Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 6 (2019) 316–325

journal homepage: www.elsevier.com/locate/inpa

# An innovative intelligent system based on remote sensing and mathematical models for improving crop yield estimation

*Mohamad M. Awad*

*National Council for Scientific Research, Remote, Sensing Center, P.O. Box 11-8281, Beirut, Lebanon*

ARTICLE INFO

*Article history:*
Received 19 July 2018
Received in revised form
4 March 2019
Accepted 3 April 2019
Available online 5 April 2019

*Keywords:*
Crop yield
Environment
Remote sensing
Image processing
Evapotranspiration
Intelligent system

ABSTRACT

There are many crop yield estimation techniques which are used in countries around the world, but the most effective is the one based on remote sensing data and technologies. However, remote sensing data which are needed to estimate crop yield is incomplete most of the time due to many obstacles such as climate conditions (percentage of cloud cover), and low temporal resolution. These problems reduce the effectiveness of the known crop yield estimation techniques and render them obsolete. There was many attempts to solve these problems by using high temporal resolution and low spatial resolution images. However, this type of images are suitable for very large homogeneous crop fields. To compensate for the lack of high spatial resolution satellite images, a new mathematical model is created. Based on the new mathematical model an intelligent system is implemented that includes the use of energy balance equation to improve the crop yield estimation. To verify the results of the intelligent system, several farmers are interviewed and information about their crops yield is collected. The comparison between the estimated crop yield and the actual production in different fields proves the high accuracy of the intelligent system.


## 1. Introduction

Roughly one third of Earth's land is today deployed for agricultural purposes, with more than ten percent used for growing crops and the reminder for pasture. The fast increase in population mean more demands are being put on agriculture than ever before [1]. Monitoring of agricultural activities faces special problems not common to other economic sectors [2]. Firstly, Agricultural activities follows seasonal patterns related to the crop phenology. Secondly, crop production depends on the physical landscape, climatic parameters, and agricultural practices. All these factors are highly variable in space and time. The Food and Agriculture Organization (FAO) [2] indicated that there is need for timeliness agricultural statistics associated with effective monitoring system. Information is worth little if it becomes available too late. Remote sensing can significantly contribute to providing a timely and accurate picture of the agricultural sector, as it is very suitable for gathering information over large areas with high revisit frequency.

In remote sensing, multispectral and hyperspectral satellite images play a major role in crop management, their ability to represent crop growth condition on the spatial and temporal scale is remarkable. These images can describe the crop development, photosynthetic active radiation (PAR), biomass accumulation (Bio), leaf area index (LAI), and actual evapotranspiration (ETa).

E-mail address: mawad@cnrs.edu.lb
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2019.04.001

Many approaches have been developed to translate remote sensing data into crop yields, and several reviews of such approaches exist [3–5]. Some of these approaches have faced several problems in estimating crop yield. One of these problems is the scarcity of remote sensing data suitable for use in crop management because of climatic conditions such as clouds [6,7]. There were several attempts to solve this matter by replacing the lack of data with information from high temporal and low spatial resolutions images such as the method implemented by Petitjean et al. [8].

Most of the time, climatic conditions and low temporal resolution are the main obstacles that prevent decision makers from using remote sensing data to map crops and to estimate crops yield.

There are several articles that use vegetation indices from one date image or multi-temporal images to estimate crop yield. Kasampalis *et al.* [9] in their review of the current available crop yield estimation models they concluded that the main limitations of crop growth models are the cost of obtaining the necessary input data to run the model, the lack of spatial information in some cases, and the input data quality.

Haig [10] conducted a study on a satellite based NDVI to predict crop yield at field level in India. He investigated the relationship between NDVI calculated from satellite images and irrigated rice yield. The results of the study also showed that the correlation between NDVI and rice yield is weak with $R \leq 0.52$. This is due to the fact that a long period of rice growth is covered with water which makes NDVI index an obsolete one.

Prasad et al. [11] combined several parameters such as soil moisture, NDVI, surface temperature, rainfall data of Iowa state in USA for over nineteen years for crop yield calculation and prediction using piecewise linear regression method with breakpoint. In his work, a non-linear multi-variate optimization was utilized that minimizes discrepancy and errors in yield prediction. The method works well for large agriculture area with homogeneous crops.

Bastiaanssen and Ali [12] used Monteith's model [13] to calculate the Absorbed Photosynthetically Active Radiation (APAR), Stanford's model for determining the light use efficiency, and Surface Energy Balance Algorithm for Land (SEBAL) to describe the temporal and spatial variabilities in land wetness conditions. The result of the research showed that there were gaps between the estimated and the actual yield of about 1075 and 1246 kg/ha for wheat and rice. This due to lack of remote sensing data and SEBAL requirements for detailed climate data.

In this paper, the method in [12] is modified and improved with new techniques and models. The modifications are essential and it concerns the energy balance model such as using another model named Mapping EvapoTranspiration at high Resolution with Internalized Calibration (METRIC) [14]. This model depends less on climatic measurements to compute actual evapotranspiration. In addition, the Monteith model is modified with new mathematical model.

The new crop yield estimation system deploys models, techniques, and remote sensing data that are based on using Landsat 7 and 8 satellite images. In addition, the system includes the development of new mathematical model to compensate for the absence of satellite data due to climatic factors and low temporal resolution.

## 2. Data and methods

Agricultural areas in different countries around the world are characterized by different farming practices and diversified natural features. This diversity add to the complexity of handling the problem of crop yield estimation and in turn encourage scientists to find one common solution that can adapt to climate conditions and to the availability of remote sensing data.

### 2.1. *Study area and data type*

The selected study area is located in the largest agricultural area in Lebanon known as Bekaa valley (Fig. 1). The valley has an area size of more than 1200 $km^2$ and the major crops in the valley are wheat and potato. It is also important to note that two Bowen Ratio stations are installed in the area in order to measure different climate parameters which are necessary for irrigation management and for estimating crop yield. The use of two Bowen Ratio one in the middle of the valley and another in the south of the valley is sufficient since in the middle the topography and climate are almost the same for a large area and this applies to the south part of the valley.

Many studies showed success in crop yield estimation when one weather station was used for a large agriculture area with homogeneous topographic land [15].

