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

S. Venkatramanan[1], S. Wu[1,2], B. Shi[3], A. Marathe[1,4], M. Marathe[1,2], S. Eubank[1,5,6], 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[4], A. Adiga[1]

[1]Biocomplexity Institute of Virginia Tech,

[2]Department of Computer Science, Virginia Tech,

[3]Department of Economics, Virginia Tech,

[4]Department of Agricultural and Applied Economics, Virginia Tech,

[5]Department of Population Health Sciences, Virginia Tech,

[6]Department of Physics, 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. Preliminary analyses indicate that *T. absoluta* will invade most major tomato production regions within a year of introduction and the economic impact of invasion could range from $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 [54], it is critical to understand human-mediated pathways to prevent introduction and mitigate immediate impact [8, 26]. Increasingly,

the need for a comprehensive outlook to tackle this problem is being highlighted [12, 15, 45]. Yet studies that analyze trade and travel in the context of food safety are only beginning to emerge [14, 17, 18, 40, 56].

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 [35]. 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 [25]. Further 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 [40, 55]. Network dynamical processes such as the SEIR (Susceptible → Exposed → Infected → Recovered) model from epidemiology [43] 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 [17]. 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 [31] and largely an agrarian economy, which recently reported *T. absoluta* invasion in 2016 [7].

Indigenous to South America, *Tuta absoluta* or the South American tomato leaf miner was accidentally introduced to Spain in 2006 [16]. Since then, it has rapidly spread throughout Europe, Africa, Western Asia, Indian subcontinent and parts of Central America [11] over the past decade. Since tomato is among the top two traded vegetables (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 [41]. Karadjova et al. [29] observed that the spread pattern in Bulgaria was correlated with prime trade routes. With tomato being a commercially important crop [22], *T. absoluta* has had significant global impact. For example, in Turkey, the annual estimated intervention cost is €167M per year [42]. Due to extensive insecticide treatment in Europe [20], insecticidal resistance have been recently observed in populations of *T. absoluta* [23]. Lack of effective natural predators has made integrated pest management quite challenging.

In Nepal, after its introduction in 2016 in Kathmandu district [7], it has continued to spread to several districts (Figure 2c). Tomato is among the major cash crops that Nepalese farmers depend on for their livelihood [38]. Therefore, stresses such as the *T. absoluta* invasion can have a huge impact on Nepal's agrarian economy and fragile ecosystem.

Summary of contributions. From a methodological perspective, our main contributions are as follows:

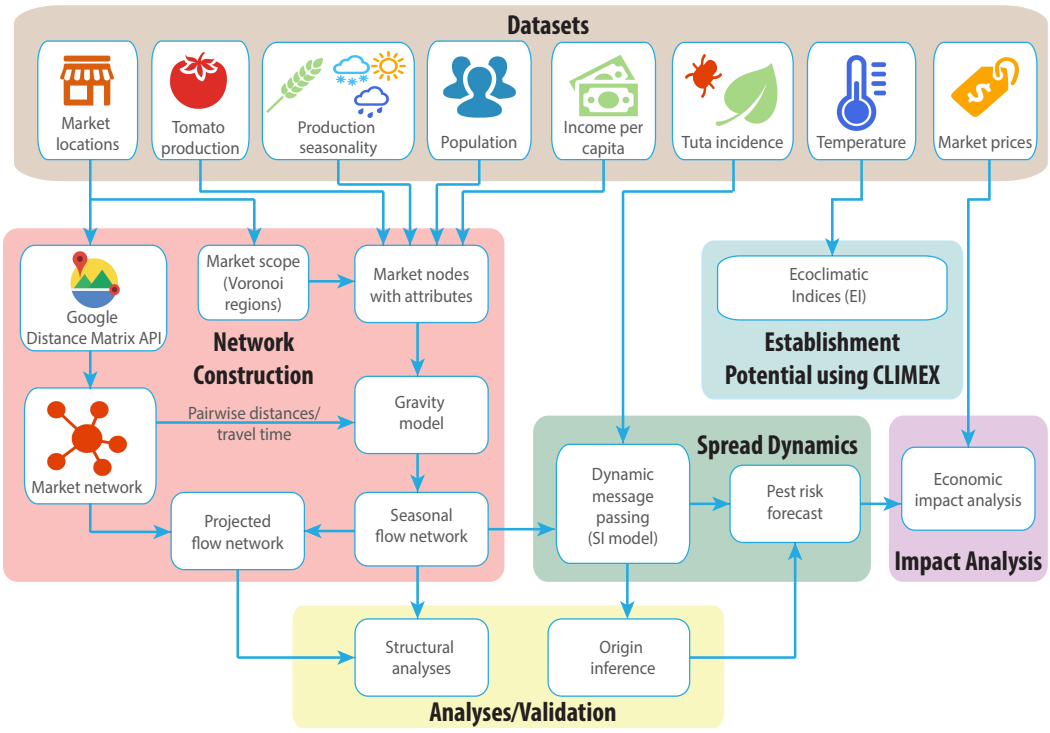


Figure 1: **Modeling framework.** The different modules and variety of data used are outlined. We show in detail the process of constructing the seasonal flow networks and modeling the spread over these networks. This is summarized in Methods and described in detail in Supplementary Information.

1. 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. We also develop methods to identify important markets in the context of pest spread.

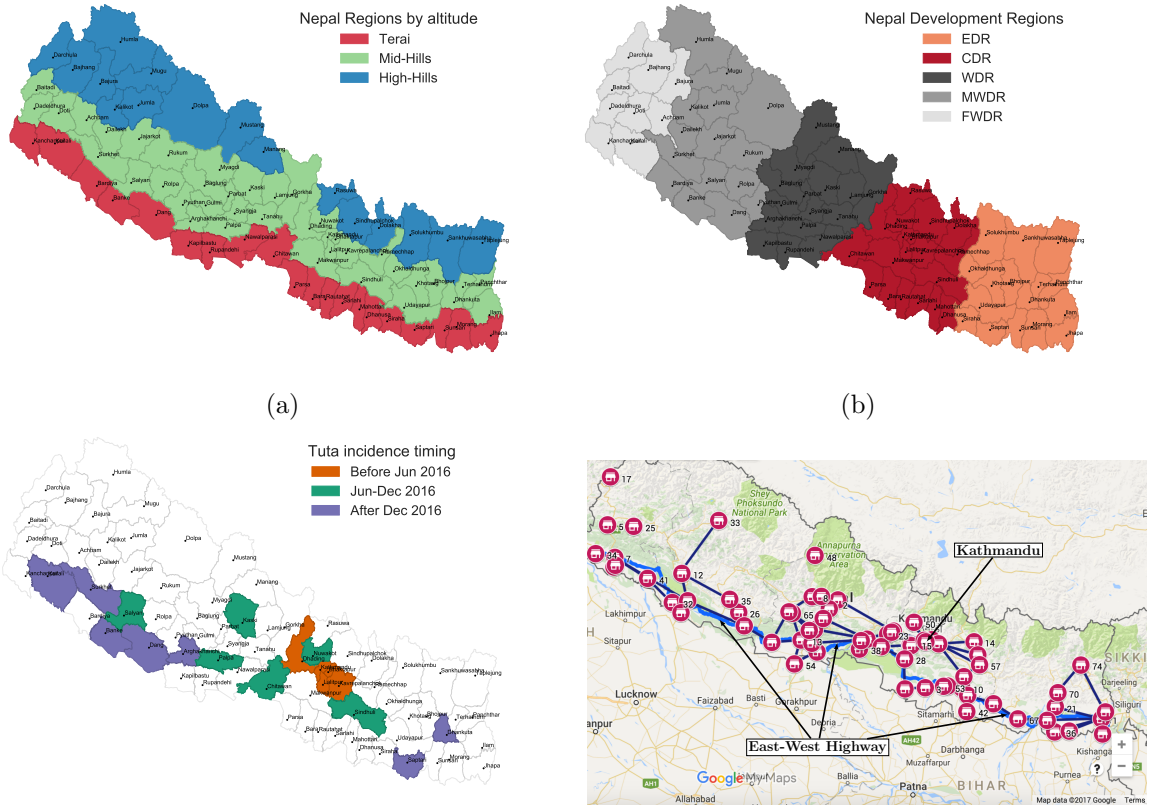
2. Under the hypothesis that trade is the primary driver of spread, pest dispersal is modeled as a stochastic dynamical process over the seasonal flow networks. Applying techniques from an emerging field in network science, namely *epidemic origin inference*, we establish a strong correlation between the expected spread over the flow network and observational data. This is the first application of the method in the context of invasive species spread.

