Biology-Informed inverse problems for insect pests detection using pheromone sensors

Thibault Malou, Nicolas Parisey, Katarzyna Adamczyk-Chauvat, Elisabeta Vergu, Béatrice Laroche, Paul-André Calatayud, Philippe Lucas, Simon Labarthe

Biology-Informed inverse problems for insect pests detection using pheromone sensors

Thibault Malou[1,§], Nicolas Parisey[2], Katarzyna Adamczyk-Chauvat[1], Elisabeta Vergu [1,†], Beatrice Laroche[1,3], Paul-Andre Calatayud[4,5], Philippe Lucas[6], and Simon Labarthe[1,7,8,¶]

[1]Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France
[2]INRAE, Institute of Genetics, Environment and Plant Protection (IGEPP—Joint Research Unit 1349), Le Rheu, France
[3]Université Paris-Saclay, Inria, 91120, Palaiseau, France
[4]Institut de Diversité, Ecologie et Evolution du Vivant (IDEEV), Université Paris-Saclay, CNRS, IRD, UMR Evolution, Génome, Comportement et Ecologie (EGCE), Gif-sur-Yvette, France
[5]International Center of Insect Physiology and Ecology (icipe), Nairobi, Kenya
[6]Institute of Ecology and Environmental Sciences, INRAE, Versailles, France,
[7] Univ. Bordeaux, INRAE, BIOGECO, 33610, Cestas, France
[8]Univ. Bordeaux, Inria, INRAE, 33400, Talence, France,

[§]First author: thibault.malou@inrae.fr
[†]deceased during the development of this study.
[¶]corresponding author: simon.labarthe@inrae.fr

Abstract

Most insects have the ability to modify the odor landscape in order to communicate with their conspecies during key phases of their life cycle such as reproduction. They release pheromones in their nearby environment, volatile compounds that are detected by insects of the same species with exceptional specificity and sensitivity. Efficient pheromone detection is then an interesting lever for insect pest management in a precision agroecological culture context. A precise and early detection of pests using pheromone sensors offers a strategy for pest management before infestation.

In this paper, we develop a biology-informed inverse problem framework that leverages temporal signals from a pheromone sensor network to build insect presence maps. Prior biological knowledge is introduced in the inverse problem by the mean of a specific penalty, using population dynamics PDE residuals. We benchmark the biological-informed penalty with other regularization terms such as Tikhonov, LASSO or composite penalties in a simplified toy model. We use classical comparison criteria, such as target reconstruction error, or Jaccard distance on pest presence-absence. But we also use more task-specific criteria such as the number of informative sensors during inference. Finally, the inverse problem is solved in a realistic context of pest infestation in an agricultural landscape by the fall armyworm (*Spodoptera frugiperda*).

keywords: Inverse problem, data assimilation, biology-informed estimation, PDE, pest detection, pheromone.

1. Introduction

Insect pests represent a major threat for agricultural systems, causing direct or indirect damages to crops (Oerke, 2006), with an even increasing burden in a warming climate (Deutsch et al., 2018). World trade, travel and climate changes increase the arrival of alien species that challenge existing pest control management for native insects (EFSA PLH Panel et al., 2018). Intensified insecticide use is not sustainable, due notably to the negative impact of pesticides on biodiversity, environment and human health. Alternative solutions must be developed: a promising perspective is the development of precision agriculture which aims in particular to leverage sensors and information systems to reduce the environmental impact of agricultural practices (Gebbers et al., 2010). In that perspective, early detection of pests before infestation settlement can help designing low-pesticide crop management and epidemiological surveillance policies.

Insect pheromones are a good target for early detection of pests (Conchou et al., 2019). Indeed, insects pheromones are specie-specific, guaranteeing a good specificity of any pheromone-based detection. Pheromone specificity has long been leveraged for insect detection with the mean of pheromone traps: pheromone diffusion modifies the odor landscape, attracting the targeted insects toward traps where they can be counted (McNeil, 1991). However, if pheromone traps allow for presence/absence detections, they do not allow to accurately locate the insects in their natural habitat. Recently, new advances in sensor developments paved the way towards field-based tools to explore *in situ* pheromone plumes. These sensors have different operating principles: there are insect antenna-based biosensors (Martinez et al., 2014; Myrick et al., 2011; Pawson et al., 2020), olfactory receptor-based biosensors (Khadka et al., 2019), OBP-based biosensors (Lu et al., 2014), and e-nose technologies based on conducting polymers, metal oxide semiconductor and other devices (Moitra et al., 2016; Negri et al., 2008; Steffens et al., 2014; Wehrenfennig et al., 2012). The use of pheromone sensors could allow to directly measure the odor landscape and track back pheromone plumes towards emitters, in order to detect pests at an early stage of invasion.

From a mathematical point of view, the problem of locating the emission source of a molecule knowing measures of the emission plume belongs to the class of inverse problems (Isakov, 2017), the direct problem of which is a plume propagation model. Atmospheric dispersion, plume propagation, and plume-related inverse problems have a large literature, in particular for environmental applications, such as pollution detection (Lushi et al., 2010; Stockie, 2011). Inverse problems, and more specifically data assimilation problems, are most often formulated as an optimization problem aiming to minimize the discrepancies between the direct model outputs and in-field observations by fitting the model parameters (Isakov, 2017). In the specific case where insects are the emitters, locating the emission source amounts to searching for continuous density maps of pests, which induces a high dimensional parameter space to be explored during the inverse problem. Variational approaches are well suited for such high-dimensional optimization problem, for fast computation of gradients of the objective function (Bocquet, 2014). Inverse problems are most of the time ill-posed and must be regularized by additional terms in the loss function that constrain the parameter exploration, *e.g.* with LASSO regularization term (Adam et al., 2016). Other regularization can be searched for, that could take into account prior biological knowledge incorporated into models, like in physic or biology-informed machine learning approaches (Yazdani et al., 2020), through weak or strong constraints (Trémolet, 2007).

In conventional agricultural practices, farmers often use pesticide prophylactically: knowing the potential occurrence of an infestation, they preventively apply pesticide before a pest is observed. If we interpret this strategy in the context of a classification problem, it is as if they were

classifying all their fields as "infected" whatever their actual state of infestation, thus minimizing their miss rate, i.e. a non detected true presence of pest. In so doing, they also maximize their false discovery rate. In order to minimize pesticide use, we would like to minimize the false discovery rate, so that pesticides are only used where and when a detected pest is actually present in the field. But, according to a classical trade-off in classification problems, reducing the false discovery rate comes with an increase of the miss rate, so that the detection problem could be reformulated as follows: find the minimal false discovery rate that keeps the miss rate under an acceptable risk, from the point of view of crop protection.

The main objective of this paper is to introduce and test a data assimilation framework, namely the biology-informed data assimilation, which seeks to combine a pheromone propagation model and pheromone sensors data, like in classical data assimilation, but also biological information. In particular, we specifically aim to introduce biology-informed regularization terms, on which the biology-informed data assimilation is based, in order to assess how adding biological priors improves the accuracy of the data assimilation. Several comparison criteria will be used. In particular, we will introduce a specific criteria designed to compare the number of active sensors involved during the inference process.

2. Material and methods

A sketch of the global methodological framework can be found in Fig. 1. The method contains three blocks that will be detailed hereafter:

1) the direct problem that aims to model the pheromone propagation in the atmosphere over an agricultural landscape, taking into account environmental data such as land use or wind field, see Sec. 2.1;
2) the pheromone sensors that provide data of the observed variable and an observation operator that maps the state variable of the direct model to the observed variable, see Sec. 2.3.1;
3) the corresponding inverse problem that aims to infer pheromone emissions by the insects and enables the construction of pests risk maps, knowing the data and the pheromone propagation model, see Sec. 2.2.

The proposed framework is illustrated with a toy case and a more realistic case, set up in an agricultural landscape, see Sec. 2.3.2. In the different test cases, the datasets are synthetic and constructed from a spatio-temporal sampling of pheromone concentration and a realistic model of the observation process.

2.1. 2D reaction-convection-diffusion model for the pheromone propagation

One of the most classical model to simulate propagation of chemical species in the atmosphere is a 3D Chemical-Transport Model (CTM), that has been extensively reviewed (Leelössy et al., 2014; Sportisse, 2007). The CTM are obtained from the reactive Navier-Stokes equations by using the dilution hypothesis. This hypothesis is valid when the chemical species of interest is present in very small amounts. In this case, the compound is assumed to be chemically passive, which results in a decoupling between atmospheric dynamics and compound concentration dynamics. As a consequence, dynamical field such as the wind or the diffusion coefficients can be computed separately, offline, and given as parameters of the CTM. It results that this (linear and offline-coupled) CTM is much less complex than the full (non-linear) reactive Navier-Stokes equations, or than a CTM coupled online with a fine scale Navier-Stokes model (Leelössy et al.,

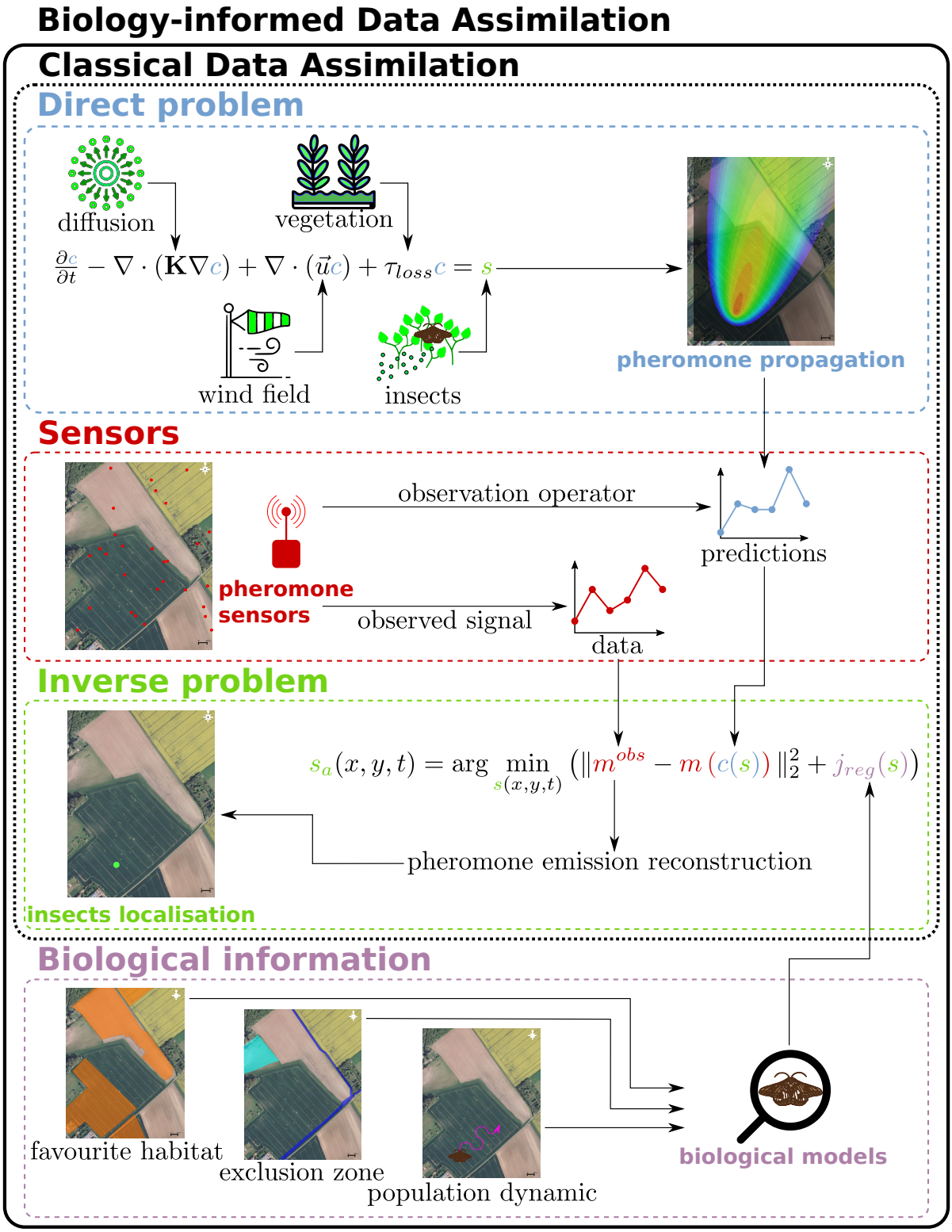


Figure 1 – Graphical abstract: summary of the proposed Biology-Informed Data Assimilation method for the pest insects localization. A classical Data Assimilation (DA) framework is first built, combining a pheromone propagation model (blue rectangle), pheromone sensors (red rectangle) into an inverse problem (green rectangle). Additionally to this classical DA framework, biological information can be supplemented through specific regularization terms of the DA problem (purple rectangle). These regularization terms are the main insights of this study.

2014), while being a reliable approximation: it is able to model the main physical phenomena at play during the transportation of pheromones.

In the present context, only the lower layer of the atmosphere, where insects and sensors are located, is of interest. Hence, for sake of computation load and data availability in real applications, a 2D depth-integrated CTM is used to model the total concentration of pheromone over the depth of the layer $c(x, y, t)$ [g m^{-2}] on a domain $\Omega \in \mathbb{R}^2$ and during a time window $[0; T]$. This depth-integrated CTM results in a 2D reaction-convection-diffusion model, that can be derived from the classical 3D CTM, see Sec. B.1 in the annex for details, and reads:

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

The coefficient $\mathbf{K}(x, y, t)$ [m^2 s^{-1}] is the diffusion tensor. This tensor models the turbulent diffusion at an unresolved scale (classically, scale smaller than the scale of the numerical grid) using the K theory, see *e.g.* (Panofsky et al., 1984; Sportisse, 2007). As mentioned above, in CTM, the wind velocity field $\vec{u}(x, y, t) = (u(x, y, t), v(x, y, t))$ [m s^{-1}] is given as parameter, where (u, v) are the velocity field components along x and y. For example, $\vec{u}$ can be computed using a large (meso-scale) meteorological model and then, taken as input data in the present CTM, see an example in Sec. 2.3.2 and on Fig. 3a. The loss coefficient $\tau_{loss}(x, y, t)$ [s^{-1}] models the loss of pheromone due to vertical phenomena, including pheromone settlement on the ground and vertical diffusion spreading the pheromone out of the layer of interest. Hence, τ_{loss} depends on the land cover and on the stability of the atmosphere, see estimations of this coefficient in Sec. B.1.3. Other features can also be included in this parameter, such as humidity-dependent deposition. In the following, the solution of the CTM model is also referred to as the pheromone plume.

In Eq. (1), the source term $s(x, y, t)$ [g m^{-2} s^{-1}] is the quantity of pheromone emitted by the insects located in (x, y, t) per unit of area and time. We highlight that this source term s is the quantity to infer to deduce the localization of the insects. Namely, this quantity can be decomposed in

$$s(x, y, t) = q(t) p(x, y, t) \tag{2}$$

with $q(t)$ [g s^{-1}] the pheromone emission rate, and $p(x, y, t)$ [m^{-2}] the insect density on the ground, so that there is a direct relationship between s and the local presence of pests. The emission rate q can incorporate biological information, such as insect-specific emission patterns.

The 2D model (1) is closed by the following initial and boundary conditions:

- a null initial condition $c(x, y, t = 0) = 0 \; \forall (x, y) \in \Omega$,
- a null diffusive flux $\mathbf{K} \nabla c \cdot \vec{n} = 0 \; \forall (x, y) \in \partial\Omega$,
- a null convective influx $\vec{u} c \cdot \vec{n} = 0 \; \forall (x, y) \in \partial\Omega \cap \{(x, y) | \vec{u}(x, y, t) \cdot \vec{n} < 0\} \; \forall t \in]0; T]$ with $\vec{n}$ the outgoing normal vector.

The pheromone propagation model (1) is solved using an implicit upwind finite volume scheme on a Cartesian grid implemented in the git repository of the PheroSensor project at `https://forgemia.inra.fr/pherosensor/pherosensor-toolbox` and (Malou et al., 2024b) (see Sec. A). The companion code to reproduce the different computations and figures of this study is available at `https://forgemia.inra.fr/pherosensor/companion-code-bi-da` and (Malou et al., 2024a).

2.2. Variational Data Assimilation for insect localization

We aim to approximate insect risk maps by inferring the spatial distribution of emitted pheromones. This inverse problem consists in estimating the source term s of the pheromone propagation

model (1), using time-series of pheromone concentration. To this purpose, we use Variational Data Assimilation (VDA) techniques.

2.2.1. Minimization problem. VDA consists in finding the value of the control variable, here s, that enables the prediction made by the direct model, here the pheromone propagation model (1), to be close to some observations, here the dataset of pheromone concentration time-series. Therefore, the inference problem is defined as a minimization problem over a cost function j (Bocquet, 2014) that can be split in an observation and a regularization term:

$$\begin{cases} \text{Find } s_a(x,y,t) \text{ such that:} \\ s_a(x,y,t) = \arg\min_{s(x,y,t)} j(s) \text{ with } j(s) = j_{obs}(s) + j_{reg}(s) \end{cases} \tag{3}$$

The term j_{obs} aims to minimize the gap between model predictions and data (Bocquet, 2014; Le Dimet et al., 1986; Sasaki, 1958). It reads

$$j_{obs}(s) = \| m(c(s)) - m^{obs} \|^2_{\mathbf{R}^{-1}} \tag{4}$$

where m^{obs} are the observations, $c(s)$ is the direct model output obtained for a given control s, $m(c)$ the observation operator that maps the direct model state variable towards the observation space and $\mathbf{R}$ is the covariance operator of the observation error. Note that the equivalent statistical notation of eq. (4) reads

$$m^{obs} = m(c(s)) + \varepsilon$$

where ε is a gaussian noise with zero mean and covariance $\mathbf{R}$.

The regularization term j_{reg} is commonly added in order to ensure the well-posedness of the optimization problem, by improving the convexity of j (Bocquet, 2014). Classical regularization terms are based on a rough background estimate of the control variable, either by a model with an insufficient accuracy or by coarse data.

The novelty and the results of this study lies on the introduction of Biology-Informed (BI) regularization terms that enable to drive the DA method toward an optimum s_a consistent with the biological knowledge introduced in j_{reg}, leading to a Biology-Informed Data Assimilation framework. In general terms, the regularization term reads:

$$j_{reg} = \sum_{i \in \mathcal{L}} \alpha_{reg,i} j_{reg,i} \tag{5}$$

with $\mathcal{L}$ the list of considered BI regularization terms $j_{reg,i}$ and $\alpha_{reg,i}$ the associated weight coefficient. This allows for composite regularization terms that take account of several biological information. The definition of these BI regularization terms is introduced in the Results in Sec. 3.1.

2.2.2. Adjoint model for gradient computation. To solve the minimization problem (3), we use a gradient-based method, suitable for our convex and differentiable optimization framework. Thus, at each descent step, the gradient $\nabla_s j$ must be computed and reads:

$$\nabla_s j(s) = \nabla_s j_{obs}(s) + \nabla_s j_{reg}(s)$$

Note that we denote ∇_s the gradient with respect to the control s to avoid confusion with ∇ the classical gradient with respect to the space variables. Since s is defined in a high-dimensional space and $c(s)$ involves the computation of a PDE model, the numerical computation of this gradient is challenging, in particular for its observation term $\nabla_s j_{obs}$. Noting the adjoint operator with

$*$, we compute this gradient by using the adjoint model of the CTM (Lions, 1971; Pudykiewicz, 1998) that reads:

$$(6)\quad \partial_t c^* + \nabla \cdot (\mathbf{K}^T \nabla c^*) + \nabla(\vec{u} c^*) - (\nabla . \vec{u}) c^* - \tau_{loss} c^* = \left(\frac{dm}{dc}(c(s)) \right)^* \cdot 2\mathbf{R}^{-1} \left(m(c(s)) - m^{obs} \right)$$

with c^* the adjoint state.

The adjoint model (6) is closed by the following final and boundary conditions:

- a null final condition $c^*(x, y, t = T) = 0 \; \forall (x, y) \in \Omega$,
- a null diffusive flux $\mathbf{K}^T \nabla c^* \cdot \vec{n} = 0 \; \forall (x, y) \in \partial\Omega$,
- a null outgoing convective flux $\vec{u} c^* \cdot \vec{n} = 0 \; \forall (x, y) \in \partial\Omega \cap \{(x, y) | \vec{u}(x, y, t) \cdot \vec{n} > 0\} \; \forall t \in]0; T]$ with $\vec{n}$ the outgoing normal vector.

Further details on the derivation of the adjoint model (6) can be found in Sec. B.3.1. This adjoint model (6) can be interpreted as the retro-propagation with diffusion of the error between the prediction of the observed variable and the observations $m(c(s)) - m^{obs}$ from the final time $t = T$ to the initial time $t = 0$. The output of the adjoint model is also referred to as the retro-plume in the following. Let us note that the adjoint model (6) requires the observation operator to be differentiable.

Once the adjoint state is computed, the gradient $\nabla_s j$ is computed using the expression (see more details in Sec. B.3.1):

$$(7)\quad \nabla_s j(s) = \nabla_s j_{reg}(s) - c^*(s)$$

The gradient-descent algorithm is initiated with a null control $s_0 = 0$.

The adjoint model (6) is solved using an implicit downwind finite volume scheme on the same Cartesian grid as the direct problem. A special care has been put on the time discretization so that the discretized adjoint problem is effectively the adjoint of the discretized direct problem. The gradient-descent algorithm is implemented in a custom script. The whole VDA process is implemented in the git repository of the PheroSensor project (see Sec. A).

2.2.3. Sensor contribution to gradient evaluation. We stress that in our context of a network of independent sensors dispatched over the landscape, we can assume that the observation error covariance $\mathbf{R}$ is block-diagonal. Moreover, let us note that the left-hand side of Eq. (6) is linear by definition of the adjoint model. In addition, we assume that the observations operator can be split into a sum of one-sensor observation operator $m(c) = \sum_i m_i(c)$ (avoiding the degenerate case of overlapping sensors). From this assumption, we can inject $m = \sum_i m_i$ in the adjoint model (6) and split the source term to get the following one-sensor adjoint model:

$$(8)\quad \partial_t c_i^* + \nabla \cdot (\mathbf{K}^T \nabla c_i^*) + \nabla(\vec{u} c_i^*) - (\nabla . \vec{u}) c_i^* - \tau_{loss} c_i^* = \left(d_c m_i(c(s)) \right)^* \cdot 2\mathbf{R}^{-1} \left(m_i(c(s)) - m_i^{obs} \right)$$

with c_i^* the one-sensor adjoint state related to the i^{th} sensor. Since the adjoint model (6) is linear, it results that the adjoint state c^* can also be expressed as the sum of the c_i^*:

$$(9)\quad c^* = \sum_i c_i^*.$$

It results that c_i^* is the contribution of the i^{th} sensor to c^*, and thus to the gradient of j_{obs}. This observation will be used later to compare the contribution of each sensor to the data assimilation. More details on the derivation can be found in the annex Sec. B.3.1. Let us note that in our case, the assumption that the observations operator can be split into a sum of one-sensor observation operator is true, see Sec. 2.3.1.

2.2.4. Stopping criteria. As per Morozov's discrepancy principle, the optimization algorithm should be continued until the gap between the observations and the system state achieves the level of the accuracy of the observations. Hence, the gradient descent algorithm is terminated and the optimum s_a is assumed to be reached when the cost j_{obs} attains a value lower than the noise norm $\|\epsilon\|^2_{L^2(\Omega_{obs})}$.

2.3. Pheromone sensors and test cases

2.3.1. Observation operator of the pheromone sensors. The state variable c is a function defined over $[0,\, T] \times \Omega$. We assume that we have at disposal N_s sensors that are dispatched throughout the landscape and remain in the same position (x_i, y_i) all along the time domain. We furthermore assume that, given $[t_1^{obs}, \cdots, t_{N_T}^{obs}]$ a time sampling of the time domain, every sensor produces a data at time t_k^{obs}, $k = 1, \cdots, N_t$, so that the observation space (O) has dimension $N_s \times N_t$. Let us note $e_{i,k}$ the vector, for sensor i and time k, of the canonical basis of $mathcalO$.

Then, we define the observation operator $c \mapsto m(c)$ mapping the state variable c towards the observation space $\mathcal{O}$ with the equation

$$m(c) = \sum_{i=1}^{N_s} m_i(c) \text{ with } m_i(c) = \sum_{k=1}^{N_t} e_{i,k} \int_{t_k^{obs}-\delta_t}^{t_k^{obs}} \int_{\Omega} \delta_{(x_i,y_i)}(x,y)c(\tau,x,y)d\tau dxdy \tag{10}$$

where $\delta_{(x_i,y_i)}$ is the Dirac function at the localization of the i^{th} sensor and δ_t a time window. For observation k of sensor i, this expression averages over a time window of width δ_t before the observation time the pheromone concentration at location (x_i, y_i). This models a pre-concentrator that actively filters air during a time window δ_t and releases the accumulated quantity into the sensor for the mesure. Let us note that the time sampling of the observations must satisfy $\delta_t < t_1^{obs}$ so that $t_1^{obs} - \delta_t > 0$. Let us note that in this case, the operator $c \mapsto m(c)$ is linear and is split in the sum of one-sensor observation operators m_i, as mentioned in Sec. 2.2.3.

2.3.2. Setting up test cases. All along the paper, we will consider two test cases. The first one is a simplified yet illustrative scenario, referred to as the Toy case, that will be used to test and analyze different biology-informed regularization strategies. The second one depicts a realistic and representative scenario for invasion of an alien insect in an agricultural landscape. We chose the Fall ArmyWorm (FAW), a.k.a. *Spodoptera frugiperda*, as the model insect considered in this study. The FAW is a moth that belongs to the family Noctuidae. This insect pest induces the highest yield loss for maize (Savary et al., 2019), but can grow on hundreds of plant hosts, causing severe damage in grasses (IPPC Secretariat, 2021). It is expected that global warming reduces the populations in the southern America area due to warmer and drier climates (IPPC Secretariat, 2021), but could favour the emergence of the species in the European Union (EFSA PLH Panel et al., 2018). It is then a suitable test case for a realistic scenario in an epidemiological surveillance framework in Europe. In this section, we will detail the set up of the environmental and landscape parameters needed for solving the pheromone propagation model (1) in the two test cases. The inference method and the additional biology-informed regularization terms are tested using synthetic dataset generated from a reference source term, hereinafter referred to as the target source term s_t. Further details for data generation and inference will be added in this sections.

Toy case. The computation spatial domain of the Toy case is a rectangular grid of size $40\text{m} \times 60\text{m}$ (i.e. 0.24ha) over a time window of $T = 20\text{s}$. It is built up upon a fictive insect that constantly releases pheromone at an emission rate of $q = 2.7 pg.s^{-1}$ per insect. The insects density is

constant $p_0^{toy} = 0.2m^{-2}$ in a fixed area of the domain located in a circular support $\Omega_{supp} = \mathcal{S}$ centred in $(x_s, y_s) = (10, 10)$ and of radius $R = 5m$, see red circle in Fig. 2, which corresponds to approximatively 16 individuals. The wind field u is constructed from a constant wind speed $|\vec{u}| = 8m.s^{-1}$ and a spatially-homogeneous time-dependent wind direction that linearly varies from $\theta = \frac{\pi}{4}$ at $t = 0$ to $\theta = \frac{\pi}{2}$ at $t = T$. The loss coefficient τ_{loss} is computed to approximate realistic vertical fluxes in a 3D simplified CTM model (see more details in Sec. B.1), resulting in a constant value of $\tau_{loss} = 0.01\text{s}^{-1}$. The diffusion tensor $\mathbf{K}$ is assumed to be isotropic and homogeneous over time and space. As the diffusion value $\mathbf{K} = 50\text{m}^2\,\text{s}^{-1}$ is a typical value to model an unstable atmosphere (BEA Fisher et al., 1981), we use the value $\mathbf{K} = 10\text{m}^2\,\text{s}^{-1}$, which corresponds to a moderately unstable atmosphere.

$N_s = 35$ synthetic sensors are randomly placed each at a different position in the landscapes with a uniform distribution, see Fig. 2. Moreover, the accumulation time step is set to $\delta_t = 2s$. Noise is subsequently added. We take $m_{obs}(t_i, x_j, y_j) = max(0, m(c(s_t); t_i, x_j, y_j) + \epsilon_{i,j})$, in order to avoid negative values, with $c(s_t)$ the results of the direct model (1) computed from the source term s_t and $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma_{i,j}^2)$ a Gaussian noise. Since pre-concentration uncertainties scale with pheromone density, we scale the standard deviation with the model observations: $\sigma_{i,j} = 0.01m(c(s_t); t_i, x_j, y_j)$.

A favorable habitat, used in the Tikhonov regularization, see Sec. 3.1.1, is defined as a square with a width of 20m, overlapping the source support $\mathcal{S}$, see zone delimited by the dotted red line on Fig. 2.

