# Geographic Determinants of China's Urbanization*

Peter Christensen†

Gordon C. McCord‡

December 19, 2013

## Abstract

We employ a unique set of satellite-based and other biophysical measurements to identify the impact of two exogenous parameters – constraints to agriculture and distance to ports – on urban location and growth. The setting is the most expansive urban growth in history: China in the 1990's. Our results indicate that while land quality for agriculture played an important role in determining the location of urban areas, it has a negative effect on urban growth during the contemporary period (1990-2000). This suggests that the opportunity cost of urbanizing agriculturally fertile land today is non-negligible and that the process of agglomeration may create a natural source of tension between land uses. Market access and land quality together explain nearly 80% of the variation in urban development across China. We examine regional heterogeneity in the marginal effects of geography on urban growth and present the predicted spatial distribution of Chinese urban areas through 2030.

Keywords: urbanization, China, satellite, spatial econometrics, geography

*The authors offer a special thank you to Dr. Jiyuan Liu and colleagues at the Chinese Academy of Sciences for contributing critical expertise and data on land use/land cover change in China. Helpful feedback from Ken Gillingham, Joshua Graff Zivin, Matthew Grant, Gordon Hanson, Solomon Hsiang, Chandra Kiran Bangalore Krishnamurty, and William Masters was much appreciated.

†peter.christensen@yale.edu, Yale University 380 Edwards St. New Haven, CT 06511

‡gmccord@ucsd.edu, University of California, San Diego

# 1 Introduction

Many now refer to the 21st century as the urban century. The global urban population is expected to grow from 2.8 billion in 2000 to 6.25 billion by 2050 before stabilizing at the end of the century (UN Urbanization Prospects 2012). It is clear that much of this growth will occur in emerging Asia and Africa, however far less is known about the distribution and types of settlements that are developing in those regions. Rapid urbanization around the developing world and the trade-offs over land use have generated enormous interest in the study of how geographic features influence the spatial distribution of urban growth. There is evidence that locational fundamentals are first-order determinants of city size rank (David and Weinstein 2002), and recent research seeks to understand how location matters: for example examining the effect of geography on the timing of urban development (Motamed et al. 2009), urban activity as a function of access to primate cities (Storeygard 2012), housing market responses to land constraints (Saiz 2010), and how physical geography conditions urban sprawl in the U.S. (Burchfield et al 2006). This paper employs a unique set of satellite and spatially explicit data on land quality, ports, rivers and topography to study the impact of market access and land suitability for agriculture in determining urban location and expansion during the most rapid build-out in history: China in the 1990's. Taking into account the spatial nature of urbanization, we employ a variety of econometric techniques to identify the total effect of these exogenous geographic characteristics, providing new estimates onhow the biophysical landscape conditions human society, and advancing urban growth projections for planners and policy-makers .

Our work is in the spirit of models that relate agglomeration and urban dynamics to underlying geography. Given that modern societies locate most of the labor, production, and high-productivity activity in cities, economics has long been concerned with the geographic constraints that determine the location and growth of urban areas. One geographical feature extensively discussed in the literature is access to markets. The "new economic geography" implicitly linked geography to the virtuous cycles of agglomeration by postulating that declining transport costs can stimulate industrial production by increasing the purchasing power of rural consumers and the market for manufactured goods (Fujita et al 1996, 1998, Krugman 1991). Importantly, the growth of the industrial sector and further reduction in transport costs ultimately fuels a spatial re-organization as industrial firms respond to returns associated with agglomeration. A growing urban center is the spatial product of this emerging production hub and simultaneous migration of workers (Murata 2008, Ottaviano 2002). A somewhat distinct and long-standing body of theory emerged to explain the spatial patterns of growth within and across urban regions (Alonso 1964, Mills 1967, Muth 1969). As with the new economic geography models, the growth of urban areas will depend on access to global and regional markets

– trade will increase both the incomes of urban producers as well as potential in-migrants.

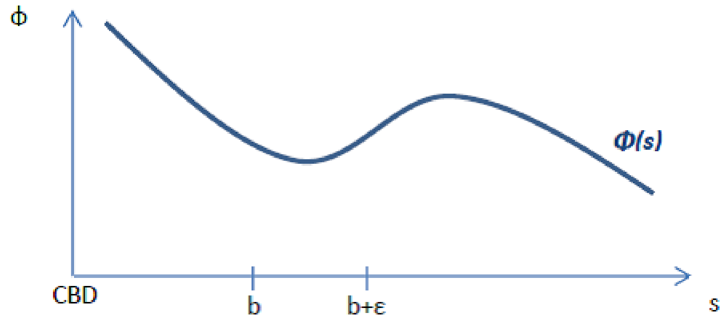
A second geographical feature relevant for urban dynamics is agricultural production potential. One strand of the literature that addresses agriculture and urban growth focuses on the transition from agricultural to industrial production, explaining the agglomeration of firms and population as a part of a multifaceted structural change described by Kuznets (1973) and Lucas (2000). When agricultural productivity rises beyond subsistence, labor is freed up for the manufacturing and service sectors, both of which achieve higher productivity in urban areas, particularly in those with access to local and global markets to competitively buy inputs and produce exports. The effect of geography in affecting the timing of structural change and urbanization was proposed by Gollin et al (2002) and empirically tested by Motamed et al (2009), who find that areas with higher agricultural productivity and navigable river access were more likely to urbanize sooner. On the other hand, the Alonso-Mills-Muth models mentioned above predict that highly productive agricultural areas present a high opportunity cost to urbanization. Cities grow away from a Central Business District (CBD) and trade off benefits of urban activity (urban wages net of rent and commuting costs to the CBD) with the agricultural land rents forgone by expanding the urban area. Empirical work has measured a negative relationship agricultural land rents and urbanization in the US (Brueckner and Fansler 1983, McGrath 2005) and more recently in China (Deng et al. 2008).

The question of geographic contraints to urbanization is particularly relevant for China, where contemporary economic reforms have catalyzed urbanization at an unparalleled rate. China added more that 150 million inhabitants to its cities between 1990-2000 and another 452 million are expected between 2000 and 2030 (UN Urbanization Prospects 2012). An enormous amount of policy attention has been paid to determining the optimal location for urban growth within the country as well as to the preservation of farmland around rapidly growing urban centers. A growing literature has examined China's urbanization processes and the policies that regulate them (see Lichtenberg and Ding 2008 for review, Ding and Lichtenberg 2011). Deng et al (2008) test and confirm classic predictions from urban economic theory about what drives the spatial structure of urban growth, including the role of agricultural land value (as explained below, this paper extends this line of inquiry by exploring the role of exogenous physical characteristics affecting agricultural potential). Au and Henderson (2006a) examine the relationship between worker output and the spatial size of cities, finding that China's cities appear to be undersized relative to the optimum. Other studies have looked at the impacts of political incentives (Lichtenberg and Ding 2009) and institutional constraints such as the Hukou system (migration restrictions) on China's process of urbanization (Au and Henderson 2006b, Boskar et al., 2012). Ongoing work focuses on the effects of transportation infrastructure on city population decentralization (Baum-Snow et al. 2012) and on incomes (Banerjee et al. 2012).

This paper exploits a unique landcover dataset for China and employes a range of estimation techniques to understand how exogenous physical geography conditions urban location and growth. We begin with a simple illustrative model describing the ambiguous prior regarding the empirical relationship between land quality for agriculture and urban land use. We then move to the empirical question of interest: whether biophysical land quality for agriculture and access to global markets continue to affect urban area location and growth in the late 20th century in China. We employ a variety of estimation strategies; clean identification is challenging due to spatial dependence inherent in agglomeration processes and the endogeneity of key parameters such as the agricultural land value, income levels, productivity growth, and transportation infrastructure. This study contributes to the literature by exploiting a unique set of satellite and other spatially explicit data to examine the role of two exogenous factors in determining and constraining the location of China's new urban lands: (1) biophysical land suitability for agriculture and (2) distance to major ports. We consider the issue of spatial dependence and employ a range of spatial econometric techniques to improve estimation and to examine spillover effects of urbanization on nearby urbanization. We then examine regional heterogeneity in the effect of geographical constraints across China's provinces and finally employ a localized technique (geographically weighted regression) to explore this heterogeneity very flexibly. We conclude by using our models to predict the spatial distribution of urban expansion in 2010, 2020, and 2030.

## 2 Theoretical Motivation

Two-sector models predict that city genesis will occur in or very near agriculturally favorable lands, particularly in historic times when transport costs were high. If urban growth is a process that exhibits some increasing returns to agglomeration, then we would expect to see a strong association between urban area location and agricultural land quality. Moreover, to the extent that land quality is a spatially autocorrelated physical attribute, then the fact that city genesis differentially occurs near good land means that we expect to continue to see urban growth occurring in areas better suited for agriculture. On the other hand, to the extent that land use change on the margins of cities takes into account the opportunity cost of urbanizing agriculturally productive land, we might expect that better lands would be differentially less urbanized as city growth proceeds. This ambiguity in the relationship between urbanization and agricultural land quality is described mathematically below:




Following the Alonso-Mills-Muth tradition, we begin with a circular city in which distance from the central business district (CBD) is measured by $s$, and $b$ represents the radius of the city. The biophysical land quality for agriculture (a function of soil types, moisture availability, growing degree days, etc.) in location $s$ is given by $\phi(s)$ where a higher value of $\phi(s)$ indicates superior suitability for agriculture (we simplify by making the units of $\phi$ to be tons of food, which implies that farmers are optimally allocating inputs). Note that $\phi(s)$ is exogenous to social dynamics. We analyze the marginal plot of land of size $\varepsilon$ and posit that the land use change from agricultural to urban use will occur first where the profits from urban activities minus the profits from agricultural activities (the opportunity cost) is greatest; where agricultural profits are greater than urban activity profits, urbanization won't occur. Both food and urban products (which we will call manufactures for simplicity) must be transported to the CBD, which means that urban incomes at the city margin are a function of transport costs (where $\tau$ represents a cost per ton per unit of distance), the price of manufactures ($P_M$) relative to the price of food($P_F$), and urban productivity. We allow for increasing returns by allowing urban productivity to grow with city size, and follow the literature in international technological diffusion by allowing cities with better access to international markets to enjoy higher urban productivity (due to faster technological diffusion, or alternatively to the presence of more exporting firms, which have been found to be more productive [Bernard et al., 2003]). This access to international markets is proxied by the city's distance to one of China's 22 largest ports ($\rho$). Results are robust to using distance to coastline instead of distance to port; we opt for distance to port because not every place along the coast is suitable for a major port.

Urbanization of the marginal plot of size ε occurs, then, if urban incomes from the plot would outweigh rural incomes:

$$P_M \cdot f\left(\int_0^b \phi(s)ds\right) \cdot (\frac{\mu}{\rho}) \int_b^{b+\varepsilon} s(1-\tau s)ds > P_F \cdot \int_b^{b+\varepsilon} \phi(s)(1-\tau s)ds \tag{1}$$

The left side of the inequality shows urban incomes on the plot. Notice that it includes urban productivity $\mu$ modified by an increasing function $f$ of the agricultural land quality of the already urbanized area $b$. This represents the path-dependent nature of the urbanization dynamic: in historic times with high transport costs for food, cities formed in areas good for farming; the better the biophysical conditions for agriculture, the higher the population density attainable. Later on, this higher population density generates the increasing returns for urban productivity growth represented here. The basic urban productivity parameter $\mu$ is in per distance units, such that $\mu\varepsilon$ is in tons, and the term in parentheses reduces income by the appropriate transport cost.

