ISSN: 2047 - 0398
SPECIAL ISSUE – V|05|N|01| June | 2015

EVALUATING MASS OF REAL ESTATE USING GEOSTATISTICAL AND BAYESIAN KRIGING FOR THE MUNICIPAL ADMINISTRATION CAN HAVE A FAIR AND EQUITABLE COLLECTION OF LAND TAX

Ricardo André Hornburg
E-mail: *ricardo_andreh@yahoo.com.br*

Norberto Hochheim
E-mail: *hochheim@ecv.ufsc.br*

ABSTRACT

One of the great difficulties that have in evaluating mass of real estate is to find a model that shows the reality of the real estate market so you can build a Plant of Generic Values. This paper presents a method for assessing mass properties that will be applied to buildings in the center district of the city of Balneario Camboriú in the State of Santa Catarina - Brazil. The objective is to find a geostatistical model that is able to estimate the value of real estate. The Bayesian geostatistical methods are used in this work in a combined way to estimate the value of the location of the property. Bayesian Kriging technique that allows estimating variable values spatially distributed from adjacent values considered as interdependent is used. Thus, the kriging is considered a method of moving averages. The semivariogram is the basic tool holder the kriging techniques, allowing quantitatively represent the variation of a regionalized phenomenon in space.

Keywords: *Evaluation of Mass. Geostatistics. Bayesian kriging.*

1. INTRODUÇÃO

A partir da aprovação da lei de responsabilidade fiscal dos Estados e municípios brasileiros impondo o controle dos gastos, condicionado à capacidade de arrecadação de tributos desses entes políticos, gerou a necessidade um maior planejamento, em especial, no que se refere aos municípios, o imposto sobre a propriedade urbana, também conhecido como Imposto Predial e Territorial Urbano (IPTU), o qual é calculado como um percentual do valor venal dos imóveis.

Na avaliação dos imóveis encontram-se dificuldades para se obter variáveis explicativas que possam determinar com precisão o valor dos imóveis. Muitos municípios no Brasil utilizam métodos estatísticos descritivos, usando para o cálculo do valor dos imóveis fatores de homogeneização e métodos empíricos predeterminados, deixando muitas das vezes de encontrar modelos estatísticos adequados que considerem as reais condições e fatores locais do mercado imobiliário.

As variáveis encontradas para compor o valor de um determinado subconjunto de imóveis, não são necessariamente as mesmas para um outro subconjunto de imóveis, para tanto, em alguns casos deve-se retirar variáveis e acrescentar outras, pois cada subconjunto de imóveis pode ter suas próprias características que não são necessariamente iguais aos outros subconjuntos.

Na engenharia de avaliação, os modelos estatísticos são de suma importância para a elaboração da Planta de Valores Genéricos (PVG), que permite com que as prefeituras possam determinar o Imposto Predial Territorial Urbano (IPTU) com máxima eqüidade.

Uma grande dificuldade que se tem na utilização de métodos econométricos, na busca de modelos de avaliação, está em considerar a variável localização que pode valorizar ou desvalorizar os imóveis.

A exclusão da localização da modelagem pode causar sérios problemas de predição, pois dados localizados espacialmente (que é o caso dos imóveis), em geral, apresentam autocorrelação ou covariância espacial. Preditores que não levam isso em consideração estão muito longe da realidade (CRESSIE, 1993).

As variáveis de localização que não são representadas e modeladas corretamente podem gerar problemas nos modelos de regressão. Além da perda do poder de explicação do modelo, pode provocar a autocorrelação espacial nos resíduos do modelo pelo tratamento incompleto ou inadequado dos fatores de localização.

Buscando soluções teóricas e de cunho metodológico para estes problemas, existem duas técnicas estatísticas diferentes para o tratamento e modelagem dos efeitos espaciais nos dados de mercado que são as técnicas de regressão espacial e as técnicas de geoestatística.

Para tanto, pretende-se, neste trabalho, apresentar um método adequado para avaliação em massa de imóveis, combinando o uso da econometria espacial e da geoestatística bayesiana.

2. AVALIAÇÃO EM MASSA DE IMÓVEIS

2.1 Avaliação Imobiliária

A função principal de uma avaliação é assegurar o valor de algum tipo de imóvel sob um determinado conjunto de condições. Os valores das propriedades variam consideravelmente de um local para outro (GONZÁLEZ, 2002).

Atribui-se valor a tudo que é útil ou escasso. Cabe à avaliação traduzir essa utilidade ou escassez e associar a necessidade e/ou desejo de possuir um bem numa quantia monetária (AYRES, 1996).

O valor de um imóvel depende diretamente das características do entorno, tais como: tipos de imóveis existentes, ruas, utilidades, conveniências. Além o entorno imediato, o imóvel relaciona-se com a cidade inteira. Todavia, nem o declínio econômico de uma cidade afeta todas as suas partes igualmente (CAN, 1998).

A avaliação de imóveis urbanos deve se basear na ABNT 14653-2 -2011. Para esta norma, a avaliação de imóveis deve preferencialmente se fundamentar na pesquisa de mercado. Devem ser colhidas amostras com preços de imóveis ofertados e comercializados, além de atributos que afetem o valor. Tais atributos serão ponderados ou por inferência estatística ou por homogeneização. A amostra deve ser representativa do mercado imobiliário em análise.

Estudos sobre Avaliação Imobiliária, como de Cesare (2004), apontam três principais metodologias para avaliação em massa para fins fiscais: Método de Custo de Reprodução, Método de Renda e Método Comparativo de Dados de Mercado.

Para Silva (2006), o objetivo da avaliação em massa de imóveis é obter o valor de todos os imóveis localizados em determinada área. Emprega métodos que devem ter respaldo legal. A metodologia empregada deve evitar o máximo o subjetivismo, tanto dos dados quanto dos procedimentos, além de adequar os mesmos ao mercado real (POLI et al., 2011).

Möller (1995) ressalta que a avaliação em massa de imóveis deve ser vinculada ao estudo do Código Tributário Municipal - CTM no que se refere aos impostos sobre a propriedade (IPTU e ITBI, impostos da esfera municipal). Dessa forma o resultado final da avaliação se adequará às exigências legais.

A avaliação em massa de imóveis geralmente é obtida pelos seguintes métodos de avaliação: método evolutivo e de comparação de dados de mercado. O primeiro requer as estimativas do valor do terreno e o corrente custo de reprodução das edificações. O segundo método faz a estimativa do valor de mercado com base em preços de grupos de imóveis vendidos em um período que antecede a data de lançamento dos tributos (SILVA, 2006).

