Stokes Inversion Techniques with Neural Networks: Analysis of Uncertainty in Parameter Estimation

Lukia Mistryukova\textsuperscript{1} · Andrey Plotnikov\textsuperscript{2,3} · Aleksandr Khizhik\textsuperscript{1} · Irina Knyazeva\textsuperscript{2,4} · Mikhail Hushchyn\textsuperscript{1} · Denis Derkach\textsuperscript{1}

Received: 1 November 2022 / Accepted: 25 July 2023 / Published online: 28 August 2023 © The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract
Magnetic fields are responsible for a multitude of solar phenomena, including potentially destructive events such as solar flares and coronal mass ejections, with the number of such events rising as we approach the peak of the 11-year solar cycle in approximately 2025. High-precision spectropolarimetric observations are necessary to understand the variability of the Sun. The field of quantitative inference of magnetic field vectors and related solar atmospheric parameters from such observations has been investigated for a long time. In recent years, very sophisticated codes for spectropolarimetric observations have been developed. Over the past two decades, neural networks have been shown to be a fast and accurate alternative to classic inversion methods. However, most of these codes can be used to obtain point estimates of the parameters, so ambiguities, degeneracies, and uncertainties of each parameter remain uncovered. In this paper, we provide end-to-end inversion codes based on the simple Milne-Eddington model of the stellar atmosphere and deep neural networks to both parameter estimation and their uncertainty intervals. The proposed framework is designed in such a way that it can be expanded and adapted to other atmospheric models or combinations of them. Additional information can also be incorporated directly into the model. It is demonstrated that the proposed architecture provides high accuracy results, including a reliable uncertainty estimation, even in the multidimensional case. The models are tested using simulations and real data samples.

Keywords Magnetic fields · Inverse problem · Spectral lines · Deep learning

\textsuperscript{1} HSE University, Myasnitskaya Street 20, 101000, Moscow, Russia
\textsuperscript{2} Central Astronomical Observatory of the Russian Academy of Sciences at Pulkovo, Pulkovskoye Shosse 65/1, 196140, Saint-Petersburg, Russia
\textsuperscript{3} Crimean Astrophysical Observatory, Russian Academy of Sciences, Nauchny, 298409, Crimea
\textsuperscript{4} N.P. Bechtereva Institute of the Human Brain, Academica Pavlova Street 9, 197376, Saint-Petersburg, Russia
1. Introduction

Modern solar physics relies, to a great extent, on the spectropolarimetric observations of the Sun (Ramos et al., 2016; Gafeira et al., 2021; Li et al., 2022; Leka et al., 2022). The data on the spectral and polarization state of the solar light, in conjunction with an appropriate atmosphere model, allows one to derive the thermodynamic, dynamic, and magnetic properties of the solar plasma (Viticchié and Almeida, 2011). Over the past two decades, many approaches aimed at deriving the solar atmosphere parameters have been developed. Despite the great capabilities provided by these techniques, the derivation of the atmospheric parameters from the observed spectra – the inverse problem – requires huge computing resources in many cases.

The formation of the spectral line in the solar atmosphere is described by the radiation transfer equation (RTE, Landi Degl’Innocenti and Landolfi, 2004). In turn, the polarization of light, which appears due to Zeeman splitting of the spectral line in an external magnetic field, can be described by four components I, Q, U, and V of the observed Stokes vector (Kuckein et al., 2021).

In the general case, the complexity of the RTE makes it impossible to solve the inverse problem analytically, that is, to obtain the solar atmosphere parameters from the observed Stokes profiles. Certain simplifications and approximations are often used to solve the RTE, for instance, the Milne-Eddington (ME) model of the atmosphere (Unno, 1956; Landi Degl’Innocenti and Landolfi, 2004; del Toro Iniesta Carlos and Ruiz Cobo, 2016). The model assumes a local thermal equilibrium and independence of the atmosphere parameters with height. Nevertheless, in such a case, the solution of the RTE is still non-linear and transcendent.

To fill the gap between theoretical models and complex simulations, special tools called inversion codes have been developed. These codes are classified according to the optimization strategies used to find the optimal set of atmospheric parameters that generate spectra closest to the observed ones (Lites et al., 2006). In most cases, the procedure is based on non-linear least-squares optimization with the simplified atmospheric model. More complicated models exist, that take into account height-dependent distribution of atmospheric parameters (non-local thermal equilibrium).

Data for training, validation, and testing contain spectropolarimetric images where each pixel of the image corresponds to an area on the Sun’s surface with its spectral profile. These images are converted into a three-dimensional data set $(x, y, \lambda)$, in which $\lambda$ is the width of the pixel spectral line with coordinates $(x, y)$. In this context, a data dimensionality problem arises: each pixel of an image is equal to one independent inversion problem. This makes the methods based on optimization techniques computationally expensive. As an example, to analyze data from the Helioseismic and Magnetic Imager (HMI) onboard of the Solar Dynamics Observatory (SDO) using Very Fast Inversion of the Stokes Vector (VFISV, Borrero et al., 2011), 50 CPUs were used in parallel to reach a 10-minute cadence, including specific limitations such as additional assumptions for the atmospheric model and low spectral resolution. Also, the result depends on the closeness of the initial approximation to the true value of the parameters. Furthermore, the atmospheric model itself is a multi-dimensional problem. In the simple basic ME model, there is a set of 11 atmospheric parameters which form a manifold in the parameters space. As a result, the inversion task became ill-conditioned and it may lead to instability of estimations or even multiple solutions. That is why a proper estimation of the uncertainty becomes an important issue.