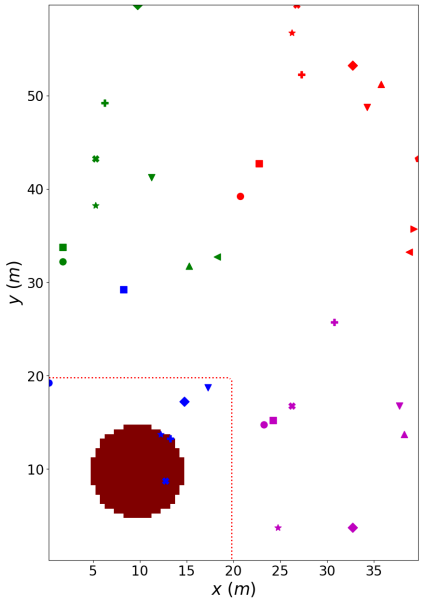


Figure 2 – Map of the target source term s_t circular support (constant within the red circle), the favorable habitat (square zone delimited by the red dotted line) and the sensor location (colored markers) for the Toy test case.

Realistic set up: the FAW case. The FAW case is based on a real landscape in central France (Adamczyk-Chauvat et al., 2024), see Fig. 3a for the computational domain (dashed lines). The same homogeneous diffusion tensor ($\mathbf{K} = 10\text{m}^2\,\text{s}^{-1}$) than the Toy case is used. The wind field $\vec{u}$ is recovered from outputs of the French weather forecast model Arome, which is the regional numerical weather forecast model from MétéoFrance (French weather forecast service, data and documentation available at `https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=131&id_rubrique=51`) corresponding to the actual wind field in area of interest at noon

the 2022/11/16. As the spatial sampling of the Arome model is about $0.01°$ ($\approx 1.3km$), see Fig. 3a, the wind field is projected on the computation grid by linear interpolation, and kept stationary over the computation time window.

As the loss coefficient τ_{loss} depends on the land cover, the land use heterogeneities in the landscape are taken into account in the FAW case using land cover data, see Fig. 3b. The different values of τ_{loss} depending on the land cover as recapitulated in Tab. 1, see Sec. B.1 for computations. Note that in the Toy case, the constant value 0.01 corresponds to bare soil.

land use	sum. crop	maize	wint. cereals	grassland	forest	bare soil	urban	road
τ_{loss} (s^{-1})	0.0097	0.011	0.0105	0.0099	0.0095	0.01	0.009	0.009

Table 1 – Loss coefficient τ_{loss} values depending on the land use. Sum. crop = Summer crop, wint. cereals = winter cereals.

FAW has a specific pattern of pheromone emission. During the mating period, the FAW females search for a suitable location and stay there until they lay their eggs. Each night, they emit a sexual pheromone that attracts the males. Pheromone emission is stopped just at the beginning of the mating process, so that at the population level, the pheromone emission exponentially decreases while females meet a mating partner and stop emitting. This pattern is repeated every night. We start the time window for pheromone recording 4h after nightfall, when emissions are at their peak (Tumlinson et al., 1986).

Recall that $s = pq$ with p the density of insects and q the quantity of pheromone emitted per time unit per insect. We use a constant emission rate per insect $q = 2.7pg.s^{-1}$. The reference density p_{ref}^{FAW} of emitter insects is chosen to satisfy a realistic population dynamics model, see later Sec. 3.7, and is defined as

$$p_{ref}^{FAW}(t,x,y) = p_0^{FAW}(x,y)\exp\left(-\int_0^t \gamma(s)ds\right) \tag{11}$$

with $\gamma(t)$ the mating rate function representing the proportion of female that encounter a male at a given time and stop emitting (see Fig. 3c) and p_0^{FAW} the initial density of females. In that regard, $t \mapsto \exp\left(-\int_0^t \gamma(s)ds\right)$ is the function representing the proportion of females still available for mating and that keep emitting at a given time. Initially, we suppose that approximately 60 females are uniformly distributed within a circular domain of radius $10m$. This corresponds to a constant initial density of $\bar{p}_0 = 0.2m^{-2}$ within $\mathcal{S} = \{(x,y) \in \Omega \mid (x-x_s)^2 + (y-y_s)^2 < R^2\}$ a circular support centred in (x_s, y_s) and of radius $R = 10m$, see the red circular domain on Fig. 3b:

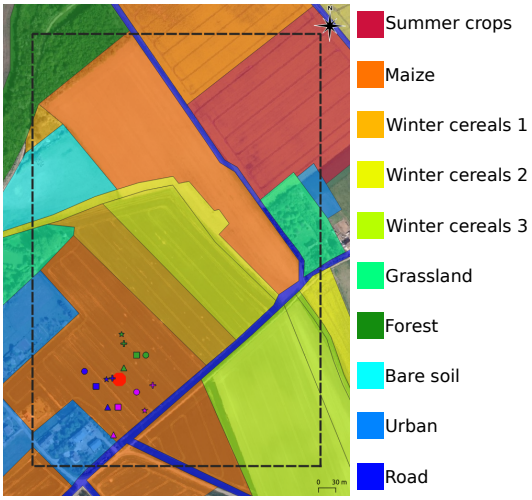
$$p_0^{FAW}(x,y) = \bar{p}_0 \mathbb{1}_{\mathcal{S}}(x,y). \tag{12}$$

The FAW females usually emit when laying on plants favorable for the larva. In the FAW case landscape, we assume that maize is the most favorable plant. Therefore, the support $\mathcal{S}$ is located within a domain region covered by maize, see land cover in dark orange on Fig. 3b. Moreover, we assume that the mating rate function γ is a logistic function, see blue curve on Fig. 3c:

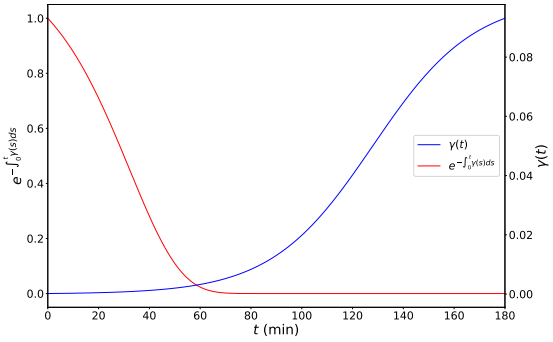
$$\gamma(t) = \frac{K_{mat}}{1+\exp\left(-\lambda(t-\tau)\right)} \tag{13}$$

where K_{mat} is the maximal mating rate, τ the midpoint time value and λ the logistic growth rate (see Fig. 3c). Let us note that τ is related to the average time a male necessitates to reach a female while tracking back the pheromone source. Based on biological expertise, τ is taken such that 50% of the females have mated with a male within the 30 first minutes. Hence, $p_{ref}^{FAW}(t = 30min, x, y) = \frac{1}{2}p_0^{FAW}(x,y)$. Using the mating maximal rate $K_{mat} = 0.1$ and a logistic growth rate $\lambda \approx 8.3 \times 10^{-3}$, it results that $\tau = 127.94min$.

(a) Wind field as obtained from the weather forecast model, before interpolation on the computing grid (length and color of the arrows normalized by the wind speed). Nine vectors in total are available, at the center of the figure, at the corners and in the cardinal directions.



(b) Fragment of agricultural landscape, land cover and sensor positions (see sec. 3.6 and fig 13).



(c) FAW test case, mating rate $\gamma(t)$ as in Eq. (13) (blue) and total proportion of female still emitting pheromones $\exp\left(-\int_0^t \gamma(s)ds\right)$ (red, see Eq. (11)) vs time t.

Figure 3 – Map of the of the computations domains (dashed lines, black for the total domain and red for the domain of the inference) with the wind field computed by the Arome model at a given time (left) and the division of agricultural parcels and soil occupation (right) for the FAW case.

In a first time, 15 sensors are randomly dispatched with uniform distribution in the maize field of interest, see the land cover in dark orange on Fig. 3b. Then, the 6 sensors that contribute the less (6 last in the contribution order for the one-sensor criteria defined in Sec. 2.4.3) are randomly relocated inside the convex hull of the 9 remaining sensors. This relocation method is detailed in the Results in sec. 3.6 and illustrated in Fig. 13.

Note that the computation domain is narrowed around the favorable zone of maize field to reduce computation time, see domain delimited by the red dashed line on Fig. 3a.

The accumulation time step is set to $\delta_t = 30min$, in the order of reasonable pre-concentration time. Noise is subsequently added with the same noise model than in the Toy case: $m_{obs}(t_i, x_j, y_j) = max(0, m(c(s_t); t_i, x_j, y_j) + \epsilon_{i,j})$ with $c(s_t)$ the results of the direct model (1) computed from the source term s_t and $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma_{i,j}^2)$ a Gaussian noise with standard deviation $\sigma_{i,j} = 0.05m(c(s_t); t_i, x_j, y_j)$.

2.4. Benchmarking criteria

Once the optimization algorithm is terminated, the optimum s_a is compared against the target s_t, that is used to generate the data, using the criteria presented below to evaluate the accuracy of the localization of the insects.

2.4.1. RMSE-type criteria. A first class of criteria is based on the Root x Mean Square Error (RxMSE) between the target s_t and a value of s, that measures the accuracy of the data assimilation process. The x indicates here that we average over the spatial dimension only. In this study, the errors of interest are the quadratic error on the whole domain Ω and on Ω_{supp}, the support of the target s_t. We set the total RxMSE criteria by

$$RxMSE_{tot}(t;s) = \sqrt{\frac{1}{|\Omega|}\int_{\Omega}(s(X,t)-s_t(X,t))^2 dX} \tag{14}$$

and the support RMSE criteria by

$$RxMSE_{supp}(t;s) = \sqrt{\frac{1}{|\Omega_{supp}|}\int_{\Omega_{supp}}(s(X,t)-s_t(X,t))^2 dX}. \tag{15}$$

While the error on the whole domain gives a global assessment of the accuracy of s_a, by also measuring false emissions estimated outside of the target support, the error on the target support enables to quantify the error made on the specific zone where the pheromone are emitted, and thus where the insects are located.

2.4.2. False discovery rate and miss rate. The presence of insects at a given time and point of the domain is assessed when the quantity of pheromone emitted is above a given threshold ϵ_s. This threshold depicts the minimal quantity of pheromone emitted by a single insect over the area of a numerical mesh cell. According to this presence criteria, the subdomain where the insects are located given a source term is the superlevel set

$$\Omega_{sls}(t;s,\epsilon_s) = \{(x,y)\in\Omega \mid s(x,y,t) > \epsilon_s\} \tag{16}$$

This implies that the superlevel set $\Omega_{sls}(t;s_a,\epsilon_s)$ (set in blue dashed lines on Fig. 4) is the subdomain where the insects are predicted to be present. On the other hand, the superlevel set $\Omega_{sls}(t;s_t,\epsilon_s)$ (set in red dashed lines on Fig. 4) is the subdomain where the insects are truly present.

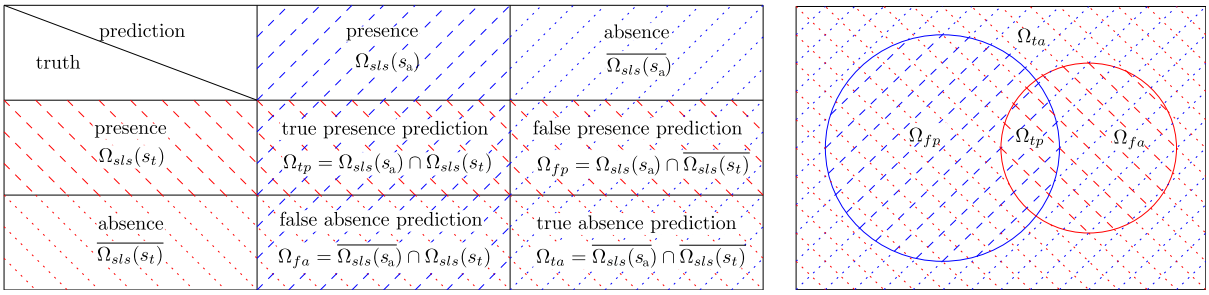


Figure 4 – Illustration of the predicted and true superlevel sets and the false absence and presence predictions

These two sets will be compared using dissimilarity criteria in a same way as the Jaccard similarity coefficient or the confusion matrix in supervised statistical learning, see *e.g.* (Stehman, 1997). The subdomain $\Omega_{fp}(t;\epsilon_s) = \Omega_{sls}(t;s_a,\epsilon_s) \cap \overline{\Omega_{sls}(t;s_t,\epsilon_s)}$ corresponds to the subdomain

where the insects are wrongly predicted to be present (false presence prediction, set in red dotted lines and blue dashed lines on Fig. 4) and $\Omega_{fa}(t;\epsilon_s) = \overline{\Omega_{sls}(t;s_a,\epsilon_s)} \cap \Omega_{sls}(t;s_t,\epsilon_s)$ the subdomain where the insects are wrongly predicted to be absent (false absence prediction, set in red dashed lines and blue dotted lines on Fig. 4). Noting $|\Omega|$ the area of the set Ω, we then define $\rho_{fa}(t)$ the false absence prediction rate (or miss rate, *i.e.* the proportion of insect true presence that is predicted as absence):

$$\rho_{fa}(t) = \frac{|\Omega_{fa}(t)|}{|\Omega_{sls}(t;s_t,\epsilon_s)|} = \frac{|\Omega_{fa}(t)|}{|\Omega_{fa}(t) \cup \Omega_{tp}(t)|} \tag{17}$$

and $\rho_{fp}(t)$ the false presence prediction rate (or false discovery rate, *i.e.* the proportion of predicted presence that is incorrect):

$$\rho_{fp}(t) = \frac{|\Omega_{fp}(t)|}{|\Omega_{sls}(t;s_a,\epsilon_s)|} = \frac{|\Omega_{fp}(t)|}{|\Omega_{fp}(t) \cup \Omega_{tp}(t)|}. \tag{18}$$

In a context where we aim to prevent and control the infestation, this miss rate must be kept the smallest possible, or even 0. On the other hand, in a context of precision agriculture, the false discovery rate will lead to useless application of pesticide. Thus, the objectif is to have the smallest possible false discovery rate while keeping the miss rate close to 0. In the following, we use the threshold value $\epsilon_s = 0.065$.

2.4.3. One-sensor adjoint state-based criteria. The purpose of this criteria is to assess how the regularization strategies influence indirectly the contribution and importance of the sensors and their data, and if some strategies require less sensors to achieve the inference with a similar accuracy on s_a.

Let $(s_k)_{k\in\mathbb{N}}$ denotes the sequence of source term given the trajectory taken by the optimization algorithm in the control space. The gradient descent algorithm iterations can then be expressed as $\forall k > 0$:

$$s_{k+1} = s_k - \eta_k \nabla_s j(s_k) \tag{19}$$

with $(\eta_k)_{k\in\mathbb{N}}$ the sequence of step sizes. From this expression of one iteration, the optimum s_a can then be expressed as function of the sequence $(s_k)_{k\in\mathbb{N}}$:

$$s_a = s_0 - \sum_k \eta_k \nabla_s j(s_k) \tag{20}$$

Using the expression (7) of $\nabla_s j$ and the splitting of the adjoint state c^* into the one-sensor adjoint state c_i^* as in Eq. (9), the optimum s_a can then be expressed as:

$$\begin{aligned} s_a &= s_0 - \sum_k \eta_k \Big(\nabla_s j_{reg}(s_k) - c^*(s_k) \Big) \\ &= s_0 - \sum_k \Big(\eta_k \nabla_s j_{reg}(s_k) \Big) + \sum_i \sum_k \Big(\eta_k c_i^*(s_k) \Big) \\ &= s_0 - \sum_k \Big(\eta_k \nabla_s j_{reg}(s_k) \Big) + \sum_i c_{i,tot}^* \end{aligned}$$

where

$$c_{i,tot}^*(x,y,t) = \sum_k \eta_k c_i^*(s_k;x,y,t) \tag{21}$$

This illustrates that $c_{i,tot}^*$ enables to quantify the contribution of the i^{th} sensor to the computation of the optimal s_a. The comparison of the regularization strategies with regard of the sensors contribution will be done by looking at $\|c_{i,tot}^*\|_{L^\infty(\Omega\times[0,T])}$, the maximal contribution of

the i^{th} sensor, and the sensors will be ordered by decreasing $\|c^*_{i,tot}\|_{L^\infty(\Omega\times[0,T])}$, hereinafter referred to as the contribution order. This representation allows to compare the contribution of the sensors with each other and to decipher if the maximal total contribution drops from a certain rank in the contribution order, thus indicating that the sensors located upwards in the ranking may have lower, or even negligible, impact on the estimation of s_a. Let us note that the quantity $c^*_{i,tot}$ can be directly computed using a trajectory-integrated one-sensor adjoint model in eq. (69) and does not require to compute $c^*_i(s_k)$ for each k separately, see more details in Sec. B.3.1.

3. Results

The core methodological improvement of this study is the Biology-Informed Data Assimilation (BI-DA) framework. It is the reason why we present it in the results and not in the material and methods section. This framework is based on the classical data assimilation optimization problem (3) but with in addition regularization terms that are informed with biological knowledge of the insect. These BI-DA framework is tested in the two test cases defined in Sec. 2.3.2, i.e. the Toy case, characterized by small spatio-temporal scales, homogeneous landscape, time-varying but space-uniform wind field and constant emission pattern of the emitting insects, and the FAW case taking into account a real agricultural landscape, observed wind field and complex emission pattern mimicking the behavior of the insect pest targetted in this study: the FAW.

We first present different biology-informed regularization terms in Sec. 3.1. In Sec. 3.2, the pheromone propagation is illustrated in the Toy case, before computing a classical VDA without regularization in Sec. 3.3, to obtain an optimal pheromone emission s_a of reference. In Sec. 3.4, the BI-DA is extensively benchmarked and compared to regularization-free VDA in the Toy case. Finally, BI-DA is used to track emission source in the FAW case. In Sec. 3.5, the pheromone propagation is illustrated in this realistic setting. In Sec. 3.6, a classical VDA used as reference is performed without regularization, raising the need for a better sensor placement. In Sec. 3.7, BI-DA is performed to infer the localization of the insects.

3.1. Biology-Informed DA: adding biological priors in pheromone emission inference

As mentioned in Sec. 2.2, regularization terms based on additional coarse estimation of the control are often considered to improve the convexity of j (Kaltenbacher et al., 2008; Tarantola, 2005). In the classical VDA framework, the optimization problem (3) is strongly constrained by the pheromone propagation model (1): the pheromone distribution c is explicitly obtained from the direct problem, given the source s. Regularization terms can be seen as additional weak constraints, *i.e.* constraints introduced in the minimization problem as penalties. Inference problems that rely only on optimization problems weakly-constrained by a dynamical model are already treated in VDA (Trémolet, 2007) and machine learning contexts (Raissi et al., 2019).

We specifically introduce regularization terms based on biological models, and thus term the resulting VDA problem as BI-DA. Let us assume that s satisfies a biological model that can be written $\mathcal{M}(s) = \sigma$, where σ is a known biological prior. Then, one can consider a regularization term of the form

$$j_{reg}(s) = \|\mathcal{M}(s) - \sigma\|^2_{L^2} \tag{22}$$

The regularization term is then designed to minimize the residual of the model $\mathcal{M}(s) = \sigma$.

The model's operator $\mathcal{M}$ can be any kind of operator, including partial differential operator. For instance, recall that the emission source can always be written $s = qp$ as in Eq. (2), and thus, the model $\mathcal{M}(s) = \sigma$ can be resulting from a population dynamics model on the insect density p. In the VDA minimization problem (3), we work in a differentiable optimization framework.

Thus, the loss function $j(s)$ must be differentiable, which implies that $j_{reg}(s)$ and the underlying operator $s \mapsto \mathcal{M}(s)$ have to be differentiable as well. However, this framework can be extended by working in a non-smooth and subdifferentiable optimization framework (Nesterov, 2018), e.g. by using the proximal-gradient algorithm (Parikh et al., 2014). This allows us to consider subdifferentiable $j_{reg}(s)$ and $\mathcal{M}$, such as a LASSO regularization term, *i.e.* of the form $j_{reg}(s) = \|\mathcal{M}(s) - \sigma\|_{L^1}$ or alike, see more details in Sec. 3.1.2. Recall from Eq. (5) that a weighted sums of multiple regularization terms can be considered in order to take into account multiple biological models.

We now introduce several biology-informed regularization terms that can be used to regularize the optimization problem (3). First, classical regularization terms will be introduced, such as Tikhonov (a.k.a. ridge), LASSO and group-LASSO regularization. For these regularization terms, we will pay a specific attention to the biological information we can add through these terms. Then, a population dynamic-informed j_{reg} will be presented, in order to explicitly model biological mechanisms.

3.1.1. Biology-based Tikhonov regularization term. The most classical j_{reg} is the Tikhonov regularization term (Kaltenbacher et al., 2008):

$$j_{reg,T}(s) = \|s - s_b\|^2_{\mathbf{C}^{-1}} \tag{23}$$

with s_b a background value of s and $\mathbf{C}$ the background error covariance operator. Classically, s_b is given by a prior estimate of the true state, *e.g.* based on additional data that are not accurate enough and/or too sparse or based on a rough physical estimate , see *e.g.* (Bocquet, 2014). From a biological point of view, a Biology-Informed Tikhonov j_{reg} ($\mathcal{M}(s) = s$ and $\sigma = s_b$) enables to take into account prior knowledge on the insects life habits, for instance their preferred habitat and insect life cycle. In this case, the term s_b describes the expected spatial distribution and density of the insects with respect to the land use or the pheromone emission pattern.

3.1.2. LASSO-type regularization term and corresponding biological input. LASSO-type regularization terms are often used to ensure sparsity and perform variable selection, see (Tibshirani, 1996). In the VDA experiments based only on the data, see Secs. 3.3 and 3.6, some spurious sources are detected, including near several sensors even though they are far away from the target source. This residual detection, particularly visible near final time (see Fig. 6), can be due to numerical error, to noise or to diffusive phenomena in the adjoint model. When computing the superlevel set in eq. (16) for locating insects from the inferred quantity of pheromone emitted, these spurious pheromone sources can lead to false presence prediction.

LASSO regularization. In our context, we assume that the data are collected at the beginning of the infestation, so that we can expect that only a few insects are present. Hence, in order to select important local pheromones sources and to mitigate false presence predictions, a LASSO penalty is considered:

$$j_{reg,LASSO}(s) = \|s\|_1 \tag{24}$$

Recall that the optimization problem (3) with this j_{reg} is no longer differentiable. A proximal gradient algorithm, namely the ISTA method, will be used to solve it, see more details in Annex B.3.3. Note that the ISTA is a Iterative Soft-Threshold Algorithm, meaning in particular that at each optimization iterations it thresholds any value under $\alpha_{reg,LASSO}\eta$, where $\alpha_{reg,LASSO}$ is the regularization weight of $j_{reg,LASSO}$ and η is the descent step of the gradient-descent algorithm. Hence,

assuming that we have a prior knowledge of a detection threshold under which pheromone emission can be discarded, a magnitude of $\alpha_{reg,LASSO}$ can be computed from η in order to shrink the estimate of s_a under this threshold at each optimization iterations.

Group-LASSO regularization. It is common that the insects do not move during the pheromone emission time period. It is *e.g.* the case for FAW, see Sec. 2.3.2. Therefore, the total quantity of pheromone emitted during this emission period can not be lower than the total amount of pheromone emitted by one insect on this time window. Conversely, if the total amount of emitted pheromone is lower, it implies that these pheromone sources are spurious and can be discarded. The following group-LASSO term is adapted to eliminate spatial positions with lowest accumulated pheromone emission:

$$j_{reg,gLASSO}(s) = \|s\|_{L^1(\Omega;L^2(0,T))} = \int_\Omega \left(\int_0^T s(t,x,y)^2 dt \right)^{1/2} dxdy \tag{25}$$

Similarly to the LASSO regularization, a proximal gradient algorithm will be used and more details are provided in B.3.3. For the group-LASSO regularization, this algorithm thresholds at each optimization iterations $s(t,x,y)$ for all t if $\|s(\cdot,x,y)\|_{L^2(0,T)}$ is under $\alpha_{reg,gLASSO}\eta$, where $\alpha_{reg,gLASSO}$ is the regularization weight of $j_{reg,LASSO}$ and η is the descent step of the gradient-descent algorithm.

3.1.3. Population dynamic-informed regularization term. Population dynamics models that model the spatio-temporal evolution of the density of insects p can also be used to derive biological models for the quantity of pheromone emitted s. These population dynamics models can be written in a generic way as follows:

$$\partial_t p + \nabla \left(\sum_i F_i(p) \right) - \sum_i R_i(p) = 0 \tag{26}$$

with F_i and R_i are resp. flux and reaction terms that model the behavior of the insect specie of interest. Recall that $s(x,y,t) = p(x,y,t)q(t) \Leftrightarrow p(x,y,t) = \frac{s(x,y,t)}{q(t)}$ with $q(t) > 0$ the quantity of pheromone emitted per insect, which is assumed to be a biological knowledge. A population dynamics PDE equation of the form (26) can be expressed as the following PDE equation, in which s is the state variable:

$$\frac{1}{q}\partial_t s - \frac{\partial_t q}{q^2} s + \nabla \left(\sum_i \mathcal{F}_i(s) \right) + \sum_i \mathcal{R}_i(s) = 0$$

with $F_i(p = \frac{s}{q}) = \mathcal{F}_i(s)$ and $R_i(p = \frac{s}{q}) = \mathcal{R}_i(s)$. From this PDE equation, the population dynamics-informed regularization term is the following:

$$j_{reg,PD}(s) = \alpha_{reg,PD} \| \partial_t s - \frac{\partial_t q}{q} s + q \nabla \left(\sum_i \mathcal{F}_i(s) \right) - q \sum_i \mathcal{R}_i(s) \|_{L^2}^2 \tag{27}$$

Several potential terms F_i and R_i (and the associated $\mathcal{F}_i$ and $\mathcal{R}_i$) are possible and have different biological meaning. The flux terms F_i can be used to model the motion of the insects in a random direction or in a specific direction (*e.g.* toward the preferred habitat). The reaction terms can be used to model birth-death processes or the fact that certain insects stop emitting pheromones. An extensive list of additional possible terms is added in the supplementary material Sec. B.2, including advection, diffusion, quorum sensing, Keller-Segel term to model pheromone sensing, and additional reaction terms such as linear birth-death process, logistic growth, Allee effect, or non-local terms. Furthermore, the corresponding gradients are explicitly computed in the supplementary Sec. B.3.2, so that the reader has all the necessary elements to

implement the population dynamic-informed penalty fitting with the behavior of other insects of interest.

For instance, the females of the model insect that we consider, FAW, start emitting pheromones once found an emission site. During the emission time lapse, the insect does not move in order to efficiently attract a mating partner and stops emitting when it founds a mating partner and the mating starts. Furthermore, a proportion of female encounter a male at a given time following a mating rate function, see Sec. 2.3.2. Thus, the dynamics of the population of the insects emitting pheromone is modelled using only a simple linear birth-death process:

$$R_{bd}(p;x,y,t) = -\gamma(x,y,t)p(x,y,t) \Leftrightarrow \mathcal{R}_{bd}(s;x,y,t) = -\frac{\gamma(x,y,t)}{q(t)}s(x,y,t) \tag{28}$$

with γ [s^{-1}] the birth (if $\gamma > 0$) or death (if $\gamma < 0$) rate. In general manners, this term models a linear birth-death process where a proportion $\gamma(t)$ of the population dies or reproduces at each time t.

3.1.4. Choices of hyperparameter and hot start. As mentioned previously, we will use the different regularization terms, which includes using them simultaneously. In this regard, recall that the overall regularization term can be expressed as the weighted sum of BI regularization terms (5) and $\mathcal{L}$ is the list of the regularization terms that we take into account. In the present study, we set $\mathcal{L} = \{'T','LASSO','gLASSO','PD'\}$ where T stands for Tikhonov, PD for population dynamics and $gLASSO$ for group-LASSO. In the following, several combination of $j_{reg,i}$ will be tested and analysed, starting from only one $j_{reg,i}$ ($\alpha_j = 0\ \forall j \neq i$) and up to three $j_{reg,i}$. The composite cases $'T+LASSO'$, $'T+PD'$ and $'LASSO+PD'$ designate the cases when, additionally to $\alpha_{gLASSO} = 0$, we have respectively $\alpha_{PD} = 0$, $\alpha_{LASSO} = 0$ and $\alpha_T = 0$. Finally, the case $'T + LASSO + PD'$ is when only $\alpha_{gLASSO} = 0$ and is designated as $'all\ reg'$ in the Toy case. Similarly, the case $'T + gLASSO + PD'$ is when only $\alpha_{LASSO} = 0$ and is designated as $'all\ reg'$ in the FAW case. Moreover, in the following, $'no\ reg'$ refers to the case without regularization terms ($\alpha_i = 0\ \forall i \in \mathcal{L}$).

