Computing geographical networks generated by air-mass movement

Hervé Richard, Davide Martinetti, D. Lercier, Y. Fouillat, B. Hadi, M. Elkahky, J. Ding, Lucie Michel, Cindy E. Morris, K. Berthier, et al.

GeoHealth

RESEARCH ARTICLE
10.1029/2023GH000885

Key Points:

- The movement of air masses within the troposphere generate air highway connecting distant areas from local to global scales
- Our approach and the accompanying web application tropolink quantifie how distant locations are connected by air-mass movement
- Our approach can be used to investigate the spread of small biological particles and organisms

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Soubeyrand,
samuel.soubeyrand@inrae.fr

Citation:

Richard, H., Martinetti, D., Lercier, D., Fouillat, Y., Hadi, B., Elkahky, M., et al. (2023). Computing geographical networks generated by air-mass movement. *GeoHealth*, *7*, e2023GH000885. https://doi.org/10.1029/2023GH000885

Received 13 JUN 2023
Accepted 22 SEP 2023

Author Contributions:

Conceptualization: H. Richard, S. Soubeyrand
Data curation: M. Elkahky, J. Ding, S. Soubeyrand
Formal analysis: S. Soubeyrand
Investigation: B. Hadi, M. Elkahky, J. Ding, L. Michel, C. E. Morris, K. Berthier, F. Maupas, S. Soubeyrand
Methodology: H. Richard, D. Martinetti, C. E. Morris, S. Soubeyrand
Software: H. Richard, D. Lercier, Y. Fouillat, S. Soubeyrand
Supervision: H. Richard, S. Soubeyrand

Computing Geographical Networks Generated by Air-Mass Movement

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], and S. Soubeyrand[1]

[1]INRAE, BioSP, Avignon, France, [2]Autonomens, Lectoure, France, [3]Makina Corpus, Nantes, France, [4]Plant Production and Protection Division (NSP), Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, [5]Plateforme ESV, INRAE, BioSP, Avignon, France, [6]INRAE, Pathologie Végétale, Montfavet, France, [7]ITB, Paris, France

Abstract As air masses move within the troposphere, they transport a multitude of components including gases and particles such as pollen and microorganisms. These movements generate atmospheric highways that connect geographic areas at distant, local, and global scales that particles can ride depending on their aerodynamic properties and their reaction to environmental conditions. In this article we present an approach and an accompanying web application called `tropolink` for measuring the extent to which distant locations are potentially connected by air-mass movement. This approach is based on the computation of trajectories of air masses with the `HYSPLIT` atmospheric transport and dispersion model, and on the computation of connection frequencies, called connectivities, in the purpose of building trajectory-based geographical networks. It is illustrated for different spatial and temporal scales with three case studies related to plant epidemiology. The web application that we designed allows the user to easily perform intensive computation and mobilize massive archived gridded meteorological data to build weighted directed networks. The analysis of such networks allowed us for example, to describe the potential of invasion of a migratory pest beyond its actual distribution. Our approach could also be used to compute geographical networks generated by air-mass movement for diverse application domains, for example, to assess long-term risk of spread from persistent or recurrent sources of pollutants, including wildfire smoke.

Plain Language Summary The movement of air masses within the atmosphere generates air highways connecting distant areas at the local, continental, and global scales. In this article we present an approach and an accompanying web application for measuring how distant locations are potentially connected by air-mass movement, that is, for representing the geographic networks generated by the above-mentioned air highways. The approach is illustrated for different spatial and temporal scales with three case studies related to plant epidemiology; one of them concerns the potential invasion routes of a migratory pest. The web application that we developed allows the user to easily perform such studies. It could also be used to compute geographical networks generated by air-mass movement for other application domains, for example, to assess long-term risk of spread from sources of pollutants.

1. Introduction

Earth's troposphere is characterized by large-scale circulation patterns (Schneider, 2006). The movement of air masses in the troposphere transport moisture and particles from natural and anthropogenic sources, such as volcanic ash (Weber et al., 2012), dust (Gonzalez-Martin et al., 2014), pollen (Kappen & Straka, 1988), and radioactive chemical elements (Masson et al., 2019). The trajectories of some of these particles, when emitted into the atmosphere in massive quantities, can be visualized with satellites (Kellogg & Griffin, 2006). For more inconspicuous particles, large-scale trajectories must be deduced from indirect ground-based observations and punctual air samples where those particles are in sufficient concentrations to be detected with current sampling technology (Masson et al., 2019; Schmale & Ross, 2015). This state of the art in tracing and predicting large-scale trajectories of airborne particles is a considerable handicap, particularly for assessing invasion risk or seasonal dynamics of airborne crop pests and to design surveillance strategies accordingly (Morris et al., 2021; Schmale & Ross, 2015; M. Wang et al., 2021).

Concerning living organisms, windborne long distance dispersal (typically larger than 10 km) has been largely explored for bacteria and fungi (Aylor, 2003, 2017; Barberán et al., 2015; Brown & Hovmøller, 2002; Morris

Visualization: S. Soubeyrand
Writing – original draft: H. Richard, S. Soubeyrand
Writing – review & editing: H. Richard, D. Martinetti, D. Lercier, Y. Fouillat, B. Hadi, M. Elkahky, J. Ding, L. Michel, C. E. Morris, K. Berthier, F. Maupas, S. Soubeyrand

et al., 2008; Soubeyrand et al., 2009), pollen and seeds (Nathan et al., 2011; Williams, 2010; Williams & Barneoud, 2021), aphids (M. Irwin & Thresh, 1988; Isard & Gage, 2001), lepidoptera (Westbrook et al., 2016; Wu et al., 2022), and midges (Aguilar-Vega et al., 2019). Such small organisms, especially microscopic ones, are not easy to track individually all along their aerial trajectories. However, as illustrated by some of the aforementioned references, they can be spotted in the atmosphere at some specific locations and times, or their origin and flow can be somewhat inferred using genetic data or isotopic profiles. Such information can then be associated to air-mass trajectories reconstructed with existing software, for example, the `HYSPLIT` atmospheric transport and dispersion model (Stein et al., 2015), the `TAPPAS` living organism dispersal model, which is based on `HYSPLIT` (Durr et al., 2017), and other models providing complementary analysis capabilities (Pretorius et al., 2023; Schmale & Ross, 2015). These software are useful for punctual estimations of dispersal over several days. They can also be used to assess long-term connectivity between distal areas as recently demonstrated by Pretorius et al. (2023), who identified attracting Lagrangian coherent structures (LCS) for invasive species and bushfire smoke between Australia and New Zealand (LCS resulting in predictable atmospheric pathways for such entities); see also Radici et al. (2022) for an illustration of long-term and large scale connectivity assessment with `HYSPLIT` for a fungal plant pathogen. However, these software do not facilitate (for users with moderate programming skills) the analysis of cumulative patterns of recurring atmospheric movements for large spatial networks over extended periods. Estimating long-term connectivity (e.g., over the entire cultural season of a given crop, or over several years) between many sites is key to understanding the spatial dynamics of populations over large scales (Leyronas et al., 2018), determining the likelihood of an air migration route (Pretorius et al., 2023; Radici et al., 2022; Zadoks, 1967), assessing the capacity of an epidemic to invade a fragmented space (Strona et al., 2017), or identifying sentinel points to be favored within the framework of epidemiological surveillance (Martinetti & Soubeyrand, 2019).

To reach these goals, one of the main challenges remains to be able to work with large networks of sites. The formalism of networks (or graphs) makes it possible to represent both the space through a set of nodes and the flows within this space by considering the links connecting the nodes. This formalism can be used to represent population dynamics or epidemics (Bajardi et al., 2011; Beaunée et al., 2015; Brooks et al., 2008; Hufnagel et al., 2004; Moslonka-Lefebvre et al., 2011): network nodes may correspond to regions, sites or host units; network links, which may be quantitatively weighted and oriented, may depend on geographical distance (or its derivatives such as resistance and least-cost distances, McRae, 2006; Picard et al., 2017), on trade flow, or on transport processes over land, air or sea. For airborne organisms, being able to infer a network, whose links would be weighted by the probabilities of atmospheric connection between sites, would offer the possibility of understanding and predicting dynamical spatial processes and species patterns across time, and to design relevant observation and control strategies (Morris et al., 2021). More or less simple frameworks for linking geographical sites using sequences of inferred movement of air masses have been recently proposed (Choufany, Martinetti, Senoussi, et al., 2021; Choufany, Martinetti, Soubeyrand, & Morris, 2021; Radici et al., 2022; M. Wang et al., 2021). The general idea consists of defining some target sites, computing sets of air-mass trajectories arriving to or departing from these sites (with a software such as `HYSPLIT`), and calculating the frequency at which these trajectories go through distant sites (which may coincide with other target sites). A spatial tolerance (i.e., a spatial buffer) is generally accounted for in the definition of "go through"; properties of the air masses along their trajectories can be accounted for to weight the trajectories linking two sites (Radici et al., 2022); and properties of windborne particles (typically their mass and shape) can also be considered to assess their ability to stay in the air mass or to be deposited along its trajectory (M. Wang et al., 2021).

Grounded on the proposal made by Choufany, Martinetti, Senoussi, et al. (2021) and Choufany, Martinetti, Soubeyrand, and Morris (2021), on the `HYSPLIT` model and on gridded meteorological data archives, the new web application `tropolink` presented in this article allows the construction of geographical networks from trajectories of air masses corresponding to sites and time periods specified by the user. In the name `tropolink`, "tropo" stands for troposphere, which is the lower layer of the atmosphere (ranging from the ground to 6–18 km, depending on the latitude and the season) and the realm of transportation of organisms that is pertinent to their ecology, evolution and impact on the environment. In addition, "link" evokes the between-site connectivity that we aim to evaluate. The architecture of `tropolink` is roughly illustrated by the simplified scheme provided in Figure 1, which describes the input (meteorological data archives, network structure and parameters), the output (air mass trajectories and connectivity matrices) and the mobilized tools (`HYSPLIT` as well as codes and notebooks for connectivity calculation and output exploration). `tropolink` allows the user to transparently

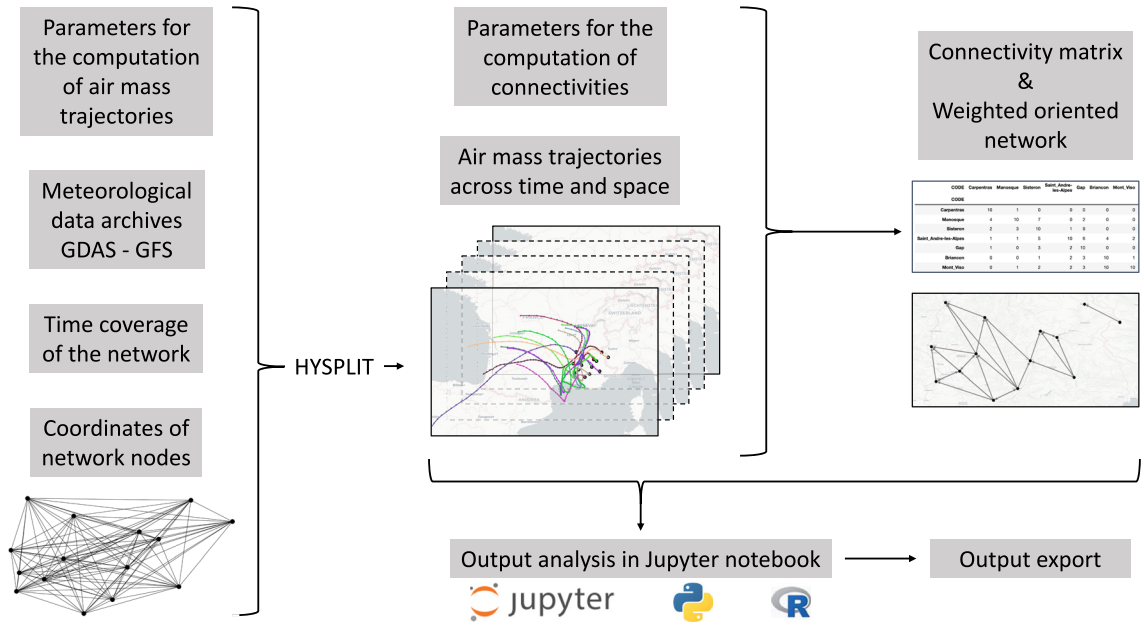

Figure 1. Simplified architecture of `tropolink`. After providing input parameters, dates, and node locations, `HYSPLIT` is executed using open meteorological data archives to generate sets of trajectories. Subsequently, connectivities between sites resulting from these trajectory sets can be computed by specifying connectivity parameters. Sets of trajectories and connectivities can then be explored within `jupyter` notebooks embedded in `tropolink` and exported for publication and in-depth analysis.

use massive and global gridded meteorological data and perform intensive computation on a distant computer cluster. It also allows the user to explicitly explore calculated trajectories and connectivity matrices in `jupyter` notebooks with `Python` and `R` environments. The notebooks give a hand to the user to further investigate `tropolink` output on their own computation system.

