Updated report of trilateral collaborative study among Japan, Korea and Taiwan for producing joint abundance indices for the yellowfin tunas in the Indian Ocean using longline fisheries data up to 2020

Toshihide Kitakado[1], Sheng-Ping Wang[2], Keisuke Satoh[3], Sung Il Lee[4], Wen-Pei Tsai[5], Takayuki Matsumoto[6], Hiroki Yokoi[3], Kei Okamoto[6], Mi Kyung Lee[4], Jung-Hyun Lim[4], Youjung Kwon[4], Nan-Jay Su[2], Su-Ting Chang[7] and Feng-Chen Chang[7]

[1] Tokyo University of Marine Science and Technology, 5-7, Konan 4, Minato, Tokyo, 108-8477, Japan (Email: kitakado@kaiyodai.ac.jp)

[2] National Taiwan Ocean University, No. 2, Beining Rd., Zhongzheng Dist., Keelung City 20224, Taiwan

[3] National Research and Development Agency, Japan Fisheries Research and Education Agency, Fisheries Resources Institute, 2-12-4 Fukuura, Kanazawa-ku, Yokohama-shi, kanagawa-ken, 236-8648, JAPAN

[4] National Institute of Fisheries Science, 216 Gijanghaean-ro, Gijang-eup, Gijang-gun, 46083 Busan, Korea

[5] National Kaohsiung University of Science and Technology, No. 415, Jiangong Rd., Sanmin Dist., Kaohsiung City 80778, Taiwan

[6] National Research and Development Agency, Japan Fisheries Research and Education Agency, Fisheries Resources Institute, 5-7-1, Orido, Shimizu, Shizuoka-shi, 424-8633, Japan

[7] Overseas Fisheries Development Council, 3F., No.14, Wenzhou St., Da'an Dist., Taipei City 10648, Taiwan

ABSTRACT

Three distant-water tuna longline countries, Japan, Korea and Taiwan, have started a collaborative study since December 2019 for producing the joint abundance indices using integrated fishery data of these fleets to contribute to the upcoming stock assessments of yellowfin tuna in the Indian Ocean. The intention is to produce reliable indices by increasing the spatial and temporal coverage of fishery data. In this paper, results using data up to 2020 fisheries were provided to update the WPTT on the progress of this activity. As an underlying analysis, a clustering approach was utilized to account for the inter-annual changes of the target in each fishery in each region. For this purpose, a hierarchical clustering method with "fastcluster" was used, and the outputs of the finalized cluster were then used to assign the cluster label on fishery target to each catch-effort data. For standardizing the catch-per-unit-effort data, the conventional linear models and delta-lognormal linear models were employed for data of monthly and 1° grid resolution in each region. In addition to the implicit target species through the clustering, geographical and temporal covariates were used in the regression structures. The models were diagnosed by the standard residual plots and influence analysis.

INTRODUCTION

Tuna-RFMOs, including the IOTC, recommended that the joint CPUE of longline fisheries be developed to improve the stock assessments for tropical tunas, and thus the IOTC has conducted collaborative works for several years to produce an abundance index by combining CPUEs data from major longline fleets. An ensemble approach of fishery data from multiple longline fleets has been applied to the tropical and temperate tuna species for their stock assessments (e.g. Hoyle et al. 2018, Hoyle et al. 2019a, 2019b).

However, it was found that the fishing technologies, data formats, spatial-temporal coverage were different among the fleets, and therefore it is important to discuss and exchange the information among countries using ample time in order to improve the analysis and index. To this end, three longline countries, Japan, Korea and Taiwan, have been conducting a collaborative study for developing the abundance index since December 2019. Some preliminary results using data up to 2019 fisheries have been reported at WPTT-23(DP) to update the WPTT on the progress of this activity (e.g. Kitakado et al. 2021).

In WPTT-23(DP), several suggestions were made:

- Data period for Taiwanese data: due to data quality problems, we used data from 2005 onwards;
- Vessel screening: in our earlier analysis, vessels which have more than or equal to 20 data set (one data set means 5-degree grid by month by year by cluster) were used for standardization of CPUE, but this time top 50% vessels of the number of efforts were also used as requested by the WPTT(DP);
- Clustering method: in our earlier analysis, a two-step approach with K-means and hierarchical clustering was used to sort some dimension problems, but this time "fastcluster" (http://danifold.next/fastcluster.html) was used according to the guidance;
- Target information: in addition to the clustering approach, data of HBF was used for the target index;
- Target in tropical area: another request was to conduct clustering with BET&YFT as a single species, but based on the conversations among member countries, it does not reflect to the YFT target, so finally we used the these as separate species.

In this paper, results using data up to 2020 fisheries were provided. Since some detailed information on distributions of fishing efforts etc. has been shown in our previous report (Kitakado et al. 2021), main outcomes were reported here.

MATERIALS

Data sharing protocol

Initially, the analysis was planned to conduct in a series of in-person meetings through data sharing in an intranet system to ensure the data security. However, after a face-to-face meeting in Busan in December 2019, we have been holding only webinar meetings because of COVID-19 pandemic. Under this circumstance, a data sharing protocol was finalized among the three countries with a restriction of data access only by the Chair of the group (Toshihide Kitakado) for reduced resolution of data set (not operational data but some aggregated data over 1° square grid by month by vessel). The data set combined for yellowfin CPUE standardization were available from 1975 to 2020, with data fields of year and month of operation, location to 1° of latitude and longitude, vessel id, number of hooks, and catch by species in number. We classified the species into albacore (ALB), bigeye (BET), yellowfin (YFT), southern bluefin tuna (SBT), black marlin (BLM), blue marlin (BUM), swordfish (SWO), other billfishes (BIL), sharks (SKX) and others (OTH).

METHODS

Analytical procedures

For standardizing the catch-per-unit-effort data, the conventional linear models and delta-lognormal linear models were employed for data of monthly and 1° grid resolution in each region. In addition to the implicit target species through the clustering, geographical and temporal covariates were used in the regression structures. The models were diagnosed by the standard residual plots and influence analysis and compared via the model selection criteria. Besides these conventional regression methods, analyses using an advanced spatio-temporal model, VAST, were attempted for developing abundance indices with additional consideration of spatio-temporal correlations and targets as well as the life stage of yellowfin tuna. So, in a nutshell, the approaches are as follows:

1) investigation of better approaches to account for changes in targeting within each country;

2) analyses using conventional regression models with geographical, environmental and fishery (including targeting) information for continuity from the previous approaches.

Cluster analysis

In our previous version of analysis, a two-step approach with K-means and hierarchical clustering was used to sort some dimension problems. In WPTT-23(DP), it was advised that a non-two-step approach "fastcluster" be used, and therefore the analysis was conducted to follow such a guidance. The data were aggregated by 10-days duration (1st-10th, 11th-20th, and 21st~ for each month) based on the agreement of the trilateral collaborative working group. The number of clusters was determined when the relative improvement of SS within-clusters was less than 10%. See some details shown in Wang et al. (2021).

Regression analyses

Log-normal (LN) regression models with a constant adjustment

Given that around 8.4% of the catch data are 0, we used an adjustment factor (here 10% of mean of CPUE) to the CPUE data to employ conventional log-normal distributions as follows:

$$\log(CPUE + c) = Main\,effects + Interactions + Error$$

Potential covariates used in the analysis were shown below:

- Temporal component (year, quarter, year*quarter)
- Spatial component (5° squared longitudinal and latitudinal grid)
- Vessel ID
- Cluster category
- HBF averaged within aggregated data: Shallow (<=7), Medium (8<=HBF<=13) and Deep (14 <= HBF)

The error terms are assumed to be independently and identically distributed as the normal distribution with mean 0 and standard deviation σ. The constant adjustment factor, *c*, is 10% of the overall mean as has been used in previous analyses.

Delta-lognormal (DL) regression model

For R3, a delta-lognormal model was also tested to account for “zero data” statistically as has been used in previous analyses (see e.g. Hoyle et al. 2018). For the first component of “zero” or “non-zero” is expressed as a binomial distribution with a probability of “non-zero” catch as a logistic relationship with some explanatory variables, and the second component for positive catch assumed the same regression structures used in the LN regression models with a constant adjustment. The number of hooks was also used in the first component of analysis.

Diagnosis and impacts of covariates (Residual plots, Q-Q plots, influence plots)

The standard residual plots were for the diagnosis for fitting of models to the data and Q-Q plots (only for the positive catch component in DL models). In addition, we used influence plots (Bentley et al. 2011) to interpret the contribution of each covariate to the difference between nominal and standardized temporal effects.

Extracts of abundance indices from models with interactions

Once the model fitting and model evaluation were conducted, the final output of the abundance index is extracted through an exercise of the least square means (so-called LS means) to account for heterogeneity of amount of data over covariate categories (as well as the standardized probability of "non-zero" catches in DL models).

RESULTS

Cluster analysis

The results of cluster analysis were shown in Figure 1. In each country’s analysis, the number of clusters ranged from 3 to 5, and among the clusters, there were some clear patterns in cluster groups in which the target was clearly yellowfin and some other patterns of mixed targets of yellowfin and bigeye tunas. However, as shown in Figure 2 as an example of Taiwanese analysis, when mixing yellowfin and bigeye into a single species, the cluster sign and geographical pattern was somewhat lost. Therefore, cluster results with the separate two species were used.

Conventional regression analysis

Full evaluation of models thought the model selection criterion has not yet reached, but comparison of results over the following models are shown in Figure 3. Also, the diagnostics and influence plots were shown in Figures 4-7.

