The Astrophysical Journal, 786:19 (13pp), 2014 May 1 doi:10.1088/0004-637X/786/1/19

HELIOSEISMOLOGY OF PRE-EMERGING ACTIVE REGIONS. III. STATISTICAL ANALYSIS

G. Barnes, A. C. Birch[1], K. D. Leka, and D. C. Braun
NorthWest Research Associates, 3380 Mitchell Lane, Boulder, CO 80301, USA; graham@nwra.com
Received 2013 July 7; accepted 2014 March 1; published 2014 April 10

ABSTRACT

The subsurface properties of active regions (ARs) prior to their appearance at the solar surface may shed light on the process of AR formation. Helioseismic holography has been applied to samples taken from two populations of regions on the Sun (pre-emergence and without emergence), each sample having over 100 members, that were selected to minimize systematic bias, as described in Paper I. Paper II showed that there are statistically significant signatures in the average helioseismic properties that precede the formation of an AR. This paper describes a more detailed analysis of the samples of pre-emergence regions and regions without emergence based on discriminant analysis. The property that is best able to distinguish the populations is found to be the surface magnetic field, even a day before the emergence time. However, after accounting for the correlations between the surface field and the quantities derived from helioseismology, there is still evidence of a helioseismic precursor to AR emergence that is present for at least a day prior to emergence, although the analysis presented cannot definitively determine the subsurface properties prior to emergence due to the small sample sizes.

Key words: methods: statistical – Sun: helioseismology – Sun: interior – Sun: magnetic fields – Sun: oscillations – sunspots

Online-only material: machine-readable tables

1. INTRODUCTION

Models for the formation of solar active regions (AR) tend to fall into one of several classes, largely dependent on the volume in which strong magnetic fields are generated. In one of these, magnetic flux tubes generated near the base of the convection zone become buoyant and rise through the convection zone (for a review, see Fan 2009), with an AR emerging when a flux tube passes through the solar surface. Another possibility is that ARs are formed as the result of the coalescence of magnetic fields generated in the bulk of the convection zone or near the solar surface (Brandenburg 2005 and references therein). Each of these scenarios has a distinct signature in the velocity of the plasma in the convection zone. Local helioseismology (Gizon & Birch 2005; Gizon et al. 2010) potentially can be used to determine the subsurface dynamics associated with AR formation and thus could provide evidence for or against either of these theories. This would also indirectly shed light on the location of the solar dynamo.

Previous studies of AR formation using local helioseismology have tended to focus on only a small number of regions (e.g., Braun 1995; Chang et al. 1999; Jensen et al. 2001; Zharkov & Thompson 2008; Kosovichev 2009; Hartlep et al. 2011; Ilonidis et al. 2011). The small number of regions considered and the lack of a control group of areas of the Sun where no AR was emerging make it difficult to identify any subsurface properties *unique* to the emergence of ARs. The exception to this is the study of Komm et al. (2009, 2011) that considered subsurface flows of a large sample of existing ARs undergoing episodes of magnetic flux emergence compared with a control group of ARs that had comparatively constant flux. However, this study did not include the pre-emergence stage of AR formation.

The present study is based on applying helioseismic holography (Lindsey & Braun 2000) to samples of more than 100 areas of the Sun where an AR subsequently emerged and an equal number where no AR emerged. The selection of these regions was described in Leka et al. (2013, hereafter Paper I), whereas an initial analysis of the travel times inferred from helioseismic holography, focusing on the average travel-time shifts, was presented in Birch et al. (2013, hereafter Paper II). It was found that there are statistically significant differences in the average travel times, as well as in the surface magnetic flux, between the samples of pre-emergence areas and quiet Sun areas. These included a reduction in the mean travel-time shift of a few tenths of a second as well as spatially antisymmetric features in both the east–west and north–south travel-time differences. The antisymmetric features are qualitatively consistent with what would be expected from a flow converging on the site of emergence, although it appears that it is not a simple converging flow. One possible interpretation is that emergence preferentially occurs at the boundaries between supergranules. In this scenario, the emergence is not at the center of a converging flow but between neighboring diverging flows. This could also account for the difference in the surface magnetic field because flux tends to concentrate in the boundaries between supergranules.

As interesting as what was found in Paper II is what was not found: any signature of a strong retrograde flow or any travel-time shifts greater than an order of 1 s. Simulations of rising flux tubes (e.g., Fan 2008) predict retrograde flows with magnitudes of order 100 m s^{-1}, whereas Ilonidis et al. (2011) found mean travel-time reductions of the order of 10 s. In both of these cases, the results are much larger effects than were found in Paper II but for layers in the Sun significantly below the roughly 20 Mm maximum depth considered here. Thus it may be that there are significant changes in the emergence process between depths of approximately 60 Mm and 20 Mm.

In the present paper, we briefly review the selection of the data and the analysis performed in the previous papers in this series before proceeding to an analysis of the data based on discriminant analysis. We use the full distribution of the travel-time shifts to determine the relative ability of

[1] Max-Planck-Institut für Sonnensystemforschung, Justus-von-Leibig-Weg 3, 37077 Göttingen, Germany.

Table 1
Sample Sizes

Time Interval	Time Before Emergence (hr)	NE	PE
TI-0	$t_0 - 24.5$	81	89
TI-1	$t_0 - 19.2$	85	88
TI-2	$t_0 - 13.9$	85	89
TI-3	$t_0 - 8.5$	82	87
TI-4	$t_0 - 3.2$	83	86

different parameters to discriminate between the samples of pre-emergence and non-emergence. We compare the ability of a measure of the surface magnetic field to distinguish the samples with the ability of the helioseismic parameters, and we examine how the surface field is influencing the helioseismic parameters. An important caveat is that here, as in Paper II, care should be taken in interpreting the nature of the holography travel-time shifts. For example, without modeling, the variation of depth of any flows or other perturbations producing the shifts is not known.

2. THE DATA AND HELIOSEISMIC ANALYSIS

The overall design of this study, including the data selection, preparation, and treatment, was presented in Paper I. In brief, samples from two populations are considered: "pre-emergence" targets (PE) that track a $32^\circ \times 32^\circ$ patch of the Sun prior to the emergence of a NOAA-numbered AR and "non-emergence" targets (NE) selected for lack of emergence and lack of strong fields in the central portions of the tracked patch. The PE sample size comprises 107 targets obtained between 2001 and 2007, matched to 107 NE targets drawn from an initially larger sample and selected further to match the PE distributions in time and observing location on the disk. The emergence time was determined using the Michelson Doppler Imager (MDI; Scherrer et al. 1995) 96 min cadence observations of the line-of-sight magnetic field. No selection was made for minimum size of the numbered NOAA regions that result, and limits were placed to avoid extreme observing angles.

Data for the helioseismology originate from the Global Oscillations Network Group (GONG) project (GONG; Harvey et al. 1998; Hill et al. 2003); a full GONG day (1664 min) of data prior to the emergence time was tracked and divided into five time intervals, each 6.4 hr long but starting every 5.3 hr with just over an hour overlap between them. Table 1 shows how many NE/PE had an acceptable duty cycle (>80%) for each time interval; once duty cycle is accounted for, the sample sizes are no longer equal.

We measured wave travel times from each of the time intervals of the GONG data using surface-focusing helioseismic holography (Lindsey & Braun 2000; Gizon & Birch 2005), a technique very similar to time–distance helioseismology (Duvall et al. 1993). In particular, the GONG Dopplergrams were first tracked and Postel projected, then phase speed filters were applied. The filters, described in Table 1 of Couvidat et al. (2005), isolate waves with particular ranges in lower turning points; these filters cover the range in lower turning-point depths from about 1.4 Mm (filter TD1) to about 23.3 Mm (filter TD11). The full list of depths is given in Table 1 of Paper II. After filtering, center-annulus and center-quadrant local-control correlations were used to measure travel-time shifts. From these, travel-time differences and proxies for the vertical component of the flow vorticity and the horizontal flow divergence were constructed. The result was, for each time interval and each region, a spatial map of the travel times listed in Table 2.

Table 2
Helioseismology and Magnetic Variables

Variable	Description
$\delta\tau_x$	East–west travel-time difference
$\delta\tau_y$	North–south travel-time difference
$\delta\tau_{\rm in}$	Annulus-to-center travel-time shift
$\delta\tau_{\rm out}$	Center-to-annulus travel-time shift
$\delta\tau_{\rm oi}$	"Out minus in" travel-time difference $\delta\tau_{\rm out} - \delta\tau_{\rm in}$
$\delta\tau_{\rm mn}$	Mean travel-time shift $[\delta\tau_{\rm out} + \delta\tau_{\rm in}]/2$
vor	Vertical component of vorticity: $\partial_x\delta\tau_y - \partial_y\delta\tau_x$
div	Horizontal flow divergence: $\partial_x\delta\tau_x + \partial_y\delta\tau_y$
B_r	Radial component of (potential) magnetic field

Note. All measures of the magnetic field are averaged over the corresponding time interval.

In order to reduce the spatial maps to a small number of parameters characterizing each region during each time interval, each travel-time map was spatially averaged over a 45.5 Mm (equal to 30 pixels in the Postel projected image) disk, centered at the emergence location for PE. This size disk just encompasses the signal seen in the ensemble averages presented in Figure 5 of Paper II, thus it is likely to capture any significant differences in the travel-time maps.

As in Paper II, three weightings were used in the averaging: a uniform weighting and $\sin\theta$ and $\cos\theta$ weightings, where θ is the angle measured counterclockwise from the direction of solar rotation (the $+\hat{\mathbf{x}}$ direction). To be consistent with Paper II, we will continue to denote the spatial average with an overline but note that only spatial averages are considered here. Using this combination of weighting factors makes the analysis sensitive to both spatially symmetric and antisymmetric features in the travel-time maps. The uniform weight average will be sensitive to spatially symmetric features, whereas both the $\sin\theta$ and $\cos\theta$ weightings will result in a zero average. Conversely, any spatially antisymmetric feature will result in a zero average with uniform weight but will have a signal in the $\sin\theta$ weighted average if it is antisymmetric about an east–west line passing through the center of the averaging disk, or a signal in the $\cos\theta$ weighted average if it is antisymmetric about a north–south line. A feature with a more complicated distribution could manifest in several of the weighted averages.