The point to note is that $\varphi(\varepsilon)$ – the biophysical quality of the marginal plot – appears on the right hand side, while the historical $\varphi$ of the city's urbanized areas appears on the left hand side. At a coarse resolution (where the unit of observation is at a spatial unit larger than $\varepsilon$), we have an ambiguous prior on the relationship between the quality of land for agriculture and urban land use. We fully expect that proximity to major ports, however, will favor city location and city growth. The model motivates the following reduced-form specification to measure the association between urban land area and our geographical characteristics of interest:

$$urbanarea_i = \beta_0 + \beta_1 \cdot landsuit_i + \beta_2 \cdot ln(costdistance_i) + \varepsilon_i \tag{2}$$

where the percentage of urbanized land in area i is modeled as a function of the land suitability for agriculture and the cost-weighted distance to the nearest sea ports (as explained below, cost-weighting takes into account topography and ocean-navigable rivers). Our prior coming from the illustrative model above is that $\beta_2$ will be negative, given that larger distance from the coast increases $\rho$ in the model, thus decreasing the profitability of urbanization. The model gives no guidance on sign on $\beta_1$, however, since better land suitability for agriculture means that the area likely urbanized earlier and could achieve higher density ($\varphi$ on the left of the inequality) while contemporary city growth might avoid the opportunity cost of urbanizing the best agricultural land ($\varphi$ to the right of the inequality). The specification is in reduced form, of course, since many other variables will influence the location of urban land (infrastructure decisions, zoning, broader policy environments such as hukou[1] ) and these endogenous variables are partly determined

[1]Land use right reform in China has proceeded at varying speeds across provinces (Deng et al 2008), while Chinese urbanization and broader demographic dynamics are constrained by the hukou system. Since 1958, this household registration system has tightly controlled how many rural workers move into urban areas, and workers outside of their authorized area could not access government- or employer-issued rations, housing, or health care. As part of its economic liberalization, China has relaxed the hukou restrictions over the last two decades (rural residents can now buy urban residency permits). Nevertheless, the hukou system remains an obstacle to the free movement of people within the country, with measurable consequences for

by the geographical characteristics of interest. For our purposes of measuring associations and eventually doing prediction, however, we are interested in the total effect of the geographical variables on urban land distribution, as opposed to the effects conditional on partly endogenous variables such as infrastructure.

Similarly, we estimate these relationships in the case of urban expansion during 1990-2000, conditional on the initial distribution of urban land in 1990:

$$urbangrowth_i = \beta_0 + \beta_1 \cdot urbanarea_i^{1990} + \beta_2 \cdot landsuit_i + \beta_3 \cdot ln(costdistance_i) + \varepsilon_i \tag{3}$$

In the sections that follow, we describe the data, present the estimation methods that account for the spatial structure of the data and urbanization processes, discuss our results and conclude by projecting our estimates forward to offer predictions of the spatial distribution of urban land in China during the coming decades.

## 3 Data

The China Land Cover Dataset (CLCD) provides spatially explicit measurements of urban land cover change for the period 1990-2000 and is considered the country's primary national inventory of land cover (Liu et al., 2002, 2005). The dataset was derived from a mosaic of scenes from the Landsat TM/ETM sensors – 524 scenes from 1988-1990 and 512 scenes from 1999-2000. Scenes were selected for the CLCD based on low atmospheric effects and from the peak agricultural season in order to maximize the vegetation signal (Liu et al., 2005). The Landsat TM data have been aggregated into a 1X1 $km^2$ grid at the national level. Each cell in the grid indicates the proportion of its total (1X1 $km^2$ land area) that has been converted to urban infrastructure during the period 1990-2000[2]. The dataset also contains measurements of total existing urban land cover (also defined as a proportion of the 1X1 $km^2$ cell) for the years 1990 and 2000.

While previous work on urbanization processes has generally focused on the growth of urban population, this measure of the conversion of urban land directly captures the long-term capitalization of land rents that motivates our study. The measure also provides direct and spatially explicit information about the distribution of urban infrastructure, which is likely to persist due to the high fixed costs of installation and the durability of the housing stock; moreover the urban land use measure more directly relates to long-term ecological implications of urbanization. Satellite data provide consistent measurements across large

agglomeration (Au and Henderson 2006b).

[2]A value of 1 = 1% of a 1$km^2$ grid cell = 10,000 $m^2$

regions and jurisdictional boundaries[3]. At decadal time scales, measurements of urban growth captured by the Landsat and MODIS sensors have generally produced more consistent measurements than DMSP/OLS ("nightlights") due to saturation effects (for many dense urban areas the light sensor saturates at a maximum reading above which variations in light emissions are not detectable), sensor degradation, and inconsistencies across nightlights data captured by different sensors (each spanning between 2-8 years) (Small and Elvidge, 2013, Zhang and Seto, 2011). Multitemporal image processing algorithms, like those used to produce the measure of change in the CLCD, reduce the magnitude of measurement errors that are propagated when classified images are combined through simple differencing.

Other researchers have examined the relationship between agricultural fertility and urban expansion in China and elsewhere. In particular, we build on the work of Deng et al (2008) who estimate the relationship between agricultural land rents[4] and urban land expansion using the CLCD measurements. Causal inference based on these estimates is limited by the fact that agricultural land values are almost certainly endogenous to demand for agricultural products emanating from the urban areas themselves. The identification strategy employed in this paper is designed to deal with the endogeneity of agricultural rents in the process of urban expansion. We employ an index of land suitability for agriculture (which we will shorten to "land suitability" or "agricultural suitability" throughout the paper) created by Ramankutty et al. (2002) that combines data on climate (temperature and moisture availability) and soil (soil carbon density and soil pH) to model inherent suitability for cultivation. Importantly, this index does not measure agricultural production itself (which could be affected by demand from nearby urban centers), but instead measures the exogenous physical determinants of agricultural productivity. The index is mapped in Figure 1 for China.

We proxy for access to global markets using a purely biophysical measure of the cost-weighted distance to sea ports, which reflects the transport costs associated with travel across China and captures topographical constraints. Note that this measure is only a function of topography and location relative to ports and rivers, thus independent of infrastructure. The cost distance is calculated to the world's largest ports (226 ports) using Containerisation International's rank of the largest ports by traffic volume (Fossey, 2008). For each land cell, we calculate the nearest distance to a port, and allow transport from a location to a port to happen over land, over sea, or on a navigable river or lake. However, in order to find the optimal path to a port, the relative transport cost between land and water transport is required. Limão and Venables (2001)

[3]When scenes from high and medium resolution sensors are compiled in mosaics, discontinuities are sometimes observed across tile boundaries. The location of tile boundaries are independent of the processes of ecnomic interest and can be treated as classical measurement error.

[4]Deng et al (2008) use a measure of investment in the agricultural sector

use cost data on shipping a standard 40-foot container from Baltimore to different destinations around the world in 1990, and find that an extra 1,000 km by sea adds $190, whereas 1,000 km by land adds $1,380. This indicates roughly a 1:7 ratio between the cost of sea and land travel, which we use to construct the index. Using a map of navigable lakes and rivers, each land cell is assigned a transport cost of 7, and cells over oceans, seas, and navigable lakes and rivers are assigned a transport cost of 1. Finally, the index adjusts for terrain slope by using an FAO/IIASA map of median terrain slope; we impose an increasing penalty with slope, with a maximum fourfold penalty[5] when slope is at the highest category ($> 45\%$). The cost distance for China is mapped in Figure 2 below, and highlights the geographic isolation of China's western regions as well as the importance of river systems which open up parts of the interior of the country to oceanic trade. It is possible that port location is partially determined by city location, which poses a problem for identification. We develop a separate parameter that measures the cost-weighted distance to coastlines. The correlation between the ports variable and the coastlines variable is .98 and results are robust to these two versions of the variable.

All of our variables are aggregated to half-degree cells across China, since this is the resolution of the index of land suitability for agriculture. Visual inspection of the two independent variables of interest (land suitability for agriculture and cost-adjusted distance to ports) might lead one to think that they are highly correlated and therefore problematic for estimation. However, the correlation coefficient between the two after the premultiplication for weighted least squares is only -0.52. Table 1 presents the summary statistics for the variables in the study. Urban land cover and growth are expressed as percent of cell area. Note that many cells have very small (or zero) urban land cover, and the highest value for urban land cover is 35.5% (assuaging econometric concerns of upper censoring in the dependent variable). We note that while a tremendus amount of urban land expansion occured between 1990-2000, there are inherent limitations in modeling growth processes on decadal time scales. In particular, the variance in urban land cover growth is quite small ($s^2 = .48$) relative to that of the total observed land cover in 1990 ($s^2 = 12.18$).

## 4 Results and Discussion

We begin studying the spatial distribution of urban land area in 1990 with weighted least squares, followed by the MLE-based general spatial model, and finally explore spatial heterogeneity in the coefficients of interest first by interacting them with province dummies and alternatively by employing geographically weighted

[5] Results are all robust to reducing the penalty for steepest terrain from fourfold to twofold.

regression. Our results provide robust evidence confirming the basic prediction that urban areas tend to locate where land is better suited for agriculture and with lower transport costs to ocean-based trade. However, agricultural suitability is negatively associated with urban expansion during the contemporary period (1990-2000). The models that allow for spatial heterogeneity show a more nuanced picture: the positive association between urban area and agricultural productivity occurs mainly in the coastal provinces and an opposing relationship is found in some of the most urbanized provinces. This suggests that the opportunity cost of urbanizing good agricultural land becomes more of a factor at higher levels of urban coverage. Province-level and local GWR estimates suggest that the power of the geographical variables in explaining urban growth from 1990-2000 is limited compared to their power in explaining distribution of urban areas in 1990.

## 4.1 Least Squares Estimates

We begin by modeling the level of urbanization in 1990. Our basic (spatially naïve) model is specified in equation (2). We employ Weighted Least Squares in all specifications since there is significant variation in cells' land area (fixed 0.5 degree cells decrease in land area in proportion to the cosine of latitude, moreover coastal cells and border cells will have smaller areas, sometimes almost no land). It is likely that these smaller areas exhibit greater variance in outcomes, so we weigh each observation by its land area to address heteroskedasticity. This is implemented by premultiplying all of the data by the square root of each observation's land area. Note that results are qualitatively unchanged if Ordinary Least Squares is used instead.[6] Our dependent variable of interest – urbanized percentage of the cell land area – is a limited dependent variable with a minimum value of zero and a theoretical maximum value of 100, although no observations are actually 100% urbanized. We report the results of a Tobit specification with a lower bound of zero.[7]

Least squares results are reported in columns (i)-(iii) of Table 2, with and without the Tobit specification . Column (iii) includes province fixed effects that absorb any province-level omitted variables that might be correlated with our independent variables of interest and that might affect urbanization. All estimates support the hypothesis that urban areas are more likely to be located in land characterized by prime agricultural conditions (climate and soil) and access to international markets (low transport costs to sea ports). In the theoretical framework described earlier, this would indicate that agglomeration forces driving

[6] Results available upon request

[7] We also estimate the regression omitting all cells with zero urbanization; the results are unchanged.

cities to expand into nearby agricultural land have empirically dominated the "opportunity cost effect" that might have steered urban expansion away from fertile agricultural land. The effect is interpreted as follows: a grid cell with optimal biophysical conditions for agriculture exhibits more urban land cover (from 1.5-4% of grid cell area) as compared to a completely infertile cell; alternatively, a one standard deviation better land suitability is associated with 0.2-0.5 of a standard deviation in urban land cover. Meanwhile, a 100% increase in the cost-adjusted distance to ports is associated with a lower level of urban land cover (by 1 – 1.6% of grid cell area, or 0.3-0.5 of a standard deviation).