New_clu_use0.5: YrQ + LonLat + Cluster + Vessel [top 50% vessels in the number of efforts]

New_clu_use20: YrQ + LonLat + Cluster + Vessel [vessels having more than or equal to 20 data set]

New_hbf_use0.5: YrQ + LonLat + HBF + Vessel [top 50% vessels in the number of efforts]

New_hbf_use20: YrQ + LonLat + HBF + Vessel [vessels having more than or equal to 20 data set]

General and specific observations are given below:

- Decreasing patterns were similar over the four different models in each region. Data screening methods (use0.5 and use20) did not influence the results.
- In tropical regions, R1 and R4, where two different target indices were matter of concern, general pattens of the standardized CPUEs including the recent trend were similar between the two analyses while the relative scale over time was slightly different. Some missing information in HBF cannot compare the models formally.
- With some difference, decreasing patterns were also similar to the previous CPUEs used in 2018 assessment.
- For region R3, there were much less catch there and quite less YFT targeted vessels, and our data were aggregated one, so information might have been a bit poor, and we have not reached very much reasonable results in the earlier period of years. So, for the assessment specification, a down-weighted option relative to other areas might be worth considering. We have also tried delta-LN analyses as
  DL[for 0/1: YrQ + LonLat + Cluster + LnHooks, for positive catch: YrQ + LonLat + Cluster + Vessel],
  but this did not solve the problems reasonably (see Figure 8) although the residua pattern was improved. If R3 results shown in this paper are regarded as usable, another way is to use previous index developed by operational data although the recent 3 years' information is lost.

Acknowledgement

The authors thank the WPTT-23(DP) participants for their useful comments on our earlier version of paper. The authors also thank Simon Hoyle and Dan Fu for their helpful and specific comments for improving the analysis.

REFERENCES

Bentley, N., Kendrick, T.H., Starr. P.J. and Breen P.A. 2012. Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations. ICES J. Mar. Sci. 69(1): 84-88.

Hoyle, S.D., Chang, S.T., Fu. D., Kim, D.N., Lee, S.I., Matsumoto, T., Chassot, E. and Yeh, Y.M. (2019a) Collaborative study of bigeye and yellowfin tuna CPUE from multiple Indian Ocean longline fleets in 2019, with consideration of discarding. IOTC-2019-WPM10-16.

Hoyle, S.D., Chassot, E., Fu, D., Kim, D.N., Lee, S.I., Matsumoto, T., Satoh, K., Wang, S.P. and Kitakado, T. (2019b) Collaborative study of albacore tuna CPUE from multiple Indian Ocean longline fleets in 2019. IOTC-2019-WPTmT07(DP)-19.

Hoyle, S.D., Kitakado, T., Yeh, Y.M., Wang, S.P., Wu, R.F., Chang, F.C., Matsumoto, T., Satoh, K., Kim, D.N., Lee, S.I., Chassot, E., and Fu, D. (2018) Report of the Fifth IOTC CPUE Workshop on Longline Fisheries, May 28th–June 1st, 2018. IOTC–2018–CPUEWS05–R. 27 pp.

Hoyle, S.D., Langley, A.D. 2020 Scaling factors for multi-region stock assessments, with an application to Indian Ocean tropical tuans. Fish. Res. 228 https://doi.org/10.1016/j.fishres.2020.105586

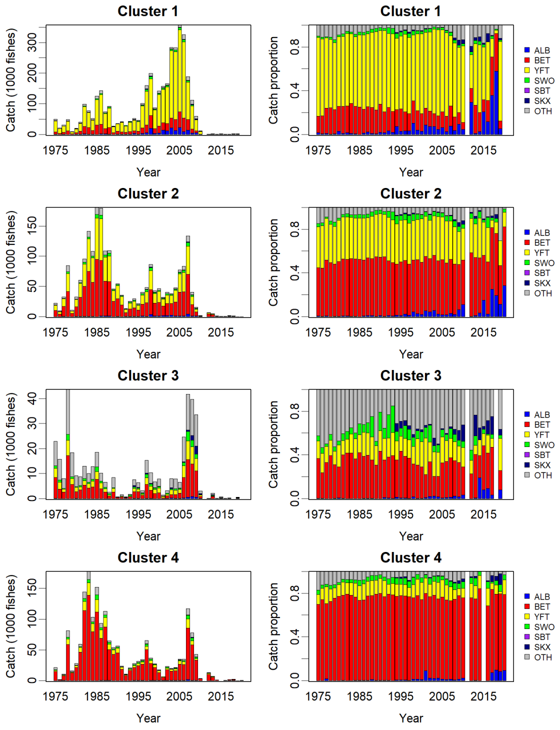
Kitakado, T., Satoh, K., Matsumoto, T., Yokoi, H., Okamoto, K., Lee, S., Lee, M., Lim, J.H., Wang, S.H., Su, N., Tsai, W., Chang, S.T. 2020. Plan of trilateral collaborative study among Japan, Korea and Taiwan for producing joint abundance index with longline fisheries data for the tropical tuna species in the Indian Ocean. Collect. IOTC-2020-WPTT22(SA)-09.

Kitakado, T., Wang, S.H., Satoh, K., Lee, S.I., Tsai, W.P., Matsumoto, T., Yokoi, H., Okamoto, K., Lee, M.K., Lim, J.H., Kwon, Y., Su, N.J. and Chang, S.T. Report of trilateral collaborative study among Japan, Korea and Taiwan for producing joint abundance indices for the yellowfin tunas in the Indian Ocean using longline fisheries data up to 2020. IOTC-2021-WPTT23(DP)-14.

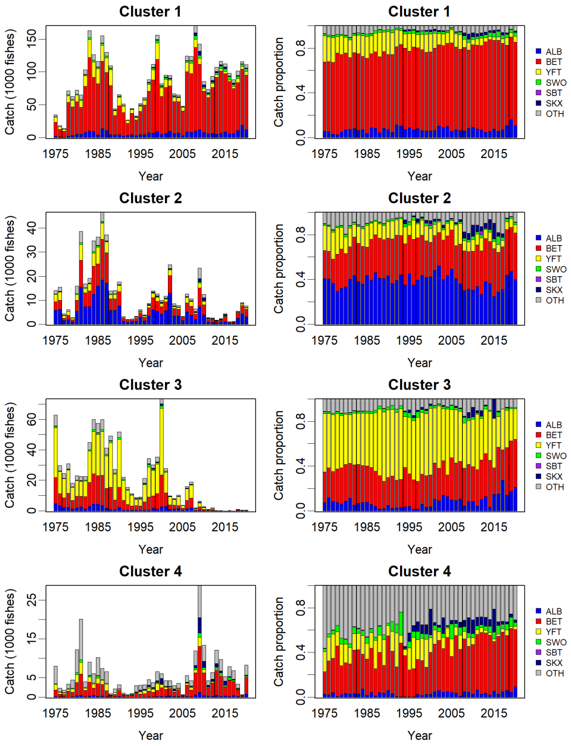
Wang, S.H., Xu, W.Q., Lin, C.Y. and Kitakado, T. (2021) Analysis on fishing strategy for target species for Taiwanese large-scale longline fishery in the Indian Ocean. IOTC-2021-WPB19-11.

(a) Japan

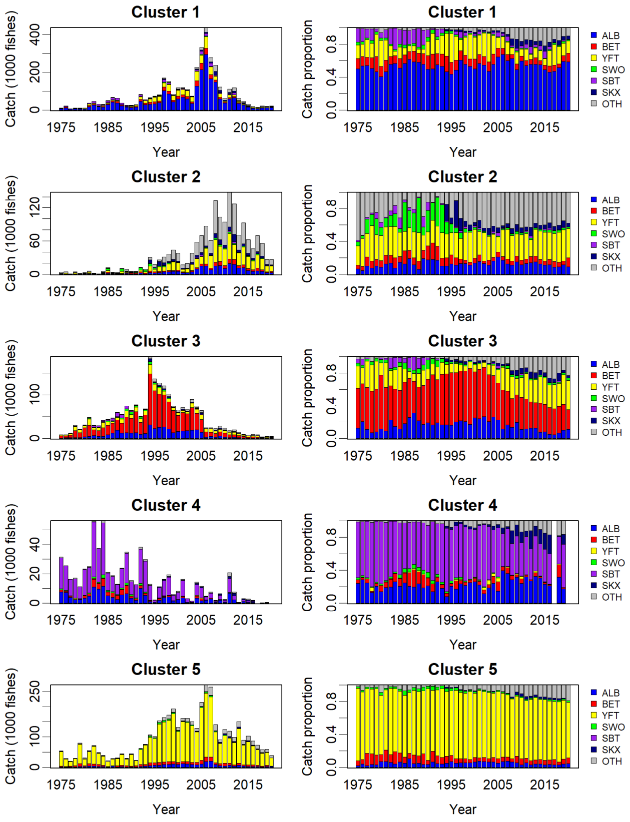
R1



R4



R2



R3

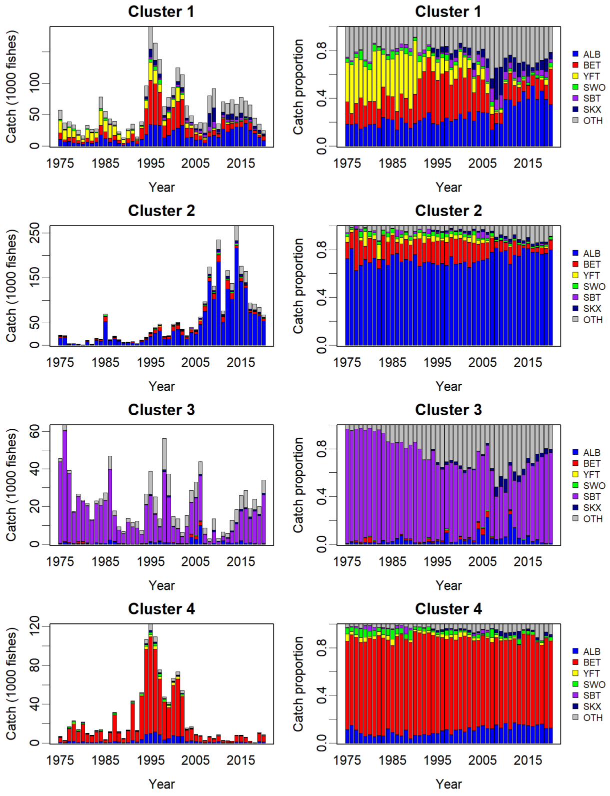
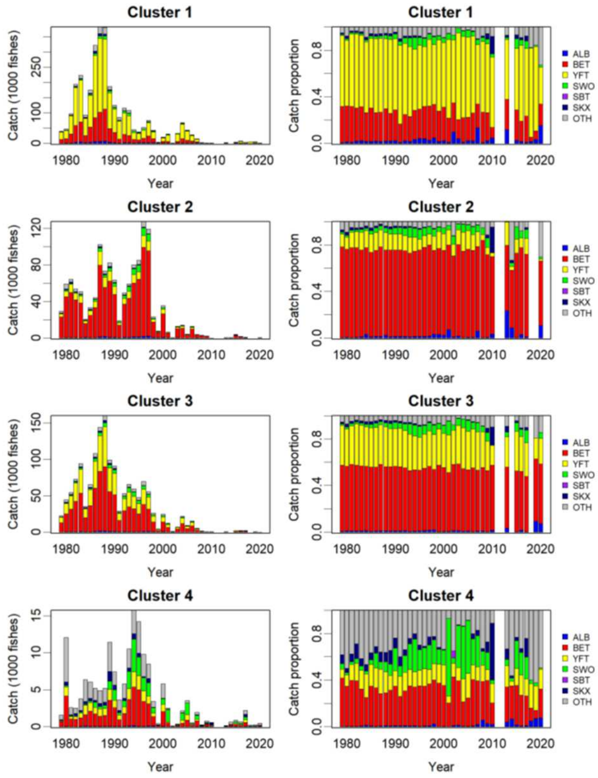


