Early-detection surveillance for stem rust of wheat: insights from a global epidemic network based on airborne connectivity and host phenology

Andrea Radici, Davide Martinetti and Daniele Bevacqua

1 BioSP UR 546, INRAE, 228 route de l’Aérodrome, CS 40509 Domaine Saint-Paul—Site Agroparc, 84914 Avignon, France
2 PSH UR 1115, INRAE, 228 route de l’Aérodrome, CS 40509 Domaine Saint-Paul—Site Agroparc, 84914 Avignon, France
* Author to whom any correspondence should be addressed.
E-mail: davide.martinetti@inrae.fr

Abstract
Stem rust of wheat, caused by the airborne pathogen *Puccinia graminis*, is a re-emerging crop disease representing a major concern to global food security. Potential long-distance transport by wind over a worldwide distributed host represents a challenge to effective surveillance and control of this disease. To monitor this disease, we have created a global epidemic network for stem rust of wheat combining (a) Lagrangian simulations of air-mass trajectories computed with the NOAA’s HYSPLIT model; (b) land use from the Map Spatial Production Allocation Model and (c) meteorological and environmental conditions that are known to affect bio-physical processes involved in the biology of *P. graminis* spores. Our findings are in agreement with the well known north-American ‘*Puccinia* pathway’ and suggest the existence of other sub-continental pathways at the global scale. We used network theory to conceive surveillance strategies aimed at early detection of outbreaks while minimizing the number of nodes to be surveilled (also referred to as sentinels). We found that the set cover algorithm, due the high average connectivity of the network (density = 0.4%), performs better than a number of other network metrics and permits us to identify an optimal sentinel set (1% of the network nodes) to surveil 50% of the network. Our results also show that effective surveillance plans for stem rust of wheat can be designed, but that they need to account for the actual geographical scale of the underlying epidemiological process and call for an international and trans-boundary approach.

1. Introduction
Epidemics caused by airborne pathogens represent an invertebrate challenge to agricultural management [1]. Moreover, the transition to monocultures characterized by low genetic diversity has made the global farming system less resilient to pathogens [2]. Airborne pathogens, capable of long-distance transport, create a network of connections among globally diffused crops. For example, *Puccinia graminis*, causing stem rust of wheat, is seasonally dispersed from northern Mexico up to Canada, along the ‘*Puccinia* pathway’ [3]. On the other hand, a single-leap event, the hurricane Ivan, transported *Phakopsora pachyrhizi* spores from Colombia to Alabama, introducing soybean rust in North America [4, 5]. The risk of losses in food production requires to take action against the diffusion of alarming pathogens, such as *P. graminis* [2, 6, 7] and species of the same genus [8]. Its aerobiology has been largely studied in the last century [9–12] and experimental procedures have been recently developed for studying three-dimensional transport and detection of spores [13, 14]. However, the use of mathematical models to simulate spore transport by wind at the continental scale is very recent [15–20]. In fact, the use of such models would allow researchers, policy makers and farmers to design surveillance strategies [21–23].
2.2. Geographic domain and air mass trajectories
We extracted the worldwide wheat distribution from the MapSPAM database [29] and computed the percentage of wheat cover on a regular grid with a resolution of 0.5°. We accounted only for cells containing at least 2% of wheat land cover. The resulting database contains 7814 cells (see section SI 1.1).
We calculated backward air-mass Lagrangian three-dimensional trajectories of 120 hours [16] with the HYSLPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) model [30] from 1st January 2013 to 31st December 2018, using atmospheric GDAS data at a spatial resolution of 0.5°. We ran simulations from the centroids of each cell at 00:00, 06:00, 12:00 and 18:00 UTC +0, at an above-ground altitude set to the minimum between the base of the cloud and the mixed layer depth. Along each simulated trajectory, we recorded atmospheric variables at an hourly frequency. Overall, out of more than $6.8 \times 10^7$ potential trajectories, we computed only those $1.6 \times 10^6$ satisfying the criteria of host availability and environmental suitability specified in sections 2.3.1 and 2.3.4 (see also section SI 1.2).

2.3. The connectivity network
We built a time-variant connectivity network for stem rust explicitly considering host (i.e. wheat) phenological phases, environmental conditions affecting spore aerobiology [31] and infection. In this network, nodes represent cells composing the worldwide wheat producing regions, while edges are computed from air-mass trajectories in order to model the dispersal of airborne spores. The nodes (or cells) of the networks remain fixed, while edges are re-computed for each of the 8764 (4 times a day for 6 years) simulations.

More specifically, edges result from the application of a set of biophysical filters to identify those air-mass movements that correspond to transport events (figure 1). We assume that, for a given time $t$, an edge exists between an arrival cell $j$ and any cell $i$ if the following conditions are satisfied:

(a) In cell $i$, which at time $t-x$ ($x \in \{1, \ldots, 120\}$ h) is crossed by the air-mass trajectory arriving at $j$ at time $t$:
   - The host is ‘available’, i.e. it is present and in a favourable phenological state for infection and sporulation, see section 2.3.1.
   - Environmental conditions are compatible for spore release, see section 2.3.2.
   - The altitude of the trajectory is lower than the planetary boundary layer.

