Characterising the Landscape in the Analysis of Urbanisation Factors: Methodology and Illustration for the Urban Area of Angers

Julie Bourbeillon*, Thomas Coisnon, Damien Rousselière** and Julien Salanié*****

Abstract – Urbanisation is usually modelled to account for the trade-off between the rent from an agricultural and urban land-use in a location. In this article, we propose a model that includes a characterisation of the land in respect of not only its economic and physical aspects, but also using variables in relation to landscape perception. To that end, we develop an original two-stage approach consisting of estimating a probability of urbanisation and then taking the uncertainty of urbanisation into account using an internal meta-regression method. The landscape descriptors, constructed based on a textual analysis of the Landscape Atlases, are introduced in this second stage. The application of this method to the urban area of Angers shows the importance of these elements in analysing urbanisation.

JEL Classification: C25, R14
Keywords: landscape atlas, internal meta-regression, perceptions, urbanisation

**Institut Agro, University of Angers, INRAE, IRHS, SFR 4207 QuaSaV (julie.bourbeillon@agrocampus-ouest.fr); **SMART-LERECO, INRAE, Institut Agro (thomas.coisnon@agrocampus-ouest.fr, damien.rousseliere@agrocampus-ouest.fr); ***University of Lyon, UJM Saint-Etienne, GATE UMR 5824 (julien.salanie@univ-st-etienne.fr)*
The authors would like to thank two anonymous reviewers for their constructive comments. This work was funded by the French Ministry of Ecology (at the time Ministère de l'Écologie, du Développement Durable, des Transports et du Logement, MEDDTL) within the framework of the Paysage et Développement Durable programme (agreement n°11-MUTS-PDD-3-CVS-019).

Received in December 2018, accepted in December 2020. Translated from "Caractériser le paysage dans l'analyse des facteurs de l'urbanisation : méthodologie et illustration pour l'aire urbaine d'Angers".

Citation: Bourbeillon, J., Coisnon, T., Rousselière, D. & Salanié, J. (2021). Characterising the Landscape in the Analysis of Urbanisation Factors: Methodology and Illustration for the Urban Area of Angers. *Economie et Statistique / Economics and Statistics*, 528-529, 109–128. doi: 10.24187/ecostat.2021.528d.2062

In France, as in the rest of Europe and in North America, most of the increase in the footprint of urban land use is taking place on land used for agriculture. The scale of the phenomenon is the result of two forces: firstly, urban growth, under the natural effect of population growth (Grekousis & Mountrakis, 2015) and the rural exodus generated by the differences in standards of living between town and country (Polèse & Shearmur, 2005); secondly, urban deconcentration, as illustrated by the United States, where the population of cities living in the suburbs – these mixed areas made accessible in particular by the advent of the private car – rose from 40% to 60% between 1950 and 1990 (Couch *et al.*, 2007). In Europe, while almost 75% of the population lives in cities, built-up land covers just under 5% of the territory but continues to expand its spatial footprint steadily, albeit at a slower pace than in the early 2000s (EEA, 2019). CORINE Land Cover data,[1] which bring together geographical data for 39 European countries, make it possible to assess the extent of urban pressure on agricultural, forest and natural areas: between 2012 and 2018 in this group of countries, urban land use (residential, commercial, etc.) led to the taking of almost 496,000 ha of agricultural land, forests and natural areas (i.e. the surface area of an average French department, such as the Jura or Haute-Loire, for example). At European level, about 42% of this land take occurred on arable land, 27% on grassland, nearly 20% on forests and the rest (about 11%) on various natural areas (moorland, wasteland and wetlands), with, of course, a very high degree of heterogeneity inherent in the characteristics of the various countries. The French situation is fairly close to the European average: over the same period, of the more than 47,000 ha of land taken for development, 50% came from arable land, 31% from grassland and 15% from forest areas.

In this article, we measure land take through the conversion of a plot of land from its original agricultural or forestry use to so-called urban use. The main determining factors are well-known: for a plot of land to be converted, its alternative use must become relatively more attractive than its original use. In the suburbs, the main alternative use is residential. The theoretical model developed in Coisnon *et al.* (2014) shows how the profitability of the two main alternative uses, agricultural and residential, is changing spatially. It also shows that amenities and the living environment can play an important role, in addition to the classic determining factors of incomes associated with these uses.

Therefore, we ask the following research question:[2] how can an empirical model of land use change include elements relating to landscape perception in addition to the usual determining factors? Indeed, although there is a substantial amount of empirical literature on the formation of agricultural land prices, urban land prices and the determining factors of land use change, it does not, to our knowledge, take into account the "cultural" dimension of the living environment, which underpins landscape analysis in cultural geography (Cosgrove, 2003). The purpose of our contribution is both methodological and operational. We propose linking a land use model to descriptors of this dimension that we take from the textual analysis of Landscape Atlases. In order to test its operational nature, we apply this innovative methodology to the case of the urban area of Angers over the period 2000-2010.

Landscape Atlases are created at departmental or regional level, by a generally multidisciplinary team. In 1994, the Directorate of Architecture and Urban Planning proposed a drafting method, which includes an analysis of the sensitive dimension, so as to ensure that these Atlases constitute "a shared state of reference" (Brunet-Vinck, 2004). This methodology suggests three parts in particular: the delimitation of landscape units, thus defining the study level (Roche, 2007), perceptions and changes to landscapes. The aim is to translate the European Convention's definition of landscape: "an area, as perceived by people, whose character is the result of the action and interaction of natural and/or human factors" (Council of Europe, 2000). Landscape Atlases can therefore be considered to be "a tool for identifying and classifying landscapes [...]" (Ambroise, 2010). They thus represent a body of knowledge on landscapes and, more specifically, the way in which they are perceived, which fits into the framework of our study.

The rest of this article is organised in the following manner. First, we present our two-stage econometric strategy. The latter is an original contribution to the literature, aiming to quantify the importance of the landscape variables introduced in the model while taking into account the uncertainty linked to the selection of models within a reasonable estimation time. We

1. Provided by the European Environment Agency (EEA). https://www.eea.europa.eu/data-and-maps/dashboards/land-take-statistics, last accessed on 05/02/2020.

2. This article follows on from research on the links between landscape and urban sprawl (PAYTAL, 2014).

also specify the method used to extract landscape data from a textual analysis of the Landscape Atlases. We then move on to the digital application of our methodological proposal for the urban area of Angers. We conclude by discussing the limitations of our approach and the opportunities for future research that arise from this work.

1. Empirical Strategy: A Two-Stage Estimation

In order to assess the role of cultural and perceptual elements of landscapes in urbanisation, we carry out a two-stage econometric approach, inspired by various recent works (Bryan & Jenkins, 2016; Coisnon *et al.*, 2019).[3] We then provide details on the construction of the landscape variables.

1.1. A Model for Estimating Land Use Change

The model proposed by Polyakov & Zhang (2008) and taken up in Nery *et al.* (2019), on land use change taking into account the initial situation, with the latter being seen as a proxy for conversion costs. This model, which is also referred to as a short-run model in the literature (Ay *et al.*, 2017), is estimated using a multinomial logit model (see Online Appendix C1 for a presentation; link at the end of the article). As the assumptions of the multinomial logit model may be restrictive with respect to the data (in particular the assumption of independence of irrelevant alternatives), we also estimate multinomial probit models and binomial models (logit and probit). Within the framework of a Bayesian model selection procedure, with the function of the first stage being the prediction of a marginal effect, we rely on the Akaike Information Criterion (AIC), which is particularly well adapted to this predictive objective (Gelman *et al.*, 2014).

The AIC then makes it possible to calculate the probability of each model approximating the true data generation process and which is therefore considered to be the best competing model among all the estimated models (Burnham & Anderson, 2004). It also makes it possible to calculate the Ockham window composed of the set of models with a probability that is reasonably different from zero (Tsai & Li, 2008).

In addition to numerous robustness checks, we introduce indicators relating to belonging to a geographical area that is homogeneous from a landscape point of view. For this purpose, we use the landscape units (LU) as defined in the Landscape Atlases.

In the robustness checks, we introduce variables capable of defining the physical dimensions of landscapes, such as landscape metrics or indicators of agricultural and topographical specificities at LU level. We describe these in subsection 2.3. In this way, we can extract the effect of belonging to each LU of a Landscape Atlas, independently of the physical characteristics of the landscapes of these landscape units, which have been extracted separately through the estimation of the econometric model.

We can thus estimate the marginal effect of belonging to a given landscape unit LU^m, where $m = \{1, ..., M\}$, on the probability for a plot i to be allocated to a land-use k urban (k=u) at time t knowing that it was allocated to a land-use j non-urban (j≠u) in the previous period. This marginal effect, noted $\widehat{P_{kui}^m}$ is given by:

$$\widehat{P_{kui}^m} = \frac{\partial Prob_i\left(k = u \middle| j \neq u, t\right)}{\partial LU^m} = \\ Prob_i\left(k = u \middle| j \neq u, t, LU^m = 1\right) \\ -Prob_i\left(k = u \middle| j \neq u, t, LU^m = 0\right) \quad (1)$$

We can therefore describe these estimated marginal conditional probabilities through their first two empirical moments:

$$\begin{cases} \text{the sample mean for } \widehat{P_{kui}^m} \text{ is} \\ \mu_P^m = \frac{1}{N^m} \sum_{i=1}^{N^m} \widehat{P_{kui}^m} \\ \text{the sample covariance of } \widehat{P_{kui}^m} \text{ is} \\ (\sigma_P^m)^2 = \frac{1}{N^m - 1} \sum_{i=1}^{N^m} \left(\widehat{P_{kui}^m} - \mu_P^m\right)^2 \end{cases} \quad (2)$$

where N^m is the number of non-urban plots located in LU m at the beginning of the period.

