CAN ACCURACY ISSUES OF LOW-COST SENSOR MEASUREMENTS BE OVERCOME WITH DATA ASSIMILATION?

Monique P. G. de Oliveira[1*], Thais Q. Zorzeto-Cesar[1], Romis R. de F. Attux[2], Luiz H. A. Rodrigues[1]

[1*]Corresponding author. Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas/Campinas - SP, Brasil.
E-mail: moniquepgoliveira@gmail.com | ORCID ID: https://orcid.org/0000-0001-7167-4473

ABSTRACT

The use of mechanistic plant growth models relies on the availability of high-quality inputs to reduce uncertainty in estimates. Measurements of photosynthetically active radiation inside a protected environment are either more expensive to obtain or dependent on assumptions regarding external measurements. This study aimed to reduce the influence of uncertainty in the measurements of low-cost lux meters by using a data assimilation strategy. We first determined, by simulation, the impact of different sensors on the estimates. We then used the Ensemble Kalman Filter to assimilate artificial observations of tomato growth in the Reduced-State Tomgro model, in simulations for which the solar radiation inputs were obtained from a low-cost lux meter. We compared the assimilated estimates to the simulations that used solar radiation obtained with a scientific-grade quantum sensor. For periods of larger radiation intensity, in which the differences in measurements from both instruments are larger, assimilation of observations with low errors lead to estimates that are closer to the ones obtained by scientific grade sensors. These results suggest that low-cost sensors could be used to obtain inputs for growth models in protected environments, provided there are also imperfect observations of the state.

INTRODUCTION

One limitation often associated with the use of crop models is the availability of input data and, in particular, of input data with the required quality (Dias & Sentelhas, 2021; Ramirez-Villegas & Challinor, 2012). While most uncertainty quantification studies focus on climate models as sources of uncertainty in climate inputs, there is another factor of great relevance: the various sources of weather datasets (Chapagain et al., 2022). However, differently from field crops, in which the availability is connected to the presence of meteorological stations that should represent large areas, greenhouses and other protected environments may be more easily monitored, which allows a certain degree of control of environmental conditions. In the context of Internet of Things, monitoring temperature and relative humidity in protected environments is a procedure often mentioned in the literature (Tzounis et al., 2017). Nevertheless, solar radiation data for modeling is not as easily available as data related to these two other factors.

In greenhouse tomato modeling studies, for various reasons, the photosynthetically active radiation (PAR) that reaches the plants inside the greenhouse has often been estimated based on external global radiation (Berrueta et al., 2020; Righini et al., 2020). Using external data requires not only the measurement to be available but also an estimate of the transmissivity of the material, which, for plastic greenhouses, may change depending on condensation or dirt (Montero et al., 2019). Scientific grade alternatives to external radiation measurements, obtained inside the greenhouse, could be obtained by a net radiometer, a pyranometer, a quantum sensor, or a spectroradiometer (Both et al., 2015). Each instrument leads to a different measurement connected to solar radiation: net radiation, solar irradiance, photosynthetically active radiation, and flux in discrete wavebands, respectively.

As different models require different specifications of inputs, depending on how they were developed, conversions are often used. For example, for the conversion from global radiation to photosynthetically active radiation, approximations are used. (Berrueta et al., 2020) and (Righini et al., 2020) assumed PAR as 50% of the global radiation, while (Impron et al., 2007) assumed 43.4%. This ratio varies depending on cloud cover, atmospheric water and aerosols content, clearness and sky brightness, diffuse fraction, dew point temperature, and solar zenith angle (García-Rodríguez et al., 2021). Nonetheless, some conversions and the resulting approximations add to the uncertainty of the measurements. The use of low-cost sensors as substitutes to more accurate instruments, which has been proposed inter alia for control systems in greenhouses (Pisanu et al., 2020; Sumalan et al., 2020), can also be a source of uncertainty.

Assessments of uncertainty in weather forecasts impacting control systems' operational returns have already been performed (Kuijpers et al., 2022), and, while the impact of uncertainty in solar radiation datasets for current and future field crop estimates has been studied (Zhang et al., 2022), there has been no assessment regarding how inputs obtained by low-cost sensors in greenhouses could affect yield estimates. And since uncertainty in measurements of low-cost sensors could lead to larger uncertainty in model predictions, more accurate estimates would also be desirable. Uncertainty in model estimates may be reduced by the adoption of data assimilation. By including additional measurements of the state variable, i.e., additional information concerning the model, this technique allows for lower prediction errors as well as lower uncertainty in the outputs (Wallach et al., 2019).

This study aims to: i) compare the outcomes of simulations of growth and development of greenhouse tomatoes, with different sources for input radiation — quantum sensor and a low-cost lux meter; and ii) assess uncertainty in estimates from low-cost sensors by assimilating fruit images.

