

RESEARCH

Open Access

# Optimal surveillance network design: a value of information model

Matteo Convertino<sup>1,2,3,4\*</sup>, Yang Liu<sup>1</sup> and Haejin Hwang<sup>1</sup>

\*Correspondence:

matteoc@umn.edu

<sup>1</sup>School of Public Health,  
Environmental Health Sciences  
Division and Public Health  
Informatics Program, HumNat Lab,  
University of Minnesota,  
Minneapolis, MN, USA

<sup>2</sup>Institute on the Environment,  
University of Minnesota,  
Minneapolis, MN, USA

Full list of author information is  
available at the end of the article

## Abstract

**Purpose:** Infectious diseases are the second leading cause of deaths worldwide, accounting for 15 million deaths – that is more than 25% of all deaths – each year. Food plays a crucial role, contributing to 1.5 million deaths, most of which are children, through foodborne diarrheal disease alone. Thus, the ability to timely detect outbreak pathways via high-efficiency surveillance system is essential to the physical and social well being of populations. For this purpose, a traceability model inspired by wavepattern recognition models to detect “zero-patient” areas based on outbreak spread is proposed.

**Methods:** Model effectiveness is assessed for data from the 2010 Cholera epidemic in Cameroon, the 2012 foodborne Salmonella epidemic in USA, and the 2004–2007 H5N1 avian influenza pandemic. Previous models are complemented by the introduction of an optimal selection algorithm of surveillance networks based on the Value of Information (Vol) of reporting nodes that are subnetworks of mobility networks in which people, food, and species move. The surveillance network is considered the response variable to be determined in maximizing the accuracy of outbreak source detections while minimizing detection error. Surveillance network topologies are selected by considering their integrated network resilience expressing the rewiring probability that is related to the ability to report outbreak information even in case of network destruction or missing information.

**Results:** Independently of the outbreak epidemiology, the maximization of the Vol leads to a minimum increase in accuracy of 40% compared to the random surveillance model. Such accuracy is accompanied by an average reduction of 25% in required surveillance nodes with respect to random surveillance. Accuracy in systems diagnosis increases when system syndromic signs are the most informative in a way they reveal linkages between outbreak patterns and network transmission processes.

**Conclusions:** The model developed is extremely useful for the optimization of surveillance networks to drastically reduce the burden of food-borne and other infectious diseases. The model can be the framework of a cyber-technology that governments and industries can utilize in a real-time manner to avoid catastrophic and costly health and economic outcomes. Further applications are envisioned for chronic diseases, socially communicable diseases, biodefense and other detection related problems at different scales.

**Keywords:** Value of information; Surveillance network; Transmission network; Network resilience; Topology; Uncertainty; Biosafety; Biodefense; Outbreak sources

## Background

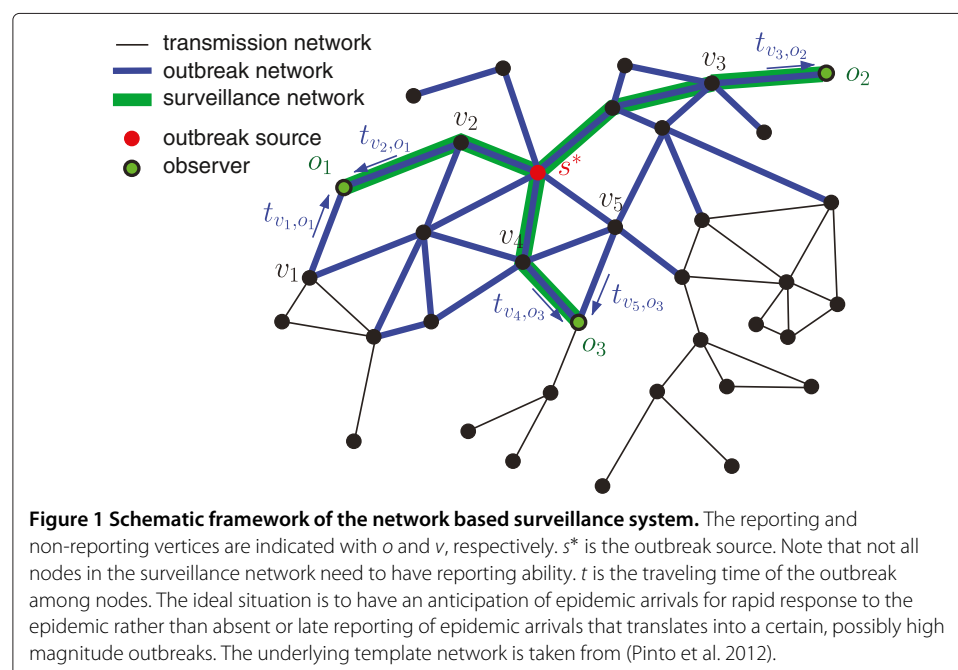
### Surveillance and uncertainty

Increased movements of people, and expansion of international trade in food and other commodities, enhanced by social and environmental changes linked to urbanization, are all manifestations of the rapidly-changing nature of the world we live in. Such changes contribute to the rapid adaptation and movement of microorganisms, which has facilitated the return of old communicable diseases, the emergence of new ones, and the evolution of antimicrobial resistance. Food, in many aspects the “connectome” of most issues related to infectious diseases and for this motivation experts are now talking about “emerging foodborne infectious diseases” whose complex dynamics embraces a larger spectrum of than the one traditionally considered for infectious diseases. The majority of the pathogens causing the significant foodborne disease burden - estimated as 1.9 million people annually at the global level - are now considered to be zoonotic (Schlundt et al. 2004). For this motivation a one-health approach is now embraced (NAS 2012). To put it more simply, food is in many cases involved in many infectious disease propagations as the vehicle of pathogen spread in the transmission chain (e.g., environment-animal-food-humans) leading to infections (e.g. for cholera and tuberculosis) beyond the classical foodborne illnesses where there is no human-human transmission. As for foodborne diseases, in USA result in 37.2 million illnesses, 228,744 hospitalizations, and 2,612 deaths each year (Batz et al. 2011; Hanson et al. 2012; NAS 2012; Scallan et al. 2011a, 2011b). Foodborne illnesses can be caused by a variety of microbial pathogens, chemicals, and parasites that contaminate food at different points in the food production and preparation process. Although most of these diarrheal deaths occur in poor countries, foodborne diseases are neither limited to developing countries nor to children. In general, infectious diseases other than foodborne are a huge problem also for creating susceptibility to chronic diseases. Because of the necessity to decrease such burden, surveillance is the primary public health strategy to emphasize, but a set of questions arise spontaneously.

What is the optimal set of surveillance nodes that best inform about the spread of an epidemic accurately? Is it possible to gauge the shape of transmission networks based on epidemic outbreaks? The accuracy in such estimations is a function of the source of information reliability. This means that a node, or a set of nodes performing surveillance, can report or not report whether a case of an infectious disease case is observed. Such information tremendously affects the estimation of outbreak path and sources. Thus, the functioning of surveillance networks is crucial for early detection and rapid response of epidemics which is possibly the only way to respond to emerging/re-emerging infectious diseases for which we are unprepared (Stachenko 2008). The function of surveillance networks is the ability to observe epidemic information (outbreak occurrence) moving along the mobility network. Therefore, the design of surveillance networks via the maximization of reporting accuracy (i.e., the accuracy in surveillance function) is a key process for saving lives of human and animal populations (Helbing et al. 2014). Economical and environmental outcomes are also associated to such optimal design: for instance industries and government costs related to commodity recalls and health care, and footprints on the environment related to pathogen and/or chemical spills can be highly reduced. To face such interconnected problems systemic risk approaches are welcome in order to design globally resilient systems (Haldane and May 2011; Helbing 2013; Helbing et al. 2014; Park et al. 2012).

For surveillance design, the schematization of surveillance as a network is very convenient and realistic (Figure 1). Nodes are sites where information about cases is observed and reported and these nodes may share information in real-time. The network scheme is related to the simplification of the system into a physical structure (the network) that is much simpler and controllable than the whole “3D domain” that is unnecessary to simulate. The mathematical problem of network search is easier than the one in which nodes are considered as separate nodes. The realism of the network scheme is related to the fact that surveillance - that is a network function - is occurring on a subset of the transmission network. Yet, surveillance and transmission networks are multiplex networks (Estrada and Gómez-Gardeñes 2014; Newman 2003) in which information and pathogens are spreading, respectively. The backbone of such networks are mobility networks of people (e.g., subway, train, airline networks, etc), food (e.g., supply chains), wild animals (e.g., migration routes), and environmental vehicle networks (e.g., watercourses and urban drainage networks) that connect portions of ecosystems and countries worldwide by contributing to the spread of infectious diseases (den Broeck et al. 2011; Goncalves et al. 2013; Knobler et al. 2006; On Effectiveness of National Biosurveillance Systems: BioWatch C, the Public Health System NRC 2011; Pastore Y Piontti et al. 2014). Thus, it is important to implement optimal surveillance on mobility networks (that can be transmission networks when a pathogen spread) via personnel and/or sensors whose reported information can be integrated into a cyber-infrastructure that supports public health authorities and industries.

The consideration of deep uncertainty in reported outbreaks is a very rarely addressed topic in epidemiology. Recently large attention has been placed on detecting outbreak sources (Convertino and Hedberg 2014a; Convertino and Hedberg 2014b; Pinto et al. 2012). However, in these studies uncertainty was neglected in reported outbreaks. Convertino and Hedberg (Convertino M, Hedberg C: Multisite outbreak detection: how many foodborne outbreak are attributable to farmer’s markets?, submitted) considered



large variability in food trade network factors to capture the deep uncertainty related to attribution of foodborne outbreaks to different local food establishments (versus food imported globally), with particular focus on farmers' markets. While this approach considered deep uncertainty, reported outbreaks were considered as "perfect information". Other studies, such as (Pinto et al. 2012) considered random and non-random placement of surveillance observers which partially captures reporting uncertainty; nonetheless these studies did not formulate any optimization algorithm maximizing detection accuracy even in presence of uncertainty.

