



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/240340/>

Version: Published Version

Article:

Jain, R., Kumar, A. and Tripathy, S.C. (2026) Machine learning–based characterization of solar p-mode frequency shifts during solar cycle 25. *Solar Physics*, 301 (4). 62. ISSN: 0038-0938

<https://doi.org/10.1007/s11207-026-02660-y>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Machine Learning–Based Characterization of Solar p -Mode Frequency Shifts During Solar Cycle 25

Rekha Jain¹ · Akash Kumar² · Sushanta C. Tripathy³

Received: 23 February 2026 / Accepted: 10 April 2026
© The Author(s) 2026

Abstract

The solar interior is probed by the properties of the Sun's acoustic oscillations (p -modes) observed on the solar surface. The frequencies of these p -modes measured in the last three decades show long term variation similar to the ~ 11 year cyclic behaviour exhibited by 10.7 cm radio flux, sunspot numbers, and other solar activity indices. It is also now established that the cyclic behavior of some of the solar proxies is connected with geomagnetic activities and has implications for space weather. Hence, in recent years efforts have been made using machine-learning methods to forecast these solar proxies with a view to improve our understanding of space weather. Developing a comparable method for forecasting p -mode frequency shifts is therefore of interest for two reasons. Firstly, it will facilitate future investigations into its potential role in tracing energy drivers from the Sun's interior to the geospace response by improving models of solar interior dynamics to coronal and heliospheric plasma conditions. In other words, it will help establish a more robust and quantitative link between the Sun's interior and its exterior. Secondly, it may provide us with an independent indicator or an early indicator of ascending and descending phase of solar activity which might be useful for space weather forecasting.

In this article, we develop and apply the standard time-series analysis and machine-learning based methods to characterise p -mode frequency shifts for the remaining Solar Cycle 25.

Keywords Sun: helioseismology · Activity · Magnetic · Sunspots · Radio flux

1. Introduction

The 11 year activity cycle of the Sun (Schwabe 1843) is characterized by many proxies such as sunspot numbers, 10.7 cm radio flux, flares, geomagnetic storms etc. Ever since the sunspot-records began, the cycles have been numbered starting from Solar Cycle 1 for

✉ R. Jain
R.Jain@sheffield.ac.uk

¹ School of Mathematical and Physical Sciences, University of Sheffield, Sheffield, S3 7RH, UK

² School of Mechanical, Aerospace and Civil Engineering, University of Sheffield, Sheffield, S1 3JD, UK

³ National Solar Observatory, Boulder, CO 80303, USA

the years 1755–1766. The current solar activity cycle is Solar Cycle 25. These activity cycles are studied with great interest for understanding the Sun and their influence on space weather (e.g. Mursula, Usoskin, and Maris 2007; González Hernández et al. 2014; Baker 2000; Hajra et al. 2021).

The oscillation frequencies observed on the Sun's surface also rise and fall in step with the Sun's activity cycle. These oscillations, commonly referred to as p -modes, are acoustic resonances within the Sun, driven by turbulent convection in its convection zone. Their main restoring force is pressure gradient.

These acoustic oscillations were first detected by Leighton, Noyes, and Simon (1962) (see also, Ulrich 1970). However, the variation in frequencies due to changes in magnetic flux were reported much later in 1985 by Woodard and Noyes (1985). It is now well established that the p -mode oscillations of low-degree modes have a slightly higher frequencies upto ~ 3.7 mHz, during the increased solar magnetic activity compared to when the magnetic activity is low (see for example, Libbrecht and Woodard 1990; Elsworth et al. 1990). Above ~ 3.7 mHz, a downturn is observed (Salabert et al. 2004) and the frequency shifts start to decrease with even a reversal in sign above the acoustic cut-off (Rhodes, Reiter, and Schou 2002). The increase in the frequency shifts below the ~ 3.7 mHz suggests changes near the photosphere with solar activity. The decrease above 5.0 mHz is likely due to changes in the chromosphere. Theoretical models have been developed with a view to understand the physical mechanism for such rise and fall in the frequencies. For example, Goldreich et al. (1991) have suggested changes in the mean square magnetic-field strength in the photospheric region. However, as pointed out by these authors such mean field strength perturbations necessary to explain the observed frequency shift would require magnetic-flux tubes to continue to strengthen to larger depths than expected. Jain and Roberts (1993, 1996) attributed the observed frequency shifts as a consequence of both magnetic and thermal changes, the ascending phase due to an increase in the mean photospheric magnetic field and the descending phase from a combination of an increase in mean chromospheric magnetic-field strength and an increase in chromospheric temperature. Such a combination is able to produce results similar to what is seen in the observations but the required temperature changes in their model seem quite large. Although the physical cause of the perturbation has not yet been firmly established, p -mode frequency shifts are believed to be caused by the changes in the plasma properties through which the waves propagate or due to direct effect of magnetic fields on the waves through Lorentz force or both (Goldreich et al. 1991; Jain and Roberts 1993). In addition to the frequencies, the mode line widths and amplitudes have been found to vary with magnetic activity cycle (Kiefer et al. 2018, and references therein).

It is now well established that magnetic activity observed at the solar surface is intimately coupled to the physical processes and dynamical conditions operating within the solar interior. However, there is no full consensus on how exactly magnetic fields are generated in the interior, transported to the surface and then some magnetic flux gathering into organized and compact sunspots. Even the location and precise nature of the dynamo is still under debate (Charbonneau 2020). Some theories suggest that the dynamo site is located at the base of the convection zone, generally referred to as tachocline. The position and the physical conditions of the base of the convection zone have been shown to change with the solar cycle (Basu 2021). As mentioned earlier, it is also argued that there is another site of dynamo, in addition to tachocline which is operating near the Near-Surface Shear Layer (NSSL: Benevolenskaya 1998; Jain et al. 2022). The inference of two distinct dynamo sites is further motivated by the presence of shorter-period variations—on the order of 2–3 years—in helioseismic observations (Fletcher et al. 2010; Mehta et al. 2022; Jain, Chowdhury, and Tripathy 2023) and other solar activity indices (for a review, see Bazilevskaya et al. 2014),

superimposed on the dominant 11-year cycle. These shorter-term modulations, known as quasi-biennial oscillations (QBOs), have gained increased attention because such periodicities may provide a significant diagnostic for evaluating solar dynamo models.

Since both, frequencies of p -mode oscillations and the strong magnetic-flux regions seen on the solar surface have their origin in the interior and show cyclic and in-phase variation with time, it raises an important question: do strong magnetic-flux regions contribute more significantly to the observed p -mode frequency shifts than other sources of perturbation? It is therefore, interesting to examine other perturbations that connect p -mode properties and magnetic activity cycle. Frequencies of solar p -modes have been used in measuring changes in the large-scale flows. The meridional flows that transport near-surface material towards the poles are an essential component of some solar-dynamo theories and may play a role in filtering dynamo models. It has been shown from helioseismic measures that meridional-flow speed changes with the Sun's magnetic activity cycle. Torsional oscillations which are helioseismic measures of near-surface zonal flows (Howard and Labonte 1980; Howe et al. 2018) also vary with solar cycle. Kosovichev and Pipin (2019) suggest that variation in zonal flows may carry the signature of a dynamo wave generated at the base of the convection zone (see also, Mandal, Kosovichev, and Pipin 2024). Thus, changes in p -mode frequencies with solar cycle may hold keys for constraining solar dynamo models.

