

UNIVERSIDADE FEDERAL DO PARANÁ

HENRIQUE PESCHL

DETERMINAÇÃO DO POTENCIAL DE TRANSIÇÃO COM RANDOM FOREST EM
PROCESSOS DE EXPANSÃO URBANA EM CURITIBA E VETOR LESTE DA RMC

CURITIBA
2024

UNIVERSIDADE FEDERAL DO PARANÁ

HENRIQUE PESCHL

DETERMINAÇÃO DO POTENCIAL DE TRANSIÇÃO COM RANDOM FOREST EM
PROCESSOS DE EXPANSÃO URBANA EM CURITIBA E VETOR LESTE DA RMC

Dissertação apresentada ao curso de Pós-Graduação
em Ciências Geodésicas, Setor de Ciências da Terra,
Universidade Federal do Paraná, como requisito
parcial à obtenção de título de Mestre em Ciências
Geodésicas.

Orientador: Prof. Dr. Hideo Araki

CURITIBA
2024

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

Peschl, Henrique

Determinação do potencial de transição com random forest em processos de expansão urbana em Curitiba e vetor leste da RMC / Henrique Peschl. - Curitiba, 2024.

1 recurso on-line: PDF.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-Graduação em Ciências Geodésicas.

Orientador: Prof. Dr. Hideo Araki

1. Crescimento urbano. 2. Imagens digitais. 3. Autômato celular.
I. Araki, Hideo. II. Universidade Federal do Paraná. III. Programa de Pós-Graduação em Ciências Geodésicas. IV. Título.



TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **HENRIQUE PESCHL** intitulada: **DETERMINAÇÃO DO POTENCIAL DE TRANSIÇÃO COM RANDON FOREST EM PROCESSOS DE EXPANSÃO URBANA EM CURITIBA E VETOR LESTE DA RMC**, sob orientação do Prof. Dr. HIDEO ARAKI, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 21 de Fevereiro de 2024.

Assinatura Eletrônica

22/02/2024 18:18:54.0

HIDEO ARAKI

Presidente da Banca Examinadora

Assinatura Eletrônica

27/02/2024 16:19:15.0

CRISTINA DE ARAÚJO LIMA

Avaliador Externo ()

Assinatura Eletrônica

23/02/2024 17:25:38.0

RODRIGO DE CAMPOS MACEDO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Este trabalho é dedicado:

Aos meus pais.

Ao meu irmão.

A minha companheira.

Por todo amor, carinho,

paciência e apoio.

AGRADECIMENTOS

Ao meu orientador, Prof. Dr Hideo Araki, por todo ensinamento, pelo aprendizado, pela confiança e incentivo durante toda minha jornada acadêmica.

Aos professores da banca avaliadora, Prof.^a Dra. Cristina de Araújo Lima e Prof. Dr. Rodrigo de Campos Macedo, pelas sugestões, correções e contribuições.

Aos amigos Bruna e Guilherme, por todo carinho e suporte ao longo de toda esta jornada.

A todos os colegas e amigos, em especial ao Lucas, Nathan, Leonardo, Thais, Lisandra pela convivência e companheirismo.

Aos professores e técnicos do curso de Pós-Graduação em Ciências Geodésicas por toda a ajuda e ensinamentos.

A Universidade Federal do Paraná (UFPR) e ao curso de Engenharia Cartográfica e de Agrimensura, por toda a experiência e oportunidades oferecidas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos.

“The very nature of science is discoveries,
and the best of those discoveries are the
ones you don’t expect.”

Neil deGrasse Tyson

RESUMO

A partir do século XVIII a humanidade presenciou um deslocamento significativo de pessoas para os centros urbanos. Em 2010 a população da Região Metropolitana de Curitiba (RMC) era de 3,2 milhões de pessoas, em 2022 esse número chegou a 3,5 milhões de pessoas (IBGE, 2022). Maior ainda tem sido o crescimento da população excluída, alocada em condições vulneráveis de habitação. A rápida urbanização é motivo de uma crescente preocupação. Há uma necessidade urgente de desenvolver estratégias adequadas para gerenciar o crescimento urbano de forma sustentável (UN-Habitat, 2020). Modelos de simulação da expansão urbana é uma forma de compreender as cidades para fornecer dados para tomadores de decisões e planejamento urbano. Entre estes, os autômatos celulares (AC) têm sido amplamente aplicados (ZHOU et al., 2019). A região leste da RMC e Curitiba são um cenário relevante para estudos de expansão urbana devido ao rápido crescimento destas cidades e locais de importantes mananciais. O objetivo principal deste trabalho é desenvolver um modelo de simulação da expansão urbana com autômatos celulares, obtendo regras de transição por meio do *Random Forest* (RF). Uma das variáveis de entrada no modelo são a de área urbana, esta foi obtida da classificação das imagens *Landsat* da região, classificação que foi realizada por meio do classificador RF. Por meio da literatura e testes, outras variáveis, chamadas fatores direcionadores, foram produzidas. Estas variáveis e fatores direcionadores foram utilizados para determinar a regressão RF, e deste obter o potencial de transição de todos os pixels não urbanos de se tornarem urbanos depois de um período. Esse potencial de transição foi utilizado como regra de transição no modelo de AC, implementado em linguagem de programação *Python*. O modelo AC também inclui uma restrição para áreas onde não deve ocorrer o crescimento urbano, que neste caso são os corpos de água, e deste modelo obteve-se a simulação final da expansão urbana para um período de 5 anos em 4 diferentes períodos. Analisando as curva ROC (*Receiver Operating Characteristic*) para os resultados obtidos da regressão RF, é possível afirmar que o desempenho do modelo é bom, com o melhor valor de Área Sob a Curva (AUC) sendo 0,94. A matriz de erros entre a paisagem real e simulada mostra uma acurácia total alta, a menor sendo de 97,26%, e uma acurácia aceitável para novos *pixels* urbanos, a maior sendo de 45%. O valor da figura de mérito (FoM) se situa em patamares próximos a outros estudos (LV et al., 2021; KAMUSOKO e GAMBIA, 2015; CHEN, 2022), atingindo o melhor valor de 0,29. A finalidade deste estudo foi de simular a expansão urbana com AC e RF e buscar diferentes fatores direcionadores que melhor modelam a expansão. Dentre as diversas métricas de avaliação utilizadas pode-se afirmar que os resultados do estudo foram bons, mostrando uma grande capacidade do RF de produzir um potencial de transição fiel à realidade.

Palavras-chave: Modelagem Urbana, Expansão Urbana, Autômatos Celulares, *Random Forest*.

ABSTRACT

Since the XVIII century, humanity has witnessed a movement of people to urban centers. In 2010, the population of the Metropolitan Region of Curitiba (RMC) was 3.2 million people, in 2022 this number reached 3.5 million people (IBGE, 2022). Even greater has been the growth of the excluded population, allocated to vulnerable housing conditions. Rapid urbanization is a reason for growing concern. There is an urgent need to develop appropriate strategies to manage urban growth sustainably (UN-Habitat, 2020). Simulation models of urban expansion are a way of understanding cities to provide data for decision makers and urban planning. Among these, cellular automata (CA) have been widely applied (ZHOU et al., 2019). The eastern region of RMC and Curitiba are a relevant scenario for urban expansion studies due to the rapid growth of these cities and locations of important water sources. The main objective of this work is to develop a simulation model of urban expansion with cellular automata, obtaining transition rules through Random Forest (RF). One of the input layers in the model is the urban area, this was obtained from the classification of Landsat images of the region, a classification that was carried out using the RF classifier. Through literature and testing, other layers, called driving factors, were produced. These layers and driving factors were used to determine an RF regression, and from this obtain the transition potential of all non-urban pixels to become urban after a period. This transition potential was used as a transition rule in the CA model, implemented in the Python programming language. The CA model also includes a restriction for areas where urban growth should not occur, which in this case are bodies of water, and from this model the final simulation of urban expansion was obtained for a period of 5 years in 4 different seasons. Analyzing the ROC curves for the results obtained from the RF regression, it is possible to state that the model's performance is good, with the best AUC value being 0.94. The error matrix between the real and simulated landscape shows a high total accuracy, the lowest being 97.26%, and an acceptable accuracy for new urban pixels, the highest being 45%. The value of the figure of merit (FoM) is close to other studies, reaching the best value of 0.29. The purpose of this study was to simulate urban expansion with CA and RF and look for different driving factors that best model expansion. Among the various evaluation metrics used, it can be stated that the results of the study were good, showing the RF's great capacity to produce a transition potential faithful to reality.

Keywords: Urban Modeling, Urban Expansion, Cellular Automata, Random Forest

LISTA DE FIGURAS

FIGURA 1 - PRIMEIRA IMPLEMENTAÇÃO DO CONSTRUTOR AUTO-REPLICANTE DE VON NEUMANN.....	27
FIGURA 2 - EXEMPLO DE EXECUÇÃO DO JOGO DA VIDA.....	28
FIGURA 3 - ESTRUTURA DO SIMLUCIA.....	31
FIGURA 4 - EXEMPLO DE ÁRVORE DE DECISÃO.....	33
FIGURA 5 - FUNCIONAMENTO DO RANDOM FOREST.....	34
FIGURA 6 - PROPORÇÃO ENTRE A MUDANÇA OBSERVADA E PREVISTA.....	35
FIGURA 7 – EXEMPLO DE MATRIZ DE ERROS.....	36
FIGURA 8 - EXEMPLO DE CURVA ROC.....	37
FIGURA 9 - ÁREA DE ESTUDO.....	42
FIGURA 10 - FLUXOGRAMA GERAL DO ESTUDO.....	43
FIGURA 11 - FLUXOGRAMA DA CLASSIFICAÇÃO.....	43
FIGURA 12 - EXEMPLO DA COLETA DE AMOSTRAS.....	44
FIGURA 13 - RECORTE DE UM EXEMPLO DE ÁRVORE DE DECISÃO DO RF.....	47
FIGURA 14 - RECORTE DA IMAGEM DE SATÉLITE, SUA CLASSIFICAÇÃO E A CLASSIFICAÇÃO MELHORADA POR MAIS PONTOS DE AMOSTRA.....	51
FIGURA 15 - RECORTE DA IMAGEM CLASSIFICADA E A MESMA COM FILTRO APLICADO.....	52
FIGURA 16 - IMPORTÂNCIA DAS VARIÁVEIS PARA AS CLASSIFICAÇÕES DE 2000, 2005, 2010, 2015, 2020.....	53
FIGURA 17 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2000.....	54
FIGURA 18 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2005.....	55
FIGURA 19 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2010.....	55
FIGURA 20 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2015.....	56
FIGURA 21 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2020.....	56
FIGURA 22 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2000.....	59
FIGURA 23 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2005.....	59

FIGURA 24 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2010.....	60
FIGURA 25 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2015.....	60
FIGURA 26 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2020.....	61
FIGURA 27 – VARIÁVEL DE MANCHA URBANA DO ANO 2000, 2005, 2010, 2015, 2020.....	63
FIGURA 28 – MANCHA URBANA QUE CRESCEU NO PERÍODO DE 5 ANOS.....	64
FIGURA 29 – VARIÁVEL DE DISTÂNCIA DAS RODOVIAS.....	65
FIGURA 30 – VARIÁVEL DE DISTÂNCIA DAS RUAS PRIMÁRIAS.....	65
FIGURA 31 – VARIÁVEL DE DISTÂNCIA DAS RUAS SECUNDÁRIAS.....	66
FIGURA 32 – VARIÁVEL DE DISTÂNCIA DAS RUAS TERCIÁRIAS.....	66
FIGURA 33 – VARIÁVEL DE DISTÂNCIA DAS ZONAS CENTRAIS.....	67
FIGURA 34 - VARIÁVEL DE DISTÂNCIA DA MANCHA URBANA.....	67
FIGURA 35 - VARIÁVEL DE VIZINHOS URBANOS.....	68
FIGURA 36 - DENSIDADE POPULACIONAL PARA O ANO DE 2000.....	68
FIGURA 37 - DENSIDADE POPULACIONAL PARA O ANO DE 2010.....	69
FIGURA 38 - DECLIVIDADE EM GRAUS DA ÁREA DE ESTUDO.....	69
FIGURA 39 - CORPOS D'ÁGUA DA ÁREA DE ESTUDO.....	70
FIGURA 40 - IMPORTÂNCIA DAS VARIÁVEIS PARA AS MODELAGENS DE 2005, 2010, 2015, 2020.....	72
FIGURA 41 - POTENCIAL DE TRANSIÇÃO PARA 2005.....	73
FIGURA 42 - POTENCIAL DE TRANSIÇÃO PARA 2010.....	74
FIGURA 43 - POTENCIAL DE TRANSIÇÃO PARA 2015.....	74
FIGURA 44 - POTENCIAL DE TRANSIÇÃO PARA 2020.....	75
FIGURA 45 - CURVAS ROC PARA AS MODELAGENS DE 2005, 2010, 2015 E 2020.....	76
FIGURA 46 - COMPARAÇÃO ENTRE O SIMULADO E O REAL DE 2000 PARA 2005.....	77

FIGURA 47 - COMPARAÇÃO ENTRE O SIMULADO E O REAL DE 2005 PARA 2010.....	77
FIGURA 48 - COMPARAÇÃO ENTRE O SIMULADO E O REAL DE 2010 PARA 2015.....	78
FIGURA 49 - COMPARAÇÃO ENTRE O SIMULADO E O REAL DE 2015 PARA 2020.....	78

LISTA DE TABELAS

TABELA 1 - NÚMERO DE HABITANTES DE CURITIBA E VETOR LESTE.....	18
TABELA 2 - MODELOS E FERRAMENTAS DE MUDANÇA NO USO DO SOLO....	29
TABELA 3 - IMAGENS UTILIZADAS.....	39
TABELA 4 - CARACTERÍSTICAS DO LANDSAT 5.....	39
TABELA 5 - CARACTERÍSTICAS DO LANDSAT 8.....	39
TABELA 6 - ARQUIVOS VETORIAIS UTILIZADOS.....	40
TABELA 7 - PONTOS DE AMOSTRAS COM SUAS RESPECTIVAS REFLECTÂNCIAS EM DIFERENTES BANDAS E ÍNDICES.....	50
TABELA 8 - VALORES DE RMS PARA DIFERENTES PROFUNDIDADES DE ÁRVORES DO RF E SUA PLOTAGEM EM UM GRÁFICO.....	52
TABELA 9 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2000..	57
TABELA 10 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2005	57
TABELA 11 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2010	58
TABELA 12 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2015	58
TABELA 13 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2020	58
TABELA 14 - ÁREA URBANA NA ÁREA DE ESTUDO.....	62
TABELA 15 - TABELA USADA NO TREINAMENTO PARA O PERÍODO 2000 - 2005..	71
TABELA 16 - MATRIZ DE ERROS PARA A SIMULAÇÃO DE 2005.....	79
TABELA 17 - MATRIZ DE ERROS PARA A SIMULAÇÃO DE 2010.....	79
TABELA 18 - MATRIZ DE ERROS PARA A SIMULAÇÃO DE 2015.....	80
TABELA 19 - MATRIZ DE ERROS PARA A SIMULAÇÃO DE 2020.....	80
TABELA 20 - VALORES DE FOM PARA OS 4 PERÍODOS SIMULADOS.....	80
TABELA 21 - COMPARAÇÃO DE VALORES DE FOM COM OUTROS ESTUDOS	80

LISTA DE ABREVIATURAS

AC	- Autômatos Celulares
AUC	- <i>Area Under the Curve</i>
COMEC	- Coordenação da Região Metropolitana de Curitiba
FoM	- <i>Figure of Merit</i>
IAT	- Instituto Água e Terra
IBGE	- Instituto Brasileiro de Geografia e Estatística
ONU	- Organização das Nações Unidas
RF	- <i>Random Forest</i>
RMC	- Região Metropolitana de Curitiba
RMS	- <i>Root Mean Square</i>
ROC	- <i>Receiver Operating Characteristic</i>

SUMÁRIO

1 INTRODUÇÃO.....	17
1.1 OBJETIVOS.....	20
1.1.1 Objetivo geral.....	20
1.1.2 Objetivos específicos.....	20
2 REVISÃO DE LITERATURA.....	21
2.1 ÁREAS URBANAS E ESPAÇOS URBANOS.....	21
2.2 URBANIZAÇÃO.....	21
2.2.1 Urbanização de Curitiba e região leste da RMC.....	22
2.3 MODELOS DE SIMULAÇÃO DE DINÂMICA ESPACIAL.....	25
2.3.1 Autômatos Celulares.....	27
2.4 EXEMPLOS DE MODELOS PARA O CRESCIMENTO URBANO COM AUTÔMATOS CELULARES.....	28
2.4.1 SLEUTH.....	29
2.4.2 SimLucia.....	30
2.4.3 CLUE-S.....	31
2.4.4 Dinamica EGO.....	32
2.5 ÁRVORE DE DECISÃO E FLORESTA ALEATÓRIA.....	32
2.6 MÉTODOS DE AVALIAÇÃO.....	35
2.6.1 Figura de Mérito.....	35
2.6.2 Matriz de Erros.....	35
2.6.3 Curva ROC.....	37
2.7 ÍNDICES ESPECTRAIS.....	38
3 MATERIAL E MÉTODOS.....	39
3.1 MATERIAIS.....	39
3.1.1 Área de Estudo.....	41
3.2 MÉTODOS.....	42
3.2.2 Classificação.....	43
3.2.3 Criação das Variáveis e Fatores Direcionadores.....	44

3.2.4 Determinação do Potencial de transição com Random Forest.....	47
3.2.5 Desenvolvimento do Modelo de Autômato Celular.....	47
4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS.....	49
4.1 CLASSIFICAÇÃO DAS IMAGENS LANDSAT.....	49
4.1.1 Melhorando a Classificação com Novas Amostras.....	50
4.1.2 Melhorando a Classificação Testando Diferentes Profundidades no RF...	52
4.2 AVALIAÇÃO DOS RESULTADOS DA CLASSIFICAÇÃO.....	53
4.3 VARIÁVEIS E FATORES DIRECIONADORES.....	61
4.4 SIMULAÇÃO.....	70
5 CONCLUSÃO.....	82
5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	83
REFERÊNCIAS.....	84

1 INTRODUÇÃO

A partir do século XVIII, a humanidade presenciou um deslocamento significativo de pessoas para os centros urbanos. No ano de 2018 mais de 55% da população mundial vivia em áreas urbanas e a projeção é que esse valor chegue a 60% em 2030 segundo a Organização das Nações Unidas (ONU, 2018). A rápida urbanização é motivo de uma crescente preocupação dos planejadores urbanos e tomadores de decisão. Desmatamento, incremento de consumo de energia, poluição do ar (LI et al., 2021), expansão urbana em áreas importantes de biodiversidade e mananciais (MESTA et al., 2022) causam impactos ambientais que afetam a saúde, segurança e o bem-estar das pessoas. Em 2018, 60% dos municípios com 500.000 ou mais habitantes estavam altamente expostas a pelo menos um de seis desastres naturais (ciclones, inundações, secas, terremotos, deslizamentos de terra e erupções vulcânicas), e esse número está crescendo (ONU, 2018).

No solo brasileiro, a urbanização experimentou um salto significativo, aumentando de 36% para 84% da população que passou a ser urbana entre 1950 e 2015, conforme indicado pelos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD). Na Região Metropolitana de Curitiba (RMC) em 2010 a população era de 3,2 milhões de pessoas, em 2022 esse número chegou a 3,5 milhões de habitantes (IBGE, 2023). As cidades de Piraquara, Pinhais e São José dos Pinhais, com importantes mananciais que abastecem a RMC, tiveram juntas um aumento de 100 mil habitantes nos últimos 12 anos (aumento de 21%). Apenas São José do Pinhais teve um aumento de 65 mil habitantes.

O processo de ocupação e configuração da aglomeração metropolitana foi intenso e rápido no século XX, transpondo as fronteiras administrativas de Curitiba e criando áreas contíguas de ocupação nos municípios vizinhos (Sanepar, 2013). Contudo, maior ainda tem sido o crescimento da população excluída, alocada em condições vulneráveis de habitação. No município de Piraquara o índice de crescimento de condições vulneráveis de habitação entre 1992 e 1997 aumentou em 69,81% e em São José dos Pinhais aumentou 34,93% neste mesmo período. Já na Região Metropolitana de Curitiba, o aumento foi de 6,90% no mesmo período (COMEC, 1999). Processo impulsionado por uma legislação de parcelamento do solo mais flexível, oferecendo estas áreas a preços reduzidos contudo sem

infraestrutura (SILVA, 2012). Conforme Silva (2012), na metrópole de Curitiba, os assentamentos informais subiram de 167 assentamentos e 33.778 domicílios em 1990 para 341 assentamentos e 59.064 domicílios, em 2005.

TABELA 1 - NÚMERO DE HABITANTES DE CURITIBA E VETOR LESTE

Cidade	Ano		Diferença	Aumento
	2010	2022		
Curitiba	1.751.907	1.773.733	21.826	1,25%
Pinhais	117.008	127.019	10.011	8,56%
Piraquara	93.207	118.730	25.523	27,38%
São José Dos Pinhais	264.210	329.222	65.012	24,61%
Total	2.226.332	2.348.704	122.372	5,50%

FONTE: IBGE, 2010 e 2022.

Além disso, segundo Hallegatte (2020 citado por MESTA et al., 2022) foi evidenciado que as pessoas com baixa renda sofrem desproporcionalmente os efeitos de desastres induzidos por riscos naturais devido à menor resiliência socioeconômica (por exemplo, menos apoio de instrumentos financeiros, como seguros e esquemas de proteção social).

Portanto, há uma necessidade urgente de desenvolver estratégias adequadas para gerenciar o crescimento urbano de forma sustentável (UN-Habitat, 2020), a ênfase na pesquisa detalhada sobre áreas urbanas pode ser atribuída ao fato de que essas regiões estão destinadas a abrigar a maior parte da população do planeta. É fundamental termos dados para conhecer regiões urbanizadas e em urbanização, de modo que pesquisadores possam compreender melhor os impactos ambientais da urbanização (LI et al., 2021), desenvolver ferramentas para antecipar prováveis eventos que terão lugar no ambiente urbano em um futuro próximo (ALMEIDA, 2004), e criar estratégias de inclusão da população economicamente vulnerável e de mitigação de desastres naturais.

Simulações urbanas podem oferecer diretrizes para planos setoriais em nível intra urbano, incluindo planos de localização industrial, loteamentos residenciais, equipamentos sociais, e infraestrutura urbana, como rotas de ônibus, pavimentação de vias e expansão da rede de água e esgoto, expansão do fornecimento de energia, entre outros (ALMEIDA, 2004).

A comunidade de pesquisadores da Ciências da Terra reconhece os modelos como ferramentas poderosas, capazes de contribuir significativamente para a

compreensão do comportamento de um sistema. Esses modelos desempenham um papel crucial ao auxiliar na tomada de decisões e no desenvolvimento de políticas que impactam nosso planeta e sociedade em escalas que variam do local ao global (GOMEZ & JONES, 2010; citado por VIEGAS et al. (2021).

Com os rápidos avanços nas técnicas de sensoriamento remoto e sistemas de informação geográfica (SIG), uma gama diversificada de modelos foi desenvolvida e aprimorada para simular e prever o crescimento urbano nas últimas décadas. Entre estes, os autômatos celulares (AC) têm sido amplamente aplicados (ZHOU et al., 2019).

Modelos de AC tornaram-se populares em grande parte devido ao fato de apresentarem uma incrível simplicidade operacional, gerarem uma dinâmica que pode reproduzir processos tradicionais de mudanças por difusão, e conterem complexidade suficiente para simular mudanças inesperadas como as observadas em fenômenos emergentes (ALMEIDA, 2004). Um modelo relevante neste sentido é o SLEUTH.

O modelo SLEUTH destaca-se como uma ferramenta amplamente reconhecida para simulações de crescimento urbano e mudança no uso do solo (CHAUDHURI & CLARKE, 2013). Este modelo incorpora diversas variáveis que influenciam a dinâmica das mudanças no uso do solo como declividade, uso do solo, áreas excluídas, áreas urbanas, mapas de transporte e áreas montanhosas.

Em estudos de simulação de expansão urbana observa-se a aplicação do algoritmo *Random Forest* (RF), utilizado tanto para classificação quanto para regressão, visando inserir a estocasticidade no modelo. O RF permite a obtenção de um indicador conhecido como potencial de transição que representa, em uma escala de 0 a 1, a chance de uma área não urbana se tornar urbana após um período. As variáveis presentes no SLEUTH podem ser empregadas no RF para determinar esse potencial, sendo utilizado no modelo de AC para realizar a simulação, que é a proposta neste estudo.

Todo o processo para atingir esta proposta será desenvolvido usando uma linguagem de programação. Diferente de softwares de modelagem da mudança do uso do solo, onde o desenvolvimento do estudo se limita às condições implementadas no software, um criado por meio de programação oferece um controle mais amplo, permite ajustes e personalizações conforme necessário, além de facilitar a integração de novos algoritmos. A compreensão detalhada do

funcionamento interno do modelo é outra vantagem, proporcionando maior transparência. Além disso, a facilidade de reprodução por outros pesquisadores devido à explicitação de todos os passos no código.

A seguir, serão delineados o objetivo geral e os objetivos específicos deste estudo. No capítulo 2, é realizada uma revisão bibliográfica que abrange um breve contexto sobre a urbanização mundial e na área de estudo, uma contextualização sobre Autômatos Celulares (AC) e modelos de simulação de alteração do uso do solo, Random Forest, métricas de avaliação utilizadas e uma breve seção sobre índices espectrais. No capítulo 3, são detalhados os materiais utilizados e a metodologia empregada para a condução deste estudo. O capítulo 4 abrange a apresentação dos experimentos realizados, bem como os resultados obtidos, acompanhados de suas respectivas discussões. No capítulo 5, são apresentadas as considerações finais sobre o estudo, além de serem levantadas questões que podem ser exploradas em futuras pesquisas

1.1 OBJETIVOS

1.1.1 Objetivo geral

Desenvolver um modelo de simulação da expansão urbana com autômatos celulares, obtendo o potencial de transição com o Random Forest e verificar sua aplicabilidade para Curitiba e região leste da RMC (Pinhais, Piraquara e São José dos Pinhais).

1.1.2 Objetivos específicos

1. Analisar a importância das variáveis que potencialmente influenciam nas mudanças da paisagem da região;
2. Classificar as imagens *Landsat* da área de estudo para usar na modelagem;
3. Determinar o potencial de transição para a classe Urbano da região de estudo;
4. Avaliar os desempenhos da classificação e do modelo para simulação da expansão urbana.

