Statistical analysis and characterization of Landsat 8 satellite images of forest wildfires regions

Análisis estadístico y caracterización de imágenes de satélite Landsat 8 sobre regiones con incendios en bosques

HERNANDEZ-LOPEZ, Sandra Paola †*[1], YAÑEZ-VARGAS, Juan Israel[1], GONZÁLEZ-RAMIREZ, Andrea[2] and TORRES-ROMAN, Deni[2]

[1]Universidad Politécnica de Juventino Rosas, Master's degree in intelligent systems engineering, Mexico.
[2]CINVESTAV del IPN, Unidad Guadalajara, Telecommunications Laboratory, Mexico.

ID 1st Author: *Sandra Paola, Hernández-López* / **ORC ID:** 0000-0003-2312-2695 **CVU CONACYT ID:** 1002324

ID 1st Coauthor: *Juan Israel, Yañez-Vargas* / **ORC ID:** 0000-0001-5749-8442, **CVU CONACYT ID:** 295711

ID 2nd Coauthor: *Andrea, Gonzalez-Ramirez* / **ORC ID:** 0000-0001-9961-4763, **CVU CONACYT ID:** 1001207

ID 3rd Coauthor: *Deni, Torres-Roman* / **ORC ID:** 0000-0002-9813-7712, **CVU CONACYT ID:** 20075

Abstract

The increase in the increase in wildfires throughout the world is largely due to increases in temperature and even to an increase in the carelessness of the population in leaving a large amount of the garbage in forests. Using Python and Matlab programs were as working medium. We performed the preprocessing on multispectral images obtained by the Landsat 8 satellite with and without wildfires, which consists of three steps: alignment, characterization and normalization, with the intention of standardization the images. From obtaining the spectral signatures of wildfires and metallic structures, boxes and whiskers diagrams, Shannon entropy and mutual information from the images, there are similar behavior in bands 6, 7, 8, 10 and 11, with more relevant information, taking into account that each image is formed by 11 bands, and in bands 1, 2, 3, 4, 5, 8 and 9 there is less information, SVD decomposition allows to have the best k-rank approximation to the original data matrix. The purpose of this analysis is to reduce the computational complexity.

Landsat 8, Wildfires, Characterization, Data analysis

Resumen

El incremento en el número de incendios forestales a lo largo del mundo se debe en gran medida a los aumentos de la temperatura e incluso al incremento de los descuidos de la población al dejar una gran cantidad de basura en bosques. Utilizando los programas de Python y Matlab como medio de trabajo, se realizó el pre procesamiento en imágenes multiespectrales obtenidas por el satélite Landsat 8 con y sin incendios, el cual consiste en tres pasos: alineación, caracterización y normalización, esto con la intención de estandarizar las imágenes. A partir de obtener las firmas espectrales de incendio y estructuras metálicas, los diagramas de cajas y bigotes, la entropía de Shannon e información mutua de las imágenes, se tienen similares comportamientos en las bandas 6,7,10 y 11, con mayor información relevante, tomando en cuenta que cada imagen está formada por 11 bandas, y en las bandas 1,2,3,4,5,8 y 9 se tiene una menor información, la descomposición SVD permite tener la mejor aproximación de rango-k de la matriz de datos originales. Al realizar este análisis se busca reducir la complejidad computacional.

Landsat 8, Incendio, Caracterización, Análisis de datos

* Correspondence to Author (e-mail: shernandezl_pa@upjr.edu.mx)
† Researcher contributing first author.

1. Introduction

The number of wildfires around the world are caused by increases in temperature and even by carelessness of the population leaving a large amount of garbage in the forest. For this reason, in recent years, the nations in the world have been focused on studies that seek on time wildfires detections, to arrive before the fires consume more hectares of forest. Several techniques have been used to accomplish wildfire studies, among them the use of Remote Sensing (RS) (Lira Ch-Vez, 2012).

