
**SCIENTIFIC COMMITTEE**
**ELEVENTH REGULAR SESSION**
Pohnpei, Federated States of Micronesia
5-13 August 2015

# Longline CPUE indices for bigeye tuna based on the analysis of Pacific-wide operational catch and effort data

**WCPFC-SC11-2015/SA-WP-02**

**S McKechnie[1], L Tremblay-Boyer, and S J Harley**

[1]Oceanic Fisheries Programme, Secretariat of the Pacific Community

# Contents

# Executive Summary

The importance of access to 'raw' or 'operational' level longline catch per unit effort (CPUE) effort data for stock assessment is increasingly accepted within the WCPFC membership and this paper describes the first examination of the largest operational longline data set likely ever assembled for any tuna fishery. It includes fishing fleets from 31 countries, most notably the Distant Water Fishing Nations (DWFNs) of China, Chinese Taipei, Japan, Korea, and the United States of America, and covers the entire breadth of the Pacific Ocean from around 45° N to 40° S, over sixty years of fishing (1952–2013) – 10.2 million sets and 22 billion hooks.

These data were assembled to support an investigation of the impacts of including eastern Pacific Ocean data on the model estimates of bigeye stock size for the western Pacific Ocean. The provision of data by the DWFNs was part of a collaborative arrangement specifically for this analysis – the current agreement is that these data are only available for the Pacific-wide comparison and must be deleted at the conclusion of SC11 in August 2015.

The assembly of the data set was not straightforward and included considerable work combining different sources of data for some DWFN fleets to ensure full coverage without inclusion of duplicate records. Through this process, some gaps in the available data for critical variables were found, and recommendations for future data rescue work made. Ultimately these data issues shaped the landscape of possible analyses and led to two CPUE series being constructed for each region.

Analysing such a large data set was computationally challenging and this limited the level of investigation and analysis which was possible. Therefore, we view the resulting indices as preliminary and not necessarily suitable for inclusion in a full stock assessment, but sufficient for the exercise at hand.

Nevertheless, considerable insights have been gained though this work and, with the approval of the WCPFC and Parties to the data sharing arrangement, we believe that there is a clear plan of data rescue, statistical modelling, and simulation work which would benefit not just future bigeye tuna assessments, but any assessment which includes longline CPUE data.

# 1 Introduction

Indices of catch-per-unit-effort (CPUE) for longline (LL) fisheries are among the most important inputs for almost all stock assessment models for the Western and Central Pacific Ocean (WCPO), in particular that of bigeye tuna. Past stock assessments have used a variety of methods to produce indices that were considered to be the best possible given the data available, however availability has not always been complete. Operational-level longline (OPLL) data is generally considered to be superior to aggregated data (Hoyle and Okamoto, 2011; Ianelli et al., 2012) as these data should provide the most information to take into account targeting practices and other factors influencing fishing success (e.g., new more efficient vessels).

While there is technically no argument about the value of these data, several Distant Water Fishing Nations (DWFNs) have indicated that domestic legal constraints make it difficult for them to provide these data to the WCPFC. Therefore, a significant proportion of operational-level data existing for fisheries operating in the WCPO is not held in SPC databases, though access to data for some DWFNs has been made available through specific collaborations (e.g., with Japan: Hoyle and Okamoto, 2011; Chinese Taipei: McKechnie et al., 2014b; and the United States: Bigelow and Hoyle, 2012).

The current modelling exercise aimed at assessing the suitability of undertaking a WCPO bigeye tuna assessment in isolation of the EPO (McKechnie et al., 2015) – a recommendation from the independent review of the 2011 bigeye tuna assessment (Ianelli et al., 2012) which has been the catalyst for the DWFNs in providing SPC temporary access to operational-level data which were previously unavailable, or under restricted access. Operational-level data for this study has been made available for analyses in Noumea, under several conditions as stated the appendix of OFP (2015), up until the end of the Eleventh Regular Session of the WCPFC Scientific Committee (SC11) in August 2015, at which time the data, including intermediate products which can restore the data, shall be deleted. Consequently the data sets presented herein represent the full (or very close to it) set of operational-level data across all fishing nations–both DWFNs and Pacific Island Countries and Territories (PICTs) in the Pacific.

This paper describes the first analyses of these data within a methodological framework which largely follows the work of McKechnie et al. (2014b). Within the paper there is considerable focus on a series of recommendations reflecting the many lessons learned and insights obtained through this work.

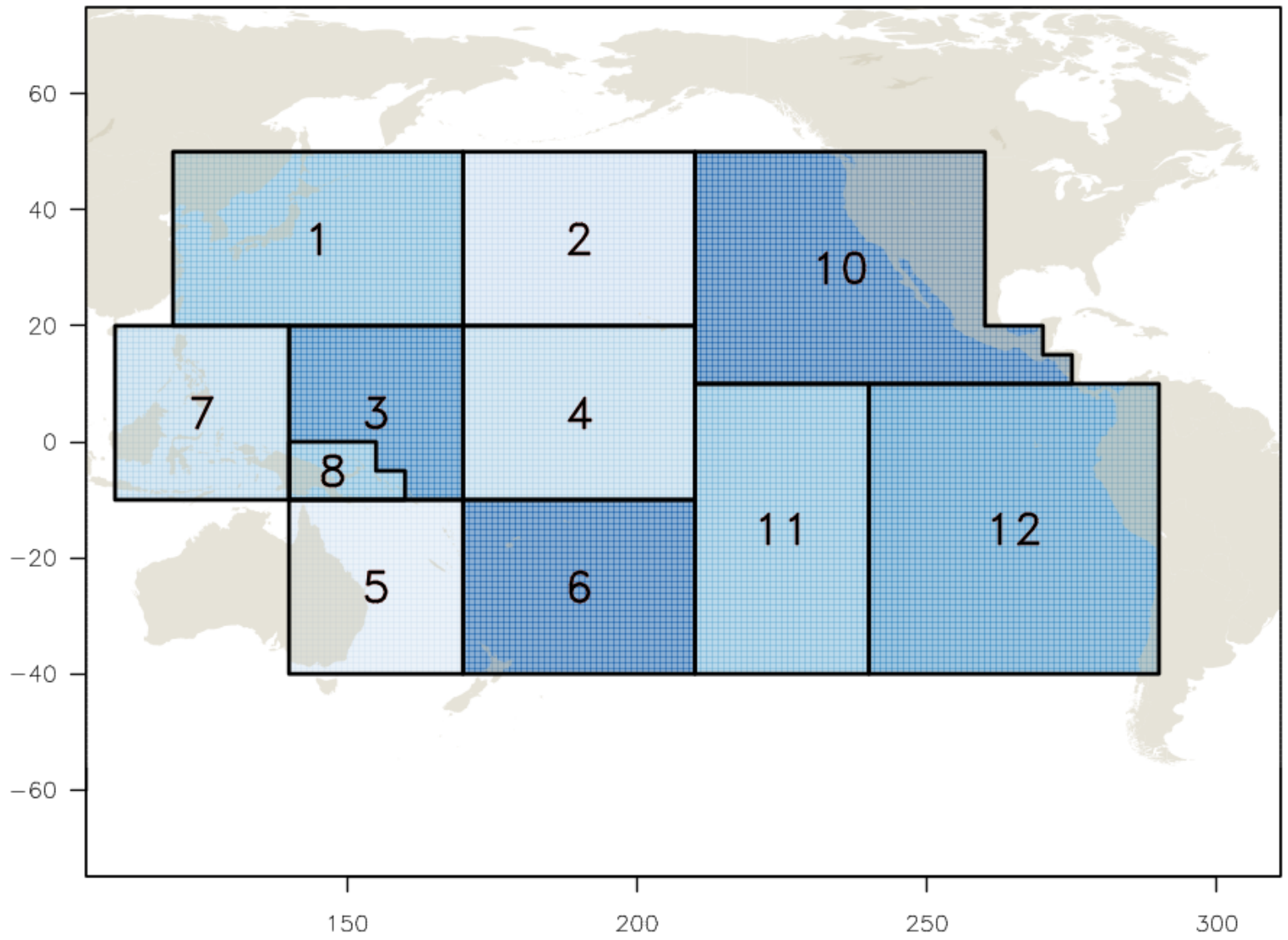


Figure 1: Eleven spatial regions used for the CPUE analysis. Note: a 12th region, region 9 is included within region 5, and will be a separate region in the Pacific-wide analysis.

# 2 Methods

In this section we will outline the many steps taken to get from the raw data to the standardized abundance indices. These steps involved data compilation and reconciliation, data grooming, exploratory data analyses, stratification into spatial strata consistent with the Pacific-wide analysis, application of clustering techniques to help identify targeting practices and their changes in space and time, selection of final data sets through the identification of targeting clusters to include and criteria for a standard 'core-vessel', application of Generalized Linear Models (GLMs) with alternative error structures to determine the most appropriate models, and finally, estimation of temporally-varying standard errors for the resulting abundance indices. We briefly describe the data used in the analysis and each of these steps in the sections below.

## 2.1 Data

The operational-level longline data used in this analysis refer to individual records of fishing activity, whereby on a given day and time a longline was set by a vessel in a particular location, subsequently hauled a few hours later, and the numbers of fish caught of various species reported. The longline set itself would be characterized by a total number of hooks set and by the number of hooks that were set between each intermediate float deployed along the length of the line (hooks-between-floats or HBF). These intermediate floats are used to maintain the fishing gear at a particular depth in the water column. The data included catches of the four species most likely to be targeted in longline operations: albacore tuna (ALB), bigeye tuna (BET), yellowfin tuna (YFT), and swordfish (SWO). Whilst we are only deriving CPUE for bigeye tuna, the catches of all four species are critical to the interpretation of BET CPUE trends *via* the understanding of targeting behaviour. The specific data fields present in the data set are provided in Table 1.

Table 1: Description and availability of different variables important for standardising CPUE

| Variable | Description | % records avail. | GLM Coefficient |
|---|---|---|---|
| cell | $5 \times 5$ spatial cell within which the first hook of the set was fished | 100 | $\beta_{S[i]}$ |
| hbf | Number of hooks between floats for the set | 74 | NA |
| vessel | The individual identity of the vessel fishing the set | 76 | $\beta_{V[i]}$ |
| hooks | Measure of effort: the total number of hooks fished in the set | 100 | $\beta_H$ |
| BET | Number of BET caught in the set | 100 | NA |
| yrqtr | The year and quarter of during which the set was fished | 100 | $\beta_{T[i]}$ |
| cluster | Categorical variable attempting to assign the set to general targetting activity | NA | $\beta_{C[i]}$ |

Aside from the United States and its territory of American Samoa, SPC holds all OPLL data

for its members and also has significant holdings for DWFNs that fish within SPC member waters. However, the coverage of these DWFN OPLL data held by SPC for within-EEZ fishing is incomplete, especially historically, and coverage for the vast areas of High Seas is very low. Therefore the provision of all available operational data by the DWFNs—China (CN), Chinese Taipei (TW), Japan (JP), Korea (KR) and United States (US)—is critical to getting the best possible spatio-temporal coverage of longline fishing.

### 2.1.1 Data reconciliation

The first critical step was the integration of the DWFN holdings with those of SPC. This work was initiated during the operational data CPUE workshop held in Noumea in February 2015 (OFP, 2015) and continued after the workshop. Particularly challenging was the integration of the DWFN data collected at Pago Pago (American Samoa) which covered fishing in the early years of the fishery (pre-1980) and included fishing activities by KR, TW, and JP vessels that were not covered in the operational data held by these flag-states (i.e., the US held data for these DWFNs that the DWFNs did not hold themselves). A further complication stemmed from the risks of duplicates when combining data holdings, since unique vessel identifiers were not always consistent within and amongst fleets (e.g., vessel names are notorious for different spellings or reporting because they are often based on similar names and numbers). Overall this task took several weeks and required considerable liaison with the DWFNs.

This data set, covering the entire breadth of the Pacific Ocean from around 45° N to 40° S and over sixty years of fishing (1952–2013), comprised 10.2 million sets and 22 billion hooks (Figure 2).

### 2.1.2 Data cleaning

Not all of the data were suitable for subsequent analysis and there were several filters applied before it was deemed appropriate for inclusion in the standardization. At the first basic level, these filters included:

1. Missing key attributes such as location, date, or number of hooks set.
2. No catch of any of the four key species for which catch data was provided.
3. Fishing outside of the Pacific-wide analysis boundaries (see Figure 1).

4. Very unlikely gear characteristics suggestive of either errors or very different fishing techniques, e.g., outliers for number of hooks per set or HBF.

This initial screening removed only a relatively small number of records (∼21,000 records representing $< 1\%$ of the overall data set). Due to time limitations no attempt was made to correct / fill-in any of the gaps or outliers. In the future it may be possible to do so by examining other records for the same vessel and determine an appropriate value (e.g., there may have been simple transcription errors).

A second level of screening occurred through the construction of alternative data sets that were included in the modelling process, i.e., further filtering was applied for some analyses, but not all. The three key areas of additional screening were: (1) absence of a unique vessel identifier; (2) fishing sets assigned to a particular mode of fishing / target species (from the cluster analysis described below) that were not considered appropriate in a particular standardization; and (3) fishing sets from vessels deemed to have had too little fishing in a particular region, i.e., the number of data points they contributed to the analysis did not outweigh the cost of having to estimate an individual vessel effect.

The three filters are discussed in further detail below as each of them reflected a "fork" in the structure of the overall analysis.

### 2.1.3 Restriction of data to those with vessel identity

It has been long recognized that the variability in fishing efficiency among different fishing vessels is one of the most important factors describing variation in CPUE and as such is critical to any CPUE analysis. This is an important variable in standardization as it allows for some of the changes in fishing efficiency of vessels entering and exiting the fishery (and data set) to be accounted for (Hoyle and Okamoto, 2011), and allows data from multiple fleets with different catchabilities to be included in analyses (McKechnie et al., 2014b).

In the current analysis we are faced with the challenge that individual vessel identifiers for Japan are not available in the electronic records prior to 1978 (there is a rapid increase in coverage of vessel identities after this time, see Figure 3). This is particularly important as Japanese fishing activity dominated the early period of the fishery and represents almost all of the operational data that exist prior to 1978. There was also a small number of records from the Korean fishery for which vessel identifiers were missing.

Such was the magnitude and importance of the issue that it was not considered appropriate to simply exclude all records without vessel identifiers. Instead we decided to proceed with

Table 2: Summary of parameters used in filtering data, by region

| Region | Min Y-Q core vessels | Start-year core fleet | # clusters identified | Cluster removed |
|---|---|---|---|---|
| 1 | 10 | 1979.125 | 4 | SWO1 |
| 2 | 10 | 1979.125 | 3 | SWO1 |
| 3 | 10 | 1979.125 | 3 | ALB1 |
| 4 | 20 | 1963.375 | 3 | ALB1 |
| 5 | 10 | 1979.375 | 3 | SWO1 |
| 6 | 10 | 1963.625 | 3 | SWO1 |
| 7 | 10 | 1979.125 | 3 | ALB1 |
| 8 | 10 | 1979.875 | 3 | ALB1 |
| 10 | 10 | 1979.125 | 4 | SWO1 |
| 11 | 15 | 1963.625 | 2 | ALB1 |
| 12 | 10 | 1979.125 | 3 | ALB1 |

two parallel lines of analysis. At this stage we constructed two data sets: (1) the **full − long series** data set covered the period 1952–2013, but did not include vessel as a potential explanatory variable; and (2) the **vessel − short series** which only included sets where a unique vessel identifier was available, and depending on the region, typically started between the mid-1960s and 1980. It was predominantly records from Pago Pago (i.e., largely albacore targeting) which allowed us to go earlier than 1978 with identified vessel data (see Table 2 for region-specific start dates for the short data set).

#### 2.1.4 Restriction of data to those vessels with sufficient fishing activity

While vessel is a powerful explanatory variable, it is costly to include in models in terms of computational resources. One must consider if a vessel has undertaken sufficient fishing to allow for a robust estimate of its own catchability, and also whether it provides sufficient information about changes in BET abundance to warrant inclusion. For the key regions in the analysis, it was necessary to have quite stringent criteria for a vessel to be included just in order to allow the models to run within the 48GB memory limitation of the computer.

Given the time constraints, it was not possible to fully examine a wide range of cut-offs or analytically select the optimal cut-offs to use in each of the 12 regions. Instead, we aimed to retain as much data as was computationally feasible while also excluding vessels for which it was impractical to estimated effects due to their short fishing activity. For the models presented here, we thus removed all sets for vessels fishing for under 10 quarters in a given region. In regions 4 and 11 this filter was increased to 20 and 15 quarters (Table 2 to reduce the number of estimable parameters to a practical level.

## 2.2 Clustering

### 2.2.1 General approach to clustering

We used clustering analyses, a classification technique, to assign fishing activity to general categories representing the different targeting practices of fishing vessels (He et al., 1997; Bigelow and Hoyle, 2012). Clustering was based on the relative abundance of key target species (the species composition in proportions of ALB, BET, SWO and YFT) caught in sequential groups of sets (a rule-based designation of "trips", defined in section 2.2.2). We used k-means clustering with the Hartigan-Wong algorithm, as implemented in the R package stats, since it both handled well large data sets and succeeded at capturing key spatial and temporal trends in targeting. The k-means algorithm assigns observations to a user-defined number of groups ("clusters") to maximize the Euclidean distance between the group means. Each run of the algorithm was given 10 random starts to ensure that the final clusters were stable. For each region we ran the algorithm with 2, 3 and 4 clusters and selected the minimum number that was able to isolate both overall and decade-specific non-bigeye targeting events (see Table 2).

There are numerous alternative clustering algorithms that can be used in such analyses, and differences in the assignment of sets to targeting clusters cannot be ruled out. Hierarchical clustering using the hclust algorithm available in R package stats was also conducted but it was too computationally costly to fit to many of the regions with large sample sizes.

Instances where none of the 4 focal species were caught in a set where excluded early on from the analyses, and as such not included when building clusters. These occurrences were most prevalent in fishing activity occurring towards the edge of BET range where vessels were presumably targeting species such as Southern Bluefin Tuna (SBT).

### 2.2.2 Definition of a targeting "trips"

Before running clustering algorithms, we had to define a time scale over which targeting activity would be identified by clustering. If clustering was undertaken at the individual set-level, the computational burden would be extreme and random variation in catch might lead to sets from vessels targeting a certain species being erroneously assigned to another cluster. In contrast, if too many sequential sets are grouped together, then a vessel may switch its target species between sets and all sets in that trip will be assigned to one targeting cluster despite very different between-set CPUE of BET (this is particularly a problem with DW

fishing vessels where the usual definition of trips may be all sets between unloadings, which might be months or longer).

We attempted to address the difficulties of defining "trip" by using two grouping methods, with the aim of avoiding both trips that were too short (and thus susceptible to random variation in species composition) and too long (susceptible to target switching). In the first approach we grouped sets into trips if they were fished by the same individual vessel in the same month. This was the definition of trip that was considered desirable when the identity of vessel was known, and was thus applied to the vessel short data set where the vessel identity of all sets was known.

For the full long data set where vessel identity was frequently missing, we were forced into an alternative definition of trip. We therefore grouped sets into trips if they were fished in the same month in the same $1 \times 1$ spatial cell under the same flag (known for all sets). This definition of trip was also applied to the short series data set to allow a direct comparison with the vessel-month grouping. Final classification under the two methods were similar in most instances, which is expected since most vessels operate in the same general area in a month and there are strong large scale patterns in cluster membership over the Pacific (e.g. see Figures 55 to 60).

### 2.2.3 Deciding on which cluster 'scenario' to use for each region

An important issue when including a cluster variable in standardisation models is deciding on the number of clusters that best describes the targeting activities of the fleet. We used multiple diagnostics including visual inspection of plots of the relationship between the number of clusters assumed and the proportion of variance explained. Where this increasing curve shows evidence of inflection suggests the point where adding additional clusters has little impact on improving the fit to the data. This was supported by refitting k-mean clustering to data from each decade separately to confirm that no important targeting activity was missed by using too small a number of clusters when applying the algorithm to the full time period.

## 2.3 General GLM approach

The GLM approaches used herein closely follows those previously investigated for OPLL data in the WCPO (Hoyle and Okamoto, 2011, McKechnie et al., 2014b). Expected distributions

of response variables, tractability and computational costs were considered when formulating a preliminary set of distributional families, link functions and the structure of linear predictors of GLMs to fit. For example it is computationally challenging to fit zero-inflated (ZI) models (e.g. ZI-Poisson, ZI-negative binomial) to these large data sets and convergence issues have been encountered for regions where data set size has allowed these models to be fitted. The distributional form of CPUE data is expected to be region-dependent owing to BET habitat preference and spatial patterns in targeting practices of LL vessels, and so several different models were considered.

OPLL data in the WCPO is characterized by the high frequency of records with zero catches for the focal species and have previously been fitted using both regular catch-based GLMs (e.g., negative-binomial, NB, Bigelow and Hoyle (2012)) and hurdle models (e.g. delta-log-normal models, DLN, Hoyle and Okamoto (2011)). Herein we initially fit NB, DLN and delta-gamma models (DG) and select the most appropriate models for further investigation by way of model diagnostics and metrics of model comparison.

### 2.3.1 Negative binomial GLMs

The counts of BET, $n_i$, were modelled using the negative binomial GLM

$$n_i \sim \text{NegBin}(\mu_i, \theta) \tag{1}$$

$$\log(\mu_i) = \beta_0 + \beta_{T[i]} + \beta_{S[i]} + \beta_{V[i]} + \beta_{C[i]} + \beta_H \log h_i \tag{2}$$

where $\mu_i$ is the mean count and $\theta$ is the size parameter. The log link is used to express $\mu_i$ in terms of the linear predictor, where the $\beta_0$, $\beta_{T[i]}$, $\beta_{S[i]}$, $\beta_{V[i]}$, and $\beta_{C[i]}$ are the coefficients for the intercept, time period (year-quarter), spatial cell ($5 \times 5$ square), vessel, and cluster, respectively, for the factor levels of each variable observed for set $i$. Parameter $\beta_H$ is the coefficient for the relationship between logarithm of mean count and $\log h_i$, the logarithm of hooks (divided by 100) fished for set $i$. The negative binomial likelihood was parameterised by the mean $\mu_i$ and variance $V_i = \mu_i + \mu_i^2/\theta$.

### 2.3.2 Hurdle models

The proportion of positive catches of BET at the set level were modelled using binomial GLMs with a binary response variable ($y_i$ ; $1 =\geq 1$ fish caught, or a $0 =$ zero fish caught in

set $i$)

$$y_i \sim \text{Bernoulli}(p_i) \tag{3}$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{T[i]} + \beta_{S[i]} + \beta_{V[i]} + \beta_{C[i]} + \beta_H h_i \tag{4}$$

where $p_i$ is probability of at least one BET being caught in set $i$, the logit link function is used to express this probability in terms of the linear predictor and the model coefficients have the same interpretation as in section 2.3.1, although note that $\beta_H$ is the coefficient for the continuous variable of hooks fished (divided by 100) for set $i$ on the natural- rather than log-scale.

The positive component of the hurdle model was modelled using either Gaussian or gamma GLMs which are given by

$$\begin{aligned} \log c_i &\sim \text{Normal}(\log \mu_i, \sigma^2) \\ \log \mu_i &= \beta_0 + \beta_{T[i]} + \beta_{S[i]} + \beta_{V[i]} + \beta_{C[i]} \qquad \text{, or} \end{aligned}$$

$$\begin{aligned} c_i &\sim \text{Gamma}(\mu_i, \theta) \\ \log(\mu_i) &= \beta_0 + \beta_{T[i]} + \beta_{S[i]} + \beta_{V[i]} + \beta_{C[i]} \end{aligned}$$

where $\log c_i$ and $c_i$ are the CPUE of BET (number caught divided by hundred hooks) on the log- and natural-scale, respectively, and the parameters in the linear predictor again have the same interpretation.

### 2.3.3 Weighted GLMs

Once final standardised indices were produced for all regions, we modified the GLM likelihoods to allow unequal prior weightings for individual datapoints, as suggested by Campbell (2015), to assess the sensitivity of indices to unequal distribution of effort between cells over time. Weights, $w_{t,s}$, in year-quarter $t$ in cell $s$, were assigned to individual sets in these strata by calculating

$$w_{t,s} = \frac{N_T}{N_Q} \times \frac{1}{n_{t,s}} \tag{5}$$

where $N_T$ is the total number of sets in the dataset being analysed, $N_Q$ is the number of unique strata (year-quarter–cell combinations), and $n_{t,s}$ is the number of sets in a strata comprising year-quarter $t$ and cell $s$. We propose further investigation of weighting schemes and development of appropriate diagnostics for these models which can complicate routine ex-

amination of residuals and simulation-based diagnostics. The simulation tests of operational datasets outlined in section 4.8 will be valuable in this respect.

## 2.4 Model fit and comparison

A range of model summaries were calculated from the fitted GLM models which were fitted using Template Model Builder (TMB; Kristensen et al., 2014), including many of those suggested by Hoyle et al. (2014). In addition, we add several new techniques not previously utilized for WCPO CPUE analyses. These included the use of quantile residuals (Dunn and Smyth, 1996) which are an attempt to overcome the problems encountered for binary data modelled using binomial GLMs (e.g. eqns 3–4) and count-based GLMs (e.g. the negative binomial GLM given by equations 1–2) with low mean counts, such as occur in our data sets.

We also simulate data from the fitted GLM model and compare the simulated and the observed data. If the model is adequate then the simulated data will be similar to those observed, at least for the characteristics of the data that are considered to be important for standardizing CPUE indices. Unfortunately, a common characteristic of the OPLL standardization is poor fit to at least some of the data in several regions (e.g., see Hoyle and Okamoto, 2011), which illustrates the difficulty of fitted general models to data from an extremely wide range of fishing activity. Future work will aim to reduce this problem and simulation will ideally be used to test the consequences of aspects of poor model fit. Some of this potential work is discussed in section 4.5.

## 2.5 Step-wise approach followed for model selection

In achieving the goal of producing standardized CPUE indices for the 12 regions used in the Pacific-wide analysis (McKechnie et al., 2015), we went through several steps which have been variously described above and are displayed in Table 3, but will be summarized again below:

**Step 0a:** Undertake clustering analysis and determine the appropriate number of clusters that provide both (a) the most parsimonious description of the data, and (b) identify any particular clusters for which exclusion of the data within them might be warranted because either the fishing behaviour was very different, or a model assuming a common error distribution would be unable to adequately fit the data for each cluster.

**Step 0b:** Through an iterative approach considering both data coverage and computational feasibility, determine the most appropriate region-specific definition for inclusion in the core-fleet (e.g., the criteria for the number of quarters that a vessel must have fished to be included). This was done with the short time series which only included data with vessel identifiers meeting region-specific minimum length of fishing activity.

**Step 1:** Apply the following error distributions to model CPUE using the data set (a; short data set, no data removed for any clusters) from Step 0 above: delta-lognormal, delta-gamma and negative binomial. Selection of the most appropriate error distribution to use for subsequent analyses was based on model diagnostics (see section 4.9 for key diagnostics). This was a "full" model with all data and variables included.

**Step 2a:** Continuing with only the delta-lognormal and negative binomial GLMs, models were rerun with a new data set with records belonging to specific clusters removed (i.e., the cluster thought to be most different to others in terms of non-BET targeting). Two models were run (1) with cluster as a explanatory categorical variable, and (2) without cluster as a variable.

**Step 2b:** Delta-lognormal models were then fitted to the data set in 2a, but with cell, and vessel removed in turn.

**Step 3:** Applying a similar approach, but with only the delta-lognormal distribution given it generally had the best diagnostics for the short dataset, models with the full set of explanatory variables were fitted to the long time series data set (i.e. long data set, no data removed for any clusters).

**Step 4:** The process of reducing the data set to exclude data for specific clusters for the short data set was also adopted for the long data set. The delta-lognormal model was fitted to these data with (1) cluster as a categorical variable, and (2) without cluster as a variable.

A comparison among the various models is then made to assess the effects of the different choices made at each step.

Table 3: Description of the different models and datsets used to standardise CPUE

| Step | Dist | Datset | Data removed | Formula | Index | Figure |
|---|---|---|---|---|---|---|
| 1 | NB | Short | N | count = yrqtr+cell+cluster+vessel+log(hook) | NB-sh-full | 72 |
| | DLN | Short | N | Bin: pos = yrqtr+cell+cluster+vessel+hook<br>LN: log(cpue) = yrqtr+cell+cluster+vessel | DLN-sh-full | 72, 76 |
| | DG | Short | N | Bin: pos = yrqtr+cell+cluster+vessel+hook<br>LN: cpue = yrqtr+cell+cluster+vessel | DG-sh-full | 72 |
| 2a | DLN | Short | Y | Bin: pos = yrqtr+cell+cluster+vessel+hook<br>LN: log(cpue) = yrqtr+cell+cluster+vessel | DLN-sh-drm-full | 74, 75 |
| 2b | DLN | Short | Y | Bin: pos = yrqtr+cell+vessel+hook<br>LN: log(cpue) = yrqtr+cell+vessel | DLN-sh-drm-clst | 74, 77 |
| | DLN | Short | Y | Bin: pos = yrqtr+cluster+vessel+hook<br>LN: log(cpue) = yrqtr+cluster+vessel | DLN-sh-drm-cell | NA |
| | DLN | Short | Y | Bin: pos = yrqtr+cell+cluster+hook<br>LN: log(cpue) = yrqtr+cell+cluster | DLN-sh-drm-ves | 75 |
| 3 | DLN | Long | N | Bin: pos = yrqtr+cell+cluster+vessel+hook<br>LN: log(cpue) = yrqtr+cell+cluster+vessel | DLN-lg-full | 76 |
| 4 | DLN | Long | Y | Bin: pos = yrqtr+cell+cluster+vessel+hook<br>LN: log(cpue) = yrqtr+cell+cluster+vessel | DLN-lg-datrm-full | 78 |
| | DLN | Long | Y | Bin: pos = yrqtr+cell+vessel+hook<br>LN: log(cpue) = yrqtr+cell+vessel | DLN-lg-datrm-clst | 78 |

## 2.6 Construction of standardized indices

From the basic GLM analysis, two further steps are necessary to produce relative abundance indices in the format required for the Pacific-wide analysis. The first is the construction of the year effects in the appropriate parameter space based on the link function used, and the second is the calculation of estimates of relative uncertainty in these year effects. These steps are further described below.

### 2.6.1 Calculation of point estimates of relative abundance

Indices for the individual components (binomial and log-normal) were estimated by predicting CPUE for each year-quarter for the most common level for each categorical variable and the mean value across all sets for continuous variables. The two components were then combined by multiplying them on the appropriate scale to produce the DLN index (Maunder and Punt, 2004). CPUE indices were normalized by dividing their mean across the time period they were calculated. Note that for the negative binomial model this process yields identical indices to those produced by using only the year effects (once indices have been normalized by their mean). When indices are directly compared in figures below but their length differs, the mean CPUE used in the normalization of both indices is calculated over a common set of year-quarters (the set available for the shorter index).

### 2.6.2 Estimation of coefficients of variation

Time-variant coefficients of variation (CVs) for the relative abundance component of the likelihood in MULTIFAN-CL (MFCL) have been implemented in previous stock assessments of BET (e.g., Harley et al. 2014). Calculation of these CVs is partly determined by the estimates of standard errors for the year effects of the fitted GLMs. The absolute values of these SEs are unimportant as the non-independence of the OPLL data makes them extremely conservative, and observation error is only one component of the uncertainty represented in the relevant likelihood in MFCL. The relative variation between years, however, is considered valuable. There are issues with estimating year-specific SEs from GLMs (Ianelli et al., 2012) and so we use Francis (1999)s canonical method of estimating relative CVs from the output of the log-normal component of the standardized DLN indices.

# 3 Results

## 3.1 Data available

A total of just under 10 million usable sets were available within the PW-BET assessment boundaries, with JP (6.3) the largest contributor of data with the other DWFNs contributing between 0.4 (US) and 1.1 (KR) million sets. Other fleets (OT), mainly consisting of PICTS, contributed 1.1 million usable sets. The size of data sets available in different regions varied considerably with some general features being; largest sample sizes were available for regions 1–4 (up to $\sim 60,000$ sets per year) and 11–12, JP dominated the available data for all regions until about the late 1980s when effort began to decline in all regions apart from region 1, KR became a very significant component of the data in region 4 (and to a lesser degree region 11) from the 1980s onward, a similar pattern was seen for US in region 2, PICTS vessels became very important in regions in the southern hemisphere (particularly regions 5 and 6) from the 1990s (Figure 2).

## 3.2 Cluster analyses

The results of the clustering analyses were extremely region-specific (Figures 44–54) due to the large differences in abundance of the focal species, the consequent differences in fishing methods and targeting activity, and the different fleets operating in different areas. The overall spatial distribution of clusters, colour-coded by the dominant species in the catch, is shown in Figures 55–60. In general, the process of selecting the number of clusters to proceed with best supported three clusters, with only two of exceptions. These clusters generally represent targeting of BET/YFT in the tropical regions, often with an ALB-targeting cluster contributing a small number of sets, and ALB targeting in temperate regions with moderate numbers of sets attributed to BET- and SWO-dominated clusters.

In many regions the clustering identified large changes in putative targeting practices over time, with the nature of target switches again region specific (Figures 61–71). Two broad patterns of time variations in clustering are changes in the presence of clusters dominated by species with relatively low contributions to overall catch in that region - often caught by vessels using quite different fishing methods, and temporal changes in the ratio of two or more of the numerically dominant clusters. These are handled in different ways.

Examples of the former include the appearance of clusters with a high proportion of ALB in tropical regions (e.g. region 3; Figure 63) and SWO in some of the temperate regions

(e.g. region 5; Figure 65). One approach we utilised was removing these clusters (one in each region, these result in the drm datasets) under the assumption that this fishing activity provides limited information on relative abundance of BET and removal of these data could potentially improve the distributional assumptions of our models. For each region, the cluster for which data were removed is displayed in Table 2. Note that models are fitted to the full dataset and the data-removed dataset, both with cluster variable included and not included, and so the sensitivity of these assumptions can be assessed (see section 3.4).

Examples of the latter include changes in the ratio of BET- and YFT-dominated clusters (e.g. regions 3 and 7, Figures 63 and 67), the ratio of clusters with different levels of ALB-dominance in southern temperate regions (e.g. region 6; Figure 66) or the ratio of ALB- and BET-dominated clusters in other temperate regions (e.g. region 11; Figure 70). None of these clusters were removed from the datasets as they 1) provide most of the data available of those regions, and so retaining these clusters maximises temporal and spatial coverage, 2) the fishing methods represented by the clusters are in some cases similar, and 3) it is less clear whether these cluster switches represent actual target switches and/or changes of relative abundance of the species involved. These alternative possibilities are explored by fitting models with and without the cluster variable included (section 3.4).

## 3.3 Step 1: Selection of the probability distribution

The negative binomial GLM appeared inadequate for modelling the distribution of counts of BET in most regions (see Figure 126 as an example). It tended to estimate the counts in the upper tail of the observed distribution because it also had to fit the number of zero, and low counts adequately. This is most obvious when observed data and data simulated under the model are compared for different targeting clusters (Figure 126). Often one or more of the clusters would be well fitted but fit in the other clusters would be poor due to the model-estimated dispersion parameter preventing adequate fit across the whole distribution of counts. In general the DLN and DG performed better in most cases due to more flexibility in the modelling of zeros and positive counts, although there was still a lack of fit in some regions. The resulting standardised indices from the DLN and DG models were near identical in all regions (Figure 72) and so only DLN models were considered in further development of modelling.

## 3.4 Step 2: Removal of data from specific clusters

Data for one selected "non-target" cluster was removed from each region and the GLMs were refitted with a cluster variable retained in the model, then without the this variable, in turn. The resulting indices for the data removed indices (drm models) were generally similar to the models for the full dataset (Figure 73), and no substantial differences were observed between the data-removed cluster, and no cluster models (DLN-sh-drm-full vs DLN-sh-drm-clst, Figure 74). In several regions (e.g. region 2 and 8) there was some minor variation between the latter two models in certain year-quarters but the overall dynamics were extremely similar. Larger differences were detected between models with and without the vessel variable, most notably in regions 6, 8 and 10, with moderate differences in several other regions (Figure 75). The affects of this variable were not always in the same direction, e.g. removing the vessel variable increased and decreased the decline in CPUE in regions 3 and 4, respectively, and is dependent on the efficiency of vessels and fleets entering and exiting the dataset.

## 3.5 Step 3 and 4: Standardised indices for the long dataset

Standardised indices for the long dataset with no cluster data removed showed the same general pattern of ongoing declines in CPUE for most regions, over most of the time-series, as were observed for the full short dataset (DLN-lg-full vs DLN-sh-full, Figure 76, and DLN-lg-drm-clst vs DLN-sh-drm-clst, Figure 77). For several regions (2, 7, 10 and 12) the short and long indices were extremely similar over the period of overlap, however there were notable differences in regions 4 (higher CPUE for DLN-sh-full in the early time-series), 6 (significantly higher rate of decline for DLN-sh-full) and 8 (DLN-sh-full shows more temporal variation).

Comparison of the data-removed indices with (DLN-lg-drm-full) and without (DLN-lg-drm-clst) the cluster variable showed similar patterns to the same comparison made on the short dataset (Figure 78). In some regions (1, 4, 5 and 12) the indices were extremely similar, while in others (most notably regions 7, 8 and 10) the two sets of models produced indices with moderately different trends in CPUE, with the direction of the change varying between regions based on the nature of the putative targeting clusters and changes in their balance over time.

For most of the scenarios above, the full model (all predictor variables included) is assumed (apart from the removal of the cluster and vessel variables for the short dataset, in turn, and the cluster for the long dataset). Table 4 shows the deviance explained (using McFadden's method of calculating $1 - D_f/D_n$ where $D_f$ and $D_n$ are the residual deviances for the focal

and null models, respectively) for the DLN-lg-drm models. The models for both components of the DLN showed varying levels of deviance explained among regions for the full model, and in all except two cases, dropping the cell and cluster variables from the model, in turn, led to a decrease in percentage deviance explained of greater than one percentage point, which is often used as a rule for allowing the variable to be retained in the model.

The weighted GLM models equivalent to models DLN-lg-drm-clst are displayed in Figure 79, and show moderate differences in temporal variation but on the whole display very similar long-term trends in CPUE.

### 3.6 Goodness of fit

In general the binomial GLM components of the DLN models had adequate fit to the data as evidence by plots of residuals, and through comparisons of the observed datasets and datasets simulated under the fitted model (not shown). While fit of the LN components was also often adequate there were occasions where poor fit was observed, especially in the tails of the distributions which can be seen in the residual plots and plots comparing observed and simulated datasets (e.g. Figure 130). For the lower tail this can sometimes be attributed to the lower bound of zero on the natural scale, while there is a tendency for the model to underestimate some of the very high CPUEs in the observed data, a feature that is also common to the gamma and negative binomial models. Further discrepancies between models and data come from the apparent clumping of data which is a result of some counts of hooks occurring with higher frequency than others (either rounding by data collectors or a tendency to use a rounded counts when setting up fishing gear).

## 4 Discussion

The analyses undertaken here likely among represent the largest analysis of operational longline data undertaken anywhere in the world. They reflect the general recognition of the importance of operational data for deriving the best available indices of abundance — a key driver of many tuna stock assessments worldwide (Hoyle and Okamoto, 2011; Ianelli et al., 2012). While it is somewhat unfortunate that this data assembly is not supporting an actual stock assessment (rather a sensitivity analysis on the impact of including the Eastern Pacific Ocean (EPO) dynamics on WCPO depletion estimates) it has lead to many advances in knowledge and techniques which will be valuable if collaborative access to these

data continues in the future. In this section we felt it timely to reflect on many of the challenges encountered in the analyses to date, some findings that we think should direct future activities, and some areas of investigation that we were unable to get to – but warrant consideration in future endeavours.

## 4.1 Collaborative arrangements

One of the critical aspects to the success of this collaboration was holding a workshop involving the parties where data was reviewed and exploratory data analyses undertaken. While there was considerable time pressure given the unfamiliarity with some of the data – discussions among all parties lead to a greater understanding of the data and its peculiarities, and there was considerable transfer of knowledge about the data and methodologies between the parties and SPC, and vice-versa. We would strongly recommend that future collaborations also include some dedicated time to consider preliminary analysis of the data – either in a dedicated workshop or as part of the Pre-assessment Workshop (PAW) that SPC often convenes as part of its assessment process.

The workshop also allowed reconciliation of various data holdings. In particular we were able to identify data for DWFN fleets that was held by SPC, but not held by the DWFN, and also data held by the United States from Pago Pago which was not held by other DWFNs. We were able to make some progress on merging these data sets for this analysis and also subsequent to the workshop – communications occurred between DWFNs seeking to obtain missing data.

The restrictions around how SPC analysts were able to use these data warrants further discussion. The restrictions around who had access to the data and where they could be stored, precluded our traditional "distributed-computing" approach whereby analyses are "farmed" out across a range of machines. As a consequence the analyses were done consecutively, i.e., one after the other in a queue, and this greatly reduced the amount of analysis and investigation that could be undertaken. This was exacerbated in this instance by the necessity of analysing data for the EPO – which while it will be a "nuisance" region in the subsequent Pacific-wide analysis, contained regions in our analysis here with large numbers of records and complex fisheries interactions.

## 4.2 Data problems encountered

While the operational longline data set assembled represents an improvement on the dataset available for previous analyses (McKechnie et al., 2014b), the final data set was far from perfect. There were three specific aspects of the data that led to "less than ideal" modelling approaches, and we focus on these below.

With a few exceptions, the only data available for the first 25 years of the fishery came from Japanese flagged vessels for which vessel identifiers were unavailable prior to 1978. The importance of vessel effects (and therefore vessel identifiers) has been previously demonstrated (Hoyle and Okamoto, 2011; McKechnie et al., 2014b) and for the purposes of defining trips for cluster analysis it is highly desirable to be able to classify sets into vessel-time strata which obviously requires information on vessel identity. It was considered inappropriate to attempt to combine CPUE indices from two periods, e.g., an early index with no vessel variable and clustering based on small time/area strata, and a later index with vessel variable and vessel/trip clustering. This required us to have a "fork" in the progression of our analyses where we constructed two "internally consistent" data sets — a long time series where clustering trips were defined as sets within time/area strata and the subsequent CPUE standardization excluded a vessel variable, and a short time series which only includes records with vessel. Therefore the long time series could be viewed as analogous to aggregate catch and effort data – albeit on a finer time scale, and the with the possibility that some impacts of target switching can be controlled for.

There are several other variables that would likely be informative for identifying targeting practices and explaining variability in catchability but are unavailable or have poor coverage. These gaps, summarized in Annex 4 of OFP (2015), primarily relate to an absence of start set time and hooks-between-floats (HBF) for early data for Chinese Taipei and several fleets held by SPC, and the coverage was very low for all other fleets. Ideally these data could be used to assist in distinguishing between spatio-temporal patterns in clusters that are driven by true changes/differences in targeting versus those driven by changes in the relative abundance of one or more of the four species for which we had data.

The final issue of concern is that, with the exception of Japan, most of the early DWFN LL data was collected through the opportunistic sampling programme operated in Pago Pago. There was considerable data for both Korean and Chinese Taipei fishing activities and a small amount of Japanese effort. These data, whilst covering several of the regions in our analysis, were primarily from vessels targeting albacore. This suggests that there was other fishing at the same time in other areas where logsheets were likely completed, but there are

an absence of records (i.e., it is unlikely that the only Korean vessels fishing were unloading in Pago Pago and that these were also the only vessels using logsheets).

## 4.3 Data rescue

OFP (2015) covered some of the important data rescue activities that are reinforced by analyses described in this report. We see considerable value in efforts to reconcile the early Japanese data with unique vessel identifiers and there is data on HBF and set start time which exists, but has not yet been entered into national data bases (see Annex 4 of OFP, 2015). Once these activities have been undertaken, and if a long-term view is taken to the construction of a "best" overall data set, we see value in attempts from SPC to examine missing and extreme values within the data set. This would identify where obvious errors or omissions exist would have the dual aims of producing the overall data set and improved individual DWFN data sets which integrate all available data.

Alternatively, it may be possible to develop dynamic models that use information on spatio-temporal proximity of sets and information on gear characteristics (number of hooks set, HBF etc.) to assign sets to putative "vessels". Estimates of the number of vessels operating in the data set in a given year are already available and would provide guidance of the number of putative vessels to "create". The efficacy of such methods remains unknown and potentially the most challenging issue will be linking individual vessels between time-periods as this is essential for "vessel effects" to be of use for standardizing CPUE data. The methods could be calibrated on data from the 1980s onward where model-based predictions could be compared to the known vessel identities. If the same operational data set becomes available for future stock assessments then undertaking these analyses should be of a high priority.

While such activities will improve both the standardization and clustering analyses, we can see further benefits to the clustering by including additional species – there are some target fisheries, aside from the four species for which we have data, that are unlikely to be identified with just these four species, e.g., target fisheries for sharks in the North Pacific Ocean and southern bluefin tuna in the south. Noting problems with the reporting of shark catches we suspect that inclusion of striped marlin may be sufficient to help differentiate billfish (e.g. swordfish or marlin) and shark targeting.

## 4.4 Computational challenges

Faced with an extremely large data set, eleven regions, and consideration of issues such as the appropriate structure for the models, statistical assumptions for the error distributions, and the inability to use distributed computing, we recognized that the standard approach of relying solely on R was unlikely to be sufficiently fast and instead chose to use a new modelling package called Template Model Builder (Kristensen et al., 2014) which is run from within R, but relies on C++ libraries and the minimisation algorithms used in ADMB (Fournier et al., 2012). This approach can solve statistical models much faster and uses less RAM than its R counterpart, especially with large data sets, and is especially important for modelling random or mixed-effect models, although we have not yet considered that functionality.

There were two conditions which led to a lack of the necessary computing power to undertake all the analyses we would have liked to in the given time frame. On the one hand, we had anticipated, as agreed at WCPFC11 in Apia, Samoa, that specific funding would be made available for the Pacific-wide analysis to support collaborations with the IATTC and the purchase of high-performance computer hardware to tackle both the CPUE analyses and the subsequent MFCL modelling. Unfortunately, due to some confusion, these funds were removed from the overall budget and no funding was available for purchasing a dedicated computer. Therefore, to undertake the work, while respecting the restrictions around storage and use of these data, we established a secure 48GB virtual machine (VM). The large memory requirements were necessary due to the modelling of vessel effects and the large number of records in several assessment regions (1, 4 and 11). While VMs can offer both security and flexibility (e.g., you can specify memory requirements), the computing power of the machine remained limited to a single core such that analyses have to be run one at a time. The overall computer also had to be shared with other users who require high-computing power to run MFCL and were also affected by the reduction in the planned hardware-dedicated funding.

This computing issue could have been somewhat alleviated by using the CONDOR-based approach to distribute analyses among the SPC fleet of network computers, as is done during MFCL assessment runs and in previous analyses of OPLL data (McKechnie et al., 2014b), but the confidentiality conditions prevented us from using this approach.

In short, while we were able to run sufficient analyses for this round of standardization, too much energy was still spent on solving computing issues *vs.* performing actual data analyses and thinking about fisheries (and not informatic) problems. Future work on data sets of this size will necessitate specific computing hardware and while it is possible to "rent" fast and

efficient VMs through various cloud computing providers, we argue that cloud-computing may offer unnecessary risks given the confidential nature of these data.

## 4.5 Further analyses

As noted above, the analyses described here represent a first examination of the Pacific-wide operational data set. There are many areas of further investigation that would undoubtedly improve upon the indices presented here. These are described in further detail below and can be divided into those related to improved analysis of targeting, inclusion of oceanographic data, alternative statistical modelling approaches, and the necessity of building a simulation model to test the robustness of the indices to the assumptions underlying the various approaches used.

### 4.5.1 Targeting and cluster analyses

Arguably the most critical area for future investigation is an in-depth consideration of deliberate changes in targeting strategies and how these can be inferred from operational records. Currently, the clustering approach assumes that differences in the relative prevalence of a given target species in sets reflects a change in targeting strategy. However, if the abundance of a species in a region changed, we would also expect to see a concurrent change in this species' prevalence in sets, independently of targeting strategy. An example of this is the transition from the yellowfin to bigeye clusters in region 3 and 7 (e.g. see Figures 63 and 67). Under the present approach, it is assumed that this transition is caused by targeting and there was no change in the relative abundance of bigeye to yellowfin over the time period. In fact, we suspect that there were actually changes in the relative abundance of these species but with the current clustering method we do not have the means to partition out changes in abundance from changes in targeting. There are at least two approaches that are likely to address this. The preferable approach is to have as much relevant information of gear characteristics to help calibrate the cluster composition—data rescue on HBF and leader material would be the priorities in this regard. A second approach, which we considered briefly in this analysis, is to undertake the clustering on time blocks of the data, e.g., by decade, to recognize that the catch composition for a bigeye target fishery might change over time due to differential changes in the relative abundance of the four species. It is then necessary to link together the clusters from different time periods. These temporal clusters would be used in a similar way as the whole-period clusters are used now, that is, either to isolate/remove particular target behaviour, or when using cluster as a variable in the GLM.

Another potential approach to targeting classification is to consider targeting on a continuous rather than categorical scale (e.g., Winker et al., 2013). This gets rid of the somewhat arbitrary boundaries that is placed around the proportion of species required in a set for it to belong to a certain targeting group. It might also reduce imbalance in some of the current models where cluster factors have little overlap in either time or space. We are still left, however, with the more important problem of partitioning out changes in targeting from changes in species relative abundance, which is why we believe that additional access to gear characteristics would be so valuable, and also why we are exploring the inclusion of oceanographic variables (see below).

While not specifically related to clustering, we also see the routine calculation of a variety of fishery "concentration indices" (e.g., Harley, 2009) could be of value. Quantities such as Gullands Index are calculated on an annual time step and could be of assistance in determining the suitability of using temporally defined clusters.

### 4.5.2 Oceanographic information

The relationship between the abundance and availability of tuna to fishing gears has been the focus of considerable research for many years, including within various CPUE modelling frameworks (Hinton and Nakano, 1996). One careful consideration if using oceanographic data in a CPUE standardization is whether the variable is describing the local abundance of the species or simply the availability of the fish to the fishing gear. We would only want to standardize for the effect of a variable in the second case. As an illustration of this concept, bigeye tuna longline CPUE might be low in areas of low oxygen and/or high thermocline depth, but in the first instance low CPUEs directly reflect abundance, while in the second instance they are a function of lower catchability because bigeye individuals are denser in deeper waters such that there is less overlap between their habitat and the fishing gear (Bigelow and Maunder, 2007).

Notwithstanding the importance that has been given to the inclusion of these variables, our approach here of using 5x5 degree cells and modelling the Pacific with 11 subregions does reduce the influence of this variation. However, we believe that inclusion of a variable such as thermocline depth in association with HBF data for the fishery would be critical for the "regional-weight" analyses that help inform relative abundance across the overall assessment region (McKechnie et al., 2014a), including models of spatial smoothing (e.g., Walsh et al., 2006).

## 4.6 Statistical modelling approaches

The limited resources (time and computing) available for this analysis precluded detailed examination of statistical modelling approaches. We did consider three models with different error assumptions, but none of them fitted particularly well and we would argue that we have not progressed from the work of Hoyle and Okamoto (2011) in improving our modelling in this regard. Alternative error distribution such as the Tweedie are being increasingly applied in CPUE analyses (Candy, 2004; Shono, 2008) and there is value in considering this approach in the future.

We also encountered a specific problem when modelling cluster as a variable using the negative binomial. We found that the expected proportion of zeros and distribution of positive values could sometimes differ significantly among clusters, e.g., comparing a bigeye and albacore cluster in a tropical region. We applied a simulation approach to test the goodness of fit and found that the model would not fit all groups well – instead it fits "through the middle" and this was one of the main motivations for removing data from particular clusters from the analysis. This approach is not ideal as we reduce our coverage in time and space. The flexibility of fitting GLMs in TMB or other languages where the likelihood is manually defined, allows further complexity to the model, and as an example we suggest investigating negative binomial models where the dispersion parameter $\theta$ is specific to the factor (e.g. cluster) over which the model with constant $\theta$ has poor fit.

Future work should also explore the inadequacies of the log normal component of the delta-log-normal models in some regions. These datasets come from an extremely diverse range of fishing activity where CPUE comes for mixture of a large number of distributions determined by factors for which we have no information. It is not surprising that our simple models not not fit all the data well, and our use of the log normal is partly due to its suitability to these situations by calling on the central limit theory, and it is known to be relatively robust to departures in normality when large sample sizes such as ours available. We however see value in exploring the data that are not fitted well in detail, to investigate whether there are reasons why this data may not suitable for inclusion in analyses, or whether there might be modifications to the model that can fit these data better - further covariates, or alternative models with heavy tails etc.

Potential problems relating to "gaps" in the coverage of fishing effort through time has been recognized for some time as a potential source of bias in standardized CPUE indices (Walters, 2003; Campbell, 2004). Some have approached the problem through imputation methods (e.g. Carruthers et al., 2011) while others have considered spatial smoothing techniques (e.g.

He et al., 1997; Tremblay-Boyer et al., 2014). We view the compilation of this data set as providing the best coverage of the region in both time and space and therefore the best opportunity for application of spatial smoothing approaches using tools such as Generalised Additive Models (GAMs), but note that patterns in targeting will require careful attention. Better utilization of HBF data, continuous cluster values, and oceanographic data are all likely to be key to the success of such approaches which may ultimately provide better indices of abundance.

## 4.7 Considerations for the Pacific-wide analysis

The purpose of these analyses is to provide CPUE data for the modelling exercise (McKechnie et al., 2015) aimed at assessing the validity of undertaking a WCPO bigeye tuna assessment in isolation of the EPO (Ianelli et al., 2012). The first observation to make is that the indices produced here are ones that we believe are suitable to make this comparison, but we have not reached the level of analysis that we would consider necessary if these were to support a full formal stock assessment. Therefore we stress that it is the **relative** estimates of depletion from the Pacific-wide population model analyses (i.e., comparisons within the analysis) that should be of attention rather than any absolute estimates of depletion. This point is also stressed by McKechnie et al. (2015).

Due to the problems encountered with the lack of early LLOP data in general, and the absence of individual vessel for the Japanese data which represented most pre-1980 data, we produced CPUE series of two lengths: long indices in which no accounting for fleet characteristics and turnover was possible, and short indices which did account for these factors. The long series can in some respects be viewed as similar to previous aggregated CPUE data analyses, and are not dissimilar to those estimated by Hoyle and Okamoto (2011) and used in the 2011 BET stock assessment. We recommend using the DLN-lg-drm-clst set of models for the 2013 PW BET stock assessment analyses, as they have maximum temporal coverage, and they do not assume that the changes in the proportion of sets in different clusters (e.g. BET *vs.* YFT clusters in regions 3 and 7) are entirely due to targeting practices. We do however recommend that full stock assessments of BET in the future should explore a range of CPUE indices, covering the different assumptions about targeting, and indices of different lengths. If this operational dataset is again available for those assessments then strong emphasis should be placed on obtaining indices that better account for changes in fishing efficiency over the entire stock assessment period.

## 4.8 Need for a simulation study

Many of the fundamental approaches underpinning the CPUE analyses have not been subject to rigorous simulation testing to determine if they are appropriate or robust to departures from model assumptions. Examples could include the ability to account for changes in the distribution of fishing effort, within-region differences in patterns of abundance, appropriateness of cluster as a categorical variable, combining binomial and lognormal components, etc. We strongly recommend that the SC prioritizes two parallel streams of CPUE work: 1) further analysis of operational data based on the recommendations above, and 2) development of a simulation model for testing the appropriateness and robustness of the techniques used therein.

After our consideration of the data set assembled for this analysis, we are better placed to determine the best design for a simulation model and the types of tests that should be undertaken. Simply stated, the data set assembled has multiple instances of confounding (e.g., different targeting in different areas) and imbalance (e.g., fleets entering or leaving the fishery at different times rather than fishing consistently), as well as key explanatory variables that are missing in specific spatial/temporal strata. It should be relatively easy to emulate "real" data given that we have one of the worst surveys ever "designed".

The overall concept is quite simple and would be comprised of the following four components:

**Population abundance model:** Build a spatial model of the abundance of the four species (or however are available for the clustering analysis). In the simplest case, abundance could be uniform in time in space, or all species could follow the same trajectory through time. Added complexity could include differences in the trends in abundance for a given species in different locations and/or differences in population trends across the different species. The most important aspect of this model is that at every time step it is possible to "sum-up" the biomass across the spatial domain to have the true abundance of each species—this is the critical quantity to be compared to CPUE series that comes from the CPUE standardization model described below.

**Fishery model:** A fishery would then deploy effort across the spatial domain. This effort could be uniform or preferentially deployed with respect to one or more target species. Added complexity could include multiple fleets with different targeting and/or spatial effort patterns, temporal changes in targeting, and patterns in overall effort deployed. Targeting differences could be manifested in a species/fleet/time catchability value. A further level of complexity – which seems particularly important given some of our data issues – would be to extend the "fishery model" to a "vessel model", where individual

vessels with a given fishing power come and leave the fleet. We could add trends in the power of a vessel over its fishing "life-time" and/or there could be a relationship between a vessel's power and the time it entered the fishery.

**Sampling error model:** While the step above will produce an expected catch for a given set, it is equally important to have an underlying model that generates variability in catch around the expectation. Here one can use different probability distributions – similar to those used in the underlying GLMs—to determine how sensitive CPUE trends are to a misspecification of the error distribution.

**CPUE standardization model:** The final step is to take the data generated above and go through the steps of calculating the standardized indices. This could include a clustering approach as an intermediate step. One could check the robustness of different error distributions and also the impact of missing explanatory variables or data on the standardization. One could also examine the performance of the standard errors around the abundance indices, in particular whether the temporal variation in the CVs have any relation to the performance of the CPUE model in predicting true abundance levels. Currently estimated CVs relate more to the number of observations, but it might be possible to derive CVs that better reflect the actual observation error.

## 4.9 Conclusions

The months since the assembly of the first full compilation of the available operational data for longlining in the Pacific Ocean has been extremely busy – with many challenges faced and addressed, but also many lessons learned and ideas generated. While this particular analysis is to support a recommendation by Ianelli et al. (2012) to consider the sensitivity of WCPO bigeye status to the inclusion of EPO dynamics, it has generated ideas that should benefit all tuna assessments that depend heavily on longline CPUE. We believe the progress made in this short time has been significant, but the sheer magnitude of the work – including the need to analyse the considerable (and complex) data holdings for the EPO – mean that we do not have definitive indices which we believe would represent all the information that these data have to offer. In this paper we have tried to both report our findings to date and build upon the recommendations of Hoyle (2011) in outlining our vision for continued analyses if these data were to become available again into the future.

# References


Bigelow, K. A. and Hoyle, S. D. (2012). Standardized CPUE for South Pacific albacore. Technical Report WCPFC-SC8-2012/SA-IP-14, Busan, Republic of Korea, 7–15 August 2012.

Bigelow, K. A. and Maunder, M. N. (2007). Does habitat or depth influence catch rates of pelagic species? *Canadian Journal of Fisheries and Aquatic Sciences*, 64:1581–1594.

Campbell, R. A. (2004). CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. *Fisheries Research*, 70:209–227.

Campbell, R. A. (2015). Constructing stock abundance indices from catch and effort data: Some nuts and bolts. *Fisheries Research*, 161:109–130.

Candy, S. G. (2004). Modelling catch and effort data using generalised linear models, the Tweedie distribution, random vessel effects and random stratum-by-year effects. *CCAMLR Science*, 11:59–80.

Carruthers, T. R., Ahrens, R. N., McAllister, M., and Walters, C. J. (2011). Integrating imputation and standardization of catch rate data in the calculation of relative abundance indices. *Fisheries Research*, 109:157–167.

Dunn, K. P. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5:1–10.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielson, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249.

Francis, R. I. C. C. (1999). The impact of correlations in standardised CPUE indices. NZ Fisheries Assessment Research Document 99/42, National Institute of Water and Atmospheric Research. (Unpublished report held in NIWA library, Wellington.).

Harley, S. J. (2009). Spatial distribution measures for the analysis of longline catch and effort data. Technical Report WCPFC-SC5-2009/SA-IP-2, Port Vila, Vanuatu, 10–21 August 2009.

Harley, S. J., Davies, N., Hampton, J., and McKechnie, S. (2014). Stock assessment of bigeye tuna in the Western and Central Pacific Ocean. Technical Report WCPFC-SC10-2014/SA-WP-01, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

He, X., Bigelow, K. A., and Boggs, C. H. (1997). Cluster analysis of longline sets and fishing strategies within the hawaii-based fishery. *Fisheries Research*, 31:144–158.

Hinton, M. G. and Nakano, H. (1996). Standardizing catch and effort statistics using physiological, ecological, or behavioural constraints and environmental data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from the Japanese longline fisheries in the Pacific. *Inter-American Tropical Tuna Commission Bulletin*, 21:171–200.

Hoyle, S. D. (2011). Research outline for longline catch per unit effort data. Technical Report WCPFC-SC7-2011/SA-IP-07, Pohnpei, Federated States of Micronesia, 9–17 August 2011.

