# Does ENSO Significantly Affect Rice Production In Indonesia? A Preliminary Study Using Computational Time-Series Approach

Sandy H. S. Herho [a,1,*], Ferio Brahmana [b,2], Katarina E. P. Herho [c,3], Dasapta E. Irawan [d,4]

[a] Department of Geology, College of Computer, Mathematical, and Natural Sciences, University of Maryland, College Park, MD, 20742, USA

[b] Pacmann Academy, Algoritma Cerdas Indonesia Inc., South Jakarta, Special Capital Region of Jakarta, 12540, Indonesia

[c] Geological Engineering Study Program, Faculty of Earth Technology and Energy, Trisakti University, West Jakarta, DKI Jakarta, 10110, Indonesia

[d] Applied Geology Research Group, Faculty of Earth Sciences and Technology, Bandung Institute of Technology (ITB), West Java, 40132, Indonesia

[1] herho@umd.edu; [2] feriobrahmana5@gmail.com; [3] katarina.evelynn@gmail.com; [4] r-win@office.itb.ac.id

* corresponding author

**ARTICLE INFO**

**ABSTRACT**

ENSO is a phenomenon that is suspected to influence rice production in Indonesia. In this study, we try to find direct correlation between ENSO and rice production in this region by using various latest computational time series methods, such as Dynamic Time Warping, Wavelet Coherence, and Bayesian Structural Time Series to quantify the statistical relationship between the Multivariate ENSO Index on annual rice production in 1961- 2019. We did not find a direct correlation between these two variables, which may be due to the local influence of ENSO on different rice production areas in Indonesia. This study would also point out the importance of shifting the theme of research in Indonesia from mapping to monitoring and freely share the data. This step would bring science to progress further and faster.

## 1. Introduction

Indonesia is the world's third biggest rice producer in 2019. The rice crop, which is also known as paddy, itself is one of the crops heavily influenced by rainfall. For this reason, it is very important for Indonesia to know the effect of climate on rice production. This article hopes to find relationship between Multivariate ENSO Index (MEI), a phenomena which will influenced rainfall periodically, and rice production. With a simple model, due to the needs of its simplicity so that it can be digest easily by researchers from the various field of studies, we try to find if there is any correlation between these two variables. Aside from using simple model, consistent with our accessibility information, we also try to execute our study by using open-source software to maximize the reproducibility and replicability.

One of the most similar study by, which also inspires this paper, had found that ENSO has strong connections with Java's rice production. However, this study is using year-to-year change of rainfall and 4-month sea surface temperature anomalies (SSTA) which represent ENSO's effects and try to fit it with year-to-year rice production in Java. This study did not use MEI directly while the index is the ENSO's intensity representation. There are also another recent study had try to build and find any correlation between ENSO phenomena and rice production specifically in Asia and South East Asia (SEA) where rice is one of the primary food.

The first one is tackling the same issue in Philippines with same method as in Java's study with addition of ecosystems and fitting target variability. The study also find association between rainfall, SSTA, and rice production, harvested area, also the yield in one of the Philippines' Island, Luzon. There is also one other study conducted in China which adopts this method (using SSTA rather than MEI) for their work which do not find any strong correlation between ENSO event and rice production even when the SSTA has a strong correlation with rainfall.

Then, there is one study by that we could find which are using MEI as one of other variables in their pooled mean group estimation to find the correlation between the ENSO event and rice production in SEA. However, the MEI is used together with irrigation area, fertilizer consumption, and harvested area. We believe those other variables are arising more complexities which will affect the difficulties of the result interpretation. This study too find that there is an association between ENSO intensity and rice production in SEA.

With all of those precedents, as stated before, this study will try to differentiate by building a simpler model that investigate the correlation between MEI Index and rice production only. As for the method, to test the robustness of our result, we are using three kind of correlation analysis which consist of Dynamic Time Wrapping (DTW), Wavelet Coherence (WTC), and Bayesian Structural Time Series (BSTS). After estimating their correlation, we also provide further discussion about the results.

## 2.    Materials and Methods

### 2.1.    Data and Code

The analysis was based rice production data provided by FAO (https://www.fao.org/faostat/en/#home) and the Multivariate Enso Index (MEI) provided by NOAA (https://psl.noaa.gov/enso/mei/).

