Renewable and Sustainable Energy Reviews 70 (2017) 1–12

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

The spatio-temporal dynamics of ethanol/gasoline price ratio in Brazil☆

Márcio Poletti Laurini

FEA-RP USP, Av. Dos Bandeirantes 3900, 14040-905, Ribeirão Preto, SP, Brazil

ARTICLE INFO

Keywords:
Spatio-temporal model
Ethanol/gasoline price ratio
Spatial random effects
Co-integration

ABSTRACT

We use an innovative method to analyze the spatio-temporal evolution of the ethanol/gasoline price ratio to end consumers in Brazil. This model allows estimating the distribution of prices throughout the country using a continuous space model estimated by Bayesian methods. We use data from the National Petroleum, Natural Gas and Biofuels Agency's price survey system, a weekly sample of 10% of fuel suppliers in Brazil, to estimate continuous projections of the price ratio for the entire country for the 2007–2014 period. We use this model to build an indicator of ethanol advantage in the country and show how this advantage has declined since 2009.

1. Introduction

In this work we analyze the dynamics of spatial dispersion of ethanol/gasoline price ratio in Brazil. Using the consumer price data obtained weekly by the National Petroleum, Natural Gas and Biofuels Agency (*Agência Nacional de Petróleo, Gas Natural e Biocombustíveis* - ANP), we propose a spatio-temporal model to capture the spatial distribution of the relative ratio of consumer prices. This model is based on a continuous representation of the spatial random effects, and thus allows estimating the price distribution process continuously in space, and also the temporal dynamics. The results of this procedure can be used to analyze fuel demand, market competition and spatial price arbitrage.

Our model allows us to analyze the evolution of relative prices in the country by performing an interpolation procedure for all locations not sampled in the ANP price survey. Based on the estimated model, we get continuous projections of the estimated price ratio between the prices of ethanol and gasoline, and construct an indicator of the probability of economic advantage of using ethanol, employing the posterior distribution of the price ratios. The obtained results indicate a change in the pattern of relative prices from 2010. In the 2007–2009 period, ethanol fueling was advantageous in most Brazilian territory, while since 2010 this advantage has declined steeply so that the most recent data show that ethanol is advantageous only in part of the Southeast region.

The space-temporal analysis of fuel prices using the available sample has non-trivial difficulties in the process of inference. In particular, the data are not sampled on a regular grid, i.e., in each period we have observations from different locations, making it difficult to use the usual methods of econometrics and spatial statistics, based on regular grids and neighborhood/lattice structures, e.g., [1,2]. These methods also fail to capture adequately the continuous pattern of spatial dependence observed in this phenomenon. To the best of our knowledge, there is no work properly exploring the spatial and temporal dimensions of this price dispersion pattern.

The methodology is based on the use of a spatial covariance matrix of the Matérn class, projected continuously in space. The representation used, introduced by [3], is based on the equivalence between the solution of a stochastic partial differential equation and a Markov random field. This representation allows designing the spatial random effects continuously in space, using a finite elements representation, and exploring a continuous version of the spatial Markov property that allows us to represent this process as a Gaussian Markov random field. This formulation also enables the use of a hierarchical representation and Bayesian inference methods, based on integrated nested Laplace approximations [4], which are accurate analytical approaches for the estimation of parameters and latent variables in Gaussian Markov random fields.

The model also permits obtaining an estimate of the spatial random effects for each point in space, i.e., we can get estimate the impact of the spatial component on the price ratio continuously in the territorial space analyzed. This space-time model allows analyzing the patterns of spatial and temporal dependence of this component, and directly analyzes the distribution of relative prices in each moment in time and space from the sampled prices. The results show that this spatial component is very persistent in time, confirming and quantifying the importance of the spatial aspects in the no-arbitrage relationship between the prices of ethanol and gasoline.

☆ I thank the comments from the editor Lawrence L. Kazmerski and the anonymous referees. The author acknowledges the assistance of CNPq (Grant number 303738/2015-4) and FAPESP (Grant Number 13/00506-1).

E-mail address: mplaurini@gmail.com.

http://dx.doi.org/10.1016/j.rser.2016.11.195
Received 23 September 2015; Received in revised form 15 October 2016; Accepted 12 November 2016
1364-0321/

Our analysis further allows analyzing some relevant aspects in the fuel market, and is a way to measure the aggregate effect of variables such as production and transportation costs, consumer preferences, the proportion of flex-fuel vehicles and market concentration on fuel prices in each locality. As some of these variables are complex to measure, especially in a high spatial disaggregation level, our method allows measuring the impact of these variables on the observed spatial and temporal distribution of prices.

This article has the following structure – in Section 2 we present some important facts about the ethanol market in Brazil, the sample of consumer fuel prices used in this work and some descriptive statistics of this process. In Section 3 we show the key elements of the proposed space-time model. The results are shown in Section 4, and final conclusions in Section 5.

2. Material and methods

2.1. Ethanol market in Brazil

The use of ethanol as fuel for light vehicles is one of the greatest technological innovations developed in Brazil. Indeed the country is considered a pioneer in production of biofuels and use, which together are a key component of the recent Brazilian economic dynamics. Ethanol use as fuel is related to the existing production of sugarcane, and the origins of its use as vehicle fuel blended with gasoline dates to the 1930s, e.g. [5]. In turn, the use of hydrous ethanol as the sole automotive fuel started in response to petroleum crises in 1973 and 1979 and the sharp reduction in sugar prices, at a time when Brazil imported most of its crude oil, as discussed, for example, in [6]. The key step was the creation of the Proálcool program in 1975, a large-scale replacement program of oil-based fuels using ethanol. Detailed analysis of the Brazilian ethanol evolution can be found in [7–9]. Prospects for the future of ethanol fuel can be seen in [10], and further aspects conditioning the market penetration of ethanol are discussed in [11].

The Proálcool program and the further incentives for adoption of ethanol were based on price subsidies, the technological development of engines powered by that fuel and a combination of various economic incentives. Among them was the incremental addition of ethanol to gasoline, reaching 25% (E25 Blend),[1] price subsidies for ethanol, economic benefits to both end consumers and the supply chain via credit to producers and tax incentives for the purchase and maintenance of vehicles running on ethanol. Also essential were the mandatory sale of ethanol in service stations and the creation of strategic reserves to manage prices in off-season periods, as summarized in [5]. See [9,13] for a recent overview of biofuel policies.

Ethanol production can also be seen in the context of a backstop technology, being a renewable fuel substitute for exhaustible fossil fuels. In this case, the technology becomes economically viable when the average cost of production is below the spot cost of the exhaustible fuel, as discussed in [14–17]. The use of ethanol by the end user becomes advantageous when the relative price of ethanol in relation to gasoline is advantageous in terms of energy efficiency. Since the end user price also involves the optimal taxation of each fuel, a potential way to encourage the backstop renewable fuel consumption is the introduction of additional charges for the use of exhaustible fuels or subsidies for renewable fuel consumption. A discussion of the optimal taxation of a backstop technology can be found in [18].

The ethanol incentive policy in Brazil had increasing success until 1985, when ethanol-fueled vehicles reached a share of 85% in the sale of new vehicles. However, from 1985, with the reduction in international oil prices and the large increase in sugar prices, the economic advantages of ethanol production declined substantially. This effect was amplified by the reduction of tax incentives in 1986 and in the following years there were significant problems in ethanol production, such as large supply shortages in 1989. These supply problems led to a decrease in the percentage of addition of ethanol to gasoline and the need to import bioethanol and methanol. These effects had a major impact on the use of ethanol, with a reduction in the registration of new vehicles using this fuel to 11.4% of sales in 1989, as discussed in [5].

This negative outlook lasted until 2003, with the development of vehicles designed to run on either gasoline or ethanol or any blend of the two, known as flex fuel vehicles. The flex fuel engines are the result of incentive policies for biofuel production and innovations in the automotive industry (e.g., [5,6,19]). This technological development had a major impact on demand and dynamics of fuel prices in Brazil. Vehicles with flex fuel motors accounted for 88.5% of new licensing of light vehicles in Brazil and they represented about 64% of the total fleet of light vehicles in 2013 [20]. Ethanol is the fuel with the third largest demand in Brazil, only after diesel and gasoline [6]. Detailed statistics on production and ethanol demand can be found at the ANP's website.

The introduction of flex fuel motors has had a major impact on the demand for fuel due to the no-arbitrage relationship that is possible with the free choice of ethanol/gasoline in the fueling process. Due to issues of energy efficiency, fueling with ethanol is advantageous to the average consumer when its price is less than 70% of the price of gasoline (e.g. [5,19]),[2] and thus the consumption of gasoline and ethanol is indifferent to consumers when the price reaches this level. See [19] for an empirical analysis of the impact of flex fuel engines on fuel consumption in Brazil. [21] discuss this no-arbitrage relationship, and show there are other important questions about preferences that affect this choice, and [22,12,11] also discuss the energy efficiency aspects related to ethanol/gas substitution. Detailed studies of ethanol demand in Brazil can be found in [6,23,24,19,25], and studies of price elasticities can be found in [26–28].

