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

Article (Accepted Version)

Barrett, Adam B, Duivenvoorden, Steven, Salakpi, Edward E, Muthoka, James M, Mwangi, John, Oliver, Seb and Rowhani, Pedram (2020) Forecasting vegetation condition for drought early warning systems in pastoral communities in Kenya. *Remote Sensing of Environment*, 248. a111886 1-10. ISSN 0034-4257

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/91240/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

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

Adam B. Barrett^{a,b}, Steven Duivenvoorden^{a,c}, Edward E. Salakpi^a, James M. Muthoka^d, John Mwangi^e, Seb Oliver^{a,c} 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*

^e *The National Drought Management Authority (NDMA), Lonrho House, Nairobi, Kenya*

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

Abstract

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, reducing the financial and human cost. However, existing EWS tend only to monitor current, rather than forecast future, environmental and socioeconomic indicators of drought, and hence are not always sufficiently timely to be effective in practice. Here we present a novel method for forecasting satellite-based indicators of vegetation condition. Specifically, we focused on the 3-month Vegetation Condition Index (VCI3M) over pastoral livelihood zones in Kenya, which is the indicator used by the Kenyan National Drought Management Authority (NDMA). Using data from MODIS and Landsat, we apply linear autoregression and Gaussian process modeling methods and demonstrate high forecasting skill several weeks ahead. As a bench mark we predicted the drought alert marker used by NDMA ($VCI3M < 35$). 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 thus identify a deteriorating vegetation condition well and sufficiently in advance to help disaster risk managers act early to support vulnerable communities and limit the impact of a drought hazard.

Keywords: Drought; Forecasting; Early Warning Systems; Disaster Risk Reduction; Landsat; MODIS

1. Introduction

Droughts are a major threat globally as they can cause substantial damage to society, especially in regions that depend on rain-fed agriculture. They particularly impact food security by significantly reducing agricultural production (Lesk et al., 2016) and raising food prices (Nelson et al., 2014; Brown and Kshirsagar, 2015), which often leads to increased levels of malnutrition, migration, disease, and other health concerns (Piguet et al., 2011; Stanke et al., 2013). The majority of droughts occur in sub-Saharan Africa (EM-DAT, 2019) where many communities rely on predictable rainfall patterns for their livelihood.

In East Africa, the main economic activity in the arid and semi-arid lands (ASAL) is subsistence rain-fed agriculture, as well as livestock farming using pastures and grasslands as the main source of fodder. As a result, the pastoral and agro-pastoral communities who live in these drylands are particularly vulnerable to drought (Nyong et al., 2007; Orindi et al., 2007), especially since their existing coping strategies have been compromised by population growth and land use change in recent years (Galvin et al., 2001). Governments and donor agencies in the region have thus developed several tools and early warning systems (EWS) to mitigate the impact of droughts on pastoralists.

Most EWS tend to monitor current key biophysical and socio-economic factors to assess the possible exposure of vulnerable people to specific hazards.

21 However, once the impacts are visible, it may be too late to mitigate the conse-
 22 quences (Kogan et al., 2013). Hence there is growing interest in moving toward
 23 a proactive humanitarian approach to disasters by developing preparedness ac-
 24 tions based on climate forecasts (Coughlan de Perez et al., 2015; Lopez et al.,
 25 2018; Wilkinson et al., 2018). Additionally, it is estimated that being better pre-
 26 pared before a drought hits significantly reduces the costs and losses from these
 27 disasters (Venton et al., 2012). Hence, EWS now increasingly include expert
 28 knowledge and qualitative assessments of seasonal climate forecasts to assess
 29 the future development of food security, and define actions to mitigate possible
 30 losses (Coughlan de Perez et al., 2015; Tozier de la Poterie and Baudoin, 2015).
 31 However for drought conditions, a meteorological drought does not always lead
 32 to negative agricultural outputs (Bhuiyan et al., 2006). There is thus a growing
 33 interest to include forecasts of the impacts of these hazards (WMO, 2015; Sai
 34 et al., 2018; Sutanto et al., 2019).

35 In Kenya, following several periods of intense drought, the government es-
 36 tablished the National Drought Management Authority (NDMA) in 2016, to
 37 set up and operate a drought EWS, as well as to establish drought prepared-
 38 ness strategies and contingency plans. The NDMA provides monthly bulletins
 39 assessing food security in the 23 ASAL regions using current biophysical (e.g.,
 40 rainfall, vegetation condition) and socio-economic (production, access, and util-
 41 isation) factors. One key biophysical indicator used by the NDMA drought
 42 phase classification is based on the Vegetation Condition Index (VCI, Kogan,
 43 1995; Klisch and Atzberger, 2016; Rulinda et al., 2011; Rojas et al., 2011).

44 The VCI, which expresses the Normalized Difference Vegetation Index (NDVI)
 45 in terms of where it currently lies within its expected range for the given pixel,
 46 is one of a number of satellite-based indicators that have been developed to de-
 47 tect and monitor drought (Zargar et al., 2011). While there is little agreement

between VCI and precipitation-based meteorological drought indicators (Quiring and Ganesh, 2010; Bhuiyan et al., 2006), it is strongly linked to agricultural production and widely used to identify drought onset, intensity, duration, and impact (Jiao et al., 2016). The NDMA uses the 3-month averaged VCI (VCI3M) in its operational EWS (Klisch and Atzberger, 2016). Once the VCI3M goes below a threshold of 35, the NDMA triggers a rapid food security assessment and has access to the National Drought Contingency Fund in order to implement its preparedness strategies and contingency plans.

The main goal of this paper is to explore machine-learning techniques to forecast the vegetation indices that are commonly used in the pastoral areas of Kenya to monitor droughts. In order to provide useful information to drought risk managers, we aim to identify the right balance between forecast lead time and uncertainty. To this end, we evaluated the performance of our approaches up to ten weeks ahead.

Based on NDMA’s experience, we particularly focused on the pastoral livelihood zones as the VCI3M is more reliable in identifying drought condition for grazing and browsing in the more arid regions of the country. Several studies have developed statistical and machine-learning approaches (Udelhoven et al., 2009; Meroni et al., 2014; Zambrano et al., 2018; Vrieling et al., 2016) to predict end-of-season crop, forage and biomass production. Recently, Matere et al. (2019) developed a decision support tool based on a mechanistic model to estimate 6-monthly forecasts of forage condition. Here, we specifically focus on Gaussian Process modelling (GP, Rasmussen and Williams, 2006), and linear autoregressive (AR) modelling (e.g. Hamilton, 1994) to forecast NDVI and VCI3M, which are derived from both Landsat (every 16 days at 30 m resolution) and the MODerate resolution Imaging Spectroradiometer (MODIS - daily data at 500 m resolution). GP modelling uses kernel-based non-parametric Bayesian

inference on the structure of correlations between observations, and is widely applied to classification, interpolation, change detection and forecasting problems (Brahim-Belhouari and Vesin, 2001; Chandola and Vatsavai, 2011; Camps-Valls et al., 2016; Upreti et al., 2019). Linear AR is the regression of future observations on past observations, assuming a linear dependence. Previously it has been performed on monthly (i.e. temporally more sparse) NDVI data, see for example Asoka and Mishra (2015) and Papagiannopoulou et al. (2017), with mixed results in terms of forecasting potential (R^2 -scores between 0 and 0.4 at a lead time of one month).

2. Study area

In Kenya, the livestock sector accounts for 13% of the national GDP and 43% of its agricultural GDP. Livestock farming mainly occurs in the ASAL which cover about 80% of the country (UNDP, 2013; FAO, 2014). In these regions, the pastoral communities rely on pastures and grasslands as the main source of fodder (Behnke and Muthami, 2011). Thus, providing information on pasture productivity to these communities is key in times of drought.

For the ASAL regions, the NDMA reports every month the VCI3M value at county level as well as over the different livelihood zones within the county. This study focused on the 10 (agro)-pastoral livelihood zones (see Fig. C.8), which cross 15 counties. The names of the 29 livelihood zone county intersections can be found in Fig. 1; these are our regions-of-interest, which we refer to simply as ‘regions’.

Figure 1: Maps of Kenya showing (a) Livelihood Zones and County intersections (Regions of Interest (ROI)) from which pixels were sampled for analysis, and (b) land-cover classification (according to the MODIS MCD12Q1 data). Analyses were performed for 29 regions, defined by pastoral livelihood zone and county intersections. A map showing the livelihood zones can be found in Fig. C.8 in the Supplementary Material.

3. Methods

This research is based on two satellite-based Earth observation datasets, Landsat and MODIS. Description and justification of data selection, and a comparison between the two datasets can be found in Supplementary Material A. It should be noted that the analysis is based on a random subsample of the pixels within each of the 29 regions (Fig. 1). A summary of the entire work from data preparation to forecasting drought can be seen in Fig. 2.

3.1. Data preprocessing

3.1.1. Landsat

Temporal gridding and gap-filling on the Landsat data was done using Gaussian Process (GP) regression. For a given pixel, the GP regression took raw data as input, fit a temporal correlation structure to the data, and used this to output a time series of expected NDVI values, with observations provided

Figure 2: A flow chart of the data processing and analysis.

Figure 3: Illustration of the GP approach used for the Landsat data. In “forecast mode”, the correlations in the data up to a given date furnish a GP model, which can then be used for forecasting. In the “non-forecast mode”, the entire time series is used to train the GP, and provide a ground truth for the forecast.

110 every Saturday over the studied time period; see Figure 3 for an illustration
 111 and Supplementary Material B.1 for details. Two versions of GP gap-filling
 112 were carried out, which we refer to as forecast mode and non-forecast mode.
 113 For the non-forecast mode, the full time series from the given pixel were used
 114 to train the GP. The non-forecasting mode was used as the “ground truth” to
 115 test forecasts against. The forecast mode, by contrast, only used data up to a
 116 certain date, whichever date a forecast was being attempted from - since when
 117 doing forecasting with a near real-time data stream, one does not have access
 118 to future data.