MATERIAL AND METHODS

Data

Plant growth data was obtained during one growth cycle of tomatoes, cultivar Seminis – DRC 564, in a protected environment, from March 16, 2021 to June 11, 2021. Approximately every two weeks, plants were subjected to destructive analyses to determine the dry biomass of plant organs — leaves, fruits, and stems — as well as plant leaf area. Dry biomass was determined by weighting the organs after drying for four days at 100 °C. Leaf area was determined by scanning the leaves along with a reference of known size and subsequently processing the digitized leaves to determine the area corresponding to leaves, measured in pixels, and the appropriate conversion to square meters.

Weather data corresponds to solar radiation and to air temperature recorded in three different tomato growth cycles, including the one previously mentioned, also in research greenhouses. The other cycles took place from July 12, 2019 to October 28, 2019 and from November 05, 2020 to February 12, 2021. Solar radiation was recorded as photosynthetically active solar radiation (PAR) by a quantum sensor Licor LI-190SA with a datalogger Licor LI-1400, and as luminosity by BH1750 sensors connected to Raspberry Pi model B computers. The first were recorded every fifteen minutes, and the latter, every five minutes. Concerning temperature, the data was obtained by SHT75 transducers installed in a hardware platform for wireless sensor networks (Radiuino BE900), as well as by DHT-22 sensors also connected to Raspberry Pi model B computers. The transducers were protected by porous capsules, which in turn were protected by polyvinyl chloride tubes coated with aluminum foil. The tubes included downstream fans. In both cases, data was recorded every five minutes. As the different sensors were not directly equivalent, unit conversion was required. Data from the lux meters was converted into photosynthetically active radiation [μmol s^{-1} m^{-2}] by multiplying measured value by 18 x 10^{-3} μmol s^{-1} m^{-2} lux^{-1} (Hall & Scurlock, 1993).

Data obtained by sensors connected to the Raspberry Pi computers will be considered, for the purpose of this study, as low-cost, and the other sensors will be defined as scientific grade. Two sensor nodes were placed in different positions in the greenhouse, and while data from both nodes will be presented to characterize the measurements, simulations will use only data from one of them.

Plant growth observations, along with input data from the scientific grade sensors, were used to obtain calibration for the model. Parameters were obtained by optimization.

Model and calibration

The Reduced State-Variable Tomgro (RT) model (Jones et al., 1999) was used to simulate growth and development of the tomatoes. Model parameters were calibrated by using an optimization algorithm with the destructive data described and using weather data from the scientific-grade sensors as inputs. Non-calibrated simulations used parameters from the original Gainesville calibration. Regardless of calibration, input data such as maximum leaf area or plant density referred to data from the evaluated cycle.

Observation data

Simulations using the calibrated model with the weather data from all four cycles were treated as the truth values of an artificial dataset. The simulations of fruit and mature fruit biomass were perturbed by Gaussian noise sampled from a distribution of zero mean and standard deviation corresponding to 10%, 30%, and 50% of the simulated truth. These were treated as observations of the truth. Twenty observation datasets were generated.

Data assimilation

The Ensemble Kalman Filter (Evensen, 1994, 2003) was used to assimilate the artificial observations, with the model estimates being obtained by the simulations run with inputs from the low-cost sensors. In this case, ensembles were generated by perturbation of weather inputs corresponding to 30% of the measured input. Weather inputs were provided by the low-cost sensors. The procedure was repeated 20 times, with each observation dataset, to avoid biasing the results due to sampling.

As the perturbation of inputs leads to variance in the difference equation, instead of the state itself, the uncertainty ascribed to observations should take this into account so that they will not be disproportionately larger. Therefore, while errors in measurements are supposed to correspond to fractions of 10%, 30%, and 50% of the simulated observations, the uncertainty associated with them will refer to the difference between the observation and the previous value multiplied by the respective fractions.

An additional random perturbation — N(1, 0.09) — was included in the observations, thus accounting for the variability of sampling the noise. The outcomes of the assimilation of different observations with different noise levels were compared to simulations using the calibrated model.

All data used for this study is available at Oliveira et al. (2021) and all code developed for model implementation and analyses is published in Oliveira (2023).

