

## DAS COISAS CERTAS E INCERTAS

O inventor e político americano Benjamin Franklin costumava dizer que na vida só existem duas coisas certas: a morte e os impostos. Tivesse vivido no Brasil e chegasse a conhecer certos senadores e dirigentes de clubes de futebol, reduziria suas certezas quanto aos impostos. Mas Ben Franklin estava no caminho certo: esta vida é mesmo incerta e por boas razões. Basta considerar nossa realidade e pensar no mundo que nos cerca. Os quasares, placas tectônicas, luas, baratas, bactérias e átomos de hidrogênio do universo tem o direito à sua própria existência, de forma independente de nossa consciência. Como conhecer o mundo, senão através de nossos sentidos e de nossos instrumentos? E aí mora o perigo: qualquer instrumento de medida, seja ele um olho humano, um aparelho GPS ou um sensor de imageamento a bordo de satélite, tem suas limitações. Pior ainda: nossos sensores apenas nos fornecem amostras incompletas dos fenômenos da natureza, e temos que apelar para nossa imaginação, ou melhor, para nossa capacidade de construir modelos abstratos de como o mundo funciona.

Em “Sobre Nuvens e Relógios”, um texto clássico sobre o problema do conhecimento, o grande filósofo da ciência Karl Popper apresenta duas grandes posturas complementares do homem frente ao mundo. A visão determinística considera que a natureza é composta de fenômenos perfeitamente mensuráveis (os *relógios*) e que o conhecimento consiste em sua descrição, preferencialmente na forma de leis. Já a abordagem estatística considera que muitos fenômenos naturais são cognoscíveis apenas de forma aproximada (*as nuvens*), e tudo o que podemos saber sobre eles é sua descrição em termos de distribuições de probabilidade.

Esta dicotomia entre a visão determinística e a estatística reflete-se diretamente na produção de medidas de qualidade para dados geográficos. A visão determinística (os relógios) supõe que é possível conhecer elementos do mundo de maneira precisa; a medida de qualidade de nossa informação será determinada pela melhor ou pior aproximação destes elementos em nossos produtos. Este é o fundamento dos *padrões de exatidão cartográfica*. Conforme define o Manual de Normas Técnicas da Cartografia Brasileira (disponível no site [www.concar.ibge.gov.br](http://www.concar.ibge.gov.br)): “Noventa por cento dos pontos bem definidos numa carta, quando testados no terreno, não deverão apresentar erro superior ao Padrão de Exatidão Cartográfica Planimétrico estabelecido”. O conceito chave é a noção de “pontos bem definidos”, que

são aqueles que facilmente visíveis e recuperáveis no solo, como monumentos, cruzamentos de estradas, cantos de grandes prédios.

No entanto, em muitos casos, o que estamos tentando medir não é composto de objetos bem definidos, mas fenômenos com variação contínua no espaço e no tempo. O caso mais simples é a topografia, e neste caso o mesmo Manual nos diz: “*Noventa por cento dos pontos isolados de altitude, obtidos por interpolação de curvas-de-nível, quando testados no terreno, não deverão apresentar erro superior ao Padrão de Exatidão Cartográfica Altimétrico estabelecido.*” Evidentemente, trata-se de uma norma concebida para cartas em papel, mas seu uso de concepções exatas da realidade (os *relógios*) apresenta problemas. O conceito de “pontos isolados de altitude” não tem a mesma clareza semântica dos monumentos e estradas usados para avaliação da exatidão planimétrica. Para muitas aplicações, a qualidade de representação dos pontos isolados não é o fator determinante da qualidade da carta, mas sim sua capacidade de capturar a declividade do terreno. Adicionalmente, a norma não explicita como deve ser feita a interpolação de curvas de nível.

