

# A Multi-pathway Model to Assess the Threat of *Tuta absoluta* in Southeast Asia

Joseph McNitt<sup>1</sup>, Young Yun Chungbaek<sup>1</sup>, Henning Mortveit<sup>1</sup>, Madhav Marathe<sup>1</sup>,  
Mateus Ribeiro de Campos<sup>2</sup>, Nicolas Desneux<sup>2</sup>, Thierry Brévault<sup>3,4,5</sup>, Rangaswamy Muniappan<sup>4</sup>,  
and Abhijin Adiga<sup>1</sup>

<sup>1</sup>*Biocomplexity Institute of Virginia Tech*, <sup>2</sup>*French National Institute for Agricultural Research*, <sup>3</sup>*BIOPASS, CIRAD-IRD-ISRA-UCAD, Dakar, Senegal*, <sup>4</sup>*CIRAD, UPR AIDA, F-34398 Montpellier, France*, <sup>5</sup>*AIDA, Univ Montpellier, CIRAD, Montpellier, France*, <sup>6</sup>*Feed the Future Integrated Pest Management Innovation Lab*

## Abstract

*Tuta absoluta* is a devastating pest of tomato which has spread in Europe, Africa, and Asia over the last decade. There is strong evidence of both natural and human-assisted pathways for its rapid range expansion. We developed a data-driven epidemiological model to study this spread accounting for biology, seasonal production, trade and demographic information. We applied it to study the dynamics of *T. absoluta* spread in South and Southeast Asia – a region at the frontier of its current range. Our analysis of spread in Bangladesh suggests that both natural spread and city-to-city vegetable trade are significant pathways of spread. We assessed the possible routes of introduction and predicted the spread pattern in the rest of the study region. There is a strong chance that within five years *T. absoluta* will invade all major vegetable growing areas of Mainland Southeast Asia if no steps are taken to mitigate the spread. Further, we showed that monitoring and effective interventions in major production regions could greatly reduce the speed of the spread.

## 1 Introduction

The world is witnessing a rapid increase in trade and travel [11]. Due to this increased connectivity, both international and domestic, no region is spared of the threat from exotic species invasion [18]. Climate

change and detrimental impact of intensive agriculture on natural resources are likely to further aggravate the problem. As a result, global food security, human health and social welfare will be adversely impacted. The South American tomato leafminer, or *Tuta absoluta*, is a representative example of biological invasion that has significantly affected food production worldwide in the last decade. Indigenous to South America, *T. absoluta* was accidentally introduced to Spain in 2006 [4, 9]. Since then, it has rapidly spread throughout Europe, Africa, Western Asia, the Indian subcontinent, and parts of Central America. With tomato being a commercially important crop, this invasion has had significant global impact. For example, in the Netherlands and Turkey alone, the annual estimated intervention costs are €4 and €167 million per year, respectively [25, 26]. Due to extensive insecticide treatment in Europe, insecticidal resistance has been recently observed in populations [15].

Since tomato is among the top two traded vegetables in the world<sup>1</sup>, it is strongly suspected that trade played a critical role in *T. absoluta*'s rapid spread. Indeed, on multiple occasions it has been discovered in packaging stations ([10] for example) and its spread pattern is correlated with prime trade routes [21]. The Animal and Plant Health Inspection Service of the United States Department of Agriculture (USDA-APHIS) has instituted quarantine regulations for imports from regions where the pest is present [37]. Rapid increase in protected cultivation methods such as green houses and tunnel farming have allowed it to overwinter as well as survive the wet season.

We describe a multi-pathway propagation model for invasive species, which is applied to study the possible spread of *T. absoluta* in the region of South and Southeast Asia comprising of 10 countries, including members of the Association of Southeast Asian Nations (ASEAN) and Bangladesh. To develop this model, we identified, analyzed, and fused disparate datasets corresponding to natural factors, biology, production seasons, trade dynamics, and demographic factors. Learning methods such as Classification and Regression Trees (CART) [7] and Random Forests were used for parameter space exploration and sensitivity analysis. To our knowledge, this is the first study that explicitly considers multiple pathways of introduction and spread of *T. absoluta*. Earlier modeling efforts have only accounted for ecological aspects and self-mediated spread [9, 16, 36]. A major hurdle in modeling the spread of emerging pests and pathogens in data of poor regions is the lack of accurate incidence reports and poor knowledge of production and trade dynamics. To address this data gap, apart from using standard datasets and knowledge from local experts, we collected quantitative and qualitative information from more than 50 research articles

---

<sup>1</sup><http://www.fao.org>

and reports that analyze vegetable production in the region.

This is a timely and important study for several reasons. In recent years, there has been a push to improve vegetable production in all the countries of this region [1]. With the pest having already spread to major tomato producing areas in Bangladesh [17], there is a high chance that it will be introduced to the remaining countries in the near future. Such invasions can have a devastating effect on the economy and livelihood of farmers. Moreover, invasion in Mainland Southeast Asia in particular is a serious threat to China, the largest producer of tomato, and Australasian neighbors.

**Summary of results.** We studied *T. absoluta* invasion of Bangladesh using the proposed model. Both natural and human-mediated pathways played a significant role in the spread of *T. absoluta* in Bangladesh. In particular, our analysis indicates that even with slow self-mediated dispersal, the pest could have rapidly expanded its range aided by domestic tomato trade. There is also a possibility of multiple introductions of the pest to this country. For the rest of the study region our models predict a faster southward spread to Thailand, Malaysia and Singapore than eastward spread in Mainland Southeast Asia due to higher trade activity in the former region. Country specific analysis shows that once introduced to a major production area, the pest will spread all over the country within two to three years. However, gaps in data and understanding of trade interactions between countries makes it difficult to predict pest movement between countries. We also demonstrated the efficacy of intervening at the trade level; monitoring and, in the case of invasion, quarantining few key production areas can contain the spread or significantly reduce the spread rate.

## 2 Results

**Multi-pathway spread model.** We developed a stochastic propagation model to simulate the multi-pathway spread of *T. absoluta*. Key concepts are illustrated in Figure 1. The study region is divided into cells by overlaying a grid. Each cell is in one of the three states: susceptible ( $S$ ) denoting pest free state, exposed ( $E$ ) denoting that the pest has been introduced but the population has not yet built up to influence other cells, and infectious ( $I$ ) denoting that the pest has established and the cell can influence its neighbors. The cell states are updated in discrete time steps, where each step  $t$  corresponds to a month. The probability that a cell  $v$  transitions from state  $S$  to  $E$  is determined by (i) suitability of the cell for *T. absoluta* to establish at that time step  $\epsilon(v, t)$  and (ii) influence of “neighboring” cells in state  $I$  depending on the pathway. An

