Discretized Migration Flow: A Vector Field Based Tool for Avian Mobility in Patchy Mechanistic Models of Early Pathogenic Spread

Raphael Chiemezie Kelly

Project advisors: Peter D. Harrington[†], Dr. Mark A. Lewis[†]

Abstract. Various pathogens are spread through avian hosts. The spread of these pathogens can have massive economic and health consequences. Spatially explicit models of spread are needed in order to anticipate where and when diseases will spread. However, making predictions from models for such diseases has traditionally been challenging due to the complexity of bird movements, and a lack of comprehensive data on them. This paper proposes a model for the directional movement of birds between patches within epidemiological models. This model considers bird mobility in two ways: directed migration and random diffusion. Migration is incorporated through a vector field that represents average movements each migratory season, generated based on continental flyways. Diffusion is incorporated between neighbouring patches and segmented between each of the major flyways. Migration and diffusion are combined into a large, temporally varying mobility matrix that represents the movement of each bird in one patch to another. The mobility matrix is then used with a system of susceptible-infected (SI) differential equations to determine the spread of disease. The system was solved and results verified against infection data on the West Nile virus (WNv) outbreak in the US in 1999 and *Turdus migratorius* distributions, demonstrating the model's ability to accurately predict both the major spatio-temporal phases of WNv spread as well as the phases of American robin migration. This approach, here called discretized migration flow (DMF), can be further developed and explored for application in early stage emerging disease models.

1. Introduction.

There are constantly novel emerging infectious diseases (EIDs) that come with significant economic and public health consequences. Approximately 60% of all EIDs are zoonoses, diseases that originate in animals [1]. Bird movement, especially migration, is crucial to the spread of many pathogens, including West Nile virus (WNv), Avian Influenza (H5N1), and Eastern Equine Encephalitis virus (EEEv) [2, 3]. Zoonoses carried by migratory birds pose an additional threat due to the accelerated rate of disease spread and the added risk of poultry contamination. However, bird migration is an inherently complex process which is difficult to model, especially due to the lack of comprehensive data and the variety of potential host species. Nevertheless, understanding the spread of EID zoonoses during the early-stages of an outbreak can be key in order to strategically allocate resources and mitigate impacts.

Mechanistic models, such as ordinary differential equations (ODEs), are useful for early stage epidemiological modelling, as they allow for useful predictions when data is lacking (as opposed to many statistical models) and for information on parameters to be inferred through retroactive model fitting. Additionally, spatiality can be incorporated through a system of ODEs in a patchy modelling framework. In order to make a spatially predictive model, it is useful to make the patches of equal sizing and finely distributed, as opposed to creating fewer larger regions. However, this approach makes it difficult to include bird migration, since mobility coefficients cannot be easily added into the model manually (due to the high number of patches). Mechanistic models such as ODEs often do not normally

---

[†]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB.

incorporate bird migration or spatiality; models that do often utilize a low spatial resolution [4]. On the other hand, statistical methods like those used for climate or environmental analyses are not mechanistic and, thus, are often not ideal for use during the early stages of a disease outbreak, especially of an novel disease, due to a lack of data. Large scale bird migration data is also sparse, with radar data, geolocator tracking tags, bird sightings, and distribution analyses all having their drawbacks. Tracking tag (such as geolocator) studies can study individual birds and populations. However, it is difficult to apply tracking tags to a continental scale and impossible to directly track movement from every patch to every other patch, ensuring every single patch is equally accounted for. Such efforts would also be impractical near the start of a novel outbreak. Thus, it is not feasible to rely on such studies alone. Radar data has been used to examine the direction and strength of bird movements but is data intensive and does not monitor individual species (which is of importance to many epidemiological models). In addition, radar data can be susceptible to error through precipitation. Bird sightings and distribution analyses have become popular and can be used to model the locations of birds. However, they cannot inherently model the direction movement from each patch to every other patch, and thus cannot be used alone either.

This paper addresses the challenges listed above by developing and describing an approach to modelling bird movement in a way that can be incorporated into patchy ODE models. A solution was developed that is computationally efficient, flexible, and easily incorporated into existing models. The approach, hereby called discretized migration flow (DMF), models bird mobility as both migration and diffusion. Migration is represented as a velocity vector field that is constructed using location coordinates along the migratory flyway of the bird species. Diffusion is represented as constant and segmented according to migratory flyway. These movements are averaged in a state-transition matrix where elements represent proportional movement from patch to patch, taking both migration and diffusion into account. In this way, the model is analogous to a discretized convection-diffusion equation with no sources or sinks and with boundary conditions on diffusion. The resulting matrix of DMF can then be incorporated into a disease model, such as the SI model exemplified here, and solved for movement. The benefit of this approach is that it incorporates enough spatial complexity to be predictive and can be easily incorporated into existing epidemiological models or developed further.

The paper is organized as follows: section 2 describes the base epidemiological equations (subsection 2.3), DMF model (subsection 2.4), and other implementation details. Evaluation of results are found in section 3, and the discussion and conclusions follow in section 4 and section 5.

2. Methodology.

2.1. Case Study: West Nile virus.
West Nile virus (WNv) is a dangerous, mosquito borne disease that first originated in Uganda and is now commonly found throughout most continents [5]. WNv rapidly spread in the years following its initial North American case in New York, 1999. At present, WNv is endemic, with bird species such as *Turdus migratorius* (American robin) as primary hosts, and mosquito species such as *Culex pipiens, Cx. tarsalis, and Cx. restuans* as important vector transmitters [6]. In humans, it can also cause various life-threatening neuroinvasive diseases such as encephalitis and meningitis [7]. In the initial

years following North America's first case in New York, phylogenetic studies have shown that West Nile spread has been related to migratory bird flyways [3, 8]. WNv was chosen as a case study for application and testing of the approach described in this paper due to the accessibility of data to test model predictions. The United States was chosen as the loose study boundary, due to the availability of mapped, yearly data on early WNv transmission. The rest of this paper describes and evaluates the efficacy of the DMF model in relation to WNv data.

