



Ministério da
Ciência e Tecnologia



MÉTRICAS DE QUALIDADE DE MODELOS DE MUDANÇA DE USO E COBERTURA DA TERRA

Luiz Gustavo Diniz de Oliveira Veras

Tese de Doutorado ou Dissertação de Mestrado do Curso de Pós-Graduação em
Computação Aplicada, orientada pelo Dr. Gilberto Câmara.

URL do documento original:

<<http://urlib.net/xx/yy>>

INPE
São José dos Campos
AAAA

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3945-6923/6921

Fax: (012) 3945-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):**Presidente:**

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr^a Regina Célia dos Santos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr. Horácio Hideki Yanasse - Centro de Tecnologias Especiais (CTE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Vivéca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



Ministério da
Ciência e Tecnologia



MÉTRICAS DE QUALIDADE DE MODELOS DE MUDANÇA DE USO E COBERTURA DA TERRA

Luiz Gustavo Diniz de Oliveira Veras

Tese de Doutorado ou Dissertação de Mestrado do Curso de Pós-Graduação em
Computação Aplicada, orientada pelo Dr. Gilberto Câmara.

URL do documento original:

<<http://urlib.net/xx/yy>>

INPE
São José dos Campos
AAAA

RESUMO

No estudo de processos de mudança de cobertura da terra, métricas de qualidade de modelos, também chamadas de goodness-of-fit, são comumente usadas para se estimar a precisão de um modelo computacional. Coeficientes que representem o grau de correspondência entre dois mapas são determinados a partir da comparação entre o mapa de uso da terra real e o mapa resultante de uma simulação. Para se estimar estes coeficientes, procedimentos usam uma variedade de resoluções espaciais para se considerar erros de localização da alocação de mudança de uso. Entretanto, estas abordagens consideram e comparam apenas dois mapas referentes a um único instante no tempo. Este trabalho possui a hipótese que um único tempo é insuficiente para se medir a qualidade de um modelo de uso e cobertura da terra, buscando evoluir na definição de métricas espaço-temporais para se estimar a qualidade de modelos de mudança de uso e cobertura da terra.

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Separação dos dados de calibração e validação por meio do tempo.	3
2.2 Separação dos dados de calibração e validação por meio do espaço.	4
2.3 Extrapolação linear em um procedimento de calibração e validação.	5
2.4 Exemplo do procedimento de ajuste de múltiplas resoluções para duas cenas 10x10 com quatro categorias.	7
2.5 Exemplo do procedimento de ajuste de múltiplas resoluções para duas cenas 10x10 com quatro categorias.	8
2.6 Equações matemáticas para cinco medidas definidas pela combinação de informações sobre quantidade e localidade.	8
2.7 Comparação de porcentagem correta estimada entre os modelos Geomod, Nulo e Randômico.	13
3.1 Possível estrutura da métrica de qualidade temporal aplicada a 5 instan- tes no tempo.	15
3.2 Dados de desmatamento da Amazônia de 2009.	17

LISTA DE TABELAS

	<u>Pág.</u>
2.1 Informações geradas combinando informação de quantidade com informação de localidade.	11
4.1 Cronograma de execução do trabalho	19

SUMÁRIO

	<u>Pág.</u>
1 INTRODUCAO	1
2 TÉCNICA DE CALIBRAÇÃO E VALIDAÇÃO	3
2.1 Métricas de goodnes of fit	5
2.1.1 Método de Múltiplas resoluções	5
2.1.2 Métricas de estimativa de erro	9
2.2 Validação	11
2.2.1 Resultados da validação	12
2.2.2 Resolução Nula	14
3 METODOLOGIA	15
3.1 Estudo de Caso	16
3.2 Implementação	17
4 CRONOGRAMA	19
REFERÊNCIAS BIBLIOGRÁFICAS	21

1 INTRODUCAO

Modelagem computacional é um método que consiste em representar, por meio de descrições formais e não ambíguas, diversos elementos do mundo real em computadores. Ela consiste em uma simplificação - menor, menos detalhada, menos complexa, ou com todas estas características unidas - de algum fenômeno em específico (GILBERT; TROITZSCH, 2005). Seu exercício permite estudar mais profundamente os estados, comportamentos, relações e possíveis feedbacks envolvendo um ou mais processos ao longo do tempo (BLILIE, 2007).

Processos de mudança de uso e cobertura da terra agregam aspectos tanto das ciências naturais quanto das ciências humanas. Os estudos destes processos estão relacionados à possível influência destes sobre o sistema terrestre em escala local, regional ou até global, podendo ter impactos na biodiversidade, nos recursos hídricos ou até no clima, alterando a capacidade de suprirem as necessidades humanas (LAMBIN et al., 2006). Existem diversos tópicos de estudo envolvendo processos de uso e cobertura da terra. Por exemplo, de intensificação da agricultura, de crescimento urbano e de desmatamento, como no caso dos utilizados nos estudos sobre a Amazônia (AGUIAR et al., 2007; LAURANCE et al., 2001).

Devido à dependência espaço-temporal desses processos e à grande variedade de fatores que podem influenciar na alteração dos padrões de uso da terra, o seu entendimento é bastante complexo, fazendo-se necessário o uso de modelagem computacional. Este tipo de modelagem permite explorar o funcionamento desses processos através do uso de diversas técnicas tais como modelagem baseada em agentes ou modelos de alocação top-down. A partir destes modelos, cenários hipotéticos podem ser elaborados para o estudo de configurações de uso da terra alternativos. (VERBURG et al., 2006).

