Physics-informed inference of animal movements from weather radar data

Anonymous Author(s)
Affiliation
Address
email

Abstract

Studying animal movements is essential for effective wildlife conservation and conflict mitigation. For aerial movements, operational weather radars have become an indispensable data source in this respect. However, partial measurements, incomplete spatial coverage, and poor understanding of animal behaviours make it difficult to reconstruct complete spatio-temporal movement patterns from available radar data. We tackle this inverse problem by learning a mapping from high-dimensional radar measurements to low-dimensional latent representations using a convolutional encoder. Under the assumption that the latent system dynamics are well approximated by a locally linear Gaussian transition model, we perform efficient posterior estimation using the classical Kalman smoother. A convolutional decoder maps the inferred latent system states back to the physical space in which the known radar observation model can be applied, enabling fully unsupervised training. To encourage scientific consistency, we additionally introduce a physics-informed loss term that leverages known mass conservation constraints. Our experiments on synthetic radar data show promising results in terms of reconstruction quality and data-efficiency.

1 Introduction

Doppler weather radars provide high-resolution information about the distribution and movement of objects in the atmosphere. Although designed to monitor weather, radar beams are reflected not only by precipitation but also by animals passing the airspace around the radar. This offers invaluable opportunities for ecologists to study mass movements of birds, bats and insects that otherwise remain hidden due to low light conditions and high flight altitudes [3]. However, since the measurement range of radars is limited, the spatial coverage of operational weather radar networks is typically incomplete [14]. Moreover, while the amount of energy reflected back to the radar antenna can be translated directly into animal density estimates, movements can be captured only partially by measuring the *radial* velocity, i.e. the component of movement along the direction of the radar beam. Inferring the complete underlying density and velocity fields from partial observations of weather radar networks remains a challenging inverse problem that requires additional knowledge to constrain the solution space.

Unfortunately, the movement of animals is much harder to understand than that of physical particles: individual behaviours depend on a wide range of environmental and social factors, and can vary substantially within a population. That means that, unlike meteorologists, ecologists are lacking well-established mathematical equations that can be used as dynamical prior in data assimilation frameworks such as [20, 1]. Instead, strong spatio-temporal smoothness assumptions have been used

Submitted to NeurIPS 2021 AI for Science Workshop.

to reconstruct continental-scale movement patterns [2, 15], leaving more fine-scale local movements unresolved. Machine learning opens up new opportunities in this respect: deep neural networks can learn physically consistent spatio-temporal dynamics from high-dimensional data sets by integrating (partial) domain knowledge into the training process [13, 7, 22]. However, the application to ecological systems, and animal movements in particular, is hindered by the size of available data sets (particularly in Europe [19]), as well as the lack of ground-truth data.

Inspired by recent work on combining deep learning with exact inference in linear Gaussian state space models [4, 12, 18, 17], we propose a data-efficient approach to reconstruct high-dimensional time-varying animal density and velocity fields from partial and noisy radar data. We jointly model densities and velocities in a learned feature space in which the dynamics are well approximated by a locally linear Gaussian transition model. In this space, inference can be performed with the classical Kalman filter or smoother [9, 16]. The mapping between physical space and feature space is learned together with the latent dynamics in an unsupervised way, minimizing both the reconstruction loss and a physics-informed loss that exploits known mass conservation constraints based on the continuity equation. This allows us to reconstruct high-resolution movement patterns while having (i) incomplete knowledge about the underlying process, (ii) no access to ground-truth fields, and (iii) only a limited number of training samples to learn a surrogate for the complex spatio-temporal dynamics.

2 Background

In this section, we briefly describe relevant prior knowledge about mass movements of animals and discuss how weather radars can be used to quantify such processes.

Modelling mass movements The coordinated movement of large number of animals, e.g. during migration, can be described in terms of individuals moving along a shared velocity field $\mathbf{v}$ that changes over time depending on weather conditions, landscape features, food availability, social interactions, etc [15]. The resulting spatio-temporal distribution of animals is restricted by the continuity equation

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\mathbf{v}\rho), \tag{1}$$

where ρ denotes animal density. The spatio-temporal dynamics of the velocity field $\mathbf{v}$ and the underlying drivers, however, remain largely unknown and need to be estimated from data.