(b) Environmental conditions along the trajectory between $i$ and $j$ allow spore survival, see section 2.3.3.

(c) The host in cell $j$ is ‘susceptible’, which means:
   - The host is ‘available’.

2. Materials and methods

2.1. Case study

*Puccinia graminis* f. sp. *tritici* is a heteroecious airborne fungal pathogen responsible for stem rust of wheat. Wheat, the main host of *P. graminis*, constitutes a staple food for a great proportion of human population. It covers $2.15 \times 10^8$ hectares worldwide (1.4% of earth surface), representing the most abundant agricultural type of land cover [25].

In recent decades, yield losses caused by stem rust have been limited by planting resistant cultivars and fungicides application. Nonetheless, the emergence of new strains overcoming plant resistance may provoke severe yield losses, accounting for up to 50%–90% of the wheat production at the regional scale [19, 26]. An extraordinary outbreak in Ethiopia in 2013–2014 caused a complete yield loss in some cultivars [27], while concerns were raised in Europe after the detection of new virulent strains capable of infecting previously resistant cultivars [28].

Meyer et al [15] assessed the risk of long distance dispersal of *P. graminis* in East Africa and the Middle East via a Lagrangian particle dispersion model. The same modelling framework was then used to investigate the possible origin of virulent strains found in Ethiopia [16]. Allen-Sadder et al [18] developed a decision support system integrating spore dispersal to help optimize fungicide allocation. Prank et al [19] investigated the impacts of climate change on worldwide spores transport patterns. Sutrave et al [24] designed a network-based surveillance-system for *P. pachyrhizi*, at the subcontinental scale, by including wind direction and intensity. Despite these advances, to our knowledge, models of spore dispersal have never been coupled to network analysis to design optimal surveillance strategies for *P. graminis* at the global scale.

In this current study we constructed a time-variant connectivity network for *P. graminis* spores at the worldwide scale explicitly considering wind patterns, host phenology and meteorological conditions affecting spores aerobiology. We used biophysical models and graph theory to identify the most susceptible worldwide regions and to reconstruct the epidemic movement through different subcontinents. We validated our findings using available knowledge regarding the North-American ‘Puccinia pathway’ [3]. Lastly, we identified those nodes of the network (i.e. regions of the world) that, when monitored, enable early detection of an outbreak. We are confident that our approach is sufficiently generic to be applied to other airborne plant pathogens, provided that basic knowledge on the pathogen aerobiology, host physiology and distribution are available.

In *Puccinia graminis* f. sp. *tritici* a stable food for a great proportion of human population. It covers $2.15 \times 10^8$ hectares worldwide (1.4% of earth surface), representing the most abundant agricultural type of land cover [25].

In recent decades, yield losses caused by stem rust have been limited by planting resistant cultivars and fungicides application. Nonetheless, the emergence of new strains overcoming plant resistance may provoke severe yield losses, accounting for up to 50%–90% of the wheat production at the regional scale [19, 26]. An extraordinary outbreak in Ethiopia in 2013–2014 caused a complete yield loss in some cultivars [27], while concerns were raised in Europe after the detection of new virulent strains capable of infecting previously resistant cultivars [28].

2. Materials and methods

2.1. Case study

*Puccinia graminis* f. sp. *tritici* is a heteroecious airborne fungal pathogen responsible for stem rust of wheat. Wheat, the main host of *P. graminis*, constitutes a staple food for a great proportion of human population. It covers $2.15 \times 10^8$ hectares worldwide (1.4% of earth surface), representing the most abundant agricultural type of land cover [25].

In recent decades, yield losses caused by stem rust have been limited by planting resistant cultivars and fungicides application. Nonetheless, the emergence of new strains overcoming plant resistance may provoke severe yield losses, accounting for up to 50%–90% of the wheat production at the regional scale [19, 26]. An extraordinary outbreak in Ethiopia in 2013–2014 caused a complete yield loss in some cultivars [27], while concerns were raised in Europe after the detection of new virulent strains capable of infecting previously resistant cultivars [28].

2.2. Geographic domain and air mass trajectories
We extracted the worldwide wheat distribution from the MapSPAM database [29] and computed the percentage of wheat cover on a regular grid with a resolution of 0.5°. We accounted only for cells containing at least 2% of wheat land cover. The resulting database contains 7814 cells (see section SI 1.1).
We calculated backward air-mass Lagrangian three-dimensional trajectories of 120 hours [16] with the HYSLPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) model [30] from 1st January 2013 to 31st December 2018, using atmospheric GDAS data at a spatial resolution of 0.5°. We ran simulations from the centroids of each cell at 00:00, 06:00, 12:00 and 18:00 UTC +0, at an above-ground altitude set to the minimum between the base of the cloud and the mixed layer depth. Along each simulated trajectory, we recorded atmospheric variables at an hourly frequency. Overall, out of more than $6.8 \times 10^7$ potential trajectories, we computed only those $1.6 \times 10^6$ satisfying the criteria of host availability and environmental suitability specified in sections 2.3.1 and 2.3.4 (see also section SI 1.2).