An important issue on the demand and the pricing of ethanol and its relation to the price of gasoline is the location of production, distribution and consumption centers (e.g., [29–31]). As discussed in [23], ethanol production is concentrated in the Center-South region of the country, with secondary production in the Northeast states. The concentration in the Center-South region is given by advantages of higher productivity of sugarcane and ethanol, and also due to proximity to the main consumer centers, although there is a larger relative demand for flex-fuel vehicles in the North and Northeast regions due to lower per-capita income in these regions.

2.2. Description of sample

We use data from the price collection system of the National Petroleum, Natural Gas and Biofuels Agency (ANP), available at http://www.anp.gov.br/preco/. This system collects a weekly sample of consumer prices of fuels in about 10% of Brazilian municipalities. Table 1 shows the number of municipalities that were surveyed in each year of our sample. In each week a personal visit is paid to a rotating sample of service stations in each municipality, and other price data are confirmed by phone. The ANP price collection system stores the purchase and sale prices of gasoline, ethanol, diesel, s10 diesel and natural gas, and also the name of the station, address, city and state.

From the available addresses in each sample, we realized a geocoding process using the API from Google Maps. This geocoding tool returns the addresses in terms of latitude and longitude with a precision of 7 digits. The geocoding system could identify about 93% of the addresses available, which represents a significant portion of the

[1] See [12] for environmental impacts and sustainability aspects of ethanol/gasoline blend.

[2] The 70% threshold is an approximation, but it became the usual pattern of price comparison. More recent estimates suggest that the 70–80% range captures the variation of energy efficiency of more modern engines, the air-fuel mixture, and variations in the octane of fuels.

Table 1
Number of municipalities in the sample.

Year	Municipalities
2007	588
2008	573
2009	586
2010	583
2011	573
2012	567
2013	590
2014	589

sample. We applied a series of filters with the returned locations to eliminate possible geocoding errors, duplicate addresses and other possible problems. Then we filtered the sample to contain only stations with full information on ethanol and gasoline prices, required to construct the price ratio variable. Also, we filtered for possible annotation errors in prices.

After this geocoding and filtering procedure, the final sample corresponded to a total of 2,835,818 weekly observations, with the sample beginning on February 1, 2007 and ending on June 26, 2014. Although could analyze the data with the weekly frequency, due to computational limitations and to facilitate the interpretation of the results we aggregated the data to a quarterly sample, taking as observation in each location the average price ratios for all samples from the same service station in the same quarter. This quarterly aggregation procedure reduced the sample to 326,916 observations with distinct prices. Our sample has a total of 21,198 distinct locations, as each location can be sampled more than once during the survey period.

In Fig. 1 we show the spatial distribution of all service stations in the sample between 2007 and 2014, where the vertical and horizontal coordinates are latitude and longitudes obtained by the geocoding process. In Fig. 1 we also show the territorial division of Brazilian municipalities.

Figs. 2 and 3 show the distribution of gasoline and ethanol prices and also the price ratio using boxplots for each observed year. The results show a large price dispersion across the sample, reflecting the large spatial dispersion in prices. We can also note that on average in years 2007–2009 the supply of ethanol was advantageous, since the average ratio is below 0.7 (Fig. 4).

We present some descriptive statistics of the dependent variable (ethanol/gasoline price ratio at each station) in Table 2. In addition to information about the quantiles, minimum and maximum values, median and mean, we present the proportion of price ratios observed in each year which are below the threshold of 0.7, which indicates the advantage of fueling with ethanol. It can be seen that until 2009 the ethanol fueling was advantageous in most of the observations in the sample, a pattern that reversed from 2011 onward. This trend can be explained by the relative increase in ethanol prices, due to the restrictions on gasoline price transfers adopted by the federal government to fight aggregate inflation acceleration, which represents a gap of about 20% in gasoline prices in 2014, and the elimination of the levy called CIDE (*Contribuição de Intervenção no Domínio Econômico*), or Contribution for Interventions in the Economic System. The rate was 28 centavos (R$ 0.28) for each liter of gasoline from 2010, as discussed in [32]. According to this source, these two effects had a very negative impact on the production chain of ethanol, leading to temporary shutdown or closure of 44 ethanol distilleries in the past five sugarcane growing seasons before 2014, and impacts on the renewal and maintenance of sugarcane plantations, significantly reducing the output of the ethanol sector.

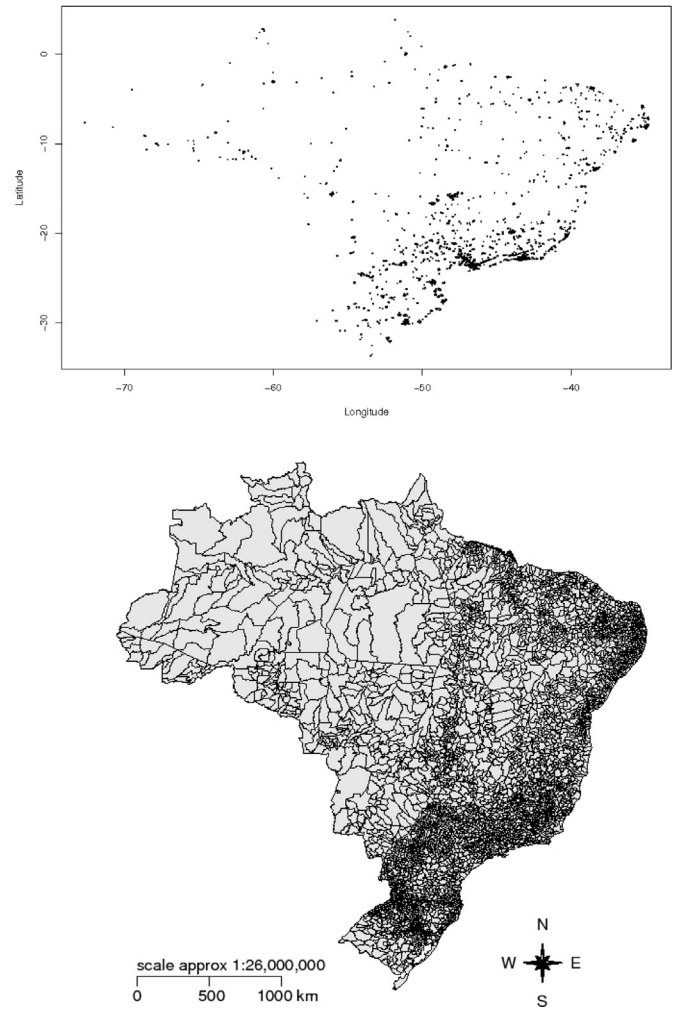

Fig. 1. Spatial distribution of service stations and territorial division of Brazilian municipalities.

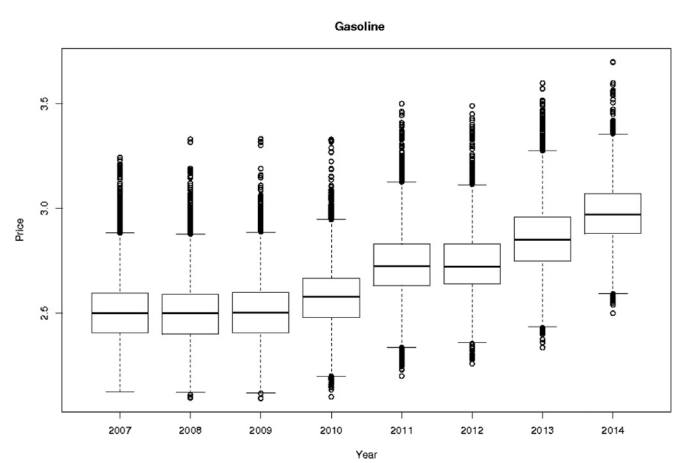

Fig. 2. Distribution of gasoline prices – boxplots.

3. Theory

3.1. Continuous space-temporal model

To obtain a space-time model for the price ratios, we use a representation of continuous spatial models proposed in [3], using the fundamental results obtained by [33,34] on the relationship between the solution of a stochastic partial differential equation and

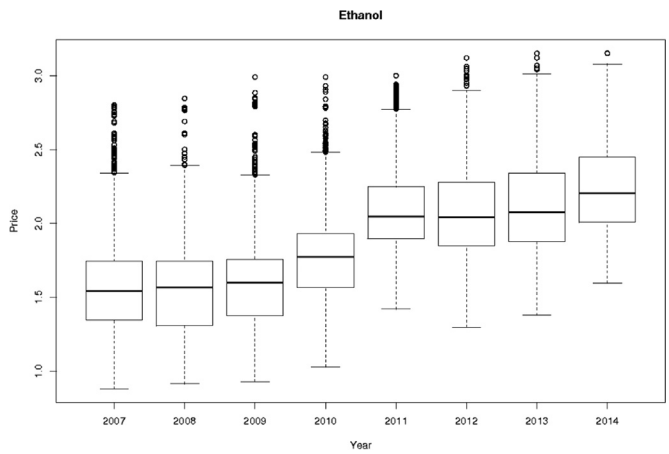

Fig. 3. Distribution of ethanol prices – boxplots.

