

PUBLICAÇÃO DE DADOS CONECTADOS PARA MODELOS DE USO E COBERTURA DA TERRA

Linked data publication for land change models

Deyvison Aguiar Garcia
Pós-Graduando em Geoprocessamento do Instituto Federal do Maranhão (IFMA) São Luís - MA.
deyvisongarcia025@gmail.com

Sérgio Souza Costa
Docente da Universidade Federal do Maranhão (UFMA) São Luís - MA.
prof.sergio.costa@gmail.com

Evaldinolia Gilbertoni Moreira
Docente do Instituto Federal do Maranhão (IFMA) São Luís - MA.
evaldinolia@gmail.com

Recebido: 01/06/2019

Aceito: 10/12/2019

RESUMO: Os modelos de mudanças de uso e cobertura da terra utilizam um grande volume de dados. Esse artigo apresenta a integração de diferentes ferramentas e conceitos em uma metodologia que permite o reuso e recriação de bases de dados entre diferentes ambientes computacionais. A metodologia foi aplicada na construção e na publicação de uma base de dados para a Amazônia Brasileira. Nesse artigo usou-se um pequeno número de variáveis, mas a metodologia pode ser aplicada independentemente da quantidade de variáveis. O sistema de informação geográfica QGIS foi usado para demonstrar a metodologia. Porém, diferentes sistemas computacionais poderão ser usados com pequenas adequações. Todos os dados e códigos utilizados, estão publicados para permitir replicação dos experimentos descritos aqui.

Palavras-chave: Modelos dinâmicos, dados conectados, Amazônia, Cobertura da terra.

ABSTRACT: Land change models use a large amount of data. This article integrates different tools and concepts in a methodology that allows the reuse and re-creation of databases between different computational environments. The methodology was applied in the construction and publication of a database for the Brazilian Amazon. In this paper, a few variables was used, but the methodology can be applied independently of the number of variables. In addition, the QGIS geographic information system was used to demonstrate the methodology. However, different computational systems can be used with small adaptations. All data and codes used are published to allow replication of the experiments described here.

Keywords: Dynamic models, linked data, Amazon, Land cover.

INTRODUÇÃO

Mudança de uso e de cobertura da terra é o resultado de interações complexas entre sistemas sociais e biofísicos (TURNER et al., 1995). Onde cobertura da terra se refere aos atributos físicos da superfície terrestre (por exemplo, floresta, água, gramíneas e áreas construídas). Enquanto o termo uso da terra se refere ao uso humano de tais atributos (por exemplo, recreação, proteção, pastagem, área residencial e área comercial). Então, mudanças de uso e cobertura da terra se referem tanto à conversão entre classes (por exemplo, processos de desertificação e desflorestamento), quanto às alterações nessas classes (por exemplo, intensificação de uso agrícola e degradação da cobertura vegetal).

O interesse por modelos espaciais dinâmicos já tem mais de uma década e pode ser observado em vários trabalhos, como (ESPINDOLA et al., 2012, 2012; FILATOVA et al., 2013; MEIYAPPAN et al., 2014; MOREIRA et al., 2009; SUBEDI; SUBEDI; THAPA, 2013; VAN ASSELEN; VERBURG, 2013). Isso se deve aos impactos de fatores locais e globais nos sistemas socioambientais. Esses impactos podem levar a extinções de comunidades, fauna e flora, além de alterações climáticas. Compreender como esses fatores afetam essas mudanças são importantes para definir indicadores confiáveis que permitirão subsidiar a elaboração de políticas públicas. Modelos computacionais são ferramentas úteis para complementar a capacidade mental de modelagem, permitindo assim tomadas de decisões. Modelos de mudanças terrestres podem então ajudar na avaliação dos impactos possíveis e das regras alternativas através da construção de cenários e simulações de hipóteses. Uma grande variedade de modelos pode ser encontrada na literatura, com objetivos, técnicas, embasamento teórico e tradições de modelagem distintas (BRIASSOULIS, 2000; PARKER et al., 2002; PONTIUS et al., 2018; VERBURG, 2006).

Este trabalho foca nos modelos que utilizam um espaço celular como meio para integrar as variáveis socioambientais. Esses modelos podem usar espaços celulares explicitamente (ESPINDOLA et al., 2012; MOREIRA et al., 2009), ou através de imagens (VELDKAMP; FRESCO, 1996; VERBURG et al., 2002). Independentemente da abordagem, esses modelos geralmente representam a superfície terrestre como uma matriz, onde cada célula tem um estado próprio que pode variar ao longo do tempo como resultados de regras de transição (LONGLEY *et al.*, 2009). Essas regras podem envolver regressões, redes neurais, autômatos celulares ou agentes. Independentemente, todos esses modelos passam por uma longa etapa onde será construído a base de dados. Em resumo, vários dados serão agregados e/ou extrapolados para serem usados em uma dada resolução. Esse processo demanda algoritmos com alto custo computacional e uma grande quantidade de variáveis. Por exemplo, em (ESPINDOLA et al., 2012), os autores usaram 30 variáveis para representar o uso do solo, o acesso a mercados, as políticas públicas, a estrutura agrária e os dados ambientais. Então, a recriação dessa base de dados em outro ambiente computacional irá exigir muito tempo, tornando muito difícil ou impraticável a sua replicação. Além da replicação em diferentes ambientes computacionais,

esperava-se que os dados já calculados pudessem ser reusados. Entretanto, isso não é possível pois os valores calculados são associados as células criadas dentro de um espaço celular específico. Não podendo ser reusado em outro espaço celular, mesmo que a resolução espacial seja idêntica. Esse problema específico será discutido e exemplificado no decorrer desse artigo.

Este artigo tem como objetivo apresentar uma metodologia para construção e publicação de dados sobre modelos espaciais dinâmicos. Com esta metodologia é possível publicar os dados utilizados em uma dada modelagem, reusá-los e replicar os experimentos em diferentes ambientes computacionais. Para alcançar esse objetivo, foi usado o conceito de dados conectados, proposto por Tim Berners-Lee (BERNERS-LEE, 2006). Esse conceito entende a web como uma grande base de dados conectados. Esta arquitetura permite conectar dados entre diferentes bases de dados e criadas por diferentes produtores. Além deste conceito, foi usado um repositório global de espaços celulares denominado DBCells (COSTA, S. S et al., 2017) e o vocabulário para dados estatísticos, denominado Data Cube, (CYGANIAK; REYNOLDS; TENNISON, 2014).

