




Article

Feasibility of Predicting Vietnam's Autumn Rainfall Regime Based on the Tree-Ring Record and Decadal Variability

Yan Sun ^{1,*} , S.-Y. Simon Wang ^{2,3} , Rong Li ³, Brendan M. Buckley ⁴ , Robert Gillies ^{2,3} and Kyle G. Hansen ⁴

¹ Department of Mathematics and Statistics, Utah State University, Logan, UT 84322, USA

² Department of Plants, Soils, and Climate, Utah State University, Logan, UT 84322, USA; simon.wang@usu.edu (S.-Y.S.W.); Robert.Gillies@usu.edu (R.G.)

³ Utah Climate Center, Utah State University, Logan, UT 84322, USA; lirong18@gmail.com

⁴ Tree Ring Lab, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA; bmb@ldeo.columbia.edu (B.M.B.); khansen@ldeo.columbia.edu (K.G.H.)

* Correspondence: yan.sun@usu.edu; Tel.: +1-435-797-2861

Received: 24 March 2018; Accepted: 10 May 2018; Published: 16 May 2018



Abstract: We investigate the feasibility of developing decadal prediction models for autumn rainfall (R_A) over Central Vietnam by utilizing a published tree-ring reconstruction of October–November (ON) rainfall derived from the earlywood width measurements from a type of Douglas-fir (*Pseudotsuga sinensis*). Autumn rainfall for this region accounts for a large percentage of the annual total, and is often the source of extreme flooding. Central Vietnam's R_A is characterized by a pronounced decadal oscillation signal. We use the decadal mode of R_A along with its notable autocorrelation and significant cross-correlation with basin-wide Pacific sea surface temperature (SST) variability, to develop four discrete time-series models. The sparse autoregressive model, with Pacific SST as an external variable, accounts for most of the autoregressive R_A , while taking advantage of the predictability from the basin-wide Pacific climate oscillation. Using this model, the decadal prediction of R_A can be reasonably achieved with a 10-year-ahead forecasting skill score (SS) about 0.46. We therefore suggest, with this paper, that forecasting R_A for Central Vietnam for multiple years ahead is possible using a time-series model.

Keywords: Central Vietnam; time-series model; tree-ring; decadal prediction

1. Introduction

Precipitation in Central Vietnam is characterized by an autumn regime associated with seasonally increased occurrences of tropical cyclones [1,2]. In autumn, Central Vietnam undergoes frequent flooding that poses serious risks to both lives and properties [3–5]. Li et al. [1] observed a pronounced decadal oscillation within the 8–11 year frequency revealed from the autumn precipitation in Central Vietnam (Figure 1a), which appears to be modulated by the so-called East Pacific–North Pacific (EP–NP) teleconnection (Figure 1b). The negative phase of the EP–NP pattern is associated with a positive sea surface temperature (SST) anomaly in the South China Sea (SCS) that induces low-level convergence that, in turn, leads to increased precipitation over Central Vietnam. This regional circulation is embedded in a large-scale circulation associated with SST anomalies across the Pacific Ocean, with widespread cooling in the eastern and central tropical Pacific sandwiched by warming in the North, South, and Western Pacific Ocean (Figure 1b). Reversely, the positive phase of the EP–NP features opposite SST and circulation anomalies, leading to reduced rainfall in central Vietnam. The EP–NP

index also reveals a pronounced quasi-decadal oscillation (Figure 1c) that is in-sync with the central Vietnam autumn precipitation [1].

