

Forecasting vegetation condition for drought early warning systems in pastoral communities in Kenya

Adam B. Barrett^{a,b}, Steven Duivenvoorden^{a,c}, Edward Salakpi^a, James M. Muthoka^d, Seb Oliver^{c,a} and Pedram Rowhani^{d*}

^a *The Data Intensive Science Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*

^b *Sackler Centre for Consciousness Science, Department of Informatics, University of Sussex, Brighton BN1 9QJ, UK*

^c *Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*

^d *School of Global Studies, Department of Geography, University of Sussex, Brighton, BN1 9QJ, UK*

Corresponding author: P.Rowhani@sussex.ac.uk

Abstract

Live version, post 1st review, pre resubmission This is such an amazing paper.

Droughts are a recurring hazard in sub-Saharan Africa, that can wreak huge socioeconomic costs. Acting early based on alerts provided by early warning systems (EWS) can potentially provide substantial mitigation in terms of money and lives lost. However existing EWS tend only to monitor, rather than forecast, the environmental and socioeconomic indicators of drought, and hence are not always sufficiently timely to be effective in practice. Here we make a first attempt [is it really first?, Seb] at forecasting satellite-based indicators of vegetation condition that are commonly monitored. Specifically, we forecast the Normalized Difference Vegetation Index (NDVI) and the Vegetation Condition Index (VCI) over pastoral livelihood zones in Kenya as these are the common indicators used by the National Drought Management Authority (NDMA). Using data from MODIS and Landsat, we apply linear autoregression and Gaussian processes modeling methods and demonstrate accurate forecasting several weeks ahead. We explored predicting the drought alert marker used by NDMA

(3 month VCI< 35). Both of our models were able to predict this alert marker four weeks ahead with a hit rate of around 89% and a false alarm rate of around 4%, or 81% and 6% respectively six weeks ahead. The methods developed here can identify a deteriorating vegetation condition well in advance and thus help disaster risk managers act early to support vulnerable communities and limit the impact of a drought hazard.

Keywords: Landsat; MODIS; Gaussian Processes; Drought; NDVI; VCI

1 1. Introduction

2 Droughts are a major threat globally as they can cause substantial damage
3 to society, especially in regions that depend on rain-fed agriculture. They par-
4 ticularly impact food security by significantly reducing agricultural production
5 (?) and raising food prices (??), which often leads to increased levels of mal-
6 nutrition, migration, disease, and other health concerns (??). Since 2000, there
7 have been 319 drought events reported (EMDAT 2019), which together have
8 killed over 21,000 people and affected almost 1.4 billion others. The majority
9 of these events took place in sub-Saharan Africa where many communities rely
10 on predictable rainfall patterns for their livelihood.

11 In East Africa, the main economic activity in the arid and semi-arid lands
12 (ASAL) is subsistence rain-fed agriculture, as well as livestock farming using
13 pastures and grasslands as the main source of fodder. The pastoral and agro-
14 pastoral communities who live in these drylands have dealt with rainfall vari-
15 ability and drought over centuries by developing extensive adaptation and miti-
16 gation strategies to reduce their vulnerability to these shocks (??). However, in
17 recent years these communities have seen their coping strategies compromised
18 by population growth and land use change (?). Additionally, while there is some
19 uncertainty in the climate models (IPCC, ?), rainfall variability is expected to
20 increase in the region (??). These factors in combination will make it harder

21 for indigenous knowledge systems to deal with droughts, and exacerbate the
22 problems created by droughts. Governments and donor agencies in the region
23 have thus developed several tools and early warning systems (EWS) to mitigate
24 the impact of droughts on pastoralists.

25 Most EWS tend to monitor current key biophysical and socio-economic fac-
26 tors to assess the possible exposure of vulnerable people to specific hazards.
27 However, once the impacts are visible, it will be too late to mitigate the con-
28 sequences (?). **As a consequence**, being better prepared before a drought hits
29 significantly reduces the costs and losses from these disasters (?). Hence, EWS
30 now increasingly include expert knowledge and qualitative assessments of sea-
31 sonal climate forecasts to assess the future development of food security, and
32 define actions to mitigate possible losses (??).

33 Within East Africa, the Famine Early Warning Systems Network (FEWS
34 NET) monitors food security through data collection and a deep understanding
35 of the livelihood patterns in the region. A team of experts and analysts will also
36 look at seasonal climate forecast to estimate future food security outcomes using
37 scenario development (?). In Kenya, the drought EWS operated by the National
38 Drought Management Authority (NDMA) provides monthly bulletins assessing
39 food security in the 23 ASAL regions **using current** environmental (rainfall,
40 vegetation condition) and socio-economic (production, access, and utilisation)
41 factors. Based on these factors, the bulletins include a qualitative evaluation
42 of food security outcomes in the months ahead. **Shouldn't the next sentence**
43 **be in the discussion, not here, because we are not addressing consequences in**
44 **this paper and also our premise is that the current EWS do not forecast at**
45 **all [Seb].** However, EWS should move from forecasting hydro-meteorological
46 events toward estimating the expected consequences of hazards, i.e. impact-
47 based forecasting, to identify more effective early action protocols (??).

48 Pastoralists strongly rely on forage availability to keep their livestock. Existing
49 EWS provide information on pasture condition through the use of satellite-based
50 Earth observation derived vegetation indices, such as the Normalized
51 Difference Vegetation Index (NDVI) and the Vegetation Condition Index (VCI,
52 ????). These products are commonly used as drought indicators as they provide
53 timely and regular assessment of vegetation health over large spatial areas.
54 EWS should include a forecast of these indicators as it would allow local and national
55 stakeholders to act early and support the pastoralists in times of drought.
56 Recent studies have highlighted the potential of satellite-based earth observation
57 data to forecast agricultural productivity (?) and seasonal forage availability
58 (?).

59 The main goal of this paper is to explore methods to forecast (up to six weeks
60 ahead) the vegetation indices that are commonly used in the pastoral areas of
61 Kenya to monitor droughts. We specifically aim to estimate the potential to
62 forecast NDVI and the 3-month VCI (VCI3M), as used by the NDMA in their
63 monthly bulletins. The indices are extracted from data derived from the Landsat
64 mission (every 16 days at 30 m resolution and the MODerate resolution imaging
65 Spectroradiometer (MODIS - daily data at 500 m resolution).

66 Machine-learning techniques offer a data-driven, empirical route to these
67 forecasts. Many different data inputs could be used to forecast these vegetation
68 indices (e.g. precipitation and precipitation forecasts. However, perhaps the
69 most simple is to use the past history of the indices themselves. This has the
70 practical benefits of readily available data over large areas. It is also likely
71 to work as these indices are subject to plant growth and climate cycles giving
72 periodic behaviour on large temporal scales that can be empirically modelled
73 while external perturbations, such as water availability, have persistent impact
74 providing correlations on short temporal scales. The existing EWS itself implies

75 that the indices have forecasting power as moderately low indexes are labelled
76 “alert” implying they might precede lower “alarm” levels. something about
77 the Boku paper here

78 The machine-learning techniques we attempt are Gaussian Process regression
79 (GP, ?) w), and linear autoregressive (AR) modelling (e.g. ?). GP regression
80 uses kernel-based non-parametric Bayesian inference on the structure of correla-
81 tions between observations, and is widely applied to classification, interpolation,
82 change detection and forecasting problems (????). For an overview on the prin-
83 ciples of GPs, and how they have previously been applied throughout remote
84 sensing, see ?. Linear AR is the regression of future observations on past ob-
85 servations, assuming a linear dependence. This has previously been performed
86 on monthly (i.e. temporally more sparse) NDVI data, see for example ? and ?,
87 with mixed results in terms of forecasting potential (R^2 -scores between 0 and
88 0.4 at a lead time of one month).

89 I think this introduction needs something to end on that leads into the paper
90 better

91 **2. Study area**

92 About 80% of Kenya lies within the ASALs, and the main economic activity
93 in these regions is livestock farming (??). The livestock sector accounts for 13%
94 of the national GDP and 43% of its agricultural GDP. Pastures and grasslands
95 in the ASAL serve as the main source of fodder for the pastoral communities
96 in Kenya (?). Following several periods of intense drought, the government in
97 Kenya established the NDMA in 2016, to set up and operate a drought early
98 warning system (DEWS), as well as to establish drought preparedness strategies
99 and contingency plans (GoK). One key biophysical indicator used by the NDMA
100 drought phase classification is the VCI (?), which is measured at county level as

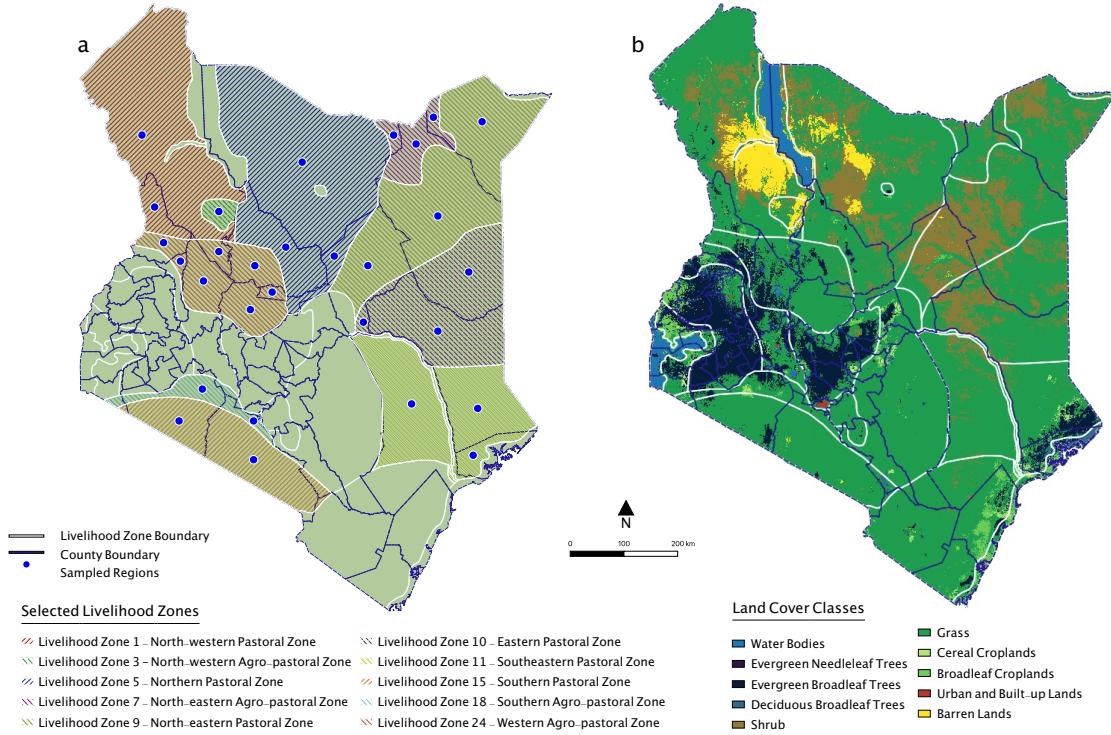


Figure 1: Maps of Kenya showing (A) the livelihood zones from which pixels were sampled for analysis, and (B) land cover classification (according to the MODIS MCD12Q1 map). Analyses were performed for 29 regions, defined by pastoral livelihood zone and county intersections. For the MODIS data only those pixels identified as being from grassland are used.