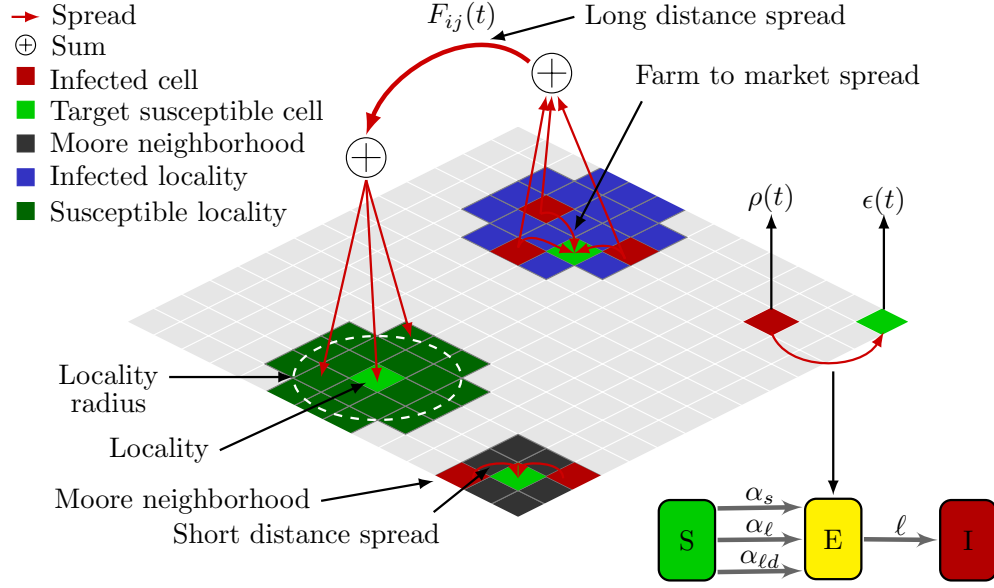


Figure 1: **Outline of the multi-pathway model.** The focus region is overlayed with a grid of cell size  $0.25$  arc degree  $\times 0.25$  arc degree, which is approximately  $27.8\text{km} \times 27.8\text{km}$  at the equator. ( $25\text{km} \times 25\text{km}$ ). We identified major urban centers and production centers in the study region. A locality corresponding to each such center consists of all grid cells for which this is the nearest city and the distance is at most the *locality radius* (in kilometers), with the additional constraint that each cell belongs to the same country as this city. Local human-mediated dispersal is modeled as the interaction between cells belonging to a city. Long distance human-mediated dispersal is modeled as trade flows of considered host crops from one city to another. Details of locality construction are in Section S2.1. The three pathways of spread are illustrated. Also shown are the states and factors that influence state transitions.

exposed cell transitions to state  $I$  after a latency period of  $\ell$  time steps. This is the time required for the population to build up to infect other cells. Once the pest has established in a cell, the cell remains infected forever, a fair assumption considering that, historically, eradication of *T. absoluta* has not been successful<sup>2</sup>. The infectiousness of a cell  $\rho(v, t)$  is modeled as a linear function of host presence at time  $t$ , for which we use the weighted sum of production volume of tomato, eggplant, and potato in that cell at time  $t$ . The weights correspond to relative carrying capacity of *T. absoluta* on the three hosts [35]. There are three pathways by which a cell can become infected: short-distance dispersal, local human-mediated dispersal and long-distance dispersal. Short-distance dispersal captures the spread through natural means from an infested cell to cells in its Moore neighborhood of range  $r_M$ . The probability that a susceptible cell gets exposed (E) at time step  $t$

<sup>2</sup>The only exception is United Kingdom where the pest was detected early (<https://gd.eppo.int/reporting/article-340>).

through short-distance spread is as follows:

$$p_s(v, t) = \epsilon(v, t) \left( 1 - \exp \left( - \alpha_s \sum_{v' \in M_v(r_M)} \rho(v', t) \right) \right), \quad (1)$$

The probability depends on the suitability of the cell  $\epsilon(v, t)$ , infestation level of each neighboring cell (Moore neighborhood of range  $r_M$  and the scaling factor,  $\alpha_s$ , which is the transmission rate for this pathway. The function form is explained in Section S2.5.

For human-assisted spread we identified large urban areas in the region which we refer to as *localities* (Figure 1) and modeled the interactions within and between localities. These areas act as attractors of vegetable flows due to high consumption and house the necessary infrastructure: wholesale markets, traders and distributors. Local human-mediated dispersal captures the spread due to “local” or farm–city–farm interactions, where the pest can spread from infested cells in a locality to other cells in the same locality. Since local farm–market dynamics are hard to model [27], we used a simple approach. For a cell in a locality (say  $L$ ), its neighbors are all cells that belong to the locality. Every cell  $v$  is influenced by cells in its locality  $L$  based on their infectiousness. The expression is similar to that in (1), but with cells in the locality instead of the Moore neighborhood.

$$p_\ell(v, t) = \epsilon(v, t) \left( 1 - \exp \left( - \alpha_\ell \sum_{v' \in L} \rho(v', t) \right) \right), \quad (2)$$

where  $\alpha_\ell$  is the scaling factor for this pathway.

Long-distance human-mediated dispersal corresponds to spread through trade between localities. For this purpose, we considered only tomato trade as there is not much evidence of *T. absoluta* spreading through trade of other hosts. We modeled domestic trade using a gravity model approach accounting for tomato production, processing, imports and exports in each locality and the travel time between localities (details are in Methods). Trade between the countries in the study region was accounted for only in those cases where the volumes were comparable to major domestic flows and there was information on ports and routes of trade. These include imports from India to Bangladesh and Malaysia to Singapore (see Figure 2 for more details). The probability of spread is directly proportional to the trade flow  $F_{ij}$  from one locality ( $i$ ) to another ( $j$ ). Suppose cell  $v$  belongs to locality  $i$ . Then, the probability of cell  $v$  transitioning from  $S$  to  $E$  due to

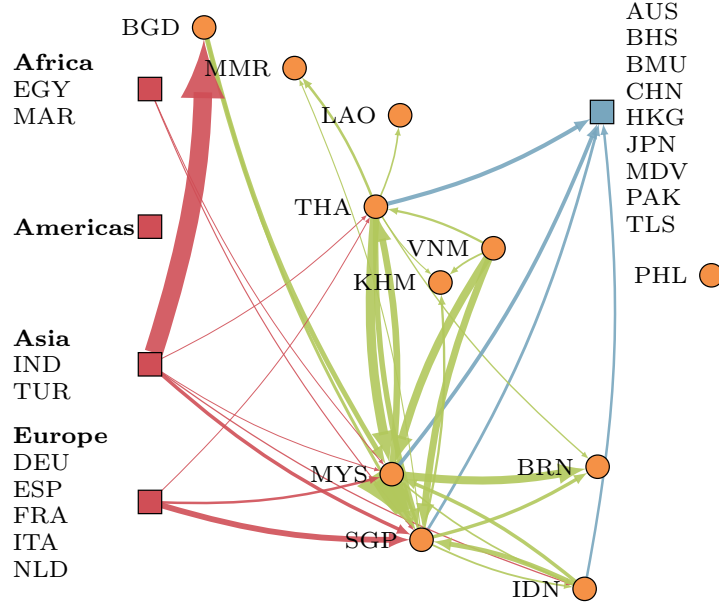


Figure 2: **The trade of tomato in the focus region (2013).** The network was constructed using data from FAOSTAT Trade matrix (Table S1). The trade within the focus region is represented by green edges. Imports from *T. absoluta* affected countries categorized by region is represented by red edges, and the blue edges correspond to exports from the focus region to countries yet to report the presence of the pest. The edge thickness is a function of the trade volume. We note that with the exception of Philippines, there is considerable trade activity among countries in the focus region. However, the majority of trading happens between neighboring countries and except for a few flows, the volume is negligible compared to production within. Most significant flow is from Malaysia to Singapore ( $\approx 30,000$  tonnes), India to Bangladesh ( $\approx 20,000$  tonnes), and Thailand and Vietnam to Malaysia (between 1000 to 2,000 tonnes). Philippines does not report any fresh tomato trade with other countries. We also analyzed how the network structure evolved across years. While we did not observe much variation in the network structure, there is some change in the countries importing to and exporting from this region (Figure S6).

long-distance dispersal is given by:

$$p_{\ell d}(v, t) = \epsilon(v, t) \left( 1 - \exp \left( - \alpha_{\ell d} \sum_{j \neq i} \sum_{v' \in L(j)} F_{ji} \rho(v', t) \right) \right), \quad (3)$$

where  $\alpha_{\ell d}$  is the scaling factor for this pathway. The model parameters and their values are summarized in Table 1.

**Assessing the role of each pathway in the spread of *T. absoluta* in Bangladesh.** The goodness of fit of a parameter instance was determined by comparing the simulation output with *T. absoluta* incidence reports from Bangladesh. The spread was simulated with infestation starting from the location of first report (Panchagarh, Bangladesh) and 100 repetitions. For each cell, empirical probability that it is in state *I* at time *t*

Table 1: Model parameters and their values.