No caso de dados altimétricos num sistema de informações geográficas, o aplicação do conceito de *exatidão cartográfica* fica ainda mais complicada. Para começar, há diferentes alternativas de armazenamento de dados, como grades regulares e malhas triangulares. Diferentes algoritmos de interpolação irão produzir resultados distintos a partir dos mesmos dados. Mais ainda, com dados em formato digital, não precisamos estar limitados aos recortes e às escalas cartográficas. Em consequência, o conceito de *exatidão cartográfica*, concebido para produtos em papel, precisa de substancial revisão para ser aplicado ao caso de tratamento digital de dados altimétricos.

Consideremos agora tipos de dados ainda mais complicados que altimetria, como poluição, temperatura e fertilidade do solo. Nestes casos, medidas diferentes num mesmo ponto são decorrentes da variabilidade intrínseca do fenômeno subjacente; a visão de *nuvens* é freqüentemente a abordagem mais adequada, o que implica em buscar uma caracterização estatística. Para este tipo de dados, o conceito de *exatidão* deve ser substituído pelo de *incerteza*. A idéia de *incerteza* é buscar expressar a extensão de nossa ignorância, a partir de uma caracterização estatística dos dados geográficos.

Para entender melhor a idéia de incerteza, considere a situação de um mapa de fertilidade de solos, classificados em 4 classes: “alta”, “média”, “baixa” e “muito baixa”. Imagine que conseguimos obter, para cada ponto da região

considerada, uma estimativa da probabilidade associada a cada classe. Por exemplo, para um ponto, se estas probabilidades fossem, respectivamente,  $\{(70\%), (20\%), (5\%), (5\%)\}$ , poderíamos dizer que tal localização pode ter sua fertilidade classificada como “boa” com incerteza de 30%. Neste caso, a incerteza foi calculada com a soma das probabilidades das demais classes.

O leitor pode indagar: mas como será que eu consigo computar probabilidades, se o que eu tenho muitas vezes são apenas algumas poucas amostras de um único levantamento? Aí entra a primeira lei da geografia: *“todas as coisas se parecem, mas coisas mais próximas são mais parecidas que aquelas mais distantes”*. A idéia aqui é que, para fenômenos contínuos no espaço, como altimetria e fertilidade de solo, os vizinhos podem fornecer informação importante sobre cada localização de estudo e sobre a incerteza associada. Esta concepção pode ser formalizada pelas técnicas da geoestatística, principalmente através da krigeagem por indicação. Para uma descrição detalhada do uso destas técnicas, o leitor deve referir-se à tese de Doutorado de Carlos Felgueiras, do INPE: “Modelagem Ambiental com Tratamento de Incertezas em Sistemas de Informação Geográfica: O Paradigma Geoestatístico por Indicação” (disponível em [www.dpi.inpe.br/teses/carlos](http://www.dpi.inpe.br/teses/carlos)).

O uso da incerteza em dados ambientais é muito importante para distinguir a qualidade dos diferentes tipos de dados. Em uma tese de mestrado recente no INPE, analisamos dois diferentes tipos de dados ambientais, utilizados no Zoneamento Pedoclimático para Agricultura em Santa Catarina, desenvolvido pela EMBRAPA: dados climáticos obtidos a partir de estações meteorológicas e dados pedológicos, obtidos a partir de perfis e amostras de solo. Os resultados foram esclarecedores, mas não de todo surpreendentes: a incerteza associada aos dados climáticos é muito inferior à incerteza dos dados de solo (Vejam-se as figuras 1 e 2). Esta disparidade decorre de vários fatores: (a) uma menor variabilidade espacial dos fenômenos de temperatura e precipitação, principalmente se medidos a partir de séries históricas; (b) uma maior disparidade dos critérios de levantamento de solo, realizados por equipes distintas.

Em resumo, nem sempre é possível pensar nos dados geográficos como relógios, e quando tratamos com nuvens, é importante dispor de critérios que nos permitam aferir a qualidade e confiabilidade dos dados em que trabalhamos. Sempre é permitido sonhar: quem sabe um dia todos os dados geográficos terão medidas de qualidade adequadas. E não está tão longe....