The two Bowen Ratio stations are located in two different areas one in an agricultural research institute while the other is located in a field owned by a potato chips manufacturer. The distance between both is more than 15 km. The two stations provide many valuable climate data such as net radiation, wind speed and direction, temperature at different

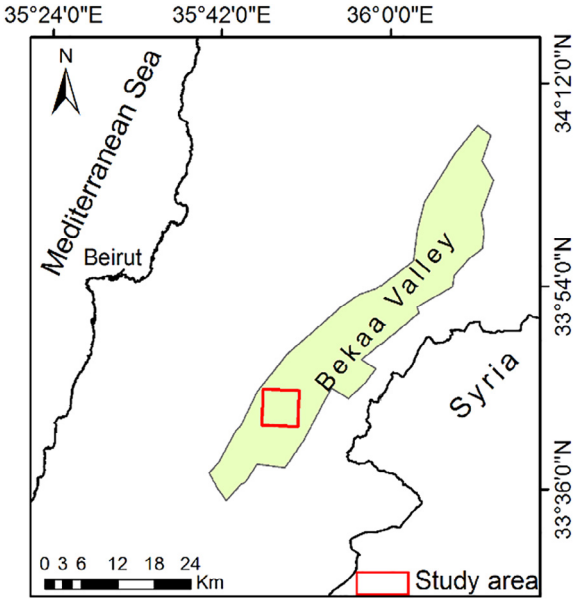


**Fig. 1 – Study area.**

elevations, humidity, global radiation, soil heat flux, and soil humidity. The images of Landsat 7 ETM+ and 8 are used in this research because they are available to download free of charge from USGS Web site [16].

In addition, they are the only free satellite images that have a thermal infrared band with high spatial resolution (60 m compared to 1000 m for Modis satellite). Where the long wave infrared band is used later in computing important parameters for crop yield estimation. Availability of Landsat images is related to the date the crop was planted and to the date that was harvested. Using both Landsat 7 and 8 improves the temporal resolution of image availability from 16 days to 8 days.

### 2.2. Monteith biomass model

Monteith model [13] is based on the photosynthetically active radiation (PAR) (0.4–0.7 m) which is part of the short wave solar radiation (0.3–3.0 m) that is absorbed by chlorophyll for photosynthesis in the plants. According to [17] a value of approximately 45–50% is generally accepted to represent the 24 h average conditions (Eq. (1)).

$$PAR = 0.48K_{inco} \tag{1}$$

Photosynthetically active radiation (PAR) is thus a fraction of the incoming solar radiation ($K_{inco}$). Then a fraction of PAR is absorbed by the plant for carbon assimilation.

Absorbed photosynthetically active radiation (APAR) in $(Wm^{-2})$ can be computed directly from PAR using the following equation

$$APAR = fPAR \tag{2}$$

The factor $f$ can be estimated using Normalized Differences Vegetation Index (NDVI) as is explained in [12].

$$\text{Where } f = -0.161 + 1.257NDVI \tag{3}$$

*NDVI* is the normalized difference vegetation index computed as the difference between near-infrared and red spectrums divided by their sum. The accumulation of biomass is according to the Monteith model proportional to accumulated APAR.

$$Bio_{act}^{tot} = \varepsilon \sum (APAR(t)(t) \tag{4}$$

where $Bio_{act}^{tot}$ in (kg/m$^2$) is the accumulated biomass in period t, $\varepsilon$ is the light use efficiency in gram per mega joules (g MJ$^{-1}$). Light use efficiency $\varepsilon$ varies, if not water short, with *C3* crops [13]. This means that there is no needs to know exact crop type. Some of the *C3* crops vary between wheat, rice, oats, alfalfa, pastures, sugar beet, and potato. This fact has an important inference: conversion for most *C3* crops can be done with the same $\varepsilon$. A more comprehensive global ecology model for computing net production was created by Field *et al.* [18] where they used the following equation for light use efficiency $\varepsilon$.

$$\varepsilon = \varepsilon' T_1 T_2 W \tag{5}$$

where $\varepsilon'$ is the maximum conversion element for above ground biomass when the environmental conditions are optimal, it is equal to 0.29 g/MJ for *C3* crops [19]. W is a function of the effective fraction of the available soil moisture.

The evaporative fraction is totally influenced by soil moisture inside the root zone [20]. The complete details for the calculation of W are explained later in the text.

$$T_1 = 0.8 + 0.02T_{opt} - 0.0005T_{opt}^2$$
$$T_2 = \frac{1}{1 + \exp(0.2T_{opt} - 10 - T_{mon})} \times \frac{1}{\exp(0.3(-T_{opt} - 10 + T_{mon}))} \tag{6}$$

$T_1$ *and* $T_2$ are two different heat functions, $T_{opt}$ (°C) which is the mean air temperature in the month with maximum leaf area index, and $T_{mon}$ (°C) is the mean monthly air temperature. One can notice easily that $T_1$ depends completely on $T_{opt}$ which is in turn affected by canopy biomass. On the other hand, $T_2$ depends on both $T_{opt}$ and $T_{mon}$.

The above values can be obtained from remote sensing images using very complex mathematical models or they can be obtained directly using Bowen Ratio stations.

### 2.3. Enhancement and correction of the images

There are still problems when using Landsat 7 products such as “no data” strips which are caused by Scan Line Corrector (SLC) failure. Normally, Landsat “SLC-off” data refers to all Landsat 7 images collected after May 31, 2003, when SLC failed. These products have data gaps, but are still useful and maintain the same radiometric and geometric corrections as data collected prior to the SLC failure. To fix this problem, a method created by Scaramuzza *et al.* [21] is used to fill gaps in one scene with data from another Landsat scene. This method is available ENVI [22] and is applied as a preprocessing step. More details about how to install and use this method as part of ENVI environment can be found in [23]. In this method, linear transform is applied to the “*filling*” image to adjust it based on the standard deviation and mean values of each band.

The at-surface reflectance for the visible to short-wave is corrected on a band-by-band basis following Tasumi *et al.* method [24]. It works first on correcting images at sensor (Atmospheric correction) and then at surface correction.

