USING LONG SHORT-TERM MEMORY MODEL FOR CLOUD FOREST VEGETATION GROWTH STATUS PREDICTION - A CASE STUDY IN SHEI-PA NATIONAL PARK

G. G. Wang [1], H. P. Tsai [1,2] *

[1] Dept. of Civil Engineering, NCHU, 145 Xingda Rd., Taichung 402, Taiwan –d110062002@mail.nchu.edu.tw (G. G. Wang); huiping.tsai@nchu.edu.tw(H. P. Tsai);
[2] Pervasive AI Research (PAIR) Labs, Hsinchu 300, Taiwan

KEYWORDS: Cloud Forest, NDVI, LSTM, Climate Change, time series.

ABSTRACT:

Cloud Forests (CFs) are characterized by their persistent foggy environment, in which fog can save two times the amount of precipitation in the dry season and increase water storage by 10% in the rainy season. CFs play an important role in ecosystems as high biodiversity and abundant endemic species live within CFs. However, CFs are sensitive to environmental changes, especially in current global climate warming conditions. Therefore, a typical cloud forest in Taiwan, Shei-Pa National park, was chosen as the study area. Specifically, the Normalized Difference Vegetation Index (NDVI) with meteorological factors including rainfall, average temperature, maximum temperature, and minimum temperature were obtained to assess the overall CFs trend from 2001 to 2017. Moreover, the Long Short-Term Memory neural network model (LSTM) was implemented to predict the future vegetation status. Preliminary results have shown that vegetation condition in Shei-Pa National park was getting better; rainfall, average temperature, and minimum temperature represented an upward trend while maximum temperature showed a downward trend. Furthermore, the LSTM- maximum temperature model displayed the highest prediction power with the MAPE index of 4.84%. The results provide a valuable reference for forest resource conservation and future climate adaptation strategies in Taiwan.

1. INTRODUCTION

According to the general statistics of the United Nations Environmental Programme (UNEP) and the World Conservation Monitor Centre (WCMC), The CFs area are accounted for approximately 0.26% of the global land area and 2.5% of the tropical forest area (Aldrich *et al*., 1997). However, CFs are unevenly distributed globally, with 60% in Asia, 25% in the Americas, and 15% in Africa (Bubb *et al*., 2004).

Cloud Forests (CFs) are characterized by foggy conditions throughout the year (Bruijnzeel & Proctor, 1995). The foggy condition produces significant horizontal precipitation, which becomes occult precipitation in CFs. The occult precipitation in CFs can save more than two times the precipitation in the dry season and increase 10% of the forest water storage capacity in the rainy season, which is very important for conserving moisture in the soil and water (Bruijnzeel *et al.*, 2011). According to the results of the fourth national forest resources survey in Taiwan and the research results from Schulz (Qiu *et al*., 2015; Schulz *et al.*, 2017), the CFs area accounted for approximately 25% of the forest area in Taiwan with a relatively high proportion of the endemic species.

In recent years, global warming and extreme climates have affected forests, thereby affecting the functions of forest ecosystems (Bubb et al., 2004). For instance, Chang et al. (2008) found that when the temperature rises, the soil will release the organic carbon stored in the past, making soil change from a carbon sink to a carbon source, which increases the carbon in the atmosphere. Furthermore, Los et al. (2019) found that changing cloud base height will impact the forest area. They calculated cloud base height and occult precipitation. They concluded that under the circumstance of 2°C warming in the future, the cloud base height would increase by 250 m, which will reduce the forest area by about 50% or even disappear completely. Therefore, understanding the current status and future changes in CFs is important for ecological conservation and biodiversity maintenance.

2. METHOD

2.1 Research area

Taiwan is rich in forest resources, with 51 high mountains over 3,000 meters above sea level (ASL). Based on a study conducted by Schulz et al. (2017), Taiwan has more than 5,500 km^2 of CFs, of which 98% of them are located at 1,000-3,000 ASL. Therefore, the Shei-Pa National Park, located in central-north Taiwan with 675 km^2 spatial coverage, is chosen as the study area due to its specific geology terrain, mountain ecology resources, and rich endemic species. The location of research area is shown in Figure 1 (a), and the vegetation species in research area, such as plantations, coniferous forests, and broadleaf forests, are shown in Figure 1 (b) , and elevations in research area between at 1,000-3,000 ASL is shown in Figure 1 (c).