2.2. Spatial Grid Creation. In order to implement DMF, the study area was first divided into 1274 equal area grid cells extending slightly beyond the contiguous United States. These were created using the Albers Equal Area Conic coordinate reference system (EPSG:5070), and implemented using Python. Various libraries were utilized, such as NumPy, Pandas, GeoPandas, MatPlotLib, GeoPy, Plotly, Shapely, and Folium.The side length of each grid cell was set to be 100 km, about half of the average distance travelled by a migrating American robin per day [9]. The finely scaled grid helped allow the model to be spatially predictive.

2.3. Mathematical Model. For this model, the American robin was chosen to be the host species, and *Culex pipiens, Cx. tarsalis, and Cx. restuans* were collectively chosen to be the vector species. In order to model WNv spread, a patchy system of ordinary differential equations was used. The base structure of the model, based on coupled ODEs, was largely adapted from [10, 12]. For each species, two compartmental classes were used: susceptible (S) and infected (I). S_{Bi} is the size of the susceptible hosts class, I_{Bi} the infected hosts class, S_{Mi} the susceptible vectors class, and I_{Mi} the infected vectors class for a given patch, i. Additionally, $N_{Bi} = S_{Bi} + I_{Bi}$ and $N_{Mi} = S_{Mi} + I_{Mi}$, representing the total size of the susceptible and infected classes for each species. The variable t is the time in days. This model assumes reservoir frequency dynamics, a common assumption for West Nile virus [10]. Parameter values and ranges were either derived or approximated from literature. Details on the model and parameters can be found in Figure 1 and Table 1.

Figure 1. *System of Ordinary Differential Equations Model for WNv.*

	Natural Births	Natural Deaths	WNv Deaths	Compartmental Transmission	Regional Influx	Regional Efflux
$\frac{dS_{Bi}}{dt} =$	$b_B N_{Bi}$	$-b_B \frac{S_{Bi} N_{Bi}}{K_B}$		$-rp I_{Mi} \frac{S_{Bi}}{N_{Bi}}$	$+\sum_{k=1}^{n}[(1-d)\beta_{ik} + dD_{ik}]S_{Bk}$	$-\sum_{k=1}^{n}[(1-d)\beta_{ki} + dD_{ki}]S_{Bi}$
$\frac{dI_{Bi}}{dt} =$		$-b_B \frac{I_{Bi} N_{Bi}}{K_B}$	$-v I_{Bi}$	$+rp I_{Mi} \frac{S_{Bi}}{N_{Bi}}$	$+\sum_{k=1}^{n}[(1-d)\beta_{ik} + dD_{ik}]I_{Bk}$	$-\sum_{k=1}^{n}[(1-d)\beta_{ki} + dD_{ki}]I_{Bi}$
$\frac{dS_{Mi}}{dt} =$	$b_M N_{Mi}$	$-b_B \frac{S_{Mi} N_{Mi}}{K_{Mi}}$		$-rq S_{Mi} \frac{I_{Bi}}{N_{Bi}}$		
$\frac{dI_{Mi}}{dt} =$		$-b_B \frac{I_{Mi} N_{Mi}}{K_{Mi}}$		$+rq S_{Mi} \frac{I_{Bi}}{N_{Bi}}$		

Where S_{Bi}, I_{Bi}, N_{Bi} and S_{Mi}, I_{Mi}, N_{Mi} are the numbers of susceptible, infected, and total birds and mosquitoes respectively.

2.3.1. Logistic Growth. In order to limit the population growth of both the bird and mosquito species, a logistic growth assumption was incorporated into the model. Because

Param.	Value	Meaning	Ref./Calc.
b_B	0.022	Bird birth rate	Assuming eight yearly births per bird
b_M	0.07	Mosquito birth rate	[12]
K_B	240000.0	Bird maximum population size (per region)	[14] averaged per square km
K_M	5760000.0	Mosquito maximum population size (per region)	Set to match the proportion of mosquitoes N_M to birds N_B found in [12]
r	0.52	Mosquito bite rate	[12]
q	0.16	Proportion of transmission $I_B \rightarrow S_M$	[12]
p	0.88	Proportion of transmission $I_M \rightarrow S_B$	[12]
v	0.07	Bird WNv death rate	[12]
d	0.01	Diffusion percentage	N/A

there are only births in the susceptible class, but deaths in both the susceptible and infected classes, the logistic growth assumption was split into a "birth" term and "death" term.

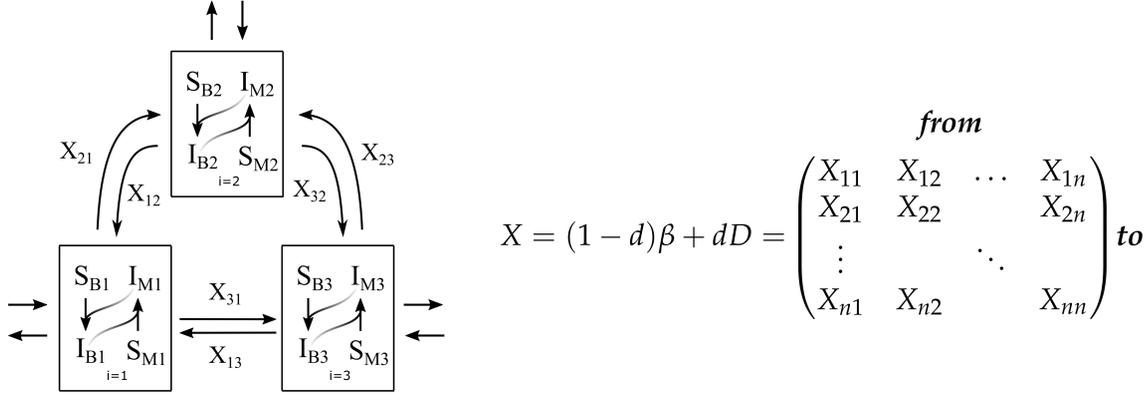For a general population N with a maximum population size of K, the logistic growth assumption is:

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}) = rN - r\frac{N^2}{K},$$

where rN can be considered to be the birth term and $r\frac{N^2}{K}$ the death term, so that K is the carrying capacity. The birth rate remains proportional to the population size for all population sizes. On the other hand, the death rate starts at zero and grows closer to the birth rate as the population reaches its theoretical maximum. The birth term was included for only susceptible individuals, and the death term for both susceptible and infected individuals. The death rate was scaled proportionally to the class size. Since $N = S + I$, the susceptible and infected death rates will add to be the total death rate, maintaining the logistic growth assumption. Thus, the model for susceptible and infected individuals is:

$$\begin{array}{ll} \frac{dS}{dt} = rN - r\frac{N^2}{K} \cdot \frac{S}{N} & \frac{dS}{dt} = rN - r\frac{SN}{K} \\ \frac{dI}{dt} = -r\frac{N^2}{K} \cdot \frac{I}{N} & \rightarrow \quad \frac{dI}{dt} = -r\frac{IN}{K} \end{array}$$

2.3.2. Spatial Framework. In this model, mosquitoes are assumed to be stationary within a patch. Only susceptible and infected birds move between regions. To incorporate the spatial movement of birds, a meta-population framework was used. Both regional influx and efflux were incorporated into the equations for $\frac{d\,S_{Bi}}{dt}$ and $\frac{d\,I_{Bi}}{dt}$. Sparse matrices β and D are generated in the upcoming sections, to represent patch to patch avian migration and diffusion respectively. The dimensionality of these matrices are $n \times n$, where n represents the number of distinct patches. The elements β_{ik} and D_{ik} represent the proportion of birds starting within patch k that end within patch i, either by migration or diffusion (within a day). Due to this,

Figure 2. *Structural overview of the ODE spatial framework and matrix structure. Normal disease transmission within patches occurs as birds move according to their mobility coefficients. Here, X_{ik} represents the percentage of birds starting in region k that end in region i, due to both migration and diffusion (in a day).*


$$X = (1 - d)\beta + dD = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & & X_{2n} \\ \vdots & & \ddots & \\ X_{n1} & X_{n2} & & X_{nn} \end{pmatrix} to$$

the column sums of both matrices are equal to 1;

$$\sum_{i}^{n} \beta_{ik} = \sum_{i}^{n} D_{ik} = 1 \ (for \ all \ k).$$

The migration matrix by itself, β, incorporates long distance migration at a relatively small rate and assumes that a high proportion of birds stay within the same patch in a given time step (and is, thus, more similar to the identity matrix I). On the other hand, the diffusion matrix by itself, D, represents local movements where equal portions of birds starting in patch k end within a neighbouring patch (or remain stationary). As a result, D is given a very small weighting compared to β. This small weighting is characterized by the parameter d. In addition, the migration matrix changes with respect to time, for different parts of the season/simulation. The low and high pre-breeding season migration matrices are $\beta_{\text{pre-I}}$ and $\beta_{\text{pre-II}}$ respectively. $\beta_{\text{post-I}}$ and $\beta_{\text{post-II}}$ are the low and high post-breeding season migration matrices respectively. Within the differential equations, the influx and efflux terms are the sum of all movement from each patch to every other patch. An overview diagram of the meta-population and matrix structures can be found in Figure 2.

2.3.3. Seasonality.
The model was solved on a seasonal, impulsive basis. Seasonality was incorporated into the model through various stages. The migratory seasons were defined according to definitions from eBird [11], with gaps between season dates interpolated. Based on these seasons, a period of "high" migration was defined to be the range covering the middle third of dates for the pre-breeding and post-breeding migratory seasons (the remaining dates are periods of "low" migration). Thus, β was defined piece-wise to be:

Figure 3. *General diagram of model seasonality and implementation. The x-axis lists months of the year. The birds row signifies the different date ranges that are considered migratory and stationary seasons. The mosquitoes row shows a simplified version of the total mosquito population size, as a function of time. The shaded grey region represents what is considered WNv season. The start and end dates, as well as the date when WNv is introduced are shown on the graph diagram as well.*
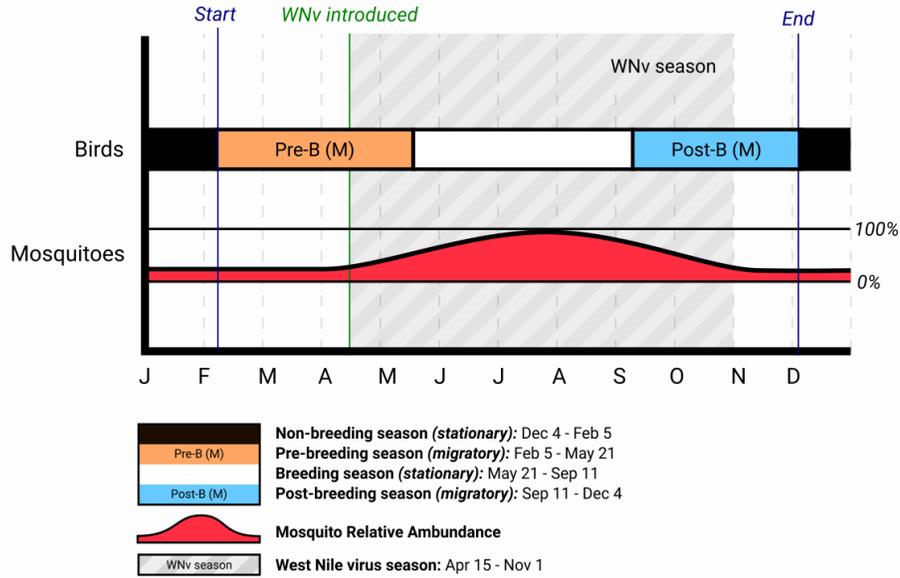
$$
\beta(t) = \begin{cases}
\beta_{pre-I} & (036 \leq t < 071,\, 106 \leq t < 141) \\
\beta_{pre\text{-}II} & (071 \leq t < 106) \\
I & (141 \leq t < 254,\, t < 036,\, t > 338) \\
\beta_{post\text{-}I} & (254 \leq t < 293,\, 321 \leq t < 338) \\
\beta_{post\text{-}II} & (293 \leq t < 321)
\end{cases}
$$

