Exploiting satellite-based rainfall for weather index insurance: the challenges of spatial and temporal aggregation

Conference or Workshop Item

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access


It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://dx.doi.org/10.3390/ecrs-1-f002
To link to this article DOI: http://dx.doi.org/10.3390/ecrs-1-f002

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.
CentAUR
Central Archive at the University of Reading
Reading's research outputs online
Exploiting Satellite-Based Rainfall for Weather Index Insurance: The Challenges of Spatial and Temporal Aggregation

Emily Black ¹,²*, Elena Tarnavsky ¹,³†, Ross Maidment ¹†, Helen Greatrex ⁴†, Agrotosh Mookerjee ⁵†, Tristan Quaife ¹,³† and Jennifer Price ¹,³†

¹ Department of Meteorology, University of Reading; ²Climate Division of the National Centre for Atmospheric Science (NCAS-Climate); ³National Centre for Earth Observation; ⁴International Research Institute for Climate and Society; ⁵MicroEnsure, Cheltenham, UK

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed: e.c.l.black@reading.ac.uk

Published: 25 June 2015

Abstract: Lack of access to insurance exacerbates the impact of climate variability on smallholder farmers in Africa. Unlike traditional insurance, which compensates proven agricultural losses, weather index insurance (WII) pays out in the event that a weather index is breached. In principle, WII could be provided to farmers throughout Africa. There are two data-related hurdles to this. First, most farmers do not live close enough to a rain gauge with sufficiently long record of observations. Second, mismatches between weather indices and yield may expose farmers to uncompensated losses, and insurers to unfair payouts – a phenomenon known as basis risk. In essence, basis risk results from complexities in the progression from meteorological drought (rainfall deficit) to agricultural drought (low soil moisture). In this study, we use a land-surface model to describe the transition from meteorological to agricultural drought. We demonstrate that spatial and temporal aggregation of rainfall results in a clearer link with soil moisture, and hence a reduction in basis risk. We then use an advanced statistical method to show how optimal aggregation of satellite-based rainfall estimates can reduce basis risk, enabling remotely sensed data to be utilized robustly for WII.

Keywords: weather index insurance; Africa; Satellite-based rainfall estimates; TAMSAT
1. Introduction

In Africa, climate variability can limit development and deepen poverty [1][2]. For example, banks are unlikely to lend to farmers if the crop failure caused by drought will result in defaults, as many farmers are likely to default in the same year [3]. In the last 10 years, a new type of insurance has been designed to mitigate climate-related risk. Weather index-based insurance (WII) is linked to a weather-based index such as rainfall, rather than to a physical outcome, such as a low yield [4]. WII has the potential to be cheap to administer and transparent to operate. In principle, WII provides a means of insuring smallholder farmers throughout Africa against perils, including drought.

There are many operational and technical hurdles to the widespread uptake of drought based WII and two major data-related hurdles. Currently, most schemes are centered around individual rain gauges. The African gauge network is, however, sparse, and the rainfall climate, spatially variable [5]. Schemes that rely on gauge data thus cannot provide insurance to the majority of farmers in Africa.

Secondly, WII schemes are based on an index, rather than a proven loss. Mismatch between insured weather based indices and agricultural losses can lead either to unfair payouts, or to uncompensated losses. Such mismatches are termed basis risk. In essence, basis risk stems from complexities in the progression from low rainfall (meteorological drought), to deficit in root zone soil moisture (agricultural drought). Low rainfall is not necessarily a precursor to soil moisture deficit, and conversely, soil moisture deficits may occur, even when rainfall is near normal [6][7].

These issues are closely linked. Basis risk is exacerbated when the insured weather index does not accurately represent local meteorological conditions. Even a few kilometers from a station, the weather may be significantly different to that observed. An alternative approach is to use remotely sensed data, such as satellite-based rainfall estimates (SRFEs), which can provide local information on meteorological conditions in near real time [8].

