Recurrence quantification analysis with wavelet denoising and the characterization of magnetic flux emergence regions in solar photosphere

National Institute for Space Research, São José dos Campos, São Paulo 12227-010, Brazil

E. E. N. Macau
National Institute for Space Research, São José dos Campos, São Paulo 12227-010, Brazil
and Federal University of São Paulo, São José dos Campos, São Paulo 12247-014, Brazil

J. Palacios
Leibniz-Institut für Sonnenphysik (KIS), Freiburg im Breisgau, 79104, Germany

M. O. Domingues†
National Institute for Space Research, São José dos Campos, São Paulo 12227-010, Brazil

(Received 15 December 2017; published 29 July 2019)

Solar systems complexity, multiscale, and nonlinearity are governed by numerous and continuous changes where the sun magnetic fields can successfully represent many of these phenomena. Thus, nonlinear tools to study these challenging systems are required. The dynamic system recurrence approach has been successfully used to deal with this kind of challenge in many scientific areas, objectively improving the recognition of state changes, randomness, and degrees of complexity that are not easily identified by traditional techniques. In this work we introduce the use of these techniques in photospheric magnetogram series. We employ a combination of recurrence quantification analysis with a preprocessing denoising wavelet analysis to characterize the complexity of the magnetic flux emergence in the solar photosphere. In particular, with the developed approach, we identify regions of evolving magnetic flux and where they present a large degree of complexity, i.e., where predictability is low, intermittence is high, and low organization is present.

DOI: 10.1103/PhysRevE.100.012217

I. INTRODUCTION

Recurrence plot (RP) is considered to be one of the most efficient methods to deal with nonlinear and nonstationary time series [1,2]. It allows us to properly characterize the underlying system, following its changes over time [2,3]. As RP extracts the invariant properties of the system, it can be used to understand the relationship between interactive systems.

The main tool to analyze an RP is the recurrence quantification analysis (RQA), which was introduced by Zbilut and Webber [4] and is very effective to properly characterize the system dynamics and even to keep track of changes in the dynamics over time. However, it may be very sensitive to the presence of noise [5]. Additive noise or inbound noise may disturb the data series so that real recurrences are washed up, and so RQA presents numerical artifacts in many cases to pointing wrong results. To deal with this problem, we introduce a new preprocessing approach, based on the wavelet formalism for denoising. This new approach takes advantage of the well-known denoising ability based on the amplitude and local regularity detection of the wavelet coefficients [6–8].

To verify our methodology we apply it to the characterization of the solar magnetic field. The solar magnetic field presents a variety of phenomena in different time and spatial scales. The magnetic field is important to describe the solar activity and complex dynamics of the solar atmosphere. The dynamics in the solar photosphere from small-scale flux emergence to active regions shows signs of the complex behavior of magnetic fields below the surface.

The characterization of magnetic flux emergence can give some ideas about the physical mechanisms that are responsible for solar atmospheric phenomena. The relationship between the flux emergence regions and active regions has been widely studied; however, due to its complexity, many questions are still open [9]. The complex behavior of the solar atmosphere, such as the interaction of emerging flux with preexisting magnetic fields can lead to the creation of current sheets and magnetic reconnection in these regions [10]. Additionally, events from the smallest scales of the solar magnetism, such as small-scale magnetic flux intensification, coalescence, or splitting of small magnetic elements—such as bright points [11]—are fundamental to understand the surface dynamics.

*barbara.reis@inpe.br
†margarete.domingues@inpe.br
In summary, solar systems complexity, multiscale, and nonlinearity usually disobey comprehensive and deterministic modeling, except from certain statistical points of view. These multidimensional systems are governed by numerous and continuous changes [9]. The sun magnetic fields can present some alternations between deterministic trajectories and terminal dynamic phenomena. Such transitions can be discontinuous, being also responsible for all solar activity, such as coronal mass ejections and flares [12].

Our aim is to present an approach capable of detecting regions presenting low predictability, high intermittency, and/or less organization in a real short-time-images data set. Basically, the proposed approach consists of three steps: (i) preprocessing with a wavelet denoising technique, (ii) computing RP and RQA measures, and (iii) identification of the perturbed regions. In particular, the fine-scale events presented in the data set become very suitable examples for the application of our new methodology. Our results show that the preprocessing with wavelet denoising allows the RQA measurements to distinguish perturbed and unperturbed regimes present in the solar data. From this, we can demonstrate that our approach is efficient in handling short sequences of magnetic field images or similar sequences.

