Computing geographical networks generated by air-mass movement

—

Supporting Information

H. Richard[1], D. Martinetti[1], D. Lercier[2], Y. Fouillat[3], B. Hadi[4], M. Elkahky[4], J. Ding[4],
L. Michel[5], C. E. Morris[6], K. Berthier[6], F. Maupas[7], S. Soubeyrand[1]

[1]INRAE, BioSP, 84914 Avignon, France
[2]Autonomens, 32700 Lectoure, France
[3]Makina Corpus, 44000 Nantes, France
[4]Plant Production and Protection Division (NSP), Food and Agriculture Organization of the United Nations (FAO), Viale delle Terme di Caracalla, 00153 Rome, Italy
[5]INRAE, Plateforme ESV, BioSP, 84914 Avignon, France
[6]INRAE, Pathologie Végétale, F-84140, Montfavet, France
[7]ITB, 75008 Paris, France

September 29, 2023

Table of contents

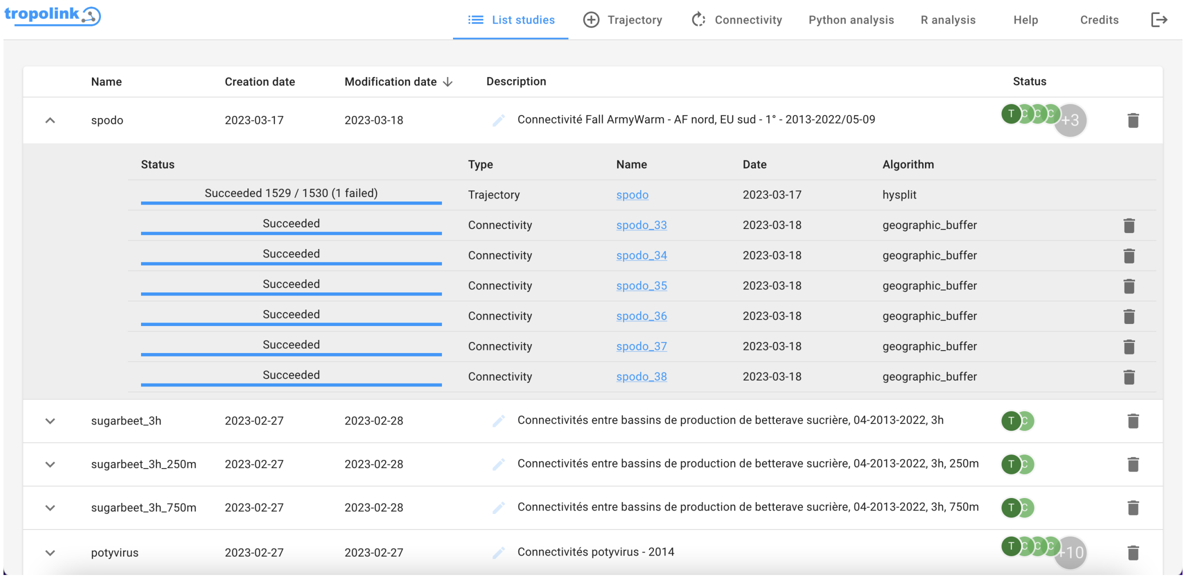
S1 How to use tropolink, step by step

This section progressively and practically describes the steps the user has to perform to run an analysis with `tropolink`. Table A1 lists input data and parameters that must be provided by the user. Note that meteorological data used as input by `HYSPLIT` have not to be provided.

Table A1: Input data and parameters provided by the user of `tropolink`.

Form section	Input/parameter	Type
Study initialization	Name of the study	Character string
	Description of the study	Character string
Trajectory	Starting/arrival dates	YYYY-MM-DD
	Starting/arrival locations	Code Name Lat Long Height
	Trajectory duration	Number b/n -144h and 144h
	Starting/arrival hour	Number b/n 0h and 23h UTC
	Top of model domain	Positive number (height in m)
	Vertical motion method	Character string in a list
Connectivity	Trajectory study	Character string in a list
	Buffer type	Character string in a list
	Buffer radius	Positive number (in km or °)

Logging in. The application, accessible at tropolink.fr, requires a login and a password obtained from the support team. Once logged in, the user can see a dashboard with the list of their previous studies/run, as illustrated by Figure A1.



Name	Creation date	Modification date ↓	Description	Status
spodo	2023-03-17	2023-03-18	Connectivité Fall ArmyWarm - AF nord, EU sud - 1° - 2013-2022/05-09	T C C C +3
sugarbeet_3h	2023-02-27	2023-02-28	Connectivités entre bassins de production de betterave sucrière, 04-2013-2022, 3h	T C
sugarbeet_3h_250m	2023-02-27	2023-02-28	Connectivités entre bassins de production de betterave sucrière, 04-2013-2022, 3h, 250m	T C
sugarbeet_3h_750m	2023-02-27	2023-02-28	Connectivités entre bassins de production de betterave sucrière, 04-2013-2022, 3h, 750m	T C
potyvirus	2023-02-27	2023-02-27	Connectivités potyvirus - 2014	T C C C +10

Status	Type	Name	Date	Algorithm
Succeeded 1529 / 1530 (1 failed)	Trajectory	spodo	2023-03-17	hysplit
Succeeded	Connectivity	spodo_33	2023-03-18	geographic_buffer
Succeeded	Connectivity	spodo_34	2023-03-18	geographic_buffer
Succeeded	Connectivity	spodo_35	2023-03-18	geographic_buffer
Succeeded	Connectivity	spodo_36	2023-03-18	geographic_buffer
Succeeded	Connectivity	spodo_37	2023-03-18	geographic_buffer
Succeeded	Connectivity	spodo_38	2023-03-18	geographic_buffer

Figure A1: Screenshot of the list of studies in the web application `tropolink`.

Creating a study. The user initiates a study by providing a name and a description in the trajectory tab (Figure A2). The study will be actually initiated when the trajectory calculation will be run. It will then appear in the dashboard listing the studies, from which its description can be updated or the study eventually removed.