Esse artigo está organizado da seguinte maneira: Seção 2 apresenta a metodologia do trabalho, Seção 3 apresenta os resultados e discussões do emprego da metodologia e Seção 4 apresenta as conclusões.

MATERIAIS E MÉTODOS

Para compreender a metodologia, é necessário compreender quais modelos espaciais dinâmicos que esta abordagem tem potencial para usar. Como apresentado na Seção 1, esse trabalho tem foco nos modelos espaciais dinâmicos que usam espaços celulares como meios de integração entre dados espaciais e estatísticos. Destacam-se as seguintes ferramentas e modelos: CLUE (VELDKAMP; FRESCO, 1996), CLUES (VERBURG et al., 2002), DINAMICA (SOARES-FILHO; CERQUEIRA; PENNACHIN, 2002), RIKS (WHITE; ENGELEN, 2000), CA_Markov (SUBEDI; SUBEDI; THAPA, 2013), TerraME (CARNEIRO et al., 2013) e LuccME¹.

Estes modelos e ferramentas permitem seguir uma abordagem denominada de top-down (VERBURG, 2006). Basicamente eles possuem um módulo de demanda, potencial e de alocação (COSTA et al., 2009; VAN SCHROJENSTEIN LANTMAN et al., 2011). A Figura 1 destaca como esses módulos se comunicam.

¹ Site oficial: <http://luccme.ccst.inpe.br/>

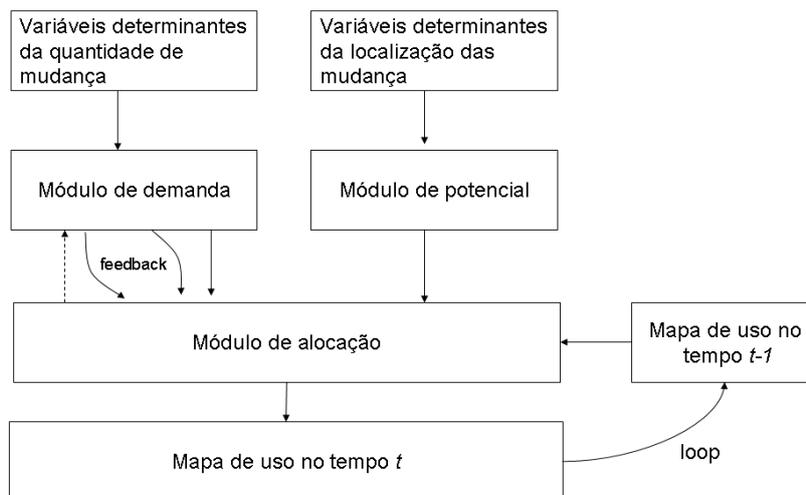


Figura 1. Estrutura geral dos modelos top-down. **Adaptado** de (VERBURG, 2006).

Por ser um modelo dinâmico, o processo ocorre através de diversas iterações que refletem os anos para o qual aquele modelo está sendo aplicado. Então, no módulo de demanda é calculada a quantidade de mudança para cada iteração. Essa quantidade de mudança é passada para o módulo de alocação, onde algoritmos específicos de alocação irão efetuar as mudanças de acordo com o mapa de potencial de mudança. Este mapa é produzido pelo módulo de potencial através de algum modelo matemático que relaciona as mudanças com um conjunto de variáveis. Tanto o módulo de demanda, quanto o módulo de potencial tem como entrada um conjunto de variáveis. Algumas dessas são usadas para o cálculo de quantidade de mudança e outras para o cálculo da localização das mudanças.

Para entender a contribuição deste trabalho, é preciso compreender o módulo de potencial. Esse módulo é responsável por responder onde ocorrerão as mudanças. O principal pressuposto desse módulo é que as mudanças de cobertura e uso do solo não ocorrem de forma aleatória no espaço, ou seja, existe uma forte relação entre as mudanças e algumas variáveis socioeconômicas e ambientais, como conexão a mercado, declividade e acesso a estradas. Usualmente, utiliza-se algum método estatístico para estabelecer as relações entre as mudanças terrestres e um conjunto de variáveis.

O método frequentemente utilizado para o relacionamento das variáveis é a estatística multivariada, mais especificamente a análise de regressão. Este método permite explorar e inferir a relação de uma variável dependente (uso da terra) com variáveis independentes específicas (fatores explanatórios). O método de regressão utilizada depende ainda da representação dos dados. Regressão linear é usada em modelos que operam sobre dados contínuos (por exemplo, o CLUE) enquanto regressão logística é empregada em modelos que operam sobre dados discretos (por exemplo, o DINAMICA e CLUE-s). Outro método estatístico empregado é o método de pesos de evidência. Este método é baseado no teorema da probabilidade condicional de Bayes (AN; MOON; BONHAM-CARTER, 1994). No contexto dos modelos de uso da

terra, este teorema concerne em calcular a propensão de um uso, por exemplo, uso não-urbano para residencial, dado que uma evidência (por exemplo, suprimento de água), também chamada variável explicativa, já ocorreu (ALMEIDA; CZADO, 2012; MARIA DE ALMEIDA et al., 2003). Existem ainda diversos outros métodos para o relacionando das variáveis explicativas com as mudanças terrestres, que podem variar de métodos simples de agregação (média ou produto) a métodos mais complexos como redes neurais. Em (LESSCHEN; VERBURG; STAAL, 2005), os autores discutem em detalhes vários destes métodos que são aplicados a modelos de mudança terrestre. O ponto principal para esse artigo é compreender que todos esses modelos e ferramentas demandam uma coleção de variáveis socioeconômicas e ambientais como entrada.

Além de entender como o modelo funciona, é importante compreender como ocorre o processo de desenvolvimento dos modelos espaciais dinâmicos. Nesse artigo, considerou-se quatro grandes etapas, onde três delas são cíclicas como apresentado na Figura 2.