This paper is structured as follows. Section II presents a summary of the nonlinear techniques: wavelet denoising, recurrence plot, and recurrence quantification analysis. Section III shows the obtained results, and Sec. IV presents the final discussions and conclusions. Appendix A contains the description of solar data. Appendix B describes the denoising method and includes some practical examples of the techniques used here. Appendix C contains cases studies concerning the magnetic flux analysis of perturbed and unperturbed regions in the solar photosphere.

II. METHODOLOGY

The study is designed by using the procedures described as follows. First, we obtained a sequence of 167 grayscale images, with $496 \times 309$ pixels, regularly spaced in time (4.5 min) from the same region of the Sun (as described in Appendix A). Then we reduced the Gaussian white noise present in each image using the wavelet denoising technique. After that, we generated a time series, with 167 points, for each position ($x, y$). In total, we created 153,264 unidimensional time series. For each time series, we computed the RP, calculated the RQA measures, and placed the results in the respective ($x, y$) position of the output image (see Fig. 1). For each measure we made a histogram and defined a threshold capable of distinguishing calm and turbulent behavior. As a final result, we had a binary mask in which the black and white colors indicate the positions whose values are above and below the threshold, respectively. In the following subsection we briefly describe these techniques.

A. Recurrence plot and recurrence quantification analysis

Developed by Eckmann et al. [1], the recurrence plot is a graphical tool used to identify characteristics in nonlinear data. It is a visualization of a binary square matrix defined by

$$R_{i,j}^{m,\rho} = \theta(\rho - ||\vec{x}_i - \vec{x}_j||); \vec{x} \in \mathbb{R}^m; i, j = 1, \ldots, N,$$

where $N$ is the number of states $\vec{x}$, $m$ is the embedding dimension, $\rho$ is a radius of the neighborhood, $|| \cdot ||$ is a norm, and $\theta(\cdot)$ is the Heaviside function [2]. If $R_{i,j}^{m,\rho} = 1$, then the states $\vec{x}_i$ and $\vec{x}_j$ are recurrent and a black dot is placed at position $(i, j)$. If $R_{i,j}^{m,\rho} = 0$, then the states $\vec{x}_i$ and $\vec{x}_j$ are not recurrent and a white dot is placed at position $(i, j)$. The patterns presented in the RP can give a qualitative interpretation of the data under analysis such as stationary, periodicity, high fluctuations, evolution of similar states at different times, and slow changes.

There is a lot of discussion on how to obtain the RP parameters as reviewed for instance in Ref. [2]. However, as discussed in Refs. [13–16], these parameter choices do not significantly alter the characterization process. For the sake of reproducibility, we use the number of states as 167, the radius of the neighborhood as 5 (which is equivalent to 0.05% of phase space diameter), the Euclidean norm, and the embedding dimension and time delay as 1 for RP calculations.

It is well known that noise in observational data affects the RP [5]. Therefore, it is important to consider a noise reduction of the data. In this work we use the wavelet denoising technique to reduce the noise in the images in order to improve the quality of the RP.

The recurrence quantification analysis can be used to aid the interpretation of the RP [2,4,17–19]. In this technique the dynamics of the system is quantified by statistics extracted from vertical, diagonal, or horizontal lines, isolated points,
structures, and clusters present in the RP. In this work we used the program available on http://www.recurrence-plot.tk to calculate the following RQA measures.

1) **Recurrence rate** ($\mathcal{R}$) is a measure of the density of recurrence points in the recurrence plot. It is defined as

\[ \mathcal{R} = \frac{1}{N^2} \sum^{N}_{i=1} \sum^{N}_{j=1} R^{n,p}_{i,j}. \]  

2) **Determinism** ($\mathcal{D}$) quantifies how predictable the system is. This measure is the ratio of recurrence points that form diagonal lines, at least length $\ell_{\text{min}}$, to all recurrence points:

\[ \mathcal{D} = \frac{\sum_{\ell=\ell_{\text{min}}}^{N} \ell \ P(\ell)}{\sum_{\ell=1}^{N} \ell \ P(\ell)}, \]  

where $P(\ell)$ is the histogram of the diagonal lines lengths.