The variation of solar p -mode frequencies is revealing information about various aspects of flows and magnetism in the Sun and is therefore, important to study. In this article, we forecast variation of p -mode frequency shifts for Solar Cycle 25 by using standard time-series analysis techniques such as wavelet and a combination of locally estimated scatterplot smoothing (LOESS), Fast Fourier Transform (FFT). Deep-learning method with neural basis expansion analysis (N-BEATS) is also used separately. In the wavelet transform method, we use `pywt.waverec` (function in python library `PyWavelets`) to extract low frequency trend and high frequency noise. After smoothing the trend, we use continuous wavelet transform `pywt.cwt` (function in python library `PyWavelets`) to search for repeating patterns both at short and long term cycles. For an independent comparison, we also examine the predictions by using a combination of LOESS and FFT methods, the figures for which are denoted by `LOESS_FFT`. The `LOESS_FFT` method involves combining LOESS for smoothing and FFT for frequency analysis. The prediction is then the average of the two. For nonlinear pattern fitting in both cases, we use the Light Gradient-Boosting Machine (LGBM). Deep-learning methods are also known for powerful learning capabilities trained on long time series to fit complex nonlinear relations. Su et al. (2023) recently used this method with the neural basis expansion analysis (N-BEATS) for solar-cycle prediction. We also employ this method for the longer time series of sunspot numbers and radio flux. These are applied to three different datasets: (i) p mode frequency variations measured in Solar Cycles 23, 24 and 25, i.e. 7 May 1995 to 31 July 2025 (ii) prediction of 10.7 cm radio flux using the flux dataset from 9 January 1954 to 31 July 2025 and then forecasting the p mode frequency shifts using its best-fit relationship with the flux (iii) prediction of sunspot numbers using the sunspot number (SSN) data from 9 January 1954 to 31 July 2025 and then forecasting the p mode frequency shifts using its best-fit relationship with the SSN. In Section 2, we describe the datasets. In Section 3, we briefly mention the data analysis techniques that are used here. The various results obtained from the three datasets are then shown in Section 4 with Discussion and Conclusions in Section 5.

2. Data

We use p -mode frequency data from Global Oscillation Network Group (GONG). GONG is a ground based network that measures Doppler velocities of the solar surface from six different stations located around the world (Harvey et al. 1996). GONG has been providing coherent data for mode frequencies from 1995 to now. The GONG data analysed here covers the period from 7 May 1995 to 31 July 2025; this period covers two complete Solar Cycles 23 and 24 and part of the Solar Cycle 25.

GONG routinely provides 108-day mode frequency tables with 36-day overlaps. To investigate short-term variations, we instead use frequencies computed from nine-day time series corresponding to the Legendre's polynomial expansion.¹ Details of the frequency-extraction procedure are described in Tripathy et al. (2007). Because the measured mode frequencies depend on the mode mass, we follow the method of Woodard et al. (1991) and define the mean frequency shift ($\delta\nu$) as the mode-inertia-weighted sum of the individual frequency measurements,

$$\delta\nu(t) = \sum_{n,\ell} \frac{Q_{n\ell}}{\sigma_{n,\ell}^2} \delta\nu_{n,\ell}(t) / \sum_{n,\ell} \frac{Q_{n,\ell}}{\sigma_{n,\ell}^2}, \quad (1)$$

where $\sigma_{n,\ell}$ is the uncertainty in the frequency measurement for a given mode defined by the radial order n and spherical harmonic degree ℓ , $\delta\nu_{n,\ell}(t)$ is the change in a given mode, and $Q_{n,\ell}$ is the mode inertia ratio defined by Christensen-Dalsgaard and Berthomieu (1991). The reference frequency is chosen to correspond to the minimum value in the time series. The analysis uses common modes with frequencies in the range $1500 \leq \nu \leq 5500$ μHz and degrees spanning $0 \leq \ell \leq 100$.

Since the discovery of the relationship between solar radio emissions and sunspot numbers in 1947 by Covington (1947), 10.7 cm radio flux (hereafter referred to as F10) has been used as a well known indicator of solar activity. F10 corresponds to a wavelength of 10.7 cm and gives a measure of the total daily emissions of radio waves within frequencies $2700 \leq \nu \leq 2900$ MHz, observed from the solar disk² (Tapping 2013) and is measured in solar flux units (sfu), where $1 \text{ sfu} = 10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$. It mainly represents the contributions from sunspots and radio plagues in the upper chromosphere in addition to the quiet-Sun background emission (Covington 1969).

The location and the number of sunspots observed on the solar surface vary as solar cycle progresses. These are recorded daily in both hemispheres from the visible part of the solar disk³ (Clette and Lefèvre 2015).

Figure 1 displays the Sunspot numbers (top), 10.7 cm radio flux (middle) and p -mode frequency shift, $\delta\nu$ (bottom) as a function of time starting from May 1995. All three dataset shown are averaged over the same 9 day intervals. It is evident that all these quantities exhibit similar temporal variation.

3. Data Analysis Techniques

The p -mode frequency shifts have long been known to correlate with 10.7 cm radio flux (Woodard and Noyes 1985) and the sunspot numbers (see Rhodes, Cacciani, and Korzennik

¹ gong2.nso.edu/ftp/TSERIES/v1z.

² www.spaceweather.gc.ca/forecast-prevision/solar-solaire/solarflux/sx-5-en.php.

³ www.sidc.be/silso/datafiles.

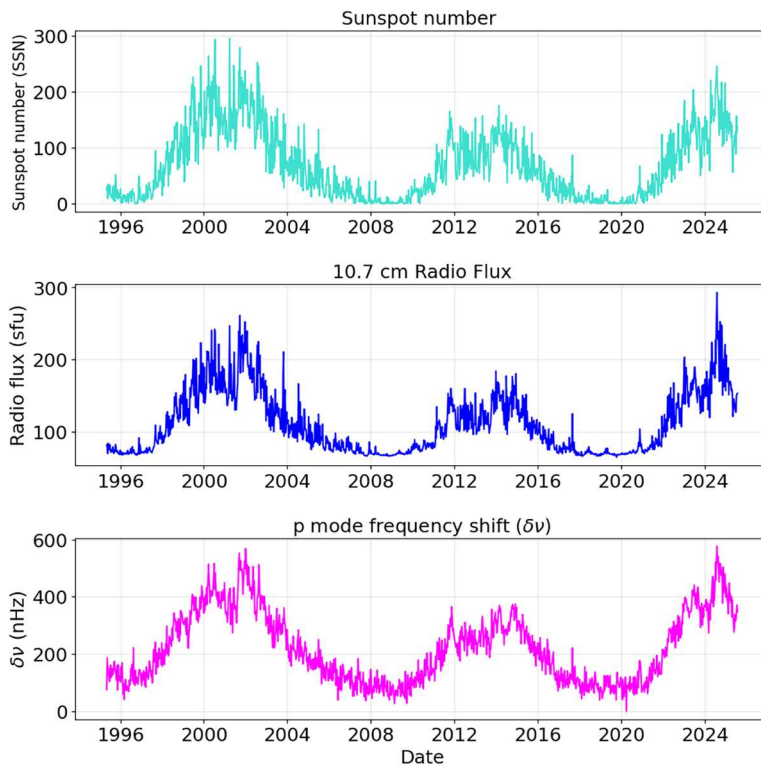


Figure 1 Temporal variation of Sunspot numbers, 10.7 cm radio flux and p -mode frequency shifts, all averaged over the same 9 day period.

1991). The data we use for the radio flux and sunspot numbers recorded are averaged over a 9-day time period to match up with the GONG time series for the p -mode frequencies. Note that the sensitivity of the p mode frequencies to the change in the magnetic activity and other activity proxies have also been explored and found to be similar with other instruments (see for example Salabert et al. 2004; Rabello-Soares, Korzennik, and Schou 2006; Chaplin et al. 2007; Jain, Tripathy, and Hill 2009; Larson and Schou 2018) making them quite robust. Figure 2 displays frequency shifts and the 10.7 cm radio flux (left panel) and sunspot numbers (right panel) with red filled circles. While most of the studies involving p -mode frequency shifts and solar activity indices report high linear correlation, deviations from a simple linear relation have also been reported indicating a saturation effect at high solar activity (Rabello-Soares, Korzennik, and Schou 2006; Chaplin et al. 2007; Rabello Soares 2019). Since Figure 2 also suggests a saturation effect during periods of maximum activity, we fit a polynomial of degree 2 to each dataset. The resulting fitting is displayed as a blue solid line (left panel) for F10 and a turquoise solid line (right panel) for the sunspot numbers. It is clear that the quadratic fit reproduces the higher-magnitude data more accurately than the linear fit (the deviation is small); therefore, we adopt the quadratic form for subsequent analysis.

