Managing Traditional Solar Greenhouse With CPSS: A Just-for-Fit Philosophy

Mengzhen Kang, *Member, IEEE*, Xing-Rong Fan, Jing Hua, Haoyu Wang, Xiujuan Wang, and Fei-Yue Wang, *Fellow, IEEE*

***Abstract*—The profit of greenhouse production is influenced by management activities (e.g., environmental control and plantation scheduling) as well as social conditions (e.g., price fluctuation). In China, the prevailing horticultural facility is the traditional solar greenhouse. The key existing problem is the lack of knowledge of growers, which in turn leads to inefficient management, low production, or unsalable products. To secure effective greenhouse management, the production planning system must account for the crop growing environment, grower's activities, and the market. This paper presents an agricultural cyber-physical-social system (CPSS) serving agricultural production management, with a case study on the solar greenhouse. The system inputs are derived from social and physical sensors, with the former collecting the price of agricultural products in a wholesale market, and the latter collecting the necessary environmental data in the solar greenhouse. Decision support for the cropping plan is provided by the artificial system, computational experiment, and parallel execution-based method, with description intelligence for estimating the crop development and harvest time, prediction intelligence for optimizing the planting time and area according to the expected targets (stable production or maximum gross profit), and prescription intelligence for online system training. The presented system fits the current technical and economic situation of horticulture in China. The application of agricultural CPSS could decrease waste in labor or fertilizer and support sustainable agricultural production.**

***Index Terms*—Big data, cropping plan, greenhouse management, Internet of Things (IoT), modeling plant development and growth, nonlinear programming, smart agriculture, solar greenhouse, tomato.**

Manuscript received February 4, 2018; revised May 22, 2018; accepted July 13, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61533019, Grant 31400623, and Grant 31700315, and in part by the Basic and Frontier Research Project of Chongqing Science and Technology Commission under Grant cstc2018jcyjAX0587. This paper was recommended by Associate Editor H. Li. *(Corresponding author: Xiujuan Wang.)*

M. Kang and J. Hua are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Innovation Center for Parallel Agriculture, Qingdao Academy of Intelligent Industries, Qingdao 266109, China (e-mail: mengzhen.kang@ia.ac.cn; jing.hua@ia.ac.cn).

X.-R. Fan is with the School of Computer Science and Information Engineering, Chongqing Technology and Business University, Chongqing 400067, China (e-mail: xingrongfan@gmail.com).

H. Wang is with the Innovation Center for Parallel Agriculture, Qingdao Academy of Intelligent Industries, Qingdao 266109, China, and also with the Technology Research and Development Department, Qingdao AgriTech Company, Ltd., Qingdao 266000, China (e-mail: wanghaoyu@smartagritech.cn).

X. Wang is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Beijing Engineering Research Center of Intelligent Systems and Technology, Beijing 100190, China (e-mail: xiujuan.wang@ia.ac.cn).

F.-Y. Wang is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, also with the School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Research Center for Military Computational Experiments and Parallel Systems Technology, National University of Defense Technology, Changsha 410073, China (e-mail: feiyue.wang@ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2018.2858264

I. Introduction

THE RAPIDLY updating and massive applications of information communications technology (ICT) are impacting the production mode of agriculture. Traditionally, production, logistics, and sale processes were implemented sequentially and independently, with clear boundaries. In the age of the Internet and Big Data, these processes are becoming interlinked with information flow among them, with each being a component of an agricultural chain system. Data play an important role, including not only the management log data and the environmental data during the production process [1] but also market and social information that reflect the balance between product demand and supply. The impact established by big data is being paid great attention in different countries [2].

In China, the farming organization is experiencing a huge reform. The central government has set up a series of policies that support the development of modern agriculture, including supply side reform, the separation of three rights (ownership, management, and contraction) in rural land. These policies are stimulating the scaled agricultural production. In the past, the unit of production was mostly at the farmer family level, which occupies limited planting area, whereas in recent years, the gathering of small pieces of land into a scaled farm has become a trend. With the expanding farming scale, personal experience is no more sufficient for controlling management risk, and decisions should be made with more caution to obtain products with not only good quality but also good profit.

Trans-regional and even trans-national sales of agricultural products have already become popular. For example, tomatoes produced in Spain are sold to North Europe, potatoes produced in North China are sold to the South, and conservable products such as corn circulate at a global level. Therefore, growers care about the market-related information in the target area to estimate the gross profit. Thanks to the rich information

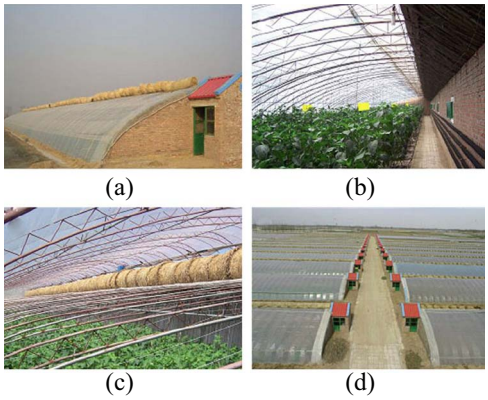


Fig. 1. Photographs of a single-slope solar greenhouse. (a) Exterior, (b) interior, (c) upper side of the double-arch solar greenhouse, and (d) a group of solar greenhouses. Photographs are from the literature of Gao *et al.* [7].

provided by the Internet, daily prices can be collected with a Web-crawling technique [3]. Convenient logistics and information dissemination place a greater demand on agricultural production management: growers need to not only produce products with good quality but also be able to plan the planting well according to the market need. However, speculative movements and the lack of information [4] are hurting the profit of farmers and customers [5]. In China, typical cases include the strong fluctuation in the price of ginger [6].

While in western countries the glass greenhouses are widely used, the prevailing greenhouses in China are the solar greenhouses (Fig. 1), accounting for approximately 14 millions in 2016. They are featured by low construction and management costs, and the wisdom of conserving the solar heat with thick walls in the day time for cold nights in regions between latitude 32°N and 43°N. Being energy-efficient, solar greenhouses have become the most important type of infrastructure for growing horticultural crops in China [7]. With the scaled production [as seen in Fig. 1(d)], challenges in solar greenhouse management lie in the insufficient technical support guiding production management (such as the dates of planting, irrigation, and fertilization) and limited knowledge for meeting market requirements. Since horticultural products such as tomatoes and cucumbers are difficult to be preserved, a good yield does not necessarily mean good profit. The information asymmetry (between market and growers) and the lack of knowledge in turn lead to soil pollution, unsalable products, and economic loss. The challenge to the new generation of farmers is how to react swiftly to the market and environment by making effective management decisions, which require multidisciplinary knowledge.

