PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

Development of models and methods for creating a digital twin of plant within the cyber-physical system for precision farming management

View the article online for updates and enhancements.

Development of models and methods for creating a digital twin of plant within the cyber-physical system for precision farming management

P O Skobelev[1], I V Mayorov[1], E V Simonova[2], O I Goryanin[3], A A Zhilyaev[1], A S Tabachinskiy[1] and V V Yalovenko[4]

[1] Institute of Automation and Information Technologies, Samara State Technical University, 443100, Samara, Molodogvardeyskaya str., 244, Russia
[2] Dept. of Information Systems and Technologies, Samara National Research University, 443086, Samara, Moskovskoye shosse, 34, Russia
[3] Samara Federal Research Scientific Center RAS, Samara Scientific Research Agriculture Institute named after N.M. Tulaykow, 446250, Karl Marx str., 41, Samara region, Bezenchuk, Russia
[4] Peschanokopskaya Agrarian Laboratory, 347561, Rostov region, Peschanokopskiy district, Razvilnoe village, Usadba SKhT str., building 3, office 12, Russia

E-mail: petr.skobelev@gmail.com, simonova@smartsolutions-123.ru

Abstract. The paper discusses creation of a digital twin (DT) of plant for an intelligent cyber-physical system for managing precision farming. A new approach to formalization of DT knowledge is proposed to form expert knowledge within the subject area based on the ontological specification of stages of plant growth and development and multi-agent technology for creating stage agents and coordinated dynamic recalculation of stage duration and yield forecast based on events in the environment. The paper proposes a method for calculating the forecast for duration of plant development stages and yield based on expert knowledge. A "tube" model of the range of changes in parameters of plant development for each stage has been developed. The paper also introduces a method for calculating the yield forecast, as well as the dates of beginning and end for each plant development stage within the "tube" during their normal development and in case of critical situations, for example, frost or drought. Ontology of plant development is constructed for implementation of the "tube" model of environmental parameters, which is converted into a digital form within the ontology editor, available for use by agents. The paper describes the structure and functions of a smart plant DT, built on the basis of a knowledge base and a module for multi-agent planning of plant development stages (for example, wheat), integrated with external weather forecast and fact services. A brief description of the created prototype of the intelligent plant DT system in Java is given. Using the system, agronomists can create their own knowledge bases and DTs of the cultivated plants for each field or even field section. The system will be useful in modern crop production for precision farming, not only "place-wise" but also "time-wise", i.e. in terms of the best time for performing agrotechnical operations.

1. Introduction

The plant growing industry is currently undergoing significant transformation associated with introduction of precision farming technologies [1]. A new trend in precision farming is development and implementation of cyber-physical systems (CPS) for managing plant growth and development. Digital twins (DT) are becoming an important part of CPS, which provide a display of states of real-world objects and the possibility of advanced modeling of their behavior in the future.

Intelligent cyber-physical systems that provide automation of not only physical processes, but also decision-making processes, are already widely used in manufacturing industry within Industry 4.0 [2]. However, in comparison to smart systems for managing machine-building production already available on the market [3], where planning is based on a rigid, tried and tested technological process, in crop production it is usually very difficult to plan work, even with accuracy to the composition and order of operations, which is due to significant lack of knowledge about the life of plants, characterized by high complexity, uncertainty and dynamics, especially caused by climate changes [4].

Plants are characterized by complex processes of growth and development, which are significantly influenced by weather and climatic conditions, biological, chemical, physical processes in the plants themselves and in soils; diseases and pests, and many other factors. At the same time, for example, there are 100 phases of development for wheat (microstages comprising 10 stages) according to the Zadoks scale [5] or the BBCH decimal scale [6].

This paper proposes a new approach to creation of smart digital twins of plants, in which the DT is built as an intelligent system that generates management decisions based on the knowledge base on plant development stages.

Such an intelligent digital twin of plant (IDT) must accumulate and represent knowledge that is currently acquired by agronomists throughout their whole lives, as well as make it possible to model development of a plant under different conditions. This can be done, to begin with, at least at a qualitative level, being the basis for making planning decisions about agrotechnical measures of precision farming. Knowledge about microstages of plant development should help the less qualified agronomists to more accurately model and forecast plant growth and change plans of agrotechnical measures in a timely manner, developing management actions in case of deviations of the actual plant development from the norm, for example, by applying fertilizers.

The paper discusses the principles of building the plant DT, as well as models, methods and specific features of its implementation, which is the basis of an intelligent cyber-physical system (ICPS) for managing precision farming. The first chapter shows the main directions of digitalization in agriculture associated with the start of development of cyber-physical systems for precision farming. The second chapter presents the problem statement for creating a DT in the form of an intelligent decision-support system using a detailed formalized knowledge representation about each stage of plant development. The third chapter provides an overview of existing developments in the use of CPS and DTs. The fourth chapter describes the structure and functions of the plant IDT, based on multi-agent technology and ontological knowledge representation. The fifth chapter proposes a new model for assessing the yield and duration of plant development stages based on the range of change for the most important parameters of plant development at each stage. The sixth chapter examines the method for calculating duration of plant development stage depending on changes in temperature and crop yield. The seventh chapter describes a prototype system for an intelligent plant DT. The eighth chapter discusses the main results and prospects for further development of the system with plant IDT. Finally, the ninth chapter provides conclusions about the advantages of introducing the developed plant IDT as part of ICPS into the daily practice of agronomists.

2. Problem statement

At present, the "digital twin" of a physical object or process is usually understood as a certain set of its virtual (digital) computer models and related processes and/or services, presented in a form that can be integrated with automated systems for implementing the functions of managing, modeling or forecasting behavior of the object in the future [7]. A smart plant DT can be an autonomous intelligent subsystem

that makes it possible to calculate and forecast the yield and the possible start and end dates of each stage of plant growth and development, as well as their duration, based on the expert knowledge of agronomists and incoming data on the weather and phenomenological signs of plant development.

The Plant Intelligent Digital Twin (IDT) is an intelligent computer system that includes the knowledge base and methods for reasoning or decision making. IDT is developed both for online control of plant development in real time and for advanced simulation of plant behavior synchronously with development of the real plant. The main purpose of IDT is forecasting and modeling the process of plant growth in fields for making decisions on agrotechnical measures in order to assess the possible yield and terms of plant vegetation under various external conditions, as well as making management decisions on conducting agrotechnical measures based on the process and fact of plant growth and development within the season, taking into account the forecast of changes in the main environmental factors.

In general, the plant DT can use real data, mathematical or simulation models, machine learning models, and any other methods that can model or predict plant behavior under various conditions in the future. However, the accuracy of these estimations of the plant state directly depends on the adequacy of models used in the DT for describing the real processes of plant development. It is necessary to develop a hybrid approach, at first trying to use the expert knowledge of agronomists who have been observing plant development for many decades. In the future, it is possible to apply neural networks and machine learning methods to combine this knowledge with data collected over many years of observation.

The paper proposes a new principle for creating a digital twin in the form of an intelligent system that uses detailed formalized knowledge representation about each stage of plant development to calculate plant parameters as a control object and develop action measures for future programming of the yield. Based on the classifier of microstages according to the Zadoks and BBCH scales, it becomes possible to construct detailed formalized specifications for each stage in order to plan the stages of plant development in the ICPS depending on changes in the environment.

Using the system, agronomists of different farms can create their own knowledge base and DT of the cultivated plants for each field or even field sections. The system will be useful in modern crop production for precision farming, not only "place-wise" but also "time-wise", i.e. in terms of the best time for performing agrotechnical operations.

