Approximate Bayesian Inference for Reconstructing Velocities of Migrating Birds from Weather Radar

Daniel Sheldon[1], **Andrew Farnsworth**[2], **Jed Irvine**[3], **Benjamin Van Doren**[2], **Kevin Webb**[2], **Thomas G. Dietterich**[3] and **Steve Kelling**[2]

[1]University of Massachusetts Amherst, Amherst, MA, USA; sheldon@cs.umass.edu

[2]Cornell Lab of Ornithology, Ithaca, NY, USA; {af27,bmv25,kfw4,stk2}@cornell.edu

[3]Oregon State University, Corvallis, OR, USA; {irvine,tgd}@eecs.oregonstate.edu

Abstract

Archived data from the WSR-88D network of weather radars in the US hold detailed information about the continent-scale migratory movements of birds over the last 20 years. However, significant technical challenges must be overcome to understand this information and harness its potential for science and conservation. We present an approximate Bayesian inference algorithm to reconstruct the velocity fields of birds migrating in the vicinity of a radar station. This is part of a larger project to quantify bird migration at large scales using weather radar data.

1 Introduction

The National Weather Service operates the WSR-88D (Weather Surveillance Radar-1988 Doppler) network of 159 radars in the United States and its territories. The network covers nearly the entire US and was designed to detect and study weather phenomena such as precipitation and severe storms. However, WSR-88D radars also detect biological phenomena, including movements of birds, bats, and insects in the atmosphere (Kunz et al. 2008). WSR-88D data are archived and available from the early 1990s to present. They provide a rare opportunity to study large-scale movements of birds to advance science and conservation. However, despite the fact that a number of bird migration studies have used WSR-88D data (Gauthreaux, Belser, and Blaricom 2003; Buler et al. 2012), these have all been limited in scope: for example, to a subset of the radar stations or to a limited number of nights within (or among) migration seasons.

A primary challenge that has prevented scientists from using these data to their full potential is that interpretation of the data can be difficult. Radar signatures of birds are similar to those of bats, insects, weather, and even airborne dust. Consequently, data must be interpreted manually by a highly-trained expert (Gauthreaux, Belser, and Blaricom 2003; Buler and Diehl 2009). This is not feasible unless the data used are significantly restricted in scope, either spatially or temporally. We estimate there are well over 100 million archived volume scans from single radar sites, and a single night during peak migration season will produce about 15,000 scans nationwide. If we can develop AI tools to automate the interpretation steps, this vast store of data could enable important advances in our understanding of many phenomena, including bird migration.

An accurate algorithm to estimate the velocities of birds and other targets detected by radar is critical to unlocking the potential of the data. Velocity information is important for understanding the biology of bird migration. In addition, there is growing evidence that the structure of the velocity field is a key to discriminating between birds, which fly under their own power, and other targets such as precipitation, insects, and dust, which are primarily carried by the wind (Gauthreaux, Belser, and Blaricom 2003; Dokter et al. 2011). In this paper, we formulate a novel probabilistic model and develop an approximate inference algorithm based on Expectation Propagation (Minka 2001) to reconstruct the velocity field of migrating birds from the partial velocity information collected by Doppler radar.

Our model is based on wind profiling algorithms from the weather community. However, unlike previous methods, we develop a novel likelihood function based on the wrapped normal distribution to simultaneously reason about the three sources of noise and incomplete information that are inherent in this problem. First, a Doppler radar can only detect *radial velocity*, the speed at which a target is approaching or departing the radar. The other velocity component is unknown, so it must be inferred from the global properties of the velocity field. Second, the true radial velocity value a is *aliased* or *wrapped* to the value $a \mod 2V_{\max}$ by the measurement process, where $V_{\max}$ is the Nyquist velocity, so the correct dealiased value must again be inferred from global properties. Finally, the individual measurements are subject to noise caused by instrument error and fine-scale variability in the true target velocities.

Two of these problems are exacerbated specifically by conditions present during bird migration. First, one of the clear-air radar operating modes that is commonly used during these periods has the lowest Nyquist velocity (11.61 ms^{-1}) of all operating modes, and this leads to the greatest loss of information due to aliasing. Second, the fine-scale variability of the velocity field of migrating birds is much greater than that of wind-borne particles (Dokter et al. 2011), which means that the data are inherently noisier.

Our method avoids difficulties faced by most previous

methods, which treat the problems of *dealiasing* (correcting aliasing errors) and velocity reconstruction separately. Doing so risks compounding of errors as they propagate from one problem to the next. Furthermore, solving one problem often assumes that the other has already been solved, which creates a chicken-and-egg problem unless external information is available. Reference wind measurements are often used for this purpose, but these will inaccurately characterize the velocities of migrating birds.

Among the previous algorithms, two notable exceptions are the gradient-based velocity profiling methods of Tabary, Scialom, and Germann (2001) and Gao and Droegemeier (2004), which do not require the radial velocity data to be dealiased first. Our first contribution, in Section 3, is to improve and analyze the gradient-based velocity azimuthal-display (GVAD) method of Gao and Droegemeier. We show theoretically that it is subject to inherent tradeoffs that limit its accuracy. This motivates our new approach.