Moreover, outbreak uncertainty is certainly biased, thus (Pinto et al. 2012)'s and others' approaches, based on regular schemes of surveillance underestimated such bias related to the anisotropic spreading of epidemics. In addition to such considerations, (Pinto et al. 2012)'s approach considered the problem of surveillance and outbreak source detection as coupled problems. In this study and in (Convertino and Hedberg 2014a) it is shown that such problem can be mathematically decoupled and the optimization of surveillance implies the optimal detection of outbreak sources but not vice versa. Pinto et al. (2012)'s model, analogous to the method used by telecommunication towers to pinpoint cell phone users, focused on arrival times of outbreaks - thought as information arrivals - that are correctly reproduced by the correct definition of outbreak sources. Here, we propose a simpler model, based on the original idea of (Brockmann and Helbing 2013) and subsequently developed by (Convertino and Hedberg 2014a; Convertino and Hedberg 2014b; (Convertino M, Hedberg C: Multisite outbreak detection: how many foodborne outbreak are attributable to farmer's markets?, submitted)) and (Manitz et al. 2014), that considers the minimization of the error in the distance from real outbreak multi-sources with the most informative surveillance networks. All explored surveillance networks correspond to different reported outbreak patterns. Other approaches for designing surveillance have been developed in the past; for instance (Bajardi et al. 2012) developed a model for robust outbreak cluster detection based on network features. While such model has an optimization component, deep uncertainty in outbreak reporting (that is uncertainty related to report cases at a selected site and time) is not considered and the detection of clusters do not detect outbreak sources. The optimal surveillance network maximizes outbreak information, or equivalently minimizes the uncertainty related to the average distance between real and predicted outbreak sources.

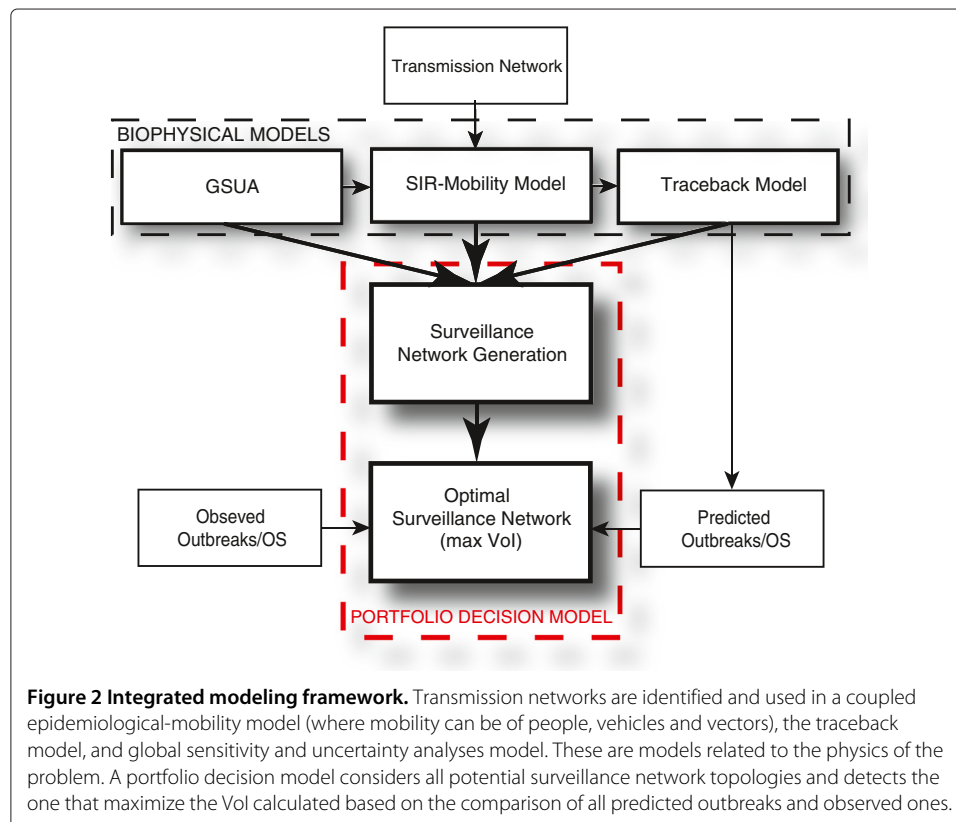
As a side note we want to emphasize that outbreak sources are oftentimes confused with "onset areas" which are areas where first cases occur. Those areas are not necessarily where the first contamination of the environment and/or human infection occur considering environmental and social determinants that contribute to the spread of the pathogen in space. However, considering the typical short range of dispersal of most pathogens (Mundt et al. 2009), onset areas tend to coincide with outbreak sources. In some cases one of the onsets is the outbreak source and all other onsets are generated by epidemic propagation. For other outbreaks in which vehicle or vector (e.g, wild birds, food, and people via airplanes) of pathogens travel for long distances the proposed model works well because of the large resolution at which outbreaks are investigated and because of the uniqueness of the outbreak source. We argue that the algorithm of (Pinto et al. 2012) is more likely to fail in detecting real outbreak sources because it does not consider uncertainty in reported outbreaks, thus the implemented matching function of the observed delay in reported information detects onset areas. Certainly, the definition of outbreak

sources is always dependent on the spatial and temporal scale of the system considered and care has to be placed when planning control strategies based on supposed outbreak sources.

### Proposed approach

The model here proposed is explicitly based on a multiobjective optimization in which the value of information of reporting nodes is maximized. This maximization determines a minimization of the uncertainty related to the detection of the correct outbreak source. In order to test the model independently of the epidemiological dynamics determined by the interplay of socio-environmental and biological factors, we consider three well-known epidemics with different geographic range, pathogen features, and transmission routes. In this way we focus on macro-linkages between patterns and process on networks that allows us to maximize detection via an optimally design surveillance. Figure 2 shows the major steps of the proposed algorithm. Specifically, the novelties of the integrated model are the following.

- Global Sensitivity and Uncertainty Analyses (GSUA) (Convertino et al. 2014a; (Convertino M, S L, M A, Morris S: Importance, Interaction, and Scale-dependence of Cholera Outbreak Drivers: Metamodeling Predictions, submitted); Saltelli et al. 2008) as a method for evaluating the systemic uncertainty in reported outbreaks for the determination of outbreak sources, importance and interaction of transmission network variables. GSUA fully attributes uncertainty to any input factor of the model via probability distributions, and such uncertainty is propagated to study how it



**Figure 2 Integrated modeling framework.** Transmission networks are identified and used in a coupled epidemiological-mobility model (where mobility can be of people, vehicles and vectors), the traceback model, and global sensitivity and uncertainty analyses model. These are models related to the physics of the problem. A portfolio decision model considers all potential surveillance network topologies and detects the one that maximizes the Vol calculated based on the comparison of all predicted outbreaks and observed ones.

affects the uncertainty of the output that is the effective distance from observed outbreak sources (Brockmann and Helbing 2013).

- Value of Information (VoI) portfolio model (Convertino and Valverde 2013; Convertino et al. 2014c, 2014d; Trainor-Guitton et al. 2012) with Pareto optimization for the design of optimal surveillance networks by selecting observers in network topologies with the highest information for detecting outbreak sources.
- Network based (Brockmann and Helbing 2013; Convertino and Hedberg 2014a; Newman 2003) versus random node surveillance design, and effective distance based prediction of outbreak spread with source detection. The effective distance implementation (Brockmann and Helbing 2013) does require only the knowledge of outbreak occurrence versus the number of cases.