Specifying trajectory calculation in one step. The user can specify a trajectory calculation in several ways via the form of the trajectory tab (Figure A2). The fastest way is to specify the trajectory calculation by loading the specifications of an existing study or by loading a `json` file containing the specifications. Some of the input can be modified if necessary (e.g., removing a date, or changing the trajectory duration) as soon as they are displayed on the same page by the application. The `json` file can be initially downloaded from an existing study (via the dashboard listing the existing studies) and edited with any text editor prior to uploading. It can also be produced independently using the appropriate format.

Specifying trajectory calculation sequentially. Alternatively, the user can specify each of the input required for trajectory calculation (Figure A2), namely the dates, the locations and the `HYSPLIT` tuning parameters. The specification of dates (of arrival or departure of trajectories) as well as geographical locations (of arrival or departure of trajectories) is quite intuitive: they can be either uploaded with well formatted files or directly provided in the form. Specifying the `HYSPLIT` tuning parameters consists in selecting a trajectory duration (negative values for the backward trajectories, positive ones for forward trajectories, -24h by default), a starting or arrival hour in universal time (12h UTC by default), an altitude limit beyond which the trajectories are no longer calculated (10,000 m by default) and the vertical motion method.

Running the trajectory calculation. When the above-mentioned specifications have been made, the calculation of trajectories can be launched by clicking the 'RUN' button at the bottom of the page.

Specifying connectivity calculation. The calculation of the connectivity matrix for a given set of trajectories is specified by selecting the corresponding trajectory study (via a drop-down menu listing all the available studies), selecting the method used to define the buffer zone around geographical locations and the 'radius' of the buffer (in km for circular buffers or the half angle of the sector for geographic latitude-longitude buffers); see Figure A3.

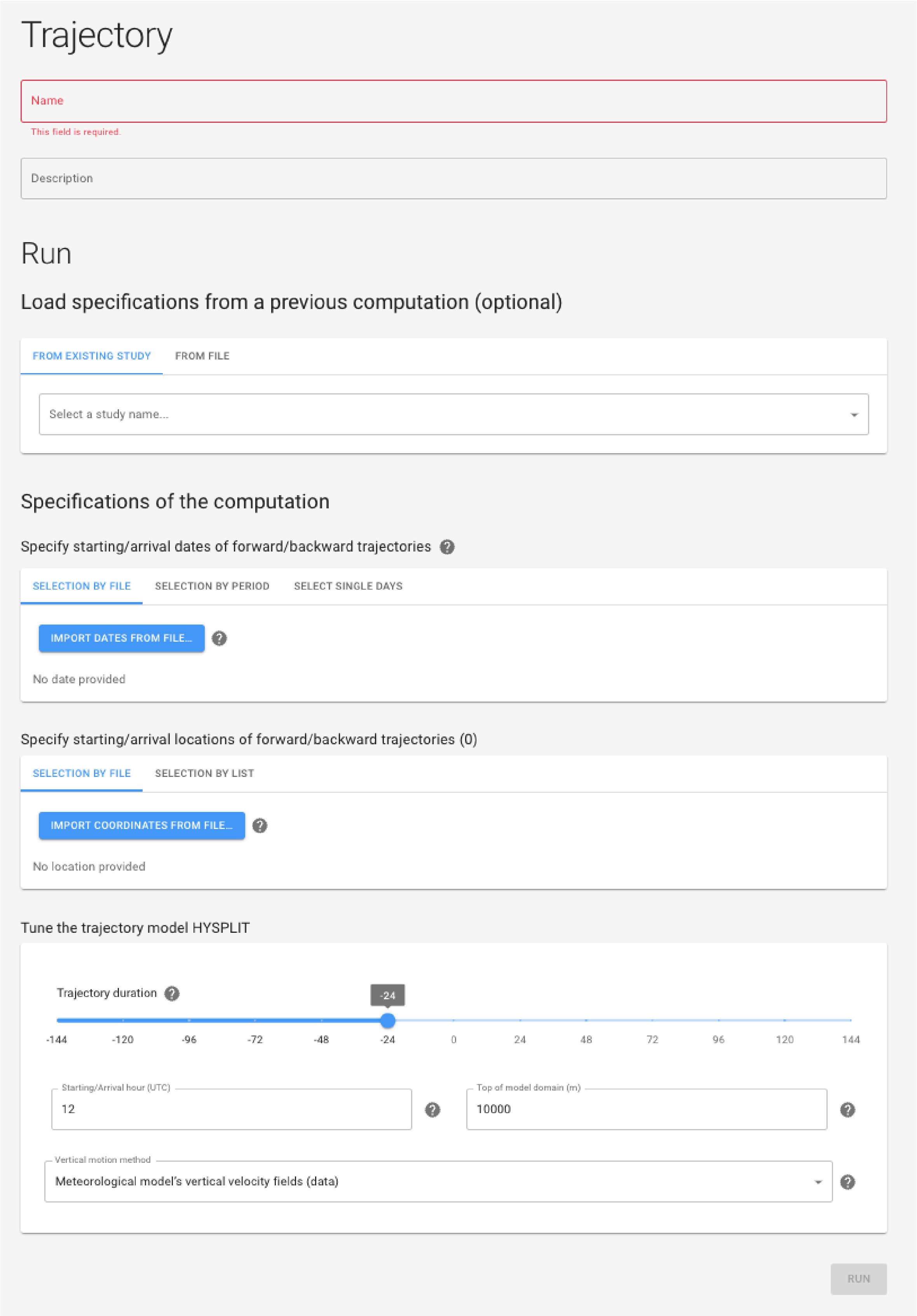
The user can also specify the connectivity calculation in one step by loading a `json` file containing the specifications, and modifying some of the input if necessary as soon as they are displayed on the same page by the application. The `json` file can be initially downloaded from an existing connectivity study (via the dashboard listing the existing studies) and edited with any text editor prior to uploading. It can also be produced independently using the appropriate format. The use of a `json` file is in particular useful when the user want to select some dates and locations among a long list of dates and locations.

