Bayesian Uncertainty Quantification in Greenhouse Modeling

Juan Daniel Molina Muñoz & J. Andrés Christen

Centro de Investigación en Matemáticas, CIMAT

Guanajuato, Gto, México

emails: juan.molina@cimat.mx / jac@cimat.mx

CIMAT

Introduction

A greenhouse is an enclosed space in which plants are grown, where it is possible to control the climate conditions. It is wanted that these conditions are suitable for the development of plants.

Some advantages of growing in greenhouses: Control of environmental conditions inside the greenhouse to reduce risks over the crop; Efficient use of land and water; And the crop is protected from external influences (wind, rain, low temperatures, pests).

It is possible to characterize the operation of a commercial greenhouse from 3 base models: The Greenhouse Climate Model; The Photosynthesis Model; and The Growth Model: The Greenhouse Climate Model is an ODE system that describes the dynamics in time of the main thermodynamic variables of the greenhouse; The Photosynthesis Model describes how the assimilates (plants food) are produced from the photosynthesis process; And The Growth Model describes the growth dynamics of the crop of interest each day.

Methodology

The inference procedure was based on the databases presented in Hemming et al. (2019): Data from August to December 2018 from 6 independent greenhouses with identical technology located in The Netherlands.

The inference procedure was divided into two stages: Climate Model and Growth Model. In each one: An error model was proposed; Parameter calibration was performed from a database; And the error model was validated from another database.

Climate Model Inference

An error model is proposed with the following structure:

$$T_2(t) = T_{2M}(t) + \varepsilon_T,$$

$$V_1(t) = V_{1M}(t) + \varepsilon_V,$$

$$C_1(t) = C_{1M}(t) + \varepsilon_C,$$

where T_2 (the greenhouse air temperature), V_1 (the vapor pressure of the greenhouse air) and C_1 (the CO_2 concentration in the greenhouse air) are the true values and T_{2M}, V_{1M} and C_{1M} are the calculated for the model. Also, $\varepsilon_T \sim \mathcal{N}(0\,,\,\sigma^2_{T_2})$, $\varepsilon_V \sim \mathcal{N}(0\,,\,\sigma^2_{V_1})$ and $\varepsilon_C \sim \mathcal{N}(0\,,\,\sigma^2_{C_1})$. So, $T_2(t) \sim \mathcal{N}(T_{2M}(t)\,,\,\sigma^2_{T_2})$, $V_1(t) \sim \mathcal{N}(V_{1M}(t)\,,\,\sigma^2_{V_1})$ and $C_1(t) \sim \mathcal{N}(C_{1M}(t)\,,\,\sigma^2_{C_1})$.

Considering observations in the times $\{t_1, \cdots, t_k\}$ the likelihood function is

$$f(T_2, V_1, C_1 \,|\, \theta) = \prod_{i=1}^{k} f(\,T_2(t_i)\,)\, f(\,V_1(t_i)\,)\, f(\,C_1(t_i)\,),$$

where $\theta = (\,\alpha_1,\, \phi_2,\, \psi_2,\, \sigma^2_{T_2},\, \sigma^2_{V_1},\, \sigma^2_{C_1}\,)$.

For α_1, ϕ_2 and ψ_2 we assigned gamma distributions as priors, and for $\sigma^2_{T_2}$, $\sigma^2_{V_1}$ and $\sigma^2_{C_1}$ inverse-gamma distributions. Assuming independence between the parameters

$$\pi(\theta) = \pi(\alpha_1)\,\pi(\phi_2)\,\pi(\psi_2)\,\pi(\sigma^2_{T_2})\,\pi(\sigma^2_{V_1})\,\pi(\sigma^2_{C_1}).$$

To calculate the posterior distribution, we considered

$$\pi(\theta \,|\, T_2, V_1, C_1) \propto f(T_2, V_1, C_1 \,|\, \theta)\,\pi(\,\theta\,).$$

Growth Model Inference

We assumed for n_k the average number of fruits harvested on day k in one square meter of the greenhouse:

$$n_k \sim \mathrm{Gamma}(\,\alpha_k, \beta_k\,), \quad \text{with} \quad \mathbb{E}(n_k) = \frac{\alpha_k}{\beta_k} = m_k,$$

where m_k is the result of the model. If the signal-to-noise ratio is set to 4, $\alpha_k = 16$ and $\beta_k = \frac{16}{m_k+\delta}$.