119 3.1.2. MODIS

120 Weekly NDVI composites were obtained for each pixel by taking the mean
 121 (after cloud masking) of all available data over a 7-day time period. Gaps
 122 in the weekly time series were then filled using quadratic interpolation. Gaps
 123 longer than 6 weeks were left unfilled, see Supplementary Material B.2 for
 124 details. The gap-filled time series were then smoothed using the Savitzky-Golay
 125 method (Savitzky and Golay, 1964) to filter high-frequency measurement noise.
 126 The smoothing involved fitting, for each pixel, a polynomial to a window centred
 127 on the observation, and then replacing that observation with the output of the
 128 polynomial fit. The polynomial order was set to 2 (i.e. quadratic function)
 129 and the window length to 7 weeks. (Note that the combined interpolation and
 130 smoothing procedure does two rounds of quadratic interpolation where there are
 131 gaps, but that these are distinct: the interpolation fills a gap of up to 6 weeks
 132 with one quadratic function, while the smoothing modifies only one observation
 133 per fitted quadratic function.)

134 3.2. Indices

On both datasets, VCI time series were constructed from the NDVI time series according to the formula:

$$\text{VCI}_i = 100 \times \frac{\text{NDVI}_i - \text{NDVI}_{\min,i}}{\text{NDVI}_{\max,i} - \text{NDVI}_{\min,i}}, \quad (1)$$

135 where $\text{NDVI}_{\min,i}$ and $\text{NDVI}_{\max,i}$ are the minimum and maximum observed val-
 136 ues for the NDVI of the pixel for the week of the year at time point i . The
 137 data within each region were aggregated taking the mean of the sampled pixels
 138 at each time point. Thus forecasting was applied on a single time series for
 139 each region. Finally, VCI3M was calculated as the mean VCI across the 12
 140 weeks leading up to the given time point. Additionally, aggregate time series of
 141 NDVI anomaly were constructed (i.e., seasonal mean-subtracted NDVI, some-
 142 times referred to as absolute anomaly; results for forecasting this can be found
 143 in Supplementary Material C).

144 With Landsat data, the mean, maximum and minimum value for the NDVI
 145 in (1) was computed using the non-forecast mode GP interpolated time series.
 146 Then forecast mode and non-forecast mode versions of each index were created.
 147 With the MODIS data, since large gaps were unfilled, whenever there were fewer
 148 than 25 individual pixel observations from a particular region at a given time,
 149 it was decided that there should be no datum in the aggregate VCI time series
 150 for that region (i.e., there should be a gap in the time series). Additionally, if
 151 the current aggregate NDVI observation was not present, a gap was placed in
 152 the VCI3M time series. Else, the mean was taken over all present observations
 153 from the most recent 12 weeks.

154 3.3. Forecasting

155 Machine-learning techniques offer a data-driven, empirical route to forecast-
 156 ing. Many different data inputs could be used to forecast these vegetation
 157 indices (e.g. precipitation and precipitation forecasts). However, perhaps the
 158 most simple is to use the past history of the indices themselves. This has the
 159 practical benefits of readily available data over large areas. Additionally, this
 160 approach will also take advantage from the fact that these indices are subject
 161 to plant growth and climate cycles giving periodic behaviour on large temporal
 162 scales that can be empirically modelled, while external perturbations, such as
 163 water availability, have persistent impact providing correlations on short tem-
 164 poral scales. Forecasts of NDVI anomaly and VCI3M were made using two
 165 separate methods, respectively based on Gaussian Process modelling (GP) and
 166 linear autoregressive (AR) modelling.

167 GP forecasting was performed by fitting a GP to the forecast mode aggregate
 168 time series for the index in question, and then using the GP to extrapolate.
 169 For details on GP modelling, see Supplementary Material B.1. The key step
 170 involved fitting a temporal correlation structure to the time series, i.e. a kernel
 171 $k(t, t')$ that describes the covariance between the index at any two times t and
 172 t' . The kernel with the highest evidence was the Radial Basis Function (RBF):

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

173 where σ_{RBF}^2 and l_{RBF} are the signal variance and the length scale, respectively,
 174 and the modelling was carried out with the best fit version of this.

175 AR forecasting was performed with the following model-fitting and extrap-

176 olation method. For forecasting n weeks ahead, the following model was fit:

$$X_{t+n} = \sum_{i=0}^{p-1} a_i X_{t-i} + \epsilon_t, \quad (3)$$

177 where X is the index in question, subscripts denote the date (week), a_i are
 178 model coefficients, ϵ_t are the residuals (i.e. the errors), and p is called the model
 179 order. (This model assumes zero mean, so for VCI3M, the mean was removed
 180 prior to fitting the model, and then added back again after using the model to
 181 forecast the deviation from the mean.) Fitting the model to a segment of data
 182 involved finding the model coefficients that gave the minimum sum-square error,
 183 i.e. led to residuals with the minimum variance. To make a forecast, the model
 184 was fit using the most recent T consecutive observations (where T is called the
 185 training segment length), and then used to predict the observation n weeks after
 186 the most recent observation. This forecasting method was carried out along the
 187 entire available time series, fitting a distinct model to each segment of length
 188 T . The model order was set to $p = 3$ and the training segment length was set
 189 to $T = 200$, since forecast skill plateaued at these values.

190 3.4. Forecast assessment

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

$$R^2\text{-score} = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (4)$$

$$S = 100 \times \frac{\sqrt{\sum_i (y_i - f_i)^2}}{\sqrt{\sum_i (y_i - \bar{y})^2}}, \quad (5)$$

Table 1: Performance statistics of VCI3M forecasts with lead times of 2, 4 and 6 weeks. Data for slope and intercept show ordinary least squares estimates $\pm$ standard error.

	Landsat GP			MODIS AR		
	2	4	6	2	4	6
	weeks			weeks		
R^2 -score	0.99	0.95	0.87	0.99	0.96	0.88
RMSE	1.8	4.2	6.8	1.8	4.3	7.0
slope	1.0 ± 0.0	1.1 ± 0.0	1.1 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
intercept	-1 ± 0	-2 ± 0	-3 ± 0	0 ± 0	0 ± 0	0 ± 0

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

These performance indicators were also used to assess the sensitivity of our methods in space (comparing the results by region) and in time (to account for seasonality). Additionally, the forecast methods were evaluated for various drought categories (Klisch and Atzberger, 2016), compared against a persistence forecast (i.e. forecast obtained by taking the most recent observation to be the forecast value), and the impact of data gaps on forecast performance was analysed.

Forecasts on the MODIS data were assessed from January 1st 2004 onward, which was approximately the earliest date for which there were sufficient prior data for the AR method to be applied. Forecasts on the Landsat data were assessed only from January 1st 2014 onward, since the GP gap-filling method required training on data up to this date.

Figure 4: Contour plots of VCI3M against two, four and six weeks VCI3M forecasts. (a,c,e) show forecast performance for the GP method on Landsat data, and (b,d,f) show forecast performance for the AR method on MODIS data (across the 19 regions for which a 4 week forecast was possible more than 50% of the time, see main text for details).

211 4. Results

212 4.1. Forecast value accuracy

213 The GP and AR forecasting methods were applied, on each of the two
214 datasets, to regional aggregate VCI3M time series. We focus on performance
215 results of GP forecasting on Landsat data and AR forecasting on MODIS data
216 since these two combinations of data and forecasting method performed the
217 best (as measured by R^2 -score). We looked at lead times of up to ten weeks
218 (see Figures C.4 and C.5). However, due to increasing uncertainty, the results
219 provided here focus on two to six weeks forecasts of VCI3M.

220 Contour plots of forecast against actual data for two, four and six week
221 forecasts are shown in Fig. 4. Table 1 shows the R^2 -scores, RMSE, slope and
222 intercept from each of these plots, and demonstrates that there is substantial
223 forecast skill from each method at each lead time (R^2 -scores are substantial), and
224 that the forecasts are unbiased (slopes are all approximately 1, and intercepts
225 approximately 0). Corresponding results for NDVI anomaly time series can be
226 found in the Supplementary Material, in Fig. C.3 and Table C.5. The much
227 higher R^2 -scores for VCI3M compared to NDVI anomaly is due to the fact
228 that VCI3M is the 12 week composite of weekly VCI observations, and thus
229 there is much greater correlation between VCI3M at time t and VCI3M at time
230 $t + (2, 4, 6)$ weeks due to overlap in the composited weeks. Notwithstanding
231 this, each forecasting method performed substantially better than a persistence
232 forecast of VCI3M (i.e. taking the most recent observation to be the forecast
233 value). For example, the AR method on the MODIS data achieved an RMSE
234 of approximately half that of the persistence forecast for a lead time of 4 weeks,
235 see Fig. C.6 in the Supplementary Material.

236 For both methods, the forecast time series sometimes lag behind the true
237 time series, since changes not foreseen by the models are incorporated only once

Figure 5: Sample aggregate VCI3M time series from the intersection of Baringo county and livelihood zone 24 from (a) Landsat and (b) MODIS (solid lines). Dotted lines show forecasts at a lead time of 4 weeks, as given by the GP method on the Landsat data, and the AR method on the MODIS data.

they are observed, see Fig. 5 for examples. The two methods sometimes make different types of error. The GP method is more likely to predict a value that is closer to the long-term mean than the true value. This is because *a priori* to taking into account the most recent observations, the model assumes the forecast observation will be equal to the long term mean. By contrast, the AR method is more likely to predict a continuation of the recent trend. This is because the model assumes a continuation of the recent frequency profile, so if a faster-than-average trend is seen in either direction, the trend will be predicted to continue (Hamilton, 1994).

Due to the presence of non-interpolated gaps in the MODIS time series, there were weeks when a forecast assessment was not carried out on these data, see Table B.2 for details. For 15 of the regions, a 4 week forecast could be made on more than 90 percent of weeks; however, for some regions, a forecast could

Table 2: RMSE in VCI3M forecast, for the true vegetation condition belonging to the different categories of drought, at lead times of 2, 4 and 6 weeks. Drought categories are defined by the VCI3M index: wet by $VCI3M > 50$; normal by $35 < VCI3M < 50$; moderate drought by $20 < VCI3M < 35$; severe drought by $10 < VCI3M < 20$; and extreme drought by $VCI3M < 10$. (The extreme drought criterion was not met in any of the Landsat data.)

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

251 rarely be made, see Fig. C.9e in the Supplementary Material.

