Available online at www.sciencedirect.com

ScienceDirect

Advances in Space Research 65 (2020) 677–683

ADVANCES IN SPACE RESEARCH (a COSPAR publication)

www.elsevier.com/locate/asr

Sunspot cycle prediction using Warped Gaussian process regression

Ítalo G. Gonçalves [a,*,1], Ezequiel Echer [b], Everton Frigo [a,c]

[a] Geophysical Signals Analysis Laboratory – Universidade Federal do Pampa, Brazil
[b] Instituto Nacional de Pesquisas Espaciais, Brazil
[c] Instituto de Geociências, Universidade Federal do Rio Grande do Sul, Brazil

Received 31 August 2019; received in revised form 15 October 2019; accepted 8 November 2019
Available online 18 November 2019

Abstract

Solar cycle prediction is a key activity in space weather research. Several techniques have been employed in recent decades in order to try to forecast the next sunspot-cycle maxima and time. In this work, the Gaussian process, a machine-learning technique, is used to make a prediction for the solar cycle 25 based on the annual sunspot number 2.0 data from 1700 to 2018. A variation known as Warped Gaussian process is employed in order to deal with the non-negativity constraint and asymmetrical data distribution. Tests using holdout data yielded a root mean square error of 10.0 within 5 years and 25.0–35.0 within 10 years. Simulations using the predictive distribution were performed to account for the uncertainty in the prediction. Cycle 25 is expected to last from 2019 to 2029, with a peak sunspot number about 117 (110 by the median) occurring most likely in 2024. Thus our method predicts that solar Cycle 25 will be weaker than previous ones, implying a continuing trend of declining solar activity as observed in the past two cycles.

Keywords: Sunspot number; Solar cycle; Machine learning; Gaussian process

1. Introduction

The Sun presents a variability in its magnetic activity on several timescales, the most noticeable being the 11-year solar cycle observed in sunspot numbers and several other solar parameters (Eddy, 1976; Hoyt and Schatten, 1997; Hathaway, 2010). The Earth's magnetosphere and ionosphere/atmosphere are strongly affected by variations in solar activity, *e. g.* solar irradiance, energetic particles, and solar wind (Hargreaves, 1992; Baker, 2000; Siscoe, 2000; Echer et al., 2004, 2005, 2011). The study of the solar-terrestrial coupling and its variation, and possible prediction, has been called space-weather science in recent decades (Baker, 2000; Siscoe, 2000; Echer et al., 2005). Because of this large scientific and practical importance, many works have applied different techniques, *e. g.* climatology, spectral analysis, non-linear techniques, precursor methods and dynamo models, in the attempt to predict solar-cycle behavior (Labonville et al., 2019; Singh and Bhargawa, 2019; Pesnell, 2008; Pesnell, 2016 and references therein).

The most basic parameter measuring the Sun's variability over long term scales and widely used in forecasting is the sunspot number S_N. This parameter was first compiled in the 19th century by R. Wolf (Eddy, 1976; Hoyt and Schatten, 1997) and includes information on individual sunspot numbers plus ten times the number of sunspot groups weighted by a coefficient related to the observer (Hoyt and Schatten, 1997). Group sunspot numbers are available since the beginning of telescopic observations with Galileo's first sporadic solar observations in 1610 (Eddy, 1976; Hoyt and Schatten, 1997). Annual averages of S_N are available since 1700. However, several investigations have shown that the earlier sunspot number had a number of accuracy problems (Hoyt and Schatten, 1998;

* Corresponding author at: Av. Pedro Anunciação, 111, Caçapava do Sul, RS, Brazil.
E-mail address: italogoncalves@unipampa.edu.br (Í.G. Gonçalves).
[1] Source code is available on https://github.com/italo-goncalves/-geoML.git.

https://doi.org/10.1016/j.asr.2019.11.011
0273-1177/

Hathaway, 2010; Clette et al., 2016). Therefore a new sunspot-number time series was compiled to provide a more accurate solar-activity index, the sunspot number 2.0 version (Clette et al., 2016,).

