MAGNETIC NONPOTENTIALITY IN PHOTOSPHERIC ACTIVE REGIONS AS A PREDICTOR OF SOLAR FLARES

XIAO YANG1,2, GANGHUA LIN1, HONGQI ZHANG1, and XINJIE MAO1,3
1 Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100012, China; yangx@nao.cas.cn
2 Graduate University of Chinese Academy of Sciences, No.19A Yuquan Road, Beijing 100049, China
3 Department of Astronomy, Beijing Normal University, No.19 Xinjiekouwai Street, Beijing 100875, China

ABSTRACT

Based on several magnetic nonpotentiality parameters obtained from the vector photospheric active region magnetograms obtained with the Solar Magnetic Field Telescope at the Huairou Solar Observing Station over two solar cycles, a machine learning model has been constructed to predict the occurrence of flares in the corresponding active region within a certain time window. The Support Vector Classifier, a widely used general classifier, is applied to build and test the prediction models. Several classical verification measures are adopted to assess the quality of the predictions. We investigate different flare levels within various time windows, and thus it is possible to estimate the rough classes and erupting times of flares for particular active regions. Several combinations of predictors have been tested in the experiments. The True Skill Statistics are higher than 0.36 in 97% of cases and the Heidke Skill Scores range from 0.23 to 0.48. The predictors derived from longitudinal magnetic fields do perform well, however, they are less sensitive in predicting large flares. Employing the nonpotentiality predictors from vector fields improves the performance of predicting large flares of magnitude \( \geq M5.0 \) and \( \geq X1.0 \).

Key words: methods: statistical – Sun: activity – Sun: flares – Sun: photosphere

Online-only material: machine-readable table

1. INTRODUCTION

Solar flares are sudden processes that release tremendous energy in a short period of time in the solar atmosphere. They lead to transient heating of local regions and the dramatic enhancement of electromagnetic radiation and high-energy particle ejection. Some large eruptions toward the Earth have an impact on normal human activities. It is worthwhile to make short-term predictions of solar flares to reduce losses. For a long period of time, solar physicists have been trying to understand the physics of flares, in order to make predictions by simulating the evolutions of magnetic fields in the solar atmosphere and by obtaining information from the solar interior. At present, however, it seems relatively feasible to make predictions based on the statistical relationships between solar eruptions and the evolution of other solar phenomena. Some authors predict flares based on morphological parameters or remote information from different sources (e.g., Gallagher et al. 2002; Qahwaji & Colak 2007; Li et al. 2007; Colak & Qahwaji 2009; Bloomfield et al. 2012). Such predictors require manual intervention before entering the prediction process, and therefore are not suitable for automatic operations. There are some other flare-prediction studies adopting the measures deduced from longitudinal magnetic fields (e.g., Georgoulis & Rust 2007; Yu et al. 2009; Song et al. 2009; Mason & Hoeksema 2010; Yuan et al. 2010; Ahmed et al. 2013).

The accumulation of magnetic nonpotential energy is of importance for solar eruptions. Mason & Hoeksema (2010) mentioned the importance of the vector-field data to obtain the most promising flare-predictive magnetic parameters. Leka and Barnes (2007) contributed a great amount to the exploration of the differences of magnetic-field properties between flare-imminent and flare-quiet active regions, however, the number or the time spans of their samples were quite restricted. Lacking long-term consistent observations of vector magnetic fields, the magnetic nonpotentiality was rarely used in solar flare predictions. The vector magnetograms obtained at the Huairou Solar Observing Station over more than 20 yr make the experiments possible. Yang et al. (2012, hereafter Paper I) have calculated the statistical relations between magnetic nonpotentiality and solar flares. By means of the prediction experiments described in this Letter, we can predict the occurrence of flares in particular active regions based on their magnetic properties alone, and also can estimate the starting time and eruption magnitude of the flares. In addition, several classical verification measures of dichotomous predictions are discussed to call for more serious concerns on the verification issue (Doswell et al. 1990). The Heidke Skill Scores (HSS) and the True Skill Statistics (TSS; see Section 3.2) of our 100 group experiments are in the ranges 0.23–0.48 and 0.32–0.82, respectively. Our results show that the nonpotentiality predictors improve the performance of predicting more powerful flares.