The values of α_i are chosen such that they are the largest possible while satisfying the Morozov's discrepancy principle (see Sec. 2.2.4) in a reasonable computation time. In the following, in order to facilitate the solving of optimization problem, the regularization terms are progressively added. This process is also known as a hot start. In a first time, we optimize without regularization. Then, we add one by one the regularization terms initiating the optimization with the optimum obtained previously. Similarly, in order to reach the highest values possible, the weight coefficients α_i are increased step by step, restarting each time with hot start and repeating until the Morozov's discrepancy principle is no longer satisfied.

3.2. Numerical illustration of pheromone propagation in the Toy case

We first present numerical results for the pheromone propagation model (1) in the Toy case, as set up in Sec. 2.3.2, in order to illustrate the pheromone plumes computed by the model (see Fig. 5). As expected, the convection is dominant: the pheromones mainly propagate in the direction of the wind, with little additional diffusive effect. The pheromone crosses the domain within a couple of seconds during which a front propagates until reaching a quasi stationary plume. As expected, time variations of the wind field, that is north-east at the beginning and north at the end, induces deformation of the plume.

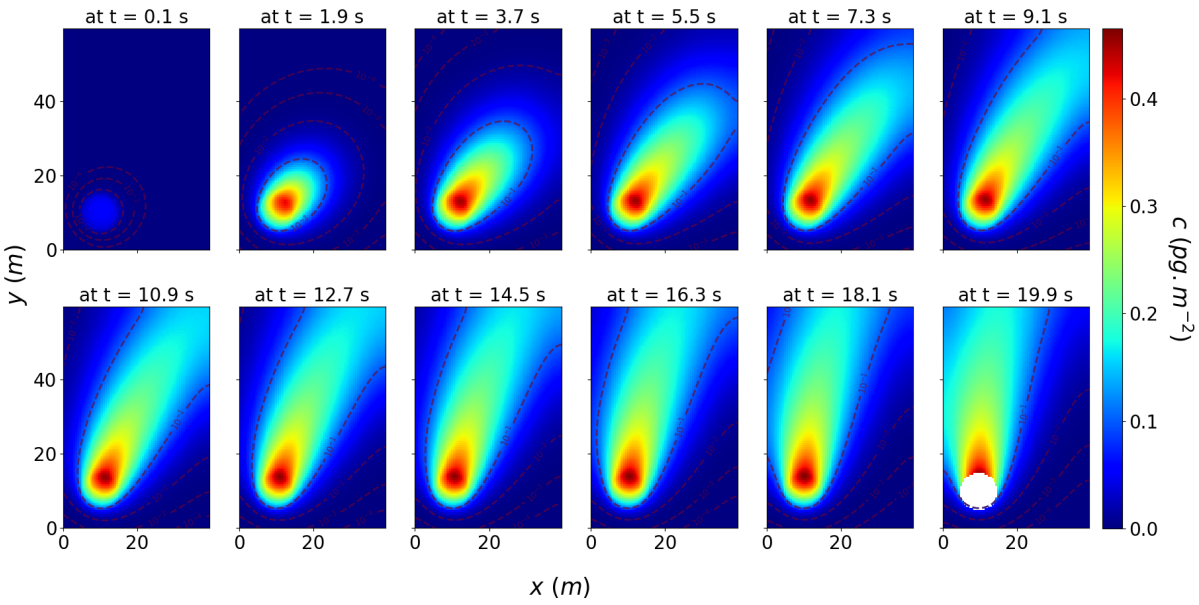


Figure 5 – Map of the concentration in pheromone $c(s_t)$ $(pg.m^{-2})$ computed from the depth-integrated CTM model (1) with the reference source term s_t for the Toy case presented in Sec. 2.3.2 at several times; in addition iso-lines of c are indicated with black dashed lines, and the reference source s_t support is displayed in white at final time.

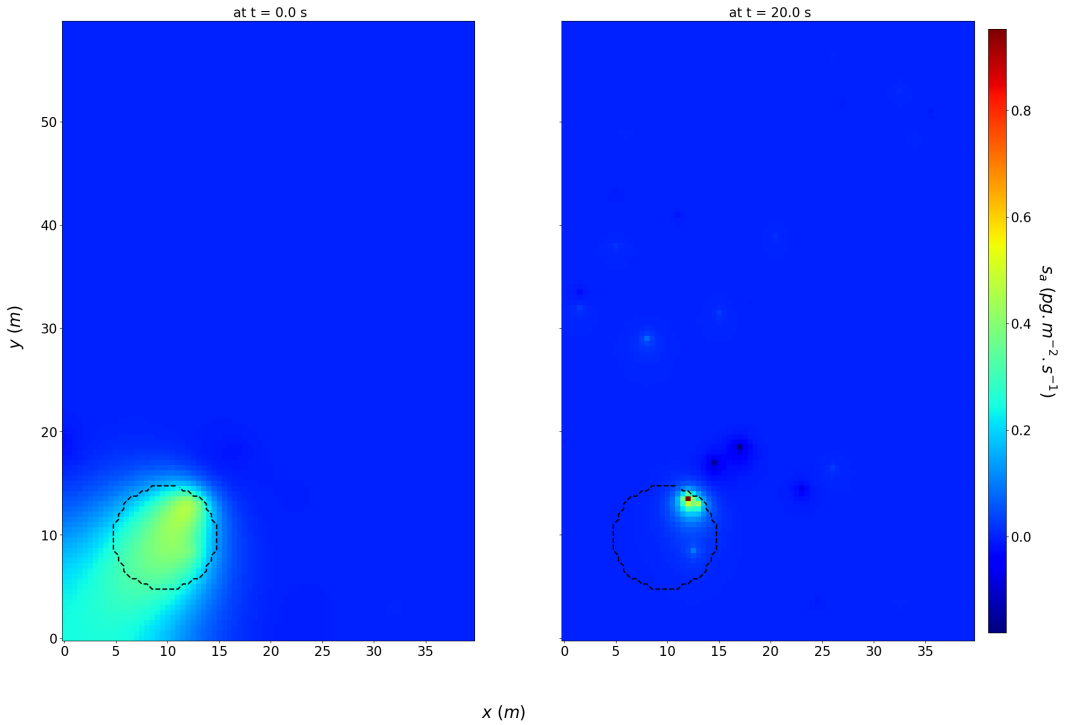


Figure 6 – **Classical VDA without regularization for the Toy case.** The optimal quantity of pheromone emitted s_a is obtained by inference without regularization term for the Toy case at $T = 0s$ (left panel) and $T = 20s$ (right panel). The support of the target quantity of emitted pheromone s_t (see Fig. 2) is delimited by the dashed lines. Some spurious detection spots can be seen, as sources (light dot) or sinks (dark blue dots) in the Toy case at $T = 20s$.

3.3. Classical VDA without regularization used as reference in the Toy case

We now illustrate the inference of emission sources in the Toy case without regularization term in Fig. 6. We can observe that the optimum s_a exhibits positive values, corresponding to pheromone emission sources, in areas spreading out upwind of the sensor locations (see Fig. 2 for sensor locations. Some sensor locations are also visible in Fig. 6, indicated by spots). This detection is referred to as the retro-plume. Indeed, the optimum s_a is computed by descending the gradient of the cost function (7), which is evaluated using the adjoint model (6). This equation describes the retro-propagation with diffusion of the discrepancy between predictions and observations $m(c(s)) - m^{obs}$ from final time $t = T$ to initial time $t = 0$. Consequently, the gradient descent algorithm mostly adjusts the initial prior s_0 in diffuse areas (attributable to the diffusion term in the adjoint model) upwind (attributable to the advection term) of the sensor locations (attributable to the source term).

Another consequence of the retro-propagation feature can be seen in Fig. 6: if at initial time, the estimated source s_a covers a large area, at final time $t = 20s$, the retro-plume is mostly confined into a small vicinity of several active sensors. Indeed, near final time, as the retro-plume is only advected for a little time interval, the propagation of the adjoint state is mostly due to diffusion.

We also note that the reconstructed source s_a exhibits negative non-physical values, that can be identified by dark blue dots near some sensors (see e.g. sensor locations $X = (17, 18.5)$ and $X = (14.5, 17.0)$ at time $t = 20s$, in Fig. 6, right panel). This phenomena is not due to numerical errors: it is a feature that directly comes from the noise in the data. The VDA intends to reproduce the data, including the noise even when the noise is not consistent with physical mechanisms. Therefore, during the retro-propagation of the noisy data, the optimal source s_a includes pheromone sinks, i.e. negative values, in order to approximate the noisy target, even if it is not physical.

3.4. Benchmark of the different regularization terms for BI-DA in the Toy case

We now aim to benchmark the different regularization terms introduced in Sec. 3.1 for BI-DA. We use the same supervised framework as in Sec. 3.3: a target source term s_t is used to produced the synthetic data, and are compared to the estimated source s_a. Recall that $s_t = qp_{ref}^{toy}$ with q the constant emission rate, and p_{ref}^{toy} the pest insect density, which remains constant over time on a circular support in the Toy case, see Sec. 2.3.2. This lead to the following population dynamics model:

$$\partial_t p_{ref}^{toy}(t, x, y) = 0 \tag{29}$$

which corresponds to the population dynamic-informed regularization term

$$j_{reg,PD}^{toy}(s) = \|\partial_t s\|_{L^2}^2. \tag{30}$$

We also recall that a favorable zone for the emitting insect is defined in a square Ω_{square} that includes the support p_{ref}^{toy}, see Fig. 2. Moreover, from the typical density of insects and the *per capita* pheromone emission rate given by the literature, we compute a surface emission rate $\bar{s}_b^{toy} = 0.54 pg.s^{-1}m^{-2}$ that is used as biology-informed background estimate

$$s_b^{toy}(t, x, y) = \bar{s}_b^{toy} \mathbb{1}_{\Omega_{square}}(x, y),$$

that will be taken into consideration through the Tikhonov regularization term (23), see Sec. 3.1.1.

We include in the benchmark the Tikhonov (T), LASSO, Population Dynamic (PD) regularization, and all combination of them, in order to understand how each regularization term operates and impacts the BI-DA. The optimum s_a is then compared with the target source s_t using the different criteria introduced in Sec. 2.4, i.e. total and support RMSE criteria (see resp. Eqs. (14) and (15)), miss and false discovery rates (see resp. Eqs. (17) and (18)), and the one-sensor adjoint state-based criteria (see Eq. (21)).

3.4.1. Visual comparison of estimated pheromone emissions s_a. The estimated sources s_a are displayed for each regularization term, including composite terms, in Fig. 7 at initial ($t = 0s$) and final times ($t = 20s$), for one single (top panel) and composite (bottom panel) regularization terms (see Sec. 3.1), together with the case without regularization ('no reg') taken as reference inference experiment.

When only a single regularization term is considered ($'T'$, $'PD'$, $'LASSO'$ on top panel of Fig. 7), the estimate s_a covers the main part of the target source s_t at $t = 0.0s$, as well as when no regularization is added ('no reg'). At final time, when fewer data are available for estimation, single regularization terms do not significantly improve, or even worsen, the estimate of s_a made in the reference case ('no reg'), except with the population dynamic-informed regularization term ($'PD'$) that pulls the estimation towards constant solutions. As expected, the Tikhonov regularization ($'T'$ on top panel of Fig.7) tends to redistribute the emissions estimate on the square zone favorable for pests, making it visible. This includes enhancing spurious sources that are located in the favorable square but out of the circular target, and diminishing the sources in the circular target to compensate the spurious sources. At final time, the redistribution of the sources on the favorable square does not lead to a significant improvement compared to the reference case ('no reg'). Let us note that the Tikhonov regularization term also diminishes the estimate out of the favorable square. For the LASSO regularization ($'LASSO'$ on top panel of Fig.7), the retroplume is less diffuse, near both the initial and final time, compared to the reference ('no reg') case since small values have been thresholded by the regularization, which also corresponds to an expected feature. With the population dynamic-informed regularization ($'PD'$ on top panel of Fig. 7), the emission estimate is slightly higher inside the target support compared to the other single regularization terms ($'T'$ and $'LASSO'$) and to the reference case ('no reg'). Moreover, near the final time, the estimate s_a covers the main part of the target source s_t.

When several regularization terms are considered (on bottom panel of Fig. 7), as long as they involve population dynamics regularization (namely $'PD + LASSO'$, $'PD + T'$ and $'all\ reg'$), the results are visually comparable to the one obtained with the population dynamic-informed regularization ($'PD'$). We can note that spurious sources are estimated far from the target source at final time with the combination of population dynamics and LASSO ($'PD + LASSO'$ on the bottom panel of Fig.7). When the combination of population dynamics and Tikhonov is used ($'PD + T'$ on the bottom panel of Fig.7), we can see that the population dynamics regularization largely counterbalance the Tikhonov regularization, since few visual differences can be noticed with the case with only population dynamics ($'PD'$). When population dynamics, Tikhonov and LASSO terms are combined ($'all\ reg'$ on the bottom panel of Fig.7), they correctly interact and slightly improve the results at both initial and final times compared to the case of population dynamics regularization alone ($'PD'$), even though spurious sources in the favorable zone can be seen due to the Tikhonov regularization. On the contrary, interactions between Tikhonov and LASSO regularizations ($'T + LASSO'$ on the bottom panel of Fig.7) does not seem beneficial, since they hardly improve the results obtained from the reference case ('no reg').

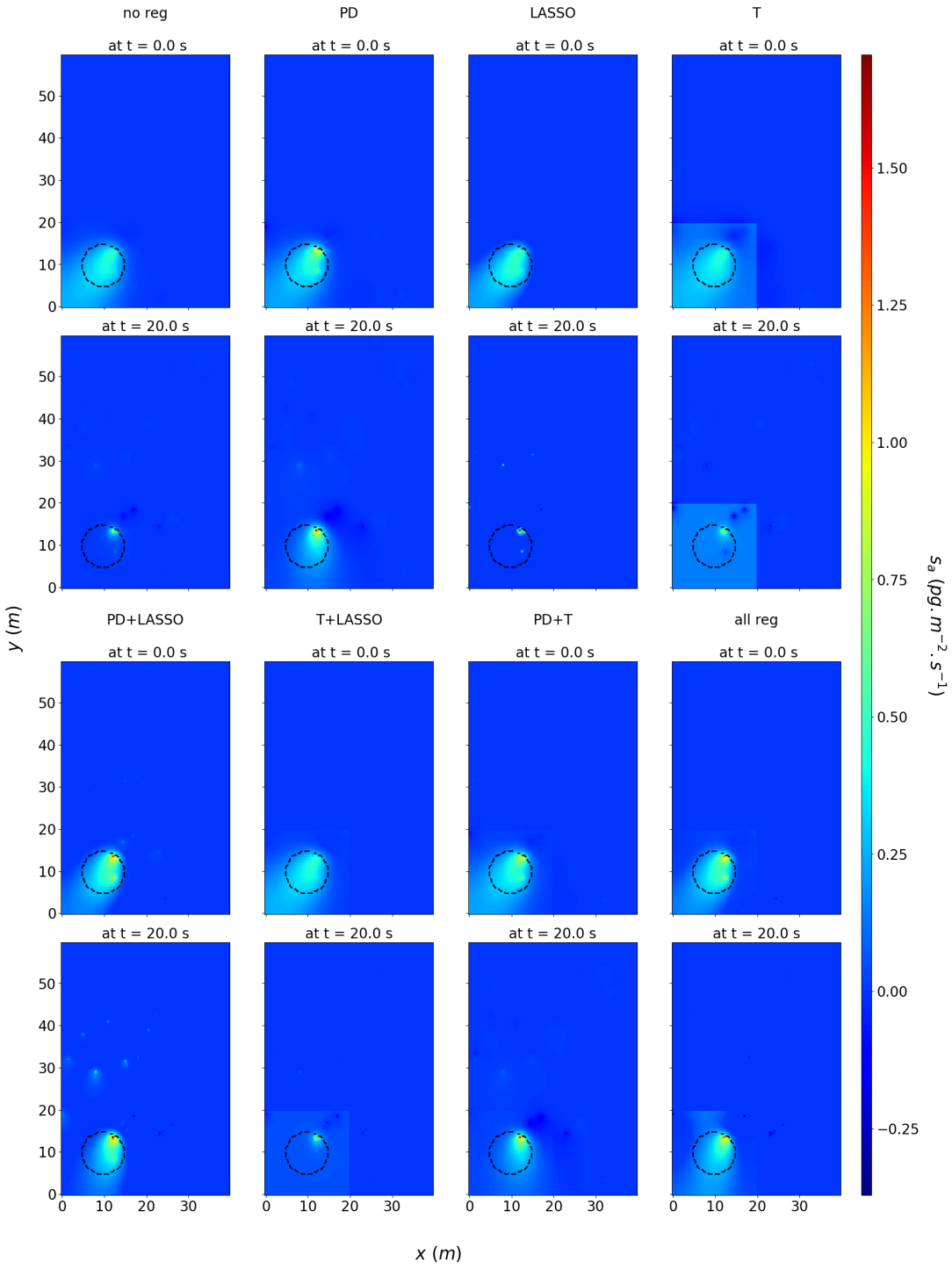


Figure 7 – Optimal sources estimated with different regularization terms. Optimal quantity of pheromone emitted s_a obtained by inference with several regularization strategies for the Toy case. Upper panel, from left to right: with no regularization term (no reg), with the population dynamic-informed regularization term (30) (PD), with the LASSO regularization term (24) (LASSO) and with the Tikhonov regularization term (T) (23). Lower panel, from left to right: with combination of population dynamic-informed and LASSO regularization terms (PD+LASSO), with a combination of Tikhonov and LASSO regularization terms (T+LASSO), with combination of population dynamic-informed and Tikhonov regularization terms (PD+T) and a combination of all three population dynamic-informed, LASSO and Tikhonov regularization terms (all reg). In both panels, the optimal sources are indicated at $t = 0.0s$ (upper row) and $t = 20.0s$ (lower row).

3.4.2. RMSE criteria. In order to have a more quantitative assessment of the discrepancies between s_a and s_t in the different cases (single and composite regularization terms), we introduced in Sec. 2.4.1 $RxMSE_{tot}$ (14), the error computed on the whole computational domain, and $RxMSE_{supp}$ (15), the error computed on the support of the target source only. The latter enables to assess the inference accuracy on the specific zone where the insects are located. We display in Fig. 8 the time evolution of $RxMSE_{tot}$ (top panel) and $RxMSE_{supp}$ (bottom panel) for the different (single and composite) regularization terms and compared to the reference case without regularization ('no reg').

We first note that all the regularization terms provide sensibly equal estimation error inside the target in a large part of the time domain, until $t \approx 17.5s$, see the bottom panel of Fig. 8. This indicates that the main differences come from reconstruction outside the target support and at the end of the time window, when data are lacking.

The Tikhonov regularization (Fig. 8, upper panel, green curve) gives higher $RxMSE_{tot}$ than the reference estimate (blue curve), except at the extreme end of the time domain. These discrepancies seems due to the additional emissions added on the favorable square. It is confirmed by the $RxMSE_{supp}$ criteria, which is sensibly the same with Tikhonov regularization (Fig. 8, lower panel, green curve) and without regularization (blue curve), except for the last time points where the additional emissions introduced by the Tikhonov regularization improves a little the estimate on the target support compared to the reference case.

The LASSO regularization (Fig. 8, upper panel, pink curve) allows a sensible reduction of $RxMSE_{tot}$ compared to the reference case (blue curve), all along the time domain except at the end. By the soft-thresholding feature it provides, the LASSO regularization thresholds the small values toward zero, specifically outside the target support. However, on the target support, the LASSO regularization (Fig. 8, lower panel, pink curve) does not significantly improve the $RxMSE_{supp}$ criteria compared to the reference case (blue curve). At the end of the time domain, the LASSO regularization keeps on thresholding, even within the target support where insects are emitting, thereby impairing the accuracy of the estimate.

Both outside and inside the target support, the population dynamics regularization (Fig. 8, both panels, orange curve) improves the estimation at the end of the time domain compared to the reference case (blue curve). This indicates that the population dynamics adds additional priors that counterbalance the lack of data from the sensor at the end of the time series. Specifically, it allows obtaining a solution that is consistent with known population dynamics, which enables the propagation of well-reconstructed pheromone emission to zones and times where the estimate of the pheromone emission is poor due to the location of the sensors and the intrinsic weaknesses of the retro-propagation process.

The use of several regularizations, in particular the combination of population dynamics and Tikhonov regularizations (Fig. 8, brown curves) and the combination of all three regularizations (Fig. 8, red curve), take benefit of the enhancement provided by the LASSO regularization outside the target support and by the population dynamics regularization at the end of the time window to improve the prediction accuracy all along the time interval.

3.4.3. Presence prediction, false discovery and miss rates. In a practical view point, the important information for cultural practice is not the distribution of pheromone emission, but the presence of insect pests that we aim to control. Therefore, we introduced in Sec. 2.4.1 a prediction of the insects presence derived from the pheromone emission map and based on the super level set of Eq. (16) of minimal emission value, see Fig. 4.

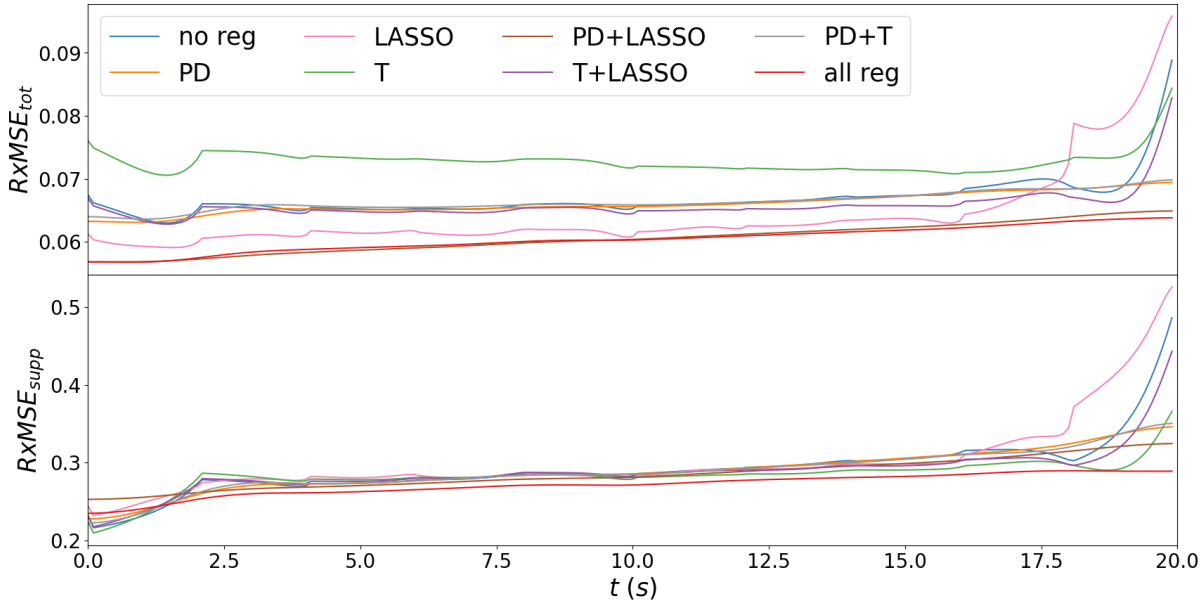


Figure 8 – Evolution of total and support RMSE of s_a against s_t. Time evolution of the root space-mean square errors between the optimum s_a obtained by inference with several regularization strategies and the target s_t, on the whole domain ($RxMSE_{tot}$, Eq. (14), upper panel) and on the target support ($RxMSE_{supp}$, Eq. (15), lower panel) in the Toy case: with no regularization term (no reg, blue), with the population dynamic-informed regularization term (30) (PD, orange), with the LASSO regularization term (24) (LASSO, pink), with the group LASSO regularization term (25) (group LASSO, yellow), with the Tikhonov regularization term (23) (T, green), with combination of population dynamic-informed and LASSO regularization terms (PD+LASSO, brown), with a combination of Tikhonov and LASSO regularization terms (T+LASSO, purple), with combination of population dynamic-informed and Tikhonov regularization terms (PD+T, grey) and a combination of all three population dynamic-informed, LASSO and Tikhonov regularization terms (all reg, red).

To get a better view of the amount of insects that are detected and missed during the time window, we use false discovery rate ρ_{fp} (18), i.e. the proportion of area where the insects are predicted to be present but are actually absent, and the miss rate ρ_{fa} (17), i.e. the proportion of area where the insects are actually present but the prediction is missed. We recall that the objective here is to reduce the false discovery rate, equivalent in our context to a useless application of pesticide in the corresponding zone, while keeping the miss rate as small as possible, or even null, to control the infestation.

We represent in the upper panel of Fig. 9 the presence maps of insects computed with different regularization terms and given by the target s_t, at three different times ($t = 0.0s$, $t = 10.0s$ and $t = 20.0s$). Pests are present inside the delineated color-coded zones. As already anticipated in the visual comparison (see Sec. 3.4.1), the Tikhonov regularization overestimates the presence of pests, while the LASSO regularization, combination of LASSO and Tikhonov regularization ($'T + LASSO'$) and the reference case ('no reg') strongly underestimate pest presence at final time.

Cases involving population dynamics regularization ($'PD'$, $'PD + LASSO'$, $'PD + T'$, 'all reg') keep covering the target zone all along the time domain. The presence predicted with $j_{reg,PD}$ at final time benefits from the propagation by the population dynamics regularization of the presence predicted at initial time. We can however notice a slight overestimation at final time with the case combining all three regularizations ('all reg'), probably due to the Tikhonov term

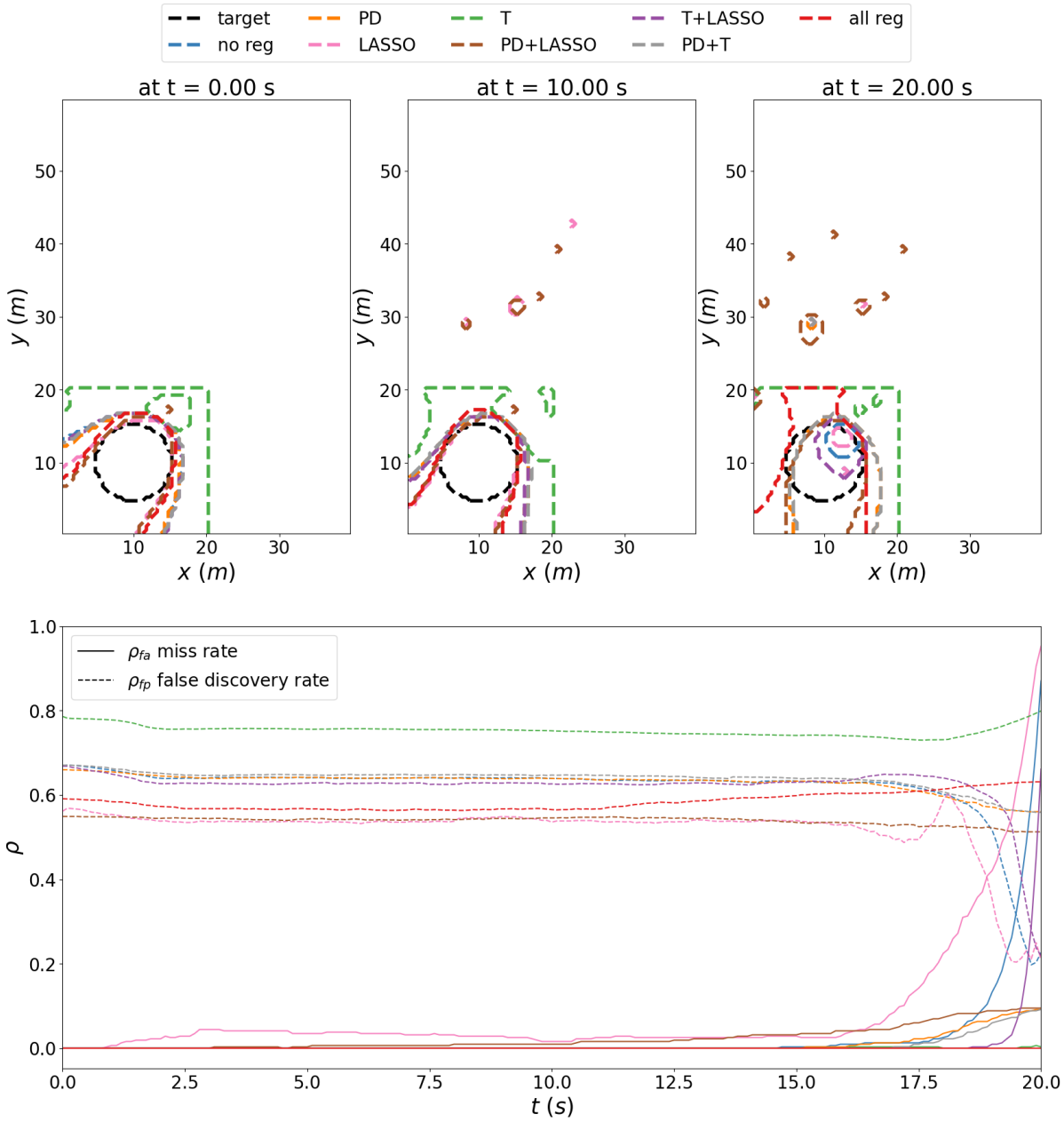


Figure 9 – Presence prediction and time evolution of false discovery and miss rates. Prediction of insects presence with several regularization strategies: with no regularization term (no reg, blue), with the population dynamic-informed regularization term (30) (PD, orange), with the LASSO regularization term (24) (LASSO, pink), with the group LASSO regularization term (25) (group LASSO, yellow), with the Tikhonov regularization term (T, green), with combination of population dynamic-informed and LASSO regularization terms (PD+LASSO, brown), with a combination of Tikhonov and LASSO regularization terms (T+LASSO, purple), with combination of population dynamic-informed and Tikhonov regularization terms (PD+T, grey) and a combination of all three population dynamic-informed, LASSO and Tikhonov regularization terms (all reg, red). Top panel: map of presence prediction at three different time. The presence is obtained from the reconstructed source s_a by computing a level set with Eq. (16). The true presence (i.e. the target support) is indicated in black.
Lower panel. Time evolution of the miss rate ρ_{fa} (solid lines), see Eq. (17), and false discovery rate ρ_{fp} (dashed lines), Eq. (18).

as the boundary of the square favorable zone are slightly visible (Fig. 9, upper panel, $T = 20.0s$, red line). We also observe spurious presence predictions with the single population dynamics regularization, the single LASSO regularization and the combination of both (Fig. 9, upper panel, resp. $'PD'$, $'LASSO'$ and $'PD + LASSO'$) far from the target, likely because of the noise.

