

Inferring depth-dependent plasma motions from surface observations using the DeepVel neural network

Benoit Tremblay^{1,2,3,*}, Jean-François Cossette^{3,4}, Maria D. Kazachenko^{2,5}, Paul Charbonneau³, and Alain Vincent³

¹ National Solar Observatory, 3665 Discovery Dr., Boulder, 80303 CO, USA

² University of Colorado, Astrophysical and Planetary Sciences, 2000 Colorado Avenue, Boulder, 80303 CO, USA

³ Université de Montréal, Département de Physique, 1375 Avenue Thérèse-Lavoie-Roux, Montréal, H2V 0B3 QC, Canada

⁴ Environment & Climate Change Canada, Dorval, H9P 1J3 QC, Canada

⁵ University of California, Space Sciences Laboratory, 7 Gauss Way, Berkeley, 94720 CA, USA

Received 5 June 2020 / Accepted 26 November 2020

Abstract—Coverage of plasma motions is limited to the line-of-sight component at the Sun’s surface. Multiple tracking and inversion methods were developed to infer the transverse motions from observational data. Recently, the DeepVel neural network was trained with computations performed by numerical simulations of the solar photosphere to recover the missing transverse component at the surface and at two additional optical depths simultaneously from the surface white light intensity in the Quiet Sun. We argue that deep learning could provide additional spatial coverage to existing observations in the form of depth-dependent synthetic observations, i.e. estimates generated through the emulation of numerical simulations. We trained different versions of DeepVel using slices from numerical simulations of both the Quiet Sun and Active Region at various optical and geometrical depths in the solar atmosphere, photosphere and upper convection zone to establish the upper and lower limits at which the neural network can generate reliable synthetic observations of plasma motions from surface intensitygrams. Flow fields inferred in the photosphere and low chromosphere $\tau \in [0.1, 1)$ are comparable to inversions performed at the surface ($\tau \approx 1$) and are deemed to be suitable for use as synthetic estimates in data assimilation processes and data-driven simulations. This upper limit extends closer to the transition region ($\tau \approx 0.01$) in the Quiet Sun, but not for Active Regions. Subsurface flows inferred from surface intensitygrams fail to capture the small-scale features of turbulent convective motions as depth crosses a few hundred kilometers. We suggest that these reconstructions could be used as first estimates of a model’s velocity vector in data assimilation processes to nowcast and forecast short term solar activity and space weather.

Keywords: active region / chromosphere / convection zone / deep learning / granulation / photosphere / velocity fields

1 Introduction

Prediction of space weather critically depends on how we can use observations to numerically simulate the physical processes in the solar photosphere and above. *Data-inspired* numerical models of the Sun, which are built upon our best understanding of radiative magnetohydrodynamics (MHD), have produced increasingly realistic simulations of the Quiet Sun (e.g., Stein & Nordlund, 1998; Vögler et al., 2005; Abbett, 2007) and of Active Regions (e.g., Rempel & Cheung, 2014; Jiang et al., 2016). Recent models even generate energy releases consistent with flaring events (Cheung et al., 2018). *Data-driven* simulations or the implementation of more complex methods

like *data assimilation* could further bridge the gap between MHD models of the upper convection zone to corona and observations, as suggested in Abbett & Fisher (2010), and implemented by e.g. Fisher et al. (2015) and Hayashi et al. (2018) for the solar atmosphere. Data assimilation is defined as adjusting the initial condition of a model (not its parameters!) to minimize errors between model predictions and satellite/ground-based observations within a window of time (i.e., the assimilation window; Bouttier & Courtier, 2002). Generated from the new initial condition is an improved representation of a given epoch of the Sun and from which to validate the physics encompassed by the model within the assimilation window (i.e., nowcasting) and to forecast short term solar activity (e.g., the evolution of an active region and the onset of space weather events) beyond the assimilation window. Remaining errors

*Corresponding author: btremblay@nso.edu

between observations and the adjusted model predictions can be interpreted as a measure of the degree of realism of the simulation. In this sense, the assimilation of data within a time window containing a space weather event could improve our understanding of the physical mechanisms and physical conditions associated with its onset and, by extension, improve our ability to forecast future events.

Performing data assimilation in a MHD model is not without difficulties. Firstly, an estimate of the initial three-dimensional model state (i.e., physical values at the beginning of the assimilation window) is required to begin the process. Secondly, the aforementioned simulations solve for physical quantities that, due to physical or instrumental limitations, cannot be derived from direct observations or can only be derived at select depths. One such model variable is the velocity vector. For example, Dopplergrams can only capture the line-of-sight component of plasma motions at the Sun's surface. Various methods have been developed to reconstruct the missing transverse component from tracking algorithms and observed intensitygrams (e.g., [November & Simon, 1988](#); [Potts et al., 2004](#); [Rieutord et al., 2007](#)) or magnetograms (e.g., [Fisher & Welsch, 2008](#)), or from magnetograms and physical principles (e.g., [Longcope, 2004](#); [Schuck, 2008](#); [Kazachenko et al., 2014](#); [Attie & Innes, 2015](#)) in the photosphere. Recently, neural network computing has been used in conjunction with numerical models of solar granulation to be able to recover the transverse velocity vector in photospheric plasma of the Quiet Sun. The fully-convolutional DeepVel neural network ([Asensio Ramos et al., 2017](#)) infers instantaneous transverse plasma motions at three simultaneous optical depths $[\tau] \approx \{1, 0.1, 0.01\}$, including the surface, photosphere and chromosphere, from pairs of intensitygrams. A comparison between intensity-based methods identified DeepVel as the method best suited to generate instantaneous synthetic observations of plasma motions at the Quiet Sun's surface, i.e., plasma motions that emulate the physics of a numerical model but are made to look as though they were observed by a specific instrument ([Tremblay et al., 2018](#)). Beyond assimilation, plasma motions impact the amount of free energy in the solar corona that can then be released in the form of flaring events, e.g. through the transport by flows of magnetic energy from surface layers to the solar atmosphere or the shearing and twisting of field lines ([Welsch, 2006](#)). The shuffling of magnetic footpoints or the coherent twisting of magnetic loops by photospheric motions can lead to an accumulation of twist in coronal loops and subsequent reconnections (e.g., [Parker, 1988](#)).

We propose that reconstructions of the transverse velocity vector could benefit data assimilation in a MHD model by providing additional coverage of the state variables of the model, either as instantaneous synthetic observations or first estimates of the model state. Inferred plasma motions or the electric field computed from it (e.g., [Kazachenko et al., 2014](#)) could also be used as boundary conditions from which to drive MHD simulations of the solar atmosphere ([Fisher et al., 2015](#)). Alternatively, these could be combined with other reconstructions to estimate quantities such as the Poynting flux and study their evolution in the context of short term solar activity. In this paper, we aim to provide additional spatial coverage for the velocity vector by generating depth-dependent reconstructions in the Quiet Sun (hereafter QS) and Active Regions (hereafter ARs) with the DeepVel neural network.

The data preprocessing and training of DeepVel are described in [Section 2](#). In [Section 3](#), continuum intensity and velocity fields computed by simulations are used to test the flows inferred above and below the surface. We conclude and discuss future work in [Section 4](#).

2 Training sets and procedures

The DeepVel neural network builds an approximation through a series of transformations (i.e., convolutions) that relates the transverse velocity vector field $[v_\perp(x, y, z, t_i)]$ (i.e., the output) at time $[t = t_i]$ to pairs of maps of the continuum intensity $I_c(x, y, \tau \approx 1, t = \{t_i, t_i + \Delta t\})$ at the surface (i.e., intensitygrams; the input), where $[\Delta t]$ is the timestep between images. Training of DeepVel is performed using computations of the input and output quantities from numerical simulations (i.e., through supervised learning). We refer to [Asensio Ramos et al. \(2017\)](#) for a detailed description of the neural network architecture. Depths $[z]$ are single or multiple optical depths $[\tau]$ or geometrical depths $[d]$ or a combination of both. Geometrical depths d are measured from the radius at which an average optical depth of $\tau \approx 1$ is achieved ($d = 0$ km) towards the solar interior (i.e., $d > 0$ km below the surface). Optical depth is the natural scale for radiative transfer and by extension for derivations from observations. Fixed geometrical heights are difficult to derive in the context of radiative transfer, but they are the natural scale in simulation meshes and are thus used when training and testing DeepVel against synthetic data. The interpretation of geometrical heights when using real observational data as input is expected to be more challenging.

Outputs are dependent on the properties of the data presented during training. Different versions of DeepVel will thus need to be trained depending on the physics of the model used for training and the preprocessing of the inputs that will be used for execution. This extends to the horizontal pixel sizes Δx and Δy , and the cadence Δt . The number of output layers must also be modified according to the depths z to probe.

The training and validation sets for DeepVel consist of subimages $I_c(x = [x_i, x_i + 50\Delta x], y = [y_i, y_i + 50\Delta y], \tau \approx 1, t = \{t_i, t_i + \Delta t\})$ and $v_\perp(x = [x_i, x_i + 50\Delta x], y = [y_i, y_i + 50\Delta y], z, t_i)$, where x_i and y_i are randomly-selected positions in the full field-of-view images (i.e., sub-images of dimensions 50 by 50 pixels² at each depth), z is a set of optical or geometrical depths, and t_i is a randomly-selected time within the sequence covered by the simulation data. Weights and biases in the network architecture are adjusted in an attempt to optimize the mean squared errors (i.e., the cost function) between the flows inferred by DeepVel and the training set data. Their values are updated after each training epoch only if the mean squared errors for the validation set are improved. The number of training epochs $[n_e]$ refers to the number of times the training and validation sets are presented to the neural network. This process is repeated until the cost function has achieved convergence or no longer varies significantly. Trained convolutional neural networks are capable of generalizing their outputs for any input image dimensions. The test sets that are used to evaluate DeepVel post training feature full field-of-view images. Intensitygrams are normalized by the median of the continuum intensity over the time sequence covered by the dataset. Each

component of the velocity vector is shifted by its minimum value in the dataset and is then normalized by the difference between its maximum and minimum values so that $v_x, v_y \in [0, 1]$. No data augmentation was performed for the training and validation sets, although it could further improve performance and generalization.

A version of DeepVel was trained in Tremblay et al. (2018) to generate synthetic velocity maps at $\tau \approx 1$ that are consistent with the spatial and temporal scales of level-2 data from the *Helioseismic Magnetic Imager* (HMI: Schou et al., 2012) onboard the *Solar Dynamics Observatory* (SDO: Hoeksema et al., 2014). The synthetic training set, hereafter referred to as the STAGGER dataset, was derived from slices of the STAGGER¹ magneto-convection simulation of the QS (Stein, 2012; Stein & Nordlund, 2012). These slices cover a patch of granulation at select optical depths $\tau \approx \{1, 0.1, 0.01\}$ within a field-of-view of dimensions $L_x \times L_y = 96.768 \times 96.768 \text{ Mm}^2$ that is sampled at a horizontal spatial resolution $\Delta x = \Delta y = 96 \text{ km}$, downsampled by a factor two from the native resolution of the simulation (i.e., $\Delta x = \Delta y = 48 \text{ km}$), and cadence $\Delta t = 60 \text{ s}$. The STAGGER simulation data was convolved with the SDO/HMI point spread function (PSF) from Wachter et al. (2012) to emulate the instrument and was then resampled to the SDO/HMI spatial resolution near disk center ($\Delta x = \Delta y \approx 368 \text{ km pixel}^{-1}$) using nearest-neighbor sampling. We refer to Tremblay et al. (2018) for a more detailed description of the pre-processing steps. A total of 2000 examples were presented to the network over 10 training epochs, with an additional 200 examples for the validation steps. The test set is comprised of a 30-min sequence of full field-of-view images at all three optical depths. Although the analysis presented in this work was limited to $\tau \approx 1$, the neural network was actually trained to infer $v_i(x, y, \tau, t_i)$ at $\tau \approx \{1, 0.1, 0.01\}$ simultaneously. Intensitygrams $I_c(x, y, \tau \approx \{0.1, 0.01\}, t_i)$ in the STAGGER dataset have undergone the reversal of the granulation pattern with respect to $I_c(x, y, \tau \approx 1, t_i)$ (not shown) and are thus located at least a few hundred kilometers above the surface (e.g., Cheung et al., 2007). Optical depth $\tau \approx 0.1$ is close to the transition between the photosphere and chromosphere (e.g., Cranmer et al., 2007). This version of DeepVel is thus used to evaluate QS flow inversions in the photosphere and chromosphere from surface data.