3) **Entropy** ($\mathcal{E}$) refers to the Shannon entropy [20] of the probability to find a diagonal line of exactly length $\ell$ in the RP. It is described as

\[ \mathcal{E} = - \sum_{\ell=\ell_{\text{min}}}^{N} p(\ell) \ln[p(\ell)], \]  

where $p(\ell) = P(\ell)/N_{\ell}$ and $N_{\ell}$ is the number of diagonal lines in the recurrence plot. This measure reflects the complexity of the recurrence plot [21].

4) **Laminarity** ($\mathcal{L}$) quantifies an intermittency that results in vertical structures in the recurrence plot. This measure is the ratio of recurrence points that form vertical lines, at least length $\nu_{\text{min}}$, to all recurrence points:

\[ \mathcal{L} = \frac{\sum_{\nu=\nu_{\text{min}}}^{\nu_{\text{max}}} \nu \ P(\nu)}{\sum_{\nu=1}^{\nu_{\text{max}}} \nu \ P(\nu)}, \]  

where $P(\nu)$ is the histogram of the vertical lines lengths.

**B. Wavelet denoising**

Our objective is to highlight the regimes present in the magnetograms using the RP. However, the noise can affect the computation of the RP and the RQA. Therefore, it is necessary to reduce the noise in the observational data to improve the quality of the analysis [22].

In this work we use a nonlinear wavelet denoising technique called VISU Shrink [6,8] to reduce the Gaussian white noise present in the solar data. Basically, this technique consists of four steps: (i) apply the discrete wavelet transform to obtain the wavelet coefficients of $n$ decomposition levels; (ii) calculate the universal threshold given by $\sigma \sqrt{2 \log(N)}$, where $N$ is the number of data points and $\sigma$ is the standard deviation of noise estimated by median absolute deviation on the wavelet coefficients in the first decomposition level divided by factor 0.6745; (iii) apply a soft thresholding algorithm to all wavelet coefficients; and (iv) obtain the denoised signal from the modified wavelet coefficients. An example of the application of this procedure is presented in Appendix B.

It should be emphasized that the nonlinear wavelet denoising technique was applied in each of the 167 images. We have chosen the spatial denoising instead of temporal denoising because the time series registers are shorter (167 points) than the space image ($496 \times 309$ pixels) registers.

For the multilevel wavelet transform methodology, the larger the series, the better the methodology performance is and even its feasibility, because at each level decomposition the records in the data are reduced by a factor two, as described in Ref. [23].

We have used the wavelet coefficients related to the Cartesian product of the directional analyzing wavelet functions, in rows and columns, of the first level to compute the noise standard variation. The selected analyzing wavelet function is the Daubechies wavelet function with local cubic polynomial approximations as described in Appendix B.

**III. ANALYSIS OF PERTURBED REGION CANDIDATES OBTAINED BY RQA**

Aiming at a blind comparison and further skill score computing with RQA parameters, we have defined the current solar magnetic activity as the benchmark, generating a binary mask of perturbed and unperturbed regions. Unperturbed regions are quiet regions and perturbed regions are locations where magnetic activity is present at least temporarily. Our benchmark (Fig. 2) is defined by selecting two different intervals by a threshold. Magnetograms in absolute value are thresholded individually and then the series average is computed. The perturbed region correspond to values above $3\sigma$ and unperturbed regions to values below $3\sigma$. The $\sigma \approx 10.2$ G set is from Ref. [24] for typical noise level of helioseismic and magnetic imager (HMI) magnetograms with cadence of 45 s.

In Fig. 3 we present examples of typical candidates for unperturbed and perturbed magnetic time series. This figure shows the density of magnetic field [Fig. 3(B), in gauss] versus time, where the frame cadence is 4.5 min (with a total of 167 frames). The perturbed series case is located at [72,116] arcsec and unperturbed series case at [25,62] arcsec. The RP obtained from the time series is a square binary matrix of size $167 \times 167$. The patterns in black identify the features for analysis. It is worthwhile to note that the time series has very different amplitudes, although that does not influence the RP analysis.