where I is the identity matrix and t is the day (according to the Julian calendar). Additionally, the maximum mosquito population size, K_M, is also parameterized, differing by both region and time of year. This is explained later on in subsection 2.7.

The West Nile virus season was determined to be from April 15 to October 31, based on [11]. The disease is only introduced into the model at the start of the WNv season, by infecting 50% of birds and mosquitoes within patches of the most infection from last year. This proportion was chosen arbitrarily, primarily so that the disease does not die out before it is given a chance to spread. The model itself was solved within the date range: Feb 5 to Dec 4, for the years 2000, 2001, and 2002. The same initial distribution of birds and mosquitoes was used for every year of the simulation. A diagram overview of the seasonality within the model is available in Figure 3.

Figure 4. *A simplified image overview of the vector field creation process. Coordinate points were visually created along a map of the major US migratory flyways. Then, these locations coordinates were connected tip to tail with vectors that were then interpolated to create a numerical vector field.*

2.4. Migration Matrix (β). American robin migration is represented and modelled in the form of two velocity vector fields that are used to generate the required β matrices. The pre-breeding and post-breeding season velocity vector fields were derived from a set of ordered longitude/latitude coordinates that were created based on the major migratory flyways of the United States. It was assumed that, for the American robin, the pre-breeding and post-breeding migrations made use of the same flyways and corridors. Coordinates correspond to locations along migratory routes, approximating the path an individual may follow during either a pre-breeding or post-breeding migration. Two to four sets of point locations were created per migratory flyway, resulting in 13 coordinate sets. An overview of this process can be found in Figure 4.

These point locations were converted into coordinates in the Albers (equal area conic) projection (here represented as x_i and y_i). Coordinate-vector pairs were generated between points of the same migration path (here, r) using a successive algorithm, with coordinates representing the origin of each vector. Vectors were directed northward and southward for the pre-breeding and post-breeding migratory seasons respectively. The magnitude of each vector was set to be 100 km, about half of the average distance travelled by an American robin in a day. Variation in travel speed would later be incorporated by use of factors which alter the magnitude of each vector. These coordinate-vector pairs were combined into arrays for both the pre-breeding and post-breeding season. Represented as matrices (where V_r and W_r are the pre-breeding and post-breeding season matrices for a given migratory route (r) respectively and $\vec{v}_i$ and $\vec{w}_i$ are the pre-breeding and post-breeding season vectors for r):

$$\mathbf{V_r} = \begin{pmatrix} x_1 & y_1 & \vec{v_{1x}} & \vec{v_{1y}} \\ x_2 & y_2 & \vec{v_{2x}} & \vec{v_{2y}} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n-1} & y_{n-1} & \vec{v_{(n-1)x}} & \vec{v_{(n-1)y}} \end{pmatrix}, \; \mathbf{W_r} = \begin{pmatrix} x_2 & y_2 & \vec{w_{2x}} & \vec{w_{2y}} \\ x_3 & y_3 & \vec{w_{3x}} & \vec{w_{3y}} \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & \vec{w_{nx}} & \vec{w_{ny}} \end{pmatrix}$$

where,

$$\vec{v}_i = 100km \angle \texttt{atan2}(y_{i+1} - y_i, \, x_{i+1} - x_i) \quad (1 \leq i < n) \text{ and}$$
$$\vec{w}_i = 100km \angle \texttt{atan2}(y_{i-1} - y_i, \, x_{i-1} - x_i) \quad (1 < i \leq n).$$

This process was repeated for all migratory routes considered ($r = 1 - N$) and combined to create larger arrays, represented as matrices (where $\mathbf{V}$ and $\mathbf{W}$ represent the complete set of pre-breeding and post-breeding season coordinate-vectors respectively)

$$\mathbf{V} = \begin{bmatrix} \mathbf{V_1} \\ \mathbf{V_2} \\ \vdots \\ \mathbf{V_N} \end{bmatrix}, \, \mathbf{W} = \begin{bmatrix} \mathbf{W_1} \\ \mathbf{W_2} \\ \vdots \\ \mathbf{W_N} \end{bmatrix}$$

.

These coordinate-vector pairs were then used as the basis of two dimensional, bi-cubic interpolations, resulting in two velocity vector fields. This process can be represented as (where $\vec{\mathbf{M}_{\mathbf{pre}}}$ and $\vec{\mathbf{M}_{\mathbf{post}}}$ represent the pre-breeding and post-breeding season migratory vector fields respectively.),

$$\vec{\mathbf{M}_{\mathbf{pre}}}(x_i, y_i) = \texttt{bicubic\_interp}(x_i, y_i | col_1(\mathbf{V}), col_2(\mathbf{V}), col_3(\mathbf{V}))\vec{\mathbf{i}}$$
$$+ \, \texttt{bicubic\_interp}(x_i, y_i | col_1(\mathbf{V}), col_2(\mathbf{V}), col_4(\mathbf{V}))\vec{\mathbf{j}}$$

$$\vec{\mathbf{M}_{\mathbf{post}}}(x_i, y_i) = \texttt{bicubic\_interp}(x_i, y_i | col_1(\mathbf{W}), col_2(\mathbf{W}), col_3(\mathbf{W}))\vec{\mathbf{i}}$$
$$+ \, \texttt{bicubic\_interp}(x_i, y_i | col_1(\mathbf{W}), col_2(\mathbf{W}), col_4(\mathbf{W}))\vec{\mathbf{j}}$$

$\texttt{bicubic\_interp}(x, y | X_{inp}, Y_{inp}, Z_{inp})$ is the bicubic interpolation, taking vectors X, Y, and Z as coordinate and data parameters. The function outputs a scalar value for every coordinate input. A diagram of the interpolation process can be found in Figure 5.