252 Additional checks were included to test the sensitivity of the methods to
253 drought severity and seasonality. The methods, when computed separately for
254 each of the five categories on the NDMA drought scale (Klisch and Atzberger,
255 2016), perform better in terms of absolute RMSE when there was a state of
256 drought than when the vegetation condition was normal or wet (Table 2). This
257 could be explained by the fact that when conditions are relatively normal, the
258 subsequent conditions could go in various directions. However, when there is
259 an extreme drought, it is likely that it will persist (because vegetation cover is
260 already below-normal). Note though that the relative RMSE as a proportion
261 of the VCI3M value appears to be similar during drought than during normal
262 vegetation condition. Neither method exhibited large seasonal differences, see
263 Fig. 6. However, for the GP method applied to Landsat data at a 6-week lead
264 time, RMSE is high at the start of the season but drops noticeably during the
265 course of the season, suggesting that it is harder to predict how a season will
266 start but easier to forecast how it will proceed once started.

267 The fact that seasonal differences in RMSE are not substantial provides re-
268 assurance that the forecast accuracy estimates are not inflated by the gap-filling
269 during preprocessing. If this were the case, there would be a sustained drop in

Figure 6: RMSE of VCI3M forecast for each week of the year. (a) GP forecasting on Landsat data. (b) AR forecasting on MODIS data. Grey shading indicate the rainy seasons, March-May and October-December.

RMSE during the more overcast months of the year (March to May and October to December). While the GP forecasts on the Landsat data were computed from time series on which no future data were used for the interpolations (see 3.1), for the MODIS data interpolations did make use of future data. Therefore, to obtain further reassurance that performance estimates of AR forecasting on the MODIS data were not inflated, a plot was made of RMSE at 4 weeks lead time against percentage of pixels from which a good observation was obtained on the date of the forecast, see Fig. C.7 in the Supplementary Material. There was no apparent correlation (Pearson coefficient was 0.01), and hence it was concluded that the gap-filling was not leading to inflated forecast performance.

4.2. Drought event forecast: ROC curves

To assess the usefulness of the AR and GP methods for drought forecasting, we tested their ability to detect specific drought events, as defined by the NDMA's alert threshold ($VCI3M < 35$, Klisch and Atzberger, 2016). ROC curves were plotted for the detection of $VCI3M < 35$ at lead times of two, four and six weeks, see Fig. 7(a, b). 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. We fur-

Figure 7: (a) ROC curve for drought detection (VCI3M < 35) for lead times of 2, 4 and 6 weeks using the GP method on Landsat data. (b) ROC curve for drought detection using the AR method on MODIS data. (c, d) Respectively for the GP method on Landsat data and the AR method on the MODIS data, hit rate versus false alarm ratio for forecasting a transition to drought (VCI3M < 35) given that the vegetation condition is normal (VCI3M > 35) on the date of the forecast. 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.

Table 3: False alarm rate and hit rate (respectively, expressed in percent) for different regions in Kenya and at different lead times. This is based on forecasting drought if the predicted VCI3M is less than 35 (different performances could be obtained with different warning thresholds (see Figure 7. Regions are composed of the following zones: North – Z1,3 and 5; East – Z7, 9, 10 and 11 and South – (Z15 and 18))

Regions	Landsat GP						MODIS AR					
	2	4	6	2	4	6	2	4	6	2	4	6
	weeks			weeks			weeks			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	1	97	2	88	3	76	2	98	6	93	11	87
East	3	94	5	85	6	77	3	97	6	91	10	85
South	1	96	3	88	4	77	2	98	6	94	11	90

ther tested the ability to detect the onset of drought, see Fig. 7(c, d). For this, the hit rate was defined as the probability of predicting a transition from the normal condition ($VCI3M > 35$) to the drought condition ($VCI3M < 35$), given that this transition occurs; and the false alarm ratio was defined as the probability that a prediction of this transition is incorrect, given that this transition has been predicted. These curves give an indication that both methods have high skill at forecasting droughts, as measured by the VCI3M, as far as six weeks ahead. The AR forecast appears to do better than the GP forecast at predicting transitions from normal conditions to drought, which can be explained by the tendency of the AR forecast to predict a continuation of the recent trend, while the GP forecast is more likely to predict a reversion to the long term mean value.

The ROC curve performance is not highly dependent on the region (see Table 3). 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. C.10 in the Supplementary Material.

308 5. Discussion

309 Droughts are complex and hence inherently difficult to define and measure
 310 (Mishra and Singh, 2010). A large number of satellite-based indicators have been
 311 developed to identify meteorological, hydrological, and agricultural droughts
 312 (Zargar et al., 2011; AghaKouchak et al., 2015) with each performing well in
 313 space and time to a certain degree (Zhang et al., 2017). This paper uses two
 314 machine-learning methods to provide short-term forecasts of the 3-month VCI
 315 (VCI3M), which is used by Kenya’s National Drought Management Authority
 316 (NDMA) in their drought Early Warning System (EWS). We have investigated
 317 the skill and robustness of our forecasts in a number of ways. Both of our meth-
 318 ods showed high sensitivity and specificity for prediction of drought conditions
 319 ($VCI3M < 35$) as well as the onset of drought, at lead times of 2, 4 and 6 weeks
 320 (see Fig. 7). They also perform better than a persistence forecast (a factor of
 321 two in RMSE for VCI3M, and R^2 improvement of 0.12). Compared to a similar
 322 study that used a Artificial Neural Network model to predict future VCI for
 323 four Kenyan counties (Adede et al., 2019), our forecasts provide higher skill, as
 324 they showed R^2 -scores of 0.78 for a 1-month VCI3M forecast compared to 0.95
 325 and 0.94 for our methods. Moreover, our two methods provide robust results
 326 with either dataset (i.e., MODIS and Landsat), and are not impacted by the
 327 preprocessing steps. Finally, the methods present a high skill in forecasting
 328 drought irrespective of the region, the drought category, and the season.

329 A very important strength of our methods is the high level of skill. It is
 330 instructive to understand the origin of this skill; particularly as it may be sur-
 331 prising since the methods are rather simple and do not include other variables
 332 (e.g., rainfall, precipitation). Part of the explanation is that, by using the in-
 333 dicators themselves to determine the forecast, we do track all the factors that
 334 impact the vegetation (e.g., disease, soil memory and land-use change) and not

335 just meteorological factors. Additionally, the natural growth cycles of vegeta-
336 tion and their response to environmental factors introduce temporal correlations
337 (persistence) in the indices which can be exploited in short-range forecasting.
338 Furthermore, the VCI3M metric used in this study, and by the NDMA, is ad-
339 ditionally smoothed over three months. This smoothing adds temporal depen-
340 dency, which in turn increases the measured skill. It is important to recognise
341 that this last improvement in skill comes from the inclusion of current data,
342 so it is not a pure forecast and skill metrics should not be directly compared
343 with skills of e.g. VCI forecasts. Nevertheless, the high apparent skill is ex-
344 tremely valuable to disaster risk managers who need to make decisions based
345 on uncertain information (Wilkinson et al., 2018).

346 There are an increasing number of forecasting studies and methodologies for
347 pastures that focus on several indices, with different lead times, and with vary-
348 ing skill (Matere et al., 2019; Adede et al., 2019; Papagiannopoulou et al., 2017;
349 Meroni et al., 2014). Often, new forecast information developed by scientists to
350 help the development and humanitarian sectors enhance disaster preparedness
351 and response goes unused due to a “usability gap” between knowledge produc-
352 ers and users (Lemos et al., 2012). In our study, we aim to bridge this gap by
353 focusing on VCI3M, a drought indicator that is currently used by the NDMA
354 to classify drought severity in the arid and semi-arid regions of Kenya. Addi-
355 tionally, decision makers need reliable forecasts to develop robust anticipatory
356 actions in order to mitigate the impacts of drought with limited financial re-
357 sources (Wilkinson et al., 2018). Our methods provide skillful VCI3M forecasts
358 with detailed information of hit rates and false-alarm ratios, which are often used
359 to define anticipatory actions (see the Red Cross/Red Crescent Forecast-based
360 Financing (FbF) manual for more information¹). Finally, the methodology is

¹http://fbf.drk.de/fileadmin/Content/Manual_FbF/01_Manual/01_Manual_For_Forecast-

361 also rather simple and easy to implement as it only relies on one data input
362 derived from satellites, which are available globally. The methods can thus be
363 applied everywhere, providing there is sufficient capacity and calibration data.
364 Such co-production strategies allow us to bridge the usability gap (Dilling and
365 Lemos, 2011; Lemos et al., 2018) and provides confidence that our forecasting
366 methods may be used.

367 We have also concentrated on methods that produce accurate short-term
368 forecasts, rather than less-certain, but longer-range forecasts. We can speculate
369 that while the latter might have greater value, the former might be more readily
370 adopted in the monthly county bulletins released by the NDMA. Indeed, the
371 forecasts developed here could, for example, help establish a new drought phase
372 classification (‘Early Alert’) which, along with adequate preparedness actions
373 developed by the disaster risk managers, would minimise the risk of a worsen-
374 ing drought condition. Anticipatory drought management strategies based on
375 this ‘Early Alert’ could for example focus on livestock vaccination programmes,
376 livestock movement monitoring, or the repair of strategic water sources which
377 enhance the resilience of these communities before a drought hits.

378 Forecasts alone do not necessarily lead to good anticipatory actions. Whilst
379 acting ahead of disasters is on average more financially effective than responding
380 to an event (Venton et al., 2012), traditionally the humanitarian agencies tend
381 to respond to disasters as financial resources are only available during or after an
382 event. Additionally, due to the uncertainty in the models, anticipatory actions
383 based on such forecasts do raise the risk of “acting in vain”, which may have
384 substantial negative impact on the humanitarian sector in the short term (Lopez
385 et al., 2018). These agencies thus need access to adequate financial resources,
386 e.g. FbF (Coughlan de Perez et al., 2015) to fund anticipatory actions based on

387 skill-assessed forecast in order to factor in the possible negative consequences
388 of acting in vain. For the forecast methods developed in this study, the chance
389 of acting in vain will be low due to the high level of skill, which will ultimately
390 lower the barriers to uptake.

391 **6. Caveats and Future Work**

392 As discussed above, our methods are already sufficiently skillful that they
393 are usable as they stand. However, we have identified some minor limitations
394 and relevant improvements to enhance the functionality, skill, lead-time and
395 impact of our forecasts.

