A robust and adaptive approach to modelling physical phenomena

Mwitondi[1], K.; Said[2], R.; Yousif[3], A.; Moustafa[4], R.; Wang[5], K., Zimin[6], Z. and Mmasi[7], R

1. *Sheffield Hallam University, Department of Computing: k.mwitondi@shu.ac.uk*
2. *Al Ain University of Science and Technology: raeed.tawfeq@aau.ac.ae*
3. *Qatar University, Department of Maths, Stats and Physics; aeyousif@qu.edu.qa*
4. *George Washington University, Department of Statistics: shalash@gwu.edu*
5. *Radio and Space Services, Australian Bureau of Meteorology: m.wang@ips.gov.au*
6. *WDC for Space Science - Chinese Academy of Sciences: mzou@nssc.ac.cn*
7. *Tanzania Commission for Science and Technology: rmmasi@costech.or.tz*

Abstract

Modelling the behaviour and evolution of the physical phenomena which surround us remains a major challenge to the data science community. Modern enhancements in data acquisition, storage, processing and transmission, highlight the need for more accurate and reliable tools, techniques and skills for extracting knowledge from the available and highly dynamic large volumes of data. Typically, modelling of natural phenomena rely on the deployment of mathematical models quite often built on the foundations of stringent assumptions. In many applications some of the underlying assumptions are violated and the models fail to yield closed form or unique solutions. We propose a generic approach to modelling sunspots numbers using integrated adaptive unsupervised and supervised models. We adopt the data's natural Gaussian distributional properties and use the early patterns as the basis for unsupervised and supervised modelling. Comparing multiple early patterns for each recorded cycle extracted at different time periods to the corresponding full cycles reveals that the first 3 years provide a sufficient basis for predicting the cycle's peak. Based on multiple simulations we develop a binary cut-off point of low and high solar activity which we use to label the data and apply Support Vector Machines (SVM) for predicting new cycles. Repeated SVM runs using repeatedly improved data parameters show that the approach yields greater accuracy and reliability than conventional approaches mainly because it simultaneously traces anomalies and provides a robust basis for model selection. Finally, we describe how the method can be adapted to other unsupervised and supervised methods with different applications.

Key Words: *Data Mining, K-Means, Predictive Modelling, Solar magnetic activity, Sunspots, Supervised Modelling, Support Vector Machines, Unsupervised Modelling*

1 Introduction

The overall behaviour of the solar magnetic activity cycles has attracted the attention of scientists for many years. Solar flares affect our planet in different ways - including ejecting plasma and energetic particles and potentially causing geomagnetic storms and damaging satellites (Reames, 2002). The interactions between the sun's surface plasma and its magnetic field are known to generate sunspots - clustered patterns in non-random positions above and below the equator (Schwabe, 1843 and Wolf, 1852). Tracking the general behaviour of sunspots provides solar scientists with a way of monitoring and measuring solar activity and particularly how it impinges on life on earth. Correlations between space and terrestrial weather have been indicated in solar studies dating back many years such as those by Siscoe (1978), Pielke *et al.,* (1998) and Rycroft *et al.,* (2000). Glasby (2002) used observational data from a set of three eleven-year sunspot cycles to develop comprehensive hypotheses on how the planets trigger natural phenomena such as sunspots and earthquakes. Various methods have been used in studying the cycle's strength and duration. Notable examples are in Kitiashvili and Kosovichev (2009) who used the data assimilation method and in Choudhuri *et al.,* (2007) and Dikpati *et al.,* (2006) who used a rotational solar dynamo-based approach in to predict the 24th cycle. Qahwaji and Colak (2007) used a variety of machine learning techniques for short-term predictions of solar flares. However, the issue of model complexity for disparate methods on disparate data sources is best addressed via adaptability rather than comparability. Consequently, minimising inherent randomness in training and test data requires novel adaptive methods of data analysis (Mwitondi and Said, 2011 and Mwitondi and Bugrien, 2010).

We propose an adaptive and robust approach to modelling capable of providing real-time predictions of the solar activity cycles based on its 11-year frequency of sunspots. The method seeks to uncover naturally arising structures in data by searching for generalising parameter levels and adapts them to supervised modelling of the data. The paper is organised as follows. Methods are described in Section 2 followed by data analyses and discussions in Section 3 and concluding remarks and potential new directions in Section 4.

2 Methods

The main modelling strategy combines sunspots historical data's distributional parameter estimates with specific hypothesised conditions as the basis for supervised modelling.

2.1 Data description and visualisation

We adopt the beginning and ending periods of solar cycles in as in Kane (2002) in which the first cycle starts in March 1755 and ends in June 1766 with the onset of the current (24th) cycle being December 2008 with each cycle lasting approximately 11-years. The LHS panel of Figure 1 provides a spider plot for the mean and variation patterns of the 23 cycles and the onset of the 24th cycle. The RHS panel exhibits the intensity of the cycles (ordered bottom-up). Our strategy adapts the discernible high and low solar activities as model inputs for detecting solar activity patterns over both short and long periods of time.

