Address for correspondence:
Anton Vrieling
University of Twente – Faculty ITC
P. O. Box 217
7500 AE Enschede
The Netherlands
T: +31 53 4874452
E: a.vrieling@utwente.nl

Historical extension of operational NDVI products for livestock insurance in Kenya

Anton Vrieling[a], Michele Meroni[b], Apurba Shee[c], Andrew G. Mude[c], Joshua Woodard[d], Kees de Bie[a], Felix Rembold[b]

[a] University of Twente, Faculty of Geo-information Science and Earth Observation, P.O. Box 217, 7500 AE Enschede, The Netherlands. E-mail: a.vrieling@utwente.nl; c.a.j.m.debie@utwente.nl

[b] Institute for Environment and Sustainability, Joint Research Centre, European Commission, Via E. Fermi 2749, I-21027 Ispra (VA), Italy. E-mail: michele.meroni@jrc.ec.europa.eu; felix.rembold@jrc.ec.europa.eu

[c] International Livestock Research Institute, P.O. Box 30709, Nairobi 00100, Kenya. E-mail: a.shee@cgiar.org; a.mude@cgiar.org

[d] Cornell University, Dyson School of Applied Economics and Management, 236 Warren Hall, Ithaca, NY 14853, United States of America. E-mail: joshua.woodard@cornell.edu

October 2013, submitted to JAG

December 2013, revised version submitted to JAG

Abstract

Droughts induce livestock losses that severely affect Kenyan pastoralists. Recent index insurance schemes have the potential of being a viable tool for insuring pastoralists against drought-related risk. Such schemes require as input a forage scarcity (or drought) index that can be reliably updated in near real-time, and that strongly relates to livestock mortality. Generally, a long record (>25 years) of the index is needed to correctly estimate mortality risk and calculate the related insurance premium. Data from current operational satellites used for large-scale vegetation monitoring span over a maximum of 15 years, a time period that is considered insufficient for accurate premium computation. This study examines how operational NDVI datasets compare to, and could be combined with the non-operational recently constructed 30-year GIMMS AVHRR record (1981-2011) to provide a near-real time drought index with a long term archive for the arid lands of Kenya. We compared six freely available, near-real time NDVI products; five from MODIS, and one from SPOT-VEGETATION. Prior to comparison, all datasets were averaged in time for the two vegetative seasons in Kenya, and aggregated spatially at the administrative division level at which the insurance is offered. The feasibility of extending the resulting aggregated drought indices back in time was assessed using jackknifed R^2 statistics (leave-one-year-out) for the overlapping period 2002-2011. We found that division-specific models were more effective than a global model for linking the division-level temporal variability of the index between NDVI products. Based on our results, good scope exists for historically extending the aggregated drought index, thus providing a longer operational record for insurance purposes. We showed that this extension may have large effects on the calculated insurance premium. Finally, we discuss several possible improvements to the drought index.

Keywords: NDVI, AVHRR, SPOT, MODIS, index insurance, intercalibration

1 Introduction

Coping with drought is a major challenge for pastoralists in the arid and semi-arid parts of Kenya (Little et al., 2001; Nkedianye et al., 2011). During dry years many animals die because of insufficient feed and water, and from drought-related epidemic diseases (Onono et al., 2013). Such losses can have severe, long-term consequences on pastoralist households if their herd sizes fall below specific thresholds (Barrett et al., 2006).

Insurance against the risk of livestock mortality may reduce the negative consequences of drought-induced livestock loss, and avoid families falling into poverty (Chantarat et al., 2013). As opposed to traditional agricultural insurance, requiring expensive verification of individual losses by the insurer, a more cost-effective insurance approach is to base payouts on a transparent and objectively measured variable, such as total seasonal rainfall (Barnett et al., 2008). This is referred to as index-based insurance. Recently, index-based insurance received much attention as it could make important contributions to agricultural growth and reduction of poverty (Hazell and Hess, 2010; Brown et al., 2011). Despite concerns regarding the demand for insurance by poor farmers (Binswanger-Mkhize, 2012), and challenges of reaching sufficient scale among numerous pilot projects, the risk-management potential that index insurance could offer poor farmers fuels continued interest and efforts to improve product design (Barrett et al., 2007; Barnett et al., 2008).

A main limitation to index-based insurance is the possibility for households to experience a loss, but no payment, or alternatively not experience a loss, but yet receive a payment (Barnett et al., 2008). This is referred to as ‘basis risk’ and is caused by the imperfect relationship between the index and incurred losses. For index-based insurance schemes to be effective, they require an index that:

1) strongly correlates with what is insured (such as livestock or crop losses);
2) is independently verifiable, i.e. based on well-described data sources and processing methods;
3) can reliably be delivered into the future (at least for the duration of the insurance contract) and is available in near real-time, so that shortly after losses are incurred, payments can be made;
4) is available for sufficiently long records to properly represent the climatic variability for estimating the probability of a payout (Bell et al., 2013), and thus accurately pricing of the insurance product.

Time series of the normalized difference vegetation index (NDVI) have been used for the purpose of index-based insurance (Turvey and McLaurin, 2012; Leblois and Quirion, 2013). A number of near real-time composite NDVI products are freely available from sensors such as MODIS (Moderate Resolution Imaging Spectroradiometer) and SPOT-VGT (Système Pour l'Observation de la Terre - VEGETATION). These sensors offer a relatively coarse spatial resolution (250-1000m), but provide observations of the same area on a daily basis. This last aspect is important to reduce cloud and atmospheric effects in the composite products, and to effectively compare vegetation conditions within and between years. Given that droughts generate spatially-correlated covariate risks that simultaneously affect a larger number of neighbouring households, pixel-level NDVI values are generally spatially aggregated. In most cases this aggregation is also a necessity for modelling crop and livestock losses, because data on production or mortality are often only available for administrative regions. As a consequence, each administrative unit has different premium specifications and payouts are equal for all insurance customers within a given unit.

In the absence of reliable station rainfall data, the index-based livestock insurance (IBLI) project in Kenya uses NDVI as a proxy for forage scarcity – a key determinant of livestock mortality in pastoral production systems (Chantarat et al., 2013). The insurance design for the Marsabit district of northern Kenya was extensively described by Chantarat et al. (2013). While they used rectangular clusters, the IBLI project currently uses administrative divisions for spatial aggregation. Since 2010, the IBLI project has operated in the Marsabit district, and between 2013 and 2014 the project plans to expand to cover about 60 per cent of Kenya's land surface that constitutes its so-called arid lands. In the four years (eight seasons) during which pastoralists in Marsabit have purchased insurance they received three times insurance payouts following drought. For operational purposes MODIS was selected as the main data source, following the suspension in the delivery of AVHRR (Advanced Very High Resolution Radiometer) NDVI composites by the Famine Early Warning Systems Network (FEWS-NET) due to the degradation of the NOAA-17 AVHRR sensor. A main drawback of MODIS is that it covers only the years 2000 to present, hence insufficient to capture the full range of climatic variability and the related drought probability, needed to properly price insurance contracts. Uncertainties regarding this probability due to data restrictions would lead insurers to add risk-loading to the premium prices, thus making the insurance more expensive and consequently less attractive to pastoralists (Biener, 2013).

The creation of a long-term consistent NDVI time series from multiple sources is not a trivial task due to differences in sensor characteristics and algorithms used to generate products (Miura et al., 2006). Differences in spectral response functions between sensors are a key characteristic responsible for the variation in NDVI (Trishchenko et al., 2002; Trishchenko, 2009). Based on spectral convolution of hyperspectral Hyperion data, Miura et al. (2006) reported that the NDVI relationship among MODIS, AVHRR and ETM+ instruments is non-

linear and largely dependent on how much the green peak (550 nm) and red edge (680-780 nm) regions are included in the red band. Despite that they find near-linear NDVI relationships by direct comparison of AVHRR and MODIS (in correspondence to the empirical study by Gallo et al., 2005), they indicate that higher-order polynomials may be more accurate in modelling cross-sensor NDVI relationships. Additional factors that cause cross-sensor variability of NDVI include atmospheric and bi-directional reflection effects, which are also wavelength dependent (Myneni and Asrar, 1994; Sandmeier et al., 1998). This combination of factors complicates a straightforward joining of NDVI series derived from multiple sensors.

Many attempts have been made to construct a single long-term NDVI record from AVHRR sensors onboard multiple satellites, which effectively corrects for effects like sensor degradation, orbital drift, and atmospheric variability (James and Kalluri, 1994; Tucker et al., 2005). Recently, the Global Inventory Monitoring and Modeling System (GIMMS) project released a 30-year record of the so-called NDVI3g, i.e., third generation GIMMS NDVI from AVHRR sensors. While effectively combining data from various AVHRR sensors already presents a big challenge, spectral response functions are even more dissimilar in comparison to SPOT-VGT and MODIS that have narrower spectral bands (Gao, 2000). Proposed corrections include empirically-derived linear functions (Steven et al., 2003; Gallo et al., 2005; Song et al., 2010) and second-order polynomial regression equations (Trishchenko et al., 2002). Swinnen and Veroustraete (2008) found a strong linear relationship between SPOT-VGT and 1-km^2 AVHRR NDVI for Southern Africa after rigorous reprocessing of spectral reflectance data using the same atmospheric correction and compositing approaches. They effectively accounted for differences in the dynamic range between SPOT-VGT and AVHRR using the adjustment functions of (Trishchenko et al. (2002)). Alternatively, neural

networks, which incorporated data layers reflecting atmospheric conditions, have been used to account for the differences between AVHRR and MODIS (Brown et al., 2008). However, despite various suggestions regarding the achievability of an intercalibrated, sensor-independent NDVI record (e.g. Steven et al., 2003; Brown et al., 2006), and recent efforts towards delivering this to the public (Pedelty et al., 2007; Gutman and Masek, 2012), no universally-accepted multi-sensor NDVI record exists to date that both covers a long (> 25 year) time frame and is available in near real-time.

The aim of this study is to provide a pragmatic solution for combining NDVI composite products derived from multiple sensors (i.e. AVHRR, MODIS, and SPOT-VGT) for the purpose of the livestock insurance programme in Kenya. Rather than analysing cross-sensor NDVI differences per pixel and composite period, we first aggregate the NDVI in space and time to provide an appropriate index in the framework of the IBLI project. This implies aggregation over administrative divisions and for two periods within each year, corresponding to the two growing seasons occurring in the region. We first evaluate if a global regression model (taking all divisions and periods together) can accurately map the aggregated index from one NDVI product to another. As our overall purpose is to have a long record that accurately displays drought-related risk for each administrative unit, which can be updated in near real-time and serve as an input to model livestock mortality, we subsequently perform a cross-sensor comparison at the division level, considering the two seasons together and separately, to examine if this increasing level of disaggregation improves the intercalibration performances with respect to the global model. Besides comparing merely with the non-operational historic AVHRR record, we also compare operational products to evaluate to what extent these datasets can be used interchangeably. This last issue may be important in case of satellite sensor failure in the future. Finally we evaluate if and how the

availability of a longer intercalibrated time series will affect the premium rate for the livestock insurance product.

2 Study area

The study area comprises the nine counties of Kenya that are planned to be covered by the IBLI project over the next one to two years, and are referred to by the Government of Kenya as the arid lands (Figure 1). Since 1996, the government has collected household-level livestock mortality data in representative locations across the study area, in the framework of the Arid Land Resource Management Project (ALRMP, http://www.aridland.go.ke). The nine counties together cover approximately 62 per cent of Kenya's land area. According to Peel et al. (2007), the area contains three Köppen-Geiger climate zones in approximately equal amounts, i.e., tropical savannah climate (Aw), hot steppe climate (BWh), and hot desert climate (BSh). Based on 1998-2012 data of the Tropical Rainfall Measurement Mission (3B43 product), average annual rainfall ranges from less than 300 mm in the dry parts of Isiolo, Marsabit, Turkana, and Wajir Counties, to more than 1,000 mm only in the south-western part of Baringo County. Two rainfall seasons can be discerned: the so-called long rains (March-May) and the short rains (October-December) separated by clear dry seasons. Following Chantarat et al. (2013), we term this bi-modal seasonal pattern as Long Rains Long Dry (LRLD) covering March to September and Short Rains Short Dry (SRSD) covering October to February. Livestock keeping is the main rural livelihood in the region. Livestock includes camels (in the driest parts), goats, sheep, and cattle. To standardize across the livestock types, and to facilitate the development of a single livestock-based insurance product, livestock numbers owned by households are expressed in Tropical Livestock Units (TLU); 1 cattle equals 1 TLU, 1 camel is 1.4 TLU, and a goat or sheep equals 0.1 TLU.

Our analysis focussed on the division-level, as this is the basic unit for which insurance premium and payout are determined. The nine counties comprise 108 divisions. Given the small size of some divisions and the consequent difficulty of obtaining a representative division-level drought index, especially from the 8-km resolution AVHRR series, we set a minimum threshold for division size. Starting from the smallest division, we iteratively aggregated divisions smaller than 1,000 km^2 to the neighbouring division within the same county that had the nearest centroid coordinates. This resulted in 84 spatial units that we further refer to in this paper simply as divisions. The red lines in Figure 1 show the resulting division boundaries.

3 NDVI data sets

To select potential sources of operational NDVI time series data we considered the two following criteria: i) archive and near-real time data should be freely available, and ii) no or minimum post processing should be required to facilitate their use by less-specialized users. As a result, a non-exhaustive list of six operational products was compiled: five derived from MODIS instruments onboard Terra and Aqua platforms, and one from SPOT-VGT. In addition, the new long-term non-operational dataset derived from AVHRR (NDVI3g) was used to create a longer historic record. The main characteristics of the products are summarized in Table 1, and Appendix 1 provides a detailed description of each.

4 Methods

4.1 NDVI processing

For the unfiltered datasets, i.e., GIMMS, SPOT-VGT, $MODIS_{T\text{-}NASA}$, and $MODIS_{A\text{-}NASA}$ (Table 1), we applied an iterative Savitzky-Golay filter (Savitzky and Golay, 1964) as described by Chen et al. (2004) to reduce remaining atmospheric effects in the time series. To do that, we first created a mask to discard any NDVI values that were cloudy or otherwise of poor quality. For this we used the quality information delivered with the SPOT-VGT, $MODIS_{T\text{-}NASA}$, and $MODIS_{A\text{-}NASA}$ data (Appendix A), while for GIMMS we masked out any NDVI values below 0 and with an increase of more than 0.30 in 15 days. The filter was subsequently applied using a third-order polynomial and a moving window of three observations prior to, and after the data point to be filtered. Visual analysis of the resulting time series showed that this procedure substantially reduced noise in the series, effectively interpolated missing values, while retaining short-term variations that relate to real changes in greenness.

Besides introducing the temporal filtering, we further adapted the NDVI processing sequence from the original IBLI design (Chantarat et al., 2013) to provide improved metrics of the season performance, which should also allow for better comparison between different sensors. In the original design, Chantarat et al. (2013) first transformed the 10-daily NDVI images to standard scores (or z-scores). The z-scored NDVI indicates how many standard deviations the pixel's NDVI is above or below the multi-annual mean pixel value of the same 10-day period (e.g., 1-10 January). They then spatially aggregated the z-scored NDVI, and subsequently cumulated the aggregated values over time for two periods, i.e., long rains-long dry (LRLD, March-September) and short rains-short dry (SRSD, October-February). The idea behind aggregating z-scores of 10-day periods is that adverse forage conditions may

occur at any time during the season; however, forage is not produced during the entire season (as defined by LRLD and SRSD). A drawback of directly calculating z-scores for each time step is that small deviations during relatively dry moments of the season can translate to large z-scores, which get equally weighted with smaller z-scores during wet moments (that can however represent stronger absolute deviations) when cumulating over time. To prevent this problem and get a better measure of seasonal forage production, we first performed temporal aggregation, then spatial aggregation, and finally z-scoring.

We performed temporal aggregation for each pixel for both LRLD and SRSD. The average seasonal NDVI was used for this aggregation, which is in the temporal context functionally similar to the cumulative NDVI value, a suitable proxy of seasonal biomass production (e.g., Bonifacio et al., 1993; Funk and Budde, 2009). The advantage of using the average compared to the cumulative value is that 1) values for the two seasons of different length are in the same units and range, and 2) it is insensitive to the different length of the compositing period of the different NDVI products. We then spatial aggregated the temporally-averaged NDVI by calculating the average value for each division (see section 2 on the divisions used). Given the coarse resolution of GIMMS, for this product we calculated a weighted average that reflects the amount of overlap a pixel has with a division.

While the z-scored values are the input for calculating insurance premiums (section 4.4), the basis in this paper for the intercalibration between NDVI products are the NDVI values, aggregated in space and time. We further refer to them as *NDVI**.

4.2 Intercalibration

To evaluate if different NDVI products perform similarly in identifying division-level drought conditions, all data comparisons are based on *NDVI**. For all dataset combinations, we compared results using only the overlapping period between the seven datasets, i.e. the period between July 2002 and December 2011. This period contains a total of 18 seasons, i.e. 9 LRLD seasons and 9 SRSD seasons.

Given the relatively small sample size available for intercalibration we limited our analysis to the linear component of the relationship between NDVI products. Visual inspection of scatterplots (Figure 2) and the residuals following linear regression (data not shown) suggest a slight deviation from linearity between GIMMS-derived *NDVI** and *NDVI** derived from different sensors. However, given the relatively small sample size available for intercalibration we limited our analysis to the linear relationship between NDVI products. We tested three calibration models that use different levels of pooling of the division-level and season-level *NDVI** data (Equations 1-3, discussed below). The aim of this was to evaluate 1) which NDVI-products show highest correlation with the long-term GIMMS dataset, 2) which level of pooling across season and space is most efficient in transforming *NDVI** between one source and another, 3) for which divisions/regions in Kenya the various products lead to the same seasonal vegetation condition (and hence drought) assessment.

4.2.1 Global model (DSP)

We first assessed the performances of a global calibration model (combining all divisions and seasons) in translating GIMMS-derived *NDVI** to the *NDVI** obtained from other NDVI products, and evaluated which operational product yielded the closest agreement. We refer to this model as the *DSP*-model, meaning "Divisions and Seasons Pooled". It takes the following form:

$$NDVI\_M^*_{d,s} = \beta_0 + \beta_1 * NDVI\_S^*_{d,s} + \varepsilon_{d,s} \quad (1)$$

where $NDVI\_M^*_{d,s}$ is the average *NDVI* for division *d* and season *s* for the NDVI-series that is used as the master (or dependent variable), while $NDVI\_S^*_{d,s}$ is the average *NDVI* for a division that will be mapped to the master (i.e., the slave, or independent variable). For example, to create longer time series for GIMMS that are compatible with MODIS, GIMMS is considered the slave and MODIS the master. The parameters β_0 and β_1 are the regression coefficients to be estimated and ε is error term. The global model is parsimonious in terms of number of parameters to be estimated (i.e., two with a sample size of 18 seasons x 84 divisions).

4.2.2 Division-specific season-pooled model (SP)

Despite its parsimonious nature, the *DSP*-model may not be able to model division-level specificities in the relationship between products. Cross-division differences may arise because of different NDVI dynamic ranges interacting in a complex way with sensor-specific NDVI saturation levels, different soil background affecting NDVI as a result of sensor-specific spectral response functions, and finally, interaction of local climatology with differences in NDVI processing chains (such as cloud screening and atmospheric correction) affecting locally the relationship between products. These issues justify the evaluation of a less parsimonious division-specific regression model (two parameters to be estimated with a sample size of 18 seasons, for a total of 2x84 parameters), referred to here as *SP*-model ("Season Pooled"). The *SP*-models can be written as:

$$NDVI\_M^*_{d,s} = \beta_{0,d} + \beta_{1,d} * NDVI\_S^*_{d,s} + \varepsilon_{d,s} \quad (2)$$

where the parameters $\beta_{0,d}$ and $\beta_{1,d}$ are now the division-specific regression coefficients to be estimated.