Therefore, multispectral images from RS systems has shown to be a more complete system in studies of detection and classification of earth elements. One of the clearest examples is "Landsat 8 was launched on February 11, 2013"(USGS Science for a changing world, 2021)

In the study by (Corrales Andino, 2017) a model is developed for the correction of images with the values of Digital Numbers (DN) to the physical parameters as a satellite reflectance, which allows the reduction of errors caused by atmospheric effects. The methodology applied was the calculation of physical magnitudes for the conversion top-of-atmosphere (TOA) reflectance with angular correction. They mention that the model will be used for satellite images calibration of the Landsat sensor Operational Land Imager (OLI) program, to validate the model they perform comparative with a research work in the postgraduate. Master's Degree in Land Planning and Management of the Faculty of Space Sciences, under the ítem: Multitemporal Analysis of the Land Cover of the R Sub-basin

(Simonetti et al., 2015) Developed a fully automatic phenology-based synthesis (PBS) classification algorithm, which performs a mapping of terrestrial coverage. The seasonality of the vegetation in the dry regions is made the classification, with conventional classifiers, using images of the same dates in which they carry out the preprocessing, where they apply reflectance to, the part in which they mention that they did not process bands 1 and 9, thermic and panchromatic. A PBS has been developed for the Landsat 8 sensor and works fully automatically without additional input requirements, as supplementary images or training data.

PBS implementation in Google Earth Engine (GEE) provides a fast and efficient mapping tool applicable.

Remote Sensing

"Remote Sensing (RS) obtaining information about a scene, using visible light, infrared and microwave radiation, by means of automated analysis of data obtained from a distance using a system formed by a group of remote sensors" (Lira Ch-Vez, 2012)

RS provides a temporary overview in large areas: this allows us to observe changes over time that could range from minutes to years. This can be used to study changes in the stage of the earth's surface and atmosphere as mentioned by (Curlander, 1991)

Landsat 8

"Landsat 8 (formally the Landsat Data Continuity Mission, LCDM) was launched on February 11, 2013 from Vandenberg Air Force Base, California on an Atlas-V 401 rocket, with the extended payload fairing (EPF) from United Launch Alliance, LLC" (USGS Science for a changing world, 2021)

"The Landsat 8 satellite payload consists of two science instruments: the Operational Land Imager (OLI) and the Thermal InfraRed Sensor (TIRS). These two sensors provide seasonal coverage of the global landmass with a spatial resolution of 30 meters (visible, Near Infrared Red (NIR), Short-Wave-Infrared (SWIR)), 100 meters (thermal), and 15 meters (panchromatic)". The spectral bands of the OLI sensor are an improvement over previous Landsat instruments, with the addition of two additional spectral bands: a visible deep blue channel (band 1) specially designed for water resources and coastal zone research, and new shortwave infrared channel (band 9) (Landsat Science, 2021).

2. Contribution

In previous works, normalization is only performed in the preprocessing part, in the present work, characterization and alignment of the images are added to the preprocessing step, by combining these three aspects, a more appropriate image for the classification process is obtained.

In this study case, the classes of wildfires and metallic structure pixels is more concentrated.

3. Problem Statement

Given x where $x \in R^{I_1 \times I_2 \times I_3}$ is a third-order tensor, representing the multispectral image, which is obtained no aligned, unclassified. Using algorithms of alignment, atmospheric correction and normalization, you will get a new tensor x aligned, calibrated and normalized, in turn, to see the possibility of verifying a decrease of error in the classification in addition to having images standardization.

4. Pre-Processing

"Image Processing is the set of techniques applied to digital images in order to improve their quality", it encompasses a series of techniques that comprise operations where the origin is an image, and the result is another, already processed image. (Gonzalez, 2017)

The images obtained from the Landsat 8 system must have the necessary characteristic to be able to stand processed, that is to say they must have the following steps to be used, which are: alignment, characterization and normalization.

Image alignment

Because the sensors located on the satellite have a different capture distribution and the angle of incidence of the signals, it is sometimes necessary to record the images so that the 11 bands provided by a single multispectral image have the objects in the same positions by (James Storey, 2016). When performing the alignment between the bands, a test difference is performed, then the alignment between the bands is done and configured according to following the parameters.

Normalization

"An alternative approach to Z-score normalization (or standardization) is the so-called Min-Max scaling (often also simply called normalization a common cause for ambiguities). In this approach, the data is scaled to a fixed range, usually 0 to 1".

Is a process that changes the range of pixel intensity values, in contrast to standardization is that we will end up with smaller standard deviations, which can suppress the effect of outliers (Raschkan, 2014)

$$Xnorm = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Where X is an original value, $Xnorm$ is the normalized value, X_{min} is the minimum value of the variable and X_{max} is the maximum value of the variable.