101 well as over the different livelihood zones within the county (FEWS NET 2011).
 102 This study focused on the 10 (agro)-pastoral livelihood zones (see Fig ??), which
 103 cross 15 counties. The names of the 29 livelihood zone county intersections can
 104 be found in Appendix ???. FEWS NET shape files were used to define these
 105 regions, and to demarcate the Landsat and MODIS pixels from which to sample
 106 data.

107 3. Data

108 This section describes the two datasets. A comparison between them, and
 109 justification of data selection can be found in Section XXX in the Supplementary

110 Material.

111 *3.1. Landsat*

112 Landsat-5, 7 and 8 (?) red and near infrared (NIR) surface reflectances and
113 quality assessment (QA) data over the 10 pastoral livelihood zones of Kenya,
114 from 1/1/2000 to 1/2/2019, were obtained using the United States Geological
115 Survey (USGS) EarthExplorer. Specifically, data were drawn from the Level-1
116 Precision Terrain (L1TP) processed dataset, which has well-characterized ra-
117 diometry and is inter-calibrated across the different Landsat sensors. The QA
118 data indicate whether or not each observation is affected by cloud; only ob-
119 servations classified as clear were kept. The spatial resolution of these data is
120 30m and the repeat interval is 16 days. Landsat-5 data were available up until
121 November 2011 (albeit with several large gaps), Landsat-7 data were available
122 for the whole time-period, and Landsat-8 data were available from March 2013.
123 The surface reflectances were combined to obtain NDVI according to the usual
124 formula. From each of the 29 pastoral livelihood zone and county intersections
125 (see Fig. ??), 1000 random pixels were selected for analysis, out of those for
126 which at least half of the observations were present and labelled good quality.¹

127 *3.2. MODIS*

128 NDVI data were also obtained from the MODIS Terra/Aqua Nadir BRDF-
129 Adjusted Reflectance (NBAR) product (MCD43A4,v006; ?) via the NASA
130 Land Processes Distributed Active Archive Center (LP DAAC) using Applica-
131 tion for Extracting and Exploring Analysis Ready Samples (AppEEARS), from
132 the start date, 22/2/2000, up to 1/2/2019. The spatial resolution of this prod-
133 uct is 500m and the repeat interval is 1 day, with the daily data being 16 day

1 Except for a few regions, for which this threshold had to be dropped in order to be able to obtain 1000 pixels, see Table ?? in the Appendix.

134 composites of raw observations at 1 to 2 day intervals. From each of the 29 pas-
135 toral livelihood zone and county intersections, 100 random pixels were selected
136 for analysis, out of those pixels identified as being from grassland according to
137 the MODIS land cover classification maps (MCD12Q1,v006).

138 **4. Pre-processing**

139 *4.1. Temporal gridding and gap-filling*

140 To prepare the datasets for the testing of forecasting, each NDVI time series
141 was processed into a form containing, wherever possible, precisely one observa-
142 tion every 7 days. For the MODIS data, the mean of all reliable data was taken
143 for each week. For the Landsat data, this required compilation and interpolation
144 of data sampled every 16 days from up to 3 different Landsat missions. Further,
145 gap-filling had to be carried out on both datasets whenever there was a lack
146 of reliable observations due to cloud cover and/or instrumental malfunction.
147 The following subsections describe briefly the gridding and gap-filling meth-
148 ods employed on the Landsat and MODIS data, respectively. Further details
149 and justification of these methods can be found in the Supplementary Material,
150 Section XXX.

151 *4.1.1. Gridding and gap-filling on Landsat data using Gaussian Processes*

152 Gridding and gap-filling on the Landsat data was done using Gaussian Pro-
153 cess (GP) modelling; for details see Section XXX in the Supplementary Mate-
154 rial. For a given pixel, the GP modelling took raw data as input, fit a temporal
155 correlation structure to the data, and used this to output a time series of ex-
156 pected NDVI values, with observations on every Saturday from 1/1/2000 to
157 2/2/2019. Two versions of GP gap-filling were carried out, which we refer to as
158 forecast mode and non-forecast mode. For the non-forecast mode all of the data
159 (1/1/2000 to 1/2/2019) from the given pixel were used to train the GP. The

160 non-forecasting mode was used as the “ground truth” to test forecasts against.
161 The forecast mode, by contrast, only used data up to a certain date, whichever
162 date a forecast was being attempted from - since when doing forecasting with a
163 near real-time data stream, one does not have access to future data.

164 *4.1.2. Gridding and gap-filling on MODIS data*

165 As mentioned above, gridding of the MODIS data was done by taking the
166 mean of all reliable [data](#) from each week. Gaps were then filled using temporal
167 quadratic interpolation. Quadratic interpolation fills each gap using the best fit
168 quadratic function to the two observations before the gap and the two observa-
169 tions after the gap (via least squares). The Savitzky-Golay (SG, [?](#)) smoothing
170 method was then used to filter high-frequency measurement noise. SG smooth-
171 ing involved, for each observation, fitting a polynomial to a window centred
172 on the observation, and then replacing that observation with the corresponding
173 observation on the polynomial fit. [To determine the optimal window length, the](#)
174 [polynomial order was set 2 and different windows lengths were tested as done](#)
175 [by ??](#). The polynomials were fit, for each [each window length](#), to the original
176 data. [Based on previous applications of this method to filter MODIS data \(??\).](#)
177 A sliding windows of length 7 time-steps (weeks) and a polynomial of order 2
178 (i.e. quadratic function) were chosen and fitting done using least squares.

179 [A sliding windows of length 7 time-steps \(weeks\) and a polynomial of order](#)
180 [2 \(i.e. quadratic function\) were chosen because, a good amount of the noise was](#)
181 [filtered out and resulting smooth NDVI time were far off the original data..](#)

182 A summary of the entire work from data preparation to forecasting drought
183 can be seen in figure([??](#))

184 Since there were occasional long segments of interpolated data, where large
185 gaps had been filled, and NDVI values sometime took unrealistic values (greater
186 than 1 or less than 0), it was decided to remove all interpolations longer than 6

¹⁸⁷ weeks, see Section ?? in the Supplementary Material for details.

188 **5. Methods**

189 *5.1. Forecast indices and aggregation*

The forecast indices were ([absolute](#)) NDVI anomaly and VCI3M. NDVI anomaly is the mean-subtracted NDVI, where the mean is the average NDVI for the given week of the year over all years covered by the data. VCI3M is the mean VCI across the 3 months leading up to the given observation time point, with VCI at time point i being defined as

$$VCI_i = 100 \times \frac{NDVI_i - NDVI_{min,i}}{NDVI_{max,i} - NDVI_{min,i}}, \quad (1)$$

190 where $NDVI_{min,i}$ and $NDVI_{max,i}$ are the minimum and maximum values for the
191 $NDVI$ of the pixel in the given week over all years covered by the data. On
192 Landsat pixels, the mean, maximum and minimum value for the $NDVI$ for each
193 week of the year was computed using the non-forecast mode GP interpolated
194 time series. Then forecast mode and non-forecast mode versions of each index
195 were created.

196 With both the Landsat and MODIS datasets, individual pixel time series
197 for each index were aggregated, by taking the mean, within each livelihood
198 zone and county intersection. Thus forecasting was carried out on a single time
199 series for each region (see Fig. ?? for an example time series). With the MODIS
200 data, since not all gaps were interpolated, whenever there were fewer than 25
201 individual pixel observations from a particular region at a given time point, it
202 was decided that there should be no datum in the aggregate time series, i.e. there
203 should be a gap.²

²VCI3M is the mean of the most recent 12 weeks of VCI observations; if some of these were missing, the mean was taken over just the ones that were present. If the current observation was not present, a gap was placed in the VCI3M time series.

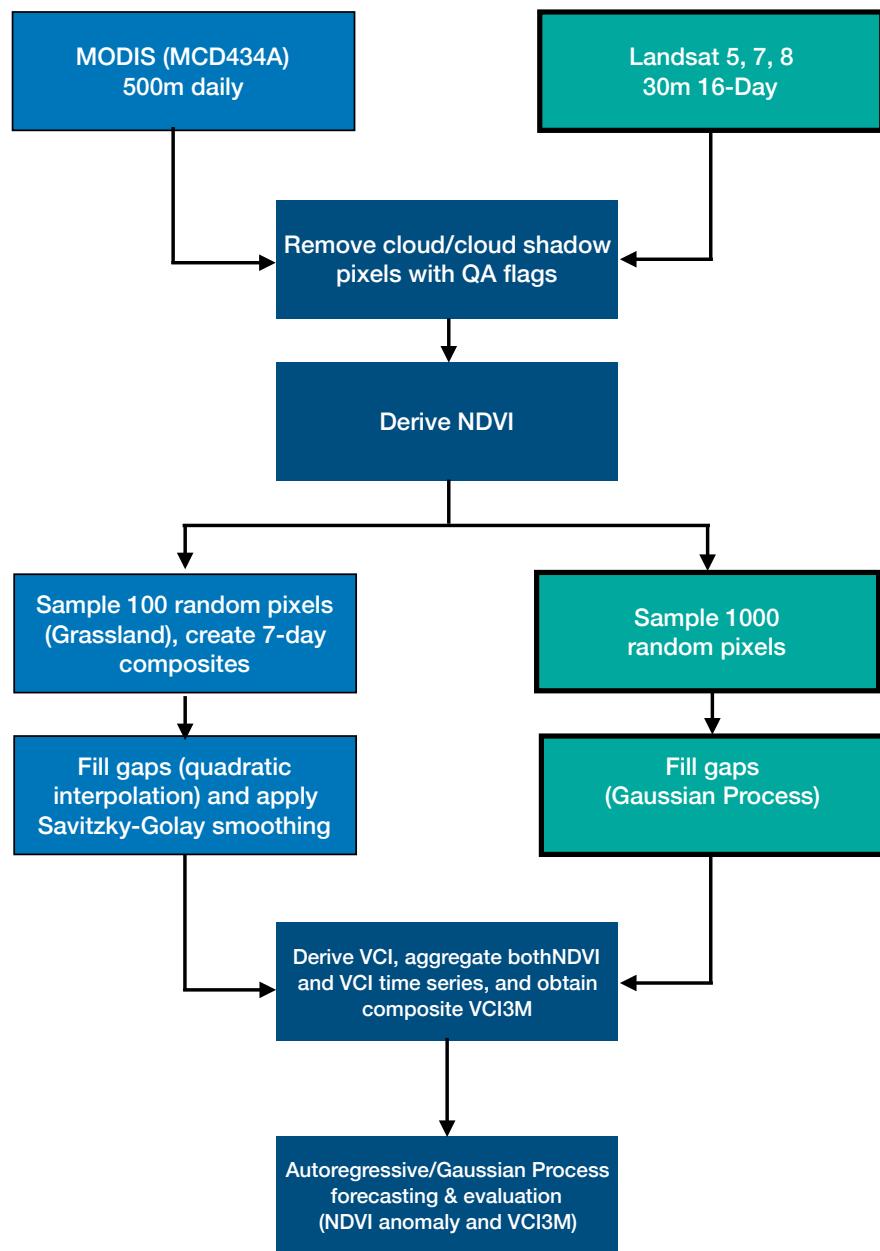


Figure 2: A Flow chart of the data processing and analysis

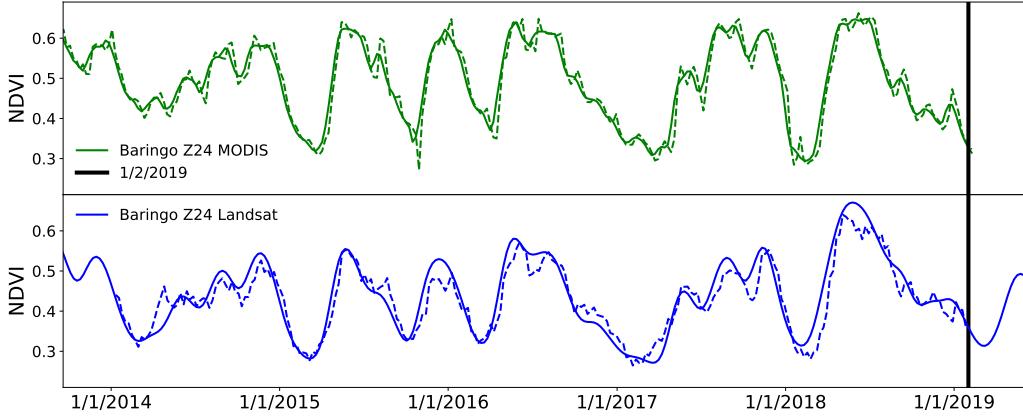


Figure 3: Aggregated NDVI time series from the intersection of Baringo county and livelihood zone 24. The top panel shows MODIS data, and the bottom panel shows Landsat data (solid lines), processed using the methods described in the text. The dotted lines in each panel show forecasting at a lead time of 2 weeks, using the AR method on the MODIS data, and the GP method on the Landsat data.