From the perspective of *T. absoluta* invasion of Nepal, our main analytical results are as follows:

1. In general, 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.

2. We predict that, in the absence of effective control measures, *T. absoluta* will spread to all major tomato producing regions of Nepal within a year.

3. Our assessment of suitability for *T. absoluta* establishment based on growth promoting/stress factors and observational data is that except for the mountainous regions in the north,



(a) (b) (c) o (d)

Figure 2: **Nepal: geography, *T. absoluta* incidence and market infrastructure.** (a) Physiographic division of Nepal. Within a distance of 180km from south to north, the altitudes range from 67 meters above sea level (masl) corresponding to tropical Terai region bordering India to the mountain region exceeding 3000 masl [3]. (b) Division based on development regions from east to west. There is significant variation in the revenue generated by and expenditures dedicated to these regions (see Methods). (c) District-level report of *T. absoluta* incidence. Details on *T. absoluta* monitoring efforts are in the Supplementary Information. (d) Major domestic markets and their adjacency network imposed by Nepal's road network.

T. absoluta can survive year-round in this region. However, the highest damage potential is in the south-eastern parts. We also analyze the shift in ecoclimatic suitability that could result due to climate change.

4. Our assessment of the economic impact based on partial equilibrium approach predicts a social welfare loss of $17.5-24.7 million due to this invasion.

2 Results

This section is organized as follows. First, we analyze the possible ways in which *T. absoluta* could have been introduced into Nepal. In doing so, we also set up the main assumptions underlying our models, experiment design and evaluation. Next, we describe our methodological contributions: flow network construction, modeling the pest spread over these networks and the evaluation based on origin inference framework. Next, we present analytical results based on these models: structural properties of the flow networks, source inference results, forecasts of current and future spread, and sensitivity analyses. Finally, we discuss the long-term establishment potential and present the economic impact analysis. The symbols and abbreviations used henceforth are summarized in Table 1.

Table 1: **Notation and abbreviations.**

Symbol/Abbrev.	Description
S1	Season 1: June to November
S2	Season 1: December to May
β	Power-law exponent of gravity model
κ	Cutoff time of gravity model
γ	Per capita income parameter
σ	Gaussian kernel parameter for spatial seeding
t	Time step for the spread model
EDR	Eastern Development Region
CDR	Central Development Region
WDR	Western Development Region
MWDR	Mid Western Development Region
FWDR	Far Western Development Region

Possible pathways of *T. absoluta* invasion of Nepal. Nepal extensively imports vegetables from India [59] where the pest was reported in 2014 [53]. Even though the import quantity is relatively small when compared to domestic production, the inflow into specific markets is significant (see Methods for details). There are several reasons to believe that the pest entered through vegetable imports rather than through natural dispersion. *T. absoluta* was first discovered near the Kathmandu region, which is not only far from the Indian border, but also imports tomatoes all round the year from this neighbor (www.kalimatimarket.com.np) [59]. Even though Nepal shares a porous border with India, the Terai as well as adjacent regions of India did not report pest incidence (see Figure 2c).

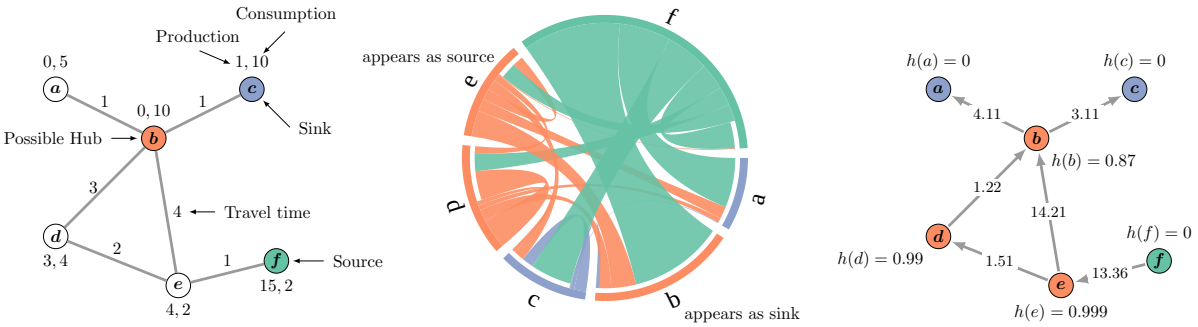
Therefore, our model design and analysis are centered around the following hypotheses: (*i*) *T. absoluta* was introduced in the region around Kathmandu, and (*ii*) domestic trade has been a primary driver of its spread within Nepal.

Flow network construction. We model the flow of agricultural produce among markets based on the following premise: The total outflow from a market depends on the amount of produce in its surrounding regions, and the total inflow is a function of the population it caters to and the corresponding per capita income. The main assumptions in this model are: (*i*) imports and exports are not significant enough to influence domestic trade, (*ii*) fresh tomatoes are mainly traded for consumption, and (*iii*) the higher the per capita income, the greater the consumption. As discussed in the Methods section, these are fair assumptions in the case of Nepal. The flows are estimated using a doubly constrained gravity model [6, 28].

