



Universidad del Azuay

Facultad de Ciencias de la Administración

Carrera de Ingeniería de Sistemas y Telemática

**APLICACIÓN DE RANDOM FOREST PARA
IDENTIFICAR ANOMALÍAS DE
PRECIPITACIÓN EN RELACIÓN A ÍNDICES
CLIMÁTICOS ENSO EN ECUADOR**

Trabajo de titulación previo a la obtención del grado en Ingeniera
en Sistemas y Telemática

Autora:

María Sol Contreras Villavicencio

Directores:

Daniela Elisabet Ballari

Cuenca – Ecuador

2021

DEDICATORIA

A mis padres, que no me han dejado sola.
Y a mis amigos, por lo mismo, pero con cerveza.
A mi familia por estar en todo este proceso, siempre
pendientes y apoyandome para lograr el éxito.



AGRADECIMIENTO

Agradezco a mis padres, por todo lo que han entregado para que esta meta se cumpla.

A mi directora, Dani Ballari, por su apoyo y comprensión para poder culminar con éxito este trabajo.

Agradezco a la Universidad del Azuay, por la beca brindada durante estos 5 años de estudio y por las enseñanzas que me ha dado.

A la Escuela de Ingeniería en Sistemas y Telemática, por la apertura y apoyo que han dado para todos los proyectos que me he planteado.

Y a mis amigos, que me levantaron en cada caída y estuvieron conmigo en cada proyecto y batalla que han traído estos 5 años de vida Universitaria.

A mi novio, por obligarme a creer en mi mismo.

Al Dr. Carlos Siguenza por permitir que llegue a este día.

ÍNDICE

DEDICATORIA.....	II
AGRADECIMIENTO.....	III
ÍNDICE.....	IV
RESUMEN	VII
ABSTRACT	VIII
1. INTRODUCCIÓN	1
1.1 Motivación	1
1.2 Estado del Arte.....	1
1.3 Problemática	2
1.4 Objetivos	2
1.4.1 General	2
1.4.2 Específicos.....	3
2. MÉTODO.....	3
2.1 Área de estudio	3
2.2 Materiales	3
2.2.1 Datos satelitales de precipitación con reducción de escala.....	3
2.2.2 Índices climáticos.....	4
2.3 Métodos.....	5
2.3.1 Cálculo de SPI de anomalías de precipitación (3 meses)	6
2.3.2 Random Forest para clasificación	7
3. RESULTADOS.....	8
3.1 Análisis de un píxel	8
Datos de entrenamiento y validación. Estratificación de la data	8
Modelos de Random Forest	9
Importancia de Variables.....	10
Validación y matriz de confusión.....	11
3.2 Análisis Global (todos los píxeles).....	12
3.2.1 Proceso.....	13
Función para obtener la clasificación de SPI	13

Función de paralelización	13
Función para analizar los pixeles	14
Lista resultante	15
3.3 Resultado para Random Forest	16
3.3.1 Gráfico de barras	16
Mapas	17
Mapas de anomalías negativas.....	17
4. CONCLUSIONES	19
4.1 Limitaciones.....	19
4.2 Beneficios	20
4.3 Trabajos futuros	20
BIBLIOGRAFÍA	21
Anexos	23
Código Fuente en formato Markdown de R	23
Preparacion de Datos.Rmd	23
Flood.Rmd	26
dry.Rmd	28
Global_floo.Rmd.....	31
Global_dry.rmd.....	33
read_results_dry.Rmd	35
read_results.Rmd	37

Índice de tablas y figuras

Tabla 1: Detalle de índices utilizados. Período 1981-2018. Extraído de (Avila, 2019)	5
Tabla 2. Rangos de nivel de precipitación. Tomado de: (World Meteorological Organization, 2012)	6

Figuras

Figura 1 Imágenes CHIRPS. La figura muestra la media mensual para el período de estudio.	4
Figura 2 Código de estratificación de la edad, por nivel	8
Figura 3 Código de estratificación de la edad, por nivel	9
Figura 4 Tabla superior de dato de Entrenamiento. Tabla inferior de la data de prueba. Data estratificada.....	9
Figura 5 Árbol de Clasificación. Gráfico para ejemplificar los posibles árboles, del bosque de decisión	10
Figura 6 Variables de importancia del píxel seleccionado	11

Figura 7 Matriz de confusión, resultante de aplicar el modelo generado sobre la data de prueba	12
Figura 8 Función de Clasificación de los registros de las precipitaciones	13
Figura 9 Función de paralelización.....	14
Figura 10 Función de analizar pixel.....	15
Figura 11 Código de aplicación de Random Forest.....	15
Figura 12. Lista de DataFrames	16
Figura 13. Detalles de un DataFrame de la lista anterior, donde se observan los tres índices de mayor importancia.	16
Figura 14. Frecuencia de importancia de índices climáticos de las anomalías negativas de precipitación.....	17
Figura 15. Frecuencia de importancia de índices climáticos de las anomalías positivas de precipitación.....	17
Figura 16 . Ejemplos de mapas de influencia espacial para anomalías negativas.	18
Figura 17 Ejemplos de mapas de influencia espacial para anomalías positivas.....	19

RESUMEN

Analizar las precipitaciones en un territorio es de vital importancia para tomar decisiones en los sectores productivos y energéticos. Comprender las relaciones de las anomalías de precipitación; es decir, exceso o ausencia de la misma, con índices climáticos permite tanto predecir estas situaciones climáticas, como tomar decisiones para el control de las mismas basadas en evidencia. El objetivo de este trabajo es identificar la relación de estas anomalías sobre el territorio ecuatoriano con diferentes índices climáticos. Para ello se utilizó Random Forest como método de clasificación. Los datos fueron obtenidos de imágenes satelitales CHIRPS del Ecuador (1981 al 2018).

Palabras clave: anomalías, precipitaciones, Ecuador, ENSO, índices climáticos, machine learning.

ABSTRACT

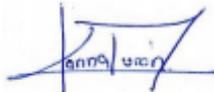
Analyzing precipitation within a region is of considerable value for decision-making in sectors involved with productivity and energy. Understanding the relationships between the factors that cause precipitation anomalies—that is, either an excess or a lack of precipitation—by using climate indices is important both for predicting those climate conditions, as well as for decision-making that controls for these factors based on collected evidence. The goal of this project was to identify the relationships between the factors for these anomalies within the country of Ecuador through different climate indices. To achieve this, the Random Forest method of classification was used. Data was collected from Ecuador’s CHIRPS satellite images (from 1981 to 2018).

Keywords: anomalies, precipitation, Ecuador, ENSO, climate indices, machine learning.

Ma Sol Contreras U.

Translated by:

Ma. Sol Contreras Villavicencio



1. INTRODUCCIÓN

1.1 Motivación

El Ecuador es un país que basa su economía sobre altos porcentajes de producción agropecuaria (INEC, n.d.). Esta producción depende directamente de la situación climática del país. Al presentarse anomalías de precipitación la producción agropecuaria resulta fuertemente afectada. Esto se traduce en afecciones directas a la situación económica del país. De igual manera, otro sector afectado directamente por las anomalías de precipitación es el sector energético. El Ecuador depende en gran medida de la producción energética hidroeléctrica, que representa el porcentaje más alto sobre el resto de energías renovables y no renovables entregadas al servicio público del país (Electricidad, n.d.). Así, el análisis de las precipitaciones en el territorio es de vital importancia para tomar decisiones dentro de los sectores productivo y energético existentes.

1.2 Estado del Arte

Uno de los fenómenos climáticos que más afecta al país, es El Niño – Oscilación del Sur (ENSO). ENSO es considerado como la mayor variabilidad climática, por la cantidad de territorio al que afecta alrededor del globo, produciendo cambios en precipitaciones y temperatura en las diferentes zonas (Del et al., 2007). Los estudios sobre este fenómeno son muy numerosos. Podemos encontrar análisis del fenómeno en zonas muy distantes al Ecuador, como lo es Sri Lanka, donde se determinó la influencia del fenómeno sobre las precipitaciones, utilizando modelos de clasificación (De Silva M. & Hornberger, 2018) o Asia (Wu et al., 2003), o en lugares muy cercanos, como son Perú o Chile. De estos estudios se puede resaltar que no sólo se ha trabajado sobre precipitaciones, sino también se ha analizado la influencia que tiene ENSO sobre la precipitación (Escobar & Aceituno, 1998) o la situación de los glaciares presentes (Francou & Ribstein, 1995). Existen también estudios enfocados específicamente en la predicción de las precipitaciones en la zona de estudio, utilizando métodos como el modelado de redes neuronales (Jiménez-Carrión et al., 2018).

Existe también un grupo de estudios que está enfocado específicamente en el estudio de la situación del Ecuador frente a este fenómeno. Algunos de ellos lo hacen de forma general, trabajando con datos de todo el territorio y utilizando, por ejemplo, métodos de clustering para determinar los diferentes patrones de precipitación (Tobar & Wyseure, 2018). Otros, se enfocan en zonas más pequeñas y específicas como es el área costera del país para determinar de manera temprana posibles sequías (Zambrano Mera et al., 2018). Existen, además, investigaciones que proporcionan nuevas formas de obtención de datos e información, como lo es la evaluación de imágenes satelitales del país para determinar las precipitaciones ocurridas en el Ecuador continental (Castro Loayza, 2014). Aunque anteriormente se ha utilizado el método de Random Forest sobre datos e información de precipitaciones (De Silva M. & Hornberger, 2018) el uso de este método sobre información obtenida de imágenes satelitales del Ecuador no se ha registrado anteriormente.

1.3 Problemática

Conocer las anomalías que pueden presentarse en determinadas épocas o bajo determinadas condiciones, permite que los sectores productivos y energéticos sean capaces de construir planes de mitigación. Sin embargo, en el Ecuador, realizar estos análisis supone un trabajo de gran complejidad. Su geografía produce variaciones climáticas muy notorias entre las diferentes regiones del país, y dificulta que el monitoreo sea denso en el territorio, por lo que en general se cuenta con escasas estaciones pluviométricas. Además, en el estado del arte se detallan trabajos realizados en esta línea; sin embargo, para el Ecuador, no se han localizado estudios previos que profundicen sobre la identificación de anomalías de precipitación en relación a índices climáticos como lo es ENSO.