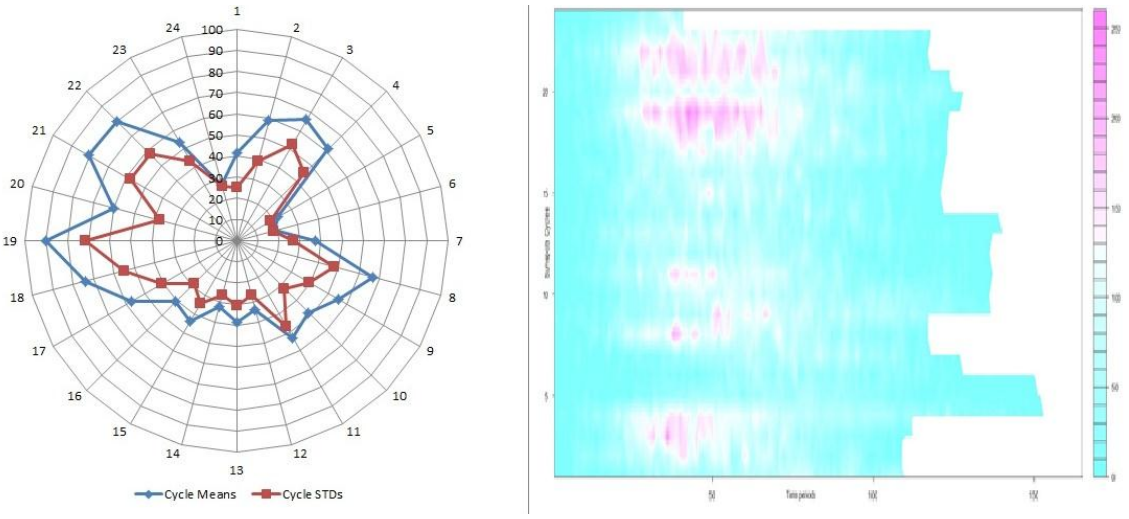

Figure 1: The spider web (LHS) and contour image (RHS) showing sunspots intensity

The relative sunspot number $s = \nu(10g + \lambda)$ where g denotes the number of sunspot groups, λ is the total number of distinct spots and $\nu < 1$ is scale factor that accommodates specific conditions of the observer, derives from the definitions and refinements in Wolf (1848, 1852 and 1861) and detailed in Izenman (1983). Wolf (1852) used eight estimates of sunspots periods for each set of "clear minima" and six "clear maxima" to determine the duration of the cycles' periodicity as $\hat{p} = \frac{\sum_{j=1}^{16} p_j {w^*}_j}{\sum_{j=1}^{16} {w^*}_j} = 11.11$ years. The 16 periods are denoted by $p_{j=1,2,\dots,16}$ and weighted by $w_j^* = {}^{1}\!/_{\epsilon_j}$ where ϵ_j is the estimated error.

Although the 11-year solar cycle has since been adopted as a given constant, interest in averages of sunspots numbers over short periods of time has continued to grow. Ross (2009) reports that in 2008 the sun experienced one of the lowest numbers of sunspots in many years - the 7th lowest since 1749 next only to sunspot numbers recorded back in 1913. Apparently, accurate and reliable monitoring of the highly complex solar magnetic activity variations require, not only large amounts of data taken over a long period of time, but also adaptive and robust modelling techniques a variant of which we propose below.

2.2 Sunspots cycles vector parameterisation and key assumptions

In a 2-D space, sunspots numbers form bell-shaped distributions, suggesting that the cycles follow a fairly similar distribution – describable by the multivariate Gaussian distribution as shown in the LHS panel of Figure 2 representing two K-Means (MacQueen, 1967) clusters. The RHS panel is the corresponding density based on the Mclust(.) model-based algorithm (Fraley and Raftery, 2006) and the Gaussian mixture model

$$\prod_{i=1}^{N}\sum_{k=1}^{K}\pi_k\, f_k(S_i|\mu_k,\Sigma_k) \leftrightarrow \prod_{i=1}^{N} f_{S_i\in k}\left(S_i|\mu_{S_i\in k},\Sigma_{S_i\in k}\right) \quad (1)$$

where S_i denotes the sunspots numbers, K is the number of components, $f_{**}(.)$ is a Gaussian mixture normal distribution, π_k is the prior probability of class membership and $S_i \in k$ are class allocations. Figure 2 derives from the original data in NOOA (2012) and exhibits two cluster patterns with lower and upper means 28.88 and 110.60 respectively. Thus, if we let the cycles form a binary pattern of "lows" and "highs" we can define

$$f(s_l,s_h) = \frac{\exp\left\{-\frac{1}{2}(\sigma_{ll}(s_l-\mu_l)^2 - 2\sigma_{lh}(s_l-\mu_l)(s_h-\mu_h) + \sigma_{hh}(s_h-\mu_h)^2)\right\}}{2\pi\sqrt{|\Sigma|}} \quad (2)$$

where $s_{l.h}$ denote low and high cycles, we can use the parameter estimates $\Theta = \{\mu,\Sigma\}$ to track the behaviour of the cycles. If the cycles are correlated the density in (2) becomes

$$f(s_l,s_h) = \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(s_l-\mu_l)^2}{\sigma_l^2} - 2\rho\frac{(s_l-\mu_l)(s_h-\mu_h)}{\sigma_l\sigma_h} + \frac{(s_h-\mu_h)^2}{\sigma_{hh}}\right)\right\} \quad (3)$$