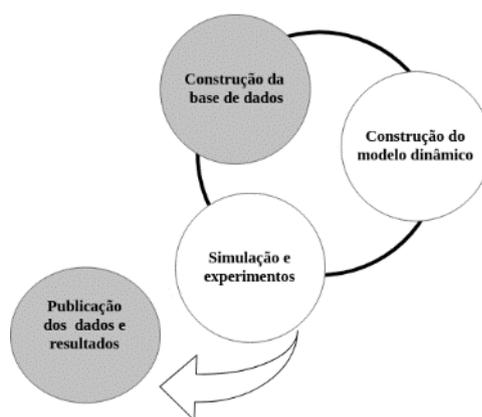


Figura 2. Ciclo de desenvolvimento de modelos espaciais dinâmicos.

A primeira etapa é a construção da base de dados, quando será definido a área de estudo e criado um banco de dados espacial com as variáveis que serão utilizadas pelo modelo. Na segunda etapa é construído então um modelo que irá relacionar as variáveis com o uso ou cobertura da terra. Na terceira, as simulações são executadas e os resultados avaliados. Essas três etapas são cíclicas, pois o processo de simulação é também um processo de aprendizado onde se pode identificar a necessidade de adicionar ou desconsiderar variáveis. Com a conclusão do processo, os dados e resultados são publicados geralmente em artigos científicos.

A contribuição deste artigo está nas duas etapas em destaque na Figura 2, que são a construção da base de dados e a publicação dos dados e resultados. A metodologia proposta para essas etapas, será detalhada a seguir.

Construção da base de dados

O desenvolvimento de um modelo espacial dinâmico inicia-se pela definição da escala espacial de estudo. Nesse passo será necessário definir a extensão e a resolução

espacial. Como nossa proposta é sobre os modelos baseados em células, é preciso definir qual o tamanho de cada célula, ou seja, a resolução. Pode-se utilizar a própria resolução das imagens de satélites. Contudo, em muitos casos a simulação é aplicada em grandes extensões espaciais e com resoluções mais baixas. Por exemplo, em (AGUIAR; CÂMARA; ESCADA, 2007), os autores usaram duas escalas uma de 100 x 100 km, e outra de 25 x 25 km. Então, é preciso criar artificialmente uma grade com a resolução espacial adequada ao modelo. Isso pode ser feito com recursos presentes nos sistemas de informação geográfica. Por exemplo, no QGIS² existe uma opção para criar grades com uma dada resolução, Figura 3.

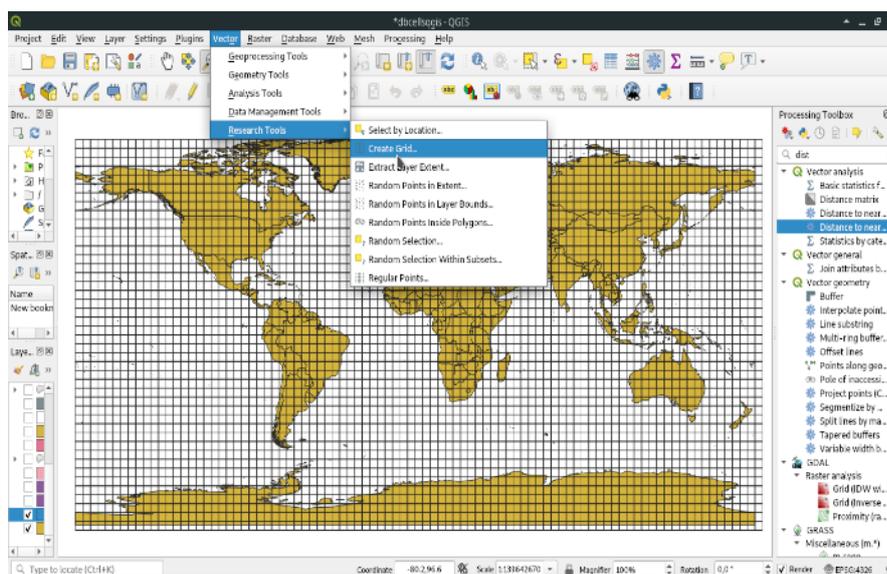


Figura 3. Criando grades (espaços celulares) no QGIS.

O problema com essa abordagem é que o sistema irá criar as grades (ou espaços celulares) tendo como referência uma determinada área de estudo. Então ao mudar a área de estudo, um novo espaço celular é gerado e as novas células não terão nenhuma relação com as criadas anteriormente. Por exemplo, é comum gerar uma grade para uma dada área de estudo A1 e depois uma diferente área de estudo A2. Onde A2 pode estar contido em A1, como apresentado pelo Diagrama Venn da Figura 4.



Figura 4. Diagrama Venn ilustrando o problema das áreas de estudos relacionadas.

² Sistema de informação geográfica livre e aberto, mas informações acesse <https://www.qgis.org/>.

Em um cenário ideal, deveria ser fácil reutilizar os dados já calculados para A1, em A2. Contudo, isso não ocorre, pois, as células de cada grade não possuem uma relação direta entre si. Além disso, elas ficarão geralmente deslocadas espacialmente. Por exemplo, dados calculados para Amazônia legal, não poderão ser reusados para um estado Amazônico como o Pará. Para entender esse problema, criou-se uma grade com 4 graus de resolução para a América do Sul (em cor preta) e outra para o Brasil (em cor azul). Observe na Figura 5 o problema de deslocamento. Isso irá ocorrer até mesmo com bases criadas pelo mesmo pesquisador e na mesma ferramenta.

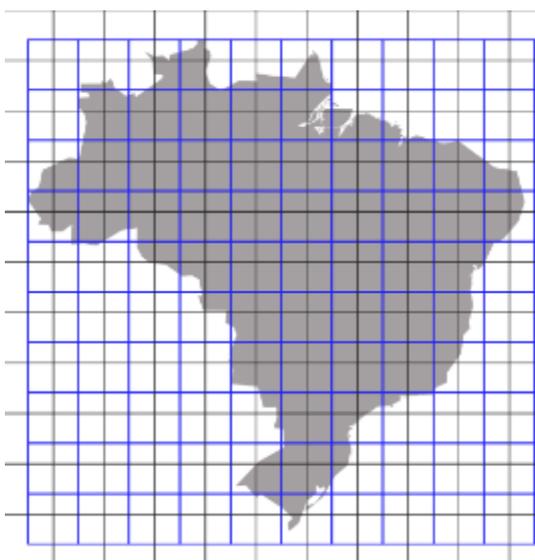


Figura 5. Deslocamento entre células criadas para diferentes áreas de estudo.