In industry, the term cyber-physical systems (CPSs) have been coined to describe the tight conjoining and coordination between computational (or cyber) and physical resources—that is, systems that feature a tight integration between computation, communication, and control in their operations and interactions, leading to a so-called intelligent enterprise [8]. For greenhouse production, preliminary CPS has been applied: the greenhouse ventilation facilities can be controlled automatically through sensing the greenhouse environment and intelligent computation [9]. In China, the wide coverage of communication infrastructures in the countryside provides the conditions for the application of ICT technologies [Internet of Things (IoT), cloud computing, etc.]. However, most agricultural IoT systems simply collect environmental data without sufficient data analysis or decision support, and their values require further investigation.

Analyzing and predicting the effects of the environment on crop growth is the task of crop growth models. Crop growth is decided jointly based on the genetic background and environmental conditions; the latter is influenced by human management. To meet the market need, customer information must also be taken into consideration. In the other domain, due to the profound impact of human behavior, with a human and social dimension in CPS, the human and social dynamics have been considered as an integral part of an effective CPS design and operation, leading to the concept of cyber-physical-social-systems (CPSSs) [8], [10]. The necessity lies in the observation that the choices of human beings (including customers and growers) affect the behavior of the system. For example, if many growers choose not to grow ginger, the price of ginger may escalate, which in turn changes the price curve and farmers' choice. It is very common in the countryside that a farmer decides what to plant according to the neighbor's field, similar to a driver's behavior in traffic. Such a system, the behavior of which cannot be analytically defined by an equation, is named the Merton system [11] and requires big data to support online decisions.

This paper presents an agricultural CPSS serving agricultural production management, with a case study of the solar greenhouse. This paper is organized as follows. The frame of CPSS for agriculture is defined, with the artificial system, computational experiment, and parallel execution (ACP)-based method for online decision support (Section II). A case study on the Chinese solar greenhouse (Section III) is provided for planting scheduling, aiming at constant production for contracted agriculture, or maximum gross profit. The results obtained for the collected information, data analysis, and optimization, are presented in Section IV. The discussion and conclusion are presented in Sections V and VI, respectively.

II. System Frame

An agricultural CPSS generally contains not only physical information concerning the environment and the crop, but also social information, such as the grower's experience and behavior, public opinion and habits, holiday schedules, etc. The system frame is shown in Fig. 2.

The physical world, including the crops, their environment and the control devices, is mapped to the cyber space (i.e., through various IoT facilities). The mental world includes the experience and knowledge of growers, as well as the preferences of customers, whose information is mapped through a management log system, including the timing, amount, and approach of applying fertilizers and other inputs (about growers), or through Web-crawlers, including the product price and preferences (of consumers). The artificial world, as a product of human's spiritual activities, includes the crop models and

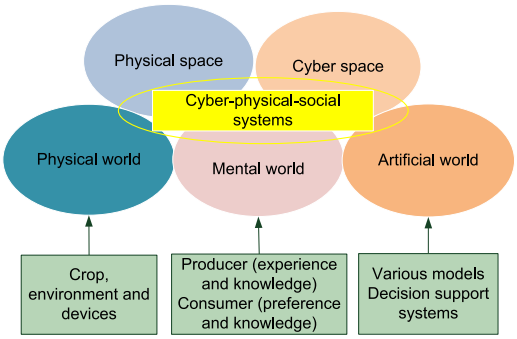


Fig. 2. Frame of an agricultural CPSS.

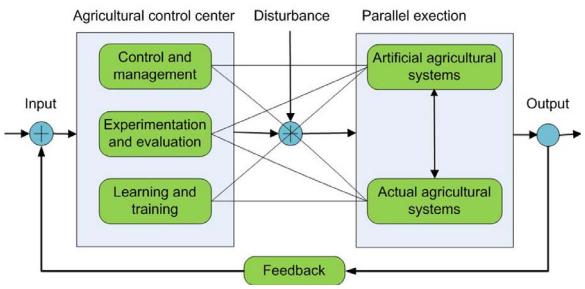


Fig. 3. Parallel agricultural system.

decisions, which are expressed in cyberspace and influence the physical world with control devices, the growers with decision support systems and the consumers with public opinion induction.

To address an agricultural system that features uncertainty, diversity, and complexity, the decision support process is modeled with a parallel agricultural system, which is composed of an ACP [12], as shown in Fig. 3. An artificial agricultural system is built by describing the dynamics of the actual one, including the crop, environment and humans. Accordingly, an artificial system is composed of the models in crop, environmental, human and market. Depending on how the description is generated, two kinds of systems can be distinguished, namely, the Newton system and the Merton system [11]. The former has determined the output with the given input, while the latter is composed of many individuals and the system behavior is a result of emergence. The computational experiment can be performed by configuring different populations, as well as using the optimization method as in the work. With computational experiments, the artificial system can predict and provide recommendations for the future. Prescription intelligence plays a role in parallel execution to provide online decision support: since the actual situation is complex and does not necessarily produce the predicted result, e.g., the growers choose the planting date based on their intuitions, or the environmental prediction is biased, the decision support is to be trained dynamically. The parallel execution (online calibration and optimization) of the artificial and actual system provides a prescription for the management and control of the actual system, as well as a platform for learning and training.

III. Case Study on CPSS for the Solar Greenhouse

The complexity of a CPSS can be chosen and tailored, according to the aim. Below we present the components of an agricultural CPSS designed for the management of a solar greenhouse. Because of the low environmental controllability of the solar greenhouse, the focus is on the cropping plan based on social and physical information. The importance of the cropping plan is similar to a business plan for growers, which could be very tedious work for a farm of a certain scale.

A. Sensing

1) Social Information: The price of vegetables is used as an indicator of social information. The most direct and realistic information is the price of vegetables in wholesale markets. Take for example the tomato from 2014 to 2017. The price data for tomato were collected through Web-crawler in three large cities (Xinfadi market, Beijing; PudongXinqu market, Shanghai; and Baiyunshan market, Guangzhou). These cities were chosen because they are typical consumption cities for horticultural products. Beijing and Shanghai in particular are two main target cities for Shandong province to sell horticultural products. Since 2015, daily prices for various products have been recorded [13]. The price data for more than one year were preferably collected to acquire historical information.

2) Physical Information: Physical information includes those for the environment (inside and outside the solar greenhouse) and crop yield. Data were from Qingdao, with a distance of approximately 170 km from Shouguang, a famous production and logistics center for horticultural products.

Environmental Data: The local environmental conditions in the greenhouse are obtained remotely based on IoT technologies. The air temperature, air humidity, and light intensity in a solar greenhouse were collected over 15 months (from August 3, 2016 to September 15, 2017) as they have direct effects on crop development, growth and diseases. The data were collected using the *Pennellii*-1 environmental monitoring system (named after the oldest wild tomato species) developed by the authors, uploaded to a cloud server every 15 min. Fig. 4 shows the system terminals and structure. The device can be charged either using the rechargeable battery or by a solar panel, so that it can be easily installed even if no external power is supplied. The data transmission occurs through GPRS, which is very easy to start and does not require the installation of a Web cable. These choices are made according to the situation of most solar greenhouses. Data can be viewed using a mobile phone or PC.