1.4 Objetivos

1.4.1 General

Explorar el uso de Random Forest para clasificación sobre información obtenida de imágenes satelitales mensuales de precipitación del Ecuador durante el período 1981

– 2018, para determinar las anomalías presentadas y su relación con los índices climáticos ENSO.

1.4.2 Específicos

- Calcular el índice estandarizado de precipitación (SPI) para imágenes satelitales de precipitación.
- Relacionar temporalmente un conjunto de píxeles de entrenamiento con datos SPI que indiquen anomalías con los diferentes índices climáticos del fenómeno ENSO.
- Implementar Random Forest para clasificación.
- Identificar los índices climáticos que tengan mayor relevancia para la clasificación de anomalías de precipitación.

2. MÉTODO

2.1 Área de estudio

Ecuador se ubica al noreste de América del Sur, compartiendo frontera al norte con Colombia, al sur y este con Perú y al oeste con el Océano Pacífico. Su extensión territorial es aproximadamente 252,000 km². De norte a sur se encuentra atravesado por la cordillera de los Andes. Estas características, además de su ubicación, sobre la línea ecuatorial, limita la estacionalidad que se presenta en el año. Así, se reconocen solamente dos estaciones, invierno (humedad) y verano (sequía). La topografía del país influye en el nivel de precipitación pluvial. Así, en zonas de mayor elevación y baja temperatura, el nivel de precipitación es menor. Otro factor importante que influye sobre la zona de estudio son las corrientes marítimas (Varela & Ron, 2018).

2.2 Materiales

2.2.1 Datos satelitales de precipitación con reducción de escala.

Este trabajo utilizó las imágenes satelitales de precipitación CHIRPS a 5.4km de resolución espacial (<https://www.chc.ucsb.edu/data/chirps>) (Funk et al., 2015). Las imágenes utilizadas se extrajeron para la región geográfica del Ecuador continental y a escala mensual para el período (1981 - 2018). La tendencia lineal fue eliminada de las

series temporales de las imágenes, lo cual se aplicó pixel a pixel. La figura 1 resume la media mensual de las imágenes CHIRPS para el período de estudio.

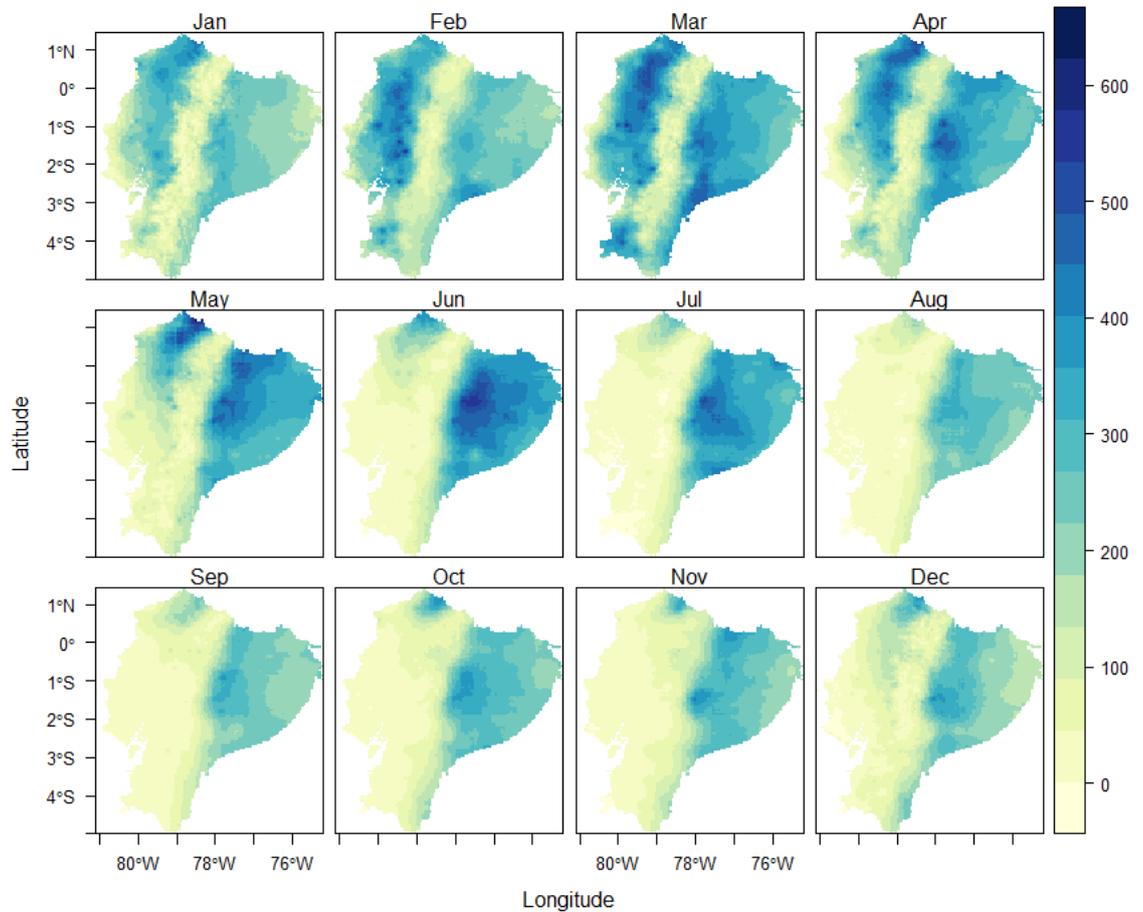


Figura 1 Imágenes CHIRPS. La figura muestra la media mensual para el período de estudio.

2.2.2 Índices climáticos

Los índices climáticos permiten observar la variabilidad climática y tendencias en diferentes regiones del planeta. Los índices para este trabajo fueron obtenidos de: “Climate Indices: Monthly Atmospheric and Ocean Time-Series” (Physical Sciences Division, n.d.), que a su vez fueron seleccionados a partir de los trabajos (Liu et al., 2018, Fierro, 2014; Córdoba-Machado et al., 2016). La tendencia lineal fue eliminada de las series temporales de los índices.

Los índices climáticos a utilizar en este trabajo son:

Tabla 1: Detalle de índices utilizados. Período 1981-2018. Extraído de (Avila, 2019)

Índice	Nombre	Acceso
Relacionados con ENSO		
Niño 1+2	El Niño región 1+2	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Niño 3	El Niño región 3	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Niño 4	El Niño región 4	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Niño 3.4	El Niño región 3,4	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Relacionados con el Océano Pacífico		
TNI	Trans-Niño	https://www.esrl.noaa.gov/psd/data/climateindices/list/
ONI	Niño oceánico	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Relacionados con Océano Atlántico		
TNA	Tropical Northern Atlantic	https://www.esrl.noaa.gov/psd/data/climateindices/list/
TSA	Tropical Southern Atlantic	https://www.esrl.noaa.gov/psd/data/climateindices/list/
CAR_ersst	Caribbean SST Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/
NTA_ersst	North Tropical Atlantic Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Amon	Atlantic multidecadal Oscillation	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Ammsst	Atlantic Meridional Mode	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Relacionados con índices atmosféricos		
AO	Artic Oscillation	https://www.esrl.noaa.gov/psd/data/climateindices/list/
AAO	Antarctic Oscillation	https://www.esrl.noaa.gov/psd/data/climateindices/list/
QBO	Quasi-Biennial Oscillation	https://climatedataguide.ucar.edu/climate-data/qbo-quasi-biennial-oscillation
SOI	Southern Oscillation Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Relacionados con teleconexiones		
PDO	Pacific Decadal Oscillation	https://www.esrl.noaa.gov/psd/data/climateindices/list/
WP	Western Pacific	https://www.esrl.noaa.gov/psd/data/climateindices/list/
PNA	Pacific North American	https://www.esrl.noaa.gov/psd/data/climateindices/list/
NAO	North Atlantic Oscillation	https://www.esrl.noaa.gov/psd/data/climateindices/list/
Otros		
Solar	Radiacion solar	https://www.esrl.noaa.gov/psd/data/climateindices/list/

2.3 Métodos

El procedimiento general se resume de la siguiente manera: Para el desarrollo de esta investigación se divide al territorio de estudio en una cuadrícula que facilite su análisis, siguiendo para esto la división en píxeles de las imágenes satelitales de precipitación. Cada píxel representa y almacena la serie temporal mensual de precipitación, en dicha localización, desde 1981 al 2018. A continuación, se calcula el índice estándar de precipitación (SPI), el período utilizado para el cálculo es de tres meses (SPI.3).

Una vez calculado, se procede a preparar tablas de entrenamiento, para cada pixel, que relaciona cada registro mensual de precipitación y los diferentes índices climáticos, se trabajará con el 70% del total de los datos. Estas tablas son generadas por cada pixel de la zona de estudio, en cuyas columnas contienen los diferentes índices, y en las filas la serie temporal mensual SPI.3. Se continua con la ejecución de Random Forest para clasificación por cada tabla generada, es decir para cada píxel. Con base en los resultados, se puede conocer el impacto de cada índice climático o la relación del mismo sobre el nivel de anomalía de la precipitación. Para validar los resultados, se procede a ejecutar el mismo proceso sobre el 30% restantes del total de datos.