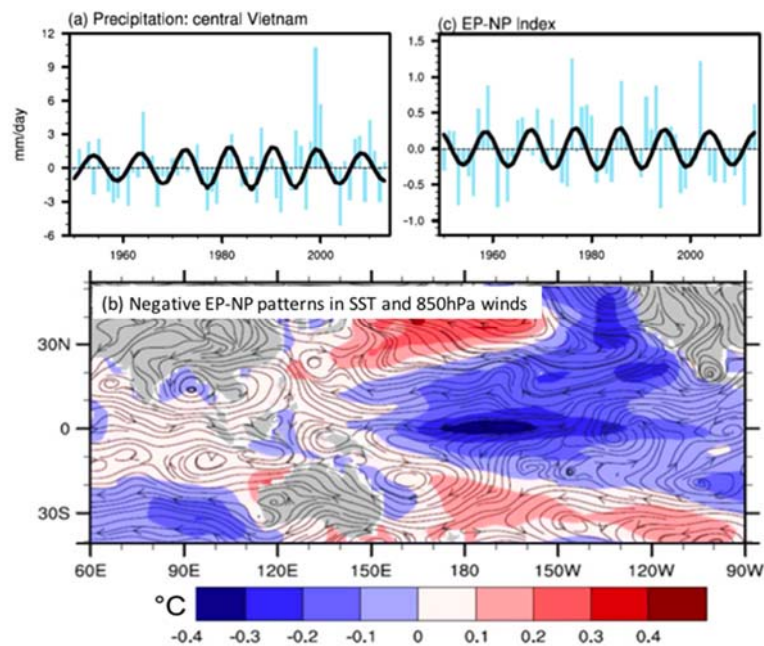


Figure 1. (a) Autumn precipitation from central Vietnam, (b) sea surface temperature (SST) and 850 hPa wind anomalies as shading and streamline, respectively, associated with the East Pacific–North Pacific (EP–NP) pattern, and (c) the EP–NP index. The thick solid lines in (a,c) are bandpass-filtered by 8–11 years; figures modified from Li et al. [1].

The decadal cycle identified by Reference [1] may be useful for developing a decadal prediction, which if proven to be feasible, could help mitigate hydroclimate extremes frequently experienced over Central Vietnam (e.g., [6]). In this study, we explored the decadal predictability of autumn precipitation, based on the previously identified decadal rainfall cycle and its EP–NP climate mode connection. However, the limited data length in the observational records prohibits the utilization of sufficiently long “training data” for models. Here, we utilized a newly developed tree-ring reconstruction of precipitation for Central Vietnam produced by Reference [7]. Additional reconstructions of SST were also adopted to depict the EP–NP climate mode. Four time-series models were developed for this study. The details of these models and data used are described in Section 2, and the major results of this study are presented in Section 3. A conclusion is provided in Section 4.

2. Materials and Methods

The tree-ring data used for this study are from a type of Douglas-fir (*Pseudotsuga sinensis*) collection from northern Vietnam, developed by Reference [7]. The trees that these authors sampled were growing on steep limestone mountains at the Kim Hy Natural Reserve in Bac Kan province (22°24' N, 106° 03' E), and the growth parameter used for the reconstruction was the width of the earlywood. Climate over this region is characterized by a warm and wet summer season through to autumn and a cold and dry season from December to April. This mountain region in Bac Kan province shows a similar climate regime to Central Vietnam. Cold surge-related wind anomalies in the lower troposphere are frequent during the winter shoulder season (ON), and can often bring anomalously high rainfall to Central Vietnam [4], further increasing moisture into the study region. While the mean average annual temperature for Bac Kan province is 25 °C, the minimum temperature drops to near freezing during the coldest months and may exceed 35 °C during much of the warm season. These climatic conditions enable the growth of Douglas-firs to be sensitive to climate/precipitation variability. In the ensuing

analysis, we use the index of ON rainfall reconstructed by Reference [7] from the earlywood widths of the Douglas-fir, denoted as autumn (R_A).

Additional data includes the NOAA (National Oceanic and Atmospheric Administration) Extended Reconstruction SST version 3 used to depict the EP–NP climate mode, following [1]; the SST data source is <https://climatedataguide.ucar.edu/climate-data/sst-data-noaa-extended-reconstruction-ssts-version-3-ersstv3-3b>.

In this paper, we construct various time-series models for predicting R_A (ON rainfall index over Central Vietnam) and compare their performances. The sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) of R_A are shown in Figure 2. Both functions exhibit significant autocorrelation, based on which a univariate time-series model is developed for R_A , from an autoregressive moving average (ARMA) model. The significant autocorrelations reflect the documented low-frequency/decadal climate variability as noted for Central Vietnam by Reference [1]. In addition, a sparse pattern is shown in the PACF, however the partial autocorrelation is only significant at a few lags such as lag 1, 6, 13, and 15. Thus, we construct a sparse autoregressive (SpsAR) model that utilizes the sparse feature to improve prediction accuracy. Figure 3 plots the sample cross-autocorrelation function (CACF) of the series R_A and EP–NP index, where a significant cross-autocorrelation structure is shown. This motivates us to include EP–NP as an exogenous variable to enhance the univariate model, thus developing an ARMAX (ARMA with exogenous variables) type of model. Finally, a sparse autoregressive with exogenous variables (SpsARX) model is constructed that simultaneously takes advantage of both the sparsity and the exogenous variable referred to as EP–NP.