396 Our analysis has been based on relatively small samples of the available
397 pixels, aggregated, spatially, at the level of the pastoral livelihood zone and
398 county intersections. This limits the localisation specificity of our predictions.
399 Additionally, our methodology using Landsat merges data from various land
400 covers which may reduce accuracy. The processing of all pixels can be achieved
401 within reasonable computational constraints and will allow us to aggregate over
402 specific regions of interest. For example, one could perform the analysis for
403 specific land covers within a county, or for individual grazing units, which would
404 provide greater accuracy and additional functionality.

405 Our forecasts are currently unavailable, or are less accurate, in periods during
406 or following cloud-cover gaps. More subtly, our validation will have favoured
407 dry season observations, which are less affected by cloud cover, and this will
408 have an impact on the validation of the forecast performance. However, as we
409 found little variation in performance throughout the seasons we do not think
410 these are significant problems. The impact of cloud cover will be reduced when
411 all pixels are processed and aggregated.

412 As with any machine-learning method, the forecast and its estimated skill

413 are only appropriate for the types of vegetation and environments for which it
414 has been calibrated. The quality that we have obtained in different regions of
415 Kenya gives us some confidence that the skills will not be substantially different.
416 Nevertheless, the calibration, validation and skill assessment of the forecast will
417 be an essential element of a practical and general tool. Future work should also
418 explore how long a temporal baseline is required for good calibration. This will,
419 in principle, allow a truly global forecasting tool.

420 The error estimates we currently provide in our forecasts are derived from the
421 global validation of our performance. They should thus be correct on average
422 but the errors will be overestimated in some situations while, correspondingly,
423 underestimated in others. Future development can provide error estimates that
424 are tailored for the specific conditions and data availability.

425 The indicators we have chosen are well motivated through their use in the
426 existing EWS operated by NDMA. Identifying the most appropriate and useful
427 indicators of such hazards is the subject of much debate and investigation (see
428 e.g., Zargar et al. (2011); AghaKouchak et al. (2015); Zhang et al. (2017)).
429 But they may not be the most suitable to quantify the relevant socio-economic
430 impacts of droughts (Sutanto et al., 2019). Subject to data availability, similar
431 machine-learning approaches could be applied to more direct socio-economic
432 indicators tracking food insecurity, such as malnutrition, food prices, or livestock
433 condition.

434 Perhaps the most-significant limitation of our methods is that they are only
435 appropriate for relatively short lead-times. Although a 4-week lead time can be
436 useful, most contingency plans and drought preparedness policies are developed
437 over seasonal timescales. It is thus key to extend this lead-time. While current
438 observations of precipitation and temperature had little impact, including other
439 observed climate variables (e.g., ENSO, sea surface temperature, Funk et al.

440 (2008)) or seasonal climate forecasts may enhance skill and lead times.

441 Future research will also be required on the effectiveness of the practical
442 implementation of forecasts in EWS (Wilkinson et al., 2018). Clearly-defined
443 triggers (e.g., threshold values based on forecasts, which may vary in time and
444 space) will need to be defined and assessed and optimised against suitable per-
445 formance metrics. Similarly, effective anticipatory actions need to be defined
446 by the decision makers in relation to these triggers. Adequate policy and in-
447 stitutional arrangements will be needed to allow the various actors to engage
448 and interact with a long-term perspective on risk management. This in turn,
449 requires financial systems that can be accessed based on such forecasts to be
450 able to act across various timescales before the disaster occurs (i.e. Forecast-
451 based-Finance).

452 **7. Conclusion**

453 In conclusion we have developed two new forecasting methods which exploit
454 the inherent temporal correlation in vegetation indices to provide highly skill-
455 ful, short-range forecasts of VCI. The choice of input data, output indicators,
456 simplicity of implementation, and demonstrated skill argues that these methods
457 will be useful for drought early warning systems. We have identified ways this
458 can be improved, but there is clear evidence here that our statistical persistence
459 model provides strong skill over useful lead times. This can be an important
460 contribution to anticipatory drought risk management in Kenya

461 **Authors responsibilities**

462 A.B.B., S.D. and E.S. are lead authors as they contributed equally to the
463 paper and the order of the three names is alphabetical. A.B.B was responsible
464 for developing and running the AR method. S.D. was responsible for developing

465 and running the GP method, and for the accumulation and processing of the
466 Landsat data. E.S. was responsible for the MODIS data accumulation and pre-
467 processing. JMw, SO and PR developed the initial idea and provided feedback
468 throughout. All authors wrote, reviewed and edited the final manuscript.

469 **Acknowledgements**

470 This research was funded by the STFC through the following projects: “As-
471 troCast: Applying Astronomy Data Analysis to enhance disaster forecasting”
472 – grant number ST/R004811/1; and “STFC Official Development Assistance
473 (ODA) Institutional Award” attached to the same grant; “A UK-Africa Data
474 Science Network: Capturing the SKA-Driven Data Transformation” grant num-
475 ber ST/R001898/1; and by the Science for Humanitarian Emergencies and
476 Resilience (SHEAR) consortium project “Towards Forecast-based Prepared-
477 ness Action” (ForPAC, www.forpac.org), grant numbers NE/P000673/1. This
478 project was initiated through pump-priming funding from the University of
479 Sussex’s “Sussex Research” thematic programme and carried out as part of the
480 interdisciplinary Data Intensive Science Centre at the University of Sussex (DIS-
481 CUS). We acknowledge early contributions to that pilot work from Peter Hurley,
482 Philip Rooney, Martin Jung, and Jörn Scharlemann. Martin Todd and Alexan-
483 der Antonarakis provided useful feedback which improved the manuscript.

484 **References**

485 **References**

486 Adede C, Oboko R, Wagacha PW, Atzberger C. A Mixed Model Ap-
487 proach to Vegetation Condition Prediction Using Artificial Neural Networks
488 (ANN): Case of Kenya’s Operational Drought Monitoring. Remote Sens-

489 ing 2019;11(9):1099. URL: <https://www.mdpi.com/2072-4292/11/9/1099>.
 490 doi:10.3390/rs11091099.

491 AghaKouchak A, Farahmand A, Melton F, Teixeira J, Anderson M, Wardlow
 492 BD, Hain C. Remote sensing of drought: Progress, challenges and opportu-
 493 nities. *Reviews of Geophysics* 2015;53(2):452–80.

494 Asoka A, Mishra V. Prediction of vegetation anomalies to improve food
 495 security and water management in India. *Geophysical Research Letters*
 496 2015;42(13):5290–8. doi:10.1002/2015GL063991.

497 Behnke R, Muthami D. The Contribution of Livestock to the Kenyan
 498 Economy. Technical Report; 2011. URL: https://cgspace.cgiar.org/bitstream/handle/10568/24972/IGAD_{_}LPI_{_}WP_{_}03-11.pdf?sequence=1{&}isAllowed=y.

501 Bhuiyan C, Singh R, Kogan F. Monitoring drought dynamics in the Aravalli
 502 region (India) using different indices based on ground and remote sensing
 503 data. *International Journal of Applied Earth Observation and Geoinforma-*
 504 *tion* 2006;8(4):289 – 302. URL: [http://www.sciencedirect.com/science/](http://www.sciencedirect.com/science/article/pii/S030324340600016X)
 505 [article/pii/S030324340600016X](http://www.sciencedirect.com/science/article/pii/S030324340600016X).

506 Brahim-Belhouari S, Vesin JM. Bayesian learning using Gaussian process for
 507 time series prediction. In: *Proceedings of the 11th IEEE Signal Processing*
 508 *Workshop on Statistical Signal Processing* (Cat. No.01TH8563). 2001. p. 433–
 509 6. doi:10.1109/SSP.2001.955315.

510 Brown ME, Kshirsagar V. Weather and international price shocks on food
 511 prices in the developing world. *Global Environmental Change* 2015;35:31–
 512 40. URL: [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S0959378015300248)
 513 [S0959378015300248](http://www.sciencedirect.com/science/article/pii/S0959378015300248). doi:[https://doi.org/10.1016/j.gloenvcha.2015.](https://doi.org/10.1016/j.gloenvcha.2015.08.003)
 514 [08.003](https://doi.org/10.1016/j.gloenvcha.2015.08.003).

515 Camps-Valls G, Verrelst J, Munoz-Mari J, Laparra V, Mateo-Jimenez F, Gomez-
516 Dans J. A survey on Gaussian processes for earth-observation data analysis: A
517 comprehensive investigation. *IEEE Geoscience and Remote Sensing Magazine*
518 2016;4(2):58–78. doi:10.1109/MGRS.2015.2510084.

519 Cao R, Chen Y, Shen M, Chen J, Zhou J, Wang C, Yang W. A simple method
520 to improve the quality of NDVI time-series data by integrating spatiotempo-
521 ral information with the Savitzky-Golay filter. *Remote Sensing of Environ-*
522 *ment* 2018;217:244–57. URL: [http://www.sciencedirect.com/science/](http://www.sciencedirect.com/science/article/pii/S0034425718303985)
523 [article/pii/S0034425718303985](http://www.sciencedirect.com/science/article/pii/S0034425718303985). doi:[https://doi.org/10.1016/j.rse.](https://doi.org/10.1016/j.rse.2018.08.022)
524 2018.08.022.

525 Chandola V, Vatsavai RR. A scalable Gaussian process analysis algorithm for
526 biomass monitoring. *Statistical Analysis and Data Mining: The ASA Data*
527 *Science Journal* 2011;4(4):430–45. URL: [https://onlinelibrary.wiley.](https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.10129)
528 [com/doi/abs/10.1002/sam.10129](https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.10129).

529 Coughlan de Perez E, van den Hurk B, van Aalst MK, Jongman B,
530 Klose T, Suarez P. Forecast-based financing: an approach for cat-
531 alyzing humanitarian action based on extreme weather and climate fore-
532 casts. *Natural Hazards and Earth System Sciences* 2015;15(4):895–
533 904. URL: [https://www.nat-hazards-earth-syst-sci.net/15/895/](https://www.nat-hazards-earth-syst-sci.net/15/895/2015/)
534 2015/. doi:10.5194/nhess-15-895-2015.

535 Dilling L, Lemos MC. Creating usable science: Opportunities and constraints
536 for climate knowledge use and their implications for science policy. *Global*
537 *environmental change* 2011;21(2):680–9.