Running the connectivity calculation. When the above-mentioned specifications have been made, the calculation of connectivities can be launched by clicking the ‘RUN’ button at the bottom of the page.

Analyzing trajectories and connectivities in a notebook. The user can analyze trajectories and connectivities with `Python` or `R` language within a `Jupyter` notebook by selecting the corresponding tab. Each of the two notebooks provides a specific `tropolink`-adequate environment and codes for starting the analysis. The codes especially help the user to access and display the result objects of trajectory and connectivity calculations. They propose four steps:

- Importing the required libraries;
- Loading the study of interest;
- Displaying air mass trajectories for a given date;
- Plotting the connectivities and comparing them with geographical distances.

Visualizing studies and downloading specifications and results. The users can visualize each of their studies in a dedicated dashboard. Each study (either a trajectory calculation or a connectivity calculation) can be expanded from the main dashboard listing the studies to see the underlying specifications (dates, locations and other input parameters). The specifications can be downloaded in `json` format (for re-use or to get a template). The calculation output can also be downloaded in diverse formats: original `HYSPLIT` text format for trajectories; `json`, `geojson` or `csv` format for connectivity matrices.


Trajectory

Name

This field is required.

Description

Run

Load specifications from a previous computation (optional)

FROM EXISTING STUDY FROM FILE

Select a study name...

Specifications of the computation

Specify starting/arrival dates of forward/backward trajectories

SELECTION BY FILE SELECTION BY PERIOD SELECT SINGLE DAYS

IMPORT DATES FROM FILE...

No date provided

Specify starting/arrival locations of forward/backward trajectories (0)

SELECTION BY FILE SELECTION BY LIST

IMPORT COORDINATES FROM FILE...

No location provided

Tune the trajectory model HYSPLIT

Trajectory duration: -24

Starting/Arrival hour (UTC): 12

Top of model domain (m): 10000

Vertical motion method: Meteorological model's vertical velocity fields (data)

Figure A2: Screenshot of the form to be filled in the trajectory tab of the web application `tropolink`.

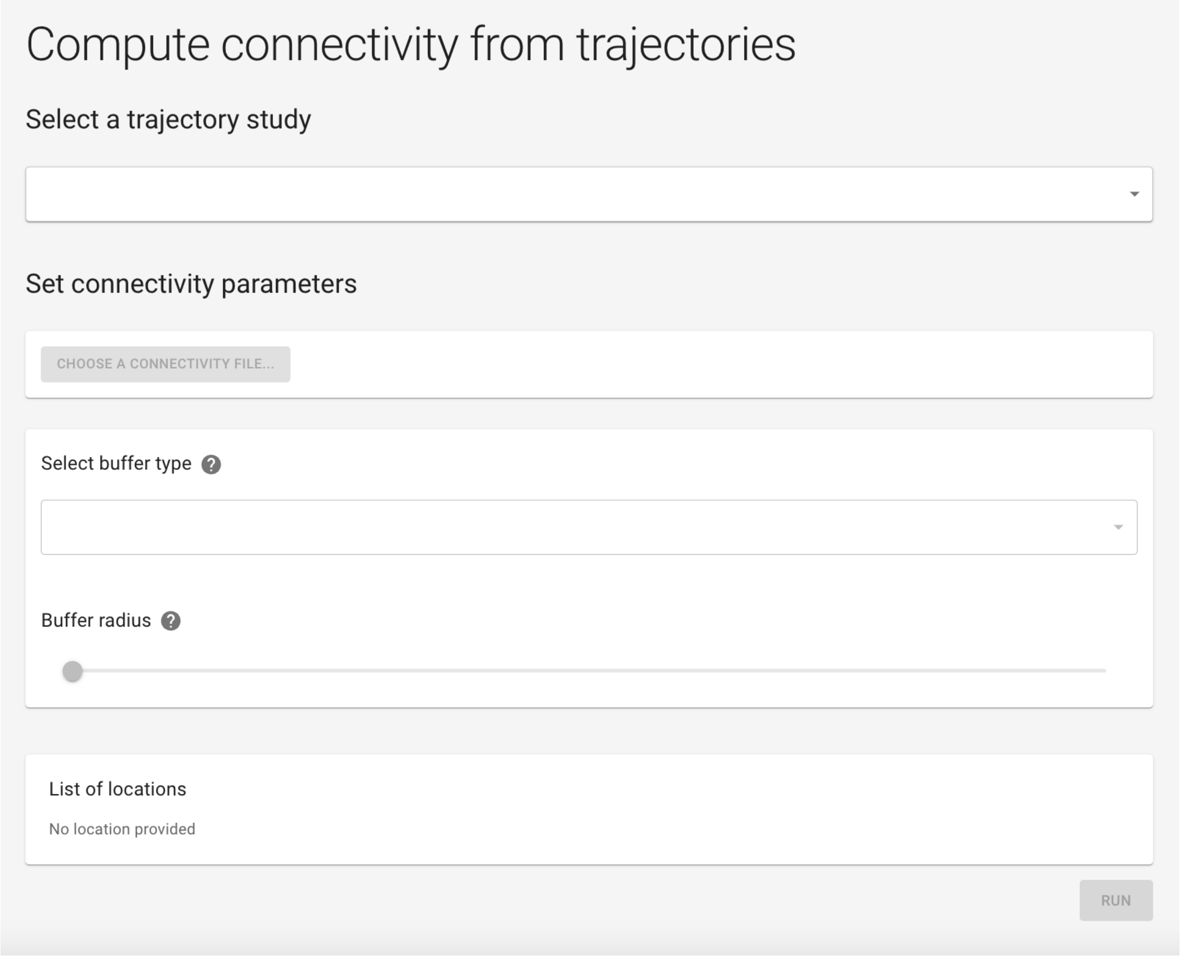


Figure A3: Screenshot of the form to be filled in the connectivity tab of the web application `tropolink`.

S2 Conceptual description of tropolink

S2.1 Technical stack

The application is made up of five components represented schematically in Figure B1:

- a `React` frontend application;
- a `Django` backend application;
- a `JupyterHub` server;
- an OpenID Connect authentication server (`Keycloak`);
- a reverse-proxy that links these elements (`HAProxy`).

