

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Resenha 2

Everton Batista da Rocha

Trabalho apresentado ao Curso de Pós-Graduação em Estatística e Experimentação Agronômica, da Universidade de São Paulo, Campus Piracicaba, como requisito parcial para a obtenção de crédito na disciplina de Geoestatística.

Piracicaba

2011

BANERJEE, S., **On Geodetic Distance Computation in Spatial Modeling.** *Biometrics*, 61, 617-625, June 2005.

A análise estatística de dados espaciais sempre necessita da detecção e ajuste de modelos associados a distâncias na superfície da terra, de forma que a computação dessas distâncias torna-se indispensável na modelagem espacial. Diante disto, objetivo deste artigo é explorar o uso de métricas planares e investigar o impacto do seu uso sob a modelagem espacial.

O artigo foi dividido pelo autor nas seguintes seções: introdução, revisão de modelos de regressão espacial, cálculo de distâncias, ilustração e discussão.

Na seção **introdução** o autor justifica o estudo/trabalho citando que o cálculo de distâncias são indispensáveis na análise espacial, pois sua precisão afeta na obtenção e análise de variogramas, auxilia na especificação de distribuições a priori no parâmetro de alcance na modelagem bayesiana e servem de valores iniciais nos algoritmos de mínimos quadrados lineares. O autor afirma que tratar as coordenadas geodésicas como um plano pode induzir a uma enganosa anisotropia.

Na seção **revisão de modelos de regressão espacial** o autor considera o modelo $Y(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s})$, em que $Y(\mathbf{s})$ é a variável resposta ou dependente, observada nas localizações \mathbf{s} , referenciada pela sua latitude e longitude, ao longo de um vetor de covariáveis $\mathbf{x}(\mathbf{s})$. Os resíduos são particionados entre um processo espacial $w(\mathbf{s})$ e um processo independente $\epsilon(\mathbf{s})$. Os objetivos inferenciais incluem a estimação dos coeficientes de regressão, variância espacial e *nugget* e no tamanho da associação espacial com o aumento da distância.

Em **cálculo de distâncias**, o autor afirma que enquanto a geometria esférica sugere métricas naturais, ele não recomenda nenhuma *true distance metric* porque elas podem não ser apropriadas para a análise de dados científicos. O autor então compara o uso dos seguintes métodos para o cálculo de distâncias: *geodetic*, *naive euclidian*, *chord*, *mercator*, *sinusoidal* e *centroid-based*. Observa-se uma super-estimação e uma sobre-estimação nas métricas *naive euclidian* e *chordal*, entretanto esta última tem uma importante implicação teórica na modelagem espacial. Há uma elevação na estimação pelo método de *mercator*, e os demais apresentam valores próximos. Dada esta discrepância de valores, uma alternativa para o uso de métricas euclidianas é o uso de uma projeção planar no domínio espacial, o que é comum entre usuários de GIS, entretanto o autor aponta que existem muitos pacotes/software estatísticos que trabalham apenas com as coordenadas espaciais em duas dimensões, citando o programa WinBUGs e o pacote geoR.

Ainda em relação aos métodos de cálculo de distâncias, no tocante a projeção de mapas, o autor aponta que projeção baseada na métrica *sinusoidal* e *centroid-based* tendem a distorcer as distâncias muito menos que a métrica de *Mercator*, que é ainda pior que a métrica *naive euclidian*.

E estas considerações tem impacto sobre o processo de predição espacial. Em particular, a projeção considerando a métrica *centroid-based*, tem a desvantagem de que a dependência dos dados não muda com a adição de novas áreas ao estudo. Entretanto, essas considerações são mais pertinentes do ponto de vista geográfico ou geodésico do que do ponto de vista estatístico espacial, pois eles não refletem como a estimação estatística é afetada, onde pontos que estão muito próximos tem grande influência na análise.

Na seção **ilustração**, o autor analisa a modelagem espacial sobre diferentes computações geodésicas em um conjunto de dados de temperatura obtidos da *National Center for Atmospheric Research (NCAR)*, em Boulder, Colorado, com temperatura média (em unidades de 10°C) obtida para 50 áreas no mês de janeiro de 1997, de forma que esta é a variável dependente do estudo, $Y(\mathbf{s})$. No estudo é também fornecida a altitude (em unidades de 100m) em cada área, que é considerada como uma possível covariável. O processo subjacente então foi a temperatura média registrada em cada área. Um modelo univariado espacial, seguindo o modelo apresentado pelo autor, mostra uma explicação da temperatura dada a elevação da altitude, explícita na correlação espacial dos dados. Verifica-se que as estimativas de regressão são quase não afetadas pelas métricas; em cada caso há um intercepto positivo significativo e, muito excepcionalmente, um efeito significativo negativo da elevação da temperatura. Nota-se ainda que o alcance espacial estimado usando as métricas *chordal*, *sinusoidal* e *centroid-bases* são semelhantes aos da faixa estimada a partir da métrica geodésica. Em comparação, com as métricas *euclidian* e *Mercator*, a estimativa da projeção do alcance por um fator excede a 1,5 vezes o intervalo geodésico para a função de correlação exponencial, isto é ainda mais drástico para o *Matérn*. Aparentemente, esta discrepância parece ser consistente com os efeitos puramente geográficos comentados anteriormente. As estimativas das projeções *sinusoidal* e *based-centroid* também estão muito perto da geodésica, corroborando a afirmação do autor de alternativas viáveis. Esta seção encerra com a comparação dos modelos obtidos pelas diferentes métricas, através da medida DIC, considerando o ajuste bayesiano, e os resultados são similares, apontando os ajustes com as métricas *naive euclidian* e *mercator* como os menos favoritos, tanto para a correlação exponencial como a *Matérn*.

Na seção **discussão** o autor apenas aponta que uma métrica natural não é necessariamente a mais apropriada e deve-se considerar um método métrico e que ainda, métricas que influenciam/diferem do ponto de vista geográfico/geodésico **podem** alterar a análise estatística espacial.

O artigo é de fundamental importância na análise estatística espacial, pois trata com clareza do impacto do uso de métricas geodésicas para o cálculo de distâncias. Uma vez que o cálculo

das distâncias influencia sobretudo na análise canônica geoestatística, que serve de base para a classificação de um conjunto de dados como isotrópico ou anisotrópico e ainda, em dados com dependência espacial ou não dependência espacial. Aqui, o uso de uma métrica não adequada pode levar a uma interpretação errônea a respeito do comportamento espacial, da dependência espacial, do processo em estudo.

E ainda, apesar de não conclusivo, o autor elucidou que métricas que influenciam no cálculo das distâncias, do ponto de vista geográfico e geodésico, **podem** influenciar na inferência. Desta forma, mostra a necessidade de mais estudos a respeito desta influência, a fim de confirmar (ou não) este(a) indício (hipótese).