Pitch to Master Thesis

GreenSloth: A Curated Web Resource for Validating and Comparing Peer-Reviewed Computational Models of Photosynthesis

Elouën Corvest

Date: February 2, 2026

Food, clothes, medicine and many more products of human labour have at least their start using plants and their fantastic way of primary metabolism, photosynthesis. While the chemical reaction can be easily written down (Eq. 1), the biological process is quite complex. It includes several proteins and complexes that come together to utilise energy absorbed by light, powering an electron transfer chain that enables the production of a few ATP molecules while transforming water into oxygen. The ATP and several other products produced by photosynthesis are used to convert breathed-in CO_2 into sugars by the Calvin-Benson-Bassham cycle (Stirbet et al., 2020).

$$6\,CO_2 + 12\,H_2O + \text{light energy} \longrightarrow C_6H_{12}O_6 + 6\,O_2 + 6\,H_2O \tag{1}$$
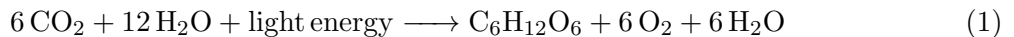
Enhancing the efficiency of photosynthesis has long been a key way to increase crop yield (Murchie et al., 2009). However, the standard methods that introduced the Green Revolution have reached a plateau (Long et al., 2006), which is why researching photosynthesis in even greater detail has gained increased importance. Of course, the basics of photosynthesis have been known since the 18th century (Ingenhousz, 1779), but the pursuit of knowledge is still not complete. While experimental methods continue to be used to gain a deeper understanding of photosynthesis, the significant increase in computational prowess over the last few decades has also made it possible to model this process.

Using experimental data, it is possible to gather information on enzymatic, chemical, topographic, and other types of properties. With this information, it is possible to identify mathematical connections in biological networks, which in turn give rise to calculations. As Biology is living chemistry, as chemistry is molecular physics, and as physics is interpretive mathematics, it should, in theory, be possible to translate every biological network into pure mathematics. This is no different with photosynthesis, and has been proven to be viable in many cases.

As photosynthesis is such a complicated process, modelling it in its entirety is no easy task. There have been attempts (Zhu et al., 2013), but none have been successful enough to be chosen as the definitive model of photosynthesis. The struggle not only lies in insufficient computational power, but also in still-unexplained areas of photosynthesis. Therefore, the modelling community has introduced simplifications in the process to bridge the knowledge gaps, hone in on the model's task to a specific part of photosynthesis, or do both. This has allowed the emergence of many models of photosynthesis, not only in terms of what they depict, but also in how they are constructed.

As models vary on so many aspects, it can be daunting to pursue using one. Especially, as most model builders are experts in that exact field and not in experimental studies, many models suffer from being built in a way that is not user-friendly for people who may not even know how to code. Contrarily, the goal of most models is not to be used by other modelling experts, but by experimentalists to see where they should focus their work next. Therefore, it is in the model's best interest to be as user-friendly as possible. However, it is not only the model builders at fault, but also the lack of guidelines for publishing models.

Publishing experimental work has semi-strict guidelines regarding reproducibility. All publications must include a "Methods" section that explains in detail how each experiment and statistical conclusion was made. Often, they include the names of used chemicals, step-by-step guides, names of companies where resources have been allocated, and much more. These restric-

tions have been implemented to ensure that science is as factual as possible. Furthermore, it is customary for a publication to undergo a reviewing process, which involves consulting other experts in the field to solicit their opinions on the publication. Sometimes, the author may need to validate their conclusion further or reconsider their work from a different perspective. This process, in theory, should enhance the quality of scientific work. These tried and tested methods, although not perfect, have been a staple of the scientific world, which is why they have been incorporated into scientific work in modelling. However, one should not overlook the flaw behind it.

