

Mathematical Modelling in Crop Production to Predict Crop Yields

Marzhan Anuarbekovna Sadenova^{a,*}, Nail Alikuly Beisekenov^a, Marzhan Rakhymberdina^a, Petar Sabev Varbanov^b, Jiří Jaromír Klemeš^b

^a Center of Excellence «Veritas», D. Serikbayev East Kazakhstan technical university, 19 Serikbayev str. 070000 Ust-Kamenogorsk, Kazakhstan

^b Sustainable Process Integration Laboratory, NETME Centre, FME, Brno University of Technology – VUT Brno, Technická 2896/2, 616 00 Brno, Czech Republic
MSadenova@ektu.kz

In this study, for remote recognition of crops of agroecosystems in Kazakhstan by methods of comparative and historical analogy with the active use of mathematical modelling, the yield indicator of agricultural crops was determined, their dynamic characteristics were studied to predict productivity. The parameters of the dynamic-statistical biomass model were determined separately for each region of the Republic of Kazakhstan based on training data for 21 y (2000 – 2021). The correlation coefficient between the calculated yield values and the official statistics is 0.84. According to the results of cross-validation, the correlation coefficient between the actual and predicted yield of spring wheat was ~ 0.70, which indicates a sufficient resistance of the model to the variability of meteorological conditions for the formation of the crop.

1. Introduction

In the agriculture of the Republic of Kazakhstan, the share of agricultural producers using digital technologies is still not very high, which limits productivity growth and cost savings. Agricultural land is either not used for its intended purpose or is used inefficiently, and it is difficult to control due to the large territory, low population density and the lack of the necessary infrastructure for monitoring the condition and use of land with analysis and forecasting in the short and long term. An effective solution to the issue of monitoring the condition and use of land is the use of various digital technologies, methods of remote sensing of the Earth (ERS) and others. This approach has already been used in the agro-industrial complex of the USA, Canada, the countries of the European Union, India, and Japan. At the present stage, the preparation of statistical data on the soil and vegetation cover, the compilation of maps of fields of agricultural crops, the forecast of yield and other "big data" to increase the efficiency of agricultural production requires the use of the IT, i.e. development of smart agriculture. Ustin et al. (2021) argue that in the next decade, based on "data cubes" compiled for individual and time periods using various visualisations such as hyperspectral images, multiband Thermal Infra Red (TIR) atmosphere sensors and a number of other data, it will be possible to obtain standardised products for solving environmental issues, climate etc. Most agricultural tasks consist of repetitive and standardised activities, that is, suitable for the use of robotics. For example, programmed drones from DroneSeed and BioCarbon can plant seeds correctly. It is known that there is a linear relationship between the analysis of aboveground biomass per square meter and plant height. In addition, on the basis of correlation and regression analyses, after generalising the results of statistical methods for predicting productivity, a correlation was revealed between the characteristics of the environment and productivity. To identify such dependence, methods of linear and nonlinear regression analysis, neural networks are usually used; investigated the dependence of yield on soil composition, soil characteristics, meteorological parameters. Typical input data for modelling plant growth under the influence of meteorological parameters, soil characteristics and fertiliser composition; temperature, precipitation, amount of solar radiation; amount and type of fertilisers, planting density, irrigation and processing parameters, type, depth of the top layer, humus content in the soil. Yang et al. (2021) used a unique dataset that included continuous measurements of corn yield for active leaf fluorescence, vegetation cover Solar-