2 REVISÃO DE LITERATURA

2.1 ÁREAS URBANAS E ESPAÇOS URBANOS

O urbano é uma expressão da organização social, com foco principalmente nas cidades. Suas características abrangem desde a singularidade da paisagem até o perfil demográfico e ocupacional dos residentes e trabalhadores locais (IBGE, 2023).

A definição de áreas urbanas pelos governos geralmente segue quatro abordagens principais: estabelecendo limiares de tamanho (por exemplo, considerando todos os assentamentos com 2.000 habitantes ou mais como urbanos); combinando limiares de tamanho com outros critérios (como densidade populacional ou proporção da população envolvida em atividades não rurais); baseando-se no status administrativo ou político (por exemplo, considerando áreas urbanas todos os assentamentos designados como capitais municipais, estaduais ou nacionais); ou utilizando listas de assentamentos nomeados em censos como áreas urbanas (Almeida, 2004).

Por sua vez, as áreas urbanas de alta e baixa densidade são classificadas segundo o número de domicílios, com aquelas entre 150 e 250 sendo de baixa densidade e as com mais de 250 sendo de alta densidade (IBGE, 2023).

Para a definição dos espaços urbanos, a Coordenação de Geografia do IBGE utilizou como base áreas com alta densidade de população, construções e arruamentos, onde a paisagem é intensamente alterada e a maioria das pessoas está envolvida em atividades secundárias e terciárias.

2.2 URBANIZAÇÃO

Os seres humanos sempre tendem a viver em comunidades, isto traz o auxílio mútuo da proteção, alimentação e reprodução. O homem aprendeu cultivar plantas e domesticar animais, dando origem aos primeiros assentamentos estáveis. Aproximadamente 5.000 anos atrás alguns assentamentos se transformaram em cidades (ALMEIDA, 2004).

A partir da revolução industrial, no século XVIII, a capacidade produtiva aumentou e as cidades se desenvolveram, com o objetivo do crescimento econômico. Os meios de transporte e as leis sanitárias se tornaram os indutores e moderadores da expansão urbana. Anteriormente, as cidades eram pequenas e o

deslocamento feito por caminhadas, mas após a Revolução Industrial observou-se a formação de cidades ferroviárias organizadas em torno das estações. Posteriormente o automóvel disseminou a ocupação e preenchendo os espaços entre estações (ALMEIDA, 2004).

A partir de 1940 até 1980 o PIB brasileiro teve um crescimento excepcional sob qualquer perspectiva (MARICATO, 2003). Neste período a urbanização brasileira foi marcada por elevados índices de natalidade, redução de mortalidade e intensificação da industrialização, estes fatores são decisivos na metropolização, que passou a desenvolver distintas aglomerações urbanas do País nos meados do século XX (SILVA, 2012).

2.2.1 Urbanização de Curitiba e região leste da RMC

Fazendo parte da Nação Guarani, os Tinguis foram os habitantes originais dos campos de Curitiba. Desde a chegada dos colonizadores portugueses os arredores do vilarejo de Curitiba foram ocupados lentamente, chegavam paulistas e aventureiros de diversas origens à procura de ouro (LIMA, 2000). A elevação de Curitiba à categoria de vila, em 1693, representou avanço significativo desse período histórico, no qual se encontra a origem dos limites físicos dos quatro municípios estudados no presente trabalho (Curitiba, Pinhais, Piraquara e São José dos Pinhais).

Após 1850 se dá um incremento populacional expressivo, por meio da imigração de colonos europeus, como resultado de uma política de colonização preconizada pelo império, sob D. Pedro II e que incluía várias regiões (LIMA, 2000). Em sua grande maioria, os imigrantes são encaminhados aos núcleos coloniais, não permanecendo em Curitiba.

As colônias de imigrantes deram origem a vários dos atuais municípios regionais. Desde então, as atividades agrícolas de pecuária e, em menor proporção, a indústria de extração mineral foram se afirmando e se mantiveram como atividades produtivas principais da região, exceto na capital, onde já naquela época iniciavam-se atividades de pequena escala nas áreas comerciais, industriais, administrativas e de serviços. A conformação regional apresentava um núcleo urbano em Curitiba que, ao início do século XX, beirava os 50.000 habitantes (LIMA, 2000).

A partir dos anos 1940, em decorrência da redução da oferta de trabalho nas áreas rurais, verifica-se no plano nacional um deslocamento expressivo da população interiorana em direção aos centros urbanos em busca de emprego (LIMA, 2000). O Brasil entra em nova fase, o país eminentemente rural passa a adquirir características mais urbanas.

No Paraná esse processo foi influenciado pela crise do café e modernização da agricultura, na década de 1960 e a partir de 1970, com o desenvolvimento industrial do Estado (NUNES, 2014). A RMC foi instituída pela Lei Complementar n.º 14/73, que concretizou as oito primeiras regiões metropolitanas do país.

Na década de setenta, o processo de urbanização na Região de Curitiba passou a apresentar alterações em maior volume nas áreas rurais, periféricas ao pólo, de baixo custo no mercado imobiliário, enquanto a cidade “legal” expandia-se centralizada por Curitiba que, a partir de 1965, passou a ser o primeiro município da futura região metropolitana detentor de um Plano Diretor de Urbanismo (LIMA, 2000). Uma ocupação intensificada vem a acontecer ao final dos anos setenta, quando o Paraná tornou-se uma das unidades da federação em que os impactos da modernização se fizeram mais expressivos.

O crescimento urbano regional passa a assumir configuração condicionada pela dinâmica de ocupação da cidade-pólo, principalmente em relação às atividades econômicas geradoras de oportunidades de trabalho (LIMA, 2000).

Na década de 1950, são aprovados e ocupados loteamentos em áreas inadequadas como áreas íngremes, inundáveis, próximas a eixos rodoviários, ou a rios de mananciais de abastecimento público, sendo os exemplos mais expressivos as áreas periféricas nos municípios de Colombo e Piraquara (LIMA, 2000).

A novidade ao final da década de 1980 e principalmente na década seguinte é que as áreas periféricas, além de serem formadas predominantemente por loteamentos de baixa qualidade técnica, passam a ser invadidas ou ocupadas ilegalmente, a característica comum sendo a falta de infra-estrutura ou serviços públicos implantados adequadamente (COMEC, 1999). Algumas dessas áreas, no caso dos loteamentos aprovados, foram atendidas com serviços básicos de abastecimento de água, energia elétrica, telefonia, transporte coletivo e pavimentação básica (LIMA, 2000). Outro evento impactante para a região foi deflagrado entre 1995 e 1996, pelas medidas adotadas pelo governo estadual para

transformação da RMC no segundo pólo automotivo brasileiro que gerou atração de mão-de-obra e modificação da área ocupada (LIMA, 2000).

No quadro de crescimento populacional (dados do IBGE segundo o Censo de 1991 e Contagem Populacional de 1996), na Região Metropolitana de Curitiba a taxa de crescimento era de 2,34%. Ao mesmo tempo, o crescimento da periferia era ainda maior que o dobro do crescimento do pólo, com 4,86% (DESCHAMPS, 2004).

A COMEC estabeleceu uma política de preservação dos mananciais regionais em duas linhas; a execução do Programa de Saneamento Ambiental para a RMC e a implementação de legislação de preservação de mananciais, com a criação do Sistema Integrado de Proteção aos Mananciais pela Lei n.º 12.248/98 (LIMA, 2000). De acordo com o órgão gestor do abastecimento público de água, a Companhia de Saneamento do Paraná (SANEPAR), os mananciais da região obrigatoriamente devem ser objeto de atenção, devido à progressiva ocupação urbana que se instala em suas bacias e o conseqüente comprometimento da qualidade hídrica por contaminantes (LIMA, 2000).

Em 1996, por decreto estadual, foi estabelecido a Área de Proteção Ambiental na área de manancial da bacia hidrográfica do rio Iraí, denominada APA Estadual do Iraí com objetivo a proteção e a conservação da qualidade ambiental e dos sistemas naturais ali existentes, em especial a qualidade e quantidade da água para fins de abastecimento público (PARANÁ, 1996). Em 1999 é finalizada a obra da barragem do Rio Iraí. Neste período Piraquara já abrigava a represa Cayuguava, que posteriormente chamada de Piraquara I, e em 2010 o município passou a abrigar em seu território mais uma barragem a Piraquara II (SANTOS, 2024).

Criado em 2009, o Programa Minha Casa Minha Vida, institui normas, regulamentos, legislação e subsídios capazes de possibilitar o acesso à moradia aos habitantes com faixas de renda anteriormente pouco ou nunca contempladas (CARVALHO, 2014). Somente para Curitiba, nas faixas 1 e 2 do programa foram liberadas a construção de pouco mais de 10 mil unidades habitacionais, garantindo moradia para mais de 30 mil habitantes. Nas últimas décadas, é justamente o entorno pobre da capital que apresenta os maiores índices de crescimento populacional. Enquanto a cidade pólo cresceu menos de 1% ao ano entre 2000 e 2010, os municípios periféricos crescem em uma média anual de 2,25%, sendo os limítrofes à área Sul, a mais pobre da cidade, os que possuíam maiores índices de crescimento (CARVALHO, 2014). Antes do boom imobiliário, a partir de 2009, São

José dos Pinhais já tinha o mercado imobiliário voltado para população de baixa renda bastante aquecido, devido à valorização dos imóveis em Curitiba. São José dos Pinhais, assim como outros municípios da Região Metropolitana, absorveu grande parte dos habitantes excluídos pela valorização dos imóveis em Curitiba a partir dos investimentos públicos nas transformações no espaço urbano da capital (GODOY, 2012).

A perda da qualidade da água dos mananciais da RMC é uma realidade, sendo que alguns deles podem ter sua utilização inviabilizada em breve, tamanha sua degradação (GARCIAS, 2009). Paralelamente à situação de degradação dos recursos hídricos, há na RMC uma expansão desordenada da malha urbana. A expansão vem ocorrendo, pressionando assim cada vez mais os mananciais abastecedores da região, localizados principalmente na região leste. Também nesses locais há a presença de muitas áreas de ocupações irregulares, sem qualquer infraestrutura adequada para moradia, agravando assim a vulnerabilidade do meio e contribuindo ainda mais para a degradação dos recursos hídricos, com o lançamento direto de esgotos e resíduos nos corpos da água (GARCIAS, 2009). Em março de 2022 estabeleceu-se o decreto nº 10.499 com o objetivo de regulamentar o ordenamento territorial em áreas de mananciais de abastecimento público, situadas na RMC.

2.3 MODELOS DE SIMULAÇÃO DE DINÂMICA ESPACIAL

Modelo de simulação dinâmica espacial é uma categoria de modelo sistêmico. Um modelo sistêmico procura descrever o sistema como um todo, buscando representar as interações entre todos seus componentes. Uma característica fundamental desse tipo de modelo é a capacidade de lidar com a dimensão espacial, implementando relações de vizinhança e suportando o uso combinado de múltiplas escalas (MEIRELLES, 2007).

Além de representar o sistema como um todo, os modelos de simulação dinâmica espacial fazem uso de ferramentas com a capacidade de acomodar a heterogeneidade espacial e processos humanos de tomada de decisão. Uma forma de elaborar modelos de simulação dinâmica espacial é através da representação da paisagem em uma matriz de células e aplicar equações matemáticas sobre cada uma das células simultaneamente. Isto simplifica o mecanismo de predição, pois

cada célula do modelo está conectada com suas células vizinhas. Desta forma, se uma célula tem três vizinhos em um estado “x”, é provável que o estado dessa célula venha a ser “x” também. No entanto, esse raciocínio pode ficar mais robusto utilizando-se de regras de transição (MEIRELLES, 2007).

Modelos podem ser classificados como determinísticos ou estocásticos, depende da incorporação de efeitos probabilísticos. Modelos determinísticos sempre se comportam da mesma forma para uma situação inicial, sem presença de incerteza ou aleatoriedade. Modelos estocásticos incorporam a aleatoriedade e a incerteza de vários eventos ao longo do tempo, os resultados são expressos em termos de probabilidade. Alguns exemplos que serão abordados no capítulo 2.2 são: SLEUTH, SimLucia, CLUE-S, DINAMICA EGO.

Modelos podem ser classificados também como teóricos ou empíricos. Modelos teóricos são concebidos como aqueles cuja gama de suposições, premissas e equações derivadas (quando existentes) explicativas do comportamento do sistema são definidas antecipadamente. Por outro lado, os modelos empíricos baseiam-se nos dados disponíveis para emitir conclusões sobre o sistema posteriormente, usando métodos estatísticos empíricos ou até mesmo regras determinísticas simples (ALMEIDA, 2004).

Similar ao trabalho realizado por Almeida (2004) neste trabalho o espaço será tratado por uma grade regular, bidimensional, consistindo de unidades quadriláteras homogêneas, os pixels, de 30 m² por unidade. O tempo será considerado sob um ponto de vista absoluto como uma entidade unidimensional independente, e no caso particular desta pesquisa, será considerado como constituído de sucessivos intervalos discretos de 5 anos. O modelo deste trabalho se encaixa na categoria de um modelo empírico e estocástico.

A tecnologia de sensoriamento remoto mostrou um potencial para mapear a expansão mundial da extensão urbana nas últimas décadas, as abordagens baseadas em modelos são úteis para projetar a expansão urbana futura ou reconstruir a expansão urbana (LI et al. 2021).

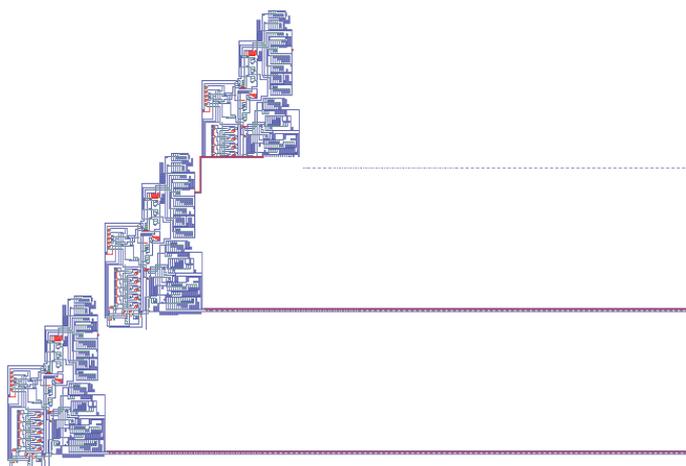
Devido à complexidade da urbanização, as condições de crescimento urbano estimadas para uma região são geralmente aplicáveis apenas a essa região, onde a taxa de urbanização pode ser considerada homogênea (ZHOU et al., 2019). Desta

forma, grande parte das pesquisas de previsão do crescimento urbano são limitadas a um único município ou região metropolitana.

2.3.1 Autômatos Celulares

Nos anos 1950 John Von Neumann estava interessado em encontrar uma organização lógica para um autômato pudesse produzir uma cópia dele mesmo (CASTRO, 2008). O ambiente desenvolvido é formado por uma grade regular retangular de células, cada célula pode ter 29 estados diferentes, estado que depende de regras de transição aplicadas à vizinhança. Neste ambiente Von Neumann obteve êxito em criar um autômato auto-replicante, como na FIGURA 1 onde apresenta a execução deste ambiente onde se tem um componente replicado duas vezes. Este ambiente foi denominado Autômato Celular.

FIGURA 1 - PRIMEIRA IMPLEMENTAÇÃO DO CONSTRUTOR AUTO-REPLICANTE DE VON NEUMANN



FONTE: PESAVENTO (1995)

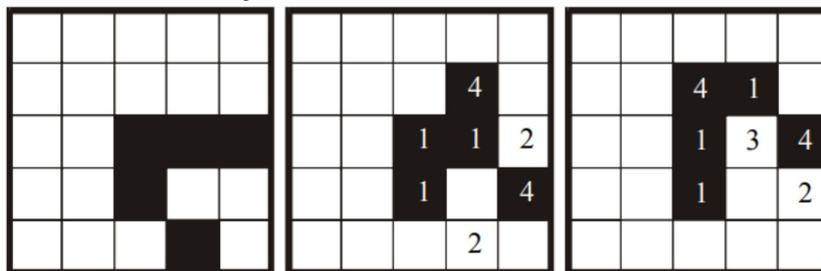
Em 1970, o matemático John Holton Conway criou o “Jogo da vida”, um autômato celular que simulava alterações em populações de seres vivos baseados em regras locais simples. Neste autômato celular, cada célula nasce ou morre de acordo com as células vizinhas (CASTRO, 2008). O fato de uma célula viver ou morrer depende da sua vizinhança e de algumas regras:

- Qualquer célula com dois vizinhos vivos continua no mesmo estado para a próxima geração (regra 1);

- Qualquer célula viva com menos de dois vizinhos vivos morre de solidão (regra 2);
- Qualquer célula viva com mais de três vizinhos vivos morre de superpopulação (regra 3);
- Qualquer célula com exatamente três vizinhos vivos se torna uma célula viva (regra 4);

A FIGURA 2 apresenta um exemplo de transição do “jogo da vida”, os números representam a aplicação de regras e os quadros a evolução ao longo do tempo (CASTRO, 2008):

FIGURA 2 - EXEMPLO DE EXECUÇÃO DO JOGO DA VIDA



FONTE: Castro (2008)

Meirelles (2007) discorre a respeito das aplicações dos autômatos celulares, dizendo que o conceito tem sido utilizado para modelar fenômenos físicos e urbanos. Neste sentido, o espaço é representado por um mosaico de células de tamanho e forma regulares e semelhantes. A cada uma destas células são aplicadas regras de transição, que determinam quando e por que o estado de uma célula se altera, além disso, as regras de transição podem ser qualitativas ou quantitativas.

2.4 EXEMPLOS DE MODELOS PARA O CRESCIMENTO URBANO COM AUTÔMATOS CELULARES

Existem vários exemplos de modelos e ferramentas que utilizam autômatos celulares para simular mudanças no uso do solo, na TABELA 1 é possível encontrar alguns destes exemplos .

TABELA 2 - MODELOS E FERRAMENTAS DE MUDANÇA NO USO DO SOLO

Modelo	Autor
Geomod	PONTIUS et al., 2001
SLEUTH	CLARKE et al., 1997
Land Use Scanner	HILFERINK; RIETVELD, 1999
Environment Explorer	ENGELEN et al., 2003
SAMBA	BOISSAU; CASTELLA, 2003
LTM	PIJANOWSKI et al., 2000
CLUE-S	VERBURG et al., 2002
DINAMICA EGO	RODRIGUES, 2007
SimLucia	ENGELEN et al., 1997

Em 1979, Waldo Tobler associou AC a processos sócio-espaciais, propondo seu uso na modelagem e predição de crescimento urbano. Nas seções a seguir, serão apresentados alguns dos modelos e ferramentas de simulação com autômatos celulares, incluindo o SLEUTH, SimLucia, CLUE-S e o Dinamica EGO.

2.4.1 SLEUTH

Os modelos desenvolvidos nos anos 1990 permitiram a representação mais acurada da realidade. Dentre tais modelos destaca-se o SLEUTH, acrônimo de *Slope, Land Use, Exclusion, Urban, Transportation e Hillshade*, desenvolvido por Clark em 1997 (LINARES, 2018). SLEUTH é um modelo de previsão baseado em CA usado para Modelagem de Crescimento Urbano e modelagem de uso do solo. O modelo considera seis variáveis de entrada, que são as variáveis que influenciam no desenvolvimento urbano: declividade, uso da terra, área excluída, área urbana, mapa de transporte e áreas montanhosas.

SLEUTH tornou-se um dos modelos de simulação mais populares de crescimento urbano e mudança de uso do solo (CHAUDHURI & CLARKE, 2013). A aplicação do modelo de Modelagem de Crescimento Urbano é mais popular. Existem pesquisas recentes realizadas sobre a previsão do crescimento urbano usando o modelo SLEUTH globalmente (KUMAR; AGRAWAL, 2023).

2.4.2 SimLucia

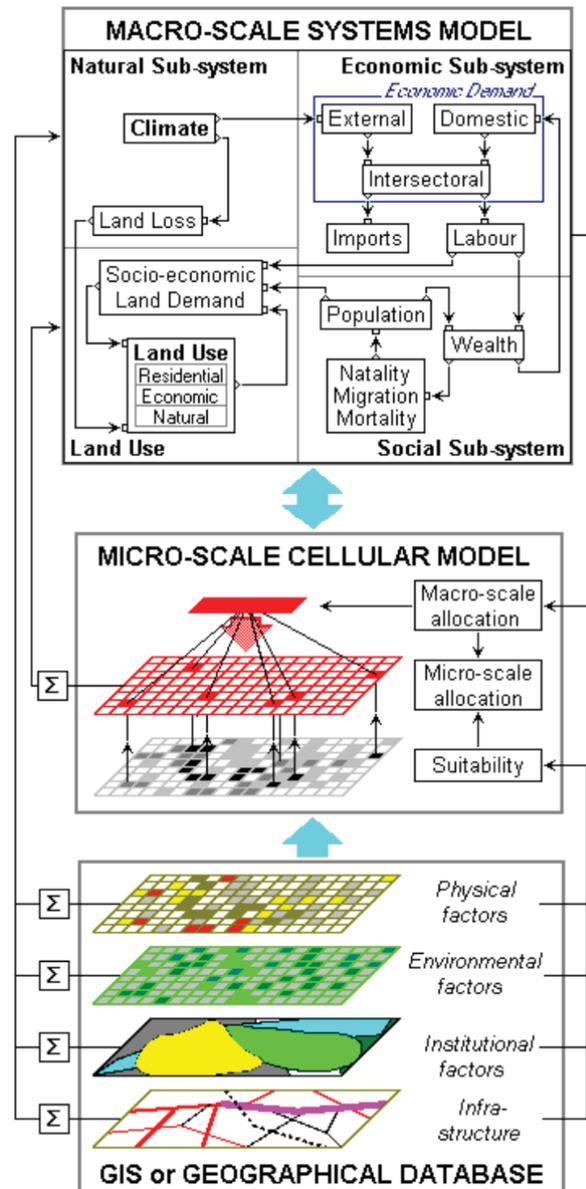
SimLucia é um *framework* de apoio à decisão desenvolvido para auxiliar planejadores e tomadores de decisão na análise de uma variedade de políticas e seus padrões espaciais associados (ENGELEN et al., 1997). Foi desenvolvido pela UNEP (Programa das Nações Unidas para o Meio Ambiente) como parte do projeto “Avaliação da vulnerabilidade das zonas costeiras baixas e pequenas ilhas às alterações climáticas e à subida do nível do mar”. O projeto busca a longo prazo criar uma metodologia e procedimento aplicáveis a países costeiros ou ilhas como Santa Lúcia. O modelo e software resultante foram nomeados de "SimLucia" (ENGELEN et al., 1997).

O núcleo central desse *framework* é composto por modelos espaciais dinâmicos que operam em escalas tanto micro quanto macro (ENGELEN et al., 1997). No nível macro, o *framework* de modelagem integra diversos submodelos que representam os sistemas naturais, sociais e econômicos característicos da área sob estudo, como apresentado na FIGURA 3 pelo primeiro retângulo. Estes submodelos estão interligados em uma rede de influência recíproca (ENGELEN, 1997).

No nível micro, modelos baseados em autômatos celulares determinam o destino de parcelas individuais de terra com base em suas características institucionais e ambientais específicas, bem como no tipo de atividades presentes em suas proximidades, como apresentado na FIGURA 3 pelo segundo retângulo. Esses modelos autômatos celulares são definidos com vizinhanças maiores e uma variedade maior de estados celulares que representam usos socioeconômicos da terra e cobertura natural (ENGELEN, 1997).

A dinâmica global desses modelos é guiada pelos modelos em escala macro. Essa abordagem possibilita a incorporação direta de variáveis detalhadas de natureza física, ambiental e institucional, assim como particularidades da infraestrutura de transporte (ENGELEN, 1997).

FIGURA 3 - ESTRUTURA DO SIMLUCIA



FONTE: Engelen et al. 1997

2.4.3 CLUE-S

O framework de modelagem Conversão de Uso da Terra e seus Efeitos (CLUE) foi desenvolvido para simular mudanças no uso da terra utilizando relações empiricamente quantificadas entre o uso da terra e seus fatores direcionadores, combinados com modelagem dinâmica (VERBURG et al., 2002). Este modelo foi desenvolvido para uma abordagem de modelagem em nível nacional e continental.

Para aplicações regionais é usado o modelo denominado CLUE-S (Conversão de Uso da Terra e seus Efeitos em Pequena Extensão Regional). O modelo requer uma variedade de dados de entrada, como mapas de uso da terra

atual, informações socioeconômicas, dados demográficos, informações ambientais, políticas de uso da terra, entre outros. Além disso, parâmetros específicos são configurados para guiar o comportamento do modelo (VERBURG et al., 2002).

2.4.4 Dinamica EGO

Dinamica EGO foi concebido como um software genérico de modelagem de mudanças ambientais, por isso, ele tem sido aplicado a uma variedade de estudos que incluem a modelagem de expansão urbana e dinâmica intra-urbana (RODRIGUES, 2007). Em 2019, FREITAS (2019) desenvolveu seu Projeto Final de graduação em Engenharia Cartográfica e de Agrimensura modelando a dinâmica espacial da cidade de Curitiba no ambiente de modelagem Dinâmica EGO. Usado também em outros estudos como os de Kamusoko e Gamba (2014) para simular o crescimento urbano de Harare, capital do Zimbábue.

O Dinamica EGO é um ambiente de modelagem que usa "functors" como operadores para realizar tarefas específicas sobre conjuntos de dados. Cada "functor" executa operações como cálculos de custo, sorteio de pontos prováveis para fenômenos, cálculos de caminho mais curto e expressões lógicas/aritméticas em mapas. Esses "functors" são essenciais para o modelo e representam os blocos de construção básicos de um modelo cartográfico. O sistema inclui operadores de análise espacial comuns em SIG comerciais, além de operadores especialmente desenvolvidos para simulações espaciais, incluindo calibração e validação de métodos (RODRIGUES, 2007).

2.5 ÁRVORE DE DECISÃO E FLORESTA ALEATÓRIA

Floresta Aleatória, em inglês *Random Forest* (RF), é um classificador e regressor que consiste em um agregado de árvores de decisão onde estão são construídas com um subconjunto aleatório dos dados de treinamento (BREIMAN, 2001).