Os resultados das simulações de processos de uso e cobertura da terra são mapas que descrevem as alterações ocorridas sobre a superfície terrestre em um período de tempo. Para se medir a qualidade dessas informações geradas é necessário estimar os erros de alocação das mudanças ocorridas na simulação com os dados reais. Diferentes tipos de erro podem ser estimados, tais como erros por localização ou por quantidade de demanda final. Métricas de qualidade de ajuste, também chamadas de *goodness-of-fit*, podem ser usadas para comparar estes mapas, através da sumarização dos erros espaciais em coeficientes que expressam o grau de ajuste dos resultados

de uma simulação (COSTANZA, 1989). Esses coeficientes podem ser utilizados tanto no procedimento de calibração quanto no de validação dos modelos (PONTIUS et al., 2004).

Alguns métodos de métricas de qualidade aplicam medidas estatísticas somente aos dados quantitativos de uso e cobertura da terra correspondente entre um mapa de simulação e um de referência, acarretando em erros e mascarando dados referentes aos padrões espaciais e temporais existentes (WILLMOTT, 1981; NASH; SUTCLIFFE, 1970). Extensões destas abordagens usam uma variedade de resoluções espaciais para se considerar erros de localização da alocação de mudança de uso e cobertura da terra (COSTANZA, 1989; PONTIUS et al., 2001). Porém, estes métodos consideram e comparam apenas dois mapas referentes a um único instante no tempo, desprezando quaisquer erros de alocação no tempo.

Atualmente, uma grande quantidade de dados espaciais, relativos ao uso e cobertura da terra em diferentes instantes do tempo, se encontra disponível. Essas informações são de grande importância para que procedimentos de calibração e validação de modelos possam usar dados referentes a mais de um instante de tempo para calcular os seus ajustes. Dessa maneira, erros de alocação podem ser detectados não somente no espaço, mas também no tempo. Este trabalho possui a hipótese que um único tempo é insuficiente para se medir a qualidade de um modelo dinâmico espacial, buscando evoluir na definição de métricas espaço-temporais para se estimar a qualidade de modelos de mudança de uso da terra. Portanto, é proposto o estudo de uma expansão do método de múltiplas resoluções, buscando uma abordagem de qualidade de ajuste espaço-temporal, onde os resultados gerados por um modelo sejam avaliados para diversos tempos distintos dentro do período de simulação.

O trabalho está estruturado da seguinte maneira. No capítulo 2 alguns métodos existentes na literatura são apresentados. No capítulo 3 é proposta uma nova metodologia para a implementação e análise dos métodos. No capítulo 4 um cronograma do trabalho é apresentado.

2 TÉCNICA DE CALIBRAÇÃO E VALIDAÇÃO

Procedimentos de calibração e validação visam melhorar e avaliar o desempenho das simulações de modelos de uso e cobertura da terra. Esses procedimentos aumentam a capacidade de compreender o funcionamento desses modelos. Assim, eles permitem, fornecer maneiras de como melhorá-los, definir um domínio apropriado para uma aplicação particular do modelo, apresentar resultados em uma adequada resolução espacial, validar as estimativas do modelo e definir metas que priorizem as questões mais importantes no estudo dos sistemas de uso e cobertura da terra. (PONTIUS et al., 2004).

Na maioria dos procedimentos de calibração e validação, a simulação de modelos de uso e cobertura da terra trabalha com uma paisagem em um ponto inicial (t_1) e então estima a mudança ocorrida para um ponto subsequente no tempo (t_2). Esse dado apresentado em t_2 é o mapa de simulação do modelo, o qual é usualmente comparado com um mapa de referência (i.e., mapa gerado a partir de dados reais) de t_2 , a fim de avaliar o desempenho da simulação. Se o mapa de simulação se mostrar similar ao mapa de referência, então se conclui que a simulação obteve um bom desempenho. Quando se está avaliando o poder de estimativa de um modelo, deve haver uma clara distinção entre validação e calibração. Caso contrário, o modelo pode apresentar erroneamente a demanda dos usos da terra para o tempo final da simulação. Uma maneira eficiente de melhorar o desempenho da simulação é conhecer como distinguir entre esses dois processos.

Calibração "é a estimativa e ajuste dos parâmetros e restrições do modelo para melhorar a correspondência entre o seu resultado e um conjunto de dados" enquanto validação "é uma demonstração de que um modelo dentro de seu domínio de aplicabilidade possui um satisfatório alcance de acurácia consistente com a aplicação pretendida do modelo" (RYKIEL, 1996). Os conjuntos de dados usados na calibração e na validação não devem ser os mesmos, existindo diferentes maneiras de separá-los.

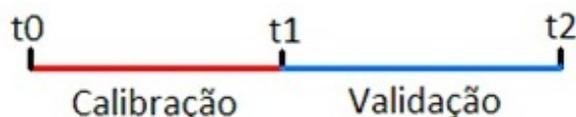


Figura 2.1 - Separação dos dados de calibração e validação por meio do tempo.

A separação por meio do tempo é a mais comumente utilizada (PONTIUS et al., 2004). Se o objetivo do modelo é obter a mudança da paisagem após o tempo t_1 , então qualquer informação em t_1 ou anterior é válida para ser usada na calibração. Um método típico desta abordagem de calibração é a regressão estatística na mudança entre o tempo t_0 e o tempo t_1 , sendo seu resultado os parâmetros de estimativas ajustadas. O importante é que nenhum dado após o tempo t_1 deve ser usado. Os dados após esse período são utilizados na validação, comparando o mapa de previsão de t_2 com o mapa de referência de t_2 .

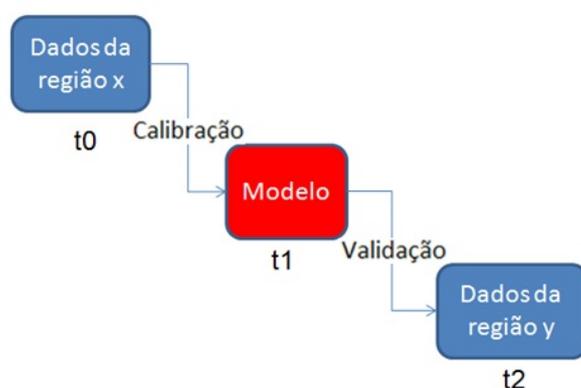


Figura 2.2 - Separação dos dados de calibração e validação por meio do espaço.

