

AUTOMATIC GENERATION OF GLOBAL AGENT-BASED MODEL OF MIGRATORY WATERFOWL FOR EPIDEMIOLOGICAL ANALYSIS OF AVIAN INFLUENZA AND OTHER ZOONOTIC DISEASES

Dhananjai M. Rao
CSE Department
Miami University
Oxford, OH 45056, USA.
email: raodm@miamioh.edu

Alexander Chernyakhovsky
Massachusetts Institute of Technology
Cambridge, MA 02139, USA.
email: achernya@mit.edu

KEYWORDS

Migratory Flyways, Tessalation, Agent-based Modeling, Simulation, Epidemiology, Avian Influenza (H5N1)

ABSTRACT

Seasonal migration of waterfowl, in which avian influenza viruses are enzootic, plays a strong role in the ecology of the disease and has been implicated in several zoonotic epidemics and pandemics. Recent investigations have established that with just 1 mutation current avian influenza viral strains gain the ability to be readily transmitted between humans. These investigations further motivate the need for detailed analysis, in addition to satellite surveillance, of migratory patterns and its influence on the ecology of the disease to aid design and assessment of prophylaxis and containment strategies for emergent epidemics. Accordingly, this paper proposes a novel methodology for generating a global agent-based stochastic epidemiological model involving detailed migratory patterns of waterfowl. The methodology transforms Geographic Information Systems (GIS) data containing global distribution of various species of waterfowl to generate metapopulation for agents that model collocated flocks of birds. Generic migratory flyways are suitably adapted to model migratory flyways for each waterfowl metapopulation. Migratory characteristics of various species are used to determine temporal attributes for the flyways. The resulting data is generated in XML format compatible with our simulation-based epidemiological analysis environment called SEARUMS. Case studies conducted using SEARUMS and the generated models for high-risk waterfowl species indicate good correlation between simulated and observed viral dispersion patterns, demonstrating the effectiveness of the proposed methodology.

INTRODUCTION

Humanity continues to face a multitude of global socio-economic challenges due to annual epidemics and punctuated pandemics of highly virulent zoonoses such as avian influenza (H5N1, H7N9) and the 2009 swine flu (H1N1)

pandemic. The 2009 swine flu (H1N1) pandemic virus involved segments from avian serotype (Trifonov et al. 2009). Moreover, these diseases are of global importance because less virulent forms of the viruses annually affect 90 million people, causing 500,000 global human fatalities and billions of dollars of annual losses due to recurrent epidemics in poultry.

Importance of Ecological Analysis of Avian Influenza: Avian influenza outbreaks in humans routinely occur in several parts of the world as reported by the World Health Organization (WHO 2013) and as shown in Figure 1. The outbreaks have been sporadic and unsustainable. However, several recent investigations have established that with just 1 mutation the H5N1 virus gains the ability to be readily transmitted between humans and consequently a global pandemic is imminent (Tharakaraman et al. 2013). Their inference is also corroborated by Smith et al. (2004) who use antigenic cartography to demonstrate that the antigenic evolution of influenza A viruses is punctuated: there are three to eight years of relative stasis followed by rapid antigenic change. Therefore, there is a strong probability that the current highly pathogenic strains of H5N1 viruses isolated since 2004, whose human-to-human transmis-

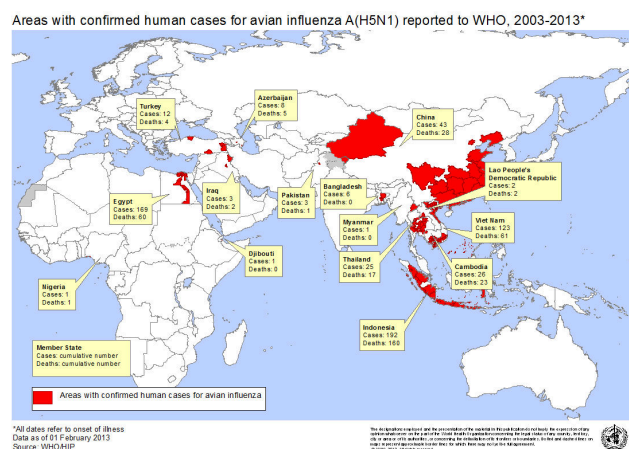


Figure 1: Avian influenza (H5N1) outbreaks in humans. Image obtained from WHO (2013) (URL in citation)

sions have been unsustainable, could potentially mutate resulting in a pandemic. Consequently, there is heightened urgency to shift the focus of investigations from studying possibilities to analyzing probabilities of outbreaks to proactively mitigate or even preempt emergent pandemics (Rappuoli and Dormitzer 2012) in contrast to delayed responses to the 2009 H1N1 pandemic as discussed in the report by the executive office of the president of the United States (President’s Council of Advisors on Science and Technology 2010).

Need for analysis of migratory waterfowl: The global disease ecology of the diseases, specifically avian influenza, is summarized in Figure 3. Various polymorphic strains of Influenza A viruses with different antigenic characteristics are endemic and enzootic in different species but vary in different parts of the world (Russell et al. 2008). It has been established that migratory waterfowl play a central role in the global ecology of avian viruses (USDA 2013). Moreover, migratory waterfowl have been implicated as natural reservoirs, mixing vessels, and intercontinental vectors for various serotypes of avian viruses (USDA 2013). However, the degree and extent of the influence of waterfowl in global disease propagation is poorly understood and is an active area of research (Gilbert et al. 2006; USDA 2013). Accordingly, several multinational surveillance efforts have been initiated to collate data on various characteristics of migratory species (USDA 2013). The surveillance efforts include: large scale satellite tracking, banding and tracking of individual birds in conjunction with satellite telemetry, and various biological and bioinformatics sampling and cataloging efforts (USDA 2013).