Top of atmosphere correction (At Satellite) works as follow:

$$\rho'_{t,b} = Mp * Qcal + Ap \tag{7}$$

where $\rho'_{t,b}$ = TOA reflectance, without correction for solar angle. $Mp$ = Band-specific multiplicative rescaling factor, $Ap$ = Band-specific additive rescaling factor, and $Qcal$ = Quantized and calibrated standard product pixel values (DN). The constants $Mp$ and $Ap$ are extracted from the metadata file that comes with the image. While $Qcal$ value is extracted from the image pixel values. To correct of the TOA for the sun angle the following is applied

$$\rho_{t,b} = \frac{\rho'_{t,b}}{Sin(\theta_{Se})} \tag{8}$$

where $\rho_{t,b}$ is the corrected TOA reflectance, $\theta_{Se}$= Local sun elevation angle. Next step is to correct with respect to at surface reflectance which is done as follow:

$$\rho_{s,b} = \frac{\rho_{t,b} - C_b(1-\tau_{in,b})}{\tau_{in,b}\tau_{out,b}} \tag{9}$$

$$\tau_{in,b} = C_1 \exp\left[\frac{C_2 P_{air}}{K_t Sin(\theta_{Se})} - \frac{C_3 Wp + C_4}{Sin(\theta_{Se})}\right] + C_5 \quad (10)$$

$$\tau_{out,b} = C_1 \exp\left[\frac{C_2 P_{air}}{K_t Cos(0)} - \frac{C_3 Wp + C_4}{Cos(0)}\right] + C_5 \quad (11)$$

$$Wp = 0.14 e_a P_{air} + 2.1 \quad (12)$$

$$e_a = \frac{R_h e_s}{100} \quad (13)$$

$$e_s = 0.6108 * \exp\left(\frac{(17.27 T_{air})}{T_{air} + 237.3}\right) \quad (14)$$

where $\rho_{s,b}$ is at-satellite reflectance for band b, and $C_b$is a band $b$ specific given constant. $\tau_{in,b}$ and $\tau_{out,b}$ are narrowband transmittances for incoming solar radiation and for surface reflected shortwave radiation. $K_t$ is a unit less "clearness" coefficient $\leq$1.0 where $K_t = 1.0$ for clean air and $K_t = 0.5$ for extremely turbid, dusty or polluted air, $P_{air}$ is air pressure (kPa), $Wp$ is perceptible water in the atmosphere (mm), $e_a$ is the actual vapor pressure (kPa), $e_s$ is the saturated vapor pressure (kPa), $R_h$ is the relative humidity in percent, and $T_{air}$ is the air temperature. Air pressure is calculated using a form of the universal gas law equation as standardized by [25,26].

$$P_{air} = 101.3\left(\frac{293 - 0.0065z}{293}\right)^{5.26} \quad (15)$$

### 2.4. Using METRIC to compute evaporative fraction

The computation of the evaporative fraction is based on using Eq. (16) which in turn depends on the energy balance Eq. (17):

$$W = \frac{\lambda E}{R_n - G_0} \quad (16)$$

$$\lambda E = R_n - G_0 - H \quad (17)$$

where $W$ is the evaporative fraction unit less, $H$ is the sensible heat flux in $(Wm^{-2})$, $\lambda E$ latent heat flux in $(Wm^{-2})$, $R_n$ is the net radiation in $(Wm^{-2})$, and $G_0$ is the soil heat flux in $(Wm^{-2})$. The equation can be solved using METRIC which was developed by Allen *et al.* [14]. METRIC is a sort of "hybrid" between pure remotely sensed energy balance and weather-based evapotranspiration methods. Where energy balance is calculated from satellite image which delivers spatial information that includes the available energy, and the sensible heat fluxes for a large area.

In this research, METRIC is used as part of the intelligent system to calculate evaporative fraction and actual evapotranspiration. Both are compared with the values computed from the Bowen Ratio stations measurements. METRIC foundation, principles and techniques are based on SEBAL [27,28]. METRIC works on solving Eq. (16) by calculating each variable separately such that net radiation is calculated based on the following Eq. (18)

$$R_n = RS\downarrow - \alpha RS\downarrow + RL\downarrow - RL\uparrow - (1 - E_0)RL\downarrow \quad (18)$$

where $R_n$ is the net radiation in $(Wm^{-2})$ where $RS\downarrow$ is the incoming shortwave radiation in $(Wm^{-2})$, $\alpha$ is the surface albedo (dimensionless), $RL\downarrow$ is the arriving longwave radiation in $(Wm^{-2})$, $RL\uparrow$ is the emitted longwave radiation in $(Wm^{-2})$, and $E_o$ is the surface thermal emissivity.

$$H = \rho C_p \frac{(T_a - T_s)}{r_H} \quad (19)$$

$H$ is the sensible heat flux in $(Wm^{-2})$, $T_a$ and $T_s$ are the air and surface temperatures in (°C), $\rho$ is the air density (kg $m^{-3}$) and $C_p$ heat capacity of air (Kj $kg^{-1}$) are constants, and $r_H$ is the transfer resistance (s $m^{-1}$) depends on wind speed and surface characteristics.

$$G_0(0.05 + 0.10e^{-0.521LAI})Rn \quad (LAI \geq 0.5) \quad (20)$$

$$G_0 = \max(0.4H, 0.15Rn) \quad (LAI < 0.5) \quad (21)$$

where $G_0$ is the soil heat flux in $(Wm^{-2})$ and LAI is the leaf area index unit less. For more details about the above equations one may refer to the paper of Allen *et al.* [29].

### 2.5. Creating the new mathematical model

Since the estimation of crop yield depends on few satellite images there will be some periodic gaps between the date of planting and harvesting. For this reason, the information obtained from each image is used to create a mathematical model to compensate for missing data.

Mathematical models analyze the observation from the real world such as satellite images in order to predict unforeseen incidents, behavior or productivity. Here they are used to enhance data availability and to predict missing data in a specific information process such as biomass yield estimation. The model works on interpolating missing data from existing ones with minimum error.

Normally, nonlinear curve-fitting (data-fitting) problems are solved using least-squares method [30]. However, the data extracted from satellite images represent specific crop biomass which is considered a complex non polynomial (*NP*) type of problems. For this reason, there is a need for a more reliable method that can solve the problem and which can provide an optimal solution.