Devido a esse problema, em (COSTA, S. S et al., 2017), os autores propuseram uma arquitetura denominada DBCells que trata cada célula como um objeto único e identificável. Nessa abordagem, uma grade (ou espaço celular) pode ser entendida como uma coleção de geo-objetos disjuntos e regulares. Tratar as células como geo-objetos pode favorecer o reuso de dados e a replicação de modelos (COSTA, S. S et al., 2017). Nesta arquitetura as células são identificadas a partir de uma URI (*Uniform Resource Identifier*), o que possibilitará ligá-las as outras bases de dados em diferentes repositórios na Web. Esse conceito é denominado de dados conectados, proposto pelo criador da Web Tim Berners-Lee em (BERNERS-LEE, 2006). A ideia básica é tratar a web como uma grande base de dados conectada, o que ele denominou como a web dos dados. Para se ter uma noção do que significa esse conceito, um grupo de pesquisadores tem gerado um gráfico que apresenta as conexões entre as bases de dados na web, Figura 6.

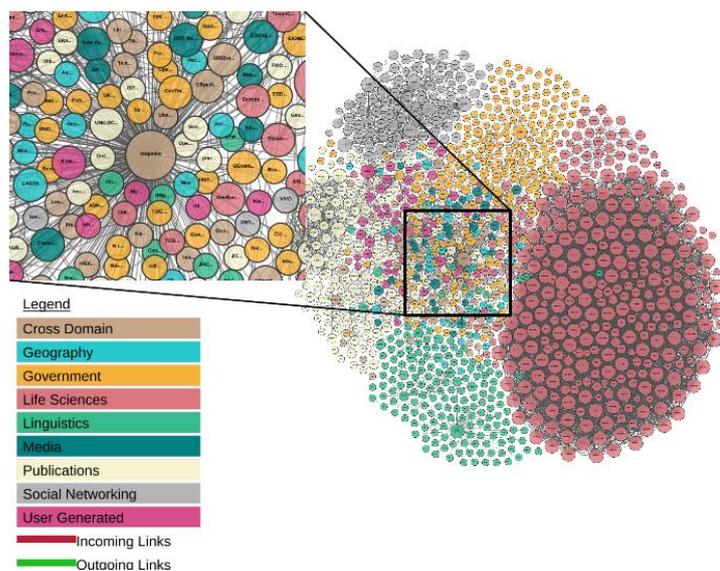


Figura 6. Base de dados conectada em 2017. Fonte (ABELE et al., 2017)

A arquitetura DBCells já está em funcionamento através de um repositório³ de espaços celulares de diferentes projeções e resoluções para todo o globo terrestre. Atualmente é provido grades de múltiplas resoluções na projeção WGS84 (EPSG4326). De acordo com o site oficial, outras projeções deverão ser suportadas no futuro, por exemplo, o *Discrete Global Grid* que já é uma especificação OGC (PURSS et al., 2016; SAHR; WHITE, 1998). As resoluções suportadas atualmente são de 4, 2, 1, 0.50 e 0.25 graus. A Figura 7 apresenta um espaço celular com uma resolução de 4 x 4 graus. O repositório DBCells tem como objetivo servir como ponto comum, onde diferentes bases de dados usadas em modelos espaciais dinâmicos poderão se relacionar.

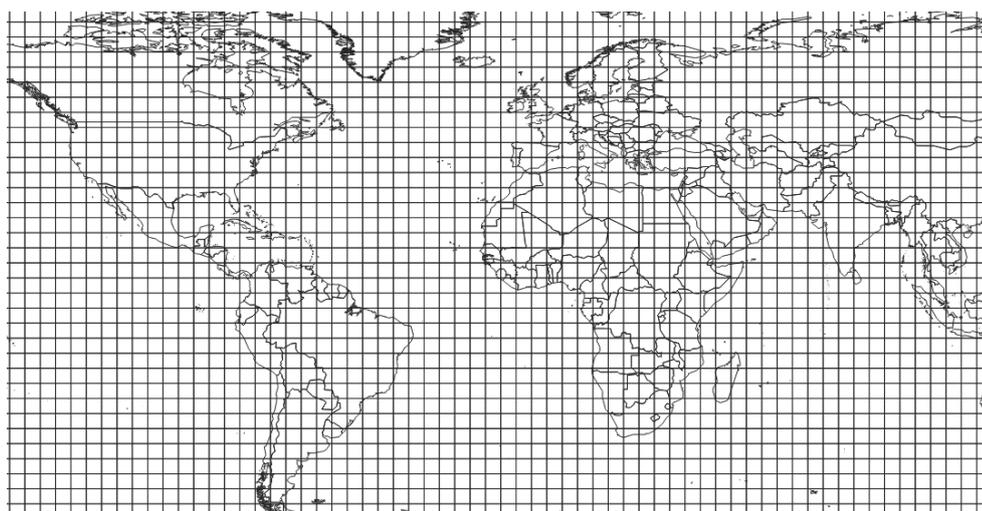


Figura 7. Espaço celular com resolução de 4 x 4 graus.

³ Repositório DBCells: www.dbcells.org

Esse artigo avança em uma metodologia que utiliza os conceitos propostos em (COSTA, S. S et al., 2017), para construção da base de dados. Onde considera-se os seguintes passos:

1. Importar os dados do repositório DBCells para uma base local, onde a URI do repositório será usada como identificador de cada célula.
2. Fazer o recorte da área de estudo, usando os recursos do sistema de informação geográfica.
3. Opcionalmente, pesquisar por dados já existentes, publicados em algum servidor de dados conectados.
4. Utilizar um sistema de informação geográfica ou linguagem de preferência para gerar as variáveis usadas pelo modelo.

Observe, que a principal restrição para usar essa metodologia é importar as células do repositório ao invés de criá-las diretamente pelo sistema de informação geográfica. Como o DBCells é um repositório aberto de dados conectados, os dados podem ser acessados através da linguagem padrão para consulta de dados conectados, denominada SPARQL⁴. A próxima Seção irá apresentar a ferramenta utilizada neste trabalho, porém essa abordagem pode ser aplicada similarmente com outras ferramentas.