3. State of the Art

Let us analyze some papers on the use of various approaches to creating a plant DT.

[8] describes a project of the Lawrence Berkeley National Laboratory (Berkeley Lab, US) on development of software based on machine learning methods for integrating ecological databases of Department of Energy with databases for monitoring and remote sensing of fields. This software is used to study soil-plant interactions and other ecosystem properties on a variety of scales, from farm to state. Farmers are provided with a means of assessing sprout quality and biomass density using UAV (unmanned aerial vehicles) images. These estimates make it possible to make real-time adjustments to farmers' decisions for improving productivity.

[9] describes the results of research based on machine learning methods - One Class Classification, One Class Support Vector Machines, Hierarchical Self Organizing Maps and Active learning algorithms, as well as their application in precision farming. Hierarchical Self Organizing Map derived models using supervised learning and novelty detection methodology make it possible to classify plant reflectance data for identifying different types of weeds and differentiating them from Zea mays (corn). The One Class Classification approach consists of a learning procedure based on analysis of various plant tissue samples belonging to four crop conditions: healthy or infected with one of the three types of diseases. The presented application is capable of evaluating images obtained in real conditions without additional processing and can reliably determine an infection even at an early stage using an unlabeled feature vector as input data to One Class Classification. This study makes it possible to identify leaf diseases in several crops, although the training selection consisted of samples of vine leaves.

[9] also describes Active Learning Algorithms that combine high-resolution multilayer soil and crop data, creating a sensor fusion neural network model that will learn to estimate the geospatial location of wheat crops with higher productivity than other modern methods. Thus, it will promote crop yield growth.

[10] decribes a model using multilayer perceptron ANN technology for evaluating the biomass of Mentha crops based on spectral images of the Landsat 8 OLI (Operational Land Imager) satellite and derived spectral indices together with field data.

[11] dwells on application of digital twins for management of water resources of a farm via the Smart Water Management Platform running under IoT (Internet of Things). The digital twin uses the Penman-Monteith equation to calculate the reference evapotranspiration (ET) based on soil moisture in the fields, measured by a soil probe, and forecast data from a weather station.

[12] presents an empirical spatial model for assessing the spatial variability of the vine phenology at the field scale using a single measurement of a given plant parameter, performed in the field within a reference plot. The model is based on a system of mathematical equations with a combination of plot-specific coefficients calculated based on historical information.

From the point of view of creating a DT of wheat, [13] should be of specific interest as it presents modeling wheat root system and realistic visualization of root growth in various soil conditions. Based on topology of wheat root system, as well as historical and experimental data, a morphological model of wheat root was preliminarily developed, including submodels of root emergence, root growth speed and curvature of the root axis. Based on technology of 3D visualization, a visualization model for root axis of wheat was developed using the Visual C+ (.Net) platform and the OpenGL library. As a result of integration of root morphological model and visualization model, a 3D visualization of wheat root system under various soil conditions was implemented.

This approach in the future makes it possible to create an intelligent system for combining models of individual parts of a plant (root, leaves, stem, etc.). However, this topic is not considered in the paper.

Based on the analysis performed, the following conclusions can be drawn. Systems of mathematical equations are used to connect individual parameters of plant growth and development with yield productivity without taking into account the process of plant development at each stage. For systems of mathematical equations for each stage, in addition to the accuracy of universal mathematical description of each microstage, a system of systems of equations is needed to describe the rules for linking different microstages to each other, which would significantly increase the complexity of the model for practical use.

When using machine learning models, a test selection is required, which has to be obtained under certain unchanged conditions. This is of course quite difficult due to global climate changes. In addition, when there are heterogeneous conditions of data collection, reconciliation procedures are needed to enable comparison of such observations, which will complicate and slow down the data analysis process. At the same time, the proposed approaches can be used to calibrate expert decision-making models.

The review also showed that, although work on creation of plant DTs is already underway, they do not take into account knowledge about stages of plant growth and development and effect of results of one stage on other stages. Novelty, relevance and significance of the proposed IDT lie in formalization of knowledge about each stage of plant development in terms of identifying its main parameters and building rules for transition between stages, taking into account environment parameters and based on the proposed methods for calculating the yield forecast and duration of stages developed during collection and analysis of expert knowledge of agronomists.

4. Structure and functions of intelligent digital twin of plant

Based on the knowledge about stages of wheat growth and development, it is proposed to create IDT for virtual sowing in each field, which will mirror growth and development of a plant in the field or even in its individual sections for precision farming, taking into account the data coming from online weather services and from field inspections by agronomists. If DT is synchronized with a real plant, i.e. it can adequately reflect its state, for example, through regular inspections of real plants by agronomists, it can

be used by agronomists to develop and make management decisions on carrying out agrotechnical measures based on planning and modeling of possible problem situations and finding ways to resolve them.

The presence of plant DT in each field, and even in its individual "cells", makes it possible not only to build a forecast of yield and timing of stages, but also to model situations ("What if") in order to justify for the business the developed decisions on agrotechnical measures, for example, on volumes of fertilizer application, since this is usually quite costly, requiring loans that can only be returned after selling the crop.

Figure 1 presents the enlarged principle of creating IDT, based on two main modules - a knowledge base about the plant and a multi-agent system (MAS) for planning its stage-by-stage development.
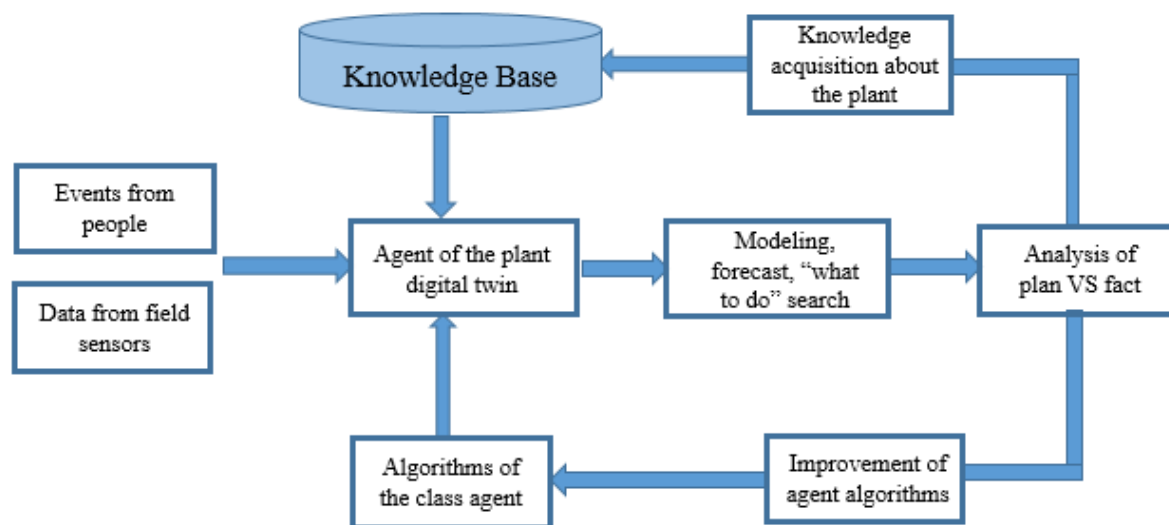

Figure 1. General scheme of a plant DT.

IDT is built using new technologies for formalization of domain knowledge [14] and contains a knowledge base with ontology for formalizing and accumulating this knowledge and its subsequent use in calculations to make forecasts for timing of plant development and yield, which becomes critically important for introduction of precision farming.