We are interested in how the marginal cost of port distance might decay.We estimate a polynomial fixed-effects specification where the ln(cost distance) variable is replaced by linear and quadratic terms. The resulting coefficients of -198 and 4413 indicate decreasing marginal effects of distance on urban land cover, with a minimum at 22,430 units. There are 992 cells (25% of the country's land area) that have a higher value for cost-distance; we interpret this nonlinearity as an indication that for these inland regions each marginal unit of distance is not affecting variation in urban land cover. Though the quadratic specification yields strong significance, we opt for the log specification because it yields higher R-squared values and is thus a tighter fit for prediction. Proximity to sea ports and agricultural conditions play an important role in determining the amount of urban development in any given location across China. This extremely parsimonious (and spatially naïve) model suggests that the two geographic variables explain 46% of the variance in urbanization levels across China. [8]

The results above indicate that agricultural potential and market access played an important role in determining where urban areas are located and their historical growth. However, it is unclear whether these factors will determine where modern-day urban growth will occur. Models that emphasize the importance of agricultural fertility as a catalyst of urban agglomeration generally feature a regional economy where trade is limited. In regions undergoing fundamemental urbanization processes in the 20th and 21st centuries, modern transport costs have altered the tradeoff between natural growth of cities in fertile areas and the opportunity cost of urbanizing fertile areas, since food can be brought in from afar. This is a critical question for China, where the majority of growth will occur in the modern period and the ultimate distribution of urban land will reflect modern constraints.

As with urban levels, we estimate the growth model using OLS, the suite of spatial models, and geographically

[8] The fixed effects model has an R-squared of 0.62, but the within R-squared excluding the explanatory power of the dummies is 0.14

weighted regression; the results for OLS are shown in columns (i) and (ii) of Table 3 below. The results reported in Table 3 indicate that urban growth is more likely to occur in cells that already contain urban areas and more likely to occur in areas with good market access (close to large ports). Unlike the levels regression above, however, the land suitability coefficient has a negative sign, indicating that areas with favorable conditions for agriculture are less likely to experience urban growth from 1990-2000. This result reflects the ambiguity expressed in our motivational model. The explanation from that model is that urbanization occurred historically in areas where food could be grown plentifully nearby, resulting in a positive correlation between urban location and agricultural productivity. Large improvements in transport technology during the last century may have altered the nature of this constraint. The opposing force is the opportunity cost of converting the most productive agricultural lands, where agricultural activities have concentrated. The observed effect is that lands with higher agricultural potential in regions near existing urban centers are differentially slower to urbanize than equally proximate and less agriculturally productive lands. This illustrates a critical tension in China's land economy – productive agricultural lands must compete with increasing returns to the growth of historical centers whose initial location depended on fertility. Column (ii) adds province dummies to absorb province-wide omitted variables. The result differs from column (i) in that land suitability is no longer significant, and the coefficient on cost-distance to ports doubles in magnitude. Since the fixed-effects specification reduces identification to within-province variation, it may be the case that more than a single decade of observations of land cover change is required to construct a sufficiently powerful test.

There is potential concern with regards to temperature: comfortable weather is an amenity that might drive urbanization and has been shown to be positively corrlated with urban sprawl in the Uinited States (Burchfield et al 2006). Since temperature is part of agricultural suitability, then the interpretation of the positive impact of agricultural suitability on urban land expansion might be misplaced. It might be that the temperature amenity, and not agriculture, is the relevant condition for urbanization processes. The correlation coefficient between temperature and the land suitability for agriculture variable was 0.54, which suggests that land suitability captures much more than just temperature. We test the independent effect of temperature in growth and levels specifications and find results unchanged. Given that temperature is not of key interest in our study and that including temperature does not improve model fit, we exclude temperature from the models discussed throughout the paper[9].

The least squares model is likely to present biased estimates and/or inflated t-statistics in the presence

[9]Results available upon request

of spatial dependence and autocorrelation of errors. A Lagrange Multiplier test provides evidence of spatial dependence in the OLS residuals (significant to 1%) by comparing models with and without a spatial dependence term. Meanwhile, the spatial Durbin-Watson statistic, which tests the correlation of residuals to nearest neighbors, finds strong presence of spatial autocorrelation (values below 2.0 indicate positive autocorrelation [Gujarati 2003] and as a rule-of-thumb values below 1.0 indicate problematic autocorrelation). The Moran's I statistic also indicates strong positive autocorrelation in both levels and growth models (within 1% statistical confidence). Following Conley (1999), we adjust standard errors using a nonparametric estimator that is spatial heteroskedasticity autocorrelation consistent (SHAC) and find that our results are robust to this correction. These are presented in square brackets in the least squares regressions of Tables 2 & 3.

## 4.2 Spatial Dependence

The distribution of urban growth is determined by geographic conditions as well as a variety of highly localized processes. These include the development of transportation infrastructure, the effect of changing amenities and zoning regulations on the housing market, and highly localized mechanisms underlying agglomeration, such as input sharing, learning effects, innovation spillovers, employment pooling and matching (Amiti and Cameron 2007, Costa and Kahn 2000, Duranton and Overman 2005, Duranton and Puga 2004, Ellison et al 2010). These (unobserved) mechanisms present challenges for estimating our empirical models, as they imply that patterns of urban expansion are generated by a spatially dependent process. For example, a cell twice-removed from a city boundary is unlikely to urbanize due to the underlying cost-benefit of conversion. However, once the city grows and reaches the neighboring cell, the market access of the non-urbanized cells increases and the likelihood of urbanization changes dramatically. Under spatial dependence, the spatially lagged dependent variable causes the OLS estimator to become biased and inconsistent (Anselin, 1988). Our diagnostics reveal that OLS fails to take into account both (1) spatial autocorrelation in the residuals (using both a spatial Durbin-Watson statistic and Moran's I statistic) as well as (2) spatial dependence (using a Lagrange Multiplier test, following LeSage [1999]). The SHAC standard errors (Conley 1999) presented in the least squares regressions of Tables 2 & 3 are robust to (1) but not (2).

We test the ability of a series of spatial autoregressive models proposed by LeSage (1999, 2009) to correctly identify spatial dependence underlying patterns of urban expansion. These models include a spatial error model (SEM), which allows for spatial structure in the error term, and a mixed autoregressive-regressive

(SAR) model, which allows for direct effects on neighbors. One concern about the use of spatial autoregressive models is that results are sensitive to the spatial dependence structure specified by the econometrician; since the spatial structure is seldom given by theory or easily measured, misspecification can lead to inconsistency in parametric estimators (Conley and Molinari 2007, Griffith and Lagona 1998). We acknowledge the concerns regarding these MLE-based spatial econometric models and proceed by employing them alongside OLS with SHAC standard errors to evaluate the robustness of our primary findings. Results are qualitatively consistent across OLS and MLE-based spatial models.

We first augment the basic models by allowing for autocorrelated errors (the spatial error model) and then we allow dependency across cells (the mixed autoregressive-regressive model). Following LeSage (1999), we use Lagrange Multiplier and Likelihood Ratio tests to compare across models. In both cases (results for urban levels in Table 4 and for urban growth in Table 5), Lagrange Multiplier and Likelihood Ratio tests on model residuals indicate that these models fail to properly account for the spatial structure in the data[10] [11]. We therefore estimate the general spatial model specified below, in the case of urban levels in 1990:

$$urbanarea_i = \rho W_1 urbanarea + \beta_0 + \beta_1 \cdot landsuit_i + \beta_2 \cdot ln(costdistance_i) + u_i \tag{4}$$

$$u_i = \gamma W_2 u + \epsilon_i \tag{5}$$

$$\epsilon \sim N\left(0, \sigma_\epsilon^2\right) \tag{6}$$

In this model, $\rho$ is a spatial lag parameter that evaluates the effect of immediate neighbors' urbanization and is estimated using a standardized spatial contiguity matrix $W_1$. The errors from (4) are assumed to have a spatial structure, which is estimated by the parameter $\lambda$ using a spatial weight matrix W. We test two different matrices for $W_2$, allowing us to account for multiple assumptions regarding the nature of spatial structure in urbanization patterns (LeSage 2002). We first construct a distance-to-neighbor matrix using cells within 300 km. This is unlike the binary $W_1$ matrix which simply codes for immediate neighbors. As an alternative, we construct a $W_2$ coding for second-degree neighbors. Of course the growth of economic centers is endogenous to regional urbanization processes and the $\lambda$ parameter has no causal interpretation.

[10] In the case of the model on urban levels, a Lagrange multiplier test on the residuals of the spatial autoregressive model yields a test statistic of 678156.57***. Likelihood ratio tests on the residuals of the two spatial autoregressive models yield test statistics of 4137.03*** and 3327.54***, respectively (***, p<.01).

[11] In the case of urban growth, a comparison of Moran's I statistics indicate that OLS, SAR, and SEM model residuals exhibit high spatial autocorrelation (near 1) whereas the residuals from general spatial models in table 3 are distributed randomly (near 0). A Lagrange multiplier test on the residuals of the spatial autoregressive model yields a test statistic of 21905.15***. Likelihood ratio tests on the residuals of the two spatial autoregressive models yield test statistics of 1860.89*** and 971.51***, respectively (***, p<.01).

It does, however, provide insight into the nature of spatial dependence in a model of urban expansion.

Results from the general spatial model of level of urbanization are shown in columns (iv)-(v) of Table 2. Column (iv) reports estimates from a general spatial model that specifies $W_2$ using the distance-to-neighbor weighting matrix and column (v) uses a second-degree contiguity matrix. The results are robust to both specifications. Urbanization is more likely to have occurred in areas with good land suitability for agriculture (a very fertile cell is around 1% more urbanized than an infertile cell), and better access to ports (cells twice as far from ports are $\sim$1.3% of cell area less urbanized, or $0.4\sigma$). The magnitude of the point estimate on land suitability is about 50% as large as the OLS estimate and 70% as large as the fixed-effects estimate, which suggests a possible upward bias due to spatial dependence in OLS. Points estimates for port access are similar to those of fixed effects models. The validity of the general spatial model is supported by lambda (the parameter for spatial error structure) and rho (the spatial dependence parameter), which capture spatial structure in the model and in the errors. Moran's I statistics for both general spatial models are highly significant and near-zero, indicating that residuals from this model exhibit spatial randomness (comparing them to a Moran's I of 0.7 for the least squares results illustrates the value added of spatial econometric methods). By capturing the impact of market access and agricultural fertility across China and accounting for the spatial dependence that operates at a more local level, this spatial model can explain nearly 80% of the variation in urban development across China, which is encouraging for prediction applications.