Hoyle, S. D., Langley, A. D., and Campbell, R. A. (2014). Recommended approaches for standardizing CPUE data from pelagic fisheries. Technical Report WCPFC-SC10-2014/SA-IP-10, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Hoyle, S. D. and Okamoto, H. (2011). Analyses of Japanese longline operational catch and effort for Bigeye and Yellowfin Tuna in the WCPO. Technical Report WCPFC-SC7-2011/SA-IP-01, Pohnpei, Federated States of Micronesia, 9–17 August 2011.

Ianelli, J., Maunder, M. N., and Punt, A. E. (2012). Independent review of the 2011 WCPO bigeye tuna assessment. Technical Report WCPFC-SC8-2012/SA-WP-01, Busan, Republic of Korea, 7–15 August 2012.

Kristensen, K., Thygesen, U. H., Andersen, K. H., and Beyer, J. E. (2014). Estimating spatio-temporal dynamics of size-structured populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 71:326–336.

Maunder, M. N. and Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, 70(2):141–159.

McKechnie, S., Hampton, J., Abascal, F., Davies, N., and Harley, S. J. (2015). Sensitivity of WCPO stock assessment results to the inclusion of EPO dynamics within a Pacific-wide analysis. Technical Report WCPFC-SC11-2015/SA-WP-03, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

McKechnie, S., Harley, S. J., Davies, N., Rice, J., Hampton, J., and Berger, A. (2014a). Basis for regional structures used in the 2014 tropical tuna assessments, including regional weights. Technical Report WCPFC-SC10-2014/SA-IP-02, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

McKechnie, S., Harley, S. J. Chang, S.-K., Liu, H.-I., and Yuan, T.-L. (2014b). Analysis of longline catch per unit effort data for bigeye and yellowfin tunas. Technical Report WCPFC-SC10/SA-IP-03, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

OFP (2015). Report of the workshop on operational longline data. Technical Report WCPFC-SC11-2015/SA-IP-01, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

Shono, H. (2008). Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fisheries Research*, 93:154–162. doi:10.1016/j.fishres.2008.03.006.

Tremblay-Boyer, L., Harley, S. J., and Pilling, G. M. (2014). Relationship between abundance and range size in longline target species. Technical Report WCPFC-SC10-2014/MI-WP-06, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Walsh, W., Howell, E., Bigelow, K., and McCracken, M. (2006). Analyses of observed longline catches of blue marlin, *Makaira nigricans*, using generalized additive models with operational and environmental predictors. *Bulletin of Marine Science*, 79(3):607–622.

Walters, C. (2003). Folly and fantasy in the analysis of spatial catch rate data. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(12):1433–1436.

Winker, H., Kerwath, S. E., and Attwood, C. G. (2013). Comparison of two approaches to standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line fishery. *Fisheries Research*, 139:118–131.

Table 4: Table of deviance explained of the full DLN-lg-drm-full model using the formula $1 - D_f/D_n$, where $D_f$ and $D_n$ are the residual deviances for the focal and null models, respectively, for both the binomal and log normal models. "No cell" and "no cluster" percentage loss show the percentage point decrease in deviance explained when the cell and cluster variables are removed from the model, in turn. Note that the full model for region 11 does not have a cluster variable as only one cluster remains in the DLN-lg-drm models for this region, and hence there are NA's in the "no cluster" row. The other NA's indicate where unstable estimates of residual deviance were encountered for the binomial model – this indicates that these models are unstable when these predictor variables are not included in the model.

| | | Region | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Family | Metric | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 11 | 12 |
| **Binomial** | Full model % expl | 0.27 | 0.30 | 0.12 | 0.16 | 0.13 | 0.09 | 0.14 | 0.27 | 0.86 | 0.16 | 0.28 |
| | No cell % loss | NA | 7.1 | 1.0 | 1.6 | 5.7 | NA | 1.5 | 1.2 | NA | 6.3 | NA |
| | No cluster % loss | 2.2 | 5.4 | 3.0 | 4.7 | 0.0 | 3.4 | 3.5 | 2.0 | NA | NA | NA |
| **LN** | Full model % expl | 0.18 | 0.29 | 0.14 | 0.19 | 0.21 | 0.23 | 0.15 | 0.21 | 0.39 | 0.15 | 0.23 |
| | No cell % loss | 2.5 | 2.9 | 1.8 | 1.3 | 7.4 | 7.3 | 2.3 | 0.4 | 3.8 | 1.6 | 3.3 |
| | No cluster % loss | 1.8 | 5.8 | 3.7 | 4.7 | 0.3 | 3.5 | 3.8 | 5.0 | 3.9 | NA | 3.6 |

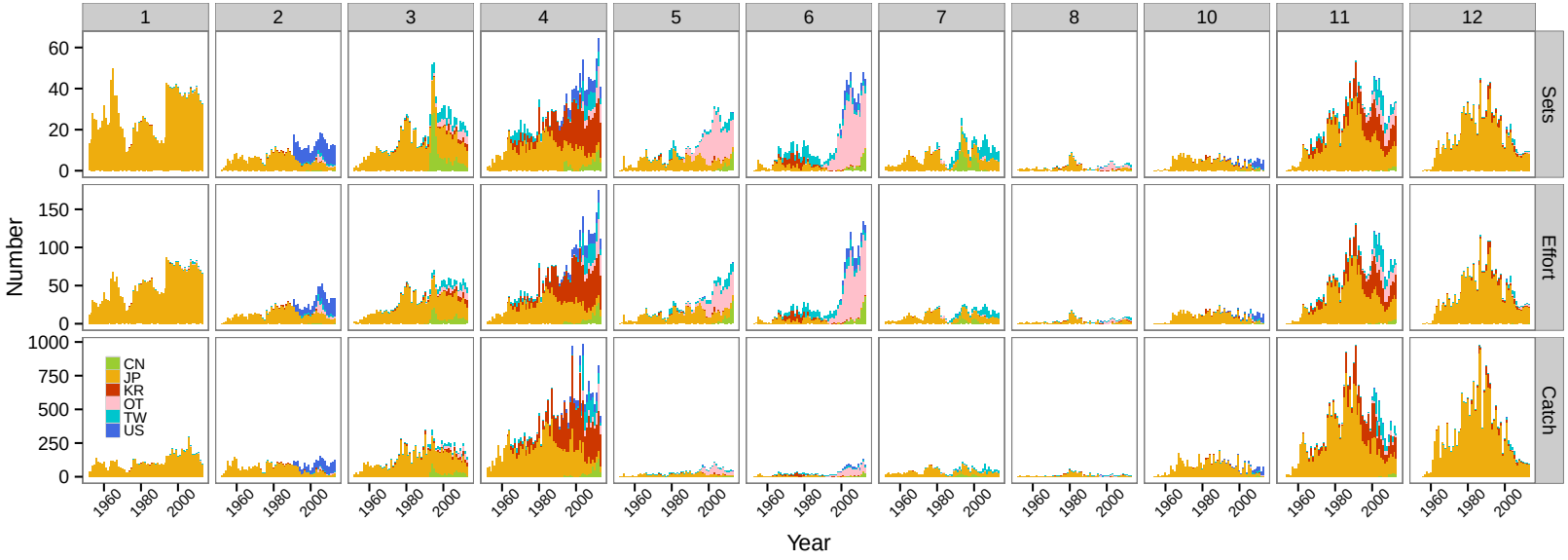


Figure 2: The number of sets, effort (hooks) and catch-in-numbers of BET by region and flag available before filtering.

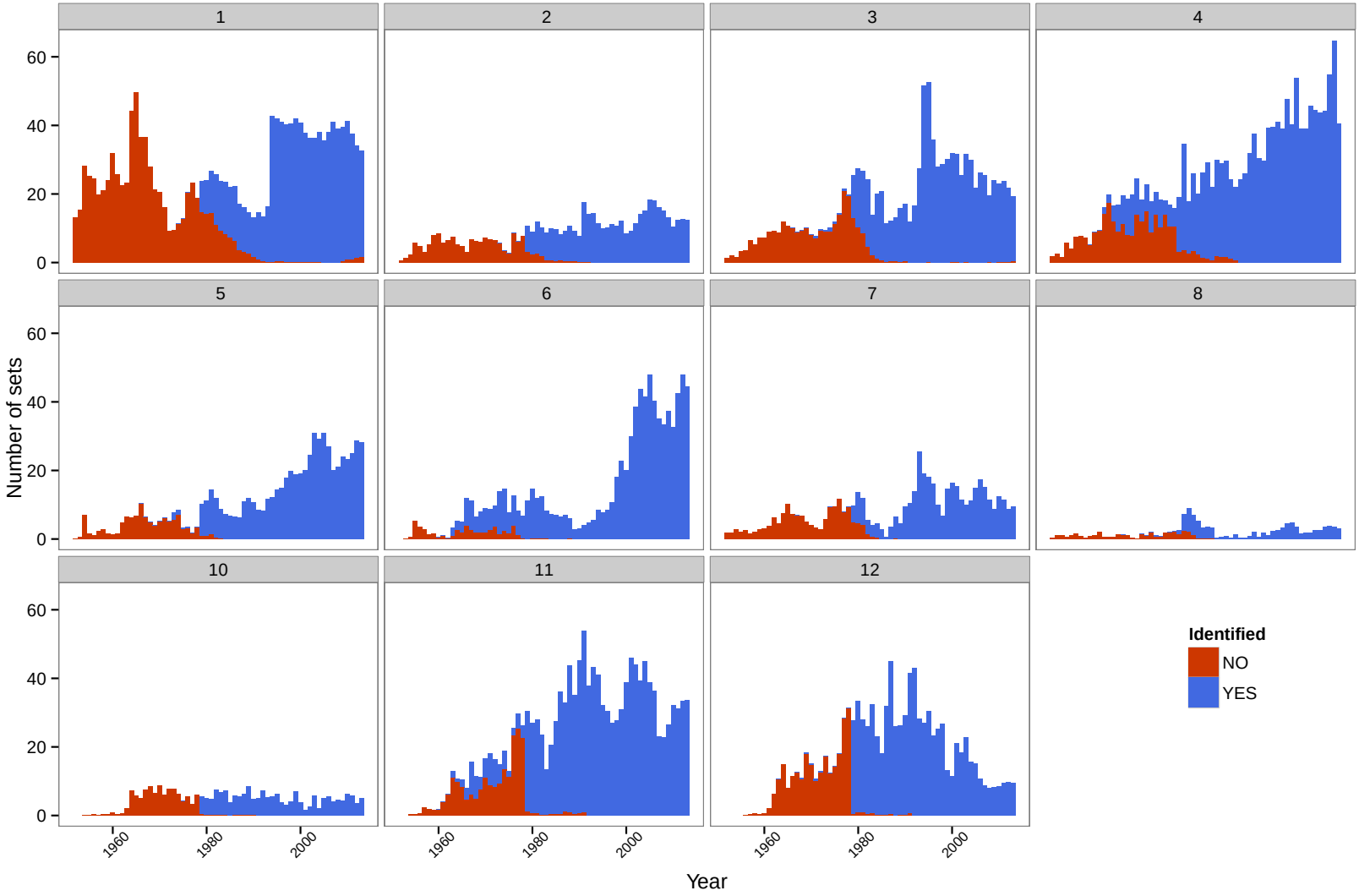


Figure 3: The number of sets for which the individual vessel undertaking the fishing could or could not be indentified, by region.

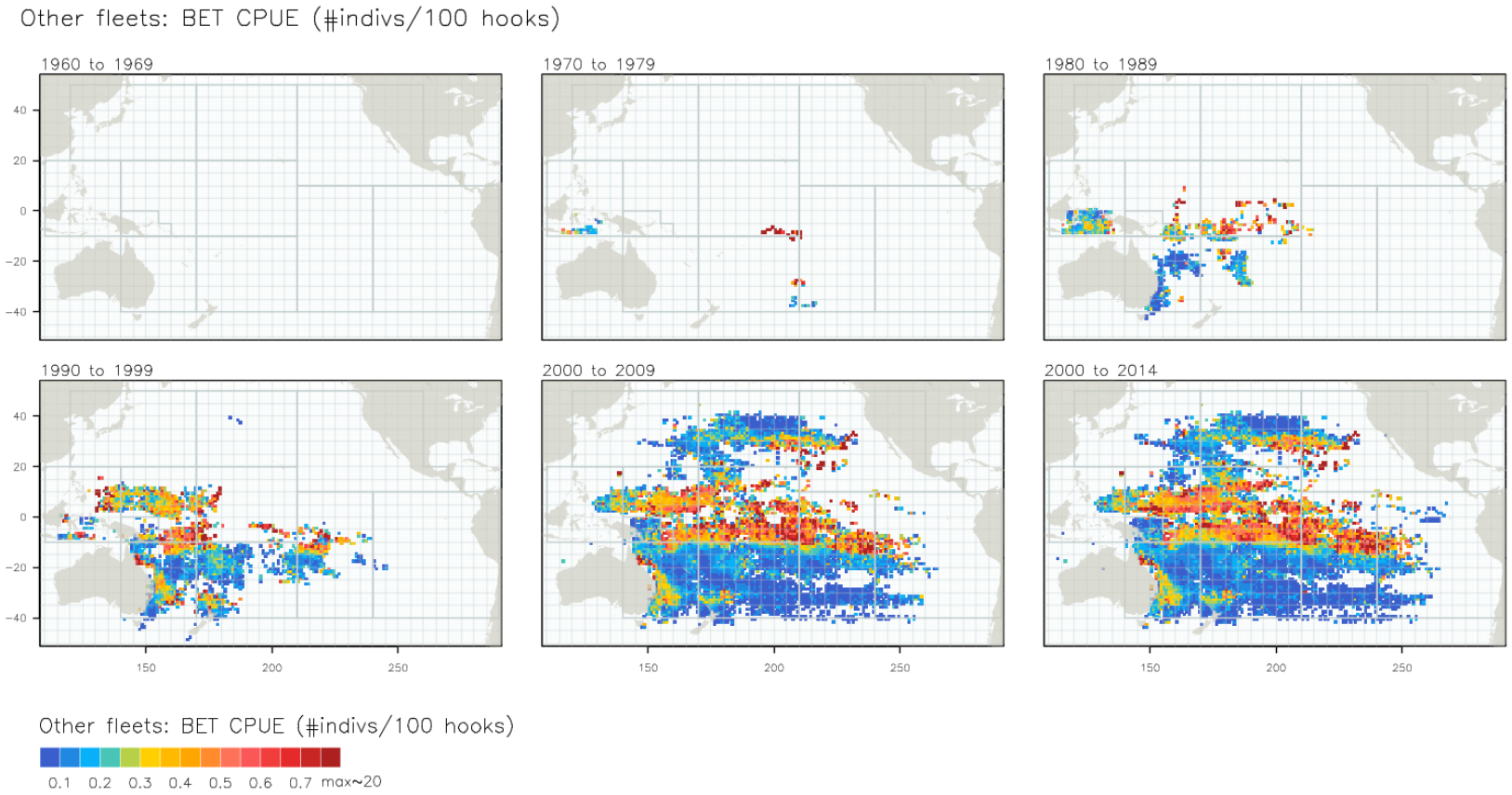


Figure 4: Aggregated bigeye tuna CPUE (in individuals per hundred hooks) by decade for the "other" fleet (non-DWFNs) from the operational data set.

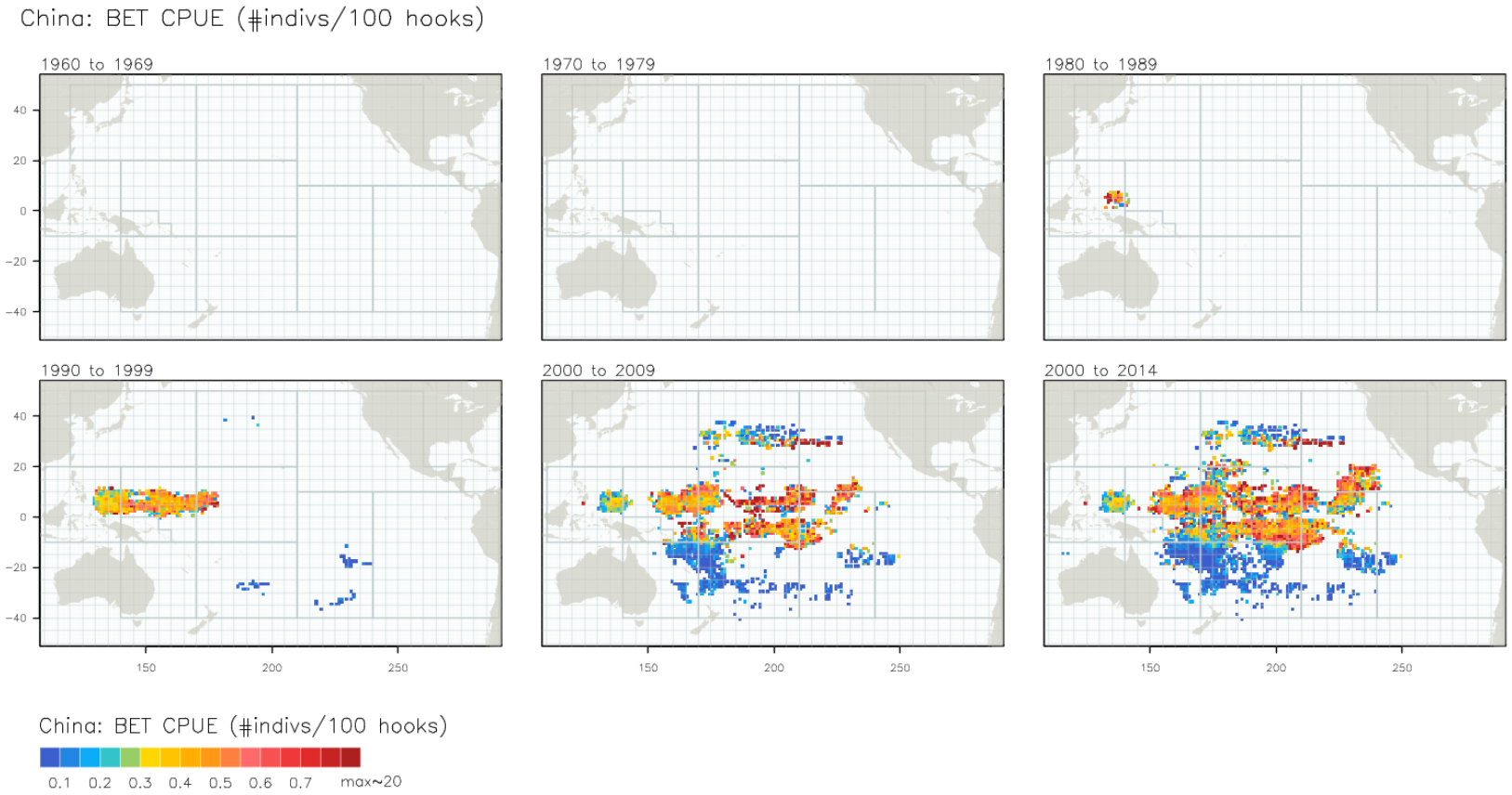


Figure 5: Aggregated bigeye tuna CPUE (in individuals per hundred hooks) by decade for the Chinese fleet from the operational data set.

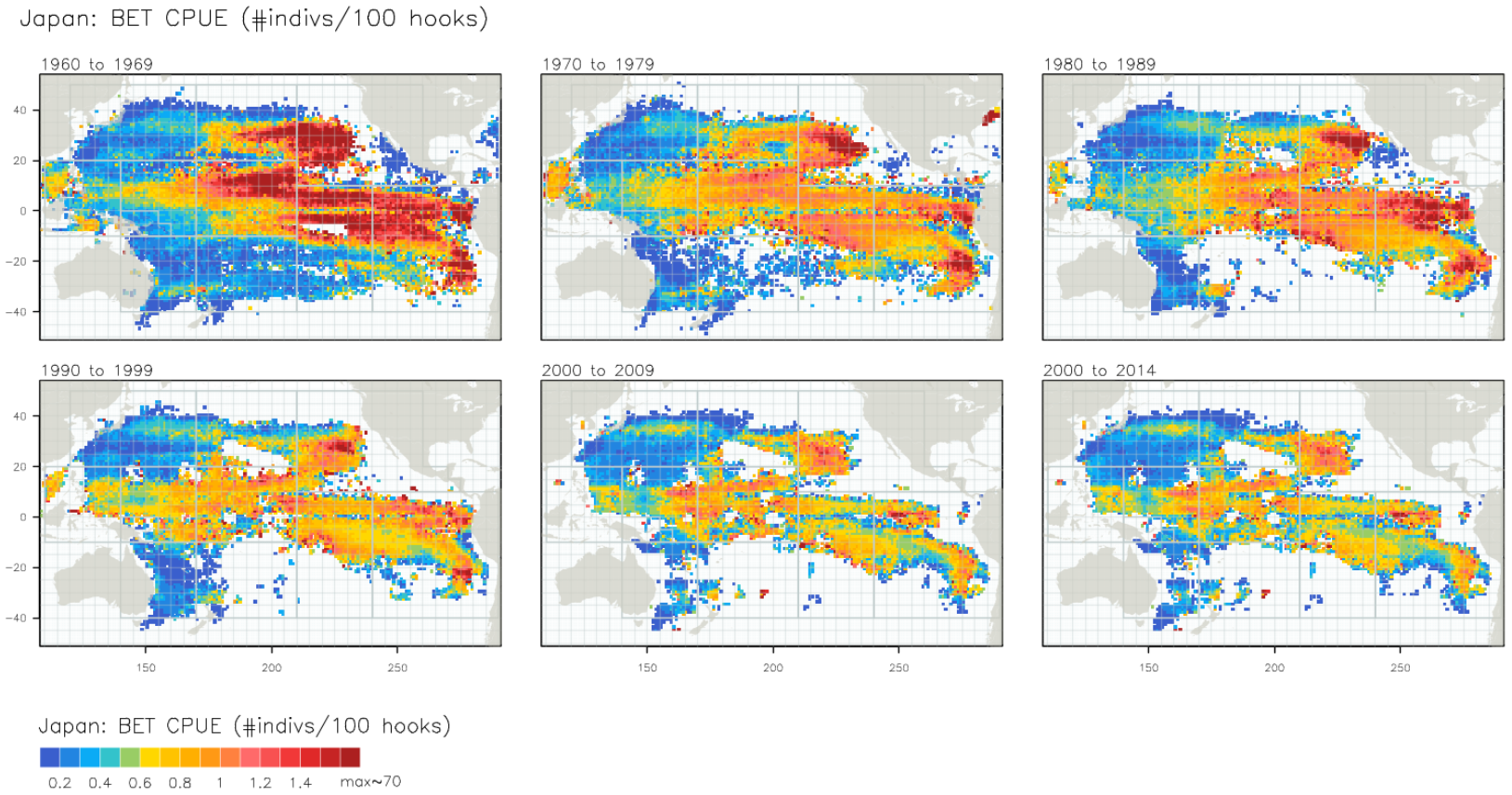


Figure 6: Aggregated bigeye tuna CPUE (in individuals per hundred hooks) by decade for the Japanese fleet from the operational data set.

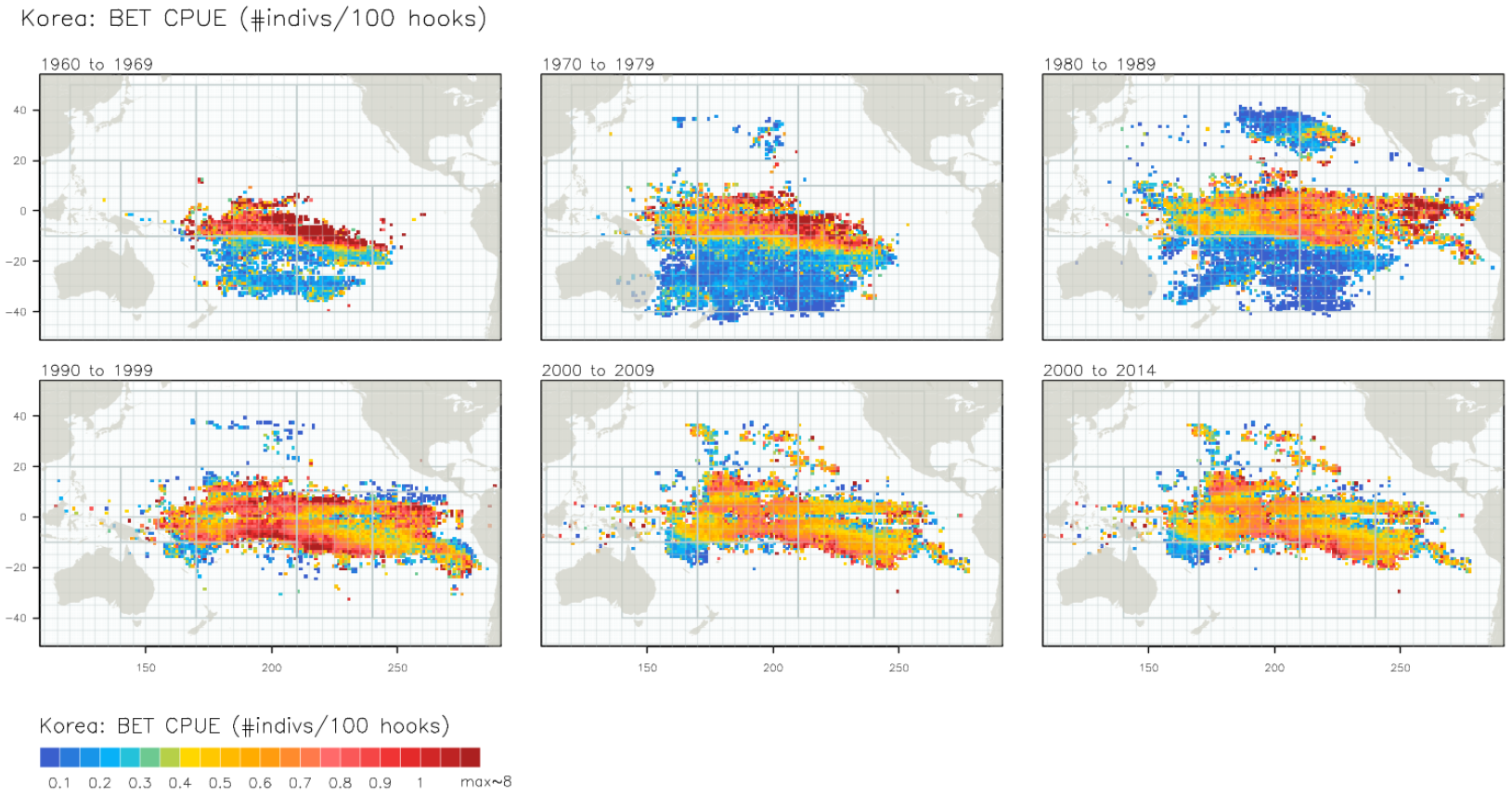


Figure 7: Aggregated bigeye tuna CPUE (in individuals per hundred hooks) by decade for the Korean fleet from the operational data set.

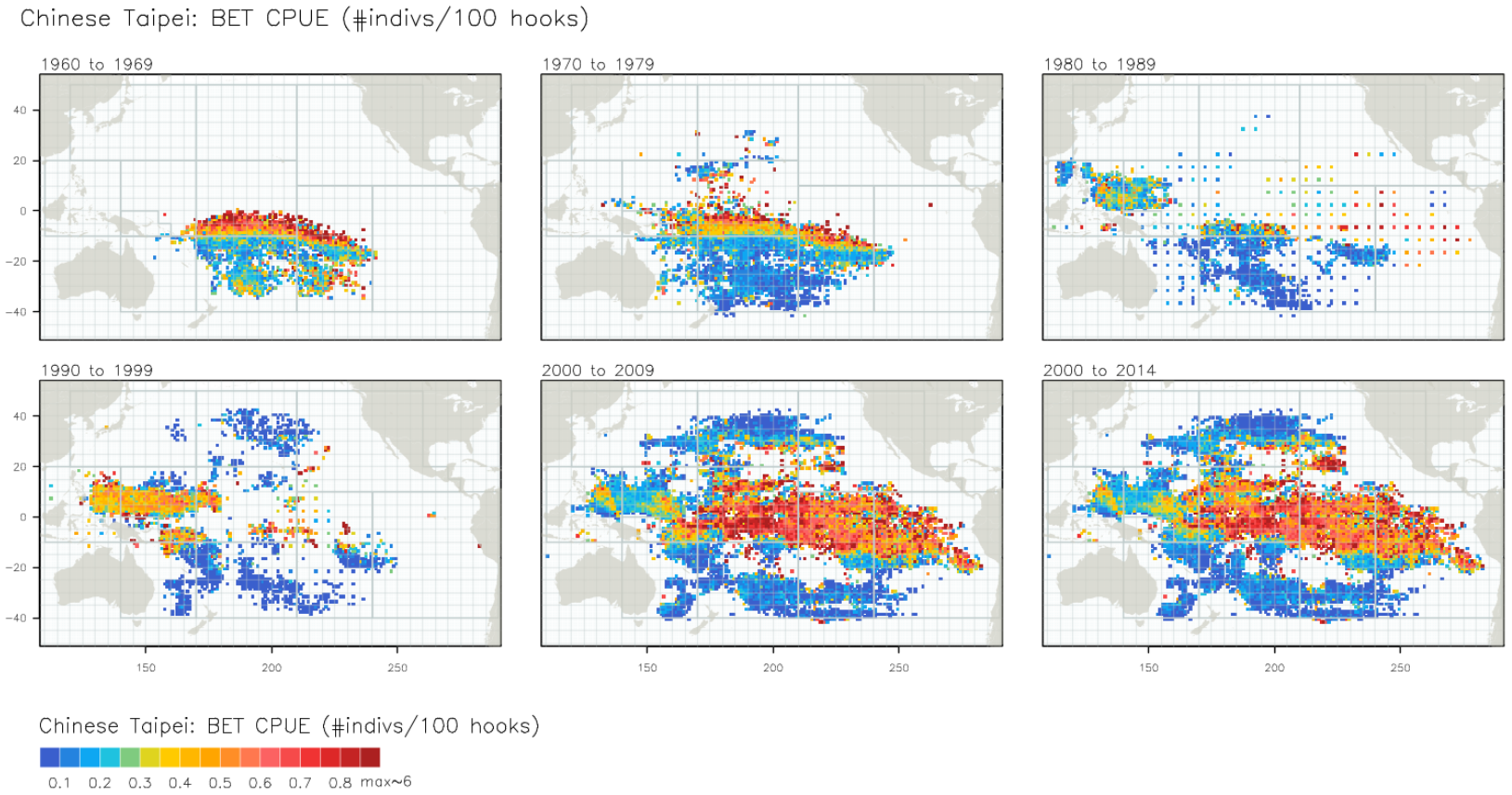


Figure 8: Aggregated bigeye tuna CPUE (in individuals per hundred hooks) by decade for the Chinese-Taipei fleet from the operational data set.

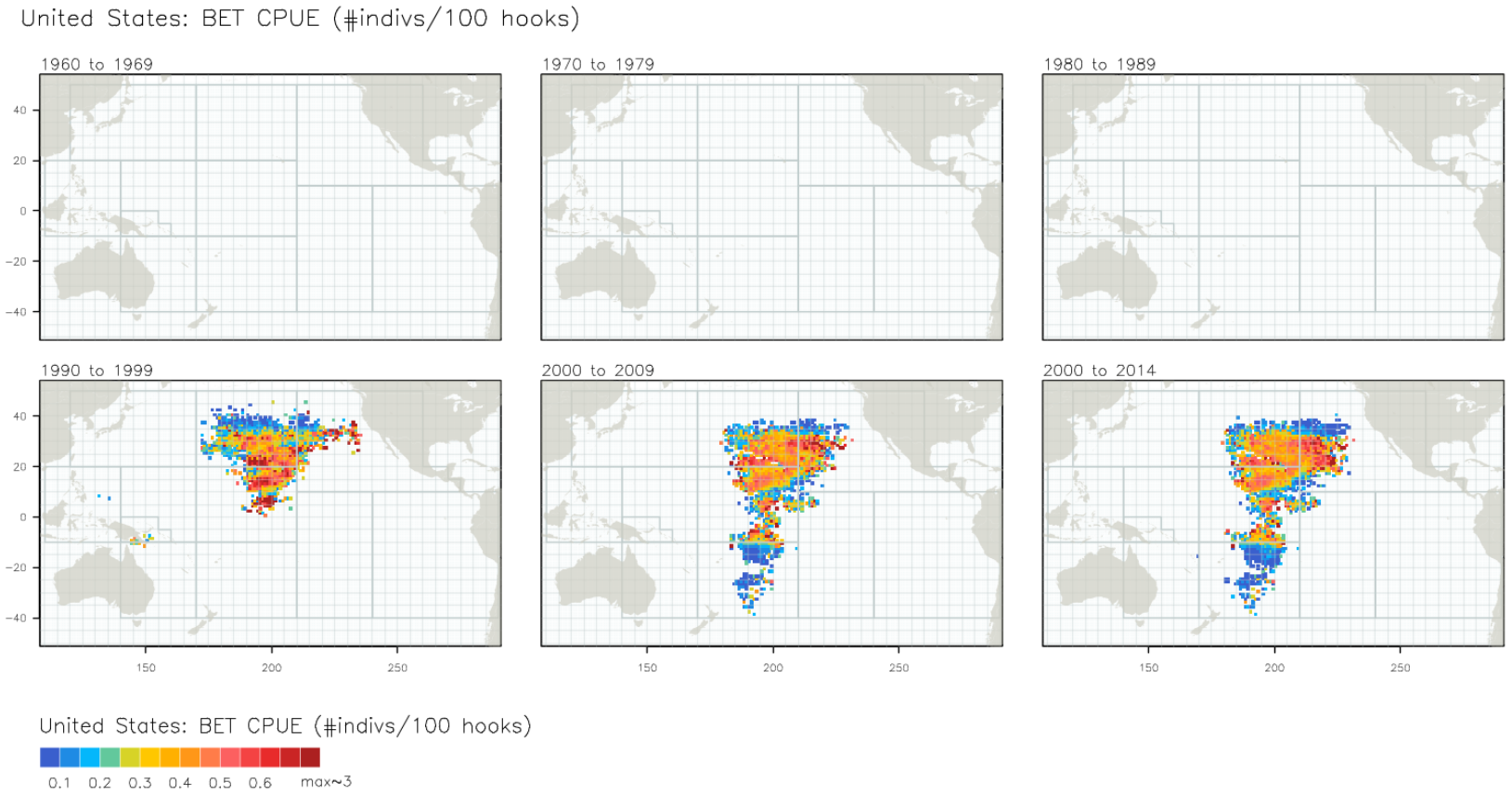


Figure 9: Aggregated bigeye tuna CPUE (in individuals per hundred hooks) by decade for the United States fleet from the operational data set.

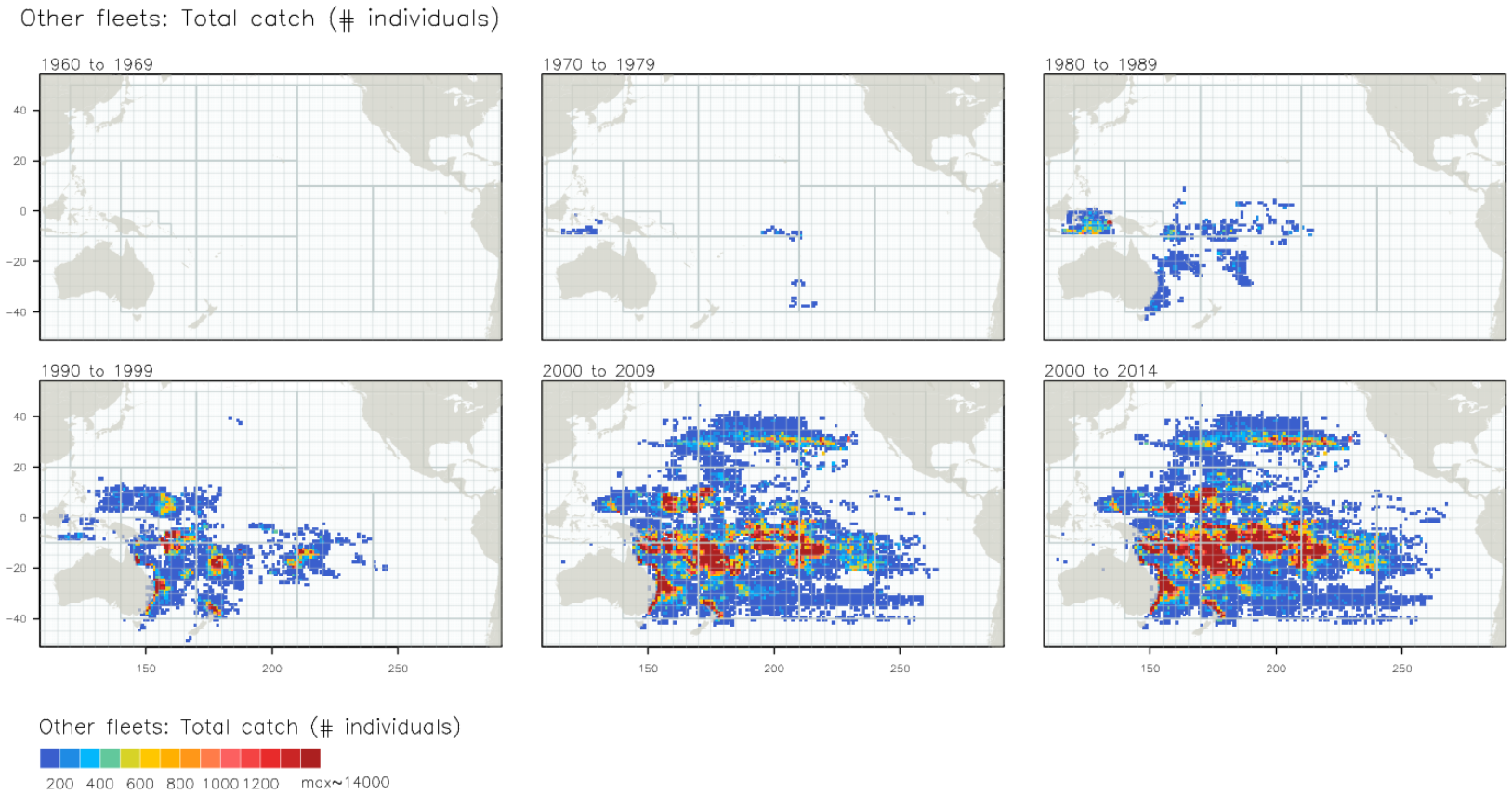


Figure 10: Bigeye catch (numbers of fish) by decade for the "other" fleet (non-DWFNs) from the operational data set.

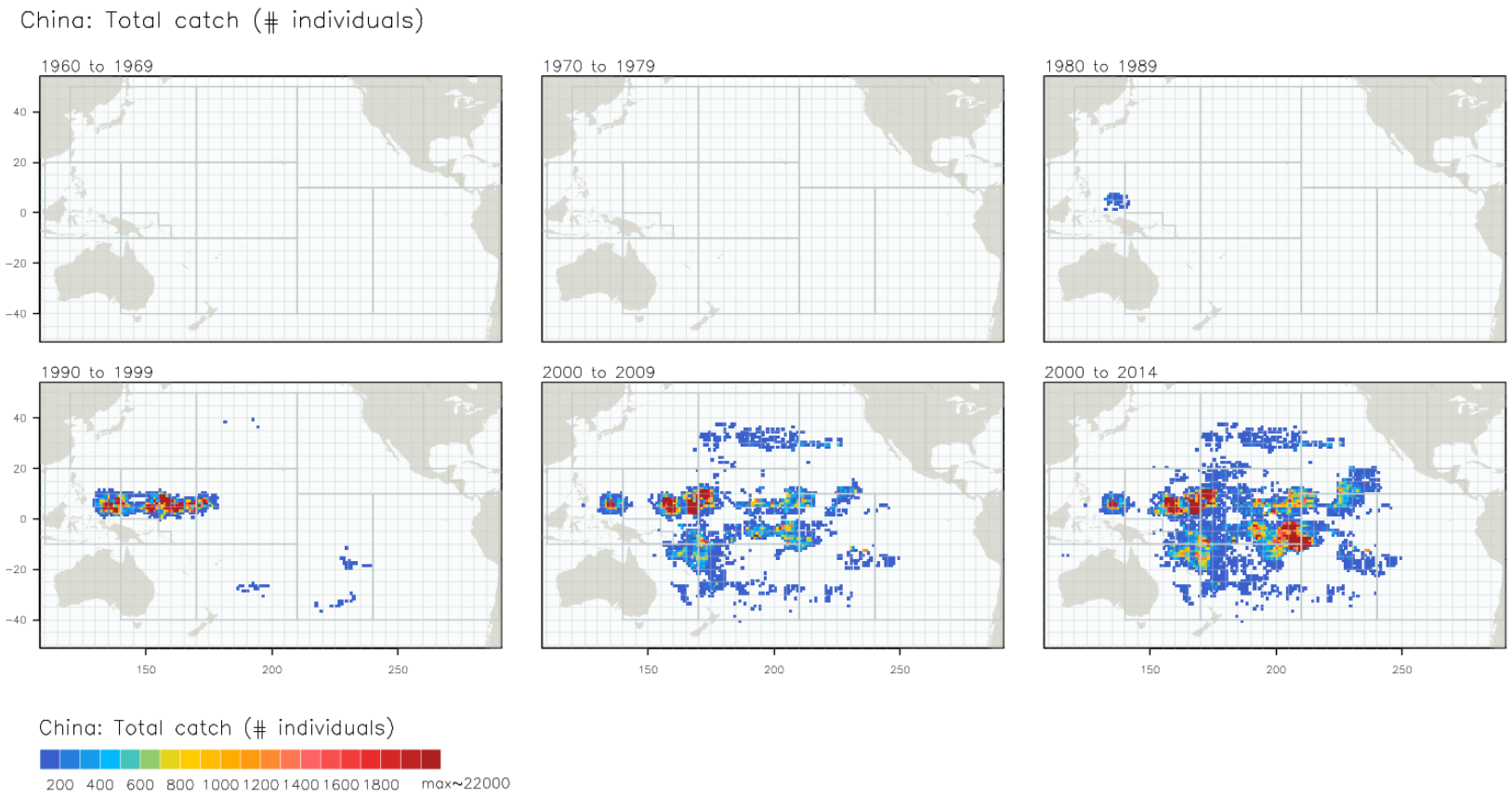


Figure 11: Bigeye catch (numbers of fish) by decade for the Chinese fleet from the operational data set.

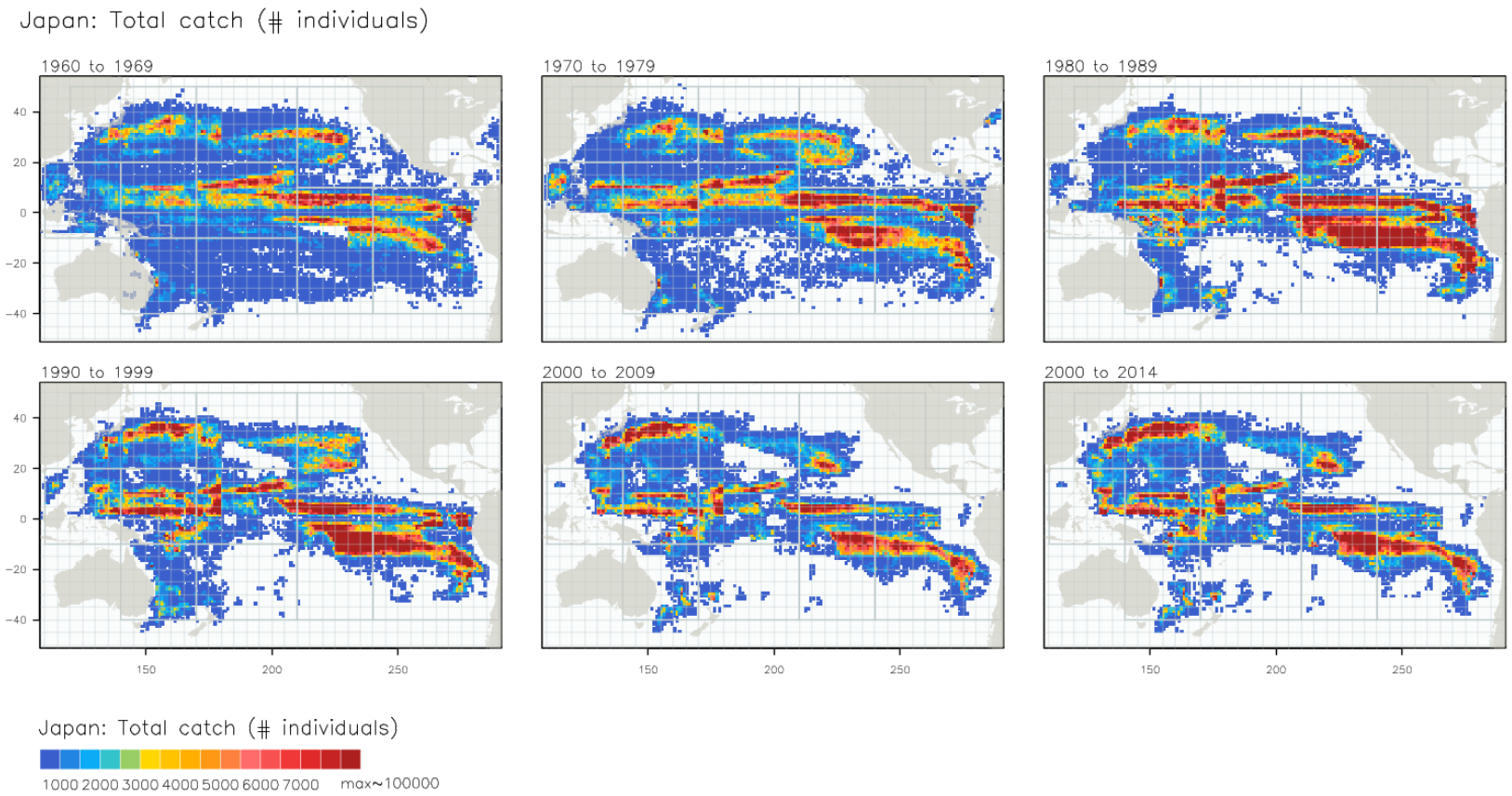


Figure 12: Bigeye catch (numbers of fish) by decade for the Japanese fleet from the operational data set.

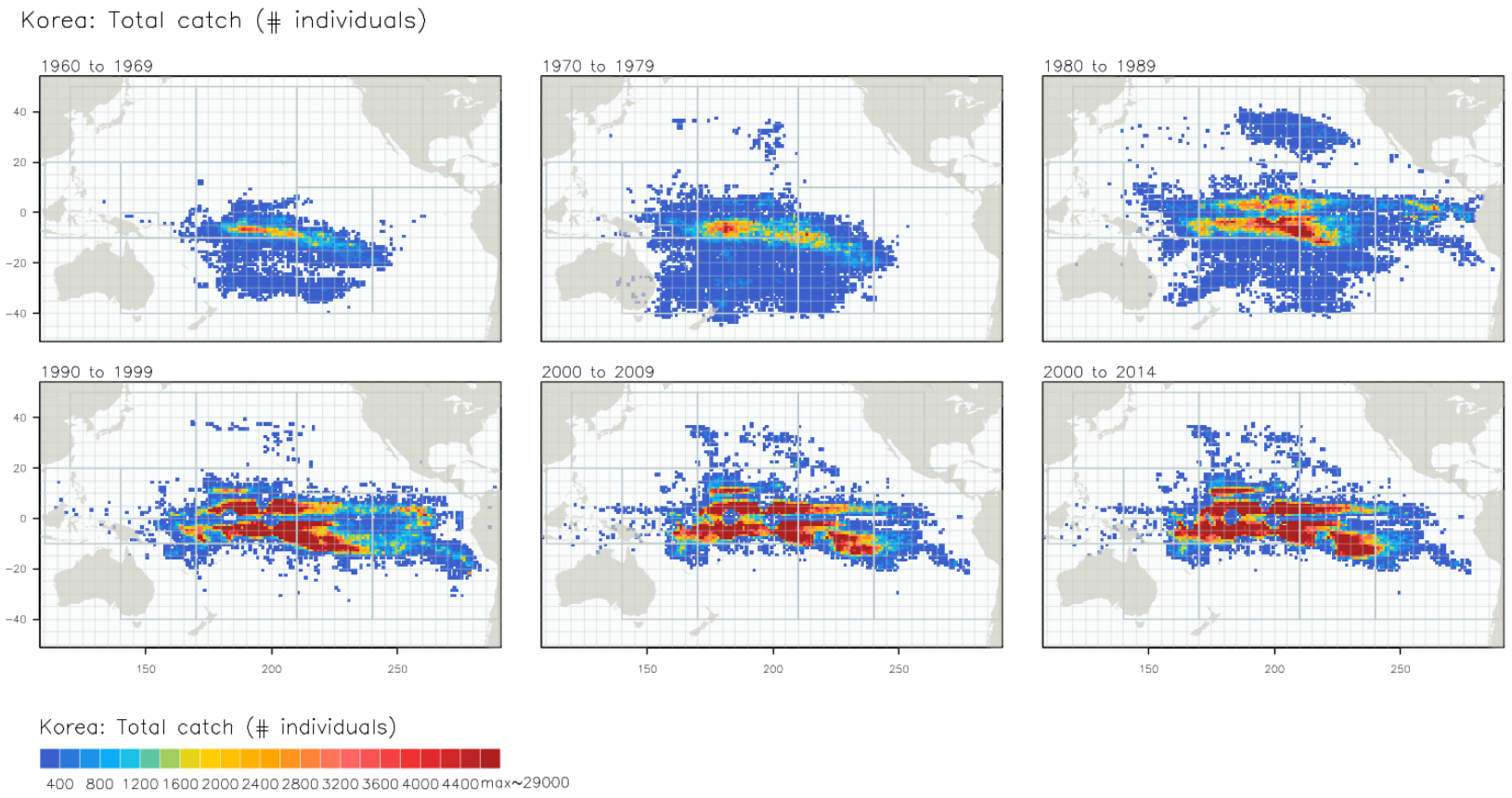


Figure 13: Bigeye catch (numbers of fish) by decade for the Korean fleet from the operational data set.

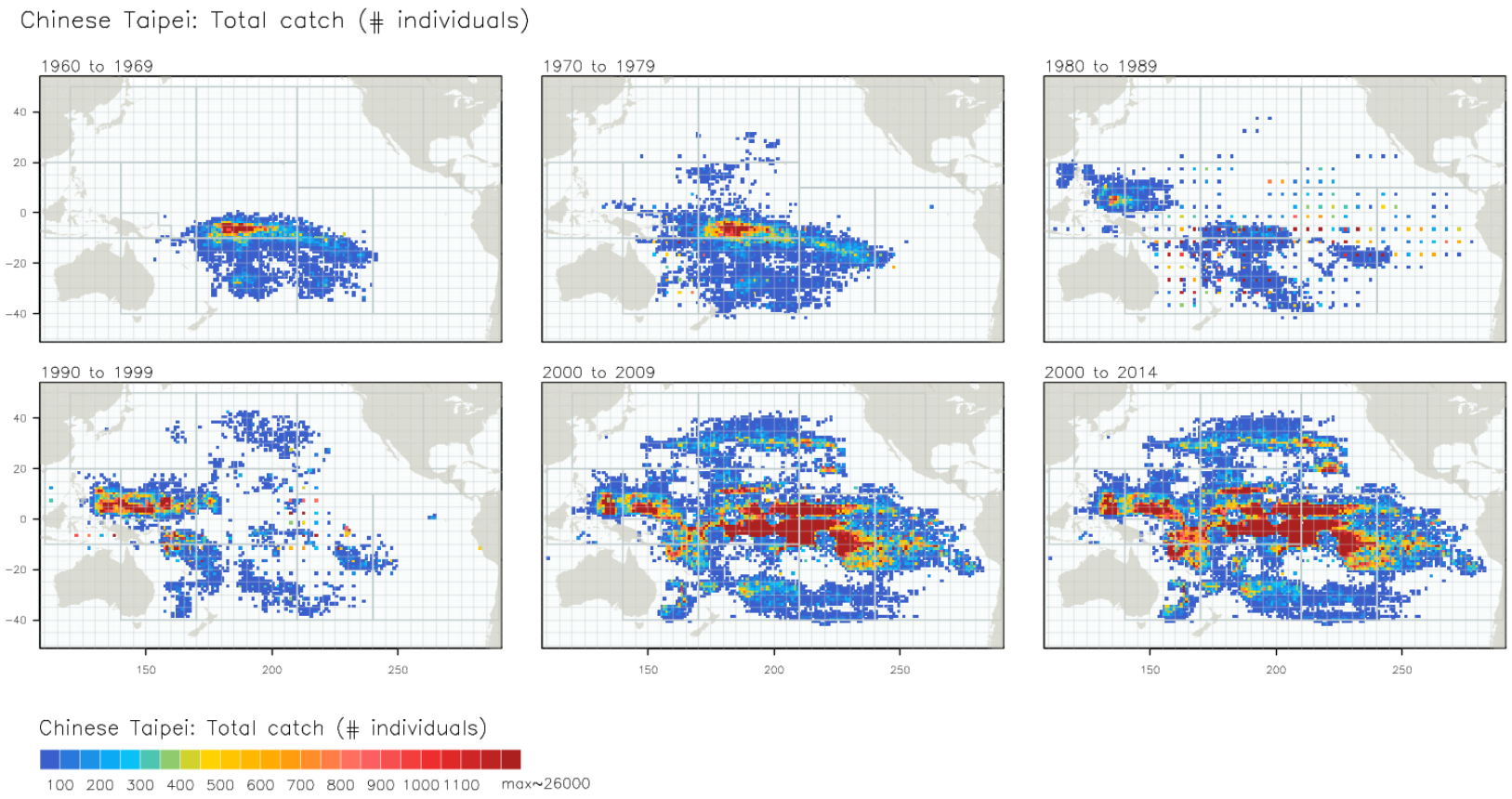


Figure 14: Bigeye catch (numbers of fish) by decade for the Chinese-Taipei fleet from the operational data set.

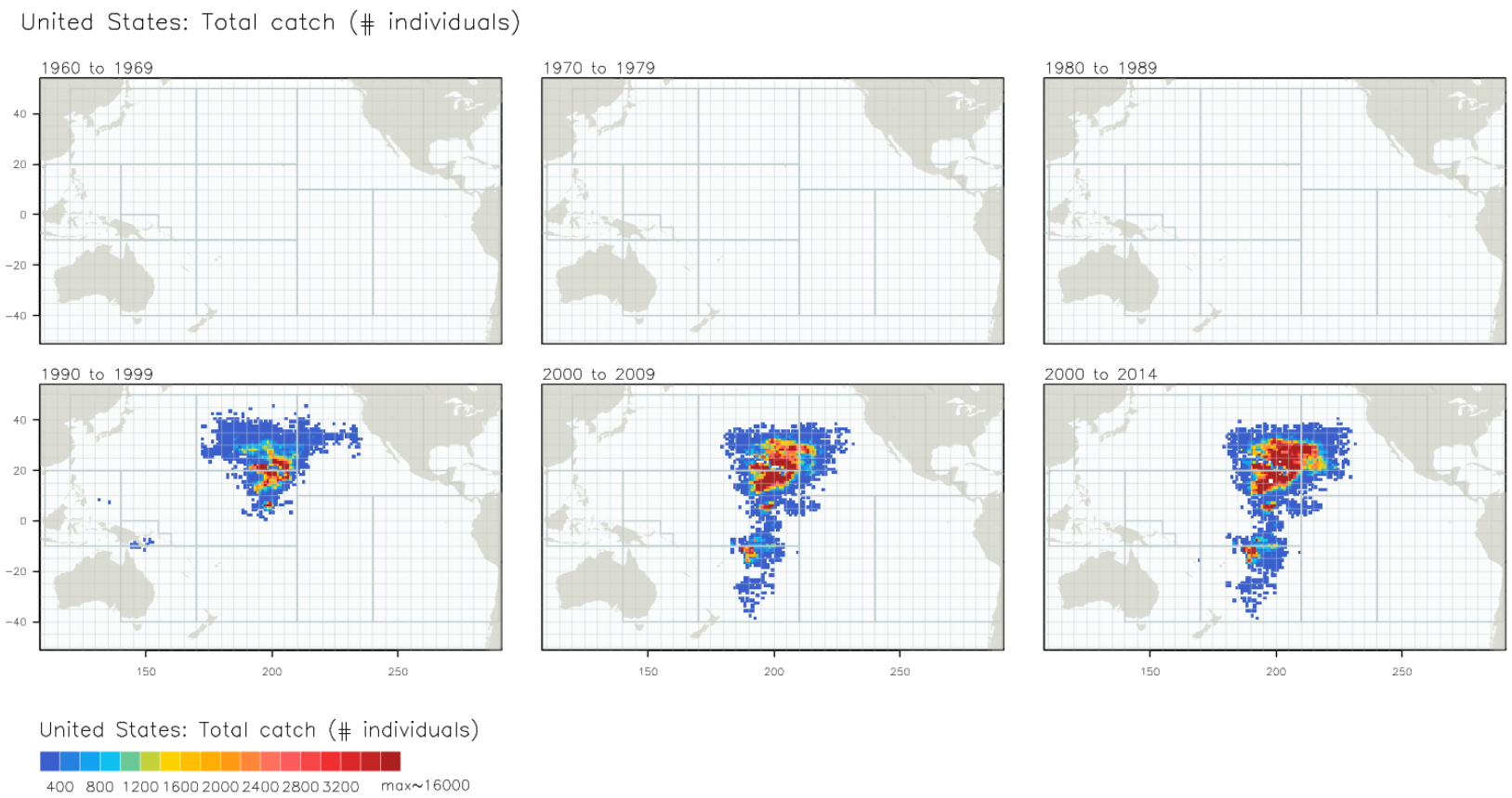


Figure 15: Bigeye catch (numbers of fish) by decade for the United States fleet from the operational data set.

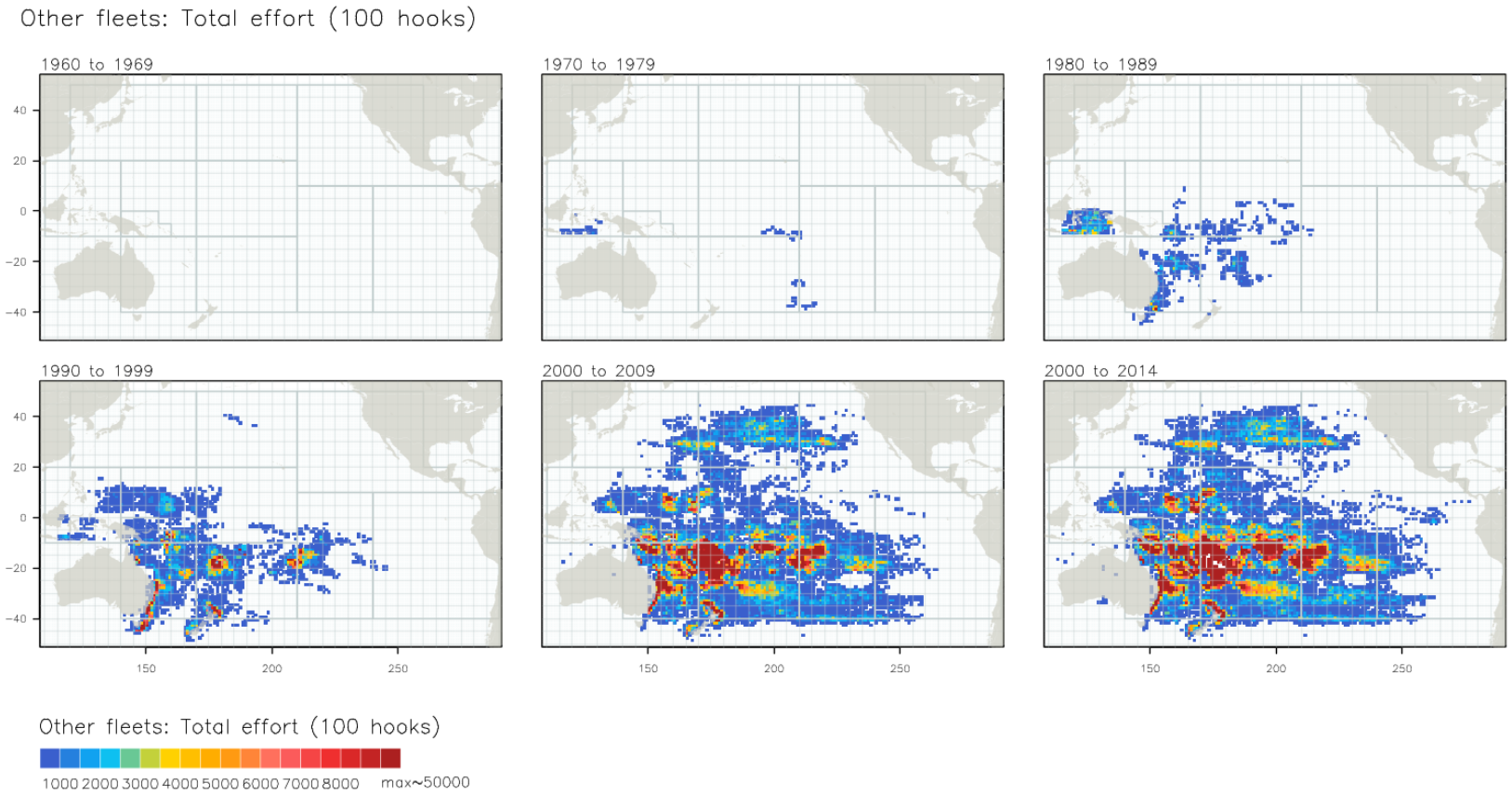


Figure 16: Longline effort (in hundred hooks) by decade for the "other" fleet (non-DWFNs) from the operational data set.