Árvores de Decisão operam sob o princípio de 'dividir para conquistar': dividindo um conjunto de dados maior em subconjuntos menores (BIAU; SCORNET, 2016). A construção da árvore baseia-se na escolha de atributos que servirá para dividir os dados em conjuntos menores para associá-lo a uma classe, chamado de nó folha, e para o resto do conjunto define-se outro atributo para dividir os dados

novamente para associar a uma classe, repetindo este procedimento até todos os dados estarem inseridos numa classe (OSHIRO, 2013). A biblioteca *Scikit-Learn* usa o método CART com critério Gini para a construção destas árvores. As árvores de decisão nada mais são que uma sequência de testes simples onde, cada teste implementa uma comparação de valores de um determinado atributo, como mostrado na FIGURA 4:

FIGURA 4 - EXEMPLO DE ÁRVORE DE DECISÃO



FONTE: Didática Tech (2019)

Com uma extensa fonte de amostras, uma árvore de decisão pode encontrar dificuldades em classificar corretamente, devido à significativa quantidade de níveis na árvore há uma maior probabilidade de ocorrer overfitting resultando em previsões imprecisas. O RF apresentado por Breiman (2001) é capaz de contornar esse problema criando uma floresta de árvores de decisão, criadas com parte das amostras escolhidas randomicamente.

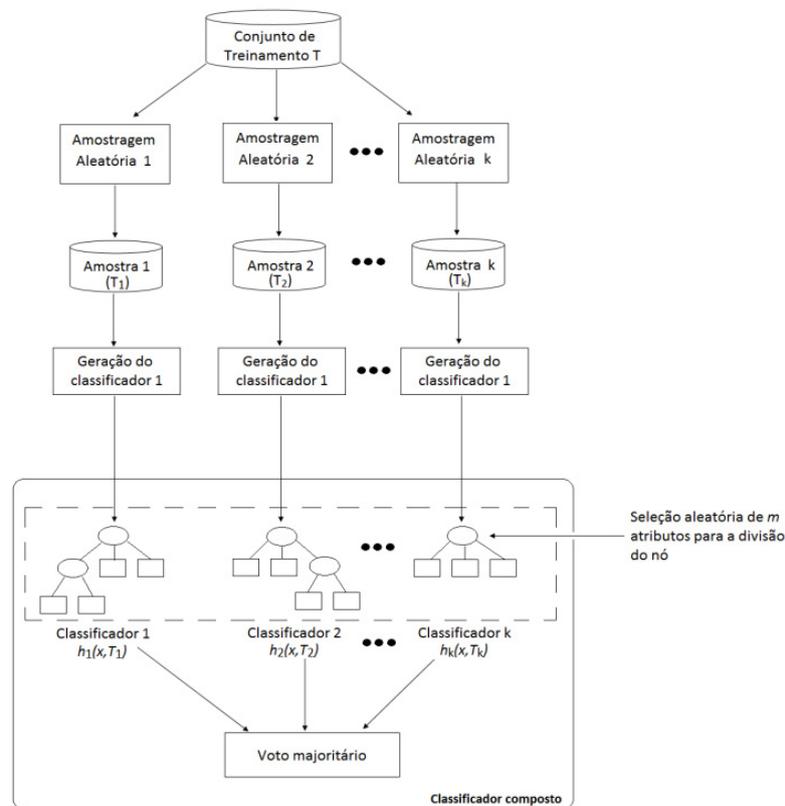
Cada árvore é construída por uma amostra do conjunto de treinamento obtida por meio do *bootstrap*, uma técnica de amostragem com reposição que seleciona aleatoriamente exemplos para um novo conjunto.

A cada divisão da árvore os atributos são avaliados, o melhor é escolhido para dividir o nó. O uso do *bootstrap* faz com que as árvores sejam diferentes, portanto, diminui a correlação, que tende a diminuir a taxa de erro da classificação (OSHIRO, 2013).

Devido a reposição, exemplos iguais podem ser escolhidos ou não serem escolhidos. Esses não escolhidos formam o conjunto *out-of-bag* e com este obter o erro *out-of-bag* usado para calcular o erro de previsão e avaliar as medidas de

importância das variáveis (SHEYKHOUSA; MAHDIANPARI, 2020). A FIGURA 5 mostra um fluxograma de como funciona o RF.

FIGURA 5 - FUNCIONAMENTO DO RANDOM FOREST



FONTE: OSHIRO, 2013

Na classificação, cada árvore obtém um resultado de classe. A classe obtida por todas as árvores são contados e a classe que tem o maior número é escolhida (voto majoritário). Isto se aplica para valores discretos de classe, para valores contínuos é possível utilizar a regressão linear para obter um valor.

O RF é usado em vários estudos com AC para simular a expansão urbana (ZHOU et al., 2020; SHAFIZADEH-MOGHADAM et al. 2021; CHEN, 2022; RIENOW et al. 2021). No estudo realizado por Kamusoko e Gamba (2015) são destacadas as vantagens do RF, como: a capacidade de lidar com um grande número de amostras, exigem menos tempo de treinamento em comparação com outros classificadores de aprendizado de máquina, estão livres de premissas de distribuição normal, são robustos ao lidar com *outliers* e ruídos, e quantificam cada variável de entrada em uma medida de importância.

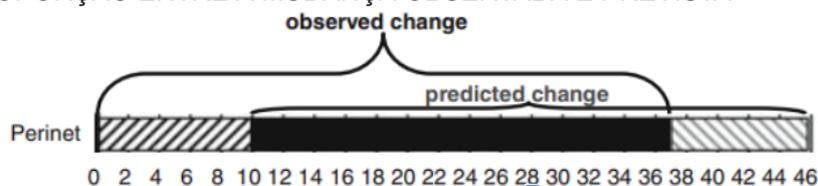
2.6 MÉTODOS DE AVALIAÇÃO

Neste tópico, serão delineados os métodos de avaliação empregados para analisar os resultados da classificação e da simulação da expansão urbana, os quais incluem a figura de mérito, a matriz de erros e a curva ROC.

2.6.1 Figura de Mérito

Proposto por Pontius et al. (2008) para ser empregado em imagens destinadas à modelagem de mudança no solo, a Figura de Mérito (FoM) é uma métrica estatística que avalia a concordância entre a realidade observada e a simulação realizada. Na FIGURA 6, o FoM é representado como a proporção da interseção entre a barra 'mudança do uso do solo observada' (realidade) e a barra 'mudança do uso do solo predita' (simulada) para Perinet em Madagascar, conforme calculado pela EQUAÇÃO 1.

FIGURA 6 - PROPORÇÃO ENTRE A MUDANÇA OBSERVADA E PREVISTA



FONTE: Adaptado de Pontius et al.(2008)

Na equação 1, A representa a área incorretamente prevista como inalterada, B é a área corretamente identificada como mudança, e C é a área equivocadamente considerada inalterada, mas prevista como mudança. Em outras palavras, esse número indica a proporção daquilo que foi previsto ou simulado corretamente.

$$FoM = B/(A + B + C) \quad (1)$$

2.6.2 Matriz de Erros

Para construir uma matriz de erros, é necessário comparar a realidade com as previsões geradas pelo modelo. Nesse processo, serão calculadas as ocorrências de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos.

- Verdadeiro Positivo: expansão urbana corretamente prevista (VP);
- Verdadeiro Negativo: não expansão urbana prevista corretamente (VN);
- Falso Positivo: expansão urbana erroneamente prevista (FP).
- Falso Negativo: expansão urbana não prevista (FN);

Desta forma pode-se construir a matriz de erros aplicado à classificação binária bem como para problemas de classificação multiclasse (KULKARNI et al., 2020). Para a simulação foi usada uma matriz de erros binária e para a classificação de imagens foi usado uma matriz de erros multiclasse. Um exemplo de matriz de erros para classificação multiclasse é mostrado na FIGURA 7.

FIGURA 7 – EXEMPLO DE MATRIZ DE ERROS

B		Dado de Classificado			
		A	B	C	Total
Dado de Referência	A	VA	FN	FN	\sum_r^A
	B	FP	VB	FN	\sum_r^B
	C	FP	FP	VC	\sum_r^C
	Total	\sum_c^A	\sum_c^B	\sum_c^C	\sum_T

FONTE: SOLVED, 2022

As matrizes de erros representam contagens de valores preditos e reais. A diagonal principal representa os valores verdadeiros, concordantes entre cada classe, FP mostra o valor de Falsos Positivo, e FN significa o valor de Falsos Negativo. Com estes valores podemos obter a acurácia do usuário determinada pela EQUAÇÃO 2:

$$Acurácia\ Usuário = X_{ii} / \sum X_{Coluna} \quad (2)$$

Onde X_{ii} é o elemento da diagonal principal e X_{Coluna} são os elementos da coluna i . Podemos calcular também a acurácia do produtor determinada pela EQUAÇÃO 3:

$$\text{Acurácia Produtor} = X_{ii} / \sum X_{Linha} \quad (3)$$

Onde X_{ii} é o elemento da diagonal principal e X_{Linha} são os elementos da linha i . A acurácia total é determinada pela EQUAÇÃO 4.

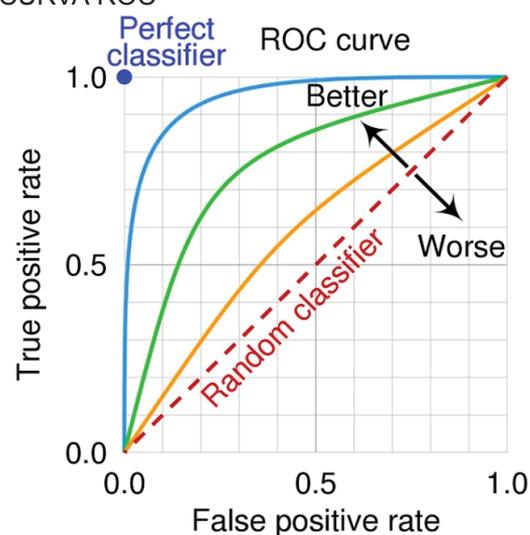
$$\text{Acurácia Total} = \frac{V}{V+FN+FP} \quad (4)$$

Sendo V o total de valores verdadeiros da diagonal principal.

2.6.3 Curva ROC

A característica operacional relativa (ROC) é um método comumente utilizado para validar simulações da mudança do uso da terra. A curva ROC representa a taxa de verdadeiros positivos versus a taxa de falsos positivos obtida a partir da comparação entre a variável booleana e os diagnósticos múltiplos derivados dos limiares aplicados ao índice (PONTIUS; PARMENTIER, 2014). A FIGURA 8 apresenta diferentes curvas, desde o classificador ideal até o menos eficiente.

FIGURA 8 - EXEMPLO DE CURVA ROC



FONTE: Gomedes (2023)

Com base nas estatísticas ROC, uma medida com poder preditivo perfeito produziria um valor de 1,0, uma sem poder preditivo produziria um valor de 0,5, e

valores inferiores a 0,5 indicam uma medida sistematicamente incorreta (KAMUSOKO; GAMBA, 2015).

2.7 ÍNDICES ESPECTRAIS

Com o sensoriamento remoto, é possível obter uma ampla gama de informações, uma delas são os índices espectrais, que por uma álgebra entre bandas é possível realçar características nas imagens que sejam relevantes. Dentre os diversos índices existentes, vale destacar o NDVI, o NDBI, o NDWI e o BLFEI. O NDVI (*Normalized Difference Vegetation Index*), popularizado por Huete(1988), é destaca as áreas vegetadas na imagem, útil para avaliar a saúde e a extensão da vegetação, permitindo a classificação de diferentes tipos e densidades de cobertura vegetal, é calculado pela equação 7.

$$NDVI = \frac{NIR - Vermelho}{NIR + Vermelho} \quad (7)$$

O NDBI (*Normalized Difference Building Index*), proposto por Zha et al(2003), é um índice espectral formulado para identificar áreas urbanizadas construídas, sendo aplicado no mapeamento dessas regiões, é calculado a partir da equação 9.

$$NDBI = \frac{(SWIR1 - NIR)}{(SWIR1 + NIR)} \quad (9)$$

O NDWI (*Normalized Difference Water Index*), proposto por McFeeters (1996), realça os corpos d'água na imagem e é calculado a partir da equação 8.

$$NDWI = \frac{Verde - NIR}{Verde + NIR} \quad (8)$$

O BLFEI (*Built-Up Land Features Extraction Index*), proposto por Bouhennache et al (2018), realça áreas impermeáveis e é calculado a partir da equação 10.

$$BLFEI = \frac{(Verde + Vermelho + SWIR2)/3 - SWIR1}{(Verde + Vermelho + SWIR2)/3 + SWIR1} \quad (10)$$

3 MATERIAL E MÉTODOS

Neste capítulo serão apresentados os dados vetoriais e raster utilizados, e a contextualização do processo de utilização do Random Forest integrado com Autômatos Celulares para a previsão do crescimento urbano.

3.1 MATERIAIS

Imagens da região que compreende os municípios de Curitiba, Pinhais, Piraquara e São José dos Pinhais, do ano de 2000, 2005, 2010, 2015 e 2020 foram usadas neste estudo. São imagens do *Landsat 5* e 8 com processamento de nível 2 conforme a TABELA 3.

TABELA 3 - IMAGENS UTILIZADAS

ANO	LANDSAT 5
2000	LT05_L2SP_220078_20000819_20230212_02_T1
2005	LT05_L2SP_220078_20050902_20200901_02_T1
2010	LT05_L2SP_220078_20100831_20200823_02_T1
LANDSAT 8	
2015	LC08_L2SP_220078_20150829_20200908_02_T1
2020	LC08_L2SP_220078_20200826_20200905_02_T1

Estes satélites têm sensores com características conforme mostrado a TABELA 4 e 5:

TABELA 4 - CARACTERÍSTICAS DO *LANDSAT 5*

Sensor	Bandas Espectrais	Resolução Espectral	Resolução Espacial	Resolução Temporal	Área Imageada	Res. Radiométrica
TM (Thematic Mapper)	(B1) AZUL	0.45 - 0.52 μm	30 m	16 dias	185 km	8 bits
	(B2) VERDE	0.52 - 0.60 μm				
	(B3) VERMELHO	0.63 - 0.69 μm				
	(B4) INFRAVERMELHO PRÓXIMO	0.76 - 0.90 μm				
	(B5) INFRAVERMELHO MÉDIO	1.55 - 1.75 μm	120 m			
	(B6) INFRAVERMELHO TERMAL	10.4 - 12.5 μm	30 m			
	(B7) INFRAVERMELHO MÉDIO	2.08 - 2.35 μm				

FONTE: Embrapa, 2020

TABELA 5 - CARACTERÍSTICAS DO *LANDSAT 8*

Sensor	Bandas Espectrais	Resolução Espectral	Resolução Espacial	Resolução Temporal	Área Imageada	Res. Radiométrica
OLI (Operational Land Imager)	(B1) COSTAL	0.433 - 0.453 μm	30 m	16 dias	185 km	12 bits
	(B2) AZUL	0.450 - 0.515 μm				
	(B3) VERDE	0.525 - 0.600 μm				
	(B4) VERMELHO	0.630 - 0.680 μm				
	(B5) INFRAVERMELHO PRÓXIMO	0.845 - 0.885 μm				
	(B6) INFRAVERMELHO MÉDIO	1.560 - 1.660 μm				
	(B7) INFRAVERMELHO MÉDIO	2.100 - 2.300 μm				
	(B8) PANCROMÁTICO	0.500 - 0.680 μm	15 m			
	(B9) Cirrus	1.360 - 1.390 μm	30 m			

FONTE: Embrapa, 2020

Os arquivos vetoriais utilizados estão listados na TABELA 6, incluindo o zoneamento das quatro cidades da área de estudo fornecido pela COMEC, os vetores das vias (rodovias, ruas primárias, ruas secundárias e ruas terciárias) do *OpenStreetMap*, dados do IBGE (Censo) para os anos 2000 e 2010, a malha censitária disponibilizada pelo IBGE e as curvas de nível fornecidas pelo IAT.

TABELA 6 - ARQUIVOS VETORIAIS UTILIZADOS

Dados	Fonte
Zoneamento	COMEC
Vias	OSM
População 2000	IBGE
População 2010	
Malha censitária 2000	IBGE
Malha censitária 2010	
Curvas de Nível	IAT

Neste estudo foi usado:

- Software:
 - QGIS (v. 3.22): para a manipulação dos dados geográficos e produção de mapas finais;
- Linguagem de Programação:
 - Python (v. 3.10.12);
- Ambiente de Programação:
 - Google Colaboratory: para o desenvolvimento do algoritmo;
- Recursos:
 - Biblioteca GDAL (v. 3.3.2): para ler e gravar dados geográficos no formato raster e vetores;
 - Biblioteca NumPy (v. 1.23.5): para realizar operações matemáticas em matrizes.
 - Biblioteca Scikit-Learn (v. 1.2.2): para utilizar o algoritmo Random Forest.
 - Biblioteca Pandas (v. 1.5.3): criação e formatação de tabelas.

- Biblioteca Matplotlib(v. 3.8.2): plotagem de gráficos.

3.1.1 Área de Estudo

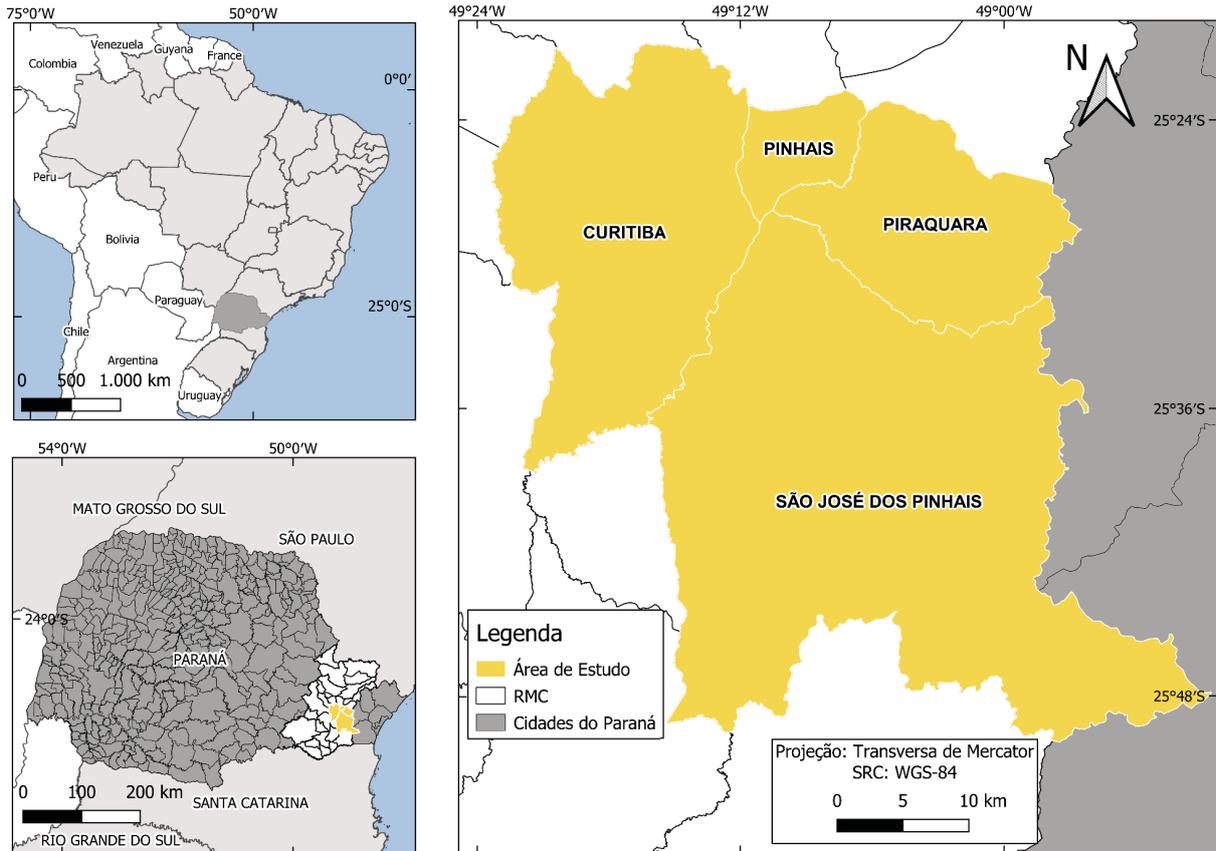
A área de estudo foi definida partindo do critério que sejam consideradas as regiões onde a expansão da área urbana acontece com expressividade, um exemplo onde a urbanização acontece de forma rápida e desordenada é o da cidade de Curitiba e Região metropolitana, nesta região é onde ocorrem os maiores índices de crescimento da RMC.

Em 2005, Curitiba tinha 341 espaços informais de moradia, aumentando de 174 entre 1996 e 2005, totalizando 59.064 domicílios. Em 2009, Pinhais tinha 38 assentamentos informais, com um crescimento de 58% entre 1997 e 2009, totalizando 1.303 domicílios. Em São José dos Pinhais, a densidade dos espaços informais diminuiu de 45,87 para 25,91 domicílios por hectare entre 1997 e 2009, tal redução é explicada pelo aumento da área dos assentamentos do que do número de domicílios, e permitiu constatar que a dinâmica é marcada por um processo extensivo de produção informal do espaço (SILVA, 2012).

Levando-se em conta essas premissas foi considerado realizar o estudo para Curitiba e região leste da RMC, compreendidas pelos municípios de Piraquara, Pinhais e São José dos Pinhais.

Em termos físico-geográficos, a Região Metropolitana está situada no Primeiro Planalto Paranaense em quase a totalidade da sua superfície, a região apresenta uma altitude uniforme na cota 850-900m na maior parte da região. O clima é caracterizado, como subtropical úmido, com temperaturas médias de aproximadamente 18°C conforme registrou o INMET (Instituto Nacional de Meteorologia) para o ano de 2020. A FIGURA 9 apresenta a localização da área de estudo.

FIGURA 9 - ÁREA DE ESTUDO



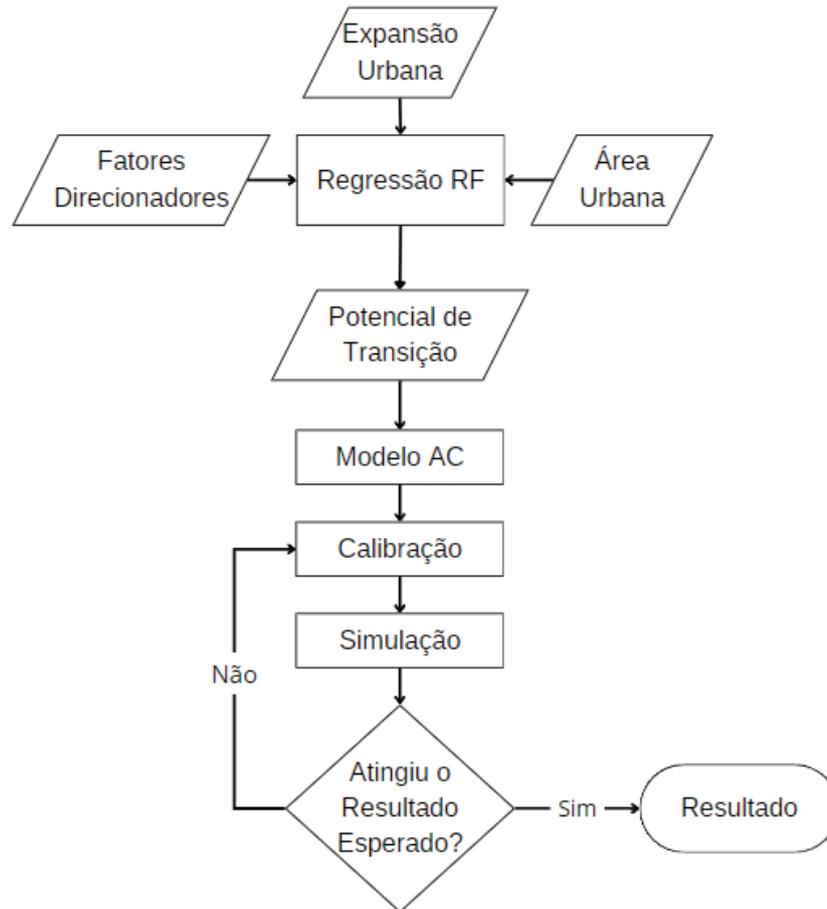
3.2 MÉTODOS

A metodologia proposta é composta pelas etapas listadas a seguir.

- Classificação das imagens *Landsat*;
- Criação das variáveis Fatores Direcionadores;
- Determinação do Potencial de Transição;
- Desenvolvimento do Modelo de Autômato Celular.

Por meio dessas etapas, obtêm-se resultados que são incorporados à metodologia do trabalho, conforme ilustrado no fluxograma apresentado na FIGURA 10.

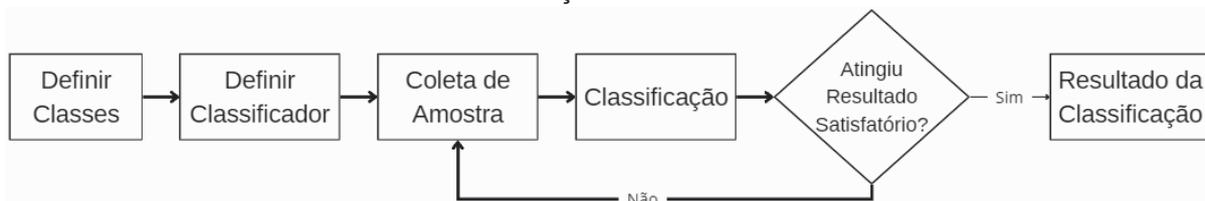
FIGURA 10 - FLUXOGRAMA GERAL DO ESTUDO



3.2.2 Classificação

Foram empregadas para a classificação imagens provenientes dos satélites *Landsat 5* e *Landsat 8*, obtidas nos anos de 2000, 2005, 2010, 2015 e 2020. A etapa de classificação é ilustrada no fluxograma da FIGURA 11.

FIGURA 11 - FLUXOGRAMA DA CLASSIFICAÇÃO

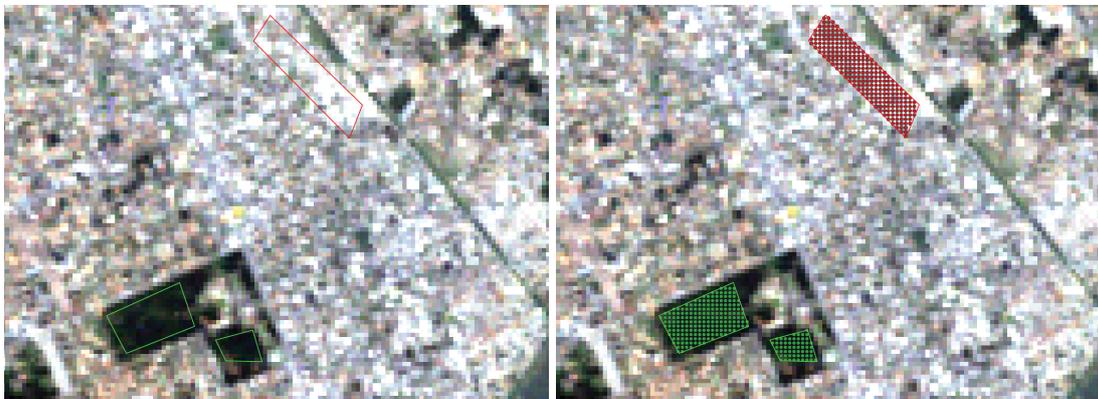


Primeiramente foram definidas as classes para este estudo e posteriormente escolhido o classificador a se usar. Depois foram coletados pontos amostrais.