The General Spatial Model results for growth between 1990-2000 are shown in columns (iii) and (iv) of Table 3, with the two $W_2$ matrices reported as above. The rho parameter is significant in both regressions, indicating a positive spillover effect of urban growth on the urban growth of neighboring cells. As with the levels results, the Moran's I statistics improve markedly when moving from least squares to the spatial models. While the signs of the point estimates are consistent and as expected (urban growth is positively associated with baseline urban coverage, and negatively associated with land suitability for agriculture and with distance to ports). However, the significance of the two independent variables of interest is not robust to the choice of the weighting matrix. Magnitudes are comparable to the point estimates of least squares. Column (iii) suggests that cells with a 100% higher cost-distance to ports have less new urban areas (by around 0.08% of cell land area) after controlling for initial urban area; column (iv) suggests that an area perfectly suitable for agriculture will have less growth (by around 0.12% of cell land area) compared to a cell that is completely inadequate for agriculture. These parsimonious models explain 57% of the variation in urban growth rates, which again is encouraging for use in prediction.

### 4.3 Spatial Heterogeneity in Effects

The theory that motivates our study expresses ambiguity in the empirical relationship between agricultural suitability and urban development. Results presented above indicate the role of agricultural fertility in determining the geography of urban expansion has changed in China. However, there are a range of unobserved variables that vary across China's diverse regions that might bias these estimates, such as provincial government economic and social policies, variation in enforcement of national laws, or variation in social and cultural norms (some provinces were created along cultural boundaries). We therefore estimate a model including province fixed effects to absorb unobserved variation at the provice level, as well as interacting province dummies with land suitability for agriculture in order to look at spatial heterogeneity in marginal effects. The results are displayed in Figure 3.

The differences across provinces are stark. Land suitability for agriculture is significant only in the eastern part of the country, and the province-level coefficient goes as high as 43 (keep in mind that the estimates above which assume spatial homogeneity were on the order of 1.0 in the MLE spatial models and 1.5 in least squares model with province fixed effects). Moreover, three coastal provinces indicate a negative association, which would suggest that the opportunity cost of urbanizing good land is a first-order concern. It is notable that these three provinces (Beijing, Jiangsu and Tianjin, with 9%, 12% and 14% urban land cover, respectively) are three of the six most urbanized provinces. This suggests a threshold of urban land cover above which the opportunity cost of urbanizing good agricultural land truly constrains spatial patterns of urban growth. Meanwhile, the cost-distance to ports also displays a systematically different association with urban land cover in 1990. The variable is significant in all provinces, but the elasticity is stronger at the coasts by an order of magnitude (up to -18.9 in Beijing, compared to the -1.6 estimated from the fixed effects model and -1.3 in the MLE spatial models). Clearly, relaxing the assumption of homogenous effects across the country will provide further understanding of the dynamics underlying urban land use.

Province-level results for growth from 1990-2000 are presented in Figure 4. The spatial patterns are less systematic than in the case of the levels regression; favorable land suitability for agriculture is positively associated with urban growth in around half of the provinces, though northern provinces and Beijing, Tianjin, Hubei and Shandong display negative associations. Meanwhile, cost-distance to ports is negatively associated with urban growth in many provinces (with Beijing and Jiangsu having the steepest elasticities), but much of the country has no significant relationship and two provinces have a surprising (perhaps spurious) positive association between distance to ports and urban growth. Shandong's positive coefficient defies the theories we have discussed here, while Xinjiang's coefficient might simply be a border effect: areas in that province farther from a port are nearer to the international border and perhaps closer to overland trading partners abroad. Note, however, that the magnitude of the surprising positive coefficient in Shandong and Xinjiang is very small, even compared to the negative counterparts.

The estimates above relax the homogenous effects assumption at the national level, but continue to assume homogenous effects within a province. We have no theoretical prior over our assumption that the province is the dominant spatial unit determining the structure of the urbanization phenomenon and we view this as an empirical question. We allow for more localized heterogeneity in marginal effects using geographically weighted regression (GWR). This technique provides local estimates of the model by iteratively defining a sample at every grid cell and using an adaptive spatial kernel to define the geographic sample that maximizes efficiency in terms of the Akaike Information Criterion (Fotheringam et al., 2002). Note that GWR implies a tradeoff: local models are hindered in regions where there is little variation in the data. This applies to western China, where there are vast areas of desert with no urbanization. With a lack of variation, the GWR technique does not identify any effect of geography on urbanization, whereas the global model makes it obvious that China's western deserts are important predictors of the lack of urbanization in those areas.

The optimal kernel was determined to be 206 observations, which corresponds to a circle with a radius of 4 decimal degrees (around 440 km). The two bottom maps in Figure 4 show the parameter estimates at every location for those areas where they were significant. Critical values for the t-statistic increase under GWR, since any method that involves multiple hypothesis tests (one for each data point) must adjust for the expected false positive rate of doing a statistical test multiple times. The adjustment therefore keeps a population-wide error of 5% by adjusting the test-level critical level. In GWR, this "Fotheringham adjustment" varies by the degrees of freedom, and for our model the critical value is 3.56 (Byrne et al. 2009).

The bottom-left map in Figure 4 shows that increased cost distance to ports has a marginal effect in reducing the urban land cover in 1990, but only up to a maximum distance. Near Tianjin and Beijing, for example, a 100% higher cost-adjusted distance to ports is associated with less urbanization by up to 7.8% of cell area. Beyond 1000 km inland (the left edge of the blue area on the bottom left map) the port distance variable has no explanatory power in explaining local variation in urbanization. In this central area of the country, terrain changes significantly (notice how the agricultural suitability index in Figure 1 changes drastically in this area) and cost-weighted distance is so large that the marginal effect of an extra kilometer is zero.

The bottom-right map in Figure 5 shows the local coefficient for land suitability for agriculture. The index is positively associated with urbanization in most coastal areas of the country. In fact, the highest elasticity is in Shandong province, where a fertile area with an index of one has up to 49% more of its cell area urbanized than a completely infertile area, a much larger effect size than estimates with "global" models that assume homogenous marginal effects. Interestingly, the impact of agricultural suitability exhibits spatial heterogeneity in both magnitude and in sign. GWR results suggest that agricultural suitability in Hebei province actually has a negative association with urban location, suggesting that the opportunity cost of developing on productive agricultural lands may have had an important impact on urban growth in this region.

In terms of urban expansion between 1990-2000, GWR estimation performs well relative to OLS estimation [12] – results are mapped below in Figure 6. This provides further evidence of the importance of spatial heterogeneity in the affects of geographic parameters on urban land expansion. The optimal kernel size was computed to be the 206 nearest neighbors, which corresponds to a circle with a radius of 4 decimal degrees (around 440 km). The map on the top left displays the local coefficient of the 1990 urban land cover variable. As expected, new urban areas will tend to occur in cells that already contain some urbanization. The power of agglomeration accelerates this process as the urban areas gets larger, hence the positive coefficient on that variable. Initial urban land cover is a strong determinant of urban growth in much of urbanized China. Urban growth has the highest elasticity to initial urban area in Hebei, Sichuan and Guangdong provinces. Notably, no areas showed a negative coefficient providing no evidence that disammenities of urban areas have led to saturation in China's largest urban areas.

It is important to note that the region of statistical significance for the two variables of interest is relatively limited, largely due to the fact that the majority of China's land surface experienced little or no urban land cover change during the 1990s and thus there is little variance for the model to explain in those regions. This is an important empirical consideration in work on urban land expansion, given the very small portion of urban land cover on the terrestrial surface. One common approach is to constrain the sample of observations to areas based on a priori assumptions about regions of importance or areas of greatest variance in the dependent variable. Local estimation allows the econometrician to test any assumptions about study boundaries. Furthermore, in this case GWR demonstrates countervailing marginal effects of

[12] The residual sum of squares drops from 2,937,374 to 1,700,790, the AIC drops from 37,793 to 35,942, and the adjusted R-squared increases from 0.33 to 0.59.

both land suitability and distance to ports in regions where urban growth was concentrated, indicating that an assumption of homogenous marginal effects may not be tenable in the case of urban growth.

The impact of land suitability for agriculture and the cost distance to ports on contemporary expansion are pronounced in particular regions. We find significant a positive marginal effect of port access on urbanization in the Yangtze River Delta region of greater Shanghai (Jiangsu, Zhejiang and Anhui provinces), but not in most of the rest of the country. Land suitability, meanwhile, has opposing effects in the north and south. In Guandong and Jiangsu, agricultural suitability is associated with greater urban land conversion, while the opposite is true in Shandong and Hebei. Urban areas in the Pearl River Delta region grew faster near agriculturally suitable lands in the 1990s, while productive agricultural areas may have presented a high opportunity cost to urbanization in most of the North. Finally, the map displayed at the bottom right indicates low clustering for nonzero residuals.

Recognizing and understanding these heterogeneous effects is critical for testing hypotheses about what drives urbanization as well as for modeling future land change. Visual comparison of the data to the predicted urban land area (the top two maps in Figure 4) suggests that the GWR model performs rather well in explaining the spatial distribution of Chinese urban areas. Comparison of the models using the residual sum of squares, AIC, and the adjusted R-squared indicates that the GWR model provides a better fit to our data than OLS.[13] The GWR output also reports the results from a Monte Carlo simulation which rearranges the data spatially to test how likely the measured spatial heterogeneity would occur from random distributions of the data. In this case, the Monte Carlo significance test rejects the null of no spatial heterogeneity for each of the parameters to the 1% confidence level.

These results are consistent with the hypothesis that geographical variables play a strong role in determining city location (as in Motamed et al., 2009). With regards to city growth, the distribution of existing urban lands clearly determines the geography of land conversion. Conditional on the distribution of China's urban lands in 1990, the effects of exogenous physical factors on decade-on-decade expansion vary substantially across and within regions.

## 4.4 Future Urban Growth in China

We conclude our analysis by using the growth model calibrated using Geographically Weighted Regression to extend predicted values for urban land percentage beyond 2000. The maps in Figure 7 below present

[13] Residual sum of squares of 27,846,168 for GWR as opposed to 71,015,456 for OLS; Akaike Information Criterion of 47,063 as opposed to 50,545 for OLS; and adjusted R-squared of 0.74 as compared to 0.46 for OLS

the actual and predicted values for urban land area in 2000, and illustrates the fit generated by this local estimation method. We highlight areas where the model has little predictive power by outlining in red cells where neither the beginning-of-period urbanization level, the land suitability variable nor the cost distance to ports variable are significant. These red zones, therefore, can be thought of as areas where our local model has no predictive power. Although most of the country is in the red zone, in fact we expect that models of urban growth will be relevant for very specific geographic regions since urban areas occupy between 0.5-3% of Earth's terrestrial surface (Schneider et al., 2009).

We generate projections by substituting the observed 2000 urbanization levels as the beginning-of-period variable in the fitted model to generate projected urbanization levels for 2010, 2020, and 2030.[14] The maps suggest increasingly concentrated urbanization in the northeast as well as significantly more urban areas of Chengdu and Chongqing in the West. Table 6 reports projections for 2020 and 2030 by province. Shandong, Hebei, and Jiangsu are expected to experience the largest absolute increases in urban land expansion over the next two decades. Sichuan and Chongqing are expected to see the highest relative growth rates, with urban land areas more than doubling in both provinces between 2000-2030. Our analysis demonstrates that agricultural suitability, access to international markets, and the historical distribution of urban areas all are important determinants of urbanization patterns and that a model that allows for heterogeneous effects may capture the impact of these variables more fully. This evidence supports the role of local estimation for advancing methods for forecasting urban growth, which play a critical role in planning for 1 billion urban residents in China by 2050.

