DETERMINISTIC MODELS OF PLANT ENVIRONMENT

O. D. Sirotenko
All Russian Institute of Agricultural Meteorology, Obninsk, Russia

Keywords: Simulation, plant environment, agroecosystem, energy-mass exchange, evapotranspiration, hydrometeorological regime, plant productivity

Contents

Summary

Ecological health of the biosphere is governed by the Earth's plant cover which in its turn depends on the plant habitat, i.e. the surface atmospheric layer and soil. The rate and direction of energy-mass exchange processes in this system stimulate not only the intensity of productive plant processes but also the level of soil fertility, gas composition of the atmosphere, and, to a certain extent, the Earth's climate.

The paper describes the physical bases of plant environment and presents corresponding deterministic models. Associated energy-mass exchange processes are considered in complex multicomponent closed media, i.e. the surface air layer and the root layer which are penetrated by phytomass elements and, by virtue of this, bind these media into a single energy-mass exchange system. Calculations of a hydrometeorological regime and photosynthetic productivity of plant cover are presented as a boundary problem of mathematical physics. The ways to simplify this problem are considered. Results are presented of numerical experiments on an assessment of the effect of external meteorological conditions, parameters of plant cover and soil on a microclimatic regime and agrocenosis productivity.

In addition to deterministic models, a statistical approach is considered for the description of relationships between hydrometeorological conditions and crop productivity.

1. Introduction

Ecology and related scientific disciplines (agronomy, soil science, plant physiology and others) are undergoing a rapid evolution. Mathematics, rather than being just a useful tool for describing research results and correlating them with theoretical concepts, has found growing support as the only possible approach to some problems. In terms of scientific methodology and philosophy, the current situation may be defined as the commencement of the maturity stage with the "normal" ways of conducting scientific research being dominant. One of the essential problems in this field of research is the description of environmental impact on plant conditions and productivity.

Three stages of development may be discerned in the ways of studying relationships between the physical environment and crop production. In the first stage descriptive approaches prevailed, with relationships between weather (climate) conditions and crop production being explained on the basis of agronomy and plant physiology concepts and the laws of soil and atmospheric physics. This qualitative approach was replaced with an empirical statistical one which emphasized mostly the search for direct relations between input and output of a black box - the "Crop-Weather system". A transition to the third current stage had been initiated in the middle of the 20th century by theoretical works and only in 2-3 decades the results of these investigations had affected the practice. The essence of a new approach, called a dynamic one, lies in the fact that the problem of describing environmental impact on plant productivity is considered as a process and calculation of crop productivity is formulated as a problem of mathematical physics. It means that a system is differential equations describing energy-mass exchange in the Soil-Plant-Atmosphere system is developed, the corresponding boundary conditions are given, and the final agrocenosis productivity is calculated as a result of this system integration from plant emergence to maturity.

2. Static models: empirical-statistical approach

A static model is a mathematical structure which explicitly does not include time. In addition, the system's behavioral characteristics which are essentially time-dependent have to be discarded. Since everything in this world is changing, either rapidly or slowly, any static model is always an approximation. Static models describing environmental impact on plant productivity are usually developed as empirical-statistical models which use a minimum of known (*a priori*) information about the physical mechanisms of corresponding phenomena or processes.

The English mathematician R.Fisher was the first to consider the problems arising in the development of such models and to pay attention to the fact that in these problems the number of effective environmental factors and the number of years of observation are of the same order of magnitude. He suggested methods for compressing information on weather conditions based on a polynomial approximation of temporal variations in meteorological parameters.

Let us give another approach to overcome the difficulties associated with multidimensionality. A model has been developed to describe the weather impact on winter rye yields in dry areas near the Volga (Russia). The spring and summer crop growing season is split into nine sub-periods and calculation is made of coefficients of correlation between the "purified" crop yield series and meteorological factors for each isolated time interval. The corresponding coefficients of correlation are given below:

How can the model considering the effects of all 18 factors be developed with only 23 years of observation available? Weights are attributed to each meteorological variable proportional to the corresponding coefficient of correlation given above. As a result, the number of independent variables in the equation of regression was reduced from 18 to 3. The effect of winter conditions was excluded with the help of a regression equation, connecting crop yield with visual crop assessment when the snow has melted.

