Rules of the Rule Block "Sanitation_Level"

Rule Block "Transportation_Services"

Highways +2

Busses +2

Road_Materials +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	3
Number of Outputs:	1
Number of Rules:	27

IF			THEN	
Busses	**Highways**	**Road_Material**	**DoS**	**Transports**
low	low	low	1.00	very_low
low	low	medium	1.00	very_low

IF			THEN	
low	low	high	1.00	low
low	medium	low	1.00	very_low
low	medium	medium	1.00	low
low	medium	high	1.00	medium
low	high	low	1.00	low
low	high	medium	1.00	medium
low	high	high	1.00	high
medium	low	low	1.00	very_low
medium	low	medium	1.00	low
medium	low	high	1.00	medium
medium	medium	low	1.00	low
medium	medium	medium	1.00	medium
medium	medium	high	1.00	high
medium	high	low	1.00	medium
medium	high	medium	1.00	high
medium	high	high	1.00	very_high
high	low	low	1.00	low
high	low	medium	1.00	medium
high	low	high	1.00	high
high	medium	low	1.00	medium
high	medium	medium	1.00	high
high	medium	high	1.00	very_high
high	high	low	1.00	high
high	high	medium	1.00	very_high
high	high	high	1.00	very_high

Table 37: Rules of the Rule Block "Transportation_Services"

Rule Block "Water_Services"

Drink Water +2

Motor Wells +2

Water Ponds +2

Parameter

Aggregation:	MINMAX
Parameter:	0.00
Result Aggregation:	BSUM
Number of Inputs:	3
Number of Outputs:	1
Number of Rules:	27

Table 38: Rules of the Rule Block "Water_Services"

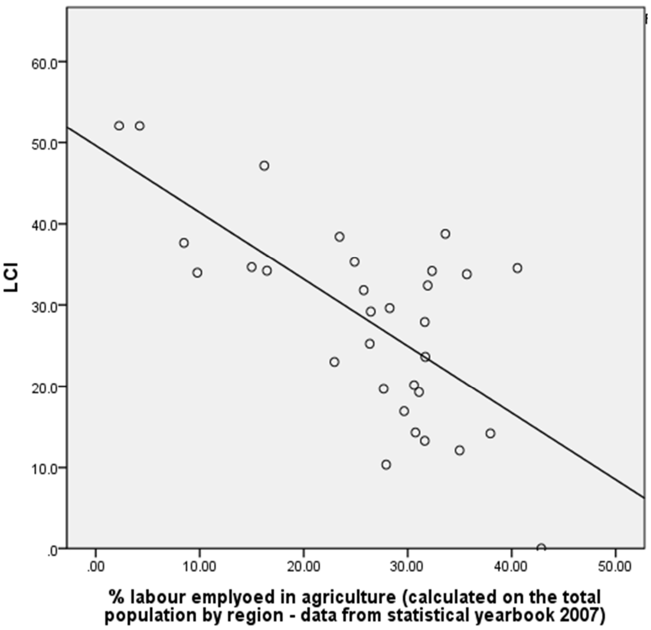
IF			**THEN**	
Drink_Water	**Motor_Wells**	**Water_Ponds**	**DoS**	**Water_Services**
low	low	low	1.00	very_low
low	low	medium	1.00	very_low
low	low	high	1.00	low
low	medium	low	1.00	very_low
low	medium	medium	1.00	low
low	medium	high	1.00	medium
low	high	low	1.00	low
low	high	medium	1.00	medium
low	high	high	1.00	high
medium	low	low	1.00	very_low
medium	low	medium	1.00	low
medium	low	high	1.00	medium
medium	medium	low	1.00	low
medium	medium	medium	1.00	medium
medium	medium	high	1.00	high
medium	high	low	1.00	medium
medium	high	medium	1.00	high
medium	high	high	1.00	very_high
high	low	low	1.00	low
high	low	medium	1.00	medium
high	low	high	1.00	high
high	medium	low	1.00	medium
high	medium	medium	1.00	high

