Modeling Commodity Flow in the Context of Invasive Species Spread: Study of *Tuta absoluta* in Nepal

S. Venkatramanan[1], S. Wu[2], B. Shi[4], A. Marathe[1,5], M. Marathe[1,3], S. Eubank[1,5], L. P. Sah[7,8,9], A. P. Giri[7,8,9], L. A. Colavito[7,8,9], K. S. Nitin[10], V. Sridhar[10], R. Asokan[10], R. Muniappan[7], G. Norton[6], A. Adiga[1]

[1]Biocomplexity Institute & Initiative, University of Virginia

[2]Department of Computer Science, Virginia Tech

[3]Department of Computer Science, University of Virginia

[4]Department of Economics, Virginia Tech

[5]Department of Public Health Sciences, University of Virginia

[6]Department of Agriculture and Applied Economics, Virginia Tech

[7]Feed the Future Integrated Pest Management Innovation Lab

[8]Feed the Future Asian Vegetable and Mango Innovation Lab

[9]International Development Enterprises, Nepal

[10]Indian Institute of Horticultural Research

Trade and transport of goods is widely accepted as a primary pathway for the introduction and dispersal of invasive species. However, understanding commodity flows remains a challenge owing to its complex nature, unavailability of quality data, and lack of systematic modeling methods. A robust network-based approach is proposed to model seasonal flow of agricultural produce and examine its role in pest spread. It is applied to study the spread of *Tuta absoluta*, a devastating pest of tomato in Nepal. Further, the long-term establishment potential of the pest and its economic impact on the country are assessed. Our analysis strongly indicates that trade plays an important role in the spread of *T. absoluta*. The economic impact of this invasion could range from USD 17-25 million. The proposed approach is generic and particularly suited for data-poor scenarios.

1 Introduction

Invasive alien species spread is a complex phenomenon driven by various natural and anthropogenic factors. While the knowledge of biology and climate is essential to assess establishment risk and devise sustainable management strategies (Sutherst 2000), it is critical to understand human-mediated pathways to prevent introduction and mitigate immediate impact (Hulme 2009; Banks et al. 2015). Increasingly, the need for a comprehensive outlook to tackle this problem is being highlighted (Cunniffe et al. 2015; Robinet et al. 2012; Carrasco et al. 2010); yet studies that analyze trade and travel in the context of food safety are only beginning to emerge (Colunga-Garcia and Haack 2015; Tatem 2009; Ercsey-Ravasz et al. 2012; Nopsa et al. 2015; Early et al. 2016).

As is the case with many built infrastructures, trade of goods naturally yields to network representations. Typically, nodes of the network represent locations–ranging from markets to continents depending on the context–connected by transportation infrastructure. The influence of one node on another is captured by an edge with weight that is a function of the transaction volume across that edge. A major challenge in constructing such networks is estimating the temporal flows. The intricate web of supply chain logistics makes it hard to document transactions, and even in economically developed countries, obtaining commodity specific flow data is a challenge (Magarey et al. 2011). On the other hand, it is comparatively easy to obtain datasets on production, population, and economic indicators at finer spatial resolution, thus allowing the use of spatial interaction models such as the gravity model (Haynes et al. 1984). Such an approach is also better suited for data-poor regions.

A representative flow network can yield valuable insights into the phenomenon. The network structure helps identify possible entry points and hubs (Nopsa et al. 2015; Sutrave et al. 2012). Network dynamical processes such as the SEIR (Susceptible → Exposed → Infected → Recovered) model from epidemiology (Pastor-Satorras et al. 2015) are applied to capture the spatio-temporal evolution of the invasion. Model selection and validation is challenging due to the lack of accurate ground surveillance, since very few countries have the capacity to effectively react to impending and emerging invasions (Early et al. 2016). Therefore, there is a need for a robust modeling framework that can operate with limited observational data.

This work addresses three key issues that current modelers encounter: (*i*) lack of a systematic approach to investigate the role of human-mediated pathways, (*ii*) the difficulty in evaluating models in the absence of accurate validation datasets, and (*iii*) the need to synthesize disparate datasets and modeling methodologies in order to provide a comprehensive assessment of the situation. Our framework is outlined in Figure 1. We apply it to study the spread dynamics of *Tuta absoluta* (*Gelechiidae, Lepidoptera*) (Meyrick,1917), a devastating pest of the tomato crop. The region of interest is Nepal, a biodiversity hotspot (Kindlmann 2011) and largely agrarian economy, which recently reported *T. absoluta* invasion in 2016 (Bajracharya et al. 2016).

Indigenous to South America, *Tuta absoluta* or the South American tomato leaf miner was accidentally introduced to Spain in 2006 (Desneux et al. 2010). It has rapidly spread throughout Europe, Africa, Western Asia, the Indian subcontinent, and parts of Central America (Biondi et al. 2018) over the past decade. Since tomato is among the top two traded vegetables worldwide (http://www.fao.org), 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, e.g., in Netherlands (NPPO 2009). Karadjova et al. (2013) observed that the spread pattern in Bulgaria was correlated with prime trade routes. With tomato being a commercially important crop (Grousset et al. 2015), *T. absoluta* has had a significant global impact. For example, in Turkey, the annual estimated intervention cost is €167M per year (Oztemiz 2014). Due to extensive insecticide treatment in Europe (Gontijo et al. 2013), insecticidal resistance has been recently observed in populations of *T. absoluta* (Guedes and Siqueira 2013). Lack of effective natural enemies has made integrated pest management challenging.

We propose a network-based model to represent seasonal flow of agricultural produce among major markets accounting for production, population, per capita income, and transportation infrastructure. Under the hypothesis that trade is the primary driver of spread, pest dispersal is modeled as a stochastic dynamical process over the seasonal flow networks. Applying techniques from an emerging field in network science, namely *epidemic origin inference* (see for example (Shah and Zaman 2011), we establish a strong correlation between the expected spread over the flow network and observational data. This is the first application of this method in the context of invasive species spread.

Our analysis strongly indicates that given its dependency on vegetable imports and the pattern of agricultural produce flow unraveled by our model, Nepal is extremely vulnerable to attacks from pests and pathogens that can spread through trade. Our assessment of suitability for *T. absoluta* establishment based on growth and stress factors and observational data is that except for the mountainous regions in the north, *T. absoluta* can survive year-round in Nepal. However, the highest damage potential is in the southeastern parts. Our assessment of the economic impact based on a partial equilibrium approach predicts a social welfare loss of $17.5-24.7 million due to this invasion.