Machine learning models are starting to be adopted for sunspot number prediction (Pala and Atici, 2019; Covas et al., 2019), especially those based on neural networks. A technique that has yet to be applied to this problem is the Gaussian process (GP), which is starting to see applications in geophysics (Sarkar et al., 2019). Compared to other machine learning models, its main advantages are its applicability to small datasets and the ability to make predictions with a confidence interval. In this article we aim to apply this technique to the sunspot-number data in order to predict the next Solar Cycle (25) amplitude and its timing. The data consists of annual averages of the sunspot-number 2.0 version (Clette and Lefèvre, 2016) from 1700 to 2018, downloaded from the Solar Influences Data Analysis Center (sidc.oma.be). The method is applied first to subsets to assess its accuracy. Results are compared with previous works.

2. Gaussian process

Rasmussen and Williams (2006) define the Gaussian process (GP) as "a collection of random variables, any finite number of which have a joint Gaussian distribution". It can be seen as a distribution over functions, specified by a mean function $m(\boldsymbol{x})$ and a kernel function[2] $k(\boldsymbol{x}, \boldsymbol{x}\prime)$, where $\boldsymbol{x}$ and $\boldsymbol{x}\prime$ are points in the space of independent variables. A function $f(\boldsymbol{x})$ that is distributed as a GP is written as:

$$f(\boldsymbol{x}) \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')) \tag{1}$$

The function $f(\boldsymbol{x})$ is the quantity of modeling interest. However, one can only measure a noisy version of f denoted as y, so that $y(\boldsymbol{x}) = f(\boldsymbol{x}) + \epsilon$ and $\epsilon \sim N(0, \sigma_0^2)$ is the measurement error (N denotes the Gaussian or normal distribution). The error variance σ_0^2 includes all possible sources of error, such as sampling and measurement errors, as well as fluctuations with frequency higher than the sampling rate. It can also be interpreted as the amount of variance in the data that a model cannot explain.

Given a number M of coordinates containing measurements of y (called *training* points) and a number T of coordinates where one wishes to make predictions of y (the *test* points), let the vector of observations be denoted as $\boldsymbol{y}_M$ and the vector with the values to be modeled as $\boldsymbol{y}_T$. The joint distribution of $\boldsymbol{y}_M$ and $\boldsymbol{y}_T$ under the GP is given by

$$\begin{bmatrix} \boldsymbol{y}_M \\ \boldsymbol{y}_T \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{m}_M \\ \boldsymbol{m}_T \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{MM} + \sigma_0^2 \mathbf{I}_M & \mathbf{K}_{MT} \\ \mathbf{K}_{TM} & \mathbf{K}_{TT} + \sigma_0^2 \mathbf{I}_T \end{bmatrix} \right) \tag{2}$$

where $\mathbf{K}_{MM}, \mathbf{K}_{MT}, \mathbf{K}_{TM}$, and $\mathbf{K}_{TT}$ are covariance matrices with entries $\mathbf{K}_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\mathbf{I}_M$ and $\mathbf{I}_T$ are identity matrices, and $\boldsymbol{m}_M$ and $\boldsymbol{m}_T$ are evaluations of the mean function on the M and T locations. By conditioning the unobserved values on the known data one obtains:

$$\boldsymbol{y}_T | \boldsymbol{y}_M \sim N(\boldsymbol{\mu}_T, \Sigma_T), \text{with} \tag{3}$$

$$\boldsymbol{\mu}_T = \mathbf{K}_{TM}\left[\mathbf{K}_{MM} + \sigma_0^2 \mathbf{I}_M\right]^{-1}(\boldsymbol{y}_M - \boldsymbol{m}_M) + \boldsymbol{m}_T \tag{4}$$

$$\Sigma_T = \mathbf{K}_{TT} - \mathbf{K}_{TM}\left[\mathbf{K}_{MM} + \sigma_0^2 \mathbf{I}_M\right]^{-1}\mathbf{K}_{MT} + \sigma_0^2 \mathbf{I}_T \tag{5}$$

The conditional mean and covariance matrix are then given by $\boldsymbol{\mu}_T$ and Σ_T, respectively. The mean function $m(\boldsymbol{x})$ is usually assumed to be constant and subtracted from the data.

