

Deteção de outliers derivados do processamento de séries temporais InSAR usando algoritmos de Deep Learning

Pedro Tiago Lima Aguiar

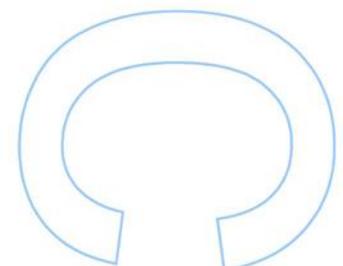
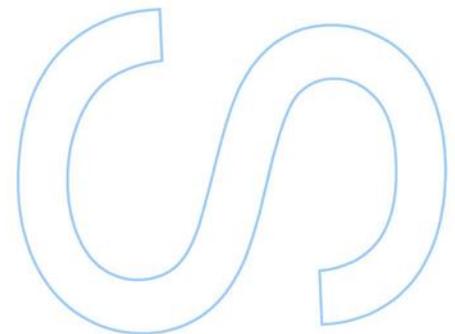
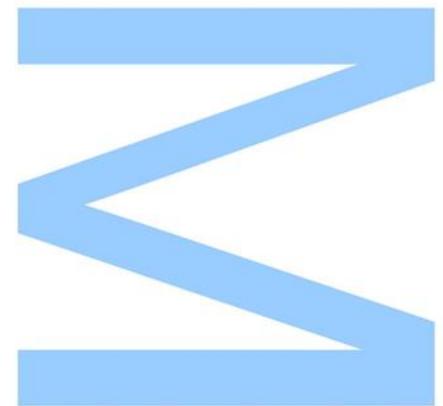
Mestrado em Engenharia Geográfica
Departamento de Geociências, Ambiente e Ordenamento do Território
Faculdade de Ciências da Universidade do Porto
2020

Orientadora

Professora Doutora Maria Clara Gomes Quadros Lázaro da Silva, Professora
Auxiliar, Faculdade de Ciências da Universidade do Porto

Coorientador

Professor Doutor António Manuel Trigueiros da Silva Cunha, Professor Auxiliar,
Universidade de Trás-os-Montes e Alto Douro





Todas as correções determinadas pelo júri, e só essas, foram efetuadas

O Presidente do Júri,

Porto, 6 / 7 / 2020

Resumo

A Interferometria Diferencial SAR (DInSAR) é atualmente uma técnica bem estabelecida e fundamentada, a qual permite uma monitorização contínua e global de subsidência de terrenos e atividades/estruturas naturais e antropogénicas da superfície terrestre com um rigor centimétrico. Para tal, são usadas séries temporais de dados SAR (*Synthetic Aperture Radar*, Radar de Abertura Sintética), sendo possível a obtenção de milhares de “medidas”, designadas por dispersores permanentes (*Permanent Scatterers*, PSs), por km². Existem, no entanto, vários fatores indutores de erro que resultam na geração, em parte, de PSs *outliers*, sendo necessária a posterior identificação e remoção dos mesmos. Devido ao facto do número de PSs ser frequentemente na ordem dos milhares, a correspondente deteção de *outliers* torna-se computacionalmente exigente em termos de complexidade e tempo de execução. Assim, torna-se crucial o desenvolvimento de métodos que possibilitem a identificação de PSs outliers de uma forma mais célere e exata.

A presente dissertação foi realizada no âmbito de uma bolsa de investigação, com a referência BIL/UTAD/35/2018, do projeto/instituição de I&D PC 0073 – Projeto ESA (PC 0073), financiada pela Agência Espacial Europeia (ESA), na Universidade de Trás-os-Montes e Alto Douro (UTAD). Este projeto teve como objetivo aplicar algoritmos de Machine Learning (ML) a séries temporais InSAR, com vista a determinar o potencial destes algoritmos. A presente dissertação, em particular, teve como objetivo avaliar a aplicabilidade de algoritmos de Deep Learning (DL) de classificação supervisionada na identificação de PSs *outliers*, derivados do processamento de séries temporais de dados SAR. A abordagem apresentada consiste em considerar para cada PS os correspondentes parâmetros multi-temporais, como a velocidade de subsidência/elevação, a coerência temporal, entre outros, e os correspondentes parâmetros dos PSs vizinhos mais próximos, bem como as suas posições relativas ao PS a classificar.

Com base nos resultados obtidos em duas bases de dados de PSs de regiões distintas, relativas às regiões da cidade de Bratislava e os subúrbios de Prievidza, Eslováquia, foi possível concluir que tais algoritmos de classificação oferecem uma abordagem robusta e com baixa exigência computacional. Verificou-se, porém, que a correta aplicabilidade de tais algoritmos exige o conhecimento prévio de diversos fatores, tais como o padrão de deformação da região a classificar, a respetiva densidade de PSs, entre outros.

Palavras-chave: InSAR, Deep Learning, Dispersores Permanentes, Deteção de *outliers*

Abstract

Differential SAR Interferometry (DInSAR) is nowadays a successful and well-established technique, which allows continuous and global monitoring of terrain subsidence, as well as natural and anthropogenic activities/structures in the Earth's surface with centimeter precision. To this end, time series of SAR data (Synthetic Aperture Radar) are used, making it possible to obtain thousands of "measures", designated by permanent scatterers (PSs), per km². There are, however, several error-inducing factors that result in the generation, in part, of outlier PSs, requiring their subsequent identification and removal. Since the number of PSs is often in the order of thousands, the corresponding outlier's identification becomes computationally demanding in terms of complexity and execution time. Thus, it is crucial to develop methods that make it possible to identify outlier PSs in a faster and more accurate way.

The present work was carried out as part of a research grant, under the reference BIL/UTAD/35/2018, of the I&D project/institution PC 0073 - ESA Project (PC 0073), financed by the European Space Agency (ESA), at University of Trás-os-Montes e Alto Douro (UTAD). This project aimed to apply Machine Learning (ML) algorithms to InSAR time series, in order to determine the potential of these algorithms. The present work, in particular, aims to evaluate the applicability of supervised Deep Learning (DL) classification algorithms in the identification of outlier PSs, derived from multi-temporal SAR data processing. The approach presented consists in considering for each PS the corresponding multi-temporal parameters, such as subsidence/uplift velocity, interferometric coherence, among others, and the corresponding parameters of the nearest neighboring PSs, as well as their positions relative to the PS to classify.

Based on the obtained results in two PSs datasets from different regions, relative to the regions of Bratislava city and the suburbs of Prievidza, Slovakia, it was possible to conclude that such classification algorithms offer a robust approach with low computational requirements. However, the results showed that the correct applicability of such algorithms requires prior knowledge of several factors, such as the deformation pattern of the region to be classified, the respective density of PSs, among others.

Keywords: InSAR, Deep Learning, Permanent Scatterers, Outlier Detection

Índice

Resumo.....	iii
Abstract.....	iv
Lista de Tabelas.....	vii
Lista de Figuras.....	xii
Lista de Abreviaturas.....	xvii
1. Introdução.....	1
2. Enquadramento Teórico.....	3
2.1. Radar de Abertura Sintética.....	3
Geometria de operação dos sistemas SAR.....	4
Efeito de <i>speckle</i>	11
Processamento de dados SAR.....	12
2.2. Interferometria SAR.....	14
Geração de Modelos Digitais do Terreno por interferometria <i>across-track</i>	15
Interferometria Diferencial.....	18
2.3. Machine Learning e Deep Learning.....	21
Machine Learning.....	21
Deep Learning e Redes Neurais.....	23
Técnicas de regularização e generalização de Redes Neurais.....	27
3. Estado da arte.....	29
3.1. Estudos de deformação superficial.....	29
3.2. Classificação do uso do solo com base na coerência temporal.....	30
3.3. Filtragem de ruído em Interferogramas SAR.....	31
3.4. Outras aplicações.....	31
4. Metodologias.....	33
4.1. Preparação dos dados.....	34
4.2. Criação, treino e avaliação dos modelos de classificação.....	37

5. Resultados e Discussão.....	39
5.1. Resultados relativos à base de dados de dispersores permanentes da cidade de Bratislava	39
5.2. Resultados relativos à base de dados de dispersores permanentes dos subúrbios de Prievidza.....	48
6. Conclusão.....	53
7. Referências	55
8. Anexo	63
8.1. Resultados relativos à base de dados de dispersores permanentes da cidade de Bratislava	63
8.2. Resultados relativos à base de dados de dispersores permanentes dos subúrbios de Prievidza.....	85

Lista de Tabelas

Tabela 1. Bandas de frequência mais comuns nos sistemas SAR e correspondentes valores de frequência (f) e comprimento de onda (λ).....	4
Tabela 2. Principais sistemas SAR, operados a partir de satélite, desenvolvidos até à atualidade.	9
Tabela 3. Semi-eixo maior (a) e inverso do achatamento (f^{-1}) do elipsoide WGS84.	35
Tabela 4. Número de PSs e a respetiva percentagem dos conjuntos de PSs de Treino, Validação e Teste resultantes da divisão da base de dados de PSs da cidade de Bratislava.	42
Tabela 5. Resultados médios dos modelos de classificação, em função do número de PSs vizinhos considerados na classificação de cada PS, relativos à base de dados de PSs da cidade de Bratislava, em comparação com os resultados dos modelos de base.	43
Tabela 6. Percentagem média de classificações incertas por classe, em função do número de PSs vizinhos considerados na classificação de cada PS, relativa à base de dados de PSs da cidade de Bratislava.	43
Tabela 7. Desvio padrão médio, em metros, da distância aos n PSs vizinhos mais próximos considerados, para os três tipos de classificações, relativo à base de dados de PSs da cidade de Bratislava.....	44
Tabela 8. Distância média ao n -ésimo PS vizinho mais próximo para os três tipos de classificações, em função do número de PSs vizinhos considerados na classificação de cada PS, relativa à base de dados de PSs da cidade de Bratislava.....	44
Tabela 9. Resultados médios dos modelos de classificação, em função do número de PSs vizinhos considerados na classificação de cada PS, relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, em comparação com os resultados dos modelos de base.	50
Tabela 10. Percentagem média de classificações incertas por classe, em função do número de PSs vizinhos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza. ...	51
Tabela 11. Desvio padrão médio, em metros, da distância aos n PSs vizinhos mais próximos considerados, para os três tipos de classificações, relativo à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza.....	51
Tabela 12. Distância média ao n -ésimo PS vizinho mais próximo, para os três tipos de classificações, em função do número de PSs vizinhos considerados na classificação de	

cada PS, relativa à base de dados da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza.....51

Tabela 13. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava.73

Tabela 14. Matriz de confusão relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos.73

Tabela 15. Matriz de confusão relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos.73

Tabela 16. Matriz de confusão relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos.73

Tabela 17. Matriz de confusão relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos.73

Tabela 18. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava.76

Tabela 19. Matriz de confusão relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos.76

Tabela 20. Matriz de confusão relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos.76

Tabela 21. Matriz de confusão relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos.76

Tabela 22. Matriz de confusão relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos.76

Tabela 23. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava.79

Tabela 24. Matriz de confusão relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos.79

Tabela 25. Matriz de confusão relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos.79

Tabela 26. Matriz de confusão relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos.79

Tabela 27. Matriz de confusão relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos.79

Tabela 28. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava.82

Tabela 29. Matriz de confusão relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos.82

Tabela 30. Matriz de confusão relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos.82

Tabela 31. Matriz de confusão relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos.82

Tabela 32. Matriz de confusão relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos.82

Tabela 33. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de

Prievidza, obtida usando os correspondentes modelos de classificação treinados na divisão de dados 1 da base de dados de PSs da cidade de Bratislava.85

Tabela 34. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos, relativa à divisão de dados 1.85

Tabela 35. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos, relativa à divisão de dados 1.85

Tabela 36. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos, relativa à divisão de dados 1.86

Tabela 37. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos, relativa à divisão de dados 1.86

Tabela 38. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtida usando os correspondentes modelos de classificação treinados na divisão de dados 2 da base de dados de PSs da cidade de Bratislava.88

Tabela 39. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos, relativa à divisão de dados 2.88

Tabela 40. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos, relativa à divisão de dados 2.88

Tabela 41. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos, relativa à divisão de dados 2.89

Tabela 42. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos, relativa à divisão de dados 2.89

Tabela 43. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de

Prievidza, obtida usando os correspondentes modelos de classificação treinados na divisão de dados 3 da base de dados de PSs da cidade de Bratislava.91

Tabela 44. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos, relativa à divisão de dados 3.91

Tabela 45. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos, relativa à divisão de dados 3.91

Tabela 46. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos, relativa à divisão de dados 3.92

Tabela 47. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos, relativa à divisão de dados 3.92

Tabela 48. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtida usando os correspondentes modelos de classificação treinados na divisão de dados 4 da base de dados de PSs da cidade de Bratislava.94

Tabela 49. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos, relativa à divisão de dados 4.94

Tabela 50. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos, relativa à divisão de dados 4.94

Tabela 51. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos, relativa à divisão de dados 4.95

Tabela 52. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos, relativa à divisão de dados 4.95

Lista de Figuras

Figura 1. Geometria de operação de um sistema SAR.....	5
Figura 2. Imagem SAR da cidade de Milão, Itália, obtida pelo satélite ERS-1 da ESA.....	6
Figura 3. Imagem SAR do Monte Vesúvio (Itália) ilustrativa do efeito de <i>foreshortening</i> , obtida pelo satélite ERS-2 da ESA.	7
Figura 4. Figura ilustrativa do modo de operação do sensor SRTM.	10
Figura 5. Imagem SAR do satélite ERS-2 da ESA do Aeroporto de Linate, na região Este de Milão, Itália. O ruído ‘sal-e-pimenta’ nas regiões circundantes do aeroporto é devido ao efeito de <i>speckle</i>	11
Figura 6. Esquematização do processo de convolução do sinal reamostrado em alcance e azimute com as respetivas funções de referência.	12
Figura 7. Geometria de operação da técnica InSAR.....	15
Figura 8. Interferogramas SAR e MDT do deserto do Atacama (Chile) gerados a partir de dados do sensor TanDEM-X: (a) antes da remoção do efeito de flat-earth; (b) depois da remoção do efeito de flat-earth; (c) MDT gerado após o processo de desenrolamento de fase e correspondente conversão em valores de altitude.....	16
Figura 9. Arquitetura habitual de uma RN.	23
Figura 10. Gráfico da função de ativação <i>relu</i>	25
Figura 11. Gráfico da função de ativação <i>sigmoid</i>	26
Figura 12. Figuras ilustrativas da abordagem seguida, considerando na classificação de cada PS os 3, 5, 10, e 15 PSs vizinhos mais próximos.	34
Figura 13. Esquematização das etapas envolvidas na preparação dos dados.....	34
Figura 14. Modelos de classificação implementados, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS.....	37
Figura 15. Mapa de velocidade de subsidência/elevação da base de dados de PSs da cidade de Bratislava.....	39
Figura 16. Mapa da coerência temporal da base de dados de PSs da cidade de Bratislava.	40
Figura 17. Mapa de classes da base de dados de PSs da cidade de Bratislava.	40
Figura 18. Abordagem de validação cruzada aplicada na base de dados de PSs da cidade de Bratislava.	41
Figura 19. Percentagem média de classificações incertas e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada	

PS, considerando os 5 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.....46

Figura 20. Exatidão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 5 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.46

Figura 21. Precisão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 5 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.47

Figura 22. Recall médio e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 5 PSs vizinhos mais próximos, relativo à base de dados de PSs da cidade de Bratislava.....47

Figura 23. Mapa de velocidade de subsidência/elevação da base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza.....48

Figura 24. Mapa da coerência temporal da base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza.49

Figura 25. Mapa de classes da base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza.49

Figura 26. Número de PSs outliers classificados como PSs inliers, e vice-versa, em função do valor de threshold aplicado à saída dos modelos de classificação relativos à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando para cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos.63

Figura 27. Número de PSs outliers classificados como PSs inliers, e vice-versa, em função do valor de threshold aplicado à saída dos modelos de classificação relativos à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando para cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos64

Figura 28. Número de PSs outliers classificados como PSs inliers, e vice-versa, em função do valor de threshold aplicado à saída dos modelos de classificação relativos à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando para cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos.65

Figura 29. Número de PSs outliers classificados como PSs inliers, e vice-versa, em função do valor de threshold aplicado à saída dos modelos de classificação relativos à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando para cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos.66

Figura 30. Percentagem média de classificações incertas e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 3 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.....67

Figura 31. Exatidão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 3 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.67

Figura 32. Precisão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 3 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.68

Figura 33. Recall médio e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 3 PSs vizinhos mais próximos, relativo à base de dados de PSs da cidade de Bratislava.....68

Figura 34. Percentagem média de classificações incertas e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 10 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.....69

Figura 35. Exatidão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 10 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.69

Figura 36. Precisão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 10 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.70

Figura 37. Recall médio e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 10 PSs vizinhos mais próximos, relativo à base de dados de PSs da cidade de Bratislava.....70

Figura 38. Percentagem média de classificações incertas e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 15 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.....71

Figura 39. Exatidão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 15 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.71

Figura 40. Precisão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 15 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.72

Figura 41. Recall médio e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 15 PSs vizinhos mais próximos, relativo à base de dados de PSs da cidade de Bratislava.....72

Figura 42. Mapas de classificação relativos à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos.....74

Figura 43. Segunda versão dos mapas de classificação ilustrados na Figura 42, demonstrando apenas classificações erradas e classificações incertas.75

Figura 44. Mapas de classificação relativos à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos.....77

Figura 45. Segunda versão dos mapas de classificação ilustrados na Figura 44, demonstrando apenas classificações erradas e classificações incertas78

Figura 46. Mapas de classificação relativos à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos.....80

Figura 47. Segunda versão dos mapas de classificação ilustrados na Figura 46, demonstrando apenas classificações erradas e classificações incertas.81

Figura 48. Mapas de classificação relativos à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos.....83

Figura 49. Segunda versão dos mapas de classificação ilustrados na Figura 48, demonstrando apenas classificações erradas e classificações incertas.84

Figura 50. Mapas de classificação relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtidos usando os modelos de classificação treinados na divisão de dados 1 da base de dados de PSs da cidade de Bratislava.....86

Figura 51. Segunda versão dos mapas de classificação ilustrados na Figura 50, demonstrando apenas classificações erradas e classificações incertas.87

Figura 52. Mapas de classificação relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtidos usando os modelos

de classificação treinados na divisão de dados 2 da base de dados de PSs da cidade de Bratislava.....89

Figura 53. Segunda versão dos mapas de classificação ilustrados na Figura 52, demonstrando apenas classificações erradas e classificações incertas.90

Figura 54. Mapas de classificação relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtidos usando os modelos de classificação treinados na divisão de dados 3 da base de dados de PSs da cidade de Bratislava.....92

Figura 55. Segunda versão dos mapas de classificação ilustrados na Figura 54, demonstrando apenas classificações erradas e classificações incertas.93

Figura 56. Mapas de classificação relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtidos usando os modelos de classificação treinados na divisão de dados 4 da base de dados de PSs da cidade de Bratislava.....95

Figura 57. Segunda versão dos mapas de classificação ilustrados na Figura 56, demonstrando apenas classificações erradas e classificações incertas.96

Lista de Abreviaturas

- CNNs – Redes Neurais de Convolução (*Convolutional Neural Networks*)
- CRs – Refletores Artificiais (*Artificial Corner Reflectors*)
- DInSAR – Interferometria Diferencial SAR (*Differential SAR Interferometry*)
- DL – *Deep Learning*
- DR – Deteção Remota
- DSs – Dispersores Dispersos (*Distributed Scatterers*)
- DTs – Árvores de Decisão (*Decision Trees*)
- EM – Espetro Eletromagnético
- ESA – Agência Espacial Europeia (*European Space Agency*)
- GBs – *Gradient Boosting Machines*
- IA – Inteligência Artificial
- InSAR – Interferometria SAR
- logreg* – Regressão Logística (*Logistic Regression*)
- LOS – Linha de visão (*Line-of-Sight*)
- LR – *Learning Rate*
- MDTs – Modelos Digitais do Terreno
- ML – *Machine Learning*
- NASA – Agência Espacial Norte-Americana (*National Aeronautics and Space Administration*)
- PRF – Frequência de emissão de pulsos (*Pulse Repetition Frequency*)
- PRI – Intervalo de repetição de pulsos (*Pulse Repetition Interval*)
- PSs – Dispersores Permanentes (*Permanent Scatterers*)
- RF – *Random Forest*
- RNs – Redes Neurais
- SAR – Radar de Abertura Sintética (*Synthetic Aperture Radar*)
- SBAS – *Small BAseline Subset*
- SGD – *Stochastic Gradient Descent*
- SLAR – *Side-Looking Airborne Radar*
- SRTM – *Shuttle Radar Topography Mission*
- SVM – *Support Vector Machine*
- TEC – *Total Electron Content*

1. Introdução

A Detecção Remota (DR) é uma área com um crescimento exponencial nos últimos quinze anos, verificando-se, em particular, um aumento significativo de missões integrando o Radar de Abertura Sintética (*Synthetic Aperture Radar*, SAR), cujo modo de operação permite a geração de imagens de alta resolução da superfície da Terra de forma independente de condições climáticas e de iluminação (Ferretti, Monti-Guarnieri, Prati, Rocca, & Massonnet, 2007; Moreira et al., 2013; Pepe & Calò, 2017).

Uma das técnicas SAR mais usadas atualmente é a Interferometria Diferencial SAR (DInSAR), na qual são usadas séries temporais de dados SAR para monitorização de subsidência de terrenos e atividades/estruturas naturais e antropogénicas com um rigor centimétrico (Ferretti, Prati, & Rocca, 2000, 2001). Esta técnica permite a obtenção de milhares de “medidas”, designadas por dispersores permanentes (*Permanent Scatterers*, PSs), por km², podendo as mesmas ser atualizadas num curto intervalo de tempo, permitindo, desta forma, a deteção de deformações do terreno numa fase inicial. No entanto, tal é apenas possível pela capacidade de processar e atualizar cada PS sempre que estão disponíveis novos dados SAR o que, devido ao frequente elevado número de PSs existentes, é computacionalmente exigente, implicando longos períodos de execução. Adicionalmente, tais medidas são, em parte, ambíguas devido à existência de PSs afetados por ruído de descorrelação, designados por *outliers*, sendo necessário a identificação e remoção dos mesmos *a posteriori*.

Uma prática comum para a identificação de PSs *outliers* consiste na avaliação da coerência temporal $\in [0,1]$ de cada PS, classificando como *outliers* os PSs que apresentam uma coerência inferior a um determinado valor pré-definido, geralmente 0.7. A aplicação da técnica anterior pode, no entanto, gerar a perda de informação em zonas caracterizadas por padrões de deformação mais complexos. Desta forma, torna-se necessário o desenvolvimento de metodologias alternativas para a identificação de PSs *outliers* e exploração de PSs com baixa coerência temporal, considerando outros parâmetros multi-temporais, tais como a velocidade de subsidência/elevação do terreno, deslocamento cumulativo, entre outros, e a correlação espacial existente entre PSs vizinhos próximos. Tal abordagem foi desenvolvida por Bakon et al. (2017), a qual permite uma melhor avaliação de PSs de baixa coerência, habitualmente descartados pelo método anterior, e uma melhor análise de áreas com padrões de deformação mais complexos, aumentando desta forma a densidade espacial de PSs. No entanto, no panorama atual de missões SAR, com elevada frequência de aquisição de dados de uma forma contínua e global, o elevado número de PSs a analisar originam longos períodos de execução das metodologias propostas pelos autores e uma complexidade considerável, sendo impraticável a sua aplicação frequente. Desta forma, torna-se necessária a definição de metodologias alternativas que, de uma forma semelhante às metodologias propostas pelos autores, efetuem a identificação de PSs *outliers* de uma forma mais célere e exata.

Nos últimos anos verificou-se um forte crescimento de popularidade da Inteligência Artificial (IA), em particular na aplicação de algoritmos de *Machine Learning* (ML) e *Deep Learning* (DL), subcampos da IA, maioritariamente em problemas de classificação e regressão. À data de realização da presente dissertação, no âmbito da DR as técnicas anteriores foram aplicadas essencialmente para classificação (Liu et al., 2019; W. Wang, Dou, Jiang, & Sun, 2018; Zeng, Chen, Chen, & Li, 2018; Zhu, Zhong, Liu, Zhang, & Li, 2018), segmentação (Arief, Strand, Tveite, & Indahl, 2018; Lyu et al., 2018; Sa et al., 2018; Wu et al., 2018; Xu, Wu, Xie, & Chen, 2018) e deteção/reconhecimento de objetos/estruturas naturais e antropogénicas (Chen, Zhan, & Zhang, 2018; H. Wang, Jiang, & Zhang, 2018; Yang et al., 2018; Zhang et al., 2018).

Especificamente na Interferometria SAR, a aplicação de algoritmos de ML e DL é atualmente de menor dimensão relativamente à DR em geral. Nos últimos dois anos surgiram, no entanto, alguns estudos neste âmbito para deteção de deformações vulcânicas em interferogramas SAR (Anantrasirichai, Biggs, Albino, & Bull, 2019^a, 2019^b; Anantrasirichai, Biggs, Albino, Hill, & Bull, 2018; Valade et al., 2019), deteção de deformações superficiais em zonas urbanas (Ma, Zhang, & Lin, 2019; Y. Wang et al., 2019), classificação do uso do solo com base na coerência temporal (Sica, Pulella, Nannini, Pinheiro, & Rizzoli, 2019; Vicente-Guijalba et al., 2018), filtragem de ruído em interferogramas SAR (Ichikawa & Hirose, 2017; Kottayil et al., 2018; Li, Xu, Gao, & Li, 2019), entre outras aplicações (Costante, Ciarfuglia, & Biondi, 2018; Kerkhof, Pankratius, Chang, Swol, & Hanssen, 2018; Rongier, Rude, Herring, & Pankratius, 2019; Shamshiri, Motagh, Nahavandchi, Haghshenas Haghghi, & Hoseini, 2019).

A presente dissertação foi realizada no âmbito de uma bolsa de investigação, com a referência BIL/UTAD/35/2018, do projeto/instituição de I&D PC 0073 – Projeto ESA (PC 0073), financiada pela Agência Espacial Europeia (ESA), na Universidade de Trás-os-Montes e Alto Douro (UTAD). Este projeto teve como objetivo aplicar algoritmos de ML a séries temporais InSAR, com vista a determinar o potencial destes algoritmos. A presente dissertação, em particular, tem por base o trabalho desenvolvido por Bakon et al. (2017), tendo como objetivo avaliar a aplicabilidade de algoritmos de DL de classificação supervisionada na identificação de PSs *outliers* derivados do processamento de séries temporais de dados SAR, de forma a automatizar e reduzir o tempo de processamento na deteção dos mesmos. Na abordagem apresentada foi considerada a distribuição espacial dos PSs e a correspondente correlação entre PSs vizinhos próximos, apresentando-se os resultados obtidos em duas bases de dados de PSs de regiões distintas relativamente à eficiência dos algoritmos e a respetiva comparação com modelos de base.

No Capítulo 2 encontra-se descrito um enquadramento teórico relativamente ao SAR, à Interferometria SAR, e ao ML e DL. O Capítulo 3 é referente à revisão do estado da arte. Nos Capítulos 4 e 5 encontram-se descritas, respetivamente, as metodologias e os resultados e consequente discussão. Por fim, no Capítulo 6 encontra-se apresentada a conclusão do trabalho.

2. Enquadramento Teórico

2.1. Radar de Abertura Sintética

O SAR é um sensor de imagem ativo de alta resolução que opera na região das micro-ondas do espectro eletromagnético (EM) (Ferretti et al., 2007; Moreira et al., 2013; Pepe & Calò, 2017). Pelo facto de se tratar de um sensor ativo, isto é, ser independente de fontes de iluminação externas, e de operar na região das micro-ondas, possui a capacidade de atuar em quaisquer condições climáticas e de iluminação. A sua capacidade de monitorização de processos dinâmicos na superfície da Terra de uma forma contínua e global levou a um forte crescimento de popularidade do mesmo nos últimos quinze anos, havendo atualmente inúmeros estudos de propriedades geofísicas da superfície da Terra baseados na análise de dados SAR (Alpers, Ross, & Rufenach, 1981; Bürgmann, Rosen, & Fielding, 2000; Henderson & Zong-Guo, 1997; Moran, Vidal, Troufleau, Inoue, & Mitchell, 1998; Qi, Yeh, Li, & Lin, 2012; Tomiyasu, 1978).

O SAR opera acoplado a uma plataforma em movimento, geralmente um satélite, movendo-se à velocidade deste último ao longo da sua órbita. Tal como outros sistemas radar, o seu modo de funcionamento consiste na emissão de um pulso EM e na receção do sinal retro-difundido (o eco), sendo a frequência de emissão de pulsos (*Pulse Repetition Frequency*, PRF) da ordem de 10^3 Hertz. O pulso transmitido interage com a superfície da Terra, e apenas uma porção do mesmo é refletido na direção da antena do sensor, que pode ser a mesma que a antena de transmissão, no caso de radar mono-estático, ou uma antena diferente, no caso de radar bi- ou multi-estático. A sequência de transmissão e receção de pulsos é repetida a cada PRI segundos (Intervalo de Repetição de Pulsos, *Pulse Repetition Interval*), sendo este último o recíproco da frequência de emissão, dado pela equação (1):

$$PRI = \frac{1}{PRF} \quad (1)$$

Relativamente à emissão e receção do pulso EM, o mesmo pode ser emitido e/ou recebido com polarização vertical e/ou horizontal, dando origem a quatro casos possíveis:

- emissão e receção do pulso com polarização vertical (VV);
- emissão do pulso com polarização vertical e receção com polarização horizontal (VH);
- emissão e receção do pulso com polarização horizontal (HH);
- emissão do pulso com polarização horizontal e receção com polarização vertical (HV).

No caso do SAR, o intervalo de tempo entre a transmissão e recepção do sinal, proporcional à distância entre o sensor e a área à superfície observada, traduz-se em diferentes posições do primeiro devido ao facto de o mesmo se encontrar numa plataforma em movimento. Uma combinação coerente dos sinais recebidos permite a construção de uma abertura virtual, com um comprimento muito maior que o comprimento real da antena do sensor. Este simples atributo deu origem ao seu nome, “Abertura Sintética”, conferindo-lhe a propriedade de ser um radar de imagem.

Adicionalmente ao mencionado anteriormente, dispersores à superfície a diferentes distâncias do sensor introduzem diferentes desfasamentos entre a transmissão e recepção do sinal. Devido à natureza sinusoidal do pulso EM transmitido, este desfasamento temporal τ é equivalente a uma diferença de fase $\Delta\phi$, dada pela equação (2):

$$\Delta\phi = \frac{4\pi}{\lambda} * R \tag{2}$$

onde R é a distância entre o sensor e a superfície e λ o comprimento de onda do sensor. No entanto, valores de R que diferem entre si de um múltiplo do comprimento de onda do sensor introduzem a mesma diferença de fase $\Delta\phi$. Desta forma, a fase do sinal SAR fornece apenas uma medida da fração de distância percorrida pela radiação inferior ao comprimento de onda do sensor.

As bandas de frequência mais comuns nos sistemas SAR são a banda L, C e X. Os valores de frequência e comprimento de onda para cada banda encontram-se na Tabela 1.

Banda	L	C	X
f (GHz)	1 – 2	3.75 – 7.5	7.5 – 12
λ (cm)	30 – 15	8 – 4	4 – 2.5

Tabela 1. Bandas de frequência mais comuns nos sistemas SAR e correspondentes valores de frequência (f) e comprimento de onda (λ).

Dependendo da frequência de banda do sistema, pode ocorrer uma penetração considerável do sinal emitido nos objetos observados. Geralmente este efeito é mais considerável em sistemas SAR a operar em frequências de banda menores, como a banda L. A amplitude e fase do sinal retro-difundido dependem das propriedades físicas e elétricas da região/objeto observada(o).

Geometria de operação dos sistemas SAR

O SAR opera segundo uma geometria lateral, ilustrada na Figura 1. O sensor, a bordo de um satélite, desloca-se ao longo da órbita deste último, com a antena orientada para a superfície da Terra no plano perpendicular à órbita. No caso do SAR, a antena não tem uma orientação nadiral, como é o caso, por exemplo, do radar altímetro, na qual o sensor aponta diretamente para a

superfície da Terra na direção perpendicular à órbita do satélite. Alternativamente, a antena do sensor SAR encontra-se ligeiramente inclinada para a direita, sendo a inclinação da mesma relativamente ao nadir designada por ângulo *off-nadir*, tomando valores entre os 20° e os 50° em sistema SAR modernos (Ferretti et al., 2007). Devido à curvatura da Terra, tipicamente o ângulo de incidência da radiação é maior que o ângulo *off-nadir*. A direção de voo é designada por direção azimutal, ou azimute, e a direção ao longo da linha de visão (*line-of-sight*, LOS) é designada por direção de alcance inclinado.

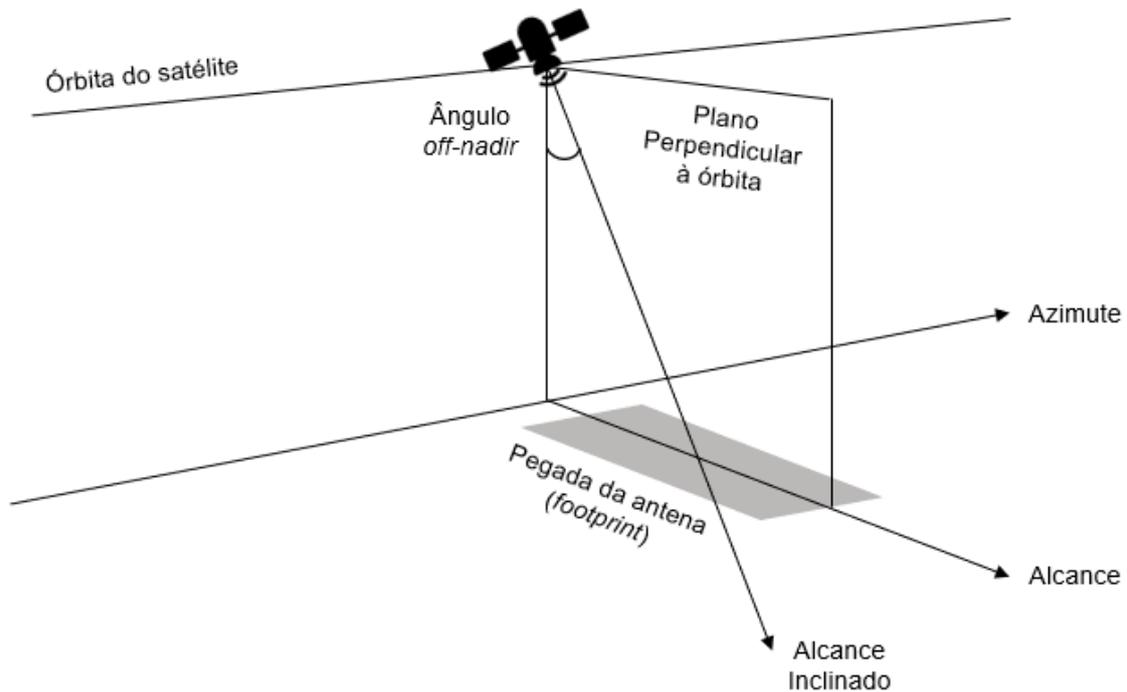


Figura 1. Geometria de operação de um sistema SAR (figura adaptada de Ferretti et al., 2007).

No seu modo de operação fundamental, designado por modo *Stripmap*, o sensor cobre em cada instante uma área aproximadamente retangular da superfície da Terra, designada por pegada da antena (*antenna footprint*). O comprimento e largura da pegada variam de sensor para sensor. O primeiro define o alcance à superfície coberto pelo radar, variando habitualmente entre os 30 km e os 500 km; a largura da pegada, por outro lado, é dependente do intervalo de tempo em que o sensor recolhe os sinais retro-difundidos à superfície.

Sendo um sensor de imagem, o SAR possui uma resolução bidimensional. A resolução em alcance, σ_r , é inversamente proporcional à largura de banda do sistema, B , através da equação (3):

$$\sigma_r = \frac{c}{2B} \quad (3)$$

onde c é a velocidade da luz. A resolução azimutal, σ_a , dada pela separação mínima entre dois objetos à superfície que o sistema é capaz de detetar, é obtida pela construção da abertura sintética, definida pela distância percorrida pelo sensor na qual o mesmo recolhe os sinais retro-difundidos à superfície. Considerando uma antena de tamanho d , a resolução azimutal é dada pela equação (4), sendo independente da altitude do sensor:

$$\sigma_a = \frac{d}{2} \quad (4)$$

De uma forma simples, os sistemas SAR fornecem imagens de alta resolução representativas da refletividade da área à superfície observada. Cada pixel da imagem contém informação da refletividade da correspondente região à superfície, designada por célula de resolução, e fornece um número complexo com informação da amplitude e fase da radiação EM retro-difundida à superfície na correspondente célula de resolução. No caso da componente de intensidade, as imagens SAR obtidas são geralmente visualizadas através de níveis de tons de cinzento. Pixels com cores mais claras estão associados a zonas à superfície com forte refletividade, como é o caso de zonas urbanas; por outro lado, pixels com cores mais escuras estão associados a regiões com baixa refletividade, geralmente superfícies planas e lisas, como por exemplo rios e oceanos, nas quais a radiação EM é maioritariamente refletida na direção contrária à da antena do sensor.

Na Figura 2 encontra-se representada uma imagem SAR, obtida pelo satélite ERS-1 da Agência Espacial Europeia (ESA), da cidade de Milão, Itália. O tamanho da imagem é de aproximadamente 25 km em alcance e azimute (orientação vertical e horizontal, respetivamente).



Figura 2. Imagem SAR da cidade de Milão, Itália, obtida pelo satélite ERS-1 da ESA (imagem obtida de Ferretti et al., 2007).

A área coberta à superfície por um sistema SAR em cada célula de resolução depende da correspondente topografia. Tal deve-se ao facto de o sensor medir apenas a projeção do terreno num sistema bidimensional, definido pelas coordenadas alcance e azimute. Mais concretamente, é fortemente depende do declive do terreno nas direções em alcance (circunscrita no plano perpendicular à órbita) e azimute. Em particular, o tamanho da célula de resolução em alcance relaciona-se com o declive do terreno através de uma perspectiva de deformação pouco convencional.

Com o aumento do declive do terreno em relação a uma superfície plana, o tamanho da célula de resolução em alcance aumenta, um efeito conhecido como *foreshortening*. O limite máximo é atingido com a proximidade do declive do terreno relativamente ao ângulo *off-nadir* do sistema, na qual a célula de resolução torna-se muito larga e todos os detalhes são perdidos. No caso em que o declive do terreno é maior que o ângulo *off-nadir* do sistema, os objetos à superfície são observados em ordem inversa de proximidade ao sistema SAR e somados às contribuições de outras células de resolução. Este efeito é conhecido como efeito de *layover*.

