

Article

A seasonality and a zoning model for Puerto Rico extracted from ten years of MODIS data

Israel O Dilán-Pantojas ^{1,†}, Marcelo Francia ^{2,‡}, and Manuel J. La Torre-Poueymirou ^{3,‡}

¹ University of Puerto Rico, Rio Piedras Campus; israe.dilan@upr.edu

² University of Puerto Rico, Rio Piedras Campus; marcelo.francia@upr.edu

³ University of Puerto Rico, Rio Piedras Campus; manuel.latorre@upr.edu

* Correspondence: israel.dilan@upr.edu; Tel.: +1-787-241-4148

† Current address: 14 Ave. Universidad Ste. 1401, San Juan, PR 00925-2534

‡ These authors contributed equally to this work.

Version May 15, 2019 submitted to *Remote Sens.*

Abstract: This work attempts to create a zoning and seasonality model of Puerto Rico by utilizing 10 years of MODIS data. The motivation for this work comes from the necessity to mitigate the potential damage that climate change and other atmospheric disturbances can effect on the island. Through spatial analysis of distinct variables using exploratory analysis and unsupervised clustering methods, the markedly distinct grouping of zones across the island was observed. A temporal analysis consisting of an exploratory analysis, statistical pattern analysis and trend analysis of the average value of the same variables over time, yielded evidence of abrupt breaks and some changes in the seasonal patterns of the Puerto Rico region, along with a model that proposed an increasing trend in LST and a slight reduction in the EVI trend. Interactions between these variables may prove useful for the construction of preventative strategies to protect vulnerable regions.

Keywords: MODIS; Enhanced Vegetation Index; Land Surface Temperature; Unsupervised Clustering; Autoregressive Integrated Moving Average model; Seasonality; Zoning; Puerto Rico

1. Introduction

Over the last decade, the world has experienced immense climatological changes with the continuous growth of industrialization, and human population. Humanity has been conscient, or at least warned, of the world's environmental changes along with the divulgence of the idea of what global warming is, through social media and other means communication. Today experiencing longer winters, longer drought seasons as well as discerning the increasing number of major storm formations all around the globe, consciousness on this matter has become relatively common to the average individual. The abnormal climatological events Puerto Rico has experienced in the last decade are an example of our environment's general reaction to the adversities it has faced since the accumulation of human waste as a result of the industrial growth along with human population increase. In the last 10 years, Puerto Rico has undergone persisting drought seasons, forest fires, as well as experience re-occurring catastrophic landfall events, like Hurricane Maria and Irma in 2017. Puerto Rico has also experienced a redistribution of age groups with the younger populations migratory patterns out of the island, which has left Puerto Rico with an increasingly elderly population. (*)

Remote sensing databases are currently being used for characterization of ecosystem structure, diversity, and function [27]. Due to the spatial and temporal dimensions that these databases contain, they have facilitated the study of ecological spatial distributions [11] and seasonal trends [2] of vegetation variables across land regions. Integrating these types of data sets into land models enables the monitoring and prediction of biomass production [17], study vegetation response to growth during

32 climatological disturbing events [20], as well as study climate and land change [5]. The MODIS
33 instrument in the NASA Terra spacecraft expresses features of the land that are commonly used for
34 wildfire monitoring, global temperature changes and ecosystem dynamics [7]. According to Wulder,
35 MODIS enables regular monitoring of global vegetation productivity, but not a spatially detailed
36 characterization of vegetation. Thus, the MODIS data can be used to describe and study general trends
37 of vegetation state and global temperature among other characteristics of our biome. With these tools
38 and data sets, available to the world's general public, through NASA's AppEEARS portal [4,18,25],
39 we will express a 10-year zoning and seasonality model of the archipelago of Puerto Rico. 10 years'
40 worth of data were selected due to data set availability from 2008 to 2018 in NASA's AppEEARS portal
41 database [4,18,25]. No evidence presenting a zoning and seasonality model based on remote sensing
42 data has been found in the existing literature. With this model, we hope to spread awareness of Puerto
43 Rico's previous, current and future ecological and climatological status to the numerous governmental
44 agencies, with the purpose of contributing to the optimization of preparedness of forest preservation
45 and emergency management teams when dealing with natural catastrophes.

46 Seasonality is a distinctive behavioral characteristic of a time series data summary, where a graph
47 may present predictable change when recurring over time. Zoning, containing numerous forms of
48 definition; in our study zoning, is defined as the classification of land area by characteristics pertaining
49 to the remote sensing variables being used in this study. This zoning and seasonality model will be
50 generated by analyzing the Enhanced vegetation index (EVI) and land surface temperature (LST) data
51 sets of a plotted area containing Puerto Rico's archipelago.

52 LST provides information regarding land surface temperature in Kelvin, that we then converted
53 to Celsius. Data sets for LST are generated by the averaging 8-day per-pixel 1000m resolution scans
54 of land through NASA's Terra MODIS spacecraft, MOD11A2 [14]. NDVI provides consistent spatial
55 and temporal comparisons of vegetation canopy greenness, a composite property of leaf area, and
56 canopy structure [4,15]. The Terra spacecraft generates vegetation indices 16-day composite eight days
57 apart, and are retrieved from daily, atmosphere-corrected, bidirectional surface reflectance scan by
58 the NASA's Terra MODIS spacecraft, MOD13Q1 [4]. Clouded pixels are removed from the produced
59 NDVI and EVI data packages with the purpose of eliminating low quality and absent data recordings
60 with the purpose of eliminating non-significant memory space [15]. NDVI has previously been used to
61 assess the vegetation response to episodic disturbance events [20]. Considering possible relationships
62 between LST and NDVI, both have been previously used to quantify live fuel moisture content in
63 forests to estimate fire danger rating [3]. Additionally, both are currently used for agricultural drought
64 monitoring and food security status assessment in various countries [8]. Both variables, have the
65 potential to be used for projecting ecological and climate risk factor situations, with the purpose of
66 ameliorating preparedness when facing natural anomalies in the archipelago of Puerto Rico.

67 2. Materials and Methods



Figure 1. The polygon utilized to extract data for the region of interest, from AppEEARS.

68 In order to extract patterns and create models which may be used to optimize Puerto Rico's public
69 policy, a significant amount of relevant data is necessary. This research focuses on utilizing some of the
70 vast amounts of freely available, high-quality data obtained through the MODIS instrument cluster
71 from the Terra satellite. These instruments record moderate resolution images of our planet's surface
72 light reflectance captured by a spectroradiometer, from which different information variables are
73 extracted. This study centers on the Enhanced Vegetation Indices (EVI) and Land Surface Temperature
74 (LST) variables as recorded by MODIS, through a ten-year time-period from January 1st, 2018 through
75 December 31st, 2018. This time period was chosen because it contained an appropriate window of
76 observations within the constraints of available data for both variables.

77 EVI was used instead of NDVI due to enhanced sensitivity for high biomass regions and reduced
78 atmospheric influence [10] on measurements, that EVI presents over traditional NDVI. EVI is a
79 composite of near-infrared, red and blue wavelength measures, used to calculate chlorophyll density
80 contained in existing vegetative cover[9]. EVI is commonly used for monitoring vegetation condition
81 in various parts of the world[21]. We used the MOD13Q1 product, which provides EVI data every
82 16 days at a 250-meter spatial resolution. Measurements were scaled by the 0.001 factor required for
83 appropriate use of the data, as appointed by NASA's AppEEARS portal(*). The LST data was obtained
84 from the MOD11A2 product. LST is a measure of the average surface temperature over an 8 days time
85 window, expressed as Kelvin units, within a 1-kilometer spatial resolution. This data is specific to the
86 temperature of the land surface, not taking into consideration air temperature when measurements
87 are made. Measurements were scaled by the 0.02 factor required for appropriate use of the data, as
88 appointed by NASA's AppEEARS portal(*). Kelvin where converted to degrees Celsius by subtracting
89 273.15, for our analysis.

90 Our Zoning model consisted of an exploratory data analysis and an unsupervised clustering
91 approach in order to extract information about trends in the spatial data. The exploratory analysis
92 consisted of observing the mean pixel value over time and frequency analysis of pixel value for the
93 whole plot. For the clustering of zones across the Puerto Rico archipelago, unsupervised clustering
94 was performed grouping the zones from 2 to 11 clusters, scoring a stratified sample of 2,000 pixels
95 within each zone, and then selecting the cluster with the best overall score. This method was used
96 to obtain the best spatial representation of different possible clusters. This unsupervised clustering
97 method was chosen because it does not require training samples, and it has been previously used
98 for the classification of remotely sensed images[1]. Two algorithms were used, K-means (KM) and
99 Clustering LARge Application (CLARA), *****which would output clusters where within-group
100 distances are minimized and between-group separation is maximized.*****The K-means algorithm
101 does this through defining the centroid of each cluster with its mean value, assigning the data to the
102 cluster with the nearest mean. The CLARA algorithm uses medoids, which are the most centrally
103 located point in the cluster, from different randomly selected data-sets and comparing them to the
104 original data, looking for the arrangement with less dissimilarity. Each approach was compared using
105 the silhouette index to obtain the best number of clusters. The silhouette index value for an object
106 indicates how well it is matched to its own cluster and how poorly it relates to other neighboring
107 clusters.

108 The Seasonality analysis was performed on the average value for the whole Puerto Rico region on
109 each time-point and it consisted on three phases; an initial exploratory analysis, an examination of
110 abrupt changes on seasonal trend patterns over time and the creation of a model/forecast with the
111 information obtained from the downloaded data products. The exploratory analysis on time-series data
112 consisted on fitting a line through linear regression and a curve in order to express clearer observable
113 trends and patterns contained in the data

114 One of these more granular analysis is the Breaks For Additive Season and Trend (BFAST)
115 approach to detect tendencies and seasonal changes in our time series data [24]. This method
116 iteratively estimates the number and the time of drastic changes within a time series, allowing the
117 characterization of the change by its magnitude and direction. Finally, for model/forecast preparation

the Auto-Regressive Integrated Moving Average(ARIMA) algorithm was used. This model is utilized in order to detect tendencies and seasonal changes in our averaged time series data. This kind of model is applied to non-stationary data and it uses past values in the regression equation for the time series in order to inform automatic model creation. The size of our 240 data points for EVI and 480 data points for LST of time series data exploits the ARIMA's high reliance on previous data points. This model creation approach has already been used to forecast NDVI in coniferous areas [6] and for drought monitoring [8].

All analysis were performed utilizing R-Statistical Software and the code for this project is found in Appendix D.Code.

3. Results

3.1. Zoning analysis

Results for the zoning analysis are divided into two segments, one is based on an exploratory analysis of the data and the other part is based on the unsupervised clustering approach. The first, exploratory analysis refers to the EVI in the spatial dimension, shown in Figure 2.A. Figure 2.B presents the average amount of vegetation on the island is of substantial density as observed around the 0.51 EVI frequency.

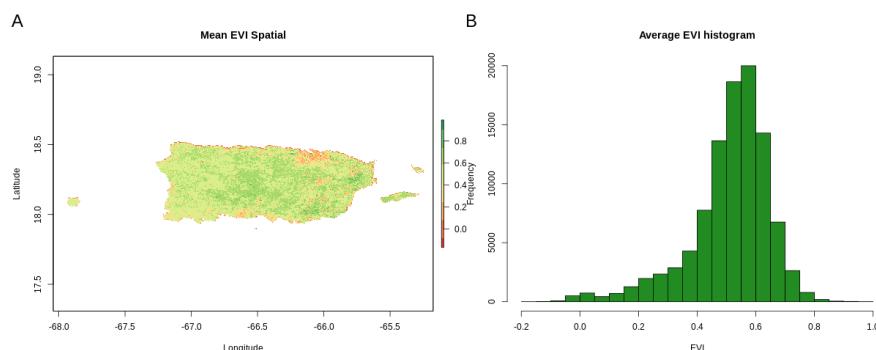


Figure 2. 2.A Spatial representation average EVI over 10 years. 2.B EVI frequency centers around 0.51 value.

Similarly, Figure 3.A shows 2 significantly distinct spatial regions between high-temperature areas and low-temperature areas. Figure 3.B shows the average frequency of LST in the whole archipelago, which over the last decade ranges from 24° to 26° degrees Celsius.

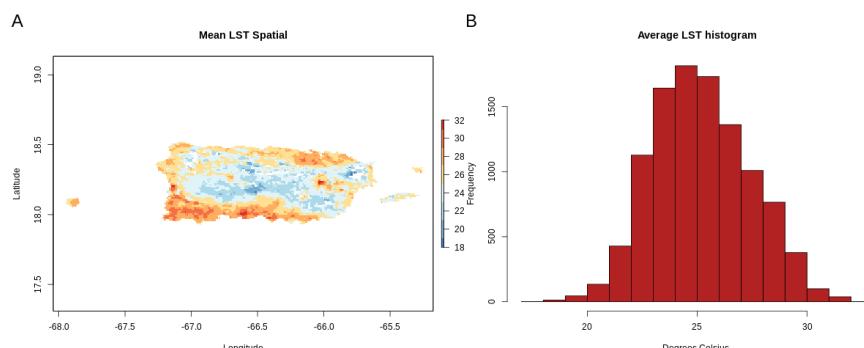


Figure 3. 3.A Spatial representation average LST over 10 years. 3.B LST frequency centers around 25.19° Celsius value.