In the following sections, we briefly describe the theoretical background underlying `HYSPLIT` and the computation of between-site connectivity from sets of trajectories. We also present the steps the user must execute when using the web application, illustrate the analysis of trajectories and networks generated by `tropolink`, from local to continental scales, and discuss further development and usage of the web application. In the illustrations, we show how knowledge of geographical networks of air-mass movement is pertinent to biological questions concerning the epidemiology of plant diseases and pests. We have chosen to apply `tropolink` to investigating potential dispersal of four pests, vectors and pathogens for which there is strong evidence for or interrogation about long distance aerial dissemination. The illustrations deal with (a) quantifying tropospheric connectivities between areas of production of sugar beet that are affected by yellows viruses transmitted by aphids, (b) exploring the relationship between tropospheric connectivities between distal populations of a plant pathogen and the epidemiological links between these populations inferred from genomic sequence data, (c) assessing the risk of long-distance windborne spread of the fall armyworm (*Spodoptera frugiperda*), a destructive migratory pest that is currently expanding globally, and (d) computing the tropospheric connectivities along the (largely documented) *Puccinia* pathway in the USA situated at the 97° West longitude, along which wheat stem rust spores spread northward annually (Aylor, 2003).

2. Theoretical Background

From a modeling and mathematical viewpoint, `tropolink` relies on two main components, the reconstruction of air-mass trajectories, and the computation of connectivity between network nodes grounded on (possibly large) sets of trajectories. The former component is obtained from the Hybrid Single-Particle Lagrangian Integrated Trajectory model (`HYSPLIT`, Draxler & Hess, 1998; Stein et al., 2015) developed by the Air Resources Laboratory of the National Oceanic and Atmospheric Administration. The latter component is based on a recently developed framework for constructing networks from trajectory data (Choufany, Martinetti, Senoussi, et al., 2021). This section sketches the theoretical background of `HYSPLIT` and trajectory-based networks. The reader who

wants to skip this section only has to know that `tropolink` generates networks whose nodes are fixed and known and whose edges between nodes are calculated from trajectories of air masses occasionally connecting nodes: the directed edge from node A to node B is weighted by the number (or proportion) of daily trajectories simulated over a specified time period that start from (or arrive to) B and that go through A in a given time interval.

2.1. Reconstruction of Air-Mass Trajectories

Equations and algorithms underlying `HYSPLIT` are described by Draxler and Hess (1997, 1998) and Stein et al. (2015) and the following section only presents the main elements of its trajectory compartment. In `HYSPLIT`, air-mass trajectories are computed from time integrated advection in a Lagrangian framework, which requires the three-dimensional velocity field. This input is obtained from gridded outputs of a meteorological model (namely and at least, horizontal wind components, temperature, height or pressure, and pressure at the surface), which are pre-processed to be used in `HYSPLIT`. Pre-processing includes the reconstruction and interpolation of 3D-wind velocities at any height from the ground to the top layer that is considered using the meteorological model output. Interpolation is performed to re-map data into a specific terrain-following vertical coordinate system; See Rolph et al. (2017) who describe how terrain is taken into account, with an illustration in the online `HYSPLIT` Basic Tutorial: https://www.ready.noaa.gov/documents/Tutorial/html/traj_isob.html.

Once the 3D-velocity field $\mathbf{V} = (U, V, W)$ has been processed and interpolated to the internal model grid, `HYSPLIT` implements the advection in two stages: the position at time t of an air mass, say $P(t)$, is transformed into a first-guess position $P'(t + \Delta t)$ after a duration Δt using the following equation:

$$P'(t + \Delta t) = P(t) + \mathbf{V}(P, t)\Delta t,$$

then the final value of the position at time $t + \Delta t$ is given by:

$$P(t + \Delta t) = P(t) + \frac{1}{2}\left\{\mathbf{V}(P, t) + \mathbf{V}\left(P', t + \Delta t\right)\right\}\Delta t.$$

The integration time step Δt can vary along the computation of the trajectory and is computed from the requirement that the advection distance per time-step should be less than 0.75 of the meteorological grid spacing. The variable nature of Δt requires to interpolate $\mathbf{V}$ at the positions $P(t)$ and $P'(t + \Delta t)$, which is performed in a linear way. Trajectories are terminated if they exit the top layer considered in the model, but advection continues along the surface if the trajectories intersect the ground.

2.2. Trajectory-Based Networks

Spatial and spatio-temporal networks generated from `tropolink` have fixed and known nodes typically consisting of fixed-radius circles around a set of geographical sites as well as directed edges weighted by a connectivity measure grounded on the movement of air masses during the time period of interest. The connectivity measure is derived from the work of Choufany, Martinetti, Senoussi, et al. (2021) as described below.

Consider two spatial domains A and B at the surface of the Earth (i.e., two sets of latitudes-longitudes) and a period of time T, eventually made of disjoint temporal intervals. Choufany, Martinetti, Senoussi, et al. (2021, Definition 2.3) defines the δ-lag directed connectivity from A to B over T by:

$$\Psi^{(2)}_{\delta,\nu}(A \to B, T) = \int_{T\times B} \psi(A|s + \delta, s, x)\nu(ds, dx), \tag{1}$$

where $\nu(ds, dx)$ is a probability measure on $T \times B$, δ is a time lag (positive or negative), and ψ is a scalar operator called *pointwise connectivity* (see Definition 2.2 in Choufany, Martinetti, Senoussi, et al. (2021)). A simple interpretable form of Equation 1 is obtained when ν is the Lebesgue measure $\nu(ds, dx) = ds \times dx$, and ψ is the binary function indicating whether the trajectory $\Gamma(s + \delta, s, x)$ of the air mass located at x at time s went through A between s and $s + \delta$. If $\delta > 0$, the trajectory is forward in time (it starts at x at time s and we consider it up to time $s + \delta > s$). If $\delta < 0$, the trajectory is backward in time (it arrives at x at time s and we consider it since time $s + \delta < s$). Hence, Equation 1 becomes:

$$\Psi^{(2)}_{\delta,\nu}(A \to B, T) = \int_{T\times B} \mathbb{1}(A \cap \Gamma(s + \delta, s, x) \neq \varnothing)dsdx, \tag{2}$$

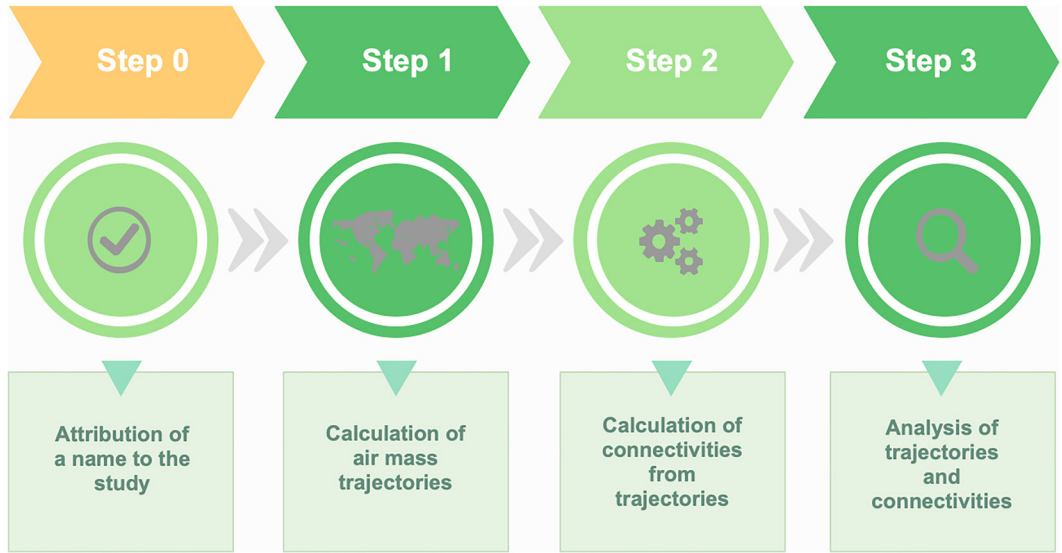

Figure 2. Workflow concept of `tropolink`.

where $\mathbb{1}(E) = 1$ if the event E is true, $\mathbb{1}(E) = 0$ otherwise. Equation 2 can be interpreted as the frequency of passage through A of trajectories of duration $|\delta|$. If one approximates this integral by taking only the centroid of B, say $\bar{x}_B$, and only one time per day over T, for example, 12:00 UTC (assuming that T is the union of a set of whole days), one obtains the following estimate of Equation 2:

$$\hat{\Psi}^{(2)}_{\delta,\nu}(A \to B, T) = \frac{|T||B|}{n_T} \sum_{i=1}^{n_T} \mathbb{1}(A \cap \Gamma(s_i + \delta, s_i, \bar{x}_B) \neq \varnothing), \tag{3}$$

where n_T is the number of days in T, s_i is the unique time taken for the ith day in T, $|T|$, and $|B|$ are the length and area of T and B, respectively. Hence, Equation 3 is simply the proportion of sampled trajectories intersecting A, multiplied by $|T||B|$. In the applications, we set $|T||B| = 1$ for facilitating the interpretation of results.

3. Tropolink Workflow

The `tropolink` workflow can be summarized in four steps illustrated by Figure 2 and detailed in this section. Texts S1 and S2 in Supporting Information S1 respectively give a step-by-step description of the use of the application and technical details of its development.

3.1. Step 0: Attribution of Study Name

As a preliminary step, the user is prompted to give a name and description to the study they want to carry out. The name will be used to identify the set of trajectories to be computed. The name extended by an underscore and a number will be used to identify each connectivity calculation to be carried out based on the aforementioned set of trajectories.

3.2. Step 1: Trajectory Calculation

In step 1, the user provides input information consisting of the set of dates at which trajectories and connectivities have to be computed (T in Section 2.2), the geographical sites forming the nodes of the network ($\bar{x}_B$ in Section 2.2), and parameters used for `HYSPLIT` tuning. These input are provided by the user by filling out a form embedded in the web application. In this form, temporal and spatial information can be provided as external files in an application-specific format (text files with specific characteristics grounded on ISO standards). Popup windows indicate the format to be used for each field-of-the-form/file when necessary.

Concerning the parameters used for `HYSPLIT` tuning, namely "trajectory duration" (with negative or positive value for backward or forward trajectories, respectively), "starting/arrival hour," "top of model domain," and

"vertical motion method," the help page refers the user to the original and complete documentation of HYSPLIT. Once these four parameters are provided, the use of HYSPLIT is transparent: the software and wrapper codes as well as input meteorological data are on distant data storage and computer cluster queried by the web application tropolink. Input meteorological data that are used by tropolink (and HYSPLIT) are global and cover the period from September 2007 to present. They are extracted from the Global Data Assimilation System half-degree archive from September 2007 to June 2019, and from the Global Forecast System quarter-degree archive from June 2019 (https://www.ready.noaa.gov/archives.php).