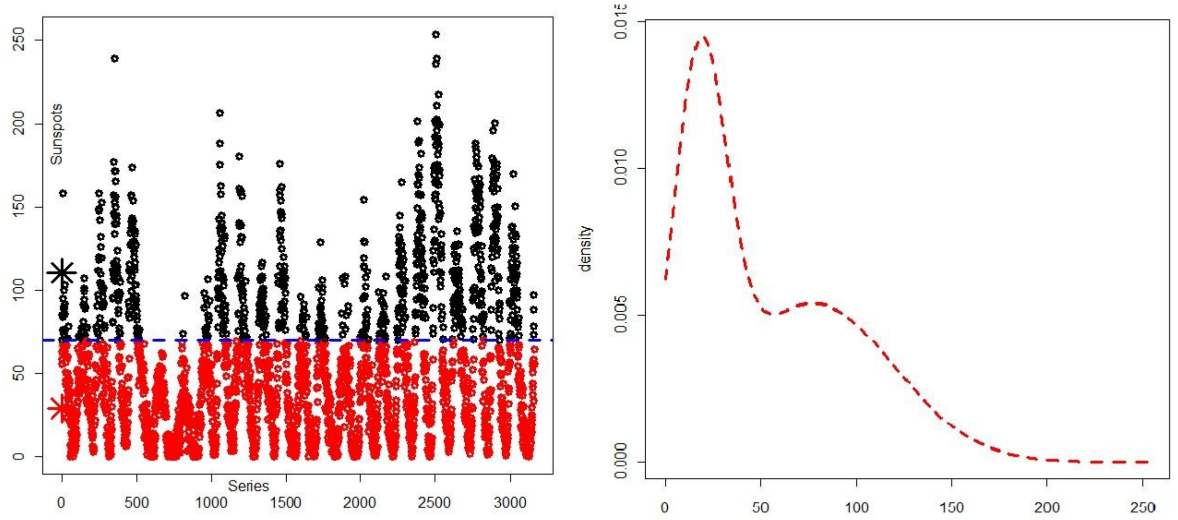

Figure 2: K-Means partitioned sunspot cycles (LHS) and corresponding densities (RHS)

We build the concept of a notionally infinite data (an infinitely long vector S) with parameters $\mu_s = \int_{-\infty}^{\infty} sf(s)\,ds \leftrightarrow \sigma_s^2 = \int_{-\infty}^{\infty}(s-\mu_s)^2 f(s)ds$. Statistically, the high-peaked (more than normal) and low-peaked (less than normal) cycles imply high and low solar activities respectively while those skewed to the right imply few increases and frequent decreases in solar activity and vice versa. Thus, we examine the initial and subsequent patterns of the cycles in order to separate the “lows” from the “highs”. Finally, the maximum likelihood estimates (MLEs) of these random finite mixture densities, are estimated and passed on to a predictive model as outlined in the algorithm below.

```
Begin
    Cluster $S_{i=1,2,\dots,N}$ into finite groups, say, $S_{i.l}$ and $S_{i.h}$: $l$ are lows and $h$ are highs
    Extract $\Theta_{l.h} = \{\mu_{l.h} \text{ and } \Sigma_{l.h}\}$
    $\omega = \arg\max_{S_i} \lambda\mu_h := \{\forall S_i |\, \mu_l \ll \mu_h: \lambda \text{ a predefined constant and } k = \{l, h\}$
    Extract $\theta_{l.h} = \left\{\mu_{S_{i.l.h[1:n]}} \text{ and } \Sigma_{S_{i.l.h[1:n]}}\right\}: n = \{n_l, n_h\}$ are cases in each cluster
    Set $\Theta = \{\Theta_{l.h}, \theta_{l.h}\}$
    Initialise $\mathrm{z_i} = \{0,1\}^2 = (\mathrm{z_{ik}}, \mathrm{z_{i\bar{k}}})$ so that $\mathrm{z_{ik}} = \begin{cases} 1 \text{ if } \left(S_{i.l.h[1:n]} \in k\right) > \omega \\ 0 \qquad\qquad \text{Otherwise} \end{cases}$
        Obtain MLE of $f_{S_i \in k}(S_i|\Theta_{l.h}, \theta_{l.h})$
            For *m:=1 to M* (Large positive integer)
                $\hat{\pi}_k := \frac{\sum_{i=1}^{N} z_{ik}}{N}$; $\hat{\mu}_k := \frac{\sum_{i=1}^{N} z_{ik} S_i}{\sum_{i=1}^{N} z_{ik}}$ and $\hat{\sigma}_k = \sqrt{\frac{\sum_{i=1}^{N} z_{ik}(S_i - \hat{\mu}_k)^2}{\sum_{i=1}^{N} z_{ik}}}$
                Update $\Theta := \Theta[m]$
                Update $z_{ik} := z_{ik}[m]$
                Update MLE of $f_{S_i \in k}(S_i|\Theta) := f_{S_i \in k}(S_i|\Theta)[m]$
            End For
        Select best model
    Use the model to predict new cycles $S^*_{i>N}$

    Output best model parameters
    Determine whether $S^*_{i>N} \in S_l$ or $S^*_{i>N} \in S_h$ based on, say, $\psi = \frac{\sum_{i=1}^{N} \mathbf{I}(s_{l.h} \le s^*_{i>N})}{N} \propto \hat{\pi}_k$
End
```

The above algorithm adapts the EM converging features described in McLachlan and Krishnan (1996) and in McLachlan and Peele (2000). Its generic form suits virtually any supervised model. This paper adopts Support Vector Machines (SVM).

2.3 Support Vector Machines (SVM) for supervised modelling

Support Vector Machines (SVM) describe a kernel-based discriminant function the mechanics of which rely on supervised learning of the underlying discriminating rules from the training data Cortes and Vapnik (1995). To put it in context, let the “high” and “low” cycles in our modified set $\{S_i, y_i : i = 1, \dots, N\}$, $y_i \in \{-1,1\}$ and $S_i \in \mathbb{R}^2$ be separable as in Figure 3. Then the points lying on H satisfy the equation $wS + a = 0$ where w is normal to the hyper-plane, $\frac{|a|}{\|w\|}$ is the perpendicular distance from H to the origin and $\|w\|$ is the Euclidean norm of w. Note that the points on the hyper-planes $\{H1, H2\}$ satisfy the equations $wS + a = \pm 1$ both with normal w and distance to the original $\frac{|\pm 1 - a|}{\|w\|}$ which means that the gap $\{H1, H2\} = \frac{|2|}{\|w\|}$. We need to find hyper-planes maximising the gap (minimising $\|w\|^2$) subject to $y_i(S_i w + a - 1 \geq 0)$. The numbers lying on $\{H1, H2\}$ are called support vectors - the core “supporters” of the optimal location of the decision surface and the hardest to classify. Intuitively, the SVM allocation rule is

$$\begin{cases} S_i \mathrm{w} + \mathrm{a} \geq +1 & \text{for } \mathrm{y_i} = +1 \\ S_i \mathrm{w} + \mathrm{a} - 1 \leq & \text{for } \mathrm{y_i} = -1 \end{cases} \leftrightarrow y_i(S_i w + a - 1 \geq 0)\ \forall_i \quad (5)$$