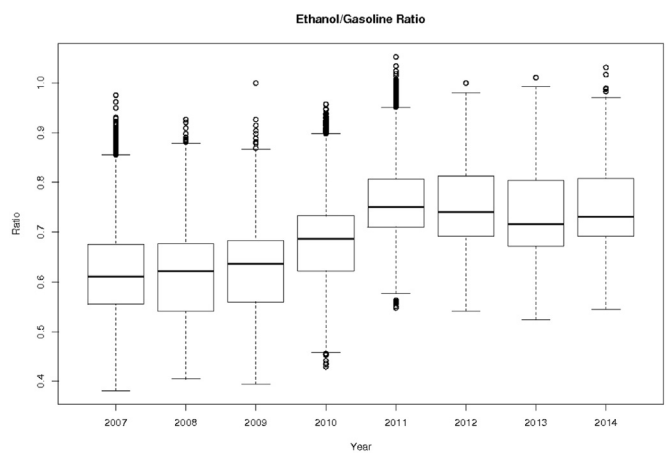

Fig. 4. Distribution of ethanol/gasoline price ratio – boxplots.

a Markov random field. Coupled with a continuous version of the spatial Markov property, the representation proposed in [3] allows representing a broad class of spatial and spatio-temporal processes as Gaussian Markov random fields. This formulation also has some computational advantages, in particular the sparse matrix representation and a continuous projection using finite elements and a hierarchical representation that allows the use of Bayesian inference based methods and Laplace approximations for inference procedures.

As stated above, this procedure is based on Gaussian Markov random field (GMRF) stochastic processes (see e.g. [35,36]). The key result in the representation of the class of continuous spatial models proposed in [3], based on the results of [33,34], is the relationship between a covariance function in a random field $x(u)$ and the solution of the following stochastic partial differential equation:

$$(\kappa - \Delta)^{\alpha/2} x(u) = W(u), \quad u \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{1}$$

with $(\kappa - \Delta)^{\alpha/2}$ denoting a pseudo-differential operator with Δ a Laplacian associated with:

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$$

and $W(u)$ an innovation process corresponding to a Gaussian spatial white noise, with marginal variance given by:

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}$$

The major contribution introduced in [3] in to exploit this equivalence to build a continuous representation of the spatial covariance structure. This representation exploits the fact that the equivalent stochastic partial differential equation can be discretized using finite element methods (e.g. [37]) using triangulation of the analyzed space through basis expansions. The relationship between this Gaussian field and the finite-dimensional solution for the partial differential equation given by Eq. (1) is obtained with the distribution of weights that solves the equation for a specific set of test functions. The second contribution of the method proposed in [3] is the possibility of using this representation via a hierarchical formulation. In particular [38] show that it is possible to represent continuously indexed spatio-temporal models $Y(s, l)$ in the form:

$$Y(s, l) = \{y(s, l)): (s, l) \in \mathcal{D} \subseteq \mathbb{R}^2 \times \mathbb{R}$$

with s being a vector of spatial locations and l a time index. This model is characterized by the space-temporal covariance function $Cov((s, l)(s', l')) = \sigma^2 C((s, l)(s', l'))$, defined for each (s, l) and $(s', l') \subseteq \mathbb{R}^2 \times \mathbb{R}$. We assume that this process is covariance stationary and depends only on the distance between the location and time via distance vectors $h = (s - s')$ and time lags $l = (t - t')$. We represent this process through a hierarchical form of a space-time model given by:

$$y(s, t) = z(s, t)\beta + \xi(s, t) + \varepsilon(s, t)\xi(s, t) = a\xi(s, t-1) + \omega(s, t) \tag{2}$$

In this process $z(s, t)$ represent observed covariates in (s, t), $\xi(s, t)$ denotes the (random) effects at each spatial location and time with the associated loading β, and analogously $\varepsilon(s, t)$ represents the component of idiosyncratic innovations, with $\varepsilon(s, t) \sim N(0, \sigma_\varepsilon^2)$. This component σ_ε^2 is the called the *nugget effect* in geospatial statistics. The temporal dynamics is given by the state equation $\xi(s, t) = a\xi(s, t-1) + \omega(s, t)$ for $t = 2, \ldots, T$ and $\xi(s, 1) \sim N(0, \sigma_\varepsilon^2/(1 - a^2))$. In this case $\omega(s, t)$ is a random field given by:

$$Cov(\omega(s_i, t)(s_j, t')) = \begin{Bmatrix} 0 & \text{if } t \neq t' \\ \sigma_\omega^2 C(h) & \text{if } t = t' \end{Bmatrix} \tag{3}$$

where $C(h)$ is given by the Matérn covariance function:

$$C(h) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(kh)^\nu K_\nu(kh) \tag{4}$$

with K_ν denoting a modified Bessel function of the second type. The marginal variance of this process is given by:

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}$$

Table 2
Descriptive statistics.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max	Sd.	Proportion < 0.7
2007	0.3807	0.5554	0.6113	0.6109	0.6753	0.9757	0.0879	0.8273
2008	0.4046	0.5415	0.6221	0.6138	0.6766	0.9267	0.0775	0.8645
2009	0.3943	0.5598	0.6363	0.6248	0.6828	1.0000	0.0789	0.8339
2010	0.4289	0.6223	0.6866	0.6794	0.7326	0.9576	0.0809	0.5734
2011	0.5477	0.7100	0.7499	0.7596	0.8067	1.0530	0.0652	0.1894
2012	0.5415	0.6918	0.7410	0.7539	0.8133	1.0000	0.0749	0.3034
2013	0.5245	0.6725	0.7164	0.7372	0.8044	1.0110	0.0776	0.4219
2014	0.5450	0.6922	0.7307	0.7506	0.8078	1.0310	0.0695	0.3266

Table 3
Estimated space-temporal model.

	Mean	sd	0.025q	0.5q	0.975q	Mode
D2007	0.6722	0.0015	0.6692	0.6722	0.6752	0.6722
D2008	0.6766	0.0015	0.6737	0.6766	0.6794	0.6766
D2009	0.6750	0.0014	0.6721	0.6750	0.6778	0.6750
D2010	0.7227	0.0014	0.7199	0.7227	0.7255	0.7227
D2011	0.7818	0.0014	0.7790	0.7818	0.7847	0.7818
D2012	0.7884	0.0014	0.7856	0.7884	0.7912	0.7884
D2013	0.7798	0.0014	0.7770	0.7798	0.7826	0.7798
D2014	0.7821	0.0019	0.7784	0.7821	0.7858	0.7821
D2 quarter	0.0049	0.0010	0.0030	0.0049	0.0068	0.0049
D3 quarter	−0.0130	0.0011	−0.0151	−0.0130	−0.0108	−0.0130
D4 quarter	−0.0030	0.0010	−0.0050	−0.0030	−0.0010	−0.0030
Precision	1072.0051	2.7476	1066.6299	1072.0015	1077.4002	1071.9945
log - τ	2.7794	0.0237	2.7330	2.7794	2.8258	2.7794
log - κ	0.8558	0.0289	−0.9125	−0.8558	−0.7992	−0.8558
AR(1)	0.9305	0.0037	0.9230	0.9306	0.9374	0.9307
Marginal Lik.	654214.47	n obs	326916			

The posterior distribution of this process is given by:

$$\pi(\theta, \xi|y) \propto \pi(y|\xi, \theta)\pi(\theta)$$

Assuming independent priors for $\pi(\theta)$, and remembering that the y_i elements are conditionally independent due to the Markov property, this posterior is given by:

$$\pi(\theta, \xi|y) \propto \left(\prod_{t=1}^{T} \pi(y|\xi, \theta)\right)\left(\pi(\xi_1|\theta) \prod_{t=2}^{T} \pi(\xi_t|\xi_{t-1}, \theta)\right)\pi(\theta)$$

To perform the Bayesian estimation procedure, we use the method proposed in [4], which is based on the use of a sequence of Laplace analytical approximations to accomplish the estimation procedure without the need of using simulation-based methods such as Markov chain Monte Carlo. This procedure allows for efficient and accurate computational approaches to estimate the posterior distributions of interest.

We assume a set of independent priors in the Bayesian estimation procedure, assuming log-gamma distributions for the parameters of the spatial covariance function, Gaussian distributions for the parameters of the conditional mean and autoregressive parameters and gamma distributions for the precision parameters, the usual choices in estimation of Bayesian models. In general, the estimated parameters are robust to the hyperparameters used in priors, due to the high number of observations in the sample.

The analyses are performed using the R software (www.r-project.org using the r-inla package (www.r-inla.org for the estimation and continuous representation procedures, and we also use the raster, rgdal and rgeos packages in R to perform geographical data processing. A complete description of the inference procedures and computational aspects of the method proposed by [3] can be found in [39,40].