2. DATA AND METHOD

2.1. Data and Preprocessing

We use the observational data of photospheric active region vector magnetograms obtained by the Solar Magnetic Field Telescope (SMFT; Ai & Hu 1986) at the Huairou Solar Observing Station, National Astronomical Observatories of China. SMFT is a 35 cm aperture vector magnetograph with a tunable birefringent filter. The working spectral line for the vector magnetograms is Fe 1 5324.19, which is a strong and broad line with an equivalent width of about 0.334 Å and a Landé factor of 1.5 (Ai et al. 1982). The data employed are selected from all the vector magnetograms during the period from 1988 to 2008 subject to the following criteria: (1) the active regions are located within 30° from the solar disk center, and (2) only one magnetogram is used for each active region in one observation day. The final data set, which is also used in Paper I, consists...
of 2173 photospheric vector magnetograms involving 1106 active regions. The detailed descriptions of the data and their distributions during the two solar cycles are in Paper I, as well as the calibration for the vector magnetograms and the determination of the 180° ambiguity of the transverse field. The records of soft X-ray flares are available from NOAA’s National Geophysical Data Center.\textsuperscript{4}

2.2. Magnetic Nonpotentiality Parameters as Predictors

The magnetic nonpotentiality parameters as predictors, the inputs for the prediction model, are the mean planar magnetic shear angle $\Delta \phi$, mean shear angle of the vector magnetic field $\Delta \psi$, mean absolute vertical current density $J_z$, mean free magnetic energy density $E_m$, and mean density of longitudinal magnetic field $d_{bL}$, longitudinal-field weighted effective distance $d_{Em}$ (Paper I), mean horizontal gradient of the longitudinal field $\nabla hB_z$, maximum horizontal gradient ($V_h B_{zm}$), length of strong-gradient ($>0.05$ G km\(^{-1}\)) inversion lines $L_{gnt}$, and mean density of longitudinal magnetic energy dissipation $\epsilon(B_z)$ (Cui et al. 2006; Jing et al. 2006). All of the above measures are macroscopic and averaged quantities, which indicate the magnetic nonpotentiality or magnetic complexity of a whole active region. In the calculations, each magnetogram is represented as $(x_i, y_i)$, where $x_i \in \mathbb{R}^n$ is the predictor array and $y_i \in \{1, -1\}$ is the class label of the magnetogram ($y_i = 1$ for flaring instances and $y_i = -1$ for non-flaring ones, according to the labeling scheme stated in Section 3.1).

We have tried five combinations of predictors:

- V06 $(\Delta \psi, |J_z|, |hB_z|, |\alpha_{av}|, \rho_{free}, d_{Em}),$
- V08 $(\Delta \phi, \Delta \psi, |J_z|, |hB_z|, |\alpha_{av}|, \rho_{free}, d_{Em}, d_{Em}),$
- L05 $(d_{Em}, \nabla hB_z, (V_h B_{zm})_m, L_{gnt}, \epsilon(B_z)),$
- A10 $(\Delta \psi, |J_z|, |hB_z|, |\alpha_{av}|, \rho_{free}, d_{Em}, \nabla hB_z, (V_h B_{zm})_m, L_{gnt}, \epsilon(B_z)),$
- A12 $(\Delta \phi, \Delta \psi, |J_z|, |hB_z|, |\alpha_{av}|, \rho_{free}, d_{Em}, \nabla hB_z, (V_h B_{zm})_m, L_{gnt}, \epsilon(B_z)).$

2.3. Prediction Method: Support Vector Classification

Predicting whether or not an active region will flare within a certain time interval can be transformed into a classification problem. The support vector machine (SVM) first introduced by Vapnik (Boser et al. 1992; Cortes & Vapnik 1995; Vapnik 1995) is now a widely applied statistical learning theory used to solve classification and regression problems. In recent years, SVM has been applied to the field of astronomy (e.g., Zhang & Zhao 2003; Woźniak et al. 2004; Wadadekar 2005; Gao et al. 2008; Beaumont et al. 2011; Peng et al. 2012) including solar physics (e.g., Qu et al. 2003; Qahwaji & Colak 2007; Li et al. 2007; Al-Omari et al. 2010; Labrosse et al. 2010; Alipour et al. 2012). A machine learning system for classification is able to learn and construct a model (from the existing training data with definite category labels) which can classify the training data and predict upcoming ones whose categories are unknown. The maximum margin principle and the kernel function are the two core concepts of the SVM. By solving an optimization problem, the SVM classifier is obtained as an optimal separating hyperplane $w \cdot x + b = 0$ that separates the two-class data with the maximum distance. When in a linearly non-separable case, a kernel function is employed, then the training vectors $x_i$ are mapped into a higher-dimensional feature space in which the data can be linearly separated.