For example, the east–west travel-time map averaged with uniform weight will be sensitive to a spatially constant (symmetric) prograde or retrograde flow, but such a flow will not have a signature in the $\sin\theta$ and $\cos\theta$ weightings. A flow converging on or diverging from the center of the averaging disk would be antisymmetric about a north–south line in the $\delta\tau_x$ map and would have a signature in the $\cos\theta$ weighted average but not the other averages of $\delta\tau_x$, whereas a shear flow in which the magnitude of the east–west flow changes from north to south, passing through zero at the center of the averaging disk, would be antisymmetric about an east–west line in the $\delta\tau_x$ map and would have a signature in the $\sin\theta$ weighted average but not the other averages of $\delta\tau_x$.

These same flows would also produce signatures in other variables, so the converging flow would also have a signature in the uniform weight average of $\delta\tau_{oi}$, for example, whereas the shear flow would have a signature in the uniform weight average of vor. However, if the shear flow consists of only an east–west flow, the signature in the vertical component of the vorticity

would be weaker than in the east–west travel-time difference because there would be no contribution to the vorticity from a north–south flow.

Each combination of travel-time map and weighting is most sensitive to a different geometry of flow or sound speed perturbation. Because the goal of the present work is to search for *any* possible signatures in the helioseismic measurements from Paper II, all of these measures are included in the analysis.

The accompanying magnetic data derive from MDI observations: a potential field was calculated that matches the observed line-of-sight component provided by MDI, hence providing the potential field approximation of the radial field present over the course of the GONG data, for comparison with the results of the helioseismology. The absolute value of the radial magnetic field was spatially averaged over the same 45.5 Mm disk as the travel times and temporally averaged over each time interval. Note that after accounting for duty cycle, the magnetic field variables have different sample sizes than the seismology variables because they were computed from MDI data; only time interval 0 has a sample size not equal to 107, where the NE sample is reduced to 106.

3. STATISTICAL TESTS: DISCRIMINANT ANALYSIS AND SKILL SCORES

The analysis presented in Paper II suggests that helioseismic holography is able to detect a signature prior to the emergence time as defined in Paper I. To quantify this ability, and in particular to determine whether there is any more information available from the holographic signatures than there is from direct measurements of the surface magnetic field, discriminant analysis (e.g., Kendall et al. 1983) was used. This technique classifies a measurement as belonging to the group with the highest probability density. Provided the probability density is estimated accurately, it maximizes the overall rate of correct classification. In this case, the two groups are the PE and NE regions, and a region would be classified as emerging whenever the probability density estimate for the emerging regions exceeds the probability density estimate for quiet regions, for the specified property of the new region.

For the results presented here, the probability density was estimated using a kernel method with the Epanechnikov kernel and the smoothing parameter set based on its optimum value for a normal distribution (Silverman 1986). For the average unsigned flux, which is a positive definite quantity, the probability density of the logarithm was estimated. This ensures that the density estimate is zero for values of the flux less than zero and better captures the typical tail to high values. Example density estimates are presented in Section 4.

At any randomly selected point on the solar disk, the probability of an AR emerging in a one-day window is extremely small. However, for this analysis, the prior probabilities for pre-emergence and non-emergence were set equal. Thus, we are not truly testing the ability of the parameters to predict the emergence of an AR, but rather we are testing whether there is a signal of emergence in the seismic analysis. The advantage to this choice is that it avoids the problem that the presence of even quite a strong signal can be masked by an extremely small prior probability.

The first step in quantifying the performance of the discriminant analysis was to construct a contingency table, as shown in Table 3. From a contingency table, there are many ways to quantify the performance of a classification scheme. Because prior probabilities are assumed to be equal, but the sample sizes are not equal after accounting for the duty cycle (see Section 2 and Table 1), we use the Peirce skill score (Peirce 1884), also known as the true skill score or Hanssen and Kuipers's discriminant (see Woodcock 1976, for a comparison of this with other skill scores). It is given by

Table 3
Contingency Table

		Classified	
		PE	NE
Observed	PE	n_{pp}	n_{np}
	NE	n_{pn}	n_{nn}

$$\mathrm{PSS} = \frac{n_{pp}}{n_p} - \frac{n_{np}}{n_n}, \tag{1}$$

where n_{pp} is the number of regions that were classified by the discriminant analysis to be emergences and did emerge, n_p is the number of PE regions, n_{np} is the number of regions that were classified by the discriminant analysis to be non-emergences but did emerge, and n_n is the number of NE regions. As expressed above, the Peirce skill score is the probability of detection (hit rate) minus the probability of false detection (false alarm rate). Changing the sample size of events or nonevents does not change this score provided the rates have been accurately estimated. Positive values of this skill score indicate improvement of the forecasts over both uniform (unskilled) forecasts and random forecasts (Woodcock 1976), with a maximum score of 1.0 for perfect forecasting, whereas negative scores indicate worse performance than uniform or random forecasts.

To further confirm the independence of the results on the varying sample sizes, the analysis was repeated using only the subset of regions that have good duty cycles for all time intervals. Sample results of this investigation are shown in Appendix A. The main result of using this subset is to increase the uncertainty estimates, which is a consequence of the sample sizes being substantially reduced to 45/48 for NE/PE; the significance of the results is thus reduced, but the interpretations remain the same.

An unbiased estimate of the skill score with an error estimate was obtained by using cross-validation (Hills 1966) and a bootstrap approach (e.g., Efron & Gong 1983). For each sample (PE and NE), a bootstrap sample was constructed by drawing with replacement from the full sample. That is, a PE (NE) region was selected at random from the full set of PE (NE) regions, and this was repeated n_p (n_n) times to construct a bootstrap sample. Because all the draws are from the full sample, the same region may be drawn more than once or not at all.

To remove bias, each member of the bootstrap sample was classified by using the remaining $n - 1$ points to determine the probability density at the removed point, and repeating for all n points in the sample, from which a contingency table and skill score were constructed. This was repeated for 1000 bootstrap samples, with the mean and standard deviation of the resulting skill scores used to estimate the skill score value and error.

4. RESULTS

To determine the variables with the greatest ability to distinguish between PE and NE regions, the unbiased estimate of the Peirce skill score and its uncertainty from nonparametric discriminant analysis were computed for all of the variables in Table 2 for each time interval and phase speed filter (except the

Table 4
Best-Performing Variables

Variable	Filter	Depth (Mm)	Peirce SS
TI-0: time $= t_0 - 24.5$ hr			
$\overline{\lvert B_r \rvert}$	...	0.0	0.38 ± 0.07
$\overline{\delta\tau_y \sin\theta}$	TD2	2.2	0.34 ± 0.08
$\overline{\delta\tau_y \sin\theta}$	TD3	3.2	0.32 ± 0.08
$\overline{\mathrm{vor} \sin\theta}$	TD8	15.7	0.29 ± 0.09
$\overline{\mathrm{vor} \sin\theta}$	TD4	6.2	0.27 ± 0.09
$\overline{\delta\tau_{\mathrm{out}}}$	TD5	9.5	0.26 ± 0.08
$\overline{\delta\tau_x \cos\theta}$	TD4	6.2	0.26 ± 0.08
TI-1: time $= t_0 - 19.2$ hr			
$\overline{\lvert B_r \rvert}$	...	0.0	0.38 ± 0.07
$\overline{\delta\tau_y \sin\theta}$	TD2	2.2	0.29 ± 0.08
$\overline{\delta\tau_{\mathrm{io}}}$	TD6	11.4	0.28 ± 0.08
$\overline{\delta\tau_{\mathrm{in}} \cos\theta}$	TD5	9.5	0.28 ± 0.09
$\overline{\delta\tau_x \cos\theta}$	TD3	3.2	0.26 ± 0.08

Notes. Depth refers to the lower turning point of the waves in the filter used. Time is relative to the emergence time t_0. Shaded variables also appear in Table 5 and have significant ability to discriminate the populations after controlling for $\overline{\lvert B_r \rvert}$, as discussed in Section 4.5.

(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.)

magnetic field, for which no filters were used). Table 4 lists all of the variables with a skill score of more than 0.27 for time intervals 0 and 1.[2] The results of a Monte Carlo experiment (Appendix B) show that it is very likely that a variable with a skill score of greater than 0.27 can truly differentiate between the populations. This cutoff is arbitrary in the sense that there are variables just below the cutoff that have essentially the same ability to discriminate NE from PE regions as variables that appear in the table. However, skill scores up to at least 0.2 can reasonably be expected due to chance for variables that have no difference between the populations. Thus the threshold chosen means that the variables in the table are ones that are very likely to have a real ability to discriminate; others with real ability may be excluded.

In each time interval, the variable that is best able to distinguish the PE from the NE regions is the average unsigned field, $\overline{\lvert B_r \rvert}$. In the time interval immediately prior to emergence (centered 3.2 hr before the emergence time), the mean travel-time shift in a variety of filters shows significant ability to distinguish PE from NE, as do the center-to-annulus and annulus-to-center travel times. In almost all of the time intervals, antisymmetric-weighted averages of the east–west ($\overline{\delta\tau_x \cos\theta}$) and the north–south ($\overline{\delta\tau_y \sin\theta}$) travel-time differences measured in filters with shallow lower turning points appear. These same measures are highlighted in Figures 4 and 5 of Paper II and are interpreted as being consistent with a converging flow. The other variable that appears in multiple time intervals is $\overline{\mathrm{vor} \sin\theta}$ in filters with a moderate-depth lower turning point. Several other variable and filter combinations appear in only one time interval, such as the difference between the center-to-annulus and the annulus-to-center travel time, $\overline{\delta\tau_{\mathrm{oi}}}$, at moderate depth, and an antisymmetric average of the annulus-to-center travel time, $\overline{\delta\tau_{\mathrm{in}} \cos\theta}$, at moderate depth.