Radiometric correction in Landsat 8 images

Landsat is a passive satellite that emits signals at certain frequencies and the captures the bounces of the signals, so those signals come with errors due to disturbances of the environment, including the sum of signals caused by the radiation from the sun, therefore the images must be adjusted according to the radiation characteristics of the sun, inclination and even the temperature, so the process is called calibration.

"The data from the OLI sensor can be converted to reflectance values, Top-Of-Atmosphere (TOA) in the ceiling of the atmosphere, using the reworked reflectance coefficients provided in the product metadata file (MTL file). The following equation is used to convert DN digital levels to reflectance values for those obtained by the OLI sensor" (Geospatial, 2021):

$$\rho\lambda' = M\rho Qcal + A\rho, \qquad (2)$$

where "$\rho\lambda'$ this is TOA planetary reflectance, without correction by solar angle. Note that $\rho\lambda'$ does not contain a correction for the solar angle, $M\rho$ is the Band-specific multiplicative rescaling factor from the metadata, it $Qcal$ quantized and calibrated standard product pixel values (DN)".

This value refers to each of the bands of the image and $A\rho$ is the specific scaling additive factor per band obtained from the metadata (Using the USGS Landsat Level-1 Data Product, s. f.).

5. Data analysis

Spectral Signature

The spectral signatures refer to the quantitative measurement of object properties in spectral bands, which refer to the variation of the reflectance or radiance emitted by objects, all geographic elements (forest, crops, rivers, lakes, buildings, etc.) each of the materials has a unique signature

Box and Whiskers

A box and whiskers diagram (or box plot) is a graphical representation of data, showing some of its key statistical features. The lines that extend parallel are whiskers, are used to indicate the variability outside the lower and upper quartiles (3d,3e), the first quartile has the lowest 25% of the data instances (3a), Median or also second quartile has 50% (3b) (the distribution of data is divided into two equal parts the distribution), and the third quartile has the lowest 75% of the data instances (3c) (Chalmers & Gilbey, 2018).

Quartiles

$$\text{First Quartile } (Q_1) = (N+1)/4 \qquad (3a)$$

$$\text{Second Quartile} = N/2 \qquad (3b)$$

$$\text{Third Quartile } (Q_3) = 3(N+1)/4 \qquad (3c)$$

Lower and upper admissible limits (*LI* y *LS*)

$$LI = Q_1 - 1.5 * IQR \qquad (3d)$$

$$LS = Q_3 + 1.5 * IQR, \qquad (3e)$$

where N = number of data instances and IQR = interquartile range. Using the box and whiskers diagrams, you get a lot of information from a probability distribution. This type of graph can identify outliers, at the same time making a distribution comparison.

Entropy

"The entropy which is a measure of the uncertainty of random variable. Let X be a discrete random variable with alphabet X and probability mass function $p(x) = Pr\{X = x\}, x \in X$.

We denote the probability mass function by $p(x)$ rather than $p_x(x)$, for convenience. Thus, $p(x)$ and $p(y)$ refer to two different random variables and are in fact different probability mass functions" (Cover, 2006).

The Shannon entropy $H(X)$ of a discrete random variable X is defined by

$$H(X) = -\sum_{x \in X} p(x) \, logp(x). \qquad (4)$$

We also write $H(p)$ for the above quantity. The log is to the base 2, then the entropy is expressed in bits. If the log is taken to be the natural log, then entropy is expressed in nats. Where $x \in X$ possible results belonging to X,

Mutual Information

"Consider two random variables X and Y with a joint probability mass function $p(x,y)$ and marginal probability mass function $p(x)$ and $p(y)$. The mutual information $I(X;Y)$ is the relative entropy between the joint distribution and the product distribution $p(x) \, p(y)$" (Cover, 2006).

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \, log \, \frac{p(x,y)}{p(x)p(y)} \qquad (5)$$

With the equation 5 can be written

$$I(X;Y) = H(X) - H(X|Y) \qquad (5a)$$

6. Singular Value Decomposition (SVD)

"The Singular value decomposition (SVD) is a theorem of linear algebra, where it mentions that a rectangular matrix A can be broken down into the product of the three matrices – an orthogonal matrix U, a diagonal matrix S, and the transpose of an orthogonal matrix V, where $m \times n$ size of the matrix, where the matrix S is the matrix of singular values, which ordered identify the direction of greatest variance, as well find the best approximation to the original data using fewer dimensions" (Decomposition, 2021, chapter-4).

$$A_{m \times n} = U_{m \times m} S_{m \times n} V_{n \times n}^T, \qquad (6)$$

where $U^T U = I, V^T V = I$; the columns of U are orthonormal eigenvectors of AA^T, the columns of V are orthonormal eigenvector of $A^T A$, and S is a diagonal matrix containing the square roots of eigenvalues from U or V in descending order (tutorial, 2021).