Parameter	Description	Value/range
$r_M$	Range of Moore neighborhood	$\{1, 2, 3\}$ corresponding to spread per month of approximately 25kms, 50kms and 75kms respectively [16, 23].
$\ell$	Latency period to transition from $E$ to $I$	$\{1, 2, 3\}$ months based on time for the pest to complete life cycle ( <i>T. absoluta</i> biology in Methods).
Monthly production	Disaggregation of annual production to monthly values	<i>Uniform</i> throughout the year or <i>seasonal</i> based on regression analysis (Methods).
$\beta$	Gravity model distance function exponent	$\{0, 1, 2\}$
$\kappa$	Gravity model distance function cut-off	Between 4 to 16 hours of travel time.
Seed	Location and time of initial infestation	Scenarios based on countries (Table S6)
Locality radius	Determines cells assigned to a locality	100kms
Start month	Time step of initial infestation	$\{3, 4, 5\}$ corresponding to March, April and May respectively based on first report in Bangladesh [17].
$\alpha_s$	Short-distance spread scaling factor	
$\alpha_\ell$	Local human-mediated dispersal scaling factor	
$\alpha_{\ell d}$	Long distance spread scaling factor	

was computed. The output was compared with ground truth using a maximum likelihood method adapted from [8]. Let  $v$  be a reporting cell and  $t_v$  denote the month of actual report of pest presence. To account for uncertainty in reporting, we consider a time window  $U_\tau = [t_v - \tau, t_v + \tau]$  during comparison, where  $\tau$  is the uncertainty parameter. We set  $\tau = 2$ , i.e., error within  $\pm 2$  months is tolerated. Supposing  $\mathcal{C}_R$  is the set of cells corresponding to ground truth, and  $p(v, t)$  is the empirical probability that cell  $v$  is infected at time  $t$  in the model, then the likelihood  $\mathcal{L}$  is given by,

$$\mathcal{L} = \sum_{v \in \mathcal{C}_R} \left( \sum_{t \in U_\tau} p(v, t) + \sum_{t \notin U_\tau} (1 - p(v, t)) \right). \quad (4)$$

We have incidence information from eight locations in Bangladesh where pheromone traps were installed (Figure 3a and Table S3). Therefore, the maximum possible value of  $\mathcal{L}$  is 8 (exact match with ground truth) and the minimum is 0. Details of parameter space exploration and experiment design are in Methods.

Focusing on models with likelihood  $\mathcal{L} \geq 6$ , we observed two distinct spread patterns determined by the presence or absence of the long-distance spread component. In the first class of models (Figure 3a) which we refer to as Class A, the pathway parameter  $\alpha_{\ell d}$  is zero. It is characterized by brisk spread between geographically adjacent cells. In contrast, in Class B the long-distance pathway ( $\alpha_{\ell d}$ ) plays a significant role (Figure 3b) and is characterized by relatively slow spread between geographically adjacent neighbors. Unlike

Class A, we observe non-radial spread with jumps from one locality to another.

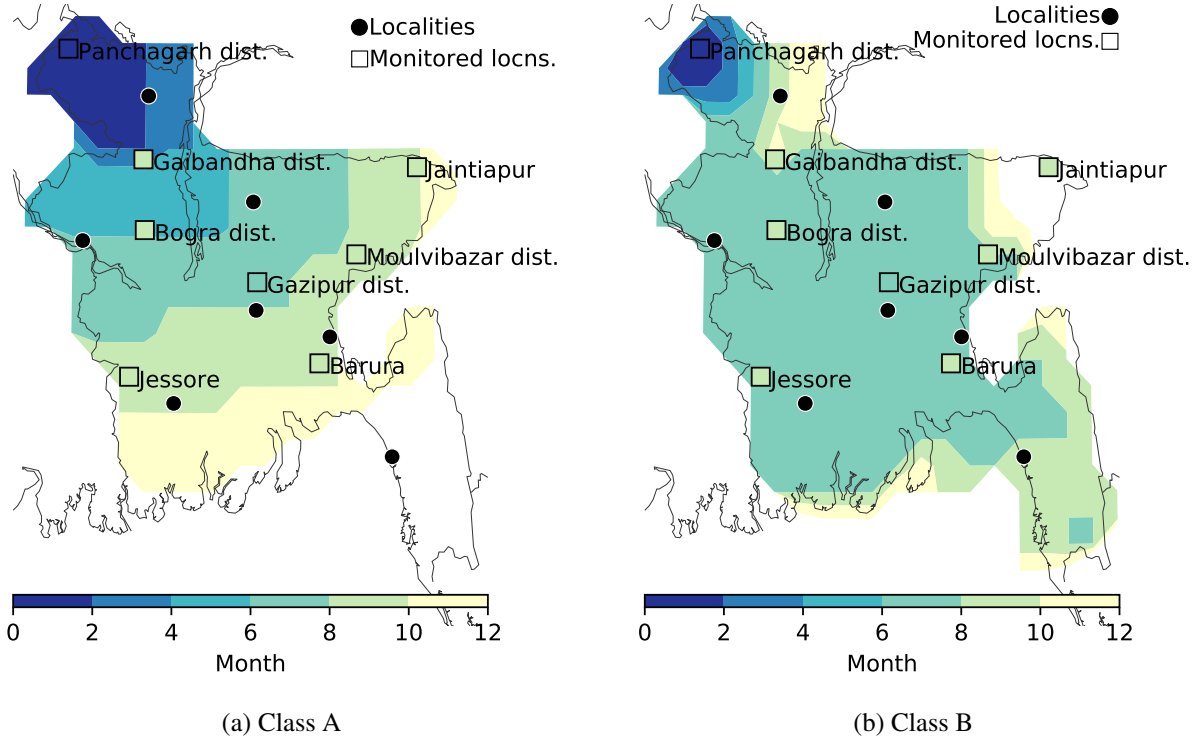


Figure 3: **Spread in Bangladesh.** The contour plots show the simulated spread starting from the location of first report in Panchagarh district for 12 months. For the purpose of plotting, the time of infection for a cell is the minimum time step  $t$  such that the empirical probability that the cell is infected by  $t$  is  $\geq 0.8$ . Also highlighted are the eight monitored locations and the localities applied in the model. The colors of the monitored locations correspond to the month of report relative to the first report (Panchagarh). These are representative spreads observed for each class. The likelihood ( $\mathcal{L}$ ) in each case was  $> 6.5$ . (a) **Class A.** Spread pattern without long-distance component. We observe a steady radial spread. (b) **Class B.** Spread pattern with long-distance jumps. Here, the radial spread is much slower. But, the long-distance jumps aid in the range expansion, particularly between the 6th and 8th time steps.

Class A spread pattern does not capture the gap between the time of first report (Panchagarh) and the report in Gaibandha district (Figure 3a). Even though the distance between the two locations is only 185kms, the latter reported the presence only after 10 months of first report suggesting that self-mediated spread might have been much slower. In the model output on the other hand, the corresponding cell gets infected between the second and fourth months. In Class B, we observe a good match for this location. However, the eastward spread towards the location Jaintiapur is slower than what was observed (Figure 3b). Even though Panchagarh is quite far from this location, pest presence was reported by February 2017, just nine months from the first report. It is possible that we have not accounted for long-distance flows that might explain this spread. A more plausible explanation is that a second invasion happened from India. In January 2017,



*T. absoluta* was reported from Umiam [32], Meghalaya about 100kms from Jaintiapur. It also happens to be near an important trade route from India to Northeastern Bangladesh [12].

We also simulated the spread using the cellular automata model developed by Guimapi [16] for Bangladesh. The spread pattern is similar to Class A as the model does not account for long-distance hops. However, the predicted rate of range expansion is much higher than our models (see Section S3 for model details and results).

**Importance of model parameters.** Calibration and sensitivity analysis of large-scale, high-resolution models such as the one proposed is a challenging task. Methods based on machine learning surrogates are being increasingly used to accomplish this efficiently for complex agent-based models [22]. Motivated by this, we studied parameter importance by applying Random Forest learning method on the simulation outputs. Among the sampled parameter instances, we analyzed the set of instances with likelihood  $\mathcal{L} > 5.5$ . The results are in Figure 4.