The current version of tropolink uses the version v4.2.0 of HYSPLIT in compliance with its End-User License Agreement (Draxler & Rolph, 2010; Rolph et al., 2017). This version is installed on a Slurm-based computer cluster (version 22.05.2) operated with the Debian GNU/Linux 10.03 (Buster) system and administered by the INRAE BioSP research unit.

From a biological perspective, the tuning of trajectory calculations must be performed in a manner that aligns, whenever possible, with the properties or behavior of the organism transported by air mass movements. For example, the typical takeoff time of the insect or particle of interest may serve as the starting point for trajectory calculations. The selected dates for trajectory calculations may coincide with favorable weather conditions for insect or spore takeoff, while the specified locations may correspond to areas featuring high densities of host habitats. Further discussion on this point can be found in Section 5.

3.3. Step 2: Connectivity Calculation

Once trajectories of air masses have been calculated, the user can select the corresponding study, parameterize the connectivity calculation by selecting a type of buffer around nodes and a buffer radius, and run the connectivity calculation. The selection of a buffer around nodes allows the user to specify a spatial tolerance in the computation of the frequency of trajectories arriving at (or departing from) a given node that go through another node or, more precisely, that go through a buffer zone around this node. The buffer can be "circular" (actually, a spherical cap), that is, it contains all points at a geographical distance lower than a specified radius (up to 100 km). It can also correspond to a latitude-longitude geographical sector centered around the node with half angle to be specified (up to 2°). Moreover, the user can select a subset of dates and locations for which connectivities have to be calculated. The specification of a connectivity calculation can be provided as a json file instead of filling each element of the form.

3.4. Step 3: Jupyter Notebooks

The last step of the tropolink workflow allows the user to perform a first analysis of computed air-mass trajectories and connectivities in a preconfigured Jupyter environment. The application offers two Jupyter notebooks: the first one is based on a Python kernel; the second one is based on a Python kernel with rpy2 Python package allowing the user to enrich the notebook with R commands. These notebooks show some more or less basic commands to help the users to launch their analyses on their own computer system. In particular, the user can download, in the notebooks, data frames containing trajectory or connectivity data with an easy-to-use format in Python or R, plot air mass trajectories with a background map, and display connectivity matrices as oriented networks.

3.5. Additional Features

The application contains a dashboard with a list of the user's studies. This list gives the status of each calculation (in progress, succeeded or failed), it gives access to the specifications for each calculation, which can be downloaded in a json format (for re-use, modification and sharing), as well as calculation outputs (either trajectories or connectivity matrices).

In addition, tropolink includes a help-tab complementing information provided by popup windows as well as a tab with credits.

Table 1
Specifications of `tropolink` *for Running the Examples About Sugar Beet, Potyvirus, Fall Armyworm (FAW), and Puccinia Pathway*

	Sugar beet	Potyvirus	FAW	*Puccinia*
Dates	Each day of April over 2013–2022	Each day of month m in 2014, for m from January to December	Each day of May–September over 2013–2022	Each day of April–July over 2019–2023
Number of locations	40	27	2,365	20
Longitude extent	(0.3°, 4.4°)	(4.57°, 5.08°)	(−25°, 55°)	−97°
Latitude extent	(46.0°, 51.0°)	(43.63°, 43.73°)	(10°, 50°)	(29°, 48°)
Starting/arrival altitude (m)	500	500	1,000	200
Trajectory duration (hr)	3	−3	12	48
Starting/arrival hour (hr UTC)	14	14	20	17
Top of model domain (m)	10,000	10,000	15,000	10,000
Vertical motion method	Meteorological model's vertical velocity fields			
Buffer type	Circular	Circular	Geographic	Circular
Buffer radius	23 km	1 km	0.5°	25 km

4. Applications

4.1. Estimating Airborne Connectivity Between Areas of Production of Sugar Beet

Here, we present a typical example of estimation of a connectivity matrix that could be used, for instance, as an input for a disease propagation model or in comparative studies of putative influencing factors.

Sugar beet yellowing is caused by several viruses whose main vectors are aphids. Aphids can be transported by air masses over long distances and an important question concerns the contribution of air mass movement to the regional circulation of the aphids and/or the disease with respect to other factors, including more local factors that are known to play a role in aphid and disease dynamics (e.g., see Qi et al. (2004), for the role of winter temperature). Here, we do not address this question beyond the scope of this article, but we aim to assess how spatially heterogeneous the tropospheric connectivity between different production areas of sugar beet is.

Thus, using `tropolink`, we computed connectivity between 40 production areas, which were identified with a standard hierarchical clustering method (namely the default method implemented in the `hclust` function in `R`) applied to the coordinates of 621 sugar beet fields in France; see Figure S1 in Supporting Information S1. These fields, whose locations were provided by ITB (https://www.itbfr.org/) and which were monitored from 2017 to 2021, are considered here to form a representative sample of sugar beet crops in central and northern France. Connectivities between the 40 production areas were inferred from data covering all dates in April from 2013 to 2022, using `tropolink` with tuning features given by Table 1 and justified in Text S3 in Supporting Information S1.

The distribution of connectivities shows a large proportion of weak values, most of them corresponding to very distant production areas, and the maximum connectivity is almost 0.4; see Figure 3. Changing the starting altitude of trajectories (either 250 or 750 m instead of 500 m) approximately leads to similar connectivity values. Figure 4 gives an example of forward trajectories obtained for a given date, namely 09 April 2013, and displays the main edges of the network generated by trajectories computed over a set of 300 dates specified in Table 1. When only edges with connectivities larger than 0.15 are shown, we mostly observe diagonal links South-West–North-East (SW-NE) in both directions. We observe a similar pattern when only edges with connectivities larger than 0.05 and lengths larger than 80 km are shown. These observations suggest that an eventual epidemiological dynamics dependent on the tropospheric network may display elongated SW-NE-diagonal shapes.

4.2. Investigating Drivers of Epidemiological Links

The second example consists of estimating a connectivity matrix based on air mass movement and assessing whether or not it matches a matrix of epidemiological links grounded on genetic information.

Alamil et al. (2019a) analyzed genomic data to assess epidemiological links within a pathogen metapopulation. The pathogen, a potyvirus, was sampled in 2014 from wild host plants (*Tragopogon pratensis*) located in 27 patches within a 40 × 10 km area in south-eastern France. The epidemiological links between the 27 pathogen populations were inferred with the `SLAFEEL` approach (Alamil et al., 2019a, 2019b) and are displayed in Figure 5, top panel. The potyvirus is transmitted by aphids that, as in the example above, may be transported by wind over relatively long-distance. Hence, one of the hypotheses possibly underlying the spatial structure of the potyvirus genetics (measured by the epidemiological links inferred with `SLAFEEL`) is that patches frequently connected by wind contain close genetic variants of this virus. Several other hypotheses

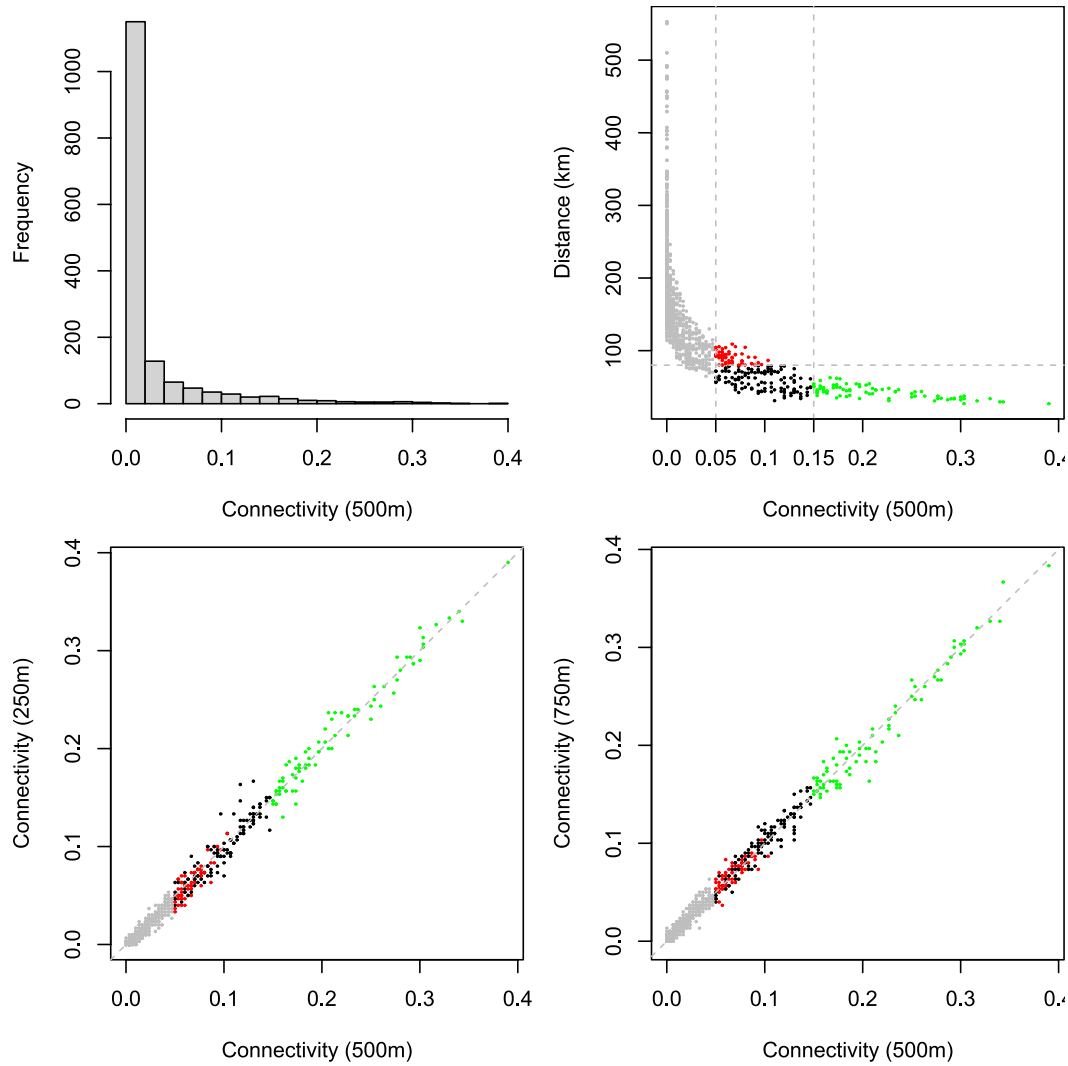

Figure 3. Sugar beet application: Distribution of non-diagonal connectivities based on trajectories starting at 500 m, say $\hat{\Psi}^{(2)}_{500}$ (top left); Geographical distance versus $\hat{\Psi}^{(2)}_{500}$ (top right); $\hat{\Psi}^{(2)}_{250}$ versus $\hat{\Psi}^{(2)}_{500}$ (bottom left), and $\hat{\Psi}^{(2)}_{750}$ versus $\hat{\Psi}^{(2)}_{500}$ (bottom right). Red points: connectivities in $\hat{\Psi}^{(2)}_{500}$ larger than 0.05 and geographical distances larger than 80 km; Green points: connectivities larger than 0.15; Black points: connectivities between 0.05 and 0.15 and geographical distances lower than 80 km; and Gray points: connectivities lower than 0.05.

have been proposed by Alamil et al. (2019a) including environmental conditions as well as host demography and genetics at the level of each patch. Here, we aim to assess whether or not epidemiological links and tropospheric connectivities coincide, at least partially.

In this aim, connectivities were inferred using `tropolink` applied to the 27 patches. A connectivity matrix was obtained for each month of the year 2014 during which genetic data were collected (tuning features are provided in Table 1 and justified in Text S3 in Supporting Information S1), and a global connectivity matrix was obtained for the whole year 2014. Given the weak spatial resolution of meteorological data at the spatial scale considered here, the trajectories computed by `HYSPLIT` (and `tropolink`) mainly provide an indication of the dominant wind direction at each specified date; see Figure S2 in Supporting Information S1.