137 After performing the clustering analysis figures for all clustering runs presented in supplementary
 138 material on Appendix A and evaluating the results corresponding figures and tables in Appendix B,
 139 the highest scoring clusters are presented in Figure 5 and Figure 6.

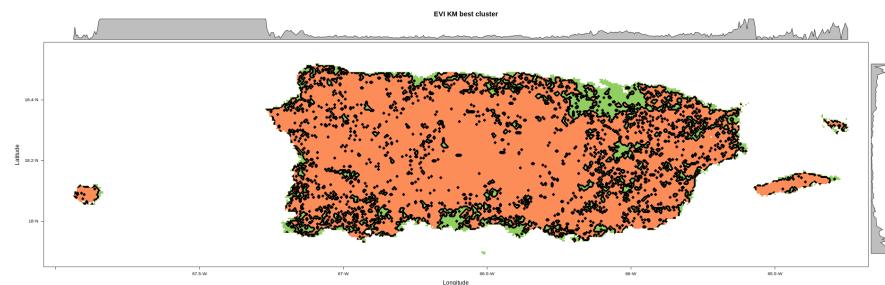


Figure 4. Best cluster number distribution according to the silhouette index for enhanced vegetation index across the Puerto Rico archipelago for the whole time series. Produced by the K-means algorithm

140 Figure 4 depicts the cluster classification with the highest silhouette index for EVI, achieved by
 141 the k-means algorithm. Two main clustered regions were identified, one in orange representing areas
 142 with low EVI, and a green region representing areas with high EVI. White represents areas without
 143 EVI cluster classification.

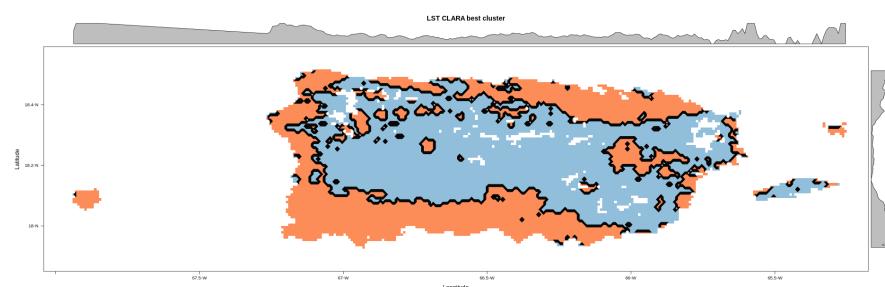


Figure 5. Best cluster number distribution according to the silhouette index for land surface temperature across the Puerto Rico archipelago for the whole time series. Produced by the CLARA algorithm

144 In figure 5 we present the cluster classification with the highest silhouette index for LST, obtained
 145 with the CLARA algorithm. Two main clustered regions were identified, blue zones cover areas with
 146 low LST, and orange zones cover areas with high LST. White represents areas without LST cluster
 147 classification.

148 3.2. Seasonality analysis

149 Seasonality analysis results are divided into three segments, the first segment is based on the
 150 exploratory analysis of the data focused on observing the yearly trends, the second segment contains

151 the results for the BFEST analysis of abrupt pattern deviations, and the third segment that is based on
 152 the utilization of the ARIMA approach to presenting a model and a forecast for seasonality behavior of
 153 EVI and LST in the next two years, namely 2019 and 2020.

154 The linear models presented, suggest a decrease in average EVI value Figure 6.A and an increase
 155 in average LST value Figure 6.B, while the curves suggest a seasonal and periodic repeat of patterns
 156 for both variables. It is important to note the variation in the amplitude and frequency of the EVI curve
 157 in the periods late 2009, late 2015, and late 2016 when compared to the rest of the curve in Figure 6.A,
 158 similarly it is important to observe the variation in the amplitude and frequency the LST curve in the
 159 periods 2008-2019 and late 2015 in Figure 6.B. The ARIMA model analysis are presented in Appendix
 160 C.

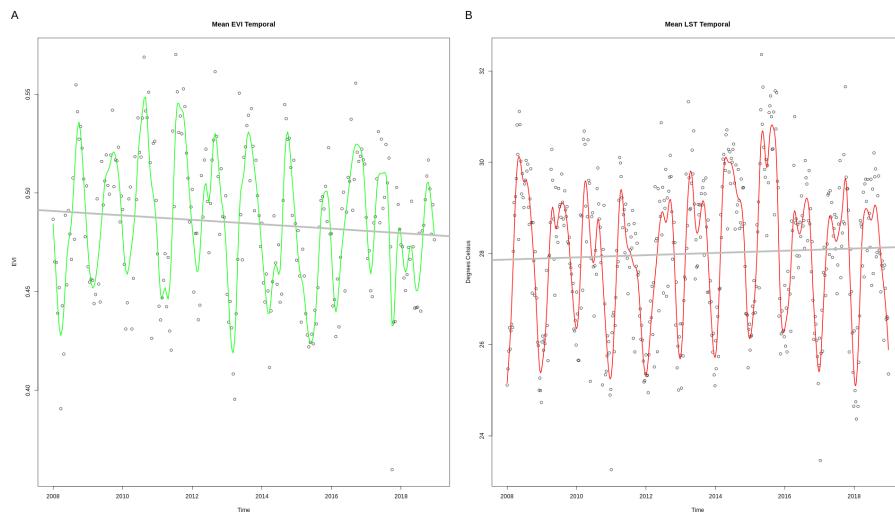


Figure 6. Lineal trends across the whole time series for EVI(A) and LST(B)

161 Closer examination and comparison of yearly behavior show deviations from yearly trends in
 162 years 2010, 2011, 2015, 2017, and 2018, further analysis was required in order to determine if this
 163 changes where significant when compared to the rest of the trend Figure 7.A. Similarly, deviations
 164 from yearly trends in years 2008, 2010, 2011, 2014, and 2015, further analysis was required in order to
 165 determine if this changes where significant when compared to the rest of the trend Figure 7.B.

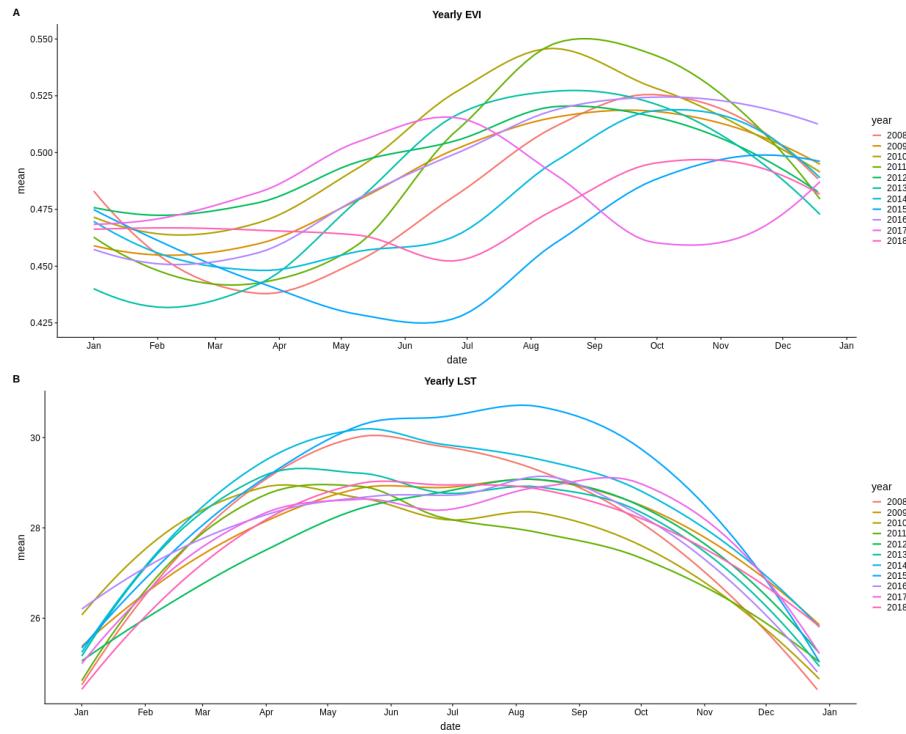


Figure 7. Yearly average change for EVI(A) and LST(B) for each year of the whole time series

166 In order to observe the significance of the changes, the BFAST analysis was performed. The
 167 results are contained in figure 8 for EVI trends and Figure 9 for LST trends. It is shown in the trend
 168 analysis, Figure 8.Tt, that significant changes in EVI trend occurred during the years 2015 and late 2017.
 169 Similarly, the trend analysis, Figure 9.Tt, shows significant deviations from the trend of LST average
 170 values that occurred during the years 2010 and late 2015.

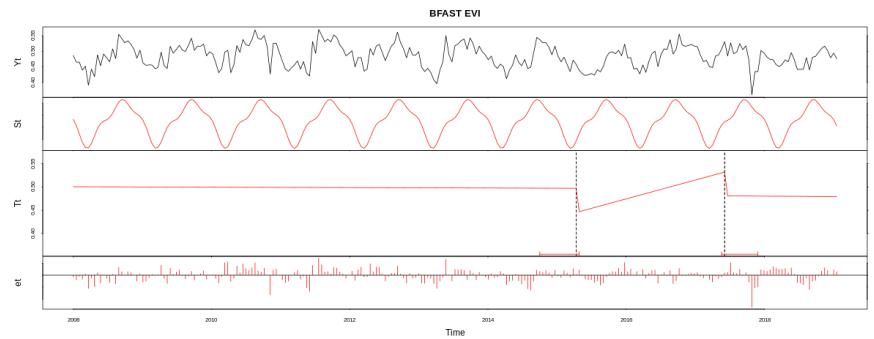


Figure 8. BFAST analysis results on EVI.

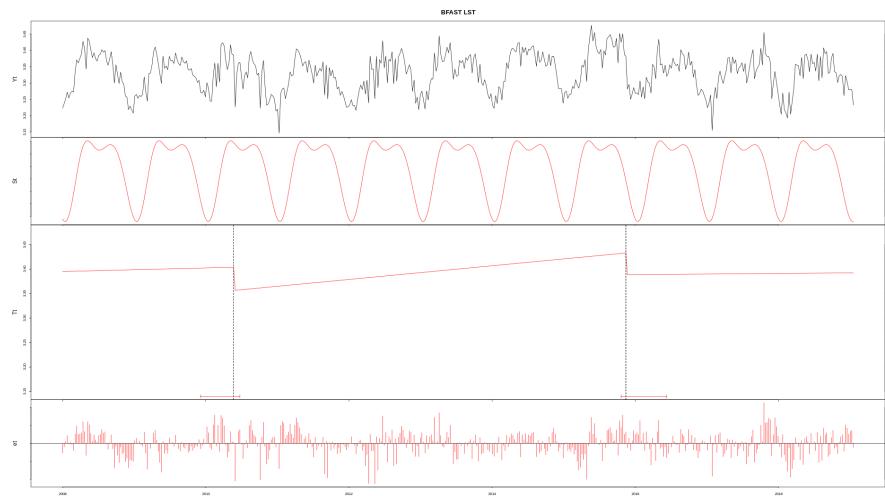


Figure 9. BFAST analysis results on LST.

171 Model creation was automatized with the utilization of the ARIMA algorithm, this model was
 172 then used to create a forecast. The forecast results are presented in Figures 10 for the EVI model and
 173 Figure 11 for the LST model, suggesting a possible behavior of the trends for the next year in the
 174 colored regions of the graph.

175 In Figure 10. EVI suggests a trend for the next year with an 80% confidence in darker purple
 176 regions. However, a 90% confidence, for the behavior of the trend is suggested with lighter purple
 177 regions on the graph. These purple regions show the possible directions the EVI trend might take in
 178 the next year. These projections are extracted from our accounted data considering 10 years of previous
 179 EVI in the archipelago of Puerto Rico that had been used for our exploratory analysis.

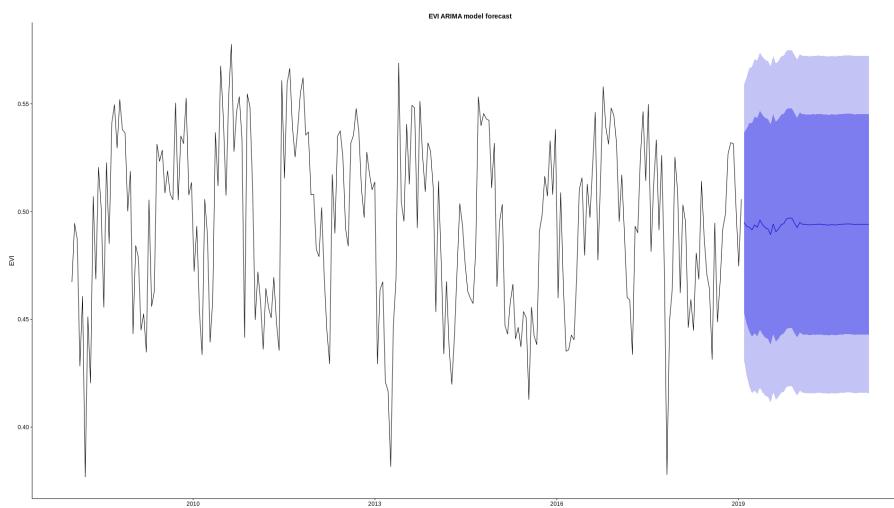


Figure 10. ARIMA EVI seasonal model shows the observed behavior of average EVI (Black) and a forecast (Blue) for the following two years.