The heart of the application is a `Django` API, which is especially in charge of performing trajectory calculations on a `Slurm`-based computer cluster. The management of the different application tasks is done using `Celery` (worker management for `Python` tasks) and `RabitMQ` (messaging broker).

User management is done using `Keycloack` (based on the openID Connect protocol). All the studies carried out are saved in a `PostgreSQL` database with the aim of reusing the calculations in a virtuous spirit (and for saving costs). Finally, the generated objects (such as the connectivity matrices) are provided to the user in a `Jupyter` environment embedding either `Python` or `R`, depending on the user's preference.

S2.2 Software modules

The application is developed in three modules corresponding to the three main groups of functions : 'API', 'Front' and 'Notebook' modules. These modules are governed by an open-source license; they are hosted by the GitLab server operated by the MathNum department of INRAE (forgeMIA). The documentation is available under the project wiki.

S2.3 API Module

The API module contains the business functions of the application. These different functions are distributed by the `Celery` asynchronous task manager via the `RabbitMQ` messenger.

The first group of functions contains everything needed to calculate air mass trajectories with `HYSPLIT`. This mainly consists of preparing input information and codes in the appropriate format and then executing the `HYSPLIT` calculation scripts on the computer cluster to which `tropolink` is associated. This also includes functions for monitoring and controlling the proper execution of calculations by the computer cluster. This stage is fully automated, which saves valuable time while guaranteeing a high level of script robustness and security.

The purpose of the second group of functions is to calculate the connectivity matrices from the trajectories of air masses previously computed. To do this, a subset is dedicated to the methods for determining the buffer zones that the trajectories are possibly intersecting. The results are saved in a global object.

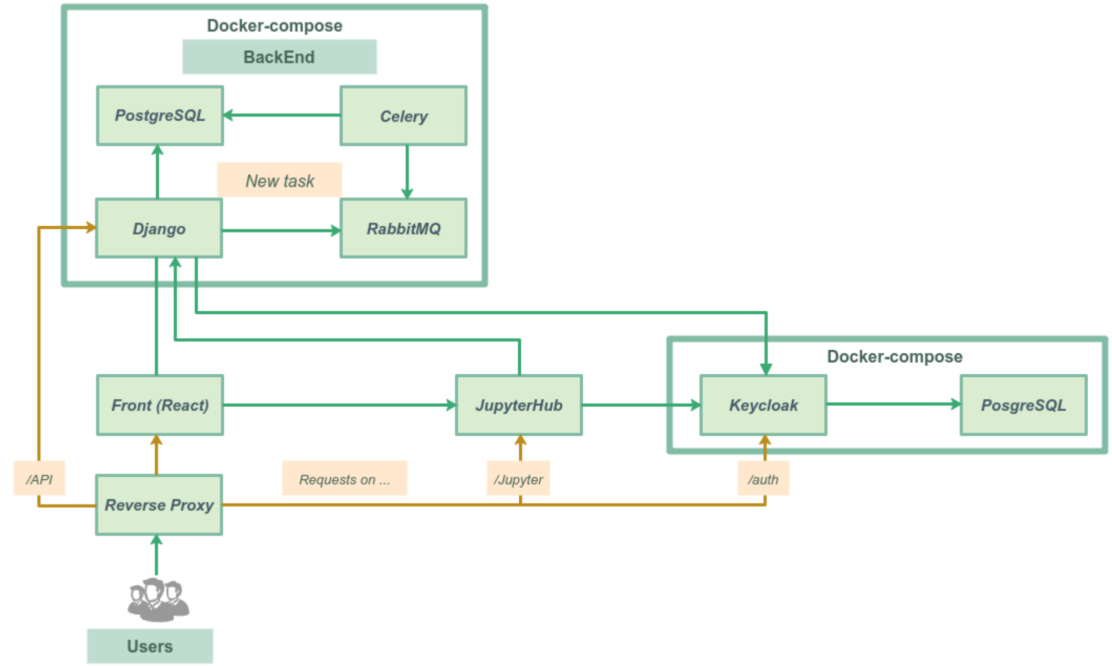


Figure B1: `tropolink` technical stack scheme.

S2.4 Front Module

The Front module has the task of managing the Web interface of the application. It is developed with the `React JS` library. The interface is designed to correspond to the mental scheme of the user when he wishes to carry out a `tropolink` study, so each stage of the study corresponds to a tab. The main tab which plays the pivotal role of the interface is the dashboard tab. In this tab the user can manage and monitor the studies already carried out or in progress. The main functions of this module are:

- managing user authentication;
- managing the display of web pages (dashboard, folding panels, ...);
- controlling and routing requests triggered by the user.

S2.5 Notebook module

The last module provides a suitable environment for the final objective of the user when he runs `tropolink`: analyzing spatial connectivities generated by air mass movement. This environment obviously relies on all the previous steps, which would be tedious if performed manually. It was built from `Jupyter` for the following reasons:

- `Jupyter` fits into the logic of a web application by offering a rich workshop without specific installation on the workstation;
- It offers an interface that allows the user to associate interactive script areas with text, graphics, etc.;
- There is a wide choice of kernels (i.e., command interpreters), which allows the user to develop his code according to his preferences or skills.

S3 Justification of tropolink specifications in the examples

Tables C1–C4 give justifications of `tropolink` specifications used in the four case studies. For the studies about sugar beet and potyvirus, we used characteristics of the aphid *Myzus persicae* as references, when available. For the fall armyworm study, we used information available for this insect. For the *Puccinia* pathway, we used information provided by Aylor (2003) and Radici et al. (2022).

Table C1: Justification of the specifications of `tropolink` for running the sugar beet application.