Radar basics For birds, bats and insects, continent-wide networks of Doppler weather radars have become an invaluable tool to quantify mass movements over large spatial and temporal extents [3]. Every 5-15 min, Doppler weather radars perform multiple $360°$ sweeps at different elevation angles to sample the three dimensional airspace around the antenna. With every sweep, several data products are obtained. In this work, we are specifically interested in *reflectivity*, from which the animal density ρ can be estimated, and *radial velocity*, describing associated movements towards or away from the antenna derived from the phase shift of reflected radio waves. More precisely, the radial velocity $r(\mathbf{x}) = \mathbf{a}(\mathbf{x})^T\mathbf{v}(\mathbf{x})$ of a moving object at location $\mathbf{x}$ is the projection of the full velocity vector $\mathbf{v}(\mathbf{x})$ onto the unit vector $\mathbf{a}(\mathbf{x})$ pointing in the direction of the radar beam. In general, the geometry of radar measurements in combination with scattering effects restricts data collection to some maximum distance from the antenna. Hence, we can define the forward model $\mathcal{H}$ mapping the true fields $\mathbf{v}$ and ρ to measurements taken by radar n:

$$\mathcal{H}(\mathbf{v}, \rho, \mathbf{A}_n, d_n) = [f_{<d_n}(\mathbf{A}_n\mathbf{v}), f_{<d_n}(\rho)] = [\tilde{\mathbf{r}}_n, \tilde{\boldsymbol{\rho}}_n] \tag{2}$$

where $f_{<d_n}(\cdot)$ represents the partial (i.e. masked) observation of a quantity up to distance d_n from the radar, and the rows of matrix $\mathbf{A}_n$ contain the unit vectors $\mathbf{a}_n(\mathbf{x})$ associated with radial velocity measurements taken by radar n.

Velocity field reconstruction Unless measurements of multiple radars are spatially overlapping, the reconstruction of $\mathbf{v}$ from $\{\tilde{\mathbf{r}}_n\}_{n=1}^N$ presents an under-determined inverse problem. To constrain the

solution space, ecological studies traditionally assume movements to be uniform. The vector field $\mathbf{v}$ then reduces to a single vector which can easily be estimated with *velocity volume profiling* techniques [21, 8]. In practice, however, animal movements are heterogeneous: local weather conditions, barriers, food availability, etc. lead to high variability in both movement directions and speeds. Angell et al. [2] made a first step towards reconstructing spatially detailed velocity fields by simultaneously modelling radial velocity measurements from many spatially distributed radars with a Gaussian process model that encodes smoothness assumptions about the latent velocity field. While this allows ecologists to recover large-scale movements across continents, more fine-scale patterns are neglected.

3 Method

Consider a time series of radar observations $(\mathbf{o}_1, ..., \mathbf{o}_T)$ with $\mathbf{o}_t = [\tilde{\mathbf{r}}_t, \tilde{\mathbf{q}}_t]$ consisting of radial velocity measurements $\tilde{\mathbf{r}}_t \in \mathbb{R}^{K \times L \times N}$ and log-transformed density measurements $\tilde{\mathbf{q}}_t = \log(\tilde{\boldsymbol{\rho}}_t) \in \mathbb{R}^{K \times L \times N}$ on a two-dimensional $K \times L$ grid, taken by N different radars with known projections $\mathbf{A} = [\mathbf{A}_1, ..., \mathbf{A}_N]$ and ranges $\mathbf{d} = [d_1, ..., d_N]$. Given this partial and noisy data, we aim at reconstructing the true underlying velocity and log-density fields $\{\mathbf{v}_t, \mathbf{q}_t\}_{t=1}^T$.

Encoding radar measurements Inspired by traditional reduced order modelling approaches for high-resolution fluid dynamics [5, 6], we assume that animal movements can be compressed to a much lower dimensional space. To learn this unknown space from data, we use a convolutional encoder network $f_{\text{enc}}(\mathbf{o}_t, \mathbf{A}) = \mathbf{w}_t$ that maps high-dimensional measurements $\mathbf{o}_t$ together with the associated radar projections $\mathbf{A}$ to a latent measurement $\mathbf{w}_t \in \mathcal{W} \subseteq \mathbb{R}^M$ capturing key spatial features of the measured fields.

Linear Gaussian state space model Apart from spatial features, animal movements are characterised by strong temporal dependencies. Moreover, radar measurements can be noisy or contain artefacts due to reflections by other objects in the atmosphere. To account for both temporal dynamics and measurement errors, we model latent measurements $(\mathbf{w}_1, ...\mathbf{w}_T)$ with a linear Gaussian state space model (LGSSM)

$$p(\mathbf{z}, \mathbf{w}) = \mathcal{N}(\mathbf{z}_1 \mid \mu_1, \boldsymbol{\Sigma}_1) \prod_{t=2}^{T} \mathcal{N}(\mathbf{z}_t \mid \mathbf{F}_t \mathbf{z}_{t-1}, \mathbf{P}) \prod_{t=1}^{T} \mathcal{N}(\mathbf{w}_t \mid \mathbf{H}\mathbf{z}_t, \mathbf{R}), \tag{3}$$

which links latent measurements to a latent representation $\mathbf{z}_t \in \mathcal{Z} \subseteq \mathbb{R}^D$ of the true physical state $(\mathbf{v}_t, \mathbf{q}_t)$. Here, $\mathbf{H}$ and $\mathbf{P}$ are predefined, $\mu_1, \boldsymbol{\Sigma}_1$, and $\mathbf{R}$ are learned parameters, and $\mathbf{F}_t$ is a linear combination of C constant transition matrices $\mathbf{F}^{(k)}$ with coefficients defined by a small neural network that takes the current latent state $\mathbf{z}_t$ as input.