Por outro lado, com a diminuição do declive do terreno relativamente a uma superfície plana, o tamanho da célula de resolução em alcance diminui. O limite mínimo que pode ser observado por um sistema SAR é obtido quando o terreno é paralelo à LOS. Para declives inferiores à LOS, o terreno fica em oclusão.

O efeito de *foreshortening* tem um forte impacto na amplitude das imagens SAR. As áreas à superfície afetadas por este efeito são mais brilhantes na imagem SAR resultante, dado que a célula de resolução é maior, o que por sua vez resulta numa maior quantidade de radiação EM retro-difundida para o sensor. Este efeito encontra-se ilustrado na imagem SAR da Figura 3 do satélite ERS-2 da ESA, relativa ao Monte Vesúvio (Itália).

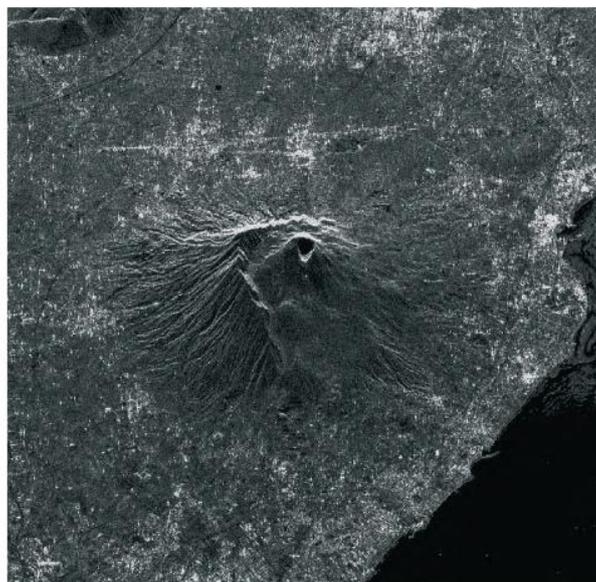


Figura 3. Imagem SAR do Monte Vesúvio (Itália) ilustrativa do efeito de *foreshortening*, obtida pelo satélite ERS-2 da ESA (imagem obtida de Ferretti et al., 2007).

A direção de alcance é vertical na imagem; a direção azimutal tem, desta forma, uma orientação horizontal e está localizada na parte superior da imagem. Regiões brilhantes correspondem a zonas urbanas. O declive positivo do vulcão na parte superior da imagem, mais brilhante, encontra-se reduzido relativamente à correspondente região voltada para o mar, menos brilhante (esta última corresponde à região escura na metade inferior direita da imagem).

Para além do modo *Stripmap*, os sensores SAR podem operar em outros modos, como é o caso dos modos *ScanSAR* e *Spotlight*. No modo *ScanSAR* a orientação da antena do sensor varia ciclicamente na direção em alcance, permitindo a obtenção de múltiplas sub-pegadas adjacentes. Cada sub-pegada é iluminada periodicamente, mas por um menor intervalo de tempo relativamente ao modo *Stripmap*. As imagens SAR resultantes deste modo de operação possuem, por conseguinte, maior comprimento na direção em alcance em detrimento de menor resolução azimutal.

No modo *Spotlight*, a orientação da antena do sensor varia na direção azimutal de forma a iluminar continuamente uma determinada região à superfície, obtendo-se desta forma uma maior resolução azimutal relativamente ao modo *Stripmap*. Contrariamente aos modos de operação *Stripmap* e *ScanSAR*, no modo *Spotlight* o sensor não recolhe uma pegada contínua ao longo da sua órbita, mas sim diferentes fragmentos de áreas iluminadas.

Alternativamente aos três modos de operação mencionados anteriormente, existem outros modos, cada um melhorando determinadas características em detrimento de outras. Contrariamente aos primeiros sensores desenvolvidos, os sensores SAR modernos possuem a capacidade de operar em diferentes modos consoante o objetivo e região a observar.

Até aos anos 50 do século passado, os radares de imagem eram designados por SLAR (*Side-Looking Airborne Radar*) e não faziam uso do princípio da abertura sintética, ainda não desenvolvido. Tais sistemas operavam acoplados a bordo de aviões e possuíam baixa resolução azimutal, que se deteriorava com o aumento da distância aos objetos observados à superfície. Esta característica foi a principal desvantagem do sistema SLAR, superada através do desenvolvimento de um radar coerente e do princípio de *Doppler Beam Sharpening* por Carl Wiley em 1951 (Wiley, 1985, citado por Moreira et al., 2013). Nos anos seguintes, este conceito foi estendido ao princípio de abertura sintética conhecido atualmente, o qual representou um marco histórico no desenvolvimento de sistemas radar de imagem nas décadas seguintes (Cutrona et al., 1961, citado por Moreira et al., 2013).

Os primeiros desenvolvimentos de sistemas SAR foram realizados em contexto militar nos anos 50 e 60, para deteção de alvos artificiais e para fins de reconhecimento. Nos anos 70 e 80 surgiram vários sistemas SAR operados a partir de avião para aplicações civis, com o intuito de estudar parâmetros geo/biofísicos da superfície da Terra. Este tipo de sistemas permitiu o desenvolvimento de novas técnicas e aplicações, mais tarde aplicadas em sistemas SAR operados a partir de satélite.

As principais missões SAR operadas a partir de satélite, passadas e atuais, desenvolvidas em contexto civil encontram-se resumidas na Tabela 2. Em 1978 foi lançado o sensor Seasat da agência

espacial norte-americana (NASA), o primeiro sensor SAR civil. A este seguiu-se nos anos 90 o lançamento dos sensores ERS-1 e ERS-2, JERS-1 e Radarsat-1 das agências espaciais Europeia, Japonesa e Canadiana, respetivamente.

Sensor	Agência	Período de operação	Banda
Seasat	NASA/JPL	1978	L
ERS-1/2	ESA	1991 – 2000 1995 – 2011	C
JERS-1	JAXA	1992 – 1998	L
SIR-C/ X-SAR	NASA/JPL DLR ASI	1994	L, C, X
Radarsat-1	CSA	1995 – 2013	C
SRTM	NASA/JPL DLR ASI	2000	C, X
ENVISAT/ASAR	ESA	2002 – 2012	C
Radarsat-2	CSA	2007 –	C
TerraSAR-X TanDEM-X	DLR/Astrium	2007 – 2010 –	X
COSMO-SkyMed 1-4	ASI /MoD	2007 – 2007 – 2008 – 2010 –	X
Sentinel-1 ^a /1B	ESA	2014 – 2016 –	C

Tabela 2. Principais sistemas SAR, operados a partir de satélite, desenvolvidos até à atualidade.

Nos anos 80 e 90 foram desenvolvidas diversas técnicas SAR, como é o caso da Interferometria SAR (InSAR), que permite a derivação de Modelos Digitais do Terreno (MDTs), e as técnicas diferenciais de interferometria (DInSAR), que permitem o estudo de deslocamentos superficiais da superfície terrestre (Bamler & Hartl, 1998; Bürgmann et al., 2000; Gabriel, Goldstein, & Zebker, 1989; Massonnet & Feigl, 1998; Massonnet et al., 1993; Rosen et al., 2000). Pouco tempo depois do lançamento do sensor ERS-2, os sensores ERS-1 e ERS-2 foram colocados no mesmo plano orbital, permitindo a obtenção de um tempo de revisita de apenas 24 horas (Ferretti et al., 2007). A órbita quase-polar dos sensores ERS, em combinação com o movimento de rotação da Terra, permitiu a aquisição de pares de imagens SAR obtidas de duas geometrias ligeiramente distintas.

A par da elevada frequência de aquisição, tal permitiu a geração de MDTs detalhados e a análise de variações na superfície terrestre num curto espaço de tempo.

Após o lançamento dos sensores SIR-C/X-SAR e SRTM (*Shuttle Radar Topography Mission*), em 1994 e 2000, respetivamente, ambas as técnicas anteriores sofreram um crescimento exponencial. O sensor SIR-C/X-SAR foi o primeiro sistema radar imagem de satélite capaz de obter simultaneamente imagens radar com multifrequência e multipolarização (Guo, Dou, Zhang, Han, & Yue, 2016). Construído na base deste último, o sensor SRTM era composto por duas antenas mono-estáticas acopladas ao sensor e um mastro com 60 metros de comprimento, na qual foram acopladas no final do mesmo duas antenas passivas (isto é, apenas para receção da radiação EM retro-difundida à superfície). Tal configuração, ilustrada na Figura 4, possibilitou pela primeira vez a aquisição simultânea de radiação EM segundo duas geometrias ligeiramente diferentes, sendo possível a derivação de MDTs de alta resolução de 80% da superfície terrestre com apenas uma passagem do sensor.

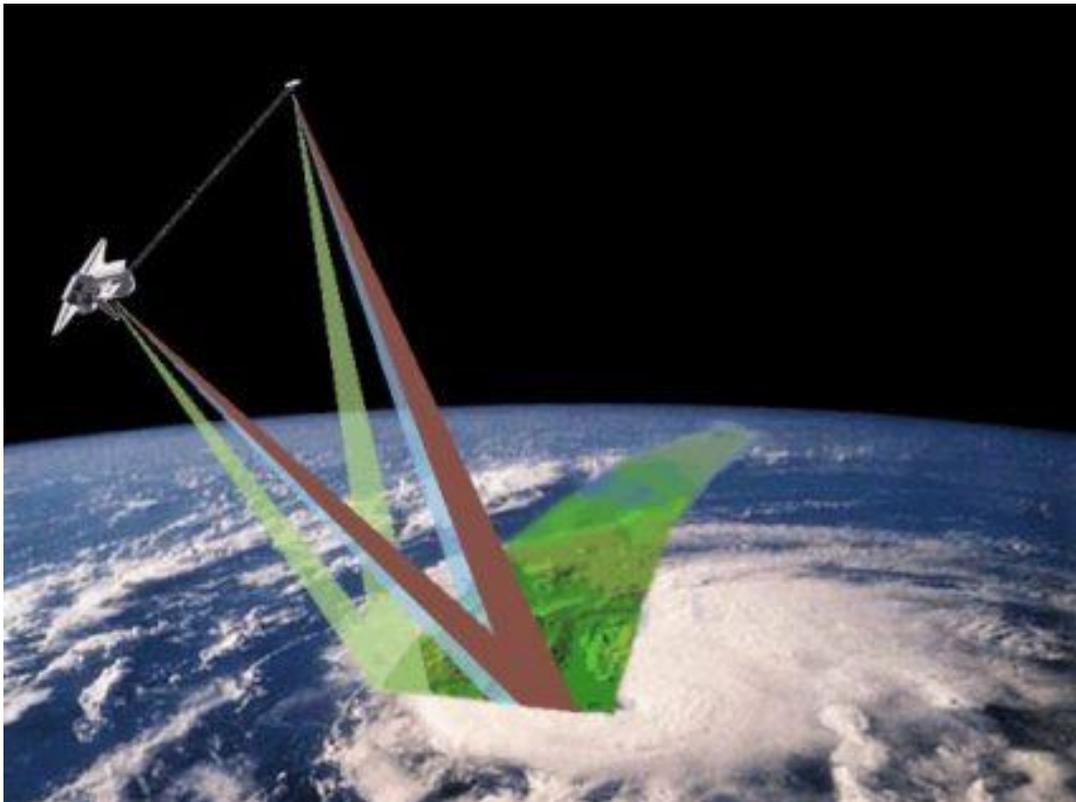


Figura 4. Figura ilustrativa do modo de operação do sensor SRTM (imagem obtida de: <https://gisgeography.com/srtm-shuttle-radar-topography-mission/>; acedido em 18/05/2020).

Lançado em 2002, o sensor ENVISAT da ESA foi o primeiro sensor com a capacidade de operar em diferentes modos de imagem. Juntamente com os sensores ERS-1 e ERS-2, os respetivos dados adquiridos permitiram o desenvolvimento de técnicas DInSAR usando PSs, para monitorização de subsidência do solo (Ferretti et al., 2000, 2001).

Com o lançamento dos satélites TerraSAR-X, TanDEM-X, Radarsat-2 e da constelação de satélites COSMO-SkyMed, foram exploradas uma nova classe de imagens SAR com uma resolução no terreno de ordem de grandeza métrica.

No panorama atual de missões operacionais, destaca-se a missão Sentinel-1. Esta missão, tal como as restantes missões Sentinel, é constituída por dois satélites no mesmo plano orbital – Sentinel-1A e Sentinel-2B. Tal configuração permite a resolução de questões relacionadas com o tempo de revisita de uma forma contínua e global (Torres et al., 2012). O destaque desta missão advém também da disseminação gratuita de dados, tal como em outras missões Sentinel, havendo atualmente inúmeros estudos com base em dados dos satélites Sentinel.

Efeito de *speckle*

Um efeito característico das imagens SAR é o chamado efeito de *speckle*. Este efeito é causado pela presença de inúmeros dispersores pontuais, com dimensões de ordem de grandeza semelhante ao comprimento de onda do radar, circunscritos na mesma célula de resolução. A soma dos sinais refletidos por cada um destes dispersores dá origem a uma flutuação de intensidade entre células de resolução adjacentes. Visualmente, a imagem SAR resultante é caracterizada pela presença de um ruído ‘sal-e-pimenta’, conforme ilustrado na imagem SAR da Figura 5.



Figura 5. Imagem SAR do satélite ERS-2 da ESA do Aeroporto de Linate, na região Este de Milão, Itália (figura obtida de Ferretti et al., 2007). O ruído ‘sal-e-pimenta’ nas regiões circundantes do aeroporto é devido ao efeito de *speckle*.

Para sistemas SAR de alta resolução este efeito tende a diminuir, visto que o número de dispersores pontuais circunscritos numa célula de resolução diminui.

O efeito de *speckle* está presente nas imagens SAR, mas não nas imagens da região do visível, tendo um impacto significativo na qualidade das primeiras. Para mitigar este efeito é habitualmente

utilizada uma técnica de *multi-look*, que consiste na aquisição de várias imagens SAR da mesma região em diferentes alturas ou de ângulos ligeiramente diferentes. Realizando uma média das várias imagens obtidas, é possível reduzir de forma eficaz este efeito, à custa da degradação de resolução.

Processamento de dados SAR

Ao contrário dos sensores óticos, os dados em bruto recolhidos pelos sensores SAR não fornecem qualquer tipo de informação relativamente à área observada. O eco recolhido pelo sensor tem de ser processado de forma a gerar uma imagem representativa da refletividade da área observada (este procedimento consiste na criação da abertura sintética). Numa primeira fase, o sinal recolhido pelo sensor é reamostrado em alcance e azimuth, formando uma matriz bidimensional de valores complexos. Cada valor da matriz é dado pela sua componente real e componente imaginária, representando os dois tipos de observáveis de um sistema SAR: amplitude e fase.

Após a operação de reamostragem do sinal, é realizada uma operação de convolução do mesmo com as respetivas funções de referência em alcance e azimuth. Esta operação, ilustrada na Figura 6, é realizada no domínio das frequências em detrimento do domínio temporal, devido ao facto de desta forma o esforço computacional ser muito menor.

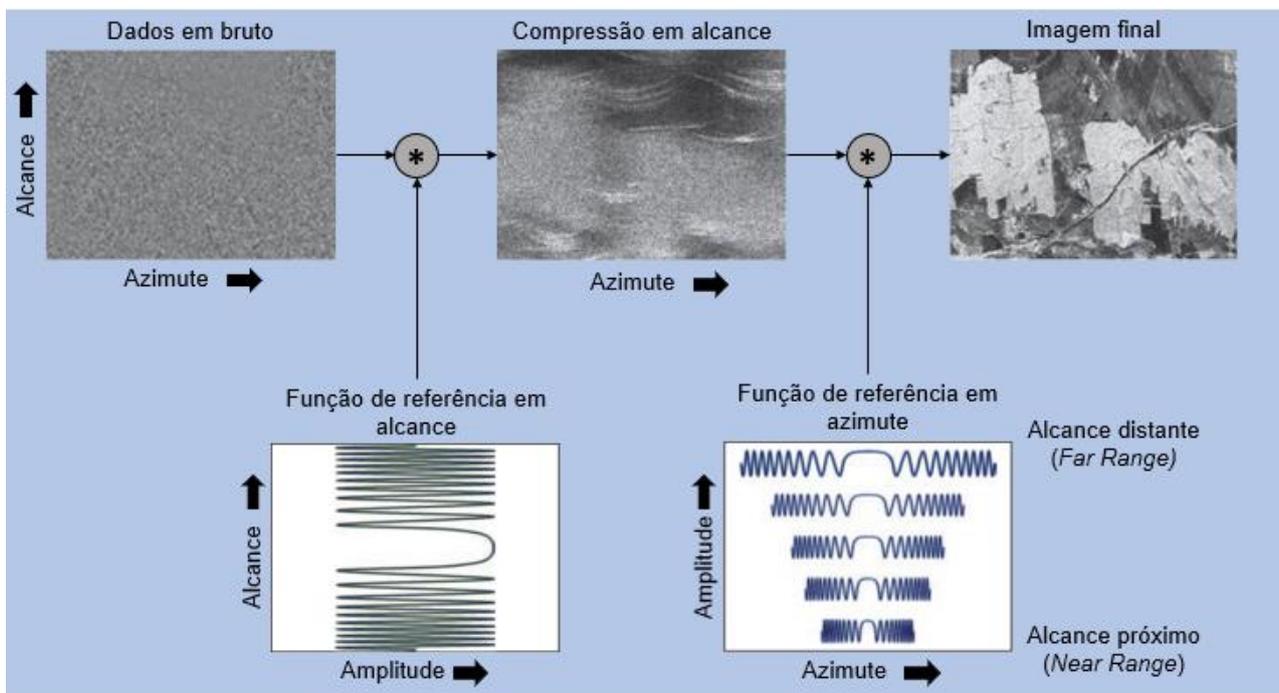


Figura 6. Esquematisação do processo de convolução do sinal reamostrado em alcance e azimuth com as respetivas funções de referência (imagem obtida de Moreira et al., 2013). O símbolo "*" na figura representa a operação de convolução.

O sinal previamente reamostrado em alcance é multiplicado pela respetiva função de referência, que consiste no complexo conjugado do sinal transmitido, resultando numa imagem comprimida em alcance. A imagem assim obtida é limitativa em termos visuais, fornecendo apenas informação da distância relativa entre o sensor e a superfície. A convolução em azimute é realizada numa segunda etapa e segue o mesmo princípio: o sinal reamostrado em azimute é multiplicado pela respetiva função de referência, que consiste no complexo conjugado da previsão do sinal retro-difundido por apenas um alvo pontual à superfície.

Conforme mencionado anteriormente, as imagens SAR são geralmente visualizadas através de valores de intensidade, onde cada pixel contém informação da refletividade da correspondente área à superfície. Tal exige a realização de duas etapas adicionais sobre a imagem resultante do processo de convolução: calibração e geocodificação.

O processo de calibração garante que cada valor de intensidade está relacionado com o valor de σ_0 , uma medida da força do sinal retro-difundido por um alvo pontual. Este processo envolve tanto a calibração interna do sensor como também uma calibração externa recorrendo a alvos pontuais cuja refletividade é conhecida (Freeman, 1992; citado por Moreira et al., 2013). A geocodificação consiste num processo de georreferenciação, garantindo assim que a localização de cada pixel na imagem SAR está diretamente associada à correspondente região à superfície.

2.2. Interferometria SAR

A Interferometria SAR (InSAR) é uma técnica de DR que permite, entre outras aplicações, a geração de MDTs e a medição de movimentos superficiais do terreno com precisões da ordem do comprimento de onda do radar (Bamler & Hartl, 1998; Bürgmann et al., 2000; Ferretti et al., 2007; Gabriel et al., 1989; Massonnet & Feigl, 1998; Pepe & Calò, 2017; Rosen et al., 2000). De uma forma simples, a técnica InSAR consiste na comparação, para uma dada região, da fase de pelo menos duas imagens SAR obtidas segundo geometrias ligeiramente diferentes e/ou em diferentes alturas. As diferentes geometrias de aquisição permitem a formação de configurações interferométricas distintas. No caso de um radar mono-estático, cujo modo de operação se encontra descrito no subcapítulo anterior, são usadas duas antenas para mapear uma dada região:

- simultaneamente de diferentes posições espaçadas na direção ortogonal à direção de voo (direção *across-track*);
- em diferentes alturas e da mesma posição;
- em diferentes alturas e de diferentes posições orbitais.

A primeira técnica é conhecida como interferometria *across-track*, sendo a segunda técnica designada por interferometria *along-track* ou DInSAR. A terceira configuração mencionada é uma variante da primeira e é designada por interferometria *repeat-pass across-track* (*repeat-pass across-track interferometry*).

Sensores bi- e multi-estáticos, nos quais, conforme referido no subcapítulo anterior, a transmissão e receção do sinal SAR é realizada por duas ou mais antenas, possibilita novas áreas de investigação (Duque, Lopez-Dekker, & Mallorqui, 2010; Krieger, Fiedler, Mittermayer, Papathanassiou, & Moreira, 2003; Krieger & Moreira, 2006). Tais sistemas permitem a exploração de novas geometrias de aquisição, melhoria de resolução em alcance e azimute, geração de MDTs de elevada qualidade e a superação das limitações interferométricas de sistemas SAR mono-estáticos (descritas mais á frente), tais como a descorrelação temporal e atrasos atmosféricos. No caso da monitorização de subsidência e elevação de terrenos, sistemas SAR multi-estáticos permitem a observação da mesma área simultaneamente de diferentes geometrias, sendo possível a extração de um modelo tridimensional e a superação dos efeitos de *foreshortening* e *layover* característicos dos sensores mono-estáticos. Uma vantagem adicional deste tipo de sistemas é o baixo custo relativamente a sistemas SAR mono-estáticos.

Desde o seu desenvolvimento nos anos 80 e 90 do século passado, surgiram diversos estudos usando a técnica InSAR para estudos de deformação sísmica e vulcânica (Juliet Biggs et al., 2009; Briole, Massonnet, & Delacourt, 1997; Massonnet, Briole, & Arnaud, 1995; Massonnet et al., 1993; Peltzer & Rosen, 1995; Ruch, Anderssohn, Walter, & Motagh, 2008), monitorização de subsidência,

elevação e deslizamentos de terrenos (Calò, Calcaterra, Iodice, Parise, & Ramondini, 2012; Ferretti et al., 2000; Guzzetti et al., 2009; Hunstad et al., 2009; Lauknes et al., 2010; Nof et al., 2019; Osmanoğlu, Dixon, Wdowinski, Cabral-Cano, & Jiang, 2011; Trasatti et al., 2008), entre outras aplicações. Adicionalmente, pela análise da coerência temporal, que descreve o grau de correlação entre as imagens SAR, é possível a aplicação da técnica InSAR para classificação do solo (Askne & Hagberg, 1993; Engdahl & Hyypä, 2003; Khalil & Saad ul, 2018; Sica et al., 2019; Vicente-Guijalba et al., 2018; Wegmuller & Werner, 1997; Wegmuller & Werner, 1995; Yun, Kim, Choi, & Lin, 2019).

Geração de Modelos Digitais do Terreno por interferometria *across-track*

A geometria de operação de um sistema SAR implica a projeção da área observada à superfície (a três dimensões) num plano bidimensional definido pelas coordenadas SAR azimute e alcance (Moreira et al., 2013). A resolução de cada pixel da imagem SAR resultante é dada pelas resoluções azimutal e em alcance, respetivamente, não fornecendo a imagem gerada informação da altitude do terreno.

O objetivo da interferometria *across-track* (e da sua variante *repeat-pass*) é precisamente a resolução da altitude do terreno, mapeando uma determinada área á superfície com uma segunda antena, conforme ilustrado na Figura 7.

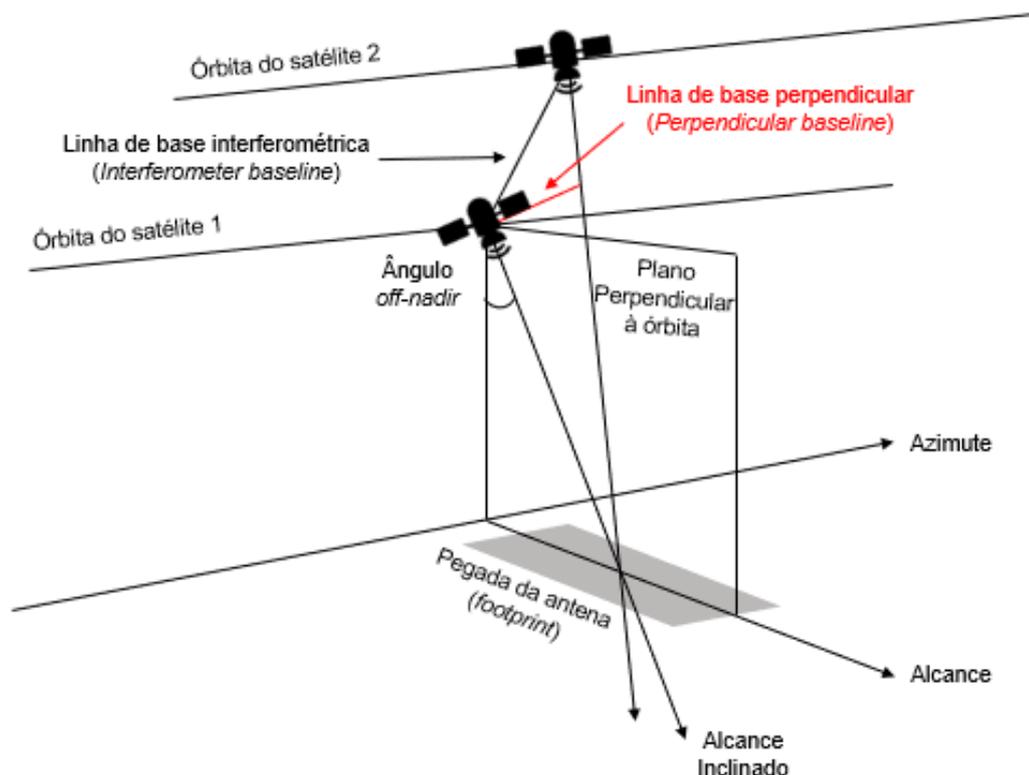


Figura 7. Geometria de operação da técnica InSAR (figura adaptada de Ferretti et al., 2007).

O desfasamento entre as geometrias de aquisição resulta numa diferença de distância entre cada antena e a superfície, podendo a mesma ser usada para inferir a altitude do terreno. A distância entre as duas antenas (ou entre as duas órbitas) no plano perpendicular à órbita é designada por linha de base interferométrica (*interferometer baseline*), e a projeção desta última na direção perpendicular à direção de alcance inclinado (*slant-range*) é designada por linha de base perpendicular (*perpendicular baseline*), um dos parâmetros chave da técnica InSAR.

O interferograma SAR, representativo da diferença de fase entre as duas imagens SAR, designada por fase interferométrica, é obtido através da multiplicação (pixel a pixel) da primeira imagem SAR, designada por imagem *master* (*master acquisition*), pelo complexo conjugado da segunda, designada por imagem *slave* (*slave acquisition*) (Bamler & Hartl, 1998; Massonnet & Feigl, 1998; Pepe & Calò, 2017; Rosen et al., 2000). A amplitude do interferograma, por outro lado, é obtida através da multiplicação (pixel a pixel) da amplitude da primeira imagem SAR pela amplitude da segunda. Previamente ao processo de obtenção do interferograma, as duas imagens SAR têm de ser alinhadas geograficamente, num processo designado por co-registo, garantindo assim que pixels correspondentes nas duas imagens sejam referentes à mesma região à superfície (Sansosti, Berardino, Manunta, Serafino, & Fornaro, 2006). O processo de co-registo consiste essencialmente na estimação da função de transformação que permite a reamostragem da imagem *slave* no sistema de referência da imagem *master*.

Como exemplo da geração da fase interferométrica, na Figura 8a encontra-se ilustrado um interferograma obtido a partir de um par de imagens SAR do sensor TanDEM-X da região do deserto do Atacama (Chile).

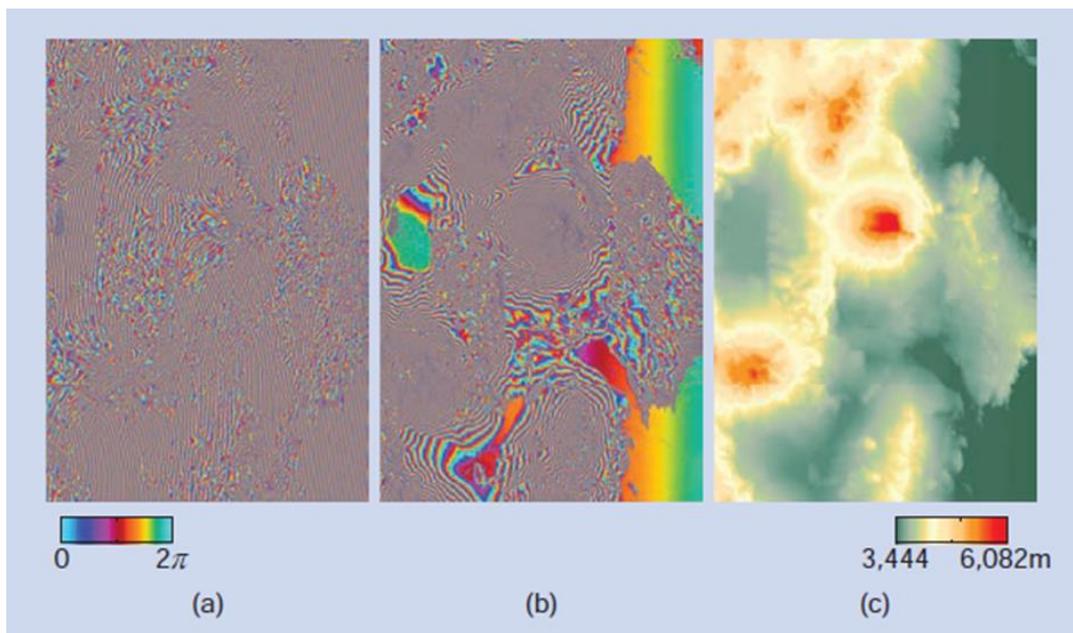


Figura 8. Interferogramas SAR e MDT do deserto do Atacama (Chile) gerados a partir de dados do sensor TanDEM-X (figura obtida de Moreira et al., 2013): (a) antes da remoção do efeito de *flat-earth*; (b) depois da remoção do efeito de *flat-earth*; (c) MDT gerado após o processo de desenrolamento de fase e correspondente conversão em valores de altitude.

Tipicamente, o interferograma SAR gerado é caracterizado pela presença de franjas interferométricas na direção em alcance (direção horizontal da esquerda para a direita na Figura 8a; a direção azimutal é vertical na figura e no sentido decrescente). Tal se deve maioritariamente pela diminuição da fase interferométrica com o aumento da distância à superfície e ângulo *off-nadir*. Este efeito, designado por efeito de *flat-earth* (Terra plana), é geralmente removido pela subtração da fase prevista relativamente a uma superfície de referência totalmente plana. Como exemplo, na Figura 8b encontra-se ilustrado o interferograma da Figura 8a após remoção do efeito de *flat-earth*. Pela respetiva análise, é possível verificar que as franjas interferométricas resultantes se assemelham nesta fase a curvas de nível.

A precisão do interferograma gerado é dependente da coerência temporal que, conforme referido anteriormente, descreve o grau de correlação entre as imagens SAR *master* e *slave*. Na prática, diversos fatores contribuem para a diminuição da coerência temporal, sendo a perda total de coerência dada pelo produto dos diversos fatores que contribuem para a redução da mesma (Rodriguez & Martin, 1992; citado por Moreira et al., 2013). A descorrelação temporal e a descorrelação geométrica são os principais fatores que contribuem para a diminuição da coerência temporal. A primeira ocorre na interferometria *repeat-pass*, e está relacionada com variações temporais da região observada entre as aquisições SAR. O segundo fator está relacionado com as diferentes geometrias de aquisição das imagens *master* e *slave*.

A descorrelação temporal é uma das maiores limitações da técnica InSAR, particularmente sobre áreas de vegetação, nas quais a coerência temporal diminui com o aumento da cobertura de vegetação. Neste caso em particular, a descorrelação temporal é mais significativa em sensores de baixa frequência (banda L) do que em sensores de alta frequência (banda C), sendo, por isso, vantajoso o uso de altas frequências na resolução da altitude do terreno neste tipo de regiões (Wei & Sandwell, 2010; citado por Pepe & Calò, 2017).

Outros fatores relevantes que contribuem para a diminuição da coerência temporal são, entre outros, o incorreto alinhamento das imagens SAR no processo de co-registo e a descorrelação interferométrica devido à linha de base. A estimativa da fase interferométrica e, por conseguinte, da altitude da área observada, é tanto melhor quanto maior for o valor da linha de base perpendicular. No entanto, o seu incremento sucessivo (e da correspondente linha de base perpendicular) resulta numa diminuição da coerência temporal entre as imagens SAR. O tamanho da linha de base que resulta numa perda total de coerência entre as imagens SAR *master* e *slave* é designada por linha de base crítica (*critical baseline*), sendo a mesma inversamente proporcional à resolução em alcance. Assim, é necessário estabelecer um compromisso relativamente ao tamanho da linha de base interferométrica para aplicações InSAR. A descorrelação relacionada com o tamanho da linha de base interferométrica é, em todo o caso, um problema menor em sistemas SAR modernos (Moreira et al., 2013).

A diminuição da coerência temporal pode causar erros de fase significativos nos interferogramas gerados. De forma a diminuir o ruído presente nos mesmos, é habitualmente usada

uma técnica de *multi-look* (descrita no subcapítulo anterior) previamente à extração dos valores de fase. Embora seja eficiente na remoção de erros de fase, a técnica de *multi-look* gera uma perda de resolução espacial por um fator n , sendo n o número de geometrias (*looks*) usados no processo.

O último passo para a geração de um MDT consiste na conversão dos valores de fase em altitudes e num processo de geocodificação, este último já descrito anteriormente. O processo de conversão, fortemente dependente da topografia do terreno, requer um conhecimento preciso da linha de base interferométrica entre as antenas radar e da diferença de distância entre cada dispersor à superfície e as mesmas. Uma medida frequentemente usada para a avaliação do MDT gerado é a chamada altitude de ambiguidade, representativa da diferença de altitude que origina uma variação de 2π na fase interferométrica (Pepe & Calò, 2017).

Conforme referido no subcapítulo anterior, as medidas de fase são ambíguas em múltiplos do comprimento de onda do sensor, resultando na existência de ambiguidades nas medidas de altitude do terreno. Tipicamente tais ambiguidades são resolvidas considerando informações externas da área observada e princípios relativos ao movimento do solo (por exemplo, assumindo que os movimentos de subsidência ou de elevação do terreno ocorrem de forma linear), um processo designado por desenrolamento de fase (*phase unwrapping*) (Ghiglia & Pritt, 1998; Goldstein, Zebker, & Werner, 1988; citados por Moreira et al., 2013). Na Figura 8c encontra-se ilustrado o MDT obtido através do interferograma da Figura 8b, após o processo de desenrolamento de fase.

Interferometria Diferencial

A técnica DInSAR é uma técnica capaz de detetar deslocamentos superficiais da superfície da Terra com precisões centimétricas (Moreira et al., 2013; Pepe & Calò, 2017). Semelhante à técnica InSAR para geração de MDTs, são adquiridas duas imagens SAR, espaçadas temporalmente, de forma a gerar um interferograma representativo da diferença de fase entre as duas aquisições. Na medição de movimentos superficiais do terreno, os movimentos de subsidência e elevação correspondem a diferenças de fase positivas e negativas, respetivamente (Osmanoğlu, Sunar, Wdowinski, & Cabral-Cano, 2016).

Idealmente, as duas imagens SAR estariam perfeitamente alinhadas geograficamente, resultando numa linha de base interferométrica nula. Nesta configuração a fase interferométrica seria apenas relativa a deslocamentos do terreno na direção de *slant-range* (i.e., da linha de visão do sensor). No entanto, em contexto real a linha de base interferométrica toma sempre valores não nulos, resultando num interferograma dependente também da topografia da região. Usando um MDT externo, esta dependência pode ser removida, subtraindo o primeiro ao interferograma gerado, obtendo-se, desta forma, um interferograma diferencial representativo de movimentos de subsidência e/ou elevação do terreno na direção em alcance.

Na técnica DInSAR o problema central continua a ser o desenrolamento e calibração da fase diferencial obtida, de forma a obter valores absolutos de deslocamentos superficiais. No âmbito desta técnica, dois fatores principais condicionam esta operação. Em primeiro lugar, a precisão do MDT externo mencionado anteriormente tem de ser da mesma ordem de grandeza do que a precisão da fase interferométrica medida. O segundo fator condicionante deve-se à presença de ruído no interferograma diferencial relativo à decorrelação temporal e variações no meio de propagação da radiação. Esta última contribuição é referente ao atraso atmosférico, maioritariamente devido ao vapor de água presente na troposfera, sendo possível a sua estimação usando medidas externas e modelos atmosféricos. Em bandas de frequência mais baixas, como é o caso da banda L, a densidade total de eletrões na ionosfera (*total electron content*, TEC) contribui adicionalmente para o efeito mencionado, resultando em medidas de deformação menos precisas. Em contrapartida, a decorrelação temporal é um fenómeno menos severo neste tipo de bandas do que em bandas de alta frequência. Erros orbitais, relacionados com um conhecimento impreciso da linha de base interferométrica, influenciam adicionalmente a precisão dos interferogramas diferenciais.

A separação das diferentes contribuições anteriormente mencionadas e as suas consequentes eliminações pode ser conseguida usando séries temporais de imagens SAR, ao invés de apenas duas aquisições. Tal abordagem consiste na seleção dos pixels com qualidade suficiente para serem usados no processo de estimação de deformações, sendo as abordagens mais comuns baseadas na técnica de PSs (Ferretti et al., 2000, 2001) e na análise da coerência temporal (Lanari et al., 2004; Mora, Mallorqui, & Broquetas, 2003; Z. Wang, Li, & Mills, 2018).

Um PS consiste num pixel nas imagens SAR caracterizado por uma dispersão de amplitude baixa no tempo, correspondendo habitualmente a alvos pontuais como edifícios ou rochas. A definição dos PSs com base na amplitude tem como base o cálculo da dispersão temporal de amplitude, dada pela razão entre o desvio padrão e a média da mesma (Ferretti et al., 2001). Habitualmente, são considerados como PSs dispersores cuja dispersão temporal é menor do que 0.25. Sendo os PSs mais abundantes nas áreas urbanas, leva a que as técnicas DInSAR usando PSs sejam particularmente eficazes na monitorização de movimentos superficiais do terreno neste tipo de áreas (Soergel, U., 2010; citado por Osmanoğlu et al., 2016). Nas áreas com baixa densidade de PSs podem adicionalmente ser implementados refletores artificiais (*Artificial Corner Reflectors*, CRs), permitindo uma melhor análise de deformação das mesmas (Crosetto, Monserrat, Cuevas-González, Devanthéry, & Crippa, 2016).

