Monthly Sunspot Number Forecasting using Gaussian Process Regression with Brownian Oscillator Kernel

Konstantinos Bekos[1], Dionissios Hristopulos[1]

[1]School of Electrical & Computer Engineering, Technical University Of Crete, Chania, Greece

The monthly averages of daily measurements of the number of sunspots is one of the most characteristic indicators of solar activity. It has been observed and recorded for more than two centuries, exhibiting patterns that approximately repeat every eleven (11) years. Forecasting the number of sunspots is useful in various fields affected by solar activity, such as satellite communications and high-frequency telecommunications.

We approach the problem of forecasting the number of monthly sunspots within the framework of Supervised Machine Learning, by applying Gaussian Process Regression, using kernels with periodic components. The Gaussian Process Regression method, in addition to generating accurate point predictions, also computes corresponding prediction intervals. These intervals represent a range of values that is expected to contain the true value of the process with a predefined probability. Gaussian processes exploit existing knowledge about the behavior of the data and their mutual correlations in order to interpolate this information at intermediate and future time points, thereby achieving a better approximation of the target function. The specification of the correlations between values of the modeled function is accomplished through the kernel function, which constitutes the central "engine" of the trained model. Suitable kernels must possess properties compatible with the function being modeled.

We investigate kernels with periodic characteristics that are typically used in machine learning, as well as a kernel based on the Brownian oscillator model. This kernel comprises harmonic functions with exponentially decaying amplitude. The estimation of the kernel parameters from the data is performed using Maximum Likelihood, as well as two other methods that exploit the sample autocorrelation function of the training data. Furthermore, we introduce an innovative prediction framework that utilizes ensemble learning via a weighted-average aggregation of weak learners.

In summary, the models that we present effectively capture —via the respective kernels— the short-time fluctuations (over small temporal scales) and the periodic correlations (across consecutive Solar Cycles) exhibited by the time series. Notably, the best-performing model attains a Root Mean Square Error of 8.37 and a Relative Root Mean Square Error of 3.61 % over a validation set that contains 40 monthly values.

Acknowledgment This research was funded by the research project SLIMNETS, which is implemented in the framework of H.F.R.I call "3rd Call for H.F.R.I.'s Research Projects to Support Faculty Members & Researchers" (H.F.R.I. Project Number: 25024).

References:

1. C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. MIT Press, 2006, ISBN: 026218253X.

2. F. Clette and L. Lef'evre, SILSO Sunspot Number V2.0, https://doi.org/10.24414/qnza-ac80, Published by WDC SILSO - Royal Observatory of Belgium (ROB), July 2015. DOI: 10.24414/qnza-ac80.

3. Dionissios T. Hristopulos, "Non-Separable Covariance Kernels for Spatiotemporal Gaussian Processes Based on a Hybrid Spectral Method and the Harmonic Oscillator", IEEE Transactions on Information Theory, vol. 70, n. 2, pp. 1268–1283, 2024. DOI: 10.1109/TIT.2023.3321215.