To predict the environmental conditions based on the outdoor data, open field weather data were obtained from the national meteorological department (from August 2016 to September 2017), including daily daytime and night weather conditions (sunny, cloudy, rainy, foggy or snowy), maximal and minimal temperature, maximal day length and night wind speed. Data for a full year are needed for model training and prediction with different seasons.

Crop Data: The crop (tomato) data were collected in greenhouse in order to analyze the effect of environment on crop

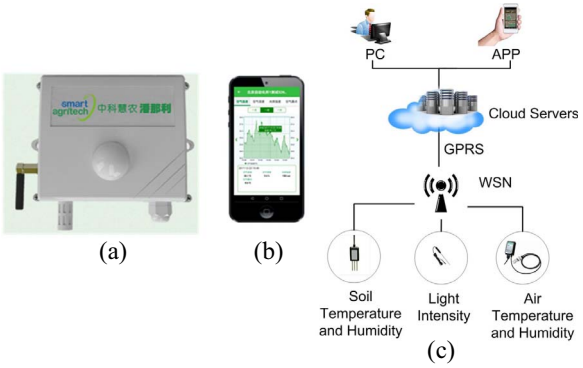


Fig. 4. IoT-based *Pennellii*-1 environmental monitoring system. (a) Device, (b) App, and (c) system structure.

growth. The plantation duration was from November 2014 to February 2015. Three plants were sampled (the weights of the internodes, leaves and fruits were measured) each time (in total five times) to record the plant growth and development.

B. Modeling and Description

1) Environmental Model: To predict crop behavior, the future environment in the greenhouse must be predicted. The national meteorological bureau provides historical data, which can serve as a reference for the long-term prediction of local climate estimations. Moreover, short-term (one week) predictions are available, and some startups are even providing climate prediction services on finer temporal and spatial scales, using open remote sensing images and machine-learning methods. Inside solar greenhouses, the local environment is more heterogeneous, depending on the greenhouse structure and types of crops. The environmental conditions (light intensity and temperature) inside the greenhouse are predicted from open-field data. This is feasible in the solar greenhouse because heating is seldom used. Moreover, in a given area, the same type of crop is often planted, which makes the internal environmental conditions more homogeneous among greenhouses. A neural network was chosen to build the link between indoor and outdoor conditions. The input is the open field weather data, and the output is the average temperature in the greenhouse. Tenfold cross-validation was used by dividing the training data into 10 groups. The mean and variance of the prediction errors were calculated.

2) Crop Model: Crop models simulate two basic processes of plants: 1) development and 2) growth. The importance of developmental stage lies in its link to the nutritional needs of the crop (timing of fertilization), occurrence of diseases (timing of spaying) and contracted agriculture (timing of marketing). In contrast, the prediction of crop growth (yield) plays a key role in estimating the quantity of fertilizer and the economic profit, especially for widely planted and batched-harvested crops. Therefore, in agricultural CPSS, the prediction of plant development and growth is important for planning and management.

Development Model: Crop development is influenced by the genetic factors and the environmental conditions. For the specific species, the temperature and photoperiod are the main influencing factors. The crop phenology can be measured by the physiological development time (PDT) [14], which is the duration needed for a key developmental stage under the optimal temperature and light conditions. The PDT is the sum of the relative physiological development effectiveness (RPDE), as

$$\mathrm{PDT} = \mathrm{SUM}(\mathrm{RPDE}) \tag{1}$$

where the RPDE is computed according to

$$\mathrm{RPDE} = \begin{cases} \mathrm{RTE}(e) & \mathrm{PDT} \le \mathrm{GER} \\ \mathrm{RTE}(e) \cdot \mathrm{RPE}(e) & \mathrm{GER} < \mathrm{PDT} \le \mathrm{FLO} \\ \mathrm{RTE}(e) & \mathrm{PDT} > \mathrm{FLO} \end{cases} \tag{2}$$

where $\mathrm{RTE}(e)$ is the relative thermal effectiveness, related to the air temperature; $\mathrm{RPE}(e)$ is the relative photoperiod effectiveness, related to the light condition; e encompasses the environmental variables, including light intensity and air temperature; GER is the PDT of the germination period; FLO is the PDT from the seeding to flowering period. The values of GER and FLO are two parameters to be adjusted.

Crop Growth Model: Many crop growth models have been developed over past decades, including the process-based (mechanistic) model [15] and statistics-based black-box model [16]. The choice on the spatial and temporal scale depends on the purpose of the model, which will not to be detailed here. In this paper, a knowledge-and-data-driven model [17], [18] is used to compute the yield of tomato, due to its features of considering plant growth mechanisms with higher plasticity. The advantage of the model is that it has fewer parameters to identify compared with the classical process-based model; additionally, the tomato yield and the biomass of leaves and stems can also be computed, which can support data circulation in the agricultural economy.

The knowledge-and-data-driven modeling (KDDM) consists of a knowledge-driven (KD) submodel and a data-driven (DD) submodel, as

$$y = f(e,\ \theta) = f_{\mathrm{k}}(e, \theta_{\mathrm{k}}) \oplus f_{\mathrm{d}}(e, \theta_{\mathrm{d}}) \tag{3}$$

where e and y are the input and output variables of KDDM, representing environmental conditions and yield, respectively; f is a function for a complete model relationship between e and y; f_{k} and f_{d} are the functions associated with the KD and DD submodels, respectively; θ is the parameter vector of the function f; and θ_{k} and θ_{d} are parameters associated with the functions f_{k} and f_{d}, respectively. The symbol "$\oplus$" represents a coupling operation between the two submodels, which is a composition coupling operator herein.

In this paper, the DD submodel was the radial basis function networks (RBFN) [17], which typically has three layers: 1) an input layer; 2) a hidden layer with a nonlinear RBF activation function; and 3) a linear output layer, as

$$E(e) = f_{\mathrm{d}}(e, \theta_{\mathrm{d}}) = \varphi(e)\theta_{\mathrm{d}} \tag{4}$$

where E is the average potential biomass production; $\theta_d = [w_1, \ldots, w_h]T$ represents the weights of the network, h is the number of neurons in the hidden layer, and $\varphi(e) = [\varphi_1(e), \varphi_2(e), \ldots, \varphi_h(e)]$.

The KD submodel was chosen as the GreenLab model, which is a generic functional–structural plant model simulating the dynamics of plant organogenesis, biomass production and allocation [19], expressed in a compact way as

$$y = f_{\mathrm{k}}(E(e), \theta_{\mathrm{k}}) \tag{5}$$

where y denotes the output of the KD model (e.g., the total weights of tomato fruits). Parameter values of θ_{d} of the DD submodel (i.e., RBFN) were obtained from [17]. Once the model parameter θ_{k} was estimated based on the observed data, given the environmental conditions e, the RBFN output (E) directly affects the Greenlab model output.

C. Computational Experiments

1) Aiming at Stable Supply: Farms with a certain scale may contract with a dedicated customer that has an estimate of weekly and whole consumption, such as a chain fast food restaurant. In that case, with the price already fixed, a stable production to meet the requirement is the primary aim of management. This raises the need for a production arrangement to achieve the best fit to this need.