A separação por meio do espaço é outra maneira comum para diferenciar a informação para calibração da informação para validação. Nesse procedimento o método usa dados de um local de estudo (região x) para ajustar os parâmetros, e então o modelo ajustado é aplicado a um diferente local (região y) para simular a mudança (PONTIUS et al., 2004).

Um exemplo de como são aplicadas as técnicas de calibração e validação é apresentada na Figura 1. Os dados de calibração correspondem aos tempos $t_0 = 1971$ e $t_1 = 1985$, os quais calibram o modelo por meio do ajuste dos componentes do mesmo. Por meio de uma simples extrapolação linear, é estimada a quantidade de demanda a ser alocada entre os tempos $t_1 = 1985$ e $t_2 = 1999$, o qual representa o período de validação.

Diversas métricas de qualidade podem ser utilizadas para comparação dos resultados dos procedimentos de calibração e validação. Os tópicos descritos a seguir mostram

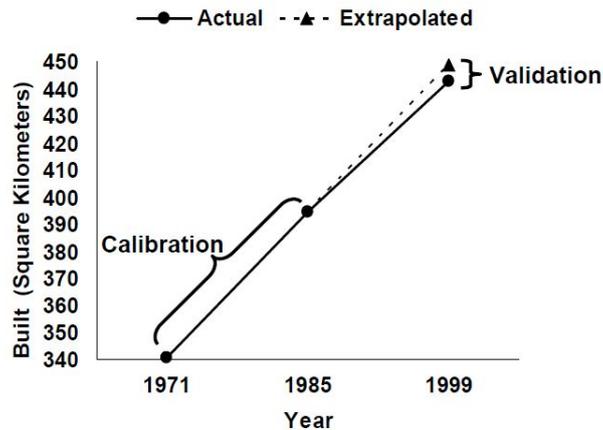


Figura 2.3 - Extrapolação linear em um procedimento de calibração e validação.
 Fonte: (PONTIUS et al., 2004)

algumas dessas métricas.

2.1 Métricas de goodness of fit

Métricas de qualidade avaliam o desempenho das simulações de modelos de uso e cobertura da terra, por meio de um conjunto de métodos aplicados sobre dados reais e estimados. Essas métricas sumarizam a discrepância entre um conjunto de valores observados e um conjunto de valores estimados. Elas podem ainda quantificar erros e identificar semelhanças entre padrões de dados.

Técnicas de calibração e validação, por meio dessas métricas, avaliam a acurácia de um modelo de uso e cobertura da terra em relação a um conjunto de dados reais. Duas métricas aplicadas ao reconhecimento de padrões e de erros de alocação espacial desses modelos serão apresentadas nos próximos tópicos.

2.1.1 Método de Múltiplas resoluções

O Método de Múltiplas Resoluções consiste em avaliar o grau de correspondência entre dois mapas em uma variedade de escalas e interpretar como a percepção dos padrões de usos e cobertura da terra são alterados com a mudança da resolução espacial (Costanza, 1989). O resultado da aplicação desse método é um conjunto de coeficientes que representam o ajuste das diferenças entre dois modelos de uso e cobertura da terra, avaliando assim sua eficiência.

Para propósitos relativamente simples de comparação entre modelos de usos e cobertura da terra, existem medidas estatísticas bem estabelecidas de métricas de qualidade. Entretanto, os modelos de uso e cobertura da terra necessitam de uma quantificação do grau de correspondência e similaridade entre os complexos padrões espaciais e temporais. A maioria dos testes estatísticos foca no desvio padrão das distribuições aleatórias, tendendo a errar ou mascarar detalhes desses padrões existentes nos dados. Isso pode gerar informação sobre se o número total de *pixels* em cada categoria foi significativamente diferente entre dois mapas, mas não se existe uma significativa diferença em seus padrões. Sendo assim, os mapas podem conter o mesmo número total de *pixels* em cada categoria, mas o arranjo entre eles pode ser totalmente diferente (??). O Método de Múltiplas Resoluções soluciona este problema utilizando janelas que aumentam gradualmente a resolução de comparação entre os mapas, chamadas janelas de amostragem.

A cada uma dessas janelas de amostragem são aplicados métodos estatísticos que agregam os dados à medida que a resolução diminui. O método diminui a resolução de comparação em detrimento do aumento do tamanho da janela de amostragem, que é deslocada sobre os mapas para as quais o ajuste será calculado. Fazendo isso, o padrão espacial se torna aparente.

Existem muitas maneiras de calcular o ajuste em uma janela de um tamanho particular (resolução), dependendo da natureza dos dados. A abordagem de múltiplas resoluções não depende do cálculo aplicado para estimar o ajuste, podendo ser atribuído consistentemente sobre todas as resoluções. Se dados numéricos estão disponíveis, um coeficiente padrão de correlação (R^2) pode ser usado. No caso de dados categóricos, outros procedimentos devem ser utilizados.

A (??) ilustra um exemplo com as categorias 1, 2 e 3, representadas em duas cenas de resolução 10x10. Olhando para as cenas é possível verificar que seus padrões são semelhantes, mas não idênticos. Por exemplo, o bloco de categoria 3 possui maior área na cena 1 do que na cena 2. Humanos conseguem reconhecer esses padrões de similaridade por demonstrarem possuir uma habilidade intuitiva para isso. Com base nisso, o problema consiste em sistematizar esta capacidade de tal maneira que os algoritmos computacionais possam imitá-la (??).

