A digital twin for arable and dairy farming

Precision Agriculture '21

Evert, F.K.; Berghuijs, H.N.C.; Hoving, I.E.; Wit, A.J.W.; Been, T.H.

https://doi.org/10.3920/978-90-8686-916-9_110

110. A digital twin for arable and dairy farming

F.K. van Evert[1*], H.N.C. Berghuijs[2], I.E. Hoving[3], A.J.W. de Wit[4] and T.H. Been[1]
[1]*Agrosystems Research, Wageningen University & Research, P.O. Box 16, 6700 AA Wageningen, the Netherlands; frits.vanevert@wur.nl*
[2]*Plant Production Systems, Wageningen University & Research, P.O. Box 430, 6700 AK Wageningen, the Netherlands*
[3]*Livestock & Environment, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands*
[4]*Earth Observation and Environmental Informatics, Wageningen University & Research, P.O. Box 47, 6700 AA Wageningen, the Netherlands*

Abstract

The Digital Future Farm (DFF) is a modeling framework that allows models of arable and dairy farms to be assembled from sub-models such as crop, soil and livestock models. The DFF is also a digital twin (DT): a model of a physical object with emphasis on (1) the connection between the real-world object and its virtual counterpart and (2) the use of real-time data from sensors to keep the model synchronized. In this study, an Ensemble Kalman Filter was used to synchronize a grass model and a potato model in the DFF with observations made in experiments. Results indicate that special care must be taken to prevent divergence of the Ensemble Kalman Filter when it is used with a crop growth model.

Keywords: crop growth model, data assimilation, observational data, recommendations for farmers, digital twin

Introduction

In precision agriculture, farmers need precise, real-time information about the status of crops, soils and livestock, as well as information about the likely outcome of management decisions. Demand is growing for dynamic crop growth models (CGM) in helping farmers to make decisions about, for example, irrigation and the application of N fertilizer. Akkerweb (Van Evert *et al.*, 2018) and Beregeningssignaal (https://www.zlto.nl/beregeningssignaal) are web-based platforms where CGMs are used to monitor commercial farms. This paper focuses on two requirements for the operational deployment of CGMs.

The first requirement arises from the fact that a CGM typically consists conceptually of sub-models: a sub-model for the crop, one for the soil water balance and another one for soil nitrogen. There are often two or more alternatives for each sub-model, each alternative having strengths and weaknesses. For example, one soil water balance model may use a simplified representation of soil water dynamics, compute quickly, but perform well only on sandy soils; while a second soil water balance model implements Richards' equation and is capable of representing capillary rise from a shallow groundwater table, but computes more slowly. In an operational setting, one must be able to use the simple model where possible and the more complex model where necessary. One way to achieve this is a framework that allows interchangeable sub-models.

The second requirement arises from the widespread availability of real-time observational data at the field level, particularly remote sensing data obtained from satellites, but also, for example, from tractor-mounted equipment and from sensors installed in the soil. This observational data must be combined with the output from simulation models to achieve the best possible estimate of the status of livestock, crops, and soils.

DOI 10.3920/978-90-8686-916-9_110, © Wageningen Academic Publishers 2021

In this paper the Digital Future Farm (DFF) is introduced. The DFF is a modeling framework that allows models of arable and dairy farms to be assembled from sub-models such as crop, soil and livestock models. The DFF is also a digital twin (DT): a model of a physical object with emphasis on (1) the connection between the real-world object and its virtual counterpart and (2) the use of real-time data from sensors to keep the model synchronized (Grieves, 2014). In this paper, the framework is described and preliminary results from data assimilation with the framework are given.

Materials and methods

Simulation framework

The DFF is based on the MODCOM simulation framework (Hillyer *et al.*, 2003), which allows linking of sub-models and handles numerical integration, events and communication between sub-models. Essentially, MODCOM is an implementation of the Discrete Event System Specification (DEVS) (Zeigler, 1976). The DFF used the .NET (https://dotnet.microsoft.com/) version of MODCOM which was developed in the EU-funded project SEAMLESS (Van Evert and Lamaker, 2007) and which inspired both SIMPLACE (Enders *et al.*, 2010) and BioMa (Donatelli *et al.*, 2012).