In the cyber world, since the yield of a crop can be simulated, computational experiments can be conducted to optimize the system behavior with a dedicated target. The aim of this experiment is to fit best the needs of the contract by choosing a suitable planting time and area (expressed as the number of solar greenhouses, with each greenhouse in the same farm usually occupying the same area).

This need can be expressed as an optimization problem as in (6). Suppose T (kg) is the weekly tomato yield to be supplied to the client, and the grower owns N solar greenhouses. If the cultivar is given and there is no water or nutritional stress, the yield of tomato is dependent mainly on the indoor environment, which is indirectly indicated by the planting date. Let $f(e, \theta; x_i)$ be the tomato yield per plant (kg/plant), corresponding to the planting date x_i (day) in the greenhouse i. To simplify, let Dur be the growth duration (day) of tomato. For plants grown on day x_i, $x_i + \mathrm{Dur}$ is the harvest time (suppose the plants are harvested once, but the formula is extendable for continuous harvest). The aim of optimization is to minimize the variance of weekly production, as

$$\min \sum_{j=1}^{M} \left(\sum_{0 \le t_j - (x_i + \mathrm{Dur} < 7)} (g(e, \theta; x_i) \cdot d \cdot S) - T \right)^2 \tag{6}$$

subject to

$$g(e, \theta; x_i) = \begin{cases} f(e, \theta; x_i) & x_i > 0 \\ 0 & x_i = 0 \end{cases}$$

$$\delta(x_i) \begin{cases} 1 & x_i > 0 \\ 0 & x_i = 0 \end{cases}, \quad i \epsilon [1, \ N]$$

where d is the planting density (set to 3 plant/m^2); S is the planting area of a solar greenhouse (set to 1 mu, a Chinese unit of area, 0.0667 hectares or 666 m^2); M is the expected duration in number of weeks that meet the balanced supply; t_j is the day after planting in the jth week; and $y_j = \sum_{0 \le t_j - (x_i + \mathrm{Dur}) < 7} \left(g(e, \theta; x_i) \cdot d \cdot S\right)$ is the total harvested fruit yield in the jth week (kg). The sum $n(x_0) = \sum_{x_i = x_0} \delta(x_i)$ gives the number of solar greenhouses to be planted on day x_0, the output information supporting the growers.

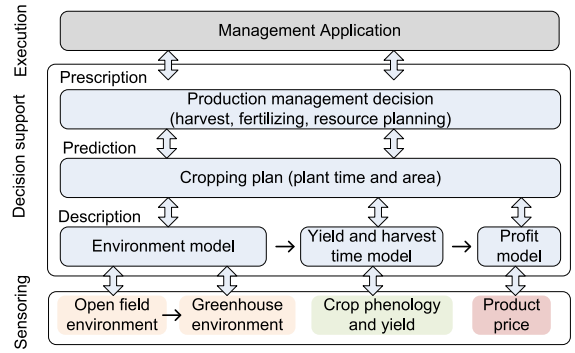


Fig. 5. Components of a CPSS for the solar greenhouse.

2) Aiming at Maximizing Gross Profit: In the case that the product is sold to the wholesale market, instead of a contracted customer, the profit of the grower is then dependent on whether the product is sold with a good price. Suppose there are two growing seasons each year in the same greenhouse, as is typical in Shandong province. With a given planting area (total number of greenhouses, N), the optimization problem is how to decide on the planting dates according to the product price and environmental information, to achieve the gross profit maximization. The problem can be expressed as

$$\max P_{x+\mathrm{Dur}} \cdot F_{x+\mathrm{Dur}} \tag{7}$$

subject to

$$\begin{cases} \mathbf{D}_1 \le x \le D_2 \\ G_1 \le \mathrm{Dur} \le G_2 \\ x, \ \mathrm{Dur} \ \epsilon \ N^* \end{cases}$$

where x is the planting dates, to be decided. D_1 and D_2 are the boundary dates for each growing season (Table I), known by the experience; Dur is the duration of the growth season (day), another variable that must be optimized, with a boundary of G_1=84, G_2=126; and $F_{x+\mathrm{Dur}}$ is the yield (fresh weight, kg) of tomato fruit planted on day x, and harvested on day $x + \mathrm{Dur}$, determined by the plant model: $F = f(x, \theta)$. Correspondingly, $P_{x+\mathrm{Dur}}$ is the price of tomato fruit on day $x + \mathrm{Dur}$ in the market (CNY/kg), currently taken from the average historical wholesale price data.

The full flowchart for the above work is summarized as in Fig. 5. The sensing level includes the physical and social information, which are the input for the ACP-based decision support, using description intelligence, prediction intelligence, and prescription intelligence [20]. At the application level is the management system providing tools for plantation scheduling or fertilization recommendations, similar to the functions of resource planning system or manufacturing execution system.

IV. Results

A. Tomato Price Collected by Web-Crawler

Fig. 6 shows the price of tomato over 2.5 years in three large cities in China based on source data from websites of

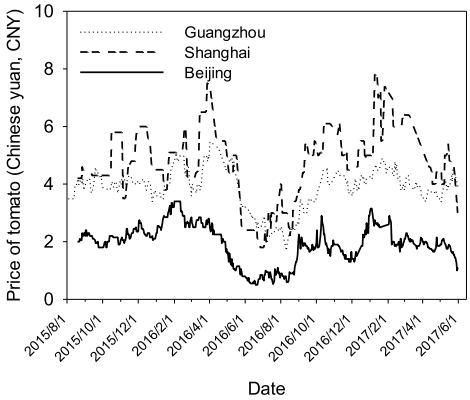


Fig. 6. Collected price of tomato by Web-crawler from August 2015 to May 2017 in Beijing, Shanghai, and Guangzhou in China.

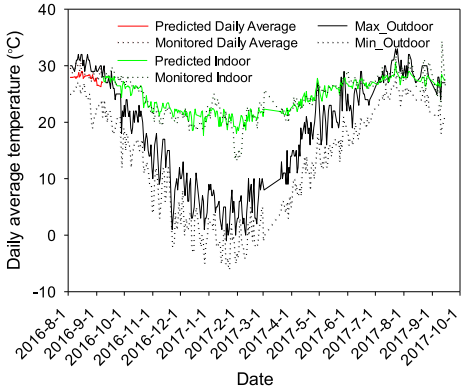


Fig. 7. Monitored outdoor maximal (solid) and minimal (dotted) temperatures (black lines), as well as monitored (dotted) and predicted (solid) temperature inside the greenhouse, for training (one month, red lines) and testing (one year, green lines).

local wholesale markets. Interestingly, although these cities are located in different places (from North to South), the pattern of price variation was common: lowest in summer, increasing from September (returning to school season), and reaching the peak around spring festival (February 8, 2016 and January 28, 2017, respectively). The price in Beijing is obviously lower than the other cities, partly because of its closer distance to the large source area of tomato, mainly Shandong province. These data are used as a reference to arrange the planting time.