For the study area considered in this example, we notice a moderate variation of connectivity across the months of the year: the correlation between monthly connectivity matrices (without diagonal terms) ranges between 0.85 and 0.95; Figure S3 in Supporting Information S1 shows how the 12 months of the year are clustered in a relatively consistent manner with respect to the seasons. Figure 5, bottom panel, represents the tropospheric network obtained for the whole year 2014 (only connectivities larger than 0.05 are plotted). To assess the relationship

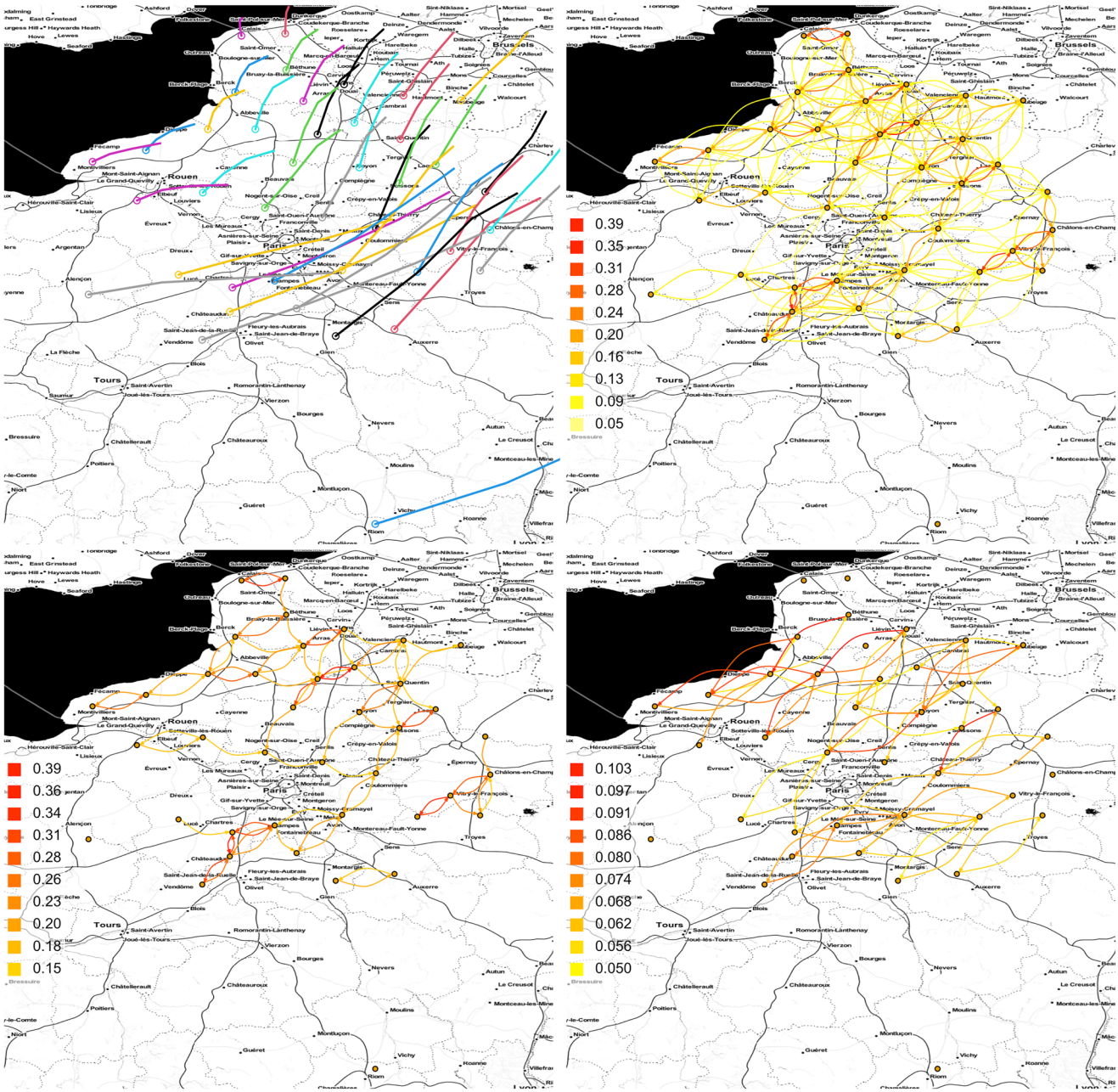

Figure 4. Sugar beet application: trajectories computed on 09 April 2013 (top left); network edges corresponding to connectivities larger than 0.05 (top right), to connectivities larger than 0.15 (bottom left), and to connectivities larger than 0.05 and lengths larger than 80 km (bottom right).

between the matrices of epidemiological links and tropospheric connectivities, we computed the correlation ρ of non-diagonal terms and we compared it to the correlations ρ^* obtained by permuting 10^4 times the rows of the matrix of epidemiological links (permuting the rows of this matrix means that the vectors of putative sources are switched between the virus populations). This permutation approach allowed us to compute a p-value given by the proportion of correlations ρ^* larger than or equal to ρ. Figure 6 shows that the p-value is relatively large, either when considering monthly connectivities, whatever the month, or the annual connectivities. P-values are weaker for June, July, and August, but remain above the classical 0.05 significance level (the network obtained for August leading to the lowest monthly p-value is shown in Figure S4 in Supporting Information S1). We computed a similar p-value grounded on the correlation between the matrices of epidemiological links and geographical distances. Here, the p-value is the proportion of correlations ρ^* lower than or equal to ρ (because under the

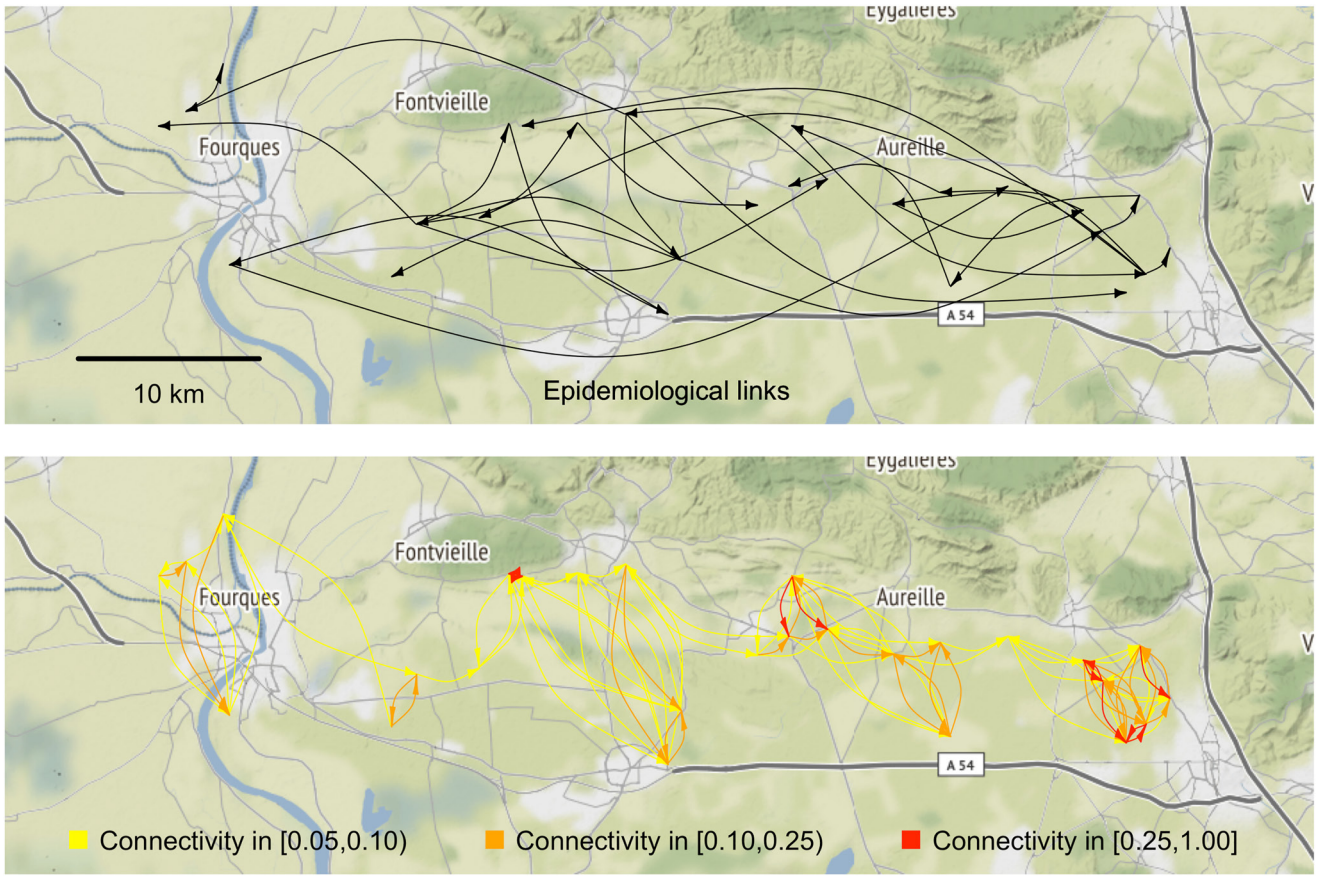

Figure 5. Application of air mass connectivity to potyvirus epidemiology: Epidemiological links between the 27 potyvirus populations inferred by Alamil et al. (2019a) from genomic data (top panel; the links are oriented from the main putative sources for each population), and tropospheric connectivities obtained from `tropolink` for the whole year 2014 (bottom panel; connectivities lower than 0.05 are not plotted).

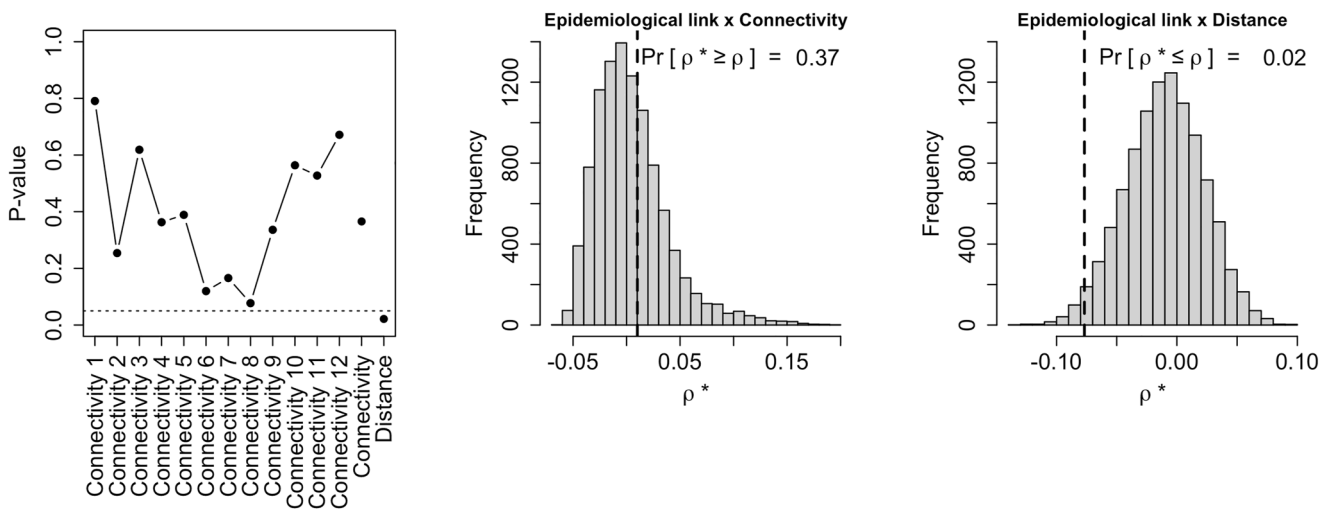

Figure 6. Application of air mass connectivity to potyvirus epidemiology: *P*-values for testing the absence of relationship between epidemiological links and either monthly connectivities (from month 1, i.e., January, to month 12, i.e., December), annual connectivities or geographical distances (left); Distribution of the correlation ρ^* between permuted epidemiological links and annual connectivities (center), the value ρ obtained without permutation is given by the dashed line; Distribution of the correlation ρ^* between permuted epidemiological links and geographical distances (right), the value ρ obtained without permutation is given by the dashed line.

alternative hypothesis, epidemiological links and geographical distances are negatively correlated). The small *p*-value (namely 0.02) obtained in this case suggests that epidemiological links and geographical distances are negatively correlated, whereas an eventual relationship between epidemiological links and tropospheric connectivities is not significant based on available data. More data should be collected to test with a larger power an eventual link with air mass connectivities in summer.

