*Journal of Agricultural Science* (2015), **153**, 306–321. 
doi:10.1017/S0021859614000215

CROPS AND SOILS RESEARCH PAPER

# Water availability and crop growth at the crop plot level in South Africa modelled from satellite imagery

E. BLANC[1]* AND E. STROBL[2]

[1] *Massachusetts Institute of Technology, Joint Program on the Science and Policy of Global Change, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA*
[2] *Ecole Polytechnique, Department of Economics, Cedex 91128, Palaiseau, France*

(*Received 11 March 2013; revised 19 December 2013; accepted 19 February 2014; first published online 8 April 2014*)

## SUMMARY

Although the effect of weather on crop growth has been studied widely, the contribution of other water sources has been less well studied, mainly due to data limitation. To address this gap, the current analysis considers the importance of water availability on crop growth by taking advantage of crop field boundaries and information on South Africa's four major grain producing provinces (Northwest, Mpumalanga, Free State and Gauteng) provided by the Agricultural Geo-referenced Information System dataset. To capture crop growth along the crop growing cycle at the plot level, the MODIS's MOD13Q1 dataset of 16-day normalized difference vegetation index (NDVI) was used. To estimate the determinants of crop growth, weather effects were considered and represented by rainfall and reference evapotranspiration satellite derived data provided by the National Oceanic and Atmospheric Administration's RFE and GDAS dataset, respectively. Hydrologic and irrigation determinants were estimated based on the HYDRO1K river network dataset produced by the US Geological Survey. The results show that although weather is an important explanatory factor, other sources of water, such as irrigation, proximity to perennial and ephemeral rivers, and stream flow are also influential. Taking into account the interaction effects between weather and water availability related factors is also important to determine the effect of water availability on crop growth.

## INTRODUCTION

Agriculture is an important sector of South Africa's economy, representing 0·03 of GDP and 0·065 of total national exports (van Niekerk 2012). However, while 0·12 of South Africa's land can be used for crop production, only 0·22 of this is considered as high-potential arable land. The greatest limitation in this regard is the availability of water, with uneven and unreliable rainfall over South Africa's seven climatic regions, causing at times severe agricultural losses to commercial and subsistence farmers. Thus, understanding the role that water scarcity plays in crop productivity is a necessary step to ensuring food sufficiency and export earnings in the region.

* To whom all correspondence should be addressed. Email: eblanc@mit.edu

Crop growth studies considering the role of climatic factors generally can be categorized as either biophysical or statistical models. Biophysical models rely on controlled environments (growth chambers and/or field trial) experiments over small areas (Azam-Ali & Squire 2002), whereas statistical models are based on time series of single crop plots or temporally pooled cross-sectional data. Traditionally, the statistical models have been based on country or regional level data and thus the unit of analysis is spatially relatively aggregated. However, advances in satellite technologies have provided crop biomass production indicators, such as net primary production (NPP) or the normalized difference vegetation index (NDVI), at a high resolution, which have led to a rising number of crop development statistical models at a much lower unit of analysis than those using information collected by national statistical agencies. Benedetti & Rossini (1993) show that such an index provides a relatively low-cost and useful crop monitoring tool. For instance, a study by Yang *et al.* (1997) of Nebraska at a 1-km resolution showed a strong correlation between NDVI temporal change and temperature. Karnieli *et al.* (2010) use NDVI data at an 8 km resolution over North America and also find a significant correlation between NDVI and land surface temperature. In Western Africa, Malo & Nicholson (1990) showed a linear relationship between precipitation and NDVI at a 3 × 5-km resolution, whereas Schultz & Halpert (1993) showed at a global 1° grid scale that the influence of climatic factors on NDVI differ regionally.

Although the recent surge of satellite-data-based statistical models have enabled researchers to study the climatic determinants of crop production at a small unit of analysis while still covering large spatial areas, they have generally been limited to examining cropland without being able to distinguish between crop types over space and time – see, for example, Milesi *et al.* (2010) and Blanc & Strobl (2013). Additionally, existing analyses have not been able to consider water availability, which is location-specific. In the present paper, these weaknesses are addressed using detailed spatial crop field information in the four major grain producing provinces of South Africa – Northwest, Mpumalanga, Free State and Gauteng – representing >0·90 of the total summer grain growing area in the country (Ferreira *et al.* 2006). More specifically, in contrast to the usual gridded cropland data, the current dataset not only identifies exact crop field boundaries, but also the type of crop grown and whether the field is irrigated. Moreover, knowing the specific location of the crop plot enables one to estimate their distance from water sources. For each of these fields, a satellite-based measure of crop biomass productivity is calculated every 16 days over the crop's growing cycle and the impact of water availability is estimated in a regression framework. Water availability in this regard is measured along a number of dimensions, including local precipitation, river flow and distance to the nearest river.

The remainder of the paper is organized as follows. The next section describes the data and provides some summary statistics, whereas the Modelling Framework and Results section outlines the regression methodology and provides results. Conclusions, limitations and future work are discussed in the final section.

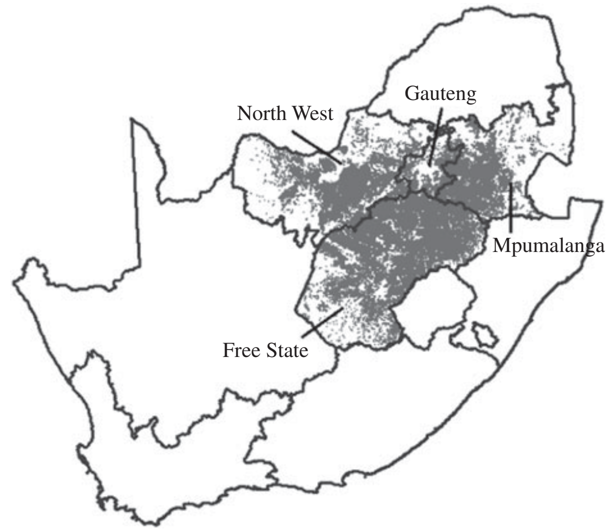


**Fig. 1.** Crop plot location in South Africa.

## DATA

### Spatial delineation: crop plot boundaries

Crop field boundaries were provided by the Agricultural Geo-referenced Information System (AGIS) developed by the South African National Department of Agriculture (available online from www.agis.agric.za). These data were available for the provinces of Free State, Gauteng, Northwest and Mpumalanga. The field boundaries in this dataset were determined using the Producer Independent Crop Estimate System (PICES) which combines satellite imagery, Geographic Information System (GIS), point frame statistical platforms and aerial observations (Ferreira *et al.* 2006). Satellite imagery of crop field of cultivated fields was obtained from the SPOT 5 satellite at a 2·5-m resolution. Crop field boundaries were then digitized using GIS. When cloud-free satellite images were not available, field polygons hidden by clouds were removed before processing. This ensured a more accurate dataset. Over the four regions of interest, PICES distinguished *c.* 280 000 plots covering an area of *c.* 6·5 million hectares.

To approximately match the resolution of the crop growth indicator data described below, which were only available at the 250-m resolution, the analysis was limited to plots larger than 6·25 ha. This restricted the sample to 213 110 crop fields. A geographical representation of the crop fields' location is presented in Fig. 1.

### Crop types

Within each crop plot described above, AGIS provided information on the crop cultivated. Crop types were determined using the digitized satellite images described above. Sample points were selected randomly and surveyed by trained observers from a very light aircraft in order to determine crop type (Ferreira *et al.* 2006). Crop information collected during the aerial surveys on the sample points was used as a training set for crop type classification for each field and for accuracy assessment. For crops planted sparsely, field verification on certain sample points was not possible and therefore larger areas were considered to include these crops. These estimated crop classifications were checked against a producer based survey for the Gauteng region. The Gauteng census survey showed that less than 1·8% of crop types had been misclassified.

All in all, seven summer crops were distinguished (classification of winter crops was generated for the season 2007/08 only, therefore these crops were not considered in the analysis): cotton, dry beans, groundnuts, maize, sorghum, soya and sunflower. Given the small number of cotton observations (110 plots over only one province), these were excluded from the analysis. Although the data also identified fallow and pasture plots, these were also not considered given the focus of the analysis on cropland.