4.2.3 Division-specific season-specific model (NP)

Finally, in order to evaluate if any season-specific effect on the relationship is present, we also evaluated at the division level if separating the LRLD and SRSD seasons improves our regression estimates. This is referred to here as the *NP*-models (for "No Pooling") and, in terms of number of parameters, it is the least parsimonious model considered (two parameters to be estimated with a sample size of nine seasons, for a total of 2x2x84 parameters). The *NP*-model can be expressed as:

$$NDVI\_M^*_{d,s} = \beta_{0,d,s} + \beta_{1,d,s} * NDVI\_S^*_{d,s} + \varepsilon_{d,s} \quad (3)$$

where $\beta_{0,d,s}$, $\beta_{1,d,s}$ are now specific for each combination of division and season.

4.3 Performance evaluation

The increased level of specificity going from the *DSP*-model, via *SP*-, to *NP*-models is achieved at the expense of a reduced sample size on which the model is calibrated, giving rise to a trade-off between the capacity of the calibration model to take spatial heterogeneity into account and data availability. In fact, although the performances in fitting increase by definition when a more specific model is employed, this may not happen in prediction because of model overparameterization. Overparameterization occurs when the amount of information contained in the calibration data is not enough to estimate the model parameters. The resulting model fits the calibration dataset, but produces large errors when used in prediction. Conversely, underparameterization refers to a situation in which the available information is not fully exploited by the restricted set of model parameters. Therefore,

over/under-parameterization must be minimized to achieve the best predictive capacity. In order to choose the best modelling solution with the data at hand, we assessed the prediction performance of different NDVI product pairs for the different regression options using a cross-validation jackknifing technique, where one full year of data was left out at a time.

For each jackknifed year, regression coefficients were estimated on the retained dataset and subsequently applied to estimate $NDVI\_M^*_{d,s}$ of the year left apart. Performances were then evaluated using the cross-validated R^2 (i.e., R^2_{cv}). The R^2_{cv} measures the fraction of total NDVI variability that is explained by the model in prediction, in all the dimensions of the database under consideration. For example, for the global *DSP*-model (Equation 1) the total variability is characterised by the spatial (division), seasonal, and interannual dimensions. As the main objective of our intercalibration is to accurately reconstruct the interannual variability of *NDVI** at division and seasonal level, we are not interested in the ability of our model to explain the variability in the spatial and seasonal dimensions. For the *DSP*-model evaluation, we therefore compute the R^2_{cv} within division and season ($R^2_{cv(wd,ws)}$), informing us on the temporal prediction capability only. This can be expressed as:

$$R^2_{cv(wd,ws)} = 1 - \frac{\sum_i^I \sum_d^D \sum_s^S \left(NDVI\_M^*_{i,d,s} - \widehat{NDVI\_M}^*_{i,d,s}\right)^2}{\sum_i^I \sum_d^D \sum_s^S \left(NDVI\_M^*_{i,d,s} - \overline{NDVI\_M^*_{d,s}}\right)^2} \quad (4)$$

where $\widehat{NDVI\_M}^*_{i,d,s}$ is the *NDVI_M** predicted by the model in year *i*, division *d*, and season s; and $\overline{NDVI\_M^*_{d,s}}$ is the average *NDVI_M** over the years for division *d* and season *s*. For clarity, in this study I=9 years, D=84 divisions, and S=2 seasons. By using the division- and season-specific *NDVI_M** averages in the denominator of Equations 4 instead of global average used in the standard R^2_{cv}, we measure to what extent the selected model performs

better than a naïve model that every year predicts a *NDVI_M** that equals the multi-annual average for each division and season.

4.4 Calculation of premium rates

To evaluate the impact of having longer *NDVI** time series (following intercalibration) on insurance pricing, we calculated premium rates using data for both the 2001-2012 period for $eMODIS_T$ as well as for the augmented period of 1981-2012 using the intercalibrated data. We estimated the premium rate as the expected value of the insurance payout rates using the historical distribution of z-scored *NDVI** (*zNDVI**). For clarity, *zNDVI** indicates how many standard deviations *NDVI** is above or below its division- and seasonal mean value. For illustration purposes we only present results for the SRSD season.

The insurance is structured as a simple index insurance contract that pays when livestock mortality predicted by *zNDVI** exceeds a predefined mortality level (called strike level). Explicitly, the payout rate (or indemnity) in any year, division, and season is calculated as follows:

$$indem_{i,d,s} = \max\left(0, f\left(zNDVI^*_{i,d,s}\right) - K\right) \quad (5)$$

where $f\left(zNDVI^*_{i,d,s}\right)$ is the response function yielding an index between 0 and 100 per cent that represents predicted livestock mortality conditional on $zNDVI^*_{i,d,s}$, and K is the strike level. The strike level is the value above which the contract will begin to indemnify and is selected by the insured at the inception of the contract. In the current IBLI implementation, the insured may select a strike level of either 10 or 15 per cent. The premium rate can then be calculated as the average of the historical predicted indemnities provided by the application of Equation 5 to the time series of $zNDVI^*_{i,d,s}$. Currently in the IBLI project, specific response

functions to predict livestock mortality from $zNDVI^*_{i,d,s}$ are created for each division and season, based on collected livestock mortality data and taking into account the spatial relationships between divisions. The precise procedure for this will be described in a forthcoming paper by Woodard et al. (in preparation). Here, to show the impact of the longer time series availability on the premium rates, we use a generic response function that describes mortality as an exponential decay function of $zNDVI^*_{i,d,s}$, i.e.:

$$M = f\left(zNDVI^*_{i,d,s}\right) = e^{-2.5-0.3zNDVI^*_{i,d,s}+0.3(zNDVI^*_{i,d,s})^2} \quad (6)$$

where M represents predicted livestock mortality. As we used a strike level (K) of 10 per cent, according to Equation (5) and (6), an indemnity would be granted when $zNDVI^*_{i,d,s}$ is smaller than -0.45 (i.e., when NDVI is smaller than the average value minus 0.45 standard deviations).

5 Results

5.1 Effect of division- and season-pooling on intercalibration

Table 2 presents the results for the global *DSP*-model that maps division- and season-level *NDVI_S** from GIMMS (slave) to *NDVI_M** from any of the other NDVI products (master) for all seasons, years, and divisions. When jointly analyzing all 84 divisions and 18 seasons, the high R^2_{cv} values (above 0.91 for all products) indicate that *NDVI** from GIMMS well correlates to that obtained from other products. When we remove the contribution related to the model's ability to explain the variability in the spatial and seasonal dimensions (i.e., $R^2_{cv(wd,ws)}$), the values decrease to the range 0.622-0.672, depending on which dataset GIMMS is mapped to (Table 2). This value expresses the average capacity of the *DSP*-model to properly model the interannual variability of *NDVI_M** for an individual division and

season. $MODIS_{A\text{-}NASA}$ performs best according to R^2_{cv} and $R^2_{cv(wd,ws)}$, although differences with other products are small. $MODIS_{T\text{-}BOKU}$ showed the poorest overall relation with GIMMS in terms of the R^2-measures and the $RMSE_{cv}$. A possible explanation for poorer agreement is that in this product the quality flags delivered with the original MODIS data are not considered, because Vuolo et al. (2012) assume that poor observations have low NDVI-values and are corrected by the filtering technique. During periods with more persistent cloud cover, this assumption may not be correct. For $MODIS_{T+A\text{-}BOKU}$ the non-consideration of quality flags is less of a problem due to the higher number of good NDVI observations available, as both the Terra and Aqua satellite are used. Figure 2 shows the corresponding scatterplots between *NDVI** from GIMMS and the other products. All regression lines are below the 1:1 line, indicating that a negative bias exists, i.e., *NDVI** from GIMMS is on average higher than *NDVI** from the other products. This data bias (Table 2) is originated by both slope and intercept of the linear regression: in all cases slope is significantly different from 1 and intercept from 0 ($p<0.01$). This bias of GIMMS is also found for Europe (Atzberger et al., 2013). The calibration could efficiently remove the existing data bias judging from the small model bias ($bias_{cv}$) reported in Table 2. From the global *DSP*-model we may conclude that *NDVI** from the operational NDVI products behave similar, and all have a strong correlation with GIMMS.

As the insurance contract is applied at the division level, we evaluated if a less parsimonious division-specific model (Equation 2) could provide a more accurate calibration. Figure 3 shows the frequency distribution of the division-level R^2_{cv} difference between applying the global *DSP*-model and the division-specific *SP*-models to individual divisions, in this case for $MODIS_{A\text{-}NASA}$ ($MODIS_{A\text{-}NASA}$ is shown here as an example, as it performed best for the *DSP*-model, Table 2). For 69 of the 84 divisions, the *SP*-model yielded better predictive

performances than the *DSP*-model, i.e. the *SP*-model better captures the interannual variability of *NDVI** for these divisions. Similar results were obtained for the other operational NDVI products (Table 3). It is beyond the scope of this paper to pinpoint the precise causes of such spatial heterogeneity in the relationship between NDVI products (see section 4.2 for possible explanations), but this finding clearly indicates that cross-division differences in the relationship of *NDVI** are important, and consequently the intercalibration can be achieved more accurately with a division-specific model. Despite the large reduction in sample size (n=1,512 for *DSP* versus n=18 for *SP*), our cross-validated results show that, on average, the more specific SP-model is not overparameterized and GIMMS *NDVI** can be more accurately mapped to *NDVI** from the operational products using the *SP*-model, as compared to the *DSP*-model.

The least parsimonious (n=9) division- and season-specific *NP*-model outperforms the *SP*-model in more than 50 per cent of the divisions for most operational NDVI products (Table 3). This implies the presence of a seasonal effect on the relationship of *NDVI** derived from GIMMS and operational products. This difference in performance was clustered in space, with the *SP*-model performing better in areas in the west (Turkana County) and the *NP*-model in eastern counties (Figure 4). The better performance of the *NP*-model in eastern counties coincides with areas that have a high dynamic range of NDVI during the SRSD season, but a lower dynamic range during the LRLD season (Figure 5). This suggests that if seasonal NDVI characteristics strongly diverge between both seasons, season-specific (*NP*) models are more effective in mapping *NDVI** derived from GIMMS to *NDVI** from operational products. This could partly result from the impact of low signal-to-noise ratios on *NDVI** during relatively dry seasons. For products other than $\text{MODIS}_{\text{A-NASA}}$ (as in the example of Figure 4) the general pattern is the same despite some changes in the values (not shown).

Note that on average the *SP*-model performs better for the SRSD as compared to the LRLD season, possibly due to the reduced NDVI dynamic range during LRLD in many divisions.

Figure 4a suggests that for each division a different model may be selected to obtain optimal relationships between *NDVI** derived from different sources. Arguably, this may in fact be an option for creating the long time series of drought indices for each division. Which model should be used for each individual division would depend on the operational NDVI time series selected (i.e. Figure 4a may deviate for other products). Here, in order to select one single modelling solution, we pragmatically evaluated the magnitude of performance improvement achieved increasing the specificity of the modelling solution (and thus, reducing its parsimony). For the example of $\text{MODIS}_{\text{A-NASA}}$, only for three divisions the *DSP*-model's R^2_{cv} was more than 0.10 higher than that of the *SP*-model; for most divisions the *SP*-model performed at least similar to *DSP*, if not much better (Figure 3). On the contrary, when comparing the *SP*- to the more specific *NP*-model, only 23 per cent of the divisions for LRLD, and 6 per cent for SRSD show an improved performance by over 0.10 (ΔR^2_{cv}) for *NP*-models. Therefore, given the relatively close similarity between *NP*- and *SP*-models performances, we chose to confine ourselves in the further analysis to the more parsimonious *SP*-model. The use of a single model-type would also imply a simpler and more consistent solution for insurance design. At the same time, we acknowledge however that no single models 'wins' across all divisions, and that the selection of different models for each division may be a better option based on purely empirical grounds.

5.2 Comparison of NDVI products

Figure 6 shows the division-level R^2_{cv} results for the *SP*-model for all NDVI product pairs analysed. The first row depicts the relationship of *NDVI** from GIMMS (slave) with *NDVI**

from all operational NDVI products. Although differences are small, on average for all divisions SPOT-VGT has the highest (0.85) and $MODIS_{T\text{-}BOKU}$ the lowest R^2_{cv} (0.79). The spatial pattern of the relationship is very similar for all NDVI products in relation to GIMMS. R^2_{cv}-values below 0.50 are found for all products in the driest divisions of Turkana in the north-west. The NDVI dynamic range is extremely low in this area (Figure 5), leading to low signal-to-noise ratios that negatively affect the relationships between products and moreover question the usability of the uncalibrated *NDVI** for insurance purposes in these regions. The NDVI dynamic range is illustrated in Figure 7 for one poorly-performing division in Turkana (Lokichar), and a good-performing division in Moyale (Obbu). Despite the poor performances of some divisions, in many divisions the performance can be considered good. For example between GIMMS and SPOT-VGT, nearly half of all divisions have an R^2_{cv} above 0.90, and R^2_{cv} above 0.80 are found in 82 per cent of the divisions (Table 4). $eMODIS_T$ shows similar figures, while the other MODIS products demonstrate a somewhat poorer relationship with GIMMS.

Compared to the relationship between operational NDVI products and GIMMS, the relationships among operational products showed a higher R^2_{cv} (Figure 6). This may be largely attributable to the closer similarity of the spectral response functions of SPOT-VGT and MODIS. Because all MODIS products are based on the same sensor (although flown on two satellites), it is logical that the relationship between SPOT-VGT and MODIS shows stronger deviations as between individual MODIS products. Nonetheless, between SPOT-VGT and $eMODIS_T$, 73 per cent of all divisions have an R^2_{cv} above 0.95, and 94 per cent above 0.90, indicating overall good comparability of *NDVI**. $MODIS_{T\text{-}BOKU}$ has the poorest relationship with SPOT-VGT (21 and 67 per cent above R^2_{cv}'s of 0.95 and 0.90, respectively), and also with other MODIS products. The different filtering used, and the fact

that the quality flags are not used for this product, may explain this behaviour, given that $MODIS_{T\text{-}NASA}$ (based on the same MOD13Q1 product) performs much better. Still we can conclude from Figure 6 that most operational NDVI products provide comparable *NDVI** values for the majority of divisions. In an operational context of division-level drought monitoring, this finding would allow their interchangeable use, which can be important in case one satellite sensor fails.

While the calibration performance in terms of R^2_{cv} informs us about the overall correlation between GIMMS and the operational products, a key interest for insurance payouts is whether different products are capable of identifying droughts and their relative severity. Despite relative poor calibration, it may still be possible that various NDVI products identify droughts similarly. As an example, Figure 8 compares the time series of z-scored *NDVI** of GIMMS and three operational NDVI products for three divisions in Wajir County, each characterized by a different quality of the calibration against GIMMS (as indicated by the R^2_{cv}-values for the *SP*-model). For the Gurar-Bute division, we can observe that three major droughts are identified by all NDVI products: in order of decreasing severity these are 2011 LRLD, 2005 LRLD, and 2010 SRSD. For Buna and Eldas (R^2_{cv} for GIMMS versus SPOT-VGT of 0.82 and 0.66) five seasons are identified as dry by all products with z-scores below -0.5. Despite that the major drought (2011 LRLD) is equally identified by all products, the severity ranking of these seasons differs slightly. Nonetheless, this comparison shows that even for relatively lower R^2_{cv}-values, which are not very common in our analysis (Table 4), we may attain a reasonably comparable estimate of drought occurrence.

Overall, our findings suggest that good perspectives exist for extending the operational NDVI products back in time to create longer time series of drought indices for livestock insurance in

Kenya. The only exceptions are the few poorly-performing divisions, notably the very arid divisions in Turkana with low signal-to-noise ratios.

5.3 Effect on premium rates

Figure 9 shows the premium rates for the SRSD season as calculated using Equation (5) and (6) for the period 2001-2012 from $eMODIS_T$ and for the period 1981-2012 using the intercalibrated data. The average premium rate across all divisions equals 2.54 per cent for the period 2001-2012, and 2.95 per cent for the longer period. For individual divisions, the premium rate estimate changes significantly when using the longer period of intercalibrated data. For example, in the Kirisia division in the east of Samburu County, the estimated premium rate more than doubled when using the 1981-2012 period (4.6%) as compared to using only 2001-2012 (2.2%). Higher rates imply a higher expected livestock mortality, and consequently a higher cost for the pastoralist to purchase the insurance (assuming equal risk-loading by the insurer), but simultaneously this could benefit the sustainability of an insurance scheme from the insurer's perspective. The standard deviation of the difference across all divisions is 1.14 per cent, which represents a significant amount of rate volatility in insurance terms. Such volatility may be expected in a dryland pastoralist system with strong deviations in asset losses. This example illustrates that the addition of 20 years of data has a strong impact on the premium.

In statistical terms, the premium rates are expected to be more efficient and robust for a larger sample size (that is more seasons), provided that the relationship between mortality and the index is stable over time and that the intercalibration between sensors is effective. The stronger statistical basis may motivate insurers to reduce the risk-loading (Biener, 2013), which could partly off-set the increased premiums that were calculated based on the longer

time period for a large number of divisions. Further study should reveal whether the premium rates based on longer time series are also more efficient in ascertaining a sustainable insurance scheme. In this context, sustainable implies that in the long run, the scheme is attractive for both the insured and the insurer. In section 6 we will further discuss limitations of long time series, and their assumed stationarity, for effectively representing livestock mortality risks.

6 Discussion

Our study confirms that the GIMMS product is the most dissimilar among all NDVI products tested, a finding that can be attributed largely to the fact that the AVHRR sensor was not designed specifically for vegetation studies and consequently has much broader spectral bands for measuring red and NIR reflection (Trishchenko et al., 2002; Miura et al., 2006). We did not find evidence of markedly different performance among the operational NDVI products: pre-processing algorithms (i.e., temporal filtering) are partly responsible for differences that did occur. Our pragmatic approach of first focussing on the relevant aggregate drought indices (*NDVI**), and subsequently comparing these across products, provided comparable measures across different NDVI products for most divisions of arid Kenya.

Despite the high overall R^2_{cv} (>0.95) between GIMMS and operational products for the global *DSP*-model (divisions and seasons pooled), division-specific models better predicted the temporal variability in *NDVI** of different sensors per division. This outcome is not trivial as the more specific models are tuned on a reduced sample size (as compared to the global *DSP*-model) and are thus more exposed to potential over-fitting problems. This finding is supported by other studies that indicate location-specific dependencies on the relationship

between NDVI derived from different sensors (e.g., Miura et al., 2006; Swinnen and Veroustraete, 2008). Except for approximately 10 to 20 per cent of the divisions that have a very limited NDVI dynamic range, seasons with poor vegetation characteristics (due to drought conditions) could be properly identified by all NDVI products using the season-pooled division-specific *SP*-model. Further disaggregation achieved by the *NP*-model, where the calibration is performed by each division and separately for the two seasons of interest, yielded relatively small improvements in prediction for more than 50 per cent of the divisions. Our results demonstrate that longer operational time series of the drought index can effectively be constructed for most divisions by mapping *NDVI** from GIMMS to that of operational products. We fully acknowledge, however, that other approaches to achieve this may be identified, and that further improvements can be envisaged. Here we non-exhaustively discuss a few possible improvements or adaptations to our approach.