Motivation for simulation-based analysis: Despite the advancements in technologies and improvements in economies of scale, data from surveillance has significant limitations. Specifically, the data will continue to remain relatively coarse and sparse. Furthermore, field observations and satellite telemetry only provide a snapshot of various natural processes that influence global ecology of the disease. Consequently, the surveillance data is typically combined with computational analysis methods such as simulations to generate com-

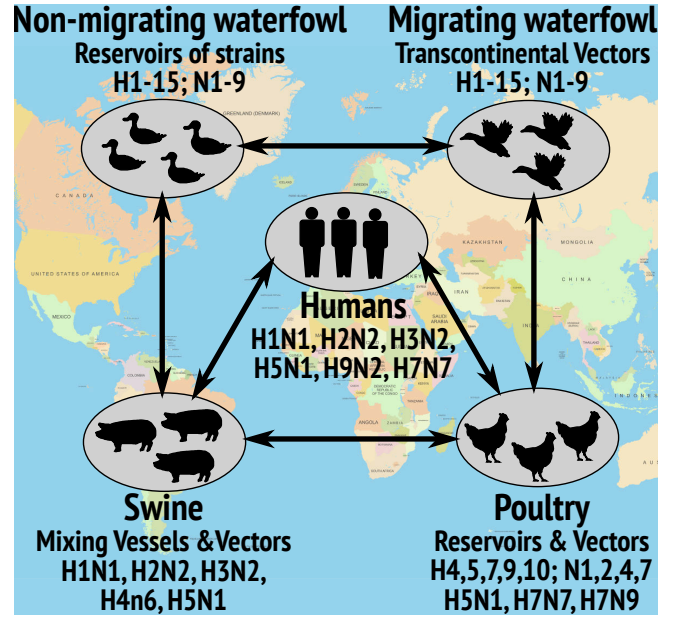
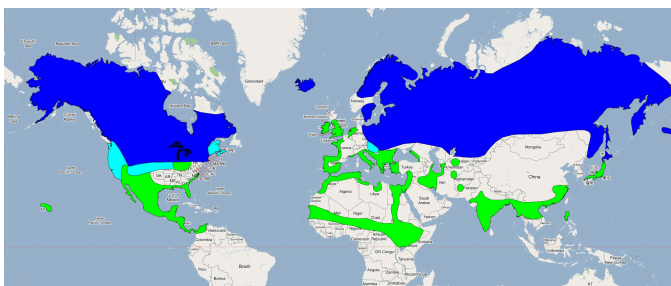


Figure 3: Overview of Global Ecology of Avian Influenza: Viral serotypes, hosts, and vectors

prehensive, multifaceted information and draw actionable inferences. Computer-based simulation, specifically discrete time simulation, is the dominant method for epidemiological analysis of various diseases and simulations play a pivotal role in design and implementation of national and multinational prophylactic strategies and policies.

Proposed Research & Paper Outline

Motivated the foregoing discussions, this research proposes a novel methodology for utilizing aggregate satellite telemetry data on various waterfowl species for conducting detailed and comprehensive simulation-based epidemiological analyses to assess the role of migratory species in intercontinental spread and ecology of avian influenza. Specifically, the methodology transforms Geographic Information Systems (GIS) data con-



(a) Raw GIS data from GROMS



(b) Metapopulations from collocated flocks

Figure 2: Overview of first phase in GIS data processing. Sub-figure (a): Example of an aggregate GIS data from the GROMS database. Sub-figure (b): The generated spatially explicit metapopulations for agent-based simulation.

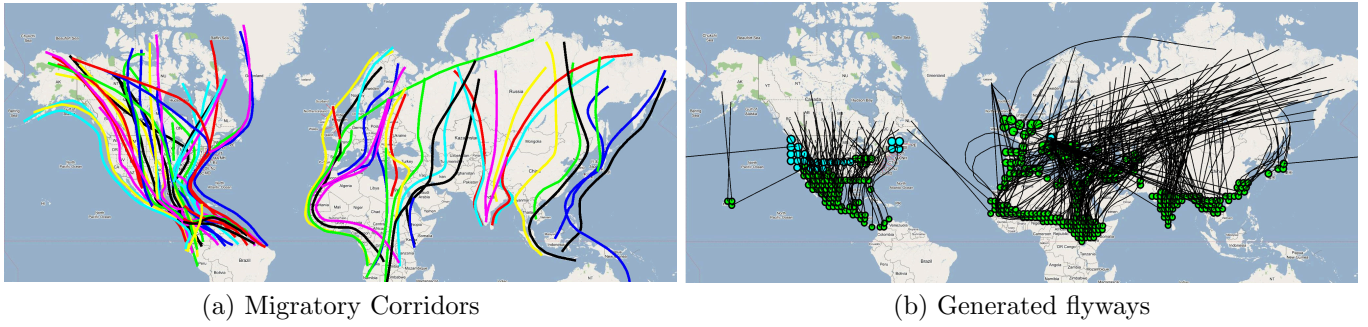


Figure 4: Overview of second phase of GIS data processing. Sub-figure (a): Generic migratory corridors (colored for clarity; colors not significant). Sub-figure (b): Flyways generated by adapting generic migratory corridors.

taining global distribution of various species of waterfowl to generate metapopulation for agents that model collocated flocks of birds. Figure 2(a) shows an example of the GIS data for *Anas acuta* (Northern Pintail) from the Global Register of Migratory Species (GROMS) database (GROMS 2013). Figure 2(b) shows an example of spatially explicit metapopulations generated for agent-based simulation. A more detailed discussion on the methodology for generating metapopulations is discussed in Section Method.

Migratory corridors for various species have been recorded using various surveillance methods and are also supplied as GIS data as shown in Figure 4(a). The generic migratory corridors are suitably adapted to model migratory flyways for each waterfowl metapopulation as shown in Figure 4(b). Migratory characteristics of various species are used to determine temporal attributes for the flyways. The Method section discusses the details involved in generating flyways while Section Related Work contrasts the proposed research with other closely related investigations.