As long as M and T are small ($< 10^4$ data points), the conditional distribution can be sampled by computing the Cholesky decomposition $\Sigma_T = \mathbf{L}\mathbf{L}^T$ and making $\boldsymbol{y}_s = \mathbf{L}\boldsymbol{n} + \boldsymbol{\mu}_T$, where $\boldsymbol{n} \sim N(\boldsymbol{0}, \mathbf{I}_T)$ and $\boldsymbol{y}_s$ is a sample from the conditional distribution.

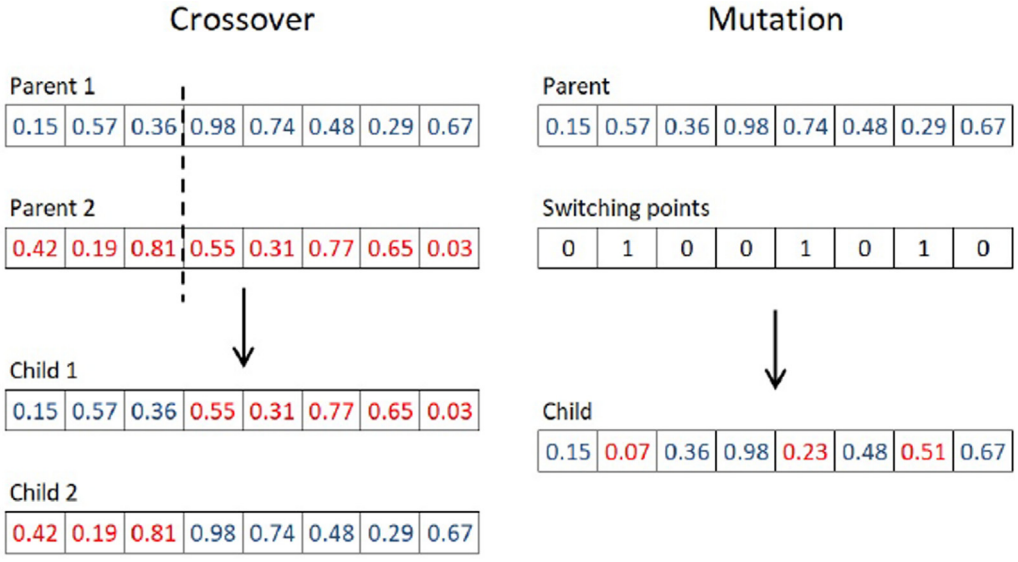

Fig. 1. An illustration of the crossover and mutation operations.

[2] Also known as covariance function.