The relationship is calculated only for datasets from May 1995 onwards since the p -mode frequencies from GONG are only available for this period. We will assume that this relationship is also valid before 1995. On that basis, even though the F10 data is available

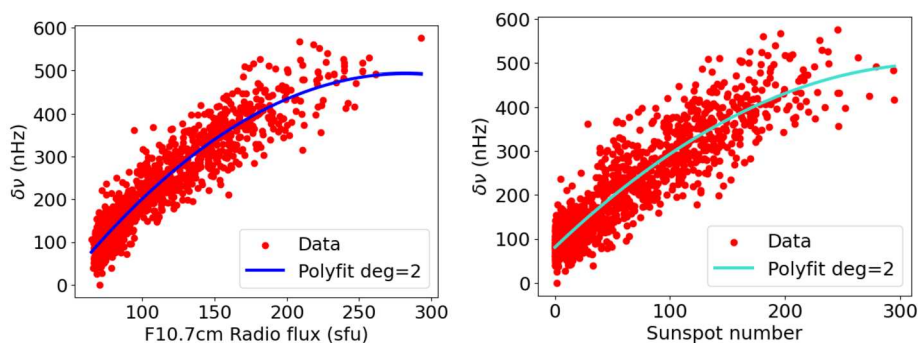


Figure 2 Left: Relation between the p -mode frequency shifts and F10.7 cm radio flux. Right: Relation between the p -mode frequency shifts and sunspot numbers.

for a longer period starting from February 1947, we use datasets of F10 and SSN, from 9 January 1954 (since the minimum of the cycle starts around this period) to 31 July 2025 to forecast the p -mode frequency shift in the current Solar Cycle 25.

The three parameters also have different kinds of errors as they are measured by different instruments; only data points from ~ 2.5 cycles can (at this stage) be used for the best fit. There is no way to quantify the scatter in a meaningful way, so no error bars are shown in the figures and the results should be interpreted with caution. The fits are expected to improve as more data will become available in the future. Since we are interested in qualitative behaviour, we will proceed with the quadratic fit that we have obtained and use the following techniques for forecasting.

3.1. Machine Learning-Based Frameworks

In order to predict the timing of the next minimum in the p -mode frequencies, two methodologies have been applied to the time series processing (i) a standalone Wavelet-LGBM model (ii) an ensemble consisting of LOESS-LGBM and FFT-LGBM models. Both methodologies consist of signal decomposition, feature engineering techniques and model training and evaluation.

3.1.1. Signal Decomposition

The 9-day averaged time series of 10.7 cm radio flux and the sunspot numbers, each has a periodicity of approximately 11 years with dissimilar fluctuations. As is customary, we decompose the signal into trend and residual components so that models can learn the temporal and seasonal dependencies more effectively. We applied three decomposition techniques to each of the time series. A brief description of each method in the context of our data is given below.

Wavelet-LGBM: Wavelet transformation decomposes signals into multiple frequency components using localised wavelets, providing high resolution time and frequency information (see Shao et al. 2017). In wavelet-LGBM model, as the name suggests a wavelet technique has been used for decomposition. The trend has been obtained using the Discrete Wavelet Transform (DWT) using ‘db4’ (Wasilewski 2012). The residual component is obtained by subtracting the obtained trend from the original timeseries.

LOESS_LGBM: LOESS (locally estimated scatterplot smoothing) is a non-parametric technique that fits low-order polynomials to extract smooth trends in both stationary and non-stationary signals. It has been used widely in utility forecasting (Zhang and Li 2021), traffic prediction (see for example Chen et al. 2019; Zhao et al. 2022; Yuan et al. 2025), economic analysis (Kehinde, Ayinde, and Lukman 2020) etc. In LOESS_LGBM model, we decomposed the signal into trend and residuals by using the optimally selected smoothing fraction based on the minimum error. For each smoothing fraction, the root mean squared error (RMSE) was calculated, and the fraction yielding the minimum error was selected as the final model parameter. The residual component is then obtained by subtracting the best fit trend from the original timeseries.

FFT_LGBM: In this model, Fast Fourier Transform (FFT) is used as low pass filter to obtain the trend of the timeseries. It filters out spectral components exceeding 0.01 cycles/9 day and zeroing them, thus removing noise or very small fluctuations. Such a cut-off frequency enables us to retain variations longer than 2.5 years in this data set. With these filtered spectral coefficients, Inverse FFT is used to reconstruct the trend component. The residual component is obtained in the same manner as performed for wavelet or LOESS-based decomposition.

3.1.2. Feature Engineering

The extracted components are further processed through feature engineering to incorporate temporal lags and seasonality, providing models with more context for better predictions.

Temporal Lags: Lag features are generated for both the trend and residual components such that at each time t , each component was augmented with n previous observations ($t - 1, t - 2, \dots, t - n$).

Seasonality: In Wavelet and Loess-based models, seasonality is accounted for by incorporating sinusoidal terms having a period determined by identifying the top spectral peaks in the wavelet power spectrum, which are then averaged to provide a mean period. In the FFT-based model, Power Spectral Density (PSD) was obtained using FFT to get the dominant frequencies. These are then converted into time period using $1/f$ and subsequently averaged to get the mean period for the seasonal term.

Uncertainty Quantification: Due to inherent uncertainties in the time series of $\delta\nu$, SSN or F10, prediction intervals are also created using LGBM-based quantile regression for all three models. Specifically, the 5th and 95th percentiles of the data are used to predict the lower and upper bounds of the forecast, respectively (creating a 90% prediction interval).

Model Training and Validation: For each of the methods described in Section 3.1.1, separate LGBM models are trained; one for the trend component, and another for the residual component. For Wavelet-LGBM methodology, the final forecasted values are obtained by adding the predicted trend and the predicted residual value. For ensemble methodology of LOESS_LGBM and FFT_LGBM, the forecasted values are obtained for each model by adding their predicted trend and residual value. Then, the final forecast (Y_{ensemble}) from LOESS (Y_{LOESS}) and FFT (Y_{FFT}) based models is averaged by averaging the predicted

$$Y_{\text{ensemble}} = \frac{Y_{\text{LOESS}} + Y_{\text{FFT}}}{2}.$$

To preserve the temporal dependencies, k-fold validation which involves shuffling, was avoided (Su et al. 2023). Instead hold-out validation strategy was employed using an 85:15 split ratio for the train and test data, ensuring the training dataset precede test dataset. Models' performances were evaluated on test data using RMSE. Upon satisfactory results, models are retrained on the entire timeseries data to get future predictions.

3.1.3. Deep Learning: N-BEATS

We use deep learning N-BEATS method which is a multi-layered prediction method (Oreshkin et al. 2019) and has two components: trend forecast component and residual forecast component. For trend forecast, stacks of 'trend' and 'seasonality' were specifically chosen as trend and periodic oscillations. For residual forecast, since it is inherently noisy, stack of 'identity' is used to model the fluctuation. Forecasts from both the models are aggregated back to give complete forecast. Such a method has been recently implemented by Su et al. (2023) for the prediction of the variation of the sunspot numbers for Solar Cycle 25. We only use this method for $\delta\nu$ forecast using SSN and F10 correlations. We do not use this method to forecast $\delta\nu$ directly from $\delta\nu$ dataset as $\delta\nu$ dataset is not long enough to use deep-learning method.