Grass model

The grass model is programmed in Delphi (https://www.embarcadero.com) and is made available as a Component Object Model (COM) class (https://docs.microsoft.com/en-us/windows/win32/com/ component-object-model--com--portal) (Vellinga, 2006; Vellinga *et al.*, 2004).
The model takes an XML-string as input and produces another XML-string as output. The input string contains all input necessary to run the simulation: soil parameters, grass parameters, management (fertilization, grazing, mowing), weather and depth to groundwater.
A simulation with the grass model must start on 1 January and can run to any time in the year but, once stopped, it cannot be resumed. The implementation of the grass model does not fit neatly with the framework of the DFF where each sub-model takes one time-step, exchanges information with other sub-models, and only then takes the next time-step. This mismatch was solved by writing a wrapper class which at each time-step creates the XML-input string for the grass model and runs the model from 1 January until the current simulation time. Following that, the XML output string is used to update the state of the wrapper. Events such as fertilizer application, grazing and mowing are captured by the wrapper and transformed into a format understood by the grass model.

Potato model

WOFOST is a simulation model for the quantitative analysis of the growth and production of annual field crops (De Wit *et al.*, 2019). It is a mechanistic, dynamic model that explains daily crop growth based on the underlying processes, such as photosynthesis, respiration and how these processes are influenced by environmental conditions.
The reference version of WOFOST is written in Python (https://www.python.org/). A wrapper class was written which uses IronPython (https://ironpython.net/) to call the original Python code of WOFOST. Fortunately, WOFOST can be stopped after every time-step; the wrapper uses this functionality to instruct WOFOST to perform a single time-step whenever the simulation time in the framework is advanced. Running WOFOST in DFF in this way is not much slower than running it stand-alone. Events such as fertilizer application and haulm killing are captured by the wrapper and translated to WOFOST events.

Weather

Weather data are provided by a sub-model which reads data from file and which can be linked to the other sub-models. The grass model and the potato model both make use of this weather model.

Data assimilation

Models are imperfect due to limitations in the model structure and processes, uncertainty about model parameters and initial conditions, as well as uncertainty about the driving variables that the models require. Therefore, the simulated states often diverge from reality and this uncertainty tends to increase with every time-step of the model. Particularly when the digital twin is used to forecast the impact of management decisions, this can become a serious problem for the reliability of the outputs from the digital twin.

Data assimilation provides a tool to adjust the model whenever observations are available that provide information on the system state. Data assimilation considers the uncertainty in observations (how likely are outliers) as well as the uncertainty in simulation outcome. The combined 'knowledge' of the simulated system plus the observations allow to make a better estimate of the state of the system and consequently adjust the model states and reduce uncertainty on the system. Data assimilation algorithms have been highly successful in many disciplines, including numerical weather prediction, robotics and many other engineering areas. Three main methods of data assimilation are distinguished, namely calibration (where model parameters are adjusted), forcing (where observations replace one or more state variables that would otherwise be simulated), and updating (where observations and simulations are combined into a new, optimal estimation of the state of the system) (e.g. Jin *et al.*, 2018). Here, the Ensemble Kalman Filter (EnKF) method (De Wit and Van Diepen, 2007; De Wit *et al.*, 2012), which is an updating method, was implemented for the DFF.

Data

Data on crop variables (LAI, biomass, grass height, etc.) measured using destructive sampling were used. For grasslands, data from a four-year experiment with grass in the Netherlands were used (Hoving *et al.*, 2019). For potato, data from a two-year experiment in the Netherlands were used (Ten Den *et al.*, 2020).

Results

Grass

The grass model is well-calibrated for the Netherlands but there is a parameter to which the model is sensitive and the value of which is difficult to quantify. This parameter specifies the annual amount of soil organic nitrogen that is mineralized. For the peat soil for which the simulation was run, a standard value is 250 kg/ha. The ensemble of simulations was generated by drawing values for this parameter from a Normal distribution with mean=250 and sd=50. To demonstrate the effect of data assimilation, it was assumed that grass yield was accurately observed on 18 April, 30 May, 10 July and 21 August, and that the standard error of measurement was either 10 or 20%. Figure 1 shows that observations with a small standard error (s.e.) have a larger effect on the ensemble than the more uncertain observations.