The long distance pathway parameter  $\alpha_{ld}$  is by far the most important parameter (Figure 4a); both spread rate and pattern critically depend on it. Class A models are more sensitive to start time step of simulation than Class B models as the spread is mostly radial in the case of the former. The main reason why it is not at all critical in the latter class is because the pest was reported just before the rainy season; subsequent months have lean trade flows, and as a result it does not affect the spread much. Also, we observe that seasonal production is very important in Class B models as it influences long distance jumps. However, it does not seem to have much effect on the radial spread component as is the case with Class A models. In both cases, the spread is very sensitive to the local human-mediated pathway parameter  $\ell$ . In the case of Class A models, This pathway leads to jumps within localities, albeit shorter than possible in Class B models that can aid in faster spread. In Class B models, the higher the value, the greater the chance of long-distance spread as it quickly leads to infection of large portions of the localities, and therefore, their total infectiousness. Moore range ( $r_M$ ) is an important parameter in both cases. Later, we will discuss how it influences the spread rate for the rest of the region.

**Predicted spread in Mainland Southeast Asia.** We applied both model classes to simulate the spread in the rest of the study region. Given that more than two years have passed since the time of first report in

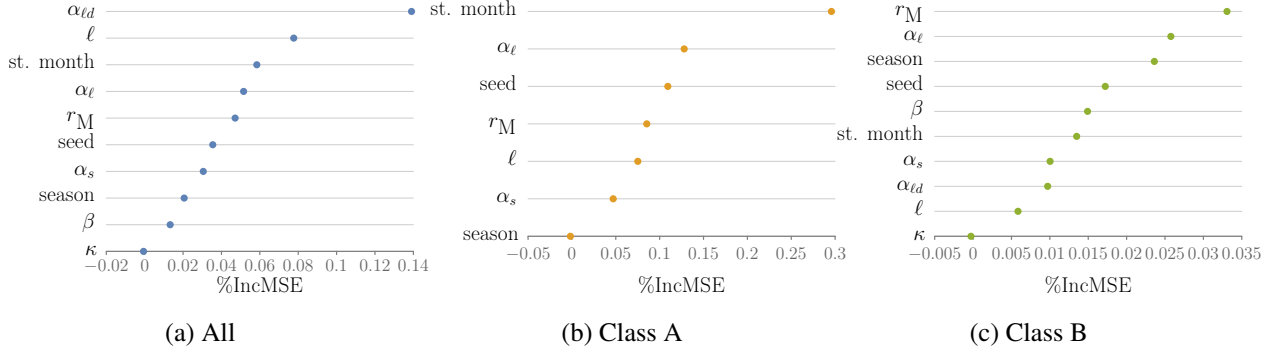


Figure 4: **Parameter importance.** Random Forest (RF) [6] method was applied to study parameter importance. With likelihood ( $\mathcal{L}$ ) as the dependent variable, the analysis was performed on the set of model instances which yielded a likelihood  $\mathcal{L} \geq 5.5$  ( $\approx 80\%$  match with ground truth), a total of 17, 000 samples from the parameter space. The results for the entire set (Plot (a)) and separately for Class A (Plot (b)) and Class B models (Plot (c)) are shown. The parameters are ordered by their importance based on percentage increase in mean square error from permuting the variable—higher, the more important.

Bangladesh and unofficial reports of *T. absoluta* presence in Myanmar<sup>3</sup>, cells in northern Myanmar bordering India and Bangladesh were seeded (Methods). Representative simulation outputs are shown for Mainland Southeast Asia in Figure 5. We note that in the case of Class A models, the eastward spread is faster than southward spread. This is mainly because the Moore neighborhood is smaller at the narrow region in the south of Myanmar and Thailand as it connects to Malaysia. However, in the case of Class B, the spread is much faster in the same region aided by domestic trade flows from northern and central Thailand to the southern region. Both models indicate that the pest will enter Thailand from Myanmar, and subsequently move to Laos and Vietnam as it spreads eastwards and to China when it spreads northwards.

Class A spread pattern predicts that within the next 4-5 years much of the northern part of Mainland Southeast Asia will be invaded. Class B spread pattern predicts that in the same period *T. absoluta* will spread all over Malaysia and Singapore. However, the rate of spread observed is slower than that observed in Bangladesh for both classes. Also, even though the models exhibited similar rate of spread for Bangladesh, we observed high variance in intensity of infestation as well as range expansion for the rest of the region. Results are in Figure 6. The reason for slow spread is as follows. Bangladesh has the highest tomato volume per country surface area ( $\approx 2.5$  tonnes/km<sup>2</sup>). The next country is Vietnam ( $\approx 1.5$  tonnes/km<sup>2</sup>). Therefore, in the case of Bangladesh, not only is the extent of infestation in a cell  $\rho(\cdot)$  typically high, but also, since it is a densely populated country, most cells have vegetable production. Hence, the rate of spread is much higher

<sup>3</sup>The Wikipedia entry on *T. absoluta* ([https://en.wikipedia.org/wiki/Tuta\\_absoluta](https://en.wikipedia.org/wiki/Tuta_absoluta)) indicates that it is present in the northern region of Myanmar since April 2017. No official confirmation is available yet.

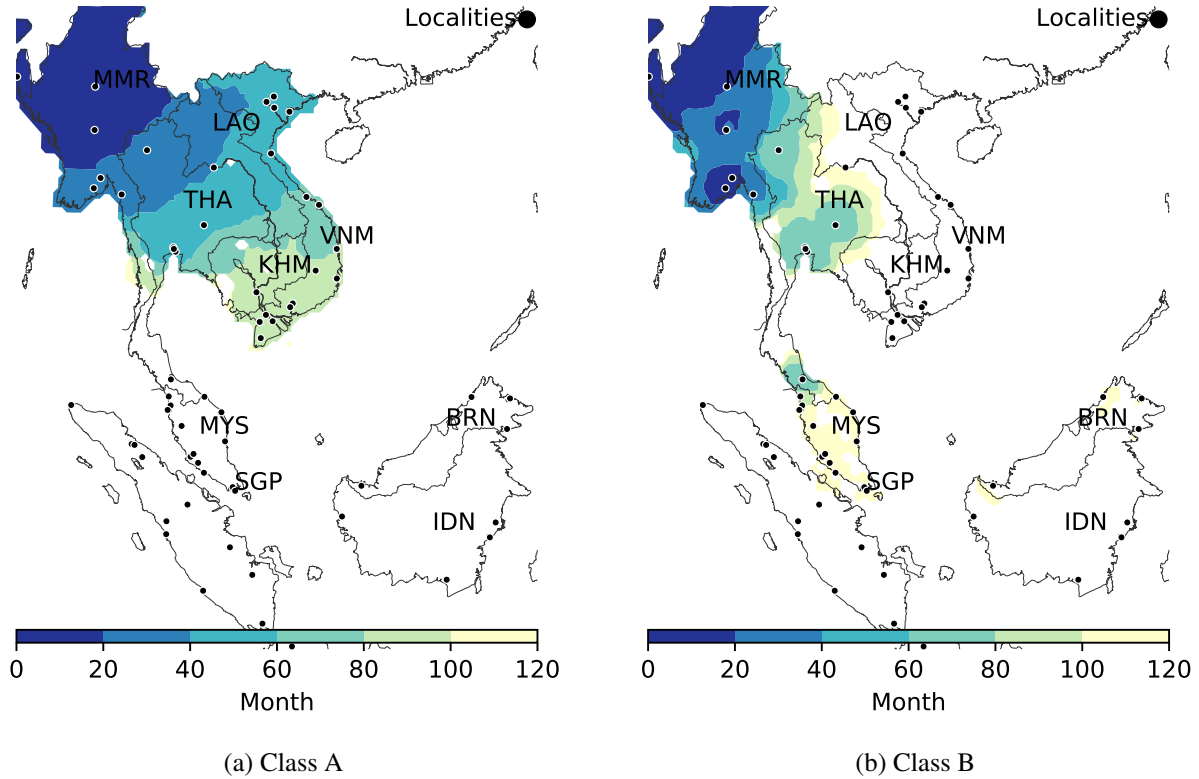


Figure 5: **Predicted spread pattern in Mainland Southeast Asia.** The contour plots show the simulated spread starting from northern Myanmar for 120 time steps or 10 years. Representative simulation outputs for Class A and Class B models are shown. With Class A, spread is faster eastward than southward and the other way round in the case of Class B.

for relatively lower values of pathway parameters and Moore range. Also, we observed a strong dependence on Moore range (Figure 6b). In countries with larger area, the production is scattered. Therefore, the lower the Moore range, the slower the spread.