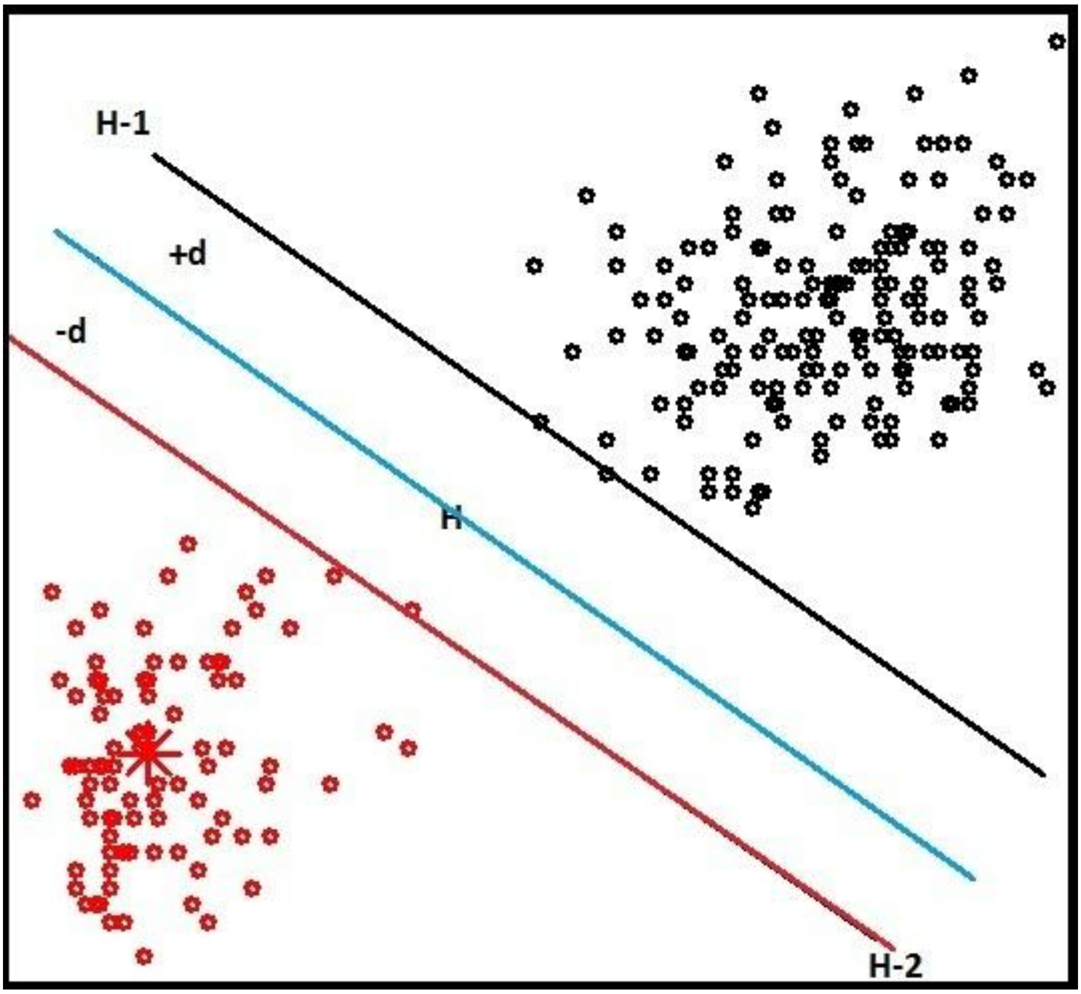

Figure 3: A graphical illustration of an SVM classifier margin in 2-D

The general formulation of SVM discriminating kernel due to Cortes and Vapnik (1995) is

$$F(S) = \sum_{i=1}^{V} \alpha_i \, \Phi(S, S_i) + a \ \ (6)$$

in which α_i represents the Lagrange multiplier summed over the values for which $\alpha_i > 0$. The upper index V denotes the number of support vectors as described above. SVM solution relies on the Lagrangian formulation of the problem – an optimisation method requiring $V \in N$ positive multipliers $(\alpha_{i=1,2,\ldots,V})$ for each of the inequalities on the RHS of Equation 5. The general formulation of the Lagrangian is as follows

$$L = \frac{\|w\|^2}{2} - \sum_{i=1}^{V} \alpha_i y_i (S_i w + a - 1 \geq 0) + \sum_{i=1}^{V} \alpha_i \ \ (7)$$

SVM solution is obtained by minimising Equation 7 with respect to w and a and simultaneously requiring that $\frac{dL}{d\alpha_i} = 0 \, \forall_i$ or equivalently maximising L and require that both w and a disappear. The latter implies that $w = \sum_i \alpha_i y_i S_i$ and $\sum_i \alpha_i y_i = 0$ transforming Equation 7 into its dual equivalent $L_d = \sum_i \alpha_i - \frac{1}{2}\sum_i \alpha_i \alpha_j y_i y_j S_i . S_j$. The SVM model weights are calculated as the product of the support vector coefficients and their values and used in forming the allocation rule. Other than the support vectors ($\alpha_i > 0$) the remaining data points have $\alpha_i = 0$ – these are those lying on the two hyper-planes $\{H1, H2\} \rightarrow y_i(S_i w + a - 1 = 0)$or beyond them if $y_i(S_i w + a - 1 > 0)$.