4. Results

As discussed in Section 2, we estimated the model using a quarterly temporal aggregation of the average price ratio between ethanol and gasoline for each fuel station sampled in this period, representing 326,916 observations between 2007 and the first two quarters of 2014. As the data are quarterly, this model has a total of 30 periods. The first step in the analysis is constructing a triangulation of the territorial space containing the spatial locations of the sampled points (Fig. 1). To construct the triangulated mesh we use Delauney triangulation with 522 triangles.[3] This triangulation was chosen to be a balanced representation of the spatial distribution space of the service stations, and also due to computational limitations in the representation of the space-time model.

We use a representation of a geostatistical process [2], which decomposes the observed variable into a conditional mean and two error processes, one being the space-time random effects and the other the process of idiosyncratic innovations, with the representation given by Eq. (2).

Since the database provided by the ANP does not have a set of covariates for each location and time (s, l), we performed a decomposition of fixed effects for years and quarters to capture the conditional effects of possible covariates. In this case, we have fixed effects for each year in the sample, and fixed effects for quarters 2, 3 and 4 to capture seasonality aspects, wherein the first quarter is used as reference. In this way, we can obtain the average value of the price ratios for each period in the sample as the sum of fixed effects of year and the seasonal pattern given by the quarterly fixed effects. In this specification, we can interpret the spatial random effects as the deviation from the average value of the ethanol/gasoline price ratio in each period.

The results of the statistics on the distribution of the posterior estimated parameters are shown in Table 3. The values of fixed effects for years 2007–2014 (D2007–D2014) show values consistent with the mean values found in Table 2. We can also note that the quarterly dummies also capture the effects of the sugarcane harvest, which is concentrated from June to November, and has a negative sign in these periods. We can also observe high accuracy of the precision of idiosyncratic effects, showing the importance of the spatial random effects in this problem. The parameters of spatial covariance are presented in logarithmic formulation, and to have an interpretable

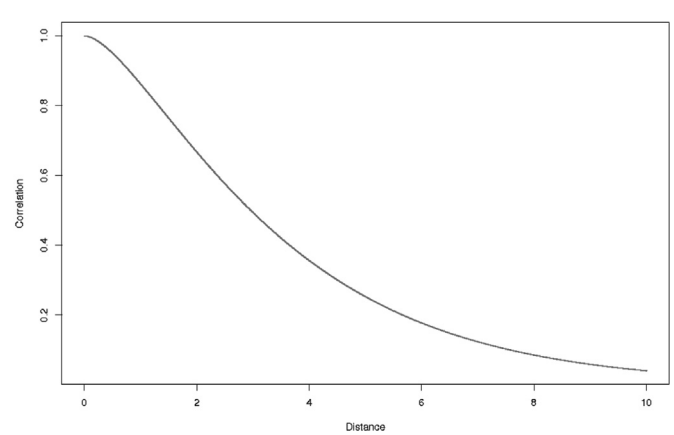

Fig. 5. Fitted spatial correlation.

[3] All details about the estimation procedure and full results for each year in the sample can be obtained with the author, and are not shown for reasons of space.

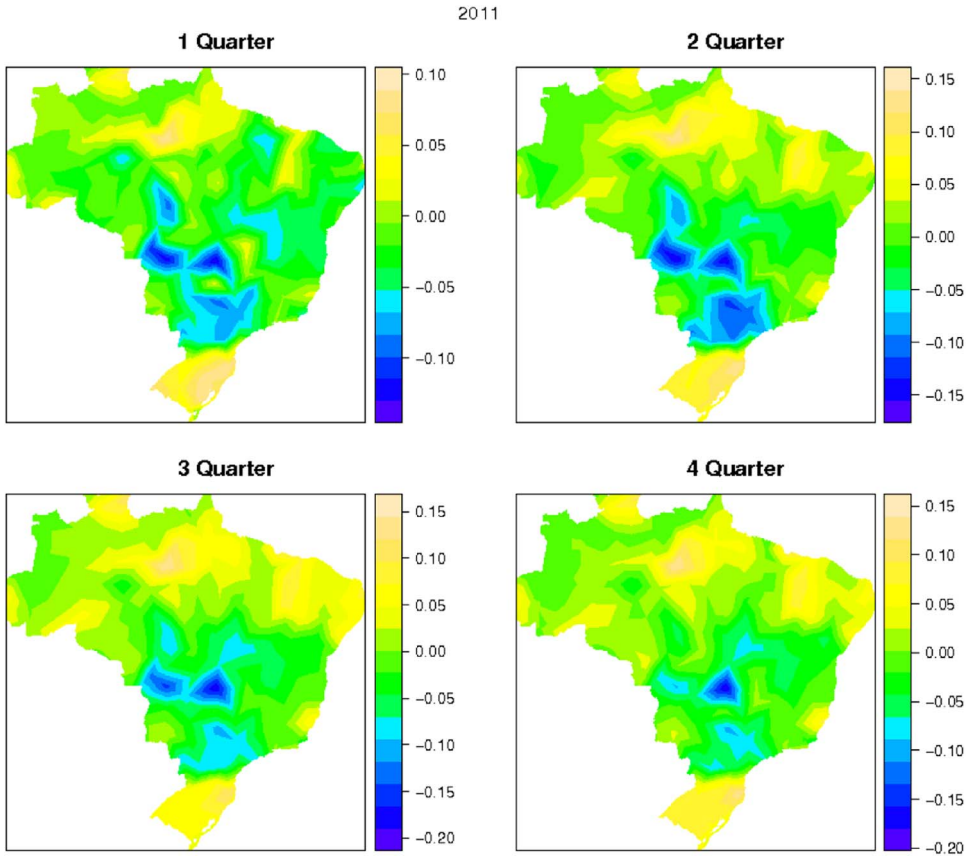

Fig. 6. Spatial random effects – 2011.

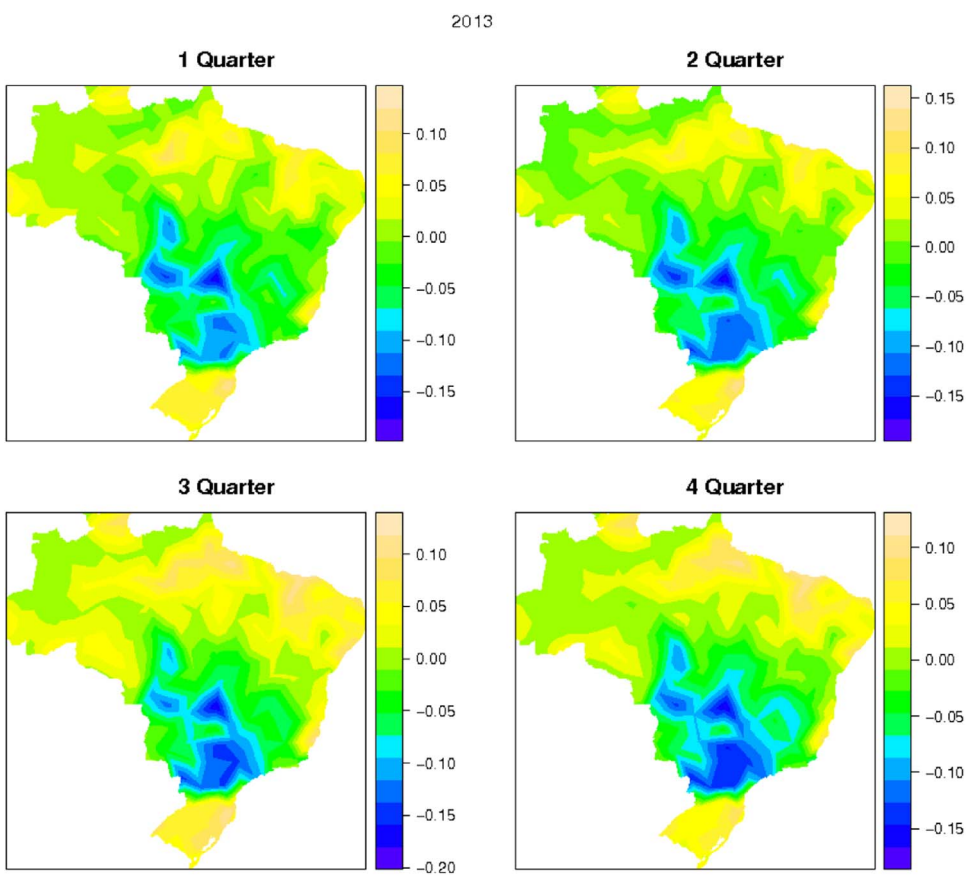

Fig. 7. Spatial random effects – 2013.