	I	II	III	IV	V	VI	VII	VIII	IX
Air temperature (°C)	0.39	0.20	-0.28	-0.53	-0.58	-0.10	-0.41	-0.34	-0.34
Precipitation sum (mm)	0.02	0.18	0.54	0.81	0.55	0.26	0.36	0.19	0.04

In a more general form this problem is presented as follows. The effect of meteorological conditions varies significantly in the process of plant growth and development. Suppose that a plant development period is equal to 100 days. How many meteorological factors are to be taken into account? At least three: air temperature, a moisture indicator and incoming solar radiation. Thus, to set up a linear equation for calculating the crop yields, 3x100=300 regression coefficients have to be identified. When setting up a nonlinear equation that takes account of relations between different hydrometeorological factors, the number of coefficients to be identified increases several fold. The problem of multidimensionality may be solved completely only as a result of transition from a static to a dynamic form of model description.

Even so, within an empirical-statistical approach there are ways to reduce the burden of the "curse of dimensionality". To this end, the meteorological variables affecting plant productivity are presented as discrete functions of time which are expanded into an orthogonal series of Chebyshev polynomials or, to be more effective, in the eigenvectors of covariance matrix. Among the $z_1, z_2, ... z_n$ expansion terms the coefficients are selected which are most significantly connected with the output variable y (yield). As a rule, the number of such variables turns out to be much less than the initial dimensionality of a problem. The greater the correlation of initial meteorological variables, the more effective is this method (a component analysis) of initial information compression. When the initial variables $x_1, x_2, ..., x_n$, are non-uniform and/or weakly correlated with each other, the considered method may appear to be insufficiently effective.

A universal way to improve the empirical-statistical models is more complete application of information which can be obtained by analyzing the physical essence of a simulated system. It is shown that using even the simplest form of linear constraints $\alpha_i \geq 0$ as a linear equation of regression with certain constraints on parameters, calculations can be appreciably improved. For example, in obtaining the regression equations for deriving 10-day totals of precipitation for individual stations based on the data from the nearest stations, it was found that, as a rule, several regression coefficients turn out to be negative, contradicting the basic rules of physics. Therefore, in addition to the conventional regression equations, new equations were obtained by the quadratic programming and by means of component analysis. Comparative simulation schemes were verified with independent samples. The mean values of conjugate correlation coefficients between calculated and actual rainfall totals were found to be as follows: for classical regression analysis – 0.373; for component analysis – 0.578; for quadratic programming ($\alpha \geq 0$) – 0.685. The account for limitations in the form of the requirements for non-negative regression coefficients, led to an increase in the variance from 14% to 47 %, i.e. an improvement in the calculation of more than 3 times.

Rather effective is the application of "alternative" methods of regression analysis, such as robust regression, ridge regression, etc. Thus, robust regression makes it possible to overcome the difficulties related to violating the normality of the predicted distribution and to the availability of "anomalous" data. Coefficient estimates for robust regression equations result from minimizing the special goal function (Huber function). It is a symmetrical error function ξ, which is presented by a parabola in the segment $[-h, h]$ and then continued by two straight lines. The h value defines a threshold after which the rate of error 'weight' growth decreases and characterizes the degree of estimate robustness. At rather high h-values, the obtained estimates coincide with the conventional least square estimates. Usually the h value is chosen so as to decrease the contribution of significant deviations of actual values from the calculated ones.

Ridge regression is aimed at overcoming difficulties related to predictors being closely correlated. Estimates of ridge regression coefficients may be obtained by solving

$$y = Q^{*}\left(C + L\right)^{-1} \cdot X,$$

where Q^{*} is the vector formed of the coefficients of covariance between the predicted y and components of the vector-predictor x (* is for transposition); C is the covariance matrix of predictors; L is the diagonal matrix (i.e. such that $l_{ij} = 0$, if $i = j$). Coefficients l_i are called the ridge parameters and defined by fitting an independent sample. In defining l_i *a priori* information concerning supposed signs for the desired regression coefficients can be used.