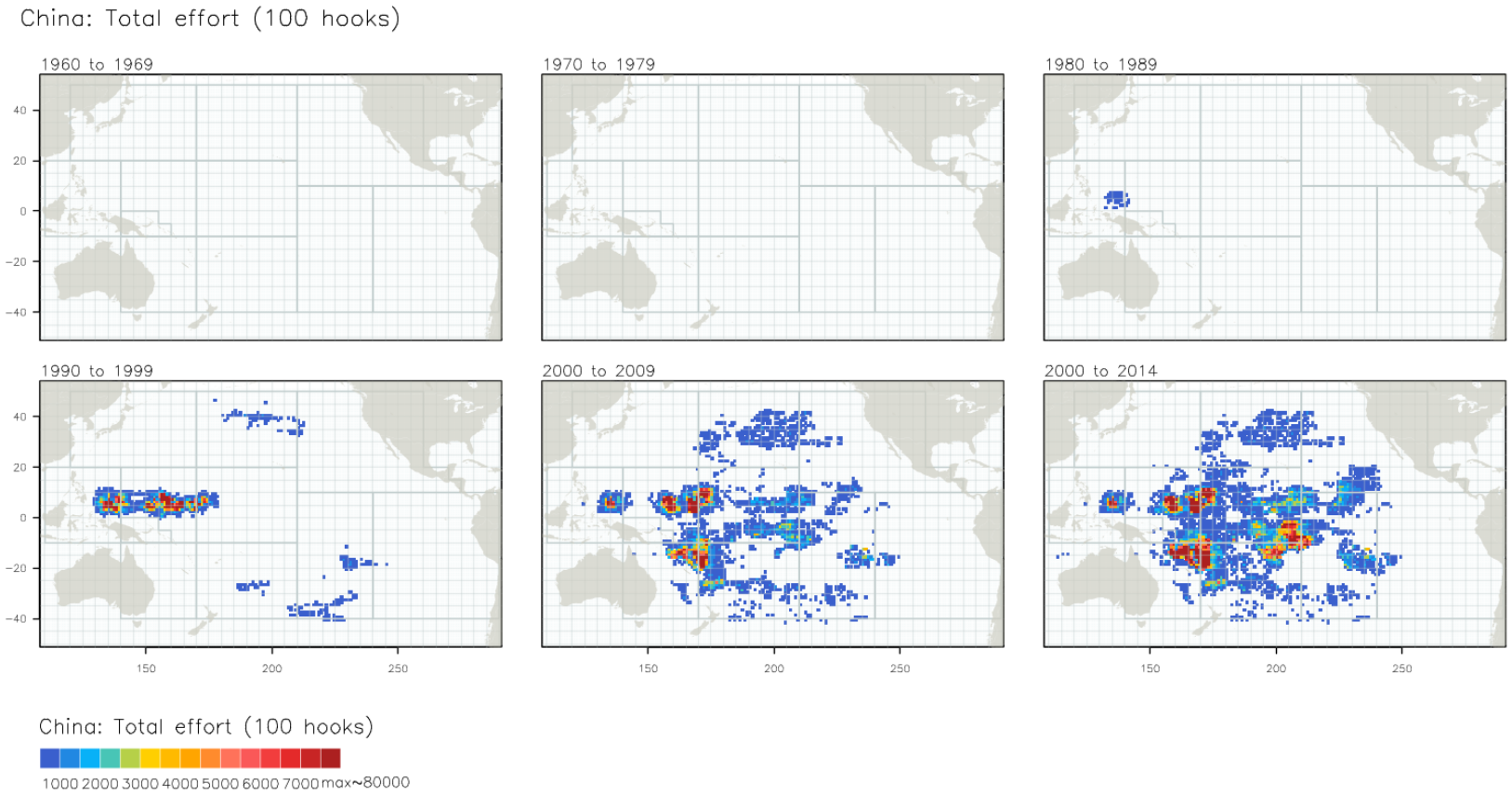


Figure 17: Longline effort (in hundred hooks) by decade for the Chinese fleet from the operational data set.

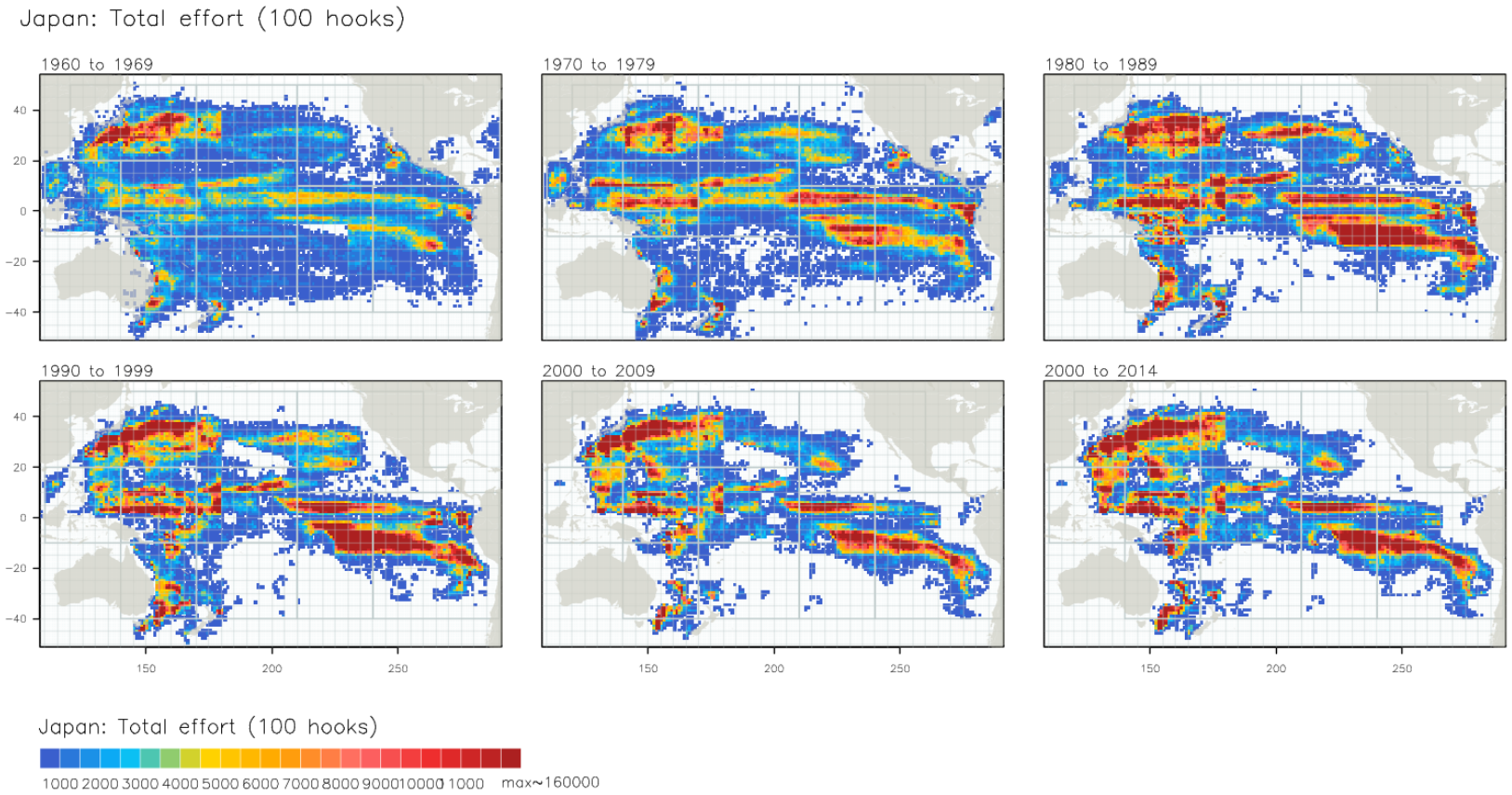


Figure 18: Longline effort (in hundred hooks) by decade for the Japanese fleet from the operational data set.

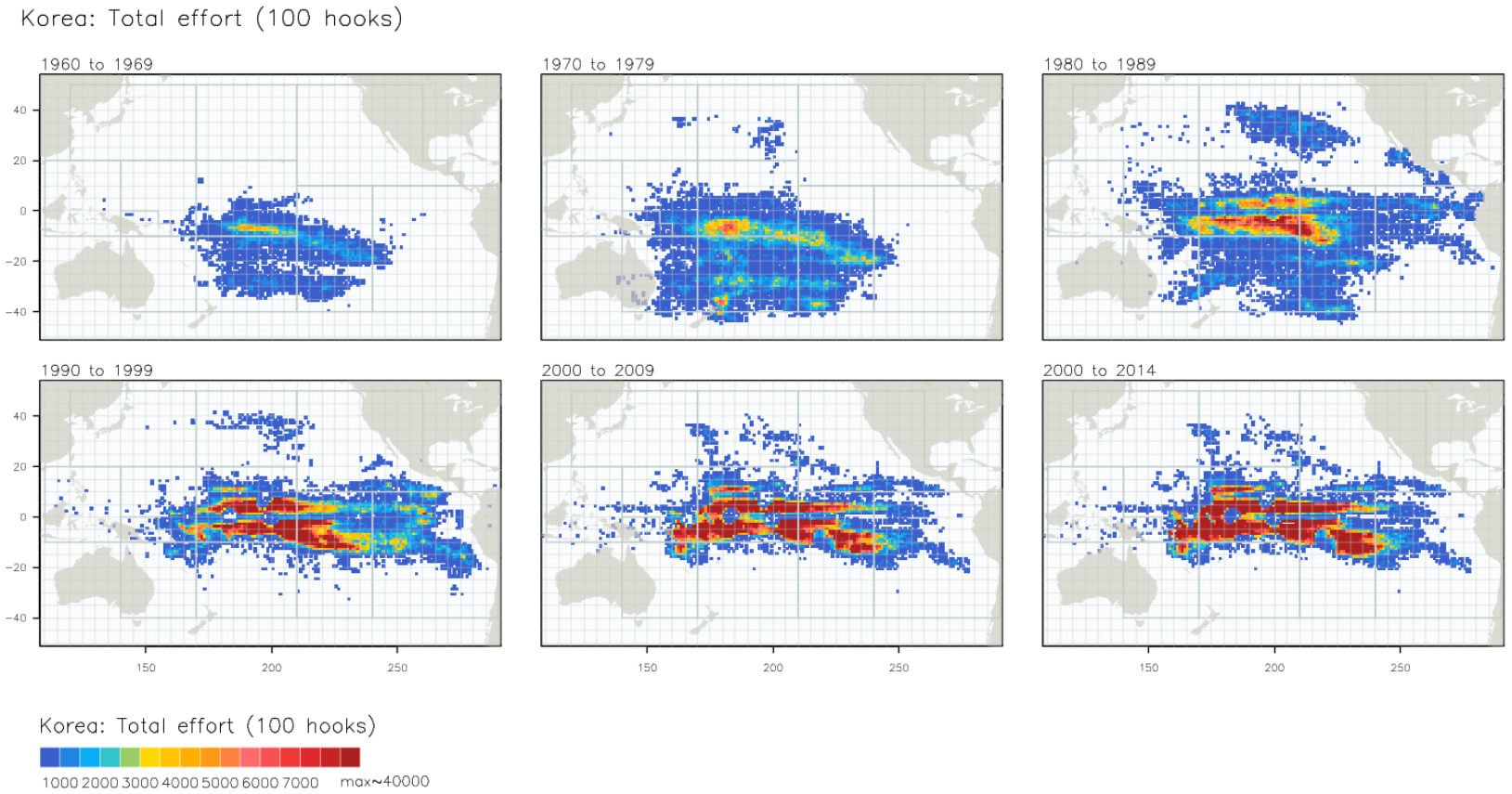


Figure 19: Longline effort (in hundred hooks) by decade for the Korean fleet from the operational data set.

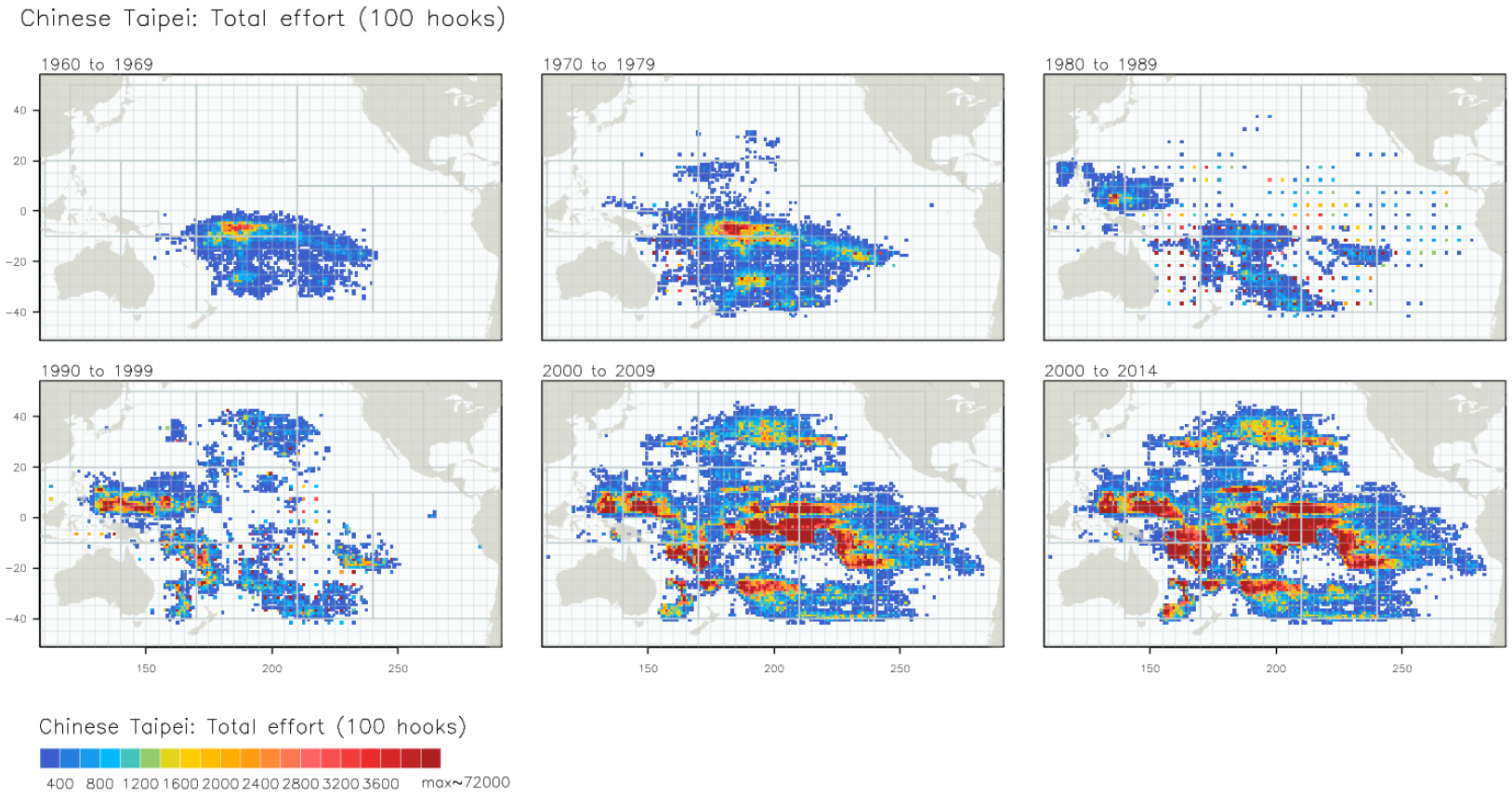


Figure 20: Longline effort (in hundred hooks) by decade for the Chinese-Taipei fleet from the operational data set.

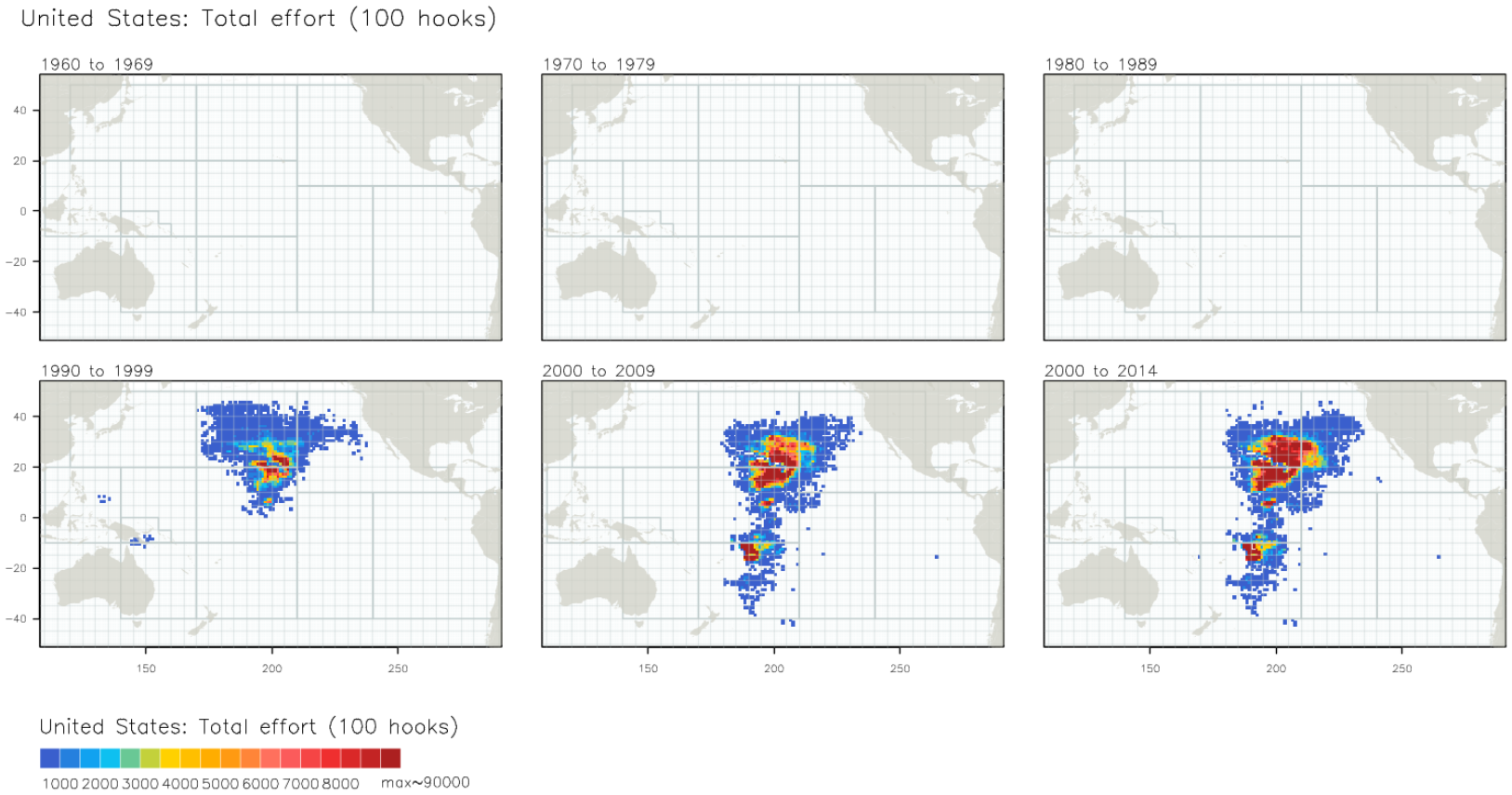


Figure 21: Longline effort (in hundred hooks) by decade for the United States fleet from the operational data set.

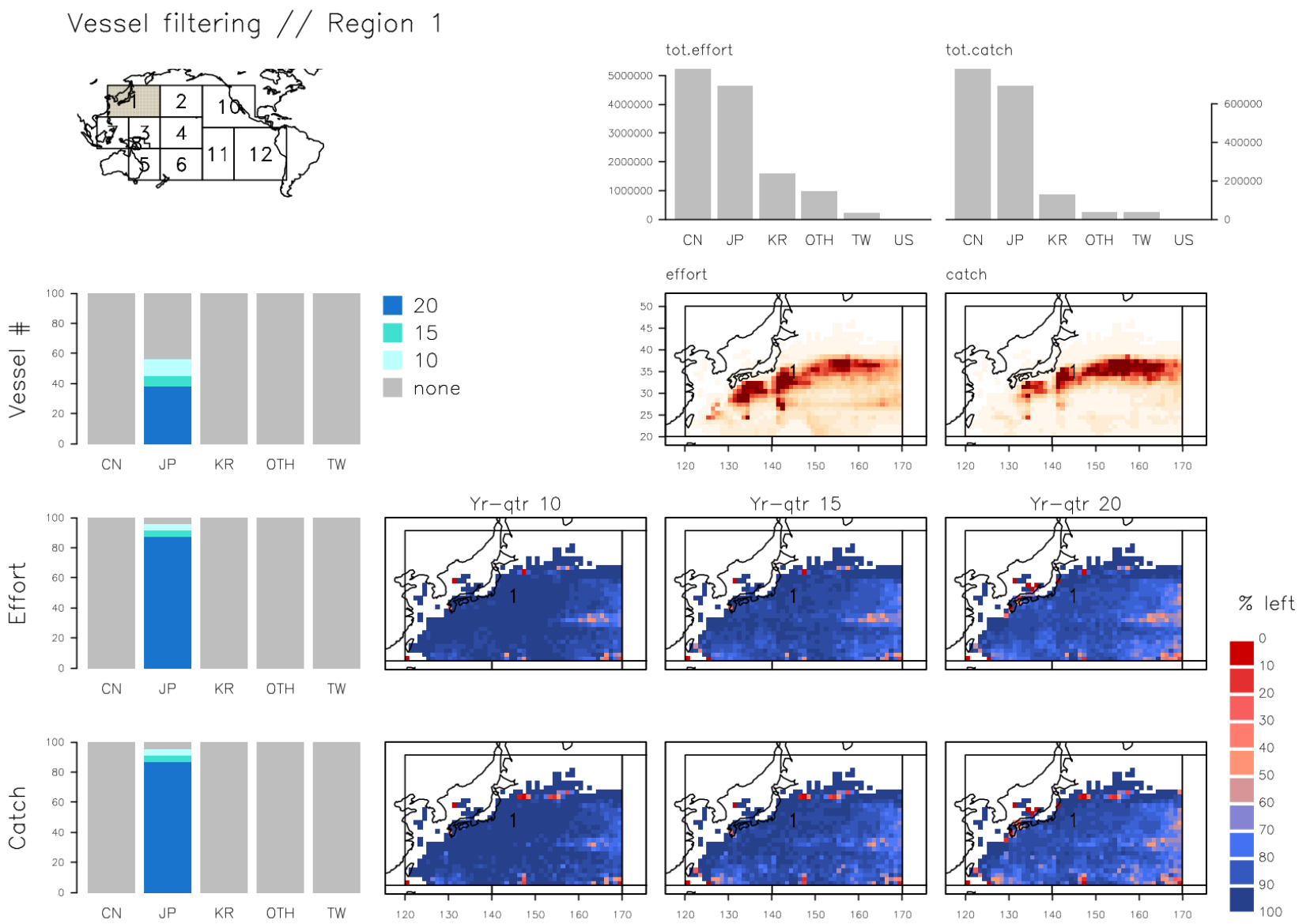


Figure 22: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

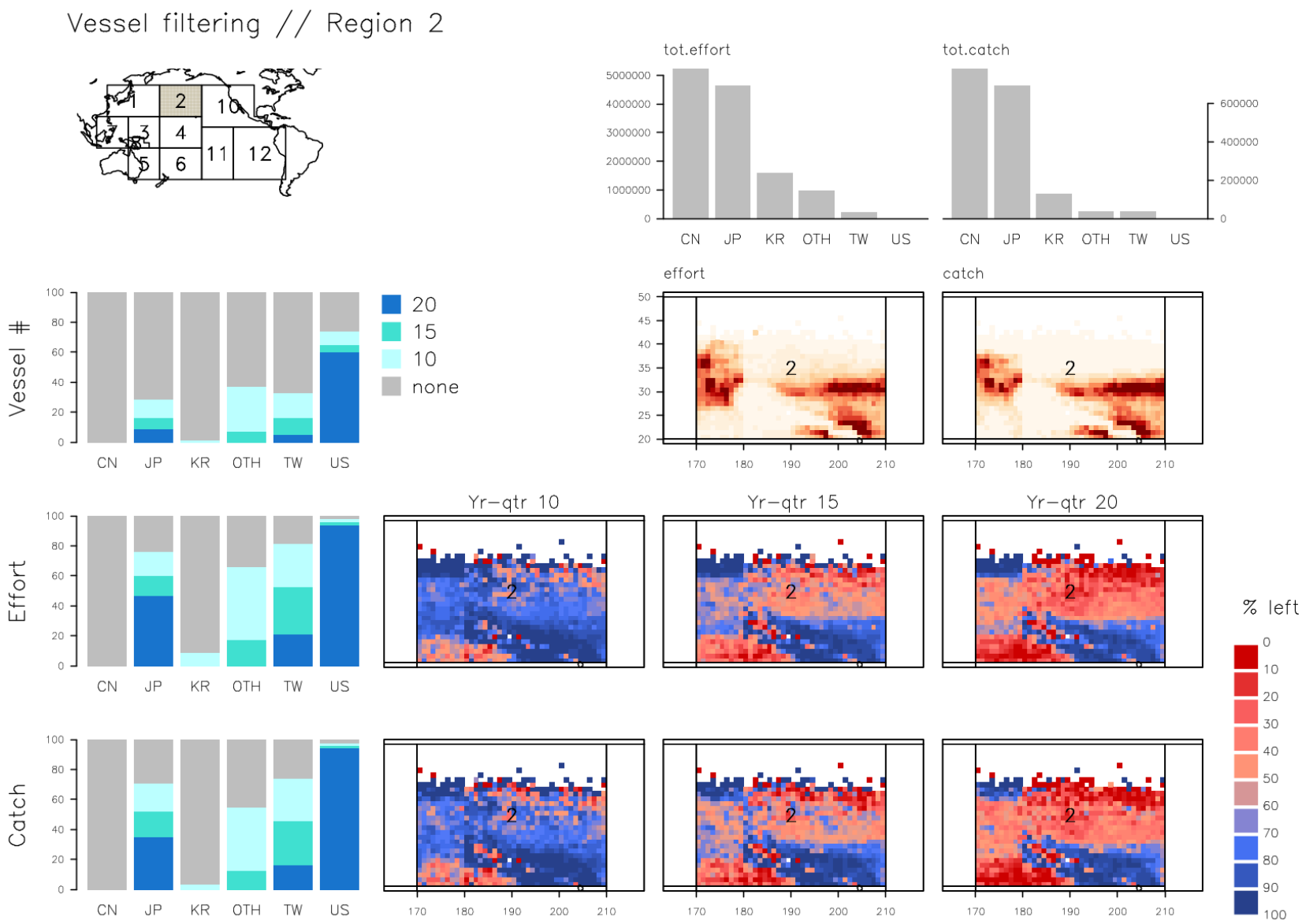


Figure 23: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

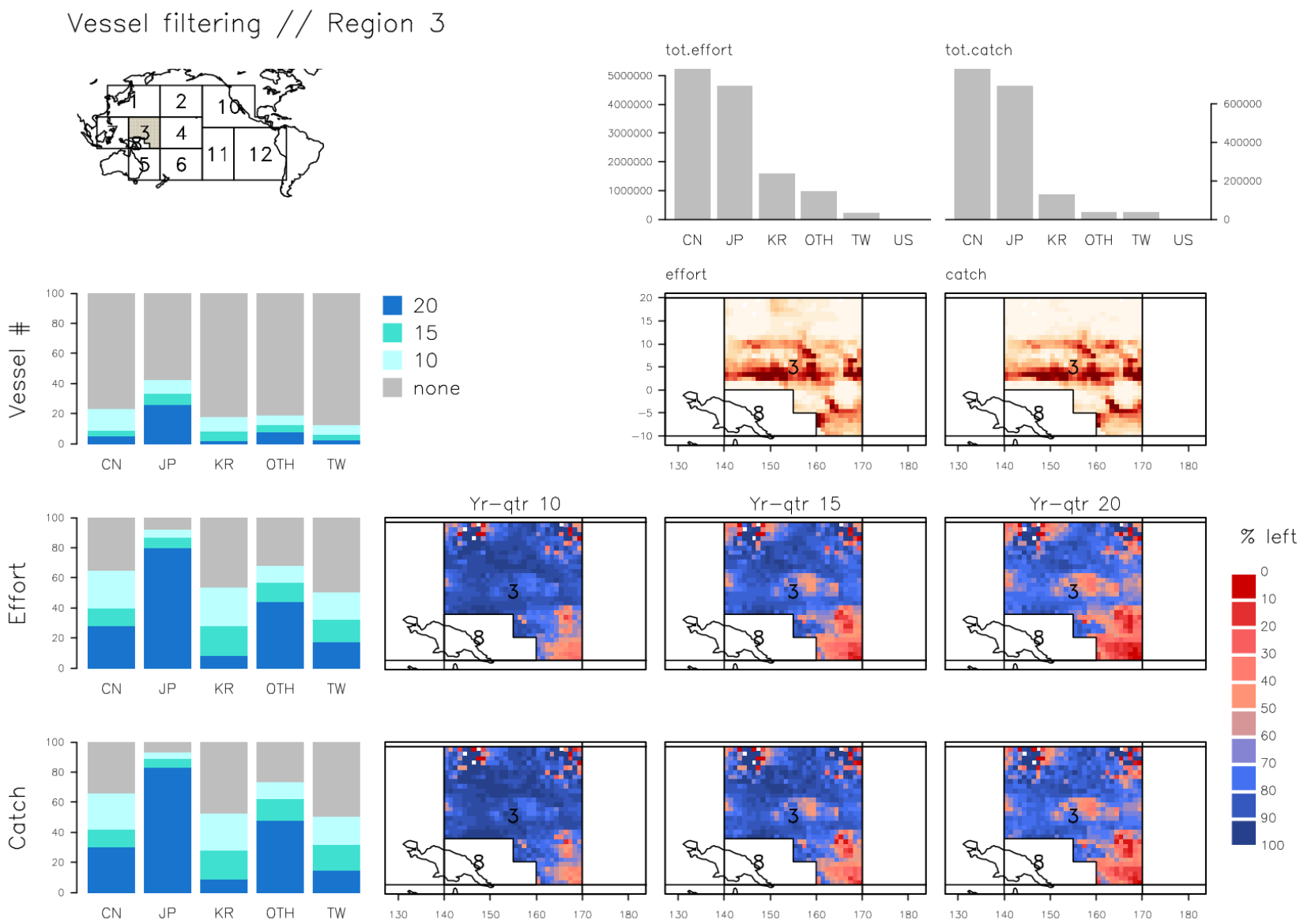


Figure 24: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

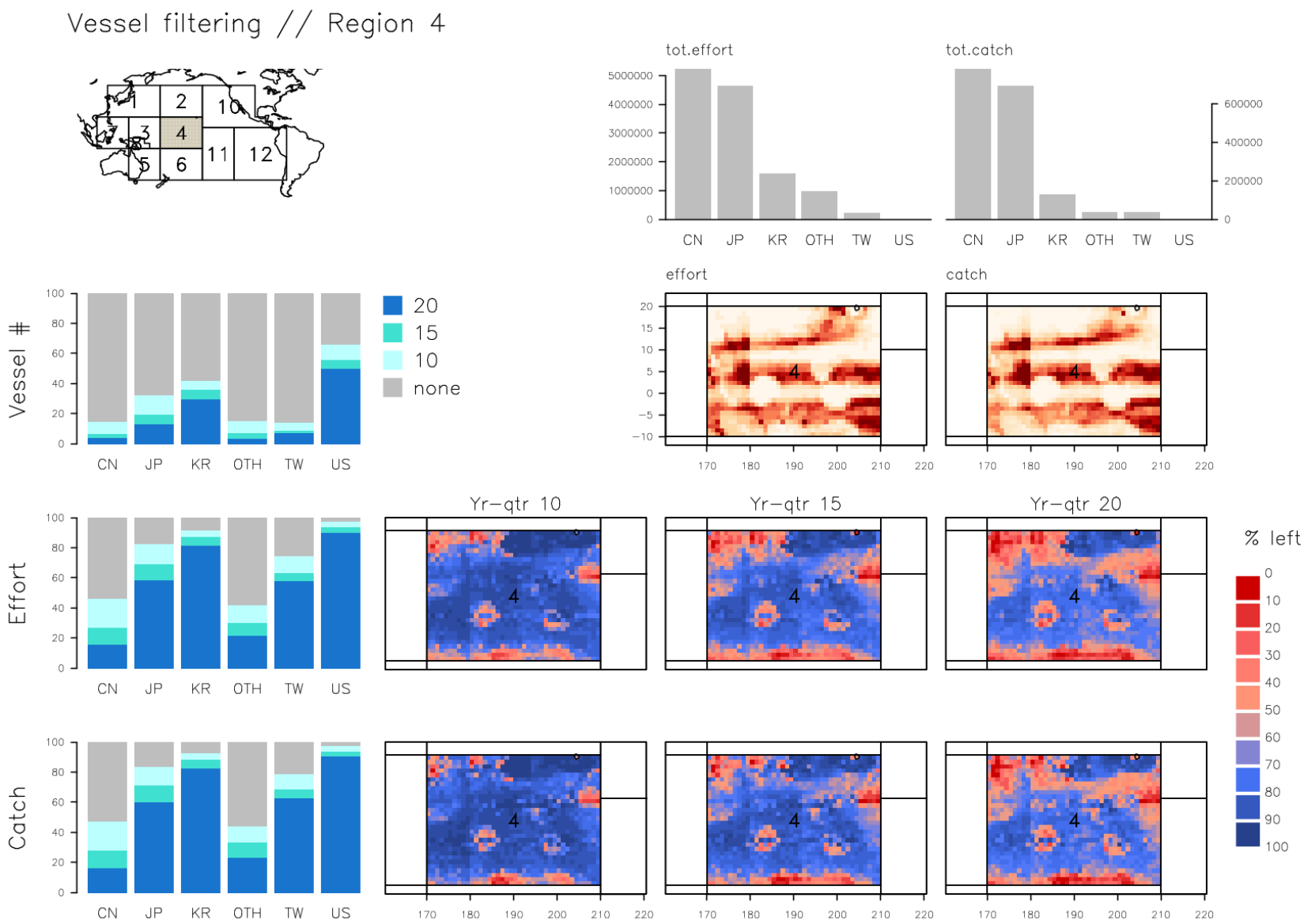


Figure 25: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

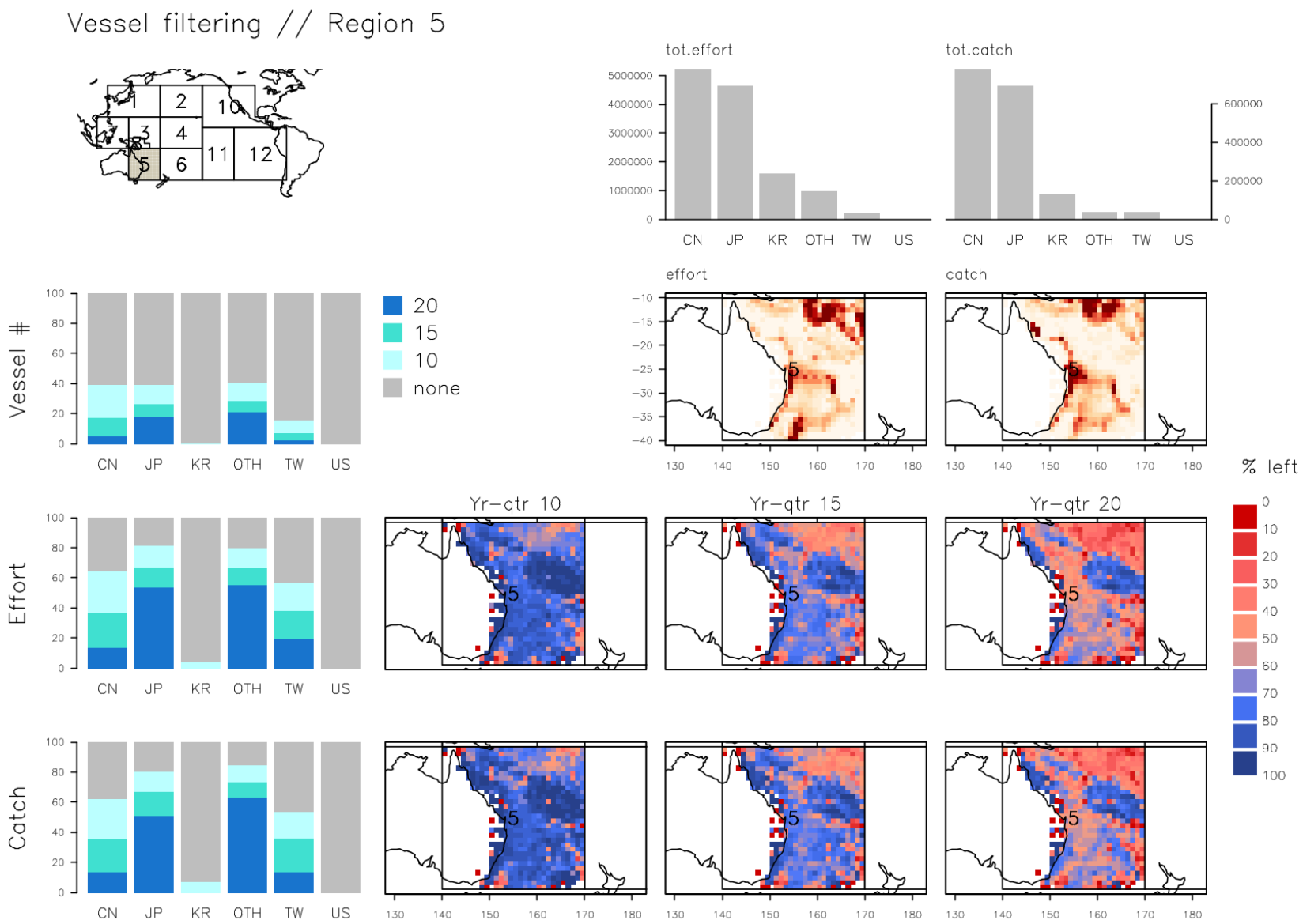


Figure 26: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

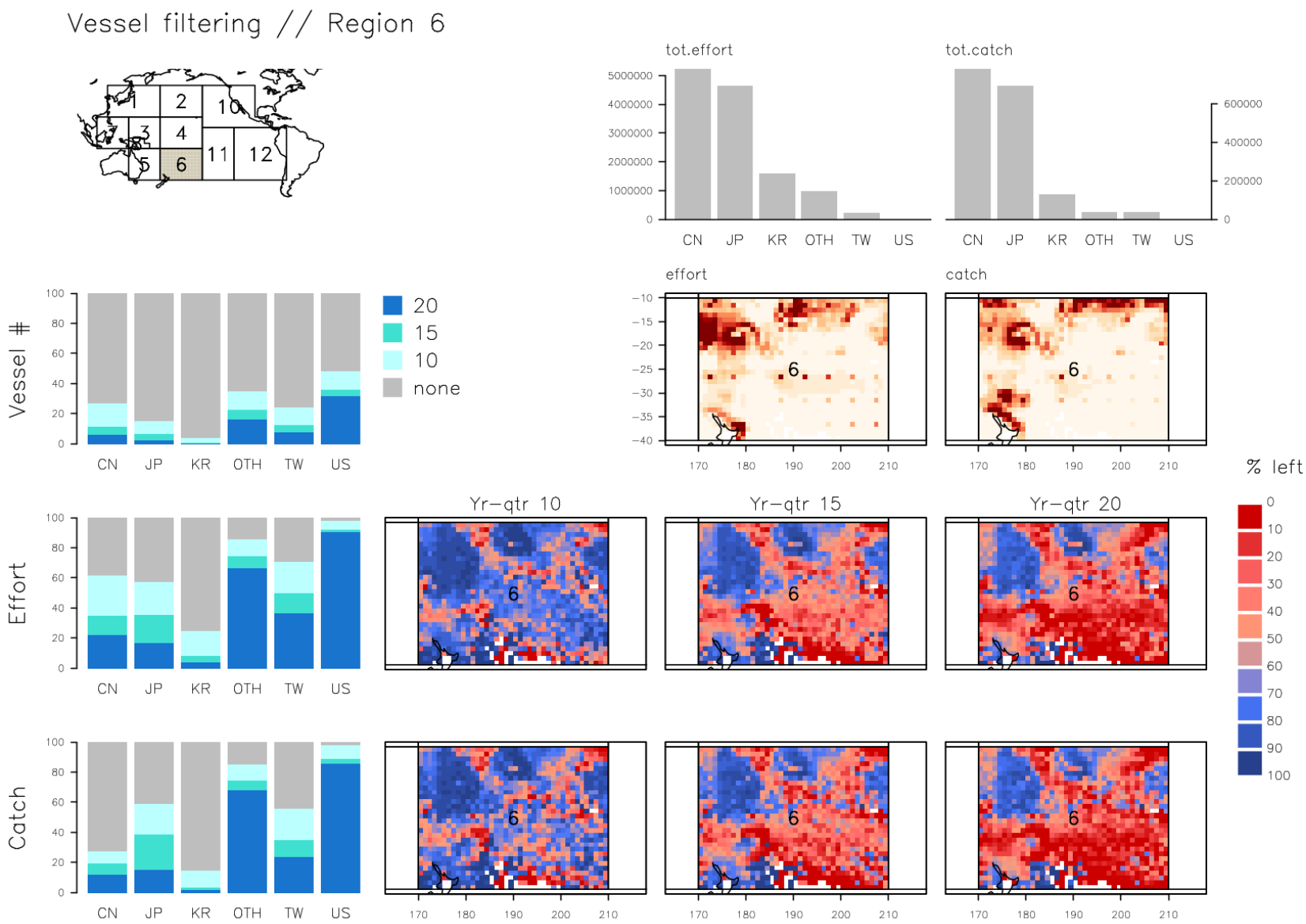


Figure 27: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

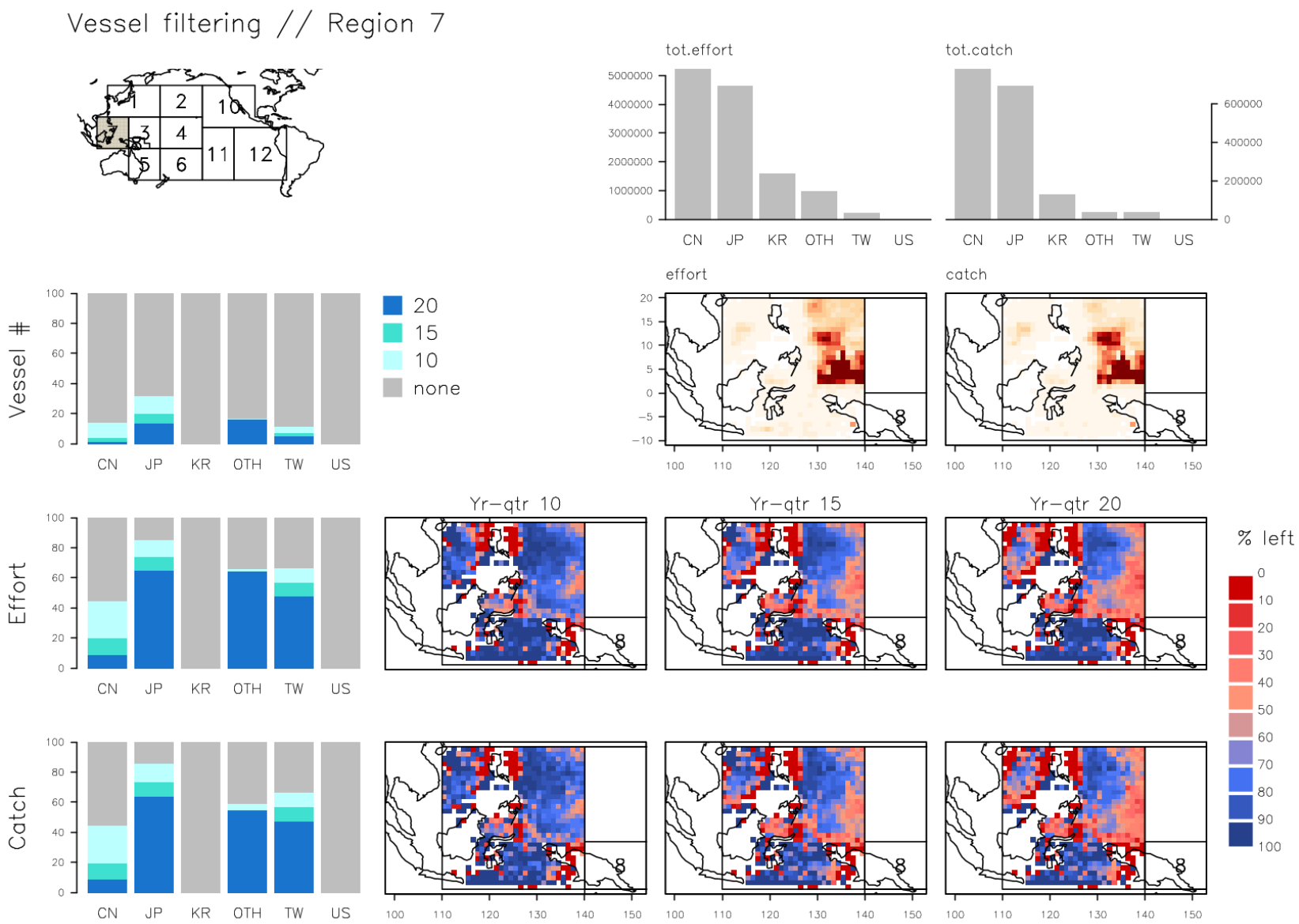


Figure 28: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

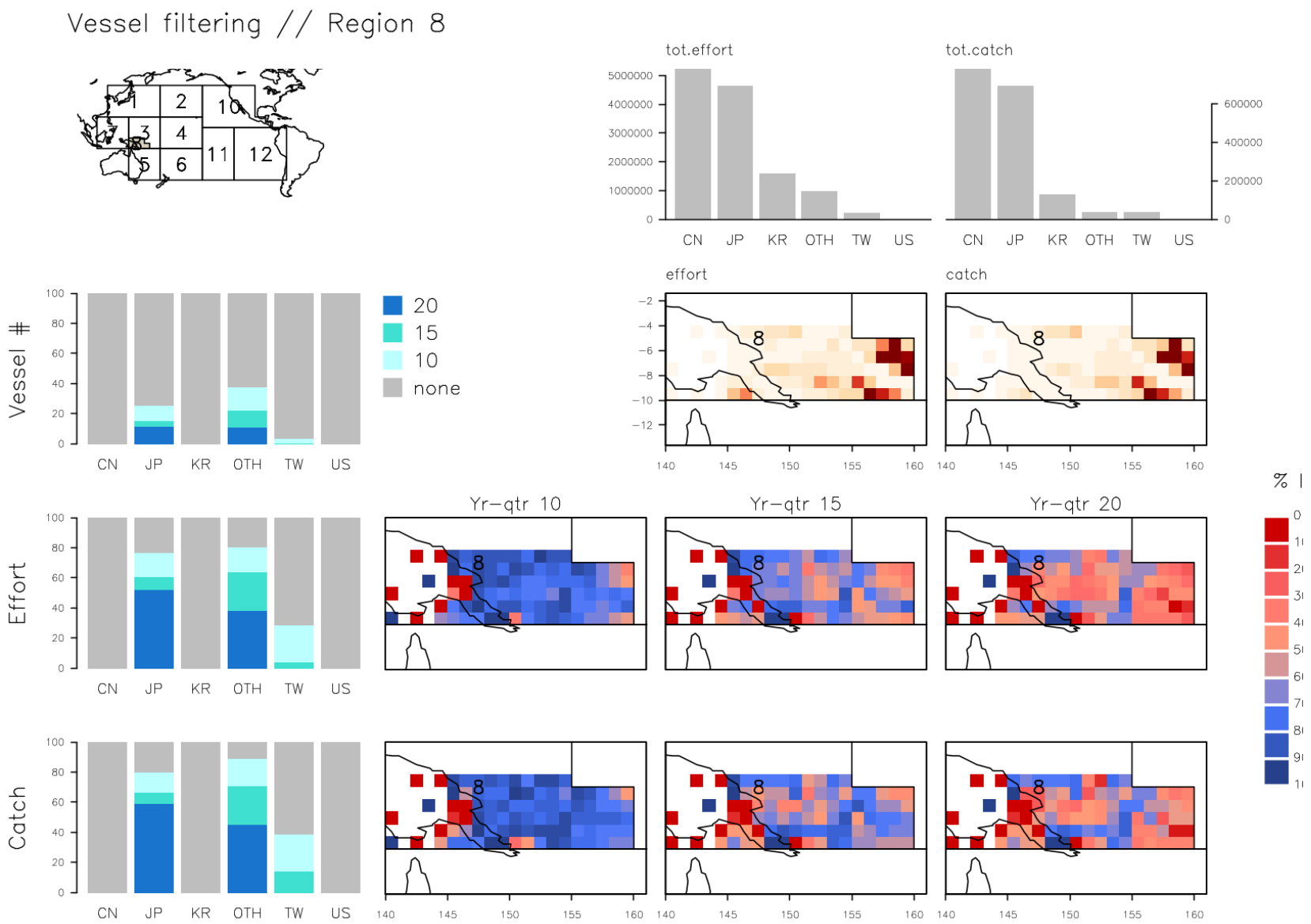


Figure 29: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

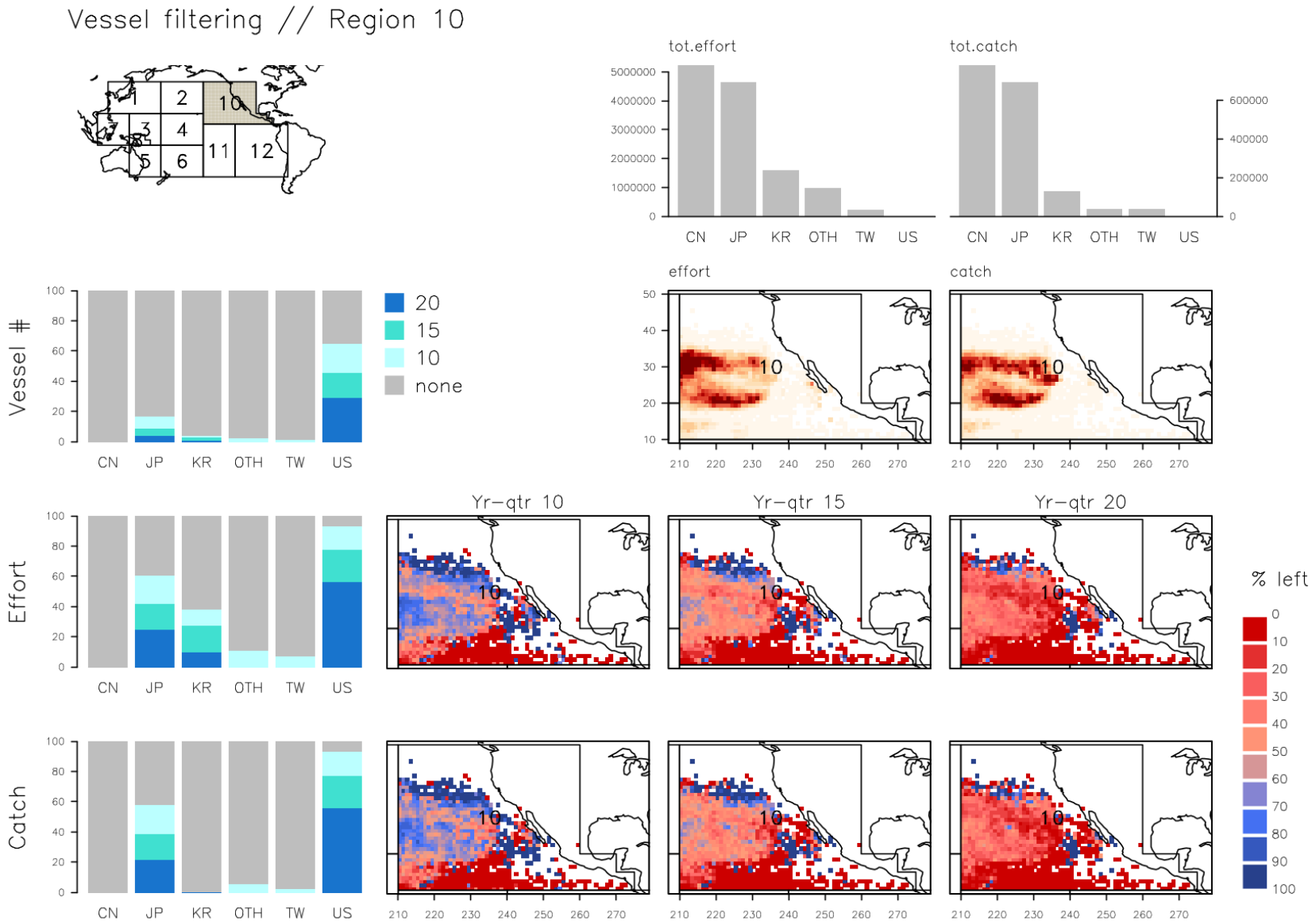


Figure 30: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

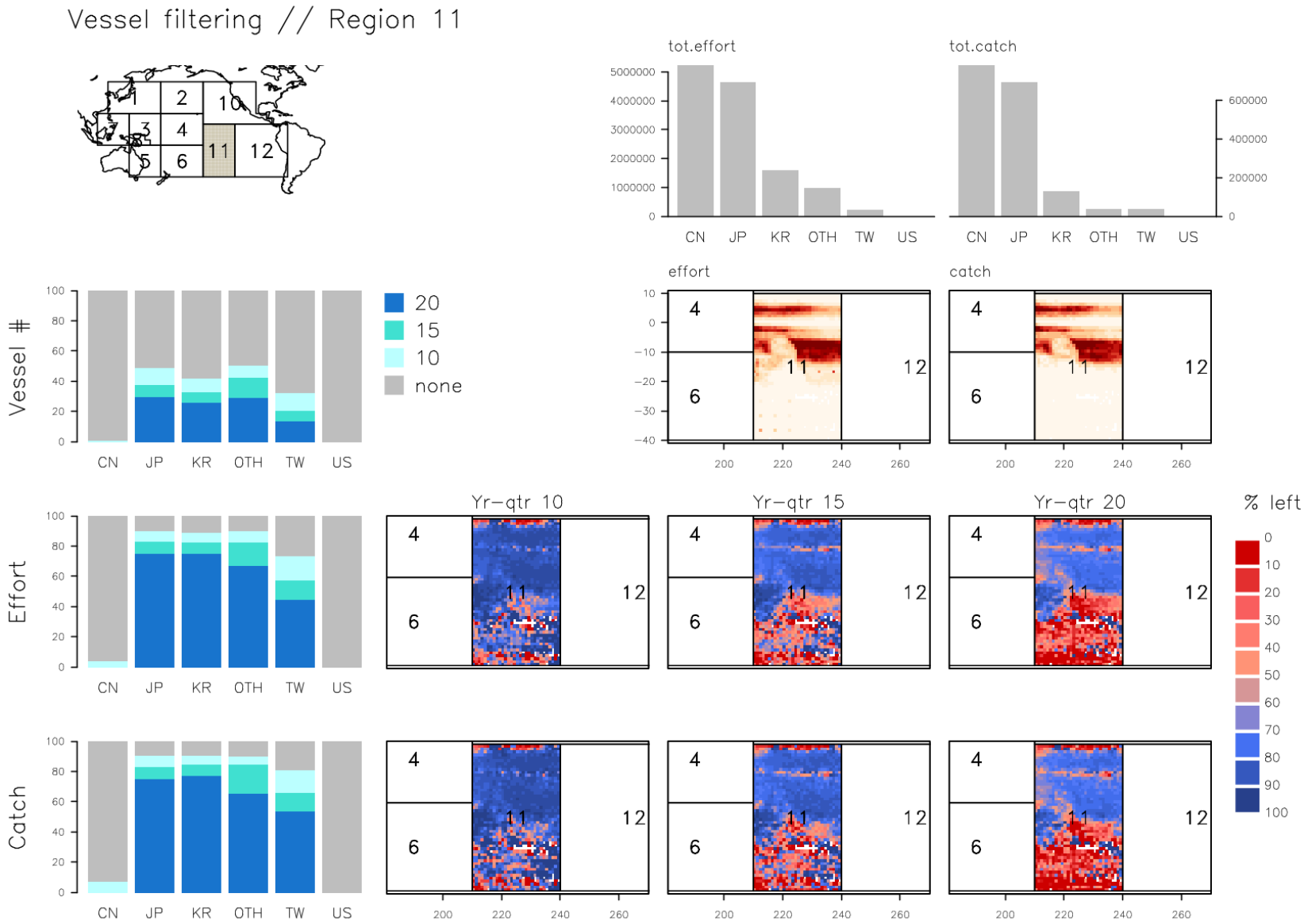


Figure 31: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

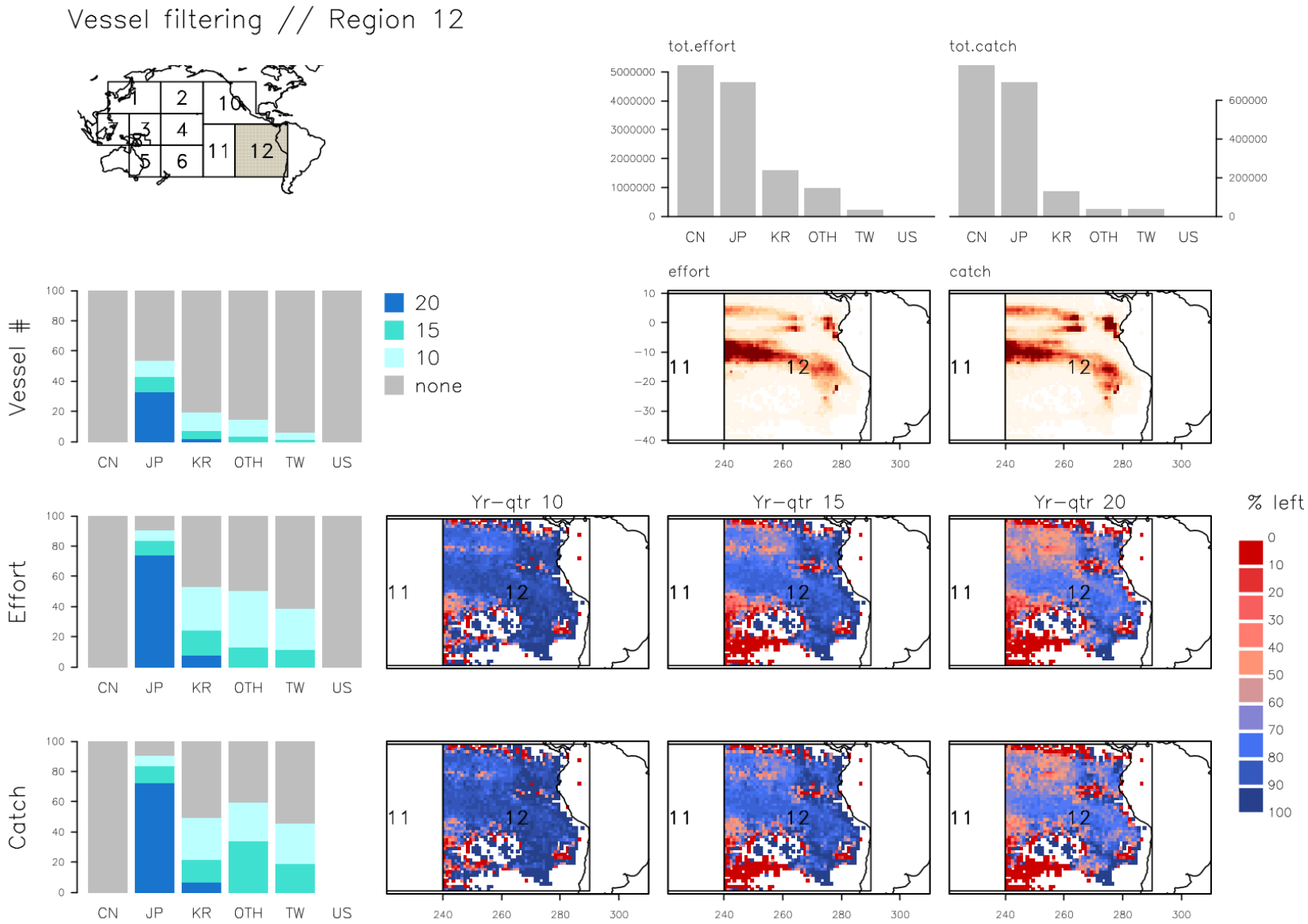


Figure 32: Effect of thresholds of minimal year-quarter presence for core fleet membership on the proportion of vessels remaining by fleet, and the spatial distribution of retained effort and catch.