As modelling is theoretical work, which should only be validated with experimental work, the "Methods" section is usually quite empty. It is customary, however, to include an explanation of how the model was built, thereby detailing the mathematical formulations behind it. However, as most models are pretty complex, writing numerous mathematical equations may introduce errors in formulation, just as it often happens with grammatical mistakes in written text. The reviewing process frequently overlooks these errors, as they are not mathematical errors, which would require the reviewers to compare the publication with the code of the model itself, a task they often do not undertake. This problem, combined with the fact that it is not compulsory to publish the code with the publication, makes working with the model quite difficult. These problems persist in many modelling publications, making it nearly impossible for someone inexperienced to delve into the world of modelling.

This issue is widely acknowledged, and numerous attempts have been made to refute it. From new ways to share a model (Hucka et al., 2003), to an easier way to find models (Malik-Sheriff et al., 2019), and more. However, most have been only partially successful and are missing out on many aspects. This is where the idea of GreenSloth came into play.

GreenSloth will be an open-source database that will focus on popular kinetic models of photosynthesis. It will support a list of curated models, all written in Python and utilising the mxlpy package (Van Aalst et al., 2025) for consistency. The models will provide a summary in text form, as well as an overview of their mathematical background. Furthermore, GreenSloth will alleviate the hassle of finding a model that fits the needs of the searchee by employing a catalogue of tags that link models together. These tags range from simple components of photosynthesis to the type of data that can be extracted from the model. Finally, a direct comparison between models will be made available, which not only includes the numerical data but also newly developed tests that will validate the model in many different ways. To bring all these features together and make them as accessible as possible, GreenSloth will be made available through a website that features a lovely and user-friendly design, further facilitating the user's work. GreenSloth is a solution that needs to exist, and a first iteration can already be found online at https://greensloth.rwth-aachen.de/.

References

Stirbet, A., Lazár, D., Guo, Y., & Govindjee, G. (2020). Photosynthesis: Basics, history and modelling. *Annals of Botany*, *126*(4), 511–537. https://doi.org/10.1093/aob/mcz171

Murchie, E. H., Pinto, M., & Horton, P. (2009). Agriculture and the new challenges for photosynthesis research. *New Phytologist*, *181*(3), 532–552. https://doi.org/10.1111/j.1469-8137.2008.02705.x

Long, S. P., Zhu, X.-G., Naidu, S. L., & Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant, Cell & Environment*, *29*(3), 315–330. https://doi.org/10.1111/j.1365-3040.2005.01493.x

Ingenhousz, J. (1779). *Experiments upon vegetables, discovering their great power of purifying the common air in the sun-shine, and of injuring it in the shade and at night: To which is joined, a new method of examining the accurate degree of salubrity of the atmosphere* [Google-Books-ID: JUMZAAAAYAAJ]. P. Elmsly, H. Payne.

Zhu, X.-G., Wang, Y., Ort, D. R., & Long, S. P. (2013). *e*-photosynthesis: A comprehensive dynamic mechanistic model of c3 photosynthesis: From light capture to sucrose synthesis [Publisher: Wiley]. *Plant, Cell & Environment*, *36*(9), 1711–1727. https://doi.org/10.1111/pce.12025

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., . . . Wang, J. (2003). The systems biology markup language (SBML): A medium forrepresentation and exchange of biochemical network models [Publisher: Oxford University Press (OUP)]. *Bioinformatics*, *19*(4), 524–531. https://doi.org/10.1093/bioinformatics/btg015

Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., Vu, M. T., Men, J., Maire, M., Kananathan, S., Fairbanks, E. L., Meyer, J. P., Arankalle, C., Varusai, T. M., Knight-Schrijver, V., Li, L., Dueñas-Roca, C., Dass, G., Keating, S. M., . . . Hermjakob, H. (2019). BioModels—15 years of sharing computational models in life science [Publisher: Oxford University Press (OUP)]. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz1055

Van Aalst, M., Nies, T., Pfennig, T., & Matuszyńska, A. (2025). MxlPy - python package for mechanistic learning in life science. https://doi.org/10.1101/2025.05.06.652335