Depois deste passo inicial, o processo de construção da base de dados poderá continuar localmente, utilizando o sistema de informação geográfica de preferência. Esse processo irá seguir até que a base de dados tenha todas as variáveis que serão usadas nos passos seguintes. Esses passos seguem cíclicos até a última etapa que é a publicação dos dados e resultados, como descrita a seguir.

Publicação dos dados e resultados

Durante a criação da base de dados, o que pode chegar a ter vários gigabytes, incluindo os dados brutos como imagens de satélites e dados vetoriais. Entretanto, essa metodologia não trata sobre a publicação dos dados brutos, que na sua maioria já estão disponíveis em algum repositório público. A metodologia é sobre a publicação dos dados e resultados usados pelo modelo, e que já estão associados a cada célula. Essa etapa é definida basicamente em dois passos:

1. Exportar as variáveis de cada célula, para o formato de dados conectados, onde cada célula possui uma URI para o repositório DBCells.

⁴ Mais informações em: <https://www.w3.org/TR/rdf-sparql-query/>

2. Publicar estes dados em um repositório público, que pode ser gerido por uma instituição, ou em repositórios globais, como o Data.World⁵.

Para a realização do primeiro passo, será necessário um conhecimento sobre o conceito de dados conectados, que pode ser encontrado em diversas referências como (BERNERS-LEE, 2006; HEATH; BIZER, 2011; KAUPPINEN; DE ESPINDOLA, 2011; KUHN; KAUPPINEN; JANOWICZ, 2014). Como observado em (KUHN; KAUPPINEN; JANOWICZ, 2014), dados conectados são coleções de afirmações sobre algo, feito por alguém, em algum lugar e em um dado momento. No caso de dados espaciais, essas afirmações são feitas geralmente com base em medições realizadas por sensores remotos. Por exemplo, a partir dos valores de reflectância capturados pelos sensores de um dado satélite. Em (COSTA, S. S et al., 2017), os autores não definiram como construir esses dados conectados. Então, esse trabalho propõe o uso de um vocabulário específico para representar dados estatísticos, denominado *Data Cube Vocabulary* (CYGANIAK; REYNOLDS; TENNISON, 2014). A utilização deste vocabulário já havia sido considerada como um trabalho futuro em (KAUPPINEN et al., 2014). Uma grande vantagem desse vocabulário é separar observações, conceitos e dimensões. O conceito é o que será medido, como a distância a estradas, conexão a portos e distância a reservas minerais. As dimensões no caso de modelos dinâmicos espaciais são usualmente o tempo e o espaço. Neste caso, o espaço é a célula para qual aquela observação foi medida, e o tempo é usualmente a data de aquisição do dado.

A próxima Seção exemplifica essa metodologia, permitindo uma melhor compreensão dessas etapas, tanto a construção da base de dados quanto a publicação dos dados e resultados.

RESULTADOS E DISCUSSÕES

Para exemplificar a metodologia proposta, foram realizadas as duas etapas discutidas propostas pela metodologia que são (1) a construção de uma base de dados e (2) a publicação dos dados e resultados. A seguir essas etapas serão descritas em mais detalhes.

Construção da base de dados

A construção da base de dados se inicia pela definição da escala espacial, ou seja, a extensão e a resolução. A extensão utilizada neste experimento é a Amazônia Legal Brasileira em uma resolução de 0.25 x 0.25 graus, ou seja, aproximadamente 27 x 27 km. Na metodologia proposta, este espaço celular precisa estar vinculado ao repositório DBCells descrito anteriormente. Esse é um repositório de espaços celulares, então cada célula tem informações mínimas, como o tipo, resolução, além das relações entre as diferentes resoluções. Adicionalmente, cada célula está relacionada ao continente a que pertence. Essas informações são úteis para limitar a busca por um dado continente e em uma dada resolução (Quadro 1). Como discutido

⁵ Acessado em <https://data.world/>

anteriormente, o repositório utiliza a linguagem de consulta SPARQL para selecionar o espaço celular requerido.

```
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbco: <http://purl.org/ontology/dbcells/cells#>.
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
SELECT ?cell ?wkt
WHERE {
    ?cell geo:asWKT ?wkt.
    ?cell geo:sfWithin dbr:South_America.
    ?cell dbco:resolution 0.25.
}
```

Quadro 1: Consulta SPARQL usada para carregar o espaço celular inicial.

Essa consulta irá retornar a URI de cada célula, com a sua respectiva geometria, descrita no formato WKT⁶. Independentemente do sistema de informação geográfica, linguagem ou ferramenta de simulação, é possível enviar essa consulta ao repositório DBCells. Com essa consulta, o servidor irá retornar os dados que serão manipulados localmente. Por exemplo, esse experimento usou como sistema de informação geográfica o software livre e aberto QGIS e a linguagem de programação Python⁷. Então, foi escrito um código que envia essa consulta diretamente ao repositório DBCells. Ao receber os dados, esse código constrói um espaço celular onde cada célula tem como seu identificador a URI para o repositório. O código está disponibilizado publicamente em: <http://bit.ly/2Vo2k4z>. A execução desse código pode demandar alguns minutos, dado que será realizada uma consulta em uma grande base de dados. Além disso, o repositório DBCells ainda está em uma versão preliminar. Para tornar o processo mais eficiente, é possível salvar a consulta em um outro repositório, como o Data.World. Com isso, pode ser escrito um código que não enviará uma consulta, mas irá solicitar uma base de dados já armazenada, reduzindo o tempo de execução para alguns segundos. O código indicado para a reprodução desse trabalho pode ser acessado em: <https://bit.ly/2QcApUk>.

Com os dados importados para um sistema de informação geográfica, o usuário irá manipular esses dados localmente. Nesse experimento, para exemplificar, gerou-se apenas três variáveis para cada célula. Lembrando que uma simulação usualmente irá utilizar dezenas de variáveis, porém esta metodologia poderá ser aplicada independentemente do número de variáveis. A primeira variável, foi o percentual de floresta em 2017 para cada célula. Essa variável foi gerada a partir da imagem disponibilizada no repositório do PRODES⁸. A Figura 8 apresenta, a área de estudo recortada com 6371 células e com o percentual de floresta calculado a partir da imagem do PRODES. Observe que o identificador de cada célula é uma URI para o repositório DBCells.

⁶ WKT do inglês Well-known text, é uma representação geométrica no formato de texto.