Para coletar os pontos de amostra, inicialmente, desenharam-se polígonos que delimitam áreas representando as classes. Nessas áreas foi determinada as coordenadas para cada *pixel*, conforme apresentado na FIGURA 12, estes *pixels* são as amostras de treinamento.

No recorte da área de estudo mostrado na FIGURA 12, à esquerda estão os polígonos delimitando áreas urbanas e de vegetação, enquanto à direita destaca polígonos preenchidos.

FIGURA 12 - EXEMPLO DA COLETA DE AMOSTRAS



Destes *pixels* pode-se obter suas coordenadas no referencial da imagem coluna e linha (C e L) aplicando para as coordenadas UTM (E_{PIXEL} , N_{PIXEL}) de cada *pixel* as fórmulas 5 e 6. Os valores E_0 e N_0 são as coordenadas UTM do limite esquerdo da imagem e do extremo superior da imagem, e 30 é o tamanho do *pixel*.

$$C = \frac{E_{PIXEL} - E_0}{30} \quad (5)$$

$$L = \frac{N_{PIXEL} - N_0}{30} \quad (6)$$

De posse das imagens com nível 2 de processamento, portanto com valores de reflectância, e de todas as coordenadas (linha e coluna) destes *pixels* numa única tabela, basta associá-los aos seus respectivos valores de reflectâncias em diferentes bandas e índices para utilizar no classificador.

3.2.3 Criação das Variáveis e Fatores Direcionadores

Para modelos de expansão urbana são necessárias variáveis denominadas Fatores Direcionadores. A maior parte dos fatores foram escolhidos baseando-se no estudo realizado por Kumar e Tripathy (2019). A variável de distância para a

mancha urbana foi escolhida com base no estudo de Kamusoko e Gamba (2015).

Os fatores direcionadores utilizados são:

- distância para as vias;
- distância para o centro urbano;
- distância para a mancha urbana;
- densidade populacional;
- e declividade.

Essas variáveis foram criadas no ambiente do QGIS. Elas operam de tal forma que, à medida que se aproximam de ruas, centros urbanos e mancha urbana, a probabilidade de ocorrer crescimento urbano aumenta. Da mesma forma, à medida que a densidade populacional for maior, a tendência de crescimento urbano nessa área é maior. A variável de declividade atua de modo que, quanto maior a declividade, menor a probabilidade de ocupação. A quantificação dessa tendência será determinada pelo modelo.

As variáveis de distância são criadas de forma que cada *pixel* tem um valor de distância para o atributo em questão que são:

- rodovias;
- ruas primárias;
- ruas secundárias;
- ruas terciárias;
- zonas centrais;
- e mancha urbana.

Vias é dividido em 4 camadas vetoriais em diferentes tipos de classificação do OSM (OpenStreetMap Wiki):

- ruas primárias: via asfaltada com duas ou mais faixas por sentido, velocidade máxima maior que 50 km/h e menor que 70 km/h;
- ruas secundárias: via asfaltada com duas faixas por sentido e velocidade máxima de até 50 km/h;
- ruas terciárias: via pavimentada ou não de apenas uma faixa em cada sentido ou velocidade máxima de 40 km/h;
- Rodovias: constituído da classe *Motorway* e *Trunk*, com via asfaltada com três ou mais faixas por sentido, velocidade máxima acima de 70 km/h, e presença de cruzamento em nível, semáforo e faixa de pedestres.

As variáveis de densidades populacionais foram geradas usando o arquivo vetorial da malha censitária do ano de 2000 e 2010. Na tabela do censo do ano de 2000 e 2010, fornecidos pelo IBGE, foi feita uma união pelo campo da tabela do censo e da tabela de atributos da malha censitária. Uma variável de áreas restritas à ocupação foi criada composta dos corpos d'água da região de estudo. Estas variáveis vetoriais passaram pelo processo de rasterização, com *pixels* de 30 metros para compatibilizar com as imagens de satélite.

A variável de declividade foi gerada a partir das curvas de nível do Paraná, de 20 em 20 metros. Foi realizada a interpolação com os valores de elevação das curvas de nível. A variável raster resultante da interpolação foi transformada para valores de declividade em graus.

Outra variável produzida foi a de vizinhos urbanos, com base no estudo de Karimi et al. (2019), calculado por:

$$V = \frac{\sum cont(c=urbano)}{n(n-1)} \quad (11)$$

onde $cont(c = urbano)$ é a contagem de células da classe mancha urbana na vizinhança 3x3 (vizinhança de Moore), n é o número de células na vizinhança.

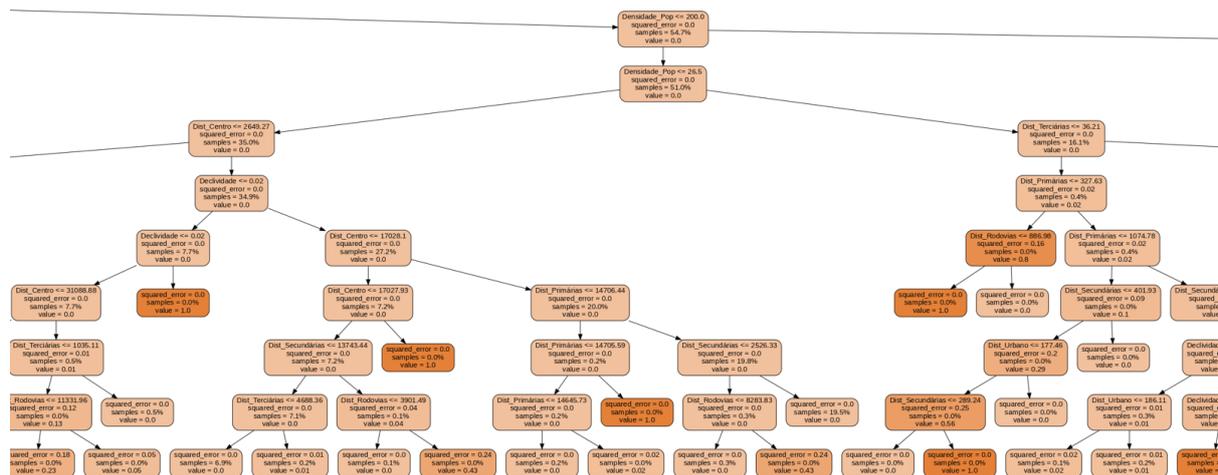
Nestes modelos também precisamos de uma variável com a mancha urbana, para indicar o estado inicial da simulação, e uma variável que nos mostra onde ocorreu a expansão urbana. Com a imagem classificada foi derivada uma variável apenas com as áreas de mancha urbana, utilizando programação *Python*, para todos os anos considerados. Com estas foram criadas outras variáveis da expansão urbana que ocorreu em um período de 5 anos.

Os parâmetros, em formato vetorial, foram convertidos para formato raster. Teve-se o cuidado para todas as variáveis raster terem a mesma resolução espacial, dimensão e projeção.

3.2.4 Determinação do Potencial de transição com Random Forest

O algoritmo de regressão Random Forest (RF) foi usado para a determinação do potencial de transição. O potencial de transição representa o valor da chance, entre 0 e 1, de uma célula (*pixel*) inicialmente classificada como não urbana se tornar urbana depois de um período.

FIGURA 13 - RECORTE DE UM EXEMPLO DE ÁRVORE DE DECISÃO DO RF



O RF opera de maneira a selecionar aleatoriamente uma parte dos dados de treinamento usando amostragem bootstrap. Com essa amostra, cada árvore de decisão é construída, utilizando as variáveis para dividir os nós de forma a maximizar a precisão da classificação na folha. As classes foram definidas como 0 para a ausência de expansão urbana e 1 para sua ocorrência.

Esse processo é repetido para construir mil árvores, todas distintas devido à amostragem bootstrap. Cada posição de pixel tem armazenado todos os valores de variáveis e é processado por essas mil árvores, gerando valores de 0 ou 1 para cada uma. Em seguida, calcula-se a média desses valores, que representa o potencial de transição para aquela posição de pixel.

3.2.5 Desenvolvimento do Modelo de Autômato Celular

Um autômato celular constitui-se, de:

- ambiente de execução;
- n estados de células;
- vizinhança;
- regras de transição.

O ambiente de execução de estudo é uma grade regular de células que são os *pixels* da imagem. Os estados considerados são apenas dois: urbano e não urbano. A vizinhança é a de Moore, uma matriz 3x3, e neste caso foi incorporada pelo algoritmo RF como uma variável de vizinhos urbanos. Foi considerado regras de transição com duas componentes: o potencial de transição e a restrição.

$$e_{ij} = (P_{RF})_{ij} * (rest)_{ij} \quad (13)$$

onde P_{RF} é a probabilidade de transição baseada pelos fatores direcionadores nesse estudo calculado com Random Forest, e o *rest* é uma condicional de restrição que são os corpos d'água. E e_{ij} é o estado (urbano ou não urbano) para o *pixel* na posição i e j.

Foi definido um limiar para o potencial de transição baseado no valor de área de crescimento urbano que realmente ocorreu num período de 5 anos, se o *pixel* atingir esse valor limiar e não estiver numa área restrita ele sofrerá transição para a classe urbana. E desta forma podemos determinar para toda a região uma simulação da expansão urbana.

4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS

4.1 CLASSIFICAÇÃO DAS IMAGENS *LANDSAT*

A etapa de classificação das imagens da área de estudo se divide em cinco partes. Primeiramente definiu-se as classe, são quatro no total:

1. Urbano - representando elementos como: Casas, Prédios, calçadas, asfalto, entre outros;
2. Vegetação - para representar qualquer tipo de vegetação como: árvore, arbusto, grama, plantações e entre outros similares;
3. Água - qualquer corpo d'água: rio, lago ou represa;
4. Solo - para representar o solo exposto.

Em segundo, estabeleceu-se o classificador, sendo escolhido o Random Forest (RF). Na terceira etapa, realizou-se a coleta de pontos de amostra, no ambiente QGIS, para as quatro classes. Em seguida, procedeu-se à execução do algoritmo Random Forest utilizado com a biblioteca Scikit-Learn em linguagem Python como quarta etapa.

De posse das imagens com nível 2 de processamento, portanto com valores de reflectância, e de todas as coordenadas (linha e coluna) destes *pixels* numa única tabela, para ser utilizada no classificador RF, basta associá-los aos seus respectivos valores de reflectâncias em diferentes bandas e índices como na TABELA 7. Foram utilizadas as bandas do azul, verde, vermelho, infravermelho próximo (NIR), ambas do infravermelho médio (SWIR 1 e SWIR 2) e do termal, e utilizados os índices do NDVI, NDWI, NDBI e o BLFEI.

A quinta etapa consistiu na avaliação dos resultados da classificação, e, caso não atingisse o desempenho desejado, retornava-se à terceira etapa para repetir a sequência.

TABELA 7 - PONTOS DE AMOSTRAS COM SUAS RESPECTIVAS REFLECTÂNCIAS EM DIFERENTES BANDAS E ÍNDICES

	Index	classe	linha	coluna	B1	B2	B3	B4	B5	B6	B7	ndvi	ndbi	BLFEI	ndwi	termal
0	0	1.0	715.0	420.0	9750.0	10236.0	11015.0	11825.0	13358.0	16280.0	15829.0	0.060874	0.098590	-0.116228	-0.192892	8766.0
1	1	1.0	714.0	421.0	10322.0	10947.0	11856.0	12051.0	14147.0	16632.0	17641.0	0.080006	0.080737	-0.091291	-0.167650	8731.0
2	2	1.0	715.0	421.0	10018.0	10503.0	11325.0	12340.0	13889.0	17119.0	16493.0	0.059057	0.104167	-0.122373	-0.203699	8752.0
3	3	1.0	712.0	422.0	10194.0	10483.0	12211.0	13149.0	15008.0	17744.0	17500.0	0.066023	0.083537	-0.107938	-0.184710	8735.0
4	4	1.0	713.0	422.0	10041.0	10732.0	11614.0	12744.0	14884.0	18715.0	17734.0	0.077458	0.114021	-0.143052	-0.234132	8725.0
...
21413	21413	4.0	221.0	1061.0	7720.0	7922.0	8402.0	8554.0	11269.0	13541.0	12463.0	0.136962	0.091576	-0.159961	-0.234198	8501.0
21414	21414	4.0	217.0	1062.0	7673.0	7841.0	8352.0	8338.0	11666.0	12959.0	11820.0	0.166367	0.052508	-0.153843	-0.216179	8486.0
21415	21415	4.0	218.0	1062.0	7697.0	7883.0	8303.0	8461.0	10916.0	13013.0	12059.0	0.126697	0.087634	-0.150541	-0.220961	8510.0
21416	21416	4.0	219.0	1062.0	7742.0	7967.0	8408.0	8592.0	11198.0	13373.0	12234.0	0.131683	0.088519	-0.156951	-0.227951	8516.0
21417	21417	4.0	220.0	1062.0	7793.0	8002.0	8502.0	8656.0	12508.0	13487.0	12038.0	0.182007	0.037661	-0.161721	-0.226704	8499.0

21418 rows x 17 columns

4.1.1 Melhorando a Classificação com Novas Amostras

Parte da qualidade da classificação foi aferida de forma visual. Como exemplo, a FIGURA 14 apresenta um recorte da imagem de 2020 com parte da represa do Iraí com uma grande porção de área com solo exposto (devido a seca da época) sendo classificado erroneamente como urbano.

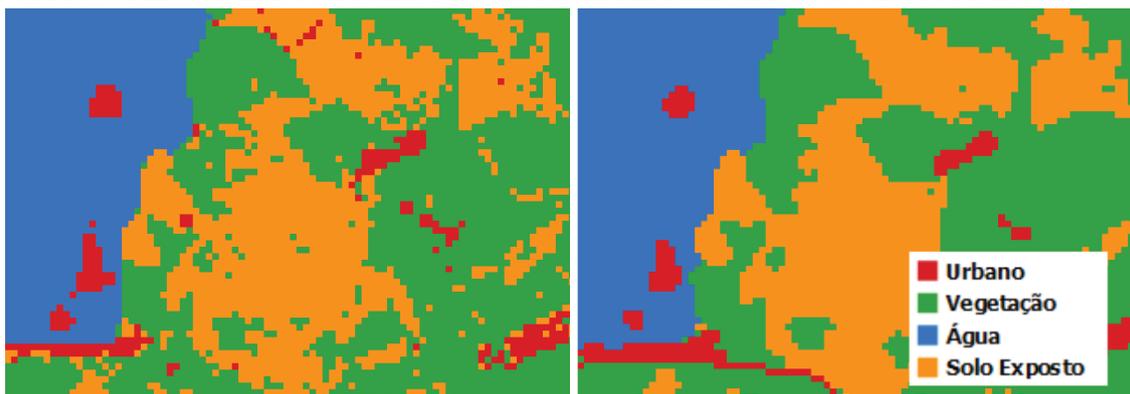
Coletando mais pontos amostrais na região é possível perceber visualmente a melhora na classificação na região. Após a inserção da banda termal do *Landsat* no classificador notou-se uma melhora para distinguir entre o solo exposto e urbano. De maneira similar todo esse processo visual foi feito para toda a imagem.

FIGURA 14 - RECORTE DA IMAGEM DE SATÉLITE, SUA CLASSIFICAÇÃO E A CLASSIFICAÇÃO MELHORADA POR MAIS PONTOS DE AMOSTRA



Como pode-se na FIGURA 15 observar a classificação apresenta ruído, *pixel* isolados classificados erroneamente. Foi aplicado um filtro de mediana com tamanho 3x3 para reduzir esse ruído com o uso da biblioteca *OpenCV*.

FIGURA 15 - RECORTE DA IMAGEM CLASSIFICADA E A MESMA COM FILTRO APLICADO

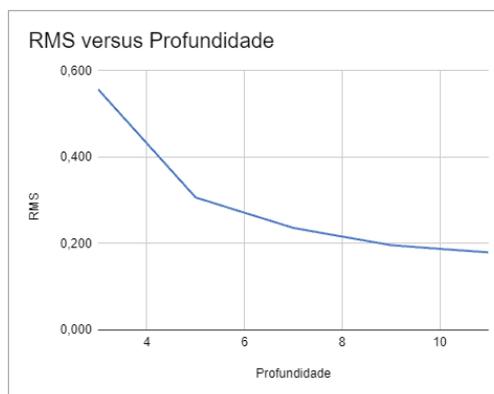


4.1.2 Melhorando a Classificação Testando Diferentes Profundidades no RF

Na execução do Random Forest, tanto no classificador como no regressor, é preciso definir a profundidade das árvores, neste estudo foi definido com base no valor do RMS (erro médio quadrático) para as execuções com diferentes valores de profundidade. Os valores obtidos com testes realizados para uma classificação de uma imagem de 2010 são apresentados na TABELA 8.

TABELA 8 - VALORES DE RMS PARA DIFERENTES PROFUNDIDADES DE ÁRVORES DO RF E SUA PLOTAGEM EM UM GRÁFICO

Profundidade	RMS
3	0,557
5	0,307
7	0,236
9	0,196
10	0,188
11	0,180



Observa-se uma melhoria significativa no valor do RMS à medida que a profundidade aumenta, tendendo a se estabilizar em um patamar inferior a 0,2. Embora seja possível aumentar ainda mais a profundidade para obter valores de RMS mais baixos, é importante considerar não apenas o custo computacional, mas também as recomendações de Kamusoko e Gamba (2015), que desaconselham árvores muito profundas devido ao risco de *overfitting*. Overfitting é quando um modelo se ajusta muito bem ao conjunto de dados, mas se mostra ineficaz para

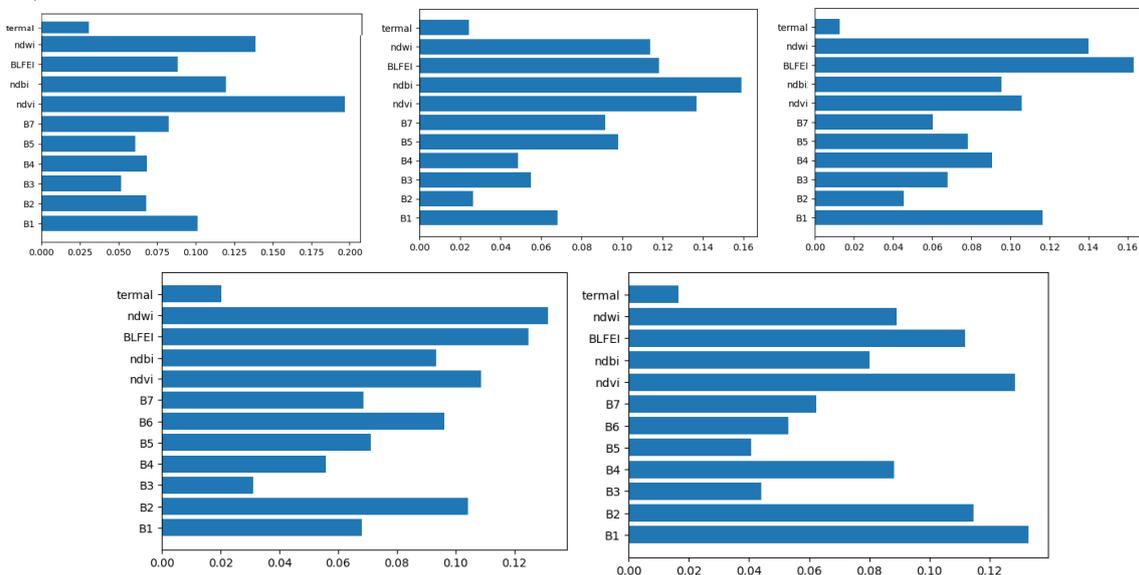
prever novos resultados. Empiricamente chegou-se à adoção de árvores com 10 níveis nas execuções do Random Forest, tanto para o classificador quanto para o regressor.

4.2 AVALIAÇÃO DOS RESULTADOS DA CLASSIFICAÇÃO

Os períodos 2000, 2005, 2010, 2015 e 2020 se dão pela disponibilidade de criar uma variável densidade populacional em 2000 e 2010 com os dados do Censo demográfico no país com disponibilidade da malha censitária em formato vetorial. Os intervalos de 5 anos são escolhidos de forma que o modelo pode trazer melhores resultados para uma simulação conforme no estudo de Peschl (2021) e, Tripathy e Kumar (2019).

A classificação RF foi efetuada com a biblioteca Scikit-learn. Dos pontos de amostra 75% foram utilizados para o treinamento e 25% para a validação, modelo composto de 10 mil árvores com profundidade de 10 níveis. Foi computada a importância de cada variável (banda e índice espectral) usada na classificação, apresentada na FIGURA 16.

FIGURA 16 - IMPORTÂNCIA DAS VARIÁVEIS PARA AS CLASSIFICAÇÕES DE 2000, 2005, 2010, 2015, 2020.



A qualidade da classificação foi aferida de forma visual, por matriz de erros e pela curva ROC. A matriz de erros foi computada com a biblioteca *Scikit-Learn* com uma coleção de pontos não usada na classificação para ter uma métrica de qualidade isentada de tendência. A curva ROC foi plotada usando a biblioteca *Matplotlib*.

Uma forma de avaliar o nível de desempenho da classificação é comparar visualmente a imagem classificada com a imagem de satélite. Uma comparação visual desses mapas mostra a semelhança relativamente alta. Isto sugere que o algoritmo RF foi relativamente preciso na atribuição das classes na área de estudo. As figuras 17 a 21 mostram as imagens classificadas resultantes.

