Supplementary material: Assessing the Multi-pathway Threat from an Invasive Agricultural Pest: *Tuta absoluta* in Asia

Joseph McNitt[1], Young Yun Chungbaek[2], Henning Mortveit[2], Madhav Marathe[2], Mateus R Campos[3], Nicolas Desneux[3], Thierry Brévault[4,5,6], Rangaswamy Muniappan[7], and Abhijin Adiga[2]

[1]Epic Systems Corporation, United States

[2]Biocomplexity Institute & Initiative, University of Virginia

[3]French National Institute for Agricultural Research

[4]BIOPASS, CIRAD-IRD-ISRA-UCAD, Dakar, Senegal

[5]CIRAD, UPR AIDA, F-34398 Montpellier, France

[6]Université de Montpellier, CIRAD, Montpellier, France

[7]Feed the Future Integrated Pest Management Innovation Lab, Virginia Tech

S1 Data

Table S1 lists all publicly available datasets incorporated in our model and analysis. Table S2 provides country-specific data sources, literature and reports used in the model and for validation. They are further tagged based on the information they provide for different aspects of the model. Table S3 provides incidence reports from Bangladesh. Code is available in Github [2].

S2 *T. absoluta* biology.

The tomato leafminer exhibits a short life cycle of about 24–38 days (temperature at $25 \pm 3^{\circ}$C), from egg to adult, as it is a multivoltine species with overlapping generations in the field. It causes serious damage to numerous solanaceous

Table S1: Global data sets used for model construction.

Data	Source	Spatial or temporal resolution
Temperature	WorldClim [25]	2.5 arc min., monthly.
Precipitation	GPCC [13]	2.5 arc deg., monthly.
Humidity	NASA [46]	1 arc deg., monthly.
NDVI	NEO [45]	0.1 arc deg., monthly.
Country Production	FAO [15]	country, yearly.
Shape files of study region	ADC Worldmap [4]	country/province/district.
Cell Production	Mapspam [61]	5 arc min., yearly. Gives vegetable and potato data.
Consumption	FAO [17]	country, yearly.
Population	Landscan[36]	5 arc min.
Cities	MAXMIND [39]	city resolution
Distance between cities	Google API [19]	city resolution

Table S2: Country-specific data obtained from reports, peer-reviewed articles and experts' inputs.

Country	Seasons	Production	Consumption	Domestic trade		Processing	International trade
				Markets	Flows		
Bangladesh	[6]	[6]	[16]	[6]	–	[6, 59]	[14]
Cambodia	[8, 18, 52]	[18, 52]	[52]	[52]	[52]	–	[44, 52]
Indonesia	[5, 20]	[5, 20]	[16]	–	–	–	[16]
Laos	[33]	[33]	[33]	[33]	[33]	–	[33, 44]
Malaysia	–	[22, 41]	[16]	–	–	–	[16]
Myanmar	–	[35, 50]	[43]		[35]		
Philippines	[7, 49]	[49]	[10]	[11]			ignored [11]
Singapore	–	[16]	imp.-exp.	city	ignored		
Thailand	[29]	[29, 42, 56]	[56]	[29, 56]	[29]	[42]	
Vietnam	[27, 58, 60]	[60, Table 16] (2003)	[60, Table 23]	[9, 44, 60]	[9, 44, 60]	[3, 60]	[44, 60]

Table S3: Data of *T. absoluta* infestation for Bangladesh (source: Bangladesh Agricultural Research Institute) and neighbouring region of India.

Location	Coordinates	Month of first detection of *T. absoluta*
Chaklarhat, Panchagarh district	26.19° N, 88.43° E	May 2016 [26]
Umiam [51], BARI Research Field, Gazipur district	23.99° N, 90.42° E	January 2017
RARS, Akbarput, Moulvibazar district	24.25° N, 91.45° E	February 2017
Lalakhal, Jointiapur, Sylhet district	25.07° N, 92.09° E	February 2017
Palashbari, Gaibandha district	25.15° N, 89.23° E	March 2017
Battara, Bogra district	24.48° N, 89.25° E	February 2017
Barura, Comilla district	23.22° N, 91.06° E	March 2017
Patengali, Jessore district	23.09° N, 89.10° E	March 2017
Umiam, Meghalaya India		January 2017 [51]

crops such as eggplants, potatoes, and especially tomato crops [54]. It penetrates into tomato leaves, stems, or fruits, wherein it feeds and develops by creating conspicuous mines as well as galleries. 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 [55] 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 [38]. 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 boundaries. The pest spread behaviour among seasonal crop resources is often non-random and directional [38]. Sylla et al. [54] analysed 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.

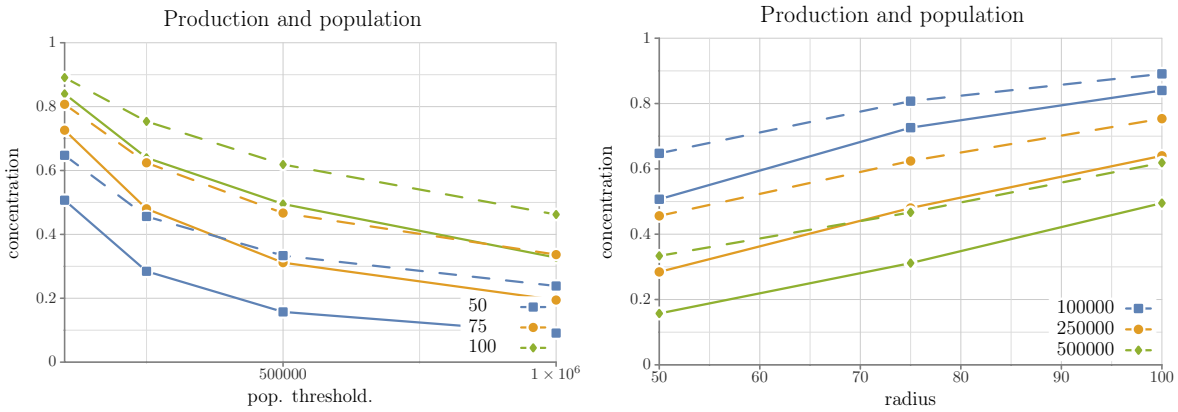
S3 Multi-pathway Model

S3.1 Locality construction

In the model, *localities* are centres of consumption or production. From the perspective of consumption, we selected cities with population greater than a certain population threshold (as per [39]) in the entire study region. The number and size of localities is controlled by population threshold and locality radius parameters. We considered a range of threshold values, and chose $250,000$ as the threshold for the model with the main criterion for the choice being coverage of population and knowledge of major wholesale markets (Figure S1). Then, we added major production centres if their population did not meet the threshold. The locality radius was chosen to be 100km since local production in an urban area was within 50–60km from the city centre for several countries in this region [8, 52, 60]. To obtain long distance trade flows, the travel times between pairs of cities were computed using Google API [1].

S3.2 Seasonal production

We estimated monthly production volume of tomato, eggplant and potato for each cell using production data available at the country/state/province level and the Spatial Production Allocation Model [61]. The latter uses a generalised cross-entropy approach to allocate crop production in administrative units into individual pixels, through judicious interpretation of all accessible evidence such as production statistics, farming systems, satellite image, crop biophysical suitability, crop price, local market access and prior knowledge. The monthly production volume was estimated in two steps: (i) Estimation of cell-level annual production, and (ii) disaggregation of annual production to monthly production.



(a) (b)

Figure S1: **Locality construction.** The number of localities and the locality area is controlled by population threshold and locality radius. We analysed the effect of these parameters on the amount of production and population captured in the region. Setting a threshold of 250,000 and radius of 100km, 80% of the production and population is covered by localities.

Annual production. From SPAM [61], we obtained annual production estimates for each cell. However, there are several issues with directly using this data. Firstly, these are estimates for the year 2005, and secondly, tomato and eggplant production estimates are not available. Instead, total vegetable production volume is available. Also, for countries where data was available, we did not find any correlation between reported tomato (eggplant) production and total SPAM vegetable production for that region. Therefore, for each country, we obtained the most recent production data available (2013 or later) at the highest spatial resolution (region/province/country) (Table S1). The production of a particular vegetable type at a cell was computed as follows:

$$\frac{\text{Total production in the region}}{\text{Total SPAM production for cells in the region}} \times \text{SPAM production in the cell} .$$

For tomato and eggplant, we used vegetable production as the surrogate. There were also cases where no data was available (Cambodia, Myanmar and Laos for example). In such cases, for potato, we used SPAM data as is. For tomato and eggplant, the SPAM value for vegetables was scaled by a scaling factor which was determined as follows. For countries where data was available, we computed the ratio of total tomato production and total SPAM vegetable production for the country. The median value (≈ 0.05) was used as the scaling factor. The same procedure was used for eggplant.

Monthly production. In order to estimate the seasonal production rates of tomatoes, we considered quarterly production data from Philippines [49] for the 16 regions of the country and studied its relationship between precipitation,

elevation and temperature. We first assigned average monthly precipitation, temperature and elevation data to the cells by latitude and longitude. Next, for each region, we obtained the seasonal relative tomato production (or production rate). This was obtained by dividing each of the quarterly production volume by total annual production volume. We conducted a linear regression with the product rate as a dependent variable and the precipitation as an independent variable in SPSS 24.0. Since the dependent variable was highly skewed, we implemented a log-transformation for it. To control elevation, we classified the elevations into two groups, high and low, using K means in SPSS 24.0. Due to the small sample size, we excluded the samples in the high-elevation group and conducted a linear regression analysis for the group of low elevation (< 235).

The regression results showed that precipitation was a statistically significant predictor ($p < 0.001$, $R^2 = 0.54$). When we accounted for temperature along with precipitation, the R value increased to 0.58 with temperature exhibiting weak correlation with production. Though the results with both variables included gave a slightly stronger correlation, in the validation step, the regression function corresponding to precipitation was a better match for the rest of the study region. Thus we decided to use only precipitation as a predictor for seasonality of tomato production. Table S4 provides information on these regression results.

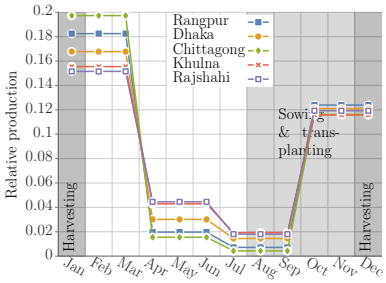
Table S4: Linear regression results for seasonal production as a function of precipitation.