These elements allow us to assess the differences between LUs. In particular, they enable us to set up a second stage in which we explain the observed differences in the average μ_P^m estimated marginal conditional probabilities for each LU. We relate these average marginal effects to measurements derived from a textual analysis of the Landscape Atlases and we regress μ_P^m on indicators of lexical richness or results of automatic language processing that describe the main semantic fields appearing in the LU descriptions.

In order to study the role of perception variables in a second stage, we introduce objective descriptors of landscapes that can be correlated with their sensitive aspects (see Uuemaa *et al.*, 2009,

3. The underlying theoretical model is detailed in the PAYTAL report (2014).

whose literature review suggests a link between objective landscape descriptors and their subjective counterparts) as control variables in the first stage. We will use landscape metrics borrowed from ecology, agricultural zoning (in small agricultural regions, *Petites Régions Agricoles*, PRA), the technical-economic orientations of farms (*orientations technico-économiques des exploitations agricoles*, OTEX) and administrative divisions (cantons).

1.2. A Meta-Model for Analysing the Role of Perceptions in Controlling Model Selection Uncertainty

In order to assess the impact of the modelling options in the first stage on the measurement of the parameters of interest and the results of the second stage, we carry out an internal meta-analysis, following the method suggested by Banzhaf & Smith (2007). In practice, as Banzhaf & Smith (2007), Kuminoff *et al.* (2010) or Klemick *et al.* (2018), have done, a set of models corresponding to the inclusion/exclusion of different variables can be estimated. In this way, a meta-regression is established which explains the effect obtained in the first stage as a function of the different modelling options chosen (i.e. inclusion/exclusion of a particular variable) and as a function of the quality of the model in question (Sutton & Higgins, 2008). We also introduce the AIC as a simple additional variable in the regression of this second stage. If the original model contains a set of K explanatory variables, there are then 2^K potential models to estimate. The result, derived from the calculation of marginal effects on a large number of observations, leads to the type of problem described as intractable in econometrics; extremely costly in terms of calculation time, it requires adapted algorithms in order to be performed within a reasonable time (Moral-Benito, 2015). We therefore restrict the candidate models to those that contain variables likely to represent other landscape aspects than those approximated by the Landscape Atlases. If we use four variables (OTEX, PRA, landscape metrics and cantons), for example, taking into account the possibility of an estimation via a probit or logit link and a dichotomous or categorical response, this leads to estimating 2^P=64 models[4] representing all possible inclusion/exclusion combinations for these six variants of the model.[5]

We calculate the moments μ_P^m and $(\sigma_P^m)^2$ of the M marginal effects calculated for each model, i.e. M×64 measurements of marginal effects. It is on these measurements that we perform a meta-regression to explain the significance of the measured effect according to the LU descriptors and the modelling options used in the first stage.

In this second stage, we use the following random effects model:

$$\mu_{Pr}^m = \theta_R R_r + \theta_D D_r + u_r + \epsilon_r \quad \text{with} \quad u_r \sim N(0, \tau^2)$$
$$\text{and } \epsilon_r \sim N(0, (\sigma_{Pr}^m)^2) \tag{3}$$

where the index r denotes the r-th of the 64 models estimated in the first stage, R_r is a vector of variables describing the LUs and D_r is a vector of variables comprising the descriptors of the model, i.e. the presence or absence of a variable in the first stage and θ_R and θ_D are two corresponding vectors of parameters to be estimated. u_r is a random term specific to each regression of the first stage and τ^2 therefore represents the inter-regression variance to be estimated. ϵ_r is a traditional random term representing the variance of the result of the first stage. We therefore have $\mu_{Pr}^m \sim N(\theta_R R_r + \theta_D D_r, \tau^2 + (\sigma_{Pr}^m)^2)$, which makes it possible to show that the variability of the results is linked to the specific characteristics of the LUs (vector R_r) and to the modelling options of the first stage (D_r). The variability of the results can also be explained by two components, the variability specific to each regression of the first stage ($(\sigma_{Pr}^m)^2$) and a residual inter-regression variability (τ^2).

The parameters θ_R, θ_D, and τ^2 of (3) are estimated by restricted maximum likelihood (REML), with standard errors corrected according to the method of Knapp & Hartung (2003). The combination of these two methods has been shown to be particularly effective.[6]

As shown by Bryan & Jenkins (2016), based on the original idea in Saxonhouse (1976), this two-stage method is conceptually equivalent to the sequential estimation of a random-effects model (i.e. a multinomial multilevel model). In addition to its econometric efficiency (with a reasonable estimation time), this approach has two further advantages. The first is, as already mentioned, that it makes it possible to control all the uncertainties related to the selection of the models in a simple way, which would be totally unrealistic if we had to do it in the framework of a one-stage model approach. The second is that it makes it possible to quantify the

4. Here, P=6 variants of the model, so 64 models to be estimated.
5. For the sake of robustness, we also estimated 64 other models without including the initial states. These models, which do not take into account these conversion costs, are largely outside the Ockham window and are therefore not taken into account in the meta-regression in the second stage.
6. For this purpose we used the metareg procedure developed in Stata.

importance of the variables used in the second stage (here, the landscape perception variables) through two traditional meta-regression indicators: the share of inter-estimation variance (measured by the adjusted R^2 coefficient) and the total variance attributable to the differences between the studies (measured by the I^2 coefficient). We return to Coisnon *et al.* (2019) for a recent example of the implementation of this method.

1.3. Characterisation of the Landscape Perception Variables

To construct the landscape data, we relied on a textual analysis of the Landscape Atlases.[7] Theoretically, there are two possible approaches: a lexicographical approach, which relies on *ad hoc* dictionaries and/or reducing *a priori* the meaning of a text to the sum of the words that compose it, and a semantic approach, corresponding to a more global approach that aims to preserve the meaning of the text (see Lebart, 1994). These approaches largely correspond to different textual analysis tools that have been greatly refined in recent years: massive text mining, neural networks, sentiment analysis through word embedding, etc. (Loughran & McDonald, 2016; Nowak & Smith, 2017; Kozlowski *et al.*, 2019).

We have chosen to test these two input methods. For lexicographical input (Nowak & Smith, 2017; Blanc *et al.*, 2019), indicators for richness of vocabulary have been estimated for all the territories covered by the Landscape Atlases at our disposal. We have used various large *ad hoc* dictionaries using thesauri such as Eurovoc or Gemet in relation to certain themes present in the texts.[8] The indicator used is the frequency of terms related to these different dictionaries (architecture, botany, economics, animal husbandry, mineralogy, urban planning, forestry, geology, countryside, viticulture, religion and water).[9] It was then standardised across all of the digitised LUs. In this way, we have an indicator that allows us to compare the relative richness for the same dictionary across different LUs.

For semantic input (Loughran & McDonald, 2016; Maire & Liarte, 2019), we have used the Tropes software developed by Molette (2009), which is part of the field of Natural Language Processing (NLP), a discipline that brings together linguistics, IT and artificial intelligence. Each text (article, speech, publication, etc.) is analysed in order to reveal the skeleton of the text, its meaning. To do this, Tropes relies on a set of theoretical models, which aim to remove the subjectivity of the user from the analysis. The study of texts is based on a morphosyntactic analysis, a lexicon and a semantic network. It makes it possible to evaluate, among other things: the styles and settings of the text, the remarkable propositions, the global context (“the reference universes”), the references used, the relations between elements, the lists of verbs and adjectives used (and their frequencies), etc. The Tropes terminology extraction method is based on taxonomies called scenarios. These scenarios are designed to enrich and filter the classes of equivalents (the associated concepts and terms) in accordance with an analysis strategy. Once the analysis has been completed, it is possible to generate a full report of the text studied. The reference dictionary, called the “concept scenario”, contains a very broad lexicon of 28 basic categories.[10] The software thus allows the analysis of any type of discourse through more than 60,000 terms of basic French vocabulary, organised hierarchically according to these categories. The text made up of descriptions of all the LUs was classified according to these concepts defined in the “concept scenario”. The variables created are, again, relative vocabulary richness variables for each of the basic categories and have been standardised.