FIGURA 17 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2000

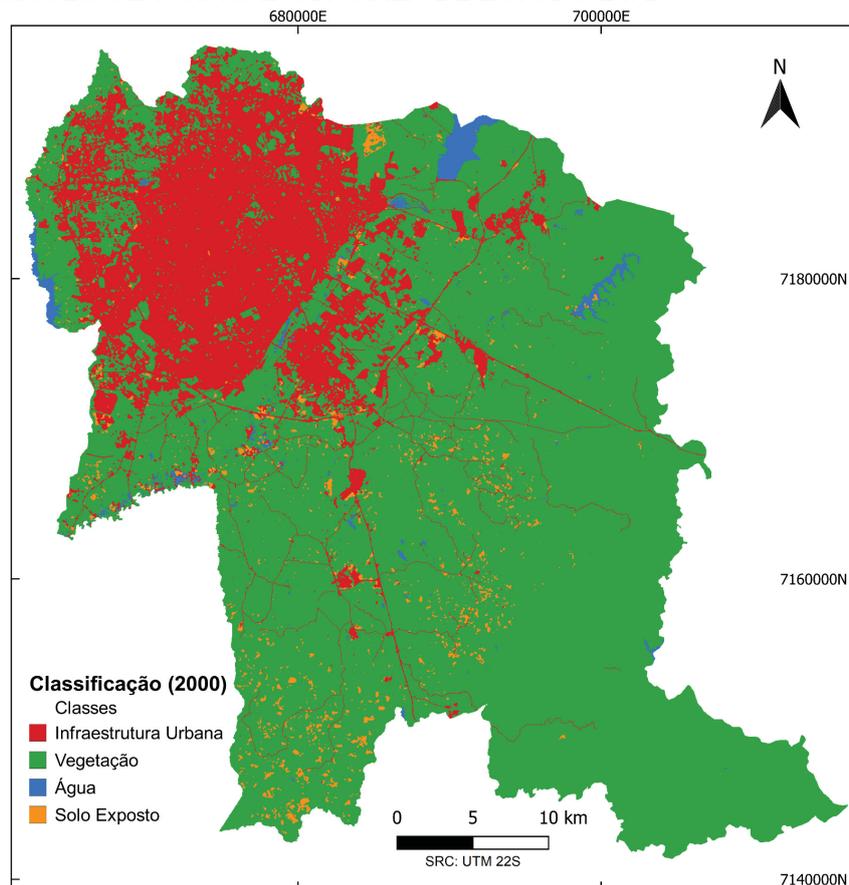


FIGURA 18 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2005

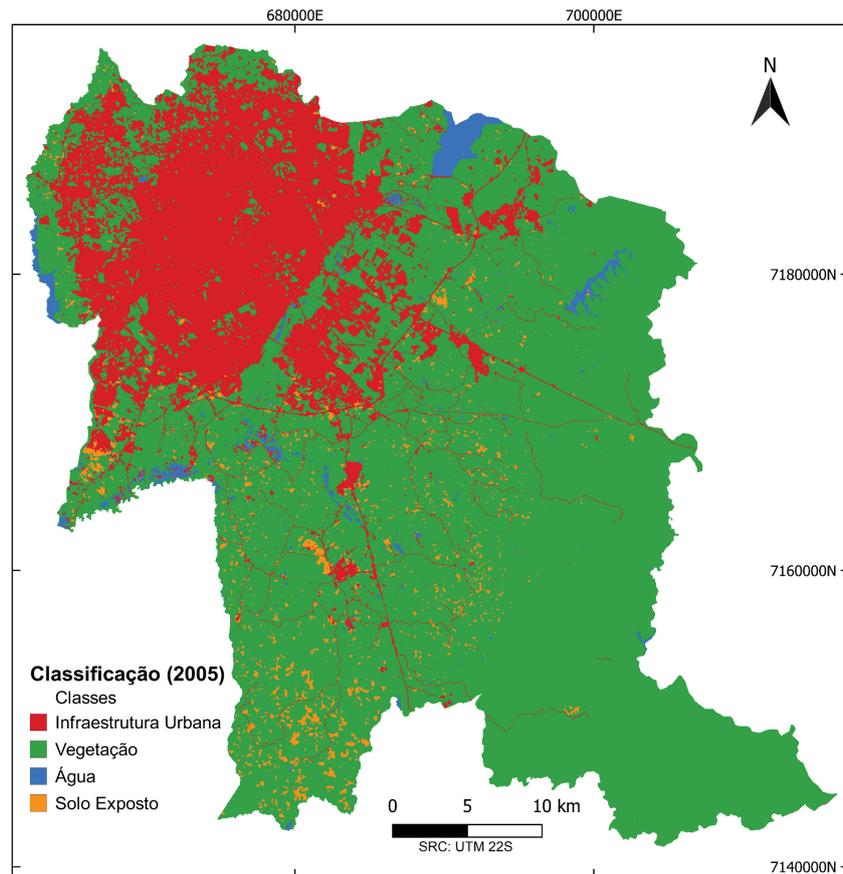


FIGURA 19 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2010

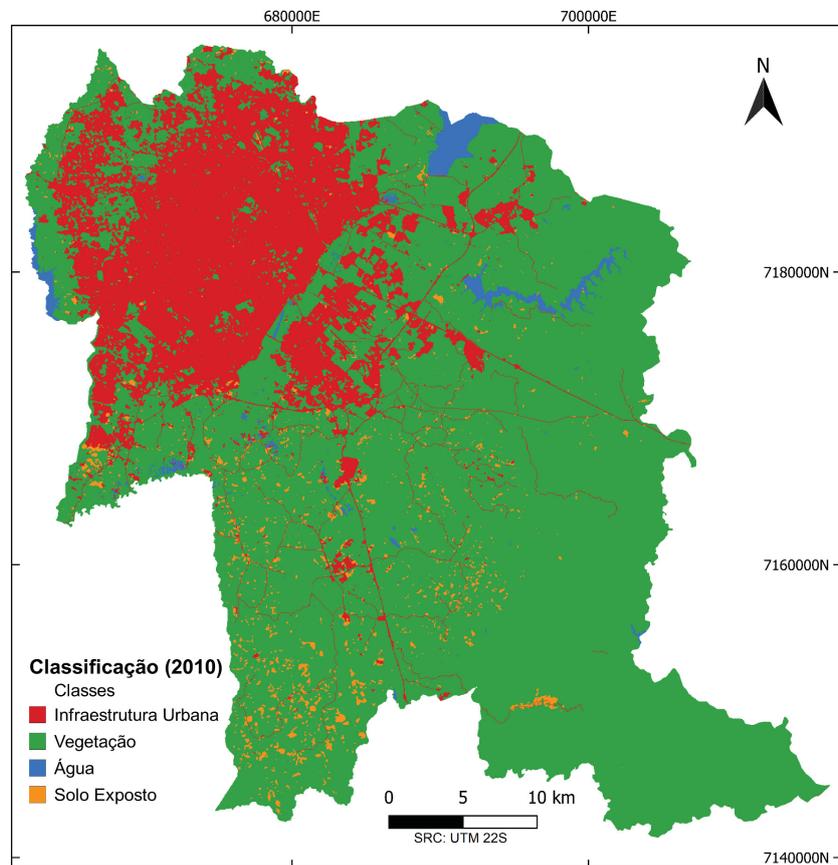


FIGURA 20 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2015

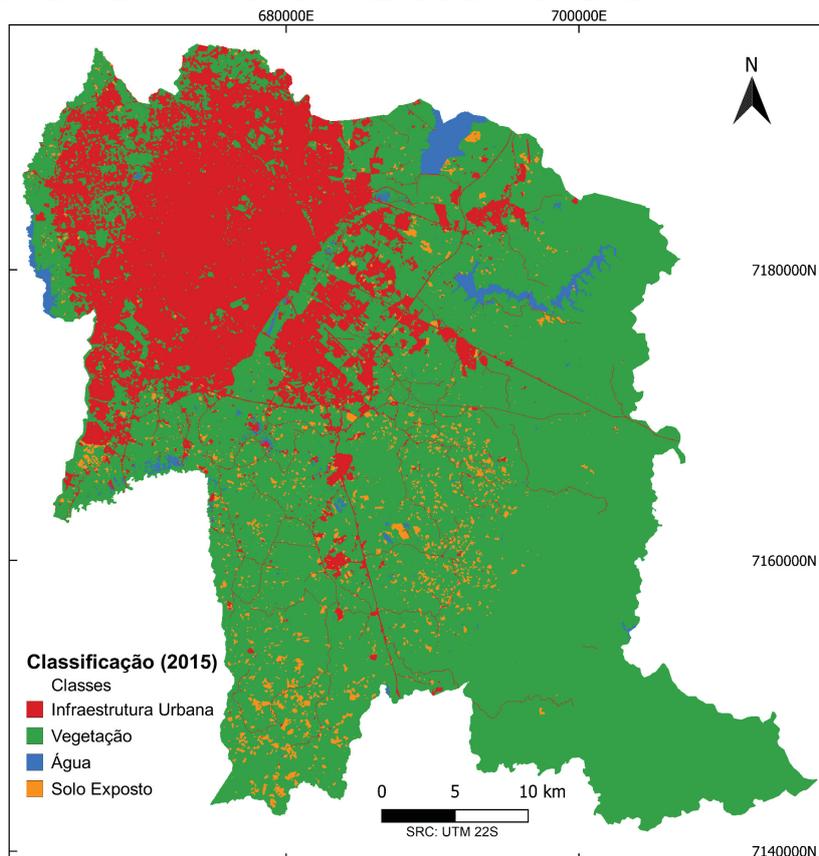
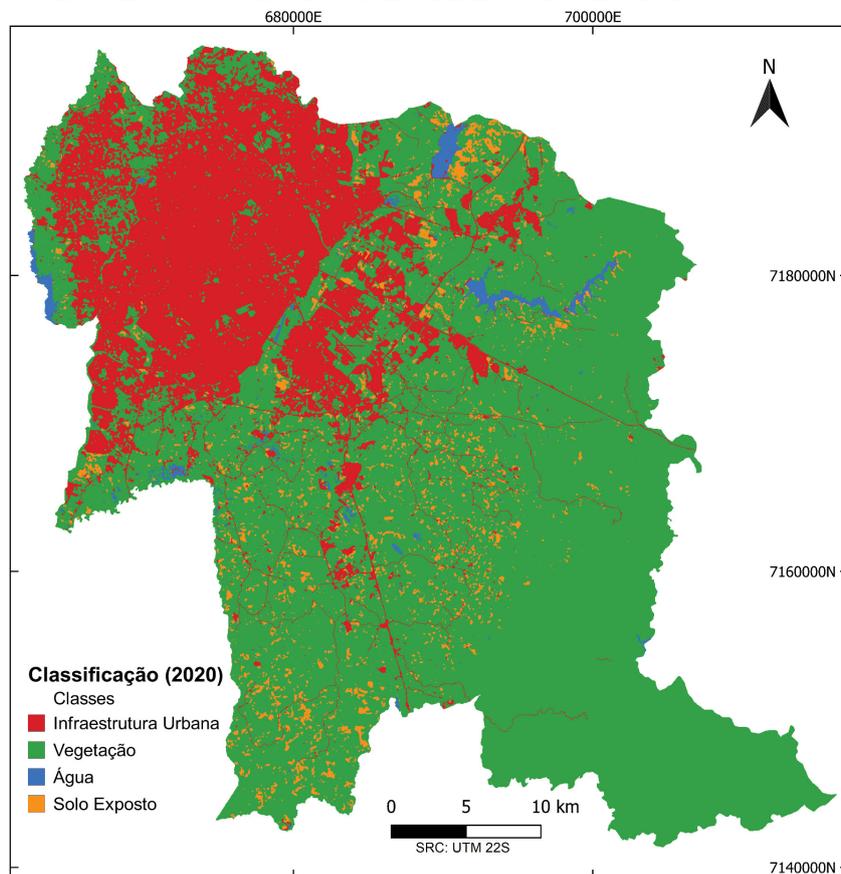


FIGURA 21 - IMAGEM CLASSIFICADA DA ÁREA DE ESTUDO EM 2020



As tabelas 9 a 13 apresentam as matrizes de erros das classificações. Tais matrizes revelam o nível de concordância por classe entre os pontos de amostra coletados da imagem (parte não usada no treinamento) e a imagem classificada. A acurácia geral obtida foi todos acima de 98% e as acurácias do Usuário e do Produtor para todas as classes variaram de 79,5% a 100%. Em relação às divergências, fica evidente a dificuldade de diferenciar as áreas de mancha urbana e solo exposto, sendo estas com os menores valores de acurácia.

TABELA 9 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2000

		Real (2000)				Total	Acurácia Produtor
		Urbano	Vegetação	Água	Solo		
Classificado (2000)	Urbano	2.186	18	9	9	2.222	98,38%
	Vegetação	16	4.165	6	20	4207	99,00%
	Água	2	1	1.391	1	1395	99,71%
	Solo	1	5	1	116	123	94,31%
Total		2205	4189	1407	146	7947	
Acurácia Usuário		99,14%	99,43%	98,86%	79,45%		
Acurácia Total		98,88%					

TABELA 10 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2005

		Real (2005)				Total	Acurácia Produtor
		Urbano	Vegetação	Água	Solo		
Classificado (2005)	Urbano	2.248	8	0	12	2.268	99,12%
	Vegetação	6	4.087	0	13	4106	99,54%
	Água	0	0	2.269	0	2269	100,00%
	Solo	0	13	0	268	281	95,37%
Total		2254	4108	2269	293	8924	
Acurácia Usuário		99,73%	99,49%	100,00%	91,47%		
Acurácia Total		99,42%					

TABELA 11 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2010

		Real (2010)				Total	Acurácia Produtor
		Urbano	Vegetação	Água	Solo		
Classificado (2010)	Urbano	2.323	12	0	10	2.345	99,06%
	Vegetação	12	1.634	2	19	1667	98,02%
	Água	0	0	2.190	0	2190	100,00%
	Solo	2	4	0	210	216	97,22%
Total		2337	1650	2192	239	6418	
Acurácia Usuário		99,40%	99,03%	99,91%	87,87%		
Acurácia Total		99,05%					

TABELA 12 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2015

		Real (2015)				Total	Acurácia Produtor
		Urbano	Vegetação	Água	Solo		
Classificado (2015)	Urbano	2.302	6	0	21	2.329	98,84%
	Vegetação	9	1.582	1	2	1594	99,25%
	Água	0	0	2.189	0	2189	100,00%
	Solo	8	6	0	234	248	94,35%
Total		2319	1594	2190	257	6360	
Acurácia Usuário		99,27%	99,25%	99,95%	91,05%		
Acurácia Total		99,17%					

TABELA 13 - MATRIZ DE ERROS PARA A CLASSIFICAÇÃO DA IMAGEM DE 2020

		Real (2020)				Total	Acurácia Produtor
		Urbano	Vegetação	Água	Solo		
Classificado (2020)	Urbano	2.345	10	2	21	2.378	98,61%
	Vegetação	14	1.595	0	9	1618	98,58%
	Água	0	2	1.021	0	1023	99,80%
	Solo	2	12	0	322	336	95,83%
Total		2361	1619	1023	352	5355	
Acurácia Usuário		99,32%	98,52%	99,80%	91,48%		
Acurácia Total		98,66%					

A curva ROC das classificações é representada nas figuras 22 a 26, esta é usada para avaliar o desempenho do classificador RF. Nota-se que este se aproxima da melhor curva possível, em todos os períodos com AUC próximo a 100%, contudo, a curva correspondente a classificação do solo exposto indica que teve a pior desempenho dentre as demais classes.

FIGURA 22 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2000

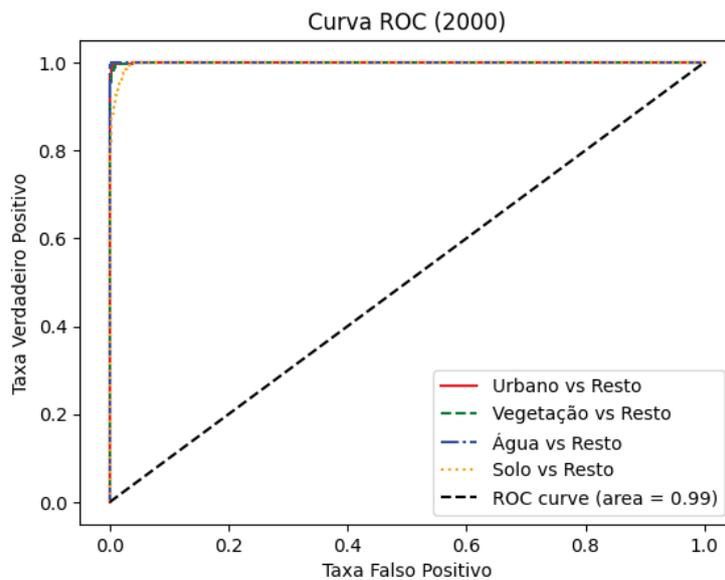


FIGURA 23 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2005

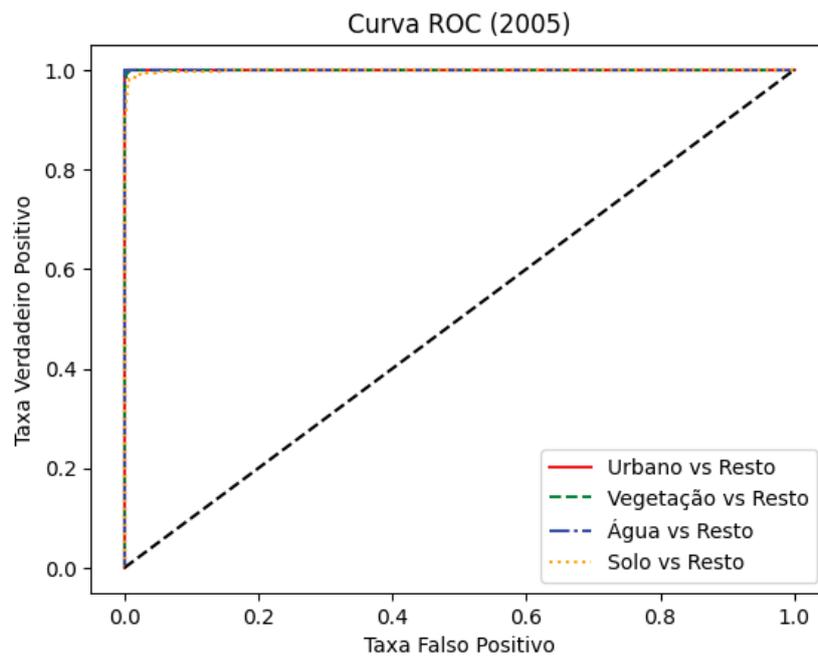


FIGURA 24 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2010

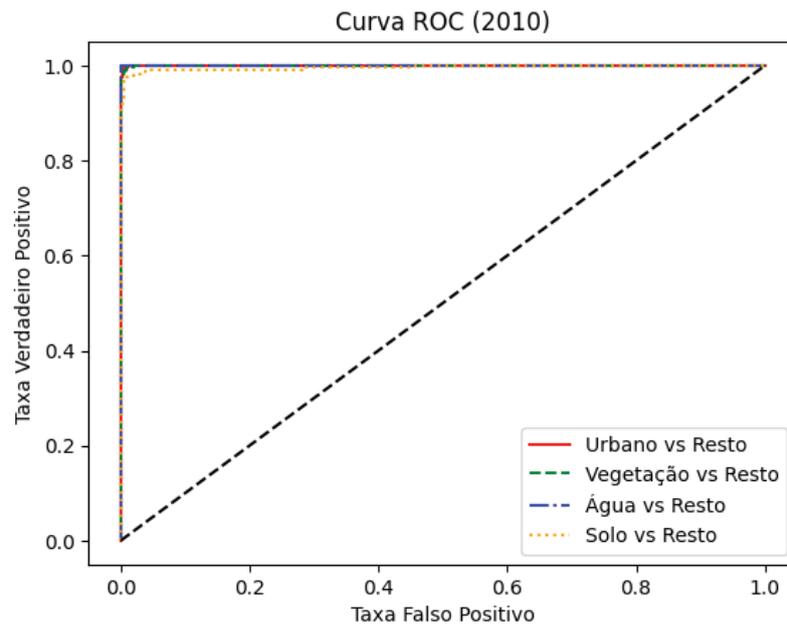


FIGURA 25 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2015

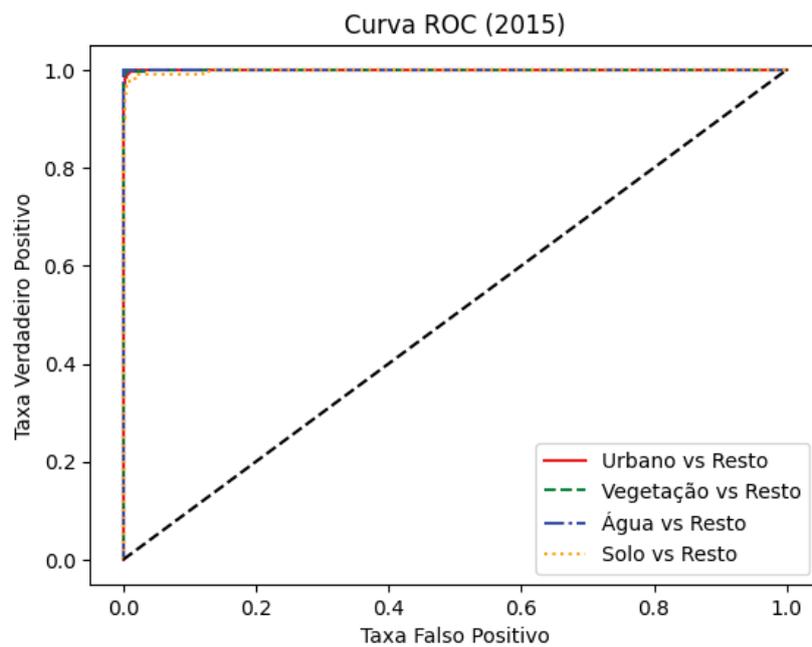
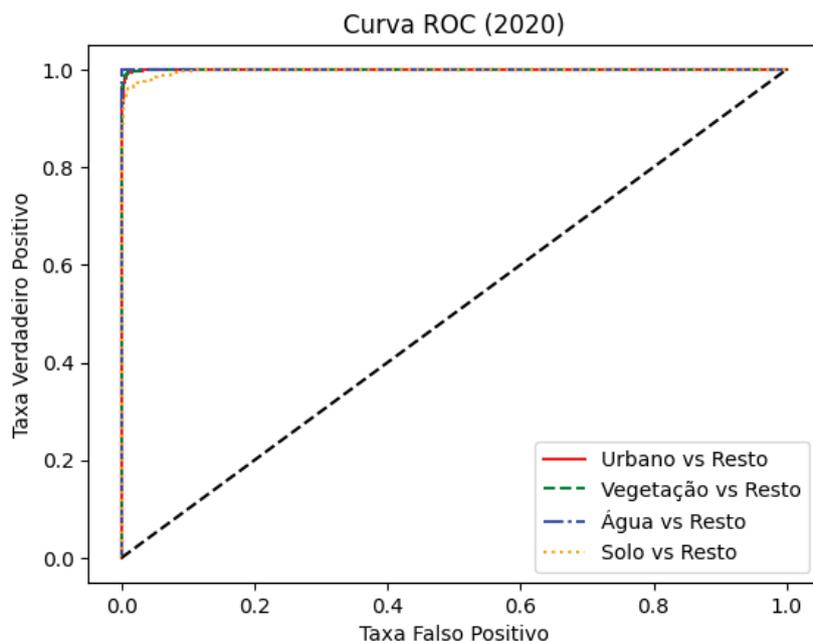


FIGURA 26 - CURVA ROC PARA TODAS AS CLASSES DA CLASSIFICAÇÃO DA IMAGEM DE 2020



4.3 VARIÁVEIS E FATORES DIRECIONADORES

As variáveis e fatores direcionadores, resultantes do processo apresentado na seção metodológica deste documento, são apresentados neste tópico.

A FIGURA 27 representa as variáveis de mancha urbana obtidas da classificação para os anos de 2000, 2005, 2010, 2015 e 2020. A FIGURA 28 representa a mancha urbana nova em um período de 5 anos resultante de uma álgebra de imagens das variáveis de mancha urbana. As figuras do 29 ao 32 representam os diferentes tipos de ruas da região, utilizados como fator impulsionador, foram obtidas do Open Street Maps através do plugin do QGIS, OSMDownloads. A FIGURA 33 representa o fator impulsionador de distância das zonas centrais do zoneamento, este obtido da COMEC e no QGIS gerado um centróide para cada zona central. As figuras 34 e 35 representam a distância da mancha urbana e a contagem de vizinhos urbanos respectivamente, usados como fator impulsionador. A FIGURA 36 e 37 representam a densidade populacional da região para o ano 2000 e 2010 respectivamente, criado no ambiente QGIS onde foi feita uma união pelo campo da tabela do censo e da tabela de atributos da malha censitária disponibilizado pelo IBGE. A FIGURA 38 representa a declividade da

região e a FIGURA 39 apresenta os corpos d'água da região com os dados fornecidos do IAT.

Considerando as manchas urbanas das Figuras 27 e 28, a tabela 14 apresenta os valores de áreas, em Km², da mancha urbana na região e o quantitativo de aumento da mancha urbana que aumentou em 5 anos

TABELA 14 - ÁREA URBANA NA ÁREA DE ESTUDO

Ano	Área Urbana (Km²)	Diferença de Área (Km²)
2000	389,0502	
2005	420,9291	31,8789
2010	437,9508	17,0217
2015	457,3629	19,4121
2020	472,5648	15,2019