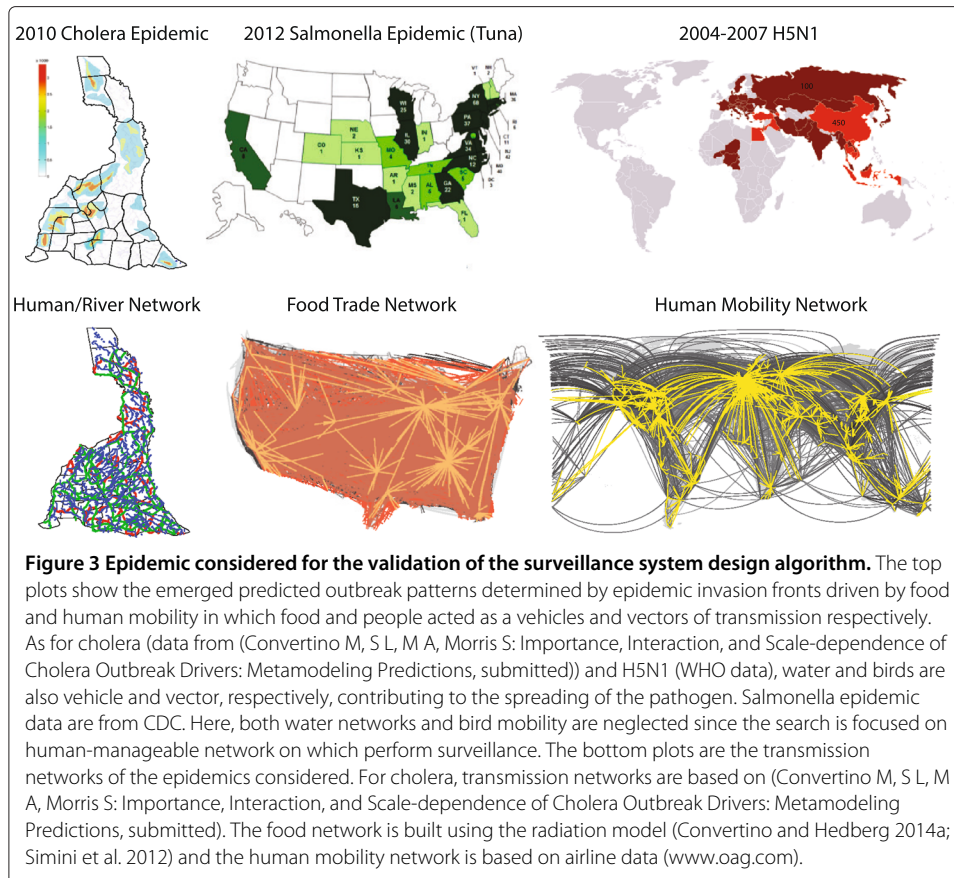
## Materials

### Case-study outbreaks

The choice about outbreaks considered in this study is related to the objective of capturing the salient links between epidemic patterns and processes regardless of disease etiology that is a result of peculiar epidemiological dynamics in terms of socio-environmental drivers, velocity, and geographical coverage (Convertino et al. 2009). This is with the aim of a general theory and cyber-infrastructure development for automated surveillance (Convertino and Hedberg 2014a). The goal is to show how observed outbreak patterns can be analyzed with the same physical-based models to detect outbreak sources and design surveillance networks considering the underlying transmission networks; thus, linking structure and functions of networks (Newman 2003) related to “invisible” outbreak dynamics over space and time. We chose among the fastest and high incidence rate epidemics: the 2004-2007 H5N1 pandemic (Kilpatrick et al. 2006), the 2010 cholera epidemic in Cameroon ((Convertino M, S L, M A, Morris S: Importance, Interaction, and Scale-dependence of Cholera Outbreak Drivers: Metamodeling Predictions, submitted); Guevart et al. 2006; Njoh 2010; Tatah et al. 2012), and the 2012 Salmonella epidemic in USA due to contaminated tuna (GATS 2014a). 34%, 27%, and 12% is the incidence rate for H5N1, cholera, and Salmonella, respectively. The estimated traveling velocity of these epidemics is 11, 22, and 122 km/month, respectively. The velocity is estimated using epidemic arrival times and distance covered over time. Additional file 1 report more information about the outbreak considered, and Figure 3 shows the observed outbreak patterns and transmission networks.

### Mobility and surveillance networks

Mobility networks are the backbone of both transmission and surveillance networks that are both subnetworks of the mobility network. Mobility networks can be related to the mobility of vehicles, vectors, or pathogens themselves. For salmonella in tuna and H5N1 in humans we consider the food and human mobility network that are the primary networks responsible for the transmission of the pathogen. For cholera, previous studies such as (Pinto et al. 2012), have considered only the river network; however, considering the higher importance of human mobility in the transmission of cholera infections here we take into account human mobility and water-flow networks, simultaneously. However, only one network is enough for detecting outbreak sources. Surveillance networks



can be though to coincide with human mobility networks rather than with environmental networks unless sensors are placed along those.

Human and food mobility network fluxes are estimated using the radiation model of (Simini et al. 2012) (Additional file 1) that is based on the population distribution. The population dataset is from the Web sites of the Gridded Population of the World and the Global Urban-Rural Mapping projects (CIESIN 2014; GRUMP 2014), which are run by the Socioeconomic Data and Application Center (SEDAC) of Columbia University. According to this dataset, the surface of the world is divided into a grid of cells that can have different resolution levels. Each of these cells has been assigned an estimated population value. Out of the possible resolutions, we have opted for cells of  $15 \times 15$  minutes of arc to constitute the basis of our model. This corresponds to an area of each cell approximately equivalent to an area of  $25 \times 25$  km<sup>2</sup> along the Equator. The dataset comprises 823,680 cells, of which 250,206 are populated. Since the coordinates of each cell and those of the airports in the World Airport Network are known, the distance between the cells and the airports can be calculated.

For the H5N1 we consider only the estimated mobility for the World Airport Network (WAN) (Colizza et al. 2006; den Broeck et al. 2011). WAN is composed of 3362 commercial airports indexed by the International Air Transport Association (IATA) that are located in 220 different countries. The database contains the number of available seats per year for each direct connection between two of these airports. The coverage of the dataset is estimated to be 99% of the global commercial traffic. The WAN can be seen as

a weighted graph comprising 16,846 edges whose weight, represents the passenger flow between airports. The network shows a high degree of heterogeneity both in the number of destinations per airport and in the number of passengers per connection.

The International Agro-Food Trade Network (IFTN) (Convertino and Liang 2014; Ercsey-Ravasz et al. 2012) for the USA is built using data of (ComTrade 2014) for the worldwide linkages among countries, and data from the United States Department of Agriculture, Foreign Agricultural Service's Global Agricultural Trade System (GATS) (FAS 2014; GATS 2014b) for the trade network of imported food commodities in USA. Additional data is used from the Agricultural Marketing Service (AMS 2014). The use of both datasets for food allows one to cross check erroneous data and to complement missing data missing. Data of imported and produced food that is locally consumed is obtained from FAOSTAT food balance sheets (FAOSTAT 2014). Such information is useful for calibrating the model factor that determines how many individuals consume the contaminated food.

## Methods

The integrated modeling used in this paper consists in the following main sequential steps.

1. Generation of potential surveillance networks within network topologies;
2. Model-based detection of onset/outbreak sources with uncertainty assigned to the reported outbreaks for all feasible surveillance networks from (1); the determination of outbreak sources is performed by minimizing the uncertainty in the effective distance estimations; and,
3. Determination of the most informative surveillance network by maximizing the VoI considering all combinations of feasible surveillance networks corresponding to potential reporting patterns.

### Effective distance and traceback model

The model is based on the optimal inference of epidemiological factors that better explain the spatio-temporal occurrence of foodborne outbreaks for any potential outbreak location to which different food trade paths correspond. In this study, differently from (Convertino and Hedberg 2014a), Manitz et al. (2014), and (Brockmann and Helbing 2013), deep uncertainty in reported outbreaks is introduced. The possible outbreak sources are determined by perturbing the reported outbreaks considering different surveillance networks. All feasible outbreak sources are tested considering their likelihood to satisfy the relationship between outbreak arrival times and velocity at any time step of the epidemic and for all infected communities simultaneously.

We define the effective distance (Brockmann and Helbing 2013) of directly connected nodes as a function of the most probable food trajectories defined by the radiation model (Section S2.1 in Additional file 1) that can be derived from the connectivity matrix  $P$  that specifies which nodes are connected; specifically  $d_{mn} = 1 - \log P_{mn}$  where for any couple  $m - n$  there are multiple distances (Brockmann and Helbing 2013). For any candidate outbreak source we use the definition of "shortest path" effective distance  $D_{mn}$  (Brockmann and Helbing 2013) from an arbitrary reference node  $n$  to another node  $m$  in the network (not necessarily directed connected) as the length of the shortest path from  $n$  to  $m$ ,



$D_{mn} = \min_{\Gamma} \lambda(\Gamma)$ , where  $\lambda(\Gamma)$  is the directed length of an ordered path  $\Gamma = n_1, \dots, n_L$  as the sum of effective lengths along the path Brockmann and Helbing (2013). The average shortest path tree is calculated among all effective distances that are associated to potential candidate outbreak sources. Thus, both average effective distance ( $d_{mn}$ ) and average shortest path tree (that is the shortest path among many effective distances) responsible of outbreak spreading ( $D_{mn}$ ) depend only on the static mobility matrix weighted by the food fluxes. There is a family of effective distances and there one of shortest path distances between community  $m$  and  $n$ . Theoretically, there can be more than one shortest path distance but the dimension of the ensemble of shortest path distances is always smaller (or equal) than the ensemble of effective distances.

The above topological considerations allow us to say that, on a spatial scale described by the metacommunity model, the complexity of spatio-temporal outbreak patterns is largely determined by the structure of the mobility network (in this case of food and people movement) and not by the nonlinearities of the pathogen/-food/-people couples and epidemiological factors. According to the second Newtons law the arrival time of outbreaks is defined as

$$T_a = D_e(P)/v_e(\alpha, R_0, \gamma, \epsilon) . \quad (1)$$

This equation states that effective outbreak distances  $D_e$  can be computed with high fidelity based on outbreak arrival times on and effective spreading speed  $v_e$ , and that each factor depends on different factors of the dynamical system considered (Brockmann and Helbing 2013). Note that here we indicate  $D_e$  as  $D_{mn}$  previously defined. The epidemiological factors associated with the classical SIR model determine the effective speed, whereas effective distance depends only on the topological features of the static underlying network, i.e., the matrix  $P$ . Because,  $T_a$  and  $v_e$  are know and estimated from data (that can be generated in real-time), respectively, it is easy to estimate  $D_e$  of an outbreak considering also the uncertainty in epidemiological dynamics and food trade network. Thus, Eq. 1 can be used as a test of hypotheses where hypotheses are about all candidate outbreak sources responsible for the observed outbreak patterns. The most likely geographical distance can be assessed after determining the effective distance and considering all potential food trade paths. The SIR dynamics (Section “Epidemiological dynamics and information preading: predictive metacommunity model”) is calibrated on the range of the outbreak velocity and arrival times, and is simulated by maximizing the prediction accuracy of the observed outbreak patterns in terms of cumulative outbreaks. Such accuracy is maximized when the uncertainty about the effective distance from the real outbreak source is minimized, that determines the maximization of the concentricity of outbreak spreading waves (Section “Outbreak detection: linking patterns and processes”). The concentricity is maximized only when the accuracy in the reported outbreaks is maximized that is equivalent to the maximum of the value of information (Section “Value of information and Pareto optimization”) considering any uncertainty source in outbreak patterns, supply chain and disease dynamics (Section “Global sensitivity and uncertainty analyses”).