WOFOST Potato

The WOFOST model was calibrated for potential production level using the 2019 field trials for cultivar Innovator. Despite the calibration efforts, the WOFOST model still showed considerable discrepancies between observed and simulated results for this cultivar (Brouwer, 2020). Particularly, leaf biomass and leaf area index during the second half of the growth cycle were overestimated leading to late senescence and an overestimation of the biomass and tuber growth rate. These discrepancies are probably related to structural aspects of the model which, being a generic crop growth model, may not properly represent this potato cultivar. It was hypothesized that assimilating the leaf area index by means of an EnKF will lead to simulation results that correspond more closely to the measurements.

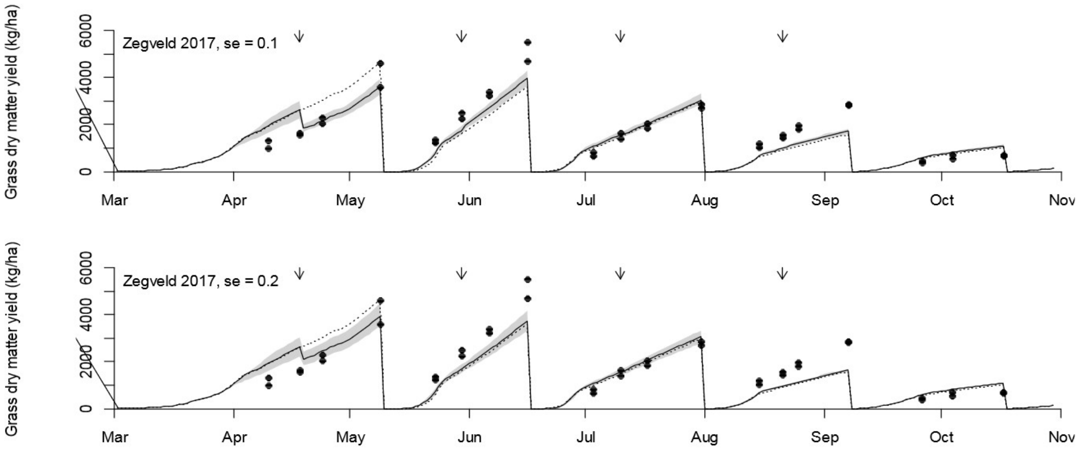

Figure 1. Data assimilation with the grass model for an experiment in Zegveld in 2017 where total N applied was 280 kg/ha. Filled symbols: destructive measurement of grass dry matter yield. Dotted line: simulation with the calibrated model. Black line: median of ensemble simulations. The grey area delineates the 0.1 and 0.9 quantiles of the ensemble simulation. Arrows: indicate when observations were assimilated. Top: s.e. of observations set to 10% of the mean. Bottom: s.e. of observations set to 20% of the mean.

Setting up the EnKF requires estimating the variance of measurements and of model predictions. Variance of measured LAI was estimated from the three replicates in the field trial. The estimated variance varies over the growth cycle with low variance in the first half and larger variance in the second half of the growth cycle (Figure 2).