Inference Given a sequence of encoded measurements $(\mathbf{w}_1, ..., \mathbf{w}_T)$ and the LGSSM from eq. 3, we can perform efficient inference in latent space using the standard Kalman smoother [16]. This results in latent state posterior estimates $\mathbf{z}_1^+, ..., \mathbf{z}_T^+$ and corresponding covariances $\boldsymbol{\Sigma}_1^+, ..., \boldsymbol{\Sigma}_T^+$. Finally, to recover the physical system states from some latent state $\mathbf{z}_t$, we use a convolutional decoder network to learn the inverse mapping $f_{\text{dec}}(\mathbf{z}_t) = [\mathbf{v}_t, \mathbf{q}_t]$ from the latent state space $\mathcal{Z}$ back to the physical space $\mathcal{P} \subseteq \mathbb{R}^{K \times L \times 3}$.

Training Since in practice it is impossible to obtain data on the true physical states $\{\mathbf{v}_t, \mathbf{q}_t\}_{t=1}^T$, we rely on unsupervised learning. That means, we minimize the reconstruction loss

$$\mathcal{L}_{\text{recons}} = \frac{1}{T} \sum_{t=1}^{T} ||\mathbf{o}_t - \mathcal{H}(f_{\text{dec}}(\mathbf{z}_t^+), \mathbf{A}, \mathbf{d})||_2^2, \tag{4}$$

where we first map each posterior estimate $\mathbf{z}_t^+$ back to physical space, and then apply the known forward model $\mathcal{H}$ (see eq. 2) to obtain the corresponding reconstruction of radar measurements.

Incorporating physical constraints To encourage scientific consistency, we regularize our model based on the continuity equation (see section 4). In particular, we map log-densities $\mathbf{q}_t$ back to densities ρ_t, and discretize equation 1 in time (using forward Euler) and in space (using finite differences):

$$\rho_{t+1}^{(k,l)} - \rho_t^{(k,l)} \approx -\Delta t \cdot \left(\frac{\partial j_{t,x}}{\partial x} + \frac{\partial j_{t,y}}{\partial y} \right)^{(k,l)} \tag{5}$$

$$\approx -\Delta t \cdot \left(\frac{j_{t,x}^{(k+1,l)} - j_{t,x}^{(k-1,l)}}{2\Delta x} + \frac{j_{t,y}^{(k,l+1)} - j_{t,y}^{(k,l-1)}}{2\Delta y} \right), \tag{6}$$

with $\mathbf{j}_t^{(k,l)} = \left[j_{t,x}^{(k,l)}, j_{t,y}^{(k,l)} \right]^T = \mathbf{v}_t^{(k,l)} \cdot \rho_t^{(k,l)}$ the flux at time t and grid cell (x_k, y_l). The resulting physical loss term

$$\mathcal{L}_{\text{physics}} = \frac{1}{T} \sum_{t=1}^{T} \sum_{k=2}^{K-1} \sum_{l=2}^{L-1} \left\{ \rho_{t+1}^{(k,l)} - \rho_t^{(k,l)} + \Delta t \left(\frac{\mathbf{j}_t^{(k+1,l)} - \mathbf{j}_t^{(k-1,l)}}{2\Delta x} + \frac{\mathbf{j}_t^{(k,l+1)} - \mathbf{j}_t^{(k,l-1)}}{2\Delta y} \right) \right\} \tag{7}$$

is combined with the reconstruction loss to form the final training objective

$$\min_{\theta} \; \mathcal{L}_{\text{recons}} + \mathcal{L}_{\text{physics}}, \tag{8}$$

where θ includes encoder, decoder, and LGSSM parameters.