IF			THEN	
high	medium	high	1.00	very_high
high	high	low	1.00	high
high	high	medium	1.00	very_high
high	high	high	1.00	very_high

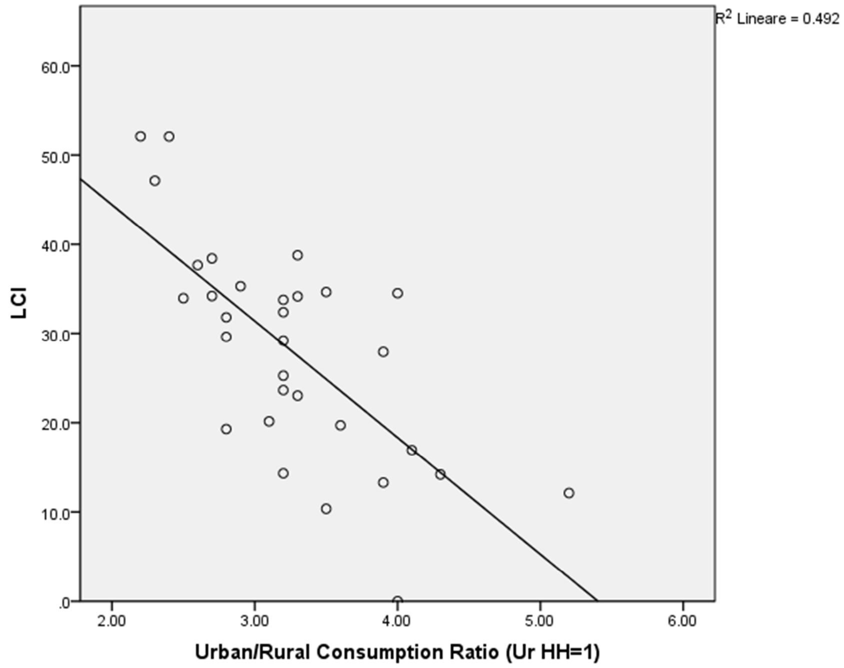
Appendix 3

		Rate of employed labour in agriculture
LCI	Correlation Pearson coefficient	-.680**
	Sig. (2-code)	.000
	N	31

**. The correlation is signifcant at a level of 0,01 (2-code).



		Urban/Rural Consumption Ratio (Ur HH=1)
LCI	Correlation Pearson coefficient	-.701**
	Sig. (2-code)	.000
	N	31



R2 Lineare = 0.492
60.0
50.0
40.0
30.0
20.0
10.0
.0
LCI
2.00
3.00
4.00
5.00
6.00
Urban/Rural Consumption Ratio (Ur HH=1)

Bibliography

Addabbo T., G. Facchinetti, T. Pirotti (2011), Measuring the Capability of Living a Healthy Life with Fuzzy Logic in a Gender Perspective, IJCCI 2011 – FCTA, 24-26 Ottobre 2011, Parigi, Francia. ISBN: 978-989-8425-83-6

Addabbo T., G. Facchinetti, A. Maccagnan, G. Mastroleo, T. Pirotti (2010), Sen's Capability and Functionings: A fuzzy system to evaluate how parents interact with their children, Metody Informatyki Stosowanej, Vol. 3/2010(24) ISSN1898-5297

Ahlquist J., C. Breunig (2011), Model-Based Clustering and Typologies in the Social Sciences, «Political Analysis», Vol. 20, No. 1, pg 92-112

Antle J. M. (1983), Infrastructure and Aggregate Agricultural Productivity: International Evidence, «Economic Development and Cultural Change», 31, 609-43

Biggeri L. (2001), *China Agricultural Census Design and Data Collection: Issues and Lessons*, Proceedings of The International Seminar on China The Agricultural Census Results, China Statistics Press, Beijing