2.3.1 Cálculo de SPI de anomalías de precipitación (3 meses)

El índice estándar de precipitación (SPI, por sus siglas en inglés), es un índice basado en la probabilidad de precipitación en cualquier escala de tiempo. SPI permite clasificar en diferentes rangos el nivel de precipitación (World Meteorological Organization, 2012). Según la guía de uso de este índice (World Meteorological Organization, 2012), se tiene la siguiente tabla de clasificación:

Tabla 2. Rangos de nivel de precipitación. Tomado de: (World Meteorological Organization, 2012)

RANGO	DESCRIPCIÓN
2.0+	Extremadamente húmedo
1,5 a 1,99	Muy húmedo
1,0 a 1,49	Moderadamente húmedo
-0,99 a 0,99	Normal
-1,0 a -1,49	Moderadamente seco
-1,5 a -1,99	Severamente seco
-2 o menos	Extremadamente Seco

Donde se puede observar que, si el valor es positivo y mayor a 1, el nivel de precipitaciones en la zona estudiada es mayor a la media esperada. Por el contrario, si es negativo y menor a 1, entonces el nivel de precipitación es menor a la media esperada.

El SPI se calcula con la librería de R <https://cran.r-project.org/web/packages/SPEI/SPEI.pdf>, tomando como referencia los tres meses anteriores al mes en cuestión, es decir se calculará el SPI.3. Además, los valores de SPI.3 mayores a 1, serán considerados como anomalías positivas moderadas (moderately wet, moderadamente húmedo), severas (very wet, muy húmedo), o extremas (extremely wet, extremadamente húmedo); mientras que los valores menores a -1, serán considerados como anomalías negativas moderadas (moderately dry, moderadamente seco), severas (very dry, muy seco), o extremas (extremely dry, extremada seco).

2.3.2 Random Forest para clasificación

Random Forest es una técnica que utiliza una colección de árboles de clasificación, estructurados con base en sub-muestras seleccionadas de forma aleatoria. Esta técnica fue seleccionada para esta investigación porque permite trabajar con grandes números de datos y variables (Medina-Merino & Ñique-Chacón, 2017). Esta característica es de vital importancia por los materiales con los que se trabajará. La fortaleza de Random Forest se relaciona con la fuerza de cada uno de los árboles que conforman el “bosque”, y la correlación que existe entre ellos. Esta técnica nos permite determinar la importancia que tienen las variables en una relación (Breiman, 2008).

En este trabajo, se utiliza Random Forest con la librería de R <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>. El objetivo es clasificar el nivel de anomalía de precipitación, tomando como referencia los valores de los índices climáticos y cuantificar la importancia de cada índice para la clasificación (o identificación del tipo y severidad de la anomalía).

3. RESULTADOS

3.1 Análisis de un píxel

Se cargan los datos de los pixeles corresponde al territorio ecuatoriano, cada píxel tiene 454 observaciones/registros mensuales. El archivo cargado contiene las observaciones de SPI diferente de cero, así se analizan únicamente las observaciones o meses que tienen anomalías positivas o negativas de precipitación. La clasificación de Random Forest se realiza con el objetivo de determinar las variables (índices climáticos) de mayor importancia en cada anomalía. En la siguiente figura [2](#) se puede observar un conjunto de observaciones pertenecientes al píxel seleccionado para trabajar a modo de ejemplo. Podemos ver que la mayoría de los registros tiene un factor SPI igual a cero y alguno de registros con factor negativo, que equivalen a anomalías de ausencia de precipitación o sequía. Las diferentes columnas muestran los valores de los índices para los registros SPI.

	spi <fctr>	aao <dbl>	ao <dbl>	qbo <dbl>	soi <dbl>	nino1 <dbl>	nino3 <dbl>
43	0	1.079121362	0.018544815	-26.327640190	0.702315023	21.25221	25.01800
44	0	-0.359898160	0.465547269	-28.832585526	1.100146209	20.37263	24.69806
45	0	0.514082318	-0.413450278	-26.367530862	0.597977395	20.58305	24.54813
46	0	0.177062796	-0.270447824	-22.512476198	-0.104191419	20.43347	24.29819
47	0	0.161043274	-0.966445370	-14.107421534	0.793639767	21.44389	24.09825
48	0	-1.291976249	0.445557084	-9.072366869	0.191470952	22.33431	23.94832
49	0	-0.611995771	-2.806440462	-1.277312205	-0.110697862	23.71473	24.38838
50	-2	0.396984707	-1.440438008	3.307742459	2.387133324	24.64515	25.48845
51	-3	0.046965185	0.550564445	5.402797123	1.684964510	25.52557	26.40851
52	-1	0.210945663	0.651566899	10.217851787	2.282795696	24.08598	26.51858

Figura 2 Código de estratificación de la edad, por nivel

Datos de entrenamiento y validación. Estratificación de la data

Para extraer el 70% de los datos para entrenamiento se clasifica o estratifica aleatoriamente los datos, para así tener representatividad proporcional en cada una de las categorías o clases de SPI. La figura 3 muestra el código de estratificación para extraer el 70% de los datos de entrenamiento y el restante 30% de validación. Mientras tanto, la figura 4 muestra la frecuencia para las categorías de sequía (-3, -2, -1, 0), en la primera fila para entrenamiento y en la segunda para validación. Como puede observarse en dicha

figura, los datos más numerosos son de cero indicando precipitaciones cercanas a la media esperada.

```
data_train<-fifer::stratified(data, c("spi"), 0.7) #Estratifica la data
basada en la variable SPI. Tomando el 30% de cada nivel registrado para
Test y el 70% para entrenamiento.
data_test<- data[-as.numeric(rownames(data_train)),]
```

Figura 3 Código de estratificación de la edad, por nivel

-3	-2	-1	0
8	13	27	270
-3	-2	-1	0
5	9	8	116

Figura 4 Tabla superior de dato de Entrenamiento. Tabla inferior de la data de prueba. Data estratificada

Modelos de Random Forest

Para ejemplificar de mejor manera la construcción del modelo se reconstruye uno de los árboles que pueden formar parte del bosque de clasificación. En la siguiente figura (Figura 5) podemos que la clasificación se realizó a partir del el índice climático niño34. Esta clasificación nos muestra que el 4% de la data presenta sequia extrema y el 96% restante no presenta ninguna anomalía. Claramente este bosque contiene más árboles similares a éste pero que permite realizar una clasificación usando todos los niveles del factor SPI.

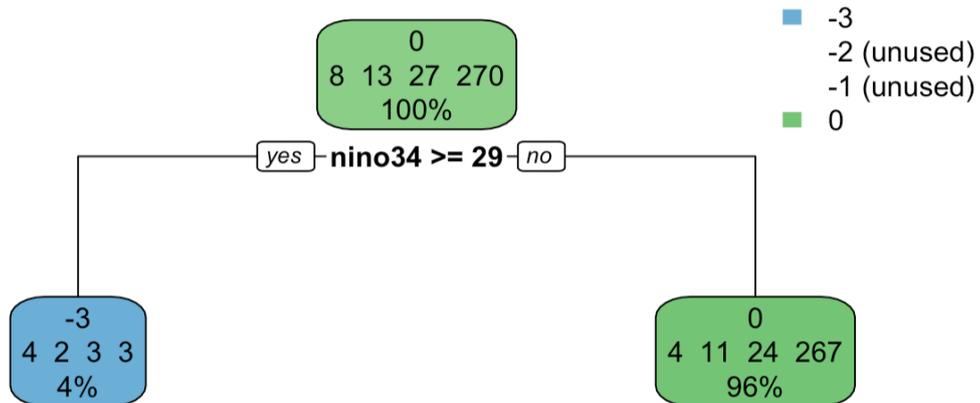


Figura 5 Árbol de Clasificación. Gráfico para ejemplificar los posibles árboles, del bosque de decisión

Importancia de Variables

Un resultado obtenido de aplicar todo el proceso de análisis sobre 1 pixel, nos permite determinar los índices climáticos con mayor influencia sobre las anomalías de precipitación, es decir que tienen mayor importancia para la clasificación de la severidad. En este pixel de ejemplo, los índices que aparecen en la parte superior son los más importantes para la clasificación, estos son niño34, niño4, soi, CAR_ersst y ONI.

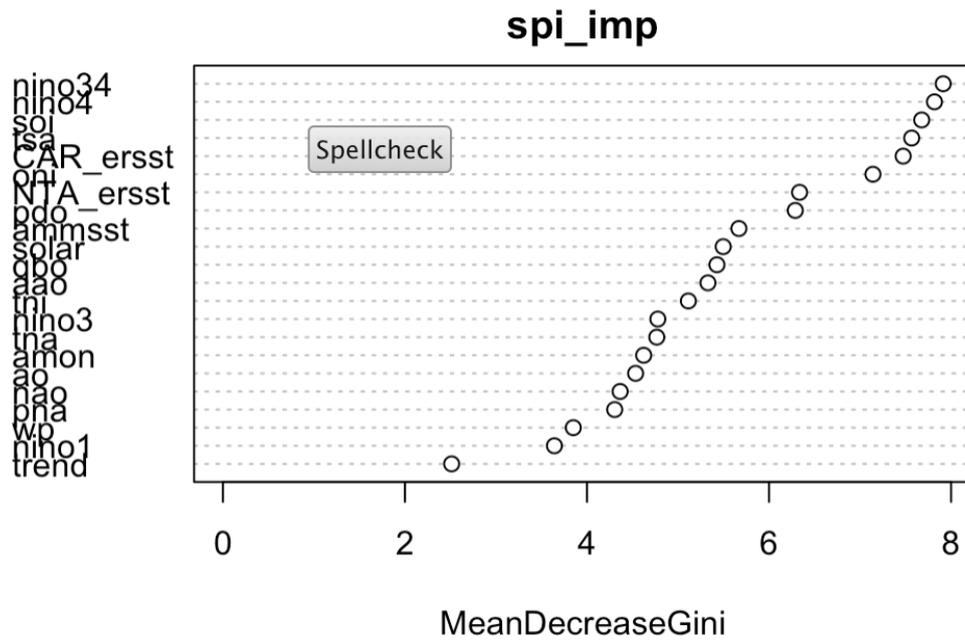


Figura 6 Variables de importancia del píxel seleccionado

Validación y matriz de confusión

Otro resultado obtenido al realizar la validación, es la matriz de confusión. Esta se obtiene de aplicar con los datos de validación (30%) el modelo de Random Forest aprendido con la data de entrenamiento. De estos resultados se observa los valores de precisión (cuanto más cercano a 1 mejor), los valores de Kappa (cuanto más cercano a 1 mejor, aunque más conservador que la precisión), y el nivel de significancia expresado por el valor-P (menor a 0.001, 0.05 o 0.1). La figura 7 muestra un resultado de ejemplo para un píxel con un nivel de significancia menor a 0.1.