Para uma janela de amostragem de tamanho 1, uma medida conveniente para ajuste de dados categóricos seria a proporção de células correspondentes (80

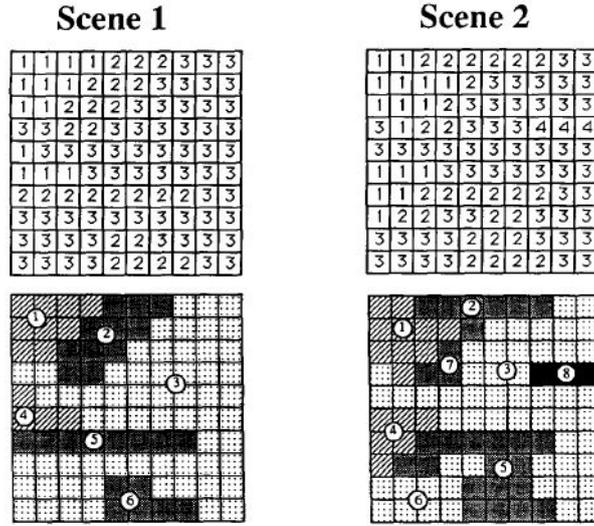


Figura 2.4 - Exemplo do procedimento de ajuste de múltiplas resoluções para duas cenas 10x10 com quatro categorias.
 Fonte: (??)

$$F_w = \frac{\sum_{s=1}^{t_w} \left[1 - \frac{\sum_{i=1}^p |a_{1i} - a_{2i}|}{2w^2} \right]_s}{t_w}$$

onde w a dimensão de um lado da janela de amostragem, F_w é o ajuste para a janela de amostragem de tamanho w , a_{ki} o número de células da categoria i na cena k na janela de amostragem, p o número de diferentes categorias na janela de amostragem, s a janela de amostragem de dimensão w por w a qual desliza através da cena uma célula por vez, e t_w o número total de janelas de amostragem na cena para uma janela de tamanho w (??).

A Figura 2 mostra um exemplo do cálculo dos ajustes para cada resolução espacial. O algoritmo diminui a resolução de comparação pelo aumento do tamanho de uma janela de amostragem que desliza sobre a cena para o qual o ajuste é calculado (COSTANZA, 1989). Os padrões de reconhecimento são dependentes da escala, o que significa que eles dependem da resolução e da extensão (ALLEN; STARR, 1982). Uma visão aproximada das cenas possui uma maior resolução, ao contrario da extensão, que é comprometida, prejudicando o reconhecimento do padrão. Já com uma maior extensão, a resolução é sacrificada, mas o padrão se torna aparente.

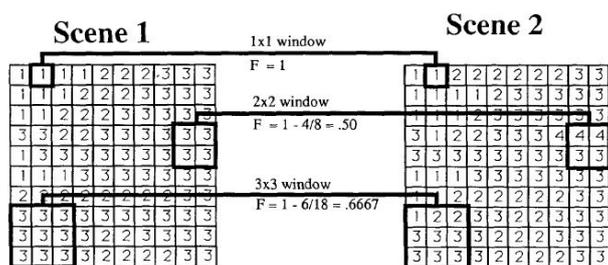


Figura 2.5 - Exemplo do procedimento de ajuste de múltiplas resoluções para duas cenas 10x10 com quatro categorias.

Fonte: (??)

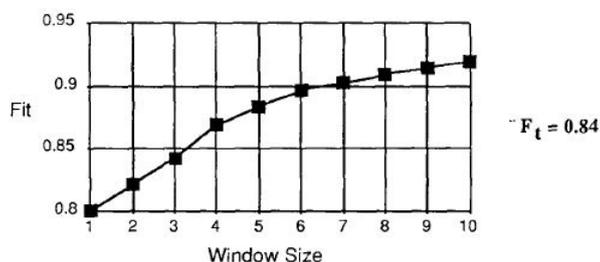


Figura 2.6 - Equações matemáticas para cinco medidas definidas pela combinação de informações sobre quantidade e localidade.

Fonte: (PONTIUS et al., 2004)

A Figura 3 apresenta um gráfico das diversas janelas de amostragem (w) em função de seus respectivos ajustes. O gráfico mostra que à medida que o tamanho da janela aumenta o ajuste cresce consideravelmente. Isto indica que há uma correspondência muito boa entre as duas cenas, embora o ajuste inicial para a janela de amostragem de tamanho 1 seja baixa. Do mesmo modo, se o gráfico inicia com um ajuste relativamente baixo e se mantém contínuo para os demais tamanhos de janela, então haverá baixa correspondência entre os padrões espaciais.

Para determinar um índice geral de ajuste entre dois mapas, as informações do gráfico da Figura 3 devem ser resumidas. Costanza (1988) propõe uma média ponderada dos ajustes dos diferentes tamanhos de janela que permita maior peso à tamanhos de janelas menores, embora janelas de tamanhos maiores não sejam totalmente ignoradas. Para este propósito pode-se usar a seguinte fórmula:

$$F_t = \frac{\sum_{w=1}^n F_w e^{-k(w-1)}}{\sum_{w=1}^n e^{-k(w-1)}}$$

onde F_t é a média ponderada dos ajuste sobre todos os tamanhos de janela, f_w o ajuste para as janelas de amostragem de dimensão linear w , k é uma constante, e w a dimensão linear de uma janela de amostragem. Esta fórmula fornece exponencialmente uma perda de peso para o ajuste em resolução mais baixa. O valor de k representa o grau de participação de uma janela de amostragem na determinação do índice geral, sendo o valor de k proporcional ao tamanho da janela.

2.1.2 Métricas de estimativa de erro

Com a computação das fontes de erros na técnica de validação, o modelador poderá melhorar aspectos do modelo que possuem grande potencial para benefícios do mesmo. Procedimentos que fornecem apenas um parâmetro de ajuste, como a quantidade de porcentagem correta, falham em prover informações relativas à como melhorar o modelo, pois não existe uma estimativa de erro oriunda de várias fontes (PONTIUS et al., 2004). Com base nisso, esta métrica estima diversos tipos de erros calculando separadamente a informação de quantidade da informação de localidade dos dados espaciais.