In Section 4, we present our new method based on the wrapped normal likelihood, and also introduce a Gaussian Markov prior to encourage smoothness in the recovered velocity field. We develop an approximate inference method based on Expectation Propagation to recover the velocity field. In Section 5, we evaluate the algorithms and show that our extensions to GVAD and our new EP algorithm each improve significantly over the previous methods.

2 Background and Problem Definition

Radar Basics. Each radar in the WSR-88D network collects data by conducting a sequence of *volume scans*, which complete every six to ten minutes. Each volume scan consists of a sequence of *sweeps* during which the antenna rotates 360 degrees around a vertical axis while keeping its elevation angle fixed. The result of each sweep is a set of raster data products summarizing the radar signal returned from targets within discrete *pulse volumes*, which are the portions of the atmosphere sensed at a particular antenna position and range from the radar. The coordinates (r, ϕ, ρ) of each pulse volume are measured in a three dimensional polar coordinate system: r is the distance in meters from the antenna; ϕ is the *azimuth*, which is the angle in the horizontal plane between the antenna direction and a fixed reference direction (typically degrees clockwise from due north); and ρ is the elevation angle, which is the angle between the antenna direction and its projection onto the horizontal plane.

The two primary data products collected from each sweep are *reflectivity* $z(r, \phi, \rho)$, a measure of the total amount of power returned to the radar from targets within each pulse volume, and *radial velocity* $a(r, \phi, \rho)$, an estimate of the average speed at which the targets in the pulse volume are approaching or departing the radar, which is measured using the Doppler shift of the radar signal. For more details, Doviak and Zrnic (1993) is a comprehensive reference.

Figure 2 (a,b) shows example reflectivity and radial velocity sweeps during a heavy migration event at the KBGM radar in Binghamton, NY on the night of September 11, 2010. The radial velocity data demonstrate several key velocity profiling concepts: targets approaching the radar have negative radial velocities (green), whereas those departing the radar have positive radial velocities (red). For any given pixel, we have no additional information about the component of velocity orthogonal to the radar beam. However, the overall pattern of the sweep clearly shows that targets to the NNE are approaching the radar, and those to the SSW are departing the radar, so we can infer that the targets (in this case, predominantly migrating birds) are moving uniformly in a SSW direction. The spiral pattern is due to changes in target velocity with height, which are driven by changes in wind direction. The beam samples higher points in the atmosphere as it travels away from the radar, due to its upward angle and the earth's curvature. In this case, winds at lower elevations are from the north, whereas those at higher elevations have more of an easterly component, which explains why the boundary between positive and negative radial velocities rotates clockwise with distance from the radar.

Aliasing significantly complicates the interpretation of radial velocity data. Our understanding of Figure 2(b) relies on the fact that the data shown are correctly dealiased. The raw data for the same sweep are shown in Figure 2(c), in which aliasing is clearly evident. The fastest moving inbound targets to the NE of the radar are erroneously measured as having outbound velocities (red), while the fastest outbound targets to the SW are erroneously measured as having inbound velocities (green). An automatic velocity profiling algorithm must properly account for aliasing to make proper inferences.

Velocity Model. To reconstruct a complete velocity field from radial velocity data, we use a uniform wind model (Doviak and Zrnic 1993) that assumes target velocity is constant in a given elevation range. Let the true components of the velocity in the x, y, and z directions be u_0, v_0, and w_0, respectively. For a fixed elevation angle ρ and range r (and thus a fixed height z), the radial velocity as a function of azimuth is given by (Doviak and Zrnic 1993; Dokter et al. 2011):

$$a(\phi) = u_0 \cos\phi \cos\rho + v_0 \sin\phi \cos\rho + w_0 \sin\rho.$$

Our primary interest is the horizontal velocity (u_0, v_0), so we neglect the final term. This is reasonable for our purposes because w_0 is expected to be small for migrating birds, and because our analyses are restricted to those sweeps at elevation angles of 5 degrees or smaller, for which $\sin(\rho) < 0.09$. We also incorporate a Gaussian noise term $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ to obtain the model:

$$a(\phi) = u_0 \cos\phi \cos\rho + v_0 \sin\phi \cos\rho + \varepsilon. \qquad (1)$$

Note that this describes a noisy sinusoidal function of ϕ. The model is linear in the parameters u_0 and v_0, so these can be recovered by a least squares fit. Variants of this fitting approach are referred to either as *velocity-azimuthal display* (VAD) algorithms or *velocity volume profiling* (VVP) (Doviak and Zrnic 1993).