Miss rates ρ_{fa} are comparable for all regularization at the beginning of the simulation, the differences being more visible in the false discovery rates ρ_{fp}. In all cases, both ρ_{fp} and ρ_{fa} are quite stable throughout the time domain until $t \approx 17.5s$, time at which ρ_{fa} without regularization (reference case 'no reg') strongly increases (Fig. 9, lower panel, blue curves) and where discrepancies occur. As expected, with the Tikhonov regularization (Fig. 9, lower panel, green curves), ρ_{fa} stays close to zero, but the ρ_{fp} remains high. On the contrary, with the LASSO regularization (Fig. 9, lower panel, pink curves), ρ_{fp} is the lowest but ρ_{fa} is among the highest and strongly increases after $t \approx 15s$, which is earlier compared to the reference case (blue curve). With the combination of Tikhonov and LASSO regularizations (Fig. 9, lower panel, purple curves), ρ_{fa} also strongly increases but after $t \approx 19s$, which is later compared to the reference case (blue curve). As soon as population dynamics regularization is involved (Fig. 9, lower panel, orange, brown, grey and red curves), ρ_{fa} is rather low and kept under the one obtained in the reference case ('no reg', blue curves) especially at final time, except with the combination of population dynamics and LASSO regularizations (brown curves) that misses some part of the true presence of insects. In particular, with the combination of all three regularizations ('all reg', red curves), the miss and false discovery rates are very stable, ρ_{fa} remains zero on the whole time window while ρ_{fp} is the third smallest one on the main part of the time window. With this combination, the three terms interact in an efficient way: the LASSO term seems to keep ρ_{fp} small, the Tikhonov term reduces the ρ_{fa} while the population dynamics term smooths the rates evolution at the end of the time domain, when the data are scarce.

3.4.4. Estimating active sensors with the one-sensor adjoint state criteria. The visual comparison of optimal pheromone emission s_a (see Sec. 3.4.1) indicates that sensors near the emission source seem to have greater impact on the retro-plume than sensors located further away. In order to better quantify this feature, we introduced in Sec. 2.4.3 the maximal total contribution $\|c^*_{i,tot}\|_\infty$ of the i^{th} sensor, which is the total contribution of the data obtained with the i^{th} given sensor to the decrease of the cost function along the optimization algorithm. This metric enables to understand how a specific sensor impact the estimation of s_a, and to rank the sensors accordingly for each regularizations, see Fig. 10. It allows in particular to understand if a regularization is able to supplement or amplify the information collected by the sensors, enabling to use less sensors to achieve a given level of accuracy in the estimate of s_a. Note that the sensors are not involved in the regularization terms, so that the impact of the BI-DA regularization terms on the total contribution $\|c^*_{i,tot}\|_\infty$ is indirect: the regularization changes the trajectory taken by the optimization algorithm in the control space, hence modifying the points where the one-sensor adjoint state c^*_i is evaluated along the optimization path.

We can observe that in the reference case (Fig. 10, 'no reg' and blue line), the 4 first sensors are particularly impactful. Then, a plateau value is attained until the 31st sensor, where a drop is observed , indicating that the last sensors could possibly be removed without impairing the estimate of s_a. The LASSO regularization (Fig. 10, pink line) tends to amplify the contribution of all the sensors and mitigates the drops of contribution at the beginning and the end of the contribution order. Indeed, the LASSO regularization tends to threshold the small value of s and the contribution of the sensors are amplified to compensate this thresholding. With Tikhonov and

population dynamics regularizations (Fig. 10, green and orange lines), the decrease in the contribution is greater for the ten first sensors compared to the reference case (blue line), which shows that the first sensors in the contribution order concentrate most of the information. Moreover, the maximal total contribution is almost always lower than the one obtained with the reference case (blue line). One can conclude that these regularizations bring information that supplement the signal carried by the sensors data.

In composite regularizations, while combining Tikhonov and LASSO or population dynamics and Tikhonov regularizations (Fig. 10, purple and grey lines) give a similar pattern than the reference case (blue line), combining population dynamics and LASSO regularizations (Fig. 10, brown line) provides results close to the single LASSO one (pink line). The combination of all three regularizations (Fig. 10, red line) globally enhances the overall contribution of each sensor. Moreover, the drops of contribution at the beginning and the end of the ranking are also significantly mitigated compared to any other composite regularization. Overall, all the sensors have very homogeneous contributions: the trend of contribution is similar to that with the LASSO regularization alone, but amplified both in terms of values and variations.

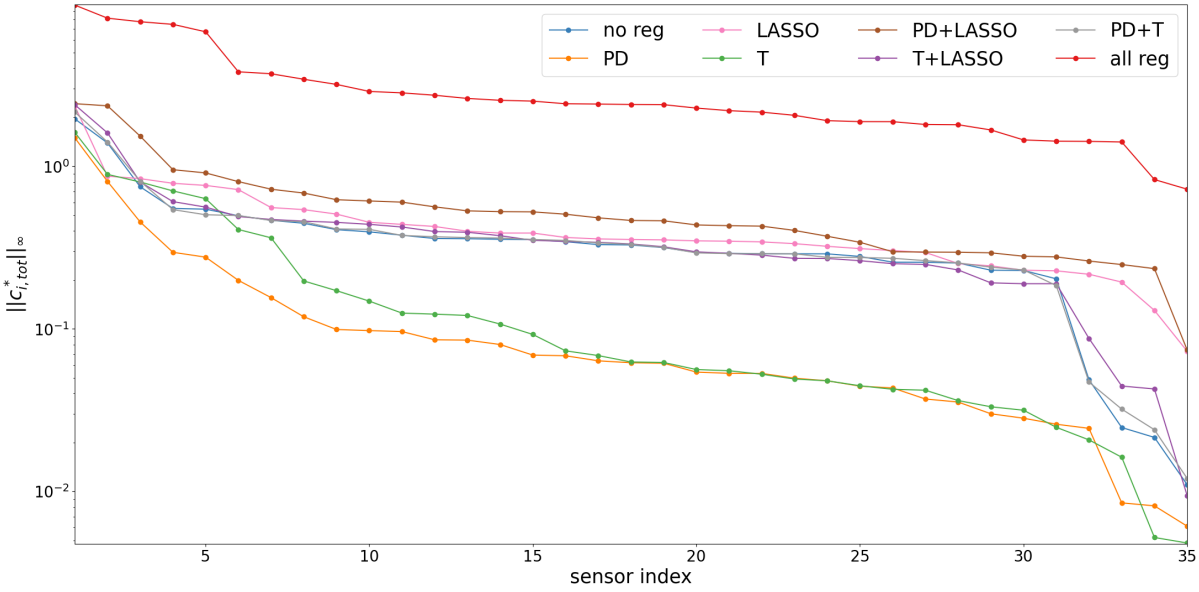


Figure 10 – Ranking sensor contribution in BI-DA. Maximal total contribution $\|c^*_{i,tot}\|_\infty$ vs sensor rank when sorted by contribution order for several regularization strategies in the Toy case: with no regularization (no reg, blue), population dynamic-informed (PD, orange), LASSO (LASSO, pink), Tikhonov (T, green), combination of population dynamic-informed and LASSO (PD+LASSO, brown), combination of Tikhonov and LASSO(T+LASSO, purple), combination of population dynamic-informed and Tikhonov(PD+T, grey) and a combination of all three population dynamic-informed, LASSO and Tikhonov regularization terms (all reg, red). The quantity $\|c^*_{i,tot}\|_\infty$ indicates the maximal (with respect to time and space) total contribution of sensor i all along the optimization path.

3.5. Numerical illustration of pheromone propagation in the FAW case

The direct model (1) is now computed on the FAW case, as described in Sec. 2.3.2. The FAW case strongly differs from the Toy case as it relies on a realistic landscape, with real application time and space scales, taking into account a pest, the FAW, harbouring a characteristic emission pattern, see Sec. 2.3.2 for details. Namely, only the females emit pheromones to attract male, but stop emitting at the start of mating. The emission model hence describes the population dynamics of female and male pairing.

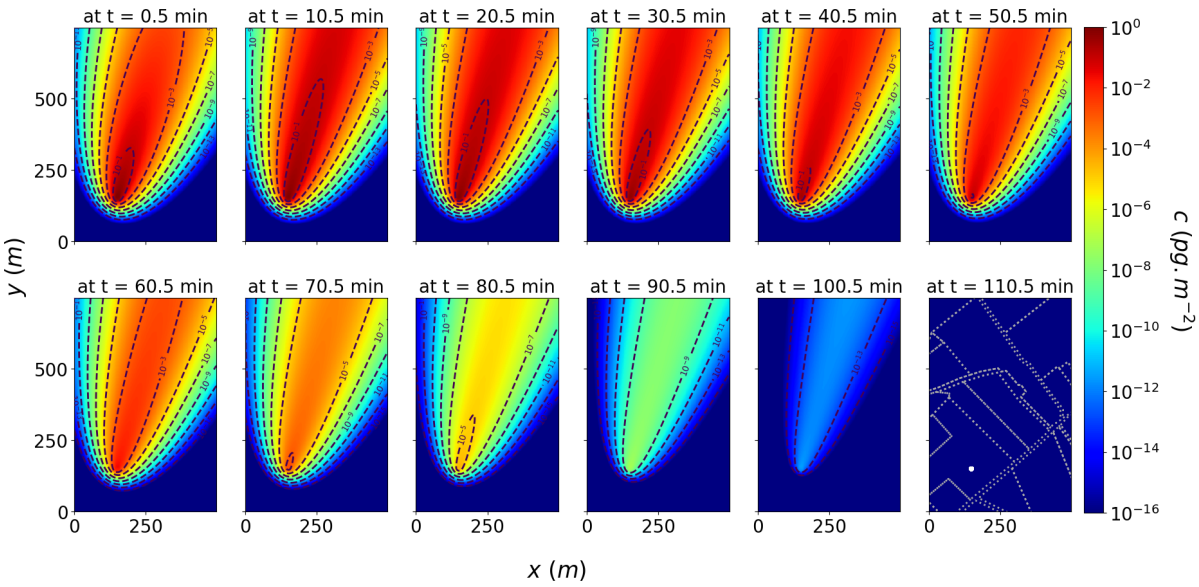


Figure 11 – Map of the concentration in pheromone $c(s_t)$ $(pg.m^{-2})$ computed from the depth-integrated CTM model (1) with the reference source term s_t for the FAW case presented in Sec. 2.3.2, in logarithmic scale and at several times. In addition iso-lines of c are indicated with black dashed lines. Furthermore,the agricultural parcels are displayed with grey dashed lines together with the reference source s_t support in white at $T = 110.5min$ (lower right panel).

Direct model outputs are displayed in Fig. 11. As in the Toy case, convection is also largely predominant in the FAW case: the pheromones are also transported along the wind field, with little additional diffusive effect. Note that the wind field is taken from real weather broadcasting data. The pheromones spread across the domain within a couple of minutes until reaching a quasi-steady state plume within about ten minutes. After that, the decrease in time of s which reflects the emission pattern of FAW that stops releasing pheromones at the start of mating (see Fig. 3c and Sec. 2.3.2), implies that c decrease everywhere until $t \approx 110min$. Finally, s becomes negligible and c does not variate significantly any longer. The landscape composition (see the grey dashed lines on the bottom right panel of Fig. 11), which is taken into account through the loss rate τ_{loss}, does not seem to influence significantly c.

3.6. Classical VDA and sensor positioning improvement in the FAW case

A first classical VDA without any regularization term is first conducted in the FAW case with sensors randomly dispatched in the field of interest, see left panel in Fig. 12. We highlight that this estimation problem is by far more difficult to solve than the Toy case for several reasons: first, there is a change in the spatio-temporal scales by one order of magnitude (from tens of meters and seconds in the Toy case to hundreds of meters and hours in the FAW case); next, the pheromone emission pattern mimics the biological behavior of the FAW and is not stationary in the FAW case unlike in the Toy case; finally, the sensor density is much higher in the Toy case ($\sim 10^{-2}$ sensor/m^2) than in the FAW case ($\sim 10^{-4}$ sensor/m^2). In this context, classical VDA fails to detect the position of the emitting insects (left panel in Fig. 12). The VDA algorithm does not sufficiently retro-propagate the discrepancies between data and predictions: during the optimization, the VDA could be stuck in a flat zone of the objective function due to insufficient signal of the sensors.

To partially counterbalance this effect, we repositioned several sensors by leveraging the one-sensor adjoint state-based criteria (see Sec. 2.4.3) for the classical VDA. Namely, we computed

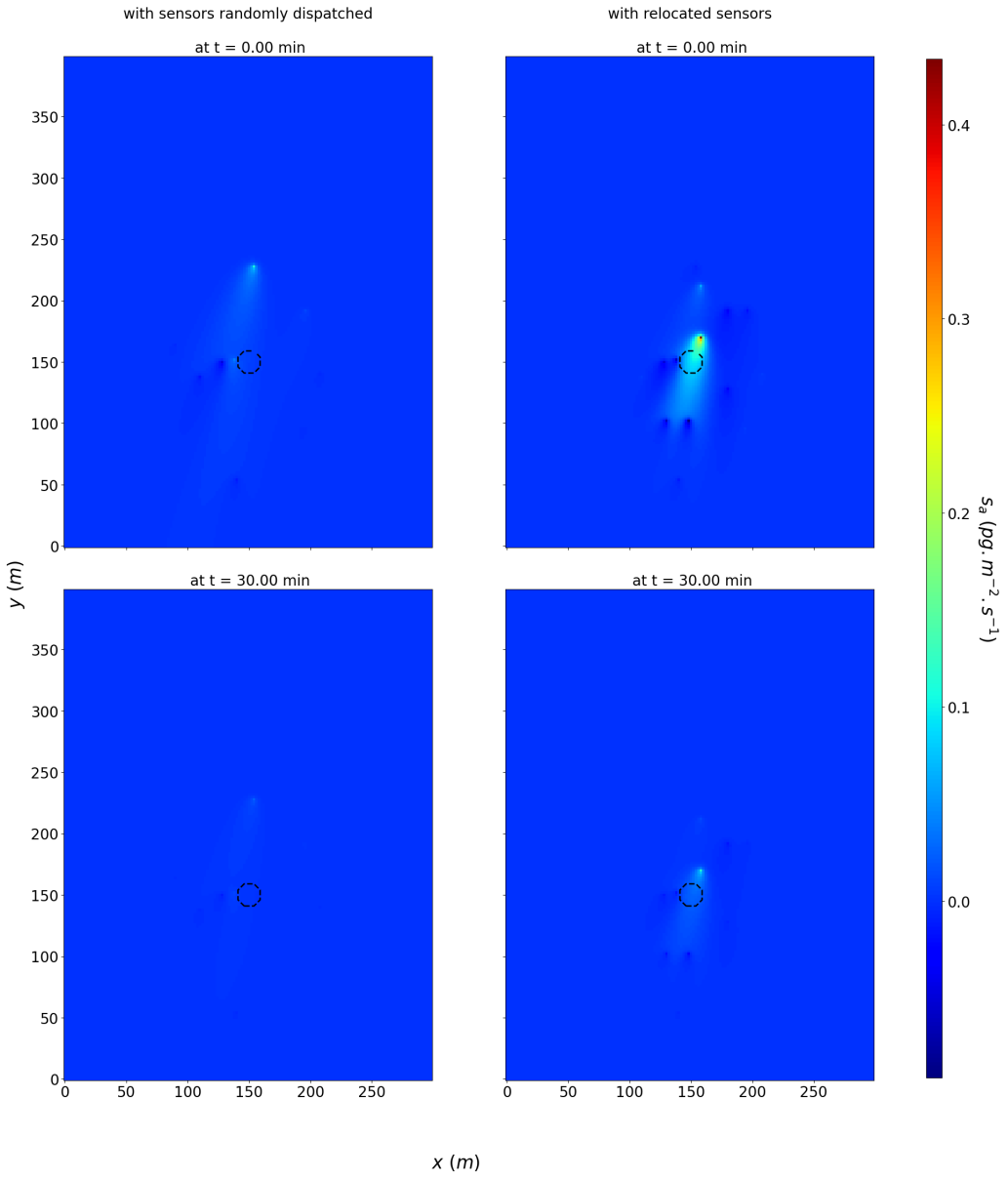


Figure 12 – VDA without regularization for the FAW case. The optimal quantity of pheromones emitted s_a is obtained by inference without regularization term for the FAW case at $T = 0min$ (upper panels) and $T = 30min$ (lower panels). The support of the target quantity of emitted pheromone s_t (see Fig. 3b) is delimited by the dashed lines. The left panel represents the source reconstruction with the initial random sensor distribution (uniform law over the maize field), whereas the right panel represents the same estimation after sensor re-positioning (see Sec. 3.6)

$c^*_{i,tot}$ in eq. (21) resulting from this first VDA without regularization, and selected the $2/5$ of the sensors with the lowest contribution to the inference of s. These sensors were randomly repositioned within the convex envelop of the $3/5$ sensors that were kept fixed, with uniform distribution (see top right panel of Fig. 13). We highlight the fact that no knowledge of the actual position of the target pheromone emission (i.e. of the emitting insects) has been used during repositioning, but only a criteria derived from the signal received by the sensors: in that sense, the repositioning method is unsupervised. On the top right panel of Fig. 13, we can observe that the sensors that have been repositioned were those located far or upwind from the targeted emitting insects, indicating that the criteria $c^*_{i,tot}$ is efficient in selecting sensors that would have

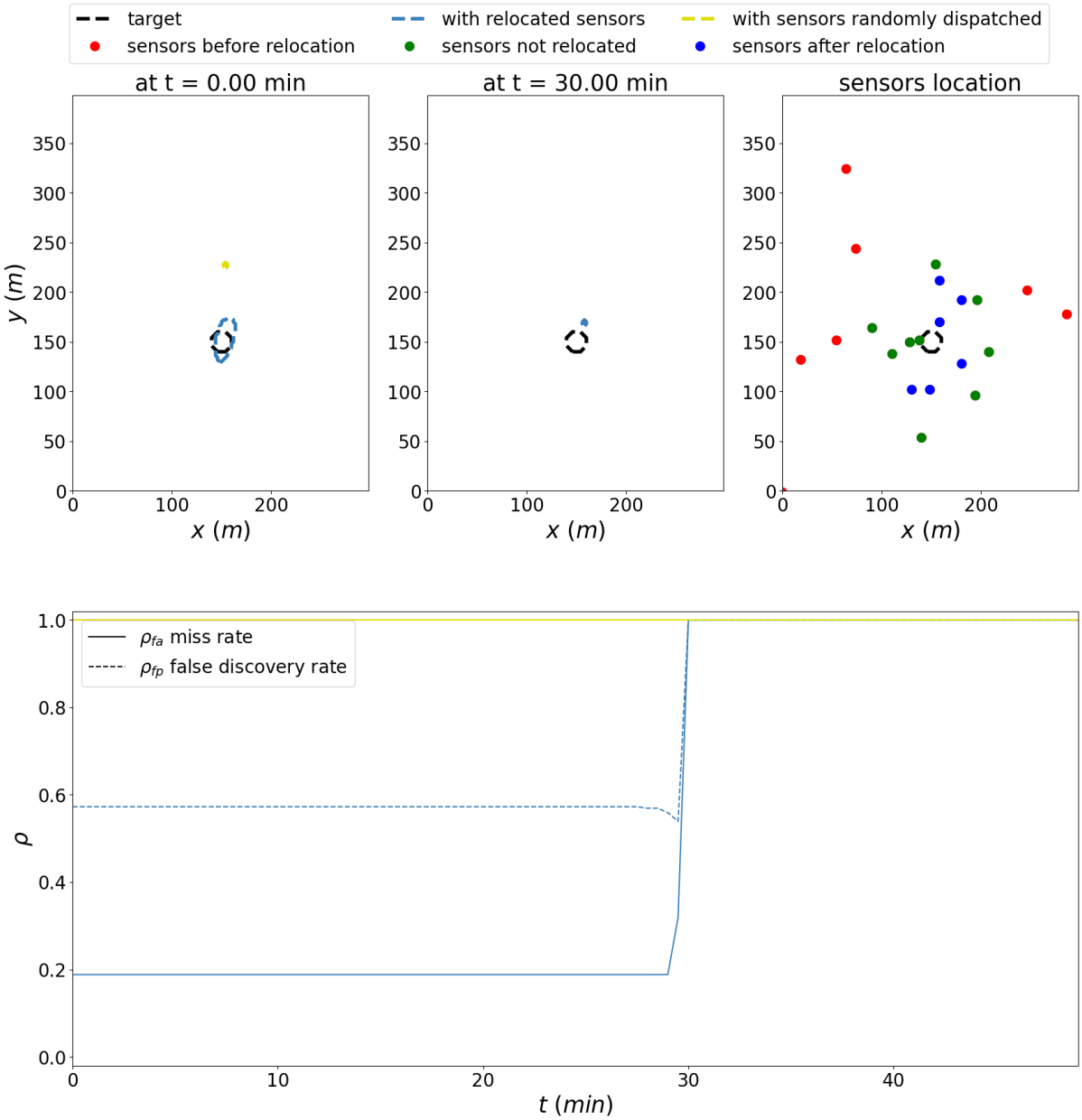


Figure 13 – Presence prediction and time evolution of false discovery and miss rates in the FAW case before and after sensor repositioning. The prediction of insects presence is obtained by classical VDA without regularization before and after sensors repositioning. Before repositioning, the sensors have been randomly dispatched in the maize field. After computation of $c^*_{i,tot}$, the two-fifth of the sensors with lower contribution to the source reconstruction were randomly re-allocated in the convex envelop of the three-fifth of the sensors with higher contribution. *Upper panel*: presence predictions at $T = 0\text{min}$ and $T = 30\text{min}$ before (green line) and after (blue line) sensor repositioning. The position of the sensors before and after relocation is indicated in the right figure. The presence is obtained from the reconstructed source s_a by computing a level set with Eq. (16) (see Sec. 2.4.2). The true presence (i.e. the target support) is indicated in black. *Lower panel*. Time evolution of the miss rate ρ_{fa} (solid lines), see Eq. (17), and false discovery rate ρ_{fp} (dashed lines), in eq. (18).

been *a priori* identified as potentially uninformative in a supervised context where the target is known.

After sensor repositioning, the reconstructed s_a by classical VDA now overlaps the target position at $T = 0\text{min}$, see right panel of Fig. 12. Furthermore, the VDA without regularization now enables to detect the insects in most of the target s_t support at initial time (blue curves on top left panel of Fig. 13), unlike before relocation where no insect were correctly located (green curves). These observations were reflected in the miss and false discovery curves, that were equal to 1 before repositioning, and dropped to about 0.2 and 0.6 respectively after sensor displacement (13 lower panel).

We note that, at $T = 30\text{min}$, insects presence prediction vanishes, and the miss and false discovery rates go up to 1, indicating inoperative presence prediction. This feature is completely different than the drop of presence prediction accuracy observed at final time in the Toy case (see Sec. 3.4 and Fig. 9), that was inherent to boundary effects near final time and due to the retro-propagation feature of the adjoint model. Here, the collapse of discovery accuracy comes from the emission pattern of the FAW, rather than from a limitation of the mathematical method. Actually, as FAW females stop emitting the sexual pheromone at the beginning of the mating, the quantity of pheromone measured decreases, impairing insect prediction although the insects are still present, and, what's more, fertilized. For a given level of pheromone, the VDA, and more specifically the adjoint model, retro-propagates the quantity of pheromone measured by the sensor: when the signal is too low, the retro-plume does not reach the target, and the insects are predicted outside the emission zone, see the results at $t = 30min$ on the lower right panel of Fig. 12 and on the upper middle panel of Fig. 13.

3.7. Pheromone emission inference and pest localization with BI-DA in the FAW case

We now test the BI-DA framework in the FAW case. We first characterize the biology-informed population dynamics regularization term, that must reflect the emission pattern of the FAW. Let us note $p_{ref}^{FAW}(t,x,y)$ the proportion of females that have not mated yet, and therefore keep emitting pheromones. The dynamics of the p_{ref}^{FAW} population is the following:

$$\partial_t p_{ref}^{FAW}(t,x,y) = -\gamma(t) p_{ref}^{FAW}(t,x,y) \tag{31}$$

with γ the mating rate at a given time that is represented by the logistic function (13). Let us note that p_{ref}^{FAW} defined in eq. (11) is an analytic solution of the population dynamics model (31). Thus, in the FAW case, the following population dynamic-informed regularization term will be considered:

$$j_{reg,PD}^{FAW}(s) = \|\partial_t s + \gamma s\|_{L^2}^2 \tag{32}$$

As mentioned in Sec. 2.3.2, the FAW females usually emit pheromone on plants favorable for the larvae, which in this case is assumed to be maize. However, maize covers most of the domain considered and so, it does not provide relevant information. From the biological knowledge of the insects, one can deduce an expected density of insects and a quantity of pheromone emitted per female insect, and thus a local expected value of pheromone emission $\bar{s}_{b,loc}^{FAW}$. Without any additional information, this value will be used on the whole domain at the initial time. This leads to the following Tikhonov regularization:

$$s_b^{FAW}(x,y) = \bar{s}_{b,loc}^{FAW} \quad \text{so that} \quad j_{reg,T}^{FAW}(s) = \|s(t=0) - s_b^{FAW}\|_{L^2(\Omega)}^2. \tag{33}$$

Finally, we note that the emission pattern of FAW is ought to be particularly impactful on the classical LASSO term. The soft-threshold will preponderantly mitigate the source reconstruction

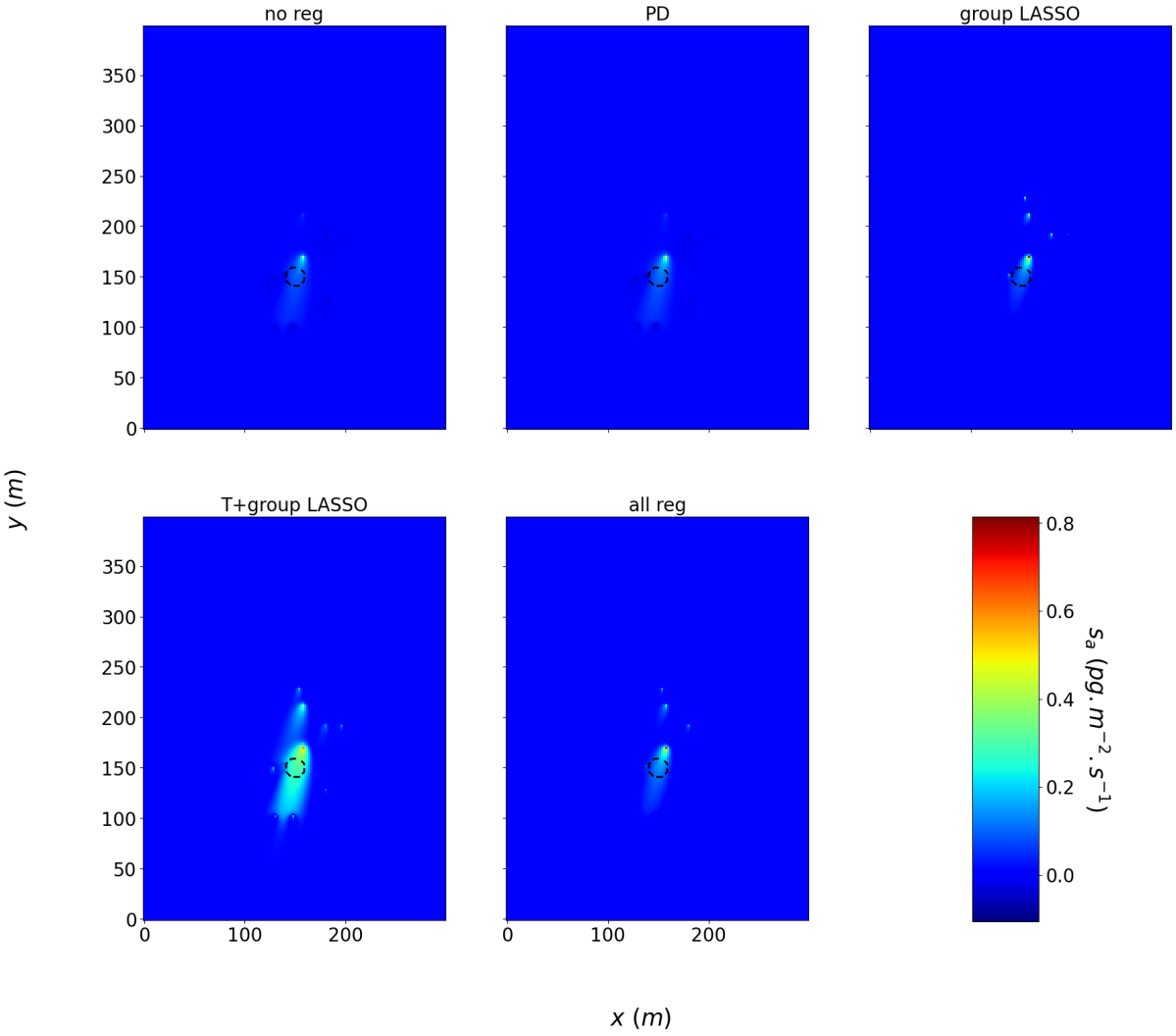


Figure 14 – Optimal quantity of pheromone emitted s_a obtained by inference with several regularization strategies for the FAW case. Upper panel, from left to right: no regularization term (no reg), population dynamic-informed (PD), group-LASSO. Lower panel, from left to right: combination of Tikhonov and group-LASSO (T+group LASSO), combination of all three population dynamic-informed, group-LASSO and Tikhonov regularization terms (all reg).

when the emission is low, accelerating miss rate at the end of the simulation when the number of emitting females decreases. Conversely, we can expect that the LASSO penalty will not filter out spurious reconstructions at the beginning of the simulation when levels of emitted pheromones are still high since these spurious zones could represent higher quantities than the shrinking threshold. To counterbalance these effects, we decided to switch the LASSO penalty for a group-LASSO penalty by filtering out spatial areas that received a total amount of pheromones, during the whole time domain, under the shrinking threshold (see eq. (25)), so that

$$j^{FAW}_{reg,gLASSO}(s) = \int_\Omega \left(\int_0^T s(t,x,y)^2 \right)^{1/2} dxdy.$$

A short comparison of LASSO and group-LASSO on the Toy case is made in the supplementary Sec. B.4.