4 Experiments and results

Synthetic data generation Besides noisy and partial measurements, working with real weather radar data comes with several additional challenges, such as aliasing, clutter effects, and discriminating biology from weather. Moreover, the lack of ground truth fields prevents a thorough evaluation of model reconstructions. We thus generate a synthetic radar data set to train and evaluate our proposed methodology (see appendix for details). In particular, we generate 1000 training sequences and 100 test sequences of time-varying 2D velocity and animal density fields with a resolution of 32×32 and length $T = 20$ (each corresponding to ca. 2-6h of radar data). The associated partial and noisy measurements $\{\tilde{\rho}_n, \tilde{\mathbf{r}}_n\}_{n=1}^3$ of 3 randomly positioned radars are obtained by applying the respective forward model and adding independent Gaussian noise with $\sigma = 0.001$.

Evaluation and baselines We compare our method against the following baselines: A convolutional VAE [11] is used to jointly model velocities and log-densities, without capturing temporal dependencies. The encoder and decoder architecture as well as the dimensionality of the latent space matches our proposed model. Further, *velocity volume profiling* (VVP) [21] is used to reduce the set of radial velocity measurements of each radar to a single velocity vector describing a uniform velocity field around the radar. An estimate of the full velocity field is obtained by linearly interpolating the resulting N velocity vectors in space. All methods are evaluated based on their reconstruction quality, measured by the root mean squared error (RMSE) between model reconstruction and ground-truth, for velocities and log-transformed densities respectively.

Experiments We evaluate all methods for radar ranges varying from $d_n = 1$ ($64 \pm 13\%$ coverage) to $d_n = 2$ ($98 \pm 4\%$ coverage) to $d_n = \infty$ (100% coverage). The results, summarized in table 1, show that for all considered ranges our method yields the lowest RMSE for both velocities and log-densities. Fig. 1 visually compares the different methods based on an example velocity field from the test set. Our method is well able to capture the patterns of both direction and speed of movement, whereas the two baseline methods fail to reconstruct some of the local structures.

In a second experiment, we evaluate the data-efficiency of our method by gradually changing the number of training sequences from 100 to 1000. As before, radar ranges $d_n \in \{1, 2, \infty\}$ are considered. While in general the RMSE increases as the number of training sequences decreases,

	d_n	VVP	VAE	Ours
velocities $\mathbf{v}$	1	0.1704	0.2193 ± 0.0032	**0.1157** ± 0.0015
	2	0.1704	0.2139 ± 0.0007	**0.0885** ± 0.0047
	∞	0.1704	0.2031 ± 0.0046	**0.0600** ± 0.0050
log-densities $\mathbf{q}$	1	–	0.1368 ± 0.0075	**0.1146** ± 0.0028
	2	–	0.1222 ± 0.0031	**0.0618** ± 0.0058
	∞	–	0.1249 ± 0.0062	**0.0317** ± 0.0012

Table 1: RMSE (mean $\pm$ standard deviation using 3 different seeds) for reconstructed velocities and log-densities, evaluated for three different radar ranges d_n.

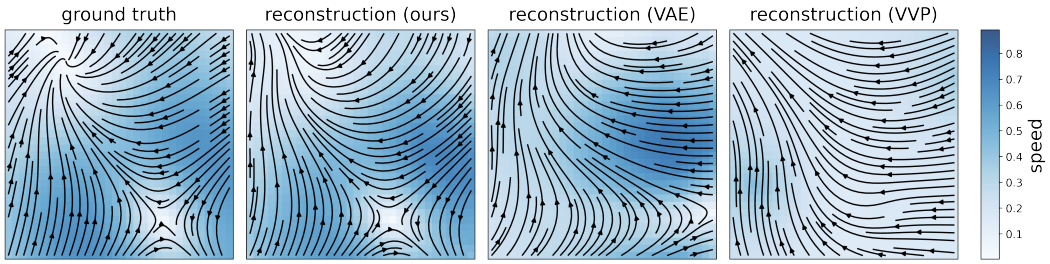

Figure 1: Qualitative comparison of three different reconstruction methods.

we are able to obtain more accurate velocity reconstructions than the VVP baseline (which does not require any training) using as little as 200 training sequences for ranges $d_n \in \{2, \infty\}$ and 500 sequences for $d_n = 1$ (see fig. 2). In comparison, the VAE cannot beat the VVP baseline, even when trained on 1000 sequences.