FIGURA 27 – VARIÁVEL DE MANCHA URBANA DO ANO 2000, 2005, 2010, 2015, 2020

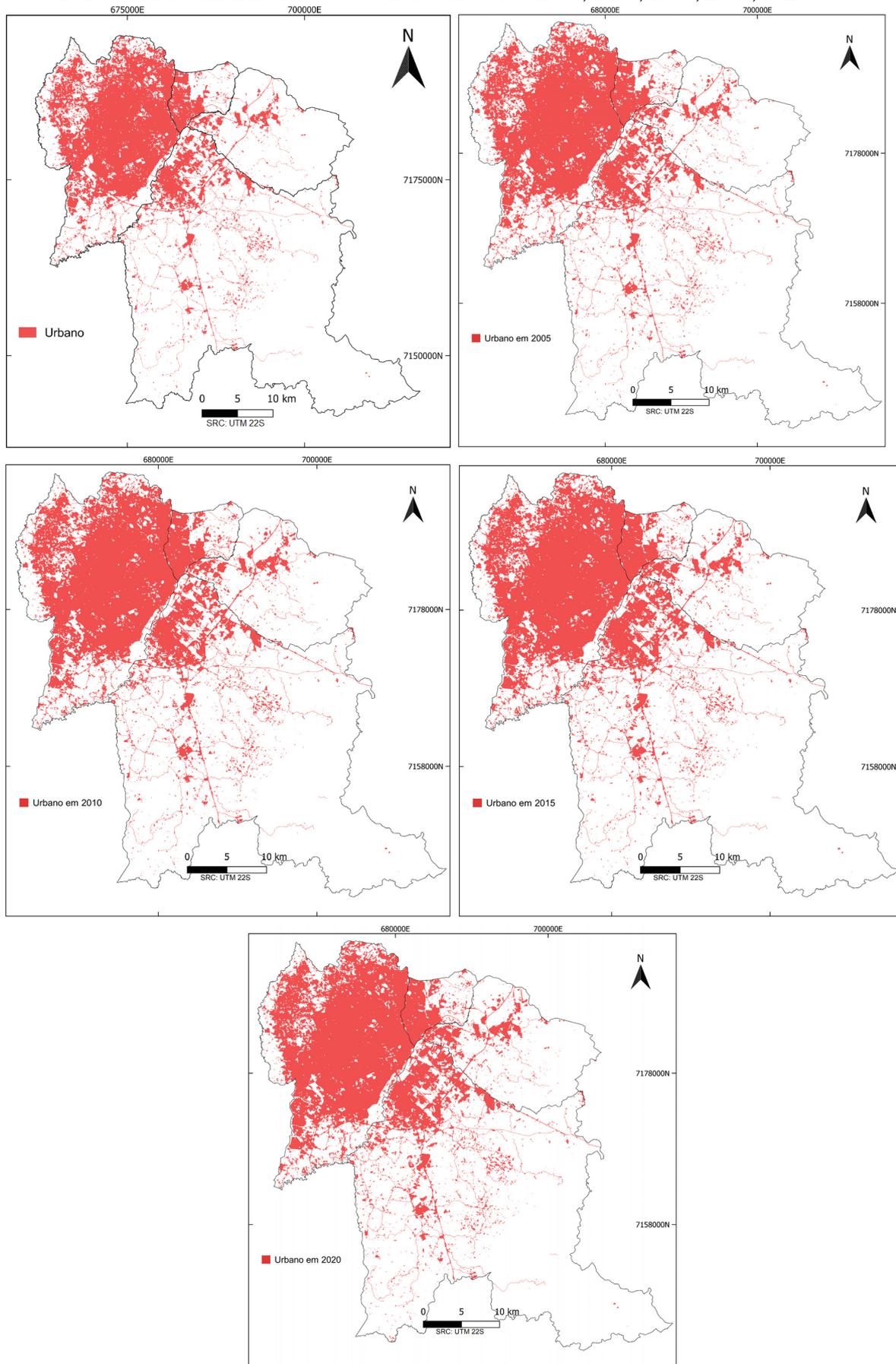


FIGURA 28 – MANCHA URBANA QUE CRESCEU NO PERÍODO DE 5 ANOS

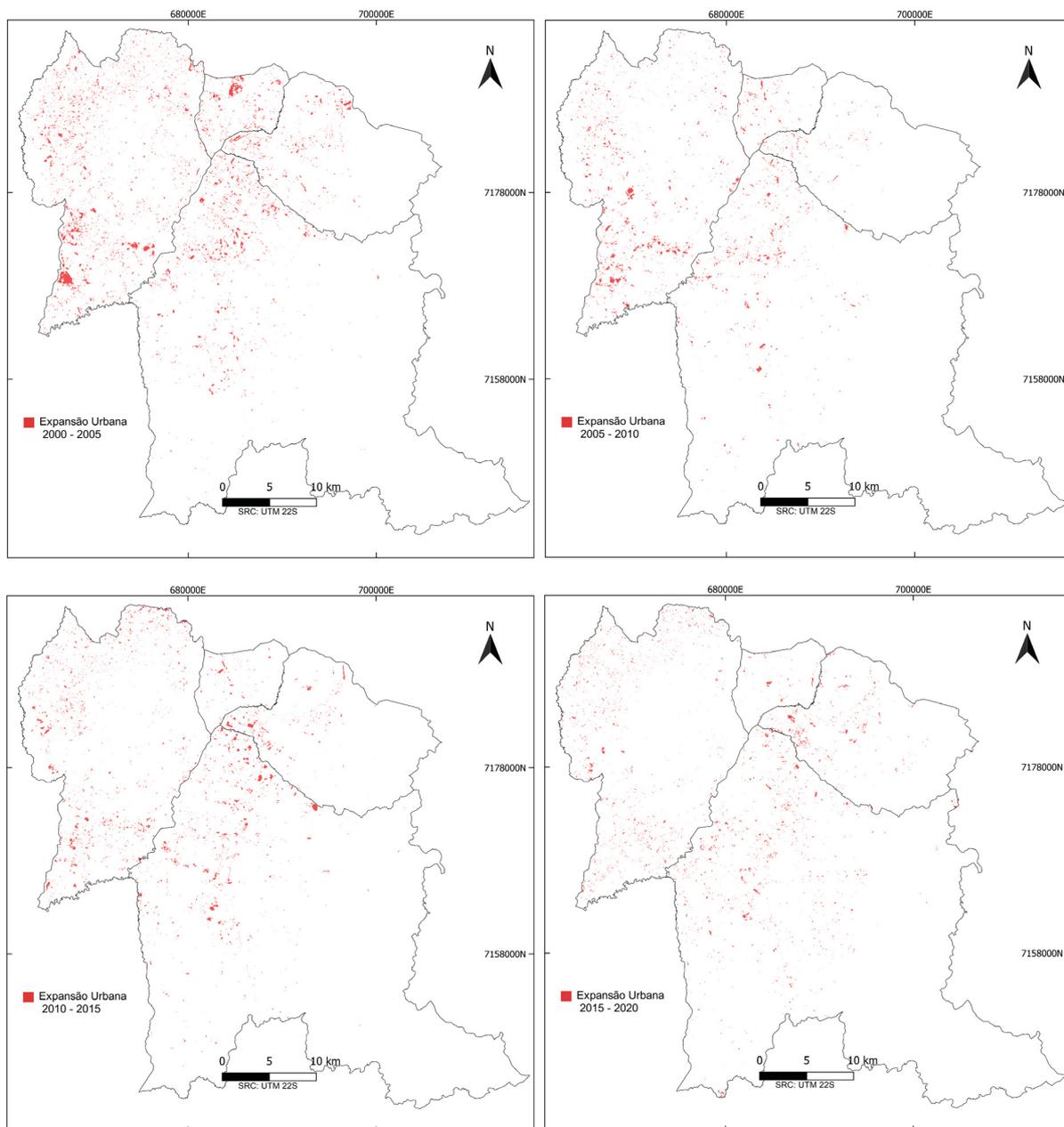


FIGURA 29 – VARIÁVEL DE DISTÂNCIA DAS RODOVIAS

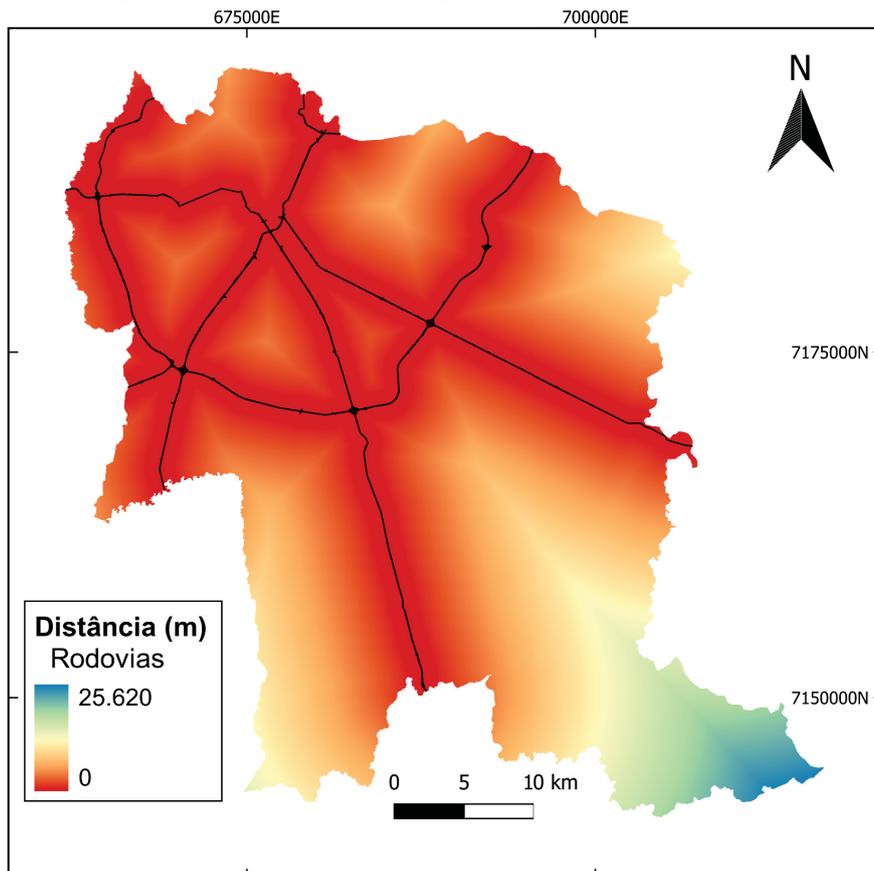


FIGURA 30 – VARIÁVEL DE DISTÂNCIA DAS RUAS PRIMÁRIAS

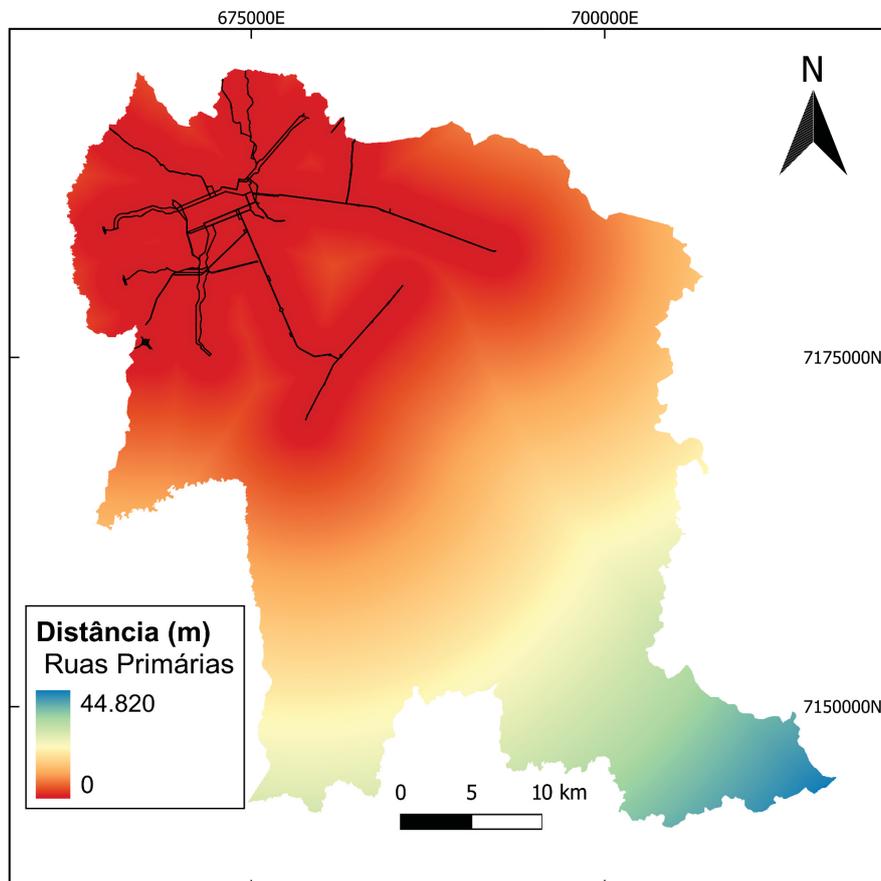


FIGURA 31 – VARIÁVEL DE DISTÂNCIA DAS RUAS SECUNDÁRIAS

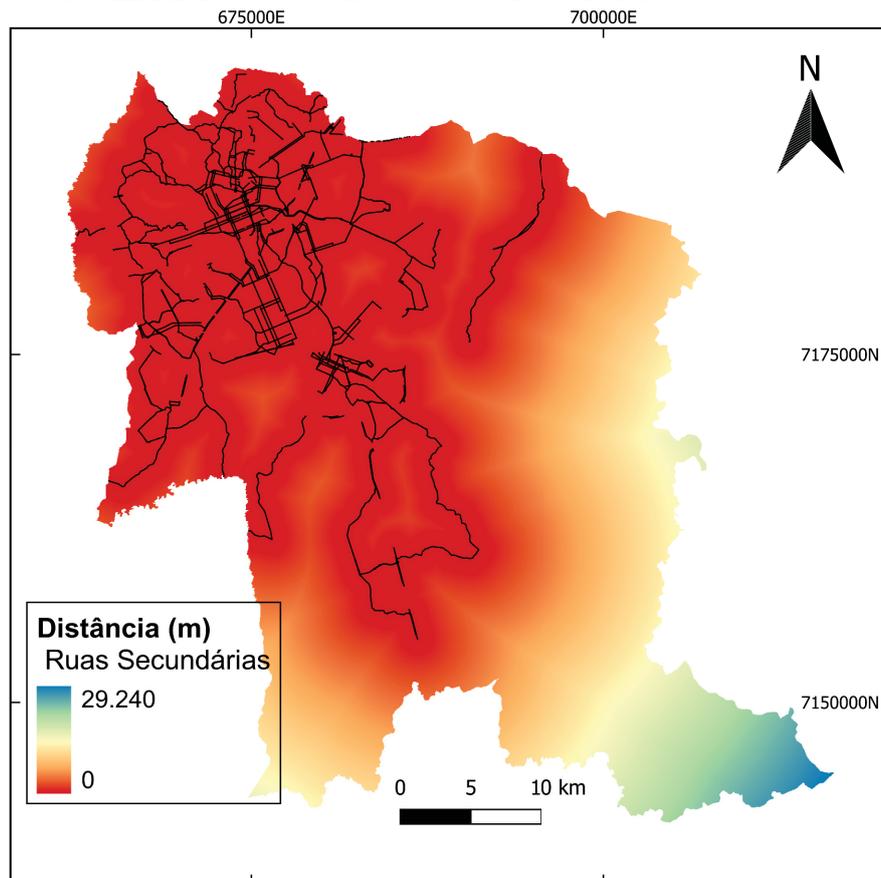


FIGURA 32 – VARIÁVEL DE DISTÂNCIA DAS RUAS TERCIÁRIAS

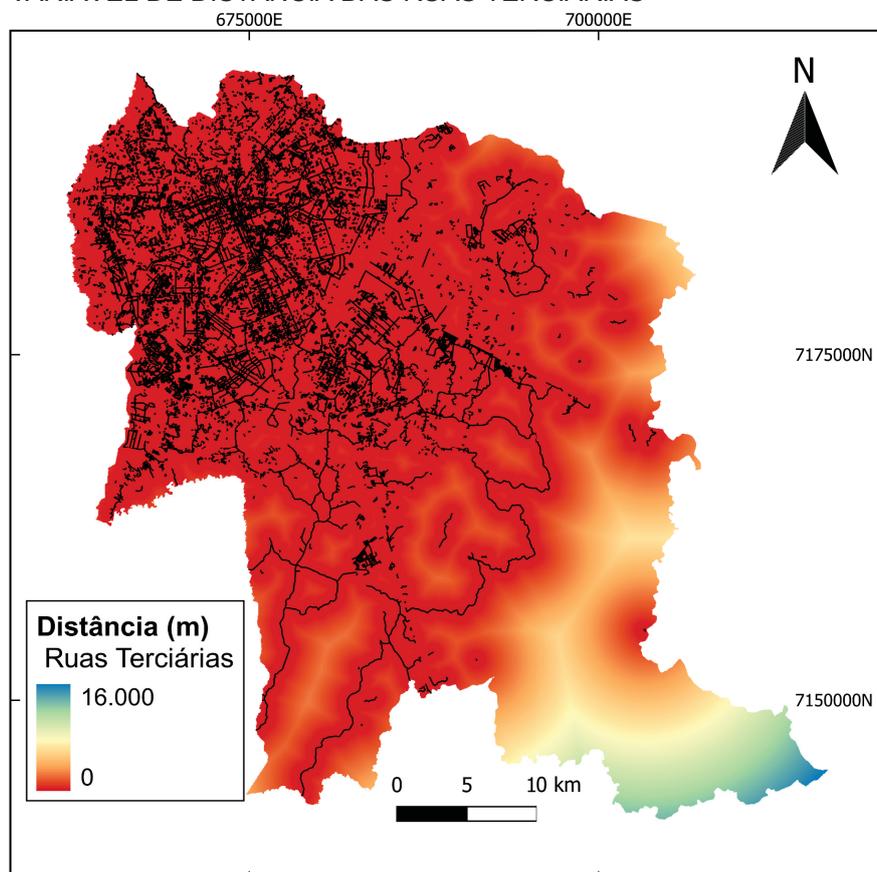


FIGURA 33 – VARIÁVEL DE DISTÂNCIA DAS ZONAS CENTRAIS

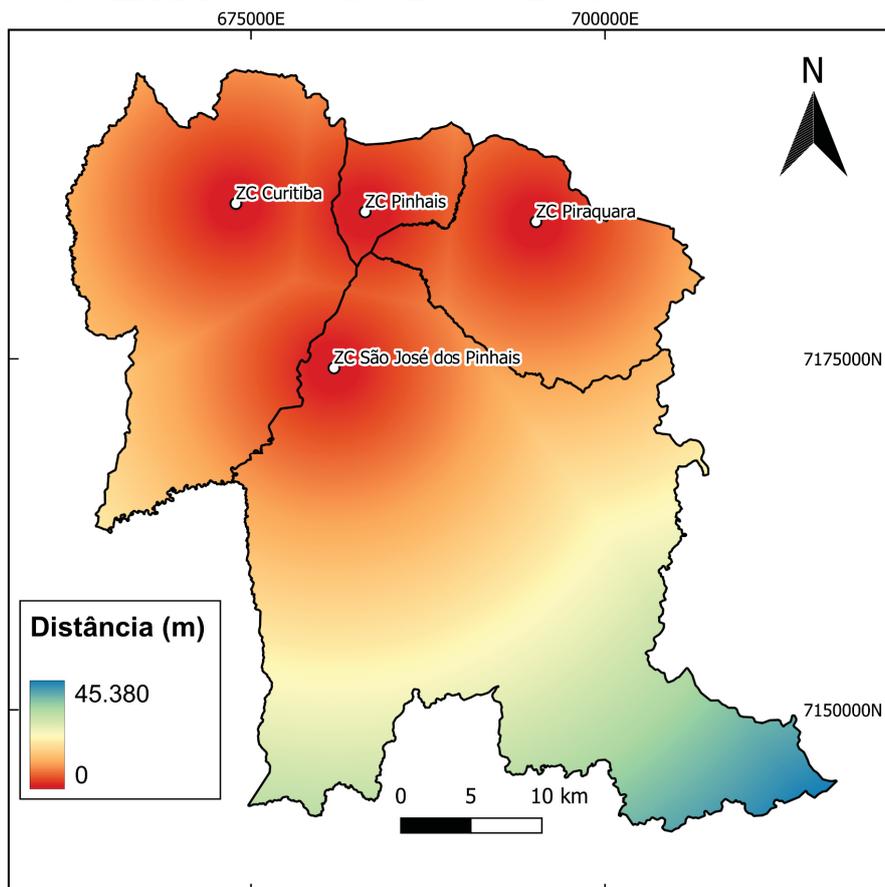


FIGURA 34 - VARIÁVEL DE DISTÂNCIA DA MANCHA URBANA

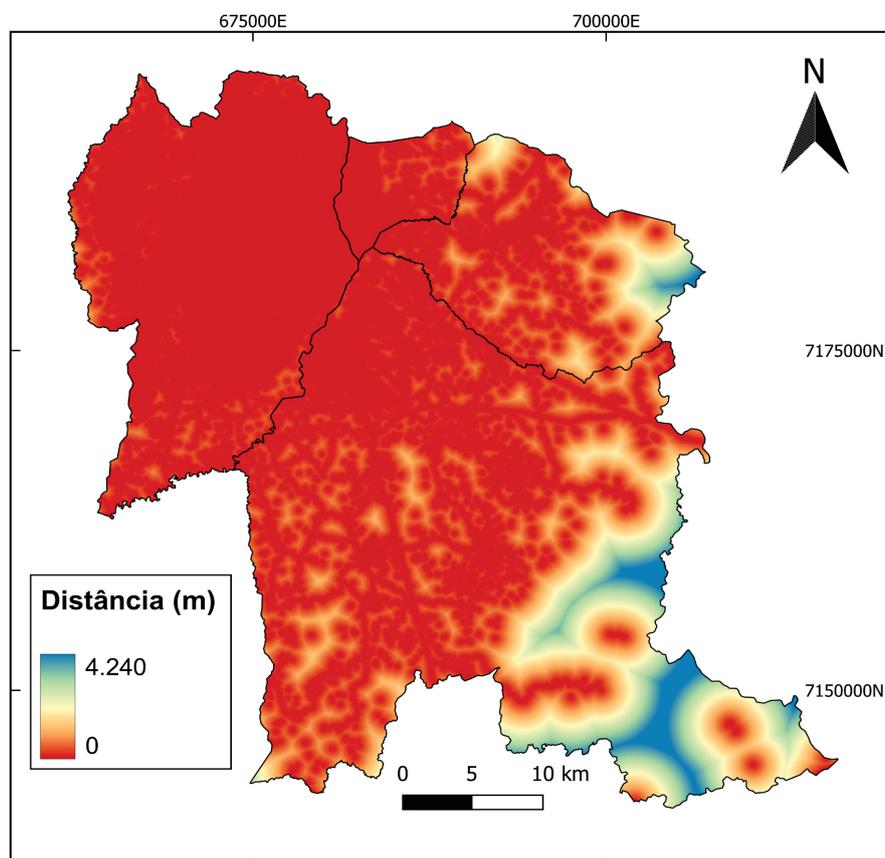


FIGURA 35 - VARIÁVEL DE VIZINHOS URBANOS

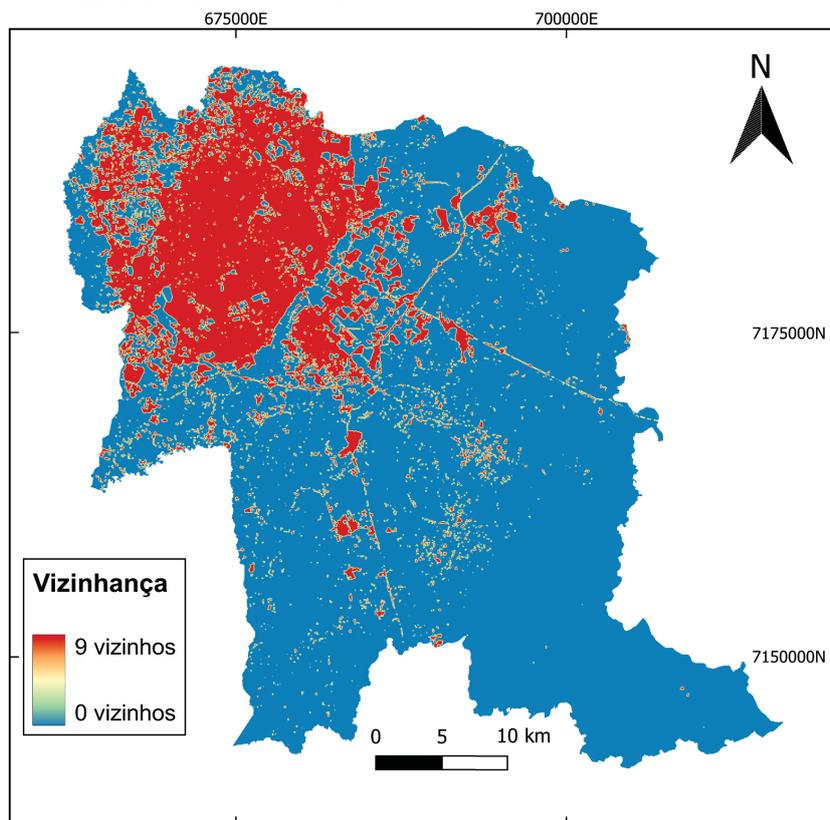


FIGURA 36 - DENSIDADE POPULACIONAL PARA O ANO DE 2000

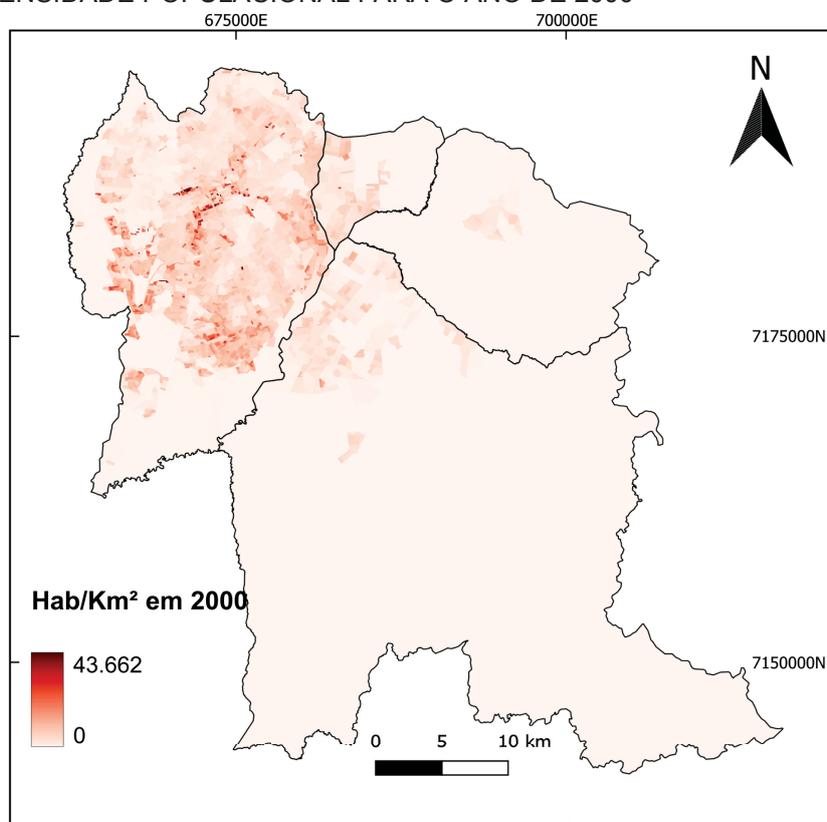


FIGURA 37 - DENSIDADE POPULACIONAL PARA O ANO DE 2010

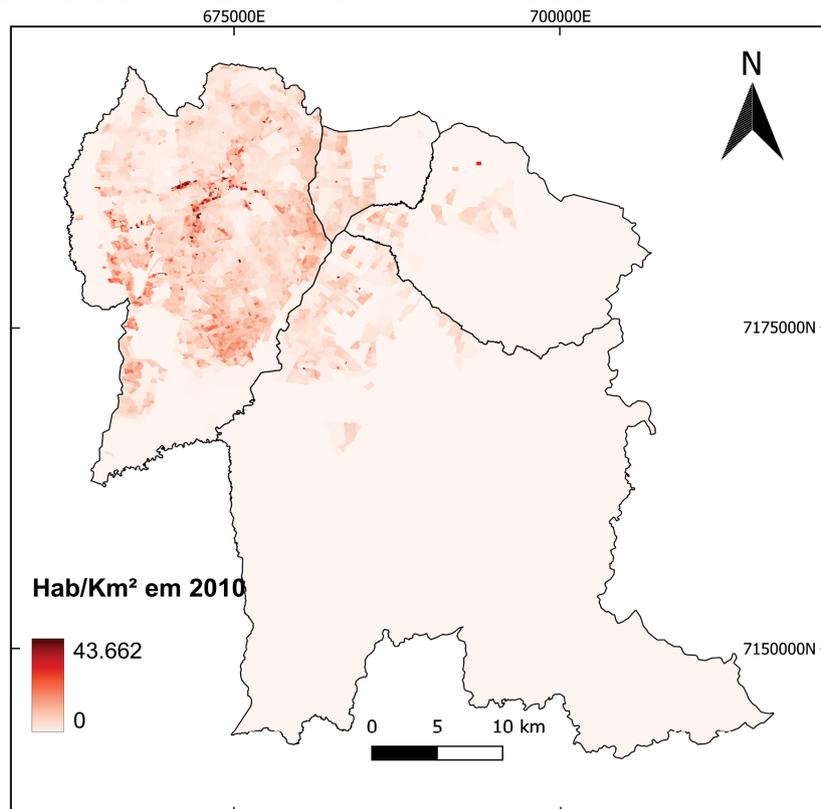


FIGURA 38 - DECLIVIDADE EM GRAUS DA ÁREA DE ESTUDO

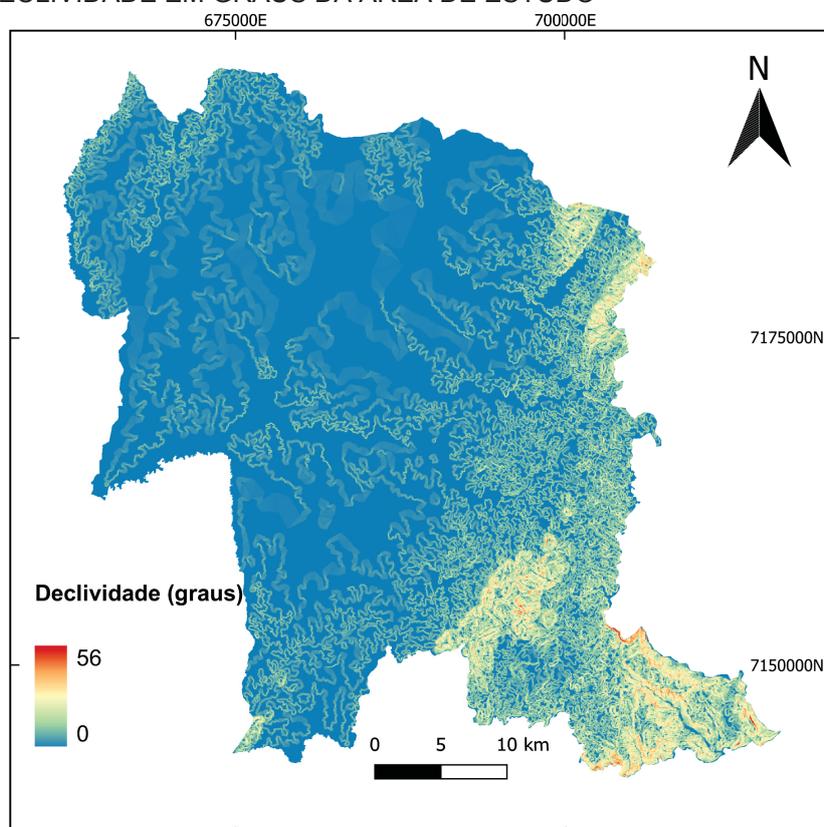
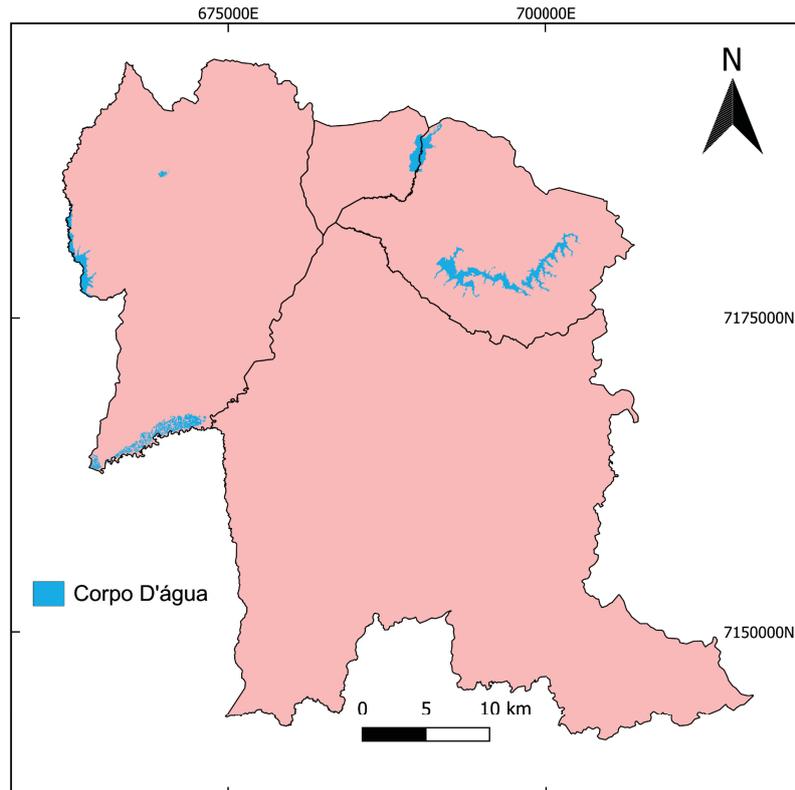


FIGURA 39 - CORPOS D'ÁGUA DA ÁREA DE ESTUDO



4.4 SIMULAÇÃO

A etapa de simulação consistiu em utilizar as variáveis e os fatores direcionadores no algoritmo de regressão RF para obter o potencial de transição e posteriormente utilizar o modelo de AC para obter o resultado da simulação.