2.3. The connectivity network
We built a time-variant connectivity network for stem rust explicitly considering host (i.e. wheat) phenological phases, environmental conditions affecting spore aerobiology [31] and infection. In this network, nodes represent cells composing the worldwide wheat producing regions, while edges are computed from air-mass trajectories in order to model the dispersal of airborne spores. The nodes (or cells) of the networks remain fixed, while edges are re-computed for each of the 8764 (4 times a day for 6 years) simulations.

More specifically, edges result from the application of a set of biophysical filters to identify those air-mass movements that correspond to transport events (figure 1). We assume that, for a given time $t$, an edge exists between an arrival cell $j$ and any cell $i$ if the following conditions are satisfied:

(a) In cell $i$, which at time $t-x$ ($x \in \{1, \ldots, 120\}$ h) is crossed by the air-mass trajectory arriving at $j$ at time $t$:
   - The host is ‘available’, i.e. it is present and in a favourable phenological state for infection and sporulation, see section 2.3.1.
   - Environmental conditions are compatible for spore release, see section 2.3.2.
   - The altitude of the trajectory is lower than the planetary boundary layer.

(b) Environmental conditions along the trajectory between $i$ and $j$ allow spore survival, see section 2.3.3.

(c) The host in cell $j$ is ‘susceptible’, which means:
   - The host is ‘available’.
A simplified overview of the transport process between a source (i.e., release) and a sink (i.e., arrival) cell. Pathogen is assumed to be available in the release cell when the host is ‘available’. Susceptibility in the arrival cell is determined by the concurrence of host ‘availability’ and environmental conditions favorable to infection. Conditions for spore release, transport and deposition are summarized in the three boxes along the trajectory.

- Environmental conditions are favourable for spore deposition and host infection, see section 2.3.4.

2.3.1. Host availability
Depending on local climate, wheat is cultivated and harvested at different times across the world. We used a growing degree-day model [32, 33] to compute the period of host availability for each year in each cell of the domain as a function of the year temperature conditions. We used a model conceived for the US [32] and we adjusted it for the Southern Hemisphere (i.e., initializing it at July 1st). For tropical countries, we consulted case by case the calendar of the prevailing season provided in the FAO country briefs [25] (see section SI 1.3 (available online at stacks.iop.org/ERL/17/064045/mmedia)).

2.3.2. Spore release
Spore release in the air column is promoted under unstable atmospheric conditions [34, 35]. Following Meyer et al [15], we assumed that release occurs between 9:00 and 15:00, provided precipitations are lower than 2.54 mm h$^{-1}$ [18]. Furthermore, the rate of spore release is affected by temperature [19]. Eventually, a probability $P_1(T)$ of release is defined as a function of temperature (see section SI 1.5).

2.3.3. Aerial spore transport and survival
Spores are released at canopy level, but only a fraction actually enters in the atmospheric layers leading to long-distance transportation. The rest may be dispersed in the air column within few kilometers under turbulent atmospheric conditions [34, 36]. Hence, we considered as suitable release sites those that are crossed by an air-mass trajectory within the mixed layer depth (therefore, able to drag spores distributed in the air column). Furthermore, during the aerial phase, spores must endure critical environmental conditions [34]. The limiting factors affecting the survival of $P. graminis$ spores are the exposure to UV radiation [15] and the washout by rain. We then defined survival probabilities to UV radiation $P_2(UV)$ and to washout $P_3(R)$ (see section SI 1.5).

2.3.4. Spore deposition and host infection
Since precipitation is the main responsible for spore deposition and infection [34, 37–41], we assumed that an edge can point an arrival cell $j$ at time $t$ only if precipitation occurs. In fact, even if spore deposition might occur also in dry conditions [42, 43], wet deposition provides a better environment for the development of infection and it is usually treated as the most important element. Following deposition, infection requires specific environmental conditions [15, 18, 44]: in the three following days, spores should enter the phase of germination, requiring dark conditions and mild temperatures, and appressorium formation, requiring sunlight and warm temperature [39]. We then used these conditions to identify times and sites where infection can occur after deposition (see section SI 1.4).
2.3.5. Time-variant connectivity networks

In order to represent the connectivity between wheat producing regions, we took advantage of the mathematical formalism of networks [45], by describing the global epidemic network of stem rust via 6-hourly connectivity networks \( A^t \) whose elements \( A^t_{ij} \) indicate the probability \( P^t_{ij} \) of a successful transport event occurring at time \( t \) from node \( i \) to node \( j \). The probability \( P^t_{ij} \) is computed as the product of the mutually independent probabilities \( P^t_i(T) \) of sporulation at release site, \( P^t_j(UV) \) of survival to harmful UV radiation and \( P^t_j(R) \) of washout, provided the concurrence of host availability at site \( i \) and precipitation at site \( j \). Finally, we aggregate the 8764 networks to obtain a weighted, directed air-mass connectivity network \( C \) (see section SI 1.5) and used a cluster detection algorithm [46] to identify groups of nodes with similar connectivity patterns.