3 Analyses and discussions

This section provides a two-stage analysis of the sunspots data in Figure 2 using a combination of graphical visualisation and predictive modelling techniques. It seeks to establish whether sunspots follow identifiable patterns which can be used as inputs in predicting future sunspots or indeed other related phenomena.

3.1 Initial sunspots patterns and maximisation of internal parameters

Figure 4 exhibits the low and high cycles separation based on the cut-off points above alongside their corresponding overall bi-modal densities. It is based on the maximum number of sun spots reached by the full cycles and the number reached in the first 30 and 40 months. The cut-off point in the LHS panel is set to the mean of the averaged maximum early sun spots which, in this case, is 109 - separating the low cycles 1, 5, 6, 7, 9, 10, 12, 13, 14 and 16 from the highs 2, 3, 4, 8, 11, 15, 17, 18, 19, 20, 21, 22 and 23. The densities in the RHS panel exhibit the emerging bi-modality as a function of time.

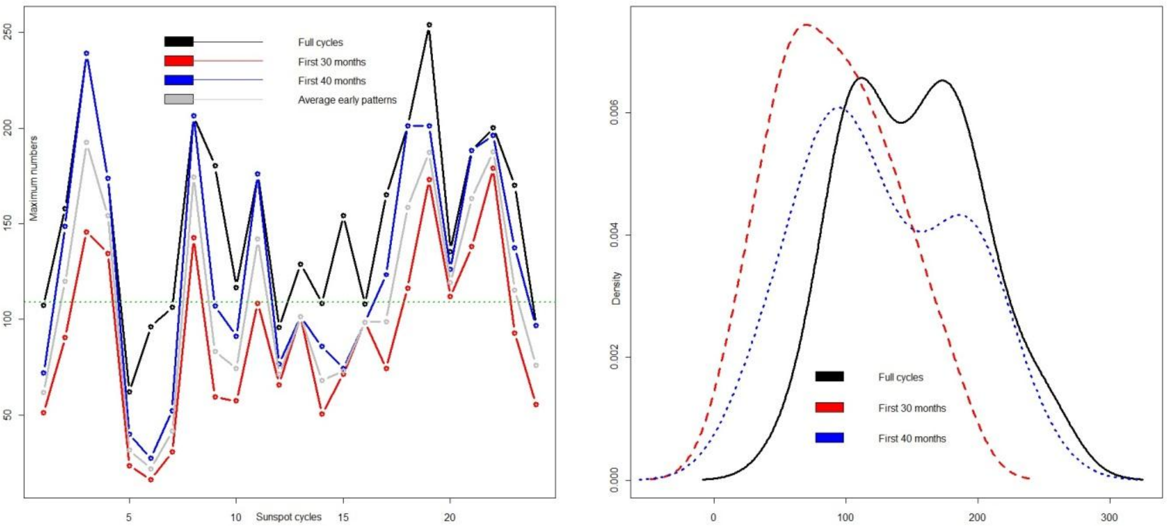

Figure 4: Omega cut-off (LHS) and the corresponding bi-modal density (RHS)

It clearly emerges from Figure 4 that each solar activity cycle can be predicted via graphical visualisation of its early patterns. In particular, the maximum values reached by each cycle appear to provide an insight into the overall activity of the cycle before it starts to subside. The foregoing structural detection of patterns in the sunspots data amounts to unsupervised modelling. Adopting these patterns as a guide to data labelling rule yields the two class priors as $\hat{\pi}_l = \frac{\sum_{i=1}^{N} z_{il}}{N} = 0.46 \leftrightarrow \hat{\pi}_h = 0.54$ with $\theta_{l.h}$ computed as above.

Both plots in Figure 5 exhibit a strong positive correlation between the sunspots mean and standard deviation vectors. Intuitively, we can reasonably focus on only one of these parameters. Implementation of the algorithm in Section 2.2 is based on those premises.