Aliasing. Aliasing significantly complicates the fitting process. For any number $a \in \mathbb{R}$, let $\overline{a} = a \mod 2V_{\max}$ be the aliased measurement of a, where we follow the convention that $\overline{a}$ lies in the interval $[-V_{\max}, V_{\max}]$ instead of $[0, 2V_{\max}]$. The values $\overline{a} + 2kV_{\max}$, $k \in \mathbb{Z}$ all produce

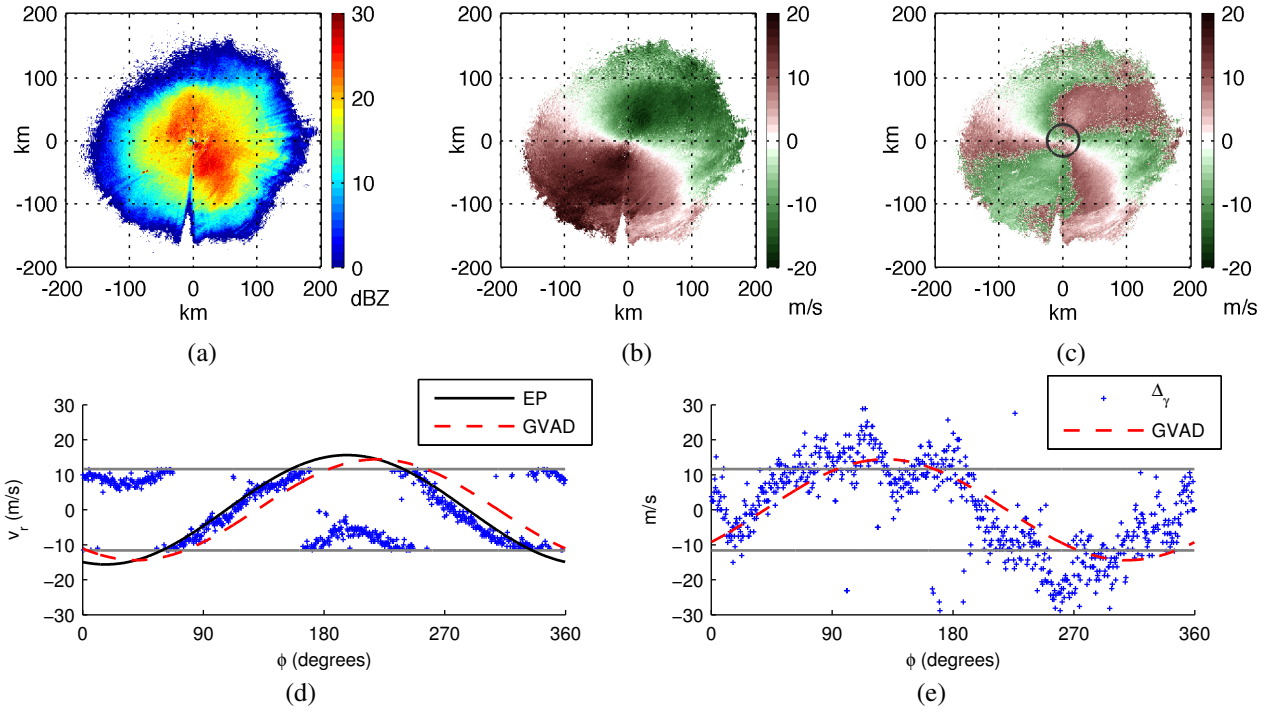

Figure 1: Radar data from KBGM station in Binghamton, NY during a heavy period of bird migration on 9/11/20100. Top row: (a) reflectivity data, (b) correctly dealiased radial velocity data, (c) original radial velocity data with clear aliasing. Bottom row: (d) radial velocity vs. azimuth at fixed range of 25 km (charcoal circle in panel (c)), together with predictions from models fitted by our method (EP) and GVAD, (e) the GVAD response variable $\Delta_\gamma(\phi)$ for $\gamma = 0.2$, together with the fitted model predictions.

the same measurement, and are called *aliases*. Note that the aliasing operation satisfies all of rules of modular arithmetic, in particular: $\overline{\overline{a} + \overline{b}} = \overline{a + b}$.

Figure 2(d) shows the aliased radial velocity as a function of azimuth at 25 km for the scan from Figure 2(c), with $V_{\max} = 11.61$ ms^{-1}. The noisy sinusoidal model is an excellent fit, but aliasing causes the curve to wrap around at $\pm V_{\max}$. A least squares fit of Equation (1) directly to the aliased data will give very poor results, which is why most existing algorithms assume that the data has already been properly dealiased.

Noise. The velocity fields of birds are more variable than those of wind-borne targets. This will have a significant impact on the fitting methods discussed in Section 3. For example, in Figure 2(c), the estimated residual standard deviation is $\sigma = 2.25$. Dokter et al. (2011) report that typical values for bird migration are $\sigma \in [2, 6]$ ms^{-1}, while typical values for precipitation are $\sigma < 2$ ms^{-1}.

3 Gradient-Based Estimation

In this section, we review, extend, and analyze gradient-based fitting methods that do not require data to be dealiased first. Tabary, Scialom, and Germann (2001) observed that the derivative of Equation (1) is also linear in (u_0, v_0):

$$a'(\phi) = -u_0 \sin\phi \cos\rho + v_0 \cos\phi \cos\rho + \varepsilon'.$$

Thus, (u_0, v_0) can be fit by least squares after first estimating $a'(\phi)$ from the data. The advantage of this approach is that $a'(\phi)$ can be estimated by fitting a locally linear model or by finite differences (Gao and Droegemeier 2004) using the *aliased* data. Within a small window around ϕ, changes in $a(\phi)$ are expected to be small, so large changes in $\overline{a(\phi)}$ can be identified as places where an aliasing boundary has been crossed and discarded from the analysis.