2. Application to the Urban Area of Angers for the Period 2000-2010

2.1. Presentation of the Territory Studied

The urban area of Angers corresponds to that used by INSEE,[11] updated in 2011 based on data from 2010. Our study area contains 133 municipalities and offers a certain landscape diversity, due in particular to a specific and highly spatialised agricultural dynamic. It includes, for example, wine-growing landscapes in the Layon and Aubance valleys to the south of the urban area, a more concentrated area of horticulture and market gardening within the horticultural triangle bordered by the Loire and Maine rivers, a wooded plateau in the Haut-Anjou to the north and a denser, hilly wooded area to the west. The east of the urban

7. Pre-processing operations are detailed in the PAYTAL report (2014).
8. The ad hoc *nature of these dictionaries, although expertly chosen, partly justifies the fact that we do not focus the econometric analysis of urbanisation in what follows on this type of indicator.*
9. See, for example, https://www.eionet.europa.eu/gemet/en/theme/40/concepts/ for the dictionary relating to water.
10. The exact list of categories or themes in this scenario is available from the Tropes software installation page (https://www.tropes.fr/)
11. The definition of urban areas used by INSEE is essentially based on commuters, i.e. individuals who do not work in the municipality in which they live. In our study area, for each municipality, at least 40% of the working age population works in the municipality of Angers.

area is characterised by fruit-growing landscapes and a dynamic of opening up the landscape along the Authion valley, characterised by the development of large-scale farming. The diversity of landscapes within the urban area of Angers is thus relevant for the empirical application of our methodology.

Land use data were obtained through remote detection;[12] they describe three alternative land uses (forest, agricultural, urban) for pixels with sides measuring 100 m. In the estimation of model (1), the "plots" thus correspond to these square pixels. We have more than 200,000 observations for any given date.

The land-use transition matrices show that urbanisation is taking place mainly on agricultural land: the share of agricultural land fell from 84% in 1990 to 82% in 2000 and 78% in 2010 (Table 1). For these same years, urbanisation increased from 6.9% to 8.7% and 10.6%; this urbanisation concerns the whole of the territory studied, with greater conversion pressure on the outskirts of Angers (see Figure). Urbanisation is virtually irreversible, since only 0.03% of the areas urbanised in 1990 have been returned to agricultural or forestry use (Chakir & Parent, 2009, make a similar observation for the Rhône department).

The shapes of urban areas, as measured by landscape metrics, have also changed between 1990, 2000 and 2010. The first two metrics (number of patches and perimeter) are measurements of the fragmentation of land use classes. We see that there is a general trend for agriculture to be less fragmented: the number of patches decreases and their perimeter increases (Table 2). The perimeter/area ratio increases for agriculture and forests but decreases for urban areas. This indicates that, overall, agricultural and forest areas tend to be less compact (less round shapes) and the urban patches tend to agglomerate, because the urbanisation takes place next to existing urbanised patches. The shape index corrects the assessments that can be made by the perimeter/area ratio by taking into account the fact that the pixels are square. We then see that the shape of the urban patches has also become more complex, but less so than with the other two classes.

To summarise, we can say that at the level of the Angers urban area, urban sprawl has essentially taken place on agricultural land, mainly by filling in the gaps between existing urbanised areas, and that this urbanisation has been accompanied by larger and more complex patches of forest and agricultural areas.

2.2. Socio-Economic Data

The conversion of land is determined by the incomes from its alternative uses and by the costs of conversion. In the absence of precise agronomic and pedological data, we assess the income from agricultural use through the slope

12. *The original raster provides land use data for pixels with sides measuring 20 m. Since the urban area is about 60 km high and 60 km wide, this leads to the observation of more than 3 million pixels. To facilitate data processing, we have aggregated these data on a raster with a base pixel size of 100 m by 100 m, i.e. 25 original pixels. When aggregating, we gave priority to urban pixels (a pixel is classed as urban if at least 1/5e of it is urbanised, otherwise it is classed as being used for the most prevalent purpose); this reduces the dataset to just over 220,000 observations, with minimal loss of information.*

Table 1 – **Land use transition matrices for the urban area of Angers (number of pixels and % contribution of previous uses)**

		Land use in 1990			
		Urban	Agriculture	Forest	Total
Land use in 2000	Urban	15,292	3,916	148	19,356
		(79.0 %)	(20.2 %)	(0.8 %)	(100.0 %)
	Agriculture	11	177,876	3,511	181,398
		(0.0 %)	(98.0 %)	(1.9 %)	(100.0 %)
	Forest	8	3,994	16,331	20,333
		(0.0 %)	(19.6 %)	(80.3 %)	(100.0 %)
		Land use in 2000			
		Urban	Agriculture	Forest	Total
Land use in 2010	Urban	19,221	4,054	169	23,444
		(82.0 %)	(17.3 %)	(0.7 %)	(100.0 %)
	Agriculture	125	169,915	2,634	172,674
		(0.0 %)	(98.4 %)	(1.5 %)	(100.0 %)
	Forest	10	7,429	17,530	24,969
		(0.0 %)	(29.7 %)	(70.2 %)	(100.0 %)

Reading note (first line): In 2000, of the 19,356 pixels detected as urban, 15,292 were already urban pixels in 1990 (i.e. 79.0%), 3,916 were agricultural pixels (i.e. 20.2%) and 148 were forest pixels (i.e. 0.8%).

Figure – **Changes in the urbanisation of the urban area of Angers**

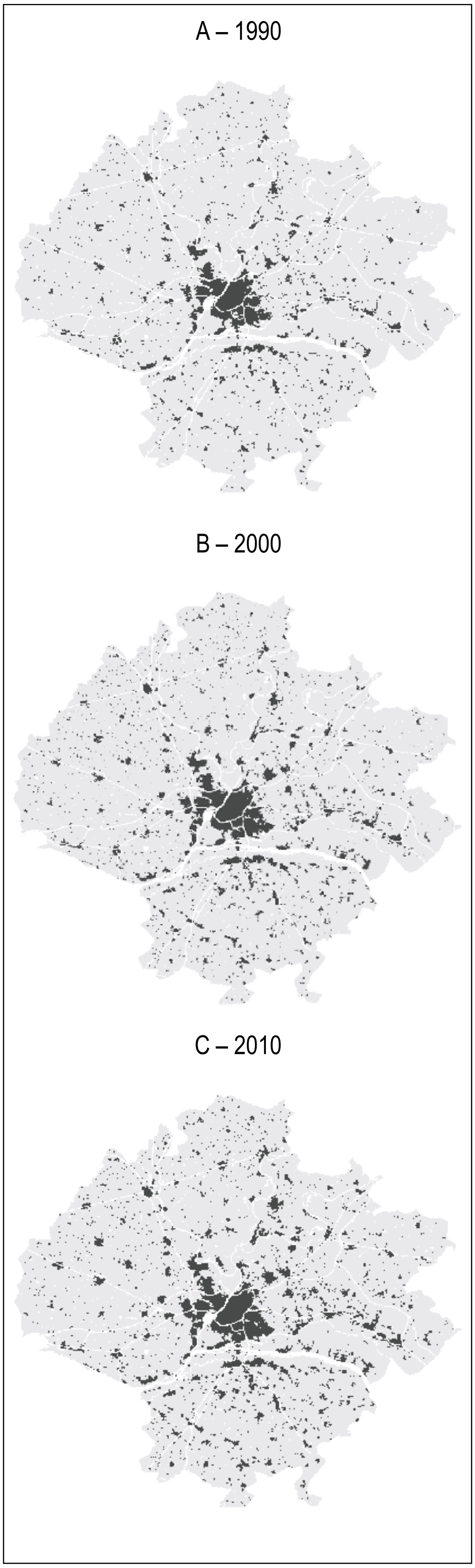

Table 2 – **Evolution of some landscape metrics for the urban area of Angers**

Land use	1990	2000	2010
	Number of patches		
Agriculture	128,838	122,368	107,300
Forest	5,704	6,223	6,763
Urban	4,839	6,220	8,027
	Perimeter		
Agriculture	55,076	56,872	64,284
Forest	24,122	22,914	31,240
Urban	17,672	22,112	25,250
	Perimeter/area ratio		
Agriculture	0.296	0.314	0.372
Forest	1.207	1.127	1.251
Urban	1.154	1.142	1.077
	Shape index		
Agriculture	31.910	33.376	38.632
Forest	42.618	40.059	49.274
Urban	35.629	39.627	41.124

measured on our pixels with sides measuring 100 m, based on the IGN's BDALTI digital terrain model (DTM) with metric precision for the Maine-et-Loire department. We define the slope as the difference between the highest and lowest points of each of our pixels. We also assess agricultural income through the differences in the technical-economic orientation (OTEX) of the municipalities in the 2000 Agricultural Census, which includes a set of considerations such as soil quality, market opportunities, the price of agricultural inputs and products and the agglomeration economies that influence farmers' choices. This can cause endogeneity problems if the same variables influence these choices and the conversion.[13] However, we believe that this risk is minor here because the OTEX is determined at the aggregated level (the municipality), independently of individual decisions. For robustness, different estimates are carried out with and without this variable. We also use the municipal share of arable land in the utilised agricultural area in 1988. Finally, we use the zoning of areas as Small Agricultural Regions (PRA) as an indicator of agricultural potential.

Urban income is strongly dependent on accessibility to jobs and services. We use three accessibility indicators: distances to Angers city centre and to the nearest main town of a municipality[14] and distance to the main inter-city road network.[15] The data available from the IGN for

13. Impermanence syndrome is a known manifestation of this problem (Lopez et al.*, 1988). It is seen in areas subject to heavy urbanisation when farmers under-invest in the expectation of an increase in land value when a plot is converted.*

calculating distances by road date from 2010, i.e. once the urbanisation decisions have been taken; for distances to Angers (CBD) and to the main towns of the municipalities (SBD), we therefore use the traditional option of calculating distances as the crow flies, as suggested by Chomitz & Gray (1996).