We processed the data using two programming languages: Python with Numpy, Matplotlib and Pandas package and R with biwavelet, bsts, and readr packages. Data and code are stored in the project' Github page (https://github.com/sandyherho/ensoIndoRice).

### 2.2.    Computational Analysis

The first step we take is to calculate annual rice yield index, which is a standardized comparison of total rice production P (in tonnes) with the area of production A (in hectare) based on FAO data (equation (1)).

$$I = \frac{P}{A} \tag{1}$$

We define the annual standardized rice yield index (ASRYI) $\hat{I}$ through equation (2) below,

$$\hat{I} = \frac{I_t - \bar{I}}{SD} \tag{2}$$

where $\bar{I}$ and $SD$ are respectively the mean and standard deviation of the data defined according to equation (3) as follows,

$$\bar{I} = \frac{1}{N} \sum_{i=1}^{N} I_t$$

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (I_t - \bar{I})^2} \tag{3}$$

where N is the number of data points, i.e., 59. The results of the time-series reconstruction of this index can be seen in Figure 1 below,

In order to measure the similarity between ASRYI $x$ and annual average of MEI $y$ signals, we use a dynamic time warping (DTW) algorithm. DTW is an algorithm that is commonly used to determine the optimal distance between two sequential data pieces to determine the similarity between the two time-series patterns. Identification of the similarity of the sequential data is done by comparing the optimal computation time to find the shortest distance between the two sequences.

The first step is to input both time-series $x_{1:N}$ and $y_{1:N}$ data, which have the same sequence length, so that we can form a cost matrix $D \in \mathbb{R}^{(N+1) \times (N+1)}$. Then, we initialize the cost matrix values as follows, $D_{0,0} = 0$; for every $i^{\text{th}}$ up to $N^{\text{th}}$ row, $D_{i,0} = \infty$; and for every $j^{\text{th}}$ up to $N^{\text{th}}$ column, $D_{i,0} = \infty$. Then, for each row component $i = 0$ and column $j = 1$ to the $N^{\text{th}}$ row and column, the components of the cost matrix are calculated using the following equation,

$$D_{i,j} = d(x_i, y_j) + \min \begin{cases} D_{i-1,j-1} & (match) \\ D_{i-1,j} & (insertion) \\ D_{i,j-1} & (deletion) \end{cases} \tag{4}$$
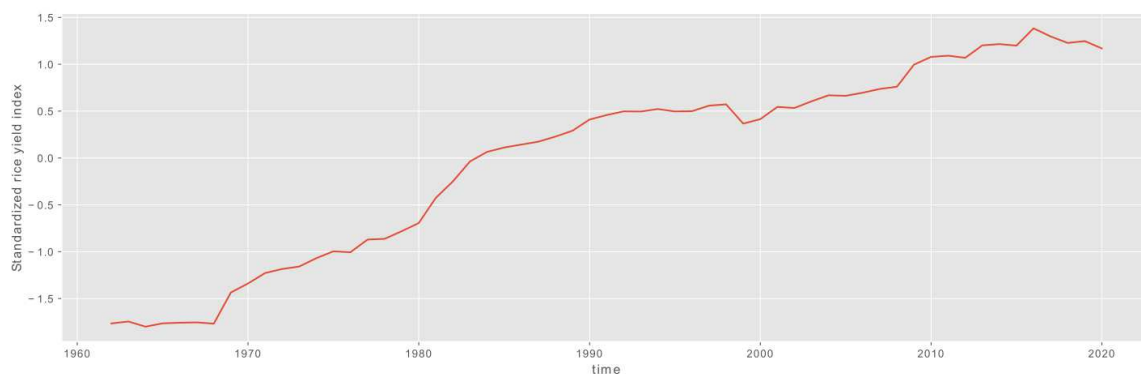


**Fig. 1** ASRYI in Indonesia from 1961 to 2019. In general, there was a rapid increase from 1961 to 2019, which is in accordance with a study conducted by

In this study we use Euclidean distance to measure $d(x_i, y_i)$. We need to trace back the components of the cost matrix from $D_{N,N}$ to $D_{0,0}$ to get the alignment of these two sequences. To automate this computation, we use an implementation of the FastDTW algorithm which is implemented in the Python computing environment via the **dtw-python** library. The 2D cost matrix representation is shown in Figure 2.