Crop classification data were available for the summer season 2006/07 for the provinces of Free State, Gauteng, Northwest and Mpumalanga, and for one province only, Free State, for the summer season 2007/08. A geographical representation of crop types for each season is provided in Fig. 2. Figure 3 provides a close-up representation of crop type change across seasons. As can be seen, crops showed patterns of alternation between crop types and fallow/pasture.

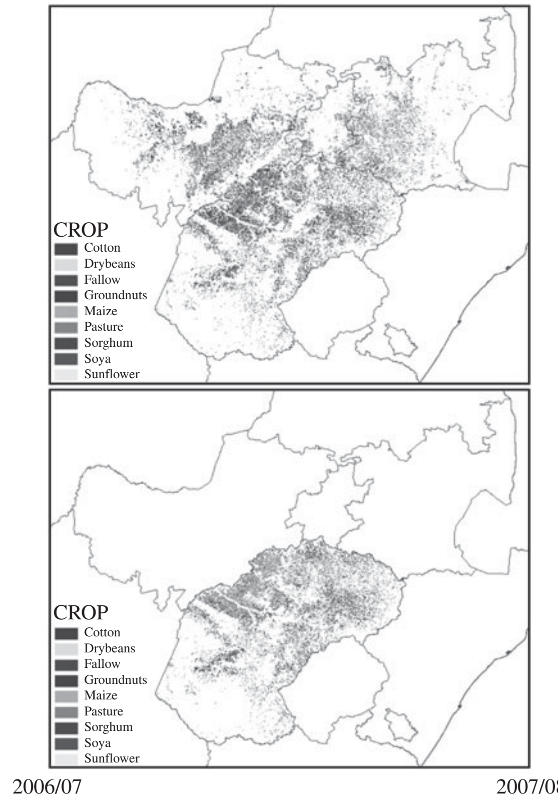


**Fig. 2.** Crop type location for the summer seasons 2006/07 and 2007/08.

### Crop growth measure

Crop biomass production was estimated using the NDVI. Vegetation indices such as the NDVI are particularly attractive as they provide consistent spatial and temporal representations of vegetation conditions. As a matter of fact, numerous studies have demonstrated that NDVI values are significantly correlated with crop yields including wheat (Das *et al.* 1993; Gupta *et al.* 1993; Doraiswamy & Cook 1995; Hochheim & Barber 1998; Labus *et al.* 2002), sorghum (Potdar 1993), corn (Hayes & Decker 1996; Prasad *et al.* 2006), rice (Quarmby *et al.* 1993; Nuarsa *et al.* 2011), soybean (Prasad *et al.* 2006), barley (Weissteiner & Kühbauch 2005), millet (Groten 1993) and tomato (Koller & Upadhyaya 2005). Moreover, NDVI has also been shown to provide a very good indicator of crop phenological development (Benedetti & Rossini 1993).

The NDVI index is calculated using ratios of vegetation spectral reflectance over incoming radiation in each spectral band. More specifically, NDVI can be formulated as:

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{VIS})/(\mathrm{NIR} + \mathrm{VIS})$$

where the difference between near-infrared reflectance (NIR) and visible reflectance (VIS) values are normalized by the total reflectance and vary between −1·0 and 1·0 (Eidenshink 1992). Negative and very low values corresponding to water and barren areas were excluded from the analysis by design. The NDVI data were extracted from the MOD13Q1 dataset (available online from: https://lpdaac.usgs.gov/lpdaac/content/view/full/6652), which regroups reflectance information collected by the MODerate-resolution Imaging Spectroradiometer (MODIS) instrument

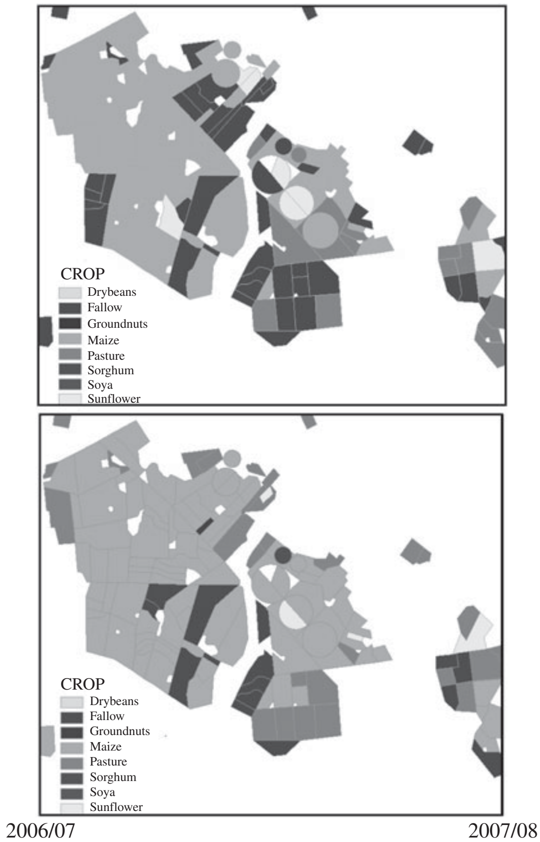


**Fig. 3.** Crop type changes across the summer seasons 2006/07 and 2007/08.

operating on NASA's Terra satellite (Huete *et al.* 2002). The NDVI estimates were available as 16-day composite indices at a resolution of 250 m. Area-weighted averages of the 16-day NDVI values for each crop field were calculated using the 'zonal statistics' tool in ArcGIS.

## Growing season

Crop growing seasons are characterized by the planting date and the phenology cycle, which determine the length of the season. In South Africa, planting dates are spread from October to December in order to reduce the vulnerability to erratic rainfalls (Ferreira *et al.* 2006). At the same time, however, phenology cycles also differ among crops. The TIMESAT program (Jönsson & Eklundh 2002, 2004) was used to determine crop- and field-specific growing seasons. The algorithm within the software is commonly used to extract seasonality information from satellite time-series data. Within this context, it allows one to approximate the start and end of growing

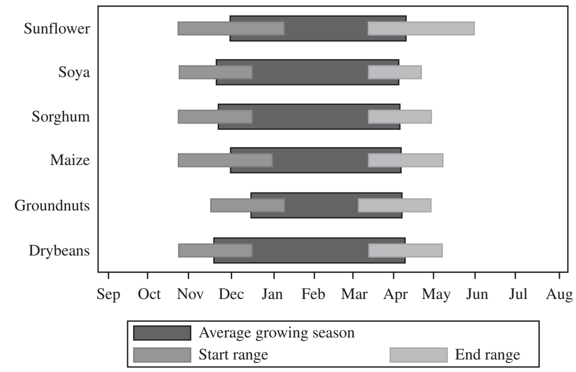


**Fig. 4.** Crop growing season estimated using Timesat. Notes: Average growing season bars represent the average start and end of growing season for each crop. Start and End range bars represent respectively start and end dates of growing seasons comprised between the 10th and the 90th percentile.

seasons based on distribution properties of NDVI. The repartition of growing season for each crop is presented in Fig. 4. This bar chart shows that within the area considered, groundnuts had the shortest growing season, whereas sunflower had the more widespread planting period.

Using these growing season estimates for each crop plot, it was also possible to determine the stage of the biomass development within the cropping season. In this exercise, the position of each 16-day period was determined relative to the length of the cropping season. For instance, the first 16-day period of a 160-day cropping season was assigned the value 0·1, and the last was assigned the value 1. This measure therefore accounts for the difference in growing season length of each crop plot.

## Weather

Daily rainfall data were extracted from the rainfall estimation algorithm RFE (version 2.0) dataset implemented by the National Oceanic and Atmospheric Administration (NOAA) – Climate Prediction Center (CPC). These data, which were generated from a combination of rain gauges and satellite observations, were available at FEWS NET Africa Data Portal (available online from: http://earlywarning.usgs.gov/fews/africa/index.php) at the 0·1° resolution (*c*. 10 km).