In this investigation the models are generated in XML format compatible with our simulation-based epidemiological analysis environment called SEARUMS (Rao et al. 2009) summarized in Section SEARUMS. SEARUMS is used to conduct various epidemiological simulations, visualize analytics from the simulations, and draw suitable inferences. In this study, SEARUMS has been used to conduct case study using *high risk* waterfowl species for verification and validation (V&V) of the generated model and proposed methodology, as discussed in Section Experiments. Section Conclusions concludes the paper by summarizes the findings and discussing ongoing efforts.

RELATED RESEARCH

The proposed investigations aim to automatically generate spatially-explicit model for migratory waterfowl with application to epidemiological analysis of avian influenza. Accordingly, this section presents a literature survey of closely related investigations. Kilpatrick et al.

(2006) analyze the global spread of H5N1 using phylogenetic relationships between virus isolates, waterfowl migration patterns, and poultry trade. In their model, waterfowl migration flyways between summer and wintering zones are indirectly used to determine the regions through which the waterfowl fly. In contrast, in the proposed methodology, waterfowl migration patterns are explicitly simulated in detail to identify areas of infections.

Lebarbenchon et al. (2009) discuss the use of individual-based spatially explicit model, in which space was divided into squared meshes of 50x50 km, to discuss the role of *Anas crecca* (Common Teal) in spreading avian influenza in Europe. Movement of birds between meshes was modeled using random distances with destinations based on flight distances and preference determined from surveillance data. Unlike their research, the proposed research focuses on multiple bird species involving intercontinental migration.

Takekawa et al. (2010) analyze East Asian flyway in the context of migrations between Poyang and Qinghai lakes to identify any temporospatial relationship between H5N1 outbreaks and the specific migratory corridor. Fuller et al. (2013) utilize statistical analysis of surveillance data on H5N1 outbreaks in poultry and a multivariate regression model to predict regions where human and avian viral serotypes could potentially reassort to give rise to pandemic strains. These two investigations uses surveillance-based approaches to analyze specific regions while this paper proposes the use of modeling and simulation for analyzing global ecology of migratory waterfowl and consequent spread of etiological agents of zoonoses.

SEARUMS

The verification and validation of the automatic model generation methodology proposed in this investigation has been conducted using an Eco-modeling and biosimulation environment called SEARUMS (<http://www.searums.org/>) (Rao et al. 2009). Currently, it is optimized to enable study and analysis of global epi-

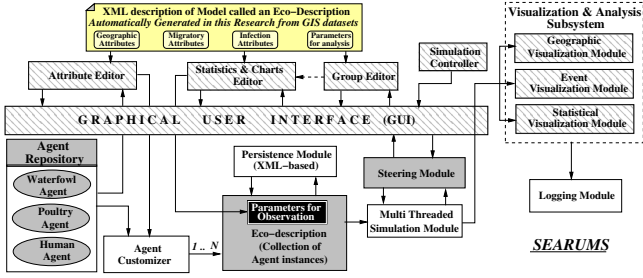


Figure 5: Architectural overview of SEARUMS

demology of avian influenza. However, the design of its modeling and simulation framework is sufficiently generic and can be adapted for analysis of other diseases. SEARUMS is designed to be a user friendly, integrated, graphical modeling, simulation, visualization, and analysis environment for conducting epidemiological analysis. These design goals have been achieved by composing the system using a collection of interdependent but loosely coupled modules as shown in Figure 5.

Each module shown in Figure 5 has a well defined functionality that can be accessed and utilized via a set of Application Program Interface (API) method calls. This approach permits seamless “plug and play” of modules and the environment is composed by loading suitable modules dynamically on-demand via Java’s reflection API. A comprehensive and intuitive modeling, visualization, and steering subsystems ease user interactions with the system.

The models for simulation is supplied to SEARUMS are called Eco-descriptions. This investigation aims to automatically generate Eco-descriptions in XML compatible with SEARUMS. The generated XML can be converted to other schema definitions such as the Ecological Metadata Language (EML) (KNB 2013). The primary information encapsulated by the Eco-description are utilized by various agents in SEARUMS. Each agent has its own behavior that is parameterized by data from the Eco-description to reflect various characteristics of its real-world counterpart. The current investigation specifically utilizes the *Waterfowl Agent* shown in Figure 5.

The agents in SEARUMS implement the bio-mathematical compartmental models that are widely used in epidemiology (Brauer and Castillo-Chavez 2013). In a compartmentalized model the population being analyzed is partitioned into non-intersecting subsets called *compartments*. Compartments are defined such that the sub-population within a compartment exhibits a vital disease characteristic, such as: Susceptible (S), Exposed (E), Infected (I), and Recovered (R). The epidemiological characteristics of the classical *SEIR* model is modeled using the following system of differential equations:

$$dS/dt = \mu N - [\lambda + \mu]S(t) \quad (1)$$

$$dE/dt = \lambda S(t) - (\beta + \mu)E(t) \quad (2)$$

$$dI/dt = \beta E(t) - (v + \mu)I(t) \quad (3)$$

$$dR/dt = vI(t) - \mu R(t) \quad (4)$$

The constants μ , λ , β , and v represent the birth/death rate, the force of infection, latency period, and recover rates respectively. These constants are determined based on the characteristics of the disease being modeled and are supplied via the Eco-description. In SEARUMS spread of infection to various agents occurs when agents overlap with each other. The system is modeled as discrete time Markov processes and the **simulation module** performs the tasks associated with Discrete Event Simulation (DES) using the Eco-description. This module utilizes a multi-threaded DES kernel that manages and schedules the discrete events generated by the Agents. Various phenomena occurring during simulation are logged for further analysis. A more detailed description of SEARUMS is available in the literature (Rao et al. 2009).