* Corresponding author

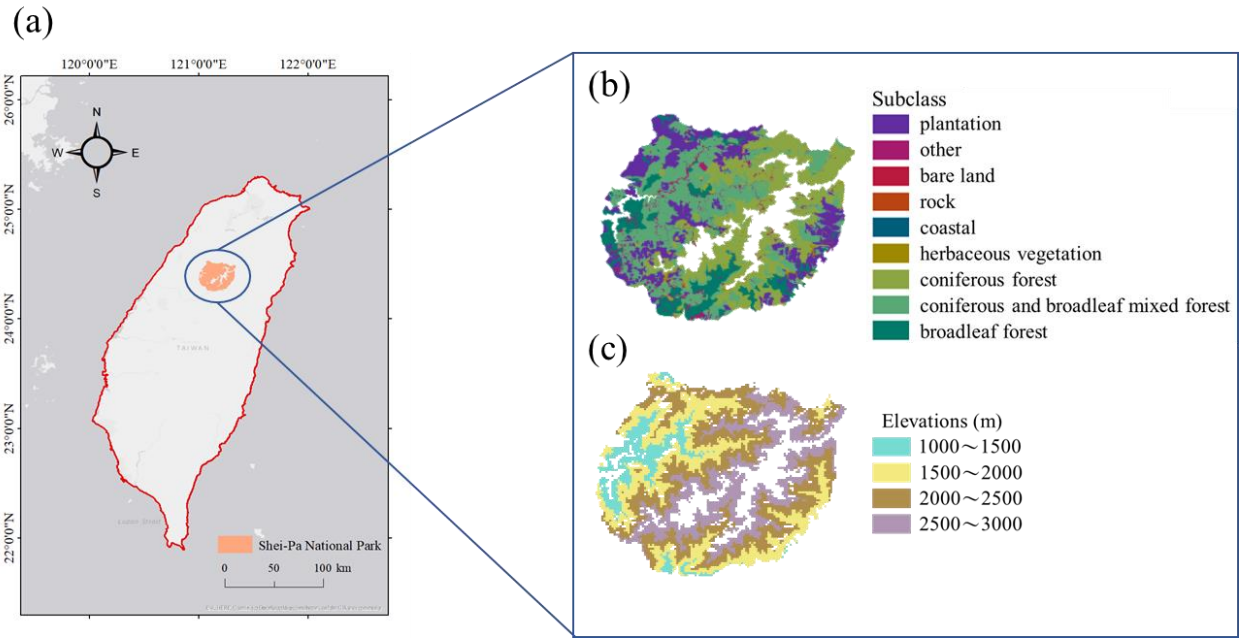

Figure 1 (a) Research area (b) Research area vegetation species (c) Research area elevations

2.2 Research flow

The research flow is shown in Figure 2. First, satellite-based vegetation data, meteorological data, and digital elevation data were collected. The Normalized Difference Vegetation Index (NDVI) images of the MOD13Q1 product were obtained from the Moderate-resolution imaging spectroradiometer (MODIS), and four meteorological factors, including rainfall, average temperature, maximum temperature, minimum temperature, are attained from Taiwan Climate Change Projection and Information Platform (TCCIP). Next, the Department of Land Administration produced the Digital Elevation Model (DEM), Ministry of the Interior. The NDVI and meteorological factors were rearranged in monthly time series. Then the second step referred to the CFs range defined by Schulz et al. (2017) and extracted the CFs area from 1,000 to 3,000 meters ASL in Shei-Pa National Park. Next, the Mann-Kendall and Seasonal Mann-Kendall analyses were implemented on NDVI and meteorological factors. In the final step, we would like to establish a prediction LSTM model for CFs. With NDVI and meteorological factors, DEM information was also considered in the LSTM prediction model, and the prediction capability was evaluated.