First, for several operational NDVI products we have more years of overlap available with the GIMMS dataset than the 18 seasons between October 2002 and September 2011 used here. These 18 seasons overlap between all products evaluated in our study, and were selected to provide a fair comparison among products. However, for example for SPOT-VGT, eight more seasons are available. Incorporating these seasons in the regression may improve the estimation of the regression coefficients, and thus results in improved consistency of the combined long-term time series.

Second, the observed deviation from linearity in the relationship between GIMMS and other NDVI products may be approached using a quadratic regression (see also Miura et al., 2006) to test the trade-off between the benefit of a potentially more appropriate model and the drawbacks of an increased parameterization.

Third, besides the three calibration options that we tested here (Equation 1-3), other intermediate levels of division- and season-pooling of *NDVI** can be envisaged and may have benefits. For example, all divisions within a county could be used in a single regression equation. Within that county, depending if the characteristics of the two seasons are very distinct, also the seasons could be separately analysed. Another intermediate possibility would be to use all divisions simultaneously in a fixed effects panel regression model, in which a single slope is obtained for all divisions and different intercepts for each, thus reducing the amount of parameters to be estimated as compared to *SP*-models while increasing sample size (Baltagi, 2008). A more drastic consideration is whether we should stick to division-boundaries, or use a better ecological stratification of the area, possibly based on NDVI series as well (de Bie et al., 2011). For IBLI this would not be a good option, given that livestock mortality data are available at the division level, and divisions are for insurers and pastoralists the most logical unit for having the same insurance premiums and payout. Alternatively, divisions could be pooled based on ecological characteristics, for example through similarity-based clustering of their average NDVI profiles. The main advantage of this would be the increase of the sample size, possibly leading to more reliable estimates of the regression coefficients, while still keeping homogeneity of insurance contract within each division. While further empirical testing of different pooling levels could lead to slight improvements of calibration performances, it is likely that bigger benefit can be obtained by improving the design of the drought index itself.

Two ways to improve the drought index (currently defined as the z-score of the division spatial average of the mean NDVI over two fixed time periods corresponding to the growing seasons, and jointly covering a whole calendar year without gaps) can be envisaged. A first

way would be through adapting the spatial aggregation step. Currently all pixels within a division are incorporated when calculating *NDVI**. However, many pixels may have low signal-to-noise ratios (because vegetation is absent throughout the year, for example) affecting the calibration reliability. In addition, these pixels may not represent locations where livestock is actually grazing. Such pixels are less relevant concerning their effect on livestock conditions and could be excluded 1) by setting a threshold on the pixel's mean NDVI and/or its temporal variability, or 2) by incorporating land cover maps such as AfriCover (e.g., Genovese et al., 2001; Rojas et al., 2011).

A second way to enhance the performance of the drought index, both in the construction of the long term archive and in the actual application of the index in the insurance scheme, could be through changing the definition of the seasons under consideration. Currently, in analogy to Chantarat et al. (2013) and the current IBLI design, the LRLD and SRSD together cover a full year. This implies, however, that a significant proportion of each season contains a (relatively) dry period. During this period, biomass is not developing and consequently NDVI values provide information of limited relevance regarding grazing opportunities for livestock. A better and more realistic tuning of the considered period is thus also expected to increase the correlation of the index with actual livestock mortality and therefore to further reduce the insurance basis risk. Moreover, given that exposure of bare soil impacts reflection differently depending on the spectral response functions, low biomass conditions tend to decrease the signal-to-noise ratio of the NDVI measurement, and as such decrease the comparability of *NDVI** across sensors. A straightforward solution could be to shorten the seasons by removing the final one to three months of each season that is consistently dominated by dry conditions (see also Figure 7). However, the optimal time period for aggregation could change from division to division. A more appropriate approach could rely

on the automated identification of start- and end-of-season from the NDVI time series at pixel-level (Meroni et al., 2013; Vrieling et al., 2013) or aggregated per division (Rojas et al., 2011; Vrieling et al., 2011).

This study started from the premise that longer time series can better capture the full range of climatic variability and the related drought probability, resulting in improved pricing of insurance contracts. This premise is based on the fact that stationarity of NDVI can be assumed over the considered period. We should place two critical notes however. First, human-induced land use changes within the past 30 years (e.g., Brink and Eva, 2009) may have changed NDVI levels, which do not relate to drought. Second, if trends are present in the NDVI data (possibly due to climatic changes), the longer record may not help to better define drought probability for the upcoming season(s), unless the trends are accounted for. While land use could be relatively stable, and trends may be absent for many divisions, more detailed analysis may be needed for future use of long NDVI records in index-insurance. In this respect, comparison with other indices could also be explored, for example using tree ring data (Bell et al., 2013) that are now available for some parts of Africa (Gebrekirstos et al., 2009).

This discussion contends that ample scope exists for further improving the remote sensing component of the IBLI project. While we achieved to compare several NDVI sources, and provided regression coefficients for creating longer time series of the *NDVI** drought index, the final usefulness of the index (as derived from different sources) can only be ascertained when it can effectively model what is insured, i.e., livestock losses. Although drought is the main cause of livestock mortality in arid Kenya, above-normal wet conditions may also trigger livestock diseases such as East Coast Fever (Homewood et al., 2006), and Rift Valley

Fever (Anyamba et al., 2009). While this is somewhat acknowledged in the current IBLI scheme through the introduction of a quadratic term (Equation 6), better predictions of such disease outbreaks may be possible. In this regard, NDVI series could also provide useful information (e.g., Norval et al., 1991; Anyamba et al., 2009; Trevennec et al., 2012). Upcoming household surveys within the IBLI project should reveal the importance of these diseases for livestock mortality in the region, and efforts are underway within the International Livestock Research Institute (and elsewhere) for NDVI-aided outbreak prediction. Eventually, NDVI-derived outbreak probabilities could be incorporated in the IBLI design to better account for increased mortality during above-normal wet conditions as well.

7 Conclusions

Index-based livestock insurance requires an accurate estimate of livestock mortality from an index to determine the insurance premium, and up-to-date information for determining payouts. Given that most mortality is drought-related, spatially- and temporally-aggregated NDVI is used as drought index input to livestock insurance in Kenya. With the aim of creating a long (>30 year) operational record of division-level seasonal drought indices for the arid lands in Kenya, we compared the non-operational 30-year GIMMS AVHRR record with six operational NDVI products (from MODIS and SPOT-VGT) and three modelling options. Based on cross-validated results, we conclude that division-specific models are the most effective in linking the division-level variability of the drought index (*NDVI**) between the various products. In relation to the long-term GIMMS record, the *SP*-model explained over 80 per cent of the *NDVI** variance for more than 80 per cent of all divisions for SPOT-VGT and eMODIS$_T$. This implies that for most divisions, good scope exists for historically extending the aggregated drought index, thus providing a longer operational record for

insurance purposes. Using a longer record has a significant influence on the insurance premium rates, as shown in this paper. We defined several possible future improvements to the drought index, which may also have a positive impact on the comparability of the resulting drought index time series.

While our work specifically focussed on the demands of the IBLI project in Kenya, the need for long time series of drought indices is not specific to this project. Although not always effective or successful (Binswanger-Mkhize, 2012), index-based insurance of crop or livestock is seen by many as having a great potential for increasing agricultural production among smallholder farmers (Hazell and Hess, 2010; Coe and Stern, 2011), and many initiatives and pilot projects currently exist. Despite limitations (Turvey and McLaurin, 2012), NDVI series are frequently used or considered in these projects, including for example for livestock insurance in Mongolia (Mahul and Skees, 2007) and the ongoing project "Evaluating remote sensing for index insurance" of the Weather Risk Management Facility (http://www.ifad.org/ruralfinance/wrmf/). Given the need for long time series for insurance design and pricing, our current work may further guide other index-insurance projects that seek to combine NDVI series.

References

Anyamba, A., Chretien, J.-P., Small, J., Tucker, C.J., Formenty, P.B., Richardson, J.H., Britch, S.C., Schnabelf, D.C., Erickson, R.L., Linthicum, K.J., 2009. Prediction of a Rift Valley fever outbreak. Proceedings of the National Academy of Sciences of the United States of America 106, 955-959.

Atzberger, C., Eilers, P.H.C., 2011. A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of South America. International Journal of Digital Earth 4, 365-386.

Atzberger, C., Klisch, A., Mattiuzzi, M., Vuolo, F., 2013. Phenological metrics derived over the European continent from NDVI3g data and MODIS time series. Remote Sensing under review.

Baltagi, B.H., 2008. Econometric Analysis of Panel Data, 4th edition. John Wiley & Sons Ltd, Chichester, West Sussex, UK.

Barnett, B.J., Barrett, C.B., Skees, J.R., 2008. Poverty Traps and Index-Based Risk Transfer Products. World Development 36, 1766-1785.

Barrett, C.B., Barnett, B.J., Carter, M.R., Chantarat, S., Hansen, J.W., Mude, A.G., Osgood, D., Skees, J.R., Turvey, C.G., Ward, M.N., 2007. Poverty Traps and Climate Risk: Limitations and Opportunities of Index-Based Risk Financing. IRI Technical Report No. 07-02.

Barrett, C.B., Marenya, P.P., McPeak, J., Minten, B., Murithi, F., Oluoch-Kosura, W., Place, F., Randrianarisoa, J.C., Rasambainarivo, J., Wangila, J., 2006. Welfare dynamics in rural Kenya and Madagascar. Journal of Development Studies 42, 248-277.

Bell, A.R., Osgood, D.E., Cook, B.I., Anchukaitis, K.J., McCarney, G.R., Greene, A.M., Buckley, B.M., Cook, E.R., 2013. Paleoclimate histories improve access and sustainability in index insurance programs. Global Environmental Change 23, 774-781.

Biener, C., 2013. Pricing in microinsurance markets. World Development 41, 132-144.

Binswanger-Mkhize, H.P., 2012. Is there too much hype about index-based agricultural insurance? The Journal of Development Studies 48, 187-200.

Bonifacio, R., Dugdale, G., Milford, J.R., 1993. Sahelian rangeland production in relation to rainfall estimates from Meteosat. International Journal of Remote Sensing 14, 2695-2711.

Brink, A.B., Eva, H.D., 2009. Monitoring 25 years of land cover change dynamics in Africa: A sample based remote sensing approach. Applied Geography 29, 501-512.

Brown, M.E., Lary, D.J., Vrieling, A., Stathakis, D., Mussa, H., 2008. Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS. International Journal of Remote Sensing 29, 7141-7158.

Brown, M.E., Osgood, D.E., Carriquiry, M.A., 2011. Science-based insurance. Nature Geoscience 4, 213-214.

Brown, M.E., Pinzón, J.E., Didan, K., Morisette, J.T., Tucker, C.J., 2006. Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. IEEE Transactions on Geoscience and Remote Sensing 44, 1787-1793.

Chantarat, S., Mude, A.G., Barrett, C.B., Carter, M.R., 2013. Designing index-based livestock insurance for managing asset risk in northern Kenya. Journal of Risk and Insurance 80, 205-237.

Chen, J., Jönsson, P., Tamura, M., Gu, Z.H., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sensing of Environment 91, 332-344.

Coe, R., Stern, R.D., 2011. Assessing and addressing climate-induced risk in sub-Saharan rainfed agriculture: lessons learned. Experimental Agriculture 47, 395-410.

de Bie, C.A.J.M., Khan, M.R., Smakhtin, V.U., Venus, V., Weir, M.J.C., Smaling, E.M.A., 2011. Analysis of multi-temporal SPOT NDVI images for small-scale land-use mapping. International Journal of Remote Sensing 32, 6673-6693.

Funk, C., Budde, M.E., 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. Remote Sensing of Environment 113, 115-125.

Gallo, K., Li, L., Reed, B., Eidenshink, J., Dwyer, J., 2005. Multi-platform comparisons of MODIS and AVHRR normalized difference vegetation index data. Remote Sensing of Environment 99, 221-231.

Gao, B.C., 2000. A practical method for simulating AVHRR-consistent NDVI data series using narrow MODIS channels in the 0.5-1.0 mu m spectral range. IEEE Transactions on Geoscience and Remote Sensing 38, 1969-1975.

Gebrekirstos, A., Worbes, M., Teketay, D., Fetene, M., Mitloehner, R., 2009. Stable carbon isotope ratios in tree rings of co-occurring species from semi-arid tropics in Africa: patterns and climatic signals. Global and Planetary Change 66, 253-260.

Genovese, G., Vignolles, C., Nègre, T., Passera, G., 2001. A methodology for a combined use of normalised difference vegetation index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain. Agronomie 21, 91-111.

Gutman, G., Masek, J.G., 2012. Long-term time series of the Earth's land-surface observations from space. International Journal of Remote Sensing 33, 4700-4719.

Hazell, P.R., Hess, U., 2010. Drought insurance for agricultural development and food security in dryland areas. Food Security 2, 395-405.

Homewood, K., Trench, P., Randall, S., Lynen, G., Bishop, B., 2006. Livestock health and socio-economic impacts of a veterinary intervention in Maasailand: Infection-and-treatment vaccine against East Coast fever. Agricultural Systems 89, 248-271.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83, 195-213.

Jacobs, T., Borstlap, G., Bartholomé, E., Maathuis, B.H.P., 2008. DevCoCast in support of environmental management and sustainable development in Africa. Proceedings of the 7th AARSE Conference, Accra, Ghana, pp. 1-6.

James, M.E., Kalluri, S.N.V., 1994. The Pathfinder AVHRR land data set: an improved coarse resolution data set for terrestrial monitoring. International Journal of Remote Sensing 15, 3347-3363.

Jenkerson, C.B., Maiersperger, T., Schmidt, G., 2010. eMODIS: a user-friendly data source. Open-File Report 2010–1055. USGS, p. 10 pages.

Leblois, A., Quirion, P., 2013. Agricultural insurances based on meteorological indices: realizations, methods and research challenges. Meteorological Applications 20, 1-9.

Little, P.D., Smith, K., Cellarius, B.A., Coppock, D.L., Barrett, C.B., 2001. Avoiding disaster: Diversification and risk management among east African herders. Development and Change 32, 401-433.

Mahul, O., Skees, J., 2007. Managing agricultural risk at the country level: the case of index-based livestock insurance in Mongolia. Policy Research Working Paper 4325. Washington D.C., p. 37.

Meroni, M., Verstraete, M.M., Rembold, F., Urbano, F., Kayitakire, F., 2013. A phenology-based method to derive biomass production anomaly for food security monitoring in the Horn of Africa. International Journal of Remote Sensing accepted.

Miura, T., Huete, A., Yoshioka, H., 2006. An empirical investigation of cross-sensor relationships of NDVI and red/near-infrared reflectance using EO-1 hyperion data. Remote Sensing of Environment 100, 223-236.

Myneni, R.B., Asrar, G., 1994. Atmospheric effects and spectral vegetation indexes. Remote Sensing of Environment 47, 390-402.

Nkedianye, D., de Leeuw, J., Ogutu, J.O., Said, M.Y., Saidimu, T.L., Kifugo, S.C., Kaelo, D.S., Reid, R.S., 2011. Mobility and livestock mortality in communally used pastoral areas: the impact of the 2005-2006 drought on livestock mortality in Maasailand. Pastoralism 1, 17.

Norval, R.A.I., Perry, B.D., Gebreab, F., Lessard, P., 1991. East Coast Fever: a problem of the future for the Horn of Africa. Preventive Veterinary Medicine 10, 163-172.

Onono, J.O., Wieland, B., Rushton, J., 2013. Productivity in different cattle production systems in Kenya. Tropical Animal Health and Production 45, 423-430.

Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., Roy, D., Junchang, J., Vermote, E., Prince, S., Nagol, J., Justice, C., Schaaf, C., Jicheng, L., Privette, J., Pinheiro, A., 2007. Generating a long-term land data record from the AVHRR and MODIS Instruments. Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, pp. 1021-1025.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Koppen-Geiger climate classification. Hydrology and Earth System Sciences 11, 1633-1644.

Pinzón, J.E., Tucker, C.J., 2013. A non-stationary 1981-2012 AVHRR $NDVI_{3g}$ time series. Remote Sensing under review.

Rahman, H., Dedieu, G., 1994. SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. International Journal of Remote Sensing 15, 123-143.

Rojas, O., Vrieling, A., Rembold, F., 2011. Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. Remote Sensing of Environment 115, 343-352.

Sandmeier, S., Muller, C., Hosgood, B., Andreoli, G., 1998. Physical mechanisms in hyperspectral BRDF data of grass and watercress. Remote Sensing of Environment 66, 222-233.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36, 1627-1639.

Song, Y., Ma, M., Veroustraete, F., 2010. Comparison and conversion of AVHRR GIMMS and SPOT VEGETATION NDVI data in China. International Journal of Remote Sensing 31, 2377-2392.

Steven, M.D., Malthus, T.J., Baret, F., Xu, H., Chopping, M.J., 2003. Intercalibration of vegetation indices from different sensor systems. Remote Sensing of Environment 88, 412-422.

Swets, D.L., Reed, B.C., Rowland, J.D., Marko, S.E., 1999. A weighted least-squares approach to temporal NDVI smoothing. Proceedings of the 1999 ASPRS Annual Conference. American Society of Photogrammetric Remote Sensing, Portland, Oregon, pp. 526-536.

Swinnen, E., Veroustraete, F., 2008. Extending the SPOT-VEGETATION NDVI time series (1998-2006) back in time with NOAA-AVHRR data (1985-1998) for southern Africa. IEEE Transactions on Geoscience and Remote Sensing 46, 558-572.

Trevennec, C., Pittiglio, C., Wainwright, S., Plee, L., Pinto, J., Lubroth, J., Martin, V., 2012. Rift Valley fever: vigilance needed inthe coming months. EMPRES Watch 27, 1-8.

Trishchenko, A.P., 2009. Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors: Extension to AVHRR NOAA-17,18 and METOP-A. Remote Sensing of Environment 113, 335-341.

Trishchenko, A.P., Cihlar, J., Li, Z.Q., 2002. Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors. Remote Sensing of Environment 81, 1-18.

Tucker, C.J., Pinzón, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F., El Saleous, N., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. International Journal of Remote Sensing 26, 4485-4498.

Turvey, C.G., McLaurin, M.K., 2012. Applicability of the Normalized Difference Vegetation Index (NDVI) in Index-Based Crop Insurance Design. Weather, Climate, and Society 4, 271-284.

Vrieling, A., de Beurs, K.M., Brown, M.E., 2011. Variability of African farming systems from phenological analysis of NDVI time series. Climatic Change 109, 455-477.

Vrieling, A., de Leeuw, J., Said, M.Y., 2013. Length of growing period over Africa: variability and trends from 30 years of NDVI time series. Remote Sensing 5, 982-1000.

Vuolo, F., Mattiuzzi, M., Klisch, A., Atzberger, C., 2012. Data service platform for MODIS NDVI time series pre-processing at BOKU Vienna: current status and future perspectives. SPIE 8538, Earth Resources and Environmental Remote Sensing/GIS Applications III, 85380A, p. 85380A.

Woodard, J., Shee, A., Mude, A.G., in preparation. A spatial econometric approach to scalable index insurance against drought related livestock mortality in Kenya.