4. Results

4.1. Direct Prediction of p -Mode Frequency Shift ($\delta\nu$)

Figure 3 displays the temporal variation of frequency shift, $\delta\nu$. The pink shaded region displays the forecasted $\delta\nu$. These forecasted $\delta\nu$ are obtained from the original $\delta\nu$ dataset covering the period 7 May 1995 – 31 July 2025. The 95 percentile confidence level is shown in grey. The top panel shows the full dataset, original observed dataset in magenta and forecasted $\delta\nu$ from Wavelet+LGBM technique in black. The bottom panel is same as the top except the method used is LOESS_FFT+LGBM. The forecasted $\delta\nu$ suggests that the Solar Cycle 25 is already in the descending phase and the cycle minimum is expected in the year 2031. Since the original dataset comprises of only two complete solar cycles, Solar Cycles 23 and 24, and partial Solar Cycle 25, predictions beyond the next few years should be interpreted with caution.

4.2. Prediction of $\delta\nu$ Based on the Relationship with 10.7 cm Radio Flux

Here, we first find the prediction of the 10.7 cm radio flux and then compute the scaled-forecast of $\delta\nu$ from this prediction. Recall that the radio flux data used in this study covers the period from 9 January 1954 to 31 July 2025. Note that each data point represents the average value over a consecutive 9-day interval. The forecast is obtained using various methods: wavelet+LGBM, LOESS_FFT+LGBM and N-BEATS. These are shown in Figure 4. The vertical dashed blue line separates the training (magenta) and test (blue) dataset. The test dataset is in very good agreement with the original data in the top and middle panel. The N-BEATS method, shown in the bottom panel, underestimates the ascending part. However, the predicted F10 displayed here in black within pink shaded area is qualitatively similar in all three panels. We, then make use of the polynomial fit obtained in Figure 2 along with these forecasts for F10 to compute the scaled-forecast for frequency shifts $\delta\nu$. The computed scaled-forecast is displayed in Figure 5 in the grey shaded region

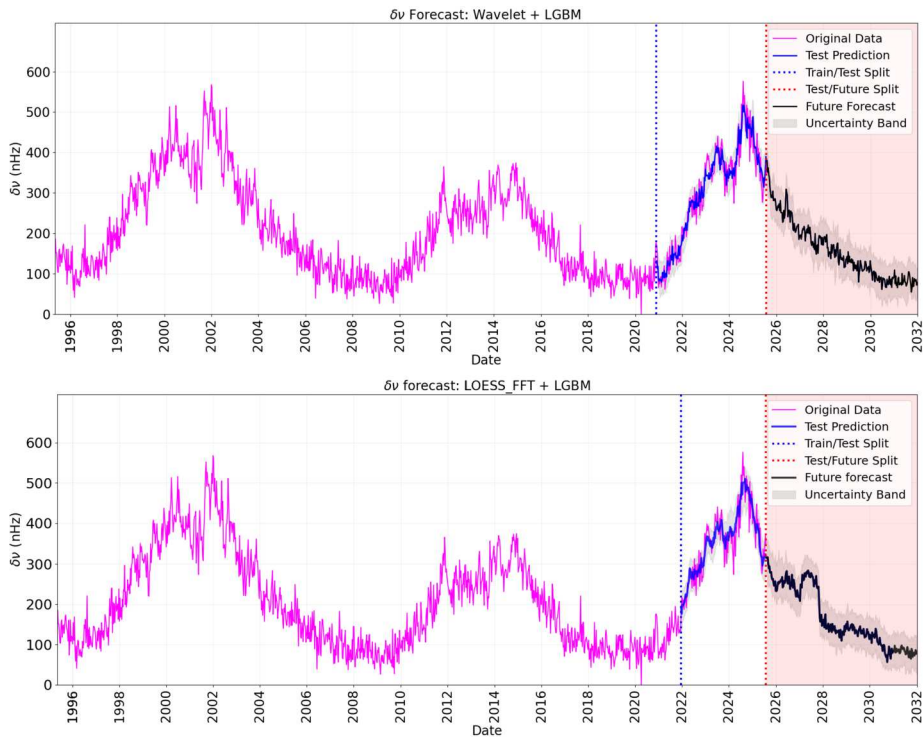


Figure 3 Temporal variation of averaged p -mode frequency shift, $\delta\nu$. The pink shaded region displays the forecasted $\delta\nu$ with 90 percentile confidence level in grey. The top panel shows forecast using wavelet+LGBM, the bottom panel is obtained from using LOESS_FFT+LGBM. The forecasted $\delta\nu$ in all panels are obtained by using $\delta\nu$ data from 7 May 1995–31 July 2025.

with black solid line. The top, middle and bottom panels correspond to wavelet+LGBM, and LOESS_FFT+LGBM and N-BEATS methods respectively. The pink shaded region around the black curve indicates 90 percentile confidence interval. It is apparent from Figure 5 that the fluctuations in the scaled-forecast are different for different methods. The wavelet+LGBM (top) shows smoother descent compared to LOESS_FFT+LGBM (middle) which is steeper. The N-BEATS method (bottom) shows slightly different decline to the other two but all three methods suggest a minimum $\delta\nu$ around 2030.

4.3. Prediction of $\delta\nu$ Based on the Relationship with Sunspot Number

In order to compare the scaled-forecast of $\delta\nu$ between the SSN and F10, we use SSN dataset from the same period i.e. from 9 January 1954 to 31 July 2025. We also average SSN within this period over the same consecutive 9-day interval as F10 and the p mode frequencies. Figure 6 shows the forecasted SSN in SC25. Using the correlation of SSN with p -mode frequencies from Figure 2, we now compute the scaled-forecast of $\delta\nu$ and display it in Figure 7. The three methods yield scaled-forecast which is different in detail. The minimum is between 2030–32 in the top and bottom panel but the middle panel has slightly larger fluctuations. Overall, the minimum is similar (~ 2030) to the one obtained using F10 shown in Figure 5.