Biggeri M. (1998), *I Fattori dello Sviluppo dell'Agricoltura Cinese dopo le Grandi Riforme, Analisi Empiriche su un Panel di Dati a Livello Provinciale*, Tesi di dottorato, Università degli Studi di Siena

Biggeri M. (2003), Key Factors of Recent Chinese Provincial Economic Growth, Journal of Chinese Economic and Business Studies, 1:2, 159-183. Available at: http://dx.doi.org/10.1080/1476528032000066703A

Deutsch J., J. Silber (2005), Measuring Multidimensional Poverty: an Empirical Comparison of Various Approaches, in «Review of Income and Wealth», Series 51, No. 1

Du C., J. Cheng (2009), Soft Cluster Analysis of Circular Economy Development Level in China, 2009 International Joint Conference on Computational Sciences and Optimization, 978-0-7695-3605-7/09 © 2009 IEEE, DOI 10.1109/CSO.2009.114

Dunteman G. (1989), Principal Components Analysis, Sage

Economist (The), *Welcome Home*, 25th February 2012, The Economist Newspaper Limited, London

Everitt B. S. (2011), Unresolved Problems in Cluster Analysis, International Biometric Society, Biometrics, Vol. 35, No. 1, Perspectives in Biometry (Mar., 1979), pp. 169- 181. Available online at: http://www.jstor.org/stable/2529943.

Fan S., Zhang X. (2004), Infrastructure and regional economic development in rural China, «China Economic Review», 15, 203-14

Fanfani R., Brasili C. (2001), *The New Geography of Chinese Agriculture*, Proceedings of The International Seminar on China The Agricultural Census Results, China Statistics Press, Beijing

Fanfani R., Brasili C. (2003), *Regional Differences in Chinese Agriculture: Results from the 1997 First National Agricultural Census*, in «Journal of Peasant Studies», Vol. 30, no.3/4.

Fanfani R., Brasili C. (2005), “Regional differences in land holdings and land use: analyzing the first agricultural census of China”, in P. Ho (ed.), *Developmental Dilemmas: Land reform, Property rights and environment in China*,Routledge ed. London and New York

Fosu A. K. (2010), Inequality, Income and Poverty- Comparative Global Evidence, World Institute for Development Economics Research, Working Paper No. 2010/93

Fraley C. and A. Raftery (1998), How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis, The Computer Journal, Vol. 41, No. 8

Fraley C. and A. E. Raftery (1998), How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis, THE COMPUTER JOURNAL, Vol. 41, No. 8, 1998. Available at: http://comjnl.oxfordjournals.org

Gollin D., S. Parente, R. Rogerson (2002), The Role of Agriculture in Development, The American Economic Review/ Association, Vol. 92, No. 2, Papers and Proceedings of the OneHundred Fourteenth Annual Meeting of the American Economic Association (May, 2002), pp.160-164

Gulati A., Fan S. (ed.) (2007), *The Dragon &The Elephant: Agricultural and rural reforms in China and India*, The John Hopkins University Press

Guo Y., P. Chen (2011), Digital Divide and Social Cleavage: Case Studies of ICT Usage among Peasants in Contemporary China”, in «The China Quarterly», vol. 207, p. 580-599, Cambridge University Press

Han J., Kamber M. and Tung A. (2001), *Spatial clustering methods in data mining: A survey*, In Miller H., Han J. (eds.), Geographic Data Mining and Knowledge Discovery, Taylor & Francis

Heilig G. (2003), *Rural Development or Sustainable Development in China: Is China’s Rural Development Sustainable?*, paper presented at the ECOLINK Workshop, Montpellier, France, 18-19 september 2003

Heilig G. K. (2003), Rural Development or Sustainable Development in China: Is China’s Rural Development Sustainable? , International Institute for Applied Systems Analysis (IIASA),

Hubert L., P. Arabie (1985), Comparing partitions, Journal of Classification, 2, 193-218.