B. Sensing and Describing Greenhouse Air Temperature

Fig. 7 shows the maximal and minimal temperatures outside the greenhouse (black solid and dotted lines, from the local weather information station), the monitored daily average temperature inside the greenhouse (green dotted line, by the IoT-based monitoring facility). It can be seen that even if no heating system was used, the indoor temperature was much more stable and higher than the outdoor temperature, especially in winter. The difference could reach as high as 20 °C. Although the outdoor temperature plummeted below zero, the indoor temperature was always above 10 °C, the approximate base temperature for tomato growth.

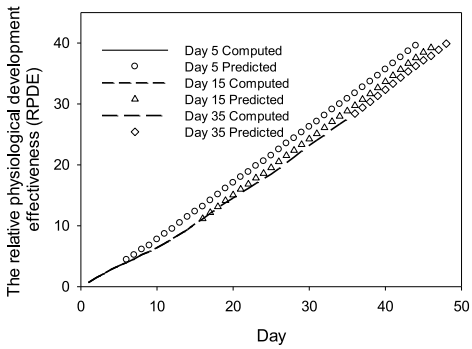


Fig. 8. Predicted PDTs of flowering time in tomato according to the predicted temperatures at days 5, 15, and 35.

For the prediction of indoor temperature based on the open field, data from August 3, 2016 to September 7, 2016 (approximately one month) were used as the training set, which is indicated by a red line. Data from September 8, 2016 to September 15, 2017 (approximately one year) were used as the testing sets (green dots). The results show that the trained neural network can well predict the temperature inside the greenhouse. During the process of cross-validation, we calculated the mean and standard deviation of the prediction errors, which were 1.82 °C and 1.54 °C, respectively. This result is very promising because temperature is one key factor in predicting plant developmental stage.

C. Describing the Crop Developmental Stage

Here, we take the prediction of the flowering stage as an example, which is a key indicator for the timing of corresponding horticultural activities. The cumulative RPDE value is computed for each day. It is computed based on the predicted indoor environmental data using past local climate data as the input. With the real-time monitoring of the actual data, the predicted phenology is updated according to the monitored data.

Fig. 8 shows the daily cumulative RPDE value, predicted (with lines) and actual (with dotted) values. The predicted PDTs for the beginning of the flowering stage were 43, 45, and 47 according to the predicted temperatures on day 5, 15, and 35, respectively. The results indicate that the phenology model can well predict the flowering time of tomato (the difference is 4 days), even with the high variation of environmental conditions. This finding can explain why the experienced farmer can predict phenology rather well. However, with online prediction, a more precise prediction can be achieved to support growers in arranging duties in the upcoming days.

D. Describing and Predicting the Crop Yield

Using a trained model [17] and having it trained again using the data collected in 2014, total fruit fresh weights of tomato plants were predicted for the different planting dates during the years 2014–2016 (January 1, 2014 is designated day 1), as in Fig. 9. Being the input of a crop model, the greenhouse indoor environmental data were simulated from the open field data using the environmental model. Two planting dates were

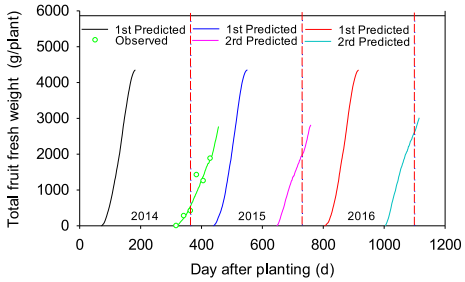


Fig. 9. Tomato yield prediction in the greenhouse (2014–2016).

simulated each year as in the solar greenhouse, as is generally the case: one in spring and another in summer. The yield curves are different because the plant experienced different environment conditions in solar greenhouses, as computed in Fig. 7. Dots represent the observed data, and the solid lines represent the predicted result. These descriptions are necessary for evaluating yield with the given planting dates, and they can be updated according to the collected yield data during production.

E. Optimizing the Planting Date and Area to Achieve Constant Supply

The weekly tomato yield (T) to be supplied is 20 000 kg, according to an empirical data from an agricultural company in Qingdao. The aim of optimization is to maintain the yield to be as stable as possible over 24 weeks continuously ($M = 24$), corresponding to a contract of nearly six months (from September 20, 2014 to February 22, 2015). The number of solar greenhouses (N) is 200, which corresponds to a middle-scale farm in Shandong province. Each solar greenhouse occupies one mu. The growth duration (Dur) is 84 days. The variables to be determined are planting area (expressed as the number of greenhouses) and the corresponding planting dates.

Fig. 10 shows the results of optimized variables: the number of greenhouses for planting tomatoes on different dates [Fig. 10(a)], and the corresponding yield curves [Fig. 10(b)]. The simulation period started from September 20, 2014 (day 263, autumn) to May 22, 2015 (day 507), with January 1, 2014 being day 1. It is noteworthy that the summer period (July and August) is not suitable for planting because it is too hot inside the solar greenhouse.

The resulting weekly yield is shown in Fig. 10(b) (dotted line). Each symbol refers to the accumulated yield during a week, for example, the yield on day 346 represents the total weight of the fruits harvested during the day 340 to the day 346. The mean production is 20274.6 kg, with a standard deviation of 438.3 kg. The coefficient of variation is 2.16%.

F. Optimizing the Planting Date and Duration to Achieve Maximum Gross Profits

Table I shows the optimization results for maximizing the gross profit by choosing the best planting dates and duration for each growing season. The boundary planting dates, D_1 and D_2, are shown in Table I. In total, there are 100 greenhouses ($N = 100$). A Gantt chart can be drawn according to

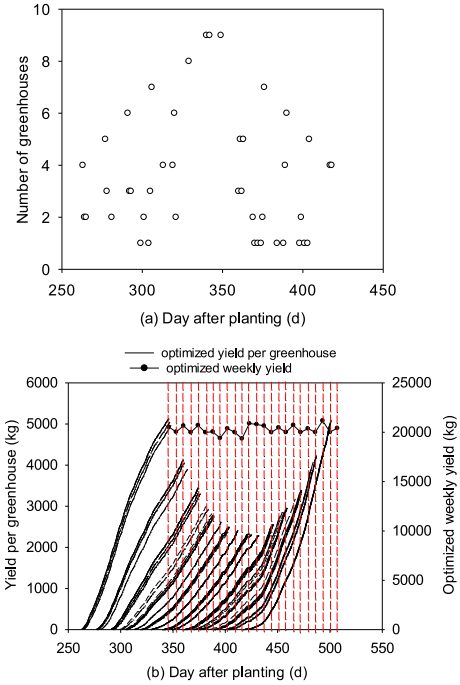


Fig. 10. Optimized planting date and number of solar greenhouses. (a) Numbers of greenhouses to plant at different dates. (b) Simulated production of each greenhouse (lines) and optimized weekly yield (dotted line).