Furthermore, we assumed for h_k the average of the total weight of the fruits harvested that day:

$$h_k \,|\, n_k \sim \mathcal{N}(\,380 n_k\,,\, n_k\,\sigma^2_F\,).$$

Given observations $D = [D_1, \cdots, D_k] = [(n_1, h_1), \cdots, (n_k, h_k)]$, and assuming independence of production between days

$$f(D \,|\, \theta) = \prod_{i=1}^{k} \frac{\left(\frac{16}{m_i+\delta}\right)^{16} n_i^{15}}{15!\sqrt{2\pi n_i}\,\sigma_F} \exp\left[-\frac{16 n_i}{m_i+\delta} - \frac{1}{2}\frac{(h_i - 380 n_i)^2}{n_i \sigma^2_F}\right]$$

where $\theta = (\,\nu,\, a,\, b,\, \sigma^2_F\,)$.

For ν, a and b we assigned gamma distributions as priors, and for σ^2_F inverse-gamma distribution. Assuming independence between the parameters

$$\pi(\theta) = \pi(\nu)\,\pi(a)\,\pi(b)\,\pi(\sigma^2_F).$$

To calculate the posterior distribution, we considered

$$\pi(\theta \,|\, D) \propto f(D \,|\, \theta)\,\pi(\,\theta\,).$$

Results

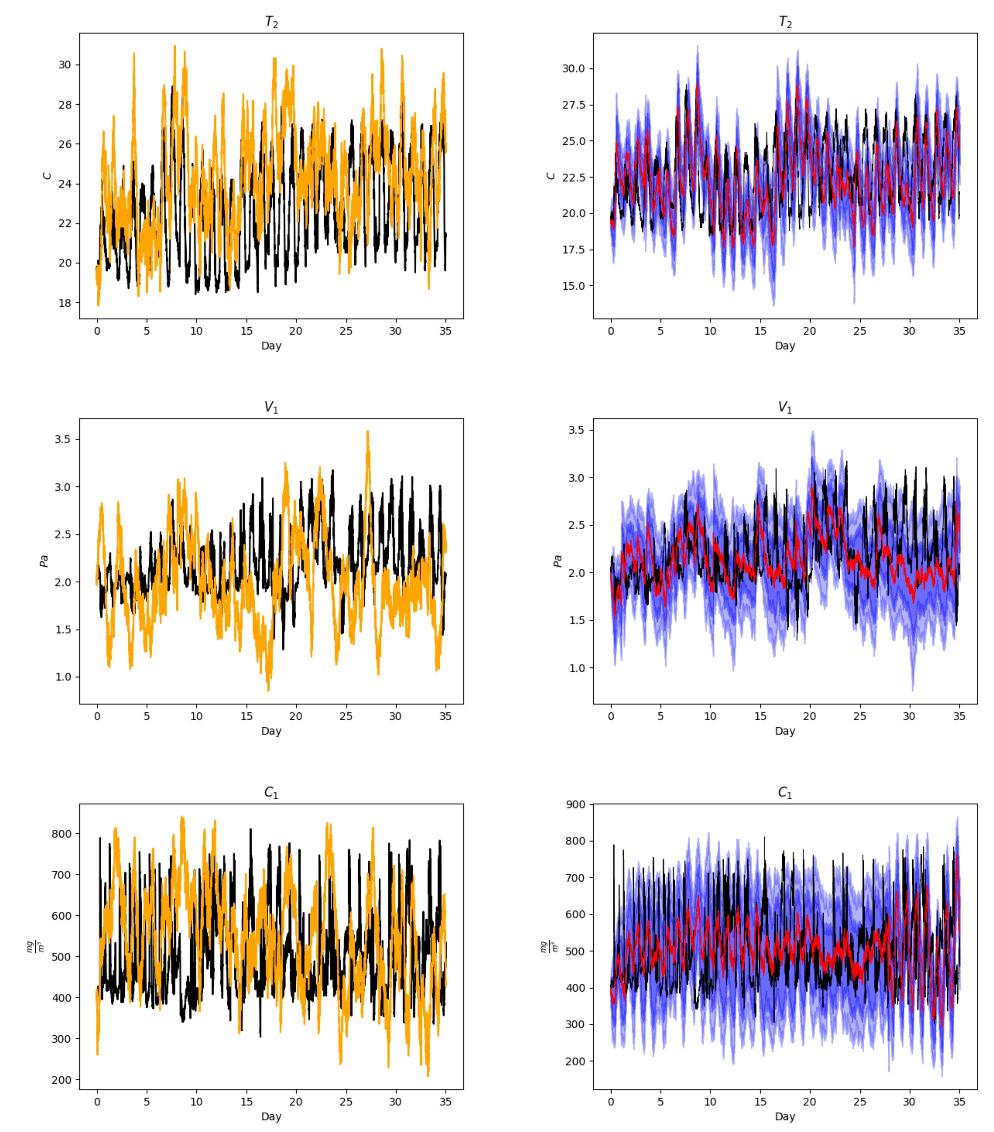

Figure 1: Validations results for the Climate Model. On the first column, in orange an example solution calculated using the posterior median and in black the true data. On the second column, the uncertainty of the solution by simulating from the posterior predictive distribution, blue shadows are the 10% - 90% and 25% - 75% quantiles and the red line is the median.