Thus, studies of the cereal crop formation in Kazakhstan, for example, showed that under deficient moisture conditions the coefficients for rainfall R should be non-negative and at temperature T_i – non-positive. However, a model obtained by the least squares method is

$$\begin{aligned} y = &-0.105R_1 - 0.091R_2 + 1.294R_3 + 0.500R_4 \\ &+0.559t_1 - 0.904t_2 + 0.138t_3 - 0.508t_4 \end{aligned} \tag{1}$$

Here R_1, R_2, R_3, R_4 and t_1, t_2, t_3, t_4 are, respectively, the rainfall and temperature for the third decade of May and three decades of June. Theoretical assumptions are violated for the following variables: R_1, R_2, t_1 and t_3. The reason is poor stipulation of the corresponding correlation matrix. To establish the "actual" dependence, the ridge regression analysis was used, which at the minimum value of a ridge parameter $l = 0.3$ resulted in the equation where the signs of all coefficients totally agree with the physical representation

$$y = 0.041R_1 + 0.070R_2 + 0.038R_3 + 0.288R_4 - 0.036t_1 \\ -0.249t_2 - 0.044t_3 - 0.249t_4 - 0.015 \qquad (2)$$

Verification against independent data has shown that a mean square prediction error for the first equation is 2.21 and for the second is only 0.911, i.e. the regression equation error was reduced to less than half the original value.

This is a typical example. So far as the methods of robust and ridge regression "defend" from different violations of input prerequisites of regression analysis, their combined utilization seems to be effective within one problem.

The volume of used *a priori* information may be increased later on by developing specialized physico-statistical models. Consider this approach with Baier's model as an example which takes the following form:

$$Y = \sum_{t=0}^{m} V_1 \cdot V_2 \cdot V_3 , \qquad (3)$$

where Y is the dependent variable representing the final yield; V_1, V_2, V_3 are functions of the selected input variables presented as:

$$V = A_0 + A_1 x + A_2 x^2 ,$$

where A_0, A_1, A_2 are the polynomials of the fourth degree in biometeorological time t (t =0 at the date of sowing , t = 1 at the date of emergence, t = 2 at the date of tilling, etc.). Coefficients of the function V are evaluated by numerical optimization methods; in V_1, V_2, V_3, x stands for one of the following characteristics of weather conditions over a 24-hour period: minimum and maximum air temperatures, relative humidity of the soil, the ratio of actual to potential evaporation, total radiation. The following combination of 3 factors, out of 5 enumerated, was found to have the highest information content: solar radiation, minimum air temperature and the ratio of actual to potential evaporation. The model parameters were evaluated on the basis of the data on spring wheat yields in Canada.

Baier's model (3) for the spring wheat accounts for about 80 % of the yield variance. This model is a good example which shows how, by comparatively simple means, it is possible to devise a much better method for investigating the CROP-WEATHER system, as compared with the classical regression methods. The model considered has 12 parameters for each variable but enables a composite quantitative estimate to be made of the daily crop yield formation conditions during the whole vegetation period.

In addition to the CROP-WEATHER systems, the models of soil fertility are decisive for agriculture. Consider one model of this type. The effect has been studied of variables characterizing impact of the initial soil state, climatic factors and agrotechnical indices on humus dynamics of the arable soils in the European nonchernozem zone of the former USSR. Results of long-run field experiments with different crop rotations and fertilizers were used for model development. As a resulting variable ΔC_n the difference was used between the carbon content in soil at the end and in the beginning of the observation period, divided by years of observation n. The problem of developing humus transformation model is reduced to identification of parameters for the following nonlinear function:

$$\Delta C_n = f\left(C_H, L, H, N, pp, MT, R_v, PEBT\right),$$

where C_H is the humus content in the beginning of observations (%), L is the content of physical clay (%), H is the annual mean norm of manure ($T \cdot ha^{-1}$), N is the annual mean dose of mineral nutrients with regard for nitrogen ($kg \cdot N \cdot ha^{-1} \cdot year^{-1}$), pp is the fraction of clean-cultivated crops and bare fallow in crop rotation (%), MT is the fraction of perennial grass in crop rotation (%), R_v is the mean precipitation sum (mm) for a vegetation period, $PEBT$ is the potential evapotranspiration (mm) for a vegetation period.

The model sought for ΔC_n calculations was presented as an arbitrary-degree polynomial, the coefficients of which and their standard errors are given in Table 1.