Based on the altitude-induced agricultural cycle, we have divided the entire year into two seasons: S1 (June to November) and S2 (December to May). During season S1, mainly the Mid Hills and High Hills (see Figure 2a for definitions) contribute to the production, while in S2, Terai region is the major producer. As a result we have two flow networks, one for each season. The total outflow from each market i, O_i is the amount of produce that arrives at the market in the specified season. The total inflow I_i is proportional to the population catered to by the market and a function of its average per capita income η_i, η_i^γ, where γ is a tunable parameter. 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, d_{ij} is the distance to travel 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 [28] (see Supplementary Information). The methodology for assigning production and consumption to nodes and computing travel times is outlined in Figure 1 and described in Methods. Figure 3 shows a simple example of a flow network given production and consumption at each node and the road network weighted with travel times. Validating the derived flows is difficult since sample flow data is absent. However, the flows are analyzed with reference to data from a single market in Kathmandu (see Methods and Figure 8 for details).

Modeling the pest spread. We use a discrete-time SI (Susceptible-Infected) epidemic model on directed weighted networks [43] to model pest dispersal. Each node is either susceptible (free from pest) or infected (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: $\lambda_{i,j} = \frac{F_{i,j}}{\max_{i,j} F_{i,j}}$. 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*).

$P_I(i, t, f_0)$ denotes 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. In general, computing P_I is hard. Efficient methods have been proposed to estimate this probability [5, 34]. Here, we adopt the *dynamic message passing algorithm* by Lokhov et al. [34], a generalization of the Belief Propagation algorithm [44] (described in Supplementary Information).



(a) (b) (c)

Figure 3: **An example market network and corresponding flow network and projected flow network.** (a) Node attributes and transportation network. Production, consumption and the travel time between the nodes in the underlying transportation network. (b) Flow network (FN). The end-to-end flux for each pair of markets, referred to as the flow network. (c) Projected flow network (PFN). The network obtained from FN by projecting the end-to-end flows onto the road network (see Methods for definition). The flow network and the projected flow network represent in some sense, two extreme scenarios. While the flow network represents direct end-to-end flow between two markets, the projected flow network represents the scenario where this flow enters every market that lies in the path connecting these two markets. The score h(·) used to separate sources and sinks from hubs is defined in Methods. Since a flow network captures end-to-end flows, it is insufficient to identify potential sources, sinks or hubs, motivating the use of the projected flow network for this purpose. For example, b and e appear as sink and source respectively in FN, but have high in and out flows (high h(·) value) in the PFN. More details on ascertaining the roles of market nodes is described in Methods.

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 distance d from the seed as measured on the road network, 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.

Model evaluation through origin inference framework. Whether to establish goodness of fit of the model given the observed spread, or to forecast, we are faced with the challenge of translating t of the SI model to a real-world equivalent time period. Estimating this requires spatio-temporal distribution of the pest at a high resolution. Despite considerable monitoring efforts in Nepal (Supplementary Information), the incidence reports have poor time information. Monitoring is resource intensive: the placement of traps is largely determined by accessibility and availability of trained personnel. Given these constraints, the pest might remain undetected for a long time. Secondly, in the absence of monitoring, the pest's presence will become apparent only during the growing season. Therefore, there is a possible delay of several months before it is reported and confirmed. To circumvent this problem, we make use of SI model's monotonic

property: For any $t' > t$, $P_I(i, t', f_0) \geq P_I(i, t, f_0)$. Rather than attempt to match the time step to observational data, and thus estimate the absolute infection probabilities, we analyze the spread using relative vulnerabilities of the nodes. For this analysis to be meaningful, it is important that the rank list changes slowly with t with other parameters fixed. We establish that this is indeed true through rigorous sensitivity analysis.

The incidence reports in Nepal and ground surveillance strongly suggest that *T. absoluta* was first introduced to the area around Kathmandu [7]. By December 2016, several areas had reported its presence (Figure 2c). With this information, we evaluate our model based on the following backward inference problem: given an observation of node states at time t, what is the most likely origin of invasion? This is precisely the source detection problem, which has been addressed extensively in recent years [5, 50]. 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 5). Suppose $\mathcal{O}$ is the observation criteria; it consists of pairs (v, s) where v is a node and s is a state. For each candidate initial condition f_0, ideally, we would like to compute the joint probability of $\mathcal{O}$ at a time step t, $P(\mathcal{O} \mid f_0, t)$. However, there are two issues. Firstly, it is not tractable to compute this probability. Following Lokhov et al. [34], we approximate this joint probability 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

$$\phi(\mathcal{O} \mid f_0, t) = -\log\left(\prod_{(i,I)\in\mathcal{O}} P_I(i, t, f_0) \prod_{(i,S)\in\mathcal{O}} (1 - P_I(i, t, f_0))\right).$$

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.

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.

In the following sections, we highlight the analytical results based on our modeling framework.

Structural properties of the flow network. In what follows, we discuss the structure of the flow network corresponding to $\beta = 2$, $\kappa = 500$ and $\gamma = 1.0$. However, these observations are stable across the specified range of values of these parameters described in Table 2. The general trends of tomato trade between markets is depicted in Figure 4(a-d). We recall that our model accounts for the fact that the 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 (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 (recall Figure 2b for definitions). This can be partially explained by the variability in vegetable production across the Development Regions. The hilly regions of Central Development Region (CDR) and Eastern Development Region (EDR) are the primary vegetable producers during season S1 [59]. The other factor that contributes to this flow is the arterial East-West Highway which covers almost the entire breadth of the country (Figure 2d). In season S2, production is more evenly distributed among several markets in Terai, but consumption is concentrated in CDR. Hence, there is a strong flow from other regions into CDR.

Identifying nodes that could play a crucial role in the spread dynamics is valuable in preparing for or stifling contagions on the network [40]. We characterize markets based on their outflows, inflows and position in the transportation network. Markets with high outflow (inflow), but negligible inflow (outflow) are potential *sources* (*sinks*). Markets with high inflow and outflow are the *hubs*. Due to high connectivity, such markets are not only highly vulnerable, but can also play a critical role in spread. As explained in Figure 3, the flow network is not enough to determine roles of markets. We used additional information such as the town's population and its position in the transportation network. It is described in Methods.

The major sources, sinks and hubs for each season are shown in Figures 4e and 4f. As is evident from the figures, many sources of S1 switch to being sinks in season S2 and vice versa. This seasonal structure makes the overall network very vulnerable to attacks. For example, if Dhankuta is infested, then, the pest invades nearby sinks in season S1, and these in turn as sources during season S2 allow the pest to establish and invade major sinks in that season. This cascading effect can facilitate quick spread and sustenance of the pest across seasons. In addition, each development region has at least one prominent hub and there may be additional big hubs like Kathmandu close to major production areas. Our analysis suggests that Kathmandu is the biggest hub in both seasons. As discussed earlier, its markets [59] also receive significant imports throughout the year. In season S2, the nearby markets are sinks and therefore depend on Kathmandu. However, by the end of the season, tomatoes are produced in large amounts. Therefore, at the start of season S1, there is ample opportunity for the pest to establish. This could also explain why EDR was not invaded during the same time, despite being a major producer.