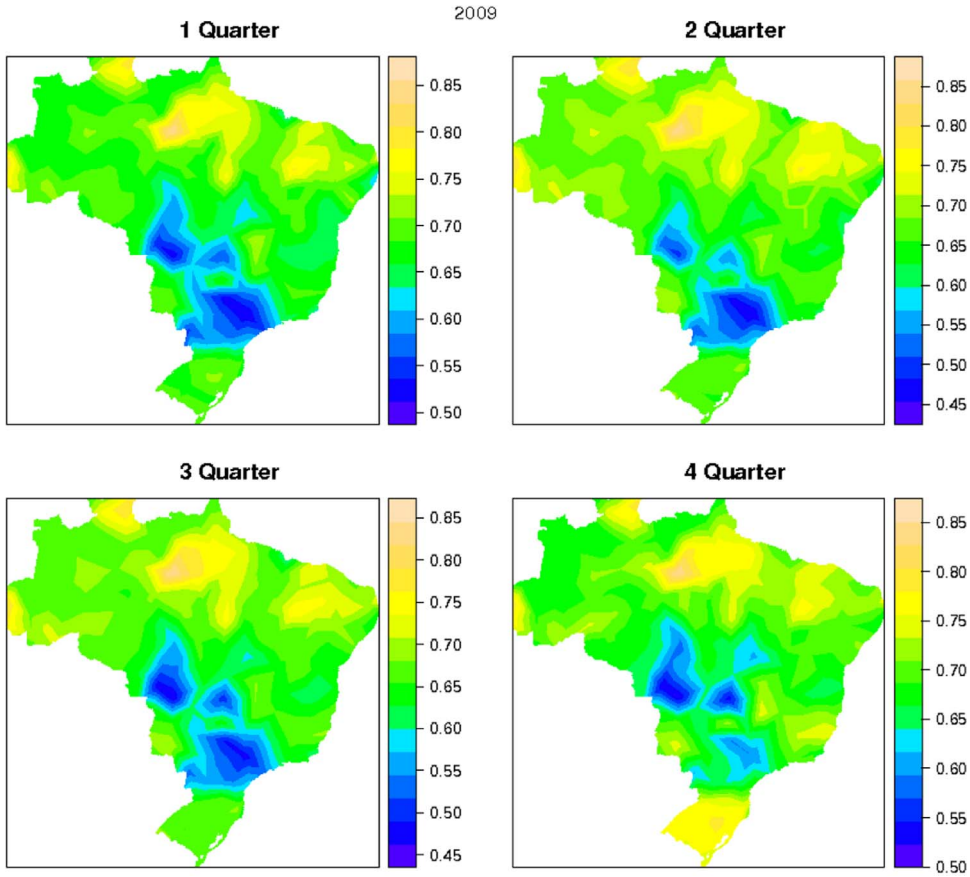

Fig. 8. Fitted price ratio – 2009.

scale we present in Fig. 5 the pattern of spatial correlation equivalent to these parameters in an observation scale measured in degrees. There is high spatial correlation, consistent with the observed results of the spatial distribution of prices.

The autoregressive component that captures the temporal persistence of spatial random effects has posterior mean of 0.9305, with a 95% credible interval between the values of 0.9230 and 0.9307. This result also shows the high persistence of spatial random effects, reflecting the enduring patterns in production and pricing of ethanol in the country. To show the evolution of spatial random effects, we present in Figs. 6 and 7 the spatial random effects for the Brazilian space continuum for the years 2011 and 2013.[4]

The results observed in these figures confirm the spatial patterns of price ratios. We can see that in general prices in the North and Northeast regions are systematically higher (positive spatial effects of about 0.10–0.20) than the average price of each quarter, and in the opposite way prices in the Southeast and in part the Center-South regions are consistently lower than the average in each year, with spatial random effects of about −0.10 to −0.15 below average. We can also observe the general spatial effects in each year. In 2011, marking the beginning of the crisis period in ethanol usage, except for some parts of the South there are few regions with prices well below the average for the year.

To show the continuous prediction for the ratio of prices across the Brazilian territorial space, we present in Figs. 8–10 the price ratio predicted by the model for the years 2009, 2011 and 2014, as the sum of fixed effects and spatial random effects for each location in the continuum. The figures show that the model is able to capture adequately the spatial patterns of prices observed in this period in Brazil, with higher prices in the Northeast and North and lower prices in the Southeast and Center-South. The results also clearly indicate the changes in the ethanol/gasoline price ratio in time. In 2009, the price ratio is close to or less the 0.7, the equilibrium value in the fueling of ethanol, for most of the country, while from 2011 most of the country presents ratios higher than 0.7, favoring the use of gasoline.

We calculate an auxiliary measure to help visualization of the economic advantage of ethanol fueling. Using the means and standard deviations from the estimated posterior distribution of price ratios for each spatial point in the continuum, and assuming that the ratio prices can be approximated by a Gaussian distribution with these posterior parameters, we calculate the probability of predicted values being below the equilibrium value of 0.7. As discussed in the previous sections, on average it is advantageous to supply a flex-fuel vehicle with ethanol if the ethanol/gasoline price ratio is below a threshold of 0.7, and thus we estimate the probability of this condition for locations not sampled in each period. Note that the choice of the 0.7 threshold is based on an average estimate of the energy efficiency of flex-fuel engines, and the exact value depends on the specifications of each vehicle as well as consumer preferences, as discussed, for example, in [21]. Values between 0.7 and 0.8 also could be considered depending on the consumer's preferences and engines efficiency. Analogous results can be obtained with other threshold values. The results of this measure, which we call the Ethanol Advantage Probability, are shown in Figs. 11–13 for the years 2009, 2011 and 2014.

These figures reflect the great change that occurred in the ethanol/gasoline price ratio during this period. The probability of economic advantage with ethanol fueling is near 1 in most of the country in the years until 2009, with low probabilities only in parts of the North and Northeast regions. However, from the last quarter of 2009, this pattern began to change, so that the areas with high probability of advantageous fueling with ethanol are reduced, an increasing pattern in the

[4] The results for the other years can be obtained with the author.

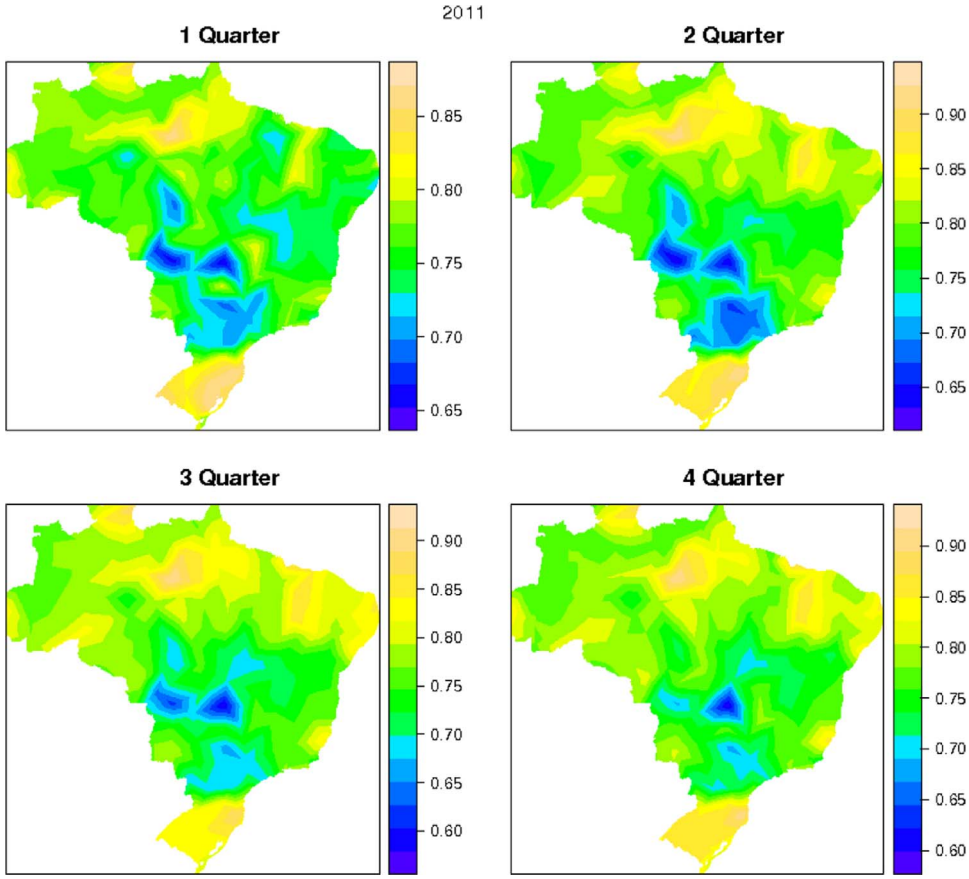

Fig. 9. Fitted price ratio – 2011.

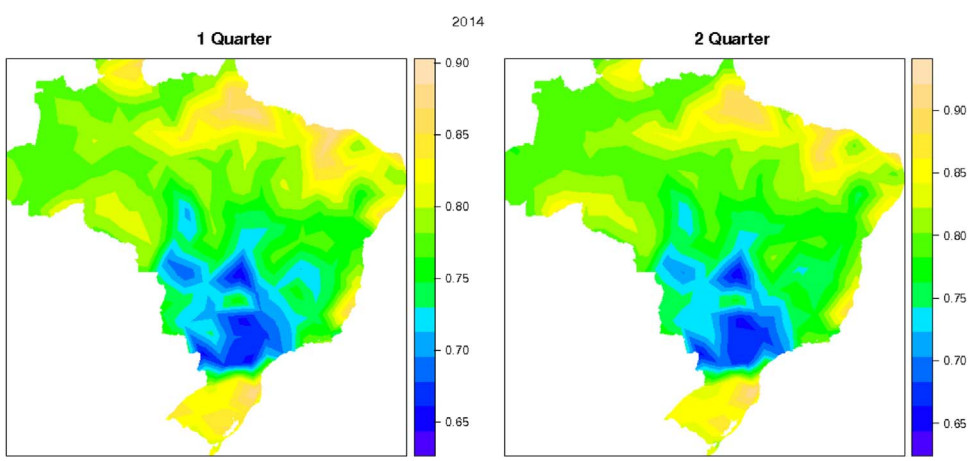

Fig. 10. Fitted price ratio – 2014.

ensuing years.