The primal optimization problem can be written as

$$\min_{w \in \mathbb{H}, b \in \mathbb{R}, \xi \in \mathbb{R}^l} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i,$$

subject to $y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, \ldots, l,$

$$\xi_i \geq 0, \quad i = 1, \ldots, l.$$

$C > 0$ is the penalty parameter for the sum of slack variables $\xi_i$. $(1/2)\|w\|^2$ corresponds to the distance maximization of the two classes. Taking the reciprocal, the square, and the factor 1/2 are for mathematical convenience. $\phi(x_i)$ denote the training vectors in the higher-dimensional space after employing the kernel function. The kernel function is denoted by $K(x_i, x_j)$, and the corresponding dual optimization problem, which is easier to solve, is

$$\min_a \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j K(x_i, x_j) - \sum_{j=1}^l a_j,$$

subject to $\sum_{i=1}^l y_i a_i = 0,$

$$0 \leq a_i \leq C, \quad i = 1, \ldots, l,$$

where $a_i$ are the Lagrange multipliers. Then the coefficients $a_i^*$ for the optimal hyperplane are solved from the dual problem. The training vectors $x_i$ with $a_i^* \neq 0$ are the support vectors that contribute to the final discriminant function

$$f(x) = \text{sgn}(w^* \cdot \phi(x) + b^*) = \text{sgn}\left(\sum_{i=1}^l a_i^* y_i K(x_i, x) + b^*\right),$$

where $w^*$ and $b^*$ are the corresponding solutions of the primal problem ($b^* = y_j - \sum_{i=1}^l y_i a_i^* K(x_i, x_j)$ taking any $0 < b^* < C$). The plus and minus signs of $f(x)$ indicate the two different classes.

There are a few commonly used kernels like the polynomial kernel, Gaussian radial basis kernel, sigmoid kernel, etc. After trying several kernels in the calculations, we accept the Gaussian kernel function. The kernel function is denoted by $K(x_i, x_j)$, and the corresponding dual optimization problem, which is easier to solve, is

$$\min_a \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j K(x_i, x_j) - \sum_{j=1}^l a_j,$$

subject to $\sum_{i=1}^l y_i a_i = 0,$

$$0 \leq a_i \leq C, \quad i = 1, \ldots, l,$$

where $a_i$ are the Lagrange multipliers. Then the coefficients $a_i^*$ for the optimal hyperplane are solved from the dual problem. The training vectors $x_i$ with $a_i^* \neq 0$ are the support vectors that contribute to the final discriminant function

$$f(x) = \text{sgn}(w^* \cdot \phi(x) + b^*) = \text{sgn}\left(\sum_{i=1}^l a_i^* y_i K(x_i, x) + b^*\right),$$

where $w^*$ and $b^*$ are the corresponding solutions of the primal problem ($b^* = y_j - \sum_{i=1}^l y_i a_i^* K(x_i, x_j)$ taking any $0 < b^* < C$). The plus and minus signs of $f(x)$ indicate the two different classes.

There are a few commonly used kernels like the polynomial kernel, Gaussian radial basis kernel, sigmoid kernel, etc. After trying several kernels in the calculations, we accept the Gaussian radial basis kernel, the mathematical expression of which is $K(x_i, x_j) = \exp(-||x_i - x_j||^2/\sigma^2)$, where $\sigma$ is a kernel parameter. The SVM software LIBSVM (Chang & Lin 2011) is used in our experiments.

Note that this classification or prediction model is based on statistical relations with no obvious physical meanings; nevertheless, the physical parameters closely related to solar flares must make positive effects to the performance of the model. This is exactly why we adopt magnetic nonpotentiality and complexity parameters as predictors.