The objective is to fit a curve to the crop yield data using unconstrained nonlinear optimization algorithm such as Trust-Region Methods for Nonlinear Minimization [31]. Trust-region methods are efficient, and can solve easily ill-conditioned problems. The basic idea is to approximate a function $F$ with a simpler function $Q$, which reasonably reflects the behavior of function $F$ in a neighborhood $N$ around the point $x$. This neighborhood is the trust region, a trial step $s$ is computed by minimizing over $N$ $((\ sMin)\{Q(s), s \in N\})$. This is the trust-region sub-problem, mathematically the Trust-Region sub-problem is typically stated as:

$$Min\left\{\frac{1}{2}s^T S + S^T\ \Omega \ \ such\ that\ |Ds \leqslant \psi\right\} \quad (22)$$

where $\Omega$ is the gradient of a given function $F$ at the current point $x$, $\phi$ is the Hessian matrix (the symmetric matrix of second derivatives), $D$ is a diagonal scaling matrix, $\psi$ is a positive scalar, and $\| \ . \ \|$ is the 2-norm.

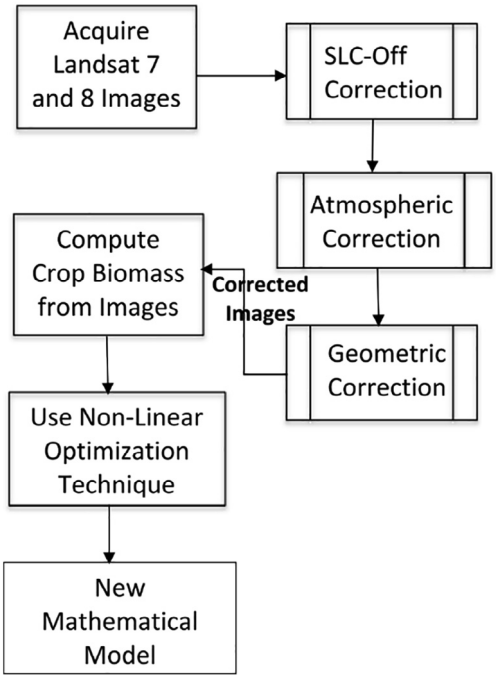


**Fig. 2 – The new mathematical model.**

After calculating the biomass data for each satellite image by using Monteith model, the biomass data with day of the year for each image are used in the optimization process to create the new mathematical model.

At the end a new biomass yield equation is obtained (Eq. (23)) which can compute the total accumulated biomass from the plantation phase until crop maturity phase.

$$NewBio = \sum_{i=1}^{N} Q(x_i) \tag{23}$$

where *NewBio* is the new total estimated biomass in kg/ha, $N$ is the number of days, $x_i$ is the day $i$ during the crop growth progress, and $Q(x_i)$ is the new mathematical model which can be created from the estimated crop yield of the available Landsat image or any other remote sensing images. The advantage of the new method is the ability to show the complete crop growth period even with the lack of complete remote sensing images. Fig. 2 shows in detail the steps needed to create the mathematical model from few Landsat images.

### 2.6. *The intelligent system*

The decision makers, scientists, and famers need a reliable system that can estimate crop yield accurately. The system should be available all the time this means that no cause whether it is natural or manmade should prevent it from providing the needed information.

The system is normally made of different components that interact between each other to complete a specific task. It should be able to take decision and perform specific tasks accordingly. The following schema (Fig. 3) shows the component of the *intelligent system* (IS) and the different tasks it performs. The intelligent system can be considered as a first step toward estimating the yield for all crop types. The role of each component in the *intelligent system* is essential for the success of the process. The IS starts its processes by running the acquisition of data component, then checking the data integrity and completeness.

Then another component recompenses for the missing data using a model based on the optimization of the final