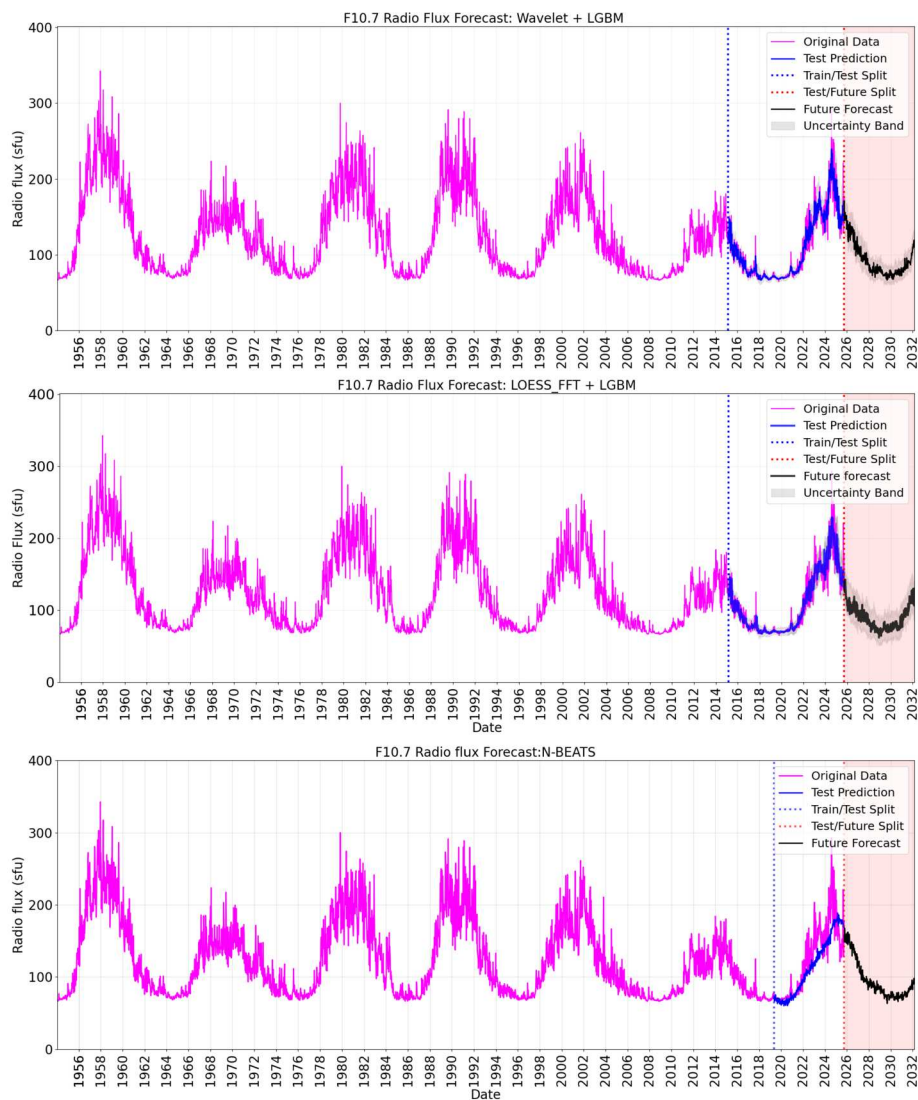


Figure 4 Forecast of 10.7 cm radio flux using Wavelet+LGBM technique (top), LOESS_FFT+LGBM technique (middle) and N-BEATS (bottom).

5. Discussion and Conclusion

Solar p mode frequency shifts exhibit ~ 11 year periodicity similar to many magnetic and related activity cycles. This is evident for Solar Cycle (SC) 23 and 24. In this paper we forecast the frequency shifts for the remaining part of Solar Cycle 25 by making use of the standard time-series analysis and machine-learning methods. We obtained forecasts in three different ways: (i) direct forecast using measured frequency shifts of Solar Cycles 23, 24 and 25 (partial), (ii) scaled-forecast using the best fit with 10.7 cm radio flux data and (iii) scaled-forecast using the best fit of frequency shifts with sunspot numbers. From Figures 3,

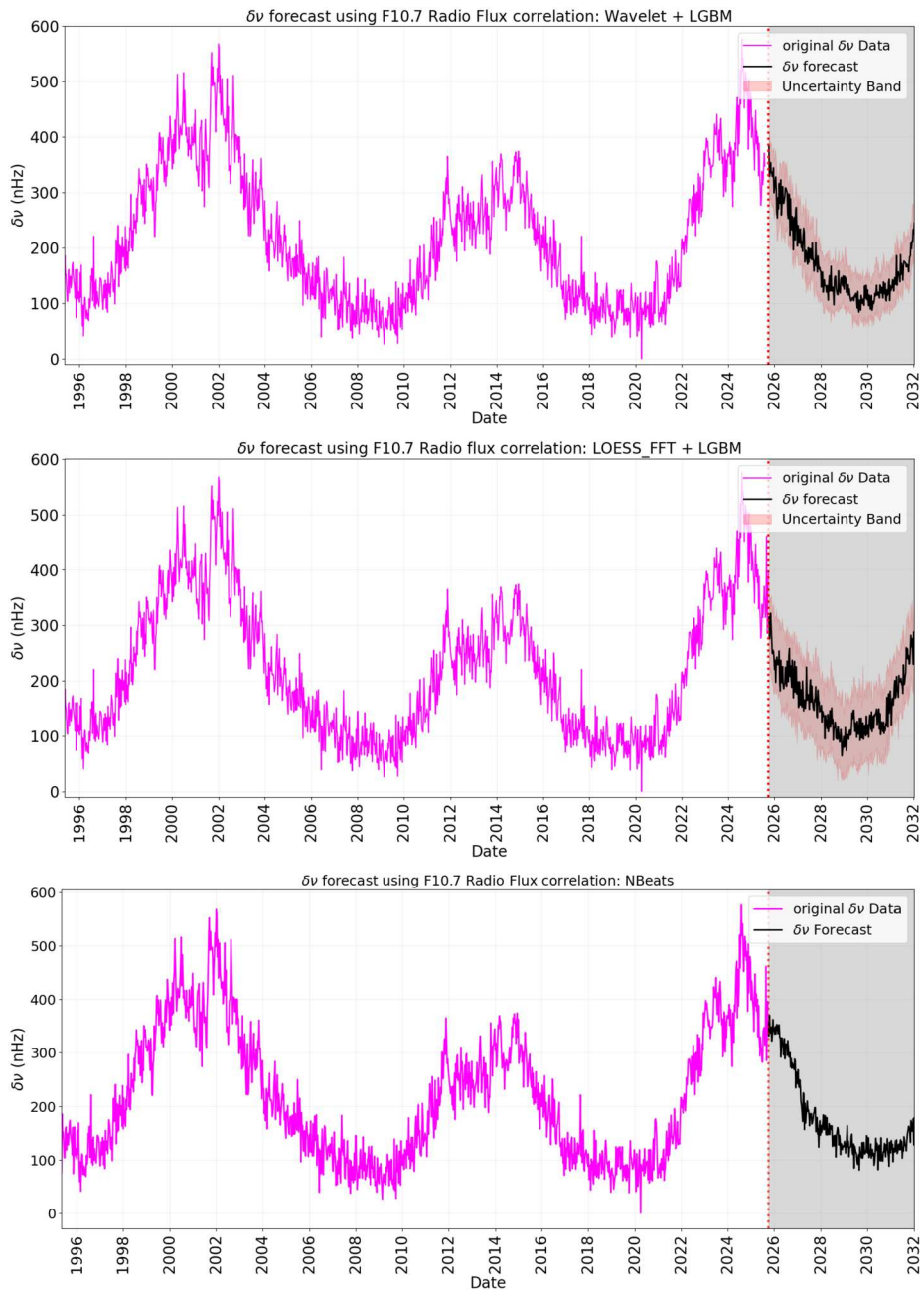


Figure 5 Scaled-forecast of p mode frequency shift using the relationship between F10 and $\delta\nu$. The forecast is obtained using Wavelet+LGBM technique (top), LOESS_FFT+LGBM technique (middle) and N-BEATS (bottom).

