Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **539,** 2498–2512 (2025) https://doi.org/10.1093/mnras/staf576
Advance Access publication 2025 April 11

Recovering coherent flow structures in active regions using machine learning

Matthew G. Lennard [1★], Suzana S. A. Silva,[1] Benoit Tremblay,[2] Andrés Asensio Ramos,[3] Gary Verth,[4] Istvan Ballai,[4] Haruhisa Iijima,[5] Hideyuki Hotta [5], Matthias Rempel,[3] Sung-Hong Park[6] and Viktor Fedun[1]

[1] *Plasma Dynamics Group, School of Electrical and Electronic Engineering, The University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK*
[2] *High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO, 80307, USA*
[3] *Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[4] *Plasma Dynamics Group, School of Mathematics and Physical Sciences, The University of Sheffield, Hicks Building, Hounsfield Road, Sheffield, S3 7RH, UK*
[5] *Institute for Space-Earth Environmental Research, Nagoya University, Furocho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan*
[6] *Space Science Division, Korea Astronomy and Space Science Institute, Daejeon 305-348, Republic of Korea*

Accepted 2025 April 4. Received 2025 April 4; in original form 2024 November 13

ABSTRACT

Analysing high-resolution solar atmospheric observations requires robust techniques to recover plasma flow features across different scales, especially in active regions. Current methodologies often fall short in capturing subgranular-scale flows, and there is limited research on the errors introduced by velocity estimation techniques and analysing the properties of recovered flows in the presence of kG magnetic flux density. This study concentrates on validating the effectiveness of the DeepVel neural network in recovering subgranular to mesogranular-scale topological plasma flow features throughout the total evolution of a simulated active region by tracking tracers, and reproducing coherent patterns. The neural network was trained on the R2D2 radiative MHD simulation depicting the emergence and decay of a magnetic flux tube. DeepVel achieved strong correlations (exceeding 0.7) with flows from an unseen MURAM simulation, despite being trained on a model with a simpler radiative transfer and lacking thermal resistivity. DeepVel was able to capture the detailed topology well, e.g. the structure of vortical and diverging structures across all scales present in the flows. DeepVel performed slightly less well in the umbra, this is likely explained by magnetic field suppression and reduced contrast. Differences in velocities introduced by DeepVel did not affect Lagrangian analysis; consequently, we demonstrate for the first time that the DeepVel-recovered velocities accurately reflected the flow's transport barriers. These findings highlight the precision and reliability of the DeepVel and its ability to emulate plasma flows surrounding and within active regions.

Key words: methods: data analysis – Sun: activity – Sun: granulation – Sun: magnetic fields – Sun: photosphere – sunspots.

1 INTRODUCTION

The appearance and development of large-scale energetic phenomena in the solar atmosphere, such as active regions (ARs) and flaring events, are strongly dependent on the interplay of plasma and magnetic fields. We adopt the definition of an AR as defined by van Driel-Gesztelyi & Green (2015), namely, we will refer to ARs as being pores in the photosphere created by the presence of more than kG magnetic flux strength. These phenomena are driven by changes in the magnetic field topology, i.e. twisting of the magnetic field as well as compression of magnetic flux, which are in part consequence of the influence of advective motions of plasma throughout sub-photospheric layers. Photospheric flow topology is, therefore, expected to manifest distinct changes due to the presence and restructuring of the magnetic field topology, when compared with flows in regions with little or no magnetic flux present (see, e.g. Attie et al. 2018). Thus, analysis of photospheric surface plasma flows and their dynamics is an essential part in the way of understanding the formation and evolution of complex magnetic structures in the solar atmosphere (see, e.g. Driel-Gesztelyi & Green 2015).

Presently, direct observation of plasma motions is restricted to line-of-sight component via the Doppler effect, however these measurements are not precise due to measuring asymmetric spectral lines that are a result of a range of phenomena which occur in the highly stratified atmosphere of the Sun (see, e.g. Martínez-Sykora et al. 2011; Löhner-Böttcher et al. 2019). The identification of the horizontal velocity components from observational data remains a challenging task (see Rempel et al. 2022) and several techniques have been proposed to deal with this problem (Tziotziou et al. 2023).

Optical flow tracking methods such as local correlation tracking (LCT) (see November & Simon 1988) have been used for the estimation of velocity fields. These methods assume that features in images are advected by the flow, hence one can produce a velocity field by cross-correlating features within time-consecutive sub-images. LCT has been shown to provide velocities that are

★ E-mail: mlennard97@gmail.com

consistent with the proper motions of granules at length scales $\gtrsim$2.5 Mm and for times $\gtrsim$30 min (see, e.g. Rieutord et al. 2001), however it produces velocities that are underestimated by a factor of 1/3 (Verma, Steffen & Denker 2013). For optical tracking, images with distinct and highly contrasting features are required. Spatial smoothing and time-averaging of images is used to reach the optimal spatial and temporal scales, which also has the added benefit of removing small-scale perturbations from the images. Thus these are best suited for recovering the motions of photospheric granules.

Other methods such as physical inversions like DAVE4VM (Schuck 2008) have also been produced. This solves the ideal induction equation

$$\partial_t B = \nabla \times (\boldsymbol{v} \times B), \tag{1}$$

for the horizontal part of the velocity vector using measurements from vector magnetograms. Inversions such as these are computationally expensive and sensitive to noise in magnetic field vector measurements, thus there are limited data sets that may be used with this method.

Artificial neural networks, often just referred to as neural networks (NNs) present methods for learning patterns in data. NNs achieve this by attempting to mimic biological NNs, weighting connections between layers of neurons in the network in order to produce a model relating some input data to an output. Therefore NNs can be designed to be highly effective at performing a specified task. This does however introduce a limitation, it is a possibility that a NN will only learn the training data and not be able to generalize to new data. This is called overfitting.

Recently, the DeepVel network (DV; see Asensio Ramos, Requerey & Vitas 2017) has been developed for the use of recovering two- or three-dimensional velocity fields from intensity or magnetogram data. Typically, simulations provide a ground truth for the training set of the network, which may be processed to match telescope imagery for use with real world observations. In the current literature, DV has been shown successful in recovering velocities from synthetic images that mimic the Solar Dynamics Observatory/Helioseismic and Magnetic Imager (SDO/HMI) (Tremblay et al. 2018), and SUNRISE Imaging Magnetograph eXperiment (IMaX) (Asensio Ramos et al. 2017), which have resolutions of 1 arcsec (or 368 km pix^{-1}; e.g. see Tremblay et al. 2018) and 0.11 arcsec (or 39.9 km pix^{-1}; Orozco Suárez et al. 2010; Asensio Ramos et al. 2017).

The study by Tremblay et al. (2018) compared the effectiveness of DV with the optical flow methods LCT, Fourier-based LCT (FLCT) and coherent structure tracking (CST). These methods were applied to synthetic photospheric observations at a resolution of the SDO/HMI instrument ($\delta x = 0.5$ arcsec pix^{-1}) averaged over 30 min. DV was able to reproduce velocities best at granular and subgranular scales in the quiet Sun (QS). However, FLCT performed better at the mesogranular scales. With this and the fact that DV is limited to QS data, we are unable to see the full capabilities of DV and the range of scenarios it may be applied. Later studies improved the studies by introducing a new architecture to DV (Tremblay & Attie 2020) and testing this and the original DV architecture on recovering flows from a sunspot at the same resolution as HMI (Tremblay et al. 2021). In both instances, only the 30-min time-averaged velocity fields are considered in the analysis, despite DV being trained on significantly higher cadences of $\approx$45 s. The sunspot used in the AR study is from a MURAM simulation (see Rempel 2015), which presents the decay of an axis-symmetric cylinder of magnetic flux embedded vertically in the box that covers 18 Mm depth. After an initial relaxation period of around 5.5 h the flux tube is left to decay under the influence of magnetoconvection, however due to the nature of the setup, only a partial decay takes place over the 100 h runtime of the simulation. With this and the fact that the DV network was only tried and tested on the same simulation used for training, there is still a gap in understanding the applicability of DV to new data at finer resolutions (<1 arcsec) and in environments, which present a realistic evolution/decay of intense magnetic flux in the solar atmosphere. Furthermore, whilst the success of velocities was presented in terms of correlations and error, it remains to be see how useful these flows are, i.e. whether or not they still can be used to identify coherent flow structures and their evolution.

In addition to the simulated sunspot presented by Rempel (2015), using the MURAM code, there exists a wealth of simulated solar data from myriad codes. Examples of these numerical simulations are the MURAM (Vögler et al. 2005), CO5BOLD (Freytag et al. 2012), STAGGER (Stein & Nordlund 1998), BIFROST (Gudiksen et al. 2011) and R2D2 (Hotta & Iijima 2020) codes. These codes present working models of magnetoconvection up to the photosphere and beyond. Some of these codes have been modified to simulate the evolution of a strong magnetic flux of varying initial configurations. The MURAM simulation by Rempel & Cheung (2014) introduces an untwisted semi-torus of magnetic flux, which is advected through a bottom boundary <16 Mm below the photosphere. The paper by Bjørgen et al. (2018) highlights simulations of an AR using the BIFROST code, which imposes an pre-formed bipolar AR into the uppermost layers of the convective region. These simulations, and others in the literature, have opted to model magnetoconvective processes using a shallow box that only covers up to 16 Mm below the surface (see, e.g. Rempel, Schüssler & Knölker 2009; Beeck et al. 2012; Rempel & Cheung 2014; Chen et al. 2023). These setups give only a short time for the magnetic field to evolve naturally under the influence of magnetoconvection leaving limited information on the effects of the interplay between plasma and strong magnetic fluxes.

A more recent model for flux emergence was presented by Hotta & Iijima (2020), in which the entire convection zone (200 Mm depth) is simulated. This allows an initial force-free magnetic flux tube, positioned at a depth of around 35 Mm below the surface, to evolve under a more realistic boundary condition than simulations with shallow boxes. The simulation places a twisted magnetic flux tube, where the force-free parameter $\alpha = 2.43 \times 10^{-7}$ m^{-1}, with > 10 kG strength at a depth of $\approx$35 Mm depth below the photosphere. In this simulation the α is determined as a ratio of the first root of the Bessel function J_0 and the characteristic length scale of the flux tube, chosen to be the radius. This simulation excludes the effects of physical resistivity, i.e. assumes the plasma to be physically ideal, however, the minimum scale in our case is the pixel size, 96km. Khomenko et al. (2014) highlights that the effects of physical diffusivity and resistivity are negligible at the scales we are interested in, only having a significant influence at scales around 100 m in both quiet and active Sun conditions at photospheric heights. This setup also better simulates the evolution of the magnetic field under the influence of a realistic bottom boundary, since the depth of the box is much larger than the length scale of the flux tube.

Synthetic observations calculated from realistic simulation data provide a sufficient testing ground for the success of flow recovery methods. In order to accurately recover the flows, one has to consider two major aspects: the magnitude of the speed at a given point and its direction. Estimated flows will always carry some error and therefore the question remains of whether or not this error inhibits the ability to study the flow dynamics of a system. In particular, whether we can identify coherent flow features which govern the transport of material via the flow field.

The transport of material by fluids can be pinpointed by distinguishing material surfaces within the plasma flow that act as transport barriers. These transport barriers prevent the passing of material, such as magnetic flux, through them. In the context of solar plasma, the magnetic field is typically strongly linked to the motions of the plasma in the photosphere where the magnetic Reynolds number is large (e.g. see Parker 1963). To this end, the transport of magnetic flux across the photosphere is dominated by advection through the plasma.