	B	Std. Error	Beta	t	Sig.
Intercept	-0.208	0.229		-0.908	0.368
Precipitation	-0.008	0.001	-0.734	-7.935	0.000

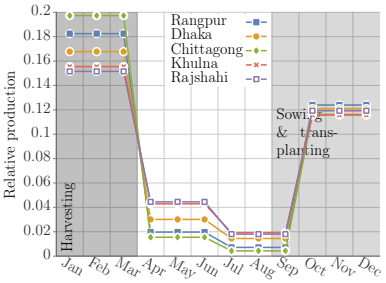
Validation. With the exception of data from Philippines on tomato and eggplant, there is no regional seasonal production information for the three crops considered for other countries. However, there are several reports that provide qualitative information. See Table S2, column "Seasons". In some cases seasonal production data is available at the country level. For example, for Bangladesh, eggplant production volume is available by season. Similar information is available for Cambodia. In Figure S2, we have compared seasonal production predicted by the regression model with available qualitative and quantitative information. We note that the precipitation based regression function captures the general trend of seasonal production in the different regions considered.

S3.3 Network construction

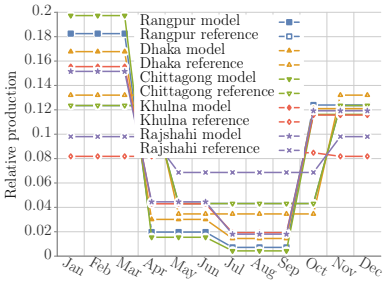
For each country, the domestic flow was estimated using a doubly constrained gravity model [31, 57]. For a city i, let O_i and I_i denote total outflow and total inflow respectively, and $\mathrm{L}(i)$ denotes all cells which are assigned to it. The flow F_{ij} from city i to city j is given by $F_{ij}(t) = a_i(t)b_j(t)O_i(t)I_j(t)f(d_{ij})$, where, d_{ij} is the time to travel



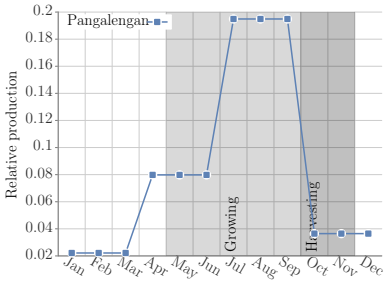
(a) Bangladesh tomato production compared with growing season information [6]



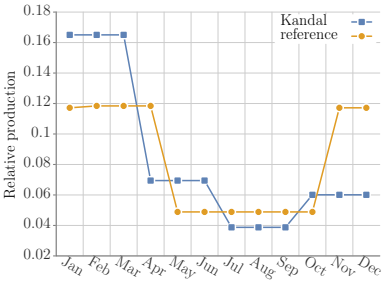
(b) Bangladesh potato production compared with growing season information [6]



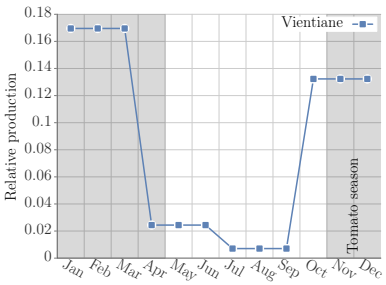
(c) Bangladesh eggplant production compared with growing season information [6]



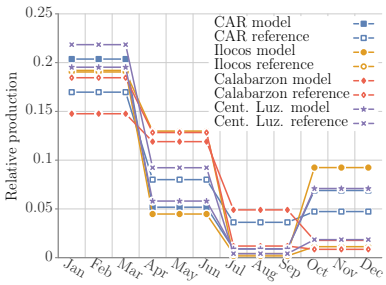
(d) Indonesia tomato production compared with growing season information [5]



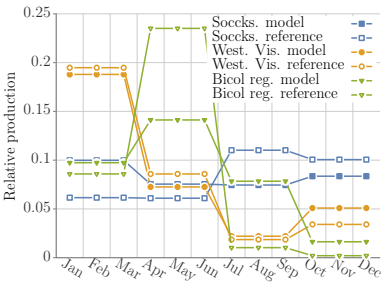
(e) Cambodia tomato production compared with seasonal production information [18]



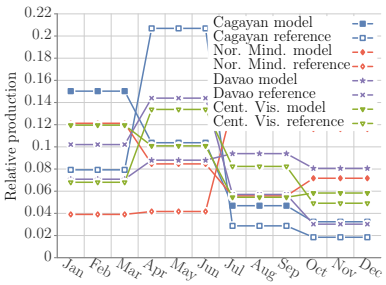
(f) Laos tomato production compared with growing season information [33]



(g) Philippines tomato production compared with growing season information [49] (North)



(h) Philippines tomato production compared with growing season information [49] (Central and South)



(i) Philippines tomato production compared with growing season information [49]: Regions with mismatch.

Figure S2: Comparing relative production as per the regression model with quantitative/qualitative reports for different regions. For the plots corresponding to the regression model we chose representative cells for each administrative region.

from i to j, and $f(\cdot)$ is the *distance deterrence function*: $d_{ij}^{-\beta}\exp(-d_{ij}/\kappa)$, where β and κ 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 [31]. Overall, we have 12 networks representing flows for each month. The outflows

and inflows are calculated as follows:

$$O_i(t) = \mathrm{Prod}(i,t) + \mathrm{Import}(i,t) - \mathrm{Export}(i,t) - \mathrm{Proc}(i,t), \tag{1}$$

$$I_i(t) = \mathrm{Pop}(i)\,. \tag{2}$$

Here, $\mathrm{Prod}(i,t)$ is the monthly production and $\mathrm{Pop}(i)$ 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. Export and Import are the monthly total export to and import from outside the country respectively. $\mathrm{Proc}(i,t)$ is the tomato produced for processing. Since for this purpose tomato is typically cultivated and consumed locally, we subtract this volume from the outflow. Country specific details of how locality attributes were estimated is in Section S3.4.

S3.4 Locality production, consumption, imports, exports and processing

Monthly tomato production (consumption) at a locality was obtained by aggregating production (population) at all cells that belong to it. Population data was obtained from Landscan (Table S1). For most countries, imports and exports are a small fraction of domestic production. The main exceptions were the significant tomato imports from India to Bangladesh and trade between Malaysia and Singapore. We identified major routes of trade from India to Bangladesh [14] and the total imports from India (FAOSTAT) was distributed uniformly between three cities close to the border with India along these routes. Finally, these imports were evenly distributed for the later half of the year since Bangladesh imports mostly during the rainy season. To capture the significant trade between Singapore and Malaysia, we included Singapore in the domestic flow network of Malaysia as there is high interaction between the two countries. The resulting flow from the gravity model from Malaysia to Singapore was obtained by aggregating network flows across months and across edges with Singapore as destination. This flow was comparable to the annual imports from Malaysia to Singapore. With the exception for Thailand [42], Vietnam [60] and Cambodia (pers. comm.), there is no information on the amount of tomato production consumed by the processing industry even though there is evidence of processing industry in Malaysia and Indonesia. For each locality in Thailand, we scaled the monthly production by the ratio annual production of fresh tomatoes over total annual production (fresh and processed).

For consumption, we used country-level production [17], which was available for half of the countries. For Singapore, we estimated it as the difference between total inflow (production and imports) and total outflow (exports) based on FAO data. For Vietnam, Wijk et al. [60] provides this information. For Myanmar, Cambodia and Laos, we found no information. We used median consumption for the region. We also analysed consumption with respect to per capita gross domestic product (GDP). However, we did not find any correlation between GDP and consumption both globally

as well as restricted to the study region. In Venkatramanan et al. [57], consumption was modelled as a function of population and Gross Domestic Product (GDP). In the study region, analysing FAO data on consumption [17] we did not find any relation between GDP and consumption.