2.4. Model validation: recovering the ‘Puccinia pathway’ in North America

Since the direct observation of long-distance spores transport is beyond current technological capacities, we tested the validity of our model checking if we could use it to identify the well known North-American ‘Puccinia pathway’.

We considered the yearly northward movement of the stem rust infection front in North America by tracking the average onset date (±1 standard deviation) during the 1922–1992 period along the 97th meridian west [3]. Then, we computed the weekly cumulative in-strength values of the nodes located along the same meridian in the US from the weekly connectivity networks \( C_w \). The in-strength is a network metric considering the sum of the weights of the edges pointing a node \( i \). In this case, the in-strength of a node \( i \) corresponds to the sum of the average weekly frequencies of connection from the other nodes toward \( i \). We used the cumulative version of this metric, by assuming that local emergence of a disease is observable after sufficiently abundant inoculum has been deposited in that node. We graphically compared observed onset dates and average weekly cumulative in-strength in figure 3 (see section SI 1.6). Here, we built the frequency distribution of the cumulative in-strength values intersected by 1922–1992 onset date observations and computed the values corresponding to the interquartile range. Lastly, cumulative in-strength values falling in the interquartile range are shaded on the heat map.

2.5. Network surveillance

2.5.1. Exploring ‘Puccinia pathways’ worldwide

In order to reconstruct the trajectories of propagation of stem rust outbreaks at world scale, we analyzed how the center of mass of the monthly connectivity networks \( C_m \) in-strength moves across different continents. Hence, for a given set of nodes representing a subcontinent (and the corresponding geographical coordinates), we defined the center of mass of the monthly connectivity networks in-strength as the geographical point whose coordinates are given by the average latitude and longitude weighted by the nodes monthly in-strengths.

2.5.2. Definition of efficient surveillance strategies

We considered the problem of establishing a reduced set of sentinels, i.e. nodes where the presence of the pathogen is monitored systematically, that should guarantee the largest coverage of the domain and provide an early-warning system for the appearance of new pathogen strains [22]. First of all, we defined the coverage of a node \( i \) as the set of nodes that points towards \( i \), under the assumption that monitoring the presence of the pathogen in \( i \) implies observing all those nodes that are pointing to it. In this case, node \( i \) is referred as the sentinel of its coverage. We considered hence the network \( C \) generated by considering those edges of the yearly connectivity networks \( C_y \) recurring at least three times over a 4 year interval 2013–2016 (i.e. \( \geq 75\% \) of the times), in order to account only for the most frequent connections. The problem of finding the smaller set of sentinels that guarantees the complete coverage is formally equivalent to the set cover problem in graph theory, that happens to have NP-complete computational complexity [24, 47]. Nonetheless, we used an iterative greedy algorithm, providing a sub-optimal minimum sentinel set. To validate the efficiency of the sentinel set, we assessed its performance in terms of ratio of surveilled domain using the network obtained by the intersection of the yearly connectivity networks of 2017 and 2018. We compared such performances with the ones given by sets of nodes chosen via other network metrics, namely in-strength, betweenness [48], PageRank [49] and random walk generalized accessibility [50], calculated on the aggregated 2013–2016 networks, and a set of 20 random samplings of nodes (see definitions and procedures in section SI 1.7).

2.5.3. Measuring early-detection performance of the sentinel set

Finally, we assessed the performance of different sentinel sets in terms of early detection by means of a compartmental Susceptible-Infected (SI) model based on the intersections of the weekly networks \( C_w \) in the years 2017 and 2018. Within this framework, a node can be either susceptible (S) or infected (I), its state depending on the state of the neighbors in the previous time steps. Namely, in this simplified approach, we assume that, in a given time step \( t \), a node pass from the state S to I if it has at least one neighbor in state I at time \( t-1 \), with no recovery. We run 7814 model simulations, each time letting the outbreak start from a different node (inoculum).

Then, we defined the Disease detection ratio (DDR) of a given sentinel set as the fraction of the total number
of simulations for which the sentinel set intercepted the epidemics at least once before the end of the fourth iteration (i.e. before one month after the first node has been inoculated). This new metric allows to compare the performance of different sentinel sets even when they fail to achieve detection within the first month, something that may happen when the first inoculated node is located in a rather isolated part of the network (see section SI 1.7).

