

PROGRAMA INSTITUCIONAL DE BOLSAS - INPE/CNPq

FORMULÁRIO PARA SOLICITAÇÃO DE BOLSA

(Este formulário deverá ser preenchido e encaminhado à secretaria com toda a documentação exigida. A falta de um dos documentos implicará na NÃO ACEITAÇÃO do pedido).

Modalidade: PIBIC	<input checked="" type="checkbox"/> Nova	<input type="checkbox"/> Renovação
Modalidade: PIBITI	<input type="checkbox"/> Nova	<input type="checkbox"/> Renovação

Orientador(a): Pedro Ribeiro de Andrade Neto	
CPF: 803.448.525-04	Data de Nascimento: 02/11/1981
Coordenação/Departamento: CCST	Ramal: 7798
E-mail: pedro.andrade@inpe.br	

Bolsista: Fabricio Rodrigues Segui Aparisi			
Título do Projeto: Mapas de Custo de Transporte de <i>Commodities</i> Agrícolas para o Brasil			
Palavras-Chaves:	1. GPM	2. Estradas	3. Algoritmos genéticos

Objetivos Específicos:

Um dos principais fatores que levam a decisões de conversão de uso da terra é o acesso a mercados (Aguiar et. al, 2007). Modelos computacionais de uso da terra, como por exemplo o Dinâmica (Soares-Filho et al., 2002) e o CLUE-S (Verburg et al., 2002), usam estes dados tanto para simular o passado quanto para estudar cenários de uso da terra dadas possíveis alterações na infraestrutura de transporte. Para determinar o acesso a mercados é necessário gerar mapas de custos de transporte para a região de estudo que descrevem o custo de transporte de cada localidade do mapa para possíveis destinos de consumo.

O principal objetivo deste projeto é gerar mapas de custos de transporte rodoviário de diferentes commodities agrícolas para o Brasil. Para tanto, os seguintes objetivos se fazem necessários:

1. Construir um banco de dados espaço-temporal contendo a infraestrutura de transporte rodoviário atual e planejada do Brasil.
2. A partir do banco de dados produzido, gerar dados de custo de transporte usando uma implementação da Matriz de Vizinhança Generalizada (Aguiar et. al, 2003).
3. Calibrar e validar os dados gerados usando como referência dados reais de custos de transporte entre municípios.

Metodologia:

1) Banco de dados

Neste projeto, será construído um banco de dados espaço-temporal contendo as principais infraestruturas de transporte rodoviário do Brasil relativos ao modal rodoviário. Serão coletados dados do Plano Nacional de Logística e Transportes (PNLT, 2007), as obras do PAC e PAC 2, bem como dados de planejamento dos governos estaduais. Este banco de dados também conterà os dados das sedes de municípios e portos de escoamento de produção necessários por serem os destinos finais de transporte e também para o processo de calibração e validação dos dados gerados.

Para a construção do banco de dados, será usado o software aRT (API R-TerraLib) (Andrade et al., 2005), que

permite gerar bancos de dados espaço-temporais TerraLib (Câmara et al., 2008) a partir do programa estatístico R (R Development Core Team, 2008). Serão implementados algoritmos para a verificação da consistência deste banco de dados, como por exemplo para garantir que toda a rede esteja conectada, pois muitos dos dados de infraestrutura de transporte são desenvolvidos para serem visualizados, não possuindo a consistência necessária para se executar algoritmos sobre grafos.

2) Matriz de Vizinhança Generalizada

Para gerar dados de custo de transporte a partir da infraestrutura de transporte, este projeto usará a ideia de Matriz de Vizinhança Generalizada (*Generalized Proximity Matrix - GPM*) (Aguiar et. al, 2003). A GPM parte do pressuposto que um espaço Euclidiano não é suficiente para descrever as complexas relações que ocorrem em espaços geográficos. A GPM apresenta então um conjunto de algoritmos para computar distâncias, usando redes de conectividade representadas através de grafos que distorcem o espaço Euclidiano. Neste projeto usaremos a estratégia da GPM denominada *rede aberta*, permitindo que uma caminho entre a origem e o destino possa entrar e sair da rede em qualquer ponto. A GPM possui uma série de parâmetros que podem ser modificados conforme o objetivo do trabalho, como por exemplo ponderação do caminho dentro e fora da rede e ponderação do custo dentro da rede de acordo com propriedades dos seus objetos (usada, por exemplo, para diferenciar estradas pavimentadas e não pavimentadas).

