

# Support Tools for Agricultural Production Simulation Processing

Jan Pavlik<sup>1</sup>, Jiri Vanek<sup>1</sup>, Jan Masner<sup>1</sup>, Michal Stoces<sup>1</sup>, Vladimir Ocenasek<sup>1</sup>

<sup>1</sup>Department of Information Technologies, Faculty of Economics and Management, Czech University of Life Sciences Prague, Czech Republic

**Abstract.** The APSIM software has proven to be extremely valuable decision support tool when it comes to optimizing agricultural management practices in order to maximize yield. Conducting high amount of simulations naturally requires processing of large volumes of data, therefore the available time and hardware resources create a limit for the scale of production simulation modelling. If APSIM is to be effectively used at a local level utilizing pre-existing hardware infrastructure, an assessment of available resources must be conducted in order to optimize the scale of the simulation. Another issue is the availability of personnel with enough information technology skills and experience to conduct the processing. The focus of this paper are software automation and other assistance tools that are therefore required for production modelling to be successfully utilized by small to medium enterprises.

**Keywords:** APSIM; scalability; hardware requirements; data processing; automation software.

## 1 Introduction

Maximizing yields of agricultural production is one the critical issues for society today. Growing worldwide population exacerbates the need for sufficient food production while increasing climate anomalies such as droughts can pose a great risk for crops. One of the approaches to maintain and improve agricultural yields is to utilize information technology to optimize managerial strategies and genotype selection by conducting multi-factor analysis in form of simulations or modelling (Holzworth et al., 2014).

The development of information technology hardware provides increasingly more technological resources to conduct simulation processing on larger scales, however as shown by (Li and Li, 2014) the increase of available data, such as higher resolution of geographical data, creates hardware limitations when scaling up the processing. This is especially important when trying to utilize agricultural simulations on a local level, for instance in small to medium agricultural companies. Due to the lack of financial resources for purchasing dedicated hardware there is a need to utilize pre-existing infrastructure. Most of the computing capacity in these companies is provided by out of date machines, meaning that any large-scale data processing involving high level of

---

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Proceedings of the 9th International Conference on Information and Communication Technologies in Agriculture, Food & Environment (HAICTA 2020), Thessaloniki, Greece, September 24-27, 2020.

parallelization such as described by (Zhao et al., 2013) is out of the question. The main approach to utilize such hardware requires individual machine optimization alongside parallelization as shown by (Bartonek, 2017). Another option would be to utilize cloud computing, but as pointed out by (Szufel, Czupryna and Kaminski, 2017) it necessary to highly optimize cloud processing in order to maintain low costs.

Second issue is the lack of qualified employees. Especially in agriculture there is a distinct lack of IT proficient workers as pointed out by (Reinmuth and Dabbert, 2017). This results in a need for easy to use support tools that would automate the simulation setup and processing. Some of these tasks are already incorporated within the APSIM software and as stated by (Holzworth et al., 2018) the ease of use and focus on automation will be one of the integral parts of next versions of APSIM.

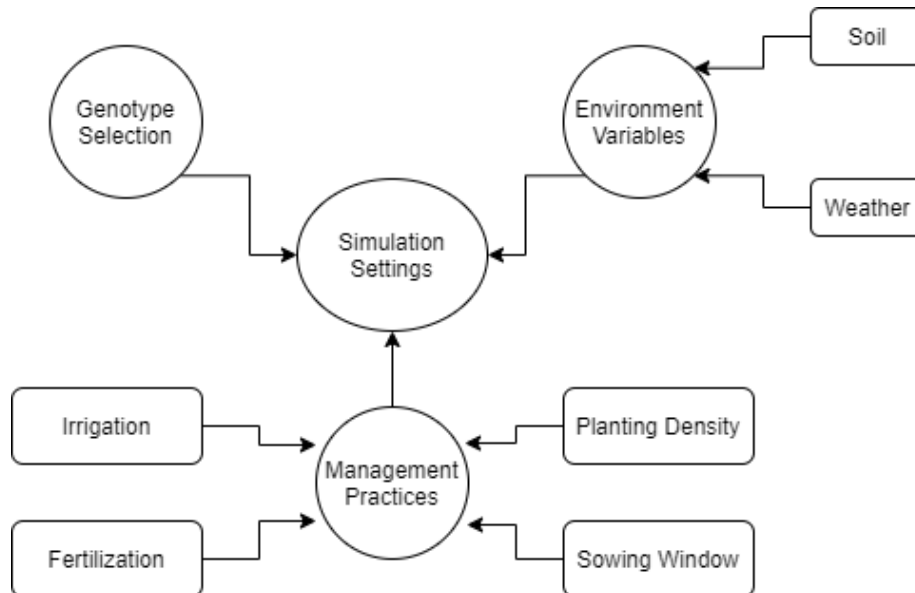
## 2 Simulation Workflow

As shown in Figure 1, the main workflow of agricultural simulation processing can be divided into four main steps. Firstly, it is necessary to establish the correct scale of the processing. When adjusting the number of options for each factor, the total number of simulations that needs to be processed changes accordingly. Since the number of simulations is essentially a product of the number of varying options, restraining variability only to the most relevant settings is the main way to reduce the processing time.



