TECHNICAL NEWSLETTER
#39 - June 2017

GUIDE TO AGRICULTURE INSURANCE PART III

Risk modelling aspects

INTRODUCTION

While the last issues of the Guide to Agriculture Insurance[1] focused on the use of new technologies for the development of insurance products, this technical newsletter will highlight some of the challenges we face in assessing production risk in agriculture.

Agricultural producers buy insurance to protect themselves against two main loss causes. Firstly, there is the risk of production loss caused by natural disasters (such as hail, drought, and flooding), disease, pests or even human error. Secondly, producers are faced with the risk of a loss of revenue due to declines in the prices of agricultural commodities.

Agriculture insurance has some particular characteristics to its coverage which sets it apart from most other P&C lines of business and which contribute to the complexity of risk modelling in agriculture:

- In crop insurance it is often difficult to identify a single event (or date) as a driver for an observed yield shortage. It is even possible that only a specific series of different events causes a loss. This is also the reason why most reinsurance contracts in crop insurance are defined on an annual aggregate rather than on an event loss basis.
- The timing of a meteorological event (like frost, drought or excessive moisture) has vastly different effects depending on the stage of plant growth at which it occurs. An event can lead to either a huge loss or have little to no impact as the plant vulnerability changes over the growing period.
- The value of the insured object (e.g. the crop) will usually increase during the coverage period, as the plant matures.
- With the high socio-economic importance of the agriculture industry in many countries, governments are involved (through subsidies for instance) in shaping the insurance framework which can add more uncertainty into the modelling.

These above risk specificities all influence the risk modelling in agriculture which the remainder of this newsletter will focus on. Despite there being many different sublines of business covered in agriculture, this newsletter will only cover risk modelling aspects of the largest agriculture insurance class, crop insurance. It will elaborate on three key elements of agriculture risk modelling:

- **Data standardisation:** There is still no standardised exposure data format agreed by the agriculture insurance market. One reason for this may be the existence of only very few agriculture vendor models which had driven the data standardisation in property insurance. Below we will outline how SCOR has tackled this data standardisation challenge with the development of its own exposure collection and assessment tool, SEED.
- **Use of meteorological data:** Crop insurance results are often linked to weather conditions. If one can find a strongly correlated, risk-specific weather-based index, this additional information can be used to expand our view of risk from pure historic losses to an additional risk measure, historic weather data. The second section below will introduce SCOR's weather data analysis tool, STRATUS.
- **Detrending of historical crop yield data:** Trends are often present in insurance losses even after accounting for inflation, and they are a topic of discussion across lines of business. In agriculture trends in losses typically emanate from trends in the underlying yield data, which in turn can be the result of various factors, for example improving farming practices. In this newsletter we discuss various detrending approaches and their implications in insurance pricing.

SCOR believes that the consideration of each of these aspects will help to further improve the assessment of agricultural production risk. However, note that the above listing is not exhaustive for all critical and important risk modelling steps required in agriculture risk modelling.

1. For a copy visit www.scor.com/en/scor-global-pac/pac-publications.html

SCOR The Art & Science of Risk | P&C

SLICE AND DICE THE DATA CUBE: THE DATA STANDARDISATION CHALLENGE

WHERE DO WE STAND?

Data used for insurance-related assessment of agriculture risks belong to various domains and are provided by insurance companies, national governments, international organisations, and vendor companies. The data collection techniques range from direct in-field sampling to remote sensing. Even data from the same provider group can use different resolution, classification schemes, languages, spelling and also differ in sources of errors. As a consequence the quality and structure of those data is highly heterogeneous. With technological advancement, for instance in computing and sensor technique, datasets available for insurance-related studies show a clear tendency towards a wider range of parameters, improving spatial and temporal resolution and coverage, more frequent updates of parameter values, and an increasing number of data sources.

WHAT ARE WE LOOKING FOR?

With this amount of data available, we can leverage considerably our risk assessment capability. Because in the agriculture insurance industry a comparable uniform data standard for information exchange does not yet exist, we first have to create a common reference basis by standardising and consolidating data; this means that we have to glue the available data strings together where common dimensions such as time, location or crop type exist.

WHAT ARE THE CHALLENGES?

- **Maintaining consistency of the data model:** Data need a common reference basis in order to be analysed. The integration of new data bears the risk of breaking the consistency of a conceptual data model (or the "common format"). Thus the data model has to be flexible enough to extend the number of dimensions as business grows, but also to adapt to a changing business environment.
- **Computational performance:** Along with increasing data volume and complexity computational struggles may occur.
- **Keeping the data structure simple:** Large and complex data structures require simplification in order to make them usable and understandable. For instance, data have to be classified, indexed and visualised following pre-defined classification schemes (e.g. taxonomic classification).
- **Interpreting values correctly:** Alphanumeric values collected globally from various sources are usually not consistent, show spelling mistakes and are provided in different languages.
- **Varying data quality:** Data coming from different providers often vary in data quality which can add an additional layer of uncertainty.
- **Preserving information content:** Data harmonization aims to retain as much information from the data as possible, while at the same time trying to keep data consistent and hence giving up certain details. Thus, the benefit of standardisation must be greater than the loss of a certain percentage of information, knowing that the degree of loss in data quality is not easily measurable.

HOW CAN IT BE DONE? THE GENERIC DATA STANDARDISATION PROCESS

A data model is basically a framework for holding data. It determines the data structure and standardises how the data relate to each other. Creating a data model as a n-dimensional cube (a hypercube) with hierarchical dimensions helps us to handle multiple hierarchies of data (e.g. the different administrative zone levels at which data are collected), to extend the number of dimensions and data volume as business grows, but also to adapt to a changing business environment.

A sound data model is a prerequisite for the data standardisation process itself. The standardisation process, at the end of which data are incorporated into the data model, can be subdivided into four steps and is visualised in Figure 1.