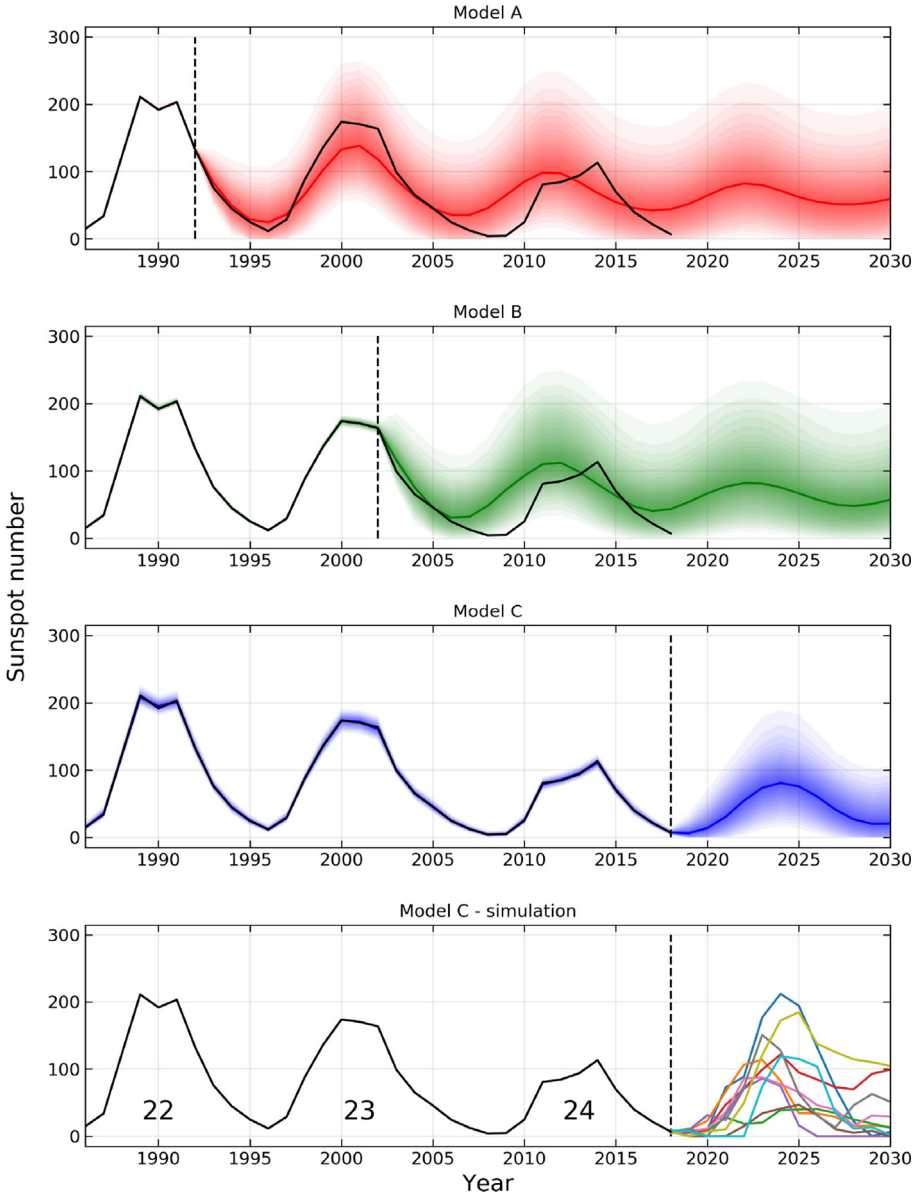

Fig. 2. Warped GP prediction of the sunspot number. *Full line*: real measurements. *Dashed line*: limit of training data. *Colored lines*: median prediction. From top to bottom: predictions using data up to 1992, 2002, 2018, and samples from the predictive distribution. The colored ribbons represent the predicted probability distribution, with opacity proportional to the probability density. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.1. Warped Gaussian process

The standard GP outputs a normal distribution as a prediction for each testing point, whose domain comprises the whole real line. For certain variables, such as the sunspot number, negative predictions are nonphysical. In order to avoid this, the predictive distribution is mapped through a nonlinear, monotonic function, the so-called "warping" (Snelson et al., 2004). The warping function must be monotonic in order to allow an inverse mapping. Snelson et al. (2004) use a series of tanh functions to parameterize the warping, while Rios and Tobar (2018) use a Box-Cox transformation to obtain a simple and computationally fast inverse mapping. The approach employed here uses a set of knots (y_p, z_p), $p \in 1, 2, \ldots, P$, where y and z are the warped and original variables, respectively. The warping function $y = f(z)$ is then obtained by interpolating these knots with a monotonic spline (Fritsch and Carlson, 1980), with linear extrapolation. The inverse warping is easily obtained by interpolating the swapped coordinates (z_p, y_p).

Non-negativity can be enforced by applying another function with non-negative domain before the spline, such as $g(t) = \log(t)$ or the inverse softplus function $g(t) = \log(\exp(t) - 1)$ (the one used in this work). This way the warping becomes $y = f(g(z))$. Besides avoiding negative values, another advantage of the warped GP is the possibility of passing the predictive distribution through the inverse warping function to obtain an

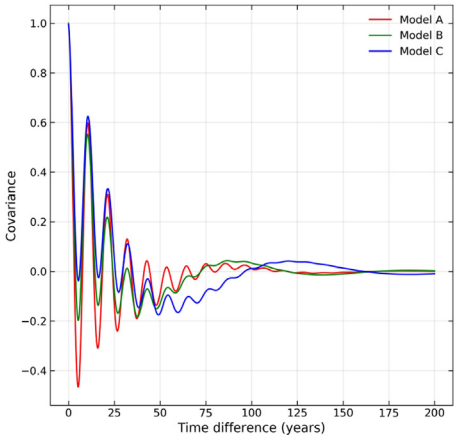

Fig. 3. Covariance function learned by the model.