204 5.2. Forecasting

GP forecasting was performed by GP modelling and extrapolation, see Section ?? in the Supplementary Material for details. AR forecasting was performed with the following model-fitting and extrapolation method. For forecasting n weeks ahead, the following model was fit:

$$X_{t+n} = \sum_{i=0}^{p-1} a_i X_{t-i} + \epsilon_t , \quad (2)$$

205 where X is NDVI anomaly (or VCI3M with mean removed), subscripts denote
206 the observation (week), a_i are model parameters, ϵ_t are the residuals (i.e. the
207 errors), and p is the model order. Fitting the model to a segment of data involved
208 finding the model parameters that gave the minimum sum-square error, i.e. led
209 to residuals with the minimum variance. To make a forecast, the model was fit
210 using the most recent T consecutive observations, and then used to predict the

211 observation n weeks after the most recent observation. This forecasting method
 212 was carried out along the entire available time series, fitting a distinct model
 213 to each segment of length T . A search for optimal model orders and training
 214 segment lengths found that forecast quality, as measured by root mean square
 215 error (RMSE), plateaued at $T = 200$ and $p = 3$.³

216 *5.3. Metrics for assessing forecasts*

Several metrics were used to assess the performance of the forecast methods tested on the data. In addition to RMSE, the R^2 -score and the percentage of standard deviation remaining, S , were used. These are given by:

$$R^2\text{-score} = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (3)$$

$$S = 100 \times \frac{\sqrt{\sum_i (y_i - f_i)^2}}{\sqrt{\sum_i (y_i - \bar{y})^2}}, \quad (4)$$

217 where the y_i are the true data, and the f_i are the forecasts. Note that $S \equiv$
 218 $100 \times \sqrt{1 - R^2\text{-score}}$. To test for bias, linear regression of actual index on
 219 forecast index was performed, and slope and intercept computed. Finally, receiver
 220 operating characteristic (ROC) curves were constructed for forecast-based
 221 drought-alert detection.

222 **6. Results**

223 *6.1. Forecast values compared with true values*

224 The GP and AR forecasting methods were applied, on each of the two
 225 datasets, to regional aggregate NDVI and VCI3M time series. We focus on
 226 performance results of GP forecasting on Landsat data and AR forecasting on

³High values of p up to 20 were also explored with LASSO regression (a regularisation procedure), but this led to inferior forecasting.

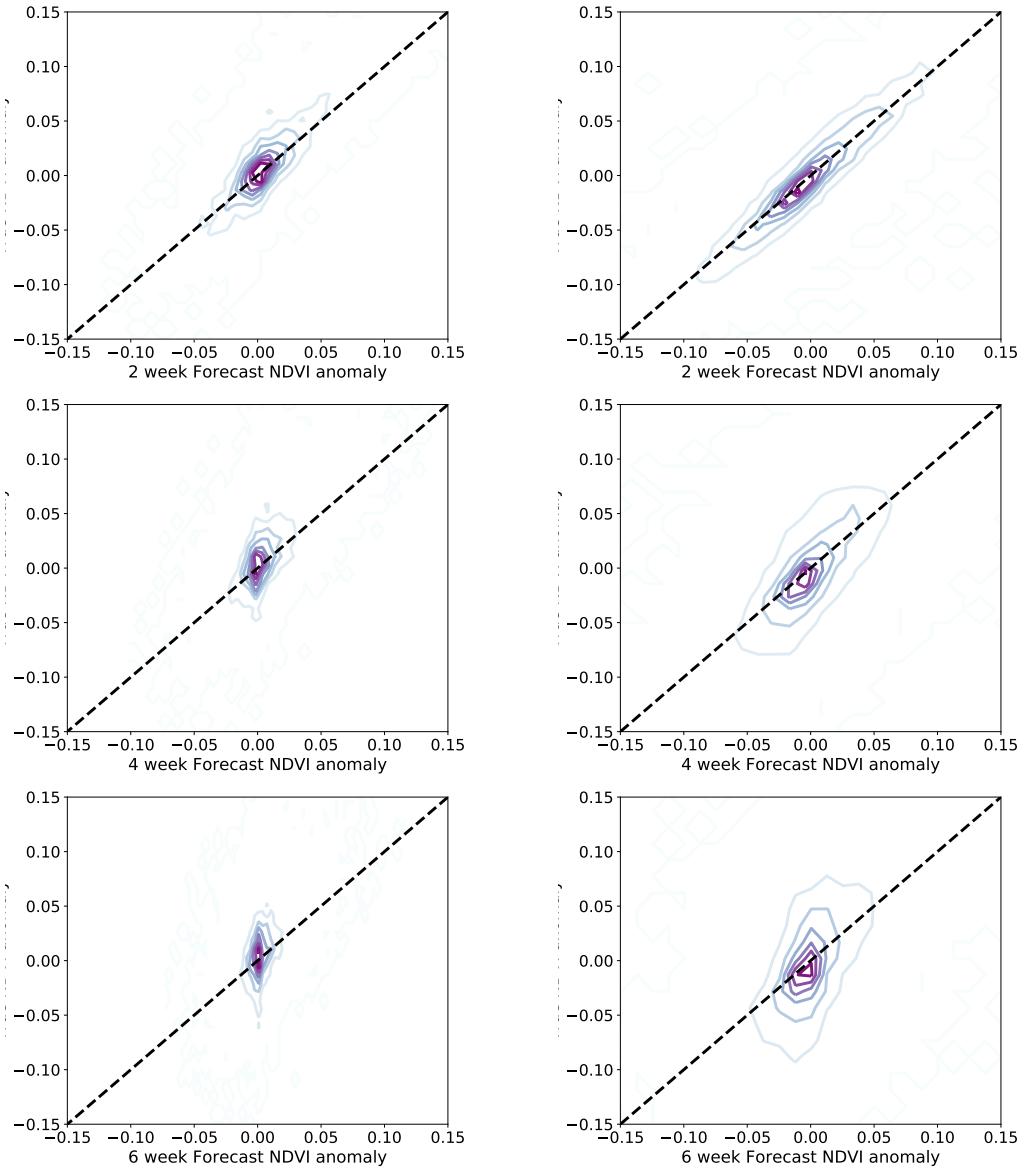


Figure 4: TO DO: EDWARD, REFORMAT SO Y AXIS LABEL IS VISIBLE. Contour plots of NDVI anomaly against our two, four and six weeks NDVI anomaly forecast. The left three plots show our forecast performance for the GP method on Landsat data, and on the right the contours show the forecast performance for the AR method on MODIS data, across the 19 regions for which a forecast is possible more than 50% of the time.

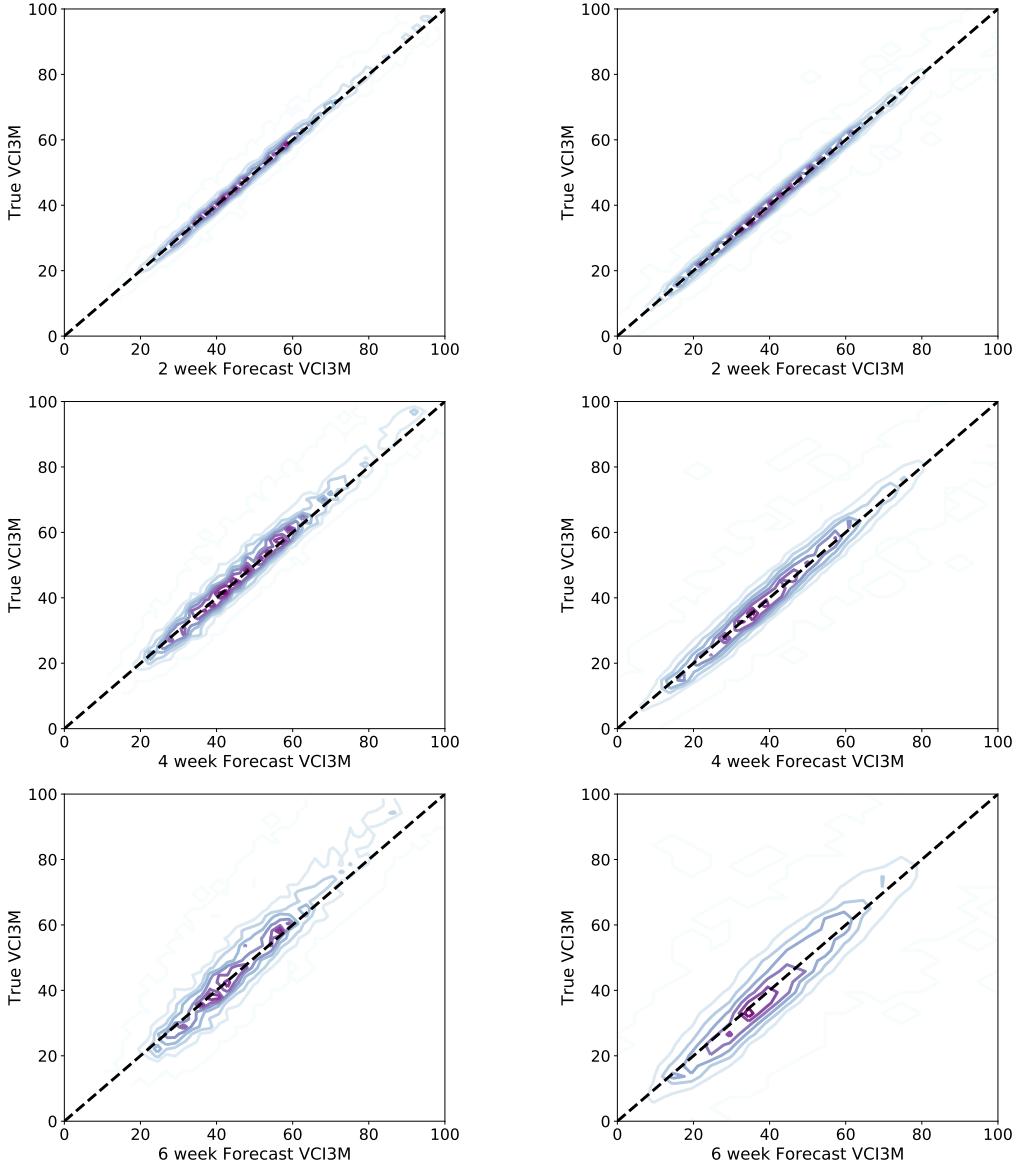


Figure 5: Contour plots of VCI3M against our two, four and six weeks VCI3m forecast. The left three plots show our forecast performance for the GP method on Landsat data, and on the right the contours show the forecast performance for the AR method on MODIS data, across the 19 regions for which a forecast is possible more than 50% of the time.