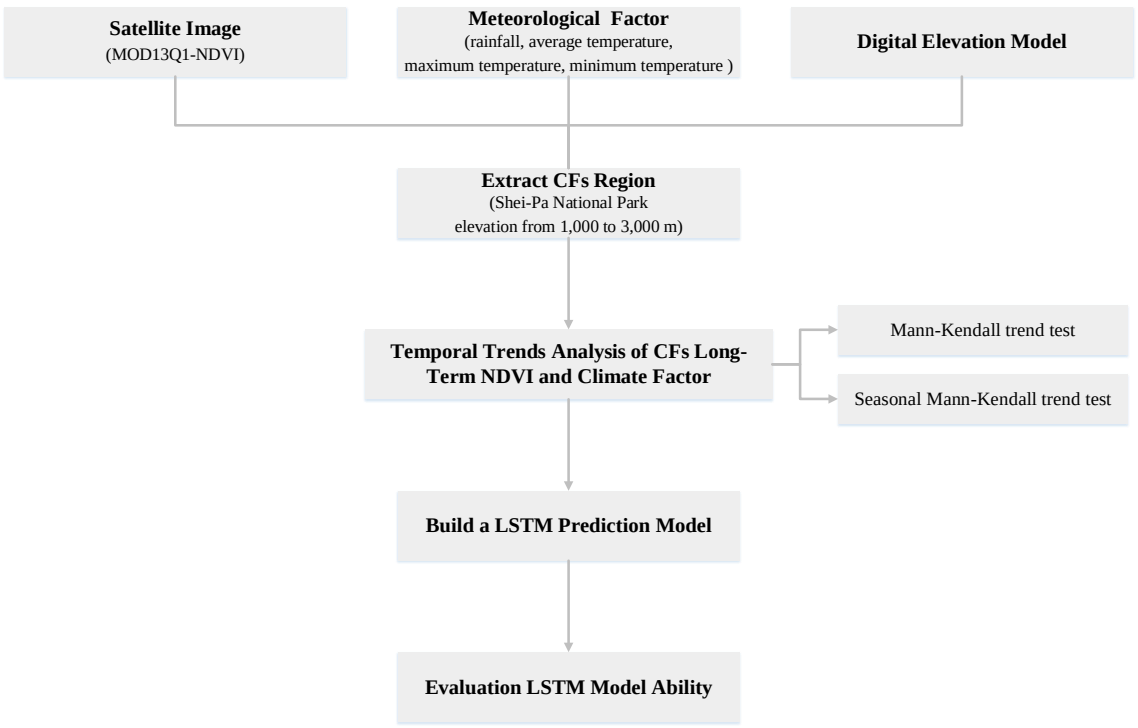

Figure 2 Research flow

2.3 Normalized Difference Vegetation Index (NDVI)

This study obtained a time series of Normalized Difference Vegetation Index (NDVI) values from 2001 to 2017 derived from the MOD13Q1 product of Moderate-resolution imaging spectroradiometer (MODIS). With a spatial resolution of 250 meters × 250 meters and one data entry every 16 days, total of 391 NDVI images were obtained. The NDVI algorithm formula is as follows (Rouse et al., 1974):

$$NDVI = \left(\frac{\rho NIR - \rho RED}{\rho NIR + \rho RED}\right) \quad (1)$$

where ρ = reflectance rate of band
NIR = Near Infrared
RED = Red

2.4 Meteorological factor

The meteorological factors used in this study include rainfall (Rain), average temperature (Tavg), maximum temperature (Tmax), and minimum temperature (Tmin). All meteorological factors data were derived from the Taiwan Climate Change Projection and Information Platform (TCCIP) in monthly gridded format with a spatial resolution of 5 km × 5 km (Lin et al., 2011). We have obtained those meteorological data from 2001 to 2017 for the study area.

2.5 Digital Elevation Model data

The Digital Elevation Model (DEM) is a gridded data produced by the Department of Land Affairs, Ministry of the Interior in Taiwan with a spatial resolution is 20 m × 20 m. Each grid point has the plane coordinates and elevation information of the point. The plane coordinates are based on the Ministry of the Interior 1997. The Taiwan Geodetic Datum Coordinate (TWD1997), the elevation coordinate datum is the Ministry of the Interior 2001 Taiwan Elevation Datum (TWVD2001) (Ministry of the Interior, Department of Land Affairs, 2016).