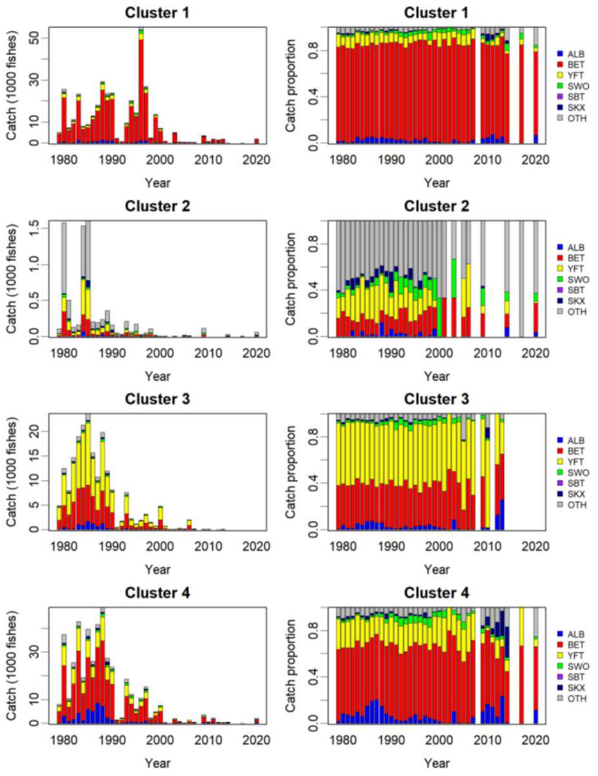
Figure 1(a): Species composition for each cluster in Japanese fisheries.

(b) Korea

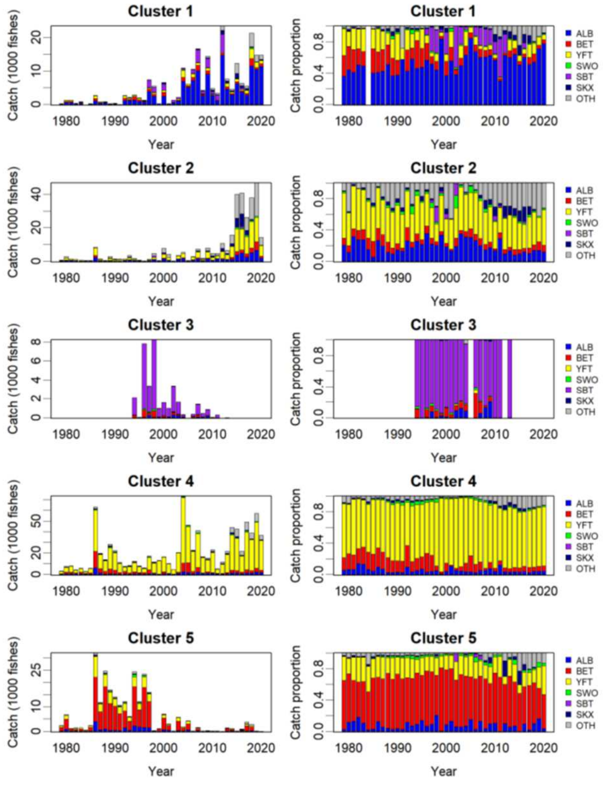
R1



R4



R2



R3

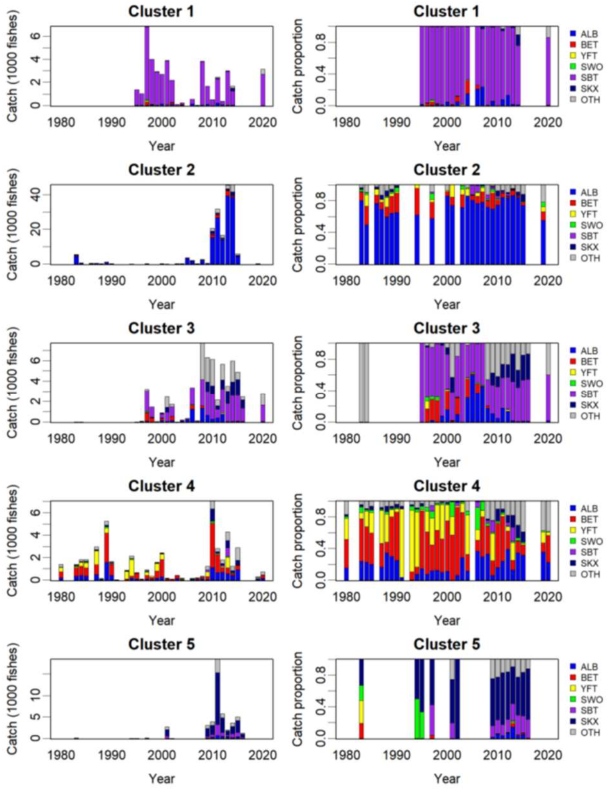
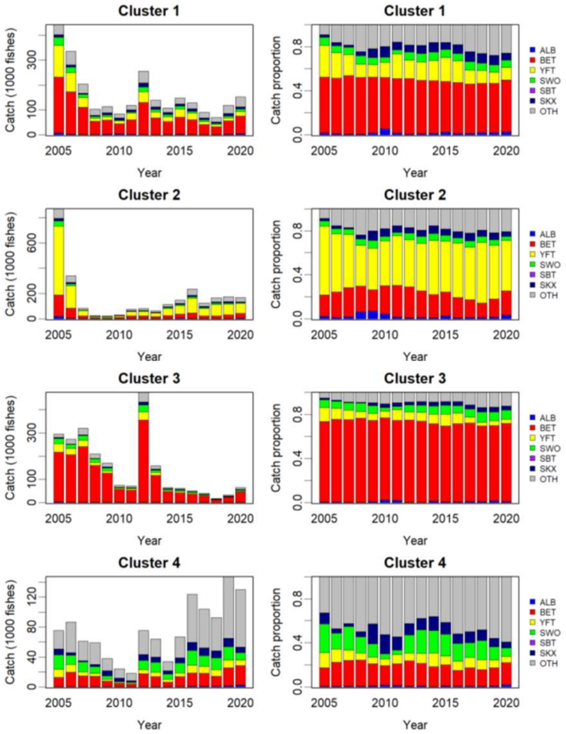


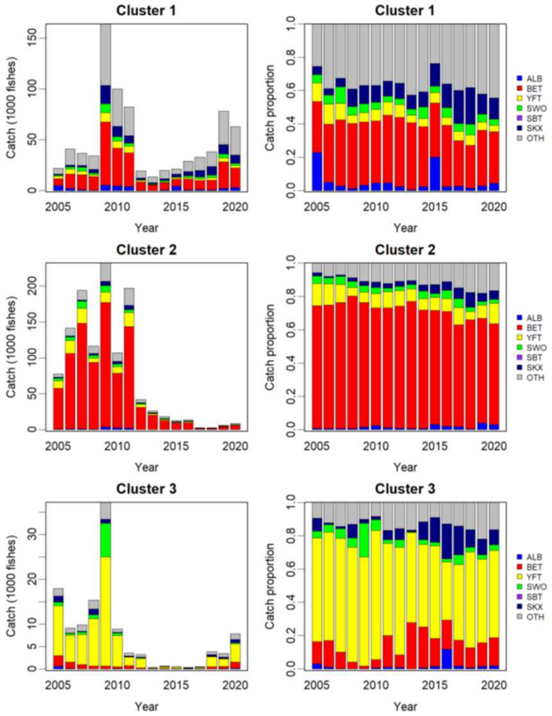
Figure 1(b): Species composition for each cluster in Korean fisheries.