180 LST forecast for years 2019 and 2020 results contained in Figure 11, closely follow the observed
 181 pattern of previous years, it is worth noting that the expected average LST peaks and valleys both at
 182 the 80% and 90% confidence levels seem to accommodate peaks and valleys close to the highest and
 183 lowest observed in previous years.

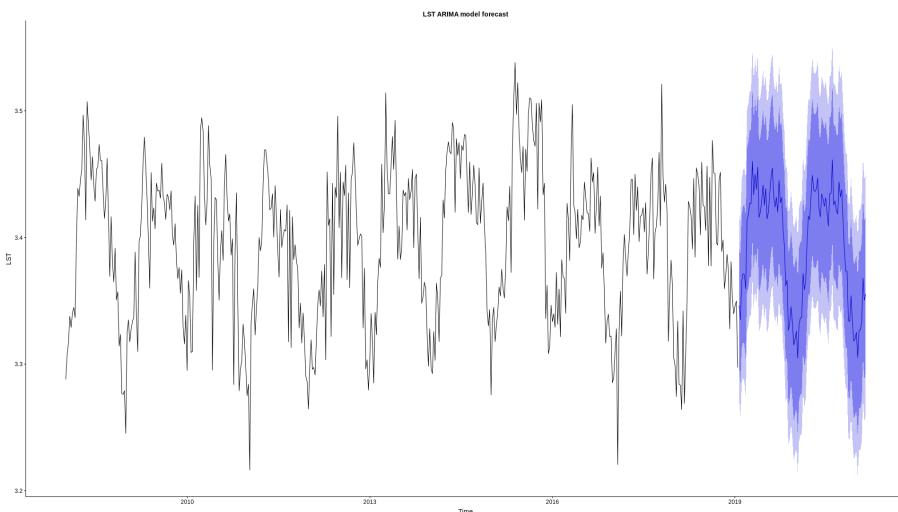


Figure 11. ARIMA LST seasonal model shows the observed behavior of average LST (Black) and a forecast (Blue) for the following two years.

184 4. Discussion

185 The zoning model was created to asses the adequate number of zones which should be considered
 186 by the Puerto Rican government when executing public policy. The exploratory analysis of the EVI
 187 in figure 2.A showed us vegetation rich areas in the interior of the island, which is mainly composed
 188 of forest and herbaceous pasture or agriculture[26]. It also suggests a grouping of lower vegetation
 189 areas such as the ones surrounding the coasts, and some in densely populated areas such as the San
 190 Juan, Caguas, and Ponce areas. Figure 2.B illustrates the frequency of all pixel values of EVI across the
 191 Puerto Rico archipelago. This also shows that there is a larger amount of areas with high EVI when
 192 compared with areas of low EVI.

193 The exploratory analysis of LST in Figure 3.A shows two significantly distinct spatial regions
 194 between high-temperature areas and low-temperature areas. This difference in zones happens to be
 195 very similar to the grouping observed when exploring the EVI data. Coastal LST is on average higher
 196 than that for the center of the island, except for densely populated areas such as the San Juan, Caguas,
 197 and Ponce areas. However, the smaller the patch of land of the composing archipelago, the more
 198 density for higher LST may be present. It is important to observe the overlap between a region around
 199 17.9° to 18.2° North and 67.3° to 66.3° West of moderately high EVI values near 0.4 and high LST
 200 nearing the 30° to 32° Celsius, this region seems more prone to spontaneous wildfires.

201 From this exploratory observations, we can infer the presence of very distinct zones which might
 202 share common characteristics that could play a major factor on the general ecological health on the
 203 island, when it comes to the effects of climate change as experienced through seemingly abnormal
 204 seasonal patterns.

205 Figure 4 depicts the cluster classification for the EVI. We can mainly identify two regions, a large
 206 area with high EVI and areas with low EVI, located either in highly urbanized regions[12] or in the
 207 shore. This kind of cluster classification does not capture the gradient of vegetation that other studies
 208 have shown[26], but this is because the MODIS does not provide enough definition for vegetation
 209 classification[27]. It does, however, illustrate the general trend of vegetation distribution between high
 210 urbanized and low urbanized areas.

211 For the cluster classification with the highest silhouette index for LST in Figure 5, we observe that
 212 highly urbanized zones and their surrounding areas are classified within the high LST cluster. Other
 213 areas, which coincide with areas with high EVI in figure 4, had lower LST.

214 In the seasonality analysis for the whole time series(Figure 17) we can observe a decreasing trend
 215 for EVI(Figure 17.A), and an increasing trend for LST (Figure 17.B) in the Puerto Rico Archipelago.
 216 This constant lowering of EVI is consistent with the increasing urbanization and deforestation in

217 Puerto Rico. The increase in LST could be attributed to the current rise of temperature that the earth is
218 experiencing.

219 When comparing the different variables across the 10 years(Figure 18), we observe for EVI(Figure
220 18A) a tendency to rise during the months of April to August, although this trend was not observed
221 during the year 2017. In 2017 we can see a decrease in EVI that started by July and extended the
222 month of September, by the time that hurricane Irma and Maria passed the Puerto Rico archipelago. A
223 lower EVI implies lower photosynthetic activity, which would also be reflected by a lower exchange of
224 plant water with the atmosphere. Examining the annual trends of LST(Figure 18.B) we can observe a
225 bimodal distribution, where there is a small decrease followed by an increase in temperature between
226 the months of June and July.

227 These results are in no way conclusive and therefore require further and more granular analysis.

228 **Author Contributions:** conceptualization was brought about by all authors; methodology was devised in
229 discussion with our professors as acknowledged; software utilized was R-Statistical Software, managed
230 by Dilán-Pantojas; validation was observed by all authors; formal analysis was performed by all authors;
231 investigation, X.X.; resources, X.X.; data curationmanaged by Dilán-Pantojas and La Torre-Poueymirou; the
232 writing—original draft preparation, was performed by all authors; writing—review and editing, was performed
233 by all authors; visualization, managed by Dilán-Pantojas; supervision, was carried by Dilán-Pantojas; project
234 administration, was performed by all authors;

235 **Funding:** This research received no external funding.

236 **Acknowledgments:** We would like to acknowledge the contributions to this study made by our professors Ph.D. Jose E.
237 Garcias Arraras & Ph.D. Carla Restrepo. We would also like to thank our individual mentors Ph.D. Humberto Ortiz-Zuazaga,
238 M.S. Eileen Poueymirou-Yunque, Ph.D. Jose Luis Agosto, for all the mentorship and guidance they've provided us. We
239 would also like to thank Jose A. Reyez-Zayas for voluntarily donating some of his time in proofreading our article and script.
240 Finally, we would like thank and acknowledge our fellow classmates in this course who were very supportive and embarked
241 on a similar journey to ours' in producing this creative work.

242 **Conflicts of Interest:** The authors declare no conflict of interest.

243 Abbreviations

244 The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
AppEEARS	Application for Extracting and Exploring Analysis Ready Samples
NASA	National Aeronautics and Space Administration
LP DAAC	Land Processes Distributed Active Archive Center
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized difference vegetation index
EVI	Enhanced vegetation index
LST	Land surface temperature
KM	K-means algorith
CLARA	Cluster LARge Applications
BFAST	Breaks For Additive Season and Trend
ARIMA	Auto-Regressive Integrated Moving Average

246 Appendix A. Clustering

247 K-means clustering of EVI was performed generating 2 to 11 clusters. These cluster images would
248 be generated randomle for each pixel contained in our EVI dataset. Thus, randomizing our EVI image

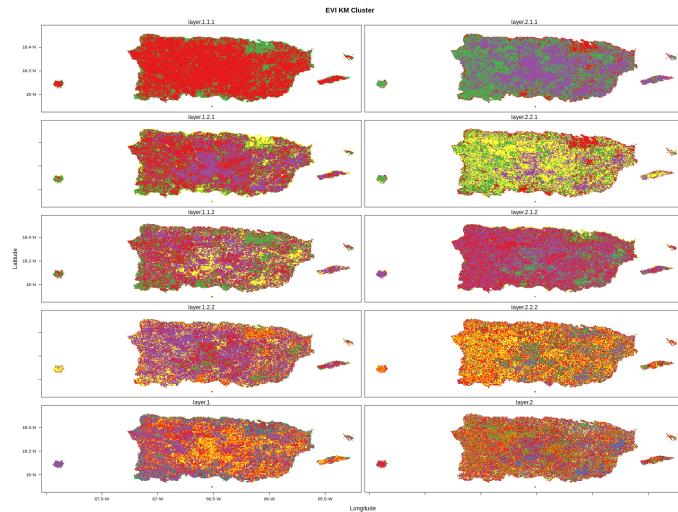


Figure A1. Results for K-means clustering of EVI from 2 to 11 clusters.

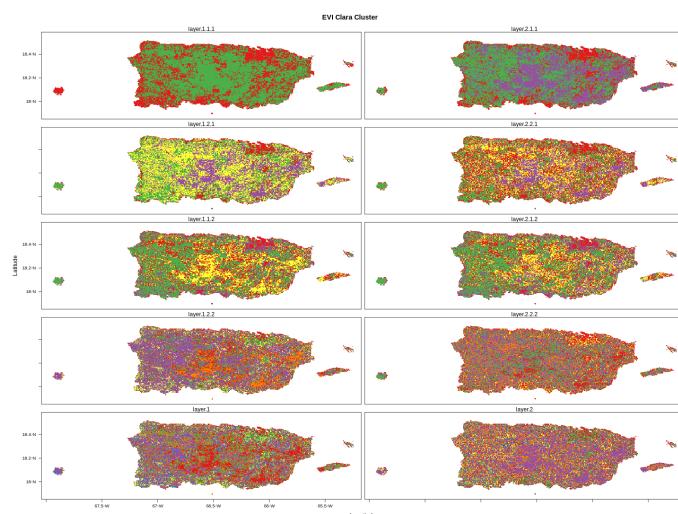


Figure A2. Results for CLARA clustering of EVI from 2 to 11 clusters

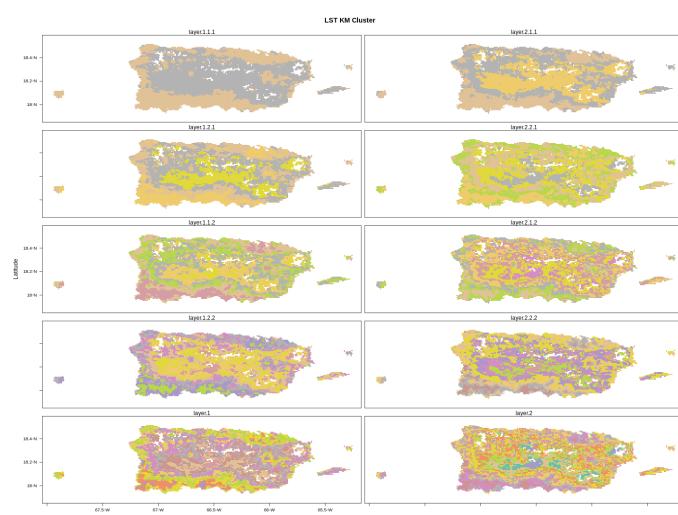
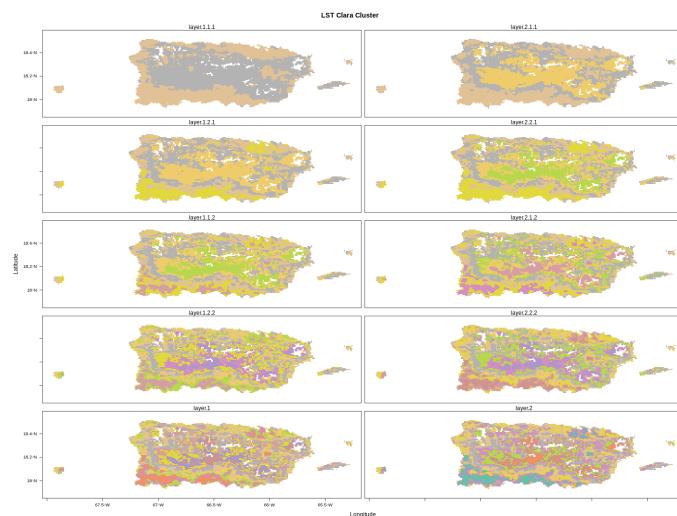


Figure A3. Results for k-means clustering of LST from 2 to 11 clusters

Table A1. Silhouette index for EVI Clusters.