⁷ Site oficial: <https://www.python.org/>

⁸ A imagem utilizada pode ser baixada em <http://www.dpi.inpe.br/prodesdigital/dadosn/2017/>.

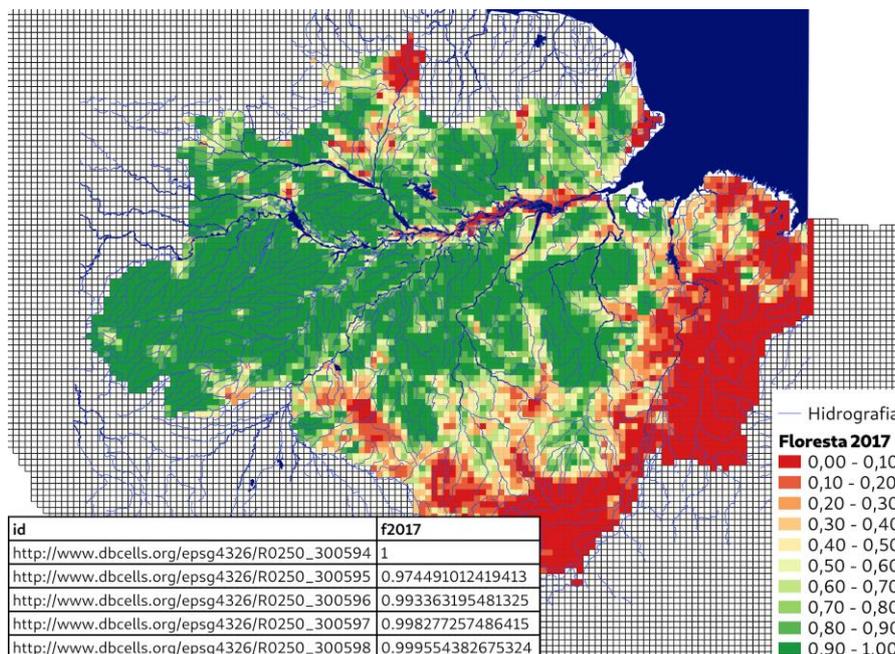


Figura 8. Área de estudo após gerar os dados de cobertura de floresta em 2017.

As outras duas variáveis geradas foram: a menor distância entre o centroide de cada célula para uma sede urbana e para uma estrada. Como o objetivo desse trabalho é testar a metodologia, não houve uma preocupação em validar essas geometrias. Ou seja, não foi verificado o ano e nem mesmo a sua correteude. Considerando os dados existentes, gerou-se então as variáveis de distância para sedes urbanas e de distância para estradas, como apresentada na Figura 9.

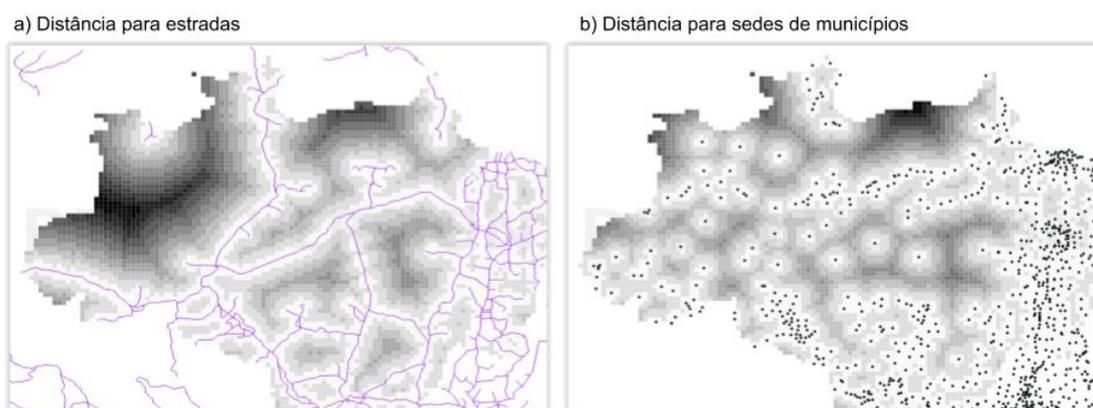


Figura 9. (a) Distância para estradas e (b) distância para sedes de municípios.

Publicação de dados e resultados

Em um cenário real, após a criação de um banco de dados com dezenas de variáveis, os pesquisadores publicam os resultados em artigos científicos. Contudo, como discutido anteriormente, é um grande desafio reconstruir a base de dados usada em um dado trabalho. De acordo com a metodologia proposta, as variáveis serão descritas usando o vocabulário *Data Cube*⁹. Porém, é necessário que as variáveis como porcentagem de floresta, distância a estrada e distância a sedes urbanas sejam definidas através de uma ontologia. Essa ontologia irá ligar esses conceitos a observações no vocabulário *Data Cube* além de definir qual a faixa de valores que eles poderão assumir. No futuro, é esperado que já exista um vocabulário criado por outro usuário e que esse poderá ser reusado. Diferentemente do que ocorre em (KAUPPINEN et al., 2014) esses conceitos são definidos independentes do tempo. O conceito não é a **distância a estradas em 2015**, mas sim a **distância a estradas**. Tornando assim o reuso desta ontologia mais fácil.

Como ainda não existem ontologias adequadas ao nosso trabalho, foi necessário criar uma ontologia específica¹⁰ para as variáveis usadas no modelo, como pode ser observado em Quadro 2.

```
### http://purl.org/ontology/dbcells/amazon#Percentage_of_Forest
:Percentage_of_Forest rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf sdmx-measure:obsValue ;
  rdfs:range xsd:decimal .

### http://purl.org/ontology/dbcells/amazon#Road_Distance
:Road_Distance rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf sdmx-measure:obsValue ;
  rdfs:range xsd:decimal .

### http://purl.org/ontology/dbcells/amazon#Urban_Distance
:Urban_Distance rdf:type owl:DatatypeProperty ;
  rdfs:subPropertyOf sdmx-measure:obsValue ;
  rdfs:range xsd:decimal .
```

Quadro 2: Ontologia para as variáveis usadas no experimento.