Additional versions of DeepVel were trained to infer flows above and below the surface using the high-resolution and high-cadence MURaM simulation of solar granulation (hereafter the MURaM-QS simulation; Vögler et al., 2005). A patch of QS is evolved within a volume $V \equiv L_x \times L_y \times L_z = 6.144 \times 6.144 \times 4 \text{ Mm}^3$ which spans from 2.3 Mm in the upper convection to 1.7 Mm above the surface. The field-of-view is ≈ 244 times smaller in area than in the STAGGER dataset, but it is much more detailed. MURaM-QS achieves a pixel-size $\Delta x = \Delta y \equiv 16 \text{ km}$ per pixel in the horizontal direction, a value that is close to the Daniel K. Inouye Solar Telescope (DKIST: Rimmele et al., 2020) spatial resolution ($\Delta x \approx 20 \text{ km}$: Warner et al., 2018). Training was performed using simulation data at its native resolution and timestep $\Delta t = 2 \text{ s}$. Slices were extracted at constant optical depths $\tau \approx \{1, 0.1, 0.01\}$ and constant

geometrical depths $d = \{0, 144, 560\} \text{ km}$. Unlike for the STAGGER dataset, subimages cover an area that is smaller than the average granule. We therefore increase the size of the training and validation sets to 30,000 and 3000 patches each and perform 30 training epochs.

Current applications of DeepVel have been limited to the QS (e.g., Asensio Ramos et al., 2017; Tremblay et al., 2018). We introduce a generalization of the algorithm for ARs that was trained from computations of the MURaM simulation of a sunspot (hereafter MURaM-ARs; Rempel & Cheung, 2014) at optical depths $\tau \approx \{1, 0.1, 0.01, 0.001\}$ and geometrical depth $d = 1000 \text{ km}$. The sunspot, which is $\approx 25 \text{ Mm}$ in diameter, is located at the center of the field of view which spans 98.304 by 98.304 Mm^2 . The MURaM-ARs data shares the same spatial resolution as the raw STAGGER data (i.e., $\Delta x = \Delta y = 96 \text{ km pixel}^{-1}$) and was resampled to the SDO/HMI spatial resolution following the same procedure (hereafter the ARs dataset). The cadence $\Delta t = 45 \text{ s}$ coincides nicely with the timestep between consecutive SDO/HMI level-2 products (Hoeksema et al., 2014). The training and validation sets include 3600 and 900 examples each, with each subimage containing a random fraction of QS and ARs. The test set is comprised of a 30-timestep sequence of full field-of-view images.

Computations were performed on a NVIDIA-GTX 960 GPU and a NVIDIA-GTX 1080 Ti GPU using the *Keras* library with the *Tensorflow* backend.

3 Results

3.1 Statistical metrics

Each version of the DeepVel neural network is tested against full field-of-view images that were generated by the same numerical simulation from which the training and validation sets were derived. These best case scenarios are used to identify the range of depths at which flow inversions can be performed reliably. Maps of the velocity fields are shown to provide visual aid for spatial features. Only patches (i.e., sub-fields) are displayed for clarity. Scatterplots are used to study the transverse velocity amplitudes within those patches. Metrics for the test sets are computed over the entire field-of-view. They include the root mean squared errors

$$\text{RMSE}(\mathbf{v}_{\text{ref},t}, \mathbf{v}_{D,t}) = \sqrt{\langle (\mathbf{v}_{\text{ref},t} - \mathbf{v}_{D,t}) \cdot (\mathbf{v}_{\text{ref},t} - \mathbf{v}_{D,t}) \rangle}, \quad (1)$$

the spatially averaged absolute errors

$$E_{\text{abs}}(\mathbf{v}_{\text{ref},t}, \mathbf{v}_{D,t}) = \left\langle \sqrt{(\mathbf{v}_{\text{ref},t} - \mathbf{v}_{D,t}) \cdot (\mathbf{v}_{\text{ref},t} - \mathbf{v}_{D,t})} \right\rangle, \quad (2)$$

where $\langle \cdot \rangle$ is the spatial averaging operator in the horizontal plane, and the spatially averaged relative errors

$$E_{\text{rel}}(\mathbf{v}_{\text{ref},t}, \mathbf{v}_{D,t}) = \left\langle \sqrt{\frac{(\mathbf{v}_{\text{ref},t} - \mathbf{v}_{D,t}) \cdot (\mathbf{v}_{\text{ref},t} - \mathbf{v}_{D,t})}{\mathbf{v}_{\text{ref},t} \cdot \mathbf{v}_{\text{ref},t}}} \right\rangle, \quad (3)$$

between the height-dependent transverse velocity fields $[\mathbf{v}_{D,t}]$ inferred by DeepVel from simulation intensitygrams $[I_{\text{ref}}]$ and the plasma motions $[\mathbf{v}_{\text{ref},t}]$ computed by the STAGGER simulation, the MURaM-QS simulation or the MURaM-ARs simulation. The correlation coefficient $[C(\mathbf{v}_{\text{ref},t}, \mathbf{v}_{D,t})]$ from

¹ Slices of the STAGGER simulation at constant optical depth are available for download: <http://steinr.pa.msu.edu/~bob/96averages/>.

Schrijver et al. (2006) is used as a measure of similarity between maps of the velocity

$$C(\mathbf{v}_{\text{ref}}, \mathbf{v}_{D,t}) = \frac{\langle \mathbf{v}_{\text{ref},t} \cdot \mathbf{v}_{D,t} \rangle}{\sqrt{\langle \mathbf{v}_{\text{ref},t} \cdot \mathbf{v}_{\text{ref},t} \rangle \langle \mathbf{v}_{D,t} \cdot \mathbf{v}_{D,t} \rangle}}. \quad (4)$$

The global spatial distribution of the orientation of the velocity vectors are measured through the spatially-averaged normalized dot product

$$A(\mathbf{v}_{\text{ref},t}, \mathbf{v}_{D,t}) \equiv \left\langle \frac{\mathbf{v}_{\text{ref},t} \cdot \mathbf{v}_{D,t}}{\|\mathbf{v}_{\text{ref},t}\| \|\mathbf{v}_{D,t}\|} \right\rangle, \quad (5)$$

where $A = \pm 1$ for parallel and anti-parallel vectors and $A = 0$ for perpendicular vectors. To summarize, the ideal reconstruction of transverse velocity field satisfies: $C(\mathbf{v}_{\text{ref},t}, \mathbf{v}_{D,t}) = 1.0$ and $A(\mathbf{v}_{\text{ref},t}, \mathbf{v}_{D,t}) = 1.0$.

Knowledge of plasma velocity, $\mathbf{v} = (v_t, v_z)$, allows us to estimate the proxy for the vertical component of the magnetic energy Poynting-flux vector, $[S_z]$. Specifically, if we assume to have an ideal photospheric electric field, \mathbf{E} , then from observed magnetic field vector \mathbf{B} and derived velocity field vector \mathbf{v} we can find $c\mathbf{E} = -\mathbf{v} \times \mathbf{B}$. The Poynting-flux vector

$$\mathbf{S} = \frac{1}{4\pi} (\mathbf{E} \times \mathbf{B}), \quad (6)$$

measures the flow of electromagnetic energy where we estimate \mathbf{E} and \mathbf{B} . The vertical component of \mathbf{S} measures the amount of energy flowing into and out of the Sun and depends upon the transverse components of both the electric field and the magnetic field:

$$S_z = \frac{1}{4\pi} (E_x B_y - E_y B_x) = \frac{1}{4\pi c} ((\mathbf{B}_t \cdot \mathbf{B}_t) v_z - (\mathbf{v}_t \cdot \mathbf{B}_t) B_z). \quad (7)$$

Here, following e.g. Liu & Schuck (2012), Parnell & De Moortel (2012) and Welsch (2015), we conceptually divide the vertical Poynting flux into an ‘‘emergence’’ term with v_z , and a ‘‘shear’’ term with v_t . This distinction though, is not very precise, since both terms involve the emergence of magnetized plasma across the surface. If we ignore the ‘‘emergence’’ term, we could describe the shear component of the vertical Poynting flux S_z (Yeates et al., 2014; Welsch, 2015):

$$S_z \approx -\frac{1}{4\pi c} (\mathbf{v}_t \cdot \mathbf{B}_t) B_z, \quad (8)$$

where we set $\mathbf{B}_t = \mathbf{B}_{\text{ref},t}$, $B_z = B_{\text{ref},z}$ and $\mathbf{v}_t = \{\mathbf{v}_{\text{ref},t}, \mathbf{v}_{D,t}\}$.

To compare the quality of S_z reconstruction we derive the ratio between the integrals of the unsigned shear components of the Poynting fluxes over the full field-of-view M

$$S \equiv \frac{\int |S_z(\mathbf{v}_{D,t})| dM}{\int |S_z(\mathbf{v}_{\text{ref},t})| dM}. \quad (9)$$

3.2 Flows above the surface of the Quiet Sun

We first expand upon the analysis presented in Tremblay et al. (2018) which focused on surface flow inversions but featured a version of the DeepVel neural network that was trained with the STAGGER dataset to perform inversions at $\tau \approx \{1, 0.1, 0.01\}$ from intensitygrams. The vertical component of the plasma motions $v_{\text{ref},z}(\tau)$, as computed by the STAGGER simulation and downsampled to the SDO/HMI

resolution, is displayed as background in Figure 1 with downflows ($v_{\text{ref},z} < 0$) outlining the granulation pattern at $\tau \approx 1$ (Fig. 1a) and the reverse granulation pattern at $\tau \approx \{0.1, 0.01\}$ (Figs. 1d and 1g). The spatial distribution of plasma motions $\mathbf{v}_{D,t}(\tau \approx 1)$ inferred by DeepVel is consistent with $\mathbf{v}_{\text{ref},t}(\tau \approx 1)$ from the STAGGER test set (Table 2). Diverging velocity vectors are produced at the center of granules ($v_{\text{ref},z}(\tau \approx 1) > 0$) and converging vectors in the intergranular lanes ($v_{\text{ref},z}(\tau \approx 1) < 0$; Fig. 1b), achieving $A = 0.79$ for the inferred angles. This behavior is akin to but does not satisfy the law of conservation of mass which is encompassed in the training data. Further analysis of Figure 1b revealed that downflow regions where $v_{\text{ref},z}(\tau \approx 1) < 0$ are the largest sources of errors in the vector orientations (not shown). Downflows in the intergranular lanes result in spatially-confined, and thus not as well resolved, transverse flow structures that are more complex than the patterns produced by upflows in the center of granules and are thus harder to reproduce. Figure 1c suggests that there is a slight tendency to underestimate the inferred velocity amplitudes $\|\mathbf{v}_D(\tau \approx 1)\|$ despite the neural network’s ability to match fairly well their spatial distribution (see the coefficient of determination in Figure 1c and the correlation coefficient C in Table 2). This, in addition to the vector orientations, contributes to the underestimation of the total Poynting flux from $\mathbf{v}_{D,t}$ inside the field of view (Table 2). Similar conclusions are drawn for inversions higher in the photosphere ($\tau \approx 0.1$: Figs. 1e and 1f) and in the chromosphere ($\tau \approx 0.01$: Figs. 1h and 1i). A comparable level of quality is achieved by the convolutional neural network at those optical depths (see metrics in Table 2) because the structures in $\mathbf{v}_{\text{ref},t}(\tau \approx \{0.1, 0.01\}, t_i)$ are spatially correlated with the reversed granulation pattern in $I_c(\tau \approx \{0.1, 0.01\}, t_i)$ and, by extension, with $\mathbf{v}_{\text{ref},t}(\tau \approx 1, t_i)$ and the input data $I_c(\tau \approx 1, t_i)$. Table 1 quantifies the similarity between flow fields from the STAGGER simulation at the surface and at optical depths $\tau \approx \{0.1, 0.01\}$. Metrics C and A decrease with τ , but the velocity fields remain correlated enough with the surface data to allow for reliable inversions from the patterns in $I_c(\tau \approx 1)$ (Table 2).