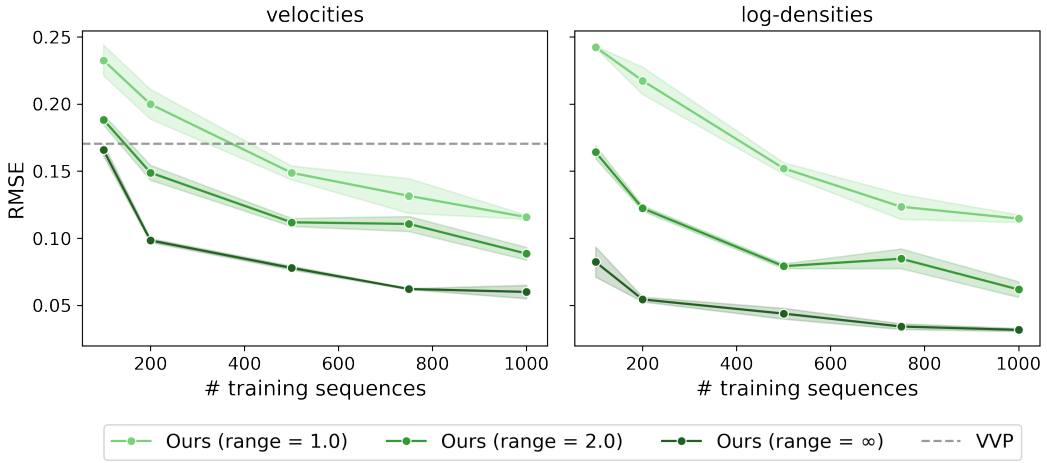

Figure 2: RMSE for velocities (left) and log-densities (right) reconstructed by our method, using a varying number of training sequences. Shaded regions correspond to one standard deviation around the mean using 3 different seeds.

5 Conclusion

In this work, we combined deep spatial feature learning with exact inference in linear Gaussian state space models to reconstruct high-dimensional time-varying animal movement patterns from partial and noisy Doppler weather radar data, without requiring in-depth knowledge about the underlying processes or access to ground-truth data. Our results indicate that this approach is able to reconstruct

movement patterns in more detail than traditional approaches based on uniformity assumptions, while requiring less data than models neglecting spatio-temporal movement dynamics. Our experiments are limited to synthetic radar data, and several additional challenges need to be tackled before applying the proposed method to real weather radar data. However, we are hopeful that future work in this direction will eventually uncover aerial movement patterns of birds, bats and insects at various spatial scales. In the long term, this opens up opportunities for ecologists to obtain novel scientific insights into the underlying behaviours and movement strategies, and to develop more effective strategies for species conservation and mitigation of human-wildlife conflicts.

References

[1] A. Aksoy, D. C. Dowell, and C. Snyder. A multicase comparative assessment of the ensemble kalman filter for assimilation of radar observations. part i: Storm-scale analyses. *Monthly Weather Review*, 137:1805–1824, 2009.

[2] R. Angell and D. Sheldon. Inferring latent velocities from weather radar data using gaussian processes. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2018.

[3] S. Bauer, J. W. Chapman, D. R. Reynolds, J. A. Alves, A. M. Dokter, M. M. H. Menz, N. Sapir, M. Ciach, L. B. Pettersson, J. F. Kelly, H. Leijnse, and J. Shamoun-Baranes. From agricultural benefits to aviation safety: realizing the potential of continent-wide radar networks. *BioScience*, 67:912–918, 2017.

[4] P. Becker, H. Pandya, G. Gebhardt, C. Zhao, C. J. Taylor, and G. Neumann. Recurrent kalman networks: Factorized inference in high-dimensional deep feature spaces. In *International Conference on Machine Learning (ICML)*, pages 544–552. PMLR, 2019.

[5] J. Burkardt, M. Gunzburger, and H.-C. Lee. Pod and cvt-based reduced-order modeling of navier–stokes flows. *Computer methods in applied mechanics and engineering*, 196(1-3):337–355, 2006.

[6] K. Carlberg, C. Bou-Mosleh, and C. Farhat. Efficient non-linear model reduction via a least-squares petrov–galerkin projection and compressive tensor approximations. *International Journal for numerical methods in engineering*, 86(2):155–181, 2011.

[7] H. Gao, L. Sun, and J. X. Wang. Super-resolution and denoising of fluid flow using physics-informed convolutional neural networks without high-resolution labels. *Physics of Fluids*, 33, 7 2021.

[8] H. V. Gasteren, I. Holleman, W. Bouten, E. V. Loon, and J. Shamoun-Baranes. Extracting bird migration information from c-band doppler weather radars. *Ibis*, 150:674–686, 2008.

[9] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82:35–45, 1960.

[10] D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[11] D. P. Kingma and M. Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations (ICLR)*, 2014.

[12] A. Klushyn, R. Kurle, M. Soelch, B. Cseke, and P. van der Smagt. Latent matters: Learning deep state-space models. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

[13] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, S. Hoyer, and A. L. Bertozzi. Machine learning-accelerated computational fluid dynamics. *PNAS*, 118, 2021.

[14] B. Kranstauber, W. Bouten, H. Leijnse, B. Wijers, L. Verlinden, J. Shamoun-baranes, and A. M. Dokter. High-resolution spatial distribution of bird movements estimated from a weather radar network. *Remote Sensing*, 12, 2020.