RESULTS AND DISCUSSION

Weather data

The curves in Figure 1 show the summaries of daily environmental data in the greenhouses. They correspond to the final value ascribed to the instruments, after processing, which were used in the simulations. In all three growth cycles, measurements of temperatures are reasonably close, except for maximum temperatures in Cycle 3. On the other hand, solar radiation integrals show visible differences across cycles. External solar radiation suggests there may have been additional noise in measured values, especially in Cycle 1, but possibly in Cycle 2, indicated by the different trends of the measurements. The differences may be ascribed to interference, such as the height in which the sensor was placed being lower than the plant maximum height, subjecting the sensors to shadows, or to the conversions of PAR and global radiation not being very precise and relying on more information than the approximations used allowed for.
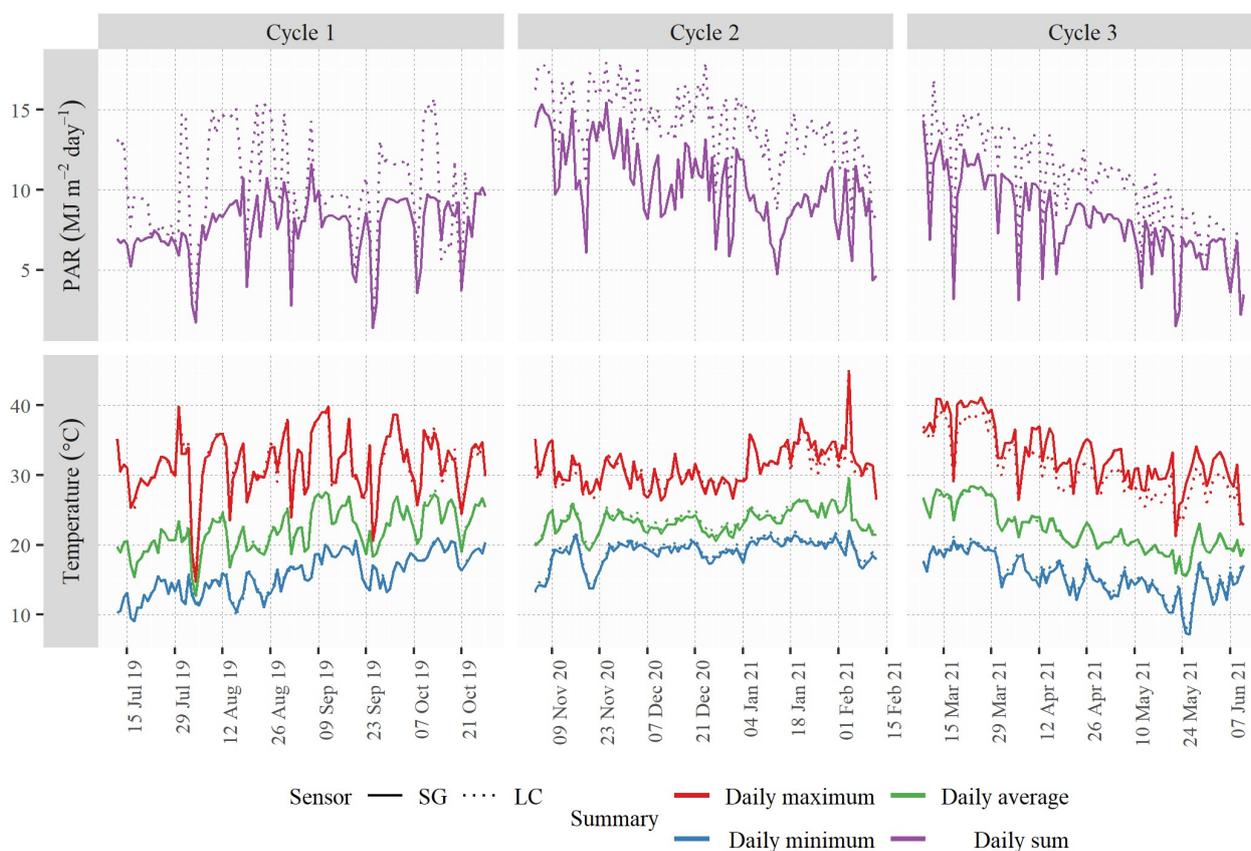

FIGURE 1. Summary of environmental data. Different line types refer to the different sensor types, scientific grade (SG) for quantum sensor and transducers and low-cost (LC) for the capacitive sensors connected to Raspberry.

Since quantum sensors for measuring PAR and lux meters are not equivalent, Figure 2 shows the relationship between the raw measurements from both. Both et al. (2015) reports nominal accuracy for quantum sensors of ± 10% and the technical note for the BH1750 lux meter, of 20%. While possible, the detection of the BH1750 was not adjusted, so the maximum value detected was of 65,535 lx, which corresponds to the horizontal lines at the maximum value of 1,000 on Figure 2.

The average ratio between the measurements of the quantum sensor and the lux meter were calculated as 0.8, 0.7, and 0.8 for cycles 1, 2, and 3, respectively, with standard deviations of 0.3, 0.3, and 0.4. These values were used as reference for determining the 30% of perturbation in the ensembles (item 4.4).