3. Results

3.1. Worldwide susceptibility and connectivity patterns

The duration of host susceptibility periods, i.e. the concurrence of the host availability and environmental conditions favorable to infection, varies considerably across the world (figures 2(a) and SI6). The majority of cells is susceptible for a period between a week and a month. The nodes where susceptibility occurs for more than a month per year are located in northeastern America, southern Brazil and Paraguay, central Europe and central China. The only nodes with susceptibility lasting for more than three months per year are located in Ethiopia (see section SI 2.1).

Regions located in the Northern Hemisphere are generally well connected, in such a way that Europe, Asia and North Africa create a unique connected component (figure 2(b)). In spite of the obstacle represented by the Atlantic Ocean, extremely long distance connections may occur from North America to the Mediterranean basin. Conversely, clusters in the Southern Hemisphere are isolated between them, although internally connected. A first representation of the role played by each node within the epidemic network is given by the sum of the probabilities associated to trajectories going into (in-strength: figure 2(c)) and out from (out-strength: figure 2(d)) that node. In biological terms, in-strength (out-strength) is a proxy of the extent to which a region acts as a sink (source) of spore. In-strength appears to be higher in specific regions located in Europe and in the northeastern US, southern Brazil, the Himalaya and central China. Some regions exhibit great variability within relatively short distances, namely Ethiopia, Middle East and Central Asia. Other regions show a smoother and regular gradient, like the US. In this sense, a continuum of intermediate in-strength values extends from Europe to Western Siberia. Out-strength separates more sharply those regions characterized by high and low values and, in particular, Europe is characterized by large regions associated with high out-strength values.

3.2. Reconstructing Puccinia pathways in North America and elsewhere

Our global epidemic network permitted to recover the well known North-American Puccinia pathway. In figure 3 we compare the average onset date of outbreak along the 97th meridian west observed between 1922 and 1992 in North America [3], with the cumulative in-strength of network nodes along the same meridian. Most of the observed onset dates occur within a certain interval of values of cumulative in-strength (the interquartile range [0.6, 0.9]), suggesting that the cumulative in-strength network metric can proxy the spatio-temporal progression of the ‘Puccinia pathway’ in North America. In other words, in this continent one would expect to observe the first signs of an epidemics when the cumulative in-strength has values between 0.6 and 0.9.

Globally, we found that the airborne transport estimated via the center of mass of the monthly in-strength always moves poleward from tropical and temperate regions (figures 4 and SI10), except for the Ethiopian pathway, that follows a southward movement even if it is located in the Northern Hemisphere, likely due to its cropping calendar.

3.3. Optimal sentinel set

The sentinel set obtained on the 2013–2016 network using the set cover algorithm allows for a complete coverage of the worldwide wheat producing regions by monitoring 1007 nodes, i.e. less than 13% of the total (figure SI12). Moreover, a reasonable coverage of 50% of the domain can be obtained monitoring 64 nodes (figure 5(a)), i.e. less than 1% of the total. The first selected nodes are those assuring the greatest coverage, and they do not distribute uniformly across continents (figure SI12—see section SI 2.4).

The sentinel set selected to optimally cover the 2013–2016 epidemic network, via the set cover algorithm, provides satisfactory results also when applied to surveil the epidemic network obtained for the 2017–2018 period (figure 5(b)). In fact, it provides a coverage of 50% of the domain by monitoring 114 nodes (1.5% of the total). For comparison, the same coverage of 50% would require 234 (3%) random sentinels (on average: interquartile range [227; 242]), 475 (6.1%) sentinels if ordered for increasing values of betweenness, or 611 (7.8%) of PageRank.

3.4. Early detection capabilities of the sentinel set

In terms of DDR, the set cover algorithm outperformed all the other methods for sentinel sets containing between 20 and 650 nodes (figure 5(c)). In-strength provides better results for very small sentinel sets, while random sentinel sets are more suitable when larger sentinel sets have to be designed. The DDR associated to 275 sentinels is 19.2%, which means that 275 sentinels are able to detect an epidemic process started from any of 19.2% of the world producing regions within a month. A similar DDR of
19.1% is obtained with 350 nodes chosen according to their betweenness, or more than 500 nodes according to their PageRank. Between 350 and 500 random sentinels are needed to achieve a DDR around 18%–22%.

We estimated the DDR associated to larger values of detection delay (3 month, 6 month, 1 year), showing that the set cover strategy improves its performances against the random sentinels sets (figure SI14).
Figure 3. Horizontal error bars, denoted with capital letters, represent the observed average onset dates (± 1 standard deviation) of stem rust epidemics along the 97th meridian west in the 1922–1992 period (source [3]). The heat map in the background represents the weekly cumulative in-strength of nodes along the same meridian. In the right panel by the side of the legend, we computed the interquartile range of the cumulative in-strength values intersected by the observed average onset dates and their error bars. The shaded area on the heat map highlights the intervals where the cumulative in-strength value falls within the interquartile range [0.6, 0.9]. The top-right panel represents the cells crossed by the 97th meridian and the actual wheat distribution in North America.