Material surfaces that are identified by tracking the flow are called Lagrangian coherent structures (LCSs) and they partition regions that exhibit similar dynamics, orchestrating the flow into discernible coherent patterns. The transport barriers are determined by advecting particles along a time series of the flow, which will compound errors produced when generating velocity fields. Thus, the LCS theory facilitates a novel and precise analysis of flow properties by identifying surfaces that locally maximize attraction, repulsion, and shearing (Haller & Yuan 2000; Haller 2015). In other words, the LCS provides a skeleton for the plasma flow dynamics, decomposing complex flow behaviour into dynamic building blocks. Analyzing the ability of a method to recover LCSs can give a significant insight into the effectiveness of a method to predict the precise flow topology and hence the evolving dynamics of photospheric flows. They also provide a key insight into the transport of material by the flow, i.e. the influence that plasma flows have on transporting magnetic flux.

Here, we examine the ability of DeepVel NN to recover coherent plasma flow structures from synthetic continuum intensity maps of the photosphere, corresponding to the Planck function at $\tau = 1$, generated by the MURAM ARs simulation (see Rempel & Cheung 2014). The network was trained using continuum intensity and horizontal velocity fields from the R2D2 magnetoconvection simulation (see Hotta & Iijima 2020). Both the MURAM and R2D2 simulations present the evolution of magnetic flux under the influence of magnetoconvection in the upper layers of the solar interior up to the photosphere and beyond. Both present similar models with the distinguishing features being the initial topology of the magnetic field, the thermal evolution of the flow, with only MURAM accounting for resistivity in the plasma in the form of a constant turbulent diffusion, and R2D2 presenting a more realistic boundary condition. Thus together, both simulations present a large range of features present throughout the evolution and decay of an AR, presenting a good testing ground for how well DV can emulate the flow topologies present in and around ARs even when some features are missing from the training simulation. Since we are only validating the applicability of DV to new scenarios, we are not presently concerned with applying to observations since there is no ground truth data for velocity on the Sun. Thus we provide evidence of DVs adaptability to ARs with different setups and flow topologies when trained only on one simulation.

The paper is structured as follows: the numerical data, velocity recovery methods, and analysis techniques are presented in Section 2. Section 3 describes the results including the success of the recovered flows, methods as well as their ability to accurately preserve Lagrangian features in the reconstructed flows. Finally, a detailed discussion of the key results and conclusions of our analysis is presented in Section 4.

2 METHODS

2.1 Velocity field recovery with DeepVel

DV is a convolutional NN (CNN), which can be used to retrieve 3D velocity fields from solar spectra, magnetograms or radiative intensity maps (Asensio Ramos et al. 2017). CNNs apply convolutions to inputs to reduce the number of parameters, e.g. convolutions may be used to identify key features in images such as the granules, intergranular lanes and the edges of pores; this makes CNNs better at generalizing spatially. We trained DV to recover the flows from simulated ARs. The numerical data used for the training and analysis are the 2D horizontal velocities, which we define as $\boldsymbol{v}_{\mathrm{h}} = (v_x, v_y, 0)$, and continuum intensity of the $\tau = 1$ surface (i.e. the photospheric layer). The horizontal velocities used for training and testing our model were extracted from the formation heights of the continuum intensities from the simulations. The heights of the surfaces vary by $\approx$100 km, over distances of $\approx$1 Mm in regions of QS granulation, and $\approx$500 km, over distances of $\approx$1 Mm over the boundary of the pore. The continuum intensity data was chosen as they contain the most significant insight into motions at all scales without relying on well-defined magnetic structures, i.e. subverting the problems presented by Rempel et al. (2022) when using magnetogram data.

2.1.1 *Training set*

The numerical data used for training DV in the present study is from the R2D2 simulation presented by Hotta & Iijima (2020). This run of the R2D2 simulation models the evolution of a twisted force-free flux tube placed, at a depth of 35 Mm below the photosphere, under the influence of magnetoconvection. The simulation box has dimensions 98×98 Mm^2 with a periodic boundary in the horizontal direction and uniform grid spacing of 96 km pix^{-1}. In the vertical direction, the simulation covers a depth of 200 Mm spaced over a non-uniform grid, and height up to 700 km in the lower photosphere. The depth of this simulation allows the evolution of the flux tube to remain minimally affected by the bottom boundary. The R2D2 code uses a set of radiative MHD equations which, in this setup, assumes an ideal plasma (no physical resistivity) and no background magnetic field, so the flux tube evolves solely under the effects of magnetoconvection. The temporal resolution of the data is 120 s over 180 h of real-time simulation, this provides 37 h of time before the emergence of the magnetic flux and over 100 h of the decay process from the peak of the magnetic flux. Flow processes, therefore have a long time to evolve naturally with the rising magnetic flux prior to the emergence of the AR. Hence, the R2D2 simulation provides a realistic model of the interaction of convective plasmas and the magnetic field. An example intensity map and magnetogram, obtained from the R2D2 simulation at a time of 60 h after the start of the simulation (i.e. the time of peak magnetic flux in the photosphere), are shown in Fig. 1.

For the training process, 3000 pairs of 4.8×4.8 Mm^2 (50×50 pix^2) continuum intensity snapshots at the simulations cadence of 120s, along with their corresponding horizontal velocity fields. These snapshots were chosen at purely random times and positions across the entirety of the $\tau = 1$ surface of the simulation, and represent $<$1 per cent of the total data available. We suggest that this 120s cadence is suitable for testing the identification and analysis of flows at the (sub)granular scale since granules have a lifetime of $\approx$10 min (e.g. see Bahng & Schwarzschild 1961). Whilst this is a relatively low cadence in comparison to some of the state-of-the-art telescope data available, it is a highly composite number, so most telescope imagery can be used at this lower cadence by selecting frames at this time apart. These pairs were filtered to ensure that there were no overlaps in the training data to avoid overfitting, i.e. to help ensure that DV is able to classify velocities well for new data just as well as the training set.

The training samples were split into sets of 2000, 500, and 500 labelled as training, validation, and testing. The training and

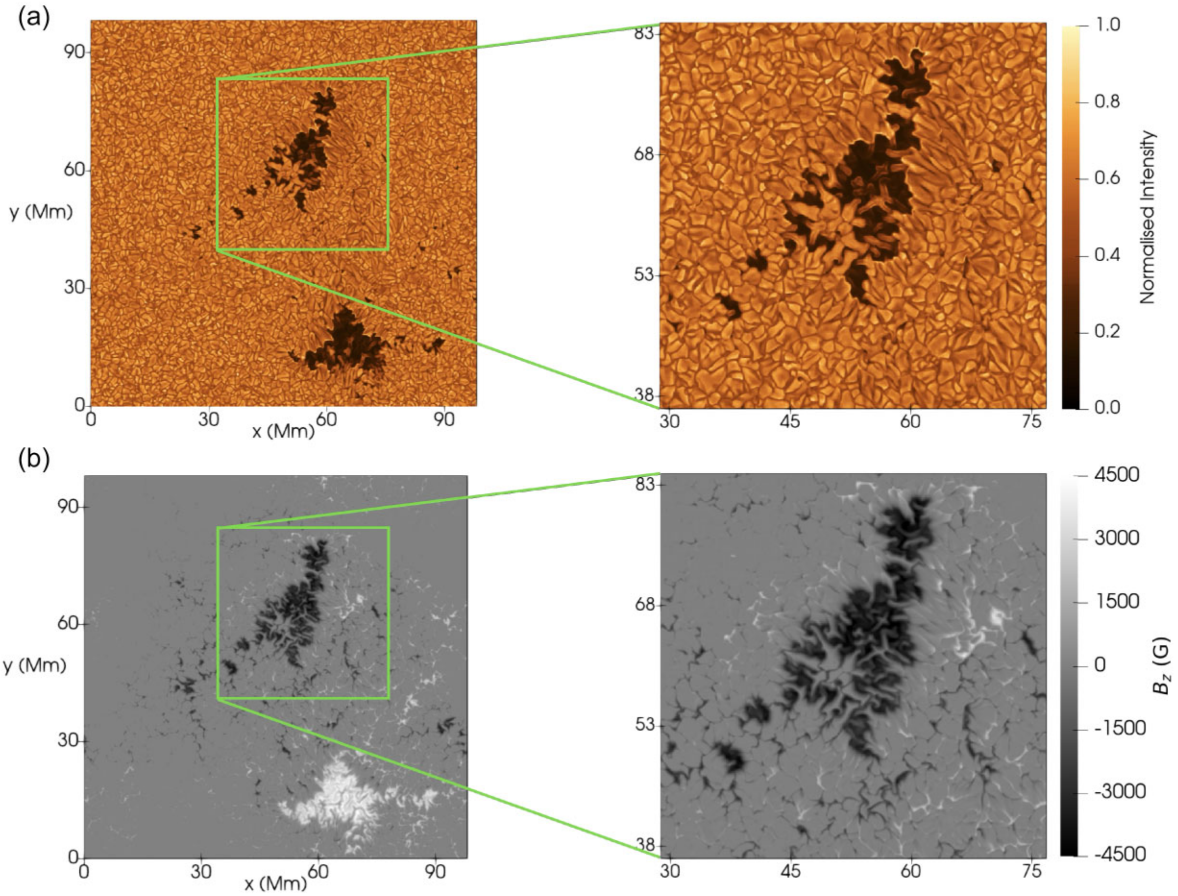

Figure 1. An example of the fully developed AR at the time of peak magnetic flux (i.e. 60 h after the start of simulation), from the R2D2 magnetoconvection simulation. On the left-hand side, two isometric panels show the white light intensity (top panel) and the magnetogram (bottom panel). The right-hand side panels show a close-up view of the intensity (top panel) and magnetogram (bottom panel) of the region bounded by the square. The intensity was normalized by the maximum intensity of the entire region.

validation set are used throughout training to improve the parameters of the network. The network parameters are only updated if the predictive performance improves on the validation set, which is checked at the end of each training epoch. Given the random nature of the sample selection, the snapshots should be representative of the flows present in all scenarios in the simulation.

It should be noted that, as with any supervised learning method, this process poses the risk of introducing overfitting. If the number of samples from the training simulation is too large, the NN may learn only the images and flows from the training data and therefore be unable to measure the velocities accurately from new images. Because of this, we kept the proportion of training samples small to reduce this risk and reduce the computational cost of training the network. The additional 500 testing samples were used for independently testing DeepVel to ensure it is able to generalize to new data.

The network was trained for 100 epochs (i.e. the network weights were adjusted over the entire set of training samples 100 times) using the Adam optimizer and the mean-squared error (MSE) loss defined as

$$\mathrm{MSE} = \frac{\sum_{i=0}^{n} |\boldsymbol{v}_{\mathrm{R2D2},i} - \boldsymbol{v}_{\mathrm{DV},i}|^2}{n}, \tag{2}$$

where i represents the observation (pixel) where the velocity is produced and n is the total number of observations. The training took approximately 100 min using NVIDIA V100 GPU and 10 h on an Apple M2 card with 10 GPU cores.

The loss from the network throughout the training process is shown in Fig. 2. This figure highlights that the network converges quickly in the first few epochs of training, reaching close to its minimum value after around just 20 epochs. The network was allowed to continue training in which the learning stabilised at around 50 epochs before the validation loss started to increase despite the training loss beginning to decrease more rapidly around 60 epochs, suggesting that the network is now overfitting to the training samples, however the network reached its optimum performance at the 67th epoch. In order to overfitting, DV does not update the network parameters unless the validation loss decreases. Thus the tradeoff between number of epochs and the performance of the network drop off after just a few iterations.

After the completion of the training process, individual velocity fields for full 1 Mp images took under 1 s to recover on both devices. When compared to other methods, if we ignore the time taken for training, this is a 40× speedup compared with FLCT; the benefit of the speedup outweighs the time taken for training when considering large time series.