Zancan (1996) alerta que a desvantagem da metodologia usualmente utilizada na avaliação em massa, que calcula separadamente os valores do terreno e das edificações, está em não existir um mercado de edificações separado dos terrenos sobre os quais foram construídas.

2.2 Ordinary Least Squares

Mínimos Quadrados Ordinários, ou OLS (Ordinary Least Squares) é uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados que tentam minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (tais diferenças são chamadas de resíduos). E isso pode ser dividido em regressão simples e múltipla:

- Regressão Simples – para estimar valores da variável dada é considerado os outros valores das variáveis x acreditanto terem poder explicativo de $\mathcal{Y}$ de acordo com a fórmula::

$$\mathcal{Y} = \alpha + \beta x + \varepsilon,$$

Where:
α – é o parâmetro do modelo chamado constante;
β – é o parâmetro do modelo chamado de coeficiente da variável;
ε – erro – é a variação que não é explicada pelo modelo.

- Regressão Múltipla – mostra uma operação semelhante com a regressão simples, no entanto, leva em consideração diversas variáveis explanatórias x que influenciam $\mathcal{Y}$ ao mesmo tempo:

$$\mathcal{Y} = \beta_0 + x_1 \beta_1 + x_2 \beta_2 + x_3 \beta_3 + \ldots + x_k \beta_k + \varepsilon$$

2.3 Modelos de Regressão Espacial

Geralmente em uma análise de regressão procura-se encontrar um bom ajuste do modelo aos dados, no sentido de reduzir a diferença entre os valores preditos pelo modelo e os valores observados da variável dependente. Também se procura descobrir quais das variáveis explicativas contribuem de forma significativa para o relacionamento linear. Uma hipótese é que as observações não sejam correlacionadas e, portanto, os termos aleatórios (resíduos) □□ do modelo são independentes e não-correlacionados entre si, além de apresentar distribuição normal com média zero e variância constante, isto no modelo clássico. No caso de dados onde está presente a dependência espacial, é muito pouco provável que esta hipótese de observações não correlacionadas seja verdadeira. E no caso mais comum, os termos aleatórios continuam apresentando a autocorrelação espacial presente nos dados, que pode se manifestar por diferenças regionais sistemáticas, ou ainda por uma tendência espacial contínua (LOPES; BRONDINO; SILVA, 2006).

Segundo Serrano e Valcarce (2000), quando se trabalha particularmente com dados de natureza espacial podem aparecer os denominados efeitos espaciais como a heterogeneidade e a autocorrelação espacial. A heterogeneidade aparece quando os dados utilizados para explicar um mesmo fenômeno são de unidades espaciais muito distintas, sendo que os problemas mais freqüentes são a instabilidade estrutural e a heterocedasticidade. A heterocedasticidade espacial ocorre pela omissão de variáveis ou outras formas de especificação que levam à aparição dos denominados erros de medidas. A dependência ou autocorrelação espacial surge sempre que o valor de uma variável em um lugar do espaço está relacionado com seu valor em outro ou outros lugares do espaço.

Para Paiva e Khan (2010), a Análise Exploratória de Dados Espaciais (AEDE), ou Exploratory Spatial Data Analysis (ESDA):

> “é feita com base em um conjunto de ferramentas gráficas e descritivas cujo objetivo é identificar propriedades espaciais dos dados. Está baseada nos aspectos espaciais das informações, ou seja, trata diretamente de questões como dependência espacial e heterogeneidade espacial. O objetivo é descrever a distribuição espacial, os padrões de associação espacial, verificar a existência de diferentes regimes espaciais ou outras formas de instabilidade espacial e identificar agrupamento de valores semelhantes (clusters), ou de observações atípicas (outliers). O cluster espacial é um agregado de ocorrências no espaço ou a ocorrência de valores semelhantes em áreas próximas; já os outliers espaciais são dados cuja localização pode exercer uma forte influência, especialmente nas estimações.”

A presença de autocorrelação espacial é medida usualmente por meio de estatísticas globais. Os indicadores globais constituem uma aproximação mais tradicional do efeito da dependência espacial, em que a estrutura geral de dependência fica resumida em um único valor. (LEMOS et al., 2005a).

Segundo Anselin (2005), a dependência espacial pode ser incorporada nos modelos clássicos de regressão de duas formas: como um regressor adicional na forma de uma variável dependente espacialmente defasada (Wy), ou uma estrutura espacialmente defasada no erro da regressão (We). O primeiro modelo é conhecido como modelo de defasagem espacial ou da variável dependente defasada e o segundo é o modelo do erro espacial ou do erro espacialmente correlacionado.

2.4 Modelo de Defasagem Espacial

No modelo de defasagem espacial, em inglês **SAR** *(Spatial Auto Regressive ou Spatial Lag Model)*, a autocorrelação espacial ignorada é atribuída à variável dependente Y.

$$Y = \rho WY + X\beta + \varepsilon$$

onde:

Y = variável dependente;
X = variáveis independentes;

β = coeficientes de regressão;

ε = erros aleatórios com média zero e variância σ^2 constante;
W = matriz de vizinhança espacial ou matriz de ponderação espacial;
ρ = coeficiente espacial autoregressivo.

2.5 Modelo do Erro Espacial

O modelo autoregressivo do erro, em inglês **CAR** *(Conditional Auto Regressive ou Spatial Error Model)*, pode ser expressado formalmente da seguinte maneira:

$$Y = X\beta + \varepsilon, \quad \varepsilon = \lambda W\varepsilon + \xi$$

onde:

$W\varepsilon$ = erros com efeito espacial;

ξ = erros aleatórios com média zero e variância σ^2;

λ = coeficiente autoregressivo.

Segundo Anselin (1999a), o método de estimação dos parâmetros do modelo normalmente usado é o de máxima verossimilhança, entretanto outros métodos também têm sido propostos, por exemplo, como os de variáveis instrumentais, mínimos quadrados espaciais, método dos momentos, método dos códigos, métodos bayesianos, entre outros.

2.6 Estatística Bayesiana

Thomas Bayes foi um reverendo presbiteriano que viveu no início do século 18 (1701-1761) na Inglaterra. Estudou teologia na Universidade de Edimburgo (Escócia). Em 1737 publicou seu primeiro e único livro de matemática, chamado The doctrine of fluxions (A doutrina dos fluxions) (PENA, 2006).

Para Gangsei (2013), a estatística bayesiana oferece alguns benefícios em comparação com métodos alternativos, sendo um deles, a capacidade de utilizar "todos" os dados disponíveis, e também a sua utilidade para lidar com dados faltantes, formando uma base adequada a fim de aproveitar uma versão ligeiramente modificada do modelo em áreas carentes em dados.