Appendix A: Detailed description of NDVI data sets

A.1 GIMMS AVHRR

To obtain long time series of NDVI data, we used the 8-km resolution NDVI dataset that was constructed by the GIMMS project. This 15-day (two maximum-value composites per month) product covers July 1981 to December 2011. The AVHRR sensors used to construct the dataset were flown on six satellites. The GIMMS AVHRR dataset has been corrected for factors that do not relate to changes in vegetation greenness, and the latest version (NDVI3g) applies an improved cloud masking as compared to older versions of the GIMMS dataset (Tucker et al., 2005; Pinzón and Tucker, 2013).

A.2 SPOT-VGT

We obtained ten-daily SPOT-VGT NDVI composites (S10 product) with a 1-km spatial resolution for 1998-present through the Flemish Institute for Technological Research (VITO). Three composites cover exactly one month, i.e. for day 1-10, 11-20, and 21-last day of each month. Following atmospheric correction (Rahman and Dedieu, 1994), a constrained view-angle maximum value composite rule is applied. For Africa, the data are available in near-real time through the VGT4Africa and GEONETCast projects (Jacobs et al., 2008). We used the quality flags provided with the data to discard observations affected by clouds or shadows, or that otherwise have a bad radiometric quality in the red or NIR band.

A.3 MODIS

We used two series of 16-day NDVI constrained view-angle maximum value composites from the 250-m resolution global MODIS vegetation indices product Collection 5, i.e. for Terra (MOD13Q1) and for Aqua (MYD13Q1). Similarly to the SPOT-VGT product, the

maximum-value compositing employed for these products selects the highest NDVI values, but constrains the candidate pixels by the view angle (Huete et al., 2002). We refer to this Terra-derived product here as $MODIS_{T\text{-}NASA}$ and the Aqua-derived product as $MODIS_{A\text{-}NASA}$. Quality flags provided with the data were used to mask out unreliable observations (i.e., MODIS quality flag greater than 1 and usefulness flag greater than 5).

A.4 MODIS – Whittaker filter

The University of Natural Resources and Applied Life Sciences (BOKU) in Vienna, Austria, provides on-demand temporally-filtered MODIS NDVI composites, based on the Terra- and Aqua-derived products described above (section A.3). Their service includes temporal filtering, mosaicking, sub-setting, and reprojection, and can deliver data requests within a day, including near-real time acquisitions (Vuolo et al., 2012). The temporal filtering is achieved with a modified Whittaker filter (Atzberger and Eilers, 2011). This least squares approach incorporates a 'penalty' criterion regarding the smoothness of the resulting NDVI profile. Currently, MODIS quality indicators are not used to mask NDVI observations prior to filtering, following the assumption that poor observations have low NDVI values and will be corrected by the temporal filter (Vuolo et al., 2012). Exploiting the availability of NDVI products from Terra and Aqua platforms (both originally composited from NASA at 16-day, but with temporal compositing window shifted of 8 day), BOKU provides both a standard 16-day composite based on Terra only (here referred to as $MODIS_{T\text{-}BOKU}$) and Terra plus Aqua combined product produced every 8 days (here $MODIS_{T+A\text{-}BOKU}$). The product based on Aqua only was not considered in this paper, but is also processed by BOKU.

A.5 eMODIS

The eMODIS data (*e* for 'enhanced', 'expedited', and 'expandable') for Africa contain 10-day NDVI composites at 250-m resolution that are constructed using similar algorithms as the Collection-5 MODIS products (Jenkerson et al., 2010). The United States Geological Survey (USGS) has produced these composites since 2010 to better respond to user needs (regarding for example projections and compositing periods). Both historical archive data and near-real time composites are freely available online. The 10-day composites are produced every five days resulting in six composites per month. Here we only took the composites that cover days 1-10, 11-20, and 21-last day of each month, i.e. the same composite periods as for SPOT-VGT. While unfiltered composites can be obtained for the full Africa window, for this study we used the filtered eMODIS product for the East Africa window that is employed operationally for food security monitoring activities of FEWS-NET. The temporal filtering is based on a weighted least-squares regression approach that gives highest weights to local peaks in the NDVI profile, and lowest weights to local valleys (Swets et al., 1999). The filtered data are available for January 2001 until present. For clarity in this paper we add subscript "T" (for Terra) to refer to this dataset, i.e., $eMODIS_T$.

Tables

Table 1. Main characteristics of the NDVI products used in this study.

*Name dataset**	*Sensor*	*Platform*	*Data provider*	*Start*	*Resolution (m)*	*Composite period (days)*	*Temporal filtering***
GIMMS	AVHRR	NOAA (7 satellites)	NASA	1981 (-2011)	8,000	15	-
SPOT-VGT	VEGETATION	SPOT 4 and 5	VITO	1998	1,000	10	-
$MODIS_{T\text{-}NASA}$	MODIS	Terra	NASA	2000	250	16	-
$MODIS_{A\text{-}NASA}$	MODIS	Aqua	NASA	2002	250	16	-
$MODIS_{T\text{-}BOKU}$	MODIS	Terra	BOKU	2000	250	16	Atzberger and Eilers (2011)
$MODIS_{T+A\text{-}BOKU}$	MODIS	Terra+Aqua	BOKU	2002	250	8	Atzberger and Eilers (2011)
$eMODIS_T$	MODIS	Terra	USGS	2001	250	10	Swets et al. (1999)

** The names refer to the abbreviations for the datasets used in this article*
*** We here indicate here if the original data sources were filtered. Unfiltered datasets were smoothed by us using an iterative Savitzky-Golay filter (see section 4.1).*

Table 2. Statistics from the global *DSP*-model, where division- and season-level *NDVI** from GIMMS is mapped to *NDVI** of each of the six datasets listed in the table. $R^2_{cv(wd,ws)}$ is an aggregate measure of the temporal prediction capability for each division and season (section 4.3).

Name dataset	R^2_{cv}	$R^2_{cv(wd,ws)}$	$RMSE_{cv}$	*data bias*	$model\ bias_{cv}$ *(x1,000)*
SPOT-VGT	0.921	0.647	**0.029**	-0.065	0.136
$MODIS_{T\text{-}NASA}$	0.923	0.662	0.031	-0.039	0.122
$MODIS_{A\text{-}NASA}$	**0.929**	**0.672**	0.030	-0.044	0.142
$MODIS_{T\text{-}BOKU}$	0.917	0.622	0.033	**-0.037**	**0.110**
$MODIS_{T+A\text{-}BOKU}$	0.924	0.652	0.031	-0.046	0.127
$eMODIS_T$	0.922	0.658	0.030	-0.058	0.145

Table 3. Percentage of the divisions where the division- and season-specific models (*SP* and *NP*) outperform the pooled models (*DSP* and *SP*), respectively.

Name dataset	$R_{cv}^2(SP) > R_{cv}^2(DSP)$	$R_{cv}^2(NP) > R_{cv}^2(SP)$	
		LRLD	***SRSD***
SPOT-VGT	88	44	51
$MODIS_{T\text{-}NASA}$	82	61	57
$MODIS_{A\text{-}NASA}$	82	68	54
$MODIS_{T\text{-}BOKU}$	76	63	62
$MODIS_{T+A\text{-}BOKU}$	79	61	56
$eMODIS_T$	89	58	50

Table 4. Percentage of the divisions where R_{cv}^2 from the *SP*-model is higher than the indicated thresholds.

Name dataset	R_{cv}^2**(SP)**				
	>0.50	***>0.70***	***>0.80***	***>0.90***	***>0.95***
SPOT-VGT	96	89	82	46	11
$MODIS_{T\text{-}NASA}$	95	90	77	38	6
$MODIS_{A\text{-}NASA}$	94	88	77	38	5
$MODIS_{T\text{-}BOKU}$	94	82	70	8	1
$MODIS_{T+A\text{-}BOKU}$	95	86	77	30	4
$eMODIS_T$	95	89	81	46	11

Figures

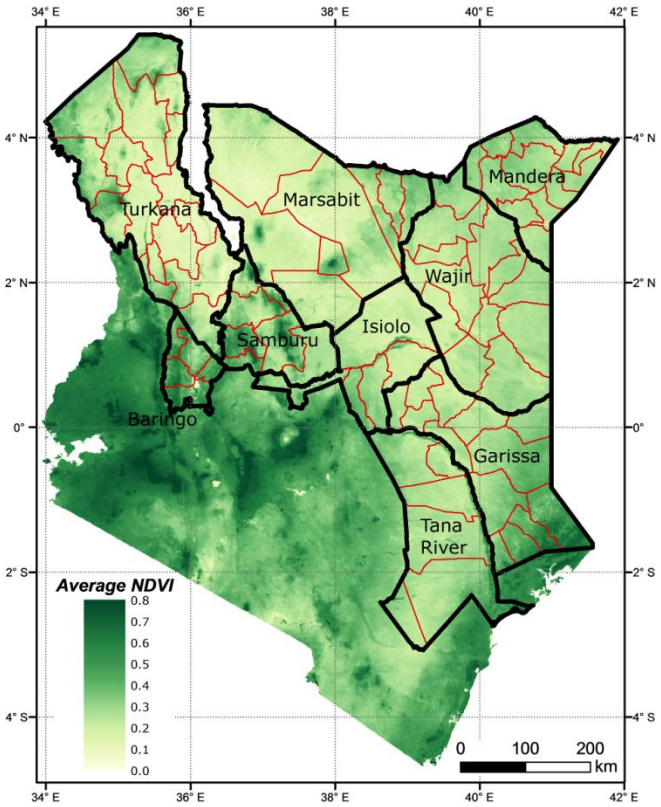

Figure 1: Overview of the study area where the image shows the average NDVI from SPOT-VGT for March 1999 until February 2013 (i.e. for 14 LRLD and 14 SRSD seasons) for Kenya. The black polygons are the 9 Kenyan counties that were considered in this study, with the 84 aggregated divisions (red lines). The aggregation departed from the original 108 divisions in these counties, where small divisions (<1,000 km^2) within the same county were joined.

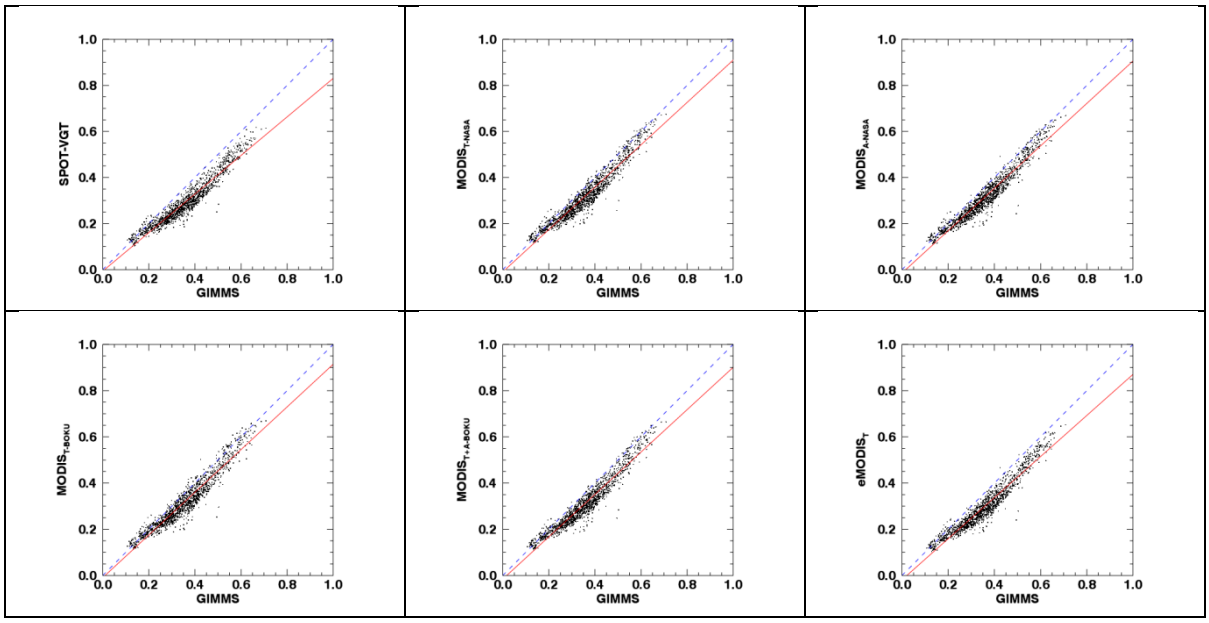

Figure 2: Scatterplots showing *NDVI** (seasonally-averaged NDVI and aggregated per division) derived from GIMMS (x-axis) against *NDVI** from each of the other NDVI products. Each plot contains a total of 1,512 data points (84 divisions x 9 years x 2 seasons per year).

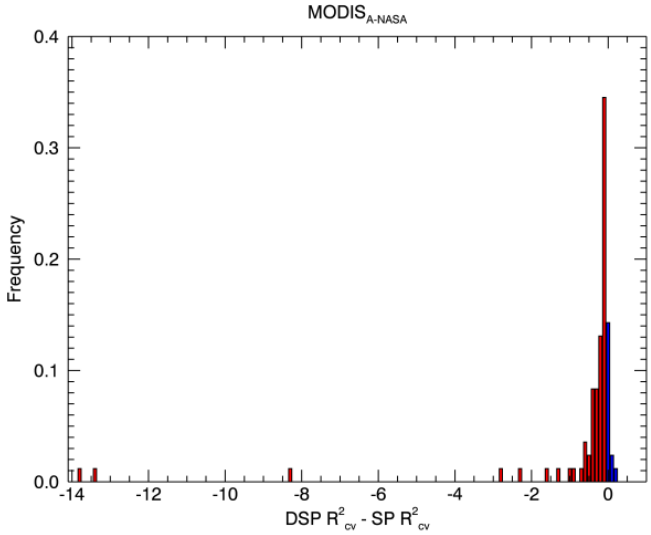

Figure 3: Frequency distribution of the division-level R^2_{cv} difference between the *DSP*- and *SP*-models. Negative values (red bars) account for 82 per cent of the divisions and indicate that the *SP*-model outperforms the *DSP*-model. The example shown is for $\text{MODIS}_{\text{A-NASA}}$.

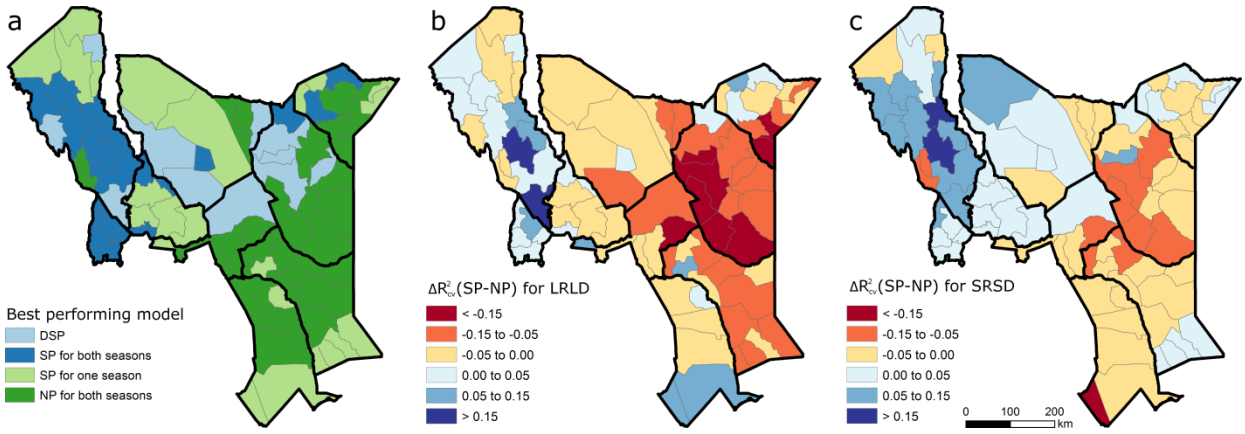

Figure 4: Comparison of models with different levels of pooling at division level: (a) displays the model that has the highest R^2_{cv} for each division. Note that for this the R^2_{cv} of the *DSP*-model was compared with the *SP*-model's R^2_{cv}, and with the average R^2_{cv} for both seasons (LRLD and SRSD) of the *NP*-model. The maps (b) and (c) show the difference in R^2_{cv} (SP-NP), separately for LRLD and SRSD. Example for $\text{MODIS}_{\text{A-NASA}}$.

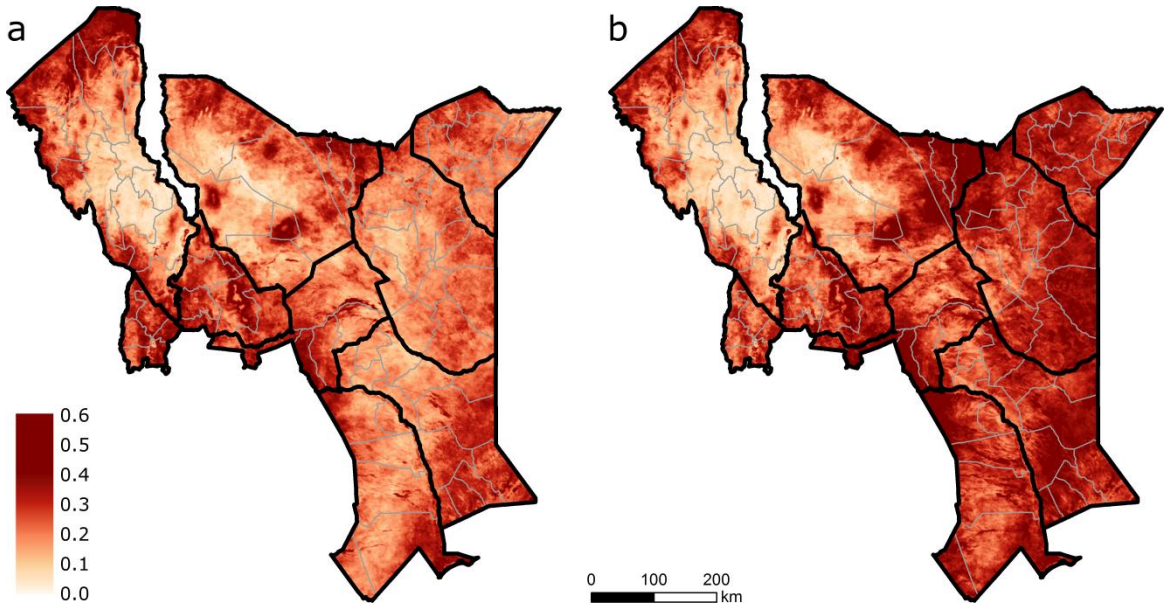

Figure 5: The difference between the 95th and 5th percentile of 10-daily filtered NDVI from SPOT –VGT for October 2002 – September 2011. The difference was calculated separately for the LRLD season (a: March-September) and the SRSD season (b: October-February). The black lines are country boundaries and grey lines are the aggregated division boundaries.