Confusion Matrix and Statistics

	-3	-2	-1	0
-3	3	1	0	1
-2	0	5	0	4
-1	0	0	7	4
0	0	0	0	112

Overall Statistics

Accuracy : 0.927
95% CI : (0.8699, 0.9644)
No Information Rate : 0.8832
P-Value [Acc > NIR] : 0.06527

Kappa : 0.7298

Mcnemar's Test P-Value : NA

Statistics by Class:

	Class: -3	Class: -2	Class: -1	Class: 0
Sensitivity	1.0000	0.83333	1.00000	0.9256
Specificity	0.9851	0.96947	0.96923	1.0000
Pos Pred Value	0.6000	0.55556	0.63636	1.0000
Neg Pred Value	1.0000	0.99219	1.00000	0.6400
Prevalence	0.0219	0.04380	0.05109	0.8832
Detection Rate	0.0219	0.03650	0.05109	0.8175
Detection Prevalence	0.0365	0.06569	0.08029	0.8175
Balanced Accuracy	0.9925	0.90140	0.98462	0.9628

Figura 7 Matriz de confusión, resultante de aplicar el modelo generado sobre la data de prueba

3.2 Análisis Global (todos los pixeles)

¿Qué es el análisis global?

Realizar este análisis, partimos desde la preparación de datos, donde se realiza una clasificación de SPI en anomalías positivas y negativas de precipitación. Así se generan dos archivos de entrenamiento que corresponden a excesos de humedad (positivo) y a excesos de sequías (negativo). Éstos dos archivos resultantes, son la entrada principal del análisis de esta sección. Para realizar la descripción del proceso se utilizará como ejemplo la clasificación de ausencia de precipitación o sequías, ya que el proceso para ambas clasificaciones es exactamente igual.

3.2.1 Proceso

Función para obtener la clasificación de SPI

Para poder iniciar el análisis global es necesario ejecutar el script de R de preparación de datos para obtener la entrada necesaria.

A continuación, se detalla el código, correspondiente a la clasificación en anomalías de precipitación y anomalías de ausencia de precipitación:

```
prob.drought_type.2<- function(spdf.final.sp){#spdf.final.sp raster bricl

pdf.final.sp.s_moderate<- spdf.final.sp
pdf.final.sp.s_severe<- spdf.final.sp
pdf.final.sp.s_extreme<- spdf.final.sp

pdf.final.sp.l_moderate<- spdf.final.sp
pdf.final.sp.l_severe<- spdf.final.sp
pdf.final.sp.l_extreme<- spdf.final.sp
class<-spdf.final.sp

for (i in 1:ncol(spdf.final.sp)){#i=1
#ausencia de precipitación
pdf.final.sp.s_moderate@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) & spdf.final.sp@data[,i]<= -1 & spdf.final.sp@data[,i]> -1.5, -1, NA)
pdf.final.sp.s_severe@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) & spdf.final.sp@data[,i]<= -1.5 & spdf.final.sp@data[,i]> -2, -2, NA)
pdf.final.sp.s_extreme@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) & spdf.final.sp@data[,i]<= -2, -3, NA)

#Anomalías de precipitación: lluvia intensa
#pdf.final.sp.l_moderate@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) & spdf.final.sp@data[,i]>= 1 & spdf.final.sp@data[,i]< 1.5, 1, NA)
#pdf.final.sp.l_severe@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) & spdf.final.sp@data[,i]>= 1.5 & spdf.final.sp@data[,i]< 2, 2, NA)
#pdf.final.sp.l_extreme@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) & spdf.final.sp@data[,i]>= 2, 3, NA)

union<-data.frame(l_m=pdf.final.sp.l_moderate@data[,i],l_s=pdf.final.sp.l_severe@data[,i],l_e=pdf.final.sp.l_extreme@data[,i])

#union<-data.frame(s_m=pdf.final.sp.s_moderate@data[,i],s_s=pdf.final.sp.s_severe@data[,i],s_e=pdf.final.sp.s_extreme@data[,i],l_m=pdf.final.sp.l_moderate@data[,i],l_s=pdf.final.sp.l_severe@data[,i],l_e=pdf.final.sp.l_extreme@data[,i])
union$merge<- NA
for(j in 1:nrow(union)){#j=1
a<- which(!is.na(union[,j]))
if (length(a)==0) next
else union$merge[j]<-union[j,a]}

class@data[,i]<-union$merge
}

class<- class[,~c(1,2)]
return(class) }
```

Figura 8 Función de Clasificación de los registros de las precipitaciones

En la función detallada anteriormente está configurada para obtener un resultado clasificación de ausencia de precipitación en valores -3, -2 y -1. Para el caso de las anomalías positivas de precipitación se deberá activar las líneas comentadas que corresponde a anomalías de precipitación de lluvia intensa y comentar las líneas previas de ausencia de precipitación.

Función de paralelización

A continuación, se detalla el código para aplicar la función análisis píxel, pero esta vez para todos los píxeles de la zona de estudio. En este proceso fue necesario determinar el

número de núcleos con la que la computadora con la que se trabajaba tenía disponibles para la ejecución del proceso, al ser una gran cantidad de datos analizar, era necesario realizar un procesamiento en paralelo. A continuación, se puede observar cómo se determina el número de núcleos disponibles y como con una función ejecución en paralelo se utilizan los núcleos disponibles para analizar toda la data necesaria. La función de paralelización en R se implementó con parLapply.

```
````{r}
no_cores <- detectCores()
clust <- makeCluster(no_cores)
#inputData<-RF.data[1:55]
#system.time(dataFrameResult<-parLapply(clust, inputData, analyzePixel))
time<-system.time(dataFrameResult<-parLapply(clust, RF.data, analyzePixel))
saveRDS(dataFrameResult, file = "dataFrameResult.rds")
my_data <- readRDS("dataFrameResult.rds")
write.csv(my_data, file="Prueba1.csv")
class(my_data)

plot(my_data,my_data)
stopCluster(clust)
````
```

Figura 9 Función de paralelización.

Función para analizar los pixeles

La función análisis píxel realiza el siguiente proceso: como entrada recibos las observaciones del píxel enviado desde la invocación de la función. A continuación, se verifica que dicho píxel registre al menos una observación con anomalías de precipitación, esto equivale a que al menos uno de los registros tenga un valor SPI diferente de cero. En caso de que 1 píxel tenga todas sus observaciones con el valor SPI igual a cero se almacena dicho píxel en un DataFrame vacío para que no altere el posterior análisis del píxel.

```

analyzePixel <- function(inputData){
  library(hash)
  entradaSPI<-inputData["spi"]
  if (length(entradaSPI[entradaSPI != 0]) == 0) {
    dataFrameResult <- data.frame(
      "nombres" = c('', '', ''),
      "valores" = c(0,0,0))
    return(dataFrameResult)
  }
}

```

Figura 10 Función de analizar pixel.

La segunda parte de la función se aplica sobre los pixeles que tienen al menos una observación registrada como anomalía de precipitación, sobre estos pixeles se aplica el siguiente proceso.

Se estratifican los datos, obteniendo el 70% de cada nivel de anomalía para data de entrenamiento, y el 30% para la data de validación. A continuación, utilizando la data de entrenamiento se construye el modelo de Random Forest. Con el modelo se obtienen las variables de mayor importancia que afectan al píxel analizado, se almacenan las tres variables más relevantes del píxel y se retorna un DataFrame con las tres variables de importancia por cada píxel que se almacena en una lista.

```

data_train<-fifer::stratified(inputData, c("spi"), 0.7) #Estratifica la data basada en la variable
ara Test y el 70% para entrenamiento.
data_test<- inputData[-as.numeric(rownames(data_train)),]
spi =randomForest(factor(spi) ~ ., data = data_train)# Modelo de Random Forest.
simp =randomForest(factor(spi) ~ ., data = inputData)#Variables de importancia obtenidas gracias al
importancia<-varImpPlot(simp, TRUE)
names<- rownames(importancia)
values<-importancia[,1]
dataFrameOrder <- data.frame(
  "nombres" = names,
  "valores" = values)
dataFrameOrderResult<-dataFrameOrder[order(-dataFrameOrder$valores),]
names<-dataFrameOrderResult$nombres[0:3]
values<-dataFrameOrderResult$valores[0:3]
dataFrameResult <- data.frame(
  "nombres" = names,
  "valores" = values)

```

Figura 11 Código de aplicación de Random Forest.

Lista resultante

Del código anterior se obtiene el siguiente resultado: una lista de dataframes o tablas. Donde cada elemento es un DataFrame con las tres variables de importancia por píxel.



Figura 12. Lista de DataFrames

| nombres
<chr> | valores
<dbl> |
|------------------|------------------|
| tni | 15.561909 |
| CAR_ersst | 7.014657 |
| qbo | 6.658219 |

3 rows

3 rows

Figura 13. Detalles de un DataFrame de la lista anterior, donde se observan los tres índices de mayor importancia.

3.3 Resultado para Random Forest

Una vez ejecutado el código para el análisis global, se obtiene como resultado la tabla descrita anteriormente (Figura 13), y se procede a graficarlos. Para graficar los resultados se utilizan gráficos de barras de frecuencia y mapas que permiten analizar de mejor manera los resultados obtenidos.

3.3.1 Gráfico de barras

A continuación, haremos una comparación entre el gráfico obtenido de la tabla de anomalías de precipitación negativas y contra el gráfico obtenido de la tabla de anomalías de precipitación positiva. Estos gráficos muestran la frecuencia (en cuantos píxeles) estos índices aparecieron entre los tres más importantes.