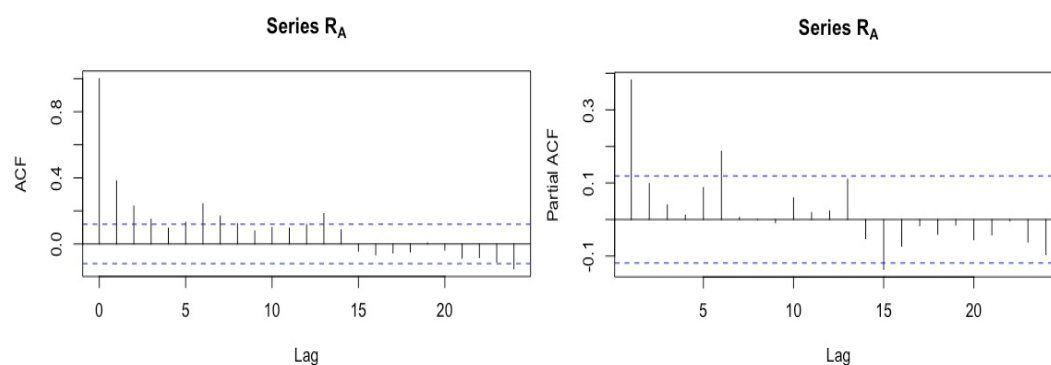


Figure 2. The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of autumn rainfall (R_A). Both of the two plots show significant autocorrelation structures that can be used to construct a time-series model.

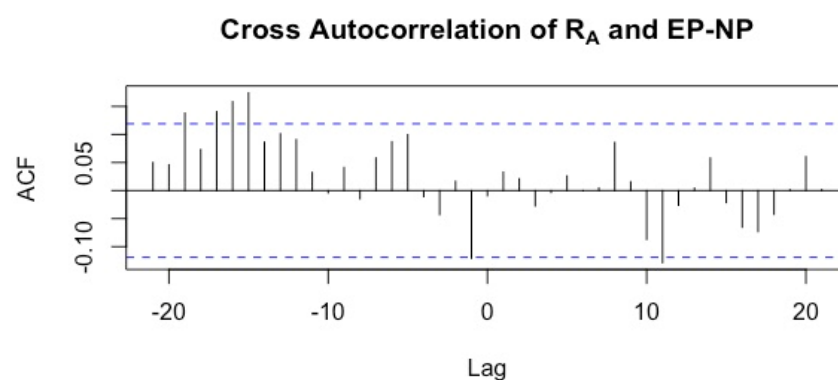


Figure 3. The sample cross-autocorrelation function (CACF) of R_A and EP–NP, both for the time span 1744–2014. Significant cross-autocorrelations are observed, which suggest that a covariate model with EP–NP being the exogenous variable might be promising.

Throughout the rest of the paper, we denote the demeaned series of R_A by $\{X_t\}$, which is observed for the 271 years between 1744–2014, so $t = 1, 2, \dots, 271$. Similarly, the demeaned exogenous series EP–NP is denoted by $\{S_t\}$, $t = 1, 2, \dots, 271$. All of the models are developed for $\{X_t\}$, and the mean is added back later at the stage of prediction. The performances of the models are evaluated for both in-sample fitting and out-of-sample predictions.

3. Results

We provide detailed results and brief interpretations of the time-series models in this section.

3.1. ARMA Model

After conducting a thorough search using various statistical criteria, the appropriate ARMA model chosen for R_A is ARMA (1,2). Precisely, the fitted model is:

$$X_t = 0.8935X_{t-1} + Z_t - 0.5759Z_{t-1} - 0.1452Z_{t-2},$$

where $\{Z_t\}$ is a white-noise process with a fixed variance of σ^2 . Plots of the residuals and their ACF are shown in Figure 4. The autocorrelations are largely removed by the model, with the residuals being very close to white noise. In Figure 5, the model fitted values are plotted along with the 95% confidence interval, in comparison with the observed values. The overall mean squared error (MSE) of fitting is calculated to be 0.8303.

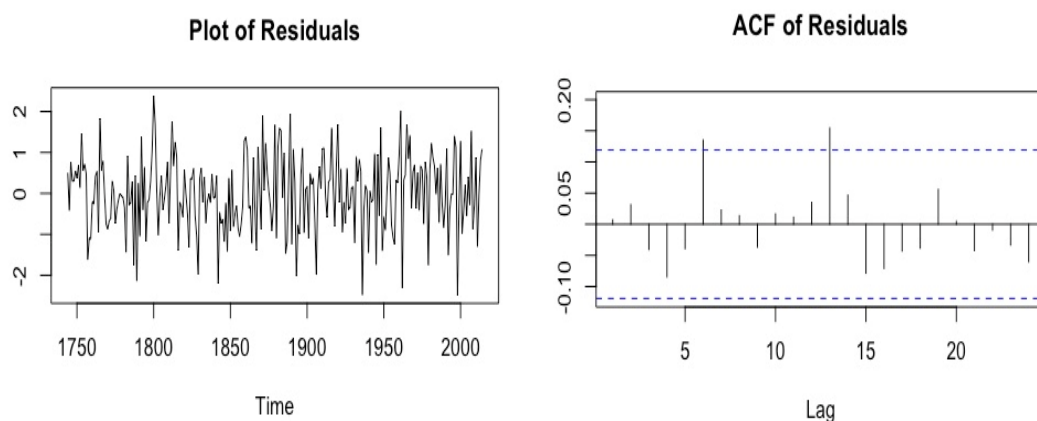


Figure 4. Plot (left), and ACF of the residuals (right), from the fitted autoregressive moving average (ARMA) (1,2) model.