#### **Epidemiological dynamics and information spreading: predictive metacommunity model**

The metacommunity model is inspired by the metacommunity concept in (Convertino 2011; Convertino et al. 2009) and (Convertino M, S L, M A, Morris S: Importance, Interaction, and Scale-dependence of Cholera Outbreak Drivers: Metamodeling Predictions,

submitted) and the mathematics in (Brockmann and Helbing 2013) and more recently in (Convertino and Hedberg 2014a). The metacommunity is an ensemble of metapopulations (not only individuals) that can be divided into different categories. The model consists in the coupling of a classical susceptible-infected-recovered (SIR) dynamic model with a radiation model (Simini et al. 2012) of food/human mobility that determines the interlinked variations of food/human mobility and population states for all communities of the system considered simultaneously. The infection of individuals occurs only via food-human and/or human-human interactions that define the infectious contact matrix. In general, the model is a reaction-diffusion transport model of information spreading with embedded bias transport in relation of the radiation model of mobility. The information is about the occurrence of a case related to the epidemic considered; yet, information can be a binary variable in which “1” stands for the occurrence of a case in a location within the system. Uncertainty in reporting means to have some areas that report the outbreak or not. Thus, such uncertainty affects the ability to reproduce correctly the observed outbreak pattern and to detect the correct outbreak source. The metacommunity model, whose mathematics is written as in (Brockmann and Helbing 2013), is

$$\begin{aligned}\delta_t j_n &= \alpha s_n j_n \sigma(j_n/\epsilon) - \beta j_n + \gamma \sum_{m \neq n} P_{mn} (j_m - j_n) \\ \delta_t s_n &= -\alpha s_n j_n \sigma(j_n/\epsilon) + \gamma \sum_{m \neq n} P_{mn} (s_m - s_n),\end{aligned}\quad (2)$$

where  $s_n = S_n/N_n$ ,  $j_n = I_n/N_n$ , and  $r_n = 1/s_n j_n$ .  $R_0 = \alpha/\beta$  that is the ratio of mean infection and recovery rate. A detailed derivation of such equations is provided in (Brockmann and Helbing 2013). The mobility parameter  $\gamma$  is the average food/human mobility rate estimated in the data (Section “Mobility and surveillance networks”), i.e.  $\gamma = \Phi/\Omega$ , where  $\Omega = \sum_n N_n$  is the total population in the system and  $\Phi = \sum_{n,m} F_{nm}$  is the total food/human flux. This yields numerical values in the range  $\gamma = 0.01150.0379 \text{ day}^{-1}$ . The matrix  $\mathbf{P}$  with  $0 \leq P_{mn} \leq 1$  quantifies the fraction of the food/human flux with destination  $m$  exported from node  $n$ , i.e.,  $P_{mn} = F_{mn}/F_n$ , where  $F_n = \sum_m F_{mn}$  (Brockmann and Helbing 2013). The sigmoid function  $\sigma(x) = x^\eta/(1 + x^\eta)$  with gain parameter  $\eta \gg 0$  accounts for the local invasion threshold  $\epsilon$  and fluctuation effects for  $j_n < \epsilon$  (Brockmann and Helbing 2013). Typical  $\epsilon$  and  $\eta$  average values are  $\eta = 4, 8, 10, 12, \dots, \infty$  and  $-\log_{10}\epsilon = 4, 6, 8, \dots$ . For human-driven contagions  $\epsilon$  is related to the number of infected people that travel within the system. For food  $\epsilon$  depends on how many individuals eat the contaminated food. Here to avoid the use of two factors, one for local consumption and another for consumption of contaminated food, we use  $\epsilon$  as the only factor determining spread of contamination. This means that we just consider the portion of consumed food that is contaminated. The fraction of produced and imported food that is locally consumed is assessed from (FAOSTAT 2014). In general, it is possible to assume that consumption is proportional to the size of the population of each community. This is one of the motivations for which the radiation model works well in reproducing food trade patterns (Section “Mobility and surveillance networks”) (Brockmann and Helbing 2013). Considering any effective path as a linear system - discretized at regular intervals  $l$  along the transmission network - the set of Equations 2 can be rewritten as

$$\begin{aligned}\delta_t j_n &= \alpha s_n j_n \sigma(j_n/\epsilon) - \beta j_n + D\delta_\Gamma^2 j_n \\ \delta_t s_n &= -\alpha s_n j_n \sigma(j_n/\epsilon) + D\delta_\Gamma^2 s_n,\end{aligned}\quad (3)$$

where  $D = l^2\gamma$  is the diffusion coefficient that is related to a classic form of a Fisher equation which has been deployed in mathematical epidemiology for describing wave spreading in reaction-diffusion system such as for epidemics (Belik et al. 2011; Campos and Mendez 2005). Note that  $s_n = s_n(\hat{s}(O), t)$  and  $j_n = j_n(\hat{s}(O), t)$  where  $(\hat{s}(O), t)$  is the surveillance network in Equation 4.

For sufficiently localized initial conditions this systems exhibits traveling waves with speed  $c = 2\sqrt{\alpha D(1 - \beta/\alpha)} \sim \sqrt{\gamma}$  (Belik et al. 2011). Such velocity can consider bias transport as in (Bertuzzo et al. 2007) in case the bias is known and elevated in the transport process. The telegraph model introduced by (Holmes et al. 1993) and simulated by (Bertuzzo et al. 2007) for cholera possesses both diffusion and wave motions. The Fisher equation of diffusion works well when using the effective distance, while the telegraph model equation works well when using the geographical distance of the transmission network. GSUA is used here to consider uncertainty in the traceback model and to assess the relative importance and interactions of factors - epidemiological and mobility network factors (Section “Epidemiological dynamics and information spreading: predictive metacommunity model” and Additional file 1: Section S2.1, respectively) - leading to outbreaks. Interactions of causal factors are typically neglected as well as the determination of the causality of such factors. The traceback model is run multiple times according to the Sobol scheme (Additional file 1: Section S1.2) (Figure 2) sampling all model factors along their probability distributions (Additional file 1: Figure S4) estimated from data or assumed as uniform according to a maximum entropy principle (Convertino et al. 2014a).

### **Outbreak detection: linking patterns and processes**

The simplest method to determine which food and which outbreak source are the most likely to determine the observed outbreak patterns is to test which feasible candidate source has the lowest variability in terms of mean and variance assessed from Eq. 1. This is essentially a portfolio problem (Convertino and Valverde 2013) in which a large set of outbreak source alternatives exist and the whole set is evaluated considering average risk and variance, where in this case the “risk” is related to the inability to predict the correct outbreak patterns in space and time. Thus, for each potential candidate outbreak location the model computes the effective distance to the subset of nodes in which an outbreak is observed. This is done for any time step of the epidemic. On the basis of this set of effective distances (denoted by different outbreak sources), we compute the mean  $\mu(D_e)$  and standard deviation  $\sigma(D_e)$  of the effective distance. As shown by (Brockmann and Helbing 2013), concentricity of epidemic waves increases with a combined minimization of mean and standard deviation of the estimated effective distance. In other words, such approach minimizes the deviation from the expected relationship arrival times-distance, or velocity-distance, equivalently. Here we perform a minimization of the sum  $\sqrt{\mu(D_e)^2 + \sigma(D_e)^2}$  as in a portfolio model where the objective function is the Euclidian distance built with the variables to minimize (Convertino and Valverde 2013). The correct outbreak source is the one that satisfies the minimum of the aforementioned sum for the whole epidemic duration. It is possible to visualize the dynamics of epidemic spread and concentricity of outbreak waves by plotting outbreak sources at the center of a network where all other nodes are placed around the central node at a radial distance equal to the effective distance. Correct outbreak sources determine cohesively evolving

outbreak waves from the center to the nodes at the boundary from the beginning to the end of the epidemic.

### Value of information and Pareto optimization

The Value of Information (VoI) is classically defined as the amount a decision maker would be willing to pay for information prior to making a decision (Convertino et al. 2014c; Keisler 2004). In our case we consider the set of observers  $\hat{s}(O)$  as the economically and epidemiologically valuable features on which stakeholders may take a decision. In the real world such decision may be related to early control strategies of the epidemic that have economic and health repercussions. In order to calculate the VoI, or the “value of spatial information” more precisely because  $\hat{s}(O)$  is a spatially explicit variable (Trainor-Guitton et al. 2012), we assess uncertainty reduction in the detection of outbreak sources when extra knowledge about outbreaks is available. The coverage of the surveillance network determines where we have information about cases. Thus, that is a macroscale uncertainty whose magnitude can be high or low, and there is an optimal coverage determined by the optimal surveillance network the proposed model aims to determine. The uncertainty in case reporting is assessed as distance  $d(\hat{s}(O))$  from the outbreak source (Eq. 1 and Section “Outbreak detection: linking patterns and processes”), that is related to the variability in the reported cases expressed by the function of cumulated cases  $C(t)$  over the epidemic. Note that in general,  $d$  can be both effective and geographical distance. The evaluation of the VoI is done by comparing the uncertainty before and after “pinching” the input, i.e., replacing it with another input (of different value) and a lower degree of uncertainty, or with no uncertainty (Oakley 2009). The former scenario is when a set of monitored communities is considered for any network topology (random, high degree, and maxVoI), and the latter is when all the monitoring communities that are reporting cases are considered. In this case study we consider such scenarios with complete presence (as reported) or absence of reported cases in USA for all states involved. The VoI in the latter scenario is commonly known as “value of perfect information” that is assumed in this study equal to the information gathered from all data. By considering the above reasoning and  $\hat{s}(O)$  with its probability structure in space and time, the VoI is defined as:

$$VoI(t)_{i,n} = [d(\hat{s}(O), t)_{i,n=1} - d(\hat{s}(O), t)_{i,n=M}] \propto [C_{i,n=1}(t) - C_{i,n=M}(t)], \quad (4)$$

where  $C(t)$  and  $d(\hat{s}(O), t)$  is evaluated for all possible networks  $n$  (among all topologies (Additional file 1: Figure S1)), communities  $i$ , and time steps  $t$ ;  $M$  is the total number of considered networks.  $\propto$  stands for “proportional to”. Hence, the VoI is the difference of  $d(\hat{s}(O))$  estimated for different networks and all communities of the system. The VoI (Eq. 4) can be theoretically defined by the difference between two estimates of  $d(\hat{s}(O))$  for different surveillance networks (Eq. 4) (i.e. a different set of observer nodes). VoI is always taken equal or greater than zero and a rational decision maker always chooses more valuable information. Note that here we do not consider any cost function in the VoI because no information is available; however, that information is easily included as function detracted to the predictive benefits in Eq. 4. In presence of a physical-based model reproducing outbreaks in space and time the VoI can also be measured as accuracy in predicting prevalence patterns. In fact, errors in detecting outbreak sources translate into errors in initial conditions of the model that reduce model predictive accuracy. Such

predictions generated by models that take in input surveillance data are used for public health decision making and system design.