2.6 Mann-Kendall trend test and Seasonal Mann-Kendall test

The Mann-Kendall test (M-K test) was used to evaluate the trend in this study; the M-K test was developed by Mann and Kendall (Mann, 1945; Kendall, 1948), which is a nonparametric statistical has been widely used and method recommended by the World Meteorological Organization (IMO). The M-K test has distribution-free characteristics and a high tolerance for extreme values in the series and missing data (Hamed, 2008; Pohlert, 2019), which is calculated as follows.

Suppose a time series is:

$$X = \{x_1, x_2, \dots, x_n\} \quad (2)$$

The M-K test statistic S is:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(x_k - x_j) \quad (3)$$

Among

$$\mathrm{sgn}(x_k - x_j) = \mathrm{sgn}(R_k - R_j) = \begin{cases} 1 & if \quad x_j < x_k \\ 0 & if \quad x_j = x_k, \\ -1 & if \quad x_j > x_k \end{cases} \quad (4)$$

R_j and R_k in formula (4) represent the corresponding order positions of x_j and x_k in the time series, sgn is a symbolic function. The M-K test assesses the order of the data rather than its true value, and it has the property that the sample must be normally distributed.

The M-K test assumes that each time series data is Independent and Identically Distributed. The mean and variance V of its statistic value S can be defined as:

$$E(S) = 0 \quad (5)$$

$$V(S) = \frac{\{n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i-1)(2t_i+5)\}}{18} \quad (6)$$

Where m is the number of groups with the same value, and t_i in each group represents the number of values in the i-th group with equal values. Kendall (1948) also found that when the total number of data in the time series is larger, the statistical value S is closer to the normal distribution. At this point, the standard normal variable Z is used with a self-defined statistical significance level (such as a 0.05 level) to judge whether the time series data has a statistically significant trend:

$$\mathrm{Z} = \begin{cases} (S-1)/\sqrt{V_0(S)} & if \quad S > 0 \\ 0 & if \quad S = 0 \\ (S+1)/\sqrt{V_0(S)} & if \quad S < 0 \end{cases} \quad (7)$$

When $|Z| > Z(\propto/2)$, it means that the time series data has a significant trend; a positive Z value indicates a significant upward trend while a negative Z value indicates a significant downward trend.

This study intends to select a significant level of 5% ($Z(\propto/2) = 1.96$), Therefore, if Z>1.96, it has a significant trend, and if Z<1.96, there is no statistically significant increase or decrease trend.

We would also like to evaluate the seasonal trend of the CFs. Therefore, the Seasonal Mann-Kendall test (S M-K test) proposed by Hirsch *et al.* (1982) was utilized. The calculation method of the S M-K test is similar to the M-K test (Hirsch & Slack, 1984); the test method is the same; the only difference is that the S M-K test divides the season or the month to calculate the M-K test and adds them together eliminate the influence of seasonality. The algorithm of the S M-K test is shown as follows (Pohlert, 2019):

$$S_g = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn\left(x_{kg} - x_{jg}\right), g = 1, 2, 3, \dots, m \quad (8)$$

2.7 Long Short-Term Memory, LSTM

The Long Short-Term Memory (LSTM), an improved version of Recurrent neural networks (RNN), was proposed by Hochreiter & Schmidhuber in 1997 (Hochreiter & Schmidhuber, 1997). The biggest difference between LSTM and RNN is that each neuron of LSTM is a memory unit and can improve the shortcomings of short-term memory in RNN. Each neuron in LSTM contains three gates, forget gate, an input gate, and an output gate. Detailed LSTM information can be found in Yu et al. (2019). The function of the forget gate is to selectively forget some unimportant information of the gates from the former one; the input gate controls the new information that it learns and decides whether to maintain the information in the memory in the cell; the output gate determines whether the information in the memory should be output or not. The mathematical function of LSTM is as follows:

$$f_t = \sigma\left(W_{fh} h_{t-1} + W_{fx} x_t + b_f\right) \quad (9)$$

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i), \quad (10)$$

$$\tilde{c}_t = tanh(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\tilde{c}}) \quad (11)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t, \quad (12)$$