Our analysis extends and refines this approach. We examine the assumptions needed to properly recognize outliers as aliasing errors, and show that discarding outliers is not necessary, because they can be dealiased and used in the fitting procedure. We also highlight and quantify a tradeoff inherent in this technique. To be as robust as possible to aliasing errors, the window around ϕ used to estimate $a'(\phi)$ should be made as small as possible. However, a smaller window will increase the variance of the estimate of the response variable $a'(\phi)$, and thus the variance of the estimated parameters (u_0, v_0).

Our starting point is the gradient-based method of Gao and Droegemeier (2004), which they called GVAD for gradient-based VAD. They estimated $a'(\phi)$ by $(\overline{a(\phi+\gamma)} - \overline{a(\phi)})/\gamma$, where γ is the increment at which the antenna advances. We observe that this method can be extended to use any step size γ as follows.

Proposition 1 (GVAD). *Let* $\Delta_\gamma(\phi) = a(\phi+\gamma) - a(\phi-\gamma)$, *where* $a(\phi)$ *follows the model in Equation* (1). *Then*

$$\frac{\Delta_\gamma(\phi)}{2\sin\gamma} = -u_0 \sin\phi \cos\rho + v_0 \cos\phi \cos\rho + \varepsilon', \qquad (2)$$

where $\varepsilon' \sim \mathcal{N}(0, \sigma^2/(2\sin^2\gamma))$.

All proofs are deferred to a longer version of the paper. Equation (2) is the basis for a least squares fitting method, GVAD, which is an alternative to VAD. The original GVAD method of Gao and Droegemeier (2004) is the special case when γ is set as small as possible, to be equal to the azimuthal increment of the antenna. We will see that, for small enough γ, GVAD is robust to aliasing errors. However, Proposition 1 also shows that the variance of the noise increases by a factor of $(2\sin^2\gamma)^{-1}$ relative to Equation (1). Thus, smaller choices of γ introduce more noise. The following proposition shows that this carries through to the parameters estimated by a least squares fit.

Proposition 2. *Assume that the values $a(\phi)$ in Equation (1) are known without aliasing errors for n evenly spaced values of ϕ, and let $\hat{u}_0$ and $\hat{v}_0$ be the least squares estimates obtained from this model. Then, var$(\hat{u}_0)$ = var$(\hat{v}_0) = \frac{2\sigma^2}{n\cos^2\rho}$, and the estimates are uncorrelated. In the analogous situation using Equation (2), when the values $\Delta_\gamma(\phi)$ are known without aliasing errors for n equally spaced points, then var$(\hat{u}_0)$ = var$(\hat{v}_0) = \frac{\sigma^2}{n\cos^2\rho\sin^2\gamma}$, and the estimates are uncorrelated.*

Figure 2(e) illustrates the increased noise in the response variable for GVAD; the detrimental effect on the model fit can be seen in Figure 2(d). In this case, γ is set to a relatively large value ($\gamma = 0.2$ radians; $\sin\gamma = 0.19$). The fit is still adequate, because there are $n = 711$ measurements. The effect can be much worse when the data are sparser, which happens when many pulse volumes do not contain enough targets to permit a radial velocity measurement.

How big can γ be? There is a clear tradeoff when selecting γ: smaller values lead to more noise in the fitting process, whereas larger values lead to more aliasing errors. We now analyze this upper limit and set conditions on γ that will limit the probability of aliasing errors. Intuitively, the analysis should depend both on γ and on how quickly the function $a(\phi)$ changes, which is determined by the wind *speed* $r_0 = \sqrt{u_0^2 + v_0^2}$. Under the appropriate assumptions, it is not necessary to discard outliers in $\Delta_\gamma(\phi)$, because they can be properly dealiased and used in the model fitting process. To set up the analysis, note that the true signal we receive is not $a(\phi)$ but $\overline{a(\phi)}$, so when we measure the change over an interval, what we really observe is $\Delta = \overline{a(\phi+\gamma)} - \overline{a(\phi-\gamma)}$, which can differ from $\Delta_\gamma(\phi)$ by any integer multiple of $2V_{\max}$. By mapping Δ into the interval $[-V_{\max}, V_{\max}]$ using the aliasing operation, we obtain the aliased value of $\overline{\Delta_\gamma(\phi)}$:

$$\overline{\Delta} = \overline{\overline{a(\phi+\gamma)} - \overline{a(\phi-\gamma)}} = \overline{a(\phi+\gamma) - a(\phi-\gamma)} = \overline{\Delta_\gamma(\phi)}.$$

Now, if we have prior knowledge that the magnitude of $\Delta_\gamma(\phi)$ is very likely to be small, we can safely assume that

$$\Delta_\gamma(\phi) = \overline{\Delta_\gamma(\phi)} \in [-V_{\max}, V_{\max}]. \tag{3}$$

In other words, to dealias the measurement, we simply alias it into the Nyquist interval. The following proposition describes conditions on the interval half-width γ, the target speed r_0, the Nyquist velocity $V_{\max}$, and the noise standard deviation σ under which Equation (3) holds with high probability.