In the first step, data have to be interpreted. The correct interpretation of data is a critical stage which needs technical knowledge as data commonly lack metadata information.

Once data are interpreted, variable names of the raw data can be mapped against variable names specified in the data

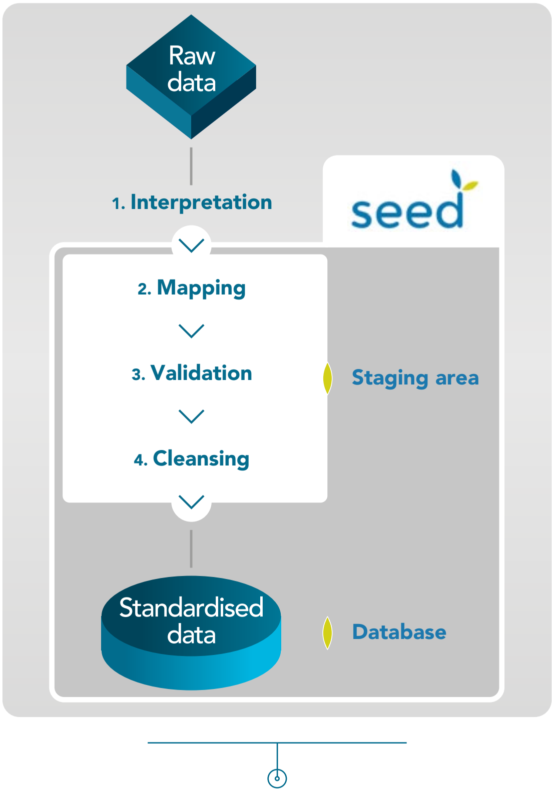


FIGURE 1: GENERIC DATA STANDARDISATION PROCESS

Source: SCOR visualisation

model in the second step. This step is not necessary, but takes out complexity in the validation process. The third step then verifies if the data comply with the requirements of the data model. Data identified as problematic in the validation step need treatment and therefore have to run through a fourth step, where data are cleansed before they can be ultimately consolidated.

WHAT HAS BEEN DONE ALREADY?

Various initiatives exist for statistics on agriculture farming practices particularly in the context of precision farming. For instance, the AG Gateway has established the common technical and statistical standard SPADE[2] for the exchange of data. Various initiatives on information management standards and related tools and services have also been set out by CIARD[3]. These include the Research Data Alliance (RDA) or the Agricultural Information Management Standards (AIMS). However, despite these efforts, a global standard has not yet emerged. Hurdles include "the resources available for data sharing, the research culture, the high price of data, the technical challenges and also issues arising from commercial and personal interest" (Rawlings 2017).

CONCLUSION

Many interesting data are available for agriculture insurance-related risk assessment studies and the ball of twine, or data, is getting bigger and denser every day. As outlined in this newsletter, it is difficult to discover any pattern and relationships in this ball because the link between those data threads is often hidden. **SCOR has therefore undertaken an effort to address major challenges in data standardisation** as shown in the box below. With the implementation of the first version of the in-house software tool SEED in 2016 we have already leveraged our risk assessment capabilities. Standardised data enable us to create exposure-specific weather indices with the in-house software STRATUS (see the box "How does SCOR use meteorological data?"), as well as allow us to improve our spatial-temporal analysis capabilities.

REFERENCES:

• **Rawlings, C.** – source: *Rothamsted Research*, March 14th 2017 – http://aims.fao.org/activity/blog/rda-igad-pre-meeting-agenda-released

2. SPADE (Standardised Precision Ag Data Exchange) is a collaboration among agricultural suppliers of hardware, software, inputs, services, implements and vehicles for improved data exchange and interoperability. It targets farm operations of seeding, tillage, fertilizing, spraying and harvest to maximise the value of precision agriculture through seamless and transparent data exchange, for more information see http://www.aggateway.org.
3. A global movement dedicated to open agricultural knowledge, see http://www.ciard.info.

THE SCOR APPROACH ON DATA STANDARDISATION

To support efficient and sound data analysis and reporting SCOR has taken the initiative of developing the in-house software SEED, the Specialty Exposure Evaluation Database. The strength of the tool is its focus on a powerful data standardisation process, whereby the guiding principle is preserving information content as far as possible.

What are the main functionalities of SEED?

Technically spoken, the software comprises a relational database and an extended set of database management functionalities to support the data standardisation process. What makes the standardisation process in SEED valuable and unique are the following characteristics:

- it recognises mappings of variables to accelerate data processing,
- it handles erroneous and uncertain data,
- it preserves the original data for auditing purposes, analysis and reporting, and finally,
- it deals with multiple languages enabling reporting in the language of our clients.

Recognizing mapping of variables

In a first step, SEED attempts to understand the raw data. This involves both the mapping of variable names of the raw data to those variables known to the system, and obtaining metadata information, such as its currency, area units, or administrative zone levels. User input and expertise is required at this stage, as even for consistent and error-free data, there often remains the challenge of interpreting the data correctly. For instance, the variable name "Premium" can refer to different financial perspectives. Thus, the user has to decide if the raw data field has to be linked to SEED's "Gross Premium" or "Net Premium" field and what the financial year basis is. To "remember" this decision for repetitive data loading (e.g. for when data having the same format is updated on a yearly basis), SEED makes use of the concept of mapping schemes which allow users to save and re-use a set of mapping rules, as shown in Figure 2.