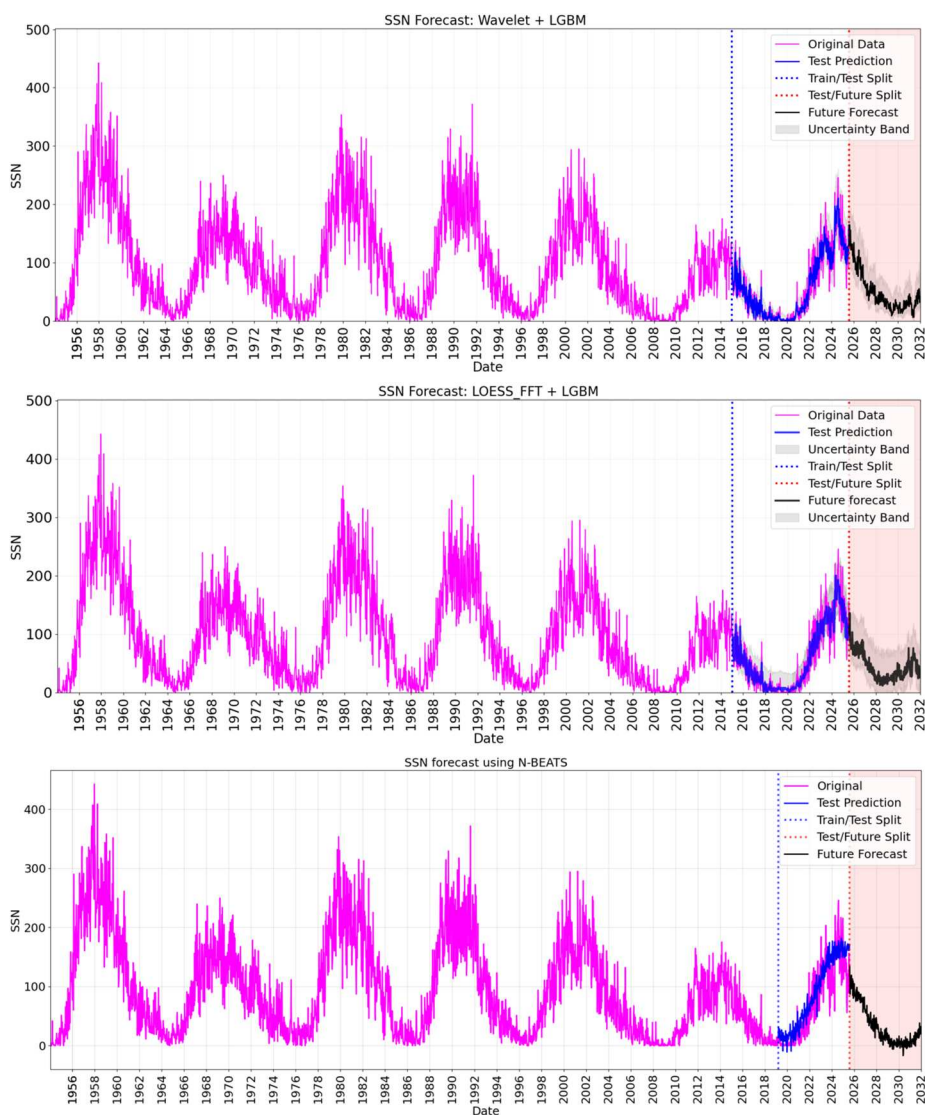


Figure 6 Forecast of sunspot numbers using Wavelet+LGBM technique (top) and LOESS FFT+LGBM technique (middle) and Deep learning-N-BEATS method (bottom).

5 and 7, it is obvious that the descending phase of SC25 lasts for a longer period of time in (i) when compared to (ii) and (iii). This is likely due to fewer number of data points in (i). The time series used to directly forecast p -mode frequency shift, from 1995–2025, is too short for the next 11 years prediction and therefore, not accurate enough to reliably point out the exact month and year of the minimum between SC25 and SC26. Based on Figure 3, an approximate time period of 2031 can be seen here. The scaled-forecast from F10 and SSN yields the minimum in the year 2030. Xu, Jain, and Xing (2024) likewise predict a comparable minimum period for SC26 for monthly sunspot numbers. In Table 1, we mention the minimum period of $\delta\nu$ for cycles 23, 24, 25 from the original data of $\delta\nu$ as shown

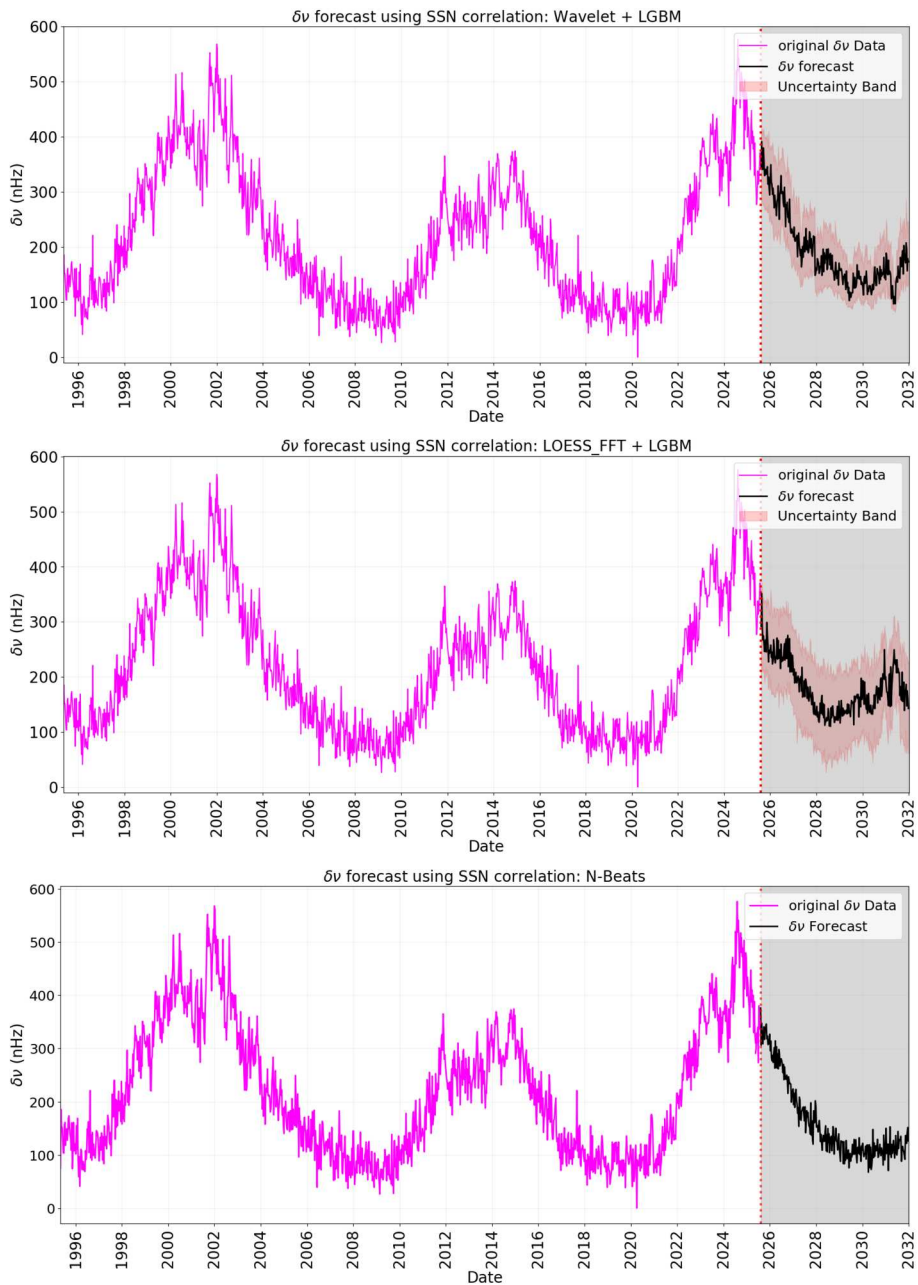


Figure 7 Scaled-forecast of p -mode frequency shift using the relationship between SSN and $\delta\nu$. The forecast is obtained using Wavelet+LGBM technique (top), LOESS_FFT+LGBM technique (middle) and N-BEATS (bottom).

Table 1 The expected year of minimum of p -mode frequency shift in Sunspot Cycle 26, based on our analysis, is between 2030 and 2031 from direct $\delta\nu$ forecast and 2030 from scaled-forecast of F10 (see top panel in Figures 5).

Sunspot cycle number	$\delta\nu$ minimum period	$\delta\nu$ maximum period
23	March 1996	January 2002
24	September 2009	October 2014
25	February 2020	August 2024
26	March 2030 (scaled forecast) January 2031 ($\delta\nu$ directly)	

in magenta color in Figure 3 and for Cycle 26 using wavelet method on the relationship between F10 and $\delta\nu$ (see Figure 5). Overall, we find from our analyses that the p -mode frequencies have already reached their maximum around the beginning of 2025 for Solar Cycle 25 and are now in the descending phase. Also, given our prediction of the SC 26 minimum near 2031, the descending phase of SC 25 is expected to last approximately seven years, which is similar to the seven-year decay phase observed for SC 23. Clearly, SC25 is stronger than SC24 and is more comparable to SC23 although the small-scale features are different in the two.