Foi compilada uma tabela, para cada período a ser simulado, com o uso da biblioteca Pandas para ser usada pelo regressor RF utilizado com a biblioteca *Scikit-learn*. Na tabela cada linha é um *pixel* da imagem, e as colunas são os valores das variáveis de fatores direcionadores. Um exemplo de tabela usada no treinamento para o período de 2000 a 2005 é mostrado na TABELA 15.

TABELA 15 - TABELA USADA NO TREINAMENTO PARA O PERÍODO 2000 - 2005

	decliv	pop_2000	dist_centro	dist_ruas	dist_urb	vizin2000	Urb2000	PoTr2005	L	C
0	6.864781	2126.0	9480.189453	1129.291748	67.082039	0.0	0.0	0.0	0.0	396.0
1	10.145716	2126.0	9480.047852	1152.562378	94.868332	0.0	0.0	0.0	0.0	397.0
2	12.119548	2126.0	9480.000000	1176.137695	108.166534	0.0	0.0	0.0	0.0	398.0
3	4.934152	2126.0	9450.761719	1064.048828	0.000000	1.0	1.0	0.0	1.0	394.0
4	9.529784	2126.0	9450.428711	1087.060181	30.000000	1.0	0.0	0.0	1.0	395.0
...
1854948	2.820058	0.0	40667.753906	19113.673828	2707.655762	0.0	0.0	0.0	1804.0	1423.0
1854949	1.772422	0.0	40685.355469	19140.093750	2737.243164	0.0	0.0	0.0	1804.0	1424.0
1854950	9.371562	0.0	40702.968750	19166.525391	2766.839355	0.0	0.0	0.0	1804.0	1425.0
1854951	0.487109	0.0	40674.480469	19101.519531	2683.281494	0.0	0.0	0.0	1805.0	1422.0
1854952	0.107145	0.0	40692.054688	19127.912109	2712.802979	0.0	0.0	0.0	1805.0	1423.0

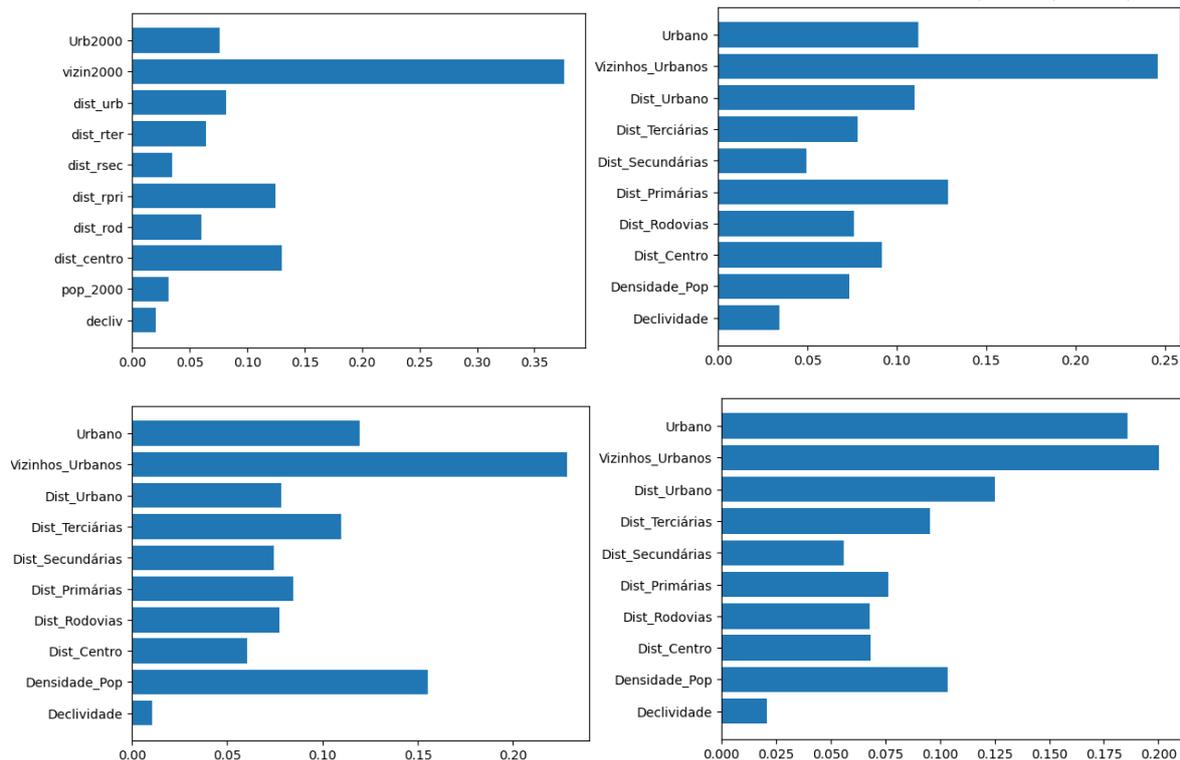
1854953 rows x 10 columns

No modelo de AC é estimado, por meio de tentativa e erro, o limiar do potencial de transição e é aplicado a área de restrição para obter o resultado da simulação.

4.5 AVALIAÇÃO DOS RESULTADOS DA SIMULAÇÃO

O algoritmo de regressão RF para a determinação do potencial de transição foi utilizado com a biblioteca Scikit-learn. Dos pontos de amostra 75% foram utilizados para o treinamento e 25% para a validação, modelo composto de 10 mil árvores com profundidade de 10 níveis.

FIGURA 40 - IMPORTÂNCIA DAS VARIÁVEIS PARA AS MODELAGENS DE 2005, 2010, 2015, 2020



O algoritmo RF permite determinar o valor da importância das variáveis, que foi computado para os 4 períodos simulados, com o uso da biblioteca *Scikit-learn*, e é apresentado na FIGURA 40.

- Em todos os casos a variável de maior importância foi a de contagem de vizinhos urbanos. E a variável de declividade apresenta ser o de menor importância, apresentando um valor menor que 0,05, isto é plausível devido ao fato que maior parte da área de estudo se concentra numa planície.
- Na modelagem para 2005, percebe-se que as variáveis de distância das ruas primárias e das zonas centrais são as próximas com maior importância.
- Para 2010, deu-se menos importância para a variável de vizinhança urbana se comparada com o anterior, e outras variáveis ganharam mais importância, como: células urbanas, distância para células urbanas, e distância para ruas primárias.
- Para 2015, destaca-se a importância dada para as variáveis: células urbanas, distância para as ruas terciárias e densidade populacional.
- E para 2020, a variável de mancha urbana foi a que recebeu maior importância dentre as 4 modelagens, com destaque também para as variáveis:

distância para a mancha urbana, distância para as ruas terciárias e densidade populacional.

As figuras 41 a 44 mostram o potencial de transição de uma célula não urbana para urbano determinado pelo RF. Nota-se a capacidade do modelo prever o desenvolvimento de novas áreas de mancha urbana além de áreas em torno de áreas urbanas já consolidadas, superando modelos mais simples de simulação da expansão urbana como o utilizado por Peschl (2021). Contudo, o modelo não consegue prever o desenvolvimento de áreas urbanas com uma certa distância de áreas urbanas já existentes, os chamados Saltos de Sapo (*Leapfrog*).

FIGURA 41 - POTENCIAL DE TRANSIÇÃO PARA 2005

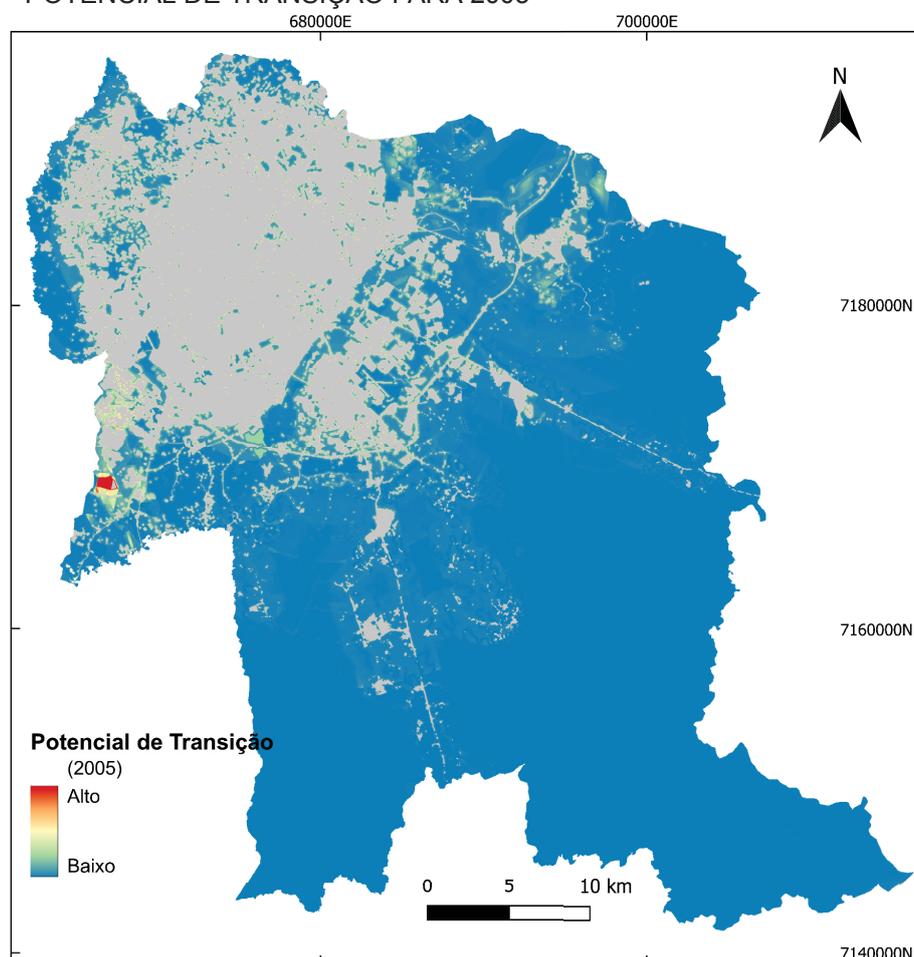


FIGURA 42 - POTENCIAL DE TRANSIÇÃO PARA 2010

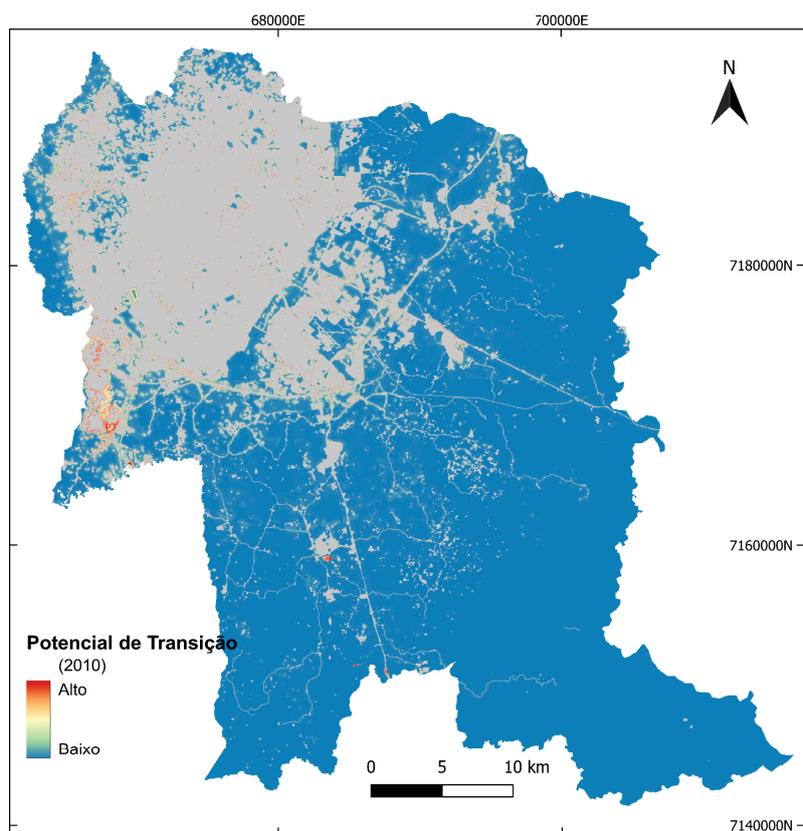


FIGURA 43 - POTENCIAL DE TRANSIÇÃO PARA 2015

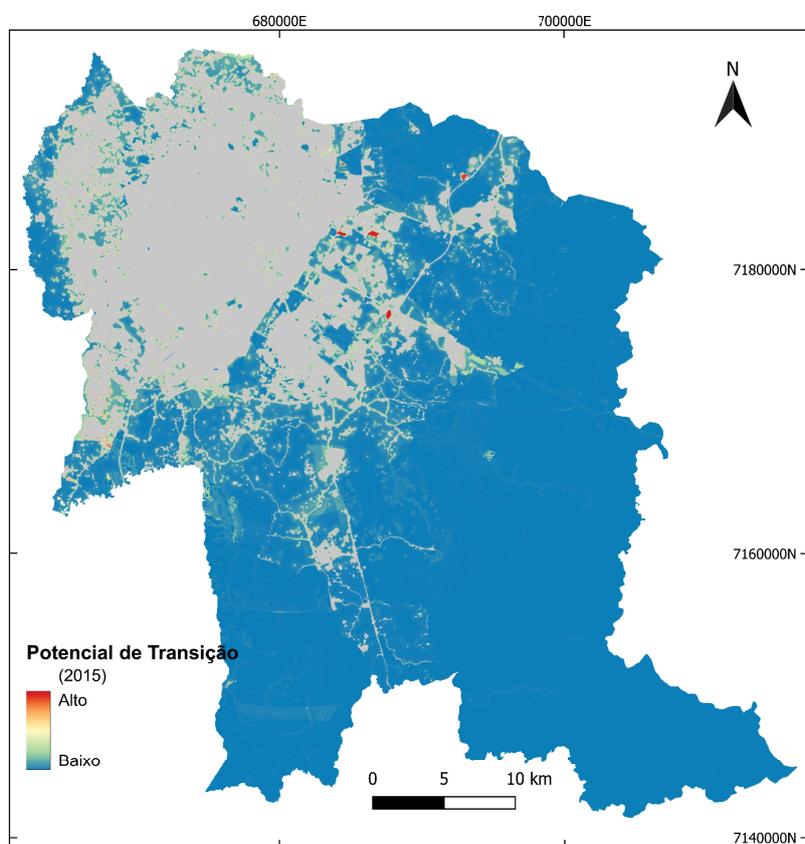
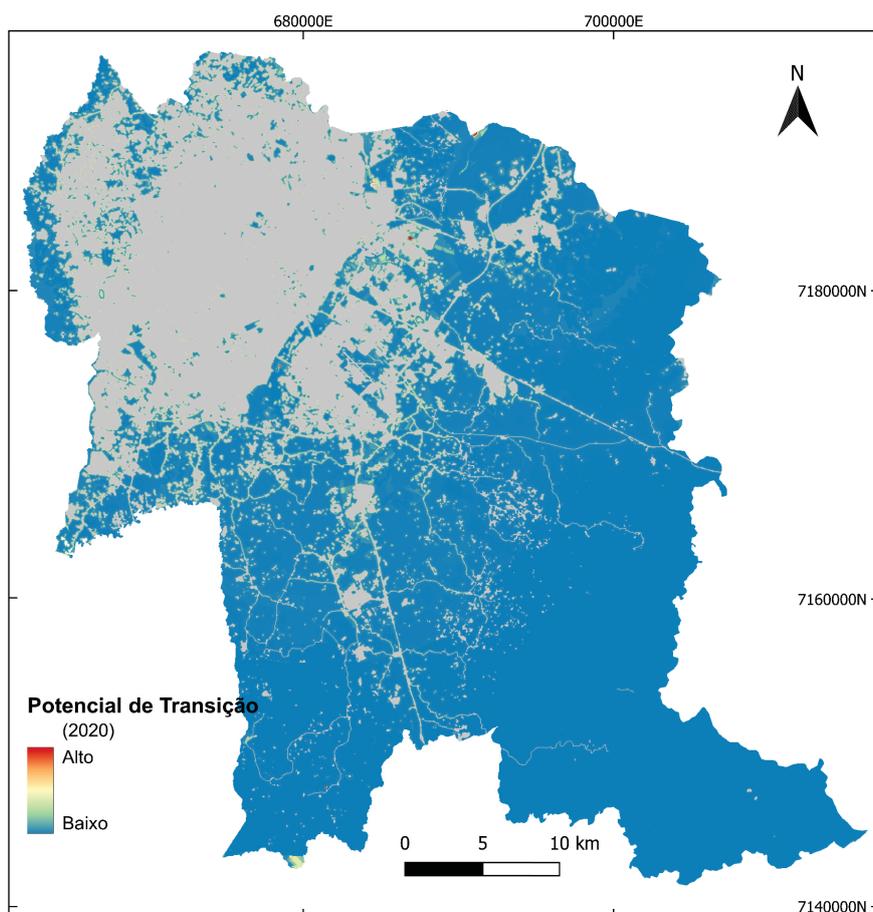
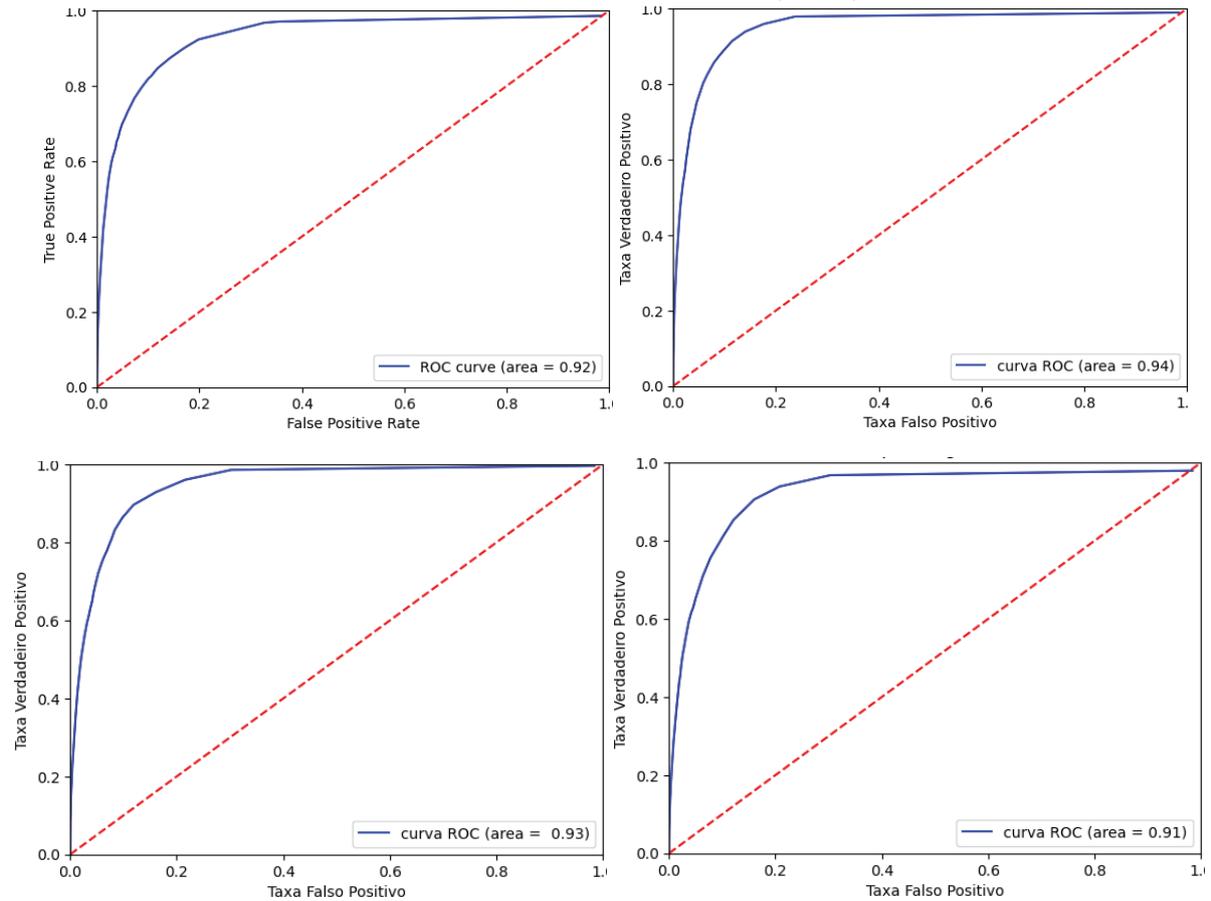


FIGURA 44 - POTENCIAL DE TRANSIÇÃO PARA 2020



A curva ROC foi calculada para os 4 modelos, plotado com a biblioteca Matplotlib, apresentados na FIGURA 45. Apesar da diferença entre as curvas serem discretas nota-se que a curva correspondente a modelagem para o ano de 2010 tem um resultado melhor que os demais, com o valor de AUC de aproximadamente de 0.94. As demais curvas mostram que as 4 modelagens tiveram bons desempenhos, sendo o menor valor da AUC sendo de 0.91.

FIGURA 45 - CURVAS ROC PARA AS MODELAGENS DE 2005, 2010, 2015 E 2020



As figuras de 46 a 49 mostram as alterações que realmente aconteceram e as mudanças simuladas. Os *pixels* em verde são as áreas onde ocorreu o crescimento urbano que o modelo corretamente previu, em vermelho são os *pixels* que o modelo previu com crescimento urbano que na realidade não ocorreu, e em amarelo são expansões que ocorreram na realidade mas o modelo não previu.