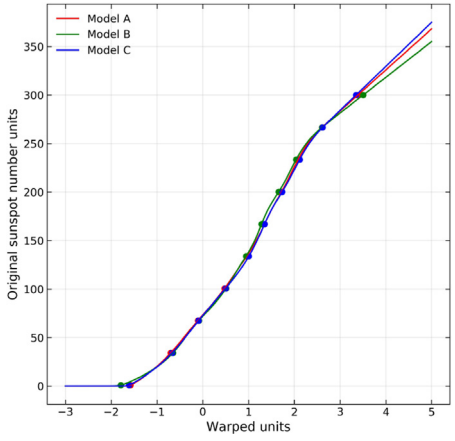

Fig. 4. Warping function learned by the model. The dots represent the learned knots (y_p, z_p).

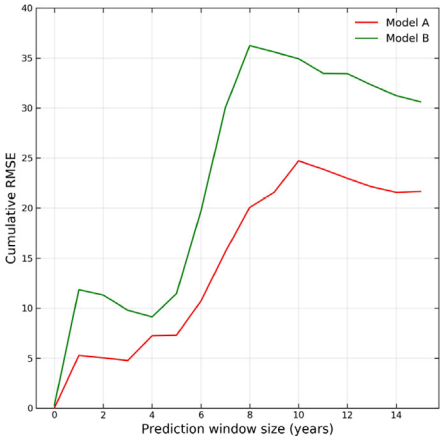

Fig. 5. Root mean square error as a function of the prediction window size.

asymmetric prediction. In terms of computational speed and analytic tractability, the spline warping possess the same advantages as the one employed by Rios and Tobar (2018).

2.2. Model training

The GP can be trained by maximizing the log-likelihood (Snelson et al., 2004):

$$L = -\frac{1}{2}\log\det\mathbf{K}_{MM} - \frac{1}{2}\boldsymbol{y}_M^T\mathbf{K}_{MM}^{-1}\boldsymbol{y}_M + \sum_{i=1}^{M}\log\left[\frac{\partial y}{\partial z}\right]_i - \frac{M}{2}\log 2\pi \tag{6}$$

The covariance matrix $\mathbf{K}_{MM}$ is a function of the kernel parameters. This work employs a combination of the exponential kernel with a periodic one:

$$k(x,x\prime) = \sum_{q=1}^{Q}\sigma_q^2\exp\left(-\frac{|x-x\prime|}{l_q}\right)\cos(2\pi(x-x\prime)f_q) \tag{7}$$

where f_q is the frequency of the kernel component q, l_q its length-scale, σ_q^2 its weight, and Q is the number of components. The weights σ_q^2 model the amplitude of each component. As the product of two kernels is a valid kernel (Rasmussen and Williams, 2006), this kernel function is able to model the somewhat chaotic evolution of the sunspot number as well as its periodic characteristic. This can also be seen as a variation of the spectral mixture kernel by Wilson and Adams (2013).

The z_p-coordinates of the warping knots are evenly spread over the range of the data, leaving the y_p as parameters to be optimized. Adding three parameters per kernel component and the noise variance σ_0^2, the total number of parameters is $P + 3Q + 1$. The warped GP is constrained to zero mean and unit variance, or $\sigma_0^2 + \sum_{q=1}^{Q}\sigma_q^2 = 1$.

2.3. Optimization

The model is trained using a genetic algorithm (Michalewicz, 1996), a stochastic optimizer that selects candidate solutions for a problem by mimicking nature's natural selection phenomenon. The problem's parameters are called *genes*, and a vector formed from all parameters is a *chromossome*.

Starting from a pool of randomly generated candidate solutions (the *population*), two individuals are selected with probability proportional to their values of the *fitness function*, the function being optimized (in this case, Eq. (6)). They undergo a *crossover* operation, where their chromossomes are crossed at a randomly chosen point, forming two children that combine the characteristics from both parents. Each child may also undergo a *mutation* operation (with a probability chosen *a priori*), where some of its genes are switched by random values. Potentially the children will present better values for the fitness function and will replace the worst performing member of the population. This procedure is repeated for any number of iterations. Fig. 1 illustrates the process.