Our analyses also warrants improvements in machine-learning based methods if we are to use them for data-driven, time series solar activity forecasts. A careful examination of the scaled-forecasts in our study shows that Deep learning-NBEATS method has been unable to reproduce the QBOs like the other methods, despite all showing similar qualitative behaviour. We leave such improvements for future.

The cyclic pattern of p -mode frequency shifts is similar to that of radio flux, sunspot numbers, and many other proxies (Chaplin et al. 2007). This suggests that solar activity proxies are influenced by different layers of the Sun from the sub-photospheric layers up to the corona. The scaled-forecasts of p -mode frequency shifts based on their relationship with activity proxies such as radio flux and sunspot numbers, exploited in this study, are thus expected to be sensitive to the type and extent of data. Different type of dataset have different sources and magnitudes of errors as they are measured by different instruments and sample different regions of the Sun. Also, very small-scale fluctuations get smoothed out with data points averaged over 9 days. We have therefore, not shown any error bars for the forecasts and no quantitative conclusion on magnitudes are emphasised here. We conclude that future improvements in forecasting model and with more data, there is potential for p -mode frequency shifts to act as an additional and independent indicator of solar activity and which will provide a valuable link from inner regions of the Sun to the outer atmosphere for space weather forecasting.

Although the first attempts to characterise p -mode frequency shifts during SC25, made in this article using either the past measured frequency shifts or their relationship with SSN and F10, are promising, their direct interpretability is limited at this stage. Our knowledge of the physical processes inside the Sun comes from many other helioseismic parameters such as amplitudes, line-widths etc. Also, there are many other solar activity indices that show similar variation with solar cycle as p -mode properties. Including these parameters and indices along with longer dataset of frequencies will be essential in establishing more accurate and more detailed machine-learning-based forecast methods.

We conclude that the $\delta\nu$ forecast for the remaining period of SC25 computed here, directly from the measured dataset of $\delta\nu$ and complemented by the scaled-frequency shifts (exploiting relationship of $\delta\nu$ with SSN and F10), using machine-learning methods provided an alternative way to study future evolution of oscillation frequencies with solar activity. However, helioseismic monitoring of the Sun needs to be continued for the foreseeable future so

that robust forecasting methods can be developed further to understand solar activity and its evolution through p -mode diagnostics. The results presented here represent a nascent step towards using oscillation frequencies as an additional and independent indicator of solar activity.

Acknowledgments RJ (PI) and AK acknowledge partial funding by the STFC (UK) Impact Acceleration Award via The University of Sheffield's Internal Knowledge Exchange Scheme. The code was developed by AK using the Google Colab cloud computing platform. For analyses and visualization, open-source Python packages NumPy, Pandas, PyWT, Matplotlib, Light GBM, and N-BEATS were used. We also acknowledge the use of Gemini (Google) and Chat GPT (OpenAI) for occasional code debugging and suggesting edits for improvement. This work also utilizes GONG data obtained by the NSO Integrated Synoptic Program, managed by the National Solar Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA), Inc. under a cooperative agreement with the National Science Foundation (USA) and with contribution from the National Oceanic and Atmospheric Administration (NOAA). The GONG network of instruments is hosted 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.

Author Contributions RJ wrote the main manuscript text with contributions from the co-authors. The code was developed by AK using the Google Colab cloud computing platform. The frequencies corresponding to 9-days were generated for this study by the co-author SCT using the GONG project pipeline. All authors reviewed the manuscript.

Data Availability The Sunspot dataset used in this study is publicly available at <https://www.sidc.be/silso/datafiles>. The 10.7 cm Radio flux data can be obtained from <https://www.spaceweather.gc.ca>. The frequencies corresponding to 9-days were generated for this study by the co-author SCT using the GONG project pipeline.

Declarations

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baker, D.N.: 2000, Effects of the Sun on the Earth's environment. *J. Atmos. Sol.-Terr. Phys.* **62**, 1669. DOI. ADS.
- Basu, S.: 2021, Evidence of solar-cycle-related structural changes in the solar convection zone. *Astrophys. J.* **917**, 45. DOI. ADS.
- Bazilevskaya, G., Broomhall, A.-M., Elsworth, Y., Nakariakov, V.M.: 2014, A combined analysis of the observational aspects of the quasi-biennial oscillation in solar magnetic activity. *Space Sci. Rev.* **186**, 359. DOI. ADS.
- Benevolenskaya, E.E.: 1998, A model of the double magnetic cycle of the Sun. *Astrophys. J. Lett.* **509**, L49. DOI. ADS.
- Chaplin, W.J., Elsworth, Y., Miller, B.A., Verner, G.A., New, R.: 2007, Solar p-mode frequencies over three solar cycles. *Astrophys. J.* **659**, 1749. DOI. ADS.
- Charbonneau, P.: 2020, Dynamo models of the solar cycle. *Living Rev. Sol. Phys.* **17**, 4. DOI. ADS.
- Chen, J., Liu, Y., Wang, Z., Wang, C.: 2019, The promotion effect of LOESS smoothing technique in short-term traffic volume clustering. In: *Proceedings of the 2019 International Conference on Modeling, Analysis, Simulation Technologies and Applications (MASTA 2019)*, Atlantis Press, 386. DOI.