These velocity vector fields were utilized in order to create the matrices β_{pre-I}, β_{pre-II}, β_{post-I}, and $\beta_{post-II}$. β_{ik} represents the percentage of birds starting in region k that end in region i. To generate these, the following was done. Nine equally spaced (x, y) coordinate pairs were defined in a grid shape within each patch. Various speed factors were also used to account for differences in flight speeds among birds. Let C be the set of all of these coordinate pairs across patches, and let A be the set of speed factors used. For a given coordinate pair $c \in C$ and speed factor $a \in A$, $p(t|c, a)$ is the position of a point at time t that starts at c and moves with a speed factor of a. If $\vec{\mathbf{M}}$ represents the migration velocity vector field,

$$p(t|c, a) = c + \int_0^t a\vec{\mathbf{M}}(p(t'|c, a))dt'.$$

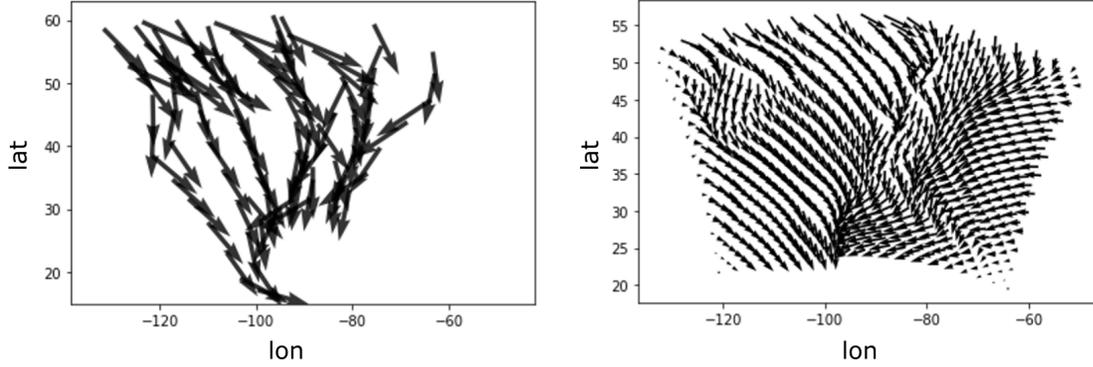Since the time step is a day, $t = 1$. The value of $p(1|c, a)$ was approximated recursively using the definition,

Figure 5. *A diagram of the interpolation process for the vector field. Left is pre-interpolation. Right is post-interpolation.*

$$p(t|c,a) \approx c + \sum_{i=1}^{4} a \frac{\vec{\mathbf{M}}(p(\frac{i-1}{4}|c,a))}{4}.$$

Then, the migratory matrix, β, was generated using the following definition,

$$\beta = (\beta_{ik}) = (\frac{n_{ik}}{9n_a})$$

where (get_patch(x,y) returns the patch that (x,y) is within).

$$n_{ik} = \left\{(c,a) : \text{get\_patch}(c) = k, \text{get\_patch}(p(1|c,a)) = i\right\}\Big|$$
$$n_a = |A|.$$

This process was repeated for both the pre-breeding and post-breeding season and using different speed factors each time. The following definitions were used:

$$
\begin{aligned}
\beta_{pre-I} \quad & \text{found using } \vec{\mathbf{M}_{pre}} \text{ and } A = \{1, 1.5, 2, 2.5\} \\
\beta_{pre-II} \quad & \text{found using } \vec{\mathbf{M}_{pre}} \text{ and } A = \{2, 2.5, 3, 3.5\} \\
\beta_{post-I} \quad & \text{found using } \vec{\mathbf{M}_{post}} \text{ and } A = \{1, 1.5, 2, 2.5\} \\
\beta_{post-II} \quad & \text{found using } \vec{\mathbf{M}_{post}} \text{ and } A = \{2, 2.5, 3, 3.5\}.
\end{aligned}
$$

Because of the way the array was constructed, $\sum_{i}^{n} \beta_{ik} = 1$.

2.5. Diffusion Matrix (D). Diffusion was considered to be the inter-regional dispersion of birds on a small scale. To incorporate diffusion, the study area was divided into four segments for each of the major migratory flyways in North America (Pacific, Central, Mississippi, and Atlantic). Next, a diffusion matrix, D, was created that represents the percentage of birds starting in region k that diffuse into region i, assuming maximal diffusion. Diffusion was assumed to only occur within the same flyway. To construct the matrix, the following definition was used:

$$D = (D_{ik}) = \begin{cases} \frac{1}{m_i+1} & k = i \text{ or } k \text{ is a neighbour to } i \\ 0 & \text{otherwise} \end{cases}$$

where, m_i is the number of neighbouring patches to i that are within the same flyway. A patch is considered a neighbour if its coordinate differs from i by 1. Because of this definition, $\sum_i^n D_{ik} = 1$.

In the ODEs, a parameter d governs the percentage of mobility that is considered to be diffusion. Since the matrix D assumes maximal diffusion, its relative weight is very small. Assuming that approximately 5% of individuals diffuse maximally over the course of 10 days, the value of d was set to be 0.005.