[2] Skill scores for all of the variables are available as a machine-readable table in the online edition of this paper.

4.1. The Average Unsigned Magnetic Field

The left panel of Figure 1 shows the probability density estimates for the mean unsigned field strength $\overline{\lvert B_r \rvert}$. The peak of the PE distribution is at a slightly higher field strength and has a substantially longer tail to large values, leading to a substantially higher mean value for the PE sample than for the NE sample. In this case, the large separation of the means is misleading because of the presence of a few strong field PE regions, whereas the distributions of PE and NE for $\overline{\lvert B_r \rvert}$ show considerable overlap. The Peirce skill score for this variable is 0.49 ± 0.06. The discriminant boundary falls at $\overline{\lvert B_r \rvert} \approx 13$ G; regions with a stronger average unsigned field strength would be classified as PE.

The right panel of Figure 1 shows the performance of $\overline{\lvert B_r \rvert}$ as a function of time prior to emergence. There is perhaps an increase in the skill score from approximately 1 day before emergence to a few hours before emergence, but the overall increase is not large in magnitude. Certainly between 24.5 hr and 8.5 hr before emergence, the variations in the skill score are less than the uncertainty; there is a small increase in the last point, 3.2 hr prior to emergence, which is likely to be a result of an incorrect emergence time for some regions, so that surface field is appearing during the final time interval (see Figure 11 of Paper I).

This constancy in the performance of the unsigned field is largely because the field itself does not evolve substantially over the time in question. Figure 2 (left) shows the field 24.5 hr before emergence versus the field 8.5 hr before emergence. There is an extremely high correlation between the field at the two times (Pearson correlation coefficients $\geqslant$0.96) for both the PE and NE regions and no clear indication of evolution. When considering the change between 24.5 hr and 3.2 hr before emergence (Figure 2, right), there is some indication of evolution of the field, consistent with there being a few regions for which emergence began in the final 6 hr before the nominal emergence time (more points lying above and to the left of the blue line). This trend is still weak compared to the variation among the regions considered, as is borne out by the small decrease in the correlations (Pearson correlation coefficients $\geqslant$0.94). Thus the evolution of the field does not greatly change the ability of the unsigned field to distinguish the PE from the NE.

The large overlap between the distributions of NE and PE regions shows that there is no clear signature in the surface field when individual regions are considered (see Figures 2 and 3 of Paper II). However, there was a bias introduced in the selection criteria for the NE compared with the PE: NE regions were required to have a magnetic field consistently $<$1000 G (see Section 3.2 of Paper I), whereas no such requirement was imposed for PE regions. It may simply be that the difference between the PE and NE samples results from the bias in the selection of NE compared to the PE regions.

It is also possible that there is a small amount of weak magnetic flux present at the surface more than a day prior to the beginning of the clear emergence phase of ARs. This field is indistinguishable from noise in individual MDI magnetograms but becomes apparent in averaging over large numbers. This could be related to the emergence process, in the form of small amounts of flux arriving at the emergence site prior to the main emergence, as seen in some simulations (e.g., Cheung et al. 2010; Stein et al. 2011). It could also be related to the known tendency for ARs to emerge in the same locations as prior ARs (e.g., Pojoga & Cudnik 2002). The latter case would be one

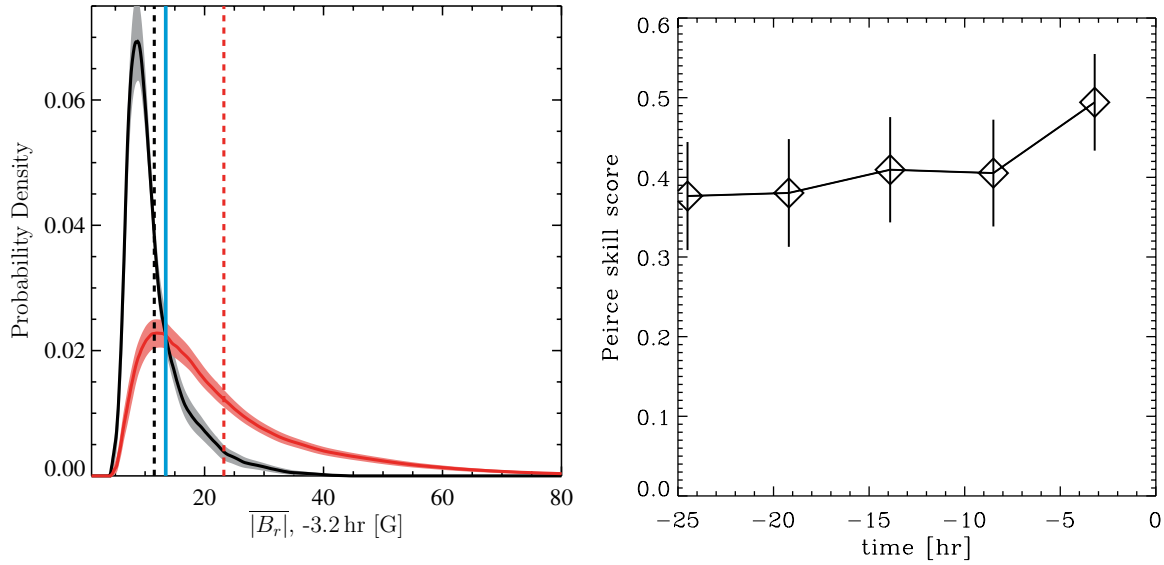

Figure 1. Nonparametric discriminant analysis for the mean unsigned magnetic flux, $\overline{|B_r|}$. Left: probability density estimates for the NE regions (black), and the PE regions (red) in time interval 4, centered 3.2 hr before the emergence time, with the mean of each sample indicated by a vertical dashed line in the corresponding color. The shaded region is a 1σ estimate of the uncertainty. The discriminant boundary, where the two probability density estimates are equal, is indicated by a vertical blue line; an observation to the right of the boundary would be classified as an emergence. There is an obvious difference between the density estimates for the NE and PE regions: the PE distribution has a longer tail to high field strength, but there is also considerable overlap of the distributions. Right: evolution of the Peirce skill score. There is perhaps a weak increase in the performance of $\overline{|B_r|}$ closer to the emergence time that is likely due to the onset of emergence during time interval 4 for some regions, but for most of the time intervals shown, the skill score is approximately constant.

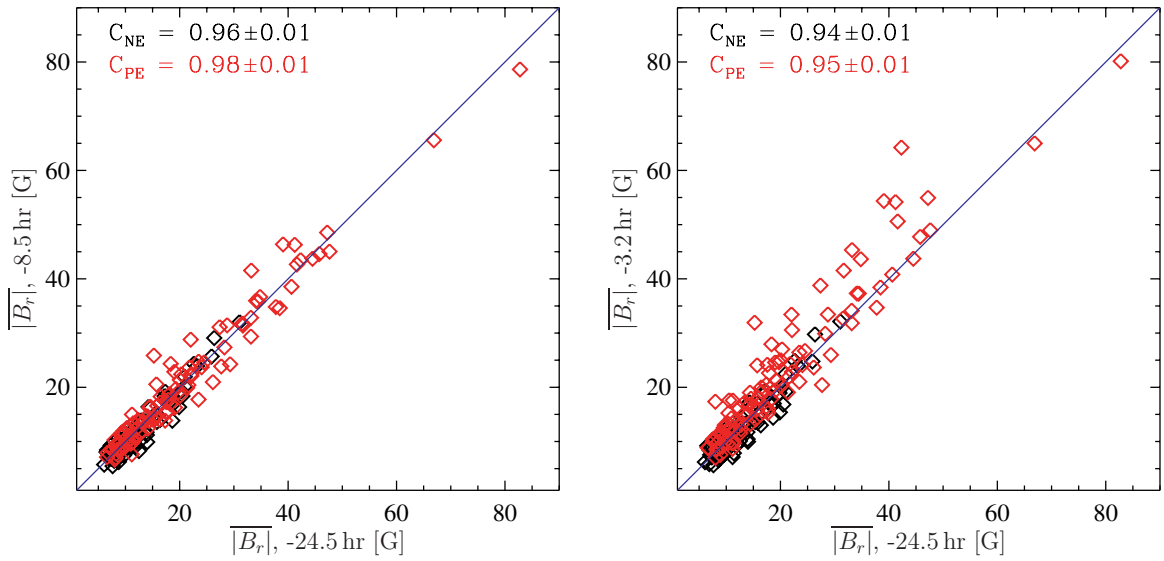

Figure 2. Left: scatter plot of $\overline{|B_r|}$ for the time interval centered 24.5 hr before emergence vs. the time interval centered 8.5 hr before emergence for NE (black) and PE (red) regions. The Pearson correlation coefficient between the PE at these times is 0.98 ± 0.01, whereas for the NE it is 0.96 ± 0.01. Right: same scatter plot but between 24.5 hr and 3.2 hr before emergence. The correlation coefficients are both slightly lower at 0.95 ± 0.01 for PE and 0.94 ± 0.01 for NE, suggesting that much of the evolution of the flux occurs between 8.5 hr and 3.2 hr before emergence, which is consistent with the emergence process beginning (for some regions) during the final 6 hr before the time of emergence.

example of how the bias manifests from a completely solar cause.