Figure 4. Panel (a) shows the poleward speed of the center of mass of the in-strength of different sub-continents. Panels (b)–(e) depict the monthly pathways for North America, China, South America and Western Europe-Maghreb, respectively.
4. Discussion

In this study we proposed an original modelling framework based on air mass movements, network analysis, meteorology, ecology and plant physiology to describe the global epidemic network of stem rust of wheat. Eventually, we used the model to identify previously unknown pathways of disease spread at the globe scale and to define a set of sentinels that should be primarily surveyed to achieve early detection of future outbreaks.

4.1. The most susceptible regions

Our results indicate that the duration of the susceptibility period greatly varies among regions (figures 2(a) and SI6). Eastern US, southern Brazil, central Europe and China, which are important areas of wheat production, experience the longest susceptibility periods. On the other hand, other important wheat-producing areas from the Middle East up to the Indian subcontinent have shorter periods of susceptibility. The reason could be the temporal mismatch between the occurrence of environmental conditions for infection and of host availability. This is particularly true for India and in Brazil, where, despite a long-lasting occurrence of environmental conditions for infection, the susceptibility period is constrained by relatively short host availability periods (figure SI6). Ethiopia, which is characterized by exceptional climatic diversity for tropical latitudes [51], exhibits a very heterogeneous behavior, with nodes that are never susceptible in close proximity to others that are highly susceptible (91–130 days per year). Since both host availability and favorable environmental conditions for spore deposition and host infection are climate dependent processes, it is likely that global crop epidemic dynamics will be affected by predicted climate changes [52]. Prank et al [19] predicted a worldwide increase in sporulation period under a RCP8.5 climate change scenario in 2100—potentially compensated by a general decrease of probability of germination. On the other hand, laboratory experiments conducted on other fungal species suggested that future climate conditions would rather inhibit sporulation, but increase mycelium growth rate [53].
4.2. On the construction of a global epidemic network

To construct an epidemic network we had to explicitly account for biophysical conditions influencing spore aerobiology. This filtering enabled to discard uninformative air-mass connections and to reduce the complexity of the network, making it simultaneously more significant and manageable.

In principle, by tracking the movement of an air mass for an indefinite long time, one would obtain a completely connected network, i.e. a network with edge density equal to 100%. On the other hand, our epidemic network has an edge density equal to 3.7%, whereas the network of only the most recurring connections (i.e. those connections found at least 3 times out of 4 in the yearly networks 2013–2016) has a density of 0.4%. For comparison, previous studies using a similar technique but filtering only over a 48 hours time [54] obtained a density of 28% over the Mediterranean basin for the period 2011–2017. In essence, our filtering allows to make epidemic networks more manageable and informative, presenting itself as alternative to other network-based methods for the extraction of the truly relevant connections [55].

4.3. From network science to global crop protection

Our results on the frequency of connections (figure 2) indicate that regions in the Northern Hemisphere are more densely connected than in the Southern one. This can be explained by the fact that 90% of wheat fields are located in the Northern Hemisphere while the clusters below the Tropic of Cancer, i.e. South America, South Africa, Ethiopia and Australia are thousands of kilometers away from each other. Furthermore, our prediction of possible long-distance connections from North America to Europe, and from Europe towards Central Asia and Russia, are in accordance with previous theoretical and empirical findings [17, 19, 20, 56]. On the other hand, due to the time lag in the respective susceptibility seasons, no connection between the Northern and Southern Hemispheres is predicted by our epidemic network. For such a connection to exists, it is necessary the presence of a ‘green bridge’ [15] in a tropical or subtropical region located between the two hemispheres. This could be the case of Ethiopia, where, according to the FAO country brief [25], there exist two wheat growing seasons: Belg, from February to July and Meher, from May to December. Furthermore, according to our estimates on the host susceptibility duration (figure 2(a)), Ethiopia features some of the nodes with the longest susceptibility period (more than 90 days/year), whereas previous studies already identified this country (and more broadly East Africa and Yemen) as an important stepping stone [56] for the long distance transport of rust spores along the Rift Valley [15].

The in- and out-strength maps of the epidemic network (figure 2(c)) provide an overview of the epidemic role played by each region, sink rather than source of spores or both. Our results indicate that out-strength tends to be stronger towards the equator compared to in-strength, in line with previous findings [3] about the ‘green’ and ‘golden’ wave: regions closer to tropics are the first where infection is possible and therefore they are more likely to act as sources of spores rather than sinks. Among the recent seasonal epidemic events moving poleward, the 2013 regional outbreak of stem rust in Germany seems to have moved to Denmark and Sweden later in the season [7].