**Alternate scenarios of invasion.** From the analysis of the international trade network (Figure 2), we observe that Malaysia and Singapore are important hubs with tomato imports from *T. absoluta* infested regions. There is a possibility that *T. absoluta* is directly introduced to these regions. However, as mentioned earlier, the import volume is very low. Also, the introduction risk depends on the preventive measure taken by the exporting countries. We considered the scenario in which *T. absoluta* establishes in the region between Kuala Lumpur and Singapore. The simulations suggest that the pest will spread to most localities in Malaysia and Sumatra, Indonesia within an year.

Since Philippines does not share its borders with any country in the region and there is no evidence of

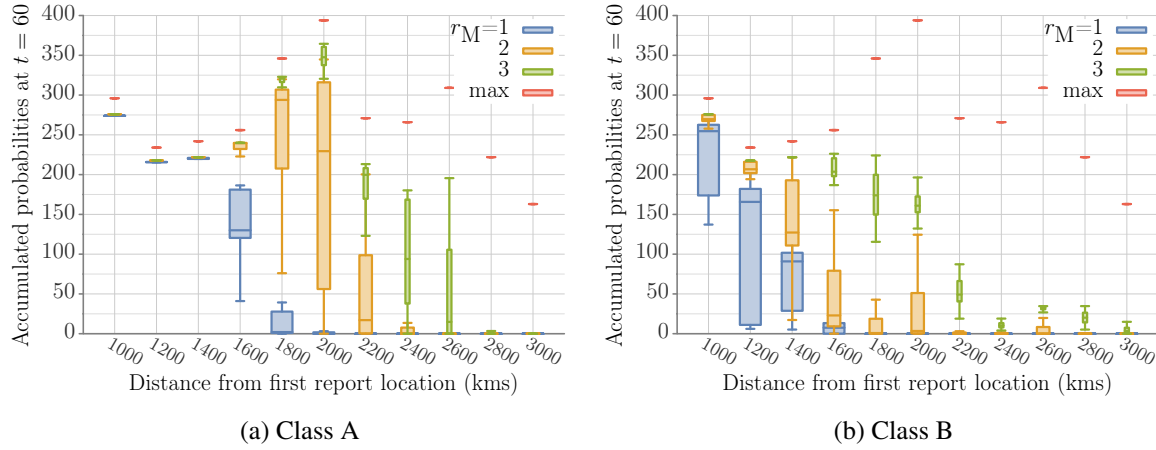


Figure 6: **Spread in Mainland Southeast Asia with respect to distance from origin.** The cells are binned based on their distance from the origin of infection (Northern Myanmar). Given time step  $t$  (60 or five years in this case), let  $\Pr(v, \leq t)$  be the probability that cell  $v$  is in state  $I$  by time  $t$ . For each parameter instance, we computed the “total infection” for every bin at time  $t$  by aggregating  $\Pr(v, \leq t)$  for each  $v$  in the bin. Parameter instances were grouped by model class and Moore range ( $r_M$ ). The average total infection for each group are plotted. The red points referred to as “max” correspond to the total number of cells in each bin, which is also the maximum possible accumulated probability for that bin. We observe that even though the models exhibit similar spread for Bangladesh, there is high variance in spread rate in both classes for spread in in the case of he spread rate when applied to the rest of the region. Also, the range of expansion is influenced by Moore range  $r_M$ .

tomato trade with rest of the countries, there is a low chance that the pest will be introduced through natural spread or trade. However, human mobility is a possible pathway. The Middle East is the top destination for Filipino workers [31]. Therefore, there is a possibility of introduction through travel. We considered two hypothetical scenarios: (i) *T. absoluta* is introduced to the Palawan region from Malaysian Borneo and (ii) it is introduced to the high production area of Northern Mindanao. While in the first case we did not observe spread beyond the Palawan region, in the second case, within two years, almost all localities were invaded. The latter case is shown in Figure 7a.

**Role of trade in spread within and between countries.** Here, we focus on long-distance dispersal, and therefore, restrict our discussion to Class B models. We observed the following pattern in the spread. When the pest is introduced to a country, dispersal is slow until the invasion front reaches a locality close to a production area, after which the spread is very fast and within a year or two, it hits almost all major localities of the country. There is a possibility that *T. absoluta* can be introduced to a country several months before it becomes pervasive. Our results show that for all countries, once the pest is introduced, within 2-3 years it will spread to almost all localities, and when it reaches a high production locality, it spreads to other regions

within a year. Past invasion records support this trend.

We recall the discussion about the slow spread rate of our models in Mainland Southeast Asia compared to the spread in Bangladesh. In the case of Class B models, another reason for the reduced spread rate could be due to ignoring some trade flows between countries. We recall that by construction, only trade flows that were comparable to major domestic flows were considered. Because of this, we observe a slow spread at the country boundaries. However, it is non-trivial to include these flows without knowledge of ports of entries and trade relations between countries. Historically, international trade has played a strong role in the spread of *T. absoluta* between countries in Europe and West Asia. This seems to apply to even South Asia. The pest was first reported by India in 2014 [19, 33]. By early 2016, it was discovered in the Kathmandu area of Nepal [3], the northern part of Bangladesh in May 2016 [17]. Both countries import significant volume of tomato from India. However, there has been no report from Pakistan, another neighbor which does not import tomato from India. It is possible that the pest is present and undetected in this country. But, it is clear that Pakistan is aware of the threat<sup>4</sup> indicating that even if it is present, it is not widespread. There are similar examples outside the region such as its slow advance from South America to Central America, or the fact that it is not reported in China despite being present in neighboring Central Asian countries since 2015.

**Interventions at the trade level.** Given the importance of the long-distance human-mediated pathway, we studied the effect of controlling this pathway in mitigating the spread. Monitoring localities by setting up pheromone traps in production areas and markets, and quarantining affected areas is one way to accomplish this. However, the feasibility and efficacy of implementing this or any other strategy in real-life is not clear. Also, since monitoring is resource intensive, developing strategies which involve few locations, yet provide near-optimal control is a goal for modelers. Nopsa et al. [24] considered this problem in the context of soybean rust in the U.S., modeling the spread as a propagation process over a network of counties. They developed efficient strategies to reduce the spread by identifying and monitoring only a subset of locations which are critical for the spread. Their strategies heavily rely on the history of infections, which is absent in our case.

We evaluated a simple strategy. Localities associated with high production areas and relatively less consumption were identified. There are at most four such localities per country. These localities have high annual outflows and as discussed earlier, pest establishment in these areas can potentially lead to rapid range

---

<sup>4</sup>The looming threat of the deadly tomato leafminer (<https://www.dawn.com/news/1420206>).

expansion. The outflow from the targeted localities was reduced to mimic control at the trade/market level. Figure 7 shows the effectiveness of this spread for the case of Philippines. We observed a significant reduction in range expansion as well as intensity of spread. Besides, as seen in Figure 7b, stifling these flows leads to a radial and more importantly, predictable spread pattern that resembles those of Class A models, but with much less intensity.