A seleção de pixels com base na coerência temporal, em comparação com as técnicas DInSAR usando PSs, permite a inclusão de dispersores dispersos (*distributed scatterers*, DSs) que se mantêm coerentes no tempo, à custa de uma perda de resolução. Os DSs diferenciam-se dos PSs em termos de tamanho e de capacidade de difusão do sinal SAR. Um exemplo deste tipo de técnica é o algoritmo SqueeSAR™, que consiste na análise de séries temporais de dados SAR pela combinação de PSs e DSs (Ferretti et al., 2011).

No âmbito da seleção de pixels com base na coerência temporal, uma abordagem habitual consiste em considerar como pixels válidos aqueles cuja coerência temporal $\epsilon [0,1]$ excede o valor de 0.7, sendo rejeitados os restantes (Ferretti et al., 2001). No entanto, na prática tal resulta numa redução da densidade de PSs em zonas urbanas e numa perda quase total de PSs em zonas rurais. Desta forma, diversas abordagens alternativas foram propostas, de forma a superar as limitações resultantes da aplicação de *threshold* na coerência temporal (Bakon et al., 2017; Perissin & Wang, 2012). Em particular, as metodologias propostas por Bakon et al. (2017), nas quais é considerada a correlação espacial entre PSs vizinhos e parâmetros multi-temporais adicionais, permitem uma melhor avaliação de PSs de baixa coerência (habitualmente descartados pelo método anterior) e uma melhor análise de áreas com padrões de deformação mais complexos.

Para além das técnicas mencionadas anteriormente, existem outros métodos de análise de séries temporais de dados SAR. A técnica StaMPS, considerada uma variante das técnicas DInSAR usando PSs, consiste na seleção de PSs com base na análise temporal e espacial da fase interferométrica, sendo esta seleção independente da amplitude (Hooper, Zebker, Segall, & Kampes, 2004). Esta independência permite o uso desta técnica para análise de deformações em áreas rurais. Outra técnica existente é o algoritmo SBAS (*Small BAseline Subset*), no qual a análise de séries temporais de dados SAR é realizada usando exclusivamente DSs. A par da técnica anterior, esta técnica possui também a capacidade de ser usada em áreas rurais.

2.3. Machine Learning e Deep Learning

Machine Learning

Nascida nos anos 50 do século passado, a IA surge da intenção de automatização de tarefas intelectuais habitualmente executadas por seres humanos (Chollet, 2018). A IA é uma área geral dividida em diferentes ramos, entre os quais o ML e o DL.

Desde a sua criação e durante décadas, o paradigma central relativamente a IA a um nível humano consistia na asserção de que tal podia ser alcançado através da codificação de um vasto conjunto de regras, abordagem conhecida como *symbolic AI*, sendo o paradigma dominante até aos anos 80. Embora esta abordagem tenha provado a sua eficiência em diversos problemas lógicos, verificou-se que a mesma era ineficiente na resolução de problemas mais complexos, como é o caso de problemas perceptuais tais como classificação de imagens, tradução de texto e reconhecimento de fala.

Conhecidas as limitações da abordagem anterior, nos anos 80 surgiu um novo paradigma, o ML. A ascensão do ML, impulsionada pela disponibilidade crescente de melhor hardware, melhores bases de dados e avanços algorítmicos, deveu-se ao intuito de automatização da codificação de regras anteriormente mencionada, na qual se baseava a abordagem anterior, facultados os dados apropriados. Desta forma, o ML possui um modo de operação distinto da abordagem anterior: em programação clássica, na qual se enquadra a *symbolic AI*, eram criadas regras e facultados dados a ser processados de acordo com as mesmas de forma a processar estes últimos; em ML são fornecidos dados e as respetivas respostas esperadas do processamento dos mesmos de forma a obter as regras apropriadas que permitem a automatização do processamento em causa. Posteriormente, estas regras podem ser aplicadas em novos dados de forma a processar igualmente os mesmos. Assim, um sistema de ML é treinado ao invés de codificado. Presenteando o sistema com um elevado número de exemplos relevantes para uma determinada tarefa, o mesmo produz de forma automática regras que permitem eventualmente a automatização da mesma.

Os algoritmos de ML (e consequentemente os de DL) são divididos em dois tipos: aprendizagem supervisionada, o mais comum, e não supervisionada. Os algoritmos de aprendizagem não supervisionada encontram-se fora do enquadramento do presente trabalho, pelo que não serão discutidos na presente dissertação. Os algoritmos de aprendizagem supervisionada são habitualmente aplicados em problemas de classificação e de regressão. Em ambas as aplicações, para além dos dados e das correspondentes respostas esperadas, é necessário proceder a uma medição entre a resposta do sistema e a resposta esperada, ajustando com base na mesma o modo de operação do sistema. Este ajuste, realizado pelos sistemas de ML e DL de forma automática, consiste essencialmente numa busca por melhores representações dos dados de entrada dos sistemas de entre um conjunto pré-definido de operações, designado por espaço de

hipóteses (*hypothesis space*), com o objetivo de minimizar o desfasamento entre a resposta fornecida pelos mesmos e a resposta esperada. Tais representações podem consistir em mudanças de coordenadas, projeções lineares, translações, operações não-lineares, entre outras.

Um dos primeiros algoritmos de ML desenvolvidos é o algoritmo de *Naive Bayes*. Este algoritmo consiste na aplicação do teorema de *Bayes* assumindo que os parâmetros nos dados de entrada do sistema são todos independentes entre si. Um algoritmo semelhante ao anterior é o algoritmo de regressão logística (*logreg*). Ao contrário do que o nome indica, trata-se de um algoritmo de classificação e não de regressão.

Em meados dos anos 90, surgiram os métodos de *kernel*, um grupo de algoritmos de classificação, entre os quais se encontra o algoritmo de suporte vetorial (*Support Vector Machine*, SVM), um algoritmo de classificação binária desenvolvido por Vladimir Vapnik e Corinna Cortes (Cortes & Vapnik, 1995). O modo de operação dos SVMs consiste na computação de linhas de fronteira entre dois conjuntos de dados de classes distintas, podendo as mesmas consistir numa reta ou plano. Uma das características fundamentais na computação das linhas de fronteira por parte dos SVMs consiste em maximizar a distância entre a linha de fronteira e as amostras de dados mais próximas de cada classe, permitindo, desta forma, a generalização da mesma a amostras de dados independentes dos dados de treino. Assim, de forma a classificar novos dados é apenas necessário verificar a que lado da linha de fronteira os mesmos pertencem.

Por volta do ano 2000, surgiram as árvores de decisão (*Decision Trees*, DTs), um algoritmo de ML com uma estrutura semelhante a um fluxograma, podendo as mesmas ser aplicadas tanto em problemas de classificação como de regressão. O algoritmo *Random Forest* (RF), em particular, permite a construção de várias DTs especializadas e no final agrupar os resultados de cada DT (via a média simples, no caso de problemas de regressão, ou a moda, no caso de problemas de classificação). De forma semelhante às RFs, as *Gradient Boosting Machines* (GBs) consistem na construção de várias DTs especializadas com a aplicação de um método iterativo que permite a construção sistemática de novos sistemas especializados em colmatar os pontos fracos de modelos anteriores. Aplicados a DTs, os algoritmos de GBs permitem frequentemente a obtenção de resultados mais eficientes relativamente àqueles obtidos por RFs.

Apesar dos avanços obtidos com os sistemas de ML, os mesmos eram limitados pela sua ineficiência de processar dados em bruto. Como consequência, tais sistemas não forneciam resultados satisfatórios em problemas percetuais, exigindo um pré-processamento sistemático dos dados de entrada de forma a obter representações mais favoráveis, operação designada por *feature engineering*, o que requeria um elevado conhecimento do problema em causa. Adicionalmente, tais sistemas comprovaram ser ineficientes no processamento de grandes conjuntos de dados.

Deep Learning e Redes Neurais

O DL, um ramo do ML, remove em grande parte a necessidade de *feature engineering*, na medida em que os sistemas inseridos neste ramo procedem de forma automática à extração de características relevantes dos dados em bruto, implementando várias camadas sequenciais de representações gradualmente mais complexas e significativas (LeCun, Bengio, & Hinton, 2015). Os sistemas de ML são habitualmente designados por sistemas de *shallow learning*, devido ao número reduzido de transformações nos dados de entrada implementadas pelos mesmos; por outro lado, os sistemas de DL atuais podem envolver dezenas ou centenas de camadas, todas ajustadas automaticamente por exposição a dados apropriados, tendo a capacidade de processar dados em bruto e oferecendo resultados notáveis em problemas percutuais.

Em DL, as implementações das camadas de transformação anteriores são habitualmente obtidas através de Redes Neurais (RNs), nas quais tais camadas são geralmente agrupadas de forma sequencial. A arquitetura habitual de uma RN encontra-se ilustrada na Figura 9.

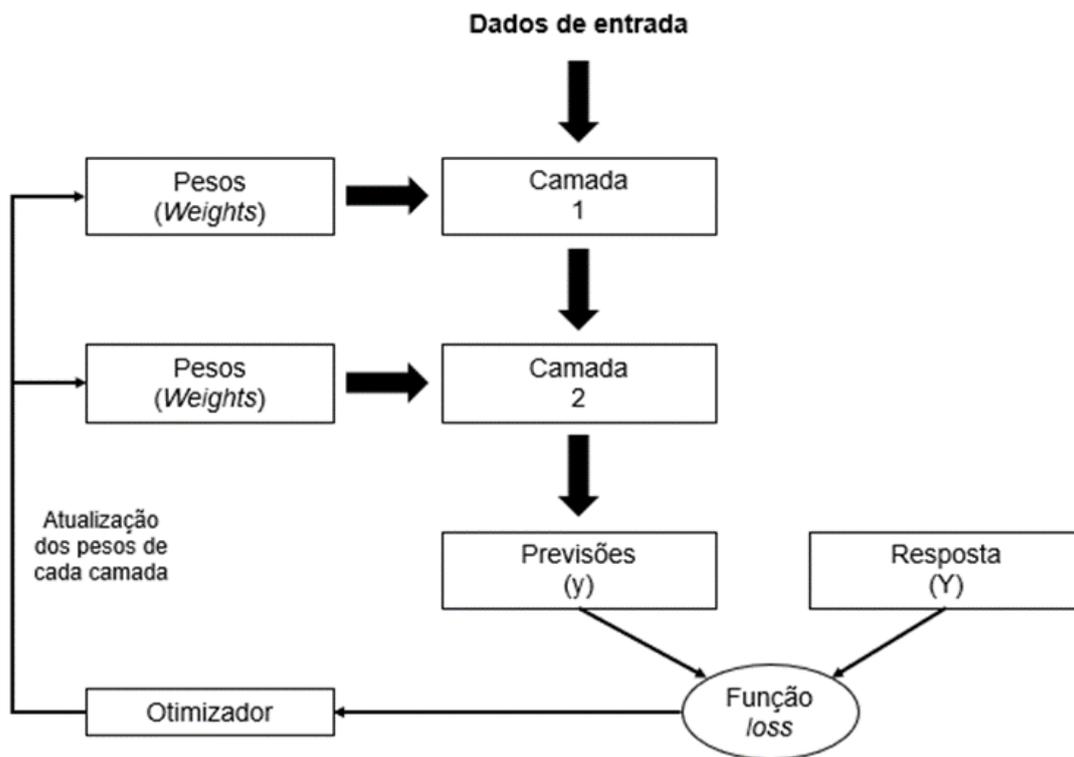


Figura 9. Arquitetura habitual de uma RN, um modelo sequencial no qual a última camada corresponde ao classificador final (figura adaptada de Chollet, 2018).

O tipo de transformação implementada por cada camada é especificado pelos parâmetros da mesma, também designados por pesos. Inicialmente estão associados valores aleatórios aos parâmetros de cada camada, originando transformações aleatórias por parte das mesmas. O treino

da RN consiste essencialmente no ajuste dos parâmetros de cada camada num processo iterativo, onde em cada iteração as previsões retornadas pela RN (y) são comparadas com a resposta esperada (Y), sendo a medição do desfasamento entre as duas respostas realizada pela função *loss*. Nesta fase encontra-se a característica fundamental do DL: a medição realizada pela função *loss* é transmitida ao otimizador do sistema que, implementando o algoritmo de *backpropagation*, ajusta os parâmetros de cada camada de forma a melhorar as previsões criadas pelo sistema.

Os sistemas de DL habitualmente não processam todos os dados de treino de uma só vez; alternativamente, os dados são divididos em pequenos conjuntos (*batches*), sendo o processamento de cada *batch* realizado de forma sequencial pelo método anterior. Na fase de treino dos sistemas, uma iteração corresponde ao processamento de todos os *batches* de dados de treino.

Na prática, as RN consistem em várias operações diferenciáveis em cadeia. A aplicação da regra da cadeia permite o cálculo do gradiente da função *loss* em função dos parâmetros do sistema, operação designada por algoritmo de *backpropagation*. A propagação do erro medido pela função *loss* das camadas finais até às camadas iniciais do sistema permite a determinação da contribuição de cada parâmetro na previsão final por parte do sistema. Assim, atualizando os pesos de cada camada na direção oposta à do gradiente da função *loss*, o desfasamento entre a resposta criada pelo sistema e a resposta esperada diminui com cada iteração. Tal operação é designada por *mini-batch stochastic gradient descent (mini-batch SGD)*, cujo modo de operação é função da taxa de aprendizagem (*learning rate*, LR) definida. A definição de um valor apropriado para a LR do sistema é um fator importante: no caso de este ser muito baixo, o treino do sistema irá requerer várias iterações, sendo possível que o mesmo fique “preso” num mínimo local; por outro lado, no caso de o LR for muito alto o eficiente treino do sistema poderá estar comprometido.

Uma variante da operação *mini-batch SGD* consiste em processar apenas uma amostra de cada vez, ao invés de processar *batches* de dados, operação designada por verdadeiro SGD (*true SGD*). No extremo oposto, seria possível processar todos os dados de treino de uma só vez, operação designada por *batch SGD*. Adicionalmente, existem outras variantes de SGD, as quais consideram na atualização dos pesos de cada camada as atualizações anteriormente implementadas, ao invés de considerar apenas o valor atual do gradiente. Na prática, é necessário determinar um valor eficiente para o tamanho de cada *batch* de dados, sendo que valores elevados associados ao mesmo previnem a generalização dos sistemas, o que se traduz em resultados ineficientes em dados independentes dos dados de treino (Hoffer, Hubara, & Soudry, 2017; Keskar, Nocedal, Tang, Mudigere, & Smelyanskiy, 2017).

De forma a melhorar o espaço de hipóteses de transformações implementadas por cada camada, é habitual a definição de funções de ativação. Tais funções permitem a cada camada a possibilidade de realizar transformações não-lineares - sem a definição das mesmas, cada camada teria apenas a capacidade de realizar transformações lineares aos dados de entrada, usufruindo, por conseguinte, de um espaço de hipóteses bastante restritivo. Uma das funções de ativação mais

eficientes e mais usadas em camadas intermédias é a função *relu*, dada pela equação (5). O respetivo gráfico encontra-se ilustrado na Figura 10.

$$\text{relu}(x) = \max(x, 0) \tag{5}$$

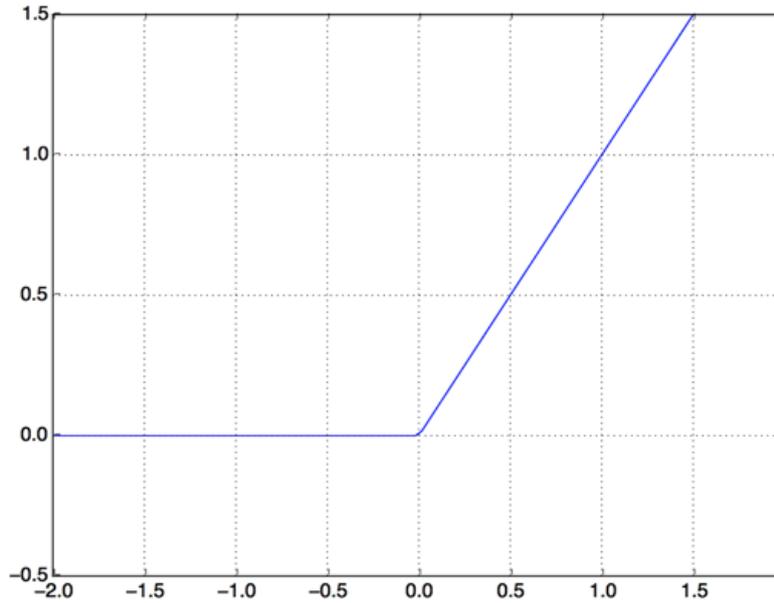


Figura 10. Gráfico da função de ativação *relu* (figura obtida de Chollet, 2018).

A função *relu* é uma função de ativação bastante simples, cujo objetivo consiste em suprimir valores negativos nos dados de entrada da camada correspondente.

A função de ativação da camada final de cada RN é dependente do tipo de problema em questão. No caso de problemas de classificação binária, é usada a função *sigmoid*, dada pela equação (6). O respetivo gráfico encontra-se ilustrado na Figura 11.

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{6}$$

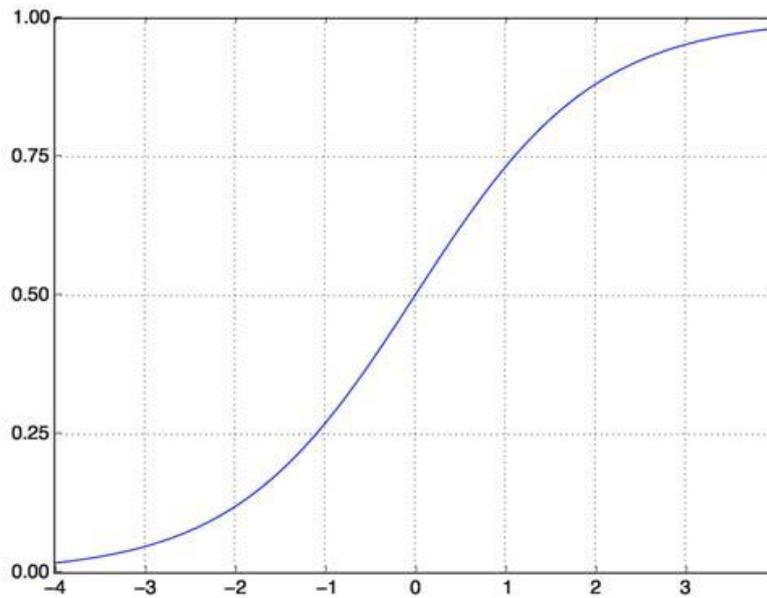


Figura 11. Gráfico da função de ativação *sigmoid* (figura obtida de Chollet, 2018).

Desta forma, a camada final da RN, isto é, o classificador final, prevê a probabilidade de cada amostra de dados pertencer a cada uma das duas classes. No caso de problemas de classificação em várias classes, é usada a função de ativação *softmax*. De forma semelhante à função *sigmoid*, neste caso o classificador final prevê a probabilidade de uma amostra de dados pertencer a cada uma das classes. Adicionalmente, em problemas de classificação em classes o classificador final possui N unidades, sendo N o número de classes (o caso $N = 2$ corresponde, desta forma, a classificação binária). No caso de problemas de regressão, o classificador final é composto apenas por uma unidade e sem função de ativação (sendo, desta forma, uma camada linear). Assim, para cada amostra de dados o mesmo prevê apenas um valor contínuo. Neste caso, a definição de uma função de ativação iria restringir a gama de valores que o mesmo é capaz de prever, pelo que a inexistência da mesma possibilita a previsão de valores contínuos sem restrições.

Relativamente à compilação dos sistemas de ML e DL, que consiste na definição do otimizador e função *loss* dos sistemas, tal operação é dependente do tipo de problema em questão. No caso de problemas de classificação, a função *loss* mais usada é a função *Binary Crossentropy*, no caso de classificação binária, e a função *Categorical Crossentropy*, no caso de classificação multi-classe. A função *Crossentropy* tem como objetivo a medição da distribuição de probabilidades prevista pelos sistemas de ML e DL e a verdadeira distribuição de probabilidades dos dados. O processo de treino de sistemas de ML e DL de classificação binária ou multi-classe consiste, assim, na minimização entre as duas distribuições. No caso de problemas de regressão, no qual a saída dos sistemas é previsto apenas um valor contínuo, é habitual definir como função *loss* a função Erro Médio Quadrático (EMQ). Como otimizador, é habitual a definição da função *rmsprop*, permitindo a mesma um treino eficiente dos sistemas independentemente do problema em questão.

Para além do tipo de arquitetura até agora descrito, consistindo em apenas um *input* e um *output*, existem outros tipos de arquitetura de sistemas de ML e DL, tais como sistemas *multi-input*, sistemas *multi-output*, sistemas *multi-input* e *multi-output*, entre outros. No caso de sistemas *multi-output*, é necessário proceder à definição de uma função *loss* para cada saída. No entanto, o processo iterativo de SGD tem de ser baseado num único valor de *loss*; assim, em sistemas *multi-output* é realizada uma média dos valores de *loss* de cada saída e o processo de SGD é implementado com base no valor médio obtido.

Técnicas de regularização e generalização de Redes Neurais

A capacidade de generalização dos sistemas de ML e DL é um conceito fundamental. Um dos maiores obstáculos neste âmbito consiste na eventual existência de um efeito de sobre-ajustamento (*overfitting*) por parte dos mesmos. Tal como o nome indica, este efeito é caracterizado por um ajustamento excessivo dos sistemas aos dados de treino, o que se traduz em resultados ineficientes em dados independentes. Por outras palavras, ocorre um ajustamento dos sistemas aos dados de treino com base em características específicas destes últimos, sendo estas inexistentes ou irrelevantes em dados independentes.

O efeito de *overfitting* pode ser minimizado de diversas formas. Uma das técnicas mais simples e mais eficientes consiste na aquisição de mais dados de treino, o que permite uma maior generalização dos sistemas. Alternativamente existem outras técnicas que atuam diretamente nos sistemas de ML e DL, designadas por técnicas de regularização. Neste âmbito, a técnica mais comum e mais simples consiste em diminuir a complexidade do sistema, reduzindo o número de camadas do mesmo e, por conseguinte, o número de parâmetros. Outras técnicas existentes consistem na regularização dos pesos do sistema (*weight regularization*) e na adição de *Dropout*. As técnicas de regularização dos pesos do sistema consistem em “forçar” os parâmetros do mesmo a tomar apenas valores pequenos, tornando a correspondente distribuição de valores mais regular. Na prática, esta regularização pode ser conseguida de duas formas – regularização L1 e L2 - adicionando à função *loss* do sistema um “custo” com base na existência de pesos com valores elevados:

- na regularização L1, o “custo” adicionado é proporcional ao valor absoluto dos pesos do sistema;
- na regularização L2, o “custo” adicionado é proporcional ao quadrado do valor dos pesos do sistema.

Por último, o *Dropout*, desenvolvido na Universidade de Toronto, é uma das técnicas mais eficazes e mais usadas na regularização de sistemas de ML e DL (Srivastava, Hinton, Krizhevsky,

Sutskever, & Salakhutdinov, 2014). Na prática, o *Dropout* consiste em definir a zero de forma aleatória uma determinada percentagem pré-definida (geralmente entre 20% e 50%) de valores do *output* de uma determinada camada. Esta operação é apenas aplicada nos dados de treino, não sendo aplicada em dados independentes. Alternativamente, na fase de avaliação dos sistemas o *output* de cada camada do sistema é minimizado por um fator igual à taxa de *Dropout* pré-definida. A ideia chave na regularização com *Dropout* consiste na introdução de ruído no *output* de cada camada, permitindo reduzir o ajustamento por parte do sistema a padrões irrelevantes dos dados de treino.

3. Estado da arte

À data de realização da presente dissertação, a aplicação de técnicas de ML e DL na Interferometria SAR, à semelhança da DR em geral, é de menor dimensão comparativamente a outras áreas. Neste capítulo será feita uma revisão das aplicações encontradas neste âmbito.

3.1. Estudos de deformação superficial

É conhecida a relação entre deformações superficiais em vulcões e a consequente erupção dos mesmos, sendo por isso crucial a contínua monitorização de padrões de deformação em vulcões a uma escala global (J. Biggs et al., 2014). Neste âmbito, Anantrasirichai et al. (2018) propuseram uma abordagem para deteção de deformações vulcânicas em Interferogramas SAR usando Redes Neurais de Convolução (CNNs). Aplicando uma técnica de *transfer learning*, os autores determinaram que a rede pré-treinada AlexNet apresenta os melhores resultados na tarefa em causa. Este estudo consistiu na primeira publicação neste âmbito usando algoritmos de ML, ficando provada a aplicabilidade dos mesmos. No entanto, e conforme constatado pelos autores, é necessário mais desenvolvimento para deteção de padrões de deformação em pequena escala ou que apresentam uma velocidade de subsidência lenta.

Em publicações mais recentes, os mesmos autores propuseram uma abordagem semelhante à anterior, usando, neste caso, Interferogramas SAR sintéticos aplicados no treino da CNN AlexNet (Anantrasirichai et al., 2019a, 2019b). O uso de Interferogramas SAR sintéticos deve-se ao facto de, conforme referido pelos autores, ser necessário possuir um número balanceado de casos positivos e negativos, isto é, que apresentam ou não deformações superficiais, de forma a eficazmente diferenciar deformações reais de ruído. Adicionalmente, apenas uma pequena percentagem de vulcões apresenta deformações superficiais, impulsionando o desfasamento entre casos positivos e negativos. Os resultados obtidos pelos autores confirmam o uso de dados sintéticos, na medida em que a CNN AlexNet treinada com tais dados supera os resultados obtidos usando apenas Interferogramas SAR reais.

Valade et al. (2019) propuseram uma abordagem semelhante à anterior, consistindo no treino de CNNs usando Interferogramas SAR sintéticos, com o propósito de tais modelos serem aplicados na deteção de deformações vulcânicas em Interferogramas SAR reais produzidos pela plataforma MOUNTS (*Monitoring Unrest from Space*), uma plataforma de monitorização global de vulcões usando dados de satélite, medidas adquiridas no terreno e IA. Tal abordagem multidisciplinar, fazendo uso de dados de múltiplos sensores, permite um estudo aprofundado de erupções vulcânicas e uma melhor compreensão da possível integração de diferentes tipos de dados na monitorização de vulcões.

No âmbito da deteção de deformações superficiais em zonas urbanas, Wang et al. (2019) propuseram a aplicação do algoritmo de *back-propagation* para estimação de deformações superficiais na zona sul da região de Erhai, China. Para tal, foram usados quatro parâmetros, nomeadamente a área de construção, nível da água, precipitação cumulativa e temperatura cumulativa. Os resultados obtidos demonstraram a potencialidade do algoritmo de *back-propagation* na previsão de deformações superficiais com elevada precisão.

Ma et al. (2019) propuseram a aplicação de CNNs para previsão de deformações superficiais com base em padrões de deformação temporais já conhecidos. A metodologia proposta consiste na extração da deformação temporal a partir de imagens SAR de forma a obter PSs e DSs, seguida da aplicação de um método de interpolação para obter uma grelha regular de deformação, esta última correspondendo aos dados de entrada das CNNs. Os resultados obtidos pelos autores demonstram que tal abordagem é aplicável na previsão de deformações temporais a curto prazo.

3.2. Classificação do uso do solo com base na coerência temporal

A coerência temporal consiste num parâmetro com potencial relevância para classificação do uso do solo, conforme já referido no Capítulo 2.2. No âmbito do projeto SEOM SInCohMap da ESA, um projeto que visa testar e analisar as potencialidades da coerência temporal derivada de dados SAR do sensor Sentinel-1 para classificação do uso do solo, Vicente-Guijalba et al. (2018) publicaram um estudo com os resultados preliminares do projeto, no qual foram aplicados os algoritmos RF e SVM para classificação do uso do solo com base em mapas de coerência (sendo, por isso, a classificação realizada pixel a pixel). Os resultados preliminares, avaliados comparativamente à carta de ocupação do solo CORINE Land Cover, revelaram-se promissores, tendo sido obtidas exatidões na ordem de 80%.

Sica et al. (2019) propuseram um estudo de classificação do uso do solo usando o algoritmo RF. A metodologia proposta consiste em modelar a descorrelação temporal, relacionada com a perda de coerência temporal entre duas ou mais aquisições SAR, e usar os parâmetros estimados como dados de entrada do algoritmo. A classificação proposta pelos autores encontra-se dividida em três classes: superfícies artificiais (como, por exemplo, zonas urbanas), zonas de floresta densa e zonas rurais, esta última constituída por zonas de pouca vegetação, zonas agrícolas, entre outras. Os resultados obtidos demonstraram um nível de concordância com a carta de ocupação do solo CORINE Land Cover superior a 90%, a qual foi usada como referência externa para o treino e teste do algoritmo de classificação.

3.3. Filtragem de ruído em Interferogramas SAR

A filtragem de ruído em Interferogramas SAR usando algoritmos de ML é outra das aplicações com publicações até à data. Ichikawa & Hirose (2017) propuseram uma abordagem consistindo na aplicação de RNs a qual, segundo os resultados obtidos, é mais eficaz na remoção de ruído do que filtros convencionais. Kottayil et al. (2018) propuseram uma abordagem semelhante, pela aplicação de CNNs para deteção de pixels em Interferogramas SAR afetados por ruído. Conforme referido pelos autores, tal abordagem é capaz de detetar ruído de forma eficaz, embora determinados fatores indutores de ruído, tais como fatores atmosféricos, possam ser considerados em trabalhos futuros.

Li et al. (2019) propuseram também um trabalho neste âmbito, dividido em duas etapas. Conforme referido pelos autores, comparativamente a outras regiões, pixels em Interferogramas SAR relativos a regiões de *layover* são caracterizados por serem de difícil filtragem de ruído, pelo que, numa primeira fase, os autores procederam à aplicação de uma CNN para segmentação de pixels relativos a regiões de *layover* dos restantes, seguida da aplicação de uma segunda CNN para remoção de ruído em pixels não relativos a regiões de *layover*. Os resultados obtidos permitem verificar que tal abordagem permite evitar a propagação de ruído de pixels relativos a regiões de *layover*, aumentando significativamente a precisão da operação de desenrolamento de fase posterior à geração do Interferograma SAR.

3.4. Outras aplicações

Para além das publicações mencionadas anteriormente, existem outras com diferentes objetivos. Conforme mencionado no Capítulo 3, existem diversos fatores indutores de erro que limitam a eficiência da técnica InSAR, exigindo pós-processamento de forma a separar cada um destes fatores. Neste âmbito, Rongier et al. (2019) propuseram a aplicação de modelos generativos de ML para geração de Interferogramas SAR sintéticos que incluem cada um dos fatores indutores de erro conhecidos, permitindo uma avaliação mais controlada da eficiência dos algoritmos de correção InSAR e uma melhor compreensão dos diferentes tipos de ruído existentes em Interferogramas SAR.

Shamshiri et al. (2019) propuseram a aplicação de processos gaussianos de ML, usando uma combinação de Interferogramas SAR e medidas GNSS (*Global Navigation Satellite Systems*), com o objetivo de mitigar as diferenças de fase causadas por efeitos troposféricos em medidas InSAR. Pela aplicação da abordagem proposta em 12 imagens SAR do sensor Sentinel-1, foi possível obter uma redução de, em média, de 83% do erro médio quadrático, um aumento significativo de 33% face à mitigação obtida por modelos troposféricos.

Costante et al. (2018) propuseram a aplicação de CNNs para geração de MDTs usando apenas uma geometria de dados SAR, ao contrário da habitual abordagem de aquisição de dados sob duas geometrias ligeiramente desfasadas, descrita no Capítulo 3. Pelos resultados preliminares obtidos em imagens SAR do sensor Sentinel-1, os autores verificaram que a abordagem proposta se revelou promissora, tendo sido gerados MDTs com exatidões na ordem dos 40 a 100 metros.

No âmbito da técnica DInSAR usando PSs, Kerkhof et al. (2018) propuseram um método de pós-processamento para melhoramento da estimação da série temporal da fase interferométrica, pela aplicação de algoritmos de ML e teste de hipóteses. Numa primeira fase, os autores procederam à aplicação do algoritmo *t-Distributed Stochastic Neighbor Embedding*, um algoritmo de ML de redução de dimensionalidade, seguida da aplicação do algoritmo DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) para agrupamento das séries temporais com comportamento semelhante, independentemente da sua distribuição espacial. A abordagem de teste de hipóteses é então aplicada em seguida para estimar qual o melhor modelo temporal para cada PS. Os resultados obtidos demonstram que a abordagem seguida permite de forma eficaz o agrupamento de séries temporais com comportamento semelhante, sendo a mesma independente de fatores como a localização espacial e complexidade temporal.

4. Metodologias

Dispondo de bases de dados de PSs, derivados do processamento de séries temporais de dados SAR, são conhecidas para cada PS as coordenadas geográficas - Latitude, Longitude e Altitude Elipsoidal - e os respetivos parâmetros multi-temporais, tais como a velocidade de subsidência/elevação e o respetivo desvio padrão, coerência temporal, deslocamento cumulativo, entre outros.

Cada PS encontra-se pré-classificado segundo as metodologias desenvolvidas por Bakon et al. (2017). Considerando a correlação espacial entre PSs, cada PS é classificado numa de três classes – *inliers*, *outliers*, e dispersores duvidosos – segundo um método sequencial envolvendo seis técnicas: *Density-based Spatial Clustering of Applications with noise* (DBSCAN), *Principal Component Analysis* (PCA), *Graph Theory Grouping*, diagramas de *Voronoi*, desvio mediano absoluto (*median absolute deviation*, MAD), e o índice de *Jaccard* (*Jaccard index*). A classe de PSs *inliers* consiste no grupo de PSs a serem mantidos na base de dados, ao passo que os *outliers*, como o próprio nome indica, são os PSs a serem removidos. Os PSs duvidosos consistem em PSs que exigem melhor análise, não tendo sido possível a sua classificação exata pelas metodologias anteriores.

No âmbito da presente dissertação, pretende-se aplicar CNNs de forma a classificar cada PS numa das três classes anteriormente mencionadas. Tal como no estudo mencionado anteriormente, no processo de classificação foi tida em consideração a correlação espacial existente entre os PSs, baseada no facto de PSs vizinhos próximos exibirem (em princípio) características semelhantes. Assim, de forma a avaliar a influência do número de PSs vizinhos considerados na classificação final de cada PS, as metodologias apresentadas em seguida foram aplicadas individualmente considerando para cada PS os respetivos parâmetros multi-temporais bem como dos 3, 5, 10, e 15 PSs vizinhos mais próximos, com base na distância euclidiana, conforme ilustrado na Figura 12. PSs *outliers* estão coloridos a vermelho, PSs *inliers* a verde e PSs duvidosos a amarelo.



Figura 12a



Figura 12b



Figura 12c

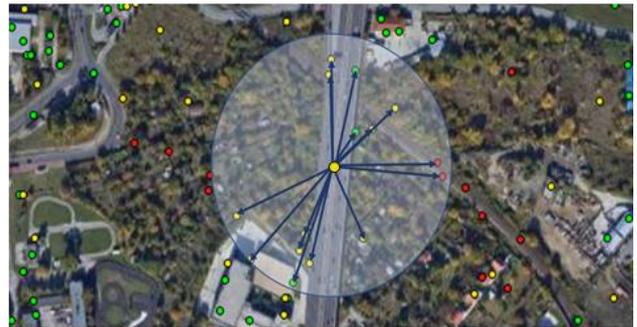


Figura 12d

Figura 12. Figuras ilustrativas da abordagem seguida, considerando na classificação de cada PS os 3, 5, 10, e 15 PSs vizinhos mais próximos (Figuras 12a a 12d, respetivamente), com base na distância euclidiana.

As metodologias seguidas são divididas em duas partes: preparação dos dados e implementação, treino e avaliação dos modelos de classificação.

4.1. Preparação dos dados

As etapas de preparação dos dados encontram-se ilustradas na Figura 13.

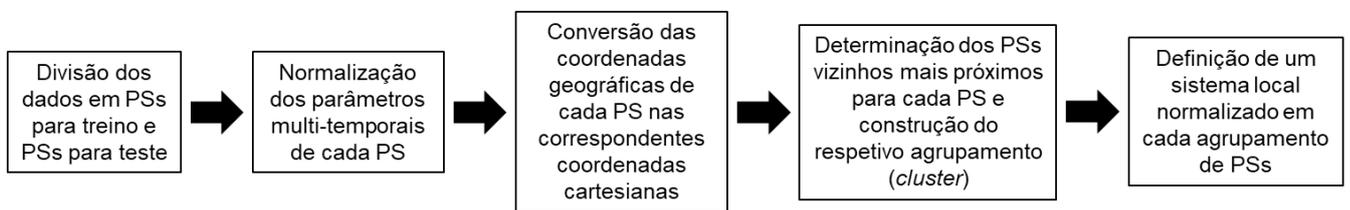


Figura 13. Esquematisação das etapas envolvidas na preparação dos dados.

Numa primeira fase, procedeu-se à divisão dos dados em PSs para treino e PSs para teste, seguida de uma operação de normalização dos parâmetros multi-temporais de cada PS. A operação de normalização consistiu no cálculo da média e do desvio padrão de cada parâmetro, considerando apenas os PSs de treino, seguida da respetiva normalização para cada PS, através da equação (7):

$$X_{i,norm} = \frac{X_i - \langle X \rangle}{\sigma_X} \quad (7)$$

onde X_i e $X_{i,norm}$ são, respetivamente, o valor do parâmetro X e o respetivo valor normalizado do i -ésimo PS, e $\langle X \rangle$ e σ_X são, respetivamente, a média e desvio padrão do parâmetro.

De forma a ser possível o cálculo da distância euclidiana entre PSs, procedeu-se à conversão das coordenadas geográficas de cada PS nas correspondentes coordenadas cartesianas, referentes ao elipsoide WGS84. São conhecidos os parâmetros semi-eixo, a , e inverso do achatamento, f^1 , do elipsoide WGS84, apresentados na Tabela 3, sendo possível o cálculo do semi-eixo menor, b , e, conseqüentemente, da excentricidade do elipsoide, e , e do raio de curvatura no primeiro vertical, N , dados, respetivamente, pelas equações (8)-(10) (no cálculo de N , o parâmetro φ corresponde à latitude geográfica de cada PS):

a (m)	6378137
f^1	298.257223563

Tabela 3. Semi-eixo maior (a) e inverso do achatamento (f^1) do elipsoide WGS84.

$$b = a * (1 - f) \quad (8)$$

$$e = \frac{\sqrt{a^2 - b^2}}{a} \quad (9)$$

$$N = \frac{a}{\sqrt{1 - e^2 * \sin^2(\varphi)}} \quad (10)$$

Conhecidos os parâmetros e , N e as coordenadas geográficas de cada PS, foram calculadas as respetivas coordenadas cartesianas, através das equações (11)-(13):

$$X = (N + h) * \cos(\varphi) * \cos(\lambda) \quad (11)$$

$$Y = (N + h) * \cos(\varphi) * \sin(\lambda) \quad (12)$$

$$Z = [N * (1 - e^2) + h] * \sin(\varphi) \quad (13)$$

onde (φ, λ, h) são, respetivamente, as coordenadas geográficas Latitude, Longitude e Altitude Elipsoidal, e (X, Y, Z) são as correspondentes coordenadas cartesianas.