Ontology describes the basic classes of concepts and relations within the subject area ("explanatory dictionary" for agents). At the top level, the digital twin ontology is defined, consisting of the most general and reusable classes of concepts and relations. Domain-specific details are described in a specialized crop ontology. The key concepts and relations of this ontology are inherited from the basic ones defined in the digital twin ontology. The presence of such a connection allows agents to interpret a new subject area in terms of an already known problem and use existing protocols and procedures to solve it.

Figure 2 shows a fragment of the plant DT ontology. The key concept here is an "order" (sowing) for receiving a "product" (crop) for which the plant needs to complete a sequence of certain "tasks" (stages). For tasks, rules are defined for calculating the duration and achieved values of product parameters (for example, for evaluating the yield). Implementation of these rules depends on environment parameters (temperature, humidity, etc.), data on which are provided by third-party services. The environment parameters are tied to certain "locations" (fields) on which "orders" are placed, i.e. each field and crop would have its own external environment.

In addition, the system uses multi-agent technology [15] and contains a multi-agent system for planning stages of plant development, in which each stage corresponds to its own software agent, which actually calculates the duration of each stage depending on the input data transmitted from the previous

stage. It also transmits the new forecast and calculation to agents of subsequent stages for further recalculation along the chain, taking into account the known weather forecast and other parameters.

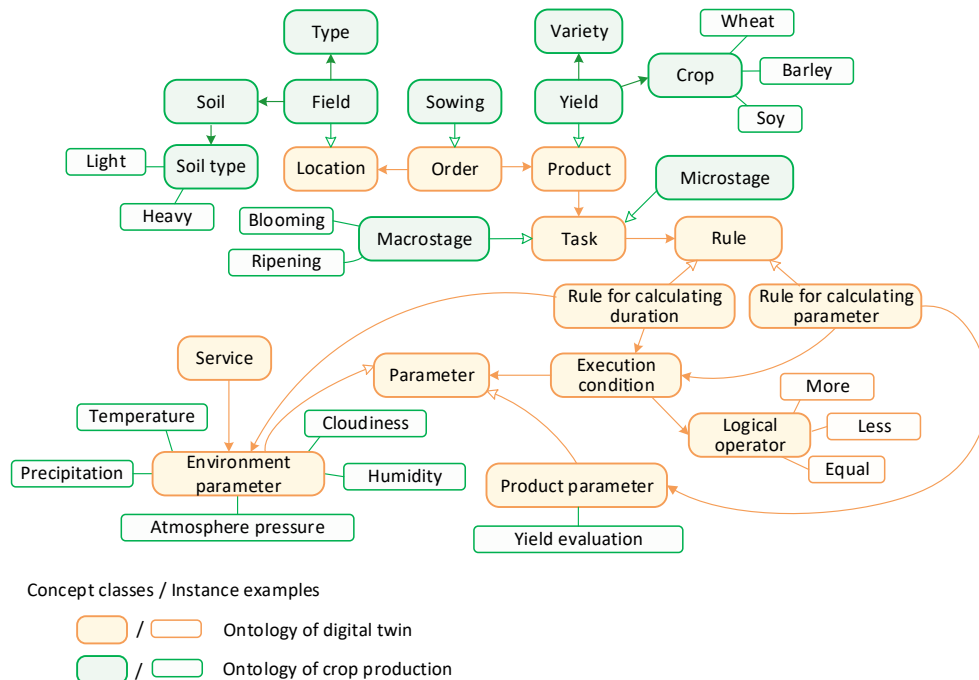A more detailed description of plant IDT as part of the ICPS is given in [16, 17].

Figure 2. Fragment of plant DT ontology.

The input data for the system include data on the weather forecast and fact, data on the plant environment, as well as phenological parameters of the state of the real plant. For this purpose, both external web services and results of plant inspections in fields by agronomists are used. Plant stage knowledge is loaded from the knowledge base.

Based on these data, for each field of the farm and for each stage of plant development, an instance of the stage software agent is created (cloned), which will calculate the yield forecast and duration of each specific stage, and transmit the results to the next agent in the chain. At each stage, agents can also check weather services and request data on field inspections or soil conditions from agronomists (figure 3).
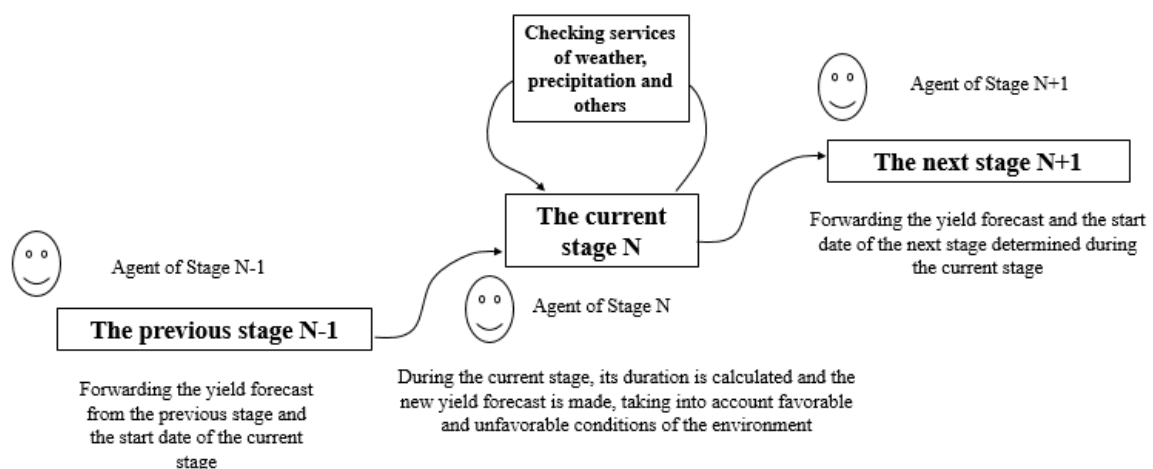

Figure 3. Representation of each stage of plant development using agents.

Based on these data, the system recalculates yield forecast, beginning and end dates for the current stage, as well as its duration. Due to the use of multi-agent technology, agents make calculations of all stages in parallel and asynchronously, which makes it possible to enter events at any stage and maintain an automatic "wave" chain of recalculations in response to any event. A plant development plan with a new yield forecast, rebuilt by MAS of stepwise planning based on data from the agronomist for each event, should allow the agronomist to timely identify dangerous deviations from the norm and develop solutions to counter threats of crop loss.

Based on the concepts and relations of ontology as an "explanatory dictionary" for agents, a model for assessing the yield and duration of plant development stages is proposed.

5. Model for assessing yield and duration of plant development stages

Precise mathematical models of plant growth and development have not yet been created. However, agronomists, throughout their lives, accumulate a lot of expert knowledge, observations and conclusions regarding the cultivated plant varieties. This knowledge is not formalized, it is usually stored as paper records, and is in no way suitable for automatic calculations and decision-making.