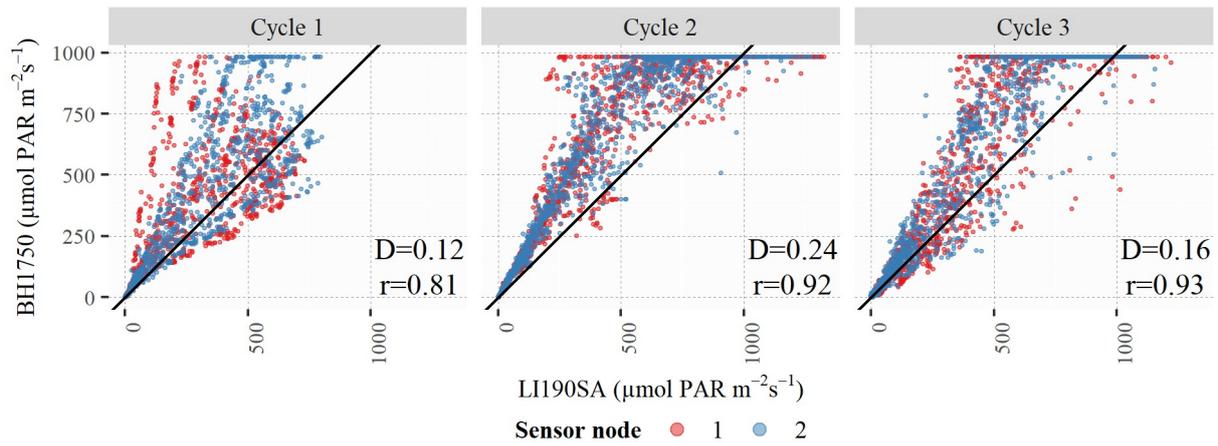Monique P. G. de Oliveira, Thais Q. Zorzeto-Cesar, Romis R. de F. Attux

FIGURE 2. Characterization of relationships between measurements of PAR converted from data obtained by BH1750 lux meters and measurements of PAR obtained by LI190SA quantum sensors. The metric r refers to the correlation coefficient and the metric D, to the two-sided Kolmogorov-Smirnov test.

Simulations

Overall, the different set of sensors led to differences only in photosynthesis and biomass variables, while leaf area and the number of nodes outcomes were the same regardless of the set of sensors used (Figure 3). The differences can be ascribed to the solar radiation inputs, since the measurements are generally similar between the temperature sensors, as noted in Figure 1, as well as leaf area and number of nodes relying exclusively on mean hourly temperature. However, even when there are differences in simulations, they are not high, except for Experiment 2, in which the differences in measurements were also higher (Figure 1). In this case, photosynthesis is more gravely affected, even if the leaf area is not, which also points to the impact of the radiation inputs. Although the ratios between sensor measurements were in general similar across cycles, the higher radiation magnitude in Cycle 2 may have led to the higher differences.

These impacts could have been captured by a sensitivity analysis, in which several thousand simulations would be run with the inputs being slightly modified, leading to differences in the outcomes that would be higher, the higher the importance of that input. Nevertheless, these analyses have two relevant requirements: calibration that would ensure the parameters are suitable for those multiple conditions, and plausible weather time-series, which would not include impossible combinations of the meteorological factors. There are few studies that meet these requirements in tomato modeling. Cooman & Schrevens (2007) assessed the sensitivity to the inputs of the second version of the Tomgro model. For this version of the model, while dry fruit biomass was very sensitive to solar radiation, temperature was the most important factor for total biomass.
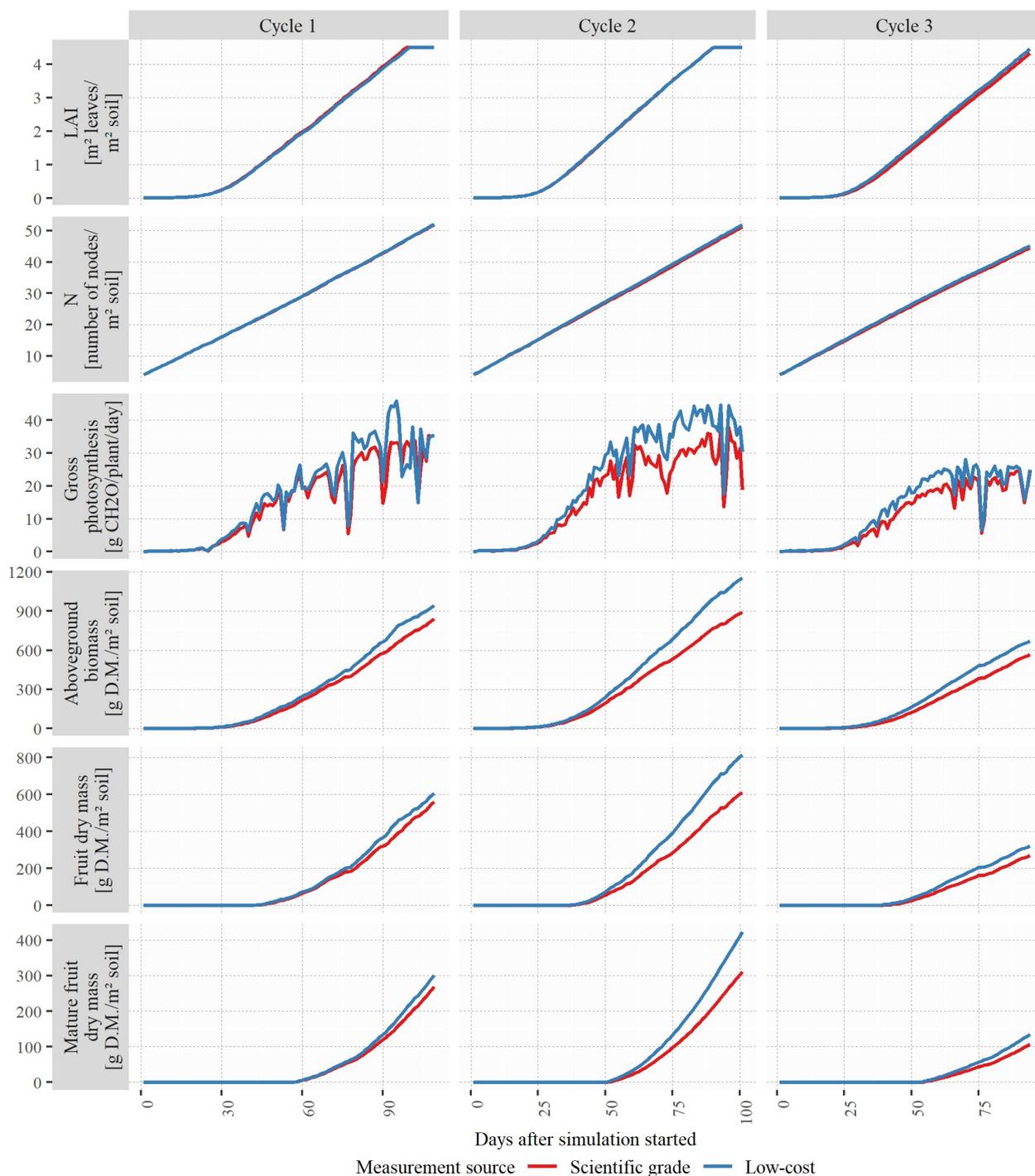
FIGURE 3. Truth curves, i. e., simulations obtained with the scientific grade sensors, compared to the simulations that used the low-cost sensors as sources of input data.