2.1.2 Test data

For testing whether DV has generalized well to predicting velocities from continuum intensity images, we have opted to use the

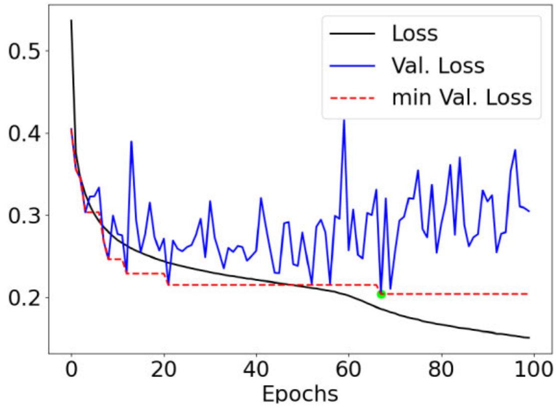

Figure 2. Loss during the first 100 epochs of training from the DV during training. The smooth solid curve represents the loss value from the training data at each epoch and jagged curve represents the loss value for the validation set (calculated at the end of each epoch). The dashed curve shows the minimum validation loss for the network at each epoch and the green point shows the minimum validation loss over the whole training process.

MURAM AR simulation presented in Rempel & Cheung (2014). This simulation presents the realistic emergence of an untwisted semi-torus of magnetic flux in a box that has a periodic horizontal boundary with dimensions $147.5 \times 73.7\,\mathrm{Mm}^2$. The depth of this box is 16.4 Mm. The simulation shares the same $96\,\mathrm{km\,pix}^{-1}$ horizontal spatial resolution as the R2D2 simulation. Along with the shallow box, relative to R2D2, this simulation includes the effects of turbulent diffusion, providing a more realistic thermal evolution of the plasma. Pairs of the continuum intensity images at the $\tau = 1$ surface, and velocity fields from the corresponding heights of these surfaces, with cadence 120 s we selected from different times in order to study the ability of DV and FLCT to recover velocities. A crucial preprocessing step for new inputs to DV is to ensure the continuum intensities are matched to the histograms of the continuum intensity images from R2D2, since DV is a supervised learning method, it is important to ensure that input images have properties similar to the training set. DV was then applied to recover velocities from the MURAM intensitygrams. An example frame of continuum intensity and magnetogram from the simulation are presented in Fig. 3.

The radiative transfer schemes between R2D2 and MURAM differ, for full details of the schemes, refer to the literature for R2D2 (Hotta & Iijima 2020) and MURAM (Vögler et al. 2005). These differences in numerical schemes may lead to differences in the intensities produced in the synthetic images. Specifically, the contrast in the intensity is greater in R2D2 data, however the normalized intensities reveal a similar distribution to that of MURAM. Testing DV on different simulations with variations in their numerical schemes, setups and included physics will provide insight into the limitations of DV when only trained on one simulation.

The R2D2 and MURAM simulations offer valuable insight into different crucial aspects of the interactions of plasma flow and magnetic flux throughout AR evolution. Thus these provide an opportunity to properly test the ability of DV to accurately reproduce flow features, from sub- to mesogranular scales, in the presence of intense magnetic flux.

2.2 Analysis of flow properties

Once DV was trained and velocities were produced using our model, we analysed the velocity fields by studying coherent flow structures by identifying transport barriers in the flow. These transport barriers act as a skeleton for the flow, dictating the general flow topology and thus the movement of material within the flow. These structures are Lagrangian in their nature as they are derived from tracing the motions of particles with the flow. Thus, these LCSs will allow us to determine how much of the flow topology has been identified correctly by DV and hence help us to understand how error is propagated throughout a time series of recovered velocity fields when taking derivatives of the outputs. Therefore, analysing what has been recovered by DV in this sense will allow us to examine how well DV is able to emulate photospheric flow physics from the R2D2 simulation and determine how well the model generalizes to more realistic and complex flows present in the MURAM simulation.

One way to identify repelling/attracting material surfaces is by means of ridges of the finite-time Lyapunov exponent (FTLE), calculated forward and backward in time (see, e.g. Haller 2015). The FTLE field can be calculated in any number of dimensions, yet it has been shown that two-dimensional surfaces (from 2D velocity fields) are enough to describe the critical influences on the distribution of the magnetic field in the solar photosphere (see e.g. Yeates, Hornig & Welsch 2012; Chian et al. 2014, 2019). Changes in the organization of the FTLE field reflect a restructuring of the flow. In the photosphere, such changes can be due to a strong emerging magnetic field that may form ARs. In this respect, based on FTLE fields of the horizontal velocity of a simulated solar surface, Silva et al. (2023) detected an objective signal of intense emerging magnetic fluxes, allowing for predictions of the short-term (of the order of a few hours) evolution of magnetic flux emergence prior to the detection of notable structures in magnetograms and intensity images of the solar photosphere. Therefore we can use the structures provided by FTLE analysis to give insight to how the magnetic field is evolving as a consequence of plasma flow. As such, comparing recovered flow FTLEs against the true flow FTLEs will give a strong indication of how physically meaningful the recovered topologies are and how well the physics of the true flows are being emulated by the recovery method.

In order to calculate the FTLEs, a uniform grid of particles is advected over a time-dependent velocity field. Here, we present the mathematical description of the FTLE field. Consider $\boldsymbol{v}$ to be a velocity field in a spatial domain and $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ to be a fluid particle at the initial position that follows a trajectory in space given by solutions of the initial value problem. Then,

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{v}(\boldsymbol{x}, t). \tag{3}$$

The displacement of particles at time $t = \tau$ from its initial position at time $t = t_0$ can be found as

$$\boldsymbol{x}_\tau - \boldsymbol{x}_0 = \int_{t_0}^{\tau} \boldsymbol{v}(\boldsymbol{x}, t)\mathrm{d}t. \tag{4}$$

Then, the flow map can be defined as

$$\boldsymbol{x}(t_0 + \tau) = \phi_{t_0}^{t_0+\tau}(\boldsymbol{x}_0), \tag{5}$$

where the operator $\phi_{t_0}^{t_0+\tau}(\boldsymbol{x}_0)$ maps the initial position of a particle at $\boldsymbol{x}(t_0)$ to a final position $\boldsymbol{x}(t + \tau)$. In the 2D case, particles are given initial positions $\boldsymbol{x}_{i,j}(t)$, where i and j are indexes of the particle's initial grid point. Distances between a particle and its nearest initial neighbours, on a 5-point grid, are tracked as they are advected by the flow.

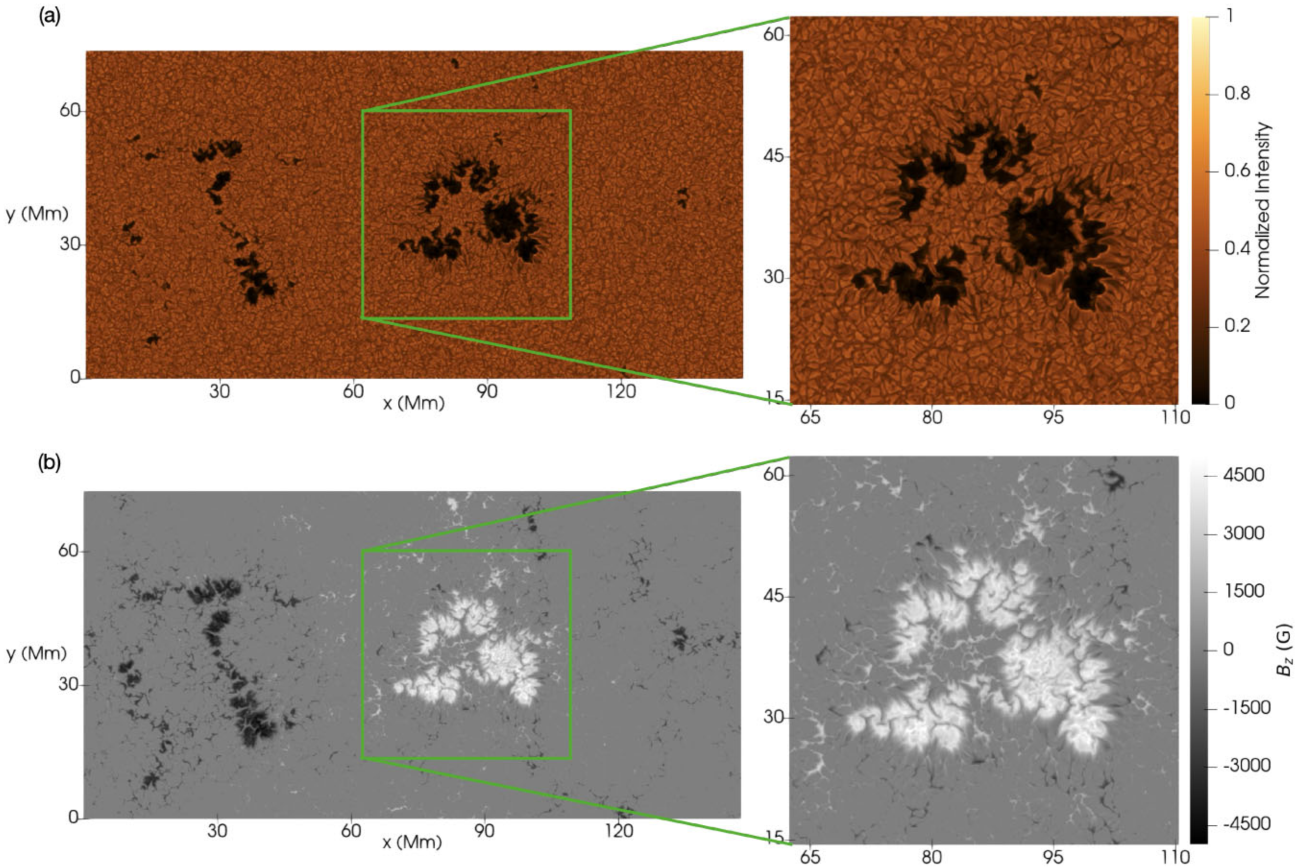

Figure 3. An example slice of the AR at 83 h after the start of the simulation, from the MURAM simulation. The panels in the left-hand column show the continuum intensity (a) and the magnetogram (b). The right-hand side panels show a close-up view of the intensity (top panel) and magnetogram (bottom panel) of the region bounded by the square. The intensity was normalized between to be on a scale between 0 and 1.

The deformation gradient after the advection is given by the Jacobian matrix of partial derivatives of the flow map

$$D\phi_{t_0}^{t_0+\tau}(\boldsymbol{x}_{i,j}) = \begin{pmatrix} \frac{x_{i+1,j}(t_0+\tau)-x_{i-1,j}(t_0+\tau)}{x_{i+1,j}(t_0)-x_{i-1,j}(t_0)} & \frac{x_{i,j+1}(t_0+\tau)-x_{i,j-1}(t_0+\tau)}{y_{i,j+1}(t_0)-y_{i,j-1}(t_0)} \\ \frac{y_{i+1,j}(t_0+\tau)-y_{i-1,j}(t_0+\tau)}{x_{i+1,j}(t_0)-x_{i-1,j}(t_0)} & \frac{y_{i,j+1}(t_0+\tau)-y_{i,j-1}(t_0+\tau)}{y_{i,j+1}(t_0)-y_{i,j-1}(t_0)} \end{pmatrix}. \qquad (6)$$

This matrix is used to compute the Cauchy–Green deformation tensor defined as

$$\Delta = \left[D\phi_{t_0}^{t_0+\tau}(\boldsymbol{x}_{i,j})\right]^{\mathrm{T}} D\phi_{t_0}^{t_0+\tau}(\boldsymbol{x}_{i,j}),$$

where the superscript T denotes the matrix transpose. Finally, the FTLE field is calculated by means of

$$\mathrm{FTLE}_i^{t_0+\tau}(\boldsymbol{x}) = \frac{1}{|\tau|} \ln \sqrt{\max(\lambda_i)}, \quad i = 1, 2 \qquad (7)$$

with λ_i being the eigenvalues of Δ.