7. Methodology

In this work, the normalization, radiometric correction and data analysis algorithms were applied. Using two images with wildfires that occurred in Mendocino County (California USA) in 2018 with a difference of one week between the images, a third image of the same region, but without wildfires, was obtained by the Landsat 8 satellite.

The images were obtained from the website United States Geological Survey (USGS), for the pre-processing of multispectral images which have a resolution of 7750 x 7530. In this work on the images, a cut was made of the area where the fire is located. We will be working on 512 x 512 images of the wildfires zone.

Alignment

The images are not aligned; this is because the sensors located on the satellite have a different capture distribution and the angle of incidence, the signals have, it is sometimes necessary to record the images so that the 11 bands provided by a single multispectral image have the objects in the same positions.

Figure 1 shows an example of an image where the non-alignment with the images can be seen.

Figure 1 Non-alignment of the images (PARROT)

The non-alignment is found in the obtained images, as a first step in the preprocessing, the alignment of the images is applied. Figure 2 shows an example of the aligned image.

Figure 2 Alignment image (PARROT)

Atmospheric correction (characterization)

With the images aligned, Top- of- Atmosphere (TOA) surface reflectance will allow for improvement when comparing multiple images of the same region, considering the atmospheric effects of aerosol dispersion and thin clouds, the correction helps in detecting and characterization of surface changes. Which is the light reflected by the earth's surface.

It is a relationship between surface radiance, surface irradiance and as such has no units, it has values between 0 and 1. By which you will be applying the equation (2) of TOA reflectance described above.

Normalization

The previously aligned and characterized images will be normalized through equation 1, before applying the formula, the maximum value, and the minimum value of each of the bands are obtained, then we apply the equation 1.

Data analysis

The images are preprocessed and the analysis of the data, in this case the images, is then carried out. In which spectral signatures, box and whiskers diagram, Shannon entropy and mutual information of the bands will be obtained.

Spectral Signature: The first step is to obtain the spectral signatures of the classes: wildfires, vegetation, smoke, water and metallic structure, to observe the behavior and similarity between the wildfires and metallic structure.

Boxers and Whiskers diagram: The second step is to create the Boxer and Whiskers graphic according to the pixel values in each multispectral band with the wildfire's scene.

Entropy: The third step is to calculate the Shannon entropy for each image bands using the equation 4.

Mutual Information: The fourth step is to compute the mutual information of each pair bands, applying the equation 5.

SVD

After the characterization and analysis of an image, perform the singular value decomposition of this image selecting *k* singular values to construct a new image of rank *k*.

8. Results

Figure 3 shows the images of the Mendocino wildfires, obtained by the Landsat 8 satellite, which has a size of 512x512 pixels, and in figure, 5 images were obtained of the same area of Mendocino without wildfires.
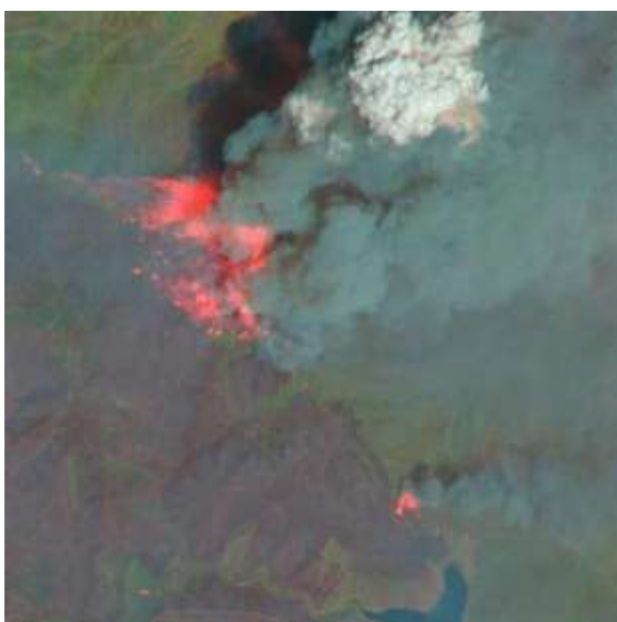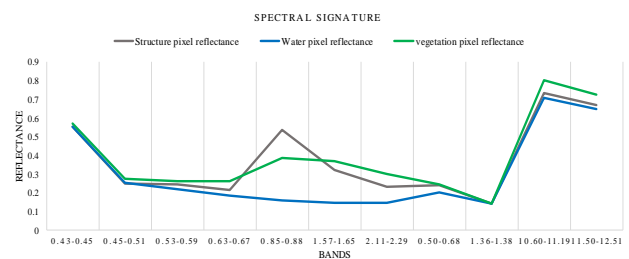