Number of clusters	K-means	CLARA
2	0.5754129	0.4972576
3	0.5325603	0.4666588
4	0.5147639	0.5038890
5	0.5140073	0.4612106
6	0.5279513	0.4745793
7	0.5340362	0.4978624
8	0.5280703	0.4883521
9	0.5327400	0.4888689
10	0.5292343	0.4936729
11	0.5257909	0.4976052

**Figure A4.** Results for CLARA clustering of LST from 2 to 11 clusters.

249 Appendix B. Unsupervised clustering analysis

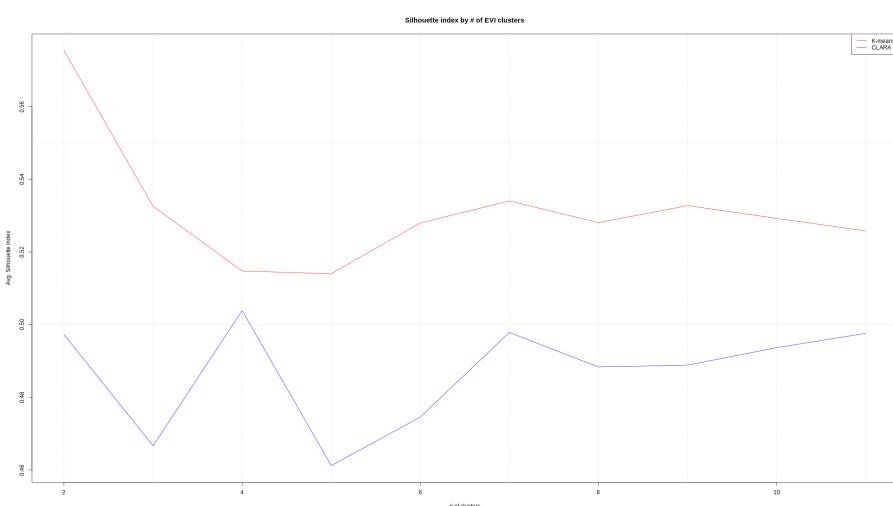
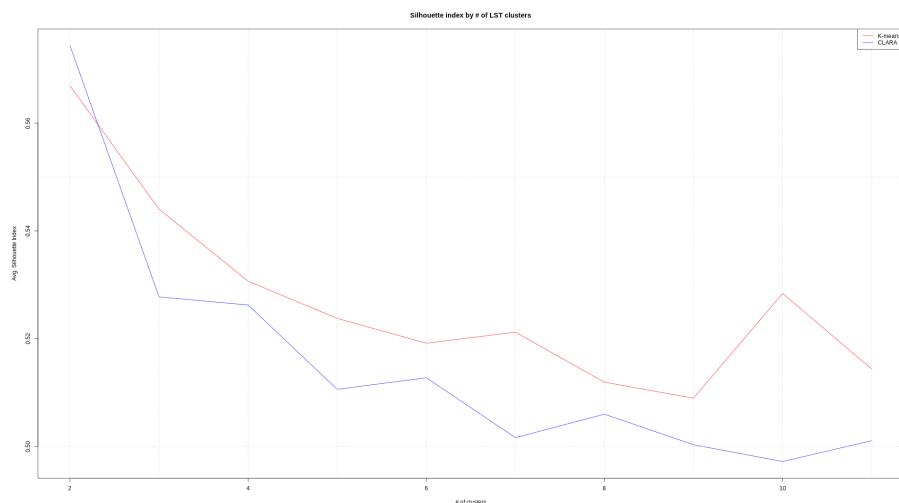
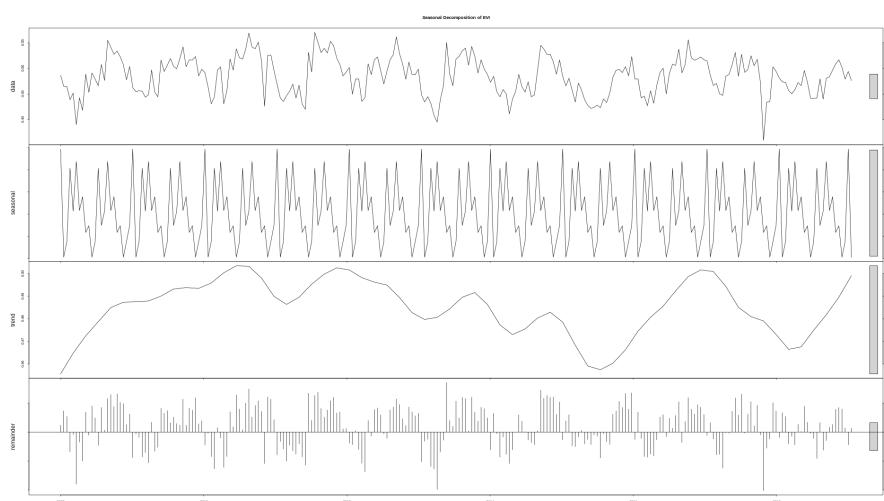
**Figure A5.** EVI clustering result comparison on Silhouette index.

Table A2. Silhouette index for LST Clusters.

Number of clusters	K-means	CLARA
2	0.5669281	0.5743742
3	0.5440286	0.5277574
4	0.5306771	0.5262508
5	0.5237698	0.5106174
6	0.5191691	0.5127648
7	0.5212166	0.5016520
8	0.5119441	0.5060061
9	0.5089707	0.5003322
10	0.5284022	0.4972170
11	0.5143890	0.5010930

**Figure A6.** LST clustering result comparison on Silhouette index.

250 Appendix C. Seasonal model evaluation

**Figure A7.** Average EVI decomposition into seasonal component and trend.

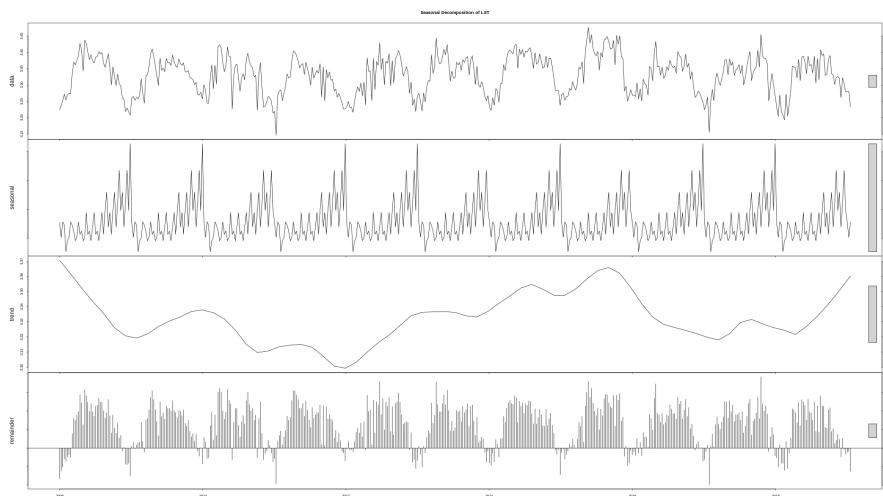


Figure A8. Average EVI decomposition into seasonal component and trend.

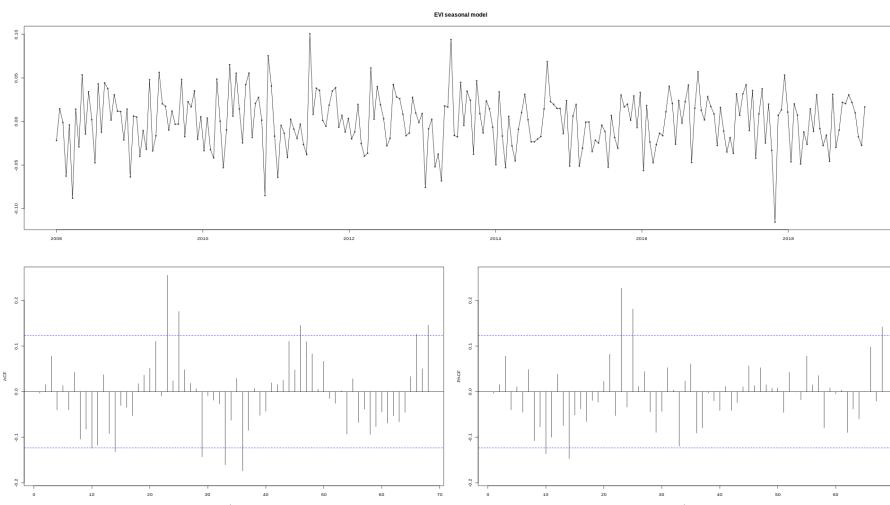


Figure A9. EVI model residuals.

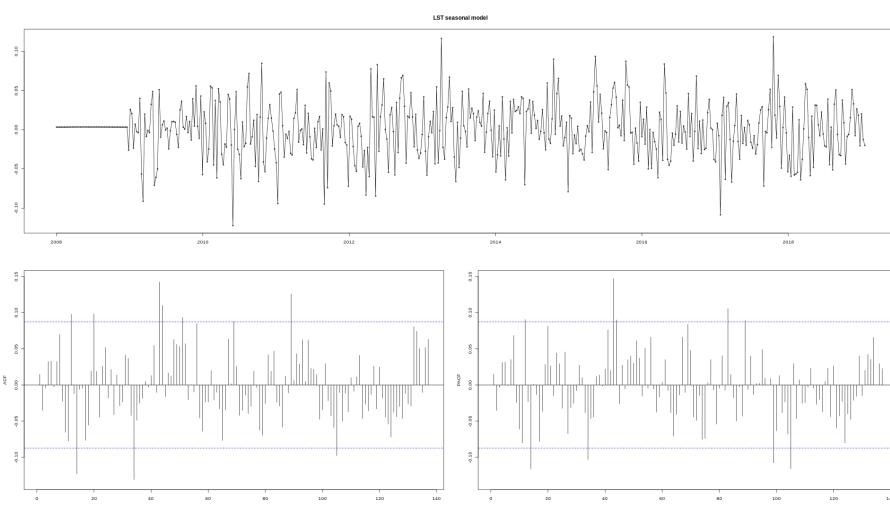


Figure A10. LST model residuals.

251 Appendix D. Code

```

252 # [plot solution]
253 fig_label <-
254   function(text,
255     region = "figure",
256     pos = "topleft",
257     cex = NULL,
258     ...) {
259   region <- match.arg(region, c("figure", "plot", "device"))
260   pos <- match.arg(
261     pos,
262     c(
263       "topleft",
264       "top",
265       "topright",
266       "left",
267       "center",
268       "right",
269       "bottomleft",
270       "bottom",
271       "bottomright"
272     )
273   )
274 }
275
276 if (region %in% c("figure", "device")) {
277   ds <- dev.size("in")
278   # xy coordinates of device corners in user coordinates
279   x <- grconvertX(c(0, ds[1]), from = "in", to = "user")
280   y <- grconvertY(c(0, ds[2]), from = "in", to = "user")
281
282   # fragment of the device we use to plot
283   if (region == "figure") {
284     # account for the fragment of the device that
285     # the figure is using
286     fig <- par("fig")
287     dx <- (x[2] - x[1])
288     dy <- (y[2] - y[1])
289     x <- x[1] + dx * fig[1:2]
290     y <- y[1] + dy * fig[3:4]
291   }
292 }
293
294 # much simpler if in plotting region
295 if (region == "plot") {
296   u <- par("usr")
297   x <- u[1:2]
298   y <- u[3:4]
299 }
300
301 sw <- strwidth(text, cex = cex) * 60 / 100
302 sh <- strheight(text, cex = cex) * 60 / 100
303
304 x1 <- switch(
305   pos,
306   topleft    = x[1] + sw,
307   left       = x[1] + sw,
308   bottomleft = x[1] + sw,
309   top        = (x[1] + x[2]) / 2,
310   center     = (x[1] + x[2]) / 2,
311   bottom     = (x[1] + x[2]) / 2,
312   topright   = x[2] - sw,
313   right      = x[2] - sw,
314   bottomright= x[2] - sw
315 )
316
317 y1 <- switch(
318   pos,
319   topleft    = y[2] - sh,
320   top        = y[2] - sh,
321   topright   = y[2] - sh,
322   left       = (y[1] + y[2]) / 2,
323   center     = (y[1] + y[2]) / 2,
324   right      = (y[1] + y[2]) / 2,
325   bottomleft = y[1] + sh,
326   bottom     = y[1] + sh,
327   bottomright= y[1] + sh
328 )
329
330 old.par <- par(xpd = NA)
331 on.exit(par(old.par))
332
333 text(x1, y1, text, cex = cex, ...)
334 return(invisible(c(x, y)))
335 }
336
337 ##########
338 # Beging Script
339 ##########
340
341 setwd("~/Hot_zones/new")
342
343 library(ncdf4) # package for netcdf manipulation
344 library(raster) # package for raster manipulation
345 library(rgdal) # package for geospatial analysis
346
347 library(lattice)
348 library(RColorBrewer)
349 library(rasterVis)