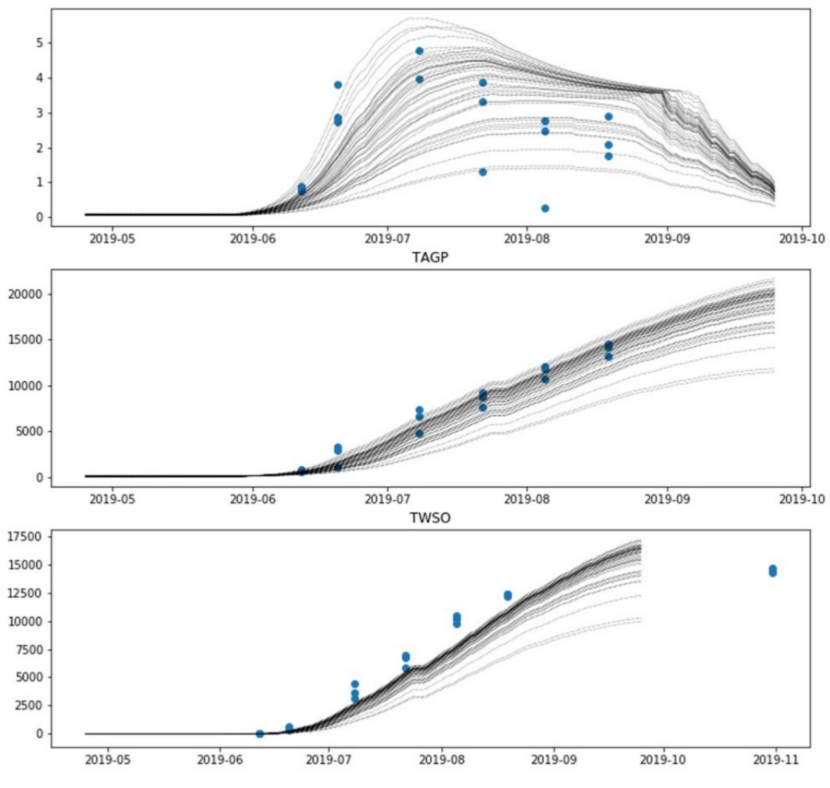

Figure 2. Results of the WOFOST Ensemble (dotted lines) for LAI, total biomass (TAGP in kg/ha) and tuber yield (TWSO in kg/ha) as well as the observed values from the field trials.

The variance of model predictions was estimated by sampling from the distributions of two important model parameters with large uncertainty: specific leaf area (SLATB), and the maximum relative growth rate for leaf area (RGRLAI). Specific leaf area was measured during the experiments and was estimated as 0.00191 ha/kg with a standard deviation of 18%. The default RGRLAI value for potato is 0.012 d^{-1} with an assumed standard deviation of 0.002. Figure 2 shows the resulting variability in the model ensemble together with the observed biomass and LAI.

The results of the assimilation of LAI with the EnKF were analysed by evaluating the resulting LAI trajectories and the impact of the EnFK on the predicted values. The results demonstrate that, despite the large variability in the ensemble, the assimilation of LAI has a limited impact, particularly during the second half of the growth cycle (Figure 3). There is a strong impact of the first set of LAI observations which has very small variance and therefore the ensemble collapses after the first assimilation interval. Although the variance in the ensemble increases again, the ensemble converges through the growth cycle. This causes the EnKF to almost completely ignore the observations from mid-July onward through the combination of low ensemble variance and high measurement variance. This phenomenon of the EnKF is called 'filter divergence' and is a known problem with crop models whose ensembles tend to collapse as the growing season continues (Huang *et al.*, 2019).

A practical solution to the problem of filter divergence is the introduction of a 'variance inflation factor' (Huang *et al.*, 2016) which enlarges the Kalman gain in the EnKF equation in order to maintain variance in the ensemble. Experimentation with the variance inflation factor shows that a value of at least 10 is required to ensure that LAI is updated based on the observed values (Figure 4). Simulation of total above-ground production (TAGP) improves and so does the final predicted tuber weight (TWSO). The time-trajectory of tuber yield still shows deviations. However, this could also be related to partitioning of newly formed dry matter.

Discussion

With both models, parameters were used as source of variation in the ensemble. For the grass model, the soil mineral nitrogen supply parameter had been calibrated earlier but it is appropriate to use filtering to incorporate year-to-year or site-to-site variation. For the potato model, the uncertainty in the parameters was related to the cultivar and in this case also it is appropriate to use filtering. For both models, uncertainty does not only consist of parameter uncertainty but also includes structural uncertainty due to model limitations; however, the EnKF was used only to update the state of the model. In future work, an effort will be made to update model parameters as well as model state.