Given a time interval of length τ, particles are integrated forward in time over the interval $[0, \tau]$ and integrated backward in time over the interval $[-\tau, 0]$, to produce the forward-FTLE and backward-FTLE fields. Ridges formed by the largest FTLEs describe the most strongly repelling structures in the plasma flow in forward time, the most strongly attracting structures in backward time.

3 RESULTS

Here, we present the key results of our analysis for the success of DV in recovering velocities from the MURAM simulation when trained on flows from R2D2. In particular we present a number of metrics which show the success of DV in recovering flows at subgranular scales ($\lesssim$100 km) up to mesogranular scales ($\approx$5 Mm) over ARs and QS regions to determine how well DV can capture the influence of >kG magnetic flux on plasma flows during the evolution of an AR.

3.1 Metrics

In this section, we discuss the effectiveness of DV at recovering AR flows both qualitatively and back these quantitatively using a number of statistical metrics. The following list of metrics have been used:

(i) Pearson Correlation Coefficient (PC)

$$\mathrm{PC} = \frac{\sum(v_{\mathrm{MURAM}} - \bar{v}_{\mathrm{MURAM}})(v_{\mathrm{DV}} - \bar{v}_{\mathrm{DV}})}{\sqrt{\sum(v_{\mathrm{MURAM}} - \bar{v}_{\mathrm{MURAM}})^2 \sum(v_{\mathrm{DV}} - \bar{v}_{\mathrm{DV}})^2}}, \qquad (8)$$

where the bar over a variable represents the mean value. This correlation coefficient provides a measurement of the strength of the linear relationship between two variables between -1 and 1, where -1 describes a perfect negative correlation and positive 1 describes a perfect positive correlation. For perfectly recovered velocities we expect $\mathrm{PC} = 1$, a strong correlation is typically identified as $0.5 < |\mathrm{PC}| < 1$.

(ii) Kolmogorov–Smirnov test (KS test)

$$\mathrm{KS} = \sup_x |\mathrm{CDF}_{\text{MURAM}}(x) - \mathrm{CDF}_{\mathrm{DV}}(x)|, \tag{9}$$

i.e. the supremum of the distance between the empirical cumulative distribution functions (CDFs) for the velocities from MURAM and DV, where sup defines the supremum function over the points x of the CDF. The test-statistic is interpreted as the maximum per cent difference in vertical between values in the distribution (Hodges 1958).

(iii) Root-mean-squared Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=0}^{n}(v_{\text{MURAM},i} - v_{\text{DV},i})^2}{n}}, \tag{10}$$

where n represents the total number pixels in the image. The RMSE provides an average of distance of the predicted values from the true values

(iv) Median Relative Error (MRE)

$$\mathrm{MRE} = \mathrm{Median}\left(\frac{|v_{\text{MURAM}} - v_{\text{DV}}|}{|v_{\text{MURAM}}|}\right). \tag{11}$$

The relative error shows the difference between the predicted and true values as a proportion of the true value, i.e. per cent error. Since the velocities can be typically large, in any location where the velocity may be, in actuality, close to 0 and is predicted to be large will cause the relative error to explode. For this reason, the median is chosen as opposed to the mean as this will ignore the extreme values.

(v) Normalized dot product

$$\text{Normalized dot product} = \frac{\boldsymbol{v}_{\text{MURAM}} \cdot \boldsymbol{v}_{\text{DV}}}{|\boldsymbol{v}_{\text{MURAM}}||\boldsymbol{v}_{\text{DV}}|}. \tag{12}$$

The normalized dot product is the cosine of the angle between the predicted and true velocity vectors. This shows the alignment of the recovered vectors as a value between -1 and 1 where 1 represents a scenario where all the predicted vectors are parallel to the true vectors, -1 are antiparallel and 0 represents a case where all the vectors are at right angles to the true vectors.

3.2 Flow recovery with DeepVel

Two-dimensional velocity fields were recovered from the MURAM simulation by applying DV to time-consecutive pairs of continuum intensity, separated by a cadence of 2 min. The x- and y-components of the velocity fields were estimated for $\tau = 1$ at various times throughout the evolution of the photosphere. At $t \approx 2$ h the magnetic flux in the photosphere has a mean value of <1 G and a maximum value of 20 G. Therefore, this highlights an example of when the simulation presents mostly QS dynamics in the photosphere. An example of a time when intense magnetic fluxes are present in an AR is presented at $t \approx 83$ h. These particular times were chosen to highlight DV's capability since they provide examples of the full range of flow dynamics in the presence of QS levels of magnetic flux and AR levels of magnetic flux. Examples of the original and recovered velocity fields are shown in Figs 4 and 5. The recovered velocity fields were compared to the original simulated velocities.

Fig. 4(a) shows the divergence of the horizontal velocity field, $\nabla \cdot \boldsymbol{v}_{\mathrm{h}}$, from the MURAM simulation at $t = 2$ h, when the magnetic flux rope is still being convected below the surface. Panel (b) presents the divergence of the horizontal velocity field recovered by DV over the same region as panel (a). Panels (c) and (d) show the line integral convolution (LIC; created by convolving a white noise filter with the velocity field and integrating over the field lines) over a zoomed-in region, distinguished by the green square. LIC was first presented by (Cabral & Leedom 1993,) and it emphasizes the streamlines of the horizontal velocity field using dark lines. The zoomed-in region highlights a region where complex magnetic structures later appear. The red and blue colouration represents regions of positive and negative horizontal divergence, which we will refer to as the divergence. From panel (b), DV is able to reproduce the location of divergent features in the flow apparently well, with the magnitude of the flow not being as high as those present in the simulation, e.g. refer to the scatter plots shown in Fig. 6. This discrepancy in speed is likely due to differences between speeds in the training set and test data. DV is able to recover topological flow features accurately from MURAM at length scales of intergranular lanes (<500 km).

Fig. 5 presents the same region as was shown in Fig. 4, much later in the simulation ($t = 83$ h) after the magnetic flux tube has emerged. Panels (a) and (c) once again show the divergence and flows from simulation, which now highlight a region at the edge of a pore where many intricate flow structures are present. In this case, we see that DV is able to distinguish flow behaviour between that over the pore and that on the edge in panels (b) and (d). The distinguishing difference here is that DV is not able to identify the exact topology of small-scale vortices (<100 km) and saddle flows within the pore. However, DV is able to to identify some of the apparent swirling motions, despite velocities having a magnitude of $\approx 1/10$ that of the flows surrounding the AR.

Figs 6 and 7 give further insight into the results discussed above; they present the distributions of different velocity field components from simulation and DV and the scattering of these quantities, respectively as a means of analysing the correlation of the recovered velocity fields.

The histograms in Fig. 7 reveal that in both active and quiet regions DV, trained on R2D2, is able to match the horizontal velocity distributions from MURAM well. It appears that DV typically under-estimates the velocities present in QS regions, but the distributions still match the shapes and general properties of the flows well. This is backed by a KS-test which highlights that the components of the horizontal velocity vector are chosen from a distribution which differs by a maximum of ≈ 10 per cent from the true distribution (see Table 1). This deviation is due to the tendency of DV to underestimate the velocities shown in panels (a) and (b), respectively. The distribution of speeds shown in panel (c) and better highlights the underestimation in velocity components, combining to show an overall reduced speed, implying that the velocity vector is recovered with a lesser magnitude and hence errors are not simply caused by a rotation, but a shift in the component distributions. The speeds are still within the same order of magnitude despite this, highlighted by Table 1. The divergence distribution of the DV velocity field also shares similar properties to the simulated velocity distribution, capturing the bimodal nature of the distribution, which is result of the mostly non-diverging nature of the flow within the AR and the positively divergent granules that dominate the QS-like regions. The key properties of the divergence match well despite the differences in the variance and errors introduced by taking derivatives of data.

The scatter plots in Fig. 6 highlight the strong correlation of the velocities over all regions, and that flow speeds are typically underestimated. The divergence is more weakly correlated over the AR, and carries more error than individual components of velocity (as to be expected from taking derivatives) however the QS and overall divergence of the recovered velocity fields are highly correlated. In fact, we see that DV performs particularly well at reproducing the distributions of AR flows, only struggling to capture the exact values of divergence in the AR, shown in panel (d) and Fig. 7.

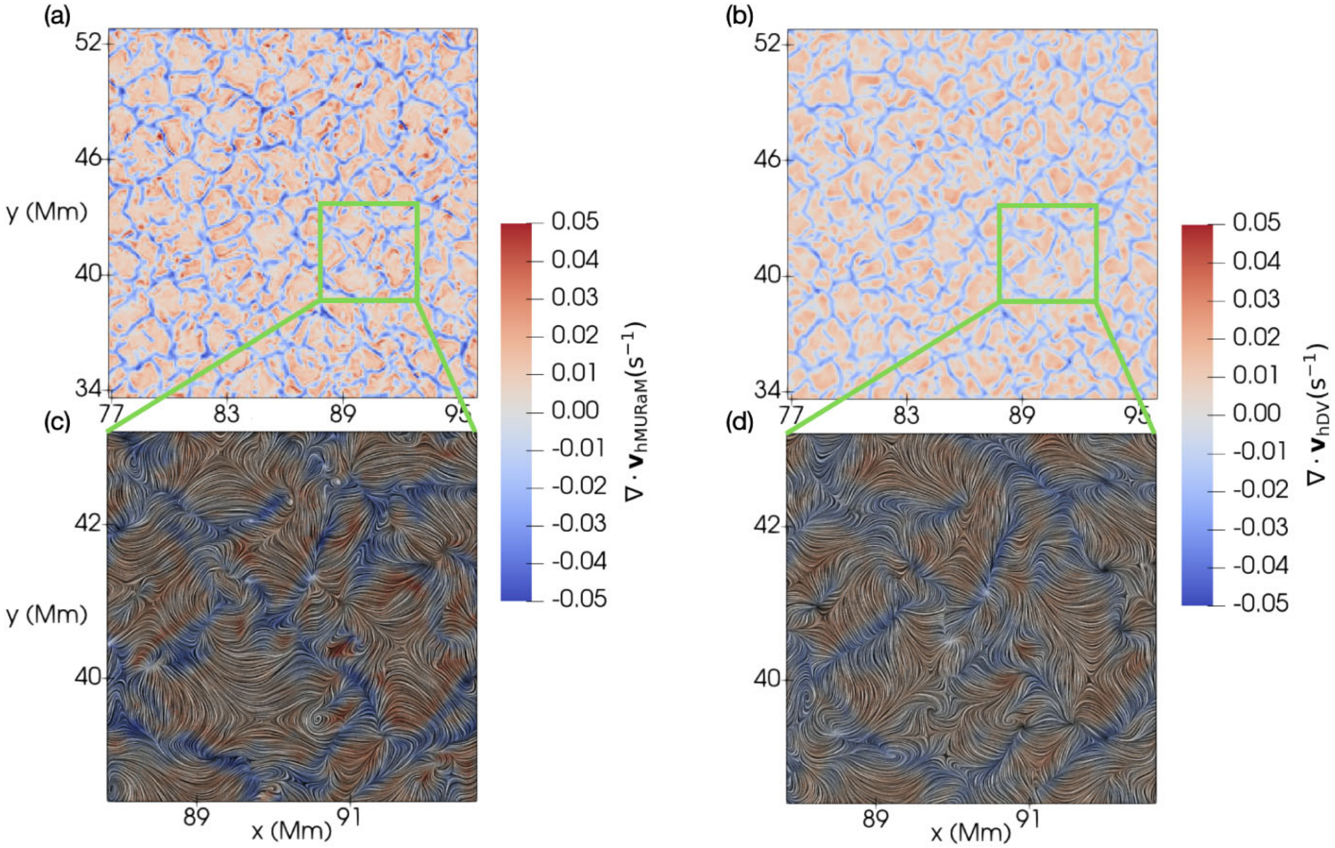

Figure 4. A close view of the divergence of the horizontal components of velocity, 2 h into the MURAM simulation. This figure highlights a region where >kG magnetic flux emerges later in the simulation. Panel (a) shows the divergence field from simulation over a zoomed-in section of the region highlighted in Fig. 3. Panel (c) shows a closeup of the LIC with the velocity field superimposed on the divergence field of the highlighted region from panel (a). Panels (b) and (d) show the same regions as panels (a) and (c), instead using velocities recovered by DV.