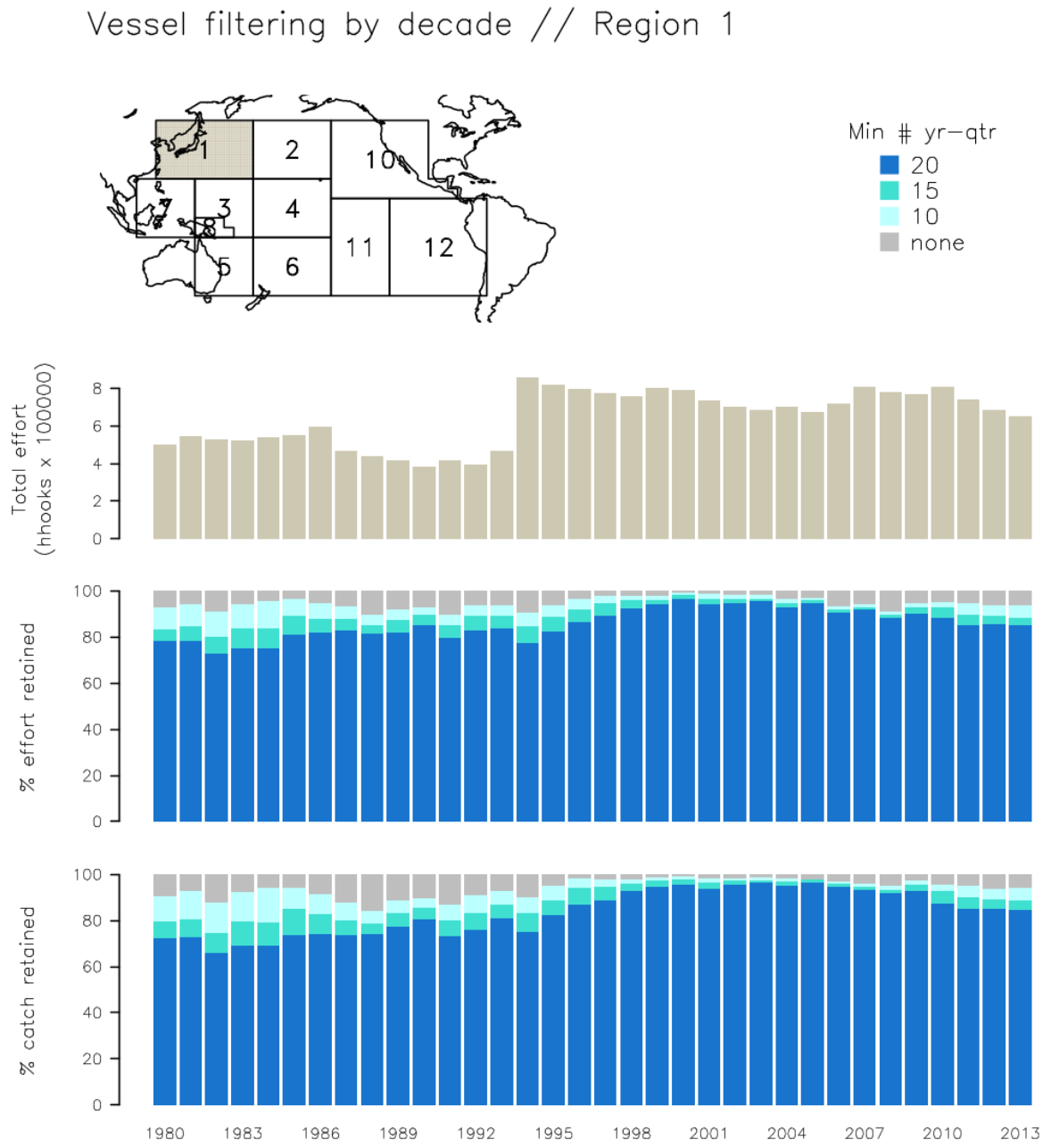


Figure 33: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

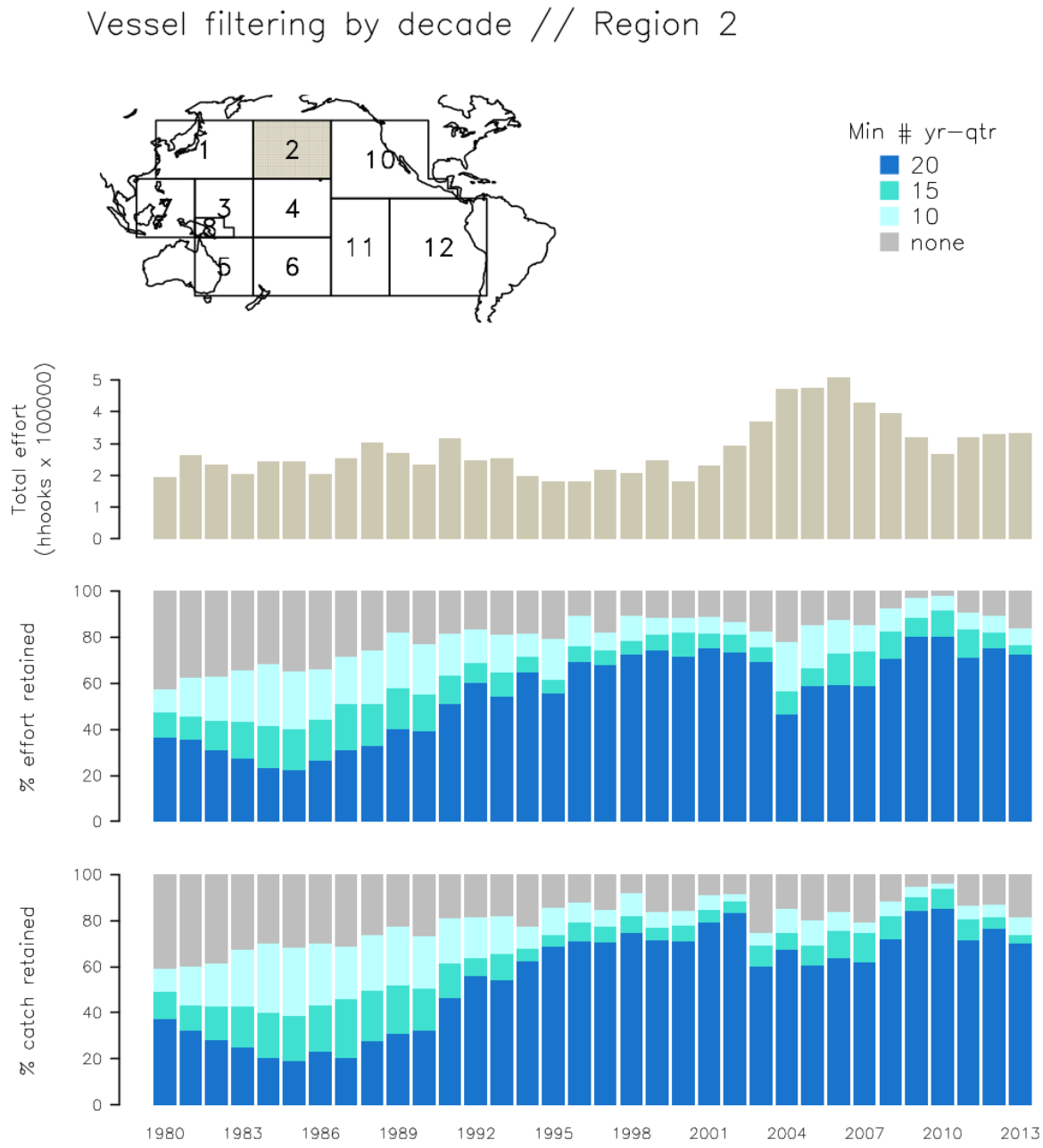


Figure 34: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

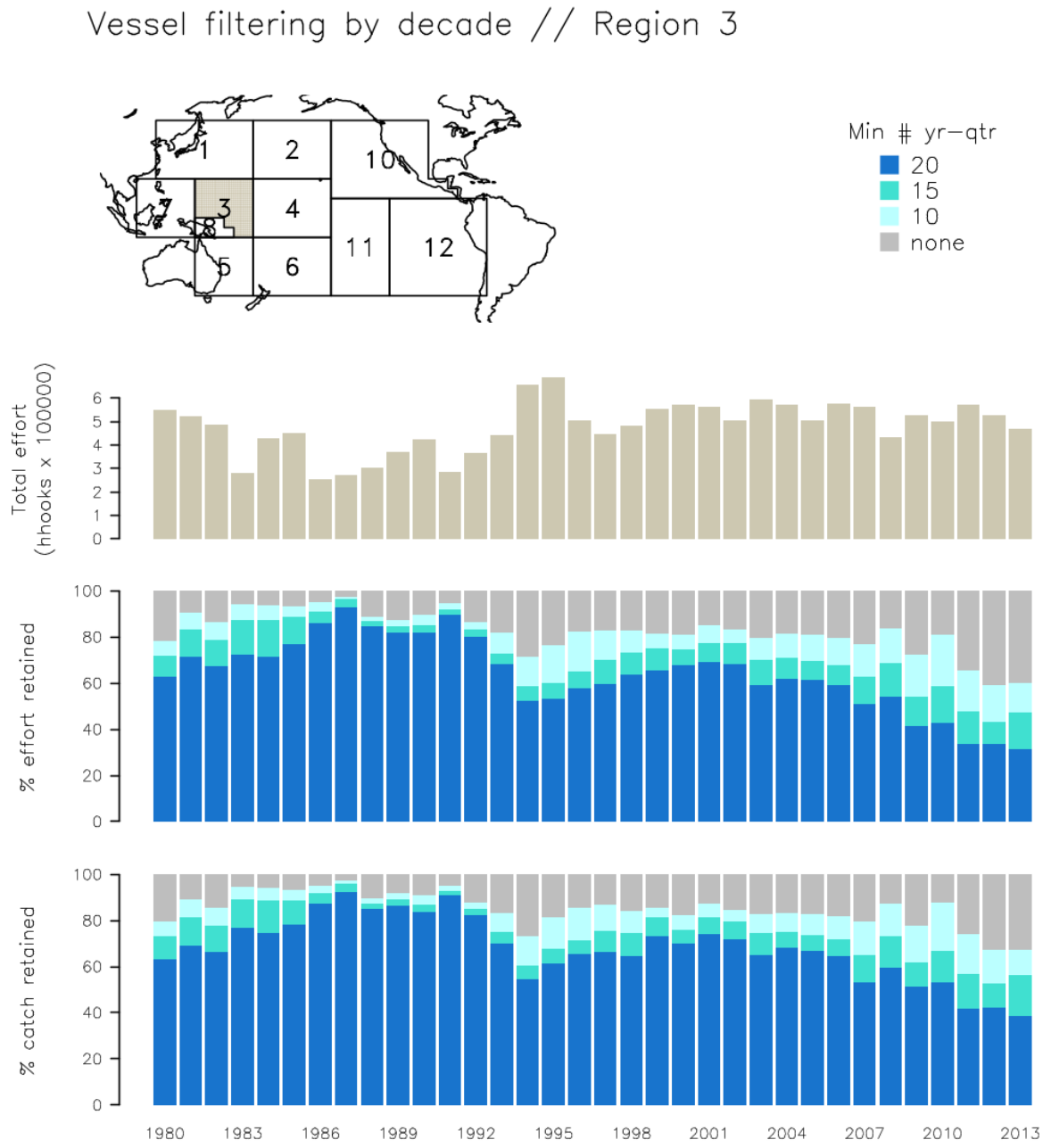


Figure 35: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

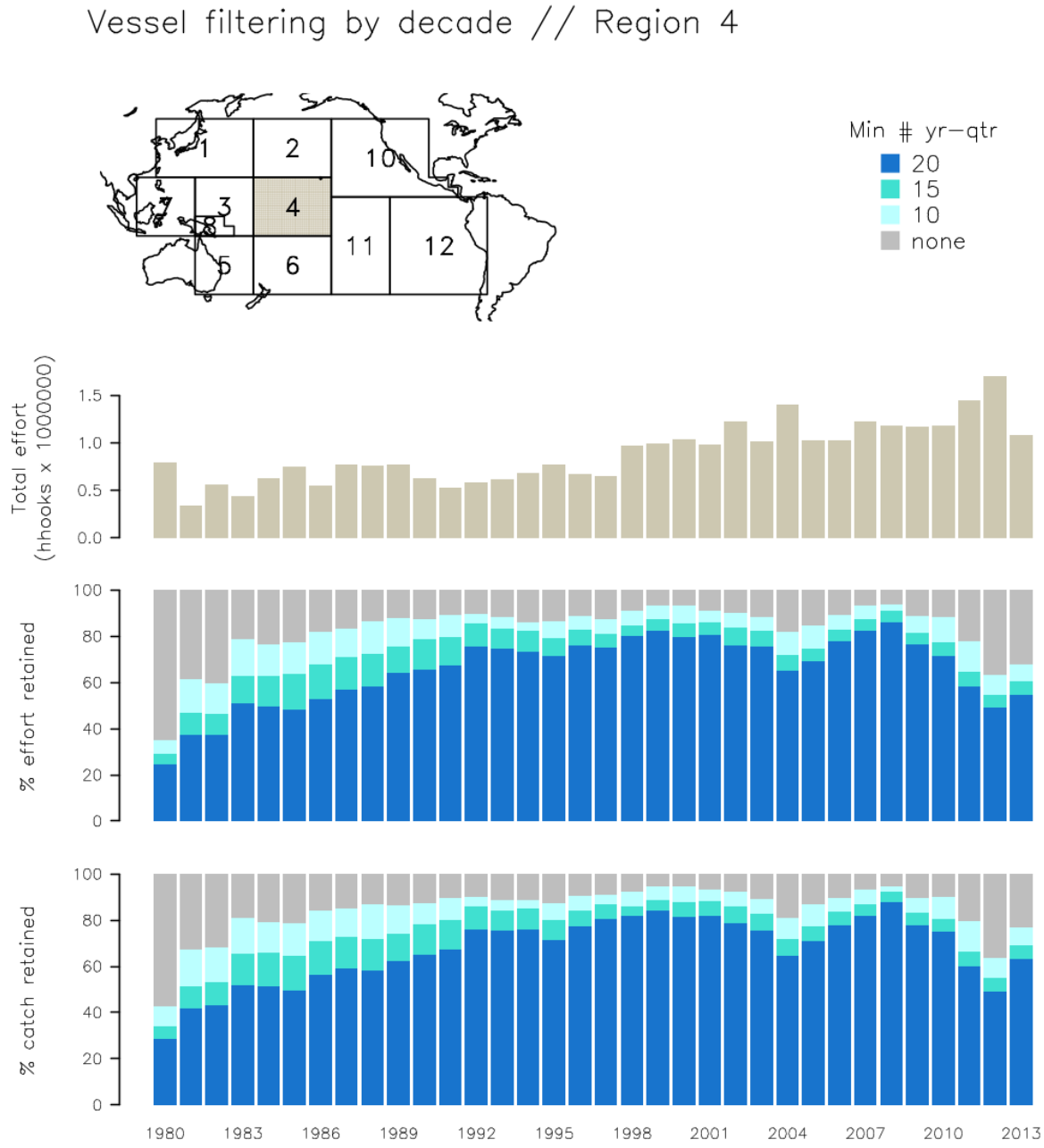


Figure 36: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

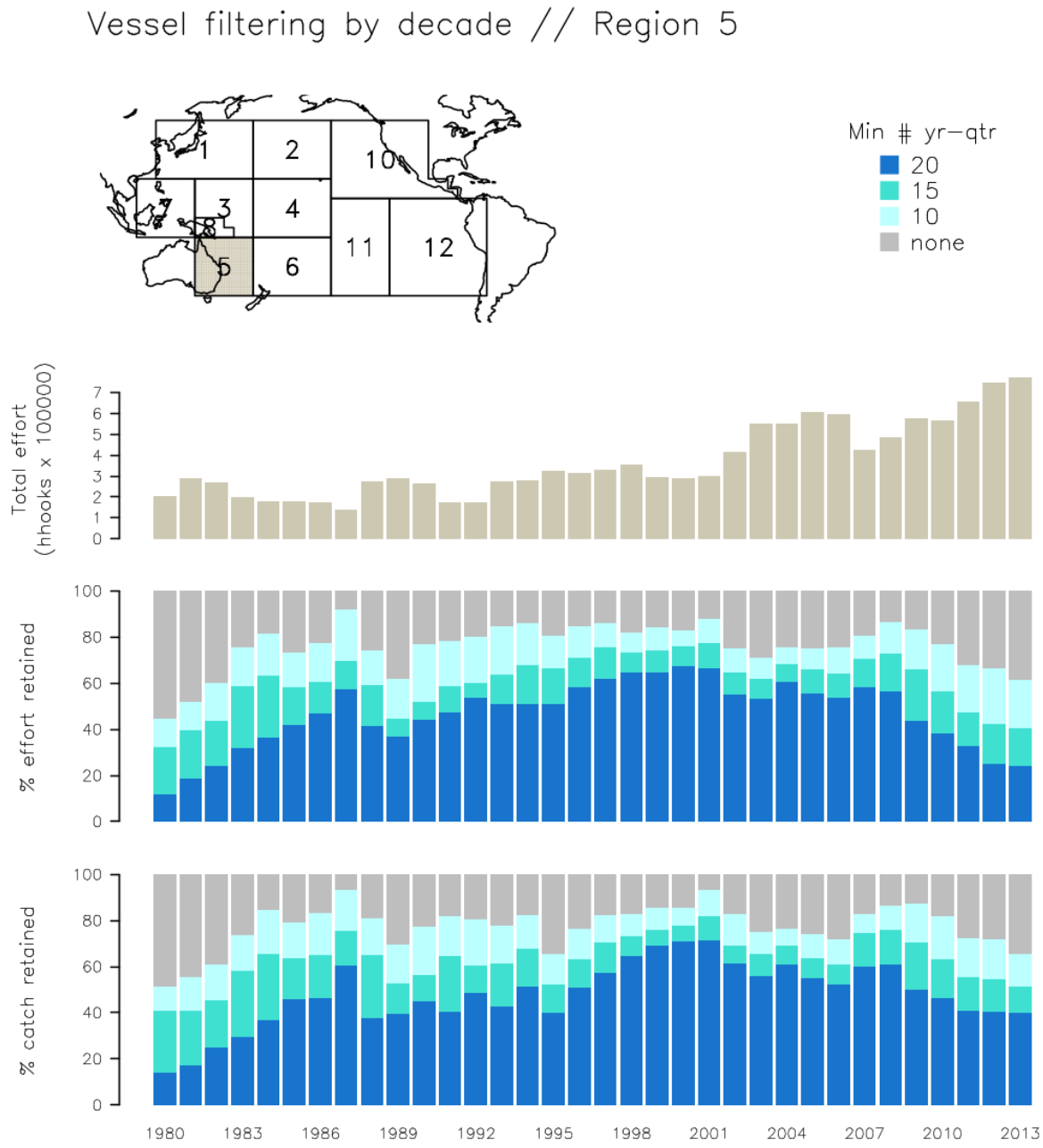


Figure 37: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

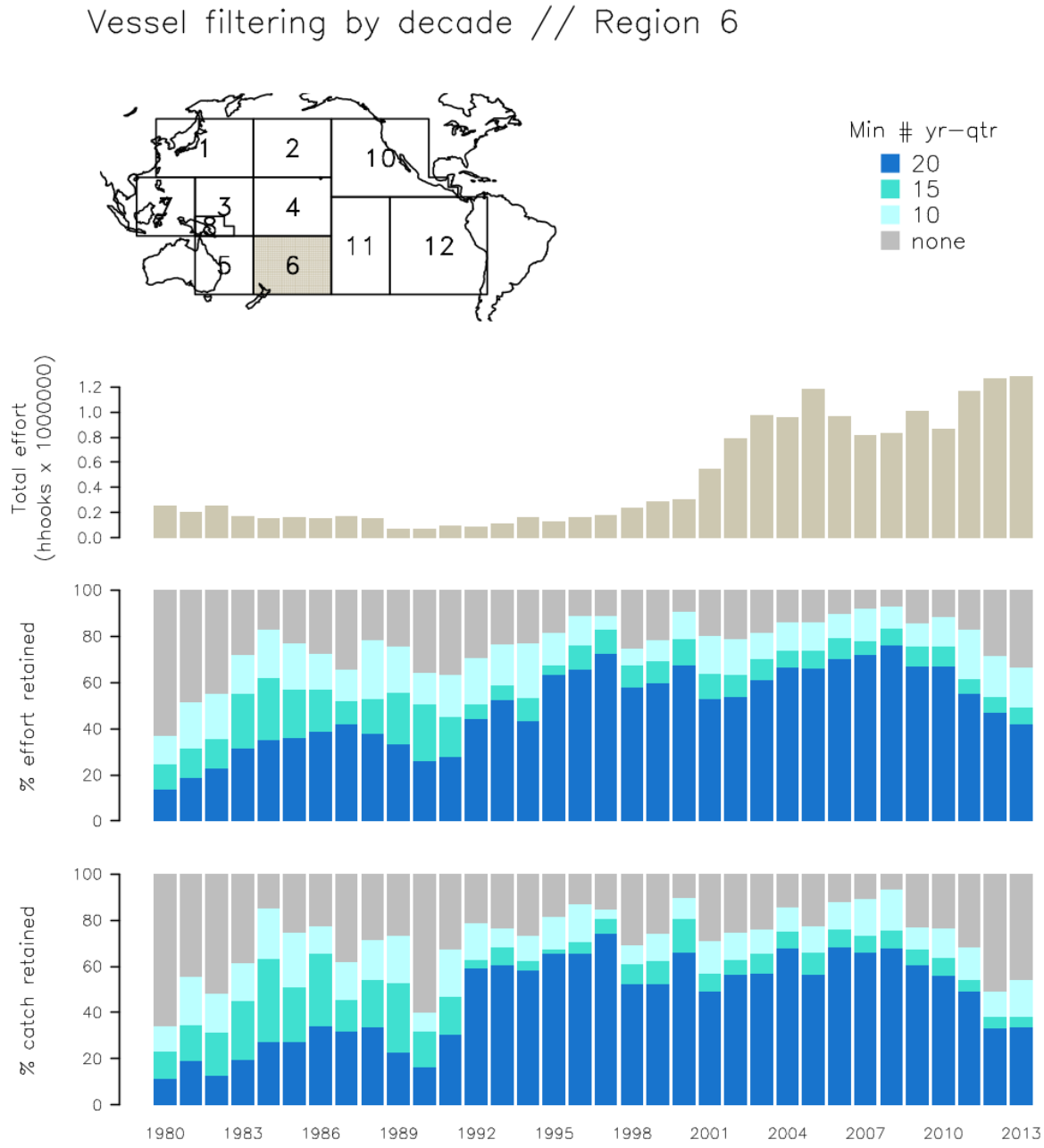


Figure 38: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

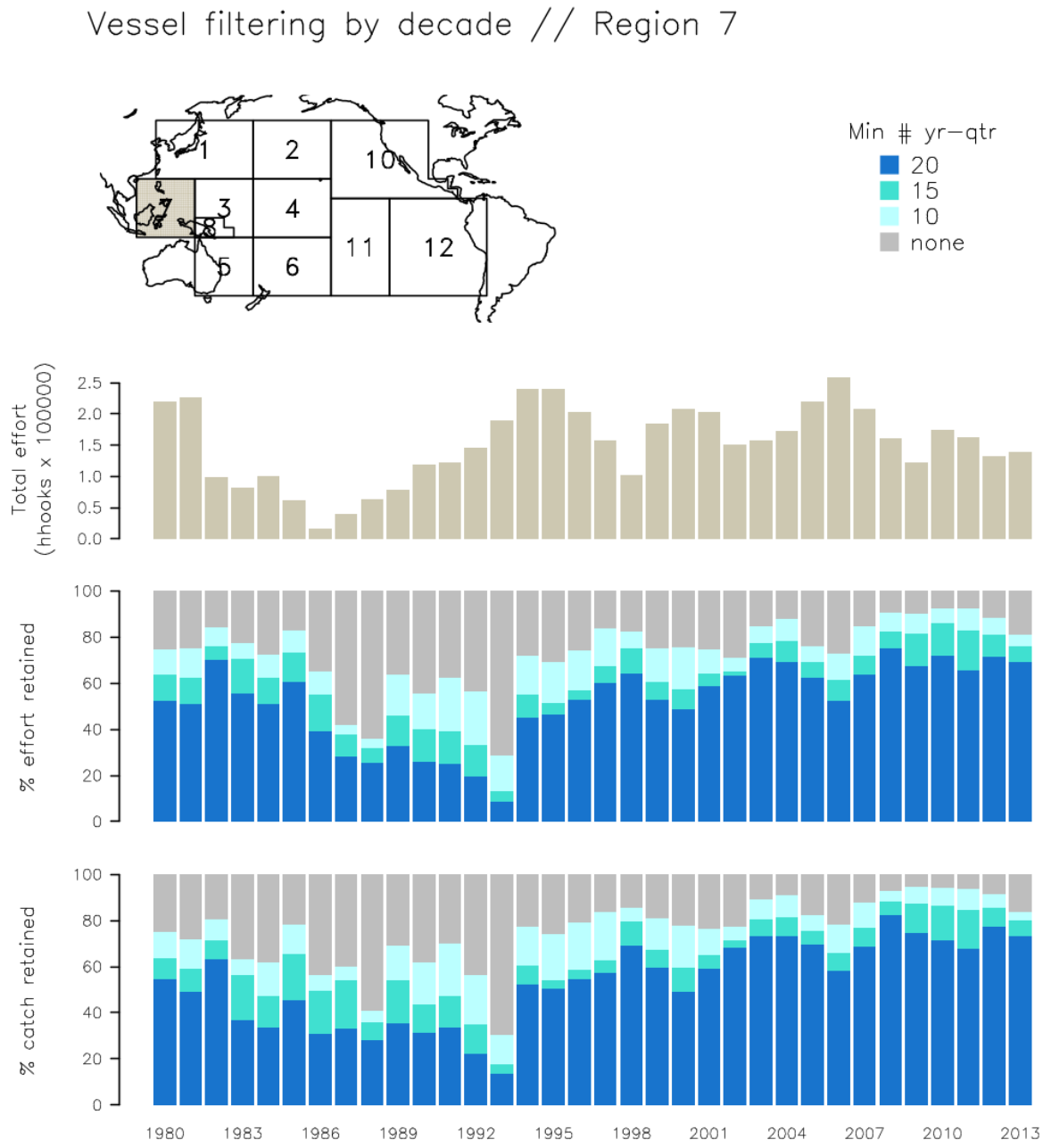


Figure 39: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

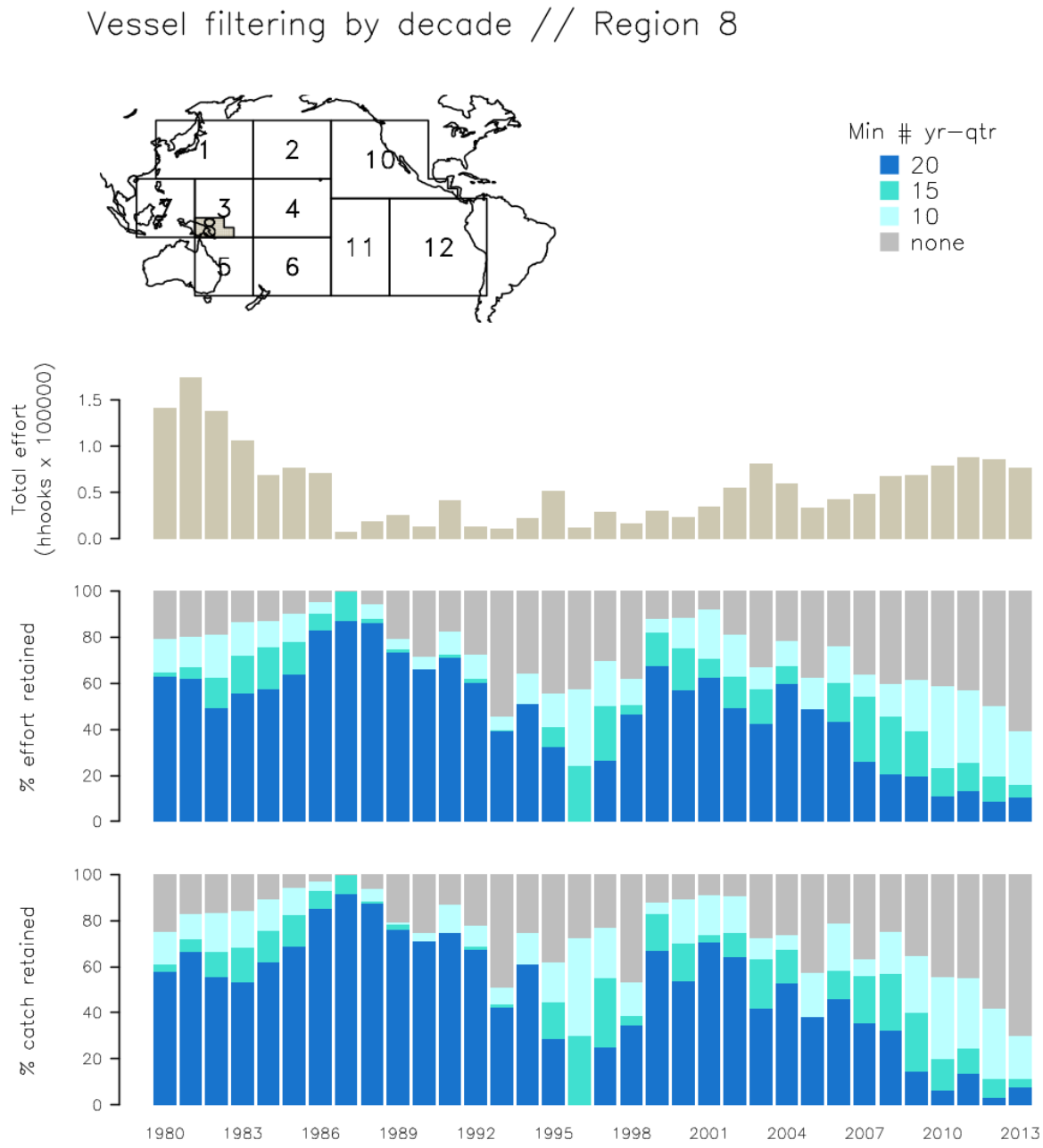


Figure 40: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

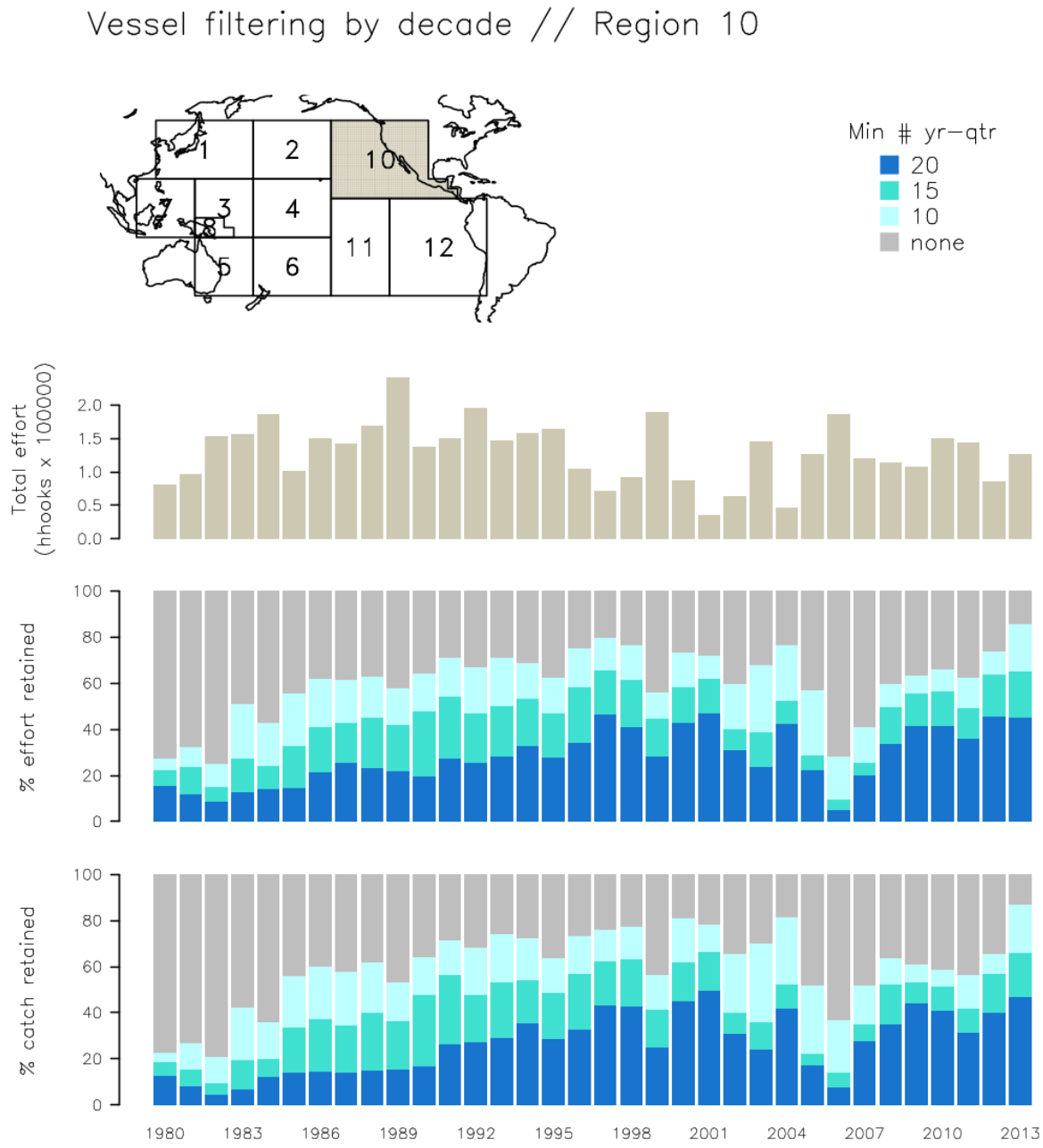


Figure 41: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

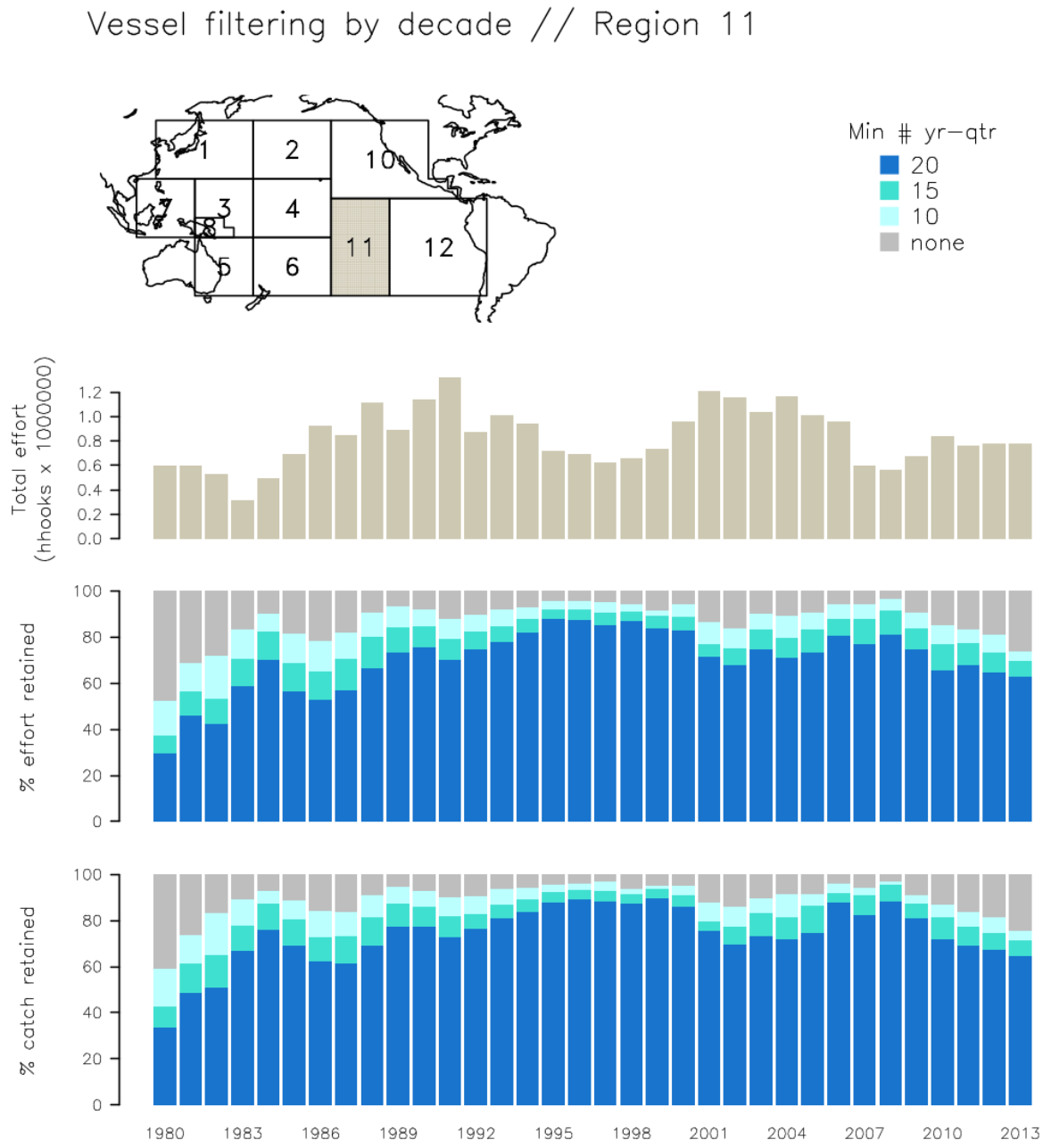


Figure 42: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

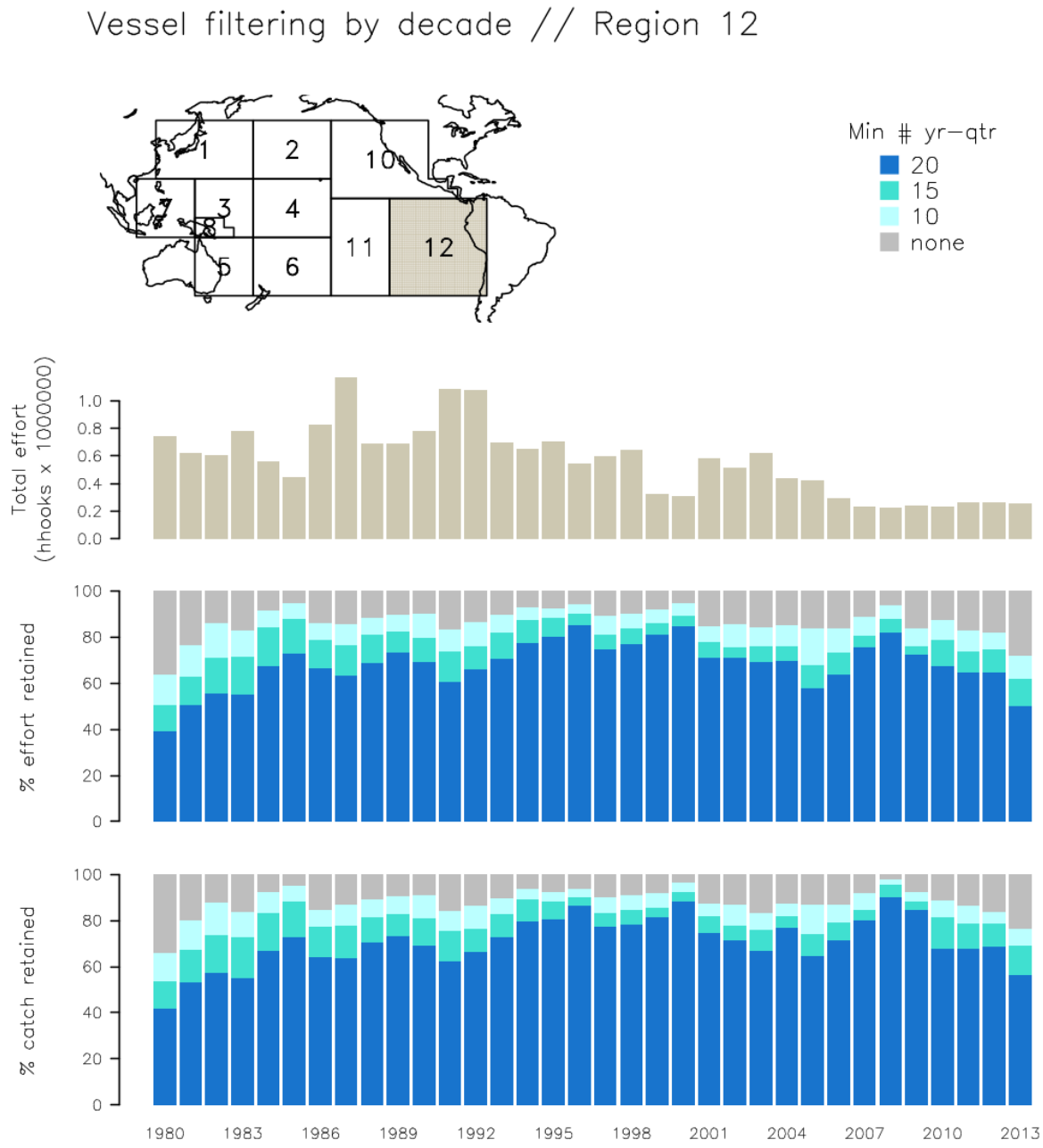


Figure 43: Effect of thresholds of minimal year-quarter presence for core fleet membership on the temporal trends in catch and effort retained for the CPUE analysis.

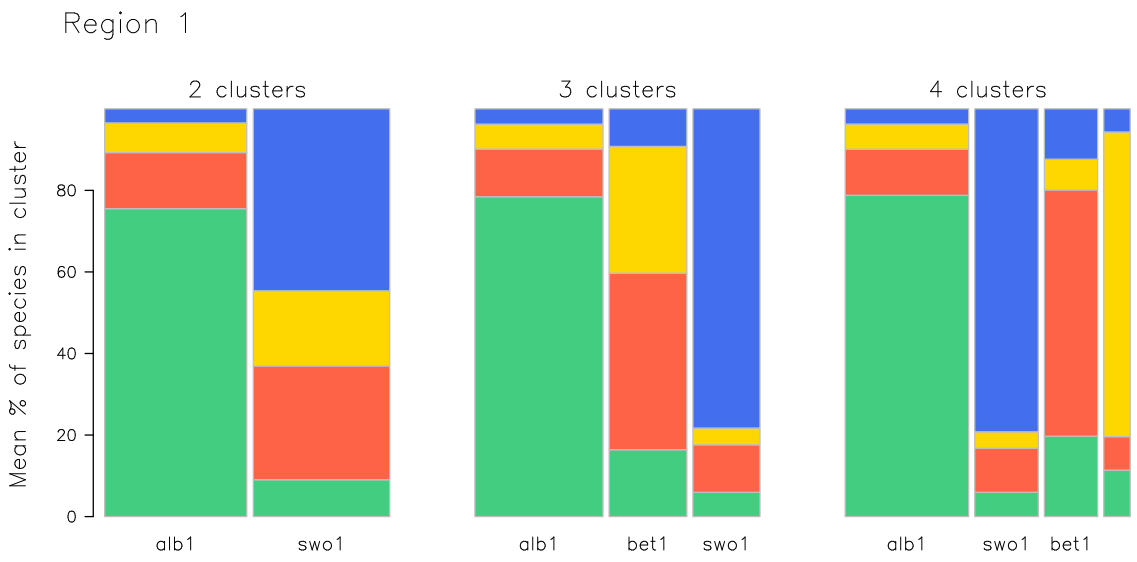


Figure 44: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

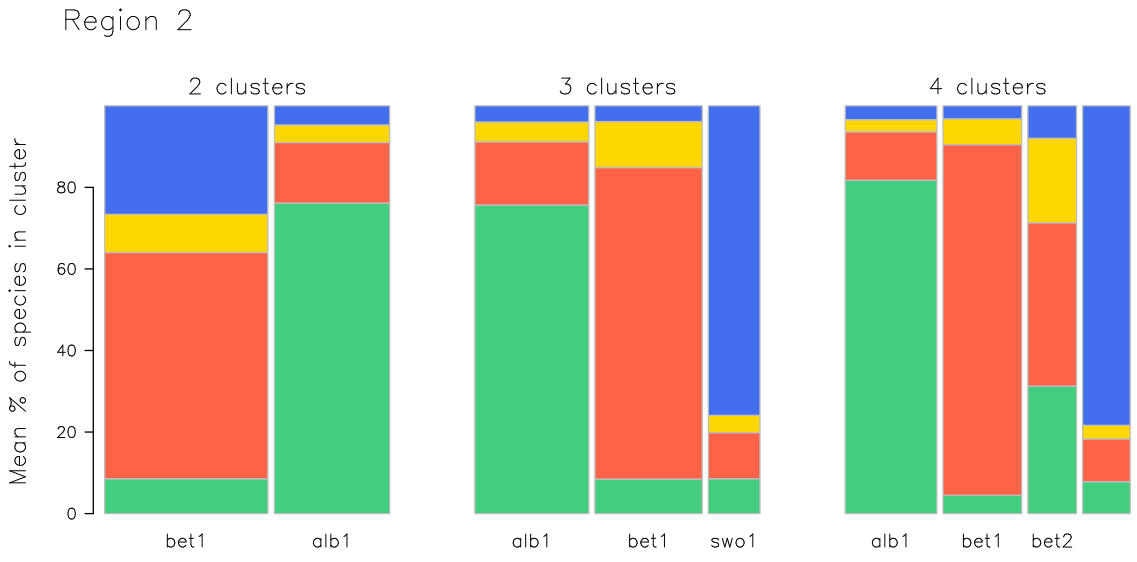


Figure 45: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

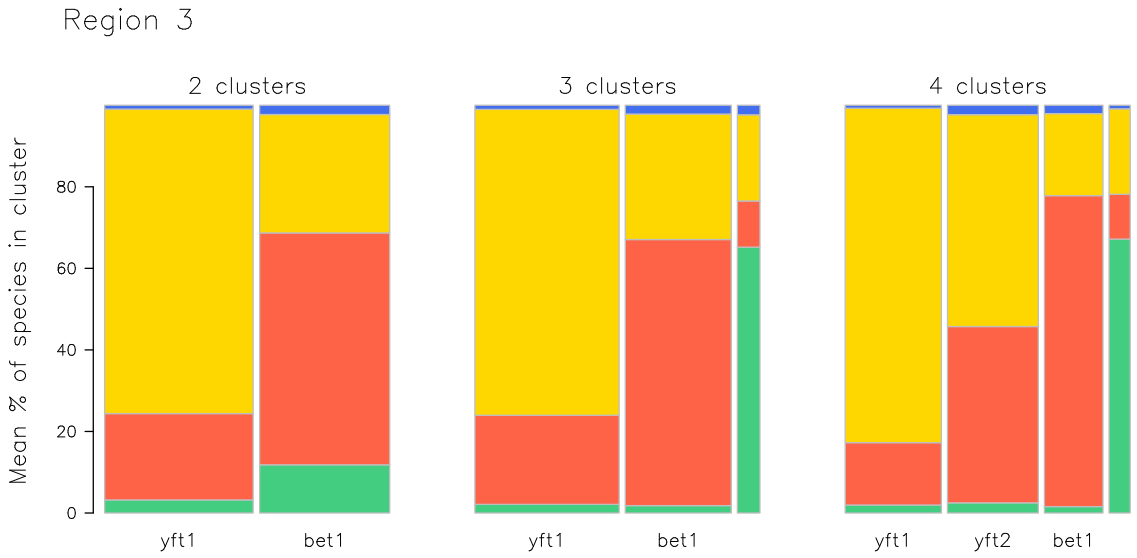


Figure 46: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

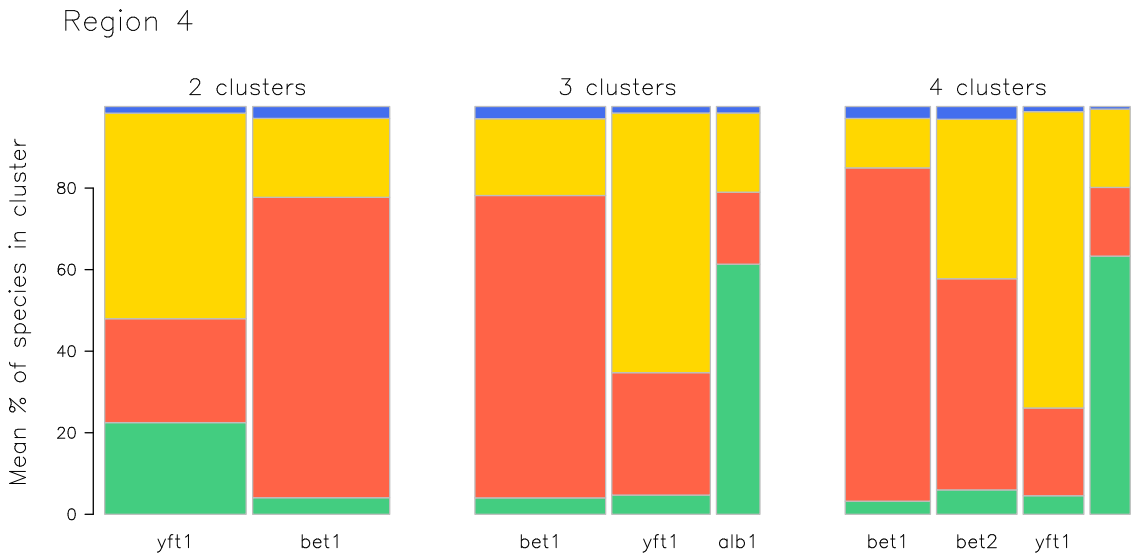


Figure 47: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

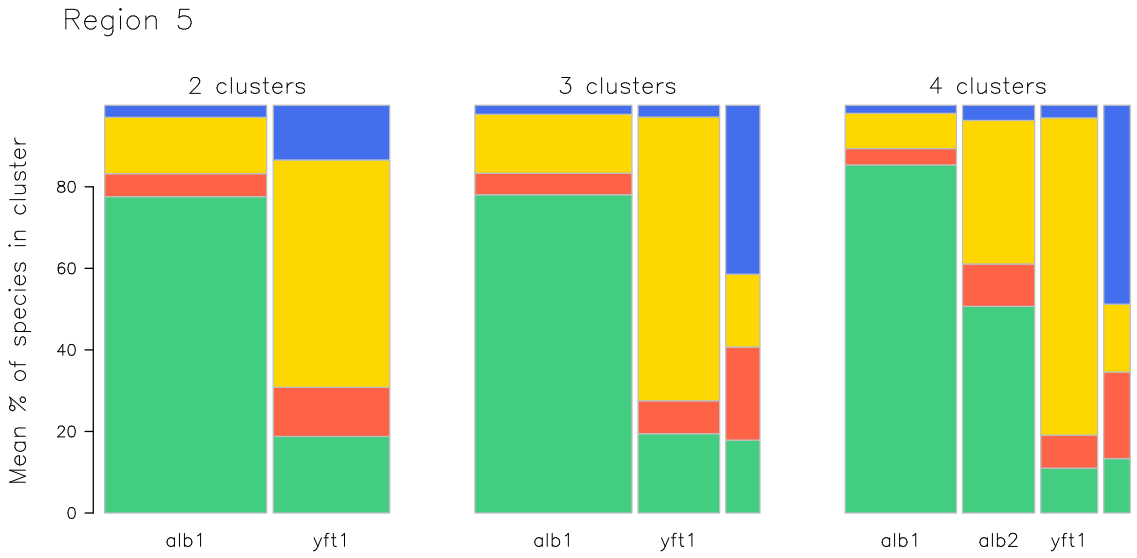


Figure 48: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

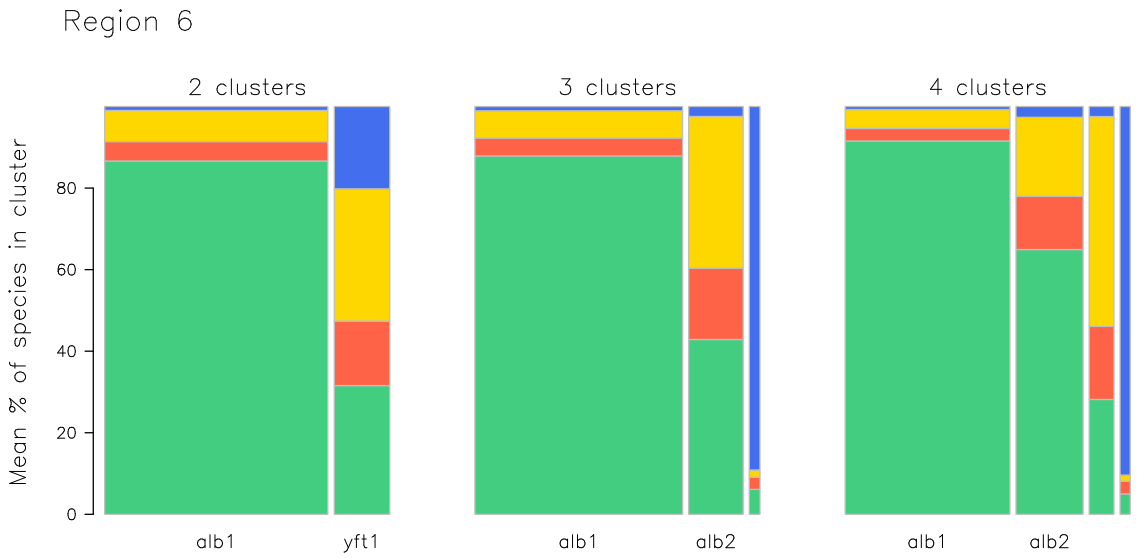


Figure 49: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.


Figure 50: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

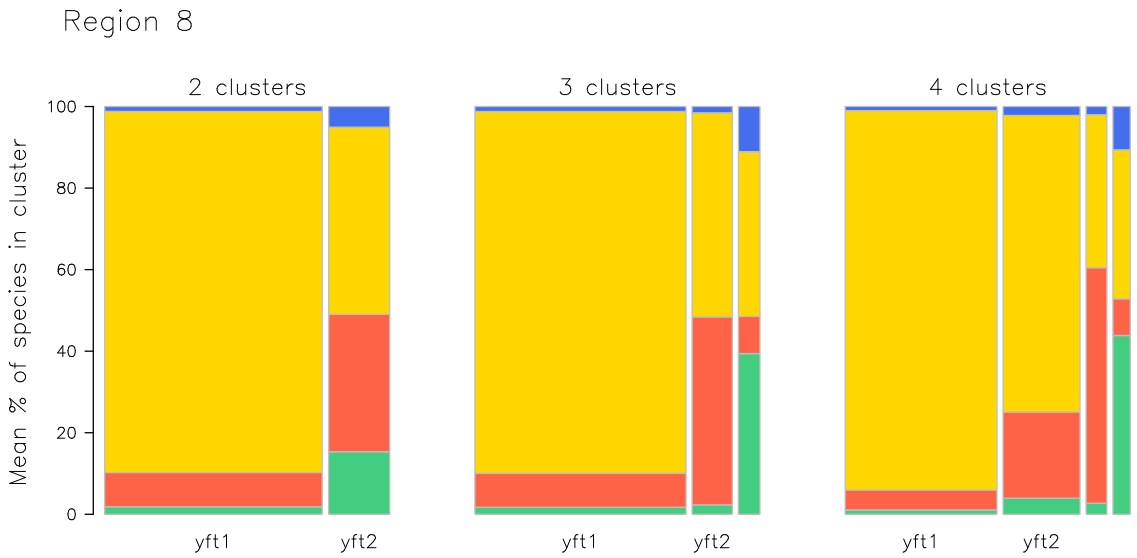


Figure 51: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

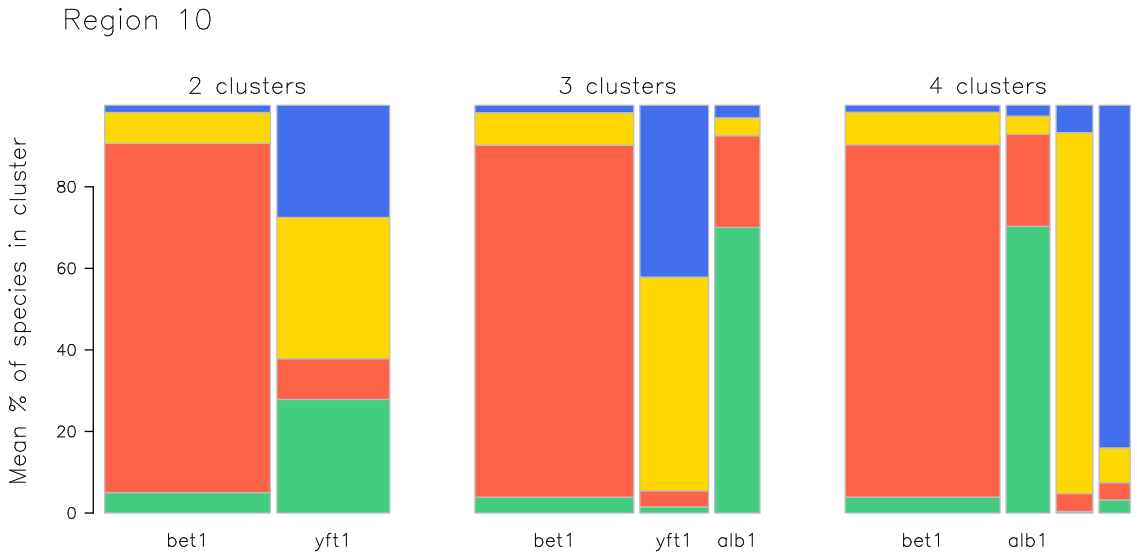


Figure 52: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

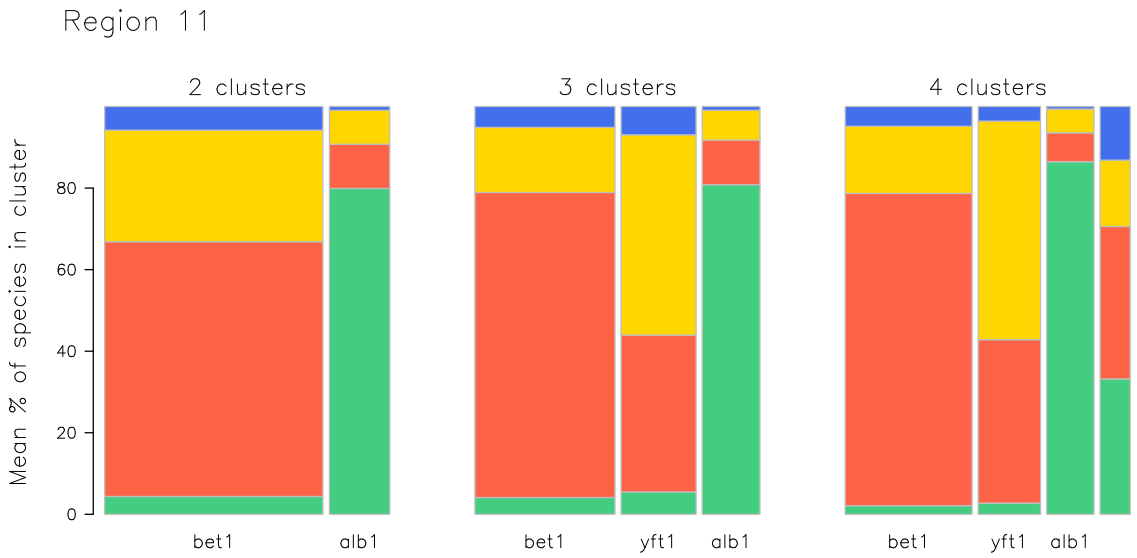


Figure 53: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

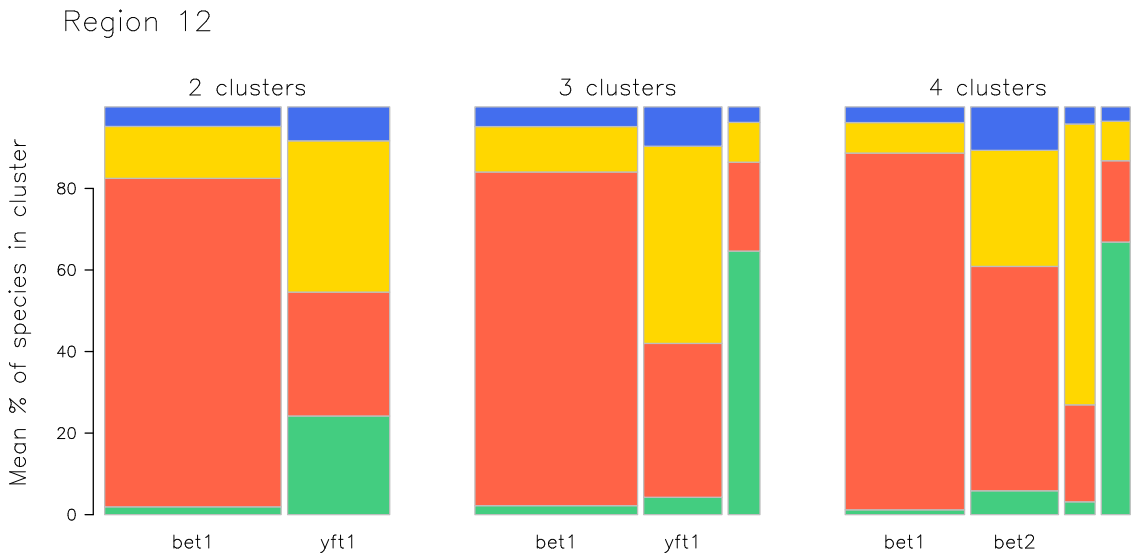


Figure 54: Mean proportion of species in the catch for each cluster - for 2, 3, and 4 cluster models. The width of the bar is proportional to the number of records in the cluster.