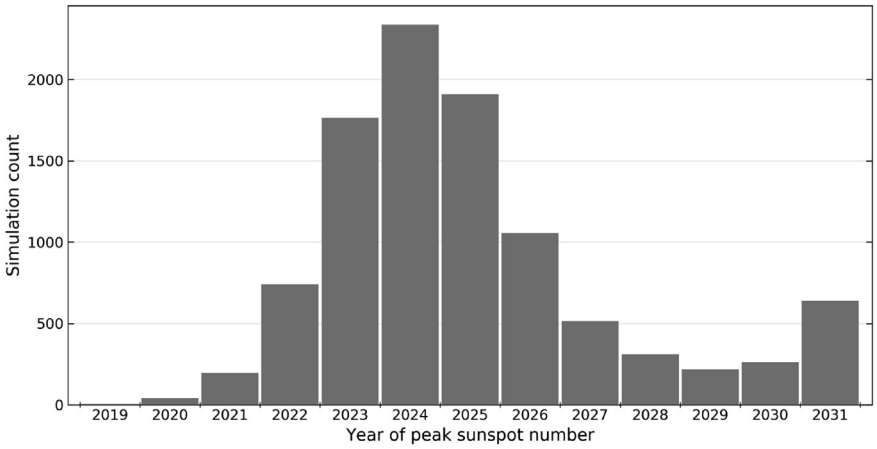

Fig. 6. Distribution of the year of peak S_N.

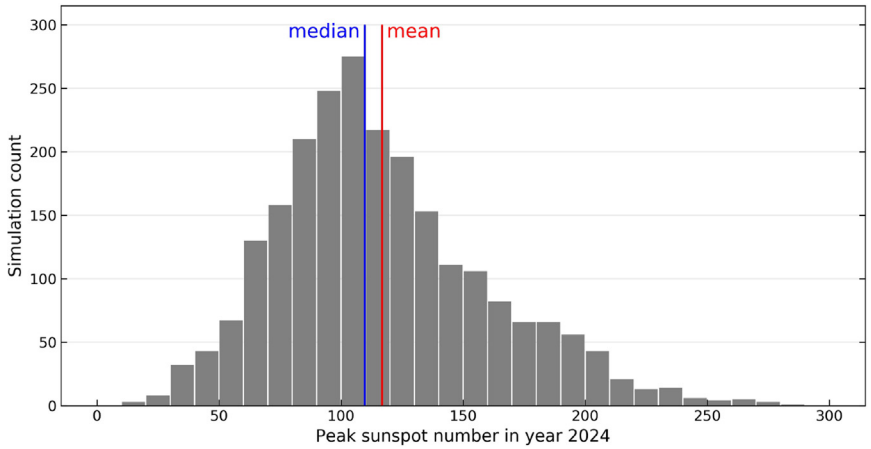

Fig. 7. Distribution of peak S_N for 2024.

All of the equations and the optimizer are implemented in a Python package (under development). The code can be found at github.com/italo-goncalves/geoML.

3. Results and discussion

In order to assess the model's predictive power, the warped GP was trained on three subsets of the data, resulting in three models: one up to 1992 (model A), one up to 2002 (model B), and one using all of the data up to 2018 (model C). In all cases we set $P = 10$ and $Q = 3$ (for higher Q, it was verified that the model tends to discard the extra components by assigning them a very low variance). After the training, the models were used to make predictions for the future. The results are displayed in Fig. 2. Fig. 3 shows the covariance function learned by the model, and Fig. 4 presents the learned warping function.

All of the models learned a similar pattern from the data, detecting major periods of 11 and 100 years, approximately (Fig. 3), with other patterns presenting negligible amplitude. For all models the amount of explained variance $(1 - \sigma_0^2)$ was greater than 99.5%. In Fig. 2 it is possible to see the asymmetric predictive density, with a positive skew. Fig. 2 (bottom) shows some samples from the predictive distribution, which follow a cyclic pattern but not one as clear as the data. Note that the model contemplates the possibility of a cycle with a very high peak or a very short duration, although such outcomes are very unlikely. In average, the samples tend to form clearer cycles as in the upper rows.