TABLE I
OPTIMIZATION RESULTS FOR MAXIMIZING ECONOMIC GROSS PROFITS. JANUARY 1, 2014 AS DAY 1

Date (D_1-D_2)	Rotation	Planting duration	Planting date (x)	Gross profit (million CNY)
2014.1.1-2014.6.60 (day 1-181)	1	126	2 (2014.1.2)	2.93
2014.8.1-2015.2.28 (day 213-424)	2	126	243 (2014.8.31)	6.03
2015.1.1-2015.6.30 (day 366-546)	1	126	416 (2015.2.19)	3.51
2015.8.1-2016.2.29 (day 578-790)	2	118	585 (2015.8.8)	4.78
2016.1.1-2016.6.30 (day 731-912)	1	112	731 (2016.1.1)	2.62
2016.8.1-2017.1.20 (day 944-1116)	2	126	973 (2016.8.30)	5.26

this result (not shown), which is typically used in farms. The result dates are quite realistic: rotation one planted in January or February, and rotation two in August, with approximately one month separating the two rotations for recovery.

V. DISCUSSION

We have proposed in this paper the concept of agricultural CPSS and concretized it with a case study of a cropping plan for the solar greenhouse, a typical kind of greenhouse in North China. The decision support for the system is based on the ACP method. The greenhouse indoor environmental model, and in turn crop development and growth model, are built based on the environment and crop data. Computational experiments are made in cyberspace to provide solutions regarding to the planting date and area aiming at a constant supply or best profit, which can be updated according to real-time monitored data. Physical information regarding to the environment

and social information (herein the product price) are collected continuously to train the components so that the real and virtual systems interplays and guide each other.

Depending on the aim of application, a CPSS for agriculture is not limited to these components, while a common feature is to add a social dimension and online training to the system. This need is becoming stronger, especially with the increase in farm's scale in recent years. With the expanded scale of planting, agricultural production becomes too complex to be handled by the grower's brain. Independently, such information collection and analysis have been conducted, while the mechanism that links them dynamically is missing. For example, price information is open to farmers; Yang *et al.* [3] have developed a system to obtain the price of agricultural products using the crawl technique. However, even with the access to information, the decision support still relies on the experience and knowledge of the grower himself, which necessitates personal capability and vision. Research has also been conducted to examine greenhouse environmental control policy aimed at maximizing profit [21]; however, such work is based on historical data, and the mechanism of self-updating according to the actual price and environment is missing. Instead, CPSS aims to solving how to schedule the cropping plan according to the social information (e.g., price) and furthermore, realize the plan by adjusting the environmental condition, ultimately contract farming. By considering the physical and social information together, decision support can better fit to the needs of growers for management. In this sense, the function of CPSS is similar to the integrated function of the enterprise resource planning system and manufacturing execution system, which remains challenging even in industry [22], [23]. The integration of social information, from both supply (grower) and demand (customer) sides, could provide a more complete view of the full agricultural production system, so that the decision is more realistic. A benefit can also be found in the psychological impact: traditionally, agriculture is regarded as a low level and poor working condition, which hampers young people from entering this area. The agricultural CPSS could improve not only management efficiency but also people's perception of agriculture, so that young people are willing to participate.

Compared with industrial production, the specialty of agricultural CPSS lies in the objects of management are live bodies with growth and product changes that are highly related to the environment. In fact, for a long time, crop models have been aimed at solving the simulation and prediction of plant growth to better management and control, or crop breeding. Typical examples in horticulture include HortiModel from Wageningen University [24], and the derived TomSim model for tomato [25]. Others include those for cucumber [26] and sweet pepper [27]. In the past decades, the spatial scale of the crop model has been further detailed to the organ level, leading to functional–structural plant models [28], which can be used to analyze the differences in cultivars under a light environment. Model parameters can be identified inversely for plants grown under different conditions [19], [29]. However, when facing reality, often the calibrated model in the laboratory is no more suitable. In contrast, there are numerous studies predicting plant yield based fully on environmental data without considering the intrinsic rule of plant development and growth. With such background, the knowledge-and-data-driven model has been proposed to add to model transparency and plasticity, as proposed in current CPSS. Since crop models are based on environmental prediction, the latter becomes part of the full system. Based on the mentioned ACP method, the behavior of the system can be updated according to the monitored data, distinguishing itself from previous types of modeling.

A feature of the solar greenhouse is poor controllability. From another point of view, it is an advantage: without water or nutritional stress, the development of a crop is mainly adjusted by temperature and light conditions [14]. Due to light conditions, the solar greenhouses are located in North China and, mostly depends on natural light, except for the occasional light supplement. For temperature, heating is typically unnecessary even in winter, with a reliance on the insulation wall. The temperature is adjusted mainly by a ventilation window for situations in which the temperature is already higher than the response threshold for plant development. Due to low controllability, the indoor light and temperature conditions can basically be predicted from the open field condition, making it possible to predict the phenology. This step is the basis for the cropping plan of production management. Regarding environmental control, two kinds of temporal scales must be considered [30], namely, the long-term crop-level scale (in days) and the short-term environment-level scale (in minutes). While the cropping plan and optimal set point for environmental control are based on the crop-level model, daily management needs a policy to achieve the set point and expected harvest time, for example, when and how to open the ventilation window in the morning and afternoon. Prevailingly, this is again highly dependent on personal experience. In CPSS, based on real-time monitoring and prediction, decision support can be provided on this short-term level environmental control, leading to a smart control facility for the greenhouse.

There are still many factors related to a successful realization and application. Some are listed as follows.

1) *The Degree of Acceptance by the Growers:* Even if the terminal CPSS includes both the popular mobile application (App) and classical desktop one, because of habit, education level, and age, growers are not necessarily willing to accept such new platforms. To solve this problem, on the one hand, the amount of data to be provided by growers should technically be as little as possible [2]. Presently, automatic recognition of crop parameters is still being tested in the laboratory [31] but not industrialized. More importantly, mechanism of pushing farming by selling should be constructed, which means the system should help to solve the key problem of farmers (first marketing). The acceptance of users is the basis of producing data, and the latter is important for the personalized updating of parallel systems.
2) *Data Security:* Being a system that integrates both social and physical information, a system aiming at linking various stakeholders, CPSS is leading to a virtual community in the Web. The credit in this community is a

relevant issue, i.e., how to validate the information and how to avoid misuse of the information. For example, in contract farming, how does one ensure that the ordered product is the one planted with high criteria, but not that from any market? Blockchain [32] provides a solution through the smart contract, avoiding distorting the data arbitrarily via decentralized management [33]. Although it is not necessary for a fully trustable community, at least the cost of cheating increases.

3) *The Availability of IoT-Monitored Data:* Although IoT for agriculture is becoming increasingly mature, supplying different sensors, and transmission methods, stability and price are key factors in the promotion of IoT products, especially for low-cost greenhouses. The choice of sensors must be made considering that the monitored data can be used for practical needs. Based on this principle, in this paper, the IoT system was composed only of sensors of temperature and light intensity.