3. EXPERIMENTS AND RESULTS

3.1. Experiment Design

According to whether the active regions produce flares exceeding a specified class within a certain time window, every magnetogram is labeled as positive (flaring) or negative (non-flaring). The “time window” in this Letter begins at

\textsuperscript{4} ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/
Table 1
Flaring (f) and Non-flaring (n-f) Sample Distributions

| Time Window | Category | Flaring Threshold |
|-------------|----------|------------------|
|             | C1.0     | C5.0             |
| 48 hr f     | 918      | 427              |
|             | 1255     | 1746             |
| 24 hr f     | 697      | 291              |
|             | 1476     | 1882             |
| 12 hr f     | 475      | 181              |
|             | 1698     | 1992             |
| 6 hr f      | 309      | 101              |
|             | 1864     | 2072             |

Table 2
Definition of the 2 × 2 Contingency Table (Confusion Matrix)

| Observed | Predicted |
|----------|-----------|
| Yes      | x         | y               |
| No       | z         | w               |

F<sub>1</sub>-measure, HSS, TSS, Critical Success Index (CSI), Gilbert Skill Score (GSS), and Clayton Skill Score (CSS), a summary of which is shown in Table 3. The perfect prediction, which is difficult to achieve in practice, corresponds to these verification measures reaching their upper bounds of 1. Though it has been more than a century since the “Finley affair” (see Murphy 1996) induced hot discussions, the study on this seemingly simple 2 × 2 problem remains ongoing (Stephenson 2000). In this work, we only consider the classical verification measures which are more intuitive to utilize in practical operations.

The percentage of correct predictions (x + w)/N (referred as ACC hereafter) is the simplest but often misleading measure to assess the prediction, especially when one side, event or non-event, is overwhelming. ACC, CSI, and F<sub>1</sub> do not exclude the correct numbers based on the stochastic prediction. The so-called skill scores indicate the relative accuracy of a prediction to some standard reference predictions. The generic form of skill score is

$$\text{SS} = \frac{S - S_{\text{ref}}}{S_{\text{perfect}} - S_{\text{ref}}} \times 100\%,$$

where S is a particular measure of accuracy, S<sub>ref</sub> a reference, and S<sub>perfect</sub> the perfect prediction. A no-skill prediction scores 0, a positive score shows a better prediction than the reference, and the perfect prediction scores 1. HSS is a skill score from ACC comparing with the random prediction. GSS is the skill-modified CSI, subtracting the expected correct predictions due to chance from x. F<sub>1</sub> is the harmonic mean of POD and FOH, and HSS happens to be the harmonic mean of skill-modified POD and skill-modified FOH (POD, and RS, in Schaefer 1990). The skill-modified ones are always lower than the original ones.

These verification measures are related to each other through the connections of x, y, z, and w. A common property of HSS, GSS, TSS, and CSS is that they all have the factor (x · w − y · z) in their numerators. This factor becomes zero in the random prediction, and thus these four skill scores all have the value 0, indicating no skill. In the constant prediction (all positive predictions, y = w = 0; or all negative, x = z = 0), this factor is also zero; CSS is meaningless in this case. The values of CSI, F<sub>1</sub>, and ACC in random situations depend on the ratio of events to non-events. Another common property of the above four skill scores is that they are all fair to both events and non-events. Considering non-events as focus, swapping x with w and y with z simultaneously, they remain unchanged; this is not the case with CSI or F<sub>1</sub>.

Keeping the numerators of the above four skill scores exactly the same, the differences of their denominators are

$$D_{\text{HSS}} - D_{\text{TSS}} = \frac{1}{2}(y - z)(y - z + x - w),$$
$$D_{\text{TSS}} - D_{\text{CSS}} = (z - y)(x - w),$$
$$D_{\text{GSS}} - D_{\text{HSS}} = \frac{1}{2}(y + z)(x + y + z + w),$$
$$D_{\text{GSS}} - D_{\text{TSS}} = y(x + y) + z(z + w).$$
introductions to the contingency table and forecast verification (Bloomfield et al. 2012). Accordingly, we compute the geometric mean of several verification measures (POD, FOH, TSS, HSS, etc.). It seems that the predictors from longitudinal fields are less sensitive in predicting large flares compared with other cases and the predictors derived from longitudinal fields perform well in the flare prediction, however, they may be less sensitive than the measures from vector fields in predicting large flares. The introduction of the prediction results. Even the unbiased TSS, which is independent of the event frequency, fails to effectively deal with rare event predictions (Doswell et al. 1990). TSS approaches POD in rare event situations, so both w and z contribute little to the results. Experientially, z rises if x’s proportion is increased. The bias ((x + z)/(x + y) ≠ 1) may be unintentionally introduced in optimizing a verification measure (Manzano 2005). Pursuing higher POD or TSS will cause higher FAR and lower FOH. Fewer misses cost more false alarms, but “crying wolf” may be undesirable. Moreover, the same TSS does not mean the same prediction performance. For instance, Table 5 lists some examples from Woodcock (1976). The prediction P1 has POD = 75% and PON = 50%, and P2 has POD = 50% and PON = 75%. TSS remains the same in the two cases and two predictions, but the results are indeed different. Therefore, only one measure might mislead the prediction verification, and multiple verification measures are probably acceptable. This point of view is as well mentioned in Schaefer (1990), Doswell et al. (1990), Marzban (1998), etc. We believe that, since each data set may have its own intrinsic properties, it is inappropriate to compare different predictions on different trial samples.