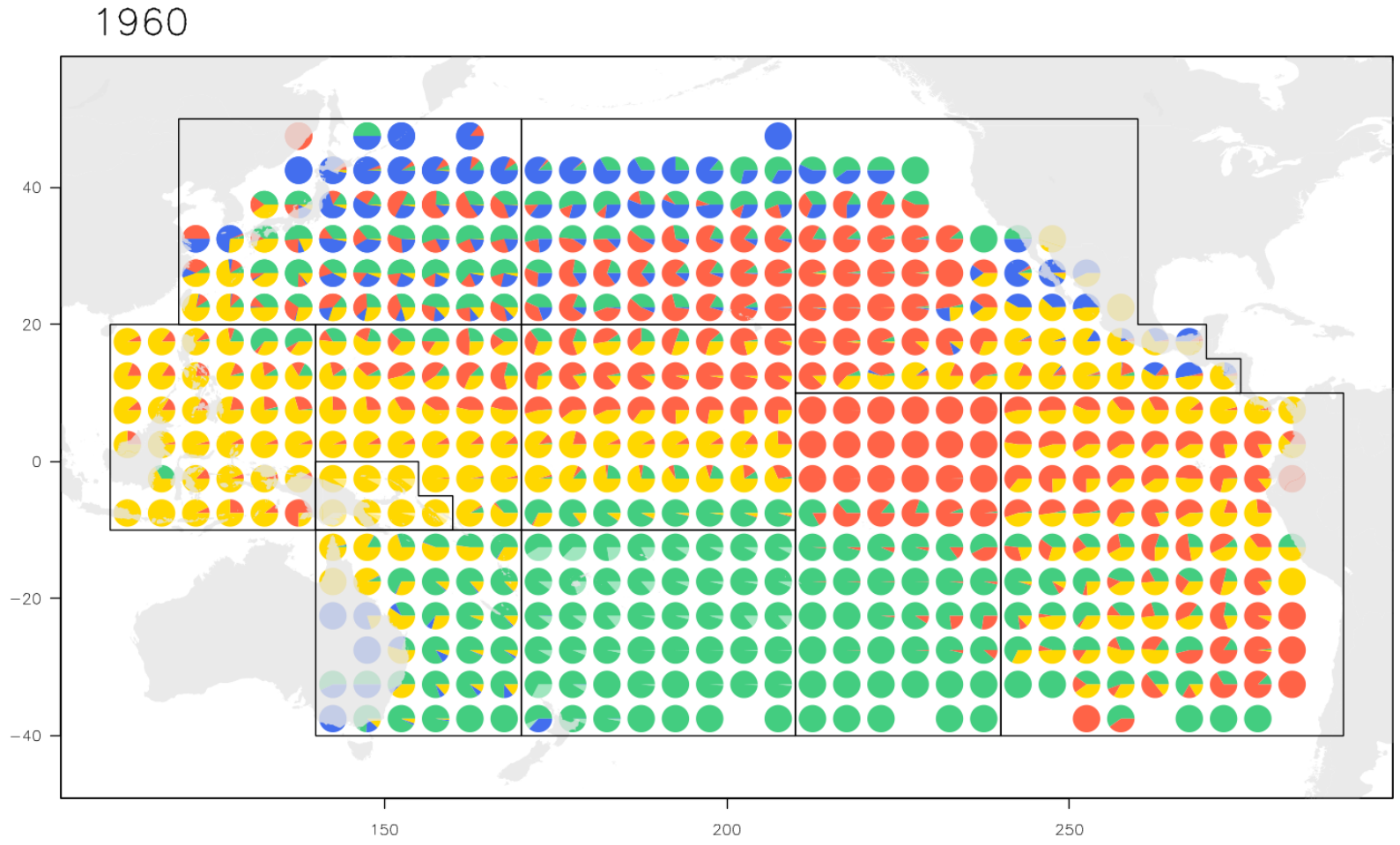


Figure 55: Map of cluster composition at the 5 degree scale, by decade, with region-specific cluster numbers as described in Table 2. ALB, BET, SWO and YFT are given by green, red, blue and yellow, respectively.

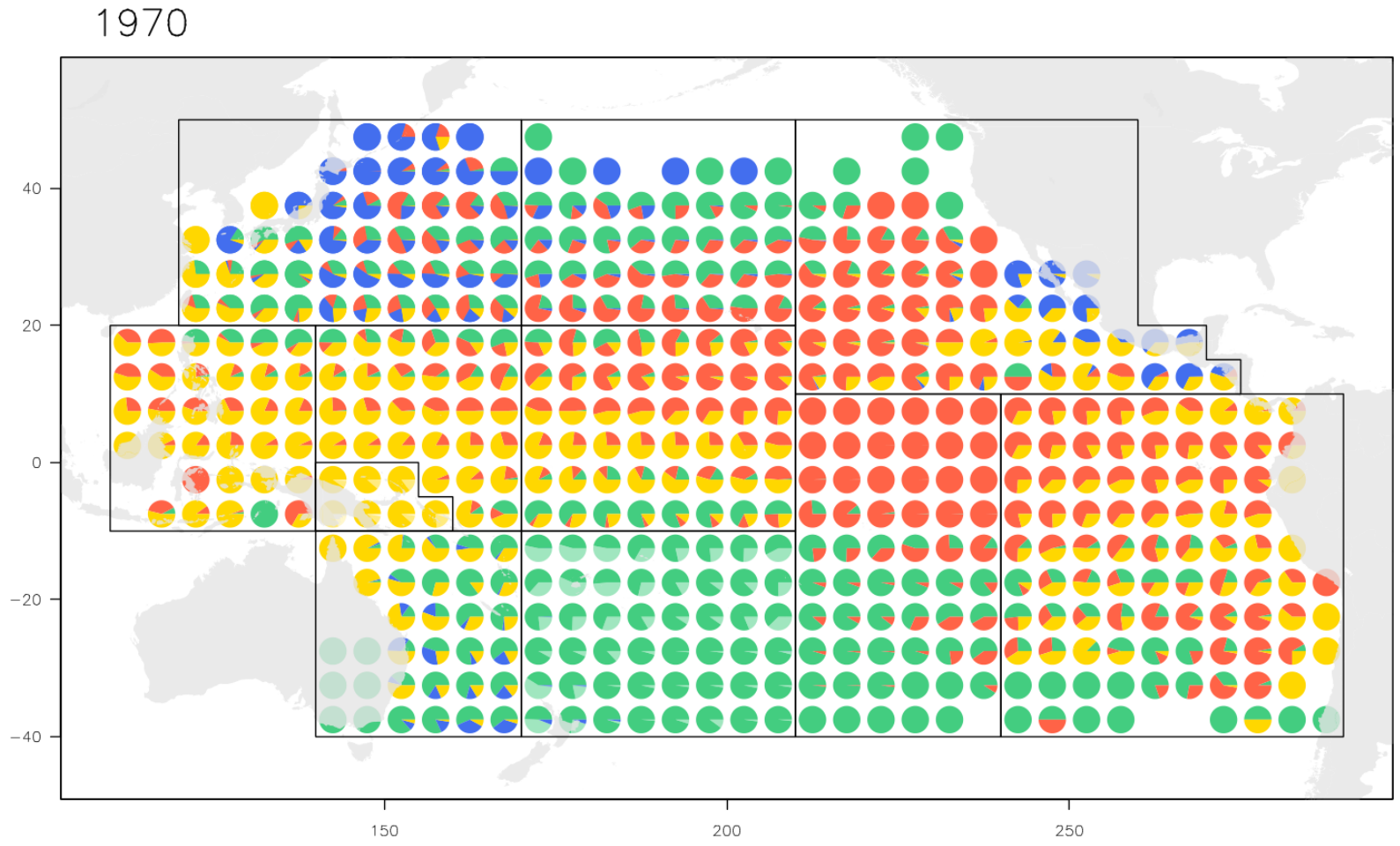


Figure 56: Map of cluster composition at the 5 degree scale, by decade, with region-specific cluster numbers as described in Table 2. ALB, BET, SWO and YFT are given by green, red, blue and yellow, respectively.

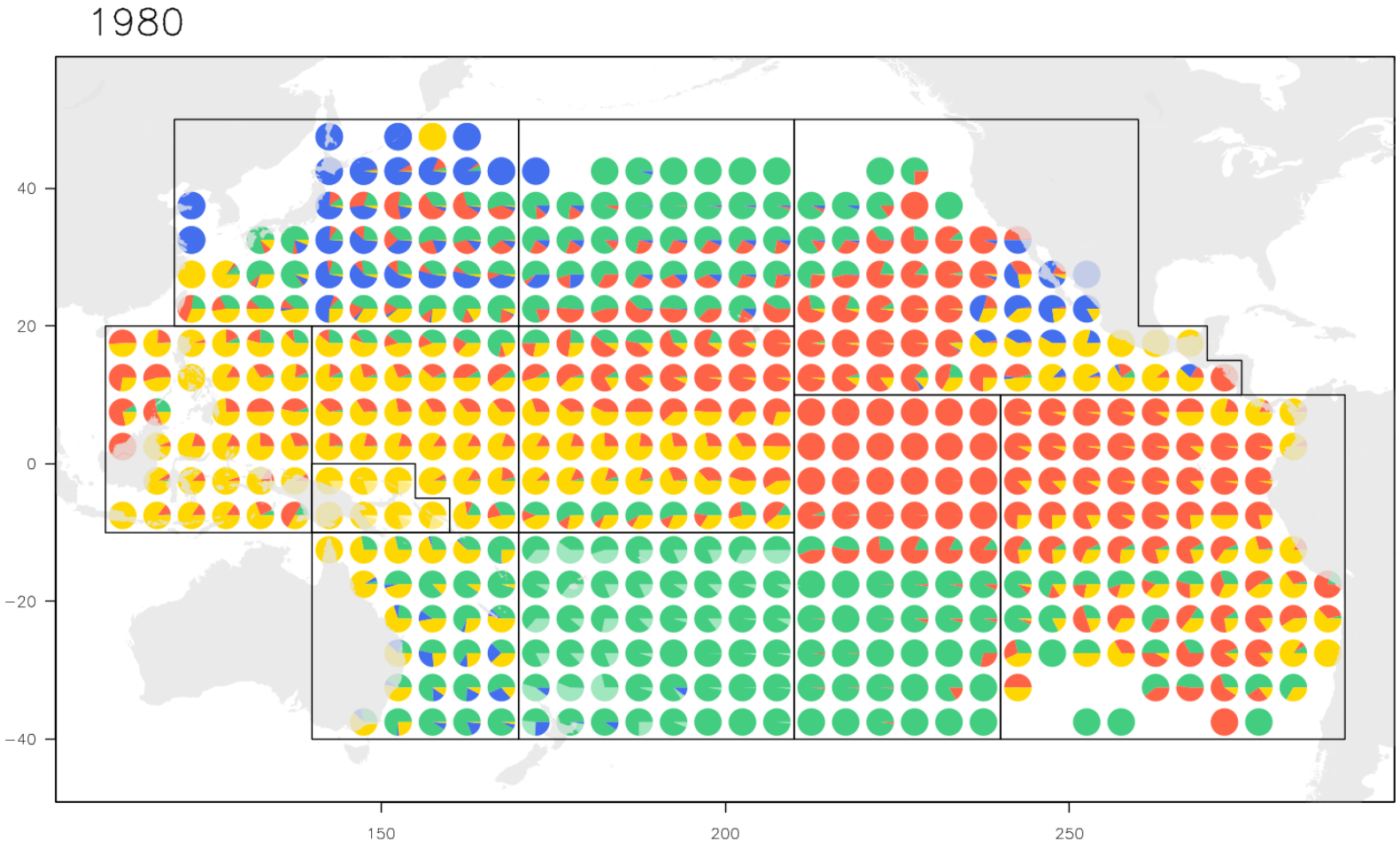


Figure 57: Map of cluster composition at the 5 degree scale, by decade, with region-specific cluster numbers as described in Table 2. ALB, BET, SWO and YFT are given by green, red, blue and yellow, respectively.

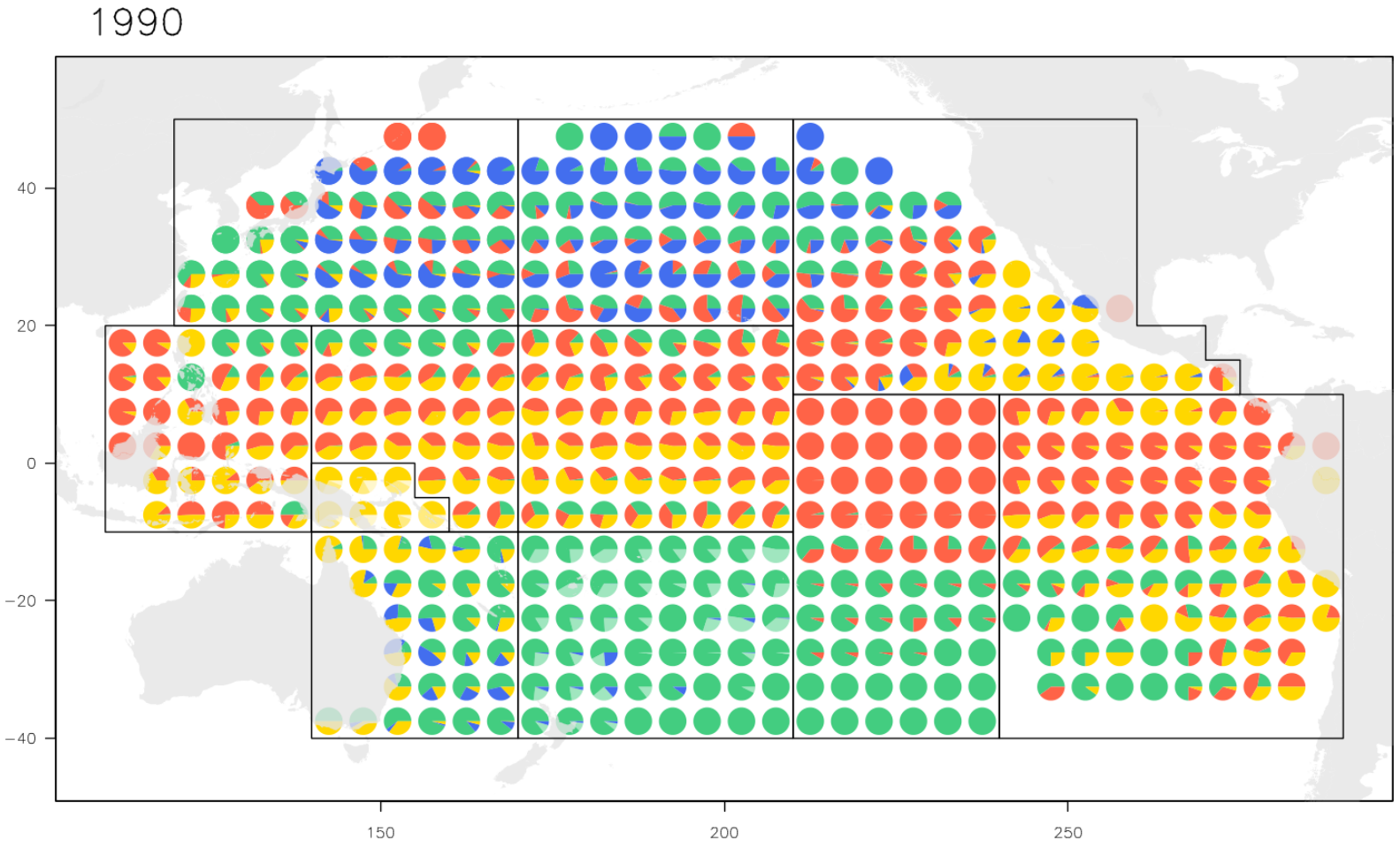


Figure 58: Map of cluster composition at the 5 degree scale, by decade, with region-specific cluster numbers as described in Table 2. ALB, BET, SWO and YFT are given by green, red, blue and yellow, respectively.

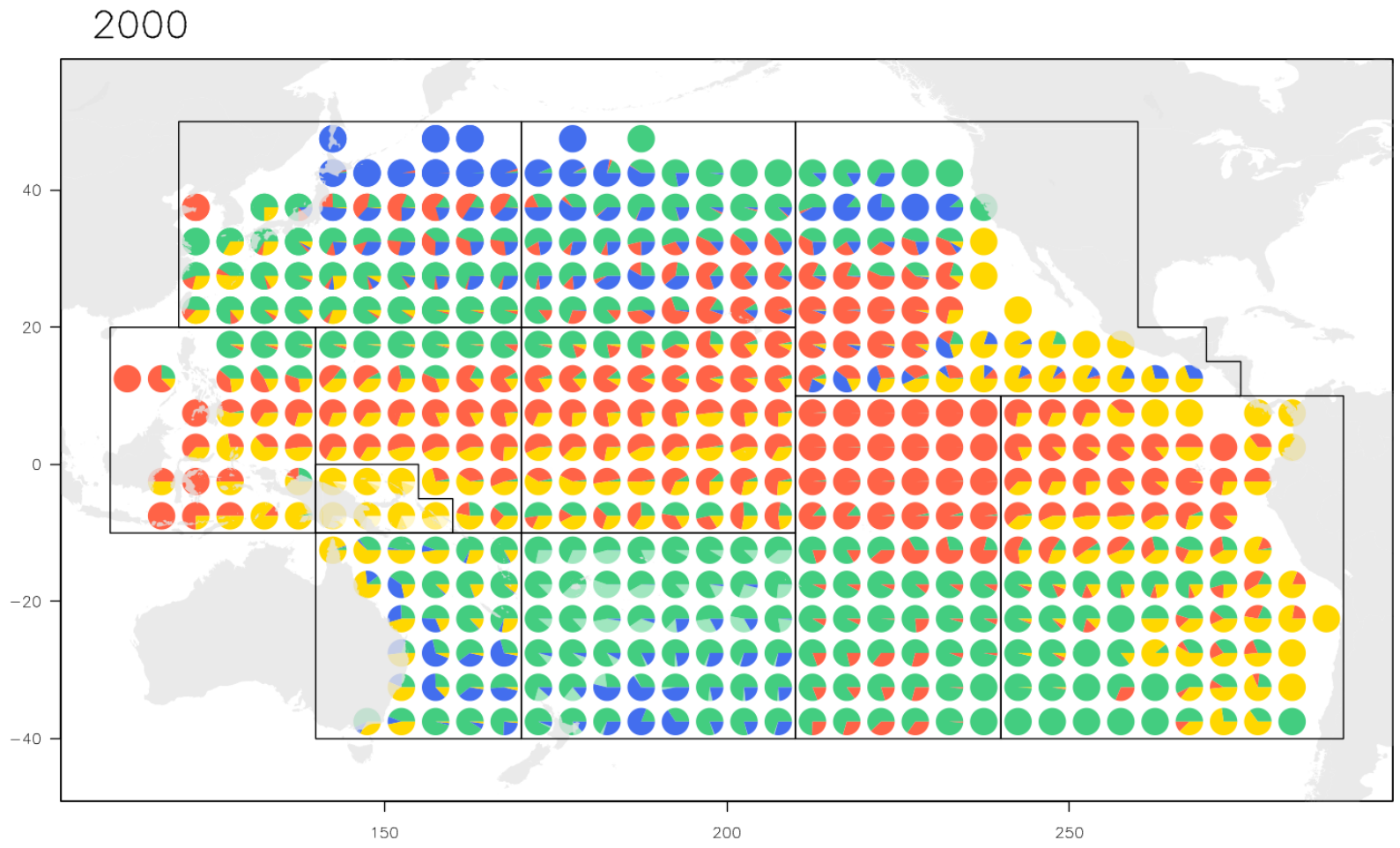


Figure 59: Map of cluster composition at the 5 degree scale, by decade, with region-specific cluster numbers as described in Table 2. ALB, BET, SWO and YFT are given by green, red, blue and yellow, respectively.

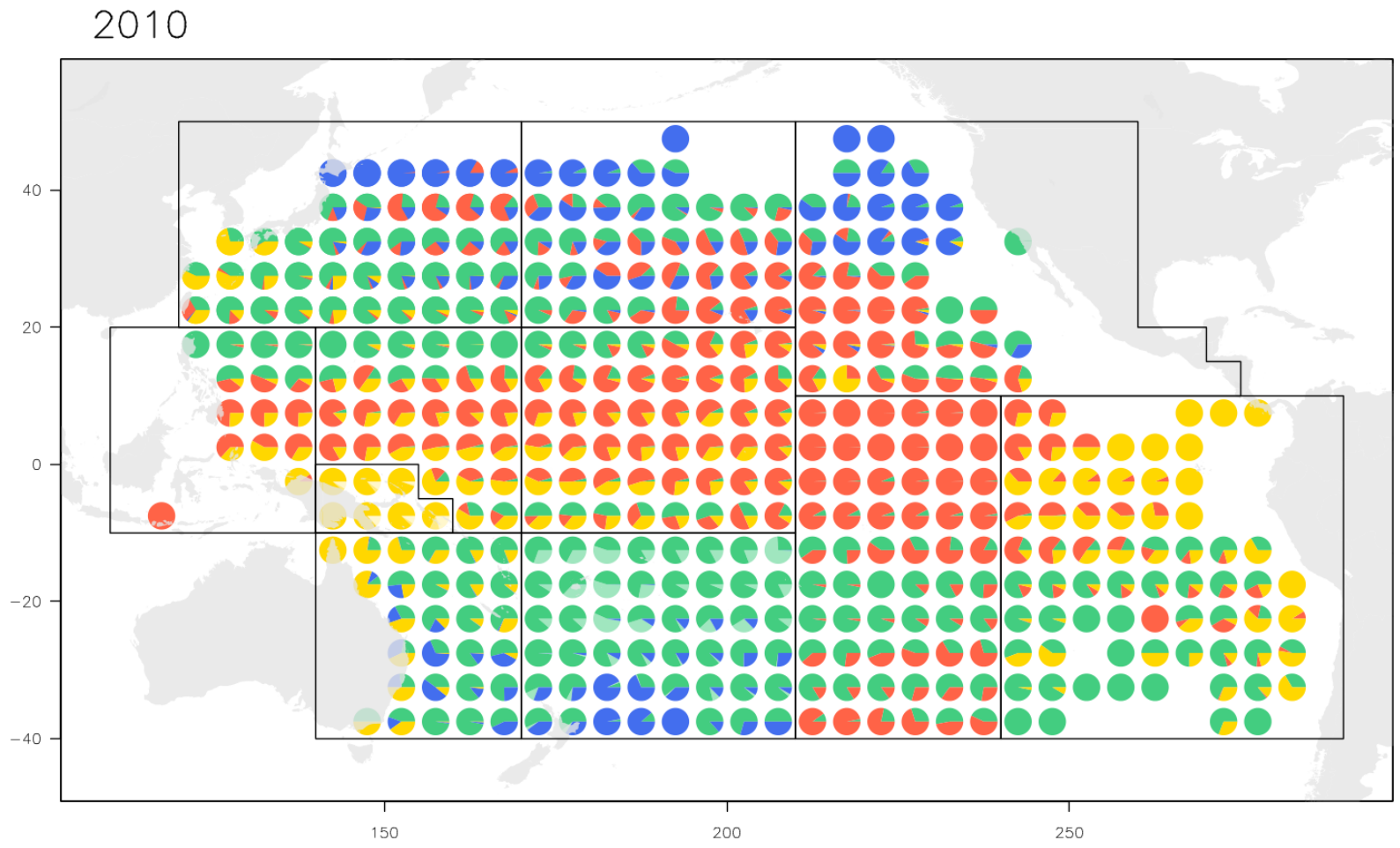


Figure 60: Map of cluster composition at the 5 degree scale, by decade, with region-specific cluster numbers as described in Table 2. ALB, BET, SWO and YFT are given by green, red, blue and yellow, respectively.

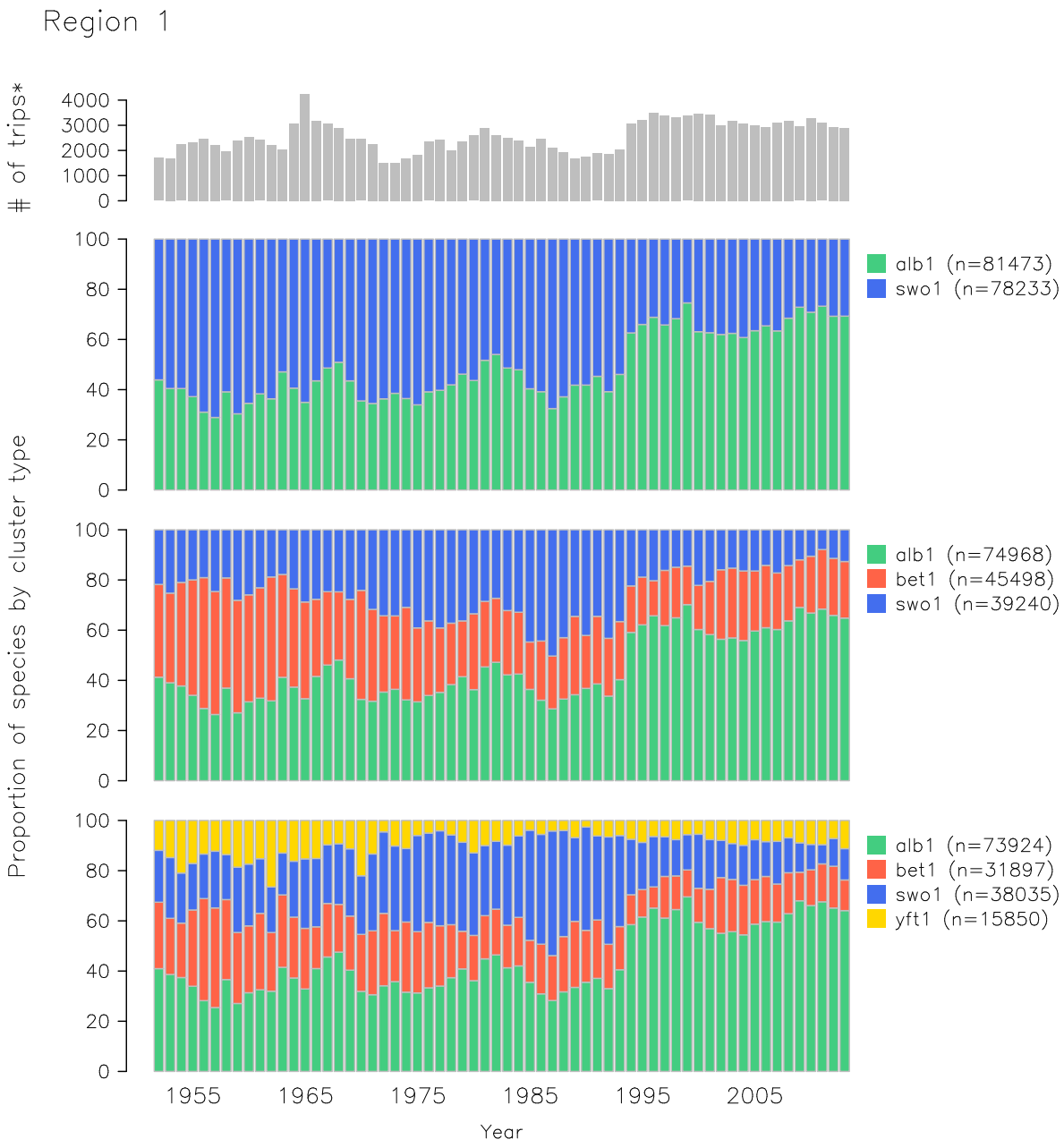


Figure 61: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

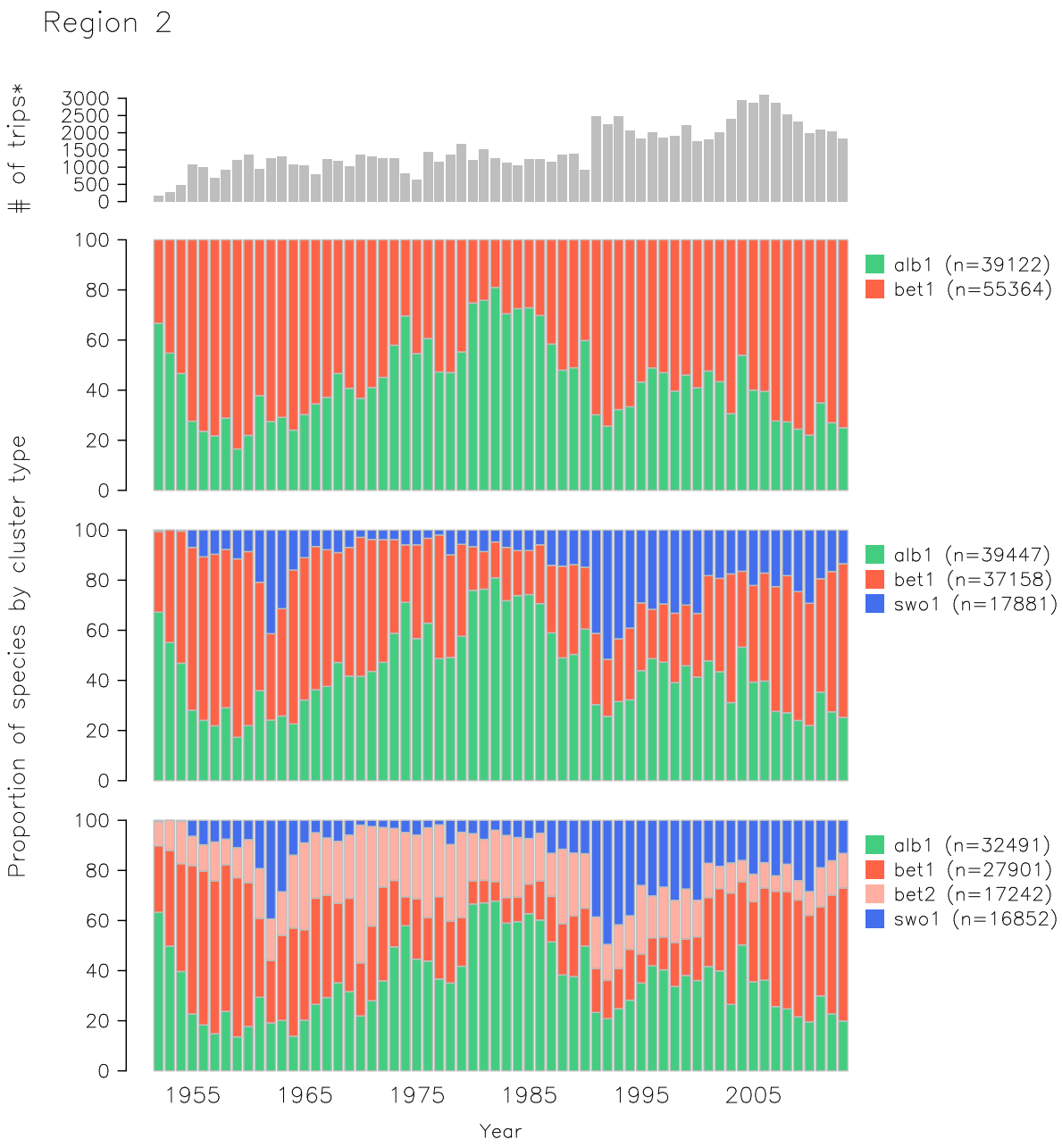


Figure 62: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

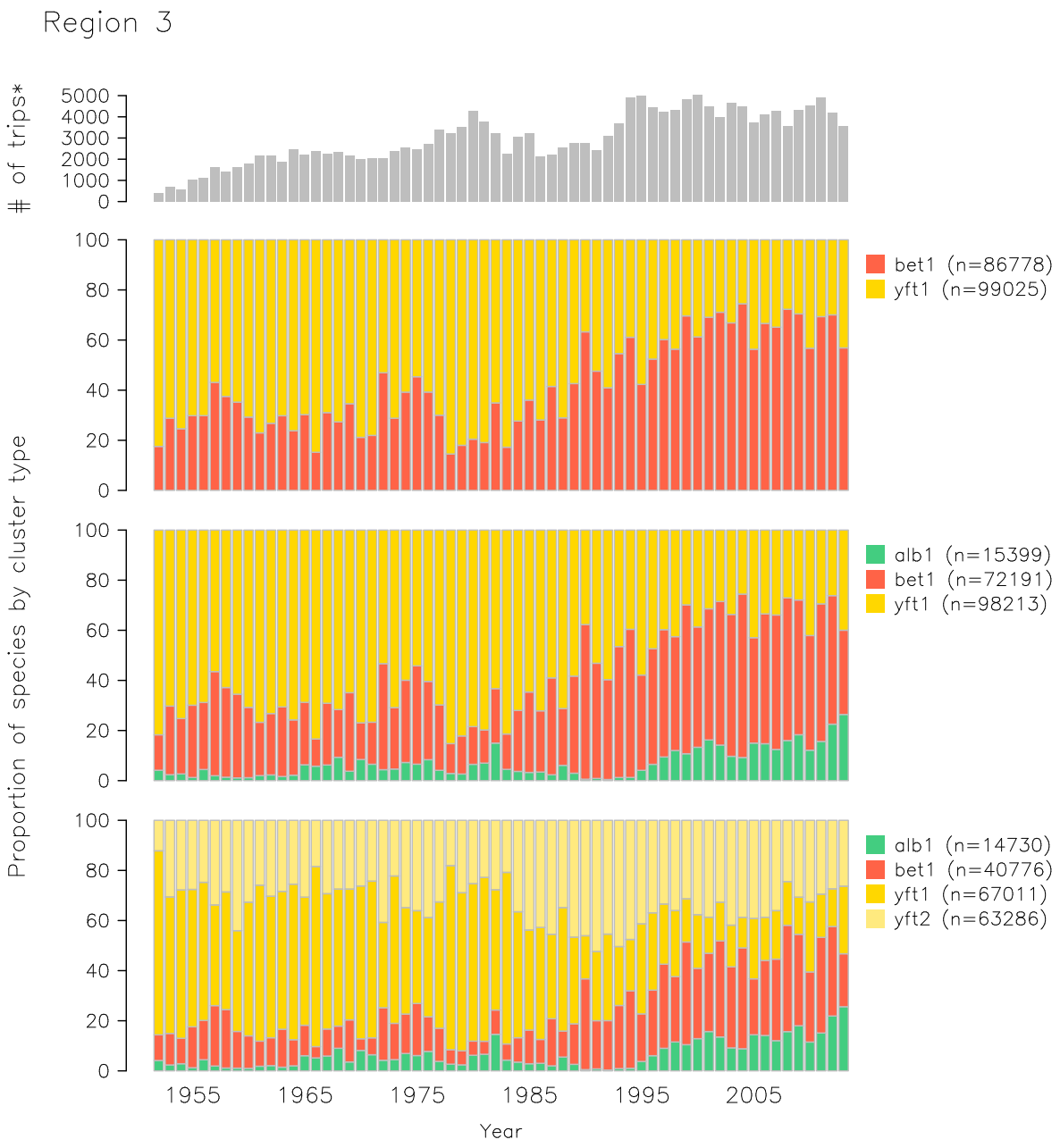


Figure 63: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

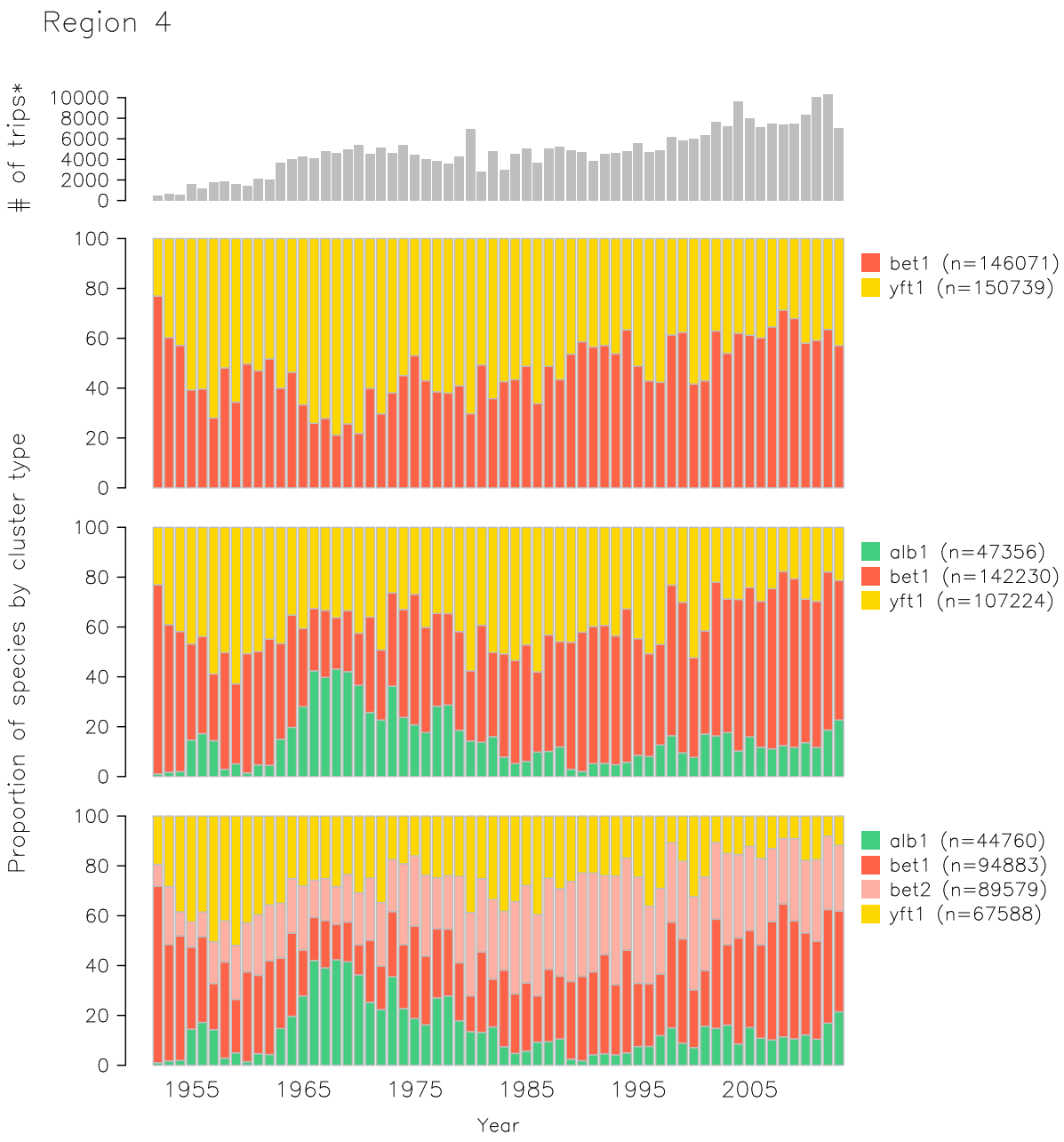


Figure 64: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

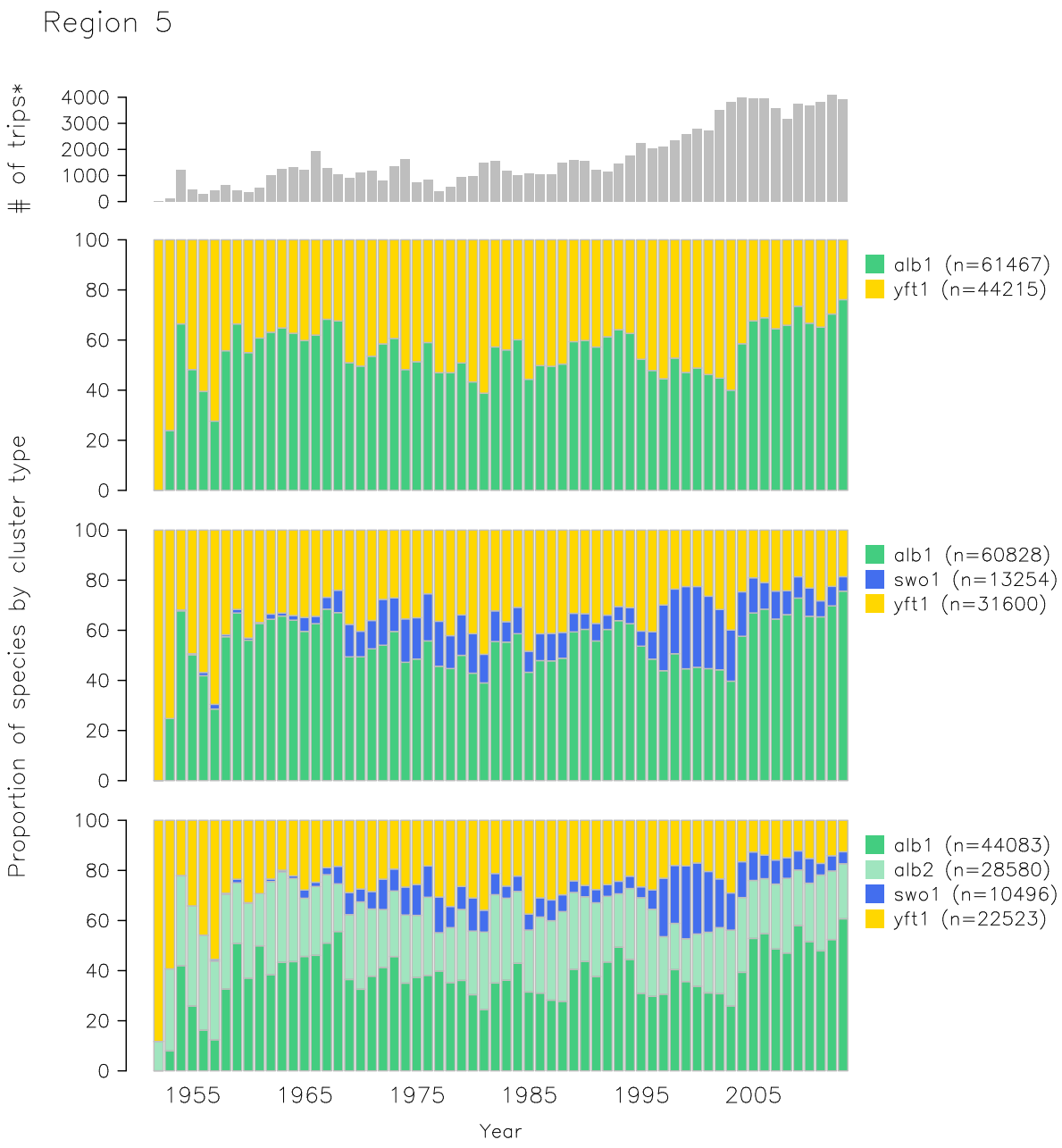


Figure 65: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

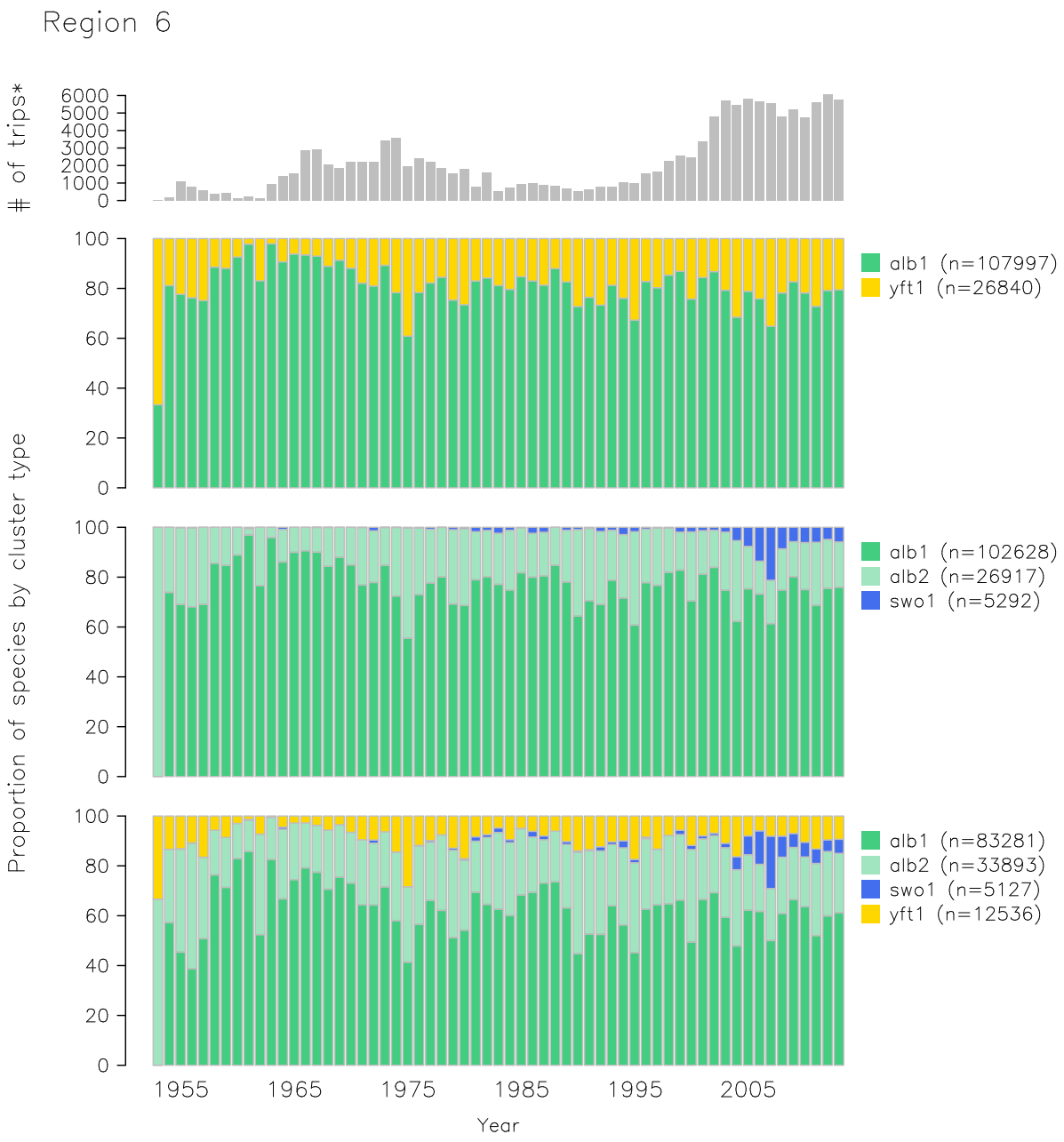


Figure 66: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

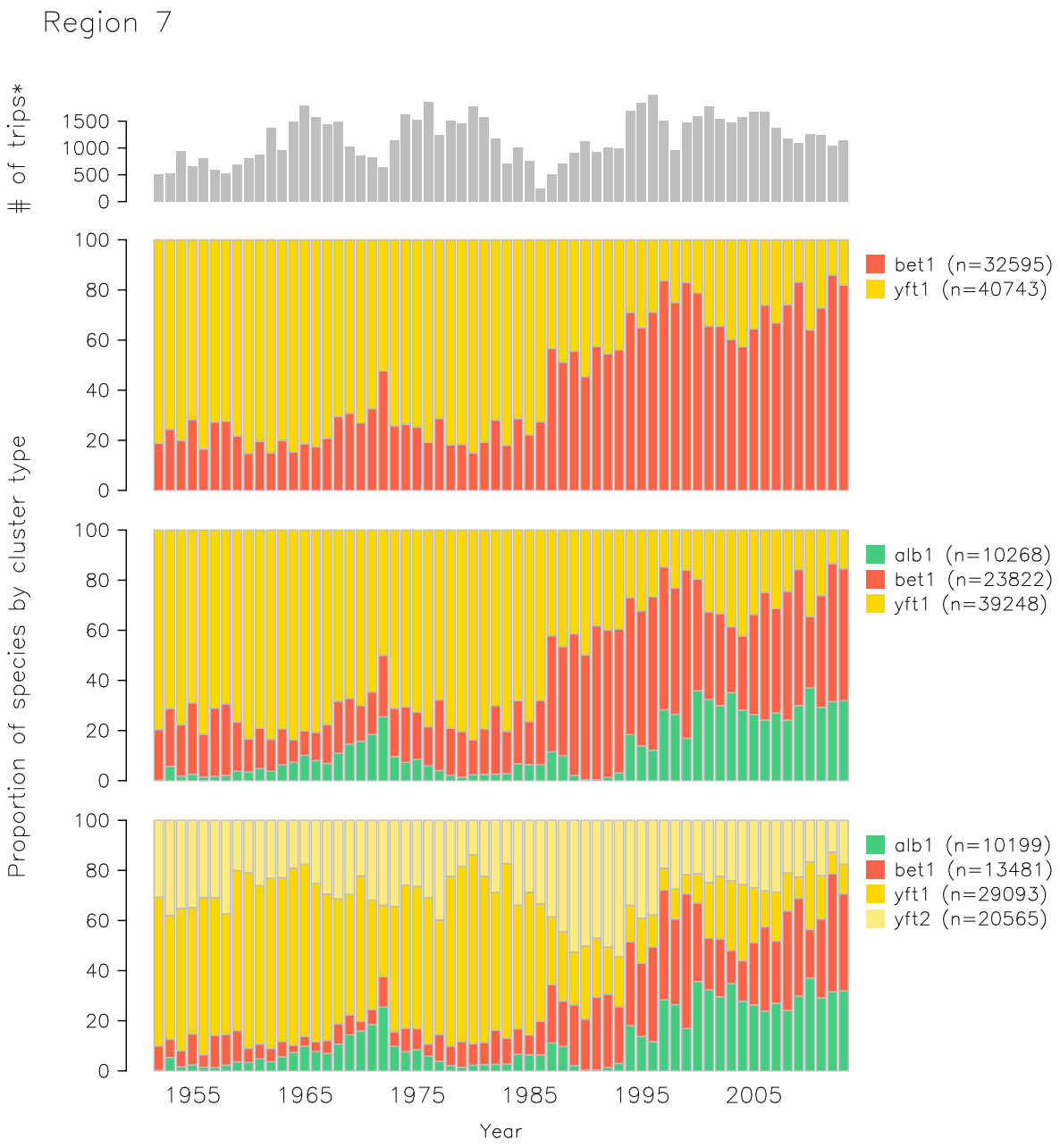


Figure 67: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

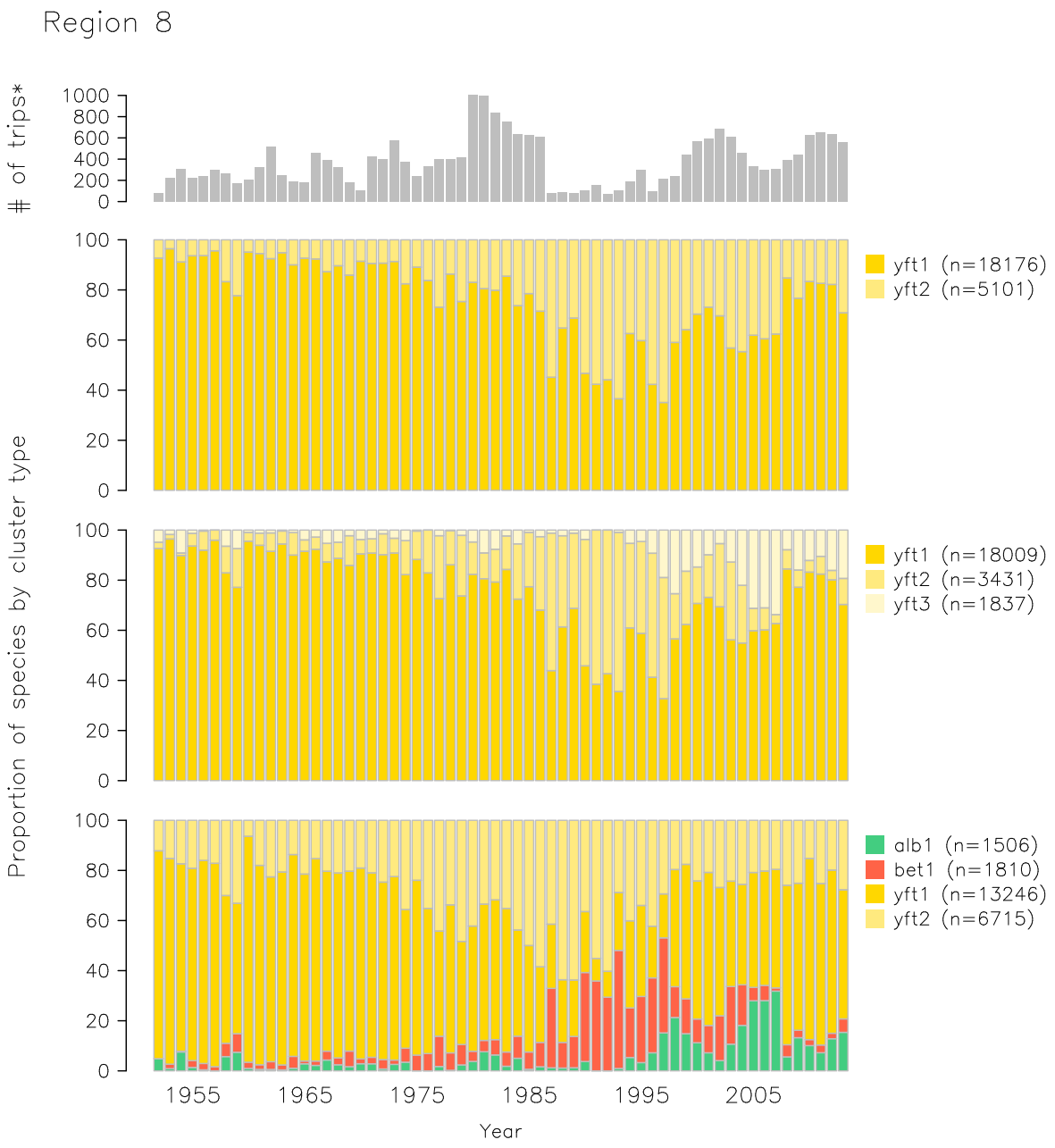


Figure 68: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

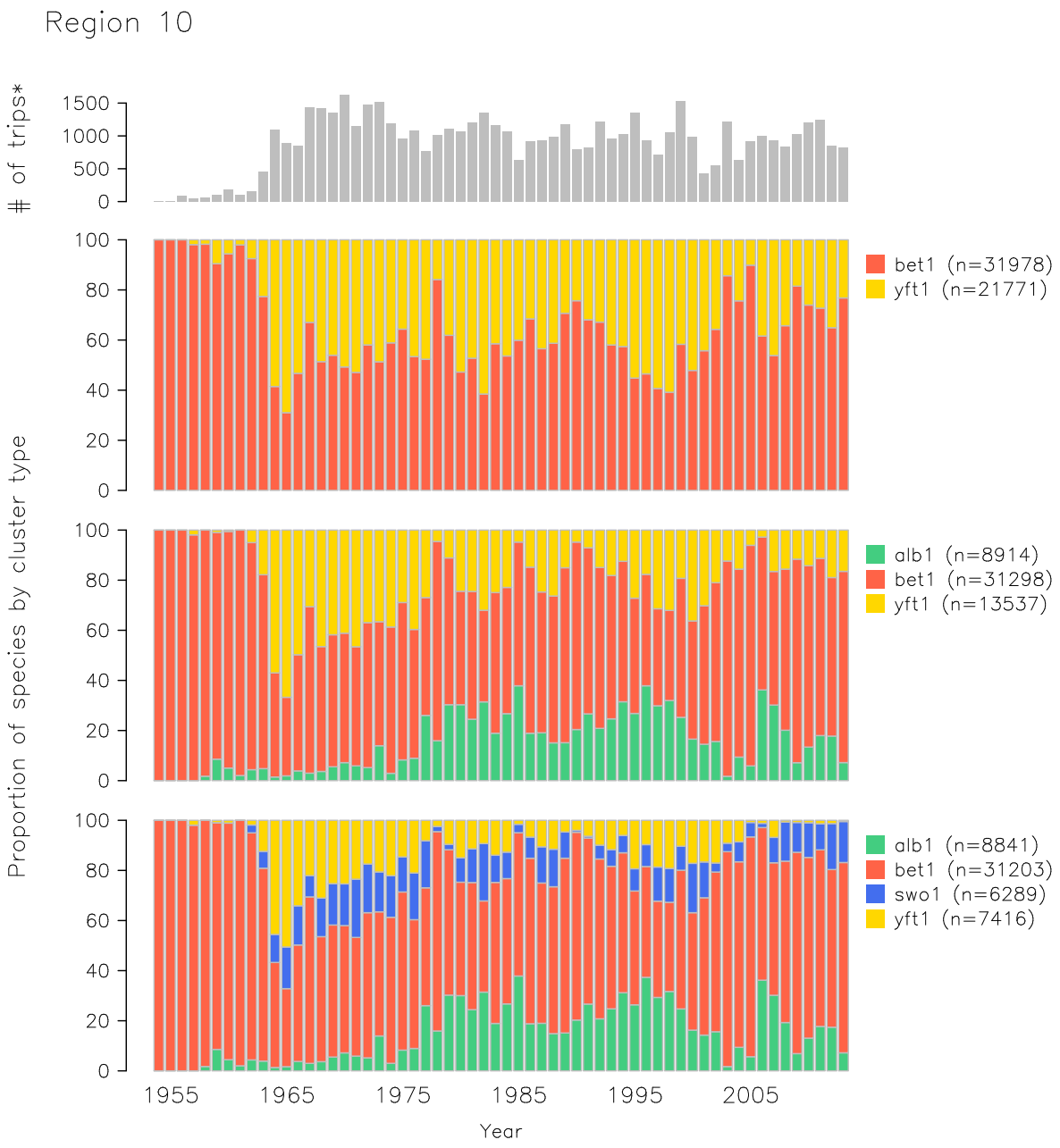


Figure 69: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

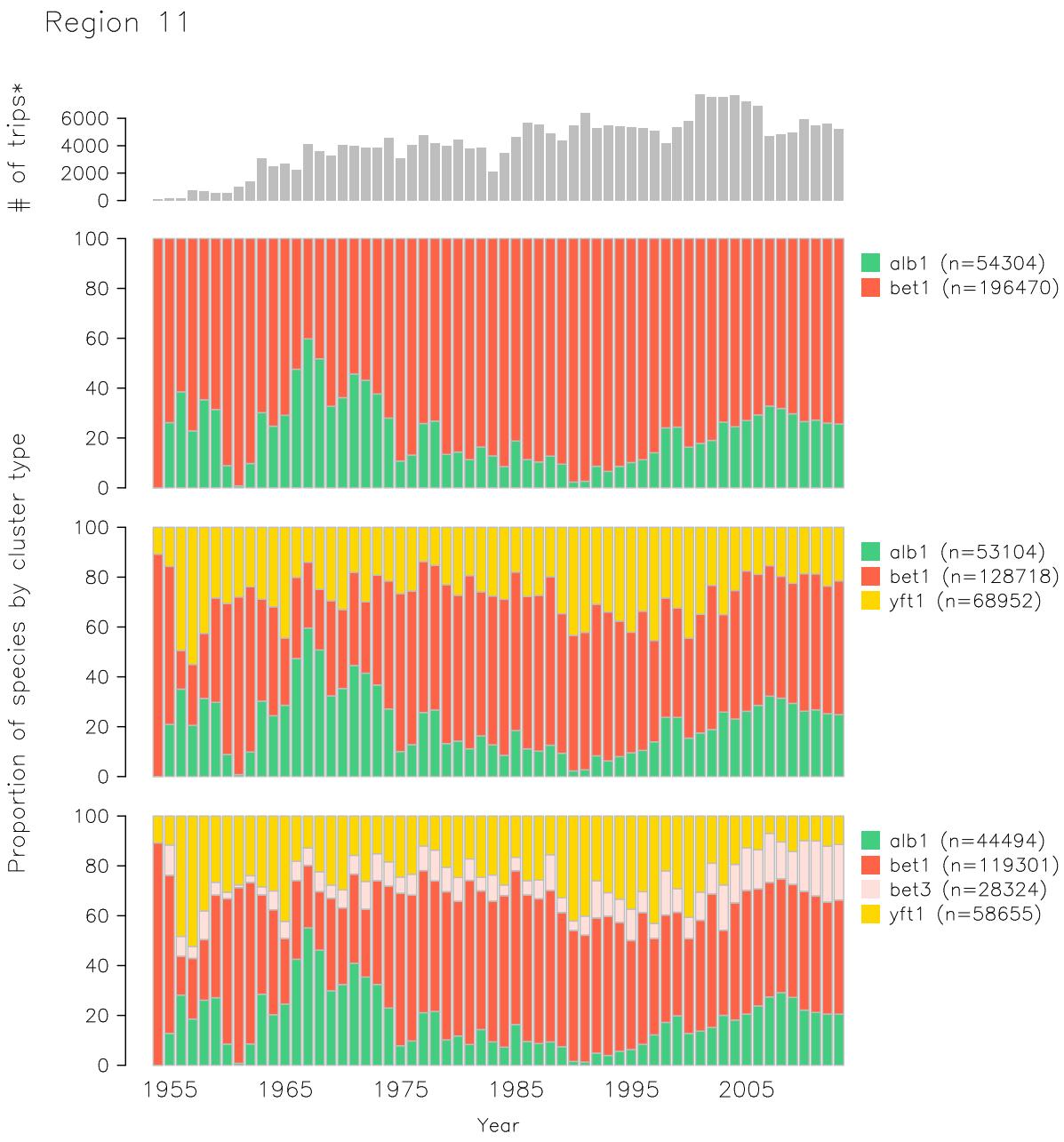


Figure 70: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

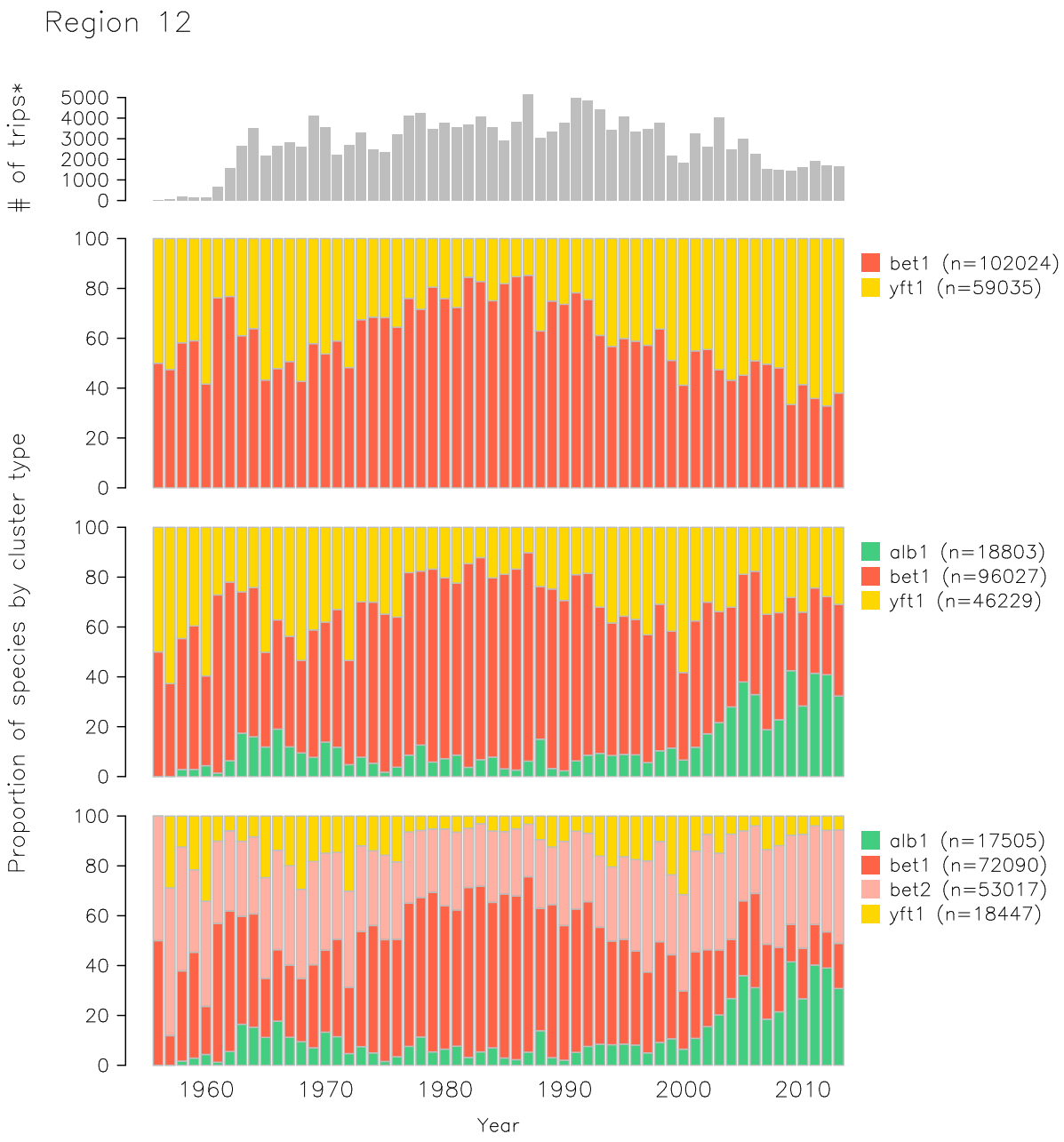


Figure 71: Time series of cluster membership for the 2, 3, and 4 cluster models, with the colour matching the dominant species in the cluster and the top panel indicating the number of records over time.