The mentioned components can be further detailed and improved. Currently, to optimize the gross profit, price information is acquired from monitored data, which is to be replaced with a prediction model using the time series analysis method [34] or neuron network algorithm [35]–[37]. In planting schedules, herein we consider the case of single type of crop. In real cases, several kinds of crops are typically planted, and the constraint of continuous cropping must be considered in nonlinear programming [38], [39]. The prediction of crop phenology is to be further combined with the recommendation of fertilization, with the latter also being related to yield prediction. The resulting decision support is to be expressed with knowledge automation; for example, embedding the recommendation into an easy-to-use resource planning application.

Returning to the topic of this paper, why does the design of the system follow a Just-For-Fit Philosophy? First, this system is designed for the solar greenhouse, the construction and maintenance cost of which fit current Chinese farmers. Second, physical information is collected with low-cost and robust sensors, and the collected environmental variables are basic for predicting plant development and growth. In our opinion, collecting but not using information is a waste. The Just-For-Fit idea also lies in the prediction of the greenhouse indoor environment based on outdoor conditions. Even if the cost of IoT facility is low, some greenhouses still may not be equipped. Thus, a trained model for neighboring greenhouses can provide predictions. Fig. 1 shows that the solar greenhouses in a farm are often densely distributed. Guiding production to fit the need of the market is another feature of this paper, either by fitting to the desired amount of product or by fitting to the need in the open market. Both can help decrease waste in labor or production. Suggestions of which crop to plant will be presented in a future work. Taken together, these findings could contribute to the sustainable agricultural production.

VI. Conclusion

An agricultural CPSS has been proposed and illustrated with the management of horticultural production using the solar greenhouse. Based on the collected product price (social information) and environmental data (physical information), the planting dates and area were optimized according to different aims: constant supply or maximum gross profit. The designed system fits the current situation and needs of solar greenhouse production. The application of intelligent management to the traditional solar greenhouse brings vitality to the horticultural production pattern.

References

[1] T. Ojha, S. Misra, and N. S. Raghuwanshi, "Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges," *Comput. Electron. Agricult.*, vol. 118, no. 3, pp. 66–84, 2015.

[2] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Bolú, "A review on the practice of big data analysis in agriculture," *Comput. Electron. Agricult.*, vol. 143, pp. 23–37, Dec. 2017.

[3] X. Yang, L. Hao, L. Yang, and J. Liu, "Research on data crawling and storage system of agricultural product price based on hadoop platform," *Comput. Appl. Softw.*, vol. 34, no. 3, pp. 76–80, 2017.

[4] M. P. L. Lu, X. Du, and M. Andrew, "Disentangling corn price volatility: The role of global demand, speculation, and energy," *J. Agricult. Appl. Econ.*, vol. 44, no. 3, pp. 401–410, 2012.

[5] A. A. Sidhoum and T. Serra, "Volatility spillovers in the Spanish food marketing chain: The case of tomato," *Agribusiness*, vol. 32, no. 1, pp. 45–63, 2016.

[6] Y. Li, J. Li, W. Zhao, and F. Zhao, "Study on early warning of fluctuation in the prices of ginger," *Adv. J. Food Sci. Technol.*, vol. 9, no. 10, pp. 802–806, 2015.

[7] L. H. Gao *et al.*, "Structure, function, application, and ecological benefit of a single-slope, energy-efficient solar greenhouse in China," *Horttechnology*, vol. 20, no. 3, pp. 626–631, 2010.

[8] F.-Y. Wang, "The emergence of intelligent enterprises: From CPS to CPSS," *IEEE Intell. Syst.*, vol. 25, no. 4, pp. 85–88, Jul./Aug. 2010.

[9] B. J. Bailey, "Constraints, limitations and achievements in greenhouse natural ventilation," *Acta Horticulturae*, vol. 534, no. 534, pp. 21–30, 2000.

[10] F.-Y. Wang *et al.*, "Parallel driving in CPSS: A unified approach for transport automation and vehicle intelligence," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 4, pp. 577–587, Mar. 2017.

[11] F.-Y. Wang, "A big-data perspective on AI: Newton–Merton, and analytics intelligence," *IEEE Intell. Syst.*, vol. 27, no. 5, pp. 2–4, Sep./Oct. 2012.

[12] F.-Y. Wang, "Toward a paradigm shift in social computing: The ACP approach," *IEEE Intell. Syst.*, vol. 22, no. 5, pp. 65–67, Sep./Oct. 2007.

[13] H. Y. Wang, J. Hua, H. Lu, and M. Z. Kang, "Collection platform of agricultural product prices-J. J. Farmer," in *Proc. Symp. Intell. Eng. Agricult. Informatization*, 2016, pp. 88–91.

[14] J. Ni *et al.*, "Simulation of the development of tomato in greenhouse," *Scientia Agricultura Sinica*, vol. 38, no. 6, pp. 1219–1225, 2005.

[15] C. A. Medina-Ruuíz, I. A. Mercado-Luna, G. M. Soto-Zarazúa, I. Torres-Pacheco, and E. Rico-García, "Mathematical modeling on tomato plants: A review," *Afr. J. Agricult. Res.*, vol. 6, no. 33, pp. 6745–6749, 2011.

[16] R. Pouteau, J.-Y. Meyer, R. Taputuarai, and B. Stoll, "Support vector machines to map rare and endangered native plants in Pacific Islands forests," *Ecol. Informat.*, vol. 9, pp. 37–46, May 2012.

[17] X.-R. Fan, M. Z. Kang, E. Heuvelink, P. de Reffye, and B.-G. Hu, "A knowledge-and-data-driven modeling approach for simulating plant growth: A case study on tomato growth," *Ecol. Model.*, vol. 312, pp. 363–373, Sep. 2015.

[18] X.-R. Fan *et al.*, "A knowledge-and-data-driven modeling approach for simulating plant growth and the dynamics of CO_2/O_2 concentrations in a closed system of plants and humans by integrating mechanistic and empirical models," *Comput. Electron. Agricult.*, vol. 148, pp. 280–290, May 2018.

[19] M. Z. Kang, E. Heuvelink, S. M. Carvalho, and P. de Reffye, "A virtual plant that responds to the environment like a real one: The case for chrysanthemum," *New Phytol.*, vol. 195, no. 2, pp. 384–395, 2012.

[20] F.-Y. Wang, "CC 5.0: Intelligent command and control systems in the parallel age," *J. Command Control*, vol. 1, no. 1, pp. 107–120, 2015.

[21] I. Ioslovich, P.-O. Gutman, and R. Linker, "Hamilton–Jacobi–Bellman formalism for optimal climate control of greenhouse crop," *Automatica*, vol. 45, no. 5, pp. 1227–1231, 2009.

[22] D. Krenczyk and M. Jagodzinski, "ERP, APS and simulation systems integration to support production planning and scheduling," in *Proc. 10th Int. Conf. Soft Comput. Models Ind. Environ. Appl.*, 2015, pp. 451–461.