In 2011 (Fig. 12) we can clearly notice the ethanol crisis, since pure ethanol fueling is advantageous only for parts of the Southeast and Center regions, where are located near or in main producing centers of sugarcane and ethanol, and thus associated with lower distribution costs. The year 2014 (Fig. 13) shows some improvement in the advantage of ethanol fueling, but the general pattern remains close to that observed in the previous years.

This movement in relative prices of ethanol can be explained in part by changes in the international price of sugar. In Fig. 14, we present a graph with sugar prices in dollars for same quarters analyzed. We note that in 2010 prices were relatively low, favoring the production of ethanol rather than production and export of sugar. The sugar price spikes are centered in 2011 and 2012, which are the years with the lowest advantage of ethanol. This effect in relative prices also can be explained by the reduction in the CIDE levy on gasoline and the political restriction on price increases of gasoline starting in 2011.

To quantify the model fit, we calculate some measures comparing the set of model fitted values for the prices effectively observed in the ANP price sample. We show in Table 4 a comparison between the average prices observed and those predicted by the model for each year. This table reports the mean, standard deviation and the ratio of prices less than 0.7 found in the sample (columns 2–4), while columns (5–7) show the equivalent values fitted by the model. It can be seen that the values predicted by the model are very similar to the values observed in the sample in terms of means and standard deviations. The proportions of values below 0.7 has a small positive bias of about 0.03 in the 2007–2009 interval, and a negative bias near 0.023 for the years 2010–2012.

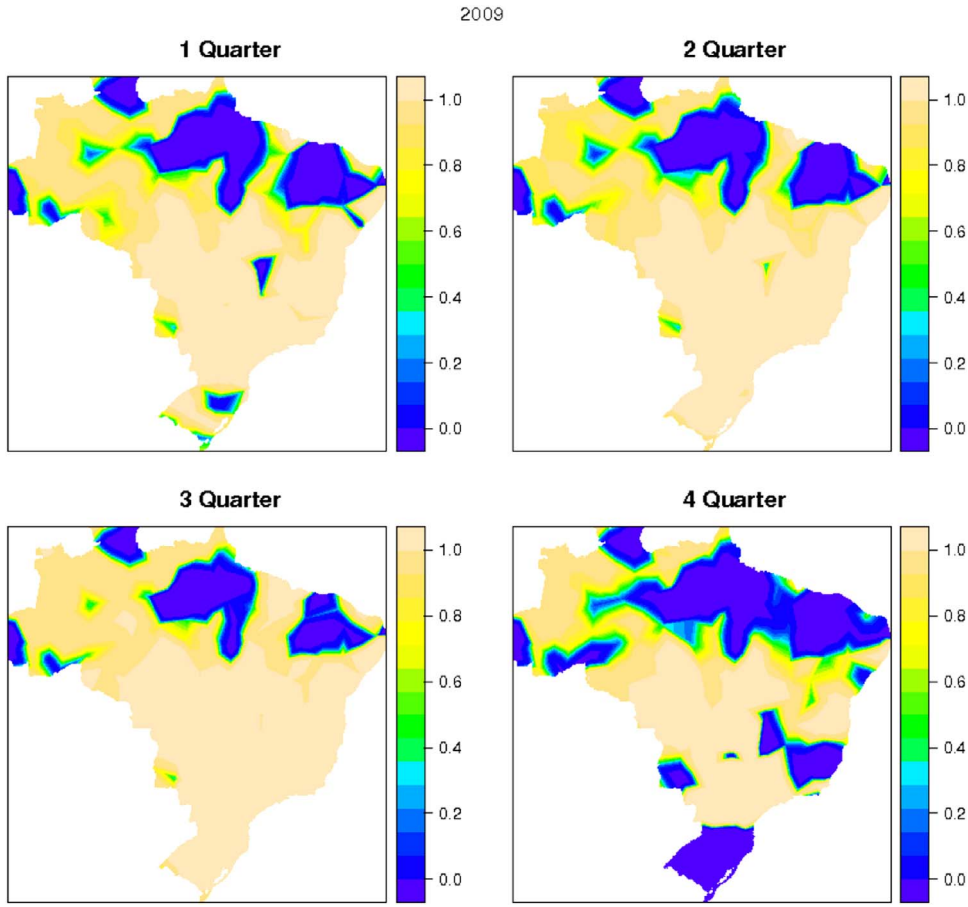

Fig. 11. Ethanol advantage probability – 2009.

We also report some more detailed adjustment measures in Table 5, with the mean error (ME), root mean squared error (RMSE), mean absolute error (MAE) and mean percentage error (MPE) for each year of the sample, comparing the data used in the estimation and the results obtained by the model. The results indicate an appropriate adjustment of the model, noting that we use in our specification only dummies for fixed effects and the spatial random effects.

In the same Table 5 we also calculate one additional model validation measure. One of the important results of the method proposed here is to permit the estimation of values for locations not observed in the sample, obtaining continuous measures for the ethanol/gasoline price ratios for the entire spatial continuum analyzed. As these values are not observed, it is not possible to make a direct measure of goodness of fit for these points.

For an indication of model fit for the data not observed in the sample, or the spatial interpolation procedure as it is known in the literature on geospatial modeling (e.g., [2,41]), we performed a cross validation procedure. For this we randomly excluded 10% of the original sample, i.e., 32,691 observations, and estimated the model again without the excluded sample and calculated the same adjustment measures, now comparing the fitted values with the data excluded from the estimation. The results are placed in the last row of Table 5 (10% omitted). These results have a similar pattern to that observed for the full sample. This similarity is an indication the good quality of fit of the spatial interpolation procedure.

A final comment on the space-temporal model presented is its relationship with another formulation widely used in the analysis of price relationships between ethanol and gasoline, based on co-integration models. In [26], an error correction model based on co-integration relationships is used to analyze the short and long-run elasticities between ethanol and gasoline. Co-integration models are also used in demand analysis in this same market (e.g., [6,23,19,28]). Although we are using an alternative specification, we note that our results can be interpreted in a manner similar to a co-integrating equation, but assuming a known co-integration vector. In this case, we assume that ethanol and gasoline prices are not stationary, but the price ratio generates a spatial stationary process.

This formulation can also be interpreted as a test of no-arbitrage relationship between the relative prices of ethanol and gasoline, related to the point of optimal price ratio between ethanol and gasoline prices that leads to end consumer indifference point. The violation of the no-arbitrage condition would be confirmed if the deviations from the equilibrium relationship are permanent, rejecting the hypothesis of stationarity of deviations and indicating the presence of some factor that precludes the reversion of the price ratio to the consumer's indifference point.[5]

This interpretation can be verified by the results in Table 3 for the autoregressive parameter estimated in the model. The spatial autoregressive component is stationary, since the posterior distribution for the parameter is below the unit value. In this way we can think of this specification as a spatial co-integration model where the spatial random effects are the residuals of the estimation of co-integration relationship, in this case formulated as a ratio that generates a stationary residual.

A more traditional formulation would be to analyze the relationship between the log of the prices, and estimate the co-integration vector. For this, we modified our specification using as dependent variable the log of ethanol prices for each location, and used as the explanatory

[5] In a related work, [25] analyze the presence of bubbles processes in the ethanol/gasoline price ratio.