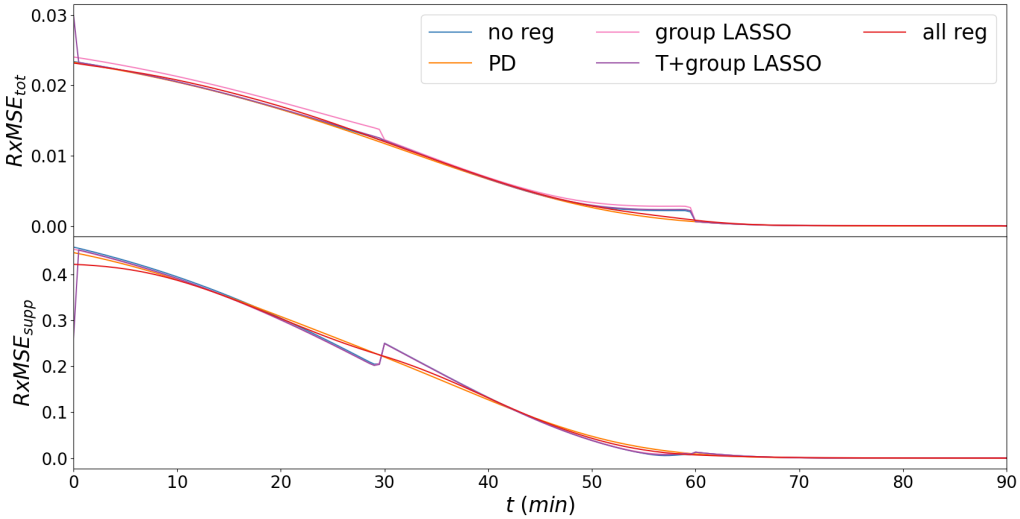


Figure 15 – Root space-mean square errors between the optimum s_a and s_t for several regularization strategies. We display the $RxMSE_{tot}$ introduced in Eq. (14) (upper panel) and the $RxMSE_{supp}$ introduced in Eq. (15) (lower panel), versus t for the FAW case: with no regularization term (no reg, blue), with the population dynamic-informed regularization term (32) (PD, orange), with the group-LASSO regularization term (25) (gLASSO, pink), with the Tikhonov regularization term and group-LASSO (T+group LASSO, brown) and with a combination of all three population dynamic-informed, group-LASSO and Tikhonov regularization terms (all reg, red).

3.7.1. Visual comparison of estimated pheromone emissions s_a. We display the map of the reconstructed pheromone emission s_a obtained by BI-DA in fig. 14 for different regularization terms, namely population dynamic-informed ('*PD*'), group-LASSO, a combination of Tikhonov and group-LASSO ('$T+groupLASSO$'), and a combination of all three population dynamic-informed, group-LASSO and Tikhonov regularization terms ('all reg'), that are compared with the reference case without regularization term ('no reg').

We can observe that s_a obtained with the population dynamics regularization (top middle panel on Fig. 14) slightly enhances the one obtained in the reference case. The group LASSO regularization (top right panel on Fig. 14) thresholds pheromone emissions at location where the total emission over time is too small, and amplifies the remaining pheromone emission to compensate, including spurious emission far from the target zone. The combination of Tikhonov and group LASSO regularizations (bottom left panel on Fig. 14) enables to reconstruct emissions in a large area, since the Tikhonov regularization pulls the emissions on the whole domain toward the righ order of magnitude, while the group-LASSO regularization counterbalances that by filtering out emission in areas with low total emission. Finally, the combination of all three regularizations (bottom middle panel on Fig. 14) proposes a mix between the population dynamics regularization and the combination of group-LASSO and Tikhonov regularizations.

3.7.2. RMSE criteria between the optimum s_a and the target s_t. The $RMSE_{tot}$ and $RMSE_{supp}$ give comparable insights for the different regularizations. The population dynamics regularization (orange curve on Fig. 15) enables to smooth the curve, especially around $t = 30\text{min}$ and $t = 60\text{min}$ where discontinuities are visible in the reference case (blue curve) due to the pre-concentration effect of the observation operator (10). With the group-LASSO regularization (pink curve on Fig. 15), as consequence of the thresholding and the resulting compensation observed previously,

$RxMSE_{tot}$ is slightly higher and $RxMSE_{supp}$ slightly lower. With the combination of Tikhonov and group-LASSO regularizations (purple curve on Fig. 15), the trend is similar. However, due to the additional Tikhonov regularization that aims to pull the emissions on the whole domain towards the right order of magnitude at the initial time, $RMSE_{supp}$ is lower and $RMSE_{tot}$ higher, due to additional spurious emissions, at only $t = 0\text{min}$. The combination of all three regularizations (red curve on Fig. 15) enables to harness the advantages offered by each of them: the Tikhonov regularization enables to recover the right order magnitude at $t = 0$, while the group-LASSO regularization filters out the spurious emissions (including the one introduced by the Tikhonov regularization) and the population dynamics regularization propagates the emissions reconstructed at $t = 0s$ through the population dynamics model (which smooths the $RxMSE$ curve). The different regularization terms achieve small errors for $t > 60\text{min}$, when the insects stop to emit, indicating that the drop of pheromone levels is correctly reconstructed.

3.7.3. Prediction of insect presence. We now focus on insects presence predictions provided by the different regularization terms together with the corresponding miss rate ρ_{fa} and false discovery rate ρ_{fp}, see respectively the upper and lower panels of Fig. 16. The population dynamics regularization (orange curves on Fig. 16) provides among the lowest ρ_{fa}, especially during the first 20 minutes, but a higher ρ_{fp}. As observed previously on the $RMSE$ criteria, the population dynamics regularization enables to smooth the curves, especially around $t = 30\text{min}$ where a discontinuity is visible in the reference case (blue curve). This smoothing feature enables also to achieve lower ρ_{fa} and ρ_{fp} just after $t = 30\text{min}$ compared to the reference case (blue curve). With the group-LASSO regularisation (pink curves on Fig. 16), ρ_{fa} and ρ_{fp} stay stable and does not vary significantly except at the $t = 30\text{min}$ discontinuity as in the reference case without regularization (blue curve). However, due to the thresholding feature of the group-LASSO regularisation, ρ_{fa} and ρ_{fp} are respectively higher and lower than the one obtained without regularization (blue curve), indicating a too stringent variable selection. With the combination of Tikhonov and group-LASSO regularizations (purple curves on Fig. 16), as observed on the $RMSE$ criteria, the presence of insects is over-predicted, ρ_{fa} is very low and ρ_{fp} very high at $t = 0\text{min}$ due to the Tikhonov regularization implying additional but potential spurious pheromone emission over the whole domain. After that, ρ_{fa} rapidly increases while ρ_{fp} rapidly drops, and the trend are similar to the trend observed with the group-LASSO regularization alone (pink curve), but still with lower values likely due to the interaction between the group-LASSO and Tikhonov regularizations. The three regularizations interact between each other in a efficient way (Fig. 16, red curves) and the combination enables to articulate the advantages offered by each regularizations: the Tikhonov regularization enables to achieve one of the lowest ρ_{fa} at $t = 0\text{min}$, the population dynamics regularization enables to propagate this good initial insect presence prediction and the group-LASSO regularization enables to threshold spurious pheromone emissions and reduces ρ_{fp}. The smoothing feature of the population dynamics regularization is also visible on ρ_{fa} around $t = 30\text{min}$ where ρ_{fa} obtained with the reference case (blue curves) is discontinuous. Let us note that with the combination of all three regularizations, in contrast to the case involving only group-LASSO regularization (pink curves), ρ_{fp}, while being among the lowest in particular around $t = 20\text{min}$, exhibits a relatively high value at $t = 0\text{min}$. Furthermore, in contrast to the case involving the combination of Tikhonov and group-LASSO regularizations (purple curves), ρ_{fa} takes time to decrease and become one of the lowest among the different regularizations.

3.7.4. Estimating active sensors with the one-sensor adjoint state criteria. We now look at the maximal total contributions $\|c^*_{i,tot}\|_\infty$ of the sensors and how the regularizations influence these sensors contributions, see Sec. 2.4.3. Similar trends at those observed in the Toy case are visible

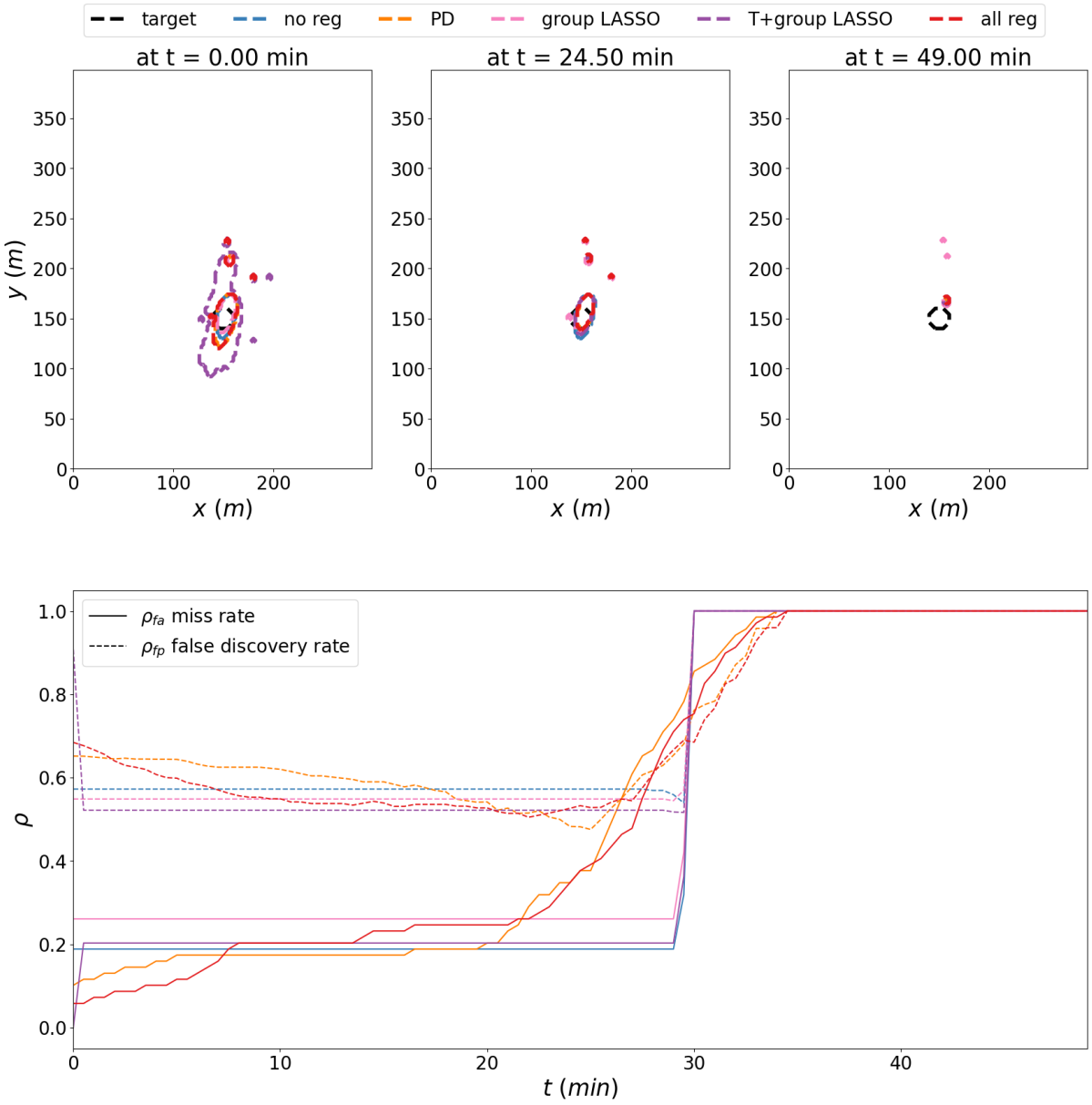


Figure 16 – Prediction of insects presence with several regularization strategies: with no regularization term (no reg, blue), with the population dynamic-informed regularization term (32) (PD, orange), with the group-LASSO regularization term (25) (group LASSO, pink), with Tikhonov and group-LASSO regularization terms (T+group LASSO, purple), and a combination of all three population dynamic-informed, LASSO and Tikhonov regularization terms (all reg, red). Upper panel: map of presence prediction at three different time, true presence in black. Lower panel, miss rate ρ_{fa} (17) (solid lines) and false discovery rate ρ_{fp} (18) (dashed lines) vs t.

here, especially without regularization (blue curve on Fig. 17): a few sensors mainly contribute, then a plateau is attained (less visible than in the Toy case but still observable) and finally, a few sensors (here approximately 4) without significant contribution.

In the FAW case, the addition of the population dynamics regularization (orange curve on Fig. 17) does not strongly influence $\|c^*_{i,tot}\|_\infty$ since the curve of $\|c^*_{i,tot}\|_\infty$ vs the sensor rank in the contribution order obtained with the population dynamic regularization is very similar to that obtained in the situation without regularization (blue curves). With the group-LASSO regularization (pink curve on Fig. 17), the contribution of the sensors that already contribute the most in the

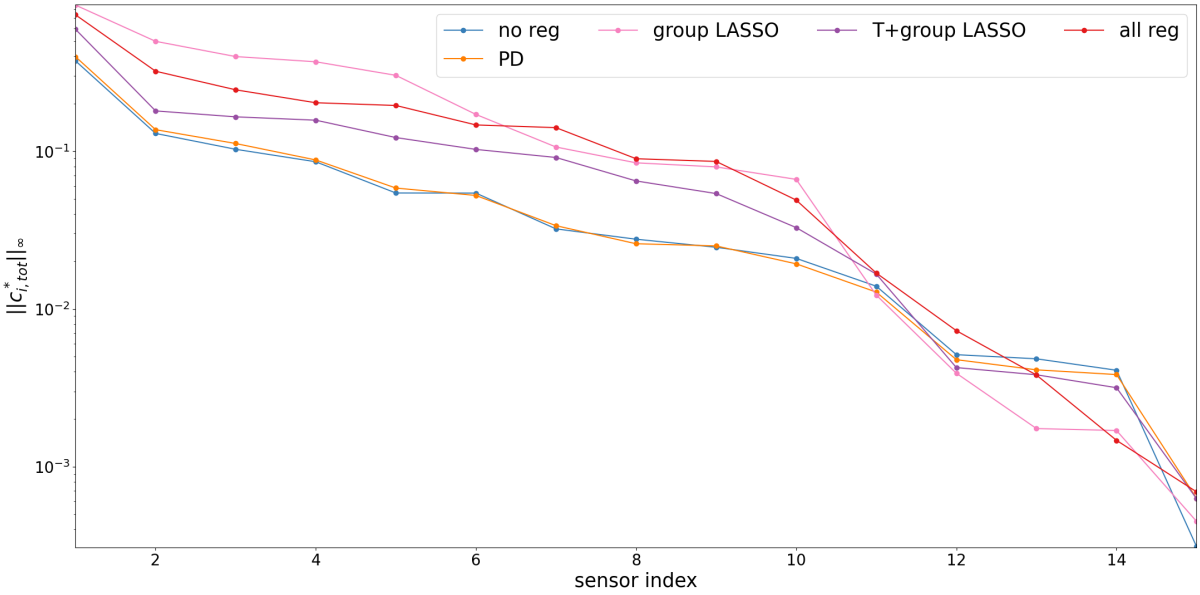


Figure 17 – Sensor importance in BI-DA. Maximal total contribution $\|c^*_{i,tot}\|_\infty$ vs sensor rank when sorted by contribution order for several regularization strategies in the FAW case. It includes no regularization (no reg, blue), population dynamic-informed (PD, orange), group-LASSO (group LASSO, pink), Tikhonov and group-LASSO (T+group LASSO, purple), and a combination of all three population dynamic-informed, LASSO and Tikhonov regularization terms (all reg, red). The quantity $\|c^*_{i,tot}\|_\infty$ indicates the maximal (with respect to time and space) total contribution of sensor i all along the optimization path

reference case (blue curve) are emphasized, while the contribution of the sensors that contribute the less in the reference case are lowered. This is due to the fact that the group-LASSO regularization thresholds all the pheromone emissions at places where the total amount of pheromone are low. Consequently to these thresholding, there is an amplification of the contributions of the major sensors to compensate for the thresholding of the locations that contribute less.

With the combination of Tikhonov and group-LASSO regularizations (purple curve on Fig. 17), the trend is close to the group-LASSO regularization only but with lower absolute values for $\|c^*_{i,tot}\|_\infty$, indicating that the Tikhonov regularization brings information at $t = 0\text{min}$ that supplement information brought by the sensors. When adding population dynamics to Tikhonov and group-LASSO (Fig. 17, red curve), the sensor contribution ranking is comparable to the situation with combination of Tikhonov and group-LASSO regularizations (purple curve).

4. Discussion

4.1. Incorporating biological knowledge with BI-DA

In this study, we introduced a methodology to infer insect pest presence maps from signals of a network of pheromone sensors dispatched over a field that must be protected against a pest. The inference method is built upon variational data assimilation techniques: the aim is to reconstruct a control variable or a hidden state (here the quantity of pheromone emitted by the insects) by combining a direct model (here the pheromone propagation model derived from the atmospheric fluid dynamics) and data (here from the pheromone sensors). This combination is done by minimizing the distance between the data and an estimate of the observed variable. Once the pheromone emission reconstructed, they are taken as a proxy of the pest presence

map. As the classical Data Assimilation (DA) problem seems ill-conditioned and enables to reconstruct the quantity of pheromone emitted in a vicinity of the sensors, we proposed several regularization terms specifically designed to add prior Biological Information (BI) of the insects to improve the accuracy of the optimal solution. In particular, we introduced a population dynamics regularization that explicitly model the behavior of the pest of interest, particularly for pheromone emission. Thus, the BI-DA reconstruction is pulled towards a solution which is consistent with the physiology of the pest. The different regularization terms have been rigorously tested in a toy case, using different criteria, such as RMSE or false discovery and miss rates. The population dynamics regularization reduces the RMSE, especially at the end of the time window when the data are missing, and reduces the miss rates without too much enhancing the false discovery rate, compared to the reference case without regularization. But using the population dynamics term conjointly with Tikhonov and LASSO terms provide the best results regarding presence/absence prediction, pulling the miss rates towards zero with lower rate of false discovery. However, if few sensors have a strong contribution to the reconstruction with the population dynamics regularization, which allows to expect to get comparable estimates with less sensors, coupling the different regularizations increase the number of sensors that have a strong contribution to the estimate.

4.2. Strong vs weak constraint in biology-informed data assimilation

One of the most classical method, referred in this study as classical VDA, consists in considering the direct model as strong constraint of the minimization problem: the estimate of the observed variable is the solution of the direct model, knowing a given estimate of the control variable. On the other hand, if the hidden state to reconstruct is the state variable of the direct model, one can use methods where the direct model acts as weak constraint of the minimization problem: the minimization problem aims to minimize the distance between the data and the estimate of the observed variable computed from the hidden state, and the residual of the direct model, see *e.g.* Trémolet, 2007.

In the BI-DA method we propose, the pheromone propagation model and the pheromone sensor data are combined with biological information, especially model of the dynamics of the insect population. Whilst the pheromone propagation model is used as strong constraint of the minimization problem, the biological information are incorporated in the minimization problem as a weak constraint through the biology-informed regularization terms, see Sec. 3.1. Thus, the combination of data, pheromone propagation model and biological information is based on a mix of strong and weak constraints of the minimization problem.

One could propose a method where the minimization problem is strongly constrained by both the pheromone propagation model and the population dynamics model. For instance, in such situation, the control to infer would be the initial density of insects $p(t = 0, x, y)$ and the observed variable would be estimated through the coupling of the population dynamics model and the pheromone propagation model in the following way. For an estimate of $p(t = 0)$, the density of insects $p(t, x, y)\ \forall(t, x, y)$ is propagated with the population dynamics model and thereby the quantity of pheromone emitted $s(t, x, y) = q(t)p(t, x, y)$ can be computed. Then, the pheromone propagation model (given $s(t, x, y)\ \forall(t, x, y)$) is used to evaluate the concentration of pheromone $c(t, x, y)\ \forall(t, x, y)$ and the observed variable can be computed. One could go even further by coupling the population dynamics model of the female insect population emitting pheromone (depending on the density of male insects), the population dynamics model of the male population (depending on the pheromone concentration and its gradient) and the pheromone propagation model (depending on the density of female insects emitting pheromone).

However, these methods based only on strong constraints require a good knowledge of the population dynamics. Actually, adding a strong constraint remains to assume that the observed data are perfectly explained by the model included in the strong constraint, with small modelling error.

On the contrary, the weak constraints resulting from the biological informations included in the BI-DA intrinsically take into account the modelling errors of the underlying biological processes. As the weak constraint tries to minimize the residual of the population dynamics model, an error model can be added to this residual to control its spatio-temporal correlation structure. Therefore, the BI-DA method does not require as much precise knowledge of the pest biology as strong constraints. It is then better suited for situations where the biological informations is only partially known or may be subject to errors or uncertainties. Moreover, strongly constraining by both the pheromone propagation model and the population dynamics model could make the minimization problem more difficult to solve and lead to identifiability issues. Indeed, the minimum of the cost function may not be unique if the population dynamics is non-linear, and adding multiple strong constraints may worsen the conditioning of the cost function.

4.3. Uncertainty sources in the model

The proposed BI-DA method and the inference of pheromone sources are subject to various source of uncertainties. These uncertainties are of several types: uncertainties related to the environmental parameters, to the pheromone propagation model, to the test cases, to the sensors and the data. In this section, these uncertainties are briefly presented and discussed.

Uncertainties related to the environmental parameters. The pheromone propagation is modelled using the 2D CTM (1). Thus, the resulting pheromone concentration depends on environmental parameters such as the wind velocity field $\vec{u}$ and the loss coefficient τ_{loss}. However, the estimation of these two environmental parameters is itself uncertain.

In the realistic FAW case, the wind velocity is estimated using the data provided by the Arome model, see Sec. 2.3.2. This model have a spatial sampling of $0.01°(\approx 1.3km)$ and a temporal sampling of $1h$. Thus, the smaller scale are not taken into account in the wind data. On the other hand, the spatial and temporal scale considered in realistic application are smaller: of the order of magnitude of a couple of minutes and of tens of meters. Therefore, in this context, uncertainties related to the small spatio-temporal scale of the wind field could be taken into account.

The loss coefficient τ_{loss} enables to parametrize all the vertical fluxes not directly modelled in the 2D model, see Sec. B.1.2. This coefficient depends in particular on the vegetation that absorbs a part of the pheromone that settles on the ground, and on the atmosphere stability. In this study and up to now, this coefficient is calibrated so the 2D CTM fits to a particular solution of the 3D CTM, see Sec. B.1.3. However, the calibration of τ_{loss} with respect to the land use remains rather empirical and source of uncertainties.

In order to understand the impact of these uncertainties on the pheromone propagation and on the pheromone source inference, an uncertainty quantification and a sensitivity analysis of the model with respect to these environmental parameters could be conducted, taking into account these different scales.

Uncertainties related to the pheromone propagation model. The 2D CTM model is derived from a 3D CTM model by averaging the 3D model over an atmospheric layer depth. This approximation is source of uncertainties since potential updrafts or vertical turbulence are roughly taken into account through the loss term. Additional modelling errors can come from inaccurate model parametrization or inaccurate modelling hypothesis. These uncertainties could be tackled by

devoting more effort to modelling pheromone propagation. This includes the use of models that takes into account the random nature of the uncertainties and of metamodels.

Uncertainties related to the test cases. To illustrate the performance of the BI-DA method compared to a classical DA method, the data are generated from a target pheromone emission that we aim to retrieve. In these test cases, the computation domain, the location of the target emission and of the sensors could be taken with much worst characteristics. For instance, in both Toy and FAW cases, it is assumed that all the pheromone emissions are inside the computation domain. If some pheromone emission are located out of the computation domain, one might expect that small spurious pheromone emissions will be retrieved inside the domain by the inference to compensate influx through the domain boundaries, that are not consider in the inverse model.

However, as done in the FAW case with the sensors reallocation, see Sec. 3.6, this type of uncertainty can be bypassed following the following steps. First, we start over a large domain and perform the inference. Then, we perform the relocation process based on the one-sensor adjoint state-based criteria as done in the FAW case and reduce the computation domain (and potentially refine the mesh) to a narrower domain around the new sensors location. These steps can be repeated several times.

Uncertainties related to the sensors and the data. The data used in this study are synthetic: they are generated from a target pheromone emission and an observation model that includes a pre-concentration phase, see Sec. 2.3.2. As soon as suitable sensors are available, the classical DA and new BI-DA methods should be tested in progressively less controlled environment and conditions. These tests should start with controlled laboratory experiment with pheromone spray and controlled wind condition, progressing to tests conducted in open fields with uncontrolled environment. This will enable us to test the proposed methods in real condition and improve them by taking into account the properties of the sensors and the statistics of their data.

4.4. Sensor placement and reallocation

In BI-DA, we try to leverage spatio-temporal data obtained through a network of sensors that measure spatio-temporal concentration maps of pheromone. The time and space sampling of these map is key for the accuracy of the inference, as illustrated in the FAW case, see Sec. 3.6. The time sampling is given by the intrinsic functioning of the sensors. Due to the small amount of pheromone in the landscape, pre-concentration chambers accumulating the signal to detect are necessary, which determine the minimal time step of the time sampling. The space sampling is given by the spatial repartition of the sensor over the landscape.

In this study, we did not investigate the sensitivity of the inverse problem to the spatial placement of the sensors. Optimal design for building sampling design minimizing the variance of the inference can be constructed in comparable spatio-temporal inference problems (Parisey et al., 2022). This optimal design indicates, given environmental and landscape characteristics, where the sensors should be placed in order to get the best possible confidence on the estimate. Optimal placement is beyond the scope of this study and will be explored in future works.