**Figure 3** Images with wildfires (Landsat 8) https://earthexplorer.usgs.gov/.

Figure 4 Images without wildfires (Landsat 8) https://earthexplorer.usgs.gov/

Graph 1 shows the spectral signatures of Figure 4 with three classes: vegetation, water and metallic structure, showing that the metallic structure has a higher reflectance in bands 6, 7 and 10.

**Graph 1** Spectral Signature images without wildfires
Own Source [Excel]

Graph 2 shows the spectral signatures of Figure 3 with five classes: wildfire, smoke, vegetation, water and metallic structure, which indicates that the wildfires show high reflectance values in band 6, 7, 10 and 11.
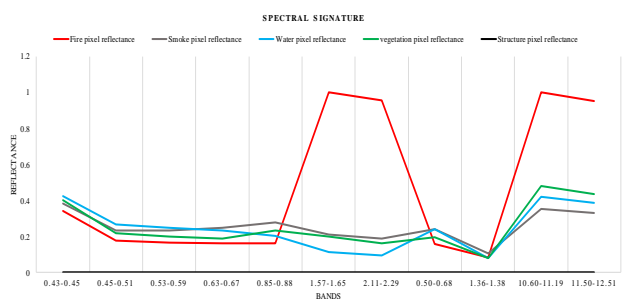
Graph 2 Spectral Signature images with wildfires
*Own Source [Excel]*

Graph 3 presents the box and whiskers diagram of the without wildfires image, showing a distribution of outliers across the 11 bands.
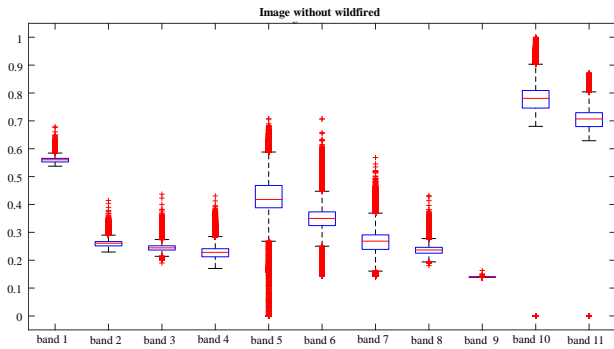
Graph 3 Box and whiskers image without wildfires
*Own Source [Matlab]*

Graph 4 presents the box and whiskers diagram of the images with wildfires, where more outliers are obtained throughout the 11 bands, with a higher concentration in bands 6,7,10 and 11.

Graph 4 presents the box and whiskers diagram of the images with wildfires, where more outliers are obtained throughout the 11 bands, with a higher concentration in bands 6,7,10 and 11.
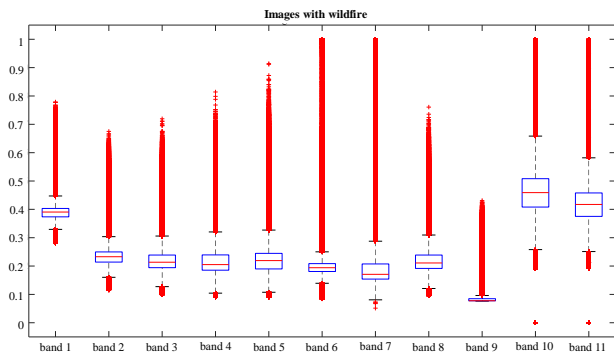

Graph 4 Box and whiskers image with wildfires
Own Source [Matlab]