4.2. Measures of the Center-to-Annulus and Annulus-to-Center Travel Times

Immediately prior to emergence, the mean travel-time shift measured in a variety of filters shows a significant ability to distinguish PE from NE regions. The left panel of Figure 3 shows the mean travel time, $\overline{\delta\tau_{\rm mn}}$, in filter TD5 with a lower turning point depth of 9.5 Mm for time interval 4 (centered 3.2 hr prior to the time of emergence). The NE sample has a mean very close to 0 s, and its distribution is symmetric and peaked close to 0 s, consistent with the differences from 0 s being simply due to noise. The mean of the PE sample is negative and close to the peak in its distribution; the PE distribution is slightly wider than the NE distribution. Although there is a distinct difference visible in the distributions, there is also substantial overlap between the two, as in the $\overline{|B_r|}$ case. This is quantified by a Peirce skill score of 0.45 ± 0.08. Regions with a mean travel-time shift less than the discriminant boundary (at approximately -0.2 s) would be classified as PE, whereas those above would

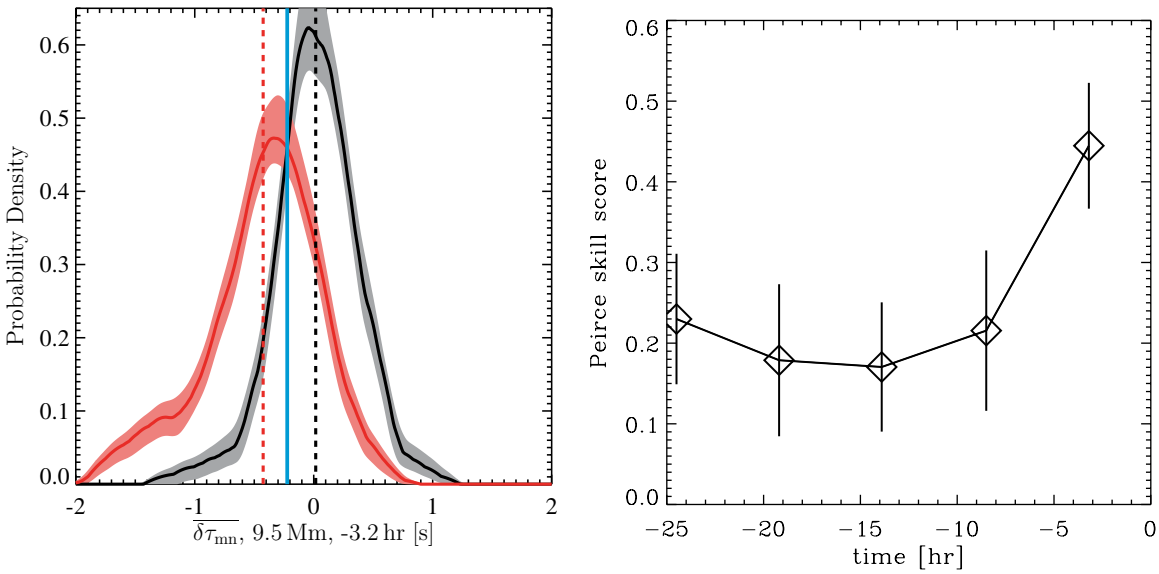

Figure 3. Nonparametric discriminant analysis for the mean travel-time shift, $\overline{\delta\tau_{\rm mn}}$, in filter TD5, in the same format as Figure 1. Left: there is an obvious difference between the density estimates for the NE and PE regions in the time interval centered 3.2 hr before the emergence time, with the PE regions typically having negative mean travel-time shifts. Right: the performance of $\overline{\delta\tau_{\rm mn}}$ shows a similar trend to $\overline{|B_r|}$, constant for most of the time considered, with an increase at the last time interval, although the skill score is consistently lower for $\overline{\delta\tau_{\rm mn}}$.

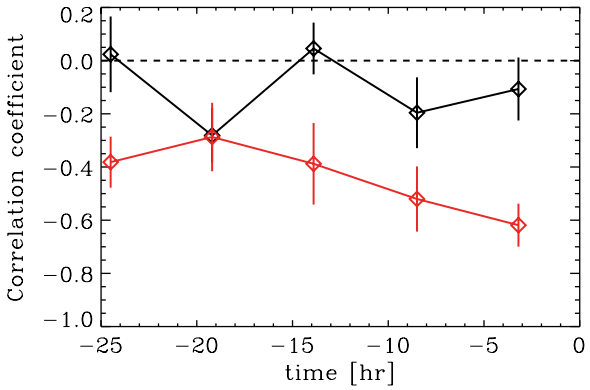

Figure 4. Pearson correlation coefficient between $\overline{|B_r|}$ and $\overline{\delta\tau_{\rm mn}}$ in filter TD5 from a bootstrap method for the NE (black) and PE (red) regions as a function of time. For the NE regions, the correlation coefficient is generally close to zero, as expected. For the PE regions, there is a weak trend toward stronger (negative) correlations at later times.

be classified as NE. That is, negative mean travel-time shifts are associated with emerging regions. Like the average unsigned field, the performance of the mean travel-time shift is better 3.2 hr before emergence than at earlier times (compare Figure 1, right, to Figure 3, right).

The mean travel-time shift is typically reduced in the presence of a surface magnetic field (e.g., Lindsey & Braun 2005; Braun & Birch 2008). Figure 4 shows the Pearson correlation coefficient between $\overline{\delta\tau_{\rm mn}}$ and $\overline{|B_r|}$ as a function of time. The correlation coefficient for the NE regions is generally close to 0, as would be expected if the mean travel-time shifts are simply due to noise, whereas for the PE regions the correlation coefficient is negative, with perhaps a weak trend toward a stronger (negative) correlation closer to the emergence time, although there is not a distinct difference between the final time interval prior to emergence and earlier time intervals. It is possible that the difference between the PE and NE regions in $\overline{\delta\tau_{\rm mn}}$ is simply an indirect result of the difference in the surface field. However, it is also possible that there is a signal in $\overline{\delta\tau_{\rm mn}}$ during all of the time intervals that is not a result of the surface magnetic field. The influence of the surface field on the travel times is investigated further in Section 4.5.

There are also many instances where $\overline{\delta\tau_{\rm in}}$ and $\overline{\delta\tau_{\rm out}}$ have a skill score only slightly less than $\overline{\delta\tau_{\rm mn}}$ in the same filter. In all time intervals and filters, there is a moderate correlation between $\overline{\delta\tau_{\rm in}}$ and $\overline{\delta\tau_{\rm out}}$. Because $\overline{\delta\tau_{\rm mn}}$ is a linear combination of $\overline{\delta\tau_{\rm in}}$ and $\overline{\delta\tau_{\rm out}}$, it is likely that the slightly better performance of $\overline{\delta\tau_{\rm mn}}$ is simply a result of a better signal-to-noise ratio than either $\overline{\delta\tau_{\rm in}}$ or $\overline{\delta\tau_{\rm out}}$ considered alone.

4.3. East–West and North–South Travel Times

In Table 4, antisymmetric averages of the east–west and north–south travel-time differences appear most frequently at shallow to moderate depths and earlier time intervals. The left panels of Figure 5 show the distributions of $\overline{\delta\tau_x \cos\theta}$ and $\overline{\delta\tau_y \sin\theta}$ in filter TD3 with a lower turning point depth of 3.2 Mm, centered 24.5 hr before emergence. For both variables, the mean and the peak of the NE distributions lie close to 0 s, whereas the mean and the peak of the PE distributions are at positive travel-time differences of approximately 1 s. Unlike previous variables considered, there are multiple discriminant boundaries; the boundaries in the tails of the distributions, at large negative values of $\overline{\delta\tau_x \cos\theta}$ and large positive values of $\overline{\delta\tau_y \sin\theta}$, are likely to be spurious results caused by a few regions having extreme values.

The evolution of the skill score (Figure 5, right) shows that the variations with time are unlikely to be real given the uncertainties in the resulting skill scores, although there is perhaps a trend for worse performance of $\overline{\delta\tau_y \sin\theta}$ at times closer to the emergence time. However, this is one example of a variable that, in one time interval, falls above the threshold to be included in Table 4, whereas in other time intervals it may be excluded from the table, despite have substantial ability to discriminate PE from NE regions.

The particular combinations that appear, namely $\overline{\delta\tau_x \cos\theta}$ and $\overline{\delta\tau_y \sin\theta}$, would be expected to have a signal from a converging