The figure also displays the median of the predicted distribution, which corresponds to the mean in warped space. Using it as a predictor yielded the root mean square error (RMSE) values displayed in Fig. 5, for prediction window sizes up to 15 years. Note that the model was able to predict the amplitude of Cycle 24, even though it is smaller than the previous cycles'. However, the long tail in Cycle 23 eluded models A and B, causing the prediction of Cycle 24 to appear out of phase and resulting in larger errors for model B. The models show good agreement with the data up to a distance of 5–6 years. Perhaps better results could be obtained by using a nonstationary GP, where the parameters σ_q^2, l_q, and f_q are also functions of time (Heinonen et al., 2016; Remes et al., 2017), but it presents its own challenges in terms of implementation, training, and especially forecasting.

In order to assess the uncertainty of the amplitude and phase of the next cycle, 10,000 samples were drawn from the predictive distribution. For each sample the peak sunspot number and the year of its occurrence were determined. Fig. 6 shows the distribution of the year of occurrence of the peak S_N. 2024 is the year with most occurrences of the peak. Years 2019 and 2029 had the smallest number of peak occurrences, which makes them good candidates for the next cycle's limits.

Fig. 7 shows the distribution of the peak S_N for 2024. The peak is expected to lie in the [45.0, 212.1] interval with 95% confidence, with values of 116.8, 109.8, and 105.0 for the mean, median, and mode, respectively.

4. Summary and conclusions

In this work a Warped Gaussian process regression has been applied to the sunspot number version 2.0 to predict the next solar-cycle duration, peak, and epoch of maximum. Evaluation on holdout data yielded a RMSE of 10.0 within 5 years and 25.0–35.0 within 10 years, proving the suitability of the model for this task. From our results it was found that the Cycle 25 peak should occur most likely in 2024 with a value about 110–117 (median and average). Considering the median of 110 as our predicted Sunspot Cycle 25 maximum, it can be concluded that Cycle 25 is predicted to be slightly smaller than Cycle 24 (maximum annual sunspot number 2.0 of 113) and much lower than Cycle 23 (maximum annual sunspot number 2.0 of 174).

Although many other solar-cycle prediction works have used the old sunspot number, it has been found elsewhere that Cycle 25 could be lower than most of modern-era solar activity. For instance, Rigozo et al. (2011) using a spectral analysis technique, have found that Cycle 25 would be 5% lower than Cycle 23, but slightly higher than cycle 24. They predicted a peak S_N of 132 around 2023. On the other hand, other authors have predicted that Solar Cycle 25 could be comparable to or even higher than Cycle 24 (Kane, 2007; Pesnell and Schatten, 2018). In this article our prediction is that Cycle 25 will be weaker than Cycles 23 and 24. If this prediction is confirmed, there will be important effects on the solar-terrestrial environment, since a decreased input of solar energy (irradiance, energetic particles, transient activity, and solar wind) would have deep effects on the magnetosphere/ionosphere and on the technological systems operating on it.

Acknowledgements

I.G. Gonçalves gratefully acknowledges the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. E. Echer acknowledges Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; process 302583/2015-7) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; process 2018/21657-1) for financial support. E. Frigo acknowledges CNPq (processes 429068/2016-6 and 160045/2018-5) for financial support.