MODEL GENERATION METHODOLOGY

In continuation with introductory discussion in Section Introduction, this section presents details on the proposed methodology for generating spatially-explicit, agent-based model of migratory waterfowl from aggregate, high resolution Geographic Information System (GIS) data. The GIS data set used in this investigation were obtained from the Global Registry of Migratory Species (GROMS) (GROMS 2013). GROMS was initiated by the UNEP/CMS Secretariat to help mitigate deficiency in estimation of migratory animals. It is supported by the German Federal Agency for Nature Conservation and operates in cooperation with the Zoological Research Museum in Bonn. The GROMS database provides GIS data for hundreds of migratory species including the *Anatidae* family which includes various waterfowl.

The GIS data for each species is organized into multiple files that essentially provide information about a collection of *rings* that identify the area where birds were detected using various surveillance methods. The GIS data is loaded and processed using the open source Geospatial Data Abstraction Library (GDAL). GDAL is a translator library for raster geospatial data formats that is maintained by the Open Source Geospatial Foundation. As a library, it presents a single abstract data model to the calling application for all supported formats. Figure 6 shows a raw GROMS GIS data consisting of 32 distinct rings. Each ring has a unique identifier associated with it which is used to reference other rings and associated metadata files. Each ring is

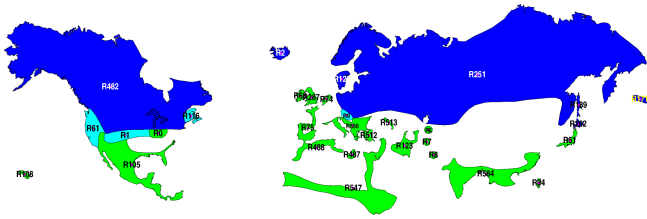


Figure 6: Sample data from GROMS illustrating rings in GIS data

essentially a closed polygon with added metadata describing its salient characteristics. The vertexes of the polygon that constitute a ring are defined by their actual latitude and longitude values that have been plotted in Figure 6 using a standard Mercator projection which tends to exaggerate areas closer to poles. The rings can either be inclusion or exclusion rings. Exclusion rings indicate that the area circumscribed by them are not part of another ring associated with them. An exclusion ring, identified by shaded region, associated with inclusion ring R482 is shown in Figure 6.

The metadata associated with each ring provides additional semantic information about the rings. Specifically, for GROMS, the rings are classified into three distinct regions, namely: summering, wintering, and all-year zones, shown in green, blue, and cyan colors respectively in Figure 6. As the terms suggest, summering zones are regions occupied by the species between April to August while wintering zones are occupied at other times to take advantage of food and nesting locations. The all-year zones are areas where the species are present throughout the year with some migratory movements in the flocks. The population of various species and other migratory characteristics such as distances traveled, speed of migration, and times of migration is obtained from BirdLife International Database (BID) (BirdLife International 2012).

Phase 1: Flock Generation

Accordingly, the process of generating spatially-explicit metapopulations for a given species commences with specification of various characteristics including GIS data from GROMS and migratory information from BID targeted for the summer or winter season as shown in Figure 7. The first phase of processing involves creation of flock metapopulations in the appropriate inclusion rings for the given season. In addition, “dummy” flocks shown in Figure 7 are also generated in the opposite season rings to ease migration path generation in the second phase. The flock metapopulations are generated as overlapping circular regions of a given radius with even density of birds. Representation of metapopulations as circles is commonly used in spatially explicit ecological models to balance computational costs versus resolution of model for various analysis (Law et al.

2005; Rao et al. 2009). Smaller circular regions provide increased resolution or level of details at the expense of increased computational cost. The resolution of flock metapopulations can vary depending on the objectives of simulation-based analysis. For example, if the objective is to analyze epidemiological details in East Asia, then species in that region can be generated at high resolution while other species can be modeled at lower resolution providing a better balance between computation overheads and resolution.

The flocks are generated in the form of a grid, similar to semi-regular tessellation, with a specified area of overlap between adjacent flocks. The overlap provides a convenient mechanism for propagating infections between metapopulations in a given area. In order to provide a consistent distribution of flocks between migratory zones, the radius of the flocks between rings are slightly adjusted based on the total surface area associated with the seasons. The population of birds specified as primary inputs (see Figure 7) is evenly distributed across all the actual flocks.

Phase 2: Migratory Flyway Generation

The second phase of the methodology involves generation of migratory flyways for the set of flocks generated in the first phase. In this phase, the generic migratory corridors shown in Figure 4(a) is loaded from GROMS GIS data. Unlike, the rings, flyways are stored as *arcs* in GIS data. However, similar to rings, arcs are represented by a collection of points whose coordinates are latitude and longitude values. Figure 7 presents an overview of the operations in this phase. The operations in this phase can be broadly classified into two major steps. The first step involves selection of a target dummy flock to which a given flock is set to migrate. Once the target dummy flock is identified, the next step involves generation of the flyways.

The first step of migratory flyways generation involves pairing a flock to a dummy flock in the destination zone. A dummy flock is selected by repeatedly applying the following set of rules in an attempt to perform best effort pairing to reflect migratory characteristics stipulated for model generation. This step commences with selection of candidate rings corresponding to target seasonal zones within stipulated migratory distance that has unpaired dummy flocks. This in step involves checks for both transatlantic and transpacific migratory distances to identify closest rings. If multiple rings are detected within migratory distance, then a round robin approach is used to select a candidate target ring. Once a candidate ring has been identified, random selection of a valid dummy flock within migratory distance is attempted. If selection of a candidate ring or target dummy flock fails, then a specified acceptable additional distance is added to migratory distance to identify a suitable ring. The process is repeated with increasing radius of migratory

