





Ci. e Nat., Santa Maria, v. 46, spe. n. 1, e87230, 2024 • https://doi.org/10.5902/2179460X87230 Submitted: 22/03/2024 • Approved: 25/06/2024 • Published: 04/11/2024

Special edition ERMAC e ENMC

# Accuracy estimates of land use and land cover classification are associated with the sensitivity of the MAXVER classifier and the holdout subsampling technique on allotments in Pelotas-RS

Estimativas da acurácia da classificação do uso e cobertura da terra associadas a sensibilidade do classificador MAXVER e da técnica de subamostragem holdout sobre loteamentos em Pelotas-RS

### Everton Rodrigues Zirbes'<sup>®</sup>, Daniela Buske'<sup>®</sup>, Diuliana Leandro'<sup>®</sup>, Andrea Souza Castro'<sup>®</sup>

<sup>I</sup> Universidade Federal de Pelotas, Pelotas, RS, Brasil

### ABSTRACT

Urban growth in the Pelotas-RS community has intensified in recent years. The number of new subdivisions and housing developments available in the city's real estate market is a testament to this trend. With this growth, potential problems related to flooding and possible inundation become a cause for concern given the topographical characteristics of the city. One way to monitor this urban expansion is through remote sensing analysis, which provides a wide range of statistical information, including precision and accuracy indices obtained from land use and land cover classification methods. The KAPPA test, for instance, has proven to be very efficient in analyzing areas of impervious surfaces, loss of vegetation cover, etc. These values are important for carrying out urban drainage calculations to determine the dimensioning of rainwater systems. In this case, the study focused on Pelotas. The classification procedure achieved an accuracy of over 90%, which is considered excellent for this type of interpretative analysis.

Keywords: Accuracy; KAPPA statistics; Remote sensing; SIG

#### RESUMO

O crescimento urbano no município de Pelotas-RS vem intensificando-se nos últimos anos. Essa tendência é evidenciada pelo número de novos loteamentos e empreendimentos habitacionais disponibilizados no mercado imobiliário da cidade. Com esse crescimento, tornam-se motivos de preocupação os possíveis problemas relacionados a ocorrência de alagamentos e eventuais inundações, tendo em vista



as características topográficas do município. Uma das formas de se monitorar essa expansão urbana é por intermédio de análises via sensoriamento remoto, sobre as quais se obtêm ampla gama de informações estatísticas, dentre elas os índices de precisão e exatidão extraídos através de processos de classificação de uso e ocupação de terra. É o caso do Teste KAPPA que se mostra muito eficiente em análises de áreas de impermeabilização da superfície, perda da cobertura vegetal, entre outros. Essas taxas são importantes para a realização de cálculos de drenagem urbana utilizados na determinação do dimensionamento das águas pluviais. Neste estudo de caso voltado a Pelotas, o processo de classificação obteve desempenho superior a 90% nos índices de acurácia, sendo considerados excelentes para esse tipo de análise interpretativa.

Palavras-chave: Acurácia; Estatística KAPPA; Sensoriamento remoto; SIG

### **1 INTRODUCTION**

Environmental monitoring through remote sensing provides data for various studies, such as spatial planning, risk and hazard area assessment, sustainable development goals, and more. It is important to note that changes in settlement forms in Brazil are neither linear over time nor homogeneous across the country. The economic, environmental, historical, and cultural factors influenced these changes (IBGE, 2020). According to Zirbes et al. (2022), in the case of the city of Pelotas-RS, the city's real estate market has seen an increase in new residential areas and available housing, reflecting recent urban growth. With this expansion, potential problems with flooding and inundation become a cause for concern given the topographical characteristics of the city, where 93% of the population lives in the urban area according to the 2010 census.

Tucci (2002) notes that the growth of the urban population in recent years has had a direct impact on the infrastructure of water resources. One of the most affected areas is the urban drainage system, leading to an increase in flooding. The author points out the changes in land cover and the hygienist perspective, commonly implemented in underdeveloped nations and accompanied by rising pollution levels, are causing modifications in the urban water cycle.

In this context, one way to identify land use and occupancy patterns is through the supervised comparison and classification of multispectral and hyperspectral satellite

images from various series. This type of data analysis involves checking the overall performance of land classification using test and training samples, i.e., the accuracy rate in identifying a particular class where the corresponding pixels represent natural or manufactured surface features.

The KAPPA test, also known as Cohen's KAPPA coefficient, is a method for measuring the accuracy of image classification. Based on Congalton (1991), this test is a discrete multivariate technique used to evaluate accuracy. The result is a KHAT statistic that estimates KAPPA, and this is a measure of agreement or precision calculated using the following equation:

$$\widehat{K} = N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} * x_{+i}) / N_2 \sum_{i=1}^{r} (x_{i+} * x_{+i}),$$
(1)

Where r is the number of rows in the matrix,  $x_{ii}$  is the number of observations in row and column *i*,  $x_{i+}$  and  $x_{i+}$  are the marginal sums for row and column *i* respectively and *N* is the total number of observations.