Induced chlorophyll Fluorescence (SIF), hyperspectral reflectance, and Gross Primary Production (GPP) throughout the growing season. Partial correlation analysis was used to estimate the contribution of Integrated Primary Analysis and Reporting (iPAR), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), and Absorbed Photosynthetically Active Radiation (APAR) to the SIF - GPP ratio at the canopy scale. In parallel, measurements of active fluorescence were used to investigate the distribution of energy in leaves in order to reveal the relationship between fluorescence and photosynthesis at the photochemical level. Rajakal et al. (2021) presented a mathematical approach to forecasting the production of Fresh Fruit Bunches (FFB), including the study of the relationship between plantation maturity and various factors of climate change on the profitability of FFB. This work aims to integrate the cumulative impacts from physiological (oil palm maturity) and climate change factors (rainfall and temperature) into yield forecasting. Most of the known models are applicable to only one type of plant, for example, SOYGRO for legumes, CERES-corn for corn, CERES-Wheat for wheat, WARM for rice. In European countries, the achievements of space technologies in agriculture are widely used, ranging from GPS, which makes it possible to determine the location of equipment, organise parallel driving, control the operation of actuators, to the use of images in the near-infrared range to determine the heterogeneity of crop growth, their further alignment using systems and units for precise fertilisation (Truflyak, 2016). In terms of the spread of precision farming technologies, South America, in particular, Brazil, is experiencing a real "boom". This is mainly due to strong economic growth and the need to reduce production costs. In Brazil, due to the introduction of resource-saving technologies in the agricultural sector of the economy (including precision farming technologies) on 60 % of agricultural land, over the past decade, grain yields have doubled with an increase in sown area by only 11 % and an additional annual income of 10,000 M\$. The precision farming system in Australia continues to improve. In the capital of Western Australia, Perth, a high-precision network based on Trimble NetR5 base stations has been created, which is the first VRS network to receive GNSS signals. The GNSS network supports both the next-generation GPS L2C and L5 signals and GLONASS signals, increasing flexibility, accelerating initialisation and providing more confident signal tracking in positioning tasks. The Industrial Internet of Things allows you to create automated farms with remote control. On the world market of software products, there is a large selection of hardware for collecting, storing, processing and presenting in the GIS system the necessary information about the area and state of crops, fertility zones, crop yields, relief. However, in connection with the development of precision farming technologies, the requirements for geographic information systems are increasing. It is necessary that the software (software) includes the decision-making systems and models necessary to predict yields, taking into account agronomic, climatic and environmental factors. Today, such GIS is less developed; however, there are a number of software products designed to analyse the collected information and calculate fertiliser doses with elements of geographic information systems. The software and equipment installed on the mobile complex allow you to create georeferenced spatial objects that are elements of a geoinformation database for the surveyed field. A well-developed logistics system and e-commerce make it possible to reduce the cost of delivering agricultural products to the final consumer, even for small farms, while maintaining their quality. This is an important factor in the preservation and development of the production of environmentally friendly products, both from the point of view of maintaining the health of the nation and from the point of view of realising the export potential.

The relevance of the research topic is caused by the need to: creation of electronic maps of fields; inventory and detailing of agricultural land; control of the volume and quality of agricultural work; operational monitoring of the state of crops; yield forecast; environmental monitoring; fire safety; other broad opportunities open up when using Unmanned aerial vehicles (UAV) in the agricultural sector of the country, especially in the context of the development of precision farming technology.

A technology was developed for monitoring agricultural land and sown areas using UAVs and Sentinel-2 satellite images. The novelty and uniqueness of this work lie in the development of a self-learning model (machine learning) for modelling the dynamics of the biomass of agricultural crops based on the Monteith equation (Pereiral, 2013).

2. Materials and methods

The modelling of the dynamics of the biomass of agricultural crops is based on the Monteith equation. For its practical use, let us specify the values included in it. The efficiency of using solar radiation for photosynthesis LUE is set in the form of the product of two functions $\eta_1(T)$ and $\eta_2(W)$, which determine the dependence of the intensity of photosynthesis on air temperature T and soil moisture W . The function $\eta_1(T)$ has the form of a one-vertex curve with a maximum value of $\eta_1(T_{opt}) = 1$ in the temperature optimum of photosynthesis T_{opt} , which depends on the type of vegetation. Several functional dependences $\eta_1(T)$ are known that are suitable for different types of vegetation. However, in order not to overload the bio productivity model with too many free parameters that have to be established on the basis of a limited-time series of empirical data, it is used the simplest temperature dependence of photosynthesis, taking into account its most characteristic features:

$$\eta_1(T) = \exp\left\{-a\left(\frac{T - T_{opt}}{10}\right)^2\right\} \quad (1)$$

where a is a positive dimensionless parameter. For the function $\eta_2(W)$, the simplest expression with a single free parameter was also chosen:

$$\eta_2(W) = 1 - \left[\frac{W - W_{opt}}{W_{opt}}\right]^2 \quad (2)$$

where W and W_{opt} are the actual and optimal moisture content of the meter layer of soil. The consumption of biomass for respiration is calculated depending on the air temperature using the following formula

$$R(T) = R_0 \cdot 2^{(T-T_R)/10} \quad (3)$$

where T_R is the basic respiration temperature, exceeding which by 10°C the air temperature is accompanied by a twofold increase in the consumption of organic matter for respiration. Due to the complexity and high cost of field measurements of soil moisture in large areas, data on it, as a rule, refer only to individual experimental sites and are not available online. However, it can be calculated based on standard meteorological parameters (such as temperature, precipitation, air humidity and wind speed) using soil – plant – atmosphere moisture transport models. However, when monitoring the state and development of agricultural crops from space, such modelling must be carried out separately for each pixel of a satellite image, which leads to unreasonably high computational costs. A way out of the situation can be the use of analytical parameterisations of the dependence of soil moisture on standard meteorological parameters. To obtain such parameterisations, representative data from field observations or calculated data from moisture transfer models are required. To calculate the moisture content of a meter layer of soil, analytical parameterizations were used, obtained on the basis of reanalysis data of the European Center for Medium-Range Weather Forecasts ERA-Interim. The initial data are presented for three soil layers (0 – 7 cm, 7 – 28 cm, 28 – 100 cm) with discreteness in time of 1 d and in geographical coordinates of 0.75° . These data were recalculated into the average volumetric moisture content of 1 m layer of soil, taking into account the contribution of each layer, and linearly interpolated to the territory of the Republic of Kazakhstan. The soil moisture values were approximated depending on the standard meteorological parameters by the following expression:

$$W_i = W_{i-1} - c_1 \cdot VPD_i \cdot W_{i-1} + c_2 \cdot P_i + c_3 \cdot VPD_i \quad (4)$$

where i is the ordinal number of the day in the year, VPD is the water vapour pressure deficit in hPa, P is the total amount of precipitation per day in mm, c_1 , c_2 , c_3 are positive definite coefficients depending on the soil and climatic features of the region. The first term in Eq(4) is the moisture reserves in the soil at the beginning of the i th calculation period (day), the second takes into account the decrease in these reserves due to evaporation, and the third and fourth - their increase due to precipitation and condensation of water vapour.

For the use of this equation, it is necessary to set the soil moisture at the beginning of the growing season, which can be taken from the reanalysis data obtained with a delay of about a month from the corresponding date. The calculation of the subsequent values of W is performed daily, depending on the amount of precipitation that fell during that day and the deficit of air humidity. An example of modelling the dynamics of W in the year 2020 for three regions of the Republic of Kazakhstan is shown in Figure 1. If we proceed from the ERA-Interim reanalysis data for the years 2000 and 2021, then the absolute error in calculating the moisture content of a meter layer of soil by the proposed method does not exceed 0.026 (or 2.6 % of the volume). Another meteorological parameter that presents significant difficulties for direct ground-based measurements is the total solar radiation flux (direct and scattered) at the lower boundary of the atmosphere Bottom of Atmosphere (FBOA). Lim et al. (2021) highlight that climate change is affecting crop yields and disrupting the global food system. To calculate it, we use a modified empirical formula from which allows us to indirectly determine the daily average solar radiation flux from the difference between the maximum and minimum surface air temperatures and the degree of atmospheric cloud cover:

$$F_{BOA} = F_{TOA}(t_d) \left[A\sqrt{T_{max} - T_{min}} + B\sqrt{1 - CF} \right] \quad (5)$$

where F_{TOA} (Top of Atmosphere), is the average solar radiation flux entering the upper boundary of the atmosphere per day t_d is the time in days since the beginning of the year, T_{min} and T_{max} are the minima and maxima daytime air temperatures, Cloud Fraction (CF) is the degree of atmospheric cloud cover in the daytime, A and B are empirical constants based on geographic coordinates.