2.6.1 Teorema de Bayes

Este teorema é uma das pedras angulares da estatística das probabilidades combinadas, e é largamente utilizada em áreas a primeira vista pouco relacionadas, como Medicina e Informática (SORENSEN et al., 1994).

Na primeira, por exemplo, o paradigma embasado em evidências é todo construído em cima do teorema de Bayes. Baseado na experiência acumulada de exames e testes para tentar diagnosticar uma doença, o médico enquadra seus pacientes e pode estimar qual a probabilidade de que uma dada doença esteja se manifestando. Ou seja, dada uma probabilidade inicial (por exemplo, o paciente é fumante) e aplicado um exame em que, se sabe, há uma probabilidade de falsos-positivos e falsos-negativos (por exemplo, uma biópsia de pulmão), o médico sabe qual a probabilidade resultante daquele paciente ter a doença (por exemplo, câncer de pulmão) (ANDRADE, 1999).

Na informática, muitos dos sistemas de classificação automática são baseados no teorema de Bayes. Inicialmente o sistema é treinado, aceitando entradas de humanos que dizem que uma dada entrada pertence a determinado grupo. Com o tempo, o sistema acumula um grande banco dessas informações e, aplicando o teorema de Bayes, consegue estimar a probabilidade de cada novo dado de pertencer a cada grupo já classificado (DIGGLE, RIBEIRO, 2007).

Sendo assim, pode-se dizer que o bayesianismo tem dois importantes alicerces epistemológicos: 1º) a visão do universo com base em graus de confiabilidade; 2º) uma regra matemática que explicita como você deve mudar suas crenças à luz de novos dados empíricos. A partir desses pilares se podem deduzir uma série de implicações filosóficas do bayesianismo (PENA, 2006).

Segundo Carroll (2010), a regra de Bayes é um teorema fundamental que pode ser trivialmente derivada dos axiomas da probabilidade. Esta lei pode ser vista como a lei fundamental e universal de aprendizagem.

Ainda segundo Carroll (2010), a estatística bayesiana envolve o uso da regra de Bayes para transformar a probabilidade de dados fornecidos alguns parâmetros para a probabilidade de os parâmetros fornecidos alguns dados.

Wetting (2013) afirma que a regra de Bayes é a base central do raciocínio bayesiano. Trata-se de uma conseqüência direta da regra da cadeia de probabilidades que afirma que a probabilidade conjunta de um conjunto de variáveis aleatórias pode ser escrito como uma cadeia de probabilidades condicionais.

Para Amaral e Inácio (2010), os dois conceitos bayesianos fundamentais são: a) Coisas que são desconhecidas são representadas por distribuições de probabilidade, e b) coisas que são conhecidas (dados) são usadas para aperfeiçoar o conhecimento acerca do problema, a partir do Teorema de Bayes.

E para que se possa chegar ao teorema de Bayes, segundo Pena, (2006), parte-se de princípios básicos:

"Assim, a probabilidade de que observemos simultaneamente um evento A e um evento B é dada por:

$$P(A \cap B) = P(A/B) \,.\, P(B) \qquad (1)$$

Por outro lado, a probabilidade de que observemos simultaneamente um evento A e um evento B também pode ser dada por:

$$P(B \cap A) = P(A \cap B) = P(B/A) \,.\, P(A) \qquad (2)$$

Combinando (1) e (2), temos:

$$P(A/B) \,.\, P(B) = P(B/A) \,.\, P(A) \qquad (3)$$

Rearranjando, chegamos ao teorema de Bayes:

$$P(A/B) = \frac{P(B/A) \,.\, P(A)}{P(B)} \qquad (4)$$

Como geralmente não conhecemos P(B), precisamos usar uma formulação alternativa, que é baseada em:

$$P(B) = P(B \cap A) + P(B \cap Ac) \qquad (5)$$

Onde Ac é o evento complementar de A, também chamado de não-A. Usando nosso conhecimento básico (equação 1 acima) e substituindo, obtemos:

$$P(B) = [P(B/A) \,.\, P(A)] + [P(B/Ac) \,.\, P(Ac)] \qquad (6)$$"[1]

Substituindo 6 em 4 obtemos a formulação alternativa:

Figura 6. Fórmula alternativa. Fonte: Pena (2006).

A regra de Bayes apresenta uma maneira de como alterar as probabilidades a priori considerando novas experiências para obter probabilidades a posteriori.

Para DiMaggio (2014), uma vantagem do método bayesiano é o ajusto que se pode dar nos dados espaciais usando uma distribuição a priori que irá explicar como que a correlação espacial irá se comportar no modelo para dar resultados mais precisos.

Talvez a vantagem prática mais importante da abordagem bayesiana para a aprendizagem é "o princípio da otimização bayesiana", que afirma que, quando usado em conjunto com a teoria da decisão, decisões baseadas em inferência usando a regra de Bayes sempre produzirão uma utilidade esperada igual ou maior do que as decisões com base em qualquer outra técnica (CARROLL, 2010).

A probabilidade a priori P(A), segundo Amaral e Inácio (2010), fornece a probabilidade dos diferentes valores possíveis da quantidade de interesse antes dos dados serem considerados. Representa o estado do conhecimento anterior aos dados.

A priori informativa é quando se conhece alguma coisa acerca do parâmetro desconhecido A ou sobre o experimento sendo realizado, usam-se essas informações no estabelecimento da função densidade de

[1] PENA, S. D. Bayes: o 'cara'!. **Ciência Hoje**, Rio de Janeiro, v.38, n.228, p. 22 – 29, jul. 2006. Disponível em: <http://cienciahoje.uol.com.br/banco-de-imagens/lg/protected/ch/228/bayes.pdf/at_download/file>. Acesso em: 21 out. 2013.

probabilidade a priori para A. Se essa densidade contiver parâmetros, estes são estabelecidos fora do modelo (hiperparâmetros) (BRASIL, 2012).

Por outro lado, a priori não informativa ocorrem em situações que se conhece muito pouco ou quando não se tem nenhuma informação disponível a priori, sendo que todos os possíveis valores de A como igualmente prováveis, isto é, com uma distribuição a priori uniforme (EHLERS, 2003).

Para Amaral e Inácio (2010), a verossimilhança P(B/A), fornece a probabilidade de obter o dado, considerando diferentes valores possíveis da quantidade desconhecida de interesse (hipótese H).