$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \quad (13)$$

$$h_t = o_t \cdot tanh(c_t) \quad (14)$$

where f_t = forget gate
i_t = input gate
$\tilde{c}_t$ = recurrent unit
c_t = cell state
o_t = output gate
h_t = hidden layer

3. RESULTS AND DISCUSSION

3.1 Mann-Kendall trend test and Seasonal Mann-Kendall test

In this study, we applied the M-K test and S M-K test to analyze trends in NDVI and meteorological factors of CF in Shei Pa National Park from 2001 to 2017. In Figure 3, the results of the M-K test show that the NDVI vegetation growth status of Shei-Pa National Park CFs was getting greener (the trend is significant with a Z value>1.96). In terms of meteorological factors, the rainfall, average temperature, and minimum temperature showed positive trends, and the maximum temperature presented a negative trend. The S M-K test results showed that the NDVI vegetation growth state was significantly greener (with a Z value>1.96). The meteorological factors, including rainfall, average temperature, and minimum temperature, all presented positive trends, while the maximum temperature was negative. Most of the trends were significant (Z value>1.96); only the rainfall trend was not significant (Z value<1.96).

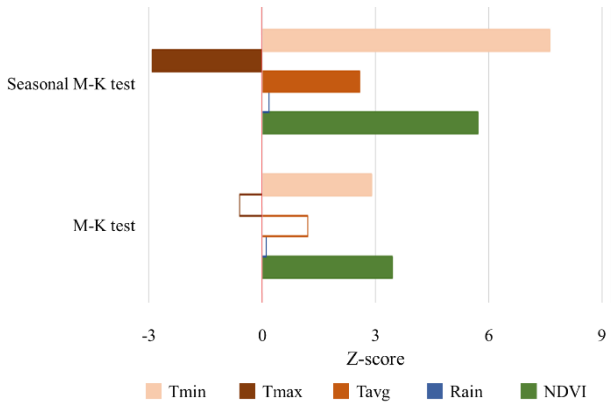

Figure 3 M-K test and Seasonal M-K test of Shei-Pa National Park NDVI and Meteorological factor

3.2 LSTM

A total of six LSTM models with different architectures of one and two hidden layers were established for testing. Three temperature factors, average temperature (Tavg), maximum temperature (Tmax), and minimum temperature (Tmin), were combined with NDVI, DEM, and rainfall to form a data set. The data from 2001 to 2014 was used for training, 80% of which was used for training and 20% for validation. The data from 2015 to 2017 were used for testing by 500 epochs. Each data set was divided into five datasets for cross-validation. The results showed that the data is evenly distributed. Furthermore, the model is stable with the maximum and minimum training loss values of 0.040 and 0.0023 and maximum and minimum validation loss values of 0.006 and 0.0023, respectively (Figure 4 and Figure 5). The results of LSTM are shown in Figure 6, with the average MAPE between 4% to 5%. The models with two hidden layers performed better than those with one. Among the six models, the two hidden layers model with the maximum temperature combined with NDVI, DEM, and rainfall performs the best with the MAPE value of 4.71% (marked with a star in Figure 6). Therefore, we presume that the maximum temperature information contributes more to the model's predictive ability. Then the model is retrained with all the data from 2001 to 2014, and its MAPE is 4.84%.

Based on the results of LSTM models, the differences between the predicted and observed NDVI in 2016 and 2017 are shown in Figures 7 and 8. The absolute values of the error less than 0.1, overestimation, and underestimation are shown in yellow, red, and blue, respectively. Most of the areas show errors less than 0.1, which indicates a good prediction capability of the LSTM model. However, a small proportion of pixels are overestimated, while some in January 2017 are underestimated. Additionally, in January, March, May, August of 2016, and January and May of 2017, some errors exceeded 5%.