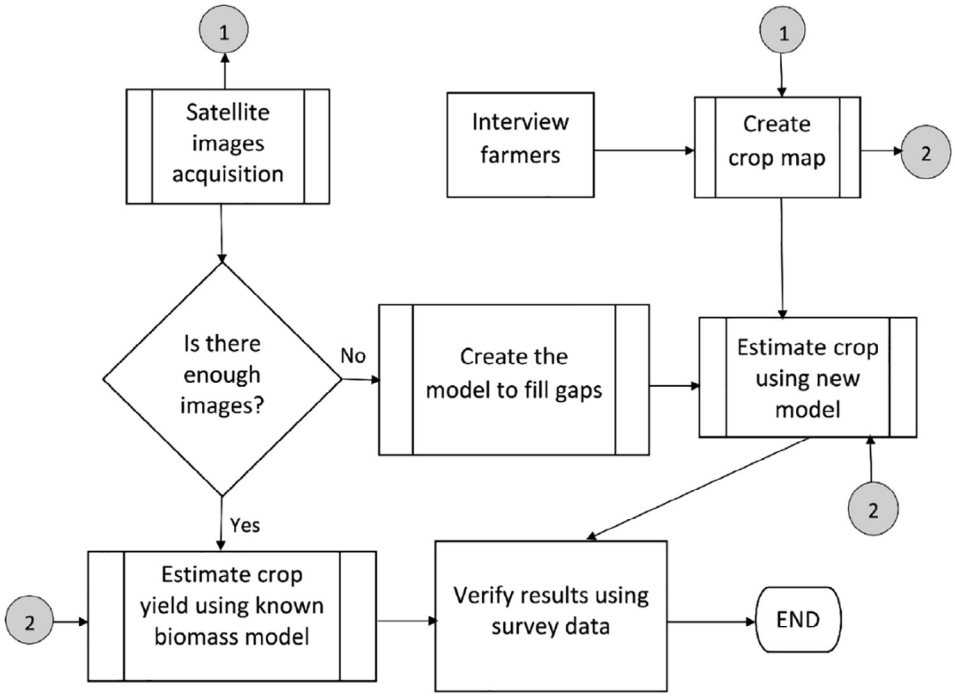


**Fig. 3 – Intelligent system for estimating crop yield.**

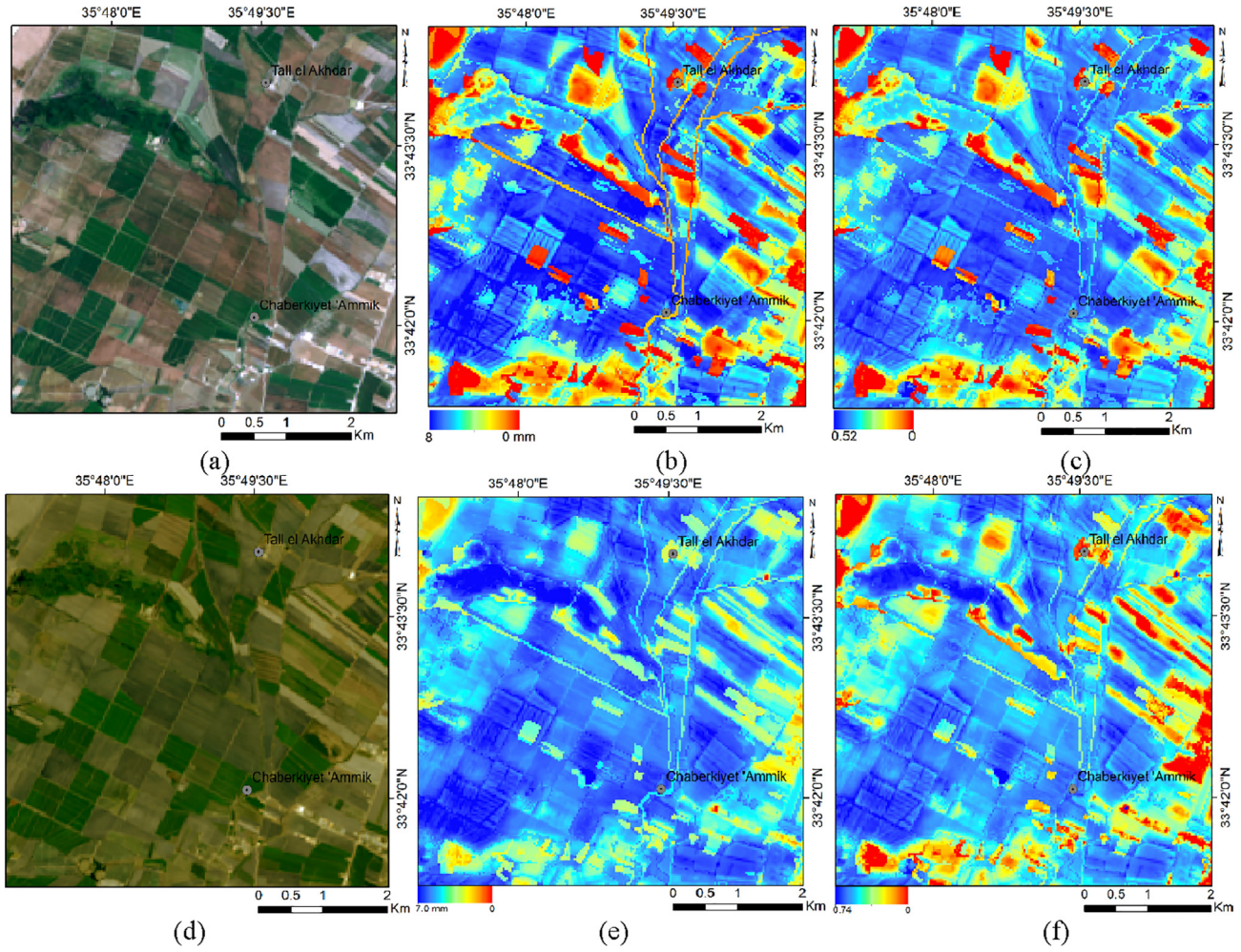


**Fig. 4 – Pilot area in Bekaa Valley 8th and 24th of June (a, d) Original images (b, e) Actual evapotranspiration maps and (c, f) Evaporative fraction maps.**

**Table 1 – Actual ET ($ET_a$) calculated using remote sensing and Bowen station data.**

| Date | Location | Bowen ratio ET (mm) | Remote sensing ETa (mm) | Absolute percentage error |
|---|---|---|---|---|
| 8/6/2015 | Qab-Elias | 7.54 | 6.4 | 15 |
| 8/6/2015 | Taanayel | 5.49 | 4.6 | 16 |
| 24/6/2015 | Qab-Elias | 8.29 | 6.88 | 17 |
| 24/6/2015 | Taanayel | 6.37 | 5.6 | 12 |
| MAPE | | | | 15 |

solution. Finally, estimating crop yield is accomplished by either using known or created method depending on the availability of data.

## 3. Experimental results

The experiments are conducted based on data type, crop type, and season length. In case of any gap in the needed data, the intelligent system can be used to prove its efficiency and robustness estimating crop yield. In this paper it is decided to estimate potato crop yield which has two different cycles, but majority of farmers plant potato in the end of winter and harvest in mid-summer. The potato leafs appears almost 20 days after seeding. Because the weather in Lebanon during winter and spring is mainly rainy and cloudy, it was very hard to get images during the months of March (the date potato was planted) and April. Seven images are collected between the months of April, May, June, and July from the date of potato sowing (20–30 days after) and near the end of the growth stage (10–30 days before harvesting). Landsat 7 was subject to corrections to remove no data stripes as was explained before in data and method section. Moreover,

**Table 2 – Evaporative fraction values for the 8th of June computed using energy balance.**

| Time | Rn W/m2 | G W/m2 | Latent Heat Flux W/m2 | Evaporative fraction |
|---|---|---|---|---|
| 9:00 | 323 | 18.72 | 220.083 | 0.72 |
| 10:00 | 725 | 46.08 | 515.380 | 0.76 |
| 11:00 | 1007 | 91.22 | 718.760 | 0.78 |
| 12:00 | 724 | 116.88 | 482.840 | 0.80 |
| 13:00 | 889 | 112.4 | 621.720 | 0.80 |
| 14:00 | 882 | 105.4 | 624.120 | 0.80 |

**Table 3 – Evaporative fraction values for the 24th of June computed using energy balance.**

| Time | Rn W/m2 | G W/m2 | Latent Heat Flux W/m2 | Evaporative fraction |
|---|---|---|---|---|
| 9 | 681 | 21.82 | 467.78 | 0.71 |
| 10 | 597 | 33.52 | 415.48 | 0.74 |
| 11 | 735 | 52.28 | 517.57 | 0.76 |
| 12 | 1093 | 66.32 | 778.40 | 0.76 |
| 13 | 1055 | 64.5 | 749.57 | 0.76 |
| 14 | 1071 | 55.3 | 767.73 | 0.76 |

at-surface reflectance is calculated by applying atmospheric correction to the at-satellite and at-surface reflectance. Potato map of Bekaa valley for spring 2015 is created using the method in [32].