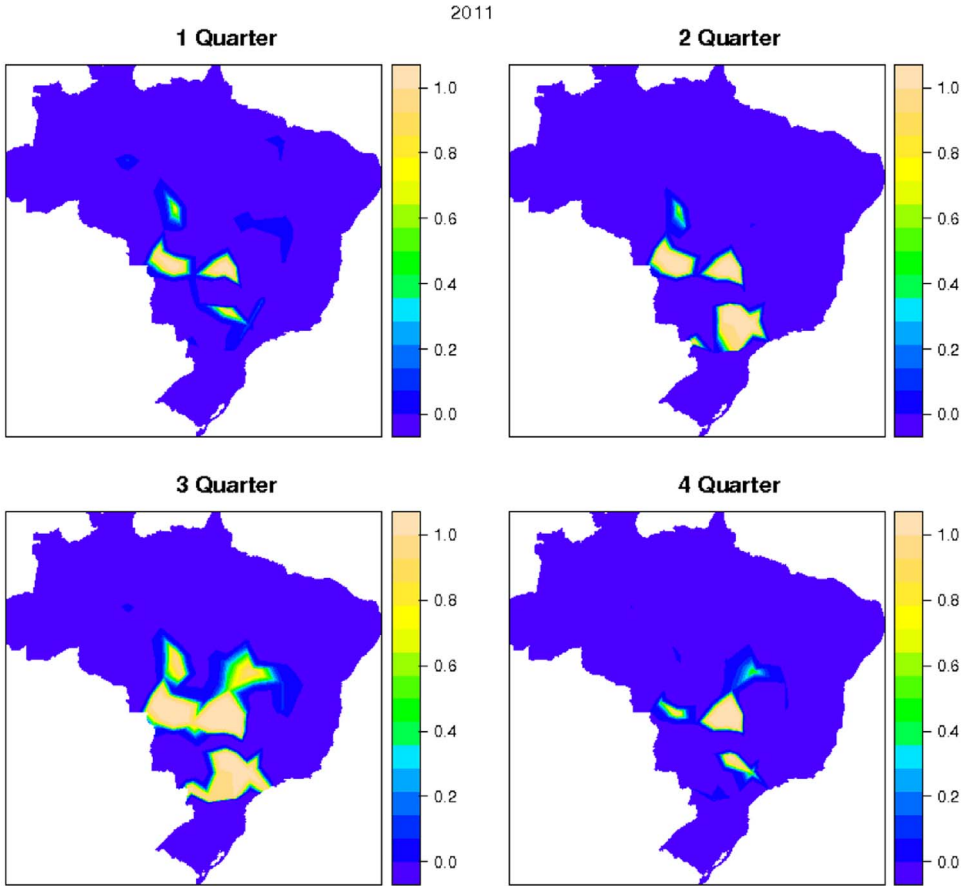

Fig. 12. Ethanol advantage probability – 2011.

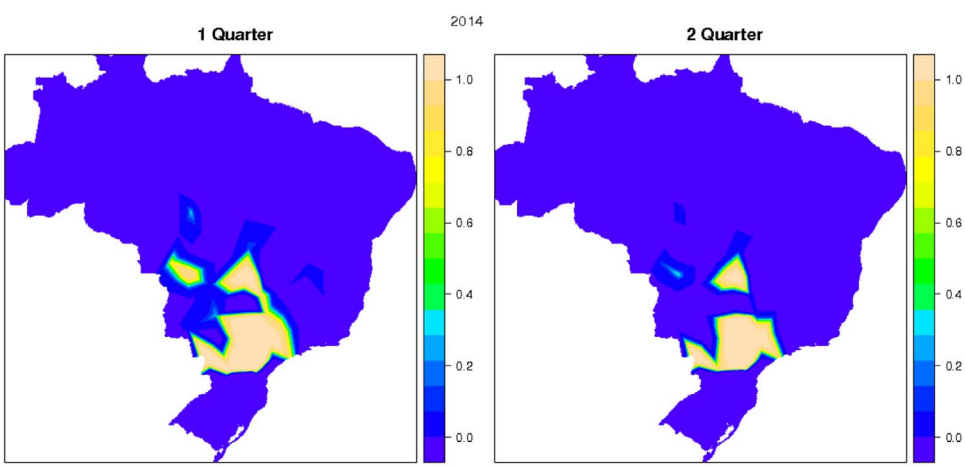

Fig. 13. Ethanol advantage probability – 2014.

variable the log of gasoline prices, the structure of spatial random effects and the idiosyncratic error components. In this way we have a state-space representation for a spatial co-integration process in the form:

$$\log y_t = \log z_t\beta + \xi_t + \varepsilon_t \xi_t = a\xi_{t-1} + \omega_t$$

with log $\log y_t$ denoting the price of ethanol and $\log z_t$ the gasoline price for all locations (stations) in the same period and with β denoting the long-run equilibrium parameter, ξ_t the spatial random effects and ε_t the idiosyncratic components. We estimated this model with and without the inclusion of fixed effects of year and quarter. The results are shown in Tables 6 and 7. These results indicate the presence of a stationary component for the spatial autoregressive process in these two procedures, and so it is consistent with the interpretation of a spatial co-integration model. Note that we can think of this specification as a generalized version of a panel co-integration model similar to that used in [23], but where we model each individual as a spatial location.

5. Discussion and conclusions

In this work we introduce an innovative method to model the spatial and temporal evolution of the ratio of ethanol/gasoline prices in Brazil. TIhis method is based on specification of a spatio-temporal model which enables representing the dynamics of spatial dispersion in prices. This formulation is based on a computationally efficient

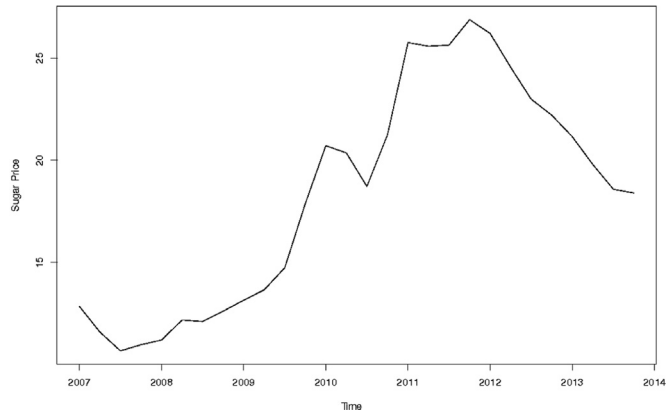

Fig. 14. International sugar prices.

prices. We make an original methodological contribution by combining the method described in [3] with co-integration models, which are important in fuel price analysis.

This method allows not only analyzing the evolution of observed prices, but also estimating/interpolating the ratio of prices and the spatial random effects for all continuous space, thus permitting analysis of the pricing process throughout Brazilian territory, and thus broadening the scope of the ANP's price survey. The results show that this model can capture adequately the relative evolution of ethanol and gasoline prices, and allows analyzing in each period the advantage of fueling with ethanol. The model captures the change in temporal price patterns, and shows results consistent with the ethanol industry's crisis in Brazil (e.g., [25,32]), showing that the supply of ethanol is only advantageous to consumers in regions near production centers in the years after 2009.

Table 4
Comparison – observed and predicted.

	Observed mean	Observed SD.	Observed < 0.7	Fitted mean	Fitted SD.	Fitted < 0.7
2007	0.6109	0.0879	0.8273	0.6109	0.0800	0.8574
2008	0.6138	0.0775	0.8645	0.6138	0.0706	0.8950
2009	0.6248	0.0789	0.8339	0.6248	0.0723	0.8672
2010	0.6794	0.0809	0.5734	0.6794	0.0733	0.5661
2011	0.7596	0.0652	0.1894	0.7595	0.0560	0.1399
2012	0.7539	0.07495	0.3034	0.7538	0.0673	0.2889
2013	0.7371	0.0776	0.4219	0.7371	0.0708	0.4369
2014	0.7506	0.0695	0.3266	0.7506	0.0633	0.3878

Table 5
Fit measures.

	EM	REQM	MAE	MPE
2007	−1.3718e−05	0.0328	0.0240	0.0265
2008	−1.4546e−05	0.0289	0.0214	0.0185
2009	−1.4111e−05	0.0289	0.0213	0.0178
2010	1.4737e−05	0.0313	0.0233	0.0177
2011	−1.6609e−05	0.0311	0.0223	0.0091
2012	−1.7153e−05	0.0303	0.0224	0.0112
2013	−1.7666e−05	0.0290	0.0210	0.0114
2014	−3.7482e−05	0.0260	0.01918	0.0059
10% Omitted	0.0001	0.0311	0.0227	0.0409

The method presented here has several applications in the design of energy policies. The spatial indicator of ethanol advantage is a visual measure of the fuel market conditions. In the presence of a large fleet of flex-fuel engines, this measure facilitates consumer decisions on the optimum choice of fuel in each location and period, increasing economic efficiency in this market. This measure is also an input for an optimal taxation policy in this market. If the goal is to encourage the use of renewable fuels, governments can implement distinct tax rates on ethanol and gasoline in each region, using the already existing CIDE levy.

Our model is based on the use of the spatial distribution of the prices as a proxy for the different factors that affect fuel prices, such as production and transport costs and conditions of demand and compe-

Table 6
Spatial co-integration.

	Mean	sd	0.025q	0.5q	0.975q	Mode
log (Gas)	0.7124	5e−04	0.7113	0.7124	0.7134	0.7124
Precision	402.3628	1.0395	400.3291	402.3614	404.4040	402.3588
log - τ	2.6550	0.0192	2.6173	2.6550	2.6926	2.6550
log - κ	−1.0335	0.0241	−1.0808	−1.0335	−0.9863	−1.0335
AR (1)	0.9231	0.0032	0.9168	0.9232	0.9291	0.9233
Marginal Lik.	478436.08	n obs	326916			

representation of spatial models proposed by [3], based on equivalence between the solution of a stochastic partial differential equation and a spatial covariance function. This representation allows a hierarchical representation of spatial and spatio-temporal processes and the use of Bayesian methods for inference of parameters and latent variables.