Moreover, it is found that most of the locations with inaccurate predictions have more than two types of forests (Taiwan national natural vegetation maps, 2014), which may indicate complicated interactions with each other; thus, it is difficult to predict accurately. In addition, the prediction results at the edges of the study area are also poor.

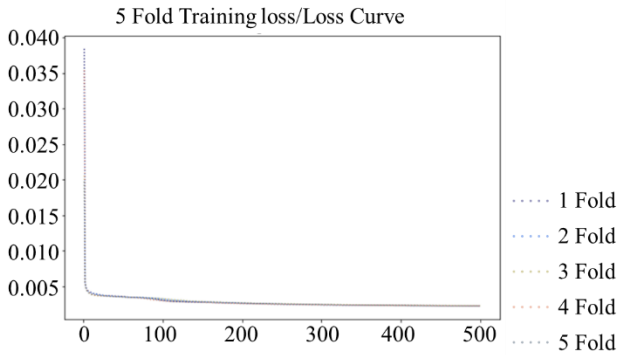

Figure 4 Results of 5 Fold Training loss

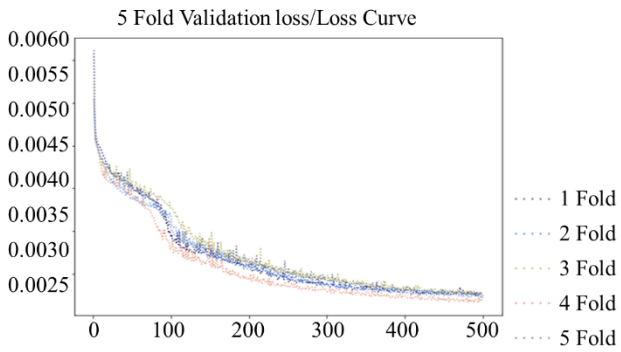

Figure 5 Results of 5 Fold Training loss

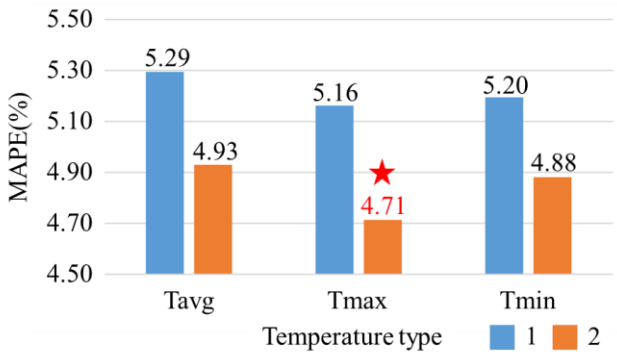

Figure 6 MAPE values of the six LSTM models

4. CONCLUSIONS

The M-K test results of the trend assessment of CFs in Shei-Pa National Park showed that its vegetation state was growing well (a significantly positive trend in NDVI). In addition, the minimum temperature increased significantly. Based on the S M-K test results, the NDVI, average temperature, and minimum temperature of Shei-Pa National Park CFs showed a significant positive trend. However, the maximum temperature dropped significantly, and the rainfall had no significant trend.

The LSTM prediction model was established based on NDVI, rainfall, DEM, and the maximum temperature data with a good prediction performance of MAPE 4.84%. However, the prediction results of the LSTM model in January and May are less accurate.

This study analyzed the trends of CFs vegetation and meteorological factors in Shei-Pa National Park. It used the LSTM model with a highly accurate prediction ability to make predictions, which accurately grasped the growth of vegetation in the past and predicted the possible conditions of vegetation in the future.

Considering possible challenges of future climate change, the prediction results of this study can provide a valuable reference for forest management authorities to support decision-making regarding forest conservation. More importantly, the results of this study serve as a scientific basis for establishing adaptation strategies, which are of positive benefit to the overall ecological environment of Taiwan.

REFERENCES

Aldrich, M., Billington, C., & Edwards, M. (1997). A global directory of tropical montane cloud forests. Draft..

Bruijnzeel, L. A., & Proctor, J. (1995). Hydrology and biogeochemistry of tropical montane cloud forests: what do we really know?, *Tropical montane cloud forests*, New York, NY, 1995, 38-78.