[15] R. Nussbaumer, S. Bauer, L. Benoit, G. Mariethoz, F. Liechti, and B. Schmid. Quantifying year-round nocturnal bird migration with a fluid dynamics model. *Journal of the Royal Society Interface*, 18, 2021.

[16] H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450, 1965.

[17] G. Revach, N. Shlezinger, R. J. van Sloun, and Y. C. Eldar. Kalmannet: Data-driven kalman filtering. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.

[18] D. Ruhe and P. Forré. Self-supervised inference in state-space models. In *International Conference on Learning Representations (ICLR)*, 2022.

[19] J. Shamoun-Baranes, S. Bauer, J. W. Chapman, P. Desmet, A. M. Dokter, A. Farnsworth, B. Haest, J. Koistinen, B. Kranstauber, F. Liechti, T. H. Mason, C. Nilsson, R. Nussbaumer, B. Schmid, N. Weisshaupt, and H. Leijnse. Weather radars' role in biodiversity monitoring. *Science*, 372:248–248, 2021.

[20] J. Sun and N. A. Crook. Real-time low-level wind and temperature analysis using single wsr-88d data. *Weather and Forecasting*, 16:117–132, 2001.

[21] P. Waldteufel and H. Corbin. On the analysis of single-doppler radar data. *Journal of Applied Meteorology and Climatology*, 18(4):532–542, 1979.

[22] R. Wang and R. Yu. Physics-guided deep learning for dynamical systems: A survey. *arXiv preprint arXiv:2107.01272*, 2021.

Appendix

A Data generation

We generate synthetic velocity fields by describing animal movements as the interplay between attraction to areas with food, mild climate, etc. and repulsion from geographical barriers, predators and hostile environmental conditions. That means, we assign each point in time and space a scalar value $\phi(t, \mathbf{x})$, representing the associated attractiveness (or potential). We initialize $\phi(0, \mathbf{x})$ as a mixture of 10 two-dimensional Gaussians with randomly sampled mean and diagonal covariance. Given $\phi(t, \mathbf{x})$, we generate auto-correlated potentials at time $t + 1$, by shifting each mode by some fixed displacement vector. Finally, the velocity fields associated with $\phi(t, \mathbf{x})$ are computed as $\mathbf{v}(t, \mathbf{x}) = -\nabla\phi(t, \mathbf{x})$.

Similar to $\phi(0, \mathbf{x})$, we initialize animal densities as a mixture of 10 two-dimensional Gaussians. Then, we simulate eq. 1 forward in time using a *forward time centered space* finite difference scheme with $dt = 0.001$ to generate a time-varying density fields $\rho(t, \mathbf{x})$ that are consistent with the time-varying velocity field $\mathbf{v}$. The resulting fields are sampled at $\Delta t = 0.025$ to obtain the final time series.

In practice, density distributions are highly skewed. To obtain approximately normally distributed inputs, we work with log-transformed densities instead. Before feeding $\mathbf{v}$ and $\log(\rho)$ to our model, we rescale them to values between -1 and 1. Measurements outside the range d_n of radar n are masked by setting them to zero. The corresponding vectors in the projection matrix $\mathbf{A}_n$ are set to $\mathbf{0}$ as well.

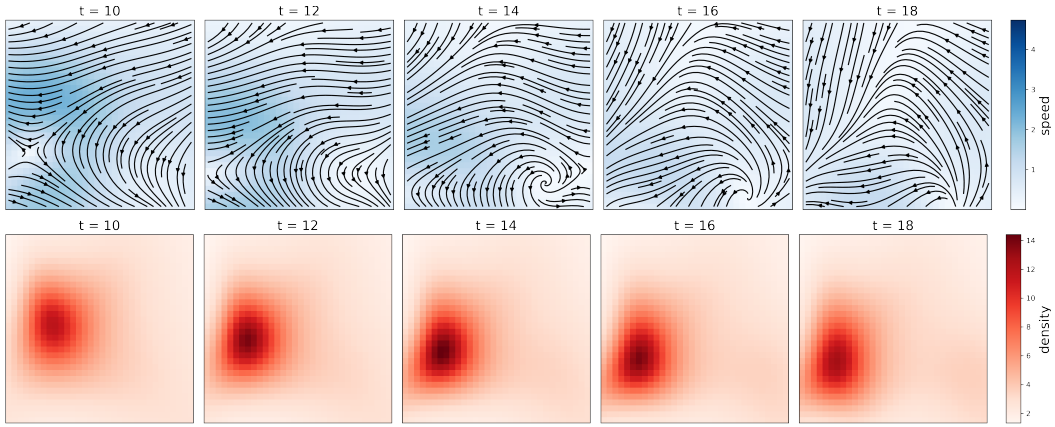

Figure 3: Samples from a time series of synthetic velocity (top row) and density (bottom row) fields.