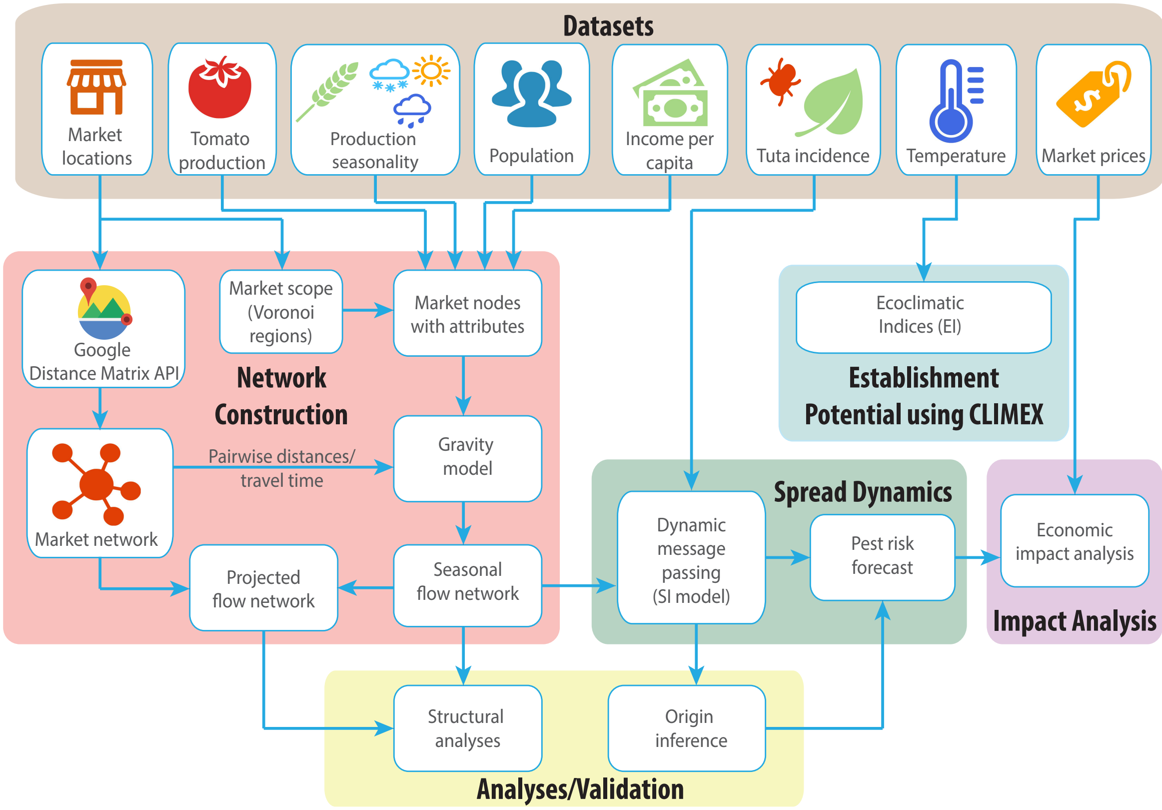


Figure 1: A schematic of the modeling framework showing the diverse datasets used and the different modules.

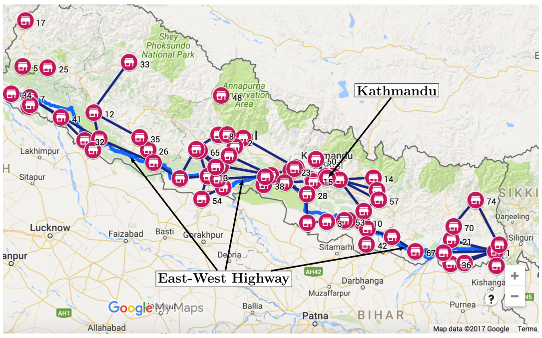
2 Methods

2.1 Data

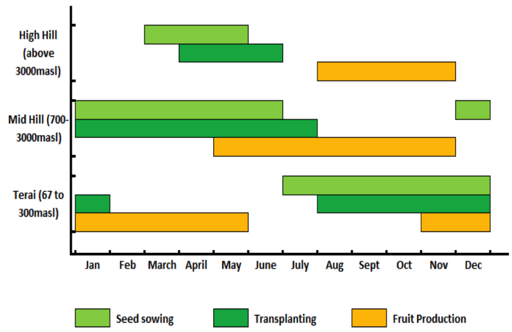
Datasets used in our framework are listed in Table ??, Supplementary Information. We link several open-source datasets along with qualitative inputs from local experts in order to model the seasonal trade of the tomato crop as well as pest dynamics. Some of the challenges arise from the fact that the datasets vary in their spatial and temporal resolution and their year of release. Owing to the unique geography of Nepal, the vegetable production cycle varies with altitude (see Figure 2b). The annual production data was combined with the knowledge of production cycle to model the spatio-temporal variation in production across seasons. Major vegetable markets were geolocated using Google Maps, and Google Distance Matrix API was used to construct the road network and compute travel times. Several organizations have been involved in the monitoring of *T. absoluta* spread in Nepal: Nepal Agricultural Research Council (NARC), United States Agency for International Development (USAID), iDE Nepal, ENBAITA, and Agricare Pvt. Ltd. The pest was monitored using pheromone traps that were installed in several Village Development Committees (VDC). In May 2016, *T. absoluta* was officially reported by NARC's entomology division in Lalitpur (near Kathmandu). More details are provided in Section ?? of Supplementary Information. For the purpose of our analysis, we used surveillance data until May 2017, a year after the first official report (Figure 2c).

2.2 Nepal's physiography, tomato production, trade and demographics

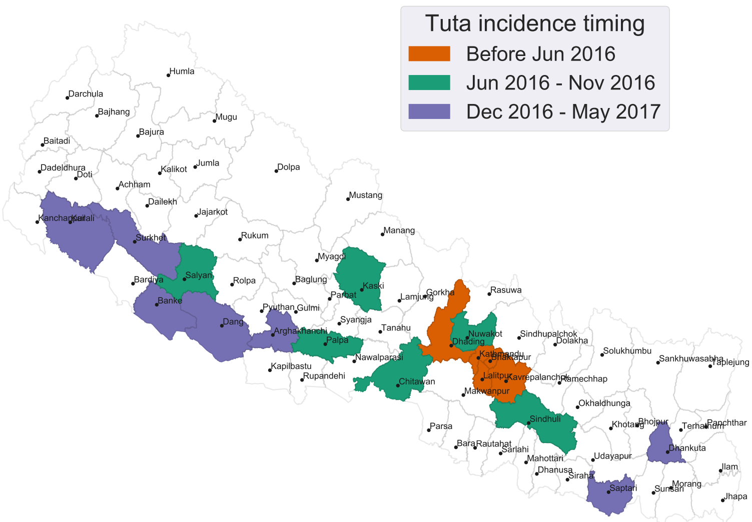
Despite being a small country (800km along the Himalayan axis, 150-250km across), Nepal has high geographic diversity owing to its altitude. Along the north-south axis, Nepal is divided into three belts, namely Terai, Mid Hills and High Hills (or Mountains) (Figure 3a). Terai is a densely populated region (with more than 400 people/sq.km. (Ali 2000)), while the High Hills are the least populated (with the exception of Kathmandu valley, which has the highest population density in Nepal). The altitude variations also impose different seasonal cycles for agricultural production (see Figure 2b). Production in the Mid Hills and High Hills is largely restricted to the summer months of May to November, while the Terai region produces during the winter months of November to May. Even though Nepal imports tomatoes from India all year round, the volume is not significant enough to influence domestic trade. For instance, in 2014 Nepal exported only 1% of its tomatoes and imported about 6-7% of its total consumption (http://www.fao.org/faostat). Also, the tomato processing industry in Nepal is not well developed, and fresh tomatoes are mostly traded for consumption (Nepal Economic, Agriculture, and Trade 2011). This motivates the use of population and per capita income as indicators of tomato consumption in a given district. Nepal is divided into development regions from east to west as shown in Figure 3e. The Central Development Region (CDR), which includes Kathmandu, generates 80% of the country's revenue, and 60% of government expenditure is allotted to it (Vandernoot and Van Hove 2014). In comparison, the revenue generated by the Far Western Development Region was less than 5% of that of CDR.