### Table 3

| VM | Derivation | Formulation | w-Dominated | Range |
|----|------------|-------------|-------------|-------|
| GSS | GSS = \(x - w^2\) \(\frac{x}{z} - 1\) | GSS = \(\frac{x}{w} + \sqrt{\frac{y}{z} - 1}\) | → CSI | \([-1/3, 1]\) |
| HSS | HSS = \(\frac{x}{w} + \sqrt{\frac{y}{z} - 1}\) | HSS = \(\frac{x}{w} + \sqrt{\frac{y}{z} - 1}\) | → F1 | \([-1, 1]\) |
| TSS | TSS = POD - FOH | TSS = POD = POD | → POD | \([-1, 1]\) |
| CSS | CSS = FOH - DFR | CSS = FOH - DFR | FOH | \([-1, 1]\) |
| CSF | CSF = \(\frac{y}{z} + \sqrt{\frac{w}{w} - 1}\) | CSI = CSI | CSI | \([0, 1]\) |
| F1 | \(F_1 = 2(POD^{-1} + FOH^{-1})^{-1}\) | \(F_1 = \frac{2(y + w) + w}{y + w}\) | \(F_1\) | \([0, 1]\) |

Notes.
- a Gilbert Skill Score (Gilbert 1884; see Schaefer 1990).
- b Heidke Skill Score (Doolittle 1888; Heidke 1926; see Woodcock 1976; Doswell et al. 1990).
- c True Skill Statistic, also called Peirce Skill Score or Hanssen-Kuipers’ discriminant (Peirce 1884; Hanssen & Kuipers 1965; see Woodcock 1976; Doswell et al. 1990).
- d Critical Success Index, also called threat score (Gilbert 1884; Donaldson et al. 1975; see Schaefer 1990).
- e Clayton Skill Score (Clayton 1934; see Wandelishin & Brooks 2002).
- f \(F_2\) measure, \(\beta = 1\) (Van Rijssbergen 1979; Chinchor 1992).
- g \(E_1 = (x + z)(x + y)/N\), the expected number of correct event predictions due to chance.
- h \(E_0 = (y + w)(z + w)/N\), the expected number of correct non-event predictions due to chance.

3.3. Experiment Results and More Comments on Verification

It is nearly impossible to optimize all the verification measures simultaneously (Manzano 2005; see also the results of Bloomfield et al. 2012). Accordingly, we compute the geometric mean of several verification measures (POD, FOH, TSS, HSS, GSS, CSI, F1, and \(\sqrt{POD \cdot FOH \cdot FOCN}\)) which we are more concerned about. A grid search process is carried out to obtain a relatively better pair of \((C, \sigma^2)\) for the final SVM classifier. The full version of the table available in the electronic version contains the V06, V08, L05, and A10 predictor results. A12’s results are shown in Table 4, in which each value with its error is the arithmetical mean of the specific verification measure in \(k\) times testing. The percentage of non-events (\(N_0/N\)) is given at the end of each row for reference. \(F_1\) is always higher than CSI, except when \(x = 0\) or \(y = z = 0\). In rare event situations, HSS is closer to \(F_1\), so HSS is likely higher than CSI. In our results, there are only two cases with HSS lower than CSI (C1.0, 48 hr; C1.0, 24 hr). These are the top two cases whose positive samples are in a larger proportion compared with other cases and \(w\) is not extremely dominant. The predictors derived from longitudinal magnetic fields (L05) perform somewhat better than those mainly involving transverse components (V06, V08) in predicting flares of \(\geq C1.0\) and \(\geq C5.0\). However, the superiority diminishes in predicting more powerful flares. For instance, in the case of \(\geq M5.0\) or \(\geq X1.0\) flares, the performance of longitudinal predictors becomes worse than that of other predictor combinations. It seems that the predictors from longitudinal fields are less sensitive in predicting large flares. Overall, there is an improvement in the prediction employing various measures derived from vector magnetic fields (A10, A12).