If one wanted to use our approach do derive effective measures for global crop protection it is worthy to note that in our exercise we privileged generality over local accuracy. Our assumptions to determine host availability and susceptibility are necessarily oversimplifications of a complex reality, as required by models. For instance, we neglected the presence of alternative hosts, such as spring wheat, whose cropping calendars would increase the host susceptibility period and the frequency of long distance connections. Also, we neglected the existence of the secondary hosts of P. graminis, the barberry (Berberis vulgaris), which is necessary for sexual reproduction, facilitating the emergence of new strains. It has been successfully eradicated in Western Europe and North America in the last century [7]. Yet, it is present in other regions which turns to be sources of new strains [27, 57] and it has been recently reintroduced in Europe [58].

Our model permitted to design an optimal surveillance strategy capable of covering 50% of wheat cultivated lands while monitoring only 64 sentinel nodes, i.e. less than 1% of the wheat cultivated cells. Although this proportion may seem optimistic for real applications, it is worth recalling the work of Sutrave et al [24] on P. pachyrhizi indicating that a reliable epidemiological status of soybean rust in the US could be achieved by reducing the sentinel set size from 500 to 12 network optimized nodes.

Of course, the success of the deployment of a sentinel system will also depend on the diagnostic ability of individual sentinels and on the communication between them. In this sense, efforts have recently been devoted to establish international protocols to improve the probability of detection and the timely communication of new outbreaks to a global network of scientists, public authorities and stakeholders (see the Global Cereal Rust Monitoring System [6, 59]). Our findings stress the importance of increasing efforts towards a trans-boundary perspective to efficiently contain the emergence of new virulent air-borne crop pathogens [23].
Data availability statement

The code that support the findings of this study is available from the corresponding author upon reasonable request. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

The authors acknowledge the support of funding from the French National Research Agency (ANR) for the BEYOND project (Contract No. 20-PCPA-0002) and SuMCrop Sustainable Management of Crop Health Program of INRAE that supported the work of all authors. We thank Cindy Morris, Renato Casagrandi, Lorenzo Mari and Samuel Soubeyrand for fruitful discussions, two anonymous referees for extremely valuable suggestions, Luke Riley for valuable language tips as well as the technical support of Loic Houde for the computation of HYSPLIT trajectories.

Conflict of interest

The authors have no competing interests.

ORCID iDs

Andrea Radici  https://orcid.org/0000-0001-5852-0015
Davide Martinetti  https://orcid.org/0000-0003-2047-1793
Daniele Bevacqua  https://orcid.org/0000-0002-3341-1696