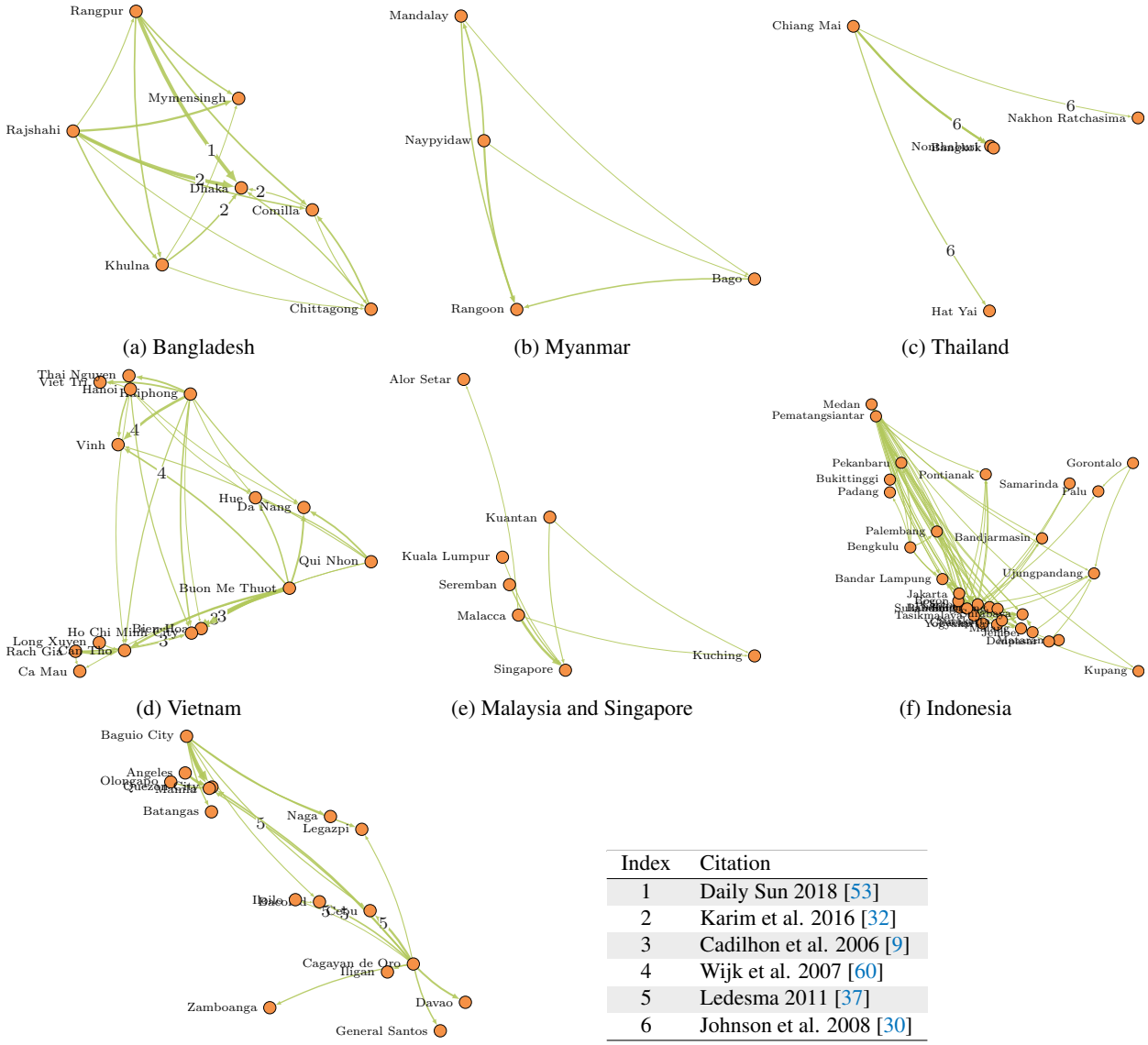
S3.5 Validation of domestic trade flows

Since no market-to-market flow data is available for the region, we searched the literature for evidence of tomato trade between cities or regions for each country. Even this information is hardly available. The next step was to further generalise the search by considering the flow of vegetables. The results of our comparison of the networks obtained using gravity flow model with literature are shown Figure S3. Our general observations as follows: (i) To obtain good representation of trade flow, at minimum, region/state level data of production is required; and (ii) densely populated urban pockets are a good representation of consumption centres. Some country specific details follow.

Bangladesh. There is evidence of vegetable flow from Rangpur region to Dhaka particularly during winter [28, 53]. The flow from Southwest Bangladesh to Dhaka [32] is captured by the edge from Khulna to Dhaka (Figure S3a). We assigned import of tomato from India to cities in the border based on information on important trade routes [14]. Since, there is no data on volume, we distributed it evenly.

Vietnam. There are primarily two consumption centres: Ho Chi Minh City in the south supplied by Central Highlands and the Mekong River Delta, and Hanoi-Haiphong urban centres in the north supplied by Red River Delta. For Ho Chi Minh City, Lam Dong is the main district that supplies vegetables [9]. In our model (Figure S3d), this flow is represented by Buon Me Thuot–a city in Central Highlands–to Ho Chi Minh City. Buon Me Thuot is less than 100km away from the Lam Dong province thus covering the production in the area. In the north, much of the production surplus is assigned to the Haiphong locality, and therefore, it acts as a big source for almost all localities in the north. Our network also has some long distance edges from north to south. This is mostly due to the large population in the south. There is not much evidence of this flow and it could be a result of not accounting for heterogeneity in consumption.

Malaysia and Singapore. We combined the locality corresponding to Singapore with Malaysia, so that the flow between these two countries is treated as a domestic flow as there is free movement of people and goods between these countries. The gravity model flows from Malaysia to Singapore is between 50,000–60,000 tonnes compared to $\approx 30,000$ tonnes according to FAOSTAT. However, our network does not capture the flow from Cameron Highlands,



(a) Bangladesh (b) Myanmar (c) Thailand

(d) Vietnam (e) Malaysia and Singapore (f) Indonesia

(g) Philippines

Index	Citation
1	Daily Sun 2018 [53]
2	Karim et al. 2016 [32]
3	Cadilhon et al. 2006 [9]
4	Wijk et al. 2007 [60]
5	Ledesma 2011 [37]
6	Johnson et al. 2008 [30]

(h) References corresponding to trade flow evidence.

Figure S3: **Long-distance annual trade flows resulting from the gravity model.** We combined monthly flows and to avoid clutter, we have displayed only major flows (> 500). Some of the edges are labelled based on information from literature. The references corresponding to the labels are listed in (h). For Laos, Cambodia and Brunei, there are no flows as there is only one locality in each country.

major tomato producer to cities of Malaysia. This shortcoming is mainly due to unavailability of state/province level tomato production information.

Philippines. The major production centres are in Cagayan de Oro and regions north of Manila. These are captured by our network (Figure S3g).

S3.6 Transmission probability

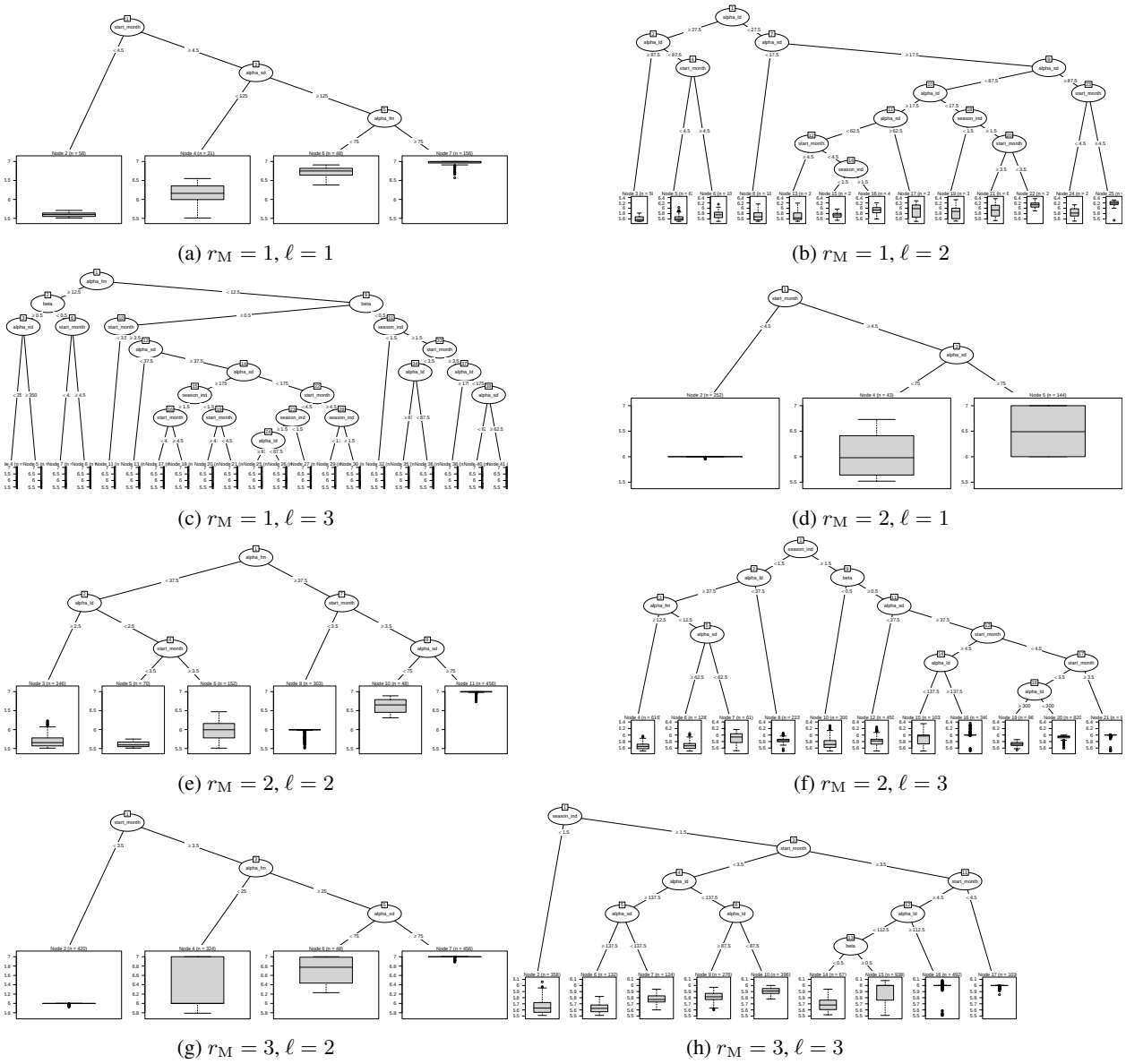
At any given time t, for a cell v and its neighbour v' corresponding to the considered pathway, let $\rho(v', t)$ be the strength of infestation at v'. Let, α be the transmission rate corresponding to the pathway. Then, the probability that the pest will be introduced to v from v' is given by $1 - \exp\big(-\alpha\rho(v', t)\big)$. The function form is similar to that in [40, 47], but with contact time between two nodes replaced with strength of infestation. Also, α can be interpreted as the transmission probability per unit infestation. Since in our case production in the cell is used as a surrogate for $\rho(v', t)$, it can be interpreted as the probability of infection per unit production.

This form of the infection probability function has been seen before in epidemiological models [12, 23, 24], both continuous and discrete time models. Diekmann et al. [12] consider a continuous time SI model, and define the probability that an individual will not be infected with a function which relates to the form above, taking the exponent as an integral of the "force of infection". In [23], we see a discrete time version of this function form, with the probability of still being exposed after t time steps as: $P(t) = e^{-\epsilon t}$, where ϵ relates to the "transfer rate". This function form was derived when considering traditional SIR models mathematically in [24].

S4 Parameterization and simulation

For each parameter setting, we evaluated the model output by comparing it with historical invasion data from Bangladesh using a similarity score as defined in the main document (equation (4)). Simulations were run with 100 repetitions. From the output we computed the empirical probability that a cell is in state I at time t, which in turn was used to compute similarity scores. We applied Classification and Regression Trees (CART) to guide parameter space exploration. Initially, the parameter space was coarsely sampled. The parameter values are listed in Table 1 in the main document. For each of these samples, simulations were run and the similarity scores computed. With the model parameters as independent variables and similarity score as the dependent variable, we analysed the results using CART (see Figure S4). We chose parameter subspaces with high similarity score (0.7 or more) and rejected those with lower values. For each chosen parameter subspace, the process was repeated.