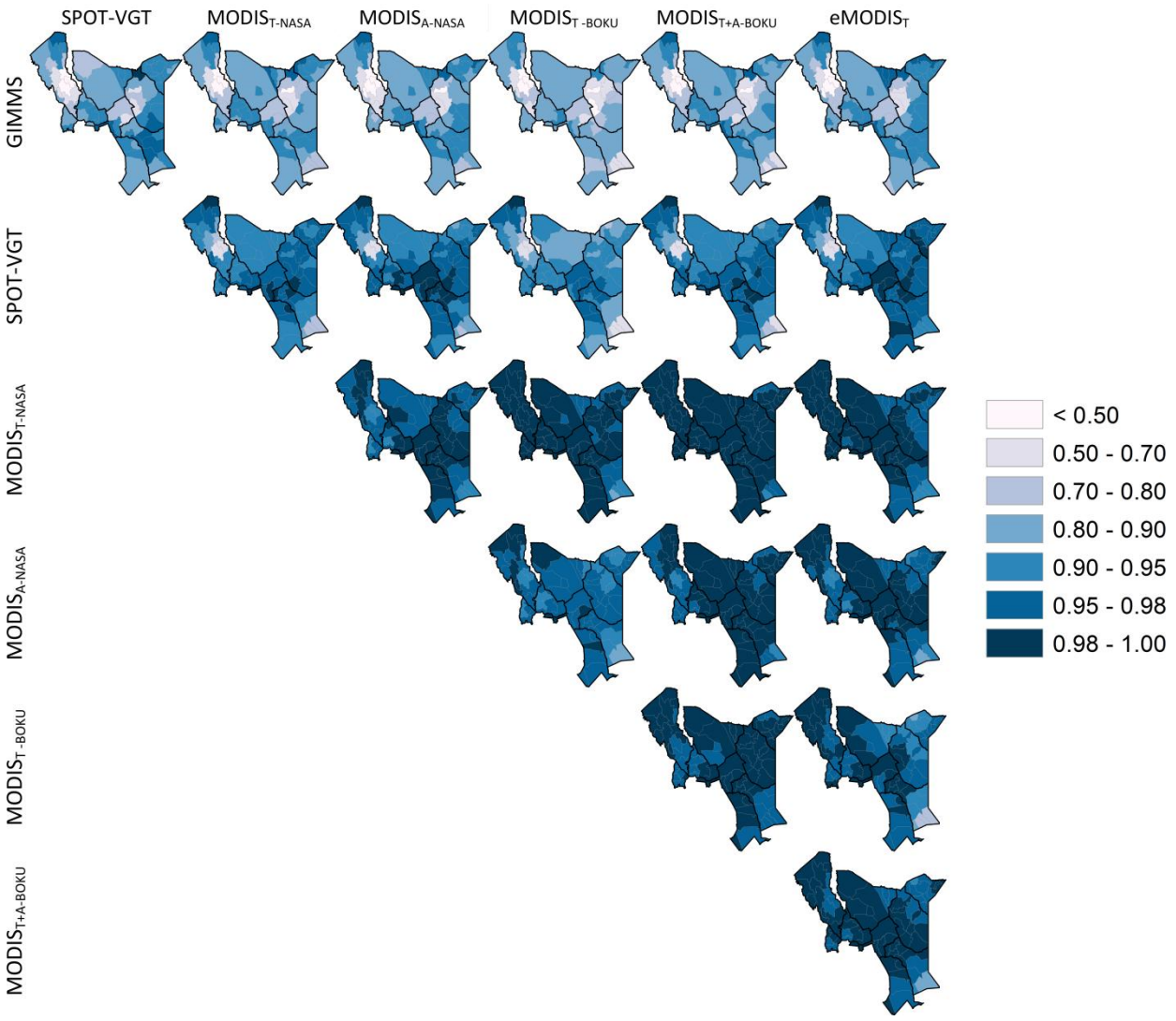

Figure 6: Division-level R^2_{cv} results between all NDVI product pairs analysed, based on the *SP*-model (Equation 2) that links division- and season-averaged NDVI (*NDVI**) for the slave dataset (rows) to that of the master dataset (columns).

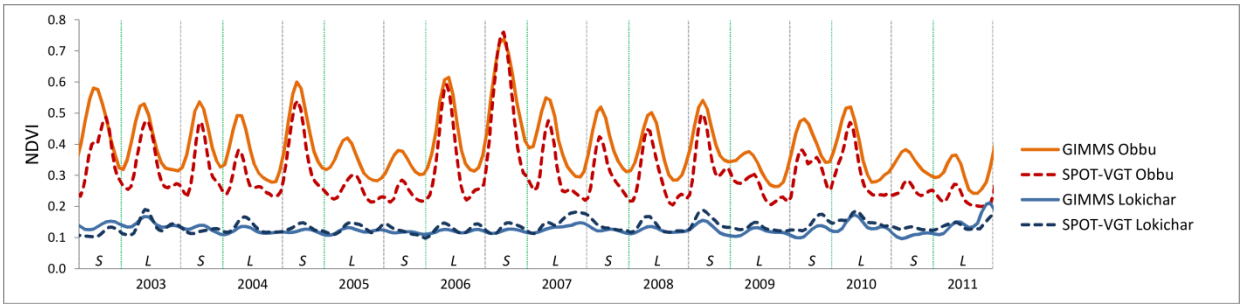

Figure 7: Division-average NDVI profiles for Obbu division (Marsabit County) and Lokichar division (Turkana County) for GIMMS and SPOT-VGT for October 2002 – September 2011. Vertical lines indicate the start of each season (L=LRLD, S=SRSD).

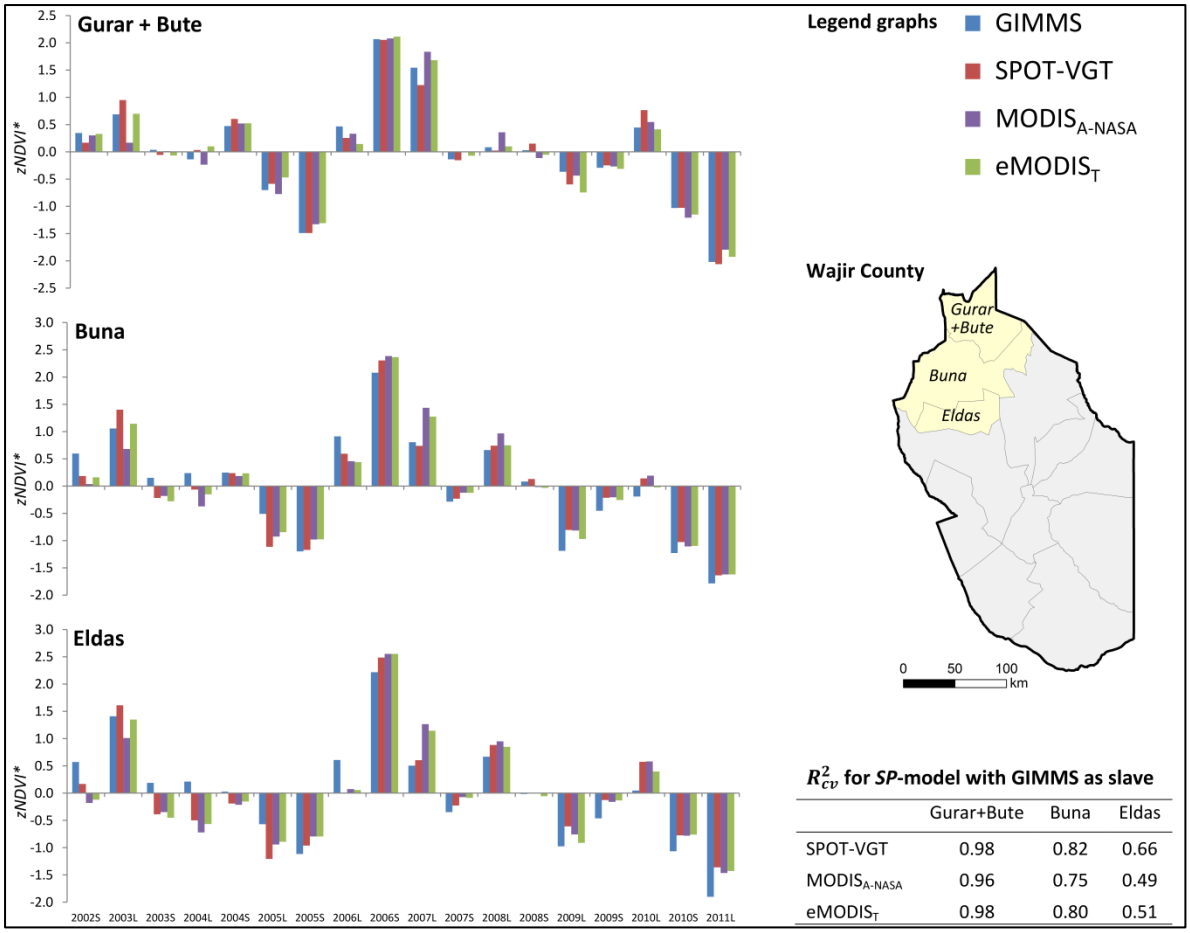

Figure 8: Comparison of z-scored *NDVI** from four NDVI products for three divisions in northern Wajir County. See Figure 1 for the location of Wajir within Kenya. The divisions represent one with a high R^2_{cv}-value (the aggregated Gurar+Bute division), a medium value (Buna), and a low value (Eldas). The base for z-score calculation was the overlapping period (2002 SRSD – 2011 LRLD), and calculations were performed separately for LRLD and SRSD.

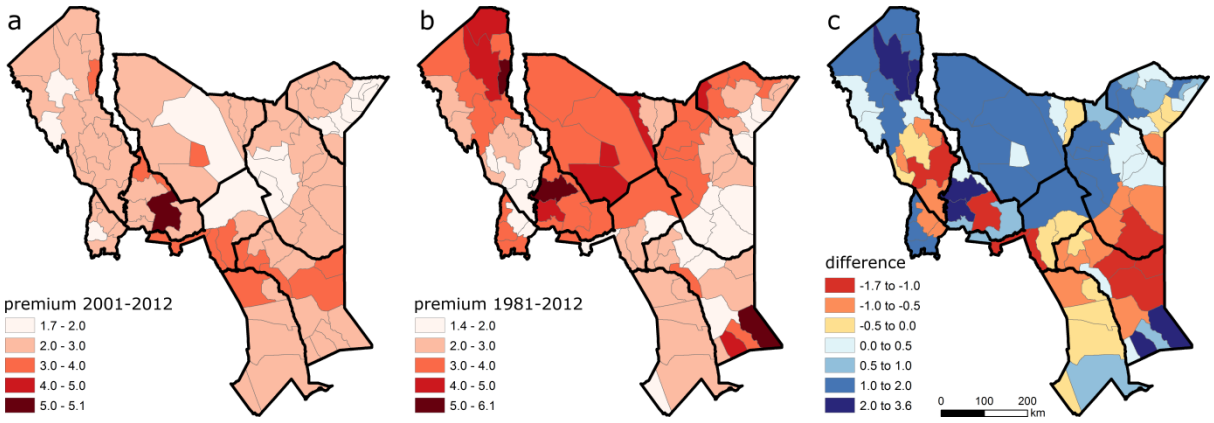

Figure 9: The insurance premium rate (%) for the SRSD season as calculated from eMODIS$_T$ for the period 2001-2012 (a), and using the 1981-2012 time series as obtained from the eMODIS$_T$ intercalibration with GIMMS (b). Map (c) shows the difference in premium between both, i.e., 1981-2012 minus 2001-2012.

Address for correspondence:
Anton Vrieling
University of Twente – Faculty ITC
P. O. Box 217
7500 AE Enschede
The Netherlands
T: +31 53 4874452
E: a.vrieling@utwente.nl

Historical extension of operational NDVI products for livestock insurance in Kenya

Anton Vrieling[a], Michele Meroni[b], Apurba Shee[c], Andrew G. Mude[c], Joshua Woodard[d], Kees de Bie[a], Felix Rembold[b]

[a] University of Twente, Faculty of Geo-information Science and Earth Observation, P.O. Box 217, 7500 AE Enschede, The Netherlands. E-mail: a.vrieling@utwente.nl; c.a.j.m.debie@utwente.nl

[b] Institute for Environment and Sustainability, Joint Research Centre, European Commission, Via E. Fermi 2749, I-21027 Ispra (VA), Italy. E-mail: michele.meroni@jrc.ec.europa.eu; felix.rembold@jrc.ec.europa.eu

[c] International Livestock Research Institute, P.O. Box 30709, Nairobi 00100, Kenya. E-mail: a.shee@cgiar.org; a.mude@cgiar.org

[d] Cornell University, Dyson School of Applied Economics and Management, 236 Warren Hall, Ithaca, NY 14853, United States of America. E-mail: joshua.woodard@cornell.edu

October 2013, submitted to JAG

December 2013, revised version submitted to JAG

Abstract

Droughts induce livestock losses that severely affect Kenyan pastoralists. Recent index insurance schemes have the potential of being a viable tool for insuring pastoralists against drought-related risk. Such schemes require as input a forage scarcity (or drought) index that can be reliably updated in near real-time, and that strongly relates to livestock mortality. Generally, a long record (>25 years) of the index is needed to correctly estimate mortality risk and calculate the related insurance premium. Data from current operational satellites used for large-scale vegetation monitoring span over a maximum of 15 years, a time period that is considered insufficient for accurate premium computation. This study examines how operational NDVI datasets compare to, and could be combined with the non-operational recently constructed 30-year GIMMS AVHRR record (1981-2011) to provide a near-real time drought index with a long term archive for the arid lands of Kenya. We compared six freely available, near-real time NDVI products; five from MODIS, and one from SPOT-VEGETATION. Prior to comparison, all datasets were averaged in time for the two vegetative seasons in Kenya, and aggregated spatially at the administrative division level at which the insurance is offered. The feasibility of extending the resulting aggregated drought indices back in time was assessed using jackknifed R^2 statistics (leave-one-year-out) for the overlapping period 2002-2011. We found that division-specific models were more effective than a global model for linking the division-level temporal variability of the index between NDVI products. Based on our results, good scope exists for historically extending the aggregated drought index, thus providing a longer operational record for insurance purposes. We showed that this extension may have large effects on the calculated insurance premium. Finally, we discuss several possible improvements to the drought index.

Keywords: NDVI, AVHRR, SPOT, MODIS, index insurance, intercalibration

1 Introduction

Coping with drought is a major challenge for pastoralists in the arid and semi-arid parts of Kenya (Little et al., 2001; Nkedianye et al., 2011). During dry years many animals die because of insufficient feed and water, and from drought-related epidemic diseases (Onono et al., 2013). Such losses can have severe, long-term consequences on pastoralist households if their herd sizes fall below specific thresholds (Barrett et al., 2006).

Insurance against the risk of livestock mortality may reduce the negative consequences of drought-induced livestock loss, and avoid families falling into poverty (Chantarat et al., 2013). As opposed to traditional agricultural insurance, requiring expensive verification of individual losses by the insurer, a more cost-effective insurance approach is to base payouts on a transparent and objectively measured variable, such as total seasonal rainfall (Barnett et al., 2008). This is referred to as index-based insurance. Recently, index-based insurance received much attention as it could make important contributions to agricultural growth and reduction of poverty (Hazell and Hess, 2010; Brown et al., 2011). Despite concerns regarding the demand for insurance by poor farmers (Binswanger-Mkhize, 2012), and challenges of reaching sufficient scale among numerous pilot projects, the risk-management potential that index insurance could offer poor farmers fuels continued interest and efforts to improve product design (Barrett et al., 2007; Barnett et al., 2008).

A main limitation to index-based insurance is the possibility for households to experience a loss, but no payment, or alternatively not experience a loss, but yet receive a payment (Barnett et al., 2008). This is referred to as ‘basis risk’ and is caused by the imperfect relationship between the index and incurred losses. For index-based insurance schemes to be effective, they require an index that:

1) strongly correlates with what is insured (such as livestock or crop losses);
2) is independently verifiable, i.e. based on well-described data sources and processing methods;
3) can reliably be delivered into the future (at least for the duration of the insurance contract) and is available in near real-time, so that shortly after losses are incurred, payments can be made;
4) is available for sufficiently long records to properly represent the climatic variability for estimating the probability of a payout~~, and thus accurately pricing of the insurance product~~ (Bell et al., 2013)~~.~~, and thus accurately pricing of the insurance product.

Time series of the normalized difference vegetation index (NDVI) have been used for the purpose of index-based insurance (Turvey and McLaurin, 2012; Leblois and Quirion, 2013). A number of near real-time composite NDVI products are freely available from sensors such as MODIS (Moderate Resolution Imaging Spectroradiometer) and SPOT-VGT (Système Pour l'Observation de la Terre - VEGETATION). These sensors offer a relatively coarse spatial resolution (250-1000m), but provide observations of the same area on a daily basis. This last aspect is important to reduce cloud and atmospheric effects in the composite products, and to effectively compare vegetation conditions within and between years. Given that droughts generate spatially-correlated covariate risks that simultaneously affect a larger number of neighbouring households, pixel-level NDVI values are generally spatially aggregated. In most cases this aggregation is also a necessity for modelling crop and livestock losses, because data on production or mortality are often only available for administrative regions. As a consequence, each administrative unit has different premium specifications and payouts are equal for all insurance customers within a given unit.

In the absence of reliable station rainfall data, the index-based livestock insurance (IBLI) project in Kenya uses NDVI as a proxy for forage scarcity – a key determinant of livestock mortality in pastoral production systems (Chantarat et al., 2013). The insurance design for the Marsabit district of northern Kenya was extensively described by Chantarat et al. (2013). While they used rectangular clusters, the IBLI project currently uses administrative divisions for spatial aggregation. Since 2010, the IBLI project has operated in the Marsabit district, and between 2013 and 2014 the project plans to expand to cover about 60 per cent of Kenya's land surface that constitutes its so-called arid lands. In the four years (eight seasons) during which pastoralists in Marsabit have purchased insurance they received three times insurance payouts following drought. For operational purposes MODIS was selected as the main data source, following the suspension in the delivery of AVHRR (Advanced Very High Resolution Radiometer) NDVI composites by the Famine Early Warning Systems Network (FEWS-NET) due to the degradation of the NOAA-17 AVHRR sensor. A main drawback of MODIS is that it covers only the years 2000 to present, hence insufficient to capture the full range of climatic variability and the related drought probability, needed to properly price insurance contracts. Uncertainties regarding this probability due to data restrictions would lead insurers to add risk-loading to the premium prices, thus making the insurance more expensive and consequently less attractive to pastoralists (Biener, 2013).

The creation of a long-term consistent NDVI time series from multiple sources is not a trivial task due to differences in sensor characteristics and algorithms used to generate products (Miura et al., 2006). Differences in spectral response functions between sensors are a key characteristic responsible for the variation in NDVI (Trishchenko et al., 2002; Trishchenko, 2009). Based on spectral convolution of hyperspectral Hyperion data, Miura et al. (2006) reported that the NDVI relationship among MODIS, AVHRR and ETM+ instruments is non-

linear and largely dependent on how much the green peak (550 nm) and red edge (680-780 nm) regions are included in the red band. Despite that they find near-linear NDVI relationships by direct comparison of AVHRR and MODIS (in correspondence to the empirical study by Gallo et al., 2005), they indicate that higher-order polynomials may be more accurate in modelling cross-sensor NDVI relationships. Additional factors that cause cross-sensor variability of NDVI include atmospheric and bi-directional reflection effects, which are also wavelength dependent (Myneni and Asrar, 1994; Sandmeier et al., 1998). This combination of factors complicates a straightforward joining of NDVI series derived from multiple sensors.