### 3.1. *Evaporative fraction calculation*

Evaporative fraction and actual evapotranspiration (Eta) maps are created using METRIC model as explained in the method section. The maps covers seven days in April, May, June and July.

The values of the evaporative fraction is less than 1 except in few cases when net radiation is very large compared to sensible and soil heat fluxes ($R_n \gg H$ and $R_n \gg G_0$) or when $H$ is zero because air and surface temperature are equal or the difference is negligible.

To validate the accuracy of METRIC model, some recorded climatic data from two Bowen Ratio stations are used to evaluate actual evapotranspiration and evaporative fraction estimated maps. Fig. 4a–f show the satellite images, the actual evapotranspiration maps and evaporative maps for a large area in Bekaa valley in the 8th and 24th of June 2015.

The following data (Table 1) shows the actual evapotranspiration values obtained from the Bowen Ratio stations

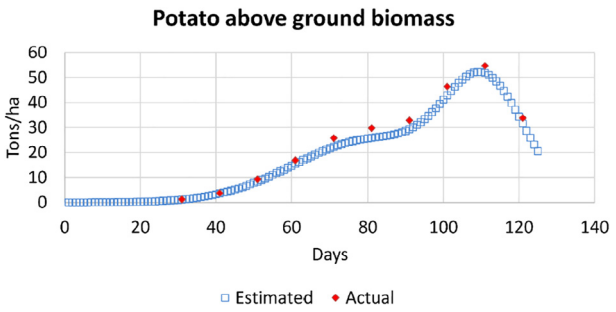


**Fig. 5 – Potato above ground biomass progress graph.**

(actual values) and the predicted values from METRIC model. The accuracy is computed based on the Mean Absolute Percentage Error (MAPE) [33].

Table 2 shows the computed evaporative fraction Wvalues for the local time between 9:00 and 14:00 on the 8th of June 2015. The value of the predicted W = 0.74 computed by METRIC model and compared with the computed one from Bowen Ratio station at time 11:00 shows that the absolute percentage error is 5%.

The estimated evaporative fraction results for day 24 of June 2015 are also verified. The following data collected by Bowen Ratio station (Table 3) shows the different parameter values used by the energy balance equation. The value of the evaporative fraction W computed from data provided by the Bowen Ratio station (highlighted row in Table 3) is compared with the value of W computed by METRIC model (value = 0.7). The result indicates that the absolute percentage error is 8%.

The results are promising ones and it proves that W computed by METRIC model is of high accuracy and can be used to solve crop yield estimation problem.

### 3.2. *Estimating potato crop yield*

To estimate potato yield several Landsat 7 and 8 images are collected. However, the number of images and their temporal coverages are not sufficient because a biomass model requires information about the complete crop growth stage. The problem can be solved if daily satellite images such as Modis are obtained, however the extracted information would be useless due to the coarse spatial resolution of this satellite (from 250 m to 1000 m) especially thermal long infrared image. So the solution is to create a mathematical model that can help in compensate for missing data.

After creating the potato map, some statistical information related to biomass are extracted from the seven

**Table 4 – Estimated potato biomass production for different periods of time in spring 2015.**

| | 10- March | 5-Apr | 7-May | 8-Jun | 16-Jun | 24-Jun | 10-July | 26-July |
|---|---|---|---|---|---|---|---|---|
| *Days* | 1 | 25 | 57 | 79 | 87 | 95 | 111 | 127 |
| *Mean* | 0 | 1.4 | 67.0 | 252.4 | 275.1 | 320.0 | 435.7 | 80.5 |
| *Min* | 0 | 0.5 | 30.0 | 103.8 | 210.0 | 275.0 | 295.4 | 15.3 |
| *Max* | 0 | 5.0 | 140.9 | 285.9 | 307.0 | 374.0 | 583.2 | 179.3 |