### Global sensitivity and uncertainty analyses

GSUA involves five steps: (1) the probability distribution functions (pdfs) for each input factor are selected (epidemiological, mobility, and network topology factors); (2) sample points are generated on the input factor distributions using the Sobol method; (3) multiple model execution using each of the sample points and a set of outputs is generated; (4) global sensitivity analysis is performed (i.e., assessment of factor interaction and importance); and, (5) the important input factors are selected. Probability distributions of mobility network factors (link length and clustering coefficient) are derived from data as in (Convertino M, S L, M A, Morris S: Importance, Interaction, and Scale-dependence of Cholera Outbreak Drivers: Metamodeling Predictions, submitted). Epidemiological factors ( $\alpha$ , and  $\beta$ ) and contamination factors ( $\sigma$  and  $\epsilon$ ) are assumed to be normally distributed within a normalized range [0,1].  $\gamma$  is related to the range of variability of the total food and human mobility flux. As for human mobility, more than 60 million people travel billions of miles on more than 2 million international flights each week. GSUA is extremely important in complex systems that are dominated by factor interactions rather than an additive contribution of each single factor. In fact, classical sensitivity analysis fails to account for factor interactions and does not offer a full probabilistic investigation of system states; yet, classical one factor sensitivity analysis provides misleading conclusions. GSUA is rarely performed and classical one-factor at a time sensitivity approaches are used to evaluate uncertainty. GSUA first assigns uncertainty and then drives sensitivity analysis via variability of model output that accounts for non-linearity among model input factors and those and model output. The Sobol method (Sobol 1993, 2001) is a variance-based method that performs a quantitative analysis of model sensitivity based on the principle of variance decomposition. According to this principle, the full variance of the model output is given by the sum of the variances of all input factors (Saltelli et al. 2008). The Sobol method has the capacity to quantify the influence of the full range of variation of each input factor as well as the interaction effects among the input factors on the model output (Saltelli et al. 2008). The Sobol method estimates sensitivity measures which summarize the model behavior. The most common measure of sensitivity is the first-order sensitivity index,  $S_i$ , that represents the main effect (direct contribution) of each input factor to the variance of the output. It is expressed as  $S_i = V_i/V$  where  $V_i$  is the part of the variance due to the input factor  $X_i$ , and  $V$  is the total variance of the model output. Another measure of sensitivity is the total factor sensitivity that includes the interactions. The total effect index,  $S_{T_i}$ , that is the result of the variance decomposition, accounts for the total contribution to the output variation due to factor  $X_i$ , i.e., its first-order effect plus all the higher order effects due to interactions (Saltelli et al. 2008). Thus, the total effect index of factor  $X_i$  can be expressed as  $S_{T_i} = S_i + S_{i,i+1} + S_{i,i+1,i+2} + S_{i,i+1,i+2,i+3} + S_{i,i+1,i+2,i+3+\dots+N_{IF}} = S_i + S_{I_i}$ , where  $N_{IF}$  is the number of input factors and  $S_{I_i}$  is the sum of all the interaction indexes. The Sobol pairwise interaction between factors can be calculated and we identify this with the notation  $S_{I_{ij}}$  where  $i$  and  $j$  are the factors that are considered. As an example, in the case of three input factors  $S_{T_1}$  is the total sensitivity index of  $X_1$ ,  $S_1$  is the main effect of  $X_1$ ,  $S_{I_{12}}$  is the interaction effect between  $X_1$  and  $X_2$ , and  $S_{I_{123}}$  is the interaction effect between  $X_1$ ,  $X_2$ , and  $X_3$ . Considering the previous expression,  $S_{T_1} - S_1$  provides a

measure of how much  $X_1$  is involved in interactions with all other input factors (Saltelli et al. 2008). The sum of all  $S_i$  is equal to one for additive models and less than one for non-additive models. The difference  $1 - S_i$  can be used as an indicator of the presence of interactions in the model and we indicate that as  $S_j$ . The number of simulations required for the Sobol method for a two-index analysis (i.e., for first order and total indices) is given as  $N = M(2k + 2)$  where  $M$  is the sample size of each index (typically taken between 500-1000) and  $k$  is the number of uncertain input factors after the Morris screening. In this study because  $k = 7$ , and we take  $M = 600$  we have 9,600 Sobol simulations. Since the Sobol method uses a pseudo-randomized multivariate sampling procedure, it can be used as a basis for a global uncertainty evaluation by constructing the pdfs and cumulative distribution functions for each of the selected outputs.

### Results and discussion

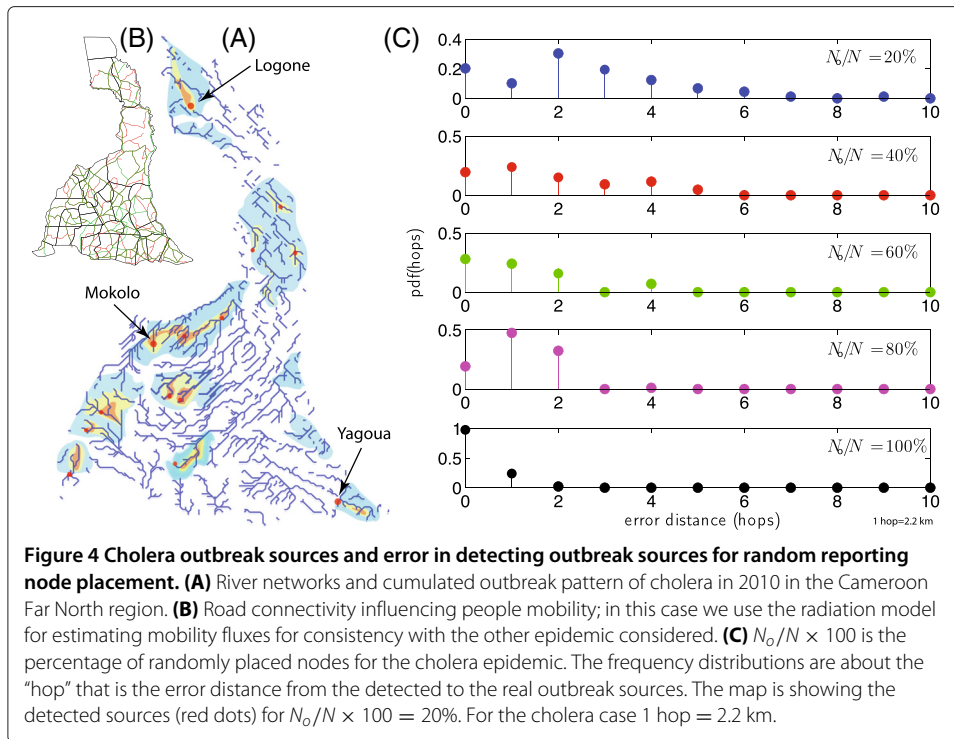
The results of the study hold when surveillance works for any network topology. Here surveillance is seen as a function of the network to contrast the spreading of infectious diseases. The random surveillance design model, i.e. placing reporting nodes of outbreaks randomly along the transmission network produces errors in the distance from the real outbreak sources as shown Figure 4 for the 2010 cholera outbreak in Cameroon. Table 1 shows the surveillance node coverage necessary for achieving an outbreak source detection higher than 90% also for other surveillance network models. The higher the percentage the higher is mismatch between the surveillance network model and the epidemic transmission process. For instance, for cholera 20% of random node coverage is sufficient for detecting precisely the outbreak source - possibly because of the extended human mobility and outbreaks - but for other highly heterogeneous epidemics such as H5N1 the percentage increase to almost 50%. For cholera, with a reporting node coverage of 20% the error is 2 hops (i.e., about 4.4. km) (Figure 4). With higher percentage of node coverage, from 40 to 80 %, the error is more or less invariant to 2 km that is a very small error considering the extension of the Far North Region of Cameroon (~ 241 km long and 145 km wide). For  $N_0/N = 100\%$  the error goes to zero and there is almost perfect detection of the three main outbreak sources.

By selecting as surveillance nodes only the nodes with the highest degree, where the node degree is defined as the the number of connections, the coverage required for achieving the same detection accuracy (90%) is lower than for the random coverage. 5, 13 and 38% are the percentages required for cholera, Salmonella, and H5N1 (Table 1). The results are quite invariant considering different thresholds - from 5 to 10 - for establishing "high degree" nodes. We believe that the high degree surveillance scheme captures

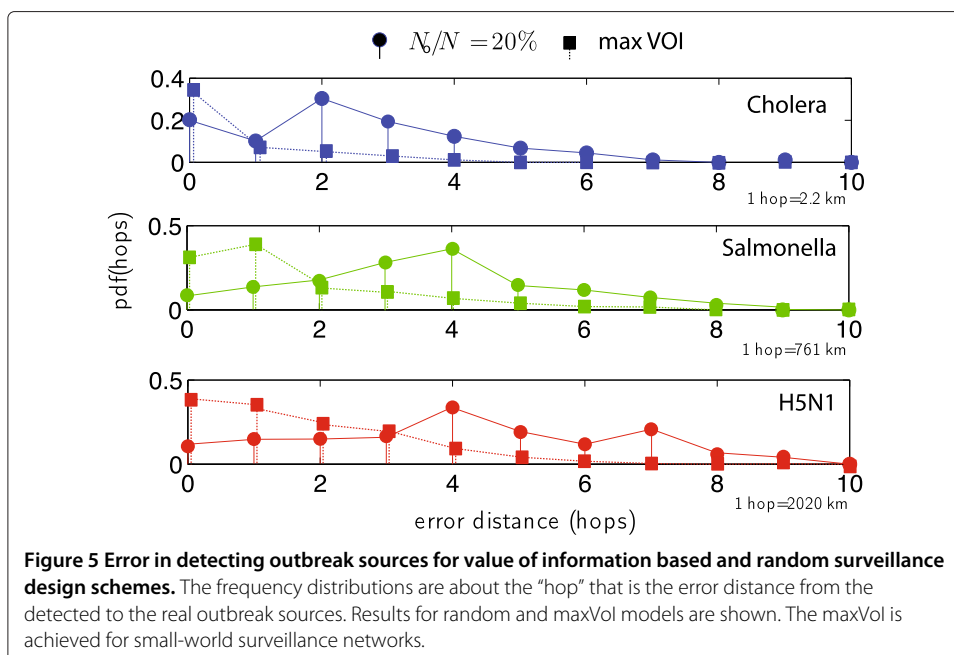
**Table 1 Surveillance node coverage for achieving an outbreak source detection higher than 90%**