With the minimum DTW path with minimum distance of 54.87, it can be concluded that the synchronization process between the two-time series is not efficient in terms of computational costs. However, we cannot conclude whether there is no statistical relationship between the ENSO and rice production in Indonesia. For this reason, we explore causal relationship between these two signals using the Wavelet Coherence (WTC) algorithm.

The wavelet transform algorithm is intended to localize the change in time ($\Delta t$) and angular frequency ($\Delta\omega$) of the time-series data into a function of frequency with respect to time, so that we can know changes in time and frequency simultaneously. One of the wavelet functions that is often used in the fields of climatology is the Morlet transform, known as the continuous wavelet transform (CWT), with the following equation,

$$\psi(\eta) = \pi^{-\frac{1}{4}} e^{i\omega_0} e^{-\frac{\eta}{2}} \tag{5}$$

where $\pi^{-\frac{1}{4}}$ is the normalized term and $\eta = -\frac{\eta}{s}$ the dimensionless temporal parameter which is the ratio of n as a time parameter and s as a scale parameter. The angular frequency parameter $\omega_0$ is defined as $\omega_0 = \omega_s$. In the Morlet transform, the period of oscillation is $P = 1.03s$. CWT for a time-series $x$ is formally defined in the equation (6) as follows,

$$W_n(s) = \sum_{n'=0}^{N-1} x'_n \psi^* \left( \frac{(n'-n)\delta \tau}{s} \right) \tag{6}$$
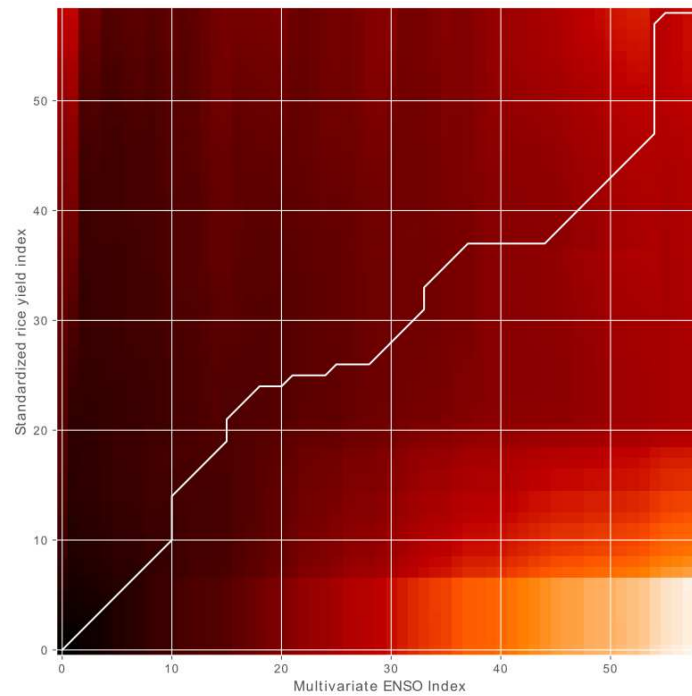


**Fig. 2** Lattice representation of cost matrix between ASRYI in Indonesia and annual average of MEI. The minimum path is shown in the white convex line.

In order to determine the relationship between the two-time series, in the context of this study between ASRYI and annual average of MEI, a complex conjugate multiplication these two CWT is needed as shown in equation (7),

$$W_n^{xy}(s) = W_n^x(s) W_n^{y^*}(s) \tag{7}$$

Phase coherence $R_n(s)^2$ in WTC is obtained by normalizing and filtering the power spectrum $|W_n^{xy}(s)|$ on the scale and time. Time smoothing is performed by normalizing using a Gaussian profile, while scale smoothing is performed using a boxcar filter. These two operations are indicated by the $\langle \dots \rangle$ expression in the following equation,

$$R_n^2(s) = \frac{\left| \langle s^{-1} W_n^{xy}(s) \rangle \right|^2}{\langle s^{-1} |W_n^x(s)|^2 \rangle \left\langle s^{-1} |W_n^y(s)|^2 \right\rangle} \tag{8}$$

To calculate the phase difference between these two-time series, we use the comparison of the imaginary term and the real term of $W_n^{xy}$ as follows,

$$\phi(s) = \arctan\left(\frac{\Im\{s^{-1}W_n^{xy}(s)\}}{\Re\{s^{-1}W_n^{xy}(s)\}}\right) \tag{9}$$