Bruijnzeel, L. A., Scatena, F. N., & Hamilton, L. S. (Eds.). (2011). *Tropical montane cloud forests: science for conservation and management*. Cambridge University Press.

Bubb, P., May, I. A., Miles, L., & Sayer, J. (2004). Cloud forest agenda.

Chang, S. C., Tseng, K. H., Hsia, Y. J., Wang, C. P., & Wu, J. T. (2008). Soil respiration in a subtropical montane cloud forest in Taiwan. *Agricultural and Forest Meteorology, 148*(5), 788-798.

Hamed, K. H. (2008). Trend detection in hydrologic data: the Mann–Kendall trend test under the scaling hypothesis. *Journal of hydrology*, *349*(3-4), 350-363.

Hirsch, R. M., & Slack, J. R. (1984). A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, *20*(6), 727-732.

Hirsch, R. M., Slack, J. R., & Smith, R. A. (1982). Techniques of trend analysis for monthly water quality data. *Water resources research*, *18*(1), 107-121.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Kendall, M. G. (1948). Rank correlation methods.

Lee, L. Y et al. (2011). Taiwan Climate Change Projection Information and Adaptation Knowledge Platform Construction Project Second Year Final Report.

Los, S. O., Street-Perrott, F. A., Loader, N. J., Froyd, C. A., Cuní-Sanchez, A., & Marchant, R. A. (2019). Sensitivity of a tropical montane cloud forest to climate change, present, past and future: Mt. Marsabit, N. Kenya. *Quaternary Science Reviews, 218*, 34-48.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society, 13*(3), 245-259.

Ministry of the Interior, Department of Land Affairs. (2016). 20-meter grid Digital Elevation model data of the Ministry of the Interior.

Pohlert, T. (2019). Nonparametric trend tests and change-point detection. 2018. *CC BY-ND*, *4*

Qiu, L. W., Huang, Q. X., Wu, J. Q., & Xie, X. T. (2015). Outline of the 4th National Forest Resources Survey Results. *Taiwan Forestry Journal, 41*(4), 3-13

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with ERTS proceeding. In *Third Earth Reserves Technology Satellite Symposium, Greenbelt: NASA SP-351* (Vol. 30103017).

Schulz, H. M., Li, C. F., Thies, B., Chang, S. C., & Bendix, J. (2017). Mapping the montane cloud forest of Taiwan using 12 year MODIS-derived ground fog frequency data. *PLoS One*, *12*(2), e0172663.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation, 31*(7), 1235-1270.