In order to take into account the neighbourhood externalities generated by urban development, we choose the percentage of urbanised pixels within a radius of 250 m[16] as done in some of the literature (e.g. Irwin & Bockstael, 2002, or Newburn *et al.*, 2006).

We also take into account the major confluences and partial embankment of the Loire river on one bank only, which create large areas subject to flooding, by introducing the zoning of areas as recognised flood zones (*zones d'inondation constatées* – ZIC) used by the Pays de la Loire Regional Directorate for the Environment, Development and Housing (*Direction régionale de l'Environnement, de l'Aménagement et du Logement* – DREAL).[17]

Finally, local public decision-makers, in particular mayors, can implement restrictive urban planning policies, specifically through land use plans (*plans d'occupation des sols* – POS), which were replaced in 2000[18] by local urban planning plans (*plans locaux d'urbanisme* – PLU). Some municipalities decide upon a slightly less restrictive municipal charter, and municipalities with little land pressure decide not to implement zoning and instead submit to the national urban planning regulations (*règlement national d'urbanisme* – RNU), which stipulate that new buildings must favour the coherence of the built environment. All of these provisions aim to combat urban sprawl.

We do not have information on the different zoning of the 133 municipalities studied, but we consider this a minor limitation. First, we analyse conversions at intervals of 10 years. Over these periods, urban planning documents are largely amended or revised to adjust to the development needs of the municipalities. Next, the delays in the implementation of the SRU law since 2000, its anticipation during the previous period and the negotiation of the Territorial Coherence Scheme (*Schéma de Cohérence de Territoriale* – SCoT) of the Loire Angers Metropolitan Hub, which covers 66 of our 133 municipalities, have certainly led to a period of instability in zoning, which has been revised in accordance with the progress made by the municipalities and communities of municipalities in defining their development strategy. For these reasons, we believe that zoning has not played a major role.[19] In contrast, we believe that it can affect the level of compactness of urban development. In particular, contiguous development and a positive effect of development on neighbouring plots are expected, due to zoning constraints, at least for low development densities. Zoning has the opposite effect to neighbourhood externalities. In addition to zoning, local taxation on land (housing tax, tax on undeveloped land and tax on developed land) can have a significant effect on urbanisation.

In order to take this into account, one option would be to introduce municipal dummy variables to identify the effects of zoning policy and municipal tax policy. However, this amounts to entering 133×2=266 variables into the model and leads to a significant increase in the time needed for estimation. This is why we have chosen a reasonable compromise consisting in entering municipal dummies:[20] without covering all the specific features of the municipalities, they identify a good part of them, notably because of the existence of strategic tax mimicry behaviours (see, for example, Cassette & Paty, 2006).

2.3. Physical and Landscape Perception Data

The landscape data are constructed for the landscape units of the Landscape Atlas of the Maine-et-Loire department. The GIS layer for this division is the one created by Groult & Roche (2009),[21] available on the CARMEN website. Each pixel of the urban area is coded in dummy form as belonging to an LU. It is the parameters estimated based on these dummies that will allow us to calculate the probability of urbanisation for each pixel, conditional on non-urbanisation in the previous period and on belonging to an LU.

We describe the physical dimensions of the landscapes using three landscape metrics

14. For the coordinates of the main town of a municipality, we use GEOFLA data from the IGN. For a point in space, depending on the spatial configuration, the nearest main town of a municipality is not necessarily that of the municipality in which the point is located.
15. I.e. the main IGN BDTOPO road network.
16. In our data, 79.9% of the pixels have 20 neighbours, 13.6% have between 15 and 19 neighbours, 6.1% have between 10 and 14 neighbours and only 0.4% of the pixels have fewer than 9 neighbours.
17. http://www.sigloire.fr/ last accessed on 2 June 2020.
18. Law on urban solidarity and renewal of 13 December 2000, known as the "SRU law".
19. In addition, the study by Kline et al. *(2001) on Oregon, a pioneering US state in terms of urban planning, finds no significant effect of zoning on the probability of parcel development.*
20. The urban area of Angers is spread over 21 cantons.
21. We are grateful to Richard Raymond for his valuable assistance in obtaining this data.

calculated at the beginning of the period: the perimeter/area ratio, the shape index and the fractal dimension index, calculated on a square grid with sides measuring 3 km. Each pixel is allocated the index values of the square within this grid in which it is located. These indices make it possible to describe the landscape shapes in the close environment of each pixel, even if not directly adjacent.

Some of the proxy variables we use reflect a landscape aspect: for example, the Small Agricultural Regions (PRA) have been established on the basis of the agronomic homogeneity of the territories - a division which certainly has a strong link with the physical aspects of the landscapes. The same applies in respect of the technical-economic orientations of the farms at municipal level (OTEX) or of the municipal division. This is why we will analyse the sensitivity of our results to the inclusion/exclusion of these variables in the economic model of the first stage. The descriptive statistics of the variables included in the model (1) are presented in the Appendix.

Table 3, on landscape perception data, shows the distribution of words for each first-level category representing at least 2% of the words. It shows the predominance of themes that are at the heart of the Landscape Atlases, such as agriculture and the environment, land use and characterisation. In contrast, perception-related elements are rare: they represent a fraction of the themes “Characteristics”, “Behaviour and feelings” or “Strengths and quantities”.

3. Results of the estimations

3.1. Stage one: estimation of the land allocation model

As the Landscape Atlas for the Maine-et-Loire department was created in 2000–2001, we decided to present only the estimates for the 2000–2010 period here. The estimation of the 128 models (without/with the initial situations) for this period is done by maximum likelihood.[22] The 64 models that do not include the initial situations, and thus the conversion costs, have a very high AIC and an almost-zero probability of reflecting the data generation process. The results are presented in Table 4 for the models with the lowest AIC (models (105) and (108) for the dichotomous models and (9) and (41) for

22. *The detailed tables of these estimates are available from the authors.*

Table 3 – **Descriptive statistics for perception variables**

Theme	Number of words	Percentage of words
Geography, countries and territories	7,625	16
Characteristics	7,173	15
Agriculture and the environment	5,077	10
Politics and society	3,628	8
Strengths and quantities	2,155	4
Construction, property and housing	2,078	4
Communications and media	1,487	3
Animals and plants	1,469	3
Behaviours and feelings	1,419	3
Other concepts	6,907	14
Other themes (<2%)	9,336	19
Total	48,354	100

Notes: ‘Other concepts’ corresponds to a residual category in Tropes grouping together “tool” concepts that do not belong to any other category; the line ‘Other themes’ groups together 18 other least frequent themes.

Table 4 – **Results of the estimation of the land allocation models**

	Model (9)		Model (41)		Model (105)	Model (108)
	Forest	Urban	Forest	Urban	Urban	Urban
Constant	-16.256***	5.482***	-9.843***	2.511**	2.375***	1.999***
	(1.083)	(1.616)	(0.687)	(0.982)	(0.715)	(0.308)
CBD distance	-0.103***	0.162***	-0.071***	0.089***	0.077***	0.068***
	(0.030)	(0.048)	(0.020)	(0.030)	(0.022)	(0.021)
(CBD distance)^2	0.000	-0.003***	0.000	-0.002***	-0.001***	-0.001***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
SBD distance	0.503***	-1.626***	0.264***	-0.966***	-0.746***	-0.748***
	(0.052)	(0.071)	(0.033)	(0.043)	(0.031)	(0.031)