FIGURA 46 - COMPARAÇÃO ENTRE O SIMULADO E O REAL DE 2000 PARA 2005

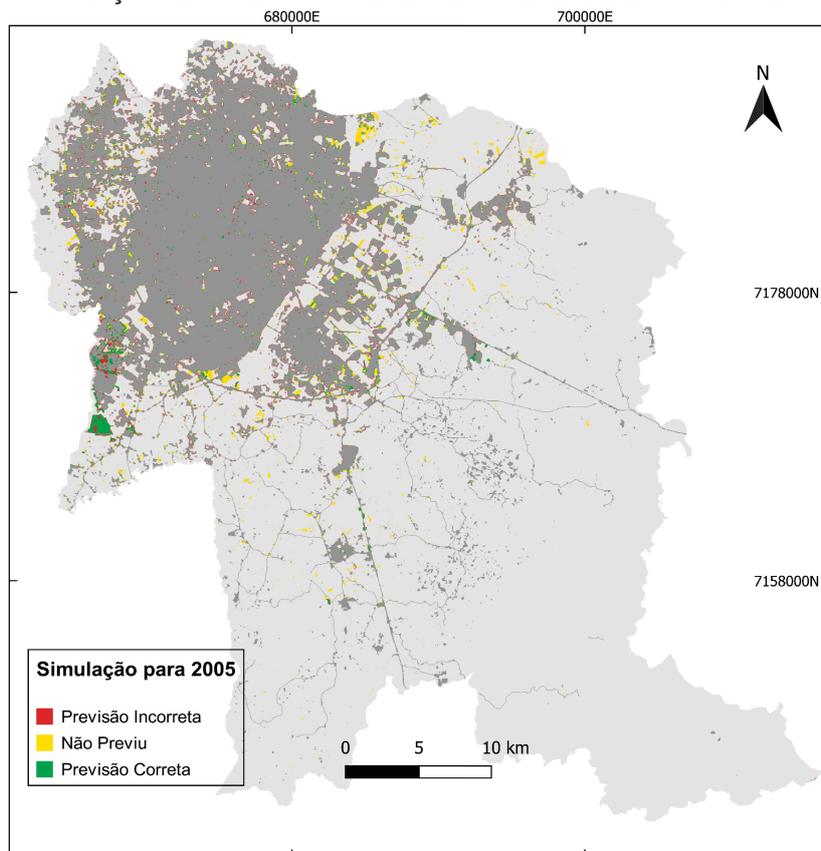


FIGURA 47 - COMPARAÇÃO ENTRE O SIMULADO E O REAL DE 2005 PARA 2010

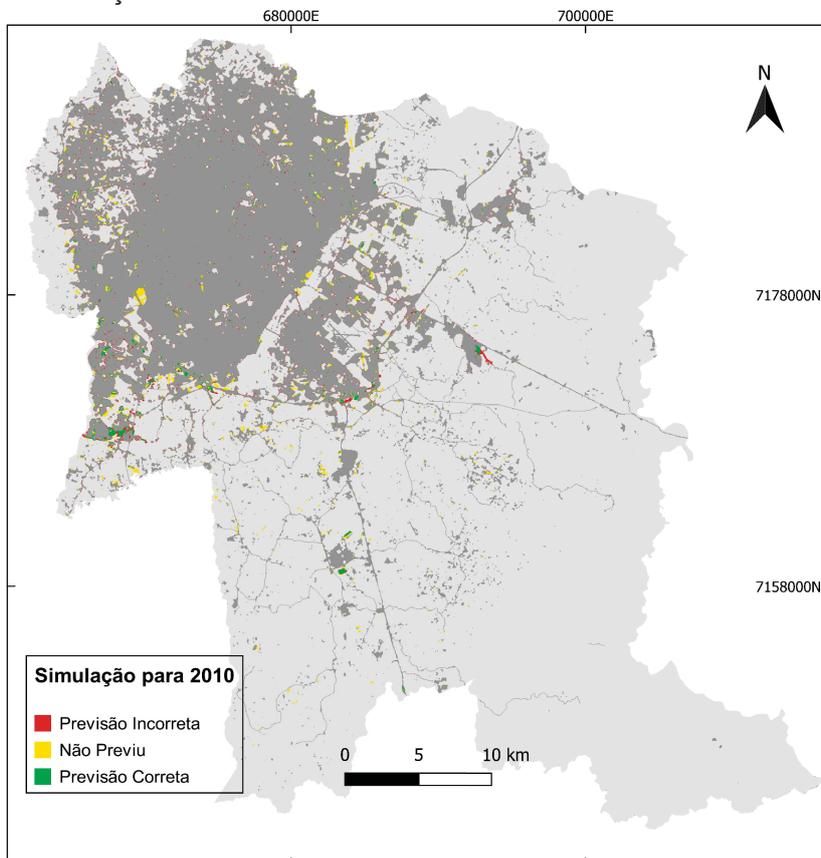


FIGURA 48 - COMPARAÇÃO ENTRE O SIMULADO E O REAL DE 2010 PARA 2015

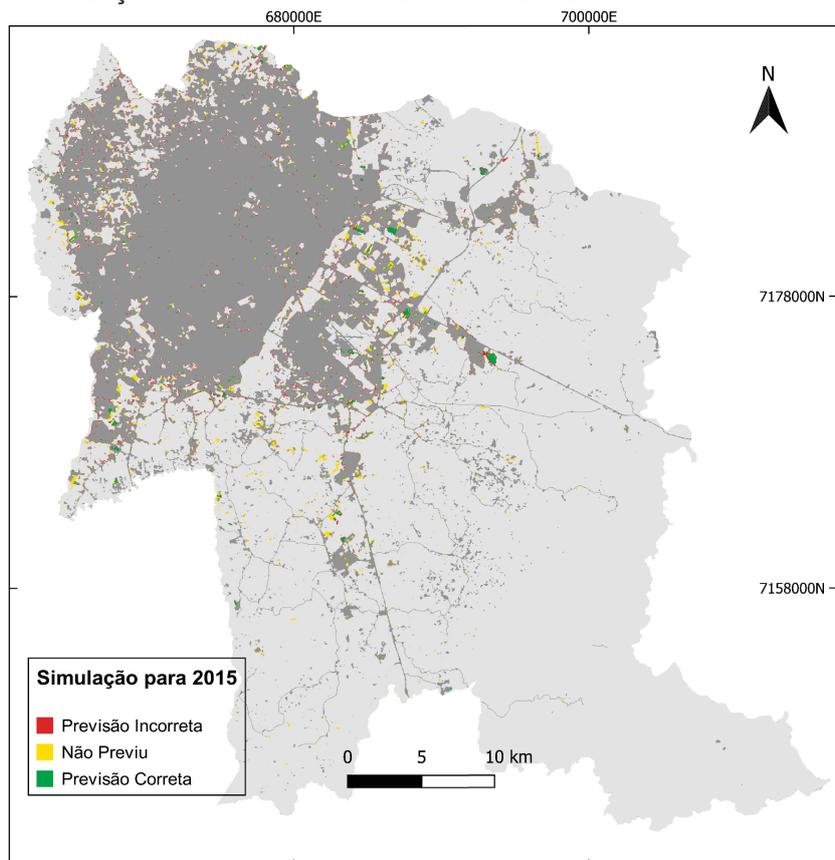
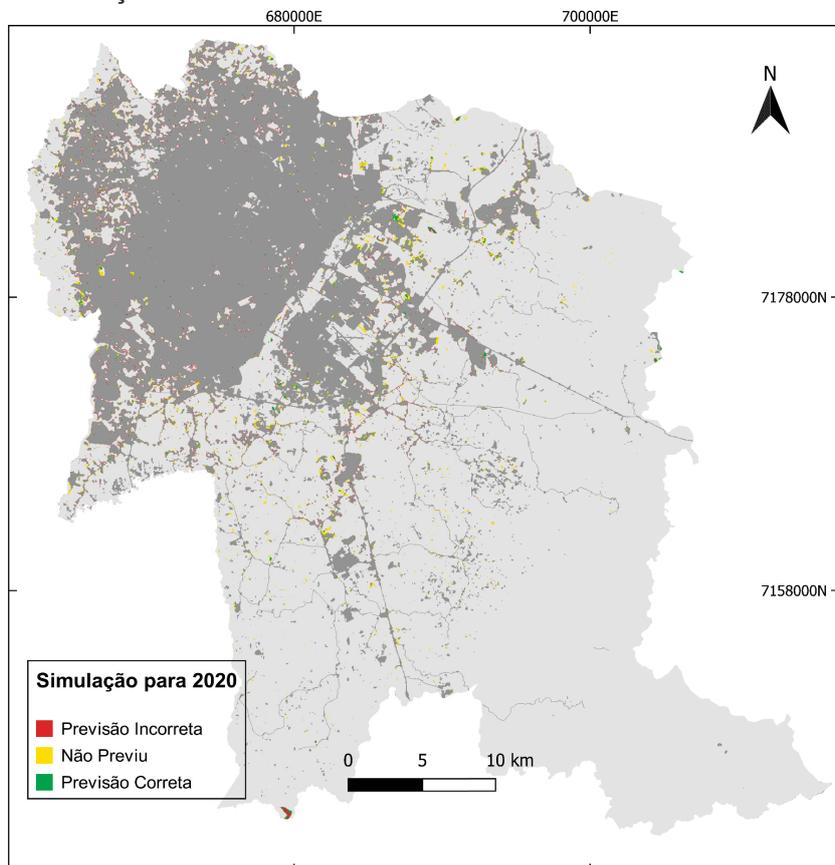


FIGURA 49 - COMPARAÇÃO ENTRE O SIMULADO E O REAL DE 2015 PARA 2020



Destaca-se dentre essas simulações a capacidade do modelo prever o crescimento de algumas áreas, como por exemplo no bairro Campo de Santana de 2000 para 2005. É possível afirmar que há muitas áreas que o modelo não previu o crescimento, isto ainda é discutido por estudos neste sentido, o fenômeno da expansão urbana é de difícil previsão e tem um alto nível de aleatoriedade.

As tabelas 16 a 19 mostram a matriz de erros, calculado com o *Scikit-learn*, do resultado da simulação para os 4 diferentes períodos. O conjunto de amostras utilizado nas matrizes de erros inclui todos os pixels da imagem resultante da simulação, com exceção daqueles que permaneceram inalterados, ou seja, aqueles que eram urbanos e continuaram urbanos depois do período de 5 anos. As previsões tiveram acurácia total alta, a menor sendo de 2005 (97,26%), contudo mostra baixa acurácia para prever corretamente os novos *pixels* urbanos, a maior acurácia sendo de 2005 (45%).

TABELA 16 - MATRIZ DE ERROS PARA A SIMULAÇÃO DE 2005

		Real 2005		Total	Acurácia Produtor
		Urbano	Nao Urbano		
Previsão	Urbano	15944	19477	35421	45,01%
	Não Urbano	19562	1367540	1387102	98,59%
Total		35506	1387017	1422523	
Acurácia Usuário		44,91%	98,60%		
Acurácia Total		97,26%			

TABELA 17 - MATRIZ DE ERROS PARA A SIMULAÇÃO DE 2010

		Real 2010		Total	Acurácia Produtor
		Urbano	Nao Urbano		
Previsão	Urbano	7042	11871	18913	37,23%
	Não Urbano	11871	1356509	1368380	99,13%
Total		18913	1368380	1387293	
Acurácia Usuário		37,23%	99,13%		
Acurácia Total		98,29%			

TABELA 18 - MATRIZ DE ERROS PARA A SIMULAÇÃO DE 2015

		Real 2015		Total	Acurácia Produtor
		Urbano	Nao Urbano		
Previsão	Urbano	7688	13882	21570	35,64%
	Não Urbano	13881	1332929	1346810	98,97%
Total		21569	1346811	1368380	
Acurácia Usuário		35,64%	98,97%		
Acurácia Total		97,97%			

TABELA 19 - MATRIZ DE ERROS PARA A SIMULAÇÃO DE 2020

		Real 2020		Total	Acurácia Produtor
		Urbano	Nao Urbano		
Previsão	Urbano	4957	11934	16891	29,35%
	Não Urbano	11934	1317984	1329918	99,10%
Total		16891	1329918	1346809	
Acurácia Usuário		29,35%	99,10%		
Acurácia Total		98,23%			

A TABELA 20 mostra os valores do FoM (Figura de Mérito) para as 4 simulações. Muitos estudos têm esse valor em torno de 0.3 como ressaltado por Liu et al. (2022). A TABELA 21 compara valores do FoM de alguns estudos encontrados que usam também o RF e com o valor médio do FoM obtido neste estudo.

TABELA 20 - VALORES DE FOM PARA OS 4 PERÍODOS SIMULADOS

	FoM
2005	0,2900
2010	0,2288
2015	0,2169
2020	0,1720

TABELA 21 - COMPARAÇÃO DE VALORES DE FOM COM OUTROS ESTUDOS

	FoM
Este Estudo	0,2263
Lv et al.(2021)	0,2040
Kamusoko e Gamba (2015)	0,4700
Chen (2022)	0,1406

Comparando a outros estudos (TABELA 21) pode-se afirmar que o valor de FoM está aceitável, mas isto é apenas para um mera comparação. Deve-se ressaltar que estes estudo foram realizados em regiões urbanas diferentes pelo mundo, diferentes espaços de tempos, diferentes variáveis consideradas, diferentes tamanhos e quantidade das árvores do RF, enfim, não se pode afirmar que uma simulação é superior a outra da TABELA 21.

5 CONCLUSÃO

O objetivo principal deste estudo foi de simular a expansão urbana com AC com potencial de transição obtidas com o uso do Random Forest para os municípios de Curitiba, Pinhais, Piraquara e São José dos Pinhais considerando diversas variáveis que influenciam no crescimento urbano. Nesse contexto, buscou-se investigar a influência de diferentes variáveis de ruas no modelo e diferentes profundidades das árvores do RF, de forma a obter os melhores resultados das simulações.

Pode-se afirmar que o objetivo geral deste estudo foi atingido. O método *Random Forest* com Autômatos Celulares mostrou uma notável capacidade de aprender com diversos fatores direcionadores e determinar um potencial de transição fiel à realidade, e capaz de prever a criação de grandes novas áreas urbanas. Os resultados das simulações da expansão urbana, demonstraram-se satisfatórios levando em conta as diversas métricas utilizadas: curva ROC, AUC e matriz de erros.

É essencial considerar os cenários futuros que essas simulações podem indicar, neste estudo não foram considerados variáveis restritivas como área de proteção ambiental ou zoneamento. Neste tipo de cenário, que ilustra uma expansão urbana sem restrições ou monitoramento, apresenta o potencial de estudar os impactos negativos deste tipo de expansão, como a de ocupação de áreas sensíveis e comprometimento de recursos naturais. Isto enfatiza a importância de políticas de planejamento urbano e um zoneamento eficaz.

Vale ressaltar a importância da disponibilização de dados geográficos para a modelagem de simulação da paisagem, sem eles a autenticidade da simulação ficaria comprometida. Neste estudo chegou-se a um conjunto de variáveis, fatores direcionadores, que possibilitaram a modelagem efetiva das mudanças na paisagem da região. Além das variáveis usadas baseadas na literatura foi observado, por meio de experimentos, que a inclusão de variáveis que representam diferentes tipos de ruas aprimorou os resultados da simulação.

Essa modelagem, embora aplicada aos municípios mencionados, possui potencial de generalização para qualquer área urbana, desde que as variáveis específicas sejam adquiridas e os parâmetros do RF e as regras de transição do AC sejam devidamente refinados.

5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Para estudos futuros, recomenda-se a exploração de novos fatores direcionadores no crescimento urbano, como dados econômicos, e outras variáveis de restrição além de corpos d'água. Poderia considerar também como um modelo se comportaria com uma nuvem de pontos da cidade e, até mesmo, prever o crescimento urbano vertical.

Outras metodologias também podem ser utilizadas ou incorporadas. Existem estudos onde se utilizam do aprendizado profundo para simulações urbanas, como o uso de CNN (Rede Neural Convolutacional) (HE et al., 2018), RNN (Rede Neural Recorrente) (PAN et al., 2022) e ANN (Rede Neural Artificial) (QIAN et al., 2020). Estes estudos têm se beneficiado destes métodos atualmente populares obtendo melhores resultados.

REFERÊNCIAS

- ALMEIDA, C. M. de. **Modelagem da dinâmica espacial como uma ferramenta auxiliar ao planejamento: simulação de mudanças de uso da terra em áreas urbanas para as cidades de Bauru e Piracicaba (SP), Brasil**. Tese (Doutorado em Sensoriamento Remoto) -. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2004.
- BIAU, G.; SCORNET, E. A Random Forest Guided Tour. **TEST**, v. 25, n. 2, p. 197–227, 2016.
- BOISSAU, S.; CASTELLA, J. C. Constructing a Common Representation of Local Institutions and Land Use Systems through Simulation-Gaming and Multiagent Modeling in Rural Areas of Northern Vietnam: The SAMBA-Week Methodology. **Simulation & Gaming**, v. 34, n. 3, p. 342–357, 2003.
- BOUHENNACHE, R.; BOUDEN, T.; CHEDDAD, A. A new spectral index for the extraction of built-up land features from Landsat 8 satellite imagery. **Geocarto International**, v. 34, p. 1–32, 2018.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.
- CARVALHO, A. de S. **Vivendo às margens: habitação de interesse social e o processo da segregação socioespacial em Curitiba**. Dissertação (Mestrado em Urbanismo, História e Arquitetura) - Florianópolis: Universidade Federal de Santa Catarina, 2014.
- CASTRO, M. L. A.; CASTRO, R. de O. AUTÔMATOS CELULARES: IMPLEMENTAÇÕES DE VON NEUMANN, CONWAY E WOLFRAM. **Revista de Ciências Exatas e Tecnologia Vol. III**, Nº. 3, p. 89-106, 2008.
- CHAUDHURI, G.; CLARKE, K. The SLEUTH land use change model: A review. **International Journal Of Environmental Resource Research**, v. 1, p. 88–105, 2013.
- CHEN, Y. An extended patch-based cellular automaton to simulate horizontal and vertical urban growth under the shared socioeconomic pathways. **Computers, Environment and Urban Systems**, v. 91, p. 101727, 2022.
- CLARKE, K.; HOPPEN, S.; J, G. A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area. **Environment and Planning B: Planning and Design**, v. 24, p. 247–261, 1997.
- COMEC. **Proposta de implantação de área de interesse social de ocupação - AISO Guarituba - Piraquara**. Curitiba, 1999.
- DESCHAMPS, M. V. **Vulnerabilidade socioambiental na região metropolitana de Curitiba/PR**. Tese (Doutorado em Meio Ambiente e Desenvolvimento) - Curitiba, Universidade Federal do Paraná, 2004.

DIDÁTICA TECH, **Como funciona o algoritmo de Árvore de Decisão (Decision Tree)**. Disponível em:

<https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao>.

EMBRAPA, **Landsat**. Disponível em: <https://www.embrapa.br/satelites-de-monitoramento/missoes/landsat>.

ENGELEN, G.; ULJEE, I.; WHITE, R. **Report & SIMLUCIA User Manual**. 1997.

ENGELEN, G. WHITE, R.; TON, A.; NIJS, T.. Environment Explorer: Spatial Support System for the Integrated Assessment of Socio-Economic and Environmental Policies in the Netherlands. **Integrated Assessment**, v. 00, p. 97–105, 2003.

FREITAS, E. V. de. **Simulação do crescimento urbano: estudo de caso para a cidade de Curitiba**. 105 f. Trabalho de Graduação (Bacharelado em Engenharia Cartográfica e de Agrimensura) - Curitiba, Universidade Federal do Paraná, 2019.

GARCIAS, C. M.; SANCHES, A. M. Vulnerabilidades socioambientais e as disponibilidades hídricas urbanas: levantamento teórico conceitual e análise aplicada à região metropolitana de Curitiba –PR. **Revista de pesquisa em arquitetura e urbanismo**, Programa de Pós-Graduação do Departamento de Arquitetura e Urbanismo, USP, São Paulo, 2009.

GODOY, J. de. **A Expansão urbana e a reorganização da área metropolitana: o caso de São José dos Pinhais no contexto da Região Metropolitana de Curitiba**. Dissertação (Mestrado em Urbanismo, História e Arquitetura) - Florianópolis: Universidade Federal de Santa Catarina, 2012.

GOMEDE, E. The ROC Curve: Application and Interpretation in the Health Context. **Medium**, 2023.

HE, J.; LI, X.; YAO, Y.; HONG, Y.; JINBAO Z.; Mining transition rules of cellular automata for simulating urban expansion by using the deep learning techniques . **International Journal of Geographical Information Science**, 2018.

HILFERINK, M.; RIETVELD, P. LAND USE SCANNER: An integrated GIS based model for long term projections of land use in urban and rural areas. **Journal of Geographical Systems**, v. 1, n. 2, p. 155–177, 1999.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, n. 3, p. 295–309, 1988.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2022: população e domicílios : primeiros resultados**. Rio de Janeiro: IBGE, 2023. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=21020>
11

IBGE - Instituto Brasileiro de Geografia e Estatística. **Pesquisa Nacional por Amostras de Domicílios (PNAD) 2015**. Rio de Janeiro: IBGE, 2015. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=759>

IBGE - Instituto Brasileiro de Geografia e Estatística. **Proposta metodológica para classificação dos espaços do rural, do urbano e da Natureza no Brasil**. Rio de Janeiro: IBGE, 2023. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102019>

KAMUSOKO, C.; GAMBA, J. Simulating Urban Growth Using a Random Forest Cellular Automata (RF-CA) Model. **ISPRS Int. J. Geo-Information**, 2015.

KARIMI, F.; SULTANA, S.; SHIRZADI, A. B. An enhanced support vector machine model for urban expansion prediction. **Computers, Environment and Urban Systems**, v. 75, p. 61–75, 2019.

KULKARNI, A.; CHONG, D.; BATARSEH, F. A., Foundations of data imbalance and solutions for a data democracy. **Data Democracy**, Academic Press, 2020

KUMAR, V.; AGRAWAL, S. Urban modelling and forecasting of landuse using SLEUTH model. **International Journal of Environmental Science and Technology**, v. 20, n. 6, p. 6499–6518, 2023.

KUMAR, A.; TRIPATHY P., Monitoring and modelling spatio-temporal urban growth of Delhi using Cellular Automata and geoinformatics, **Cities**, Volume 90, 2019.

LI, X.; ZHOU, Y; HEJAZI, M.; WISE, M.; VERNON, C.; IYER, G.; CHEN, C. Global urban growth between 1870 and 2100 from integrated high resolution mapped data and urban dynamic modeling, **Nature, Communications Earth & Environment** 2, 2021.

LIMA, C. de A. **A OCUPAÇÃO DE ÁREA DE MANANCIAIS NA REGIÃO METROPOLITANA DE CURITIBA: DO PLANEJAMENTO À GESTÃO AMBIENTAL URBANA - METROPOLITANA**. 2000. 406 f. Tese (Doutorado em Meio Ambiente e Desenvolvimento) - Curitiba, Universidade Federal do Paraná, 2000.

LINARES, S.; PICONE, N. Application of Remote Sensing and Cellular Automata Model to Analyse and Simulate Urban Density Changes. **Urban Remote Sensing, 2nd edition**. CRC Press, 2018.

LIU, X.; WEI, M.; LI, Z. Multi-scenario simulation of urban growth boundaries with an ESP-FLUS model: A case study of the Min Delta region, China. **Ecological Indicators**, v. 135, p. 108538, 2022.

LV, J.; WANG, Y.; LIANG, X. Simulating urban expansion by incorporating an integrated gravitational field model into a demand-driven random forest-cellular automata model. **Cities**, v. 109, p. 103044, 2021.

MARICATO, E. Conhecer para resolver a cidade ilegal. **Urbanização brasileira: redescobertas**, Belo Horizonte, 2003.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, v. 17, n. 7, p. 1425–1432, 1996.

MEIRELLES, M. S. P.; CAMARA, G.; ALMEIDA, C. M., **GEOMÁTICA MODELOS E APLICAÇÕES AMBIENTAIS**. Brasília, 2007

MESTA, C.; CREMEN, G.; GALASSO, C., Urban growth modelling and social vulnerability assessment for a hazardous Kathmandu Valley, **Nature, Scientific Reports** **12**, 2022.

NUNES, F. P. G. **Diretrizes de ocupação de baixo impacto ambiental para áreas de mananciais com aplicação do LiderA: o caso do Guarituba - Município Piraquara - Paraná**. Dissertação (Mestrado em Engenharia da Construção Civil) - Curitiba, Universidade Federal do Paraná, 2014.

ONU - Organização das Nações Unidas. **The World's Cities in 2018**: Data Booklet.

OpenStreetMap Wiki. **Proposta de Classificação de Vias Urbanas**. Disponível em: https://wiki.openstreetmap.org/wiki/Proposta_de_Classifica%C3%A7%C3%A3o_de_Vias_Urbanas

OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. Dissertação (Mestrado em Bioinformática) - Ribeirão Preto, Universidade de São Paulo, 2013.

PAN, X.; LIU, Z.; HE, C.; HUANG, Q. Modeling urban expansion by integrating a convolutional neural network and a recurrent neural network. **International journal of applied earth observation and geoinformation**, 2022

PARANÁ. **Decreto Nº 1.753, de 06 de Maio de 1996**. Instituída a Área de Proteção Ambiental na área de manancial da bacia hidrográfica do rio Iraí, denominada APA Estadual do Iraí. Legislação do Estado do Paraná, Curitiba, 06 de maio de 1996.

PESAVENTO, U. An Implementation of von Neumann's Self-Reproducing Machine. **Artificial Life** **2**, p. 337–354, 1995.

PESCHL, H. **Modelagem do crescimento urbano com autômatos celulares: estudo de caso para Curitiba e região**. Trabalho de Graduação (Bacharelado em Engenharia Cartográfica e de Agrimensura) - Curitiba, Universidade Federal do Paraná, Curitiba, 2021.

PIJANOWSKI, B.; GAGE, S.; LONG, D.; COOPER, W. C. A land transformation model: Integrating policy, socioeconomics and environmental drivers using a geographic information system. **Landscape Ecology: A Top Down Approach**, p. 183–198, 2000.

PONTIUS, R. G.; CORNELL, J. D.; HALL, C. A. S. Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. **Agriculture, Ecosystems & Environment**, v. 85, n. 1, p. 191–203, 2001.

PONTIUS, R. G.; BOERSMA, W.; CASTELLA, J. C. Comparing the input, output, and validation maps for several models of land change. **The Annals of Regional Science**, v. 42, n. 1, p. 11–37, 2008.

PONTIUS, R. G., PARMENTIER, B. **Recommendations for using the relative operating characteristic (ROC)**. *Landscape Ecol* 29, 367–382 (2014). <https://doi.org/10.1007/s10980-013-9984-8>

QIAN, Y.; XING, W.; GUAN, X.; YANG, T.; WU, H. Coupling cellular automata with area partitioning and spatiotemporal convolution for dynamic land use change simulation. 1 Jun. 2020. **Science of The Total Environment**. DOI 10.1016/j.scitotenv.2020.137738.

RIENOW, A. MUSTAFA, A.; KRELAUS, L.; LINDNER, C. Modeling urban regions: Comparing random forest and support vector machines for cellular automata. **Transactions in GIS**, v. 25, n. 3, p. 1625–1645, 2021.

RODRIGUES, H. O.; SOARES-FILHO, B. S.; COSTA, W. L. S. **Dinâmica EGO, uma plataforma para modelagem de sistemas ambientais**. 2007

SANEPAR, **Plano Diretor SAIC - Sistema de Abastecimento de Água Integrado de Curitiba e Região Metropolitana**. Curitiba, 2013.

SANTOS, R. C. **História - Prefeitura Municipal de Piraquara - PR**. Disponível em: <<https://www.piraquara.pr.gov.br/a-cidade/historia>>.

SHAFIZADEH-MOGHADAM, H; MINAEI, M.; PONTIUS, R. G. Integrating a Forward Feature Selection algorithm, Random Forest, and Cellular Automata to extrapolate urban growth in the Tehran-Karaj Region of Iran. **Computers, Environment and Urban Systems**, v. 87, p. 101595, 2021.

SHEYKHMOUSA, R. M.; MAHDIANPARI, M. Support Vector Machine vs. Random Forest for Remote Sensing Image Classification: A Meta-analysis and systematic review. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, 2020.

SILVA, M. N. da. **A dinâmica da produção dos espaços informais de moradia e processo de metropolização em Curitiba**. Tese (Doutorado em Geografia) - Curitiba, Universidade Federal do Paraná, 2012.

SOLVED. **Avaliação e Métricas de Acurácia**. 2022. Disponível em: <<https://solved.eco.br/avaliacao-de-acuracia-ou-concordancia/>>.

UN-Habitat. **World Cities Report 2020: The Value of Sustainable Urbanization** Disponível em: <<https://unhabitat.org/world-cities-report-2020-the-value-of-sustainable-urbanization>>

VERBURG, P. H.; SOEPBOER, W.; VELDKAMP, A.; LIMPIADA, R.; ESPALDON, V.; MASTURA, S. S. A. Modeling the Spatial Dynamics of Regional Land Use: The CLUE-S Model. **Environmental Management**, v. 30, n. 3, p. 391–405, 2002.

VIEGAS, V.; MADUREIRA, C.; SOUZA, E. Autômatos celulares no contexto da modelagem dinâmica: desafios da modelagem de espaços urbanos. **Geography Department University of Sao Paulo**, v. 41, p. e181171, 2021.

ZHA, Y.; GAO, J.; NI, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. **International Journal of Remote Sensing**, v. 24, n. 3, p. 583–594, 2003.

ZHOU, Y.; VARQUEZ, A. C. G.; KANDA, M. High-resolution global urban growth projection based on multiple applications of the SLEUTH urban growth model. **Scientific Data**, v. 6, n. 1, p. 34, 2019.