In the developed intelligent DT system, it is proposed to introduce the concept of a working "tube" for changing the most important parameters of plant development as the basis for building a knowledge base of domain expert knowledge in plant growing, meaning the ideal (nominal) range of change for each parameter (figure 4).
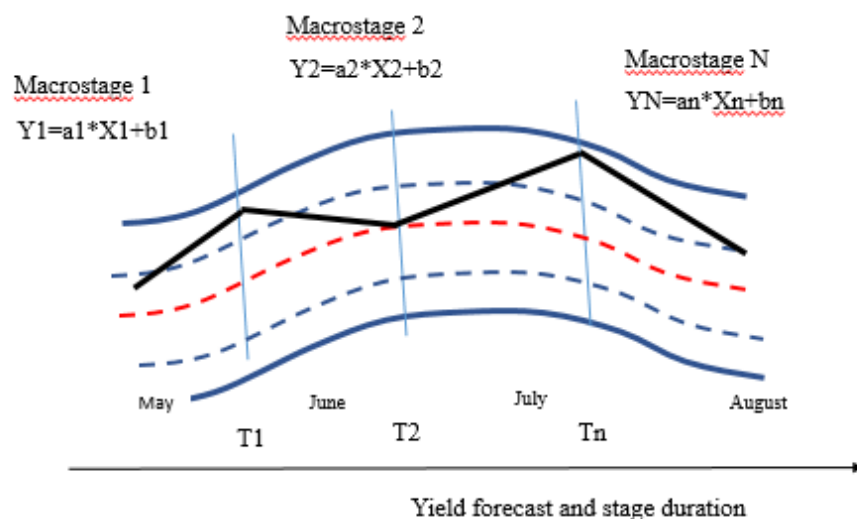

Figure 4. Model of the range of changes in plant parameters ("tube").

The growing season is considered as 240-350 days, but it takes about 70-80 days from spring awakening to heading in various climatic conditions of the country. In figure 4, the red dashed line shows the ideal (nominal) "trajectory" of change in parameters, most preferable for a plant of a given variety. Such a parameter can be, for example, air temperature which guarantees maximum yield and optimal growing season with other ideal values of important parameters (precipitation, soil characteristics, etc.). Blue dashed lines indicate range boundaries of the desired (preferred) trajectory of change of an important parameter for the process of plant development.

Within the study, it is proposed to adopt a linear model of plant growth and development at each stage in order to build a knowledge base for wheat. This means that deviation of a parameter from the ideal would cause a linear (proportional) change of yield and timing of plant stages in one direction or another. For example, for the parameter of accumulated temperature, if the value is less than the ideal, the yield forecast and the rate of plant development will decrease, and vice versa, if it is higher, they will increase. Consequently, the stage duration also changes, either decreasing or increasing.

It is proposed to use linear dependences of different types at different stages of plant development, which is shown by the black line in figure 4. In other words, if the current accumulated temperature is 10% lower than the ideal one, the entire development of the plant will be 10% worse (it will ripen longer and the yield will be less). For some parameters at different stages, changes in values in the range shown by dashed lines may not at all affect the yield and timing of stages, but this is a special case.

The main blue lines show the range of acceptable limit values for each parameter. If the values go beyond this limit, the plant will, as a rule, die and the crop will be lost. In some cases, it will be possible to indicate a nonlinear effect using an expert rule for a parameter that has gone beyond the permissible values. The type of expert rule is "If A and B, then C", where A and B are examples of conditions, and C can be either the probability of plant survival with reduced characteristics, or prediction of complete death of the plant or cessation of growth. Examples of expert rules are the following: "in case of severe frost at the moment of coleoptile germination, the yield will be completely lost"; "if at the stage of first sprouts the temperature is -15°C for a week, we can confidently assume that the plant will die".

For example, table 1 shows the temperature range for wheat cultivation in the Samara region of Russia (this information was studied and formalized on the basis of monographs and reference books).

Table 1. Parameters of the "tube" model for wheat in the Samara region of Russia.

Symbol	BBCH 0	BBCH 1	BBCH 2	BBCH 3	BBCH 4	BBCH 5	BBCH 6	BBCH 7	BBCH 8	BBCH 9
S_y, °C	110	190	100	300	300	100	100	450	150	100
T_{max}, °C	42	42	40	40	40	40	40	40	42	42
T_{min}, °C	-12	-12	-8	-7	-4	-1	5	5	5	5
y_1, °C	15	15	14	12	15	18	20	22	24	25
y_2, °C	15	14	10	15	18	20	22	24	25	25
x_2-x_1, days	7	17	53	13	15	5	5	22	7	6

Here, S_Y is the sum of accumulated temperatures of the stage, T_{max} and T_{min} are boundaries of the "tube", outside of which the expert rule is fulfilled, y_1 and y_2 are the ideal temperature line, in the area of which there is no noticeable change in the stage duration.

We believe that when it reaches the limits of the "tube", the plant growth stops (in the future, expert agronomist rules for such situations will be available). It is assumed that the specified expert rules and model coefficients at the initial stage will be set empirically by an agronomist, taking into account specific features of the region and characteristics of plant varieties, fields and other parameters. However, as actual data accumulates, they can be calibrated using neural networks or other machine learning methods.

6. Methods for calculating duration of plant development stage and crop yield
Based on the model of the range of changes in plant parameters, the following methods have been developed: a method for calculating the duration of plant development stage depending on temperature changes and a method for calculating the crop yield.

6.1. Method for calculating the stage duration depending on temperature changes
For the main 10 stages of wheat development according to BBCH, the sum of accumulated positive temperatures is considered as the defining parameter. On the basis of a survey of experts, a model of the "tube" of temperature range has been constructed. When the values exceed the tube, the model of change is replaced by an expert rule of decision making, possibly without moving to a new stage in the event of crop failure. The range is considered between + 5°C and + 35°C. Duration of the stage is counted in days from its beginning (day 0). The optimal sum of accumulated temperatures S (degrees*day) for this stage is assumed to be known.

6.1.1. Linear temperature change function. Duration of the stages varies from days to dozens of days; therefore, we assume that at each stage there is a linear temperature function that increases in spring and decreases in autumn (the ideal temperature *y* for each stage). Conventionally, two ordinates of this function are set through the upper and lower "optimal" temperatures, within which the necessary accumulated temperature was calculated. In case of simplification or unavailability of data, the accumulated temperature can be replaced by the average temperature in the region within the given month (figure 4). The temperature is calculated using the equation (1).

$$y = kx + b, \tag{1}$$

where *y* is the ideal temperature, *x* is the number of days from the beginning of the stage, *y1* is the temperature at the beginning of the stage, *y2* is the temperature at the end of the stage, $k=(y_2-y_1)/x_2$, $b=y_1$. Substituting these symbols in (1), we get the equation (2) for plotting the blue graph in figure 5.