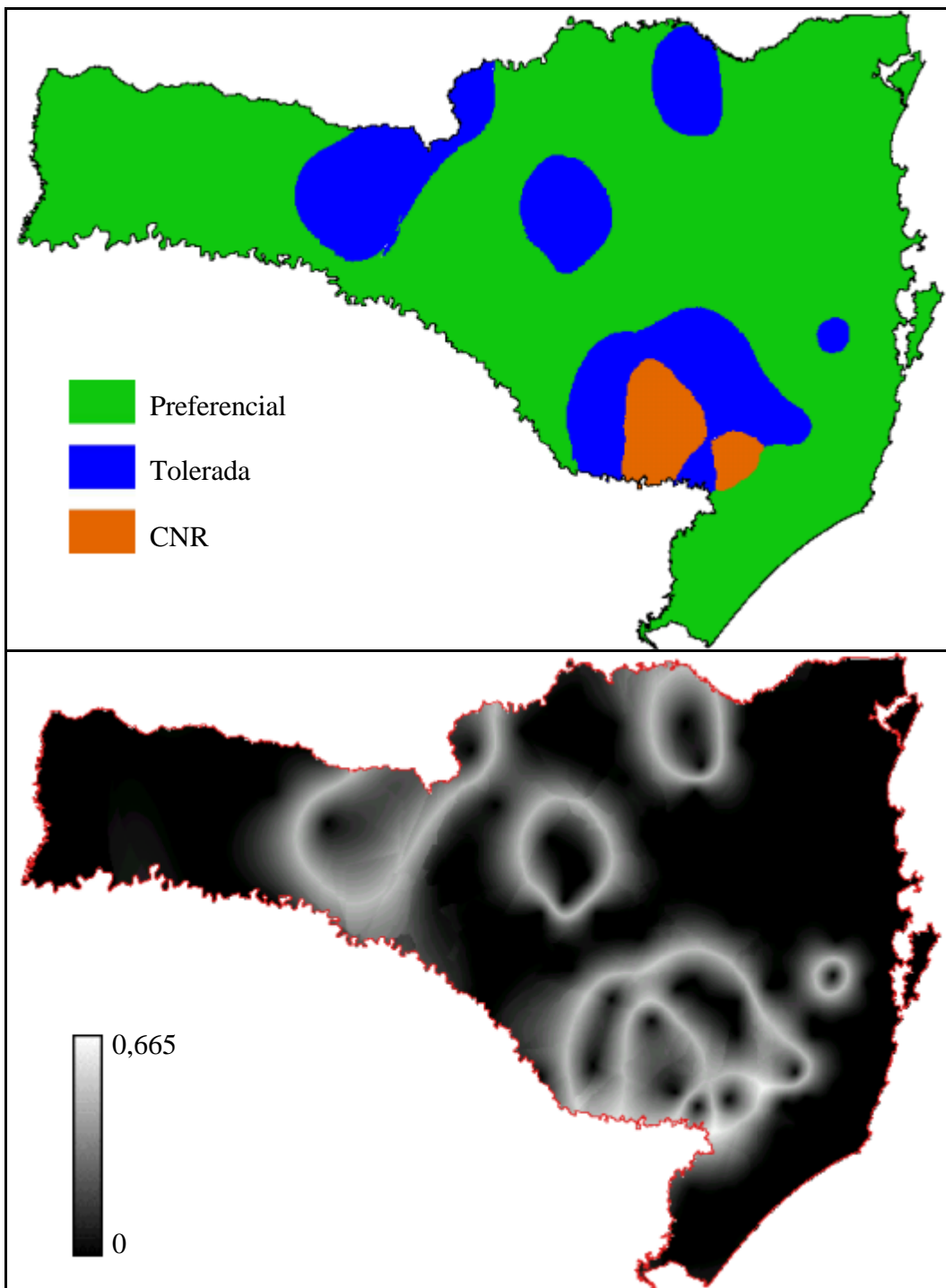


Figura 1 – Acima: Aptidão para o cultivo da soja em Santa Catarina, apenas com as restrições climáticas. Abaixo: Incerteza associada.