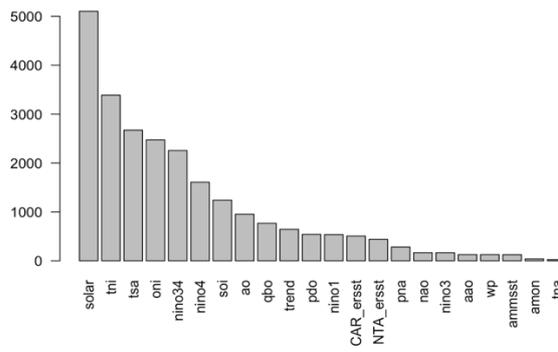


Figura 14. Frecuencia de importancia de índices climáticos de las anomalías negativas de precipitación

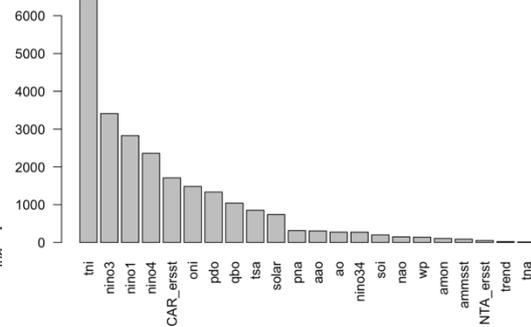


Figura 15. Frecuencia de importancia de índices climáticos de las anomalías positivas de precipitación

Como se puede observar en la figura 14, en las anomalías negativas que corresponden a la ausencia de precipitación, el índice climático que más frecuentemente fue importante para la clasificación se SPI fue solar, índice que también se encuentra presente en la gráfica de las anomalías positivas o de exceso de precipitación, pero en este la posición del índice es media. Para las anomalías negativas, también tienen importancia los índices tni, tsa, oni, niño 34, niño 4, entre otros.

Para las anomalías positivas (Figura 15), los índices más importantes fueron tni, niño 3, niño 12, niño 4, entre otros.

Mapas

Mapas de anomalías negativas

Los siguientes mapas describen el territorio o distribución espacial que se ve afectado por los índices climáticos indicados. Así podemos ver que existen índices que tienen mayor importancia en ciertos territorios de Ecuador en referencia a las anomalías negativas. Se muestran tres ejemplos, el primero de un índice con amplia influencia espacial, el segundo de influencia media, y finalmente de influencia baja.

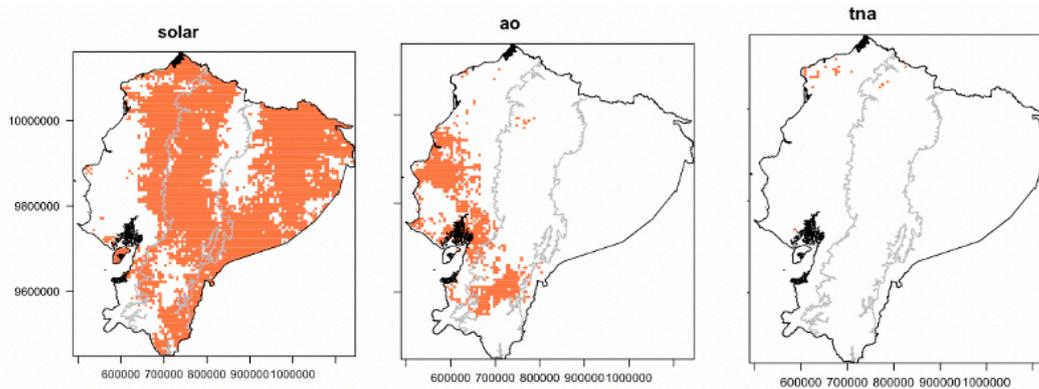


Figura 16 . Ejemplos de mapas de influencia espacial para anomalías negativas.

En el primer mapa de la figura 16 podemos observar que el territorio que es afectado por el índice climático solar que corresponde a la radiación solar es en su mayoría territorio lejano a la costa comprendido por las regiones sierra y oriente del país. En el mapa intermedio, se observa que el índice ao (Influencia del Océano Atlántico) afecta en su mayoría al territorio de la costa correspondiente a las anomalías negativas, es decir influye en la ausencia de precipitación de esta región. Asimismo, podemos observar que no tiene mayor influencia en las regiones sierra y oriente del país. Finalmente, con baja influencia o influencia muy aislada se observa el índice TNA (Tropical North Atlantic). TNA es una serie de tiempo de las anomalías de la temperatura superficial marina promediadas sobre un cuadrante comprendido entre las coordenadas 20-60 °O, 6-18 °N y de 20 a 10 °O, y de 6 a 10 °N (Jiménez-Quiroz, 2014)

Mapas de anomalías positivas

A continuación, se muestran tres ejemplos de influencia alta, media y baja para las anomalías positivas. En el primer mapa de la figura 17, se observa que tni se relaciona en casi todo el territorio del Ecuador con las anomalías positivas. Solar, en el mapa intermedio, tiene una influencia moderada en la amazonia, y finalmente TNA tiene una baja influencia sólo en píxeles aislados en la amazonia.

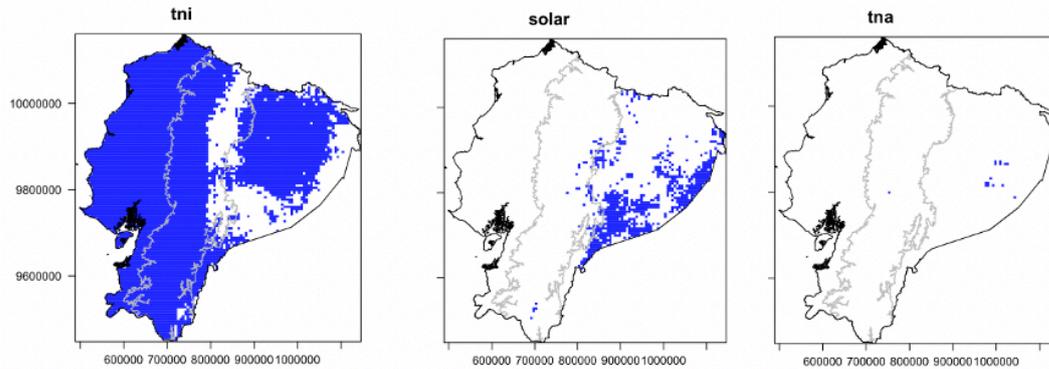


Figura 17 Ejemplos de mapas de influencia espacial para anomalías positivas.

4. CONCLUSIONES

Este trabajo desarrolló el código en R para analizar y relacionar la influencia de índices climáticos en anomalías positivas y negativas de precipitación. Este código calcula SPI.3 a partir de imágenes satelitales de precipitación para el período 1981-2018, y relaciona las anomalías positivas y negativas con diferentes índices climáticos, mediante un modelo Random Forest. A partir de la importancia de variables, se visualiza con gráficos de frecuencias y con mapas, qué índices resultaron más importantes y como es la distribución espacial de dicha importancia. El código desarrollado en este trabajo se encuentra en los anexos de este documento.

4.1 Limitaciones

- El área de estudio se limita al Ecuador continental, el territorio insular no se toma en cuenta para este análisis.
- El objetivo específico inicialmente planteado de realizar un levantamiento general del estado del arte, enfocado en los diferentes métodos de análisis que se han utilizado y contrastarlo con Random Forest, no pudo ser llevado a cabo. Sin embargo, se destaca su relevancia para ser desarrollado en un futuro tema de trabajo de titulación.
- Los datos fueron obtenidos de imágenes satelitales de precipitaciones y se recomienda aplicar este método en estaciones meteorológicas disponibles para validar los resultados aquí mostrados.
- El rango temporal en el que se encuentran los datos es desde el año 1981 al año 2018, por lo que en futuros trabajos se debe extender esta serie temporal.

- Este trabajo mostró el desarrollo metodológico y de implementación de código. En futuros trabajos se plantea analizar y discutir los resultados de la influencia de los índices en concordancia con la literatura relacionada.

4.2 Beneficios

- Se dispone de un código que puede ser reproducible para otras variables climáticas, y series temporales más actuales.
- Este código permite reconocer el grado de influencia que tiene cada índice climático sobre el nivel de precipitación de cada zona del territorio ecuatoriano, lo que puede tener incidencia en la toma de decisiones dentro de los sectores energético y agro productivo con base en la presencia de los diferentes índices climáticos.
- Este código es fácilmente aplicable a otras variables climáticas además de la precipitación, como puede ser temperatura.

4.3 Trabajos futuros

- Identificar si existen índices climáticos que, en conjunto, puedan aumentar o disminuir las precipitaciones en áreas de la zona de estudio.
- Realizar un estudio en conjunto con las áreas energéticas y agro productivas para reconocer las situaciones en las cuales pueden ser afectadas por la presencia o ausencia de precipitaciones.