Assimilation of observations

To fulfill the premise of the study, assimilation should lead to estimates that are close to the ones obtained by the simulations with inputs from scientific grade sensors. However, the results in a data assimilation study depend on the quality of observations. In Cycles 1 and 3, in which the model estimates obtained using low-cost sensors only slightly overestimated fruit and mature fruit weights, assimilation of fruit observations with low errors led to estimates that while close to the simulated truth, do not differ as much from the original estimates. In Cycle 2, in which the difference was higher, the effect of low-error observations is more pronounced and up to the added noise of 30% of the observation value, estimates are close to the ones obtained by the scientific grade sensors. In all cases, when observations had errors that were higher than 50%, the benefits disappeared.
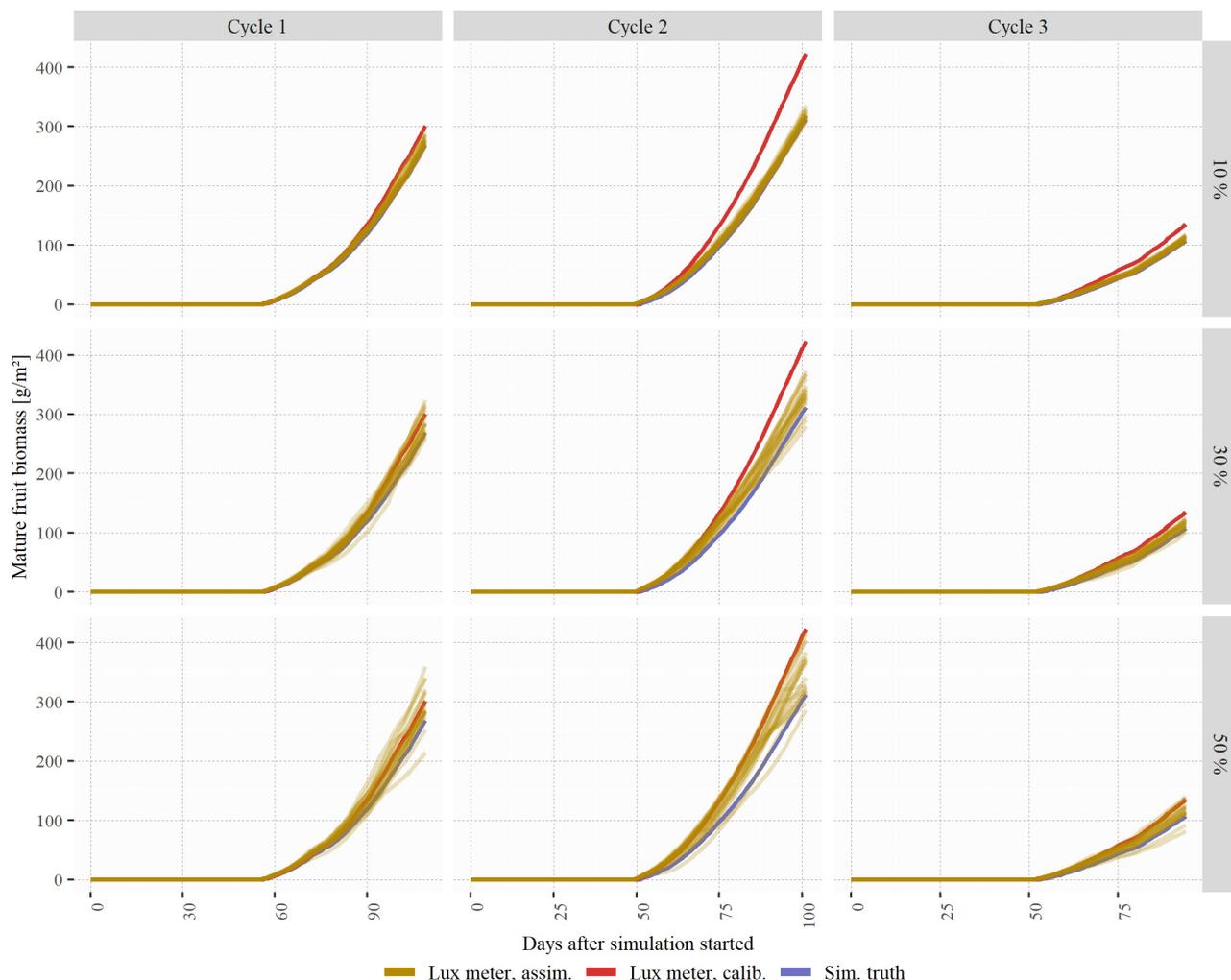
FIGURE 4. Mature fruit biomass estimates given different input sources: lux meters and assimilation, lux meters without assimilation, and scientific grade sensors as simulated truth.