Outbreak (Network)	Observers		
	Random	High degree	maxVol
Cholera Cameroon 2010 (Tree)	20%	5%	3%
Salmonella USA 2012 (Tree)	37%	13%	5%
H5N1 2007 (Graph)	46%	38%	20%

Number are the percentage ( $N_0/N \times 100$ ) of observers necessary to achieve a probability of localization of outbreak sources equal to 90%. Such percentage is for the random, high degree, and maxVol surveillance network topology irrespectively of the outbreak type. The maxVol algorithm is the one used for all potential surveillance network topologies.



better outbreak evolutions than other schemes because epidemics are very likely to pass through the most connected nodes. However, selecting only these nodes is not enough since many infections pass through other nodes. The maxVol model (Eq. 4) determines a drastic reduction in the number of surveillance nodes required to detect outbreak sources with an accuracy higher than 90%. 3, 5 and 20% are the percentages of nodes required for cholera, Salmonella, and H5N1 (Table 1). Such percentages correspond to errors

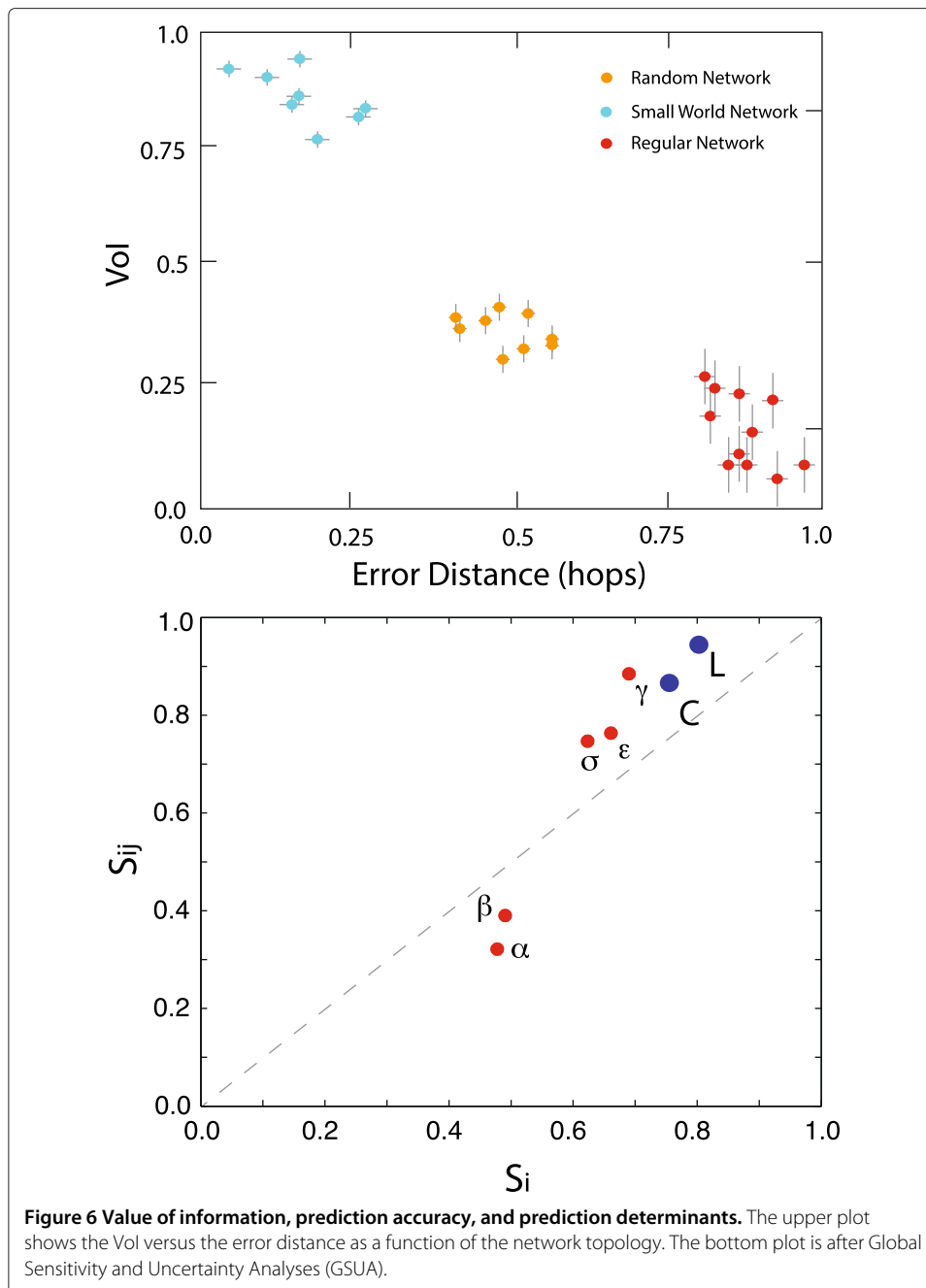


of  $0.5 \times 2.2$  km,  $1.0 \times 761$  km, and  $1.5 \times 2020$  km as distances from the correct outbreak sources (Figure 5). In case of random placement the error in km units is  $2.5 \times 2.2$  km,  $3.8 \times 761$  km, and  $4.6 \times 2020$  km. The reduction in surveillance nodes determined by the maxVoI model is 17%, 32%, and 26% with an average 25% reduction. These differences creates differences in predictions of outbreak patterns larger than 40% in terms of cumulative cases. The magnitude of reduction is significative of the both interlinked epidemic dynamics and transmission network; more space filling epidemics such as cholera show a smaller advantage of the maxVoI model than more long-range and less space filling epidemics such as Salmonella and H5N1 for which the food/human mobility network does not covert the whole system.

The ability of the maxVoI model to detect outbreak sources with high accuracy and low surveillance nodes is related to determination of the most meaningful nodes (in terms of known information for describing outbreak patterns) via a portfolio model with Pareto optimization that explores all potential combinations of observers and corresponding outbreak information. The VoI is maximized, thus the difference between two estimated distances from the real outbreak sources is maximized. The set of observer nodes for which the VoI is the maximum allows to predict the real outbreak source and the observed outbreak pattern.

For the outbreak considered the error in the estimation of the distance, and thus in the outbreak velocity, determined by the uncertainty of reporting outbreak arrivals is minimized by the maximization of the VoI. Such optimization results maximized in terms of VoI for small-world networks that likely capture both local and large scale outbreak spreading regardless of epidemiological details. Figure 6 shows that there is clearly an association between network topology, maximum VoI and error distance regardless of the epidemic considered. Additional file 1: Figure S2 shows these findings in an alternative way by plotting the normalized probability distribution of estimated distances from outbreak sources and the Integrated Network Resilience (INR) (Halpern et al. 2012; Pandit and Crittenden 2012) (see Additional file 1, Section S2.3, and Additional file 1: Figure S1) of all surveillance network topologies. Random and regular networks with medium and high INR, seem limited in maximizing the VoI; yet, with such network it is hard to predict accurately outbreak sources, and the evolution of outbreaks over space and time. INR, that is manifesting the structural resilience of a network in terms of reconfiguration ability (see Additional file 1) determines also the ability to inform about outbreak evolution. Additional file 1: Figure S3 shows the error in epidemic spreading velocity as a function of the VoI for all surveillance network topologies. The variability of the velocity is higher than the distance, nonetheless even the VoI-velocity error pattern shows a neat dependency of the epidemic dynamics on the network topology (Additional file 1: Figure S3). Small world networks have the highest VoI, thus the ability to detect outbreak sources and velocities carefully. Outbreak waves for different time steps of the epidemics considered are estimated with data related to the maxVoI surveillance model (bottom plot of Additional file 1: Figure S3). Such epidemic waves are dependent on the surveillance network and their estimation is crucial for detecting epidemic spreading and outbreak sources rapidly and carefully, respectively. Ideally, the wish is to have a surveillance system that detects and stops the epidemic spreading after the first cases that correspond to the first epidemic wave. The celerity of epidemic waves can be derived from Eq. 3 or more simply with Eq. 1 in the





effective distance domain. Wrong estimations of epidemic velocities are for inefficient outbreak reporting systems during the course of epidemics which makes very difficulty the detection of outbreak sources, attribution of outbreaks to food and/or infected individuals, and prediction of outbreak spread. The distance error from outbreak sources is anisotropic - in the sense that different errors are associated to different paths of the transmission network - because epidemic spreading is anisotropic in the geographical space. For this structured anisotropy related to the nonrandom and non regular structure of the mobility network, we believe that random and regular surveillance network are suboptimal surveillance networks. Vice versa, small-world surveillance networks such

as power-law networks capture the small-world features of mobility networks (Bajardi et al. 2012), thus they are able to monitor outbreak evolution over space and time very accurately.

Figure 6 also shows the result of GSUA in which uncertainty is considered for SIR and mobility model input factors. Results show that topological factors of the mobility network, namely the average link length and the clustering coefficient  $L$  and  $C$ , are the most important factors in explaining outbreak patterns. The average mobility rate  $\gamma$  and the invasion threshold  $\epsilon$  that defined the number of free-moving contaminated food and/or individuals are very important and interacting factors after topological factors. Epidemiological transition factors,  $\alpha$  and  $\beta$ , are secondary importance and weakly interacting factors because they are strongly dependent on the aforementioned factors. These results underlines even more the importance of knowledge and controllability of the network on which epidemic may occur. Mobility networks define where outbreaks can occur but also where surveillance should be crucially designed. Because a fully extended surveillance network is likely unfeasible considering also the highly varying mobility network of food and people, an optimally designed surveillance system via maxVoI is ideal in terms of benefits and costs.