We then increase the spatial and temporal resolution of the training data and perform inversions using the MURaM-QS dataset which features elaborate flow patterns. The subfield presented in Figure 2 is a closeup of the intersection between four granules at $\tau \approx \{1, 0.1, 0.01\}$, showcasing the complexity of the intergranular lanes and granule edges. This subfield is approximately 80 times smaller in area than the one featured in Figure 1, and each pixel is covering an area that is more than 500 times smaller. As the optical depth decreases and distance from the solar surface increases, the reference transverse velocity undergoes morphological changes which are more significant than that seen in Figure 1 and Table 1 at the same optical depths, most notably near granule edges. The scatterplot comparing the reference amplitude $\|\mathbf{v}_{\text{ref},t}(\tau \approx 0.1)\|$ (resp. $\|\mathbf{v}_{\text{ref},t}(\tau \approx 0.01)\|$) to $\|\mathbf{v}_{\text{ref},t}(\tau \approx 1)\|$ has a coefficient of determination of 0.82 (resp. 0.72) (not shown). The vertical component, displayed as background in Figure 2, undergoes similar changes. Yet, despite differences in patterns, spatial resolution, cadence and physics, the MURaM-QS-trained version of DeepVel achieves similar if not slightly improved metrics than the STAGGER-trained version in the photosphere and in the chromosphere (Table 2). The MURaM-QS training set featured fifteen times the number of examples in the STAGGER training set. Three times

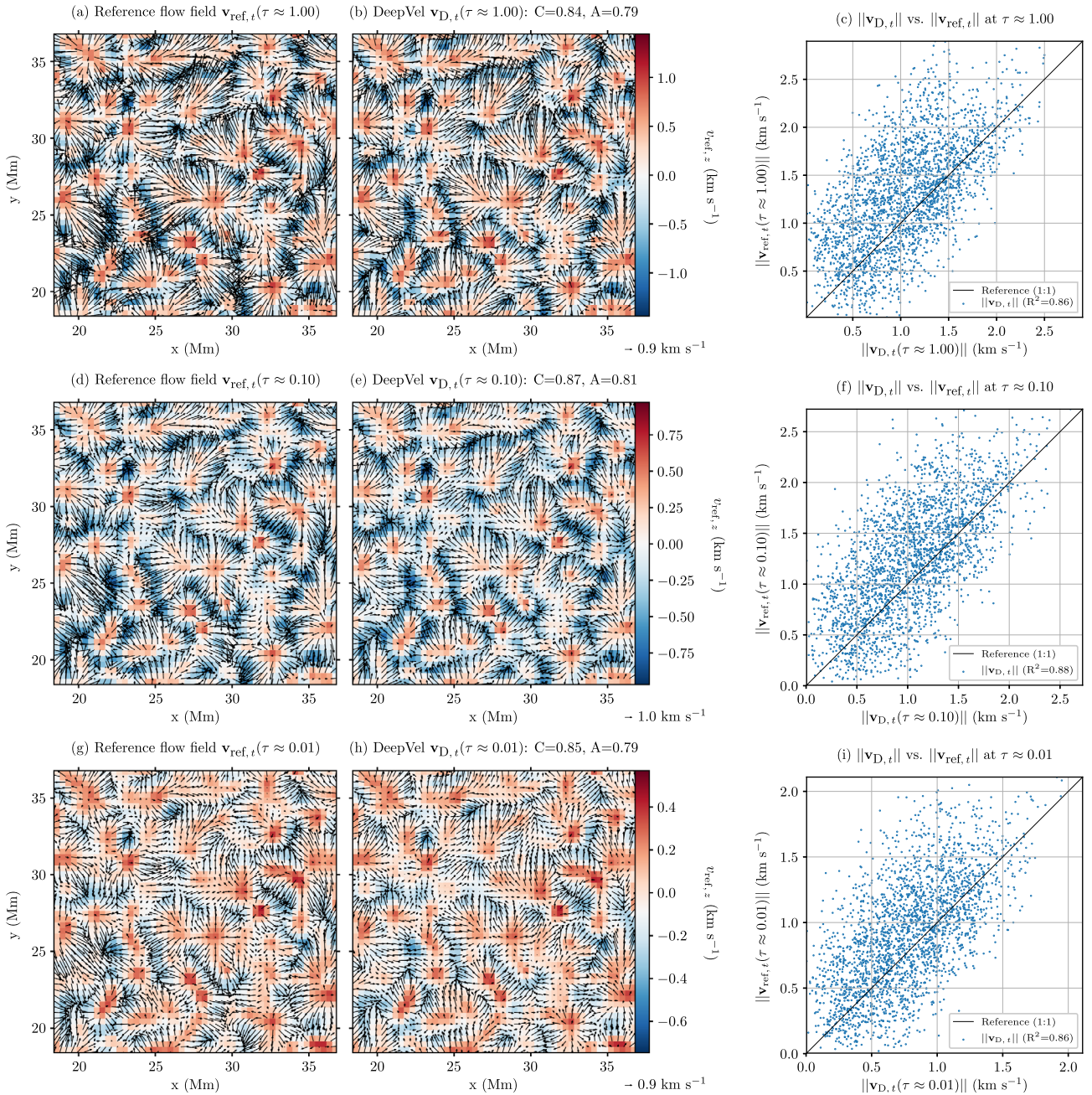


Fig. 1. Above the surface velocity inversions, STAGGER dataset. Patches of 50 by 50 pixels² extracted from the 30-min-averaged transverse velocity fields $\mathbf{v}_{D,t}(\tau, t_i)$ (left column) computed by the STAGGER simulation and (middle column) inferred by DeepVel at optical depths (top) $\tau \approx 1$, (middle) $\tau \approx 0.1$ and (bottom) $\tau \approx 0.01$. The time-averaged vertical velocity $v_{\text{ref},z}(\tau \approx \{1, 0.1, 0.01\}, t_i)$ computed by the STAGGER simulation and resampled to the SDO/HMI resolution is displayed as colored background. (Right column) Scatterplots comparing amplitudes $\|\mathbf{v}_{D,t}\|$ to $\|\mathbf{v}_{\text{ref},t}\|$. The black line represents an ideal solution (i.e., a coefficient of determination $R^2 = 1$).

as many training epochs were performed in anticipation of challenges related to the higher spatial resolution and cadence. Additionally, reference flows at $\tau \approx 0.1, 0.01$ appear to be smoother than at $\tau \approx 1$ (not shown). This results in poor metrics in Table 1 but also in flow patterns that, from the perspective of a neural network, are less complex than in $\mathbf{v}_{\text{ref},t}(\tau \approx 1)$ and still look similar to the surface granulation pattern. As for the STAGGER dataset, performances decrease in downflow regions.

3.3 Subsurface flows in the Quiet Sun

Experiments with multiple-depth inversions of flows reveal that spatial structures which are depth-specific (e.g., vortex-like structures in the upper convection zone; see Figs. 3d and 3g) can appear in the velocity inversions at other depths (e.g., the surface; not shown). These artifacts are the results of the minimization process and the network architecture which, for all intents and purposes, is simply a mapping function.

Table 1. Similarity between the reference velocity field at the surface $\mathbf{v}_{\text{ref},t}(\tau \approx 1)$ and at various optical and geometrical depth above and below the surface, τ and d , respectively. Metrics C and A correspond to the correlation coefficient (Eq. (4)) and the spatially-averaged normalized dot product A (Eq. (5)), respectively, averaged over time.

| Compared quantities\Metrics | STAGGER | | MURaM-QS | | MURaM-ARs | |
|---|---------|-------|----------|-------|-----------|-------|
| | C | A | C | A | C | A |
| $\mathbf{v}_{\text{ref},t}(\tau \approx 1), \mathbf{v}_{\text{ref},t}(\tau \approx 0.100)$ | 0.844 | 0.774 | 0.572 | 0.586 | 0.855 | 0.780 |
| $\mathbf{v}_{\text{ref},t}(\tau \approx 1), \mathbf{v}_{\text{ref},t}(\tau \approx 0.010)$ | 0.720 | 0.631 | 0.431 | 0.408 | 0.584 | 0.558 |
| $\mathbf{v}_{\text{ref},t}(\tau \approx 1), \mathbf{v}_{\text{ref},t}(\tau \approx 0.001)$ | – | – | – | – | 0.162 | 0.243 |
| $\mathbf{v}_{\text{ref},t}(\tau \approx 1), \mathbf{v}_{\text{ref},t}(d = 0 \text{ km})$ | – | – | 0.812 | 0.846 | – | – |
| $\mathbf{v}_{\text{ref},t}(\tau \approx 1), \mathbf{v}_{\text{ref},t}(d = 144 \text{ km})$ | – | – | 0.668 | 0.671 | – | – |
| $\mathbf{v}_{\text{ref},t}(\tau \approx 1), \mathbf{v}_{\text{ref},t}(d = 560 \text{ km})$ | – | – | 0.148 | 0.239 | – | – |
| $\mathbf{v}_{\text{ref},t}(\tau \approx 1), \mathbf{v}_{\text{ref},t}(d = 1000 \text{ km})$ | – | – | – | – | 0.492 | 0.438 |

Table 2. Comparison between flow fields $\mathbf{v}_{D,t}$ and $\mathbf{v}_{\text{ref},t}$ for the QS datasets (i.e., STAGGER and MURaM-QS). RMSE is the root mean squared error (Eq. (1)), E_{abs} is the mean absolute error (Eq. (2)), E_{rel} is the mean relative error (Eq. (3)), C is the correlation coefficient (Eq. (4)), A is the spatially-averaged normalized dot product (Eq. (5)) and S is the ratio between the integrals of the unsigned shear components of the DeepVel and reference Poynting fluxes (Eq. (9)).

| Training set (QS dataset) | Quantity (units) | Depth | RMSE | E_{abs} | E_{rel} (%) | $C \in [0, 1]$ | $A \in [-1, 1]$ | S (no units) |
|------------------------------------|--|----------------------|-------|------------------|----------------------|----------------|-----------------|----------------|
| STAGGER | \mathbf{v}_t (km s ⁻¹) | $\tau \approx 1.00$ | 0.777 | 0.694 | 70.6 | 0.841 | 0.786 | 0.811 |
| | | $\tau \approx 0.10$ | 0.697 | 0.620 | 65.4 | 0.867 | 0.810 | 0.793 |
| | | $\tau \approx 0.01$ | 0.553 | 0.491 | 70.3 | 0.852 | 0.791 | 0.729 |
| | $\nabla_t \mathbf{v}_t$ (10 ⁻³ s ⁻¹) | $\tau \approx 1.00$ | 0.735 | 0.577 | 195.7 | 0.912 | – | – |
| | | $\tau \approx 0.10$ | 0.783 | 0.618 | 261.3 | 0.892 | – | – |
| | | $\tau \approx 0.01$ | 0.931 | 0.733 | 199.9 | 0.884 | – | – |
| | $(\nabla \times \mathbf{v})_z$ (10 ⁻³ s ⁻¹) | $\tau \approx 1.00$ | 0.767 | 0.607 | 281.6 | 0.887 | – | – |
| | | $\tau \approx 0.10$ | 0.796 | 0.630 | 301.9 | 0.871 | – | – |
| | | $\tau \approx 0.01$ | 0.930 | 0.726 | 225.6 | 0.851 | – | – |
| MURaM-QS (three-depth version) | \mathbf{v}_t (km s ⁻¹) | $\tau \approx 1.00$ | 1.460 | 1.130 | 52.9 | 0.890 | 0.862 | – |
| | | $\tau \approx 0.10$ | 1.303 | 1.031 | 55.8 | 0.904 | 0.850 | – |
| | | $\tau \approx 0.01$ | 1.378 | 1.146 | 81.5 | 0.819 | 0.727 | – |
| | $\nabla_t \mathbf{v}_t$ (s ⁻¹) | $\tau \approx 1.00$ | 0.029 | 0.019 | 753.4 | 0.511 | – | – |
| | | $\tau \approx 0.10$ | 0.018 | 0.012 | 579.4 | 0.469 | – | – |
| | | $\tau \approx 0.01$ | 0.017 | 0.011 | 5419.6 | 0.262 | – | – |
| | $(\nabla \times \mathbf{v})_z$ (s ⁻¹) | $\tau \approx 1.00$ | 0.031 | 0.020 | 883.0 | 0.467 | – | – |
| | | $\tau \approx 0.10$ | 0.020 | 0.013 | 540.1 | 0.444 | – | – |
| | | $\tau \approx 0.01$ | 0.019 | 0.012 | 610.0 | 0.245 | – | – |
| MURaM-QS (single-depth version) | \mathbf{v}_t (km s ⁻¹) | $d = 0 \text{ km}$ | 1.386 | 1.109 | 54.2 | 0.907 | 0.872 | – |
| | | $d = 144 \text{ km}$ | 1.741 | 1.358 | 67.9 | 0.827 | 0.796 | – |
| | | $d = 560 \text{ km}$ | 2.110 | 1.677 | 117.0 | 0.415 | 0.428 | – |
| | $\nabla_t \mathbf{v}_t$ (s ⁻¹) | $d = 0 \text{ km}$ | 0.026 | 0.017 | 584.3 | 0.513 | – | – |
| | | $d = 144 \text{ km}$ | 0.036 | 0.024 | 653.6 | 0.411 | – | – |
| | | $d = 560 \text{ km}$ | 0.036 | 0.023 | 521.1 | 0.065 | – | – |
| | $(\nabla \times \mathbf{v})_z$ (s ⁻¹) | $d = 0 \text{ km}$ | 0.027 | 0.018 | 665.1 | 0.466 | – | – |
| | | $d = 144 \text{ km}$ | 0.039 | 0.025 | 617.8 | 0.379 | – | – |
| | | $d = 560 \text{ km}$ | 0.038 | 0.024 | 648.0 | 0.052 | – | – |