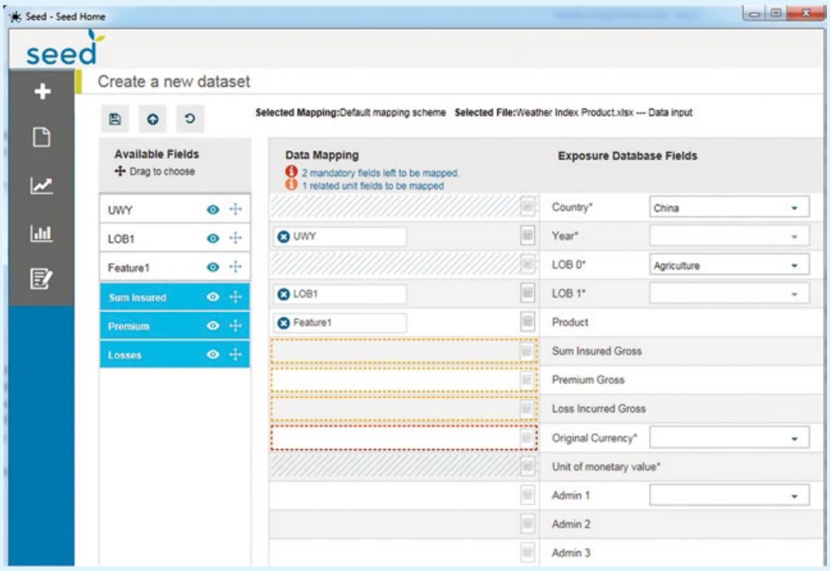


FIGURE 2: THE MAPPING STEP OF SEED

The fields of the original data are displayed in the leftmost column, and the SEED database fields in the rightmost column. The middle column shows the fields which have already been mapped (for instance, the field "UWY" from the original data has been mapped to SEED's "Year" field.

Handling of erroneous and uncertain data

In a next step, the uploaded raw data need to be validated. The validation and cleansing of numeric values is relatively simple: all numeric values must have a unit or measure value, such as “degrees Fahrenheit”, attached to it. Rules restrict the range of numeric data of a specific type to specific values. For instance, premium values coded with the currency “USD” cannot be negative. In addition, rules validate the values of one field against the values from another field, and so e.g. the value in the field “Premium Gross” is not allowed to be greater than the value in the field “Sum Insured”, as illustrated in Figure 3. Erroneous numeric data are then cleansed either by removing the record from the dataset or by allowing the user to apply “on-the-fly” correction supported by SEED’s field calculation function. For example, a Premium Rate can be specified by the user to calculate either the Sum Insured or the Premium value for selected records.

For alphanumeric data the validation and cleansing is more complex. For this, SEED utilises reference datasets, which define sets of permissible values. Reference datasets exist for many domains, such as worldwide administrative zones or the taxonomic classification of cultivars. The taxonomic classifications follow international scientific standards while related classifications are compiled internally by SCOR and focus on insurance application (such as classification of commercial use of cultivars, e.g. “fodder or oil crops”) and further risk modelling aspects.

SEED then attempts to establish a link between the input and the reference values using some analytical routines. Establishing such a link is challenging, and so this process is supported by a user-managed thesaurus. It enables the user to map a reference value against alternative names such as synonyms (corn versus maize), different spellings, repetitive spelling errors, or common abbreviations. By adding a value to the thesaurus the process follows in some the principles of supervised learning.

UWY All	Product	Sum Insured (Gross)	Premium (gross)
1999	Wheat	1,895	17,244
1999	Pumpkin + Squash	121,480	2,451
1999	Soibeans	358,421,340	20,500,603

FIGURE 3: SEED DATA INPUT VALIDATION SHOWING RAW DATA CONTAINING THREE ERRORS
1- The crop type “Pumpkin + Squash” cannot be linked to an existing crop type value in the reference dataset nor to the user-managed thesaurus; 2- the UWY 1999 Premium for “Wheat” is greater than the Sum Insured; 3- there is a spelling error identified for “Soibeans”.

Preserving the original data

Knowing that classification schemes are ever-changing (for instance new crop hybrids are developed or administrative zones are merged or split), the original alphanumeric value is always kept in the database. Beside ensuring auditability, this allows performing studies based on different classification schemes as well as running analyses on original values in case too many values had to be coded as “Unknown”.

Data input and reporting in multiple languages

Input data are collected globally, meaning that data are provided in various languages, which complicates the standardisation process significantly. Therefore, in SEED it is possible to set the proofing language before any advanced data validation and cleaning routines are applied. Currently, this functionality is available in six languages such as Spanish and Chinese besides the reference language English.

NO RAIN, NO GRAIN: THE USE OF METEOROLOGICAL DATA

INSIGHTS TO BE GAINED FROM CONSIDERING METEOROLOGICAL DATA

Most agriculture insurance contracts worldwide have a direct or indirect dependence on weather conditions, and especially on weather extremes such as droughts or cold spells. The use of meteorological, or weather, data can aid the modelling of agricultural risk in a number of ways. As a first step, it allows us to gain a qualitative estimation of a hazard, giving insight into its geographic occurrence and extent. In addition, it allows for an estimation of return periods of specific types of weather events, can aid with the pricing of parametric products[4], and may be a starting point for obtaining Probable Maximum Loss (PML) estimates for an insurance program. The use of meteorological data can be especially valuable for insurance programs with a short, unavailable, unreliable or even non-existent loss history.

On a fundamental level, the amounts and timing of meteorological "inputs" a plant receives (such as sunlight, water or temperature) directly affects its development. For example, Kaufmann & Snell (1997) developed a biophysical model of corn, and show that stressing the plant, by for example too high or too low temperatures, has a negative effect on the yield achieved by the plant at the end of the growing season. Indeed, the impact on crop yields due to deviations from optimal meteorological conditions forms one of the core principles of crop yield models such the Decision Support System for Agrotechnology Transfer (DSSAT) (Hoogenboom et al. 2015, Jones et al. 2003). Impacts of weather stress on plants on a country level may be very large – for example, the 2012 drought in the US reduced corn and sorghum yields by over 25% (Rippey 2015).