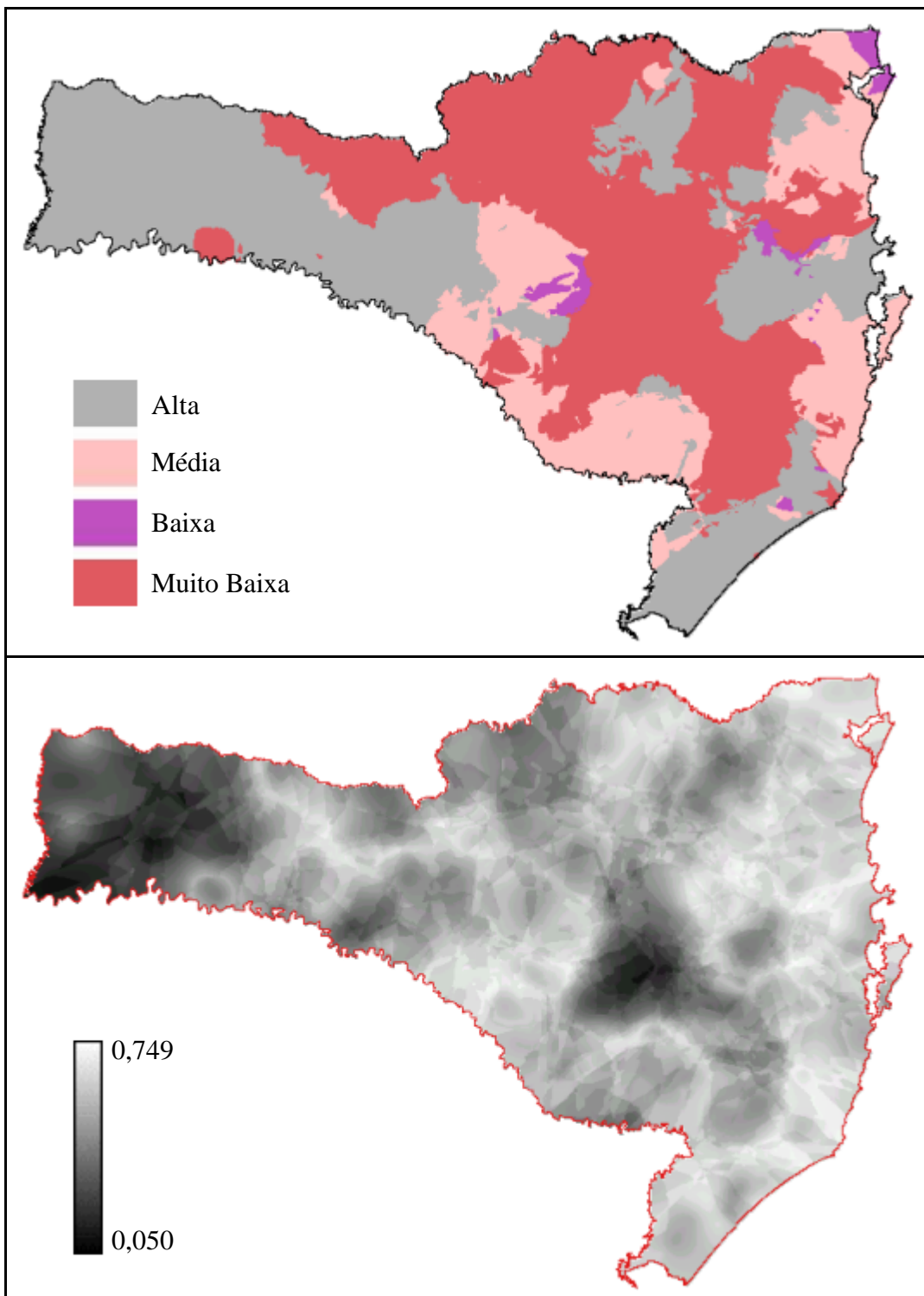


Figura 2 – Acima: Classes de Fertilidade de Solo para Cultivo de Soja em Santa Catarina; Abaixo: Incerteza associada.