4.3. Assessing the Risk of Long-Distance Spread of Fall Armyworm

In this example, we compute a connectivity matrix at a larger spatial scale than those considered in the two previous examples, and use it to assess a component of the invasion risk of a crop pest.

The fall armyworm (FAW, *S. frugiperda*) is a migratory pest insect, which is in a phase of invasion at the global scale, with recently newly colonized territories in Asia and Africa (Early et al., 2018). FAW larvae may be disseminated via the trade of vegetables and fruits, but the adults may also spread with air masses over long distances, typically during nights (Westbrook et al., 2016; Wu et al., 2022). Currently, an important issue concerns the evaluation of the risk of expansion of FAW in northern Africa and Europe. Here, we want to estimate which FAW-free areas in these regions could be reached given the flight capabilities of FAW and the areas that it presently occupies.

Thus, we computed tropospheric connectivity across a spatial grid covering land at latitudes from Burkina Faso to France (10°–50°N) and longitudes from Cabo Verde to Saudi Arabia (−25°–55°E). Tuning features used to apply `tropolink` are provided in Table 1 and justified in Text S3 in Supporting Information S1. Typical trajectories at a given date are displayed in Figure S5 in Supporting Information S1. The connectivity matrix and the vector of FAW presence/pseudo-absence (considered over the same 1°-resolution) were multiplied (with the matrix product) to estimate a heatmap measuring the risk of long distance dispersal from occupied areas. FAW presence/pseudo-absence was derived from data extracted from the FAMEWS data base provided by the Food and Agriculture Organization (FAO) and centralizing crowd-sourcing observations of the fall armyworm (using two sampling approaches: traps and scouting). In this data base, only data with latitudes and longitudes falling within the borders of the indicated countries were kept to build the vector of presence/pseudo-absence (i.e., >65,000 observations between 01 January 2018 and 10 April 2023; 47% of observed presence for traps, 72% for scouting). FAW is considered to be present in a grid node if at least one observation of FAW was made over the 2018–2023 sampling period; see Figure 7. FAW is considered as absent in the other grid nodes (either without sampling or with sampling but without FAW observation), but this absence is viewed as a pseudo-absence since FAW might be at densities under the detection threshold in areas covered by the surveillance and might be present in unsampled areas (if Europe is actually considered as FAW-free, some unsampled territories in Africa and western Asia might be occupied by FAW).

The heatmap that is shown in Figure 8 allows us to identify areas at risk of long-distance dispersal based on the partial current knowledge of FAW spatial distribution. We observe in particular some connections between already occupied areas, for example, between continental western Africa and Cabo Verde archipelago, and possible migration across seas (e.g., across the eastern Mediterranean and the southern Red Sea).

4.4. Evaluating Connectivities Along a Recurrent Migration Pathway

In this example, we compute a connectivity matrix to partially characterize a recurrent migration route of an airborne plant pathogen that has already been thoroughly studied.

Several fungus pathways, described as predictable routes for the recolonization of plant fungal pathogens on the basis of host availability, prevailing seasonal winds and favorable weather conditions, have been especially identified in the USA, India, China and Europe (Brown & Hovmøller, 2002). One such pathway has been pinpointed for the stem rust of wheat caused by *Puccinia graminis* along the 97° West longitude (Aylor, 2003; Hamilton & Stakman, 1967). In this pathway, urediniospores produced by *P. graminis* lesions overwinter in the southern USA and northern Mexico and are subsequently transported northward from April to July, allowing the infection of progressively planted wheat from Texas to North Dakota (Hamilton & Stakman, 1967). Here, we aim to quantify the prevalence level of northward winds along the 97° West longitude during the observed migration period (thus refining the long-standing knowledge that northward winds prevail in summer; Ward, 1916).

To accomplish this, we utilize `tropolink` to calculate tropospheric connectivity among 20 sites displayed in Figure 9, which also shows an example of air-mass trajectories obtained for a specific date. The parameters used to configure `tropolink` for this case study are detailed in Table 1, with justifications provided in Text S3

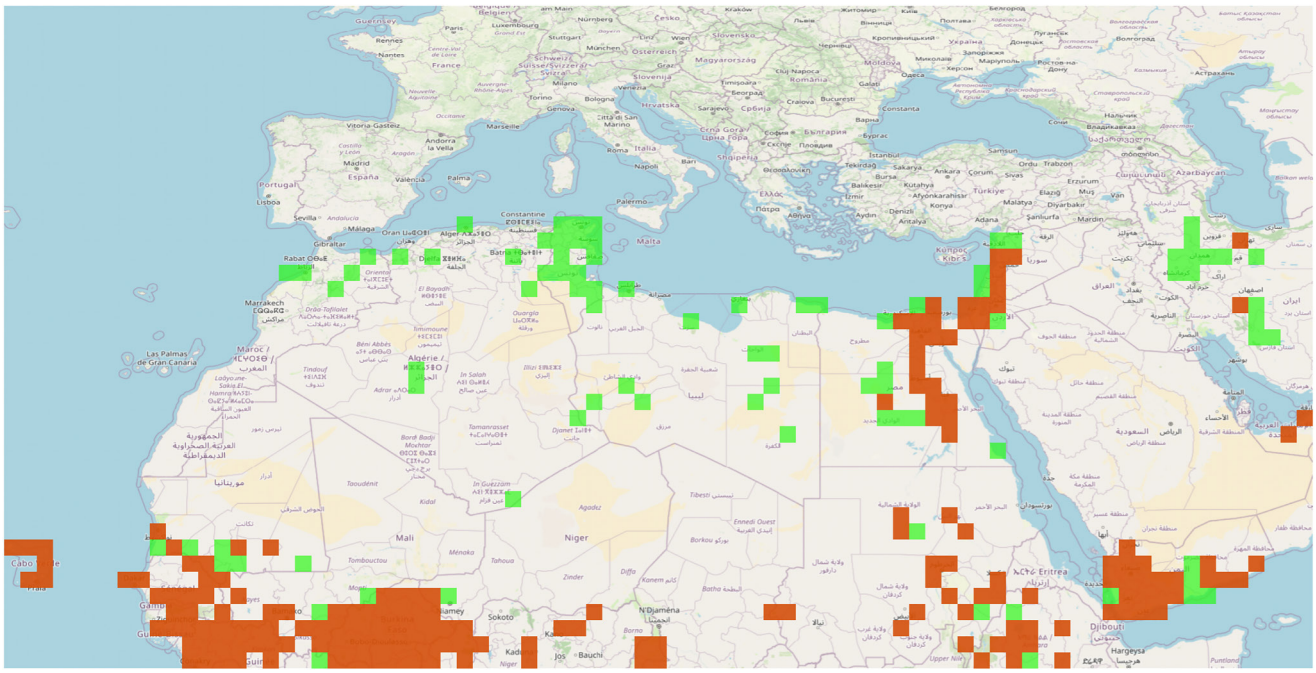

Figure 7. Fall armyworm (FAW) application: Spatial distribution of FAW presence (red grid nodes) and absence (green grid nodes) based on the Food and Agriculture Organization data base with records up to 10 April 2023.

in Supporting Information S1. The resulting connectivity matrix is presented in Figure 9. Notably, this matrix reveals that northward connections occur approximately four times more frequently than southward connections. Indeed, on average, there are 416.5 monthly connections from every sites to northern sites, whereas this average drops to 96.2 for connections to southern sites. Figure S6 in Supporting Information S1 shows that this

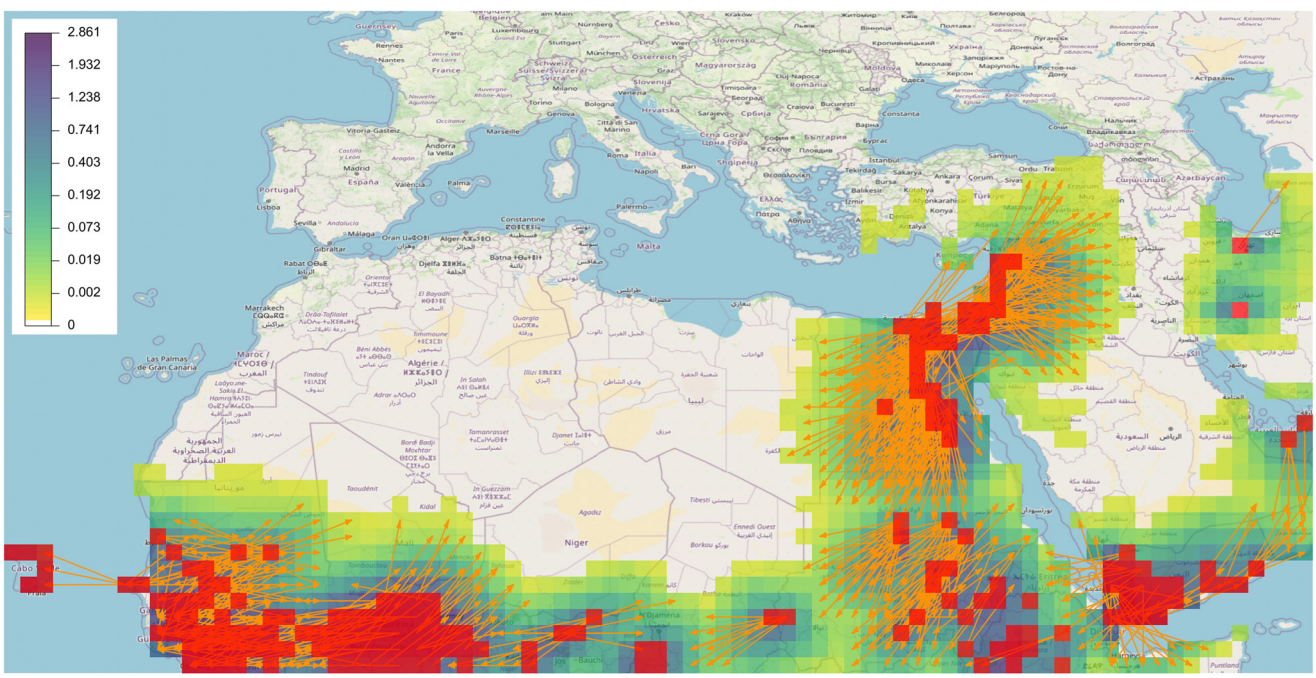

Figure 8. Fall armyworm (FAW) application: Heatmap (color palette from white to purple) corresponding to the matrix product between the connectivity matrix and the vector of FAW presence/pseudo-absence based on Food and Agriculture Organization (FAO) data. Red grid nodes: areas where FAW is present based on the FAO data base. Orange arrows: Connectivities between distant grid nodes that are larger than 0.001 (connection at least 1 day over 1,000) and longer than 500 km.