Input information	Specification	Reference
Dates	Each day of April from 2013 to 2022	In France, *M. persicae* has two flying periods in the year, the important one for sugar beet yellows transmission being spring (Luquet, 2022)
Starting/arrival altitude	500m	*M. persicae* typically flies at altitudes from 10m to 2000m (Irwin et al., 2007)
Trajectory duration	3h	Aphids are expected to fly up to 3h (Johnson, 1969; Parry, 2013)
Starting/arrival hour	14h UTC	*M. persicae* preferentially flies around mid day (Boiteau, 1986)
Buffer type	Circular	Arbitrary choice
Buffer radius	23km	Half of the 0.8-quantile of pairwise distances between fields in the same cluster

Table C2: Justification of the specifications of `tropolink` for running the potyvirus application.

Input information	Specification	Reference
Dates	Each day of month m in 2014 (for m =Jan.,. . .,Dec.)	All months considered to test multiple hypotheses
Starting/arrival altitude	500m	*M. persicae* typically flies at altitudes from 10m to 2000m (Irwin et al., 2007)
Trajectory duration	-3h	Aphids are expected to fly up to 3h (Johnson, 1969; Parry, 2013)
Starting/arrival hour	14h UTC	*M. persicae* preferentially flies around mid day (Boiteau, 1986)
Buffer type	Circular	Arbitrary choice
Buffer radius	1km	≈ 1/50 of the maximum inter-population distance

Table C3: Justification of the specifications of `tropolink` for running the fall armyworm (FAW) application. *Adult FAW are considered to be able to fly up to three consecutive nights (Westbrook et al., 2016; Wu et al., 2022). Here, we do not account for such a possibility, but our approach could be adapted to compute connectivities over several nights.

Input information	Specification	Reference
Dates	Each day of May–Sept. from 2013 to 2022	Sampling in a country such as France is organized during the risk period from May to September (Plateforme ESV, 2022)
Starting/arrival altitude	1000m	Fight altitudes of FAW adults ranges from 500m to 2250m (Wu et al., 2022)
Trajectory duration	12h*	FAW adults fly over long distances during the
Starting/arrival hour	20h UTC	night, from dusk to dawn (Westbrook et al., 2016)
Buffer type	Geographic	Allow the complete coverage of the land surface
Buffer radius	0.5°	Trade-off between high resolution and limited number of network points

Table C4: Justification of the specifications of `tropolink` for running the *Puccinia* pathway application

Input information	Specification	Reference
Dates	Each day of April–July from 2019 to 2023	Period of onset of wheat stem rust from the South to the North of the western Mississippi basin (Aylor, 2003; Hamilton and Stakman, 1967)
Starting/arrival altitude	200m	Median altitude used by Radici et al. (2022)
Trajectory duration	48h	Value typically considered for spore survival (Aylor, 2003)
Starting/arrival hour	17h UTC	i.e. 12h local time; arbitrary choice
Buffer type	Circular	Arbitrary choice
Buffer radius	25km	Arbitrary choice

S4 Additional figures

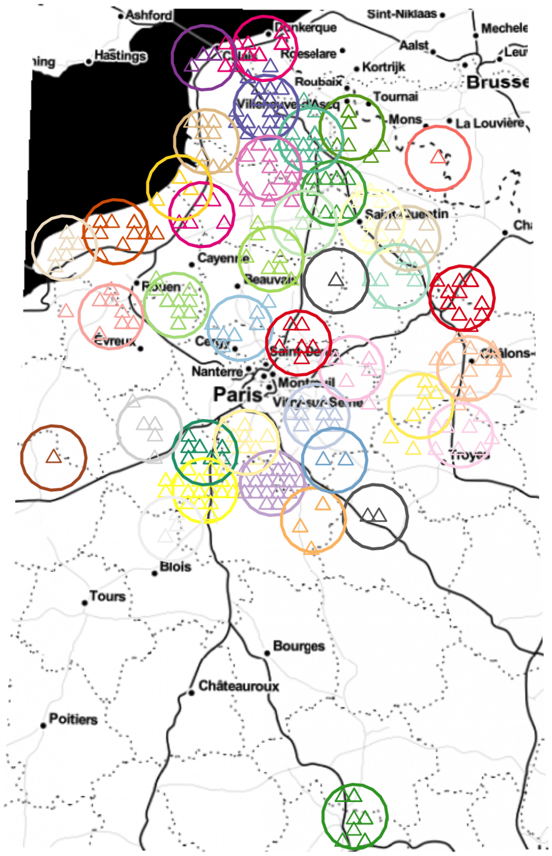


Figure S1: Hierarchical clustering of sugar beet fields based on the Euclidean distance between field coordinates in the Lambert-93 projection, which is a standard projection for displaying geographical distances in France. Locations of fields, which measure from a few hectares to about 20 hectares each, are indicated by triangles, circular buffers with 23km radius are indicated by circles (circle sizes effectively represent their extent on the map). Different colors (for triangles and circles) indicate different clusters.

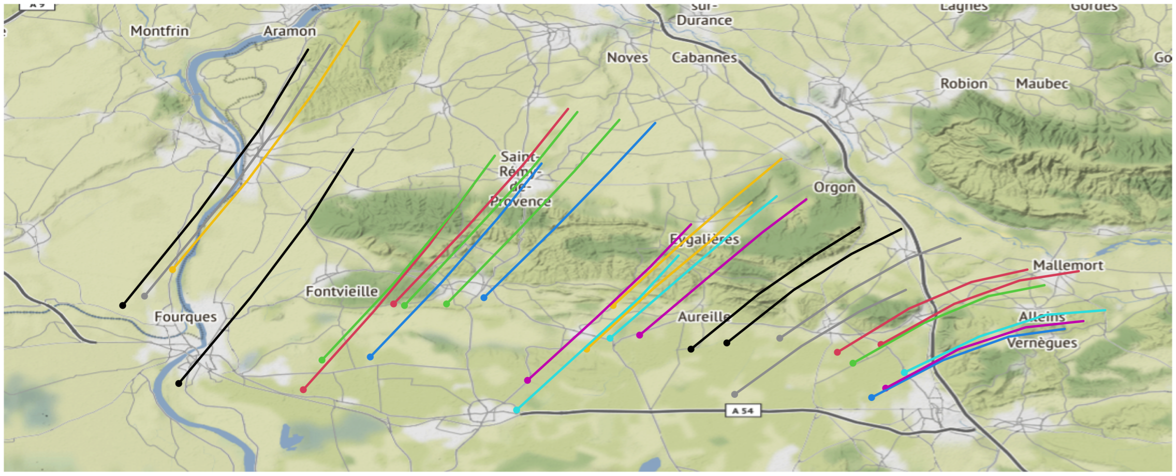


Figure S2: Trajectories computed on 2014-01-12 for the potyvirus example. Colors were randomly attributed to starting points.