Table 1 shows that DV is able to replicate the properties of the velocity fields well about the means of the distributions, seen by the means and interquartile range (IQR) matching well for all presented values. Only the divergence and curl, where the error is expected to be increased from taking derivatives of estimated data, present issues. Despite this, the divergences of recovered flows correlate strongly in QS regions. Flows still correlate well, but not as strongly, within ARs where the contrast in images becomes much weaker and the dynamics differ largely due to the suppression of plasma motions by the magnetic field. In comparison with the distributions of properties over AR, the Table 2 suggests some contradicting results, however the table presents values that are spatially dependent whereas the distributions in Table 1 are not; i.e. recovered flows in ARs are not recovered spatially as well as flows in QS regions, but the properties of the flows in these regions align better.

In order to quantify the success of the recreated flows, Table 2 displays metrics describing the accuracy of flows from DV, including their alignment and associated errors when compared against the simulation flows. The PC describes the strength of the linear relationship between the recovered velocities and the simulated velocities. The square of this coefficient describes the explained variance captured by the method. The recovered flows, as mentioned above, are highly correlated with the simulated flows. This high correlation shows that a significant amount of information in the flow is recovered. The correlation of the divergence is only slightly impacted over QS regions despite taking derivatives of velocities, which will invariably include some error. The correlation of the divergence within the pore becomes more moderate compared to the surrounding flows but still shows a statistically significant correlation to the original flows. We also observe the same trend in the alignment of the flows, measured by the normalized dot product, i.e. the cosine of the angle between the recovered flows and the simulation flows. A normalized dot product close to 1 would imply the flow is well aligned, and a value close to zero represents a flow with no alignment with the target. The DV recovered flows have a mean alignment of around 0.674 over quiet regions, and a lesser alignment over the AR of around 0.501. The inverse cosine of these values corresponds to mean differences in angles of $\approx 47^\circ$ over the QS-like regions and $\approx 60^\circ$ over ARs. However the median normalized dot product sits at 0.871 and 0.774 for both regions, thus half of the velocity vectors are within 30° and 40° of their true direction. Furthermore, the RMSE and the MRE are shown. The RMSE provides an average distance of the NN velocities from the true velocities, however this can be easily skewed where large velocities are involved. The MRE is more representative of the performance of the network, as it is less skewed by outliers in the recovered flows. Since the velocities and divergences from DV correlate well with the simulation and the flows are aligned well over the majority of the simulated surface, the apparently high relative error should be taken in context. Velocities produced by DV have less variance than the velocities in MURAM, but correlate strongly. Therefore the topology and features of the flow field may remain consistent with the simulation but show a different magnitude which produces a consistently large relative error with the speed of the flows.

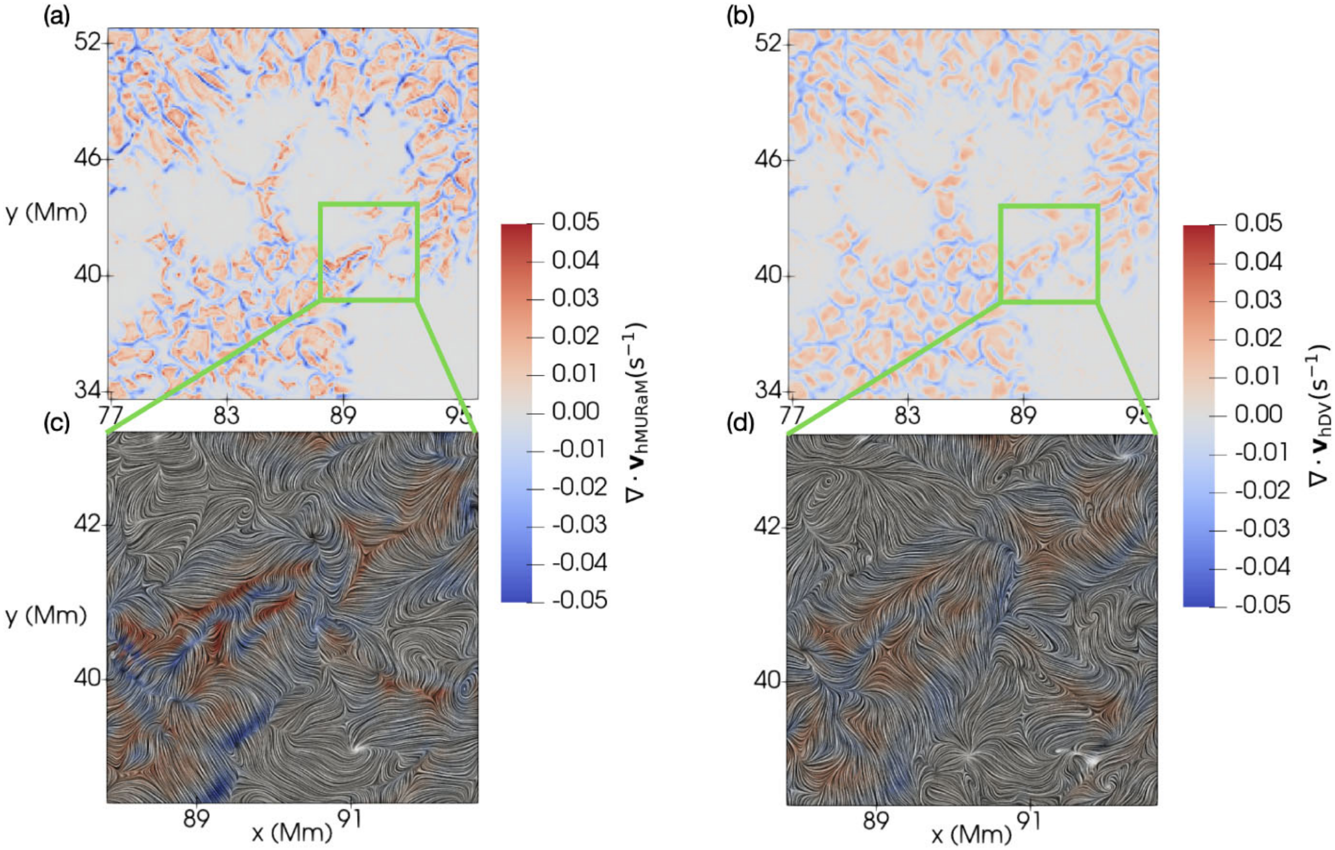

Figure 5. The same visualization as Fig. 4 is shown, at a time after the AR has emerged (83 h). The zoomed-in regions highlight the boundary of a pore, containing complex flow structures such as saddles and vortices.

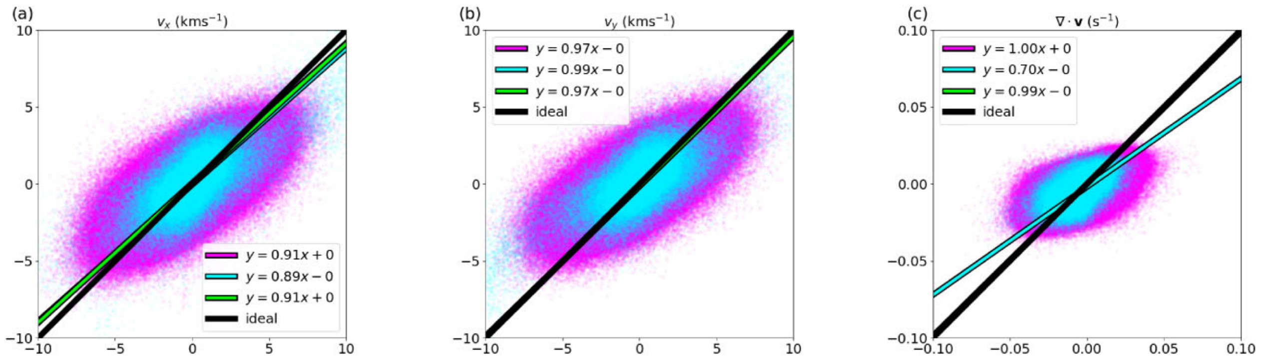

Figure 6. Scatter plots of components of velocities (a) and (b) and their divergence (c) recovered from by DV and their corresponding lines of best fit, plotted against the simulated values from MURAM, shown in cgs units. Scatter values over ARs are superposed on top and values over the QS are shown underneath. An additional line of best fit, described by the third equation in the legend, shows the overall line of best fit over all types of regions and the dark line along the diagonal shows the target line of best fit ($y = x$). Points have had their opacity lowered to remove outliers and highlight where the strongest spread of velocities lie.

This difference in magnitude is highlighted by Fig. 7(c), where there is an underestimation in the speed of flows, particularly over the QS regions. These values are consistent with the mean relative error shown by Tremblay et al. (2018), which still prove stronger than other flow recovery methods.

3.3 Finite-time lyapunov exponents

A further step to test the results from DV is to assess the capability of the recovered velocities to reproduce the natural transport barriers created by the flow interaction. To identify the flow's barriers, the FTLE field was computed by integrating over short (20 min) times and long (100 min) times, recovering granular and mesogranular scales, respectively. The 20-min FTLE fields are shown in Fig. 8 and the 100-min fields are presented in Fig. 9. Ridges in the forward-FTLE and backward-FTLE fields represent locally repelling and attracting regions in the flow, in areas where these ridges overlap, there may be complex flow behaviours present such as saddles and vortices. Regions with small FTLE values (i.e. that contain no ridges) highlight regions with simple flows.

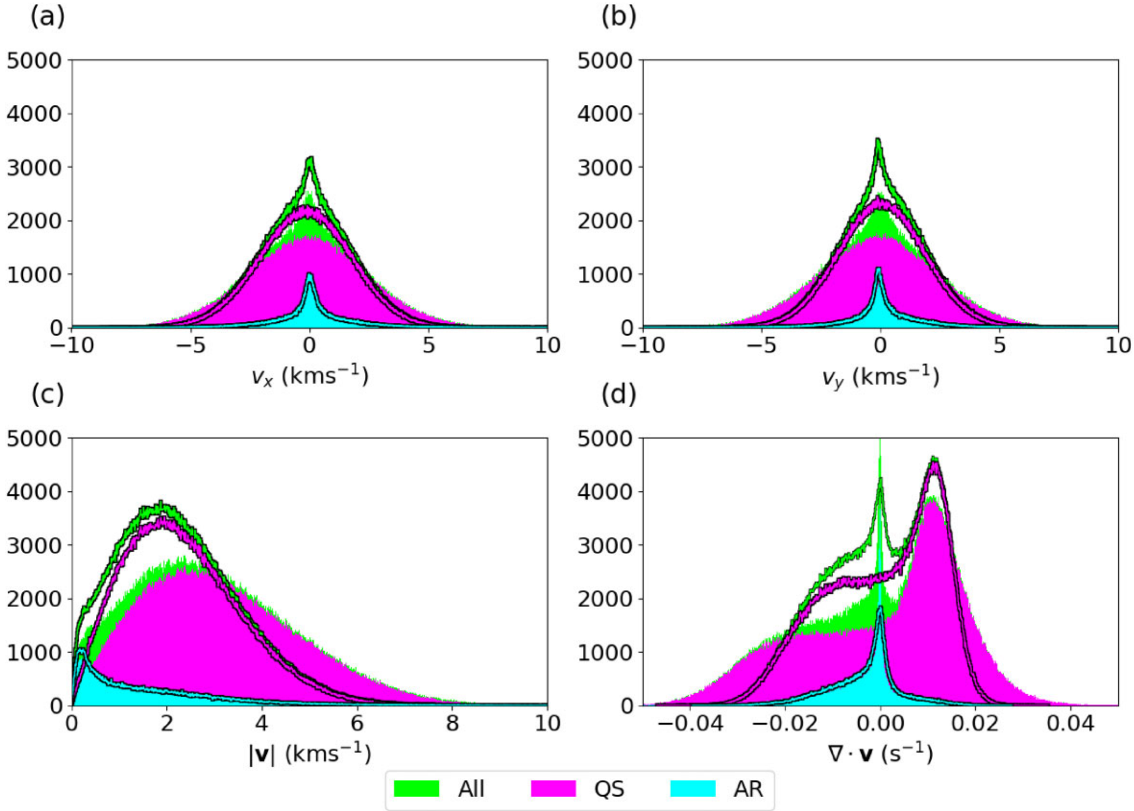

Figure 7. Histograms showing the distribution of different components of the velocity field at a time of 83 h, for the original simulation (solid coloured backgrounds) and the recovered flows (the outlined curves) over regions of >kG magnetic flux (top-most distribution), <kG flux (middle-layer distribution), and over the entire field of view (bottom-most distribution). Panels (a) and (b) show the distributions for the x and y components of velocity. Panel (c) shows the distribution of speed and (d) shows the distribution of divergence. All are shown in CGS units.