According to Congalton (1991), if an insufficient sample is collected, many zeros will appear in an error matrix, or if the classification was done exceptionally well. Still, according to the author, positions where zeros are expected tend to receive positive values in offdiagonal cells during the normalization process due to the iterative proportional fitting routine. This happens when there is spatial autocorrelation in the presence, absence, or degree of a particular feature that affects the presence, absence, or degree of the same feature in the neighboring unit; technically, this is called spectral confusion.

As per Chowdhury and Hafsa's (2022) research, the confusion matrix derived from the KAPPA coefficient is a measure of the accuracy of land cover mapping, which is essential for validating the accuracy of remote sensing image classification. Furthermore, as Bradter et al. (2020) note, the KAPPA test is a precision measure widely used in remote sensing that provides useful information, as the degree of agreement varies with the number of categories used and the different image resolutions.

Ci e Nat., Santa Maria, v. 46 spe. n. 1, e87230, 2024

The present article aims to evaluate the accuracy of the confusion matrices obtained through the classification of land use and cover during the urbanization process from 2010 to 2022, using orbital images of recent subdivisions in the neighborhood of Laranjal, in the city of Pelotas, Rio Grande do Sul, Brazil.

# 2 METHODOLOGY

Initially, the study areas of interest were defined, and the image classification method was applied. The study areas appear highlighted in Figure 1, polygonized in red (area 1) and magenta (area 2). They encompass approximately 253 hectares, comprising pre-existing subdivisions, condominiums, adjacent green areas, and those built over a period of 13 years (2010–2022). They are the following: São Conrado, Condominium Veredas — Altos do Laranjal, and Condominium Alphaville (area 1); as well as Amarílis and Vila Mariana subdivision (area 2), located in the Laranjal neighborhood.

From the orbital images of the defined study areas, georeferenced with 8-bit spectral resolution in the visible spectrum (RGB), and spatial resolutions of 65x65 cm and 41x41 cm, obtained from Google Earth PRO, it was possible to perform a temporal analysis spanning from 2010 to 2022. The analysis development was conducted using the MultiSpec Application GIS tool, which provides a processing system for interactive analysis of data from multispectral and hyperspectral Earth images obtained from aerial and space systems. Among the different statistical classification methods available in this tool, Gaussian Maximum Likelihood or MAXVER classification was chosen, applied to repeated random subsampling sets for testing and training (holdout), due to the high performance observed in preliminary classification tests of the study images.

According to Lee and Landgrebe (1993), in remote sensing, in the process of modeling classes from a limited number of samples, one must determine with the greatest precision what the location and spatial distribution of each class of samples

Ci e Nat., Santa Maria, v. 46 spe. n. 1, e87230, 2024

is likely to be. One advantage of using the Gaussian model is that it requires smaller training sets to define the desired classes accurately, especially when the spectral dimensionality is large because it uses first and second order statistics, both the mean vector and the matrix correlation or covariance. The class mean vector defines the location of the class centroid in N-dimensional space.

Second-order statistics define the shape and orientation of the class distribution. To put it another way, to train a classifier effectively, it is necessary to establish a comprehensive list of classes based on their distribution in feature space. This list should include training samples for each class that meet the following criteria: exhaustiveness, meaning that there is a logical class to assign each pixel in the scene; adequate separability using available resources; and informative value, meaning that the classes are of interest to the analyst.



Figure 1 – Location Map of Study Areas in Pelotas, Rio Grande do Sul

# **3 RESULTS AND DISCUSSIONS**

The following confusion matrices graphs displaying the performance of classes for the two areas of the study, divided between Classification of Testing and Training Fields analysis conducted for the years 2010 and 2022.



Figure 2 – Confusion matrix graph for area 1 training classification in 2010

Source: Authors (2023)

Figure 3 – Confusion matrix graph for area 1 test classification in 2010



Source: Authors (2023)







Figure 5 – Confusion matrix graph for area 1 test classification in 2022

Source: Authors (2023)

Figure 6 – Confusion matrix graph for area 2 training classification in 2010



Source: Authors (2023)

Figure 7 – Confusion matrix graph for area 2 test classification in 2010



Source: Authors (2023)





Figure 9 – Confusion matrix graph for area 2 test classification in 2022



Source: Authors (2023)

Figure 10 - Comparative map of urban expansion over the study areas in Pelotas-RS



The confusion matrices show that the accuracy and performance indices for land use and cover classification remained consistently high for both the training and test samples, except for the classification between Brasilit-tiles (*Telha-Brasilit*) and Pavements (*pavimentos*), and between Arboreal (*Arbórea*) and Vegetation-Field (*Veg-Campestre*). This is due to the spectral similarity between these classes, resulting in high color correlation. Despite this, accuracy percentages remained above 80% in the majority of cases. Both the KAPPA index and the class performance index remained above 93% in all tests, indicating high performance.