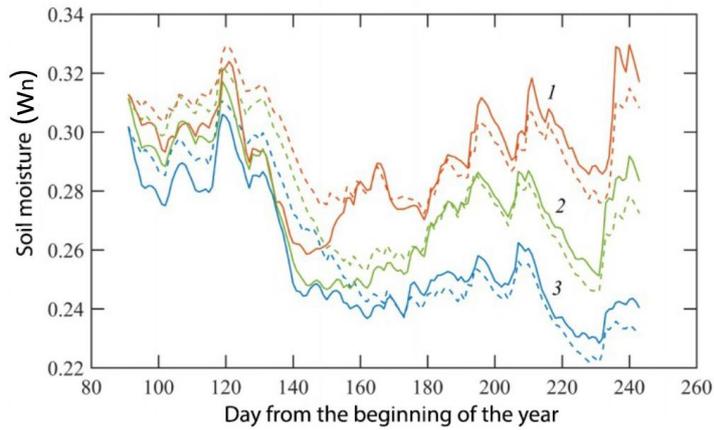


Figure 1: The result of data on modelling the dynamics of soil moisture in East Kazakhstan (1), Pavlodar (2) and Akmola (3) regions of the Republic of Kazakhstan: straight lines - ERA-Interim reanalysis, dotted line - regression model Eq (4)

The F_{TOA} flux is calculated depending on the season and the zenith angle of the Sun:

$$F_{TOA}(t_d) = F_0 \left[1 + 0.033 \cos\left(\frac{360t_d}{365}\right) \right] \langle \cos \theta_s \rangle \quad (6)$$

where $F_0 = 1,367 \text{ W/m}^2$ is the solar constant, $\cos \theta_s$ is the average cosine of the zenith angle of the Sun at a given time of the year. The average value of $\cos \theta_s$ at a given time of the year at a given latitude β is found by the following formula.

$$\overline{\cos \theta_s} = \sin \beta \sin \delta + \cos \beta \cos \delta \frac{\sin \psi^*}{\psi^*} \quad (7)$$

3 Results and discussion

Given the difficulties in identifying arable land in aerospace imagery, a correlation method to isolate informative pixels was used. A US produced spectroradiometer of medium spatial resolution MODIS, which has a fairly long observation series (2000 – 2021), was chosen as a satellite device for monitoring the conditions of crops and forecasting yield. The authors assume that the state of crops is comprehensively characterised by the normalised difference vegetation index NDVI, calculated with a spatial resolution of 500 m by the brightness coefficients of the underlying surface in the red (620 – 670 nm) and near-infrared (841 – 876 nm) channels of the MODIS instrument. Unlike MODIS information products such as Leaf Area Index (LAI) and FAPAR, NDVI determination does not require the use of a priori assumptions about vegetation architecture and radiation transfer models, which makes it less prone to measurement and calculation errors. Due to this, the random component in the NDVI time dynamics associated with inaccuracies in the atmospheric correction of satellite images is much less pronounced than the random component of LAI and FAPAR. The NDVI data obtained by the MODIS device is freely available and is provided in the form of 16 d composite maps, built on all images available during this time, taking into account the following requirements for each pixel: minimum cloud cover, smallest sighting angle, maximum NDVI value. For crop monitoring purposes, NDVI composite maps are used for eight 16 d periods from late March to late July. For the Republic of Kazakhstan, data on the yields of the main agricultural crops are available from the "Experimental Oilseed Farm". Analysis of these data shows the presence of trends in crop yields: from 2018 to 2021 (Figure 1). The detected significant change in the trend component of yield since 2018 indicates exclusively technological reasons for the phenomenon (level of technology, change of varieties, change in fertilizer doses, etc.). The yield component that remains after subtracting the linear trend is mainly determined by meteorological factors. Further, when studying the dependence of yield on various factors, it is the yield component that is determined as the most significant parameter, the forecast of which is the goal of most agrometeorological studies. Correlation coefficients are calculated between the yield of the test crop and the 16-day NDVI values. To select pixels on NDVI maps that contain information about the potential yield of a particular crop. Such correlations were considered separately for each of the fifteen regions of the Republic of Kazakhstan and each of the eight intra-seasonal NDVI composites, taking into account only those pixels that fall within the boundaries of the region under consideration. From the correlation coefficients calculated for each pixel, the maximum modulus coefficient is selected, which refers to one of the 16-day intervals during the growing season. The map of the vegetation