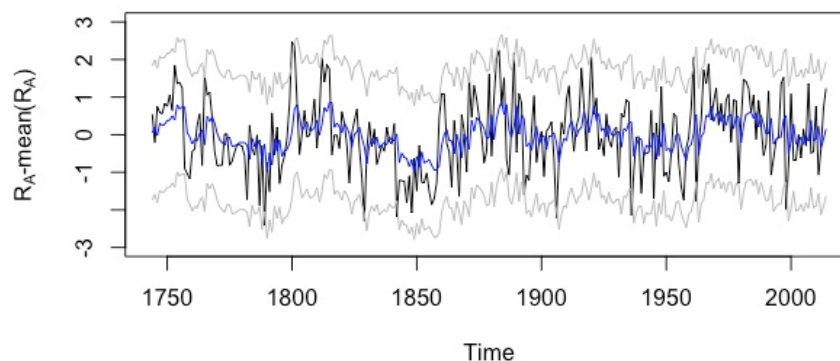


Figure 5. In-sample fitting of the ARMA (1,2) model to the demeaned series of R_A . This figure shows overlaid plots of the observed value (black), model fitted value (blue), and 95% confidence intervals of the fitted value (grey). The mean squared error (MSE) of fitting is 0.8303.

3.2. Long-Order Sparse AR (SpsAR) Model

The PACF of R_A shows some significant partial autocorrelations at large lags (e.g., lag 15), which indicates that a long-order sparse autoregressive (SpsAR) model is promising. It turns out that the appropriate model chosen is a SpsAR (15) model with significant lags at 1, 2, 6, 10, 13, and 15. The fitted model is:

$$X_t = 0.3107X_{t-1} + 0.1023X_{t-2} + 0.1781X_{t-6} + 0.0490X_{t-10} + 0.1405X_{t-13} - 0.1441X_{t-15} + Z_t,$$

with $\{Z_t\}$ being a white-noise process. By looking at the ACF of the residuals in Figure 6, the SpsAR model has improved fitting over ARMA, as all of the significant autocorrelations are completely removed. This is also reflected on the smaller MSE of fitting, which is computed to be 0.7715 (see Figure 7).

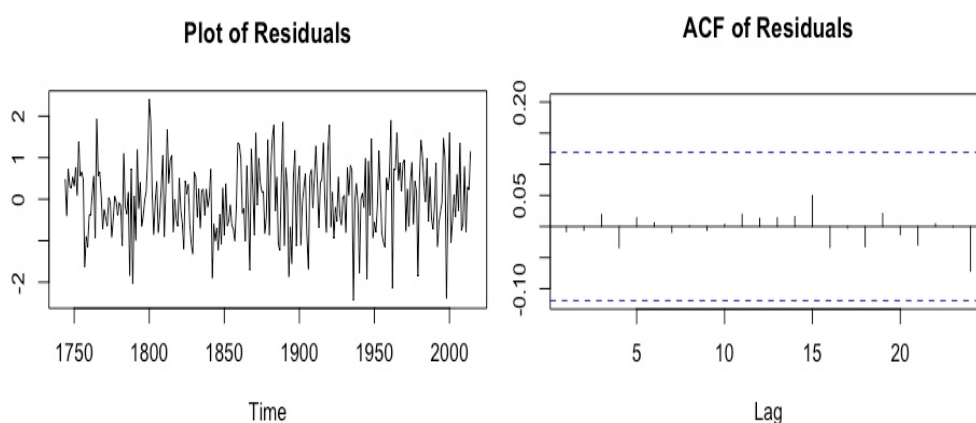


Figure 6. Plot (left) and ACF of the residuals (right) from the fitted SpsAR (15) model. The ACF gets improved from the previous ARMA model, with all of the autocorrelations removed.