We recall that the emergent outcome of the model can be classified into two classes: Class A and Class B. In Figure S4, we show the results of CART analysis on the parameter sets which yielded a similarity score of ≥ 0.7. The chosen set is partitioned by Moore range r_{M} and latency period ℓ. For (r_{M}, ℓ) values that permit rapid range expansion through short-distance spread over the contiguous landscape, including long-distance spread leads to faster spread than observed leading to lower score. Hence, for those regimes, we observe only Class A spread. The corresponding (r_{M}, ℓ)



(a) $r_{\mathrm{M}} = 1, \ell = 1$

(b) $r_{\mathrm{M}} = 1, \ell = 2$

(c) $r_{\mathrm{M}} = 1, \ell = 3$

(d) $r_{\mathrm{M}} = 2, \ell = 1$

(e) $r_{\mathrm{M}} = 2, \ell = 2$

(f) $r_{\mathrm{M}} = 2, \ell = 3$

(g) $r_{\mathrm{M}} = 3, \ell = 2$

(h) $r_{\mathrm{M}} = 3, \ell = 3$

Figure S4: CART analysis of the parameter space: The models with similarity score ≥ 0.7 were partitioned based on the Moore range r_{M} and latency period ℓ tuple and analysed. For the case $(r_{\mathrm{M}} = 3, \ell = 1)$, there were no instances with similarity score ≥ 0.7 among the sampled instances. Minimum number of observations at a node for splitting was set to 50. Minimum number of observations at each leaf node was set to 20.

values are $(1, 1)$, $(2, 1)$, $(2, 2)$ and $(3, 1)$. For the regime $(3, 1)$ we did not observe any instance with score > 5.5.

S4.1 Seeding scenarios

During the model evaluation phase, to account for both spatial and temporal observation noise, we considered six seeding scenarios: The locations Bangladesh 1 and Bangladesh 2 (see Table S5) and start times March, April or

May (the month *T. absoluta* was reported). For the prediction phase, we constructed three different seeding scenarios reflecting three different possible introductions to the region. The first corresponds to spread from Bangladesh through northeast Myanmar. This region represents a possible pathway of *T. absoluta* should it spread radially through the border with Bangladesh. The second scenario we considered is the region of Johor in Malaysia - including Singapore. This region contains major ports in Malaysia including Tanjung Pelepas [34], and Singapore is also a major trading hub for Bangladesh. We considered two scenarios of introduction for Philippines. The first is motivated by the fact that there is a large migrant Filipino worker population travelling to and from infested countries. We seeded the port of Masinloc in this scenario as it processes many of these travellers. The second was the introduction to Northern Mindanao region, one of the major tomato production areas in the country. We conducted country-specific experiments in the case of Thailand and Vietnam. The seed locations were determined by spread observed from both Class A and Class B models.

Table S5: Seeding scenarios

Region seeded	Description
Bangladesh 1	Parameterization phase: The cell corresponding to the location of first report was seeded.
Bangladesh 2	Parameterization phase: The cells corresponding to the location of first report and its adjacent cells (Moore neighbourhood 1) were seeded.
Bangladesh and Northeastern Myanmar	This seeding is the estimated current range of *T. absoluta* in the study region. It is used to study the spread in the rest of the region.
Singapore and southern (mainland) Malaysia	This reflects possible introduction through trade, as Singapore is a major port.
Philippines 1	This reflects the possibility of *T. absoluta* being introduced through workers travelling to and from India, and other countries, which have reported the pest. The port of Masinloc was seeded.
Philippines 2	Northern Mindanao region
Thailand	Northwestern region: informed by observed spread in Mainland Southeast Asia.
Vietnam	Northwester region: Informed by observed spread in Mainland Southeast Asia.

S4.2 Software and Computational aspects.

The model was implemented in Python 2.7. For data management and processing, we used PostGreSQL and SQlite. Statistical analysis was done using Python, R 3.4.3 and SPSS 24.0. The experiments were run using Discovery, a high performance computing cluster with 232 nodes (16-core Sandy Bridge-EP E5-2670 2.60GHz (3.30GHz Turbo) Dual Processor (8 Cores per Processor) nodes with 32 GB of Memory and 500 GB Internal Hard Drive) in the Biocomplexity Institute of Virginia Tech.

S5 Clustering analysis of the spread pattern

The analysis process is illustrated in Figure S5 We recall that each simulation output has the following format. Suppose, the simulations were run for T time steps. For every v, $p(v,t)$ is the empirical probability that cell v is infected at time t. Each simulation output can be viewed as a matrix of size $n \times T$, where n is the number of cells. To normalise each matrix, we introduced an extra index $T+1$ which carries the residual probability, i.e., $p(v,T+1) = \sum_{t=1}^{T} p(v,t)$. The simulation outputs for models with similarity $\mathcal{S} \geq 0.75$ were thus processed and clustered using two algorithms: hierarchical agglomerative clustering (SPSS 24.0) and k-means algorithm (using Pyclustering [48]). In each case, the number of clusters considered ranged from $k = 1$ to 10.

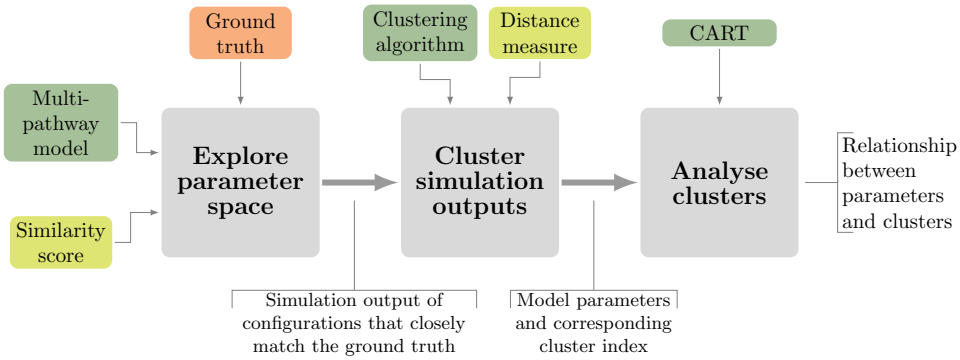


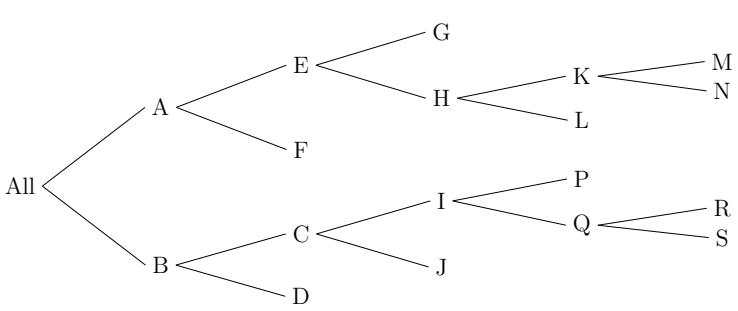
Figure S5: Outline of the process used for analysing the multi-pathway spread.

To discover relationships between model parameters and cluster membership, we cast it as a classification problem. With configurations as features and cluster index as labels, we applied CART. The resulting decision tree was used to interpret the relationship between parameters and cluster membership. The results are in Figures S6 (hierarchical) and S7 (k-means).

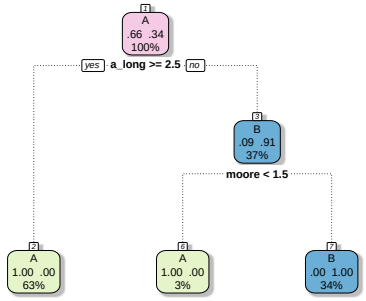
S6 More results

S6.1 International trade

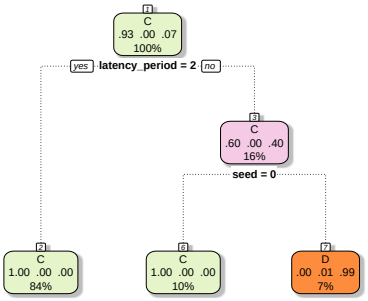
Figure S8a shows the annual tomato trade for the study region. The network was constructed using data from FAOSTAT Trade matrix (Table S1). The trade within the focus region is represented by green edges. Imports from *T. absoluta* affected countries categorised by region is represented by red edges, 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



