Pherosensor-toolbox: a Python package for Biology-Informed Data Assimilation

Thibault Malou, Simon Labarthe

The Journal of Open Source Software

Pherosensor-toolbox: a Python package for Biology-Informed Data Assimilation

Thibault Malou[1] and **Simon Labarthe**[1,2,3¶]

1 Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France **2** University of Bordeaux, INRAE, BIOGECO, 33610, Cestas, France **3** Inria, University of Bordeaux, INRAE, Talence, France ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Insect pests are a major threat for agricultural systems (Oerke, 2006), leading to an intensive use of pesticide for crop protection with non-sustainable drawbacks on environment, biodiversity and human health. Most of insects produce pheromones for conspecific communication, making pheromone sensors a good tool for early specific detection of pests, in order to reduce pesticide use in a precision agriculture context (Gebbers & Adamchuk, 2010).

`Pheromone-toolbox` is a Python package containing numerical tools for pheromone sensor data assimilation to infer the position of emitting pest insects. It contains specific tools to model pheromone propagation and solve the corresponding inverse problem to determinate emitters position taking into account the environmental context (wind, landscape, vegetation…). A specific focus is put on the integration of biological knowledge of pest behavior during inference.

Statement of need

Unlike other generic tools for data assimilation, such as DAPPER (Raanes et al., 2024), OpenDA (Ridler et al., 2014) or (Nerger et al., 2005), `Pherosensor-toolbox` is a context-specific application-oriented package specifically designed to solve the inverse problem of inferring the source term (i.e. pheromone emitters position and emission rates) within a Chemical-Transport Model (CTM) modelling pheromone propagation in an agricultural landscape. Additional feature is the possibility to inform the data assimilation with insect behavior, such as the population dynamics partial differential equations (PDE), to get Biology-Informed Data-Assimilation (BI-DA). BI-DA aims to counter-balance data scarcity with prior biological knowledge. An other additional feature to come is optimal sensor placement tools to find the most informative sensor placement for assimilating the sensors data and inferring the source term of the CTM.

Outlook

Direct CTM problem

`Pherosensor-toolbox` first contains numerical tools to solve a 2D CTM, i.e. the equation defined on a landscape Ω and a time span $(0,T)$ as

$$\frac{\partial c}{\partial t} - \nabla \cdot (\mathbf{K}\nabla c) + \nabla \cdot (\vec{u}c) + \tau_{loss}c = s \quad \forall (x,y) \in \Omega, \forall t \in (0;T) \tag{1}$$

where $c(t,x,y)$ is the local pheromone concentration, $\mathbf{K}$ is a diffusion coefficient, $\vec{u}$ is a wind field, τ_{loss} represents vertical loss of pheromone (including vertical transport and vegetation-specific deposition), and s is the quantity of pheromone emitted. Note that $\mathbf{K}$, $\vec{u}$ and τ_{loss} are known parameters, whereas $s(t,x,y)$ is the source term to estimate. The latter is related to pest density $p(x,y)$ by the relation $s = q(t)p(x,y)$ where q is a time pheromone emission per insect.

Pherosensor-toolbox includes a finite volume solver defined on a cartesian scatter grid with implicite and semi-implicite time-schemes.

BI-DA to solve the inverse problem

We define BI-DA with the following optimization problem: find the optimal quantity of pheromone emitted in time and space $s_a(t,x,y)$ such that \$

$$s_a(x,y,t) = \underset{s(x,y,t)}{\operatorname{argmin}}\ j(s) \text{ with } j(s) = j_{obs}(s) + j_{reg}(s) \tag{2}$$

\$ where j_{obs} is the observation loss and j_{reg} is a regularization term. Namely

$$j_{obs}(s) = \|m\left(c(s)\right) - m^{obs}\|^2_{\mathbf{R}^{-1}}$$

where $c(s)$ is the concentration map obtained by solving the CTM Equation 1 with second member s, m^{obs} are noisy observations with covariance $\mathbf{R}$, and $c \mapsto m$ is an observation operator.

In the BI-DA framework, the term j_{reg} involves biological priors including LASSO (pest sparsity in time and space), group-LASSO (pest sparsity in space), Tikhonov (pest favorite habitat), log-barrier (inappropriate habitat) or pest population dynamics. For population dynamics, $j_{reg}(s) = \|\mathcal{M}(s) - \gamma\|_2^2$, i.e. the regularization aims at minimizing the residual of a PDE or ODE model defined with the differential operator $\mathcal{M}$ and a background value γ.

Pherosensor-toolbox provides gradient-based (gradient descent or proximal gradient) variational optimization methods to solve Equation 2, where the gradient $\nabla_s j_{obs}(s)$ is obtained by solving the adjoint model of the CTM. It also provides tools to implement the population dynamics PDE or ODE-based regularization.

Postprocessing

Pherosensor-toolbox comes with several plotting functions to display differences and benchmarks between a ground truth and the estimate s_a including spatial maps or pest presence maps defined with level sets.

Related works

Pherosensor-toolbox has been used in a publication introducing the BI-DA framework and assessing the impact of incorporating prior biological knowledge on the estimation accuracy (Malou et al., 2024). This publication also incorporates mathematical developments to include any type of PDE-based population dynamics regularization. The optimal placement tools, that will be soon added to Pherosensor-toolbox, will be used to study the optimal placement in the landscape of pheromone sensors in order to enhance the accuracy of pest localization, and to study methodologies of sensors placement and replacement.

Acknowledgements

We acknowledge the support of the French National Research Agency (ANR) under the grant ANR-20-PCPA-0007 (PheroSensor).

The Journal of Open Source Software

References

Gebbers, R., & Adamchuk, V. I. (2010). Precision Agriculture and Food Security. *Science*, *327*(5967), 828–831. https://doi.org/10.1126/science.1183899

Malou, T., Parisey, N., Adamczyk-Chauvat, K., Vergu, E., Laroche, B., Calatayud, P.-A., Lucas, P., & Labarthe, S. (2024). *Biology-Informed inverse problems for insect pests detection using pheromone sensors*. Zenodo. https://doi.org/10.5281/ZENODO.11506617

Nerger, L., Hiller, W., & Schröter, J. (2005). PDAF - THE PARALLEL DATA ASSIMILATION FRAMEWORK: EXPERIENCES WITH KALMAN FILTERING. *Use of High Performance Computing in Meteorology*, 63–83. https://doi.org/10.1142/9789812701831_0006

Oerke, E.-C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, *144*(1), 31–43. https://doi.org/10.1017/S0021859605005708

Raanes, P. N., Chen, Y., & Grudzien, C. (2024). DAPPER: Data Assimilation with Python: A Package forExperimental Research. *Journal of Open Source Software*, *9*(94), 5150. https://doi.org/10.21105/joss.05150

Ridler, M. E., Van Velzen, N., Hummel, S., Sandholt, I., Falk, A. K., Heemink, A., & Madsen, H. (2014). Data assimilation framework: Linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI). *Environmental Modelling & Software*, *57*, 76–89. https://doi.org/10.1016/j.envsoft.2014.02.008