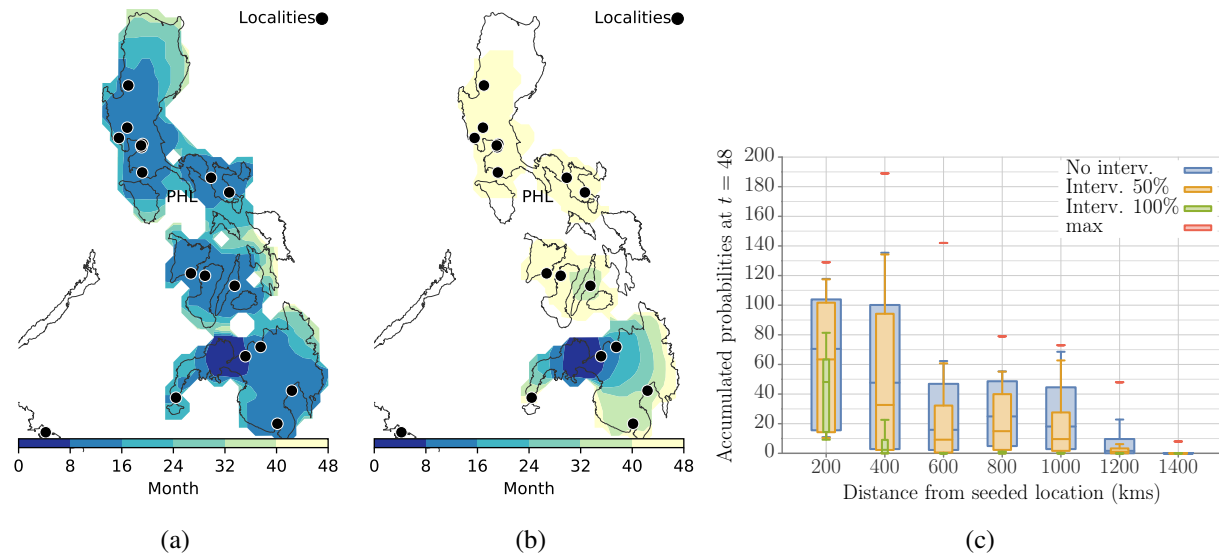


Figure 7: **Interventions.** The effect of reducing outflows from major production regions is shown for Class B models. These are representative results for Philippines. Plots for other countries are in Figure S8. In the simulations, a high production region (Northern Mindanao) was seeded. (a) **Without intervention.** Simulations indicate that for the given initial conditions, there is a high chance that all major production areas of Philippines will be affected within 24 months. The colors indicate the time interval at which there is at least a 50% chance that a location will be infected. (b) **With market level intervention.** The spread without the influence of long-distance flows from localities corresponding to high production areas (Northern Mindanao and Central Luzon). We observe a delay of more than two years in the introduction to the northern part due to this intervention. (c) **Spread with respect to origin of infection.** We compare the effect of reducing the long-distance flow the above-mentioned regions by 50% and 100%. The results shown correspond to all Class B instances with Moore range  $r_M = 1$ .

### 3 Discussion

Traditionally, crops such as tomato and eggplant have been grown in the winter during the dry season. However, over the past decade, due to rising demand and opportunities to export, there has been a push towards year-round production using protected cultivation methods and resilient varieties [1]. Our analysis of the FAO data on tomato production and trade shows a steady increase in production and internal trade. In

comparison, the export of tomato to outside of the focus region has risen steeply in the recent years (after 2011), while the imports generally indicate a downward trend (See Figure S5). Also the food industry is gradually restructuring: supermarkets are replacing traditional food chains, and there is a steady increase in processing industries. Therefore, invasions from pests such as *T. absoluta* can have a huge negative impact on the socioeconomic fabric of this region.

Although there is a general consensus that vegetable and seedling trade is a primary driver of *T. absoluta* spread, previous modeling efforts have exclusively focused on ecological aspects. Two studies [9, 36] provide risk maps using CLIMEX and take additional factors into account. Guimapi et al. [16] used a cellular automata approach to capture the global spread of the pest by factoring in temporal variations and spatial distribution of vegetation, temperature, and tomato production. A precursor to this work [38] modeled the seasonal production and trade of tomato in Nepal to study the role of trade in the spread of *T. absoluta* in Nepal using gravity model and network dynamics.

Multi-pathway models have been analyzed to study the role of human-mediated dispersal [8, 28, 30]. Robinet et al. [30] show that the distribution and spread pattern of the pinewood nematode in China is strongly correlated with density of human population and infrastructure such as railways and river ports. A similar approach was applied to the pine processionary moth [28]. Carrasco et al. [8] combine spatially explicit models with a phenology model to incorporate population dynamics of the pest (western corn rootworm). Two types of long-distance dispersals are considered – domestic and international. The domestic mode is modeled as a flow network between cities using a gravity model approach. Nopsa et al. [24] use a network science approach to studying the role of transport and storage infrastructure in the spread of pests and pathogens of wheat. Suttrave et al. [34] use a time-varying network model to study the spread of Soybean rust. However, in all these cases, there was more information available on the incidence of the pest or pathogen under study enabling rigorous validation of the developed models.

Our model is in part motivated by the hybrid approaches used in the study of infectious diseases of humans and livestock. Bradhurst et al. [5] study the spread of foot and mouth disease in livestock by using an aggregate population-level model to capture within-herd spread and an individual-based model for between herd spread. A similar approach is used by Yang et al. [40] to forecast influenza outbreaks. They use a patch network model where a compartmental model is used to simulate intra-locale spread and a gravity model based approach is used for inter-locale spread of flu in the neighborhoods of New York.

**Key challenges.** We have used a multitude of datasets—production, consumption, trade dynamics, climate, biology, etc.— to capture, to a reasonable extent, the complexity of the invasion process. For many countries, domestic production data had to be collected (or even inferred) from several publications and reports (Table S2). Therefore, data came from disparate sources (multi-type, different countries, etc.), and were misaligned in time (different years) and spatial resolution (grid to country level). Production and trade data from FAOSTAT also had gaps in them. In particular, it was hard to model seasonal production and human assisted spread. Seasonal production is dependent not only on the host, climate, and geography, but also on people’s preferences and market demand. For most countries, only qualitative information is available. To determine outflows and inflows for each locality, we had to identify major ports for imports and exports as well as estimate fraction of production which was used for processing. This data was available only for a few countries.

**Limitations.** Scarcity of data has forced us to greatly simplify or ignore some of the processes that might significantly influence the spread.

Our model uses monthly production as a surrogate for infectiousness of a cell. Complex phenology models can be used instead to model accurately the growth of the pest under specific conditions including the tritrophic interactions concerning the pest, host, and its indigenous predators (like, for example, the work by Carrasco et al. [8]). However, as cautioned by Robinet et al. [29], this would add to the complexity of the model making it near impossible to verify and validate.

The farm–market–consumer interactions (local human-mediated spread) are well understood at a conceptual level. More or less the structure remains the same in all the countries of the focus region. It involves various actors such as farmers, wholesalers, retailers, wet markets, supermarkets, etc. Nevertheless, it is nearly impossible to accurately model the dynamics. Rebaudo et al. [27], for example, use a complex agent-based model to study just the interaction between farmers of two villages in the context of the potato moth in Ecuador. Similarly, given data on actual flow of vegetables, the gravity model can be improved or replaced by more sophisticated approaches. Also, the relationship between long-distance invasion risk and trade volume is hard to determine. While a direct relationship between volume and risk is plausible, whether the relation is linear (as assumed by our model) is not clear.

It is important to account for heterogeneity in production, consumption, awareness, cultural factors, etc.

both within and between countries. Some countries are technologically more advanced than others,



which manifests as differences in yield, crop loss and trade infrastructure. Also, they are better prepared for the invasion by taking necessary steps such as identifying and stocking pesticides and conducting workshops to create awareness. It is possible that there is lot of variation in consumption within a country [39] as well as across seasons. The relationship between trade volume and distance between localities could vary from one region to another.