The EnKF was implemented directly. The use of a library such as OpenDA (https://www.openda.org/) was investigated. However, this library was not used because it was estimated that it would

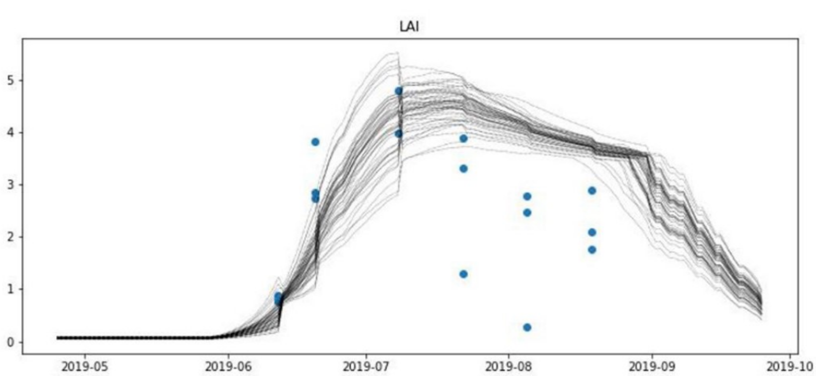

Figure 3. WOFOST model simulated LAI with assimilation of observations using the EnKF.

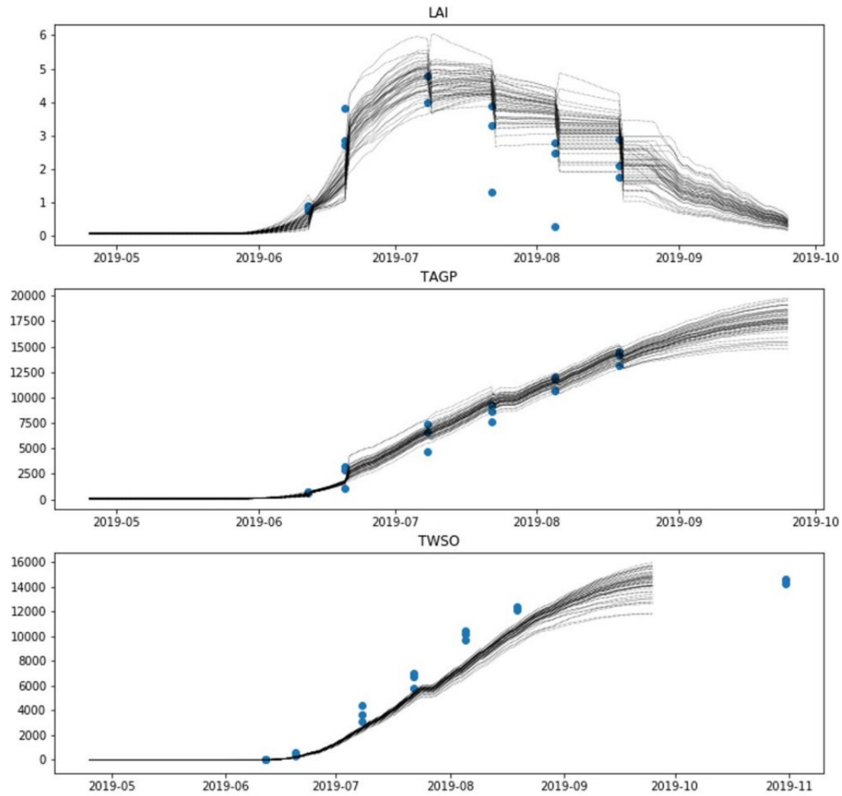

Figure 4. Results from assimilating LAI with the EnKF using a variance inflation factor of 10 for LAI, total biomass (TAGP in kg/ha) and tuber yield (TWSO in kg/ha) as well as the observed values from the field trials.

take less time to implement the EnKF than it would take to link to OpenDA. If, in future work, the need arises to use other data assimilation methods, then this decision will have to be reconsidered.

Conclusions

The Digital Future Farm (DFF) provides uniform access to several models that are commonly used at Wageningen UR. An Ensemble Kalman Filter was implemented for the DFF and tested. Results indicate that special care must be taken to prevent divergence of the Ensemble Kalman Filter when it is used with a crop growth model.

Acknowledgements

This project was funded by the Digital Twin investment theme of the Wageningen University & Research Strategic Plan 2019–2022.

References

Brouwer, S. (2020). Calibration of WOFOST by using experimental data to simulate potential production of five Dutch potato cultivars. MSc thesis, Plant Production Systems, Wageningen University. Wageningen, the Netherlands.

De Wit, A., Boogaard, H., Fumagalli, D., Janssen, S., Knapen, R., van Kraalingen, D. *et al.* (2019). 25 years of the WOFOST cropping systems model. Agricultural Systems 168, 154-167.