A figura 5 fornece as expressões matemáticas usadas na computação dos componentes de correspondência e divergência durante a comparação de dois mapas. Cada equação compara a correspondência entre os *pixels* de dois mapas, os quais são denotados de R para o mapa de referência e S para o mapa simulado. Cada expressão está em uma coluna e linha, de acordo com o nível de informação de quantidade e localidade, respectivamente. A informação de quantidade é representada por letras em negrito e a informação de localidade é representada por letras maiúsculas. Para a informação de quantidade, o n significa que não existe informação, o m significa informação média e o p informação perfeita. Já em relação à informação de localidade, o N(x) significa que não há nenhuma informação, M(x) significa informação média e P(x) significa informação perfeita. As equações na figura x possuem a seguinte notação:

- j é o índice da categoria;
- J é o número de categorias;

- n é o índice para *pixels*;
- g é a resolução como um múltiplo da largura do lado de um *pixel* da resolução inicial;
- N_g é o número de *pixels* no mapa na resolução g ;
- W_{gn} é o peso do *pixel* n na resolução g ;
- R_{gnj} é a proporção da categoria j no *pixel* n na resolução g do mapa de referência;
- S_{gnj} é a proporção da categoria j no *pixel* n na resolução g no mapa de simulação.

Em algumas expressões, o símbolo ”.” substitui o sobrescrito n . Isto indica que o termo é a média global n como mostrado nas equações 1 e 2.

$$R_{g.j} = \frac{\sum_{n=1}^{N_g} (W_{gn} R_{gnj})}{\sum_{n=1}^{N_g} W_{gn}}$$

$$S_{g.j} = \frac{\sum_{n=1}^{N_g} (W_{gn} S_{gnj})}{\sum_{n=1}^{N_g} W_{gn}}$$

A figura 5 mostra uma sequência das expressões, iniciando no canto inferior esquerdo, subindo pela coluna central e chegando ao canto superior direito. À medida que as equações são apresentadas por essa sequência, um dado referente à combinação da informação de quantidade e da informação de localidade correspondente é fornecido. A Tabela 1 descreve as informações geradas ao cruzar a informação de quantidade com a informação de localidade.

Cada expressão denota um grau de correspondência entre o mapa de previsão e um mapa ajustado, portanto usualmente $0 < N(n) < N(m) < M(m) < P(m) < P(p) = 1$. A diferença entre cada expressão subsequente fornece os seguintes componentes de correspondência ou divergência:

- $N(n)$ é a correspondência devido à aleatoriedade da alocação das categorias;

Tabela 2.1 - Informações geradas combinando informação de quantidade com informação de localidade.

	n	m	p
P(x)		P(m): Correspondência entre o mapa de referência e um mapa ajustado no qual a localização dos <i>pixels</i> do mapa de simulação são rearranjados no espaço para se adequar o máximo possível aos <i>pixels</i> do mapa de referência.	P(p): Correspondência entre o mapa de referência e o mapa que tem sido perfeitamente ajustado em termos de quantidade e localidade dos <i>pixels</i> , portanto P(p) será sempre igual a 1.
M(x)		M(m): Correspondência entre o mapa de referência e o mapa de simulação não ajustado.	
N(x)	N(n): Correspondência entre o mapa de referência e o mapa no qual todo <i>pixel</i> é idêntico e possui participação parcial de $1/J$ para cada categoria J.	N(m): Correspondência entre o mapa de referência e o mapa onde todo <i>pixel</i> é idêntico e possui participação igual para a proporção gerada para cada categoria J.	

- $N(m) - N(n)$ é a correspondência devido à quantidade gerada;
- $M(m) - N(n)$ é a correspondência devido à localidade gerada;
- $P(m) - M(m)$ é a divergência devido à localidade gerada;
- $P(p) - P(m)$ é a divergência devido à quantidade gerada.

A expressão do meio na segunda coluna e segunda linha da Figura x é a proporção correta baseada em uma comparação *pixel-a-pixel* do mapa de referência de t2 e o mapa de simulação de t2. As outras quatro expressões geram a proporção correta entre os dois mapas após o mapa de simulação ter sido ajustado para vários níveis de informação de quantidade e/ou localidade.

2.2 Validação

Um bom procedimento de validação deve ser empregado para que as simulações de uso e cobertura da terra apresentem resultados satisfatórios. A falta de rigor desses métodos pode fornecer uma falsa sensação de segurança sobre o poder de previsão

de um modelo. Portanto, alguns requisitos devem ser definidos em qualquer técnica de validação.

Pontius et al. (2004) sugerem que pelo menos três simulações devam ser comparadas no procedimento de validação (a) de um modelo nulo, (b) de um modelo aleatório, (c) de um modelo de uso e cobertura da terra. O modelo nulo admite persistência das categorias, gerando um mapa em t2 igual ao mapa de t1. Por sua vez, o modelo aleatório distribui de forma randômica no espaço, as mudanças estimadas pelo modelo de uso e cobertura da terra. É importante comparar o modelo de uso e cobertura da terra com os modelos nulo e aleatório, a fim de avaliar o potencial de simulação que aquele gera (PONTIUS et al., 2004).

Durante a validação, é importante haver uma metodologia objetiva em comparar o mapa de previsão e o mapa de referência, a fim de avaliar o nível de correspondência entre os dois. Não existe um critério geral para analisar a qualidade de ajuste da validação (RYKIEL, 1996), e nem deve existir. Cada modelo possui um propósito diferente, e o critério de validação deverá ser relacionado a cada objetivo. Mas, após os propósitos serem definidos, existem diversos critérios que são relevantes a uma aplicação particular, devendo-se usar aquele mais apto a melhorar o modelo. Uma técnica de validação é mais útil quando:

- Compara o modelo com um modelo nulo;
- Compara o modelo com um modelo aleatório;
- Estima a fonte de erro;
- Desempenha a análise em múltiplas escalas.