Although they both provide estimates of rainfall, SRFEs and gauge measurements are fundamentally different. The degree to which they agree varies spatially and temporally - depending on the meteorological regime, satellite rainfall estimation methodology and density of the gauge network[9]. If the assumptions underlying the satellite rainfall estimation methodology do not account for the dominant rain forming processes, SRFEs provide a poor approximation of rainfall variability.
In most cases, the skill of SRFEs is improved by averaging in time and space [10]–[12]. Agricultural drought is, furthermore, linked more closely to cumulative than to instantaneous rainfall, because soil moisture is affected by the accumulation of rainfall over weeks or months. In some regions, moreover, lateral as well as vertical flow of water contributes significantly to soil moisture. Total soil moisture and hence the risk of drought, is thus related to spatially, as well as to temporally distributed rainfall.

The previous discussion suggests that the aggregation of SRFE-based WII indices is likely to mitigate basis risk. The optimal scale of aggregation is a function both of rainfall variability and of the interaction between meteorological and land-surface conditions. As such, the scale is likely to vary from one region and season to another. The principles and methodologies for identifying the optimal scale for aggregation, however, are widely applicable in Africa and beyond.

In this paper, we focus on cotton in Zambia – an important cash crop, which is already insured using SRFEs. We first use a land-surface model to describe the progression from meteorological to agricultural drought in the region. We then consider how aggregation in space and time affects the capacity of SRFEs to represent local rainfall. The final section of the paper draws these threads of analysis into a discussion of how the aggregation of SRFEs relates to basis risk.

2. Data and models

This study combines land-surface model integrations with analyses of satellite imagery and survey of agricultural loss data. The land surface model chosen was the Joint UK Land Environment Simulator (JULES). JULES is used to investigate the development of drought and the scales of land-atmosphere interactions. Statistical analysis of satellite imagery is used to quantify algorithmic uncertainty, and thus to quantify the effect of aggregation in space and time on SRFE skill. The following sections describe JULES, the SRFE estimation methodology (TAMSAT) and the agricultural loss data.

2.1. Joint UK Land Environment Simulator (JULES)

JULES is a process-based land-surface model. When coupled to one of the Hadley Centre atmosphere models, it comprises the land-surface scheme of the Hadley Centre climate models. Full descriptions of JULES are available at [13], [14]. The following summarizes the features of greatest relevance to this study.

JULES divides the land-surface into nine surface types: broadleaf trees, needle leaf trees, C3 (temperate) grass, C4 (tropical) grass, shrubs, urban, inland water, bare soil, and ice. The land-
surface types are tiled to represent sub-grid heterogeneity\cite{15}. Surface fluxes of moisture and heat are calculated for each tile, and the state of the grid box is then represented by the aggregation of the tile fluxes. JULES can be run either at a point or over a grid (distributed JULES). It is important to note that the formulation of distributed JULES used for this study does not include lateral transfer of heat or moisture.

JULES includes a multi-layer representation of soil. In this study, the default four layers were used (depths 0.1, 0.25, 0.65, and 2 metres). Each soil layer is described by a set of hydraulic and thermal properties (table 3 in \cite{13}). In practice, these quantities are derived by applying pedotransfer functions to maps of soil texture \cite{16}. In distributed JULES, the soil hydraulic and thermal properties are allowed to vary at each grid point. Although it is possible to vary the soil properties with depth, for this study, they were assumed to be constant.

The water available to plants is described by beta, a dimensionless measure of water stress. Beta is related to the soil moisture by the following formula:

\[
\beta = \begin{cases} 
1 & \text{for } \theta \geq \theta_c \\
\frac{\theta - \theta_w}{\theta_c - \theta_w} & \text{for } \theta_w < \theta < \theta_c \\
0 & \text{for } \theta \leq \theta_w
\end{cases}
\] (1)

where \(\theta\) is the soil moisture in the root zone and \(\theta_c\) and \(\theta_w\) (the soil moisture at the critical and wilting point respectively) are amongst the aforementioned hydraulic soil properties. Beta is used in this study as a proxy for plant water stress and hence for agricultural drought.