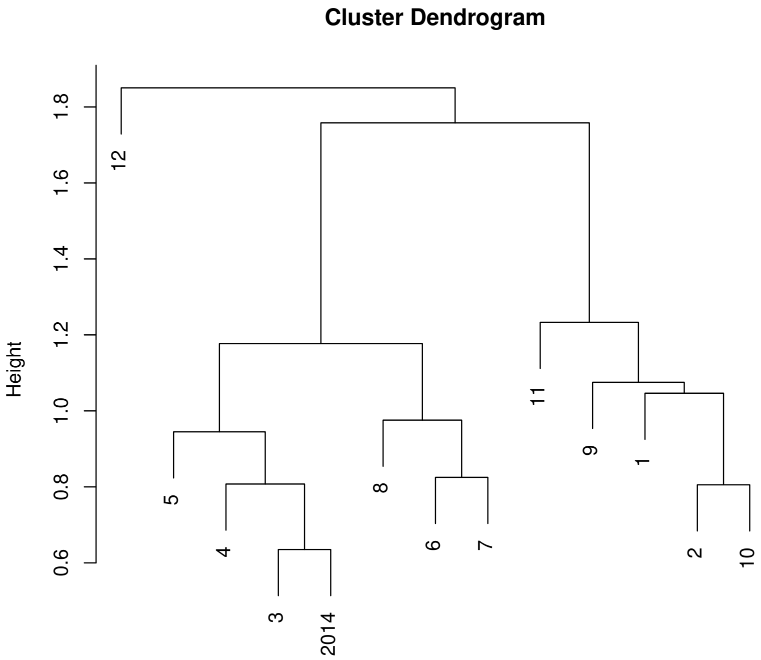


Figure S3: Dendogram corresponding to the hierarchical clustering of the monthly and the annual connectivity matrices computed in the potyvirus example. The label 1 stands for the connectivity matrix of Jan. 2014, 2 for the connectivity matrix of Feb. 2014 and so on; the label 2014 stands for the connectivity matrix of the year 2014. The clustering was performed with the default method implemented in the `hclust` function in `R` and using the Euclidean distance between non diagonal terms of the connectivity matrices.

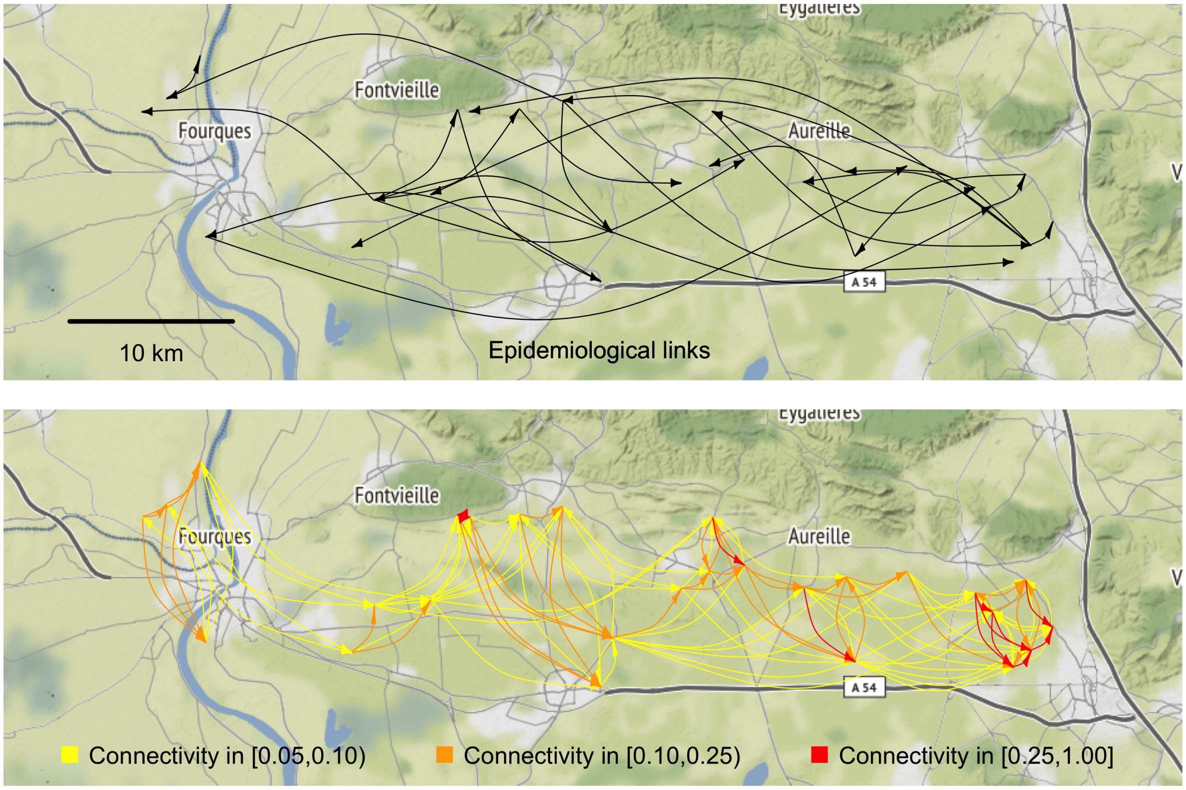


Figure S4: Application of air mass connectivity to potyvirus epidemiology: Epidemiological links between the 27 potyvirus populations inferred by Alamil et al. (2019) from genomic data (top panel; the links are oriented from the main putative sources for each population), and tropospheric connectivities obtained from `tropolink` for August 2014 (bottom panel; connectivities lower than 0.05 are not plotted). In comparison to Figure 5 (bottom panel) in the main text, we see more East – West links in both directions, which are more consistent with the epidemiological links and explains the relatively low p-value obtained for August.