The neural network’s cost function is the sum of the mean squared errors at all three depths. For a given architecture, its weights and biases are not adjusted to provide the best solution at each specific depth but rather the best overall solution at an ensemble of depths. This can explain why DeepVel favored $\mathbf{v}_{D,t}(\tau \approx \{0.1, 0.01\}, t)$ over $\mathbf{v}_{D,t}(\tau \approx 1, t)$ in Figure 1 and Table 2 as it best optimized the cost function. As discussed in Section 2, the number of depths inferred simultaneously by DeepVel can be changed through its architecture. For example, the original three-depth architecture can be replaced by three single-depth-specific architectures, potentially improving the results at each

depth with the caveat being an increase in the number of neural networks to train and run. Fortunately, the training step only needs to be performed once and the execution of DeepVel takes less than a second. From this point forward, the flow inversions presented were computed using single-depth versions of DeepVel instead of the original three-depth version. Note that this approach could further improve the results from Figures 1 to 2.

Previous velocity field reconstructions were performed in a medium where plasma motions are ultimately dictated by the magnetic field. Transitioning to below the surface brings forth

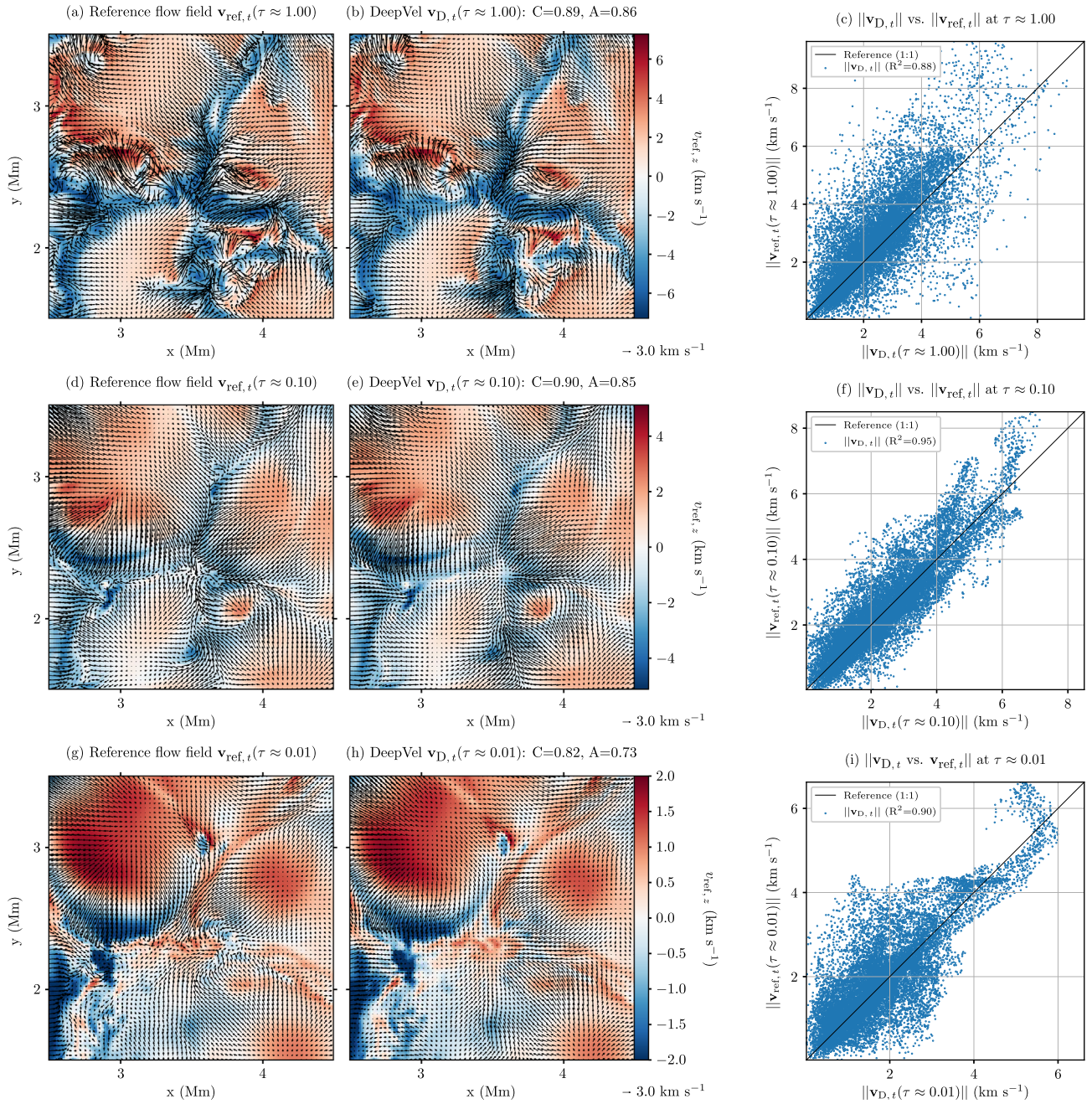


Fig. 2. Above the surface velocity inversions, MURaM-QS dataset. Patches of 125 by 125 pixels² extracted from the transverse velocity fields $v_{D,t}(\tau, t_i)$ (left column) computed by the MURaM-QS simulation (Vögler et al., 2005) and (middle column) inferred by DeepVel at optical depths (top) $\tau \approx 1$, (middle) $\tau \approx 0.1$ and (bottom) $\tau \approx 0.01$. The vertical velocity $v_{ref,z}(\tau \approx \{1, 0.1, 0.01\}, t_i)$ computed by the MURaM-QS simulation is displayed as colored background with a symmetrical threshold. (Right column) Scatterplots comparing amplitudes $\|v_{D,t}\|$ to $\|v_{ref,t}\|$. The black line represents the expected solution (i.e., a coefficient of determination $R^2 = 1$).

turbulent convection-driven plasma motions. In addition, inversions presented in this Section are performed from data at a constant optical depth $\tau \approx 1$ to probe flows at a constant geometrical depth d (i.e., radius), with $d = 0$ defined as the radius at which $\langle \tau \rangle \approx 1$ is achieved in the simulation. Figures 2a and 3a and Table 1 underline that noticeable differences arise between the velocity field associated with the DeepVel input data ($v_{ref,t}(\tau \approx 1)$) and $v_{ref,t}(d = 0)$ km.

The divergence of transverse velocity vectors is displayed as background in Figure 3, highlighting the complexity of transverse flow patterns. In particular, the high-resolution MURaM-QS simulation features vortex-like patterns between convective cells (Figs. 3d and 3g) which are more complex than that seen in the intergranular lanes above the solar surface (Fig. 2), i.e. where the neural network inversions have previously been the least accurate. Such patterns do not appear

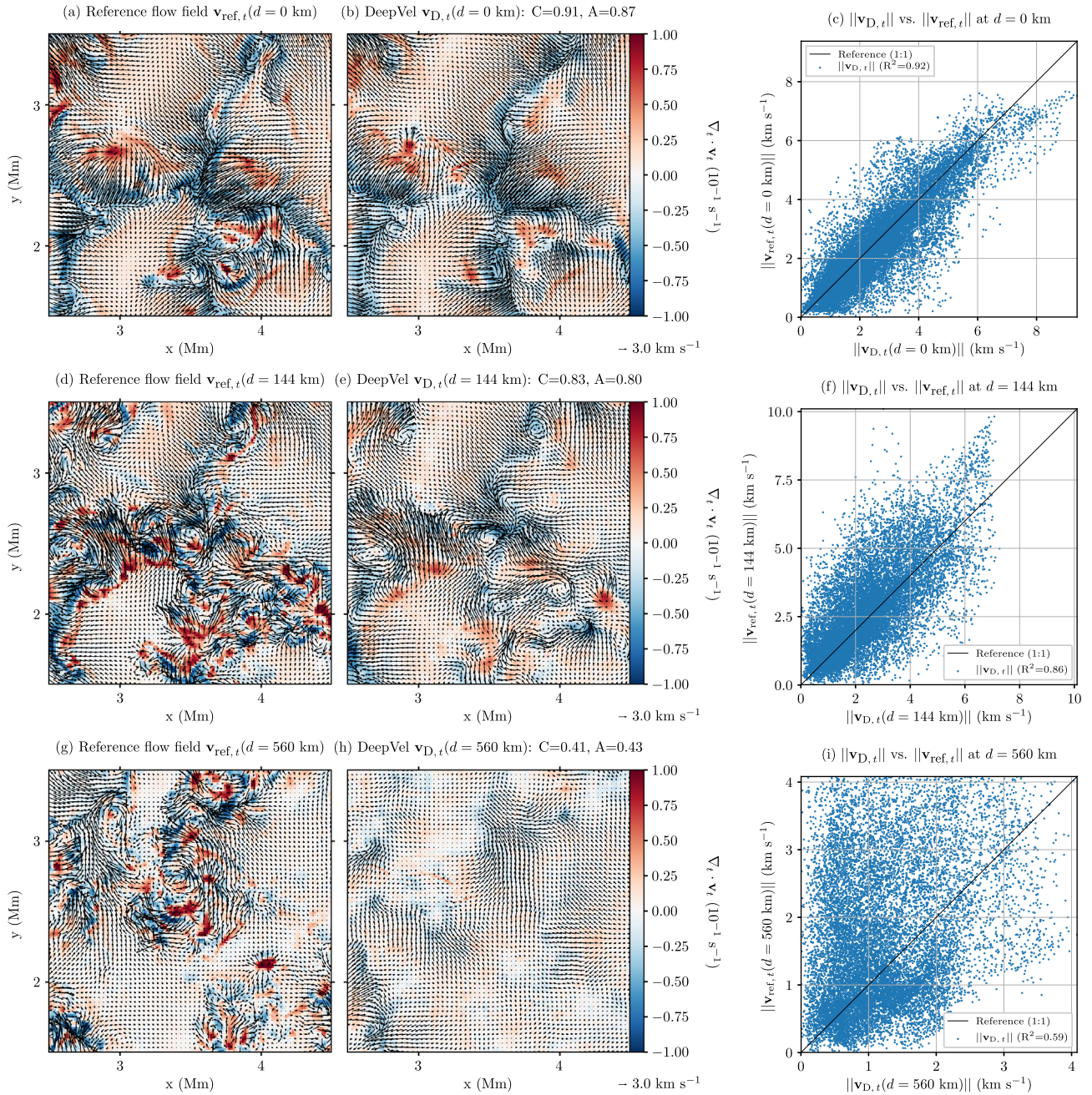


Fig. 3. Below the surface velocity inversions, MURaM-QS dataset. Patches of 125 by 125 pixels² extracted from instantaneous transverse velocity field maps which were computed at geometrical depths (top) $d = 0$ km, (middle) $d = 144$ km, (bottom) $d = 560$ km by (left column) the MURaM-QS simulation (Vögler et al., 2005) and (middle column) single-layer output versions of DeepVel. The divergence of the transverse plasma motions (i.e., $\nabla_t \cdot \mathbf{v}_t$) is displayed as colored background to emphasize the complexity of the structures that arise at increasing depths. (Right column) Scatterplots comparing amplitudes $\|\mathbf{v}_{D,t}\|$ to $\|\mathbf{v}_{ref,t}\|$ at each geometrical depth. The black line represents the expected solution (i.e., a coefficient of determination $R^2 = 1$).