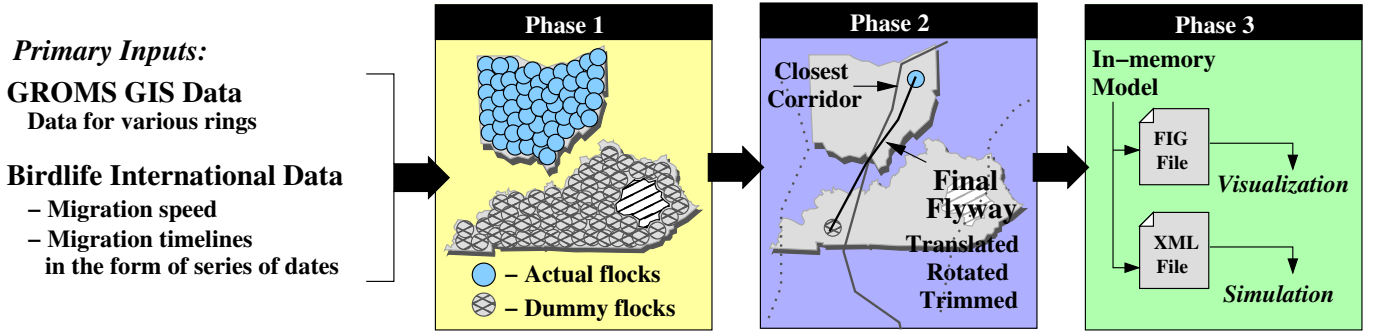


Figure 7: Overview of the three phased approach for generating flocks and migratory flyways from input data

distances to pair a flock with a suitable migratory destination. However, if the process does not yield a suitable ring, primarily due to slight differences between number of flocks in migratory zones, then a destination ring and dummy flock is chosen at random from the set of valid candidates.

Once a target dummy flock has been identified, the next step in this phase involves adapting a putative migratory corridor shown in Figure 4(a) to generate flyway between a flock to its migratory target as summarized in Figure 7. First, an attempt is made to identify a corridor that is closest to both the flock and its target based on their latitude and longitude values. This operation is performed using by identifying the nearest vertex on each corridor to both the source and target locations. Corridors that are farther than a given threshold are ignored and the closet corridor is identified. Next, the corridor is translated and logically anchored to either the source or destination flock depending on which ever one is the nearest. Next, the corridor is rotated around the anchor location until some point on the corridor is sufficiently close to the other location. Finally, a complete flyway is generated by trimming the corridor by removing unnecessary points and adding the source and target points to it.

If a suitable candidate corridor is not identified, then the methodology reverts to generating a direct flyway between the flock and its target location. Once flyways are generated, temporal attributes for the flyways are generated using species specific migratory characteristics specified as primary inputs. The average migratory distance and speed are used to set intermediate stopover locations on the flyways. Since migratory birds demonstrate amazing flyway consistency and site fidelity (A. Iverson and Esler 2006; Ketterson and Nolan 1990), the same flyway but with different temporal attributes is used for modeling the reverse direction of migratory movements.

Phase 3: Data Persistence

The first two phases of the proposed methodology are applied to generate flocks and migratory flyways which

are stored using various data structures in memory. This phase deals with persisting the generated model in various formats for visualization and simulation. Visualization of the generated models (as shown in Figure 4(b)) is accomplished in two steps. The model is persisted in a FIG file format, which provides a convenient textual representation of geometric shapes along with images. The textual FIG file format is then converted into various raster formats such as PNG, JPEG, or TIFF file formats for visualization.

Model data for simulation is persisted in XML using the schema compatible with SEARUMS. The generated XML can then be converted to other formats such as Ecological Metadata Language (EML) (KNB 2013) or binary file formats. This phase also involves combining data for multiple species into a single comprehensive model for simulation. The generated models are used for conducting epidemiological simulations and analysis using SEARUMS as discussed in the next section. The primary objective of the simulations in this study is to verify validity of the generated models.

EXPERIMENTS

This section discusses the experiments conducted to verify and validate the effectiveness of the automatic model generation methodology via hypothesis testing is presented in Method section. The hypothesis for the experiment is that the automatically generated model will reflect spread of highly pathogenic H5N1 strain, originally isolated in Southeast Asia, to various *human outbreak* regions as reported by the World Health Organization (WHO) (WHO 2013) (see Figure 1). Testing this hypothesis will accomplish verification and validation of the model and thereby the automatic model generation methodology. Furthermore, it adds further evidence to support the role of migratory species in intercontinental spread of H5N1.

Accordingly, for hypothesis testing, simulation-based epidemiological outbreak analysis was conducted. The analyses were conducted using a characteristic model generated from 22 species of high risk birds from which H5N1 is enzootic as viral strains have been consistently

Species Name	Population	#Flocks / #Agents
Aix sponsa	3500000	64
Amazonetta brasiliensis	110000	103
Anas acuta	5300000	372
Anas bahamensis	640000	78
Anas platalea	500000	47
Anas platyrhynchos	19000000	557
Anas sibilatrix	250000	30
Anas versicolor	126000	42
Anser anser	1000	4
Anser indicus	56000	11
Aythya ferina	2200000	213
Aythya fuligula	2600000	148
Aythya marila	1200000	114
Branta canadensis	5500000	169
Callonetta leucophrys	50000	3
Cygnus melanocoryphus	50000	32
Melanitta nigra	2100000	96
Mergellus albellus	130000	71
Netta peposaca	1000000	26
Anas Crecca	5900000	403
Philomachus pugnax	4,200,000	210
Porzana pusilla	21300	262

Table 1: List of high risk waterfowl species used for model generation

isolated in the past as reported by Gilbert et al. (2006). Table 1 lists the high risk waterfowl and the population of birds modeled in this study.

