CropWatch online resources:

B. Methodology [updated December 23 2015]

CropWatch analyses of global production and environmental and agricultural trends are based on remote sensing and ground data and a combination of well-established and innovative methodologies. Over the years, the CropWatch system has been updated regularly as new data and methodologies became available or matured. The latest bulletins have used a new hierarchical approach of indicators and spatial scales and a new operational methodology, while also taking advantage of several new remote sensing data sources.

The sections below provide details on the geographic units of analysis (section 1), data sources (section 2), and methodologies (section 3).

1. Geographic units of analysis

CropWatch uses sixty-five Monitoring and Reporting Units (MRU) and several Major Production Zones (MPZ), as well as 31 individual countries (including China) as the basis for its global and regional analyses. For large countries, first-level administrative divisions are also used. Indicators are further calculated for 173 countries and territories, with data provided in bulletin annexes or on the Website. (For more detail, see also the Online Resource: Definition of Spatial Units on www.cropwatch.com.cn.)

1.1 Country boundaries

Boundaries are based on the Global Administrative Units Layer (GAUL) boundaries for all national and sub-national units, except for China, where official Chinese data have been used. GAUL is available from FAO GeoNetwork(1).

Indicators are computed for the CPSZs, MPZs, as well as the selection of 173 countries and territories. The polygons constitute a subset of GAUL_0 (the national level) after exclusion of all “small” polygons defined as those smaller than a 25 km x 25 km pixel, empirically measured by polygon perimeter length.

1.2 Crop masks and arable land masks

Global crop masks were provided by JRC. The original crop masks were created for global water satisfaction indices and cover 11 crops. In the CropWatch analyses, maize, rice, soybean, and wheat masks at 0.25 degree resolution are used. Other sources include major crop areas by USDA (2) and the suitability map for rain-fed plus irrigated crops by FAO/IIASA (3)(4).

The arable land mask was created by joining the arable land from MODIS-derived land use and land cover products for 2010 and 2011 (5), Version 2 International Geosphere Biosphere Programme (IGBP) global land cover dataset (IGBP-DISCover) (6), and GlobCover 2009 (7) (8). The arable land maps for China in 2000, 2005, and 2010 were extracted from ChinaCover 2000, 2005, and 2010, provided by the Institute of Remote Sensing and Digital Earth (RADI) (9).
2 Data

2.1 NDVI
NDVI data is mostly MODIS NDVI provided by NASA, selected mainly because of its high spatial and temporal consistency. The NDVI data covers January 2002 to the end of January 2014. Only MODIS Terra Land Level 3/Level 4 16-Day Tiled Products (found on the LAADS Website (10)) with one kilometer resolution were used.

In addition, long-term average NDVI over the years from 1999 to 2012 with a resolution of 0.1875 degree based on SPOT-VEGETATION was provided by VITO (11).

2.2 Temperature
Air temperature is a global gridded (0.25x0.25 degree) 10-daily product from January 2000 to the end of the current reporting period generated based on the Global Surface Summary of the Day (GSOD) dataset, available from the Global Change Master Directory (56).

The GSOD dataset is derived from the Integrated Surface Hourly (ISH) dataset, DSI-3505 (C00532), and is produced by the National Climatic Data Center (NCDC). Its online data dates back to 1929 and the latest daily summary data are normally available 1-2 days after the date-time of the observations used in the daily summaries. Over 9000 stations’ data (including air temperature, dew point, sea level pressure, wind speed, precipitation, snow depth) are typically available in this dataset.

The mean daily air temperatures from 2000 to now in the GSOD dataset were extracted to calculate the 10-daily air temperature for each station. Then kriging interpolation was applied combined with STRM_DEM data (57) considering temperature elevation correction to generate the 0.25x0.25 degree global dekadal product.

2.3 PAR
Photosynthetically active radiation (PAR), which covers radiation in the 400 to 700 nm range, is an important biological variable. The ten-day PAR product of this bulletin from the year 2000-2013 is downloaded from MERRA(12) (Modern-Era Retrospective analysis for Research and Applications) and the original hourly PAR product was converted to global gridded (0.25x0.25 degree) ten-day’s map later.