A seguir é apresentado o resultado gerado pela técnica de validação, aplicada a um modelo nulo, aleatório e de uso e cobertura da terra. Uma análise do desempenho dos modelos também é apresentada.

2.2.1 Resultados da validação

(PONTIUS et al., 2004) avalia a eficiência da técnica de validação, aplicando os métodos sobre um tipo de modelo de uso e cobertura da terra chamado Geomod (PONTIUS et al., 2001; PONTIUS; MALANSONA, 2005). As técnicas de validação são aplicadas

também aos modelos nulo e aleatório, e após isso, uma análise dos resultados é realizada.

A Figura 10 ilustra a comparação estatística entre os modelos aleatório, nulo e Geomod em um procedimento de validação. A porcentagem de correspondência é apresentada em contraste com diversas escalas analisadas pelo método de múltiplas resoluções.

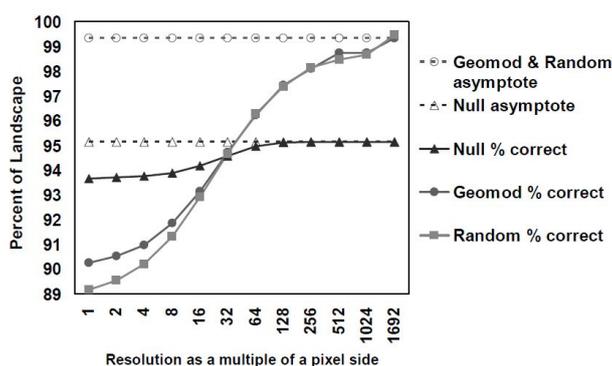


Figura 2.7 - Comparação de porcentagem correta estimada entre os modelos Geomod, Nulo e Randômico.

Fonte: (PONTIUS et al., 2004)

A linha horizontal do topo do gráfico mostrado na figura 10 corresponde à medida $P(m)$ para o modelo Geomod e o randômico. Essa medida apresenta a divergência devido à quantidade gerada, como mostrado na figura 5. As curvas de porcentagem correta do modelo Geomod e do randômico tendem a se aproximar deste limite à medida que a resolução se torna mais grosseira.

Já a linha horizontal mais abaixo é referente à medida $P(m)$ do modelo nulo, a qual mostra a divergência devido à quantidade de mudança. A curva de porcentagem correta do modelo nulo se aproxima dessa linha à medida que a resolução aumenta.

O resultado mostrou que o modelo nulo se comportou melhor do que o modelo Geomod em resoluções mais finas do que 32 vezes o tamanho da resolução inicial. Isso não é necessariamente um indicador alarmante. Isso apenas indica que a resolução do dado original é mais fina do que a da habilidade do Geomod estimar as mudanças de cobertura da terra. Se os resultados apresentados são bons, isso irá depender do

objetivo que se deseja alcançar (PONTIUS et al., 2004).

2.2.2 Resolução Nula

É muito comum a porcentagem correta de um modelo de uso e cobertura da terra ser menor do que a porcentagem correta de um modelo nulo, como também é comum que a divergência devida à demanda por mudança de categoria (quantidade estimada da mudança de um dado tipo de uso ocorrida entre dois tempos distintos) de um modelo de simulação seja menor do que a divergência da demanda de um modelo nulo. Quando estas duas condições são satisfeitas, então haverá uma resolução em que as curvas de porcentagem de acerto se cruzam, indicando que a acurácia do modelo de simulação corresponde à acurácia do modelo nulo (PONTIUS et al., 2004).

A resolução na qual as curvas de porcentagem de acerto se cruzam chama-se resolução nula. No exemplo apresentado, como mostrado na figura z, a resolução nula ocorre em um ponto com resolução 32 vezes maior do que a do dado original. Quando nas resoluções mais finas que a da resolução nula, o modelo de simulação possui acurácia menor do que a do modelo nulo. Já nas resoluções mais grosseiras que a da resolução nula ocorre o contrário, o modelo de simulação possui acurácia maior do que a do modelo nulo.

3 METODOLOGIA

A técnica de múltiplas resoluções introduzida por ??) permite expressar por meio de coeficientes a similaridade entre componentes espaciais de dois mapas. Mas essa abordagem aplica as métricas de qualidade a dois mapas em apenas um instante no tempo, e portanto não considera o grau de correspondência temporal. A proposta deste trabalho é estender o método de múltiplas resoluções apresentado por ??), estimando o ajuste para vários tempos, a fim de detectar erros de alocação não somente em escala espacial, mas também temporal.

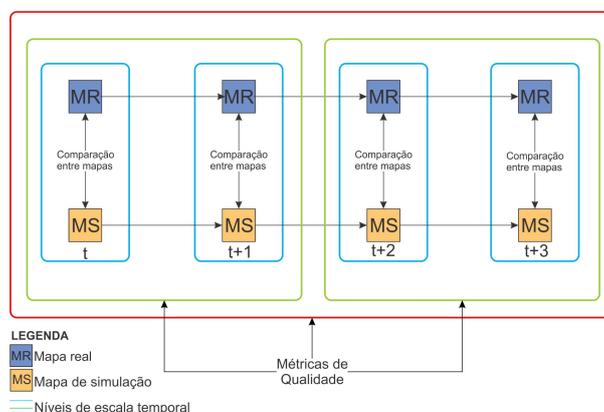


Figura 3.1 - Possível estrutura da métrica de qualidade temporal aplicada a 5 instantes no tempo.