The results shown in Figures 2-4 derive from integrations of JULES carried out for 1983-2012 at 0.5° horizontal resolution over the domain shown in Figure 2. JULES was forced with three-hourly gridded time series of radiation, precipitation, temperature, humidity, wind speed, and surface pressure, extracted from the WFDEI (WATCH Forcing Data based on ERA-Interim) forcing dataset. In this dataset, all variables apart from precipitation are extracted from the ERA-Interim reanalysis. The precipitation data are based on ERA-Interim, bias corrected using the CRU dataset. For a full description of the WFDEI forcing data, see \cite{17}.

2.2. TAMSAT and TAMSAT rainfall ensembles

The TAMSAT (Tropical Applications of Meteorology using SATellite data and ground-based observations) rainfall ensemble algorithm is used to represent the inherent uncertainty in satellite estimates of rainfall by generating ensembles of equally likely rainfall scenarios over Africa.
2.2.1 The TAMSAT Method

The TAMSAT algorithm relies upon imagery from Meteosat thermal Infra-red (TIR) imagery to determine the Cold Cloud Duration (CCD) parameter which is defined as the duration each pixel is below a predetermined threshold temperature and is used as a proxy for rainfall[8], [11], [18], [19], [20]–[22]. Such an approach is used in place of instantaneous temperature measurements often used in rainfall estimation [23][24] as TIR-only based rainfall estimates are most skillful when aggregated due to the indirect nature of the relationship between cloud top temperature and rainfall. This indirect relationship is valid for convective rainfall events (i.e. the longer a cloud top is below the threshold temperature the greater one would expect the rainfall amount to be) and does not hold for warm rain processes where the cloud top temperature is less representative of the rainfall on the ground. As such, the TAMSAT algorithm is suited for much of tropical Africa, which is dominated by convective rainfall. Given the heterogeneous nature of the African rainfall climate, CCD fields are regionally calibrated assuming a linear relationship between CCD and rainfall for each calendar month using historic gauge measurements, ensuring the resulting estimates reflect the expected local conditions. The TAMSAT method has shown high levels of skill across Africa, often exhibiting similar, or greater, performance than other, more complex satellite algorithms [10], [25]–[30].

2.2.2 TAMSAT rainfall ensembles

During the calibration stage, the relationship between CCD and rainfall can be characterised probabilistically to determine probability distributions of both rainfall occurrence and amount. Using this information, it is possible to generate an ensemble of rainfall fields by randomly sampling from the probability distributions. However, carrying out this process for each pixel independently would result in unrealistic, spatially uncorrelated fields. To overcome this, spatially independent ‘seed’ pixels are chosen from the observed CCD field and the influence of the surrounding pixels on each seed pixel’s probabilities is calculated using a geostatistical process known as sequential simulation (SS). SS is performed in two stages: (1) to delineate regions of rain and no rain and (2) to assign a rainfall amount to the rainy pixels. This process samples out of each pixel’s occurrence and rainfall amount probability distributions respectively in a manner designed to preserve spatial correlations, and is conducted until all pixels in the domain are considered. The entire process is repeated many times, producing a set of spatially coherent, equally likely rainfall scenarios that are consistent with the observed CCD field and the climatological CCD-rainfall relationship. A more detailed description of the method is given in
TAMSAT rainfall ensembles are thus an expression of the uncertainty in the TAMSAT algorithm.

2.3. Loss data
Data on agricultural losses were gathered at the 38 locations shown on the maps included in Figure 2. The loss percentage (actual/expected yield) is calculated using a combination of different sources of data, both quantitative and qualitative. Multiple sources of data are used because there is no single reliable source of loss data for all sites. The reliability/credibility of the source of loss data, moreover, varies by location.

The sources of loss data are:

1. Information on farmer experience, collected via semi-structured interviews. Inevitably these focus on the ‘worst years’, which farmers’ remember, and there is a bias towards recent years,

2. Feedback from field staff of distribution channels, including agricultural extension agencies and suppliers of agricultural inputs, and

3. A simple yield stress model, relating rainfall deviation and multiplicative crop yield factors to calculated yield deviations for cotton for different historical years (based on FAO crop yield stress factors), supplemented by occurrence of droughts reported in the scientific literature.