The standard MERRA data product is a NASA atmospheric reanalysis for the satellite era using Version 5 of the Goddard Earth Observing System model (GEOS-5) and its associated data assimilation system (DAS). For January 2014, version 5.2.0 (13)(14) radiation data was obtained from JRC (15).

2.4 Rainfall
CropWatch has assembled composite dekad rainfall grids for the period from 2000.1 to 2014.1 covering the land surface from 50 degrees south (50S) to 90 degrees north (90N) at 0.25 degree spatial resolution. Two rainfall products resources were merged: (i) version-7 TRMM rainfall (63) at 0.25 degree resolution
extending from 50S to 50N and (ii) ERA-Interim (ERA-I) and ECMWF Operational (ERA OPE) dekad rainfall products (16) downloaded from JRC for other regions. For the period between 2000.10 and 2013.3, TRMM 3B42 daily rainfall was used between 50S and 50N and other regions used ERA-I dekad rainfall products. From 2013.4 to 2013.8, dekad rainfall products were combined with daily TRMM 3B42 rainfall products from 50S to 50N and ERA OPE dekad rainfall products were used in other regions. Since August 2013, TRMM 3b42 3hours real time products were combined into dekad rainfall products from 50S to 50N and ERA OPE were used in other regions.

2.5 VHI

The Vegetation Health Index (VHI) is an effective indication of the crop growth condition. In this bulletin, the VHI was calculated (67, 68, 69) by weighting the Vegetation Condition Index (VCI) and Temperature Condition Index (TCI), which were downloaded separately from the NOAA Star Center for Satellite Applications and Research GVI-x VH dataset (70). The equation is as follows:

\[
\text{VHI} = a \times \text{TCI} + (1 - a) \times \text{VCI}
\]

The weighting factor (a) is an empirical coefficient for which the constant value of 0.5 was used.

2.6 Biomass

In the CropWatch bulletin, the net primary production potential (NPP) according to Lieth’s “Miami model” (71, 72) is used as a biomass index (referred to as “biomass” or “biomass accumulation” in the text) to describe the global potential biomass situation. Two environmental factors, temperature and precipitation are considered in Miami model with following equations:

\[
\text{NPP[\text{Rain}(\text{dek})]} = 3000(1 - e^{-0.000664\text{Rain}(\text{dek}) < 36})
\]

\[
\text{NPP[\text{Temp}(\text{dek})]} = 3000/(1 + e^{1.315 - 0.119\text{Temp}(\text{dek})})
\]

where, \(\text{Rain}(\text{dek})\) is accumulated dekad rainfall in mm, \(\text{Temp}(\text{dek})\) is average dekad temperature in ℃.

Finally, the biomass accumulation is expressed as the combined effect of rainfall and temperature accumulated during a reference period (dekad from i to n) using following equations:

\[
\text{NPP}_\text{Rain} = \sum_{\text{dek}=1}^{n} \text{NPP[\text{Rain}(\text{dek})]} / n
\]

\[
\text{NPP}_\text{Temp} = \sum_{\text{dek}=1}^{n} \text{NPP[\text{Temp}(\text{dek})]} / n
\]

\[
\text{NPP} = \min(\text{NPP}_\text{Rain}, \text{NPP}_\text{Temp})
\]

the unit of biomass is grams of dry matter per square meter over concerned period.
3 Methodology

3.1 CropWatch Agroclimatic Indicators (CWAI)

To compare the agricultural impact of environmental variables across years and geographic areas (e.g., countries), it is necessary to use an index that gives a high weight to agricultural areas. The intention is to derive “one number per year,” index so that it becomes comparable with other variables that are available at the same spatial and temporal scale, i.e., agricultural statistics and a number of socio-economic indicators and variables.

Such an index has been defined earlier for rainfall (73) and applied for synthetic studies (e.g., (74) (75)). The CropWatch bulletin expands the notion of Rainfall Index to Agroclimatic Indicators by applying basically the same approach to PAR and temperature.