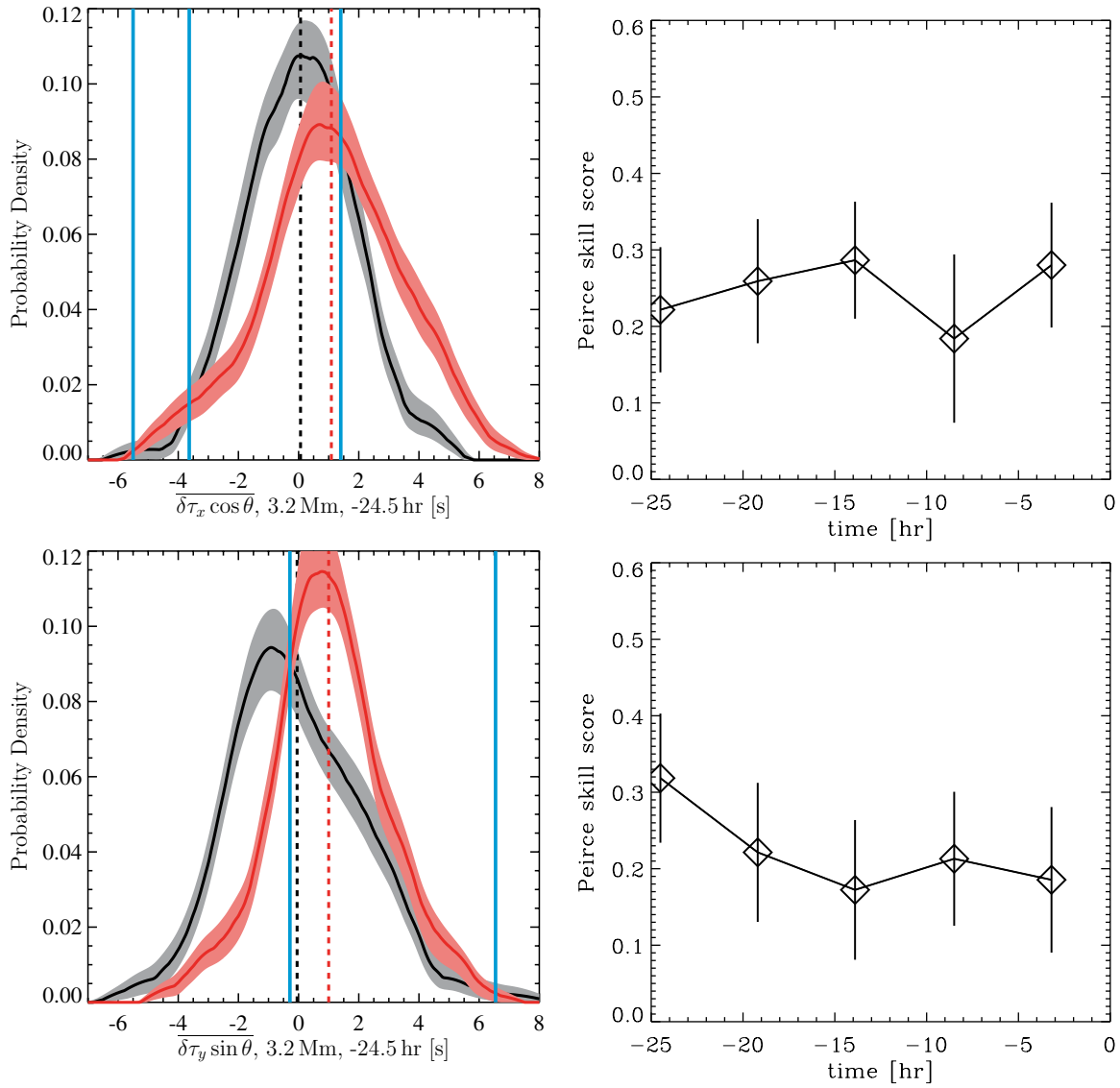

Figure 5. Nonparametric discriminant analysis for antisymmetric averages of the east–west and north–south travel-time differences, $\overline{\delta\tau_x \cos\theta}$ (top) and $\overline{\delta\tau_y \sin\theta}$ (bottom) in filter TD3, in the same format as Figure 1. Left: the density estimates for NE regions for both $\overline{\delta\tau_x \cos\theta}$ and $\overline{\delta\tau_y \sin\theta}$, 24.5 hr before the emergence time are peaked close to 0 s and are fairly symmetric; the density estimates for PE regions are both peaked close to 1 s. Right: given the uncertainties, there is no obvious trend in $\overline{\delta\tau_x \cos\theta}$ and only weak evidence for a trend toward worse performance closer to emergence for $\overline{\delta\tau_y \sin\theta}$.

flow. However, note that only one instance of $\overline{\delta\tau_{\rm oi}}$ appears in this table and in a filter with a much deeper lower turning point. The relative strength of the signals in $\overline{\delta\tau_x \cos\theta}$ and $\overline{\delta\tau_y \sin\theta}$ versus the signal in $\overline{\delta\tau_{\rm oi}}$ in general depends on the geometry of the assumed flow; detailed modeling is beyond the scope of the current work. We note, however, that Figure 5 of Paper II shows patterns in the ensemble averages of $\overline{\delta\tau_x \cos\theta}$ and $\overline{\delta\tau_y \sin\theta}$ with more structure than would be expected for a simple converging flow. As discussed in Paper II, one potential interpretation for the signals in $\overline{\delta\tau_x \cos\theta}$ and $\overline{\delta\tau_y \sin\theta}$ is a preference for emergence to occur at the boundary between supergranules, so these signals are the result of supergranular flows, not the emergence process itself.

4.4. The Vorticity

The remaining variable appearing multiple times in the list of best parameters is $\overline{{\rm vor}\sin\theta}$, particularly in filter TD4 with a lower turning point depth of 6.2 Mm. Figure 6 (left) shows a small but clear offset in the distributions of the NE and PE regions 24.5 hr before the emergence time, with PE regions more likely to have negative values of $\overline{{\rm vor}\sin\theta}$ relative to NE regions. As for the variables $\overline{\delta\tau_x \cos\theta}$ and $\overline{\delta\tau_y \sin\theta}$, this is an example of a variable that, in some time intervals, falls above the threshold to be included in Table 4, whereas in other time intervals it is excluded from the table, despite have substantial ability to discriminate PE from NE regions (see Figure 6, right).

It is known that surface magnetic fields are associated with a prograde flow (e.g., Zhao et al. 2004). A small area of prograde flow, centered in the averaging disk but diminishing toward the north and south edges, would result in the signal seen in $\overline{{\rm vor}\sin\theta}$ while likely producing a weaker signal in $\overline{\delta\tau_x}$. Thus one explanation for the signal in $\overline{{\rm vor}\sin\theta}$ is that, once again, the difference between the PE and NE regions is a result of the surface magnetic field. To determine whether there is a helioseismic signature of the emergence process not caused by the surface magnetic field, it is important to account for the

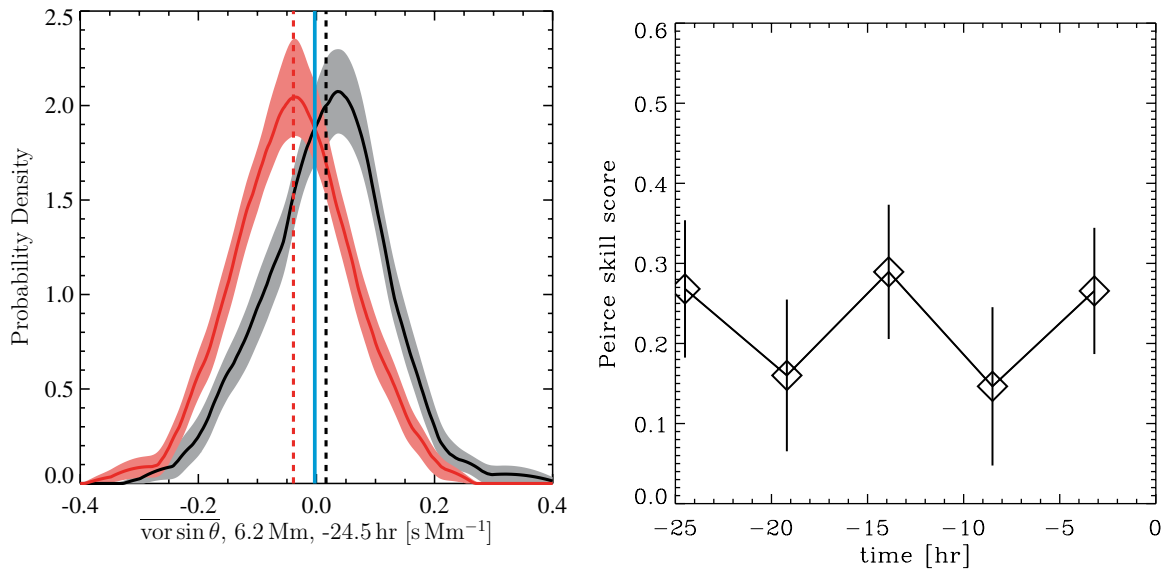

Figure 6. Nonparametric discriminant analysis for an antisymmetric average of the vertical vorticity, $\overline{\mathrm{vor}\sin\theta}$, in filter TD4, in the same format as Figure 1. Left: 24.5 hr before the emergence time, the density estimate for NE regions has a peak and a mean value that are slightly positive, whereas the peak and mean value for the PE distribution are at negative values. Right: given the uncertainties in the skill score, the performance of this variable does not clearly evolve with time.

contribution of the surface field to the differences in the travel times.

4.5. Matching the Flux Distribution

To investigate the influence of the surface magnetic field on the helioseismic parameters, we selected subsets of the NE and PE regions with matching distributions of average unsigned field, $\overline{|B_r|}$, and location on the disk. The approach to this was essentially the same as the selection of the NE regions to match the distributions in position and time of the PE regions described in Paper I: we used simulated annealing to select subsets of PE and NE regions that minimize the integrated absolute value of the difference between nonparametric density estimates for the two distributions in each time interval. Sixty-five regions from each sample were selected, which resulted in 50–55 regions with good duty cycle in each sample. This was the largest subset for which an integrated absolute difference of no more than about 0.1 could be obtained; the integrated absolute difference of two completely nonoverlapping distributions would be 2.

We performed the same analysis as for the full set of regions on these subsets of regions to rank the variables by skill score. We also repeated the Monte Carlo experiment (see Appendix B) for these sample sizes. We found that although the maximum skill scores of the helioseismology variables can reasonably be expected when there is no difference in the populations, there is still a preponderance of large skill score values in the helioseismology variables compared to what would be expected from chance. This suggests that some of the helioseismology variables have real ability to discriminate the PE regions from the NE regions but that it is difficult to determine if any specific variable has any power to discriminate. Because of this, we choose to show variables with Peirce skill scores greater than 0.24 in time interval 0 (rather than the 0.27 used for the full samples) in Table 5. This is done to highlight the variables (shaded in the table) that still have large skill scores after the flux matching.

Because of the matching of the distributions, $\overline{|B_r|}$ has virtually no ability to discriminate between the samples and thus does not appear anywhere in Table 5. To illustrate how well the distributions match, the distributions of $\overline{|B_r|}$ in time interval 0, a day before emergence, are shown in Figure 7, along with the skill score as a function of time. The skill score is consistent with zero in time intervals 0–3, and compared with Figure 1 the PE and NE distributions are very closely matched with no remaining tail to large values of $\overline{|B_r|}$ for the PE sample. The increase in skill score in time interval 4 is likely due to the start of emergence in a few regions.