Table 1: TO DO: STEVEN/EDWARD FILL THIS OUT FOR GP/LANDSAT. Performance statistics of NDVI anomaly and VCI3m forecasting.

	Landsat GP			MODIS AR		
	2 weeks	4 weeks	6 weeks	2 weeks	4 weeks	6 weeks
NDVI anomaly:						
R^2 -score				0.85	0.55	0.33
RMSE				0.025	0.043	0.053
slope				0.99	0.99	0.97
intercept				-0.00	-0.00	-0.00
VCI3M:						
R^2 -score				0.99	0.96	0.88
RMSE				1.8	4.3	7.0
slope				1.00	1.00	1.00
intercept				-0.0	-0.1	-0.0

227 MODIS data since these two combinations of data and forecasting method per-
 228 formed the best (as measured by R^2 -score, see Section ?? in the Supplementary
 229 Material). Contour plots of forecast against actual data for two, four and six
 230 week forecasts are shown in Fig. ?? for NDVI anomaly, and Fig. ?? for VCI3M.
 231 Table ?? shows the R^2 -scores, RMSE, slope and intercept from each of these
 232 plots, and demonstrates that there is substantial forecast skill from each method
 233 at each lead time (R^2 -scores are substantial), and that the forecasts are unbi-
 234 ased (slopes are all approximately 1, and intercepts approximately 0). The much
 235 higher R^2 -scores for VCI3M compared to NDVI anomaly is explained by the
 236 fact that VCI3M is a 12 week aggregate, and hence its (near) future is actually
 237 derived from both past and future NDVI values.

238 Due to the presence of non-interpolated gaps in the MODIS time series,
 239 there were weeks when a forecast assessment was not carried out on these data,
 240 see Section ?? in the Supplementary Material for details. For 15 of the regions,
 241 a 4 week forecast could be made on more than 90 percent of weeks; however,
 242 for some of the more cloudy/wet regions, a forecast could rarely be made, see
 243 Table ??.

244 To check if forecast skill depended on the true vegetation condition, RMS

Table 2: RMS error in VCI3M forecast, for the true vegetation condition belonging to the different categories of drought, at lead times of 2, 4 and 6 weeks. (The bottom category, extreme drought, did not occur according to the Landsat data.)

Drought category	Landsat GP			MODIS AR		
	2 weeks	4 weeks	6 weeks	2 weeks	4 weeks	6 weeks
Wet, VCI3M>50	2.2	5.3	9.0	2.2	4.8	7.5
Normal, 35<VCI3M<50	1.7	3.4	5.0	1.6	4.0	6.5
Moderate drought, 20<VCI3M<35	1.5	3.2	5.0	1.5	3.7	5.7
Severe drought, 10<VCI3M<20	1.1	2.5	5.5	1.4	3.3	5.4
Extreme drought, VCI3M<10				1.1	2.9	4.8

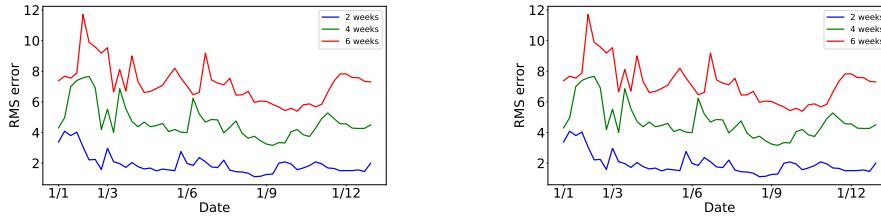


Figure 6: TO DO: STEVEN/EDWARD REPLACE LEFT HAND PLOTS WITH THE CORRESPONDING PLOTS FROM LANDSAT/GP. RMS error of VCI3M forecast for each week of the year. (Left) GP forecasting on Landsat data. (Right) AR forecasting on MODIS data.

245 error was also computed separately for each of the five categories on the NDMA
 246 drought scale. The results are shown in Table ??, where it can be seen that
 247 RMS error tended to be lower when there was a state of drought than when the
 248 vegetation condition was normal.

249 To check if forecast skill depended on the time of year, RMS error of VCI3M
 250 forecast was plotted against the week of the year, see Fig. ?? . The plot shows
 251 that the seasonal differences in RMS error are not substantial on the scale on
 252 which VCI3M varies, although RMS error was generally somewhat elevated for
 253 some of the January/February dry season.

254 6.2. Drought event forecast: ROC curves

255 To assess the usefulness of the AR and GP methods for drought forecast-
 256 ing, we tested their ability to detect specific drought events, as defined by the
 257 NDMA's alert threshold (VCI3M<35, ?). Receiver operating characteristic

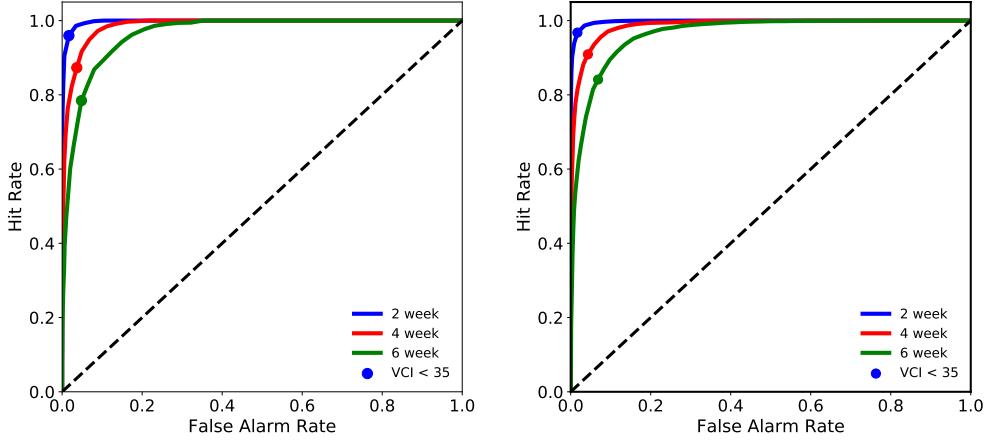


Figure 7: (Left) ROC curve for drought detection ($VCI3M < 35$) for lead times of 2, 4 and 6 weeks using the GP method. (Right) ROC curve for drought detection using the AR method. The curves are plotted from applying different thresholds to convert the continuous forecast into a binary forecast of drought or no drought, see text for details. The shaded circles show the point obtained from forecasting drought when the predicted $VCI3M < 35$. The area under the curve is 1.0, 0.98, 0.96 (GP, left) and 1.0, 0.99 and 0.96 (AR, right) for lead times of two, four and six weeks, respectively.

(ROC) curves were plotted for detection of $VCI3M < 35$ at lead times of two, four and six weeks (Fig. ??). These curves show the probability of predicting a state of drought ($VCI3M < 35$) when there will be a state of drought, i.e. hit rate, against the probability of predicting drought when there will not be drought, i.e. false alarm rate, for varying binarisation thresholds on the forecast. These curves give an indication that one can forecast droughts with these methods even as far as six weeks ahead.

The ROC curve performance is not highly dependent on the region (see Table ??). Even for the wetter Eastern regions, for which observations are sparser due to cloud cover, the hit and false alarm rates only differ by 1 to 2 percentage points compared with those computed across all regions. Further, ROC curves for predicting the NDMA drought categories of severe ($10 < VCI3M < 20$) or extreme ($VCI < 10$) drought look similar to those for detecting $VCI3m < 35$, see Fig. ??.

Table 3: False alarm rate and hit rate (respectively, in percent) for different regions in Kenya based on forecasting drought if the predicted VCI3M is less than 35.

Regions	Landsat GP			MODIS AR		
	2 weeks	4 weeks	6 weeks	2 weeks	4 weeks	6 weeks
All	2% 96%	4% 87%	5% 78%	2% 97%	4% 91%	7% 84%
Z24	2% 99%	4% 91%	5% 82%	2% 98%	5% 94%	8% 88%
North (Z1,3 and 5)	1% 97%	2% 88%	3% 76%	2% 98%	6% 93%	11% 87%
East (Z7, 9, 10 and 11)	3% 94%	5% 85%	6% 77%	3% 97%	6% 91%	10% 85%
South (Z15 and 18)	1% 96%	3% 88%	4% 77%	2% 98%	6% 94%	11% 90%

271 Together, these results demonstrate that there is a huge potential for drought
 272 forecasting, and encourage a future cost-benefit analysis of applying such a
 273 forecast in practice.

274 7. Discussion and Conclusion

275 This paper highlights the potential of two separate methods for drought fore-
 276 casting in pastoral regions of Kenya. The linear autoregression models applied
 277 to MODIS achieved an R^2 -score of 0.58 for NDVI anomaly at a lead time of
 278 4 weeks, and an R^2 -score of 0.95 for the VCI3M, the three-month vegetation
 279 condition index used within the drought early warning system developed by
 280 the National Drought Management Authority. The Gaussian Processes method
 281 was applied to Landsat and achieved an R^2 -score of 0.36 for NDVI anomaly at
 282 a lead time of 4 weeks, and an R^2 -score of 0.94 for the VCI3M. Importantly,
 283 both methods showed high sensitivity and specificity for prediction of VCI val-
 284 ues indicative of drought, at lead times of 2, 4 and 6 weeks (see Fig. ??). We
 285 have presented results at the level of livelihood zone and county intersections,
 286 however both methods can be applied at any suitable spatial unit (e.g., grazing
 287 units) due to the high spatial resolution of both satellite datasets.

288 Both methods constitute novel analyses of vegetation index time series. To
 289 our knowledge, this is the first time that the GP method for NDVI forecasting
 290 has been applied to large amounts of real data and used for gap-filling. We

291 have shown that GPs are a very useful addition to other methods for both these
292 purposes. Similarly, linear AR of NDVI, or of Granger causality of various
293 variables to NDVI, has not previously been explored at a temporal resolution
294 as fine as 1 week. That such substantial R^2 -scores can be achieved for NDVI
295 anomaly at a lead-time of several weeks just by using the past few observations
296 of NDVI anomaly in a linear AR model is a novel finding. Furthermore we
297 moved beyond fitting a single model, and rather fit models to segments of data,
298 repeatedly using refreshed models to forecast subsequent observations not used
299 in the model fitting (i.e. we had separation of training and testing models).