E a probabilidade a posteriori P(A/B), fornece a probabilidade dos diferentes valores possíveis da quantidade de interesse depois de considerar os dados, representando o estado do conhecimento posterior aos dados (AMARAL e INÁCIO, 2010).

Em resumo, a probabilidade a posteriori é a combinação da priori com a verossimilhança, sendo que a diferença entre priori e a posteriori é o aprendizado que se obtêm com os dados.

2.7 Inferência Bayesiana

Para Hainline (2013), a inferência estatística moderna pode ser dividida em duas principais escolas de pensamento: freqüentista e bayesiana.

A abordagem freqüentista é o método clássico de análise estatística em que informação prévia, obtida por estudos ou ensaios anteriores, é usada apenas durante a fase de planejamento. Estatística Bayesiana, ao contrário, constrói informação prévia para a análise formal, uma vez que se torna disponível. Ensaios anteriores, estudos em outros países, ou opiniões de especialistas são considerados fontes válidas de informação prévia. O uso de informação prévia na análise pode ser útil no planejamento do estudo e pode-se argumentar que um resultado mais preciso resulta de uma análise bayesiana. A informação prévia permite ao pesquisador a diminuir o alcance de um julgamento, o que resulta em um resultado mais específico (HAINLINE, 2013).

Em uma análise freqüentista, a interpretação dos dados depende das intenções do pesquisador. Para freqüentistas, antes que o experimento é conduzido, valores críticos devem ser determinados, e p-valores são a base para a tomada de decisões. Estatísticos bayesianos afirmam que as hipóteses devem ser comparadas com o quão bem elas explicam os dados. P-valores representam a probabilidade de observar uma estatística de teste igual ou mais extrema quanto aquela observada em uma amostra, assumindo verdadeira a hipótese nula do que o observado, assumindo que a hipótese nula é verdadeira (HAINLINE, 2013).

Um valor-p pequeno em uma análise freqüentista resultará na rejeição da hipótese nula (muitas vezes assumindo que não há associação) em favor da hipótese alternativa (muitas vezes alegando associação está presente). Estatística bayesiana permite ao pesquisador obter *odds ratio* e probabilidades de previsão que possam demonstrar a magnitude do efeito que a variável tem sobre a variável de desfecho (HAINLINE, 2013).

Com análise bayesiana, cada análise compõe-se de dois tipos de informações: os dados que estão sendo analisados e a informação prévia. Sendo assim, o pesquisador é livre para escolher qualquer informação prévia que ajuda a explicar os dados que estão sendo analisados. (HAINLINE, 2013).

Para Diggle e Ribeiro (2007), a inferência bayesiana é o processo de se encontrar um modelo de probabilidade para um conjunto de dados e resumir o resultado para uma distribuição de probabilidade sobre os parâmetros do modelo, que são tratadas como variáveis aleatórias e sobre quantidades não observadas como valores de novas observações (preditivas).

O teorema de Bayes é usado na inferência estatística para atualizar estimativas da probabilidade de que diferentes hipóteses sejam verdadeiras, baseado nas observações e no conhecimento de como essas observações se relacionam com as hipóteses (SORENSEN, 1996).

Para Amaral e Inácio (2010), os objetivos da estatística bayesiana são:

- representar o desconhecimento a priori sobre os parâmetros do modelo com uma distribuição de probabilidade (distribuição a priori);
- atualizar esse desconhecimento a priori com dados atuais - verossimilhança (likelihood);
- e produzir uma distribuição de probabilidade para o parâmetro que contenha menos desconhecimento (distribuição posteriori).

 ISSN: 2047 - 0398
SPECIAL ISSUE – V|05|N|01| June | 2015

Segundo Bernando (2001), a estatística bayesiana utiliza a probabilidade como uma medida condicional de incerteza associada com a ocorrência de um determinado evento, dada a informação disponível e os pressupostos aceitos.

> "A abordagem bayesiana para inferência permite estimativa de parâmetros utilizando informações provenientes dos dados através da função de probabilidade, bem como informações provenientes de outras fontes antes vistos os dados (ou seja, estudos anteriores, julgamentos subjetivos), que é formalizada através de distribuições anteriores. Teorema de Bayes combina a função de verossimilhança e a distribuição prévia definição de uma nova quantidade, conhecida como distribuição posterior que forma a base da inferência Bayesiana. Os parâmetros são considerados como aleatórios e as suas estimativas resulta não só num único valor, mas nas probabilidades de seus valores possíveis que são dadas pela sua distribuição de probabilidade, conhecida como a distribuição marginal posterior."[2]

Para Resende (2000), em inferência Bayesiana, certos métodos que assumem distribuições a priori não informativas, são essencialmente de inferência verossimilhança, tais como o método VEIL (ou da verossimilhança integrada de Gianola & Foulley, 1990) de estimação de componentes de variância, os quais mantêm a propriedade de conduzir a análise exata de amostra de tamanho finito.

E para Lavine (2000), a análise Bayesiana utiliza a distribuição posterior para resumir o estado de conhecimento. A distribuição posterior combina informações a partir dos dados na mão expressa através da função de verossimilhança, com outras informações expressas através da distribuição prévia.

2.8 Técnicas de Krigagem

A origem da palavra krigagem vem do nome Daniel G. Krige, que foi o primeiro a introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas de mineração (DELFINER e DELHOMME, 1975).

Segundo Câmara (2004), num primeiro momento o método da krigagem foi desenvolvido para resolver os problemas de mapeamentos geológicos, mas seu uso foi se expandindo com sucesso no mapeamento de solos, mapeamento hidrológico, mapeamento atmosférico e outros campos correlatos.

O método de Krigagem tem por fundamento a Teoria da Variável Regionalizada (TVR), desenvolvida por Matheron (1965). Uma variável regionalizada é uma variável distribuída no espaço ou tempo cujos valores são considerados como realizados de uma função aleatória.

Dantas (2003) diz que esta teoria identifica que a distribuição espacial de uma variável é expressa pela soma de três componentes:

- uma componente estrutural, tendo uma média constante ou tendência;
- uma componente aleatória espacialmente correlacionada, também chamada de variação regionalizada;
- uma componente aleatória não correlacionada espacialmente (erro residual).