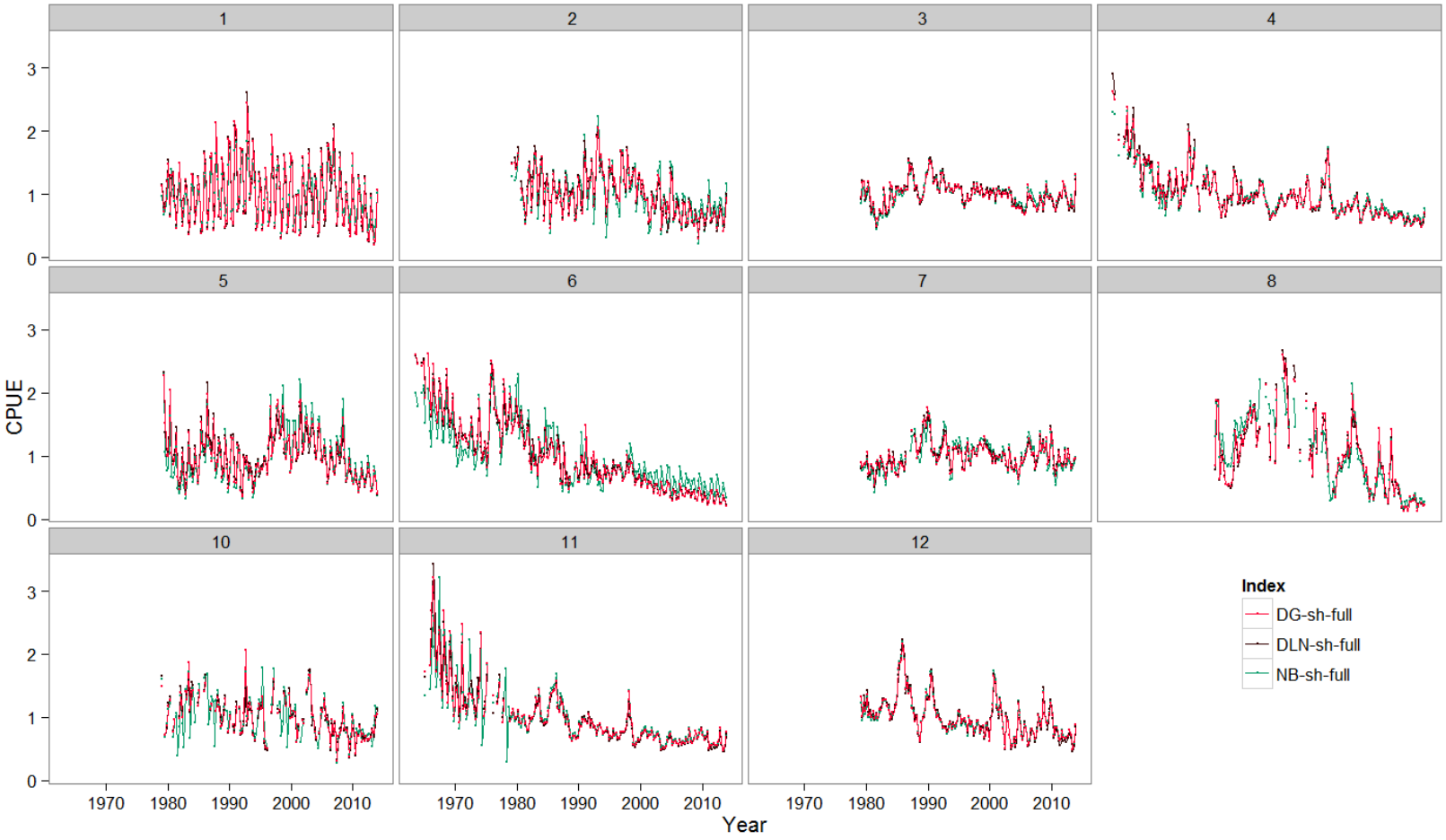


Figure 72: Comparison of standardised indices for the three different models (NB, DG and DLN) for the full short dataset (no cluster data removed) at the year-quarter scale.


Figure 73: Comparison of standardised indices for the full short dataset (no cluster data removed) with the indices for the short dataset with cluster data removed, at the year-quarter scale. Both models include the cluster variable in the linear predictor.

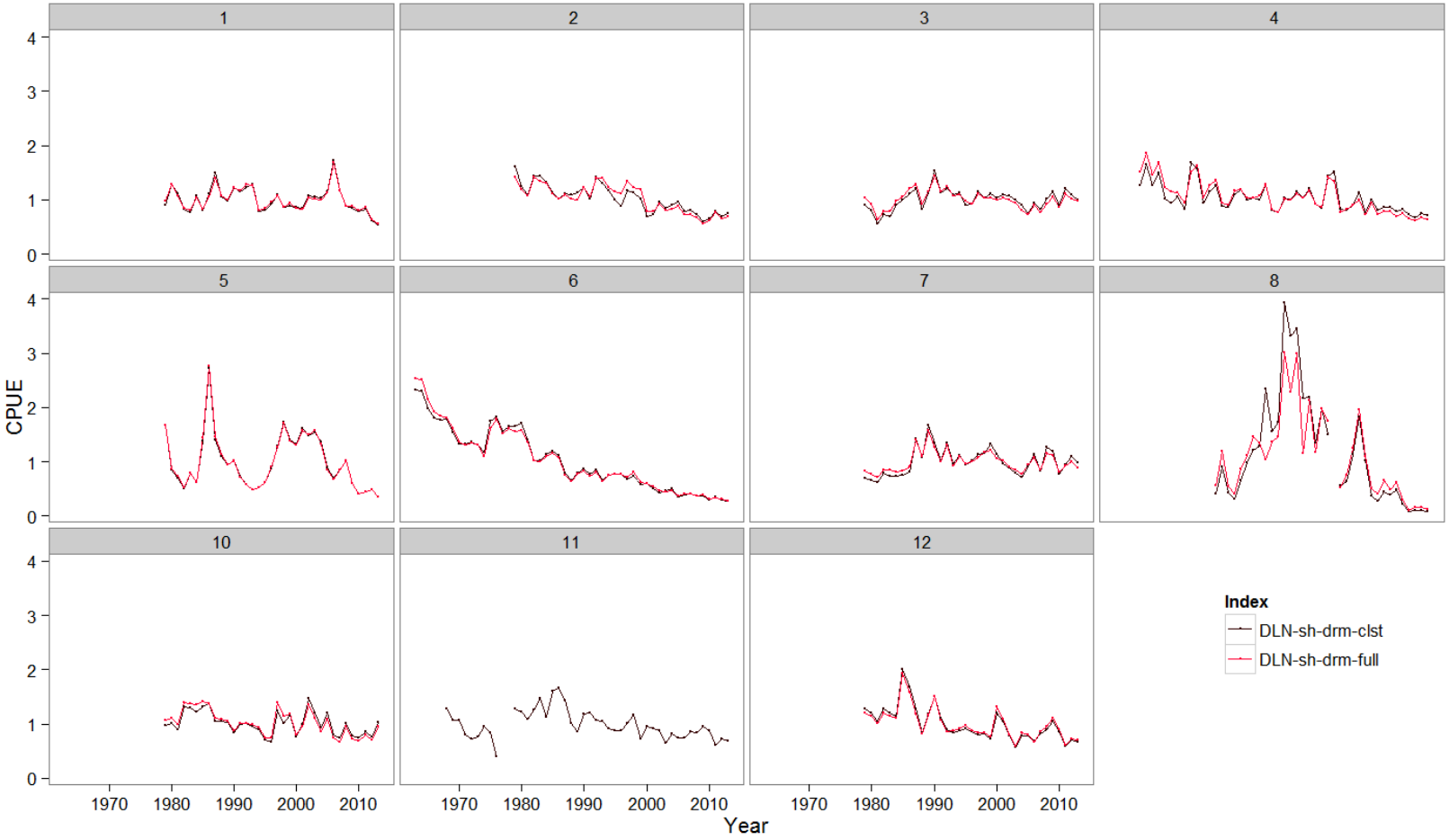


Figure 74: Comparison of standardised DLN indices with and without the cluster variable included, and the nominal indices, for the short dataset with cluster data removed, at the year scale. Note that R11 does not have a 'full' model as only one cluster remains in the dataset after data removal.


Figure 75: Comparison of standardised DLN indices with and without the vessel variable included, and the nominal indices, for the short dataset with cluster data removed, at the year scale. Note that R11 does not have a 'full' model as only one cluster remains in the dataset.

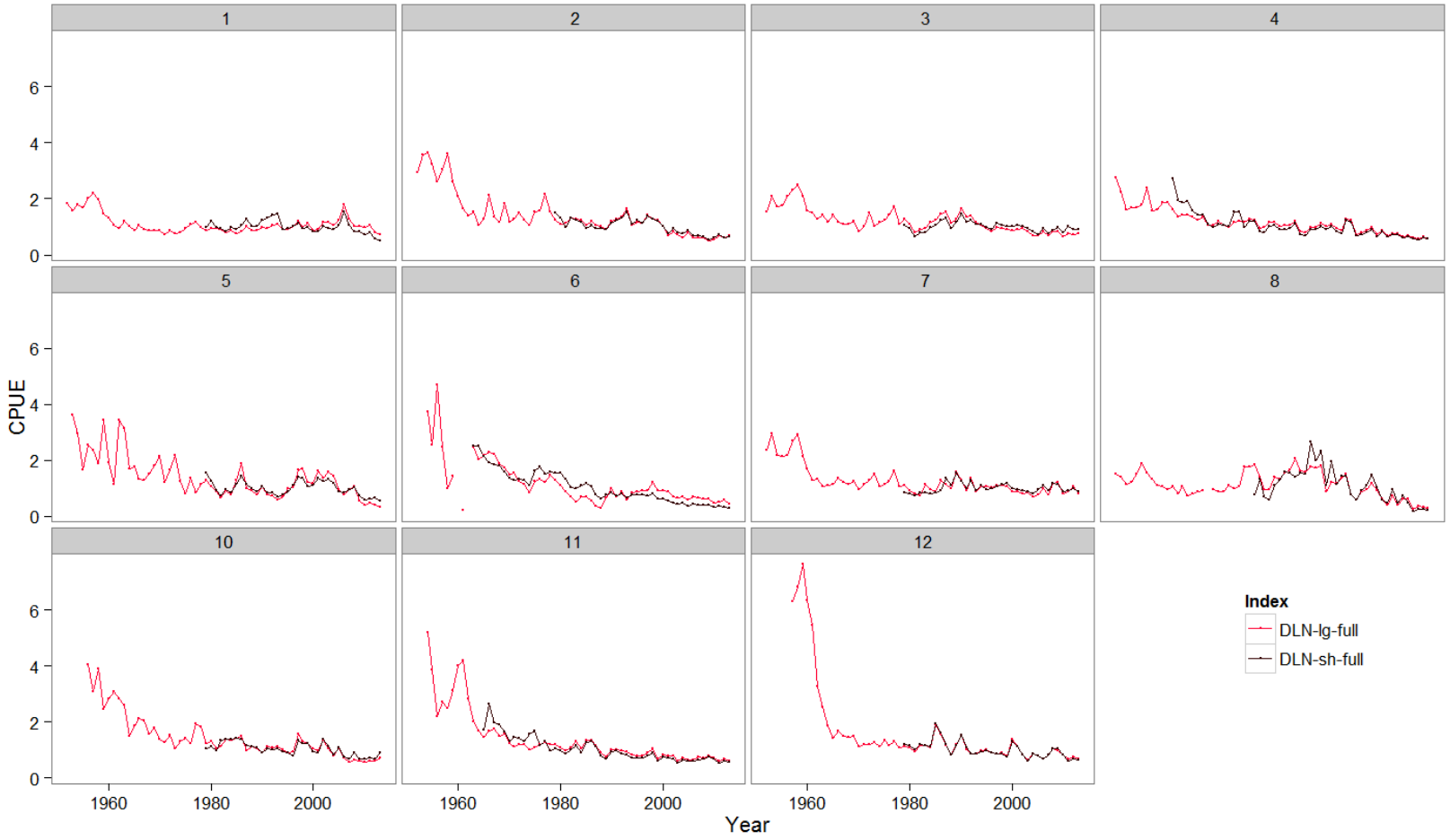


Figure 76: Comparison of standardised DLN indices for the short and long datasets (no cluster data removed) at the year scale, normalized to have a mean of 1 over the time period of the shorter indices.

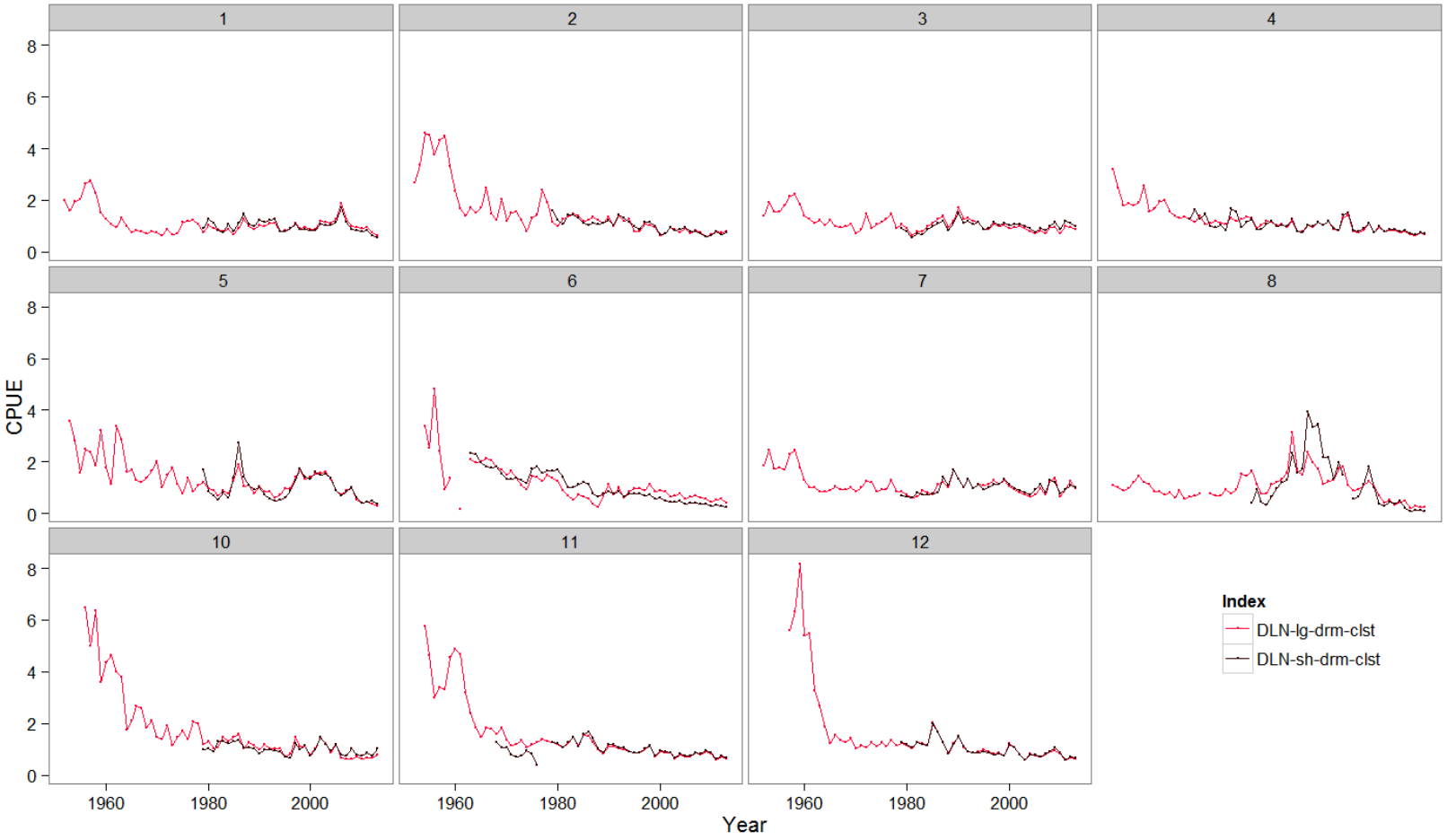


Figure 77: Comparison of standardised DLN indices for the short and long datasets (cluster data removed) at the year scale, normalized to have a mean of 1 over the time period of the shorter indices.

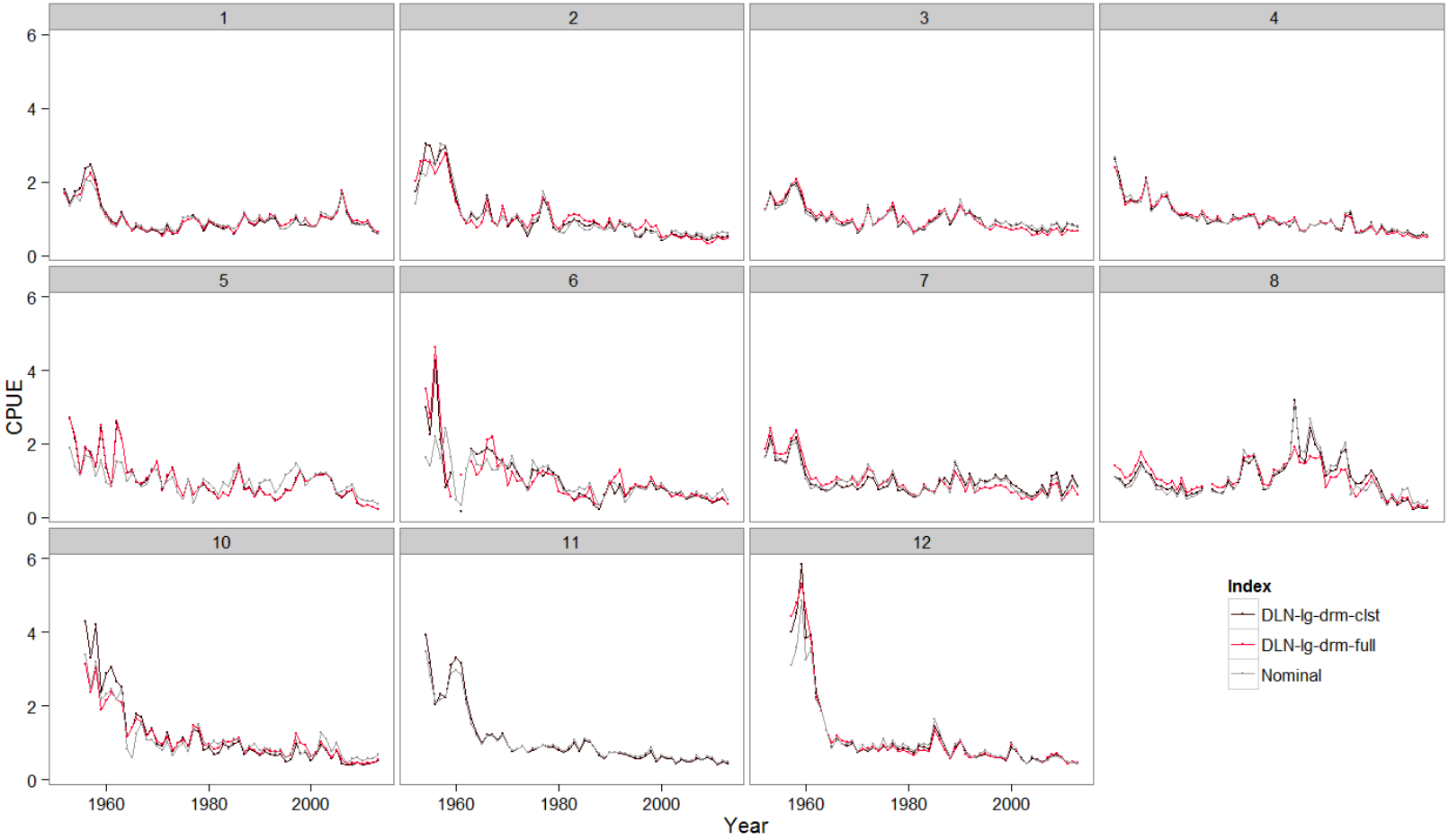


Figure 78: Comparison of standardised DLN indices with and without the cluster variable included, and the nominal indices, for the long dataset with cluster data removed, at the year scale. Note that R11 does not have a 'full' model as only one cluster remains in the dataset after data removal

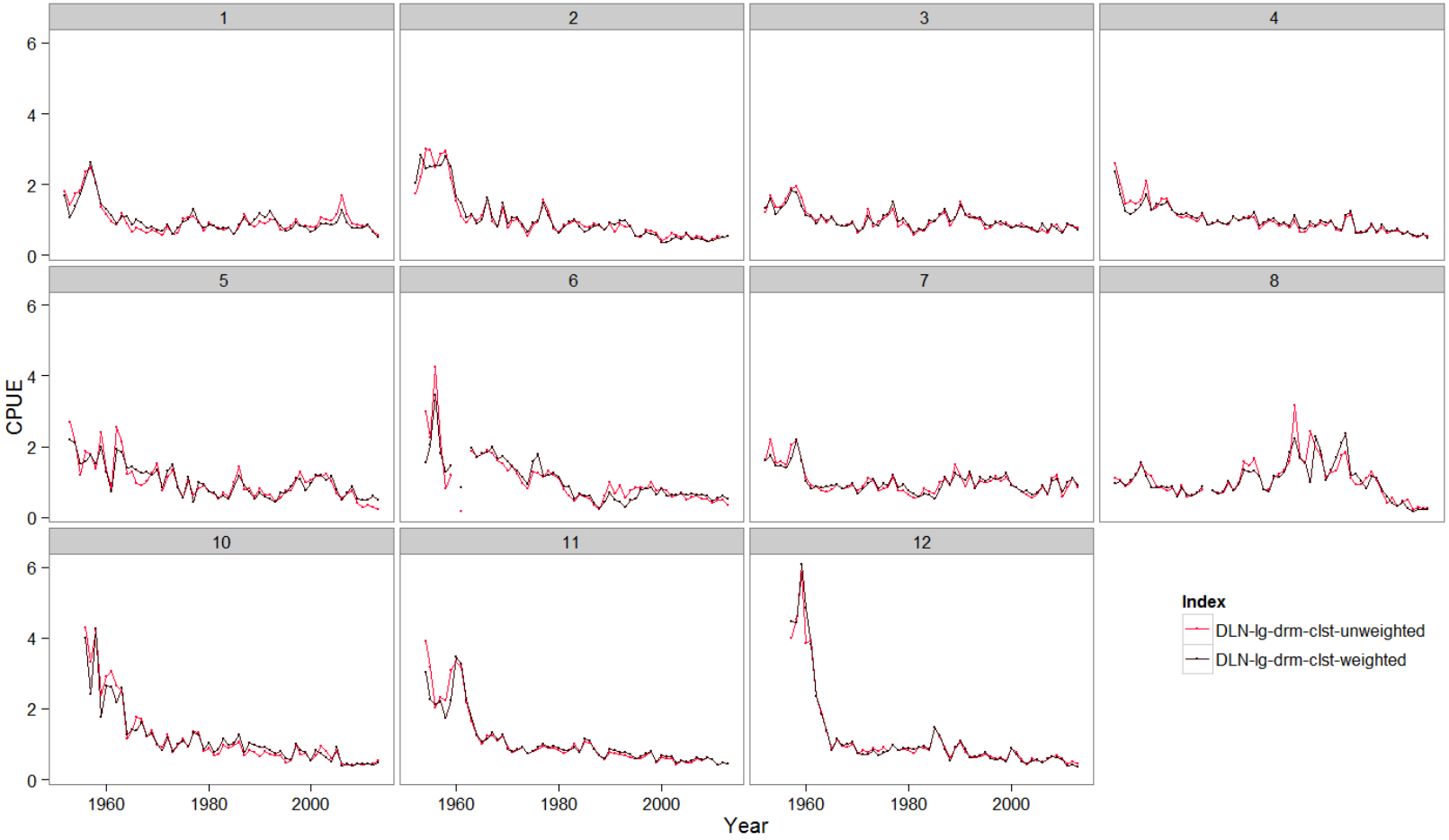


Figure 79: Comparison of standardised DLN indices for the long dataset (cluster data removed) where data is either unweighted or is weighted based on sample sizes in spatio-temporal strata. Indices are presented at the year scale, normalized to have a mean of 1.

## Diagnostic plots

The plots presented below display a range of diagnostics that describe the fit of the GLM models and the features of the model that affect the standardisation of the raw CPUE data. There are two figures for each model, a figure displaying the indices and their determinants (summary plots), and a figure showing residuals and simulated data, where appropriate (diagnostic plots). The indices figure is the same for all GLM families and displays the nominal and standardised indices in the top left panel (note that for binomial, normal and gamma the standardised index is the DLN, DLN and DG, respectively), the rest of the plots in the left column are influence plots for the models categorical variables, and the right column are plots of coefficient values for these same variables. Note that the vessel coefficients are positioned on the x axis by the first year-quarter that they appeared in the dataset for that region.

The residual figures are the same for the log normal, negative binomial and gamma models, but different for the binomial models. For the former, the top left figure plots the observed distribution (blue bars) of BET counts or cpue (on the log-scale for log normal) and the distribution of data simulated from the fitted model (red lines, one for each simulated dataset). The top right plot is a normal quantile-quantile of the observed residuals. The middle left plot shows the same residuals against the models fitted values. The remaining plots show box and whiskers of the residuals for the categorical variables in the model. The binomial figure differs in that the the top left plot the observed and fitted proportions of positive counts are compared after aggregating data at the vessel-year-quarter-cell scale. Throughout, the quantile residuals are shown for the binomial, negative and binomial models and deviance residuals are shown for the log-normal and gamma models.

The first set of figures (Figures 80 – 126) denoted step 1 in the caption are for the short dataset with no cluster data removed and all variables in the model (sh-drm-full models) while the second set (Figures 127 – 170) are for the long data set with cluster data removed and no cluster variable included in the model (DLN-sh-drm-clst).

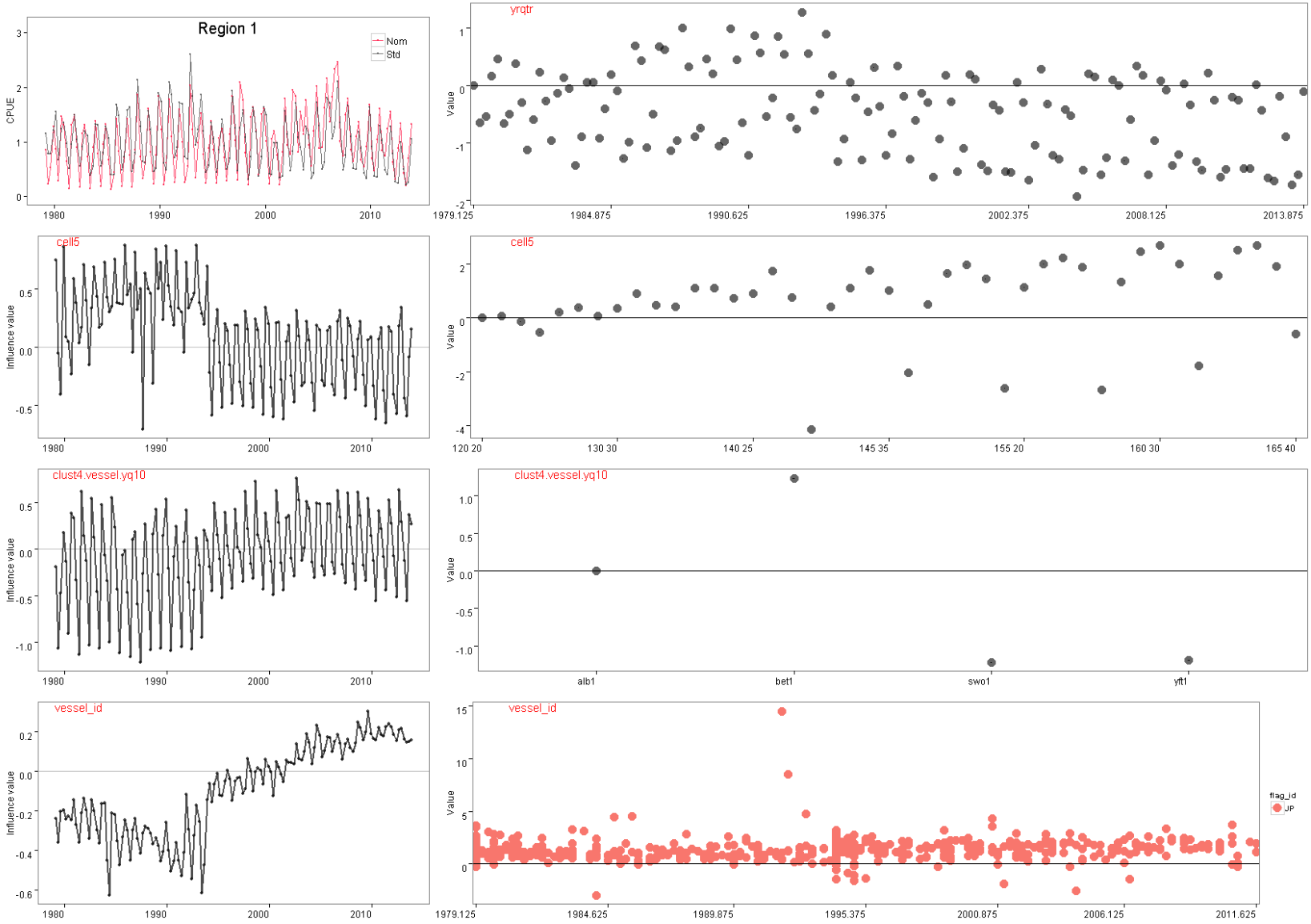


Figure 80: Summary plots of fitted GLM models for region 1, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

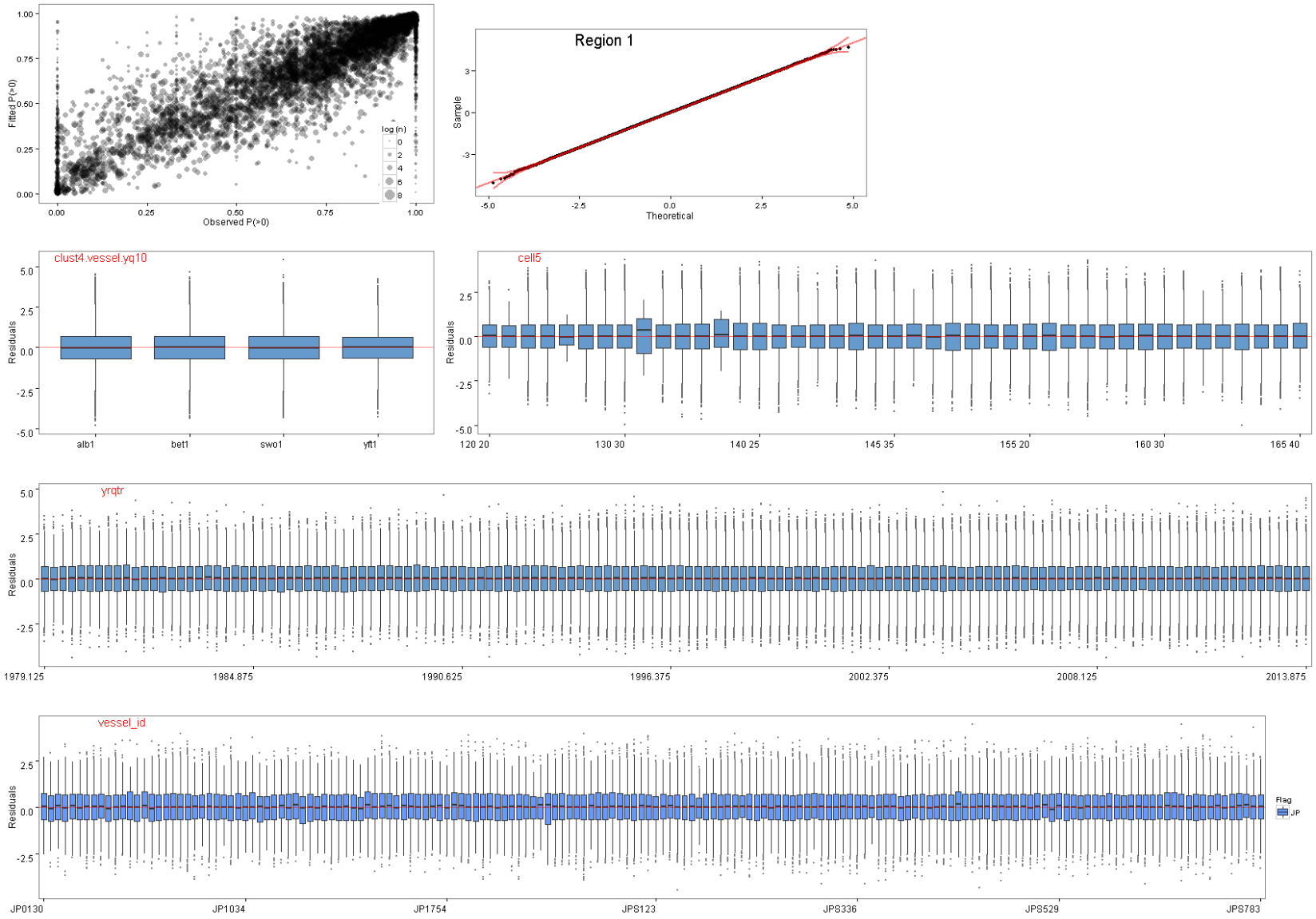


Figure 81: Diagnostic plots of fitted GLM models for region 1, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

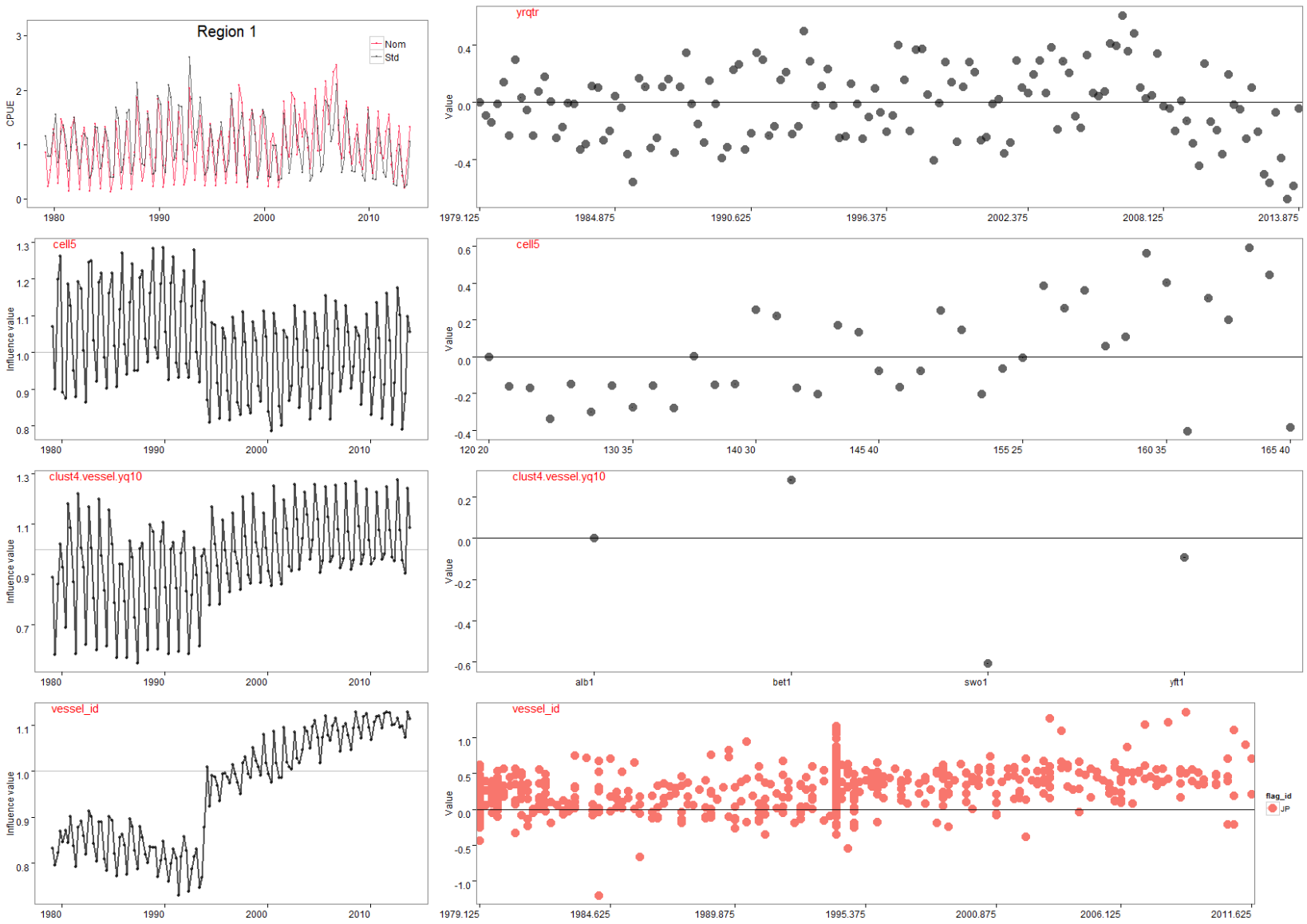


Figure 82: Summary plots of fitted GLM models for region 1, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

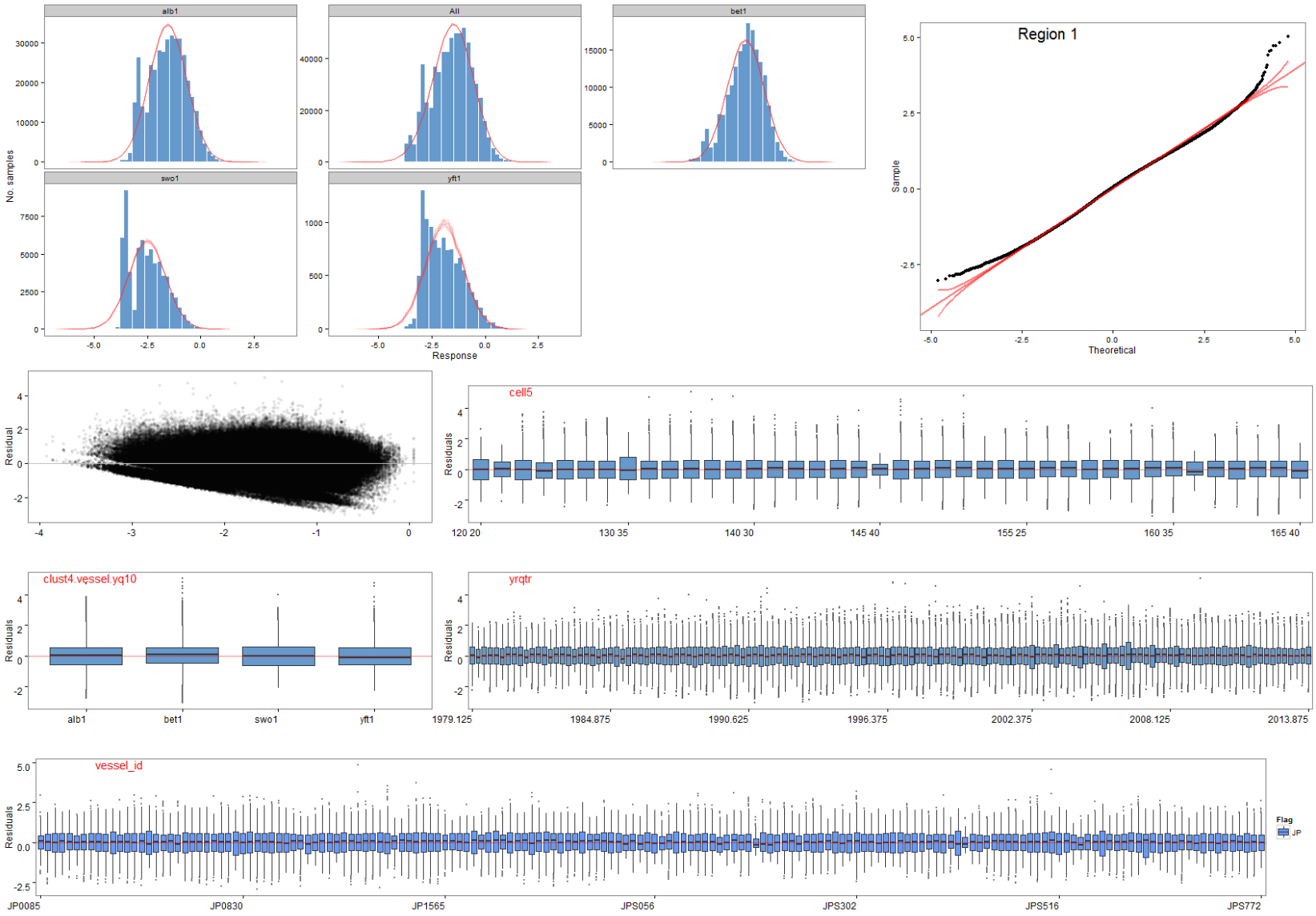


Figure 83: Diagnostic plots of fitted GLM models for region 1, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

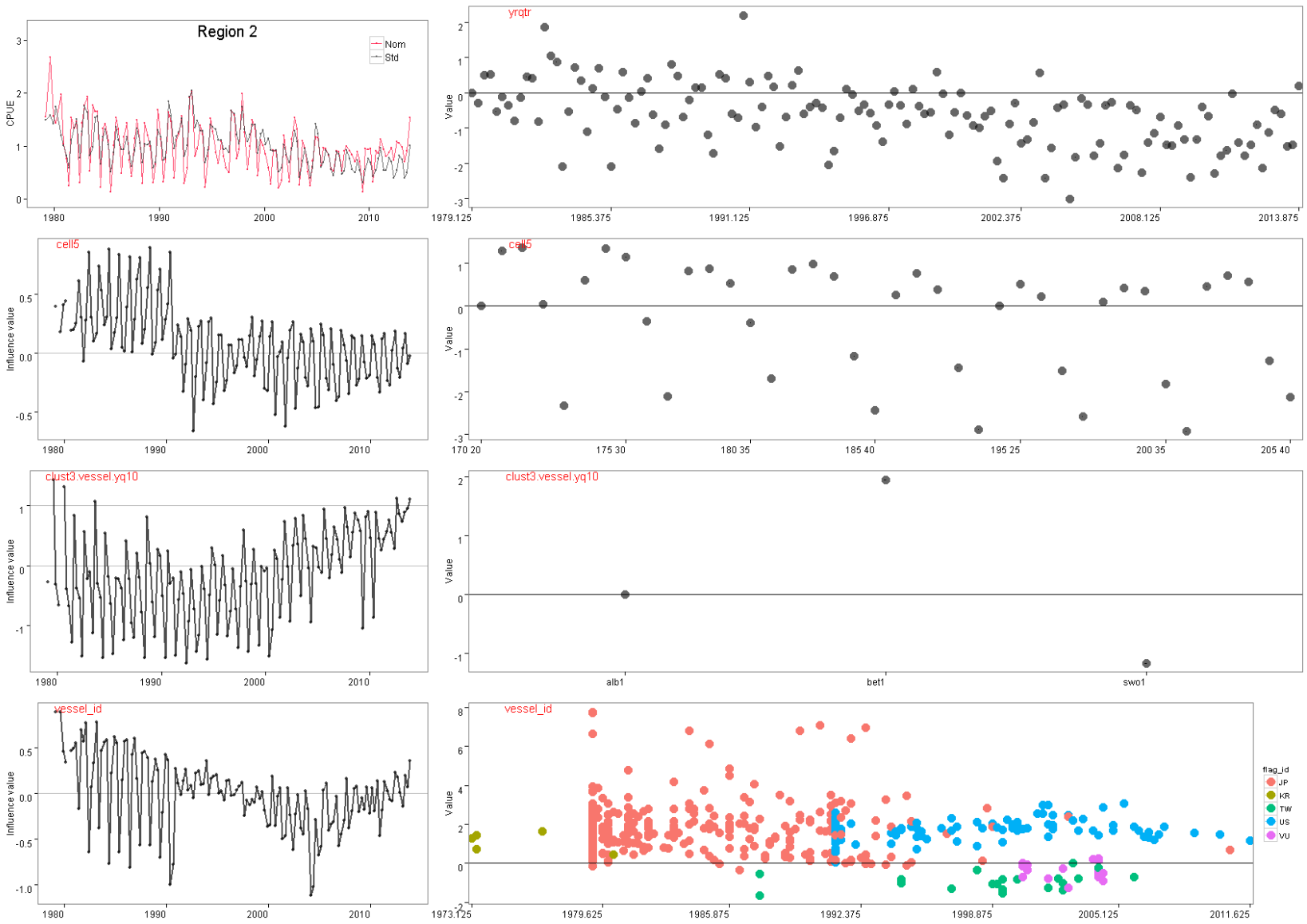


Figure 84: Summary plots of fitted GLM models for region 2, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

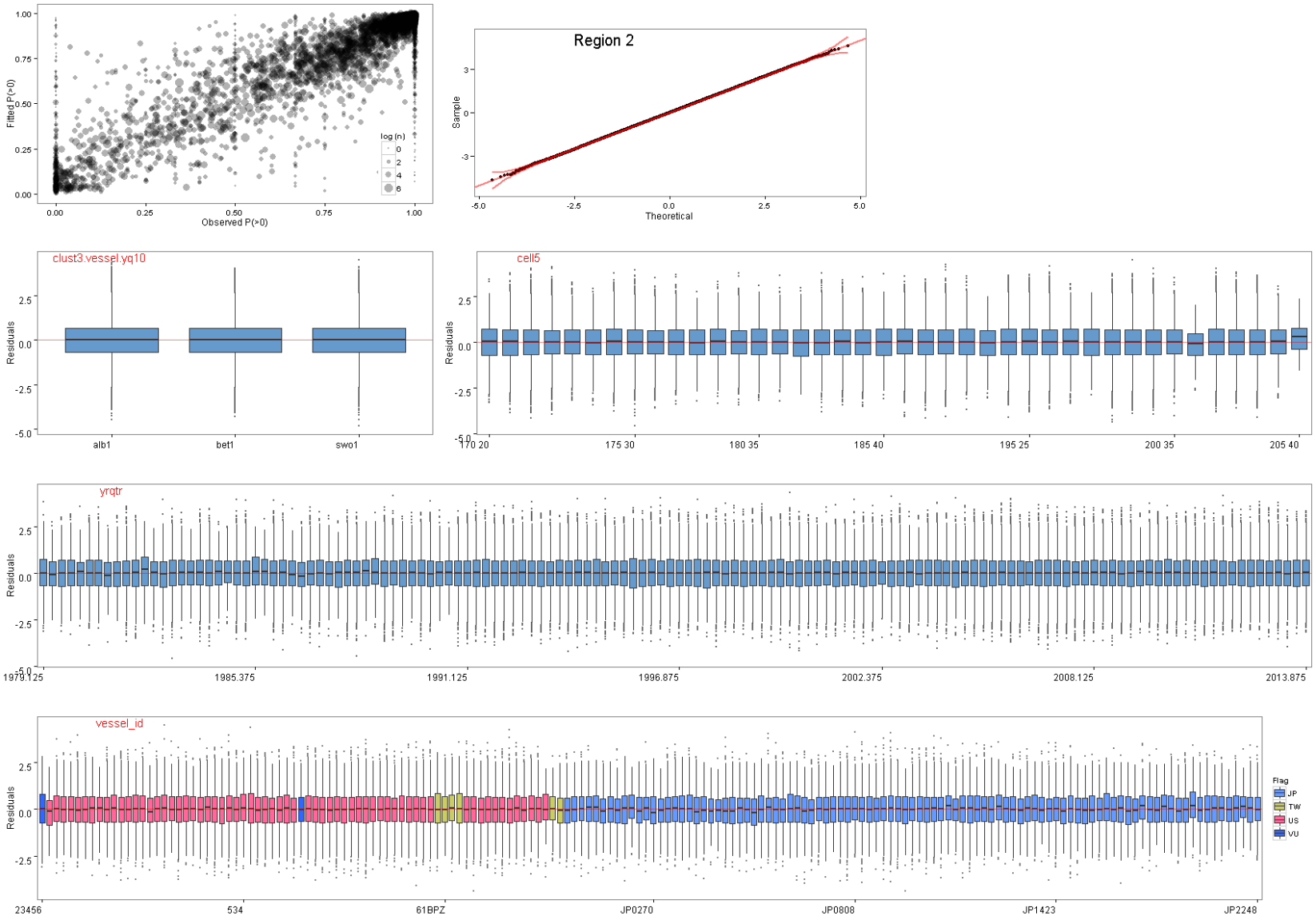


Figure 85: Diagnostic plots of fitted GLM models for region 2, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

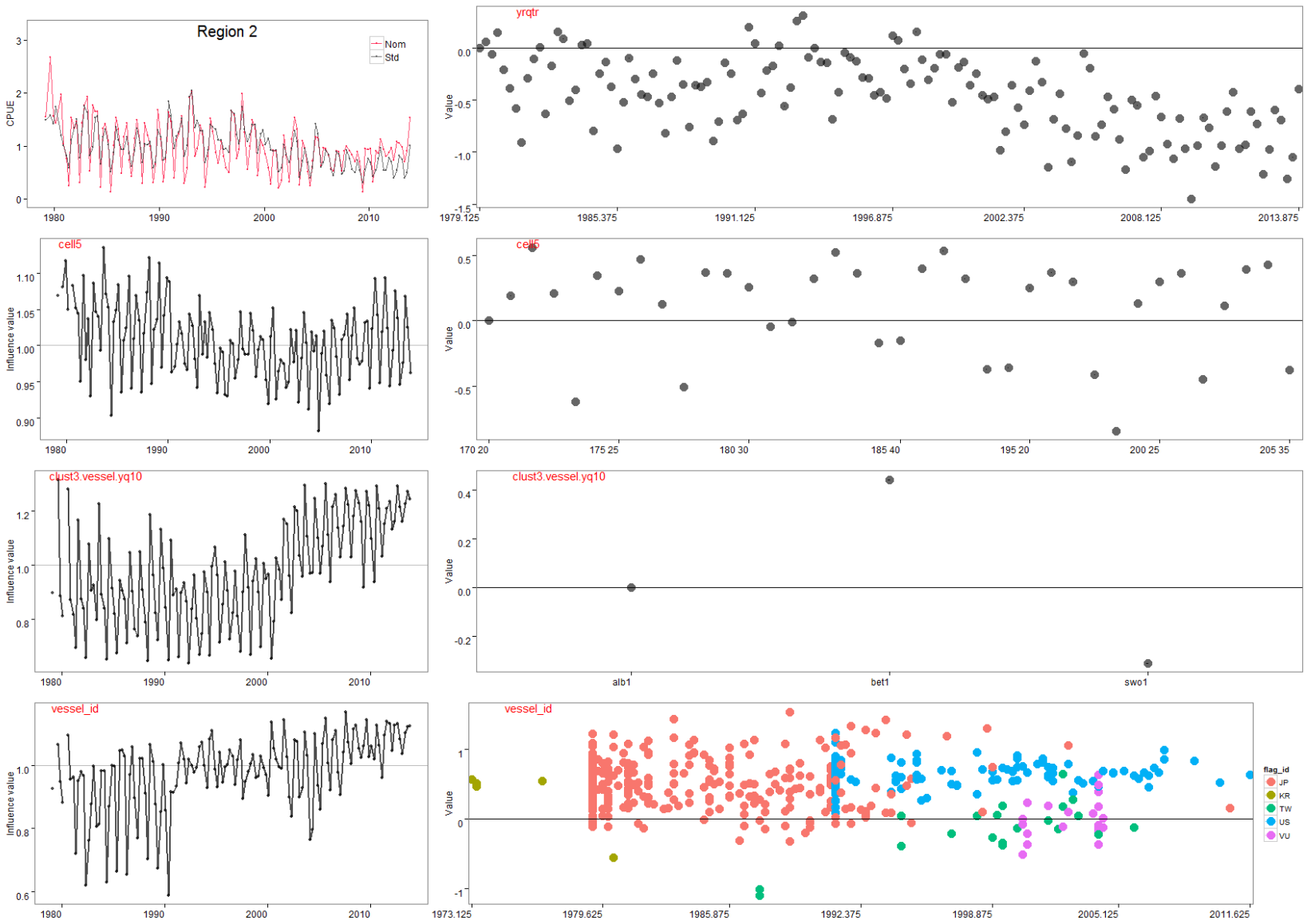


Figure 86: Summary plots of fitted GLM models for region 2, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

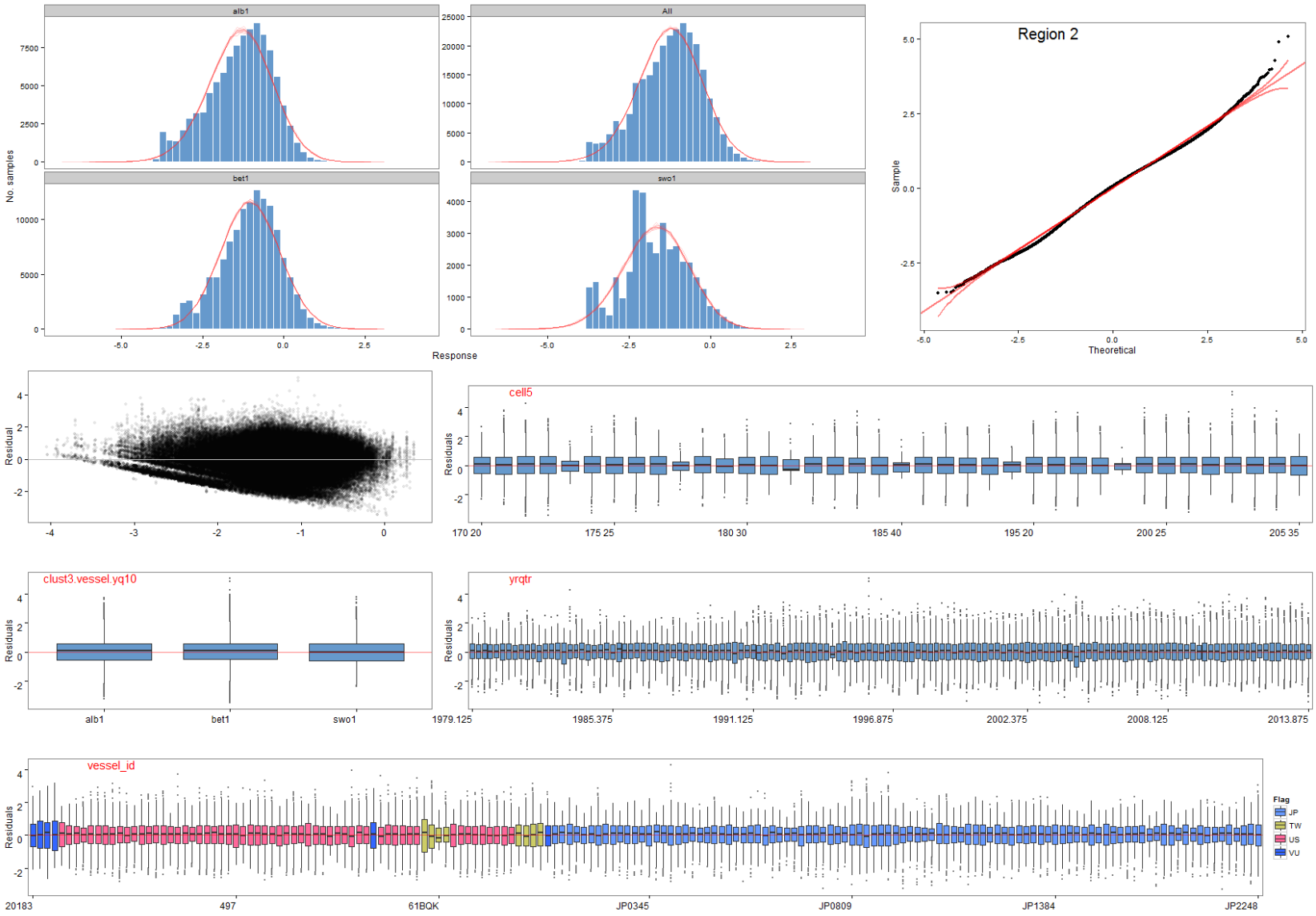


Figure 87: Diagnostic plots of fitted GLM models for region 2, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