BIBLIOGRAFÍA

- Avila, R. (2019). Universidad del Azuay Escuela de Ingeniería de Sistemas y Telemática EXPLORACIÓN DE REDES BAYESIANAS PARA DEL ECUADOR Y TELECONEXIONES CLIMÁTICAS Trabajo de graduación previo a la obtención del título de Ingeniero de Sistemas y Telemática Autor : Paúl Ren.
- Breiman, L. (2008). Random Forests - Original Paper. *Vasa*, 1–33. <http://oz.berkeley.edu/~breiman/randomforest2001.pdf> <http://medcontent.meta press.com/index/A65RM03P4874243N.pdf>
- Castro Loayza, E. S. (2014). Evaluación de Imágenes Satelitales de Precipitaciones (TRMM 3B43 mensual) en Región Costa, Sierra y Oriente del Ecuador (Tesis de pregrado). 51. <http://dspace.ucuenca.edu.ec/bitstream/123456789/20874/1/TESIS.pdf>
- Córdoba-Machado, S., Palomino-Lemus, R., Gámiz-Fortis, S. R., Castro-Díez, Y., & Esteban-Parra, M. J. (2016). Seasonal streamflow prediction in Colombia using atmospheric and oceanic patterns. *Journal of Hydrology*, 538, 1–12. <https://doi.org/10.1016/j.jhydrol.2016.04.003>
- De Silva M., T., & Hornberger, G. (2018). Identifying ENSO Influences on Rainfall with Classification Models: Implications for Water Resource Management of Sri Lanka. *Hydrology and Earth System Sciences Discussions*, June, 1–29. <https://doi.org/10.5194/hess-2018-249>
- Del, S., Atlántico, O., La, S., Millano, J., Paredes, F., & Vivas, I. (2007). Efecto de la oscilación meridional (ENSO) y la temperatura superficial del océano atlántico sobre la distribución ... 103. 103–116.
- Electricidad, A. de R. y C. de. (n.d.). Estadística Anual y Multianual del Sector Eléctrico Ecuatoriano.
- Escobar, F., & Aceituno, P. (1998). Influencia del fenómeno ENSO sobre la precipitación nival en el sector andino de Chile Central durante el invierno. In *Bulletin de l'Institut Français d'Études Andines* (Vol. 27, Issue 3, pp. 753–759).
- Fierro, A. O. (2014). Relationships between California rainfall variability and large-scale climate drivers. *International Journal of Climatology*, 34(13), 3626–3640. <https://doi.org/10.1002/joc.4112>
- Francou, B., & Ribstein, P. (1995). Balances de glaciares y clima en Bolivia y Peru: impacto de los eventos ENSO. *Bulletin de l'institut Francais d'etudes Andines*, 24(3), 661–670. http://redpeia.minam.gob.pe/admin/files/item/4df8c6fa1daa5_Balances_de_glaciares_y_clima_en_Bolivia_y_Peru_impacto_de_los_eventos_ENSO.pdf
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific Data*, 2, 1–21. <https://doi.org/10.1038/sdata.2015.66>

- INEC. (n.d.). Tabulados de la Encuesta de Superficie y Producción Agropecuaria Continua ESPAC 2018. http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_agropecuarias/espac/espac-2018/Tabulados ESPAC 2018.xlsx
- Jiménez-Carrión, M., Gutiérrez-Segura, F., & Celi-Pinzón, J. (2018). Modelado y Predicción del Fenómeno El Niño en Piura, Perú mediante Redes Neuronales Artificiales usando Matlab. *Información Tecnológica*, 29(4), 303–316. <https://doi.org/10.4067/s0718-07642018000400303>
- Jiménez-Quiroz, C. (2014). Indicadores Climáticos. Una Manera para identificar la variabilidad climática a escala global. *Inapesca*, 29.
- Liu, Y. C., Di, P., Chen, S. H., & DaMassa, J. (2018). Relationships of rainy season precipitation and temperature to climate indices in California: Long-Term variability and extreme events. *Journal of Climate*, 31(5), 1921–1942. <https://doi.org/10.1175/JCLI-D-17-0376.1>
- Medina-Merino, R. F., & Ñique-Chacón, C. I. (2017). Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *Interfases*, 0(010), 165. <https://doi.org/10.26439/interfases2017.n10.1775>
- Physical Sciences Division. (n.d.). Climate Indices: Monthly Atmospheric and Ocean Time-Series.
- Tobar, V., & Wyseure, G. (2018). Seasonal rainfall patterns classification, relationship to ENSO and rainfall trends in Ecuador. *International Journal of Climatology*, 38(4), 1808–1819. <https://doi.org/10.1002/joc.5297>
- Varela, A. L., & Ron, S. R. (2018). Geografía y clima del Ecuador. <https://bioweb.bio/faunaweb/amphibiaweb/GeografiaClima/>
- World Meteorological Organization. (2012). Standardized Precipitation Index User Guide.
- Wu, R., Hu, Z. Z., & Kirtman, B. P. (2003). Evolution of ENSO-related rainfall anomalies in East Asia. *Journal of Climate*, 16(22), 3742–3758. [https://doi.org/10.1175/1520-0442\(2003\)016<3742:EOERAI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3742:EOERAI>2.0.CO;2)
- Zambrano Mera, Y. E., Rivadeneira Vera, J. F., & Pérez-Martín, M. Á. (2018). Linking El Niño Southern Oscillation for early drought detection in tropical climates: The Ecuadorian coast. *Science of the Total Environment*, 643, 193–207. <https://doi.org/10.1016/j.scitotenv.2018.06.160>

Anexos

Código Fuente en formato Markdown de R

Preparacion de Datos.Rmd

```
---
title: "Preparacion datos"
author: "Sol Contreras, Daniela Ballari"
date: "16 de Septiembre de 2019"
output:
  html_document:
    toc: true # table of content true
    toc_depth: 6 # upto three depths of headings (specified by #, ## and ###)
    toc_float: true
    code_folding: "hide"
---

### Librerias y preparacion de datos
```{r}
library(sp)
library(raster)
library(rasterVis)
library(latticeExtra)
library(mapview)
library(sf)
library(rgdal)
library(RColorBrewer)
library(spplot)
```

## Precipitacion sin tendencia
Precipitacion mensual, resolucion 5km sin tendencia
```{r}
load("i.1981.2018.v.list.dt.RData")
indices<- i.2001.2011.v.list.dt

load("chirps_1981_2018_dt.Rdata")
r.prec5km.df.2
```

###bloque 2
Poligono del Ecuador
```{r}
load("Ecuador_shape_4326.RData") #~/teleconexiones/datos/shp/Ecuador_shape_4326.RData
ecuador<- spTransform(ecuador, CRS("+init=epsg:32717"))#transformacion sistema de
referencia a UTM
prec5km.df<- as(r.prec5km.df.2, "SpatialPointsDataFrame")
Cada lista contiene la serie temporal de cada pixel.
prec5km.df.list.dt <- as.list(as.data.frame(t(prec5km.df@data)))
```

```

#Altitude of 1000m and DEM for visualization
load("curva_nivel_1000m_4326.RData") #cn
cn_32717<- spTransform(cn, CRS("+init=epsg:32717"))

#visualizacion de los 4 primeros rasters
levelPal <- colorRampPalette(brewer.pal(n = 9, 'Blues'))

levelplot(r.prec5km.df.2[[1:4]],col.regions = levelPal, margin=FALSE)+
latticeExtra::layer(sp.polygons(cn_32717, col='grey'))+latticeExtra::layer(sp.polygons(ecuador,
col='black'))

spplot(r.prec5km.df.2[3:6], zlim=c(0,800), col.regions=colorRampPalette(c("white",
"blue"))(255))

...

```{r}
library(ggplot2)
library(SPEI)
...

###Funcion spi.3
##bloque 3
```{r}
spi.raster <- function(spdf) {#pasar un rasterbrick
 spdf <- as(spdf, "SpatialPointsDataFrame") #crear data.frame
 spdf.final<- matrix(nrow = dim(spdf)[1], ncol = dim(spdf)[2])#crea matriz vacia

 for (i in 1:dim(spdf)[1]){#i=1
 spdf.spi<- spi(ts(t(spdf@data[i,]),start=c(1981, 1), end=c(2018, 12), frequency=12),3, ref.start
=c(1981,1), ref.end =c(2018,12), na.rm = TRUE)
 spdf.final[i,]<- t(matrix(spdf.spi$fitted))# se traspone para que cada fila sea 1 pixel con los X
meses
 }
 spdf.final<- data.frame(spdf.final)#transormar matriz a data.frame
 names(spdf.final)<-names(spdf@data)#conservar nombres de columnas

 spdf.final.sp<- SpatialPointsDataFrame(coordinates(spdf),data.frame(spdf.final))
 gridded(spdf.final.sp)<- TRUE
 return(spdf.final.sp)
}
...

##bloque 4
```{r}
spi.prec <- spi.raster(r.prec5km.df.2)

save(spi.prec, file="raster_spi_prec.RData")
# spplot(spi.prec[3:6], zlim=c(0,800), col.regions=colorRampPalette(c("white", "blue"))(255))
...

##bloque 5
###Funcion para clasificar rasterbrick en 6 niveles

```

```

... {r}
prob_drougth_type.2<- function(spdf.final.sp){#spdf.final.sp raster briel

spdf.final.sp.s_moderate<- spdf.final.sp
spdf.final.sp.s_severe<- spdf.final.sp
spdf.final.sp.s_extreme<- spdf.final.sp

spdf.final.sp.l_moderate<- spdf.final.sp
spdf.final.sp.l_severe<- spdf.final.sp
spdf.final.sp.l_extreme<- spdf.final.sp
class<-spdf.final.sp

for (i in 1:ncol(spdf.final.sp)){#i=1
  #ausencia de precipitación
  spdf.final.sp.s_moderate@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) &
  spdf.final.sp@data[,i]<= -1 & spdf.final.sp@data[,i]> -1.5, -1, NA)
  spdf.final.sp.s_severe@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) &
  spdf.final.sp@data[,i]<= -1.5 & spdf.final.sp@data[,i]> -2, -2, NA)
  spdf.final.sp.s_extreme@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) &
  spdf.final.sp@data[,i]<= -2, -3, NA)

  #Anomalías de precipitación: Lluvia intensa
  #spdf.final.sp.l_moderate@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) &
  spdf.final.sp@data[,i]>= 1 & spdf.final.sp@data[,i]< 1.5, 1, NA)
  #spdf.final.sp.l_severe@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) &
  spdf.final.sp@data[,i]>= 1.5 & spdf.final.sp@data[,i]< 2, 2, NA)
  #spdf.final.sp.l_extreme@data[,i] <- ifelse(!is.na(spdf.final.sp@data[,i]) &
  spdf.final.sp@data[,i]>= 2, 3, NA)

  union<-
  data.frame(l_m=spdf.final.sp.l_moderate@data[,i],l_s=spdf.final.sp.l_severe@data[,i],l_e=spdf.
  final.sp.l_extreme@data[,i])

  #union<-
  data.frame(s_m=spdf.final.sp.s_moderate@data[,i],s_s=spdf.final.sp.s_severe@data[,i],s_e=spdf.
  f.final.sp.s_extreme@data[,i],l_m=spdf.final.sp.l_moderate@data[,i],l_s=spdf.final.sp.l_severe
  @data[,i],l_e=spdf.final.sp.l_extreme@data[,i])
  union$merge<- NA
  for(j in 1:nrow(union)){#j=1
    a<- which(!is.na(union[,j]))
    if (length(a)==0) next
    else union$merge[j]<-union[j,a]}

  class@data[,i]<-union$merge

}

class<- class[,-c(1,2)]
return(class) }

...

#bloque 6