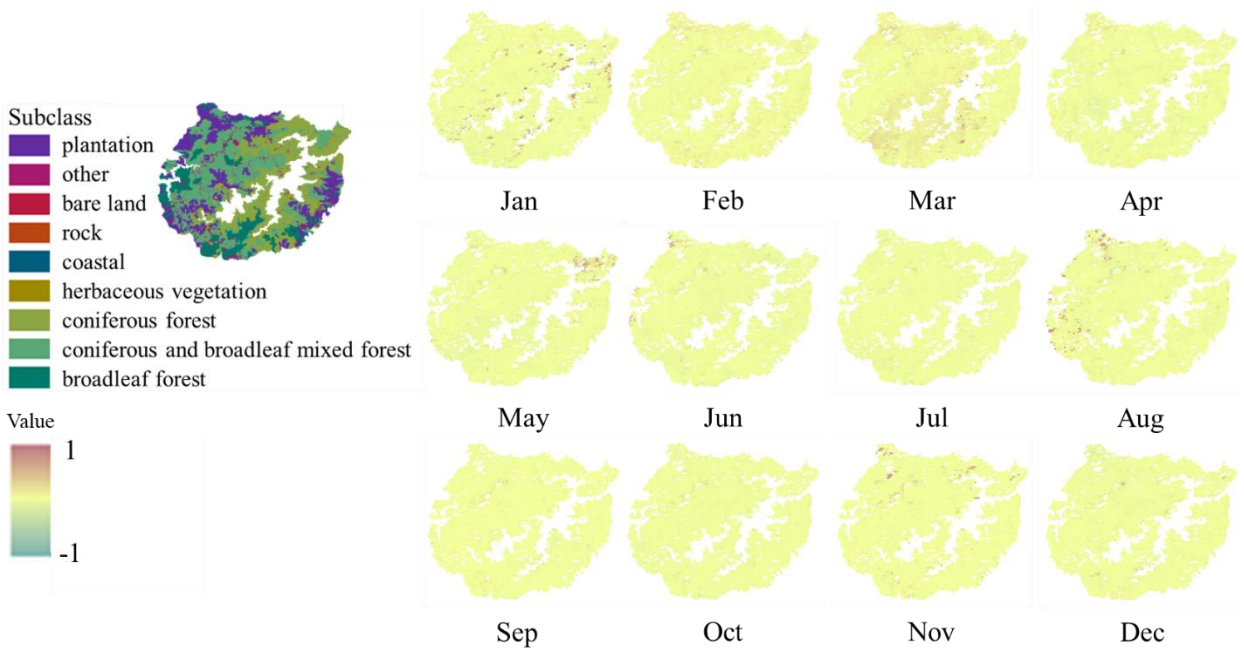

Figure 7 The differences between the predicted and observed NDVI in 2016

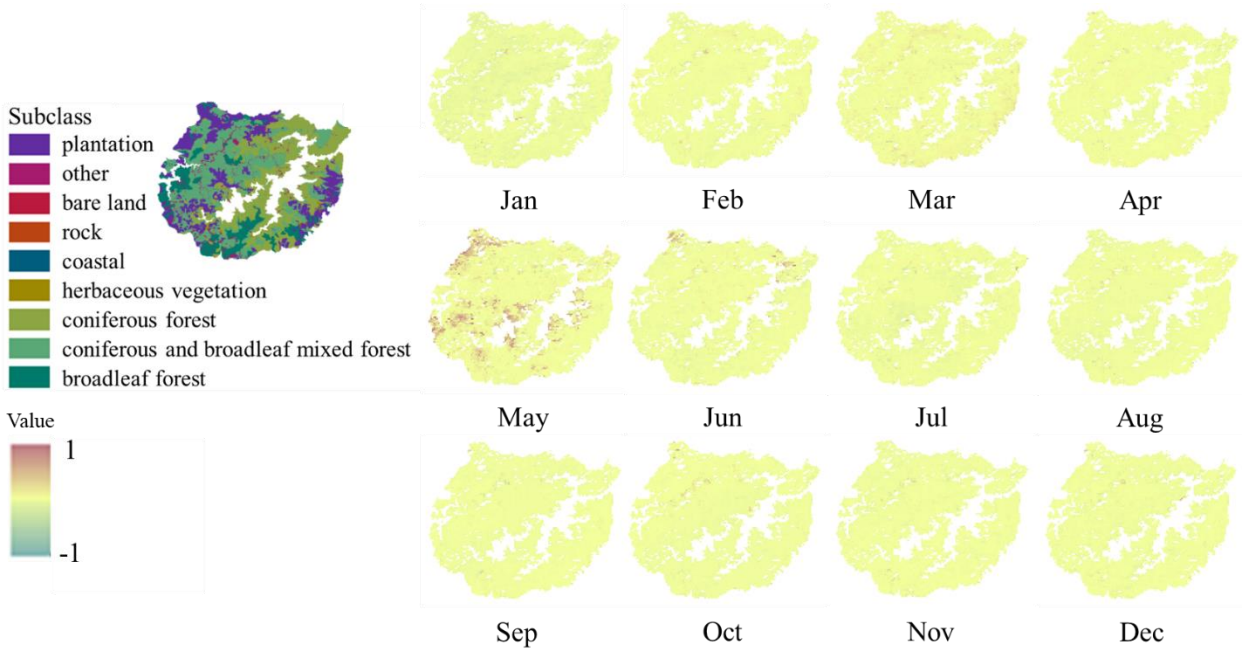

Figure 8 The differences between the predicted and observed NDVI in 2017