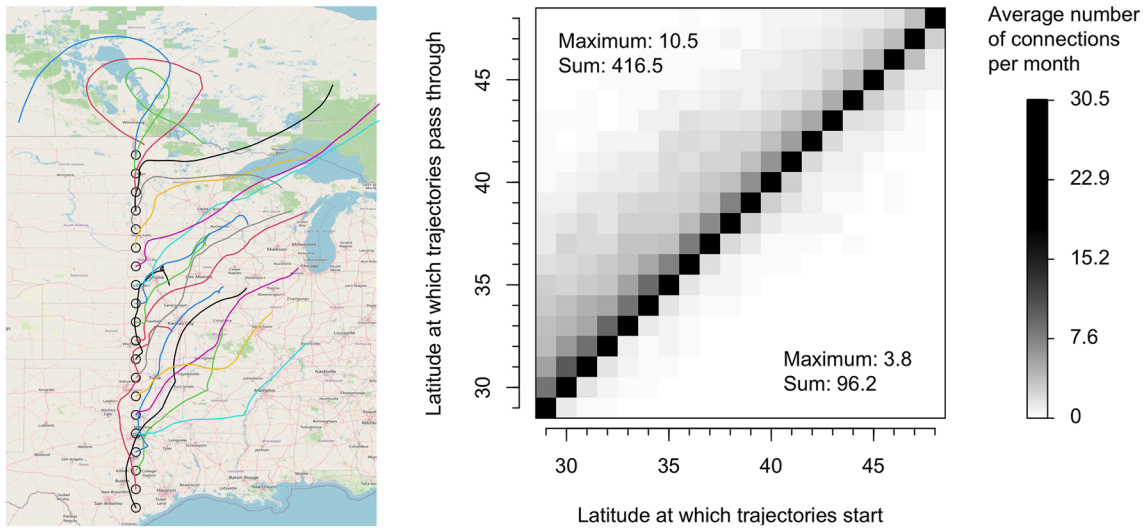

Figure 9. *Puccinia* pathway application: trajectories computed on 20 June 2020 for sites (circles) located along the 97° West longitude (left) and representation of the connectivity matrix (right). Connectivity is expressed in terms of average number of pairwise connections per month, whose theoretical maximum value is 30.5, that is, the mean number of days per month over the study period. Numbers within the plot on the right-hand-side give the maximum and the sum (averaged over 1 month) of northward connections (top left) and southward connections (bottom right).

predominance of northward connections remains consistent when we reduce the buffer radius around the sites used in the connectivity calculation from 25 to 10 km.

5. Discussion

The computation of spatial connectivities generated by air-mass movement and the resulting geographical networks offers new possibilities to test and explore the links between the atmosphere and biological systems with components that are disseminated via the air (Choufany, Martinetti, Senoussi, et al., 2021; Choufany, Martinetti, Soubeyrand, & Morris, 2021; Leyronas et al., 2018; Pretorius et al., 2023; Radici et al., 2022; M. Wang et al., 2021; J. Wang et al., 2023). The `tropolink` web application precisely contributes to this objective: it facilitates the computation of air mass trajectories and the resulting connectivity matrix for a network of sites, which relies on relatively complex software and massive data for non-specialist scientists. It should enable users to address a variety of questions related to invasion risks and spatio-temporal demo-genetic structures of populations, as well as to handle different spatial and temporal scales and varying numbers of network nodes, as exemplified by the scenarios explored in this article. However, it is important to note that other modeling approaches may be more suitable for describing and predicting the transport of plant pathogens and pests at specific scales, for particular organisms and to test particular hypotheses (Clobert et al., 2012; Jeger et al., 2007; Kenkre & Giuggioli, 2021; Nathan et al., 2011; Schmale & Ross, 2015). The aforementioned examples concern plant pests and diseases, but our approach can be used to generate connectivity data for diverse application domains: for example, the dispersal from identified sources of pollen, dust, radioactive elements, smokes, and other pollutants; see references in the introduction. For such particles, often playing the role of contaminants and xenobiotics, a significant challenge lies in predicting their dispersal locations and magnitudes at various scales (Committee of the Significance of International Transport of Air Pollutants, 2010; Kim et al., 2014; Reisen et al., 2015). In this context, `tropolink` can be applied to forecast spatial patterns related to long-term exposure risks, utilizing estimated tropospheric connectivities originating from pollutant sources. We mentioned above the practical aspects of `tropolink` for the non-specialist user. In addition, in terms of computation time, if the 40-cores computer cluster dedicated to `tropolink` is fully available, a trajectory computation with 30 dates, 40 network nodes distributed in a 500 × 500 km box and 3 hr trajectory duration takes about 3 min. A trajectory computation with around 1,500 dates, 2,300 network nodes distributed in a 5,000 × 5,000 km box and 12 hr trajectory duration takes about 3 hr. The connectivity computation for the latter trajectory study takes about 3 hr 30 min.

Concerning the examples presented above, only preliminary analyses were performed in this article. It would be of particular interest to assess whether the connectivity pattern inferred in the sugar beet application could influence the correlation structure of yellowing disease epidemics in distant production areas, or correlates with

the spatial distributions of either the viruses causing the disease or the aphids transmitting the viruses. The latter task could be accomplished by utilizing data generated from long-term suction trap networks designed to collect migratory aphids (Bell et al., 2015; Lagos-Kutz et al., 2020). Such an approach could complement current endeavors aimed at modeling and predicting the dynamics of aphids responsible for the spread of sugar beet yellowing disease (Luquet et al., 2023). Furthermore, we could employ the geographical network that we have inferred to design sampling schemes encompassing contrasting situations in terms of connectivity strength. This would allow us to test hypotheses related to aphid transport in relation to their biology with higher statistical power than with a naive sampling scheme. For the potyvirus application, additional data should be collected to increase the power of the test of an eventual relationship between epidemiological links and tropospheric connectivities in summer, as suggested in Section 4.2. Moreover, other measures of epidemiological links between populations derived from genetic distances (Alamil et al., 2022) may be proposed and analyzed with respect to the connectivity patterns provided by `tropolink` to detect an eventual signature of the inter-population spread of aphids on the virus dynamics. Regarding the *Puccinia* pathway, the quantitative information provided by `tropolink` can allow us to measure how much and where winds prevail over multiple years or for different years. This information can be used to assess whether the annual variations in propagation speed along a pathway, as described by Hamilton and Stakman (1967), are correlated with changes in annual tropospheric connectivity, in addition to fluctuations in host availability and weather conditions. It is also worth exploring whether prevailing winds are a necessary condition or if even minor yet significant winds could be sufficient. This investigation may be pursued by revisiting with `tropolink` the various plant pathogen pathways reported from around the world (Brown & Hovmøller, 2002; Meyer et al., 2017; Radici et al., 2022; Saubin et al., 2023). For the FAW application, the heatmap is a tool for orientating the surveillance of FAW when one accounts for its natural ability to disperse (note that the surveillance strategy must also depend on other factors, e.g., the risk of introduction with trade of goods). Such a map, crossed with maps of crop sensitivity to FAW larvae and climate-suitability for FAW (J. Wang et al., 2023), may be used to inform the sampling strategies with the aims of (a) delineating the actual spatial distribution of the pest and (b) enhancing the surveillance in pest-free areas.

Aim (a) (*delineating the actual spatial distribution of the pest*) is illustrated by the case of Cyprus and Türkiye: these countries do not use the voluntary FAW monitoring tool provided by FAO, resulting in the absence of FAW observations in the FAMEWS data base. Nevertheless, countries must report presence of invasive pests such as FAW to global and regional plant protection organizations such as the International plant Protection Convention and the European and Mediterranean Plant Protection Organization (EPPO). Thus, a recent map of the EPPO indicates that the pest is present in Cyprus and Türkiye (see Figure S7 in Supporting Information S1): FAW was detected for the first time in early 2023 in the Limassol district, Cyprus (EPPO Reporting Service no. 02/2023, Num. Article: 2023/034), and in 2022 in the Adana province, Türkiye (Pehlivan & Atakan, 2022). Cyprus and the South of Türkiye are precisely possible arrival points of FAW based on our heatmap, with a possible long-distance origin from northern Egypt. Hence, our heatmap may be used to orientate surveillance of FAW in the aim of improving the delineation of the FAW distribution based on the voluntary data points provided by countries. Note however that FAW is also present in Canary Islands based on EPPO data (EPPO Reporting Service no. 03/2021 Num. Article: 2021/053), whereas this territory is not highlighted by our heatmap. FAW possibly arrived in Canary Islands with trade of goods (Cock et al., 2017; Plateforme ESV, 2022), via flights of adults during several successive nights (Westbrook et al., 2016), or from an area currently supposed to be FAW-free but which is not.

Aim (b) (*enhancing the surveillance in pest-free areas*) is illustrated by the hypothetical case displayed in Figure 10. In this example, we added nine grid nodes where FAW is present based on EPPO data (namely in Cyprus, Türkiye, and Canary Islands) and two grid nodes where FAW is hypothetically present in Morocco and Tunisia, and we assessed the risk of long distance spread. The inclusion of Morocco and Tunisia is purely hypothetical: FAW has not been reported there through any channel we know of, neither through FAO nor through any regional/global plant protection organization. The presence of FAW in Cyprus and Türkiye increases the risk area in Türkiye and generates a possible dispersal toward the North-East of the Black Sea. The presence of FAW in Canary Islands generates a risk of dispersal to Western Sahara. In addition, if FAW were present in northern Morocco (resp. northern Tunisia), this would pose a new risk of introduction into southern Spain and Portugal (resp. Italy and Corsica), which should be taken into account to guide surveillance in Europe. Beyond surveillance objectives, the connectivities from occupied areas to pest-free areas may be used to target likely source populations of FAW and assess the adaptive potential of these populations to the pest-free areas under focus (Estoup et al., 2016).

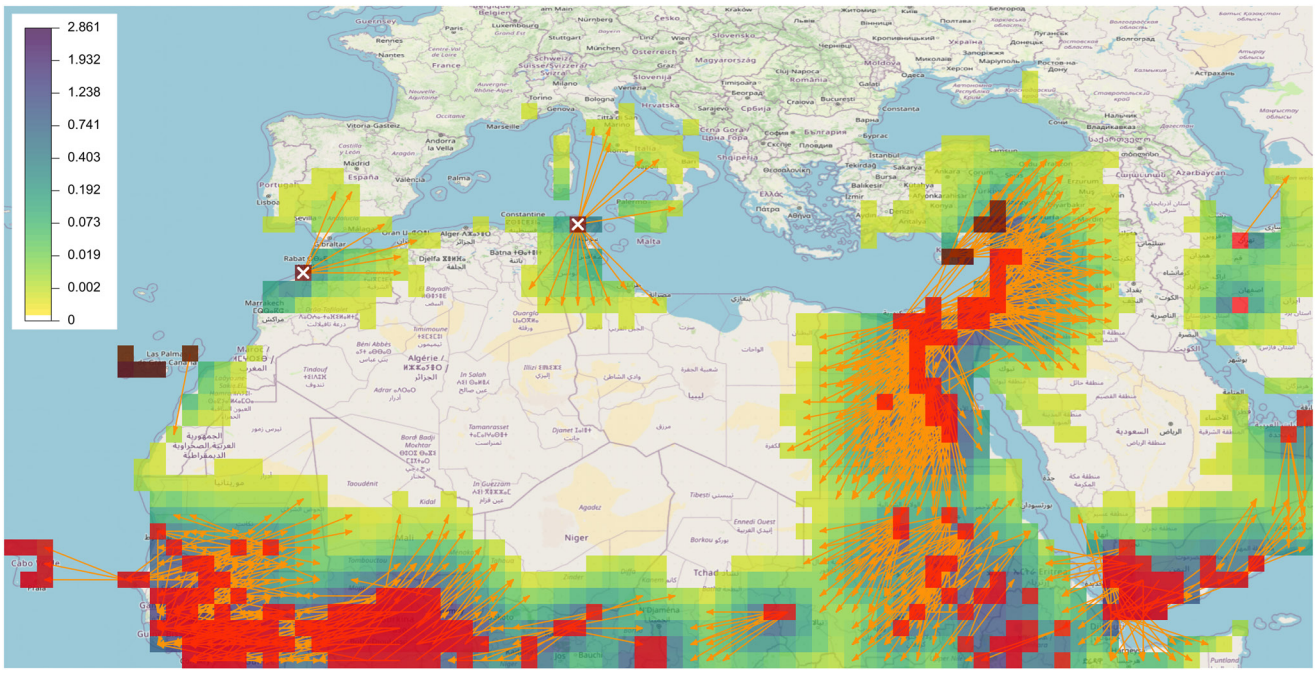

Figure 10. Fall armyworm (FAW) application: Hypothetical heatmap (color palette from white to purple) corresponding to the matrix product between the connectivity matrix and the vector of FAW presence/pseudo-absence based on Food and Agriculture Organization (FAO) and European and Mediterranean Plant Protection Organization (EPPO) data augmented by two hypothetical grid nodes of FAW presence. Red grid nodes: areas where FAW is present based on the FAO data base. Brown grid nodes: areas where FAW is present based on the EPPO data. Brown grid nodes with white crosses: hypothetical presence of FAW in Morocco and Tunisia, which are actually FAW-free areas. Orange arrows: Connectivities between distant grid nodes that are larger than 0.001 (connection at least 1 day over 1,000) and longer than 500 km.