(c) Taiwan

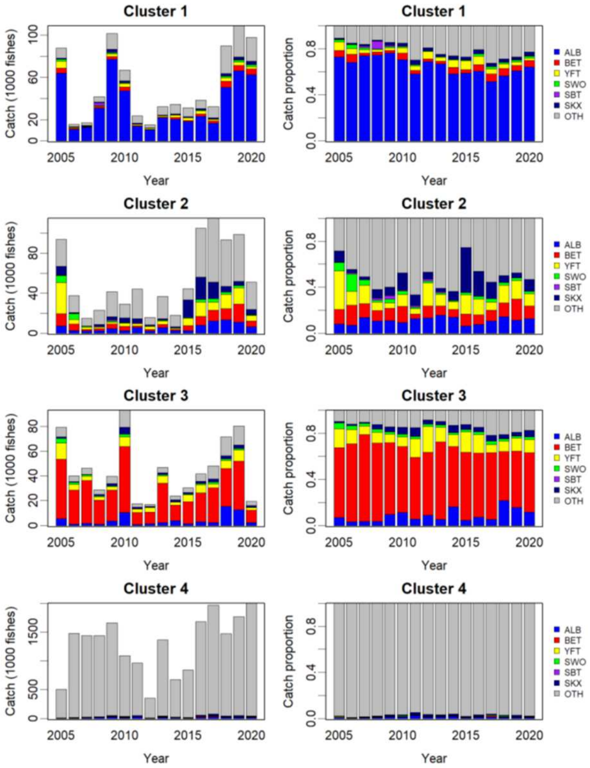
R1



R4



R2



R3

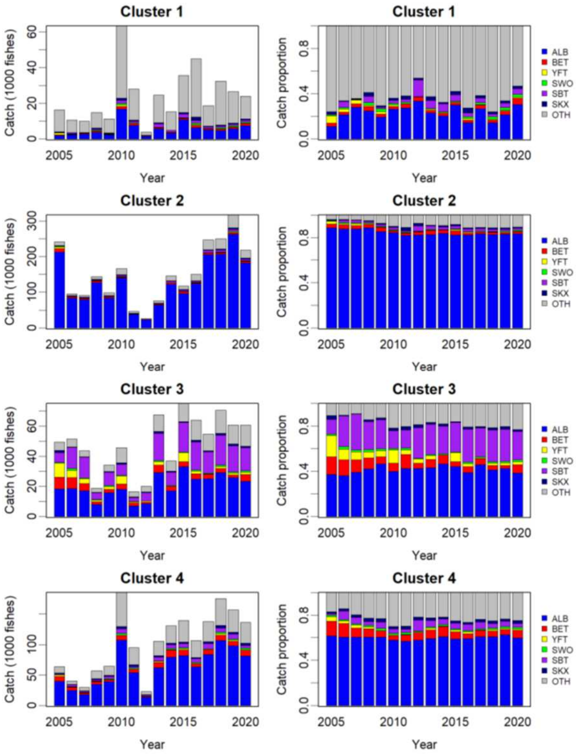
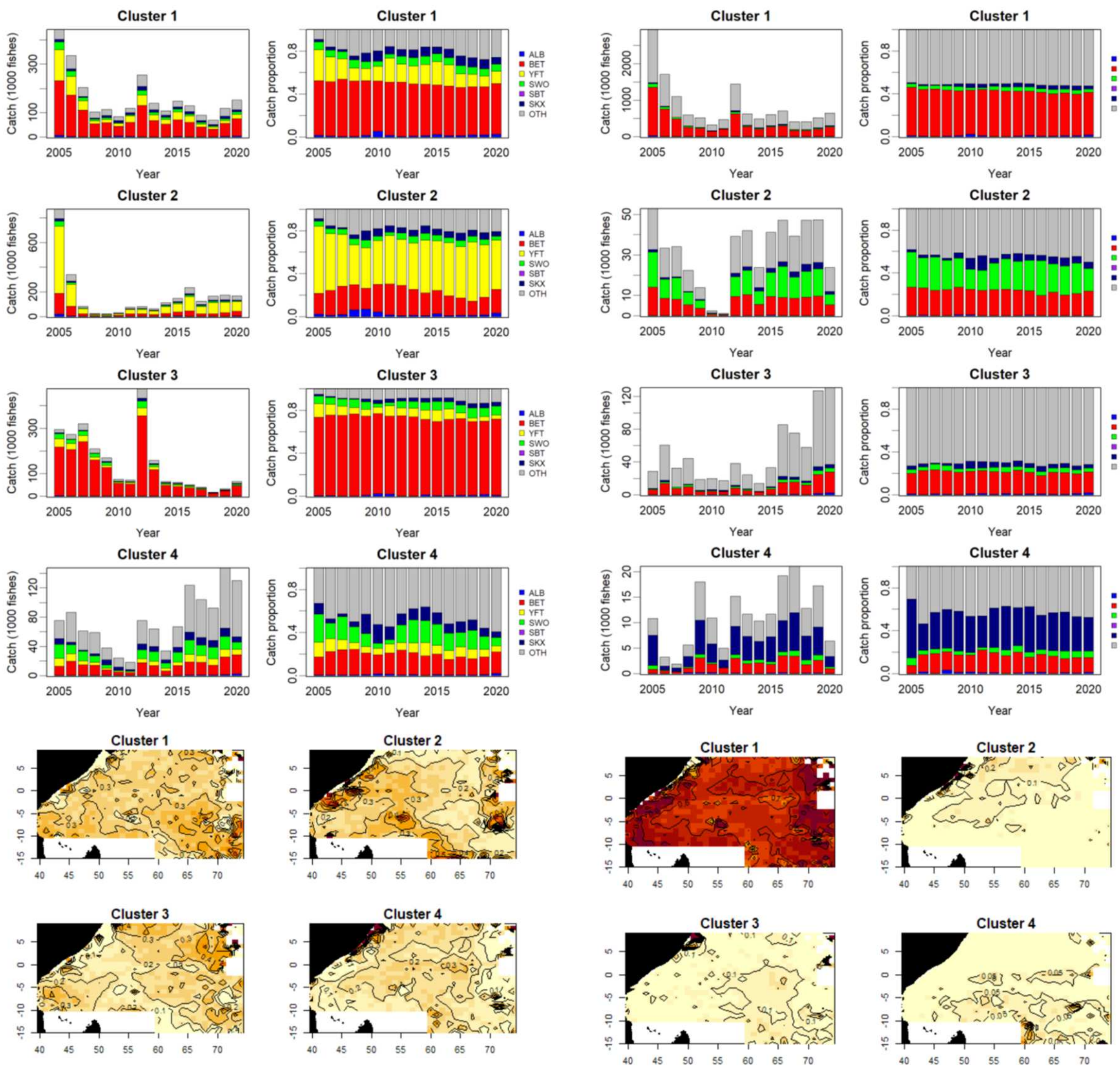


Figure 1(c): Species composition for each cluster in Taiwanese fisheries.

(REGION1)



BET/YFT separated

BET/YFT combined

Figure 2(a): Comparison of results of clustering with “BET/YFT separated” and “BET/YFT combined” for Taiwanese fishery data in R1.

(REGION4)

BET/YFT separated **BET/YFT combined**

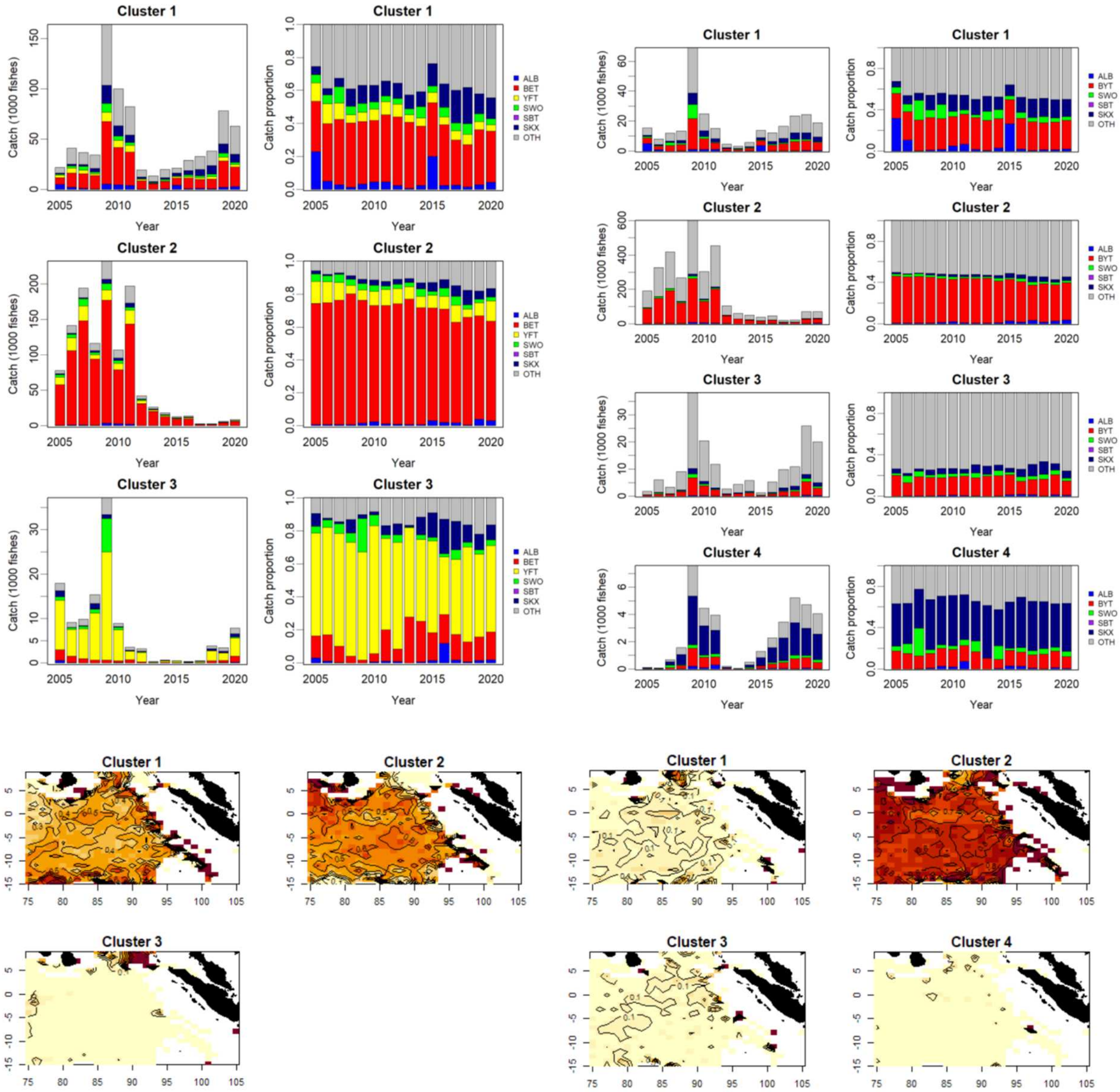
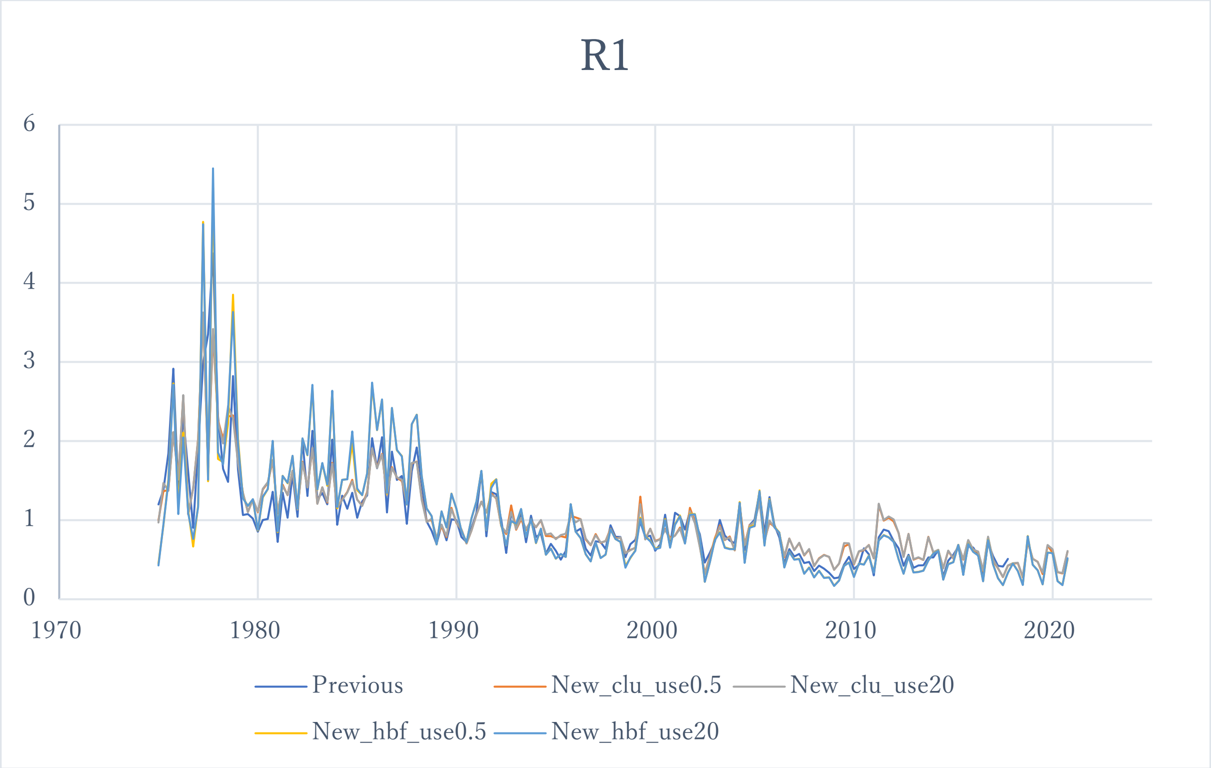