This method enables avoiding the aggregation of observed prices required for the use of spatial analysis methods based on areal/lattice models, and thus allows modeling the full observed heterogeneity in prices observed at the level of fuel stations. This method also allows incorporating the dynamic evolution of the price process, particularly the long-run equilibrium relationship between ethanol and gasoline tition in each market. As these factors have complex measurement, the proposed model provides a simple way to analyze the aggregate effects of these variables. Our specification was based on the use of fixed effects for year and quarter and spatial random effects. This model can be refined by adding other variables measured at each location, like the proportion of the flex-fuel vehicles, freight costs, income measures, or supply side variables such as the number of distributors.

Another possible application of this method is the analysis of fuel price competition. Since the model allows estimating the dispersion of prices in any region of interest, it can be used to estimate the price dispersion within the same region. Assuming that transport and

Table 7
Spatial co-integration – fixed effects for year and quarter.

	Mean	sd	0.025q	0.5q	0.975q	Mode
log (Gas)	1.2178	0.0022	1.2135	1.2178	1.2222	1.2178
D2007	−0.6122	0.0027	−0.6175	−0.6122	−0.6069	−0.6122
D2008	−0.6049	0.0027	−0.6102	−0.6049	−0.5997	−0.6049
D2009	−0.6045	0.0027	−0.6097	−0.6045	−0.5992	−0.6045
D2010	−0.5396	0.0027	−0.5449	−0.5396	−0.5343	−0.5396
D2011	−0.4654	0.0028	−0.4709	−0.4654	−0.4600	−0.4654
D2012	−0.4604	0.0028	−0.4659 −	0.4604	−0.4550	−0.4604
D2013	−0.4819	0.0029	−0.4875	−0.4819	−0.4762	−0.4819
D2014	−0.4858	0.0032	−0.4922	−0.4858	−0.4795	−0.4858
D2 quarter	−0.0038	0.0011	−0.0060	−0.0038	−0.0015	−0.0038
D3 quarter	−0.0307	0.0013	−0.0332	−0.0307	−0.0283	−0.0307
D4 quarter	−0.0138	0.0012	−0.0162	−0.0138	−0.0115	−0.0138
Precision	72.3251	1.2105	469.9570	472.3235	474.7019	472.3204
log - τ	2.6457	0.0207	2.6051	2.6457	2.6863	2.6457
log - κ	−0.8424	0.0250 −	0.8913	−0.8424	−0.7935	−0.8424
AR (1)	0.9111	0.0037	0.9036	0.9112	0.9182	0.9114
Marginal Lik.	516264.47	n obs	326916			

production costs are constant in a single region, the spatial random effects can be used to identify sub-regions with anomalous price patterns, which can be associated with pricing cartels.

References

[1] Besag J. Spatial interaction and the statistical analysis of lattice systems [with discussion]. J R Stat Soc Ser B 1974;36(2):192–225.
[2] Bivand RS, Pebesma E, Gomes-Rubio V. Applied spatial data analysis with R. Heidelberg, Germany: Springer; 2013.
[3] Lindgren F, Rue H, Lindstrom J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc Ser B 2011;73(4):423–98.
[4] Rue H, Martino S, Chopin N. Approximated Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations [with discussion]. J R Stat Soc Ser B 2009;71:319–92.
[5] BNDES. Sugarcane-based bioethanol: energy for sustainable development. first ed.. BNDES; 2008.
[6] de Freitas LC, Kaneko S. Ethanol demand under the flex-fuel technology regime in Brazil. Energy Econ 2011;33:1146–54.
[7] Goldemberg J, Macedo IC. Brazilian alcohol program: an overview. Energy Sustain Dev 1994;1(1):17–22.
[8] Moreira JR, Goldemberg J. The alcohol program. Energy Policy 1999;27(4):229–45.
[9] Rico JAP, Mercedes SS, Sauer IL. Genesis and consolidation of the brazilian bioethanol: a review of policies and incentive mechanisms. Renew Sustain Energy Rev 2010;14(7):1874–87.
[10] Cerqueira Leite RC, Leal MRLV, Cortez LAB, Griffin WM, Scandiffio MIG. Can Brazil replace 5% of the 2025 gasoline world demand with ethanol?. Energy 2009;34(5):655–61.
[11] Szulczyk KR, McCarl BA, Cornforth G. Market penetration of ethanol. Renew Sustain Energy Rev 2010;14(1):394–403.
[12] Niven RK. Ethanol in gasoline: environmental impacts and sustainability review article. Renew Sustain Energy Rev 2005;9(6):535–55.
[13] Su Y, Zhang P, Su Y. An overview of biofuels policies and industrialization in the major biofuel producing countries. Renew Sustain Energy Rev 2015;50:991–1003.
[14] Partha Dasgupta GH. The optimal depletion of exhaustible resources. Rev Econ Stud 1974;41:3–28.
[15] Heal G. Reflections – the economics of renewable energy in the United States. Rev Environ Econ Policy 2010;4(1):139–54.
[16] Grafton RQ, Kompas T, Long NV. Substitution between biofuels and fossil fuels: is there a green paradox?. J. Environ Econ Manag. vol. 64(3). 2012.p. 328–41. [2010 Monte Verita Conference on Sustainable Resource Use and Economic Dynamics(SURED)].
[17] Nepal R. Roles and potentials of renewable energy in less-developed economies: the case of Nepal. Renew Sustain Energy Rev 2012;16(4):2200–6.
[18] Strand J. Optimal fossil-fuel taxation with backstop technologies and tenure risk. Energy Econ 2010;32(2):418–22.
[19] Gomez MA, J, Legey LFL. An analysis of the impact of flex-fuel vehicles on fuel consumption in Brazil, applying cointegration and the Kalman filter. Energy. vol. 81(0). 2015. p. 696–705.
[20] ANFAVEA. Brazilian automotive industry yearbook. ANFAVEA; 2013.
[21] Salvo A, Huse C. Build it, but will they come? Evidence from consumer choice between gasoline and sugarcane ethanol. J Environ Econ Manag 2013;66:251–79.
[22] Szklo A, Schaeffer R, Delgado F. Can one say ethanol is a real threat to gasoline?. Energy Policy 2007;35(11):5411–21.
[23] de Freitas LC, Kaneko S. Ethanol demand in Brazil: regional approach. Energy Policy 2011;39:2289–98.
[24] Du X, Carriquiry MA. Flex-fuel vehicle adoption and dynamics of ethanol prices: lessons from Brazil. Energy Policy 2013;59:507–12.
[25] Montasser GE, Gupta R, Martins AL, Wanke P. Are there multiple bubbles in the ethanol-gasoline price ratio of brazil?. Renew Sustain Energy Rev 2015;52:19–23.
[26] Alves D, Bueno R. Short-run, long-run and cross elasticities of gasoline demand in Brazil. Energy Econ 2003;25:191–9.
[27] Azevedo B. Evaluation on the price and income elasticity for liquid fuels in Brazil and its geographical regions [master thesis]. IBMEC-RJ. 2007.
[28] Zafeiriou E, Arabatzis G, Tampakis S, Soutsas K. The impact of energy prices on the volatility of ethanol prices and the role of gasoline emissions. Renew Sustain Energy Rev 2014;33:87–95.
[29] Bojesen M, Birkin M, Clarke G. Spatial competition for biogas production using insights from retail location models. Energy 2014;68(0):617–28.
[30] Rezende ML, Richardson JW. Economic feasibility of sugar and ethanol production in Brazil under alternative future prices outlook. Agric Syst 2015;138:77–87.
[31] Bargos FF, de Queiroz Lamas W, Bargos DC, Neto MB, Pardal PCPM. Location problem method applied to sugar and ethanol mills location optimization. Renew Sustain Energy Rev 2016;65:274–82.
[32] Farina E. ⟨http://www1.folha.uol.com.br/poder/2014/04/1440253-politica-de-dilma-esta-quebrando-o-etanol-diz-presidente-de-entidade.shtml⟩. 2014.
[33] Whittle P. On stationary processes on the plane. Biometrika 1954;41(7):434–49.
[34] Rozanov JA. Markov random fields and stochastic partial differential equations. Math URSS Sb 1977;32:515–34.
[35] Rue H, Held L. Gaussian Markov random fields: theory and applications. Boca Raton, United States: Chapman & Hall – CRC; 2005.
[36] Arbia G. Spatial econometrics – statistical foundations and applications to regional convergence. Heidelberg, Germany: Springer; 2006.
[37] Brenner SC, Scott R. The mathematical theory of finite element methods. Heidelberg, Germany: Springer; 2007.
[38] Cameletti M, Lindgren F, Simpson D, Rue H. Spatio-temporal modeling of particulate matter concentration through the spde approach. AStA Adv Stat Anal 2013;97:109–31.
[39] Blangiardo M, Cameletti M. Spatial and spatio-temporal models with R-INLA. New Jersey, United States: Wiley; 2015.
[40] Krainski ET, Lindgren F, Simpson D, Rue H. The r-inla tutorial on spde models, disponible at ⟨www.r-inla.org⟩. 2016.
[41] Cressie N, Wikle CK. Statistics for spatio-temporal data. New Jersey, United States: Wiley; 2011.