We can also observe that the candidates to perturbed series have distinct patterns and the RP identifies the magnetic flux trend in the beginning of the series. The unperturbed candidate region presents periodic patterns in the RP.

![FIG. 2. Benchmark image indicating perturbed (black) and unperturbed (white) regions.](image-url)
Based on the distribution of the values of the RQA quantifier image before the wavelet denoising (Fig. 4), we choose the threshold as the value that obtains the best skill score when compared to the benchmark defined in Fig. 2. The skill score $S$ in % given by

$$S = \frac{\text{no. of pixels that are correctly classified}}{\text{total no. of pixels}} \times 100.$$ 

The perturbed region candidates are the regions where measures are smaller than 0.15, 0.49, 0.56, and 0.835, respectively, for the $R$, $D$, $L$, and $E$ quantifiers. For each RQA measurement matrix ($I$) we create the following mask:

$$M(x, y) = \begin{cases} 1, & \text{if } |I(x, y)| \geq \delta \\ 0, & \text{otherwise} \end{cases}$$ 

in which $M(x, y) = 0$ is shown in white to indicate that the position $(x, y)$ was in an unperturbed region and $M(x, y) = 1$ is in black to show that the position $(x, y)$ was in a perturbed region.

Figure 5 shows the binary RQA images obtained. $R$ measures are sensitive to any flux emergence even under weak and nonpersistent flux emergence regions. On the left panels of Fig. 5, we present the RQA result images for the original images and, on the right panels, the results for the denoised images. The denoised images reduce the pointwise regions that can yield false-positive candidates to perturbed regions, especially on $D$, $E$, and $L$ measurements. The $R$ measurements are not affected by the denoising. On the other hand, the denoising enables the $D$, $E$, and $L$ measures to identify regions with large perturbations.

The proposed methodology allows us to distinguish between two regimes associated to perturbed and unperturbed solar states based on the time-series analysis. These kinds of results can be modified when we use different time windows, because they define physical processes. Particularly, in this work, we have considered a time duration associated to flux emergence regions as a case study. In the future we plan to extend this study to other physical processes in solar atmosphere. With those studies we will be able to define (estimate) the uncertainty of the measurements in the physical sense. So far, the results are encouraging because we have been able
to identify these regimes and this seems a good in order to develop a complementary tool for automatic classification or alert.

In Appendix C we present the detailed study of the behavior of the magnetic flux in the four selected cases identified in this study.

IV. FINAL DISCUSSIONS AND CONCLUSIONS

The characterization of the dynamic structures and their behavior in the solar photosphere considering two-dimensional images is a challenge. In this work we present a new approach which combines the wavelet denoising with the RQA measures to help with this task.

Our goal was to identify the perturbed and unperturbed regimes presented in the solar data. Physically, the perturbed regime has low predictability, low organization, and high intermittency that can be associated to reconnection regions over the solar photosphere, while the unperturbed regime is associated to quiet sun regions and weak and nonpersistence flux emergence regions. In our approach a region \((x, y)\) is considered perturbed when all the RQA measures \((R, S, E, L)\) present values above their own thresholds. Consequently, the other cases are considered unperturbed regions. Otherwise, the region \((x, y)\) is considered unperturbed. We also observe that the \(R\) measure is the most sensitive measure to flux emergence (evolving regions).

One of the most powerful and promising future applications of this method is testing and validating the recurrence of solar active regions (ARs) appearance in long timescales. Additionally, the wavelet denoising is specially useful for HMI magnetogram noise correction, whose noise is time and disk-position dependent, as shown in Ref. [25]. An improvement on this technique could be made with the use of a more refined way to compute the wavelet threshold, as suggested in Ref. [26]. We believe that these tools can be extended to the study of other phenomena and can contribute to an objective analysis in solar physics research in special to flux emergence investigations. This methodology can also be successfully applied in similar three-dimensional real data.