300 Droughts have many adverse effects on pastoral and agro-pastoral commu-
301 nities as they mainly rely on rainfall for food and fodder availability. In order
302 to reduce drought-related damage and losses within these communities, local,
303 national, and international stakeholders often decide to act on information pro-
304 vided by EWS which may come too late (?). Indeed, these systems tend to
305 monitor current, rather than forecast future, environmental and socio-economic
306 factors in a region, and sound the alarm when the situation is already critical.
307 Some EWS now include a qualitative assessment of future rainfall. However, a
308 meteorological or hydrological drought will not necessarily lead to agricultural
309 damage (?). To mitigate the impacts on food security and nutrition, EWS need
310 to focus on monitoring and forecasting the possible socio-economic impacts of
311 future rainfall variability on agricultural drought (?). Additionally, acting ahead
312 of a disaster instead of providing humanitarian assistance once a disaster hits
313 can save money and lives (?). The methods developed in this study allow disas-
314 ter risk managers to estimate vegetation condition to access resources and limit
315 the impacts for pastoralist communities up to six weeks ahead. For example,
316 in Kenya, the emergency funds that are linked to the VCI could be accessed
317 earlier to launch livestock destocking and vaccination campaigns. Future work

³¹⁸ should focus on methods that forecast socio-economic drought indicators such
³¹⁹ as livestock mortality, milk production, or food prices.

³²⁰ Droughts are complex and hence inherently difficult to define and measure
³²¹ (?). A large number of satellite-based indicators have been developed to iden-
³²² tify meteorological, hydrological, and agricultural droughts (??) with each per-
³²³ forming well in space and time to a certain degree (?). While its limitations are
³²⁴ known, the VCI used in this study has been introduced as one of the main bio-
³²⁵ physical indicators in the drought early warning system operated by the NDMA,
³²⁶ with specific thresholds to identify different levels of drought throughout the
³²⁷ ASAL regions of Kenya (?). In future, we suggest that the performance of this
³²⁸ indicator together with the thresholds used should be linked to ground-based
³²⁹ measurements over various agro-ecological zones.

³³⁰ Droughts have devastating impacts on many people around the world. There
³³¹ are increasing efforts to develop tools and identify actions to save lives and liveli-
³³² hoods before these disasters strike. The methods developed in this study can
³³³ help policy makers, disaster risk managers and other key stakeholders to under-
³³⁴ stand up to six weeks in advance the state of vegetation in pastoral areas. This
³³⁵ will allow them to access resources and develop procedures before the impacts of
³³⁶ drought become visible to mitigate the adverse effects in these vulnerable com-
³³⁷ munities. To further strengthen EWS, future research needs to clearly identify
³³⁸ satellite-based indicators and thresholds of drought (which may vary in time
³³⁹ and space), to build a relation between observable indices and future impacts.
³⁴⁰ More work is also needed to understand how a hazard (e.g., reduced rainfall)
³⁴¹ becomes a disaster (e.g., food insecurity) so that these events can be better
³⁴² forecasted.

³⁴³ **Authors responsibilities**

³⁴⁴ A.B.B., S.D. and E.S. are lead authors as they contributed equally to the
³⁴⁵ paper and the order of the three names is alphabetical. A.B.B was responsible
³⁴⁶ for the Linear AR and Granger causality calculations, and the text describing
³⁴⁷ those methods. S.D. was responsible for the GPs used in the paper and was
³⁴⁸ responsible for the usage of all the Landsat data and the text describing those
³⁴⁹ methods. E.S. was responsible for the MODIS data accumulation, creation of
³⁵⁰ the MODIS time series and the filtering of the MODIS data and the text describ-
³⁵¹ ing those methods. SO and PR developed the initial idea and provided feedback
³⁵² throughout. All authors wrote, reviewed and edited the final manuscript. We
³⁵³ acknowledge early contributions to pilot work from Peter Hurley, Philip Rooney,
³⁵⁴ Martin Jung, and Jörn Scharlemann.

³⁵⁵ **Acknowledgements**

³⁵⁶ This research was funded by the STFC through the following projects: “As-
³⁵⁷ troCast: Applying Astronomy Data Analysis to enhance disaster forecasting”
³⁵⁸ – grant number ST/R004811/1; and “STFC Official Development Assistance
³⁵⁹ (ODA) Institutional Award” attached to the same grant and; ”A UK-Africa
³⁶⁰ Data Science Network: Capturing the SKA-Driven Data Transformation” grant
³⁶¹ number ST/R001898/1. This project was initiated through pump-priming fund-
³⁶² ing from the University of Sussex’s “Sussex Research” thematic programme and
³⁶³ carried out as part of the interdisciplinary Data Intensive Science Centre at the
³⁶⁴ University of Sussex (DISCUS)

³⁶⁵ **Appendix A. Comparison of the two datasets**

³⁶⁶ The key differences between the two datasets are the spatial and temporal
³⁶⁷ resolutions, see Table ???. The Landsat data had higher spatial resolution, whilst

Table A.4: Table comparing Landsat and MODIS products

Feature	Landsat	MODIS
Spatial Resolution	High resolution at 30 m	Medium resolution ranging from 250 m to 1 km
Temporal Resolution	16-day sampling (8-day when both Landsat-7 and 8 are used)	Daily sampling monitoring dynamic variables
Quality	Cloud coverage at 30 m	Cloud coverage at 500 m

368 the MODIS data had higher temporal resolution. Since forecasting was being
 369 attempted at the level of large scale regions (livelihood zone and county inter-
 370 sections), and at a weekly temporal resolution, the expectation was that the
 371 MODIS data would have advantages, assuming individual Landsat and MODIS
 372 observations have similar signal-to-noise ratios. The processed MODIS time
 373 series with weekly observations have less measurement noise because they are
 374 composites of 7 daily observations ([that themselves are 16-day composites of](#)
 375 [measurements taken every 1-2 days](#)), whereas the processed Landsat time series
 376 are derived from more temporally sparse data (up to 3 different Landsat mis-
 377 sions, each yielding one observation every 16 days). Landsat data would have
 378 advantages in different applications where forecasts on smaller spatial scales are
 379 required. The Landsat data also has the advantage that the quality flags and
 380 cloud masks are defined on smaller scales.

381 The differences between the MODIS and Landsat datasets produced slightly
 382 different “True” aggregate time series on which to assess the interpolation and
 383 forecasting methods. In addition to the different temporal resolution of the
 384 observations supplying the final time series, the MODIS data were aggregated
 385 across 100 random grassland pixels from each region, whereas the 1 000 Landsat
 386 pixels analysed were randomly distributed over the whole of each region. [In](#)
 387 [choosing how many pixels to analyse per region, there is a trade-off between](#)
 388 [using a larger number of pixels for higher accuracy, and a smaller number of](#)
 389 [pixels for lower computational cost.](#) Fewer MODIS pixels were used than Land-

sat pixels since they correspond to larger spatial regions. Both these choices of number of pixels should be sufficient for high accuracy of results, since for Landsat data the R^2 -score comparing the average of all pixels from a region with the average of 100 or 1 000 random pixels was 0.990 and 0.9993 respectively. The MODIS grassland classification was not available at Landsat resolution, thus unambiguous classification of the smaller Landsat pixels was not possible. This is unlikely to have made much difference to pixel selection, given that the pastoral livelihood zones are mostly grasslands (Fig. ??).

Appendix B. Further details on methods

Appendix B.1. Gaussian process modelling

A Gaussian Process is a probabilistic model defined as a collection of random variables for which any finite subset has a joint Gaussian distribution (?).

Formally, for an output y and inputs \mathbf{x} :

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (\text{B.1})$$

$$y_i = f(\mathbf{x}_i) + \sigma_r, \quad (\text{B.2})$$

where the mean function $m(\mathbf{x})$ represents the expectation $E[f(\mathbf{x})]$ and the kernel function $k(\mathbf{x}, \mathbf{x}')$ defines the covariances $\text{cov}(f(\mathbf{x}), f(\mathbf{x}'))$, which specifies how similar outputs $f(x)$ and $f(x')$ are. The sample drawn from $f(\mathbf{x})$ at locations $\mathbf{x}_{n=1}^N$ follow a joint multivariate Gaussian distribution with covariance matrix determined by the kernel function. Here, the \mathbf{x}_i are the dates, t_i , of the observations, and y_i is the NDVI at time t_i , subject to measurement error σ_r . To interpolate, the existing data were used to fit the mean m and the kernel k , and then the GP with the fitted mean and kernel provided an estimate of the probability distribution for the missing data.

The mean $m(t)$ was set to be constantly the mean of the whole time series.

To determine the kernel, compositional Kernel Search (?) was used to determine the best kernel combination, by calculating the maximum evidence⁴ (Marginal likelihood) for any product or sum of two common kernel combinations (Linear, Radial Basis Function, Periodic, Rational Quadratic and Matern). The kernel with the highest evidence for the Landsat NDVI time series was the Radial Basis Function (RBF) in addition to the Periodic kernel ($k_{RBF} + k_P$), where the period p was set to one year:

$$k_{RBF}(t, t') = \sigma_{RBF}^2 \exp\left(-0.5 \frac{|t - t'|^2}{l_{RBF}^2}\right), \quad (\text{B.3})$$

$$k_P(t, t') = \sigma_P^2 \exp\left(-2 \frac{\sin^2(\pi|t - t'|/p)}{l_P^2}\right). \quad (\text{B.4})$$

409 This GP contains 5 parameters to be fit ($\sigma_r, \sigma_{RBF}, l_{RBF}, \sigma_P, l_P$), which were
410 learned using Stochastic Variational Inference (SVI). The code was written using
411 the Deep Universal Probabilistic Programming language from **Pyro**, which is
412 written in **Python** and supported by **PyTorch**.

413 Gridding and gap-filling on Landsat data was carried out by fitting the
414 above-described GP model to individual pixel time series, and then using the
415 model to output an expected NDVI value every Saturday from 1/1/2000 until
416 2/2/2019. Note that the processed, gap-filled, time series have distinct obser-
417 vation dates from the raw data, which consisted of observations from up to 3
418 different Landsat missions, each at 16 day temporal resolution. (Technically
419 this means that when forecasting with this method, for a given lead time, the
420 actual time between the last observation and the forecasted observation will
421 vary from week to week.) Note also that Landsat data is acquired in tiles that

⁴In practice the evidence lower bound (ELBO) was used instead of the evidence (p_σ) with $\log(p_\sigma) \geq \text{ELBO}$

422 slightly overlap. Where there was overlap, both observations were retained,
423 since the GP modelling can handle multiple observations **with the same time-**
424 **stamp**. Smoothing was not required after gap-filling of the Landsat data with
425 GPs, since the kernel function used already enforces a smooth function.

426 For GP forecasting of NDVI anomaly and VCI3M, the GP model was fit to
427 these respective time series, but without a periodic component, since seasonal
428 periodicity is not present in these indices. The model was then used to output
429 **expected values at 1 to 10 weeks from the date of forecast**. Forecasting was car-
430 ried out starting from 1/1/2014, which was when Landsat-8 became operational.
431 This choice gave the GPs sufficient time to train (1/1/2000 to 31/12/2013).