538 Duvenaud D, Lloyd J, Grosse R, Tenenbaum J, Zoubin G. Structure discov-
539 ery in nonparametric regression through compositional kernel search. In:
540 Dasgupta S, McAllester D, editors. *Proceedings of the 30th International*

541 Conference on Machine Learning. Atlanta, Georgia, USA: PMLR; volume 28
 542 of *Proceedings of Machine Learning Research*; 2013. p. 1166–74. URL:
 543 <http://proceedings.mlr.press/v28/duvenaud13.html>.

544 EM-DAT . The International Disaster Database. Université catholique de Lou-
 545 vain, Belgium, 2019. URL: www.emdat.be.

546 FAO . Food and Agriculture Organization (FAO) Country programming frame-
 547 work for Kenya (2014-2017). Technical Report; 2014. URL: [www.fao.org/](http://www.fao.org/publications)
 548 [publications](http://www.fao.org/publications).

549 Funk C, Dettinger MD, Michaelsen JC, Verdin JP, Brown ME, Barlow M,
 550 Hoell A. Warming of the Indian Ocean threatens eastern and southern
 551 African food security but could be mitigated by agricultural development.
 552 Proceedings of the National Academy of Sciences 2008;105(32):11081 LP
 553 – 11086. URL: <http://www.pnas.org/content/105/32/11081.abstract>.
 554 doi:10.1073/pnas.0708196105.

555 Galvin K, Boone R, Smith N, Lynn S. Impacts of climate variability on East
 556 African pastoralists: linking social science and remote sensing. Climate
 557 Research 2001;19:161–72. URL: [http://www.int-res.com/abstracts/cr/](http://www.int-res.com/abstracts/cr/v19/n2/p161-172/)
 558 [v19/n2/p161-172/](http://www.int-res.com/abstracts/cr/v19/n2/p161-172/). doi:10.3354/cr019161.

559 Hamilton J. Time series analysis. Princeton, NJ: Princeton University Press,
 560 1994.

561 Hoffman MD, Blei DM, Wang C, Paisley J. Stochastic variational inference. J
 562 Mach Learn Res 2013;14(1):1303–1347.

563 Jiao W, Zhang L, Chang Q, Fu D, Cen Y, Tong Q. Evaluating an enhanced
 564 vegetation condition index (VCI) based on VIUPD for drought monitoring

565 in the continental united states. Remote Sensing 2016;8(3). URL: <https://www.mdpi.com/2072-4292/8/3/224>. doi:10.3390/rs8030224.

566

567 Kandasamy S, Baret F, Verger A, Neveux P, Weiss M. A comparison
568 of methods for smoothing and gap filling time series of remote sens-
569 ing observations - application to MODIS LAI products. Biogeosciences
570 2013;10(6):4055–71. URL: [https://www.biogeosciences.net/10/4055/](https://www.biogeosciences.net/10/4055/2013/)
571 2013/. doi:10.5194/bg-10-4055-2013.

572 Klisch A, Atzberger C. Operational drought monitoring in Kenya using MODIS
573 NDVI time series. Remote Sensing 2016;8(4). URL: [http://www.mdpi.com/](http://www.mdpi.com/2072-4292/8/4/267)
574 2072-4292/8/4/267. doi:10.3390/rs8040267.

575 Kogan F. Application of vegetation index and brightness temperature
576 for drought detection. Advances in Space Research 1995;15(11):91–
577 100. URL: [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S027311779500079T)
578 027311779500079T; natural Hazards: Monitoring and Assessment Using Re-
579 mote Sensing Technique”.

580 Kogan F, Adamenko T, Guo W. Global and regional drought dynamics in the
581 climate warming era. Remote Sensing Letters 2013;4(4):364–72. URL: <https://doi.org/10.1080/2150704X.2012.736033>. doi:10.1080/2150704X.
582 //doi.org/10.1080/2150704X.2012.736033. doi:10.1080/2150704X.2012.736033. arXiv:<https://doi.org/10.1080/2150704X.2012.736033>.

583

584 Lemos MC, Arnott JC, Ardoin NM, Baja K, Bednarek AT, Dewulf A, Fieseler
585 C, Goodrich KA, Jagannathan K, Klenk N, et al. To co-produce or not to
586 co-produce. Nature Sustainability 2018;1(12):722.

587 Lemos MC, Kirchhoff CJ, Ramprasad V. Narrowing the climate information
588 usability gap. Nature climate change 2012;2(11):789.

589 Lesk C, Rowhani P, Ramankutty N. Influence of extreme weather disas-
590 ters on global crop production. *Nature* 2016;529(7584):84–7. doi:10.1038/
591 **nature16467**.

592 Lopez A, Coughlan de Perez E, Bazo J, Suarez P, van den Hurk B, van Aalst
593 M. Bridging forecast verification and humanitarian decisions: A valuation
594 approach for setting up action-oriented early warnings. *Weather and Climate*
595 *Extremes* 2018;doi:10.1016/j.wace.2018.03.006.

596 Matere J, Simpkin P, Angerer J, Olesambu E, Ramasamy S, Fasina F. Predictive
597 livestock early warning system (PLEWS): Monitoring forage condition and
598 implications for animal production in Kenya. *Weather and Climate Extremes*
599 2019;:100209.

600 Meroni M, Fasbender D, Kayitakire F, Pini G, Rembold F, Urbano F, Verstraete
601 M. Early detection of biomass production deficit hot-spots in semi-arid en-
602 vironment using FAPAR time series and a probabilistic approach. *Remote*
603 *Sensing of Environment* 2014;142:57–68.

604 Mishra AK, Singh VP. A review of drought concepts. *Journal of Hydrology*
605 2010;391(1):202–16. URL: [http://www.sciencedirect.com/science/](http://www.sciencedirect.com/science/article/pii/S0022169410004257)
606 [article/pii/S0022169410004257](http://www.sciencedirect.com/science/article/pii/S0022169410004257). doi:[https://doi.org/10.1016/j.](https://doi.org/10.1016/j.jhydrol.2010.07.012)
607 [jhydrol.2010.07.012](https://doi.org/10.1016/j.jhydrol.2010.07.012).

608 Nelson GC, Valin H, Sands RD, Havlík P, Ahammad H, Deryng D, El-
609 liott J, Fujimori S, Hasegawa T, Heyhoe E, Kyle P, Von Lampe
610 M, Lotze-Campen H, Mason d’Croz D, van Meijl H, van der Mens-
611 brugghe D, Müller C, Popp A, Robertson R, Robinson S, Schmid E,
612 Schmitz C, Tabeau A, Willenbockel D. Climate change effects on agri-
613 culture: Economic responses to biophysical shocks. *Proceedings of*
614 *the National Academy of Sciences* 2014;111(9):3274–9. URL: [https:](https://doi.org/10.1073/pnas.1313028111)

615 [//www.pnas.org/content/111/9/3274](http://www.pnas.org/content/111/9/3274). doi:10.1073/pnas.1222465110.
616 arXiv:<https://www.pnas.org/content/111/9/3274.full.pdf>.

617 Nyong A, Adesina F, Osman Elasha B. The value of indigenous knowledge
618 in climate change mitigation and adaptation strategies in the African sa-
619 hel. *Mitigation and Adaptation Strategies for Global Change* 2007;12(5):787–
620 97. URL: "<https://doi.org/10.1007/s11027-007-9099-0>". doi:"10.
621 1007/s11027-007-9099-0".

622 Orindi VA, Nyong A, Herrero M. Pastoral Livelihood Adaptation to Drought
623 and Institutional Interventions in Kenya. Technical Report; 2007.

624 Papagiannopoulou C, Gonzalez Miralles D, Decubber S, Demuzere M, Ver-
625 hoest N, Dorigo WA, Waegeman W. A non-linear Granger-causality frame-
626 work to investigate climate-vegetation dynamics. *GEOSCIENTIFIC MODEL*
627 *DEVELOPMENT* 2017;10(5):1945–60. URL: [http://dx.doi.org/10.5194/](http://dx.doi.org/10.5194/gmd-10-1945-2017)
628 [gmd-10-1945-2017](http://dx.doi.org/10.5194/gmd-10-1945-2017).

629 Piguet E, Pécoud A, de Guchteneire P. Migration and climate change:
630 An overview. *Refugee Survey Quarterly* 2011;30(3):1–23. URL:
631 <https://doi.org/10.1093/rsq/hdr006>. doi:10.1093/rsq/hdr006.
632 arXiv:<http://oup.prod.sis.lan/rsq/article-pdf/30/3/1/4460951/hdr006.pdf>.

633 Tozier de la Poterie A, Baudoin MA. From Yokohama to Sendai: Approaches
634 to participation in international disaster risk reduction frameworks. *Inter-*
635 *national Journal of Disaster Risk Science* 2015;6(2):128–39. URL: [https://](https://doi.org/10.1007/s13753-015-0053-6)
636 doi.org/10.1007/s13753-015-0053-6. doi:10.1007/s13753-015-0053-6.

637 Quiring SM, Ganesh S. Evaluating the utility of the vegetation condition in-
638 dex (VCI) for monitoring meteorological drought in Texas. *Agricultural and*
639 *Forest Meteorology* 2010;150(3):330–9.

640 Rasmussen CE, Williams CKI. Gaussian processes for machine learning 2006;.

641 Rojas O, Vrieling A, Rembold F. Assessing drought probability for agri-
 642 cultural areas in Africa with coarse resolution remote sensing imagery.
 643 Remote Sensing of Environment 2011;115(2):343 –52. URL: <http://www.sciencedirect.com/science/article/pii/S0034425710002798>.
 644 doi:<https://doi.org/10.1016/j.rse.2010.09.006>.

645

646 Roy D, Wulder M, Loveland T, Woodcock C, Allen R, Anderson M, Helder D,
 647 Irons J, Johnson D, Kennedy R, Scambos T, Schaaf C, Schott J, Sheng Y, Ver-
 648 mote E, Belward A, Bindischadler R, Wohen W, Gao F, Zhu Z. Landsat-8: Sci-
 649 ence and product vision for terrestrial global change research. Remote Sensing
 650 of Environment 2014;2014:154–72. doi:10.1016/j.rse.2014.02.001.