2.6. American robin Initial Distribution. The initial distribution of the host species was determined by approximating the distribution of American robins found during non-breeding seasons present on eBird [11]. This is shown in Figure 6. (For a direct visual comparison to the non-breeding season American robin distribution given by eBird, the reader is directed to the [13].)

2.7. Mosquito Initial Distributions. The three species *Culex pipiens, Cx. restuans, Cx. tarsalis* were considered to act as a single species for this model. The relative spatial distribution of these mosquito species between locations was assumed to remain the same throughout the year. The mosquito density in any location is assumed to only vary proportionally to a scaling factor that changes throughout the year. The mosquito distribution at maximal intensity was determined through ecological niche model outputs. Ecological niche modelling is a statistical machine learning technique which uses climate data layers and presence/absence locations of a species to make raster predictions for the likeliness of finding a species at any location within the study space. The outputs of three models, each ran for one of the species, were obtained in raster format from records held in the Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution and accessed through the VectorMap data portal [18]. These outputs were raster averaged in QGIS, and subsequently averaged per cell of the study space grid. These polygonal averages were then normally scaled so that $\texttt{max}(\text{all outputs}) = 2.40 \times 10^5$ (as per Table 1).

It is assumed that the distribution of mosquitoes across the continent stays constant and that the relative density within all patches changes only with time. Thus, K_M is defined such that:

$$K_{Mi}(t) = 10K_{Mi}(t_0)f(t)$$

where $K_{Mi}(t_0)$ is the initial mosquito population size within region i. The maximum population size for any given patch, K_{Mi}, will be ten times the size of its initial population size. Working in reverse, the initial mosquito population size for each patch is set as one tenth of the desired maximum. The desired maximum is given by the normalized polygonal average that was previously calculated. A function, $f(t)$, approximates the relative distribution of mosquitoes throughout the year and is defined by the function:

$$f(t) = 0.06 + 0.57(1.3e^{\frac{-1}{4222}(t-170)^2} + e^{\frac{-1}{1222}(t-230)^2})\ (where\ 0.0 \le f(t) \le 1.0)$$
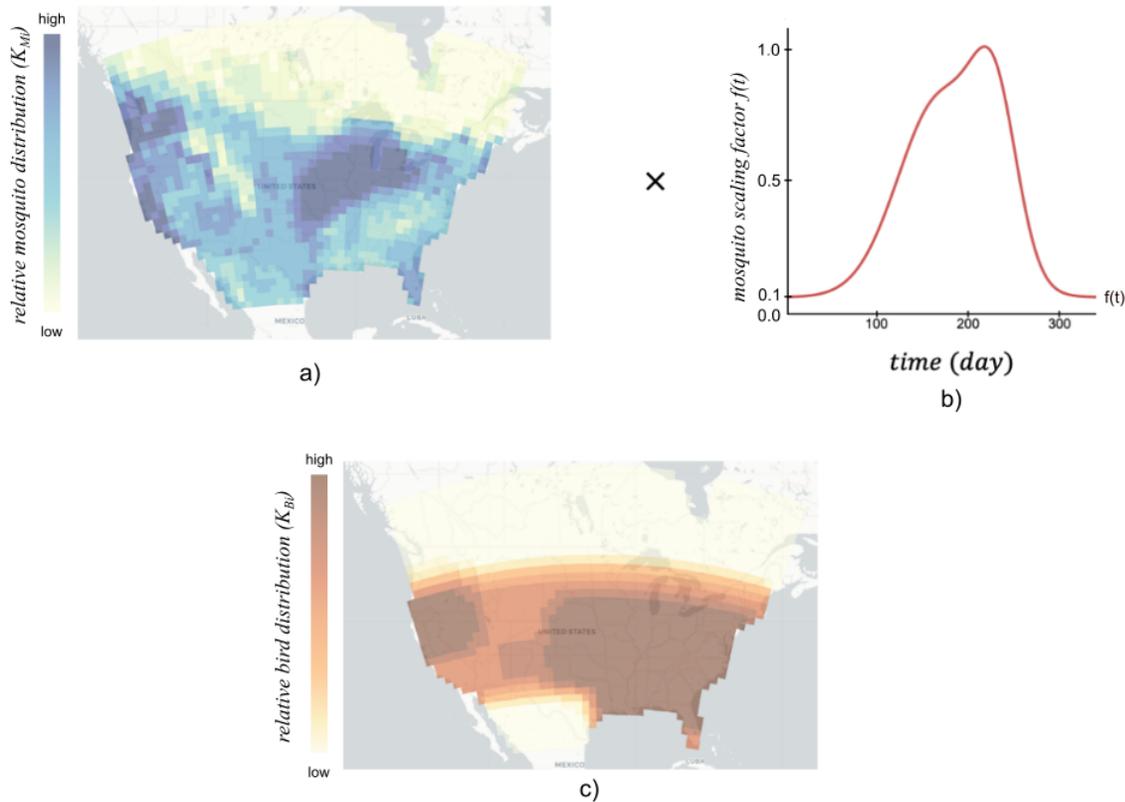
Figure 6. *a) The initial relative mosquito distribution across the study space. b) The relative mosquito scale factor versus time. c) The initial relative bird distribution across the study space.*

This function was determined by manually fitting a curve to an average of the *Culex* relative distributions versus time, as shown in [15].

An image of this is shown in Figure 6.

2.8. Solving the System of Differential Equations. The resulting system of 5096 differential equations (4 equations × 1024 patches) Figure 1 was solved numerically, using the Differential Equations library in the Julia programming language, for the years 2000, and 2001. A parallel, 5th order Runge-Kutta method was used.

3. Results. Due to the complex nature of spatio-temporal data and the goals of this project, no single statistic was used to determine the success rate of the simulation. Rather, the primary goal of the model was to predict the main stages of WNv outbreak.