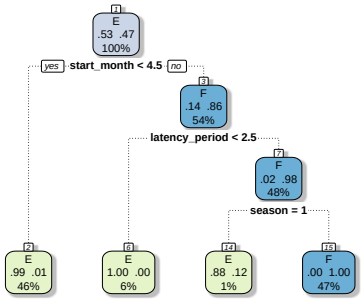
(a) Hierarchy



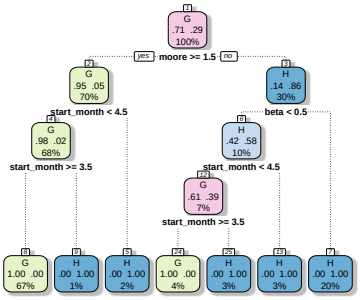
(b) $k = 2$



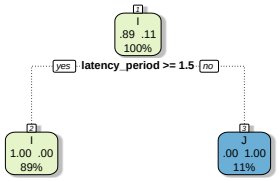
(c) $k = 3$



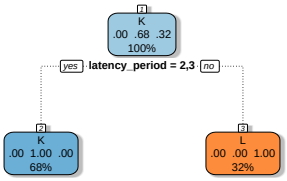
(d) $k = 4$



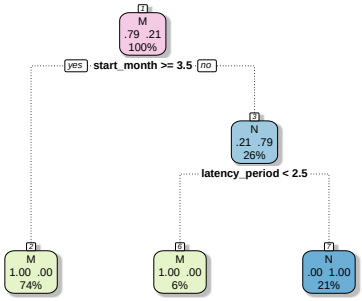
(e) $k = 5$



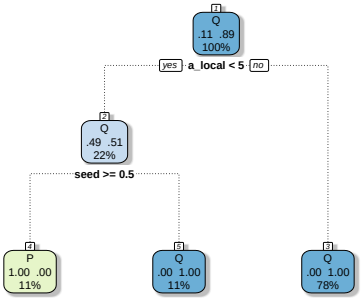
(f) $k = 6$



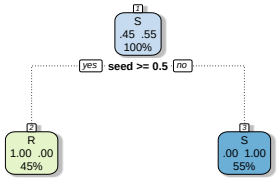
(g) $k = 7$



(h) $k = 8$

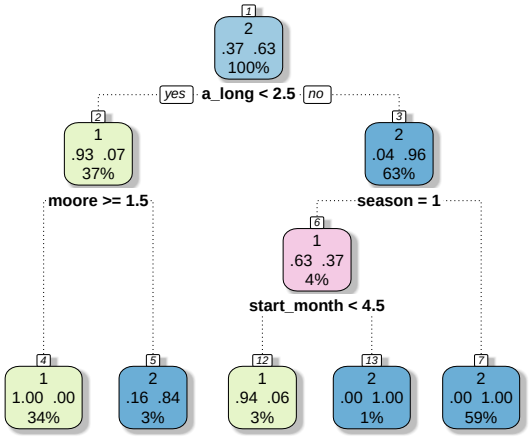


(i) $k = 9$

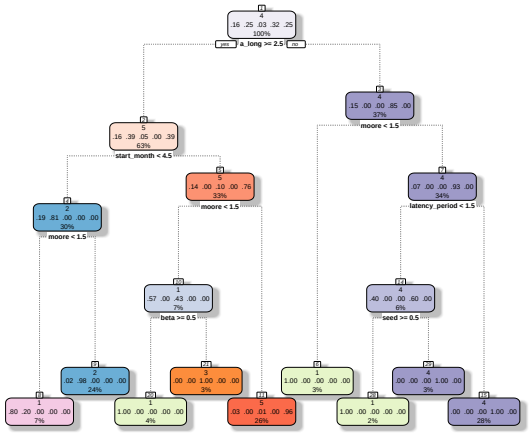


(j) $k = 10$

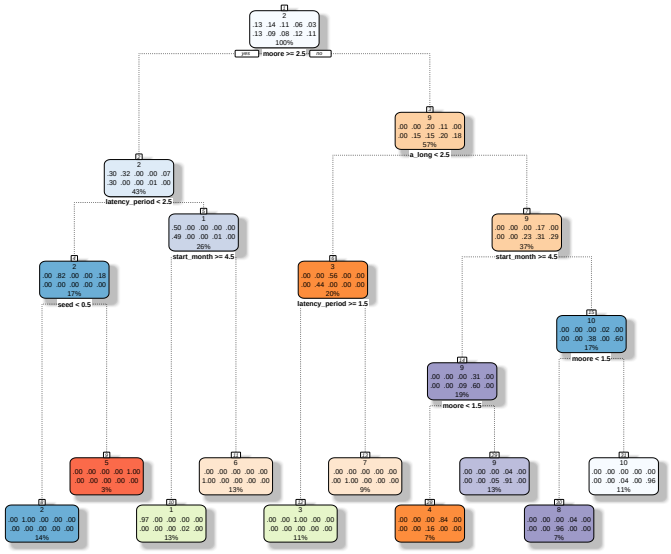
Figure S6: **Agglomerative clustering of simulation outputs.** The 9-level hierarchy of the clusters is shown in (a). At each level, we applied CART to understand which parameter was significantly influencing the split. Here, `a_long` corresponds to $\alpha_{\ell d}$, `latency_period` is ℓ and `moore` is r_{M}.



(a) $k = 2$



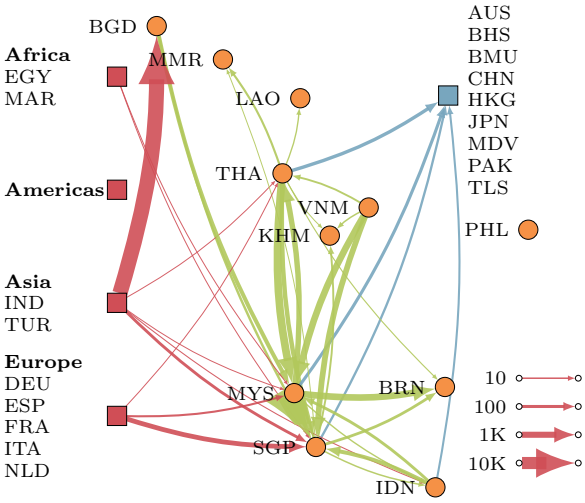
(b) $k = 5$



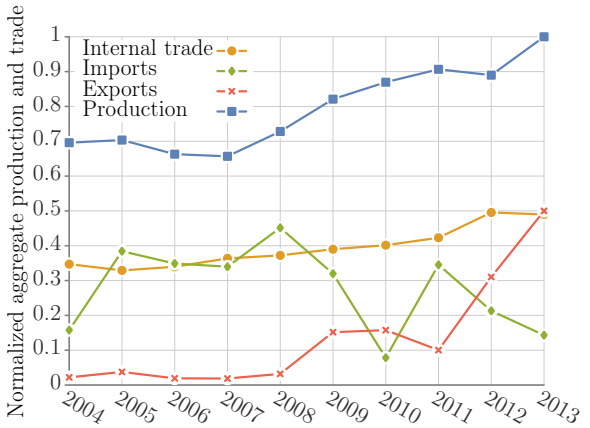
(c) $k = 10$

Figure S7: k**-means clustering of simulation outputs.** For each k, the number of clusters, we applied CART to understand which parameter was significantly influencing the split. Here, `a_long` corresponds to $\alpha_{\ell d}$, `latency_period` is ℓ and `moore` is r_{M}.

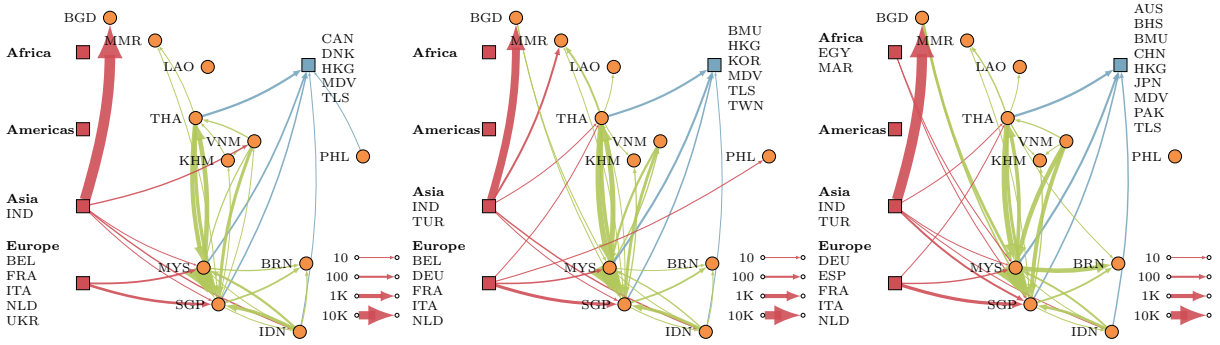
in the focus region. However, the majority of trading happens between neighbouring countries and except for a few flows, the volume is negligible compared to production within. Most significant flow is from Malaysia to Singapore ($\approx 30,000$ tonnes), India to Bangladesh ($\approx 20,000$ tonnes), 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 analysed 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



(a) Tomato trade in the focus region (2013).



(b) General trends of overall production and trade in the region over a decade based on FAOSTAT 2013.



(c) Evolution over years 2011, 2012, 2013.

General trends. Figure S8b shows the evolution of trade and production volume over time. The total tomato production in the study region obtained by summing the production for all countries is plotted for each year between 2004 to 2013 (blue). We computed the total tomato import volume per year (green) from outside the study region using FAOSTAT Trade Matrix (Table S1). This was obtained by aggregating imports where the importing country belongs to the study region and the exporting country is outside the region. Similarly export volume per year to outside the

study region (red) is shown. Finally, the amount of tomato traded within the study region was obtained by aggregating tomato trade between countries in the region for each year (orange). In order to compare these plots, each plot was normalised by dividing by the corresponding maximum value across all years. There is a steady increase in the production and amount of internal trade, with comparable rate of change for both quantities. In comparison, the export of tomato to outside of the focus region (plot "Exports") has risen steeply in the recent years (after 2011), while the imports generally indicate a downward trend (plot "Imports"). Recent efforts to increase production and trade infrastructure in these countries support these trends [8, 27, 44]. Under these circumstances, *T. absoluta*'s invasion can have a high negative impact on the economy and livelihood of the people in this region.

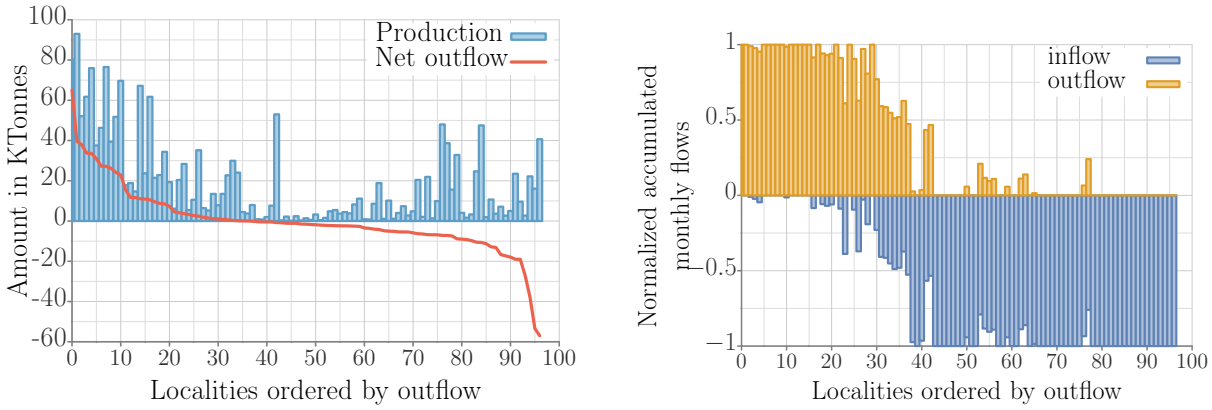
S6.2 Localities, production and trade

We identified 102 localities representing urban and production centres. For the purpose of this discussion, a *major source* (*major sink*) is a locality with net outflow (inflow) at least 10 Kilo tonnes. Among these, less than 20% of the localities have net outflow of at least 10 Kilo tonnes in the gravity model derived trade network, indicating that there are few major sources (Figure S9a). We also note that there are a few high sinks which have production comparable to some of the major sources. These are regions not only vulnerable to attacks, but also potentially sustain a huge negative impact due to loss in production.