```

```

350
351 library(cluster)
352 library(clusterCrit)
353 library(ggplot2) # packages for plotting
354 library(scales)
355
356 library(tidyverse)
357 library(lubridate)
358 library(cowplot)
359
360 library(forecast)
361 library(tseries)
362 library(bfast)
363 # library(map)
364
365 # library(zoo)
366 # library(xts)
367
368 ######
369 # Load Data
370 #####
371 # Load EVI data
372 eviStack <- stack(raster(x = "MOD13Q1.006_250m_aid0001.nc"))
373 # Apparently scale factor is already applied in NCDF4
374 # eviStack$X250m.16.days.EVI <- (eviStack$X250m.16.days.EVI * 0.0001)
375
376 # Load LST data
377 lstStack <- stack(raster(x = "MOD11A2.006_1km_aid0001.nc"))
378 # Apparently scale factor is already applied in NCDF4
379 # lstStack$X8.day.daytime.1km.grid.Land.surface.Temperature <- (lstStack$X8.day.daytime.1km.grid.Land.surface.Temperature * 0.02)
380 lstStack$X8.day.daytime.1km.grid.Land.surface.Temperature <-
381 (lstStack$X8.day.daytime.1km.grid.Land.surface.Temperature - 273.15)
382
383 # -----
384 # Zonification
385 # -----
386 #####
387 # Exploratory analysis
388 #####
389 # EVI mean and histogram plots
390 png(filename = "../Figures/Exploratory_EVI.png", width = 1240)
391 par(mfrow = c(1, 2))
392 plot(
393   mean(eviStack),
394   main = "Mean EVI Spatial",
395   xlab = "Longitude",
396   ylab = "Latitude",
397   col = brewer.pal(8, "RdYlGn")
398 )
399 fig_label("A", cex = 2)
400 hist(eviStack,
401       col = "forestgreen",
402       xlab = "EVI",
403       main = "Average EVI histogram")
404 fig_label("B", cex = 2)
405 dev.off()
406
407 # LST mean and histogram plots
408 png(filename = "../Figures/Exploratory_LST.png", width = 1240)
409 par(mfrow = c(1, 2))
410 plot(
411   mean(lstStack),
412   main = "Mean LST Spatial",
413   xlab = "Longitude",
414   ylab = "Latitude",
415   col = rev(brewer.pal(8, "RdYlBu"))
416 )
417 fig_label("A", cex = 2)
418 hist(lstStack,
419       col = "firebrick",
420       xlab = "Degrees Celsius",
421       main = "Average LST histogram")
422 fig_label("B", cex = 2)
423 dev.off()
424
425 #####
426 # Individual Countour Plots
427 #####
428 par(mfrow = c(1, 1))
429 png(filename = "../Figures/Exploratory_EVI_Mean.png",
430      width = 1920,
431      height = 1080)
432 levelplot(
433   eviStack,
434   main = "Average EVI 10 year composite",
435   col.regions = colorRampPalette(brewer.pal(8, "RdYlGn")),
436   margin = list(FUN = 'mean'),
437   contour = T
438 )
439 dev.off()
440
441 png(filename = "../Figures/Exploratory_LST_Mean.png",
442      width = 1920,
443      height = 1080)
444 levelplot(
445   lstStack,
446   main = "Average LST 10 year composite",
447   col.regions = colorRampPalette(rev(brewer.pal(8, "RdYlBu"))),
448   margin = list(FUN = 'mean'),
449 )