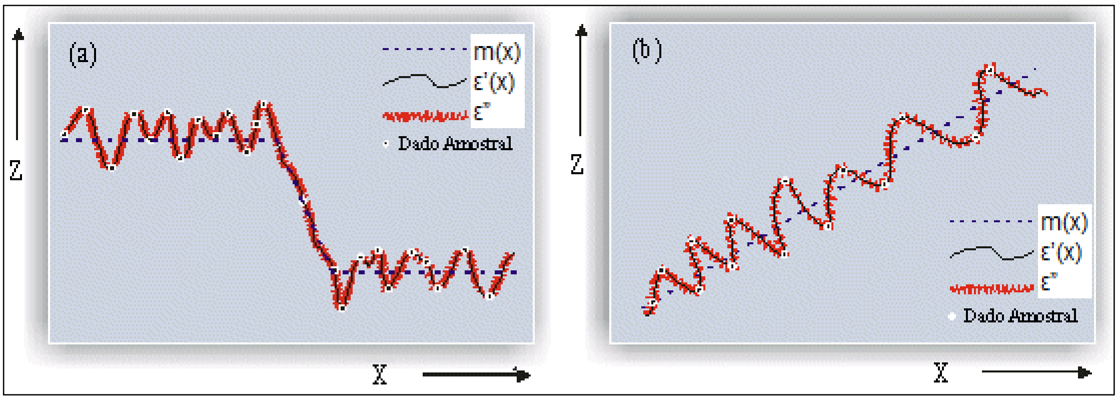

Figure 2. Distribuição espacial de uma variável. Fonte: Burrough (1987).

[2] GOȘONIU, L. **Development of Bayesian geostatistical models with applications in malaria epidemiology**. Thesis Doktors, Universität Basel, Philosophisch-Naturwissenschaftlichen Fakultät, 2008.

O valor de uma variável Z, em uma posição geográfica x, representado por Z(x), fica definido como:

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon''$$

onde:

m(x) – é uma função determinística descrita pela componente estrutural de Z em x;
ε'(x) – é a variação regionalizada;
ε" – o resíduo do modelo, do tipo gaussiano, espacialmente independente, com média zero e variância constante σ^2.

> "A krigagem engloba um conjunto de métodos de estimação, incluindo procedimentos estacionários (krigagem simples e ordinária), não estacionários (krigagem universal, funções intrínsicas de ordem k), univariados e multivariados (co-krigagem etc)"[3]

Para Soares (2002) a krigagem é:

> "É o uso de médias móveis para evitar a superestimação. Ela difere de outros métodos de interpolação pela maneira como os pesos são atribuídos às diferentes amostras. Na krigagem, o procedimento é semelhante ao de interpolação por média móvel ponderada, exceto que os pesos são determinados a partir de uma análise espacial, baseada no semivariograma experimental."

A krigagem linear engloba um conjunto de métodos de estimação.

Segundo Soares (2002) a Krigagem simples foi inicialmente utilizada como um estimador de valores de atributos numéricos, em posições não observadas, para mapeamentos por médias ponderadas dos valores existentes das amostras locais.

$$\hat{Z}_{xo} = m + \sum_{i=1}^{n} \lambda_i [Z(x_i) - m]$$

Sendo $\hat{Z}_{X0}$ o valor desconhecido que pode ser estimado a partir de uma combinação dos n valores observados, m é a média que supõe-se que é conhecida a priori e λ_i são os pesos obtidos a partir do seguinte sistema de equações, denominado sistema de krigagem simples (CARVALHO, 1997):

$$\underset{j=1}{\overset{n}{Z}} \lambda_i C(x_i, x_j) = C(x_i, x_0) \text{ para i = 1, ..., n (n equações)}$$

onde:

- $C(x_i, x_j)$ refere-se à função covariância correspondente a um vetor, h, com origem em x_i e extremidade em x_j;
- $C(x_i, x_0)$ refere-se a função covariância correspondente a um vetor, h, com origem em x_i e extremidade no ponto a ser estimado x_0.

3. RESULTADOS E ANÁLISES

3.1 Modelos de Regressão

Para encontrar um modelo de regressão que melhor explique o valor da localização na área em estudo, buscou-se uma equação com as variáveis significativas na formação do valor para os imóveis.

Foram feitas transformações das variáveis independentes e da variável dependente, e também foram analisadas as interações entre as variáveis independentes para encontrar o melhor modelo de regressão.

3.2 Estimação da matriz de pesos

Na regressão espacial foi utilizada a matriz de vizinhança de até 300 metros, pois foi a que melhor explicou a formação do valor dos imóveis do município de Balneário Camboriú/SC no bairro centro, como pode ser observado no semivariograma da Figura 14 abaixo:

[3] CÂMARA Gilberto, et al.. **Análise Espacial de Dados Geográficos. São José dos Campos**, INPE, 2003 - on-line (3ª edição, revista e ampliada). Dezembro 2004. http://www.dpi.inpe.br/gilberto/tutoriais/analise/ Acesso: 10/02/2007.

 ISSN: 2047 - 0398
SPECIAL ISSUE – V|05|N|01| June | 2015

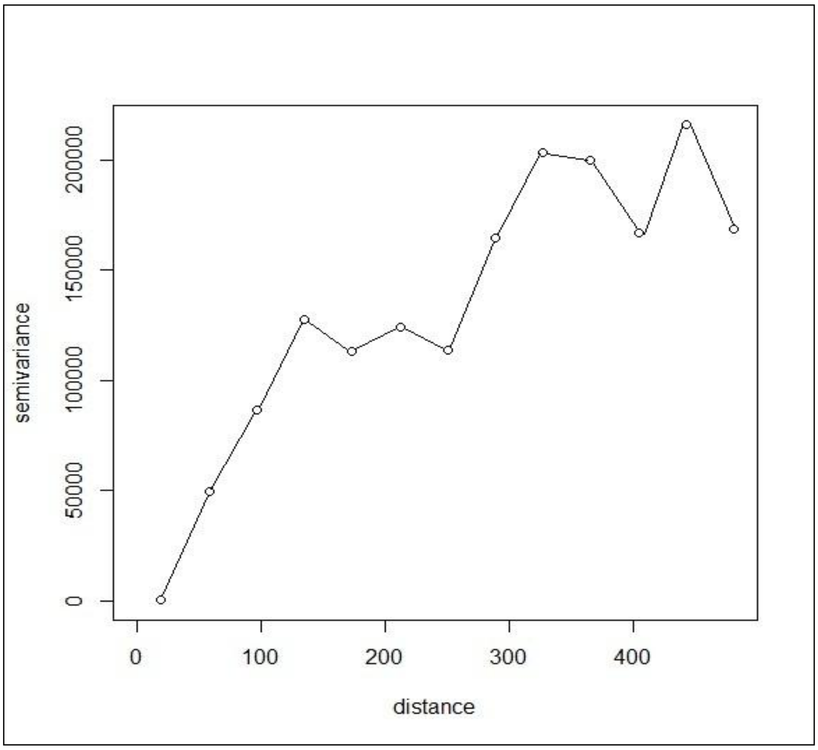

Figura 3. Semivariograma experimental omnidirecional ou isotrópico do VU. Fonte: O Autor.