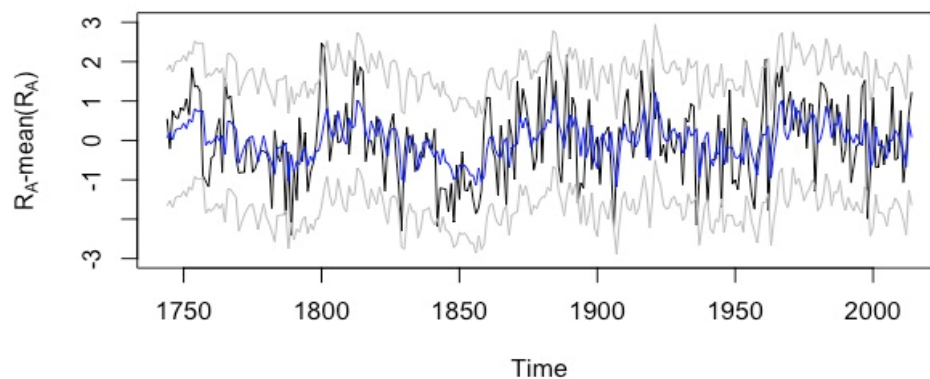


Figure 7. In-sample fitting of the long-order sparse autoregressive (SpsAR) (15) model to the demeaned series of R_A . The mean squared error (MSE) of fitting is 0.7715.

3.3. ARMAX Model

An ARMAX model for R_A with the covariate EP–NP is fitted to the data. The selected model is ARMAX (1,0,6), i.e., AR (1) with an AR (6) filtered covariate. The fitted model is:

$$X_t = 0.0479 * (S_t - 0.4216S_{t-1} - 0.4168S_{t-2} + 1.9732S_{t-3} - 0.3692S_{t-4} - 0.4354S_{t-5} + 0.9266S_{t-6}) + 0.4022X_{t-1} + Z_t,$$

where $\{Z_t\}$ is white noise. From Figure 8, the residuals are very close to white noise, except for the slightly significant autocorrelation at lag 13, which may require models with more complex structures

to be able to handle it. Notice that the MSE of fitting has improved from 0.8303 (ARMA without covariate) to 0.7948 by including the external variable EP–NP. The overall fitting is displayed in Figure 9.

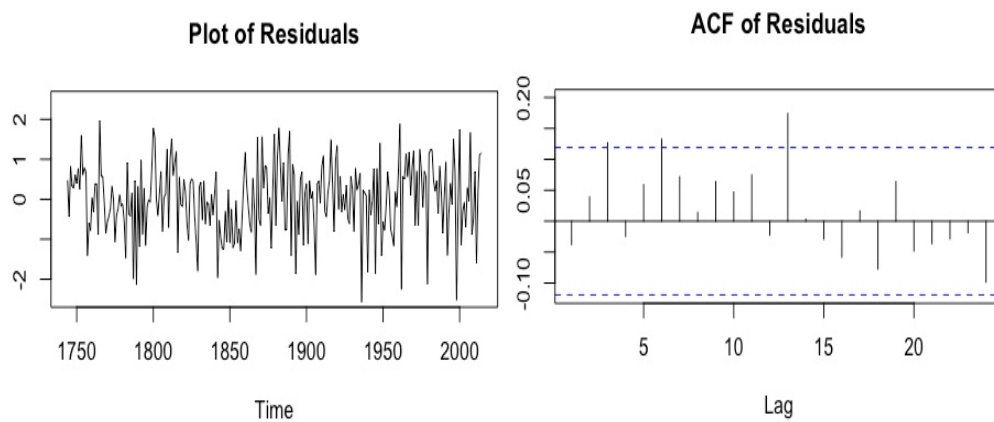


Figure 8. Plot (left) and ACF (right) of the residuals from the fitted ARMA with exogenous variables (ARMAX) (1,0,6) model.

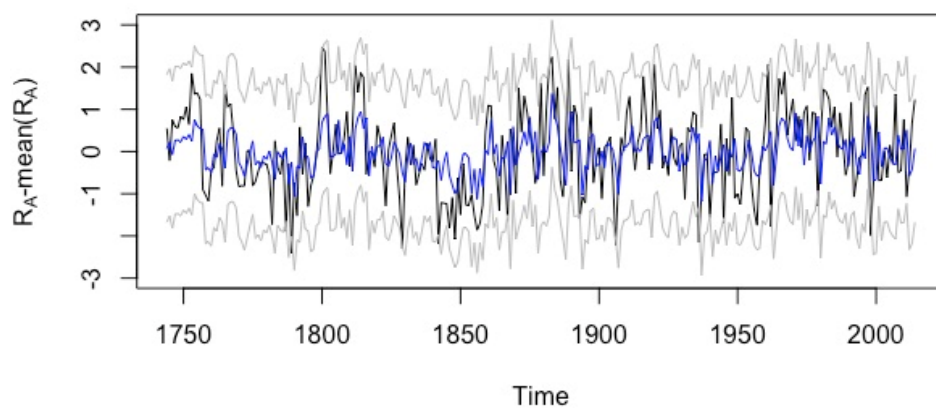


Figure 9. In-sample fitting of the ARMAX (1,0,6) model to the demeaned series of R_A (1744–2014). This figure shows overlaid plots of the observed value (black), model fitted value (blue), and 95% confidence intervals of the fitted value (grey). The mean squared error (MSE) of fitting is 0.7948.