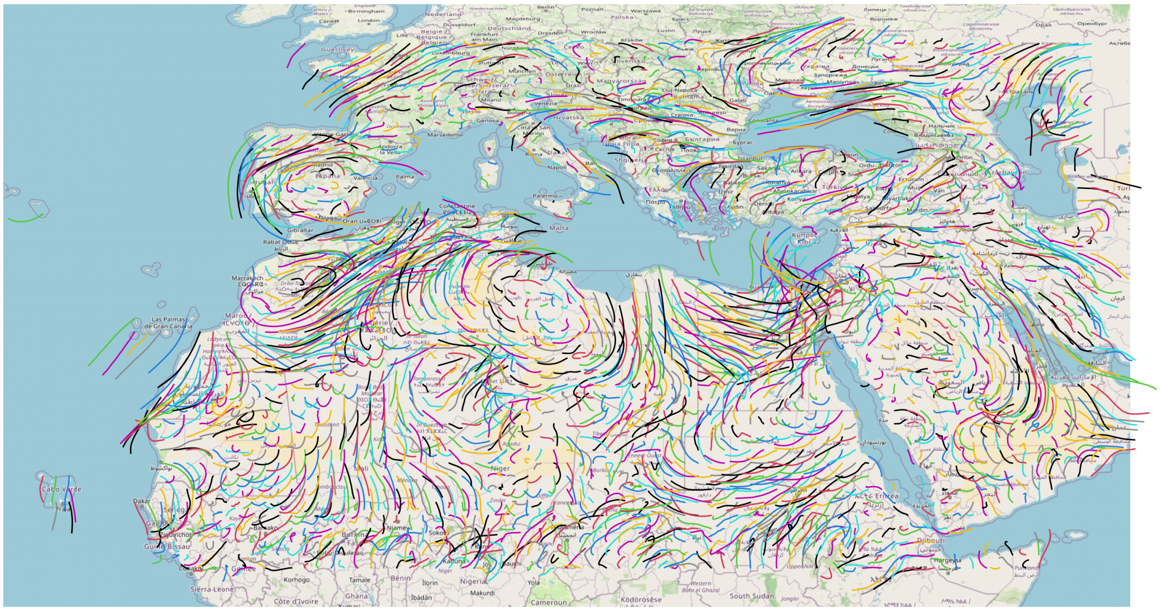


Figure S5: Trajectories computed on 2022-05-02 for the fall armyworm example. Colors were randomly attributed to starting points.

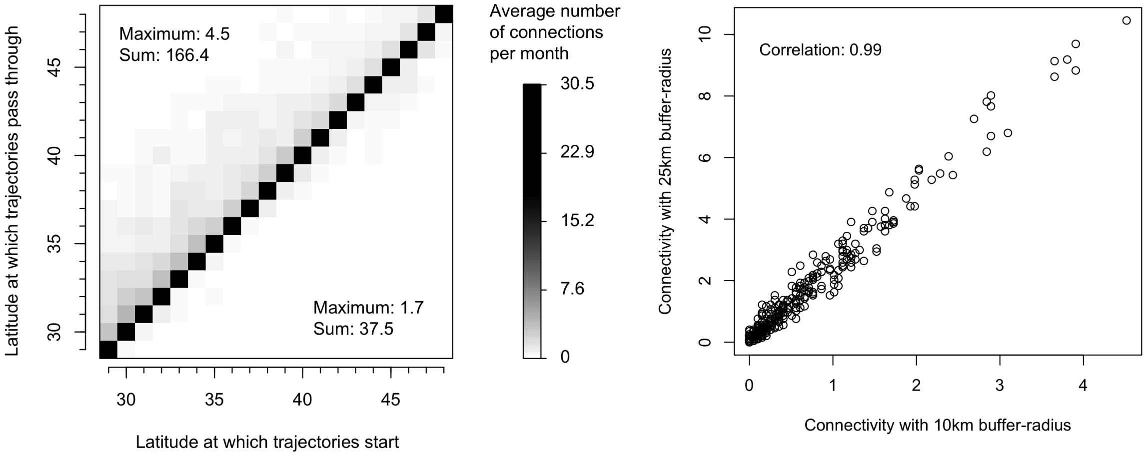


Figure S6: *Puccinia* pathway application: representation of the connectivity matrix between sites along the 97° West longitude when the buffer radius around sites is 10km instead of 25km (left) and cloud of points giving connectivities computed with 25km buffer-radius versus connectivities computed with 10km buffer-radius (right). Connectivity is expressed in terms of average number of pairwise connections per month, whose theoretical maximum value is 30.5, i.e., the mean number of days per month over the study period. Numbers within the plot on the left-hand-side give the maximum and the sum (averaged over one month) of northward connections (top left) and southward connections (bottom right).