3.1. Evaluation of WNv Spatial Predictions. The model was evaluated based on its ability to capture the major phases of WNv disease outbreak. A results metric, R, was calculated as the average number of infected mosquitoes within each polygonal region during the "WNv season". This value was mapped, and results are shown in Figure 7. As exact county-level data was not publicly available to perform spatial correlation tests, a side by
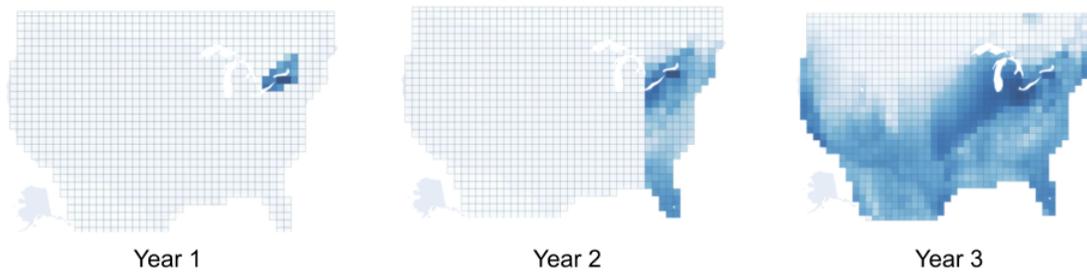
Year 1   Year 2   Year 3

Figure 7. *The predicted yearly results of the model for the first three years of simulation captured by the value* $R = \sum_i \frac{I_M}{N_M}$. *Darker patch colours represent higher relative R values. The Year 1 distribution represents the initial disease distribution that was inputted into the model. The model was then run on the distribution for Year 1 to predict the final disease distribution at the end of Year 2. Finally, the model was run on the results for Year 2 to gain a prediction of the final disease distribution at the end of Year 3.*

side visual comparison was done between Center for Disease Control and Prevention (CDC) infection data [16] and model results. The exact figures used for the visual comparison can be found in Figure 4 of [17]. This comparison demonstrated the similarities between the model predictions and the actual disease spread.

Some key observations include the accuracy of the model in recreating the stages of WNv outbreak as shown in the yearly CDC data. It should be noted that, although this model was made to model the outbreak which began in 1999, the initial distribution used for the model was recreated from the data for 2000 as there was little change in the disease distribution between these years, and the disease remained relatively localized within the year 1999. Hence, the results for year 1, year 2, and year 3 correspond most closely to the actual disease distributions for the years 2000, 2001, and 2002. The initial spread could be classified into three distinct stages. Initially, there is not much change because of a lack of movement. However, due to the level of diffusion, the disease propagates to areas with lower host-species density, leading to higher infection rates on the perimeter of the major bird population. Afterwards, the trend was that the disease spread throughout the rest of the population and became homogenized, which was also replicated by the model. This model also replicates the real world behaviour where the disease emerged at the north and south edges of the bird population, due to reservoir frequency dynamics. Overall, mosquito prevalence had a strong effect in determining the disease hotspots and prevalence whereas birds mostly acted as transmitters that introduce the virus to specific sub-regions. The density of birds is not seen to have as much of an effect on the virus prevalence. Additionally, many iterations were done to test the effect of each parameter on the model output. For parameters with ranges, certain values were iteratively adjusted. This process revealed the model to be particularly sensitive to the parameters of mosquito bite rate, bird WNv death rate, and diffusion percentage. (Formal sensitivity analyses were not performed. However, sensitivity is meant in the sense that changes of about 10% to each value led to the disease being either eliminated or spreading homogeneously within flyways.) When the mosquito bite rate was too high, the entire system became infected too quickly whereas too low led to the disease dying out. A similar effect was noticed in regards to the migration itself, which acted as the mechanism for dis-

eases to spillover into neighbouring flyways. Finally, this also validated the assumption of flyway segmented diffusion, which was what allowed this model to work properly. Without that assumption, the disease would reach all parts of the study space within a few days and in very small (fractional) percentages, that would then infect the majority of individuals as individuals migrated. With flyway segmented diffusion, a similar delayed infection process where the disease spreads west-ward could be replicated.

3.2. Evaluation of American robin Migration Dynamics Prediction. The model was also evaluated based on how well it was able to reproduce the seasonal ranges of the American robin. A visual comparison with the eBird yearly range models [11, 13] and the supplemental animation file, demonstrates that the main stages of American robin migration ranges were adequately replicated. The ranges available at eBird visually correlate with the ranges of output from this model, showing that the migration assumption was able to hold up over long periods of time.

4. Discussion. In summary, DMF with flyway segmented diffusion was presented as a successful model of bird mobility in a state-transition matrix that can be included in early-stage epidemiological models. The approach was tested against the case of WNv and shown to provide useful predictions on the spatial spread of WNv for the initial years of WNv outbreak in the US, 1999.

Overall, this model was able to replicate the real world behaviour of WNv where the disease emerged at the northern and southern edges of the bird population, due to reservoir frequency dynamics. The model outputs visually correlated with the true result. However, it is important to note that spatial correlation values were not calculated, nor was any direct numerical comparison made with real world data. Though the model displays potential as a qualitative tool for predicting the spread of an emerging zoonose, further quantitative testing would be required to measure the accuracy and utility of the model in real-world situations.

Several new insights were gained from this work. The way in which flyway segmented diffusion and migration interacted was crucial to the disease dynamics. Incorporating diffusion was necessary to explain how the disease spread so quickly to opposite ends of the same flyway but not inland at the same pace. The flyway segmentation meant that it took longer for the disease to reach the most western flyway (the Pacific flyways). The idea of migratory connectivity is relevant here because it shows how the specific locations that a migratory bird travels to is important in introducing the disease to new locations.

The distinction of neighbouring flyways and the discreteness of the barriers between them, as well as the necessity of migration to overcome that barrier are three open questions brought up by this project that warrant further investigation. Overall, the assumption of flyway segmented diffusion seems to be a crucial one in order to explain known patterns. It provides an ability for the model to recognize areas of high and low transmittable, as opposed to having either homogeneous spread or lack of spread. Without it, the velocity vector field and migratory matrix would be descriptions of where birds do fly without explaining where they do not and why.