3.4. Sparse ARX Model

A sparse AR model with covariate SST was fitted to the data, with the optimal AR order chosen to be 15. Same as in the ARMAX model, the linear filter chosen for EP–NP is AR (6). The estimated model is:

$$X_t = -0.0877 * (S_t + 0.4230S_{t-1} - 0.1949S_{t-2} - 0.1042S_{t-3} - 0.2607S_{t-4} + 0.3701S_{t-5} + 0.9892S_{t-6}) + 0.3033X_{t-1} + 0.1144X_{t-2} + 0.1977X_{t-6} + 0.0431X_{t-10} + 0.1539X_{t-13} - 0.1539X_{t-15} + Z_t,$$

where $\{Z_t\}$ is a white-noise process. As shown in Figure 10, the autocorrelations are completely removed by the model. The fitted values are plotted in comparison with the observed values in Figure 11. It is worth noting that, among all of the four models, the SpsARX model has achieved the optimal in-sample MSE of 0.7399.

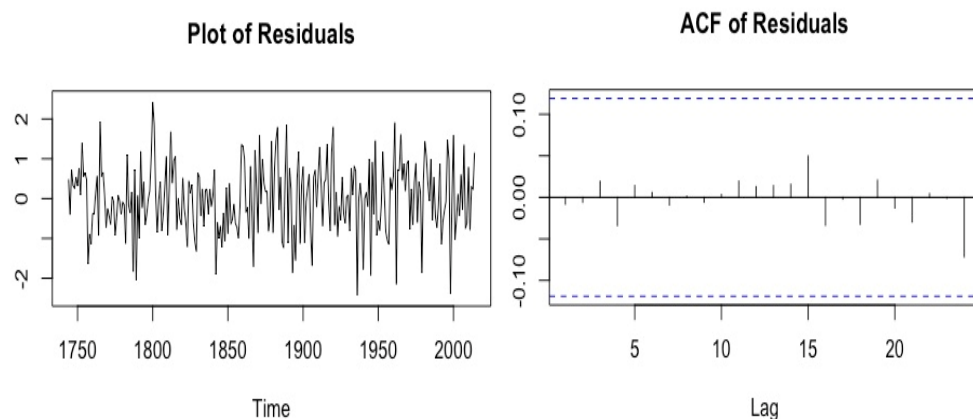


Figure 10. Plot (left) and ACF of the residuals (right) from the fitted SpsARX (15, 6) model.

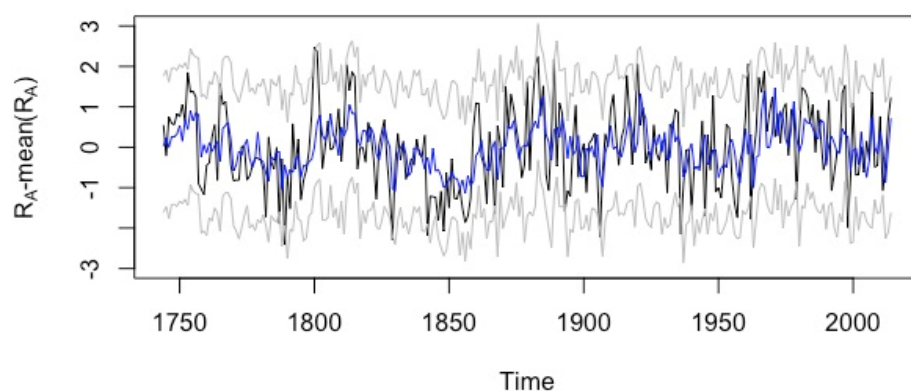


Figure 11. In-sample fitting of the SpsARX (15, 6) model to the demeaned series of R_A . This figure shows overlaid plots of the observed value (black), model fitted value (blue), and 95% confidence intervals of the fitted value (grey). The mean squared error (MSE) of fitting is 0.7399.

3.5. Forecasting Accuracy

A cross-validation study was carried out to compare the forecasting accuracy of the four selected models. We divided the entire data (1744–2014) into a 261-year training set (1744–2004) and a 10-year validation set (2005–2014). Four forecasting horizons $h = 1, 2, 5, 10$ are considered, and the model coefficients were updated every time we move forward. For example, for the 1-step-ahead prediction, we use the models that fitted with the data of the 1744–2004 period to predict the year 2005, and then the models that fitted with the data of 1744–2005 to predict the year 2006, and so on. Similarly, for the 2-step-ahead prediction, models that fitted with the data of 1744–2003 are used to predict the year 2005, and so on. For each of the three forecasting horizons, two error measures are used to evaluate the overall 10-year predictions: Root Mean Squared Error (RMSE) and Mean Standard Error (MSTE). The numerical results are reported in Table 1.