Reference evapotranspiration (ETo), which represents the evaporative demand of the air, was calculated using the Penman–Monteith equation following the FAO methodology (Allen *et al.* 1998). These data are also available from: http://earlywarning.usgs.

https://doi.org/10.1017/S0021859614000215 Published online by Cambridge University Press

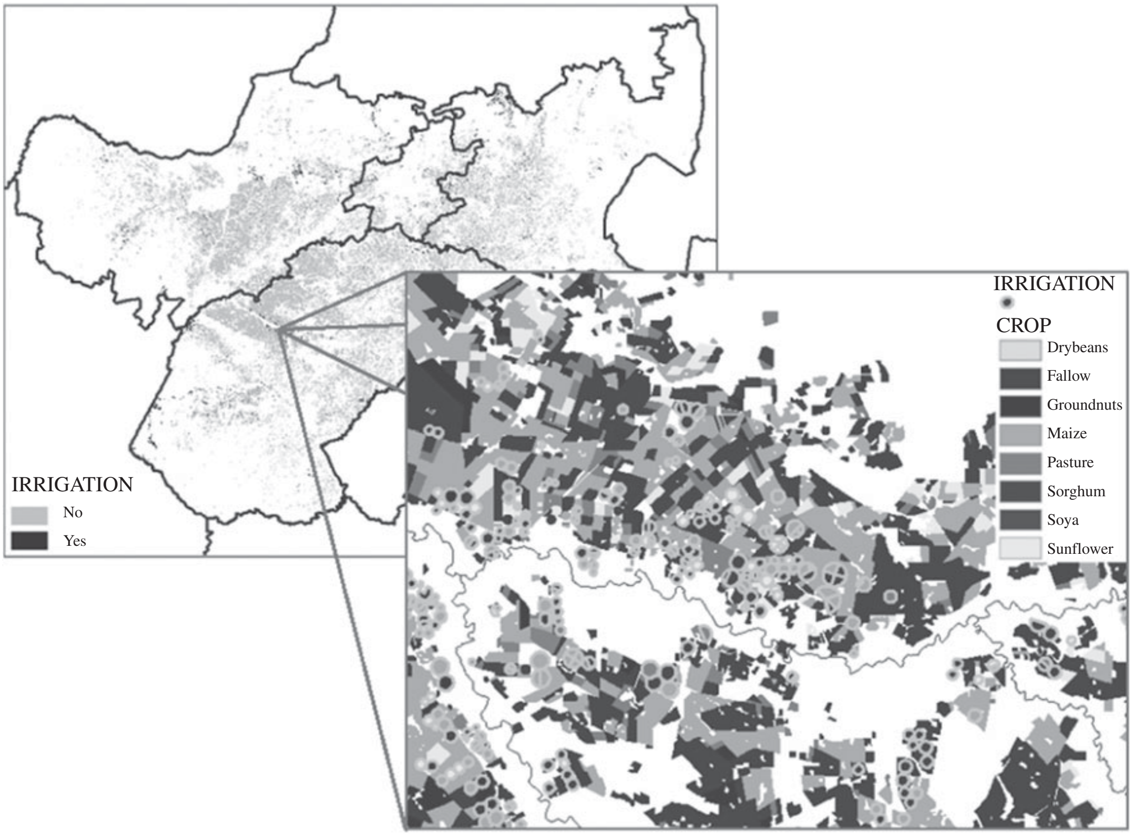


**Fig. 5.** Crop plots irrigation.

gov/fews/global/index.php: this source labels ETo as 'potential evaporation'. However, as noted in Allen *et al.* (1998), '*the use of other denominations such as potential ET is strongly discouraged due to ambiguities in their definitions.*' Daily ETo data at 1° resolution were calculated using a 6-hourly assimilation of conventional and satellite observational data of air temperature, atmospheric pressure, wind speed, relative humidity and solar radiation extracted from the NOAA's Global Data Assimilation System.

## Irrigation

The AGIS crop boundaries dataset also provides information regarding irrigation for each crop plot. As displayed on the left-hand side of Fig. 5, cropland in South Africa was mainly rain-fed. Moreover, the right-hand side of Fig. 5, which represents a detailed view of the crop plots, shows that irrigated crop plots were generally clustered around streams. It should be noted that the lines representing streams were not representative of the width of the actual streams, which may explain the 'gap' between the plots and the streams.

## Other water sources

To account for water sources other than local precipitation, the proximity of crop plots to streams and the daily flow of these were calculated. The proximity of each crop plot to a river or stream (perennial and non-perennial) was estimated using the river network of the African continent from the HYDRO1K dataset (USGS 2011). Stream flow was estimated using the Geospatial Stream Flow Model (GeoSFM) (Artan *et al.* 2008*), which* simulates the dynamics of runoff processes using spatial information on river basin and network coverage, land cover type, soil characteristics, and daily precipitation and evapotranspiration data. River basins in South Africa were delineated using the HYDRO1K dataset, which provides drainage basin boundaries data derived from river network and flow direction data. Soil characteristics (water-holding capacity, hydrologically active soil depth, texture, average saturated hydraulic conductivity) were extracted from the Digital Soil Map of the World (FAO 2011). The GeoSFM model produced daily stream flow in terms of cubic meters per second ($m^3/s$). Each crop plot was then assigned to a river basin and its corresponding river flow. For crop plots spreading over more than one river basin, the basin to which the plot belongs is determined by the largest area share located in the basin.

### Summary statistics

Summary statistics of time-invariant plot characteristics for each crop are provided in Table 1. Maize was the most prevalent crop with nearly 50 000 plots in the 2006/07 season. Groundnuts plots were on average the largest (37 ha). Irrigation summary statistics indicate that the irrigation rate for crops ranged from 4 to 7% (i.e. lowest for sorghum and highest for sunflower). Dry beans and groundnuts plots were located the furthest from perennial rivers, with an average distance of 13 km.

Crop plot time-series statistics are presented by crop type in Table 2 for all provinces over the 2006/07 season. Table 3 provides similar statistics for the Free State province over both the 2006/07 and 2007/08 seasons. The NDVI index, showing daily biomass productivity, was the largest for soya. Productivity was generally slightly larger during the 2007/08 season. One may also note that although the NDVI index is theoretically bounded, it never actually came close to either bound within this dataset.

The growing season length was the longest for dry beans and the shortest for groundnuts, with an average of 172 and 141 days, respectively. Weather variables show that sorghum and soya were grown in the wettest conditions (i.e. the highest rainfalls and the lowest ETo rates) during the 2006/07 season. Dry beans and soya were cultivated within the basin having the lowest stream flows in both seasons.

The skewness statistics show that the *Stream flow* and *Rain* were positively skewed and may need to be transformed.

## MODELLING FRAMEWORK AND RESULTS

### Regression specification

In the current analysis, a vector of various weather and irrigation factors was considered to estimate the determinants of crop growth. The base regression specification was formulated as follows:

$$NDVI_{pt} = \alpha + \delta X_{pt} + \varepsilon_{pt} \qquad (1)$$

where $p$ and $t$ are plot and time indicators, respectively, $X$ represents a vector of explaining variables, $\delta$ are the estimated coefficients representing the marginal effects and $\varepsilon$ a standard independent and identically distributed (i.i.d.) error term. The vector $X$ includes weather variables, represented by rainfall and ETo, and irrigation-related variables such as the distance to perennial and non-perennial rivers, stream flow and irrigation application. Many factors explaining crop growth, such as soil fertility, ozone concentration, crop variety, nutrient application and other management practices, were not included in the analysis. The omission of these explaining factors was an important issue for this, and many other, studies but was unavoidable due to data limitation. However, in order to control for region-specific fixed effects, a set of province factors was also included in all specifications. This regional distinction was deemed the most appropriate as it captures the geographic and political unobservable specificities. Also, it may reflect data collection discrepancies across regions as the data were collected at the province level. As a sensitivity analysis, regression (6) was re-estimated for the full sample using alternatives to the province factors. Climatic zone factors were considered first. The study area spreads over three major climate zones defined by FAO (1996): warm Sub-Tropics, cool Sub-Tropics and cold Sub-Tropics. A second alternative considered agro-ecological zones (AEZ) developed by Monfreda *et al.* (2009), who defines zones globally by overlaying six categories of growing length periods with climatic zones. In the study area, five AEZ categories are distinguished. However, neither of these alternative regional indicators changed the estimation results in any noticeable qualitative or quantitative manner. Detailed results are available from the authors upon request. It should be noted that, since the full set of these factors would be perfectly correlated if included together, the Free State factor was excluded so that any marginal effects on the others must be interpreted in terms of additional marginal productivity relative to this reference province. Additionally, as these fixed effects assume that the explanatory variables coefficients are similar across provinces, the interaction of the province factor variables with each determinant were also included to relax this condition.