For most of the variables that are not strongly correlated with the magnetic flux, the skill score values have not changed substantially, but the smaller sample sizes generally lead to larger uncertainties. Almost every filter and depth combination of $\overline{\delta\tau_x\cos\theta}$, $\overline{\delta\tau_y\sin\theta}$, and $\overline{\mathrm{vor}\sin\theta}$ present in Table 4 is also present after flux matching in Table 5, with a similar value of the skill score. For example, in time interval 0, the skill score values $\overline{\delta\tau_y\sin\theta}$ and $\overline{\mathrm{vor}\sin\theta}$ for the filters found in Table 4 lie in the range $0.27 \leqslant \mathrm{PSS} \leqslant 0.34$ for both the full samples

Table 5
Best-Performing Variables

Variable	Filter	Depth (Mm)	Peirce SS
TI-0: time $= t_0 - 24.5$ hr			
$\overline{\delta\tau_y\sin\theta}$	TD3	3.2	0.33 ± 0.10
$\overline{\delta\tau_y\sin\theta}$	TD2	2.2	0.32 ± 0.11
$\overline{\mathrm{vor}\sin\theta}$	TD8	15.7	0.31 ± 0.12
$\overline{\delta\tau_x}$	TD10	20.9	0.29 ± 0.10
$\overline{\mathrm{div}}$	TD5	9.5	0.27 ± 0.10
$\overline{\mathrm{vor}\sin\theta}$	TD4	6.2	0.27 ± 0.11
$\overline{\delta\tau_x\cos\theta}$	TD4	6.2	0.24 ± 0.11
$\overline{\delta\tau_{\mathrm{io}}\cos\theta}$	TD3	3.2	0.24 ± 0.11

Notes. Depth refers to the lower turning point of the waves in the filter used. Time is relative to the emergence time t_0. Shaded variables also appear in Table 4.

(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.)

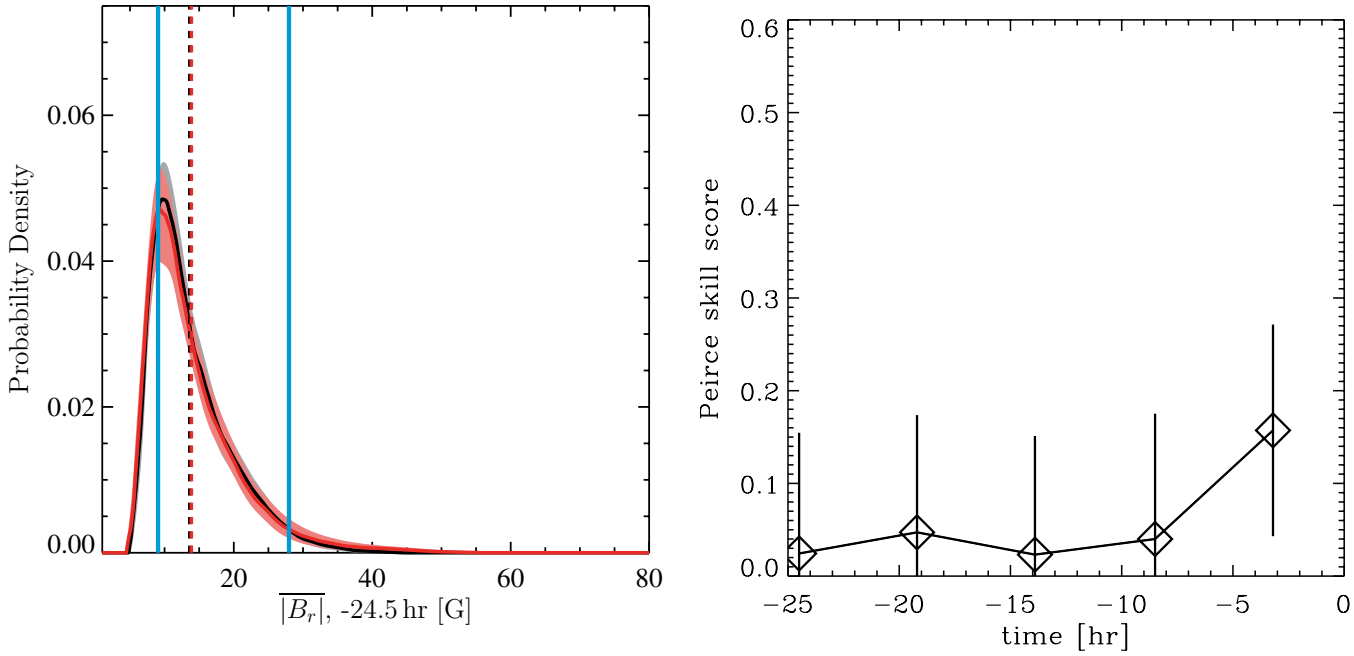

Figure 7. Nonparametric discriminant analysis for the mean unsigned magnetic flux, $\overline{|B_r|}$, for a subset of regions with matched distributions of magnetic flux, in the same format as Figure 1. By construction of the subset, the distributions of PE and NE regions are closely matched, as seen in the left panel. Thus the Peirce skill score is consistent with 0 for time intervals 0–3, as it should be for matched distributions, with a slight increase at the final time interval likely due to the onset of emergence during time interval 4 for some regions.

and the matched flux subsets. By contrast, the mean travel-time shifts in time interval 4 have consistently lower skill scores for the matched flux subset, indicating that the correlation with the magnetic flux accounts for some of the ability of $\overline{\delta\tau_{\rm mn}}$ (and $\overline{\delta\tau_{\rm in}}$ and $\overline{\delta\tau_{\rm out}}$) to discriminate between PE and NE regions. In addition to the best-performing variables for the full samples of PE and NE regions, there are a considerable number of other variables, in a range of filters, present in Table 5. Many of these are likely to be statistical anomalies, with no real ability to discriminate between the PE and NE regions.

Figure 8 shows the probability density estimates and the time variation of the skill score for $\overline{\delta\tau_y \sin\theta}$ in filter TD3 and for $\overline{\rm vor\sin\theta}$ in filter TD4. Qualitatively, the results are extremely similar to those shown in Figures 5 (bottom) and 6, where the full sets of regions were included. The distributions are peaked at similar values with similar widths, leading to discriminant boundaries in approximately the same locations. The main difference is that the uncertainty in the skill score has increased slightly. Thus, the ability of these variables to distinguish PE from NE regions is not a result of a difference in the average unsigned vertical field between the two samples, although it could still be a result of a different aspect of the surface field (e.g., the horizontal field or small areas of strong vertical field).

5. DISCUSSION

There are statistically significant differences between the properties of the pre-emergence and non-emergence samples that persist, with relatively little change, for at least a day prior to the onset of emergence. However, these differences are small, of the order of 1 s or less in the travel-time shifts on average, thus none of the variables considered can clearly distinguish an emergence from a non-emergence for any single region (see Figures 2 and 3 from Paper II). This is quite different from the results of Ilonidis et al. (2011), who found much larger travel-time reductions, although that study considered waves that propagate much deeper than were considered here.

The average unsigned magnetic flux at the surface was the best discriminator between the two samples. This could be a result of the appearance of small amounts of flux at the surface, starting at least one day prior to our definition of emergence time. The MDI instrument is unable to resolve this flux in a single magnetogram, and thus it is only distinguishable when averaging over many regions. Simulations of flux emergence (e.g., Cheung et al. 2010; Stein et al. 2011) do exhibit this type of behavior, thus our investigation shows some support for these simulations.

It is also possible that the ability of the average magnetic flux to distinguish between the two samples is a result of a bias in the samples, either of solar origin or as a result of our selection criteria. Our selection of NE regions (see Paper I) imposed a maximum field strength allowed that was not similarly applied to the PE regions. This could have resulted in a bias between the two samples in the average flux. However, it is also possible that the difference in average flux is a result of the tendency for ARs to emerge in the same location as prior ARs (e.g., Pojoga & Cudnik 2002) and not directly related to the emergence process.

While such considerations are important if the goal is to use helioseismology to predict the emergence of ARs, our goal is simply to determine if there is a helioseismic signal of emergence. The helioseismic measures that best distinguish the pre-emergence from the non-emergence regions are mean travel-time shifts, particularly immediately prior to emergence, antisymmetric averages of north–south and east–west travel-time differences, and an antisymmetric average of the vertical vorticity. The mean travel-time shifts are correlated with the presence of surface field and thus may not be related to subsurface properties of the emergence. This was confirmed by the reduced ability of mean travel-time shifts to distinguish PE from NE regions for subsets of the initial samples of PE and