Different credibility weights are assigned to the each of these sources. For example the early part of the record (before ~1995) depends more strongly on the yield stress model, while the later data incorporates more information from farmer interviews.

3. Results and discussion
Cotton in Zambia is almost exclusively rain-fed. It is to be expected, therefore, that when rain is low, losses are high. This is supported by Figure 1, which shows a convincing association between average losses over the 38 study locations and peak rainy season (November-March rainfall) averaged over the domain shown, in Figure 2 (correlation coefficient 0.65, which is significant at the 95% level).
There are, however, discrepancies - with heavy losses experienced during some years of near normal rainfall. These may be explained by the wider social context of agriculture in Zambia. As well as rainfall variability, cotton production is affected by a multitude of economic factors. It is notable that around 2002, the Zambian Kwacha started to appreciate against the US Dollar, placing the export sector under increasing pressure. In response to the failing export market, in 2006, one of the largest cotton ginning company, NWK-Agri-Services announced a 30% reduction in the price it would pay, resulting in a 40% drop in area planted during the 2007 season [34]. It is likely that price pressures, rather than rainfall deficit, explain the fall in production during the 2006 and 2007 seasons. This wider context may also, in part explain the volatility in cotton production since 2002.

Zambia's economic troubles exacerbate the population's vulnerability to climate variability, particularly drought. Robust WII could mitigate some of these risks – especially if it can be implemented at a large scale. As was described in the introduction, the fair implementation of drought WII requires a clear connection between rainfall and root zone soil moisture i.e. a consistent progression from meteorological to agricultural drought. The progression from rainfall deficit to agricultural drought is illustrated by Figure 2, which shows the correlations between rainfall, upper level soil moisture and beta. Correlation coefficients are calculated, for inter-annual variability in January, April, July, and October at every modeled grid point. The correlation between rainfall and upper level soil moisture (top row of figures) relates to the

**Figure 1:** Time series of total November-March rainfall (black) and a time series of mean losses at the 38 locations shown in Figure 2 (grey)
infiltration of rainfall into the top layer of soil. The correlation between upper level soil moisture and beta (middle row of figures) relates to percolation of water from the top layer of soil to the root zone. The correlation between rainfall and beta (bottom row) illustrates the ensuing link between rainfall and root zone soil moisture.

During the rainy season, the correlation between rainfall and upper level soil moisture is strong (>0.8 in most of the country). Out of the rainy season, in July, when there is little rainfall, the correlation is near zero. There is a slight weakening in the strength of the correlation as the rainy season progresses – probably because of the increasing proportion of time that the upper soil layer is saturated. The second row of plots in Figure 2 shows that the correlation between upper level soil moisture and beta varies both spatially and temporally – reflecting heterogeneity in climate and soil type. In the north of the country, the correlation is strongest towards the end of the rainy season (in April), while in the south it is strongest in January. The maps of the correlation between beta and rainfall (bottom row) confirm that the strength of the link between rainfall and agricultural drought is varies. This complexity highlights the need to adjust WII indices spatially, and indeed, to consider carefully which regions can be insured fairly using indices based on cumulative rainfall.
Figure 2 Correlation between rainfall and upper level soil moisture (top row); upper level soil moisture and beta (middle row); rainfall and beta (bottom row) for (from left to right) January, April, July and October. Cyan shading denotes negative correlations of <-0.1. Black circles are the localities for which loss data are available; blue circles represent the three locations shown in Figures 3 and 4 (from west to east: Chikanta, Makafu, Kalichero).

The discussion above focused on the broad links between rainfall and drought. The next sections look in more detail at three localities for which we have loss data:

- Chikanta: 26° E; 16.4°S
- Makafu: 28.8°E; 14.3°S
- Kalichero: 32.4°E; 13.5°S

The locations of these stations are shown as blue circles on Figure 2.
**Figure 3** Precipitation versus beta for (from top to bottom) Chikanta, Makafu, and Kalichero for May–September (grey circles), October–December (yellow circles), and January–April (green circles).