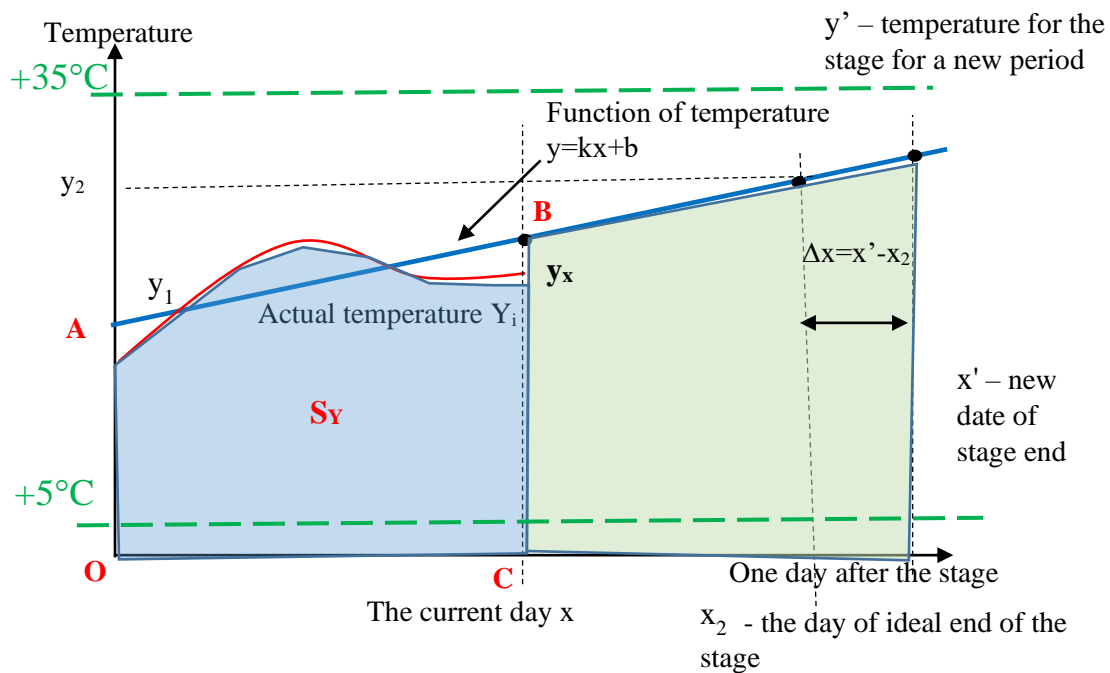$$y = \frac{(y_2 - y_1)}{x_2}x + y_1 \tag{2}$$

Figure 5. Method for calculating the end of the plant development stage.

Every day (or, in general, within irregular time intervals) the temperature is measured or the weather forecast *Yi* is available up to the current day *x* (daily readings on the red line in figure 5). The green dashed lines show the temperature "tube", beyond which the model is replaced by the expert rules for changing stage conditions. It is necessary to define a new date *x'* for the end of the stage. The sum of accumulated temperatures (3) is used. In this case, the interval between measurements is 1 day.

$$S_Y = \sum_i Y_i \cdot \delta x_i \tag{3}$$

If the area under the real temperature graph (S_Y), indicated in red, is less than the area under the linear function up to the value x, the end date of the stage shifts to the right with relation to the day of the ideal end x_2 as shown in figure 5. And vice versa, it shifts to the left if it is bigger. The final sum of temperatures should be equal to the set sum of temperatures S for the stage (polygon OABC in figure 5) and is calculated according to the equation (4).

$$S_0 = \frac{(y_2 + y_1)}{2} x_2 \qquad (4)$$

We assume that the actual temperature points arrive quite often, and the changes with respect to the temperature line are insignificant (and, consequently, the time shift is not significant either). Then the shift of the stage boundary is expressed as $\Delta x = x' - x_2$. The accumulated temperature increment must be equal to the difference between the values S_x and S_Y (5).

$$\frac{(y_2 + y')}{2} \cdot \Delta x = S_X - S_Y \qquad (5)$$

Thus, to determine Δx, we get a linear equation (6) with respect to Δx (the total sum of accumulated temperatures S_0 enters indirectly through the parameters y_1, y_2, x_2).

$$\Delta x \cdot y_2 \approx y_1 x + \frac{1}{2}\left(\frac{y_2 - y_1}{x_2}\right) \cdot x^2 - S_Y \qquad (6)$$

If the temperature line is parallel to the time axis, then equal temperature values are set throughout the entire stage: $y_1 = y_2 = y_x = y'$. Equation (5) is simplified, and the new stage boundary is determined according to the equation (7).

$$x' = x_2 + \frac{S_0 - S_Y}{y_1} \qquad (7)$$

The performed mathematical reasoning helps propose a practical calculation method that can be used by agents at each stage of plant development.

6.1.2. Examples of calculating the duration of stages. Let us consider the "Germination" (BBCH 0) stage of winter wheat in the Samara region of Russia in September.

Example 1. $S_0 = 105$ °C, $y_1 = 17$ °C, $y_2 = 12$ °C, $x_2 = 7$ days. The temperature line $y = (12-17)/7*x+12 = -0.714*x+12$. Let us suppose that the Y_i readings for x=4 days were 10°C, 10°C, 10°C, 10°C (below the temperature line). Then $y_x = -0.71*4+12=-2.85+12=9.15$, $S_Y=40$ °C.

From equation (6) it follows: $12*\Delta x=17*4+0.5(-5/7)*16-40=28-8*0.714=28-3.57=24.4$, $\Delta x=2$ days. Consequently, the end date of the stage in this case will be $7+2 = 9$ days from the beginning of the stage (compared to the initial duration of 7 days).

Example 2. $S_0 = 105$ °C, $y_1 = 17$ °C, $y_2 = 12$ °C, $x_2 = 7$ days. The temperature line $y = (12-17)/7*x+12 = -0.714*x+12$. Let us suppose that the Y_i readings for x=4 days were 20°C, 20°C, 20°C, 20°C (above the temperature line). Then $y_x = -0.71*4+12=-2.85+12=9.15$, $S_Y =80$ °C.

From equation (6) it follows: $12*\Delta x=17*4+0.5(-5/7)*16-80=28-8*0.714=-12-8*0.714=-17.7$, $\Delta x=-1.5$ days. Consequently, the end date of the stage in this case will be $7-1.5 = 5.5$ days from the beginning of the stage (compared to the initial duration of 7 days).

The calculations performed confirm practical adequacy of the model.

6.2. Method for assessing crop yield

It is necessary to develop a simple model of effect of a complex of parameters (temperature, humidity, precipitation) on the yield. Since all these data and estimations are inaccurate, fluctuating depending on the region and accounting methods, there is no fundamental possibility to build an accurate model. The same plants grown under similar conditions give different yields, hardly with more than 10% accuracy. In addition, it can be difficult to take into account the influence of external conditions. Since there can be many parameters of external environment, it is required to create a consistent, unambiguous way of accounting for them and mutual joint influence on the output parameter - yield.

Let us introduce the following simplifications. Empirical rules show that at the given stage there is a "tube" of parameter values in which development takes place, and a certain range that is considered ideal for the stage, for example, for the "Heading" stage (BBCH 5), plant development occurs from +5°C to +35°C, and the optimum temperature is from +15°C to +20°C. Below and above the "tube", the yield falls to 90% (above the upper limit) and to 80% (from 0°C to +5°C), however, all these readings are theoretical and conditional. It is proposed to take the parameter values as a basis and build a linear model in a piecewise-linear way. The influence on the yield will be assessed using the "satisfaction function". We believe that there is an ideal yield, and all deviations in ranges lead to its decrease. It is only possible to counter the yield drop by using some agricultural technology, for example, by applying top dressing, and this will lead to a drop not by 80%, but by 70%. Application of specific rules is not considered in this model.