Como pode ser observado no semivariograma experimental omnidirecional da Figura 3 os dados apresentam maior dependência espacial até uma distância aproximadamente de 300 metros, e depois desta distância a dependência diminui consideravelmente apresentando uma estrutura praticamente estacionária. Desta maneira, pode se notar que os imóveis que se encontram muito próximas exercem uma forte influência entre si e esta influência diminui consideravelmente com a distância, de forma que, aqueles imóveis que se encontram separados por distâncias maiores quase não apresentam influência entre si.

3.3 Testes de autocorrelação espacial

Foram calculadas as estatísticas correspondentes aos testes de autocorrelação espacial de Moran's I, e também os testes do Multiplicador de Lagrange (LM) com a matriz W (do peso), definida para o modelo da variável dependente e do erro e suas estatísticas robustas. Na Tabela 1 podem-se observar os resultados.

TESTE	VALOR	PROBABILIDADE
Moran's I	0,8372797	0,4024352
LM (lag)	6,0234965	0,0141167
LM robusto (lag)	5,8629350	0,0154630
LM (erro)	0,2933405	0,5880878
LM robusto (erro)	0,1327790	0,7155679

Tabela 1. Testes de autocorrelação espacial do modelo. Fonte: O Autor.

Esses resultados mostram que há uma autocorrelação espacial nos resíduos do modelo de regressão por mínimos quadrados, pois a probabilidade do teste $LM(lag)$ é menor que 10%. Sendo assim, o modelo da defasagem espacial apresenta-se significativo.

No modelo de defasagem espacial foi usado a matriz de até 300 metros, pois foi a que melhor explicou a formação do valor dos apartamentos do município de Balneário Camboriú/SC no bairro centro.

Na Tabela 2 podem ser observados os resultados encontrados.

ISSN: 2047 - 0398
SPECIAL ISSUE – V|05|N|01| June | 2015

Variável	Coeficiente	Erro Padrão	Valor Z	Probabilidade
W_ln(VU)	0,02911609	0,01145382	2,542043	0,0110207
Constante	5,307201	0,2747589	19,31585	0,0000000
1/(AT)	34,99747	11,82741	2,959013	0,0030864
1/(DPRAIA)	2,788294	0,2047816	13,61594	0,0000000
ND	0,5909657	0,1238395	4,772029	0,0000018
Andar	0,0140154	0,003373855	4,154119	0,0000327
Idade	-0,01587764	0,002556708	-6,21019	0,0000000
Frente	0,1930556	0,05445892	3,544976	0,0003927

Tabela 2. Modelo da defasagem espacial. Variáveis independentes. Fonte: O Autor.

Onde:
VU = valor unitário (R$/m²);
AT = area total (m²);
DPRAIA = distância da praia ao longo da rua (m);
ND = número de domrintórios;
Andar = andar do apartamento;
Idade = idade do apartamento;
Frente = apartamento de frente ou de fundos.

O coeficiente autorregressivo espacial é estimado como 0,02911609, e é significativo (p = 0,0110207).
Segundo Trivelloni (2005), a estatística z corresponde ao equivalente para a regressão por máxima verossimilhança ao valor t de *Student* para o método de mínimos quadrados. As probabilidades indicam o grau de significância de cada variável de forma análoga que na regressão por mínimos quadrados.

3.4 Krigagem para avaliação em massa de imóveis

Para estimar o valor unitário dos apartamentos, foi feito uma homogeneização para os terrenos avaliando, sendo que para isso foram usados três apartamentos paradigmas conforme a Tabela 3.

Paradigma	Frente	Andar	Idade	ND	AT
1	1	4	6	3	238,65
2	0	5	5	3	170,00
3	1	3	12	4	156,00

Tabela 3. Apartamentos paradigmas. Fonte: O Autor.

Após se ter encontrado o modelo de predição que melhor correspondeu com a realidade para a praia central de Balneário Camboriú/SC, foi feita uma krigagem ordinária e uma krigagem bayesiana.

Usando os Valores Unitários calculados com o modelo de regressão na defasagem foi construído um semivariograma experimental conforme Figura 4.

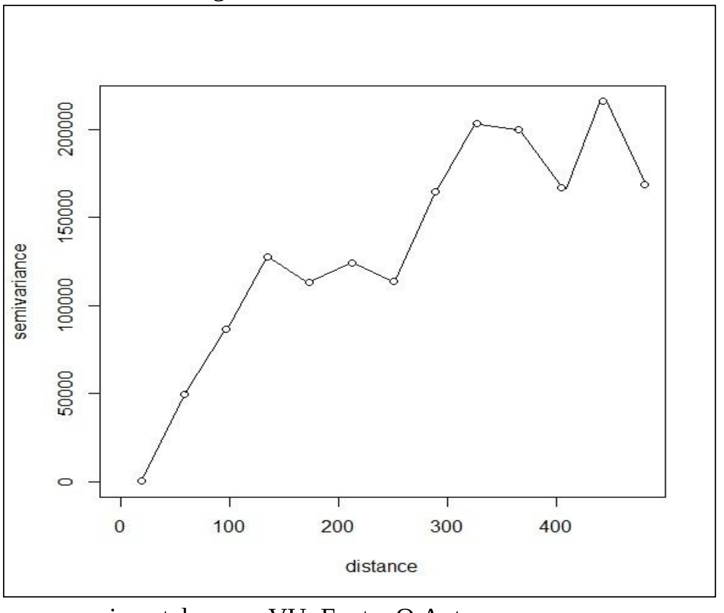

Figura 4. Semivariograma experimental para o VU. Fonte: O Autor.

ISSN: 2047 - 0398
SPECIAL ISSUE – V|05|N|01| June | 2015

Pode-se observar que o gradiente de variação se estabiliza aproximadamente a partir dos 300 metros de distância.

Em seguida para os mesmo Valores Unitários foi feita uma Krigagem Bayesiana para os imóveis avaliando conforme Figura 5. Podem ser observadas as áreas com maior e menor valor conforme o gradiente de cores que permite uma análise visual mais rápida e simples das áreas.