Imanov, Gorkhmaz (2007), The Fuzzy Approach to estimation of the Index of Population Life Quality, «Fuzzy Economic Review», Vol. XII, No. 2, p. 85-93

Jain A. K. and R. C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, 1988.

Kanbur R., X. Zhang (1999), Which Regional Inequality? The Evolution of Rural–Urban and Inland–Coastal Inequality in China from 1983 to 1995, «Journal of Comparative Economics», Volume 27, No. 4, Academic Press

Kanbur1 R., X. Zhang (2004), Fifty Years of Regional Inequality in China- A Journey through Central Planning, Reform, and Openness, World Institute for Development Economics Research, Working Paper No. 2004/50

Kannan R., S. Vempala, A. Vetta (2004), On clusterings: Good, bad and spectral, «Journal of the ACM», Vol. 51, No. 3

Kaufman L., P. Rousseeuw (1990), Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley and Sons, New Jersey

Knight J., Shi L. and Renwei Z. (2004), Divergent Means and Convergent Inequality of Incomes among the Provinces and Cities of Urban China, World Institute for Development Economics Research, Research Paper No. 2004/52.

Leisch F. (2006), A toolbox for K-centroids cluster analysis, «Computational Statistics & Data Analysis», vol. 51, No. 2

Lin J. Y. (1992), Rural Reforms and Agricultural Growth in China, «America Economic Review», Vol. 82, No.1, pp. 34-51

Lohmar B. (*et al.*) (2009), China's Ongoing Agricultural Modernization. Challenges Remain after 30 Years of Reform, Economic Information Bullettin, No. 51, USDA

Lohmar B., F. Gale, F. Tuan, J. Hansen (2009), Background: China's Economic Development and Implications for Agriculture, China's Ongoing Agricultural Modernization: Challenges Remain After 30 Years of Reform / EIB-51, Economic Research Service, USDA

Mazzocchi M. (2008), Statistics for Marketing and Consumer Research, SAGE Publications, London

Minot N. and B. Baulch (2004), Poverty Mapping with Aggregate- Census Data- What is the Loss in Precision? , World Institute for Development Economics Research, Working Paper No. 2004/38.

Monasterolo I., N. Coppola (2010), MAPPING SERBIA: MORE TARGETED RURAL AREAS FOR BETTER POLICIES, 118th seminar of the EAAE, Ljubljana 2010

Morrison W. M. (2011), China's Economic Conditions, Congressional Research Service, 7-5700

Montgomery Jr. E. B., H. Huanga and A. Assadi (2005), Unsupervised clustering algorithm for N-dimensional data, «Journal of Neuroscience Methods», Vol. 144, No. 1

Murat M., T. Pirotti (2010), The attractiveness of countries for FDI. A fuzzy approach, «Fuzzy Economic Review», Vol. XV, No. 2

NBS (2011) *China Statistical Yearbook 2010*, State Statistical Bureau People's Republic of China, Beijing Agricultural University Press

OECD (2011), Goods trade balance, Trade: Key Tables from OECD, No. 2, available at: http://www.oecd-ilibrary.org/sites/tbal-gds-table-2011-1-en/index.html;jsessionid=g0g80l3krm4lg.delta?contentType=/ns/StatisticalPublication,/ns/Table,/ns/KeyTableEdition&itemId=/content/table/tbal-gds-table-2011-1-en&containerItemId=/content/table/20743920-table2&accessItemIds=/content/tablecollection/20743920&mimeType=text/html

OECD (2009), Rural Policy Reviews: China, ISBN 978-92-64-05956-6.