To predict seasonal production, we used linear regression to model the production rate as a function of precipitation, temperature, and elevation using seasonal production data from different regions of Philippines. The results showed that precipitation was a statistically significant predictor ($p < 0.001$). More details are in Section S3.2. To study the effect of seasonality on monthly flows, we accumulated the net flow per month by directionality. For a locality i, let $F_i(m)$ denote the net flow in month m (positive means outflow, else inflow). The accumulated outflow is $F_i^+ = \sum_{m=1}^{12} I\big(F_i(m) > 0\big)F_i(m)$, where $I(\cdot)$ is the indicator function. Corresponding definition of accumulated inflow is $F_i^- = \sum_{m=1}^{12} I\big(F_i(m) < 0\big)F_i(m)$. The results in Figure S9b show that there are very few localities which switch from the role of source to sink and vice versa. This is unlike what is observed in Nepal by Venkatramanan et al. [57]. The flow between high altitude regions and low altitude regions switches direction according to season. Here, while the intensity of flows vary between seasons, we do not see such a switch in roles.

S6.3 Cellular automata model from Guimapi et al

Guimapi et al. [21] use a cellular automata approach to model the spread of *T. absoluta* through the study region of Africa, Spain, and Portugal. Here, we implemented and applied the model to study the spread in the focus region.



(a) Net annual production and outflow

(b) Effect of seasonality on monthly flows

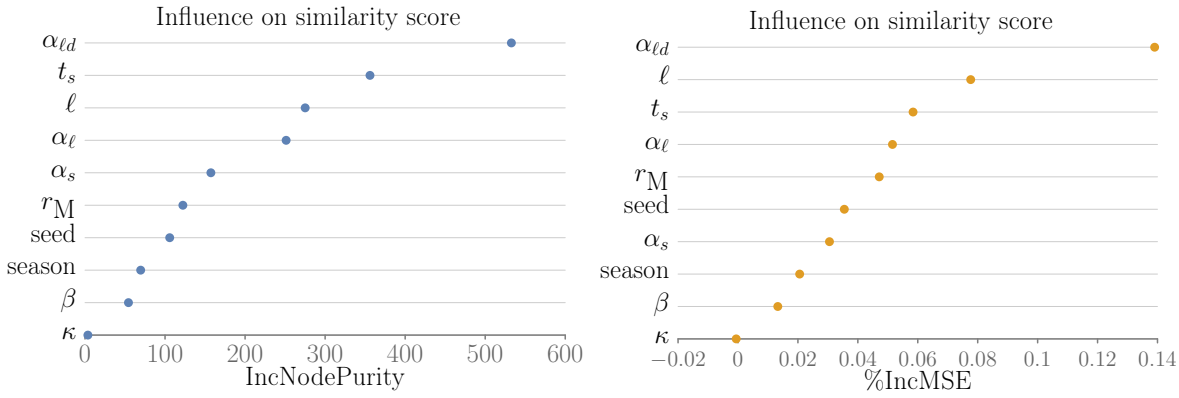
Figure S9: **Production, inflow and outflow at localities.** These are results for $\beta = 2$ and $\kappa = 500$ (i)

Methodology. Each cell in the automata is of size $25\text{km} \times 25\text{km}$ induced by a grid overlayed on the study region. Both square and hexagonal cell configurations are considered, but square cells (Moore neighbourhood) were chosen in order to cover a larger area per cell. Each cell can be in one of three states: S, E and I with very similar definitions as in our model. The spread is modelled as follows: For an infected cell to spread to another, each cell within its moore neighbourhood of respective distance is considered in each time step (one month). If this neighbour cell is suitable, it will change states. The suitability of a cell for pest establishment depends on temperature, humidity, NDVI, and tomato production in that cell at the given time step. Two Moore neighbourhoods were considered, one with a range of 50 km and another with a range of 75 km.

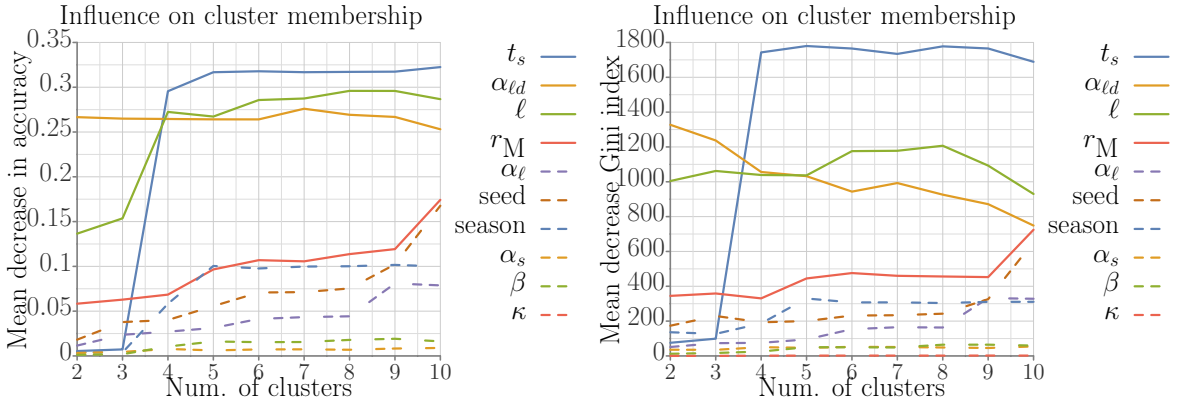
Given a cell within the Moore neighbourhood of an infected cell, its state in the next time step depends on the following rules.

- If NDVI $\geq$ NDVI threshold 1, then state becomes "Exposed".
- If NDVI $\geq$ NDVI threshold 1 and temperature $>=$ temperature threshold and humidity $>=$ humidity threshold, then state becomes "Invaded".
- If NDVI $\geq$ NDVI threshold 2 and tomato production high or very high, then state becomes "Invaded".
- Otherwise the state remains unchanged

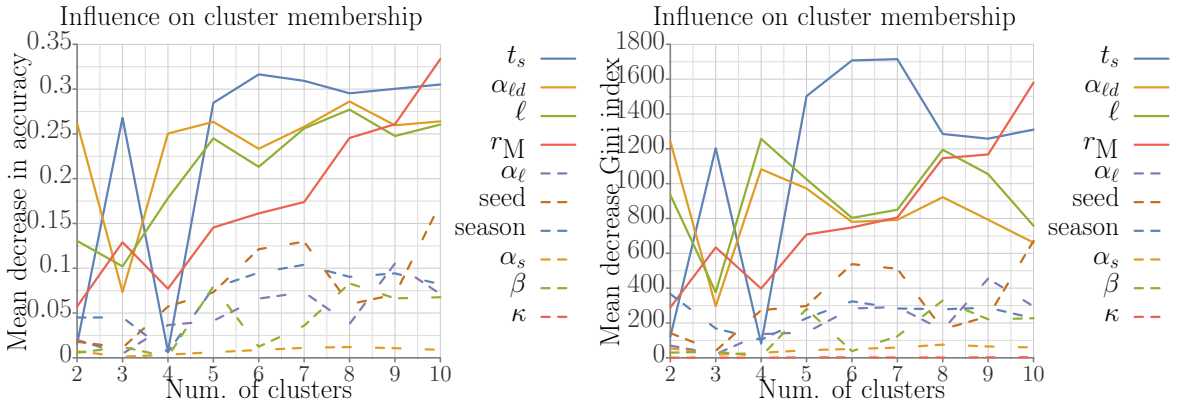
Temperature, humidity, and NDVI data is at the level of the cell, while production data is at the country-level. The temperature threshold is 22 degrees Celsius and humidity threshold is 55%. The NDVI threshold 1 is 0.1, and threshold 2 is 0.3. The production yield threshold is 10 tons/hectare to be considered high or very high. Thus if a cell is within a country with high or very high yield, that cell would be classified at high or very high yield.



(a) Similarity score ($\mathcal{S}$) based evaluation.



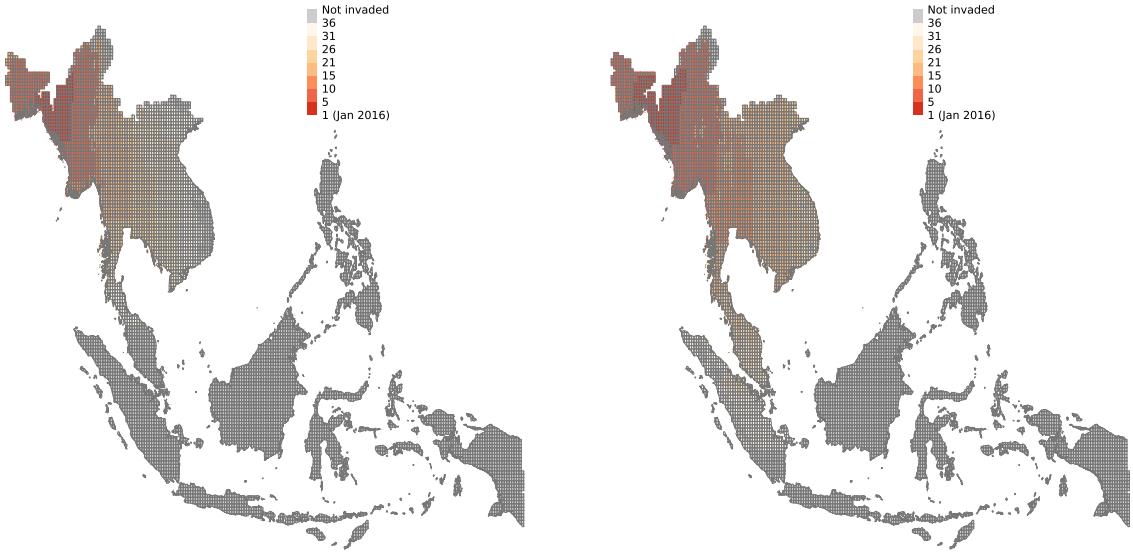
(b) Agglomerative clustering



(c) k-means clustering

Figure S10: **Parameter importance.** Continued from Figure 2c

Results. We applied the seeding scenarios described in Table S5. Each scenario was simulated with Moore neighbourhoods of 2 and 3. The results are shown in Figure S11. The model predicts a rapid radial spread in this region as vegetation and temperature are generally favourable throughout this region. The tomato production yield in each country in the focus region is in the high or very high categories, and thus did not limit the spread. Generally, the predicted range expansion is much faster compared to our models, particularly in the case where Moore range is 3.