References

[1] Mahaffee W F and Stoll R 2016 Phytopathology 106 420–31
[2] Corredor-Moreno P and Saunders D G O 2020 New Phycol. 225 118–25
[3] Aylor D E 2003 Ecology 84 1989–97
[4] Schneider R, Hollier C, Whitam H, Palm M, McKemy I, Hernandez J, Levy L and DeVries-Paterson R 2005 Plant Dis. 89 774–774
[5] Isard S A, Gage S H, Comtois P and Russo J M 2005 BioScience 55 851–61
[6] Park K, Fetch T, Hodson D, Jin Y, Nazari K, Prashar M and Pretorius Z 2011 Euphytica 179 109–17
[7] Saunders D G, Pretorius Z A and Hovmoller M S 2019 Commun. Biol. 2 9–11
[8] Hovmoller M S, Yahyaoui A H, Milus E A and Justesen A F 2008 Mol. Ecol. 17 3818–26
[9] Zadoks J C 1967 Netr. J. Plant Pathol. 73 61–80
[10] Burrage S W 1970 Annl. Appl. Biol. 66 429–40
[11] Maddison A and Manners J 1972 Trans. Br. Mycol. Soc. 59 429–43
[12] Aylor D E 1986 Agric. Forest Meteorol. 38 263–88
[13] Damialis A, Kaimakamis E, Konoulou M, Akritidis I, Traidil-Hoffman C and Gioulkelas D 2017 Sci. Rep. 7 44535
[14] Schmale III D G and Ross S D 2015 Annu. Rev. Phytopathol. 53 591–611
[15] Meyer M, Cox J A, Hitchings M D, Burgin L, Horta M C, Hodson D P and Gilligan C A 2017 Nat. Plants 3 780–8
[16] Meyer M, Burgin L, Horta M C, Hodson D P and Gilligan C A 2017 Phytopathology 107 1175–86
[17] Visser B, Meyer M, Park R F, Gilligan C A, Burgin L E, Horta M C, Hodson D P and Pretorius Z A 2019 Phytopathology 109 133–44
[18] Allen-Sader C et al 2019 Environ. Res. Lett. 14 115004
[19] Prank M, Kenaley S C, Bergstrom G C, Acedavo M and Mahowald N M 2019 Environ. Res. Lett. 14 124053
[20] Wang M, Kriticos D J, Ota N, Brooks A and Paini D 2021 New Phycol. 232 1506–18
[21] Cunniffe N J, Koskella B, Metcalf E C J, Parnell S, Gottwald T R and Gilligan C A 2015 Epidemics 10 6–10
[22] Parnell S, van den Bosch F, Gottwald T and Gilligan C A 2017 Annu. Rev. Phytopathol. 55 591–610
[23] Ristaino J B et al 2021 Proc. Natl Acad. Sci. 118 e22239118
[24] Sutrape S, Scoglio C, Isard S A, Hutchinson J S and Garrett K A 2012 PLoS One 7 e37793
[25] FAO 2021 FAO—country brief (available at: http://www.fao.org/giews/countrybrief/) (Accessed 20 July 2021)
[26] Huerta-Espinosa J, Singh R and Roffels A P 2014 Fungi From Different Substrates (Boca Raton, FL: CRC Press) pp 217–59
[27] Olivera P et al 2015 Phytopathology 105 917–28
[28] Bhattacharya S 2017 Nat. News 542 145
[29] Institute I F P R 2019 Global spatially-disaggregated crop production statistics data for 2010, version 2.0 (https://doi. org/10.7910/DVN/PRFFBV)
[30] Draxler R R and Hess G D 1998 Aust. Meteorol. Mag. 47 295–308
[31] Aylor D E 1999 Agric. Forest Meteorol. 97 275–92
[32] Mcmaster G S and Snuka D E 1988 Agric. Forest Meteorol. 43 1–18
[33] Mcmaster G S and Wilhelms W W 1997 Agric. Forest Meteorol. 87 291–300
[34] Levetin E 2015 Manual of Environmental Microbiology (Washington, DC: ASM Press) pp 3.2.8.1–20
[35] Oneto D L, Golan J, Mazzino A, Pringle A and Seminara A 2020 Proc. Natl Acad. Sci. USA 117 5134–43
[36] Aylor D 2017 Aerial Dispersal of Pollen and Spores (St, Paul, MN: The American Phytopathological Society)
[37] Morris C E, Sands D C, Glaux C, Samshy J, Assad S, Moukahel A R, Goncalves F L and Bigg E K 2013 Atmos. Chem. Phys. 13 4223–33
[38] Nagarajan S and Singh D 1990 Annu. Rev. Phytopathol. 28 139–53
[39] Roffels A P 1992 Rust Diseases of Wheat: Concepts and Methods of Disease Management (Mexico City: CIMMYT)
[40] Rowell J and Romig R et al 1966 Phytopathology 56 807–11
[41] Li X, Yang X, Mo J and Guo T 2009 Eur. J. Plant Pathol. 123 377–86
[42] Emerson E W, Hodshrie A L, DeBolt H M, Billsback K R, Pierce J R, McMeeking G R and Farmer D K 2020 Proc. Natl Acad. Sci. 117 26076–82
[43] Slinn W 1977 Water Air Soil Pollut. 7 513–43
[44] Baiocco S, Cavina F and Pradolesi G 2021 A weather-based simulation model for the development of wheat stem rust epidemics 2021 29th Conf. Open Innovations Association (FRUCT) (IEEE) pp 22–29
[45] Choudhury M, Martinetti D, Senoussi R, Morris C E and Soubeyrand S 2021 Front. Appl. Math. Stat. 6 67
[46] Clausen T, Newman M E and Moore C 2004 Phys. Rev. E 70 066111
[47] Gerey M R and Johnson D S 1979 Computers and Intractability vol 174 (San Francisco, CA: Freeman)
[48] Freeman L C 1977 Sociometry 40 35–41
[49] Page L, Brin S, Motwani R and Winograd T 1999 The paper: citation ranking: bringing order to the web Technical Report Stanford InfoLab
[50] De Arruda G F, Barbieri A L, Rodriguez P M, Rodrigues F A, Moreno Y and Costa L D F 2014 Phys. Rev. E 90 032812
[51] Pearl M, Finlayson B L and McMahon T A 2007 Hydrol. Earth Syst. Sci. 2 1133–44
[52] IPCC 2021 Climate Change 2021: The Physical Science Basis. Contribution of Working Group 1 to the Sixth Assessment
Report of the Intergovernmental Panel on Climate Change vol 1, ed Masson Delmotte et al (Cambridge: Cambridge University Press)

[53] Damialis A, Mohammad A B, Halley J M and Gange A C 2015 Int. J. Biometeorol. 59 1157–67

[54] Choufany M, Martinetti D, Soubeyrand S and Morris C E 2021 Sci. Rep. 11 11093

[55] Serrano M A, Boguná M and Vespignani A 2009 Proc. Natl Acad. Sci. 106 6483–8

[56] Mayol E et al 2017 Nat. Commun. 8 201

[57] Olivera P D, Sikharulidze Z, Dumbadze R, Szabo I J, Newcomb M, Natsarishvili K, Rouse M N, Luster D G and Jin Y 2019 Phytopathology 109 2152–60

[58] Barnes G, Saunders D G and Williamson T 2020 Plant Pathol. 69 1193–202

[59] Morris C E, Geniaux G, Nédellec C, Sauvion N and Soubeyrand S 2021 Plant Pathol. 71 86–97