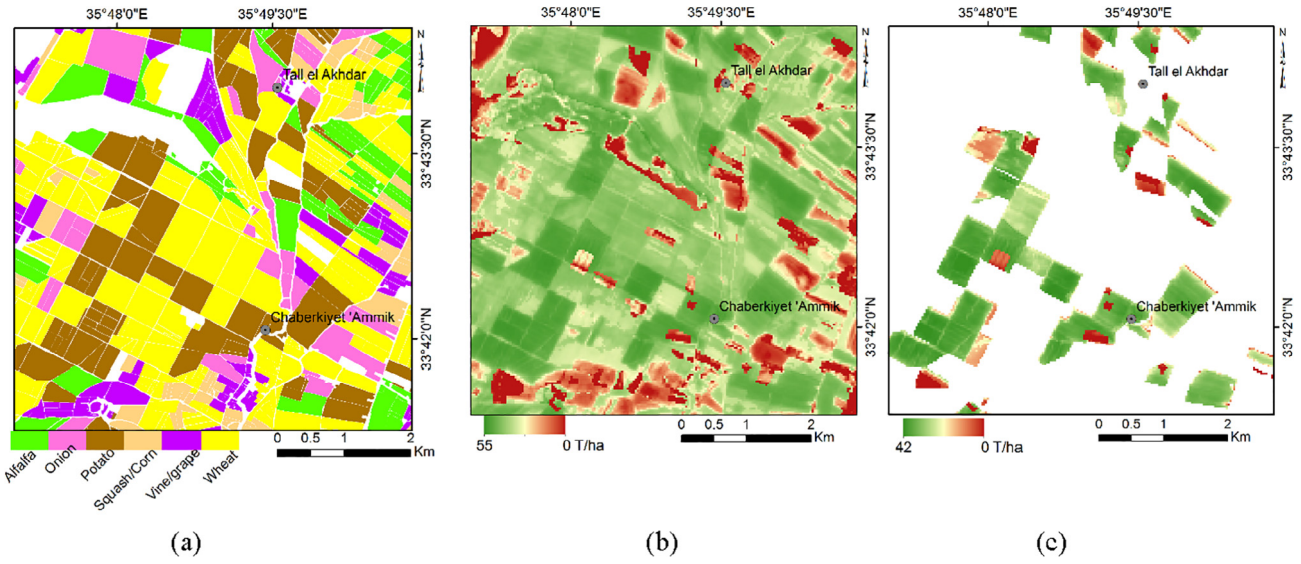


**Fig. 6 – Maps of (a) Selected crops in Bekaa crop (b) above ground crop biomass and (c) Potato tuber.**

processed satellite images. Table 4 lists the *minimum*, *maximum* and *average* biomass production kg/pixel (225 $m^2$) collected from these images for several potato fields. Although the plantation dates of spring potato is variable, but the majority of farmers start planting in different dates of March month especially in the first two weeks. In addition, most of the farmers harvest potato in July. Eq. (24) is the result of fitting a curve the given data in Table 4 using the Trust region algorithm The Parametric fitting of the given data involves finding coefficients (parameters) for one or more models that fit to data.

$$F(Xdata) = a_1 \exp\left(-\left(\frac{Xdata - b_1}{c_1}\right)^2\right) + a_2 \times \exp\left(-\left(\frac{Xdata - b_2}{c_2}\right)^2\right) \quad (24)$$

where $Xdata$ is the day number from the beginning of the crop season to harvesting, $a_1$, $b_1$, $c_1$, $a_2$, $b_2$, and $c_2$ are constants where $a_1 = 489$; $b_1 = 111.1$; $c_1 = 14.92$; $a_2 = 283.1$; $b_2 = 80.96$; $c_2 = 28.64$.

Based on the new model and Eqs. (23) and (24), a new crop yield model is obtained shown in Eq. (25).

$$NewBio = \sum_{i=1}^{N} a_1 \exp\left(-\left(\frac{Xdata_i - b_1}{c_1}\right)^2\right) + a_2 \times \exp\left(-\left(\frac{Xdata_i - b_2}{c_2}\right)^2\right) \quad (25)$$

Fig. 5 shows the progress of the potato biomass after about three weeks from plantation and few weeks before the end of the growth stage. The graph was created using the new yield model (Eq. (25)).

In addition, a map is created using the new crop yield method (Fig. 6a–c). They show the crop map, estimated above ground crop biomass, and potato dry matter (for the whole season of spring 2015). The graph was verified using real biomass data collected periodically from the potato fields after a

**Table 5 – Potato crop yield for several farmers in different areas in Bekaa valley.**

| Farmer # % | Area (ha) | Production (tons/ha) | Type of potato | Estimated (tons/ha) | Error |
|---|---|---|---|---|---|
| 1 | 50 | 30 | Spunta | 28.5 | 0.05 |
| 2 | 50 | 30 | Spunta | 29 | 0.03 |
| 3 | 15 | 35 | Agria | 33 | 0.06 |
| 4 | 75 | 40 | Spunta | 38 | 0.05 |
| 5 | 20 | 40 | Spunta | 38.5 | 0.04 |
| 6 | 10 | 44 | Spunta | 40.5 | 0.04 |
| 7 | 10 | 36 | Spunta | 35 | 0.03 |
| 8 | 50 | 38 | Spunta | 39 | 0.03 |
| 9 | 300 | 25 | Agria | 25 | 0 |
| 10 | 100 | 35 | Agria | 33 | 0.06 |
| 11 | 15 | 30 | Spunta | 31 | 0.03 |
| 12 | 20 | 35 | Agria | 34 | 0.03 |
| 13 | 100 | 25 | Spunta | 24 | 0.04 |
| 14 | 70 | 30 | Agria | 29 | 0.03 |
| 15 | 10 | 15 | Spunta | 14.5 | 0.03 |

month from sowing day and every 10 days. The comparison showed an agreement of about 96% as the graph shows.

To gain confidence in the estimated crop yield results, several potato farmers are interviewed. Table 5 shows the real potato yield and the estimated for several farmers (collected from the map according to each farmer's location). The survey showed that the average accuracy of the estimated crop yield is about 96%.

It is noticeable that there is agreement between the data in the map of Fig. 6c (tuber is normally equal up to 80% of the potato biomass [34]) and the data in Table 5 with respect to having actual values greater or equal to the estimated values. This means that the validation process is correct during the development of the model and after completing it.

In addition, one can notice that the crop yield for some fields is very low this may be due to error in crop mapping.

## 4. Conclusions

The lack of remote sensing data is no longer an obstacle for managers and decision makers of the agriculture sector. It is proved in this research that crop yield estimation can be improved if reliable intelligent system exists to help in overcoming all obstacles which faces the estimation process. The intelligent system is able to make choices whether to directly estimate crop yield or to precede it with a mathematical model that increases remote sensing data availability. The mathematical model is created by using an optimizing algorithm Trust-Region Methods for Nonlinear Minimization that fits available data to an exponential equation. The experimental results proved the accuracy and reliability of the intelligent system in estimating the crop yield in the case where remote sensing data is missing (worst case). The potato crop is used to prove the success of the objective of this research. It is planned to improve the intelligent system by adding more components which can help in managing many agriculture practices and tasks and that can make different agriculture information available to decision makers.

## Conflict of interest

The authors declared that there is no conflict of interest.

## Acknowledgement

The authors would like to thank CNRS for facilitating the achievement of this research as part of a project supported and financed by CEDRE 2018.