(a) Market network



(b) Production seasonality



(c) *T. absoluta* incidence timings

Figure 2: Market locations, production cycle and *T. absoluta* monitoring and distribution in Nepal. Data shown represents surveillance until May 2017.

2.3 Modeling the domestic tomato trade network

The domestic trade of tomato is modeled as a network of flows between major vegetable markets in Nepal based on the following premise: the total outflow from a market is a function of the amount of produce in its surrounding regions, and the total inflow is a function of consumption, which in turn is a function of the population to which it caters and the corresponding per capita income. The main assumptions in this model are based on the discussion in the previous section. They are: (i) imports and exports are not significant enough to influence domestic trade; (ii) fresh tomatoes are mainly traded for consumption as there is negligible processing; and (iii) the higher the per capita income, the greater the consumption.

The nodes of the flow network are the major markets, 69 in all (Figure 2a). The seasonal flows are estimated using a doubly constrained gravity model (Kaluza et al. 2010; Anderson 2011). The flow F_{ij} from location i to j is given by $F_{ij} = a_i b_j O_i I_j f(d_{ij})$ where, O_i is the total outflow from i, I_j is the total inflow to j, d_{ij} is the travel duration from i to j, $f(\cdot)$ is the *distance deterrence function*, and coefficients a_i and b_j are computed through an iterative process to ensure flow balance. The reader should note that the resulting network can contain self loops representing volume consumed locally. The total outflow O_i is a function of the seasonal production assigned to that market. The total inflow I_j is a function of consumption at the market. Based on Figure 2b, we identified two seasons – S1: June to November and S2: December to May. We partitioned the districts into two groups: Mid Hills and High Hills belong to group 1, while the Terai districts belong to group 2. All districts belonging to group i were assigned their respective annual production for season Si and zero for the other season. Thus, we derived two distinct flow networks, one for each season by assigning for every district, its production to one season. The details of how these quantities were estimated is in Section **??**, Supplementary Information. Also provided are the methods used for validation and sensitivity analysis.

2.4 Spread dynamics

We apply a discrete-time SI (Susceptible-Infected) epidemic model (Pastor-Satorras et al. 2015) on the domestic trade networks to simulate the spread of *T. absoluta* by the domestic trade pathway. Each node is either susceptible S (free from pest) or infected I (pest is present). Henceforth, we use the term "infected" for a node or a region frequently to imply *T. absoluta* infestation at that location. A node i in state I infects each of its out-neighbors j in the network with probability proportional to the flow F_{ij} at each time step t. The infection probabilities are obtained by normalizing flows globally: $\lambda_{ij} = \frac{F_{ij}}{\max_{i,j} F_{ij}}$. The model is based on two assumptions: (i) an infected node remains infected and continues to infect its neighbors, and (ii) the chance of infection is directly proportional to the volume traded. Considering the fact that Nepal was ill-prepared for this invasion and the lack of effective intervention methods, (i) is a fair assumption. Historically, *T. absoluta* has spread rapidly in regions where tomato trade has been the highest (parts of Europe and Middle-East for example) thus motivating assumption (ii). The quantity of interest is the probability that node i is infected by time t given the initial condition f_0 (which assigns probability of infection at time step $t = 0$ to each node). We denote it by $P_I(i, t, f_0)$. In general, computing it is difficult. We adopt the *dynamic message passing algorithm* by Lokhov et al. (Lokhov et al. 2014) to estimate this. This is described in (Venkatramanan et al. 2017).

The initial configuration f_0 is chosen to mimic a spatially dispersed seeding scenario. We first select a *central* seed node, and then use a Gaussian kernel with parameter σ around the seed node to assign initial infection probabilities for neighboring markets. A market at a geodesic distance d from the seed is assigned the infection probability $e^{-\frac{d^2}{2\sigma^2}}$. The kernel accounts for factors such as uncertainty in determining the pest location, the possibility of spread of the pest through natural means, as well as interactions between these markets.

A Bayesian inference approach for model evaluation. Given that Nepal had limited resources to monitor the *T. absoluta* invasion, the pest incidence reports are not adequate. While it is hard to determine month of invasion at any location, some confidence can be placed on whether introduction happened by the end of a particular season. We evaluated our model based on the following backward inference problem: for an observation of node states at time t, what is the most likely origin of invasion (also known as the source detection problem (Shah and Zaman 2011))? We examined the likelihood of markets or regions being the source nodes, and in particular, we compared this with the likelihood of the region around Kathmandu being the source. Suppose $\mathcal{O}$ is the observation criterion; it consists of pairs (v, X) where v is a node and $X \in \{S, I\}$ is a state. For each candidate initial condition f_0, we estimated the joint probability of $\mathcal{O}$ at a time step t as a product of the marginal probability estimates from the message passing algorithm and define an *energy function* for each tuple (f_0, t) as

$$\phi(\mathcal{O}|f_0, t) = -\log\left(\prod_{(v,X)\in\mathcal{O}} P_X(i, t, f_0)\right)$$

The lower the value of ϕ, the higher the likelihood of f_0 being the initial condition. Secondly, recalling the uncertainty in interpreting time step t, we examined the relative likelihoods of each f_0 and the stability of the ranking across a range of model parameters. This is again discussed in (Venkatramanan et al. 2017).

2.5 Long-term establishment

In order to evaluate the long term establishment potential of *T. absoluta* in Nepal, various growth and stress factors regulating the development of *T. absoluta* were computed with particula reference to Kathmandu District. The parameter fitting values used in our study for CLIMEX modeling were iteratively changed after Tonnang et al. (Tonnang et al. 2015). This was done particularly for lower temperature threshold (DV0) and upper temperature threshold (DV3) to 7°C and to 40°C, respectively, in order to accommodate areas with *T. absoluta* infestation reported in Nepal. We used EI scaled up to 100 in order to get a clear gradient picture in terms of suitability of the pest in different geographical localities within Nepal, with 0 EI indicating non-suitable areas for *T. absoluta* establishment and areas with higher EIs indicating proportionately more favorable climatic conditions for the pest. Meteorological data of Nepal over the years indicates an increase in the average temperature at a rate of 0.05°C/year between 1975 to 2005 (Marahatta et al. 2009). Rising temperatures may a have a direct impact on the spread and establishment of the insect pests. Keeping this in view, an impact of 1°C rise in temperature scenario was also modeled to determine the near future potential areas for the establishment of *T. absoluta*. The resulting differences in EI values were mapped to find the influence of a 1°C rise in temperature for different regions in Nepal for *T. absoluta* establishment.