➔

Table 4 – **Results of the estimation of the land allocation models** *(contd.)*

	Model (9)		Model (41)		Model (105)	Model (108)
	Forest	Urban	Forest	Urban	Urban	Urban
(SBD distance)^2	-0.081***	0.262***	-0.041***	0.156***	0.120***	0.120***
	(0.012)	(0.018)	(0.008)	(0.011)	(0.008)	(0.008)
Municipality income	-0.000**	0.000***	-0.000**	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Municipality income x CBD distance	0.000***	-0.000***	0.000***	-0.000**	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance to Road	0.270***	-0.103***	0.182***	-0.045**	-0.053***	-0.066***
	(0.018)	(0.030)	(0.012)	(0.018)	(0.013)	(0.013)
(Distance to Road)^2	-0.030***	0.011***	-0.021***	0.006***	0.007***	0.008***
	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
Slope	0.059***	0.057***	0.041***	0.039***	0.026***	0.027***
	(0.007)	(0.014)	(0.004)	(0.008)	(0.006)	(0.006)
(Slope)^2	0.000	-0.004***	0.000	-0.002***	-0.002***	-0.002***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
PNR	0.187	0.339**	0.149*	0.198**	0.125*	0.154**
	(0.118)	(0.146)	(0.076)	(0.093)	(0.067)	(0.067)
Municipality equipment	0.002***	0.003***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Munic. Equip. × SBD Distance	-0.000***	0.000***	-0.000**	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Vicinity urbanised in 2000	-2.298***	10.281***	-1.075***	6.274***	4.812***	4.825***
	(0.262)	(0.233)	(0.166)	(0.155)	(0.110)	(0.110)
(Vicinity urbanised in 2000)^2	3.499***	-6.804***	2.397***	-3.395***	-2.933***	-2.918***
	(0.504)	(0.350)	(0.325)	(0.261)	(0.180)	(0.179)
Agricultural 2000	-0.479	-7.766***	-1.116***	-5.441***	-3.879***	-3.878***
	(0.347)	(0.098)	(0.147)	(0.055)	(0.038)	(0.038)
Forest 2000	4.231***	-6.787***	2.630***	-4.275***	-4.240***	-4.222***
	(0.347)	(0.132)	(0.147)	(0.074)	(0.052)	(0.052)
Susceptible to flooding	0.539***	-1.700***	0.345***	-0.980***	-0.754***	-0.753***
	(0.077)	(0.171)	(0.048)	(0.095)	(0.071)	(0.071)
Loire des Promontoires	-0.176	0.051	0.025	-0.035	-0.036	-0.063
	(0.188)	(0.165)	(0.122)	(0.115)	(0.082)	(0.081)
Beaugeois	-0.258	-1.190***	-0.089	-0.751***	-0.553***	-0.586***
	(0.175)	(0.187)	(0.116)	(0.124)	(0.090)	(0.089)
Couloir du Layon	0.440***	-0.652***	0.353***	-0.364***	-0.322***	-0.349***
	(0.163)	(0.167)	(0.107)	(0.111)	(0.080)	(0.079)
Haut Anjou	-0.390*	-1.396***	-0.143	-0.933***	-0.680***	-0.700***
	(0.207)	(0.232)	(0.135)	(0.152)	(0.110)	(0.109)
Saumurois	0.688***	-0.320**	0.517***	-0.149	-0.184**	-0.214***
	(0.161)	(0.157)	(0.106)	(0.106)	(0.076)	(0.075)
Segréen	-0.759***	-1.347***	-0.302**	-0.886***	-0.629***	-0.681***
	(0.188)	(0.184)	(0.122)	(0.125)	(0.090)	(0.088)
Val d'Anjou	-0.763***	-0.179	-0.399***	-0.237	-0.161	-0.193*
	(0.216)	(0.246)	(0.141)	(0.160)	(0.116)	(0.115)
Basses Vallées Angevines	-0.228	-1.473***	-0.071	-0.863***	-0.646***	-0.669***
	(0.170)	(0.179)	(0.111)	(0.120)	(0.087)	(0.085)
Marches du Segréen	-1.298***	-0.883***	-0.763***	-0.674***	-0.452***	-0.468***
	(0.261)	(0.340)	(0.171)	(0.215)	(0.156)	(0.155)
Mauges	0.788***	-0.400***	0.585***	-0.241**	-0.246***	-0.257***
	(0.164)	(0.152)	(0.109)	(0.106)	(0.076)	(0.075)

➔

Table 4 – **Results of the estimation of the land allocation models** *(contd.)*

	Model (9)		Model (41)		Model (105)	Model (108)
	Forest	Urban	Forest	Urban	Urban	Urban
Plateaux de l'Aubance	-0.155	-0.012	0.036	-0.078	-0.104	-0.159
	(0.208)	(0.220)	(0.135)	(0.146)	(0.106)	(0.104)
Portes du Beaugeois	0.212	-0.722***	0.172	-0.451***	-0.377***	-0.420***
	(0.180)	(0.167)	(0.118)	(0.117)	(0.085)	(0.082)
PRA	Yes		Yes		Yes	Yes
Canton	Yes		Yes		Yes	Yes
OTEX	Yes		Yes		Yes	Yes
Landscape metrics	Yes		Yes		Yes	No
Methodology	Multinomial		Multinomial		Binomial	Binomial
Link	Logit		Probit		Probit	Probit
Observations	221,087		221,087		221,087	221,087
Log-likelihood	-51,477		-51,452		-15,068	-15,077
AIC	103,227		103,176		30,272	30,284

Notes: The standard errors are shown in brackets. ***, ** and * identify the significant parameters at the 0.01%, 0.05% and 0.1% thresholds.

the multinomial models). The two multinomial models differ only in respect of the link function used (logit or probit), whereas the dichotomous models are all probit models and differ only in respect of the inclusion or exclusion of landscape metrics. All of these models show consistent and very similar results. The coefficients for a final land use of agriculture (in 2010) are taken as a reference and normalised to 0. We therefore present estimated coefficients for the other two land use categories, forest and urban. A positive coefficient indicates[23] that the variable favours conversion to one of these other uses, with agriculture being used as the reference. The opposite is true of a negative coefficient.

As we are interested in the dynamics of urbanisation, we limit ourselves to a quick discussion of the positivity/negativity of the estimated coefficients for the urbanised plots in 2010. The linear term and the quadratic term for distance to Angers, i.e. distance to jobs, indicate a concave, inverted-U relationship between distance to the CBD and the probability of urbanisation. The top of the inverted U-curve is about 1.8 km from the centre of Angers. Thus, non-urbanised plots in the Angers municipality itself have a lower probability of being urbanised than plots in the immediate vicinity of Angers (around 2 km). This has two effects. Firstly, the spaces are highly valuable in their non-urbanised state, as they provide amenities and are therefore certainly protected. Furthermore, these areas essentially correspond to the banks of the Maine, which are subject to flooding. Beyond 2 km, plots have a decreasing probability of being urbanised, reflecting the effect of increased costs of transport to the workplace. The effect of distance to the city centre (SBD) is decreasing and convex. The probability of urbanisation is therefore higher in the immediate vicinity of the market towns than when further away from them, which reflects the value of proximity to services and social proximity, on the one hand, and, on the other, the unobserved effect of the planning documents which favour the contiguity of urban development. The decreasing and convex relationship for proximity to the main road network reflects the value of accessibility to Angers and the main regional centres.

Average income in the municipality has a positive effect, reflecting the preference of suburban households for better-off neighbourhoods. This is a classic manifestation of the forces at play behind segregation. The level of equipment in the municipalities also has a positive effect.

Natural amenities also play a role. Firstly, the municipalities located in the Loire-Anjou-Touraine Regional Nature Park have a greater probability of being urbanised than the others. Conversely, flood-prone areas have a much lower probability of being urbanised. The estimated parameters for slopes reflect a widely observed phenomenon: households value landforms and views, but land that is too rough has conversion costs that are too high and lower approval values.

Finally, the estimated parameters for neighbourhood externalities also show an inverted U-shaped relationship between the probability of

23. It is generally easier to interpret the coefficients of a logit or probit model, whether multinomial or dichotomous, in terms of marginal effects; however, as the coefficients of the first stage are not our focus, we limit ourselves to a quick discussion of their positivity/negativity.

conversion of a plot and its probability of urbanisation. In a sparsely urbanised area, increasing the level of urbanisation in the vicinity of a plot is favourable to the urbanisation thereof as this facilitates its conversion by lowering the costs of servicing the land. Beyond that, the probability of urbanisation decreases: the negative externalities of density (loss of sight, congestion, etc.) then become preponderant.

3.2. Stage Two: Analysis of the Role of Perceptions

For each of the estimated models and for each LU, we estimate the marginal probability that a pixel is urbanised in 2010, knowing that it was not urbanised in 2000. Table 5 shows the descriptive statistics for these effects. Belonging to the *"Agglomération angevine"* (the Greater Angers area) has the strongest effect on the probability of conversion to urban use. This effect is 3.6%, while the effect of belonging to any LU is 2.2%. Belonging to this LU therefore increases the conversion probability of a pixel by about 1.4%, compared to the average. As can be seen from the standard errors, the variability of the measured effects is low, indicating significant differences between the LUs.

To estimate the metamodel (3), we use the estimated marginal effects and their standard errors. To explain the variability of the conditional probability measurements, we regress them using indicators of vocabulary richness relating to several topics (dictionary) but also using the importance of different semantic fields (Tropes scenarios).

We focus our analysis on the semantic fields taken from the Landscape Atlases based on the Tropes scenario. The variables created are vocabulary distribution variables expressed as a percentage (also the residual category "Other concepts" grouping together mainly "tools" concepts has not been introduced in the analyses). The results can then be interpreted as measuring the impact of an increase in the share of vocabulary related to a given concept (compared to the residual category) on the probability of urbanisation in the LU. Vocabulary is then seen as an indicator of the presence or absence of amenities sought by individuals, with the implicit assumption that it is the nature and richness of the description that counts, not its positive or negative connotation. While this assumption can be debated at national level, it is reasonable in the context of our study.