HSS and TSS are often discussed and applied in forecast verification (e.g., Woodcock 1976; Doswell et al. 1990; Manzano 2005). HSS = TSS when \(y = z\); HSS ≡ TSS when \(N_1 = N_0\). Bloomfield et al. (2012) proposed using TSS instead of HSS as a standard to reliably compare flare forecasts. However, no single scalar measure can cover all the information of the prediction results. Even the unbiased TSS, which is independent of the event frequency, fails to effectively deal with rare event predictions (Doswell et al. 1990). TSS approaches POD in rare event situations, so both \(w\) and \(z\) contribute little to the results. Experientially, \(z\) rises if \(x\)’s proportion is increased. The bias ((\(x + z\))/(\(x + y\)) ≠ 1) may be unintentionally introduced in optimizing a verification measure (Manzano 2005). Pursuing higher POD or TSS will cause higher FAR and lower FOH. Fewer misses cost more false alarms, but “crying wolf” may be undesirable. Moreover, the same TSS does not mean the same prediction performance. For instance, Table 5 lists some examples from Woodcock (1976). The prediction P1 has POD = 75% and PON = 50%, and P2 has POD = 50% and PON = 75%. TSS remains the same in the two cases and two predictions, but the results are indeed different. Therefore, only one measure might mislead the prediction verification, and multiple verification measures are probably acceptable. This point of view is as well mentioned in Schaefer (1990), Doswell et al. (1990), Marzban (1998), etc. We believe that, since each data set may have its own intrinsic properties, it is inappropriate to compare different predictions on different trial samples.

4. CONCLUSIONS AND DISCUSSIONS

Based on the long-term reliable observations of the photospheric vector magnetic fields by SMFT, we adopt some non-potentiality measures which are not available from observations of only line-of-sight magnetic fields to study the prediction of solar flares. Real-time processing and no manual intervention are two advantages of our prediction system. The data for the input of the prediction model are obtained by local observations, and the key measures as predictors are available without manual operations.