However, we introduced a specific metric measuring the maximal total contribution of a given sensor. This metric allows to order the sensors by contribution in the inference, hence distinguishing between key and accessory sensors. One could then imagine an iterative method for a dynamics placement of the sensors where the marginally contributing sensors are reallocated and moved to another place where they could enhance the accuracy of pest presence/absence estimates. We investigated how this criteria was impacted for varying BI-DA regularization terms. We found that some regularization terms supplement the information taken out of most sensors

and only a few key sensors still significantly contribute to the inference. This is the case for population dynamics and Tikhonov regularization terms. As these regularization terms highlight the sensors that do not significantly contribute in the light of both the propagation of the pheromone and the biological information, they can be used to study the placement and reallocation of sensors with regards to the pheromone propagation, including the atmosphere dynamics and environmental parameter, and to the biological knowledge of the insects, including the dynamics of its population. On the contrary, other regularization terms amplify the information taken out of the sensors and the contribution of every sensors. This is particularly the case for regularizations involving a LASSO term. This stresses the fact that optimal placement strategies will depend on the selected regularization terms.

5. Conclusion

We introduced in this study a numerical framework to solve the inverse problem for pest detection from the assimilation of pheromone concentration data. We specifically worked on adding prior biological knowledge on the population dynamics and the behavior of the pest to the inference problem. Diverse regularizations to biologically inform the data assimilation have been proposed and benchmarked. The population dynamics model enables to propagate emissions well estimated by the data to places and times that are not covered by the assimilation of the data set.

To reduce the pest miss rate, synonymous of a better control of the pest infestation, without increasing too much the false discovery rate, synonymous of unnecessary pesticide application, the following conclusions can be formulated:

(1) Some regularization terms, such as LASSO or group-LASSO, filter out spurious signals, hence enhancing the information provided by the sensors. This leads to lower false discovery rate but higher miss rate.
(2) Some regularization terms, such as the population dynamic-informed or Tikhonov regularization terms, provide prior information that can partially supplement the signal provided by the sensors, leading to lower miss rate, sometimes to the cost of a higher false discovery rate.
(3) The sensor contribution to the inference can be measured with the one-sensor adjoint state criteria. Regularization terms can decrease the contribution of some sensors and enable to efficiently discriminate sensors with high and low contributions. This implies that less sensors are required to achieve equivalent accuracy in the inference. Moreover, sensors with the lowest contribution can be moved to increase inference accuracy of emission sources. In this respect, population-dynamic or Tikhonov regularization terms are convenient to benchmark sensor position.
(4) Once sensor location is fixed, cumulating regularization terms can reduce miss rate while keeping false discovery rates in reasonable values. In particular, it is worth using a composite regularization mixing knowledge on favorable habitat and population dynamics and thresholding spurious pheromone emissions.

Acknowledgments

Funding

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

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

6. References

Adam L, Branda M (2016). *Sparse Optimization for Inverse Problems in Atmospheric Modelling*. *Environmental Modelling & Software* **79**, 256–266. https://doi.org/10.1016/j.envsoft.2016.02.002. (Visited on 10/09/2023).

Adamczyk-Chauvat K, Kassa M, Papaïx J, Kiêu K, Stoica RS (2024). *Statistical Inference for Random T-tessellations Models. Application to Agricultural Landscape Modeling*. *Annals of the Institute of Statistical Mathematics* **76**, 447–479. https://doi.org/10.1007/s10463-023-00893-3. (Visited on 05/01/2024).

Bocquet M (2014). *Introduction to the Principles and Methods of Data Assimilation in the Geosciences*.

Boukal DS, Berec L (2009). *Modelling Mate-Finding Allee Effects and Populations Dynamics, with Applications in Pest Control*. *Population Ecology* **51**, 445–458. https://doi.org/10.1007/s10144-009-0154-4.

Britton N (1989). *Aggregation and the Competitive Exclusion Principle*. *Journal of Theoretical Biology*. https://doi.org/10.1016/S0022-5193(89)80189-4.

Conchou L, Lucas P, Meslin C, Proffit M, Staudt M, Renou M (2019). *Insect Odorscapes: From Plant Volatiles to Natural Olfactory Scenes*. *Frontiers in Physiology* **10**, 972. https://doi.org/10.3389/fphys.2019.00972. (Visited on 07/28/2023).

Deutsch CA, Tewksbury JJ, Tigchelaar M, Battisti DS, Merrill SC, Huey RB, Naylor RL (2018). *Increase in Crop Losses to Insect Pests in a Warming Climate*. *Science* **361**, 916–919. https://doi.org/10.1126/science.aat3466. (Visited on 06/26/2023).

EFSA PLH Panel et al. (2018). *Pest Risk Assessment of Spodoptera Frugiperda for the European Union*. *EFSA Journal* **16**. https://doi.org/10.2903/j.efsa.2018.5351. (Visited on 06/26/2023).

Fisher BEA, MacQueen JF (1981). *A Theoretical Model for Particulate Transport from an Elevated Source in the Atmosphere*. *IMA Journal of Applied Mathematics* **27**, 359–371. https://doi.org/10.1093/imamat/27.3.359. eprint: https://academic.oup.com/imamat/article-pdf/27/3/359/1735525/27-3-359.pdf.

Fisher RA (1937). *The Wave of Advance of Advantageous Genes*. *Annals of eugenics* **7**, 355–369.

Furter J, Grinfeld M (1989). *Local vs. Non-Local Interactions in Population Dynamics*. *Journal of Mathematical Biology*. https://doi.org/10.1007/BF00276081.

Ganga Prasath S, Mahadevan L (2023). *Rheomergy: Collective Behaviour Mediated by Active Flow-Based Recruitment*. *Proceedings of the Royal Society A* **479**, 20220470. https://doi.org/10.1098/rspa.2022.0470.

Gebbers R, Adamchuk VI (2010). *Precision Agriculture and Food Security*. *Science* **327**, 828–831. https://doi.org/10.1126/science.1183899. (Visited on 07/27/2023).

Genieys S, Volpert V, Auger P (2006). *Pattern and Waves for a Model in Population Dynamics with Nonlocal Consumption of Resources*. *Mathematical Modelling of Natural Phenomena*. https://doi.org/10.1051/mmnp:2006004.

IPPC Secretariat . (2021). *Scientific Review of the Impact of Climate Change on Plant Pests*. FAO on behalf of the IPPC Secretariat. https://doi.org/10.4060/cb4769en. (Visited on 06/26/2023).

Isakov V (2017). *Inverse Problems for Partial Differential Equations*. Vol. 127. Applied Mathematical Sciences. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-51658-5. (Visited on 10/09/2023).

Kaltenbacher B, Neubauer A, Scherzer O (2008). *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. Vol. 6. Walter de Gruyter. https://doi.org/10.1515/9783110208276.

Kareiva PM (1983). *Local Movement in Herbivorous Insects: Applying a Passive Diffusion Model to Mark-Recapture Field Experiments*. *Oecologia* **57**, 322–327. JSTOR: 4216967.

Keller EF, Segel LA (1970). *Initiation of Slime Mold Aggregation Viewed as an Instability*. *Journal of Theoretical Biology* **26**, 399–415. https://doi.org/10.1016/0022-5193(70)90092-5.

Keller EF, Segel LA (1971). *Model for Chemotaxis*. *Journal of Theoretical Biology* **30**, 225–234. https://doi.org/10.1016/0022-5193(71)90050-6.

Khadka R, Aydemir N, Carraher C, Hamiaux C, Colbert D, Cheema J, Malmström J, Kralicek A, Travas-Sejdic J (2019). *An Ultrasensitive Electrochemical Impedance-Based Biosensor Using Insect Odorant Receptors to Detect Odorants*. *Biosensors and Bioelectronics* **126**, 207–213. https://doi.org/10.1016/j.bios.2018.10.043. (Visited on 04/28/2024).

Le Dimet FX, Talagrand O (1986). *Variational Algorithms for Analysis and Assimilation of Meteorological Observations: Theoretical Aspects*. *Tellus A: Dynamic Meteorology and Oceanography* **38**, 97–110. https://doi.org/10.3402/tellusa.v38i2.11706.

Leelössy Á, Molnár F, Izsák F, Havasi Á, Lagzi I, Mészáros R (2014). *Dispersion Modeling of Air Pollutants in the Atmosphere: A Review*. *Open Geosciences*. https://doi.org/doi:10.2478/s13533-012-0188-6.

Lions JL (1971). *Optimal Control of Systems Governed by Partial Differential Equations*. Vol. 170. Springer Verlag.

Lu Y, Li H, Zhuang S, Zhang D, Zhang Q, Zhou J, Dong S, Liu Q, Wang P (2014). *Olfactory Biosensor Using Odorant-Binding Proteins from Honeybee: Ligands of Floral Odors and Pheromones Detection by Electrochemical Impedance*. *Sensors and Actuators B: Chemical* **193**, 420–427. https://doi.org/10.1016/j.snb.2013.11.045. (Visited on 04/28/2024).

Lushi E, Stockie JM (2010). *An Inverse Gaussian Plume Approach for Estimating Atmospheric Pollutant Emissions from Multiple Point Sources*. *Atmospheric Environment* **44**, 1097–1107. https://doi.org/10.1016/j.atmosenv.2009.11.039. (Visited on 10/09/2023).

Malou T, Labarthe S (2024a). *BI-DA-companion-code*. https://doi.org/10.5281/ZENODO.11175422. (Visited on 05/10/2024).

Malou T, Labarthe S (2024b). *Pherosensor-Toolbox*. [object Object]. https://doi.org/10.5281/ZENODO.11110047. (Visited on 05/10/2024).

Martinez D, Arhidi L, Demondion E, Masson JB, Lucas P (2014). *Using Insect Electroantennogram Sensors on Autonomous Robots for Olfactory Searches*. *Journal of Visualized Experiments*, 51704. https://doi.org/10.3791/51704. (Visited on 04/28/2024).

McMahon T, Denison P (1979). *Empirical Atmospheric Deposition Parameters—a Survey*. *Atmospheric Environment (1967)* **13**, 571–585.

McNeil JN (1991). *Behavioral Ecology of Pheromone-Mediated Communication in Moths and Its Importance in the Use of Pheromone Traps*. *Annual Review of Entomology* **36**, 407–430. https://doi.org/10.1146/annurev.en.36.010191.002203. (Visited on 07/29/2023).

Moitra P, Bhagat D, Pratap R, Bhattacharya S (2016). *A Novel Bio-Engineering Approach to Generate an Eminent Surface-Functionalized Template for Selective Detection of Female Sex Pheromone of Helicoverpa Armigera*. *Scientific Reports* **6**, 37355. https://doi.org/10.1038/srep37355. (Visited on 04/28/2024).

Murray JD (2002). *Mathematical Biology: I. An Introduction*. Springer.

Myrick AJ, Baker TC (2011). *Locating a Compact Odor Source Using a Four-Channel Insect Electroantennogram Sensor*. *Bioinspiration & Biomimetics* **6**, 016002. https://doi.org/10.1088/1748-3182/6/1/016002. (Visited on 04/28/2024).

Negri RM, Bernik D (2008). *Tracking the Sex Pheromone of Codling Moth against a Background of Host Volatiles with an Electronic Nose*. *Crop Protection* **27**, 1295–1302. https://doi.org/10.1016/j.cropro.2008.02.008. (Visited on 04/28/2024).

Nesterov Y (2018). *Lectures on Convex Optimization*. Vol. 137. Springer.

Nisbet RM, Gurney W (1982). *Modelling Fluctuating Populations*. John Wiley.

Oerke EC (2006). *Crop Losses to Pests*. *The Journal of Agricultural Science* **144**, 31–43. https://doi.org/10.1017/S0021859605005708. (Visited on 06/26/2023).

Okubo A (1986). *Dynamical Aspects of Animal Grouping: Swarms, Schools, Flocks, and Herds*. *Advances in Biophysics* **22**, 1–94. https://doi.org/10.1016/0065-227X(86)90003-1.

Panofsky HA, Dutton JA (1984). *Atmospheric Turbulence. Models and Methods for Engineering Applications*. *New York: Wiley*.

Parikh N, Boyd S, et al. (2014). *Proximal Algorithms*. *Foundations and trends® in Optimization* **1**, 127–239.

Parisey N, Leclerc M, Adamczyk-Chauvat K (2022). *Optimal Spatial Monitoring of Populations Described by Reaction–Diffusion Models*. *Journal of Theoretical Biology* **534**, 110976.

Pawson SM, Kerr JL, O'Connor BC, Lucas P, Martinez D, Allison JD, Strand TM (2020). *Light-Weight Portable Electroantennography Device as a Future Field-Based Tool for Applied Chemical Ecology*. *Journal of Chemical Ecology* **46**, 557–566. https://doi.org/10.1007/s10886-020-01190-6. (Visited on 07/28/2023).

Perthame B (2006). *Transport Equations in Biology*. Birkhäuser Basel.

Pudykiewicz JA (1998). *Application of Adjoint Tracer Transport Equations for Evaluating Source Parameters*. *Atmospheric Environment* **32**, 3039–3050. https://doi.org/10.1016/S1352-2310(97)00480-9.

Raissi M, Perdikaris P, Karniadakis G (2019). *Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations*. *Journal of Computational Physics* **378**, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

Sasaki Y (1958). *An Objective Analysis Based on the Variational Method*. *Journal of the Meteorological Society of Japan. Ser. II* **36**, 77–88. https://doi.org/10.2151/jmsj1923.36.3_77.

Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A (2019). *The Global Burden of Pathogens and Pests on Major Food Crops*. *Nature Ecology & Evolution* **3**, 430–439. https://doi.org/10.1038/s41559-018-0793-y. (Visited on 06/26/2023).

Shigesada N (1980). *Spatial Distribution of Dispersing Animals*. *Journal of mathematical biology* **9**, 85–96.

Shigesada N, Kawasaki K, Teramoto E (1979). *Spatial Segregation of Interacting Species*. *Journal of Theoretical Biology* **79**, 83–99. https://doi.org/10.1016/0022-5193(79)90258-3.

Skellam JG (1951). *Random Dispersal in Theoretical Populations*. *Biometrika* **38**, 196–218.

Sportisse B (2007). *A Review of Current Issues in Air Pollution Modeling and Simulation*. *Computational Geosciences*.

Steffens C, Manzoli A, Oliveira JE, Leite FL, Correa DS, Herrmann PSP (2014). *Bio-Inspired Sensor for Insect Pheromone Analysis Based on Polyaniline Functionalized AFM Cantilever Sensor*. *Sensors and Actuators B: Chemical* **191**, 643–649. https://doi.org/10.1016/j.snb.2013.10.053. (Visited on 04/28/2024).

Stehman SV (1997). *Selecting and Interpreting Measures of Thematic Classification Accuracy*. *Remote Sensing of Environment* **62**, 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7.

Stockie JM (2011). *The Mathematics of Atmospheric Dispersion Modeling*. *SIAM Review* **53**, 349–372. https://doi.org/10.1137/10080991X. (Visited on 10/09/2023).

Sun GQ (2016). *Mathematical Modeling of Population Dynamics with Allee Effect*. *Nonlinear Dynamics*. https://doi.org/10.1007/s11071-016-2671-y.

Takahashi LT, Maidana NA, Ferreira WC, Pulino P, Yang HM (2005). *Mathematical Models for the Aedes Aegypti Dispersal Dynamics: Travelling Waves by Wing and Wind*. *Bulletin of mathematical Biology* **67**, 509–528. https://doi.org/10.1016/j.bulm.2004.08.005.

Tarantola A (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Vol. 89. SIAM.

Tibshirani R (1996). *Regression Shrinkage and Selection via the Lasso*. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1996.tb02080.x.

Trémolet Y (2007). *Model-Error Estimation in 4D-Var*. *Quarterly Journal of the Royal Meteorological Society* **133**, 1267–1280. https://doi.org/10.1002/qj.94.

Tumlinson JH, Mitchell ER, Teal PEA, Heath RR, Mengelkoch LJ (1986). *Sex Pheromone of Fall Armyworm,Spodoptera Frugiperda (J.E. Smith): Identification of Components Critical to Attraction in the Field*. *Journal of Chemical Ecology* **12**, 1909–1926. https://doi.org/10.1007/BF01041855. (Visited on 04/28/2024).

Wehrenfennig C, Schott M, Gasch T, Sauerwald T, Düring RA, Vilcinskas A, Kohl CD (2012). *Laboratory Characterization of Metal-oxide Sensors Intended for* in Situ *Analyses of Pheromones – SOMMSA Approach*. *physica status solidi (a)* **209**, 935–939. https://doi.org/10.1002/pssa.201100586. (Visited on 04/28/2024).

Yazdani A, Lu L, Raissi M, Karniadakis GE (2020). *Systems Biology Informed Deep Learning for Inferring Parameters and Hidden Dynamics*. *PLOS Computational Biology* **16**. Ed. by Vassily Hatzimanikatis, e1007575. https://doi.org/10.1371/journal.pcbi.1007575. (Visited on 10/09/2023).

Appendix A. Data, script, code, and supplementary information availability

The toolbox containing the core functions for BI-DA is available at https://forgemia.inra.fr/pherosensor/pherosensor-toolbox and (Malou et al., 2024b). The companion codes needed to reproduce the different computation and figures of this paper are available at https://forgemia.inra.fr/pherosensor/companion-code-bi-da and (Malou et al., 2024a).

Appendix B. Supplementary material

B.1. Derivation of the depth-integrated chemical transport model

This section aims to present the derivation of the 2D model of pheromone propagation (1). In a first part, a classical 3D model of the pheromone propagation in the atmosphere, also called 3D CTM (Chemical Transport Model), is presented. Then, the 2D model, presented in Sec. 2.1 and used in this study, is derived from this 3D model. In the last part, the calibration of the 2D model using the 3D model is presented.

B.1.1. 3D chemical transport model. In this section, $C(\vec{x}, t)$ $(g.m^{-3})$ denotes the density of chemical (here the pheromone) at a given point $\vec{x} = (x, y, z) \in \mathbb{R}^2 \times \mathbb{R}^+$ and at a time $t \in \mathbb{R}^+$. The equation of mass conservation of C can be written as the following PDE, see *e.g.* reviews in (Leelössy et al., 2014; Sportisse, 2007):

$$\frac{\partial C}{\partial t}(\vec{x}, t) + \nabla_{3D} \cdot (-\mathbf{K}_{3D}(\vec{x}, t)\nabla_{3D} C(\vec{x}, t) + \vec{u}_{3D}(\vec{x}, t) C(\vec{x}, t)) = S(\vec{x}, t) \tag{34}$$

with $-\mathbf{K}_{3D}\nabla_{3D} C + \vec{u}_{3D} C$ $(g.m^{-3}.m.s^{-1})$ the mass flux of pheromone due to the diffusion and the advection of the pheromone in the atmosphere, and $S(\vec{x}, t)$ $(g.m^{-3}.s^{-1})$ the source term, *i.e.* the quantity of pheromone emitted by the insects per seconds and per air volume in the present context. In this section, ∇_{3D} denotes the 3D gradient operator $\nabla_{3D} = (\partial_x\ \partial_y\ \partial_z)$, in order to distinguish with the 2D gradient operator used in the rest of the article.

This equation can be obtained from the reactant Navier-Stokes equations using the dilution hypothesis. This hypothesis consists in assuming that the chemical of interest is present in the atmosphere in a very small quantity and does not significantly impact the dynamic of the atmosphere. It results that the dynamics of the atmosphere and the dynamics of the pheromone are decoupled.

As classically done in the K-theory, see *e.g.* Panofsky et al., 1984; Sportisse, 2007, the diffusion flux due to fluctuations of the air motion caused by the turbulence of the atmosphere at non-resolved scales is closed using the Fick's law. This law relates the diffusion flux to the gradient of the concentration and leads to the diffusion term $-\mathbf{K}_{3D}\nabla_{3D} C$ with $\mathbf{K}_{3D}(\vec{x}, t)$ the diffusion tensor. In the present case, the diffusion tensor, a priori non-homogeneous, is assumed anisotropic in the Oxy plane but isotropic along the z axis. Thus, the diffusion tensor can be written as: $\mathbf{K}_{3D} = \begin{pmatrix} K_{xx} & K_{xy} & 0 \\ K_{yx} & K_{yy} & 0 \\ 0 & 0 & K_{zz} \end{pmatrix}$.

The advection flux of mass $\vec{u}_{3D} C$ is due to the wind and the gravity, and $\vec{u}_{3D}(\vec{x}, t)$ is the velocity of the pheromones in the atmosphere. Thus, the velocity of the pheromones can be written as $\vec{u}_{3D} = (u, v, -w_{set})$. The components u and v of the velocity are given by the wind velocity. Since the vertical wind velocity is assumed to be negligible and the pheromones are heavier than the air, the vertical component $-w_{set}$ is the settling velocity, *i.e.* the velocity at which the pheromones fall due to the gravity.

The dilution hypothesis and decoupling of the pheromone and atmosphere dynamics imply that the diffusion tensor $\mathbf{K}_{3D}$ and the wind field $\vec{u}_{3D}$ can be computed separately by a meteorological model and given as environmental parameters of the CTM model (34).

As the pheromones are settled under the effect of the gravity, they end up on the ground and a part of the pheromones are absorbed by the ground. To model this phenomenon, the following boundary condition on the vertical mass flux is commonly used:

$$(K_{zz}\partial_z C + w_{set} C)\big|_{z=0} = w_{dep} C\big|_{z=0} \tag{35}$$

with w_{dep} the deposition velocity.

B.1.2. Depth integration of the 3D CTM model. The 3D CTM (34) is now integrated over a layer of the atmosphere $z \in [h_-; h_+]$ in order to obtain the 2D model used and presented in Sec. 2.1:

$$\int_{h_-}^{h_+} \frac{\partial C}{\partial t} - \nabla_{3D} \cdot (\mathbf{K}_{3D}\nabla_{3D} C) + \nabla_{3D} \cdot (\vec{u}_{3D} C)\, dz = \int_{h_-}^{h_+} S\, dz \tag{36}$$

In the following, the total concentration of pheromone in the layer is denoted $c(x, y, t) = \int_{h_-}^{h_+} C(x, y, z, t) dz$. The integral of the different terms are now computed in order to obtain a 2D model.

The integral of the time derivative becomes simply:

$$\int_{h_-}^{h_+} \partial_t C dz = \partial_t c \tag{37}$$

The integral of the advection term can be written as:

$$\int_{h_-}^{h_+} \nabla_{3D} \cdot (\vec{u}_{3D} C) dz = \underbrace{\int_{h_-}^{h_+} \partial_x(uC) + \partial_y(vC) dz}_{=I_1} + \underbrace{\int_{h_-}^{h_+} \partial_z(-w_{set} C) dz}_{=I_2}$$

Assuming that the velocity components u and v are constant or near the average value over the layer of interest, the integral I_1 becomes:

$$I_1 = \partial_x \left(\int_{h_-}^{h_+} uC dz \right) + \partial_y \left(\int_{h_-}^{h_+} vC dz \right) = \partial_x (uc) + \partial_y (vc) \tag{38}$$

The integral of the diffusion term can be written as:

$$-\int_{h_-}^{h_+} \nabla_{3D} \cdot (\mathbf{K}_{3D} \nabla_{3D} C)\, dz = -\underbrace{\int_{h_-}^{h_+} \partial_x \left(K_{xx} \partial_x C + K_{xy} \partial_y C \right) + \partial_y \left(K_{yx} \partial_x C + K_{yy} \partial_y C \right) dz}_{=I_3} - \underbrace{\int_{h_-}^{h_+} \partial_z \left(K_{zz} \partial_z C \right) dz}_{=I_4}$$

In the present case, the diffusion coefficients in the plane Oxy K_{xx}, K_{xy}, K_{yx} and K_{yy} is assumed to not depend on the depth z which is a common assumption. Thus, the integral I_3 becomes:

$$\begin{aligned} I_3 &= \partial_x \left(\int_{h_-}^{h_+} K_{xx} \partial_x C + K_{xy} \partial_y C dz \right) + \partial_y \left(\int_{h_-}^{h_+} K_{yx} \partial_x C + K_{yy} \partial_y C dz \right) \\ &= \partial_x \left(K_{xx} \partial_x c + K_{xy} \partial_y c \right) + \partial_y \left(K_{yx} \partial_x c + K_{yy} \partial_y c \right) \end{aligned} \tag{39}$$

From the integrals of the advection and the diffusion terms, it remains the vertical mass fluxes $I_2 - I_4$ that become:

$$I_2 - I_4 = -\int_{h_-}^{h_+} \partial_z \left(K_{zz} \partial_z C + w_{set} C \right) dz = -[K_{zz} \partial_z C + w_{set} C]_{z=h_-}^{z=h_+} \tag{40}$$

In the present study, these vertical fluxes, which model the quantity of pheromones leaving the layer of interest, are closed using a 1^{st} order term, *i.e.* by a reaction term weighted by a loss rate $\tau_{loss}(x, y)$:

$$I_2 - I_4 = -[K_{zz} \partial_z C + w_{set} C]_{z=h_-}^{z=h_+} \approx \tau_{loss} c \tag{41}$$

At last, the total amount of pheromone emitted by the insects in the layer of interest is denoted:

$$\int_{h_-}^{h_+} S(x, y, z, t) dz = s(x, y, t) \tag{42}$$

By introducing the expression of the integrals (37), (38), (39), (41) and (42) in the 3D CTM integrated over a layer of the atmosphere (36), we obtain the 2D CTM:

$$\frac{\partial c}{\partial t} - \nabla \cdot (\mathbf{K} \nabla c) + \nabla \cdot (\vec{u} c) + \tau_{loss} c = s \tag{43}$$

with the $\vec{u}$ and $\mathbf{K}$ resp. the wind field and diffusion tensor in the Oxy plane.

B.1.3. Calibration of the loss coefficient τ_{loss}. In the derivation of the 2D CTM from the 3D CTM, the vertical fluxes are closed by a reaction term weighted by a loss coefficient τ_{loss}. This coefficient depends on the vertical fluxes. Thus, it depends on the atmosphere (through the vertical diffusion), on the pheromone's molecules (through the settling velocity) and on the vegetation (through the deposition velocity). However, the estimation of τ_{loss} is not straightforward. In this section, a method of estimation of τ_{loss} is proposed. In the present study, τ_{loss} is estimated by

comparing the solution of the 2D CTM derived in the previous section and the integral over the layer of the Gaussian plume, the solution of the 3D CTM (34) in a particular case.

The following assumptions are made. It is assumed that the diffusion tensor is isotropic and constant ($\mathbf{K}_{3D} = KId$) and that the wind is unidirectional and constant ($\vec{u}_{3D} = (u, 0, -w_{set})$). Moreover, it is assumed that the source term is of the form $S(\vec{x}) = Q\delta(x)\delta(y)\delta(z-H)$ with Q a constant emission rate and H the height at which the source of emission is located. The Gaussian plume is the solution of the 3D CTM under these assumptions and is of the following form, see more details and a review on the topic in (Stockie, 2011):

$$C_{gp}(r, y, z) = \frac{Q}{u}\frac{1}{\sqrt{4\pi r}}\mathcal{G}_h(r, y, z)\mathcal{G}_v(r, y, z)\mathcal{F}_{bc}(r, y, z) \tag{44}$$

with:

$\mathcal{G}_h(r, y, z) = \exp\left(-\frac{y^2}{4r}\right)$,

$\mathcal{G}_v(r, y, z) = \exp\left(-\frac{w_{set}(z-H)}{2K} - \frac{w_{set}^2 r}{4K^2}\right)\left(\exp\left(-\frac{(z-H)^2}{4r}\right) + \exp\left(-\frac{(z+H)^2}{4r}\right)\right)$,

$\mathcal{F}_{bc}(r, y, z) = \frac{4w_0\pi r}{K}\exp\left(-\frac{w_{set}(z-H)}{2K} - \frac{w_{set}^2 r}{4K^2}\right)\exp\left(\frac{w_0(z+H)}{K} + \frac{w_0^2 r}{K^2}\right)\mathrm{erfc}\left(\frac{z+H}{2\sqrt{r}} + \frac{w_0\sqrt{r}}{K}\right)$,

the change of variable $r = \frac{1}{u}\int_0^x K(\xi)d\xi$, *erfc* the complementary error function and $w_0 = w_{dep} - \frac{w_{set}}{2}$.

In order to get realistic estimations of the loss coefficient τ_{loss}, the solution of the 2D CTM (43) c is empirically compared with the integral over the layer of the Gaussian plume $\int_{h_-}^{h_+} C_{gp}(r, y, z)dz$ under the assumptions presented above in a realistic case.

Based on the average of the wind speed of the meteorological data, the value $u = 12m.s^{-1}$ is used in this section. The settling velocity is set as $w_{set} = 2.7 \times 10^{-3}m.s^{-1}$ as used in (Stockie, 2011). It is assumed that the source of pheromone is located at $H = 2m$ and the layer of interest is $2m$ wide centred at the height of the pheromone source ($h_\pm = H \pm 1m$). The values of deposition velocity w_{dep} with respect to the vegetation used here are found in (McMahon et al., 1979). From this realistic case, it results the value of τ_{loss} given in Tab. 1.

B.2. List of additional population dynamics-informed regularization terms

We detail here additional regularization terms that could be used during the inference process. These regularization terms are not included in the results, since they were not included in the numerical exploration of the BI-DA problems, in order to limit the number of possibilities and to focus on the known behavior of the FallArmy Worm (FAW), our insect of interest. However, we indicate here a large list of features that could be introduced in the population dynamic-informed regularization term (27). The corresponding gradient of the terms are indicated in section B.3.2.