```

```

450   contour = T
451 }
452 dev.off()
453
454 ##########
455 # EVI Zonification Clustering
456 #####
457 par(mfrow = c(1, 1))
458 eviBrick <- brick(eviStack)
459 eviDF <- values(eviBrick)
460
461 # Check NA's in the data
462 idx <- complete.cases(eviDF)
463
464 # Max Clusters
465 n <- 11
466 # Initiate the raster datasets that will hold all clustering solutions
467 # from 2 groups/clusters up to n
468 rstKM <- raster(eviBrick[[1]])
469 rstCLARA <- raster(eviBrick[[1]])
470
471 for (nClust in 2:n) {
472   cat("-> Clustering data for nClust =", nClust, ".....")
473   km <- kmeans(eviDF[idx,], centers = nClust, iter.max = 50)
474   cla <- clara(eviDF[idx,], k = nClust, metric = "manhattan")
475
476   # Create a temporary integer vector for holding cluster numbers
477   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
478   claClust <- vector(mode = "integer", length = ncell(eviBrick))
479
480   # Generate the temporary clustering vector for K-means (keeps track of NA's)
481   kmClust[!idx] <- NA
482   kmClust[idx] <- km$cluster
483
484   # Generate the temporary clustering vector for CLARA (keeps track of NA's too ;)
485   claClust[!idx] <- NA
486   claClust[idx] <- cla$clustering
487
488   # Create a temporary raster for holding the new clustering solution
489   # K-means
490   tmpRstKM <- raster(eviBrick[[1]])
491   # CLARA
492   tmpRstCLARA <- raster(eviBrick[[1]])
493
494   # Set raster values with the cluster vector
495   # K-means
496   values(tmpRstKM) <- kmClust
497   # CLARA
498   values(tmpRstCLARA) <- claClust
499
500   # Stack the temporary rasters onto the final ones
501   if (nClust == 2) {
502     rstKM <- tmpRstKM
503     rstCLARA <- tmpRstCLARA
504   } else{
505     rstKM <- stack(rstKM, tmpRstKM)
506     rstCLARA <- stack(rstCLARA, tmpRstCLARA)
507   }
508
509   cat(" done!\n\n")
510 }
511
512 png(filename = "../Figures/Clusters_EVI_KM.png",
513      width = 1920,
514      height = 1080)
515 levelplot(
516   rstKM,
517   main = "EVI KM Cluster",
518   col.regions = brewer.pal(6, "Set1"),
519   contour = F,
520   colorkey = FALSE
521 )
522 dev.off()
523 png(filename = "../Figures/Clusters_EVI_Clara.png",
524      width = 1920,
525      height = 1080)
526 levelplot(
527   rstCLARA,
528   main = "EVI Clara Cluster",
529   col.regions = brewer.pal(6, "Set1"),
530   contour = F,
531   colorkey = FALSE
532 )
533 dev.off()
534
535 ##########
536 # EVI Zonification Clustering Analysis
537 #####
538 # Start a data frame that will store all silhouette values
539 # for k-means and CLARA
540 clustPerfSI <- data.frame(nClust = 2:n,
541                           SI_KM = NA,
542                           SI_CLARA = NA)
543
544 for (i in 1:nlayers(rstKM)) {
545   # Iterate through each layer
546
547   cat("-> Evaluating clustering performance for nClust =",
548       (2:n)[i],
549       ".....")
550
551   rstKM <- stack(rstKM, tmpRstKM)
552   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
553
554   km <- kmeans(eviDF[idx,], centers = i)
555   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
556
557   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
558   claClust <- vector(mode = "integer", length = ncell(eviBrick))
559
560   kmClust[!idx] <- NA
561   kmClust[idx] <- km$cluster
562
563   claClust[!idx] <- NA
564   claClust[idx] <- cla$clustering
565
566   values(rstKM) <- kmClust
567   values(rstCLARA) <- claClust
568
569   rstKM <- stack(rstKM, tmpRstKM)
570   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
571
572   km <- kmeans(eviDF[idx,], centers = i)
573   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
574
575   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
576   claClust <- vector(mode = "integer", length = ncell(eviBrick))
577
578   kmClust[!idx] <- NA
579   kmClust[idx] <- km$cluster
580
581   claClust[!idx] <- NA
582   claClust[idx] <- cla$clustering
583
584   values(rstKM) <- kmClust
585   values(rstCLARA) <- claClust
586
587   rstKM <- stack(rstKM, tmpRstKM)
588   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
589
590   km <- kmeans(eviDF[idx,], centers = i)
591   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
592
593   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
594   claClust <- vector(mode = "integer", length = ncell(eviBrick))
595
596   kmClust[!idx] <- NA
597   kmClust[idx] <- km$cluster
598
599   claClust[!idx] <- NA
600   claClust[idx] <- cla$clustering
601
602   values(rstKM) <- kmClust
603   values(rstCLARA) <- claClust
604
605   rstKM <- stack(rstKM, tmpRstKM)
606   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
607
608   km <- kmeans(eviDF[idx,], centers = i)
609   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
610
611   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
612   claClust <- vector(mode = "integer", length = ncell(eviBrick))
613
614   kmClust[!idx] <- NA
615   kmClust[idx] <- km$cluster
616
617   claClust[!idx] <- NA
618   claClust[idx] <- cla$clustering
619
620   values(rstKM) <- kmClust
621   values(rstCLARA) <- claClust
622
623   rstKM <- stack(rstKM, tmpRstKM)
624   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
625
626   km <- kmeans(eviDF[idx,], centers = i)
627   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
628
629   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
630   claClust <- vector(mode = "integer", length = ncell(eviBrick))
631
632   kmClust[!idx] <- NA
633   kmClust[idx] <- km$cluster
634
635   claClust[!idx] <- NA
636   claClust[idx] <- cla$clustering
637
638   values(rstKM) <- kmClust
639   values(rstCLARA) <- claClust
640
641   rstKM <- stack(rstKM, tmpRstKM)
642   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
643
644   km <- kmeans(eviDF[idx,], centers = i)
645   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
646
647   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
648   claClust <- vector(mode = "integer", length = ncell(eviBrick))
649
650   kmClust[!idx] <- NA
651   kmClust[idx] <- km$cluster
652
653   claClust[!idx] <- NA
654   claClust[idx] <- cla$clustering
655
656   values(rstKM) <- kmClust
657   values(rstCLARA) <- claClust
658
659   rstKM <- stack(rstKM, tmpRstKM)
660   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
661
662   km <- kmeans(eviDF[idx,], centers = i)
663   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
664
665   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
666   claClust <- vector(mode = "integer", length = ncell(eviBrick))
667
668   kmClust[!idx] <- NA
669   kmClust[idx] <- km$cluster
670
671   claClust[!idx] <- NA
672   claClust[idx] <- cla$clustering
673
674   values(rstKM) <- kmClust
675   values(rstCLARA) <- claClust
676
677   rstKM <- stack(rstKM, tmpRstKM)
678   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
679
680   km <- kmeans(eviDF[idx,], centers = i)
681   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
682
683   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
684   claClust <- vector(mode = "integer", length = ncell(eviBrick))
685
686   kmClust[!idx] <- NA
687   kmClust[idx] <- km$cluster
688
689   claClust[!idx] <- NA
690   claClust[idx] <- cla$clustering
691
692   values(rstKM) <- kmClust
693   values(rstCLARA) <- claClust
694
695   rstKM <- stack(rstKM, tmpRstKM)
696   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
697
698   km <- kmeans(eviDF[idx,], centers = i)
699   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
700
701   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
702   claClust <- vector(mode = "integer", length = ncell(eviBrick))
703
704   kmClust[!idx] <- NA
705   kmClust[idx] <- km$cluster
706
707   claClust[!idx] <- NA
708   claClust[idx] <- cla$clustering
709
710   values(rstKM) <- kmClust
711   values(rstCLARA) <- claClust
712
713   rstKM <- stack(rstKM, tmpRstKM)
714   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
715
716   km <- kmeans(eviDF[idx,], centers = i)
717   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
718
719   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
720   claClust <- vector(mode = "integer", length = ncell(eviBrick))
721
722   kmClust[!idx] <- NA
723   kmClust[idx] <- km$cluster
724
725   claClust[!idx] <- NA
726   claClust[idx] <- cla$clustering
727
728   values(rstKM) <- kmClust
729   values(rstCLARA) <- claClust
730
731   rstKM <- stack(rstKM, tmpRstKM)
732   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
733
734   km <- kmeans(eviDF[idx,], centers = i)
735   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
736
737   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
738   claClust <- vector(mode = "integer", length = ncell(eviBrick))
739
740   kmClust[!idx] <- NA
741   kmClust[idx] <- km$cluster
742
743   claClust[!idx] <- NA
744   claClust[idx] <- cla$clustering
745
746   values(rstKM) <- kmClust
747   values(rstCLARA) <- claClust
748
749   rstKM <- stack(rstKM, tmpRstKM)
750   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
751
752   km <- kmeans(eviDF[idx,], centers = i)
753   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
754
755   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
756   claClust <- vector(mode = "integer", length = ncell(eviBrick))
757
758   kmClust[!idx] <- NA
759   kmClust[idx] <- km$cluster
760
761   claClust[!idx] <- NA
762   claClust[idx] <- cla$clustering
763
764   values(rstKM) <- kmClust
765   values(rstCLARA) <- claClust
766
767   rstKM <- stack(rstKM, tmpRstKM)
768   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
769
770   km <- kmeans(eviDF[idx,], centers = i)
771   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
772
773   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
774   claClust <- vector(mode = "integer", length = ncell(eviBrick))
775
776   kmClust[!idx] <- NA
777   kmClust[idx] <- km$cluster
778
779   claClust[!idx] <- NA
780   claClust[idx] <- cla$clustering
781
782   values(rstKM) <- kmClust
783   values(rstCLARA) <- claClust
784
785   rstKM <- stack(rstKM, tmpRstKM)
786   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
787
788   km <- kmeans(eviDF[idx,], centers = i)
789   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
790
791   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
792   claClust <- vector(mode = "integer", length = ncell(eviBrick))
793
794   kmClust[!idx] <- NA
795   kmClust[idx] <- km$cluster
796
797   claClust[!idx] <- NA
798   claClust[idx] <- cla$clustering
799
800   values(rstKM) <- kmClust
801   values(rstCLARA) <- claClust
802
803   rstKM <- stack(rstKM, tmpRstKM)
804   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
805
806   km <- kmeans(eviDF[idx,], centers = i)
807   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
808
809   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
810   claClust <- vector(mode = "integer", length = ncell(eviBrick))
811
812   kmClust[!idx] <- NA
813   kmClust[idx] <- km$cluster
814
815   claClust[!idx] <- NA
816   claClust[idx] <- cla$clustering
817
818   values(rstKM) <- kmClust
819   values(rstCLARA) <- claClust
820
821   rstKM <- stack(rstKM, tmpRstKM)
822   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
823
824   km <- kmeans(eviDF[idx,], centers = i)
825   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
826
827   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
828   claClust <- vector(mode = "integer", length = ncell(eviBrick))
829
830   kmClust[!idx] <- NA
831   kmClust[idx] <- km$cluster
832
833   claClust[!idx] <- NA
834   claClust[idx] <- cla$clustering
835
836   values(rstKM) <- kmClust
837   values(rstCLARA) <- claClust
838
839   rstKM <- stack(rstKM, tmpRstKM)
840   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
841
842   km <- kmeans(eviDF[idx,], centers = i)
843   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
844
845   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
846   claClust <- vector(mode = "integer", length = ncell(eviBrick))
847
848   kmClust[!idx] <- NA
849   kmClust[idx] <- km$cluster
850
851   claClust[!idx] <- NA
852   claClust[idx] <- cla$clustering
853
854   values(rstKM) <- kmClust
855   values(rstCLARA) <- claClust
856
857   rstKM <- stack(rstKM, tmpRstKM)
858   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
859
860   km <- kmeans(eviDF[idx,], centers = i)
861   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
862
863   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
864   claClust <- vector(mode = "integer", length = ncell(eviBrick))
865
866   kmClust[!idx] <- NA
867   kmClust[idx] <- km$cluster
868
869   claClust[!idx] <- NA
870   claClust[idx] <- cla$clustering
871
872   values(rstKM) <- kmClust
873   values(rstCLARA) <- claClust
874
875   rstKM <- stack(rstKM, tmpRstKM)
876   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
877
878   km <- kmeans(eviDF[idx,], centers = i)
879   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
880
881   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
882   claClust <- vector(mode = "integer", length = ncell(eviBrick))
883
884   kmClust[!idx] <- NA
885   kmClust[idx] <- km$cluster
886
887   claClust[!idx] <- NA
888   claClust[idx] <- cla$clustering
889
890   values(rstKM) <- kmClust
891   values(rstCLARA) <- claClust
892
893   rstKM <- stack(rstKM, tmpRstKM)
894   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
895
896   km <- kmeans(eviDF[idx,], centers = i)
897   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
898
899   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
900   claClust <- vector(mode = "integer", length = ncell(eviBrick))
901
902   kmClust[!idx] <- NA
903   kmClust[idx] <- km$cluster
904
905   claClust[!idx] <- NA
906   claClust[idx] <- cla$clustering
907
908   values(rstKM) <- kmClust
909   values(rstCLARA) <- claClust
910
911   rstKM <- stack(rstKM, tmpRstKM)
912   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
913
914   km <- kmeans(eviDF[idx,], centers = i)
915   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
916
917   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
918   claClust <- vector(mode = "integer", length = ncell(eviBrick))
919
920   kmClust[!idx] <- NA
921   kmClust[idx] <- km$cluster
922
923   claClust[!idx] <- NA
924   claClust[idx] <- cla$clustering
925
926   values(rstKM) <- kmClust
927   values(rstCLARA) <- claClust
928
929   rstKM <- stack(rstKM, tmpRstKM)
930   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
931
932   km <- kmeans(eviDF[idx,], centers = i)
933   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
934
935   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
936   claClust <- vector(mode = "integer", length = ncell(eviBrick))
937
938   kmClust[!idx] <- NA
939   kmClust[idx] <- km$cluster
940
941   claClust[!idx] <- NA
942   claClust[idx] <- cla$clustering
943
944   values(rstKM) <- kmClust
945   values(rstCLARA) <- claClust
946
947   rstKM <- stack(rstKM, tmpRstKM)
948   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
949
950   km <- kmeans(eviDF[idx,], centers = i)
951   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
952
953   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
954   claClust <- vector(mode = "integer", length = ncell(eviBrick))
955
956   kmClust[!idx] <- NA
957   kmClust[idx] <- km$cluster
958
959   claClust[!idx] <- NA
960   claClust[idx] <- cla$clustering
961
962   values(rstKM) <- kmClust
963   values(rstCLARA) <- claClust
964
965   rstKM <- stack(rstKM, tmpRstKM)
966   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
967
968   km <- kmeans(eviDF[idx,], centers = i)
969   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
970
971   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
972   claClust <- vector(mode = "integer", length = ncell(eviBrick))
973
974   kmClust[!idx] <- NA
975   kmClust[idx] <- km$cluster
976
977   claClust[!idx] <- NA
978   claClust[idx] <- cla$clustering
979
980   values(rstKM) <- kmClust
981   values(rstCLARA) <- claClust
982
983   rstKM <- stack(rstKM, tmpRstKM)
984   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
985
986   km <- kmeans(eviDF[idx,], centers = i)
987   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
988
989   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
990   claClust <- vector(mode = "integer", length = ncell(eviBrick))
991
992   kmClust[!idx] <- NA
993   kmClust[idx] <- km$cluster
994
995   claClust[!idx] <- NA
996   claClust[idx] <- cla$clustering
997
998   values(rstKM) <- kmClust
999   values(rstCLARA) <- claClust
1000
1001   rstKM <- stack(rstKM, tmpRstKM)
1002   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1003
1004   km <- kmeans(eviDF[idx,], centers = i)
1005   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1006
1007   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1008   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1009
1010   kmClust[!idx] <- NA
1011   kmClust[idx] <- km$cluster
1012
1013   claClust[!idx] <- NA
1014   claClust[idx] <- cla$clustering
1015
1016   values(rstKM) <- kmClust
1017   values(rstCLARA) <- claClust
1018
1019   rstKM <- stack(rstKM, tmpRstKM)
1020   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1021
1022   km <- kmeans(eviDF[idx,], centers = i)
1023   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1024
1025   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1026   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1027
1028   kmClust[!idx] <- NA
1029   kmClust[idx] <- km$cluster
1030
1031   claClust[!idx] <- NA
1032   claClust[idx] <- cla$clustering
1033
1034   values(rstKM) <- kmClust
1035   values(rstCLARA) <- claClust
1036
1037   rstKM <- stack(rstKM, tmpRstKM)
1038   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1039
1040   km <- kmeans(eviDF[idx,], centers = i)
1041   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1042
1043   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1044   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1045
1046   kmClust[!idx] <- NA
1047   kmClust[idx] <- km$cluster
1048
1049   claClust[!idx] <- NA
1050   claClust[idx] <- cla$clustering
1051
1052   values(rstKM) <- kmClust
1053   values(rstCLARA) <- claClust
1054
1055   rstKM <- stack(rstKM, tmpRstKM)
1056   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1057
1058   km <- kmeans(eviDF[idx,], centers = i)
1059   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1060
1061   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1062   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1063
1064   kmClust[!idx] <- NA
1065   kmClust[idx] <- km$cluster
1066
1067   claClust[!idx] <- NA
1068   claClust[idx] <- cla$clustering
1069
1070   values(rstKM) <- kmClust
1071   values(rstCLARA) <- claClust
1072
1073   rstKM <- stack(rstKM, tmpRstKM)
1074   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1075
1076   km <- kmeans(eviDF[idx,], centers = i)
1077   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1078
1079   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1080   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1081
1082   kmClust[!idx] <- NA
1083   kmClust[idx] <- km$cluster
1084
1085   claClust[!idx] <- NA
1086   claClust[idx] <- cla$clustering
1087
1088   values(rstKM) <- kmClust
1089   values(rstCLARA) <- claClust
1090
1091   rstKM <- stack(rstKM, tmpRstKM)
1092   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1093
1094   km <- kmeans(eviDF[idx,], centers = i)
1095   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1096
1097   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1098   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1099
1100   kmClust[!idx] <- NA
1101   kmClust[idx] <- km$cluster
1102
1103   claClust[!idx] <- NA
1104   claClust[idx] <- cla$clustering
1105
1106   values(rstKM) <- kmClust
1107   values(rstCLARA) <- claClust
1108
1109   rstKM <- stack(rstKM, tmpRstKM)
1110   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1111
1112   km <- kmeans(eviDF[idx,], centers = i)
1113   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1114
1115   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1116   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1117
1118   kmClust[!idx] <- NA
1119   kmClust[idx] <- km$cluster
1120
1121   claClust[!idx] <- NA
1122   claClust[idx] <- cla$clustering
1123
1124   values(rstKM) <- kmClust
1125   values(rstCLARA) <- claClust
1126
1127   rstKM <- stack(rstKM, tmpRstKM)
1128   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1129
1130   km <- kmeans(eviDF[idx,], centers = i)
1131   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1132
1133   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1134   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1135
1136   kmClust[!idx] <- NA
1137   kmClust[idx] <- km$cluster
1138
1139   claClust[!idx] <- NA
1140   claClust[idx] <- cla$clustering
1141
1142   values(rstKM) <- kmClust
1143   values(rstCLARA) <- claClust
1144
1145   rstKM <- stack(rstKM, tmpRstKM)
1146   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1147
1148   km <- kmeans(eviDF[idx,], centers = i)
1149   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1150
1151   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1152   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1153
1154   kmClust[!idx] <- NA
1155   kmClust[idx] <- km$cluster
1156
1157   claClust[!idx] <- NA
1158   claClust[idx] <- cla$clustering
1159
1160   values(rstKM) <- kmClust
1161   values(rstCLARA) <- claClust
1162
1163   rstKM <- stack(rstKM, tmpRstKM)
1164   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1165
1166   km <- kmeans(eviDF[idx,], centers = i)
1167   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1168
1169   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1170   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1171
1172   kmClust[!idx] <- NA
1173   kmClust[idx] <- km$cluster
1174
1175   claClust[!idx] <- NA
1176   claClust[idx] <- cla$clustering
1177
1178   values(rstKM) <- kmClust
1179   values(rstCLARA) <- claClust
1180
1181   rstKM <- stack(rstKM, tmpRstKM)
1182   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1183
1184   km <- kmeans(eviDF[idx,], centers = i)
1185   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1186
1187   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1188   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1189
1190   kmClust[!idx] <- NA
1191   kmClust[idx] <- km$cluster
1192
1193   claClust[!idx] <- NA
1194   claClust[idx] <- cla$clustering
1195
1196   values(rstKM) <- kmClust
1197   values(rstCLARA) <- claClust
1198
1199   rstKM <- stack(rstKM, tmpRstKM)
1200   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1201
1202   km <- kmeans(eviDF[idx,], centers = i)
1203   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1204
1205   kmClust <- vector(mode = "integer", length = ncell(eviBrick))
1206   claClust <- vector(mode = "integer", length = ncell(eviBrick))
1207
1208   kmClust[!idx] <- NA
1209   kmClust[idx] <- km$cluster
1210
1211   claClust[!idx] <- NA
1212   claClust[idx] <- cla$clustering
1213
1214   values(rstKM) <- kmClust
1215   values(rstCLARA) <- claClust
1216
1217   rstKM <- stack(rstKM, tmpRstKM)
1218   rstCLARA <- stack(rstCLARA, tmpRstCLARA)
1219
1220   km <- kmeans(eviDF[idx,], centers = i)
1221   cla <- clara(eviDF[idx,], k = i, metric = "manhattan")
1222
12
```

```

550
551 # Extract random cell samples stratified by cluster
552 cellIdx_RstKM <- sampleStratified(rstKM[[i]], size = 2000)
553 cellIdx_rstCLARA <- sampleStratified(rstCLARA[[i]], size = 2000)
554
555 # Get cell values from the Stratified Random Sample from the raster
556 # data frame object (rstDF)
557 rstDFStRS_KM <- eviDF[cellIdx_RstKM[, 1], ]
558 rstDFStRS_CLARA <- eviDF[cellIdx_rstCLARA[, 1], ]
559
560 # Make sure all columns are numeric (intCriteria function is picky on this)
561 rstDFStRS_KM[] <- sapply(rstDFStRS_KM, as.numeric)
562 rstDFStRS_CLARA[] <- sapply(rstDFStRS_CLARA, as.numeric)
563
564 # Compute the sample-based Silhouette index for:
565 # K-means
566 clCritKM <- intCriteria(
567   traj = as.matrix(rstDFStRS_KM),
568   part = as.integer(cellIdx_RstKM[, 2]),
569   crit = "Silhouette"
570 )
571 # and CLARA
572 clCritCLARA <- intCriteria(
573   traj = as.matrix(rstDFStRS_CLARA),
574   part = as.integer(cellIdx_rstCLARA[, 2]),
575   crit = "Silhouette"
576 )
577
578 # Write the silhouette index value to clustPerfSI data frame holding
579 # all results
580 clustPerfSI[i, "SI_KM"] <- clCritKM[[1]][1]
581 clustPerfSI[i, "SI_CLARA"] <- clCritCLARA[[1]][1]
582
583 cat(" done!\n\n")
584
585 }
586
587 EVI_CLUST <- as.data.frame(
588   clustPerfSI,
589   digits = 3,
590   align = "c",
591   col.names = c(
592     "#clusters",
593     "Avg. Silhouette (k-means)",
594     "Avg. Silhouette (CLARA)"
595   )
596 )
597
598 View(EVI_CLUST)
599
600 # Sublime text 3 regex to format tabular output
# \d+\s(\d+)\n(\d\.\d+)\n(\d\.\d+)
# \1 & \2 & \3 \\\\
601
602
603
604 png(filename = "../Figures/Cluster_EVI_SI.png",
605   width = 1920,
606   height = 1080)
607 plot(
608   clustPerfSI[, 1],
609   clustPerfSI[, 2],
610   xlim = c(2, n),
611   ylim = range(clustPerfSI[, 2:3]),
612   type = "n",
613   ylab = "Avg. Silhouette Index",
614   xlab = "# of clusters",
615   main = "Silhouette index by # of EVI clusters"
616 )
617
618 # Plot Avg Silhouette values across # of clusters for K-means
619 lines(clustPerfSI[, 1], clustPerfSI[, 2], col = "red")
620 # Plot Avg Silhouette values across # of clusters for CLARA
621 lines(clustPerfSI[, 1], clustPerfSI[, 3], col = "blue")
622
623 # Grid lines
624 abline(v = 2:n, lty = 2, col = "light grey")
625 abline(h = seq(0.20, 1.0, 0.05),
626         lty = 2,
627         col = "light grey")
628
629 legend(
630   "topright",
631   legend = c("K-means", "CLARA"),
632   col = c("red", "blue"),
633   lty = 1,
634   lwd = 1
635 )
636 dev.off()
637
638 png(filename = "../Figures/Best_Cluster_EVI.png",
639   width = 1920,
640   height = 1080)
641 levelplot(
642   rstKM,
643   layer = 1,
644   main = "EVI KM best cluster",
645   col.regions = colorRampPalette(brewer.pal(3, "RdYlGn")),
646   contour = T,
647   colorkey = FALSE
648 )
649 dev.off()