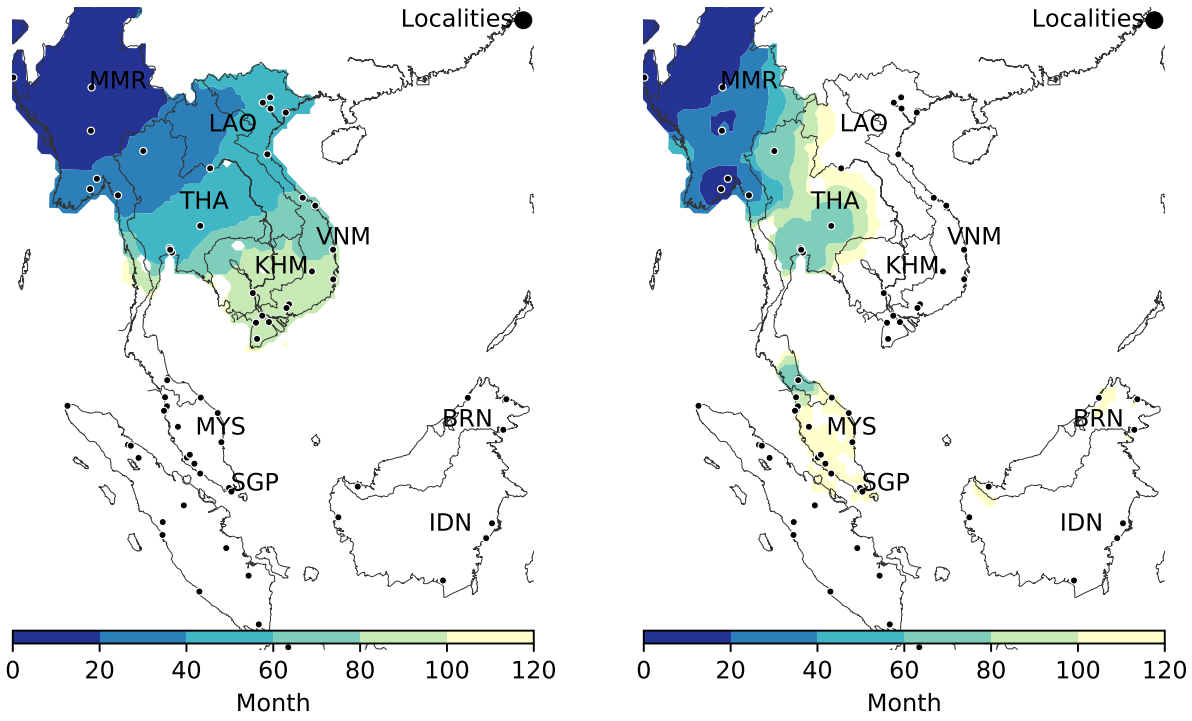


(a) Moore neighbourhood of 2

(b) Moore neighbourhood of 3, baseline model

Figure S11: Guimapi et al. [21] model results: Here the region of Chin in northeast Myanmar was seeded.

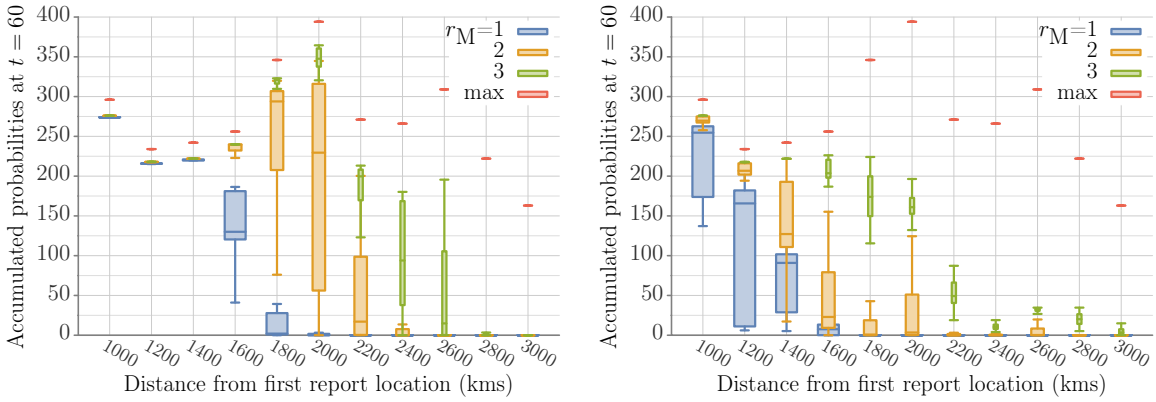
S6.4 Spread patterns and rate



(a) Class A

(b) Class B

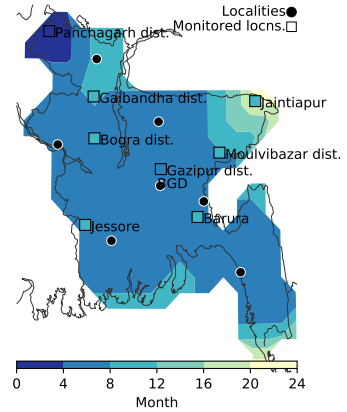
Figure S12: **Spread pattern in Mainland Southeast Asia without accounting for trade between countries.** The contour plots show the simulated spread starting from northern Myanmar for 120 time steps or 10 years. Representative simulation outputs for Class A and Class B models are shown. With Class A, spread is faster eastward than southward and the other way around in the case of Class B.



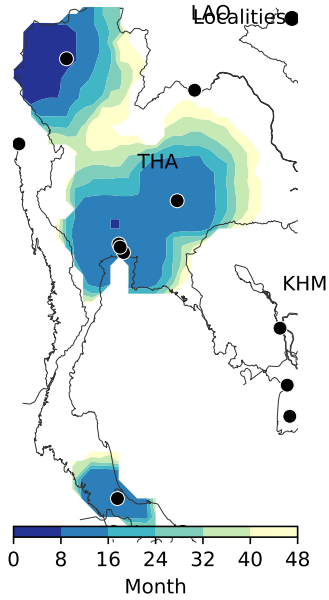
(a) Class A

(b) Class B

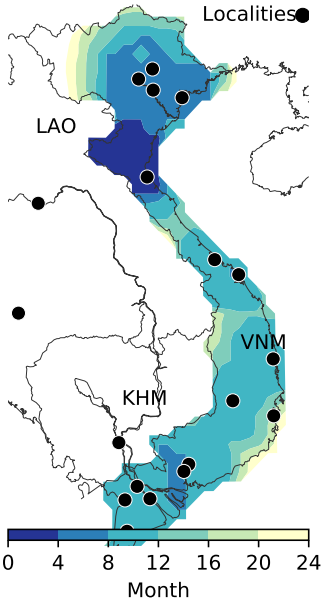
Figure S13: **Spread in Mainland Southeast Asia with respect to distance from origin.** The cells are binned based on their distance from the origin of infection (Northern Myanmar). Given time step t (60 or five years from time of start in this case), let $\Pr(v, \leq t)$ be the probability that cell 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. Configurations 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}.



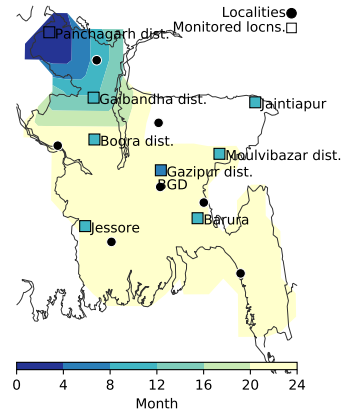
(a) Bangladesh: without intervention



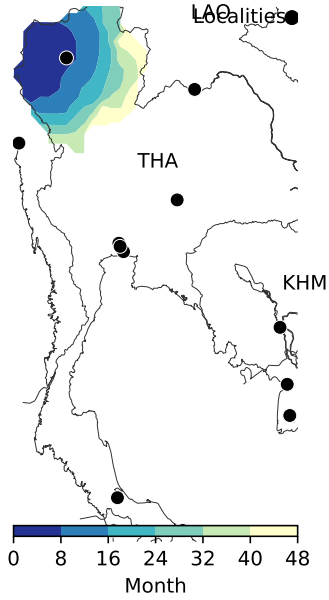
(b) Thailand: without intervention



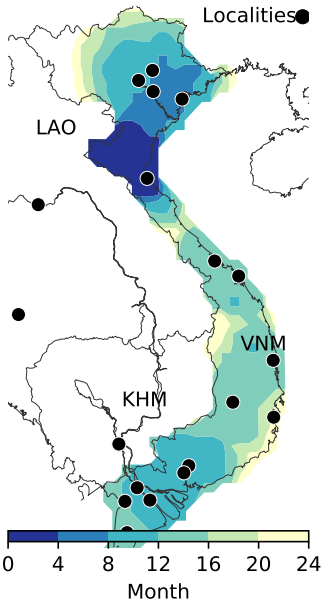
(c) Vietnam: without intervention



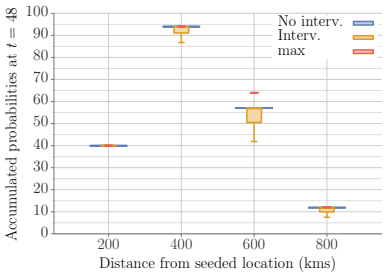
(d) Bangladesh: with intervention (100%)



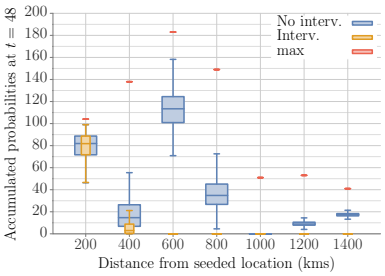
(e) Thailand: with intervention (100%)



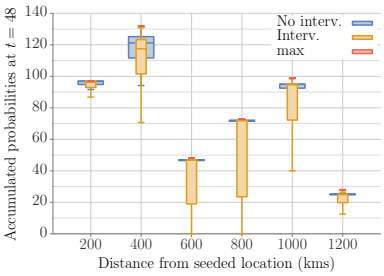
(f) Vietnam: with intervention



(g) Bangladesh: spread intensity and distance with respect to origin



(h) Thailand: spread intensity and distance with respect to origin



(i) Vietnam: spread intensity and distance with respect to origin

Figure S14: **Interventions.** This is a continuation of Figure 3 from the main document.