As a possible solution to overcome the computational problems, a neural-network-based solution was proposed (Carroll and Staude, 2001). The main idea of this approach is to use...
neural networks for direct inversion by studying the mapping between Stokes profiles and atmospheric parameters. In multiple studies, this approach has been shown to be effective, however, in most cases the uncertainty estimation problem still needs to be solved. This is connected to a large number of parameters the neural solutions operate, which requires a lot of computing power to scan (Krzywinski and Altman, 2013; Ghahramani, 2015). The list with a short description and links to the paper known to us related to the Stokes inversion problem is summarized in Table 4.

Within the last several years, various approaches have been developed to estimate the prediction uncertainty. These methods can be split into four groups, based on the number and the nature of the used neural networks (Lakshminarayanan, Pritzel, and Blundell, 2017): single deterministic methods, where the uncertainty of a prediction is computed based on one single forward pass within a deterministic network (Malinin and Gales, 2018), Bayesian methods, which contain uncertainty in their networks, assuming that parameters are defined as some probability distributions (Blundell et al., 2015), augmentation methods, which are based on the modification of the training data set so that a model learns on the extended data (Shorten and Khoshgoftaar, 2019), and ensembles, which derive a final prediction based on other predictions received from multiple ensemble members (Lakshminarayanan, Pritzel, and Blundell, 2017).

One of the probabilistic approaches to solving the inverse problem is to use algorithms based on Bayesian inference, such as the Markov Chain Monte Carlo (MCMC) and Nested Sampling (NS). Despite limitations in their applicability, even for relatively simple atmospheric models and requirements of detailed knowledge of the parameter space, Bayesian approaches have long been used in solar physics analysis (Ramos, González, and Rubiño-Martín, 2007; Li et al., 2019). An alternative to these may be variational inference methods, for example, normalizing flows, where the true distribution of the solution is approximated by a simpler analytical one (Ramos et al., 2017; Baso, Ramos, and de la Cruz Rodríguez, 2022). However, the disadvantage of such models is that their optimization is not always stable, so one has to consider the simpler posterior distribution.

Recently, it has been proposed to use a single deterministic network to quantify uncertainties in the prediction of atmospheric parameters (Higgins et al., 2021, 2022). The authors of these papers suggest obtaining confidence intervals by treating the problem as a regression by classification, such that, for each pixel of the parameter image, the model predicts the distribution of its possible values by applying the softmax function.

In this paper, we focus on uncertainty estimation using a combination of several convolutional neural networks modified in such a way that they can quantify the uncertainty in predictions (treating the observed value as a sample from a Gaussian distribution) and trained as an ensemble. To the best of our knowledge, this is the first systematic study of these approaches within the framework of the inverse solar problem.

2. Uncertainty Quantification and Metrics

Experimental measurement uncertainty plays a central role in physical sciences. The assigned uncertainties can point to the reliability of the measurement. That is why, currently, the determination of uncertainty plays a crucial role in the analysis of the experiment. In this paper, we follow the most frequently used method, which leads to an important consequence: the methods proposed should estimate the interval with a given confidence, that is, provide experimental coverage probability for a given confidence level. This in turn means
obtaining the rate at which the true value is contained in the confidence interval of an individual measurement (Pawitan, 2001).

We check the correctness of the procedure using a graphical representation, as shown later in Figures 5 and 6. We also use integral metrics to estimate how accurately we evaluate the confidence interval in our model: the normalized Mean Squared Error (nMSE) (Quinonero-Candela et al., 2005), the Negative Log Predictive Density (NLPD) (Quinonero-Candela et al., 2005), and the Prediction Interval Coverage Probability (PICP) (Shrestha Durga and Solomatine Dimitri, 2006) with two different fractions of the distribution inside the confidence interval. The nMSE is defined as

\[
\text{nMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{(t_i - m_i)^2}{\sigma_i^2}},
\]  

(1)

where \(m_i, \sigma_i^2\) are the mean and variance of the predictive distribution, respectively, \(N\) denotes the total number of pixels (the size of a test data set), and \(t_i\) is a true sample. The NLPD is defined as

\[
\text{NLPD} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{(t_i - m_i)^2}{2\sigma_i^2} + \log \sigma_i + c \right),
\]  

(2)

where \(c\) is a constant independent of \(m_i\) and \(\sigma_i\). Specifying the upper PL \(U_i\) and lower PL \(L_i\) bounds on a prediction \(i\) (the uncertainty of a single observation), one can calculate the prediction interval coverage probability metric (PICP). The PICP is defined as the probability that the real value lies within the predicted confidence interval and estimated as the following frequency:

\[
\text{PICP} = \frac{1}{N} \sum_{i=1}^{N} m_i, \quad m_i = \begin{cases} 1, & PL_i^L \leq t_i \leq PL_i^U, \\ 0, & \text{otherwise} \end{cases},
\]  

(3)

where \(t_i\) is the \(i\)-th element of the reference data set. Both the nMSE and PICP metrics most penalize the cases of incorrect forecasts made with uncertainty close to zero. However, PICP is sensitive to both when the prognosis is not sufficiently certain and when it is over-confident. The PICP value should be close to the \(\alpha\), however, over-confident predictions are worse for the model.

These metrics show the overall performance of the algorithm, while the local performance might vary depending on the point in the parameter space.