ACKNOWLEDGMENTS

The authors are grateful to the SDO HMI Data Science and Teams and Dr. Marwan for the codes used to compute the RP-RQA. The authors thank CAPES Brazil (Grant No. 1710827), CNPq Brazil (Grants No. 306038/2015-3, No. 140779/2015-9, No. 312246/2013-7, No. 304209/2014-7, No. 162480/2014-8, No. 302226/2018-4, No. 424352/2018-4, and No. 307083/2017-9), FAPESP Brazil (Grants No. 2015/25624-2 and No. 2015/50122-0), and FINEP/CT-INFRA (Grant No. 01120527-00) for financial support. J.P. thanks to the MINECO project AYA2016-80881-P (including AEI/FEDER funds, EU).

APPENDIX A: SOLAR DATA

The HMI SDO [27] provides continuous monitoring of the Sun’s magnetic field through full-disk photospheric data with high cadence and high spatial resolution [28]. We used 4.5-min (270 s) cadence line-of-sight magnetograms from HMI SDO. NOAA11855. Figure 6 shows six samples of the selected region from a time series covering about 12.5 h.

APPENDIX B: DAUBECHIES WAVELET DENOISING

The main point to use the technique of Donoho is to have a good local polynomial approximation. We have chosen the Daubechies’s wavelet because it has a high-order local polynomial reproduction that improves the quality of the denoising. Moreover, with this choice, we use an orthogonal wavelet transform. The orthogonality is also really helpful when we consider a two-dimensional transform, mainly because we avoid redundancy and unnecessarily increasing the number of wavelet coefficients. Also, this transform has fast algorithms that allow efficient computations of the data, as discussed in Ref. [29]. Furthermore, there is an isometry feature in both transformed and untransformed domains in the sense of the space of squared integrable function \(L_2\). Then, the signal energy is preserved in the sense of the Parseval theorem. Ad-
FIG. 7. Example of the original image at 18:28:30UT and zooms showing the original and denoised regions using db4 wavelet with two decomposition levels. Original image (a), denoised image (b), zoom showing the original region (c), and zoom showing the denoised region (d).

Additionally, the 167 frames with Donoho’s threshold procedure present a mean of 18% of signal energy related to the Gaussian white noise (mode 13%, std 4%, and median 19%).

In this discrete wavelet transform procedure, we use the Daubechies orthogonal wavelet filters with order four ($N = 4$), which corresponds to local polynomial reproduction of cubic degrees, and eight coefficients (e.g., $n = 0, \ldots, 7$). We adopt the $L^2$ normalization (conservation of the signal energy between the physical space and the wavelet space), i.e., the sum of coefficients is normalized by the factor $\sqrt{2}/4$ [[23], Table 6.1, p. 195]. We compute this transform using the pywt library available at http://wavelets.pybytes.com/wavelet/db4/.

One example of the denoising procedure is presented in Fig. 7 and, with zoom in, in Fig. 8. The denoising procedure slightly smooths the image, as desired for experimental data according to the suggestions in Ref. [5]. The RQA measures are affected by noise and there is a detailed theoretical discussion in Ref. [5]. We have improved the RQA results by using Donoho’s wavelet denoising technique, which only removed Gaussian white noise from the data. The threshold parameter for denoising is obtained directly from the first-level decomposition data. Choosing an analyzing wavelet that has a good polynomial reproduction, the value of the threshold will not change significantly, as discussed in Donoho and Johnstone [6]. In our case, we use a third-order polynomial reproduction.

Finally, based on Donoho’s threshold, we have removed the nonlinear white noise. The range of scale does not influence the threshold value itself as Donoho’s threshold is based on the first decomposition level. However, the application of the threshold in a multilevel transform can influence its reconstruction. The multilevel transform makes sense only if we have sufficient data. Thus, for this kind of application, the mentioned effect could be investigated in a deeper way in the future.

APPENDIX C: MAGNETIC FLUX ANALYSIS IN REGIONS DEFINED BY RQA

In this section we present case studies of perturbed or unperturbed regions where we analyze the behavior of the magnetic flux. Four regions are selected using the results of RQA measures with the wavelet denoising from Fig. 5.

FIG. 9. Magnetic flux in four selected regions from the measure $\Phi$. Upper left: Case I corresponds to a perturbed region, Case II describes an unperturbed region, and Case III and Case IV are selected regions of magnetic field variation. Upper right: Average magnetogram. Lower left: Magnetic flux in each selected region; Case I is represented by a solid line and Case II by a dotted line. Lower right: Magnetic flux computed for Case III (solid line) and Case IV (dotted line).