in surface intensitygrams. Scatterplots comparing local velocity amplitudes from Figure 2a (i.e., the surface flows) to those of Figures 3a, 3d and 3g reveal that similarity with surface flows varies significantly with d as the coefficient of determination decreases from 0.81 to 0.72 and 0.59 for geometrical depths of 0 km, 144 km and 560 km respectively (scatterplots not shown). This is supported by Table 1. In this context, subsurface

flows prove to be challenging to capture at time t_i simply from maps of the continuum intensity $I_c(\tau \approx 1, t = \{t_i, t_{i+1}\})$, as demonstrated by the velocity vectors in Figures 3b, 3e and 3h and metrics in Table 2. Scatterplots 3c, 3f and 3i follow a positive trend but exhibit deviations from a linear relationship for larger velocity amplitudes, the latter being associated with the edges of convective cells. Metrics for $\mathbf{v}_{D,t}(d = 0$ km)

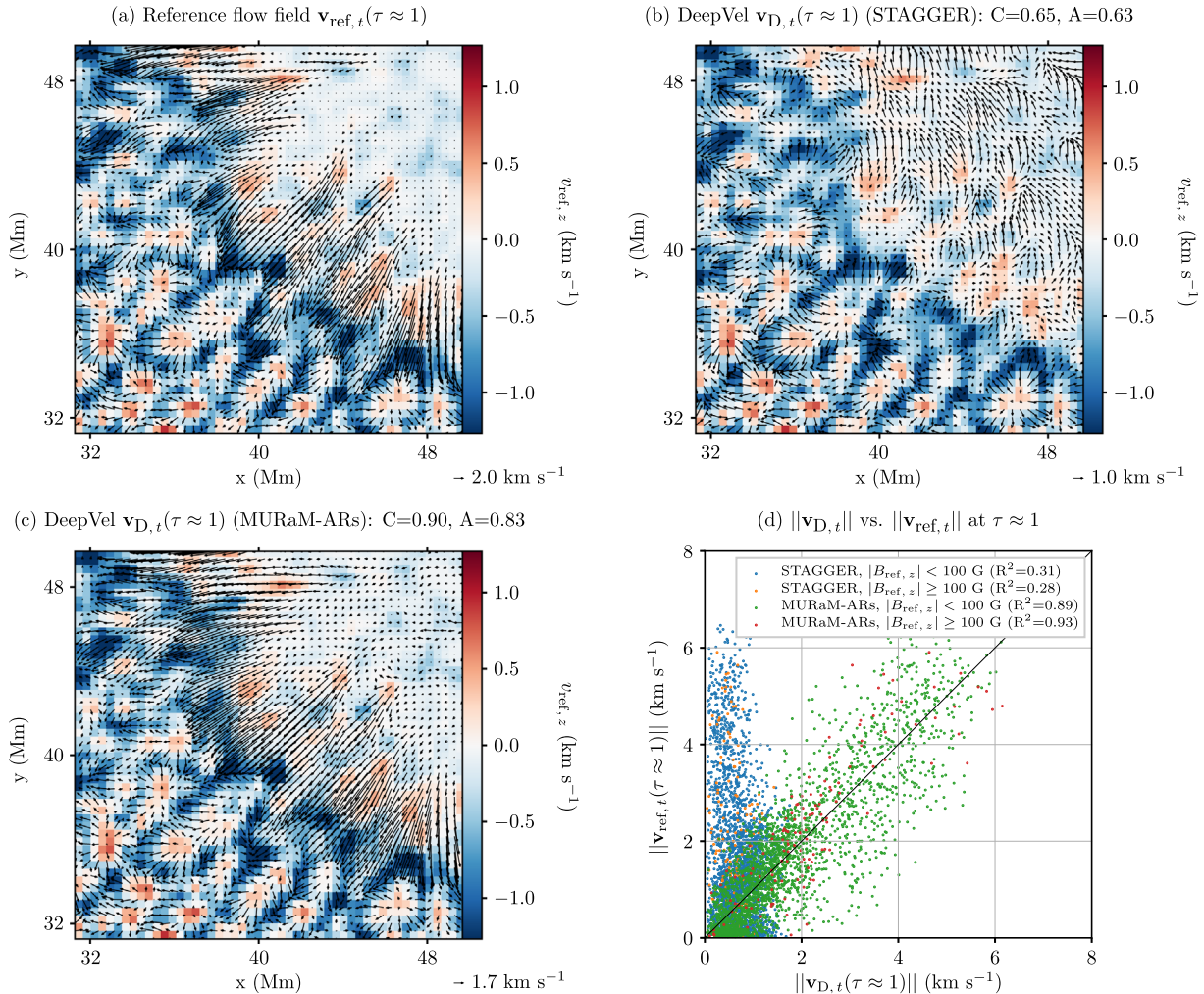


Fig. 4. At the surface velocity inversions, MURaM-ARs dataset. Patches of 50×50 pixels² extracted from 30-timestep-averaged transverse velocity field maps which were computed at optical depth $\tau \approx 1$ by (a) the MURaM-ARs simulation (Rempel & Cheung, 2014), (b) the DeepVel neural network that generated Figure 1a, and (c) a version of the DeepVel neural network which was trained using the MURaM-ARs simulation resampled at the SDO/HMI spatial resolution ($\Delta x \approx 368$ km). The vertical component of the velocity computed by the MURaM-ARs simulation is displayed as colored background. (Right column) Scatterplots comparing amplitudes $\|\mathbf{v}_{D,t}\|$ to $\|\mathbf{v}_{ref,t}\|$. The black line represents the expected solution (i.e., a coefficient of determination $R^2 = 1$).

(Figs. 3b and 3c) are on par with those achieved for $\mathbf{v}_{D,t}(\tau \approx 1)$ (see Table 2). As we increase to $d = 144$ km inside the convection zone, metrics C and A decrease to 0.83 and 0.80 respectively (Figs. 3e and 3f). The neural network underperforms in comparison to previous inversions at distances from the surface that are greater than 144 km above the surface (Fig. 2), but remain of a similar quality to what was achieved at $\tau \approx \{1, 0.1, 0.01\}$ with the STAGGER dataset in Figure 1 (see Table 2). Most notably, the complexity of inferred flow patterns between convective cells in Figure 3e is much less than that in the MURaM-QS simulation (Fig. 3d). At geometrical depth $d = 560$ km, reference and inferred flows become weakly correlated (Table 2). The velocity field is devoid of structure between adjacent convective cells, only faintly outlining their edges (see divergence in Fig. 2h and Table 2). Nonetheless, coefficients $C = 0.415$ and $A = 0.428$ (Table 2) and $R^2 = 0.59$ (Fig. 3i) improve upon the similarity between $\mathbf{v}_{ref,t}(d = 560$ km

and $\mathbf{v}_{ref,t}(\tau \approx 1)$ (Table 1). Vögler et al. (2005) reported that the MURaM-QS magnetic field lines transition from being spread-out at $d = 0$ to disrupted at $d \approx 300$ km per the turbulent convective motions inside the Sun. We expect this decorrelation with the surface to be a determining factor in DeepVel’s ability to map intensitygrams to subsurface flows. The neural network is capable of distinguishing between different convective cells, but the level of detail captured quickly decreases as d increases.

Inputs in DeepVel consist of two consecutive intensitygrams at $\tau \approx 1$. As distance from the surface increases, so does the time delay for the plasma at the surface to travel to geometrical depth d and vice versa. The native timestep of the MURaM-QS dataset ($\Delta t = 2$ s), also used by our implementation of DeepVel, is too short to account for such variations. Future experiments will be dedicated to varying the time gap between intensitygrams (i.e., the neural network inputs) as a function of the depth probed by DeepVel.

Table 3. Comparison between flow fields $\mathbf{v}_{D,t}$ and $\mathbf{v}_{\text{ref},t}$ for the ARs dataset. RMSE is the root mean squared error (Eq. (1)), E_{abs} is the mean absolute error (Eq. (2)), E_{rel} is the mean relative error (Eq. (3)), C is the correlation coefficient (Eq. (4)), A is the spatially-averaged normalized dot product (Eq. (5)) and S is the ratio between the integrals of the unsigned shear components of the DeepVel and reference Poynting fluxes (Eq. (9)).

| Training set (ARs dataset) | Quantity (units) | $ B_{\text{ref},z} $ (G) | RMSE | E_{abs} | E_{rel} (%) | $C \in [0, 1]$ | $A \in [-1, 1]$ | S (no units) |
|------------------------------------|--------------------------------|--------------------------|-------|------------------|----------------------|----------------|-----------------|----------------|
| STAGGER ($\tau \approx 1$) | \mathbf{v}_t (km s $^{-1}$) | ≥ 0 | 1.297 | 1.029 | 102.6 | 0.653 | 0.630 | 0.213 |
| | | < 100 | 1.264 | 1.006 | 78.44 | 0.692 | 0.676 | – |
| | | ≥ 100 | 1.604 | 1.267 | 353.2 | 0.166 | 0.155 | – |
| MURaM-ARs ($\tau \approx 1$) | \mathbf{v}_t (km s $^{-1}$) | ≥ 0 | 0.719 | 0.603 | 70.98 | 0.905 | 0.830 | 1.061 |
| | | < 100 | 0.701 | 0.592 | 56.02 | 0.912 | 0.855 | – |
| | | ≥ 100 | 0.878 | 0.717 | 226.2 | 0.845 | 0.563 | – |
| MURaM-ARs ($\tau \approx 0.1$) | \mathbf{v}_t (km s $^{-1}$) | ≥ 0 | 0.798 | 0.663 | 76.17 | 0.867 | 0.785 | 1.459 |
| | | < 100 | 0.793 | 0.662 | 61.51 | 0.878 | 0.824 | – |
| | | ≥ 100 | 0.840 | 0.678 | 217.9 | 0.639 | 0.410 | – |
| MURaM-ARs ($\tau \approx 0.01$) | \mathbf{v}_t (km s $^{-1}$) | ≥ 0 | 0.887 | 0.731 | 106.5 | 0.660 | 0.585 | 1.038 |
| | | < 100 | 0.880 | 0.728 | 91.03 | 0.680 | 0.619 | – |
| | | ≥ 100 | 0.956 | 0.759 | 253.9 | 0.415 | 0.261 | – |
| MURaM-ARs ($\tau \approx 0.001$) | \mathbf{v}_t (km s $^{-1}$) | ≥ 0 | 1.114 | 0.912 | 127.5 | 0.331 | 0.208 | 0.996 |
| | | < 100 | 1.104 | 0.913 | 118.6 | 0.281 | 0.199 | – |
| | | ≥ 100 | 1.214 | 0.893 | 225.9 | 0.580 | 0.306 | – |
| MURaM-ARs ($d = 1000$ km) | \mathbf{v}_t (km s $^{-1}$) | ≥ 0 | 0.961 | 0.851 | 152.9 | 0.258 | 0.164 | 1.343 |
| | | < 100 | 0.980 | 0.876 | 149.2 | 0.258 | 0.159 | – |
| | | ≥ 100 | 0.762 | 0.619 | 187.1 | 0.257 | 0.215 | – |

3.3.1 Flux emergences and sunspots

The MURaM-ARs dataset and the STAGGER dataset share the same spatial resolution. Hence MURaM-ARs intensity-grams can readily be used as input in the version of DeepVel used in Section 3.2, if the output velocities are multiplied by a factor of $\frac{45}{60}$ to account for the difference in cadence between the training and test sets. When doing so, the spatial distribution of the flows inferred in the QS are consistent with the reference flow (lower left corner of Figs. 4a and 4b, and Table 3 where $|B_{\text{ref},z}| < 100$ G in the ARs dataset), with divergent vectors in the center of granules and convergent vectors in the intergranular lanes. Note that the length of the velocity vectors in Figure 4b was re-normalized to match the length of the arrows in Figure 4a and focus on the spatial features. More specifically, for the same arrow length, the velocity amplitudes are twice as large in Figure 4a than in Figure 4b. The STAGGER simulation computes transverse velocity amplitudes of up to 3 km s $^{-1}$, whereas the MURaM-ARs simulation achieves velocities more than twice as large. For this reason, the STAGGER-trained version of DeepVel underestimates velocity amplitudes in the QS (see Fig. 4d for the scatterplot of the FOV covered by Fig. 4b). In the penumbra of the sunspot (i.e., the line joining the upper left and lower right corners of Figs. 4a and 4b), the velocities are significantly underestimated (Fig. 4d) and do not appear to be diverging away from the center of the sunspot. Inside the sunspot (i.e., the upper right corner of Figs. 4a and 4b), variations in the continuum intensity are mistakenly interpreted by the neural network as the edges of granular cells (Fig. 4b), and the vector orientations are poorly reconstructed (Table 3 where $|B_{\text{ref},z}| \geq 100$ G). This is justified by the neural network having been trained exclusively on examples of solar granulation. The generalization from the STAGGER-trained version of DeepVel to the MURaM-ARs simulation is thus limited to the QS spatial features.