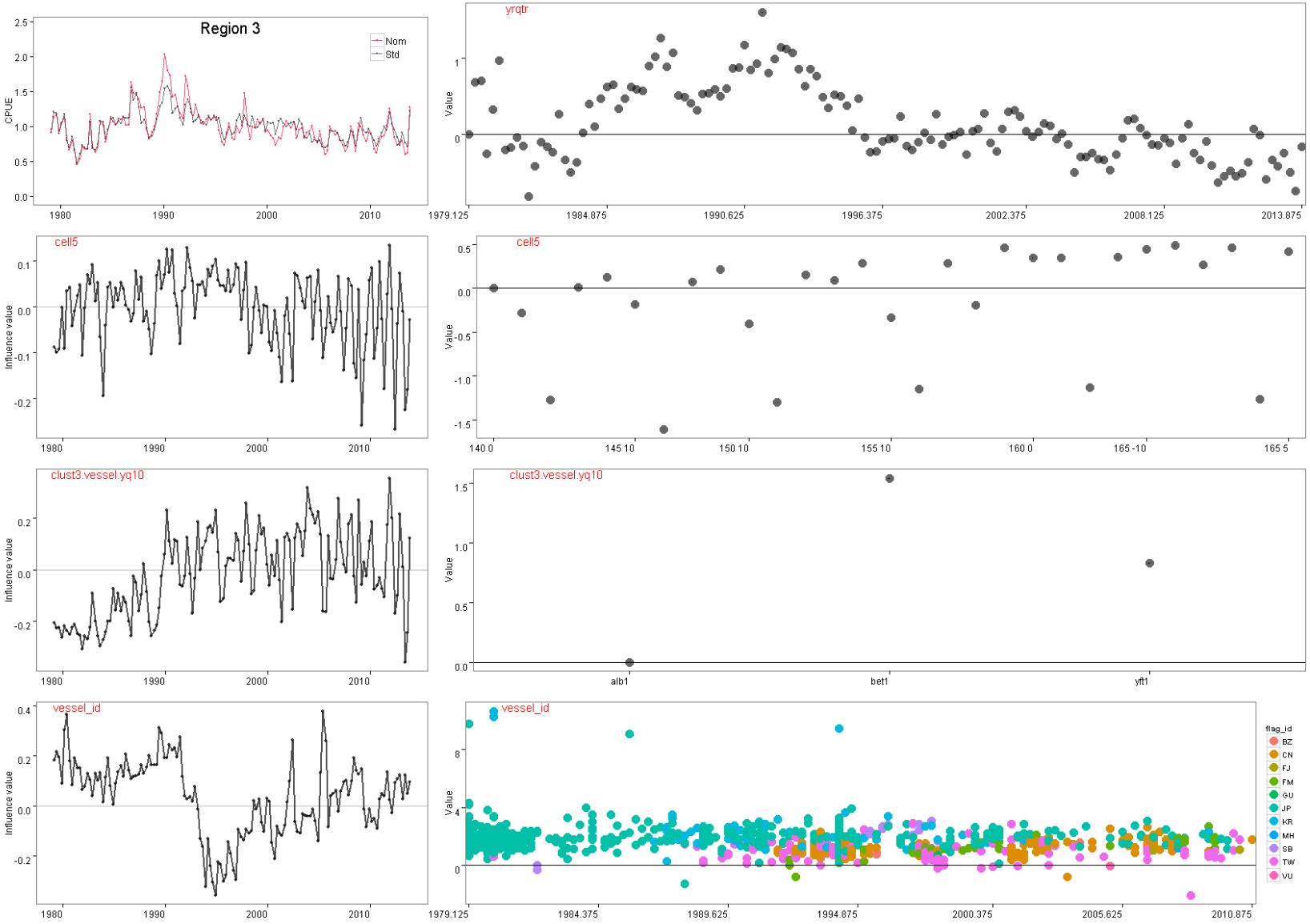


Figure 88: Summary plots of fitted GLM models for region 3, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

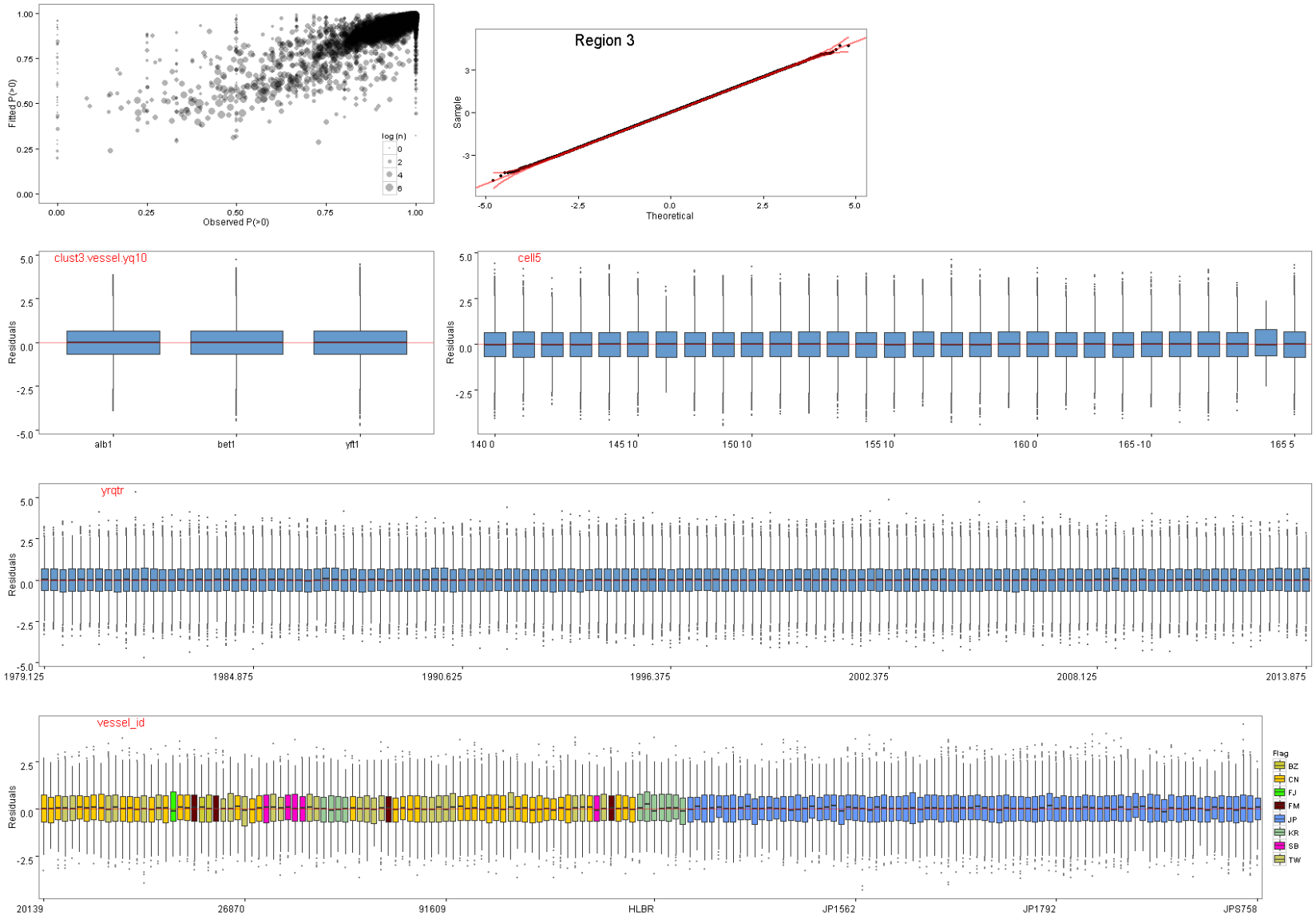


Figure 89: Diagnostic plots of fitted GLM models for region 3, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

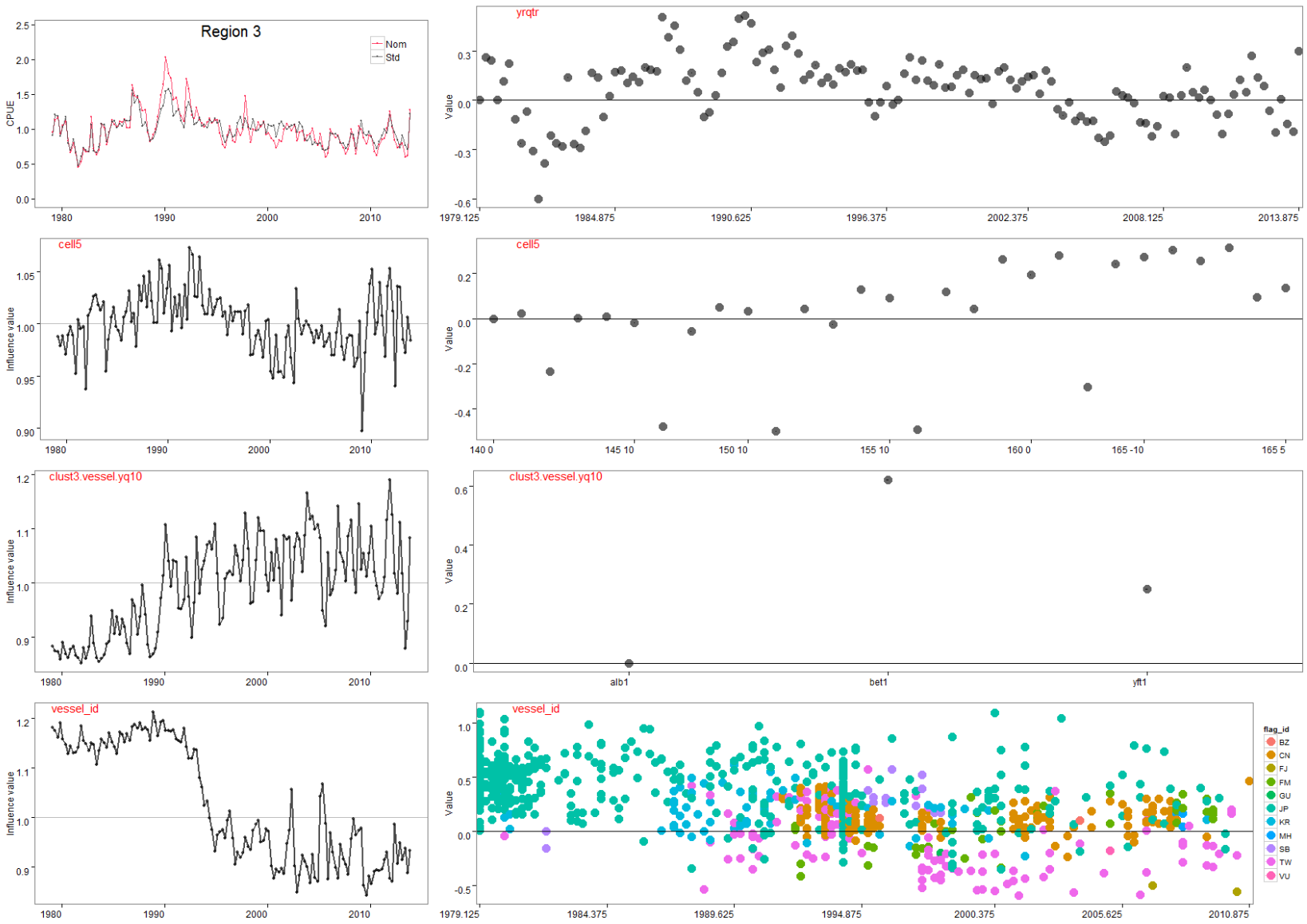


Figure 90: Summary plots of fitted GLM models for region 3, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

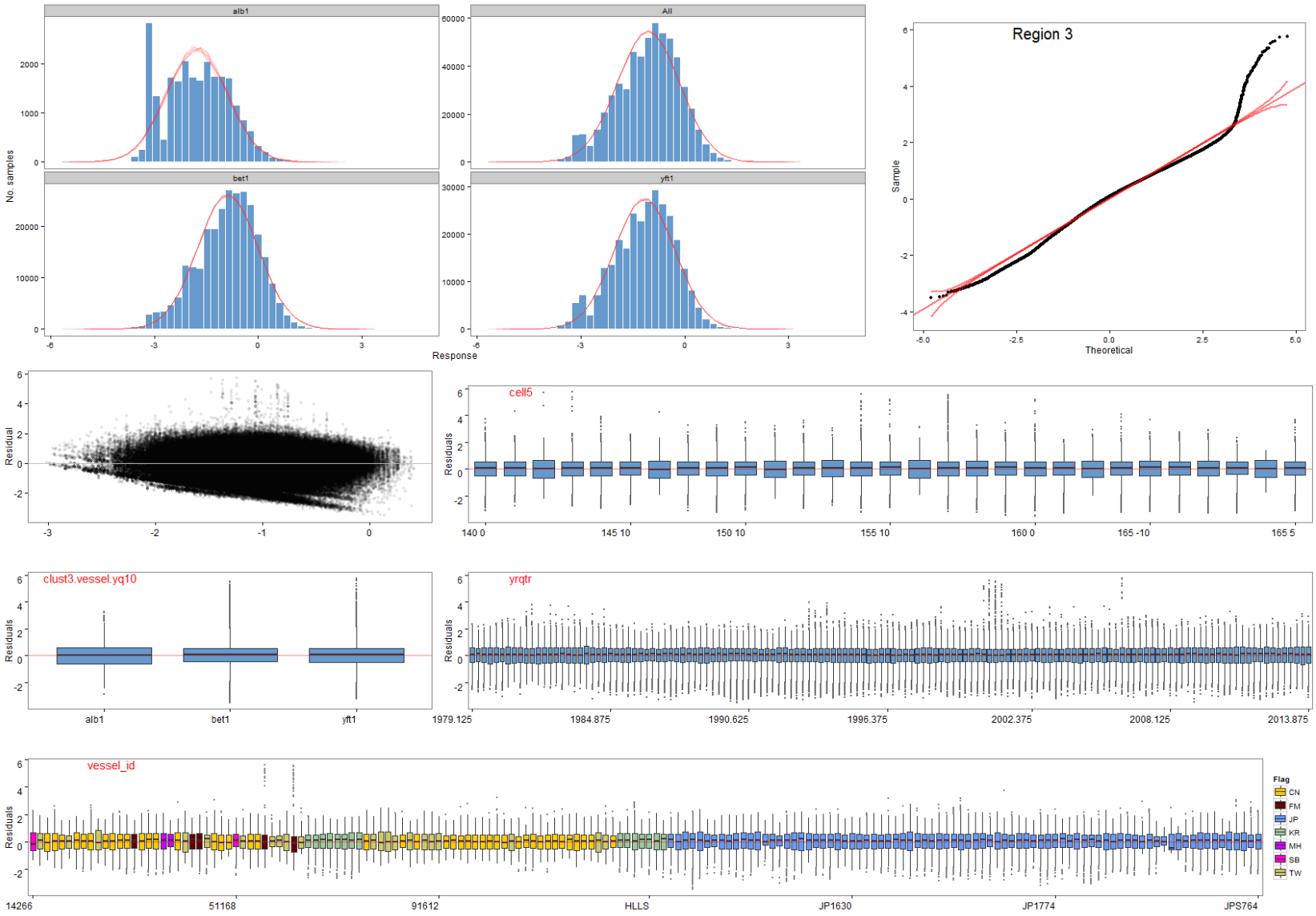


Figure 91: Diagnostic plots of fitted GLM models for region 3, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

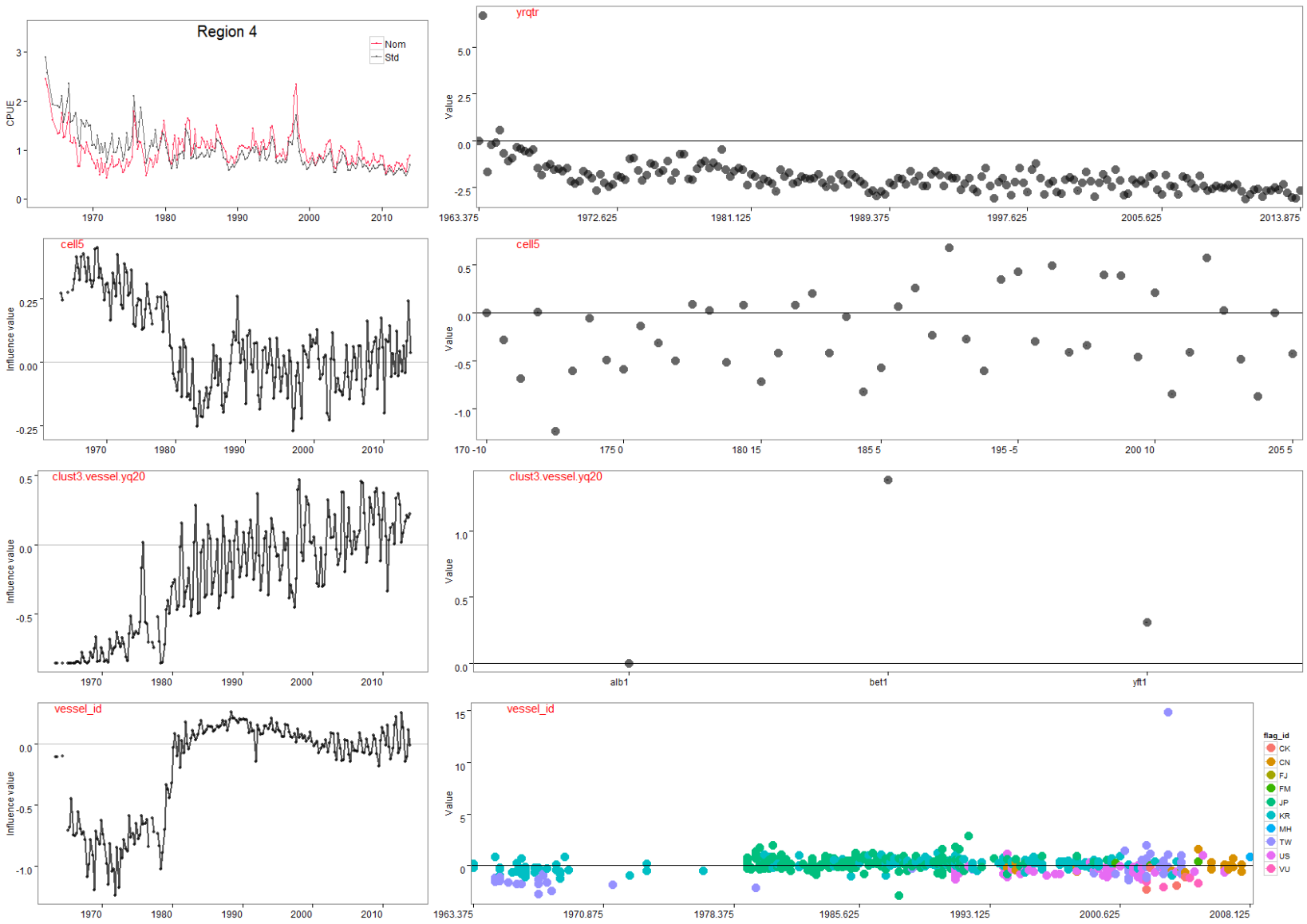


Figure 92: Summary plots of fitted GLM models for region 4, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

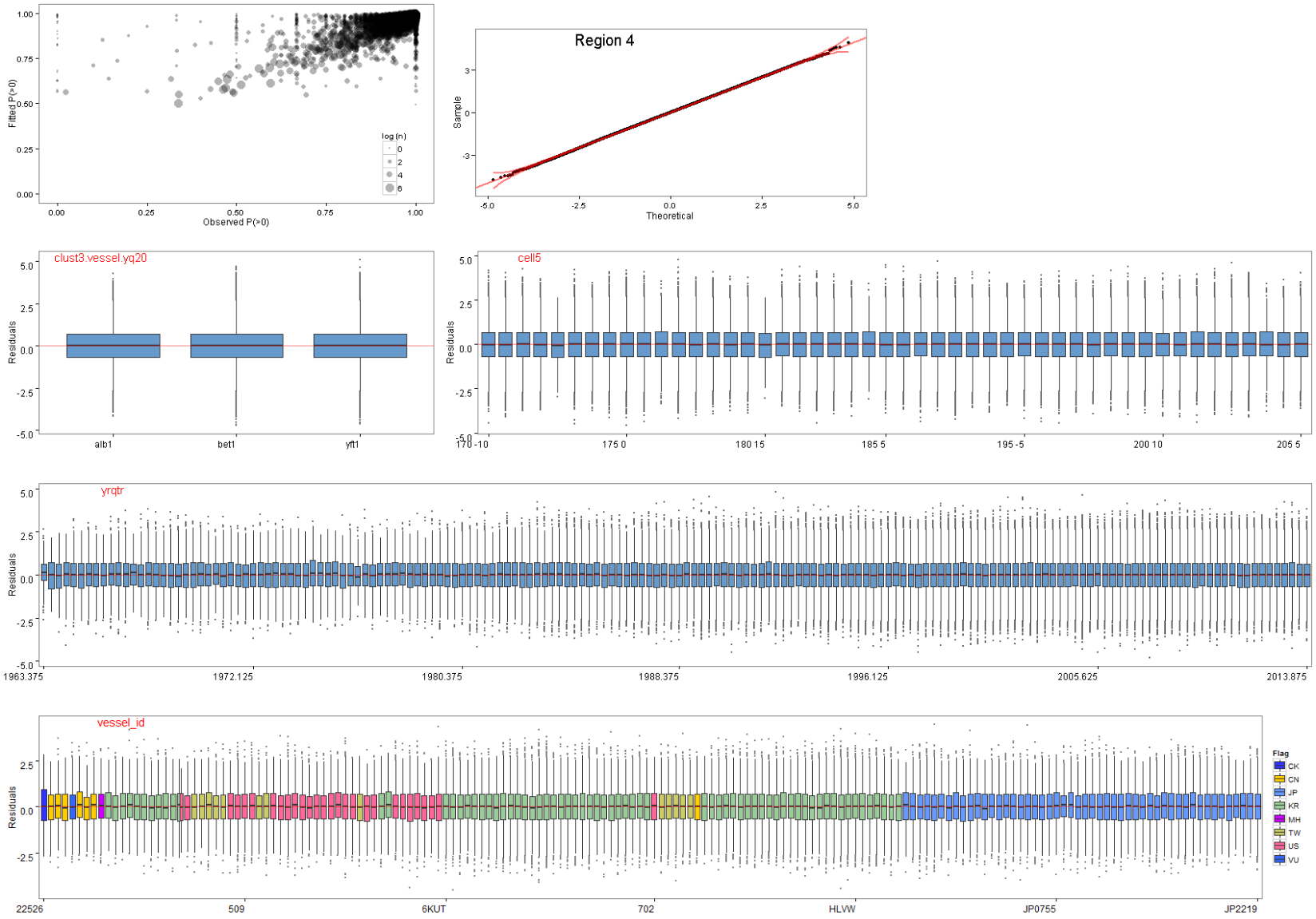


Figure 93: Diagnostic plots of fitted GLM models for region 4, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

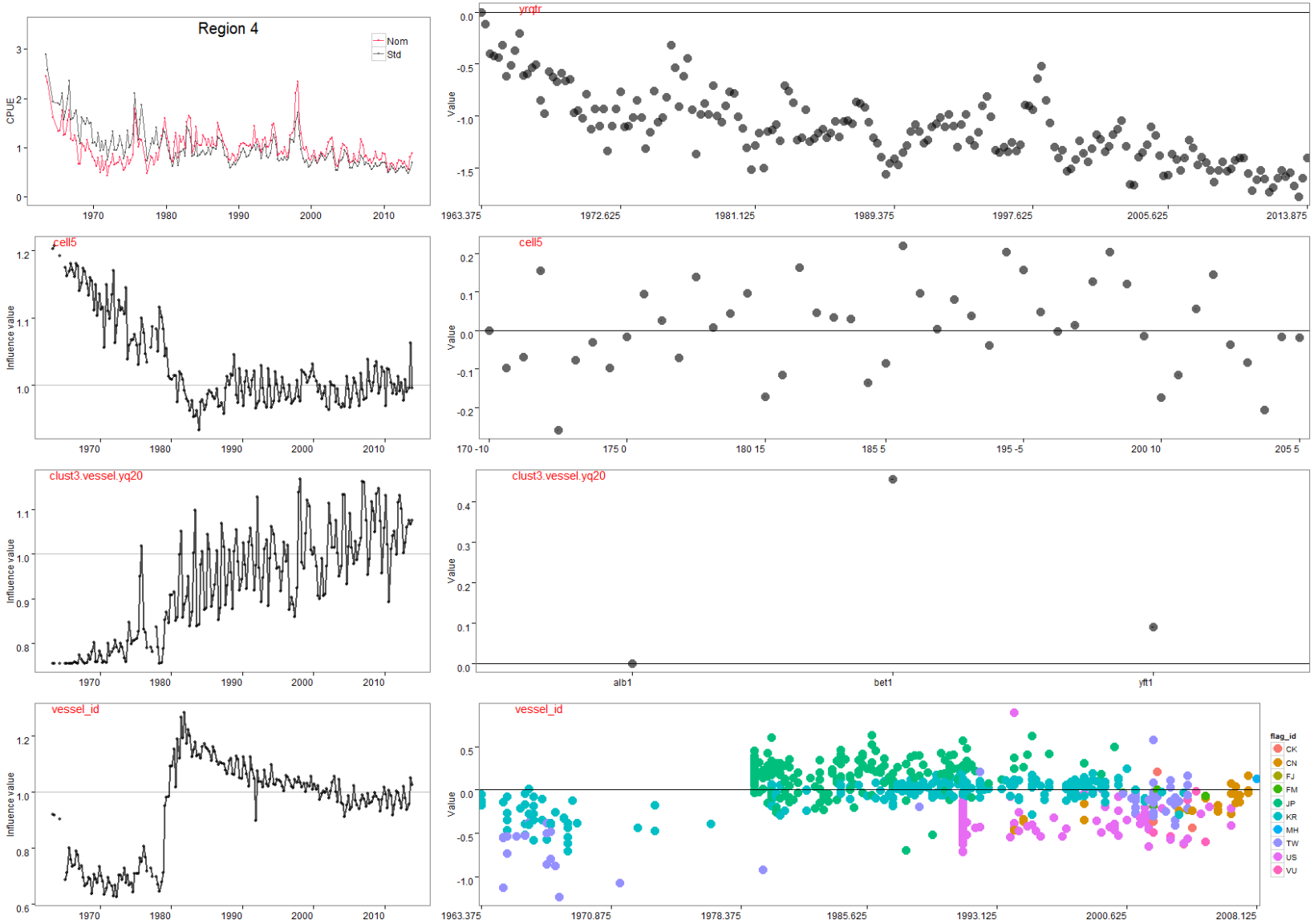


Figure 94: Summary plots of fitted GLM models for region 4, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

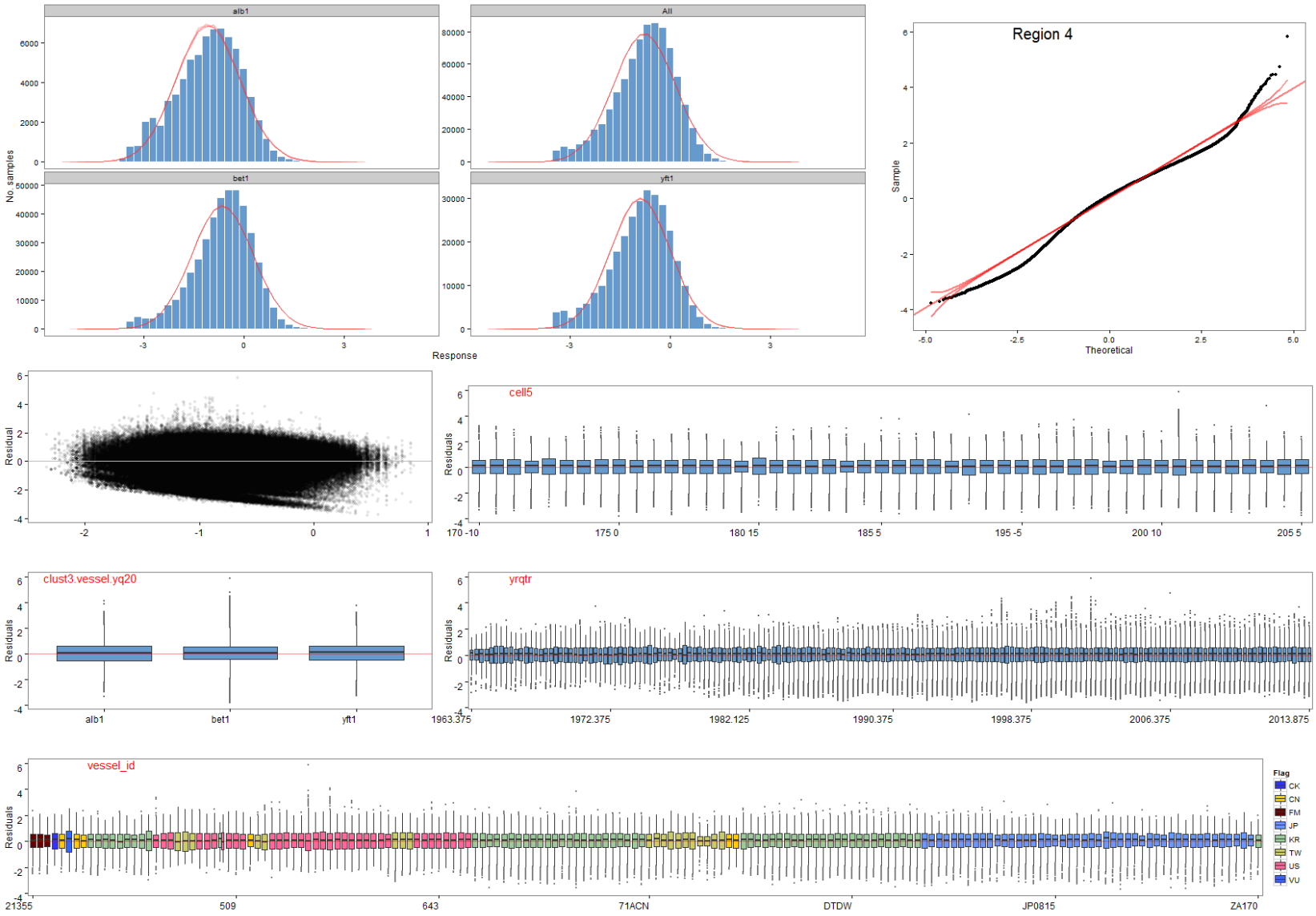


Figure 95: Diagnostic plots of fitted GLM models for region 4, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

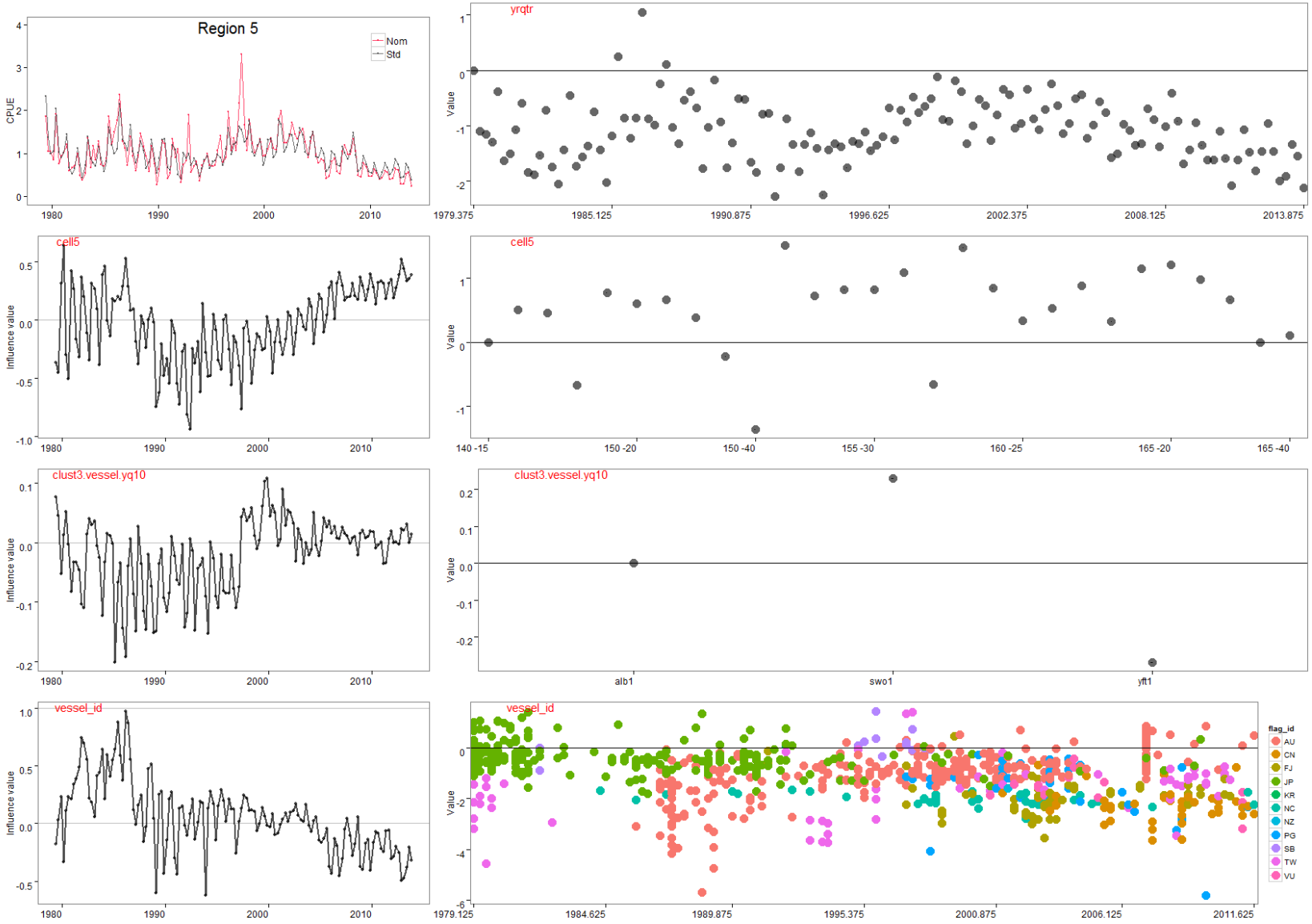


Figure 96: Summary plots of fitted GLM models for region 5, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

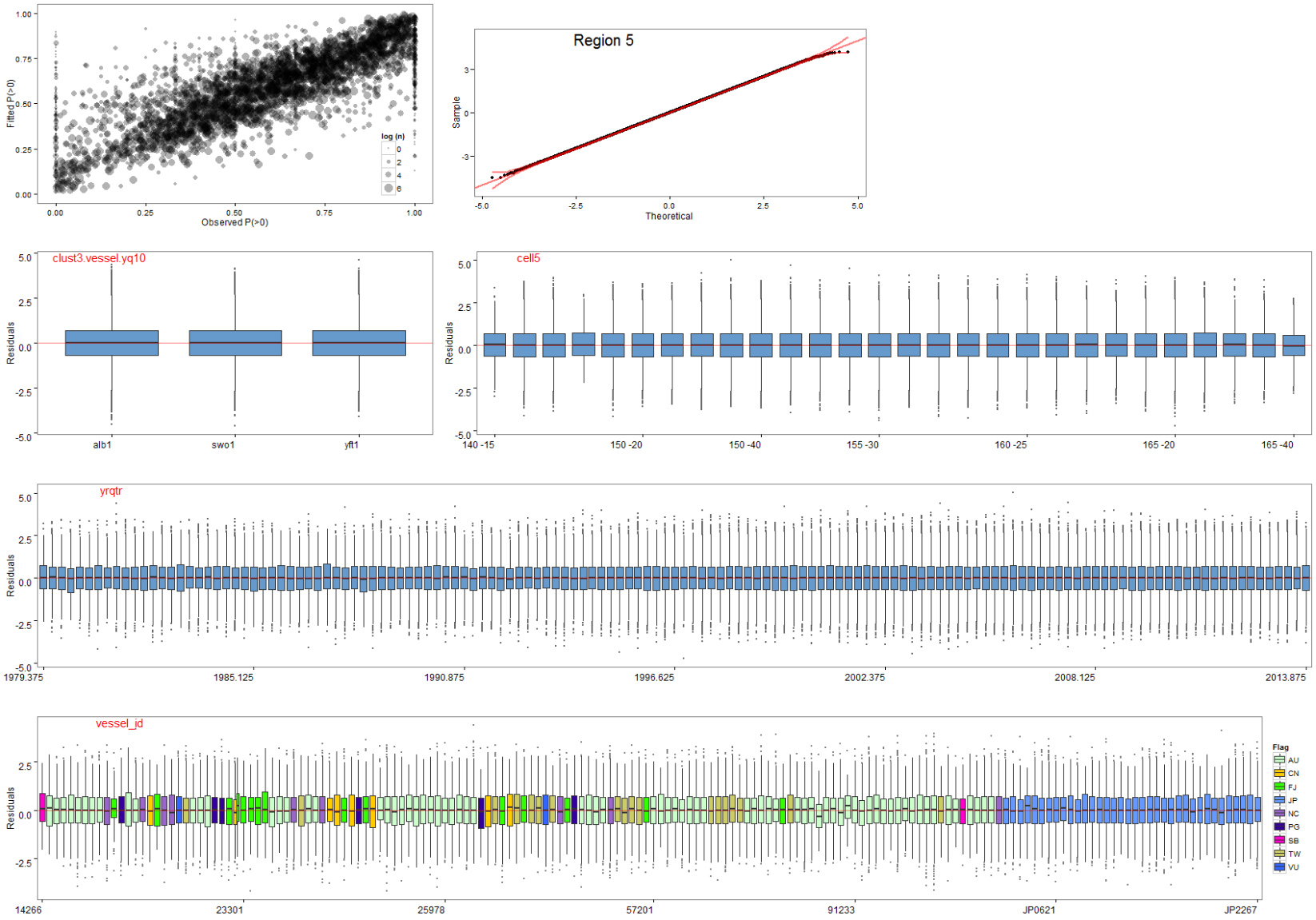


Figure 97: Diagnostic plots of fitted GLM models for region 5, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

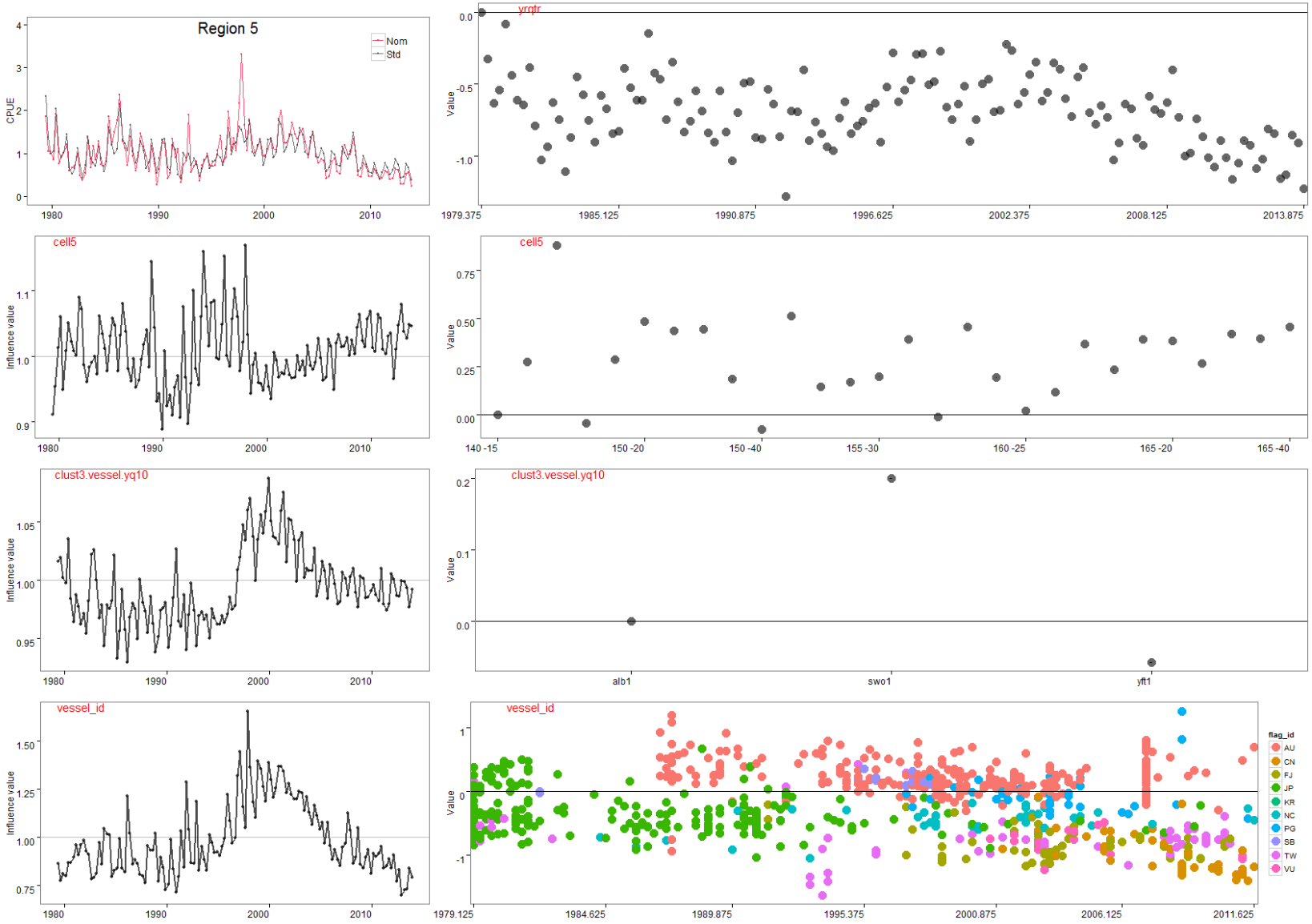


Figure 98: Summary plots of fitted GLM models for region 5, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

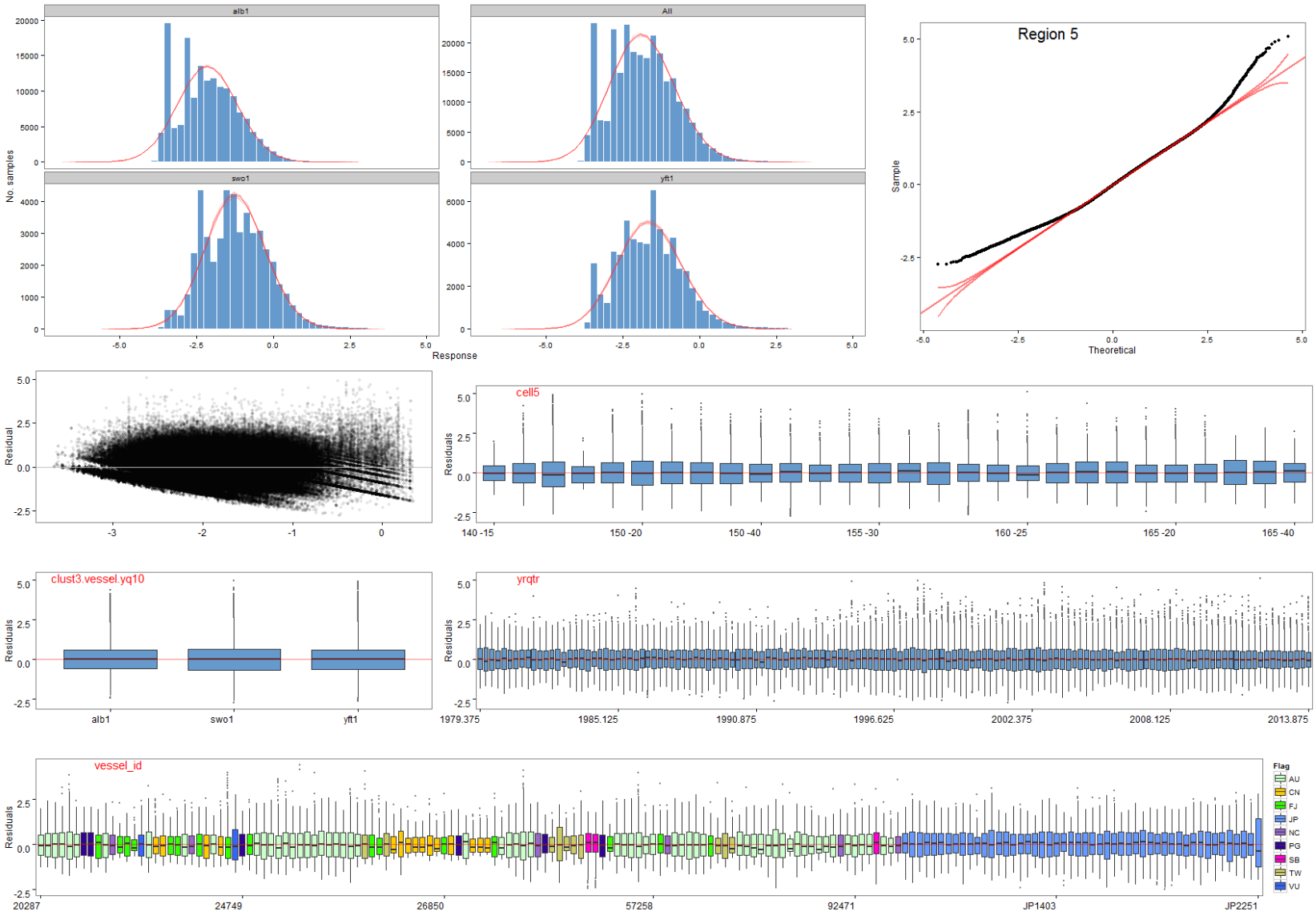


Figure 99: Diagnostic plots of fitted GLM models for region 5, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

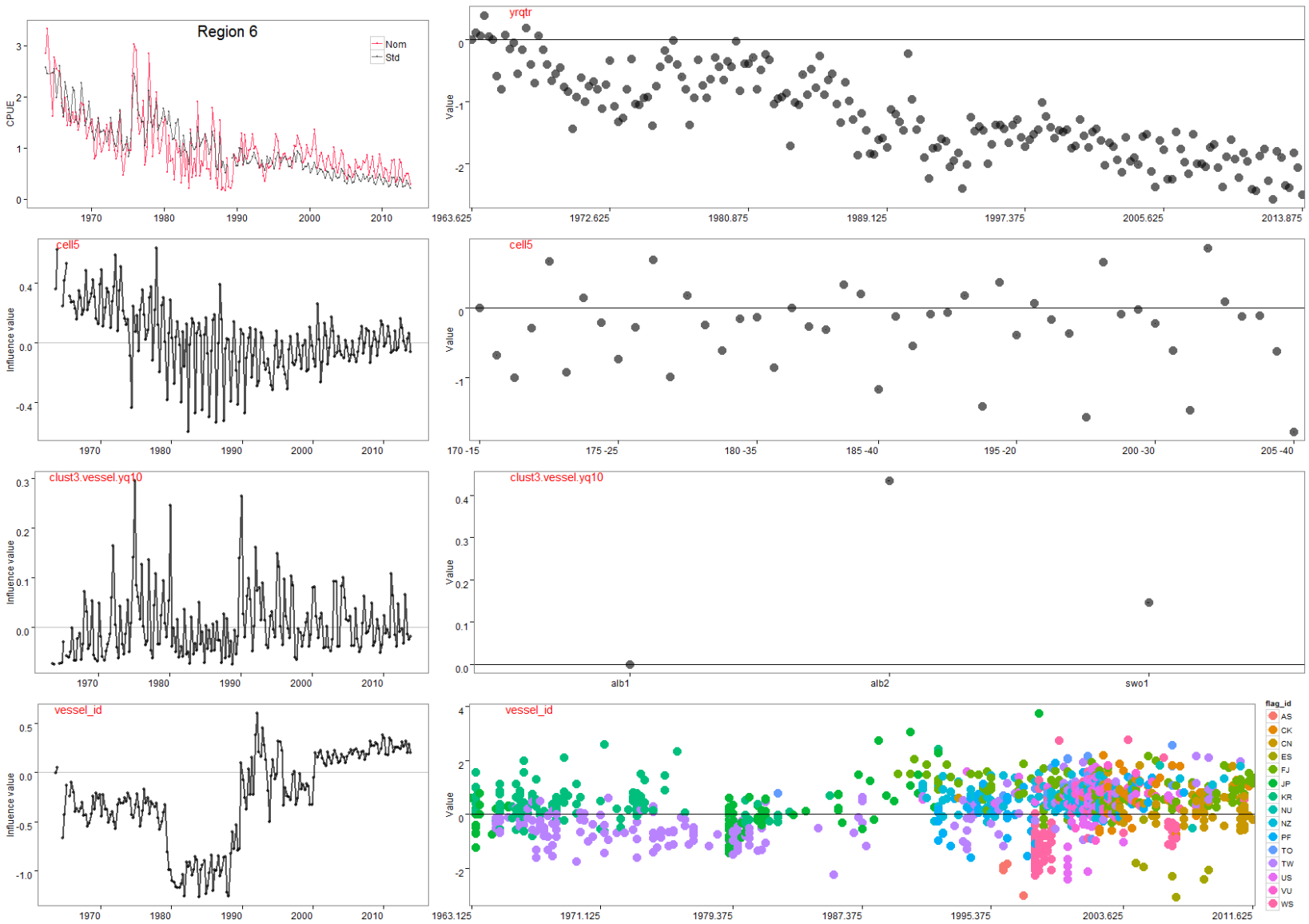


Figure 100: Summary plots of fitted GLM models for region 6, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

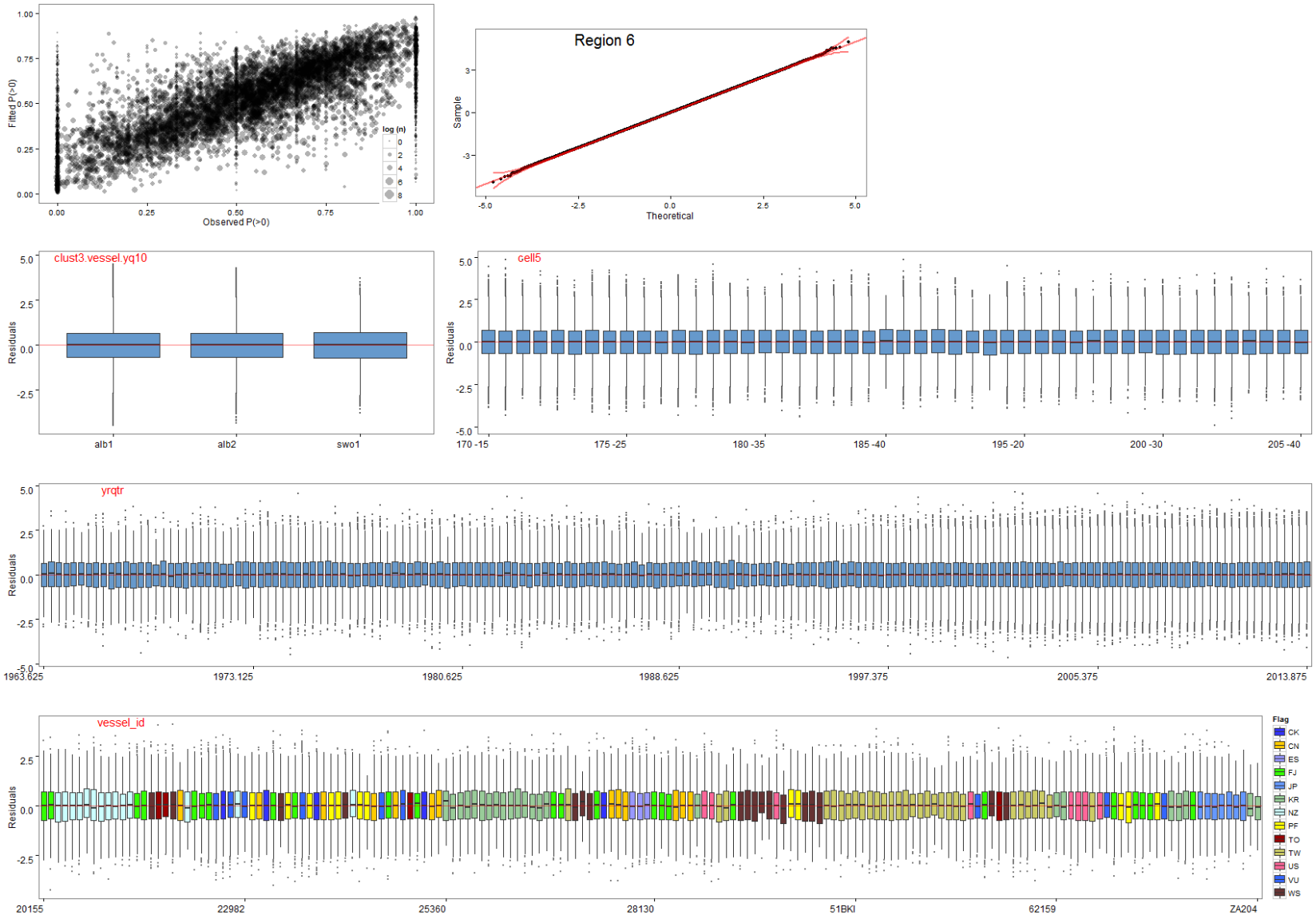


Figure 101: Diagnostic plots of fitted GLM models for region 6, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

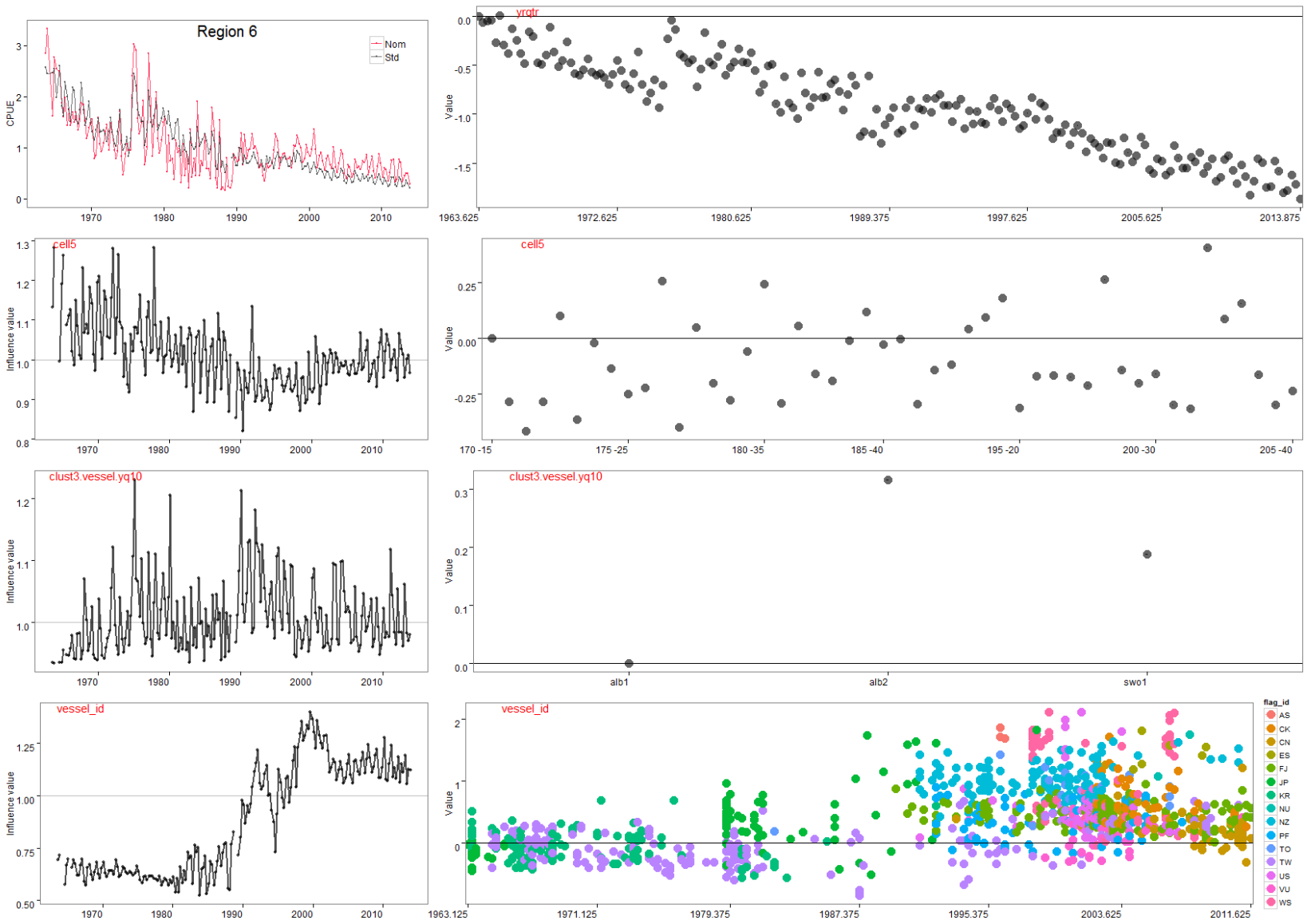


Figure 102: Summary plots of fitted GLM models for region 6, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

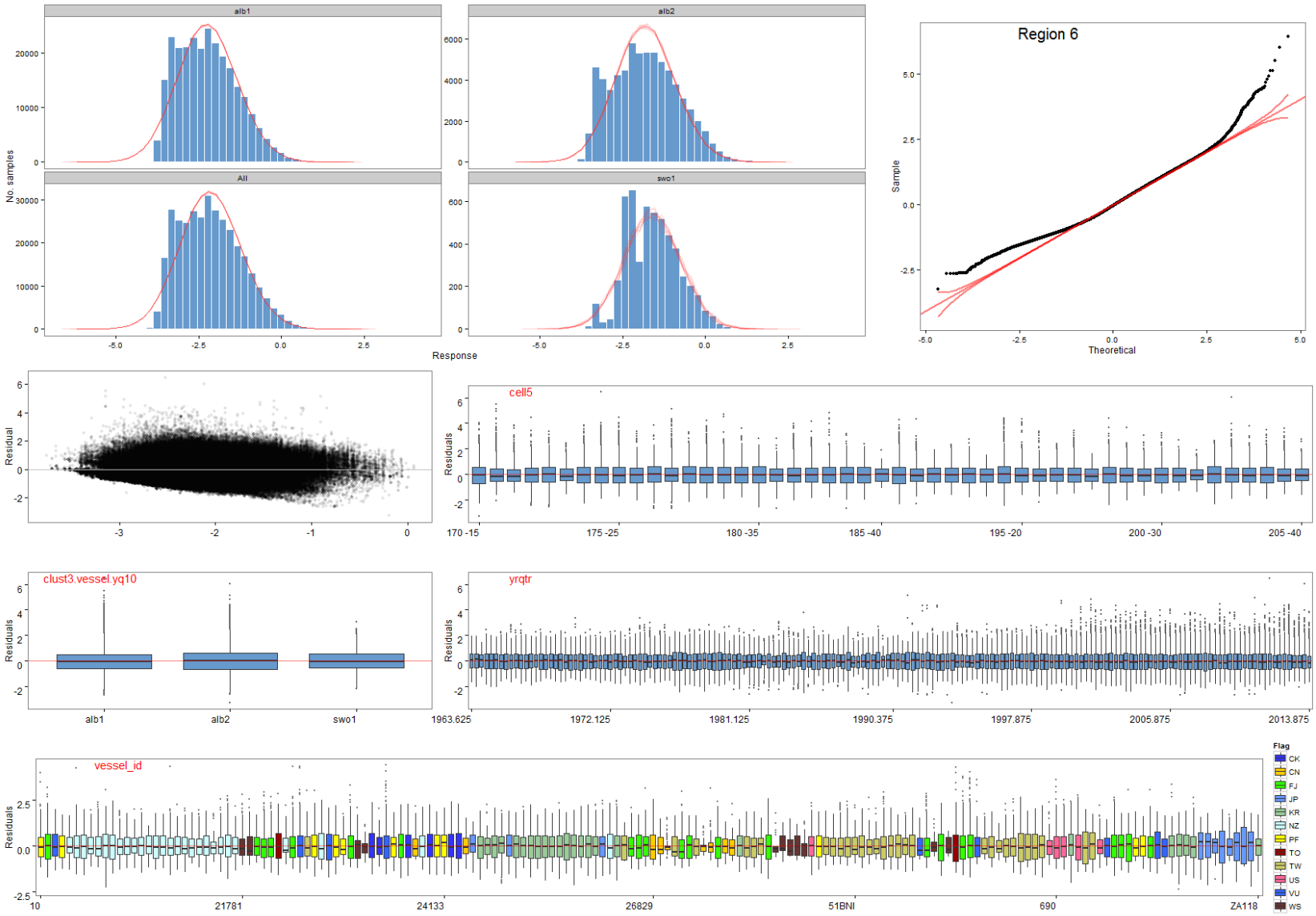


Figure 103: Diagnostic plots of fitted GLM models for region 6, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