Os passos seguintes consistiram na determinação para cada PS dos n PSs vizinhos mais próximos, com base na distância euclidiana, onde n é um número inteiro positivo variável igual a 3, 5, 10, e 15, e na definição do respetivo agrupamento (*cluster*) de PSs. Cada *cluster* consiste numa matriz bidimensional constituída na primeira linha pelas coordenadas cartesianas e parâmetros multi-temporais normalizados do PS a classificar, e nas restantes linhas pelas coordenadas cartesianas e parâmetros multi-temporais normalizados dos n PSs vizinhos mais próximos, por ordem crescente de distância. Todos os *clusters* resultantes foram em seguida agrupados numa matriz cúbica.

Na fase seguinte, foi implementado em cada *cluster* um sistema local, centrado no PS a classificar, através das equações (14)-(16):

$$x_i = X_i - X_{ref} \quad (14)$$

$$y_i = Y_i - Y_{ref} \quad (15)$$

$$z_i = Z_i - Z_{ref} \quad (16)$$

onde $(X_{ref}, Y_{ref}, Z_{ref})$ são as coordenadas cartesianas do PS a classificar, e (X_i, Y_i, Z_i) e (x_i, y_i, z_i) são, respetivamente, as coordenadas cartesianas do i -ésimo PS vizinho mais próximo, entre os n PSs vizinhos mais próximos considerados, e as correspondentes coordenadas cartesianas no sistema local. De forma semelhante à normalização dos parâmetros multi-temporais de cada PS descrita anteriormente, o passo seguinte consistiu na normalização de cada sistema local, através das equações (17)-(19):

$$x_{i,norm} = \frac{x_i - \langle x_{sl} \rangle}{\sigma_{xsl}} \quad (17)$$

$$y_{i,norm} = \frac{y_i - \langle y_{sl} \rangle}{\sigma_{ysl}} \quad (18)$$

$$z_{i,norm} = \frac{z_i - \langle z_{sl} \rangle}{\sigma_{zsl}} \quad (19)$$

onde (x_i, y_i, z_i) são as coordenadas cartesianas no sistema local do i -ésimo PS vizinho mais próximo, entre os n PSs vizinhos mais próximos considerados, e $(x_{i,norm}, y_{i,norm}, z_{i,norm})$ as correspondentes coordenadas cartesianas normalizadas; $(\langle x_{sl} \rangle, \langle y_{sl} \rangle, \langle z_{sl} \rangle)$ e $(\sigma_{xsl}, \sigma_{ysl}, \sigma_{zsl})$ são, respetivamente, o

valor médio e desvio padrão de cada coordenada cartesiana no sistema local, calculados iterativamente para cada *cluster* com base nas coordenadas dos n PSs vizinhos mais próximos considerados.

Relativamente à classe de cada PS, procedeu-se à codificação das mesmas usando a técnica *one-hot-encoding*. Desta forma, a classe de cada PS é transformada num vetor com três elementos nulos, correspondentes às três classes de PSs, exceto no índice relativo à classe, onde o respetivo elemento toma o valor 1.

4.2. Criação, treino e avaliação dos modelos de classificação

Para cada número de PSs vizinhos considerados na classificação de cada PS mencionado anteriormente, foi construída uma CNN como modelo de classificação, conforme ilustrado na Figura 14 na página seguinte.

Tal como referido anteriormente, os dados de entrada (*input*) consistem numa matriz cúbica, considerando para cada PS os respetivos parâmetros multi-temporais normalizados e os correspondentes parâmetros dos n PSs vizinhos mais próximos num sistema local normalizado, centrado no PS a classificar.

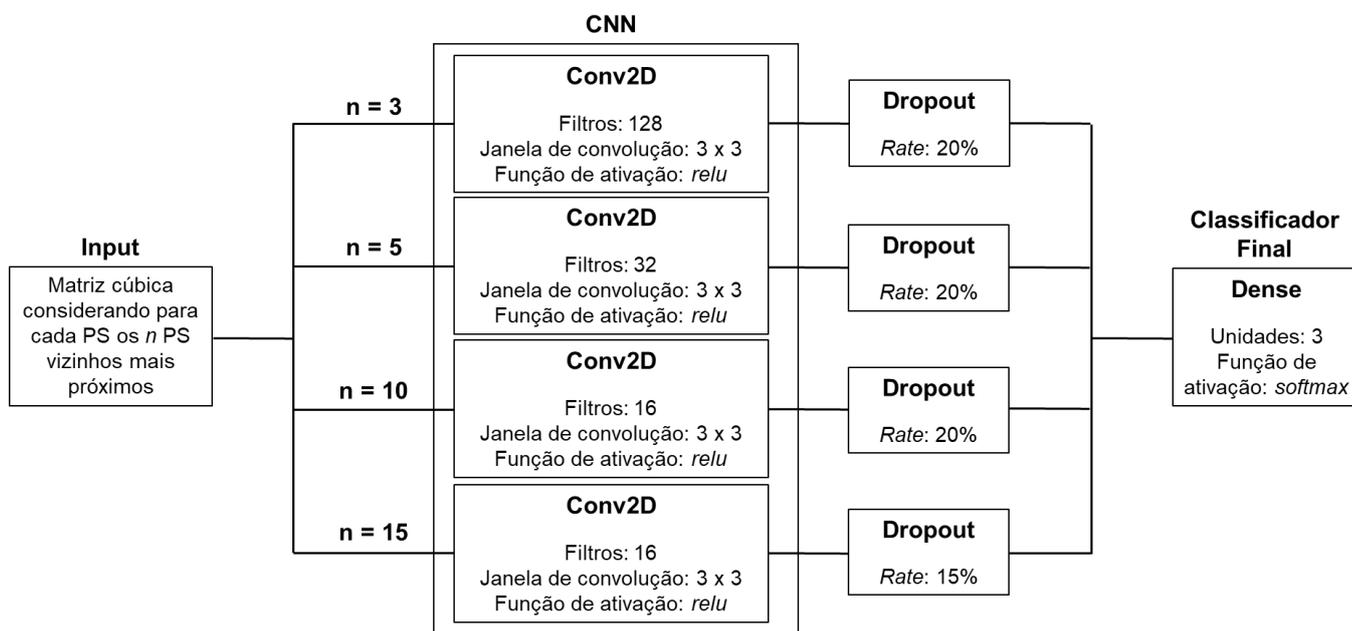


Figura 14. Modelos de classificação implementados, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS.

Cada CNN é constituída por uma camada *Conv2D*, esta última definida por uma janela de convolução de tamanho 3×3 e a função de ativação *relu*. O número de filtros de cada camada foi ajustado no processo de treino dos modelos. De forma a aumentar a capacidade de generalização

dos modelos, foi adicionada à saída de cada CNN uma camada de regularização *Dropout*, cujo parâmetro associado, *Dropout Rate*, foi igualmente ajustado no processo de treino dos modelos.

O classificador final consiste numa camada *Dense* com 3 unidades, correspondentes às três classes de PSs, e a função de ativação *softmax*. Assim, à saída, cada modelo de classificação prevê a probabilidade de um PS pertencer a cada uma das classes.

Na compilação dos modelos foi definida a função *rmsprop* como otimizador e a função *Categorical Crossentropy* como função *loss*. De forma a avaliar a eventual existência de sobreajustamento dos modelos de classificação, durante o processo de treino cada modelo foi avaliado numa base de dados de PSs de validação, esta última composta por 20% dos PSs de treino de cada classe definidos inicialmente, sendo os parâmetros de cada camada ajustados adequadamente.

Para efeitos de comparação, para cada modelo de classificação foi construído o respetivo modelo de base, que consiste em atribuir a cada PS a classe mais frequente entre os n PSs vizinhos mais próximos considerados. No caso de não haver consenso no descrito anteriormente, é atribuída uma classe aleatória entre as classes mais frequentes.

Na fase de treino dos modelos de classificação, foram testadas várias arquiteturas de modelos diferentes, consistindo na adição de camadas *Conv2D* adicionais, *MaxPooling2D*, *AveragePooling2D*, e camadas *Dense*. Verificou-se que a adição das mesmas resultava numa perda de eficiência ou não acrescentavam valor aos modelos. Na definição do número de épocas (iterações) de treino, foi determinada a época a partir da qual a aprendizagem do modelo estagnava, ou seja, a partir da qual o acrescento de épocas de treino adicionais não melhorava a eficiência de classificação dos modelos. A avaliação dos resultados obtidos foi realizada segundo cinco métricas: percentagem de classificações incertas, exatidão, precisão, *recall*, e o F_1 Score.

A exatidão consiste no cálculo da percentagem de PSs bem classificados. No entanto, devido à eventual existência de um enviesamento do número de PSs por classe, foram analisadas três métricas adicionais, que permitem uma melhor análise da eficiência dos algoritmos de classificação. Tanto a precisão, *recall* e o F_1 Score foram calculados individualmente para cada classe de PSs, sendo no final calculado o respetivo valor médio para cada métrica.

Como exemplo, considerando a classe de PSs *inliers*, o cálculo da precisão tem por base responder à seguinte questão: de todos os PSs classificados como *inliers* quantos de facto foram bem classificados? O *recall* tem por base responder a uma questão semelhante: de todos os PSs *inliers* disponíveis, quantos foram bem classificados? O F_1 Score, por fim, consiste na média harmónica entre a precisão e o *recall*, dado pela equação (20):

$$F_1 \text{ Score} = 2 * \frac{\text{Precisão} * \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (20)$$

5. Resultados e Discussão

5.1. Resultados relativos à base de dados de dispersores permanentes da cidade de Bratislava

Dispõe-se de uma base de dados de PSs contendo 107624 PSs relativos à cidade de Bratislava, Eslováquia. Para cada PS, são conhecidas as respectivas coordenadas geográficas e a Altitude Elipsoidal, a classe associada, a altitude do MDT usado no processamento InSAR e o respetivo desvio padrão, e os seguintes parâmetros multi-temporais:

- velocidade de subsidência/elevação e o respetivo desvio padrão;
- deslocamento cumulativo;
- coerência temporal e o respetivo inverso.

Os mapas de velocidade de subsidência/elevação e da coerência temporal encontram-se ilustrados nas Figuras 15 e 16, respetivamente.

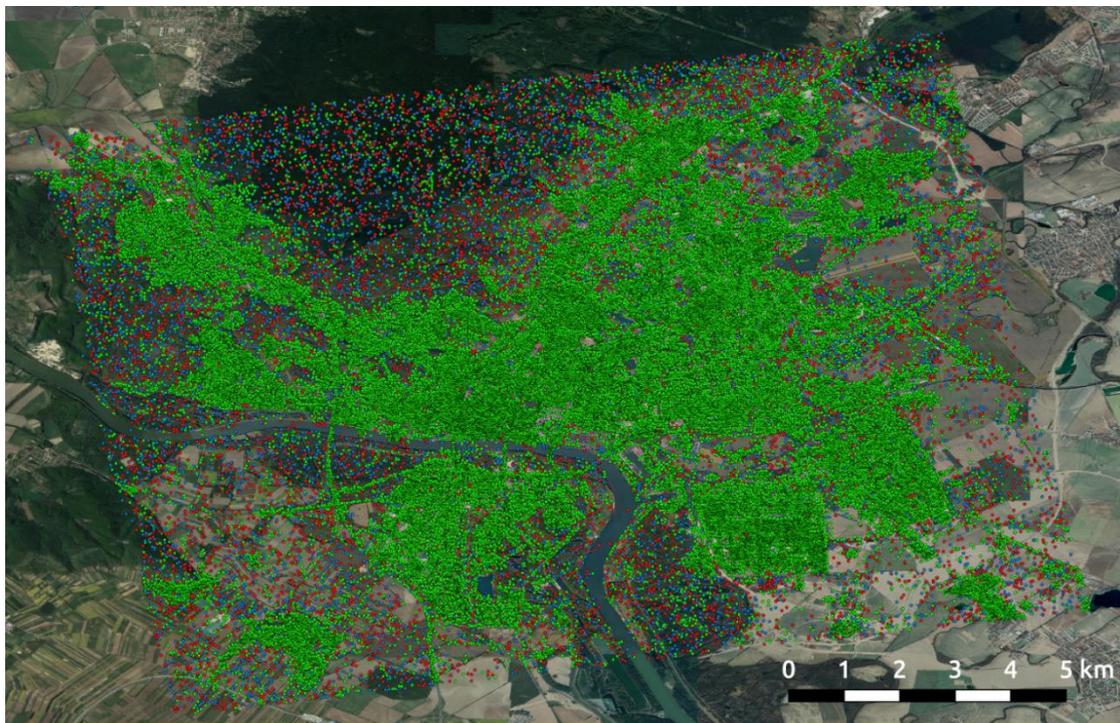


Figura 15. Mapa de velocidade de subsidência/elevação (mm/ano) da base de dados de PSs da cidade de Bratislava. PSs com velocidade inferior a -5 mm/ano estão coloridos a vermelho, entre -5 mm/ano e +5 mm/ano, inclusive, encontram-se coloridos a verde, e com velocidade superior a +5 mm/ano encontram-se coloridos a azul.



Figura 16. Mapa da coerência temporal da base de dados de PSs da cidade de Bratislava. PSs cuja coerência temporal é inferior a 0.5 encontram-se coloridos a vermelho, igual ou superior a 0.5 e inferior a 0.7 encontram-se coloridos a amarelo, e cuja coerência é igual ou superior a 0.7 encontram-se coloridos a verde.

Na Figura 17 encontra-se ilustrado o mapa de classes, na qual PSs *outliers* estão coloridos a vermelho, PSs *inliers* a verde e PSs duvidosos a amarelo.

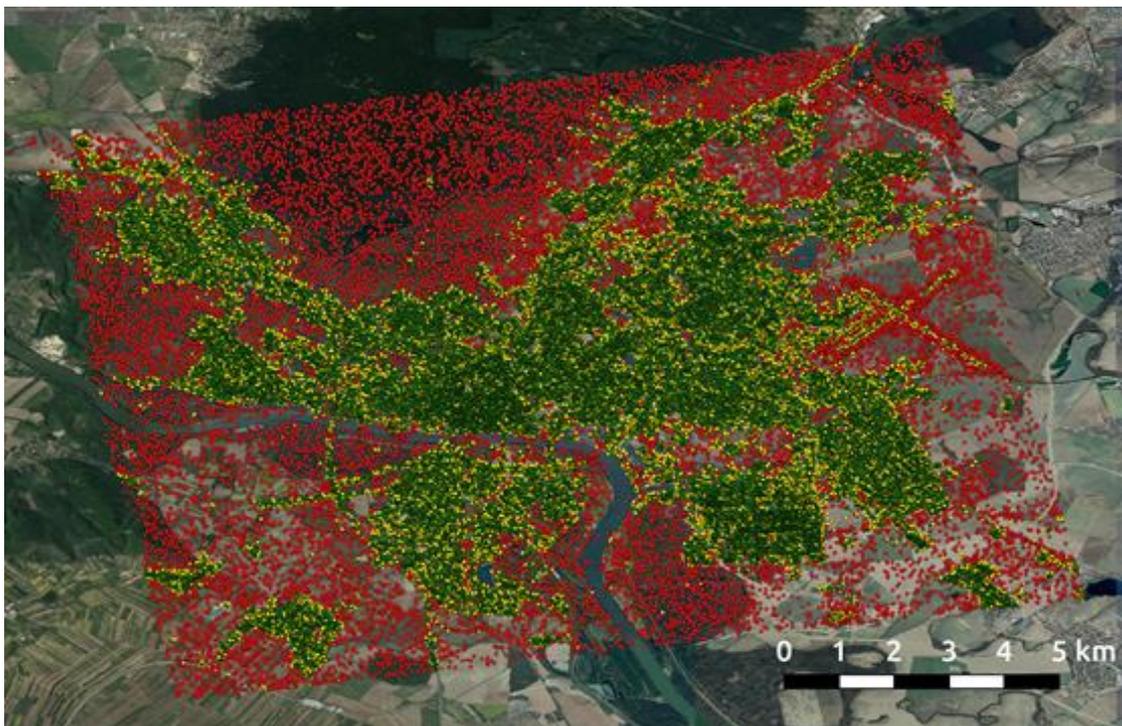


Figura 17. Mapa de classes da base de dados de PSs da cidade de Bratislava.

Pela análise das Figuras 15 e 16, verifica-se que a cidade de Bratislava, capital da Eslováquia, é uma região estável, circundada por zonas de floresta densa e de exploração agrícola. Conforme esperado, os PSs com elevada coerência temporal (superior a 0.7) situam-se nas regiões urbanas. O mapa de classes da Figura 17 sustenta o já referido: os PSs *inliers* situam-se maioritariamente nas zonas urbanas, ao passo que PSs *outliers* se situam nas zonas circundantes da cidade. É possível também verificar que os PSs duvidosos, que exigem melhor análise, se encontram maioritariamente em zonas urbanas, havendo uma pequena percentagem dos mesmos em zonas rurais.

De forma a analisar a capacidade de generalização dos modelos de classificação e a possível existência de um enviesamento dos resultados, foi aplicada uma abordagem de validação cruzada, consistindo em 4 divisões da base de dados de PSs (*4-fold cross-validation*) em PSs para treino e PSs para teste, conforme ilustrado nas Figuras 18a-18d.

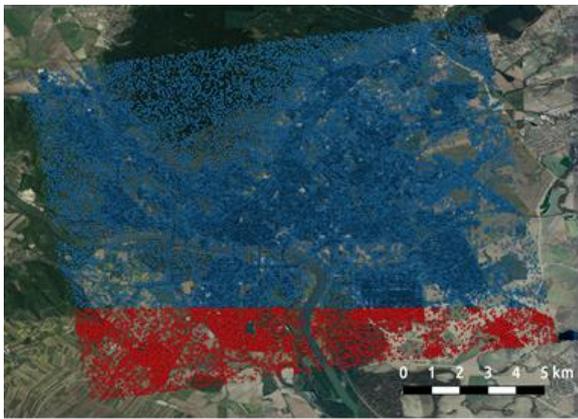


Figura 18a

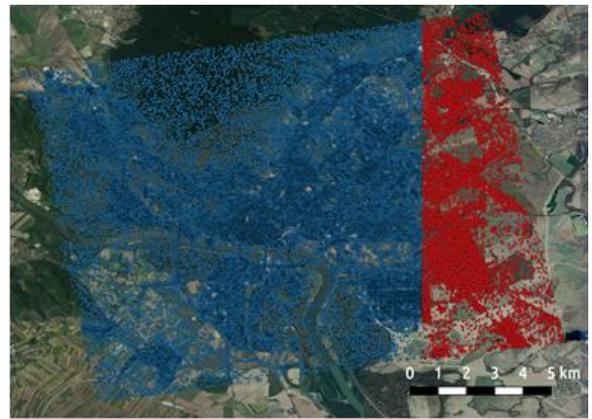


Figura 18b

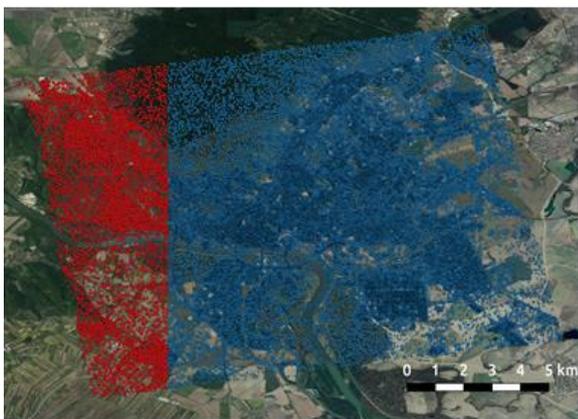


Figura 18c

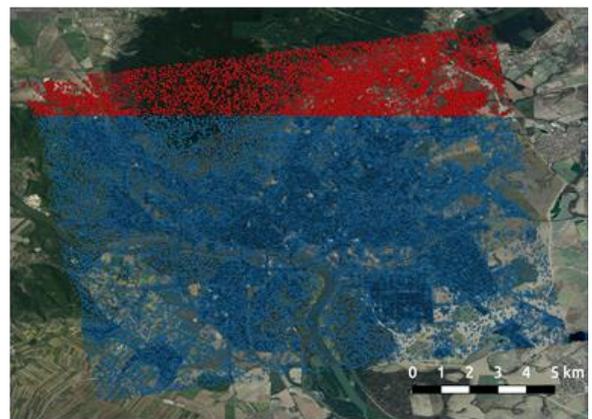


Figura 18d

Figura 18. Abordagem de validação cruzada aplicada na base de dados de PSs da cidade de Bratislava: considerando os paralelos 48.1125 °N e 48.1989 °N (Figuras 18a e 18d, respetivamente); considerando os meridianos 17.1915 °E e 17.0724 °E (Figuras 18b e 18c, respetivamente). PSs para treino estão coloridos a azul e PSs para teste a vermelho. PSs para validação correspondem a 20% dos PSs de treino.

As divisões da base de dados foram realizadas de forma a que tanto o conjunto de PSs para treino como o conjunto de PSs para teste possuíssem PSs relativos às duas zonas mencionadas anteriormente, urbanas e rurais. O número de PSs para cada um dos conjuntos de dados (Treino, Validação e Teste) e a respetiva percentagem para cada uma das divisões de dados descritas anteriormente encontram-se ilustradas na Tabela 4.

Divisão	Treino		Validação		Teste	
	Nº de PSs	Percentagem	Nº de PSs	Percentagem	Nº de PSs	Percentagem
1	75844	70	18961	18	12819	12
2	71003	66	17750	16	18871	18
3	71627	67	17906	17	18091	17
4	74668	69	18666	17	14290	13

Tabela 4. Número de PSs e a respetiva percentagem dos conjuntos de PSs de Treino, Validação e Teste resultantes da divisão da base de dados de PSs da cidade de Bratislava.

Conforme mencionado anteriormente, cada modelo de classificação retorna a probabilidade de um PS pertencer a cada uma das classes. Para uma correta avaliação dos resultados obtidos, foi necessário proceder à conversão das probabilidades fornecidas pelos mesmos em classes. Para cada PS, considerando a classe com probabilidade máxima de o mesmo pertencer, P_{max} , e um valor de *threshold*, $t \in [0,1]$, no caso de P_{max} ser maior ou igual a t , então é associado ao respetivo PS a classe correspondente e avaliada a correspondente classificação; em caso contrário, a classificação do respetivo PS é considerada incerta e o mesmo devidamente anotado.

No âmbito do presente trabalho, a estratégia para a determinação do valor t a aplicar à saída dos modelos de classificação consiste em minimizar o número de PSs *inliers* classificados como *outliers* e vice-versa, visto ser este o caso crucial a evitar (idealmente pretende-se que ambos tomem o valor zero). Para a determinação do valor global t a aplicar, para cada modelo criado foi avaliado o descrito anteriormente em função de t , considerando as previsões nos dados de treino e de validação, tendo sido determinado um valor global de 86% de *threshold*. Os gráficos correspondentes, para as divisões de dados 1 a 4 ilustradas na Figura 18, encontram-se, respetivamente, nas Figuras 26 a 29 em anexo, em função do número de PSs vizinhos considerados.

No capítulo 8.1 encontram-se apresentados os resultados relativamente às divisões de dados 1 a 4. Para cada divisão de dados são apresentados, em função do número de PSs vizinhos considerados na classificação de cada PS, a percentagem de classificações incertas por classe, matrizes de confusão e mapas de classificação representando os três tipos de classificações: incertas, corretas e erradas. De forma a permitir uma melhor visualização das classificações incertas

e erradas, neste último caso são apresentados uma segunda versão dos mapas de classificação onde as classificações corretas se encontram omitidas.

Os resultados médios obtidos, em função do número de PSs vizinhos considerados, encontram-se ilustrados na Tabela 5, em comparação com os resultados dos modelos de base.

n	Modelo	Classificações Incertas (%)	Exatidão (%)	Precisão (%)	Recall (%)	F ₁ Score (%)
3	Base	-	76.92 ± 2.29	71.94 ± 0.90	70.49 ± 0.99	70.65 ± 0.97
	CNN	13.04 ± 1.39	98.13 ± 0.38	97.42 ± 0.43	97.35 ± 0.68	97.38 ± 0.55
5	Base	-	76.82 ± 2.23	71.72 ± 0.88	69.82 ± 1.09	69.89 ± 1.07
	CNN	11.49 ± 1.31	98.30 ± 0.35	97.67 ± 0.43	97.64 ± 0.46	97.65 ± 0.44
10	Base	-	76.26 ± 2.27	70.77 ± 1.08	68.48 ± 1.33	68.21 ± 1.40
	CNN	11.39 ± 2.29	98.41 ± 0.26	97.93 ± 0.39	97.56 ± 0.38	97.74 ± 0.34
15	Base	-	75.81 ± 2.06	70.05 ± 0.80	67.31 ± 1.19	66.62 ± 1.31
	CNN	11.55 ± 0.53	98.22 ± 0.29	97.57 ± 0.21	97.59 ± 0.63	97.57 ± 0.41

Tabela 5. Resultados médios dos modelos de classificação, em função do número de PSs vizinhos considerados na classificação de cada PS, relativos à base de dados de PSs da cidade de Bratislava, em comparação com os resultados dos modelos de base. O F₁ Score encontra-se apresentado no intervalo [0,100].

A percentagem média de classificações incertas por classe, em função do número de PSs vizinhos considerados, encontra-se representada na Tabela 6.

n	PSs outliers	PSs inliers	PSs duvidosos
3	5.16 ± 0.64	2.39 ± 0.28	5.50 ± 0.89
5	4.39 ± 1.20	2.11 ± 0.33	4.99 ± 0.21
10	3.47 ± 0.85	2.11 ± 0.77	5.82 ± 0.93
15	3.80 ± 0.56	2.85 ± 0.77	4.90 ± 1.12

Tabela 6. Percentagem média de classificações incertas por classe, em função do número de PSs vizinhos considerados na classificação de cada PS, relativa à base de dados de PSs da cidade de Bratislava.

Na Tabela 7 encontra-se apresentado o desvio padrão médio da distância aos *n* PSs vizinhos mais próximos considerados, para os três tipos de classificações.

n	Classificações	Classificações	Classificações
	Erradas	Corretas	Incertas
3	11.20 ± 1.10	9.11 ± 0.75	9.49 ± 0.71
5	15.31 ± 0.99	13.29 ± 1.07	13.66 ± 0.52
10	21.56 ± 0.74	20.28 ± 1.51	19.63 ± 0.92
15	26.94 ± 0.99	25.43 ± 1.97	24.57 ± 0.63

Tabela 7. Desvio padrão médio, em metros, da distância aos n PSs vizinhos mais próximos considerados, para os três tipos de classificações, relativo à base de dados de PSs da cidade de Bratislava.

Pela análise da Tabela 5, verifica-se que os melhores resultados são obtidos considerando na classificação de cada PS os 5 PSs vizinhos mais próximos. Das três classes de PSs, aquela que apresenta um maior número de percentagem de classificações incertas é a classe de PSs duvidosos, seguida da classe de PSs *outliers*, que apresenta valores semelhantes à primeira classe. Pela análise da Tabela 7, verifica-se que não existe uma dispersão significativa da distância de cada PS aos PSs vizinhos mais próximos para os três tipos de classificações.

A análise das matrizes de confusão apresentadas no capítulo 8.1 permitem verificar que o aumento do número de PSs vizinhos considerados na classificação de cada PS permite, em média, classificar corretamente um número crescente de PSs *outliers*; no entanto, em contrapartida, verifica-se que o número de PSs *inliers* corretamente classificados e dos PSs duvidosos diminui, verificando-se que tal é válido até um número máximo de 10 PSs vizinhos mais próximos considerados. Desta forma, no processo de classificação é necessário chegar a um compromisso no número de PSs vizinhos considerados, verificando-se, no caso da base de dados disponível, que, como já mencionado anteriormente, tal consiste em considerar os 5 PSs vizinhos mais próximos na classificação de cada PS.

Na Tabela 8 encontra-se apresentada a distância média ao n -ésimo PS vizinho mais próximo para os três tipos de classificações, em função do número de PSs vizinhos considerados.

n	Classificações	Classificações	Classificações
	Erradas	Corretas	Incertas
3	61.78 ± 4.25	51.18 ± 4.92	53.44 ± 3.59
5	81.50 ± 2.55	65.86 ± 6.12	67.46 ± 2.08
10	105.86 ± 3.47	93.21 ± 7.63	90.57 ± 4.36
15	129.16 ± 4.81	114.08 ± 9.46	110.47 ± 3.72

Tabela 8. Distância média ao n -ésimo PS vizinho mais próximo para os três tipos de classificações, em função do número de PSs vizinhos considerados na classificação de cada PS, relativa à base de dados de PSs da cidade de Bratislava.

Considerando que os melhores resultados obtidos consistem em considerar os 5 PSs vizinhos mais próximos na classificação de cada PS, pela análise da Tabela 8 é possível deduzir que a abordagem descrita é aplicável até uma distância máxima entre o PS a classificar e o n -ésimo PS vizinho mais próximo de 65 metros.

Numa última fase, foi avaliada a importância de cada parâmetro na classificação final de cada PS. Para cada parâmetro, foram adicionados valores aleatórios ao mesmo, no intervalo $[-1,1]$, iterativamente para cada agrupamento de PSs, e o processo de classificação repetido. Devido à existência de correlações entre parâmetros, o descrito anteriormente foi realizado simultaneamente para as seguintes variáveis (previamente à repetição do processo de classificação):

- coordenadas cartesianas normalizadas no sistema local ($X + Y + Z$);
- altitude do MDT usado no processamento InSAR (HEIGHT DEM) e o respetivo desvio padrão (SIGMA HEIGHT);
- velocidade de subsidência/elevação (VEL) e o respetivo desvio padrão (SIGMA VEL);

No caso do parâmetro de deslocamento cumulativo (CUM DISP), o mesmo não apresenta uma correlação significativa com os restantes parâmetros, pelo que foi avaliado isoladamente. A coerência temporal (COHER) e o respetivo inverso (INV_COHER) apresentam uma relação diferente das anteriores. Neste caso, foram adicionados valores aleatórios à coerência e em seguida considerado o respetivo inverso para a segunda variável. Os resultados obtidos, considerando para cada PS os 5 PSs vizinhos mais próximos, avaliados na percentagem de classificações incertas, exatidão, precisão e *recall*, encontram-se representados nos gráficos das Figuras 19 a 22, respetivamente. Os respetivos gráficos considerando para cada PS os 3, 10 e 15 PSs vizinhos mais próximos, encontram-se representados, respetivamente, nas Figuras 30 a 33, Figuras 34 a 37 e Figuras 38 a 41 em anexo.

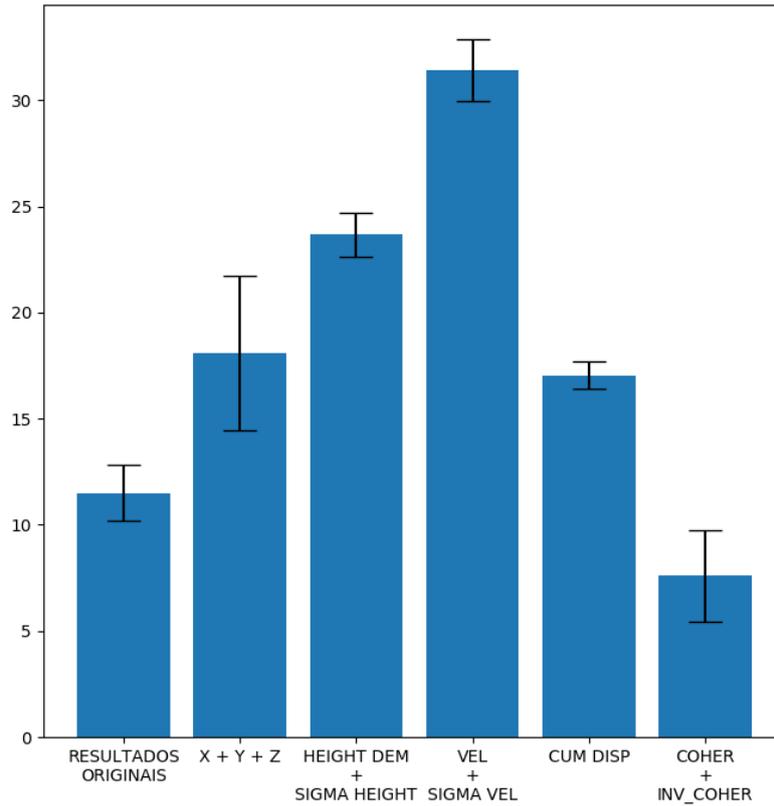


Figura 19. Percentagem média de classificações incertas e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 5 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

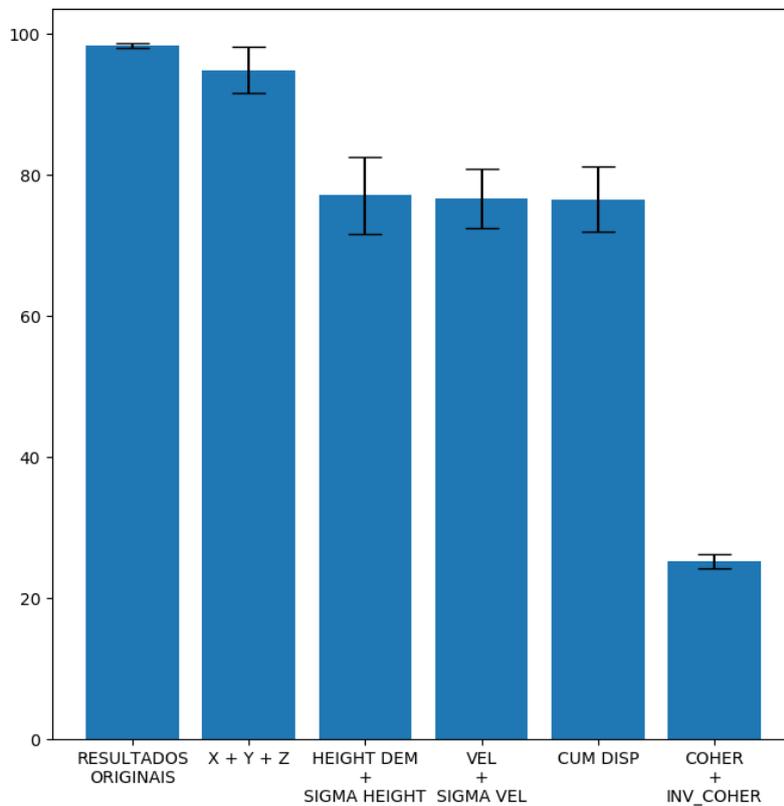


Figura 20. Exatidão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 5 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

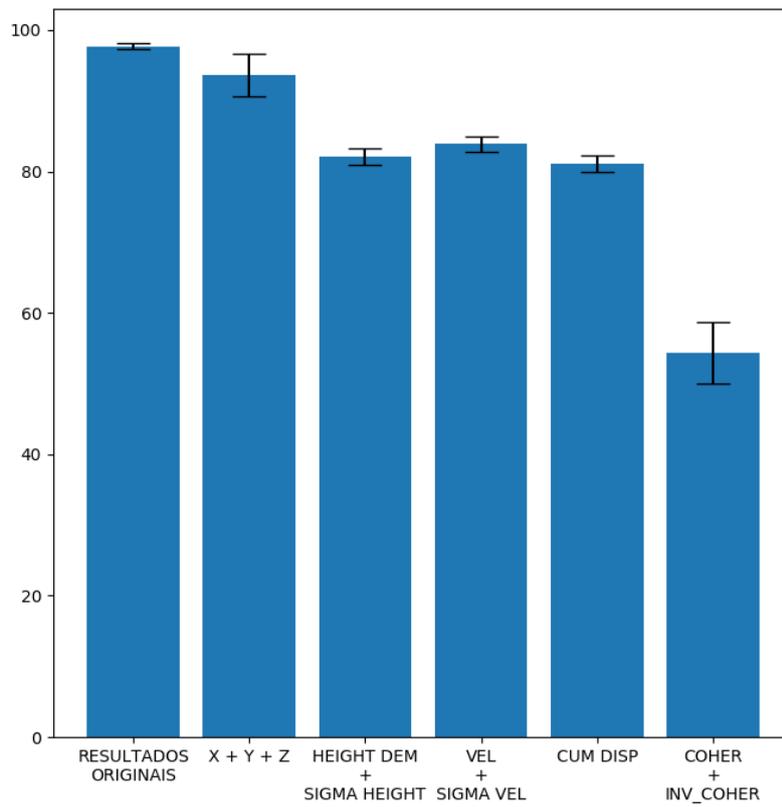


Figura 21. Precisão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 5 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

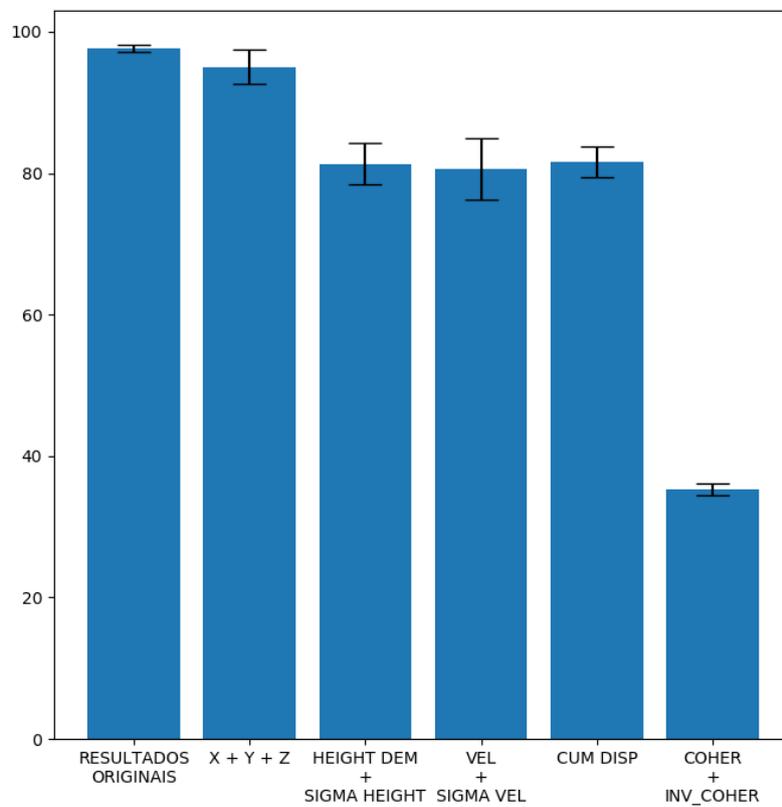


Figura 22. Recall médio e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 5 PSs vizinhos mais próximos, relativo à base de dados de PSs da cidade de Bratislava.