Com a ontologia definida, pode-se então exportar os dados para o formato de dados conectados, ou seja, uma coleção de triplas com sujeito, propriedade e valor. Novamente isso foi feito através de um código escrito na linguagem de programação Python, que exportou os dados do QGIS. Esse algoritmo também está disponível no seguinte endereço: <http://bit.ly/2XcOzY6>. Ao usar esse formato, cada variável será representada como uma dada observação, associada a um conceito, uma célula e um

⁹ Especificação em <https://www.w3.org/TR/vocab-data-cube/>

¹⁰ Acesse a ontologia criada em <http://purl.org/ontology/dbcells/amazon>

dado tempo. Por exemplo, o Quadro 3, descreve uma observação associada ao conceito de porcentagem de floresta, célula R0250_302678, e data 2017-01-01.

```
@prefix amz: <http://purl.org/ontology/dbcells/amazon#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
@prefix dbco: <http://purl.org/ontology/dbcells/cells#>.
@prefix qb: <http://purl.org/linked-data/cube#>.
@prefix lc: <http://purl.org/landchange/amazon/>.
@prefix dbce: <http://www.dbcells.org/epsg4326/>.

lc:F17_1 a qb:Observation ;
amz:Percentage_of_Forest "0.99672"^^xsd:decimal ;
dbco:refCell dbce:R0250_302678
dbco:refDate "2017-01-01"^^xsd:dateTime.
```

Quadro 3: Exemplo de uma observação.

Após a geração dos dados, é necessário então publicá-los em um repositório. Algumas instituições podem criar e manter repositórios próprios. Entretanto, existem alternativas como o repositório global de dados conectado denominado Data World¹¹. Esse portal permite que pessoas e/ou instituições compartilhem os dados diretamente no portal. Neste trabalho, criou-se um repositório de dados nesse portal, e ele poderá ser acessado através do seguinte endereço: <http://bit.ly/2WeDFVu>.

Diretamente pelo portal do Data.World é possível conectar os dados das variáveis com o repositório DBCells. Por exemplo, a consulta SPARQL da Quadro 4 retorna a porcentagem de floresta, distância a estradas, distância a sedes urbanas e a geometria de cada célula.

```
SELECT ?uriCell ?wkt ?forest ?distUrban ?distRoad
where {
  ?o1 amz:Percentage_of_Forest ?forest.
  ?o1 dbco:refCell ?uriCell.
  ?o2 amz:Urban_Distance ?distUrban.
  ?o2 dbco:refCell ?uriCell.
  ?o3 amz:Road_Distance ?distRoad.
  ?o3 dbco:refCell ?uriCell
  SERVICE <http://dbcells.org:3030/cells> {
    ?uriCell geo:asWKT ?wkt.
    ?uriCell geo:sfWithin dbr:South_America.
    ?uriCell dbco:resolution 0.25. }
}
```

Quadro 4: Conectando os dados entre o repositório DBCells e Data "World".

A Figura 10, apresenta a interface gráfica do Data World para o envio de uma consulta SPARQL.

¹¹ O repositório <https://data.world/>, pode ser usado para armazenar dados gratuitamente

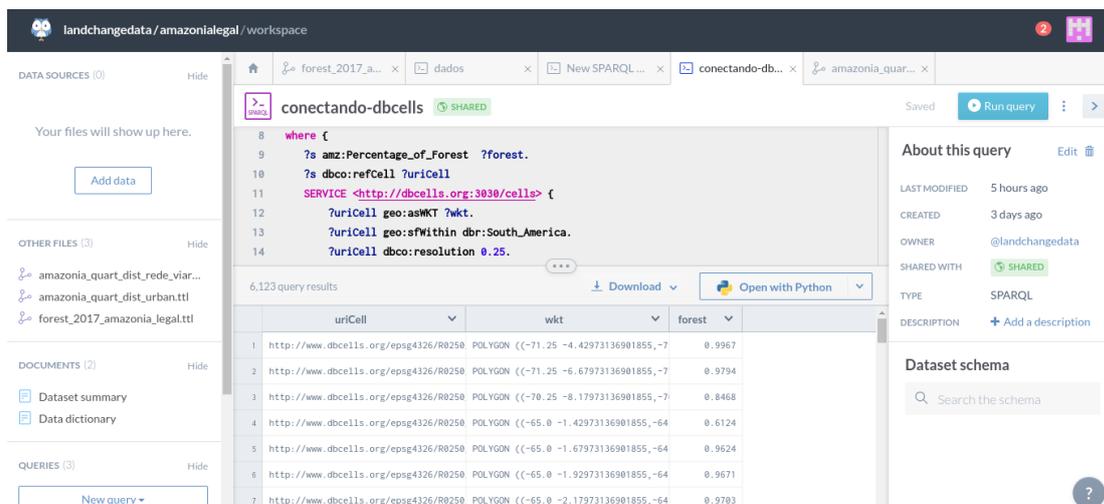


Figura 10. Enviando a consulta através da Interface do “Data World”.

O dado retornado por essa consulta pode ser baixado diretamente em um formato aberto no seguinte endereço: <https://bit.ly/2LXjkzi>. Esse dado pode ser carregado e importado para um sistema de informação geográfica e/ou banco de dados. Neste trabalho usou-se novamente um código em Python que dado esse endereço constrói a base de dados com todas as variáveis publicadas. Esse código pode ser acessado (<https://bit.ly/2LU12Pi>), baixado e executado no QGIS.

CONSIDERAÇÕES FINAIS

Modelos espaciais dinâmicos têm uma grande relevância tanto para a ciência quanto para políticas públicas, sobretudo os que estudam as mudanças de uso e cobertura da terra. Entretanto, dado o grande volume de dados, existe uma dificuldade em reconstruir uma base de dados usada em uma determinada simulação. Este trabalho avançou ao integrar alguns conceitos e ferramentas anteriores, como os repositórios DBCells e Data World com o vocabulário Data Cube. Neste trabalho os experimentos foram limitados com relação ao número de variáveis, mas a mesma metodologia pode ser aplicada independentemente da quantidade de variáveis. Na verdade, ela se torna mais relevante com maior quantidade de variáveis. Então, em um trabalho futuro pretende-se demonstrar a utilização desta metodologia em uma modelagem completa, que envolverá na publicação de uma quantidade maior de variáveis e resultados.