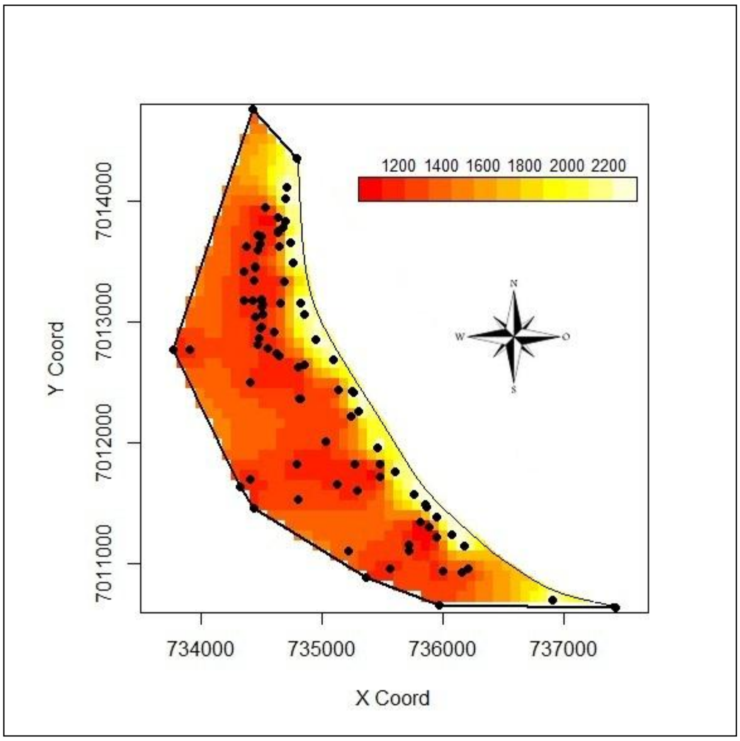

Figura 5. Krigagem para variável VU usando método SLM para os imóveis avaliando. Fonte: O Autor.

Na Figura 6 a área avaliada foi separada em três setores A, B e C, sendo assim pode-se identificar ainda melhor a variação de valores.

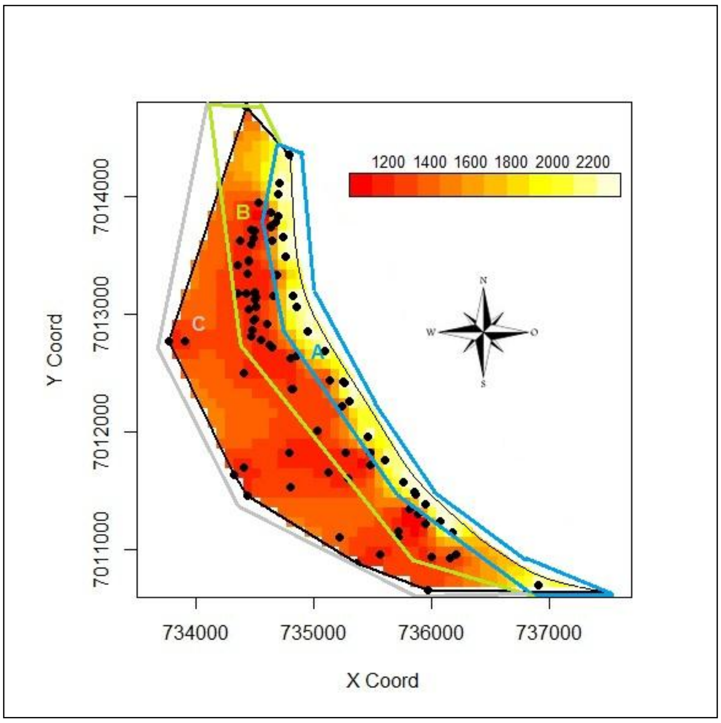

Figura 6. Krigagem para variável VU usando método SLM para os imóveis avaliando dividida em três setores. Fonte: O Autor.

A área em tom amarelado, que pode ser observada na setor A, são as áreas mais valorizadas, pois são aquelas de frente ao mar e na proximidade.

Entretanto as áreas em tom avermelhadas, os setores B e C, são aquelas áreas que ficam mais de 300 metros da praia, sendo que quanto mais afastado for o imóvel (setor C), o valor do metro quadrado diminui. O setor C é aquele que fica próximo a BR 101. Porém não diminui de forma acentuada, pois os imóveis localizados na área em que o modelo foi aplicado existem pouca variação de valor.

A cidade de Balneário Camboriú como um todo, é uma cidade valorizada e com crescimento nos valores dos imóveis nos últimos 10 anos.

Por fim, os resultados da krigagem da Figura 17 apresentaram coerência com a realidade da área de estudo possibilitando uma fácil interpretação.

4. CONCLUSÕES

O método de regressão *SLM* utilizado comprovou ser adequado pelos testes de autocorrelação espacial nos resíduos do modelo de mínimos quadrados calculados com a matriz W serem significativos e pela significância dos modelos de regressão espacial estimados.

O semivariograma experimental mostrou ser uma ferramenta muito importante para a construção da matriz de pesos espaciais para a regressão espacial, definindo assim o principal parâmetro da matriz que é a distância máxima de vizinhança em função da análise dos dados de mercado.

A krigagem bayesiana obteve melhores resultados em relação a krigagem ordinária, sendo assim, serviu para gerar valores entre os vizinhos, pois dificilmente se obtém dados na coleta em campo para toda a área estudada. Entretanto deve ser lembrado que a krigagem é um método de interpolação, que vizinhos muito afastados podem gerar erros na estimação dos valores unitários.

Uma conclusão importante deste trabalho é o uso combinado da geoestatística e da krigagem bayesiana, pois antes desse trabalho não se tem estudos sobre tal.

Os resultados foram muito bons, pois a estatística espacial ajudou a encontrar o modelo de predição de valores unitários condizentes com a realidade do local.

BIBLIOGRAPHICAL REFERENCE

ABNT (ASSOCIAÇÃO BRASILEIRA DE MORMAS TECNICAS) – Avaliação de bens – parte 2: imóveis urbanos: Norma **NBR – 14.653-2/2011**, Rio de Janeiro: ABNT, 2011.

ANDRADE, P. J. N. **Sistemas Especialistas de Apoio ao Diagnóstico em Medicina. Relações com o Teorema de Bayes e com a Lógica do Raciocínio Diagnóstico**. Fortaleza, 1999. Available in: < http://publicacoes.cardiol.br/abc/1999/7306/73060008.pdf>. Access: 21 out. 2014.

AMARAL, E. F. L. e INÁCIO, M. M. **Modelos Bayesianos**. Belo Horizonte, 2010. Available in: <http://www.ernestoamaral.com/docs/dcp859b4-102/Aula142.pdf>. Access: 15 mar. 2014.