- Christensen-Dalsgaard, J., Berthomieu, G.: 1991, Theory of solar oscillations. In: Cox, A.N., Livingston, W.C., Matthews, M.S. (eds.) *Solar Interior and Atmosphere* **401**. [ADS](#).
- Clette, F., Lefèvre, L.: 2015, SILSO Sunspot Number V2.0. [DOI](#). Published by WDC SILSO – Royal Observatory of Belgium (ROB).
- Covington, A.E.: 1947, Micro-wave solar noise observations during the partial eclipse of November 23, 1946. *Nature* **159**, 405. [DOI](#). [ADS](#).
- Covington, A.E.: 1969, Solar radio emission at 10.7 cm, 1947–1968. *J. R. Astron. Soc. Can.* **63**, 125. [ADS](#).
- Elsworth, Y., Howe, R., Isaak, G.R., McLeod, C.P., New, R.: 1990, Variation of low-order acoustic solar oscillations over the solar cycle. *Nature* **345**, 322. [DOI](#). [ADS](#).
- Fletcher, S.T., Broomhall, A.-M., Salabert, D., Basu, S., Chaplin, W.J., Elsworth, Y., Garcia, R.A., New, R.: 2010, A seismic signature of a second dynamo? *Astrophys. J. Lett.* **718**, L19. [DOI](#). [ADS](#).
- Goldreich, P., Murray, N., Willette, G., Kumar, P.: 1991, Implications of solar p-mode frequency shifts. *Astrophys. J.* **370**, 752. [DOI](#). [ADS](#).
- González Hernández, I., Komm, R., Pevtsov, A., Leibacher, J.: 2014, Solar origins of space weather and space climate: preface. *Sol. Phys.* **289**, 437. [DOI](#). [ADS](#).
- Hajra, R., Marques de Souza Franco, A., Echer, E., José Alves Bolzan, M.: 2021, Long term variations of the geomagnetic activity: a comparison between the strong and weak solar activity cycles and implications for the space climate. *J. Geophys. Res. Space Phys.* **126**, e28695. [DOI](#). [ADS](#).
- Harvey, J.W., Hill, F., Hubbard, R.P., Kennedy, J.R., Leibacher, J.W., Pintar, J.A., Gilman, P.A., Noyes, R.W., Title, A.M., Toomre, J., Ulrich, R.K., Bhatnagar, A., Kennewell, J.A., Marquette, W., Patron, J., Saa, O., Yasukawa, E.: 1996, The global oscillation network group (GONG) project. *Science* **272**, 1284. [DOI](#). [ADS](#).
- Howard, R., Labonte, B.J.: 1980, The sun is observed to be a torsional oscillator with a period of 11 years. *Astrophys. J. Lett.* **239**, L33. [DOI](#). [ADS](#).
- Howe, R., Hill, F., Komm, R., Chaplin, W.J., Elsworth, Y., Davies, G.R., Schou, J., Thompson, M.J.: 2018, Signatures of Solar Cycle 25 in subsurface zonal flows. *Astrophys. J. Lett.* **862**, L5. [DOI](#). [ADS](#).
- Jain, K., Chowdhury, P., Tripathy, S.C.: 2023, Helioseismic investigation of quasi-biennial oscillation source regions. *Astrophys. J.* **959**, 16. [DOI](#). [ADS](#).
- Jain, R., Roberts, B.: 1993, Do p-mode frequency shifts suggest a hotter chromosphere at solar maximum? *Astrophys. J.* **414**, 898. [DOI](#). [ADS](#).
- Jain, R., Roberts, B.: 1996, Discrete high-frequency p-modes. *Astrophys. J.* **456**, 399. [DOI](#). [ADS](#).
- Jain, K., Tripathy, S.C., Hill, F.: 2009, Solar activity phases and intermediate-degree mode frequencies. *Astrophys. J.* **695**, 1567. [DOI](#). [ADS](#).
- Jain, K., Jain, N., Tripathy, S.C., Dikpati, M.: 2022, What seismic minimum reveals about solar magnetism below the surface. *Astrophys. J. Lett.* **924**, L20. [DOI](#). [ADS](#).
- Kehinde, O.J., Ayinde, K., Lukman, A.F.: 2020, A hybrid model of LOESS and ARIMA for improved time series forecasting. *J. Appl. Stat.* **47**, 2212.
- Kiefer, R., Komm, R., Hill, F., Broomhall, A.-M., Roth, M.: 2018, GONG p-mode parameters through two solar cycles. *Sol. Phys.* **293**, 151. [DOI](#). [ADS](#).
- Kosovichev, A.G., Pipin, V.V.: 2019, Dynamo wave patterns inside of the Sun revealed by torsional oscillations. *Astrophys. J. Lett.* **871**, L20. [DOI](#). [ADS](#).
- Larson, T.P., Schou, J.: 2018, Global-mode analysis of full-disk data from the Michelson Doppler Imager and the Helioseismic and Magnetic Imager. *Sol. Phys.* **293**, 29. [DOI](#). [ADS](#).
- Leighton, R.B., Noyes, R.W., Simon, G.W.: 1962, Velocity fields in the solar atmosphere. I. Preliminary report. *Astrophys. J.* **135**, 474. [DOI](#). [ADS](#).
- Libbrecht, K.G., Woodard, M.F.: 1990, Solar-cycle effects on solar oscillation frequencies. *Nature* **345**, 779. [DOI](#). [ADS](#).
- Mandal, K., Kosovichev, A.G., Pipin, V.V.: 2024, Helioseismic properties of dynamo waves in the variation of solar differential rotation. *Astrophys. J.* **973**, 36. [DOI](#). [ADS](#).
- Mehta, T., Jain, K., Tripathy, S.C., Kiefer, R., Kolotkov, D., Broomhall, A.-M.: 2022, Cycle dependence of a quasi-biennial variability in the solar interior. *Mon. Not. R. Astron. Soc.* **515**, 2415. [DOI](#). [ADS](#).
- Mursula, K., Usoskin, I.G., Maris, G.: 2007, Introduction to space climate. *Adv. Space Res.* **40**, 885. [DOI](#). [ADS](#).
- Oreshkin, B.N., Carpio, D., Chapados, N., Bengio, Y.: 2019, N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv e-prints. [arXiv](#). [DOI](#). [ADS](#).
- Rabello Soares, M.C.: 2019, On the saturation of acoustic mode frequencies at high solar activity. *Mon. Not. R. Astron. Soc.* **486**, 1847. [DOI](#). [ADS](#).
- Rabello-Soares, M.C., Korzenik, S.G., Schou, J.: 2006, High-degree mode frequencies: changes with solar cycle. In: Fletcher, K., Thompson, M. (eds.) *Proceedings of SOHO 18/GONG 2006/HELAS I, Beyond the Spherical Sun, ESA Special Publication* **624**, 71. [ADS](#).

- Rhodes, E.J. Jr., Reiter, J., Schou, J.: 2002, Solar cycle variability of high-frequency and high-degree p-mode oscillation frequencies. In: Wilson, A. (ed.) *From Solar Min to Max: Half a Solar Cycle with SOHO*, *ESA Special Publication* **508**, 37. [ADS](#).
- Rhodes, E.J., Cacciani, A., Korzenik, S.G.: 1991, Observations of intermediate- and high-degree p-mode oscillations during sunspot cycles 21 and 22. *Adv. Space Res.* **11**, 17. [DOI](#). [ADS](#).
- Salabert, D., Fossat, E., Gelly, B., Kholikov, S., Grec, G., Lazrek, M., Schmider, F.X.: 2004, Solar p modes in 10 years of the IRIS network. *Astron. Astrophys.* **413**, 1135. [DOI](#). [ADS](#).
- Schwabe, H.: 1843, Sonnen-Beobachtungen im Jahre 1843. *Astron. Nachr.* **20**, 495.
- Shao, Z., Chao, F., Yang, S.-L., Zhou, K.-L.: 2017, A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting. *Renew. Sustain. Energy Rev.* **75**, 123. [DOI](#).
- Su, X., Liang, B., Feng, S., Dai, W., Yang, Y.: 2023, Solar Cycle 25 prediction using N-BEATS. *Astrophys. J.* **947**, 50. [DOI](#). [ADS](#).
- Tapping, K.F.: 2013, The 10.7 cm solar radio flux ($F_{10.7}$). *Space Weather* **11**, 394. [DOI](#). [ADS](#).
- Tripathy, S.C., Hill, F., Jain, K., Leibacher, J.W.: 2007, Solar oscillation frequency changes on time scales of nine days. *Sol. Phys.* **243**, 105. [DOI](#). [ADS](#).
- Ulrich, R.K.: 1970, The five-minute oscillations on the solar surface. *Astrophys. J.* **162**, 993. [DOI](#). [ADS](#).
- Wasilewski, F.: 2012, PyWavelets Documentation. https://pywavelets.readthedocs.io/_/downloads/en/0.2.0/pdf/#:~:text=decomposition%20using%20wavedec-,pywt. Accessed: 2026-02-05.
- Woodard, M.F., Noyes, R.W.: 1985, Change of solar oscillation eigenfrequencies with the solar cycle. *Nature* **318**, 449. [DOI](#). [ADS](#).
- Woodard, M.F., Kuhn, J.R., Murray, N., Libbrecht, K.G.: 1991, Short-term changes in solar oscillation frequencies and solar activity. *Astrophys. J. Lett.* **373**, L81. [DOI](#). [ADS](#).
- Xu, Q., Jain, R., Xing, W.: 2024, Data-driven forecasting of sunspot cycles: pros and cons of a hybrid approach. *Sol. Phys.* **299**, 25. [DOI](#). [ADS](#).
- Yuan, F., Fan, Y., Bing, X., Tian, C., Yuan, Z., Li, Y.: 2025, A decomposition-driven hybrid framework based on STL for accurate traffic flow forecasting. *Renew. Sustain. Energy Rev.* [DOI](#).
- Zhang, X., Li, R.: 2021, A novel decomposition and combination technique for forecasting monthly electricity consumption. *Front. Energy Res.* **9**, 792358. [DOI](#).
- Zhao, J., Yu, Z., Yang, X., Gao, Z., Liu, W.: 2022, Short term traffic flow prediction of expressway service area based on STL-OMS. *Phys. A, Stat. Mech. Appl.* **595**, 126937.