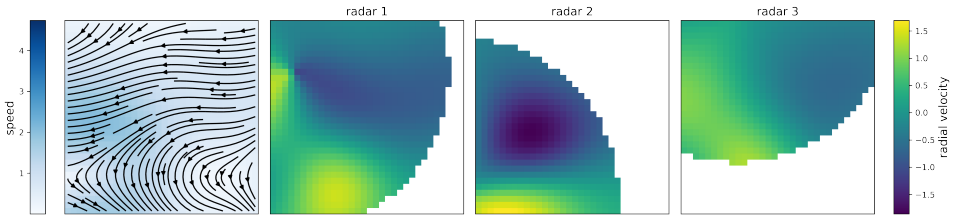

Figure 4: A full velocity field (left), and corresponding radial velocities measured by 3 different radars with range $d = 2$ (right).

B Architectures and hyperparameters

Convolutional encoder and decoder In all our experiments (including VAE baseline), we use a convolutional encoder network consisting of 3 layers with 3×3 filters, 1×1 stride, ReLU non-linearities, and 2×2 max-pooling. The number of output channels are 32, 64, and 128 respectively. A fully-connected layer with linear output maps the encoder output to a $M = 128$ dimensional latent measurement. Following the structure of the encoder, we decode latent states with dimension $D = 128$ by first applying a fully-connected layer with linear output, followed by 3 convolutional layers where in each layer we first upsample the input by factor 2, and then apply a 3×3 filter with 1×1 stride, followed by a ReLU non-linearity (for all but the last layer).

LGSSM We assume both $\mathbf{P}$ and $\mathbf{R}$ to be diagonal covariance matrices. We define $\mathbf{P} = 0.1 \cdot \mathbf{I}_D$, while the diagonal σ_R is learned. The latent measurement operator $\mathbf{H}$ is set to the identity matrix $\mathbf{I}_D$, assuming the latent measurement space $\mathcal{W}$ and the latent state space $\mathcal{Z}$ to have the same dimensionality $M = D = 128$. Further, we define the prior for time point $t = 1$ as $\mathcal{N}(\mathbf{z}_1 \mid \mathbf{0}, 10 \cdot \mathbf{I}_D)$, expressing high uncertainty about the initial state. Finally, we define the transition matrix $\mathbf{F}_t$ as

$$F_t = \sum_{k=1}^{C} \alpha_k(\mathbf{z}_t)\mathbf{F}^{(k)},$$

with $C = 8$ and $\alpha_1, ..., \alpha_C$ the outputs of a 2-layer MLP with softmax activation.

Training We trained all models for a maximum of 100 epochs using the Adam optimizer [10] with default settings. The learning rate is set to 0.001 (based on grid search).

C Additional results

Based on latent posterior estimates $\mathcal{N}(\mathbf{z}_t^+, \mathbf{\Sigma}_t^+)$, we can quantify the uncertainty associated with the reconstructed velocity and log-density fields by repeatedly sampling $\hat{\mathbf{z}}_t \sim \mathcal{N}(\mathbf{z}_t^+, \mathbf{\Sigma}_t^+)$ and decoding $\hat{\mathbf{z}}_t$ to obtain a set of physical reconstructions. Computing the average per-grid-cell standard deviation for each time point reveals higher uncertainties at the beginning and of each sequence, for velocities and log-densities respectively (see fig. 5). This matches the generally higher RMSE at the beginning and end of each sequence and indicates that our method relies on information about both past and future to generate accurate reconstructions.

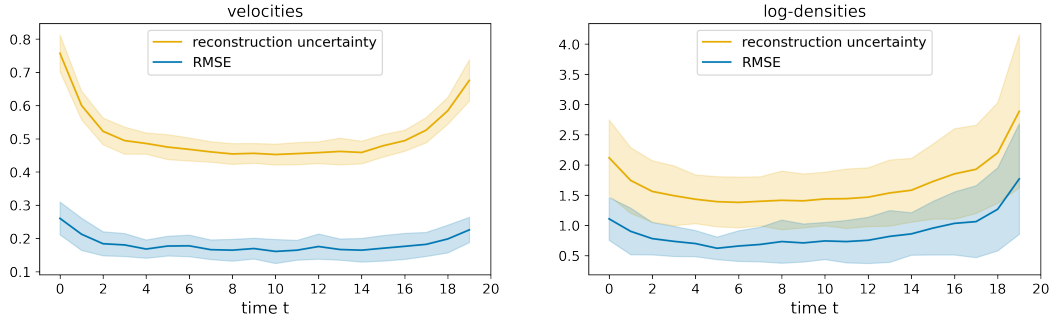

Figure 5: Reconstruction error (RMSE) and uncertainty as a function of time, using our method with radar range $d_n = 2$ and 1000 training sequences. Uncertainties are computed as the standard deviation over reconstructions based on 10 samples from the latent posterior $\mathcal{N}(\mathbf{z}_t^+, \mathbf{\Sigma}_t^+)$. Shaded regions correspond to one standard deviation around the mean computed for 100 test sequences.