The automatic model generation methodology discussed in Section Method was used for generating models for all 27 high risk species. The flocks in Table 1 were generated using a radius of 100 miles with an overlap of 10%. A larger flock size was favored based on the average size of high risk regions reported by WHO. The GIS data for the various species were obtained from GROMS database GROMS (2013) while migra-

tory characteristics were obtained from BirdLife International Database (BID) BirdLife International (2012). The XML model description for various species were merged together to obtain the final characteristic model shown in Figure 8(a). The various species in the model have been color coded. The initial starting zone was set to correspond to the wintering zone. The position of various flocks after migration to summering zones is shown in Figure 8(b). The migratory pathways are not shown in Figure 8 to improve clarity. Readers are referred to online information available at <http://www.searums.org/esm13/> for the complete model, high resolution images, and videos illustrating the migration and infection spreads discussed in this paper.

Metamorphic Testing

The aggregate model shown in Figure 8 consisting of all 22 species listed in Table 1 was constructed in an incremental fashion by introducing one species at a time and verifying the resulting model using metamorphic testing. Metamorphic testing uses well defined relationships between a given input parameter and output such that the outcomes can be tested (Pullum and Ozmen 2012; Ramanathan et al. 2012). Metamorphic testing has been used for V&V of epidemiological models similar to the agent-based approach used in this investigation Pullum and Ozmen (2012); Ramanathan et al. (2012). Specifically, each time a new species was introduced, metamorphic approach was used to verify propagation of infection occurred as expected verifying fundamental epidemiological aspects of the generated model.

Hypothesis Testing

Testing of the hypothesis, that the model shown in Figure 8 will reflect the spread of highly pathogenic H5N1 strain from Southeast Asia to regions with human outbreaks as reported by WHO, was conducted after metamorphic testing. The experiments for hypothesis test-

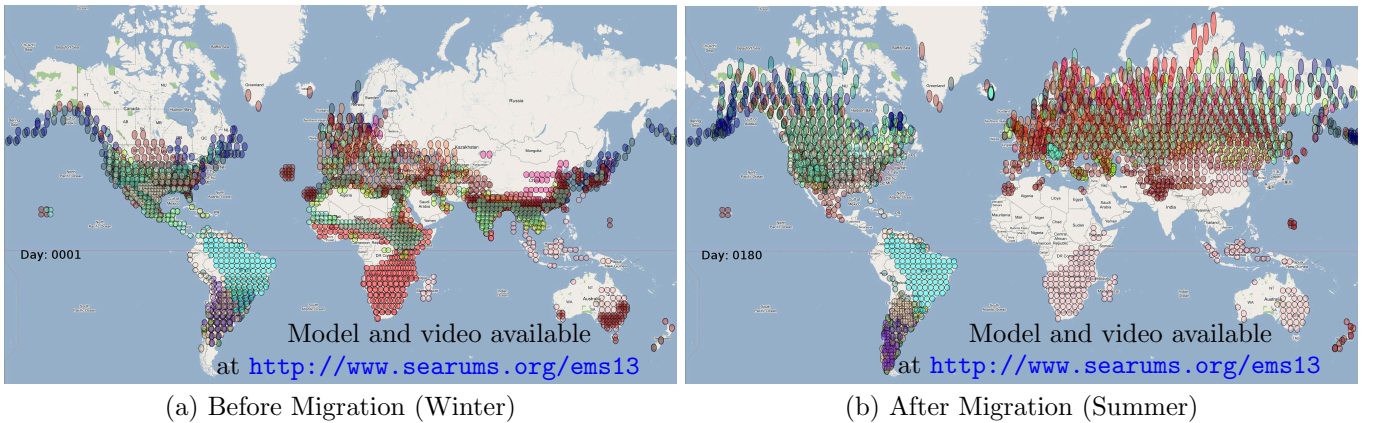


Figure 8: Location of flocks of various species in their wintering and summering zones

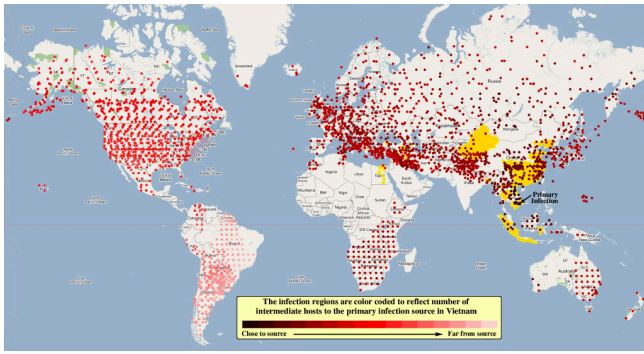


Figure 9: Dominant outbreaks from multiple simulations with high risk areas highlighted from Figure 1.

ing were conducted by seeding a single infection in one waterfowl flock in Vietnam (at 11°57,71' 50" N 108°6' 31.31"). The basic reproductive number for the infection (R_0) was set to be greater than 1 to model enzootic nature of the infection. Furthermore, the disease transmission parameters were configured to reflect a Susceptible → Infected (SI) type compartmental epidemiological model. The model was simulated for a period of two years while logging the locations of various infections occurring in the model. The stochastic nature of the simulations require the use of a Monte Carlo approach in which infections consistently occurring in multiple simulations are identified as the dominant set of infections. The dominant outbreaks identified from multiple stochastic simulations are plotted in Figure 9. The infection locations indicate potential areas for outbreaks and increased density of outbreaks in turn increase the probability of human outbreaks. The infection locations in Figure 9 are color coded to reflect the number of intermediate hosts to the source infection in Vietnam.

As illustrated by Figure 9, the infection is carried by infected waterfowl and radiate out of Vietnam to various parts of Asia and rest of the world. The high risk areas as reported by WHO (WHO 2013) and shown in Figure 1 are highlighted in bright orange in Figure 9. The figure highlights the strong correlation between dense, strong outbreaks observed in the simulation and the regions with human outbreaks reported by WHO. This observation provides strong evidence for increased confidence in the validity of the model and in turn the automatic model generation methodology proposed in this paper.