Figure 4c showcases an improved time-averaged velocity field that was obtained by training DeepVel with the ARs dataset. The spatial distribution of plasma motions where $|B_{\text{ref},z}| \geq 100$ G is now much more consistent with the MURaM-ARs test set (Table 3), most notably in the penumbra. However, velocity amplitudes remain overestimated inside the sunspot (Fig. 4d and errors in Table 3). Inversions in the QS are of similar quality to the results of Section 3.2. To sum up, this version of the neural network successfully generalizes the flow patterns at $\tau \approx 1$ for ARs and granulation. Estimates of the total Poynting flux inside the field of view are also more accurate, with the STAGGER-trained version significantly underestimating its value (Table 3).

Displayed in Figure 5 are examples of velocity fields reconstructed by the DeepVel neural network at optical depths $\tau \approx 0.1, 0.01, 0.001$ above the surface. Metric values $C = 0.867, A = 0.785$ and $R^2 = 0.79$ for $\mathbf{v}_{D,t}(\tau \approx 0.1)$ (Table 3, Figs. 5b and 5c) are similar to those achieved in Section 3.2 at the same optical depth for the Quiet Sun (Table 2), thus extending the generalization potential of DeepVel for ARs from the surface to the base of the chromosphere. As for inversions at $\tau \approx 1$, loss of performance is traced back to regions where $|B_{\text{ref},z}| \geq 100$ G, with E_{rel} exceeding 200% and metrics C and A lowering to 0.639 and 0.410 respectively (Table 3). These errors contribute to the overestimation of the total Poynting flux within the field-of-view (Table 3). This further demonstrates the difficulty of capturing plasma motions inside magnetic structures such as sunspots simply from maps of the continuum intensity and advocates for the use of additional inputs in DeepVel (e.g., magnetograms and dopplergrams: Tremblay & Attie, 2020). Higher in the solar atmosphere of the MURaM-ARs simulation, flow patterns undergo significant morphological changes in weak field regions ($|B_{\text{ref},z}| < 100$ G) with respect to surface flows $\mathbf{v}_{\text{ref}}(\tau \approx 1)$, as quantified in Table 1. This is apparent in the lower portion of the subfield presented in

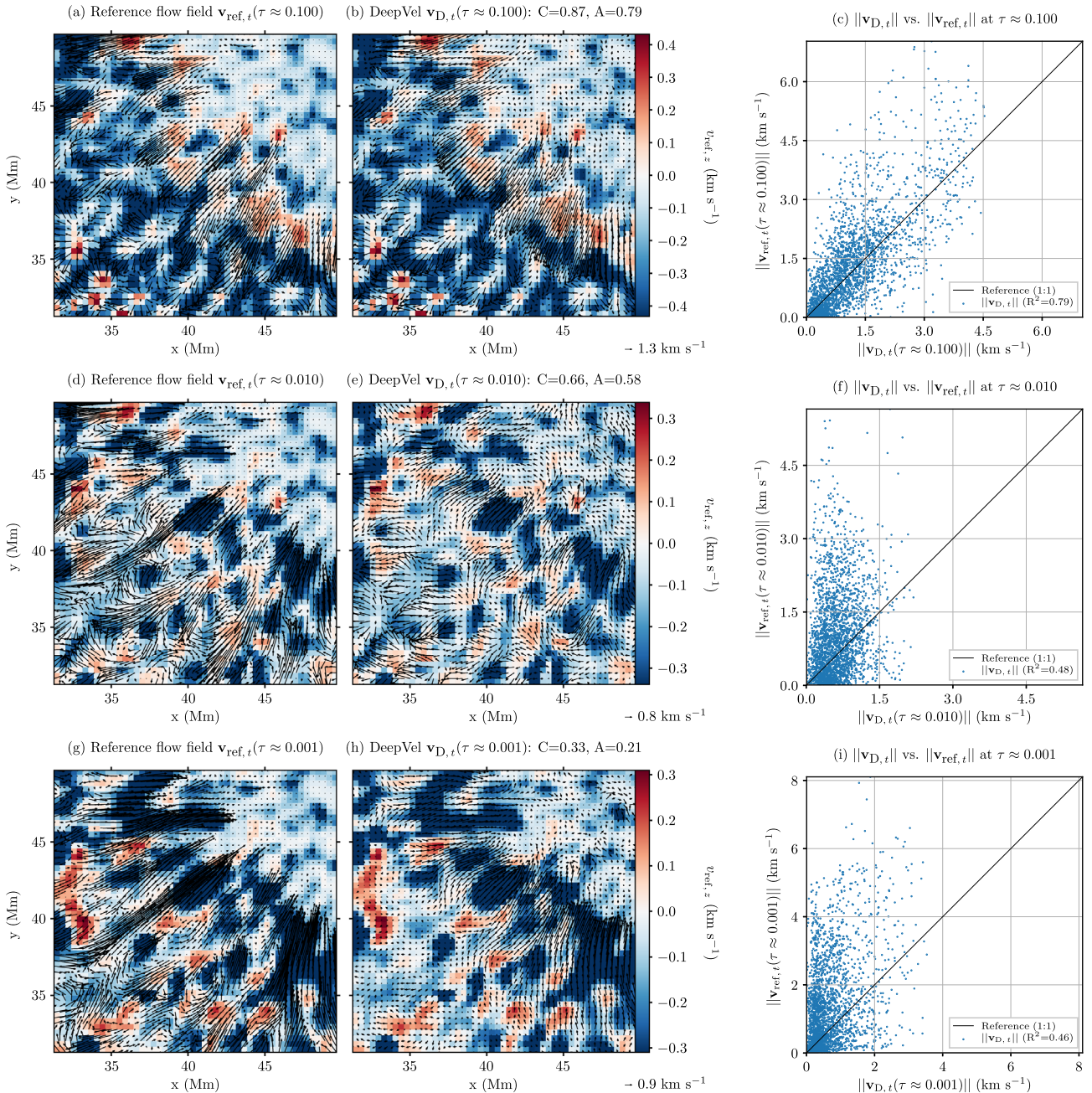


Fig. 5. Above the surface velocity inversions, MURaM-ARs dataset. (Left column) Patches of 50 by 50 pixels² extracted from the 30-min-averaged transverse velocity fields $\mathbf{v}_{D,t}(\tau, t_i)$ inferred by DeepVel at optical depths (top) $\tau \approx 0.1$, (middle) $\tau \approx 0.01$ and (bottom) $\tau \approx 0.001$. The time-averaged vertical velocity $v_{\text{ref},z}(\tau \approx \{0.1, 0.01, 0.001\}, t_i)$ computed by the MURaM-ARs simulation and resampled to the SDO/HMI resolution is displayed as colored background. (Right column) Scatterplots comparing amplitudes $\|\mathbf{v}_{D,t}\|$ to $\|\mathbf{v}_{\text{ref},t}\|$. The black line represents the expected solution (i.e., a coefficient of determination $R^2 = 1$).

Figures 5a, 5d and 5g, both in terms of the vertical velocity $v_{\text{ref},z}$ presented as background and the arrow length and orientation of the transverse velocity vectors $\mathbf{v}_{\text{ref},t}$. This loss of similarity translates to a loss of performance for $\mathbf{v}_{D,t}(\tau \approx \{0.01, 0.001\})$ in the same area (Table 3 and Figs. 5f and 5i). Training attempts at optical depths $\tau \approx \{0.01, 0.001\}$ have only slightly improved the value of C in comparison to the aforementioned similarity with $\mathbf{v}_{\text{ref}}(\tau \approx 1)$, whereas values for the normalized dot product and coefficient of determination are about the same

(see Tables 1 and 3). Scatterplots 5f and 5i show that larger velocities are significantly underestimated (see also arrow lengths in Figs. 5f and 5i). Despite the total Poynting flux being closer to its expected value than at any other optical depth, the remaining metrics suggest that this is most likely due to a fortuitous cancellation of errors.

Inversions at geometrical depth $d = 1000$ km in the upper convection zone (Fig. 6b), i.e. approximately 500 km deeper below the surface than previously probed with the MURaM-QS

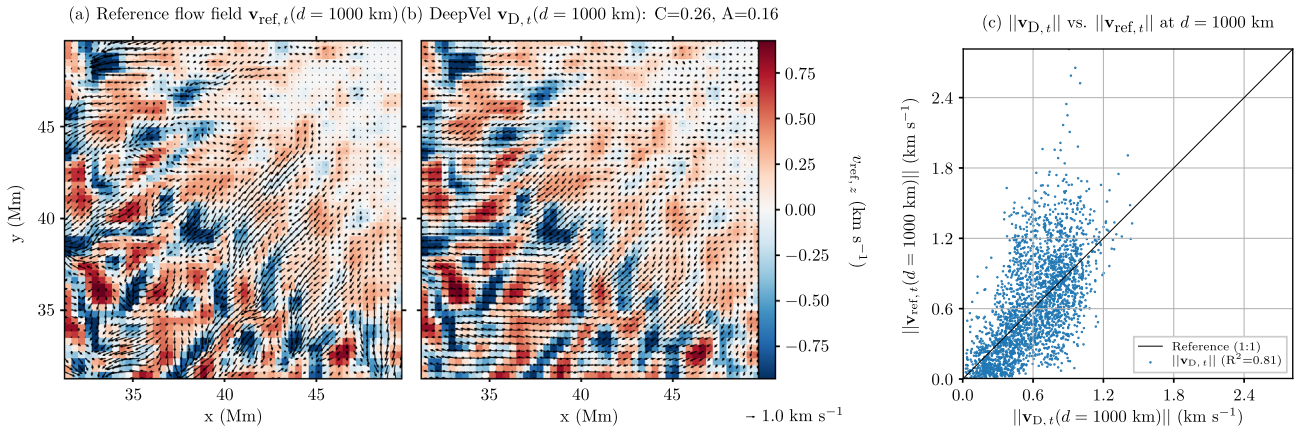


Fig. 6. Below the surface velocity inversions, MURaM-ARs dataset. (a) Patch of 50 by 50 pixels² extracted from the 30-min-averaged transverse velocity field $v_{D,t}(d, t_i)$ inferred by DeepVel at geometrical depth $d = 1000 \text{ km}$. The time-averaged vertical velocity $v_{ref,z}(d = 1000 \text{ km}, t_i)$ computed by the MURaM-ARs simulation and resampled to the SDO/HMI resolution is displayed as colored background. (Right column) Scatterplot comparing amplitudes $\|v_{D,t}\|$ to $\|v_{ref,t}\|$. The black line represents the expected solution (i.e., a coefficient of determination $R^2 = 1$).

dataset (Fig. 3h), fail to accurately capture the angle between horizontal components with $A = 0.164$ and, by extension, $C = 0.258$ (Eq. (4) and Table 3). Reconstructions are underperforming with respect to the similarity between $v_{ref,t}(\tau \approx 1)$ and $v_{ref,t}(d = 1000 \text{ km})$ ($C = 0.492$ and $A = 0.438$, Table 1 and Fig. 6a). Scatterplot 6c indicates a strong positive correlation between inferred and reference flow amplitudes with $R^2 = 0.83$, corroborating that the main source of errors at $d = 1000 \text{ km}$ are the vector orientations. As noticed with previous inversions, larger velocities are underestimated and there is a tendency to overestimate lower amplitudes. This is demonstrated by the arrow lengths in Figures 6a and 6b, most notably in the upper right corner and near the center of the subfield (i.e., where there is a concentration of magnetic flux). Training DeepVel on flow amplitudes and angles rather than the transverse components themselves may improve the results. This feature engineering approach will be tested in the future.