**Fig 1.** Basic simulation workflow

The factors and options that are determined in this first stage can be for instance different soil types, plant genotypes, different weather condition scenarios, and settings involving various managerial practices such as time of sowing, time of harvest, irrigation, fertilization, number of plants per meter square etc. (see Figure 2).



**Fig 2.** Basic concept of GxExM framework (genotype, environment, management)

Second step is the data preparation and preprocessing. Each simulation is stored as a single xml file. When APSIM is being deployed on dedicated hardware, is it possible to generate these files “on the fly” during the processing. However as pointed out by (Jarolimek et al., 2019) the hardware components, mainly processor and RAM, constitute a limit on to how many simulations at once can APSIM handle. Therefore, in order to maximize effectiveness, it is necessary to optimize simulation batch sizes on a per machine basis, that is why the data preparation and preprocessing step is unavoidable when trying to utilize sub-par preexisting hardware in smaller or medium enterprises.

The actual simulation processing itself should be conducted during downtimes such as nights, when the infrastructure is not required for other critical operations. This part is generally the most time consuming and therefore also the most likely to significantly improve the overall efficiency if automation and scheduling is utilized. The hardware dependence of the processing can result in bottlenecks if the previous two steps were not conducted properly.

The analysis of results is essentially a statistical data analysis and can therefore be conducted using tools like MS Excel or more specialized software such as SAS, STATISTICA etc. The most basic analysis would consist of simply taking the simulations that produced highest yields and finding commonalities in the simulation settings.

## **3 Automation and support tools**

### **3.1 Simulation Settings**

The number of total simulations that needs to be processed determines the scale of the processing. The calculation of the number of simulations is a simple multiplication of options for each simulation settings. The scale can be therefore decreased or increased by simply adjusting the number of options. When scaling up, it is possible to include more options or “what-if” scenarios and generally explore more combinations. When there is a need to scale down, the most likely best approach is to use hands on experience of local producers or company agronomists to narrow down the simulation option settings only to a few variations that were historically most effective or make most sense in terms of the common local managerial practices.

Therefore, the important question in step one is how many simulations we should aim for. In order to properly select the scale of simulation an estimate must be made, based on allotted time and the hardware available for the computation. And it is in this point where the absence of experienced IT employees creates first problems. Unlike bigger companies or corporations that might have dedicated IT departments, SMEs generally lack specialists that would be able to adequately estimate the capabilities of older hardware when it comes to processing large number of simulations.

The support tools for this step could be a simple web application where the user inputs basic hardware information of their machines including processing power and RAM and the application will estimate number of simulations that can be run per day or per hour.

### **3.2 Data Preparation**

The main goal of this step is to gather all input information for the APSIM software. There is a possibility for integration between APSIM and existing agricultural software to partially automate this process, similar to that outlined by (Skoogh, Michaloski and Bengtsson, 2010). Another option is to utilize integration to existing knowledge databases such as soil or weather databases as explored by (Kim et al., 2018). But due to the lower scale and therefore limited number of option settings, this is not necessary, since manual entry of the input data is not very time consuming.

The other part of preprocessing consists of generating the APSIM simulation files beforehand and grouping them into various size batches optimized on a per machine basis. The degree of detail this task needs to be performed to depends on the sophistication of the software automation tools used in the following step. As shown by (Pavlik et al., 2019) it is possible to develop an application that can combine some of the work required in steps two and three and handle both the batch APSIM file generation and the processing scheduling and automation.

### 3.3 Simulation Processing

There are four basic approaches to automate the simulation processing:

1. Use built-in APSIM capabilities
2. Process simulations ex-situ – on the cloud
3. Use existing software tools for automation
4. Develop new software specifically designed to automate APSIM simulations

As explained earlier, using existing APSIM options for automation might not be possible due to the hardware limitations. Processing on the cloud is in essence similar to purchasing dedicated hardware. It might be cheaper, but in this paper, we are focusing mainly on exploiting already existing hardware and infrastructure. This leaves us with options three and four.

There are many existing software tools to automate tasks. Whether it be task schedulers that already come with the operating system, or dedicated automation software, such as HTCondor. The advantage of using such tools is that they include many useful functions such as workload monitoring, virtualization, checkpointing and can also combine serialized batch processing with parallelization. Therefore, this option is preferable when used on hardware that also performs other day to day tasks. The main disadvantage is that the previous and following steps (data preprocessing and result analysis) will require more work since they cannot be incorporated into these existing tools, not even partially.

The last approach is to design and develop brand new tools specifically to automate APSIM simulation processing. A custom-built tool could potentially encompass more than just the processing automation. It can theoretically handle all four steps of the workflow, bridging the gaps between the parts. This could be very valuable considering the lack of experienced IT employees in smaller agricultural companies. However, it will come at a cost of lower sophistication of automation and it will be harder to combine the processing with other necessary tasks the hardware needs to perform on a daily basis. This approach is therefore better suited for situations when the company can set aside one or several computers, perhaps older or currently unused machines, and fully use them towards the simulation processing.