As can be seen in Figure 10, a large difference is noticeable in relation to build areas due to the process of intensification of urban expansion in the region when comparing the years 2010 and 2022, which marks a period of new construction projects being implemented in the Laranjal neighborhood.

Figure 11 demonstrates the high accuracy of the image classification process due to the abundance of observable features.



Figure 11 – Reclassified Map of Land Use and Occupation study areas in Pelotas-RS

# **4 CONCLUSIONS**

The results of the study show that the use of statistical analysis via the KAPPA index extracted with the GIS MultiSpec Application Tool was very satisfactory during the classification process of land use and land cover. This method accurately represented the features present in the scene, especially in the absence of an infrared spectral band. Such a band would have facilitated the detection of certain targets on the surface due to their physical characteristics, which differ in the intensity of interaction with the incident energy. When using a visible composition, the accuracy of the analysis depends on the analyst's familiarity with interpreting orbital images in that composition and the configuration of the GIS used.

# ACKNOWLEDGEMENTS

This paper was carried out with support from the Coordination for the Improvement of Higher Education Personnel — Brazil (CAPES) — Financing Code 001, from CNPq, for the subsidy and granting of scholarships and, finally, from to the Postgraduate Program in Environmental Sciences at the Federal University of Pelotas.

## REFERENCES

- Bradter, U. et al. (2020). Classifying grass-dominated habitats from remotely sensed data: The influence of spectral resolution, acquisition time and the vegetation classification system on accuracy and thematic resolution. *Science of The Total Environment*, 711, 134584, DOI: https: //doi .org/10.1016/j.scitotenv.2019.134584.
- Chowdhury, M.S., Hafsa, B. (2022). Multi-decadal land cover change analysis over Sundarbans Mangrove Forest of Bangladesh: A GIS and remote sensing based approach. *Global Ecology and Conservation*. 37, e0215. DOI: https://doi.org/10.1016/j.gecco.2022. e02151.
- Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment,* 37, 35-46. DOI https://doi-org.ez66. periodicos.capes.gov.br/10.1016/0034-4257(91)90048-B.
- IBGE. (2020). Monitoramento da cobertura e uso da terra do Brasil: 2016 2018. *Coleção IBGEANA*. Rio de Janeiro (Brasil): Biblioteca IBGE.

Ci e Nat., Santa Maria, v. 46 spe. n. 1, e87230, 2024

- Lee, C., Landgrebe, D. A. (1993). Analyzing High-dimensional Multispectral Data. *IEEE Transactions* on Geoscience and Remote Sensing, 31(4), 792-800. DOI: 10.1109/36.239901.
- Tucci, C. E. M. (2002). Gerenciamento da Drenagem Urbana. *Revista Brasileira de Recursos* Hídricos, 7.(1), 5-27. DOI10.21168/rbrh.v7n1.p5-27.
- Zirbes, E. R. et al. (2022). Análise Temporal dos Índices de Impermeabilização Urbana em Loteamentos no Município de Pelotas - RS. Anais do XXXI Congresso de Iniciação Científica – 8ª *SIIEPE*, Pelotas, Brasil.

# Authorship contributions

### 1 – Everton Rodrigues Zirbes

Universidade Federal de Pelotas - M.Sc. in Environmental Sciences https://orcid.org/0000-0002-5711-7563 - thor.zirbes@gmail.com Contribution: Conceptualization, Methodology, Resources, Investigation, Data curation, Visualization, Validation, Software, Writing – original draft, Writing – original draft

### 2 – Daniela Buske

Universidade Federal de Pelotas - PhD in Mechanical Engineer https://orcid.org/0000-0002-4573-9787 - danielabuske@gmail.com Contribution: Supervision, Funding acquisition, Project administration, Writing – review & editing

### 3 – Diuliana Leandro

Universidade Federal de Pelotas - PhD in Geodetic Sciences. https://orcid.org/0000-0002-8092-5550 - diuliana.leandro@gmail.com Contribution: Methodology, Formal analysis, Supervision, Visualization, Validation, Software, Writing – review & editing

### 4 – Andrea Souza Castro

Universidade Federal de Pelotas - PhD in Water Resources https://orcid.org/0000-0003-1989-684X - andreascastro@gmail.com Contribution: Supervision, Formal analysis, Visualization, Validation, Writing – review & editing

## How to quote this article

Zirbes, E. R., Buske, D., Leandro, D., & Castro, A. S. (2024). Accuracy estimates of land use and land cover classification are associated with the sensitivity of the MAXVER classifier and the holdout subsampling technique on allotments in Pelotas-RS. *Ciência e Natura*, 46, spe. 1, e87230. https://doi.org/10.5902/2179460X87230.