2.6 Economic impact

Direct economic impact. The direct impact measures the direct revenue loss from the tomato crop as determined by P $\sum_i hy_i l_i z_i$ which is the sum of loss encountered by each district. For each of Nepal's 75 districts, the spread dynamics module provides the probability of invasion and hence the proportion of area affected (z_i). The proportional loss in the affected area (h) is assumed to be 0.25. Here y_i is the yield per unit of land in district i before being affected and l_i is the tomato production area in the district.

Total economic impact. We use the partial equilibrium approach (Alston et al. 1995; Soliman et al. 2012), which focuses on the dynamics of the tomato market and assumes that the price for substitute and complementary goods are given, i.e. no changes occur in the substitute and complementary goods market due to changes in tomato prices. Due to *T. absoluta* infestation, the supply curve of tomato shifts leftwards in each district based on the probability of invasion. Assuming no changes occur in the downward sloping demand curve, shift in supply results in a higher equilibrium price and lower equilibrium quantity. This further impacts the consumers' and producers' welfare as measured by the consumers' surplus and producers' surplus (profits).

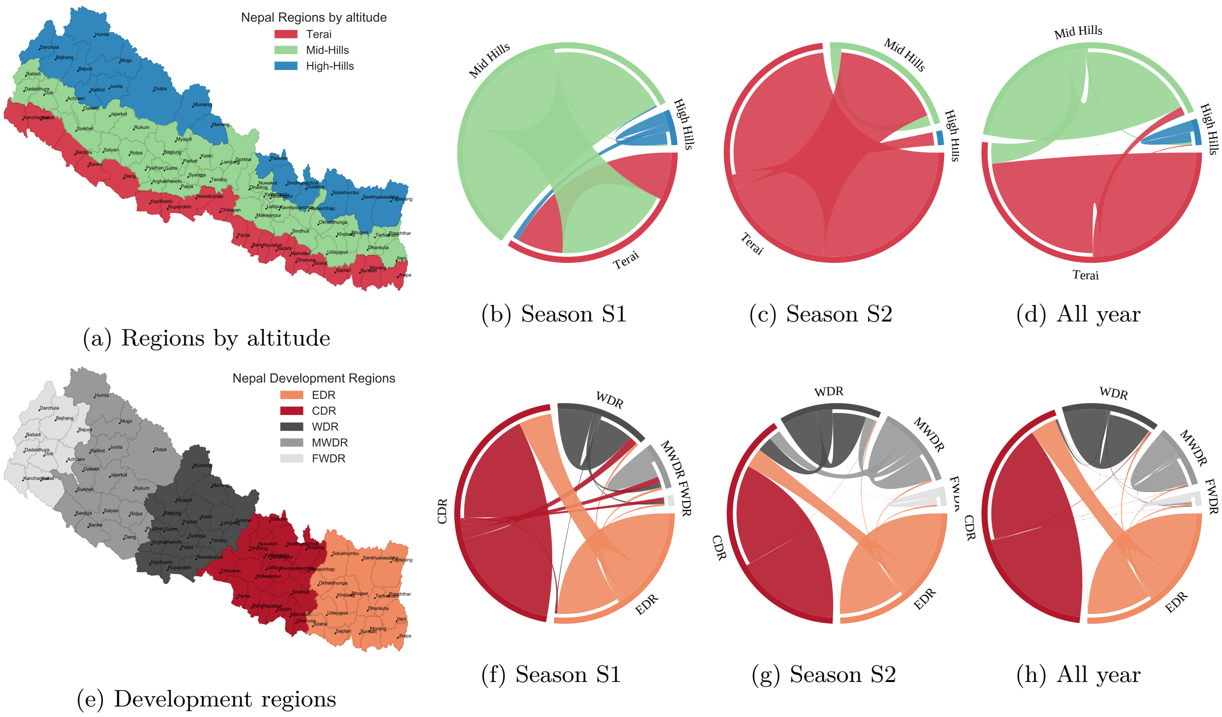
The total economic impact, or the change in social welfare, is the sum of change in consumers' and producers' surplus. To measure this, we first estimate the demand parameters given the initial price and then use it to estimate the new equilibrium price. Once both prices are known, we calculate the changes in consumers' and producers' surplus and hence the social welfare. For details on the economic impact calculation, please see the Supplementary Information.

3 Results

3.1 Structural properties of the trade network

Recall that there are two networks, one for each season S1 and S2, each comprising of 69 nodes. Analysis of the networks with respect to the gravity model parameters is discussed in (Venkatramanan et al. 2017). In Figure 3, we present representative observations of tomato trade between markets resulting from the gravity model. It is helpful to view these results in terms of the physiographic divisions (Figure 3a) and development regions (Figure 3e). Our model accounts for the fact that the High Hills/Mid Hills and the Terai are the primary sources of tomato during seasons S1 and S2 respectively. This is clearly reflected in the net flow diagram between geographic regions: north (High Hills/Mid Hills) to south (Terai) in S1 and south to north in S2. However, an interesting pattern to be noted is the significant flow from east to west during S1 as observed in the net flow diagram between the Development Regions. The East and Central Development Regions produce significantly more tomatoes than the other regions. Also, the presence of the arterial East-West highway that almost covers the entire breadth of the country (Figure 2a) helps in the east to west trade.

Comparison with the annual flow network. To evaluate the importance of seasons, we constructed the annual flow network by using the gravity model with annual production for each district. The resulting flows are shown in Figures 3h and 3d. Compared to the seasonal flows we see that annual flows are of shorter distance and thus there is not much flow between regions (either between east and west or south and north). This comparison clearly emphasizes the need to account for seasonal trade.



(a) Regions by altitude

(b) Season S1

(c) Season S2

(d) All year

(e) Development regions

(f) Season S1

(g) Season S2

(h) All year

Figure 3: **The spatio-temporal structure of the flow network.** S1 ranges from May toNovember, while S2 ranges from December to April. The first row shows the flow from east to west between development regions of Nepal. The second row depicts the flow from north to south between regions of different altitudes. While the second and third columns correspond to seasonal flow, the last column corresponds to the flows generated from annual data.