```

```

```{r}
spi.prec.clas<-prob_drougth_type.2(spi.prec)
spplot(spi.prec.clas[1:4])

save(spi.prec.clas, file="flo_dataframe_clasificado_spi_prec.RData")
load("flo_dataframe_clasificado_spi_prec.RData")#spi.prec.clas

```

##bloque 7

```{r}
brick.spi.list<- as(spi.prec.clas, "SpatialPointsDataFrame")

brick.spi.list@data[is.na(brick.spi.list@data)]<-0

brick.spi.list <- as.list(as.data.frame(t(brick.spi.list@data)))

RF.data<- lapply(1:length(brick.spi.list), function(i)
data.frame(spi=factor(brick.spi.list[[i]]),data.frame(indices)[-c(1,2),]))

table(RF.data[[1]]$spi)

save(RF.data, file="flo_RFdata.RData")
load("flo_RFdata.RData")
table(RF.data[[1]]$spi)
```

```

Flood.Rmd

```

---
title: "RandomForest Final"
author: "Zol Maria y Daniela Ballari"
date: "10/13/2020"
output: html_document
---
```{r}

install.packages("Rcpp")
library(caret)
library(randomForest)
library(caret)
library(dplyr)
library(reshape2)
library(ggpubr)
library(ggplot2)
library(ggplot)
library(reshape)

install.packages("mice")
install.packages("plotrix")
remotes::install_github("dustinfife/flexplot")
remotes::install_github("dustinfife/fifer")

install.packages("installr", dependencies = TRUE)

```

```

library(installr)
updateR()
```

```{r}
load("flo_RFdata.RData")
set.seed(3795)
```

###bloque 3

```{r}
data<-RF.data[[3795]]
#length(RF.data)
data_train<-fifer::stratified(data, c("spi"), 0.7) #Estratifica la data basada en la variable SPI.
Tomando el 30% de cada nivel registrado para Test y el 70% para entrenamiento.
data_test<- data[-as.numeric(rownames(data_train)),]

table(data_train$spi)
table(data_test$spi)

spi ~ data =
=randomForest(factor(spi) ~ data =
data_train)#sampsiz=c(100,100),strata=data_train[[50]]$spi
#Summary of the model
spi

ad <- rpart(spi~., data = data_train, method = 'class')
rpart.plot(ad, extra = 101)
```

```{r}
#plot error OOB evolution by the number of generated trees
plot(spi)
```

Observations:

##En este bloque se va a realizar la determinación de las variables de importancia que afectan directamente a las excesos de precipitación en todo el Ecuador

```{r}

data_imp<-RF.data[[3795]]
todo<- RF.data

table(data_imp$spi)

spi_imp ~ data =
=randomForest(factor(spi) ~ data =
data_imp)#sampsiz=c(100,100),strata=data_train[[50]]$spi
#Summary of the model
importancia<-varImpPlot(spi_imp)
spi_imp

```

```

ind_imp=c(rownames(importancia)[1][1],rownames(importancia)[2][1],rownames(importancia)
[3][1])
ind_imp
valor=c(importancia[1][1],importancia[2][1],importancia[3][1])
dataf= data.frame(ind_imp,valor)
```


```

t.

```{r}
#Test with independent data
CM_MA_RD<- confusionMatrix(table(data_test$spi, predict(spi,data_test)))

rf.ma<-ggplot(data.frame(melt(prop.table(CM_MA_RD$table))), aes(Var1, Var2,
fill=value))+geom_bin2d() + labs(x="Observed", y="Predicted") + scale_fill_continuous(limits
= c(0,0.7))+ geom_text(aes(label = round(value*nrow(data_test),0), colour="gray"), vjust =
1,show.legend = FALSE) + labs(x="Observed", y="Predicted") + scale_fill_continuous(limits
= c(0,0.7)) + ggtitle("")

rf.ma

ggtexttable(t(as.matrix(round(CM_MA_RD$byClass,3))))
```

```{r}
library(caret)
spi =randomForest(factor(spi) ~ ., data = data_train)
modelo<-spi

set.seed(3795)

predTrainStr2 <- predict(modelo, data_test, type = "class")

table(predTrainStr2, data_test$spi)

CM_test<- confusionMatrix(table(data_test$spi, predict(modelo,data_test)))

CM_test
```

dry.Rmd

```

---
title: "RandomForest Final"
author: "Zol Maria y Daniela Ballari"
date: "10/13/2020"
output: html_document
---
```{r}

install.packages("Rcpp")
library(caret)

```


```


```

```

library(randomForest)
library(caret)
library(dplyr)
library(reshape2)
library(ggpubr)
library(ggplot2)
library(ggplot)
library(reshape)

install.packages("mice")
install.packages("plotrix")
remotes::install_github("dustinfife/flexplot")
remotes::install_github("dustinfife/fifer")

install.packages("installr", dependencies = TRUE)
library(installr)
updateR()
```



```

```{r}
load("dry_RFdata.RData")
set.seed(1234)
```

##bloque 3

```{r}
data<-RF.data[[3795]]
length(RF.data)

data_train<-fifer::stratified(data, c("spi"), 0.7) #Estratifica la data basada en la variable SPI.
Tomando el 30% de cada nivel registrado para Test y el 70% para entrenamiento.
data_test<- data[-as.numeric(rownames(data_train)),]

table(data_train$spi)
table(data_test$spi)

spi ~ randomForest(factor(spi) ~ data =
data_train)#sampsiz=c(100,100),strata=data_train[[50]]$spi
#Summary of the model
spi
```

```{r}

data_imp<-RF.data[[3795]]

table(data_imp$spi)

```


```

```

spi_imp = randomForest(factor(spi) ~ ., data = data_imp)
#sampsiz=c(100,100),strata=data_train[[50]]$spi
#Summary of the model
importancia<-varImpPlot(spi_imp)
spi_imp

ind_imp=c(rownames(importancia)[1][1],rownames(importancia)[2][1],rownames(importancia)
[3][1])
ind_imp
valor=c(importancia[1][1],importancia[2][1],importancia[3][1])
dataf= data.frame(ind_imp,valor)
...

```{r}
#variable importance
varImpPlot(spi)
...

```{r}
#Test with independent data
CM_MA_RD<- confusionMatrix(table(data_test$spi, predict(spi,data_test)))

rf.ma<-ggplot(data.frame(melt(prop.table(CM_MA_RD$table))), aes(Var1, Var2,
fill=value))+geom_bin2d() + labs(x="Observed", y="Predicted") + scale_fill_continuous(limits
= c(0,0.7))+ geom_text(aes(label = round(value*nrow(data_test),0), colour="gray"), vjust =
1,show.legend = FALSE) + labs(x="Observed", y="Predicted") + scale_fill_continuous(limits
= c(0,0.7)) + ggtitle("")

rf.ma

ggtexttable(t(as.matrix(round(CM_MA_RD$byClass,3))))
...

#crea el modelo para el Random Forest
```{r}
spi =randomForest(factor(spi) ~ ., data = data_train)
modelo<-spi

set.seed(3795)

predTrainStr2 <- predict(modelo, data_test, type = "class")

table(predTrainStr2, data_test$spi)

CM_test<- confusionMatrix(table(data_test$spi, predict(modelo,data_test)))

CM_test
...

```

Global_floo.Rmd

```
title: "Global_floo"
author: "Zol Maria"
date: "12/7/2020"
output: html_document
```

```
```{r}
```

```
install.packages("Rcpp")
install.packages("caret")
install.packages("randomForest")
install.packages("caret")
install.packages("dplyr")
install.packages("reshape2")
install.packages("ggpubr")
install.packages("ggplot2")
install.packages("ggplot")
install.packages("reshape")
install.packages("remotes")
install.packages("hash")
install.packages("mice")
install.packages("plotrix")
remotes::install_github("dustinfife/flexplot")
remotes::install_github("dustinfife/fifer")
```

```
library(caret)
library(caret)
library(dplyr)
library(reshape2)
library(ggplot2)
library(ggpubr)
#library(ggplot) -> no hay
library(reshape)
library(data.table)
library(hash)
library(parallel)
```

```
```
```

Random Forest para sequía.