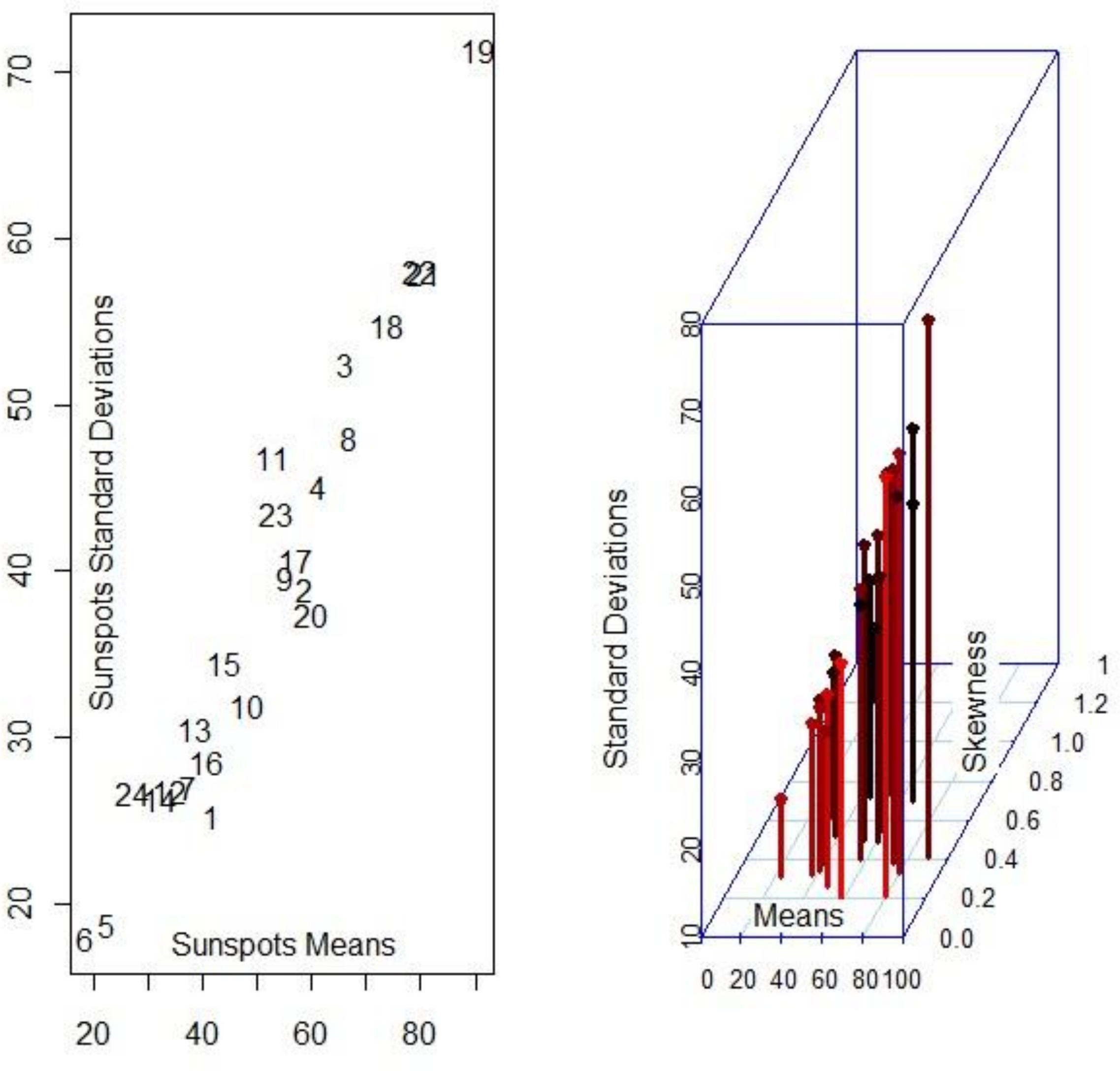

Figure 5: Positively correlated sunspots means and standard deviations

The four panels in Figure 6 are based on the maximisation of averaged estimated internal parameters – the means (north-western panel in Figure 6) and class probabilities which, in this case, are equally likely (south-eastern panel of Figure 6). Although not graphically presented here, group variances were also maximised in the same way.

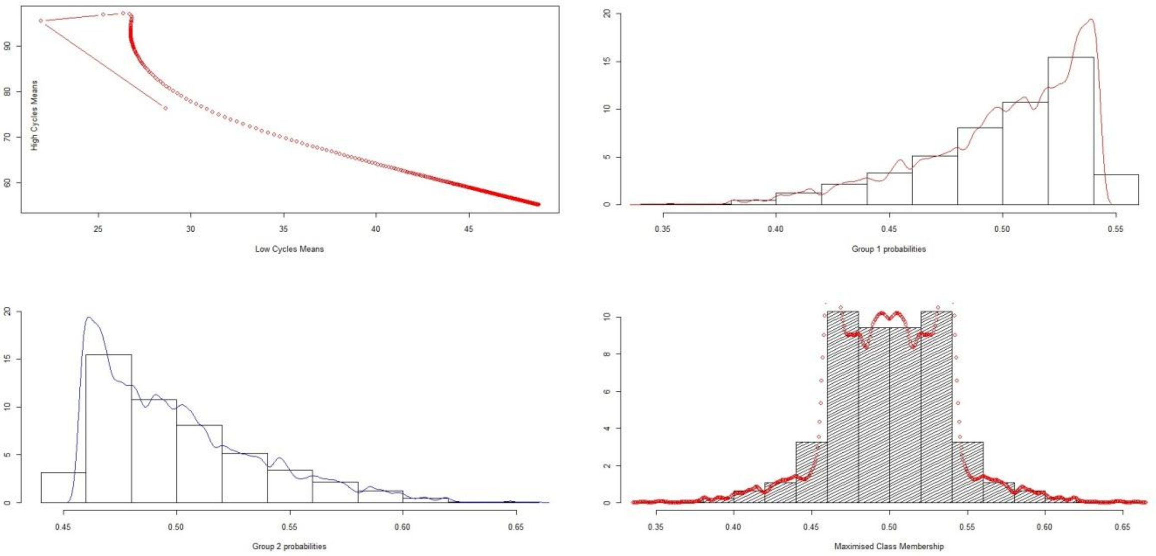

Figure 6: Maximisation of group means and class separation

Based on the established bi-modal nature of the cycles (south-eastern panel in Figure 6) and the fact that the average early patterns for cycle 24 fall below the cut-off point, it is reasonable to suggest that it will be a low activity cycle. Next, we implement SVM modelling based on the initial patterns followed by a similar implementation based on maximised estimated parameters for the purpose of re-defining the class labels.

3.2 SVM-based supervised modelling

Results from SVM modelling based on the initial class patterns with prior probabilities $\hat{\pi}_l$ and $\hat{\pi}_h$ gave an averaged accuracy of 58% on a cost range of 0.005 to 5 and a training sample of 500. Posterior class probabilities conditioned on maximised averages of the early low and high group means reached an average accuracy of 98% on the same cost range and training sample size. The resulting support vectors are shown in Figure 7 with the horizontal and vertical axes correspond to the support vectors and indices respectively.