3. Data and Methods

The solution of the inverse problem requires, at the first stage, the initial approximation of the Stokes profiles. These profiles can be generated synthetically using atmospheric models, and this method has been shown to be highly effective (Knyazeva et al., 2022). The ME codes obtained from the Hinode/SOT/SP database were used to construct 5.3 million synthetic records of the state of the solar atmosphere, collected over 12 different time periods from 2010 to 2017, of which 10% were used for validation and the rest for training. The data for testing \((x_{true})\) refer to the result of Stokes profiles inversion using the ME codes HAO.
“MERLIN” from samples collected on 26 September 2014 for NOAA 12172 active region (see Figure 1).

Generated spectra used for training were obtained directly from the Hinode/SOT/SP database from the Community Spectropolarimetric Analysis Center at HAO/NCAR using the ME codes and were made more similar to real samples by adding Gaussian noise. Assuming that the level of noise, which enters the receiver with the true signal value, is always constant and independent of the input signal, the resulting spectrum can be calculated as the sum of actual solar radiation and noise. According to the Milne–Eddington model, the parameters of the four profiles should tend toward zero as they move away from the center of the line. In this case, noise determines the value of the extreme pixels of the spectrum, and it has been shown that the distribution of the noise is close to Gaussian (Carroll, Kopf, and Strassmeier, 2008; Li et al., 2019).

In the following study, we consider 11 parameters of the solar atmosphere: the magnetic field vector, consisting of field strength component (0 ÷ 5000 G), inclination angle (0 ÷ 180°) and azimuth angle (0 ÷ 180°); line parameters, consisting of Doppler width (20 ÷ 90 mÅ), line damping (0 ÷ 1.5 Dopplerwidths) and line strength (0.01 ÷ 100); two intensity parameters, consisting of source function (SF, 0 ÷ 1) and its height gradient (0 ÷ 1); Doppler shift (DS) of the line (−10 ÷ 10 km s⁻¹); stray light Doppler shift (−10 ÷ 10 km s⁻¹) and magnetic filling factor (0 ÷ 1). Before training, true samples of parameter values were transformed by applying the logarithmic transformation to magnetic-field vector and trigonometric to magnetic-field inclination and azimuth angles, and were brought to one scale with min-max normalization. The example of color maps is shown in Figure 1.

We used a single deterministic method to evaluate the predictive uncertainty and modified architecture of the partial sharing model (Knyazeva et al., 2022) in such a way that it predicts two values in the final layer for each pixel, corresponding to the mean $\mu(x)$ and the variance $\sigma^2(x) > 0$. The proposed model consists of 11 independent MLP blocks, predicting the mean value and the standard deviation of each pixel in the case of 11 atmospheric parameters. Independent blocks took as input the result of one common MLP block, which in turn
The model has 11 independent MLP blocks, predicting the mean and the variance of atmospheric parameters, one common MLP block, and six common convolutional blocks of two types. Blocks of the second type contain the following layers: one-dimensional convolution, batch normalization, ELU activation function, and dropout. In blocks of the first type, the max-pooling layer is added. The green color indicates the features being entered during training.

followed six common convolutional blocks of two types. Blocks of the first type contain the following layers: one-dimensional convolution with kernel size 3, max pooling with kernel size 2, batch normalization, ELU activation function, and dropout. In blocks of the second type, the max pooling layer is excluded. It was reasonable to add convolutional blocks to the model since Stokes profiles are usually interconnected. The schematic representation of the architecture is shown in Figure 2.

As the observed value was treated as a sample from a Gaussian distribution, the model was trained by minimizing the negative logarithm loss function (Gawlikowski et al., 2021):

$$L(θ, y, x) = -\frac{1}{M K} \sum_{i=1}^{M} \sum_{j=1}^{K} \log \left( \frac{1}{\sqrt{2\pi} \sigma_{θij}(x)} \exp \left[ -\frac{(y_{ij} - μ_{θij}(x))^2}{2\sigma_{θij}^2(x)} \right] \right) ,$$  (4)

where $μ_{θ}(x)$ and $σ_{θ}(x)$ – predicted average value and standard deviation, respectively, $K = 11$ and $M = 128$ are the number of parameters that were considered and the batch size, respectively. We choose Gaussian distribution as a starting point for studying the problem. In addition, it was shown that the least squares method is applicable to any distribution when the noise level is less than 5% (Xu, Chen, and Liang, 2018). The Adam optimizer was used for the iterative update of the model weights. The model was found to achieve the required result in five epochs based on the degradation of performance on validation datasets.

While other methods exist (Gawlikowski et al., 2021), to obtain an estimate, a scan of the loss function can be used. This requires a lot of computing resources, which is why approximate methods are implemented. These methods allow estimation of the minimum width, however, their performance depends on the problem. We test the ensemble-based model (Lakshminarayanan, Pritzel, and Blundell, 2017), which is shown to perform well in open datasets. The model described above was used to build an ensemble of several models that were trained using bagging, thus the training and validation data were divided equally between the models. The final prediction was treated as a uniformly weighted mixture of
Table 1 Performance metrics of the convolutional model predictions made on synthetic data of Stokes profiles (from samples collected on 26 September 2014 for NOAA 12172 active region).