TYPES OF METEOROLOGICAL DATA AVAILABLE AND CHALLENGES WITH USING IT

There are several types of meteorological data available, which can be roughly divided into two categories – observational data and reanalysis data. Observational data ▸▸

4. Parametric products provide insurance pay-outs when certain meteorological conditions are met.

HOW DOES SCOR USE METEOROLOGICAL DATA?

At SCOR, we have developed a tool, STRATUS, which allows the linking of agricultural insurance data with weather data. The tool offers a range of global meteorological datasets and variables, and it uses them to calculate weather severity indices, such as for measuring drought. STRATUS takes into consideration insurance losses and geographical distribution of risk to develop exposure-specific weather indices to support risk assessment. It allows for analyses for a range of crops and perils worldwide.

As an example, STRATUS can be used for analysing early spring freeze events affecting flowering fruit trees in Switzerland. By using the cultivated area of apple, cherry and pear trees in Switzerland in 2015 (Caloz & Boehlen 2015), and the Consecutive Days Temperature Index[5], we select critical parameters (see Table 1) and perform the analysis. The results of the analysis are specific to this particular geographical location of the fruit orchards.

From the results of this analysis, shown in Figure 4, we can find an estimate of the return periods of the events occurring in historical years, including those years for which insurance loss data[6] is not available. By looking at the correlation between insurance loss data and the weather index, we can also get an indication how well the index could reflect losses in years for which no insurance data are available. Additionally, for each historical year, the value of the index is displayed on a map.

Of course, many factors play a role in agricultural production. However, the results of the analyses done in STRATUS may be a first and important step for analysing insurance programmes.

5. This index is designed to reflect the amplitude and persistence of heat and cold conditions (Klein Tank et al. 2009). The idea behind the index is that a crop will likely suffer considerable damage if it is exposed to hot or cold conditions for a long enough period without interruption.
6. Note that for data confidentiality reasons, and since a frost insurance cover for fruit trees does not exist in Switzerland, we use only fictional insurance loss data information. All other data are real.

Parameter	Value	Notes
Time period	7 April to 21 May, 2015	This approximately corresponds to the blooming dates of apple, cherry and pear trees in Switzerland (MeteoSchweiz 2016).
Dataset	CFSR	The Climate Forecast System Reanalysis (CFSR) product is created by the US National Center for Environmental Prediction (NCEP), see Saha et al. (2010) and Saha et al. (2014) for more details.
Variable	Minimum temperature	This is the lowest temperature recorded in a 24 hour period.
Minimum temperature threshold	-3°C	This corresponds to the critical temperatures for the first bloom to post bloom stages of plant development for apple, cherry and pear trees (Proebsting & Mills 1978), accounting for the smoothing of temperatures extremes in gridded datasets (Haylock et al. 2008).
Topography	1000m	Sometimes it is necessary to exclude regions above a certain altitude which otherwise could distort the results due to the lower temperature. Apple, cherry and pear trees, for instance, do not grow at such high altitudes.

TABLE 1 CRITICAL PARAMETERS USED FOR ASSESSING EARLY SPRING FREEZE EVENTS ON FLOWERING FRUIT TREES IN SWITZERLAND.

Source: SCOR

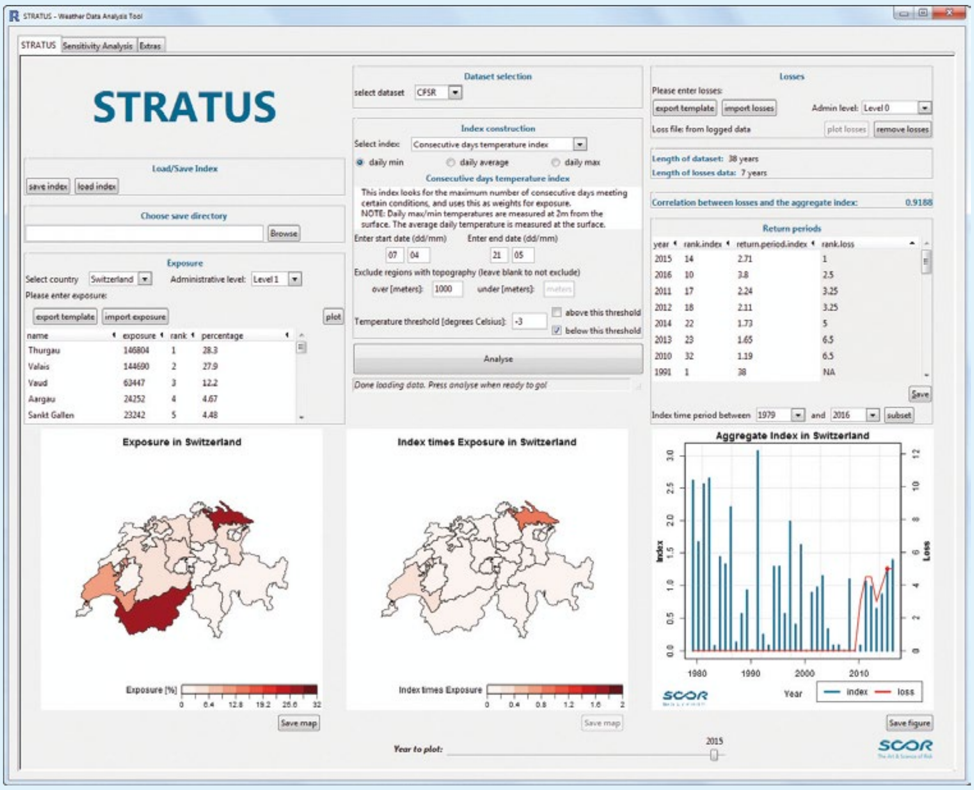


FIGURE 4: SCREENSHOT OF STRATUS

The results of an analysis of spring freeze events on flowering fruit trees in Switzerland. See Table 1 for an explanation of parameters chosen.