### 3.4 Analysis of Results

If the analysis of the results is to be conducted in a separate statistical software, it is necessary to convert the output simulation from the basic text format into .csv or .xls. This can be achieved with an extraction program or a script that will parse the output files and siphon only the necessary data.

However, when simulating agricultural production on a local level, the only important outputs are the simulation settings associated with the highest yields. A complex statistical analysis may therefore not be necessary. In case of developing new custom software for the previous processing steps, the extraction of yield data, selection of top simulations and any eventual visualization of simulation settings can be added to it, resulting in an overarching support tools that can automate vast majority of the workflow processes.

## 4 Conclusions

The paper discussed two main approaches to conducting agricultural production simulations in smaller or medium companies. The first approach is to split the process into logical parts and optimize and automate them separately, utilizing either already existing software tools, or developing a smaller programs for individual tasks, such as format conversion tools, simulation generation tools, data parsers etc. The main advantages of this approach are higher optimization and better interoperability when utilizing existing hardware that cannot be fully dedicated to the task. However, such a solution would require employee skilled in IT, prompting additional need for financial resources in order to hire or train someone.

The second approach consists of developing an overarching support tool that would incorporate all the various processing tasks. If developed as a general purpose streamlined software with focus on ease of use, it could overcome the problem with lack of experienced IT workers in small to medium companies. The disadvantages of this solution are harder incorporation with existing agricultural software running parallel on the same machines and loss of modularity options such as in-depth data analysis or conducting simulations for goals other than maximizing yields.

**Acknowledgment:** The results and knowledge included herein have been obtained owing to support from the following institutional grant. Internal Grant Agency of the Faculty of Economics and Management, Czech University of Life Sciences in Prague, grant no. 2019MEZ0005 – “Optimization of management practices in sorghum production under uncertain future weather conditions”.

## References

1. Bartonek, D. The Possibilities of Big GIS Data Processing on the Desktop Computers (2017). RISE OF BIG SPATIAL DATA Book Series: Lecture Notes in Geoinformation and Cartography, pp. 273-287. DOI: 10.1007/978-3-319-45123-7\_20
2. Holzworth, D., Huth, N. I., de Voil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G. et al. (2014) “APSIM - Evolution towards a New Generation of Agricultural Systems Simulation.” Environmental Modelling & Software, Vol. 62, pp. 327-350. ISSN 1364-8152. DOI 10.1016/j.envsoft.2014.07.009.
3. Holzworth, D., Huth, N. I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N. I., Zheng, B. and Snow, V. (2018) “APSIM Next Generation: Overcoming Challenges in Modernising a Farming Systems Model”, Environmental Modelling & Software, Vol. 103, pp. 43-51. ISSN 1364-8152. DOI 10.1016/j.envsoft.2018.02.002
4. Jarolínek, J., Pavlík, J., Kholova, J. and Ronanki, S. (2019) “Data Pre-processing for Agricultural Simulations“, AGRIS on-line Papers in Economics and Informatics, Vol. 11, No. 1, pp. 49-53. ISSN 1804-1930. DOI 10.7160/aol.2019.110105

5. Kim, K. S., Yoo, B. H., Shelia, V., Porter, C. H. and Hoogenboom, G. (2018) "START: A data preparation tool for crop simulation models using web-based soil databases", *Computers and Electronics in Agriculture*, vol. 154, pp. 256-264. ISSN 0168-1699. DOI 10.1016/j.compag.2018.08.023
6. Li, Q., Li, D. Big data GIS (2014). *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics and Information Science of Wuhan University*, vol. 39, iss. 6, pp. 641-644+666. DOI: 10.13203/j.whugis20140150
7. Pavlík, J., Masner, J., Jarolímek, J., Lukáš, M. (2019) "Data Processing for Yield Optimization", *Agrarian perspectives XXVIII. – Business Scale in Relation to Economics*, pp. 189-193.
8. Reinmuth, E. and Dabbert, S. (2017) "Toward more efficient model development for farming systems research - An integrative review", *Computers And Electronics In Agriculture*, Vol. 138, pp. 29-38. ISSN 0168-1699. DOI 10.1016/j.compag.2017.04.007
9. Skoogh, A., Michaloski, J., Bengtsson, N. Towards continuously updated simulation models: Combining automated raw data collection and automated data processing (2010). *Winter Simulation Conference*, pp. 1678-1689. DOI: 10.1109/WSC.2010.5678901
10. Szufel, P., Czupryna, M. and Kaminski, B. (2017) "Optimal execution of large scale simulations in the cloud. The case of route-To-pa sim online preference simulation", *Proceedings - Winter Simulation Conference*, pp. 3702-3703. DOI 10.1109/WSC.2016.7822408.
11. Zhao, G., Bryan, B. A., King, D., Luo, Z., Wang, E., Bende-Michl, U., Song, X. and Yu, Q. (2013) "Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining grid computing and parallel processing", *Environmental Modelling & Software*, Vol. 41, pp. 231-238. ISSN 1364-8152. DOI 10.1016/j.envsoft.2012.08.007