Figure 2(b): Comparison of results of clustering with “BET/YFT separated” and “BET/YFT combined” for Taiwanese fishery data in R4.



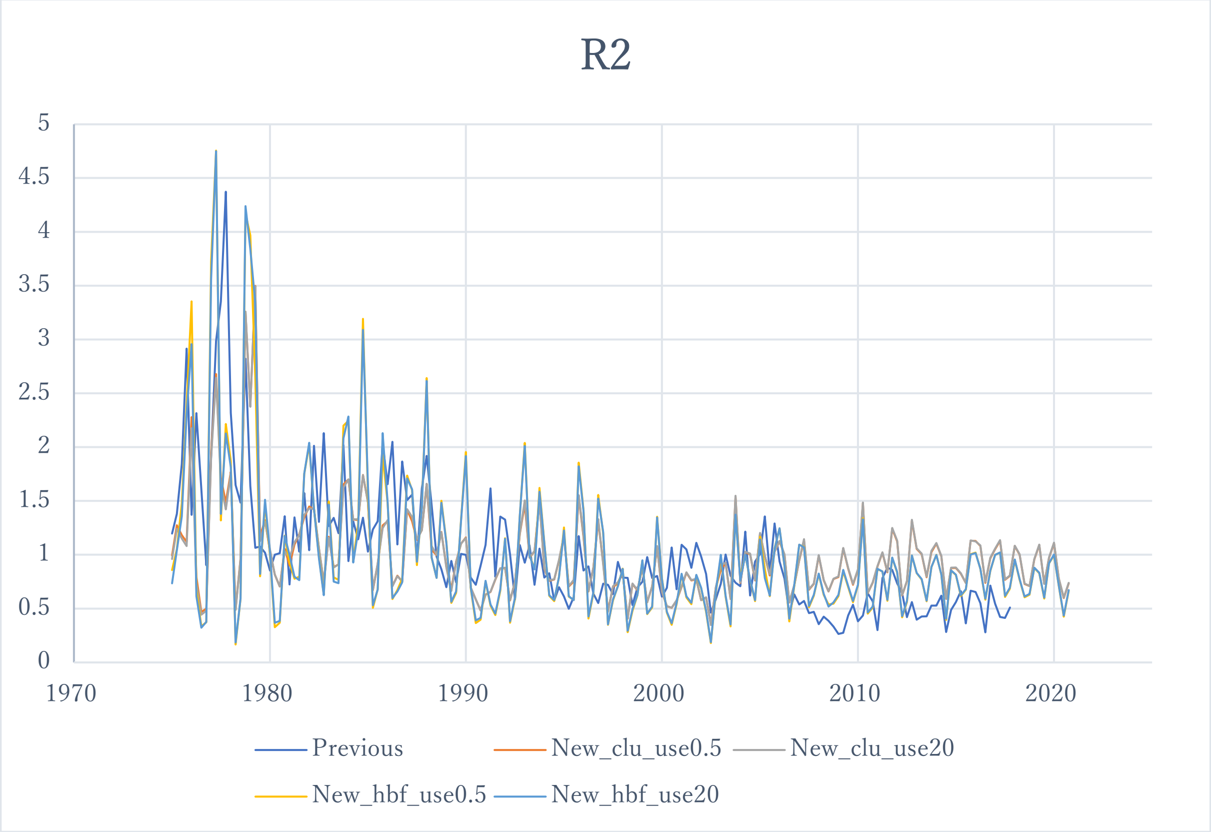
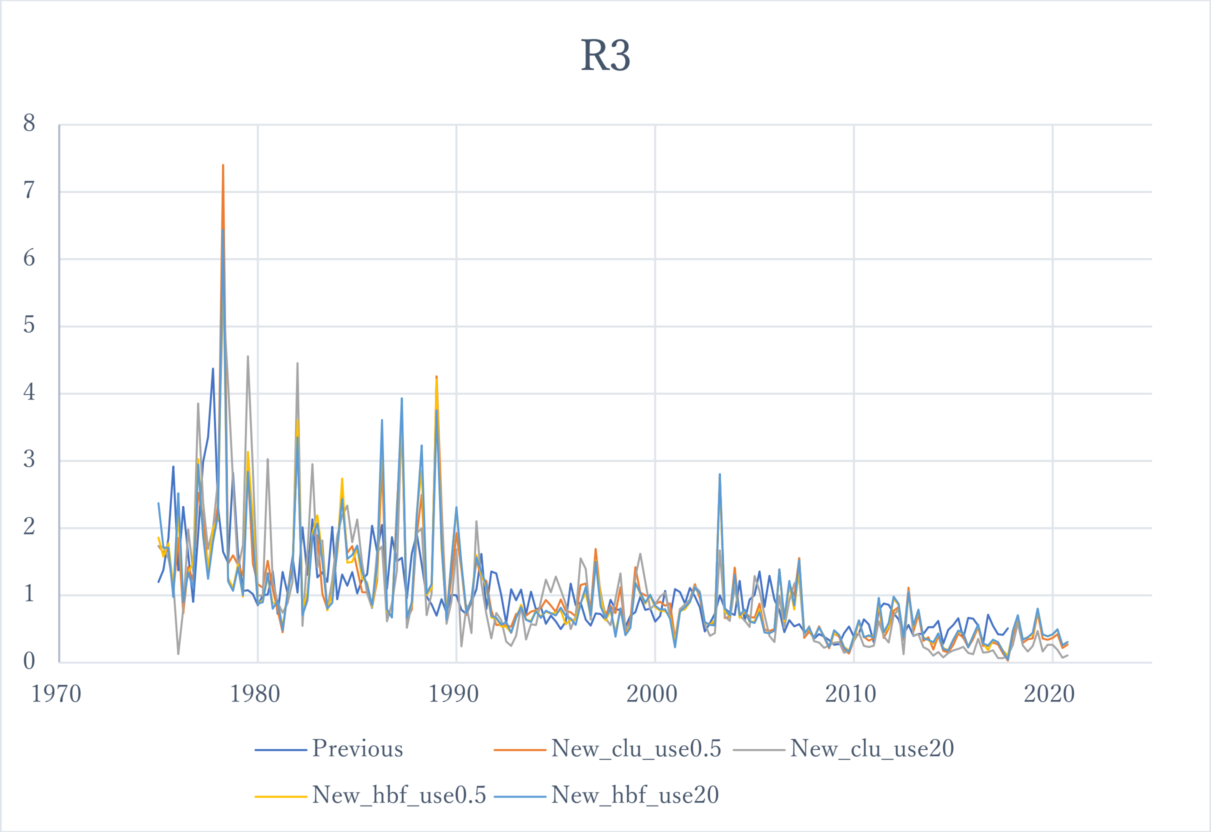


Figure 3(a): Comparison of standardized CPUEs in R1 and R2. “Previous” means the CPUE indices used in the 2018 assessment.