Graph 5 illustrates the Shannon entropy of the images with and without wildfires where we observe the variation of the uncertainty from one image to the other, which gives more information in the image with wildfires.
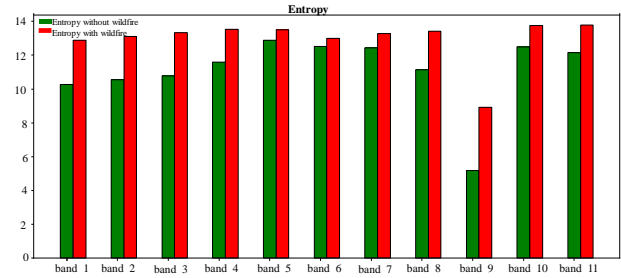

Graph 5 Shannon entropy
Own Source [Python]

Graph 6 illustrates the mutual information of the images with wildfires, where there is a greater similarity from band 1 to band 5, and band8.
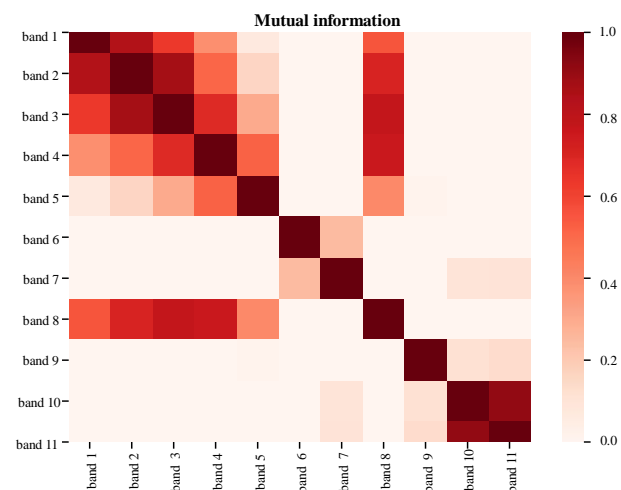
Graph 6 Mutual information image with wildfires
Own Source [Python]

Graph 7 shows the image without wildfires, where there is a greater similarity from band 1 to band 4, and bands 6, 7, 10 and 11.
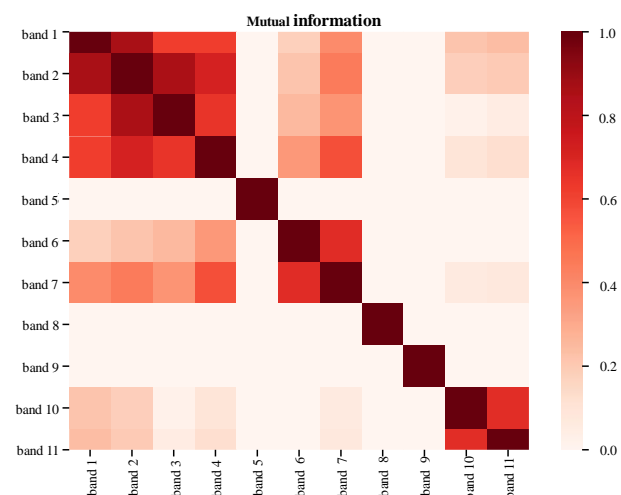
Graph 7 Mutual information without wildfires
Own Source[Python]

For each 512x512 image, its SVD was calculated. For a few singular values, the images do not have a good clarity, but the higher the number of singular values, the better image clarity, as shown in figures 5 and 6.
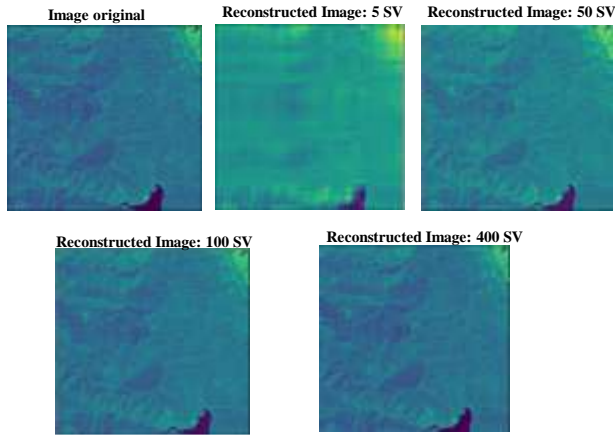