▸▸ may be either *in situ* ("on site") or from remote sensing devices. *In situ* measurements are direct measurements of meteorological variables (such as temperature, precipitation, and wind speed) by surface weather stations, radiosondes, ships, aircrafts, balloons, or buoys (see e.g. Inness & Dorling 2013 and references therein for more details). They are most useful in situations where precise, localised measurements are needed, and spatial variations are not so important. If on the other hand geospatial information is required, or measurements are needed from locations for which *in situ* observations are unavailable, data from remote sensing devices, such as radar, wind profilers, or satellites may be used. These are available for larger regions, although they are often less accurate than *in situ* measurements. The second category of data, reanalysis data, is a combination of observational data with the output from a model of the Earth's atmosphere (see e.g. Lahoz & Schneider 2014). Reanalysis data are useful where there is scarce or incomplete observational data.

The availability and accessibility of meteorological data presents exciting opportunities for modelling agricultural risk, but it is not without its challenges. For analysis, long, continuous weather time series which are homogeneous over time are necessary, and this in practice can be hard to ensure. For instance, a change of the instrument used for performing measurement may introduce systematic biases in time series of meteorological data. The measurement instrument may also break, causing a gap in the time series, or - due to human error - measurements may not be performed or an erroneous value recorded. Practical issues to address when dealing with meteorological data include choosing the correct data source for a given type of analysis, reconciling data from multiple sources and formats, as well as processing and identifying patterns in large volumes of data.

HOW TO LINK IT WITH AGRICULTURAL RISK?

It is necessary to develop a methodology to link meteorological data with agricultural risk. There are various approaches which can be taken, but here we will focus on one – weather indices. The idea behind a weather index is to isolate the driving meteorological factors over a pre-defined time period in such a way that they reflect the occurrence and severity of damage. In agriculture, this usually translates to measuring the extent to which the weather deviates from normal conditions, meaning that it is necessary to define both what is normal for a specific crop in question, and what the crucial time period is.

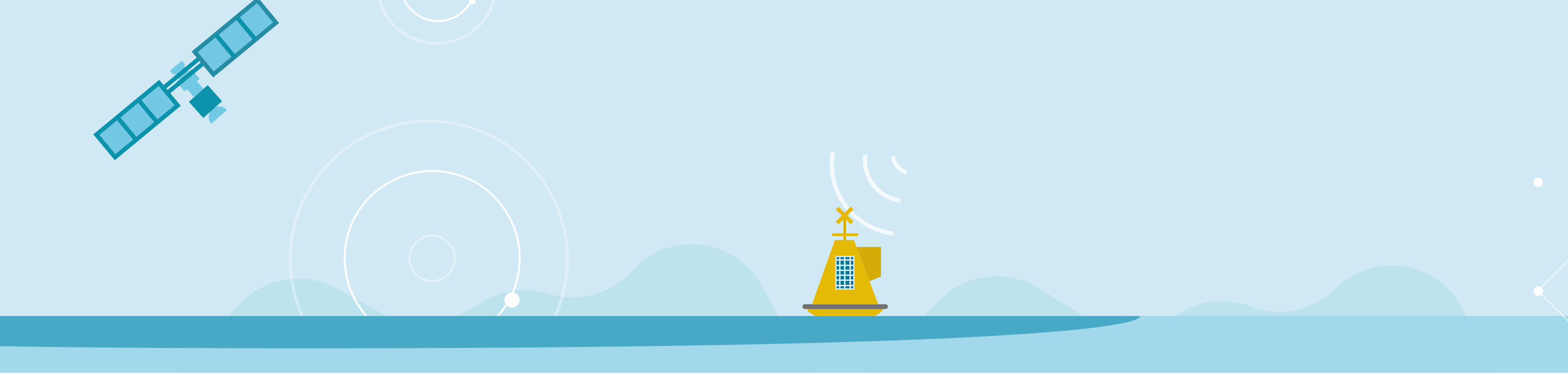
The use of weather indices is not limited to risk analysis – they can also serve as a basis of insurance schemes. Worldwide, there are several parametric insurance schemes in place, which use weather indices. For example, the African Risk Capacity[7] uses satellite rainfall observations and other data to calculate the Water Requirement Satisfaction Index (WRSI), an indicator of crop performance, and an insurance pay-out is calculated based, among others, on the estimate of WRSI. Another parametric insurance product, the Weather Based Crop Insurance Scheme (WBCIS), is in place in the Indian insurance market. Pay-outs in this scheme also depend on how much certain meteorological parameters deviate from reference values during pre-defined plant growth stages. In contrast to ARC, not only precipitation but also temperature, wind speed and solar radiation are considered. The reference values as well as beginning and end dates of growth stages vary between crops and locations.

To conclude this section, various types of weather data, both from observations and from reanalyses, are available. They can provide useful insight and analysis opportunities for modelling agricultural risk, despite difficulties associated with their use. Of course, in order to use historical loss or yield time series for meteorological analyses, they first need to be detrended to reflect present-day conditions, as discussed in the following section.