| Parameter                  | $R^2$  | MSE    | MAE    | PPMCC  | NLPD   | nRMSE  | PICP$_{68}$ | PICP$_{95}$ |
|----------------------------|--------|--------|--------|--------|--------|--------|-------------|-------------|
| Field Strength             | 0.986  | 4135.95| 37.56  | 0.993  | 5.534  | 1.399  | 0.699       | 0.947       |
| Field Inclination          | 0.938  | 87.272 | 3.719  | 0.970  | 3.181  | 1.634  | 0.678       | 0.918       |
| Field Azimuth              | 0.595  | 1069.36| 19.87  | 0.855  | 4.601  | 1.901  | 0.506       | 0.724       |
| Doppler Width              | 0.970  | 2.7360 | 1.097  | 0.979  | 1.616  | 1.043  | 0.666       | 0.947       |
| Damping                    | 0.935  | 0.0031 | 0.028  | 0.963  | −2.214 | 1.046  | 0.683       | 0.945       |
| Line Strength              | 0.810  | 0.0026 | 0.019  | 0.837  | −2.733 | 1.140  | 0.696       | 0.936       |
| SF                         | 0.899  | 0.0005 | 0.015  | 0.940  | −2.655 | 1.091  | 0.647       | 0.937       |
| Cont. SF Grad.             | 0.980  | 0.0002 | 0.010  | 0.988  | −3.073 | 1.067  | 0.659       | 0.941       |
| DS                         | 0.952  | 0.0336 | 0.095  | 0.966  | −1.149 | 1.039  | 0.682       | 0.948       |
| Filling Factor             | 0.852  | 0.0101 | 0.064  | 0.918  | −1.298 | 1.132  | 0.664       | 0.926       |
| Stray Light DS             | 0.629  | 0.9504 | 0.295  | 0.756  | 1.979  | 2.325  | 0.757       | 0.948       |

Figure 3 Two-dimensional color maps of model predictions made on synthetic data.

Gaussian distributions, and the combination of results was determined as follows:

\[
\mu_x(x) = \frac{1}{N} \sum_{i=1}^{N} \mu_{\theta_i}(x),
\]

\[
\sigma_x^2(x) = \frac{1}{N} \sum_{i=1}^{N} \left( \sigma_{\theta_i}^2(x) + \mu_{\theta_i}^2(x) \right) - \mu_x^2(x),
\]
where \( \mu_{\theta_i}(x) \), \( \sigma_{\theta_i}(x) \) are the mean and the standard deviation of the \( i \)-th model, respectively, and \( N = 6 \) is the number of models used. Predicted values were scaled back into their physical ranges after training to correctly interpret the results.

4. Results

For each parameter, we compared its ME-calculated values with our network-inferred values and computed performance metrics. Four quality metrics are the following: the coefficient of determination \( R^2 \), the mean squared error (MSE), the mean absolute error (MAE) and the Pearson product-moment correlation coefficient (PPMCC). Additionally, we show the metrics that characterize the uncertainty region, as defined in Section 2.

The data for training and validation were collected in such a way that they contained both the quiet and active areas of the Sun. The metrics are represented in Table 1. It can be seen that all the prediction parameters are covered by confidence intervals. The two-dimensional color maps of the 11 reconstructed parameters are visualized in Figure 3 and a comparison of these with the reference values can be seen in Figure 4.

To estimate the performance of the method, we also suggest plotting several dependencies. The scatter plots of the standard deviation \( \sigma_{\text{pred}} \) on the difference \( x_{\text{true}} - x_{\text{pred}} \) are presented in Figure 5. Color intensity indicates the frequency of the values encountered. As one can see, the areas with the highest concentration of points are located near the zero-error line. However, at the edges of the parameters there are significant deviations in predictions.
Figure 5 Dependence of standard deviations on the difference between predictions made on synthetic data and the true values. Each figure corresponds to one of the atmospheric parameters. Color intensity indicates the frequency of the values encountered.

from the reference values. The reason for this behavior could be the small number or complete absence of samples with such parameter values in the training data.

For further analysis, dependencies of the ratio \((x_{\text{true}} - x_{\text{pred}}) / \sigma_{\text{pred}}\) on the true values \(x_{\text{true}}\) in case of each parameter were divided into approximately 500 segments, and then each segment was fitted by a normal distribution. These fitting curves can be seen in Figure 6. In the ideal case, the mean values have to be close to 0, while the standard deviations have to be in the range \([-1; 1]\). It can be seen that for some parameters there are deviations from the ideal scenario. Regions known as the most difficult to reconstruct (such as low field strength) have some over- or under-estimation of uncertainty, but no more than 30%. Also, in this framework, the problem of azimuth disambiguation remains actual. To solve this problem, we need to apply stand-alone disambiguation methods (Rudenko and Anfinogentov, 2013).

In addition, an ensemble was created in order to improve the quality of predictions based on single convolutional models and using bagging. The performance metrics of the ensemble can be seen in Table 2. A comparison of the results obtained by one model and the ensemble of models is in Figure 6. As can be seen, the use of ensembles leads to a reduction of prediction variance and smoothing of the results.

The model was also tested on the real data, Hinode/SOT/SP observations from the Community Spectropolarimetric Analysis Center at HAO/NCAR collected on 28 June 2014 for NOAA 12096 active region. A learning step on real spectra was added before testing to fine-tune the model and avoid over-fitting on synthetic samples. Comparison between the predictions made from the real Stokes parameters and the test data is in Figure 7. Density graphs similar to those shown for synthetic spectra can be seen in Figure 8. It can be seen...
Figure 6  Analysis of the difference between predictions made on synthetic data and the true values divided by standard deviation as a function of the true values. Being segmented, these data were fitted with a normal distribution, then the dependence of the parameters of this approximation on the true data was constructed. Each figure corresponds to one of the atmospheric parameters. In the case of one single model, mean values are marked by solid orange lines and the confidence intervals are by translucent orange areas. The average curves obtained from an ensemble are marked by solid green lines, and the corresponding confidence intervals by translucent green areas.

that the new approach infers atmospheric parameters with an accuracy comparable to the ME inversion technique.