Many attempts have been made to construct a single long-term NDVI record from AVHRR sensors onboard multiple satellites, which effectively corrects for effects like sensor degradation, orbital drift, and atmospheric variability (James and Kalluri, 1994; Tucker et al., 2005). Recently, the Global Inventory Monitoring and Modeling System (GIMMS) project released a 30-year record of the so-called NDVI3g, i.e., third generation GIMMS NDVI from AVHRR sensors. While effectively combining data from various AVHRR sensors already presents a big challenge, spectral response functions are even more dissimilar in comparison to SPOT-VGT and MODIS that have narrower spectral bands (Gao, 2000). Proposed corrections include empirically-derived linear functions (Steven et al., 2003; Gallo et al., 2005; Song et al., 2010) and second-order polynomial regression equations (Trishchenko et al., 2002). Swinnen and Veroustraete (2008) found a strong linear relationship between SPOT-VGT and 1-km^2 AVHRR NDVI for Southern Africa after rigorous reprocessing of spectral reflectance data using the same atmospheric correction and compositing approaches. They effectively accounted for differences in the dynamic range between SPOT-VGT and AVHRR using the adjustment functions of (Trishchenko et al. (2002)). Alternatively, neural

networks, which incorporated data layers reflecting atmospheric conditions, have been used to account for the differences between AVHRR and MODIS (Brown et al., 2008). However, despite various suggestions regarding the achievability of an intercalibrated, sensor-independent NDVI record (e.g. Steven et al., 2003; Brown et al., 2006), and recent efforts towards delivering this to the public (Pedelty et al., 2007; Gutman and Masek, 2012), no universally-accepted multi-sensor NDVI record exists to date that both covers a long (> 25 year) time frame and is available in near real-time.

The aim of this study is to provide a pragmatic solution for combining NDVI composite products derived from multiple sensors (i.e. AVHRR, MODIS, and SPOT-VGT) for the purpose of the livestock insurance programme in Kenya. Rather than analysing cross-sensor NDVI differences per pixel and composite period, we first aggregate the NDVI in space and time to provide an appropriate index in the framework of the IBLI project. This implies aggregation over administrative divisions and for two periods within each year, corresponding to the two growing seasons occurring in the region. We first evaluate if a global regression model (taking all divisions and periods together) can accurately map the aggregated index from one NDVI product to another. As our overall purpose is to have a long record that accurately displays drought-related risk for each administrative unit, which can be updated in near real-time and serve as an input to model livestock mortality, we subsequently perform a cross-sensor comparison at the division level, considering the two seasons together and separately, to examine if this increasing level of disaggregation improves the intercalibration performances with respect to the global model. Besides comparing merely with the non-operational historic AVHRR record, we also compare operational products to evaluate to what extent these datasets can be used interchangeably. This last issue may be important in case of satellite sensor failure in the future. Finally we evaluate if and how the

availability of a longer intercalibrated time series will affect the premium rate for the livestock insurance product.

2 Study area

The study area comprises the nine counties of Kenya that are planned to be covered by the IBLI project over the next one to two years, and are referred to by the Government of Kenya as the arid lands (Figure 1). Since 1996, the government has collected household-level livestock mortality data in representative locations across the study area, in the framework of the Arid Land Resource Management Project (ALRMP, http://www.aridland.go.ke). The nine counties together cover approximately 62 per cent of Kenya's land area. According to Peel et al. (2007), the area contains three Köppen-Geiger climate zones in approximately equal amounts, i.e., tropical savannah climate (Aw), hot steppe climate (BWh), and hot desert climate (BSh). Based on 1998-2012 data of the Tropical Rainfall Measurement Mission (3B43 product), average annual rainfall ranges from less than 300 mm in the dry parts of Isiolo, Marsabit, Turkana, and Wajir Counties, to more than ~~1000~~1,000 mm only in the south-western part of Baringo County. Two rainfall seasons can be discerned: the so-called long rains (March-May) and the short rains (October-December) separated by clear dry seasons. Following Chantarat et al. (2013), we term this bi-modal seasonal pattern as Long Rains Long Dry (LRLD) covering March to September and Short Rains Short Dry (SRSD) covering October to February. Livestock keeping is the main rural livelihood in the region. Livestock includes camels (in the driest parts), goats, sheep, and cattle. To standardize across the livestock types, and to facilitate the development of a single livestock-based insurance product, livestock numbers owned by households are expressed in Tropical Livestock Units (TLU); 1 cattle equals 1 TLU, 1 camel is 1.4 TLU, and a goat or sheep equals 0.1 TLU.

Our analysis focussed on the division-level, as this is the basic unit for which insurance premium and payout are determined. The nine counties comprise 108 divisions. Given the small size of some divisions and the consequent difficulty of obtaining a representative division-level drought index, especially from the 8-km resolution AVHRR series, we set a minimum threshold for division size. Starting from the smallest division, we iteratively aggregated divisions smaller than 1,000 km^2 to the neighbouring division within the same county that had the nearest centroid coordinates. This resulted in 84 spatial units that we further refer to in this paper simply as divisions. The red lines in Figure 1 show the resulting division boundaries.

3 NDVI data sets

To select potential sources of operational NDVI time series data we considered the two following criteria: i) archive and near-real time data should be freely available, and ii) no or minimum post processing should be required to facilitate their use by less-specialized users. As a result, a non-exhaustive list of six operational products was compiled: five derived from MODIS instruments onboard Terra and Aqua platforms, and one from SPOT-VGT. In addition, the new long-term non-operational dataset derived from AVHRR (NDVI3g) was used to create a longer historic record. The main characteristics of the products are summarized in Table 1, and Appendix 1 provides a detailed description of each.

4 Methods

4.1 NDVI processing

For the unfiltered datasets, i.e., GIMMS, SPOT-VGT, $MODIS_{T\text{-}NASA}$, and $MODIS_{A\text{-}NASA}$ (Table 1), we applied an iterative Savitzky-Golay filter (Savitzky and Golay, 1964) as described by Chen et al. (2004) to reduce remaining atmospheric effects in the time series. To do that, we first created a mask to discard any NDVI values that were cloudy or otherwise of poor quality. For this we used the quality information delivered with the SPOT-VGT, $MODIS_{T\text{-}NASA}$, and $MODIS_{A\text{-}NASA}$ data (Appendix A), while for GIMMS we masked out any NDVI values below 0 and with an increase of more than 0.30 in 15 days. The filter was subsequently applied using a third-order polynomial and a moving window of three observations prior to, and after the data point to be filtered. Visual analysis of the resulting time series showed that this procedure substantially reduced noise in the series, effectively interpolated missing values, while retaining short-term variations that relate to real changes in greenness.

Besides introducing the temporal filtering, we further adapted the NDVI processing sequence from the original IBLI design (Chantarat et al., 2013) to provide improved metrics of the season performance, which should also allow for better comparison between different sensors. In the original design, Chantarat et al. (2013) first transformed the 10-daily NDVI images to standard scores (or z-scores). The z-scored NDVI indicates how many standard deviations the pixel's NDVI is above or below the multi-annual mean pixel value of the same 10-day period (e.g., 1-10 January). They then spatially aggregated the z-scored NDVI, and subsequently cumulated the aggregated values over time for two periods, i.e., long rains-long dry (LRLD, March-September) and short rains-short dry (SRSD, October-February). The idea behind aggregating z-scores of 10-day periods is that adverse forage conditions may

occur at any time during the season; however, forage is not produced during the entire season (as defined by LRLD and SRSD). A drawback of directly calculating z-scores for each time step is that small deviations during relatively dry moments of the season can translate to large z-scores, which get equally weighted with smaller z-scores during wet moments (that can however represent stronger absolute deviations) when cumulating over time. To prevent this problem and get a better measure of seasonal forage production, we first performed temporal aggregation, then spatial aggregation, and finally z-scoring.

We performed temporal aggregation for each pixel for both LRLD and SRSD. The average seasonal NDVI was used for this aggregation, which is in the temporal context functionally similar to the cumulative NDVI value, a suitable proxy of seasonal biomass production (e.g., Bonifacio et al., 1993; Funk and Budde, 2009). The advantage of using the average compared to the cumulative value is that 1) values for the two seasons of different length are in the same units and range, and 2) it is insensitive to the different length of the compositing period of the different NDVI products. We then spatial aggregated the temporally-averaged NDVI by calculating the average value for each division (see section 2 on the divisions used). Given the coarse resolution of GIMMS, for this product we calculated a weighted average that reflects the amount of overlap a pixel has with a division.

While the z-scored values are the input for calculating insurance premiums (section 4.4), the basis in this paper for the intercalibration between NDVI products are the NDVI values, aggregated in space and time. We further refer to them as *NDVI**.

4.2 Intercalibration

To evaluate if different NDVI products perform similarly in identifying division-level drought conditions, all data comparisons are based on *NDVI**. For all dataset combinations, we compared results using only the overlapping period between the seven datasets, i.e. the period between July 2002 and December 2011. This period contains a total of 18 seasons, i.e. 9 LRLD seasons and 9 SRSD seasons.

Given the relatively small sample size available for intercalibration we limited our analysis to the linear component of the relationship between NDVI products. Visual inspection of scatterplots ~~provided no indication of non~~ (Figure 2) and the residuals following linear regression (data not shown) suggest a slight deviation from linearity between GIMMS-derived *NDVI** and *NDVI** derived from different sensors. However, given the relatively small sample size available for intercalibration we limited our analysis to the linear relationship between NDVI products ~~(Figure 2).~~. We tested three calibration models that use different levels of pooling of the division-level and season-level *NDVI** data (Equations 1-3, discussed below). The aim of this was to evaluate 1) which NDVI-products show highest correlation with the long-term GIMMS dataset, 2) which level of pooling across season and space is most efficient in transforming *NDVI** between one source and another, 3) for which divisions/regions in Kenya the various products lead to the same seasonal vegetation condition (and hence drought) assessment.

4.2.1 Global model (DSP)

We first assessed the performances of a global calibration model (combining all divisions and seasons) in translating GIMMS-derived *NDVI** to the *NDVI** obtained from other NDVI products, and evaluated which operational product yielded the closest agreement. We refer to

this model as the *DSP*-model, meaning "Divisions and Seasons Pooled". It takes the following form:

$$NDVI\_M^*_{d,s} = \beta_0 + \beta_1 * NDVI\_S^*_{d,s} + \varepsilon_{d,s} \quad (1)$$

where $NDVI\_M^*_{d,s}$ is the average *NDVI* for division *d* and season *s* for the NDVI-series that is used as the master (or dependent variable), while $NDVI\_S^*_{d,s}$ is the average *NDVI* for a division that will be mapped to the master (i.e., the slave, or independent variable). For example, to create longer time series for GIMMS that are compatible with MODIS, GIMMS is considered the slave and MODIS the master. The parameters β_0 and β_1 are the regression coefficients to be estimated and ε is error term. The global model is parsimonious in terms of number of parameters to be estimated (i.e., two with a sample size of 18 seasons x 84 divisions).

4.2.2 Division-specific season-pooled model (SP)

Despite its parsimonious nature, the *DSP*-model may not be able to model division-level specificities in the relationship between products. Cross-division differences may arise because of different NDVI dynamic ranges interacting in a complex way with sensor-specific NDVI saturation levels, different soil background affecting NDVI as a result of sensor-specific spectral response functions, and finally, interaction of local climatology with differences in NDVI processing chains (such as cloud screening and atmospheric correction) affecting locally the relationship between products. These issues justify the evaluation of a less parsimonious division-specific regression model (two parameters to be estimated with a sample size of 18 seasons, for a total of 2x84 parameters), referred to here as *SP*-model ("Season Pooled"). The *SP*-models can be written as:

$$NDVI\_M^*_{d,s} = \beta_{0,d} + \beta_{1,d} * NDVI\_S^*_{d,s} + \varepsilon_{d,s} \quad (2)$$

where the parameters $\beta_{0,d}$ and $\beta_{1,d}$ are now the division-specific regression coefficients to be estimated.

4.2.3 Division-specific season-specific model (NP)

Finally, in order to evaluate if any season-specific effect on the relationship is present, we also evaluated at the division level if separating the LRLD and SRSD seasons improves our regression estimates. This is referred to here as the *NP*-models (for "No Pooling") and, in terms of number of parameters, it is the least parsimonious model considered (two parameters to be estimated with a sample size of nine seasons, for a total of 2x2x84 parameters). The *NP*-model can be expressed as:

$$NDVI\_M^*_{d,s} = \beta_{0,d,s} + \beta_{1,d,s} * NDVI\_S^*_{d,s} + \varepsilon_{d,s} \quad (3)$$

where $\beta_{0,d,s}$, $\beta_{1,d,s}$ are now specific for each combination of division and season.

4.3 Performance evaluation

The increased level of specificity going from the *DSP*-model, via *SP*-, to *NP*-models is achieved at the expense of a reduced sample size on which the model is calibrated, giving rise to a trade-off between the capacity of the calibration model to take spatial heterogeneity into account and data availability. In fact, although the performances in fitting increase by definition when a more specific model is employed, this may not happen in prediction because of model overparameterization. Overparameterization occurs when the amount of information contained in the calibration data is not enough to estimate the model parameters. The resulting model fits the calibration dataset, but produces large errors when used in prediction. Conversely, underparameterization refers to a situation in which the available

information is not fully exploited by the restricted set of model parameters. Therefore, over/under-parameterization must be minimized to achieve the best predictive capacity. In order to choose the best modelling solution with the data at hand, we assessed the prediction performance of different NDVI product pairs for the different regression options using a cross-validation jackknifing technique, where one full year of data was left out at a time.

For each jackknifed year, regression coefficients were estimated on the retained dataset and subsequently applied to estimate $NDVI\_M^*_{d,s}$ of the year left apart. Performances were then evaluated using the cross-validated R^2 (i.e., R^2_{cv}). The R^2_{cv} measures the fraction of total NDVI variability that is explained by the model in prediction, in all the dimensions of the database under consideration. For example, for the global *DSP*-model (Equation 1) the total variability is characterised by the spatial (division), seasonal, and interannual dimensions. As the main objective of our intercalibration is to accurately reconstruct the interannual variability of *NDVI** at division and seasonal level, we are not interested in the ability of our model to explain the variability in the spatial and seasonal dimensions. For the *DSP*-model evaluation, we therefore compute the R^2_{cv} within division and season ($R^2_{cv(wd,ws)}$), informing us on the temporal prediction capability only. This can be expressed as:

$$R^2_{cv(wd,ws)} = 1 - \frac{\sum_i^I \sum_d^D \sum_s^S \left(NDVI\_M^*_{i,d,s} - \widehat{NDVI\_M^*_{i,d,s}}\right)^2}{\sum_i^I \sum_d^D \sum_s^S \left(NDVI\_M^*_{i,d,s} - \overline{NDVI\_M^*_{d,s}}\right)^2} \quad (4)$$

where $\widehat{NDVI\_M^*_{i,d,s}}$ is the *NDVI_M** predicted by the model in year *i*, division *d*, and season s; and $\overline{NDVI\_M^*_{d,s}}$ is the average *NDVI_M** over the years for division *d* and season *s*. For clarity, in this study I=9 years, D=84 divisions, and S=2 seasons. By using the division- and season-specific *NDVI_M** averages in the denominator of Equations 4 instead of global average used in the standard R^2_{cv}, we measure to what extent the selected model performs

better than a naïve model that every year predicts a *NDVI_M** that equals the multi-annual average for each division and season.

4.4 Calculation of premium rates

To evaluate the impact of having longer *NDVI** time series (following intercalibration) on insurance pricing, we calculated premium rates using data for both the 2001-2012 period for $eMODIS_T$ as well as for the augmented period of 1981-2012 using the intercalibrated data. We estimated the premium rate as the expected value of the insurance payout rates using the historical distribution of z-scored *NDVI** (*zNDVI**). For clarity, *zNDVI** indicates how many standard deviations *NDVI** is above or below its division- and seasonal mean value. For illustration purposes we only present results for the SRSD season.

The insurance is structured as a simple index insurance contract that pays when livestock mortality predicted by *zNDVI** exceeds a predefined mortality level (called strike level). Explicitly, the payout rate (or indemnity) in any year, division, and season is calculated as follows:

$$indem_{i,d,s} = \max\left(0, f\left(zNDVI^*_{i,d,s}\right) - K\right) \quad (5)$$

where $f\left(zNDVI^*_{i,d,s}\right)$ is the response function yielding an index between 0 and 100 per cent that represents predicted livestock mortality conditional on $zNDVI^*_{i,d,s}$, and K is the strike level. The strike level is the value above which the contract will begin to indemnify and is selected by the insured at the inception of the contract. In the current IBLI implementation, the insured may select a strike level of either 10 or 15 per cent. The premium rate can then be calculated as the average of the historical predicted indemnities provided by the application of Equation 5 to the time series of $zNDVI^*_{i,d,s}$. Currently in the IBLI project, specific response

functions to predict livestock mortality from $zNDVI^*_{i,d,s}$ are created for each division and season, based on collected livestock mortality data and taking into account the spatial relationships between divisions. The precise procedure for this will be described in a forthcoming paper by Woodard et al. (in preparation). Here, to show the impact of the longer time series availability on the premium rates, we use a generic response function that describes mortality as an exponential decay function of $zNDVI^*_{i,d,s}$, i.e.:

$$M = f\left(zNDVI^*_{i,d,s}\right) = e^{-2.5-0.3zNDVI^*_{i,d,s}+0.3(zNDVI^*_{i,d,s})^2} \quad (6)$$

where M represents predicted livestock mortality. As we used a strike level (K) of 10 per cent, according to Equation (5) and (6), an indemnity would be granted when $zNDVI^*_{i,d,s}$ is smaller than -0.45 (i.e., when NDVI is smaller than the average value minus 0.45 standard deviations).

5 Results

5.1 Effect of division- and season-pooling on intercalibration

Table 2 presents the results for the global *DSP*-model that maps division- and season-level *NDVI_S** from GIMMS (slave) to *NDVI_M** from any of the other NDVI products (master) for all seasons, years, and divisions. When jointly analyzing all 84 divisions and 18 seasons, the high R^2_{cv} values (above 0.~~95~~91 for all products) indicate that *NDVI** from GIMMS well ~~correlated~~correlates to that obtained from other products. When we remove the contribution related to the model's ability to explain the variability in the spatial and seasonal dimensions (i.e., $R^2_{cv(wd,ws)}$), the values decrease to the range 0.~~789~~622-0.~~820~~672, depending on which dataset GIMMS is mapped to (Table 2). This value expresses the average capacity of the *DSP*-model to properly model the interannual variability of *NDVI_M** for an individual

division and season. $MODIS_{A\text{-}NASA}$ performs best according to R^2_{cv} and $R^2_{cv(wd,ws)}$, although differences with other products are small. $MODIS_{T\text{-}BOKU}$ showed the poorest overall relation with GIMMS in terms of the R^2-measures and the $RMSE_{cv}$. A possible explanation for poorer agreement is that in this product the quality flags delivered with the original MODIS data are not considered, because Vuolo et al. (2012) assume that poor observations have low NDVI-values and are corrected by the filtering technique. ~~Nonetheless,~~ During periods with more persistent cloud cover, this assumption may not be correct. For $MODIS_{T+A\text{-}BOKU}$ the non-consideration of quality flags is less of a problem due to the higher number of good NDVI observations available, as both the Terra and Aqua satellite are used. Figure 2 shows the corresponding scatterplots between *NDVI** from GIMMS and the other products. All regression lines are below the 1:1 line, indicating that a ~~positive~~negative bias exists, i.e., *NDVI** from GIMMS is on average higher than *NDVI** from the other products. This ~~positive~~data bias (Table 2) is originated by both slope and intercept of the linear regression: in all cases slope is significantly different from 1 and intercept from 0 ($p<0.01$). This bias of GIMMS is also found for Europe (Atzberger et al., 2013). The calibration could efficiently remove the existing data bias judging from the small model bias ($bias_{cv}$) reported in Table 2. From the global *DSP*-model we may conclude that *NDVI** from the operational NDVI products behave similar, and all have a strong correlation with GIMMS.