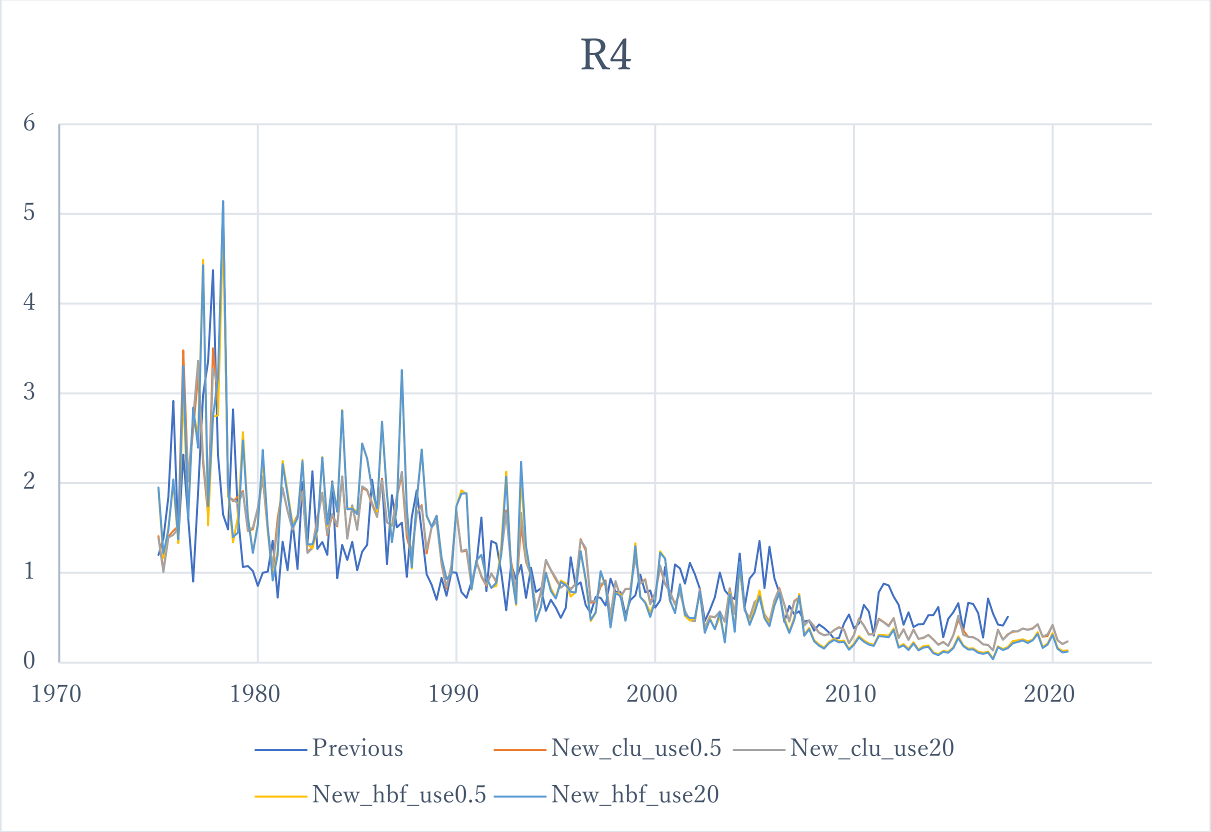


Figure 3(b): Comparison of standardized CPUEs in R1 and R2.

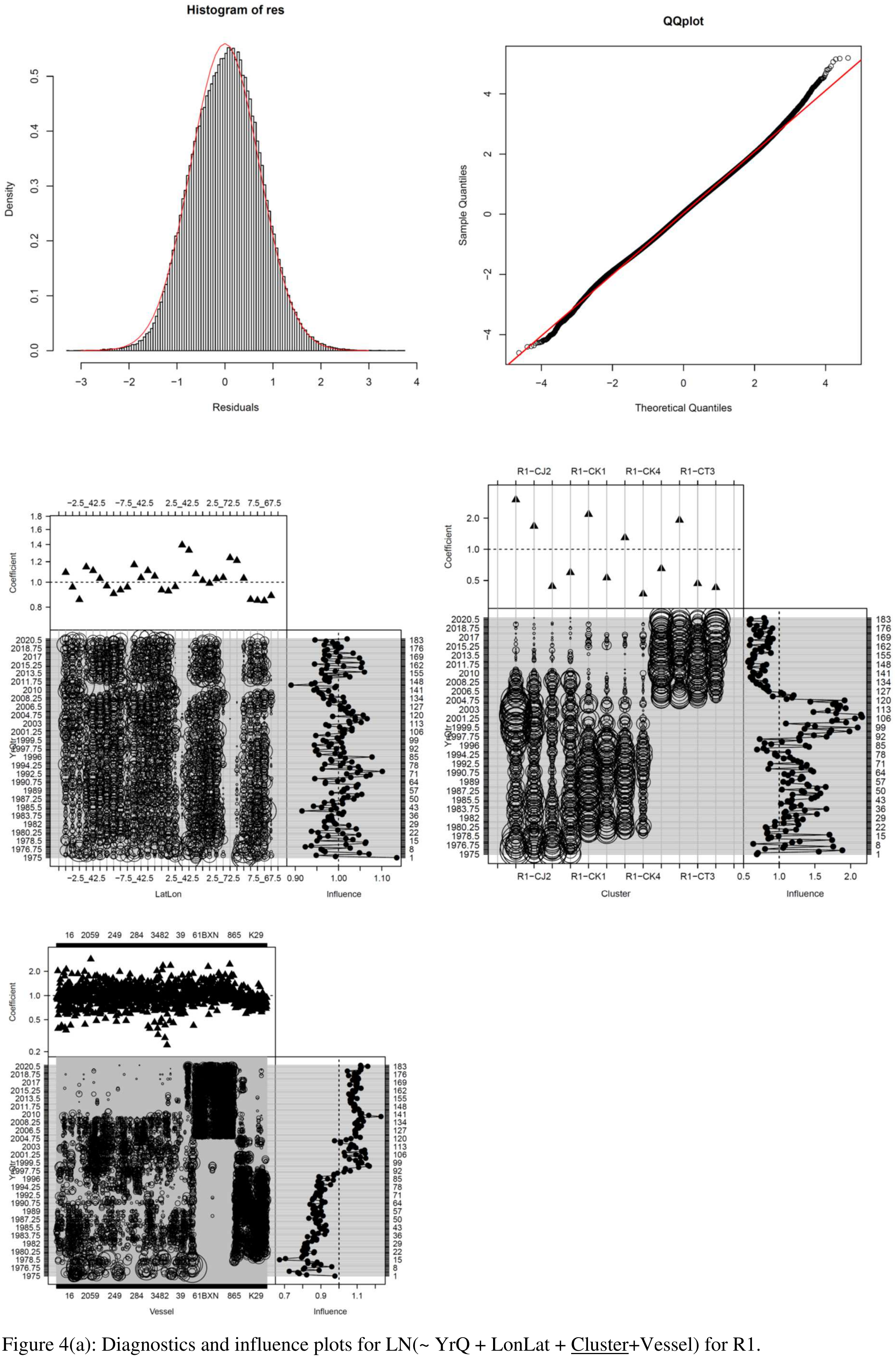


Figure 4(a): Diagnostics and influence plots for LN(~ YrQ + LonLat + <u>Cluster</u>+Vessel) for R1.

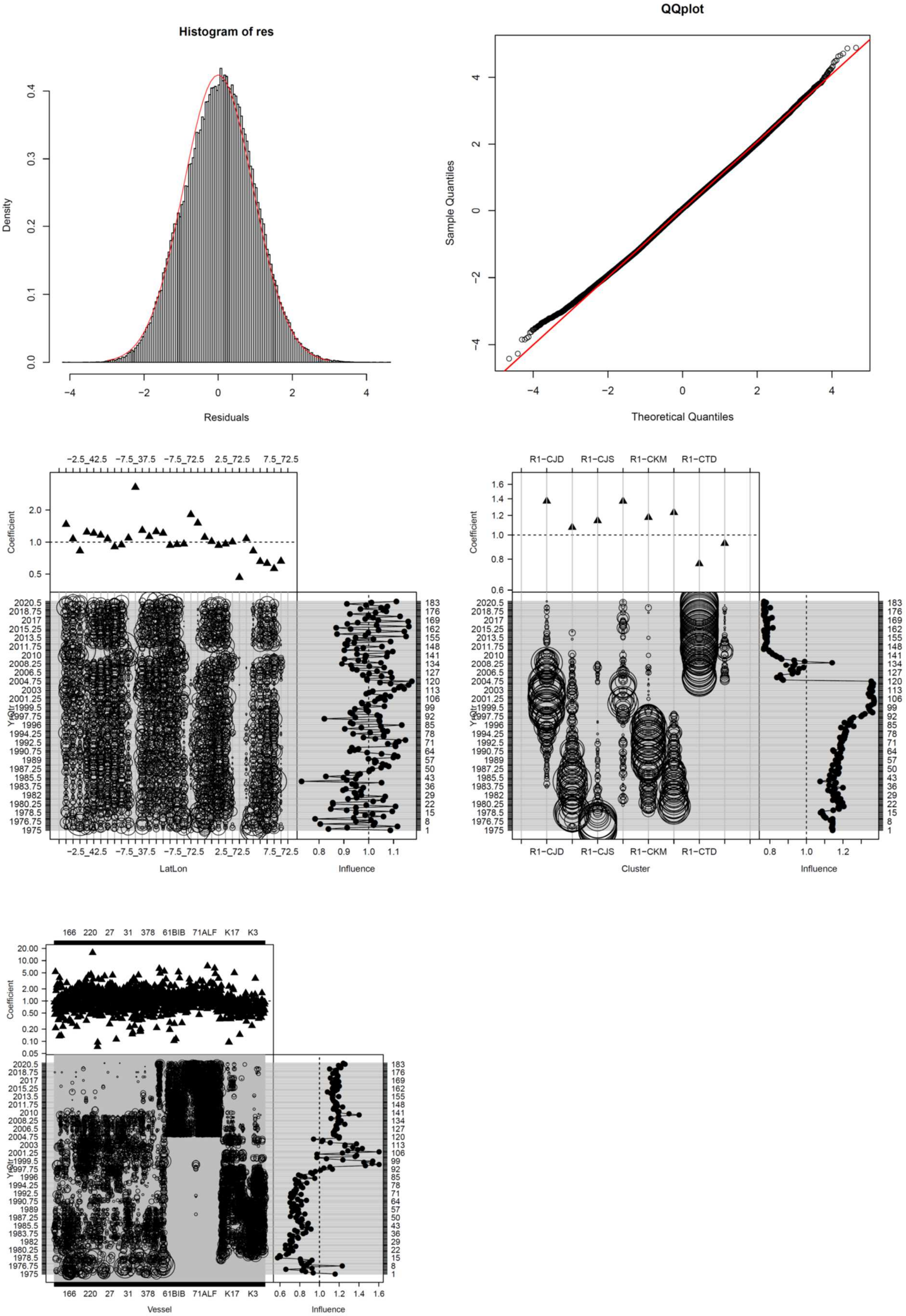


Figure 4(b): Diagnostics and influence plots for LN(~ YrQ + LonLat + HBF+Vessel) for R1.