It could also be noted that the model provides physically adequate uncertainties. For example, regions of weak magnetic field (where Stokes Q, U, and V parameters have low amplitudes and thus bad signal-to-noise ratio) correspond to predictions with larger uncertainties for the magnetic field vector. At the same time, pixels with strong magnetic field (where all four Stokes parameters suffer from a low signal-to-noise ratio, causing a lack of light intensity) match predictions with large uncertainties for almost all parameters.

The performance metrics are presented in Table 3. In some of the more noisy regions of the Sun, the model predicts less accurate and overconfident results, thus leaving space for improvement of the model as well as the method of generating synthetic spectra. We assume that performance quality and robustness to out-of-distribution samples can be improved if the synthetic generation algorithm is upgraded, since synthetic profiles are usually symmetrical, while the distribution of the real data could be uneven and asymmetrical.

The real data test shows the differences in behavior between PICP and other metrics. This difference is motivated by the fact that the PICP is aimed to produce coverage tests, while others aim to show the overall quality of fit and uncertainties obtained. The results of
Table 2  Performance metrics of predictions made by the convolutional model ensemble based on synthetic data of Stokes profiles.

| Parameter                  | $R^2$  | MSE   | MAE   | PPMCC | NLPD  | nRMSE | PICP68 | PICP95 |
|----------------------------|--------|-------|-------|-------|-------|-------|--------|--------|
| Field Strength             | 0.991  | 2177.3| 31.08 | 0.995 | 4.946 | 0.999 | 0.699  | 0.959  |
| Field Inclination          | 0.984  | 20.992| 2.479 | 0.987 | 2.316 | 1.061 | 0.701  | 0.950  |
| Field Azimuth              | 0.686  | 799.62| 15.25 | 0.907 | 4.019 | 1.739 | 0.546  | 0.784  |
| Doppler Width              | 0.976  | 1.988  | 0.963 | 0.980 | 1.492 | 1.028 | 0.676  | 0.948  |
| Damping                    | 0.966  | 0.0012 | 0.021 | 0.975 | 2.402 | 1.031 | 0.680  | 0.946  |
| Line Strength              | 0.872  | 0.0009 | 0.014 | 0.891 | 2.994 | 1.059 | 0.700  | 0.944  |
| SF                         | 0.911  | 0.0004 | 0.013 | 0.941 | 2.829 | 0.982 | 0.699  | 0.958  |
| Cont. SF Grad.             | 0.982  | 0.0001 | 0.008 | 0.988 | 3.260 | 0.970 | 0.704  | 0.960  |
| DS                         | 0.969  | 0.0205 | 0.077 | 0.977 | 1.294 | 0.982 | 0.702  | 0.957  |
| Filling Factor             | 0.881  | 0.0068 | 0.054 | 0.926 | 1.552 | 1.009 | 0.685  | 0.954  |
| Stray Light DS             | 0.906  | 0.1909 | 0.179 | 0.897 | 0.403 | 1.045 | 0.741  | 0.957  |

Figure 7  Two-dimensional color maps of the results of Stokes profiles inversion by the ME codes HAO “MERLIN”, model predictions made on real data of Stokes profiles (collected on 28 June 2014 for NOAA 12096 active region), and results of uncertainty estimation in these predictions. Each column corresponds to one of the first four atmospheric parameters. For each parameter, the true values are shown in the first line, predictions in the second line, and results of uncertainty estimation in the third line.

the test can be interpreted as the need to recalibrate the uncertainties if the model is trained on simulation and applied to real data (Kuleshov, Fenner, and Ermon, 2018).

5. Conclusion and Discussion

Machine learning, in particular the neural network approach to Stokes profile inversion, is gaining popularity due to computational efficiency, but physical models often require not
Figure 8  Analysis of predictions made on real data of Stokes profiles: dependence of predictions on the true values and dependence of standard deviations on the difference between predictions and the true values. Each plot corresponds to one of the atmospheric parameters. Red dashed lines represent pixels whose predictions are equal to the true values. Color intensity indicates the frequency of the values encountered.

Table 3  Performance metrics of the convolutional model predictions made on real data of Stokes profiles (samples collected on 28 June 2014 for NOAA 12096 active region).

| Parameter           | $R^2$ | MSE     | MAE     | PPMCC   | NLPD   | nRMSE  | PICP68 | PICP95 |
|---------------------|-------|---------|---------|---------|--------|--------|--------|--------|
| Field Strength      | 0.914 | 25·10^3 | 90.09   | 0.956   | 95.57  | 13.60  | 0.094  | 0.187  |
| Field Inclination   | 0.888 | 119.69  | 5.769   | 0.942   | 6.458  | 2.910  | 0.526  | 0.750  |
| Field Azimuth       | 0.572 | 1178.3  | 20.63   | 0.756   | 15.11  | 5.073  | 0.188  | 0.375  |
| Doppler Width       | 0.644 | 60.24   | 5.136   | 0.802   | 26.25  | 7.149  | 0.117  | 0.232  |
| Damping             | 0.535 | 0.023   | 0.102   | 0.707   | -0.661 | 0.939  | 0.767  | 0.959  |
| Line Strength       | 0.250 | 0.022   | 0.112   | 0.500   | -0.624 | 1.025  | 0.675  | 0.973  |
| SF                  | 0.623 | 0.036   | 0.170   | 0.789   | 0.001  | 1.543  | 0.277  | 0.824  |
| Cont. SF Grad.      | 0.860 | 0.037   | 0.181   | 0.927   | 0.161  | 1.679  | 0.164  | 0.740  |
| DS                  | 0.748 | 0.174   | 0.307   | 0.865   | 0.782  | 1.667  | 0.471  | 0.775  |
| Filling Factor      | 0.633 | 0.021   | 0.101   | 0.796   | -0.510 | 0.738  | 0.866  | 0.979  |
| Stray Light DS      | 0.205 | 1.697   | 1.170   | 0.453   | 7.936  | 3.979  | 0.122  | 0.336  |