3.2 Role of trade network in pest spread

The experiment was setup under the premise that *T. absoluta* was first introduced to Nepal in the Kathmandu valley. Ground experts have high confidence in this assumption since the pest was not discovered in the previous growing season in other parts of Nepal. Given the pest reports till December 2016 (Figure 2c), we evaluate our model based on the following backward inference problem: for an observation of node states at time t, what is the most likely origin of invasion? This is also known as the source detection problem (Shah and Zaman 2011). Using the energy function defined in Methods, we examine the likelihood of markets or regions being the source nodes, and in particular, we compare this with the likelihood of the region around Kathmandu being the source (see Figure 4). Secondly, recalling the uncertainty in interpreting time step t, we examined the relative likelihoods of each f_0 and the stability of the ranking across a range of model parameters.

We considered the spread during June-November (season S1) for model evaluation. Using the S1 flow network, our objective was to rank various starting configurations f_0 based on $\phi(\mathcal{O}|f_0, t)$ given $\mathcal{O}$, t. For a given σ, we evaluated the likelihood of each node being the central node. We considered two criteria based on which the likelihood of each f_0 as the starting configuration was computed: (i) $\mathcal{O}_\mathrm{G}$: this is the set of all pairs (v, I) where v is a market node that belongs to a district that reported pest presence by December 2016. (ii) $\mathcal{O}_\mathrm{B}$: this is the set of (v, I) for all nodes v, setting up the baseline which assumes no observational data.

The results are shown in Figure 4. Firstly, we observed that for both criteria $\mathcal{O}_\mathrm{G}$ and $\mathcal{O}_\mathrm{B}$, the top few ranks are relatively robust to varying network and model parameters. Also, for both criteria markets from the Central Development Region (CDR) that belong to Kathmandu and its adjacent districts are among the top ranked nodes. Interestingly, for the criterion $\mathcal{O}_\mathrm{G}$, Dhankuta (EDR), with the highest assigned production has a very low rank (Figure 4a) and a low ϕ value compared to the top market in $\mathcal{O}_\mathrm{G}$. However, for $\mathcal{O}_\mathrm{B}$, it is ranked second (Figure 4b). This clearly shows that while Dhankuta has the potential to infect a large number of areas, given what has been observed it is very unlikely that it was the source of infection. Dhankuta reported presence of the pest only towards the end of 2016 (see Figure 2c).

Spread in season S2 To study the spread from November 2016 to May 2017, we considered the dynamics on season S2 network. To set the initial conditions, we used the results of our inference study and chose Kathmandu with $\sigma = 10$ as the seed distribution. For this initial condition, we obtained the probability of infection for all nodes in S1 for T1 time steps. This distribution is used as initial condition for the S2 network spread. Figure 4c shows the infection probabilities for a particular combination of (T_1, T_2). As seen in Figure 4c, our model suggests that most of Terai and Mid Hills regions of CDR and WDR were affected by the end of May 2017, and in the subsequent seasons, the pest should affect most of the Terai and Mid Hills region. From Figure 2c, we see that regions belonging to Terai in CDR and Mid Hills of WDR and MWDR had already reported pest presence (marked in Figure 4c). The latest reports indeed corroborate our findings. With the exception of 3-4 districts, the pest is present in the entire Terai and Mid Hills regions.

While the intended usage of the origin inference formulation is to determine the source of infection, we have adapted it to compare expected spread in the model with observed data. Our results demonstrate that this framework is in general very useful in finding the likely pathways of introduction of the pest.

3.3 Establishment potential

Figure 5a shows the plot of Ecoclimatic Indices (EI). The Terai and Mid Hills show relatively favorable climatic conditions for *T. absoluta* establishment. The locations that already report pest incidence have EI ranging from 30-75. High Hills does not favor its establishment, which may be attributed to severe cold stress. Eastern parts of Nepal (districts such as Udaypur, Saptari, Sindhuli and parts of Sansari) are more favorable to *T. absoluta* establishment than the central and western parts. With a rise of 1°C in temperature, most of the Terai regions of Nepal may experience a reduction in the EI value (ranging from 15-20), though the pest can survive and establish itself in these regions (see Figure 5b). But some parts of Mid Hills may experience an increase (15-20) in EI values with a 1°C rise. However, considering the increasing adoption of protected cultivation methods (Nepal Ministry of

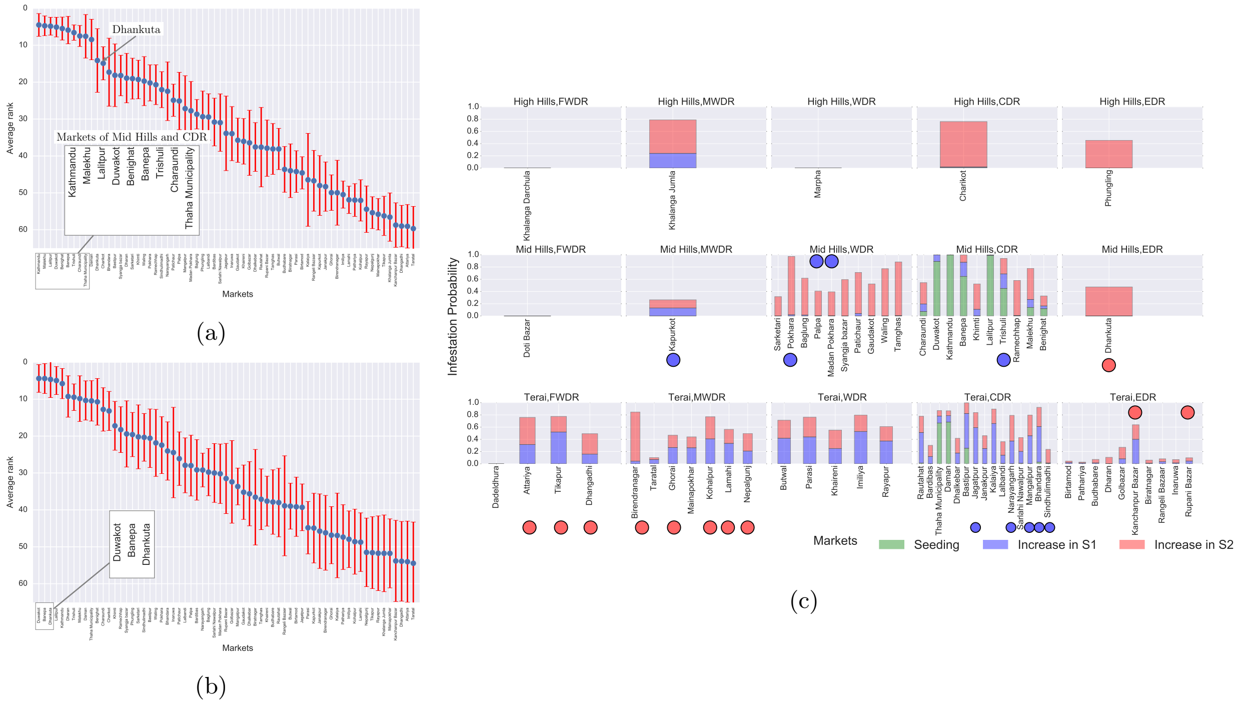


Figure 4: **Evaluating the spread model using epidemic source inference framework.** (a) The average rank of each market based on the likelihood for the criterion $\mathcal{O}_{\mathrm{G}}$ for a range of model parameters. (b) Same as (a), but for criterion $\mathcal{O}_{\mathrm{B}}$. (c) **Spread in S2.** The parameters used were $\beta = 2$, $\kappa = 500$, $\sigma = 15$, $\gamma = 1$, $T_1 = T_2 = 10$ with Kathmandu as the seed node. The blue dots correspond to markets whose districts reported *T. absoluta* presence before December 2016 (season S1), while the red dots correspond to markets which reported later. More details on sensitivity analysis are presented in (Venkatramanan et al. 2017).