Furthermore, as shown in Figure 9, infections into North America involve a large number of intermediate hosts suggesting that the infections are weak. However, the primary entry locations near the Gulf of Alaska, a known high risk area (Wilson et al. 2007), consistently show sufficiently strong infections. This observation is supported by other investigations and consequently further increases confidence in the proposed methodology. The infections into South America are propagated from

North American infections and are significantly distant from the source infections in Vietnam. Infections into South America occurred only in the second year of migration cycle. Analytics from the simulation suggest that the viral strains in South America lag behind primary infections in Southeast Asia by more than a year. A similar inference was drawn by Russell et al. (2008) who use bioinformatics-based analysis of influenza A viruses to show more than a year's lag between viruses in Southeast Asia and South America.

These experiments provide evidence to support and accept the hypothesis that the automatically generated model will reflect spread of highly pathogenic H5N1 strain, originally isolated in Southeast Asia, to various human outbreak regions as reported by the World Health Organization (WHO 2013) and shown in Figure 1. More importantly, the experiments significantly increase confidence in the validity of the model as well as verify and validate the proposed model generation methodology. Furthermore, the experiments add credence to the significance of migratory waterfowl in the global ecology of avian influenza.

CONCLUSIONS

This paper discussed a novel methodology for automatic generation of spatially-explicit models of migrating waterfowl for agent-based epidemiological modeling and simulation-based analysis. The methodology generates metapopulation of collocated flocks from aggregate Geographic Information System (GIS) data obtained from GROMS (GROMS 2013). Migratory flyways for each flock are generated from putative migratory corridors from the GROMS database. Temporal characteristics for the flyways are generated using data from BirdLife International database (BirdLife International 2012). The proposed model generation methodology was used to generate a model consisting of high risk species of migrating waterfowl. Metamorphic testing was used to incrementally verify the model. Next, a case study to analyze the spread of infection from Southeast Asia was conducted. The results from the case study showed that the generated model yields analytics that is consistent with prior investigations:

1. The locations infections observed in the simulation are correlated to regions of avian influenza outbreaks in humans as reported by WHO (WHO 2013).
2. Consistent with surveillance reports (Wilson et al. 2007; USDA 2013), the model correctly highlights the region around Gulf of Alaska as a primary gateway for infections into North America and subsequently into South America.
3. The models shows more than a year's lag between outbreaks in South America and Southeast Asia, which is an observation that has been confirmed by

several investigations (Russell et al. 2008).

The experiments used for hypothesis testing provide evidence to accept the hypothesis and significantly increase confidence in the validity of the model and the proposed model generation methodology. The validated methodology can now be used for other simulation-based epidemiological and ecological analyses. Specifically, such simulation-based analyses provides a powerful approach for design and assessment of various prophylactic and containment strategies. Furthermore, outbreak information can be used to focusing surveillance efforts. Simulation-based observations about influence of other migratory waterfowl species on the spread of infections can be used to identify other influential or high-risk species. This investigation used avian influenza as the primary case study for verification and validation of the proposed approach. However, the same methodology can be applied analyze ecological impact of other migratory species. This investigation provides support that the proposed methodology is an effective approach for using surveillance-based GIS datasets for a broad range of simulation-based epidemiological and ecological analysis to combat emergent zoonotic diseases.

REFERENCES

- A. Iverson S. and Esler D., 2006. *Site fidelity and the demographic implications of winter movements by a migratory bird, the harlequin duck *Histrionicus histrionicus**. *Journal of Avian Biology*, 37, no. 3, 219–228. ISSN 1600-048X. doi:10.1111/j.2006.0908-8857.03616.x. URL <http://dx.doi.org/10.1111/j.2006.0908-8857.03616.x>.
- BirdLife International, 2012. *BirdLife Data Zone*. URL <http://www.birdlife.org/datazone/home>.
- Brauer F. and Castillo-Chavez C., 2013. *Mathematical Models for Communicable Diseases*. SIAM, 3600 Market Street, Philadelphia, PA 19104-2688, USA. ISBN 978-1-611972-41-2.
- Fuller T.L.; Gilbert M.; Martin V.; Cappelle J.; Hosseini P.; Njabo K.Y.; Aziz S.A.; Xiao X.; Daszak P.; and Smith T.B., 2013. *Predicting Hotspots for Influenza Virus Reassortment*. *Emerging Infectious Diseases*, 19, no. 4. doi:10.3201/eid1904.120903.
- Gilbert M.; Xiao X.; Domenech J.; Lubroth J.; Martin V.; and Slingenberg J., 2006. *Anatidae Migration in the Western Palearctic and Spread of Highly Pathogenic Avian Influenza H5N1 Virus*. *Emerging Infectious Diseases*, 12, no. 11. doi:10.3201/eid1211.060223.
- GROMS, 2013. *Global Register of Migratory Species (GROMS): Summarising Knowledge about Migratory Species for Conservation*. URL <http://groms.gbif.org/>.
- Ketterson E. and Nolan V. J., 1990. *Site Attachment and Site Fidelity in Migratory Birds: Experimental Evidence from the Field and Analogies from Neurobiology*. In E. Gwinner (Ed.), *Bird Migration*, Springer Berlin Heidelberg. ISBN 978-3-642-74544-7, 117–129. doi:10.1007/978-3-642-74542-3_8. URL http://dx.doi.org/10.1007/978-3-642-74542-3_8.
- Kilpatrick A.M.; Chmura A.A.; Gibbons D.W.; Fleischer R.C.; Marra P.P.; and Daszak P., 2006. *Predicting the global spread of H5N1 avian influenza*. *Proceedings of the National Academy of Sciences*, 103, no. 51, 19368–19373. doi:10.1073/pnas.0609227103.
- KNB, 2013. *Ecological Metadata Language (EML)*. URL <http://knb.ecoinformatics.org/software/eml>.
- Law R.; Dieckmann U.; and Metz J.A., 2005. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*. Cambridge University Press, 32 Avenue of the Americas, New York, NY 10013-2473, USA. ISBN 9780521022095.
- Lebarbenchon C.; Albespy F.; Brochet A.L.; Grandhomme V.; Renaud F.; Fritz H.; Green A.J.; Thomas F.; van der Werf S.; Aubry P.; Guillemain M.; and Gauthier-Clerc M., 2009. *Spread of Avian Influenza Viruses by Common Teal (*Anas crecca*) in Europe*. *PLOS One*, 4, no. 10, e7289. doi:doi:10.1371/journal.pone.0007289.
- Normile D., 2006. *AVIAN INFLUENZA: Evidence Points to Migratory Birds in H5N1 Spread*. *Science*, 311, no. 5765, 1225. doi:10.1126/science.311.5765.1225.
- President’s Council of Advisors on Science and Technology, 2010. *Report To The President on Engineering The Influenza Vaccine Production Enterprise to Meet The Challenges of Pandemic Influenza*. URL <http://www.whitehouse.gov/sites/default/files/microsites/ostp/Vaccinology-Backgrounder.pdf>.
- Pullum L.L. and Ozmen O., 2012. *Early Results from Metamorphic Testing of Epidemiological Models*. In *BioMedical Computing (BioMedCom), 2012 ASE/IEEE International Conference on*. 62–67. doi:10.1109/BioMedCom.2012.17.
- Ramanathan A.; Steed C.A.; and Pullum L.L., 2012. *Verification of Compartmental Epidemiological Models Using Metamorphic Testing, Model Checking and Visual Analytics*. In *BioMedical Computing (BioMedCom), 2012 ASE/IEEE International Conference on*. 68–73. doi:10.1109/BioMedCom.2012.18.