Table 1. Properties of the horizontal velocity field ($\boldsymbol{v}_\mathrm{h}$) distributions from the MURAM simulation and those recovered from DV, shown in cgs units.

		Minimum	Maximum	Mean	IQR
v_x (kms^{-1})	MURAM	−11.1	12.4	-62.1×10^{-3}	3.11
	DV	−10.2	12.5	-175×10^{-3}	2.83
v_y (kms^{-1})	MURAM	−13.5	12.5	-88.3×10^{-3}	3.13
	DV	−12.2	11.0	119×10^{-3}	2.65
$\nabla \cdot \boldsymbol{v}_\mathrm{h}$ (s^{-1})	MURAM	−0.0815	−0.0728	5.52×10^{-7}	0.0238
	DV	−0.0521	0.0355	8.22×10^{-6}	0.0197
$(\nabla \times \boldsymbol{v}_\mathrm{h})_z$ (s^{-1})	MURAM	−0.0958	0.114	-5.61×10^{-7}	8.63×10^{-3}
	DV	−0.0614	0.0662	-3.61×10^{-5}	6.55×10^{-3}

Panels (b) and (c) of Fig. 8 depict a notable reduction in the strength of the ridges and hence a reduction in the complexity of the flow in the presence of high levels of magnetic flux, in contrast to the quiet regions. This behaviour is typical of the magnetic field

Table 2. Metrics comparing the DV-recovered velocity components to the original velocities from the MURAM simulation. Horizontal velocities were taken over the entire $\tau = 1$ surface at a time where an AR was present in the photosphere ($t \approx 83$ h). Compared values are split over QS conditions (<kG) and AR conditions (>kG).

	PC		KS test statistic		RMSE (kms^{-1})		Median relative err.		Normalized dot product	
	QS	AR	QS	AR	QS	AR	QS	AR	QS	AR
$\boldsymbol{v}$	–	–	–	–	–	–	–	–	0.674	0.501
v_x	0.750	0.712	0.0722	0.0538	1.69	1.47	0.710	0.886	–	–
v_y	0.752	0.717	0.0963	0.0391	1.68	1.49	0.745	0.950	–	–
$\nabla \cdot \boldsymbol{v}_\mathrm{h}$	0.723	0.549	0.112	0.0516	0.0116	8.43×10^{-3}	0.642	0.903	–	–

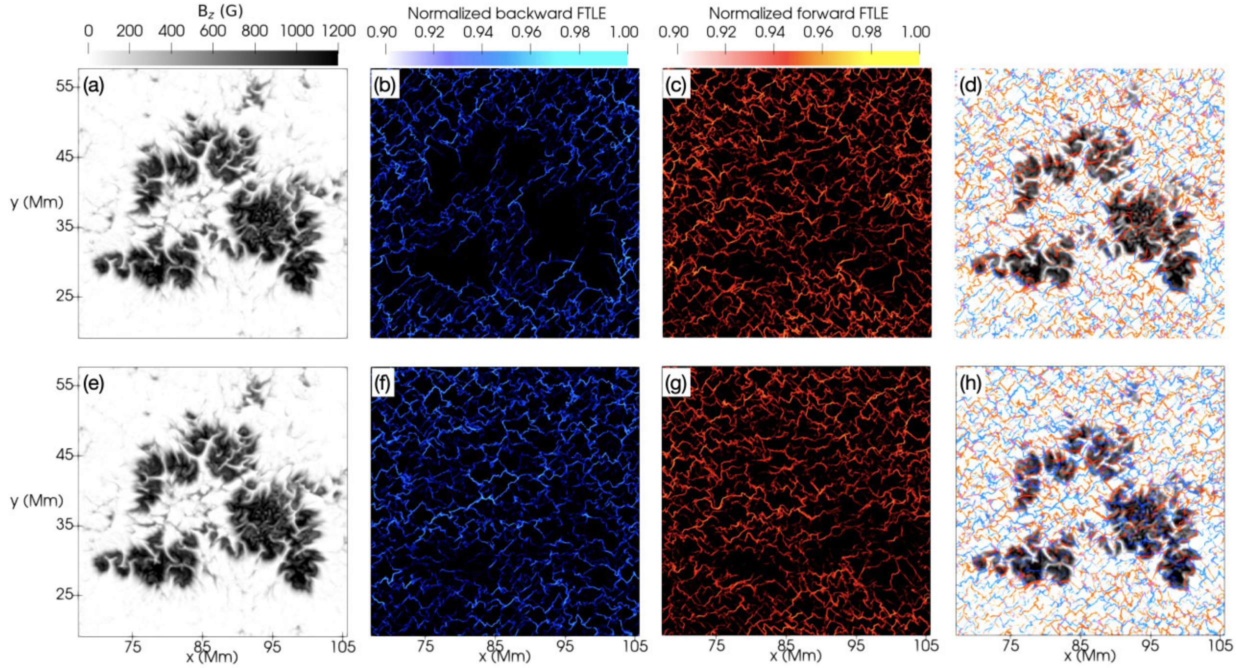

Figure 8. The normalized forward- and backward-FTLE field superimposed on the magnetic field, corresponds to the highlighted region in Fig. 3. Panels (a) and (e) show the magnetic field averaged over the time of integration. (b) and (f) show the backward-FTLE ridges resulting from a 20-min integration of the simulated and recovered velocity fields, respectively. Panels (c) and (g) show the resulting forward integrated FTLE fields. In panels (d) and (h) both the forward and backward integrated FTLE fields are superimposed on the magnetic field. The ridges resulting from the 20-min integration correspond to features on the length scales of granules.

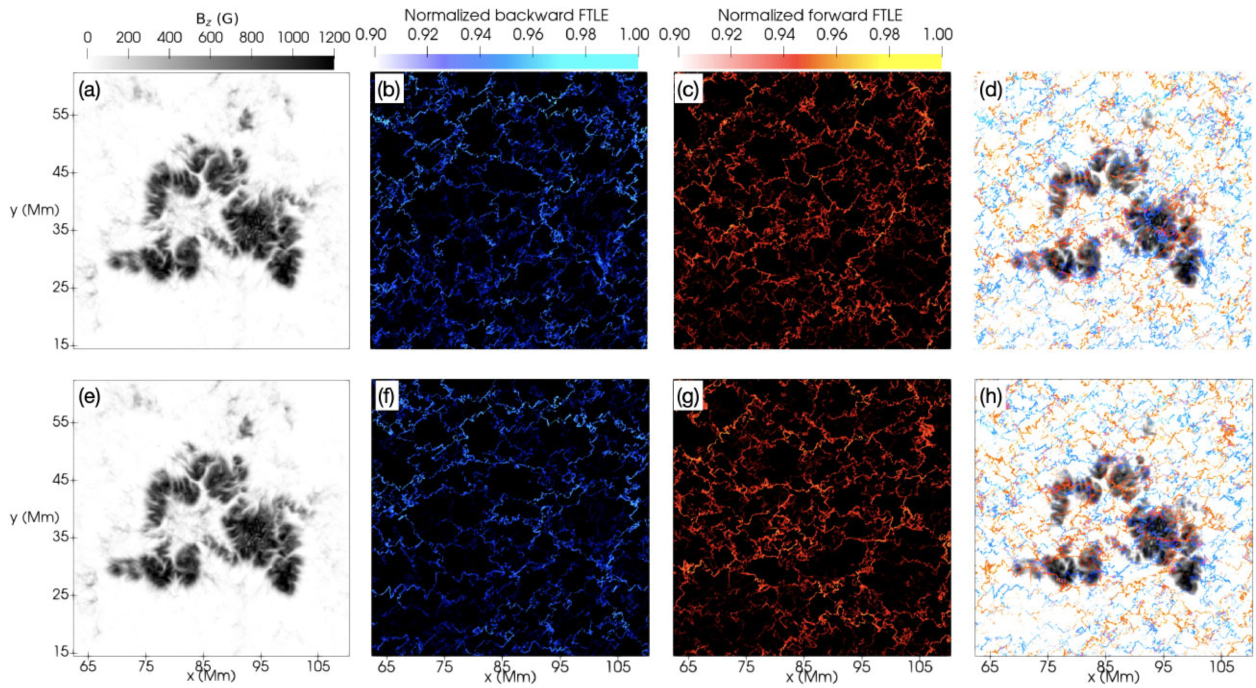

Figure 9. The same as Fig. 8 however the FTLE ridges here result from a 100-min integration corresponding to mesogranular features.

suppressing the plasma motions. Furthermore, attracting barriers delimit the entire pore indicating the presence of downflows at the umbral boundary. The red, repelling, barriers are slightly weaker over the pore but still dominate much of the region, indicating the outward transport of material by the flow, i.e. the dissipation of the magnetic field over region. suggesting that the plasma is being transported out of this region. At the time shown in the figure, the simulated AR is decaying and being is becoming more spread out. DV is able to reconstruct the repelling structures very well, maintaining most of the fine structure at granular scales. DV is also able to reconstruct the strongest attracting features, however largely overestimates much of the weaker ridges over the pore. Thus DV is able to capture short-lived and sub-granular diverging flow features well, but less so the converging flow features on these scales. In the 100-min integrated field, presented in Fig. 9, DV is able to reconstruct both attracting and repelling barriers on mesogranular scales well.

It should be noted here that the same analysis was performed on the R2D2 simulation used to train DV. The left column of Fig. 10 shows that in the R2D2 simulation, there is the presence coherent structures indicative of the flow dynamics surrounding the AR. One such example is represented by the strong attracting ridge, which delimits the pore and is further delimited by a strong repelling ridge in both the short and long-term integrated fields. This describes a strong horizontal flow into the pore, like an Evershed flow, which is typically present around strong pores and sunspots. The NN correctly identifies the strong repelling flows structures delimiting the AR and some of the attracting structures within at the granular scale, suggesting that the NN is able to emulate the flow physics within and around the AR well.

These Evershed-like flow structures are not present in the MURAM simulation, shown in Figs 8 and 9. DV also identifies that these structures are not present surrounding the AR and produces flows that are more consistent with the MURAM flows. Thus we can discern that DV has been able to generalize well to the MURAM simulation flows, since the NN does not simply mimic the flow features present in training set. Therefore we determine DV is able to consistently produce flow structures equally well across simulations. In particular the forward-FTLE fields correlate well with the original at granular scales and up, whereas the backward-FTLE field is underestimated at (sub)granular scales.