651 Rulinda CM, Bijker W, Stein A. The chlorophyll variability in Meteosat derived
 652 NDVI in a context of drought monitoring. Procedia Environmental Sciences”
 653 2011;3:32 –7. URL: <http://www.sciencedirect.com/science/article/pii/S1878029611000089>. doi:<https://doi.org/10.1016/j.proenv.2011.02.007>;
 654 02.007; 1st Conference on Spatial Statistics 2011 – Mapping Global Change.
 655

656 Sai F, Cumiskey L, Weerts A, Bhattacharya B, Haque Khan R. To-
 657 wards impact-based flood forecasting and warning in Bangladesh: a
 658 case study at the local level in Sirajganj district. Natural Hazards
 659 and Earth System Sciences Discussions 2018;2018:1–20. URL: <https://www.nat-hazards-earth-syst-sci-discuss.net/nhess-2018-26/>.
 660 doi:10.5194/nhess-2018-26.
 661

662 Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Sim-
 663 plified Least Squares Procedures. Analytical Chemistry 1964;36(8):1627–
 664 39. URL: <http://pubs.acs.org/doi/abs/10.1021/ac60214a047>. doi:10.
 665 1021/ac60214a047.

666 Schaaf C, Wang Z. MCD43A4 MODIS/Terra+Aqua BRDF/Albedo
667 Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006. 2015.
668 URL: [https://lpdaac.usgs.gov/dataset{ }discovery/modis/
669 modis{ }products{ }table/mcd43a4{ }v006](https://lpdaac.usgs.gov/dataset{ }discovery/modis/modis{ }products{ }table/mcd43a4{ }v006). doi:10.5067/MODIS/
670 MCD43A4.006.

671 Stanke C, Kerac M, Prudhomme C, Medlock J, Murray V. Health effects of
672 drought: a systematic review of the evidence. PLoS currents 2013;5. URL:
673 <https://www.ncbi.nlm.nih.gov/pubmed/23787891>.

674 Sutanto SJ, van der Weert M, Wanders N, Blauhut V, Van Lanen HAJ.
675 Moving from drought hazard to impact forecasts. Nature Communications
676 2019;10(1):4945. URL: <https://doi.org/10.1038/s41467-019-12840-z>.
677 doi:10.1038/s41467-019-12840-z.

678 Udelhoven T, Stellmes M, del Barrio G, Hill J. Assessment of rainfall and
679 NDVI anomalies in Spain (1989 – 1999) using distributed lag models. In-
680 ternational Journal of Remote Sensing 2009;30(8):1961–76. URL: [https://
681 doi.org/10.1080/01431160802546829](https://doi.org/10.1080/01431160802546829). doi:10.1080/01431160802546829.
682 arXiv:<https://doi.org/10.1080/01431160802546829>.

683 UNDP . Kenya – adaptation to climate change in arid lands (KACCAL)-
684 Kenya case study. Technical Report; 2013. URL: [http://gefonline.org/
685 projectDetailsSQL.cfm?projID=3792](http://gefonline.org/projectDetailsSQL.cfm?projID=3792).

686 Upreti D, Huang W, Kong W, Pascucci S, Pignatti S, Zhou X, Ye H, Casa R. A
687 comparison of hybrid machine learning algorithms for the retrieval of wheat
688 biophysical variables from Sentinel-2. Remote Sensing 2019;11(5). URL:
689 <http://www.mdpi.com/2072-4292/11/5/481>. doi:10.3390/rs11050481.

690 Venton C, Fitzgibbon C, Shitarek T, Coulter L, Dooley O. The economics

691 of early response and disaster resilience: lessons from Kenya and Ethiopia.
692 Independent report 2012;.

693 Vrieling A, Meroni M, Mude AG, Chantarat S, Ummenhofer CC, de Bie KC.
694 Early assessment of seasonal forage availability for mitigating the impact
695 of drought on East African pastoralists. *Remote sensing of environment*
696 2016;174:44–55.

697 Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. An
698 effective approach for gap-filling continental scale remotely sensed time-
699 series. *ISPRS Journal of Photogrammetry and Remote Sensing* 2014;98:106–
700 18. URL: [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S0924271614002512)
701 [S0924271614002512](https://www.sciencedirect.com/science/article/pii/S0924271614002512). doi:10.1016/J.ISPRSJPRS.2014.10.001.

702 Wilkinson E, Weingärtner L, Choularton R, Bailey M, Todd M, Kniveton D,
703 Cabot Venton C. Forecasting hazards, averting disasters: Implementing
704 forecast-based early action at scale. Technical Report; Overseas Development
705 Institute (ODI); 2018.

706 WMO . WMO guidelines on multi-hazard impact-based forecast and warning
707 services 2015;.

708 Zambrano F, Vrieling A, Nelson A, Meroni M, Tadesse T. Prediction of drought-
709 induced reduction of agricultural productivity in Chile from MODIS, rainfall
710 estimates, and climate oscillation indices. *Remote Sensing of Environment*
711 2018;219:15–30. URL: [http://www.sciencedirect.com/science/article/](http://www.sciencedirect.com/science/article/pii/S0034425718304541)
712 [pii/S0034425718304541](http://www.sciencedirect.com/science/article/pii/S0034425718304541). doi:[https://doi.org/10.1016/j.rse.2018.10.](https://doi.org/10.1016/j.rse.2018.10.006)
713 006.

714 Zargar A, Sadiq R, Naser B, Khan FI. A review of drought indices. *Environ-*
715 *mental Reviews* 2011;19(NA):333–49.

716 Zhang L, Jiao W, Zhang H, Huang C, Tong Q. Studying drought phe-
717 nomena in the continental united states in 2011 and 2012 using var-
718 ious drought indices. Remote Sensing of Environment 2017;190:96-
719 106. URL: [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S0034425716304813)
720 [S0034425716304813](http://www.sciencedirect.com/science/article/pii/S0034425716304813). doi:<https://doi.org/10.1016/j.rse.2016.12.010>.

721 **List of Figures**

722	1	Maps of Kenya showing (a) Livelihood Zones and County in-	
723		tersections (Regions of Interest (ROI)) from which pixels were	
724		sampled for analysis, and (b) land-cover classification (according	
725		to the MODIS MCD12Q1 data). Analyses were performed for 29	
726		regions, defined by pastoral livelihood zone and county intersec-	
727		tions. A map showing the livelihood zones can be found in Fig.	
728		C.8 in the Supplementary Material.	6
729	2	A flow chart of the data processing and analysis.	7
730	3	Illustration of the GP approach used for the Landsat data. In	
731		“forecast mode”, the correlations in the data up to a given date	
732		furnish a GP model, which can then be used for forecasting. In	
733		the “non-forecast mode”, the entire time series is used to train	
734		the GP, and provide a ground truth for the forecast.	8
735	4	Contour plots of VCI3M against two, four and six weeks VCI3M	
736		forecasts. (a,c,e) show forecast performance for the GP method	
737		on Landsat data, and (b,d,f) show forecast performance for the	
738		AR method on MODIS data (across the 19 regions for which a 4	
739		week forecast was possible more than 50% of the time, see main	
740		text for details).	14
741	5	Sample aggregate VCI3M time series from the intersection of	
742		Baringo county and livelihood zone 24 from (a) Landsat and (b)	
743		MODIS (solid lines). Dotted lines show forecasts at a lead time	
744		of 4 weeks, as given by the GP method on the Landsat data, and	
745		the AR method on the MODIS data.	16

746	6	RMSE of VCI3M forecast for each week of the year. (a) GP	
747		forecasting on Landsat data. (b) AR forecasting on MODIS	
748		data. Grey shading indicate the rainy seasons, March-May and	
749		October-December.	18
750	7	(a) ROC curve for drought detection ($VCI3M < 35$) for lead	
751		times of 2, 4 and 6 weeks using the GP method on Landsat data.	
752		(b) ROC curve for drought detection using the AR method on	
753		MODIS data. (c, d) Respectively for the GP method on Landsat	
754		data and the AR method on the MODIS data, hit rate versus false	
755		alarm ratio for forecasting a transition to drought ($VCI3M < 35$)	
756		given that the vegetation condition is normal ($VCI3M > 35$) on	
757		the date of the forecast. The curves are plotted from applying	
758		different thresholds to convert the continuous forecast into a bi-	
759		binary forecast of drought or no drought, see text for details. The	
760		shaded circles show the point obtained from forecasting drought	
761		when the predicted $VCI3M < 35$	19
762	B.1	Contour plot of Landsat observed and predicted NDVI values	
763		from the GP interpolation.	8
764	B.2	Contour plot of MODIS observed and predicted NDVI values	
765		from 2000 pixels for gap-filling by quadratic interpolation.	9
766	C.3	Contour plots of NDVI anomaly against two, four and six weeks	
767		NDVI anomaly forecasts. (a,c,e) show forecast performance for	
768		the GP method on Landsat data, and (b,d,f) show forecast perfor-	
769		mance for the AR method on MODIS data (across the 19 regions	
770		for which a 4 week forecast was possible more than 50% of the	
771		time, see main text for details).	11

772	C.4	Forecast performance with a lead time of 1 to 10 weeks using	
773		the GP method on the Landsat data, as given by percentage	
774		standard deviation remaining S , for (Left) NDVI anomaly, and	
775		(Right) VCI3M. The blue lines show results for the individual	
776		regions, and the black line shows the median across all regions. .	12
777	C.5	Forecast performance with a lead time of 1 to 10 weeks using the	
778		AR method on the MODIS data, as given by percentage stan-	
779		dard deviation remaining, for (Left) NDVI anomaly, and (Right)	
780		VCI3M. The blue lines show results for the individual regions for	
781		which a forecast is possible more than 50% of the time, and the	
782		black line shows the median across all 19 of these regions.	12
783	C.6	Comparison of AR forecast with persistence forecast on the MODIS	
784		data. For lead times of 1 to 10 weeks, the RMSE of the AR fore-	
785		cast as a percentage of the RMSE of the persistence forecast.	
786		The blue lines show results for the individual regions for which a	
787		4 week forecast is possible more than 50% of the time, and the	
788		black line shows the median across these regions.	13
789	C.7	RMSE of 4 week forecast against percentage of clear pixels at	
790		most recent observation, for the AR method on the MODIS data.	
791		Plotted points are RMSE for each integer percentage of clear	
792		pixels. The Pearson correlation here is 0.01.	13
793	C.8	Map of Kenya showing the livelihood zones from which pixels	
794		were sampled.	14