Pela análise dos gráficos das Figuras 19 a 22, verifica-se que os parâmetros mais relevantes no processo de classificação são a altitude do MDT usado no processamento InSAR e o respetivo desvio padrão, velocidade de subsidência/elevação e o respetivo desvio padrão, deslocamento cumulativo e a coerência temporal e o respetivo inverso. De todos os parâmetros, a coerência temporal e o respetivo inverso revelaram ser um parâmetro bastante relevante, na medida em que a introdução de valores aleatórios para a mesma em cada agrupamento de PSs resultou em fracos resultados de classificação.

5.2. Resultados relativos à base de dados de dispersores permanentes dos subúrbios de Prievidza

Numa segunda fase, os modelos de classificação criados foram avaliados numa base de dados de PSs relativa a uma área geográfica diferente, contendo 1388 PSs cobrindo as áreas de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, Eslováquia. Os mapas de velocidade de subsidência/elevação, coerência temporal, e o mapa de classes encontram-se ilustrados nas Figuras 23 a 25, respetivamente.

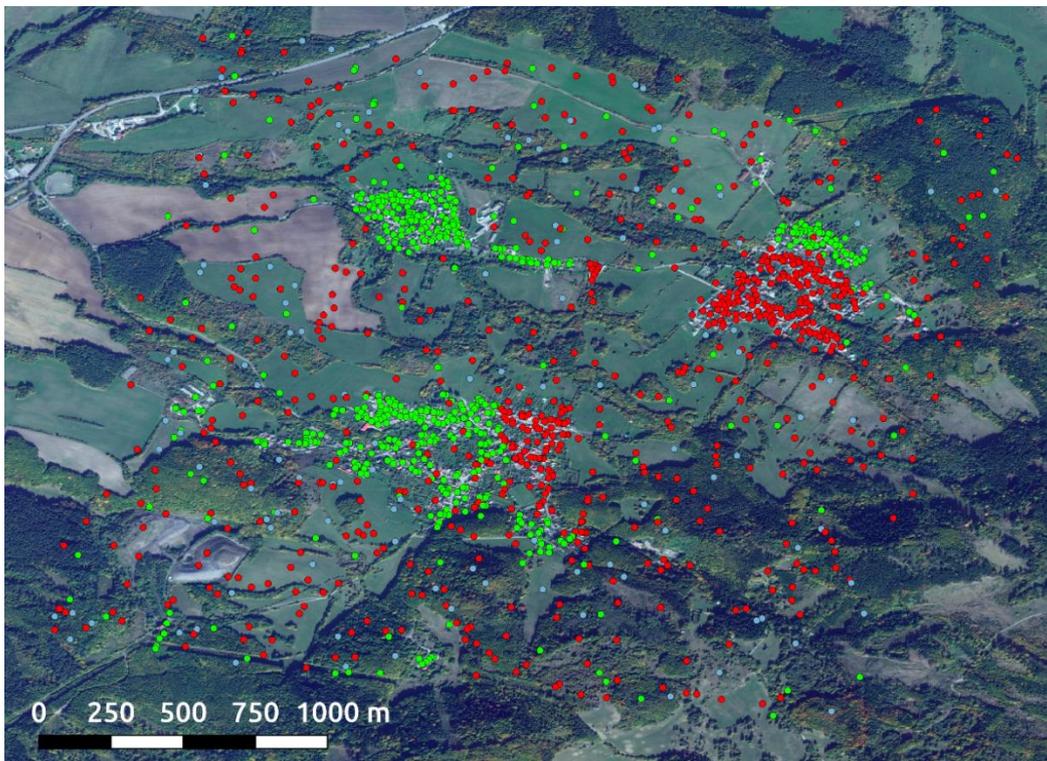


Figura 23. Mapa de velocidade de subsidência/elevação (mm/ano) da base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza. PSs com velocidade inferior a -5 mm/ano estão coloridos a vermelho, entre -5 mm/ano e +5 mm/ano, inclusive, encontram-se coloridos a verde, e com velocidade superior a +5 mm/ano encontram-se coloridos a azul.

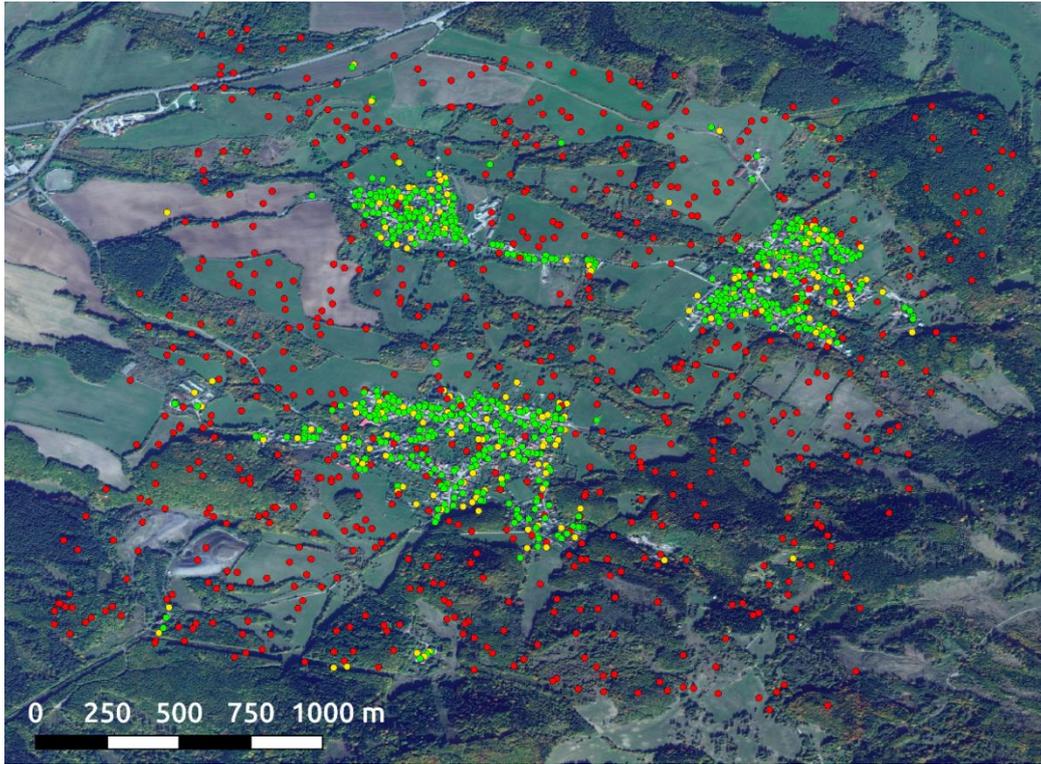


Figura 24. Mapa da coerência temporal da base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza. PSs cuja coerência temporal é inferior a 0.5 encontram-se coloridos a vermelho, igual ou superior a 0.5 e inferior a 0.7 encontram-se coloridos a amarelo, e cuja coerência é igual ou superior a 0.7 encontram-se coloridos a verde.

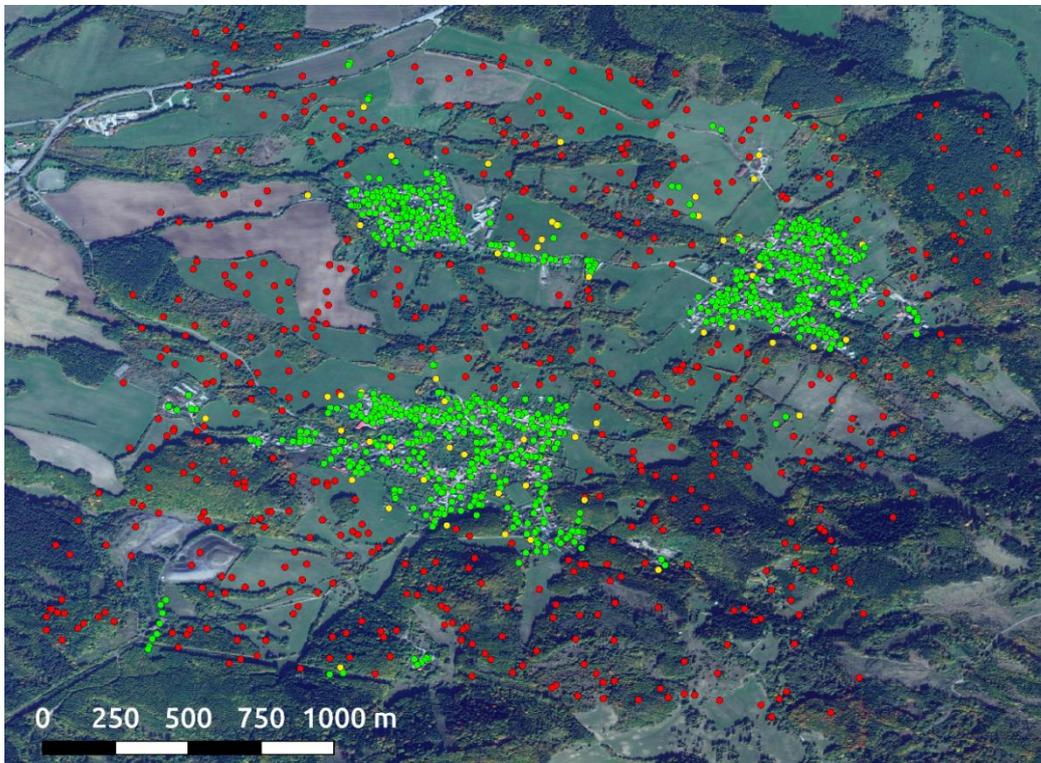


Figura 25. Mapa de classes da base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza. Os PSs estão coloridos da mesma forma que no mapa de classes da base de dados de PSs da cidade de Bratislava (Figura 17).

Comparativamente à cidade de Bratislava, esta segunda região é de muito menor extensão, sendo maioritariamente uma região rural composta pelas três zonas urbanas mencionadas anteriormente. Pela análise da Figura 23, verifica-se que toda a região se encontra em subsidência, incluindo parte das três regiões urbanas. Esta situação problemática de deformação é conhecida, sendo causada maioritariamente por deslizamentos de terra e extração mineira no subsolo. A análise das Figuras 24 e 25 permitem verificar que os PSs *outliers* se encontram maioritariamente nas regiões rurais, sendo caracterizados por baixa coerência temporal, e os PSs *inliers* encontram-se maioritariamente localizados nas três regiões urbanas, sendo caracterizados por elevada coerência temporal (superior a 0.7).

Tal como na apresentação dos resultados da base de dados de PSs anterior, no capítulo 8.2 encontram-se apresentadas, respetivamente para as divisões de dados 1 a 4 e em função do número de PSs vizinhos considerados na classificação de cada PS, as tabelas com a percentagem de classificações incertas por classe, matrizes de confusão e mapas de classificação.

Os resultados médios obtidos, em função do número de PSs vizinhos considerados, encontram-se ilustrados na Tabela 9, em comparação com os resultados dos modelos de base.

n	Modelo	Classificações Incertas (%)	Exatidão (%)	Precisão (%)	Recall (%)	F ₁ Score (%)
3	Base	-	91.5	64.32	63.65	63.5
	CNN	17.49 ± 0.39	53.01 ± 2.66	65.23 ± 0.96	53.14 ± 5.80	42.02 ± 0.53
5	Base	-	91.35	73.75	65.29	66.26
	CNN	17.62 ± 1.43	54.45 ± 1.40	63.76 ± 0.77	49.42 ± 2.11	40.87 ± 0.30
10	Base	-	89.77	60.25	61.89	60.59
	CNN	23.90 ± 7.27	58.86 ± 6.03	60.75 ± 1.80	43.11 ± 1.91	39.89 ± 2.53
15	Base	-	88.47	59.36	60.96	60.08
	CNN	21.40 ± 2.45	54.17 ± 3.78	60.34 ± 2.80	44.66 ± 5.06	37.60 ± 1.33

Tabela 9. Resultados médios dos modelos de classificação, em função do número de PSs vizinhos considerados na classificação de cada PS, relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, em comparação com os resultados dos modelos de base. O F₁ Score encontra-se apresentado no intervalo [0,100].

A percentagem média de classificações incertas por classe, em função do número de PSs vizinhos considerados, encontra-se representada na Tabela 10. Na Tabela 11 encontra-se apresentado o desvio padrão médio da distância aos *n* PSs vizinhos mais próximos considerados, para os três tipos de classificações.

n	PSs outliers	PSs inliers	PSs duvidosos
3	2.48 ± 1.40	13.67 ± 1.45	1.34 ± 0.29
5	1.21 ± 0.36	15.31 ± 1.67	1.10 ± 0.14
10	0.94 ± 0.39	21.92 ± 6.78	1.05 ± 0.38
15	1.59 ± 1.28	18.73 ± 3.23	1.08 ± 0.27

Tabela 10. Percentagem média de classificações incertas por classe, em função do número de PSs vizinhos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza.

n	Classificações	Classificações	Classificações
	Erradas	Corretas	Incertas
3	7.79 ± 0.35	17.08 ± 0.35	9.54 ± 1.25
5	11.64 ± 0.23	25.71 ± 0.08	13.06 ± 0.73
10	19.95 ± 1.79	40.67 ± 0.82	17.22 ± 0.87
15	24.53 ± 0.64	52.23 ± 0.71	24.08 ± 3.26

Tabela 11. Desvio padrão médio, em metros, da distância aos n PSs vizinhos mais próximos considerados, para os três tipos de classificações, relativo à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza.

Na Tabela 12 encontra-se apresentada a distância média ao n -ésimo PS vizinho mais próximo, para os três tipos de classificações, em função do número de PSs vizinhos considerados.

n	Classificações	Classificações	Classificações
	Erradas	Corretas	Incertas
3	40.09 ± 2.54	106.21 ± 2.15	53.12 ± 7.74
5	52.85 ± 0.87	137.36 ± 0.53	60.98 ± 3.51
10	84.64 ± 7.29	193.79 ± 3.87	75.75 ± 3.84
15	103.48 ± 2.33	239.91 ± 3.46	103.67 ± 14.69

Tabela 12. Distância média ao n -ésimo PS vizinho mais próximo, para os três tipos de classificações, em função do número de PSs vizinhos considerados na classificação de cada PS, relativa à base de dados da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza.

Pela análise dos resultados obtidos, verifica-se que nesta segunda região os resultados não foram satisfatórios, na medida em que os mesmos são inferiores aos resultados obtidos pelos modelos de base. Tais resultados insatisfatórios devem-se maioritariamente a uma classificação incorreta e incerta de PSs *inliers*, relativos às três regiões urbanas mencionadas anteriormente, havendo uma elevada percentagem de PSs *inliers* classificados como PSs duvidosos. Por outro lado, os PSs *outliers* relativos a esta região foram maioritariamente bem classificados.

Uma razão provável para tais resultados pode dever-se aos diferentes padrões de deformação entre as duas regiões. A região da cidade de Bratislava, na qual os modelos de classificação foram treinados, é uma região estável, ao passo que a segunda região nos subúrbios de Prievidza é caracterizada por um padrão de subsidência geral, muito diferente da primeira região.

6. Conclusão

As técnicas DInSAR usando PSs são atualmente bem estabelecidas e fundamentadas, permitindo um estudo dos padrões de deformação da superfície terrestre de uma forma contínua e global. Existem, no entanto, vários fatores indutores de erro que originam PSs *outliers*, sendo necessária a posterior deteção e remoção dos mesmos, o que, devido à frequente existência de um número excessivamente alto de PSs a analisar, torna tal tarefa computacionalmente exigente.

No âmbito da presente dissertação, foi estudada a possibilidade de automatização da deteção de PSs *outliers*, derivados do processamento de séries temporais de dados SAR, usando algoritmos de DL, nomeadamente através da implementação de CNNs. Com base nos resultados obtidos em duas bases de dados de PSs independentes, é possível concluir que, de uma forma geral, tal é possível, sendo o tempo de execução da abordagem proposta curto. A tarefa que exige um esforço computacional maior, embora mesmo assim aceitável, consiste na determinação dos PSs vizinhos mais próximos para cada PS, sendo dependente do volume de PSs e do número de PSs vizinhos a considerar. Todas as restantes tarefas apresentaram um tempo de execução desprezável. Na aplicação de algoritmos de DL no problema em causa, certos fatores devem ser, no entanto, considerados.

O primeiro fator a considerar (e o mais relevante) é relativo à classe de PSs duvidosos, que apresentam baixa coerência temporal. Conforme constatado por Bakon et al. (2017), é possível a existência de PSs que apresentam baixa coerência temporal sem os mesmos serem *outliers*. Assim, tal classe compreende PSs com características semelhantes tanto a PSs *inliers* como PSs *outliers*, o que, conforme se verificou, pode dificultar a tarefa de classificação. Desta forma, para tratar tal problema são propostas duas soluções.

A primeira solução consiste em descartar os PSs pertencentes a esta classe, considerando apenas PSs *inliers* e *outliers*. Porém, tal significaria uma redução da densidade e número de PSs disponíveis para treino, o que leva à segunda solução proposta, que consiste em reclassificar tal classe de PSs em PSs *inliers* e PSs *outliers*. Desta forma, o problema de classificação em questão seria puramente binário, sendo possível a eventual obtenção de resultados mais eficientes do que aqueles que foram obtidos.

Relativamente ao segundo fator a considerar, verificou-se que a aplicabilidade dos modelos de classificação na deteção de PSs *outliers* é eventualmente limitada pelo padrão de deformação existente na região onde os mesmos foram treinados. Assim, no caso geral, dispendo de uma base de dados de PSs para classificação, é necessário conhecer *a priori* o padrão de deformação da respetiva região, aplicando o modelo de classificação que melhor se adequa à mesma. Este fator exige, no entanto, melhor investigação.

Por fim, a densidade de PSs poderá também ser um fator a considerar. No caso da base de dados de PSs da cidade de Bratislava, os melhores resultados correspondem a considerar na

classificação de cada PS os 5 PSs vizinhos mais próximos, ao passo que na base de dados de PSs dos subúrbios de Priedza os melhores resultados obtidos, embora inferiores aos respetivos resultados dos modelos de base, foram conseguidos considerando os 3 PSs vizinhos mais próximos. Assim, no caso geral, para além de possuir modelos de classificação treinados em várias regiões com padrões de deformação distintos, é, eventualmente, também necessário possuir diferentes versões dos mesmos, considerando na classificação de cada PS diferentes números de PSs vizinhos. À semelhança do ponto anterior, este fator exige, no entanto, ser melhor estudado.

De forma a dar continuidade ao presente trabalho, outras abordagens poderão ser consideradas, tais como a exploração adicional de diferentes algoritmos de classificação, como é o caso de DTs, RFs, entre outros, e a construção das curvas de aprendizagem (*learning curves*). Tais curvas permitiriam verificar de que forma o processo de classificação beneficiaria do aumento de dados de treino, sendo possível determinar um eventual limite máximo de dados de treino necessário, a partir do qual considerar mais dados não teria um efeito significativo na eficiência dos modelos de classificação.

7. Referências

- Alpers, W. R., Ross, D. B., & Rufenach, C. L. (1981). On the detectability of ocean surface waves by real and synthetic aperture radar. *Journal of Geophysical Research: Oceans*, 86(C7), 6481-6498. doi:10.1029/JC086iC07p06481
- Anantrasirichai, N., Biggs, J., Albino, F., & Bull, D. (2019a). The Application of Convolutional Neural Networks to Detect Slow, Sustained Deformation in InSAR Time Series. *Geophysical Research Letters*, 46(21), 11850-11858. doi:10.1029/2019gl084993
- Anantrasirichai, N., Biggs, J., Albino, F., & Bull, D. (2019b). A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets. *Remote Sensing of Environment*, 230, 111179. doi:<https://doi.org/10.1016/j.rse.2019.04.032>
- Anantrasirichai, N., Biggs, J., Albino, F., Hill, P., & Bull, D. (2018). Application of Machine Learning to Classification of Volcanic Deformation in Routinely Generated InSAR Data. *Journal of Geophysical Research: Solid Earth*, 123(8), 6592-6606. doi:10.1029/2018jb015911
- Arief, A. H., Strand, G.-H., Tveite, H., & Indahl, G. U. (2018). Land Cover Segmentation of Airborne LiDAR Data Using Stochastic Atrous Network. *Remote Sensing*, 10(6). doi:10.3390/rs10060973
- Askne, J., & Hagberg, J. O. (1993, 18-21 Aug. 1993). *Potential of interferometric SAR for classification of land surfaces*. Paper presented at the Proceedings of IGARSS '93 - IEEE International Geoscience and Remote Sensing Symposium.
- Bakon, M., Oliveira, I., Perissin, D., Sousa, J. J., & Papco, J. (2017). A Data Mining Approach for Multivariate Outlier Detection in Postprocessing of Multitemporal InSAR Results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(6), 2791-2798. doi:10.1109/JSTARS.2017.2686646
- Bamler, R., & Hartl, P. (1998). Synthetic aperture radar interferometry. *Inverse Problems*, 14(4), 1-54.
- Biggs, J., Burgmann, R., Freymueller, J. T., Lu, Z., Parsons, B., Ryder, I., . . . Wright, T. (2009). The postseismic response to the 2002 M 7.9 Denali Fault earthquake: constraints from InSAR 2003–2005. *Geophysical Journal International*, 176(2), 353-367. doi:10.1111/j.1365-246X.2008.03932.x
- Biggs, J., Ebmeier, S. K., Aspinall, W. P., Lu, Z., Pritchard, M. E., Sparks, R. S. J., & Mather, T. A. (2014). Global link between deformation and volcanic eruption quantified by satellite imagery. *Nature Communications*, 5(1), 3471. doi:10.1038/ncomms4471
- Briole, P., Massonnet, D., & Delacourt, C. (1997). Post-eruptive deformation associated with the 1986–87 and 1989 lava flows of Etna detected by radar interferometry. *Geophysical Research Letters*, 24(1), 37-40. doi:10.1029/96gl03705

- Bürgmann, R., Rosen, P. A., & Fielding, E. J. (2000). Synthetic Aperture Radar Interferometry to Measure Earth's Surface Topography and Its Deformation. *Annual Review of Earth and Planetary Sciences*, 28(1), 169-209. doi:10.1146/annurev.earth.28.1.169
- Calò, F., Calcaterra, D., Iodice, A., Parise, M., & Ramondini, M. (2012). Assessing the activity of a large landslide in southern Italy by ground-monitoring and SAR interferometric techniques. *International Journal of Remote Sensing*, 33(11), 3512-3530. doi:10.1080/01431161.2011.630331
- Chen, S., Zhan, R., & Zhang, J. (2018). Geospatial Object Detection in Remote Sensing Imagery Based on Multiscale Single-Shot Detector with Activated Semantics. *Remote Sensing*, 10(6). doi:10.3390/rs10060820
- Chollet, F. (2018). *Deep Learning with Python*: Manning Publications Co.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. doi:10.1007/BF00994018
- Costante, G., Ciarfuglia, T. A., & Biondi, F. (2018, 4-7 June 2018). *Towards Monocular Digital Elevation Model (DEM) Estimation by Convolutional Neural Networks - Application on Synthetic Aperture Radar Images*. Paper presented at the EUSAR 2018; 12th European Conference on Synthetic Aperture Radar.
- Crosetto, M., Monserrat, O., Cuevas-González, M., Devanthéry, N., & Crippa, B. (2016). Persistent Scatterer Interferometry: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 78-89. doi:<https://doi.org/10.1016/j.isprsjprs.2015.10.011>
- Cutrona, L., Vivian, W., Leith, E., & Hall, G. (1961). Synthetic aperture radars: A paradigm for technology evolution. *IRE Trans. Military Electron*, 127-131.
- Duque, S., Lopez-Dekker, P., & Mallorqui, J. J. (2010). Single-Pass Bistatic SAR Interferometry Using Fixed-Receiver Configurations: Theory and Experimental Validation. *IEEE Transactions on Geoscience and Remote Sensing*, 48(6), 2740-2749. doi:10.1109/TGRS.2010.2041063
- Engdahl, M. E., & Hyypä, J. M. (2003). Land-cover classification using multitemporal ERS-1/2 InSAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(7), 1620-1628. doi:10.1109/TGRS.2003.813271
- Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., & Rucci, A. (2011). A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9), 3460-3470. doi:10.1109/TGRS.2011.2124465
- Ferretti, A., Monti-Guarnieri, A., Prati, C., Rocca, F., & Massonnet, D. (2007). *InSAR Principles: Guidelines for SAR Interferometry Processing and Interpretation (TM-19)*: ESA Publications.
- Ferretti, A., Prati, C., & Rocca, F. (2000). Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5), 2202-2212. doi:10.1109/36.868878

- Ferretti, A., Prati, C., & Rocca, F. (2001). Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 39(1), 8-20. doi:10.1109/36.898661
- Freeman, A. (1992). SAR calibration: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 30(6), 1107-1121. doi:10.1109/36.193786
- Gabriel, A. K., Goldstein, R. M., & Zebker, H. A. (1989). Mapping small elevation changes over large areas: Differential radar interferometry. *Journal of Geophysical Research: Solid Earth*, 94(B7), 9183-9191. doi:10.1029/JB094iB07p09183
- Ghiglia, D. C., & Pritt, M. D. (1998). *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*: Wiley.
- Goldstein, R. M., Zebker, H. A., & Werner, C. L. (1988). Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio Science*, 23(4), 713-720. doi:10.1029/RS023i004p00713
- Guo, H., Dou, C., Zhang, X., Han, C., & Yue, X. (2016). Earth observation from the manned low Earth orbit platforms. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 103-118. doi:<https://doi.org/10.1016/j.isprsjprs.2015.11.004>
- Guzzetti, F., Manunta, M., Ardizzone, F., Pepe, A., Cardinali, M., Zeni, G., . . . Lanari, R. (2009). Analysis of Ground Deformation Detected Using the SBAS-DInSAR Technique in Umbria, Central Italy. *Pure and Applied Geophysics*, 166(8-9), 1425-1459.
- Henderson, F. M., & Zong-Guo, X. (1997). SAR applications in human settlement detection, population estimation and urban land use pattern analysis: a status report. *IEEE Transactions on Geoscience and Remote Sensing*, 35(1), 79-85. doi:10.1109/36.551936
- Hoffer, E., Hubara, I., & Soudry, D. (2017). *Train longer, generalize better: closing the generalization gap in large batch training of neural networks*. Paper presented at the 31st Conference on Neural Information Processing Systems (NIPS 2017).
- Hooper, A., Zebker, H., Segall, P., & Kampes, B. (2004). A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers. *Geophysical Research Letters*, 31(23). doi:10.1029/2004gl021737
- Hunstad, I., Pepe, A., Atzori, S., Tolomei, C., Salvi, S., & Lanari, R. (2009). Surface deformation in the Abruzzi region, Central Italy, from multitemporal DInSAR analysis. *Geophysical Journal International*, 178(3), 1193-1197. doi:10.1111/j.1365-246X.2009.04284.x
- Ichikawa, K., & Hirose, A. (2017). Singular Unit Restoration in InSAR Using Complex-Valued Neural Networks in the Spectral Domain. *IEEE Transactions on Geoscience and Remote Sensing*, 55(3), 1717-1723. doi:10.1109/TGRS.2016.2630719
- Kerkhof, B. v. d., Pankratius, V., Chang, L., Swol, R. v., & Hanssen, R. F. (2018, 22-27 July 2018). *Automatic Insar Phase Modeling and Quality Assessment Using Machine Learning and Hypothesis Testing*. Paper presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium.

- Keskar, N. S., Nocedal, J., Tang, P. T. P., Mudigere, D., & Smelyanskiy, M. (2017). *On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima*. Paper presented at the 5th International Conference on Learning Representations (ICLR 2017).
- Khalil, R. Z., & Saad ul, H. (2018). InSAR coherence-based land cover classification of Okara, Pakistan. *The Egyptian Journal of Remote Sensing and Space Science*, 21, S23-S28. doi:<https://doi.org/10.1016/j.ejrs.2017.08.005>
- Kottayil, N. K., Zimmer, A., Mukherjee, S., Sun, X., Ghuman, P., & Cheng, I. (2018, 28-31 Oct. 2018). *Accurate Pixel-Based Noise Estimation for InSAR Interferograms*. Paper presented at the 2018 IEEE SENSORS.
- Krieger, G., Fiedler, H., Mittermayer, J., Papathanassiou, K., & Moreira, A. (2003). Analysis of multistatic configurations for spaceborne SAR interferometry. *IEE Proceedings - Radar, Sonar and Navigation*, 150(3), 87-96. doi:10.1049/ip-rsn:20030441
- Krieger, G., & Moreira, A. (2006). Spaceborne bi- and multistatic SAR: potential and challenges. *IEE Proceedings - Radar, Sonar and Navigation*, 153(3), 184-198. doi:10.1049/ip-rsn:20045111
- Lanari, R., Mora, O., Manunta, M., Mallorqui, J. J., Berardino, P., & Sansosti, E. (2004). A small-baseline approach for investigating deformations on full-resolution differential SAR interferograms. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7), 1377-1386. doi:10.1109/TGRS.2004.828196
- Lauknes, T. R., Piyush Shanker, A., Dehls, J. F., Zebker, H. A., Henderson, I. H. C., & Larsen, Y. (2010). Detailed rockslide mapping in northern Norway with small baseline and persistent scatterer interferometric SAR time series methods. *Remote Sensing of Environment*, 114(9), 2097-2109. doi:<https://doi.org/10.1016/j.rse.2010.04.015>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. doi:10.1038/nature14539
- Li, S., Xu, H., Gao, S., & Li, C. (2019, 28 July-2 Aug. 2019). *Non-fuzzy Interferometric Phase estimation method Based on Deep Learning*. Paper presented at the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium.
- Liu, C.-C., Zhang, Y.-C., Chen, P.-Y., Lai, C.-C., Chen, Y.-H., Cheng, J.-H., & Ko, M.-H. (2019). Clouds Classification from Sentinel-2 Imagery with Deep Residual Learning and Semantic Image Segmentation. *Remote Sensing*, 11(2). doi:10.3390/rs11020119
- Lyu, H., Lu, H., Mou, L., Li, W., Wright, J., Li, X., . . . Gong, P. (2018). Long-Term Annual Mapping of Four Cities on Different Continents by Applying a Deep Information Learning Method to Landsat Data. *Remote Sensing*, 10(3). doi:10.3390/rs10030471
- Ma, P., Zhang, F., & Lin, H. (2019). Prediction of InSAR time-series deformation using deep convolutional neural networks. *Remote Sensing Letters*, 11(2), 137-145. doi:10.1080/2150704X.2019.1692390
- Massonnet, D., Briole, P., & Arnaud, A. (1995). Deflation of Mount Etna monitored by spaceborne radar interferometry. *Nature*, 375(6532), 567-570. doi:10.1038/375567a0

- Massonnet, D., & Feigl, K. L. (1998). Radar interferometry and its application to changes in the earth's surface. *Reviews of Geophysics*, 36(4), 441-500.
- Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., & Rabaute, T. (1993). The displacement field of the Landers earthquake mapped by radar interferometry. *Nature*, 364(6433), 138-142. doi:10.1038/364138a0
- Mora, O., Mallorqui, J. J., & Broquetas, A. (2003). Linear and nonlinear terrain deformation maps from a reduced set of interferometric SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 41(10), 2243-2253. doi:10.1109/TGRS.2003.814657
- Moran, M. S., Vidal, A., Troufleau, D., Inoue, Y., & Mitchell, T. A. (1998). Ku- and C-band SAR for discriminating agricultural crop and soil conditions. *IEEE Transactions on Geoscience and Remote Sensing*, 36(1), 265-272. doi:10.1109/36.655335
- Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1), 6-43. doi:10.1109/MGRS.2013.2248301
- Nof, N. R., Abelson, M., Raz, E., Magen, Y., Atzori, S., Salvi, S., & Baer, G. (2019). SAR Interferometry for Sinkhole Early Warning and Susceptibility Assessment along the Dead Sea, Israel. *Remote Sensing*, 11(1). doi:10.3390/rs11010089
- Osmanoğlu, B., Dixon, T. H., Wdowinski, S., Cabral-Cano, E., & Jiang, Y. (2011). Mexico City subsidence observed with persistent scatterer InSAR. *International Journal of Applied Earth Observation and Geoinformation*, 13(1), 1-12. doi:<https://doi.org/10.1016/j.jag.2010.05.009>
- Osmanoğlu, B., Sunar, F., Wdowinski, S., & Cabral-Cano, E. (2016). Time series analysis of InSAR data: Methods and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 90-102. doi:<https://doi.org/10.1016/j.isprsjprs.2015.10.003>
- Peltzer, G., & Rosen, P. (1995). Surface Displacement of the 17 May 1993 Eureka Valley, California, Earthquake Observed by SAR Interferometry. *Science*, 268(5215), 1333. doi:10.1126/science.268.5215.1333
- Pepe, A., & Calò, F. (2017). A Review of Interferometric Synthetic Aperture RADAR (InSAR) Multi-Track Approaches for the Retrieval of Earth's Surface Displacements. *Applied Sciences*, 7(12), 1264.
- Perissin, D., & Wang, T. (2012). Repeat-Pass SAR Interferometry With Partially Coherent Targets. *IEEE Transactions on Geoscience and Remote Sensing*, 50(1), 271-280. doi:10.1109/TGRS.2011.2160644
- Qi, Z., Yeh, A. G.-O., Li, X., & Lin, Z. (2012). A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. *Remote Sensing of Environment*, 118, 21-39. doi:<https://doi.org/10.1016/j.rse.2011.11.001>
- Rodriguez, E., & Martin, J. M. (1992). Theory and design of interferometric synthetic aperture radars. *IEE Proceedings F - Radar and Signal Processing*, 139(2), 147-159. doi:10.1049/ip-f-2.1992.0018

- Rongier, G., Rude, C., Herring, T., & Pankratius, V. (2019). Generative Modeling of InSAR Interferograms. *Earth and Space Science*, 6(12), 2671-2683. doi:10.1029/2018ea000533
- Rosen, P. A., Hensley, S., Joughin, I. R., Li, F. K., Madsen, S. N., Rodriguez, E., & Goldstein, R. M. (2000). Synthetic aperture radar interferometry. *Proceedings of the IEEE*, 88(3), 333-382. doi:10.1109/5.838084
- Ruch, J., Anderssohn, J., Walter, T. R., & Motagh, M. (2008). Caldera-scale inflation of the Lazufre volcanic area, South America: Evidence from InSAR. *Journal of Volcanology and Geothermal Research*, 174(4), 337-344. doi:<https://doi.org/10.1016/j.jvolgeores.2008.03.009>
- Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., . . . Siegwart, R. (2018). WeedMap: A Large-Scale Semantic Weed Mapping Framework Using Aerial Multispectral Imaging and Deep Neural Network for Precision Farming. *Remote Sensing*, 10(9). doi:10.3390/rs10091423
- Sansosti, E., Berardino, P., Manunta, M., Serafino, F., & Fornaro, G. (2006). Geometrical SAR image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10), 2861-2870. doi:10.1109/TGRS.2006.875787
- Shamshiri, R., Motagh, M., Nahavandchi, H., Haghshenas Haghghi, M., & Hoseini, M. (2019). Improving tropospheric corrections on large-scale Sentinel-1 interferograms using a machine learning approach for integration with GNSS-derived zenith total delay (ZTD). *Remote Sensing of Environment*, 239, 111608. doi:<https://doi.org/10.1016/j.rse.2019.111608>
- Sica, F., Pulella, A., Nannini, M., Pinheiro, M., & Rizzoli, P. (2019). Repeat-pass SAR interferometry for land cover classification: A methodology using Sentinel-1 Short-Time-Series. *Remote Sensing of Environment*, 232, 111277. doi:<https://doi.org/10.1016/j.rse.2019.111277>
- Soergel, U. (2010). *Radar Remote Sensing of Urban Areas* (Vol. 15): Springer Science & Business Media.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15, 1929-1958.
- Tomiyasu, K. (1978). Tutorial review of synthetic-aperture radar (SAR) with applications to imaging of the ocean surface. *Proceedings of the IEEE*, 66(5), 563-583. doi:10.1109/PROC.1978.10961
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., . . . Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9-24. doi:<https://doi.org/10.1016/j.rse.2011.05.028>
- Trasatti, E., Casu, F., Giunchi, C., Pepe, S., Solaro, G., Tagliaventi, S., . . . Lanari, R. (2008). The 2004–2006 uplift episode at Campi Flegrei caldera (Italy): Constraints from SBAS-DInSAR ENVISAT data and Bayesian source inference. *Geophysical Research Letters*, 35(7). doi:10.1029/2007gl033091

- Valade, S., Ley, A., Massimetti, F., D'Hondt, O., Laiolo, M., Coppola, D., . . . Walter, T. R. (2019). Towards Global Volcano Monitoring Using Multisensor Sentinel Missions and Artificial Intelligence: The MOUNTS Monitoring System. *Remote Sensing*, 11(13), 1528. Retrieved from <https://www.mdpi.com/2072-4292/11/13/1528>
- Vicente-Guijalba, F., Jacob, A., Lopez-Sanchez, J. M., Lopez-Martinez, C., Duro, J., Notarnicola, C., . . . Engdahl, M. (2018, 22-27 July 2018). *Sincohmap: Land-Cover and Vegetation Mapping Using Multi-Temporal Sentinel-1 Interferometric Coherence*. Paper presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium.
- Wang, H., Jiang, J., & Zhang, G. (2018). CraterIDNet: An End-to-End Fully Convolutional Neural Network for Crater Detection and Identification in Remotely Sensed Planetary Images. *Remote Sensing*, 10(7). doi:10.3390/rs10071067
- Wang, W., Dou, S., Jiang, Z., & Sun, L. (2018). A Fast Dense Spectral–Spatial Convolution Network Framework for Hyperspectral Images Classification. *Remote Sensing*, 10(7). doi:10.3390/rs10071068
- Wang, Y., Guo, Y., Hu, S., Li, Y., Wang, J., Liu, X., & Wang, L. (2019). Ground Deformation Analysis Using InSAR and Backpropagation Prediction with Influencing Factors in Erhai Region, China. *Sustainability*, 11(10), 2853. Retrieved from <https://www.mdpi.com/2071-1050/11/10/2853>
- Wang, Z., Li, Z., & Mills, J. (2018). A new approach to selecting coherent pixels for ground-based SAR deformation monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144, 412-422. doi:<https://doi.org/10.1016/j.isprsjprs.2018.08.008>
- Wegmuller, U., & Werner, C. (1997). Retrieval of vegetation parameters with SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 35(1), 18-24. doi:10.1109/36.551930
- Wegmuller, U., & Werner, C. L. (1995). SAR interferometric signatures of forest. *IEEE Transactions on Geoscience and Remote Sensing*, 33(5), 1153-1161. doi:10.1109/36.469479
- Wei, M., & Sandwell, D. T. (2010). Decorrelation of L-Band and C-Band Interferometry Over Vegetated Areas in California. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7), 2942-2952. doi:10.1109/TGRS.2010.2043442
- Wiley, C. (1985). Synthetic aperture radars: A paradigm for technology evolution. *IEEE Transactions on Aerospace and Electronic Systems*, 21(3), 440-443.
- Wu, G., Shao, X., Guo, Z., Chen, Q., Yuan, W., Shi, X., . . . Shibasaki, R. (2018). Automatic Building Segmentation of Aerial Imagery Using Multi-Constraint Fully Convolutional Networks. *Remote Sensing*, 10(3). doi:10.3390/rs10030407
- Xu, Y., Wu, L., Xie, Z., & Chen, Z. (2018). Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. *Remote Sensing*, 10(1). doi:10.3390/rs10010144

- Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., & Guo, Z. (2018). Automatic Ship Detection in Remote Sensing Images from Google Earth of Complex Scenes Based on Multiscale Rotation Dense Feature Pyramid Networks. *Remote Sensing*, 10(1). doi:10.3390/rs10010132
- Yun, H.-W., Kim, J.-R., Choi, Y.-S., & Lin, S.-Y. (2019). Analyses of Time Series InSAR Signatures for Land Cover Classification: Case Studies over Dense Forestry Areas with L-Band SAR Images. *Sensors (Basel, Switzerland)*, 19(12), 2830. doi:10.3390/s19122830
- Zeng, D., Chen, S., Chen, B., & Li, S. (2018). Improving Remote Sensing Scene Classification by Integrating Global-Context and Local-Object Features. *Remote Sensing*, 10(5). doi:10.3390/rs10050734
- Zhang, Y., Sun, H., Zuo, J., Wang, H., Xu, G., & Sun, X. (2018). Aircraft Type Recognition in Remote Sensing Images Based on Feature Learning with Conditional Generative Adversarial Networks. *Remote Sensing*, 10(7). doi:10.3390/rs10071123
- Zhu, Q., Zhong, Y., Liu, Y., Zhang, L., & Li, D. (2018). A Deep-Local-Global Feature Fusion Framework for High Spatial Resolution Imagery Scene Classification. *Remote Sensing*, 10(4). doi:10.3390/rs10040568

8. Anexo

8.1. Resultados relativos à base de dados de dispersores permanentes da cidade de Bratislava

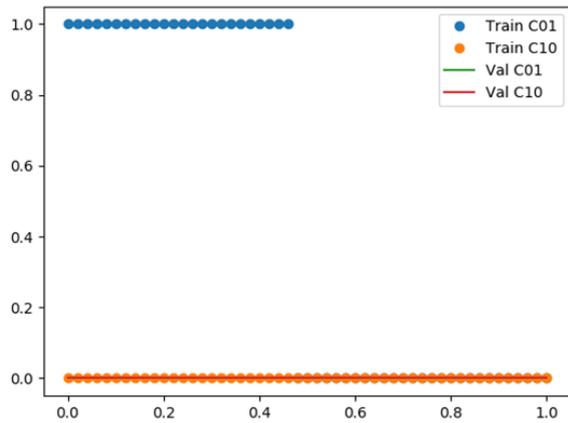


Figura 26a

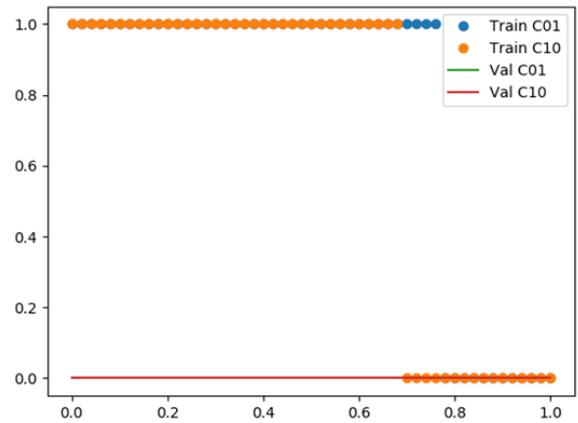


Figura 26b

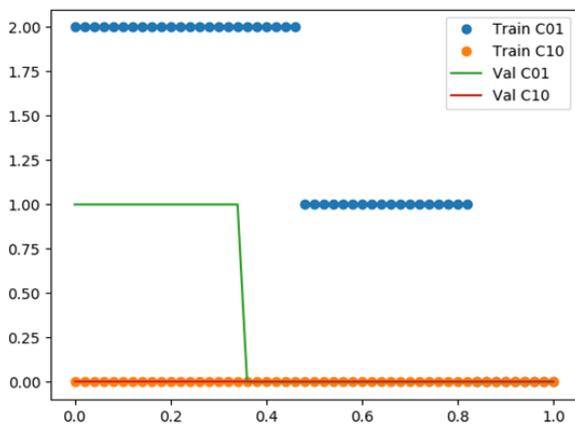


Figura 26c

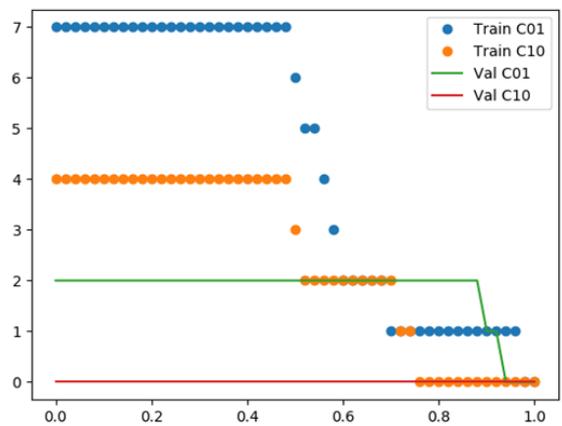


Figura 26d

Figura 26. Número de PSs outliers classificados como PSs inliers, e vice-versa, em função do valor de threshold aplicado à saída dos modelos de classificação relativos à divisão de dados 1 (Figura 18a) da base de dados de PSs da cidade de Bratislava, considerando para cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos (Figuras 26a a 26d, respetivamente). *Train C01* correspondem a PSs de treino *outliers* classificados como *inliers* e *Train C10* o contrário. *Val C01* e *Val C10* são as correspondentes métricas avaliadas nos PSs de validação.

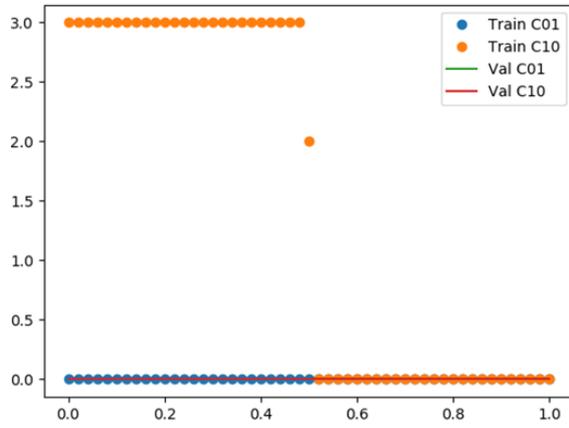


Figura 27a

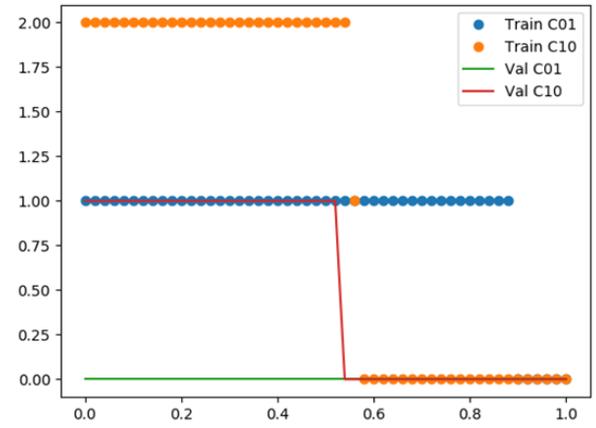


Figura 27b

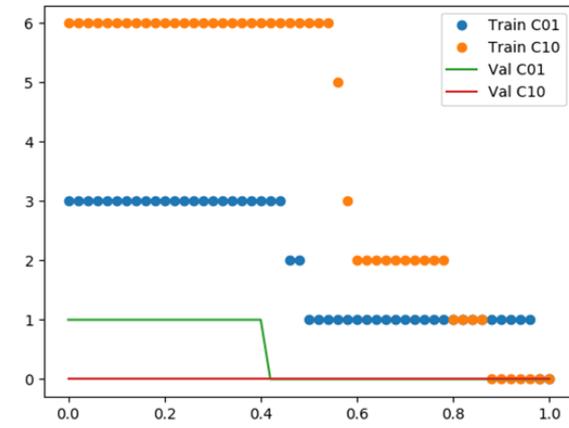


Figura 27c

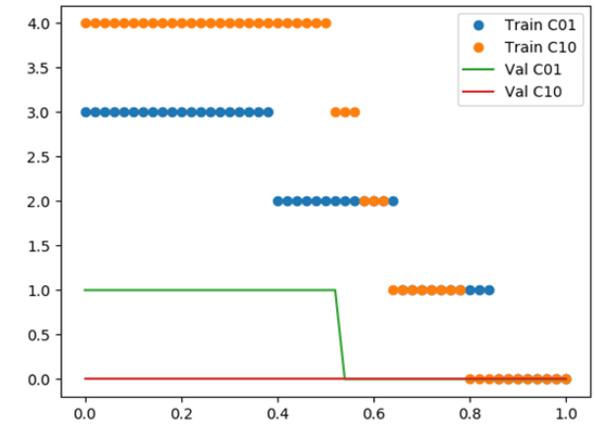


Figura 27d

Figura 27. Número de PSs outliers classificados como PSs inliers, e vice-versa, em função do valor de threshold aplicado à saída dos modelos de classificação relativos à divisão de dados 2 (Figura 18b) da base de dados de PSs da cidade de Bratislava, considerando para cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos (Figuras 27a a 27d, respetivamente). *Train C01* correspondem a PSs de treino *outliers* classificados como *inliers* e *Train C10* o contrário. *Val C01* e *Val C10* são as correspondentes métricas avaliadas nos PSs de validação.

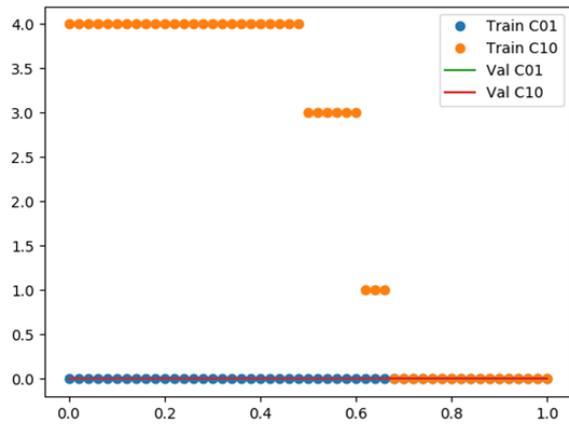


Figura 28a

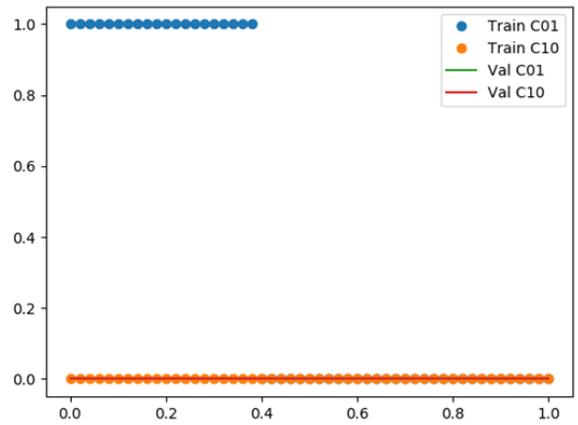


Figura 28b

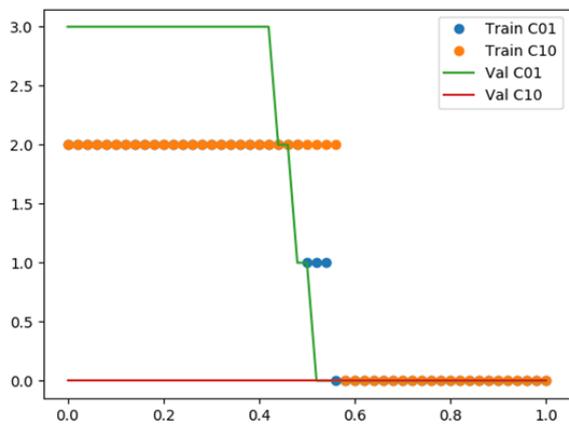


Figura 28c

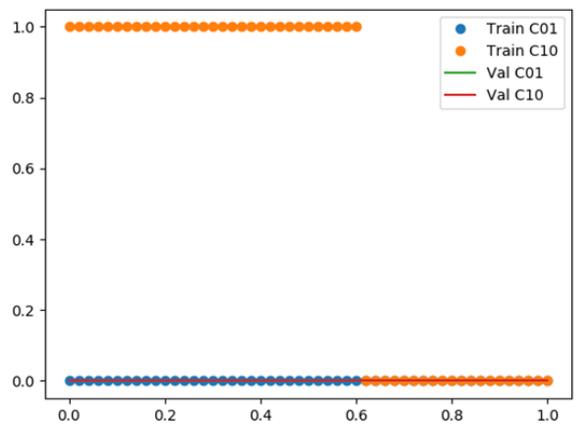


Figura 28d

Figura 28. Número de PSs outliers classificados como PSs inliers, e vice-versa, em função do valor de threshold aplicado à saída dos modelos de classificação relativos à divisão de dados 3 (Figura 18c) da base de dados de PSs da cidade de Bratislava, considerando para cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos (Figuras 28a a 28d, respetivamente). *Train C01* correspondem a PSs de treino *outliers* classificados como *inliers* e *Train C10* o contrário. *Val C01* e *Val C10* são as correspondentes métricas avaliadas nos PSs de validação.

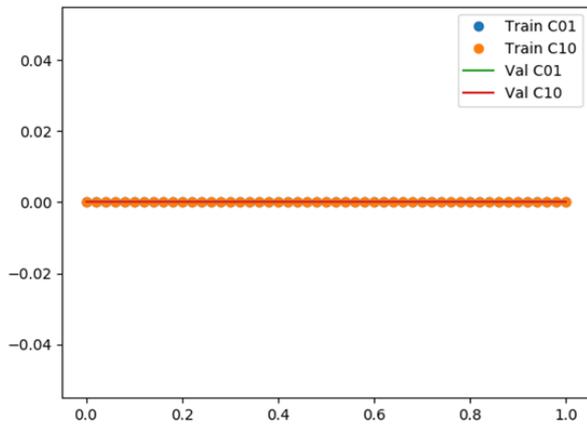


Figura 29a

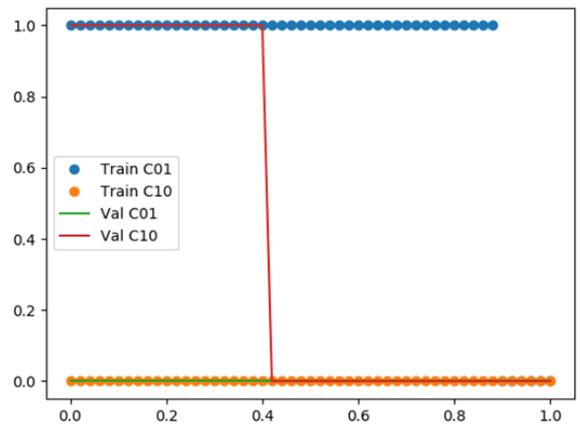


Figura 29b

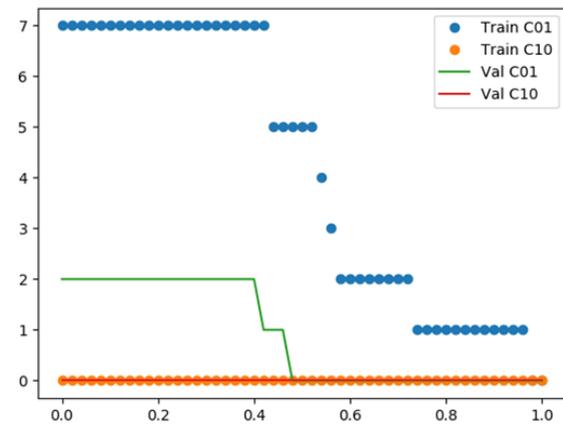


Figura 29c

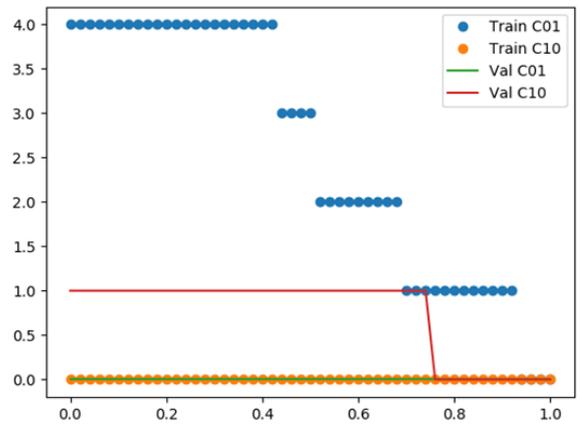


Figura 29d

Figura 29. Número de PSs outliers classificados como PSs inliers, e vice-versa, em função do valor de threshold aplicado à saída dos modelos de classificação relativos à divisão de dados 4 (Figura 18d) da base de dados de PSs da cidade de Bratislava, considerando para cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos (Figuras 29a a 29d, respetivamente). *Train C01* correspondem a PSs de treino *outliers* classificados como *inliers* e *Train C10* o contrário. *Val C01* e *Val C10* são as correspondentes métricas avaliadas nos PSs de validação.

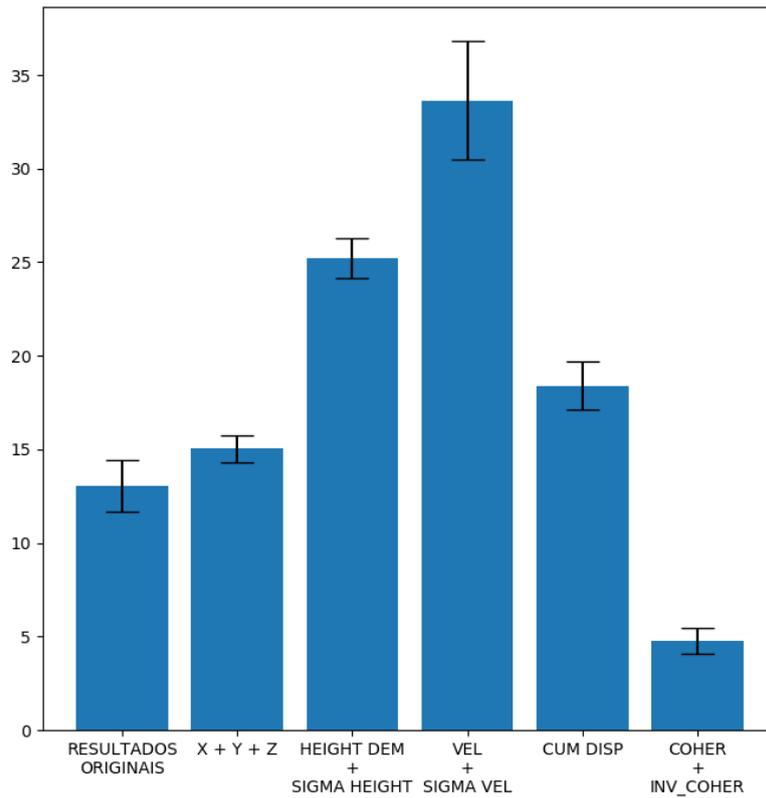


Figura 30. Percentagem média de classificações incertas e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 3 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

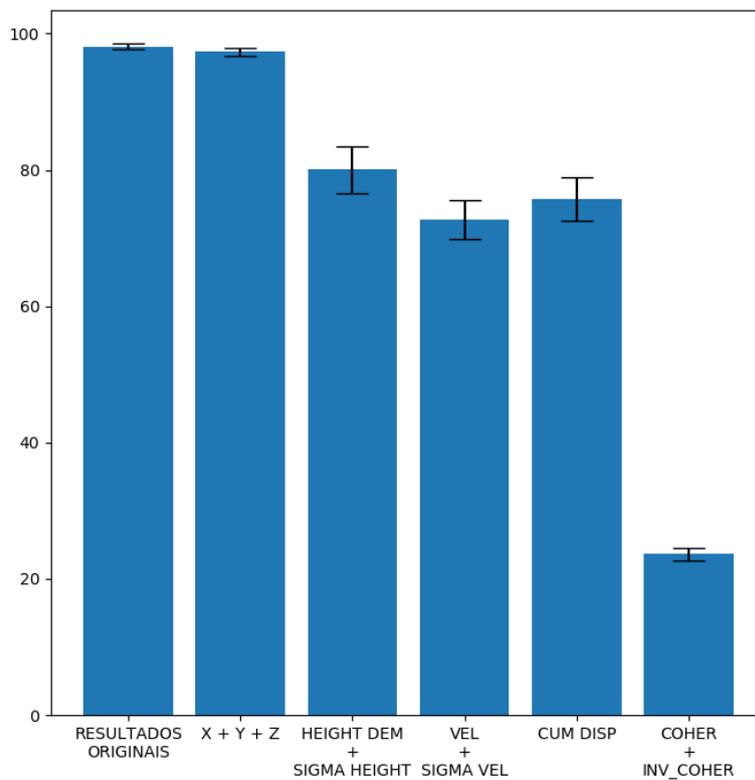


Figura 31. Exatidão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 3 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

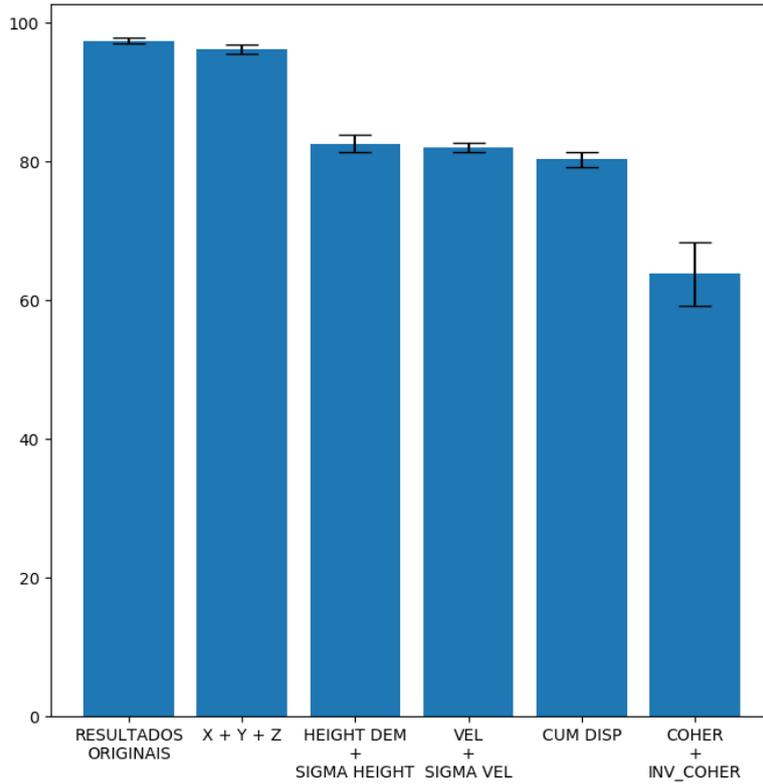


Figura 32. Precisão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 3 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

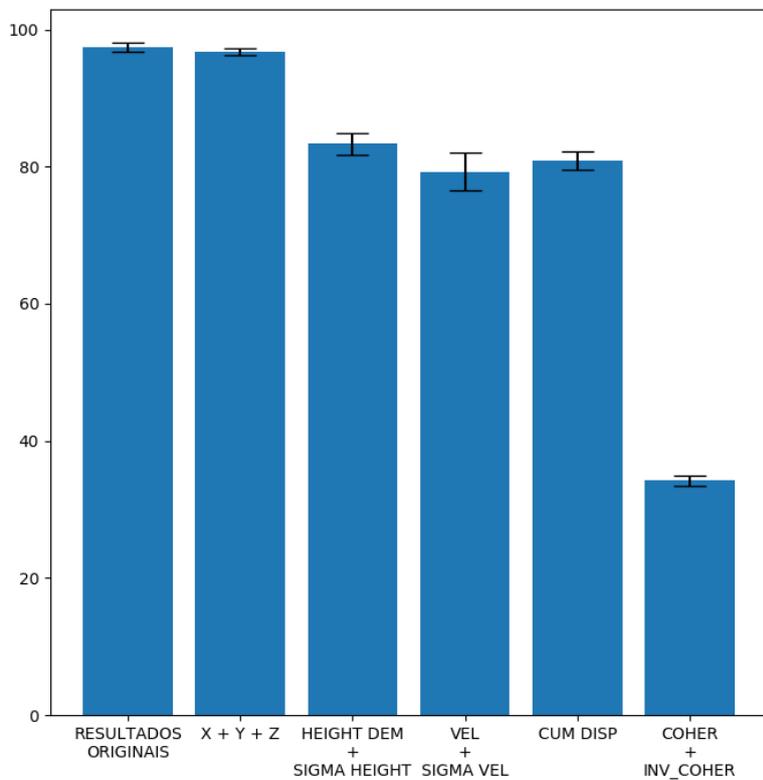


Figura 33. Recall médio e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 3 PSs vizinhos mais próximos, relativo à base de dados de PSs da cidade de Bratislava.

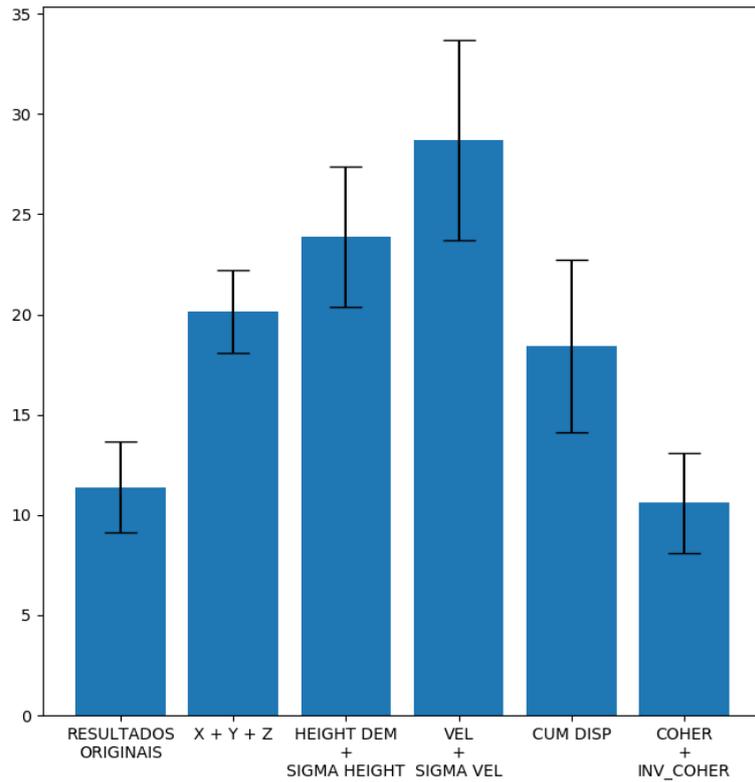


Figura 34. Percentagem média de classificações incertas e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 10 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

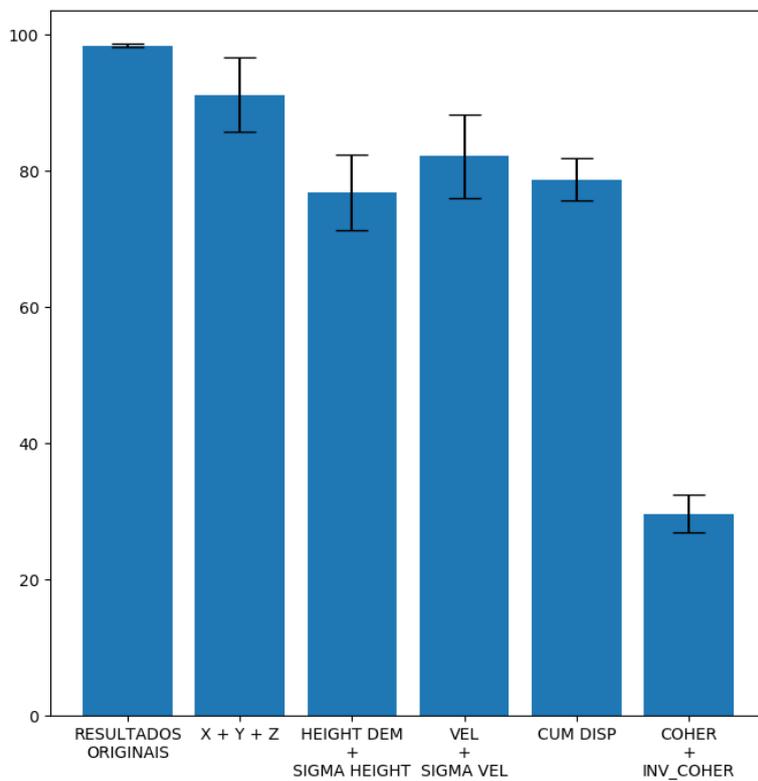


Figura 35. Exatidão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 10 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

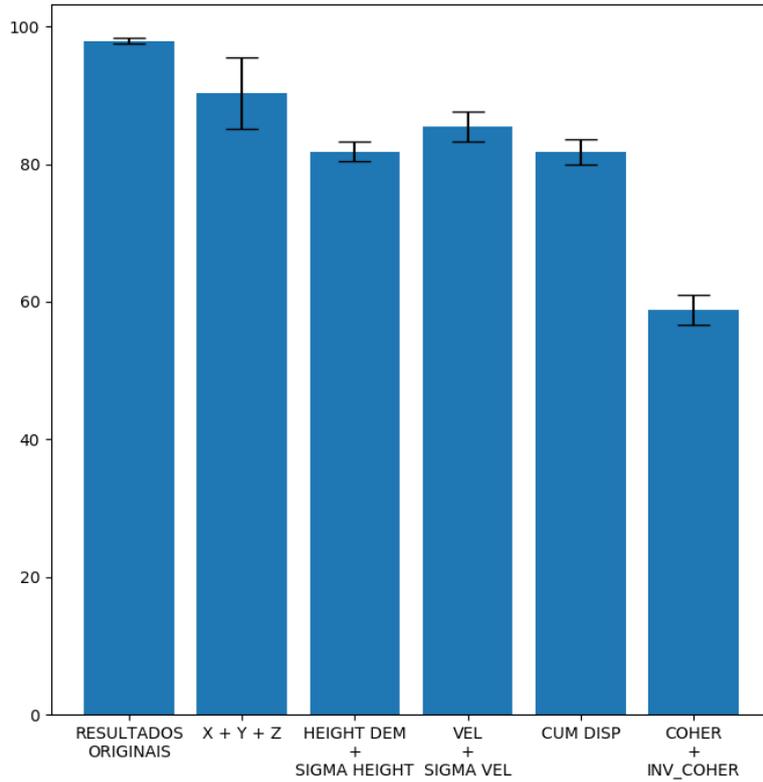


Figura 36. Precisão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 10 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

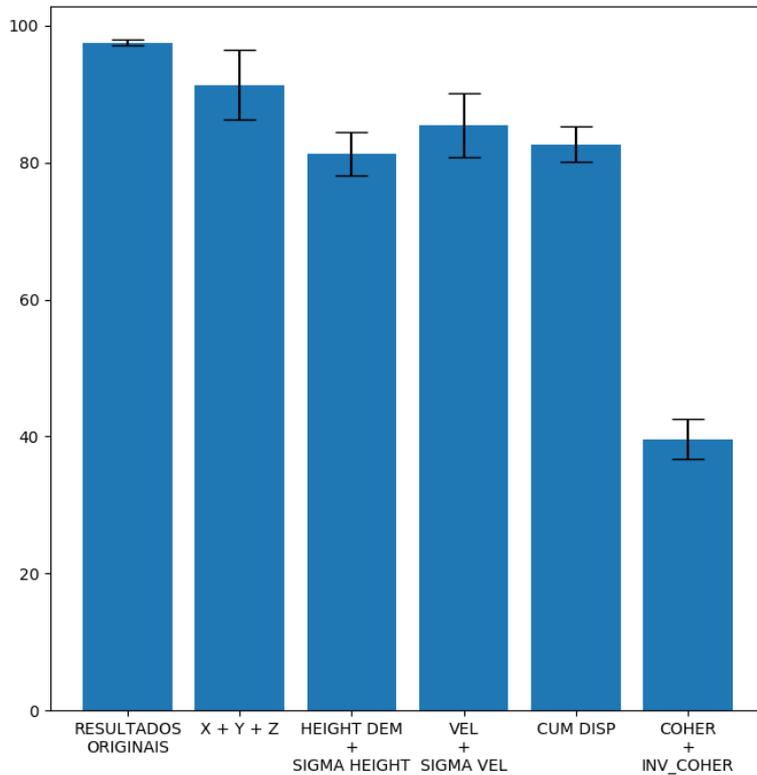


Figura 37. Recall médio e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 10 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

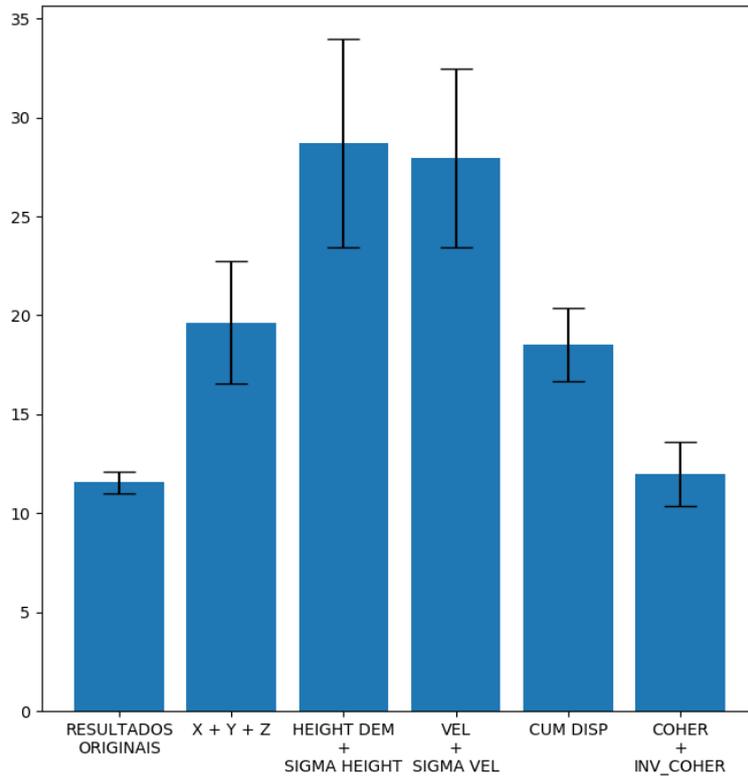


Figura 38. Percentagem média de classificações incertas e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 15 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

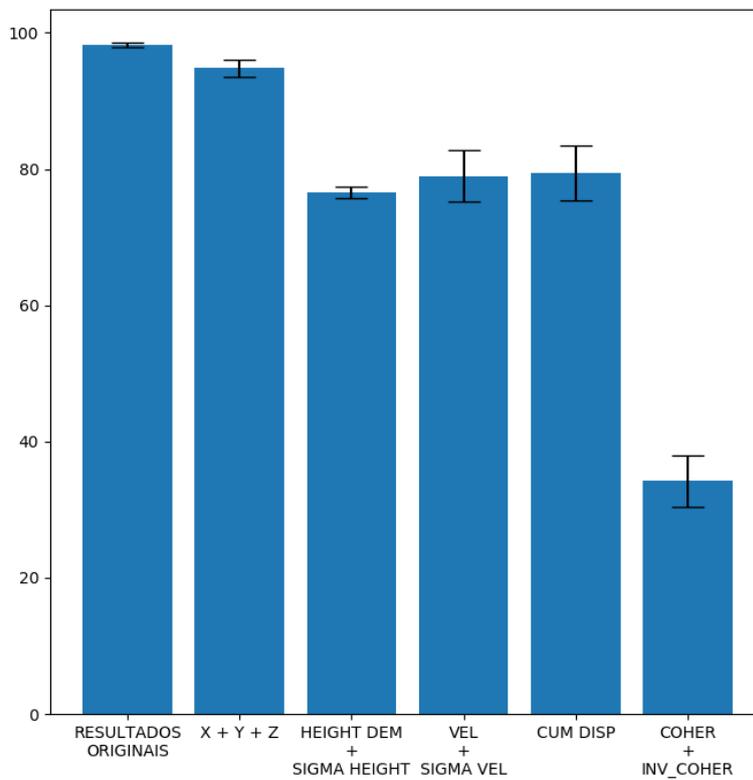


Figura 39. Exatidão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 15 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

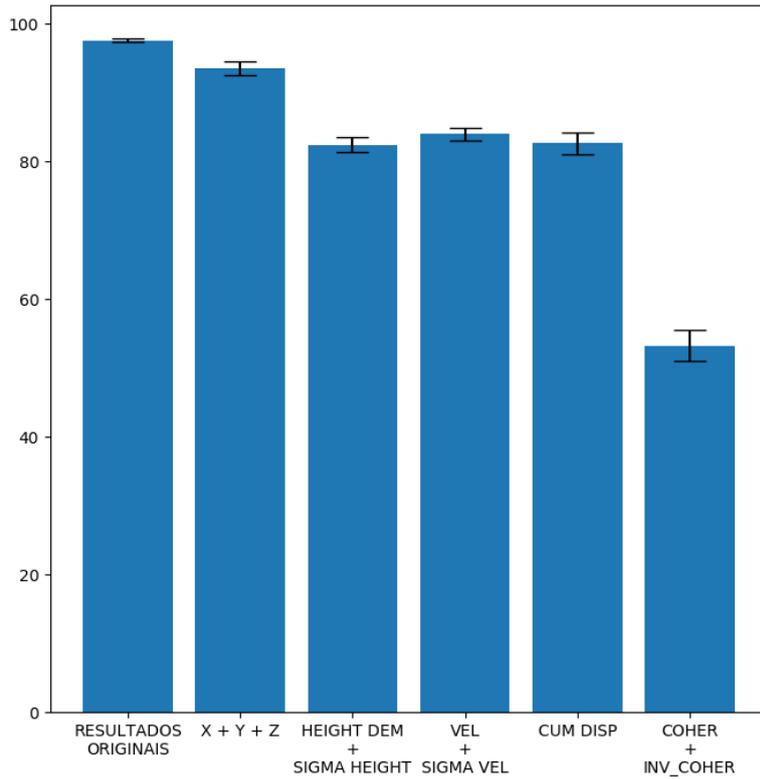


Figura 40. Precisão média e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 15 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

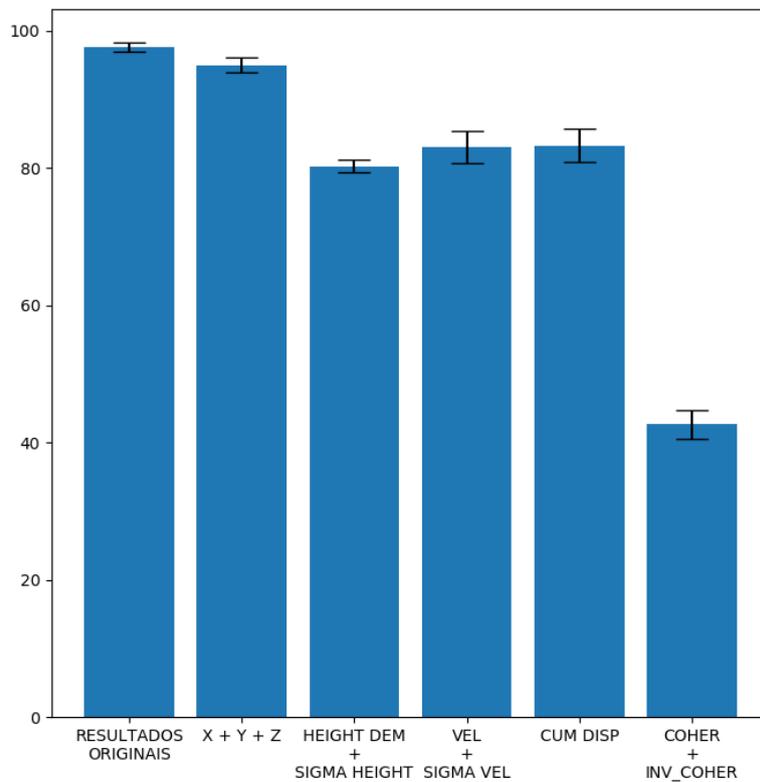


Figura 41. Recall médio e o respetivo desvio padrão resultante da avaliação da importância de cada parâmetro na classificação final de cada PS, considerando os 15 PSs vizinhos mais próximos, relativa à base de dados de PSs da cidade de Bratislava.