En este bloque se carga la data correspondiente a la sequía registrada en el país

```
```{r}
load("../flo_RFdata.RData")
#set.seed(1234)-> para el valor del pixel a trabajar
```
##bloque 3
```

```

##Función para aplicar el análisis de 1 px sobre todos los datos de Ecuador.
```{r}
analyzePixel <- function(inputData){
 library(hash)
 entradaSPI<-inputData["spi"]
 if (length(entradaSPI[entradaSPI != 0]) == 0) {
 dataFrameResult <- data.frame(
 "nombres" = c("", ""),
 "valores" = c(0,0,0))
 return(dataFrameResult)
 }

 library(randomForest)
 #inputPixel=55
 #inputData<-RF.data[[inputPixel]]

 data_train<-fifer::stratified(inputData, c("spi"), 0.7) #Estratifica la data basada en la variable
 SPI. Tomando el 30% de cada nivel registrado para Test y el 70% para entrenamiento.
 data_test<- inputData[-as.numeric(rownames(data_train)),]
 spi =randomForest(factor(spi) ~ ., data = data_train)# Modelo de Random Forest.
 simp =randomForest(factor(spi) ~ ., data = inputData)#Variables de importancia obtenidas
 gracias al modelo
 importancia<-varImpPlot(simp, TRUE)
 names<- rownames(importancia)
 values<-importancia[,1]
 dataFrameOrder <- data.frame(
 "nombres" = names,
 "valores" = values)
 dataFrameOrderResult<-dataFrameOrder[order(-dataFrameOrder$valores),]
 names<-dataFrameOrderResult$nombres[0:3]
 values<-dataFrameOrderResult$valores[0:3]
 dataFrameResult <- data.frame(
 "nombres" = names,
 "valores" = values)
 #hashResult<-hash()
 #hashResult[[paste(inputPixel, "", sep = "")]]<-dataFrameResult

 #plot(spi) --> dibuja

 #registro<-data.frame(pixel, ind_imp, valor)
 #rblid(registro,import)
 #return (hashResult)
 return (dataFrameResult)
}

```

##Creación de variable global donde se Almacenada, las tres variables de importancia, clasificada
dos por píxel

```{r}
no_cores <- detectCores()

```

```

clust <- makeCluster(no_cores)
#inputData<-RF.data[1:55]
#system.time(dataFrameResult<-parLapply(clust, inputData, analyzePixel))
time<-system.time(dataFrameResult<-parLapply(clust, RF.data, analyzePixel))
saveRDS(dataFrameResult, file = "flo_dataFrameResult.rds")
my_data <- readRDS("flo_dataFrameResult.rds")

```

```

plot(my_data,my_data)
stopCluster(clust)
```

```

```

```{r}
my_data[[3]][1]
r2<- do.call(rbind.data.frame, my_data)
barplot(table(r2$nombres), las=2)
```

```

Global_dry.rmd

```

---
title: "Global4_dry"
author: "Zol Maria"
date: "12/7/2020"
output: html_document
---

```

```

```{r}

install.packages("Rcpp")
install.packages("caret")
install.packages("randomForest")
install.packages("caret")
install.packages("dplyr")
install.packages("reshape2")
install.packages("ggpubr")
install.packages("ggplot2")
install.packages("ggplot")
install.packages("reshape")
install.packages("remotes")
install.packages("hash")
install.packages("mice")
install.packages("plotrix")
remotes::install_github("dustinfife/flexplot")
remotes::install_github("dustinfife/fifer")

```

```

library(caret)
library(caret)
library(dplyr)

```

```

library(reshape2)
library(ggplot2)
library(ggpubr)
#library(ggplot) -> no hay
library(reshape)
library(data.table)
library(hash)
library(parallel)

```

```

```

```

Random Forest para sequía.

En este bloque se carga la data correspondiente a la sequía registrada en el país

```

```{r}
load("../dry_RFdata.RData")
#set.seed(1234)-> para el valor del pixel a trabajar
```

```

##Función para aplicar el análisis de 1 px sobre todos los datos de Ecuador.

```

```{r}
analyzePixel <- function(inputData){
 library(hash)
 entradaSPI<-inputData["spi"]
 if (length(entradaSPI[entradaSPI != 0]) == 0) {
 dataFrameResult <- data.frame(
 "nombres" = c(", ", ", "),
 "valores" = c(0,0,0))
 return(dataFrameResult)
 }
}

```

```

library(randomForest)
#inputPixel=55
#inputData<-RF.data[[inputPixel]]

```

data\_train<-fifer::stratified(inputData, c("spi"), 0.7) #Estratifica la data basada en la variable SPI. Tomando el 30% de cada nivel registrado para Test y el 70% para entrenamiento.

```

data_test<- inputData[-as.numeric(rownames(data_train)),]
spi =randomForest(factor(spi) ~ ., data = data_train)# Modelo de Random Forest.
simp =randomForest(factor(spi) ~ ., data = inputData)#Variables de importancia obtenidas gracias al modelo

```

```

importancia<-varImpPlot(simp, TRUE)
names<- rownames(importancia)
values<-importancia[,1]
dataFrameOrder <- data.frame(
 "nombres" = names,
 "valores" = values)
dataFrameOrderResult<-dataFrameOrder[order(-dataFrameOrder$valores),]
names<-dataFrameOrderResult$nombres[0:3]
values<-dataFrameOrderResult$valores[0:3]
dataFrameResult <- data.frame(
 "nombres" = names,
 "valores" = values)

```

```

#hashResult<-hash()
#hashResult[[paste(inputPixel, "", sep = "")]]<-dataFrameResult

#plot(spi)--> dibuja

#registro<-data.frame(pixel, ind_imp, valor)
#rblid(registro,import)
#return (hashResult)
return (dataFrameResult)
}

...

##Creación de variable global donde se Almacenada, las tres variables de importancia, clasificada
dos por píxel

```{r}
no_cores <- detectCores()
clust <- makeCluster(no_cores)
#inputData<-RF.data[1:55]
#system.time(dataFrameResult<-parLapply(clust, inputData, analyzePixel))
time<-system.time(dataFrameResult<-parLapply(clust, RF.data, analyzePixel))
saveRDS(dataFrameResult, file = "dataFrameResult.rds")
my_data <- readRDS("dry_dataFrameResult.rds")

plot(my_data,my_data)
stopCluster(clust)
```

...

```{r}
my_data[[3]][1]
r2<- do.call(rbind.data.frame, my_data)
barplot(table(r2$nombres), las=2)
```

...

read_results_dry.Rmd

title: "read_results_dry"
output: html_document

```{r}
r = readRDS("../dry_dataFrameResult.rds")

# Asignar numero de pixel en cada data.frame
names(r)<- 1:8073

for(i in 1:8073){
  r[[i]]$pixel <- names(r)[[i]]
}

```

```

# Convertir a data frame
r2<- do.call(rbind.data.frame, r)
r2<-r2[!c(r2$nombrs==""),] #hay algunas celdas de nombres de indices que estan vacios, puede
ser que coincidan con las celdas donde no hay sequias?? Les he omitido a partir de esta fila

#grafico de barras
barplot(sort(table(r2$nombrs), decreasing=TRUE), las=2)

...

#Mapas
```{r}

library(raster)
library(rasterVis)

#Datos para recuperar la estructura raster

load("chirps_1981_2018_dt.RData")
load("Ecuador_shape_4326.RData")
ecuador<- spTransform(ecuador, CRS("+init=epsg:32717"))#transformacion sistema de
referencia a UTM

#Altitude of 1000m and DEM for visualization
load("curva_nivel_1000m_4326.RData") #cn
cn_32717<- spTransform(cn, CRS("+init=epsg:32717"))

#estructura raster
raster<- mask(r.prec5km.df.2[[1]], ecuador)

#convertir a data.frame con coordenadas
raster.df<- as(raster, "SpatialPointsDataFrame")

gridded(raster.df)<-TRUE

raster.index<- raster(raster.df)

#create pixel id en el raster
raster.df$id <- 1:8073

listado de indices
index<-unique(r2$nombrs)

i="tni"

for(i in index){

r2.index<- subset(r2, nombrs== i,)#select todos los pixeles de un indice en particular

raster.df2<-merge(raster.df[, "id"], r2.index[,c("pixel", "nombrs")], by.x="id", by.y="pixel")
raster.df2[!is.na(raster.df2$nombrs),]<-1

```

```

gridded(raster.df2)<-TRUE

raster.df3<- brick(raster.df2)
names(raster.df3[[2]])<-i

raster.index<-addLayer(raster.index, raster.df3[[2]])

}
raster.final<- raster.index[[2:23]]

for(i in 1:222) { #i=10
print(levelplot(raster.final[[i]], margin=FALSE,col.regions = c("coral"), colorkey=FALSE,
main=index[[i]]+ latticeExtra::layer(sp.polygons(cn_32717,
col='grey'))+latticeExtra::layer(sp.polygons(ecuador, col='black')))
}

```

...

## read\_results.Rmd

```

title: "Untitled"
output: html_document

```{r}
r = readRDS("floo_dataFrameResult.rds")

# Asignar numero de pixel en cada data.frame
names(r)<- 1:8073

for(i in 1:8073){
  r[[i]]$pixel <- names(r)[[i]]
}

# Convertir a data frame
r2<- do.call(rbind.data.frame, r)
r2<-r2[!c(r2$nombre==""),] #hay algunas celdas de nombres de indices que estan vacios, puede
ser que coincidan con las celdas donde no hay sequias?? Les he omitido a partir de esta fila

#grafico de barras
barplot(sort(table(r2$nombre), decreasing=TRUE), las=2)

...

#Mapas

```{r}

library(raster)
library(rasterVis)

#Datos para recuperar la estructura raster

```

```

load("chirps_1981_2018_dt.RData")
load("Ecuador_shape_4326.RData")
ecuador<- spTransform(ecuador, CRS("+init=epsg:32717"))#transformacion sistema de
referencia a UTM

#Altitude of 1000m and DEM for visualization
load("curva_nivel_1000m_4326.RData") #cn
cn_32717<- spTransform(cn, CRS("+init=epsg:32717"))

#estructura raster
raster<- mask(r.prec5km.df.2[[1]], ecuador)

#convertir a data.frame con coordenadas
raster.df<- as(raster, "SpatialPointsDataFrame")

gridded(raster.df)<-TRUE

raster.index<- raster(raster.df)

#create pixel id en el raster
raster.df$id <- 1:8073

listado de indices
index<-unique(r2$nombrs)

i="tni"

for(i in index){

r2.index<- subset(r2, nombrs== i,)#select todos los pixeles de un indice en particular

raster.df2<-merge(raster.df[, "id"], r2.index[,c("pixel", "nombrs")], by.x="id", by.y="pixel")
raster.df2[!is.na(raster.df2$nombrs),]<-1

gridded(raster.df2)<-TRUE

raster.df3<- brick(raster.df2)
names(raster.df3[[2]])<-i

raster.index<-addLayer(raster.index, raster.df3[[2]])

}
raster.final<- raster.index[[2:23]]

for(i in 1:222) { #i=10
print(levelplot(raster.final[[i]], margin=FALSE,col.regions = c("coral"), colorkey=FALSE,
main=index[[i]]+ latticeExtra::layer(sp.polygons(cn_32717,
col='grey'))+latticeExtra::layer(sp.polygons(ecuador, col='black')))
}

...

```