795	C.9	Maps of NDVI anomaly and VCI3M 4 week forecast performance	
796		region-by-region for: (a) NDVI anomaly with GP method on	
797		Landsat data; (b) NDVI anomaly with AR method on MODIS	
798		data; (c) VCI3M with GP method on Landsat data; (d) VCI3M	
799		with AR method on MODIS data. In (a), asterisks indicate re-	
800		gions where selected pixels had a minimum of 180, rather than	
801		250, clean observations. (e) shows the percentages of weeks that	
802		the AR method provided a 4 week VCI3M forecast.	15
803	C.10	ROC curves for predicting drought with drought defined at var-	
804		ious NDMA thresholds. Possible hit rates against possible false	
805		alarm rates for the AR method on the MODIS data for the detec-	
806		tion of: (Top) Any drought, $VCI3M < 35$, (Middle) Severe or ex-	
807		treme drought $VCI3M < 20$, (Bottom) Extreme drought $VCI3M < 10$. 16	
808	C.11	Granger causality of VCI3M from each region to each other re-	
809		gion, computed on the MODIS data, measured as percentage re-	
810		duction in RMSE when observations from region 'From' are added	
811		to the AR model for forecasting region 'To' at a lead time of 4	
812		weeks. Only substantial Granger causalities are shown, i.e. those	
813		with percentage reduction in RMSE of more than 5%.	18
814	List of Tables		
815	1	Performance statistics of VCI3M forecasts with lead times of 2,	
816		4 and 6 weeks. Data for slope and intercept show ordinary least	
817		squares estimates $\pm$ standard error.	13

818	2	RMSE in VCI3M forecast, for the true vegetation condition be-	
819		longing to the different categories of drought, at lead times of 2,	
820		4 and 6 weeks. Drought categories are defined by the VCI3M	
821		index: wet by $VCI3M > 50$; normal by $35 < VCI3M < 50$; moderate	
822		drought by $20 < VCI3M < 35$; severe drought by $10 < VCI3M < 20$;	
823		and extreme drought by $VCI3M < 10$. (The extreme drought cri-	
824		terion was not met in any of the Landsat data.)	17
825	3	False alarm rate and hit rate (respectively, expressed in percent)	
826		for different regions in Kenya and at different lead times. This	
827		is based on forecasting drought if the predicted VCI3M is less	
828		than 35 (different performances could be obtained with different	
829		warning thresholds (see Figure 7. Regions are composed of the	
830		following zones: North – Z1,3 and 5; East – Z7, 9, 10 and 11 and	
831		South – (Z15 and 18))	20
832	A.1	Table comparing Landsat and MODIS products	2
833	B.2	Comparison of outcomes for different choices of maximum allowed	
834		interpolation length $L_{\max}$ on the MODIS data. R^2 -score of 4 week	
835		AR forecast and the percentage of the time that it was possible	
836		to make a forecast, for $L_{\max} = 4, 6$, and 8 weeks. Numbers show	
837		the median across all regions.	5
838	B.3	Comparison of GP method with commonly used interpolation	
839		methods as candidates for gap-filling on Landsat data. At the	
840		pixel level a random observation was removed, and then interpo-	
841		lated with each of the listed methods.	7
842	B.4	Comparison of interpolation methods as candidates for gap-filling	
843		on MODIS data.	9

844	C.5	Performance statistics of NDVI anomaly forecasts with lead times	
845		of 2, 4 and 6 weeks. Data for slope and intercept show ordinary	
846		least squares estimates $\pm$ standard error.	10

847 **Supplementary Material**

848 **Supplementary Material A. Data selection and comparison of datasets**

849 *Supplementary Material A.1. Landsat*

850 Landsat-5, 7 and 8 (Roy et al., 2014) red and near infrared (NIR) surface
851 reflectances and quality assessment (QA) data over the 10 pastoral livelihood
852 zones of Kenya, from January 1st, 2000 to February 1st, 2019, were obtained
853 using the United States Geological Survey (USGS) EarthExplorer. Specifically,
854 data from 1 000 pixels within each region were drawn from the Level-1 Precision
855 Terrain (L1TP) processed dataset, which has well-characterized radiometry and
856 is inter-calibrated across the different Landsat sensors. The spatial resolution
857 of these data is 30m and the repeat interval is 16 days. Using the QA data,
858 observations classified as clear from clouds or cloud shadows were kept. Pixels
859 with fewer than half of the observations over the full time period were discarded
860 (and replaced with an alternative random selection, with a few exceptions, see
861 Fig. C.9). The surface reflectances were combined to obtain NDVI.

862 *Supplementary Material A.2. MODIS*

863 NDVI data were also gathered from the surface reflectances obtained from
864 the daily, 500-meter resolution MODIS Terra/Aqua Nadir BRDF-Adjusted Re-
865 flectance product (MCD43A4,v006; Schaaf and Wang, 2015). Data from Febru-
866 ary 22nd, 2000 up to February 1st, 2019 were acquired via the NASA Land
867 Processes Distributed Active Archive Center. QA maps files with binary qual-
868 ity flags were used to remove poor quality data resulting from cloud or unreliable
869 BRDF corrections. Data were drawn from 100 pixels within each region, out of
870 those that had been identified as grassland by the MODIS land cover classifica-
871 tion maps (MCD12Q1,v006).

872 *Supplementary Material A.3. Comparison of the two datasets*

873 The key differences between the two datasets are the spatial and temporal
874 resolutions, see Table A.1. The Landsat data have higher spatial resolution,
875 whilst the MODIS data have higher temporal resolution. Since forecasting was
876 being attempted at the level of large scale regions (livelihood zone and county
877 intersections), and at a weekly temporal resolution, the expectation was that the
878 MODIS data would have advantages, assuming individual Landsat and MODIS
879 observations have similar signal-to-noise ratios. The processed MODIS time
880 series with weekly observations have less measurement noise because they are
881 composites of 7 daily observations (that themselves are 16-day composites of
882 measurements taken every 1-2 days), whereas the processed Landsat time series
883 are derived from more temporally sparse data (up to 3 different Landsat mis-
884 sions, each yielding one observation every 16 days). Landsat data would have
885 advantages in different applications where forecasts on smaller spatial scales are
886 required. The Landsat data also has the advantage that the quality flags and
887 cloud masks are defined on smaller scales.

888 The differences between the MODIS and Landsat datasets produced slightly
889 different ‘true’ aggregate time series on which to assess the interpolation and
890 forecasting methods. In addition to the different temporal resolution of the
891 observations supplying the final time series, the MODIS data were aggregated
892 across 100 random grassland pixels from each region, whereas the 1 000 Landsat

Table A.1: 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

893 pixels analysed were randomly distributed over the whole of each region. In
 894 choosing how many pixels to analyse per region, there is a trade-off between
 895 using a larger number of pixels for higher accuracy, and a smaller number of
 896 pixels for lower computational cost. Fewer MODIS pixels were used than Land-
 897 sat pixels since they correspond to larger spatial regions. Both these choices of
 898 number of pixels should be sufficient for high accuracy of results, since for Land-
 899 sat data the R^2 -score comparing the average of all pixels from a region with the
 900 average of 100 or 1 000 random pixels was 0.990 and 0.9993 respectively. The
 901 MODIS grassland classification was not available at Landsat resolution, thus un-
 902 ambiguous classification of the smaller Landsat pixels was not possible. This is
 903 unlikely to have made much difference to pixel selection, given that the pastoral
 904 livelihood zones are mostly grasslands (Fig. C.8).

905 **Supplementary Material B. Further details on preprocessing**

906 *Supplementary Material B.1. Gaussian process modelling*

907 A Gaussian Process is a probabilistic model defined as a collection of ran-
 908 dom variables for which any finite subset has a joint Gaussian distribution (Ras-
 909 mussen and Williams, 2006). Formally, for the present application of interpo-
 910 lation or extrapolation of a time series, with observation at time t denoted by
 911 X_t , the model is

$$X_t \sim \mathcal{N} [Y(t), \sigma_r^2] , \quad (\text{B.1})$$

$$Y(t) \sim \mathcal{GP} [m(t), k(t, t')] . \quad (\text{B.2})$$

912 Here $Y(t)$ is the true value of the observed index, and the measurement noise
 913 is σ_r , so that an observation X_t is a normal random variable with mean $Y(t)$

914 and standard deviation σ_r . The true values $Y(t)$ are also normally distributed,
 915 with the mean at time t given by the mean function $m(t)$, and the covariance
 916 between values at times t and t' given by the kernel function $k(t, t')$. To carry
 917 out interpolation or extrapolation from a time series, existing data are used to
 918 fit the mean, m , kernel, k , and measurement noise σ_r , and then expected values
 919 are produced for the desired times, based on the obtained fit.

920 For gap-filling on individual Landsat pixel NDVI time series, the model was
 921 determined as follows, using the `Pyro` programming package for `Python`. The
 922 mean, $m(t)$, was assumed to be constant, and the mean of the whole time series.
 923 To determine the kernel, Compositional Kernel Search (Duvenaud et al., 2013)
 924 was used. Specifically, a search through all the following kernels, and products
 925 and sums of pairs of them was carried out: Linear, Radial Basis Function (RBF),
 926 Periodic (with period p set to one year), Rational Quadratic, and Matern. The
 927 highest marginal likelihood was achieved by Radial Basis Function (RBF) plus
 928 Periodic ($k_{\text{RBF}} + k_{\text{P}}$), so this combination was selected as the kernel:

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

$$k_{\text{P}}(t, t') = \sigma_{\text{P}}^2 \exp \left(-2 \frac{\sin^2(\pi|t - t'|/p)}{l_{\text{P}}^2} \right). \quad (\text{B.4})$$

929 There were thus 5 parameters to fit for each time series ($\sigma_r, \sigma_{\text{RBF}}, l_{\text{RBF}}, \sigma_{\text{P}}, l_{\text{P}}$).
 930 These were learned using Stochastic Variational Inference (Hoffman et al., 2013).

931 For the forecasting on the aggregated NDVI anomaly and VCI3M, a pure
 932 Radial Basis Function kernel was used, since for these anomaly indices, the
 933 periodic component is not present, see Section 3 in the main manuscript.