Epidemic source inference results. We consider the spread during June-November for model evaluation. Hence, the network corresponds to season S1 for a given set of parameters (β, κ, γ). Our objective was to rank various starting configurations f_0 based on $\phi(\mathcal{O} \mid f_0, t)$ given $\mathcal{O}$, t and flow network of season S1. Recall that every initial configuration f_0 corresponds to a central node and Gaussian kernel parameter σ, which results in the infection probabilities at time $t = 0$. 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. This is the baseline which assumes no observational data on infections.

The results of the origin inference experiments are shown in Figure 5. 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. This is discussed in more detail in Sensitivity Analysis. Also, for both criteria, markets from the Central Development Region (CDR) that belong to Kathmandu and its adjacent districts Lalitpur, Bhaktapur and Kavrepalanchok are among the top ranked nodes. This is in agreement with ground truth, despite the fact that none of these nodes are assigned the highest production in our model. This can be explained by the fact that Kathmandu and nearby districts combined together dominate the production. Also, they are centrally located in the network. Interestingly, for the criterion $\mathcal{O}_{\mathrm{G}}$, Dhankuta (EDR), with the highest assigned production has a very low rank (Figure 5a) and even lower ϕ value (Supplementary Information) compared to the top market in $\mathcal{O}_{\mathrm{G}}$. However, for $\mathcal{O}_{\mathrm{B}}$, it is ranked second (Figure 5b). 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). On the other hand, markets close to Kathmandu–Charaundi and Thaha Municipality, have average ranks comparable to

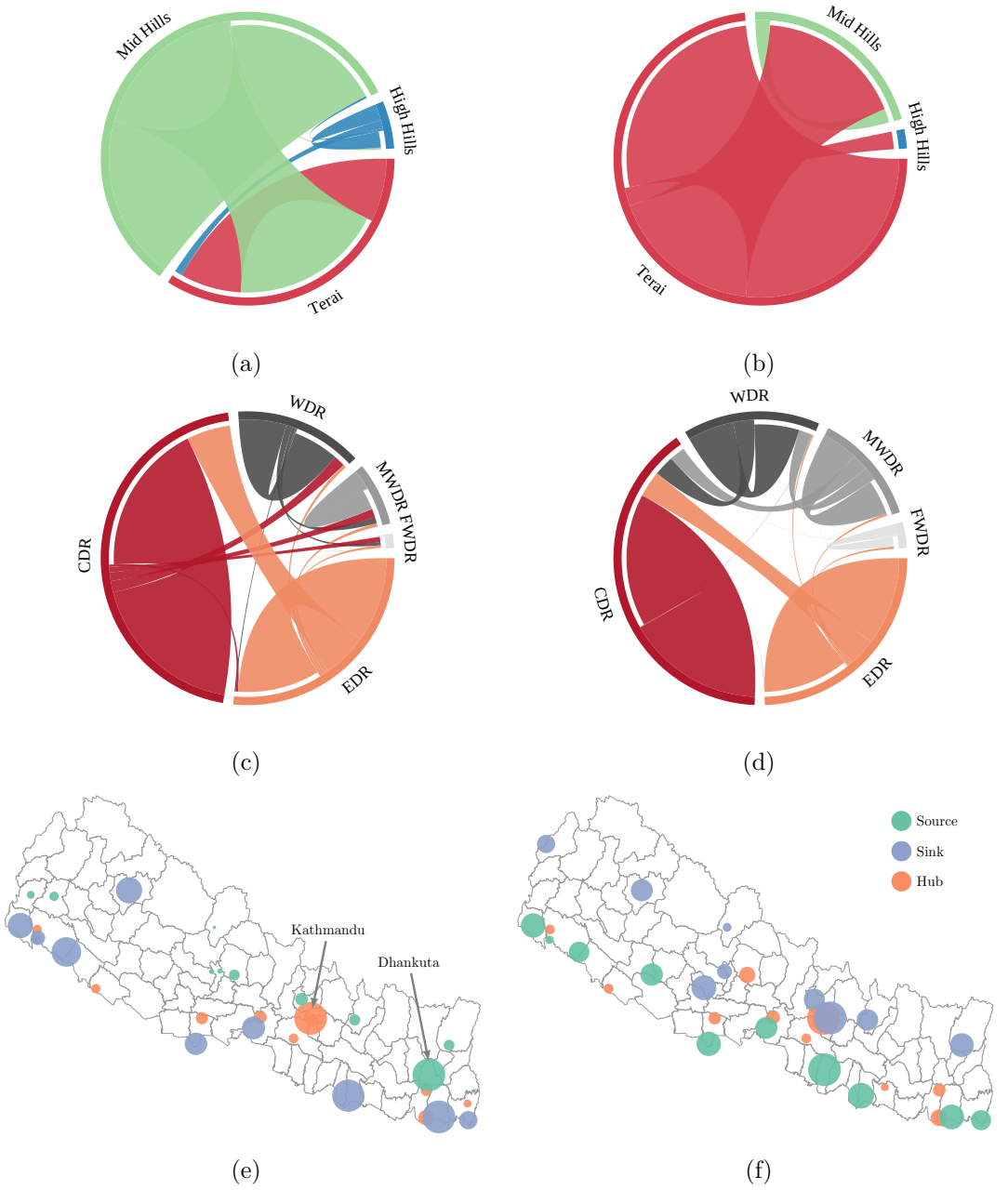
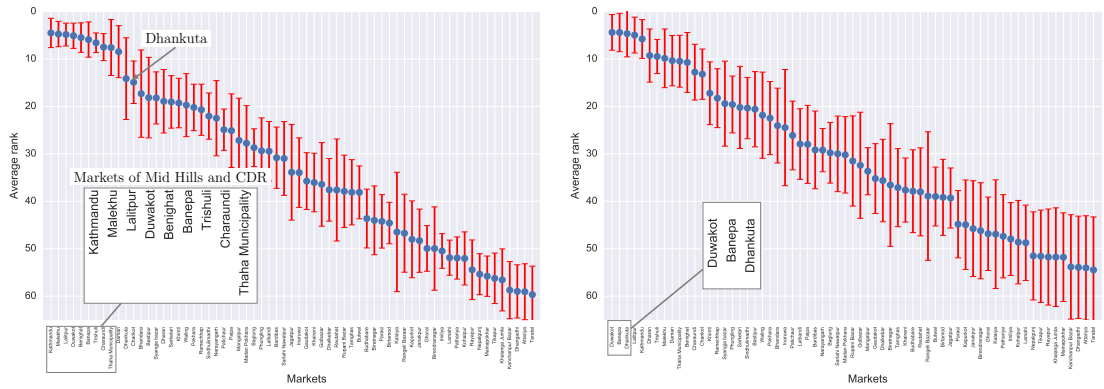


Figure 4: **Structural properties of the domestic tomato trade network.** The left part corresponds to season S1 and right to S2. (a) and (b) Net flow between physiographic regions. (c) and (d) Net flow between development regions. The flows are all relative. (e) and (f) Sources, sinks and hubs. In the plot, the sizes are all relative. The size of a source (sink) is a function of production (population), while that of a hub is a function of the product of the town's population and h(·) value.

the top markets with respect to $\mathcal{O}_{\mathrm{G}}$, but much lower average ranks with respect to $\mathcal{O}_{\mathrm{B}}$. Thus, using the principled approach of epidemic origin inference formulation, one can demonstrate the correlation between trade patterns and observational data on *T. absoluta* incidence.