7. The African Risk Capacity (ARC) is a specialised agency of the African Union, offering drought protection for 35 African countries. It aims to allow governments to launch early response activities in case of droughts, as set out in their pre-agreed contingency plan. For more information, see the ARC website http://www.africanriskcapacity.org

REFERENCES:

- **Caloz, M. and D. Boehlen,** 2015 – *"Obst- und Tafeltraubenanlagen der Schweiz 2015"*. Fachbereich Pflanzliche Produkte, 30. November 2015, Bundesamt für Landwirtschaft BLW.
- **Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones, M. New,** 2008 – *"A European daily high-resolution gridded dataset of surface temperature and precipitation"*. J. Geophys. Res (Atmospheres), 113, D20119.
- **Hoogenboom, G., J.W. Jones, P.W. Wilkens, C.H. Porter, K.J. Boote, L.A. Hunt, U. Singh, J.I. Lizaso, J.W. White, O. Uryasev, R. Ogoshi, J. Koo, V. Shelia, and G.Y. Tsuji,** 2015 – *"Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.6 (http://dssat.net)"*. DSSAT Foundation, Prosser, Washington.
- **Inness P.M. and S. Dorling,** 2013 – *"Operational Weather Forecasting"*. John Wiley & Sons: London.
- **Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijsman, and J.T. Ritchie,** 2003 – *"DSSAT Cropping System Model"*. European Journal of Agronomy, 18, pp. 235-265.
- **Kaufmann, R.K. and S.E. Snell,** 1997 – *"A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants"*. American Journal of Agricultural Economics, 79(1), pp. 178-190
- **Klein Tank, A.M., F.W. Zwiers, and X. Zhang,** 2009 – *"Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation".* Geneva: WMO.
- **Lahoz, W.A., and P. Schneider,** 2014 – *"Data Assimilation: Making sense of Earth Observation"*. Front. Environ. Sci., 2, pp. 16.
- **MeteoSchweiz,** 2016 – *"Klimareport 2015"*. Bundesamt für Meteorologie und Klimatologie MeteoSchweiz, Zürich. 84 S.
- **Proebsting, E.L and H.H. Mills,** 1978 – *"Synoptic analysis of peach and cherry flower bud hardiness"*. J. Am. Soc. Hortic. Sci. 103, pp. 842-845.
- **Rippey, B.R.,** 2015 – *"The U.S. drought of 2012"*. Weath. Clim. Extremes, 10, pp. 57–64.
- **Saha, S., et al.** 2010 – *"The NCEP Climate Forecast System Reanalysis".* Bull. Amer. Meteor. Soc., 91, pp. 1015–1057.
- **Saha, S., et al.** 2014 – *"The NCEP Climate Forecast System Version 2".* J. Climate, 27, pp. 2185–2208.



IN SEARCH OF A PATTERN: DETRENDING HISTORICAL CROP YIELD DATA

THE BACKGROUND: WHAT IS DETRENDING?

Historical yield data are often the basis for risk assessment in agriculture. They may tell us something on the past performance of a crop in a region. However, what can we derive from this historic information for today's or even future yield levels?

Yield levels are a result of many interacting factors, related to management practices, such as crop varieties, tillage, fertilizer and pesticide input, or related to environmental conditions, such as soil properties and weather conditions. These factors change in the course of time with the result of increasing or decreasing yield levels. Hence, we have to account for these changing conditions to make yield levels comparable over time. This is done by the so-called "detrending" and the resulting yields are referred to as detrended yields. The ultimate goal is to derive a series of yield data, which are representative for today's agricultural systems in today's climate.

THE CHALLENGE

To identify a trend in a yield time series, statistical models are used. However, there are several challenges associated with the application of these methods, most of them related to data constraints such as missing data and insufficient information.

- Yield time series are often short, contain outliers and inconsistencies. Outliers are data points, which deviate from the majority of observations, for instance due to an extreme weather event or a pest infestation. Similarly, in cases where only very few data points are available trend estimations are challenging. Inconsistencies, for instance resulting from human error, have to be detected and corrected (see previous section "Slice and dice the data cube – the data standardisation challenge").
- In some countries, innovation is spread rather homogenously and crop yields have increased linearly over time. In contrast, in other countries managerial abilities and access to credit markets are much more diverse among regions and over time. This uneven spatial and temporal spread leads to a high heterogeneity among farms. Hence, trend estimations on a higher aggregation level, e.g. the county or regional level, which is often done due to data constraints, might lead to severe biases.
- Ideally, we would know all the variables having a notable impact on the final yield outcome. This could include, among others, fertilizer input, timing of certain farming activities, soil information and weather variables (see previous section "No rain, no grain: the use of meteorological data"). In reality though, we often have little or no information on these factors and their interactions.
- In addition, not all changes have a uniform effect on yield level. For instance, a new variety, which boosts yield levels substantially from one year to the other, may cause a larger jump in yields than any one before it.

All these factors make it difficult to differentiate between short term yield variability and a potential trend. **Data constraints make consistent trend estimations sometimes very challenging or even impossible.**

THE APPROACH

In order to avoid misspecification of trend patterns, the choice of the detrending method should depend on the characteristics of the yield time series, as highlighted above. Many different methods exist, and here we outline just two of these. Ordinary Least Squares (OLS) is often used to account for a linear trend. This is a non-robust method which minimises the sum of the squared differences between an observed and predicted response. For time series with many outliers, robust regression techniques are favourable, as these techniques are less sensitive towards outlying observations. An example of such a robust technique is the Modified M-estimator (MM) (see e.g. Yohai et al. 1991). Besides these parametric techniques, there are as well semi- and non-parametric regression techniques (e.g. Ker and Coble 2003; Goodwin and Ker 1998).

Independent of the method used to estimate a trend, the uncertainty around the existence and the magnitude of the trend estimate should be investigated. Knowledge of this uncertainty can then help decide if, how, and to what extent detrending should be applied (Clarke et al. 2012). This is of importance as the definition of a loss in insurance heavily depends on the selected detrending method.

THE IMPACT: AN EXAMPLE

Figure 5 shows a hypothetical yield time series (black line) and a threshold value (dashed grey line). Here, the threshold is defined as 80% of a 7-year average, similarly as it is done for the yield based insurance products currently in place in India. The threshold value represents the yield level below which a loss occurs and a pay-out is triggered. The raw yield data have been detrended using a non-robust method (OLS) and a robust method (MM-estimator).