Table B.2: Comparison of outcomes for different choices of maximum allowed interpolation length $L_{\max}$ on the MODIS data. R^2 -score of 4 week AR forecast and the percentage of the time that it was possible to make a forecast, for $L_{\max} = 4, 6$, and 8 weeks. Numbers show the median across all regions.

$L_{\max}$ (weeks)	R^2 -score	Forecasts attempted (%)
4	0.60	84
6	0.58	93
8	0.63	98

934 *Supplementary Material B.2. Gap-filling for MODIS*

935 Interpolation of gaps in the raw MODIS time series was not carried out when
936 the length of the gap was longer than a certain maximum, $L_{\max}$. In choosing
937 $L_{\max}$, a trade off between quality and quantity of remaining observations had
938 to be made: a small $L_{\max}$ would lead to fewer forecasts being attempted, but
939 interpolations closer to the ground truth, while a large $L_{\max}$ would lead to
940 more forecasts being attempted, but with these forecasts being assessed against
941 interpolations that are potentially far from the ground truth. The choice $L_{\max} =$
942 6 weeks was made, after exploring a range of values and finding R^2 -score to be
943 not sensitive to the precise choice within the range between 4 and 8 weeks, see
944 Table B.2. Note that interpolation on the Landsat data was carried out for all
945 gaps, since the GP interpolation method makes use of the entire time series, and
946 interpolated values within a long interpolation take values close to the seasonal
947 mean.

948 Due to the presence of non-interpolated gaps in the MODIS time series, there
949 were weeks when a forecast assessment was not carried out on these data. The
950 criteria for being able to do AR forecasting on these data were: (i) the three
951 most recent weekly aggregated observations had to be present, since these are
952 required for making a prediction; (ii) there had to be an aggregated observation
953 present for the week being forecast, so the quality of the prediction could be

954 assessed.²

955 *Supplementary Material B.3. Comparison of other possible gap-filling methods*

956 Various gap-filling methods have been used to deal with missing values re-
957 sulting from the presence of clouds and atmospheric aerosols. These methods
958 are based on either spatial information, temporal information or some combina-
959 tion of both spatial and temporal information (Kandasamy et al., 2013; Weiss
960 et al., 2014). Temporal interpolation was chosen given that spatial interpolation
961 methods suffer from the fact that there are frequently clouds over Kenya that
962 cover large groups of neighbouring pixels (although a possible alternative, not
963 considered here, would be to make use of other pixels that historically behave
964 similarly in time (Cao et al., 2018)).

965 The performance of the temporal gap-filling methods employed, compared
966 with alternative temporal gap-filling methods, was tested by removing observa-
967 tions, applying the method, and then comparing the interpolated observations
968 with the removed observations. GP interpolation and linear, quadratic and
969 cubic polynomial interpolation methods were tested, on both the Landsat and
970 MODIS datasets. R^2 -scores were obtained for using the interpolated values to
971 predict the ‘true’ values for the missing observations.

972 For the Landsat data, one randomly chosen observation between 1/1/2014
973 and 1/2/2019 was removed from each of 2000 randomly selected individual pixel
974 time series. From the MODIS data, 2000 random individual pixel NDVI time
975 series (1/1/2014 to 1/2/2019) were chosen. 20 randomly selected NDVI values
976 were dropped from each of the time series and the various gap-filling methods
977 were used to interpolate the dropped values. The results for Landsat are shown
978 in Table B.3, and for MODIS in Table B.4. Note that with these methods, the

²GP 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.3: 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

random samples are more likely to come from periods when there are not many gaps. It is an assumption that the results are valid across all periods.

For the Landsat data, the GP method achieved the highest R^2 -score, thus showing its utility, and justifying our choosing it. The R^2 -score of 0.67, achieved by the GP method, is close to the R^2 -score of 0.76 which is obtained from using one Landsat observation to predict a second Landsat observation from the same 16-day observation period (of which there were instances in the data). Fig. B.1 shows a contour plot of the true versus interpolated NDVI observations using this method. This plot shows that the method doesn't introduce any biases- the slope and intercept are approximately 1 and 0 respectively.

For the MODIS data, GP, linear interpolation and quadratic interpolation all performed similarly well. Quadratic interpolation had the highest R^2 -score, hence this method was chosen for gap-filling on the MODIS data. The higher interpolation R^2 -scores for MODIS, compared to Landsat, imply that the MODIS data is less noisy than the Landsat data. Assuming that observations from MODIS and Landsat have similar signal-to-noise ratio, this can be explained by the higher temporal resolution of MODIS, and the compositing of multiple observations for the weekly gridded MODIS data. Fig. B.2 shows a contour plot of the true versus interpolated NDVI observations using the quadratic interpolation method. This again demonstrates that the interpolation doesn't introduce

Figure B.1: Contour plot of Landsat observed and predicted NDVI values from the GP interpolation.

Table B.4: 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

999 biases- the slope and intercept are approximately 1 and 0 respectively.

Figure B.2: Contour plot of MODIS observed and predicted NDVI values from 2000 pixels for gap-filling by quadratic interpolation.

Table C.5: Performance statistics of NDVI anomaly forecasts with lead times of 2, 4 and 6 weeks. Data for slope and intercept show ordinary least squares estimates $\pm$ standard error.

	Landsat GP			MODIS AR		
	2	4	6	2	4	6
	weeks			weeks		
R^2 -score	0.69	0.46	0.27	0.85	0.55	0.33
RMSE	0.029	0.039	0.045	0.025	0.043	0.053
slope	1.1 $\pm$ 0.0	1.2 $\pm$ 0.0	1.4 $\pm$ 0.0	1.0 $\pm$ 0.0	1.0 $\pm$ 0.0	1.0 $\pm$ 0.0
intercept	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0

Supplementary Material C. Further forecast results

Fig. C.3 shows contour plots of forecast against actual NDVI anomaly data for the two methods, and Table C.5 shows the R^2 -scores, RMSE, slope and intercept from each of these plots. Figs. C.4 and C.5 plot the forecast performance of the two methods in terms of percentage of standard deviation remaining S , for lead times of 1 to 10 weeks. For NDVI anomaly, for both methods, S approaches the baseline of 100 as the lead time approaches 10 weeks, while for VCI3M, some forecast skill is still apparent at a lead time of 10 weeks. Fig. C.6 compares the performance of the AR VCI3M forecast with that of the persistence VCI3M forecast, on the MODIS data; the persistence forecast being simply the most recent observation. The AR forecast performs substantially better than the persistence forecast, for example, achieving a RMSE of approximately half that of the persistence forecast for a lead time of 4 weeks. The GP VCI3M forecast on the Landsat data achieves a similar improvement on the persistence forecast. Fig. C.7 shows, for the MODIS/AR method, the average RMSE of a 4 week forecast against the percentage of pixels from which there was a clear observation during the week the forecast was made. Fig. C.9 shows forecast performance region by region. Fig. C.10 shows alternative ROC curves for drought prediction using the AR method on the MODIS data, based on different thresholds for defining drought.

Figure C.3: Contour plots of NDVI anomaly against two, four and six weeks NDVI anomaly forecasts. (a,c,e) show forecast performance for the GP method on Landsat data, and (b,d,f) show forecast performance for the AR method on MODIS data (across the 19 regions for which a 4 week forecast was possible more than 50% of the time, see main text for details).

Figure C.4: 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, and the black line shows the median across all regions.

Figure C.5: 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 C.6: Comparison of AR forecast with persistence forecast on the MODIS data. For lead times of 1 to 10 weeks, the RMSE of the AR forecast as a percentage of the RMSE of the persistence forecast. The blue lines show results for the individual regions for which a 4 week forecast is possible more than 50% of the time, and the black line shows the median across these regions.

Figure C.7: RMSE of 4 week forecast against percentage of clear pixels at most recent observation, for the AR method on the MODIS data. Plotted points are RMSE for each integer percentage of clear pixels. The Pearson correlation here is 0.01.

Figure C.8: Map of Kenya showing the livelihood zones from which pixels were sampled.

Figure C.9: Maps of NDVI anomaly and VCI3M 4 week forecast performance region-by-region for: (a) NDVI anomaly with GP method on Landsat data; (b) NDVI anomaly with AR method on MODIS data; (c) VCI3M with GP method on Landsat data; (d) VCI3M with AR method on MODIS data. In (a), asterisks indicate regions where selected pixels had a minimum of 180, rather than 250, clean observations. (e) shows the percentages of weeks that the AR method provided a 4 week VCI3M forecast.

Figure C.10: ROC curves for predicting drought with drought defined at various NDMA thresholds. Possible hit rates against possible false alarm rates for the AR method on the MODIS data for the detection of: (Top) Any drought, $VCI3M < 35$, (Middle) Severe or extreme drought $VCI3M < 20$, (Bottom) Extreme drought $VCI3M < 10$.

1020 *Supplementary Material C.1. Effect of including observations from other re-*
1021 *gions in the AR model*

For the MODIS data, we tested to see whether we could improve the prediction of VCI3M by incorporating the past of VCI3M from a distinct region in the AR model, i.e. Granger causality analysis was performed. Taking X as the VCI3M of the region to be forecast, as in equation (3), and Y to be the VCI3M from another region, the extended model was fit:

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

1022 and Granger causality measured as the percentage reduction in RMSE obtained
1023 when the extended model is used instead of the previous (reduced) model (3).

1024 We tested for Granger causality of VCI3M from each region to each other
1025 region (within the set of regions for which predictions could be made more than
1026 50% of the time). That is, for each pair of distinct regions, i and j , the 3
1027 most recent observations from region j were added to the AR forecast model
1028 for region i , and the RMSE was compared with that obtained without including
1029 observations from region j . There was not strong Granger causality of VCI3M
1030 between most regions. For only a few combinations was there a reduction in
1031 RMSE of more than 5%, see Fig. C.11. Nevertheless, these results suggest that,
1032 to create the optimal linear regression based forecasting method, data from all
1033 regions should be used. Future work will explore how best to extract any useful
1034 information from regions other than the one being forecast.

1035

Figure C.11: Granger causality of VCI3M from each region to each other region, computed on the MODIS data, measured as percentage reduction in RMSE 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 percentage reduction in RMSE of more than 5%.