```

```

650 #####  

651 # LST Zonification Clustering  

652 #####  

653 #####  

654 lstBrick <- brick(lstStack)  

655 lstDF <- values(lstBrick)  

656  

657 # Check NA's in the data  

658 idx <- complete.cases(lstDF)  

659  

660 # Max Clusters  

661 n <- 11  

662 # Initiate the raster datasets that will hold all clustering solutions  

663 # from 2 groups/clusters up to n  

664 rstKM <- raster(lstBrick[[1]])  

665 rstCLARA <- raster(lstBrick[[1]])  

666  

667 for (nClust in 2:n) {  

668   cat("-> Clustering data for nClust =", nClust, ".....")  

669   km <- kmeans(lstDF[idx,], centers = nClust, iter.max = 50)  

670   cla <- clara(lstDF[idx,], k = nClust, metric = "manhattan")  

671  

672   # Create a temporary integer vector for holding cluster numbers  

673   kmClust <- vector(mode = "integer", length = ncell(lstBrick))  

674   claClust <- vector(mode = "integer", length = ncell(lstBrick))  

675  

676   # Generate the temporary clustering vector for K-means (keeps track of NA's)  

677   kmClust[!idx] <- NA  

678   kmClust[idx] <- km$cluster  

679  

680   # Generate the temporary clustering vector for CLARA (keeps track of NA's too ;-)  

681   claClust[!idx] <- NA  

682   claClust[idx] <- cla$clustering  

683  

684   # Create a temporary raster for holding the new clustering solution  

685   # K-means  

686   tmpRstKM <- raster(lstBrick[[1]])  

687   # CLARA  

688   tmpRstCLARA <- raster(lstBrick[[1]])  

689  

690   # Set raster values with the cluster vector  

691   # K-means  

692   values(tmpRstKM) <- kmClust  

693   # CLARA  

694   values(tmpRstCLARA) <- claClust  

695  

696   # Stack the temporary rasters onto the final ones  

697   if (nClust == 2) {  

698     rstKM <- tmpRstKM  

699     rstCLARA <- tmpRstCLARA  

700   } else{  

701     rstKM <- stack(rstKM, tmpRstKM)  

702     rstCLARA <- stack(rstCLARA, tmpRstCLARA)  

703   }  

704  

705   cat(" done!\n\n")  

706 }  

707  

708 png(filename = "../Figures/Clusters_LST_KM.png",  

709   width = 1920,  

710   height = 1080)  

711 levelplot(  

712   rstKM,  

713   main = "LST KM Cluster",  

714   col.regions = colorRampPalette(rev(brewer.pal(8, "Set2"))),  

715   contour = F,  

716   colorkey = FALSE  

717 )  

718 dev.off()  

719  

720 png(filename = "../Figures/Clusters_LST_Clara.png",  

721   width = 1920,  

722   height = 1080)  

723 levelplot(  

724   rstCLARA,  

725   main = "LST Clara Cluster",  

726   col.regions = colorRampPalette(rev(brewer.pal(8, "Set2"))),  

727   contour = F,  

728   colorkey = FALSE  

729 )  

730 dev.off()  

731  

732 #####  

733 # LST Zonification Clustering Analysis  

734 #####  

735 # Start a data frame that will store all silhouette values  

736 # for k-means and CLARA  

737 clustPerfSI <- data.frame(nClust = 2:n,  

738   SI_KM = NA,  

739   SI_CLARA = NA)  

740  

741 for (i in 1:nlayers(rstKM)) {  

742   # Iterate through each layer  

743  

744   cat("-> Evaluating clustering performance for nClust =",  

745   (2:n)[i],  

746   ".....")  

747  

748   # Extract random cell samples stratified by cluster  

749   cellIdx_RstKM <- sampleStratified(rstKM[[i]], size = 2000)

```

```

750 cellIdx_rstCLARA <- sampleStratified(rstCLARA[[i]], size = 2000)          498
751
752 # Get cell values from the Stratified Random Sample from the raster      499
753 # data frame object (rstDF)
754 rstDFStRS_KM <- lstDF[cellIdx_RstKM[, 1], ]                            500
755 rstDFStRS_CLARA <- lstDF[cellIdx_rstCLARA[, 1], ]                         501
756
757 # Make sure all columns are numeric (intCriteria function is picky on this) 502
758 rstDFStRS_KM[] <- sapply(rstDFStRS_KM, as.numeric)                         503
759 rstDFStRS_CLARA[] <- sapply(rstDFStRS_CLARA, as.numeric)                      504
760
761 # Compute the sample-based Silhouette index for:                           505
762 #
763 # K-means
764 clCritKM <- intCriteria(                                              506
765   traj = as.matrix(rstDFStRS_KM),                                         507
766   part = as.integer(cellIdx_RstKM[, 2]),                                     508
767   crit = "Silhouette"                                                       509
768 )
769 # and CLARA
770 clCritCLARA <- intCriteria(                                              510
771   traj = as.matrix(rstDFStRS_CLARA),                                         511
772   part = as.integer(cellIdx_rstCLARA[, 2]),                                     512
773   crit = "Silhouette"                                                       513
774 )
775
776 # Write the silhouette index value to clustPerfSI data frame holding      514
777 # all results
778 clustPerfSI[i, "SI_KM"] <- clCritKM[[1]][1]                                515
779 clustPerfSI[i, "SI_CLARA"] <- clCritCLARA[[1]][1]                          516
780
781 cat(" done!\n\n")
782
783 }
784
785 LST_CLUST <- as.data.frame(                                              517
786   clustPerfSI,
787   digits = 3,                                                               518
788   align = "c",                                                               519
789   col.names = c(                520
790     "#clusters",                                                 521
791     "Avg. Silhouette (k-means)",                                         522
792     "Avg. Silhouette (CLARA)"                                              523
793   )                                                               524
794 )
795
796 View(LST_CLUST)
797
798 png(filename = "../Figures/Clusters_LST_SIL.png",
799       width = 1920,                                                       525
800       height = 1080)                                                       526
801 plot(
802   clustPerfSI[, 1],                                         527
803   clustPerfSI[, 2],                                         528
804   xlim = c(2, n),                                           529
805   ylim = range(clustPerfSI[, 2:3]),                           530
806   type = "l",                                                 531
807   ylab = "Avg. Silhouette Index",                           532
808   xlab = "# of clusters",                                    533
809   main = "Silhouette index by # of LST clusters"           534
810 )
811
812 # Plot Avg Silhouette values across # of clusters for K-means          535
813 lines(clustPerfSI[, 1], clustPerfSI[, 2], col = "red")                   536
814 # Plot Avg Silhouette values across # of clusters for CLARA             537
815 lines(clustPerfSI[, 1], clustPerfSI[, 3], col = "blue")                  538
816
817 # Grid lines
818 abline(v = 2:n, lty = 2, col = "light grey")                           539
819 abline(h = seq(0.20, 1.0, 0.05),                                         540
820   lty = 2,                                                               541
821   col = "light grey")                                                    542
822
823 legend(                                              543
824   "topright",                                              544
825   legend = c("K-means", "CLARA"),                                         545
826   col = c("red", "blue"),                                              546
827   lty = 1,                                                 547
828   lwd = 1                                                 548
829 )
830 dev.off()
831
832 png(filename = "../Figures/Best_Clusters_LST.png",
833       width = 1920,                                                       549
834       height = 1080)                                                       550
835 levelplot(                                              551
836   rstCLARA,                                               552
837   main = "LST CLARA best cluster",                               553
838   layer = 1,                                                 554
839   col.regions = colorRampPalette(rev(brewer.pal(3, "RdYlBu"))), 555
840   contour = T,                                                556
841   colorkey = FALSE                                         557
842 )
843 dev.off()
844
845 ##########
846 ## ZONING END
847 ##########
848 ##########
849 ##########

```

```

850 ## SEASONALITY BEGIN
851 ##########
852 #
853 # -----
854 # Seasonality
855 #
856 # -----
857 ##########
858 # Exploratory Seasonality
859 ##########
860 #
861 # Load EVI seasonality data
862 #
863 evi_stat <-
864   read.csv("MOD13Q1-006-Statistics.csv", stringsAsFactors = FALSE)[-1,]
865 evi_stat$date <-
866   as.Date(evi_stat$date, format = "%Y-%m-%d", deltat = 16 / 365)
867
868 #
869 # Load LST Seasonality data
870 #
871 lst_stat <-
872   read.csv("MOD11A2-006-Statistics.csv", stringsAsFactors = FALSE)[-1,]
873
874 lst_stat$Mean <- lst_stat$Mean - 273.15
875 lst_stat$date <-
876   as.Date(lst_stat$date, format = "%Y-%m-%d", deltat = 8 / 365)
877
878
879 png(filename = "../Figures/Exploratory_Seasonality.png",
880   width = 1920,
881   height = 1080)
882 par(mfrow = c(1, 2))
883 plot(
884   x = evi_stat$date,
885   y = evi_stat$Mean,
886   main = "Mean EVI Temporal",
887   xlab = "Time",
888   ylab = "EVI"
889 )
890 lines(
891   smooth.spline(
892     x = evi_stat$date,
893     y = evi_stat$Mean,
894     cv = T
895   ),
896   lwd = 2,
897   col = "green")
898 )
899 abline(
900   reg = lm(evi_stat$Mean ~ evi_stat$date),
901   col = "grey",
902   lw = 5
903 )
904 fig_label("A", cex = 2)
905 plot(
906   x = lst_stat$date,
907   y = lst_stat$Mean,
908   main = "Mean LST Temporal",
909   xlab = "Time",
910   ylab = "Degrees Celsius"
911 )
912 lines(
913   smooth.spline(
914     x = lst_stat$date,
915     y = lst_stat$Mean,
916     cv = T
917   ),
918   lwd = 2,
919   col = "red")
920 )
921 abline(
922   reg = lm(lst_stat$Mean ~ lst_stat$date),
923   col = "grey",
924   lw = 5
925 )
926 fig_label("B", cex = 2)
927 dev.off()
928
929 #####
930 # EVI Seasonality Yearly
931 #####
932 # Parse the dates, and use lower case names
933 df <- as_tibble(evi_stat) %>%
934   rename_all(tolower) %>%
935   mutate(date = ymd(date))
936
937 # Define the plot
938 p <- df %>%
939   mutate(year = factor(year(date)), # use year to define separate curves
940         date = update(date, year = 1)) %>% # use a constant year for the x-axis) %>%
941   ggplot(aes(date, mean, color = year)) +
942   scale_x_date(date_breaks = "1 month", date_labels = "%b")
943
944 # Raw montly data
945 p + geom_line()
946
947 # Hone in on a single year
948 p + geom_line(aes(group = year), color = "black", alpha = 0.1) +
949   # geom_line(data = function(x) filter(x, year == 2010), size = 1)