The CWAIIs are defined for one variable V (rainfall, PAR, temperature) and one polygon, which may be a MRU or a country or a crop mask or any other spatial unit. They are based on spatial grids at a resolution of 25 km and are computed as the spatial average over an arable land mask of the values of V accumulated over a user defined period, using the net primary production potential (1976-2000 VASClimO based NPPP) (72) as the weighting factor: the most productive pixels receive a higher weight than unproductive ones. The accumulation or averaging period coincides with the reporting period covered by each bulletin, unless specified otherwise.

3.2 Production, Area and Yield

Production

The main concept of the CropWatch methodology for estimating production is the calculation of current year production based on information about last year’s production and the variations in crop yield and cultivated area compared with the previous year. The equation for production estimation is as follows:

\[
\text{Production}_i = \text{Production}_{i-1} \times (1 + \Delta \text{Yield}_i) \times (1 + \Delta \text{Area}_i)
\]

where \(i\) is the current year, \(\Delta \text{Yield}_i\) and \(\Delta \text{Area}_i\) are the variations in crop yield and cultivated area compared with the previous year; the values of \(\Delta \text{Yield}_i\) and \(\Delta \text{Area}_i\) can be above or below zero.

Yield

For the 31 countries monitored by CropWatch, yield variation for each crop is calibrated against NDVI time series, using the following equation:

\[
\Delta \text{Yield}_i = f(\text{NDVI}_i, \text{NDVI}_{i-1})
\]

where \(\text{NDVI}_i\) and \(\text{NDVI}_{i-1}\) are taken from the time series of the spatial average of NDVI over the crop specific mask for the current year and the previous year. For NDVI values that correspond to periods after
the current monitoring period, average NDVI values of the previous five years are used as an average expectation. \( \Delta \text{Yield}_i \) is calculated by regression against average or peak NDVI (whichever yields the best regression), considering the crop phenology of each crop for each individual country.

**Area**

A different method is used for areas. For China, Australia, Canada and the U.S., CropWatch combines remote-sensing based estimates of the crop planting proportion (cropped area to arable land) with a crop type proportion (specific type area to total cropped area). The planting proportion is estimated based on an unsupervised classification of multi-temporal high resolution satellite images from HJ-1 CCD and GF-1 images or time series of MODIS NDVI data. The crop-type proportion is obtained by the GVG instrument from field transects. The area of a specific crop is computed by multiplying farmland area, planting proportion, and crop-type proportion of the crop.

To estimate crop area for wheat, soybean, maize, and rice for other 27 countries, CropWatch relies on the regression of crop area against cropped arable land fraction of each individual country (paying due attention to phenology):

\[
\text{Area}_i = a + b \times \text{CALF}_i
\]

where \( a \) and \( b \) are the coefficients generated by linear regression with area from FAOSTAT or national sources and \( \text{CALF} \) the Cropped Arable Land Fraction from CropWatch estimates. \( \Delta \text{Area}_i \) can then be calculated from the area of current and the previous years.

The production for "other countries" (outside the 31 CropWatch monitored countries) was estimated as the linear trend projection for 2014 of aggregated FAOSTAT data (using aggregated world production minus the sum of production by the 31 CropWatch monitored countries).

**3.3 Cropping intensity (CI)**

The Cropping Intensity index describes the extent to which arable is used over a growing season. It is the ratio of total crop area of all planting seasons in a year to the total area of arable land (84). It also can be used to describe the food-producing potential of an area (85). This report adopts the method proposed by Fan and Wu (86) based on an NDVI time-series derived from MODIS Terra and reconstructed by the S-G filter method (87). The following descriptors were derived from the data: number of peaks, width of each peak and peak values at pixel level.