As the insurance contract is applied at the division level, we evaluated if a less parsimonious division-specific model (Equation 2) could provide a more accurate calibration. Figure 3 shows the frequency distribution of the division-level R^2_{cv} difference between applying the global *DSP*-model and the division-specific *SP*-models to individual divisions, in this case for $MODIS_{A\text{-}NASA}$~~.~~ ($MODIS_{A\text{-}NASA}$ is shown here as an example, as it performed best for the *DSP*-model, Table 2). For 69 of the 84 divisions, the *SP*-model ~~performed~~yielded better

predictive performances than the *DSP*-model, i.e. the *SP*-model better captures the interannual variability of *NDVI** for these divisions. Similar results were obtained for the other operational NDVI products (Table 3). It is beyond the scope of this paper to pinpoint the precise causes of such spatial heterogeneity in the relationship between NDVI products (see section 4.2 for possible explanations), but this finding clearly indicates that cross-division differences in the relationship of *NDVI** are important, and ~~can better be accounted for using~~consequently the intercalibration can be achieved more accurately with a division-specific model. Despite the large reduction in sample size (n=1,512 for *DSP* versus n=18 for *SP*), our cross-validated results show that, on average, the more specific SP-model is not overparameterized and GIMMS *NDVI** can be more accurately mapped to *NDVI** from the operational products using the *SP*-model, as compared to the *DSP*-model.

The least parsimonious (n=9) division- and season-specific *NP*-model outperforms the *SP*-model in more than 50 per cent of the divisions for most operational NDVI products (Table 3 ~~and Figure 4~~). This implies the presence of a seasonal effect on the relationship of *NDVI** derived from GIMMS and operational products. This difference in performance was clustered in space, with the *SP*-model performing better in areas in the west (Turkana County) and the *NP*-model in eastern counties (Figure ~~5~~4). The better performance of the *NP*-model in eastern counties coincides with areas that have a high dynamic range of NDVI during the SRSD season, but a lower dynamic range during the LRLD season (Figure ~~6~~5). This suggests that if seasonal NDVI characteristics strongly diverge between both seasons, season-specific (*NP*) models are more effective in mapping *NDVI** derived from GIMMS to ~~that of~~*NDVI** from operational products. This could partly result from the impact of low signal-to-noise ratios on *NDVI** during relatively dry seasons. For products other than $MODIS_{A\text{-}NASA}$ (as in the example of Figure ~~5~~4) the general pattern is the same despite some changes in the values (not

shown). Note that on average the *SP*-model performs better for the SRSD as compared to the LRLD season, possibly due to the reduced NDVI dynamic range during LRLD in many divisions.

Figure ~~5a~~4a suggests that for each division a different model may be selected to obtain optimal relationships between *NDVI** derived from different sources. Arguably, this may in fact be an option for creating the long time series of drought indices for each division. Which model should be used for each individual division would depend on the operational NDVI time series selected (i.e. Figure ~~5a~~4a may deviate for other products). Here, in order to select one single modelling solution, we pragmatically evaluated the magnitude of performance improvement achieved increasing the specificity of the modelling solution (and thus, reducing its parsimony). For the example of $MODIS_{A\text{-}NASA}$, only for three divisions the *DSP*-model's R^2_{cv} was more than 0.10 higher than that of the *SP*-model; for most divisions the *SP*-model performed at least similar to *DSP*, if not much better (Figure 3). On the contrary, when comparing the *SP*- to the more specific *NP*-model, only 23 per cent of the divisions for LRLD, and 6 per cent for SRSD show an improved performance by over 0.10 (ΔR^2_{cv}) for *NP*-models ~~(see also Figure 4).~~ Therefore, given the relatively close similarity between *NP*- and *SP*-models performances, we chose to confine ourselves in the further analysis to the more parsimonious *SP*-model. The use of a single model-type would also imply a simpler and more consistent solution for insurance design. At the same time, we acknowledge however that no single models 'wins' across all divisions, and that the selection of different models for each division may be a better option based on purely empirical grounds.

5.2 Comparison of NDVI products

Figure ~~7~~6 shows the division-level R^2_{cv} results for the *SP*-model for all NDVI product pairs analysed. The first row depicts the relationship of *NDVI** from GIMMS (slave) with *NDVI** from all operational NDVI products. Although differences are small, on average for all divisions SPOT-VGT has the highest (0.85) and MODIS$_{\text{T-BOKU}}$ the lowest R^2_{cv} (0.79). The spatial pattern of the relationship is very similar for all NDVI products in relation to GIMMS. R^2_{cv}-values below 0.50 are found for all products in the driest divisions of Turkana in the north-~~east~~west. The NDVI dynamic range is extremely low in this area (Figure ~~6~~5), leading to low signal-to-noise ratios that negatively affect the relationships between products~~. This~~ and moreover question the usability of the uncalibrated *NDVI** for insurance purposes in these regions. The NDVI dynamic range is illustrated in Figure ~~8~~7 for one poorly-performing division in Turkana (Lokichar), and a good-performing division in Moyale (Obbu). Despite the poor performances of some divisions, in many divisions the performance can be considered good. For example between GIMMS and SPOT-VGT, nearly half of all divisions have an R^2_{cv} above 0.90, and R^2_{cv} above 0.80 are found in 82 per cent of the divisions (Table 4). eMODIS$_{\text{T}}$ shows similar figures, while the other MODIS products demonstrate a somewhat poorer relationship with GIMMS.

Compared to the relationship between operational NDVI products and GIMMS, the relationships among operational products showed a higher R^2_{cv} (Figure ~~7~~6). This may be largely ~~attributed~~attributable to the closer similarity of the spectral response functions of SPOT-VGT and MODIS. Because all MODIS products are based on the same sensor (although flown on two satellites), it is logical that the relationship between SPOT-VGT and MODIS shows stronger deviations as between individual MODIS products. Nonetheless, between SPOT-VGT and eMODIS$_{\text{T}}$, 73 per cent of all divisions have an R^2_{cv} above 0.95, and 94 per cent above 0.90, indicating overall good comparability of *NDVI**. MODIS$_{\text{T-BOKU}}$ has

the poorest relationship with SPOT-VGT (21 and 67 per cent above R^2_{cv}'s of 0.95 and 0.90, respectively), and also with other MODIS products. The different filtering used, and the fact that the quality flags are not used for this product, may explain this behaviour, given that $MODIS_{T\text{-}NASA}$ (based on the same MOD13Q1 product) performs much better. Still we can conclude from Figure ~~7~~6 that most operational NDVI products provide comparable *NDVI** values for the majority of divisions. In an operational context of division-level drought monitoring, this finding would allow their interchangeable use, which can be important in case one satellite sensor fails.

While the calibration performance in terms of R^2_{cv} informs us about the overall correlation between GIMMS and the operational products, a key interest for insurance payouts is whether different products are capable of identifying droughts and their relative severity. Despite relative poor calibration, it may still be possible that various NDVI products identify droughts similarly. As an example, Figure ~~9~~8 compares the time series of z-scored *NDVI** of GIMMS and three operational NDVI products for three divisions in Wajir County, each characterized by a different quality of the calibration against GIMMS (as indicated by the R^2_{cv}-values for the *SP*-model). For the Gurar-Bute division, we can observe that three major droughts are identified by all NDVI products: in order of decreasing severity these are 2011 LRLD, 2005 LRLD, and 2010 SRSD. For Buna and Eldas (R^2_{cv} for GIMMS versus SPOT-VGT of 0.82 and 0.66) five seasons are identified as dry by all products with z-scores below -0.5. Despite that the major drought (2011 LRLD) is equally identified by all products, the severity ranking of these seasons differs slightly. Nonetheless, this comparison shows that even for relatively lower R^2_{cv}-values, which are not very common in our analysis (Table 4), we may attain a reasonably comparable estimate of drought occurrence.

Overall, our findings suggest that good perspectives exist for extending the operational NDVI products back in time to create longer time series of drought indices for livestock insurance in Kenya. The only exceptions are the few poorly-performing divisions, notably the very arid divisions in Turkana~~.~~ with low signal-to-noise ratios.

5.3 Effect on premium rates

Figure ~~10~~9 shows the premium rates for the SRSD season as calculated using Equation (5) and (6) for the period 2001-2012 from eMODIS$_T$ and for the period 1981-2012 using the intercalibrated data. The average premium rate across all divisions equals 2.54 per cent for the period 2001-2012, and 2.95 per cent for the longer period. For individual divisions, the premium rate estimate changes significantly when using the longer period of intercalibrated data. For example, in the Kirisia division in the east of Samburu County, the estimated premium rate more than doubled when using the 1981-2012 period (4.6%) as compared to using only 2001-2012 (2.2%). Higher rates imply a higher expected livestock mortality, and consequently a higher cost for the pastoralist to purchase the insurance (assuming equal risk-loading by the insurer), but simultaneously this could benefit the sustainability of an insurance scheme from the insurer's perspective. The standard deviation of the difference across all divisions is 1.14 per cent, which represents a significant amount of rate volatility in insurance terms. Such volatility may be expected in a dryland pastoralist system with strong deviations in asset losses. This example illustrates that the addition of 20 years of data has a strong impact on the premium.

In statistical terms, the premium rates are expected to be more efficient and robust for a larger sample size (that is more seasons), provided that the relationship between mortality and the index is stable over time and that the intercalibration between sensors is effective. The

stronger statistical basis may motivate insurers to reduce the risk-loading (Biener, 2013), which could partly off-set the increased premiums that were calculated based on the longer time period for a large number of divisions. Further study should reveal whether the premium rates based on longer time series are also more efficient in ascertaining a sustainable insurance scheme. In this context, sustainable implies that in the long run, the scheme is attractive for both the insured and the insurer. In section 6 we will further discuss limitations of long time series, and their assumed stationarity, for effectively representing livestock mortality risks.

6 Discussion

Our study confirms that the GIMMS product is the most dissimilar among all NDVI products tested, a finding that can be attributed largely to the fact that the AVHRR sensor was not designed specifically for vegetation studies and consequently has much broader spectral bands for measuring red and NIR reflection (Trishchenko et al., 2002; Miura et al., 2006). We did not find evidence of markedly different performance among the operational NDVI products: pre-processing algorithms (i.e., temporal filtering) are partly responsible for differences that did occur. Our pragmatic approach of first focussing on the relevant aggregate drought indices (*NDVI**), and subsequently comparing these across products, provided comparable measures across different NDVI products for most divisions of arid Kenya.

Despite the high overall R^2_{cv} (>0.95) between GIMMS and operational products for the global *DSP*-model (divisions and seasons pooled), division-specific models better predicted the temporal variability in *NDVI** of different sensors per division. This outcome is not trivial as the more specific models are tuned on a reduced sample size (as compared to the global

DSP-model) and are thus more exposed to potential over-fitting problems. This finding is supported by other studies that indicate location-specific dependencies on the relationship between NDVI derived from different sensors (e.g., Miura et al., 2006; Swinnen and Veroustraete, 2008). Except for approximately 10 to 20 per cent of the divisions that have a very limited NDVI dynamic range, seasons with poor vegetation characteristics (due to drought conditions) could be properly identified by all NDVI products using the season-pooled division-specific *SP*-model. Further disaggregation achieved by the *NP*-model, where the calibration is performed by each division and separately for the two seasons of interest, yielded relatively small improvements in prediction for more than 50 per cent of the divisions. Our results demonstrate that longer operational time series of the drought index can effectively be constructed for most divisions by mapping *NDVI** from GIMMS to that of operational products. We fully acknowledge, however, that other approaches to achieve this may be identified, and that further improvements can be envisaged. Here we non-exhaustively discuss a few possible improvements or adaptations to our approach.

First, for several operational NDVI products we have more years of overlap available with the GIMMS dataset than the 18 seasons between October 2002 and September 2011 used here. These 18 seasons overlap between all products evaluated in our study, and were selected to provide a fair comparison among products. However, for example for SPOT-VGT, eight more seasons are available. Incorporating these seasons in the regression may improve the estimation of the regression coefficients, and thus results in improved consistency of the combined long-term time series.

Second, the observed deviation from linearity in the relationship between GIMMS and other NDVI products may be approached using a quadratic regression (see also Miura et al.,

2006)~~Second~~ to test the trade-off between the benefit of a potentially more appropriate model and the drawbacks of an increased parameterization.

Third, besides the three calibration options that we tested here (Equation 1-3), other intermediate levels of division- and season-pooling of *NDVI** can be envisaged and may have benefits. For example, all divisions within a county could be used in a single regression equation. Within that county, depending if the characteristics of the two seasons are very distinct, also the seasons could be separately analysed. Another intermediate possibility would be to use all divisions simultaneously in a fixed effects panel regression model, in which a single slope is obtained for all divisions and different intercepts for each, thus reducing the amount of parameters to be estimated as compared to *SP*-models while increasing sample size (Baltagi, 2008). A more drastic consideration is whether we should stick to division-boundaries, or use a better ecological stratification of the area, possibly based on NDVI series as well (de Bie et al., 2011). For IBLI this would not be a good option, given that livestock mortality data are available at the division level, and divisions are for insurers and pastoralists the most logical unit for having the same insurance premiums and payout. Alternatively, divisions could be pooled based on ecological characteristics, for example through similarity-based clustering of their average NDVI profiles. The main advantage of this would be the increase of the sample size, possibly leading to more reliable estimates of the regression coefficients, while still keeping homogeneity of insurance contract within each division. While further empirical testing of different pooling levels could lead to slight improvements of calibration performances, it is likely that bigger benefit can be obtained by improving the design of the drought index itself.

Two ways to improve the drought index (currently defined as the z-score of the division spatial average of the mean NDVI over two fixed time periods corresponding to the growing seasons, and jointly covering a whole calendar year without gaps) can be envisaged. A first way would be through adapting the spatial aggregation step. Currently all pixels within a division are incorporated when calculating *NDVI**. However, many pixels may have low signal-to-noise ratios (because vegetation is absent throughout the year, for example) affecting the calibration reliability. In addition, these pixels may not represent locations where livestock is actually grazing. Such pixels are less relevant concerning their effect on livestock conditions and could be excluded 1) by setting a threshold on the pixel's mean NDVI and/or its temporal variability, or 2) by incorporating land cover maps such as AfriCover (e.g., Genovese et al., 2001; Rojas et al., 2011).

A second way to enhance the performance of the drought index, both in the construction of the long term archive and in the actual application of the index in the insurance scheme, could be through changing the definition of the seasons under consideration. Currently, in analogy to Chantarat et al. (2013) and the current IBLI design, the LRLD and SRSD together cover a full year. This implies, however, that a significant proportion of each season contains a (relatively) dry period. During this period, biomass is not developing and consequently NDVI values provide information of limited relevance regarding grazing opportunities for livestock. A better and more realistic tuning of the considered period is thus also expected to increase the correlation of the index with actual livestock mortality and therefore to further reduce the insurance basis risk. Moreover, given that exposure of bare soil impacts reflection differently depending on the spectral response functions, low biomass conditions tend to decrease the signal-to-noise ratio of the NDVI measurement, and as such decrease the comparability of *NDVI** across sensors. A straightforward solution could be to shorten the

seasons by removing the final one to three months of each season that is consistently dominated by dry conditions (see also Figure ~~8~~7). However, the optimal time period for aggregation could change from division to division. A more appropriate approach could rely on the automated identification of start- and end-of-season from the NDVI time series at pixel-level (Meroni et al., 2013; Vrieling et al., 2013) ~~or aggregated per division (Rojas et al., 2011; Vrieling et al., 2011)~~ or aggregated per division (Rojas et al., 2011; Vrieling et al., 2011).

This study started from the premise that longer time series can better capture the full range of climatic variability and the related drought probability, resulting in improved pricing of insurance contracts. This premise is based on the fact that stationarity of NDVI can be assumed over the considered period. We should place two critical notes however. First, human-induced land use changes within the past 30 years (e.g., Brink and Eva, 2009) may have changed NDVI levels, which do not relate to drought. Second, if trends are present in the NDVI data (possibly due to climatic changes), the longer record may not help to better define drought probability for the upcoming season(s), unless the trends are accounted for. While land use could be relatively stable, and trends may be absent for many divisions, more detailed analysis may be needed for future use of long NDVI records in index-insurance. In this respect, comparison with other indices could also be explored, for example using tree ring data (Bell et al., 2013) that are now available for some parts of Africa (Gebrekirstos et al., 2009).

This discussion contends that ample scope exists for further improving the remote sensing component of the IBLI project. While we achieved to compare several NDVI sources, and provided regression coefficients for creating longer time series of the *NDVI** drought index,

the final usefulness of the index (as derived from different sources) can only be ascertained when it can effectively model what is insured, i.e., livestock losses. Although drought is the main cause of livestock mortality in arid Kenya, above-normal wet conditions may also trigger livestock diseases such as East Coast Fever (Homewood et al., 2006), and Rift Valley Fever (Anyamba et al., 2009). While this is somewhat acknowledged in the current IBLI scheme through the introduction of a quadratic term (Equation 6), better predictions of such disease outbreaks may be possible. In this regard, NDVI series could also provide useful information (e.g., Norval et al., 1991; Anyamba et al., 2009; Trevennec et al., 2012). Upcoming household surveys within the IBLI project should reveal the importance of these diseases for livestock mortality in the region, and efforts are underway within the International Livestock Research Institute (and elsewhere) for NDVI-aided outbreak prediction. Eventually, NDVI-derived outbreak probabilities could be incorporated in the IBLI design to better account for increased mortality during above-normal wet conditions as well.

7 Conclusions

Index-based livestock insurance requires an accurate estimate of livestock mortality from an index to determine the insurance premium, and up-to-date information for determining payouts. Given that most mortality is drought-related, spatially- and temporally-aggregated NDVI is used as drought index input to livestock insurance in Kenya. With the aim of creating a long (>30 year) operational record of division-level seasonal drought indices for the arid lands in Kenya, we compared the non-operational 30-year GIMMS AVHRR record with six operational NDVI products (from MODIS and SPOT-VGT) and three modelling options. Based on cross-validated results, we conclude that division-specific models are the most effective in linking the division-level variability of the drought index (*NDVI**) between

the various products. In relation to the long-term GIMMS record, the *SP*-model explained over 80 per cent of the *NDVI** variance for more than 80 per cent of all divisions for SPOT-VGT and eMODIS$_T$. This implies that for most divisions, good scope ~~exist~~exists for historically extending the aggregated drought index, thus providing a longer operational record for insurance purposes. Using a longer record has a significant influence on the insurance premium rates, as shown in this paper. We defined several possible future improvements to the drought index, which may also have a positive impact on the comparability of the resulting drought index time series.

While our work specifically focussed on the demands of the IBLI project in Kenya, the need for long time series of drought indices is not specific to this project. Although not always effective or successful (Binswanger-Mkhize, 2012), index-based insurance of crop or livestock is seen by many as having a great potential for increasing agricultural production among smallholder farmers (Hazell and Hess, 2010; Coe and Stern, 2011), and many initiatives and pilot projects currently exist. Despite limitations (Turvey and McLaurin, 2012), NDVI series are frequently used or considered in these projects, including for example for livestock insurance in Mongolia (Mahul and Skees, 2007) and the ongoing project "Evaluating remote sensing for index insurance" of the Weather Risk Management Facility (http://www.ifad.org/ruralfinance/wrmf/). Given the need for long time series for insurance design and pricing, our current work may further guide other index-insurance projects that seek to combine NDVI series.