In order to advance our ambition of guiding surveillance efforts and enhancing the biosecurity of crops in relation to migrating airborne pathogens and pests, `tropolink` could contribute to pipelines predicting probable future occupied areas. Such pipelines and accompanying web-based interactive tools for visualization have already been developed in the USA, particularly for soybean rust (https://soybean.ipmpipe.org/soybeanrust/, Isard et al., 2011) and cucurbit downy mildew (https://cdm.ipmpipe.org/, Ojiambo et al., 2011). Two potential solutions for including `tropolink` in such pipelines could be explored. First, we can determine the likely locations where an invasive species may be detected if it follows recent air-mass movements from already occupied areas. This could involve utilizing connectivities calculated by `tropolink` over the last N days, with the specific value of N depending on the context. Second, we can identify areas that should be monitored in the near future by predicting probable future air-mass movements from occupied regions. The second solution can be implemented in several ways. One approach is to employ forecast meteorological data sets that `HYSPLIT` can use instead of archived meteorological data. These forecast data set. allow for predictions ranging from 1 to 10 days, depending on the specific data set used. Alternatively, we can use connectivities calculated from historical data for the relevant period of interest (Choufany, Martinetti, Senoussi, et al., 2021; Choufany, Martinetti, Soubeyrand, & Morris, 2021, shown a certain stability of monthly connectivities across years, suggesting that connectivity is predictable based on past data).

To conclude, we discuss possible improvements of `tropolink`. This web application is grounded on meteorological data with 0.50°-resolution in latitude and longitude from September 2007 to June 2019, 0.25° from June 2019. If the resolution of these data may somewhat be considered as insufficient, they present the clear advantage of covering all the regions of the Earth. A future version of `tropolink` may exploit higher-resolution data available in certain parts of the world and compatible with `HYSPLIT` (Hernández-Ceballos et al., 2014). Concerning the uncertainty in air-mass trajectories, it can be presently taken into account by considering trajectory ensembles (Rolph et al., 2017), that is to say by noising the initial conditions (i.e., starting/arrival locations, times, and altitudes of trajectories) specified by the user (as we did in a very simple way in the sugar beet application where we computed connectivities for three starting altitudes). For further developing `tropolink` we may, (a) complement the options consisting of specifying a circular or geographic buffer around network nodes

by considering flexible polygons provided by the user (Choufany, Martinetti, Senoussi, et al., 2021). We may also (b) combine variables associated to the computation of air mass trajectories (e.g., the altitude of the air mass along its trajectory, its temperature, its humidity, the level of solar radiation…), land-use maps, and properties/behaviors of particles/insects (e.g., their shapes, their mass, their behavior with respect to the air mass conditions, their host preferences; Aylor, 2017; Isard & Gage, 2001) for weighting the trajectory contributions to connectivities and especially evaluating the likelihood of particle/insects release and deposition (Choufany, Martinetti, Senoussi, et al., 2021; Radici et al., 2022; M. Wang et al., 2021). We may also (c) generate archives incorporating `tropolink` specifications and outputs as well as associated metadata for facilitating the sharing of these data on public data repositories, and (d) offer the possibility to the users to share their studies directly within the application (e.g., a user could retrieve a trajectory study from another user and calculate new connectivities), as well as to merge trajectory studies and to split them, in particular to save computation costs. `R` codes prefiguring some of these perspectives (trajectory ensembles and weighted connectivities) are available in the wiki of the application as vignettes (https://forgemia.inra.fr/tropo-group/tropolink/-/wikis/Examples). We expect that the list of vignettes in the wiki will be enriched by the `tropolink` user community to share ideas for analyzes in `R`, `Python` and any other relevant language.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Trajectory and connectivity data generated for the preparation of this article, as well as the corresponding specifications (node coordinates, dates, and other `tropolink` options), are available in the public repository Zenodo (Soubeyrand, 2023). Potyvirus genetic data are also available from Zenodo (Alamil et al., 2019b) and FAW data are available from the FAMEWS global platform of the FAO (FAO, 2023).

Software availability statement: The web application `tropolink` is free but requires a login and a password that can be requested from the software wiki whose link is provided in the reference list (Richard et al., 2023).

Acknowledgments

We thank Maria Choufany, Christelle Lacroix, Christel Leyronas, Thierry Marcel, Cyril Dutech, Anne-Lise Boixel, Fréderic Suffert, Lydia Bousset-Vaslin, Benoît Marçais, and Cécile Robin for sharing thoughts about `tropolink`. We thank Loïc Houde for the administration of the computer cluster used by `tropolink`. We thank ITB for sharing sugar beet field locations and FAO for sharing the fall armyworm monitoring data base. The development of the application was partly funded by a Grant of the MathNum division of INRAE, a Grant of the BioSP research unit, the SMITID project (ANR-16-CE35-0006), the SPREE project (ANR-17-CE32-0004), the BEYOND project (ANR-20-PCPA-0002), and the SEPIM project (FranceAgriMer 3890396).

References

Aguilar-Vega, C., Fernández-Carrión, E., & Sánchez-Vizcaíno, J. M. (2019). The possible route of introduction of bluetongue virus serotype 3 into Sicily by windborne transportation of infected *Culicoides* spp. *Transboundary and Emerging Diseases*, *66*, 1665–1673. https://doi.org/10.1111/tbed.13201

Alamil, M., Hughes, J., Berthier, K., Desbiez, C., Thébaud, G., & Soubeyrand, S. (2019a). 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*(1775), 20180258. https://doi.org/10.1098/rstb.2018.0258

Alamil, M., Hughes, J., Berthier, K., Desbiez, C., Thébaud, G., & Soubeyrand, S. (2019b). R scripts and reformatted data analyzed by Alamil et al. (2019) [Dataset]. https://doi.org/10.5281/zenodo.2543673

Alamil, M., Thébaud, G., Berthier, K., & Soubeyrand, S. (2022). Characterizing viral within-host diversity in fast and non-equilibrium demo-genetic dynamics. *Frontiers in Microbiology*, *13*, 983938. https://doi.org/10.3389/fmicb.2022.983938

Aylor, D. E. (2003). Spread of plant disease on a continental scale: Role of aerial dispersal of pathogens. *Ecology*, *84*(8), 1989–1997. https://doi.org/10.1890/01-0619

Aylor, D. E. (2017). *Aerial dispersal of pollen and spores*. APS Press.

Bajardi, P., Poletto, C., Ramasco, J. J., Tizzoni, M., Colizza, V., & Vespignani, A. (2011). Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS One*, *6*(1), e16591. https://doi.org/10.1371/journal.pone.0016591

Barberán, A., Ladau, J., Leff, J. W., Pollard, K. S., Menninger, H. L., Dunn, R. R., & Fierer, N. (2015). Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(18), 5756–5761. https://doi.org/10.1073/pnas.1420815112

Beaunée, G., Vergu, E., & Ezanno, P. (2015). Modelling of paratuberculosis spread between dairy cattle farms at a regional scale. *Veterinary Research*, *46*, 1–13. https://doi.org/10.1186/s13567-015-0247-3

Bell, J. R., Alderson, L., Izera, D., Kruger, T., Parker, S., Pickup, J., et al. (2015). Long-term phenological trends, species accumulation rates, aphid traits and climate: Five decades of change in migrating aphids. *Journal of Animal Ecology*, *84*(1), 21–34. https://doi.org/10.1111/1365-2656.12282

Brooks, C. P., Antonovics, J., & Keitt, T. H. (2008). Spatial and temporal heterogeneity explain disease dynamics in a spatially explicit network model. *The American Naturalist*, *172*(2), 149–159. https://doi.org/10.1086/589451

Brown, J. K. M., & Hovmøller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, *297*(5581), 537–541. https://doi.org/10.1126/science.1072678

Choufany, M., Martinetti, D., Senoussi, R., Morris, C. E., & Soubeyrand, S. (2021). Spatiotemporal large-scale networks shaped by air mass movements. *Frontiers in Applied Mathematics and Statistics*, *6*, 602621. https://doi.org/10.3389/fams.2020.602621

Choufany, M., Martinetti, D., Soubeyrand, S., & Morris, C. E. (2021). Inferring long-distance connectivity shaped by air-mass movement for improved experimental design in aerobiology. *Scientific Reports*, *11*(1), 11093. https://doi.org/10.1038/s41598-021-90733-2