The statistical analysis was performed using Stata 12.0 (StataCorp 2011). To investigate the possibility that crop growth is spatially correlated, a Moran's I test (Moran 1950) was conducted. Since this test indicated that the crop growth measure was spatially correlated, the non-parametric covariance matrix estimator

Table 1. *Crop plots summary statistics*

| Variable | | Name | Dry beans | Groundnuts | Maize | Sorghum | Soya | Sunflower |
|---|---|---|---|---|---|---|---|---|
| *Plots* (number) | 2006/07 | Number | 1288 | 1404 | 49 194 | 728 | 6946 | 10 511 |
| | 2007/08 | Number | 339 | 673 | 36 038 | 1902 | 2124 | 4092 |
| *Area* (Ha) | 2006/07 | Mean | 29·79 | 36·83 | 31·28 | 24·15 | 24·31 | 28·74 |
| | | Skewness | 4·18 | 2·14 | 3·54 | 2·78 | 3·74 | 2·85 |
| | | MinMax | [6·25;341·07] | [6·26;249·12] | [6·25;602·27] | [6·26;168·5] | [6·25;321·2] | [6·25;292·25] |
| | 2007/08 | Mean | 20·86 | 37·84 | 30·61 | 27·09 | 20·13 | 29·35 |
| | | Skewness | 1·79 | 0·8 | 1·56 | 1·43 | 1·92 | 1·5 |
| | | MinMax | [6·34;90·13] | [6·59;114·81] | [6·25;335·99] | [6·26;136·43] | [6·26;100·15] | [6·26;150·11] |
| *Irrigation* (irrigated = 1, non-irrigated = 0) | 2006/07 | Mean | 0·04 | 0·04 | 0·05 | 0·02 | 0·06 | 0·07 |
| | | MinMax | [0;1] | [0;1] | [0;1] | [0;1] | [0;1] | [0;1] |
| | 2007/08 | Mean | 0·01 | 0·03 | 0·06 | 0·01 | 0·01 | 0·03 |
| | | MinMax | [0;1] | [0;1] | [0;1] | [0;1] | [0;1] | [0;1] |
| *Distance perennial river* (m) | 2006/07 | Mean | 12 583·66 | 13 557·42 | 7485·29 | 3087·84 | 2865·3 | 7886·47 |
| | | Skewness | 1·86 | 1·41 | 2·77 | 2 | 2·88 | 2·53 |
| | | MinMax | [0;76 437] | [10;69 765] | [0;127 111] | [0;22 318] | [0;38 235] | [0;81 938] |
| | 2007/08 | Mean | 2785·3 | 11 636·11 | 7908·46 | 4553·82 | 2987·54 | 9565·4 |
| | | Skewness | 1·28 | 1·09 | 2·11 | 1·36 | 1·92 | 2·14 |
| | | MinMax | [0;11 248] | [333;57 190] | [0;58 216] | [0;23 393] | [0;21 986] | [0;58 014] |
| *Distance non-perennial river* (m) | 2006/07 | Mean | 9717·64 | 7061·61 | 8743·43 | 10 789·16 | 10 577·7 | 8172·57 |
| | | Skewness | 1·00 | 1·11 | 1·76 | 0·68 | 0·88 | 1·95 |
| | | MinMax | [0;39 962] | [0;31 258] | [0;66 961] | [0;37 867] | [0;38 982] | [0;67 080] |
| | 2007/08 | Mean | 6127·68 | 6809·64 | 7204·36 | 5614·39 | 10 396·86 | 6961·04 |
| | | Skewness | 2·28 | 1·43 | 2·8 | 1·22 | 1·06 | 2·06 |
| | | MinMax | [20;36 187] | [0;31 181] | [0;67 565] | [0;24 718] | [0;39 323] | [0;61 643] |

Table 2. *Summary statistics by crop type for all provinces for the 2006/07 season*

| Variable | Name | Dry beans | Groundnuts | Maize | Sorghum | Soya | Sunflower |
|---|---|---|---|---|---|---|---|
| *NDVI* (unitless) | Number | 13 107 | 11 587 | 440 342 | 6529 | 64 058 | 96 326 |
| | Mean | 0·48 | 0·42 | 0·48 | 0·55 | 0·57 | 0·47 |
| | Skewness | 0·52 | 0·66 | 0·43 | 0 | −0·03 | 0·48 |
| | MinMax | [0·2;0·88] | [0·17;0·82] | [0·07;0·93] | [0·21;0·87] | [0·2;0·91] | [0·02;0·9] |
| *Growing season length* (number of 16-day periods) | Number | 13 107 | 11 587 | 440 346 | 6529 | 64 058 | 96 333 |
| | Mean | 10·75 | 8·83 | 9·51 | 9·27 | 9·61 | 9·79 |
| | Skewness | 0·20 | 0·67 | 0·55 | 0·49 | 0·69 | 0·57 |
| | MinMax | [4;17] | [3;17] | [3;20] | [5;15] | [3;19] | [3;21] |
| *ETo* (mm/day) | Number | 13 107 | 11 587 | 440 346 | 6529 | 64 058 | 96 333 |
| | Mean | 4·58 | 4·72 | 4·56 | 4·45 | 4·45 | 4·56 |
| | Skewness | −0·25 | −0·42 | −0·47 | −0·75 | −0·83 | −0·41 |
| | MinMax | [1·85;7·54] | [1·68;7·54] | [1·15;8·39] | [1·69;6·79] | [1·18;6·74] | [1·18;8·39] |
| *Rain* (mm/16-day period) | Number | 13 107 | 11 587 | 440 346 | 6529 | 64 058 | 96 333 |
| | Mean | 30·89 | 19·99 | 28·87 | 35·81 | 36·84 | 26·46 |
| | Skewness | 1·72 | 1·14 | 1·64 | 1·61 | 1·34 | 1·72 |
| | MinMax | [0;168] | [0;110] | [0;212] | [0;199] | [0;199] | [0;199] |
| *Stream flow* ($m^3/s$) | Number | 13 107 | 11 587 | 440 346 | 6529 | 64 058 | 96 333 |
| | Mean | 66·71 | 577·22 | 219·86 | 105·16 | 86·99 | 214·86 |
| | Skewness | 11·28 | 1·64 | 3·63 | 3·02 | 3·23 | 3·74 |
| | MinMax | [0;4049·69] | [0;4436·7] | [0;4862·66] | [0;2898·99] | [0;3003·38] | [0;4862·66] |

proposed by Driscoll & Kraay (1998) was implemented to obtain robust s.e. for all estimations.

To account for the skewness of the *Stream flow* and *Rain* variables, the log of *Stream flow* and the square root of *Rain*, which were deemed the most appropriate at reducing the skewness, were also considered. However, the results indicated that the transformed variables did not improve the fit of the regression and the distribution of the residuals. Therefore, for ease of interpretation, the *Stream flow* and *Rain* variables were kept in levels.

## Regression results

Regressions results at the crop level are presented in Tables 4–6. These regressions were obtained through a selection process to remove non-significant secondary variables (i.e. non-linear terms and interaction terms) from the most general specification of Eqn (1). In order to make the results more legible, the dependent variable was scaled by a factor of 10 000 in all specifications. The crop samples were pooled across the four provinces. Differences between provinces were accounted for by including a set of province factors (the reference case is Free State) and their interaction terms with explanatory variables.

Rainfall and evapotranspiration were considered in the model to represent the effect of weather on crop biomass productivity. The *Rain* coefficients suggest that precipitation had a direct beneficial impact on sorghum and soya biomass productivity in all provinces. For dry beans, the beneficial effect was observed in the Mpumalanga province only. Rainfall had no significant effect on sunflower and groundnuts. Increases in evapotranspiration had a negative effect on dry beans, except in the Mpumalanga province, where it was beneficial to biomass productivity growth. For groundnuts, the effect of increased evapotranspiration was detrimental in all provinces, whereas it was beneficial for maize, sorghum and sunflower. The largest evapotranspiration effect was observed for sorghum.