only point estimates but also errors. In this paper, we provide a novel neural network architecture for inferring solar atmospheric parameters together with their predictive uncertainties by modifying model architecture and loss function. The method was tested on the 11 atmospheric parameters: three components of the magnetic field vector, three line parameters, two intensity parameters, Doppler shift of the line, stray light Doppler shift, and magnetic filling factor. Several performance metrics were calculated ($R^2$, MSE, MAE, PPMCC, NLPD, nRMSE, PICP_{68}, and PICP_{95}) in case of synthetic data, as well as real Hinode observations collected on 28 June 2014 for NOAA 12096 active region. When choosing a quality metric such as MSE, errors are assumed to be normally distributed, so uncertainty is evaluated in this work also based on the same assumptions. Further analysis should include a precise determination of the associated uncertainty in the particular experimental setup.
Table 4  Overview of neural network studies in Stokes inversion problem. The following abbreviations were used in this table: magnetic field (MF), velocity (V), temperature (T), source function and its gradient (SF and SFG), Doppler width (DW), filling factor (FF), and line damping (LD).

| Reference                        | Network architecture                          | Input data                        | Output data                                      |
|----------------------------------|-----------------------------------------------|-----------------------------------|--------------------------------------------------|
| Carroll and Staude 2001          | MLP blocks                                    | Synthetic                         | MF, V, FF and other                              |
| Socas-Navarro 2005               | MLP blocks                                    | Synthetic and real from High Altitude Observatory (HAO) | MF, DW, SF, FF, LD and other                     |
| Ramos et al. 2007                | Bayesian inference                            | Synthetic                         | MF                                               |
| Ramos et al. 2017                | Variational inference                         | Real from Swedish Solar Telescope (SST) | MF                                               |
| Ramos and Baso 2019              | CNN with 2D convolutions                      | Synthetic                         | MF, V, T and other                               |
| Sainz Dalda et al. 2019          | MLP blocks                                    | Synthetic                         | V, T and other                                   |
| Li et al. 2019                   | Bayesian inference                            | Synthetic                         | MF, V, DW, SF + SFG, LD and other                |
| Liu et al. 2020                  | CNN with 1D convolutions                      | Real from Goode Solar Telescope (GST) | MF                                               |
| Milić and Gafeira 2020           | CNN with 1D convolutions                      | Synthetic                         | MF, V, T                                        |
| Gafeira 2021                     | Ensemble of CNNs with 1D convolutions          | Synthetic and real from Gregor telescope | MF, V, T                                        |
| Guo et al. 2021                  | CNN with 2D convolutions                      | Real from Hinode telescope        | MF                                               |
| Higgins et al. 2021              | CNN (U-Net), treating the problem as a regression by classification | Real from Solar Dynamics Observatory (SDO) | MF, DW, SF + SFG, and other                      |
| Baso et al. 2022                 | Variational inference                         | Synthetic and real from Swedish Solar Telescope (SST) | V, T, DW, SF + SFG and other                     |
| Knyazeva et al. 2022             | MLP blocks                                    | Synthetic                         | MF, DW, FF, SF + SFG, LD and other               |
| Higgins et al. 2022              | CNN (U-Net), treating the problem as a regression by classification | Synthetic                         | MF, FF and other                                 |
| Jiang et al. 2022                | CNN with 1D convolutions                      | Real from Goode Solar Telescope (GST) | MF, V, DW                                        |
| Present work                     | Ensemble of CNNs with 1D convolutions and uncertainty quantification | Synthetic and real from Hinode telescope | MF, DW, FF, SF + SFG, LD and other               |

The maps of Stokes profiles inversion by Milne–Eddington codes HAO “MERLIN” were taken as ground truth, and about a minute was required for the model to make predictions of one map of 512×873 pixels. Analysis showed that the proposed model represents a reliable method compared with classical methods for solving the inverse problem. In addition, it was shown that the smoothness and the accuracy of results and the width of the uncertainty intervals can be improved by ensembles. It is also important to note that there are also
uncertainties within the true data, as the experimental values are not subtracted directly, but by averaging several observations (Lites and Ichimoto, 2013). Nevertheless, on a reduced set of observations, we show that the proposed method provides reasonable results and thus, can be used to improve theoretical calculations and provide a starting point for more precise methods, thus making it possible to reduce the total computation time. Although synthetic spectra are symmetric, which is unusual for real data, the model trained on synthetic spectra showed the ability to generalize even in the case of real observations.

The method proposed can be used for analysis in various fields of astrophysics (Okamoto et al., 2009; Podladchikova et al., 2022): in the analysis of the solar cycle and prediction of a coronal mass ejection, in the analysis of the solar atmosphere itself, for example, to study the spatial distribution of parameters or of the processes in the solar atmosphere such as plasma convection, and open prospects for future studies. Further analysis raises the question of the credibility of the results obtained by the network. The model described is a simple and scalable method for quantifying uncertainty. Since it necessitates only a modification of architecture (doubling the number of output layers and changing the loss function), it requires as much learning time as a network that is not modified.