coefficient NDVI and the yield of spring wheat in the "Experimental Oilseed Farm" located in East Kazakhstan is shown in Figure 2, which shows maps for 3 y, in the period for the month of July. Positive values of the correlation coefficients correspond to the period of intensive plant growth accompanied by an increase in NDVI. Negative - the period of earing of spring wheat, when the formation and filling of grain occur, and the NDVI values decrease. The pixels for predicting yield are selected based on the requirement that the corresponding correlation coefficient between NDVI and yield exceeds a certain threshold value. The NDVI values averaged over the mask of such pixels for each area are used to calculate the dynamics of crop biomass in those areas. Figure 2 is presented only as a verification of the field study, collection, and summary analysis of NDVI vegetation data for 3 years.



Figure 2: Data on the state of the experimental site from the Sentinel-2 satellite for 2019-2021 y

The fraction of solar radiation absorbed by the vegetation cover depends on the leaf index, angular foliage distribution, leaf pigment concentrations, soil optical properties, sun zenith angle, and other factors. Their full account in determining FAPAR is possible only with the involvement of laborious field measurements. Nevertheless, the seasonal variation of FAPAR is quite adequately reflected by the NDVI index, which is quite easily determined from measurements from space. The presence of a linear correlation between FAPAR and NDVI has been shown both experimentally and on the basis of modelling the radiation transfer in vegetation. In this regard, when modelling biomass dynamics based on the Monteith equation, we use the assumption that the energy of solar radiation absorbed by vegetation per day is proportional to the product of the flux F_{BOA} and NDVI. Taking into account the assumptions made, the assessment of the yield of agricultural crops on the i^{th} day of the growing season is carried out on the basis of the following equation:

$$Y = \gamma \sum_{i=1}^N F_{TOA} \times NDVI_i \times \eta_1(T_i) \times \eta_2(W_i) \times (1 \cdot (T_i)) \quad (8)$$

where γ is an empirical coefficient that depends on the region and crops. In total, the dynamic-statistical model of yield based on Eqs (1) - (8) includes six unknown parameters: a and T_{OPT} in Eq(1), W_{OPT} in Eq(2), R_0 and T_R in (3), γ in (8). Their specific values are established separately for each region and crops grown in it by comparing the results of biomass modelling with the actual yield in that region. The paper presents a mathematical model of the dynamics of the biomass of agricultural crops, taking into account the main meteorological factors of crop formation and easily adaptable for different regions and crops in the geographic information system of a smart farmer, as shown in Figure 3. To model the dynamics of biomass, satellite measurements of the vegetation index of the underlying surface and standard meteorological observations (temperature, humidity, precipitation and cloudiness).



Figure 3: A geoinformation system for smart agriculture has been developed, which includes many tools for field management

In the calculations of biomass dynamics, only those territories of the region are used for which the vegetation index correlates with the yield of the agricultural crop under consideration, which eliminates the need to identify arable land in satellite images. Modelling biomass dynamics directly at the level of administrative units of the state also minimises the amount of information required to calibrate the model.