Independent variables	**Regression**	**Standard coefficients errors**	**t-statistics**
$pp^{2.5} \cdot Cn^{0.1}$	3.527532	1.000962	3.5241
$pp^{1.5} \cdot Cn^{3.8}$	-0.000022	6.935189	-3.1051
$N^{0.4} \cdot Cn$	0.000923	0.000328	2.8183
$Rv \cdot Cn$	-0.001932	0.00019	-10.1642
$PEBT \cdot Cn$	0.000931	0.000095	9.8336
$MT \cdot Cn$	0.003848	0.00042	9.1669
MT	-0.009203	0.000936	-9.8317
$MT \cdot L$	0.000126	0.000018	7.1229
$H^{0.7} \cdot PEBT^{0.5}$	0.000096	0.000018	5.2702
$PEBT^{0.025}$	-25.915459	2.660653	-9.7403
$Rv^{0.5}$	0.079278	0.008077	9.8159

Table 1. Regression equation for calculating humus reserve variations ΔC_n (%, $year^{-1}$) for loamy soddy-podzolic soils

This model generalizing data of 95 years of field experiments reproduces 87 % of dispersion for the independent variable ΔC_n. Its multiple correlation coefficient is estimated as 0.93. The regression coefficient estimates from the absolute value exceed their standard errors threefold and more in all cases, except one. It confirms the reliability of obtained estimates for an overwhelming majority of model coefficients. The exception is a coefficient for the independent variable $C_H N^{0,4}$, however, and for this coefficient the confidence limit exceeds 94%.

So, the examples of statistical model development for crop yield calculations as well as for humus dynamics assessment show that the observed difficulties have a great deal in common. They include a large number of factors, generating great dimensionality, nonlinearity, strong correlation of input parameters, a small volume of samples, observation contamination by errors, and significant deviations from a normal distribution law. All this violates the input theoretical prerequisites, on which the classical regression analysis is based. The application of alternative methods of component, ridge and robust analyses and regression with coefficient limitations in some cases permit the stated problems to be solved more effectively. Nevertheless, the greatest effect is achieved in the case when the individual structure of a model is successfully evaluated from preliminary analysis.

3. Dynamical models: An approach oriented to process account

More than 50 mathematical models of crop production have been developed for most important crops in different countries of the world; they may be called dynamical models. All of them are computer programs enabling the dynamics of the vegetative biomass of crops to be calculated for the entire growing period or a significant portion of it. Dynamical models differ appreciably from each other, depending on the tasks they are intended to undertake, the degree of detail in the description of individual processes and the amount of environmental data used. All dynamical models are based on a similarly structured set of difference equations of the form:

$$m_p^{j+1} = m_p^j + f_p\left(M^j, X^j, A^j\right)\cdot \Delta t, \quad m_p^{j=0} = m_p^0, \qquad (4)$$

where m_p^j ($p \in l, s, r, R$) is the dry matter of leaves, stems, roots and reproductive organs, respectively; M^j is the vector consisting of m_p^j; X^j is the vector characterizing the current state of environmental conditions (e.g., x_1^j is solar radiation; x_2^j is air temperature, x_3^j is air humidity, x_4^j is precipitation); A^j is the vector of functional and numerical parameters of the model; *j* is the current moment in time, *j*+1 is the future moment in time; Δt is the time step (for most of modern models the time step is taken to equal to 1 day).

Equation (4) is integrated in one set by the functions $f_p\left(M^j, X^j, A^j\right)$. The physical meaning of these functions is obvious: they characterize the intensity of the vegetative matter increment for individual plant organs. The vegetative matter increment of individual plant organs f_p depends on the current conditions of the crop the current environmental conditions X^j and the model parameters A^j. Consider the role of each of these factors.

The vector components of M^j specify the general crop conditions, i.e. the area of its photosynthetic apparatus, the capacity of a root system and etc. It is clear that daily phytomass increments are in fact largely depended on this set of variables. Besides, the vector M^j plays also another role: its component values are the peculiar "memory" about past environmental conditions from the initial time moment up to the current one.

An important feature of dynamical models is the small dimensionality of the "weather" condition vector X^j. Even in the largest dynamical models the number of X^j components does not exceed 5.

The third argument of the function f_p is the vector of the model parameters A^j. Its dimensionality is specified by the degree of complexity of a dynamical model. A fundamentally important property of dynamical models is the lack of a rigid relationship between the number of parameters and the duration of a simulated period (a month, a season), and also the time step (an hour, a day). The number of parameters in a dynamical model is specified only by the degree of detail in describing processes being modeled.