Clobert, J., Baguette, M., Benton, T. G., & Bullock, J. M. (2012). *Dispersal ecology and evolution*. Oxford University Press.
Cock, M. J., Beseh, P. K., Buddie, A. G., Cafá, G., & Crozier, J. (2017). Molecular methods to detect *Spodoptera frugiperda* in Ghana, and implications for monitoring the spread of invasive species in developing countries. *Scientific Reports*, *7*(1), 4103. https://doi.org/10.1038/s41598-017-04238-y
Committee of the Significance of International Transport of Air Pollutants. (2010). *Global sources of local pollution: An assessment of long-range transport of key air pollutants to and from the United States*. National Academies Press.
Draxler, R. R., & Hess, G. D. (1997). *Description of the HYSPLIT4 modeling system (revised in April 2020)* (Tech. Rep.). National Oceanic and Atmospheric Administration, ERL ARL-224.
Draxler, R. R., & Hess, G. D. (1998). An overview of the HYSPLIT4 modelling system for trajectories. *Australian Meteorological Magazine*, *47*, 295–308.
Draxler, R. R., & Rolph, G. (2010). *HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) model access via NOAA ARL READY website* (Vol. 25). NOAA Air Resources Laboratory. Retrieved from http://ready.arl.noaa.gov/HYSPLIT.php
Durr, P. A., Graham, K., & van Klinken, R. D. (2017). Sellers' revisited: A big data reassessment of historical outbreaks of bluetongue and African horse sickness due to the long-distance wind dispersion of *Culicoides midges*. *Frontiers in Veterinary Science*, *4*, 98. https://doi.org/10.3389/fvets.2017.00098
Early, R., González-Moreno, P., Murphy, S., & Day, R. (2018). Forecasting the global extent of invasion of the cereal pest *Spodoptera frugiperda*, the fall armyworm. *NeoBiota*, *40*, 25–50. https://doi.org/10.3897/neobiota.40.28165
Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is there a genetic paradox of biological invasion? *Annual Review of Ecology, Evolution, and Systematics*, *47*(1), 51–72. https://doi.org/10.1146/annurev-ecolsys-121415-032116
FAO. (2023). FAMEWS global platform [Dataset]. Retrieved from https://www.fao.org/fall-armyworm/monitoring-tools/famews-global-platform/en/
Gonzalez-Martin, C., Teigell-Perez, N., Valladares, B., & Griffin, D. W. (2014). The global dispersion of pathogenic microorganisms by dust storms and its relevance to agriculture. *Advances in Agronomy*, *127*, 1–41. https://doi.org/10.1016/B978-0-12-800131-8.00001-7
Hamilton, L. M., & 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. Retrieved from https://www.cabdirect.org/cabdirect/abstract/19671103048
Hernández-Ceballos, M., Skjøth, C., García-Mozo, H., Bolívar, J., & Galán, C. (2014). Improvement in the accuracy of back trajectories using WRF to identify pollen sources in southern Iberian Peninsula. *International Journal of Biometeorology*, *58*(10), 2031–2043. https://doi.org/10.1007/s00484-014-0804-x
Hufnagel, L., Brockmann, D., & Geisel, T. (2004). Forecast and control of epidemics in a globalized world. *Proceeding of the National Academy of Sciences of the United States of America*, *101*(42), 15124–15129. https://doi.org/10.1073/pnas.0308344101
Irwin, M., & Thresh, J. (1988). Long-range aerial dispersal of cereal aphids as virus vectors in North America. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, *321*, 421–446. https://doi.org/10.1098/rstb.1988.0101
Isard, S. A., Barnes, C., Hambleton, S., Ariatti, A., Russo, J., Tenuta, A., et al. (2011). Predicting soybean rust incursions into the North American continental interior using crop monitoring, spore trapping, and aerobiological modeling. *Plant Disease*, *95*(11), 1346–1357. https://doi.org/10.1094/pdis-01-11-0034
Isard, S. A., & Gage, S. H. (2001). *Flow of life in the atmosphere: An airscape approach to understanding invasive organisms*. Michigan State University Press.
Jeger, M. J., Pautasso, M., Holdenrieder, O., & Shaw, M. W. (2007). Modelling disease spread and control in networks: Implications for plant sciences. *New Phytologist*, *174*(2), 279–297. https://doi.org/10.1111/j.1469-8137.2007.02028.x
Kappen, L., & Straka, H. (1988). Pollen and spores transport into the Antarctic. *Polar Biology*, *8*(3), 173–180. https://doi.org/10.1007/bf00443450
Kellogg, C. A., & Griffin, D. W. (2006). Aerobiology and the global transport of desert dust. *Trends in Ecology & Evolution*, *21*(11), 638–644. https://doi.org/10.1016/j.tree.2006.07.004
Kenkre, V. M., & Giuggioli, L. (2021). *Theory of the spread of epidemics and movement ecology of animals: An interdisciplinary approach using methodologies of physics and mathematics*. Cambridge University Press.
Kim, C. S., Anthony, T. L., Goldstein, D., & Rytuba, J. J. (2014). Windborne transport and surface enrichment of arsenic in semi-arid mining regions: Examples from the Mojave Desert, California. *Aeolian Research*, *14*, 85–96. https://doi.org/10.1016/j.aeolia.2014.02.007
Lagos-Kutz, D., Voegtlin, D. J., Onstad, D., Hogg, D., Ragsdale, D., Tilmon, K., et al. (2020). The soybean aphid suction trap network: Sampling the aerobiological "Soup". *American Entomologist*, *66*(1), 48–55. https://doi.org/10.1093/ae/tmaa009
Leyronas, C., Morris, C. E., Choufany, M., & Soubeyrand, S. (2018). Assessing the aerial interconnectivity of distant reservoirs of *Sclerotinia sclerotiorum*. *Frontiers in Microbiology*, *9*, 2257. https://doi.org/10.3389/fmicb.2018.02257
Luquet, M., Sylvain, P., Buchard, C., Plantegenest, M., & Tricault, Y. (2023). *Predicting the seasonal flight activity of Myzus persicae*, the main aphid vector of Virus Yellows in sugar beet. Pest Management Science.
Martinetti, D., & Soubeyrand, S. (2019). Identifying lookouts for Epidemio-surveillance: Application to the emergence of *Xylella fastidiosa* in France. *Phytopathology*, *109*(2), 265–276. https://doi.org/10.1094/phyto-07-18-0237-fi
Masson, O., Steinhauser, G., Zok, D., Saunier, O., Angelov, H., Babić, D., et al. (2019). Airborne concentrations and chemical considerations of radioactive ruthenium from an undeclared major nuclear release in 2017. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(34), 16750–16759. https://doi.org/10.1073/pnas.1907571116
McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x
Meyer, M., Cox, J., Hitchings, M., Burgin, L., Hort, M., Hodson, D., & Gilligan, C. (2017). Quantifying airborne dispersal routes of pathogens over continents to safeguard global wheat supply. *Nature Plants*, *3*(10), 780–786. https://doi.org/10.1038/s41477-017-0017-5
Morris, C. E., Geniaux, G., Nédellec, C., Sauvion, N., & Soubeyrand, S. (2021). One health concepts and challenges for surveillance, forecasting and mitigation of plant disease beyond the traditional scope of crop production. *Plant Pathology*, *71*(1), 86–97. https://doi.org/10.1111/ppa.13446
Morris, C. E., Sands, D. C., Vinatzer, B. A., Glaux, C., Guilbaud, C., Buffiere, A., et al. (2008). The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *The ISME Journal*, *2*(3), 321–334. https://doi.org/10.1038/ismej.2007.113
Moslonka-Lefebvre, M., Finley, A., Dorigatti, I., Dehnen-Schmutz, K., Harwood, T., Jeger, M. J., et al. (2011). Networks in plant epidemiology: From genes to landscapes, countries, and continents. *Phytopathology*, *101*(4), 392–403. https://doi.org/10.1094/phyto-07-10-0192
Nathan, R., Katul, G. G., Bohrer, G., Kuparinen, A., Soons, M. B., Thompson, S. E., et al. (2011). Mechanistic models of seed dispersal by wind. *Theoretical Ecology*, *4*(2), 113–132. https://doi.org/10.1007/s12080-011-0115-3
Ojiambo, P., Holmes, G., Britton, W., Babadoost, M., Bost, S., Boyles, R., et al. (2011). Cucurbit downy mildew ipmPIPE: A next generation web-based interactive tool for disease management and extension outreach. *Plant Health Progress*, *12*(1), 26. https://doi.org/10.1094/php-2011-0411-01-rv

Pehlívan, S., & Atakan, E. (2022). First record of the fall armyworm, *Spodoptera frugiperda* (JE Smith, 1797)(Lepidoptera: Noctuidae) in Türkiye. *Çukurova Tarım ve Gıda Bilimleri Dergisi*, *37*, 139–145. Retieved from https://dergipark.org.tr/en/download/article-file/2685359

Picard, C., Dallot, S., Brunker, K., Berthier, K., Roumagnac, P., Soubeyrand, S., et al. (2017). Exploiting genetic information to trace plant virus dispersal in landscapes. *Annual Review of Phytopathology*, *55*(1), 139–160. https://doi.org/10.1146/annurev-phyto-080516-035616

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

Pretorius, I., Schou, W. C., Richardson, B., Ross, S. D., Withers, T. M., Schmale, D. G., III, & Strand, T. M. (2023). In the wind: Invasive species travel along predictable atmospheric pathways. *Ecological Applications*, *33*(3), e2806. https://doi.org/10.1002/eap.2806

Qi, A., Dewar, A. M., & Harrington, R. (2004). Decision making in controlling virus yellows of sugar beet in the UK. *Pest Management Science*, *60*(7), 727–732. https://doi.org/10.1002/ps.871

Radici, A., Martinetti, D., & 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*(6), 064045. https://doi.org/10.1088/1748-9326/ac73aa

Reisen, F., Duran, S. M., Flannigan, M., Elliott, C., & Rideout, K. (2015). Wildfire smoke and public health risk. *International Journal of Wildland Fire*, *24*(8), 1029–1044. https://doi.org/10.1071/wf15034

Richard, H., Fouillat, Y., Lercier, D., Martinetti, D., Morris, C., & Soubeyrand, S. (2023). tropolink (Version 1.0) [Software]. Retrieved from https://tropolink.fr. Associated gitlab repository: https://forgemia.inra.fr/tropo-group; Accompanying wiki: https://forgemia.inra.fr/tropo-group/tropolink/-/wikis; code components specific to tropolink licensed under GPLv3: https://www.gnu.org/licenses/gpl-3.0.html

Rolph, G., Stein, A., & Stunder, B. (2017). Real-time environmental applications and display system: READY. *Environmental Modelling & Software*, *95*, 210–228. https://doi.org/10.1016/j.envsoft.2017.06.025

Saubin, M., Coville, J., Xhaard, C., Frey, P., Soubeyrand, S., Halkett, F., & Fabre, F. (2023). Inferring invasion determinants with mechanistic models and multitype samples. bioRxiv PPR635737.

Schmale, D. G., III, & Ross, S. D. (2015). Highways in the sky: Scales of atmospheric transport of plant pathogens. *Annual Review of Phytopathology*, *53*(1), 591–611. https://doi.org/10.1146/annurev-phyto-080614-115942

Schneider, T. (2006). The general circulation of the atmosphere. *Annual Review of Earth and Planetary Sciences*, *34*(1), 655–688. https://doi.org/10.1146/annurev.earth.34.031405.125144

Soubeyrand, S. (2023). Air mass trajectory and connectivity data generated with tropolink [Dataset]. https://doi.org/10.5281/zenodo.7962177

Soubeyrand, S., Laine, A., Hanski, I., & Penttinen, A. (2009). Spatio-temporal structure of host-pathogen interactions in a metapopulation. *The American Naturalist*, *174*(3), 308–320. https://doi.org/10.1086/603624

Stein, A., Draxler, R. R., Rolph, G. D., Stunder, B. J., Cohen, M., & Ngan, F. (2015). NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society*, *96*(12), 2059–2077. https://doi.org/10.1175/bams-d-14-00110.1

Strona, G., Carstens, C. J., & Beck, P. S. (2017). Network analysis reveals why *Xylella fastidiosa* will persist in Europe. *Scientific Reports*, *7*, 1–8. https://doi.org/10.1038/s41598-017-00077-z

Wang, J., Huang, Y., Huang, L., Dong, Y., Huang, W., Ma, H., et al. (2023). Migration risk of fall armyworm (*Spodoptera frugiperda*) from north Africa to southern Europe. *Frontiers in Plant Science*, *14*, 1141470. https://doi.org/10.3389/fpls.2023.1141470

Wang, M., Kriticos, D. J., Ota, N., Brooks, A., & Paini, D. (2021). A general trait-based modelling framework for revealing patterns of airborne fungal dispersal threats to agriculture and native flora. *New Phytologist*, *232*(3), 1506–1518. https://doi.org/10.1111/nph.17659

Ward, R. D. C. (1916). The prevailing winds of the United States. *Annals of the Association of American Geographers*, *6*(1), 99–119. https://doi.org/10.1080/00045601609357048

Weber, K., Eliasson, J., Vogel, A., Fischer, C., Pohl, T., Van Haren, G., et al. (2012). Airborne in-situ investigations of the Eyjafjallajökull volcanic ash plume on Iceland and over north-western Germany with light aircrafts and optical particle counters. *Atmospheric Environment*, *48*, 9–21. https://doi.org/10.1016/j.atmosenv.2011.10.030

Westbrook, J., Nagoshi, R., Meagher, R., Fleischer, S., & Jairam, S. (2016). Modeling seasonal migration of fall armyworm moths. *International Journal of Biometeorology*, *60*(2), 255–267. https://doi.org/10.1007/s00484-015-1022-x

Williams, C. G. (2010). Long-distance pine pollen still germinates after meso-scale dispersal. *American Journal of Botany*, *97*(5), 846–855. https://doi.org/10.3732/ajb.0900255

Williams, C. G., & Barneoud, P. (2021). Live pine pollen in rainwater: Reconstructing its long-range transport. *Aerobiologia*, *37*(2), 333–350. https://doi.org/10.1007/s10453-021-09697-5

Wu, M.-F., Qi, G.-J., Chen, H., Ma, J., Liu, J., Jiang, Y.-Y., et al. (2022). Overseas immigration of fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), invading Korea and Japan in 2019. *Insect Science*, *29*(2), 505–520. https://doi.org/10.1111/1744-7917.12940

Zadoks, T. (1967). Epidemiology of wheat rust in Europe. *International Journal of Process Management and Benchmarking*, *13*(1), 29–46. https://doi.org/10.1080/05331846709432232

References From the Supporting Information

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*(6), 989–993. https://doi.org/10.1093/aesa/79.6.989

Irwin, M. E., Kampmeier, G. E., & Weisser, W. W. (2007). Aphid movement: Process and consequences. In H. F. van Emden & R. Harrington (Eds.), *Aphids as crop pests* (pp. 153–186). CABI.

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

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* (Tech. Rep.). IGEPP, INRAE, Institut Agro, Univ Rennes.

Parry, H. R. (2013). Cereal aphid movement: General principles and simulation modelling. *Movement Ecology*, *1*, 1–15. https://doi.org/10.1186/2051-3933-1-14