(a) (b)

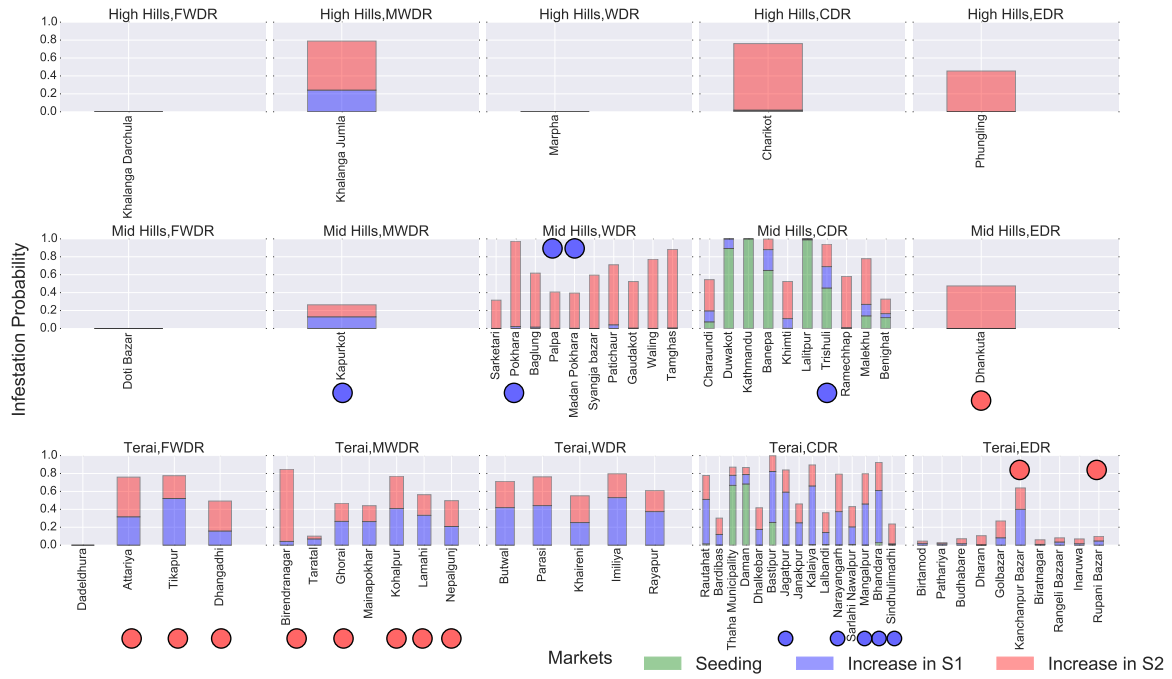
Figure 5: **Evaluating the spread model using epidemic source inference framework.** (a) The average rank of each market based on the likelihood for the criterion $\mathcal{O}_{\mathrm{G}}$ for a range of model parameters (see Table 2). (b) Same as (a), but for criterion $\mathcal{O}_{\mathrm{B}}$.

Predicting the current distribution and beyond. To study the spread from November 2016 to May 2017, we considered the dynamics in season S2. To set the initial conditions, we considered three scenarios, which differ in how we use the observational data on infected markets at the end of S1: (*i*) **Scenario A:** The infection probabilities of all markets that are reported as infected in the surveillance data at the end of November 2016 is set to 1. The rest are assigned 0. (*ii*) **Scenario B:** Based on our inference study, we 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. (*iii*) **Scenario C:** this is a combination of Scenario A and B. For every node, we take the maximum of the probability of infection in the two distributions and forecast the dynamics using season S2 network.

We find that Scenario A does not cause any secondary infections, since most of the current reported infected markets are not sources in S2. For the same reason, we do not see any distinctions between Scenario B, and the corresponding Scenario C. In particular, regions like Dhankuta do not appear as infected. This seems to suggest that *T. absoluta* has spread more than what has been reported. Figure 6 shows the infection probabilities for a particular combination (more plots are provided in the Supplementary Information). We used (5,5), (10,10) and (20,20) for (T_1, T_2) respectively representing best case, average and worst case spreads.

As seen in Figure 6, our model suggests that most Terai and Mid Hills regions of CDR, WDR would be affected by the end of May 2017, and subsequent seasons are only going to see increasing incidence of the pest throughout the country. From Figure 2c, we see that regions belonging to Terai in CDR (Chitwan) and Mid Hills of WDR (Arghakhanchi) and MWDR (Dang) have already reported pest presence (marked in Figure 6).

Establishment potential. Our assessment of *T. absoluta*'s spread to most parts of Nepal in a short period of time prompted the study of its long-term establishment potential in this country. For this purpose, a CLIMEX model [54] was used (see Methods for details). Figure 7a 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 do not favor its establishment, which may be attributed to severe cold stress. Eastern part of Nepal (districts such as Udaypur, Saptari, Sindhuli and



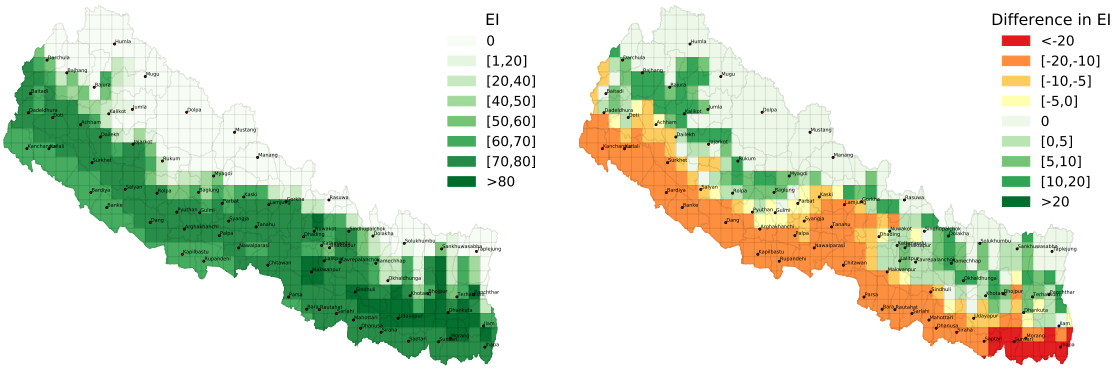
(a)

Figure 6: **Short-term forecasts of spread.** The forecasts were generated according to Scenario B described in the text. 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.