Proposition 3. *Let Φ be the cumulative distribution function of the standard normal distribution, and let $\delta(p,\sigma) = \sqrt{2}\sigma\,\Phi^{-1}(1-p/2)$. If $r_0 \le \frac{V_{\max} - \delta(p,\sigma)}{2\sin\gamma}$, then $\overline{\Delta_\gamma(\phi)} = \Delta_\gamma(\phi)$ with probability at least $1-p$.*

Thus, a safe way to choose γ is by first placing conservative upper bounds on σ and r_0, selecting an error tolerance p, and then selecting $\gamma = \sin^{-1}\left((V_{\max} - \delta(p,\sigma))/(2r_0)\right)$. We have found that this way of choosing γ works well when r_0 and σ are known and are not too large, but may not provide a useful value when conservative upper bounds are chosen. For example, a good fit to the data in Figure 2 is $r_0 \approx 15.61$, $\sigma = 2.25$, and $V_{\max} = 11.61$. Setting $p = 0.05$ in the bound gives $\gamma = .17$. However, if we do not know σ in advance, and take the not-too-conservative upper bound of $\sigma = 5$, then to get any positive bound on γ we must set $p \ge .10$; therefore, we can expect up to 10% aliasing errors in the portions of the sweep where $a(\phi)$ is changing most rapidly.

This discussion highlights the importance of having reasonable upper bounds on both the noise magnitude σ and the target speed r_0 to be able to find a value of γ that is large enough to keep noise levels under control while still being robust to aliasing. The increased variability of bird velocities (compared with wind) makes this more difficult.

4 Our Approach

In contrast with the previous approaches, we develop a probabilistic model that directly accounts for aliasing in the radial velocity values by considering the probability of each possible alias in the likelihood model. In this section, we will first present the model, and then describe our approximate inference algorithm.

Likelihood. In Equation (1), the density of the random variable $a(\phi)$ can be written as $p(a) = \mathcal{N}(a \,|\, \mu, \sigma^2)$, where $\mathcal{N}(\cdot \,|\, \mu, \sigma^2)$ is the normal density and $\mu = u_0\cos\phi\cos\rho + v_0\sin\phi\cos\rho$. The aliased random variable $\overline{a(\phi)}$ has a *wrapped normal* density (Breitenberger 1963):

$$\mathcal{N}_w(a \,|\, \mu, \sigma^2) = \sum_{k=-\infty}^{\infty} \mathcal{N}(a + 2kV_{\max} \,|\, \mu, \sigma^2),$$

which is supported on the interval $a \in (-V_{\max}, V_{\max}]$, and adds together the normal density of each alias. Our high-level approach is to use this wrapped normal density instead of the normal density in the model of Equation (1) and then fit the parameters. However, the likelihood surface is considerably more complex than that of the simple linear model, so fitting the parameters is more difficult.

For computations, it is useful to introduce the following family of finite-sum approximations to the wrapped normal density (Bahlmann 2006; Agiomyrgiannakis and Stylianou 2009):

$$\mathcal{N}_w^\ell(a \,|\, \mu, \sigma^2) = \sum_{k=-\ell}^{\ell} \mathcal{N}(\overline{a-\mu} + 2kV_{\max} \,|\, 0, \sigma^2), \tag{4}$$

which count only the contributions of the $2\ell+1$ aliases closest to the mean μ. Bahlmann (2006) reports that $\mathcal{N}_w^0(\cdot)$ is an excellent approximation when $\sigma/V_{\max} \leq 0.16$. Our examples are slightly less concentrated, so we will typically use $\mathcal{N}_w^1(\cdot)$ or $\mathcal{N}_w^2(\cdot)$.

The overall model for a complete volume scan is as follows. All pulse volumes with valid radial velocity measurements are grouped into discrete bins based on their height above ground level. Let $\mathbf{w}_i = (u_i, v_i)^T$ be the velocity vector for height level i, and let a_{ij}, ϕ_{ij}, and ρ_{ij} be the unaliased radial velocity, azimuth, and elevation angle, respectively, for the jth pulse volume at level i. Let $\mathbf{a}_i$ be the vector with entries a_{ij}, and let X_i be the matrix whose jth row is $(\cos\phi_{ij}\cos\rho_{ij}, \sin\phi_{ij}\cos\rho_{ij})$, so that $X_i\mathbf{w}_i$ is the predicted mean. By "wrapping" the normal model of Equation (1), we see that the aliased vector $\overline{\mathbf{a}}_i$ has the joint density:

$$p(\overline{\mathbf{a}_i} \,|\, \mathbf{w}_i) = \prod_j \mathcal{N}_w(\overline{a_{ij}} \,|\, (X_i\mathbf{w}_i)_j, \sigma^2). \tag{5}$$

The likelihood surface $L(\mathbf{w}_i) = p(\overline{\mathbf{a}}_i \,|\, \mathbf{w}_i)$ can be seen in examples to be multi-modal. It typically has one high peak centered near the "true" parameters, which is surrounded by an annular pattern circular ridges and valleys. Thus, any fitting method risks getting caught in a local optimum, which can cause the fitted parameters to be worse than those found by other methods despite having a better model.