The results of the estimations for (3) are shown in Table 6. As this is an internal meta-regression, using the same data and with similar models, the estimated inter-estimation variance τ^2 is low, which is normal for analyses of this type. We see that the content indicators of the landscape unit descriptions explain almost all of the variation in the measured effects. The proportion of the inter-estimation variance explained by the model is measured by the adjusted R^2 coefficient. The explanatory variables introduced in the meta-models thus make it possible to explain between 74% and 79% of the differences measured between LUs.[24] Furthermore, the I^2 coefficient provides an estimate of the total variance that can be attributed to differences between the models. The estimated meta-models

24. *This proportion should be compared with models that do not include perception variables (less than 20% for a full model without the introduction of Tropes scenarios or without the introduction of dictionary analyses, see Online Appendix C2).*

Table 5 – **Descriptive statistics for estimated marginal effects ($\widehat{P^m_{kui}}$)**

Landscape unit	Marginal effect		Minimum		Maximum	
	μ^m_P	$(\sigma^m_P)^2$	μ^m_P	$(\sigma^m_P)^2$	μ^m_P	$(\sigma^m_P)^2$
Agglomération Angevine	0.0360	0.0037	0.0258	0.0024	0.0444	0.0050
Loire des Promontoires	0.0308	0.0024	0.0239	0.0015	0.0399	0.0036
Beaugeois	0.0193	0.0012	0.0146	0.0009	0.0259	0.0016
Couloir du Layon	0.0257	0.0015	0.0217	0.0008	0.0316	0.0020
Haut Anjou	0.0112	0.0012	0.0074	0.0007	0.0147	0.0018
Saumurois	0.0258	0.0019	0.0168	0.0008	0.0302	0.0027
Segréen	0.0148	0.0011	0.0121	0.0009	0.0200	0.0013
Val d'Anjou	0.0230	0.0026	0.0154	0.0010	0.0340	0.0048
Basses Vallées Angevines	0.0170	0.0016	0.0116	0.0012	0.0227	0.0020
Marches du Segréen	0.0139	0.0031	0.0089	0.0021	0.0208	0.0050
Mauges	0.0224	0.0019	0.0143	0.0009	0.0278	0.0027
Plateaux de l'Aubance	0.0248	0.0024	0.0185	0.0011	0.0373	0.0044
Portes du Beaugeois	0.0217	0.0017	0.0159	0.0013	0.0282	0.0024
Total	0.0220	0.0020	0.0074	0.0007	0.0444	0.0050

Table 6 – **Results of the second stage estimation**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dictionary				Tropes scenario			
	base	controls	methods	complete	base	controls	methods	complete
Architecture	0.0016***	0.0017***	0.0017***	0.0017***				
	(0.0005)	(0.0005)	(0.0005)	(0.0005)				
Botany	-0.0054***	-0.0053***	-0.0054***	-0.0052***				
	(0.0016)	(0.0016)	(0.0016)	(0.0016)				
Economy	0.0016	0.0017*	0.0017	0.0017*				
	(0.0010)	(0.0010)	(0.0010)	(0.0010)				
Animal husbandry	0.0000	0.0001	0.0001	0.0001				
	(0.0018)	(0.0017)	(0.0018)	(0.0018)				
Mineralogy	0.0023	0.0025	0.0023	0.0025				
	(0.0028)	(0.0028)	(0.0028)	(0.0028)				
Urban planning	0.0137***	0.0135***	0.0137***	0.0135***				
	(0.0032)	(0.0032)	(0.0032)	(0.0032)				
Forestry	0.0087***	0.0086***	0.0087***	0.0086***				
	(0.0009)	(0.0009)	(0.0009)	(0.0009)				
Geology	0.0044*	0.0042*	0.0044*	0.0041*				
	(0.0024)	(0.0024)	(0.0024)	(0.0024)				
Countryside	-0.0014	-0.0016	-0.0014	-0.0017				
	(0.0025)	(0.0025)	(0.0025)	(0.0025)				
Viticulture	0.0028	0.0029*	0.0028*	0.0029*				
	(0.0017)	(0.0017)	(0.0017)	(0.0017)				
Religion	-0.0076**	-0.0072*	-0.0075**	-0.0071*				
	(0.0037)	(0.0037)	(0.0037)	(0.0037)				
Water	0.0047	0.0054	0.0043	0.0055				
	(0.0236)	(0.0235)	(0.0235)	(0.0235)				
Agri./Env.					0.0821***	0.0795***	0.0809***	0.0794***
					(0.0125)	(0.0123)	(0.0125)	(0.0124)
Anim./Plant.					0.0309***	0.0286***	0.0298***	0.0285***
					(0.0097)	(0.0096)	(0.0096)	(0.0096)
Arts/Culture					0.1312***	0.1286***	0.1303***	0.1286***
					(0.0278)	(0.0274)	(0.0276)	(0.0274)
Characteristics					-0.0822***	-0.0854***	-0.0839***	-0.0855***
					(0.0194)	(0.0192)	(0.0193)	(0.0192)
Comm./Media					0.3666***	0.3609***	0.3639***	0.3606***
					(0.0378)	(0.0373)	(0.0376)	(0.0373)
Behav./Feel.					-0.1511***	-0.1526***	-0.1520***	-0.1528***
					(0.0197)	(0.0194)	(0.0196)	(0.0194)
Strengths/Quantities					-0.0462*	-0.0453*	-0.0455*	-0.0451*
					(0.0238)	(0.0235)	(0.0237)	(0.0235)
Geography					0.0346***	0.0328***	0.0337***	0.0327***
					(0.0111)	(0.0109)	(0.0110)	(0.0109)
Politics/Society					0.1754***	0.1690***	0.1725***	0.1686***
					(0.0285)	(0.0282)	(0.0284)	(0.0282)
Transport					0.2086***	0.2129***	0.2105***	0.2131***
					(0.0261)	(0.0258)	(0.0260)	(0.0258)
OTEX		0.0005*		0.0005		0.0005*		0.0005
		(0.0003)		(0.0003)		(0.0003)		(0.0003)
PRA		0.0012***		0.0011***		0.0013***		0.0013***
		(0.0003)		(0.0003)		(0.0003)		(0.0004)
Cantons		0.0005**		0.0004		0.0006*		0.0005
		(0.0003)		(0.0005)		(0.0003)		(0.0006)
Metrics		-0.0001		-0.0002		-0.0001		-0.0001
		(0.0003)		(0.0004)		(0.0003)		(0.0004)

➔

Table 6 – **Results of the second stage estimation** *(contd.)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dictionary				Tropes scenario			
	base	controls	methods	complete	base	controls	methods	complete
AIC			-0.0257***	-0.0048			-0.0274***	-0.0035
			(0.0078)	(0.0185)			(0.0085)	(0.0202)
Probit			-0.0004	-0.0003			-0.0004	-0.0003
			(0.0003)	(0.0003)			(0.0003)	(0.0003)
Dichotomous			0.0001	0.0001			0.0000	0.0001
			(0.0003)	(0.0003)			(0.0003)	(0.0003)
Constant	0.0134	0.0110	-0.0116	0.0062	-0.0038	-0.0030	-0.0300***	-0.0063
	(0.0352)	(0.0350)	(0.0358)	(0.0393)	(0.0081)	(0.0080)	(0.0115)	(0.0211)
Observations	832	832	832	832	832	832	832	832
τ^2	1.04e-05	1.04e-05	1.04e-05	1.04e-05	1.28e-05	1.27e-05	1.29e-05	1.28e-05
I^2	0.805	0.806	0.806	0.807	0.818	0.818	0.818	0.819
Adjusted R^2	0.791	0.790	0.790	0.790	0.741	0.743	0.740	0.742
LR test ($\tau^2 = 0$)	4.93e-05	4.93e-05	4.93e-05	4.93e-05	4.93e-05	4.93e-05	4.93e-05	4.93e-05
Model F test	174.6	134.2	141.9	112.8	162.5	121.5	127.2	99.97

Notes: The standard errors are shown in brackets. ***, ** and * identify the significant parameters at the 0.01%, 0.05% and 0.1% thresholds.

therefore explain 81% to 82% of this variance. All of the tests show that the models are clearly significant.

We take into account the differences between the models estimated in the first stage to explain the variability of the measured effects. Indeed, if the information they convey is correlated with that conveyed by the LUs, then their exclusion may introduce an omitted variable bias. This is not the case here in the full model. The coefficient estimated using landscape metrics, OTEX or cantons are not significant. The dummies for the LUs therefore reflect a different dimension than these variables, which we interpret as the cultural component of the landscapes. We note, however, that the division into PRAs is significant. As we had anticipated (see subsection 2.3), the construction of the PRAs does reflect a cultural-historical aspect of landscapes. To account for the fact that some of the 64 models generating the data in this second stage are better models than others, we also introduce the Akaike Information Criterion (AIC), measured by the difference with the best model (which differs depending on whether the estimated model is multinomial or dichotomous). It is never significant, which indicates that the measured effects are independent of the quality of the model estimated in the first stage. Finally, it is interesting to note that there is no significant effect due to the method used (probit or logit link/categorical or dichotomous dependent variable).

Conversely, our textual and semantic indicators of the content of the descriptions of the LUs in the Landscape Atlas of the Maine-et-Loire department play a role in explaining the differences in urbanisation probabilities. LUs described with a richer vocabulary relating to agriculture and the environment had a higher probability of being urbanised. Thus a 1% increase in the share of these themes in the total vocabulary leads to an increase of about 0.08% in the probability of urbanisation. We find the same results for vocabulary related to the themes “Arts and Culture”, “Communications and Media”, “Politics and Society”, “Transport”, “Animals and Plants” and “Geography”. If we follow our assumption regarding a link between the description of landscapes and the nature of the amenities produced, we can thus assume that amenities relating to agriculture, the environment and the local social and cultural dynamics, while taking into account amenities relating to transport, have been drivers of the urbanisation of these areas. We thus see that the results validate the assumptions of the urban sprawl model in relation to environmental and agricultural amenities proposed by Coisnon *et al.* (2014). Conversely, a higher share of vocabulary related to the themes “Strengths and quantities” or “Characteristics” decreases the probability of urbanisation. One may think that this vocabulary (which includes terms such as level, mass and power) describes areas that are rather difficult to “live in” or less attractive because they produce fewer amenities sought by households. This may correspond to areas that are difficult and therefore costly to develop. Finally, the theme “Behaviour and feelings” is negatively related to the probability of urbanisation.