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 7b). 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 [38], this is in some sense a conservative estimate. As observed in Europe and Mediterranean regions [29], *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 Supplementary Information) indicates that infestation can be expected throughout the year. The weekly growth index peaks during February-March and October-November, indicating possibility of higher incidence of *T. absoluta* whereas 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.

Economic impact analysis. We evaluate the economic impact of *T. absoluta* in Nepal as we project its spread from the initial infestation in Kathmandu. Nepal's tomato production is less



(a) (b)

Figure 7: **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.

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 treat it as a small closed economy (http://www.fao.org/faostat). To compute the impact, we use two different measures: (*i*) the direct impact and (*ii*) the total impact, following the methodology in [52]. The direct impact measures the direct revenue loss from the tomato crop as the sum of loss encountered by each district, which in turn depends on proportional loss in the affected area, yield per unit of land in the district before being affected, tomato production area in the district, and the proportion of area affected by the pest which is informed by the spread model.

A 25% crop loss [7] 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 economic impact, we use the partial equilibrium approach [4, 52]. This method 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 substitute and complementary goods market due to changes in tomato prices. The comprehensive economic impact analysis shows a social welfare loss of $22.4M and a price increase of 32%. The analysis assumes '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% [7] and reduced net price due to increased cost of production from pest control to be 80% [30]. The pest risk probabilities are obtained after running the spread model for 10 time steps.

As mentioned in the model evaluation, the time duration of each time step t can be arbitrary, so we perform 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 comprehensive economic impact of $17.5M, whereas after 20 time steps these numbers increase to $21.5M and $24.7M respectively.

Sensitivity Analysis. We studied the sensitivity of individual market ranks as well as rank lists to network parameters (β, κ, γ), and diffusion model parameters (σ, t) (see Table 1 for definitions). The joint parameter effects are shown in Figure 5, while the single parameter results are provided in the Supplementary Information. We found that the market ranks are more sensitive to spatial seeding parameter σ and distance exponent β than other parameters. In particular, we observed that the sensitivity was highest when $\sigma = 0$ was included in the analysis. In this case (and in general for very low values of σ), substantial spread occurs only when the seed node is a source. Even if a node is in close proximity to several sources (such as Kathmandu), there is hardly any spread. This is unrealistic in the context of pest and pathogen dispersal. Hence, we restricted σ to be greater than 0 in our analysis. Also, we observe that the variance in rank is small for higher ranked nodes. This can be seen in Figure 5, and is more pronounced in the single parameter analyses. This property gives higher confidence in interpreting the results on top markets.

For rank list stability, we used Spearman's rank correlation coefficient (also known as Spearman's rho). It assesses monotonic relationships between rank lists and the value ranges between -1 and $+1$. A positive correlation coefficient indicates a positive relationship and vice versa, and two identical rank lists have a correlation of 1. Here we use the rank list that results from configuration ($\beta = 2,\ \kappa = 500,\ \gamma = 0,\ \sigma = 5,\ T = 10$) as the reference and calculate the Spearman's rho value with respect to it for rank lists induced by other parameter settings. Table 2 gives the Analysis of Variance (ANOVA) results. Under 95% confidence level, p-value < 0.05 means that the particular parameter has a significant effect. Therefore, we see that β and σ have significant effects, while others do not. Here, we note that this is despite not considering $\sigma = 0$ in the analysis.

Parameter	Description	Levels	t-ratio	F-value	p-value
β	Power-law exponent of gravity model	[0, 1, 2]	-5.16	26.6059	< 0.0001
σ	Gaussian kernel parameter for spatial seeding	[0, 5, 10, 15, 20]	-3.29	10.8424	< 0.0001
κ	Cutoff time of gravity model	[100, 500, 1000]	-0.42	0.1758	0.6753
γ	Per capita income parameter	[0, 0.5, 1.0]	0.89	0.7970	0.3727
T	Time step for the spread model	[5, 10, 20]	1.14	1.2976	0.2556

Table 2: **Analyzing sensitivity to model parameters using ANOVA.**

3 Discussion

Although there is general consensus that vegetable and seedling trade is a primary driver of *T. absoluta* spread [11, 16, 29, 60], previous modeling efforts have only focused on establishment potential [16, 58] and spatial dispersion [24]. This is the first work that analyzes human-mediated pathways in the context of *T. absoluta*. In recent years however, the role of human-mediated dispersal is being increasingly accounted for in the modeling community. Robinet et al. [46, 47] develop a long-distance dispersal model to study the spread of pests that accounts for heterogeneous human population densities in the study region. International [17, 18, 28] and domestic [14, 35] trade datasets have been analyzed to assess the susceptibility of countries to invasive alien species and contaminants. Nopsa et al. [40] study the structure of stored grain network induced by storage facilities and an underlying rail network. Nepal's vegetable production and trade has been extensively studied from a socio-economic perspective [1, 3, 59], But, to the best of our knowledge, there is no such work in the context of invasive species spread with focus on this region.

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 [38, 59]. Increasingly, farmers have been adapting protected cultivation methods such as tunnel farming [38] to increase yield, and reverse the trend of importing. Also, given its unique geography, Nepal is among the most susceptible regions to climate change [31, 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.

Since our study is one of the first to consider commodity flow analysis in the context of *T. absoluta*, we believe 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.

While gravity models have been applied to study a diverse set of phenomena that concern interaction between entities [10, 19, 27, 28, 32, 51, 57] they do have known shortcomings [49, 51]. The illustrative example (Figure 3) and the analysis of the flow network with respect to data on flows in Kalimati markets brought some of them to the fore (see Methods and Figure 8). 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 (a) population and production estimates at a higher spatial resolution, (b) ground survey data on each market's scope. The same holds for temporal resolution in assigning seasonal production. Further, a better understanding of production cycles, underlying infrastructure for commodity transport will greatly improve the flow estimates.

Our model predominantly focuses on the commodity flow, and does not explicitly account for natural spread or ecological parameters. A more comprehensive model will need to integrate ecological suitability directly in the diffusion process, and also account for spatial diffusion. Further studies will be needed to understand the pest's flying capacity, influence of wind direction, etc. to realistically capture natural diffusion. In this regard, more general SEIR models can be considered to account for time taken for pest establishment and interventions if available.

T. absoluta is also known to survive on alternative hosts such as eggplant, potato and pepper, at varying degrees of suitability [9]. 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".

Despite these limitations, having such a bare bones framework means that it can be quickly extended to other vegetables, pests and regions with minimal effort. We also believe that our approach provides a modular framework for integration of other models that can be refined with increased availability of data and sophisticated methods.

4 Methods

Data Basic demographic information (population, per capita income) were obtained at the district level (75 districts) as well as town level from the Nepal Central Bureau of Statistics (http://cbs.gov.np) [13]. Tomato production at district level, and the farm gate price (used for economic analysis) was obtained from the annual statistics released by the Ministry of Agricultural Development (MOAD), Govt. of Nepal [37]. Locations of major vegetable markets were obtained from the Online Marketing Information System maintained by the MOAD [2]. We used Google Distance Matrix API [21] to compute travel times between markets. Climate data for the CLIMEX model was obtained from CliMond [33]. *T. absoluta* surveillance reports were gathered with the support of National Agriculture Research Council, Nepal (NARC), United States Agency for International Development (USAID) Integrated Pest Management Innovation Lab, and USAID Expanding Nepalese and Bhutanese Access to Indian Technologies for Agriculture, iDE Nepal (ENBAITA).