Figure 3 plots beta against mean rainfall - distinguishing between the different parts of the season. The plots show how beta increases as the season progresses, reaching a maximum in January/February. For the most western locality, beta never reaches 1, and as a result, the strong correlation between beta and rainfall persists through the season. This is consistent with Figure 2, which shows that correlations remain reasonably strong through the whole season in this part of Zambia. For the other localities, beta reaches its maximum during the second half of the season, and as a result, the correlation between beta and rainfall drops. Again, this is consistent with Figure 2.

The relationship between rainfall and beta is explored further in Figure 4, which shows plots of beta versus cumulative precipitation during the peak of the rainy season (November–March). Broadly speaking, cumulative rainfall is proportional to beta, until beta reaches 1. The clarity of the link, however, varies from one locality to another. In particular, at Chikanta, there is considerably more noise than at the other locations. In one of the drought years (1994/95), beta remains near zero, which implies that the wilting point is never reached. It is notable that during the droughts, not only is cumulative rainfall lower than average at all three localities, but also that beta is unusually low, given the rainfall that has occurred. This has implications for index design, in that indices based on cumulative rainfall may underestimate the severity of losses during the most severe droughts.

Notwithstanding subtleties in the link between cumulative rainfall and beta, the previous section has demonstrated a clear link between meteorological and agricultural drought in southern Zambia. This supports the use of WII for mitigating drought-related risk. However, the gauge network is sparse, and any large scale WII scheme is likely to depend on remotely sensed data. There are a number of Africa rainfall datasets available at the required resolution and for a sufficient time period [35][19]. The rest of this study will focus on one such dataset, TARCAT, which is the historical product, based on the TAMSAT method (see Section 2 for further details). We focus on TARCAT because the underlying method has been shown to have good skill for Zambia [26], and the dataset is already used in WII schemes for this region.
Figure 4 Cumulative precipitation versus beta for (from top to bottom) Chikanta, Makafu, and Kalichero. The brown dots and lines highlight four years during which there was widespread drought and severe agricultural losses in Zambia: 1986, 1991, 1994, and 2001.

Although the TAMSAT method captures Zambia rainfall variability reasonably well, a degree of algorithmic uncertainty is inevitable. For example, the TAMSAT daily estimation method cannot distinguish rainfall over ~40mm. The ensembles method described in Section 2, provides a means of objectively quantifying algorithmic uncertainty. In this study, we have used the
ensembles to investigate the expected improvements in skill when SRFEs are aggregated in time and space[10]. Figure 5 shows how the spread of ensembles, and hence the uncertainty in rainfall estimates, reduces as they are aggregated in time. It can be seen that there is significant variability in the magnitude of the spread from one locality to another, but that the shape of the curves is very similar, with limited benefit from aggregating beyond five days. The example shown in Figure 5 is for January, but there were similar results for the whole rainy season (not shown).

**Figure 5** Time length of aggregation plotted against the mean ensemble range for each of the localities shown in Figure 2. The pale grey lines relate to individual time series for the, and the bold line is the mean.

Figure 6 illustrates an analogous effect when aggregating in space with clear increases in skill evident up to ~200 km. Beyond 400 km, there is little discernible improvement. When aggregating rainfall, however, the improvement in skill must be balanced against less accurate representation of local conditions. In other words, a 400 x 400 km aerial mean rainfall estimate may be an accurate representation of rainfall over a 400 x 400 km region, but a poor proxy for rainfall at any given point within it. This is illustrated by Figure 7, which shows the correlation between regionally averaged and local rainfall, for regions ranging from 100 x 100 km to 600 x 600 km. It can be seen that there is a rapid drop off in correlation, so that beyond 300 km, values are bordering on statistically insignificant. This means that rainfall averaged over a 300 x 300 km region is not representative of local rainfall.
The above analysis of aggregation in time and space provides useful information for index design. Figure 5 shows that the ensemble range is high when daily SRFEs are not aggregated in time. This means that at a daily time scale, TIR-based SRFEs are highly uncertain. Indices based on statistics of individual days may, therefore, not be robust. An example of such an index would be $n$, where $n$ is the number of days in a given period, for which rainfall is less than $x$ mm. Conversely, indices based on cumulative rainfall are more robust. Spatially aggregating rainfall indices also increases their skill, but it is necessary also to consider the effect of aggregation on the representation of local conditions.