We use the biwavelet package in the R computing environment to automate the WTC calculation process. This package is an implementation of the algorithm used by. This 2D representation of the WTC is shown in Figure 3. Time is displayed on the horizontal axis, while the vertical axis shows the frequency (the lower the frequency, the higher the scale). Regions in time-frequency space where the two-time series co-vary are located by the WTC. Warmer colors represent regions with significant interrelation, while colder colors signify lower dependence between the sequences. Cold regions beyond the significant areas represent time and frequencies with no dependence in the series. Arrows in this figure represents the lead/lag-phase relationship between annual average of MEI and ASRYI. A zero-phase difference means that the two sequences move together on a particular scale. Arrows point to the right when the time series are in phase, yet if the arrows are pointing to the right, then both time sequences are in anti-phase. When this two-series data are in phase, it indicates that they move in the same direction, and anti-phase means that they move in the opposite direction. Arrows pointing to the right-down or left-up indicate that annual average of MEI is leading, while arrows pointing to the right-up or left-down show that the ASRYI is leading. In Figure 3, we do not see any significant driving pattern of annual average of MEI on ASRYI, which belongs to the 95% cone of influence (COI) determined by two degrees of freedom in the $\chi^2$ distribution.
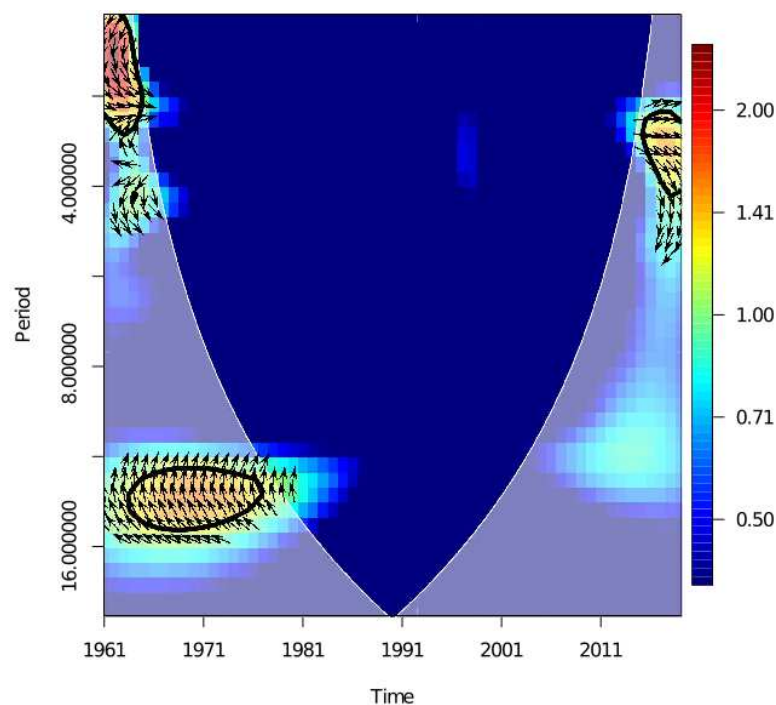


**Fig. 3** WTC between MEI and the rice yield index.

## 3.      Results and Discussions

Based on the results we get from the DTW and WTC algorithms, we do not find a causal relationship between the annual average of MEI and the ASRYI in Indonesia. However, to further ensure that there is no statistical significance between these two variables, we try to make MEI as a covariate component of the ASRYI sequence in the univariate Bayesian structural time-series (BSTS) framework. We model the ASRYI $y_t$ into a local linear trend with seasonality and regression (LLTSR) model as follows,

$$y_t = \mu_t + \tau_t + \beta^T x_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2) \qquad (10)$$

where $\mu_t$ is a local linear trend, $\tau_t$ is a seasonal component which is defined in equation (11) as follows,

$$\tau_t = -\sum_{s=1}^{s-1} \tau_{t-s} + \omega_t, \qquad \omega_t \sim \mathcal{N}(0, \sigma_\omega^2) \qquad (11)$$