Nepal's physiography, production, trade and consumption. Nepal, despite being a small country (800km along Himalayan axis, 150-250km across), has a 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 2a). Terai is densely populated region (with more than 400 people/sq.km. [3]), 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, as emphasized in our analyses. 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 [59]. 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 2b. The Central Development Region, which includes Kathmandu, generates 80% of the country's revenue, and 60% of government expenditure is allotted to it [61]. In comparison, the revenue generated by the Far Western Development Region was less than $\frac{1}{20}^{th}$ of that of CDR.

Network construction The nodes of the flow network are the major markets, 69 in all, after merging markets that belong to the same town. Recall that the amount of production is specified at the district level. Based on the physiography, we partitioned the districts into two groups: Mid Hills and High Hills (see Figure 2a) 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. The country's map was overlayed by a grid cell of size 25 sq.km. We constructed a Voronoi partition of these cells using node locations as centroids, motivated by the fact that tomato sellers and buyers will seek out the nearest market. We assumed uniform spatial distribution of production and population for each district. Each grid cell was assigned a value of production in that season (population) which was proportional to the fraction of the area of the district covered by the cell. The total inflow (outflow) to the market is the sum of population (production) of the grid cells assigned to it (see Figure 1). Owing to the diverse landscape of Nepal and varying road conditions we used travel time instead of distance between markets. The function $f(d_{ij})$ combines power-law and exponential decay with d_{ij} which

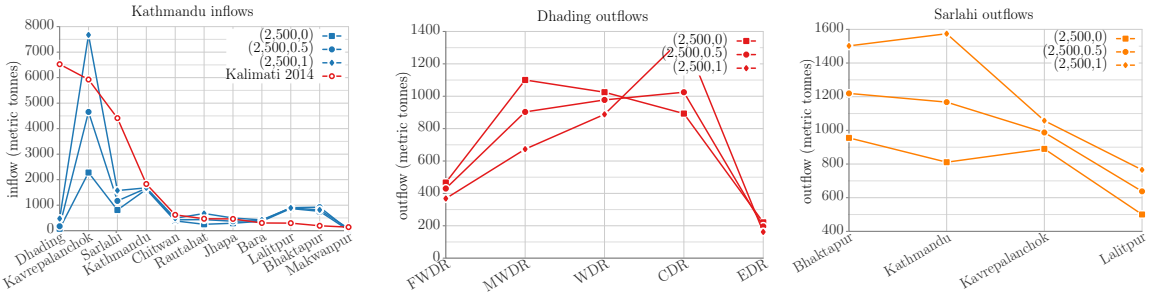
can be controlled by parameters β, the power-law exponent, and κ, the cutoff time. We explain the process of network construction in detail in the Supplementary Information.

Estimating network model parameters and validating flows. The unavailability of sample flow data makes it nearly impossible to calibrate and validate our flow network model. Therefore, we consider a wide range of values for each parameter (see Table 2). For the cutoff parameter κ even though we considered 1000 minutes, noting that the marketing infrastructure in Nepal is poorly developed [3, 39] and therefore, the produce is not transported for very long distances, for most of our analysis we restrict it to 500 (approximately 8 hours). For the exponent of the income parameter, γ, we considered the following values, $[0, 0.5, 1]$, where 0 corresponds to no effect of income and 1 to consumption directly proportional to income. For these values, the per capita annual tomato consumption for the node Kathmandu in the flow network comes to 10.87 kg, 16.57 kg and 24.54 kg respectively (for any fixed β and κ). This is the ratio of total inflow of produce aggregated across seasons to total population assigned to the node. The average consumption per person per year is reported as approximately 12 kg by Ministry of Agriculture Development [39].

Even though sample flow data is absent, the amount of inflow into one particular market is available. This is the largest wholesale market in Kathmandu, Kalimati market (kalimatimarket.com.np), for which the annual inflow of tomato from several states is documented. For a given set of network parameters (β, κ, γ), we aggregated the flow from different markets into Kathmandu at the district level and also across seasons S1 and S2. This was compared with the market data, and the analysis is presented in Figure 8.

We first note that except for inflows from two districts–Dhading and Sarlahi–the inflows are reasonably close to the reported ground truth. Upon further investigation, we find that Dhading is a major producer west of Kathmandu, and caters significant quantities to Western and Mid Western Development Regions. While the gravity model predicts that these flows will be directly delivered to WDR/MWDR, in reality, Dhading's produce reaches the central Kalimati market in Kathmandu before getting routed to rest of Nepal (reminiscent of a hub-and-spokes model of distribution) [39]. As for Sarlahi, even though there is little inflow into Kathmandu according to our model, other markets in the Kathmandu valley also receive significant flows from Sarlahi. Once again, since Kathmandu is a major hub, it is possible that the majority of this produce is routed through Kathmandu's Kalimati market. These highlight some of the limitations of a naive gravity model, which does not account for real-world trader dynamics. However, we find that, overall, when aggregated at the level of development regions, our models faithfully reproduce the regional flows.

Sources, sinks and hubs. Here, we describe the procedure used to identify important markets and their role in the commodity flow network, and in turn, the spread of the pest. Firstly, we note that high production alone is not enough to characterize a node as a source in the traditional sense: high outflow and low inflow in the flow network. We need to consider network effects. If the market is on a major highway, it is still possible that there are high inflows from other markets as illustrated in Figure 3. The same holds with sinks. Network effects alone are not enough to determine market roles either. For example, we applied the popular network measure, betweenness centrality to the road network to identify markets which are strategically placed. The results were mixed: not all documented major markets [59] were recognized (Kathmandu for example), and many nodes which have high betweenness centrality do not appear to be major markets on further inspection. Therefore, to characterize the role of each market in the trade networks, we used a different approach.



(a) (b) (c)

Figure 8: **Comparing trade network flows with available data.** (a) Inflows into Kalimati market from different districts. (b) Outflow of produce from district Dhading to various Development Regions. (c) Outflow of produce from district Sarlahi to markets in the Kathmandu valley.