Pretende-se implementar uma estrutura semelhante a da Figura 1, baseada na metodologia apresentada por Costanza (1989). Do mesmo modo que ocorre na Técnica de Múltiplas Resoluções Espaciais, pretende-se desenvolver um modelo que seja capaz de manipular múltiplas resoluções temporais, agregando os dados em janelas de amostragem.

O grau de acurácia temporal do modelo pode ser obtido comparando mapas reais de tempos diferentes e mapas de simulação também de tempos subsequentes, acarretando em mapas que expressam as diferenças de uso ou cobertura da terra entre os tempos. Medidas estatísticas de métricas de qualidade tradicionais seriam aplicadas sobre esses mapas. A sumarização dos ajustes gerados por essas métricas resultaria na métrica de qualidade temporal. Serão estudadas também as consequências desta métrica espaço-temporal sobre os modelos nulo e aleatório propostos por Pontius et

al. (2001) para estudar a qualidade do modelo.

3.1 Estudo de Caso

Como avaliação analítica, será realizada uma aplicação da nova métrica desenvolvida, que consiste em comparar o comportamento de dois modelos de desmatamento na Amazônia, um apresentado por Laurance et al. (2001) e outro por Aguiar et al. (2007), a fim de verificar a eficiência do método.

Laurance et al. (2001) apresenta dois modelos com diferentes cenários, um otimista e outro não-otimista, da degradação na Amazônia. Com base em diversos processos como, por exemplo, desmatamento, extração madeireira, mineração, rodovias e estradas, rios navegáveis, vulnerabilidade a queimadas, áreas protegidas, e existência e planejamento de projetos de infraestrutura, esses modelos estimam a distribuição espacial do desmatamento em diferentes níveis.

O modelo apresentado por Aguiar et al. (2007) é uma extensão do arcabouço de modelagem dinâmica CLUE para explorar como fatores alternativos, políticas públicas e condições de mercado que influenciam no processo de ocupação da Amazônia. Modelos de regressão linear são utilizados por esse modelo na determinação de fatores determinantes do processo estudado. Diversos aspectos do modelo são explorados, permitindo descrever uma análise da influência de diferentes fatores na dinâmica das novas fronteiras na Amazônia, possíveis impactos políticas públicas, e aumento e diminuição da demanda.

Os resultados mostrados pelos modelos de Laurance et al. (2001) sugerem que haverá uma grande área florestal a ser devastada dentro de 20 anos na Amazônia. Esse cenário é atribuído principalmente ao desenvolvimento de novas rotas de acesso a regiões em desenvolvimento da infraestrutura e da economia. Mas segundo Aguiar et al. (2007), a presença de estradas é um importante fator no surgimento de áreas de desmatamento, mas que só é possível entender este processo combinando as rodovias com outras variáveis relacionadas à organização do sistema de produtividade em diferentes regiões, tal como as condições ambientais e ao acesso ao mercado local e nacional.

Hoje já se encontram disponíveis dados do desmatamento na Amazônia para o ano de 2009. Com esses dados serão estudados o modelo de Laurance et al. (2001) e um modelo baseado nos estudos de Aguiar et al. (2007), comparando-os utilizando a

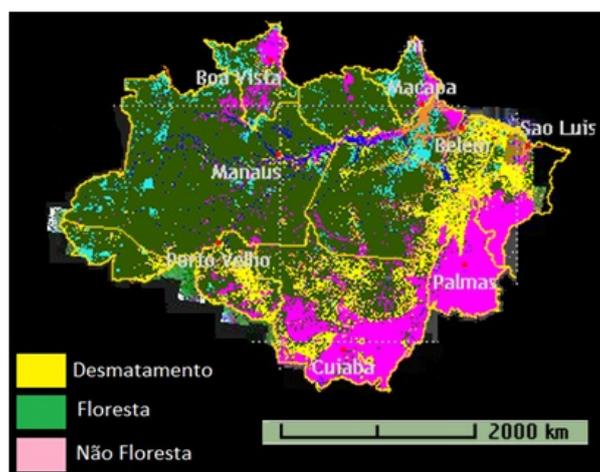


Figura 3.2 - Dados de desmatamento da Amazônia de 2009.
 Fonte: PRODES, 2009

métrica de qualidade temporal para determinar o grau de correspondência com os mapas de referência do desmatamento da Amazônia.

3.2 Implementação

Um arcabouço contendo diversas métricas de qualidade, inclusive a nova métrica temporal, serão implementados utilizando-se o framework de modelagem dinâmica TerraME, desenvolvido pelos Laboratório de Modelagem e Simulação de Sistemas da Terra - TerraLab, da Universidade Federal de Ouro Preto, com a Divisão de Processamento de Imagens - DPI e com o Centro de Ciência do Sistema Terrestre - CCST, do Instituto Nacional de Pesquisas Espaciais - INPE. Os modelos serão desenvolvidos em Lua, a linguagem suportada pelo TerraME.

O TerraME é um ambiente de modelagem que implementa o modelo de autômatos celulares aninhados que permite o desenvolvimento de modelos espacialmente explícitos onde várias resoluções e extensões temporais, espaciais e analíticas são levadas em consideração (CARNEIRO, 2006). A linguagem de modelagem do TerraME é uma extensão da linguagem LUA. O TerraME é acoplado com a biblioteca SIG (Sistema de Informação Geográfica) de código fonte aberto Terralib (??), o qual fornece serviços para o armazenamento de entrada e saída dos dados do modelo. Os modelos de Laurance et al. (2001) e um modelo aleatório do desmatamento da Amazônia, serão implementados em um framework de modelagem que combina

diferentes componentes de modelos de uso e cobertura da terra (PIMENTA, 2010), denominado LUCC-ME. O modelo de Aguiar et al. (2007) já está implementado nesse framework.