In order to investigate whether there may be non-linearities in the relationship between these weather factors and cropland biomass productivity, their squared terms were included. Accordingly, there is strong evidence that evapotranspiration had a non-linear relationship with biomass productivity of maize, sorghum and sunflower. For these crops, the regressions indicate that evapotranspiration had a biomass productivity enhancing effect up to a certain point. The coefficients suggest that this threshold was about 7·4 mm/day for maize, 8·3 mm/day for sunflower and

Table 3. *Summary statistics by crop type for the Free State province for the 2006/07 and 2007/08 seasons*

| Variables | | Statistics | Drybeans | Groundnuts | Maize | Sorghum | Soya | Sunflower |
|---|---|---|---|---|---|---|---|---|
| *NDVI* (unitless) | 2006/07 | Number | 2422 | 8556 | 199 264 | 1889 | 21 499 | 45 501 |
| | | Mean | 0·5 | 0·43 | 0·45 | 0·46 | 0·52 | 0·44 |
| | | Skewness | 0·26 | 0·48 | 0·49 | 0·33 | 0·16 | 0·55 |
| | | MinMax | [0·26;0·82] | [0·19;0·82] | [0·07;0·89] | [0·21;0·81] | [0·23;0·87] | [0·02;0·87] |
| | 2007/08 | Number | 3406 | 5440 | 335 200 | 19 631 | 21 299 | 39 101 |
| | | Mean | 0·58 | 0·52 | 0·52 | 0·51 | 0·58 | 0·48 |
| | | Skewness | 0 | 0·01 | 0·06 | 0·24 | −0·03 | 0·24 |
| | | MinMax | [0·2;0·88] | [0·18;0·86] | [0·01;0·92] | [0·21;0·88] | [0·1;0·9] | [0·12;0·87] |
| *Growing season length* (number of 16-day periods) | 2006/07 | Number | 2422 | 8556 | 199 266 | 1889 | 21 499 | 45 508 |
| | | Mean | 10·40 | 8·39 | 9·51 | 9·71 | 9·81 | 9·82 |
| | | Skewness | −0·15 | 0·92 | 0·50 | −0·03 | 0·53 | 0·53 |
| | | MinMax | [5;15] | [3;17] | [3;20] | [5;14] | [4;19] | [3;21] |
| | 2007/08 | Number | 3406 | 5440 | 335 200 | 19 631 | 21 299 | 39 101 |
| | | Mean | 10·38 | 8·5 | 9·84 | 10·77 | 10·29 | 10·17 |
| | | Skewness | 1·12 | 1·42 | 0·70 | 0·33 | 0·35 | 0·42 |
| | | MinMax | [5;19] | [4;19] | [2;22] | [5;20] | [3;18] | [3;21] |
| *ETo* (mm/day) | 2006/07 | Number | 2422 | 8556 | 199 264 | 1889 | 21 499 | 45 501 |
| | | Mean | 4·59 | 4·81 | 4·75 | 4·53 | 4·59 | 4·76 |
| | | Skewness | −0·85 | −0·51 | −0·49 | −0·64 | −0·87 | −0·46 |
| | | MinMax | [2·2;6] | [1·68;7·42] | [1·15;8·39] | [1·69;6·79] | [1·18;6·24] | [1·18;8·39] |
| | 2007/08 | Number | 3406 | 5440 | 335 200 | 19 631 | 21 299 | 39 101 |
| | | Mean | 4·24 | 4·69 | 4·53 | 4·54 | 4·33 | 4·57 |
| | | Skewness | −0·62 | −0·26 | −0·28 | −0·32 | −0·5 | −0·29 |
| | | MinMax | [1·38;6·27] | [1·59;6·81] | [1·12;7·09] | [1·38;6·98] | [1·38;6·81] | [1·15;7·02] |
| *Rain* (mm/16-day period) | 2006/07 | Number | 2422 | 8556 | 199 264 | 1889 | 21 499 | 45 501 |
| | | Mean | 40·17 | 20·15 | 28·29 | 32·66 | 40·67 | 25·85 |
| | | Skewness | 1·00 | 1·09 | 1·45 | 1·54 | 1·19 | 1·48 |
| | | MinMax | [0;138] | [0;110] | [0;212] | [0;126] | [0;143] | [0;177] |
| | 2007/08 | Number | 3406 | 5440 | 335 200 | 19 631 | 21 299 | 39 101 |
| | | Mean | 47·67 | 30·91 | 35·53 | 36·75 | 46·25 | 32·89 |
| | | Skewness | 0·24 | 0·56 | 0·76 | 0·49 | 0·49 | 0·73 |
| | | MinMax | [0;144] | [0;123] | [0;179] | [0;151] | [0;168·71] | [0;179] |
| *Stream flow* ($m^3/s$) | 2006/07 | Number | 2422 | 7754 | 194 783 | 1889 | 21 478 | 44 830 |
| | | Mean | 136·81 | 689·68 | 354·64 | 292·32 | 172·77 | 361·74 |
| | | Skewness | 0·75 | 1·33 | 2·82 | 1·36 | 1·75 | 2·77 |
| | | MinMax | [05·32;487·6] | [0·47;3471·9] | [0·47;3776] | [08·94;2326·7] | [04·85;1350·8] | [0·47;3776] |
| | 2007/08 | Number | 3406 | 4737 | 328 433 | 19 621 | 21 299 | 38 370 |
| | | Mean | 220·66 | 1032·18 | 664·44 | 1105·43 | 365·91 | 487·26 |
| | | Skewness | 1·07 | 1·93 | 2·79 | 0·45 | 4·69 | 3·37 |
| | | MinMax | [12·17;870·5] | [4·57;7698·2] | [0·16;7934·3] | [10·72;5097·1] | [9·09;7698·2] | [0·16;7934·3] |

Table 4. *Regression results*

| | Dry beans | | | Groundnuts | |
|---|---|---|---|---|---|
| | | Province factor interaction terms | | | Province factor interaction terms |
| Variables | | *Mpumalanga* | *Northwest* | | *Northwest* |
| | | −4150 (*P*<0·01) (609·4) | −1214 (*P*<0·01) (229·7) | | −698·1 (*P*<0·01) (241·5) |
| *Rain* | −12·38 (8·680) | 68·35 (*P*<0·01) (20·08) | | −17·63 (10·78) | |
| *ETo* | −239·3 (*P*<0·01) (52·24) | 757·8 (*P*<0·01) (109·1) | | −223·4 (*P*<0·05) (108·8) | |
| *ETo*$^2$ | | | | | |
| *Irrigation* | 934·2 (*P*<0·01) (179·2) | | | 605·7 (*P*<0·01) (194·4) | 693·9 (*P*<0·05) (258·0) |
| *Distance perennial river* | −0·0216 (*P*<0·01) (0·00593) | 0·0368 (*P*<0·01) (0·00996) | 0·0212 (*P*<0·01) (0·00571) | −0·00342 (0·00222) | |
| *Distance non-perennial river* | −0·00190 (0·00268) | | | −0·00229 (0·00319) | |
| *Stream flow* | 0·480 (*P*<0·1) (0·273) | −1·663 (*P*<0·01) (0·454) | | −0·0196 (0·0698) | 0·440 (*P*<0·01) (0·0917) |
| *ETo × rain* | 2·764 (1·711) | −13·95 (*P*<0·01) (4·132) | | 4·112 (*P*<0·1) (2·400) | |
| *Rain × irrigation* | | | | | |
| *Stream flow × irrigation* | −3·275 (*P*<0·01) (0·701) | | 4·983 (*P*<0·01) (0·719) | 0·237 (0·144) | −0·459 (*P*<0·05) (0·227) |
| *Stream flow × distance perennial river* | $1{\cdot}00\times10^{-5}$ ($3{\cdot}07\times10^{-5}$) | | $-5{\cdot}49\times10^{-5}$ (*P*<0·1) ($3{\cdot}23\times10^{-5}$) | $-3{\cdot}09\times10^{-6}$ ($5{\cdot}02\times10^{-6}$) | $-2{\cdot}28\times10^{-5}$ (*P*<0·01) ($5{\cdot}68\times10^{-6}$) |
| *Stream flow × distance non-perennial river* | $-2{\cdot}87\times10^{-5}$ (*P*<0·1) ($1{\cdot}67\times10^{-5}$) | 0·000135 (*P*<0·01) ($4{\cdot}18\times10^{-5}$) | $7{\cdot}87\times10^{-5}$ (*P*<0·01) ($2{\cdot}51\times10^{-5}$) | $1{\cdot}06\times10^{-5}$ (*P*<0·01) ($1{\cdot}26\times10^{-6}$) | |
| *Growing season length* | −155·1 (*P*<0·01) (26·42) | | 138·2 (*P*<0·01) (24·27) | −176·5 (*P*<0·01) (27·06) | 122·3 (*P*<0·01) (29·97) |
| *Growing season stage* | 11 810 (*P*<0·01) (932·2) | 1364 (*P*<0·01) (347·5) | −8522 (*P*<0·01) (1073) | 9590 (*P*<0·01) (1594) | −5564 (*P*<0·01) (1697) |
| *Growing season stage*$^2$ | −11 107 (*P*<0·01) (697·3) | | 7389 (*P*<0·01) (940·7) | −8971 (*P*<0·01) (1574) | 4812 (*P*<0·01) (1617) |
| Constant | 5912 (*P*<0·01) (302·5) | | | 5366 (*P*<0·01) (727·3) | |
| Observations | 16 513 | | | 15 522 | |
| Number of groups | 1625 | | | 1845 | |
| $R^2$ | 0·680 | | | 0·351 | |
| $R^2$ adjusted | 0·680 | | | 0·350 | |