Another premise of this discussion is the existence of observations that would satisfy these low errors requirements. Observations of fruits could be obtained from pictures; for instance, Fonteijn et al. (2021) obtained reasonable correlations between fruit weight and the radius measured in pixels, but with increased variance for higher fruit weights. If the premise is satisfied, the observed improvement in simulation outcomes suggests that with low-cost equipment such as cameras and lux meters it is possible to obtain a simulation that relies on solar radiation inputs in a greenhouse without depending on characterization of the cover material.

There are different approaches that would not rely on assimilation of observations. If the goal is to improve environmental observations themselves, instead of only obtaining better yield estimates, it could be possible to assimilate data from the environment into a process-based greenhouse model. van Mourik et al. (2019) used this approach, but the filters did not lead to improvement in monitoring. Furthermore, recently published greenhouse climate models were reviewed by Katzin et al. (2022), and the authors noted there is still a lot of progress to be made on them, including regarding the assessment of their performances. Nevertheless, this is a case in which, differently from the assimilation of environmental measurements, assimilation of fruit observations would directly impact uncertainties in estimates, since they include external information regarding the state of the desired variable.

As the magnitudes increase through the cycle, so do the values of the uncertainty hyperparameters used in the Ensemble Kalman Filter (Figure 5). When the observation covariance increases, the gain becomes lower, leading estimates to rely less on observations. The difference in the state's covariance also points to how much the estimates' uncertainties obtained by the perturbation in the inputs, ascribed to the expected uncertainty caused by accuracy of the observations, are reduced after the assimilation of observations.
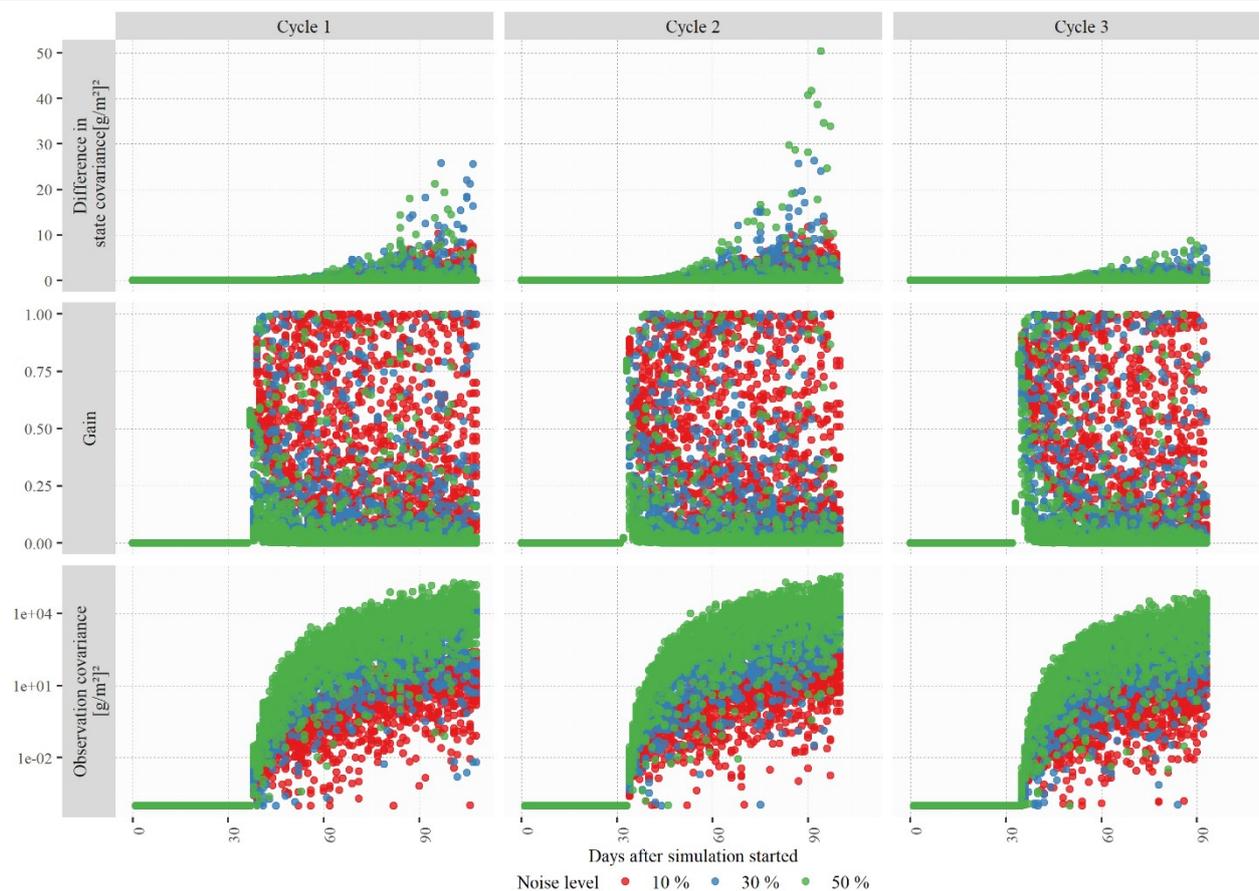