Local search. Before presenting the overall inference procedure, we first describe a particularly simple local search procedure to find an approximate mode of the likelihood surface. Let $L^0(\mathbf{w}_i)$ be the approximate log-likelihood obtained using single-term approximation $\mathcal{N}_w^0(\cdot)$ to the wrapped normal. Note that

$$L^0(\mathbf{w}_i) = \max_{\mathbf{k}\in\mathbb{Z}^n} \log \mathcal{N}(\overline{\mathbf{a}}_i + 2\mathbf{k}V_{\max} \,|\, X_i\mathbf{w}_i, \sigma^2 I).$$

The problem of maximizing $L_0(\mathbf{w}_i)$ can thus be expressed as jointly maximizing over $\mathbf{w}_i$ and $\mathbf{k}$. The problem is easy when either one of these vectors is fixed. When $\mathbf{w}_i$ is fixed, the optimal $\mathbf{k}$ selects the alias of each measurement closest to the predicted mean. When $\mathbf{k}$ is fixed, the optimal $\mathbf{w}_i$ is found by linear regression with response $\mathbf{y}_i = \overline{\mathbf{a}}_i + 2\mathbf{k}V_{\max}$. It is clear that alternating these two steps will lead to a local optimum of $L^0(\mathbf{w}_i)$. Later, we will need two small extensions. First, we will need to run the analogous local search when there is also a prior $p(\mathbf{w}_i)$, in which case the linear regression at each step is conducted with a prior. We will also want to estimate the Hessian of the log-posterior at the mode, for which we use standard formulas for the (inverse of) the posterior variance in the final linear regression.

To summarize, the local search alternates between two simple high-level steps with obvious interpretations: (1) use the current model to dealias the radial velocity measurements, and then (2) refit the model using the current dealiased values. We have found that the search is very fast and usually converges in one or two steps. Our experiments show that it shows no measurable loss in overall performance compared with a much slower numerical search over the true likelihood.

Prior model. Despite the risk of local optima, most data we have examined provide overwhelming evidence about the true mode when one looks at the complete volume. Thus, we incorporate a Gaussian smoothness prior over the velocities at different elevations to help guide the fitting process toward the mode that is most compatible with neighboring elevations:

$$p(\{\mathbf{w}_i\}) \propto \prod_{i=1}^m \psi_n(\mathbf{w}_i) \prod_{i=2}^m \psi_e(\mathbf{w}_{i-1}, \mathbf{w}_i),$$

where $\psi_n(\mathbf{w}_i) = \exp(-\frac{\|\mathbf{w}_i\|^2}{2\sigma_n^2})$ and $\psi_e(\mathbf{w}_{i-1}, \mathbf{w}_i) = \exp(-\frac{\|\mathbf{w}_i - \mathbf{w}_{i-1}\|^2}{2\sigma_e^2})$ are Gaussian potentials. The parameter σ_n^2 is typically set to a large value to provide a weak prior that encourages lower target speeds; σ_e^2 controls the strength of the smoothness prior. The full graphical model, which combines the prior and the likelihood terms, is illustrated by the factor graph in Figure 2. For brevity, we collect the unary terms into the factor $t_i(\mathbf{w}_i) = \psi_n(\mathbf{w}_i)p(\overline{\mathbf{a}}_i \,|\, \mathbf{w}_i)$.

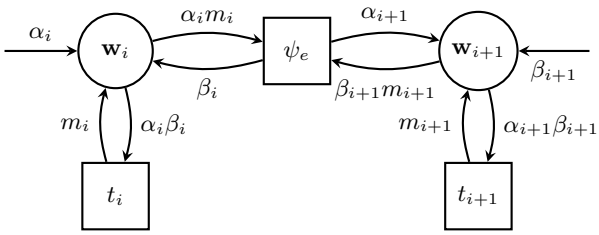

Figure 2: Message passing scheme

Message passing. The two primary operations needed to perform posterior inference by message passing in this problem are multiplication and marginalization of potentials. For two potentials $\psi_1(\mathbf{x})$ and $\psi_2(\mathbf{x})$ on the same variables, let $\psi_1\psi_2$ denote the potential obtained by pointwise multiplication, so that $(\psi_1\psi_2)(\mathbf{x}) = \psi_1(\mathbf{x})\psi_2(\mathbf{x})$. To multiply two potentials on different sets of variables, it will be understood that one first needs to expand each potential to the same set of variables, and the resulting potential is on the union of the variables. For a joint potential $\psi(\mathbf{x}_i, \mathbf{x}_j)$, let $\psi \downarrow \mathbf{x}_i$ be the marginal potential on $\mathbf{x}_i$, so that $(\psi \downarrow \mathbf{x}_i)(\mathbf{x}_i) = \int \psi(\mathbf{x}_i, \mathbf{x}_j) d\mathbf{x}_j$.