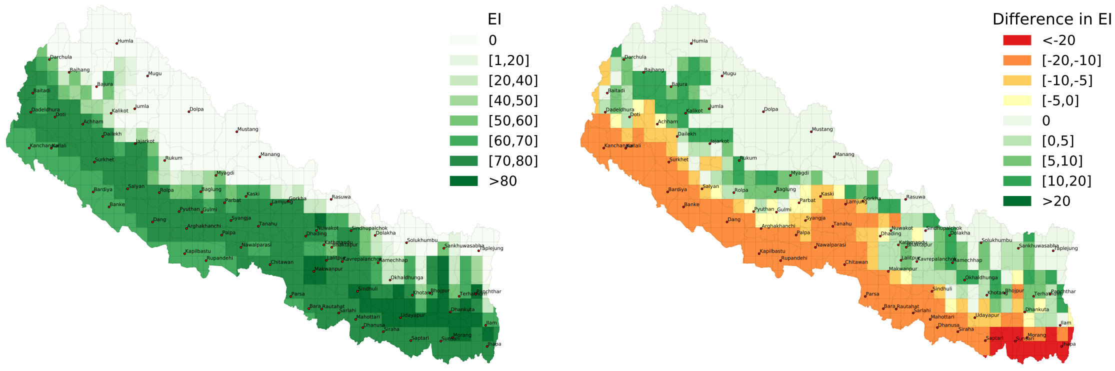
Agriculture Development 2012), this is in some sense a conservative estimate. As observed in Europe and Mediterranean regions (Karadjova et al. 2013), *T. absoluta* has been known to survive unfavorable weather conditions in greenhouses.

In particular, Kathmandu and its surrounding locations have an EI in the range of 15-45, corresponding to low to moderate level of infestation. Growth index analysis for Kathmandu region (see Figure S4 in Supplementary Information) indicates that infestation can be expected throughout the year. The weekly growth index peaks during February-March and October-November, indicating the possibility of higher incidence of *T. absoluta*; the lower weekly growth index during April-May indicates lower incidence of the pest. Another growth promoting factor for *T. absoluta* identified for Kathmandu is the moisture index which directly influences the weekly growth index of the pest.

3.4 Economic impact analysis

We evaluated the economic impact of *T. absoluta* in Nepal based on the projection of its spread from the initial infestation in Kathmandu. Since Nepal's tomato production is less than 0.2% of the global tomato production, most of which is used to meet the domestic demand for tomatoes, and it exports only 1% of its production, so we treated it as a small closed economy (http://www.fao.org/faostat). A 25% crop loss (Bajracharya et al. 2016) in cultivated areas infected by *T. absoluta* results in a direct economic impact of \$19.7M (Supplementary Information). The direct impact, however, does not account for the change in the market price of tomatoes due to loss in the crop or the impact of price change on consumers' and producers' welfare.

To calculate a more comprehensive total economic impact, we used the partial equilibrium approach (Alston et al. 1995; Soliman et al. 2012). The comprehensive economic impact analysis shows a social



(a) (b)

Figure 5: **Long-term establishment potential.** (a) Plot of the Ecoclimatic Index (EI) for current climate conditions. (b) The difference between EI's of current scenario and 1°C rise in temperature.

welfare loss of \$22.4M and a price increase of 32%. The analysis assumed 'demand elasticity to price' to be -0.7, 'supply elasticity to price' to be 0.5 (https://data.ers.usda.gov), base price at \$400 per ton (www.kalimatimarket.com.np), crop loss due to pest invasion to be 25% (Bajracharya et al. 2016), and reduced net price due to increased cost of production from pest control to be 80% (Khidr et al. 2013). The pest risk probabilities were obtained after running the spread model for 10 time steps.

As mentioned in the model evaluation, due to inaccurate and inadequate incidence reports, it is not clear what time duration each simulation step t corresponds to. So we performed a sensitivity analysis of the economic impact for two other time steps; one where the spread model is run for only 5 time steps and the other where it is run for 20 time steps. The longer run forecasts result in higher pest risk probabilities and hence higher economic impact, and vice versa. Pest risk probabilities based on 5 time steps, ceteris paribus, result in a direct economic impact of \$16M and a total economic impact of \$17.5M, whereas after 20 time steps these numbers increase to \$21.5M and \$24.7M respectively.

4 Discussion

Even though Nepal is mainly an agrarian society, its agriculture has been predominantly characterized by subsistence farming, poor marketing infrastructure, and dependency on imports from neighboring India. However, in the past decade, there has been a surge of efforts to improve this situation (Nepal Ministry of Agriculture Development 2012; Nepal Economic, Agriculture, and Trade 2011). Increasingly, farmers have been adapting protected cultivation methods such as tunnel farming (Nepal Ministry of Agriculture Development 2012) to increase yield, and reverse the trend of importing. Also, given its unique geography, Nepal is among the most susceptible regions to climate change (Kindlmann 2011, Chapter 2). Under such conditions, major invasion events such as *T. absoluta* are detrimental to its biodiversity, economy, and societal well-being in general. Therefore, there is a desperate need for concerted efforts to understand the complex food system of this country and increase its reactive capacity and resilience to attacks from invasive species. We believe that our work has taken the first steps in this direction.

Although there is general consensus that vegetable and seedling trade is a primary driver of *T. absoluta* spread (Desneux et al. 2010; Karadjova et al. 2013; USDA 2011; Campos et al. 2017), previous modeling efforts have only focused on establishment potential (Desneux et al. 2010; Tonnang et al. 2015) and spatial dispersion (Guimapi et al. 2016). This is the first work that analyzes human-mediated pathways in the context of *T. absoluta*. Some methodological aspects pertaining to this work appear in Venkatramanan et al. (2017).