Acknowledgments Denis Derkach, Lukia Mistryukova, Aleksandr Khizhik, and Mikhail Hushchyn are grateful to the HSE basic research program. This research was supported in part through computational resources of HPC facilities at HSE University (Kostenetskiy, Chulkevich, and Kozyrev 2021).

Author contributions D.D. conceived and led the research; L.M. and A.K. performed calculations and processed the results; A.P., I.K., M.H. and D.D. analyzed and discussed the results; L.M., I.K. and D.D. wrote the main manuscript text; A.P. and I.K. processed data samples. All authors reviewed the manuscript.

Data Availability In the current study, we used a collection of the Level 1 calibrated Stokes spectra (comprised by images stored in FITS format) and a collection of the Level 2 data sets (obtained from the MERLIN spectral line inversion of the Level 1 calibrated spectra) produced by the Spectropolarimeter (SP) on board the Hinode, since its launch in 2006 (collected in the Community Spectropolarimetric Analysis Center (CSAC) at HAO/NCAR). Hinode is a Japanese mission, developed and launched by ISAS/JAXA, with NAOJ as a domestic partner and NASA and STFC (UK) as international partners. It is operated by these agencies in cooperation with ESA and NSC (Norway). The Hinode has an open data policy, allowing anyone to access the data and data products. Level 1 and 2 data are available by following the data link.

Declarations

Competing interests The authors declare no competing interests.

References

Baso, C.J.D., Ramos, A.A., de la Cruz Rodríguez, J.: 2022, Bayesian Stokes inversion with normalizing flows. Astron. Astrophys. 659, A165. DOI.
Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D.: 2015, Weight uncertainty in neural networks. J. Mach. Learn. Res.. DOI.
Borrero, J.M., Tomczyk, S., Kubo, M., Socas-Navarro, H., Schou, J., et al.: 2011, VFISV: very fast inversion of the Stokes vector for the helioseismic and magnetic imager. Solar Phys. 273(1), 267. DOI.
Carroll, T., Kopf, M., Strassmeier, K.: 2008, A fast method for Stokes profile synthesis. Astron. Astrophys. 488(2), 781. DOI.
Carroll, T.A., Staude, J.: 2001, The inversion of Stokes profiles with artificial neural networks. Astron. Astrophys. 378, 316. DOI.
Dalda, A.S., de la Cruz Rodríguez, J., Pontieu, B.D., Gošić, M.: 2019, Recovering thermodynamics from spectral profiles observed by IRIS: a machine and deep learning approach. Astrophys. J. Lett. 875(2), L18. DOI.
del Toro Iniesta Carlos, J., Ruiz Cobo, B.: 2016, Inversion of the radiative transfer equation for polarized light. *Living Rev. Solar Phys.* **13**(1), 4. DOI.

Gafeira, R., Suárez, D.O., Milić, I., Noda, C.Q., Cobo, B.R., Uitenbroek, H.: 2021, Machine learning initialization to accelerate Stokes profile inversions. *Astron. Astrophys.* **651**, A31. DOI.

Gawlikowski, J., Cedriqué Rovile, N.T., Ali, M., Lee, J., Humt, M., Feng, J., et al.: 2021, A survey of uncertainty in deep neural networks. Preprint. arXiv. DOI.

Ghahramani, Z.: 2015, Probabilistic machine learning and artificial intelligence. *Nature* **521**(7553), 452. DOI.

Guo, J., Bai, X., Liu, H., Yang, X., Deng, Y., Lin, J., et al.: 2021, A nonlinear solar magnetic field calibration method for the filter-based magnetograph by the residual network. *Astron. Astrophys.* **646**, A41. DOI.

Higgins, R.E.L., Fouhey, D.F., Zhang, D., Antiochos, S.K., Barnes, G., Todd, H.J., et al.: 2021, Fast and accurate emulation of the SDO/HMI Stokes inversion with uncertainty quantification. *Astrophys. J.* **911**(2), 130. DOI.

Higgins, R.E.L., Fouhey, D.F., Antiochos, S.K., Barnes, G., Cheung, M.C.M., Todd, H.J., et al.: 2022, SynthIA: a synthetic inversion approximation for the Stokes vector fusing SDO and Hinode into a virtual observatory. *Astrophys. J. Suppl.* **259**(1), 24. DOI.

Jiang, H., Li, Q., Xu, Y., Hsu, W., Ahn, K., Cao, W., Wang, J.T.L., Wang, H.: 2022, Inferring line-of-sight mitigating bias and Doppler widths from Stokes profiles of gis/niris using stacked deep neural networks. *Astrophys. J.** 939**(2), 66. DOI.

Knazyeva, I., Plotnikov, A., Medvedeva, T., Makarenko, N.: 2022, Multi-output deep learning framework for solar atmospheric parameters inferring from Stokes profiles. In: *Advances in Neural Computation, Machine Learning, and Cognitive Research V*. 1008. Springer, Cham, 299. DOI.

Kostneteskiy, P.S., Chulkевич, R.A., Kozyrev, V.I.: 2021, HPC resources of the higher school of economics. *J. Phys. Conf. Ser.* **1740**(1), 012050. DOI.

Krzywinski, M., Altman, N.: 2013, Power and sample size. *Nat. Methods* **10**(12), 1139. DOI.