Notes: The dependent variable, NDVI, is scaled by a factor of 10 000; S.E. in parentheses; the reference case for the province factors is Free State.

https://doi.org/10.1017/S0021859614000215 Published online by Cambridge University Press

Table 5. *Regression results*

| Variables | Maize | Maize – Province factor interaction terms: *Gauteng* | Maize – Province factor interaction terms: *Mpumalanga* | Maize – Province factor interaction terms: *Northwest* | Sorghum | Sorghum – Province factor interaction terms: *Gauteng* | Sorghum – Province factor interaction terms: *Mpumalanga* |
|---|---|---|---|---|---|---|---|
| | | 495·5 ($P<0·05$) (213·6) | −392·6 ($P<0·1$) (203·5) | −755·9 ($P<0·01$) (161·1) | | 511·2 ($P<0·05$) (225·2) | 593·8 (468·1) |
| *Rain* | 3·326 (2·318) | | | | 34·54 ($P<0·1$) (18·08) | | |
| *ETo* | 584·2 ($P<0·05$) (267·0) | | | | 1248 ($P<0·05$) (521·5) | | |
| *ETo*$^2$ | −79·47 ($P<0·05$) (33·84) | | | | −120·7 ($P<0·1$) (62·67) | | |
| *Irrigation* | 1153 ($P<0·01$) (137·4) | −555·5 ($P<0·01$) (128·5) | −836·9 ($P<0·01$) (131·6) | 803·1 ($P<0·01$) (167·6) | 103·1 (130·0) | | |
| *Distance perennial river* | −0·0231 ($P<0·01$) (0·00240) | | 0·00841 ($P<0·05$) (0·00389) | 0·0110 ($P<0·01$) (0·00184) | −0·00274 (0·00779) | −0·0921 ($P<0·01$) (0·0214) | |
| *Distance non-perennial river* | 0·0143 ($P<0·01$) (0·00176) | −0·0220 ($P<0·01$) (0·00188) | −0·0144 ($P<0·01$) (0·00405) | −0·0161 ($P<0·01$) (0·00229) | 0·0376 ($P<0·01$) (0·00617) | | −0·0381 ($P<0·01$) (0·00622) |
| *Stream flow* | 0·108 ($P<0·05$) (0·0441) | −0·493 ($P<0·01$) (0·173) | −0·705 ($P<0·01$) (0·127) | | 0·111 (0·0868) | | |
| *ETo × rain* | | | | | −7·476 ($P<0·05$) (3·701) | | |
| *Rain × irrigation* | −6·810 ($P<0·01$) (2·221) | 8·232 ($P<0·01$) (2·347) | 8·688 ($P<0·01$) (2·170) | | −5·377 ($P<0·05$) (2·053) | | 9·926 ($P<0·01$) (1·773) |
| *Stream flow × irrigation* | −0·0487 (0·0420) | −1·125 ($P<0·01$) (0·248) | 0·393 ($P<0·05$) (0·185) | −0·196 ($P<0·01$) (0·0546) | 0·787 ($P<0·01$) (0·209) | | |
| *Stream flow × distance perennial river* | $-5·25\times10^{-7}$ ($1·28\times10^{-6}$) | $8·02\times10^{-5}$ ($P<0·05$) ($3·54\times10^{-5}$) | | $-9·29\times10^{-6}$ ($P<0·01$) ($1·42\times10^{-6}$) | $3·02\times10^{-6}$ ($4·63\times10^{-6}$) | | $-9·07\times10^{-5}$ ($P<0·05$) ($3·61\times10^{-5}$) |
| *Stream flow × distance non-perennial river* | $-5·76\times10^{-6}$ ($P<0·01$) ($9·27\times10^{-7}$) | | $8·23\times10^{-5}$ ($P<0·01$) ($2·29\times10^{-5}$) | $1·25\times10^{-5}$ ($P<0·01$) ($3·06\times10^{-6}$) | $-1·39\times10^{-5}$ ($P<0·01$) ($3·69\times10^{-6}$) | | $9·34\times10^{-5}$ ($P<0·01$) ($2·68\times10^{-5}$) |
| *Growing season length* | −72·25 ($P<0·01$) (19·79) | | | | −33·95 (26·61) | | −113·7 ($P<0·01$) (33·98) |
| *Growing season stage* | 8244 ($P<0·01$) (862·8) | | 6148 ($P<0·01$) (1106) | | 7980 ($P<0·01$) (731·2) | | 6377 ($P<0·01$) (1131) |
| *Growing season stage*$^2$ | −7643 ($P<0·01$) (876·5) | | −5534 ($P<0·01$) (1097) | | −7161 ($P<0·01$) (846·0) | | −5644 ($P<0·01$) (1174) |
| Constant | 3046 ($P<0·01$) (563·6) | | | | 551·3 (794·1) | | |
| Observations | 761 494 | | | | 26 150 | | |
| Number of groups | 72 900 | | | | 2608 | | |
| $R^2$ | 0·412 | | | | 0·476 | | |
| $R^2$ adjusted | 0·412 | | | | 0·476 | | |

Notes: The dependent variable, NDVI, is scaled by a factor of 10 000; S.E. in parentheses; the reference case for the province factors is Free State.

Table 6. *Regression results*

| | Soya | | | | Sunflower | | | |
|---|---|---|---|---|---|---|---|---|
| | | Province factor interaction terms | | | | Province factor interaction terms | | |
| Variables | | *Gauteng* | *Mpumalanga* | *Northwest* | | *Gauteng* | *Mpumalanga* | *Northwest* |
| | | 131·5 (156·8) | 798·1 ($P<0·01$) (220·0) | 1115 ($P<0·01$) (275·5) | | 471·6 ($P<0·05$) (202·7) | 561·6 (389·2) | 61·65 (180·9) |
| *Rain* | 16·91 ($P<0·1$) (9·255) | | | | 1·091 (6·904) | | | |
| *ETo* | 44·54 (142·3) | | | | 698·7 ($P<0·01$) (224·2) | | | |
| *ETo*$^2$ | | | | | −84·19 ($P<0·01$) (28·51) | | | |
| *Irrigation* | 418·8 ($P<0·01$) (105·5) | | | 1182 ($P<0·01$) (377·2) | 1094 ($P<0·01$) (133·0) | −514·6 ($P<0·05$) (227·2) | −1061 ($P<0·01$) (150·8) | 690·4 ($P<0·01$) (130·5) |
| *Distance perennial river* | −0·0123 ($P<0·01$) (0·00294) | | | −0·0551 ($P<0·01$) (0·00677) | −0·0177 ($P<0·01$) (0·00265) | −0·0512 ($P<0·01$) (0·0150) | 0·0185 ($P<0·05$) (0·00862) | −0·00982 ($P<0·01$) (0·00230) |
| *Distance non-perennial river* | 0·0159 ($P<0·01$) (0·00316) | −0·0205 ($P<0·01$) (0·00384) | −0·0239 ($P<0·01$) (0·00578) | −0·0650 ($P<0·01$) (0·00865) | 0·00654 ($P<0·05$) (0·00251) | −0·0297 ($P<0·01$) (0·00473) | −0·00991 ($P<0·01$) (0·00339) | −0·0357 ($P<0·01$) (0·00518) |
| *Stream flow* | 0·455 ($P<0·1$) (0·235) | | −1·294 ($P<0·01$) (0·228) | | 0·251 ($P<0·01$) (0·0689) | | −0·704 ($P<0·01$) (0·226) | −0·387 ($P<0·01$) (0·0401) |
| *ETo × rain* | −3·170 (1·982) | | | 1·320 ($P<0·05$) (0·654) | 0·146 (1·606) | 0·977 (0·766) | | 2·324 ($P<0·01$) (0·714) |
| *Rain × irrigation* | −0·282 (1·803) | | | −17·46 ($P<0·05$) (7·543) | −8·423 ($P<0·01$) (2·593) | 6·931 (4·281) | 12·90 ($P<0·01$) (2·681) | |
| *Stream flow × irrigation* | | | | | −0·144 ($P<0·01$) (0·0508) | −0·945 ($P<0·01$) (0·222) | | |
| *Stream flow × distance perennial river* | −2·43 × $10^{-5}$ ($P<0·05$) (1·07 × $10^{-5}$) | | | | −6·74 × $10^{-6}$ ($P<0·1$) (3·38 × $10^{-6}$) | | | |
| *Stream flow × distance non-perennial river* | −1·33 × $10^{-5}$ (9·37 × $10^{-6}$) | | 0·000134 ($P<0·01$) (2·55 × $10^{-5}$) | 4·29 × $10^{-5}$ ($P<0·01$) (9·59 × $10^{-6}$) | −6·77 × $10^{-6}$ ($P<0·01$) (1·43 × $10^{-6}$) | −2·90 × $10^{-5}$ ($P<0·01$) (5·51 × $10^{-6}$) | 0·000107 ($P<0·01$) (2·35 × $10^{-5}$) | 3·09 × $10^{-5}$ ($P<0·01$) (4·17 × $10^{-6}$) |
| *Growing season length* | −71·79 *($P<0·01$)* (25·24) | | | | −35·48 (22·39) | | −102·2 ($P<0·01$) (23·99) | |