4 Conclusions & future work

We explored the use of deep learning tools and observed continuum intensity maps at the photosphere to derive horizontal plasma velocity vector below and above the photosphere. Specifically, using MURaM and STAGGER synthetic data, we trained the DeepVel neural network to infer horizontal plasma motions at various optical depths $\tau \approx \{0.001, 0.01, 0.1, 1\}$ above the surface and geometrical depths $d = \{0, 144, 560, 1000\} \text{ km}$ below the surface (i.e., in the upper convection zone). We compare the properties of the DeepVel reconstructions and the actual synthetic velocity fields $v_{D,t}$ and $v_{ref,t}$ using statistical metrics and measures of similarity such as the Schrijver et al. (2006) correlation coefficient C (expected value of 1; Eq. (4)) and normalized dot product A (expected value of 1; Eq. (5)). Our findings are as follows:

- **QS flow inversions at the surface and in the solar atmosphere:** Using the STAGGER and MURaM-QS

simulations, we find that the inversions at optical depths $\tau \approx \{0.1, 0.01\}$ in the QS (i.e., near the base of the chromosphere and near the transition region, respectively) are comparable in quality to the inversions at $\tau \approx 1$ presented in Tremblay et al. (2018), with $C(\tau \approx 0.1) = 0.867$, $A(\tau \approx 0.1) = 0.810$, $C(\tau \approx 0.01) = 0.852$ and $A(\tau \approx 0.01) = 0.791$ achieved for the SDO/HMI resolution STAGGER dataset and $C(\tau \approx 0.1) = 0.904$, $A(\tau \approx 0.1) = 0.850$, $C(\tau \approx 0.01) = 0.819$ and $A(\tau \approx 0.01) = 0.727$ achieved for the high-resolution MURaM-QS dataset. The inferred flow fields are thus deemed suitable approximations for use as synthetic observations of the transverse velocity vector that reflect the physics of the carefully-preprocessed model data used to train DeepVel, if the latter is a valid representation of real observations. In particular, the STAGGER-trained version of the neural network could be implemented in a data reduction pipeline due to its fast computations, and it can readily accept SDO/HMI intensitygrams as input if the output velocities at $\tau \approx \{1, 0.1, 0.01\}$ are multiplied by a factor of $\frac{45}{60}$ to account for the difference in cadence between SDO level-2 data (i.e., 45 s) and the STAGGER data (i.e., 60 s). We refer to Tremblay et al. (2018) for an example of such application. Similarly, the MURaM-QS version of DeepVel could be retrained to match the cadence of future DKIST data products and use the latter as inputs. Velocity amplitudes, which are typically underestimated by the method, can be adjusted by multiplying the vector field by a correction factor (i.e., a scalar) while still preserving its spatial properties. Additionally, training versions of the neural network to reconstruct single-depth velocity fields from surface intensitygrams will likely further minimize the remaining errors. This would increase the amount of time to train or execute trained neural networks by a factor equal to the number of depths to reconstruct. If implemented in a data reduction pipeline, the total execution time would still remain reasonably short. Future experiments will be dedicated to identifying an upper limit on the geometrical

or optical depth at which the DeepVel neural network can produce reliable results from surface observations of the QS.

- QS flow inversions below the surface:** Inferred subsurface flows fail to capture the complexity of the turbulent convective motions at increasing depths from intensitygrams of the surface. Inversions at $d = 144$ km below the surface achieve metrics comparable to what was obtained above the surface (i.e., $C = 0.827$ and $A = 0.796$) and could potentially be used as synthetic observations, though details between convective cells are missed by DeepVel. At $d = 560$ km, the level of detail diminishes significantly, with metric values $C = 0.415$ and $A = 0.428$ suggesting that these reconstructions may be used as first estimates of the velocity vector at best. This behavior is consistent with the decorrelation observed in the MURaM-QS simulation between magnetic field lines at the surface and at geometrical depths $d \geq 300$ km as convection becomes too disruptive (Vögler et al., 2005). By extension, the spatial correlation between surface intensitygrams and subsurface plasma motions decreases significantly with depth inside the convection zone, much more so than at increasing height above the surface (see Table 1). Additionally, the timestep used between two consecutive intensitygrams does not account for how variations at the surface affect the probed depth and vice versa. Increasing the timestep Δt and/or the number of intensitygrams provided as input may improve the results. Furthermore, processing sequences of images with a recurrent neural network architecture would build a memory of how past information impacts the output (Chollet, 2017).
- ARs flow inversions at the surface and in the solar atmosphere:** Using the MURaM-ARs simulation, we find that inversions inside and around ARs at $\tau \approx \{1, 0.1\}$ (i.e., in the photosphere and near the base of the chromosphere) are comparable to our QS inversions at the same optical depths (i.e., $C(\tau \approx 1) = 0.905$, $A(\tau \approx 1) = 0.830$, $C(\tau \approx 0.1) = 0.867$ and $A(\tau \approx 0.1) = 0.785$), although the inside of the sunspot is an area in need of further improvement. As is the case for the STAGGER-trained version of DeepVel, the MURaM-ARs-trained version can readily use SDO/HMI intensitygrams as input to generate synthetic transverse velocity estimates at $\tau \approx \{1, 0.1\}$. Note that due to differences in the training data, outputs generated by the two neural networks will differ in the QS, most notably in terms of velocity amplitudes. As we approach higher regions in the atmosphere, the velocity fields in the solar atmosphere of the MURaM-ARs simulation become less correlated with the surface flows and inversions from surface intensitygrams become less reliable, with metrics decreasing to $C(\tau \approx 0.01) = 0.660$, $A(\tau \approx 0.01) = 0.585$, $C(\tau \approx 0.001) = 0.331$ and $A(\tau \approx 0.001) = 0.208$.
- ARs flow inversions below the surface:** Inferred subsurface velocity fields at $d = 1000$ km follow the trend set by QS flow inversions, with performance decreasing significantly as a function of depth. Metrics $C = 0.258$ and $A = 0.164$ underperform in comparison to the simple similarity between flow patterns at $\tau \approx 1$ and at

$d = 1000$ km, with velocity vector angles being the main source of errors. Simulation data at additional geometrical depths between 0 and 1000 km below the surface will be required to estimate up to which depth the neural network can reliably probe subsurface magnetic structures and flows from surface observations of ARs.

We conclude that the DeepVel neural network is capable of extrapolating reliable flow field approximations in the photosphere, the base of the chromosphere and upper convection zone simply from observed intensitygrams. The level of detail and, by extension, errors in reconstructions vary as a function of distance from the surface as flow patterns become decorrelated from surface intensitygrams. Based on our results and the simulation data, this trend appears to be more prominent in ARs than in the QS. Nonetheless, we believe that DeepVel synthetic data is ideal for (1) inferring depth-dependent and data-driven estimates of transverse-velocity-dependent physical quantities, (2) generating boundary conditions for data-driven simulations of the solar atmosphere, and (3) generating internal conditions for data assimilation in MHD models by providing additional spatial coverage of the transverse velocity vector. We provided examples of depth-dependent Poynting flux estimates that were derived from DeepVel transverse velocity maps and simulation magnetograms at the same depths. Such estimates are to be validated in the chromosphere with the advent of DKIST magnetograms and Dopplergrams. Data assimilation processes minimize errors between model predictions and observations within a given time window. We are now able to provide synthetic observations of the transverse velocity vector that may be included in the minimization process to guide more effectively the model towards a solution representative of a given epoch of the Sun. As suggested in Tremblay et al. (2018), the model-dependency of DeepVel flow fields may work as an advantage if the training set is emulating the same model in which the velocity vector is to be assimilated. This would also simplify the observation operator which projects model velocity vector into the observations space. We currently limited inversions to select depths to establish a range at which DeepVel can operate, but one could train versions of the neural network to generate a cube of transverse velocities at all depths in between. Errors as a function of depth are to be quantified and accounted for in the covariance matrix of observation errors (Bouttier & Courtier, 2002). An improved nowcasting step in data assimilation allows to more accurately invert other model state variables such as the magnetic field components which can be studied over the duration of assimilation window, providing insight on the QS and ARs and their influence on solar activity based on the physics encompassed by the model. This also translates in improved forecasts beyond the assimilation window of the evolution of the QS, ARs and, if the simulation allows for it, the occurrence of space weather events (e.g., flares in MURaM: Cheung et al., 2018). Additionally, an estimate of the velocity vector is required at the beginning of the assimilation window in order to initiate the data assimilation process. Hence, despite errors, DeepVel reconstructions may be used to that effect to assist in the convergence of the data assimilation method. Taking the longer view, rapid and automated reconstruction of subsurface flows from surface observations also has potential interest beyond data assimilation

towards short-term activity forecasting. For example, modelling of global surface magnetic flux transport has revealed the importance of accurately capturing the physical characteristics of emerging bipolar active regions in order to properly model their contribution to the global solar dipole (Jiang et al., 2015; Hathaway & Upton, 2016, also Petrovay et al., this volume). This is essential to dynamo model-based the prediction of overall activity levels on decadal timescales (see, e.g., Labonville et al., 2019), and possibly to understand the onset of extended “Grand Minima” epochs of reduced activity (Nagy et al., 2017). Detecting active regions through internal flow perturbations they induce prior to their emergence at the photosphere may lead to improved determination of their ultimate dynamo efficiency. Current tests have revealed that when the DeepVel neural network is trained with the computations of a simulation of an sunspot, it is capable of distinguishing between the different depth-dependent flows in the QS and ARs with varying level of detail. We speculate that if the future focus is shifted from the missing detailed spatial features to the successful detection of spatial structures such as the edges and center of granules and sunspots, this approach may be adapted for the prediction of the emergence of magnetic flux or of a sunspot prior to its occurrence at the surface. Instead of velocity maps, the outputs could be a measure of a spatially-averaged physical quantity or a simple binary signal for detection and non-detection. This is left as future work.

As we probed the upper convection zone, photosphere and chromosphere, the DeepVel neural network was subjected to a wide range of flow patterns and spatial scales (i.e., turbulence, intergranular lanes, granulation, mesogranulation, supergranulation). We were motivated by those structures to modify the DeepVel neural network architecture to that of a U-net (Ronneberger et al., 2015). This architecture allows to probe the dominant features not only at small scales (i.e., the pixel size) but also scales up the size of the sub-images presented during training (DeepVelU: Tremblay & Attie, 2020; Tremblay et al., 2019). Our recent efforts have also focused on modifying the inputs of the neural network in order to accept a combination of consecutive intensitygrams (i.e., the default inputs), magnetograms and Dopplergrams (Tremblay & Attie, 2020). This is motivated by the magnetic induction equation which relates transverse plasma motions to the magnetic field and line-of-sight plasma motions. Intensity-based methods are effective in inferring plasma motions in the Quiet Sun, however reconstructions in the Active Sun typically rely on magnetograms (e.g., Longcope, 2004; Schuck, 2005, 2006, 2008) or magnetograms and Dopplergrams (Kazachenko et al., 2014; Lumme et al., 2019; Fisher et al., 2020) to solve the magnetic induction equation. We expect that providing additional information to DeepVel regarding the physics will further improve the metrics in active regions and thus in Figure 4c and Table 3. DeepVel’s architecture and cost function could also be modified to ensure that inferred solutions are consistent with the magnetic induction equation using the framework of physics-informed neural networks (PINNs: Raissi et al., 2019) which have recently been used to solve for the velocity vector in Navier-Stokes equations in the context of fluid mechanics (e.g., NSFnets: Jin et al., 2020). All of the above may help us improve flow depth-dependent inversions. This will be explored in future work.

Finally, a similar method could be invoked to infer other physical quantities of interest that cannot yet be measured directly at the photosphere or anywhere else in the solar atmosphere, e.g. the electric field and the Poynting flux.

DeepVel codes, weights and biases used in this paper can be found in the following Github repository: https://github.com/tremblaybenoit/DeepVel_DeepVelU.

Acknowledgements. The authors would like to thank the anonymous referees for their constructive comments. The authors would also like to thank Dr. Matthias Rempel (NCAR) for providing the MURaM simulation data. The editor thanks Michael Kirk and an anonymous reviewer for their assistance in evaluating this paper.