Let us consider the satisfaction function using the example of one particular parameter - temperature. The four-point satisfaction function $(y_1, x_1) \dots (y_4, x_4)$ fits the linear deviation model. The ordinate axis is satisfaction $(0, 1)$, the abscissa axis is the parameter, in this case, temperature. The graph of satisfaction function has the form of a trapezium (figure 6): (y_1, x_1) is the left border of the range (to the left of the border, the rules are applied, not the model), $\dots (y_4, x_4)$ is the right border of the model, and $(y_2, x_2) \dots (y_3, x_3)$ is the optimality plateau. If there is no plateau, then the second and third points coincide, and the trapezium turns into a triangle. We interpret the satisfaction function as the yield.
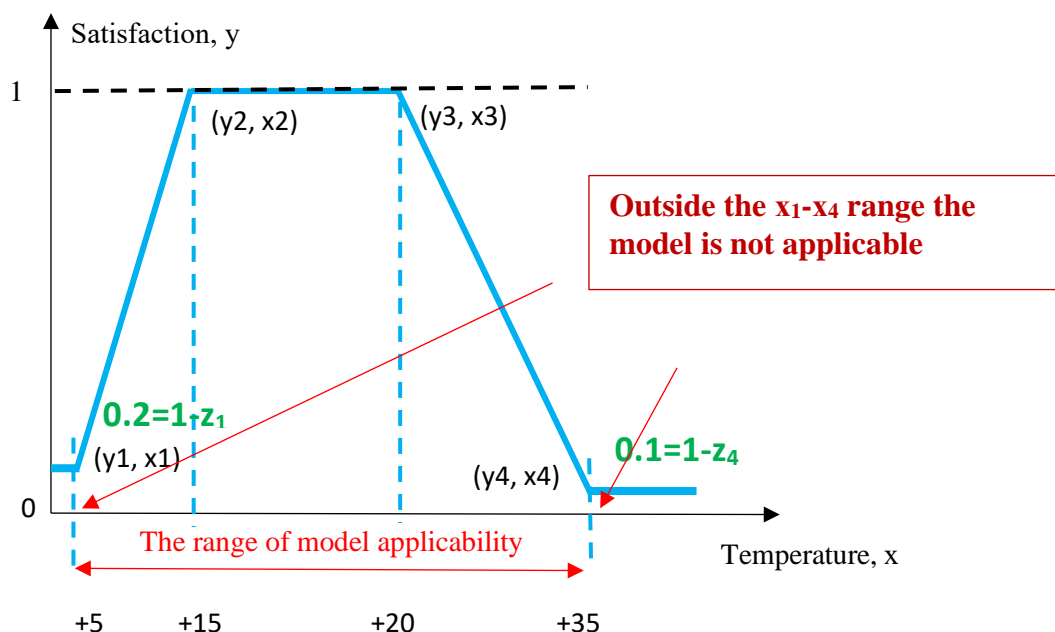

Figure 6. Satisfaction function as the function of yield.

The drop in yield will be calculated when plotting points 1 and 4 (by subtracting from one). In this case, satisfaction is equal to yield. Parameters for assessing the yield are given in table 2.

Table 2. Parameters for assessing the yield.

Parameter	Symbol	Value
The left border of the range	x1	+5 °C
The left optimality point	x2	+15 °C
The right optimality point	x3	+20 °C
The right boarder of the range	x4	+35 °C
The left drop in yield	z1	0.8
The right drop in yield	z4	0.9
The left yield point	y1=1-z1	0.2
The left point of optimal yield	y2	1
The right point of optimal yield	y3	1
The right yield point	y4=1-z4	0.1

Rules for calculating the yield are given in table 3.

Table 3. Rules for calculating the yield.

Number	Condition	Rule
1	$0<x<x1$	y=y1 and other rules
2	$x1<x<x2$	y=k1 *(x-x1)+ y1 k1=(y2-y1)/(x2-x1)
3	$x2<x<x3$ y=1	no drop of yield
4	$x3<x< x4$	y=k2*(x-x3)+y3 k2=(y4-y3)/(x4-x3)
5	$x>x4$	y=y4 and other rules

6.3. Selection of the most important parameters of the model for determining the "tube" ranges and developing expert rules for plant development

During the study, the following main significant abiotic factors have been identified, which were classified according to degree of influence on plant growth and development (table 4). These factors can be entered into the model as parameters.

Table 4. Model parameters.

Degree of Importance	Parameters
Parameters of paramount importance	Air temperature
	Air humidity
	Soil moisture
	Hydrothermic coefficient
Additional parameters	Soil temperature – can be measured using weather stations or determined indirectly through air temperature
	Plant nutrition - can be measured indirectly by leaf diagnostics: black soil, humus, mobile phosphate, exchangeable potassium, microelements, zinc, sulphur, etc.
	Granulometric composition and soil density - does not change during the season, is measured before sowing (heavy, light, etc.).
Small influence parameters	Sun radiation
	Others

To create a prototype of the system of a plant's intelligent digital twin, the following parameters are taken as initial, affecting the course of plant development: the sum of active temperatures, daily temperature and soil moisture (volume of precipitation).

7. Description of the prototype of plant IDT

To implement the proposed approach, a software package has been developed, which includes an ontology editor, a digital twin editor and a multi-agent planning module (figure 7). To work with the components, a graphical user interface has been implemented that provides data entry about the digital twin, entry of events and viewing of created plans. Interaction of components is based on the message bus.
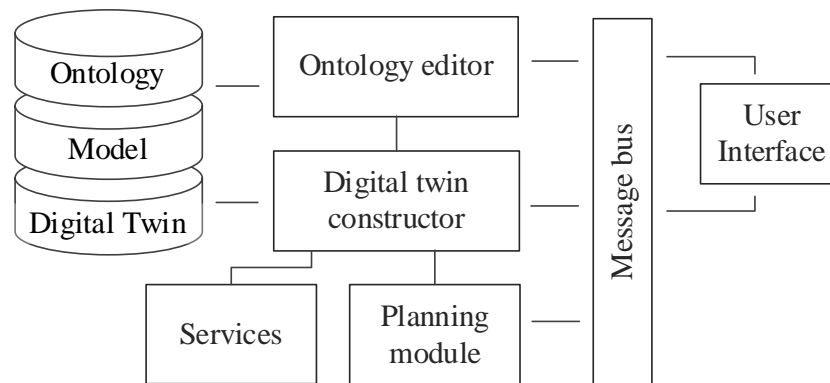

Figure 7. Architecture of the software package.

Ontology editor allows for viewing and modifying the crop knowledge base, where all the properties could be specified and added. The content of the knowledge base is presented in a table and semantic form (figure 8).
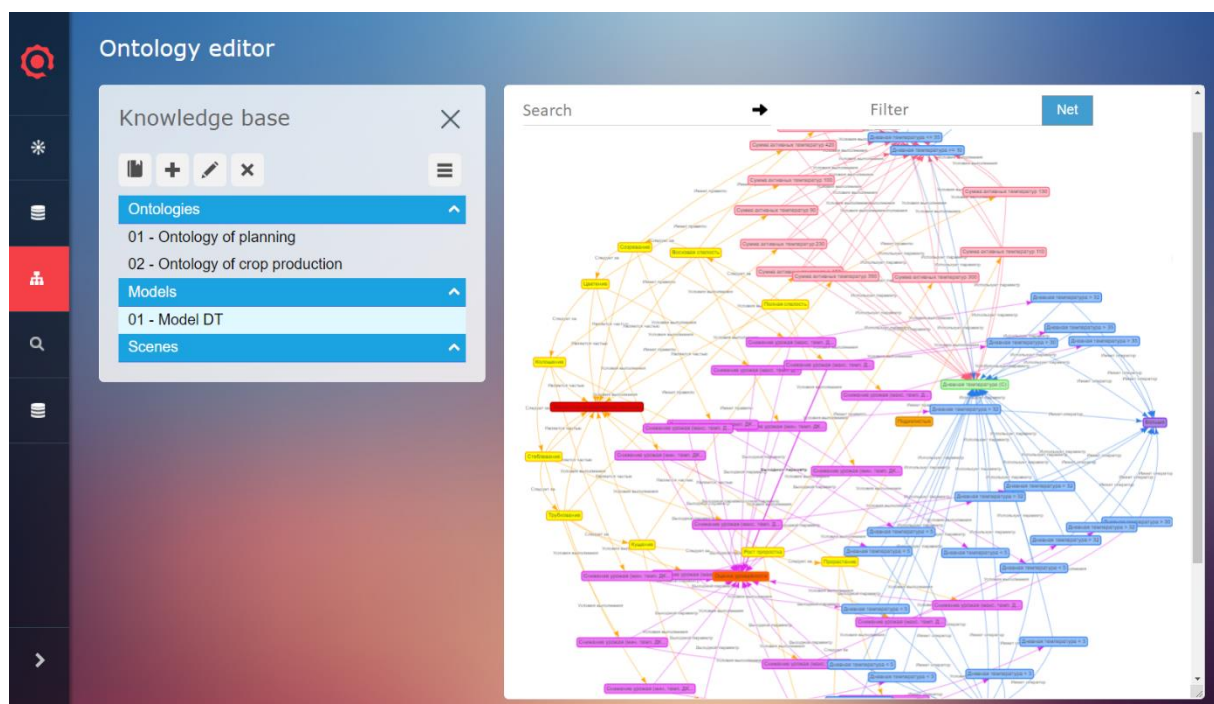

Figure 8. Ontology editor screen.