Given the flow network, we define the *projected flow network* as follows: For each edge (a, b) of the road network, $w(a, b)$ is the sum of all flows F_{ij} of FN where the identified shortest path between i and j contains edge (a, b) and the flow direction is from a to b. The final flow on (a, b) is $|w(a, b) - w(b, a)|$ with direction $a \to b$ if $w(a, b) \geq w(b, a)$, or $b \to a$ otherwise. To identify sources and sinks, we define the parameter $\mathrm{h}(v) = 4\frac{I'_v O'_v}{(I'_v + O'_v)^2}$, where I'_v and O'_v are the net inflow and outflow at v in the projected flow network. A node is a perfect source or sink if $\mathrm{h}(v) = 0$. Among nodes with low $\mathrm{h}(v)$ (≤ 0.1), we chose the top nodes with highest (lowest) assigned production to population ratio as major sources (sinks). However, for a node to be a hub, a high $\mathrm{h}(v)$ is not enough. Even small markets which occupy a central position in the road network, such as on a major highway, will appear to have high inflow and outflow. Therefore, under the assumption that bigger towns have bigger markets [48], we take into consideration the population of the town to decide on its hub status. The hub score of a node v is computed as the product of the town/municipality population and $\mathrm{h}(v)$.

Sensitivity analysis We performed sensitivity analysis for both the generated flow network, and the ranking results of the origin inference problem. Specifically, a full factorial design was performed with levels for the parameters of interest as given in Table 2, and analysis of variance (ANOVA) was used to evaluate single parameter effect. It is worth noting that assessment of parameter sensitivity depends on the choice of quantity of interest (QoI). Since the outcome of origin inference is a ranking on markets, we used Spearman's rho to test its stability across the parameter space. The experiment was set up within the GENEUS framework [62]. GENEUS is a general computational environment for experimental design, uncertainty quantification (UQ) and sensitivity analysis (SA). It supports automatic model configuration, experimental design, model execution and UQ/SA related analysis. GENEUS was used to streamline the experimental pipeline in this study.

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 particular reference to Kathmandu District. The parameter fitting values used in our study for CLIMEX modeling were iteratively changed after Tonnang et al.[58]. 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 clear gradient picture in terms of suitability of the pest in different geographical localities within Nepal, wherein 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 indicate an increase in the average temperature at a rate of 0.05°C/year between 1975 to 2005[36]. Rising temperatures may have direct impact on the spread and establishment of the insect pests. Keeping this in view, impact of 1°C rise in temperature scenario was also modeled to know the near future potential areas for the establishment of *T. absoluta*. The resulting differences in EI values were mapped to know the influence of 1°C rise in temperature on EI values for different regions in Nepal for *T. absoluta* establishment.

5 Acknowledgments

This work was supported in part by the United States Agency for International Development under the Cooperative Agreement NO. AID-OAA-L-15-00001 Feed the Future Innovation Lab for Integrated Pest Management, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, NSF BIG DATA Grant IIS-1633028, NSF DIBBS Grant ACI-1443054, NIH Grant 1R01GM109718 and NSF NRT-DESE Grant DGE-154362. G.N. was also partly supported by Virginia Agricultural Experiment Station project VA-136324.

6 Author contributions

A.A., S.V., S.W., A.M., M.M., S.E. and R.M. defined the scope of the research. L.P.S., A.P.G., L.A.C. and S.V. collected and interpreted data. A.A., S.V., S.W., A.M., M.M., V.S., and K.S.N. conceived and designed the experiments. S.V., S.W., B.S., K.S.N., A.M., V.S. and A.A. performed the analysis. G.N., R.A., S.E. and R.N. provided assistance in interpreting the results. A.A., S.V. and S.W. wrote the paper with significant inputs from A.M. and V.S.. A.A. supervised the research. All authors discussed the results and commented on the manuscript.

References

[1] R. P. Adhikari, R. Collins, and X. Sun. Segmenting consumers of tomato in Nepal: implications for value chain development. In *Proceedings of the 22nd Annual IFAMA World Forum and Symposium*, pages 1–17. International Food and Agribusiness Management Association, 2012.

[2] Agribusiness Promotion and Marketing Development Directorate. Domestic markets of Nepal. Accessed in January 2017.

[3] M. Ali et al. Dynamics of vegetable production, distribution and consumption in Asia. 2000.

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

[5] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, and R. Zecchina. Bayesian inference of epidemics on networks via belief propagation. *Physical Review Letters*, 112(11):118701, 2014.

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

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

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

[9] T. Bawin, D. Dujeu, L. De Backer, F. Francis, and F. J. Verheggen. Ability of *Tuta absoluta* (Lepidoptera: Gelechiidae) to develop on alternative host plant species. *The Canadian Entomologist*, 148(04):434–442, 2016.

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

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

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

[13] Central Bureau of Statistics. Domestic markets of Nepal. Accessed in January 2017.

[14] M. Colunga-Garcia and R. A. Haack. 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, 2015.

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

[16] N. Desneux, E. Wajnberg, K. A. Wyckhuys, G. Burgio, S. Arpaia, C. A. Narváez-Vasquez, J. González-Cabrera, D. C. Ruescas, E. Tabone, J. Frandon, et al. Biological invasion of European tomato crops by *Tuta absoluta*, ecology, geographic expansion and prospects for biological control. *Journal of Pest Science*, 83(3):197–215, 2010.

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

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

[19] S. Erlander and N. F. Stewart. *The gravity model in transportation analysis: theory and extensions*, volume 3. VSP, 1990.

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

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

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

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

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

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

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

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

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

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

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

[31] P. Kindlmann. *Himalayan biodiversity in the changing world*. Springer Science & Business Media, 2011.

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

[33] D. J. Kriticos, B. L. Webber, A. Leriche, N. Ota, I. Macadam, J. Bathols, and J. K. Scott. CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, 3(1):53–64, 2012.

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

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

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

[37] Nepal Ministry of Agricultural Development. Statistical information on Nepalese agriculture 2013/14. 2014.

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

[39] Nepal Ministry of Agriculture Development. Demand and supply situation of tomato in Nepal. http://mrsmp.gov.np/files/download/tomato%20book.pdf, 2016.

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

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

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

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

[44] J. Pearl. *Reverend Bayes on inference engines: A distributed hierarchical approach*. Cognitive Systems Laboratory, School of Engineering and Applied Science, University of California, Los Angeles, 1982.

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

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

[47] C. Robinet, C. Suppo, and E. Darrouzet. 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*, 2016.

[48] D. A. Rondinelli. Cities as agricultural markets. *Geographical Review*, pages 408–420, 1987.

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

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

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

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

[53] V. Sridhar, A. Chakravarthy, and R. Asokan. New record of the invasive South American tomato leaf miner, *Tuta absoluta* (Meyrick)(Lepidoptera: Gelechiidae) in India. *Pest Management in Horticultural Ecosystems*, 20(2):148–154, 2014.

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

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

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

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

[58] H. E. Tonnang, S. F. Mohamed, F. Khamis, and S. Ekesi. Identification and risk assessment for worldwide invasion and spread of *Tuta absoluta* with a focus on Sub-Saharan Africa: implications for phytosanitary measures and management. *PloS one*, 10(8):e0135283, 2015.

[59] USAID/Nepal. Value Chain/Market Analysis of the vegetable Sub-Sector in Nepal. 2011.

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

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

[62] S. Wu, H. Mortveit, and S. Gupta. A Framework for Validation of Network-based Simulation Models: an Application to Modeling Interventions of Pandemics. In *Proceedings of ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*. ACM, 2017.