# 5 Conclusion

The results of this study provide new and unique estimates for how physical geography impacts the current-day distribution of urban land cover through access to international markets and agricultural suitability. We highlight ambiguity in the economic geography literature regarding the expected empirical relationship between biophysical agricultural suitability and urban development and distinguish between estimating the role of geography in levels of urban land area (which reflect both city genesis and historical growth) and the role in modern-day urban expansion. Urban development in China has tended to locate in fertile areas with good access to international markets. The marginal effect of port access drops off relatively quickly and is therefore negligible for roughly 1/4 of China's land mass. However, it has played a critical role in shaping urban centers in the East. In the case of modern-day urban expansion, the direct impact of access to ports

[14]The fitted model uses locally estimated coefficients for all the variables

is less substantial and the opportunity cost of agricultural fertility may outweigh its impact as a catalyst for agglomeration. This is particularly true of northern China. This distinction has important implications for understanding land use policy: fertile lands help to catalyze agglomeration and exert significant influence over city location. However, the opportunity cost associated with converting them becomes important as urban and agricultural economies mature. Increasing returns to city growth in areas surrounded by prime farmland now creates a tension in the land market. This very real tension has led to an enormous political debate within China as the country's prime farmlands are converted into new urban lands around the most productive urban centers. The central government has attempted to regulate the process of farmland conversion, with a variety of potential impacts on productivity (see Lichtenberg and Ding 2008 for review, Ding and Lichtenberg 2011).

The empirical results presented here reflect careful consideration of econometric estimation with spatial data and inherently spatial processes, using a variety of techniques, discussing their weaknesses, and assessing robustness across them. We acknowledge some important limitiations of our data, particularly with the measurement of urban land change within a single decade. Estimation will improve substantially as optical satellite archives continues to grows. We present a series of models and tests that demonstrate that our primary findings are robust to spatial autocorrelation and spatial dependence as well as bias from omitted variables (such as temperature) or variable misspecification (distance to coast vs. distance to ports). We find a great deal of heterogeneity in the marginal effect of geographic constraints across China, both within and across provinces. This suggests that estimation of average effects at the national level might have some important limitations, the most critical being conflation of lack of statistically significant effects and opposing effects across regions. The literature on forecasting patterns of urban land cover change currently focuses on national level population and productivity growth, without a sophisticated parsing of constraints and catalysts that determine the spatial distribution within a country. We find that this parsimonious model of exogenous geographical determinants of urban growth explains a substantial amount of the variation in urban land cover across China, suggesting that these parameters may have an important role to play role in future forecasting exercises, and more generally that the geographical landscape continues to play profound role in the spatial distribution of human population and economic activity.

## References

Alonso, W. 1964. Location and Land Use. Harvard University Press: Cambridge, MA.

Amiti, Mary and Lisa Cameron. 2007. Economic geography and wages. Review of Economics and Statistics 89(1):15–29.

Anselin, L. 1998. Spatial Econometrics: Methods and Models. Dorddrecht: Kluwer Academic Publishers.

Anselin, L. and D.A. Griffith. 1988. Do spatial effects really matter in regression analysis? Papers of the Regional Science Association 65, pp. 11-34.

Au, Chun-Chung and J. Vernon Henderson. 2006a. Are Chinese Cities too Small? Review of Economics and Statistics 73, pp. 549-576.

Au, Chun-Chung and J. Vernon Henderson. 2006b. How migration restrictions limit agglomeration and productivity in China. Journal of Development Economics 80(2), pp. 350-388.

Banerjee, A., Duflo, E., and Qian., N. 2012. On the Road: Access to Transportation Infrastructure and Economic Growth in China. Working Paper.

Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., and Zhang, Q. 2012. Roads, Railroads and Decentralization of Chinese Cities. Working Paper.

Bernard, A.B., Eaton, J., Jensen, J.B., and Kortum., S. 2003. Plants and Productivity in International Trade. American Economic Review 93(4): 1268-1290.

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. 2012. Relaxing Hukou: Increased labor mobility and China's economic geography. Journal of Urban Economics, Volume 72, Issues 2–3, Pages 252–266.

Bruekner J. 1987. The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model, in Edwin Mills (ed.), Handbook of Regional and Urban Economics, Volume II, Amsterdam: Elsevier, 821-845

Brueckner, J.K. and Fansler, D.A. 1983. The economics of urban sprawl: Theory and evidence on the spatial sizes of cities, Review of Economics and Statistics 65, 479–482.

Burchfield, M., Overman, H.G., Puga, D., and Turner M.A. 2006. Causes of Sprawl: A Portrait from Space. Quarterly Journal of Economics 121(2): 587-633.

Byrne, G., Charlton, M. and Fotheringham, S. 2009. Multiple Dependent Hypothesis Tests in Geographically Weighted Regression, in Lees, B.G. and Laffan, S.W. (eds), 10th International Conference on GeoComputation, UNSW, Sydney, November-December.

Combes, PierrePhilippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. Estimating agglomeration effects with history, geology, and worker fixed effects. In Edward L. Glaeser (ed.) Agglomeration Economics. Chicago, il: Chicago University Press.

Costa, Dora L. and Matthew E. Kahn. 2000. Power couples: Changes in the locational choice of the college educated, 1940-1990. Quarterly Journal of Economics 115(4):1287–1315.

Davis, Donald R. and David E. Weinstein. 2002. Bones, Bombs, and Break Points: The Geography of Economic Activity. The American Economic Review 92(5): 1269-1289.

Deng, Xiangzheng, Jikun Huang, Scott Rozelle, and Emi Uchida. 2008. Growth, Population and Industrialization, and Urban Land Expansion of China. Journal of Urban Economics 63 (1): 96–115.

Ding, Chengri, and Erik Lichtenberg. 2011. Land and Urban Economic Growth in China. Journal of Regional Science 51 (2): 299–317.

Duranton, Gilles and Henry G. Overman. 2005. Testing for localization using microgeographic data. Review of Economic Studies 72(4):1077–1106.

Duranton, Gilles and Diego Puga. 2004. Microfoundations of urban agglomeration economies. In Vernon Henderson and Jacques François Thisse (eds.) Handbook of Regional and Urban Economics, volume 4. Amsterdam: NorthHolland, 2063–2117.

Ellison, Glenn, Edward L. Glaeser, and William Kerr. 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. American Economic Review, 100(3): 1195-1213.

Fossey, John. 2008. Containerisation International Yearbook 2008. Informa Maritime & Transport : London, UK.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Chichester: Wiley.

Fujita, M., and Thisse, J.F. 1996. Economics of Agglomeration. Journal of the Japanese and International Economies, 10: 339-378.

Fujita, M., Krugman, P., and Venables, A. 1999. The Spatial Economy: Cities, Regions and International Trade. MIT Press, Cambridge, MA.

Gollin, D., Parente, S., and Rogerson, R. 2002. The role of agriculture in development. The American Economic Review 92(4), 1205-1217.

Gujarati, D.N. 2003. Basic econometrics, 4th ed., Boston, McGraw–Hill.

Krugman, P. 1991. Increasing Returns and Economic Geography. Journal of Political Economy 99: 483-489.

Kuznets, S. 1973. Modern economic growth: Findings and reflections. The American Economic Review 63(3), 247-258.

LeSage, J., Pace, R.K. 2009. Introduction to Spatial Econometrics. Taylor and Francis Group: Boca Raton, FL.

LeSage, James P. 2002. Applied Econometrics using MATLAB. A manual to accompany the econometrics toolbox at: www.spatial-econometrics.com

LeSage, James P. 1999. The Theory and Practice of Spatial Econometrics. Mimeo. Lichtenberg, Erik, and Chengri Ding. 2008. Assessing Farmland Protection Policy in China. Land Use Policy 25 (1): 59–68.

Lichtenberg, E., and Ding, C. 2009. Local officials as land developers: Urban spatial expansion in China. Journal of Urban Economics, 66(1), 57–64.

Limão, Nuno and Anthony J. Venables. 2001. Infrastructure, Geographic Disadvantage, Transport Costs, and Trade. The World Bank Economic Review. Vol. 15, No 3.

Liu, J., Tian, H., Liu, M., Zhuang, D., Melillo, J., and Zhang, Z. 2005. China's changing landscape during the 1990s: Large-scale land transformations estimated with satellite data. Geophysical Research Letters, Vol 32: 2405.

Lucas, R. E. 2000. Some macroeconomics for the 21st century. The Journal of Economic Perspectives 14(1), 159-168.

McGrath, D.T. 2005 More evidence on the spatial scale of cities, Journal of Urban Economics 58, 1–10.

Murata, Y. 2008. Engel's Law, Petty's Law, and agglomeration. Journal of Development Economics 87(1), 161-177.

Mills, E.S. 1967. An Aggregative Model of Resource Allocation in a Metropolitan Area, American Economic Review 57(2): 197-210.

Motamed, M., Florax, R., Masters, W. 2009. Geography and Economic Transition: Global Spatial Analysis at the Grid Cell Level. Paper presented at the Agricultural & Applied Economic Association 2009.

Muth, R. 1969. Cities and Housing. Chicago: University of Chicago Press.

Ottaviano, G.I.P., Tabuchi, T. and Thisse, J.T. 2002. Agglomeration and Trade Revisited. International Economic Review 43, 2:409-435.

Ramankutty, Navin, Jonathan A. Foley, John Norman and Kevin McSweeney. 2002. The global distribution of cultivable lands: current patters and sensitivity to possible climate change. Global Ecology & Biogeography 11:377-392.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." The Quarterly Journal of Economics 125 (3) (August 1): 1253–1296. doi:10.1162/qjec.2010.125.3.1253.

Schneider, A., Friedl, M.A., and Potere. D. 2009. A new map of global urban extent from MODIS satellite data. Environmental Research Letters, 4 p. 044003

Small, Christopher and Christopher D. Elvidge. 2013. Night on Earth: Mapping decadal changes of anthropogenic night light in Asia. International Journal of Applied Earth Observation and Geoinformation. 22:40-52.

Storeygard, Adam. 2012. Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. Working Paper.

United Nations. 2012. World Population Prospects: The 2010 Revision and World Urbanization Prospects: The 2011 Revision. Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. Accessed October 12.