From our experiments, the combinations of magnetic measures derived from longitudinal fields perform well in the flare prediction, however, they may be less sensitive than the measures from vector fields in predicting large flares. The information...
| Flare Level | Time Window | POD      | FOH      | FOCN     | CSI      | $F_1$     | TSS      | CSS      | HSS      | GSS      | ACC      |
|------------|-------------|----------|----------|----------|----------|-----------|----------|----------|----------|----------|----------|
|            | 48 hr       | 0.707 ± 0.011 | 0.690 ± 0.013 | 0.782 ± 0.008 | 0.538 ± 0.013 | 0.698 ± 0.011 | 0.474 ± 0.020 | 0.472 ± 0.020 | 0.473 ± 0.020 | 0.312 ± 0.018 | 0.742 ± 0.010 | 0.578 ± 0.014 |
|            | 24 hr       | 0.677 ± 0.019 | 0.617 ± 0.013 | 0.840 ± 0.008 | 0.478 ± 0.016 | 0.645 ± 0.014 | 0.478 ± 0.022 | 0.458 ± 0.020 | 0.466 ± 0.021 | 0.306 ± 0.018 | 0.761 ± 0.010 | 0.679 ± 0.019 |
|            | 12 hr       | 0.653 ± 0.028 | 0.508 ± 0.010 | 0.895 ± 0.007 | 0.399 ± 0.013 | 0.569 ± 0.014 | 0.475 ± 0.024 | 0.404 ± 0.014 | 0.430 ± 0.016 | 0.275 ± 0.013 | 0.786 ± 0.006 | 0.781 ± 0.009 |
|            | 6 hr        | 0.560 ± 0.026 | 0.423 ± 0.022 | 0.923 ± 0.004 | 0.317 ± 0.017 | 0.479 ± 0.020 | 0.430 ± 0.026 | 0.346 ± 0.025 | 0.378 ± 0.024 | 0.235 ± 0.018 | 0.826 ± 0.009 | 0.858 ± 0.008 |
|            | 12 hr       | 0.485 ± 0.044 | 0.400 ± 0.028 | 0.952 ± 0.004 | 0.283 ± 0.028 | 0.435 ± 0.033 | 0.419 ± 0.044 | 0.352 ± 0.031 | 0.379 ± 0.035 | 0.239 ± 0.029 | 0.896 ± 0.006 | 0.917 ± 0.008 |
|            | 6 hr        | 0.595 ± 0.064 | 0.250 ± 0.029 | 0.979 ± 0.003 | 0.219 ± 0.029 | 0.351 ± 0.038 | 0.508 ± 0.065 | 0.229 ± 0.032 | 0.306 ± 0.041 | 0.187 ± 0.029 | 0.898 ± 0.006 | 0.954 ± 0.009 |
|            | 12 hr       | 0.554 ± 0.056 | 0.344 ± 0.021 | 0.979 ± 0.002 | 0.266 ± 0.025 | 0.415 ± 0.030 | 0.505 ± 0.052 | 0.323 ± 0.022 | 0.382 ± 0.031 | 0.240 ± 0.024 | 0.934 ± 0.004 | 0.956 ± 0.009 |
|            | 6 hr        | 0.523 ± 0.067 | 0.225 ± 0.028 | 0.986 ± 0.002 | 0.191 ± 0.028 | 0.312 ± 0.039 | 0.474 ± 0.067 | 0.211 ± 0.030 | 0.286 ± 0.040 | 0.173 ± 0.028 | 0.939 ± 0.004 | 0.973 ± 0.009 |
|            | 12 hr       | 0.460 ± 0.104 | 0.275 ± 0.067 | 0.994 ± 0.001 | 0.213 ± 0.056 | 0.338 ± 0.075 | 0.447 ± 0.105 | 0.269 ± 0.067 | 0.329 ± 0.076 | 0.207 ± 0.056 | 0.982 ± 0.002 | 0.990 ± 0.003 |
|            | 6 hr        | 0.667 ± 0.139 | 0.220 ± 0.038 | 0.998 ± 0.001 | 0.202 ± 0.042 | 0.327 ± 0.058 | 0.654 ± 0.139 | 0.218 ± 0.039 | 0.322 ± 0.059 | 0.198 ± 0.042 | 0.985 ± 0.002 | 0.994 ± 0.005 |
|            | 12 hr       | 0.533 ± 0.062 | 0.278 ± 0.023 | 0.996 ± 0.001 | 0.214 ± 0.010 | 0.353 ± 0.013 | 0.522 ± 0.061 | 0.274 ± 0.023 | 0.346 ± 0.013 | 0.210 ± 0.010 | 0.985 ± 0.002 | 0.990 ± 0.003 |
|            | 6 hr        | 0.700 ± 0.200 | 0.169 ± 0.055 | 0.999 ± 0.001 | 0.167 ± 0.056 | 0.270 ± 0.084 | 0.688 ± 0.199 | 0.167 ± 0.056 | 0.265 ± 0.084 | 0.164 ± 0.056 | 0.987 ± 0.003 | 0.996 ± 0.009 |

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)
mation of transverse fields makes a limited contribution to the prediction of low magnitude flares, but it does improve the prediction for large flares such as M5.0 and X1.0 ones. Thus, it is reasonable to include transverse field components in flare predictions.

To avoid misleading the optimization work or misusing the results from a single verification measure, prediction results should be assessed carefully. It is helpful to consider multiple verification measures. A step like k-fold cross-validation is necessary for improving the generalization capability of the prediction models. The intrinsic properties of various data sets may make a specific tool perform rather differently, and hence, it is then significant to make comparisons in the same data environment.

Some researchers have begun to use vector magnetograms from the Helioseismic Magnetic Imager (HMI) on board the Solar Dynamics Observatory to predict solar flares. Yet, the prediction methods founded on statistical information are restricted by the finite time span of HMI data at present. Results of statistical predictions depend on both the historical data set and prediction method employed. There is still a long way to go for the prediction of solar activities employing the exquisite HMI data.

The authors are very grateful to the anonymous referees for encouraging comments and beneficial suggestions that improved the manuscript. We are indebted to the HSOS staff and for encouraging comments and beneficial suggestions that improved the manuscript. We are indebted to the HSOS staff and X.Y. acknowledges helpful discussions with Dr. Shangbin Yang. This work is supported by the National Natural Science Foundation of China (60940030, 10921303, 41174153, 10903015, 11003025, and 11121063), the Young China (60940030, 10921303, 41174153, 10903015, 11003025, and 11121063), the Young China (60940030, 10921303, 41174153, 10903015, 11003025, and 11121063), the Young China (60940030, 10921303, 41174153, 10903015, 11003025, and 11121063), the Young China (60940030, 10921303, 41174153, 10903015, 11003025, and 11121063).