The modeling framework developed in this paper can be used to study the spread of *T. absoluta* in other regions. However, it is possible that additional factors need to be accounted for. Seedling trade and greenhouse production are two examples that apply to regions such as North America or Australia. It can also be applied to other pests. However, this would require first assessing the importance of each pathway based on evidence: flying capacity, spread by human mobility or trade, etc. Besides invasive species spread, other potential applications for this work include studies of natural or human-initiated disasters [11], climate change [13] and optimization of food flows [30].

## 4 Methods

**Data.** All the datasets used in the model and for analysis are described in Section S1. Almost all the datasets are openly available. The remaining information was provided by local contacts in Bangladesh, Vietnam and Cambodia.

***T. absoluta* biology.** The tomato leafminer exhibits a short life cycle of about 24–38 days (temperature at  $25 \pm 3^\circ\text{C}$ ), from egg to adult, as it is a multivoltine species with overlapping generations in the field [14]. This species causes serious damage to numerous solanaceae crops such as eggplants, potatoes, and especially tomato crops [35]. It penetrates into tomato leaves, stems, or fruits, wherein it feeds and develops by creating conspicuous mines as well as galleries. *T. absoluta* additionally restricts tomato plant growth by feeding on the growing tips. Considering the warm weather throughout the year, particularly in the dry season, the study region presents ideal conditions for rapid development and spread of *T. absoluta*. Pest risk analysis [36] shows that the Ecoclimatic Index for this region is above 50 (highly suitable). Spatial distribution assessment survey of *T. absoluta* eggs has shown its high dispersive capacity in tomato producing areas [23]. The dispersion in a tomato cultivation starts mainly at the periphery and the pest is able to migrate between tomato farms to generate egg aggregation at the crop edges. The pest spread behavior among seasonal crop resources is

often non-random and directional [23]. Sylla et al. [35] analyzed host preference of *T. absoluta* in France and Senegal. While the highest preference is for tomato, it can survive well on eggplant and potato, which happen to be major vegetable crops in the study region. However, since *T. absoluta* primarily attacks leaves of eggplant and potato, the chance of the pest spreading through trade of these crops seems to be low.

**Seasonal production.** We estimated monthly production volume of tomato, eggplant and potato for each cell. This was accomplished in two steps. First, we estimated annual production in each cell followed by disaggregation to monthly production. The annual production was estimated using the vegetable production available at the highest resolution for each country (at the level of province to just one value for the entire country) and a synthetic dataset called Spatial Production Allocation Model (SPAM)<sup>5</sup>. Gathering qualitative data from multiple sources, we inferred that the cropping pattern in this region depends primarily on two factors: seasons—dry and wet, and elevation—highland (upland) and lowland. Also, barring few exceptions, the production of the considered host crops peaks in the winter and is lowest during the rainy season (monsoon). To predict seasonal production, we used linear regression to model the production rate as a function of precipitation, temperature, and elevation. Regional tomato and eggplant production data is available for Philippines. The regression results showed that precipitation was a statistically significant predictor ( $p < 0.001$ ). For validation, the regression function was applied to locations of different countries where this information is available and visually compared. More details on the methods and validation are provided in Section S2.2.

**Domestic trade flows.** We modeled the flow of fresh tomato crop between markets based on the following assumptions: (i) The total outflow from a city depends on the amount of produce in its surrounding regions and imports from countries outside the focus region at time  $t$ , and (ii) the total inflow depends on total consumption, processing demand, and exports from the city to countries outside the focus region. While there is strong evidence of tomato trade as a possible pathway of *T. absoluta* spread [4], there is no evidence of a similar role of eggplant or potato trade. Hence, we modeled only tomato trade. For each country, the domestic flow is estimated using a doubly constrained gravity model [2, 20]. For a city  $i$ , let  $O_i$  and  $I_i$  denote total outflow and total inflow respectively, and  $L(i)$  denotes all cells which are assigned to it. The flow  $F_{ij}$  from city  $i$  to city  $j$  is given by  $F_{ij}(t) = a_i(t)b_j(t)O_i(t)I_j(t)f(d_{ij})$ , where,  $d_{ij}$  is the time to travel from  $i$

---

<sup>5</sup><http://mapspam.info>

to  $j$ , and  $f(\cdot)$  is the *distance deterrence function*:  $d_{ij}^{-\beta} \exp(-d_{ij}/\kappa)$ , where  $\beta$  and  $\kappa$  are tunable parameters. The coefficients  $a_i$  and  $b_j$  are computed through an iterative process such that the total outflow and total inflow at each node agree with the input values [20]. Overall, we have 12 networks representing flows for each month. The outflows and inflows are calculated as follows:

$$O_i(t) = \text{Prod}(i, t) + \text{Import}(i, t) - \text{Export}(i, t) - \text{Proc}(i, t), \quad (5)$$

$$I_i(t) = \text{Pop}(i). \quad (6)$$

Here,  $\text{Prod}(i, t)$  is the monthly production and  $\text{Pop}(i)$  is the population of the locality as a surrogate to consumption. The latter is the sum total of population in every cell that belong to the locality.  $\text{Export}$  and  $\text{Import}$  are the monthly total export to and import from outside the country respectively.  $\text{Proc}(i, t)$  is the tomato produced for processing. Since for this purpose it is typically cultivated and consumed locally [39], we subtract this volume from the outflow. Country-specific details of how locality attributes were estimated is in Section S2.3. Since no data is available on volume of trade between markets, we gathered qualitative information from reports and research articles which show evidence of tomato or vegetable flow between market pairs. The results of this analysis is in Section S2.4.

**Parameterization and experiment design.** Parameter space exploration was conducted in multiple iterations. First, we coarsely sampled the space. With model parameters as independent variables and the goodness of fit measure defined in (4) as the dependent variable, we used Classification and Regression Trees (CART) approach to identify subspaces for which likelihood was high and rejected subspaces for which likelihood was low. Based on this, in the subsequent phases, we refined our search to improve the parameterization. More details of the CART analysis is presented in Section S4. Simulations were performed on more than 500,000 parameter combinations using a high performance computing cluster.

**Data availability.** The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information file, or from the authors upon reasonable request.

**Acknowledgments** This work was supported in part by the United States Agency for International Development under the Cooperative Agreement NO. AID-OAA-L-15-00001 Feed the Future Innovation Lab for Integrated Pest Management, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, NSF BIG DATA Grant

IIS-1633028, NSF DIBBS Grant ACI-1443054, NIH Grant 1R01GM109718 and NSF NRT-DESE Grant DGE-154362. We are grateful to Yousuf Mian, Nguyen Van Hoa, and Kimhian Seng for their help with obtaining country-specific information on production, trade, and pest incidence, and Richard Beckman for useful discussions on model design.

**Author contributions.** AA defined the scope of the research. AA, JM, TB, MRC collected and interpreted data. AA, MM conceived and designed the experiments. JM, AA and YYC performed the analysis. HM, ND, TB and RM provided assistance in interpreting the results. AA and JM wrote the paper with significant inputs from MRC and YYC. AA supervised the research. All authors discussed the results and commented on the manuscript.

## References

- [1] M. Ali and V. T. B. Hau. Vegetables in Bangladesh: Economic and nutritional impact of new varieties and technologies. *Asian Vegetable Research and Development Centre*, (Technical Bulletin No. 25):1–15, 2001.
- [2] J. E. Anderson. The gravity model. *Annual Review of Economics*, 3(1):133–160, 2011.
- [3] A. S. R. Bajracharya, R. P. Mainali, B. Bhat, S. Bista, P. Shashank, and N. Meshram. The first record of South American tomato leaf miner, *Tuta absoluta* (Meyrick 1917) (Lepidoptera: Gelechiidae) in Nepal. *J. Entomol. Zool. Stud*, 4:1359–1363, 2016.
- [4] A. Biondi, R. Guedes, and F. Wan. Ecology, Worldwide Spread, and Management of the Invasive South American Tomato Pinworm, *Tuta absoluta*: Past, Present, and Future. *Annual Review of Entomology*, (63):239–258, 2017.
- [5] R. A. Bradhurst, S. E. Roche, I. J. East, P. Kwan, and M. G. Garner. A hybrid modeling approach to simulating foot-and-mouth disease outbreaks in australian livestock. *Frontiers in Environmental Science*, 3:17, 2015.
- [6] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [7] L. Breiman. *Classification and regression trees*. Routledge, 2017.