Figure 1: Index of Land Suitability for Agriculture

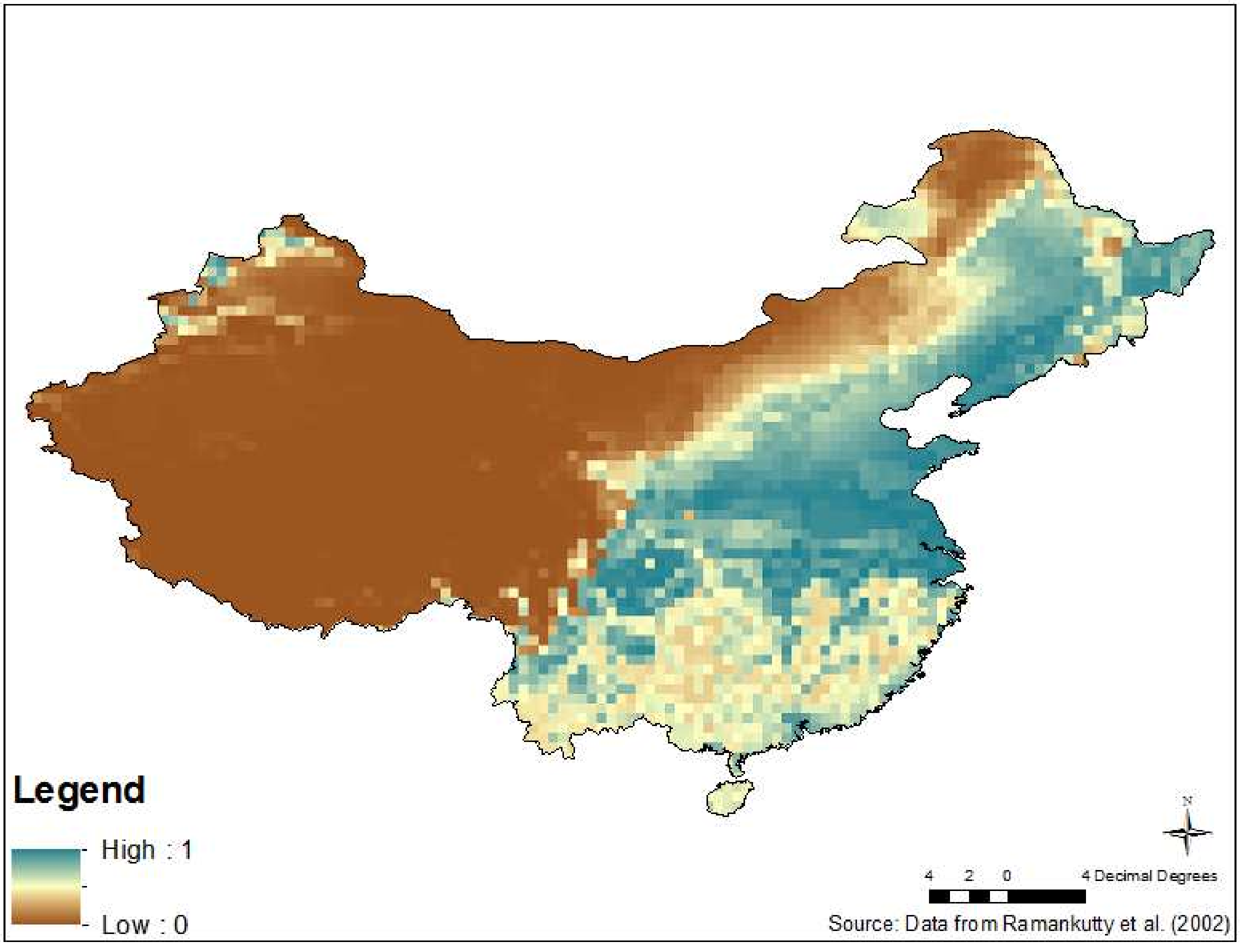


Figure 2: Cost-Adjusted Distance to Ports

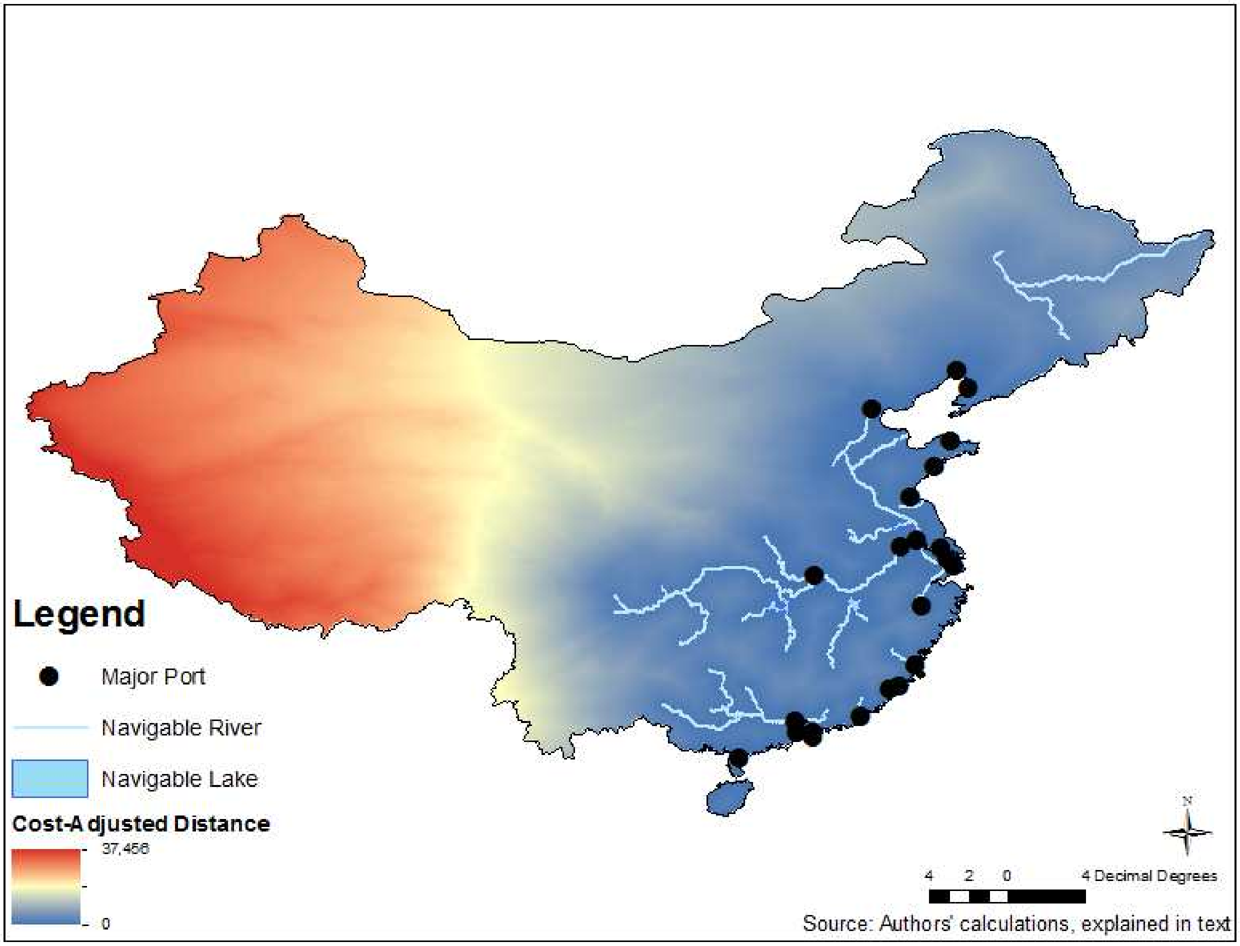


Figure 3: Marginal Effects on Urban % of Cell in 1990 (by province)

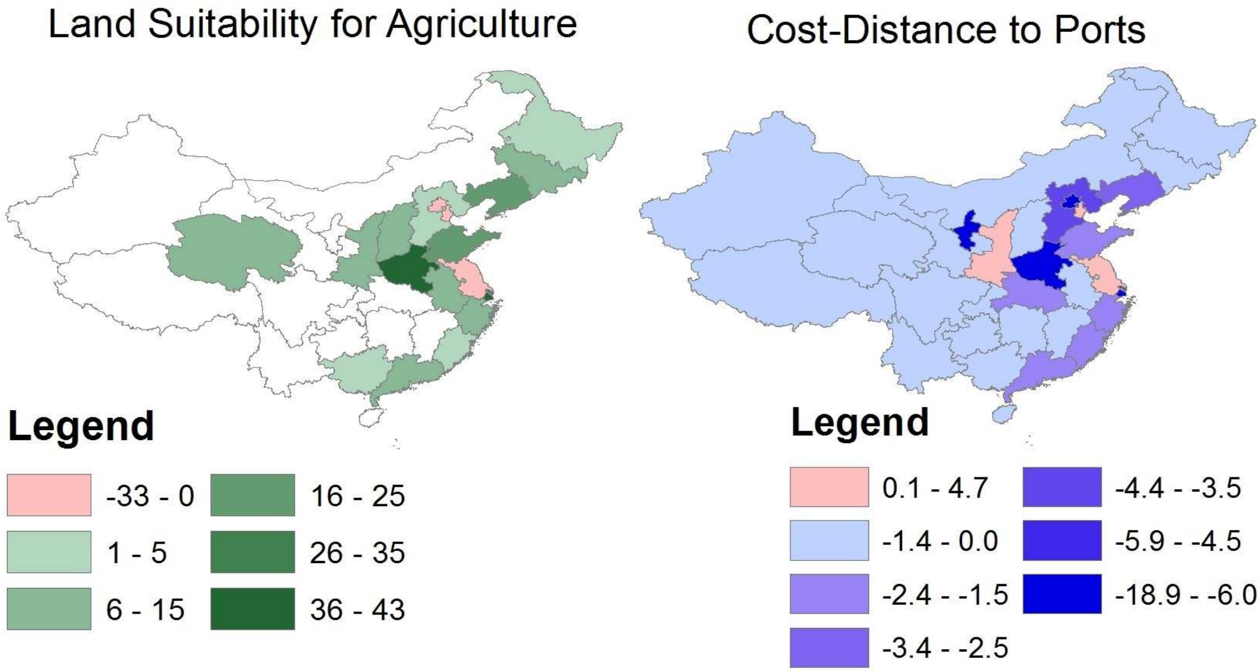


Figure 4: Results from GWR for Urban Levels

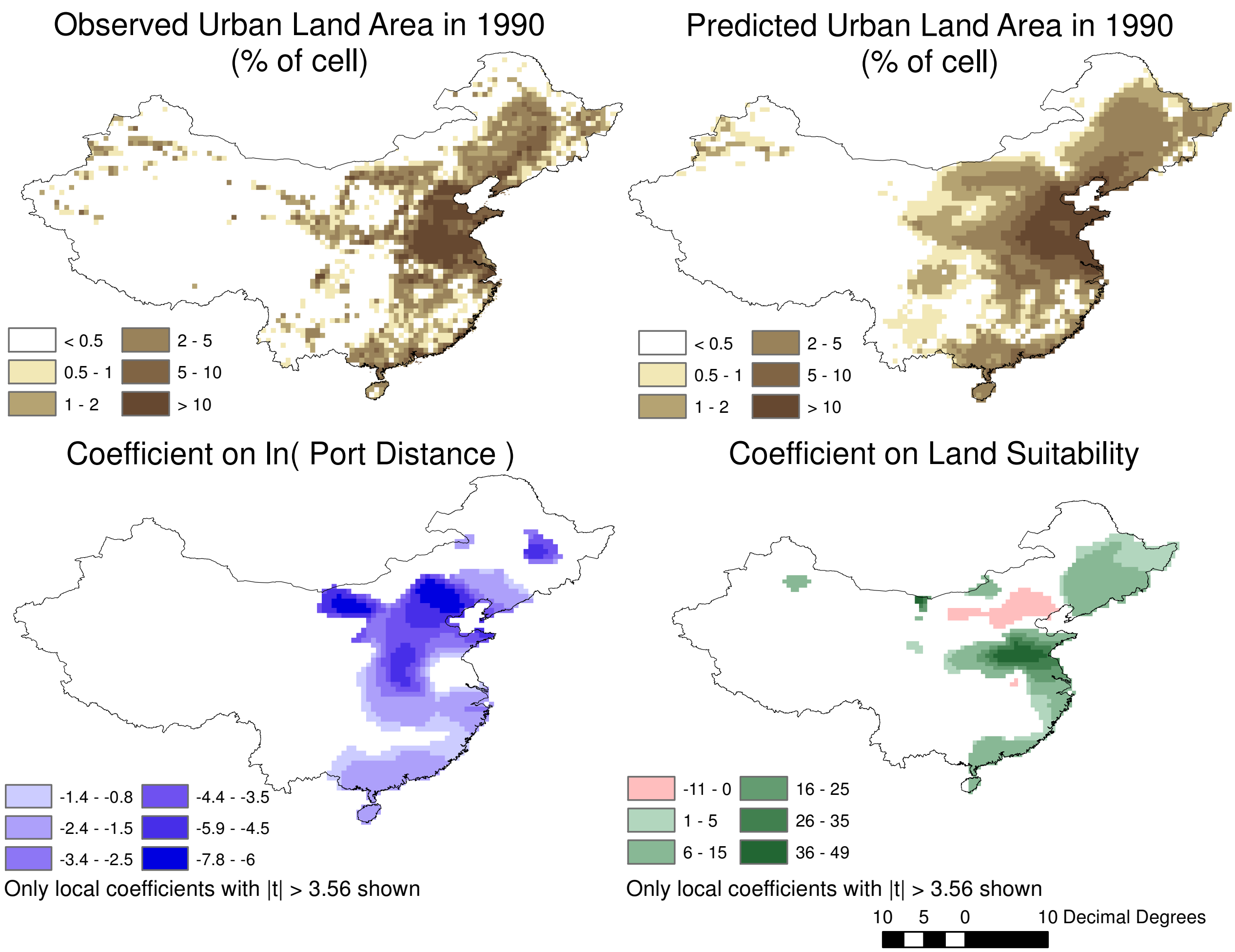


Figure 5: Marginal Effects on 1990-2000 Urban Growth as % of cell area, by province