B.2.1. Logarithmic barrier term for exclusion zone. As mentioned in Sec. 3.1.1, the biology-based Tikhonov enables to take into account the favorite habitat of the insects. Nonetheless, this approach does not preclude the possibility of locating insects in areas that they will avoid.

To do so, one can consider a j_{reg} based on a logarithmic barrier function, see *e.g.* (Nesterov, 2018). This approach, and barrier function in general, is in particular used in penalty method that replaces an inequality constraint (*e.g.* $h(s; x, y, t) < b\ \forall t \in [0; T]\ \forall (x, y) \in \Omega_c$) by a penalizing term (*e.g.* $\int_0^T \int_{\Omega_c} -\epsilon \log(b - h(s; x, y, t))dxdydt$ with Ω_c the constraint domain ϵ a very small constant) easier to handle in optimization problems.

In the present case, this penalizing term aims to preclude locating insects in areas that they will avoid or where they will not emit pheromones, *i.e.* $s = 0$ or very small on the exclusion zone Ω_{exc} derived from the biological knowledge of the insect. For instance, it is known that the FAW

females emit pheromones in a vertical position on vegetation that are favorable for the growth of the larva. Hence, the FAW females will not emit pheromones in some areas that are not suitable such as roads.

Therefore, the following regularization/penalization term is considered:

$$j_{reg,exc}(s) = \int_0^T \int_{\Omega_{exc}} -\epsilon \log(\epsilon - s(x,y,t)^2) dxdydt \tag{45}$$

Let us note that this logarithmic barrier function is smooth (and hence differentiable) and tends to the indicator function (0 if $s = 0$ locally, $+\infty$ otherwise) when $\epsilon \to 0$.

B.2.2. Population-dynamics models. Recall that the quantity of pheromone emitted s can be written as $s = pq$ with q the quantity of pheromone emitted per insect, which is assumed to be a biological knowledge, and p the density of insect. Moreover, the density p, and consequently s, follows a population dynamic model that can be written in the generic way as following:

$$\begin{aligned} &\partial_t p + \nabla \left(\sum_i F_i(p) \right) - \sum_i R_i(p) = 0 \\ \Leftrightarrow &\frac{1}{q} \partial_t s - \frac{\partial_t q}{q^2} s + \nabla \left(\sum_i \mathcal{F}_i(s) \right) + \sum_i \mathcal{R}_i(s) = 0 \end{aligned} \tag{46}$$

with F_i and R_i are resp. flux and reaction terms that model the behavior of the insect specie of interest, and $F_i(p = \frac{s}{q}) = \mathcal{F}_i(s)$ and $R_i(p = \frac{s}{q}) = \mathcal{R}_i(s)$.

As mentioned earlier, only a little number of terms have been introduced in this study. In this section, we aim to indicate a large list of features and resulting population dynamics terms that could be introduced in the population dynamic-informed regularization term (27).

Diffusion term. The most classical term modelling population dynamics is the diffusion term:

$$F_d(p; x, y, t) = -\sigma(x,y) \nabla p(x,y,t) \Leftrightarrow \mathcal{F}_d(s) = -\frac{1}{q(t)} \sigma(x,y) \nabla s(x,y,t) \tag{47}$$

with σ $[m^2.s^{-1}]$ the diffusion coefficient of the population, depending notably on the specie.

This term models the spread of the population at a macroscopic scale caused by a motion of each individual in a random direction with a probability independent of other individuals or the environment. This diffusion term can be understood as a continuous and large population limit of a discrete microscopic unbiased random walk process with a fixed probability of jump, see *e.g.* (Nisbet et al., 1982) or (Murray, 2002).

Simple models based on the diffusion term (47) are used to model insects population dynamics in (Kareiva, 1983) where the author provides constant values of σ for 12 species based on experimental data.

Advection term. This term models the advection motion of the population toward a determined direction, see *e.g.* (Murray, 2002):

$$F_a(p; x, y, t) = \vec{v}_i(x,y,t) p(x,y,t) \Leftrightarrow \mathcal{F}_a(s) = \frac{1}{q(t)} \vec{v}_i(x,y,t) s(x,y,t) \tag{48}$$

with $\vec{v}_i$ $[m.s^{-1}]$ the velocity field of the population (assumed here independent of p).

For instance, one can model the attraction of the population toward the preferred habitat by choosing the velocity field $\vec{v}_i(x,y) = \nabla \phi(x,y)$, where ϕ denotes the iso-distance line function from this habitat. This term can also be understood as an attractive force produced by the potential function ϕ representing how favorable the environment is, see *e.g.* (Shigesada et al., 1979) and in the review (Okubo, 1986). The two following terms are specific cases of advection terms.

Quorum sensing term. This term models the advection motion of the population but toward a direction depending on the population density:

$$F_{qs}(p;x,y,t) = \vec{v}_i(p;x,y,t)p(x,y,t) \tag{49}$$

with $\vec{v}_i(p)$ the density-dependent velocity field of the population.

This term enables to model the advection and spread of the population where the individuals interact with each other. The most classical density-dependent velocity field is

$$\vec{v}_i(p) = -\sigma_{qs}\nabla p \tag{50}$$

with σ_{qs} $[m^2.m^2.s^{-1}]$ the diffusion coefficient per density unit. This velocity field enables to model a motion due to avoidance of overcrowded places and population pressure, see *e.g.* (Murray, 2002). It can also be understood as the diffusion of the population at a macroscopic scale caused by a motion of each individual in a biased random direction with a probability increasing with the density of individuals, see *e.g.* (Murray, 2002; Nisbet et al., 1982). This density-dependent velocity field (50) leads to the term in variable s

$$\mathcal{F}_{qs}(s) = -\frac{\sigma_{qs}}{q^2}s\nabla s \tag{51}$$

Avoidance of overcrowded places and of population pressure is a common behavior among insects, see *e.g.* (Murray, 2002; Shigesada, 1980) and references therein. For instance, (Shigesada, 1980) proposes and applies a model with the term (49) and the velocity field (50) to an ant lion population. This model is based on experiments highlighting population pressure in this population.

Pheromone sensing term. This term models the advection motion of the population but toward a direction depending on the concentration of pheromone:

$$F_{ps}(p;x,y,t) = \chi p(x,y,t)\nabla c(s;x,y,t) \Leftrightarrow \mathcal{F}_{ps}(s;x,y,t) = \frac{\chi}{q(t)}s(x,y,t)\nabla c(s;x,y,t) \tag{52}$$

with χ $[m^2.g^{-1}.m^2.s^{-1}]$ the chemotactic gain constant ($\chi \in \mathbb{R}^{+*}$).

The insects employ pheromones to attract individuals of the same sex, such as in the case of sex pheromones for FAW, or of both sexes, as is the case with aggregation pheromones for the red palm weevils. To reach the source of the pheromone, the insects moves in the direction of the gradient of concentration of pheromones. This motion induced by a chemical signal is commonly known as chemotaxis, especially in cellular biology. A Keller-Segel-type system (Keller et al., 1970, 1971) can be used to model at the same time both the population dynamics (here of the insects) and the dynamics of the chemical attractant (here the pheromone), see *e.g.* (Murray, 2002; Perthame, 2006).

For instant, (Ganga Prasath et al., 2023) proposes a Keller-Segel-type system to model the dynamics of swarm of bees communicating using a pheromone.

Logistic term. Introduced by Verhulst in the 19^{th} century, the logistic growth is used to model a self-limiting birth-death process, see *e.g.* (Murray, 2002) or (Perthame, 2006) and reference therein:

$$R_l(p;x,y,t) = \frac{r}{\kappa}p(x,y,t)(\kappa - p(x,y,t)) \Leftrightarrow \mathcal{R}_l(s;x,y,t) = \frac{1}{q(t)^2}\frac{r}{\kappa}s(x,y,t)(\kappa q(t) - s(x,y,t)) \tag{53}$$

with r $[s^{-1}]$ the linear reproduction rate ($r \in \mathbb{R}^{+*}$) and κ $[m^{-2}]$ the carrying capacity of the environment ($\kappa \in \mathbb{R}^{+*}$).

This term enables to model a saturation phenomena that arises when the maximal occupation density (the carrying capacity) is exceeded ($p > \kappa$, the death process overcome the birth process),

and invasion front. This saturation is for instance due to intra-specific competition for a limited resource that is consumed faster than it can renewed.

The equation obtained when considering diffusion and logistic source is the well-known Fisher-KPP equation, introduced in (RA Fisher, 1937) to model the spread of favorable gene and first used for population dynamics problems in (Skellam, 1951). A Fisher-KPP-type system is for instance used in (Takahashi et al., 2005) to model the dynamics of mosquitoes populations where the clear water (required for egg-laying) and the amount of nutrient (human blood) are limited, potentially leading to saturation phenomenon.

Allee effect term. The Allee effect is characterized by the decrease of the population due to a small or sparse population. The following term is an extension of the logistic term (53) which additionally models the Allee effect, see *e.g.* a review in (Sun, 2016) and reference therein:

$$R_{Ae}(p;x,y,t) = \frac{r}{\kappa_+\kappa_-}p(x,y,t)(\kappa_+ - p(x,y,t))(p(x,y,t)-\kappa_-) \tag{54}$$
$$\Leftrightarrow \mathcal{R}_{Ae}(s;x,y,t) = \frac{1}{q(t)^3}\frac{r}{\kappa_+\kappa_-}s(x,y,t)(q(t)\kappa_+ - s(x,y,t))(s(x,y,t)-q(t)\kappa_-)$$

with r $[s^{-1}]$ the linear reproduction rate ($r \in \mathbb{R}^{+*}$), κ_+ $[m^{-2}]$ the carrying capacity of the environment ($\kappa_+ \in \mathbb{R}^{+*}$) and κ_- $[m^{-2}]$ the Allee threshold ($\kappa_- \in \mathbb{R}^{+*}$ and $\kappa_- < \kappa_+$) that is the critical density under which the population decrease.

This effect is *e.g.* due to the lack of interactions between individuals of the population. In insect populations, the Allee effect commonly occurs since lower densities reduce the chance of an individual to to locate a mate for purpose of reproduction.

Since the Allee effect can lead to the extinction of the insect population, it is often used in pest control strategy, see *e.g.* (Boukal et al., 2009).

B.2.3. Non-local terms. The non-local terms are generalization of the previously presented terms. These non-local terms enable to model non-local interaction and long range diffusion, leading for instance to spatial patterns, see *e.g.* (Furter et al., 1989), (Genieys et al., 2006) or (Murray, 2002) and reference therein. Non-local terms arise especially when the individuals of the population communicate with each other remotely, *e.g.* using chemical or visual signals, or when the individuals interact with the environment not local but surrounding.

For example, one can extend the logistic term (53) into a non-local version:

$$R_{nll}(p;x,y,t) = \frac{r}{\kappa}p(x,y,t)(\kappa - k * p(x,y,t)) \tag{55}$$
$$\Leftrightarrow \mathcal{R}_{nll}(s;x,y,t) = \frac{1}{q(t)^2}\frac{r}{\kappa}s(x,y,t)(\kappa q(t) - k * s(x,y,t))$$

with the convolution product $k * s(x,y,t) = \int_\Omega k(x-x',y-y')s(x',y',t)dx'dy'$ and k the convolution kernel.

This implies that individuals are not competing for the resource locally available, as with the term (53), but for the resource available in a vicinity. In this term, the convolution product can be seen as an average weighted by the kernel k that defines how the individuals interact with the vicinity and how large is this vicinity. Let us note that the local term (53) is retrieved if k is the Dirac kernel δ. Similarly, one can consider a non-local Allee effect term to model the decrease of the population due to a small or sparse population in a vicinity. For instance, (Britton, 1989) proposed a non-local version of the Fisher-KPP equation based on this non-local logistic term.

As mentioned before, a Keller-Segel type system can be used to model both the dynamics of the population and the dynamics of the chemical attractant, in which the quantity of chemical attractant emitted depends on the population density. In special cases, the equation of the chemical attractant dynamics can be analytically solved. The concentration of the attractant can then be expressed as a convolutional product between the density of the population and a kernel describing the attractant dynamics, which leads to a non-local flux term in the population dynamics model, see *e.g.* (Perthame, 2006).

For example, in the present case, one can approximate the pheromone dynamics, described in the direct model (1), by a Gaussian plume $c_{gp}(s) = g_{gp} * s$ with g_{gp} the Gaussian plume kernel. Injecting this approximation in the pheromone sensing term (52) leads to a non-local pheromone sensing term:

$$\mathcal{F}_{nlps}(s; x, y, t) = \frac{\chi}{q(t)} s(x, y, t) \nabla (g_{gp} * s(x, y, t)) \tag{56}$$

Non-local terms of the shape $R_{lr}(p) \propto k * p$ can be used to model long range diffusion processes. The range of the diffusion processes depends on the choice of the kernel k and its moments, see *e.g.* (Murray, 2002).

B.3. On solving the optimization problem (3)

In this section, we present all the calculation details for solving the optimization problem (3) presented in Sec. 2.2. In the present context, the cost function j is optimized using gradient descent or proximal gradient methods. Recall that the gradient of j, or at least of its differentiable part, is given by

$$\nabla_s j = \nabla_s j_{obs} + \nabla_s j_{reg} \tag{57}$$

with $j_{reg} = \sum_i \alpha_{reg,i} j_{reg,i} \Rightarrow \nabla_s j_{reg} = \sum_i \alpha_{reg,i} \nabla_s j_{reg,i}$.

In Sec. 2.2.2, it is mentioned that $\nabla_s j_{obs} = -c^*$ with c^* the adjoint state, solution of the adjoint model (6). In a first part, the derivation of the adjoint model and of the expression (7) are detailed. Moreover, related models used in this study, such that the one-sensor adjoint model, will be detailed.

In Secs. 3.1 and B.2, several Biology-Informed differentiable $j_{reg,i}$ are proposed. In a second part, the gradients of these $j_{reg,i}$ are calculated.

In Sec. 3.1.2, LASSO and group-LASSO regularization terms are proposed. However, these j_{reg} are not differentiable. Thus, a proximal gradient algorithm is used in this study. In a third part, we present the proximal-gradient algorithm and its application to the LASSO and group-LASSO regularization terms, called ISTA algorithm.

B.3.1. Derivation of the adjoint model and related. In this section, $\mathcal{K}_T$ denotes the control space and $\mathcal{K}_T = \mathcal{L}^2([0; T]; \mathcal{K})$ with $\mathcal{K}$ the space of functions from $\Omega \subset \mathbb{R}^2$ to $\mathbb{R}$ which regularity in space is required in order to solve the direct model (1). We assume that the spaces $\mathcal{K}_T$ and $\mathcal{K}$ are Hilbert spaces and we have
$< f, g >_{\mathcal{K}_T} = \int_0^T < f(t, \cdot), g(t, \cdot) >_{\mathcal{K}} dt$.
In a similar way, $\mathcal{V}_T = \{f; f \in \mathcal{L}^2([0; T]; \mathcal{V}) \text{ and } \partial_t f \in \mathcal{L}^2([0; T]; \mathcal{V})\}$ denotes the space of the state variable, solution of the direct model and we have $< f, g >_{\mathcal{V}_T} = \int_0^T < f(t, \cdot), g(t, \cdot) >_{\mathcal{V}} dt$. Moreover, $\mathcal{D}$ denotes the observations space.

As mentioned previously, the gradient of the cost function is $\nabla_s j(s) = \nabla_s j_{obs}(s) + \alpha_{reg} \nabla_s j_{reg}(s)$ with the gradient of the observation term (4) that is: $\forall \delta s \in \mathcal{K}_T$

$$
\begin{aligned}
(58) \quad < \nabla_s j_{obs}(s), \delta s >_{\mathcal{K}_T} &= 2 < \mathbf{R}^{-1} \left(m(c(s)) - m^{obs} \right), \frac{dm}{dc}(c(s)) \cdot \frac{dc}{ds}(s) \cdot \delta s >_{\mathcal{D}} \\
&= 2 < \left(\frac{dm}{dc}(c(s)) \right)^* \cdot \mathbf{R}^{-1} \left(m(c(s)) - m^{obs} \right), \frac{dc}{ds}(s) \cdot \delta s >_{\mathcal{V}_T} \\
&= 2 < \left(\frac{dc}{ds}(s) \right)^* \cdot \left(\frac{dm}{dc}(c(s)) \right)^* \cdot \mathbf{R}^{-1} \left(m(c(s)) - m^{obs} \right), \delta s >_{\mathcal{K}_T}
\end{aligned}
$$

with $s \mapsto c(s)$ the operator that given s returns the solution of the direct model (1) and $d_s c(s)$ the derivative of this operator with respect to s. However, the operator $(d_s c(s))^*$ may not be easy to calculate. The adjoint model provides a method to estimate the results of this operator.

Linear tangent model. In this section, the direct model is written as :

$$
(59) \qquad \frac{\partial c}{\partial t} + \mathcal{A}(c) = \mathcal{B}(s)
$$

In our context, the differentiable operators are:

$$
(60) \qquad \mathcal{A}(c) = \nabla \cdot (\vec{u} c - \mathbf{K} \nabla c) + \tau_{loss} c \text{ and } \mathcal{B}(s) = s
$$

with the boundary conditions:

- a null diffusive flux $\mathbf{K} \nabla c \cdot \vec{n} = 0 \; \forall (x, y) \in \partial\Omega$,
- a null convective influx $\vec{u} c \cdot \vec{n} = 0 \; \forall (x, y) \in \partial\Omega \cap \{(x, y) | \vec{u}(x, y, t) \cdot \vec{n} < 0\} \; \forall t \in]0; T]$ with $\vec{n}$ the outgoing normal vector.

Denoting $\omega = d_s c \cdot \delta s \in \mathcal{V}_T$, the linear tangent model is then:

$$
(61) \qquad \frac{\partial \omega}{\partial t} + \frac{d\mathcal{A}}{dc}(c(s)) \cdot \omega = \frac{d\mathcal{B}}{ds}(s) \cdot \delta s
$$

with the initial condition $\omega(t = 0) = 0$ (since $c(t = 0) = 0 \; \forall x, y \; \forall s$).

Let us note that, in the present case, $d_c \mathcal{A} = \mathcal{A}$ since $\mathcal{A}$ is a linear operator. By applying the scalar product with a test function $z \in \mathcal{V}_T$ and integration by part, we then have:

$$
(62) \quad < \frac{\partial z}{\partial t} - \left(\frac{d\mathcal{A}}{dc}(c(s)) \right)^* \cdot z, \omega >_{\mathcal{V}_T} - \left[< z, \omega >_{\mathcal{V}} \right]_{t=0}^{t=T} = - < \frac{d\mathcal{B}}{ds}(s) \cdot \delta s, z >_{\mathcal{V}_T} \quad \forall z \in \mathcal{V}_T
$$

Adjoint model. On the other hand, based on the expression of the gradient of j (57) and of j_{obs} (58), the gradient of j satisfies: $\forall \delta s \in \mathcal{K}_T$

$$
(63) \quad < \nabla_s j, \delta s >_{\mathcal{K}_T} = < \nabla_s j_{reg}, \delta s >_{\mathcal{K}_T} + 2 < \left(\frac{dm}{dc}(c(s)) \right)^* \cdot \mathbf{R}^{-1} \left(m(c(s)) - m^{obs} \right), \omega >_{\mathcal{V}_T}
$$

By introducing the linear tangent model (62) in the equation of gradient (63), we then have: $\forall \delta s \in \mathcal{K}_T$ and $\forall z \in \mathcal{V}_T$

(64)

$$
\begin{aligned}
< \nabla_s j, \delta s >_{\mathcal{K}_T} = \quad & < \nabla_s j_{reg}, \delta s >_{\mathcal{K}_T} - < \frac{d\mathcal{B}}{ds}(s) \cdot \delta s, z >_{\mathcal{V}_T} \\
(65) \qquad - \quad & < \frac{\partial z}{\partial t} - \left(\frac{d\mathcal{A}}{dc}(c(s)) \right)^* \cdot z - 2 \left(\frac{dm}{dc}(c(s)) \right)^* \cdot \mathbf{R}^{-1} \left(m(c(s)) - m^{obs} \right), \omega >_{\mathcal{V}_T} \\
+ \quad & \left[< z, \omega >_{\mathcal{V}} \right]_{t=0}^{t=T}
\end{aligned}
$$

Thus, the adjoint model aims to find a particular test function c^* of $\mathcal{V}_T$, so-called adjoint state, that satisfies:

$$\frac{\partial c^*}{\partial t} - \left(\frac{d\mathcal{A}}{dc}(c(s))\right)^* \cdot c^* = 2\left(\frac{dm}{dc}(c(s))\right)^* \cdot \mathbf{R}^{-1}\left(m(c(s)) - m^{obs}\right) \tag{66}$$

with a null final condition $c^*(t = T) = 0\ \forall (x, y) \in \Omega$.

For the differential operator $\mathcal{A}$ given by Eq. (60), the adjoint operator $(d_c\mathcal{A})^*$ is simply $(d_c\mathcal{A})^* \cdot z = -\nabla \cdot (\mathbf{K}^T \nabla z) - \nabla(\vec{u}z) + (\nabla.\vec{u})z + \tau_{loss} z$ and the boundary conditions of the adjoint model are:

- a null diffusive flux $\mathbf{K}^T \nabla c^* \cdot \vec{n} = 0\ \forall (x, y) \in \partial\Omega$,
- a null outgoing convective flux $\vec{u}c^* \cdot \vec{n} = 0\ \forall (x, y) \in \partial\Omega \cap \{(x, y) | \vec{u}(x, y, t) \cdot \vec{n} > 0\}\ \forall t \in]0; T]$ with $\vec{n}$ the outgoing normal vector.

Let us note that by definition, the adjoint model is a linear model.

Computation of the gradient. By introducing the adjoint state as test function in the equation (64), it returns that $\nabla_s j$ satisfies the following equation:

$$< \nabla_s j(s), \delta s >_{\mathcal{K}_T} = < \nabla_s j_{reg}(s), \delta s >_{\mathcal{K}_T} - < \left(\frac{d\mathcal{B}}{ds}(s)\right)^* \cdot c^*(c(s)), \delta s >_{\mathcal{V}_T} \quad \forall \delta s \in \mathcal{K}_T \tag{67}$$

with $(d_s\mathcal{B})^* = Id$ for the differential operator $\mathcal{B}$ given by (60). Thus, it results the expression of the gradient:

$$\nabla_s j(s) = \nabla_s j_{reg}(s) - \left(\frac{d\mathcal{B}}{ds}(s)\right)^* \cdot c^*(c(s)) \tag{68}$$

To estimate the gradient at each optimization iteration, it requires to solve in order the direct model (1) and then the adjoint model (6).

One-sensor and trajectory-integrated one-sensor adjoint models. Recall that the one-sensor adjoint model is mostly obtained by splitting the observation operator $c \mapsto m(c)$ into the sum of one-sensor observation operator $m = \sum_i m_i$. By introducing this observation operator splitting in the adjoint model (6), the adjoint model can be written the following way:

$$\partial_t c^* + \nabla \cdot (\mathbf{K}^T \nabla c^*) + \nabla(\vec{u}c^*) - (\nabla.\vec{u})c^* - \tau_{loss} c^* = \sum_i \sum_k \left(\frac{dm_k}{dc}(c(s))\right)^* \cdot 2\mathbf{R}^{-1} \cdot \left(m_i(c(s)) - m_i^{obs}\right)$$

Assuming that $\left(\frac{dm_k}{dc}(c(s))\right)^* \cdot 2\mathbf{R}^{-1} \cdot \left(m_i(c(s)) - m_i^{obs}\right) = 0$ for $i \neq k$ and as the left hand side is linear with respect to c^*, the adjoint model written (B.3.1) can be split into the one-sensor adjoint models (8). Let us note that, as mentioned in Sec. 2.2.3, the hypothesis $\left(\frac{dm_k}{dc}(c(s))\right)^* \cdot 2\mathbf{R}^{-1} \cdot \left(m_i(c(s)) - m_i^{obs}\right) = 0$ for $i \neq k$ is satisfied in our case if the sensors are independent ($\mathbf{R}^{-1}$ is block-diagonal) and if two sensors are not located in the same place.

Recall from Sec. 2.2.3 that the resulting one-sensor adjoint state is the contribution of a sensor to the gradient $\nabla_s j_{obs}$.

In this study, the sequence $(s_j)_{j\in\mathbb{N}}$ given by the trajectory taken by the optimization algorithm using different DA strategies are compared through one-sensor adjoint state-based criteria. The main one-sensor adjoint state-based criteria used, introduced in Sec. 2.4.3, is the maximal total contribution $\| \sum_j \eta_j c_i^*(s_j) \|_{L^\infty(\Omega\times[0,T])}$, with $(\eta_j)_{j\in\mathbb{N}}$ the sequence of the step sizes and $c_i^*(s_j)$ is the solution of the one-sensor adjoint model:

$$\partial_t c_{i,j}^* + \nabla \cdot (\mathbf{K}^T \nabla c_{i,j}^*) + \nabla(\vec{u}c_{i,j}^*) - (\nabla.\vec{u})c_{i,j}^* - \tau_{loss} c_{i,j}^* = \left(d_c m_i(c(s_j))\right)^* \cdot 2\mathbf{R}^{-1}\left(m_i(c(s_j)) - m_i^{obs}\right)$$

Thus, by multiplying with the step size and summing over the sequences the previous one-sensor adjoint model and due the linearity of the left hand side (that is always true by definition of the

adjoint model), $c^*_{i,tot} = \sum_j \eta_j c^*_i(s_j)$ is solution of the following trajectory-integrated one-sensor adjoint model:

$$\begin{aligned}(69)\qquad &\partial_t c^*_{i,tot} + \nabla \cdot (\mathbf{K}^T \nabla c^*_{i,tot}) + \nabla(\vec{u} c^*_{i,tot}) - (\nabla . \vec{u}) c^*_{i,tot} - \tau_{loss} c^*_{i,tot} \\ &= \sum_j \left(\eta_j (d_c m_i(c(s_j)))^* \cdot 2\mathbf{R}^{-1} \left(m_i(c(s_j)) - m_i^{obs} \right) \right)\end{aligned}$$

This trajectory-integrated one-sensor adjoint model enable to compute the criteria without having to compute $c^*_i(s_j)$ for each s_j.

Let us note that, if the one-sensor observation operator $c \mapsto m$ is linear (which is the case in the present study), the right hand side of the trajectory-integrated one-sensor adjoint model $RHS = \sum_j \left(\eta_j (d_c m_i(c(s_j)))^* \cdot 2\mathbf{R}^{-1} \left(m_i(c(s_j)) - m_i^{obs} \right) \right)$ becomes $RHS = (d_c m_i)^* \cdot 2\mathbf{R}^{-1} \cdot \left(\sum_j \eta_j \left(m_i(c(s_j)) - m_i^{obs} \right) \right)$.

Moreover, if the direct model is linear with respect to s (which is the case in the present study), the right hand side of the trajectory-integrated one-sensor adjoint model even becomes $RHS = (d_c m_i)^* \cdot 2\mathbf{R}^{-1} \cdot \left(m_i \left(c \left(\sum_j \eta_j s_j \right) \right) - \left(\sum_j \eta_j \right) m_i^{obs} \right)$.

B.3.2. Computation of the gradients $\nabla_s j_{reg,i}$. In this section, we calculate the gradient of the differentiable regularization terms $j_{reg,i}$ introduced in Sec. 3.1 and in Sec. B.2. Recall that ∇_s is the gradient operator with respect to s and $\nabla_s j_{reg,i}(s)$ is the function given by $\frac{dj_{reg,i}(s)}{ds} \cdot \delta s = \langle \nabla_s j_{reg,i}(s), \delta s \rangle \ \forall \delta s$.