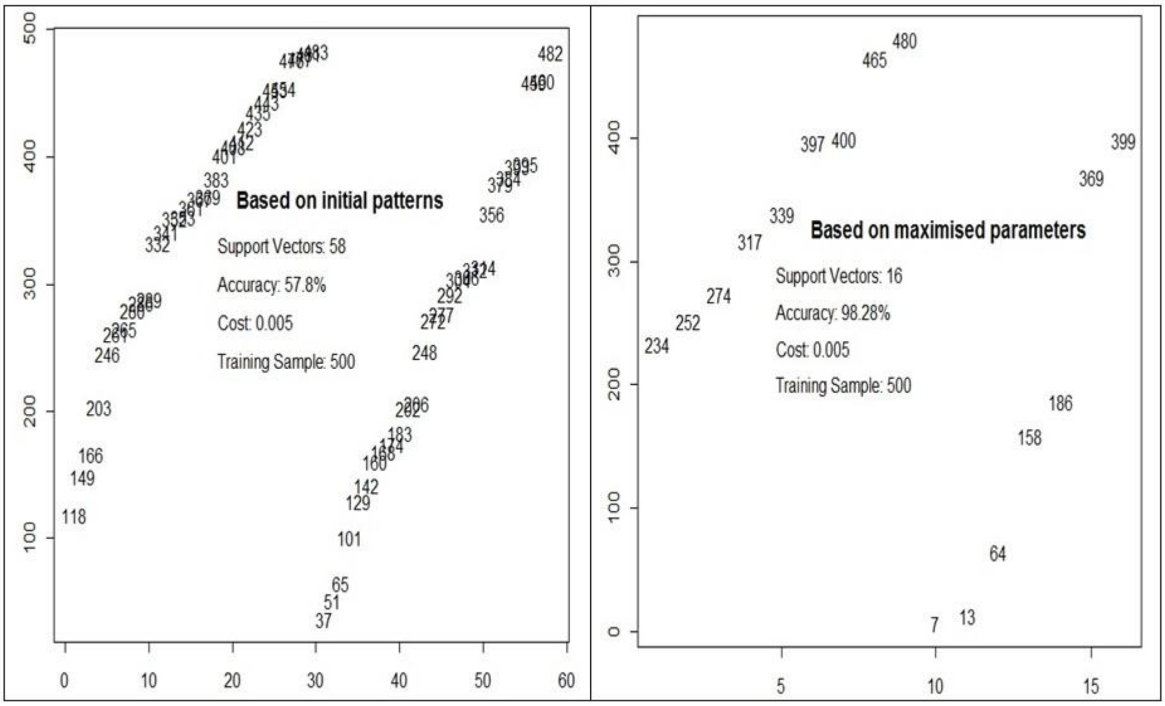

Figure 7: Support vectors for the initial patterns (LHS) and maximised parameters (RHS)

In R (2011) the SVM model weights for each of the support vectors are obtained as a cross product of the model coefficients and support vectors. Other useful SVM outputs include the individual probabilities and decision values as graphically exhibited in Figure 8. Notice the difference between the lower accuracy case (north-western panel) - highlighting the random nature of class allocation – and the higher accuracy model (south western panel) showing clear concentrations of $\hat{\pi}_l$ and $\hat{\pi}_h$ in either side of the class boundary.

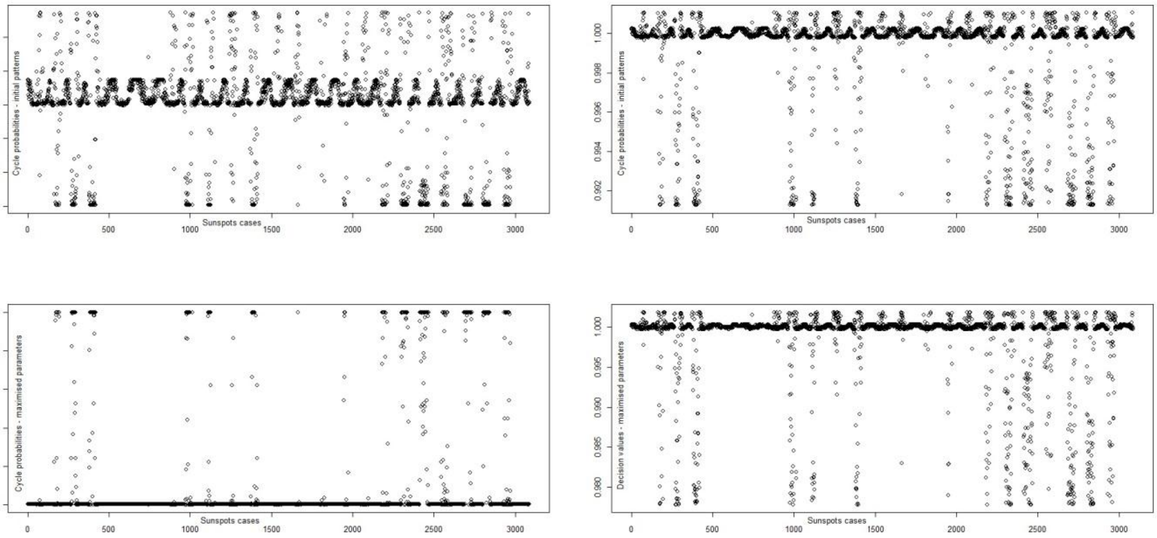

Figure 8: Sun cycles class probabilities

Both the observations corresponding to the vectors in Figure 7 and to the probabilities and decision values in Figure 8 can easily be identified by indexing.

4 Concluding remarks and potential future directions

Capturing and gaining full understanding of all the attributes which characterise a phenomenon are all it takes to describe it. Generally, empirical results are more reliable if they not only depend on the modelling tools and techniques but also on clearly defined states of the object of investigation. Predicting solar activity cycles remains one of the major challenges the scientific community faces with intricacy being compared to predicting, say, the severity of next year’s winter. In this weather analogue, if all that is available is a long vector of temperature readings over many years, the only sensible approach is to search for naturally arising structures in the data with the hope that if uncovered they may provide potentially useful information. This paper adopted the foregoing philosophy and so it sought to develop a predictive framework for modelling sunspots data using inherent Gaussian distributional properties in the data. As in Colak and Qahwaji (2009), the paper relied on a continuous flow of data for prediction, but rather than assessing model accuracy on the NOAA benchmark, an SVM model was trained and tested on a notionally infinite dataset of cycles. For the EM algorithm emphasis was on its convergence features and their relevance to the detection of solar activity cycles.