Comparison between Figure 6 and Figure 7 can provide guidance as to the optimal spatial scale for aggregation. In Zambia, it has been found that indices based on dekadal rainfall averaged over 10 days closely match user experience of losses. This is consistent with Figure 6, which shows that aggregating over 10 days markedly improves skill. Figure 7, however, indicates rainfall averaged over a wide region would have higher skill, while still representing local conditions. It is important to recognize, moreover, that while this methodology is widely applicable, the optimal scale of aggregation is likely to vary from one region/season to another.
Figure 7 Correlation between rainfall at a point at 25°E 15°S and mean rainfall in a box surrounding the point (y-axis), plotted as a function of the length of the side of the box (x-axis).

The first part of the discussion considered the progression from meteorological to agricultural drought, and the second part considered aggregation of rainfall estimates. These two factors are interlinked, and both must be considered when designing WII indices.

As well as being more reliably represented by SRFEs, cumulative precipitation is a better determinant of soil moisture deficit than instantaneous precipitation as confirmed by comparison between Figures 3 and 4. The link between soil moisture and rainfall also strengthens as these quantities are aggregated in space. This is confirmed by Figure 8, which shows how the percentage in variance in beta is explained by cumulative precipitation increases when the data are aggregated. It should be noted that although the effect is fairly small, it is coherent across the region (not shown). Because JULES does not account for lateral transfers of water (see Section 2), this effect can only be explained by the averaging out of extremes of rainfall and heterogeneities in the land surface. In reality, like any land surface hydrological variable, soil moisture is affected by aerially distributed rainfall. Figure 8 may therefore underestimate the effect of spatial aggregation on the link between rainfall and soil moisture.

These findings support the notion that an index based on spatially and temporally averaged SRFEs can be a good proxy for agricultural drought as long as the aggregation is carried out at an appropriate scale. The optimal scale depends on spatial and temporal variability in the rainfall climate, the skill of SRFEs, and the properties of the land-surface.
In summary, we have shown that on a national scale, cotton production losses in Zambia are strongly linked to variability in rainfall. A process analysis using a land surface model showed some association between cumulative rainfall, upper level soil moisture and root zone soil moisture. This is reflected in significant correlation between rainfall and root zone soil moisture opening up the possibility of implementing WII. The lack of gauge data, however, means that WII must be based on remotely sensed data, such as SRFEs.

Further analyses confirmed that when rainfall is aggregated in time and space, the skill of SRFEs improves. The aggregation also irons out heterogeneity in both the rainfall climate and the land surface, resulting in higher correlations between precipitation and root zone soil moisture. When designing WII schemes the improvements resulting from aggregation must, however, be balanced against the need to capture local conditions.

4. Conclusions

1. In Zambia, cotton production losses are associated with rainfall variability.

2. There is a significant relationship between meteorological and agricultural drought on all spatial scales and throughout Southern, Central and Eastern Zambia.

3. The high skill of SRFEs in the region open up the possibility of expanding drought WII to a national level, provided that indices are carefully chosen to be skillful, representative of local conditions and strongly linked to variability in soil moisture.
Acknowledgments

We acknowledge the support of NERC and the TSB through SatWIN-Toolbox (TS/L007762/1), and the support of NERC for SatWIN-Scale (NE/M008797/1). We are grateful to the National Centre for Atmospheric Science for supporting work carried out by Emily Black and Ross Maidment, and the National Centre for Earth Observation for supporting Elena Tarnavsky, Jennifer Price, and Tristan Quaife. We are grateful to NWK Agri-Services in Zambia for facilitating the provision of data on historical agricultural losses from farmers and field staff.

Author Contributions

This paper is the outcome of two collaborative projects (SatWIN-Toolbox and SatWIN-Scale) on the use of remotely sensed data in weather index insurance. Although EB wrote the main body of this paper, all of the authors made substantial intellectual contributions to the project, and hence to the findings reported here.

Conflicts of Interest

None

References and Notes


© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).