To illustrate the effect of detrending and difference of the two detrending methods, loss costs are calculated. The loss costs are defined as the percentage yield deficit below the threshold value. The raw yield data (black line) have a Loss Cost (LC) of 4.1% for the period from 2004 to 2014. The yields fall below the threshold value three times in this time period. If one uses the non-robust OLS detrending technique (red line), LC drops to 0%, meaning that the threshold was never crossed during the same period of time. The robust detrending technique (blue line) estimates the yield levels to lie in between the raw yields and the detrended yields based on OLS, with a resulting LC of 1.2%. The difference between the two detrending methods lies mainly in the interpretation of the very high yield in 2014. The non-robust OLS estimator takes the 2014 yield fully into account, while the robust MM-estimator confines the influence of this last data point. Further information is necessary to understand whether this 2014 yield is exceptional (e.g. due to an uncommon and exceptional combination of favourable weather conditions) and should be rather treated as an outlier or whether this yield is the "new normal" (e.g. the results of a new rice variety with higher yield levels) and we expect yield levels around this magnitude in the future.

With this example it becomes apparent that **the choice of detrending approach has to be made carefully due to its large impact on pay-out estimates and further risk analysis.**

REFERENCES:

- **Clarke, D. J., Mahul, O. and Verma, N.,** 2012 – Index Based Crop Insurance Product Design and Ratemaking: The Case of Modified NAIS in India, Policy Research Working Paper 5986, The World Bank.
- **Goodwin, B.K. and A.P. Ker,** 1998 – *"Nonparametric estimation of crop yield distributions: Implications for rating group-risk crop insurance contracts"*. American Journal of Agricultural Economics, 80, pp.139–153.
- **Ker, A.P. and K. Coble,** 2003 – *"Modeling Conditional Yield Densities"*. American Journal of Agricultural Economics, 85, pp. 291–304.
- **Yohai, V., W. Stahel, and R.H. Zamar,** 1991 – *"A procedure for robust estimation and inference in linear regression"*. In: Directions in Robust Statistics and Diagnostics. The IMA volumes in mathematics and its applications, 34. Springer-Verlag, New York.

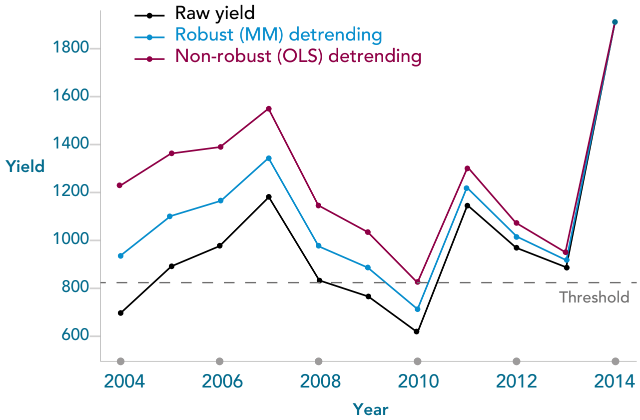
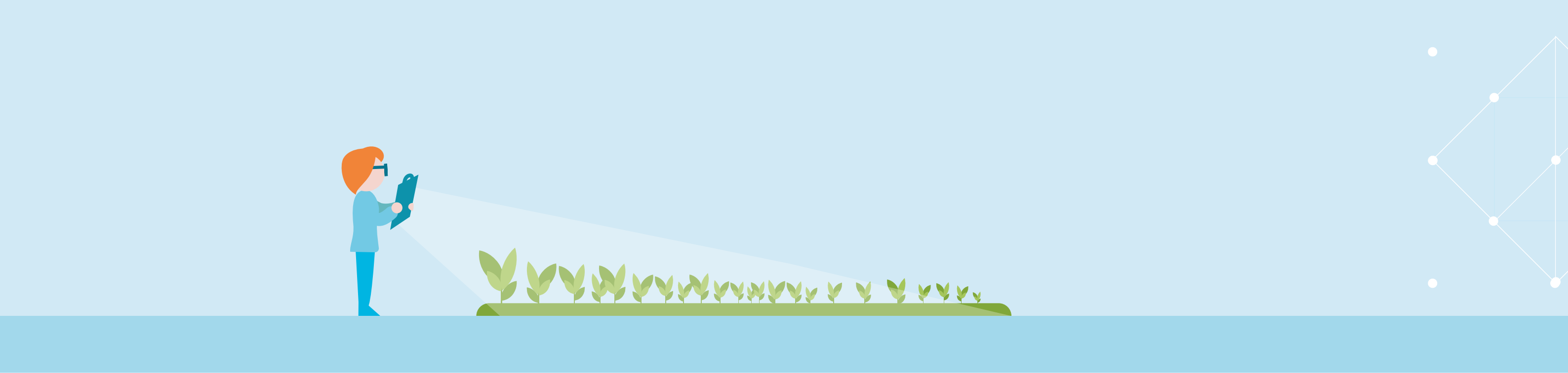


FIGURE 5: HYPOTHETICAL YIELD TIME SERIES

The lines show the raw yield data (in black, with LC 4.1%), and the detrended yields determined using a robust MM-estimator method (in blue, with LC 1.2%) and using a non-robust OLS method (in red, with LC 0%). The dashed grey horizontal line indicates the threshold yield, corresponding to 80% of the 7-year average yield. Yield values below the threshold yield indicate a loss.

Source: SCOR calculation based on hypothetical data example



A REAL CASE EXAMPLE FOR INDIA

SCOR is participating in a yield-based insurance scheme in India. As the insurance losses of this product are calculated from yield data, it is important to have in place a sound method to detrend yield data. SCOR has developed a detrending method that aims to estimate insurance losses using simple and well-known mathematical concepts.

Loss settlement in this Indian agricultural insurance product is done at village level but in practice pricing is conducted using data at the lowest available level, which is often higher than village level. In the method described below the highest available resolution data are always taken as the basis for detrending so that local particularities in terms of trends are taken into account. At the same time, higher-level (district level) trends are also considered in the detrending process, in order to mitigate data quality and data homogeneity issues, which are often encountered in village-level data.