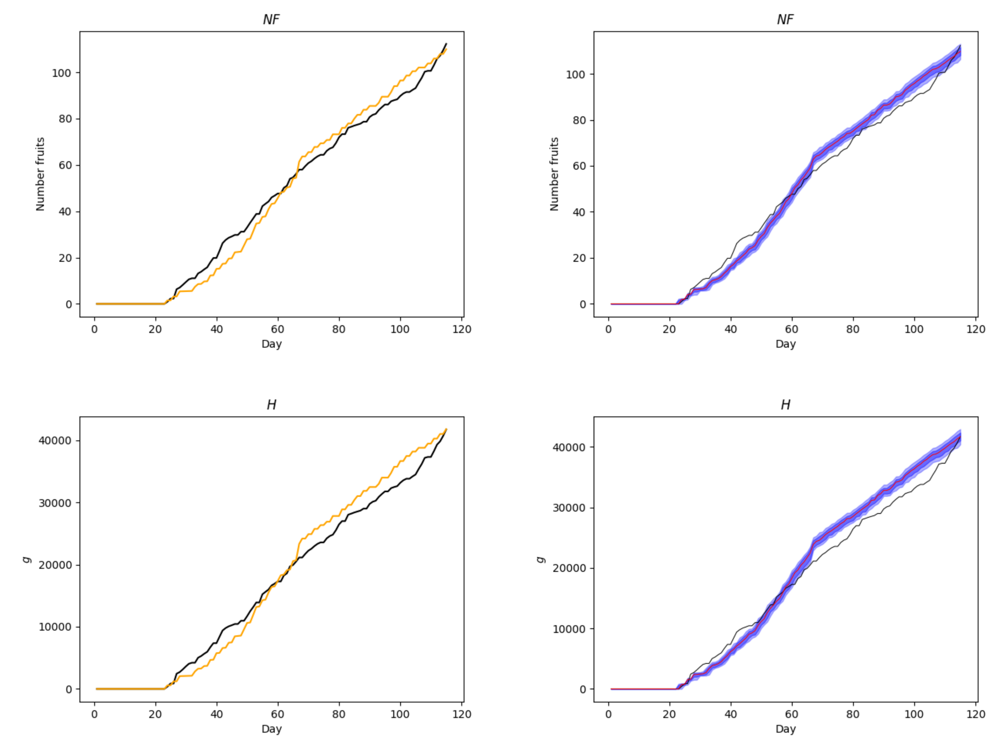

Figure 2: Validations results for the Growth Model. On the first column, in orange an example solution calculated using the MAP and in black the true data. On the second column, the uncertainty of the solution by simulating from the posterior predictive distribution, blue shadows are the 10% - 90% and 25% - 75% quantiles and the red line is the median. N_F is the accumulated number of fruits harvested and H is the accumulated weight of fruits harvested.

Discussion

By means of the proposed methodology, simulations of the production variables close to reality are obtained.

The proposed methodology has the advantage (compared to those that currently exist in the literature) that it allows quantifying the uncertainty of the climate and production variables, and of parameters of interest.

References

[1] Hemming, S., de Zwart, F., Elings, A., Righini, I. & Petropoulou, A. (2019). Remote control of greenhouse vegetable production with artificial intelligencegreenhouse climate, irrigation, and crop production. *Sensors*, 19(8),1807.