[23] R. Qin, S. Zeng, J.-J. Li, and Y. Yuan, "Parallel enterprises resource planning based on deep reinforcement learning," *Acta Automatica Sinica*, vol. 43, no. 9, pp. 1588–1596, 2017.

[24] H. Gijzen *et al.*, "HORTISIM: A model for greenhouse crops and greenhouse climate," *Acta Horticulturae*, vol. 456, pp. 441–450, Jan. 1998.

[25] E. Heuvelink, "Evaluation of a dynamic simulation model for tomato crop growth and development," *Ann. Botany*, vol. 83, no. 4, pp. 413–422, 1999.

[26] L. F. M. Marcelis, "A simulation model for dry matter partitioning in cucumber," *Ann. Botany*, vol. 74, no. 1, pp. 43–52, 1994.

[27] L. F. M. Marcelis *et al.*, "Modelling dry matter production and partitioning in sweet pepper," in *Proc. 3rd Int. Symp. Models Plant Growth Environ. Control Farm Manag. Protected Cultivation ISHS Acta Horticulturae (HortiModel)*, vol. 718, 2006, pp. 121–128, doi: 10.17660/ActaHortic.2006.718.13.

[28] J. B. Evers, V. Letort, M. Renton, and M. Kang, "Computational botany: Advancing plant science through functional–structural plant modelling," *Ann. Botany*, vol. 121, no. 5, pp. 767–772, 2018.

[29] T.-W. Chen, H. Stützel, and K. Kahlen, "High light aggravates functional limitations of cucumber canopy photosynthesis under salinity," *Ann. Botany*, vol. 121, no. 5, pp. 797–807, 2018.

[30] E. J. V. Henten and J. Bontsema, "Time-scale decomposition of an optimal control problem in greenhouse climate management," *Control Eng. Pract.*, vol. 17, no. 1, pp. 88–96, 2009.

[31] Y. Qu, J. Wang, J. Dong, and F. Jiang, "Design and experiment of crop structural parameters automatic measurement system," *Trans. Chin. Soc. Agricult. Eng.*, vol. 28, no. 2, pp. 160–165, 2012.

[32] Y. Yuan and F.-Y. Wang, "Parallel blockchain: Concept, methods and issues," *Acta Automatica Sinica*, vol. 43, no. 10, pp. 1703–1712, 2017.

[33] J. Hua, X. J. Wang, M. Z. Kang, H. Y. Wang, and F.-Y. Wang, "Blockchain based provenance for agricultural products: A distributed platform with duplicated and shared bookkeeping," in *Proc. IEEE/IFAC Conf. Blockchain Knowl. Autom. (ICBKA)*, 2018, pp. 97–101.

[34] H. C. Yuan and Y. H. Xin, "Oil price prediction model based on multiple regression analysis and time series analysis," *Sci. Technol. Ind.*, vol. 9, pp. 76–77, Apr. 2011.

[35] C. Luo, Q. Wei, L. Zhou, J. Zhang, and S. Sun, "Prediction of vegetable price based on neural network and genetic algorithm," in *Computer and Computing Technologies in Agriculture IV*. Berlin, Germany: Springer, 2011, pp. 672–681.

[36] D.-P. Li, Y.-J. Liu, S. Tong, C. L. P. Chen, and D.-J. Li, "Neural networks-based adaptive control for nonlinear state constrained systems with input delay," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2018.2799683.

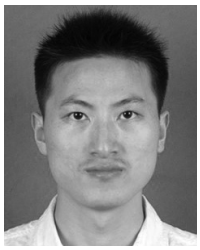
[37] L. Liu, Z. Wang, and H. Zhang, "Adaptive fault-tolerant tracking control for MIMO discrete-time systems via reinforcement learning algorithm with less learning parameters," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 299–313, Jan. 2017.

[38] Y.-J. Liu, M. Z. Gong, S. C. Tong, C. L. P. Chen, and D.-J. Li, "Adaptive fuzzy output feedback control for a class of nonlinear systems with full state constraints," *IEEE Trans. Fuzzy Syst.*, to be published, doi: 10.1109/TFUZZ.2018.2798577.

[39] Y.-J. Liu *et al.*, "Adaptive control-based barrier Lyapunov functions for a class of stochastic nonlinear systems with full state constraints," *Automatica*, vol. 87, pp. 83–93, Jan. 2018.

Mengzhen Kang (M'16) received the Ph.D. degree in pattern recognition and intelligent systems from the Institution of Automation, Chinese Academy of Sciences, Beijing, China, in 2003.

He is an Associate Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. From 2005 to 2006, she was a Post-Doctoral Fellow with Projet DigiPlante, INRIA Rocquencourt, France, and HPC, Wageningen University, Wageningen, The Netherlands. Her current research interests include parallel agriculture and computational biology. She has published over 60 papers in the above areas.



Xing-Rong Fan received the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2016, under the supervision of Prof. B.-G. Hu.

From 2016 to 2018, he was a Lecturer in computer science with the Chongqing University of Posts and Telecommunications, Chongqing, China. His current research interests include machine learning and plant growth modeling.

Jing Hua received the Ph.D. degree in applied computer science from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2012.

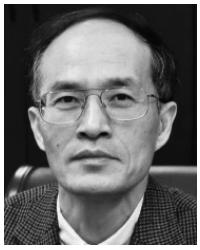
He is currently an Assistant Professor in computer science with the Institute of Automation, Chinese Academy of Sciences. His current research interest includes computer science and agriculture, including virtual plant modeling, smart agriculture, programming languages, distributed computer systems, and computer graphics.

Haoyu Wang received the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2014.

He is currently an Assistant professor in computer science with the Institute of Automation, Chinese Academy of Sciences. His current research interest includes computer science and agriculture, including virtual plant modeling, smart agriculture, programming languages, and information system.

Xiujuan Wang received the Ph.D. degrees in applied mathematics from Ecole Central, Écully, France, and in soil resource use from China Agricultural University, Beijing, China, in 2011.

She is currently an Assistant Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing. Her current research interests include plant growth modeling and parallel agriculture.



Fei-Yue Wang (S'87–M'89–SM'94–F'03) received the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

He is the Director of the State Key Laboratory of Management and Control for Complex Systems. His current research interests include methods and applications for parallel systems, social computing, and knowledge automation.

Dr. Wang was a recipient of the 2nd Class National Prize in Natural Sciences of China in 2007, the Outstanding Scientist by ACM for his research in intelligent control and social computing, the IEEE ITS Outstanding Application and Research Awards in 2009 and 2011, and the IEEE SMC Norbert Wiener Award in 2014. He is currently an Editor-in-Chief of *China's Journal of Command and Control*. Since 1997, he has been serving as the General or the Program Chair of over 20 IEEE, INFORMS, ACM, and ASME conferences. Since 2008, he has been the Vice President and the Secretary General of the Chinese Association of Automation. He is an Elected Fellow of INCOSE, IFAC, ASME, and AAAS.