Yield time series processed for pricing the agriculture insurance product in India have diverse statistical properties and are influenced by various factors, which change by crop and by location. Ideally, each time series should be analysed separately. However, the sheer number of yield time series to detrend (more than 100’000 for the whole country) would make an analysis on an individual basis very time consuming. An objective and efficient detrending method, suitable for processing large datasets has been developed instead.

As a first step in the detrending process, a straight line is fitted on each individual yield time series using the non-robust OLS regression. The trend estimates obtained by OLS have uncertainty due to the limited data (usually 10 years), the presence of data variability around the trend line (including outliers) and the often small magnitude of the trend estimate. The issues above are taken into account by use of hypothesis testing: detrending is only applied if after carrying out a hypothesis test the trend is found to be statistically significant.

Even a statistically significant trend can be the result of problems in the available data. For instance, village-level yield data can be inhomogeneous: a strong but artificial trend is introduced because the data resolution increases at some point and the more recent higher-resolution data have different statistical properties compared to the older data. To address such problems, statistically significant individual trends are compared to their corresponding district-level trends, which tend to suffer less from data quality and data inhomogeneity issues. Namely, detrending is applied if the individual and the district-level trends have the same sign. In these cases, the district-level trend (instead of the individual village-level trend) is removed. Insurance Loss Costs are then calculated using the detrended time series. While simply removing the original village-level trend often has a huge impact on the Loss Cost result, detrending partly using the district-level trend tends to have a more moderate influence. The method described above is summarised in Figure 6.

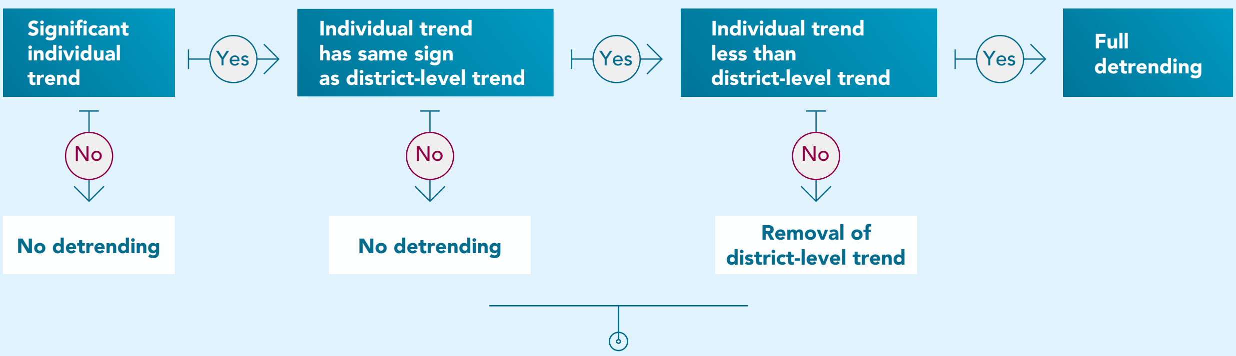


FIGURE 6: DETRENDING APPROACH
The basic steps of the partial detrending approach developed by SCOR for the yield insurance product in India.

SCOR The Art & Science of Risk | P&C

CONCLUSIONS AND OUTLOOK

This technical newsletter has introduced and discussed three important aspects of risk modelling in agriculture insurance. Standardising the wide set of available data will open a completely new range of analysis possibilities. Together with a more systematic use of weather data and a sound methodology for detrending, this contributes to an enhanced quality of risk modelling in agriculture insurance. This in turn supports sound underwriting decisions and hence guarantees sustainability of agriculture reinsurance.

While SCOR is convinced of their general importance, there are various other important questions which are not addressed in this newsletter. These include the use of satellite imagery, restating of historical portfolio information, or a better integration of crop yield models into the risk assessment, to name a few.

SCOR believes in the growing importance of risk modelling for the agriculture insurance industry as a whole. Therefore, **SCOR has invested in the creation of a dedicated Agriculture Risk Modelling Unit.** This team is not only focused on the enhancement of SCOR's in-house modelling capabilities, through developing tools like SEED or STRATUS, but also supports its clients with the development of customised risk solutions and technical advice.

This article is written by :

Modelling team:

Tobias HOFFMANN, Head of Agri & Specialty Risk Modelling
thoffmann@scor.com
Iakovos BARMPADIMOS, Risk Analyst
ibarmpadimos@scor.com
Sarah CONRADT, Risk Analyst
sconradt@scor.com
Duri FLORINETH, Senior Risk Analyst
dflorineth@scor.com
Brigitte PABST, Senior Pricing Actuary
bpabst@scor.com
Hanna PŁOTKA, Risk Analyst
hplotka@scor.com

For more information, please contact our team

Underwriting team:

René KUNZ, Chief Underwriting Officer Agriculture rkunz@scor.com
Yvonne BUSCHOR, ybuschor@scor.com
Henri DOUCHE, hdouche@scor.com
Guillermo GONSETH, ggonseth@scor.com
Fanny ROSSET, frosset@scor.com
Michael RUEEGGER, mrueegger@scor.com
Daniela SCHOCH BARUFFOL, dschoch@scor.com
Swapnil SONI, ssoni@scor.com
Wei XU, wxu@scor.com

PLEASE FEEL FREE TO VISIT US AT SCOR.COM

SCOR P&C
5, avenue Kléber - 75795 Paris Cedex 16
France
scorglobalpc@scor.com

SCOR The Art & Science of Risk | P&C

TO GET THE FULL RANGE OF TECHNICAL NEWSLETTERS, PLEASE CONTACT SCORGLOBALPC@SCOR.COM

Editor: SCOR P&C Strategy & Development
ISSN: 1967-2136