* *
*

In this article, we sought to assess the role of perceptual landscape elements on land use choices. We estimate a two-stage econometric land conversion model. In the first stage, we estimate the probability of urbanisation of a plot and then take into account its uncertainty using an internal meta-regression method. A textual analysis of the Landscape Atlases allows the introduction of landscape descriptors in a second stage. In this way, we can account for the impact of economic and landscape determining factors on urbanisation, in both their physical and perceptual aspects. Our estimates highlight the relative importance of the factors of urbanisation identified in the European and North American literature. We see that, in the case of the urban area of Angers as elsewhere, the probability of urbanisation depends on the proximity to the employment centre and to transport infrastructures, as well as on the living environment and neighbourhood externalities (average income of the municipality, public facilities, natural amenities and surrounding urbanisation).

Our estimates also show that the conditional probability of a given location being urbanised is significantly dependent on its belonging to a landscape unit. The diversity of the positivity/negativity of the associated coefficients makes it possible to highlight the heterogeneity of the preferences expressed with regard to this or that landscape unit. This probability was removed from the physical aspect of the landscapes in the first stage. It is linked, through a meta-regression, to descriptors constructed from texts describing the landscape units. This approach makes it possible to account for the uncertainty associated with model selection. We can thus confirm that cultural aspects of landscapes play a significant role in urbanisation and we can identify the components of perceptions that play the most important role. Thus, the territories described with greater richness in agricultural, political and societal terms,[25] which are therefore more likely to produce the associated amenities, seem to be more sought after by households. The relative significance of these effects, which have been revealed in the case of the urban area of Angers, could however be different for other urban areas.

Taking into account the sensitive aspect of landscapes, beyond their physical characteristics, thus provides a way to better understand the residential location choices of households. The landscape, as perceived by individuals, can explain the desertion of certain rural peripheral areas characterised by a low landscape quality or, conversely, the residential attractiveness of certain areas, entailing a risk of facing an increased urban sprawl phenomenon.

To our knowledge, this work is pioneering in respect of its incorporation of a landscape perception indicator for the analysis of urbanisation phenomena. It is fully in line with the development of so-called mixed approaches in social sciences. Other recent work also highlights the important role of perceptions. For example, Jones & Dantzler (2021) show that perceptions of different neighbourhoods in a city shape residential mobility.

Here, this mixed approach is necessary because of the complexity of the social term "landscape", which is difficult to reduce to a single aspect, either as physical elements (observable and objectifiable), or as perceptions of individuals or groups of individuals (difficult to quantify and observe). However, it must be stressed that this approach is not only methodologically demanding but also costly. Indeed, we had to create two original datasets, one using satellite images 10 years apart, the other to construct perception data, which are linked to other databases. Application to other urban areas is therefore subject to the availability of similar data. This is why spreading this work will mean it is necessary to find solutions in the absence of a Landscape Atlas.

One possibility would be to use dated and geocoded information generated by social networks. For example, using data from Twitter, Park *et al.* (2021) show that it is possible to identify areas of a city that generate feelings of happiness or dissatisfaction. Using data from the platform Yelp, Glaeser *et al.* (2018) show that the information generated by social networks can not only provide a better understanding of gentrification phenomena but can also predict them, almost in real time. Such data could be used in the model presented in this article. ☐

25. Terms related to the themes "Agriculture and Environment", "Arts and Culture", "Communications and Media" and "Politics and Society".

Link to the Online Appendix:
https://www.insee.fr/en/statistiques/fichier/6005377/ES528-529_Bourbeillon-et-al_Annexe-en-ligne_Online-Appendix.pdf

BIBLIOGRAPHY

Ambroise, R. (2010). Appréhender le paysage sous l'angle de son fonctionnement pour gérer la biodiversité. *Espaces Naturels*, 32, 20–26.
http://www.espaces-naturels.info/apprehender-paysage-sous-angle-son-fonctionnement-pour-gerer-biodiversite

Ay, J.-S., Chakir, R. & Le Gallo, J. (2017). Aggregated versus individual land-use models: Modeling spatial autocorrelation to increase predictive accuracy. *Environmental Modeling & Assessment*, 22(2), 129–145.
https://doi.org/10.1007/s10666-016-9523-5

Banzhaf, H. S. & Smith, V. K. (2007). Meta-analysis in model implementation: choice sets and the valuation of air quality improvements. *Journal of Applied Econometrics*, 22(6), 1013–1031. https://doi.org/10.1002/jae.977

Blanc, A., Peton, H. & Garcias, F. (2019). L'analyse lexicométrique des macro discours par les vocabulaires – enjeux théoriques et méthodologiques. *Finance Contrôle Stratégie*, (NS-6). https://doi.org/10.4000/fcs.3459

Brunet-Vinck, V. (2004). Méthode pour les Atlas de paysages – Enseignements méthodologiques de 10 ans de travaux. Rapport technique, Ministère de l'Ecologie et du Développement Durable.
http://temis.documentation.developpement-durable.gouv.fr/document.html?id=Temis-0070141

Bryan, M. L. & Jenkins, S. P. (2016). Multilevel modelling of country effects: A cautionary tale. *European Sociological Review*, 32(1), 3–22. https://doi.org/10.1093/esr/jcv059

Burnham, K. & Anderson, D. (2004). Multimodel inference. Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33(2), 261–304. https://doi.org/10.1177/0049124104268644

Cassette, A. & Paty, S. (2006). La concurrence fiscale entre communes est-elle plus intense en milieu urbain qu'en milieu rural ? *Cahiers d'Économie et Sociologie Rurales*, 78, 6–30.
https://ageconsearch.umn.edu/record/201679

Chakir, R. & Parent, O. (2009). Determinants of land use changes: A spatial multinomial probit approach. *Papers in Regional Science*, 88(2), 327–344. https://doi.org/10.1007/s00701-011-0993-6

Chomitz, K. M. & Gray, D. A. (1996). Roads, land use, and deforestation: a spatial model applied to Belize. *The World Bank Economic Review*, 10(3), 487–512. https://doi.org/10.1093/wber/10.3.487

Coisnon, T., Oueslati, W. & Salanié, J. (2014). Urban sprawl occurrence under spatially varying agricultural amenities. *Regional Science and Urban Economics*, 44, 38–49. https://doi.org/10.1016/j.regsciurbeco.2013.11.001

Coisnon, T., Rousselière, D. & Rousselière, S. (2019). Information on biodiversity and environmental behaviors: A European study of individual and institutional drivers to adopt sustainable gardening practices. *Social Science Research*, 84,102–123. https://doi.org/10.1016/j.ssresearch.2019.06.014

Council of Europe (2000). *Convention Européenne du Paysage – Florence, 20 octobre 2000*. Division de l'aménagement du territoire et du paysage, Bruxelles.
https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=176

Cosgrove, D. (2003). *Handbook of cultural geography*, Chapter: Landscape and the European sense of sight–eyeing nature, pp. 249–268. London: Sage Publications. http://dx.doi.org/10.4135/9781848608252.n17

Couch, C., Leontidou, L. & Petschel-Held, G. (2007). *Urban Sprawl in Europe*. Oxford: Black-Well Publishing.

EEA (2019). Land and soil in Europe: why we need to use these vital and finite resources sustainably. Report 2443-7662, European Environment Agency.
https://www.eea.europa.eu/publications/eea-signals-2019-land/download

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. & Rubin, D. (2014). *Bayesian Data Analysis, Third Edition*. Boca Raton: CRC Press. http://www.stat.columbia.edu/~gelman/book/BDA3.pdf

Glaeser, E. L., Kim, H. & Luca, M. (2018). Nowcasting gentrification: using yelp data to quantify neighborhood change. *American Economic Review: Papers and Proceedings*, 108(2), 77–82.
http://dx.doi.org/10.1257/pandp.20181034

Grekousis, G. & Mountrakis, G. (2015). Sustainable development under population pressure: Lessons from developed land consumption in the conterminous U.S. *PLOS ONE*, 10(3), e0119675.
https://doi.org/10.1371/journal.pone.0119675

Groult, M. & Roche, A. (2009). *Carte des Unités de Paysage contenue dans les Atlas de Paysage publiés entre 1993 et 2009*. CNRS – UMR 7533 LADYSS, Paris.