FIGURE 5. Evolution of filter hyperparameters on the assimilation of fruit biomass with different noise levels for the observations using the low-cost sensor as source of solar radiation input, as well as in the difference in the state covariance caused by the assimilation.

CONCLUSIONS

Sensors of different precision and accuracy levels as sources of solar radiation inputs in a tomato growth model led to different estimates of photosynthesis, which impacted estimates of biomass and yield. These differences, particularly regarding yield estimates, were mitigated by assimilation of observations of fruits. Our simulated data led us to observe that, often, noise levels of 10 and 30% in observations yielded outcomes similar to the ones obtained by scientific-grade sensors. This assimilation also reduces uncertainties in model estimates when low-cost sensors are used to provide weather inputs.

ACKNOWLEDGEMENTS

REFERENCES

Berrueta C, Heuvelink E, Giménez G, Dogliotti S (2020) Estimation of tomato yield gaps for greenhouse in Uruguay. Scientia Horticulturae 265:109250. https://doi.org/10.1016/j.scienta.2020.109250

Both AJ, Benjamin L, Franklin J, Holroyd G, Incoll LD, Lefsrud MG, Pitkin G (2015) Guidelines for measuring and reporting environmental parameters for experiments in greenhouses. Plant Methods 11(1):43. https://doi.org/10.1186/s13007-015-0083-5

Chapagain R, Remenyi TA, Harris RMB, Mohammed CL, Huth N, Wallach D, Rezaei EE, Ojeda JJ (2022) Decomposing crop model uncertainty: A systematic review. Field Crops Research 279:108448. https://doi.org/10.1016/j.fcr.2022.108448

Cooman A, Schrevens E (2007) Sensitivity of the Tomgro Model to Solar Radiation Intensity, Air Temperature and Carbon Dioxide Concentration. Biosystems Engineering 96(2):249–255. https://doi.org/10.1016/J.BIOSYSTEMSENG.2006.10.011

Dias HB, Sentelhas PC (2021) Assessing the performance of two gridded weather data for sugarcane crop simulations with a process-based model in Center-South Brazil. International Journal of Biometeorology 65(11):1881–1893. https://doi.org/10.1007/s00484-021-02145-6

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research 99(C5):10143–10162. https://doi.org/10.1029/94JC00572

Evensen G (2003) The Ensemble Kalman Filter: Theoretical formulation and practical implementation. Ocean Dynamics 53(4):343–367. https://doi.org/10.1007/s10236-003-0036-9