References

- Abbett WP. 2007. The magnetic connection between the convection zone and corona in the Quiet Sun. *Astrophys J* **665**: 1469–1488. <https://doi.org/10.1086/519788>.
- Abbett WP, Fisher GH. 2010. Improving large-scale convection-zone-to-corona models. *Mem Soc Astron Italiana* **81**: 721.
- Asensio Ramos A, Requerey IS, Vitas N. 2017. DeepVel: Deep learning for the estimation of horizontal velocities at the solar surface. *A&A* **604**: A11. <https://doi.org/10.1051/0004-6361/201730783>.
- Attie R, Innes DE. 2015. Magnetic balltracking: Tracking the photospheric magnetic flux. *A&A* **574**: A106. <https://doi.org/10.1051/0004-6361/201424552>.
- Bouttier F, Courtier P. 2002. *Data assimilation concepts and methods*. <https://www.ecmwf.int/node/16928>.
- Cheung MCM, Rempel M, Chintzoglou G, Chen F, Testa P, et al. 2018. A comprehensive three-dimensional radiative magnetohydrodynamic simulation of a solar flare. *Nat Astron* **3**: 160–166. <https://doi.org/10.1038/s41550-018-0629-3>.
- Cheung MCM, Schüssler M, Moreno-Insartit F. 2007. The origin of the reversed granulation in the solar photosphere. *A&A* **461**: 1163–1171. <https://doi.org/10.1051/0004-6361:20066390>.
- Chollet F. 2017. *Deep learning with Python, chap. 1, 2, 5, 7, 1–50, 19–143, 260–263*, 1st edn, Manning Publications Co., Greenwich, CT, USA. ISBN 1617294438, 9781617294433.
- Cranmer SR, van Ballegoijen AA, Edgar RJ. 2007. Self-consistent coronal heating and solar wind acceleration from anisotropic magnetohydrodynamic turbulence. *Astrophys J Suppl Ser* **171(2)**: 520–551. <https://doi.org/10.1086/518001>.
- Fisher GH, Welsch BT. 2008. FLCT: A fast, Efficient method for performing local correlation tracking. In *Subsurface and atmospheric influences on solar activity*, Howe R, Komm RW, Balasubramaniam KS, Petrie GJD (Eds.), Vol. **383** of Astronomical Society of the Pacific Conference Series, Astronomical Society of the Pacific, San Francisco, CA, USA, 373 p.
- Fisher GH, Abbett WP, Bercik DJ, Kazachenko MD, Lynch BJ, et al. 2015. The Coronal Global Evolutionary Model: Using HMI vector magnetogram and doppler data to model the buildup of free magnetic energy in the solar corona. *Space Weather* **13**: 369. <https://doi.org/10.1002/2015SW001191>.
- Fisher GH, Kazachenko MD, Welsch BT, Sun X, Lumme E, Bercik DJ, DeRosa ML, Cheung MCM. 2020. The PDFI_SS electric field inversion software. *Astrophys J Suppl Ser* **248(1)**: 2. <https://doi.org/10.3847/1538-4365/ab8303>.

- Hathaway DH, Upton LA. 2016. Predicting the amplitude and hemispheric asymmetry of solar cycle 25 with surface flux transport. *J Geophys Res(Space Phys)* **121(11)**: 10744–10753. <https://doi.org/10.1002/2016JA023190>.
- Hayashi K, Feng X, Xiong M, Jiang C. 2018. An MHD simulation of solar active region 11158 driven with a time-dependent electric field determined from HMI vector magnetic field measurement data. *Astrophys J* **855**: 11. <https://doi.org/10.3847/1538-4357/aaacd8>.
- Hoeksema JT, Liu Y, Hayashi K, Sun X, Schou J, et al. 2014. The helioseismic and magnetic imager (HMI) vector magnetic field pipeline: Overview and performance. *Sol Phys* **289**: 3483–3530. <https://doi.org/10.1007/s11207-014-0516-8>.
- Jiang C, Wu ST, Feng X, Hu Q. 2016. Data-driven magnetohydrodynamic modelling of a flux-emerging active region leading to solar eruption. *Nat Commun* **7**: 11522. <https://doi.org/10.1038/ncomms11522>.
- Jiang J, Cameron RH, Schüssler M. 2015. The cause of the weak solar cycle 24. *Astrophys J Lett* **808(1)**: L28. <https://doi.org/10.1088/2041-8205/808/1/L28>.
- Jin X, Cai S, Li H, Karniadakis GE. 2020. NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations. *J Comput Phys*, 109951. ISSN 0021-9991. <https://doi.org/10.1016/j.jcp.2020.109951>. <http://www.sciencedirect.com/science/article/pii/S0021999120307257>.
- Kazachenko MD, Fisher GH, Welsch BT. 2014. A comprehensive method of estimating electric fields from vector magnetic field and doppler measurements. *Astrophys J* **795**: 17. <https://doi.org/10.1088/0004-637X/795/1/17>.
- Labonville F, Charbonneau P, Lemerle A. 2019. A dynamo-based forecast of solar cycle 25. *Sol Phys* **294(6)**: 82. <https://doi.org/10.1007/s11207-019-1480-0>.
- Liu Y, Schuck PW. 2012. Magnetic energy and helicity in two emerging active regions in the Sun. *Astrophys J* **761(2)**: 105. <https://doi.org/10.1088/0004-637X/761/2/105>.
- Longcope DW. 2004. Inferring a photospheric velocity field from a sequence of vector magnetograms: The minimum energy fit. *Astrophys J* **612**: 1181–1192. <https://doi.org/10.1086/422579>.
- Lumme E, Kazachenko MD, Fisher GH, Welsch BT, Pomoell J, Kilpua EKJ. 2019. Probing the effect of cadence on the estimates of photospheric energy and helicity injections in eruptive active region NOAA AR 11158. *Sol Phys* **294(6)**: 84. <https://doi.org/10.1007/s11207-019-1475-x>.
- Nagy M, Lemerle A, Labonville F, Petrovay K, Charbonneau P. 2017. The effect of “Rogue” active regions on the solar cycle. *Sol Phys* **292(11)**: 167. <https://doi.org/10.1007/s11207-017-1194-0>.
- November LJ, Simon GW. 1988. Precise proper-motion measurement of solar granulation. *Astrophys J* **333**: 427–442. <https://doi.org/10.1086/166758>.
- Parker EN. 1988. Nanoflares and the solar X-ray corona. *Astrophys J* **330**: 474. <https://doi.org/10.1086/166485>.
- Parnell CE, De Moortel I. 2012. A contemporary view of coronal heating. *Roy Soc Lond Philos Trans Ser A* **370**: 3217–3240. <https://doi.org/10.1098/rsta.2012.0113>.
- Potts HE, Barrett RK, Diver DA. 2004. Balltracking: An highly efficient method for tracking flow fields. *A&A* **424**: 253–262. <https://doi.org/10.1051/0004-6361:20035891>.
- Raissi M, Perdikaris P, Karniadakis GE. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* **378**: 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>.
- Rempel M, Cheung MCM. 2014. Numerical simulations of active region scale flux emergence: From spot formation to decay. *Astrophys J* **785**: 90. <https://doi.org/10.1088/0004-637X/785/2/90>.
- Rieutord M, Roudier T, Roques S, Ducottet C. 2007. Tracking granules on the Sun’s surface and reconstructing velocity fields. I. The CST algorithm. *A&A* **471**: 687–694. <https://doi.org/10.1051/0004-6361:20066491>.
- Rimmele TR, Warner M, Keil SL, Goode PR, Knölker M, et al. 2020. The Daniel K. Inoué solar telescope – observatory overview. *Sol Phys* **295(12)**: 172. <https://doi.org/10.1007/s11207-020-01736-7>. <https://ui.adsabs.harvard.edu/abs/2020SoPh..295..172R>. Provided by the SAO/NASA Astrophysics Data System.
- Ronneberger O, Fischer P, Brox T. 2015. U-Net: Convolutional networks for biomedical image segmentation. In: *International conference on Medical image computing and computer-assisted intervention*, Springer, Cham, pp. 234–241.
- Schou J, Scherrer PH, Bush RI, Wachter R, Couvidat S, et al. 2012. Design and ground calibration of the helioseismic and magnetic imager (HMI) instrument on the solar dynamics observatory (SDO). *Sol Phys* **275**: 229–259. <https://doi.org/10.1007/s11207-011-9842-2>.
- Schrijver CJ, De Rosa ML, Metcalf TR, Liu Y, McTiernan J, Régnier S, Valori G, Wheatland MS, Wiegmann T. 2006. Nonlinear force-free modeling of coronal magnetic fields part I: A quantitative comparison of methods. *Sol Phys* **235**: 161–190. <https://doi.org/10.1007/s11207-006-0068-7>.
- Schuck PW. 2005. Local correlation tracking and the magnetic induction equation. *Astrophys J Lett* **632**: L53–L56. <https://doi.org/10.1086/497633>.
- Schuck PW. 2006. Tracking magnetic footpoints with the magnetic induction equation. *Astrophys J* **646**: 1358–1391. <https://doi.org/10.1086/505015>.
- Schuck PW. 2008. Tracking Vector magnetograms with the magnetic induction equation. *Astrophys J* **683**: 1134–1152. <https://doi.org/10.1086/589434>.
- Stein RF. 2012. Solar surface magneto-convection. *Living Rev Sol Phys* **9**: 4. <https://doi.org/10.12942/lrsp-2012-4>.
- Stein RF, Nordlund Å. 1998. Simulations of Solar Granulation. I. General Properties. *Astrophys J* **499**: 914–933. <https://doi.org/10.1086/305678>.
- Stein RF, Nordlund Å. 2012. On the formation of active regions. *Astrophys J Lett* **753**: L13. <https://doi.org/10.1088/2041-8205/753/1/L13>.
- Tremblay B, Attie R. 2020. Inferring plasma flows at granular and supergranular scales with a new architecture for the DeepVel neural network. *Front Astron Space Sci* **7**: 25. <https://doi.org/10.3389/fspas.2020.00025>.
- Tremblay B, Roudier T, Rieutord M, Vincent A. 2018. Reconstruction of horizontal plasma motions at the photosphere from intensitygrams: A comparison between DeepVel, LCT, FLCT, and CST. *Sol Phys* **293**: 57. <https://doi.org/10.1007/s11207-018-1276-7>.
- Tremblay B, Roudier T, Cossette J-F, Attié R, Rieutord M, Vincent A. 2019. Neural network to emulate numerical simulations of the Sun and infer synthetic observations for data assimilation. In: *Solar Heliospheric and Interplanetary Environment (SHINE 2019)*, 30 p. <https://shinecon.org/Publications.php>.
- Vögler A, Shelyag S, Schüssler M, Cattaneo F, Emonet T, Linde T. 2005. Simulations of magneto-convection in the solar photosphere. Equations, methods, and results of the MURaM code. *A&A* **429**: 335–351. <https://doi.org/10.1051/0004-6361:20041507>.

- Wachter R, Schou J, Rabello-Soares MC, Miles JW, Duvall TL, Bush RI. 2012. Image quality of the helioseismic and magnetic imager (HMI) onboard the solar dynamics observatory (SDO). *Sol Phys* **275(1–2)**: 261–284. <https://doi.org/10.1007/s11207-011-9709-6>.
- Warner M, Rimmele TR, Martinez Pillet V, Casini R, Berukoff S, et al. 2018. Construction update of the Daniel K. Inouye Solar Telescope project. In: *Ground-based and Airborne Telescopes VII*, Vol. **10700** of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, SPIE, Bellingham, WA, USA, 107000V p. <https://doi.org/10.1117/12.2314212>.
- Welsch BT. 2006. Magnetic flux cancellation and coronal magnetic energy. *Astrophys J* **638(2)**: 1101–1109. <https://doi.org/10.1086/498638>.
- Welsch BT. 2015. The photospheric Poynting flux and coronal heating. *Publ Astron Soc Jpn* **67(2)**: 18. <https://doi.org/10.1093/pasj/psu151>.
- Yeates AR, Bianchi F, Welsch BT, Bushby PJ. 2014. The coronal energy input from magnetic braiding. *A&A* **564**: A131. <https://doi.org/10.1051/0004-6361/201323276>.

Cite this article as: Tremblay B, Cossette J-F, Kazachenko MD, Charbonneau P & Vincent A. 2021. Inferring depth-dependent plasma motions from surface observations using the DeepVel neural network. *J. Space Weather Space Clim.* **11**, 9. <https://doi.org/10.1051/swsc/2020073>.