Kuckein, C., Balthasar, H., Noda, C.Q., Dierckx, A., Arjona, J.C.T., Cobo, B.R., et al.: 2021, Multiple Stokes I inversions for inferring magnetic fields in the spectral range around Cr I 5782 Å. *Astron. Astrophys.* **653**, A165. DOI.

Kuleshov, V., Fenner, N., Ermon, S.: 2018, Accurate uncertainties for deep learning using calibrated regression. DOI.

Lakshminarayanan, B., Pritzel, A., Blundell, C.: 2017, Simple and scalable predictive uncertainty estimation using deep ensembles. In: *Advances in Neural Information Processing Systems*. MIT Press, Cambridge, 6405. DOI.

Landi Degl’Innocenti, E., Landolfi, M.: 2004, *Polarization in Spectral Lines*, Springer, Dordrecht. DOI.

Leka, K.D., Wagner, E.L., Gríñón-Marín, A.B., Bonmier, V., Higgins, R.E.L.: 2022, On identifying and mitigating bias in inferred measurements for solar vector magnetic field data. *Solar Phys.* **297**, 121. DOI.

Li, H., Xu, Z., Qu, Z., Sun, L.: 2019, MCMC inversion of Stokes profiles. *Astrophys. J.* **875**(2), 127. DOI.

Li, H., del Pino Alemán, T., Bueno, J.T., Casini, R.: 2022, TIC: a Stokes inversion code for scattering polarization with partial frequency redistribution and arbitrary magnetic fields. *Astrophys. J.* **933**(2), 145. DOI.

Lites, B.W., Ichimoto, K.: 2013, The SP_PREP data preparation package for the Hinode spectro-polarimeter. *Solar Phys.* **283**(2), 601. DOI.

Lites, B., Casini, R., Garcia, J., Socas-Navarro, H.: 2006, A suite of community tools for spectro-polarimetric analysis. *MemSAIt* **78**, 148.

Liu, H., Xu, Y., Wang, J., Jing, J., Liu, C., Wang, J.T.L., Wang, H.: 2020, Inferring vector magnetic fields from Stokes profiles of GST/NIRIS using a convolutional neural network. *Astrophys. J.* **894**(1), 70. DOI.

Malinin, A., Gales, M.: 2018, Predictive uncertainty estimation via prior networks. In: *Advances in Neural Information Processing Systems*, Curran Associates, Red Hook. DOI.

Milić, I., Gafeira, R.: 2020, Mimicking spectropolarimetric inversions using convolutional neural networks. *Astron. Astrophys.* **644**, A129. DOI.

Okamoto, T., Tsuneta, S., Lites, B., Kubo, M., Yokoyama, T., Berger, T.E., et al.: 2009, Prominence formation associated with an emerging helical flux rope. *Astrophys. J.* **697**, 913. DOI.

Pawitan, Y.: 2001 *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, OUP, Oxford.

Podladchikova, T., Jain, S., Veronig, A., Sutyrina, O., Dumbovic, M., Clette, F., et al.: 2022, Maximal growth rate of the ascending phase of a sunspot cycle for predicting its amplitude. *Astron. Astrophys.* **663**, A88. DOI.

Quinonero-Candela, J., Rasmussen, C.E., Sinz, F., Bousquet, O., Schölkopf, B.: 2005, *Evaluating Predictive Uncertainty Challenge*, Machine Learning Challenges Workshop 1. Springer, Berlin. DOI.

Ramos, A.A., Baso, C.D.: 2019, Stokes inversion based on convolutional neural networks. *Astron. Astrophys.* **626**, A102. DOI.
Ramos, A.A., González, M.J.M., Rubiño-Martín, J.A.: 2007, Bayesian inversion of Stokes profiles. *Astron. Astrophys.* **476**(2), 959. DOI.

Ramos, A.A., de la Cruz Rodríguez, J., González, M.J.M., Yabar, A.P.: 2016, Inversion of Stokes profiles with systematic effects. *Astron. Astrophys.* **590**, A87. DOI.

Ramos, A.A., de la Cruz Rodríguez, J., González, M.J.M., Socas-Navarro, H.: 2017, Inference of the chromospheric magnetic field orientation in the Ca ii 8542 a line fibrils. *Astron. Astrophys.* **599**, A133. DOI.

Rudenko, G.V., Anfinogentov, S.A.: 2013, Very fast and accurate azimuth disambiguation of vector magnetograms. *Solar Phys.* **289**(5), 1499. DOI.

Shorten, C., Khoshgoftaar, T.M.: 2019, A survey on image data augmentation for deep learning. *J. Big Data* **6**(1), 60. DOI.

Shrestha Durga, L., Solomatine Dimitri, P.: 2006, Machine learning approaches for estimation of prediction interval for the model output. *Neural Netw.* **19**(2), 225. DOI.

Socas-Navarro, H.: 2005, Strategies for spectral profile inversion using artificial neural networks. *Astrophys. J.* **621**(1), 545. DOI.

Unno, W.: 1956, Line formation of a normal Zeeman triplet. *Publ. Astron. Soc. Japan* **8**, 108.

Viticchié, B., Almeida, J.S.: 2011, Asymmetries of the Stokes V profiles observed by Hinode SOT/SP in the quiet sun. *Astron. Astrophys.* **530**, A14. DOI.

Xu, W., Chen, W., Liang, Y.: 2018, Feasibility study on the least square method for fitting non-Gaussian noise data. *Physica A* **492**, 1917. DOI.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.