ANSELIN, L. **Spatial Econometrics**. Discussion paper. Bruton Center, School of Social Sciences, University of Texas at Dallas, 1999a.

AYRES, A. **Como avaliar imóveis**. São Paulo: Editora Imobiliária, 1996.

BERNARDO, J. M. **Bayesian Statistics**. Encyclopedia of Life Support Systems (EOLSS). Paris: UNESCO (to appear) 2001.

CÂMARA, G., et al. **Análise Espacial de Dados Geográficos**. São José dos Campos, INPE, 2003 - on-line (3a. edição, revista e ampliada). Dezembro 2004. Available in: <http://www.dpi.inpe.br/gilberto/tutoriais/analise/> Access: 10/02/2007.

CAN, A. GIS and Spatial Analysis of Housing and Mortgage Markets. **Journal of Housing Research**, v.9, 1998.

CARROLL, J. L. **A Bayesian Decision Theoretical Approach to Supervised Learning, Selective Sampling, and Empirical Function Optimization**. A dissertation submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, 2010. 357p.

CESARE, C. de. **Valuación de inmuebles para fines fiscales**. Impuesto a la propriedade inmobiliaria. Lincoln Institute of Land Policy, 2004.

CRESSIE, N. A. C. **Statistics for Spatial Data**. Revised Edition, John Wiley & Sons, Inc., 1993.

DANTAS, R. A. **Modelos Espaciais Aplicados ao Mercado Habitacional um Estudo de Caso Para a Cidade do Recife**. Tese de Doutorado, Universidade Federal de Pernambuco, Recife, 2003. 133p.

DELFINER, P.; DELHOMME, J. P. Optimum interpolation by Kriging. In: Davis, J. C.; McCullagh, M. J. ed. **Display and analysis of spatial data**. New York, John Wiley, 1975.

DIGGLE, P. J.; RIBEIRO JR., P. J. **Model-Based geostatistics**. New York: Springer, 2007. 228p.

GANGSEI, L. E. **A Bayesian method for estimating moose (Alces alces) population size based on hunter observations and killed at age data**. Master Thesis in the Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 2013.

HAINLINE, A. E. **Frequentist and bayesian modeling in the presence Of unmeasured confounding**. A Thesis Submitted to the Faculty of Baylor University In Partial Fulfillment of the Requirements for the Honors Program. Waco, Texas May 2013. 56p.

HALBERT, G. J. **Estimating the effects of air pollution on human health in Greater Glasgow in space and time**. Dissertation Submitted to the University of Glasgow for the degree of Master of Science. School of Mathematics & Statistics, 2013. 81p.

GONZÁLEZ, M. A. S. **Aplicação de Técnicas de Descobrimento de Conhecimento em Base de Dados e de Inteligência Artificial em Avaliações de Imóveis**. Tese de Doutorado, Universidade Federal de Rio Grande do Sul, Programa de Pós-Graduação em Engenharia Civil, 2002. 296p.

GOŞONIU, L. **Development of Bayesian geostatistical models with applications in malaria epidemiology**. Thesis Doktors, Universität Basel, Philosophisch-Naturwissenschaftlichen Fakultät, 2008. 166p.

LAVINE, M. **What is Bayesian statistics and why everything else is wrong**. Technical report, Duke University, North Carolina, 2000.

LEMOS, M. B. et al. A Organização Territorial da Indústria no Brasil. In: NEGRI, J. A.; SERGIO, M. (Orgs.). **Inovações, Padrões Tecnológicos e Desempenhos das Firmas Industriais Brasileiras**. Brasília: IPEA, 2005, p. 325-363.

LOPES, S. B.; BRONDINO, N.C.M.; SILVA, A.N.R. **Análise Do Desempenho De Modelos De Regressão Espacial Na Previsão De Demanda Por Transportes**. In: XIV Congreso Panamericano de Ingeniería de Tránsito y Transporte, 2006, Las Palmas de Gran Canaria. XIV PANAM, 2006.

MATHERON, G.: **Les Variables Regionalisées et Leur Estimation**, Masson, Paris. 1965.

MÖLLER, L. F. C. **Planta de valores genéricos: avaliação coletiva de imóveis para fins tributários**. Porto Alegre: Sagra-DC Luzzatto, 1995.

PAIVA, W. L., KHAN, A. S. DEPENDÊNCIA ESPACIAL E EMPREGO FORMAL: O que é possível afirmar para indústria cearense?. In. **Anais** do VI Encontro de Economia do Ceará em Debate. Fortaleza - CE, 05 de novembro de 2010.

PENA, S. D. Bayes: o 'cara'!. **Ciência Hoje**, Rio de Janeiro, v.38, n.228, p. 22 – 29, jul. 2006. Available in: <http://cienciahoje.uol.com.br/banco-de-imagens/lg/protected/ch/228/bayes.pdf/at_download/file>. Access: 21 out. 2013.

POLI, C. R.; FRANCO, E. R.; ROCHA, R. J. de G.; AKIN, R. G. Planta de Valores Genéricos de Zona Residencial do Município de Fazenda Rio Grande – PR. XVI COBREAP Congresso Brasileiro de Engenharia de Avaliações e Perícias, **Anais**. Manaus, 2011.

RESENDE, M. D. V. de. **Inferência Bayesiana e simulação estocástica (amostragem de Gibbs) na estimação de componentes de variância e de valores genéticos em plantas perenes**. Colombo: Embrapa Florestas, 2000.68p. (Embrapa Florestas. Documentos, 46).

SERRANO, R. M.; VALCARCE, E. V. **Técnicas econométricas para el tratamiento de dados espaciales: la econométrica espacial.** Edictions Universita de Barcelona, Barcelona, 2000.

SILVA, E. **Cadastro técnico multifinalitário: base fundamental para avaliação em massa de imóveis**. Florianópolis, 2006. Tese de doutorado em engenharia civil, Universidade Federal de Santa Catarina – UFSC, 220p.

SOARES, A. F. **Uso do Módulo "Análise" do SPRING (Versão 3.2.02) Para Estudos de Dados Climáticos**. Embrapa. Campinas – SP. 2002.

SORENSEN, D.; WANG, C. S.; JENSEN, J.; GIANOLA, D. Bayesian analysis of genetic change due to selection using Gibbs sampling. **Genetic Selection Evolution**, v. 26, p.333-360, 1994.

TRIVELLONI, C. A. P. **Método Para Determinação do Valor da Localização Com Uso de Técnicas Inferências e Geoestatísticas na Avaliação Em Massa de Imóveis**. Tese de Doutorado, Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia Civil, 2005.