Fonteijn H, Afonso M, Lensink D, Mooij M, Faber N, Vroegop A, Polder G, Wehrens R (2021) Automatic phenotyping of tomatoes in production greenhouses using robotics and computer vision: From theory to practice. Agronomy 11(8): https://doi.org/10.3390/agronomy11081599

García-Rodríguez A, Granados-López D, García-Rodríguez S, Díez-Mediavilla M, Alonso-Tristán C (2021) Modelling Photosynthetic Active Radiation (PAR) through meteorological indices under all sky conditions. Agricultural and Forest Meteorology 310:108627. https://doi.org/10.1016/j.agrformet.2021.108627

Hall OD, Scurlock JMO (1993) Biomass production and data. In: Hall DO, Scurlock JMO, Bolhàr-Nordenkampf HR, Leegood RC, Long SP. Photosynthesis and Production in a Changing Environment: afield and laboratory manual. Dordrecht, Springer Netherlands. https://doi.org/10.1007/978-94-011-1566-7

Impron I, Hemming S, Bot GPA (2007) Simple greenhouse climate model as a design tool for greenhouses in tropical lowland. Biosystems Engineering 98(1):79–89. https://doi.org/10.1016/J.BIOSYSTEMSENG.2007.03.028

Jones JW, Kenig A, Vallejos CE (1999) Reduced state-variable tomato growth model. Transactions of the ASAE 42(1):255–265. https://doi.org/10.13031/2013.13203

Katzin D, van Henten EJ, van Mourik S (2022) Process-based greenhouse climate models: Genealogy, current status, and future directions. Agricultural Systems 198:103388. https://doi.org/10.1016/j.agsy.2022.103388

Kuijpers WJP, Antunes DJ, van Mourik S, van Henten EJ, van de Molengraft MJG (2022) Weather forecast error modelling and performance analysis of automatic greenhouse climate control. Biosystems Engineering 214:207–229. https://doi.org/10.1016/j.biosystemseng.2021.12.014

Montero JI, Zhang Y, Yang Q, Ke X (2019) Advances in greenhouse design. In: Marcelis LFM, Heuvelink E. Achieving sustainable greenhouse cultivation. Cambridge, Burleigh Dodds Science Publishing. https://doi.org/10.19103/AS.2019.0052.02

Oliveira MPG de, Amaro RP, Pescarini HB, Rodrigues LHA (2021) Tomato growth in production-like setting Repositório de Dados de Pesquisa da Unicamp, V5 https://doi.org/10.25824/redu/EP4NGO

Oliveira M (2023) Leveraging high frequency data for improving crop growth estimates (0.1.1) https://doi.org/10.5281/zenodo.7632419

Pisanu T, Garau S, Ortu P, Schirru L, Macciò C (2020) Prototype of a Low-Cost electronic platform for real time greenhouse environment monitoring: an agriculture 4.0 perspective. Electronics 9(5):726. https://doi.org/10.3390/electronics9050726

Ramirez-Villegas J, Challinor A (2012) Assessing relevant climate data for agricultural applications. Agricultural and Forest Meteorology 161:26–45. https://doi.org/10.1016/j.agrformet.2012.03.015

Righini I, Vanthoor B, Verheul MJ, Naseer M, Maessen H, Persson T, Stanghellini C (2020) A greenhouse climate-yield model focussing on additional light, heat harvesting and its validation. Biosystems Engineering 194:1–15. https://doi.org/10.1016/j.biosystemseng.2020.03.009

Sumalan RL, Stroia N, Moga D, Muresan V, Lodin A, Vintila T, Popescu CA (2020) A cost-effective embedded platform for Greenhouse environment control and remote monitoring. Agronomy 10(7):936. https://doi.org/10.3390/agronomy10070936

Tzounis A, Katsoulas N, Bartzanas T, Kittas C (2017) Internet of Things in agriculture, recent advances and future challenges. Biosystems Engineering 164:31–48. https://doi.org/10.1016/j.biosystemseng.2017.09.007

van Mourik S, van Beveren PJM, López-Cruz IL, van Henten EJ (2019) Improving climate monitoring in greenhouse cultivation via model based filtering. Biosystems Engineering 181:40–51. https://doi.org/10.1016/J.BIOSYSTEMSENG.2019.03.001

Wallach D, Makowski D, Jones JW, Brun F (2019) Data Assimilation for Dynamic Models. In: Wallach D, Makowski D, Jones JW, Brun F. Working with dynamic crop models. Amsterdam, Elsevier. https://doi.org/10.1016/B978-0-12-811756-9.00014-9

Zhang Z, Li T, Guo E, Zhao C, Zhao J, Liu Z, Sun S, Zhang F, Guo S, Nie J, Yang X (2022) 20% of uncertainty in yield estimates could be caused by the radiation source. Science of The Total Environment 838:156015. https://doi.org/10.1016/j.scitotenv.2022.156015