Pagliaro V., Pirotti T., Socci L. (2010), “Distribuzione del gas in Europa. Un sistema fuzzy di valutazione del mercato”, in La cultura dell’incertezza. Fuzzy – Logica – Sanità – Arte, Ed. Eurilink

Pirotti T. (2010), Analisi congiunturale e strutturale delle economie regionali: un approccio “fuzzy” alla specializzazione produttiva, PhD thesis, University of Bologna

Ravallion M. (2003), Externalities in Rural Development- Evidence for China, World Institute for Development Economics Research, Discussion Paper No. 2003/54

Reynolds A. P., G. Richards, B. De La Iglesia and V. J. Rayward-Smith (2006), Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms, Journal of Mathematical Modelling and Algorithms (2006) 5: 475–504, DOI: 10.1007/s10852-005-9022-1

Rural Survey Organization of National Bureau of Statistics (2004), Poverty Statistics in China, “International Conference on Official Poverty Statistics: Methodology and Comparability”,Manila, Philippines
http://www.nscb.gov.ph/poverty/conference/papers/4_poverty statistics in china.pdf

Saade J. J., Diab H. B. (2004), *Defuzzification Methods and New Techniques for Fuzzy Controllers,* Iranian Journal of Electrical and Computer Engineering, Vol. 3, No. 2

Salvador S. and Chan P. (2004), Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms, Proceedings of the 16th IEE International Conference on Tools with AI (2004), pp. 576–584

Tan P. N., M. Steinbach, V. Kumar (2006), Introduction to Data Mining, Addison Wesley

Van der Laan M. J., K. S. Pollard, J. Bryan (2002), A New Partitioning Around Medoids Algorithm, paper No. 105, The Berkely Electronic Press

Wan G. and Z. Zhou (2004), Income Inequality in Rural China- Regression-Based Decomposition Using Household Data, World Institute for Development Economics Research, Working Paper No. 2004/51

World Bank (2008),*Agriculture for development,* Annual Report

World Bank (2009), Reshaping Economic Geography, Annual Report

World Bank (2009), *From Poor Areas to Poor People: China’s Evolving Poverty Reduction Agenda. An Assessment of Poverty and Inequality in China,* Report

WTO (2010), International Trade Statistics, ISBN 978-92-870-3739-8. Available online at: http://www.wto.org/english/res_e/statis_e/its2010_e/its10_toc_e.htm

Xiang Y. (2008), Rural-Urban Migration and Education in China: A Case Study, Duke East Asia Nexus. Available at: http://www.dukenexus.org/39/rural-urban-migration-and-education-in-china-a-case-study/

Xiong H., J. Wu and J. Chen (2004), K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 39, NO. 2, APRIL 2009

Yang Y., A. Hu (2008), Investigating Regional Disparities of China's Human Development with Cluster Analysis: A Historical Perspective, Social Indicators Research, Vol. 86, no. 3, 417-432

Yao S. (2000), Economic Development and Poverty Reduction in China over 20 years of Reforms, «Economic Development and Cultural Change», Vol. 48, No. 3, pp. 447-474

Yu T. T., M. M. Zhang (2007), Multivariate Analysis Of Foreign- Direct Investment In China, Journal of Applied Business Research– Second Quarter 2007, Volume 23, Number 2

Zadeh L. A. (1968), Fuzzy Algorithm, Information and Control, vol.12, pp. 94-102

Zadeh L. A. (1973), Outline of a new Approach to the analysis of complex systems and decision processes, IEEE Tras. Syst., Man, Cybern., vol. 3, pp. 28-44

Zhang Z. (2011), "Age Difference of Productivity and Lewisian Turning Point", paper presented at the XVI IEA World Congress, Beijing, July 2011

Zhang X., J. Yang, S. Wang (2010), China has reached the Lewis Turning Point, IFPRI Discussion Paper 000977

Zen X. (2011), Comparative Research on Rural Ddevelopment- Levels of 31 Provinciances and Regions in Chiana, Asian Agricultural Reserch , 3 (3) :31-33, 38

Zheng X. (2011), Comparative Research on the Rural Development Levels of 31 Provinces and Regions in China, «Asian Agricultural research», vol. 3, No. 3