Atualmente, um problema encontrado foi a instabilidade do repositório DBCells, que está em desenvolvimento e em aprimoramentos para responder de modo mais eficiente as consultas. Isso requer uma melhor configuração tanto de software quanto da infraestrutura. No portal, os desenvolvedores explicam esses problemas, e estão buscando investimentos para conseguir entregar os dados mais rapidamente. Porém, mesmo com esses problemas, este artigo conseguiu demonstrar que já é possível conectar dados de outros repositórios ao DBCells, tornando o processo de reuso e publicação de dados de modelos mais simples. A expectativa futura é que os

pesquisadores publiquem os resultados não apenas em artigos, mas deixem os dados e os modelos disponíveis em repositórios públicos, tornando a sua replicação viável.

REFERÊNCIAS

ABELE, A. et al. *Linking Open Data cloud diagram 2017*. [s.l.: s.n.].

AGUIAR, A. P. D. et al. Modeling the spatial and temporal heterogeneity of deforestation-driven carbon emissions: The INPE-EM framework applied to the Brazilian Amazon. *Global Change Biology*, v. 18, n. 11, p. 3346–3366, 2012.

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, n. 2–4, p. 169–188, 2007.

ALMEIDA, C.; CZADO, C. Efficient Bayesian inference for stochastic time-varying copula models. *Computational Statistics & Data Analysis*, v. 56, n. 6, p. 1511–1527, 1 jun. 2012.

AN, P.; MOON, W. M.; BONHAM-CARTER, G. F. Uncertainty management in integration of exploration data using the belief function. *Nonrenewable Resources*, v. 3, n. 1, p. 60–71, 1 mar. 1994.

BERNERS-LEE, T. Linked Data. *International Journal on Semantic Web and Information Systems*, v. 4, n. 2, p. 1, 2006.

BRIASSOULIS, H. *Analysis of Land Use Change: Theoretical and Modeling Approaches*. [s.l.] Regional Research Institute, West Virginia University, 2000.

COSTA, S. et al. *Common Concepts to Development of the Top-Down Models of Land Changes*. XIV Brazilian Symposium on Remote Sensing. **Anais...** In: BRAZILIAN SYMPOSIUM ON REMOTE SENSING. 1 abr. 2009.

COSTA, S. S et al. DBCells – an open and global multi-scale linked cells. *Revista Brasileira de Cartografia*, v. 69, 2017.

CYGANIAK, R.; REYNOLDS, D.; TENNISON, J. The RDF data cube vocabulary. *World Wide Web Consortium*, 2014.

ESPINDOLA, G. M. et al. Agricultural land use dynamics in the Brazilian Amazon based on remote sensing and census data. *Applied Geography*, v. 32, n. 2, p. 240–252, 1 mar. 2012.

FILATOVA, T. et al. Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental modelling & software*, v. 45, p. 1–7, 2013.

HEATH, T.; BIZER, C. Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, v. 1, n. 1, p. 1–136, 2011.

KAUPPINEN, T. et al. Linked Brazilian Amazon Rainforest Data. *Semantic Web*, v. 5, n. 2, p. 151–155, 2014.

KAUPPINEN, T.; DE ESPINDOLA, G. M. Linked open science-communicating, sharing and evaluating data, methods and results for executable papers. *Procedia Computer Science*, v. 4, p. 726–731, 2011.

KUHN, W.; KAUPPINEN, T.; JANOWICZ, K. Linked Data - A Paradigm Shift for *Geographic Information Science*. p. 173–186, 2014.

LESSCHEN, J. P.; VERBURG, P. H.; STAAL, S. J. *Statistical methods for analysing the spatial dimension of changes in land use and farming systems*. [s.l.] International Livestock Research Institute Kenya, 2005.

MARIA DE ALMEIDA, C. et al. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Computers, Environment and Urban Systems*, v. 27, n. 5, p. 481–509, 1 set. 2003.

MEIYAPPAN, P. et al. Spatial modeling of agricultural land use change at global scale. *Ecological Modelling*, v. 291, p. 152–174, 2014.

MOREIRA, E. et al. Dynamical coupling of multiscale land change models. *Landscape Ecology*, v. 24, n. 9, 2009.

PARKER, D. C. et al. Agent-based models of land-use and land-cover change. *LUCC Report Series*, v. 6, n. 6, 2002.

PONTIUS, R. G. et al. Lessons and challenges in land change modeling derived from synthesis of cross-case comparisons. In: *Trends in spatial analysis and modelling*. [s.l.] Springer, 2018. p. 143–164.

PURSS, M. B. et al. *The OGC® Discrete Global Grid System core standard: A framework for rapid geospatial integration*. In: 2016 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS). IEEE, 2016.

SAHR, K.; WHITE, D. Discrete global grid systems. *Computing Science and Statistics*, p. 269–278, 1998.

SOARES-FILHO, B. S.; CERQUEIRA, G. C.; PENNACHIN, C. L. DINAMICA—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological modelling*, v. 154, n. 3, p. 217–235, 2002.

SUBEDI, P.; SUBEDI, K.; THAPA, B. Application of a hybrid cellular automaton–Markov (CA-Markov) model in land-use change prediction: a case study of Saddle Creek Drainage Basin, Florida. *Applied Ecology and Environmental Sciences*, v. 1, n. 6, p. 126–132, 2013.

TURNER, B. et al. Land-Use and Land-Cover Change: science/research plan. [No source information available], 1 jan. 1995.

VAN ASSELEN, S.; VERBURG, P. H. Land cover change or land- use intensification: simulating land system change with a global- scale land change model. *Global change biology*, v. 19, n. 12, p. 3648–3667, 2013.

VAN SCHROJENSTEIN LANTMAN, J. et al. Core Principles and Concepts in Land-Use Modelling: A Literature Review. In: KOOMEN, E.; BORSBOOM-VAN BEURDEN, J. (Eds.). *Land-Use Modelling in Planning Practice*. Dordrecht: Springer Netherlands, 2011. p. 35–57.

VELDKAMP, A.; FRESCO, L. O. CLUE: a conceptual model to study the conversion of land use and its effects. *Ecological modelling*, v. 85, n. 2–3, p. 253–270, 1996.

VERBURG, P. H. et al. Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environmental management*, v. 30, n. 3, p. 391–405, 2002.

VERBURG, P. H. Simulating feedbacks in land use and land cover change models. *Landscape Ecology*, v. 21, n. 8, p. 1171–1183, 2006.

WHITE, R.; ENGELEN, G. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, v. 24, n. 5, p. 383–400, 1 set. 2000.