A GPM possui uma implementação no software aRT, o mesmo usado na construção do banco de dados. A linguagem R, por usar conceitos de programação funcional, facilita a parametrização da GPM, uma vez que alguns dos seus parâmetros são funções. Este projeto usará a implementação atual da GPM no pacote aRT para gerar dados de custo de transporte entre municípios.

3) Calibrar e validar os dados gerados

Para calibrar e validar os dados gerados, serão usados os dados reais de fretes rodoviários disponibilizados pelo Sistema de Informações de Fretes (SIFRECA), desenvolvido pela ESALQ/USP (SIFRECA, 2012). O SIFRECA disponibiliza dados de custo de transporte entre pares de municípios para diferentes commodities. Neste projeto, serão usados os custos das seguintes commodities: algodão, carne bovina, soja, milho e trigo. Para cada uma das *commodities* estudadas, a GPM será usada para gerar dados de custo de transporte envolvendo todos os pares de municípios disponíveis no banco de dados do SIFRECA. Os dados do SIFRECA serão divididos em dois subconjuntos. O primeiro subconjunto será usado para a calibração da GPM e o segundo para a validação. Para a calibração, será usado um método de correlação estatística para computar o ajuste entre os dados gerados pela GPM e os dados do SIFRECA, para cada uma das *commodities*. Será então executada uma calibração usando algoritmos genéticos através do pacote GA (*Genetic Algorithms*) do R (Scrucca, 2013) para determinar os parâmetros da GPM que produzem dados que mais se ajustam aos custos do SIFRECA. O segundo subconjunto será usado para validar os resultados produzidos pela GPM calibrada com o primeiro subconjunto.

Plano de Trabalho:

O plano de trabalho deste projeto de iniciação científica é composto das seguintes atividades:

1. **Estudar o software e linguagem R e bancos de dados espaciais.** Para o desenvolvimento deste projeto de pesquisa, é necessário conhecimento da linguagem de programação R, bem como as estruturas de dados para representação espacial disponíveis no pacote sp (Pebesma e Bivand, 2005) e as funcionalidades do pacote aRT (Andrade et al., 2005) para a construção de bancos de dados espaciais. Adicionalmente, é necessário compreender conceitos de geoprocessamento para construir o banco de dados e validar a sua estrutura. Estes tópicos serão estudados na primeira atividade.
2. **Construir e validar banco de dados espaço-temporal.** A segunda atividade deste plano de trabalho consiste em coletar dados de transporte, incluindo as infraestruturas existentes e novas infraestruturas planejadas pelo governo federal e por governos estaduais. Os dados serão organizados em um banco de dados TerraLib através de um script R usando o pacote aRT. Este script deverá implementar uma série de funções para validação dos dados, de forma que a garantir que a GPM possa ser executada sem erros.
3. **Computar custos entre pares de municípios, calibrar e validar a GPM.** Para cada uma das *commodities* estudadas, serão gerados dados de custo de transporte envolvendo todos os pares de municípios que possuem dados de custo de transporte disponíveis no SIFRECA. Os dados produzidos pela GPM serão então calibrados e validados conforme a metodologia descrita previamente.
4. **Gerar os mapas de transporte e cenários futuros.** A partir da validação do algoritmo com os custos de transporte entre municípios, serão gerados mapas de custo de transporte atual e cenários futuros (usando a

infraestrutura planejada). Serão gerados mapas com a resolução de 25x25km tendo como destinos o porto de exportação mais próximo, a capital mais próxima e a cidade de São Paulo, o maior centro consumidor do Brasil.