As for H5N1 the source is correctly identified in Vietnam (Hanoi as the outbreak source); subsequently the epidemic is predicted to move to Thailand, and within weeks to spread to ten countries and regions in Asia, including Indonesia, South Korea, Japan and China. As for the Salmonella outbreak in tuna, the outbreak source is detected in New York City if only the US food trade is considered, and in Kerala, India, if the whole worldwide food trade network is taken into account. For the cholera epidemic in Cameroon, the health districts of Logone (North), Mokolo (West), and Yagoua (South East) are detected as outbreak sources. Considering the epidemic starting some time before the peak of cases, the urban health district of Maroua is detected as outbreak source. This evidences that: (i) the cholera epidemic in Cameroon is very likely generated by infections in rural areas and urban cases are due to infected people mobility among cities where human-human transmission is higher (in this case Maroua that is the capital of the Far North Region); and (ii) the definition of outbreak sources depend on the time horizon considered. For instance, Maroua is a source only if previous cases are neglected. Equivalent considerations are related to the scale of analysis: outbreak sources are related to the scale of the system considered. For example, it is known that cholera is endemic in large Sub-Saharan countries and the migration of populations, in which there are also infected people, among such countries is very high. For Cameroon, a high migration is reported from Nigeria in the East and Chad in the North, thus real outbreak sources may be located outside Cameroon. These considerations highlight the need to be very careful when planning control strategies that may result very inefficient for decreasing the systemic risk of infection because of the wrong targeting of outbreak source areas.

We show that efficient surveillance in the human mobility network is sufficient to detect outbreak source rapidly; this is despite complications played by wild bird migratory pathways and infected domestic birds for H5N1, and water-driven pathogen mobility for cholera. The almost immediate detection of outbreak sources has been shown in previous papers (Brockmann and Helbing 2013; Convertino and Hedberg 2014a). Certainly, in general it is important to detect which transmission network causes the fastest spreading

infections and we believe that in our cases, for both cholera and H5N1, the fastest infections are driven by human-human contacts related to the human mobility network as shown by epidemiological studies.

## Conclusions

The design of optimal surveillance networks allows stakeholders to detect outbreak sources and paths correctly. Such detection is only possible if information about outbreak “fingerprints”, i.e. reported outbreaks, is accurate and timely. We show that independently of system scale and resolution, biological and socio-environmental dynamics the “unknowns” of epidemiological dynamics - such as epidemic spreading paths and sources - can be determined considering the same physics principles of reaction-diffusion processes and information theory. We believe that our results are robust independently of any other epidemic types, or it can provide the methodological framework to design epidemic-specific surveillance network. The following results are worth emphasizing.

- The structural resilience of the system is independent of epidemic dynamics and can be identified by the Integrated Network Resilience that combines two key network topological factors. Because of the spatio-temporal anisotropy of outbreaks the small-world network is the surveillance network with the highest value of information, that implies the most accurate outbreak source detection and outbreak pattern prediction. Structural resilience is just one component of the system to respond to outbreaks. A portion of system resilience has to be built considering training of personnel dedicated to surveillance.
- The knowledge of mobility networks and the value of information of surveillance networks as subsets of the former is determinant for early detection and response of outbreaks. The use of effective distances allow to avoid the need to use topological features - necessary in the design process - that increase the computational complexity and uncertainty related to the estimation of the most likely transmission networks.
- Optimal dynamic design and real time surveillance are tasks that can be achieved using the proposed model. The design and outbreak source detection can be performed with a portfolio model with Pareto optimization that selects the most important observer networks by maximizing the VoI, and detects the most likely outbreak sources by minimizing the uncertainty in the effective distance from the real sources. The optimal design implies accurate detection but not vice versa. Dynamic surveillance network in which surveillance change configurations, for instance taking advantage of smart sensors placed along the mobility network, are at the frontier of technological development and can implement the maxVoI model.

The model can work for other kinds of spreading processes (e.g. viruses in computer networks and biomarker activation in chronic diseases) in which information spreads on partially known networks, although networks can be back-inferred with a reverse application of the model where the question is the identification of the networks (among all candidate ones) that leads to the generation of observed epidemic patterns. For instance, an important application of the model is related to the determination of the most effective preparedness information, for example provided by the surveillance network or other information networks, with the aim to decrease the systemic risk of

outbreaks in populations. Such information can be organized considering public media and/or online networks that inform people about performing tasks related to control outbreaks. Reversely, information networks can also be used as tools to more rapidly inform surveillance networks about the spread of outbreaks in populations.

## Additional file

**Additional file 1: Specifics of case-study outbreaks, supplementary methods (Radiation Model of Mobility Fluxes, Surveillance Network Topologies, and Integrated Network Resilience) and supplementary Figures.**

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

MC designed and performed the study. YL and HH helped in the data gathering and analysis. MC wrote the paper while YL and HH provided comments to the text. All authors read and approved the final manuscript.

### Acknowledgments

M.C. acknowledges fundings from the Healthy Food Healthy Lives Institute (grant "Impact of Local and Organic Foods on Food Safety: Assessment of Supply Chain Controls and Value Outcomes for Quantitative-based Food Policies"), the Institute on the Environment "Discovery Grant" (grant "Food System Design for Resilient Population Health: The Minnesota Model"), and the Minnesota Discovery, Research and Innovation Economy (MnDRIVE) Global Food Systems Faculty Scholar Start-Up Funding. Y.L. acknowledges the School of Public Health Dean Fellowship for the academic year 2014-2015.

### Author details

<sup>1</sup>School of Public Health, Environmental Health Sciences Division and Public Health Informatics Program, HumNat Lab, University of Minnesota, Minneapolis, MN, USA. <sup>2</sup>Institute on the Environment, University of Minnesota, Minneapolis, MN, USA. <sup>3</sup>Institute for Engineering in Medicine, University of Minnesota, Minneapolis, MN, USA. <sup>4</sup>Academic Health Center, Mayo Memorial Building, Suite 1232 12th fl., MMC807, 420 Delaware St. SE, 55455 Minneapolis, MN, USA.

Received: 3 July 2014 Accepted: 23 October 2014

Published online: 27 November 2014

### References

- AMS: **Agricultural marketing service**. 2014. Tech. rep. [[http://www.marketnews.usda.gov/portal/fv?&paf\\_gear\\_id=1200002&movNavClass=FVPHN&rowDisplayMax=25&repType=movementDaily&previousVal=&lastCommodity=&paf\\_dm=full&lastLocation=&volume=40000&locName=MEXICO&locAbr=MX&startIndex=51&dr=1](http://www.marketnews.usda.gov/portal/fv?&paf_gear_id=1200002&movNavClass=FVPHN&rowDisplayMax=25&repType=movementDaily&previousVal=&lastCommodity=&paf_dm=full&lastLocation=&volume=40000&locName=MEXICO&locAbr=MX&startIndex=51&dr=1); Accessed April 2014]
- Bajardi P, Barrat A, Savini L, Colizza V: **Optimizing surveillance for livestock disease spreading through animal movements**. *J R Soc Interface* 2012.
- Batz M, Hoffmann S, Morris JJ: **Ranking the risks, the 10 pathogen-food combinations with the greatest burden on public health**. 2011. Tech. rep., University of Florida, Emerging Pathogens Institute. [<https://folio.iupui.edu/bitstream/handle/10244/1022/72267report.pdf>; Date of access: 06/06/2013]
- Belik V, Geisel T, Brockmann D: **Natural human mobility patterns and spatial spread of infectious diseases**. *Phys Rev X* 2011, **1**:011001.
- Bertuzzo E, Maritan A, Gatto M, Rodriguez-Iturbe I, Rinaldo A: **River networks and ecological corridors: reactive transport on fractals, migration fronts, hydrochory**. *Water Resour Res* 2007, **43**(4).
- Brockmann D, Helbing D: **The hidden geometry of complex, network-driven contagion phenomena**. *Science* 2013, **342**(6164):1337–1342.
- Campos D, Mendez V: **Reaction-diffusion wave fronts on comblike structures**. *Phys Rev E* 2005, **71**:051104.
- CIESIN: **The gridded population of the world version 3 (GPWv3) population grids**. 2014. Tech. rep. [<http://sedac.ciesin.columbia.edu/gpw>; Accessed April 2014]
- ComTrade: **United Nations commodity trade statistics**. 2014. Tech. rep. [<http://comtrade.un.org/db/>; Accessed April 2014]
- Colizza V, Barrat A, Barthlemy M, Vespignani A: **The role of the airline transportation network in the prediction and predictability of global epidemics**. *Proc Nat Acad Sci USA* 2006, **103**(7):2015–2020.
- Convertino M: **Neutral metacommunity clustering and SAR: River basin vs. 2-D landscape biodiversity patterns**. *Ecol Model* 2011, **222**(11):1863–1879.
- Convertino M, Valverde LJ Jr: **Portfolio decision analysis framework for value-focused ecosystem management**. *PLoS ONE* 2013, **8**(6):e65056.