- [8] L. Carrasco, J. Mumford, A. MacLeod, T. Harwood, G. Grabenweger, A. Leach, J. Knight, and R. Baker. Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. *Ecological Modelling*, 221(17):2068–2075, 2010.
- [9] N. Desneux, E. Wajnberg, K. A. Wyckhuys, G. Burgio, S. Arpaia, C. A. Narváez-Vasquez, J. González-Cabrera, D. C. Ruescas, E. Tabone, J. Frandon, et al. Biological invasion of European tomato crops by *Tuta absoluta*, ecology, geographic expansion and prospects for biological control. *Journal of Pest Science*, 83(3):197–215, 2010.
- [10] EPPO article-105. *Tuta absoluta* caught in a tomato packing station in the Netherlands. <https://gd.eppo.int/reporting/article-105>, 2009.
- [11] M. Ercsey-Ravasz, Z. Toroczka, Z. Lakner, and J. Baranyi. Complexity of the international agro-food trade network and its impact on food safety. *PloS one*, 7(5):e37810, 2012.
- [12] Export-Import Bank of India. Bangladesh: A Study of India’s Trade and Investment Potential. 2015.
- [13] FAO. Climate change and food security: A framework document. *Food and Agriculture Organization of the United Nations*, page 93, 2008.
- [14] R. Guedes and M. Picanço. The tomato borer *Tuta absoluta* in South America: pest status, management and insecticide resistance. *EPPO bulletin*, 42(2):211–216, 2012.
- [15] R. Guedes and H. Siqueira. The tomato borer *Tuta absoluta*: insecticide resistance and control failure. *Plant Sciences Reviews 2012*, page 245, 2013.
- [16] R. Y. Guimapi, S. A. Mohamed, G. O. Okeyo, F. T. Ndjomatchoua, S. Ekesi, and H. E. Tonnang. Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecological Complexity*, 28:77–93, 2016.
- [17] M. Hossain, M. Mian, and R. Muniappan. First record of *Tuta absoluta* (Lepidoptera: Gelechiidae) from Bangladesh. *Journal of Agricultural and Urban Entomology*, 32(1):101–105, 2016.
- [18] P. E. Hulme. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1):10–18, 2009.

- [19] C. Kalleshwaraswamy, M. S. Murthy, C. Viraktamath, and N. K. Kumar. Occurrence of *Tuta absoluta* (Lepidoptera: Gelechiidae) in the Malnad and Hyderabad-Karnataka Regions of Karnataka, India. *Florida entomologist*, 98(3):970–971, 2015.
- [20] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48):1093–1103, 2010.
- [21] O. Karadjova, Z. Ilieva, V. Krumov, E. Petrova, V. Ventsislavov, et al. *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae): Potential for entry, establishment and spread in Bulgaria. *Bulgarian Journal of Agricultural Science*, 19(3):563–571, 2013.
- [22] F. Lamperti, A. Roventini, and A. Sani. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, 90:366–389, 2018.
- [23] J. C. Martins, M. C. Picanço, R. S. Silva, A. H. Gonring, T. V. Galdino, and R. N. Guedes. Assessing the spatial distribution of *Tuta absoluta* (Lepidoptera: Gelechiidae) eggs in open-field tomato cultivation through geostatistical analysis. *Pest management science*, 74(1):30–36, 2018.
- [24] J. F. H. Nopsa, G. J. Daglish, D. W. Hagstrum, J. F. Leslie, T. W. Phillips, C. Scoglio, S. Thomas-Sharma, G. H. Walter, and K. A. Garrett. Ecological networks in stored grain: Key postharvest nodes for emerging pests, pathogens, and mycotoxins. *BioScience*, page biv122, 2015.
- [25] S. Oztemiz. *Tuta absoluta* povolny (Lepidoptera: Gelechiidae), the exotic pest in Turkey. *Romanian Journal of Biology*, 59:47–58, 2014.
- [26] R. Potting, D. Van Der Gaag, A. Loomans, M. Van der Straten, H. Anderson, A. MacLeod, J. Castrillón, and G. Cambra. *Tuta absoluta*, tomato leaf miner moth or South American tomato moth. Ministry of Agriculture, Nature and Food Quality. *Plant Protection Service of the Netherlands, Utrecht, The Netherlands*, 2013.
- [27] F. Rebaudo, V. Crespo-Perez, J.-F. Silvain, and O. Dangles. Agent-Based Modeling of Human-Induced Spread of Invasive Species in Agricultural Landscapes: Insights from the Potato Moth in Ecuador. *Journal of Artificial Societies and Social Simulation*, 14:1–22, 2011.

- [28] C. Robinet, C.-E. Imbert, J. Rousselet, D. Sauvard, J. Garcia, F. Goussard, and A. Roques. Human-mediated long-distance jumps of the pine processionary moth in Europe. *Biological invasions*, 14(8):1557–1569, 2012.
- [29] C. Robinet, H. Kehlenbeck, D. J. Kriticos, R. H. Baker, A. Battisti, S. Brunel, M. Dupin, D. Eyre, M. Faccoli, Z. Ilieva, et al. A suite of models to support the quantitative assessment of spread in pest risk analysis. *PLoS One*, 7(10):e43366, 2012.
- [30] C. Robinet, A. Roques, H. Pan, G. Fang, J. Ye, Y. Zhang, and J. Sun. Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One*, 4(2):e4646, 2009.
- [31] R. M. Rodriguez. Philippine migrant workers’ transnationalism in the middle east. *International Labor and Working-Class History*, 79(1):48–61, 2011.
- [32] E. Sankarganesh, D. Firake, B. Sharma, V. Verma, and G. Behere. Invasion of the South American Tomato Pinworm, *Tuta absoluta*, in northeastern India: a new challenge and biosecurity concerns. *Entomologia Generalis*, 36(4):335–345, 2017.
- [33] V. Sridhar, A. Chakravarthy, and R. Asokan. New record of the invasive South American tomato leaf miner, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) in India. *Pest Management in Horticultural Ecosystems*, 20(2):148–154, 2014.
- [34] S. Sutrave, C. Scoglio, S. A. Isard, J. S. Hutchinson, and K. A. Garrett. Identifying highly connected counties compensates for resource limitations when evaluating national spread of an invasive pathogen. *PLoS One*, 7(6):e37793, 2012.
- [35] S. E. Sylla, T. Brévault, L. Monticelli, K. Diarra, and N. Desneux. Geographic variation of host preference by the invasive tomato leafminer *Tuta absoluta*: implications for host range expansion. Technical report, 2018.
- [36] H. E. Tonnang, S. F. Mohamed, F. Khamis, and S. Ekesi. Identification and risk assessment for worldwide invasion and spread of *Tuta absoluta* with a focus on Sub-Saharan Africa: implications for phytosanitary measures and management. *PloS one*, 10(8):e0135283, 2015.
- [37] USDA. New Pest Response Guidelines: Tomato Leafminer (*Tuta absoluta*). *Animal and Plant Health Inspection Service, Plant Protection and Quarantine*, 2012.

- [38] S. Venkatramanan, S. Wu, B. Shi, A. Marathe, M. Marathe, S. Eubank, L. P. Sah, A. Giri, L. A. Colavito, K. Nitin, et al. Towards robust models of food flows and their role in invasive species spread. In *Big Data (Big Data), 2017 IEEE International Conference on*, pages 435–444. IEEE, 2017.
- [39] S. V. Wijk. The market for vegetables in North Vietnam. 2007.
- [40] W. Yang, D. R. Olson, and J. Shaman. Forecasting Influenza Outbreaks in Boroughs and Neighborhoods of New York City. *PLoS Computational Biology*, 12(11):1–19, 2016.