432 *Appendix B.2. Gaps in processed MODIS data and forecast*

433 Interpolation of gaps in the raw MODIS time series was not carried out when
434 the length of the gap was longer than a certain maximum, L_{\max} . In choosing
435 L_{\max} , a trade off between quality and quantity of remaining observations had
436 to be made. The choice $L_{\max} = 6$ was made, after exploring a range of values
437 and finding results to be not sensitive to the precise choice within the range
438 between 4 and 8, see Table ???. This meant that all interpolated observations
439 were no more than 3 weeks distant from a real observation, which is within the
440 range for which interpolation can be assumed to be reasonably accurate, given
441 the forecasting results found. Note that interpolation on the Landsat data was
442 carried out for all gaps, since GP the interpolation method makes use of the
443 entire time series, and interpolated values within a long interpolation take values
444 close to the seasonal mean.

445 Due to the presence of non-interpolated gaps in the MODIS time series, there
446 were weeks when a forecast assessment was not carried out on these data. The
447 criteria for being able to do AR forecasting on these data were: (i) the three
448 most recent weekly aggregated observations had to be present, since these are

Table B.5: Comparison of outcomes for different choices of maximum allowed interpolation length L_{\max} on the MODIS data. Percentage standard deviation remaining, S , at 4 weeks, and the percentage of the time that it was possible to make a forecast, for $L_{\max} = 4, 6$, and 8. Numbers show the median across all regions.

L_{\max}	S at 4 weeks	Forecasts attempted (%)
4	63	84
6	65	93
8	61	98

449 required for making a prediction; (ii) there had to be an aggregated observation
 450 present for the week being forecast, so the quality of the prediction could be
 451 assessed.⁵

452 *Appendix B.2.1. Comparison of other possible gap-filling methods*

453 Various gap-filling methods have been used to deal with missing values re-
 454 sulting from the presence of clouds and atmospheric aerosols. These methods
 455 are based on one of three approaches, these include the use of spatial infor-
 456 mation, temporal information within time series and a combination of both
 457 spatial and temporal (spatio-temporal) information for interpolation(?) There
 458 is a choice of methods for gap-filling (??), and these fall into the categories of
 459 temporal interpolation and spatial interpolation. Temporal interpolation was
 460 chosen given that spatial interpolation methods suffer from the fact that there
 461 are frequently clouds over Kenya that cover large groups of neighbouring pixels
 462 (although a possible alternative, not considered here, would be to make use of
 463 other pixels that historically behave similarly in time (?)).

464 The performance of the temporal gap-filling methods employed, compared
 465 with alternative temporal gap-filling methods, was tested by removing observa-
 466 tions, applying the method, and then comparing the interpolated observations
 467 with the removed observations. GP interpolation and linear, quadratic and

⁵GP forecasting was still possible when (i) failed, but was also not carried out in that case, since performance would have been worse than usual in this case.

Table B.6: Comparison of GP method with commonly used interpolation methods as candidates for gap-filling on Landsat data. At the pixel level a random observation was removed, and then interpolated with each of the listed methods.

Method	R^2 -score
GP	0.67
Linear	0.53
Quadratic	-0.07
Cubic	-1.92
Last value	0.34
Mean value	0.0

⁴⁶⁸ cubic polynomial interpolation methods were tested, on both the Landsat and
⁴⁶⁹ MODIS datasets. R^2 -scores were obtained for using the interpolated values to
⁴⁷⁰ predict the “true” values for the missing observations.

⁴⁷¹ For the Landsat data, one randomly chosen [observation](#) between 1/1/2014
⁴⁷² and 1/2/2019 was removed from [each of](#) 2000 randomly selected individual pixel
⁴⁷³ time series.⁶ From the MODIS data, 1200 random [individual pixel](#) NDVI time
⁴⁷⁴ series (1/1/2014 to 1/2/2019) were [chosen](#). 20 randomly selected NDVI values
⁴⁷⁵ were dropped from each of the time series and the various gap-filling methods
⁴⁷⁶ were used to interpolate the dropped values. The results for Landsat are shown
⁴⁷⁷ in Table ?? and for MODIS in Table ??.

⁴⁷⁸ For the Landsat data, the GP method achieved the highest R^2 -score, thus
⁴⁷⁹ showing its utility, and justifying our choosing it. The R^2 -score of 0.67, achieved
⁴⁸⁰ by the GP method, is close to the R^2 -score of 0.76 which is obtained from using
⁴⁸¹ one Landsat observation to predict another Landsat observation of the same
⁴⁸² pixel on the same day (see Section ??). For interpolation the linear method was
⁴⁸³ also somewhat effective, achieving an R^2 -score of 0.53.

⁴⁸⁴ For the MODIS data, GP, linear interpolation and quadratic interpolation

⁶We remove the mean of the individual NDVI time series for every single observed and interpolated datum before calculating the R^2 -scores. This avoids an over-estimate of the denominator (see Equation ??) due to the variance from different regions in Kenya. This also forces the mean value prediction to be zero, which it should be for a R^2 calculation.

Table B.7: Comparison of interpolation methods as candidates for gap-filling on MODIS data.

Method	R^2 -score
GP	0.92
Linear	0.93
Quadratic	0.94
Cubic	0.92
Last value	0.70
Mean value	-0.02

485 all performed similarly well. Quadratic interpolation had the highest R^2 -score,
 486 hence this method was chosen for gap-filling on the MODIS data. The higher in-
 487 terpolation R^2 -scores for MODIS, compared to Landsat, imply that the MODIS
 488 data is less noisy than the Landsat data. Assuming that observations from
 489 MODIS and Landsat have similar signal-to-noise ratio, this can be explained
 490 by the higher temporal resolution of MODIS, and the compositing of multiple
 491 observations for the weekly gridded MODIS data.

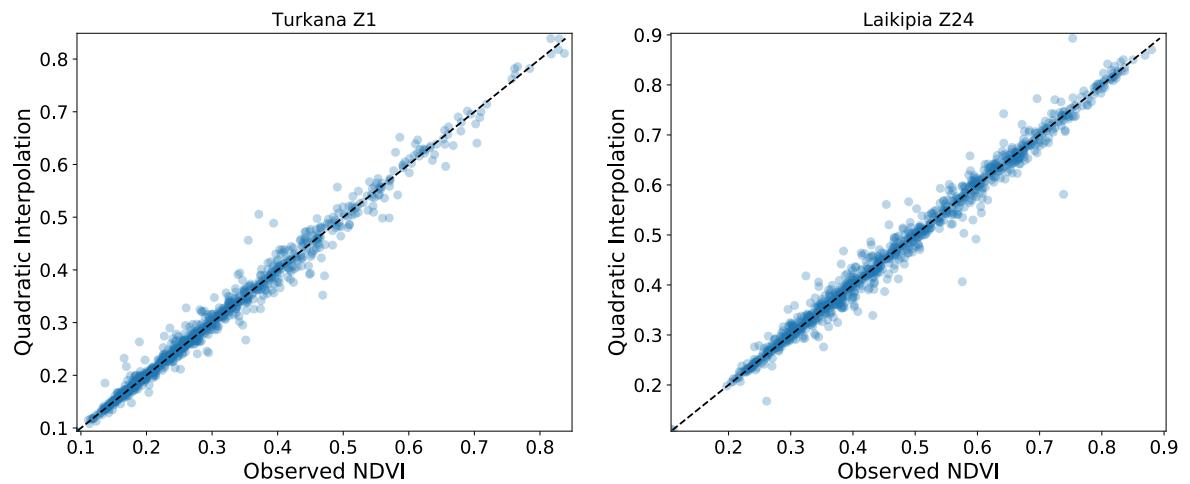


Figure C.8: Scatter plot of MODIS observed and predicted NDVI values for the Quadratic Interpolation gap-filling method for Turkana Zone 1 and Laikipia Zone 24

492 **Appendix C. Interpolation**

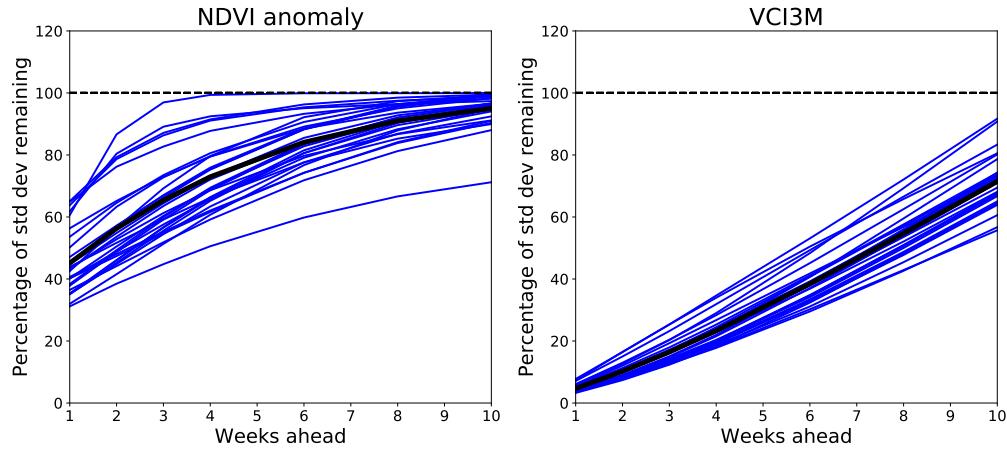


Figure D.9: Forecast performance with a lead time of 1 to 10 weeks using the GP method on the Landsat data, as given by percentage standard deviation remaining S , for (Left) NDVI anomaly, and (Right) VCI3M. The blue lines show results for the individual regions (county/livelihood zone intersections), and the black line shows the median across all regions.

493 **Appendix D. Forecast**

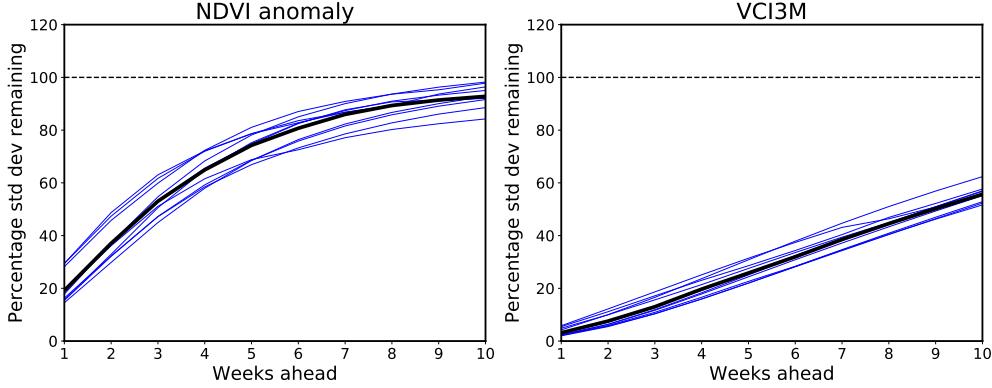


Figure D.10: Forecast performance with a lead time of 1 to 10 weeks using the AR method on the MODIS data, as given by percentage standard deviation remaining, for (Left) NDVI anomaly, and (Right) VCI3M. The blue lines show results for the individual regions for which a forecast is possible more than 50% of the time, and the black line shows the median across all 19 of these regions.