In Fig. 11, the distribution of the 20-min forward-FTLE field is shown for two different regions in the flow, over times when there is little magnetic flux present in the photosphere and when the AR is fully emerged (2 and 83 h, respectively). Panels in the left column of the figure present the forward-FTLE distribution over a section of the photosphere where $<$kG magnetic flux is present throughout the simulation. On the other hand, the right column shows the distribution over the region where $>$kG magnetic flux has emerged in the later time. The distributions are shown in blue and red for early and late times, respectively. Panel (a) of the aforementioned figure shows that the distribution in quiet regions at both early and late times remains almost constant, and panel (b) reveals that there is a difference when magnetic flux emerges into the photosphere. We can see this more clearly as the mean over the AR changes significantly, relative to the size of the FTLEs ($\approx$10 per cent difference). DV is able to reproduce this change in the shape of the distribution, however not as accurately, due to the errors in the predicted velocities. This indicates that DV is recovering key physical aspects of the flow, at the granular scale, which determine the its behaviour in the presence of $>$kG magnetic flux.

4 CONCLUSIONS

For the first time, DV was tested on the recovery plasma flow features on scales from the mesogranular (several Mm) to subgranular ($\geq$0.096 Mm) scales in simulated AR flows. We assessed the performance of DeepVel using a different simulation data set from the training set in order to remove any bias that may have come from DV overfitting to the training set. This helped us figure out how effective DV was, especially in QS regions and ARs where more intense magnetic flux is present. One of the main things we found is that DV can recover topological flow features pretty accurately, especially when it comes to things like divergence and swirling motions at granular scales. Because of this, DV can be used to uncover information about horizontal inflow and outflow in ARs at scales that have not been studied much before. Considering the performance of DV on different simulations, it is expected that proper training DV will also perform well on observational data.

In this work, DV was not trained to work with observations. In order to do so, a number of features need to be considered. The first of a long list includes atmospheric turbulence can cause the blurring of images and thus data sets from ground-based telescopes would need cleaning or deconvolving for use. Further, optics of telescopes can cause aberrations in images due the diffraction of light as it passes through a lens. The appearance of this distortion is an airy disc in the images which alters contrast. This effect can be approximated using a point spread function (PSF) if the parameters are known (see, for example, the discussion by Wedemeyer-Böhm 2008). In order to apply DV to telescope imagery, the training set would have to be convolved with the PSF or the observations would need to be deconvolved to match the training set as closely as possible. Additionally, observations usually contain the 5-min oscillations caused by p modes, which cause the brightening and dimming of images, work toward mitigating this is highlighted by Tremblay et al. (2018) by filtering high-frequencies from images. Where resolutions of images differ significantly from the training set, it is expected that DV performance may suffer, higher resolution data can be degraded and the training images can be degraded to match lower resolution observations in order to optimise performance of DV. We argue that thanks to the existence of high resolution space-based telescopes, some data sets may readily be applied with little of these modifications. For example, the balloon-borne instrument Sunrise/IMAX (see Martínez Pillet et al. 2011) and its more modern equivalent TuMag (see del Toro Iniesta et al. 2025) have wide apertures which lessens the distortion due to the lens and also sit high in the Earth's atmosphere so that atmospheric effects are of little concern.

Overall, our work shows that DV performed well, though it struggled when recovering smaller scale features like the apparent vortices in umbral regions. This is likely because strong magnetic fields in those areas suppress plasma motions and DV has trouble replicating these flows due to their low-speeds, which is accounted for by only a small amount of the data in the training set. Still, DV was able to differentiate between various flow behaviours, showing it could adapt to different scenarios within the photosphere. DV effectively reproduced the velocity field distributions, though there were some small inaccuracies in the tails of the velocity distributions; inaccuracies such as these are somewhat expected, as supervised learning methods with continuous outputs will tend to produce values that are smooth. We measured DV's accuracy using PC, RMSE, and alignment (dot product) to effectively measure DV performance on singular outputs. In QS regions, the results were good, but in ARs, the performance was a bit lower. For example, the mean alignment in QS regions was 0.674, but it dropped to 0.501 in ARs. In other words tracking flows becomes more difficult in umbral regions, limiting the quality and therefore conclusions that may be drawn from flows identified by DV within the pore of an AR. This is probably because of the impact that magnetic flux has on plasma motion as well as the inputs having a lower contrast in umbral regions as opposed to the typical contrasts present in QS regions. Additionally, the lower correlation for divergence and curl in ARs suggests that DV has a harder time recovering more complex flow structures when intense magnetic fields are involved. These issues likely arise as

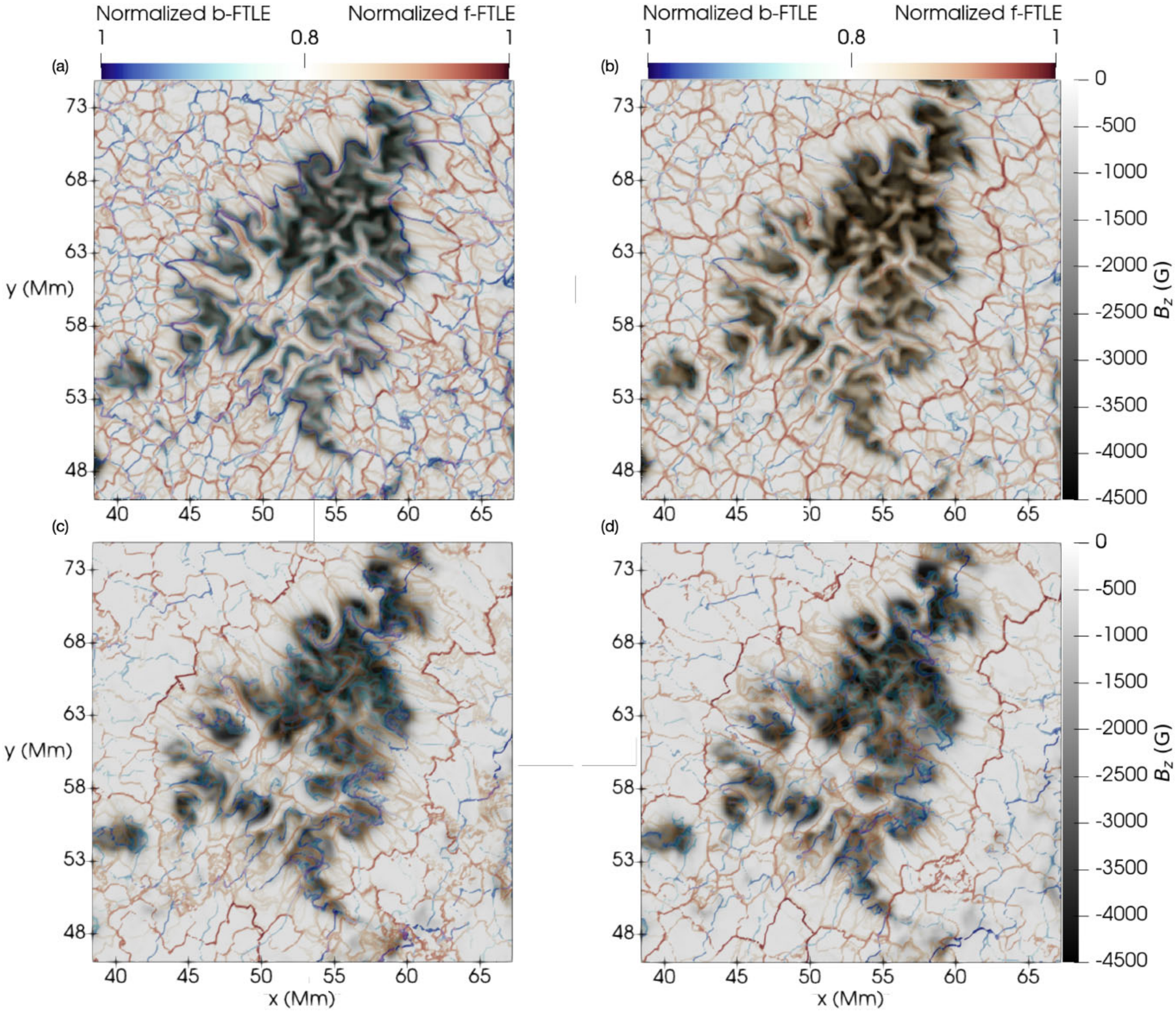

Figure 10. The normalized forward- and backward-FTLE fields superimposed on the magnetic field from the R2D2 simulation. The region shown corresponds to the highlighted region in Fig. 1. Panels (a) and (b) show the FTLE fields resulting from integrating over 20-min and (c) and (d) over 100-min. Panels (a) and (c) show the FTLE fields calculated using velocities from the R2D2 simulation and (b) and (d) the velocities recovered by DV. The 20-min fields show structures in the FTLE fields on the scale of granules and the 100-min fields show mesogranular structures.

the horizontal velocities become relatively small within the AR, and the intensities exhibit less contrast within these regions, thus small differences in how intensity maps are constructed may become significant in the production accurate velocities. Our results highlight that despite differences in boundary conditions in synthetic data and the numerical scheme for RT, DV is able to produce consistent results that provide insight into the types of features present over ARs, even if their exact topology is not captured fully.

This study is the first to present an analysis on how recovered flows are able to reproduce the transport barriers present in the original flow. DV presents a great performance in replicating the overall magnetic field transport dynamics in the simulation. The reconstruction of attracting and repelling barriers matched well with the simulation, especially at granular and mesogranular scales. This confirms that despite the inevitable errors in the recovered velocities, DV was able to reconstruct enough information such that these did not majorly impact how particles motion and trajectories in the flow, which shows that DV is a solid tool for studying the movement of small-scale magnetic fields. On top of that, DV successfully captured the major changes in flow behaviour during the formation of ARs, like when magnetic flux suppresses plasma motion and pushes magnetic fields outward. Moreover, DV did not introduce the topological features, observed in the FTLE fields, present in the training data in the analysed data from MURAM. This is a strong indicator that the model constructed by DV from R2D2 is good for generalizing to new data containing similar features and that DV is able to emulate the properties of AR flows well from the training set.