De Wit, A., Duveiller, G. and Defourny, P. (2012). Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. Agricultural and Forest Meteorology 164, 39-52. https://doi.org/10.1016/j.agrformet.2012.04.011.

De Wit, A.J.W. and van Diepen, C.A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. Agricultural and Forest Meteorology 146, 38-56. https://doi.org/10.1016/j.agrformet.2007.05.004.

Donatelli, M., Cerrani, I., Fanchini, D., Fumagalli, D. and Rizzoli, A.-E. (2012). Enhancing model reuse via component-centered modeling frameworks: the vision and example realizations. In: R. Seppelt, A.A. Voinov, S Lange and D. Bankamp (Eds). Congress on Environmental Modelling and Software. 140. Leipzig, Germany: International Environmental Modelling and Software Society (iEMSs).

Enders, A., Diekkrüger, B. Laudien, R. Gaiser, T. and Bareth, G. (2010). The IMPETUS spatial decision support systems. Impacts of Global Change on the Hydrological Cycle in West and Northwest Africa. Cham, Switzerland: Springer. pp. 360-393.

Grieves, M. (2014). Digital twin: Manufacturing excellence through virtual factory replication. White paper. Available online at https://www.researchgate.net/publication/275211047 (accessed 22 February 2021).

Hillyer, C., Bolte, J., van Evert, F. and Lamaker, A. (2003). The ModCom modular simulation system. European Journal of Agronomy 18, 333-343.

Hoving, I.E., van Riel, J., Holshof, G., Plomp, M., Agricola, S., van Boheemen, K. *et al.* (2019). Schatten van grasopbrengst op basis van spectrale reflectie, grashoogte en modellering: onderzoeksresultaten van een maaiproef op zand-klei en veengrond 2016-2017. (Estimating grass yield based on spectral reflection, grass height and modeling: research results of a mowing trial on sandy clay and peat soil 2016-2017) Wageningen: Wageningen Livestock Research. https://doi.org/10.18174/508117.

Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., *et al.* (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. Agricultural and Forest Meteorology 276-277, 107609. https://doi.org/10.1016/j.agrformet.2019.06.008.

Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., *et al.* (2016). Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. Agricultural and Forest Meteorology 216, 188-202. https://doi.org/10.1016/j.agrformet.2015.10.013.

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., *et al.* (2018). A review of data assimilation of remote sensing and crop models. European Journal of Agronomy 92, 141-152. https://doi.org/10.1016/j.eja.2017.11.002.

Ten Den, T., van de Wiel, I., de Wit, A., van Evert, F., van Ittersum, M. and Reidsma, P. (2020). The effect of potato cultivar differences on parameters in WOFOST. In: iCROPM 2020 Book of Abstracts: Second international Crop Modelling Symposium. Montpellier: CIRAD/INRA. P 319-320.

Van Evert, F.K., Been, T.H., Booij, J.A., Kempenaar, C., Kessel, G.J.T. and Molendijk, L.P.G. (2018). Akkerweb: a platform for precision farming data, science, and practice. In: 14th International Conference on Precision Agriculture. Available at https://ispag.org/proceedings/?action=abstract&id=5059 (accessed 22 February 2021).

Van Evert, F.K. and Lamaker, E.J.J. (2007). The MODCOM framework for component-based simulation. In: M. Donatelli, J. Hatfield and A.E. Rizzoli, (eds.) Farming Systems Design 2007, International Symposium on Methodologies on Integrated Analysis on Farm Production Systems. Catania, Scicily, Italy: La Goliardica Pavese s.r.l. pp. 223-224.

Vellinga, T.V. (2006). Management and nitrogen utilisation of grassland on intensive dairy farms. Available at https://library.wur.nl/WebQuery/wurpubs/348625. Wageningen University and Research, Wageningen, the Netherlands.

Vellinga, T.V., André, G., Schils, R. and Oenema, O. (2004). Operational nitrogen fertilizer management in dairy farming systems: identification of criteria and derivation of fertilizer application rates. Grass and Forage Science 59, 364-377.

Zeigler, B.P. (1976). Theory of modelling and simulation. New York, NY, USA: John Wiley & Sons Inc.