Table 6. (*Cont.*)

| | Soya | | | | Sunflower | | | |
|---|---|---|---|---|---|---|---|---|
| | | Province factor interaction terms | | | | Province factor interaction terms | | |
| Variables | | *Gauteng* | *Mpumalanga* | *Northwest* | | *Gauteng* | *Mpumalanga* | *Northwest* |
| *Growing season stage* | 12 588 ($P<0·01$) (921·7) | | | −6392 ($P<0·01$) (1131) | 6419 ($P<0·01$) (762·7) | 3202 ($P<0·01$) (918·6) | 7976 ($P<0·01$) (774·8) | |
| *Growing season stage*$^2$ | −11 441 ($P<0·01$) (802·5) | | | 5930 ($P<0·01$) (901·9) | −5835 ($P<0·01$) (664·4) | −3016 ($P<0·01$) (847·8) | −7326 ($P<0·01$) (742·1) | |
| Constant | 3209 ($P<0·01$) (722·3) | | | | 2299 ($P<0·01$) (571·1) | | | |
| Observations | | | 85 216 | | | | 132 007 | |
| Number of groups | | | 8576 | | | | 13 930 | |
| $R^2$ | | | 0·484 | | | | 0·327 | |
| $R^2$ adjusted | | | 0·483 | | | | 0·326 | |

Notes: The dependent variable, NDVI, is scaled by a factor of 10 000; S.E. in parentheses; the reference case for the province factors is Free State.

10·3 mm/day for sorghum. In contrast, the impact of rainfall was better modelled by simply including it in its linear form.

To account for the effect of other forms of water availability on crop biomass productivity, the effect of the presence of irrigation on the crop plot was investigated. The results show that irrigation enhanced cropland biomass productivity substantially for all crops except sorghum, which was the less irrigated crop of the sample (<0·02 of the plots). The size of the coefficients suggests that irrigation was the most important for maize in the Northwest. More specifically, the estimated coefficient indicates that, in the Northwest, an irrigated maize plot had a 0·2 greater NDVI than a non-irrigated plot. In contrast, irrigation of a sunflower plot in Mpumalanga only increased biomass productivity by 0·003 NDVI.

Stream-related factors such as stream flow and the distance of each plot from perennial and ephemeral rivers were also considered. Stream flow had a positive direct effect on biomass productivity for most crops, although the quantitative effect was very small (e.g. one SD of river flow entailed an increase of less than 0·00005 NDVI). In a few cases (i.e. dry beans in Mpumalanga, maize in Gauteng and soya and sunflower in Mpumalanga), however, the increased stream flow appears to have had a detrimental effect on biomass productivity. For sorghum, the direct effect of stream flow was not significant.

Distance to perennial and non-perennial rivers were also significant determinants of the dependent variable. Generally, being closer to a perennial river increased the biomass productivity of a plot but being closer to an ephemeral river decreased it. For instance, for each kilometre closer to a perennial river, sunflower biomass productivity increased by up to 0·02 NDVI in Gauteng. In Mpumalanga, productivity was slightly larger for dry beans and sunflower plots being farther away from a stream. No effect was estimated for groundnuts. In contrast, being farther away from ephemeral streams increased biomass productivity for maize, sorghum and soya in the Free State and sorghum in Gauteng only.

Thus far, it was assumed that each of the explanatory factors had potentially only isolated effects on the biomass productivity of crops. This may arguably be a rather restrictive and unrealistic assumption. For instance, the importance of rainfall in providing water to crops depends on the degree of evapotranspiration. Similarly, the effect of river flow will be related to the distance of the plot from rivers. Moreover, the existence of irrigation may make the dependence on other time-varying water resources less important. To investigate these factors, a set of interaction terms was included. As can be seen for the regressions, this unearthed some interesting features from the data. When considering the interaction effect between rainfall and evapotranspiration, the results indicate that evapotranspiration not only had a direct effect on sorghum, but also reduced the productivity enhancing impact of precipitation by about 20%. Similar impacts were found for dry beans in Mpumalanga and soya in Northwest.

As gauged by the interaction term between the *Irrigation* and *Rain* variables, irrigation also appeared to reduce the reliance on precipitation for most crops except dry beans and groundnuts. Irrigation reduced the importance of rainfall as a water source for sunflower biomass production the most. When considering the combined effect of *Stream flow* with *Irrigation*, stream flow increased sorghum biomass productivity only if the plots were irrigated. For maize in Gauteng, in contrast, the negative effect of stream flow on biomass productivity was exacerbated by irrigation. For groundnuts in the Northwest and dry beans, stream flow had a smaller beneficial impact on biomass productivity in irrigated plots. Finally, the distance to perennial and ephemeral rivers can reduce the impact of stream flow for some crops. More specifically, the interaction terms of these factors with the stream flow proxy show that being farther away from a perennial river reduced the impact of stream flow on soya and sunflower in all provinces and for maize in Free State and Northwest, sorghum in Mpumalanga, and groundnuts and dry beans in Northwest.

All regressions included the growing season length, growing season stage and its square term. The growing season length unsurprisingly reduced the biomass productivity of plots within South Africa, as crops having a longer cropping season grow more slowly and produce less biomass daily. The growing season stage coefficients indicate that crop biomass increased as the growing season advanced and then decreased past an optimal point in the growing season. This inverted U shape is representative of the biomass evolution of crops along the phenological cycle. These results are consistent across all specifications in Tables 4–6.

The inspection of the residual for each crop-specific regression confirmed that variances are constant over the fitted range and that the residuals have a normal distribution.

## DISCUSSION

The current analysis considers the determinants of crop growth at the crop plot level in four grain producing provinces of South Africa. It is unique in that it is based on a detailed high-frequency (every 16 days) dataset of satellite derived productivity measures for four crop types at the plot level covering the main agricultural areas of the country: Northwest, Mpumalanga, Free State and Gauteng. The regression results show that water availability plays an important role in cropland biomass productivity in South Africa. Weather, as represented by evapotranspiration and precipitation, has a significant and well-known impact on crop biomass productivity. The results also demonstrate the importance of other sources of water for crop biomass productivity growth by considering irrigation, stream flow and proximity to perennial and ephemeral rivers. However, the effect of these factors depends on the crop type and province considered, with the best representation obtained for dry beans.