In Fig. 9 Case I corresponds to a perturbed region and the main spot of the flux emergence of the AR; Case II shows an unperturbed region, as an example of a quiet sun area; Case III and Case IV show bright point regions and their variable magnetic field and some magnetic feature coalescence. The values of the unsigned magnetic flux from Cases II, III, and IV show the same variations, and they are lower than $0.5 \times 10^{20}$ Mx, in agreement with the quiet sun behavior.
Analysis of the signed magnetic flux. We calculate the signed (positive and negative) magnetic flux for the whole magnetogram field-of-view (FOV) during the time series. The starting time is September 29 at 06:00UT. The positive and negative fluxes increase during this period, probably caused by the emergence of the active region in the FOV. The signed magnetic flux shows an imbalance, the negative flux increases from $1.07 \times 10^{20}$ Mx to $1.3 \times 10^{20}$ Mx, while the positive flux increases from $1 \times 10^{20}$ Mx to $1.3 \times 10^{20}$ Mx, showing the increase of signed magnetic flux during the process until the flux emergence region appears.

We use the selected regions from Fig. 9 to calculate the evolution of the signed magnetic flux in a perturbed and unperturbed region (Case I and II). Figure 10 Case I shows that the magnetic flux increased in the same way. Nevertheless, after 13:12UT the negative magnetic flux is higher than the positive magnetic flux. The Case II panel shows the signed magnetic flux. Positive and negative magnetic fluxes are constant in this quiet-sun region during this selected period. In Fig. 10, Case III shows that the positive magnetic flux is higher than the negative one. The positive magnetic flux shows a constant trend in the interval of about $4 \times 10^{18}$ Mx and the negative magnetic flux shows variations in lower values and a similar constant trend of $1 \times 10^{18}$ Mx. Case IV shows larger values of negative magnetic flux, where the positive magnetic flux increased while the negative magnetic flux decreased until 13:12UT. Later, the negative magnetic flux shows some variations near $4 \times 10^{18}$ Mx, while the positive magnetic field varies near $2 \times 10^{18}$ Mx. This behavior may be related to dynamics of the magnetic feature in Case IV.

We used selected regions from Fig. 9 (upper right panel) from Cases III and IV to calculate the variation of maximum and minimum values of the magnetic density in order to check the possibility of magnetic field intensification (Fig. 11). The magnetic field intensification or convective collapse is a relevant physical phenomenon where the smallest flux tubes observed in the solar photosphere increase their flux density due to plasma evacuation outward the tube [30]. The maximum and minimum values of the average magnetic density from magnetograms show appreciable changes in intensity, but they do not reach any kilogauss (kG) values. Case III (left panel) is a positive flux feature and shows high variations on its density but does not reach kG values. The negative flux density line is stable and corresponds to the noise level. Case IV (right panel) is a negative flux feature, and its behavior is similar to Case III, with opposite signs. However, the dynamics of Case IV is higher and the increase in density is more evident in magnetograms. Both features in Cases III and IV are composed of bright points that are moving, coalescing, and separating.

These regions show an interesting behavior to test the quantifiers $R$, $E$, $D$, and $L$ on them, due to the magnetic density and flux variations in each selected box associated with different polarities or signed magnetic field, and because of the very fine dynamics observed, as merging and drifting of small magnetic elements.

The RQA measurements highlight the flux emergence, bright points dynamics, cancellation events, and coalescence (Figs. 6 and 9). To assess the statistical significance of our results, we perform the Kolmogorov-Smirnov test [31–33], in which we compare the distributions of the magnetic flux in different regions. The regimes obtained by the RQA technique can be used to highlight characteristics in the solar photosphere that may be associated with the same phenomena, such as the quiet sun regions and field intensification regions.

This technique displays significant changes in the fine structure in the magnetic field (Fig. 11). The small-scale magnetic flux elements related to the flux emergence in developing ARs are emphasized (Fig. 10). Therefore, we verify that RQA measures can indicate some characteristics that could be associated to perturbed regions with variations in distinct spatial and temporal scales on the solar photosphere, specifically changes at internetwork regions, variation of magnetic field, and flux emergence areas. This behavior is validated with computations of the unsigned and signed magnetic flux in these regions to check the possible magnetic field intensification.