The ability of DV to emulate flow properties from the provided training set means that DV will recover better the characteristics of the flow, so long as they are in line with flow features present in the training simulation. Typically, NNs may suffer from over-fitting and create features that are not present or fail to reproduce anything coherent in the worst case, when passing new data through the network. Therefore, results from NNs should be considered

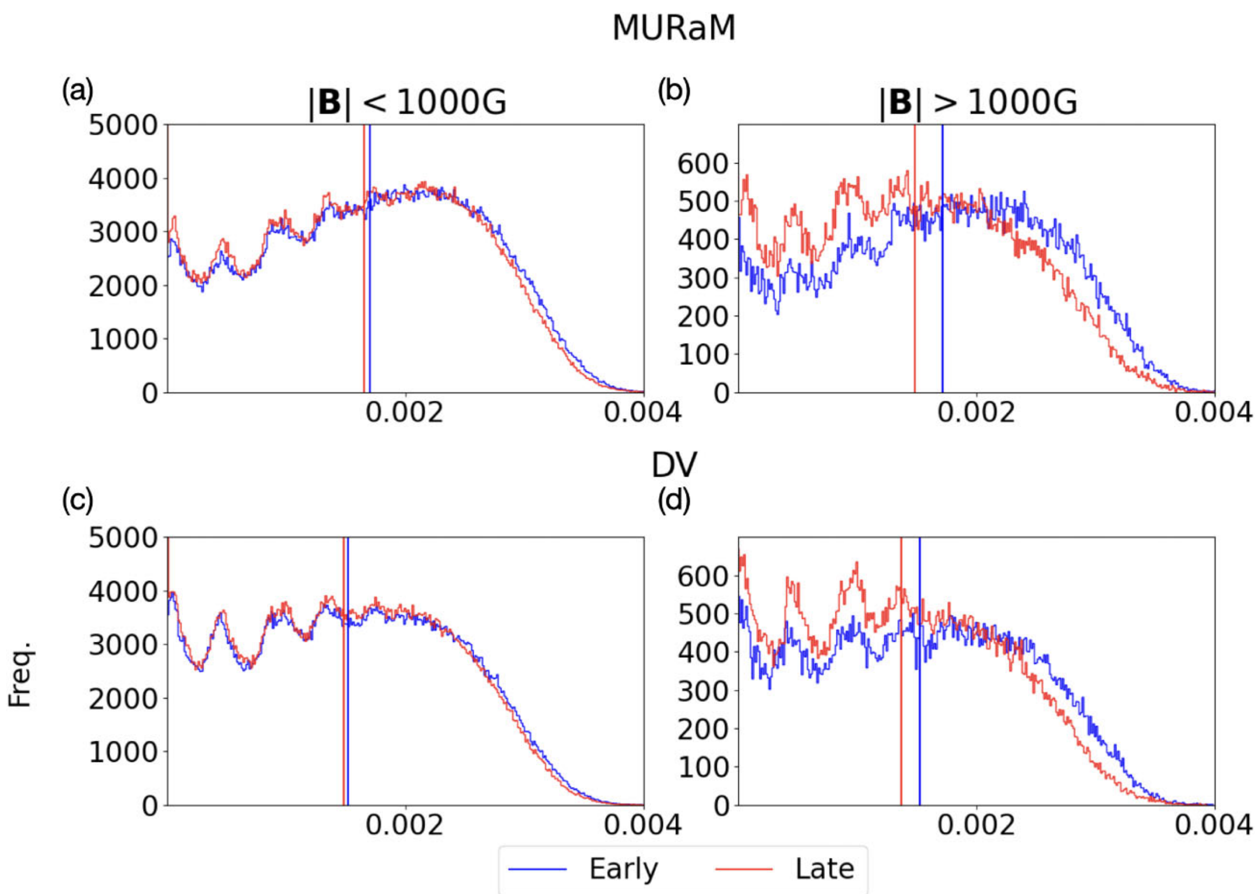

Figure 11. Distribution of the 20-min forward-FTLE field, at an early ($t = 2$ h) and a late ($t = 83$ h) time in the simulation, over a region where the AR emerges and >kG magnetic flux is present at the later time. Panels (a) and (b) show the forward-FTLE distributions from MURAM over the QS region and AR, respectively. Panels (c) and (d) show the forward-FTLE distribution from DV over the QS region and AR. The vertical lines show the means for each distribution. There are two key things to note. First, at late times when the AR becomes present, there is a spiking in the frequency of the smaller FTLE values, i.e. a reduction in the complexity of flow topology, compared to earlier times when the photosphere is mostly quiet and there is little difference in the distributions. This spiking causes a reduction in the mean forward-FTLE over the AR, which is captured by both recovery methods. Secondly, DV produces results which are consistent with the simulation at both times.

carefully. In the case of supervised learning, it is expected that NNs will perform well at reproducing information from data similar to what the NN has already seen during training. In other words, if the input data provided to the NN has similar properties to that of the training set, then the outputs produced will be consistent with the training set and hence consistent with the input data if it shares the same flow properties/physics of the training set. If the inputs have different properties to that of the training inputs, the outputted flows will be inconsistent with the training set flows, and most likely inconsistent with the test set flows as the NN has not learned to emulate the test simulation. In the case of providing new input data, the NN expects that most, if not all, of the properties of the input data will remain within the limits of the training set. Thus, if we test our model with images from a new simulation, which shares similar physics and properties of the training set, the outputs are expected to be reasonably good and will typically produce results that are in accordance with properties from the training set with some expected error. In our case, the physics and properties of the MURAM and R2D2 simulations (e.g. cadence, resolution) are in agreement with each other. Other aspects of the simulation differ, the physical influence of a shallow bottom boundary, turbulent diffusion and an untwisted flux tube with a different initial shape present a somewhat different scenario to the R2D2 simulations. Synthetic images from MURAM also differ slightly to R2D2 due to differences in the numerical schemes for how the optical surfaces are calculated (for the full descriptions of the schemes, see Rempel & Cheung 2014; Hotta & Iijima 2020). Despite these differences, once the histograms of continuum intensities were matched, the results of applying DV to MURAM synthetic images are expected to be at least consistent with the flows in R2D2. This is confirmed by the high correlation between DV and original velocity fields and by the successful identification of the material transport barriers of MURAM by DV.

A number of changes can be made to the NN itself as well as the training process in order to improve DV predictions. Since new changes to DV should be validated carefully for wider applications they are out of the scope of this work. For example, the NN can be redesigned to consider features on multiple scales such as the multiscale NN from Ishikawa et al. (2022), properties from the training set can be considered sample by sample rather than considering one set of properties for describing the whole set, and different training processes could be applied (i.e. altering the network architecture throughout the training or applying different activation functions/loss functions). In addition to this, in the current testing, our work was limited by the cadence of the simulations available, the sound speed of the simulation near the surface reaches around 10 $\mathrm{km\,s^{-1}}$, this means that features may move up to around 1 Mm between frames, close to the training patch size. Thus it may well be the case that features often disappear and appear over each patch when they are close to the edges causing there to be no discernible pattern between some of the flows and features from one frame to the next, thus training with larger patches and higher cadences would further highlight the capabilities of DV for flow recovery in the photosphere. Furthermore, DV could be trained on a wide range of simulations covering most of the phenomena present on the solar surface such as penumbral flows, the aftermath of reconnection and more, and images could even be projected onto a sphere to allow predictions of flows on the full disc. This would potentially provide an ensemble solution, useful for forecasting the true velocity within some range, however, the storage and time taken for training such a network would be very computationally and financially expensive.

DV, as we have presented, is a robust tool, which can be trained to handle complex dynamical scenarios in a solar context. The validation that we have performed shows that DV has the capability to recover the smallest scale flows present in synthesized images from simulations, instantaneously with a good level of accuracy. In conclusion, our study shows that DeepVel has a lot of potential for recovering velocity fields from continuum intensity data in ARs. While it does have some limitations within the umbra of ARs, DV is a promising tool for studying solar photospheric flows and provides a solid foundation for future improvements. Future work could focus on reducing the discrepancies in AR regions, particularly in recovering smaller scale features and better capturing how magnetic flux affects plasma flows.

ACKNOWLEDGEMENTS

We acknowledge and give our thanks to the insightful comments provided by the referee, which have improved the quality of this manuscript. ML thanks STFC for the studentship and is thankful for the patience of his supervisor and the support of his collaborators. VF, GV, and SSAS are grateful to the Science and Technology Facilities Council (STFC) grants ST/V000977/1 and ST/Y001532/1. VF, GV, and SSAS thanks the Institute for Space-Earth Environmental Research (ISEE, International Joint Research Programme, Nagoya University, Japan). VF, GV, and IB thank the Royal Society, International Exchanges Scheme, collaboration with Greece (IES/R1/221095), India (IES/R1/211123), and Taiwan (IEC/R3/233017) for the support provided. This research has also received financial support from the European Union's Horizon 2020 research and innovation program under grant agreement no. 824135 (SOLARNET). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising. We also thank Miguel for his patience during writing this paper.

DATA AVAILABILITY

The data used and generated in this analysis will be shared upon reasonable request to the corresponding author.

REFERENCES

Asensio Ramos A., Requerey I. S., Vitas N., 2017, A&A, 604, A11
Attie R., Kirk M. S., Thompson B. J., Muglach K., Norton A. A., 2018, Space Weather, 16, 1143
Bahng J., Schwarzschild M., 1961, ApJ, 134, 312
Beeck B. et al., 2012, A&A, 539, A121
Bjørgen J. P., Sukhorukov A. V., Leenaarts J., Carlsson M., de la Cruz Rodríguez J., Scharmer G. B., Hansteen V. H., 2018, A&A, 611, A62
Cabral B., Leedom L. C., 1993, Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, VOL. 1 Imaging vector fields using line integral convolution, p. 263
Chen F., Cheung M. C. M., Rempel M., Chintzoglou G., 2023, ApJ, 949, 118
Chian A. C.-L., Rempel E. L., Aulanier G., Schmieder B., Shadden S. C., Welsch B. T., Yeates A. R., 2014, ApJ, 786, 51
Chian A. C.-L., Silva S. S. A., Rempel E. L., Gošić M., Bellot Rubio L. R., Kusano K., Miranda R. A., Requerey I. S., 2019, MNRAS, 488, 3076
del Toro Iniesta J. C. et al., 2025, preprint (arXiv:2502.08268)
Freytag B., Steffen M., Ludwig H. G., Wedemeyer-Böhm S., Schaffenberger W., Steiner O., 2012, J. Comput. Phys., 231, 919
Gudiksen B. V., Carlsson M., Hansteen V. H., Hayek W., Leenaarts J., Martínez-Sykora J., 2011, A&A, 531, A154
Haller G., 2015, Ann. Rev. Fluid Mech., 47, 137
Haller G., Yuan G., 2000, Physica D: Nonlinear Phenomena, 147, 352
Hodges J. L., 1958, Ark. Mat., 3, 469
Hotta H., Iijima H., 2020, MNRAS, 494, 2523
Ishikawa R. T., Nakata M., Katsukawa Y., Masada Y., Riethmüller T. L., 2022, A&A, 658, A142
Khomenko E., Collados M., Díaz A., Vitas N., 2014, Phys. Plasmas, 21, 092901
Löhner-Böttcher J., Schmidt W., Schlichenmaier R., Steinmetz T., Holzwarth R., 2019, A&A, 624, A57
Martínez Pillet V. et al., 2011, Sol. Phys., 268, 57
Martínez-Sykora J., De Pontieu B., Hansteen V., McIntosh S. W., 2011, ApJ, 732, 84
November L. J., Simon G. W., 1988, ApJ, 333, 427
Orozco Suárez D., Rubio L., Martínez-Pillet V., Bonet J., Vargas Dominguez S., del Toro Iniesta J. C., 2010, A&A, 522, A101
Parker E. N., 1963, ApJ, 138, 552
Rempel M., 2015, ApJ, 814, 125
Rempel M., Cheung M. C. M., 2014, ApJ, 785, 90
Rempel M., Schüssler M., Knölker M., 2009, ApJ, 691, 640
Rempel E. L., Chertovskih R., Davletshina K. R., Silva S. S. A., Welsch B. T., Chian A. C.-L., 2022, ApJ, 933, 2
Rieutord M., Roudier T., Ludwig H.-G., Nordlund A., Stein R., 2001, A&A, 377, L14
Schuck P. W., 2008, ApJ, 683, 1134
Silva S. S. A. et al., 2023, ApJ, 948, L24
Stein R. F., Nordlund A., 1998, ApJ, 499, 914
Tremblay B., Attie R., 2020, Frontiers Astron. Space Sci., 7, 25
Tremblay B., Thierry R., Rieutord M., Vincent A., 2018, Sol. Phys., 293, 57
Tremblay B., Cossette J.-F., Kazachenko M. D., Charbonneau P., Vincent A., 2021, J. Space Weather Space Clim., 11, 9
Tziotziou K. et al., 2023, Space Sci. Rev., 219, 1
van Driel-Gesztelyi L., Green L. M., 2015, Living Rev. Sol. Phys., 12, 1
Verma M., Steffen M., Denker C., 2013, A&A, 555, A136
Vögler A., Shelyag S., Schüssler M., Cattaneo F., Emonet T., Linde T., 2005, A&A, 429, 335
Wedemeyer-Böhm S., 2008, A&A, 487, 399
Yeates A. R., Hornig G., Welsch B. T., 2012, A&A, 539, A1

This paper has been typeset from a TEX/LATEX file prepared by the author.