Resultados relativos aos modelos de classificação treinados na divisão de dados 1

n	PSs outliers	PSs inliers	PSs duvidosos	Total
3	5.47	2.08	4.82	12.37
5	4.74	2.15	4.87	11.76
10	4.52	2.45	5.75	12.72
15	2.99	3.07	5.72	11.78

Tabela 13. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	3487	0	37
	1	0	5488	42
	2	128	11	2040

Tabela 14. Matriz de confusão relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	3641	0	42
	1	0	5442	37
	2	96	10	1921

Tabela 16. Matriz de confusão relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	3578	0	32
	1	0	5485	44
	2	130	5	2037

Tabela 15. Matriz de confusão relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	3852	1	86
	1	2	5352	33
	2	79	19	1885

Tabela 17. Matriz de confusão relativa à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

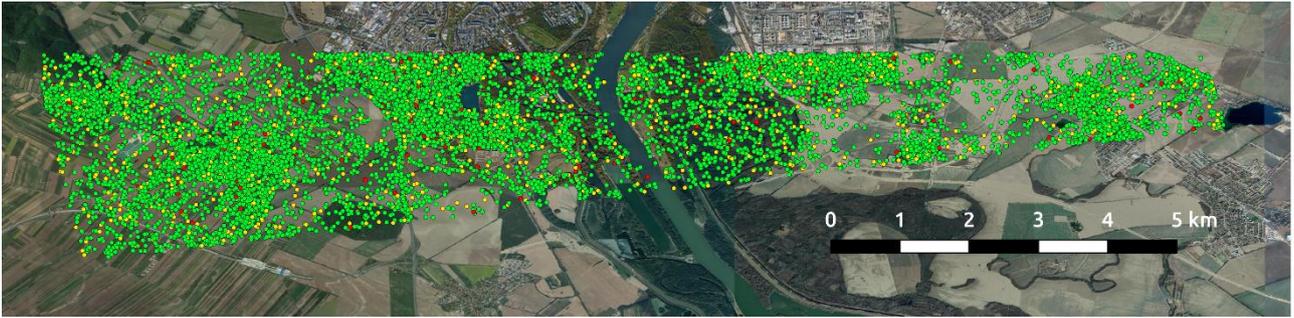


Figura 42a

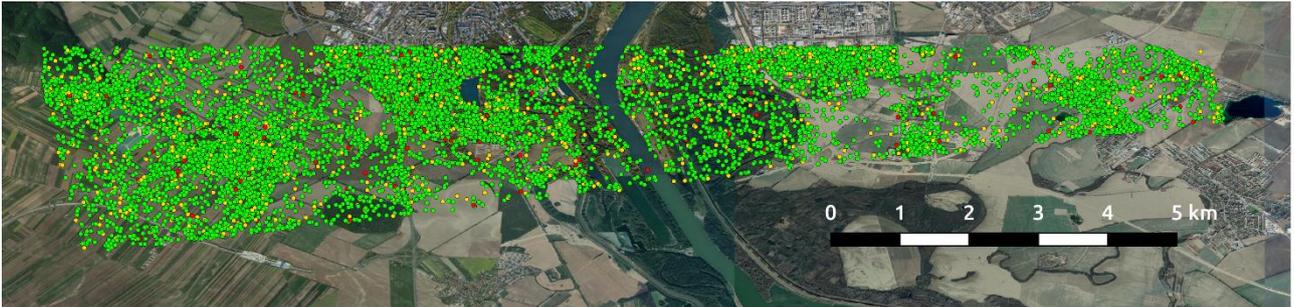


Figura 42b

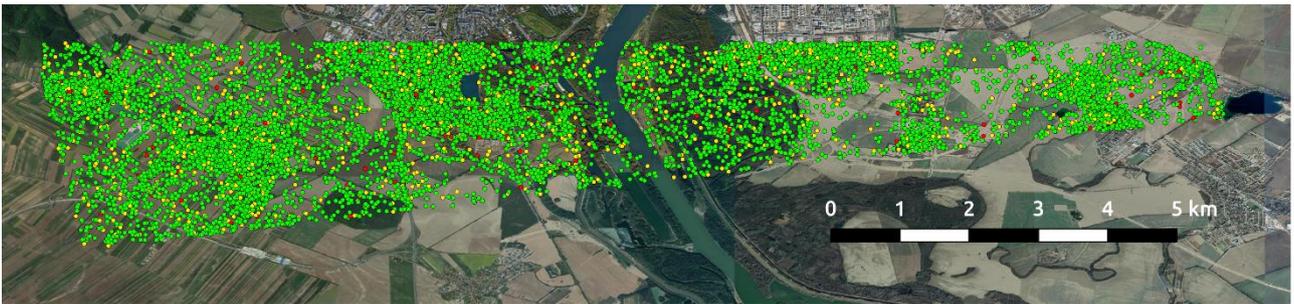


Figura 42c

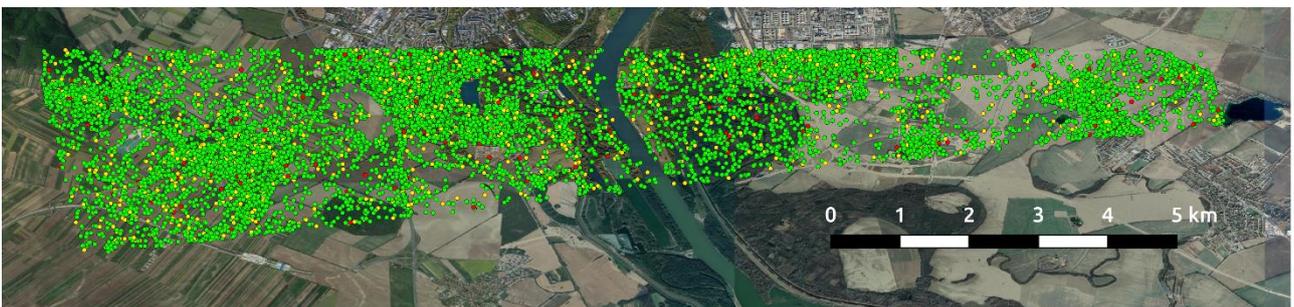


Figura 42d

Figura 42. Mapas de classificação relativos à divisão de dados 1 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos (Figuras 42a a 42d, respetivamente). PSs bem classificados estão coloridos a verde, mal classificados a vermelho, e cuja classificação é incerta estão coloridos a amarelo.

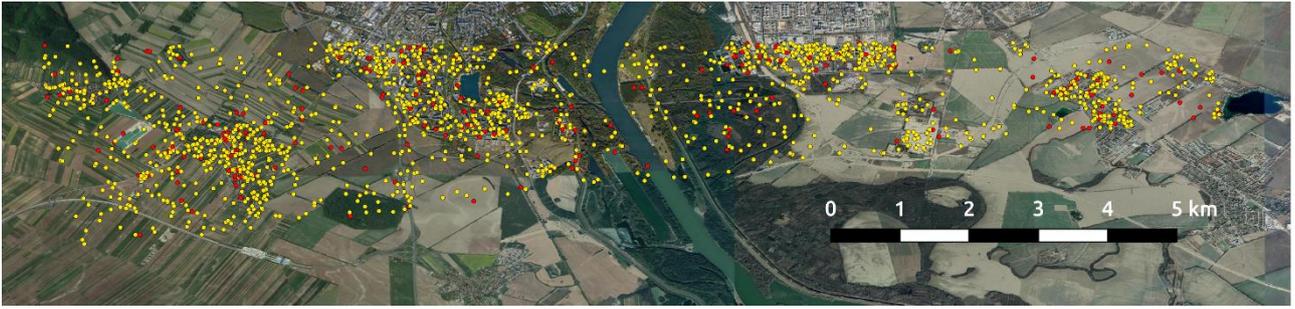


Figura 43a

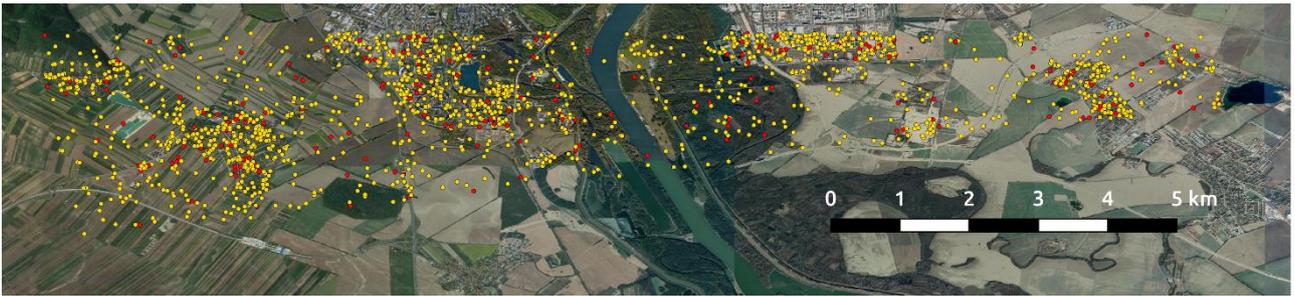


Figura 43b

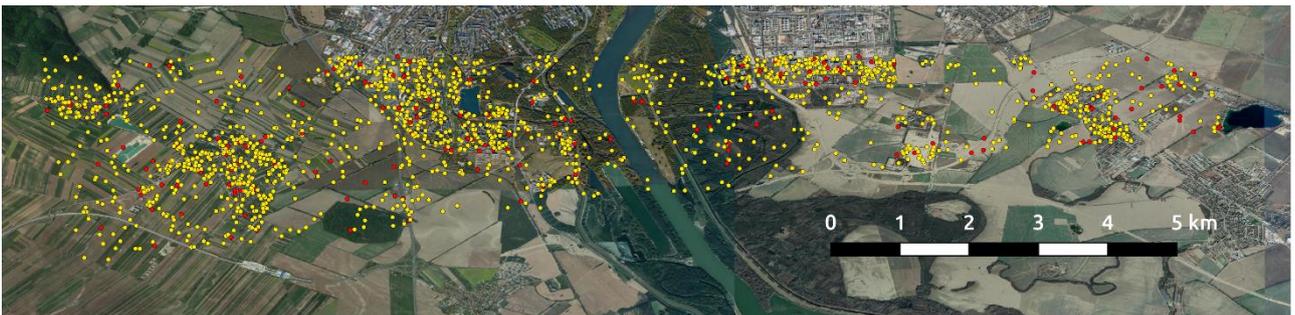


Figura 43c

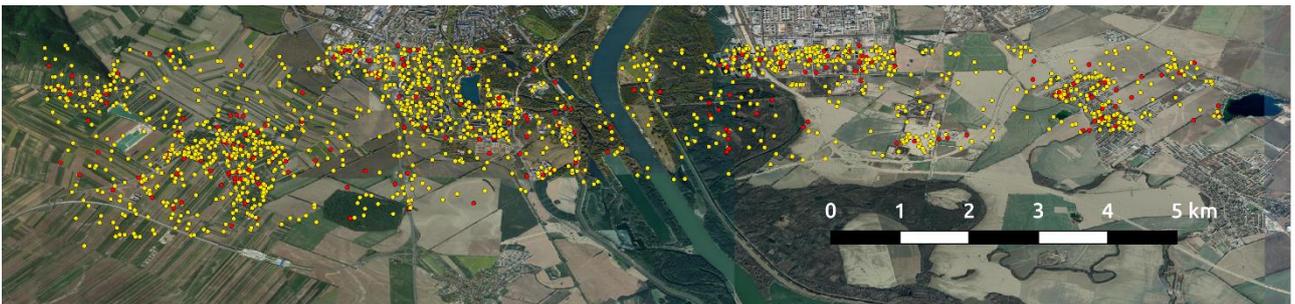


Figura 43d

Figura 43. Segunda versão dos mapas de classificação ilustrados na Figura 42, demonstrando apenas classificações erradas (a vermelho) e classificações incertas (a amarelo). As Figuras 43a a 43d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos.

Resultados relativos aos modelos de classificação treinados na divisão de dados 2

n	PSs outliers	PSs inliers	PSs duvidosos	Total
3	5.57	2.63	6.77	14.97
5	5.73	2.18	5.24	13.15
10	3.74	3.03	7.15	13.92
15	4.02	2.09	5.99	12.10

Tabela 18. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	4111	0	177
	1	0	8682	72
	2	122	6	2876

Tabela 19. Matriz de confusão relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	4481	0	138
	1	0	8608	71
	2	97	4	2845

Tabela 21. Matriz de confusão relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	4035	0	95
	1	0	8766	72
	2	167	7	3247

Tabela 20. Matriz de confusão relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	4412	0	134
	1	0	8784	80
	2	114	5	3059

Tabela 22. Matriz de confusão relativa à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.



Figura 44a



Figura 44b



Figura 44c



Figura 44d

Figura 44. Mapas de classificação relativos à divisão de dados 2 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos (Figuras 44a a 44d, respetivamente). PSs bem classificados estão coloridos a verde, mal classificados a vermelho, e cuja classificação é incerta estão coloridos a amarelo.

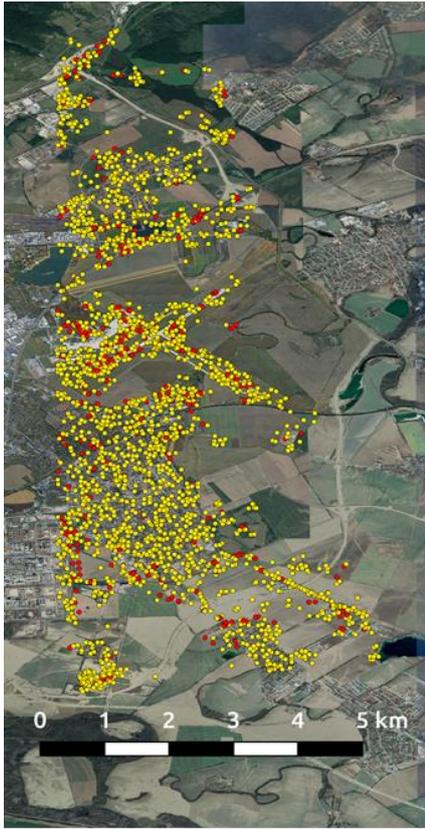


Figura 45a

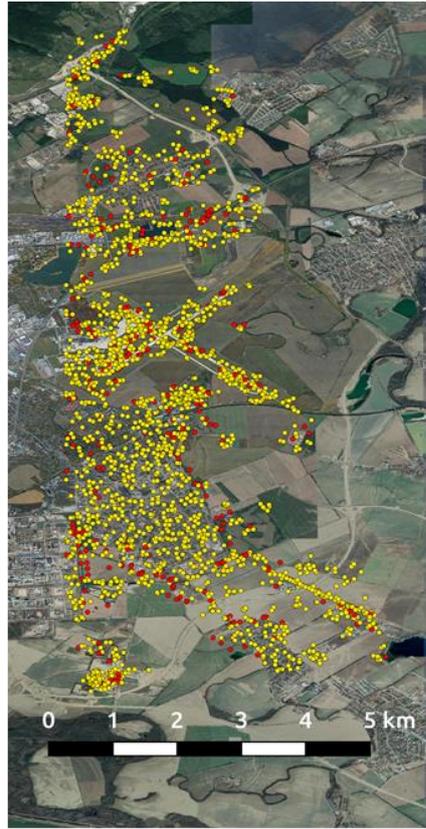


Figura 45b

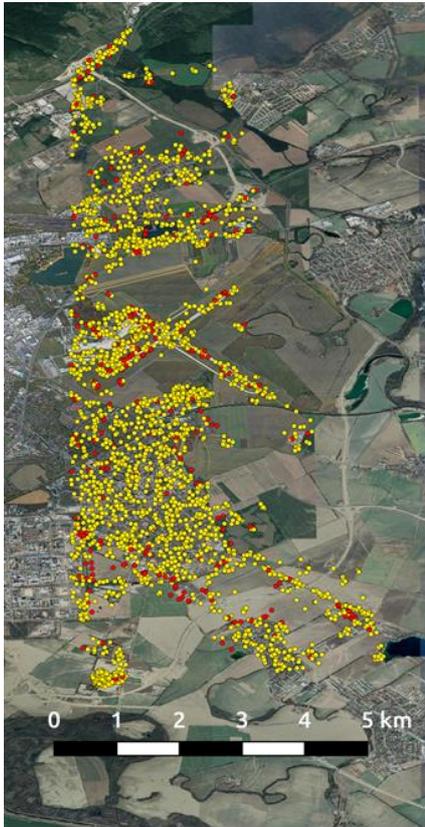


Figura 45c

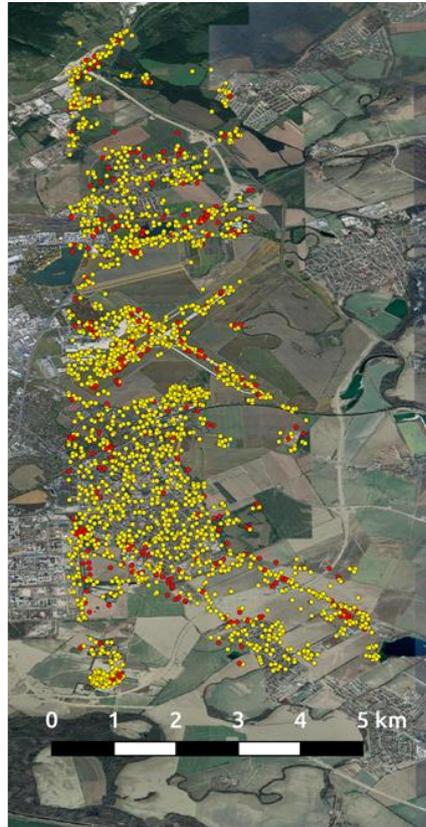


Figura 45d

Figura 45. Segunda versão dos mapas de classificação ilustrados na Figura 44, demonstrando apenas classificações erradas (a vermelho) e classificações incertas (a amarelo). As Figuras 45a a 45d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos.

Resultados relativos aos modelos de classificação treinados na divisão de dados 3

n	PSs outliers	PSs inliers	PSs duvidosos	Total
3	4.20	2.62	4.95	11.77
5	2.84	2.45	4.76	10.05
10	2.55	1.56	5.15	9.26
15	3.90	3.83	3.73	11.46

Tabela 23. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	4474	0	57
	1	0	8232	63
	2	141	25	2971

Tabela 24. Matriz de confusão relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	4809	1	63
	1	0	8440	72
	2	103	8	2920

Tabela 26. Matriz de confusão relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	4739	1	64
	1	0	8267	53
	2	121	19	3008

Tabela 25. Matriz de confusão relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	4489	1	27
	1	0	8001	46
	2	180	36	3238

Tabela 27. Matriz de confusão relativa à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

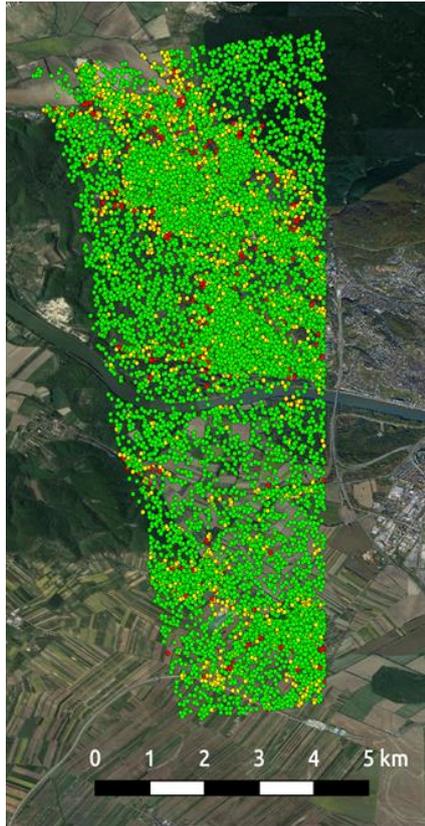


Figura 46a

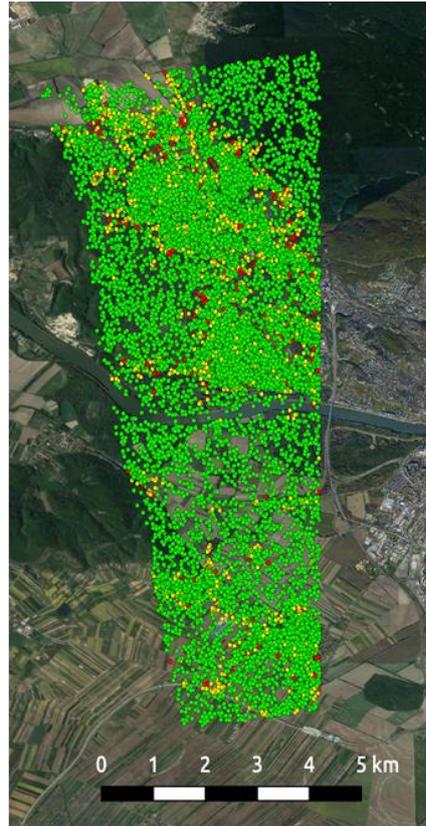


Figura 46b

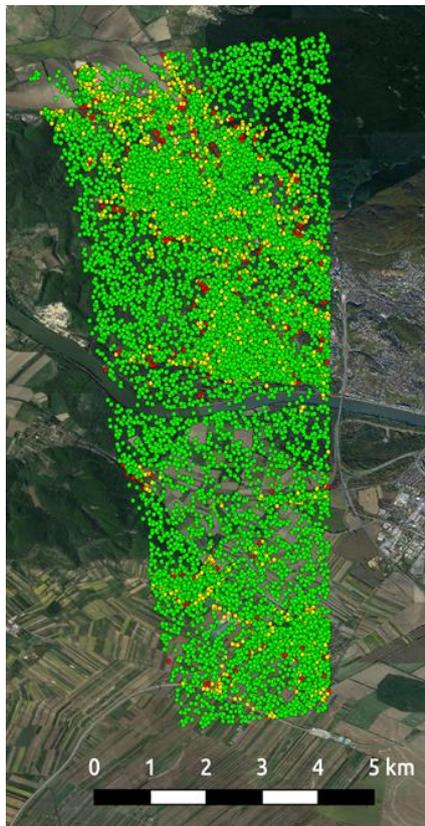


Figura 46c

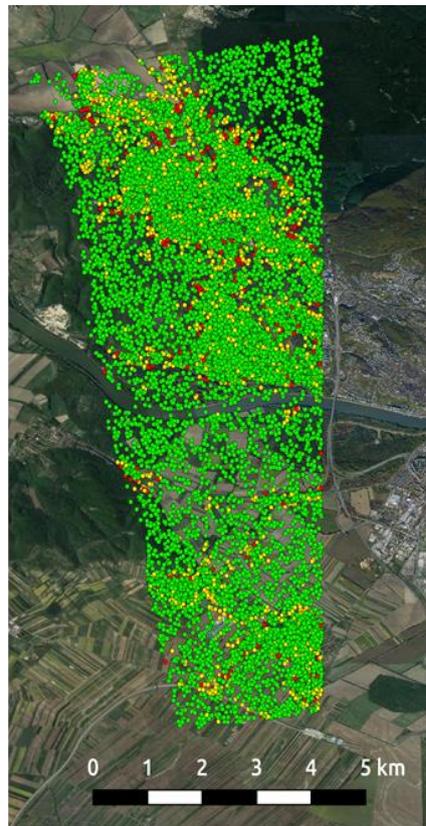


Figura 46d

Figura 46. Mapas de classificação relativos à divisão de dados 3 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos (Figuras 46a a 46d, respetivamente). PSs bem classificados estão coloridos a verde, mal classificados a vermelho, e cuja classificação é incerta estão coloridos a amarelo.

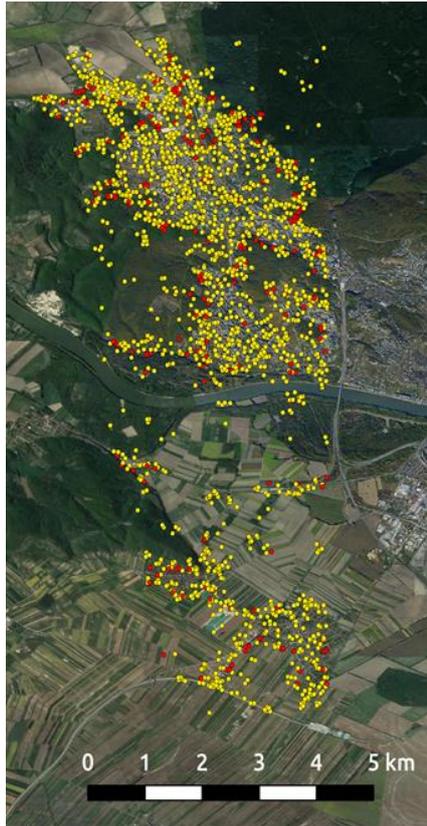


Figura 47a

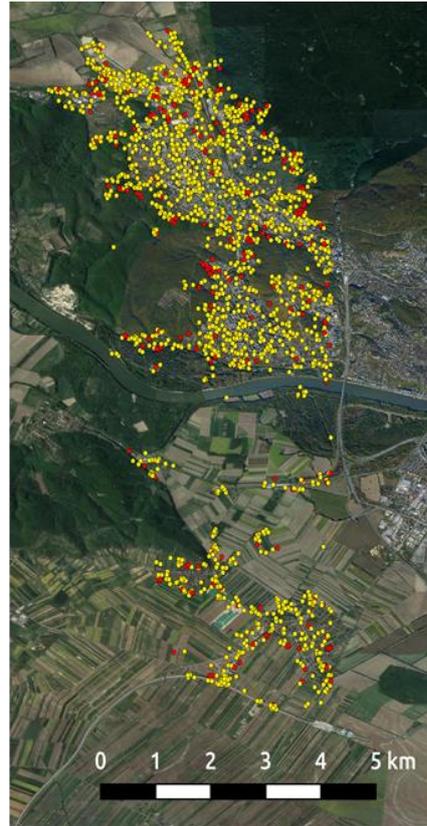


Figura 47b

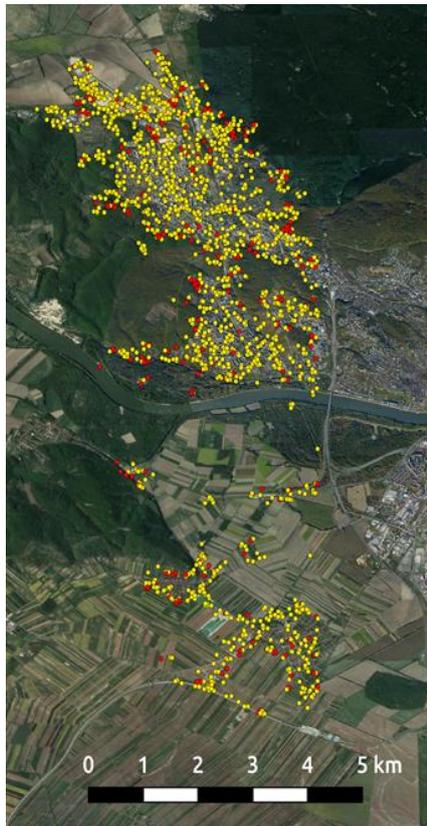


Figura 47c

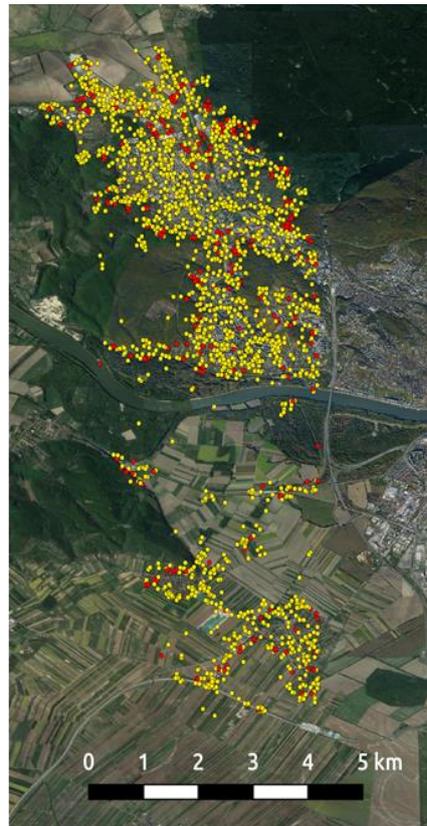


Figura 47d

Figura 47. Segunda versão dos mapas de classificação ilustrados na Figura 46, demonstrando apenas classificações erradas (a vermelho) e classificações incertas (a amarelo). As Figuras 47a a 47d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos.

Resultados relativos aos modelos de classificação treinados na divisão de dados 4

n	PSs outliers	PSs inliers	PSs duvidosos	Total
3	5.39	2.23	5.44	13.06
5	4.26	1.66	5.08	11.00
10	3.07	1.39	5.22	9.68
15	4.27	2.41	4.17	10.85

Tabela 28. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	5143	0	43
	1	0	5254	45
	2	83	5	1850

Tabela 29. Matriz de confusão relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	5516	0	76
	1	0	5373	43
	2	42	7	1851

Tabela 31. Matriz de confusão relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	5313	0	42
	1	0	5339	44
	2	74	2	1904

Tabela 30. Matriz de confusão relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	5290	0	27
	1	0	5215	32
	2	96	18	2061

Tabela 32. Matriz de confusão relativa à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

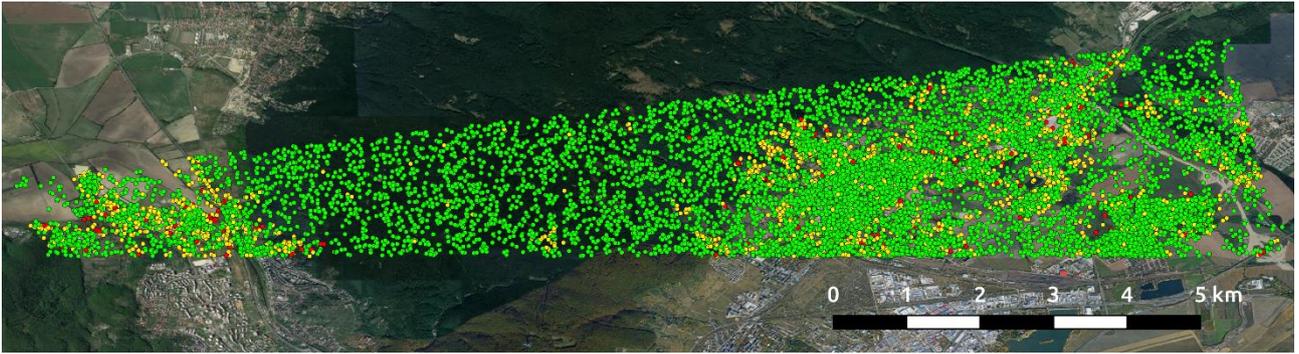


Figura 48a

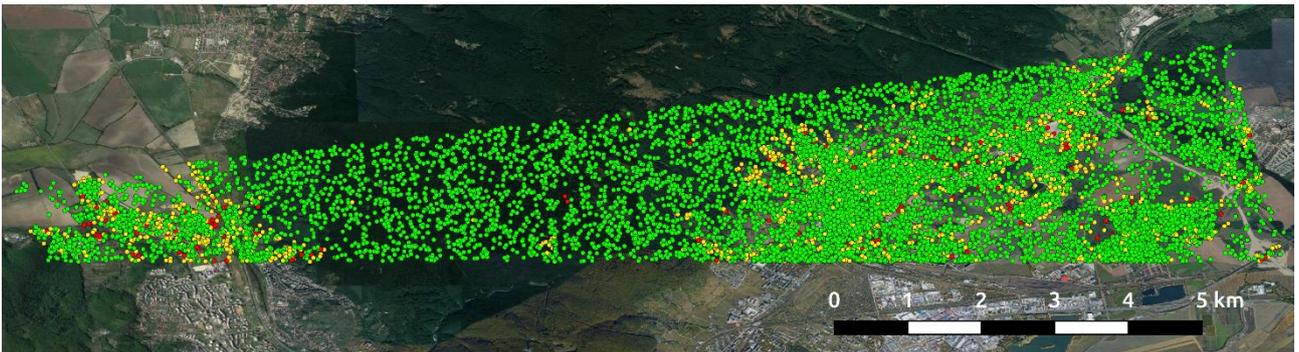


Figura 48b

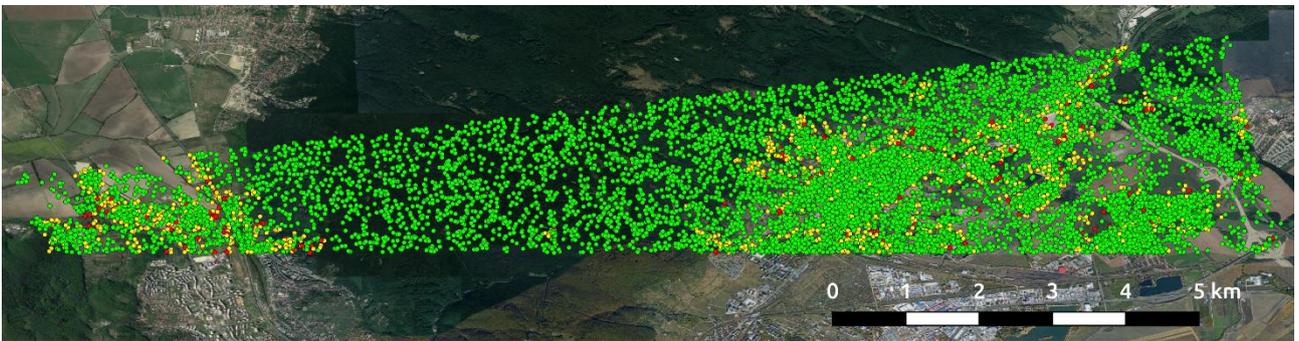


Figura 48c

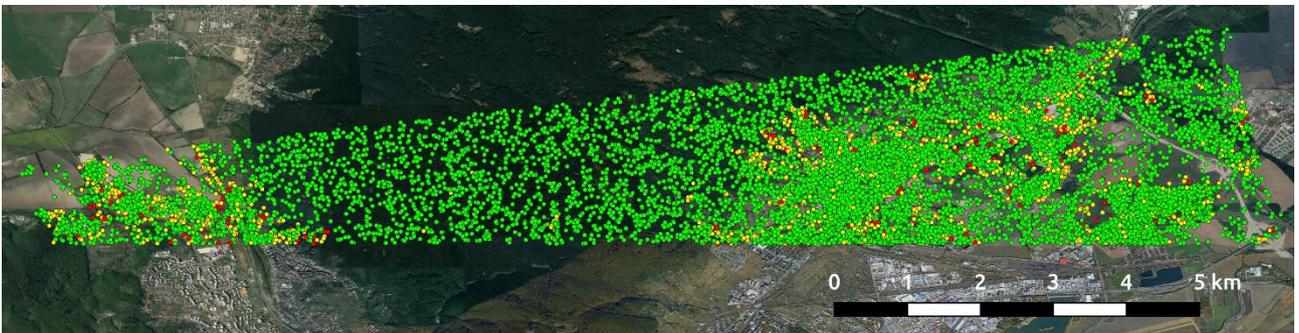


Figura 48d

Figura 48. Mapas de classificação relativos à divisão de dados 4 da base de dados de PSs da cidade de Bratislava, considerando na classificação de cada PS os 3, 5, 10 e 15 PSs vizinhos mais próximos (Figuras 48a a 48d, respetivamente). PSs bem classificados estão coloridos a verde, mal classificados a vermelho, e cuja classificação é incerta estão coloridos a amarelo.