The calculation of Cropping Intensity involves counting the number of peaks in the NDVI profile; we use the difference method to extract the maximum value of the discrete points. Assuming a pixel is in a discrete points series, the values of pixels before and after the point constitute a point series \( S1 \) which include N-1 pixels,

\[
S1=\text{DIFF}(S)
\]
where DIFF is the function used to calculate the difference among these points. Then, determine the sign of each pixel in S1, if the value of a pixel less than zero, assign the element value to minus one, otherwise, the element value is assigned to one; these values constitute a point series S2,

\[ S2 = \text{SIGN} (S1) \]

Where SIGN is the function used to determine whether the values are positive or negative. As a next step, calculate the difference of pixels before and after a point in series S2, which yields the points series S3

\[ S3 = \text{DIFF} (S2), \]

Finally, count the number of pixels with value minus two, which is the number of peaks in the NDVI profile. Based on the above equations and with other limitation, the model can generate a Cropping Intensity of one, two, and three per pixel, to illustrate areas with a single, two, or three crop seasons respectively.

### 3.4 Cropped Arable Land Fraction (CALF)

CALF was introduced to demonstrate the proportion of cropped arable land to the total arable land over a certain geographic area (MPZs, countries or sub-national units). Previous studies have shown a high correlation between NDVI and the photosynthetic biomass of cropped fields (88, 89). MODIS NDVI time series were used to identify whether an agricultural pixel is cropped or uncropped. For each pixel, time series MODIS NDVI values were extracted from time series NDVI images and smoothed using Savitzky-Golay (S-G) filter (87). Lagrange polynomials were employed to extract extreme points for the smoothed NDVI curves. Maximum NDVI peak values at each extreme point were acquired from smoothed NDVI profiles from 2001 to 2014. Average and standard deviation of seasonal NDVI peaks (NDVI\text{m} and NDVI\text{std}) were calculated based on maximum NDVI peak values over the same growing season from 2001 to 2014. A NDVI threshold method (90, 91) together with a decision tree were used to identify whether an agricultural pixel was cropped or during the reporting period. The difference between average and standard deviation of annual NDVI peaks was incorporated as a threshold to eliminate the inter-annual variability of biomass, crop phenology, and crop rotation.

For the major crop producing regions in China, multi-temporal high resolution images (HJ-1 CCD and GF-1 images) over each growing season were acquired and processed. The cropped and uncropped arable land map was generated using decision tree.

CALF over those regions was then calculated based on cropped and uncropped map and zonal statistical analysis.

### 3.5 Maximum Vegetation Condition Index (VCI)

Based on the good relationship between NDVI and plant productivity and biomass (92, 93, 94, 95), NDVI can be used as a proxy of crop biomass. Based on the Vegetation Condition Index (VCI) which was proposed by Kogan (1990) (17), the maximum VCI is adopted in CropWatch bulletins to describe the
optimal crop condition of the current period compared with historical maximum crop biomass potential using the following equation:

$$\text{Maximum VCI} = \frac{NDVI_{\text{max},c} - NDVI_{\text{min},h}}{NDVI_{\text{max},h} - NDVI_{\text{min},h}}$$

where $NDVI_{\text{max},c}$ is the maximum NDVI of a fixed period, $NDVI_{\text{max},h}$ and $NDVI_{\text{min},h}$ respectively the historical maximum NDVI and historical minimum NDVI of the same period using long term time series NDVI data sets. Considering the crop minimum NDVI may be contaminated by cloud or non-vegetation pixels, in this report, the empirical minimum vegetation NDVI value (0.15) is introduced to calculate $NDVI_{\text{min},h}$ with following equation:

$$NDVI_{\text{min},h} = \max(0.15, NDVI_{\text{min},h0})$$

where $NDVI_{\text{min},h0}$ is the original minimum NDVI of the study period from time series NDVI datasets. The value of Maximum VCI ranges from 0 to 1. The higher the maximum VCI value is, the better crop condition and larger biomass potential it indicates for a concerned period. Therefore, crop maximum VCI is more meaningful when calculated during crop growing period.

### 3.6 Cropping structure

Cropping structure is an additional variable only applied to China and some of the other large countries. It precisely illustrates the proportion of area under a given crop type to the total sown area (i.e., crop type proportion) for geographic areas (i.e., provinces). The crop type proportion was estimated by combining GPS, video, and GIS data (collectively referred to as GVG) from field transects (76). The specifically developed GVG instrument collects thousands of field photos that are used to estimate the proportion of different crop types with accuracy above 98 percent (96, 97, 98, 85).