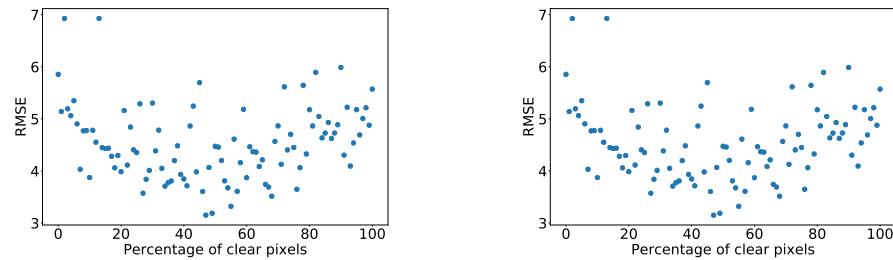


Figure D.11: TO DO: STEVEN/EDWARD REPLACE LEFT HAND PLOTS WITH THE CORRESPONDING PLOTS FROM LANDSAT/GP. ADAM: REFERENCE TO THIS FROM MAIN TEXT. RMSE of 4 week forecast against percentage of clear pixels at most recent observation. (Left) Landsat/GP (Right) MODIS/AR, for which the Pearson correlation is 0.01.

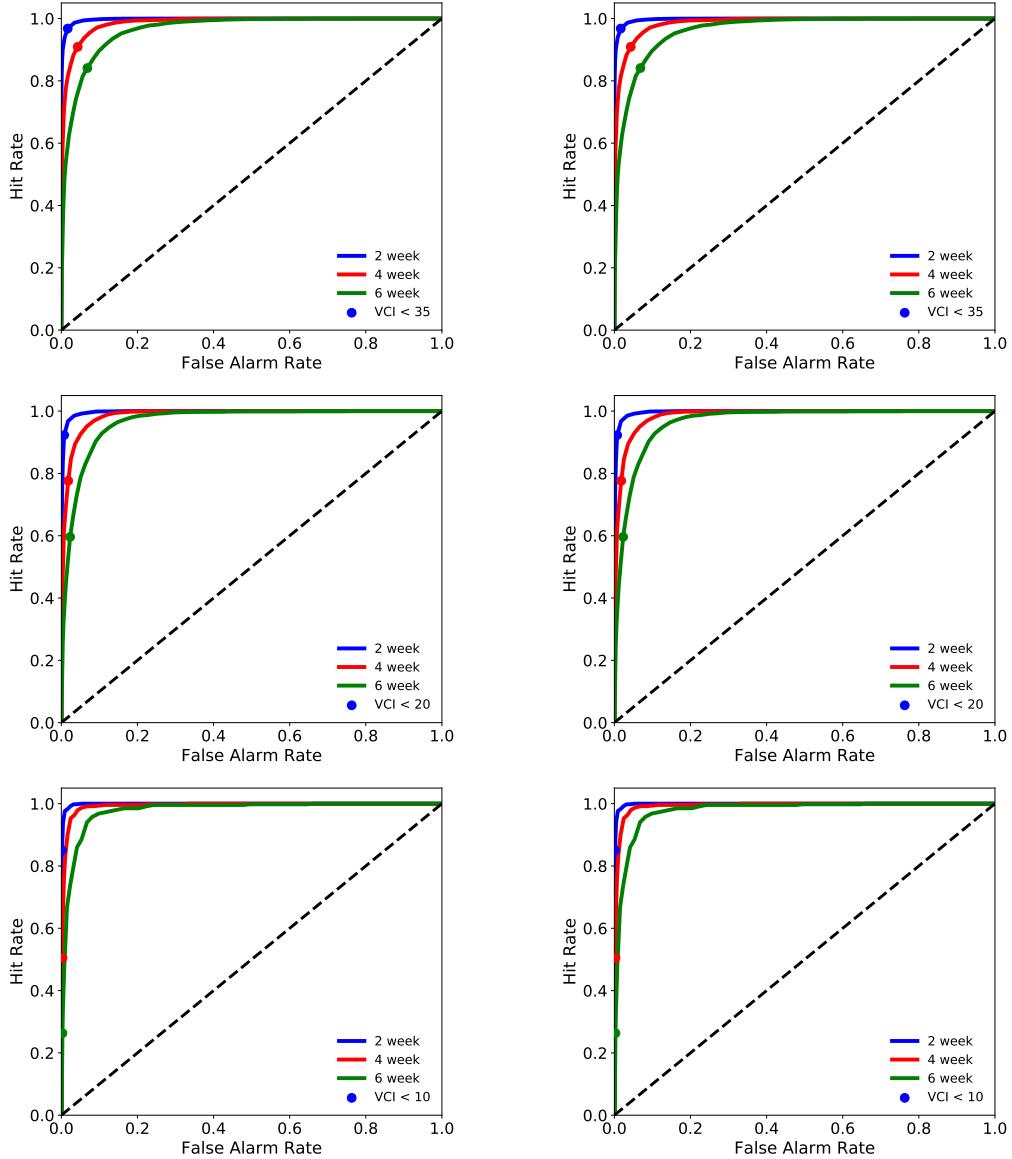


Figure D.12: TO DO: ADAM REPLACE LEFT HAND PLOTS WITH THE CORRESPONDING PLOTS FROM LANDSAT/GP. ROC curves for predicting drought with drought defined at various NDMA thresholds. For (Left) Landsat/GP (Right) MODIS/AR: (Top) Any drought, $VCI_{3M} < 35$, (Middle) Severe or extreme drought $VCI_{3M} < 20$, (Bottom) Extreme drought $VCI_{3M} < 10$.

Table D.8: NDVI anomaly forecast using Landsat data for the 29 regions. The numbers shown are the proportion of standard deviation remaining (Equation ??) and the R^2 -score for NDVI anomaly. We only used past data for the interpolation and we the average value for every pixel within the region for the region estimate. The * indicates regions where a minimum of 180 detections per pixel were used, instead of 250.

Region	2 weeks	4 weeks	6 weeks
Baringo Z24	46 0.74	66 0.46	81 0.19
Elgeyo-Marakwet Z24	49 0.74	69 0.47	83 0.22
Garissa Z10*	55 0.64	73 0.36	86 0.12
Garissa Z11*	63 0.58	80 0.33	89 0.16
Isiolo Z5	57 0.64	75 0.37	88 0.13
Isiolo Z9	65 0.53	79 0.30	89 0.13
Isiolo Z10	79 0.31	91 0.08	96 -0.02
Isiolo Z24	57 0.67	76 0.41	89 0.19
Kajiado Z15	45 0.76	63 0.52	78 0.28
Kajiado Z18*	44 0.75	59 0.55	72 0.34
Laikipia Z24	42 0.82	61 0.62	77 0.39
Lamu Z11*	80 0.33	92 0.12	95 0.07
Mandera Z7	76 0.25	88 0.00	93 -0.13
Mandera Z9	53 0.44	69 0.06	80 -0.27
Marsabit Z5	52 0.60	66 0.35	78 0.11
Marsabit Z7*	47 0.76	62 0.59	74 0.40
Narok Z15	56 0.67	80 0.34	92 0.12
Narok Z18	56 0.68	75 0.42	88 0.20
Samburu Z5	49 0.68	69 0.36	84 0.08
Samburu Z24	45 0.78	65 0.54	81 0.30
Tana River Z11*	65 0.57	81 0.33	91 0.15
Turkana Z1	54 0.56	72 0.21	84 -0.09
Turkana Z3	38 0.61	51 0.33	60 0.06
Turkana Z24	46 0.75	66 0.48	81 0.21
Wajir Z7*	48 0.71	62 0.51	74 0.30
Wajir Z9	79 0.20	91 -0.07	95 -0.18
Wajir Z10	87 0.24	99 0.01	100 0.00
WestPokot Z1	50 0.69	69 0.43	83 0.18
WestPokot Z24	49 0.68	66 0.42	79 0.16
Median	53 0.67	69 0.36	84 0.15

⁴⁹⁴ Appendix D.1. Tables of NDVI and VCI3M forecast

Table D.9: VCI3M forecast performance using GPs on the Landsat data. The numbers shown are the percentage standard deviation remaining and the R^2 score, respectively.

Region	2 weeks	4 weeks	6 weeks
Baringo Z24	9 0.99	22 0.95	38 0.86
Elgeyo-Marakwet Z24	9 0.99	21 0.96	36 0.87
Garissa Z10	10 0.99	23 0.95	39 0.85
Garissa Z11	11 0.99	25 0.94	41 0.83
Isiolo Z5	10 0.99	23 0.95	39 0.85
Isiolo Z9	11 0.99	24 0.94	38 0.85
Isiolo Z10	13 0.98	29 0.92	46 0.79
Isiolo Z24	10 0.99	23 0.95	39 0.85
Kajiado Z15	9 0.99	21 0.96	36 0.87
Kajiado Z18	9 0.99	20 0.96	34 0.88
Laikipia Z24	7 0.99	18 0.97	32 0.89
Lamu Z11	13 0.98	29 0.92	45 0.80
Mandera Z7	15 0.98	32 0.90	49 0.76
Mandera Z9	12 0.98	29 0.92	48 0.77
Marsabit Z5	11 0.99	25 0.94	41 0.83
Marsabit Z7	8 0.99	19 0.96	32 0.90
Narok Z15	10 0.99	25 0.94	41 0.83
Narok Z18	11 0.99	24 0.94	40 0.84
Samburu Z5	10 0.99	24 0.94	42 0.83
Samburu Z24	8 0.99	20 0.96	35 0.88
TanaRiver Z11	11 0.99	24 0.94	40 0.84
Turkana Z1	14 0.98	31 0.90	52 0.73
Turkana Z3	12 0.99	26 0.93	43 0.81
Turkana Z24	9 0.99	22 0.95	38 0.85
Wajir Z7	9 0.99	20 0.96	34 0.88
Wajir Z9	16 0.97	35 0.88	53 0.72
Wajir Z10	17 0.97	35 0.88	52 0.73
WestPokot Z1	9 0.99	23 0.95	39 0.85
WestPokot Z24	10 0.99	22 0.95	38 0.85
Median	10 0.99	24 0.94	39 0.85

Table D.10: VCI3M forecast performance using AR on the MODIS data. The numbers shown are the percentage standard deviation remaining and the R^2 score, respectively. In the 'Forecasts' column, the number gives the percentage of time points for which it was possible to obtain a forecast.

Region	2 weeks	4 weeks	6 weeks	Forecasts
Baringo Z24	6 0.99	18 0.96	32 0.89	100
Elgeyo-Marakwet Z24	6 0.99	18 0.96	32 0.89	95
Garissa Z10	11 0.98	20 0.95	53 0.71	15
Garissa Z11	xx xx	xx xx	xx xx	0
Isiolo Z5	12 0.98	27 0.92	42 0.81	97
Isiolo Z9	13 0.98	28 0.91	43 0.81	89
Isiolo Z10	13 0.98	28 0.91	42 0.81	71
Isiolo Z24	8 0.99	22 0.95	37 0.86	93
Kajiado Z15	12 0.98	26 0.92	39 0.84	75
Kajiado Z18	11 0.98	24 0.94	38 0.85	71
Laikipia Z24	9 0.99	23 0.94	37 0.85	93
Lamu Z11	xx xx	xx xx	xx xx	0
Mandera Z7	15 0.97	34 0.87	53 0.71	44
Mandera Z9	16 0.97	35 0.87	55 0.69	43
Marsabit Z5	9 0.99	21 0.95	34 0.88	94
Marsabit Z7	17 0.96	35 0.87	48 0.76	34
Narok Z15	11 0.98	28 0.92	44 0.79	96
Narok Z18	6 0.99	19 0.96	32 0.89	98
Samburu Z24	5 0.99	16 0.97	30 0.90	100
Samburu Z5	7 0.99	21 0.95	37 0.86	100
Tana River Z11	11 0.98	22 0.95	33 0.88	41
Turkana Z1	7 0.99	20 0.95	35 0.87	98
Turkana Z3	7 0.99	23 0.94	40 0.83	100
Turkana Z24	7 0.99	20 0.95	35 0.87	100
Wajir Z7	14 0.97	30 0.90	44 0.79	45
Wajir Z9	16 0.97	33 0.88	50 0.74	41
Wajir Z10	12 0.98	23 0.94	33 0.88	19
West Pokot Z1	7 0.99	21 0.95	38 0.85	100
West Pokot Z24	9 0.99	22 0.94	36 0.86	94
Median	11 0.99	23 0.95	38 0.85	93

Table D.11: NDVI anomaly forecast performance using AR on the MODIS data. The numbers shown are the percentage standard deviation remaining and the R^2 score, respectively.