This general approach could be developed and refined in several ways. A limitation is the data points used to construct the vector field. A refinement could incorporate observed data into the vector field creation. An example could be individual level tracking tags or

presence observations taking the place of some coordinate sets that were determined based on the diagrams of North American flyways. Another possibility could be incorporating it with a radar weather system in order to get a more actively changing vector field. Additionally, the approach of using coordinates to create the vector field is not a requirement, as alternative means could result in the same, or better, solutions. And variation can be incurred in regards to the integration of the velocity vector field, using numerical approximation. It should also be noted that the assumption of using two classes (susceptible and infectious) for the American robin was made for the sake of keeping the number of differential equations to a minimum. Nonetheless, incorporating additional compartments, such as an exposed class, may make the model more realistic as American robins do not experience high mortality from WNv. Currently, there is a gap in simulation between successive years, as the model is not simulated during the non-breeding season for the American robin. As part of this, the choice of infecting 50% of birds and mosquitoes in patches with the most infection was arbitrary. Both of these assumptions could be altered to yield more accurate spatial results, incorporating the inter-seasonal dynamics of the disease by a time-continuous simulation across years.

In connection to future work, this approach could be developed into a live-updating early-stage epidemic modelling system that updates predictions as the disease moves along. This would allow for better understanding of disease dynamics in the context of an emerging infectious disease, and could provide immediate practical usage to government and policy makers. In addition, connections to connectivity could be further researched, as the heterogeneities in bird migration seem to be largely responsible for the introduction of disease into new areas.

5. Conclusion. This project demonstrates the start to end development, testing and evaluation of a novel method to characterize avian mobility in compartmental models. It was successfully tested against the historical case of West Nile virus, and can be applied to novel outbreak situations as well. This project developed a novel approach to characterise bird mobility through migration and diffusion, intended for use in early-stage epidemiological models where there is a lack of data to utilize a statistical method. It is an easily adaptable and efficient approach that can be employed by epidemiologists and ecologists to model bird mobility through migration and non-contiguous diffusion in a realistic way for prediction-oriented models. The approach was justified by numerous numerical test results that produced results that visually correlated with the yearly disease spread, matched bird distribution estimates from eBird, and reinforced the importance of the assumption of reservoir frequency dynamics in these models, in addition to pointing to many additional interesting factors. This tool can be easily used by ecologists and epidemiologists to aid in studying and predicting the spread of new avian-host diseases as they occur in the future. Nonetheless, further testing would be required to determine the exact accuracy of the numerical results outputted by this model, as well as to test the validity of model assumptions against real-world test cases.

REFERENCES

[1] K. JONES, et al., Global trends in emerging infectious diseases, Nature, 451 (2008), pp. 990–993.

[2] K. DHAMA, et al., Pathogens Transmitted by Migratory Birds: Threat Perceptions to Poultry Health and Production, Int. J. Poult. Sci., 7(6) (2008), pp. 516–525.

[3] K. REED, et al., Birds, Migration and Emerging Zoonoses: West Nile Virus, Lyme Disease, Influenza A and Enteropathogens, Clin. Med. Res., 1(1) (2003), pp. 5–12.

[4] S. RILEY, et al., Five challenges for spatial epidemic models, Epidemics, 10 (2015), pp. 68–71.

[5] J. J. SEJVAR, West Nile Virus: An Historical Overview, Ochsner J., 5(3) (2003), pp. 6–10.

[6] B. M. DUNPHY, et al., Long-term surveillance defines spatial and temporal patterns implicating Culex tarsalis as the primary vector of West Nile virus, Sci. Rep., 9 (2019).

[7] R. L. DEBIASI and K. L. TYLER, West Nile virus meningoencephalitis, Nat. Clin. Pract. Neurol., 2(5) (2006), pp. 264–275.

[8] D. SWETNAM, et al., Terrestrial Bird Migration and West Nile Virus Circulation, United States, Emerg. Infect. Dis., 24(12) (2018), pp. 2184–2194.

[9] L. ERICKSON, Robin Migration, Journey North, https://journeynorth.org/tm/robin/facts_migration.html (10 February 2021)

[10] M. J. WONHAM and M. A. LEWIS, A Comparative Analysis of Models for West Nile Virus, Mathematical Epidemiology, (2008), pp. 365–390.

[11] Fink, D., T. Auer, A. Johnston, M. Strimas-Mackey, O. Robinson, S. Ligocki, W. Hochachka, C. Wood, I. Davies, M. Iliff, L. Seitz. 2020. eBird Status and Trends, Data Version: 2019; Released: 2020. Cornell Lab of Ornithology, Ithaca, New York. https://doi.org/10.2173/ebirdst.2019

[12] G. VRIES, West Nile Virus, Kings Center for Visualization in Science, (2019).

[13] eBird, Cornell Lab of Ornithology, https://science.ebird.org/en/status-and-trends/species/amerob/abundance-map (27 February 2021)

[14] Partners in Flight Database, https://pif.birdconservancy.org/population-estimate-database-scores/ (26 February 2021)

[15] K. LEBL, Predicting Culex pipiens/restuans population dynamics by interval lagged weather data, Parasites & Vectors, 6 (2013).

[16] Center for Disease Control and Prevention, Annual Maps & Data, (1999-2002), https://www.cdc.gov/westnile/statsmaps/finalmapsdata/index.html (27 February 2021).

[17] J. T. ROEHRIG, West Nile Virus in the United States — A Historical Perspective, Viruses, 5 (2013), pp. 3088-3108

[18] VectorMap Data Portal, Walter Reed Biosystematics Unit, http://vectormap.si.edu (15 February 2021)