5. **Escrever relatório técnico.** Escrever um relatório técnico descrevendo onde os dados foram adquiridos, como replicar o banco de dados desenvolvido, bem como todo o processo para reprodução das simulações, como parâmetros usados e estratégia de calibração e validação dos dados.
6. **Desenvolver página web.** Os resultados deste projeto serão de domínio público. Para atender a este objetivo, será desenvolvida uma webpage disponibilizando todos os dados e os resultados deste projeto científico.
7. **Escrever artigo científico.** Escrever um artigo científico para conferência nacional descrevendo os principais resultados do projeto.

Etapas Concluídas: (Preencher somente no caso de renovação)

Etapas a Concluir: (Preencher somente no caso de renovação)

Cronograma de Atividades:

	Ago/Set	Out/Nov	Dez/Jan	Fev/Mar	Abr/Mai	Jun/Jul
Estudar o software e linguagem R e bancos de dados espaciais						
Construir e validar banco de dados espaço-temporal						
Computar custos entre pares de municípios, calibrar e validar a GPM						
Gerar os mapas de transporte e cenários futuros						
Escrever relatório técnico						
Desenvolver página web						
Escrever artigo científico						
Escrever relatório final						

Resultados Esperados:

Ao término do projeto, os seguintes resultados são esperados:

1. Um banco de dados espaço-temporal contendo as principais infraestruturas de transportes do Brasil separadas por modais, bem como as principais infraestruturas planejadas.
2. Mapas de custo de transporte para o Brasil com custos para capitais e portos de exportação, para diferentes *commodities*, disponíveis gratuitamente em uma página na Internet. Estes mapas serão usados nos modelos de uso da terra desenvolvidos no próprio instituto, podendo ter outras aplicações tais como estimativas de emissão de gases do efeito estufa no transporte de *commodities*.
3. Um relatório técnico descrevendo todo o processo de construção do banco de dados e de geração dos mapas de custo de transporte.
4. Um artigo científico com os principais resultados do projeto.

Referências Bibliográficas:

- Aguiar, A. P. D., G. Câmara, and M. I. S. Escada. *Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity*. Ecological Modelling 209.2 (2007): 169-188.
- Aguiar, A. P. D., G. Camara, A. M. V. Monteiro, R. C. M. Souza. *Modelling Spatial Relations by Generalized Proximity Matrices*. GeoInfo. 2003.
- Andrade, P. R., P. J. Ribeiro Jr, K. D. Fook. *Integration of Statistics and Geographic Information Systems: the R/TerraLib Case*. GeoInfo. 2005.
- Câmara, G., Vinhas, L., Ferreira, K. R., De Queiroz, G. R. et al. *TerraLib: An open source GIS library for large-scale environmental and socio-economic applications*. Open source approaches in spatial data handling. Springer Berlin Heidelberg, 2008. 247-270.
- Pebesma, E., R. S. Bivand. *S Classes and Methods for Spatial Data: the sp Package*. 2005. http://cran.r-project.org/web/packages/sp/vignettes/intro_sp.pdf. Acessado em 22-06-2014.
- PNLT, *Plano Nacional de Logística e Transportes - Relatório Executivo* (2007). <http://transportes.gov.br/index/conteudo/id/3280>. Acessado em 25-06-2014.
- R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>. Acessado em 22-06-2014.
- Scrucca, L. *GA: A Package for Genetic Algorithms in R*. Journal of Statistical Software 53(4). 2013.
- Sifreca - *Sistema de Informações de Fretes - Anuário 2012*. Departamento de Economia, Administração e Sociologia. Escola Superior de Agricultura Luiz de Queiroz (ESALQ/USP), 2012. <http://esalqlog.esalq.usp.br/files/biblioteca/606.pdf>. Acessado em 25-06-2014.
- Soares-Filho, B. S., Coutinho Cerqueira, G., & Lopes Pennachin, C. (2002). *DINAMICA—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier*. Ecological modelling, 154(3), 217-235.
- Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., & Mastura, S. S. (2002). *Modeling the spatial dynamics of regional land use: the CLUE-S model*. Environmental management, 30(3), 391-405.

Local: _____ Data: ____ / ____ / ____

Assinatura do Orientador(a)

PARECER DO COMITÊ INTERNO: () APROVADO () REPROVADO

Nome do(a) Avaliador(a): _____

Observações: _____