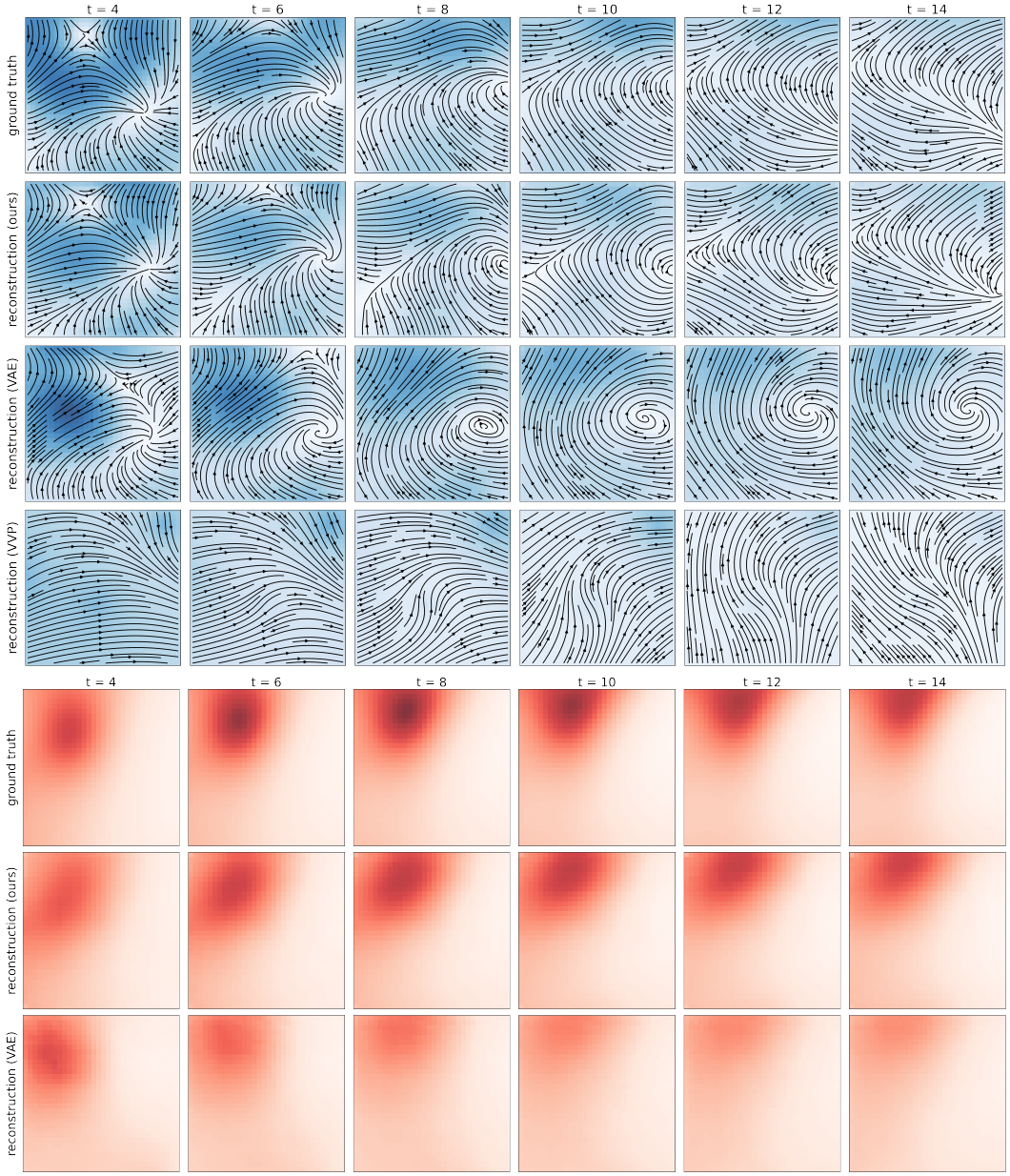

Figure 6: Example time series of synthetic velocity (blue) and density (red) fields, together with corresponding reconstructions.