- Convertino M, Hedberg C: **Epidemic intelligence cyberinfrastructure: real-time outbreak source detection and prediction for rapid response.** *PLoS Curr Outbreaks* 2014a.
- Convertino M, Hedberg C: **Intelli-food: cyberinfrastructure for real-time outbreak source detection and rapid response.** *Lecture Notes in Computer Science* 2014b, **8549**:181-196. [Smart Health book, Springer].
- Convertino M, Liang S: **Probabilistic supply chain risk model for food safety.** *GRF Davos Planet@Risk* 2014. **2**. [Special Issue on One Health].
- Convertino M, Muneeppeerakul R, Azaze S, Bertuzzo E, Rinaldo A, Rodriguez-Iturbe I: **On neutral metacommunity patterns of river basins at different scales of aggregation.** *Water Resour Res* 2009, **45**(8).
- Convertino M, Munoz-Carpena R, Chu-Agor M, Kiker G, Linkov I: **Untangling drivers of species distributions: Global sensitivity and uncertainty analyses of MaxEnt.** *Environ Model Softw* 2014a, **51**:296-309.
- Convertino M, Foran CM, Keisler JM, Scarlett L, LoSchiavo A, Kiker GA, Linkov I: **Enhanced adaptive management: integrating decision analysis, scenario analysis and environmental modeling for the everglades.** *Sci Rep* 2014c, **2922**.
- Convertino M, Munoz-Carpena R, Kiker G, Perz S: **Design of monitoring networks by value of information optimization: experiment in the Amazon.** *Stochastic Environ Res Risk Assess* 2014d. In press.
- den Broeck WW, Gioannini C, Goncalves B, Quaggiotto M, Colizza V, Vespignani A: **GLeaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale.** *BMC Infect Dis* 2011, **11**.
- Estrada E, Gómez-Gardeñes J: **Communicability reveals a transition to coordinated behavior in multiplex networks.** *Phys Rev E* 2014, **89**(4).
- Ercsey-Ravasz M, Toroczka Z, Lakner Z, Baranyi J: **Complexity of the international agro-food trade network and its impact on food safety.** *PLoS ONE* 2012, **7**(5):e37810. [http://dx.doi.org/10.1371/journal.pone.0037810]
- FAOSTAT: **Food balance sheet website of the food and agriculture organization of the United Nations.** 2014. Tech. rep. [http://faostat.fao.org/site/368/default.aspx#ancor; Accessed April 2014]
- FAS: **USDA foreign agricultural service.** 2014. Tech. rep. [http://www.fas.usda.gov/data; Accessed April 2014]
- GATS: **Multistate outbreak of salmonella bareilly and salmonella nchanga infections associated with a raw scraped ground tuna product.** 2014a. Tech. rep. [http://www.cdc.gov/salmonella/bareilly-04-12/index.html?s\_cid=cs\_654; Accessed April 2014]
- GATS: **USDA global agricultural trade system.** 2014b. Tech. rep. [http://www.fas.usda.gov/gats/default; Accessed April 2014]
- Goncalves B, Balcan D, Vespignani A: **Human mobility and the worldwide impact of intentional localized highly pathogenic virus release.** *Scientific Reports* 2013, **3**(810).
- GRUMP: **Global rural-urban mapping project alpha version population grids.** 2014. Tech. rep. [http://sedac.ciesin.columbia.edu/gpw; Accessed April 2014]
- Guevart E, Noeske J, Solle J, Essomba J, Edjenguele M, Bitá A, Mouangue A, Manga B: **Factors contributing to endemic cholera in Douala, Cameroon.** *Med Trop (Mars)* 2006, **66**(3).
- Haldane AG, May RM: **Systemic risk in banking ecosystems.** *Nature* 2011, **469**(7330):351-355.
- Halpern BS, Longo C, Hardy D, McLeod KL, Samhoury JF, Katona SK, Kleisner K, Lester SE, O'Leary J, Ranelletti M, Rosenberg AA, Scarborough C, Selig ER, Best BD, Brumbaugh DR, Chapin FS, Crowder LB, Daly KL, Doney SC, Elfes C, Fogarty MJ, Gaines SD, Jacobsen KI, Karrer LB, Leslie HM, Neeley E, Pauly D, Polasky S, Ris B, St Martin K, et al: **An index to assess the health and benefits of the global ocean.** *Nature* 2012, **488**:615-620. [http://dx.doi.org/10.1038/nature11397]
- Hanson L, Zahn E, Wild S, Dopfer D, Scott J, Stein C: **Estimating global mortality from potentially foodborne diseases: an analysis using vital registration data.** *Popul Health Metrics* 2012, **10**(5).
- Helbing D: **Globally networked risks and how to respond.** *Nature* 2013, **497**(7447):51-59. [http://dx.doi.org/10.1038/nature12047]
- Helbing D, Brockmann D, Chadefaux T, Donnay K, Blanke U, Woolley-Meza O, Moussaid M, Johansson A, Krause J, Schutte S, et al: **How to save human lives with complexity science.** *J Stat Phys* 2014. SSRN 2390049. DOI 10.1007/s10955-014-1024-9.
- Holmes EE: **Are diffusion models too simple? A comparison with telegraph models of invasion.** *Am Nat* 1993, **142**:403-419.
- Keisler J: **Value of information in portfolio decision analysis.** *Decis Anal* 2004, **1**(3):177-189. [http://da.journal.informs.org/content/1/3/177.abstract]
- Kilpatrick AM, Chmura AA, Gibbons DW, Fleischer RC, Marra PP, Daszak P: **Predicting the global spread of H5N1 avian influenza.** *Proc Nat Acad Sci* 2006, **103**(51):19368-19373.
- Knobler S, Mahmoud A, Lemon S, Pray L, on Microbial Threats EF: *The Impact of Globalization on Infectious Disease Emergence and Control: Exploring the Consequences and Opportunities, Workshop Summary - Forum on Microbial Threats: The National Academies Press*; 2006. [http://www.nap.edu/openbook.php?record\_id=11588]
- Manitz J, Kneib T, Schlather M, Helbing D, Brockmann D: **Origin detection during food-borne disease outbreaks a case study of the 2011 EHEC/HUS outbreak in Germany.** *PLoS Curr Outbreaks* 2014.
- Mundt CC, Sackett KE, Wallace LD, Cowger C, Dudley JP: **Long?distance dispersal and accelerating waves of disease: empirical relationships.** *Am Nat* 2009, **173**(4):456-466. [http://www.jstor.org/stable/10.1086/597220]
- NAS: *Improving food safety through a one health approach: workshop summary.* Washington (DC): National Academies Press (US); 2012. ISBN-13: 978-0-309-25933-0 ISBN-10: 0-309-25933-9.
- Newman M: **The structure and function of complex networks.** *SIAM Rev* 2003, **45**(2):167-256. [http://epubs.siam.org/doi/abs/10.1137/S003614450342480]
- Njoh ME: **The cholera epidemic and barriers to healthy hygiene and sanitation in cameroon: a protocol study.** 2010. Tech. rep., Umea University. [http://bvs.per.paho.org/texcom/colera/MENjoh.pdf]
- On Effectiveness of National Biosurveillance Systems: BioWatch C, the Public Health System NRC: *BioWatch and Public Health Surveillance: Evaluating Systems for the Early Detection of Biological Threats: Abbreviated Version: The National Academies Press*; 2011. [http://www.nap.edu/openbook.php?record\_id=12688]

- Oakley JE: **Decision-theoretic sensitivity analysis for complex computer models.** *Technometrics* 2009, **51**(2):121–129. [http://www.tandfonline.com/doi/abs/10.1198/TECH.2009.0014]
- Pandit A, Crittenden J: **Index of Network Resilience (INR) for Urban Water Distribution Systems.** In *Nature*; 2012. [http://www.tisp.org/index.cfm?cdid=12519&pid=10261]
- Park J, Seager TP, Rao PSC, Convertino M, Linkov I: **Integrating risk and resilience approaches to catastrophe management in engineering systems.** *Risk Anal* 2012, **33**:356–367. doi: 10.1111/j.1539-6924.2012.01885.x.
- Pastore Y Piontti, A, Gomes Da Costa, M, Samay N, Perra N, Vespignani A: **The infection tree of global epidemics.** *Netw Sci* 2014, **2**:132–137. [http://journals.cambridge.org/article\_S2050124214000058]
- Pinto PC, Thiran P, Vetterli M: **Locating the source of diffusion in large-scale networks.** *Phys Rev Lett* 2012, **109**:068702. [http://link.aps.org/doi/10.1103/PhysRevLett.109.068702]
- Saltelli A, Ratto TM, Andres, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S: *Global Sensitivity Analysis - The Primer.* John Wiley and Sons; 2008.
- Schlundt J, Toyofuku H, Jansen J, Herbst S: **Emerging food-borne zoonoses.** *Rev Sci Tech* 2004, **23**(2).
- Scallan E, Griffin P, Angulo F, Tauxe R, Hoekstra R: **Foodborne illness acquired in the United states-Unspecified agents.** *Emerg Infect Dis* 2011a, **17**.
- Scallan E, Hoekstra R, Angulo F, Tauxe R, Widdowson S, Roy J, Jones P, Griffin P: **Foodborne illness acquired in the United StatesMajor Pathogens.** *Emerg Infect Dis* 2011b, **17**(7).
- Simini F, González MC, Maritan A, Barabási AL: **A universal model for mobility and migration patterns.** *Nature* 2012, **484**(7392):96–100. [http://dx.doi.org/10.1038/nature10856]
- Sobol I: **Sensitivity analysis for non-linear mathematical model.** *Math Comput Simul* 1993:407–414.
- Sobol I: **Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates.** *Math Comput Simul* 2001, **55**:271–280.
- Stachenko S: **The role of surveillance and data use in the development of public health policies.** *Promot Educ* 2008, **15**(3):27–29. [http://dx.doi.org/10.1177/1025382308095654]
- Trainor-Guitton WJ, Mukerji T, Knight R: **A methodology for quantifying the value of spatial information for dynamic Earth problems.** *Stochastic Environ Res Risk Assess* 2012:1–15. [http://dx.doi.org/10.1007/s00477-012-0619-4]
- Tatah A, Pulcherie K, Mande N, Akum N: **Investigation of water sources as reservoirs of Vibrio cholerae in Bepanda, Douala and determination of physico-chemical factors maintaining its endemicity.** *Onderstepoort J Vet Res* 2012, **79**(2).

doi:10.1186/s40294-014-0006-8

**Cite this article as:** Convertino et al.: Optimal surveillance network design: a value of information model. *Complex Adaptive Systems Modeling* 2014 **2**:6.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)