References

Baker, D., 2000. Effects of the Sun on the Earth s environment. J. Atmos. Solar Terr. Phys. 62, 1669–1681.
Clette, F., Cliver, E., Lefèvre, L., Svalgaard, L., Vaquero, J., Leibacher, J., 2016. Preface to topical issue: recalibration of the sunspot number. Sol. Phys. 219, 2479–2486.
Clette, F., Lefèvre, L., 2016. The new sunspot number: assembling all corrections. Sol. Phys. 291, 2629–2651.
Covas, E., Peixinho, N., Fernandes, J., 2019. Neural network forecast of the sunspot butterfly diagram. Sol. Phys. 294, (3). https://doi.org/10.1007/s11207-019-1412-z.
Echer, E., Gonzalez, W., Guarnieri, F., Lago, A.D., Vieira, L., 2005. Introduction to space weather. Adv. Space Res. 35, 855–865.
Echer, E., Rigozo, N.R., Nordemann, D.J.R., Vieira, L.E.A., 2004. Prediction of solar activity on the basis of spectral characteristics of sunspot number. Ann. Geophys. 22, 2239–2243.
Echer, E., Tsurutani, B.T., Gonzalez, W.D., 2011. Extremely low geomagnetic activity during the recent deep solar cycle minimum. In: Mandrini, C.H., Webb, D. (Eds.), Proc. IAU Symp, 286, Comparative Magnetic Minima, Characterizing quiet times in the Sun and Stars. Cambridge University Press, pp. 200–209.
Eddy, J., 1976. The Maunder minimum. Science 19, 1189–1202.
Fritsch, F.N., Carlson, R.E., 1980. Monotone piecewise cubic interpolation. SIAM J. Num. Anal. 17, 238–246.
Hargreaves, J., 1992. The Solar-Terrestrial Environment. Cambridge University Press, Cambridge, Great Britain.
Hathaway, D.H., 2010. The Solar Cycle. Living Rev. Sol. Phys. 7.
Heinonen, M., Mannerström, H., Rousu, J., Kaski, S., Lähdesmäki, H., 2016. Non-Stationary Gaussian Process Regression with Hamiltonian Monte Carlo. In: AISTATS '16: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics. vol. 41. pp. 1–12. http://arxiv.org/abs/1508.04319.
Hoyt, D., Schatten, K., 1997. The Role of the Sun in Climate Change. Oxford University Press, Oxford.
Hoyt, D., Schatten, K., 1998. Group sunspot numbers: a new solar activity reconstruction. Sol. Phys. 179, 189–219.
Kane, R., 2007. Solar Cycle predictions based on extrapolation of spectral components: an update. Sol. Phys. 246, 487–493.
Labonville, F., Charbonneau, P., Lemerle, A., 2019. A dynamo-based forecast of solar cycle 25. Sol. Phys. 294, (6). https://doi.org/10.1007/s11207-019-1480-0.
Michalewicz, Z., 1996. Genetic Algorithms + Data Structures = Evolution Programs, third ed. Springer, Berlin.
Pala, Z., Atici, R., 2019. Forecasting sunspot time series using deep learning methods. Sol. Phys. 294 (5). https://doi.org/10.1007/s11207-019-1434-6.
Pesnell, W.D., 2008. Predictions of solar cycle 24. Sol. Phys. 252, 209–220.
Pesnell, W.D., 2016. Predictions of Solar Cycle 24: How are we doing? Space Weather 14, 10–21.
Pesnell, W.D., Schatten, K.H., 2018. An early prediction of the amplitude of Solar Cycle 25. Sol. Phys., 293
Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press, Cambridge, Massachusetts www.GaussianProcess.org/gpml.
Remes, S., Heinonen, M., Kaski, S., 2017. Non-Stationary Spectral Kernels (Nips). http://arxiv.org/abs/1705.08736.
Rigozo, N.R., Souza Echer, M.P., Evangelista, H., Nordemann, D.J.R., Echer, E., 2011. Prediction of sunspot number amplitude and solar cycle length for cycles 24 and 25. J. Atmos. Solar Terr. Phys. 73 (11–12), 1294–1299.

Rios, G., Tobar, F., 2018. Learning non-Gaussian time series using the box-cox Gaussian process. In: Proceedings of the International Joint Conference on Neural Networks 2018-July.

Sarkar, D., Osborne, M.A., Adcock, T.A., 2019. Spatiotemporal prediction of tidal currents using gaussian processes. J. Geophys. Res.: Oceans 124 (4), 2697–2715.

Singh, A.K., Bhargawa, A., 2019. Prediction of declining solar activity trends during solar cycles 25 and 26 and indication of other solar minimum. Astrophys. Space Sci. 364 (1), 1–7. https://doi.org/10.1007/s10509-019-3500-9.

Siscoe, G., 2000. The space-weather enterprise: past, present and future. J. Atmos. Solar Terr. Phys. 62, 1223–1232.

Snelson, E., Rasmussen, C.E., Ghahramani, Z., 2004. Warped Gaussian processes. Adv. Neural Inf. Proc. Syst. 16, 337–344.

Wilson, A.G., Adams, R.P., 2013. Gaussian Process Kernels for Pattern Discovery and Extrapolation. In: 30th Internat. Conf. Machine Learning. Vol. 28. pp. 1067–1075. http://arxiv.org/abs/1302.4245.