4. Deterministic models of energy and mass exchange for plant environment

Deterministic models of the environment where the plants grow and develop are the models of a three-link-chain SOIL-PLANT-ATMOSPHERE system. The plants are simultaneously in two environments (air and soil) and continuously exchange with them by substance and energy. The plants "sew" these environments into a united hydrodynamical system – the roots absorb water from soil and leaves in the process of transpiration return it into the atmosphere as vapor. The surface air layer and soil are separated by the inner boundary, i.e. the soil surface where transformations and conjugations of heat, moisture, carbon dioxide and solar radiation flows proceed. The problem of energy and mass exchange for the SOIL-PLANT-ARMOPSHERE system is related to the conjugated problems for two linking environments with the "inner" moisture exchange via plants. The problem of conjugated heat exchange for the atmosphere and soil is considered to be one of the most important theoretical problems of meteorology.

Dorodnitsyn A.A. was the first who formulated this problem (with no account of vegetation) in 1941. Its difficulty, apart from vegetation account, consists in formulating boundary conditions on the soil surface. It is necessary to state that for lack of the evaluated analytical theory of natural evaporation from soil the boundary conditions for moisture transfer on the active surface can be correctly formulated for rather moist or over dried soil. In other cases the semi-empirical relations are used for the problem closure. A general one-dimensional problem of energy and mass exchange and some particular cases of its significant simplification are considered below.

-
-
-

Bibliography

Bichele Z.N., Moldau H.A., Ross Yu.K. (1980) Mathematical simulation of plant transpiration and photosynthesis under soil moisture deficit. Leningrad, Hydrometeoizdat, Russia 223 pp. [The mathematical simulation principles of a plant production process, the comparative assessment of some mathematical models and the analysis of plant physiology literature used in simulation are presented].

Boiko A.P. (1993) Numerical studies of a hydrometeorological regime in the plant environment, St-Peterburg, Hydrometeoizdat, Russia 223 pp. [Studies on aerodynamical, hydrometeorological and gaseous regimes in numerical experiments with the energy and mass exchange models of a soil-plant-atmosphere system].

Dubov A.S., Bykova L.P., Marunich S.V. (1978) Turbulence in vegetation cover, Leningrad, Hydrometeoizdat, Russia 182 pp. [The investigation results of turbulent regime peculiarities in plant communities are included, the experimental data obtained under natural conditions and in aerodynamic tubes are presented, the numerical integration results of nonlinear equations as applied to the boundary layer above horizontally homogeneous vegetation cover are analyzed and the structural transformation of an air flow with its approach to forest is investigated].

France J., Thornley J.H.M. (1987) Mathematical models in agriculture, Moscow, Agropromizdat, Russia, 400 pp. [A wide spectrum of problems on developing and applying mathematical models in agriculture from deterministic dynamical models of plant productivity, pest and disease development to meat and egg production simulation and farm planning and management].

Nerpin S.V., Chudnovsky A.F. (1975) Energy and mass exchange in a plant-soil-air system, Leningrad, Hydrometeoizdat, Russia 182 pp. [Considered are all most essential processes of energy and substance exchange in a three-chain system plant-soil-atmosphere, specifying external environmental conditions in the agricultural land].

Poluektov R.A. (1991) Agroecosystem dynamics models, Leningrad, Hydrometeoizdat, Russia 312 pp. [Development and application of simulation dynamical models of energy and mass exchange and productivity of anthropogenic ecosystems – agroecosystems].

Sirotenko O.D. (1996) Mathematical models in crop bioclimatology in the former USSR (history, achievements and prospects). *Advances in Bioclimatology* V.4, 125-169. [A review of mathematical simulation methods of energy and mass exchange in the plant environment and their impact on crop productivity].

Ulanova E.S., Zabelin V.N. (1990) Methods of correlation and regression analysis in agrometeorology, Leningrad, Hydrometeoizdat, Russia, 207 pp. [Application of classical methods of regression analysis and models of bridge, robust and other regressions for calculating and forecasting crop yield].

Biographical sketch

Oleg D.Sirotenko is Head of Department of the All Russian Institute of Agricultural Meteorology. His main scientific interests include mathematical modeling of energy-mass exchange in the soil-plant-atmosphere system and development of models describing weather and climate influence on crop productivity, and assessment of global climatic change and greenhoouse impact on agriculture. He is the lead author on the assessment by the IPCC (Intergovermental Panel on Climate Change) of potential impact on agriculture, the author of the IIASA/UNEP research project on climate change and agriculture. He is Chairman of the Working Group on Relationship between Climate and Sustainable Agricultural Production within the WMO Commission for Agricultural Meteorology.