## REFERENCES

[1] European Space Agency (ESA). Agriculture overview; 2019. link: https://earth.esa.int/web/guest/earth-topics/agriculture/.
[2] Vogel F, Okidegbe N, Cooke S. Global strategy to improve agricultural and rural statistic. World Bank Econ Sector Work 2010;No:56719-GLB.
[3] Gallego J, Carfagna E, Baruth B. Accuracy, objectivity and efficiency of remote sensing for agricultural statistics. In: Benedetti R, Bee M, Espa G, Piersimoni F, editors. Agricultural survey methods. UK: John Wiley & Sons; 2010. p. 193–211.
[4] Zhiwei J, Jia L, Zhongxin C, Liang S. A review of data assimilation of crop growth simulation based on remote sensing information. In: Proc the third international conference on agro-geoinformatics. Beijing, China; 2014. p. 1–6.
[5] Lobell D. The use of satellite data for crop yield gap analysis. J Field Crops Res 2013;143:56–64.
[6] Devadas R, Denham R, Pringle M. Support vector machine classification of object-based data for crop mapping, using multi-temporal landsat imagery. In: Proc XXII ISPRS congress of remote sensing and spatial information sciences. Melbourne, Australia; 2012. p. 185–90
[7] Estimating Sivarajan S. Yield of irrigated potatoes using aerial and satellite remote sensing Doctor thesis. Utah: University of Utah; 2011.
[8] Petitjean F, Ingladac J, Gancarskia P. Assessing the quality of temporal high-resolution classifications with low-resolution satellite image time series. Int J Rem Sens 2014;35(7):2693–712.
[9] Kasampalis D, Alexandridis T, Deva C, Challinor A, Moshou D, Zalidis G. Contribution of remote sensing on crop models: a review. J Imag MDPI 2018(52):4. online.
[10] Haig S. Crop yield estimation: integrating RS, GIS, management, and land factors. A case study of Birkoor and Kortigiri Mandals-Nizamabad district, India. Master thesis. Netherland: International Institute for Geo-information Science and Earth Observation; 2003.
[11] Prasad A, Chai L, Singh R, Kafatos M. Crop yield estimation model for Iowa using remote sensing and surface parameters. Int J Appl Earth Obs Geoinform 2006;8:26–33.
[12] Bastiaanssen W, Ali S. A new crop estimating model based on satellite measurements applied across the Indus basin, Pakistan. J Agric Ecosyst Environ 2003;94(3):321–40.
[13] Monteith J. Solar radiation and productivity in tropical ecosystems. J Appl Ecol 1972;9:747–66.
[14] Allen R, Tasumi M, Trezza R. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) model. J Irrig Drain Eng ASCE 2007;133(4):380–94.
[15] Van Wart J, Kersebaumb C, Pengc S, Milnera M, Cassmana K. Estimating crop yield potential at regional to national scales. J Field Crops Res 2013;143:34–43.
[16] USGS, EarthExplorer; 2019. link: http://earthexplorer.usgs.gov/.
[17] Moran M, Maas S, Pinter P. Combining remote sensing and modeling for estimating surface evaporation and biomass production. J Rem Sens Rev 1995;12(3–4):335–53.
[18] Field C, Randerson J, Malmstrom C. Global net primary production: combining ecology and remote sensing. J Remote Sens Environ 1995;51(1):74–88.
[19] Bradford B, Hicke A, Lauenroth K. The relative importance of light-use efficiency modifications from environmental conditions and cultivation for estimation of large- scale net primary productivity. J Rem Sens Environ 2005;2005(96):246–55.
[20] Bastiaanssen W, Pelgrum H, Droogers P, de Bruin H, Menenti M. Area-average estimates of evaporation, wetness indicators and top soil moisture during two golden days in EFEDA. J Agric Meteorol 1997;87(2–3):119–37.
[21] Scaramuzza P, Micijevic E, Chander G. SLC gap-filled products phase one methodology; 2018. link: https://landsat.usgs.gov/documents/SLC_Gap_Fill_Methodology.pdf.
[22] HARRIS Geospatial Solutions. ENVI software; 2018. link: https://www.harrisgeospatial.com/SoftwareTechnology/ENVI.aspx/.

[23] Center For Earth Observation (CEO); 2018. How to fill gaps in Landsat ETM images. link: https://yceo.yale.edu/how-fill-gaps-landsat-etm-images/.
[24] Tasumi M, Allen R, Trezza R. At-surface albedo from Landsat and MODIS satellites for use in energy balance studies of evapotranspiration. J Hydrol Eng 2008;13:51–63.
[25] Allen R, Pereira L, Raes D, Smith M. Crop Evapotranspiration guidelines for computing crop water requirements. FAO Irrigation and drainage paper; 1998. No. 56.
[26] Allen R, Walter I, Elliott R, Howell T, Itenfisu D, Jensen M. The ASCE standardized reference evapotranspiration equation. USA: Task Committee on Standardization of Reference Evapotranspiration; 2005.
[27] Bastiaanssen W, Menenti M, Feddes R, Holtslag A. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. J. Hydrology. 1998;212(1–4):198–212.
[28] Bastiaanssen W, Noordman E, Pelgrum H, Davids G, Thoreson B, Allen R. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. J Irrigation Drain Eng ASCE 2005;131(1):85–93.
[29] Allen R, Burnett B, Kramber W, Huntington J, Kjaersgaard J, Kilic A, et al. Automated calibration of the METRIC-landsat evapotranspiration process. J Am Water Res Assoc (JAWRA) 2013;49(3):563–76.
[30] Dennis J, Gay D, Wolf R. An adaptive nonlinear least-squares algorithm. J ACM Trans Math Softw (TOMS) 1981;7(3):348–68.
[31] Waltz A, Morales J, Nocedal J, Orban D. An interior algorithm for nonlinear optimization that combines line search and trust region steps. J Math Program 2006;107(3):391–408.
[32] Awad M. New mathematical models to estimate wheat leaf chlorophyll content based on Artificial Neural Network and remote sensing data. In: Proc. IEEE International Multidisciplinary Conference on Engineering Technology (IMCET). Beirut, Lebanon; 2016.
[33] De Myttenaere A, Golden B, Le Grand B, Rossi F. Using the mean absolute percentage error for regression models. In: Proc 23-th European symposium on artificial neural networks, computational intelligence and machine learning (ESANN). Bruges, Belgium; 2015.
[34] Lizana C, Avila A, Tolabab A, Martinezce J. Field responses of potato to increased temperature during tuber bulking: projection for climate change scenarios, at high-yield environments of Southern Chile. J Agric Forest Meteorol 2017;239:192–201.