Region	2 weeks	4 weeks	6 weeks
Baringo Z24	32 0.90	59 0.65	75 0.42
Elgeyo-Marakwet Z24	32 0.90	58 0.66	73 0.46
Garissa Z10	42 0.82	51 0.74	56 0.68
Garissa Z11	xx xx	xx xx	xx xx
Isiolo Z5	55 0.69	79 0.37	90 0.18
Isiolo Z9	36 0.87	64 0.58	80 0.35
Isiolo Z10	37 0.86	65 0.57	82 0.32
Isiolo Z24	29 0.91	57 0.66	76 0.42
Kajiado Z15	47 0.78	71 0.48	82 0.31
Kajiado Z18	45 0.79	72 0.48	87 0.24
Laikipia Z24	38 0.85	62 0.60	77 0.40
Lamu Z11	xx xx	xx xx	xx xx
Mandera Z7	33 0.89	64 0.58	87 0.24
Mandera Z9	32 0.89	65 0.57	90 0.18
Marsabit Z5	38 0.85	64 0.59	75 0.44
Marsabit Z7	35 0.88	58 0.66	71 0.49
Narok Z15	48 0.76	72 0.48	83 0.30
Narok Z18	36 0.87	61 0.62	72 0.47
Samburu Z24	28 0.92	56 0.68	73 0.46
Samburu Z5	47 0.78	74 0.45	86 0.25
Tana River Z11	53 0.71	68 0.54	87 0.24
Turkana Z1	32 0.89	64 0.58	82 0.32
Turkana Z3	33 0.89	71 0.49	88 0.21
Turkana Z24	31 0.90	62 0.61	80 0.35
Wajir Z7	29 0.91	54 0.70	67 0.54
Wajir Z9	30 0.91	57 0.67	74 0.44
Wajir Z10	26 0.93	37 0.86	46 0.79
West Pokot Z1	37 0.86	68 0.53	85 0.28
West Pokot Z24	42 0.82	65 0.57	78 0.39
Median	36 0.87	65 0.58	80 0.35

Table D.12: NDVI anomaly forecast performance using GPs on the MODIS data. The numbers shown are the percentage standard deviation remaining and the R^2 score, respectively.

Region	2 weeks	4 weeks	6 weeks
Baringo Z24	37 0.86	73 0.44	93 0.10
Elgeyo-Marakwet Z24	37 0.85	73 0.40	91 0.07
Garissa Z10	46 0.75	59 0.50	68 0.21
Garissa Z11	xx xx	xx xx	xx xx
Isiolo Z5	60 0.60	87 0.16	96 0.02
Isiolo Z9	42 0.83	79 0.38	97 0.08
Isiolo Z10	41 0.82	77 0.36	95 0.05
Isiolo Z24	34 0.89	70 0.51	92 0.15
Kajiado Z15	48 0.75	76 0.35	90 0.09
Kajiado Z18	47 0.78	77 0.40	94 0.11
Laikipia Z24	43 0.81	72 0.46	89 0.16
Lamu Z11	xx xx	xx xx	xx xx
Mandera Z7	39 0.86	74 0.50	97 0.17
Mandera Z9	35 0.89	71 0.54	96 0.20
Marsabit Z5	42 0.79	69 0.37	80 0.11
Marsabit Z7	37 0.80	61 0.38	74 0.09
Narok Z15	50 0.72	76 0.32	87 0.09
Narok Z18	39 0.84	74 0.42	90 0.11
Samburu Z5	50 0.70	80 0.26	91 0.05
Samburu Z24	34 0.88	71 0.48	92 0.12
TanaRiver Z11	60 0.47	75 0.13	87 -0.00
Turkana Z1	36 0.87	75 0.44	95 0.10
Turkana Z3	35 0.88	80 0.37	99 0.03
Turkana Z24	35 0.87	75 0.42	95 0.07
Wajir Z7	31 0.86	57 0.45	72 0.15
Wajir Z9	36 0.85	65 0.45	79 0.15
Wajir Z10	37 0.77	53 0.37	68 -0.17
WestPokot Z1	41 0.83	77 0.39	96 0.08
WestPokot Z24	49 0.78	86 0.34	103 0.06
Median	39 0.83	74 0.40	92 0.09

Table D.13: NDVI anomaly forecast performance using AR on Landsat data. The numbers shown are the percentage standard deviation remaining and the R^2 score, respectively.

Region	2 weeks	4 weeks	6 weeks
Baringo Z24	67 0.54	89 0.19	105 -0.12
Elgeyo-Marakwet Z24	74 0.44	98 0.03	113 -0.28
Garissa Z10	77 0.39	96 0.07	108 -0.17
Garissa Z11	78 0.38	94 0.10	106 -0.13
Isiolo Z5	86 0.25	103 -0.06	114 -0.31
Isiolo Z9	96 0.06	109 -0.20	116 -0.35
Isiolo Z10	108 -0.17	120 -0.46	126 -0.60
Isiolo Z24	66 0.55	82 0.31	92 0.13
Kajiado Z15	59 0.64	78 0.38	91 0.15
Kajiado Z18	60 0.63	76 0.40	90 0.18
Laikipia Z24	54 0.70	75 0.42	93 0.12
Lamu Z11	88 0.20	97 0.05	103 -0.07
Mandera Z7	78 0.38	87 0.23	93 0.12
Mandera Z9	56 0.67	69 0.51	76 0.41
Marsabit Z5	76 0.41	91 0.15	103 -0.07
Marsabit Z7	52 0.72	66 0.55	78 0.38
Narok Z15	79 0.36	101 -0.02	109 -0.19
Narok Z18	73 0.46	91 0.15	101 -0.02
Samburu Z5	74 0.44	95 0.08	110 -0.21
Samburu Z24	60 0.63	81 0.32	97 0.05
TanaRiver Z11	90 0.17	108 -0.18	120 -0.46
Turkana Z1	80 0.35	103 -0.07	118 -0.39
Turkana Z3	75 0.43	95 0.09	111 -0.24
Turkana Z24	69 0.52	91 0.16	106 -0.14
Wajir Z7	53 0.70	63 0.59	70 0.50
Wajir Z9	77 0.40	85 0.27	84 0.27
Wajir Z10	100 0.00	111 -0.24	117 -0.37
WestPokot Z1	70 0.50	92 0.15	107 -0.15
WestPokot Z24	73 0.45	96 0.06	114 -0.31
Median	74 0.44	92 0.15	106 -0.13

495 *Appendix D.2. Assessing forecasting with the inclusion of additional variables*

For the MODIS data, we tested to see whether we could improve the prediction of NDVI anomaly by including the past of other available variables in the AR model, i.e. we performed a Granger causality analysis. Taking X as NDVI anomaly, as in equation (??), for another variable Y , the extended model was fit:

$$X_{t+n} = \sum_{i=0}^{p-1} a_i X_{t-i} + \sum_{i=0}^{q-1} b_i Y_{t-i} + \epsilon'_t, \quad (\text{D.1})$$

496 and Granger causality measured as ΔR^2 , the R^2 -score obtained from this ex-
497 tended model minus the R^2 -score obtained from the previous (reduced) model
498 (??).

499 Firstly, we tested whether including past observations of either the red band
500 or the NIR band (at the same lags as NDVI anomaly) in the regression to
501 predict NDVI anomaly could improve the quality of the forecast, and found it
502 did not. For a lead time of 4 weeks, for example, the improvement in R^2 -score
503 was generally negative; the mean improvement across regions was -0.007 for red
504 and -0.01 for NIR.

505 Secondly, we tested for Granger causality of NDVI anomaly from each region
506 to each other region (within the set of regions for which predictions could be
507 made more than 50% of the time). That is, for each pair of distinct regions, i and
508 j , the 3 most recent observations from region j were added to the AR forecast
509 model for region i , and the R^2 -score was compared with that obtained without
510 including observations from region j . There was not strong Granger causality
511 of NDVI anomaly between most regions. For only a few combinations was there
512 an improvement in R^2 -score of more than 0.05, see Fig. ???. Nevertheless, these
513 results suggest that, to create the optimal linear regression based forecasting
514 method, data from all regions should be used. Future work will explore how
515 best to extract any useful information from regions other than the one being

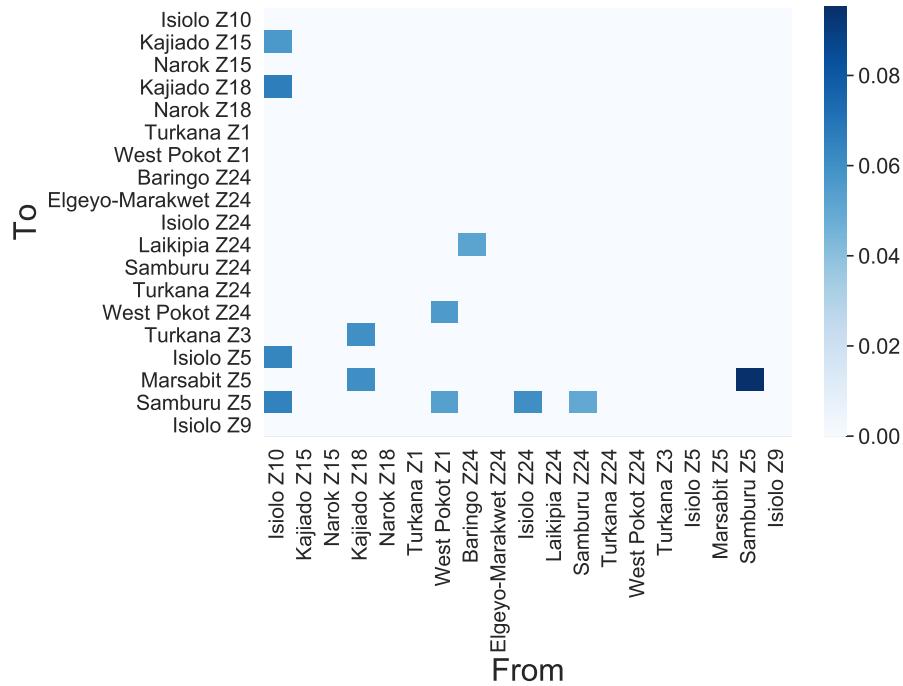


Figure D.13: Granger causality of NDVI anomaly from each region to each other region, computed on the MODIS data, measured as improvement in R^2 -score when observations from region ‘From’ are added to the AR model for forecasting region ‘To’ at a lead time of 4 weeks. Only substantial Granger causalities are shown, i.e. those with $\Delta R^2 > 0.05$.

516 forecast.

517 **References**