For Gaussian potentials, these operations can be done by simple parameter updates on the constituent potentials to obtain a new Gaussian potential (Murphy 2002). However, in our model the likelihood terms t_i are not Gaussian, so exact marginalization is difficult.

Instead, we adopt a message passing scheme based on the Expectation Propagation (EP) algorithm (Minka 2001), which approximates each likelihood term by a Gaussian potential that is a good match in the context of the current posterior distribution. The details of the message passing scheme are illustrated in Figure 2. Each of α_i, β_i, and m_i is a unary potential on $\mathbf{w}_i$. At any point during the execution of the algorithm, the approximate posterior over $\mathbf{w}_i$ is $\alpha_i\beta_i m_i$. The updates are:

$$\begin{aligned}
\alpha_i &\leftarrow \alpha_{i-1}m_{i-1}\psi_e \downarrow \mathbf{w}_i, \\
\beta_i &\leftarrow \beta_{i+1}m_{i+1}\psi_e \downarrow \mathbf{w}_i, \\
m_i &\leftarrow \text{approx}(\alpha_i\beta_i t_i)/\alpha_i\beta_i.
\end{aligned}$$

In the last line, division by a potential ψ is multiplication by its pointwise inverse, which is also accomplished by a simple parameter update for Gaussian potentials. The approx$(\cdot)$ operator returns a Gaussian potential that is a good approximation to its argument. In standard EP, this would be done by moment matching, but in our case, computing the moments of $\alpha_i\beta_i t_i$ is also a hard problem. Instead, we use Laplace's method: find a mode $\hat{\mathbf{w}}$ of $\alpha_i\beta_i t_i$ using the previously described local search procedure, and then set m_i to be the Gaussian with mean $\hat{\mathbf{w}}$ and covariance equal to the inverse Hessian of $-\log(\alpha_i\beta_i t_i)$ evaluated at $\hat{\mathbf{w}}$. This method is known as Laplace Propagation (Smola, Vishwanathan, and Eskin 2003).

Messages may be updated in any order, but a forward-backward scheme is logical given the chain structure. Two variants are discussed and evaluated in Section 5.

5 Experiments

We evaluated the accuracy of six different velocity profiling algorithms on volume scans collected by the KBGM radar station in Binghamton, NY during the month of September 2010. We screened an initial set of 351 scans (one per hour for each night of the month) to eliminate scans with clear non-biological targets (precipitation or ground clutter) within 37.5 km of the station. Of the remaining 142 scans, 102 operated in the clear-air mode with $V_{\max} = 11.61$ ms^{-1}, which is the lowest of all modes. We then used all pulse volumes with $\rho \leq 5$ degrees and $r \leq 37.5$ km to fit velocity vectors for each 20 m elevation bin below 6000 m. We tested the following algorithms:

- GVAD-0.01, GVAD-0.10: GVAD with $\gamma = 0.01, 0.1$, fit separately for each elevation bin.
- GVAD+LOCAL: GVAD with $\gamma = 0.1$, followed by the local search procedure described in Section 4.
- GVAD+KALMAN: same as above, but followed by forward message passing (update each α_i) and then backward message passing (update each β_i), while keeping the m_i messages frozen to the final approximate likelihoods reached during the local search. This is a Kalman filter in which each t_i is approximated by a Gaussian m_i *a priori*.
- EP: same as above, but followed by additional forward and backward passes that update the m_i messages using EP, with the local search procedure used to find the mode.
- EP+HESS: same as above, but MATLAB's numerical optimizer is used to find the mode and Hessian of $\alpha_i\beta_i t_i$ under the five-term approximation $\mathcal{N}_w(\cdot) \approx \mathcal{N}_w^2(\cdot)$.

In all methods, pulse volumes with radial velocities in the interval $[-1, 1]$ ms^{-1} were discarded to suppress ground clutter (Doviak and Zrnic 1993), and likelihood terms for elevations with fewer than 5 measurements were excluded. Parameter settings were: $\sigma = 4$, $\sigma_n = 30$, $\sigma_e = 0.36$.

For each scan, the fitted velocities were evaluated by making predictions for the scan immediately following it, and then calculating the root mean squared error (RMSE) between the predictions and the closest aliases at each elevation level. These values were averaged over all elevation levels to give the overall RMSE. Figure 3 shows the average RMSE and 95% confidence interval attained by each algorithm on the 142 test scans. All comparisons are highly significant ($p < 10^{-7}$; paired t-test) except GVAD+LOCAL vs. GVAD+KALMAN and EP vs. EP+HESS.