According to the total data points, the number of seasons $S$ that we use in this model is 59. We use a single vector covariate, the annual average of MEI $x_t$ as the spike in the slab-spike priors, or in other words we have the assumption that this variable will have a significant effect on ASRYI. The regression coefficient $\beta$ is calculated numerically. In this model, we also assume that the residual terms $\omega_t$ and $\varepsilon_t$ follow Gaussian distribution. We obtained the posterior density distribution of ASRYI by using the Kalman filter, Kalman smoother, and 1000 times Markov Chain Monte Carlo (MCMC) process automatically by utilizing the **bsts** package in the R computing environment. Based on the posterior density distribution of the non-residual components of the LLTSR model (Figure 4), we can see that the prior regression component that we expect to have an effect, namely the annual average of MEI, actually barely contributes to the ASRYI predictability. This is also further confirmed by the negative inclusion probability value of the $\beta$ regression coefficient which does not reach 20% (Figure 5).
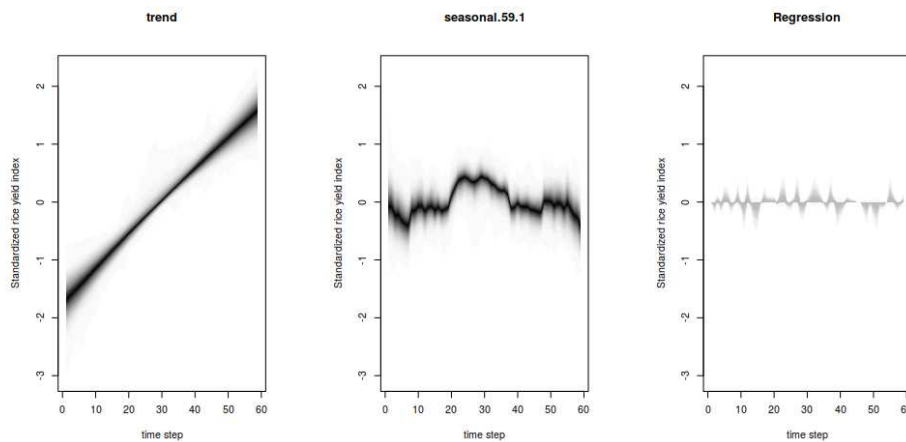


**Fig. 4** Posterior density distribution of local linear trend $\mu_t$ (left), seasonal component $\tau_t$ (center), and annual average of MEI $\beta^T x_t$ as a regression component (right).
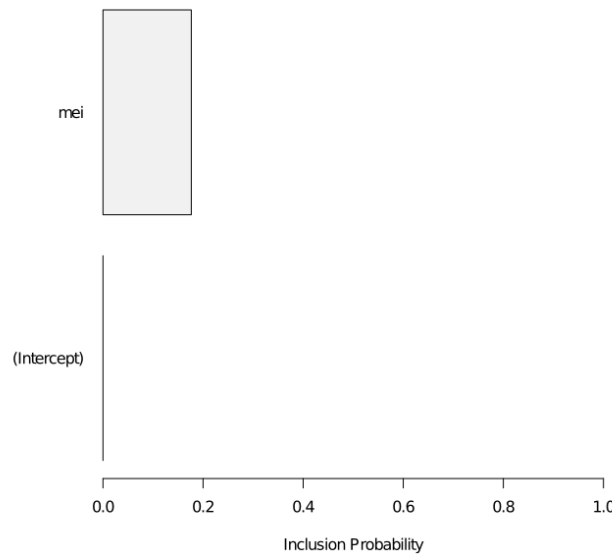


**Fig. 5** Covariate significance of the annual average of MEI. Grey bar here represents negative regression coefficient $\beta$.

More factors were likely to influence the production of rice in Indonesia. On farming and off farming techniques would also contribute to the fluctuation of rice production [26, 27, 28].

## 4. Conclusion

As the results of those three methods, we do not find any association between MEI and yearly rice production. We believe the results are caused by a few reasons. First one, the uncorrelated event may be the consequence of having yearly periodicity for rice production rather than few months in a year. Also, using a whole of Indonesian data rather than only specific region, may also arise problems. Certainly, a more detailed inter annual time series analysis is also needed to solve this problem. We hope this would open more room for research opportunities for Indonesian researchers.

## Code and data availability

Data are available through the cited sources throughout the article. Python and R scripts used in this study are accessible at *https://github.com/sandyherho/ensoIndoRice*. Rice production is provided by FAO *(http: //www.fao.org/faostat/en/#home)* and the Multivariate Enso Index (MEI) is available from NOAA (*https: //psl.noaa.gov/enso/mei/*).

## Competing interest

The authors declare that they have no conflict of interest.

## Author contribution

All authors designed the overall study. All authors contributed to the interpretation of the results and writing of the paper.