References

[1] Google API. Retrieved March, 2018.

[2] Abhijin Adiga. Spread multi-pathway model. https://github.com/rmuniappan/SPREAD_model.git, 2019.

[3] Japan International Cooperation Agency. Data collection survey on selecting the processed food to be focused and promoting foreign direct investment in food business in Vietnam.

[4] American Digital Cartography, Inc. Adc worldmap v. 7.4, 2018. Last accessed: June 2018.

[5] I. W. Arsanti and M. H. Bohme. Assessing vegetable cropping patterns in upland areas of Indonesia. *Acta Horticulturae*, 1103:37–42, 2015.

[6] Bangladesh Bureau of Statistics. Yearbook of Agricultural Statistics. http://bbs.portal.gov.bd, 2017.

[7] P. J. Batt, M. T. Lopez, J. T. Axalan, L. A.T. Hualda, M. O. Montiflor, and S. B. Concepcion. Exploring the institutional market for fresh vegetables in the Southern Philippines. *Acta Horticulturae*, 895:59–68, 2011.

[8] B. Buntong, V. Srilaong, T. Wasusri, S. Kanlayanarat, and A. L. Acedo. Reducing postharvest losses of tomato in traditional and modern supply chains in Cambodia. *International Food Research Journal*, 20(1):233–238, 2013.

[9] Jean-Joseph Cadilhon, Paule Moustier, Nigel D Poole, Thi Giac Tam Phan, and Andrew P Fearne. Traditional vs. modern food systems? Insights from vegetable supply chains to Ho Chi Minh City (Vietnam). *Development Policy Review*, 24(1):31–49, 2006.

[10] Sylvia B Concepcion. Consumer market segments in the Philippine vegetable industry. *IFAMA Forum and Symposium*, pages 1–8, 2009.

[11] Sylvia B Concepcion. Institutional markets for fresh vegetables Trends. pages 1–30, 2011.

[12] Odo Diekmann and Johan Andre Peter Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, volume 5. John Wiley & Sons, 2000.

[13] Earth System Research Laboratory. CPC Merged Analysis of Precipitation (CMAP). https://www.esrl.noaa.gov/psd/data/gridded/data.cmap.html.

[14] Export-Import Bank of India. Bangladesh: A Study of India's Trade and Investment Potential. 2015.

[15] FAO. Food supply quantity. http://www.fao.org/faostat/en/#data/CC, 2016.

[16] FAO. Production and trade. http://www.fao.org/faostat/en/#data, 2016.

[17] FAO. Per capita tomato consumption. http://www.helgilibrary.com/indicators/tomato-consumption-per-capita/world/, Accessed in February 2018.

[18] Christian Genova II, Katinka Weinberger, Srun Sokhom, Mong Vanndy, and En Chan Yarith. *Postharvest Loss in the Supply Chain for Vegetables–The Case of Tomato, Yardlong Bean, Cucumber and Chinese Kale in Cambodia*. AVRDC-WorldVegetableCenter, 2006.

[19] Google. Distance Matrix API. https://developers.google.com/maps/documentation/distance-matrix/, 2017.

[20] G J H Grubben. Timing of vegetable production in Indonesia. In *VI Symposium on the Timing of Field Production of Vegetables*, pages 261–270, 1989.

[21] Ritter YA Guimapi, Samira A Mohamed, George O Okeyo, Frank T Ndjomatchoua, Sunday Ekesi, and Henri EZ Tonnang. Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecological Complexity*, 28:77–93, 2016.

[22] S H Hengky. Competition Gaps of Tomatoes' Industries in Cameron Highland, Malaysia. *Business and Economic Research*, 6(2):368, 2016.

[23] Herbert W Hethcote. A thousand and one epidemic models. In *Frontiers in mathematical biology*, pages 504–515. Springer, 1994.

[24] Herbert W Hethcote and Simon A Levin. Periodicity in epidemiological models. In *Applied mathematical ecology*, pages 193–211. Springer, 1989.

[25] Robert J Hijmans, Susan E Cameron, Juan L Parra, Peter G Jones, and Andy Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology*, 25(15):1965–1978, 2005.

[26] M. S. Hossain, M. Y. Mian, and R. Muniappan. First record of *Tuta absoluta* (Lepidoptera: Gelechiidae) from Bangladesh. *Journal of Agricultural and Urban Entomology*, 32(1):101–105, 2016.

[27] Pham Thi Thu Huong, A. P. Everaarts, J. J. Neeteson, and P. C. Struik. Vegetable production in the Red River Delta of Vietnam. I. Opportunities and constraints. *NJAS - Wageningen Journal of Life Sciences*, 67:27–36, 2013.

[28] The Independent. Rangpur region to produce 7.02 lakh tonnes winter vegetables. http://www.theindependentbd.com/printversion/details/68010.

[29] K Itharattana. Market prospects for upland crops in Thailand. (November), 1996.

[30] Greg I. Johnson, Katinka Weinberger, and Mei-huey Wu. *The vegetable industry in tropical asia: Thailand. An overview of production and trade*. Number January 2008. 2008.

[31] Pablo Kaluza, Andrea Kölzsch, Michael T Gastner, and Bernd Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48):1093–1103, 2010.

[32] Rubayet Karim and Jony Biswas. Value Stream Analysis of Vegetable Supply Chain in Bangladesh: A Case Study. *International Journal of Managing Value and Supply Chains*, 7(2):41–60, 2016.

[33] Somsack Kethonga, Khamtanh Thadavong, and Paule Moustier. Vegetable Marketing in Vientiane. (November):56, 2004.

[34] Nazery Khalid. The development of ports and shipping sector in malaysia. *Maritime Institute of Malaysia*, pages 1–15, 2005.

[35] Frauke Kraas, Hartmut Gaese, and Mi Mi Kyi. *Megacity Yangon: Transformation Processes and Modern Developments: Second German-Myanmar Workshop in Yangon, Myanmar, 2005*. Lit Verlag, 2006.

[36] LandScan Geographic Information Science & Technology. LandScan 2016. https://landscan.ornl.gov/.

[37] Merlyn P Ledesma. Northern Mindanao vegetables. Technical report, Department of Agriculture RFU- 10, 2011.

[38] Júlio C Martins, Marcelo C Picanço, Ricardo S Silva, Alfredo HR Gonring, Tarcísio VS Galdino, and Raul NC 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.

[39] MaxMind. Cities database. https://www.maxmind.com/en/free-world-cities-database.

[40] Lauren Ancel Meyers. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*, 44(1):63–86, 2007.

[41] Ministry of Agriculture, Malaysia. Vegetable production by state. http://www.moa.gov.my/, Accessed April 2018.

[42] Ministry of Information and Communication Technology, Thailand. Agricultural census. http://web.nso.go.th/en/census/agricult/cen_agri03.htm, 2013.

[43] Zaw Aye Moe. Overview of food accessibility situation in Myanmar Zaw. *Thematic papers on Myanmar census of agriculture*, (May), 2013.

[44] Paule Moustier. Final summary report of SUSPER. *Sustainable Development of Peri-Urban Agriculture in South-East Asia*, (April):152, 2007.

[45] NASA Earth Observations (NEO). Normalized Difference Vegetation Index (NDVI). https://neo.sci.gsfc.nasa.gov.

[46] NASA Langley Research Center Atmospheric Science Data Center Surface meteorological and Solar Energy (SSE) web portal. Relative humidity. https://eosweb.larc.nasa.gov/sse/.

[47] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

[48] Andrei Novikov. annoviko/pyclustering: pyclustering 0.8.2 release, November 2018.

[49] Philippines Statistics Authority. Major vegetables and rootcrops quarterly bulletin. https://psa.gov.ph/content/major-vegetables-and-rootcrops-quarterly-bulletin, 2017.

[50] San-San-Yi, S. A. Jatoi, T. Fujimura, S. Yamanaka, J. Watanabe, and K. N. Watanabe. Potential loss of unique genetic diversity in tomato landraces by genetic colonization of modern cultivars at a non-center of origin. *Plant Breeding*, 127(2):189–196, 2008.

[51] E. Sankarganesh, D.M. Firake, B. Sharma, V.K. Verma, and G.T. 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.

[52] Chhean Sokhen, Diep Kanika, and Paule Moustier. Vegetable market flows and chains in phnom penh. 2004.

[53] Daily Sun. Bumper winter vegetable production likely in rangpur region. http://www.daily-sun.com/post/282672/Bumper-winter-vegetable-production-likely-in-Rangpur-region.

[54] Serigne Elhadji Sylla, Thierry Brévault, Lucie Monticelli, Karamoko Diarra, and Nicolas Desneux. Geographic variation of host preference by the invasive tomato leafminer *Tuta absoluta*: implications for host range expansion. *Journal of Pest Science, Accepted*, 2018.

[55] Henri EZ Tonnang, Samira F Mohamed, Fathiya Khamis, and Sunday 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.

[56] Chuthaporn Vanit-Anunchai. Possibilities and constraints of marketing environmentally friendly produced vegetables in Thailand. *Faculty of Economics and Management*, 2006.

[57] S Venkatramanan, S Wu, B Shi, A Marathe, M Marathe, S Eubank, LP Sah, AP Giri, LA Colavito, KS Nitin, et al. Modeling commodity flow in the context of invasive species spread: Study of tuta absoluta in Nepal. *Crop Protection*, 2019.

[58] Tran Duc Vien. Overview on tomato production and tomato varieties in Vietnam. (March):1–10, 2003.

[59] Katinka Weinberger and Christian A Genova II. *Vegetable production in Bangladesh: commercialization and rural livelihoods*. AVRDC-WorldVegetableCenter, 2005.

[60] Siebe Van Wijk and A. P. Everaarts. The market for vegetables in North Vietnam. Technical report, Wageningen University, 2007.

[61] L. You, U. Wood-Sichra, S. Fritz, Z. Guo, L. See, and J. Koo. Spatial Production Allocation Model (SPAM) 2005 v3.2. http://mapspam.info, 2017.