By examining multiple sets of observations from the onset of each cycle via graphical visualisation early patterns of sun cycles and their binary nature were determined. Comparing multiple early patterns for each recorded cycle extracted at different time periods to the corresponding full cycles revealed that the first 3 years provide a sufficient basis for predicting the cycle’s peak. The patterns were then adapted as inputs into an integrated unsupervised and supervised modelling algorithm. Based on multiple simulations a binary cut-off point was developed – demarcating low from high solar activity. The cut-off was then used to label the data and apply Support Vector Machines (SVM) for predicting new cycles using a novel parameter generating approach. The novel method’s mechanics are geared towards simultaneously tracing anomalies via a robust and adaptive approach. Repeated SVM runs using repeatedly improved parameters showed that the approach yields greater accuracy and reliability than conventional approaches. The paper’s main substance can be described as an enhancement of algorithmic methods for learning underlying rules from data. Although it adopted SVM for implementation, the general approach can easily be implemented in any domain-partitioning method.

References and bibliography

Choudhuri, A. R., Chatterjee, P. and Jiang, J. (2007). *Predicting Solar Cycle 24 with a Solar Dynamo Model; Physical Review Letters, Vol. 98, Issue 13, American Phys. Society.*

Colak, T. and Qahwaji, R. (2009). *Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares; SPACE WEATHER, Vol. 7, No. 12.*

Cortes and Vapnik, (1995). *Support-vector networks; Machine Learning, Vol. 20, No. 3, pp. 273-297, Kluwer Academic Publishers.*

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm; Journal of the Royal Statistical Society, Vol. 39, pages 1-38.*

Dikpati, M., de Toma, G. and Gilman, P. (2006). *Predicting the strength of solar cycle 24 using a flux-transport dynamo-based tool*

Fraley, C and Raftery, A. (2006, revised 2010). *MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering, Technical Report no. 504, Department of Statistics, University of Washington.*

Glasby, F. (2002). *Planets, Sunspots and Earthquakes: Effects on the sun, the earth and its inhabitants; iUniverse, ISBN-13: 978-0595226412.*

Izenman, A. (1983). *J. R. Wolf and H. A. Wolfer: An Historical Note on the Zurich Sunspot Relative Numbers; Journal of the Royal Statistical Society, 146, Part 3, pp 311-318.*

Kane, R. (2002). *Some Implications Using the Group Sunspot Number Reconstruction; Solar Physics, Vol. 205, No. 2, pp 383-401, Springer, ISBN 1014296529097*

Kitiashvili, I. and Kosovichev, A. (2009). *Prediction of solar magnetic cycles by a data assimilation method; Cosmic Magnetic Fields: From Planets, to Stars and Galaxies; Proceedings IAU Symposium, No. 259, Edited by Strassmeier, K, Kosovichev, A. and Beckman, J. (2009) - International Astronomical Union.*

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds L. M. Le Cam & J. Neyman, 1, pp. 281–297. Berkeley, CA: University of California Press.*

McLachlan, G. Krishnan, T. (1996). *The EM Algorithm and Extensions; John Wiley.*

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models; John Wiley.*

Mwitondi, K. and Said, R. (2011). *A step-wise method for labelling continuous data with a focus on striking a balance between predictive accuracy and model reliability; International Conference on the Challenges in Statistics and Operations Research; 08th - 10th March - 2011, Kuwait City.*

Mwitondi, K. and Bugrien, J. (2010). *Harnessing data flow potentials for sustainable applications of Science, Technology and Innovation for African Development; 22nd International CODATA Conference on Scientific Data and Sustainable Development, 24-27 October 2010, Cape Town.*

NOOA (2012). *http://www.ngdc.noaa.gov/stp/solar/ssndata.html#hemi*

Pielke, R. A., Avissar, R., Raupach, M., Dolman, A. J., Zeng, X. and Denning, A. S. (1998). *Interactions between the atmosphere and terrestrial ecosystems: Influence on weather and climate; Global Change Biology, Vol 4, Issue 5, pp 461–475.*

Qahwaji, R. and Colak, T. (2007). *Automatic Short-Term Solar Flare Prediction Using Machine Learning and Sunspot Associations; SOLAR PHYSICS, Vol. 241, No. 1, pp 195-211, ISBN 11207-006-0272-5*

Reames, D. (2002). *Magnetic topology of impulsive and gradual solar energetic particle events; The Astrophysical Journal, Vol. 571, pp 63–66.*

R (2011). *R Version 2.13.0 for Windows; R Foundation for Statistical Computing.*

Ross, J. (2009). *http://www.sott.net/articles/show/181839*

Rycroft, M. J, Israelsson, S. and Price, C. (2000). *The global atmospheric electric circuit, solar activity and climate change; Journal of Atmospheric and Solar-Terrestrial Physics Vol. 62, Issues 17–18, pp 1563–1576.*

Siscoe, G. L. (1978). *Solar–terrestrial influences on weather and climate; Climatology Supplement, Nature, Vol. 276, pp 348-352.*

Schwabe, S.H. (1843). *Astronomische Nachrichten, 20, no. 495, 234-235*

Vapnik, V. (1995). *The Nature of Statistical Learning Theory. Springer-Verlag, ISBN 0-387-98780-0*

Wolf, J. R. (1848). *Message from the observatory in Berne. Flacken solar observing.; Communications of Natural History; Society in Bern, pp 169-173.*

Wolf, J. R. (1852). *New studies of the period of Sunspots and their meanings; Communications of Natural History; Society in Bern, 255, pp 249 to 270.*