## Acknowledgements

## References

[1] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear Processes in Geophysics*, vol. 11, no. 5/6, pp. 561–566, 2004.

[2] A. H. Malian, S. Mardianto, and M. Ariani, "Faktor-Faktor yang Mempengaruhi Produksi, Konsumsi dan Harga Beras serta Inflasi Bahan Makanan," *Jurnal Agro Ekonomi*, vol. 22, pp. 119–146, Sep 2016.

[3] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sept. 2020.

[4] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological Society*, vol. 79, no. 1, pp. 61–78, 1998.

[5] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, AAAIWS'94, p. 359–370, AAAI Press, 1994.

[6] D. R. Panuju, K. Mizuno, and B. H. Trisasongko, "The dynamics of rice production in indonesia 1961–2009," *Journal of the Saudi Society of Agricultural Sciences*, vol. 12, no. 1, pp. 27–37, 2013.

[7] D. S. Bloomfield, R. T. J. McAteer, B. W. Lites, P. G. Judge, M. Mathioudakis, and F. P. Keenan, "Wavelet phase coherence analysis: Application to a quiet-sun magnetic element," *The Astrophysical Journal*, vol. 617, no. 1, pp. 623–632, 2004.

[8]   FAO, "Food and Agriculture Organization of the United Nations. FAOSTAT Statistical Database," 1997. [Online; accessed 5. Jul. 2021].

[9]   G. van Rossum, "Python tutorial," Tech. Rep. CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.

[10]  H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[11]  J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[12]  K. Wolter and M. S. Timlin, "Measuring the strength of enso events: How does 1997/98 rank?," *Weather*, vol. 53, no. 9, pp. 315–324, 1998.

[13]  M. Ambarinanti, "Analisis faktor-faktor yang mempengaruhi produksi dan ekspor beras Indonesia," *IPB (Bogor Agricultural University)*, 2007.

[14]  M. G. Roberts, D. Dawe, W. P. Falcon, and R. L. Naylor, "El niño–southern oscillation impacts on rice production in luzon, the philippines," *Journal of Applied Meteorology AND Climatology*, vol. 48, pp. 1718–1723, 2009.

[15]  Mahananto, S. Sutrisno, and C.-F. Ananda, "Faktor-faktor yang mempengaruhi produksi Padi Studi Kasus di Kecamatan Nogosari, Boyolali, Jawa Tengah," *Wacana Journal of Social and Humanity Studies*, vol. 12, no. 1, pp. 179–191, 2009.

[16]  N. V. Nguyen, "Global climate changes and rice food security," 2005. [Online; accessed 14. Jul. 2021].

[17]  N. W. Ismail and S. M. Chan, "Impacts of the el niño-southern oscillation (enso) on paddy production in southeast asia," *Climate and Development*, 2019.

[18]  R Core Team, R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[19]  R. L. Naylor, W. P. Falcon, D. Rochberg, and N. Wada, "Using el niño/southern oscillation climate data to predict rice production in indonesia," *Climatic Change*, vol. 50, pp. 255–265, 2001.

[20]  "Read Rectangular Text Data [R package readr version 2.0.1]," Aug 2021. [Online; accessed 18. Aug. 2021].

[21]  S. L. Scott, bsts: Bayesian Structural Time Series, 2020. R package version 0.9.5.

[22]  S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, pp. 70–80, 2004.

[23]  S. Scott and H. Varian, "Predicting the present with bayesian structural time series," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 5, pp. 4 – 23, 2014.

[24]  S. Wright, "Largest Rice-Producing Countries," *WorldAtlas*, Jul 2020.

[25]  T. C. Gouhier, A. Grinsted, and V. Simko, R *package biwavelet: Conduct Univariate and Bivariate Wavelet Analyses*, 2021. (Version 0.20.21).

[26]  T. Giorgino, "Computing and visualizing dynamic time warping alignments in r: The dtw package," *Journal of Statistical Software*, vol. 31, no. 7, pp. 1–24, 2009.

[27]  T. pandas development team, *pandas-dev/pandas*: Pandas, Feb. 2020.

[28]  X. Deng, J. Huang, F. Qiao, R. L. Naylor, W. P. Falcon, M. Burke, S. Rozelle, and D. Battisti, "Impacts of el niño-southern oscillation events on china's rice production," *Journal of Geographical Science*, vol. 20, no. 1, pp. 3–16, 2010.