Related work. In recent years, there has been a lot of interest in studying the role of international trade and travel in invasive species spread. Nopsa et al. (2015) studied the structure of a stored grain network induced by storage facilities and an underlying rail network. Carrasco et al. (2010) use phenology models along with gravity models for human-assisted dispersal. Nopsa et al. (2015) evaluated the structure of rail networks in the US and Australia for pest and mycotoxin dispersal. Colunga-Garcia et al. (2009) used the regional freight transport information to characterize risk of urban and periurban areas to exotic forest insect pests in the US. Robinet et al. (2009, 2016) developed a long-distance dispersal model to study the spread of pests that accounts for heterogeneous human population densities in the study region. International (Early et al. 2016; Ercsey-Ravasz et al. 2012; Kaluza et al. 2010) and domestic (Magarey et al. 2011; Colunga-Garcia and Haack 2015) trade datasets have been analyzed to assess the susceptibility of countries to invasive alien species and contaminants. Tatem (2009) showed that the world-wide airline network increases the risks of establishment by providing busy transport links between spatially distant, but climatically similar regions of the world.

Challenges. Agro-trade networks are a complex system problem. The networks depend on varied factors, including seasonal production, population distribution, cultural factors, economic activity, storage, and transport infrastructure. Furthermore, data needed to develop agro-trade networks is often sparse, noisy, misaligned in reporting time and is not openly available. Apart from quantitative datasets, there is also need for qualitative information pertaining to the study region such as cultural practices, seasonal production cycles, etc. Interpreting this data and integrating it into model design requires local knowledge. Challenges exist in obtaining high-resolution pest distribution data. Monitoring is a resource intensive task. Also, the pest might not have been detected during the off season due to host unavailability. Its reporting might be delayed by farmers due to lack of awareness or fear of quarantining. Given these constraints, there might be several months of delay in reporting pest presence. Another challenge is validating the network representations. While international trade data is available at the commodity level, domestic data is hard to come by. Even in data-rich regions such as the US, the available sample data (e.g. Freight Analysis Framework[1]) is aggregated at the commodity level. The role of these networks in the study of invasive species requires one to understand the ecological contagion processes.

Limitations and possible improvements. Since our study is one of the first to consider regional commodity flow analysis in the context of pest spread, especially *T. absoluta*, there are several avenues for improvement. While some of the limitations arise from lack of refined data, others are due to the limited understanding of the underlying complexity of pest invasions. The former may be the norm for emerging contagions in a data-poor region, whereas the latter will need several iterations of model development and validation by the scientific community. Our model predominantly focuses on commodity flow, and does not explicitly account for natural or other modes of spread (infected seedlings from nurseries for example). A more comprehensive model will need to integrate ecological suitability and biology directly in the diffusion process. Further studies will be needed to understand the pest's flying capacity, influence of wind direction, alternate hosts, etc. to realistically capture spread dynamics. Transportation of infected seedlings from nurseries to production sites and dispersal through packaging material are other possible causes of long-distance spread, thus hinting at multiple pathways that need to be accounted for within the umbrella term of "commodity flow". While gravity models have been applied to study a diverse set of phenomena that concern interaction between entities (Erlander and Stewart 1990; Kaluza et al. 2010; Bossenbroek et al. 2001; Thiemann et al. 2010; Jongejans et al. 2015; Krings et al. 2009) they do have known shortcomings (Simini et al. 2012; Rothlisberger and Lodge 2011). Further, the model could be refined by accounting for price variations and commodity varieties. The assignment of demand and supply attributes was done assuming homogeneous distribution of district level production and population information. This can be improved with (i) population and production estimates at a higher spatial resolution, (ii) ground survey data on each market's scope. The same holds for temporal resolution in assigning seasonal production.

We have described a first-principles based commodity modeling framework that integrates easily available datasets on population, production, etc. to model the flow of agricultural produce. We have demonstrated the validity of the constructed networks, and have used it to understand the impact of commodity flow on pest spread. Despite being limited by the availability of quality validation datasets, a

[1] https://ops.fhwa.dot.gov/freight/freight_analysis/faf/

bare bones framework such as ours can be quickly extended to other vegetables, pests and regions with minimal effort.

Acknowledgments

This work was suppoted 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. G.N. was also partly supported by Virginia Agricultural Experiment Station project VA-136324.

References

Ali, M. (2000). *Dynamics of vegetable production, distribution and consumption in Asia.* Shanhua (Taiwan) AVRDC.

Alston, J. M., Norton, G. W., and Pardey, P. G. (1995). Science under scarcity. *CAB International. Wallingford, Oxon, UK.*

Anderson, J. E. (2011). The gravity model. *Annual Review of Economics*, 3(1):133–160.

Bajracharya, A. S. R., Mainali, R. P., Bhat, B., Bista, S., Shashank, P., and Meshram, N. (2016). The first record of South American tomato leaf miner, Tuta absoluta (Meyrick 1917)(Lepidoptera: Gelechiidae) in Nepal. *J. Entomol. Zool. Stud*, 4:1359–1363.

Banks, N. C., Paini, D. R., Bayliss, K. L., and Hodda, M. (2015). The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecology letters*, 18(2):188–199.

Biondi, A., Guedes, R. N. C., Wan, F.-H., and Desneux, N. (2018). 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.

Bossenbroek, J. M., Kraft, C. E., and Nekola, J. C. (2001). Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. *Ecological Applications*, 11(6):1778–1788.

Campos, M. R., Biondi, A., Adiga, A., Guedes, R. N., and Desneux, N. (2017). From the Western Palaearctic region to beyond: *Tuta absoluta* ten years after invading Europe. *Journal of Pest Science.*

Carrasco, L., Mumford, J., MacLeod, A., Harwood, T., Grabenweger, G., Leach, A., Knight, J., and Baker, R. (2010). Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. *Ecological Modelling*, 221(17):2068–2075.

Colunga-Garcia, M. and Haack, R. A. (2015). Following the transportation trail to anticipate human-mediated invasions in terrestrial ecosystems. *Pest Risk Modelling and Mapping for Invasive Alien Species. CAB International, Wallingford, UK*, pages 35–48.

Colunga-Garcia, M., Haack, R. A., and Adelaja, A. O. (2009). Freight transportation and the potential for invasions of exotic insects in urban and periurban forests of the united states. *Journal of Economic Entomology*, 102(1):237–246.

Cunniffe, N. J., Koskella, B., Metcalf, C. J. E., Parnell, S., Gottwald, T. R., and Gilligan, C. A. (2015). Thirteen challenges in modelling plant diseases. *Epidemics*, 10:6–10.

Desneux, N., Wajnberg, E., Wyckhuys, K. A., Burgio, G., Arpaia, S., Narváez-Vasquez, C. A., González-Cabrera, J., Ruescas, D. C., Tabone, E., Frandon, J., et al. (2010). 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.

Early, R., Bradley, B. A., Dukes, J. S., Lawler, J. J., Olden, J. D., Blumenthal, D. M., Gonzalez, P., Grosholz, E. D., Ibañez, I., Miller, L. P., Sorte, C. J. B., and Tatem, A. J. (2016). Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications*, 7.

Ercsey-Ravasz, M., Toroczkai, Z., Lakner, Z., and Baranyi, J. (2012). Complexity of the international agro-food trade network and its impact on food safety. *PloS one*, 7(5):e37810.