Irwin, E. & Bockstael, N. (2002). Interacting agents, spatial externalities and the evolution of residential land use patterns. *Journal of Economic Geography*, 2(1), 31–54. https://doi.org/10.1093/jeg/2.1.31

Irwin, E. G. & Geoghegan, J. (2001). Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment*, 85(1), 7–24.
https://doi.org/10.1016/S0167-8809(01)00200-6

Jones, A. & Dantzler, P. (2021). Neighbourhood perceptions and residential mobility. *Urban Studies*, 58(9), 1792–1810. https://doi.org/10.1177/0042098020916440

Klemick, H., Griffiths, C., Guignet, D. & Walsh, P. (2018). Improving water quality in an iconic estuary: an internal meta-analysis of property value impacts around the Chesapeake Bay. *Environmental and Resource Economics*, 69(2), 265–292. https://doi.org/10.1007/s10640-016-0078-3

Kline, J. D., Moses, A. & Alig, R. J. (2001). Integrating urbanization into landscape-level ecological assessments. *Ecosystems*, 4(1), 3–18. https://doi.org/10.1007/s100210000056

Knapp, G. & Hartung, J. (2003). Improved tests for a random-effects meta-regression with a single covariate. *Statistics in Medicine*, 22, 2693–2710. https://doi.org/10.1002/sim.1482

Kozlowski, A. C., Taddy, M. & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5), 905–949. https://doi.org/10.1177/0003122419877135

Kuminoff, N. V., Zhang, C. & Rudi, J. (2010). Are travelers willing to pay a premium to stay at a "green" hotel? Evidence from an internal meta-analysis of hedonic price premia. *Agricultural & Resource Economics Review*, 39(3), 468–484. https://doi.org/10.1017/S1068280500007450

Lebart, L. (1994). Sur les analyses statistiques de textes. *Journal de la Société Française de Statistique*, 135(1), 17–36. http://www.numdam.org/item/JSFS_1994__135_1_17_0/

Lopez, R., Adelaja, A. & Andrews, M. (1988). The effects of suburbanization on agriculture. *American Journal of Agricultural Economics*, 70(2), 346–358. https://doi.org/10.2307/1242075

Loughran, T. & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187–1230. https://doi.org/10.1111/1475-679X.12123

Maire, S. & Liarte, S. (2019). Classifier, représenter et labelliser : Cadre et outils méthodologiques associés pour une émergence automatique de thématiques sur données textuelles. *Finance Contrôle Stratégie*, (NS-6). https://doi.org/10.4000/fcs.3423

McGarigal, K., Cushman, S. A., Neel, M. C. & Ene, E. (2002). Fragstats: spatial pattern analysis program for categorical maps. Rapport technique, University of Massachusetts, Amherst. http://www.umass.edu/landeco/research/fragstats/fragstats.html

Molette, P. (2009). De l'APD à Tropes : comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste. *In Actes du colloque Psychologie sociale de la communication*, Tarbes, Juin 2009. https://www.tropes.fr/PierreMolette-CommunicationColloquePsychoTarbesJuin2009.pdf

Molette, P. (2014). Comportements et sentiments. de l'ambiguïté dans les émotions? journées d'étude sur la mesure des émotions dans les corpus textuels. Rapport technique, Atelier Texto du Labex SMS, Toulouse.

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*. 29(1), 46–75. https://doi.org/10.1111/joes.12044

Nery, T., Polyakov, M., Sadler, R. & White, B. (2019). Spatial patterns of boom and bust forestry investment development: A case study from Western Australia. *Land Use Policy*, 86, 67–77. https://doi.org/10.1016/j.landusepol.2019.04.015

Newburn, D. A., Berck, P. & Merenlender, A. M. (2006). Habitat and open space at risk of land-use conversion: targeting strategies for land conservation. *American Journal of Agricultural Economics*, 88(1), 28–42. https://doi.org/10.1111/j.1467-8276.2006.00837.x

Nowak, A. & Smith, P. (2017). Textual analysis in real estate. *Journal of Applied Econometrics*, 32(4), 896–918. https://doi.org/10.1002/jae.2550

Park, Y., Kim, M., & Seong, K. (2021). Happy neighborhoods: Investigating neighborhood conditions and sentiments of a shrinking city with Twitter data. *Growth and Change*, *52*(1), 539–566. https://doi.org/10.1111/grow.12451

PAYTAL (2014). PAYsage et éTALement urbain: dynamiques physique et culturelle et modélisation. Athané, E., Bourbeillon, J., Coisnon, T, Guyet T., Nicolas H., Oueslati W., Rousselière, D. & Salanié, J. (coordinateur), Agrocampus Ouest.
http://temis.documentation.developpement-durable.gouv.fr/documents/Temis/0082/Temis-0082075/22025_Rapport.pdf

Polèse, M. & Shearmur, R. (2005). *Économie urbaine et régionale : introduction à la géographie économique*. Paris: Economica.

Polyakov, M. & Zhang, D. (2008). Population growth and land use dynamics along urban– rural gradient. *Journal of Agricultural and Applied Economics*, 40(2), 649–666. https://doi.org/10.1017/S1074070800023919

Roche, A. (2007). Les unités et structures paysagères dans les atlas de paysage. Rapport technique, Ministère de l'Ecologie, du Développement et de l'Aménagement Durables.

Saxonhouse, G. R. (1976). Estimated parameters as dependent variables. *The American Economic Review*, 66(1), 178–183. https://www.jstor.org/stable/1804956

Sutton, A. J. & Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27, 625–650. https://doi.org/10.1002/sim.2934

Tsai, F. T.-C. & Li, X. (2008). Inverse groundwater modeling for hydraulic conductivity estimation using bayesian model averaging and variance window. *Water Resources Research*, 44(9), W09434. https://doi.org/10.1029/2007WR006576.

Uuemaa, E., Antrop, M., Roosaare, J., Marja, R. & Mander, Ü. (2009). Landscape metrics and indices: an overview of their use in landscape research. *Living Reviews in Landscape Research*, 3(1), 1–28. http://lrlr.landscapeonline.de/Articles/lrlr-2009-1/

APPENDIX

DESCRIPTIVE STATISTICS

	Description	Mean	Standard error	Sources
Land use in 2010	Agricultural	0.781		Authors
	Forest	0.113		
	Urban	0.106		
Land use in 2000	Agricultural	0.820		
	Forest	0.092		
	Urban	0.088		
Vicinity urbanised in 2000	Number of urbanised pixels within a 200 m radius	0.088	0.212	
P/A ratio	Perimeter to area ratio in 2000	0.468	0.213	
Shape index	Shape index in 2000	2.799	0.795	
FD index	Fractal dimension index in 2000	1.312	0.101	
CBD distance	Distance to Angers (km)	18.951	7.026	IGN[a]
SBD distance	Distance to nearest town (km)	1.862	0.875	
Distance to road	Distance to main road network (km)	2.729	2.169	
Slope	Height difference within the pixel (m)	2.663	2.899	
PNR	Municipality of a regional nature park (dummy)	0.152		
Municipality income	Average income/capita in the municipality (euros) in 2000	14,713	3,467	INSEE[b]
Municipality equipment	Number of pieces of equipment in the municipality in 1998	117.6	509.1	INSEE[c]
Susceptible to flooding	Recognised flood zones	0.045		DREAL
PRA	Small agricultural regions in 1981 (dummy)			Agreste[d]
	Beaugeois	0.220		
	Angers woodland	0.430		
	Choletais	0.147		
	Saumurois	0.071		
	Vallée de la Loire	0.132		
OTEX	Technical/economic orientation of the municipality in 2000 (dummy)			Agreste[d]
	Dairy cattle	0.009		
	Mixed cattle	0.040		
	Fruit-Permanent Crops	0.007		
	Field crops	0.024		
	Mixed Granivores	0.059		
	Horticulture	0.067		
	Vegetable growing	0.012		
	Oilseeds	0.001		
	Polyculture-Animal husbandry	0.667		
	Viticulture	0.102		
	Poultry	0.013		

➔

	Description	Mean	Standard error	Sources
Canton	Belonging to a canton (dummy)			Insee[e]
	Angers	0.017	-	
	Angers-Nord-Est	0.030	-	
	Angers-Est	0.010	-	
	Baugé	0.023	-	
	Beaufort-en-Vallée	0.072	-	
	Chalonnes-sur-Loire	0.024	-	
	Châteauneuf-sur-Sarthe	0.091	-	
	Doué-la-Fontaine	0.006	-	
	Durtal	0.019	-	
	Gennes	0.029	-	
	Lion-d'Angers	0.087	-	
	Louroux-Béconnais	0.091	-	
	Ponts-de-Cé	0.086	-	
	Saint-Georges-sur-Loire	0.070	-	
	Seiches-sur-le-Loir	0.088	-	
	Thouarcé	0.111	-	
	Tiercé	0.066	-	
	Angers-Trélazé	0.022	-	
	Angers-Ouest	0.016	-	
	Angers-Nord	0.037	-	
	Angers-Nord-Ouest	0.006	-	
Landscape unit	Belonging to an LU (dummy)			LADYSS
	Agglomération angevine	0.030	-	
	Loire des promontoires	0.041	-	
	Beaugeois	0.143	-	
	Couloir du Layon	0.077	-	
	Haut Anjou	0.211	-	
	Saumurois	0.063	-	
	Segréen	0.143	-	
	Val d'Anjou	0.065	-	
	Basses vallées angevines	0.039	-	
	Marches du Segréen	0.080	-	
	Mauges	0.013	-	
	Plateaux de l'Aubance	0.054	-	
	Portes du Beaugeois	0.040	-	

(a) BDALTI, MNT500, BDTOPO, Geoportail – authors' calculations.
(b) Population census.
(c) 1998 municipal inventory.
(d) 2000 agricultural census, Zoning into Agricultural Regions.
(e) Official Geographical Code.