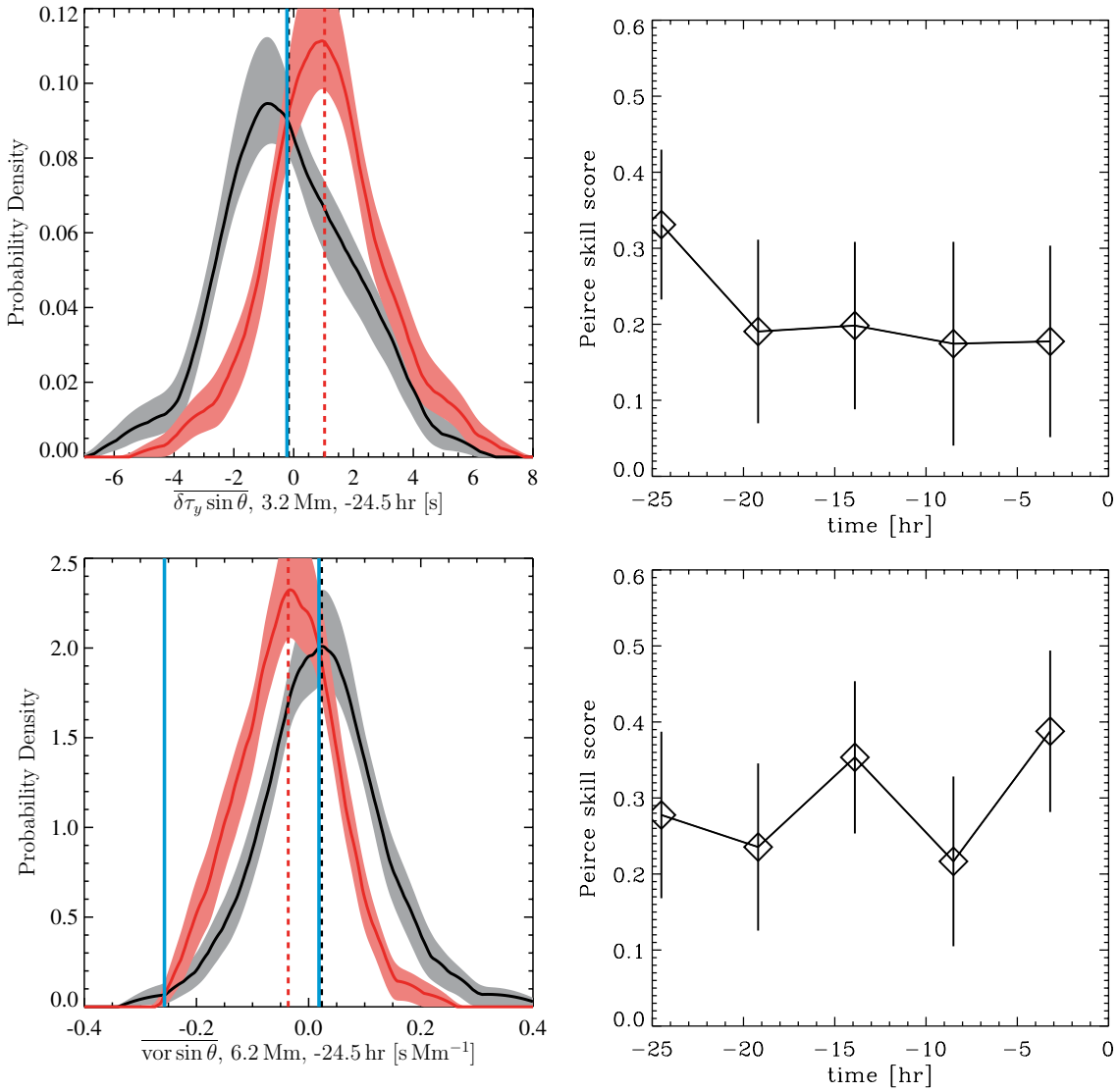

Figure 8. Nonparametric discriminant analysis for antisymmetric averages of the north–south travel-time differences, $\overline{\delta\tau_y \sin\theta}$, in filter TD3 (top), and for the vertical vorticity, $\overline{\mathrm{vor}\sin\theta}$, in filter TD4 (bottom), for a subset of regions with matched distributions of average magnetic flux in the same format as Figure 1. The results are qualitatively very similar to those seen in Figures 5 (bottom) and 6 for the full samples. The largest change is an increase in the uncertainties caused by the smaller sample sizes.

NE regions that had matched distributions of average unsigned magnetic flux.

The signals in the north–south and east–west travel-time differences and the signal in the vertical vorticity appear to not be sensitive to the surface field. Thus, we believe there are differences in the subsurface flows that can be detected by helioseismology prior to the emergence of significant magnetic flux. A converging flow could qualitatively explain the signals seen in the north–south and east–west travel-time differences, but it appears that the flow pattern is not a simple converging flow. A prograde flow of small extent in the north–south direction below the site of the emergence would produce the observed pattern in the vertical vorticity. This is perhaps related to the "small shearing flow feature" at a depth of 2 Mm described by Kosovichev & Duvall (2008) for AR 10488. There is no clear evidence for a retrograde flow, as would be expected from typical rising flux tube simulations (e.g., Fan 2008), although these simulations have an upper boundary at about 20 Mm below the solar surface. Instead, we found the vertical vorticity to be consistent with a prograde flow, and the difference in north–south travel-time differences is comparable to that in east–west travel-time differences, as would result from a converging flow. Thus our results suggest that the properties of simulated rising flux tubes must change as they approach the surface, at shallower depths than are presently simulated, if this is the mechanism by which ARs form.

As noted in Paper I, there are several ways in which a future investigation could improve on the present method. However, we have already found subtle but significant differences in the helioseismic signals from our samples of pre-emergence and non-emergence regions that suggest a detectable subsurface manifestation of AR formation prior to the appearance of significant surface magnetic flux. Our statistical results place strong constraints on models of AR formation.

This work is dedicated to Irene González Hernández for her help, enthusiasm, and scientific contribution. The authors acknowledge support from NASA contracts NNH07CD25C and NNH12CF23C. This work utilizes data obtained by the Global Oscillation Network Group (GONG) Program. GONG is managed by the National Solar Observatory, which is operated

by AURA, Inc. under a cooperative agreement with the National Science Foundation. The data were acquired by instruments operated by the Big Bear Solar Observatory, High Altitude Observatory, Learmonth Solar Observatory, Udaipur Solar Observatory, Instituto de Astrofísica de Canarias, and Cerro Tololo Interamerican Observatory. MDI data were provided by the *SOHO*/MDI consortium; *SOHO* is a project of international cooperation between ESA and NASA. We thank I. González Hernández, B. Javornik, and T. Dunn for their work in preparing the database. The Discriminant Analysis code was originally developed with funding from AFOSR under contracts F49620-00-C-0004 and F49620-03-C-0019. A.C.B. acknowledges DFG SFB 963 "Astrophysical Flow Instabilities and Turbulence" (project A18) aimed at understanding solar and stellar dynamos. The authors also thank the referee for improving the presentation of our results.

APPENDIX A

THE INFLUENCE OF DUTY CYCLE

Because of the varying duty cycle, some regions are only present in a subset of the time intervals. This could potentially influence the results if there is a handful of regions (with varying duty cycle) that are easy to classify. To check this, the analysis was repeated for the subset of regions that had a good duty cycle for every time interval. This severely reduces the sample sizes, to 48 and 45 for PE and NE, respectively. The main effect of this is an increase in the error bars, which is expected from the reduction in the sample sizes, without greatly changing the results. To illustrate this, the right panels of Figures 1, 3, and 5 have been reproduced in Figure 9 with this subset. All of the other plots exhibit the same behavior; thus we believe that this does not affect any of our conclusions.

APPENDIX B

MONTE CARLO EXPERIMENT

Given the number of variables considered compared to the number of data points, one should ask the question: are these results simply a statistical fluke? To answer this, a Monte Carlo experiment was performed. To represent one variable, two random samples of 85 points each (typical of the sample sizes in Section 4 and Table 4) were drawn from the same normal distribution. This was repeated for 66,250 variables

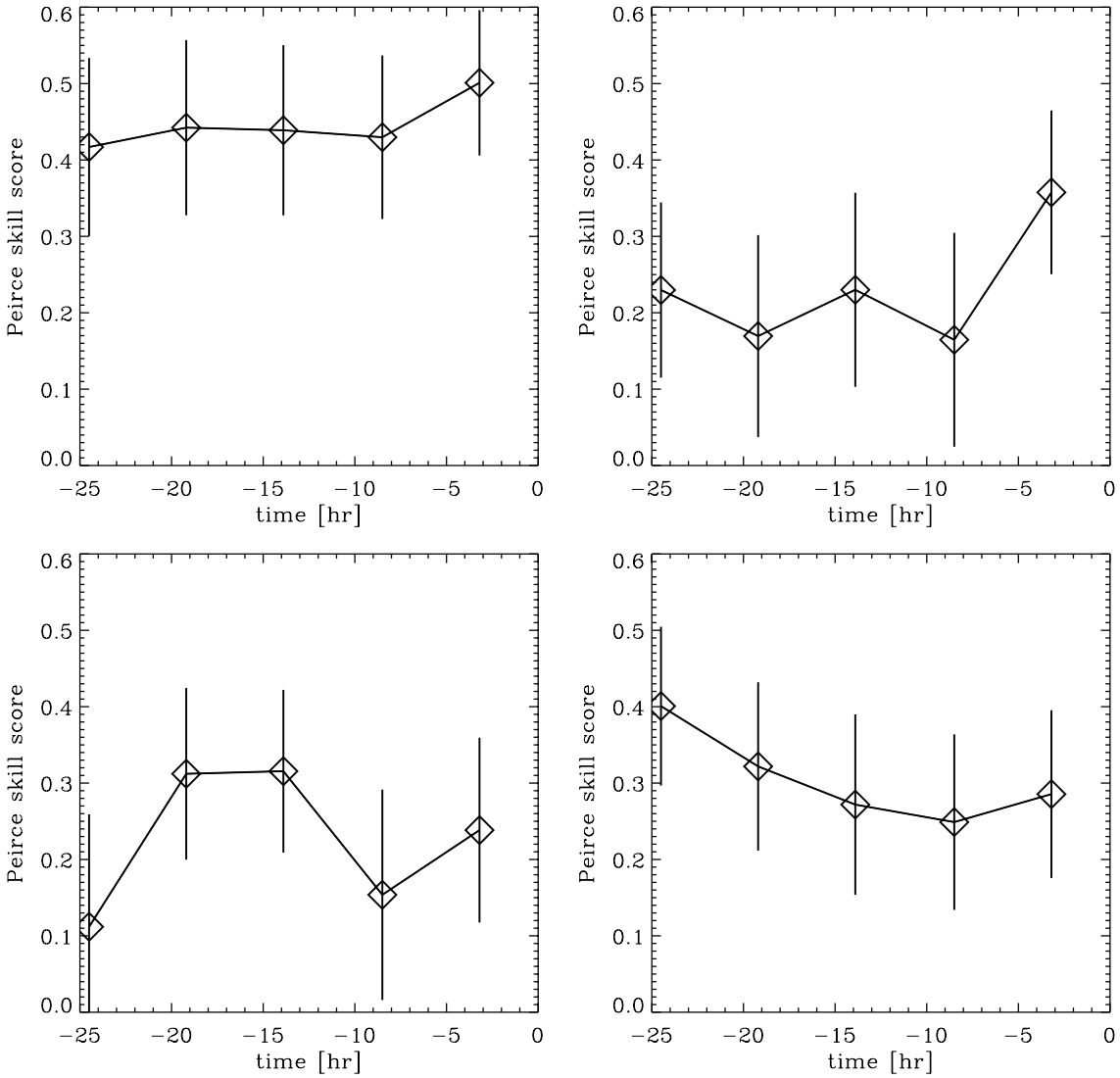

Figure 9. Evolution of the Peirce skill score for four variables: top left: $\overline{|B_r|}$, top right: $\overline{\delta\tau_{\rm mn}}$ in filter TD5, bottom left: $\overline{\delta\tau_x \cos\theta}$ in filter TD3, and bottom right: $\overline{\delta\tau_y \sin\theta}$ in filter TD3. Only regions with good data for all time intervals and seismology variables were used. Compared to the right panels of Figures 1, 3, 5, in which all available regions were used in each plot, the same trends (or lack of trends) are seen, but with larger uncertainties.