### 3.7 Time profile clustering

Based on a time series of pixel-based (raster) images, time profile clustering is a method that automatically or semi-automatically compares the time profiles of all pixels and distributes them among a limited number of "typical" behaviors (classes) that can be mapped. The method has the advantage of very synthetically describing the spatial distribution of typical time profiles (99, 100). In the bulletins, the CropWatch team uses the SPIRITS software developed for JRC/MARS by VITO (101). According to the bulletin, Rainfall, Temperature, NDVI and VHI profiles have been clustered, especially the difference between the current season and the average of the last five years or twelve years taken as reference.

### 3.8 Pests and diseases monitoring

Occurrence and development of crop pests/diseases is a result of multiple factors, such as crop type and growth status, pests/pathogens and their transmission rules, meteorological conditions and habitat. According to the large spread area and rapid development of pests/diseases, remote sensing technology with the capability of observing the earth in continuous space can meet the requirements of timely
monitoring of crop pests/diseases. In this report, we construct new models for maize and rice pests/diseases monitoring, which combined with spectral features of crop canopy, physiological properties of specific pests/diseases, meteorological conditions and habitat to quantitatively monitor occurrence and development of crop pests/diseases.

For crop pests/diseases monitoring with remote sensing technologies, priori knowledge is a kind of effective and useful background information, so extracting landscape features, pests/diseases transmission rules, sensitive meteorological conditions and habitat from historical data is normally the first step. Then, for specific type of crop pest/disease, corresponding spectral features will be constructed: 1) for armyworm (maize), given the fact that armyworm eats maize leaves that destroy crop morphological structure, so we established Armyworm Index (AI) integrating with Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974) and Modified Soil Adjusted Vegetation Index (MSAVI) (Rouse et al., 1974) to monitor armyworm with RELIEF algorithm (Robnik-Šikonja and Kononenko, 2003), for which AI is an index considering the effects of crop canopy cover and soil background information (Zhang et al., 2015); 2) for northern leaf blight (maize), it brings scabs to maize leaves, and reduction of chlorophyll and water contents, so we proposed Northern Leaf Blight Index (NLBI) combined with NDVI and Normalized Difference Water Index (NDWI) (Gao, 1996) to monitor northern leaf blight with RELIEF algorithm, where NLBI is an index considering the features and changing information of crop chlorophyll and water contents; 3) for rice planthopper, its main damage is sucking rice juice, which hinders nutrient transport efficiency, transmits virus, and then stunts crop growth. Based on these facts, we constructed Rice Planthopper Index (RPI) based on NDVI and Temperature Vegetation Drought Index (TVDI) (Sandholt et al., 2002) to monitor rice planthopper through RELIEF algorithm, while RPI is an index considering crop growth status, temperature, and crop water content; 4) for rice sheath blight, its main damage to rice is forming disease spots on leaves, which leads to leaf rot, chlorophyll and water content reduction, then we built Sheath Blight Index (SBI) with Plant Senescence Reflectance Index (PSRI) (Merzlyak et al., 1999) and NDWI to monitor rice sheath blight based on RELIEF algorithm, where SBI is an index considering crop health condition; At last, from the aspects of physiological properties of pests/diseases and their damage to crop, Pest Index (PI) or Disease Index (DI) (Huang et al., 2014) is constructed to evaluate the specific damage levels of pests/diseases, moreover, evaluation system is built to monitoring crop pests/diseases and formulating management strategies.
REFERENCES

11. Personal communication with Herman Eerens at VITO. [Online]

55. Personal communication with Herman Eerens at VITO.


93. Accuracy of the AVHRR Vegetation Index as a predictor of biomass, primary productivity and net CO 2 flux. Elgene O. Box, B. N. Holben, V. Kalb. 1989, Vegetation, Vols. 80: 71–89.
106. Gao B. C. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of