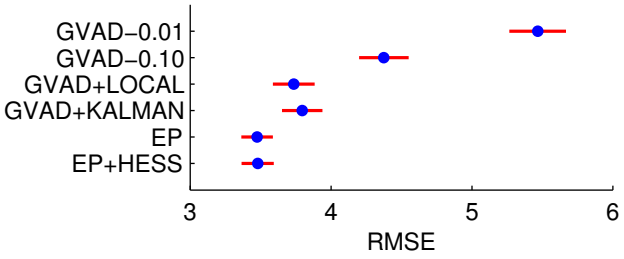

Figure 3: Performance comparison on test scans.

It is clear from the results that each of the improvements we have proposed makes a substantial improvement to performance. The fact that GVAD with $\gamma = 0.1$ is much better than GVAD with $\gamma = 0.01$ highlights the importance of our extension to allow arbitrary values of γ. By adding the simple dealias-and-refit local search procedure, GVAD+LOCAL performs much better than GVAD alone. Finally, by performing approximate Bayesian inference with the more accurate wrapped normal likelihood, EP performs significantly better than GVAD+LOCAL. The lesser performance of GVAD+KALMAN indicates that smoothing alone does not explain the better performance EP: the approach of approximating t_i in the context of the current posterior is important. Finally the performance of EP+HESS shows that EP loses no accuracy by using the approximate local search in the Laplace approximation.

We observed that the better models were generally in close agreement about the direction of travel, but the GVAD-based models seemed to underestimate speed, which is likely due to aliasing errors in the $\Delta_\gamma(\cdot)$ variables. The directions measured by the different algorithms were also in excellent agreement with human labeled directions (mean error less than 10 degrees).

6 Conclusion

Our new approach to reconstructing the velocity fields of migrating birds detected by WSR-88D is a significant improvement over previous methods. Our algorithm will allow us to overcome a fundamental challenge of analyzing radar data to tap the potential information about bird migration available from the continent-scale WSR-88D network. By creating an AI tool to automate velocity data processing, we can extract information about bird migration more accurately and at a substantially larger scale than previously possible, and make advances in our knowledge of bird movements for science and conservation.

Acknowledgments. This work was supported in part by the National Science Foundation under Grant No. 1117954 and by the Leon Levy Foundation.

References

Agiomyrgiannakis, Y., and Stylianou, Y. 2009. Wrapped Gaussian Mixture Models for Modeling and High-Rate Quantization of Phase Data of Speech. *IEEE Transactions on Audio, Speech, and Language Processing* 17(4):775–786.

Bahlmann, C. 2006. Directional features in online handwriting recognition. *Pattern Recognition* 39(1):115–125.

Breitenberger, E. 1963. Analogues of the normal distribution on the circle and the sphere. *Biometrika* 50(1):81–88.

Buler, J., and Diehl, R. 2009. Quantifying Bird Density During Migratory Stopover Using Weather Surveillance Radar. *IEEE Transactions on Geoscience and Remote Sensing* 47(8):2741–2751.

Buler, J. J.; Randall, L. a.; Fleskes, J. P.; Barrow, W. C.; Bogart, T.; and Kluver, D. 2012. Mapping wintering waterfowl distributions using weather surveillance radar. *PloS one* 7(7):e41571.

Dokter, A. M.; Liechti, F.; Stark, H.; Delobbe, L.; Tabary, P.; and Holleman, I. 2011. Bird migration flight altitudes studied by a network of operational weather radars. *Journal of the Royal Society, Interface / the Royal Society* 8(54):30–43.

Doviak, R., and Zrnic, D. 1993. *Doppler Radar and Weather Observations*. Academic Press.

Gao, J., and Droegemeier, K. 2004. A variational technique for dealiasing Doppler radial velocity data. *Journal of Applied Meteorology* (1990):934–940.

Gauthreaux, S.; Belser, C.; and Blaricom, D. V. 2003. Using a network of WSR-88D weather surveillance radars to define patterns of bird migration at large spatial scales. In Berthold, P.; Gwinner, E.; and Sonnenschein, E., eds., *Avian migration*. Springer-Verlag. 335–346.

Kunz, T. H.; Gauthreaux, S. A.; Hristov, N. I.; Horn, J. W.; Jones, G.; Kalko, E. K. V.; Larkin, R. P.; McCracken, G. F.; Swartz, S. M.; Srygley, R. B.; Dudley, R.; Westbrook, J. K.; and Wikelski, M. 2008. Aeroecology: probing and modeling the aerosphere. *Integrative and comparative biology* 48(1):1–11.

Minka, T. P. 2001. *A family of algorithms for approximate Bayesian inference*. Ph.D. Dissertation, Massachusetts Institute of Technology.

Murphy, K. P. 2002. *Dynamic bayesian networks: representation, inference and learning*. Ph.D. Dissertation, University of California, Berkeley.

Smola, A.; Vishwanathan, S.; and Eskin, E. 2003. Laplace propagation. *Neural Information Processing Systems* (Section 5).

Tabary, P.; Scialom, G.; and Germann, U. 2001. Real-Time Retrieval of the Wind from Aliased Velocities Measured by Doppler Radars. *Journal of Atmospheric and Oceanic Technology* 18(6):875–882.