Esse framework é composto pelos módulos de mudança de demanda, potencial e alocação. O módulo de mudança de demanda é responsável por calcular a quantidade de mudança para cada uso da terra. O módulo de mudança de potencial é caracterizado por traduzir de que forma as transições de uso ocorrem (e.g. através de probabilidades, propensidades, funções de verossimilhança, etc.), com o objetivo de estabelecer um grau de potencialidade para quais locais podem ser modificados em um determinado momento no futuro. O módulo de mudança de alocação pode é responsável por estabelecer uma posição espacial da demanda gerada de um uso (PIMENTA, 2010).

O modelo CLUE (Conversion of Land Use and its Effects) é utilizado por esse framework, e consiste em arcabouços de modelagem com abordagem top-down desenvolvidos na Wageningen University, na Holanda (VELDKAMP; FRESCO, 1996), para quantificar relações entre uso do solo e seus fatores de influência através de relações estatísticas quantificadas empiricamente em combinação com modelagem dinâmica (PIMENTA, 2010).

4 CRONOGRAMA

Para a execução do trabalho aqui proposto, algumas etapas serão seguidas, como mostra o cronograma da Tabela 2:

Tabela 4.1 - Cronograma de execução do trabalho

Etapa	2011								2012		
	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan	Fev	Mar
1	•	•									
2	•	•									
3		•	•								
4				•	•	•	•				
5				•	•	•	•				
6						•	•	•	•		
7										•	

- Etapa 1: Aquisição dos dados de todos os dados necessários para construção de um banco de dados referente ao desmatamento da Amazônia.
- Etapa 2: Desenvolvimento de um framework contendo métodos de métricas de qualidade no TerraME (Costanza, Pontius, etc.)..
- Etapa 3: Implementação do modelo de Laurance et al.(2001) no LUCC-ME. .
- Etapa 4: Testes dos métodos de métricas de qualidade nos resultados do modelo do Laurance e no modelo de Aguiar (CLUE, já implementado no LUCC-ME), e proposta de um índice espaço-temporal que seja capaz de estimar o grau de realismo dos resultados destes modelos.
- Etapa 5: Elaboração de artigo com os resultados Execução do procedimento de validação sobre os modelos.
- Etapa 6: Elaboração da dissertação de mestrado.
- Etapa 7: Defesa da dissertação, e finalização do documento acrescentando as possíveis sugestões da banca.

REFERÊNCIAS BIBLIOGRÁFICAS

AGUIAR, A. P. D.; CÂMARA, G.; ESCADA, M. I. S. Spatial statistical analysis of land-use determinants in the brazilian amazonia: Exploring intra-regional heterogeneity. **Ecological modelling**, v. 209, p. 169–188, 2007. 1, 16, 18

ALLEN, T. F. H.; STARR, T. B. **Hierarchy: Perspectives for Ecological Complexity**. [S.l.]: University of Chicago Press, Chicago, 1982. 7

BLILIE, C. **The Promise and Limits of Computer Modeling**. [S.l.]: World Scientific, 2007. 1

CARNEIRO, T. G. de S. **Nested-CA: A Foundation For Multiscale Modelling of Land Use and Land Cover Change**. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais - INPE, 2006. 17

COSTANZA, R. Model goodness of fit: a multiple resolution procedure. **Ecological Modelling**, v. 47, p. 199–215, 1989. 2, 7, 15

GILBERT, N.; TROITZSCH, K. G. **Simulation for the social scientist**. 2. ed. [S.l.]: Open University Press, 2005. 1

LAMBIN, E. F.; GEIST, H.; RINDFUSS, R. R. Land-use and landcover change: local processes and global impacts. In: _____. [S.l.]: Springer, Berlim, 2006. cap. Introduction: Local Processes with Global Impacts, p. 1 – 8. 1

LAURANCE, W. F.; COCHRANE, M. A.; BERGEN, S.; M.FEARNSIDE, P.; DELAMÔNICA, P.; BARBER, C.; D'ANGELO, S.; FERNANDES, T. The future of the brazilian amazon. **Science**, v. 291, p. 438–439, 2001. 1, 16, 17

NASH, J.; SUTCLIFFE, J. River flow forecasting through conceptual models. part i: a discussion of principles. **Journal of Hydrology**, v. 27, p. 282 – 290, 1970. 2

PIMENTA, P. F. **Uma proposta de um framework para modelos computacionais de uso e cobertura da terra**. Tese (Mestrado) — Universidade Federal do Pará, 2010. 18

PONTIUS, G. R.; MALANSONA, J. Comparison of the accuracy of land-change models: cellular automata markov versus geomod. **International Journal of Geographical Information Science**, v. 19 (2), p. 243–265, 2005. 12

PONTIUS, R.; CORNELL, J.; HALL, C. Modeling the spatial pattern of land-use change with geomod2: application and validation for costa rica. **Agriculture, Ecosystems & Environment**, v. 85(1-3), p. 191–203, 2001. 2, 12, 16

PONTIUS, R. G.; HUFFAKER, D.; DENMAN, K. Useful techniques of validation for spatially explicit land-change models. **Ecological Modelling**, v. 179, p. 445–461, 2004. 2, 3, 4, 5, 8, 9, 12, 13, 14

RYKIEL, E. Testing ecological models: the meaning of validation. **Ecological Modelling**, v. 90, p. 229–244, 1996. 3, 12

VELDKAMP, A.; FRESCO, L. O. Clue: a conceptual model to study the conversion of land use and its effects. **Ecological modelling**, v. 85, p. 253–270, 1996. 18

VERBURG, P. H.; KOK, K.; PONTIUS, R. G.; VELDKAMP, A. Land-use and landcover change: local processes and global impacts. In: _____. [S.l.]: Springer, Berlin, 2006. cap. Modeling Land-Use and Land-Cover Change, p. 117 – 131. 1

WILLMOTT, C. On the validation of models. **Physical Geography**, v. 2, p. 184 – 194, 1981. 2