```

```

950 # Smoothed version
951 p1 <- 698
952 p + geom_smooth(se = F) + ggtitle("Yearly EVI") # for the main title 699
953 ##### 700
954 # LST Seasonality Yearly 701
955 ##### 702
956 # Parse the dates, and use lower case names 703
957 df <- as_tibble(lst_stat) %>% 704
958 rename_all(tolower) %>% 705
959 mutate(date = ymd(date)) 706
960 ##### 707
961 # Define the plot 708
962 p <- df %>% 709
963 mutate(year = factor(year(date)), # use year to define separate curves 710
964 date = update(date, year = 1)) %>% # use a constant year for the x-axis 711
965 ggplot(aes(date, mean, color = year)) + 712
966 scale_x_date(date_breaks = "1 month", date_labels = "%b") 713
967 ##### 714
968 # Raw daily data 715
969 p + geom_line() 716
970 ##### 717
971 # Hone in on a single year 718
972 p + geom_line(aes(group = year), color = "black", alpha = 0.1) + 719
973 # geom_line(data = function(x) filter(x, year == 2010), size = 1) 720
974 ##### 721
975 # Smoothed version 722
976 p2 <- 723
977 p + geom_smooth(se = F) + ggtitle("Yearly LST") # for the main title 724
978 ##### 725
979 png(filename = "../Figures/Seasonality_All.png", 726
980 width = 1240, 727
981 height = 1000) 728
982 plot_grid(p1, 729
983 p2, 730
984 labels = "AUTO", 731
985 nrow = 2, 732
986 ncol = 1) 733
987 dev.off() 734
988 ##### 735
989 # EVI Seasonality Yearly 736
990 ##### 737
991 b <- 738
992 ##### 739
993 # EVI ARIMA Forecast 740
994 ##### 741
995 bfast( 742
996 bfast( 743
997 Yt = ts( 744
998 evi_stat$Mean, 745
999 deltat = 16 / 365, 746
1000 start = c(2008, 1) 747
1001 ), 748
1002 season = "harmonic", 749
1003 max.iter = 50 750
1004 ) 751
1005 ##### 752
1006 png(filename = "../Figures/Sesonal_BFAST_EVI.png", width = 1240) 753
1007 plot(b, main = "BFAST EVI") 754
1008 dev.off() 755
1009 ##### 756
1010 # EVI ARIMA Forecast 757
1011 ##### 758
1012 X <- 759
1013 tsclean(ts( 760
1014 evi_stat$Mean, 761
1015 deltat = 16 / 365, 762
1016 start = c(2008, 1) 763
1017 )) 764
1018 decomp <- stl(X, s.window = "periodic") 765
1019 un_seasoned <- seasadj(decomp) 766
1020 model1 <- auto.arima(un_seasoned, seasonal = T) 767
1021 f1 <- forecast(model1, 48) 768
1022 accuracy(f1) 769
1023 ##### 770
1024 accuracy(f1) 771
1025 png(filename = "../Figures/Sesonal_Decompo_EVI.png", 772
1026 width = 1920, 773
1027 height = 1080) 774
1028 plot(decomp, main = "Seasonal Decomposition of EVI") 775
1029 dev.off() 776
1030 ##### 777
1031 png(filename = "../Figures/Sesonal_Residuals_EVI.png", 778
1032 width = 1920, 779
1033 height = 1080) 780
1034 tsdisplay(residuals(model1), main = "EVI seasonal model") 781
1035 dev.off() 782
1036 ##### 783
1037 png(filename = "../Figures/Sesonal_Model_EVI.png", 784
1038 width = 1920, 785
1039 height = 1080) 786
1040 autoplot(f1, xlab = "Time", ylab = "EVI") + ggtitle("EVI ARIMA model forecast") 787
1041 dev.off() 788
1042 ##### 789
1043 # LST Seasonality Yearly 790
1044 ##### 791
1045 # LST Seasonality Yearly 792
1046 b <- 793
1047 bfast( 794
1048 Yt = ts( 795
1049 log(lst_stat$Mean), 796

```

```

1050      deltat = 8 / 365,
1051      start = c(2008, 1)
1052    ),
1053    season = "harmonic",
1054    max.iter = 50
1055  )
1056
1057 png(filename = "../Figures/Sesonal_BFAST_LST.png",
1058   width = 1920,
1059   height = 1080)
1060 plot(b, main = "BFAST LST")
1061 dev.off()
1062 #####
1063 # LST ARIMA Forecast
1064 #####
1065 X <-
1066 tsclean(ts(
1067   log(lst_stat$Mean),
1068   deltat = 8 / 365,
1069   start = c(2008, 1)
1070 ))
1071 decomp <- stl(X, s.window = "periodic")
1072 un_seasoned <- seasadj(decomp)
1073 model1 <- auto.arima(un_seasoned, seasonal = T)
1074 f1 <- forecast(model1, 96)
1075 accuracy(f1)
1076
1077 png(filename = "../Figures/Sesonal_Decompo_LST.png",
1078   width = 1920,
1079   height = 1080)
1080 plot(decomp, main = "Seasonal Decomposition of LST")
1081 dev.off()
1082
1083 png(filename = "../Figures/Sesonal_Residuals_LST.png",
1084   width = 1920,
1085   height = 1080)
1086 tsdisplay(residuals(model1), main = "LST seasonal model")
1087 dev.off()
1088
1089 png(filename = "../Figures/Sesonal_Model_LST.png",
1090   width = 1920,
1091   height = 1080)
1092 autoplot(f1, xlab = "Time", ylab = "LST") + ggtitle("LST ARIMA model forecast")
1093 dev.off()
1094
1095 # RasterVis Documentation
1096 # https://oscarperinan.github.io/rastervis/
1097
1098 # Unsupervised Clustering on Raster Data
1099 # https://www.r-exercises.com/2018/02/28/advanced-techniques-with-raster-data-part-1-unsupervised-classification/
1100
1101 # Plot Year Stacking
1102 # https://stackoverflow.com/questions/48722758/r-how-to-create-a-seasonal-plot-different-lines-for-years
1103
1104 # Seasonality ETS
1105 # https://www.statmethods.net/advstats/timeseries.html
1106
1107 # Seasonality Auto.Arima
1108 # https://www.datacamp.com/community/tutorials/time-series-r
1109
1110 # TimeSeries Analysis
1111 # https://www.datascience.com/blog/introduction-to-forecasting-with-arima-in-r-learn-data-science-tutorials
1112
1113

```

1114 References

1. Atkinson, P.M.; Lewis, P. Geostatistical classification for remote sensing: an introduction. *Computers & Geosciences* **2000**, *26*, 361–371, DOI:10.1016/S0098-3004(99)00117-X.
2. Bertin, R.I. Plant Phenology And Distribution In Relation To Recent Climate Change. *The Journal of the Torrey Botanical Society* **2008**, *135*, 1:126-1:46, DOI:10.3159/07-rp-035r.1.
3. Chuvieco, E.; Cocero, D.; Riaño, D.; Martin, P.; Martínez-Vega, J.; de la Riva, J.; Pérez, F. Combining NDVI and Surface Temperature for the Estimation of Live Fuel Moisture Content in Forest Fire Danger Rating. *Remote Sensing of Environment* **2004**, *92*, 3:322-3:31, DOI:10.1016/j.rse.2004.01.019.
4. Didan, K. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. [Data set] (Accessed April 28, 2019)
5. Eastman, J.; Sangermano, F.; Machado, E.; Rogan, J.; & Anyamba, A. Global trends in seasonality of normalized difference vegetation index (NDVI), 1982–2011. *Remote Sensing* **2013**, *5*, 10:4799-10:4818, DOI:10.3390/rs5104799.
6. Fernández-Manso, A.; Quintano, C.; Fernández-Manso, O. Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale *International Journal of Remote Sensing* **2011**, *32*, 1595-1617, DOI:10.1080/01431160903586765

- 1130 7. García-Mora, T.J.; Mas, J.; Hinkley, E.A. Land cover mapping applications with MODIS: a literature review.
1131 *International Journal of Digital Earth* **2012**, *5*:1, 63-87, DOI:10.1080/17538947.2011.565080
- 1132 8. Han, P.; Wang, P.X.; Zhang, S.Y.; Zhu, D.H. Drought forecasting based on the remote sensing data using
1133 ARIMA models *Mathematical and Computer Modelling* **2010**, *51*, 1398-1403, DOI:10.1016/j.mcm.2009.10.031
- 1134 9. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and
1135 biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* **2002**, *83*, 195-213,
1136 DOI:10.1016/S0034-4257(02)00096-2
- 1137 10. Huete, A.; Justice, C.; Van Leeuwen, W. MODIS vegetation index (MOD13) algorithm theoretical basis
1138 document. Ver. 3 NASA: Washington, DC, USA **1999**.
- 1139 11. Jennings, M.D. Gap analysis: Concepts, methods, and recent results. *Landscape Ecology* **2000**, *15*, 5-20,
1140 DOI:10.1023/A:1008184408300.
- 1141 12. Martinuzzi, S.; Gould, W.A.; Gonzalez, O.M.R. Land development, land use, and urban sprawl in Puerto Rico
1142 integrating remote sensing and population census data. *Landscape and Urban Planning* **2017**, *79*, 288-297.
- 1143 13. MODIS Evapotranspiration. <https://modis.gsfc.nasa.gov/data/dataproducts/mod16.php> (Accessed April 28,
1144 2019)
- 1145 14. MODIS Land Surface Temperature and Emissivity (MOD11).<https://modis.gsfc.nasa.gov/data/dataproducts/mod11.php>
1146 (Accessed April 28, 2019)
- 1147 15. MODIS Vegetation Index Products (NDVI and EVI). <https://modis.gsfc.nasa.gov/data/dataproducts/mod13.php>
1148 (Accessed April 28, 2019)
- 1149 16. Reiter, M.E.; Elliott, N.K.; Jongsomjit, D.; Golet, G.H.; Reynolds, M.D. Impact of extreme drought and
1150 incentive programs on flooded agriculture and wetlands in California's Central Valley. *PeerJ* **2018**, *6*:e5147,
1151 DOI:10.7717/peerj.5147.
- 1152 17. Rigge, M.; Smart, A.; Wylie, B.; Gilmanov, T.; Johnson, P. Linking Phenology and Biomass Productivity in South
1153 Dakota Mixed-Grass Prairie. *Rangeland Ecology & Management* **2013**, *66*, 579-87, DOI:10.2111/rem-d-12-00083.1.
- 1154 18. Running, S., Mu, Q., Zhao, M. MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN
1155 Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC.[Data set] (Accessed April 28, 2019)
- 1156 19. Senay, G.; Velpuri, N.; Bohms, S.; Budde, M.; Young, C.; Rowland, J.; Verdin, J. Drought monitoring and
1157 assessment: remote sensing and modeling approaches for the famine early warning systems network.
Hydro-Meteorological Hazards, Risks and Disasters **2015**, 233-262. DOI:10.1016/B978-0-12-394846-5.00009-6.
- 1158 20. Steyer, G.D.; Couvillion, B.R.; Barras, J.A. Monitoring vegetation response to episodic disturbance events by
1159 using multitemporal vegetation indices *Journal of Coastal Research* **2013**, Spring, 118-130, DOI:10.2112/SI63-011.1.
- 1160 21. Sjostrom, M.; Ardo, J.; Arneth, A.; Boulain, N.; Cappelaere, B.; Eklundh, L.; De Grandcourt, A.; Kutsch, W.L.;
1161 Merbold, L.; Nouvellon, Y.; Scholes, R.J.; Schubert, P.; Seaquist, J.; Veenendaal, E.M. Exploring the potential of
1162 MODIS EVI for modeling gross primary production across African ecosystems *Remote Sensing of Environment*
1163 **2011**, *115*, 1081-1089, DOI:10.1016/j.rse.2010.12.013.
- 1164 22. Torres-Degró, A. Envejecimiento demográfico: Un acercamiento a los métodos cuantitativos. *CIDE digital* **2010**,
1165 *1*, 79-102.
- 1166 23. Psilovikos, A.; Elhag, M. Forecasting of Remotely Sensed Daily Evapotranspiration Data Over Nile Delta
1167 Region, Egypt *Water Resources Management* **2013**, *27*, 4115-4130, DOI:10.1007/s11269-013-0368-2
- 1168 24. Verbesselt, J.; Hyndman, R.; Newnham, G.; Culvenor, D. Detecting trend and seasonal changes in satellite
1169 image time series. *Remote Sensing of Environment* **2009**, *114*, 106–115, DOI:10.1016/j.rse.2009.08.014.
- 1170 25. Wan, Z.; Hook, S.; and Hulley, G. MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3
1171 Global 1km SIN Grid V006. [Data set] (Accessed April 28, 2019)
- 1172 26. Wang, C.; Yu, M.; Gao, Q. Continued Reforestation and Urban Expansion in the New Century of a Tropical
1173 Island in the Caribbean *Remote Sensing* **2017**, *9*, 731.
- 1174 27. Wulder, M.; Michael A.; Hall, R.; Coops, N.; Franklin, S. High Spatial Resolution
1175 Remotely Sensed Data for Ecosystem Characterization. *BioScience* **2004**, *54*, 511-521,
1176 DOI:10.1641/0006-3568(2004)054[0511:HSRRSD]2.0.CO;2.

1177 **Sample Availability:** The data used for this works was obtained from and is available through NASA's
1178 AppEEARS portal where the request ID for the data is 5643f426 – f351 – 429d – 8607 – d4555eabc307. Projection
1179 and request information for the data utilized in the project can be downloaded from:
1180 <https://drive.google.com/a/upr.edu/file/d/1nS5aluj7DH0MxjrtQLywf1tgBZBsxBWJ4/view?usp=sharing>

1182 © 2019 by the authors. Submitted to *Remote Sens.* for possible open access publication
1183 under the terms and conditions of the Creative Commons Attribution (CC BY) license
1184 (<http://creativecommons.org/licenses/by/4.0/>).