Based on ontology, the digital twin of plant is described: instances of classes of basic concepts are created and their attributes and relations are added. In particular, the following information is added: composition and coordinates of fields, sowing dates, used crops and varieties.

An important function is taking into account the rate of plant growth and development. The plant develops depending on the weather conditions during each stage. At the same time, it will be possible in the future to include into the ontology concepts and relations that describe the rate of growth and development of individual parts of the plant, depending on these parameters. Then, at the output of the system, it is possible to obtain not only estimates of yield and dates of stage completion, but also numerical estimates of plant parameters (leaf size, plant height, parameters of the root system, etc.). Comparing these estimates with the observed ones, it is also possible to adjust the forecast of yield and dates of next stages.

To obtain data about the external environment, a service connection mechanism is implemented. Each service must provide a unified interface for accessing the values of the specified parameter for the selected field within a certain time interval. The system automatically routes requests for data, directing them to those services that have registered themselves as providers of the requested parameter. All the information in aggregated form is displayed in the form of a graph (figure 9).

Figure 9. Graph of changes in environmental parameters.

This information is used by the planning module. Its task is to form and adaptively restructure the plant development plan, taking into account the prevailing environmental conditions. The module creates and configures agent instances, provides a multi-threaded environment for their execution, determines the order and algorithm of their work.

As a result of the module's work, the following plant development plans are formed:

- Nominal (plan for development of plants in ideal conditions);
- Planned created for a specific field at the beginning of the season, from the moment of sowing);
- Current (observed), taking into account real weather changes;
- Simulated – for simulations like what will happen if next month there is less precipitation or if fertilizer is applied (figure 10).
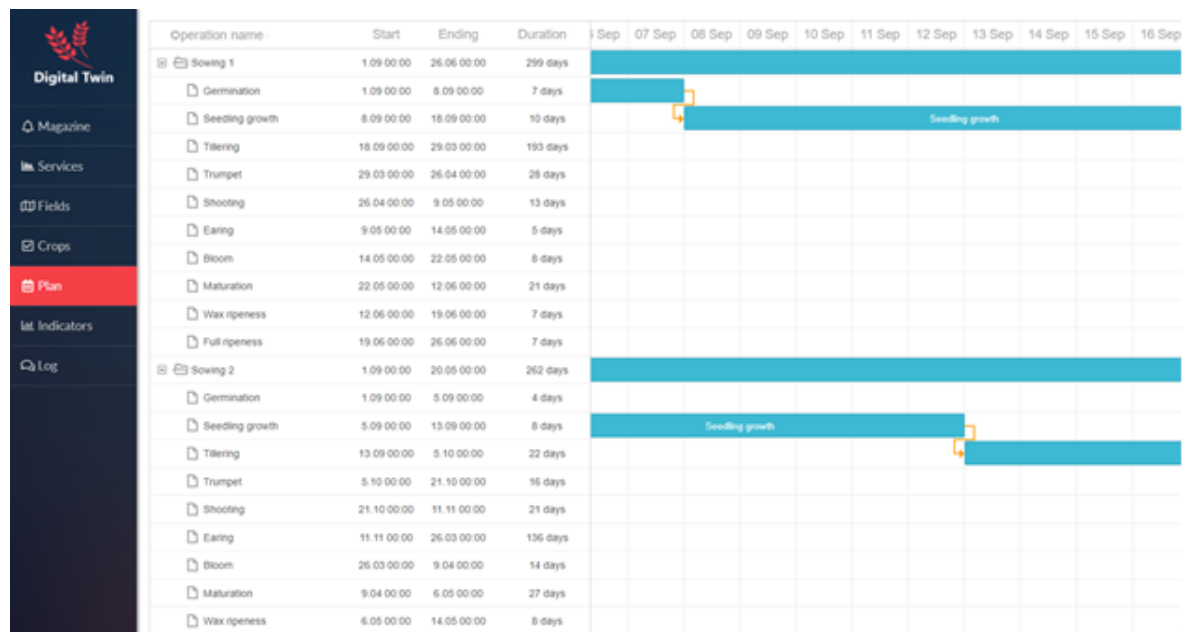
Figure 10. Plant development plans.

To implement the simulation mode, the user is given the opportunity to enter events related to changing crop parameters (for example, the date), as well as adjusting environment parameters (changing the temperature, cloudiness, applying fertilizers, etc.). When these events occur, the DT agent corrects its state, recalculating durations and results of stages affected by the event (figure 11).
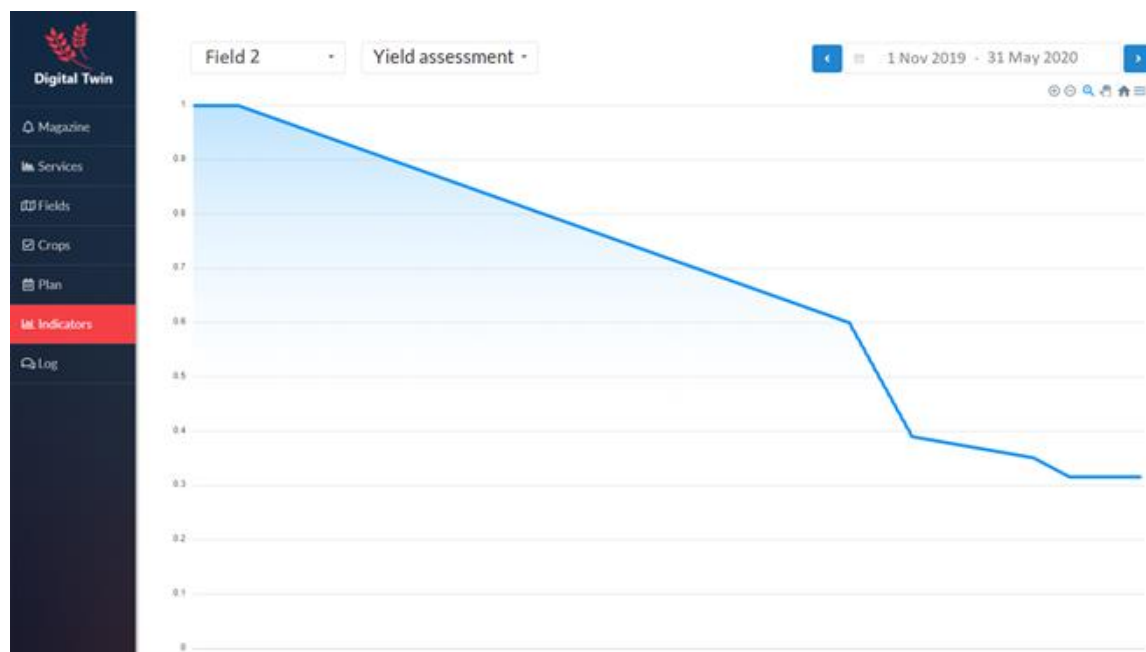

Figure 11. Graph of change in yield estimations.