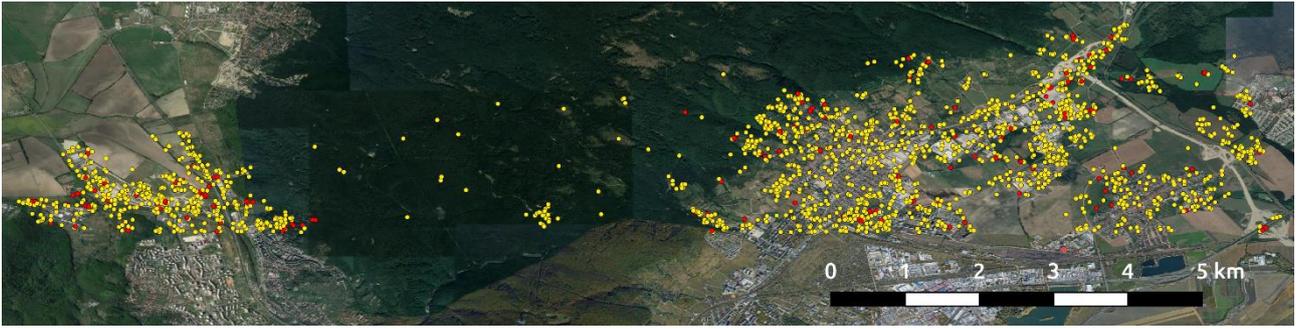


Figura 49a

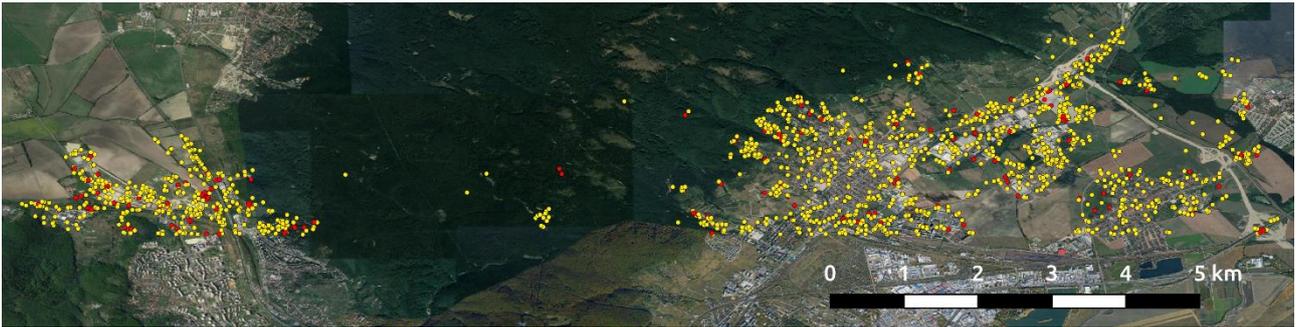


Figura 49b

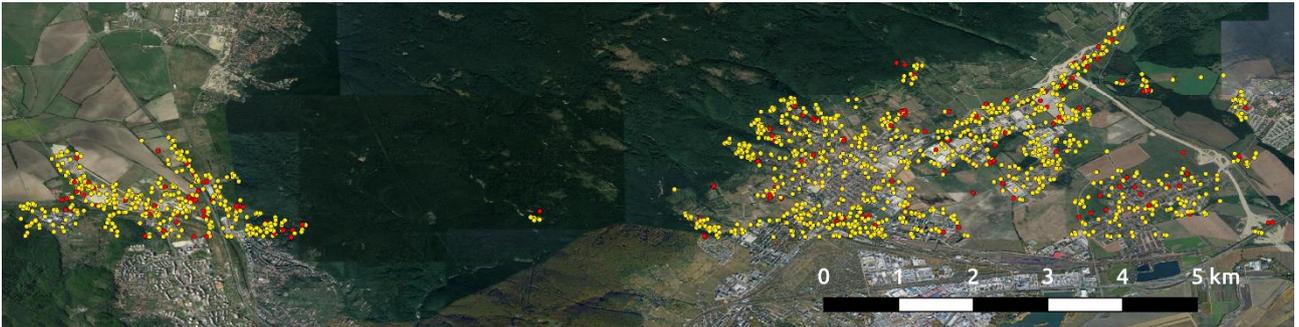


Figura 49c

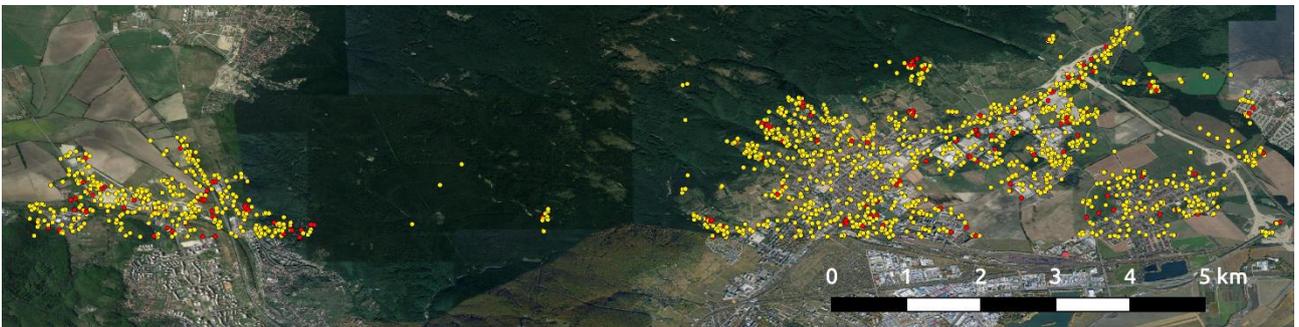


Figura 49d

Figura 49. Segunda versão dos mapas de classificação ilustrados na Figura 48, demonstrando apenas classificações erradas (a vermelho) e classificações incertas (a amarelo). As Figuras 49a a 49d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos.

8.2. Resultados relativos à base de dados de dispersores permanentes dos subúrbios de Prievidza

Resultados relativos aos modelos de classificação treinados na divisão de dados 1 da base de dados de dispersores permanentes da cidade de Bratislava

n	PSs outliers	PSs inliers	PSs duvidosos	Total
3	4.39	11.67	0.94	17.00
5	1.73	14.19	1.22	17.14
10	1.22	32.06	1.44	34.72
15	0.58	23.20	1.08	24.86

Tabela 33. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtida usando os correspondentes modelos de classificação treinados na divisão de dados 1 da base de dados de PSs da cidade de Bratislava.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	445	19	11
	1	0	91	0
	2	48	508	30

Tabela 34. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 1 da base de dados de PSs da cidade de Bratislava. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	524	38	21
	1	0	73	0
	2	6	472	16

Tabela 35. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 1 da base de dados de PSs da cidade de Bratislava. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	531	88	31
	1	0	80	0
	2	6	167	3

Tabela 36. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 1 da base de dados de PSs da cidade de Bratislava. Aos PSs *outliers* foi atribuído o valor 0, aos PSs *inliers* o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	543	190	36
	1	0	57	0
	2	3	211	3

Tabela 37. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 1 da base de dados de PSs da cidade de Bratislava. Aos PSs *outliers* foi atribuído o valor 0, aos PSs *inliers* o valor 1 e aos PSs duvidosos o valor 2.

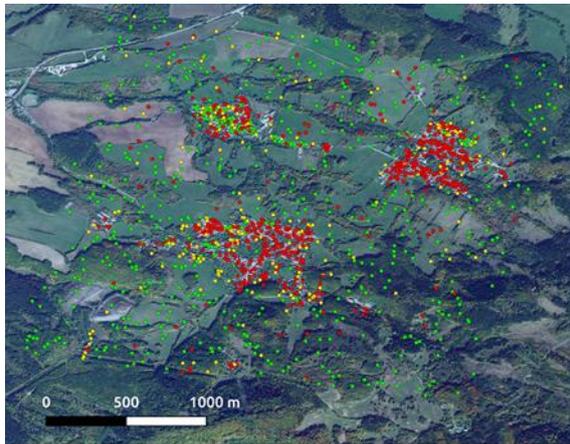


Figura 50a

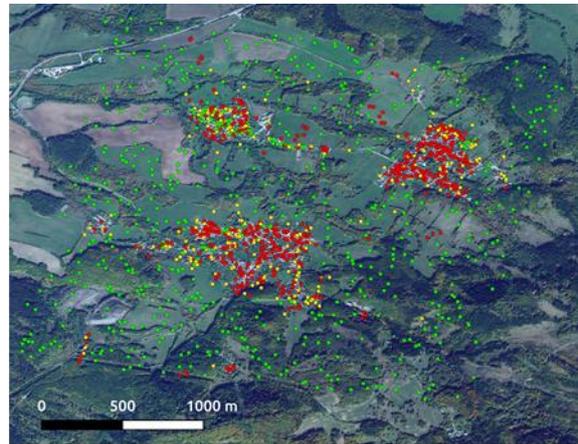


Figura 50b

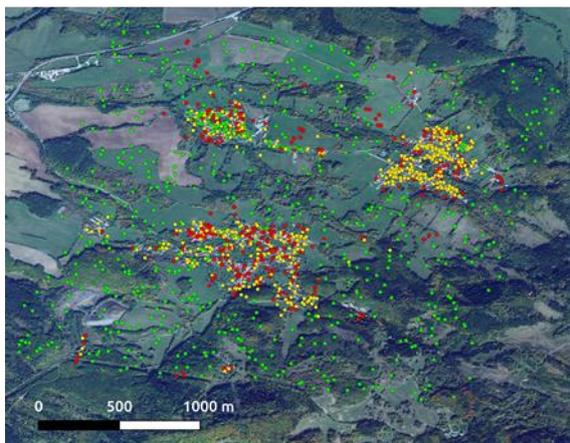


Figura 50c

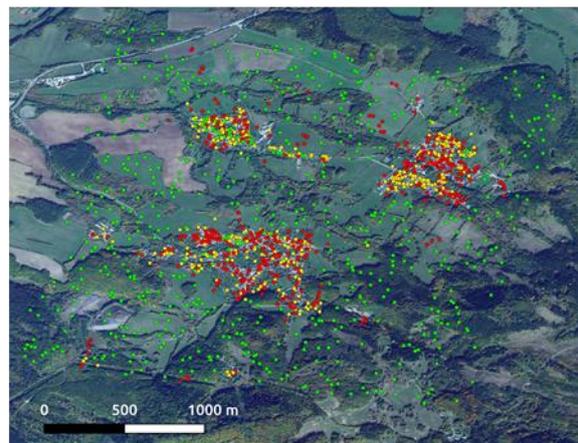


Figura 50d

Figura 50. Mapas de classificação relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtidos usando os modelos de classificação treinados na divisão de dados 1 da base de dados de PSs da cidade de Bratislava. As Figuras 50a a 50d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos. PSs bem classificados estão coloridos a verde, mal classificados a vermelho, e cuja classificação é incerta estão coloridos a amarelo.

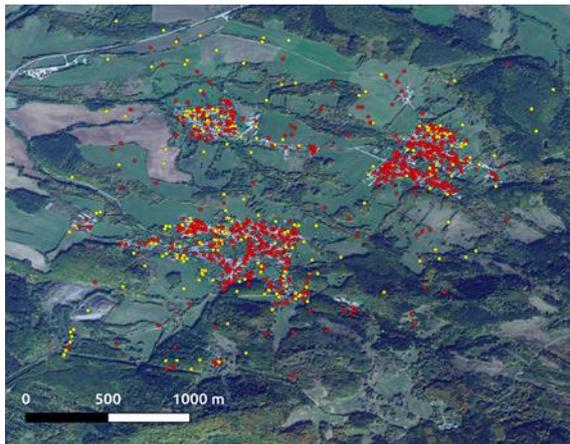


Figura 51a

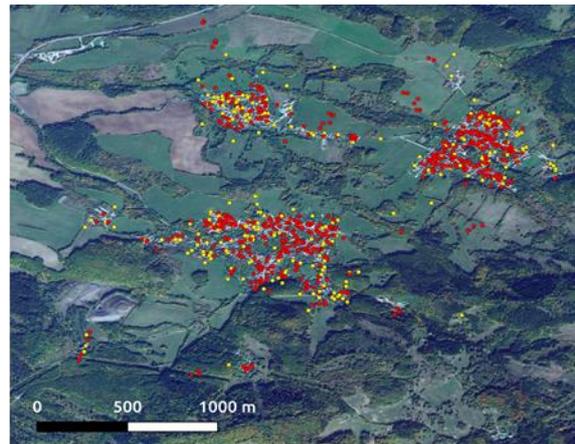


Figura 51b

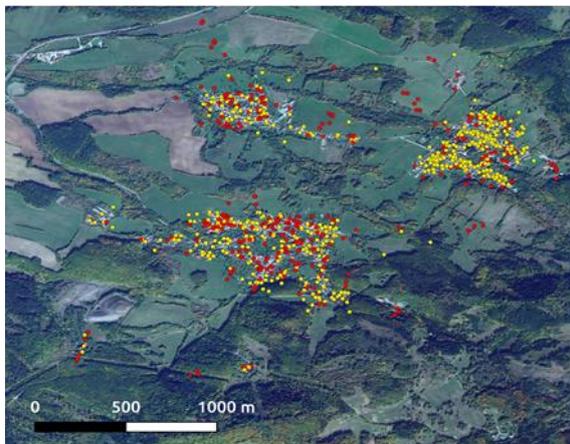


Figura 51c

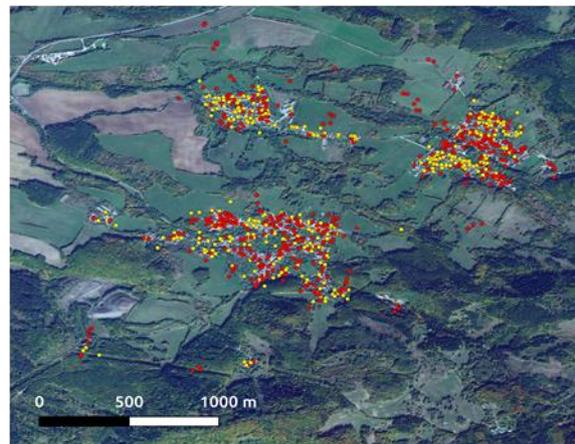


Figura 51d

Figura 51. Segunda versão dos mapas de classificação ilustrados na Figura 50, demonstrando apenas classificações erradas (a vermelho) e classificações incertas (a amarelo). As Figuras 51a a 51d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos.

Resultados relativos aos modelos de classificação treinados na divisão de dados 2 da base de dados de dispersores permanentes da cidade de Bratislava

n	PSs outliers	PSs inliers	PSs duvidosos	Total
3	1.15	15.06	1.59	17.80
5	1.15	14.55	1.22	16.92
10	1.30	18.80	1.30	21.40
15	0.65	19.02	1.01	20.68

Tabela 38. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtida usando os correspondentes modelos de classificação treinados na divisão de dados 2 da base de dados de PSs da cidade de Bratislava.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	535	29	23
	1	0	85	0
	2	3	457	9

Tabela 39. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 2 da base de dados de PSs da cidade de Bratislava. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	533	39	23
	1	0	79	0
	2	5	460	14

Tabela 40. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 2 da base de dados de PSs da cidade de Bratislava. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	533	51	27
	1	0	52	0
	2	3	416	9

Tabela 41. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 2 da base de dados de PSs da cidade de Bratislava. Aos PSs *outliers* foi atribuído o valor 0, aos PSs *inliers* o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	539	101	36
	1	0	75	0
	2	6	340	4

Tabela 42. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 2 da base de dados de PSs da cidade de Bratislava. Aos PSs *outliers* foi atribuído o valor 0, aos PSs *inliers* o valor 1 e aos PSs duvidosos o valor 2.

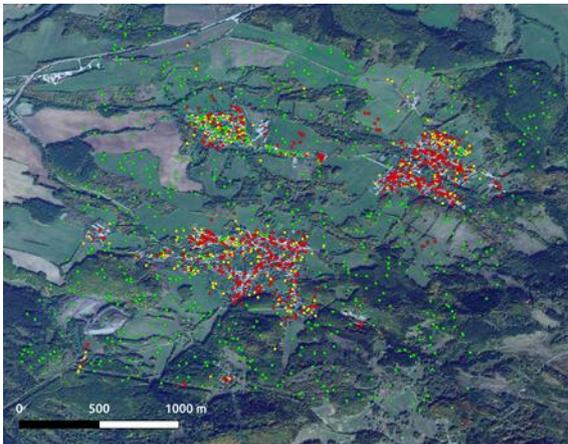


Figura 52a

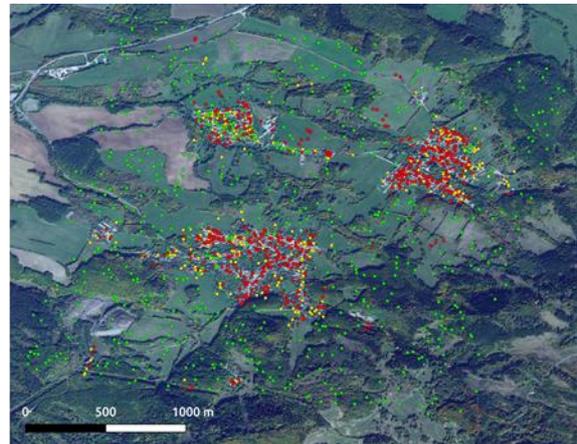


Figura 52b

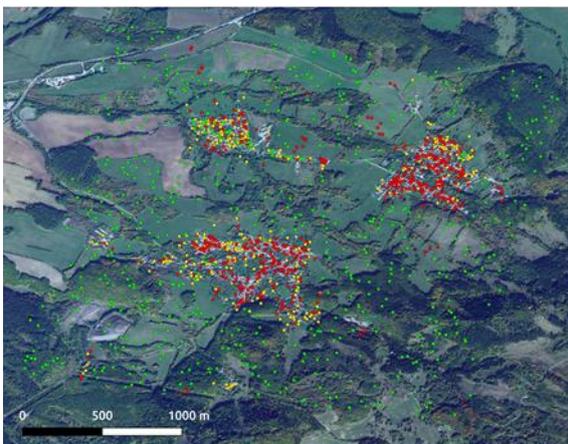


Figura 52c

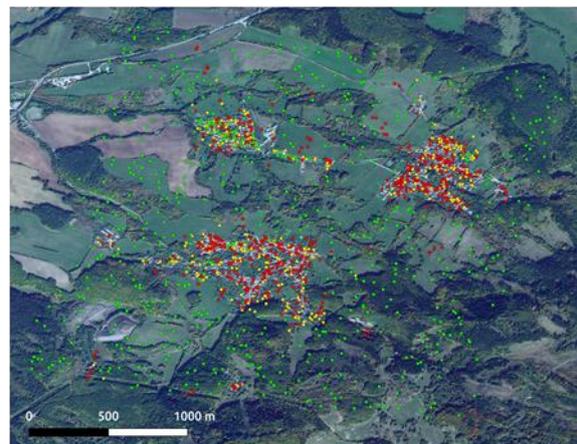


Figura 52d

Figura 52. Mapas de classificação relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtidos usando os modelos de classificação treinados na divisão de dados 2 da base de dados de PSs da cidade de Bratislava. As Figuras 52a a 52d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos. PSs bem classificados estão coloridos a verde, mal classificados a vermelho, e cuja classificação é incerta estão coloridos a amarelo.

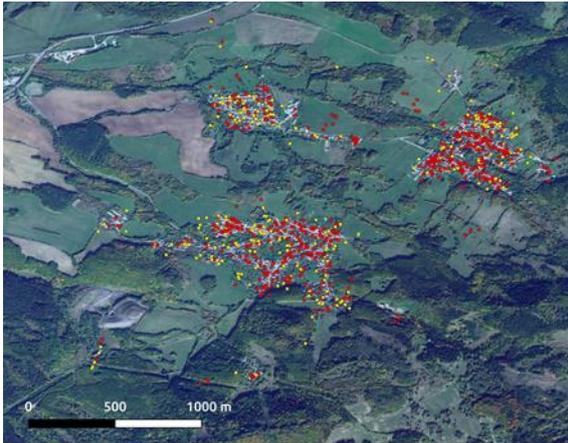


Figura 53a

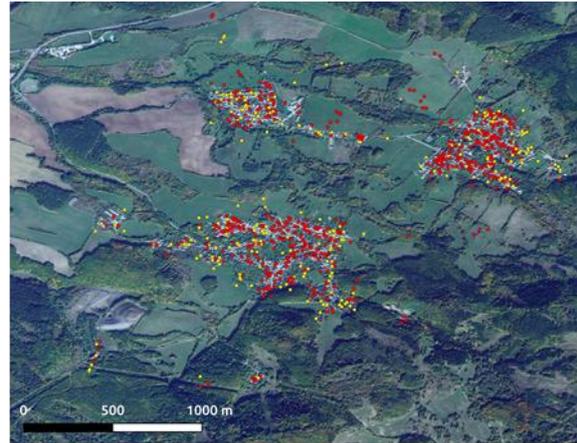


Figura 53b

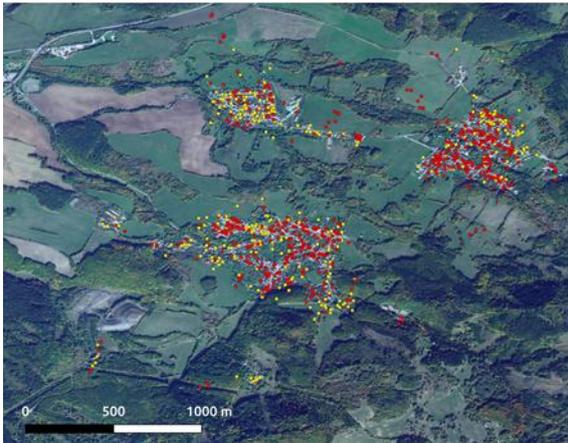


Figura 53c

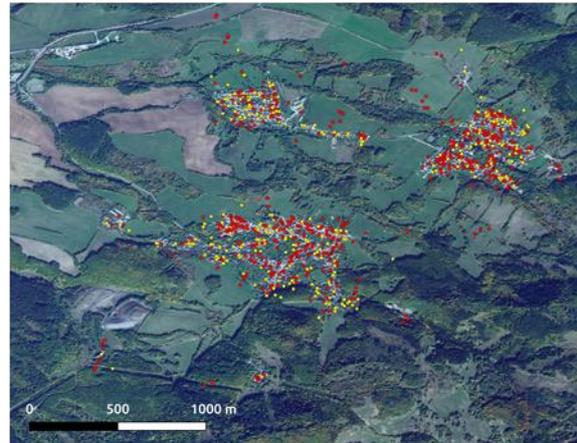


Figura 53d

Figura 53. Segunda versão dos mapas de classificação ilustrados na Figura 52, demonstrando apenas classificações erradas (a vermelho) e classificações incertas (a amarelo). As Figuras 53a a 53d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos.

Resultados relativos aos modelos de classificação treinados na divisão de dados 3 da base de dados de dispersores permanentes da cidade de Bratislava

n	PSs outliers	PSs inliers	PSs duvidosos	Total
3	2.59	13.69	1.51	17.79
5	1.01	14.70	0.94	16.65
10	0.72	17.80	0.79	19.31
15	3.31	16.21	1.44	20.96

Tabela 43. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtida usando os correspondentes modelos de classificação treinados na divisão de dados 3 da base de dados de PSs da cidade de Bratislava.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	510	20	14
	1	0	83	0
	2	8	487	19

Tabela 44. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 3 da base de dados de PSs da cidade de Bratislava. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	535	44	26
	1	0	72	0
	2	5	460	15

Tabela 45. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 3 da base de dados de PSs da cidade de Bratislava. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	540	140	37
	1	0	90	0
	2	4	303	6

Tabela 46. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 3 da base de dados de PSs da cidade de Bratislava. Aos PSs *outliers* foi atribuído o valor 0, aos PSs *inliers* o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	474	42	17
	1	0	48	0
	2	34	465	17

Tabela 47. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 3 da base de dados de PSs da cidade de Bratislava. Aos PSs *outliers* foi atribuído o valor 0, aos PSs *inliers* o valor 1 e aos PSs duvidosos o valor 2.

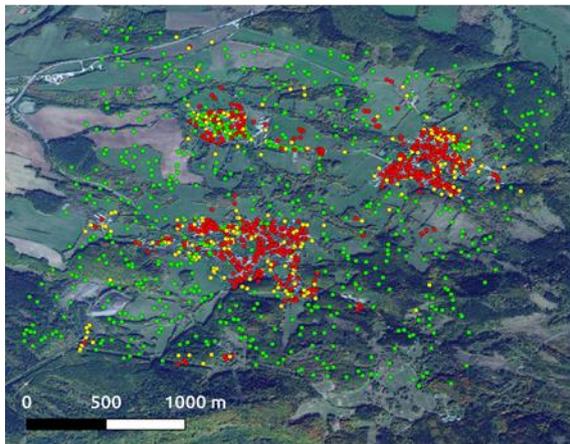


Figura 54a

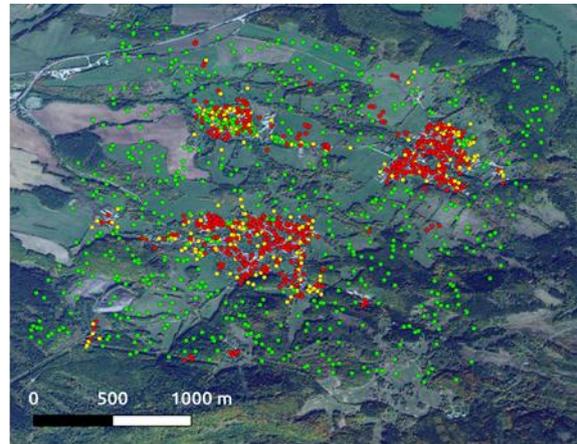


Figura 54b

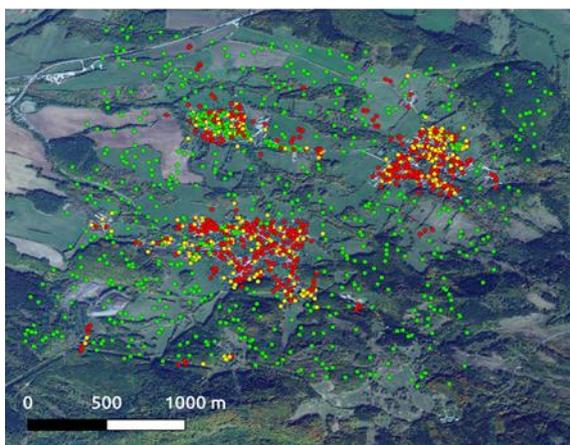


Figura 54c

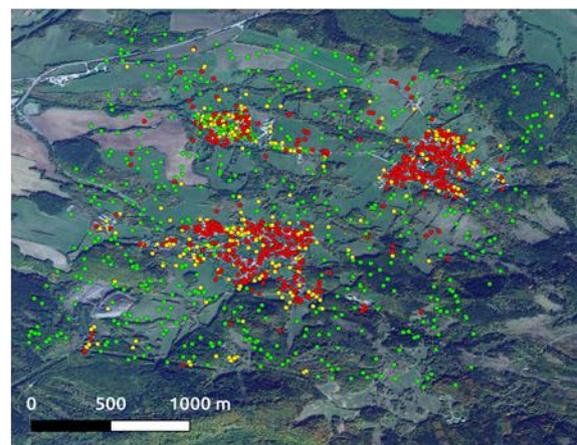


Figura 54d

Figura 54. Mapas de classificação relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtidos usando os modelos de classificação treinados na divisão de dados 3 da base de dados de PSs da cidade de Bratislava. As Figuras 54a a 54d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos. PSs bem classificados estão coloridos a verde, mal classificados a vermelho, e cuja classificação é incerta estão coloridos a amarelo.

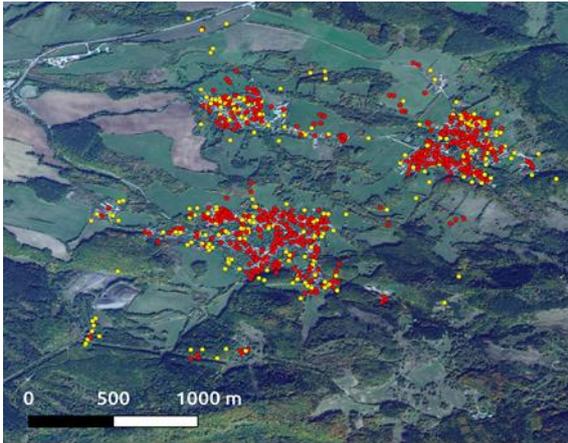


Figura 55a

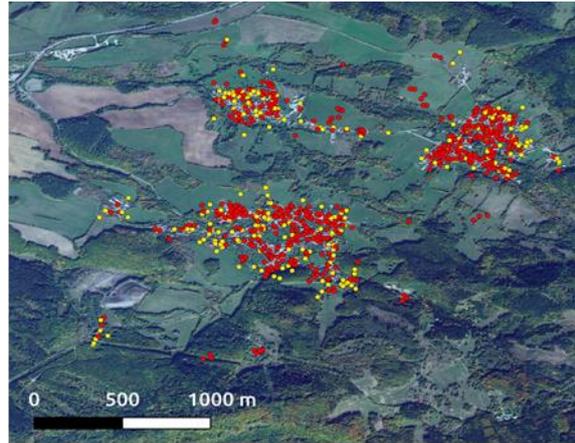


Figura 55b

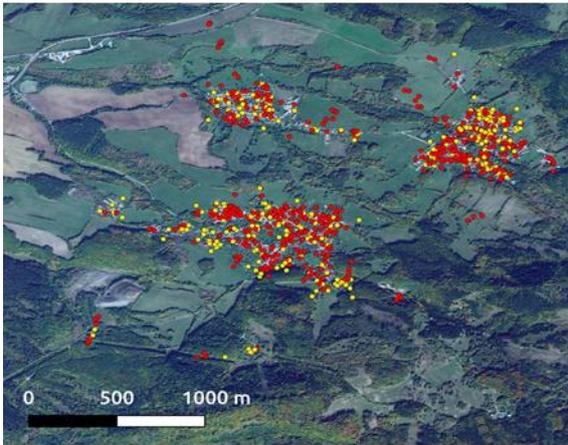


Figura 55c

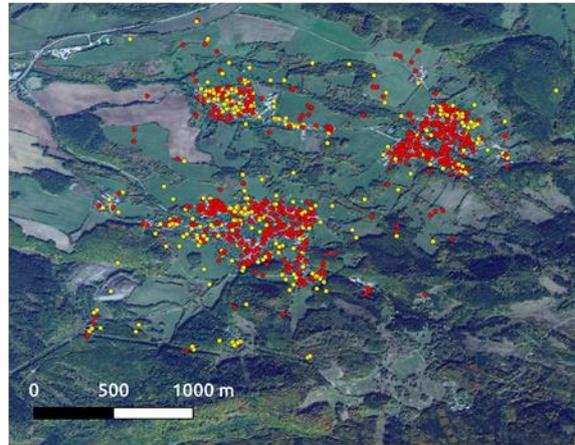


Figura 55d

Figura 55. Segunda versão dos mapas de classificação ilustrados na Figura 54, demonstrando apenas classificações erradas (a vermelho) e classificações incertas (a amarelo). As Figuras 55a a 55d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos.

Resultados relativos aos modelos de classificação treinados na divisão de dados 4 da base de dados de dispersores permanentes da cidade de Bratislava

n	PSs outliers	PSs inliers	PSs duvidosos	Total
3	1.80	14.27	1.30	17.37
5	0.94	17.80	1.01	19.75
10	0.50	19.02	0.65	20.17
15	1.80	16.50	0.79	19.09

Tabela 48. Percentagem de classificações incertas por classe, em função do número de PSs vizinhos mais próximos considerados na classificação de cada PS, relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtida usando os correspondentes modelos de classificação treinados na divisão de dados 4 da base de dados de PSs da cidade de Bratislava.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	525	28	24
	1	0	84	0
	2	4	470	12

Tabela 49. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 3 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 4 da base de dados de PSs da cidade de Bratislava. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	537	61	30
	1	0	82	0
	2	4	390	10

Tabela 50. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 5 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 4 da base de dados de PSs da cidade de Bratislava. Aos PSs outliers foi atribuído o valor 0, aos PSs inliers o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	546	111	42
	1	0	76	0
	2	1	329	3

Tabela 51. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 10 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 4 da base de dados de PSs da cidade de Bratislava. Aos PSs *outliers* foi atribuído o valor 0, aos PSs *inliers* o valor 1 e aos PSs duvidosos o valor 2.

		Classes Verdadeiras		
		0	1	2
Classes Previstas	0	514	137	30
	1	0	75	0
	2	15	339	13

Tabela 52. Matriz de confusão relativa à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, considerando na classificação de cada PS os 15 PSs vizinhos mais próximos, obtida usando o correspondente modelo de classificação treinado na divisão de dados 4 da base de dados de PSs da cidade de Bratislava. Aos PSs *outliers* foi atribuído o valor 0, aos PSs *inliers* o valor 1 e aos PSs duvidosos o valor 2.

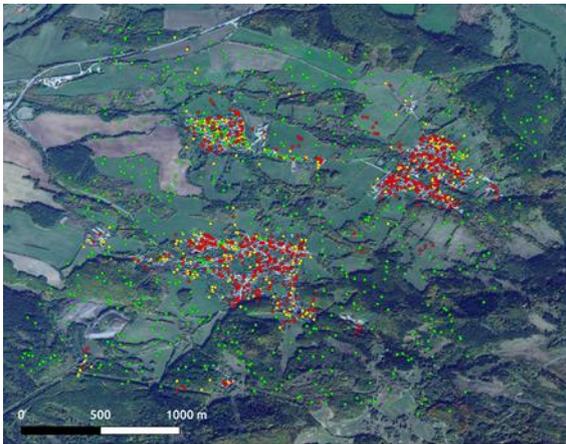


Figura 56a

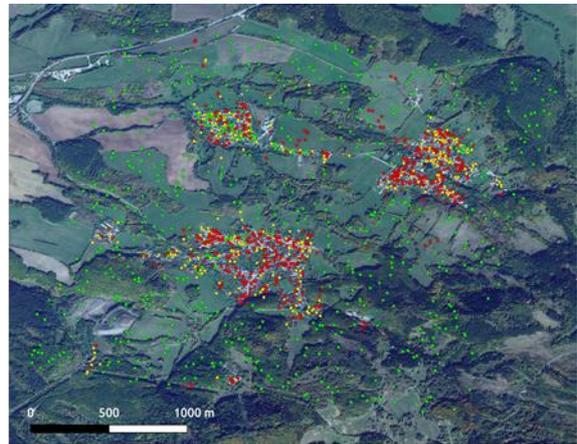


Figura 56b



Figura 56c

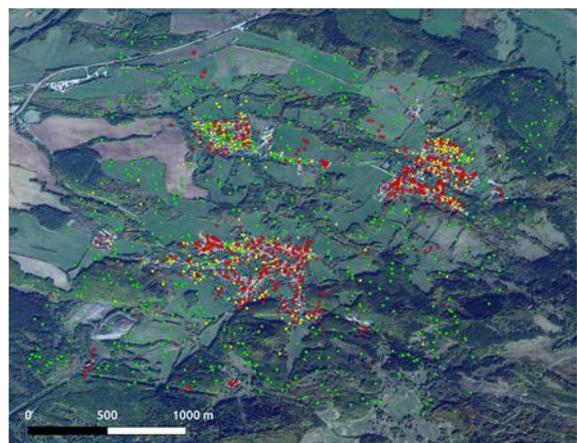


Figura 56d

Figura 56. Mapas de classificação relativos à base de dados de PSs da região de Velka Lehotka, Hradec e Mala Lehotka, nos subúrbios de Prievidza, obtidos usando os modelos de classificação treinados na divisão de dados 4 da base de dados de PSs da cidade de Bratislava. As Figuras 56a a 56d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos. PSs bem classificados estão coloridos a verde, mal classificados a vermelho, e cuja classificação é incerta estão coloridos a amarelo.

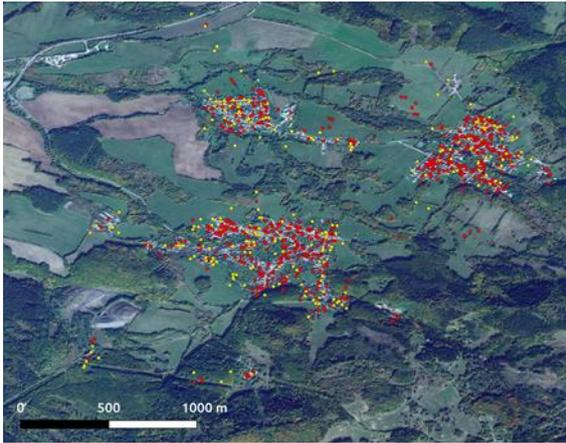


Figura 57a

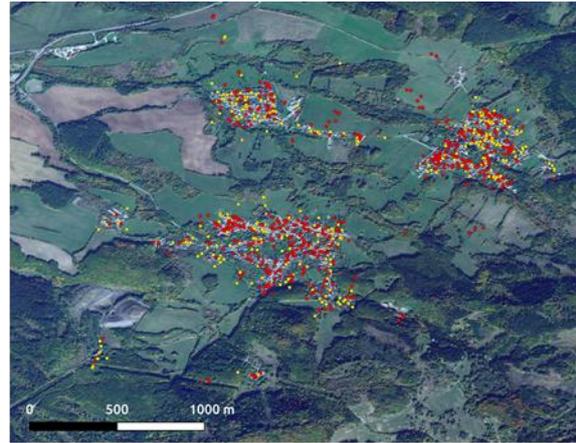


Figura 57b

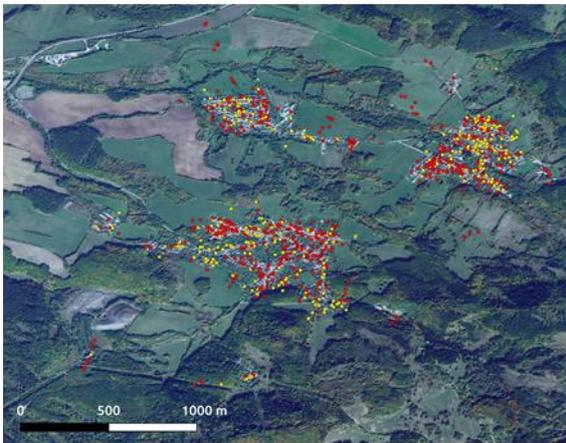


Figura 57c

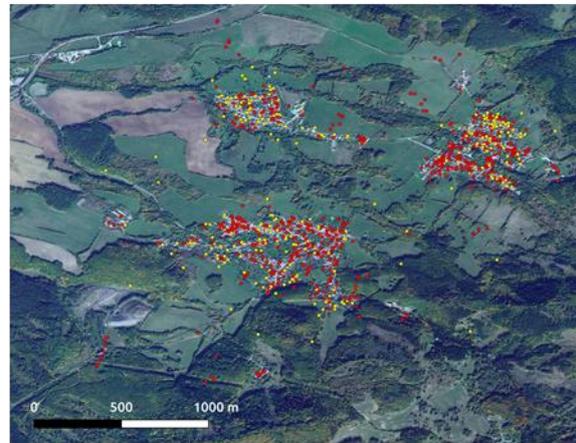


Figura 57d

Figura 57. Segunda versão dos mapas de classificação ilustrados na Figura 56, demonstrando apenas classificações erradas (a vermelho) e classificações incertas (a amarelo). As Figuras 57a a 57d correspondem a considerar na classificação de cada PS, respetivamente, os 3, 5, 10 e 15 PSs vizinhos mais próximos.