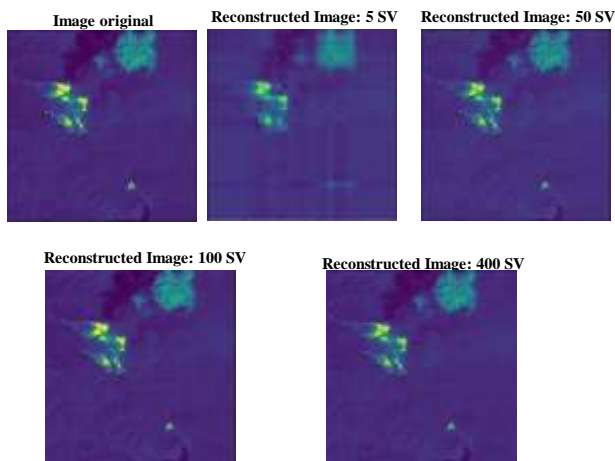**Figure 5** Singular values, image without wildfires

Figure 6 Singular values image with wildfires

Figure 7 it is observed that there is a difference in the singular values between the images with and without wildfires, and also from the image with wildfires more singular values can be used to work with a modified image in which not enough information will be lost.
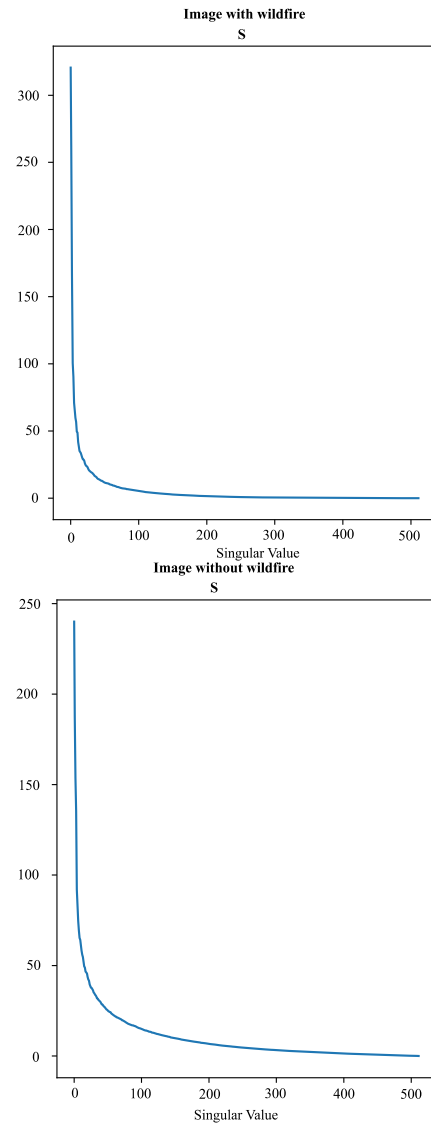

Figure 7 Singular Values
Own Source

9. Conclusions

This paper shows that when working with multispectral images obtained from the satellite Landsat 8, an alignment between the 11 bands must be performed. The characterization is also performed using the atmospheric reflectance correction to obtain an image in physical reflectance parameters from digital image values (DN), the normalize the images.

When obtaining the spectral signatures of the classes: wildfires, vegetation, smoke, water and metallic structure, it is observed that the reflectance of the wildfires and metallic structure pixels have the same behavior in bands 6,7,8,10 and 11, the spectral signatures shows in which bands there is a greater visualization of the wildfires and metallic structure pixels, this will allow us to reduce the classification error percentage of the classes in future works, when working with the classification algorithms.

HERNANDEZ-LOPEZ, Sandra Paola, YAÑEZ-VARGAS, Juan Israel, GONZÁLEZ-RAMIREZ, Andrea and TORRES-ROMAN, Deni. Statistical analysis and characterization of Landsat 8 satellite images of forest wildfires regions. ECORFAN Journal-Democratic Republic of Congo. 2021

Statistical analysis visualizes the behavior of the data in the image with and without wildfires, as shown in the box and whiskers diagram, Shannon entropy and mutual information of the images, helps us to know the amount of information provided by each of the bands, to discard the images where the information is minimal in the image with and without wildfires. From the image with wildfires in band 8 there is less information compared to band 6, 7, 8, 10 and 11 where there is more information. By using SVD a representation of the best approximation the original data.

In previous work, the preprocessing is perform by applying the atmospheric correction or normalization. What is done with the present work by adding in the preprocessing the steps of alignment, atmospheric correction and normalization to obtain a standardized image, at the same time it is sought to reduce the computational complexity.