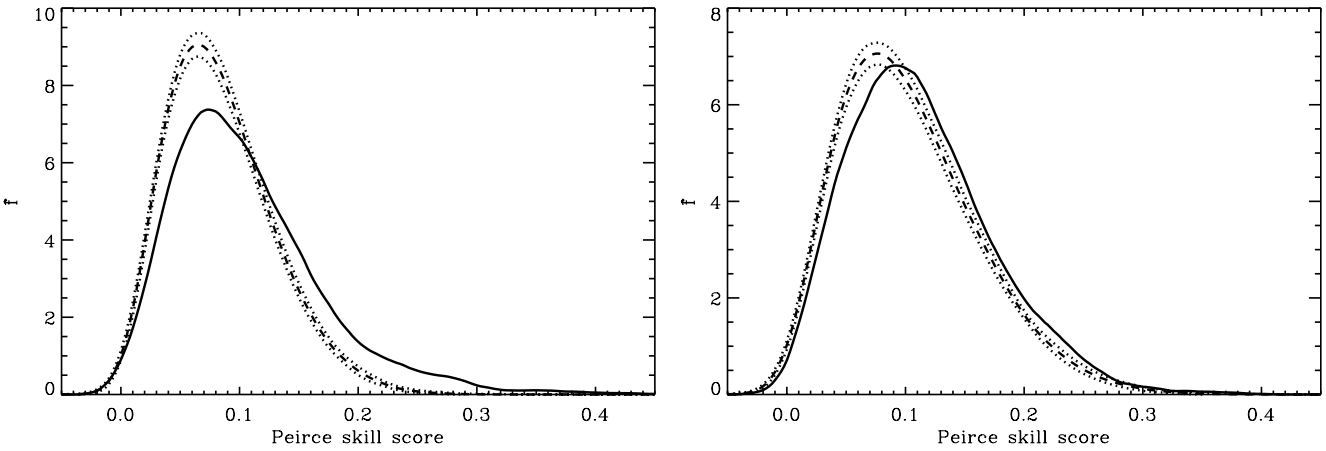

Figure 10. Probability density of Peirce skill scores from the Monte Carlo experiment (dashed curve) with 1σ error estimate (dotted curves) and from the variables considered for active region emergence (solid curve), using a bootstrap estimate. The left panel shows the results for sample sizes of 85; the right panel shows the results for sample sizes of 50. There is a clear tail of the distribution of the AR emergence variables to larger skill scores not present for the random variables, indicating that it is very unlikely that chance alone accounts for the performance of the best variables in distinguishing PE from NE regions. For the sample size of 85, the variables with skill scores above 0.27 almost certainly have a real ability to discriminate between PE and NE regions; for the sample size of 50, it is difficult to determine if any specific helioseismology variable has a real ability to discriminate between PE and NE regions, but the number of helioseismology variables with large skill scores ($\gtrsim$0.2) suggests that some can discriminate between the two.

(50 times the number of AR emergence variables), changing only the random number seed between variables. Nonparametric discriminant analysis was applied to the resulting values, and an unbiased bootstrap estimate of the Peirce skill score was made for each variable. The distribution of the resulting skill scores is shown in Figure 10, left, along with the distribution of the variables considered for AR emergence. Compared with the random variables, there is a preponderance of large skill score values for the emergence variables. Dividing the random variables into 50 sets of size equal to the number of emergence variables shows that the typical maximum skill score achieved is about 0.27, so the probability of getting a skill score greater than that if there is no information in the variable is extremely small. However, the distributions of random variables and of AR emergence variables show considerable overlap below a skill score of about 0.2.

To compare with the results when the distribution of magnetic flux was matched between NE and PE regions (Section 4.5 and Table 5), the experiment was repeated for two random samples of 50 points each. The resulting distribution of skill scores is shown in Figure 10 (right). There is still a preponderance of large skill scores for the emergence variables compared to the random variables, but it is less pronounced than for the larger sample size. Again dividing the random variables into 50 sets of size equal to the number of emergence variables shows that the typical maximum skill score achieved is now about 0.34, but it was as high as 0.43. In this case, it is no longer possible to determine whether any individual variable has any real ability to discriminate between PE and NE regions. However, it is possible to infer that there are more variables with high skill scores than would be expected from chance alone. The number of variables with skill score $\geqslant$0.27 (the value used in Table 4) is 17, compared with an expected number of 11. The chance of getting at least this many variables by chance is approximately 2%. The number of variables with skill score $\geqslant$0.2 is 120, compared with an expected number of 86. The chance of getting at least this many variables by chance is less than 2%. Thus there is reason to believe that there is a difference in the helioseismology variables between PE and NE regions.

For the results presented here, the distributions were assumed to be normal because this is a reasonable approximation to the expected noise distribution. However, to confirm the results the Monte Carlo experiments were also performed by drawing at random from a Cauchy distribution and from a cosine distribution. For the Cauchy distribution, which has longer tails than a normal distribution, the largest skill scores obtained were less than the largest skill scores for the normal distribution. For the cosine distribution, which has shorter tails than a normal distribution, the largest skill scores were very similar to those obtained from the normal distribution. In all cases, there is a clear tail of the distribution of the emergence variables to larger skill scores not present for the random variables, indicating that it is very unlikely that chance alone accounts for the performance of the best variables at distinguishing PE from NE regions.

REFERENCES

Birch, A. C., Braun, D. C., Leka, K. D., Barnes, G., & Javornik, B. 2013, ApJ, 762, 131
Brandenburg, A. 2005, ApJ, 625, 539
Braun, D. C. 1995, in ASP Conf. Ser. 76, GONG 1994. Helio- and Astro-Seismology from the Earth and Space, ed. R. K. Ulrich, E. J. Rhodes, Jr., & W. Dappen (San Francisco, CA: ASP), 250
Braun, D. C., & Birch, A. C. 2008, SoPh, 251, 267
Chang, H., Chou, D., & Sun, M. 1999, ApJL, 526, L53
Cheung, M. C. M., Rempel, M., Title, A. M., & Schüssler, M. 2010, ApJ, 720, 233
Couvidat, S., Gizon, L., Birch, A. C., Larsen, R. M., & Kosovichev, A. G. 2005, ApJS, 158, 217
Duvall, T. L., Jr., Jefferies, S.M, Harvey, J. W., & Pomerantz, M. A. 1993, Natur, 362, 430
Efron, B., & Gong, G. 1983, Am. Statist., 37, 36
Fan, Y. 2008, ApJ, 676, 680
Fan, Y. 2009, LRSP, 6, 4
Gizon, L., & Birch, A. C. 2005, LRSP, 2, 6
Gizon, L., Birch, A. C., & Spruit, H. C. 2010, ARA&A, 48, 289
Hartlep, T., Kosovichev, A. G., Zhao, J., & Mansour, N. N. 2011, SoPh, 268, 321
Harvey, J., Tucker, R., & Britanik, L. 1998, in ESA Special Publication, Vol. 418, Structure and Dynamics of the Interior of the Sun and Sun-like Stars, ed. S. Korzennik (Noordwijk: ESA), 209
Hill, F., Bolding, J., Toner, C., et al. 2003, in ESA Special Publication, Vol. 517, GONG+2002. Local and Global Helioseismology: The Present and Future, ed. H. Sawaya-Lacoste (Noordwijk: ESA), 295
Hills, M. 1966, J. R. Statist. Soc. B, 28, 1
Ilonidis, S., Zhao, J., & Kosovichev, A. 2011, Sci, 333, 993
Jensen, J. M., Duvall, T.L., Jacobsen, B. H., & Christensen-Dalsgaard, J. 2001, ApJL, 553, L193

Kendall, M., Stuart, A., & Ord, J. K. 1983, The Advanced Theory of Statistics, Vol. 3 (4th ed.; New York: Macmillan)
Komm, R., Howe, R., & Hill, F. 2009, SoPh, 258, 13
Komm, R., Howe, R., & Hill, F. 2011, SoPh, 268, 407
Kosovichev, A. G. 2009, SSRv, 144, 175
Kosovichev, A. G., & Duvall, T. L., Jr. 2008, in ASP Conf. Ser. 383, Subsurface and Atmospheric Influences on Solar Activity, ed. R. Howe, R. W. Komm, K. S. Balasubramaniam, & G. J. D. Petrie (San Francisco, CA: ASP), 59
Leka, K. D., Barnes, G., & Birch, A. C. 2013, ApJ, 762, 130
Lindsey, C., & Braun, D. C. 2000, SoPh, 192, 261
Lindsey, C., & Braun, D. C. 2005, ApJ, 620, 1107
Peirce, C. S. 1884, Sci, 4, 453
Pojoga, S., & Cudnik, B. 2002, SoPh, 208, 17
Scherrer, P. H., Bogart, R. S., Bush, R. I., et al. 1995, SoPh, 162, 129
Silverman, B. W. 1986, Density Estimation for Statistics and Data Analysis (London: Chapman and Hall)
Stein, R. F., Lagerfjärd, A., Nordlund, Å., & Georgobiani, D. 2011, SoPh, 268, 271
Woodcock, F. 1976, MWRv, 104, 1209
Zhao, J., Kosovichev, A. G., & Duvall, T. L., Jr. 2004, ApJL, 607, L135
Zharkov, S., & Thompson, M. J. 2008, SoPh, 251, 369