Table 1. Ten-year cross-validation results for the four models and the four forecasting horizons. The optimal results for each forecasting horizon are highlighted in bold.

	1-Step-Ahead		2-Step-Ahead		5-Step-Ahead		10-Step-Ahead		
	RMSE	MSTE	RMSE	MSTE	RMSE	MSTE	RMSE	MSTE	SS
ARMA (1,2)	0.87883	0.91138	0.83087	0.95757	0.80310	0.98140	0.78911	0.99966	0.00265
SpsAR (15)	0.77737	0.87834	0.81207	0.91978	0.77672	0.94083	0.78770	0.96606	0.00620
ARMAX (1,0,6)	0.98435	0.88835	0.99321	0.95844	0.94871	0.97373	0.94790	0.97407	−0.43912
SpsARX (15,6)	0.63692	0.86983	0.62585	0.90789	0.58614	0.93281	0.58306	0.96902	0.45550

It is interesting to note that, although the ARMAX model has improved fitting over ARMA, its predicting performance is very unstable, which results in the highest prediction RMSE among all the models. This suggests that the dynamics of R_A , while related to that of EP–NP, is largely driven by its own autocorrelation structure, and predictive models should be built as such. The SpsAR model improves over both the ARMAX and ARMA models, and the SpsARX model consistently and significantly outperforms all the other models for both evaluation criteria and for all forecasting horizons. Additionally, to verify decadal predictability, we compute the skill score (SS) [8] for the 10-year-ahead forecast using the sample mean of the training set as the reference forecast. It turns out that the two univariate models, ARMA and SpsAR, have very little decadal predictability—their SS's are almost zero. The ARMAX model has an even more negative SS, which, consistent with our preceding discussion, indicates again that this model has a highly unstable forecasting performance. The SpsARX model has comfortably achieved decadal predictability with an SS 0.455.

Furthermore, as a reference analysis, we repeated the above cross-validation study to a 50-year validation period. That is, we divided the data into a 221-year training set (1744–1964) and a 50-year validation set (1965–2014), and then calculated the RMSE and MSTE for the predictions over the 50 years. The results are summarized in Table 2. Generally, the errors for the 50-year validation are larger, most probably because of changes in the autocorrelation structures, as a significant portion of the data is taken for validation purposes. Nevertheless, the ranking among the four types of models is consistent with what is shown in Table 1. Especially, SpsARX continues to stand out as the best model, which enhances our finding that the tree-ring record does contribute to the rainfall predictability. Finally, to complete the analysis, Figure 12 shows the forecast values for the years 2015–2024 by the four models.

Table 2. Fifty-year cross-validation results for the four models and the four forecasting horizons. The optimal results for each forecasting horizon are highlighted in bold.

	1-Step-Ahead		2-Step-Ahead		5-Step-Ahead		10-Step-Ahead	
	RMSE	MSTE	RMSE	MSTE	RMSE	MSTE	RMSE	MSTE
ARMA (1,2)	0.94291	0.91012	0.97659	0.96582	0.97619	1.00022	0.98199	1.00556
SpsAR (15)	0.92504	0.87843	0.89998	0.91994	0.92794	0.94119	0.97126	0.96686
ARMAX (1,0,6)	0.98752	0.87852	1.03347	0.95508	1.02754	0.97091	1.03543	0.96946
SpsARX (15,6)	0.84880	0.87290	0.88449	0.91330	0.87881	0.94081	0.93695	0.97237

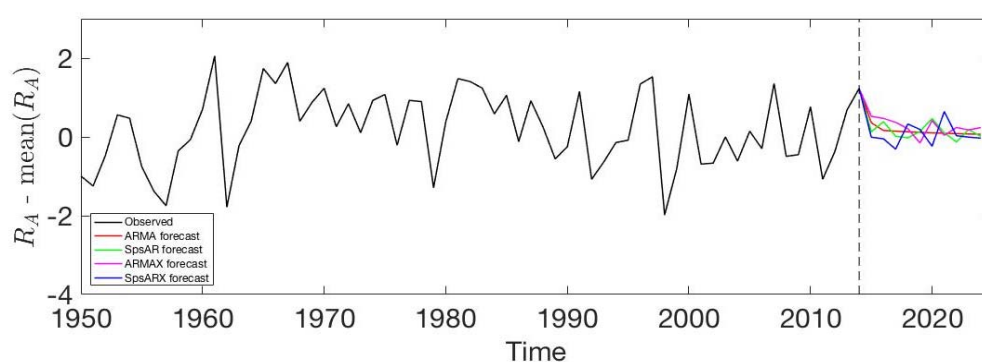


Figure 12. Plot of the 10-year forecasts (2015–2024) by the four models, together with the observed values (1950–2014).