Further, it is planned to supplement functions of the IDC system prototype as follows.

If the rules of plant development are based on observed or measured parameters of a plant, soil or other parameters that cannot be obtained automatically from weather services (for example, appearance of sprouts, first leaves or earing), an agronomist should be asked to perform examination of crops and fixation of these phenomena, followed by collection and transfer of the necessary samples to a laboratory (leaf or soil samples, etc.). Results of this inspection are received as events that clarify parameters of environment and the course of plant development, containing the obtained values of required parameters, for example, in the format of an attached and downloadable .csv file. The input of this file immediately leads to recalculation of the current stage and all subsequent stages of plant development.

To issue situationally dependent recommendations to the plant development plan for each stage, it is necessary to introduce the concept of "recommendations" into the ontology, which also depend on the values of environmental parameters coming from external services, the current stage and the actual state of plant development. In this case, advice will be given on what agrotechnical measures should be carried out in the field, including basic actions in the ideal part of the "tube" and even changing the composition of fertilizers. These recommendations can be "modeled" in the created prototype by entering the event corresponding to the change (for example, the necessary minerals have been added to the soil), and observing the result in terms of time and yield for the weather forecast that was selected for this field.

These data can be weighed against the risks and costs of losing or preserving the crop and the cost of fertilizers or plant protection products applied in order to provide an economic basis for agrotechnical measures.

8. Results and discussion

During development of IDT system, the following results have been obtained:

1. New principles for construction and implementation of plant DT;
2. A new approach to formalization of DT knowledge to form expert knowledge of the subject area based on ontological specification of stages of plant growth and development;
3. A method for calculating the forecast for yield and duration of plant development stages based on expert knowledge:
   - A model of the "tube" of plant development parameters;
   - A method for calculating the forecast of yield and the dates of beginning and end for stages of plant development within the "tube";
   - The key parameters of development of the "tube" for wheat;
   - Ranges of the "tube" in terms of the nominal and permissible trajectory of development;
   - Model calculations together with experts on model data, which confirmed the adequacy of the method;
4. Ontology of wheat development:
   - Collection of data to formalize the domain knowledge of experts and agronomists;
   - Identifying the main classes of concepts and relations to formalize expert knowledge for the "tube" of plant development parameters;
   - Ontology of plant development to implement the "tube" model;
   - Ontology of plant development based on the "tube" model transferred to the ontology editor in a digital form readily available;
5. A knowledge base for development of wheat stages:
   - Possibility to access external servers, for example, weather, and external calculation functions;
6. The structure and functions of intelligent system of plant DT (using wheat as an example), open for replenishment with other crops;
7. A prototype of an intelligent plant DT system in Java;
8. A prototype tested on model data to prove its practical applicability for further filling, developing the knowledge base, making calculations, as well as making decisions.

9. Conclusions

The system of plant IDT as part of ICPS provides the following advantages:

- systematization and formalization of domain knowledge on new technologies of plant cultivation;
- improving the quality of crop production due to a more detailed analysis and control of microstages of plant growth and development in each field;
- automation of decision support processes for agronomists to improve the efficiency of farms and accelerate introduction of precision farming technologies.

Using the IDT system, agronomists of different farms can create their own knowledge bases with digital twins of cultivated plants, which will be useful both for their own farm and for similar farms, as well as for research, educational or other organizations.

Introduction of plant IDT as part of ICPS reduces complexity and labor intensity of managing a precision farming enterprise, improves the quality of crop production on farms and raises business efficiency, as well as reduces dependence of enterprises on the knowledge of agronomists and mitigates risks of yield decrease or failure.

Acknowledgement

References

[1] Precision Agriculture Technology: The Future of Precision Farming with IoT: https://www.digiteum.com/precision-agriculture-technology

[2] Rojko A 2017 Industry 4.0 Concept: Background and Overview *International Journal of Interactive Mobile Technologies* **11** 77–90

[3] Leitao P, Karnouskos S, Ribeiro L, Lee J, Strasser T and Colombo W A 2016 Smart agents in industrial cyber–physical systems *Proc. IEEE* **104** 1086–1101

[4] Zambon I et al 2019 Revolution 4.0: Industry vs. agriculture in a future development for SMEs *Processes* **7** 36.

[5] Zadoks growth scale (Agriculture and Food): https://www.agric.wa.gov.au/grains/zadoks-growth-scale

[6] BBCH Mohograph 2001 Uwe Meier (Ed.) (Berlin and Braunschweig: Federal Biological Research Centre for Agriculture and Forestry)

[7] Verdouw C and Kruize J W 2017 Digital twins in farm management: illustrations from the FIWARE accelerators SmartAgriFood and Fractals: 10.5281/zenodo.893662

[8] Chao J 2020 Smart farms of the future: Making bioenergy crops more environmentally friendly: https://www.rdworldonline.com/smart-farms-of-the-future-making-bioenergy-crops-more-environmentally-friendly/

[9] Pantazi X E, Moshou, D and Bochtis D (Eds.) 2020 *Intelligent Data Mining and Fusion Systems in Agriculture* (Academic Press) chapter 2 330: DOI: 10.1016/B978-0-12-814391-9.00002-9

[10] Khan M S, Semwal M, Sharma A and Verma R K 2020 An artificial neural network model for estimating Mentha crop biomass yield using Landsat 8 OLI *Precision Agriculture* **21** 18-33

[11] Alves R G, Souza G, Maia R.F et al 2019 A Digital Twin for Smart Farming *Proc of IEEE Global Humanitarian Technology Conference* (IEEE: IEEE Xplore® Digital Library): DOI: 10.1109/GHTC46095.2019.9033075

[12] Verdugo-Vásquez N, Acevedo-Opazo C, Valdés-Gómez et al 2020 Towards an empirical model to estimate the spatial variability of grapevine phenology at the within field scale *Precision Agriculture* **21** 107–130

[13] Tang L, Tan F, Jiang H, Lei X, Cao W and Zhu Y 2011 *Root Architecture Modeling and Visualization in Wheat* Li D Liu Y and Chen Y (Eds.) IFIP Advances in Information and Communication Technology (Berlin-Heidelberg: Springer) **345** 479-490

[14]   Domingue J, Fensel D and Hendler J (Eds.) 2011 *Handbook of semantic web technologies* (Berlin: Springer-Verlag) 1035

[15]   Rzevski G and Skobelev P 2014 *Managing complexity* (London-Boston: WIT Press) 256

[16]   Laryukhin V, Skobelev P, Lakhin O, Grachev S, Yalovenko V and Yalovenko O 2019 The Multi-agent approach for developing a cyber-physical system for managing precise farms with digital twins of plants *Cybernetics and Physics* **8** 257–261

[17]   Skobelev P, Larukchin V, Simonova E, Goryanin O, Yalovenko V and Yalovenko O 2020 Developing a smart cyber-physical system based on digital twins of plants *Proc. Int. Conf. on Smart Trends in Systems, Security and Sustainability (WORLDS4 2020)* (IEEE: IEEE Xplore® Digital Library) (in Publishing)