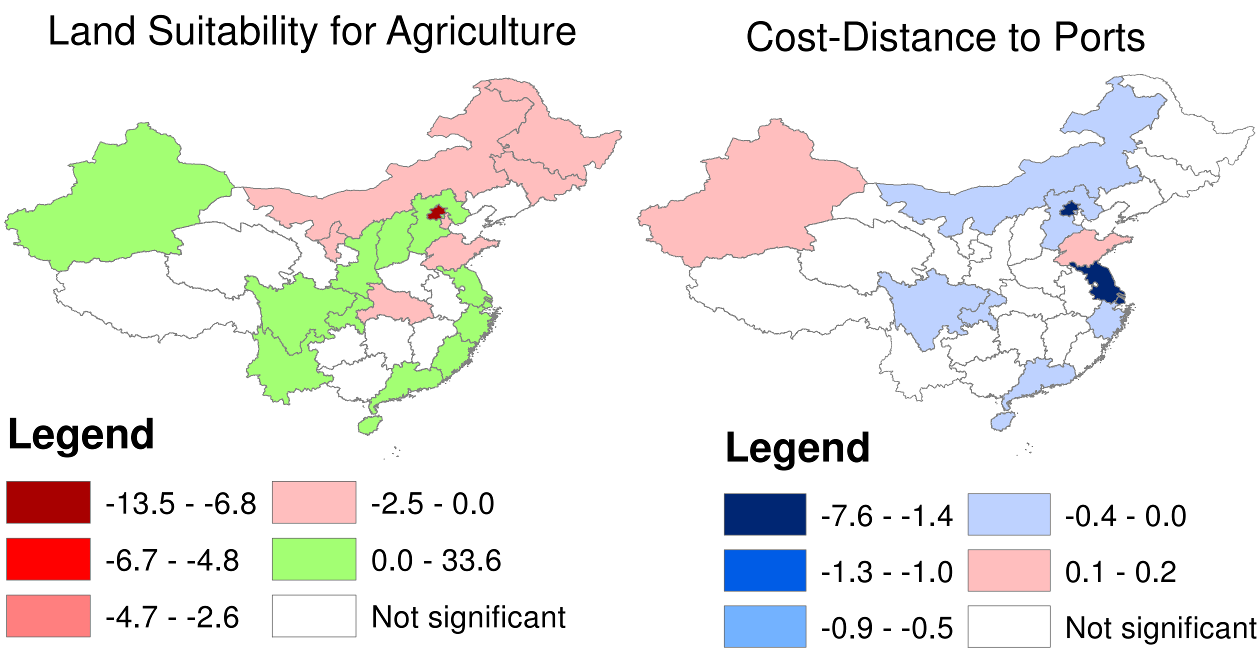


Figure 6: Results from GWR for Urban Growth

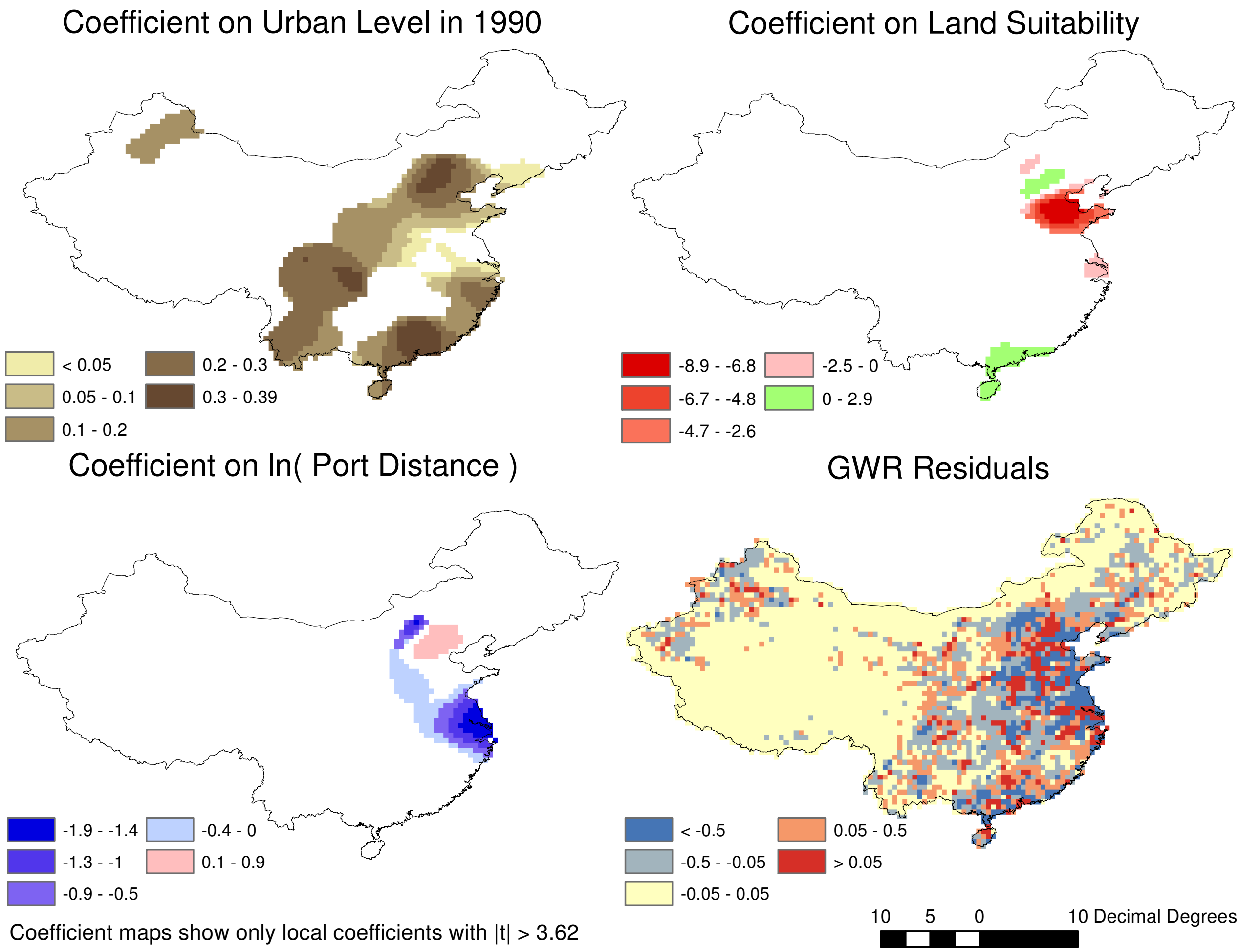


Figure 7: Projections of Future Urban Expansion

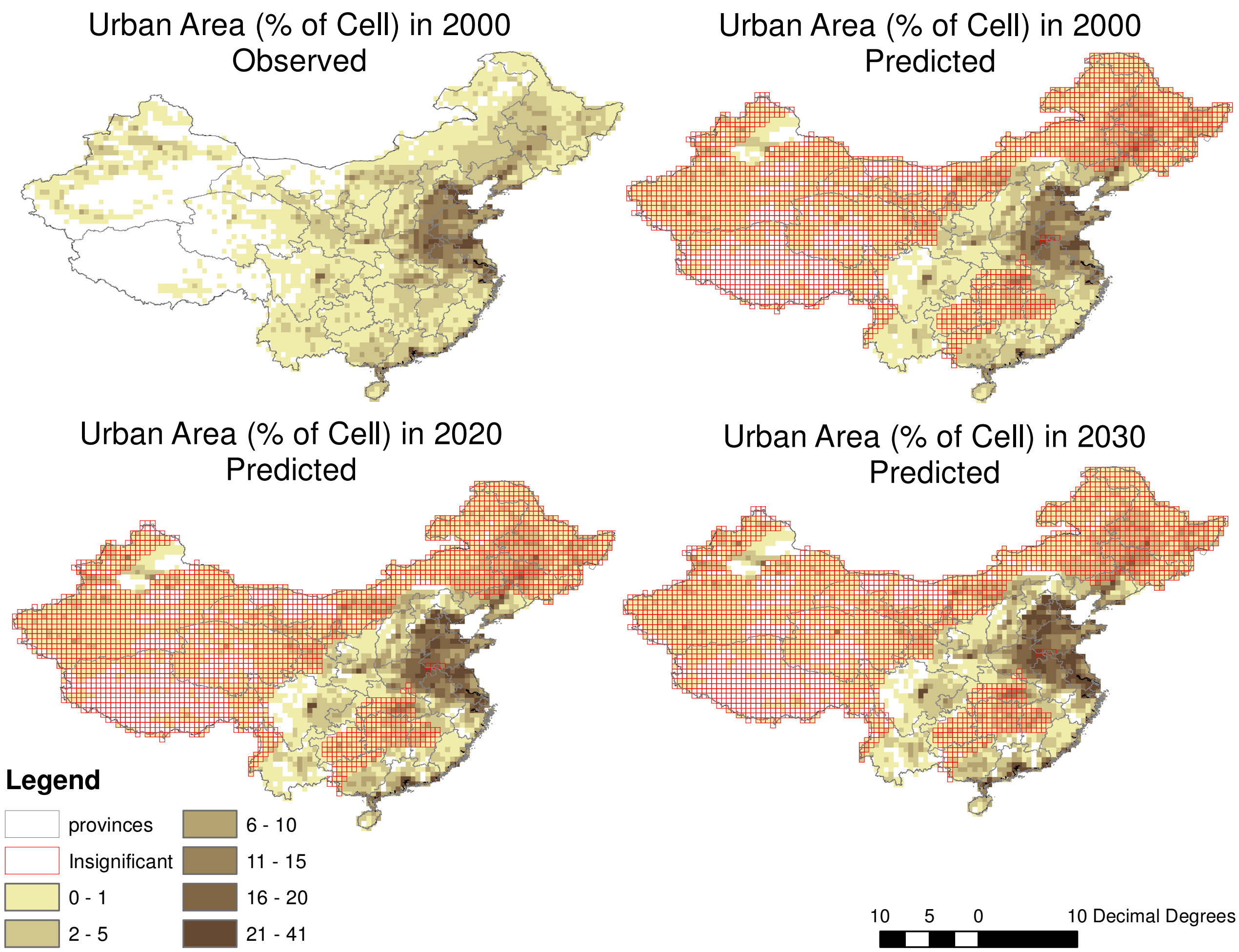


**Table 1: Summary Statistics**

| | Mean | Standard Deviation | Min | Max |
|---|---|---|---|---|
| Area (kms squared) | 2,336.5 | 483.7 | 41.9 | 2,918.0 |
| Urban Land Cover in 1990 (%) | 1.58 | 3.49 | 0 | 35.5 |
| 1990-2000 Urban Land Cover Increase (%) | 0.17 | 0.69 | 0 | 15.6 |
| Land Suitability for Agriculture | 0.35 | 0.36 | 0 | 1 |
| Cost-Adjusted Distance to Ports | 12,152.8 | 10,499.0 | 38.3 | 37,118.8 |