4. Summary and Discussion

Statistical modeling of a tree-ring reconstructed ON rainfall index (R_A) for Central Vietnam was used to investigate its decadal predictability. Considering the autocorrelations of R_A and its cross-autocorrelations with EP–NP, four specific time-series models were developed. It was found that the prediction of R_A can be reasonably achieved by an ARMA model built solely on its own

autocorrelation structures, i.e., the in-sample 95% confidence interval captures almost all of the observations including the few extreme values (Figure 5). One intrinsic restriction of ARMA model is its short memory, which limits its capability to fully seize the relatively persistent autocorrelations of R_A as shown in Figure 2. To address this issue, we then adopted a long-order AR model with a sparse coefficient structure, and it showed a significant improvement over the ARMA model. In addition to forecasting R_A by itself, we further investigated the possibility to enhance the modeling by including the variable EP–NP, derived from SST reconstructions, given their significant cross-autocorrelations. It was shown that the ARMAX model indeed improved the in-sample performance over the ARMA, but its forecasting performance is relatively unstable. This result indicates that EP–NP is a valuable predictor for R_A , but the driving factor of R_A may largely be derived from its own autocorrelations (i.e., stochastic forcing that is common in precipitation) and the added data from tree rings did help capture the decadal variability. Therefore, we selected the final model based on the sparse AR model, with EP–NP used as an external variable. We found that this model can fully account for the autocorrelations of R_A while taking advantage of the predictability from EP–NP. These findings suggest that decadal predictability is achievable by using the optimal fitting with selected stable predictor(s).

Author Contributions: Y.S. and S.-Y.S.W. conceived and designed the experiments; Y.S. performed the experiments; Y.S. and S.-Y.S.W. analyzed the data; K.G.H. and B.M.B. contributed to the data; Y.S., S.-Y.S.W., B.M.B. and R.G. wrote the paper.

Acknowledgments: This research was supported by the USU Agriculture Experimental Station paper no. 9096 and Columbia University's Lamont contribution no. 8220.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Li, R.; Wang, S.Y.; Gillies, R.R.; Buckley, B.; Troung, L.H.; Cho, C. Decadal oscillation of autumn precipitation in Central Vietnam modulated by the East Pacific–North Pacific (EP–NP) teleconnection. *Environ. Res. Lett.* **2015**, *10*, 024008. [CrossRef]
2. Nguyen, D.Q.; Renwick, J.; McGregor, J. Variations of surface temperature and rainfall in Vietnam from 1971 to 2010. *Int. J. Climatol.* **2014**, *34*, 249–264. [CrossRef]
3. Wang, S.Y.S.; Promchote, P.; Truong, L.H.; Buckley, B.M.; Li, R.; Gillies, R.; Trung, N.T.Q.; Buan, B.; Minh, T.T. Changes in the autumn precipitation and tropical cyclone activity over Central Vietnam and its East Sea. *Vietnam J. Earth Sci.* **2015**, *36*, 489–496. [CrossRef]
4. Yokoi, S.; Matsumoto, J. Collaborative effects of cold surge and tropical depression–type disturbance on heavy rainfall in central Vietnam. *Mon. Weather. Rev.* **2008**, *136*, 3275–3287. [CrossRef]
5. Li, R.; Wang, S.Y.; Gillies, R.R.; Buckley, B.; Yoon, J.-H.; Cho, C. Regional trends in early-monsoon rainfall over Vietnam and CCSM4 attribution. *Clim. Dyn.* **2018**, in press. [CrossRef]
6. Buckley, B.M.; Stahle, D.K.; Luu, H.T. Central Vietnam climate over the past five centuries from cypress tree rings. *Clim. Dyn.* **2017**, *48*, 3707–3723. [CrossRef]
7. Hansen, K.G.; Buckley, B.M.; Zottoli, B.; D'Arrigo, R.D.; Nam, L.C.; Truong, V.V.; Nguyen, D.T.; Nguyen, H.X. Discrete seasonal hydroclimate reconstructions over northern Vietnam for the past three and a half centuries. *Clim. Chang.* **2017**, *145*, 177–188. [CrossRef]
8. Wilks, D. *Statistical Methods in the Atmospheric Sciences*, 3rd ed.; Academic Press: Oxford, UK, 2011; pp. 2–676. ISBN 9780123850225.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).