It is noteworthy that the $R^2$ in the estimations suggests that the regression analyses explain biomass productivity growth better for some crops than others. For instance, *c*. 68% of dry beans productivity is explained by the regressions. When considering sunflower and groundnuts, however, the $R^2$ is less than 0·35. This may not be surprising given that the vector of crop biomass determinants is missing a number of potentially important factors as discussed above. Moreover, NDVI is inevitably only a proxy of actual crop yield and thus entails some measurement error. Finally, the growing season of a crop is determined by a decision rule, similarly inducing some measurement errors.

One should practise some caution when interpreting the determinants of crop growth in a strictly causal way given that they are derived from an observational study. As already indicated by the less than perfect explanatory power, there are likely to be many other determining factors driving crop growth that have not been able to be included, but that could be correlated with the ones that are. For example, crop location is implicitly taken as given, but farmers with greater access to fertilizers or machinery may be those that choose, or can afford to, locate in more irrigated areas. Similarly, the choice of crop for a particular location may be determined by such potentially confounding factors. Thus, it is perhaps best to view the estimated relationships, particularly for the non-climatic factors, as correlations rather than causal determinants.

With these caveats in mind, the present results are arguably most useful in terms of anticipating the impact of weather shocks on the growth of a given crop. More specifically, one could use these quantitative estimates to forecast the growth of existing crops in specific regions in South Africa once one observes the weather in a given growing season. Alternatively, one may use the results to compare short-term crop growth under different climate scenarios. Once more crop-level data – such as the one employed in the present study – become available over a longer time period, one could also then extend this analysis to model crop choice.

Finally, one should note that the availability of crop plot level information played an essential role in this analysis. These plot information combined with biomass productivity data would be also extremely useful in devising an identification scheme of crop type based on NDVI temporal evolution. More generally, and related to this, it is encouraging that satellite imagery is becoming increasingly more sophisticated, offering products at finer spatial scale and higher frequency. This will allow greater precision in terms of crop identification and growth patterns.

## REFERENCES


Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. (1998). *Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper no. 56. Rome: FAO.

Artan, G. A., Asante, K., Smith, J., Pervez, S., Entenmann, D., Verdin, J. P. & Rowland, J. (2008). *Users' Manual for the Geospatial Stream Flow Model (GeoSFM)*. U.S. Geological Survey Open-File Report 2007–1440. Reston, Virginia, USA: U.S. Geological Survey.

Azam-Ali, S. N. & Squire, G. R. (2002). *Principles of Tropical Agronomy*. Wallingford, UK: CABI Publishing.

Benedetti, R. & Rossini, P. (1993). On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sensing of Environment* **45**, 311–326.

Blanc, E. & Strobl, E. (2013). The impact of climate change on cropland productivity: evidence from satellite based products at the river basin scale in Africa. *Climatic Change* **117**, 873–890.

Das, D. K., Mishra, K. K. & Kalra, N. (1993). Assessing growth and yield of wheat using remotely-sensed canopy temperature and spectral indices. *International Journal of Remote Sensing* **14**, 3081–3092.

Doraiswamy, P. C. & Cook, P. W. (1995). Spring wheat yield assessment using NOAA AVHRR data. *Canadian Journal of Remote Sensing* **21**, 43–51.

Driscoll, J. C. & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* **80**, 549–560.

Eidenshink, J. C. (1992). The 1990 conterminous US AVHRR data set. *Photogrammetric Engineering and Remote Sensing* **58**, 809–813.

FAO (1996). *Major Climates of Africa*. Rome: FAO GeoNetwork. Available from: http://www.fao.org/geonetwork/srv/en/metadata.show?id=6&currTab=simple (accessed 19 February 2014).

FAO (2011). *Digital Soil Map of the World*. Rome: FAO. Available from: http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116 (accessed 12 February 2014).

Ferreira, S. L., Newby, T. & du Preez, E. (2006). Use of remote sensing in support of crop area estimates in South Africa. In *SPRS Archives XXXVI-8/W48 Workshop Proceedings: Remote Sensing Support to Crop Yield Forecast and Area Estimates* (Eds B. Baruth, A. Royer & G. Genovese), pp. 51–52. Ispra, Italy: Joint Research Centre.

Groten, S. M. E. (1993). NDVI-crop monitoring and early yield assessment of Burkina Faso. *International Journal of Remote Sensing* **14**, 1495–1515.

Gupta, R. K., Prasad, S., Rao, G. H. & Nadham, T. S. V. (1993). District level wheat yield estimation using NOAA/AVHRR NDVI temporal profile. *Advances in Space Research* **13**, 253–256.

Hayes, M. J. & Decker, W. L. (1996). Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *International Journal of Remote Sensing* **17**, 3189–3200.

Hochheim, K. P. & Barber, D. G. (1998). Spring wheat yield estimation for Western Canada using NOAA NDVI data. *Canadian Journal of Remote Sensing* **24**, 17–27.

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.

Jönsson, P. & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing* **40**, 1824–1832.

Jönsson, P. & Eklundh, L. (2004). TIMESAT – a program for analysing time-series of satellite sensor data. *Computers and Geosciences* **30**, 833–845.

Karnieli, A., Agam, N., Pinker, R. T., Anderson, M., Imhoff, M. L., Gutman, G. G., Panov, N. & Goldberg, A. (2010). Use of NDVI and land surface temperature for drought assessment: merits and limitations. *Journal of Climate* **23**, 618–633.

Koller, M. & Upadhyaya, S. K. (2005). Prediction of processing tomato yield using a crop growth model and remotely sensed aerial images. *Transactions of the ASAE* **48**, 2335–2341.

Labus, M. P., Nielsen, G. A., Lawrence, R. L., Engel, R. & Long, D. S. (2002). Wheat yield estimates using multitemporal NDVI satellite imagery. *International Journal of Remote Sensing* **23**, 4169–4180.

Malo, A. R. & Nicholson, S. E. (1990). A study of rainfall and vegetation dynamics in the African Sahel using normalized difference vegetation index. *Journal of Arid Environments* **19**, 1–24.

Milesi, C., Samanta, A., Hashimoto, H., Krishna Kumar, K., Ganguly, S., Thenkabail, P. S., Srivastava, A. N., Nemani, R. R. & Myneni, R. B. (2010). Decadal variations in NDVI and food production in India. *Remote Sensing* **2**, 758–776.

Monfreda, C., Ramankutty, N. & Hertel, T. W. (2009). Global agricultural land use data for climate change analysis. In *Economic Analysis of Land Use in Global Climate Change Policy* (Eds T. W. Hertel, S. Rose & R. S. J. Tol), pp. 33–48. Abingdon, Oxon, UK: Routledge.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23.

van Niekerk, L. (2012). *Pocket Guide to South Africa 2011/2012*. Pretoria, South Africa: GCIS.

Nuarsa, I. W., Nishio, F. & Hongo, C. (2011). Relationship between rice spectral and rice yield using MODIS data. *Journal of Agricultural Science (Canada)* **3**, 80–88.

Potdar, M. B. (1993). Sorghum yield modelling based on crop growth parameters determined from visible and near-IR channel NOAA AVHRR data. *International Journal of Remote Sensing* **14**, 895–905.

Prasad, A. K., Chai, L., Singh, R. P. & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation* **8**, 26–33.

Quarmby, N. A., Milnes, M., Hindle, T. L. & Silleos, N. (1993). The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *International Journal of Remote Sensing* **14**, 199–210.

Schultz, P. A. & Halpert, M. S. (1993). Global correlation of temperature, NDVI and precipitation. *Advances in Space Research* **13**, 277–280.

StataCorp (2011). *Stata Statistical Software: Release 12*. College Station, TX, USA: StataCorp LP.

USGS (2011). *HYDRO1k Dataset*. Washington, DC: U.S. Geological Survey. Available from: https://lta.cr.usgs.gov/HYDRO1K (accessed 12 February 2014).

Weissteiner, C. J. & Kühbauch, W. (2005). Regional yield forecasts of malting barley (*Hordeum vulgare* L.) by NOAA-AVHRR remote sensing data and ancillary data. *Journal of Agronomy and Crop Science* **191**, 308–320.

Yang, W., Yang, L. & Merchant, J. W. (1997). An assessment of AVHRR/NDVI-ecoclimatological relations in Nebraska, U.S.A. *International Journal of Remote Sensing* **18**, 2161–2180.