References

Anyamba, A., Chretien, J.-P., Small, J., Tucker, C.J., Formenty, P.B., Richardson, J.H., Britch, S.C., Schnabelf, D.C., Erickson, R.L., Linthicum, K.J., 2009. Prediction of a Rift Valley fever outbreak. Proceedings of the National Academy of Sciences of the United States of America 106, 955-959.

Atzberger, C., Eilers, P.H.C., 2011. A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of South America. International Journal of Digital Earth 4, 365-386.

Atzberger, C., Klisch, A., Mattiuzzi, M., Vuolo, F., 2013. Phenological metrics derived over the European continent from NDVI3g data and MODIS time series. Remote Sensing under review.

Baltagi, B.H., 2008. Econometric Analysis of Panel Data, 4th edition. John Wiley & Sons Ltd, Chichester, West Sussex, UK.

Barnett, B.J., Barrett, C.B., Skees, J.R., 2008. Poverty Traps and Index-Based Risk Transfer Products. World Development 36, 1766-1785.

Barrett, C.B., Barnett, B.J., Carter, M.R., Chantarat, S., Hansen, J.W., Mude, A.G., Osgood, D., Skees, J.R., Turvey, C.G., Ward, M.N., 2007. Poverty Traps and Climate Risk: Limitations and Opportunities of Index-Based Risk Financing. IRI Technical Report No. 07-02.

Barrett, C.B., Marenya, P.P., McPeak, J., Minten, B., Murithi, F., Oluoch-Kosura, W., Place, F., Randrianarisoa, J.C., Rasambainarivo, J., Wangila, J., 2006. Welfare dynamics in rural Kenya and Madagascar. Journal of Development Studies 42, 248-277.

Bell, A.R., Osgood, D.E., Cook, B.I., Anchukaitis, K.J., McCarney, G.R., Greene, A.M., Buckley, B.M., Cook, E.R., 2013. Paleoclimate histories improve access and sustainability in index insurance programs. Global Environmental Change 23, 774-781.

Biener, C., 2013. Pricing in microinsurance markets. World Development 41, 132-144.

Binswanger-Mkhize, H.P., 2012. Is there too much hype about index-based agricultural insurance? The Journal of Development Studies 48, 187-200.

Bonifacio, R., Dugdale, G., Milford, J.R., 1993. Sahelian rangeland production in relation to rainfall estimates from Meteosat. International Journal of Remote Sensing 14, 2695-2711.

Brink, A.B., Eva, H.D., 2009. Monitoring 25 years of land cover change dynamics in Africa: A sample based remote sensing approach. Applied Geography 29, 501-512.

Brown, M.E., Lary, D.J., Vrieling, A., Stathakis, D., Mussa, H., 2008. Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS. International Journal of Remote Sensing 29, 7141-7158.

Brown, M.E., Osgood, D.E., Carriquiry, M.A., 2011. Science-based insurance. Nature Geoscience 4, 213-214.

Brown, M.E., Pinzón, J.E., Didan, K., Morisette, J.T., Tucker, C.J., 2006. Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. IEEE Transactions on Geoscience and Remote Sensing 44, 1787-1793.

Chantarat, S., Mude, A.G., Barrett, C.B., Carter, M.R., 2013. Designing index-based livestock insurance for managing asset risk in northern Kenya. Journal of Risk and Insurance 80, 205-237.

Chen, J., Jönsson, P., Tamura, M., Gu, Z.H., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sensing of Environment 91, 332-344.

Coe, R., Stern, R.D., 2011. Assessing and addressing climate-induced risk in sub-Saharan rainfed agriculture: lessons learned. Experimental Agriculture 47, 395-410.

de Bie, C.A.J.M., Khan, M.R., Smakhtin, V.U., Venus, V., Weir, M.J.C., Smaling, E.M.A., 2011. Analysis of multi-temporal SPOT NDVI images for small-scale land-use mapping. International Journal of Remote Sensing 32, 6673-6693.

Funk, C., Budde, M.E., 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. Remote Sensing of Environment 113, 115-125.

Gallo, K., Li, L., Reed, B., Eidenshink, J., Dwyer, J., 2005. Multi-platform comparisons of MODIS and AVHRR normalized difference vegetation index data. Remote Sensing of Environment 99, 221-231.

Gao, B.C., 2000. A practical method for simulating AVHRR-consistent NDVI data series using narrow MODIS channels in the 0.5-1.0 mu m spectral range. IEEE Transactions on Geoscience and Remote Sensing 38, 1969-1975.

Gebrekirstos, A., Worbes, M., Teketay, D., Fetene, M., Mitloehner, R., 2009. Stable carbon isotope ratios in tree rings of co-occurring species from semi-arid tropics in Africa: patterns and climatic signals. Global and Planetary Change 66, 253-260.

Genovese, G., Vignolles, C., Nègre, T., Passera, G., 2001. A methodology for a combined use of normalised difference vegetation index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain. Agronomie 21, 91-111.

Gutman, G., Masek, J.G., 2012. Long-term time series of the Earth's land-surface observations from space. International Journal of Remote Sensing 33, 4700-4719.

Hazell, P.R., Hess, U., 2010. Drought insurance for agricultural development and food security in dryland areas. Food Security 2, 395-405.

Homewood, K., Trench, P., Randall, S., Lynen, G., Bishop, B., 2006. Livestock health and socio-economic impacts of a veterinary intervention in Maasailand: Infection-and-treatment vaccine against East Coast fever. Agricultural Systems 89, 248-271.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83, 195-213.

Jacobs, T., Borstlap, G., Bartholomé, E., Maathuis, B.H.P., 2008. DevCoCast in support of environmental management and sustainable development in Africa. Proceedings of the 7th AARSE Conference, Accra, Ghana, pp. 1-6.

James, M.E., Kalluri, S.N.V., 1994. The Pathfinder AVHRR land data set: an improved coarse resolution data set for terrestrial monitoring. International Journal of Remote Sensing 15, 3347-3363.

Jenkerson, C.B., Maiersperger, T., Schmidt, G., 2010. eMODIS: a user-friendly data source. Open-File Report 2010–1055. USGS, p. 10 pages.

Leblois, A., Quirion, P., 2013. Agricultural insurances based on meteorological indices: realizations, methods and research challenges. Meteorological Applications 20, 1-9.

Little, P.D., Smith, K., Cellarius, B.A., Coppock, D.L., Barrett, C.B., 2001. Avoiding disaster: Diversification and risk management among east African herders. Development and Change 32, 401-433.

Mahul, O., Skees, J., 2007. Managing agricultural risk at the country level: the case of index-based livestock insurance in Mongolia. Policy Research Working Paper 4325. Washington D.C., p. 37.

Meroni, M., Verstraete, M.M., Rembold, F., Urbano, F., Kayitakire, F., 2013. A phenology-based method to derive biomass production anomaly for food security monitoring in the Horn of Africa. International Journal of Remote Sensing accepted.

Miura, T., Huete, A., Yoshioka, H., 2006. An empirical investigation of cross-sensor relationships of NDVI and red/near-infrared reflectance using EO-1 hyperion data. Remote Sensing of Environment 100, 223-236.

Myneni, R.B., Asrar, G., 1994. Atmospheric effects and spectral vegetation indexes. Remote Sensing of Environment 47, 390-402.

Nkedianye, D., de Leeuw, J., Ogutu, J.O., Said, M.Y., Saidimu, T.L., Kifugo, S.C., Kaelo, D.S., Reid, R.S., 2011. Mobility and livestock mortality in communally used pastoral areas: the impact of the 2005-2006 drought on livestock mortality in Maasailand. Pastoralism 1, 17.

Norval, R.A.I., Perry, B.D., Gebreab, F., Lessard, P., 1991. East Coast Fever: a problem of the future for the Horn of Africa. Preventive Veterinary Medicine 10, 163-172.

Onono, J.O., Wieland, B., Rushton, J., 2013. Productivity in different cattle production systems in Kenya. Tropical Animal Health and Production 45, 423-430.

Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., Roy, D., Junchang, J., Vermote, E., Prince, S., Nagol, J., Justice, C., Schaaf, C., Jicheng, L., Privette, J., Pinheiro, A., 2007. Generating a long-term land data record from the AVHRR and MODIS Instruments. Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, pp. 1021-1025.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Koppen-Geiger climate classification. Hydrology and Earth System Sciences 11, 1633-1644.

Pinzón, J.E., Tucker, C.J., 2013. A non-stationary 1981-2012 AVHRR $NDVI_{3g}$ time series. Remote Sensing under review.

Rahman, H., Dedieu, G., 1994. SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. International Journal of Remote Sensing 15, 123-143.

Rojas, O., Vrieling, A., Rembold, F., 2011. Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. Remote Sensing of Environment 115, 343-352.

Sandmeier, S., Muller, C., Hosgood, B., Andreoli, G., 1998. Physical mechanisms in hyperspectral BRDF data of grass and watercress. Remote Sensing of Environment 66, 222-233.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36, 1627-1639.

Song, Y., Ma, M., Veroustraete, F., 2010. Comparison and conversion of AVHRR GIMMS and SPOT VEGETATION NDVI data in China. International Journal of Remote Sensing 31, 2377-2392.

Steven, M.D., Malthus, T.J., Baret, F., Xu, H., Chopping, M.J., 2003. Intercalibration of vegetation indices from different sensor systems. Remote Sensing of Environment 88, 412-422.

Swets, D.L., Reed, B.C., Rowland, J.D., Marko, S.E., 1999. A weighted least-squares approach to temporal NDVI smoothing. Proceedings of the 1999 ASPRS Annual Conference. American Society of Photogrammetric Remote Sensing, Portland, Oregon, pp. 526-536.

Swinnen, E., Veroustraete, F., 2008. Extending the SPOT-VEGETATION NDVI time series (1998-2006) back in time with NOAA-AVHRR data (1985-1998) for southern Africa. IEEE Transactions on Geoscience and Remote Sensing 46, 558-572.

Trevennec, C., Pittiglio, C., Wainwright, S., Plee, L., Pinto, J., Lubroth, J., Martin, V., 2012. Rift Valley fever: vigilance needed inthe coming months. EMPRES Watch 27, 1-8.

Trishchenko, A.P., 2009. Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors: Extension to AVHRR NOAA-17,18 and METOP-A. Remote Sensing of Environment 113, 335-341.

Trishchenko, A.P., Cihlar, J., Li, Z.Q., 2002. Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors. Remote Sensing of Environment 81, 1-18.

Tucker, C.J., Pinzón, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F., El Saleous, N., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. International Journal of Remote Sensing 26, 4485-4498.

Turvey, C.G., McLaurin, M.K., 2012. Applicability of the Normalized Difference Vegetation Index (NDVI) in Index-Based Crop Insurance Design. Weather, Climate, and Society 4, 271-284.

Vrieling, A., de Beurs, K.M., Brown, M.E., 2011. Variability of African farming systems from phenological analysis of NDVI time series. Climatic Change 109, 455-477.

Vrieling, A., de Leeuw, J., Said, M.Y., 2013. Length of growing period over Africa: variability and trends from 30 years of NDVI time series. Remote Sensing 5, 982-1000.

Vuolo, F., Mattiuzzi, M., Klisch, A., Atzberger, C., 2012. Data service platform for MODIS NDVI time series pre-processing at BOKU Vienna: current status and future perspectives. SPIE 8538, Earth Resources and Environmental Remote Sensing/GIS Applications III, 85380A, p. 85380A.

Woodard, J., Shee, A., Mude, A.G., in preparation. A spatial econometric approach to scalable index insurance against drought related livestock mortality in Kenya.

Appendix A: Detailed description of NDVI data sets

A.1 GIMMS AVHRR

To obtain long time series of NDVI data, we used the 8-km resolution NDVI dataset that was constructed by the GIMMS project. This 15-day (two maximum-value composites per month) product covers July 1981 to December 2011. The AVHRR sensors used to construct the dataset were flown on six satellites. The GIMMS AVHRR dataset has been corrected for factors that do not relate to changes in vegetation greenness, and the latest version (NDVI3g) applies an improved cloud masking as compared to older versions of the GIMMS dataset (Tucker et al., 2005; Pinzón and Tucker, 2013).

A.2 SPOT-VGT

We obtained ten-daily SPOT-VGT NDVI composites (S10 product) with a 1-km spatial resolution for 1998-present through the Flemish Institute for Technological Research (VITO). Three composites cover exactly one month, i.e. for day 1-10, 11-20, and 21-last day of each month. Following atmospheric correction (Rahman and Dedieu, 1994), a constrained view-angle maximum value composite rule is applied. For Africa, the data are available in near-real time through the VGT4Africa and GEONETCast projects (Jacobs et al., 2008). We used the quality flags provided with the data to discard observations affected by clouds or shadows, or that otherwise have a bad radiometric quality in the red or NIR band.

A.3 MODIS

We used two series of 16-day NDVI constrained view-angle maximum value composites from the 250-m resolution global MODIS vegetation indices product Collection 5, i.e. for Terra (MOD13Q1) and for Aqua (MYD13Q1). Similarly to the SPOT-VGT product, the

maximum-value compositing employed for these products selects the highest NDVI values, but constrains the candidate pixels by the view angle (Huete et al., 2002). We refer to this Terra-derived product here as $MODIS_{T\text{-}NASA}$ and the Aqua-derived product as $MODIS_{A\text{-}NASA}$. Quality flags provided with the data were used to mask out unreliable observations (i.e., MODIS quality flag greater than 1 and usefulness flag greater than 5).

A.4 MODIS – Whittaker filter

The University of Natural Resources and Applied Life Sciences (BOKU) in Vienna, Austria, provides on-demand temporally-filtered MODIS NDVI composites, based on the Terra- and ~~Terra plus~~ Aqua-derived products described above (section ~~3~~A.3). Their service includes temporal filtering, mosaicking, sub-setting, and reprojection, and can deliver data requests within a day, including near-real time acquisitions (Vuolo et al., 2012). The temporal filtering is achieved with a modified Whittaker filter (Atzberger and Eilers, 2011). This least squares approach incorporates a 'penalty' criterion regarding the smoothness of the resulting NDVI profile. Currently, MODIS quality indicators are not used to mask NDVI observations prior to filtering, following the assumption that poor observations have low NDVI values and will be corrected by the temporal filter (Vuolo et al., 2012). Exploiting the availability of NDVI products from Terra and Aqua platforms (both originally composited from NASA at 16-day, but with temporal compositing window shifted of 8 day), BOKU provides both a standard 16-day composite based on Terra only (here referred to as $MODIS_{T\text{-}BOKU}$) and Terra plus Aqua combined product produced every 8 days (here $MODIS_{T+A\text{-}BOKU}$). The product based on Aqua only was not considered in this paper, but is also processed by BOKU.

A.5 eMODIS

The eMODIS data (*e* for 'enhanced', 'expedited', and 'expandable') for Africa contain 10-day NDVI composites at 250-m resolution that are constructed using similar algorithms as the Collection-5 MODIS products (Jenkerson et al., 2010). The United States Geological Survey (USGS) has produced these composites since 2010 to better respond to user needs (regarding for example projections and compositing periods). Both historical archive data and near-real time composites are freely available online. The 10-day composites are produced every five days resulting in six composites per month. Here we only took the composites that cover days 1-10, 11-20, and 21-last day of each month, i.e. the same composite periods as for SPOT-VGT. While unfiltered composites can be obtained for the full Africa window, for this study we used the filtered eMODIS product for the East Africa window that is employed operationally for food security monitoring activities of FEWS-NET. The temporal filtering is based on a weighted least-squares regression approach that gives highest weights to local peaks in the NDVI profile, and lowest weights to local valleys (Swets et al., 1999). The filtered data are available for January 2001 until present. For clarity in this paper we add subscript "T" (for Terra) to refer to this dataset, i.e., $eMODIS_T$.

Tables

Table 1. Main characteristics of the NDVI products used in this study.

*Name dataset**	*Sensor*	*Platform*	*Data provider*	*Start*	*Resolution (m)*	*Composite period (days)*	*Temporal filtering***
GIMMS	AVHRR	NOAA (7 satellites)	NASA	1981 (-2011)	8,000	15	-
SPOT-VGT	VEGETATION	SPOT 4 and 5	VITO	1998	1,000	10	-
$MODIS_{T\text{-}NASA}$	MODIS	Terra	NASA	2000	250	16	-
$MODIS_{A\text{-}NASA}$	MODIS	Aqua	NASA	2002	250	16	-
$MODIS_{T\text{-}BOKU}$	MODIS	Terra	BOKU	2000	250	16	Atzberger and Eilers (2011)
$MODIS_{T+A\text{-}BOKU}$	MODIS	Terra+Aqua	BOKU	2002	250	8	Atzberger and Eilers (2011)
$eMODIS_T$	MODIS	Terra	USGS	2001	250	10	Swets et al. (1999)

~~*Note:~~

* *The names refer to the abbreviations for the datasets used in this article*
** *We here indicate here if the original data sources were filtered. Unfiltered datasets were smoothed by us using an iterative Savitzky-Golay filter (see section 4.1).*

Table 2. Statistics from the global *DSP*-model, where division- and season-level *NDVI** from GIMMS is mapped to *NDVI** of each of the six datasets listed in the table. $R^2_{cv(wd,ws)}$ is an aggregate measure of the temporal prediction capability for each division and season (section 4.3).

Name dataset	R^2_{cv}	$R^2_{cv(wd,ws)}$	$RMSE_{cv}$	*data bias*	*model bias (x1,000)*
SPOT-VGT	0.921	0.647	**0.029**	-0.065	0.136
$MODIS_{T\text{-}NASA}$	0.923	0.662	0.031	-0.039	0.122
$MODIS_{A\text{-}NASA}$	**0.929**	**0.672**	0.030	-0.044	0.142
$MODIS_{T\text{-}BOKU}$	0.917	0.622	0.033	**-0.037**	**0.110**
$MODIS_{T+A\text{-}BOKU}$	0.924	0.652	0.031	-0.046	0.127
$eMODIS_T$	0.922	0.658	0.030	-0.058	0.145

Table 3. Percentage of the divisions where the ~~pooled models (*DSP* and *SP*) outperform the~~ division- and season-specific models (*SP* and *NP*) outperform the pooled models (*DSP* and *SP*), respectively.

Name dataset	$R^2_{cv}(SP) > R^2_{cv}(DSP)$	$R^2_{cv}(NP) > R^2_{cv}(SP)$	
		LRLD	***SRSD***
SPOT-VGT	88	44	51
$MODIS_{T\text{-}NASA}$	82	61	57
$MODIS_{A\text{-}NASA}$	82	68	54
$MODIS_{T\text{-}BOKU}$	76	63	62
$MODIS_{T+A\text{-}BOKU}$	79	61	56
$eMODIS_{T}$	89	58	50

Table 4. Percentage of the divisions where R^2_{cv} from the *SP*-model ~~(taking GIMMS as the slave to be mapped to any of the datasets listed below)~~ is higher than the indicated thresholds.

Name dataset	R^2_{cv}(SP)				
	>0.50	***>0.70***	***>0.80***	***>0.90***	***>0.95***
SPOT-VGT	96	89	82	46	11
$MODIS_{T\text{-}NASA}$	95	90	77	38	6
$MODIS_{A\text{-}NASA}$	94	88	77	38	5
$MODIS_{T\text{-}BOKU}$	94	82	70	8	1
$MODIS_{T+A\text{-}BOKU}$	95	86	77	30	4
$eMODIS_{T}$	95	89	81	46	11