- Rao D.M.; Chernyakhovsky A.; and Rao V., 2009. *Modeling and analysis of global epidemiology of avian influenza*. *Environmental Modelling & Software*, 24, no. 1, 124–134. ISSN 1364-8152. doi:10.1016/j.envsoft.2008.06.011.
- Rappuoli R. and Dormitzer P.R., 2012. *Influenza: Options to Improve Pandemic Preparation*. *Science*, 336, no. 6088, 1531–1533. doi:10.1126/science.1221466. URL <http://www.sciencemag.org/content/336/6088/1531.abstract>.
- Russell C.A.; Jones T.C.; Barr I.G.; Cox N.J.; Garten R.J.; Gregory V.; Gust I.D.; Hampson A.W.; Hay A.J.; Hurt A.C.; de Jong J.C.; Kelso A.; Klimov A.I.; Kageyama T.; Komadina N.; Lapedes A.S.; Lin Y.P.; Mosterin A.; Obuchi M.; Odagiri T.; Osterhaus A.D.M.E.; Rimmelzwaan G.F.; Shaw M.W.; Skepner E.; Stohr K.; Tashiro M.; Fouchier R.A.M.; and Smith D.J., 2008. *The Global Circulation of Seasonal Influenza A (H3N2) Viruses*. *Science*, 320, no. 5874, 340–346. doi:10.1126/science.1154137.
- Smith D.J.; Lapedes A.S.; de Jong J.C.; Bestebroer T.M.; Rimmelzwaan G.F.; Osterhaus A.D.M.E.; and Fouchier R.A.M., 2004. *Mapping the Antigenic and Genetic Evolution of Influenza Virus*. *Science*, 305, no. 5682, 371–376. doi:10.1126/science.1097211.
- Takekawa J.Y.; Newman S.H.; Xiao X.; Prosser D.J.; Spragens K.A.; Palm E.C.; Yan B.; Li T.; Lei F.; Zhao D.; Douglas D.C.; Muzaffar S.B.; and Ji W., 2010. *Migration of Waterfowl in the East Asian Flyway and Spatial Relationship to HPAI H5N1 Outbreaks*. *Avian Diseases Digest*, 5, no. s1, e101–e102. ISSN 1933-5334.
- Tharakaraman K.; Jayaraman A.; Raman R.; Viswanathan K.; Stebbins N.W.; Johnson D.; Shriver Z.; Sasisekharan V.; and Sasisekharan R., 2013. *Glycan Receptor Binding of the Influenza A Virus {H7N9} Hemagglutinin*. *Cell*, 1016, no. 10. ISSN 0092-8674. doi:10.1016/j.cell.2013.05.034. URL <http://www.sciencedirect.com/science/article/pii/S0092867413006405>.
- Trifonov V.; Khiabani H.; and Rabadan R., 2009. *Geographic Dependence, Surveillance, and Origins of the 2009 Influenza A (H1N1) Virus*. *New England Journal of Medicine*, 361, no. 2, 115–119. doi:10.1056/NEJMp0904572. URL <http://www.nejm.org/doi/full/10.1056/NEJMp0904572>. PMID: 19474418.
- USDA, 2013. *Assessment of Introduction Pathway for Novel Avian Influenza Virus into North America by Wild Birds from Eurasia*. Tech. rep., United States Department of Agriculture (USDA)'s Animal & Plant Health Inspection Service (APHIS) Department of Veterinary Services and Centers for Epidemiology and Animal Health. URL http://www.aphis.usda.gov/newsroom/hot_issues/avian_influenza/contents/printable_version/ai_pathway_assessment.pdf. Available online.
- WHO, 2013. *World Health Organization: Global Health Observatory Map Gallery*. URL http://gamapserver.who.int/mapLibrary/Files/Maps/2003_AvianInfluenza_GlobalMap_01Feb13.png. Areas with confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003–2013.
- Wilson H.M.; Petersen M.R.; and Sexson M.G., 2007. *Relation of Highly Pathogenic Avian Influenza (H5N1) Prevalence to Migration Patterns of Pacific Common Eiders Nesting in Northwest Alaska: Summary Report, 2007–Northwest Alaska*. Tech. rep., U.S. Fish and Wildlife Service and U.S. Geological Survey. URL http://alaska.fws.gov/mbmp/mbm/waterfowl/surveys/pdf/nwak_ai_coei.pdf. Available online.