10. Open issues

–   Apply an exhaustive study in image processing information theory.

–   To carry out a study of detection/classification algorithms and their subsequent implementation.

11. Acknowledgements

The authors would like to thank the CINVESTAV del IPN, Unidad Guadalajara, Telecommunications Laboratory, Mexico. And the Universidad Politécnica de Juventino Rosas

Reference

Corrales Andino, R. O. (2017). Calculation of reflectance in Landsat OLI-8 images, on the central region of Honduras, using fre software SEXTANTE. *Ciencias especiales*, 81-96. doi: https://doi.org/10.5377/ce.v9i1.3127

Chalmers,D., & Gilbey, J.(2018). *Cambridge International as & a Level Mathematics probability & Statistics 1 Coursebook with Cambridge Online Mathematics* (2 years) [Electronic book]. Cambridge University Press.

Cover, T.M. (2006). *[Elements of Information Theory] [By: Cover, Thomas M.] [July, 2006](SECOND EDITION)*[Electronic book]. Wiley-Interscience.

Curlander, J. &. (1991). *Synthetic Aperture Radar: Systems and Signal Processing*. Wiley-Interscience.

Decomposition, S. V. (09 de 05 de 2021). *Singular Value Decomposition*. Obtained from https://www.cs.cmu.edu/~venkatg/teaching/CStheory-infoage/book-chapter-4.pdf

Geospatial, L. H. (8 de April de 2021). *Radiometric Calibration*. Obtenido de L3 Harris Geospatial: https://www.l3harrisgeospatial.com/docs/radiometriccalibration.html

Gonzalez, R. &. (2017). *Digital Image Processing*. Pearson.

James Storey, D. M. (2016). A note on the temporary misregistration of Landsat-8 Operation Land Imager (OLI) and Sentinel-2 MultiSpectral Instrument (MSI) Imagery. *Remote Sensing of Enviroment*, 1221-122. doi:https://doi.org/10.1016/j.rse.2016.08.025

Landsat Sciencie. (7 de April de 2021). *Landsat Sciencie*. Obtenido de Landsat Sciencie: https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-overview

Lira Ch-Vez, J. (2012). *The Remote Perception (Spanish edition)*. Paperback.lulu.com.

Raschkan, D. (11 de July de 2014). *Dr.Sebastian Raschkan*. Obtenido de About Feature Scaling and Normalization: https://sebastianraschka.com/Articles/2014_about_feature_scaling.html#about-min-max-scaling

Raschkan, D. S. (11 de july de 2014). *About Feature Scaling and Normalization*. Obtenido de About Feature Scaling and Normalization.: https://sebastianraschka.com/Articles/2014_about_feature_scaling.html#about-min-max-scaling

Simonetti, D. &. (2015). First Result From the Phenology-Based Synthesis Classifier Using Landsat 8 Imagery. *IEEE Geoscience and Remote Sensing Letters*, 12(7),1496-1500. doi:https://doi.org/10.1109/lgrs.2015.2409982

Simmotti, D., Simonetti, e., Szatoi, Z., Lupi, A., & Eva, H.D. (2015) First Result From the Phenology-Base Sythesis Classifier Using Landsat 8 Imagery.*IEEE Geoscience and Remote Sensing Letters, 12(7),* 1496-1500. https://doi.org/10.1109/lgrs.2015.2409982.

Tschannerl, J. R. (2019). MIMR-DGSA: Unsupervised hyperspectral band selection based on information theory and modified discrete gravitational search algorithm. *Information Fusion,,* 189-200. doi:https://doi.org/10.1016/j.inffus.2019.02.005

tutorial, S. V. (09 de 05 de 2021). *Singular Value Decomposition (SVD) tutorial*. Obtenido de Singular Value Decomposition (SVD) tutorial:https://web.mit.edu/be.400/www/SVD/Singular_Value_Decomposition.htm

USGS Science for a changing world. (7 de april de 2021). *Landsat Missions*. Obtenido de Landsat Missions: Landsat Missions. (s. f.).

USGS Science for a changing world. Recuperado 7 de abril de 2021, de https://www.usgs.gov/core-science systems/nli/landsat/landsat-8?qt-science_support_page_related_con=0#qt-science_support_page_related_con

USGS science for a changing world. (19 de junio de 2021): https://earthexplorer.usgs.gov/