Gradient of the Tikhonov regularization term. The gradient of the Tikhonov regularization term (23), presented in Sec. 3.1.1 is simply:

$$\nabla_s j_{reg,T}(s) = 2\mathbf{C}^{-1}(s - s_b)$$

Gradient of the logarithmic barrier regularization term. The gradient of the logarithmic barrier regularization term (45), presented in Sec. B.2.1 is simply:

$$\nabla_s j_{reg,exc}(s) = \frac{2\epsilon}{\epsilon - s^2} \mathbb{1}_{\Omega_{exc}}$$

Gradient of the population dynamics-informed regularization term. Recall that the population dynamics-informed regularization term (27), presented in Sec. 3.1.3, can be written as:

$$\begin{aligned} j_{reg,PD}(s) &= \| \partial_t s - \frac{\partial_t q}{q} s + q \nabla \left(\sum_i \mathcal{F}_i(s) \right) + q \sum_i \mathcal{R}_i(s) \|^2_{L^2} \\ &= \langle \partial_t s - \frac{\partial_t q}{q} s + q \nabla \left(\sum_i \mathcal{F}_i(s) \right) + q \sum_i \mathcal{R}_i(s), \partial_t s - \frac{\partial_t q}{q} s + q \nabla \left(\sum_i \mathcal{F}_i(s) \right) + q \sum_i \mathcal{R}_i(s) \rangle \end{aligned}$$

and hence, its differentiable is

$$\begin{aligned} \frac{dj_{reg,PD}(s)}{ds} \cdot \delta s = 2 \langle &\partial_t s - \frac{\partial_t q}{q} s + q \nabla \left(\sum_i \mathcal{F}_i(s) \right) + q \sum_i \mathcal{R}_i(s) \\ &, \partial_t \delta s - \frac{\partial_t q}{q} \delta s + q \nabla \left(\sum_i \frac{d\mathcal{F}_i(s)}{ds} \cdot \delta s \right) + q \sum_i \frac{d\mathcal{R}_i(s)}{ds} \cdot \delta s \rangle \end{aligned}$$

Hence $\frac{dj_{reg,PD}(s)}{ds} \cdot \delta s = \langle \nabla_s j_{reg,PD}(s), \delta s \rangle$ with :

$$\nabla_s j_{reg,PD}(s) = 2 \left(\partial_t - \frac{\partial_t q}{q} + q \sum_i \frac{d \nabla \mathcal{F}_i(s)}{ds} + q \sum_i \frac{d\mathcal{R}_i(s)}{ds} \right)^* \cdot \left(\partial_t s - \frac{\partial_t q}{q} s + q \nabla \left(\sum_i \mathcal{F}_i(s) \right) + q \sum_i \mathcal{R}_i(s) \right)$$

Recall that the adjoint operator is a linear operator: $(F_1 + F_2)^* = F_1^* + F_2^*$

We know have to calculate the adjoint operators $(\partial_t - \frac{\partial_t q}{q})^*$, $\left(\frac{d\nabla\mathcal{F}_i(s)}{ds}\right)^*$ and $\left(\frac{d\mathcal{R}_i(s)}{ds}\right)^*$ for the $\mathcal{F}_i$ and $\mathcal{R}_i$ presented in Sec. 3.1.3 and in Sec. B.2.2. Let us note that no boundary conditions are imposed in the estimation of the population dynamics model. Hence, all the boundary terms are taken into account in the adjoint operators.

Temporal dynamics terms. Here, we calculate the operator $(\partial_t - \frac{\partial_t q}{q})^*$. By applying the scalar product with a test function φ, we have:

$$\langle \partial_t \delta s - \frac{\partial_t q}{q}\delta s, \varphi\rangle = \langle -\partial_t \varphi - \frac{\partial_t q}{q}\varphi, \delta s\rangle + \int_\Omega [\varphi \delta s]_{t=0}^{t=T} dX$$

Therefore, the operator $\left(\partial_t - \frac{\partial_t q}{q}\right)^*$ is the operator such that

$$\varphi \mapsto -\partial_t \varphi - \frac{\partial_t q}{q}\varphi \; \forall (x, y, t) \in \Omega \times]0; T[$$
$$\text{for } \varphi \text{ such that } \varphi = 0 \; \forall (x, y, t) \in \Omega \times \{0; T\}$$

Diffusion term. Here, we calculate the operator $\left(\frac{d\nabla\mathcal{F}_i(s)}{ds}\right)^*$ for the diffusion operator $\mathcal{F}_i = \mathcal{F}_d$ (47). By applying the scalar product with a test function φ, we have:

$$\begin{aligned}\langle \frac{d\nabla\mathcal{F}_d(s)}{ds}\cdot \delta s, \varphi\rangle &= \langle -\nabla\cdot\left(\frac{1}{q}\sigma\nabla\delta s\right), \varphi\rangle \\ &= \langle -\nabla\cdot\left(\frac{1}{q}\sigma^T\nabla\varphi\right), \delta s\rangle + \int_0^T\int_{\partial\Omega}\left(\frac{1}{q}\sigma^T\nabla\varphi\delta s - \frac{1}{q}\varphi\sigma\nabla\delta s\right)\cdot\vec{n}d\Gamma dt\end{aligned}$$

Therefore, the operator $\left(\frac{d\nabla\mathcal{F}_d(s)}{ds}\right)^*$ is the operator such that

$$\varphi \mapsto -\nabla\cdot\left(\frac{1}{q}\sigma^T\nabla\varphi\right) \; \forall (x, y, t) \in \Omega \times [0; T]$$
$$\text{for } \varphi \text{ such that } \frac{1}{q}\sigma^T\nabla\varphi\cdot\vec{n} = 0 \text{ and } \varphi = 0 \; \forall (x, y, t) \in \partial\Omega \times [0; T]$$

Advection term. Here, we calculate the operator $\left(\frac{d\nabla\mathcal{F}_i(s)}{ds}\right)^*$ for the advection operator $\mathcal{F}_i = \mathcal{F}_a$ (48). By applying the scalar product with a test function φ, we have:

$$\begin{aligned}\langle \frac{d\nabla\mathcal{F}_a(s)}{ds}\cdot \delta s, \varphi\rangle &= \langle \nabla\cdot\left(\frac{1}{q}\vec{v}_i\delta s\right), \varphi\rangle \\ &= \langle \nabla\cdot\left(\frac{1}{q}\vec{v}_i\right)\varphi - \nabla\cdot\left(\frac{1}{q}\vec{v}_i\varphi\right), \delta s\rangle + \int_0^T\int_{\partial\Omega}\frac{1}{q}\vec{v}_i\varphi\delta s\cdot\vec{n}d\Gamma dt\end{aligned}$$

Therefore, the operator $\left(\frac{d\nabla\mathcal{F}_a(s)}{ds}\right)^*$ is the operator such that

$$\varphi \mapsto \nabla\cdot\left(\frac{1}{q}\vec{v}_i\right)\varphi - \nabla\cdot\left(\frac{1}{q}\vec{v}_i\varphi\right) \; \forall (x, y, t) \in \Omega \times [0; T]$$
$$\text{for } \varphi \text{ such that } \varphi = 0 \; \forall (x, y, t) \in \partial\Omega \times [0; T]$$

Quorum sensing term. Here, we calculate the operator $\left(\frac{d\nabla\mathcal{F}_i(s)}{ds}\right)^*$ for the quorum sensing operator $\mathcal{F}_i = \mathcal{F}_{qs}$ (51). By applying the scalar product with a test function φ, we have:

$$\begin{aligned}\langle\frac{d\nabla\mathcal{F}_{qs}(s)}{ds}\cdot\delta s,\varphi\rangle &= \langle-\nabla\cdot\left(\frac{\sigma_{qs}}{q^2}(s\nabla\delta s+\nabla s\delta s)\right),\varphi\rangle\\ &=\langle-\nabla\cdot\left(\frac{\sigma_{qs}}{q^2}(s\nabla\varphi-\nabla s\varphi)\right)-\nabla\cdot\left(\frac{\sigma_{qs}}{q^2}\nabla s\right)\varphi,\delta s\rangle\\ &\quad+\int_0^T\int_{\partial\Omega}\frac{\sigma_{qs}}{q^2}(s\nabla\varphi\delta s-s\varphi\nabla\delta s+\nabla s\varphi\delta s)\cdot\vec{n}d\Gamma dt\end{aligned}$$

Therefore, the operator $\left(\frac{d\nabla\mathcal{F}_{qs}(s)}{ds}\right)^*$ is the operator such that

$$\varphi\mapsto-\nabla\cdot\left(\frac{\sigma_{qs}}{q^2}(s\nabla\varphi-\nabla s\varphi)\right)-\nabla\cdot\left(\frac{\sigma_{qs}}{q^2}\nabla s\right)\varphi\ \forall(x,y,t)\in\Omega\times[0;T]$$
$$\text{for }\varphi\text{ such that }\varphi=0\text{ and }\frac{\sigma_{qs}}{q^2}s\nabla\varphi\cdot\vec{n}=0\ \forall(x,y,t)\in\partial\Omega\times[0;T]$$

Non local pheromone sensing term. Here, we calculate the operator $\left(\frac{d\nabla\mathcal{F}_i(s)}{ds}\right)^*$ for the non-local pheromone sensing operator $\mathcal{F}_i = \mathcal{F}_{nlps}$ (52). By applying the scalar product with a test function φ, we have:

$$\begin{aligned}\langle\frac{d\nabla\mathcal{F}_{nlps}(s)}{ds}\cdot\delta s,\varphi\rangle &= \langle\nabla\cdot\left(\frac{\chi}{q}(s\nabla(g_{gp}*\delta s)+\nabla(g_{gp}*s)\delta s)\right),\varphi\rangle\\ &=\langle\overline{g}_{pg}*\nabla\cdot\left(\frac{\chi}{q}s\nabla\varphi\right)-\nabla\cdot\left(\frac{\chi}{q}\nabla(g_{gp}*s)\varphi\right)+\nabla\cdot\left(\frac{\chi}{q}\nabla(g_{gp}*s)\right)\varphi,\delta s\rangle\\ &\quad+\int_0^T\int_{\partial\Omega}\frac{\chi}{q}\left(s\varphi\nabla(g_{gp}*\delta s)-s\nabla\varphi(g_{gp}*\delta s)+\nabla(g_{gp}*s)\varphi\delta s\right)\cdot\vec{n}d\Gamma dt\end{aligned}$$

with $\overline{g}_{pg}$ the function such that $\overline{g}_{pg}(x,y,t)=g_{pg}(-x,-y,-t)$. Therefore, the operator $\left(\frac{d\nabla\mathcal{F}_{nlps}(s)}{ds}\right)^*$ is the operator such that

$$\varphi\mapsto\overline{g}_{pg}*\nabla\cdot\left(\frac{\chi}{q}s\nabla\varphi\right)-\nabla\cdot\left(\frac{\chi}{q}\nabla(g_{gp}*s)\varphi\right)+\nabla\cdot\left(\frac{\chi}{q}\nabla(g_{gp}*s)\right)\varphi\ \forall(x,y,t)\in\Omega\times[0;T]$$
$$\text{for }\varphi\text{ such that }\varphi=0\text{ and }\frac{\chi}{q}s\nabla\varphi\cdot\vec{n}=0\ \forall(x,y,t)\in\partial\Omega\times[0;T]$$

Logistic term. Here, we calculate the operator $\left(\frac{d\mathcal{R}_i(s)}{ds}\right)^*$ for the logistic operator $\mathcal{R}_i = \mathcal{R}_l$ (53). By applying the scalar product with a test function φ, we have:

$$\begin{aligned}\langle\frac{d\mathcal{R}_l(s)}{ds}\cdot\delta s,\varphi\rangle &= \langle\frac{1}{q^2}\frac{r}{\kappa}(\kappa q-2s)\delta s,\varphi\rangle\\ &=\langle\frac{1}{q^2}\frac{r}{\kappa}(\kappa q-2s)\varphi,\delta s\rangle\end{aligned}$$

Therefore, the operator $\left(\frac{d\mathcal{R}_l(s)}{ds}\right)^*$ is the operator such that

$$\varphi\mapsto\frac{1}{q^2}\frac{r}{\kappa}(\kappa q-2s)\varphi\ \forall(x,y,t)\in\Omega\times[0;T]$$

Allee effect term. Here, we calculate the operator $\left(\frac{d\mathcal{R}_i(s)}{ds}\right)^*$ for the Allee effect operator $\mathcal{R}_i = \mathcal{R}_{Ae}$ (54). By applying the scalar product with a test function φ, we have:

$$\begin{aligned}\langle\frac{d\mathcal{R}_{Ae}(s)}{ds}\cdot\delta s,\varphi\rangle &= \langle\frac{1}{q^3}\frac{r}{\kappa_+\kappa_-}(-3s^2+2q(\kappa_++\kappa_-)s-q^2\kappa_+\kappa_-)\delta s,\varphi\rangle\\ &=\langle\frac{1}{q^3}\frac{r}{\kappa_+\kappa_-}(-3s^2+2q(\kappa_++\kappa_-)s-q^2\kappa_+\kappa_-)\varphi,\delta s\rangle\end{aligned}$$

Therefore, the operator $\left(\frac{d\mathcal{R}_{Ae}(s)}{ds}\right)^*$ is the operator such that

$$\varphi \mapsto \frac{1}{q^3}\frac{r}{\kappa_+\kappa_-}(-3s^2 + 2q(\kappa_+ + \kappa_-)s - q^2\kappa_+\kappa_-)\varphi \; \forall (x, y, t) \in \Omega \times [0; T]$$

B.3.3. On the proximal gradient algorithm for the LASSO regularization term. As mentioned in Sec. 3.1, if the regularization terms j_{reg} are convex and differentiable, then the cost j is convex and differentiable and the optimization problem (3) is solved using a gradient descent algorithm. However, as mentioned in Secs 3.1 and 3.1.2, the differentiable framework can be extended to a non-smooth and subdifferentiable framework to consider LASSO-type regularization terms. In this context, the gradient descent algorithm can be extended to the proximal-gradient algorithm, also called the Forward-Backward algorithm.

In this section, the proximal-gradient algorithm is briefly presented in the context of the optimization problem (3) with several j_{reg} including the LASSO j_{reg} (24) and group-LASSO j_{reg} (25).

The proximal-gradient algorithm, see *e.g.* (Parikh et al., 2014) is an extension of the gradient descent algorithm used when the function to minimize can be split in $j(s) = g(s) + h(s)$ with:

- $g(s)$ a convex and differentiable function, which is in the present context the differentiable part of j, *i.e.* the observation term j_{obs} and the differentiable j_{reg},
- $h(s)$ a convex and subdifferentiable function, which is in the present case the LASSO or group-LASSO j_{reg}.

For such cost function, the optimum s_a is a fixed point $s_a = prox_{\lambda h}(s_a - \lambda\nabla g(s_a))$ with $prox_f$ the proximal operator applied to the function f and $\lambda > 0$ the gradient descent step size.

The proximal-gradient algorithm is then a fixed point algorithm based on the iteration loop:

$$s_{k+1} = prox_{\lambda h}(s_k - \lambda\nabla g(s_k))$$

In the present case, the gradient ∇g is computed using the expression (7). For the LASSO regularization term $h = \alpha_{reg,LASSO} j_{reg,LASSO}$ given by Eq. (24), the proximal operator $prox_{\lambda h}$ is

$$prox_{\lambda h}(s) = sign(s)max(0, |s| - \lambda\alpha_{reg,LASSO}).$$

For the group-LASSO regularization term $h = \alpha_{reg,gLASSO} j_{reg,gLASSO}$ given by Eq. (25), the proximal operator $prox_{\lambda h}$ is

$$prox_{\lambda h}(s) = sign(s)max\left(0, 1 - \frac{\lambda\alpha_{reg,gLASSO}}{\|s\|_{L^2(0;T)}}\right)s$$

with $\|s\|_{L^2(0;T)} = \left(\int_0^T s(t)^2 dt\right)^{1/2}$. As highlighted by the proximal operator $prox_{\lambda h}$, this algorithm acts as an iterative soft-threshold, shrinking to 0 the value of s lower than the threshold defined by $\lambda\alpha_{reg,LASSO}$ for LASSO and the value of s at places where the total amount $\|s\|_{L^2(0;T)}$ is lower than the threshold $\lambda\alpha_{reg,gLASSO}$ for group-LASSO. Consequently, the adaptation of the proximal-gradient algorithm to the LASSO regularization term is commonly called *Iterative Shrinkage-Threshold Algorithm* (ISTA).

B.4. Comparison of the LASSO and group-LASSO regularization terms in the Toy case

In this section, we briefly compare the LASSO (24) and group-LASSO (25) regularization terms in the Toy case presented in Sec. 2.3.2. Recall that the effect of the LASSO regularization term in the Toy case has already been highlighted in Sec. 3.4.

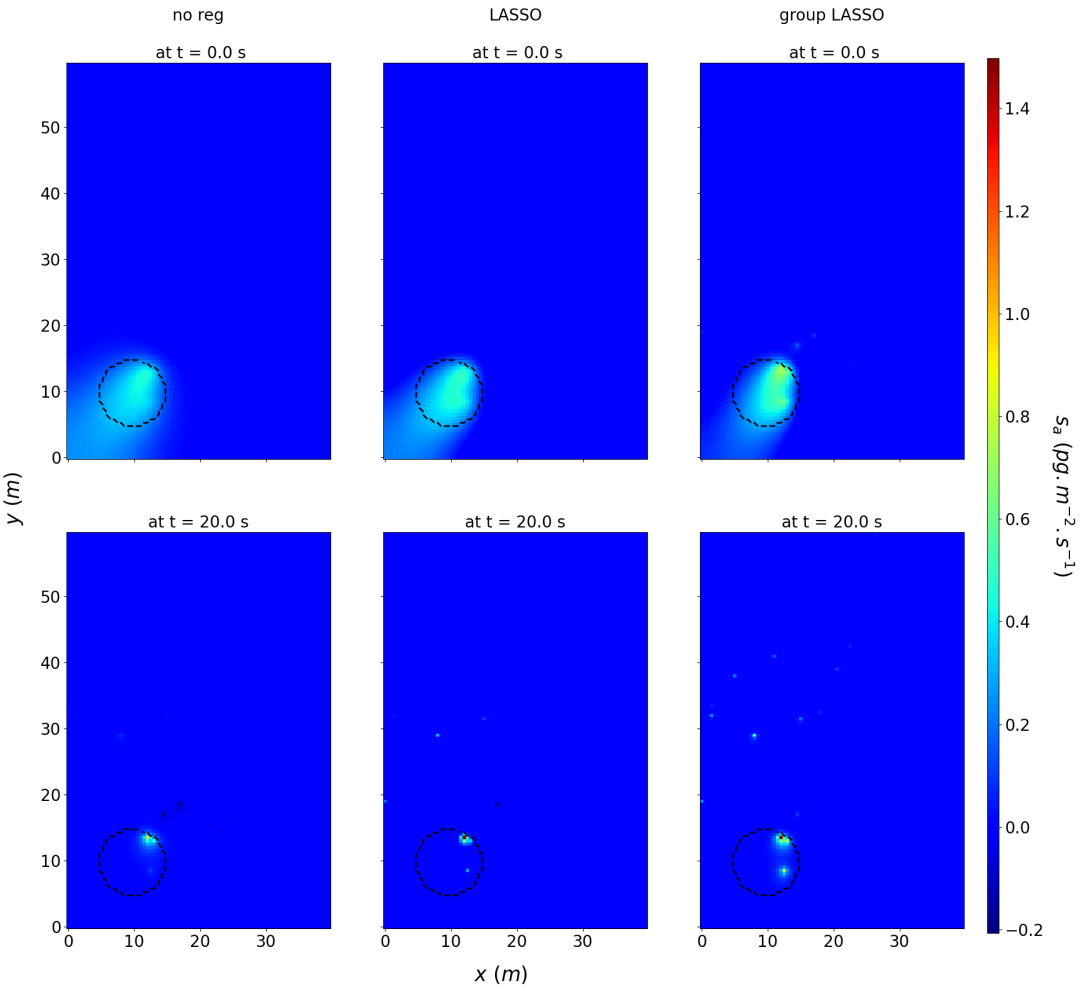


Figure 18 – Optimal sources estimated with different regularization terms. Optimal quantity of pheromone emitted s_a obtained by inference without regularization (left), with the LASSO regularization (24) (centre) and with the group-LASSO regularization (25) (right) for the Toy case. The optimal sources are indicated at $t = 0.0s$ (upper row) and $t = 20.0s$ (lower row).

B.4.1. Visual comparison of estimated pheromone emissions s_a. Recall that the LASSO regularization ('LASSO', centre of Fig. 18) implies that the retro-plume is less diffuse at both initial ($t = 0s$, top row) and final ($t = 20s$, bottom row) times. On the other hand, the group-LASSO ('group LASSO', right of Fig. 18) implies that the retro-plume is even less diffuse at initial time but more diffuse at final time. This is due to the fact that the group-LASSO regularization aims to threshold all the pheromone emissions at location where the total emissions are too low, unlike the LASSO regularization that shrinks all pheromone emissions that are too low. To compensate these retro-plumes less diffuse at initial time but more diffuse at final time, the pheromone emissions are higher at initial time but lower at final time with the group-LASSO regularization than with the LASSO regularization. Let us note that more spurious emissions far from the target support are recovered with the group-LASSO regularization than with the LASSO regularization.

B.4.2. RMSE criteria. Recall that with the LASSO regularization (pink curves on Fig. 19), the total error $RxMSE_{tot}$ is sensibly lower and the error on the target support $RxMSE_{supp}$ similar to the ones obtained without regularization (blue curve) at the beginning of the time window and up to $t \approx 17.5s$. After that time, with the LASSO regularization, both $RxMSE_{tot}$ and $RxMSE_{supp}$ significantly increase and become higher than without regularization. With the group-LASSO

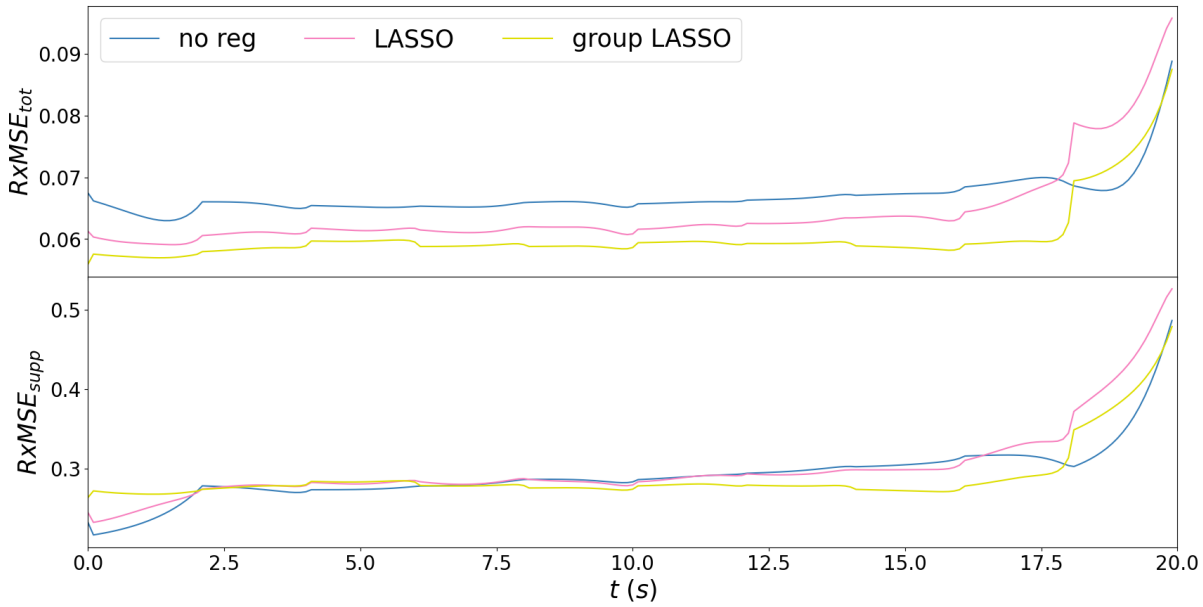


Figure 19 – Evolution of total and support RMSE of s_a against s_t. Time evolution of the root space-mean square errors between the optimum s_a obtained by inference with several regularization strategies and the target s_t, on the whole domain ($RxMSE_{tot}$, Eq. (14), upper panel) and on the target support ($RxMSE_{supp}$, Eq. (15), lower panel) in the Toy case: with no regularization term (no reg, blue), with the LASSO regularization term (24) (LASSO, pink) and with the group-LASSO regularization term (25) (group LASSO, yellow)

regularization (yellow curves on Fig. 19), the errors follow similar trend but are nearly always slightly lower. Let us note that, at the very beginning of the time window (between $t = 0s$ and $t = 2.5s$), $RxMSE_{supp}$ is higher with group-LASSO regularization than with LASSO regularization. This comes from the fact that the group-LASSO regularization thresholds all the emissions, including at the initial time, located at certain positions in the target supports.

B.4.3. Presence prediction, false discovery and miss rates. As already anticipated with the visual comparison, the presence prediction (upper row of Fig. 20) is less diffuse at initial time, similarly diffuse at middle time and more diffuse at final time with the group-LASSO regularization (yellow curves) than with the LASSO regularization (pink curve). Recall the with the LASSO regularization (pink curves), the miss rate ρ_{fa} and false discovery rate ρ_{fp} (bottom row of Fig. 20) are quite stable until $t \approx 17.5s$, time at which ρ_{fa} significantly increases. With the group-LASSO regularization (yellow curves), the trend is very similar, but two little difference can be noted. First, as observed on the errors, at the initial times, ρ_{fa} is slightly higher and ρ_{fp} is slightly lower with the group-LASSO regularization than with the LASSO regularization. With the LASSO regularization, ρ_{fa} is even null for a short amount of time, while it is always non-zero with the group-LASSO regularization. Also, one can note that ρ_{fa} significantly increases a bit later with the group-LASSO regularization than with the LASSO regularization.

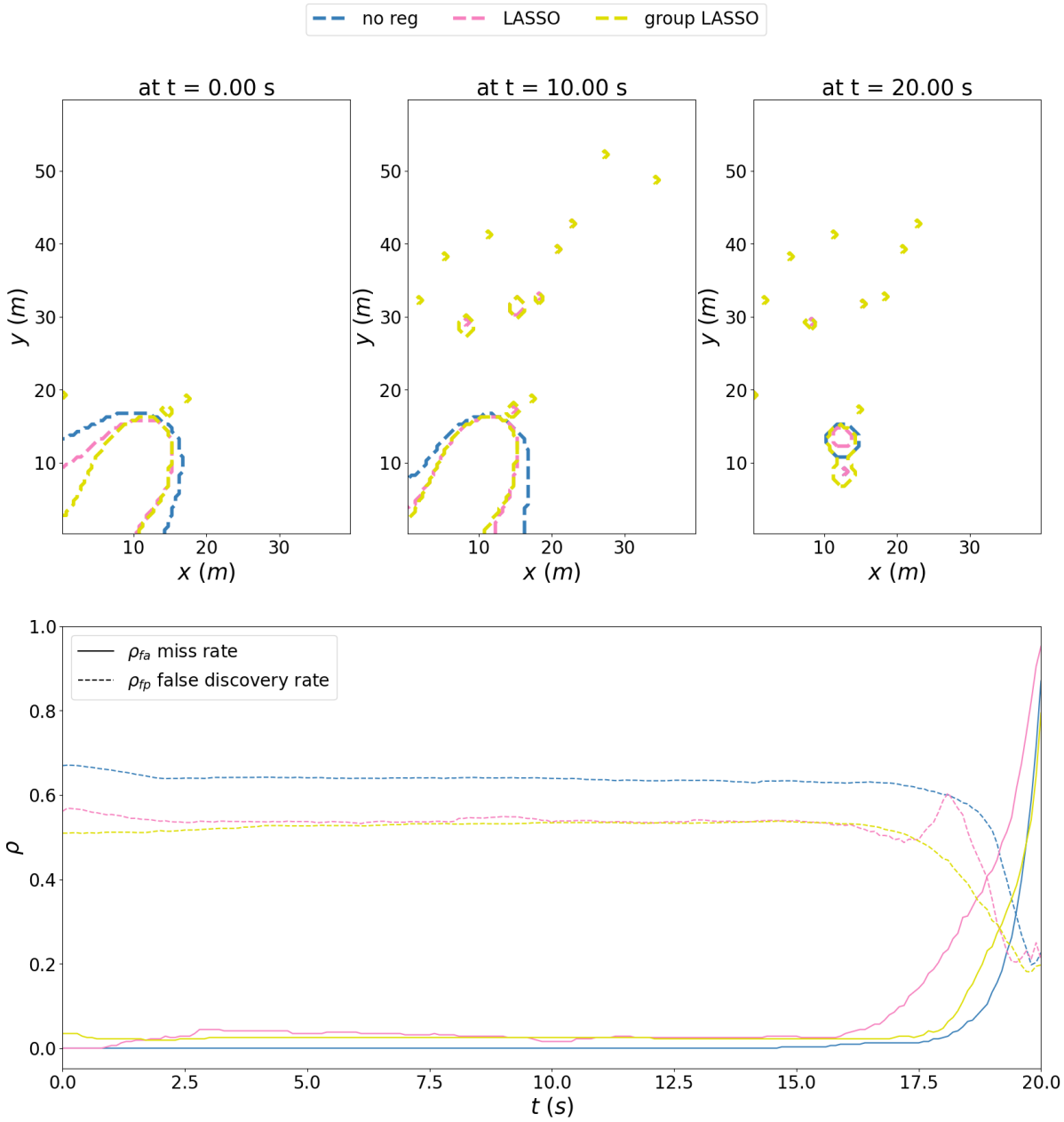


Figure 20 – Presence prediction and time evolution of false discovery and miss rates. Prediction of insects presence with several regularization strategies: with no regularization term (no reg, blue), with the LASSO regularization term (24) (LASSO, pink) and with the group-LASSO regularization term (25) (group LASSO, yellow). Top panel: map of presence prediction at three different time. The presence is obtained from the reconstructed source s_a by computing a level set with Eq. (16). The true presence (i.e. the target support) is indicated in black.
Lower panel. Time evolution of the miss rate ρ_{fa} (solid lines), see Eq. (17), and false discovery rate ρ_{fp} (dashed lines), Eq. (18).