3. Conclusions

The method for predicting crop production, described above, models the dynamics of plant biomass. It has been tested on the example of monitoring the process of forming the yield of spring crops grown in the Republic of Kazakhstan. Calculations of the dynamics of the biomass of agricultural crops were carried out with a time resolution of one day based on meteorological data ERA-Interim (temperature, precipitation, air humidity deficit) and data from the MODIS satellite spectroradiometer. Compositional indices NDVI, determined from the MODIS instrument data, were assumed to be unchanged throughout the entire 16 d period covered by them. Modelling the dynamics of biomass begins from the day the first composite image of the NDVI is obtained during the growing season. The end date of the calculation period was selected based on the analysis of the correlation between the calculated biomass and official data on crop yields. The best correlations are achieved at the end of the calculation period at the beginning of July, i.e. at least one month before the start of cleaning. The parameters of the dynamic-statistical biomass model were determined separately for each region of the Republic of Kazakhstan based on training data for 21 y (2000 – 2021). The correlation coefficient between the calculated yield values and the official statistics is 0.84. In view of the small sample size, the Leave-One-Out cross-validation procedure was used to check the quality of the model, based on the sequential exclusion of data for one year from the training sample, calibration of the model against the data for the remaining years, and testing the model on the excluded data. According to the results of cross-validation, the correlation coefficient between the actual and predicted yield of spring wheat was ~ 0.70, which indicates a sufficient resistance of the model to the variability of meteorological conditions for the formation of the crop. Deviations between the actual and calculated values of the yield may be associated with an insufficiently accurate description of the daily meteorological fields by the reanalysis data. In this regard, a further increase in the accuracy of the model is possible due to the use of data from direct meteorological measurements on a network of ground stations for its calibration. However, in this case, there are problems of missing meteorological data and heterogeneity of their coverage of the territory, the elimination of which is a separate and rather difficult task. Meteorological parameters that are difficult to measure, such as the solar flux at the lower boundary of the atmosphere and soil moisture, are calculated based on empirical formulas with coefficients obtained for the region under study from reanalysis data. On the example of "Experimental Oilseed Farm" the applicability of the developed model for predicting the yield of agricultural crops using readily available data from official statistics, Era-Interim reanalysis and NDVI satellite measurements using the MODIS device is shown.

Acknowledgements

This research has been supported by the Project IRN BR10865102 "Development of technologies for remote sensing of the earth (RSE) to improve agricultural management", funded by the Ministry of Agriculture of the Republic of Kazakhstan and the EU project "Sustainable Process Integration Laboratory - SPIL", project No. CZ.02.1.01/0.0/0.0/15_003/0000456 funded by EU as "CZ Operational Programme Research, Development and Education", Priority 1: Strengthening capacity for quality research under a collaboration agreement with D. Serikbayev East Kazakhstan Technical University.

References

- Lim L.Y., Lee C.T., Bong C.P.C., Lim J.S., Ong P.Y., Klemeš J.J., 2021, Selection of Parameters for Soil Quality. Following Compost Application: A Ranking Method, *Chemical Engineering Trans*, 83, 505-510.
- Pereira L.S., Alves I., 2013, Reference Module in Earth Systems and Environmental Sciences: Crop Water Requirements, 6.
- Rajakal J.P., Andiappan V., Wan Y.K., 2021, Mathematical Approach to Forecast Oil Palm Plantation Yield under Climate Change Uncertainties, *Chemical Engineering Transactions*, 83, 115-120.
- Truflyak E.V. The main elements of the system of precision farming. - Krasnodar: KubGAU, 2016. 39.
- Ustin S.L., Middleton E.M., 2021, Current and near-term advances in Earth observation for ecological applications. *Ecological Process*, 10, 1.
- Yang P., van der Tol C., Campbell P.K.E., Middleton E.M., 2021, Unraveling the physical and physiological basis for the solar-induced chlorophyll fluorescence and photosynthesis relationship using continuous leaf and canopy measurements of a corn crop, *Biogeosciences*, 18, 441–465.