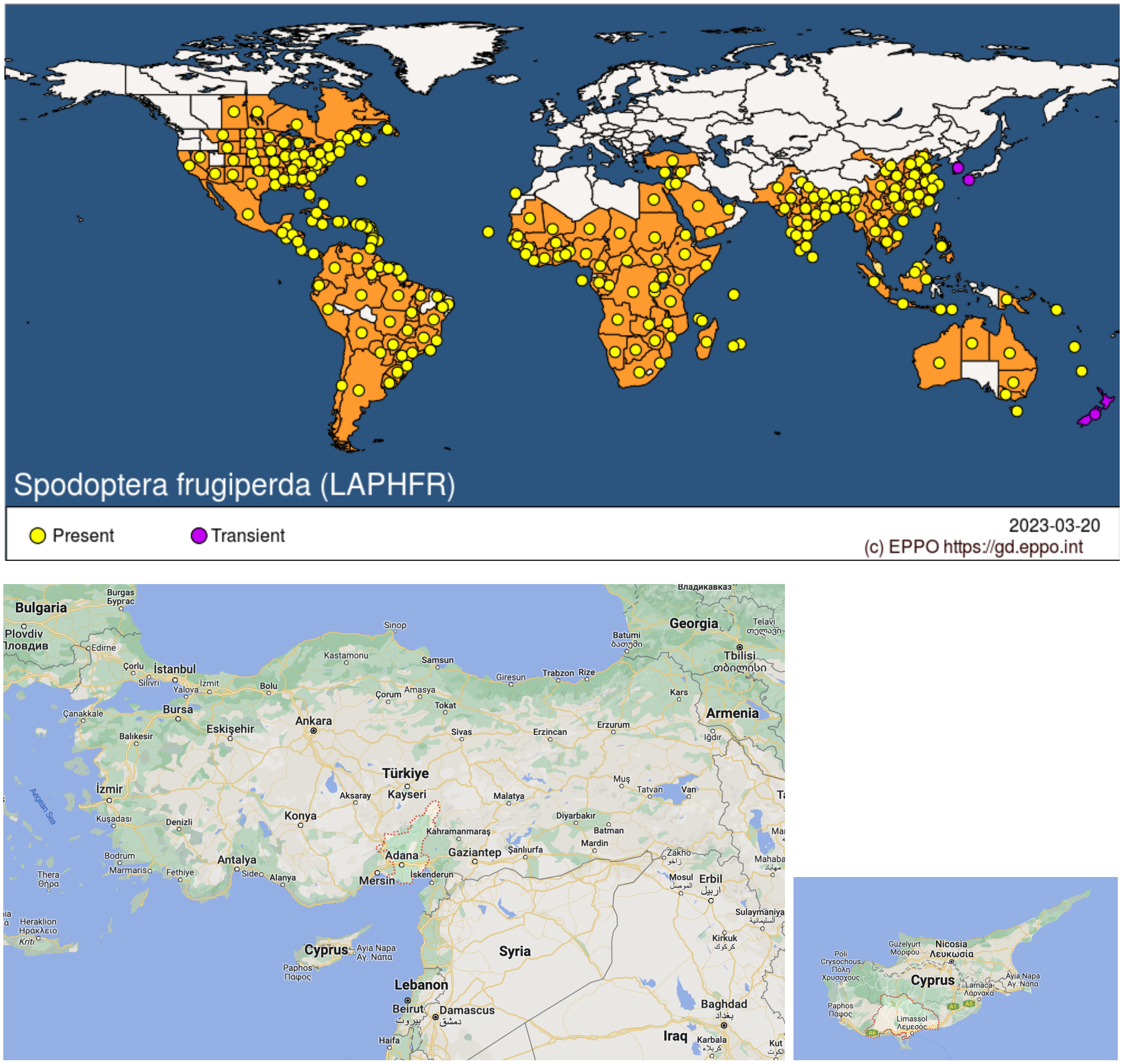


Figure S7: Top: Spatial distribution of *Spodoptera frugiperda* (fall armyworm) provided by the European and Mediterranean Plant Protection Organization (EPPO) on 2023-03-02; see original source at `https://gd.eppo.int/taxon/LAPHFR/distribution`. Bottom: Location of the Adana province in Türkiye and the Limassol district in Cyprus.

References

Alamil, M., Hughes, J., Berthier, K., Desbiez, C., Thébaud, G., and Soubeyrand, S. (2019). Inferring epidemiological links from deep sequencing data: a statistical learning approach for human, animal and plant diseases. *Philosophical Transactions of the Royal Society B*, 374:20180258.

Aylor, D. E. (2003). Spread of plant disease on a continental scale: role of aerial dispersal of pathogens. *Ecology*, 84:1989–1997.

Boiteau, G. (1986). Diurnal flight periodicities and temperature thresholds for three potato-colonizing aphids (Homoptera: Aphididae) in New Brunswick. *Annals of the Entomological Society of America*, 79:989–993.

Hamilton, L. M. and Stakman, E. C. (1967). Time of stem rust appearance on wheat in the Western Mississippi Basin in relation to the development of epidemics from 1921 to 1962. *Phytopathology*, 57:609–614.

Irwin, M. E., Kampmeier, G. E., and Weisser, W. W. (2007). Aphid movement: process and consequences. In van Emden, H. F. and Harrington, R., editors, *Aphids as Crop Pests*, pages 153–186. CABI, Wallingford.

Johnson, C. G. (1969). *Migration and dispersal of insects by flight.* Methuen & Co. Ltd., London.

Luquet, M. (2022). Synthèse bibliographique du cycle de vie des pucerons et de leurs capacités de vol pour la définition des filtres biologiques à considérer dans le cadre de la modélisation des flux de masses d'air. Technical report, IGEPP, INRAE, Institut Agro, Univ Rennes, 49045 Angers, France.

Parry, H. R. (2013). Cereal aphid movement: general principles and simulation modelling. *Movement Ecology*, 1:1–15.

Plateforme ESV (2022). Fiche de reconnaissance SORE *Spodoptera frugiperda* `https://fichesdiag.plateforme-esv.fr/fiches/Fiche_Diagnostique_LAPHFR_Spodoptera_frugiperda.pdf`. pages 1–4.

Radici, A., Martinetti, D., and Bevacqua, D. (2022). Early-detection surveillance for stem rust of wheat: insights from a global epidemic network based on airborne connectivity and host phenology. *Environmental Research Letters*, 17:064045.

Westbrook, J., Nagoshi, R., Meagher, R., Fleischer, S., and Jairam, S. (2016). Modeling seasonal migration of fall armyworm moths. *International Journal of Biometeorology*, 60:255–267.

Wu, M.-F., Qi, G.-J., Chen, H., Ma, J., Liu, J., Jiang, Y.-Y., Lee, G.-S., Otuka, A., and Hu, G. (2022). Overseas immigration of fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), invading Korea and Japan in 2019. *Insect Science*, 29:505–520.