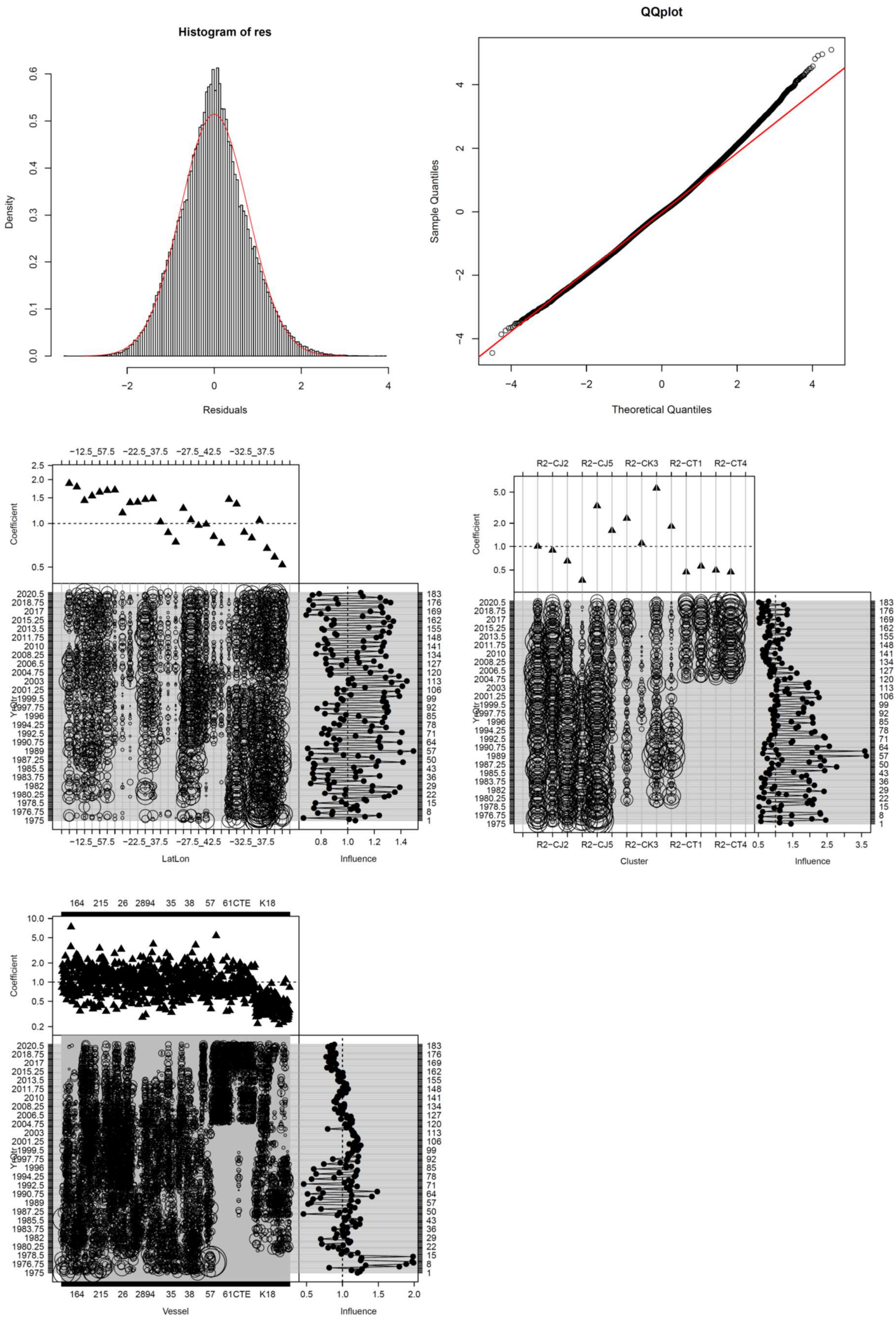


Figure 5: Diagnostics and influence plots for LN(~ YrQ + LonLat + Cluster+Vessel) for R2.

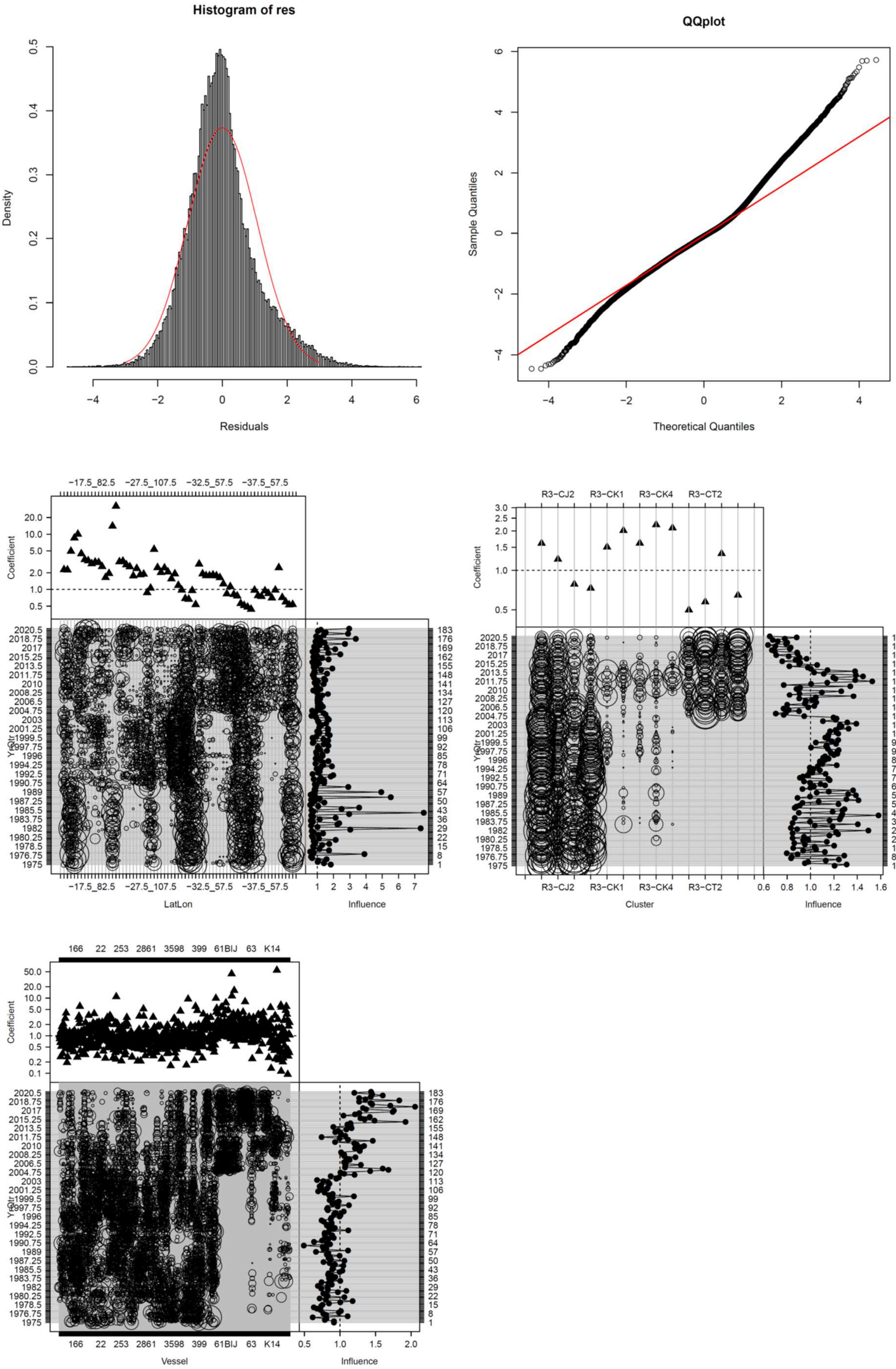


Figure 6: Diagnostics and influence plots for LN(~ YrQ + LonLat + Cluster+Vessel) for R3.