**Table 2: Least Squares & General Spatial Model for Urban Land Cover in 1990**

| Dependent variable: | Urban Land % | | | | |
|---|---|---|---|---|---|
| Independent variables: | (i) | (ii) | (iii) | (iv) | (v) |
| Land Suitability | 2.11*** (0.20), [.65] | 3.97*** (0.24) | 1.49** (0.69), [.38] | 1.00*** (0.22) | 1.04*** (0.24) |
| ln( Cost Distance to Ports ) | -0.99*** (0.05), [.19] | -1.36*** (0.06) | -1.59*** (0.41), [.22] | -1.28*** (0.09) | -1.33*** (0.11) |
| lambda | | | | 0.94*** (0.004) | 0.89*** (0.01) |
| rho | | | | 0.57*** (0.002) | 0.39*** (0.02) |
| Constant | 9.44*** (0.52), [1.87] | 10.90*** (0.62) | 16.1*** (2.69), [2.28] | 11.69*** (0.83) | 12.18*** (1.01) |
| N | 4004 | 4004 | 4004 | 4004 | 4004 |
| $W_1$ Matrix | - | - | - | Contiguous Neighbors | Contiguous Neighbors |
| $W_2$ Matrix | - | - | - | Distance to neighbor <300km | Contiguous Neighbors Quadratic |
| R-squared | 0.46 | 0.14 | 0.62 | 0.79 | 0.79 |
| log-likelihood | | | | -21799.65 | -21814.90 |
| Durbin-Watson | 0.63 | | | | |
| Moran's I | 0.69*** | 0.72*** | | .044*** | .056*** |

Standard errors in parentheses, *** significant to 1%, ** significant to 5%

(i) is least squares with (uncorrected) and [Conley SHAC corrected] standard errors, (ii) is a Tobit specification

(iii) includes province fixed effects & robust standard errors clustered standard errors by province

Models (iv) and (v) are the General Spatial Model using different neighbor matrices for error structure

All models weigh observations by land area

**Table 3: Least Squares & General Spatial Model for Urban Growth from 1990-2000**

| Dependent variable: | Growth in Urban Land Cover (% of Cell) | | | |
|---|---|---|---|---|
| Independent variables: | (i) | (ii) | (iii) | (iv) |
| Urban Cover in 1990 (%) | 0.10*** (0.003), [0.02] | 0.09*** (0.03), [0.02] | 0.11*** (0.004) | 0.12*** (0.004) |
| Land Suitability | -0.18*** (0.04), [0.1*] | -0.03 (0.10), [0.05] | -0.04 (0.06) | -0.12* (-0.06) |
| ln( Cost Distance to Ports ) | -0.08*** (0.01), [0.04**] | -0.19* (0.11), [0.05***] | -0.08*** (0.03) | -0.02 (-0.02) |
| lambda | | | 0.89*** (0.03) | 0.71*** (0.02) |
| rho | | | 0.52*** (0.002) | 0.41*** (0.02) |
| Constant | 0.78*** (0.11), [0.38**] | 1.38*** (0.87) [0.39] | 0.73*** (0.26) | 0.14 (0.19) |
| N | 4004 | 4004 | 4004 | 4004 |
| $W_1$ Matrix | - | - | Contiguous Neighbors | Contiguous Neighbors |
| $W_2$ Matrix | - | - | Distance to neighbor, up to 300km | Contiguous Neighbors Quadratic |
| Durbin-Watson | 1.16 | | | |
| R-squared | 0.33 | 0.20 | 0.58 | 0.57 |
| log-likelihood | - | - | -16729.72 | -16742.6 |
| Moran's I | .47*** | | .08*** | .09*** |

Standard Errors in parentheses, *** significant to 1%, ** significant to 5%, * significant to 10%

Model (i) is least squares with (uncorrected) and [Conley SHAC corrected] standard errors, (ii) includes province fixed effects & robust standard errors clustered by province

Models (iii) and (iv) are the General Spatial Model using different neighbor matrices for error structure

All models weigh observations by land area

## Table 4: Spatial Models for Land Cover in 1990

| Dependent variable: | Urban Land Cover (% of Cell) | | |
|---|---|---|---|
| Independent variables: | (i) | (ii) | (iii) |
| Land Suitability | .98*** | 1.37*** | .39*** |
| | (0.24) | (0.26) | (0.12) |
| Cost Distance to Ports | -1.55*** | -1.98*** | -0.20*** |
| | (0.09) | (0.10) | (0.03) |
| lambda | 0.877*** | 0.99*** | |
| | (0.007) | (0.02) | |
| rho | | | 0.84*** |
| | | | (0.01) |
| Constant | 14.70*** | 17.87*** | 1.86*** |
| | (0.87) | (1.01) | (0.30) |
| N | 4004 | 4004 | 4004 |
| $W_1$ Matrix | - | Contiguous Neighbors | Contiguous Neighbors |
| R-squared | 0.81 | 0.73 | 0.31 |
| log-likelihood | -21818.41 | -22223.18 | -21983.42 |
| Moran's I | .73*** | .61*** | .71*** |

standard errors in parentheses, *** significant to 1%, ** significant to 5%

All models weigh observations by land area

(i) and (ii) are SEM models which incorporate spatial structure in the errors

(iii) is an SAR model which incorporates a spatial dependence

**Table 5: Spatial Models for Urban Growth from 1990-2000**

| Dependent variable: | Growth in Urban Land Cover (% of Cell) | | |
|---|---|---|---|
| Independent variables: | SEM (i) | SEM (ii) | SAR |
| Urban Cover in 1990 (%) | 0.13*** | 0.12*** | 0.06*** |
| | (0.004) | (0.004) | (0.002) |
| Land Suitability | -0.13** | 0.0003 | -0.13*** |
| | (0.07) | (0.06) | (0.03) |
| Cost Distance to Ports | -0.07*** | -0.22*** | -0.01 |
| | (-0.02) | (0.03) | (0.01) |
| lambda | 0.78*** | 0.96*** | |
| | (0.01) | (0.01) | |
| rho | | | 0.64*** |
| | | | (0.002) |
| Constant | 0.62*** | 2.12*** | 0.12 |
| | (0.21) | (0.27) | (0.09) |
| N | 4004 | 4004 | 4004 |
| $W_1$ Matrix | - | Contiguous Neighbors | Contiguous Neighbors |
| R-squared | 0.63 | 0.49 | 0.23 |
| log-likelihood | -16573.4 | -17018.09 | -1416.57 |
| Moran's I | .51*** | .32*** | .50*** |

standard errors in parentheses, *** significant to 1%, ** significant to 5%

All models weigh observations by land area

(i) and (ii) are SEM models which incorporate spatial structure in the errors

(iii) is an SAR model which incorporates a spatial dependence

**Table 6: Projections of Urban Land Area Growth, by province**

| Province | Observed Urbanization in 2000 | | Predicted Urbanization in 2020 | | Predicted Urbanization in 2030 | |
|---|---|---|---|---|---|---|
| | % | km sq. | % | km sq. | % | km sq. |
| Anhui | 8.12% | 11,394 | 9.96% | 13,973 | 10.89% | 15,275 |
| Beijing | 13.37% | 2,194 | 17.04% | 2,796 | 21.51% | 3,528 |
| Chongqing | 0.71% | 584 | 1.48% | 1,216 | 2.01% | 1,660 |
| Fujian | 1.62% | 1,985 | 2.00% | 2,445 | 2.23% | 2,723 |
| Gansu | 0.72% | 2,909 | 0.84% | 3,409 | 0.91% | 3,707 |
| Guangdong | 4.31% | 7,667 | 6.73% | 11,965 | 8.59% | 15,268 |
| Guangxi | 1.74% | 4,120 | 2.14% | 5,072 | 2.39% | 5,643 |
| Guizhou | 0.32% | 559 | 0.47% | 835 | 0.56% | 979 |
| Hainan | 1.87% | 638 | 2.12% | 723 | 2.37% | 809 |
| Hebei | 7.36% | 13,812 | 9.78% | 18,344 | 11.63% | 21,807 |
| Heilongjiang | 1.82% | 8,280 | 1.89% | 8,619 | 1.94% | 8,803 |
| Henan | 10.03% | 16,599 | 11.40% | 18,876 | 12.16% | 20,130 |
| Hubei | 2.61% | 4,856 | 3.19% | 5,928 | 3.47% | 6,450 |
| Hunan | 1.17% | 2,472 | 1.37% | 2,898 | 1.50% | 3,179 |
| Jiangsu | 13.31% | 13,538 | 17.78% | 18,078 | 20.10% | 20,441 |
| Jiangxi | 1.56% | 2,611 | 1.83% | 3,064 | 2.00% | 3,332 |
| Jilin | 3.21% | 6,124 | 3.38% | 6,451 | 3.47% | 6,628 |
| Liaoning | 5.84% | 8,515 | 6.34% | 9,251 | 6.63% | 9,666 |
| Nei Mongol | 0.89% | 10,222 | 0.98% | 11,271 | 1.03% | 11,794 |
| Ningxia Hui | 1.69% | 879 | 1.99% | 1,033 | 2.20% | 1,141 |
| Qinghai | 0.13% | 952 | 0.16% | 1,171 | 0.18% | 1,302 |
| Shaanxi | 1.37% | 2,811 | 1.78% | 3,659 | 2.03% | 4,182 |
| Shandong | 12.15% | 18,900 | 14.66% | 22,803 | 15.99% | 24,860 |
| Shanghai | 17.52% | 1,206 | 24.99% | 1,720 | 29.37% | 2,022 |
| Shanxi | 2.60% | 4,064 | 3.39% | 5,312 | 3.85% | 6,019 |
| Sichuan | 0.57% | 2,735 | 1.01% | 4,906 | 1.32% | 6,408 |
| Tianjin | 15.84% | 1,853 | 24.22% | 2,833 | 29.07% | 3,401 |
| Xinjiang Uygur | 0.33% | 5,373 | 0.40% | 6,596 | 0.47% | 7,607 |
| Xizang | 0.01% | 116 | 0.02% | 218 | 0.03% | 289 |
| Yunnan | 0.52% | 1,999 | 0.74% | 2,851 | 0.89% | 3,395 |
| Zhejiang | 3.10% | 3,189 | 4.71% | 4,841 | 5.74% | 5,900 |