Erlander, S. and Stewart, N. (1990). *The Gravity Model in Transportation Analysis: Theory and Extensions.* VSP, Utrecht.

Gontijo, P., Picanço, M., Pereira, E., Martins, J., Chediak, M., and Guedes, R. (2013). Spatial and temporal variation in the control failure likelihood of the tomato leaf miner, *Tuta absoluta*. *Annals of Applied Biology*, 162(1):50–59.

Grousset, F., Suffert, M., and Petter, F. (2015). EPPO Study on pest risks associated with the import of tomato fruit. *EPPO Bulletin*, 45(1):153–156.

Guedes, R. and Siqueira, H. (2013). The tomato borer *Tuta absoluta*: insecticide resistance and control failure. *Plant Sciences Reviews 2012*, page 245.

Guimapi, R. Y., Mohamed, S. A., Okeyo, G. O., Ndjomatchoua, F. T., Ekesi, S., and Tonnang, H. E. (2016). Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecological Complexity*, 28:77–93.

Haynes, K. E., Fotheringham, A. S., et al. (1984). *Gravity and spatial interaction models*, volume 2. Sage Beverly Hills, CA.

Hulme, P. E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1):10–18.

Jongejans, E., Skarpaas, O., Ferrari, M. J., Long, E. S., Dauer, J. T., Schwarz, C. M., Rauschert, E. S., Jabbour, R., Mortensen, D. A., Isard, S. A., et al. (2015). A unifying gravity framework for dispersal. *Theoretical Ecology*, 8(2):207–223.

Kaluza, P., Kölzsch, A., Gastner, M. T., and Blasius, B. (2010). The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48):1093–1103.

Karadjova, O., Ilieva, Z., Krumov, V., Petrova, E., Ventsislavov, V., et al. (2013). *Tuta absoluta* (Meyrick)(Lepidoptera: Gelechiidae): Potential for entry, establishment and spread in Bulgaria. *Bulgarian Journal of Agricultural Science*, 19(3):563–571.

Khidr, A., Gaffar, S., Maha, S., Nada, A., Taman, A., Fathia, A., and Salem, A. (2013). New approaches for controlling tomato leafminer, *Tuta absoluta* (Meyrick) in tomato fields in Egypt. *Egyptian Journal of Agricultural Research*, 91(1):335–348.

Kindlmann, P. (2011). *Himalayan biodiversity in the changing world.* Springer Science & Business Media.

Krings, G., Calabrese, F., Ratti, C., and Blondel, V. D. (2009). Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07):L07003.

Lokhov, A. Y., Mézard, M., Ohta, H., and Zdeborová, L. (2014). Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E*, 90(1):012801.

Magarey, R. D., Borchert, D., Engle, J., Colunga-Garcia, M., Koch, F. H., and Yemshanov, D. (2011). Risk maps for targeting exotic plant pest detection programs in the United States. *EPPO Bulletin*, 41(1):46–56.

Marahatta, S., Dangol, B. S., and Gurung, G. B. (2009). *Temporal and Spatial Variability of Climate Change Over Nepal, 1976-2005.* Practical Action Nepal Office.

Nepal Economic, Agriculture, and Trade (2011). Value Chain/Market Analysis of the vegetable Sub-Sector in Nepal. Technical report, USAID/Nepal.

Nepal Ministry of Agriculture Development (2012). Value chain development plan for tomato. http://pact.gov.np/docs/publication/Value%20Chain%20Dev%20for%20Tomato%20book.pdf.

Nopsa, J. F. H., Daglish, G. J., Hagstrum, D. W., Leslie, J. F., Phillips, T. W., Scoglio, C., Thomas-Sharma, S., Walter, G. H., and Garrett, K. A. (2015). Ecological networks in stored grain: Key postharvest nodes for emerging pests, pathogens, and mycotoxins. *BioScience*, page biv122.

NPPO (2009). *Tuta absoluta* Povolny (Gelechiidae) - tomato leaf miner - in tomato packaging facility in The Netherlands, National Plant Protection Organization (NPPO), Wageningen.

Oztemiz, S. (2014). *Tuta absoluta* povolny (Lepidoptera: Gelechiidae), the exotic pest in Turkey. *Romanian Journal of Biology*.

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925.

Robinet, C., Kehlenbeck, H., Kriticos, D. J., Baker, R. H., Battisti, A., Brunel, S., Dupin, M., Eyre, D., Faccoli, M., Ilieva, Z., et al. (2012). A suite of models to support the quantitative assessment of spread in pest risk analysis. *PLoS One*, 7(10):e43366.

Robinet, C., Roques, A., Pan, H., Fang, G., Ye, J., Zhang, Y., and Sun, J. (2009). Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One*, 4(2):e4646.

Robinet, C., Suppo, C., and Darrouzet, E. (2016). Rapid spread of the invasive yellow-legged hornet in France: the role of human-mediated dispersal and the effects of control measures. *Journal of Applied Ecology*.

Rothlisberger, J. D. and Lodge, D. M. (2011). Limitations of gravity models in predicting the spread of Eurasian watermilfoil. *Conservation Biology*, 25(1):64–72.

Shah, D. and Zaman, T. (2011). Rumors in a network: Who's the culprit? *IEEE Transactions on information theory*, 57(8):5163–5181.

Simini, F., González, M. C., Maritan, A., and Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100.

Soliman, T., Mourits, M. C., Van Der Werf, W., Hengeveld, G. M., Robinet, C., and Lansink, A. G. O. (2012). Framework for modelling economic impacts of invasive species, applied to pine wood nematode in Europe. *PLoS One*, 7(9):e45505.

Sutherst, R. W. (2000). Climate change and invasive species: a conceptual framework. *Invasive species in a changing world*, pages 211–240.

Sutrave, S., Scoglio, C., Isard, S. A., Hutchinson, J. S., and Garrett, K. A. (2012). Identifying highly connected counties compensates for resource limitations when evaluating national spread of an invasive pathogen. *PLoS One*, 7(6):e37793.

Tatem, A. J. (2009). The worldwide airline network and the dispersal of exotic species: 2007–2010. *Ecography*, 32(1):94–102.

Thiemann, C., Theis, F., Grady, D., Brune, R., and Brockmann, D. (2010). The structure of borders in a small world. *PloS one*, 5(11):e15422.

Tonnang, H. E., Mohamed, S. F., Khamis, F., and Ekesi, S. (2015). 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.

USDA (2011). New Pest Response Guidelines: Tomato Leafminer (*Tuta absoluta*). *Animal and Plant Health Inspection Service, Plant Protection and Quarantine*.

Vandernoot, J. and Van Hove, C. (2014). Disparities between development regions and district development committees in Nepal. *International Advances in Economic Research*, 20(3):353–355.

Venkatramanan, S., Wu, S., Shi, B., Marathe, A., Marathe, M., Eubank, S., Sah, L. P., Giri, A., Colavito, L. A., Nitin, K., et al. (2017). 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.