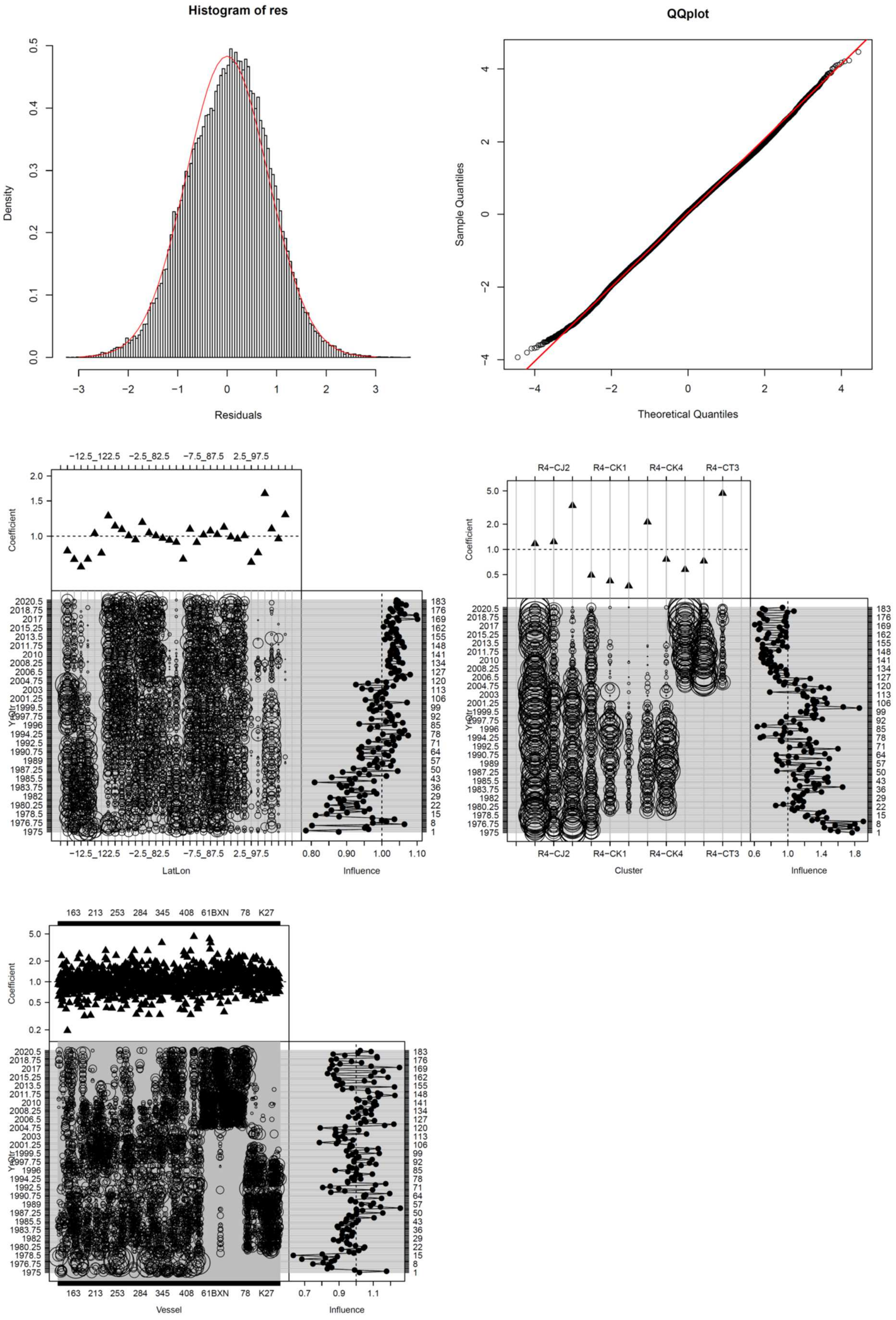


Figure 7(a): Diagnostics and influence plots for LN(~ YrQ + LonLat + Cluster+Vessel) for R4.

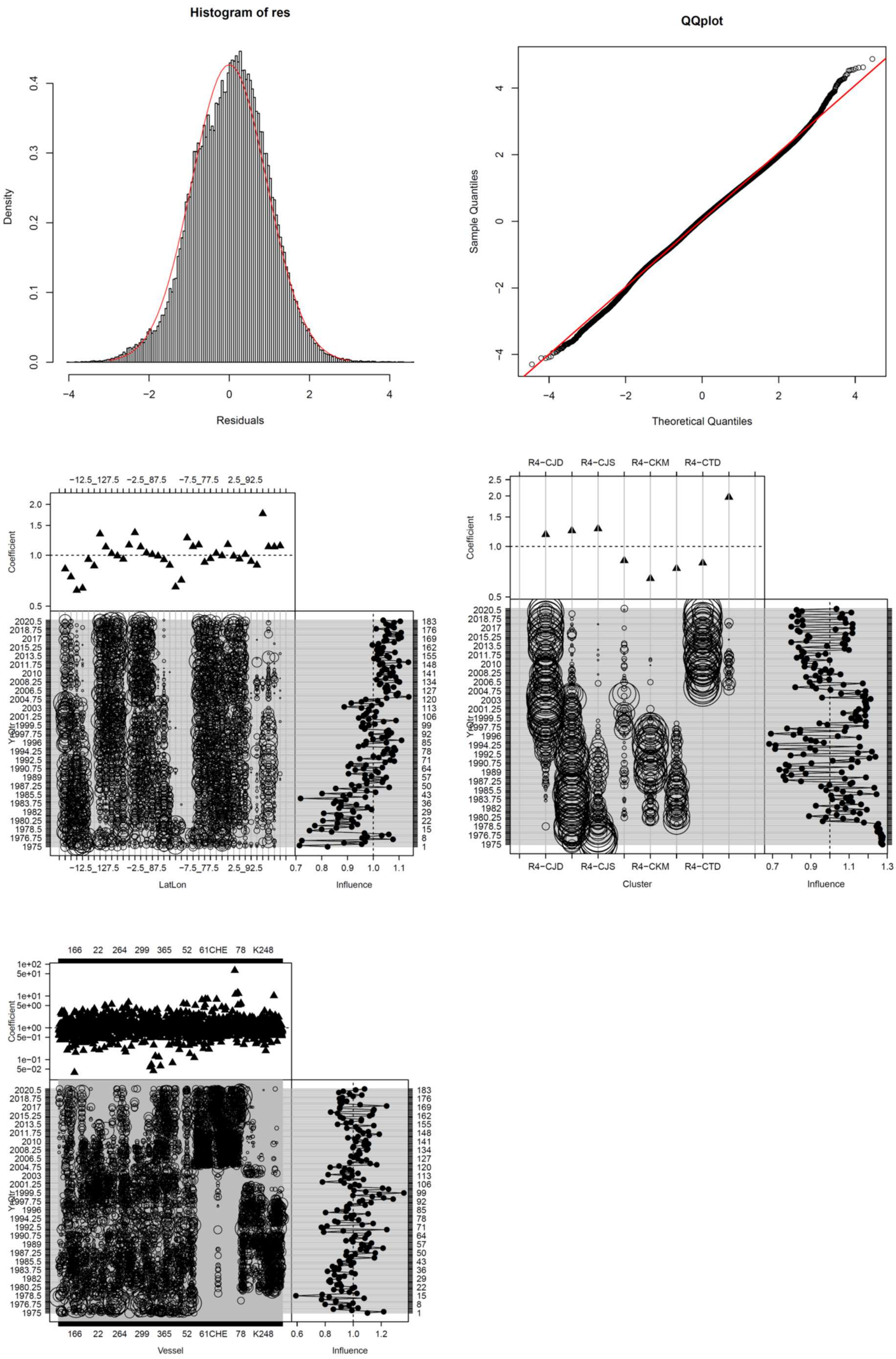


Figure 7(b): Diagnostics and influence plots for LN(~ YrQ + LonLat + HBF+Vessel) for R4.

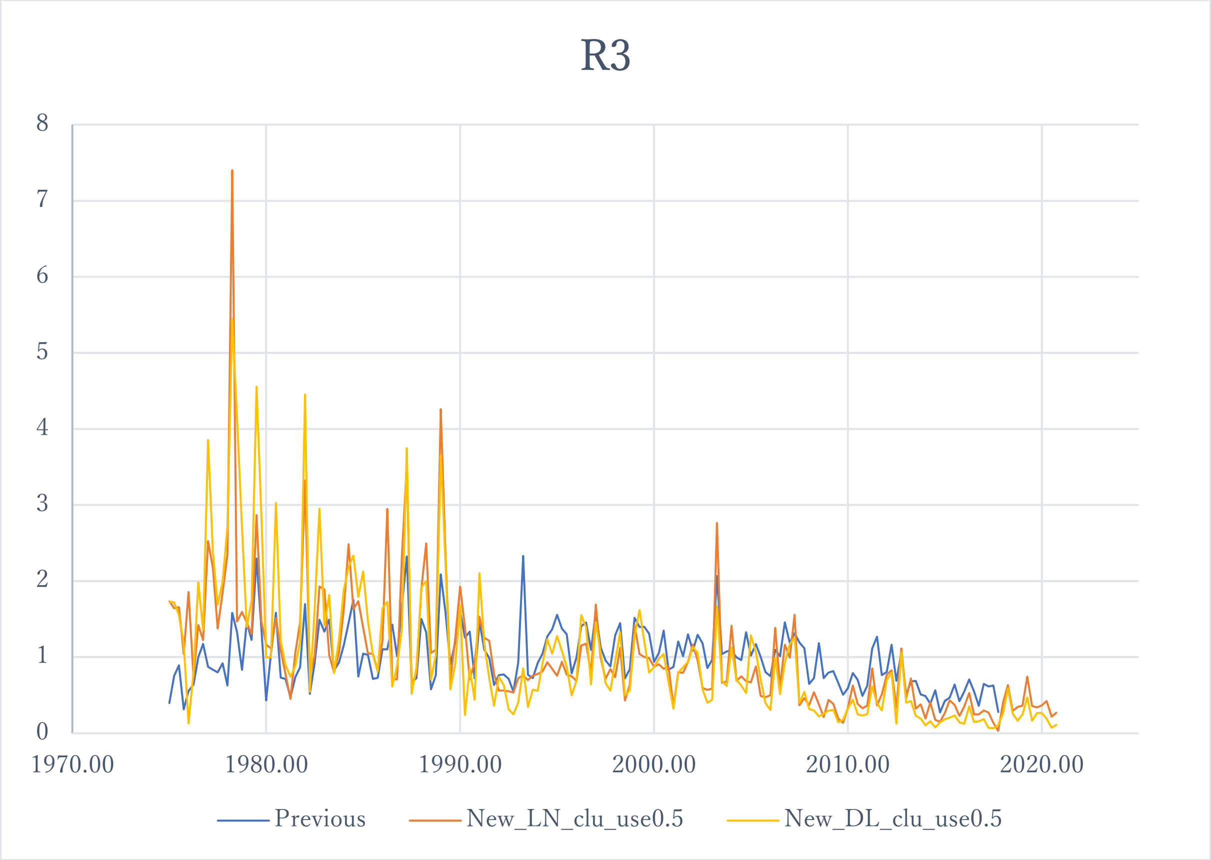


Figure 8. Comparison of CPUEs in R3.