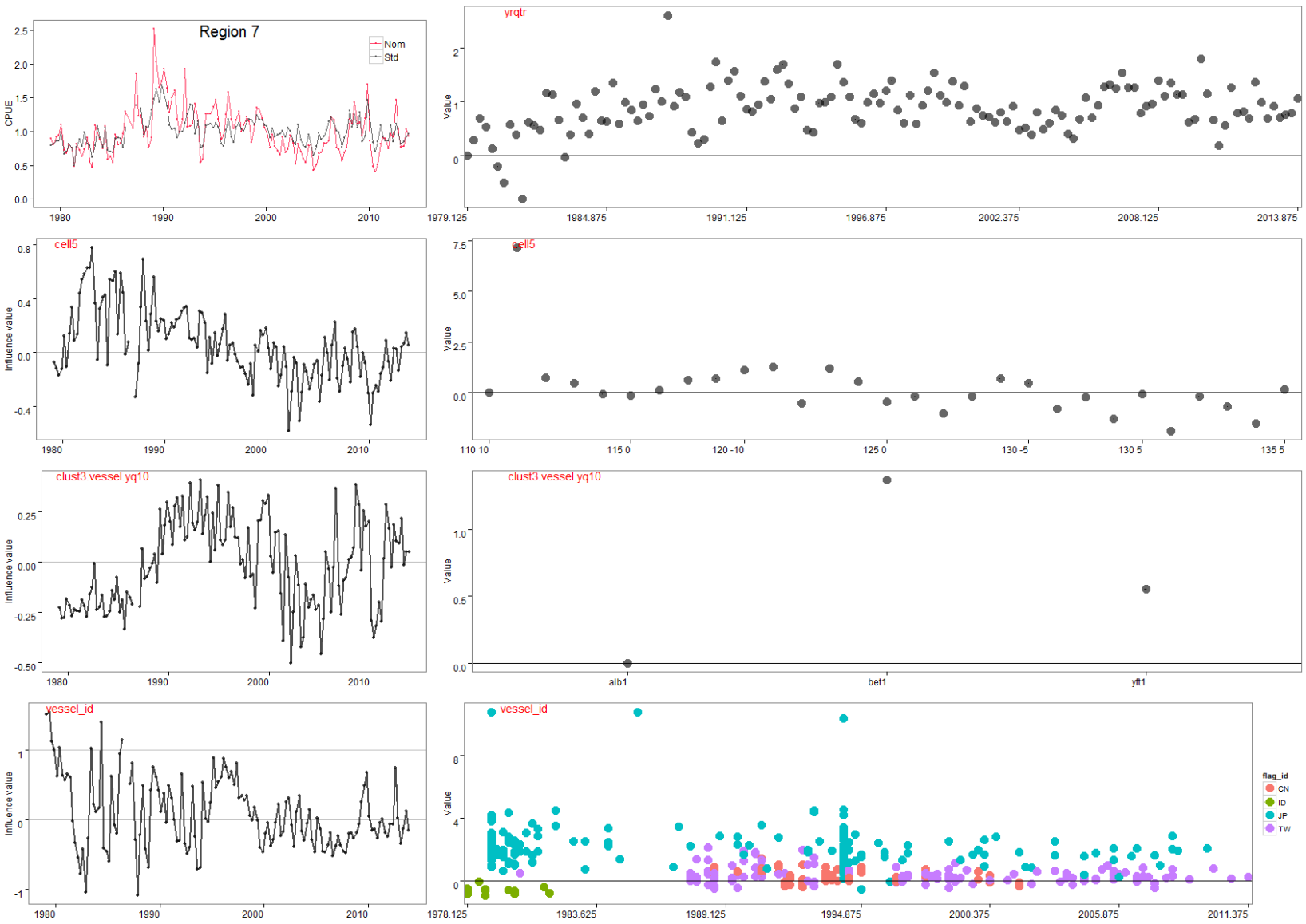


Figure 104: Summary plots of fitted GLM models for region 7, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

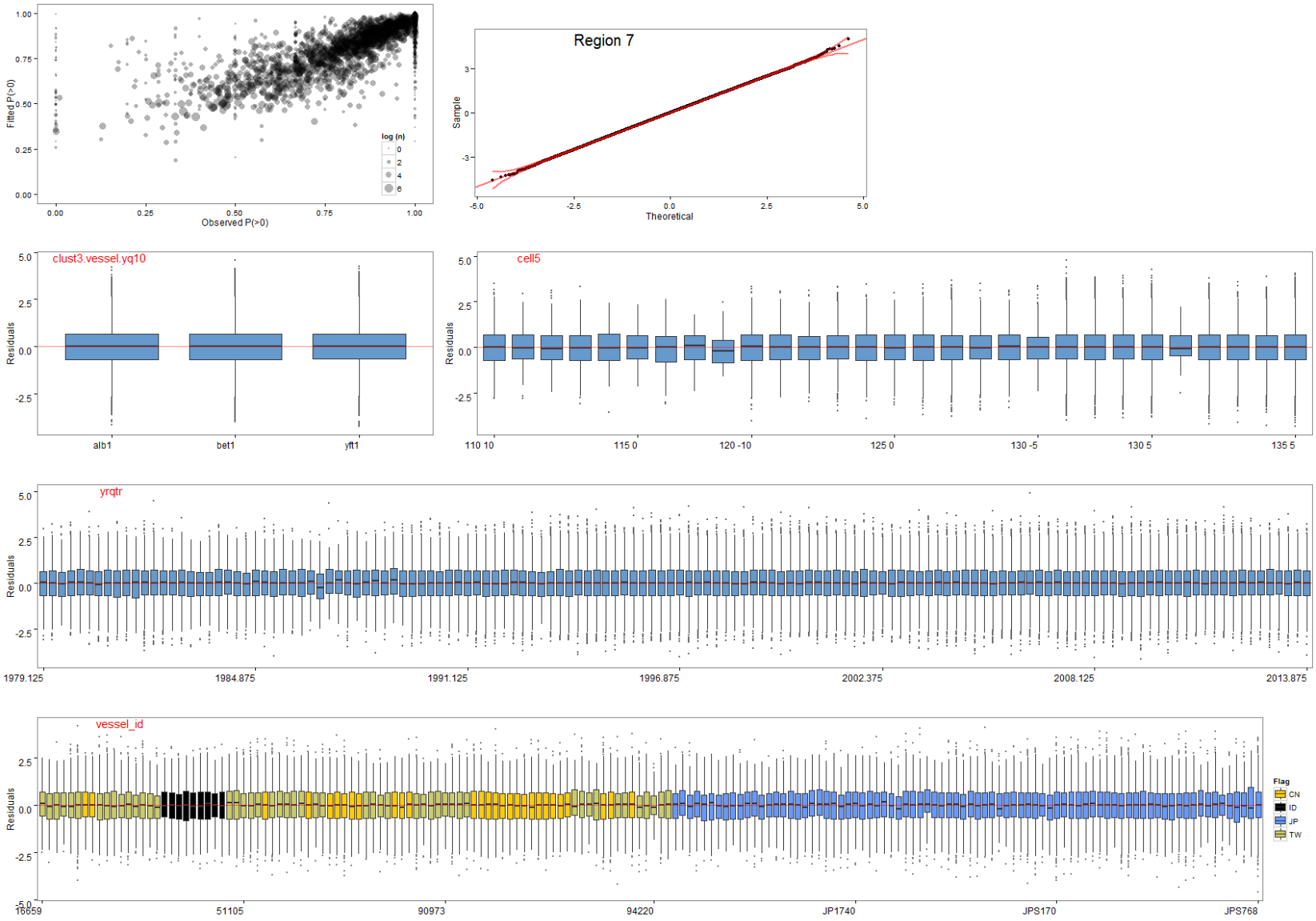


Figure 105: Diagnostic plots of fitted GLM models for region 7, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

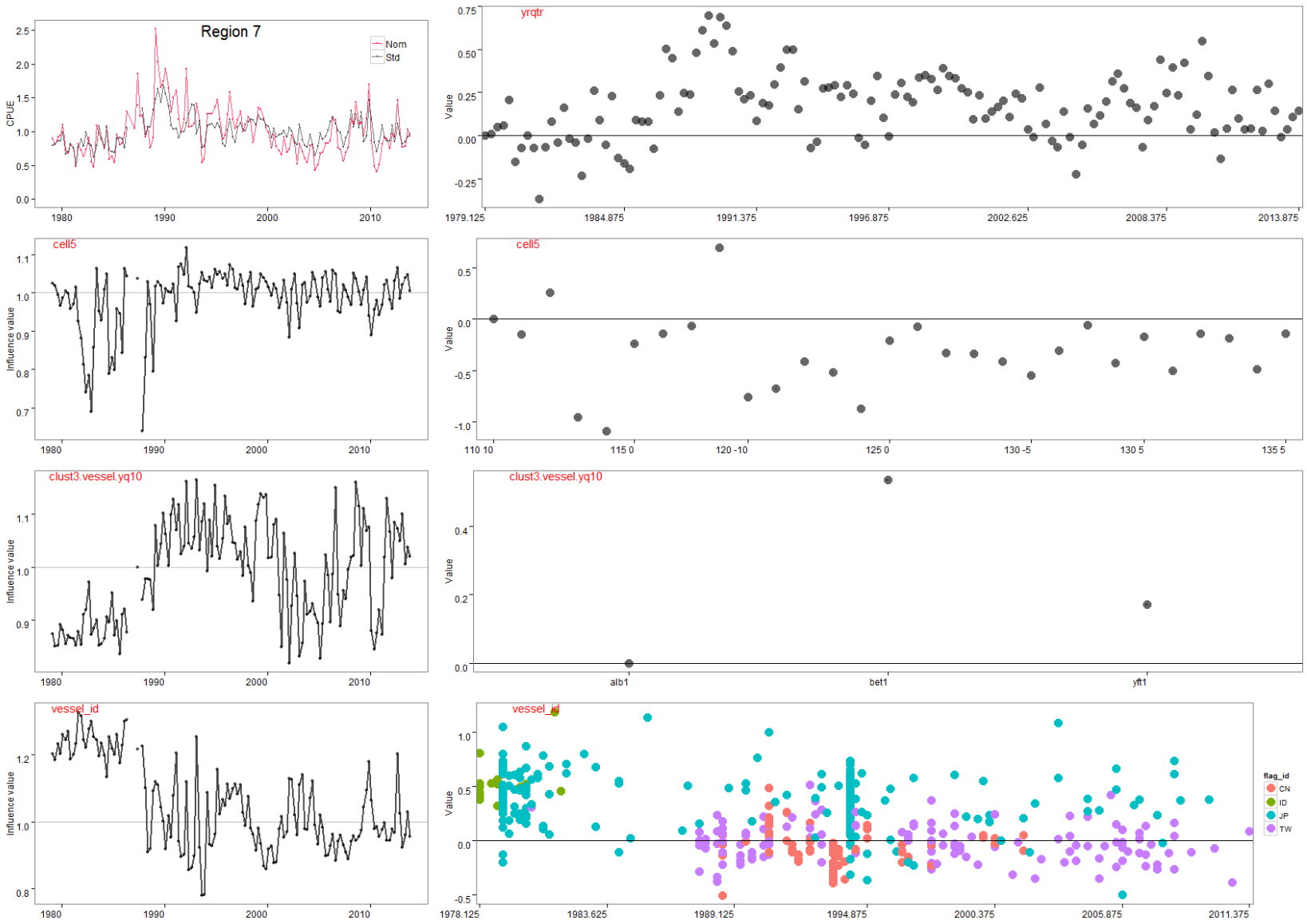


Figure 106: Summary plots of fitted GLM models for region 7, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

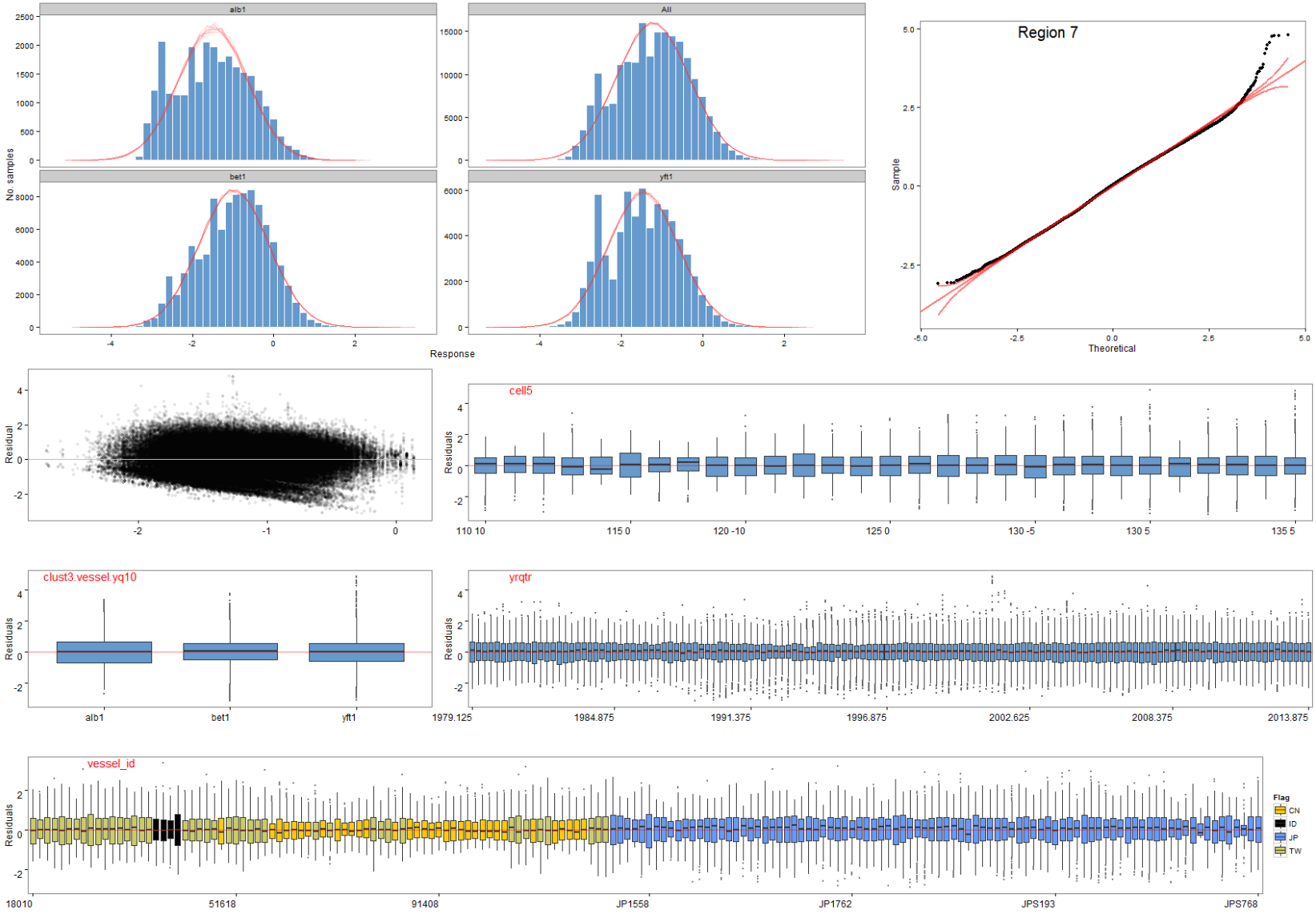


Figure 107: Diagnostic plots of fitted GLM models for region 7, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

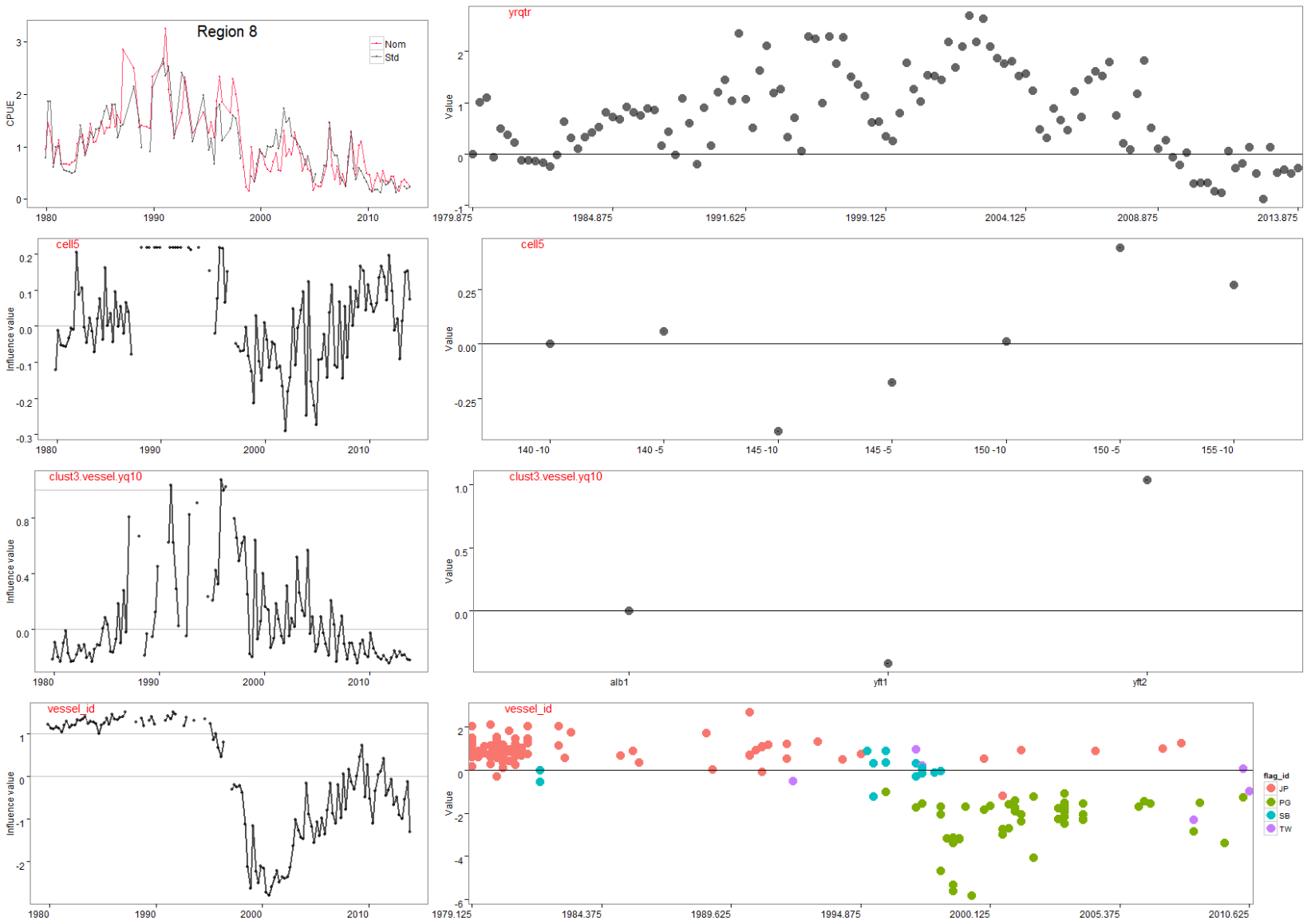


Figure 108: Summary plots of fitted GLM models for region 8, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

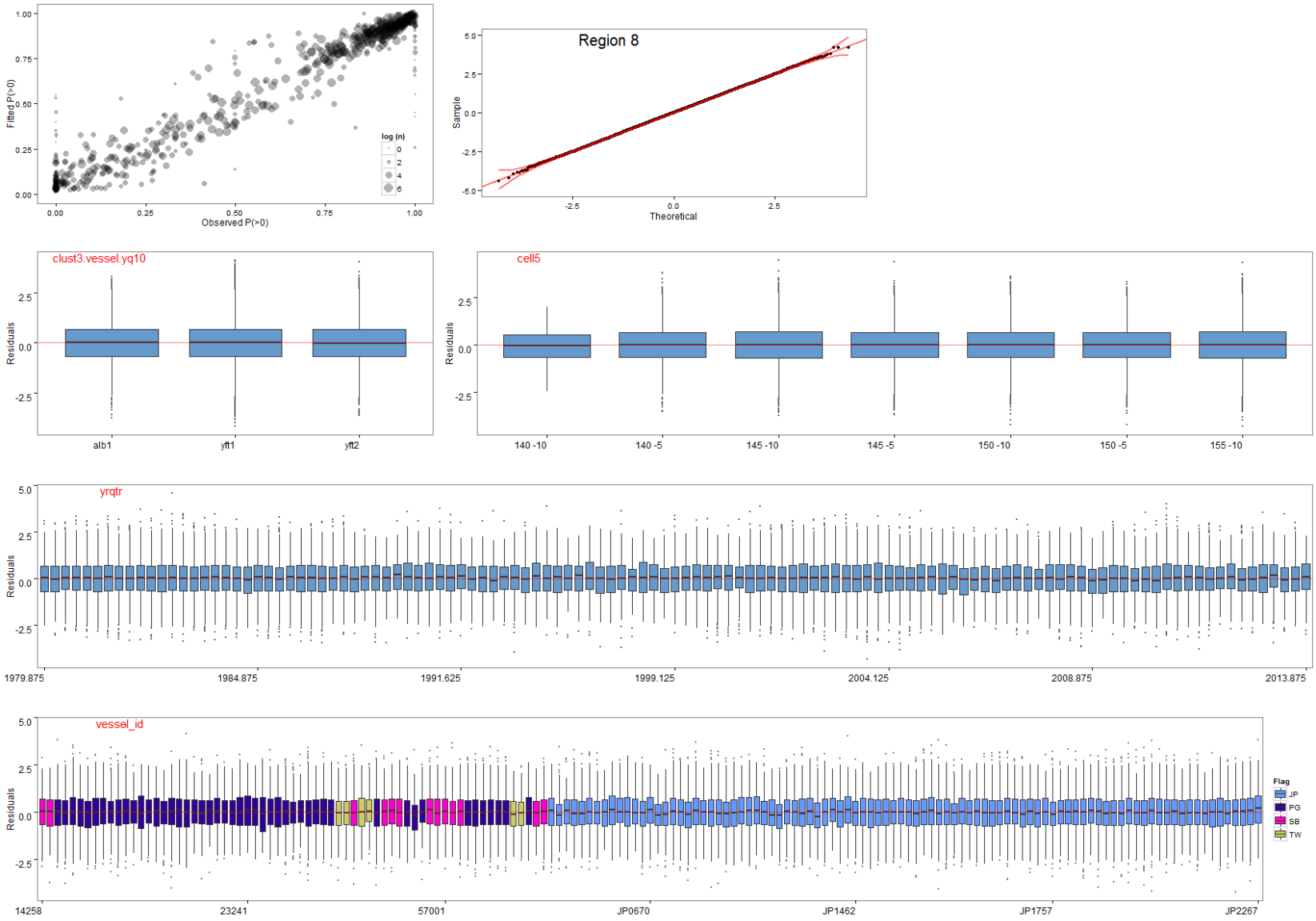


Figure 109: Diagnostic plots of fitted GLM models for region 8, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

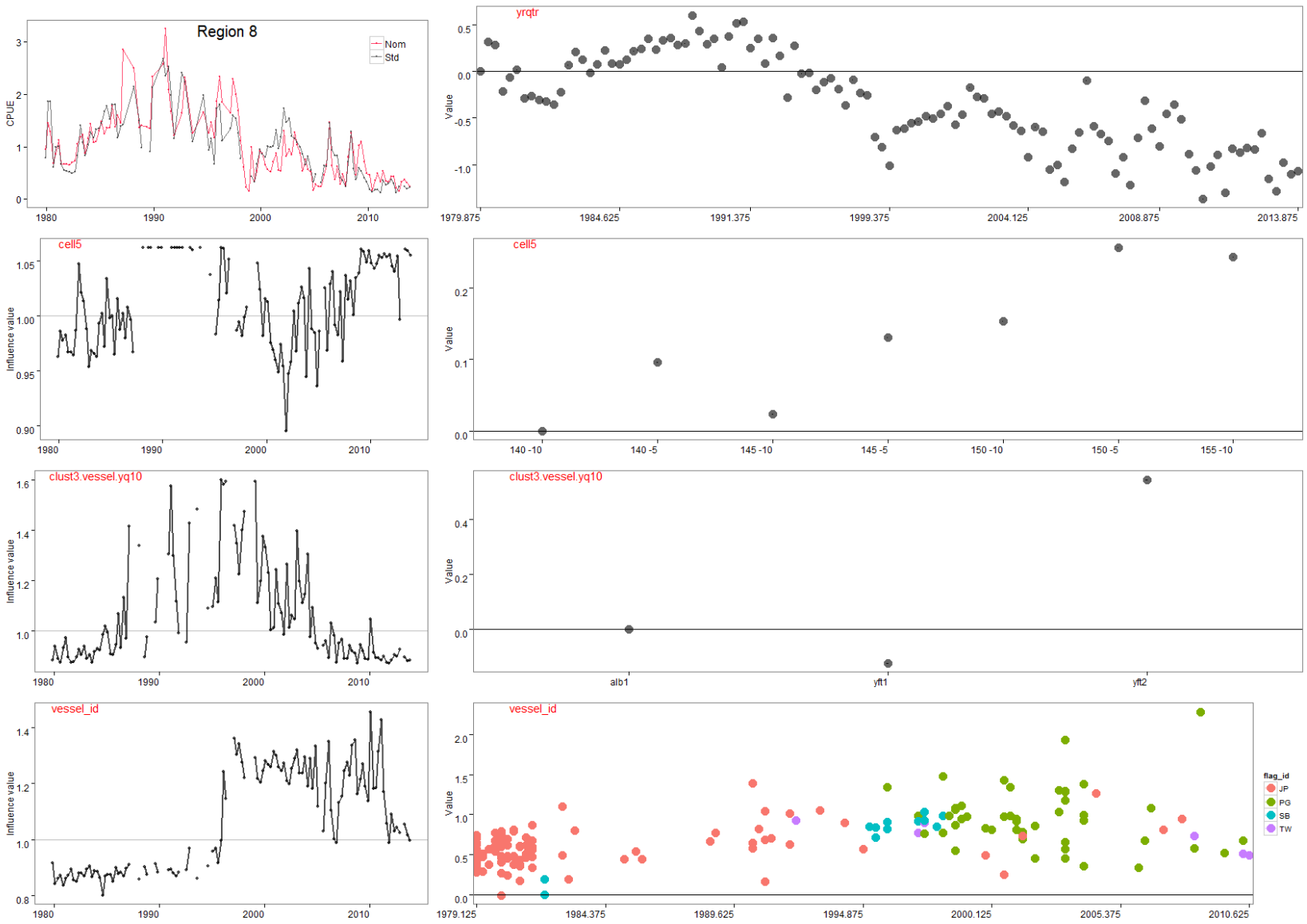


Figure 110: Summary plots of fitted GLM models for region 8, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

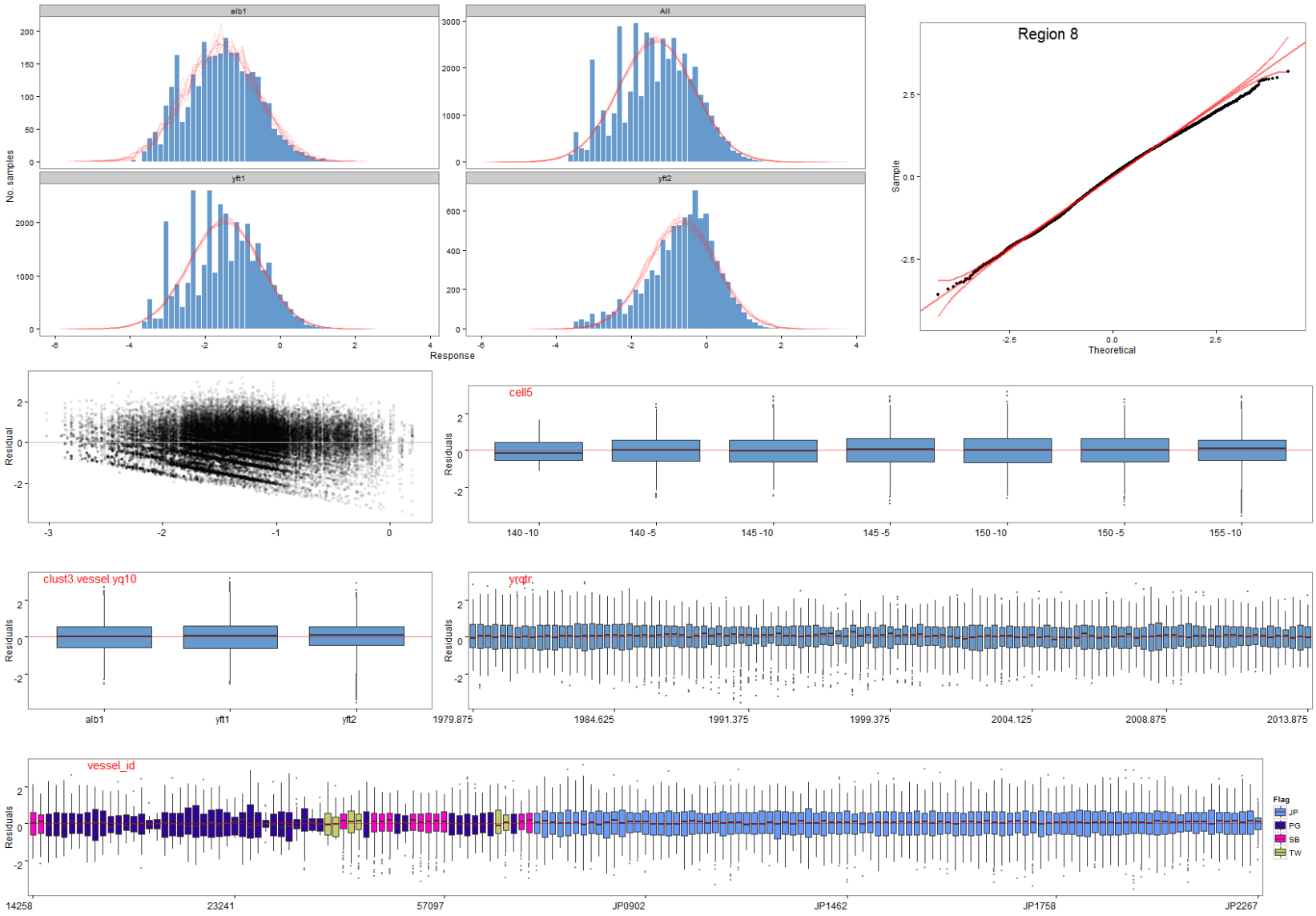


Figure 111: Diagnostic plots of fitted GLM models for region 8, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

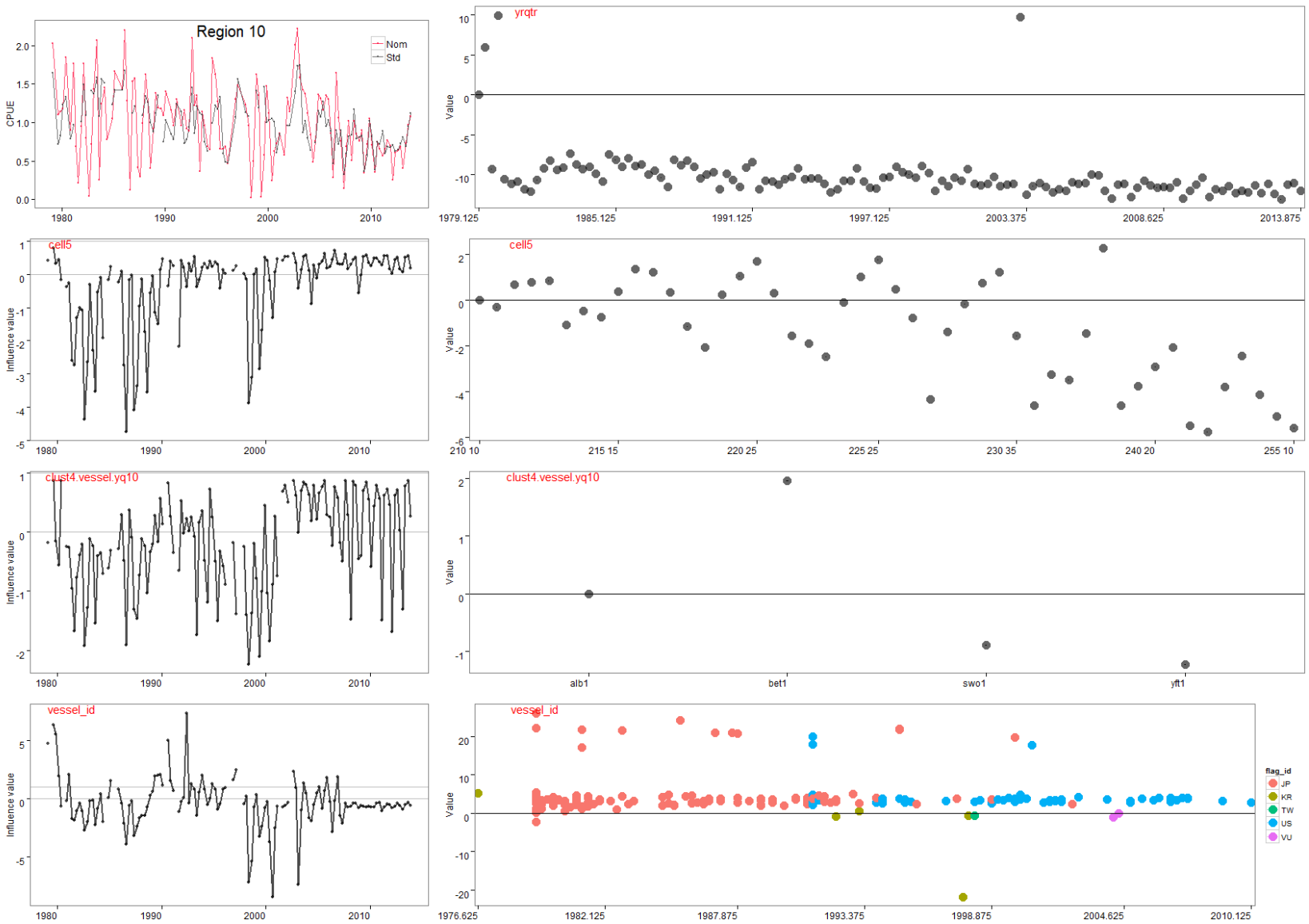


Figure 112: Summary plots of fitted GLM models for region 10, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

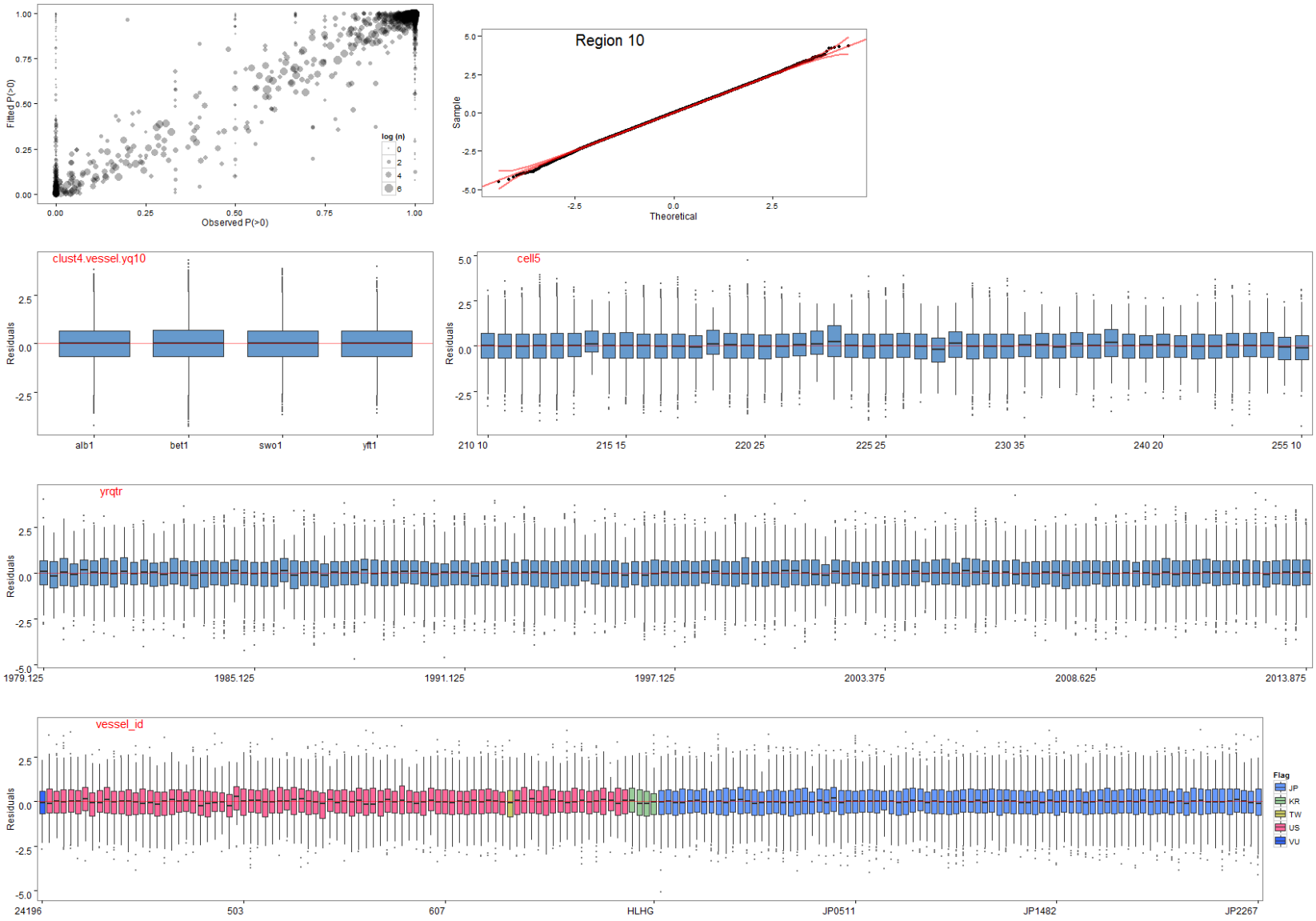


Figure 113: Diagnostic plots of fitted GLM models for region 10, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

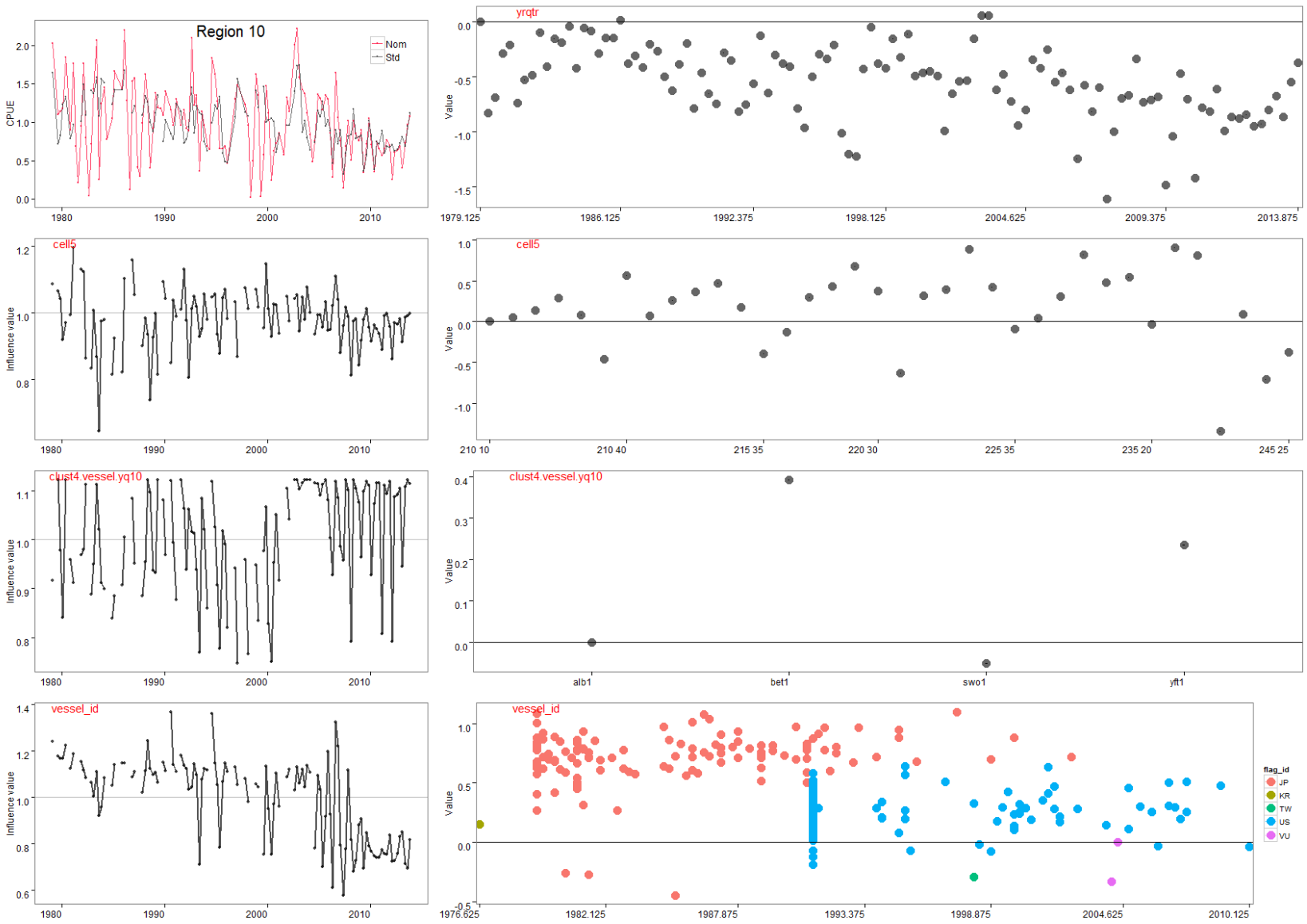


Figure 114: Summary plots of fitted GLM models for region 10, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

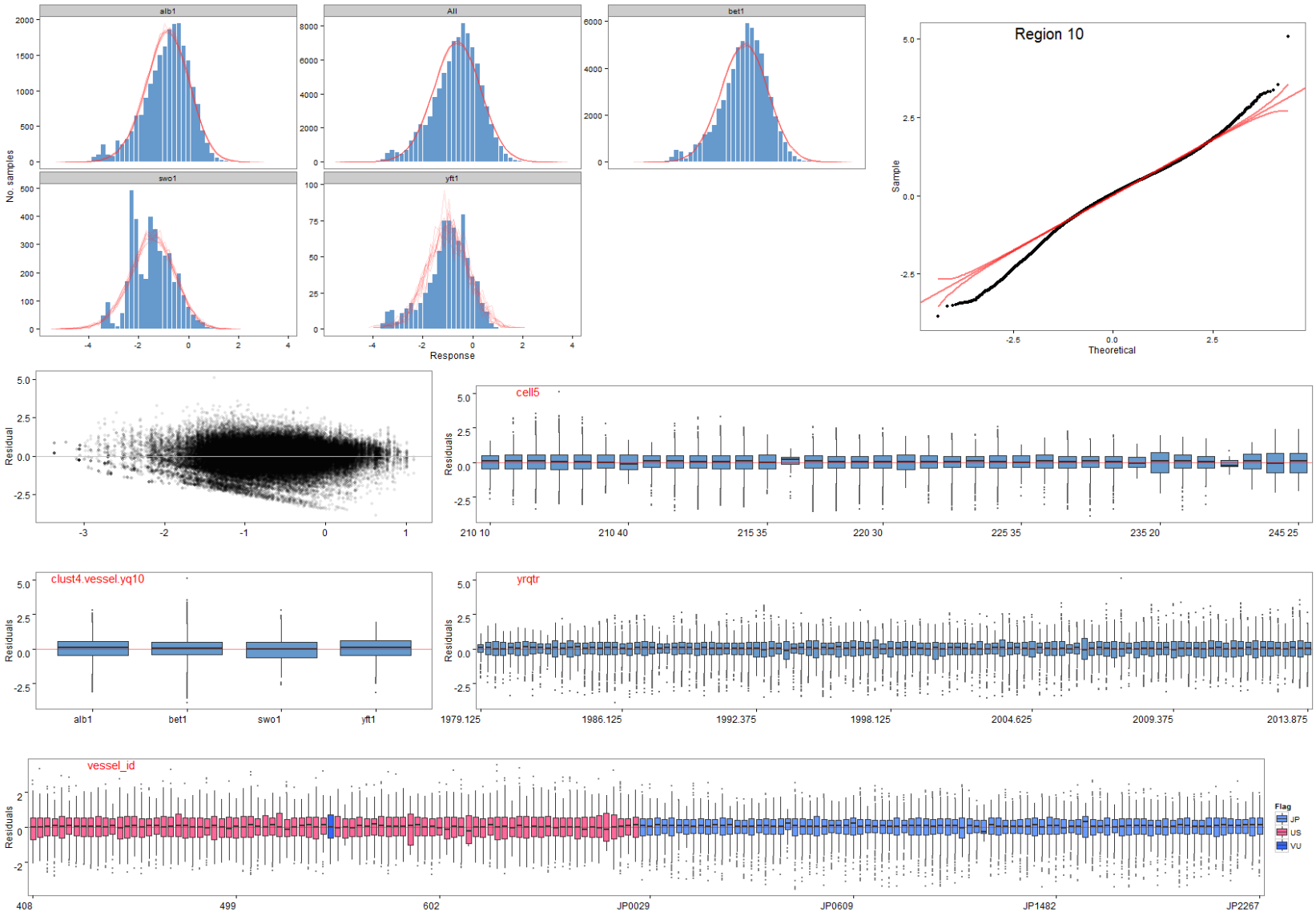


Figure 115: Diagnostic plots of fitted GLM models for region 10, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

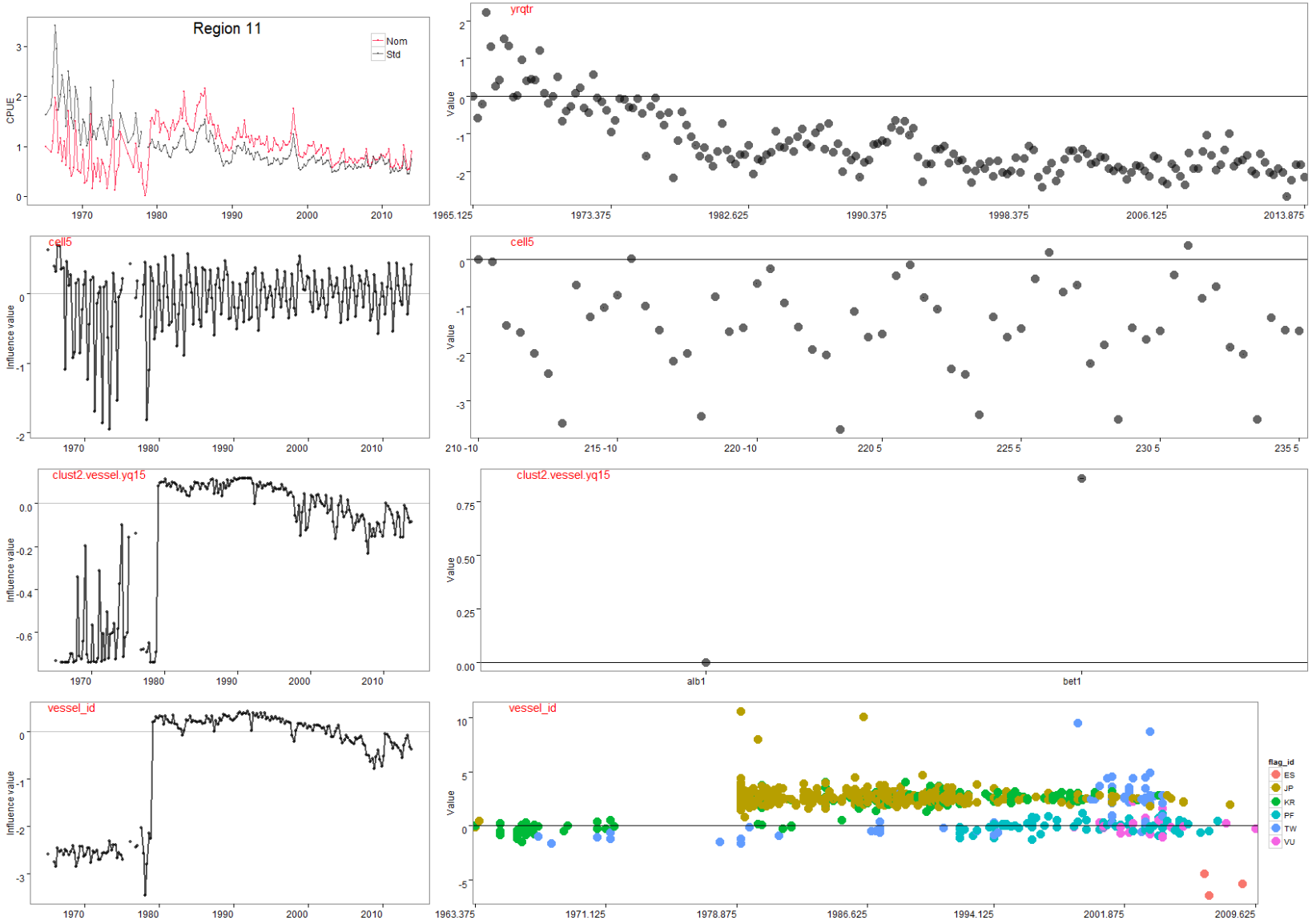


Figure 116: Summary plots of fitted GLM models for region 11, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

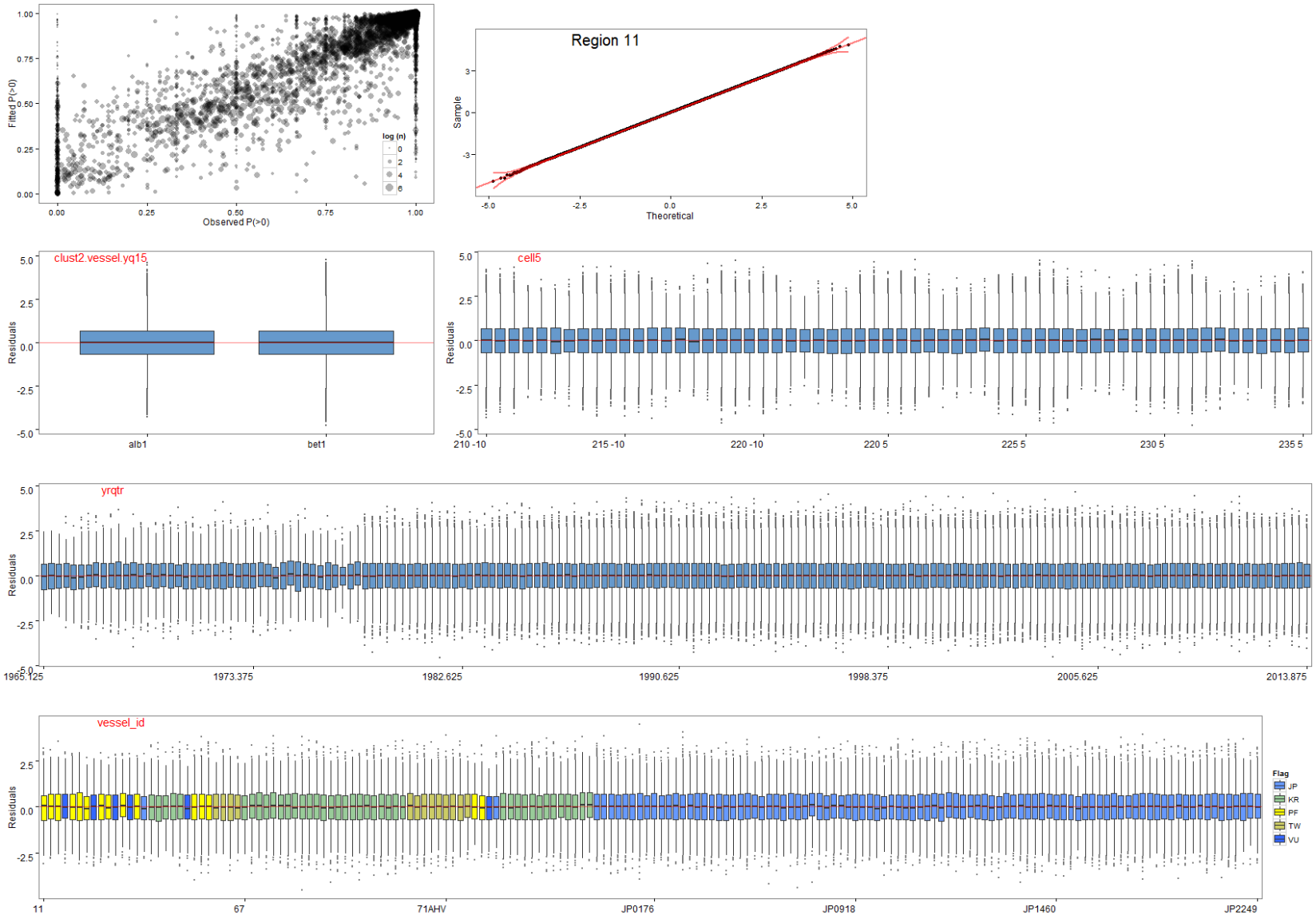


Figure 117: Diagnostic plots of fitted GLM models for region 11, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

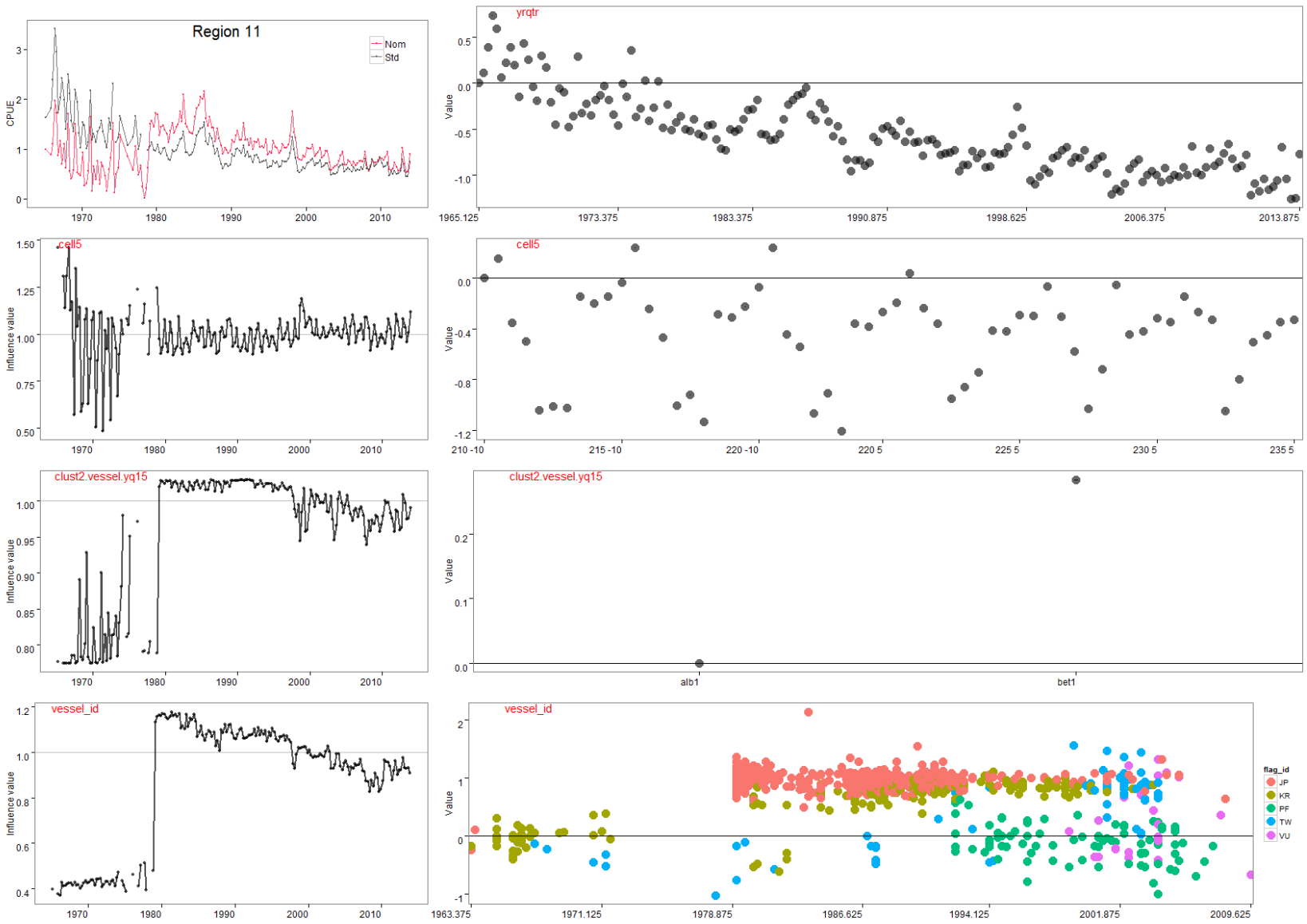


Figure 118: Summary plots of fitted GLM models for region 11, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

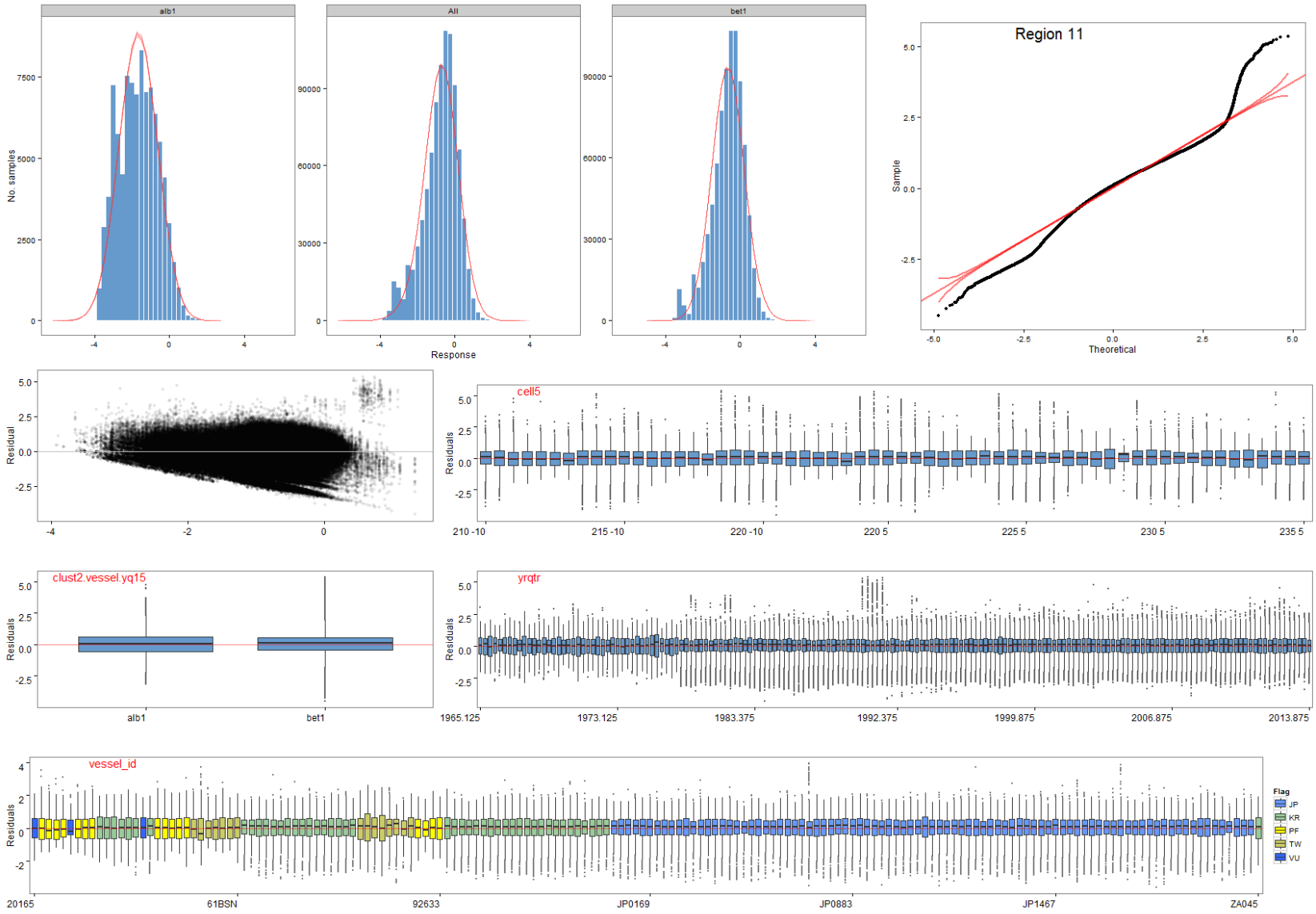


Figure 119: Diagnostic plots of fitted GLM models for region 11, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

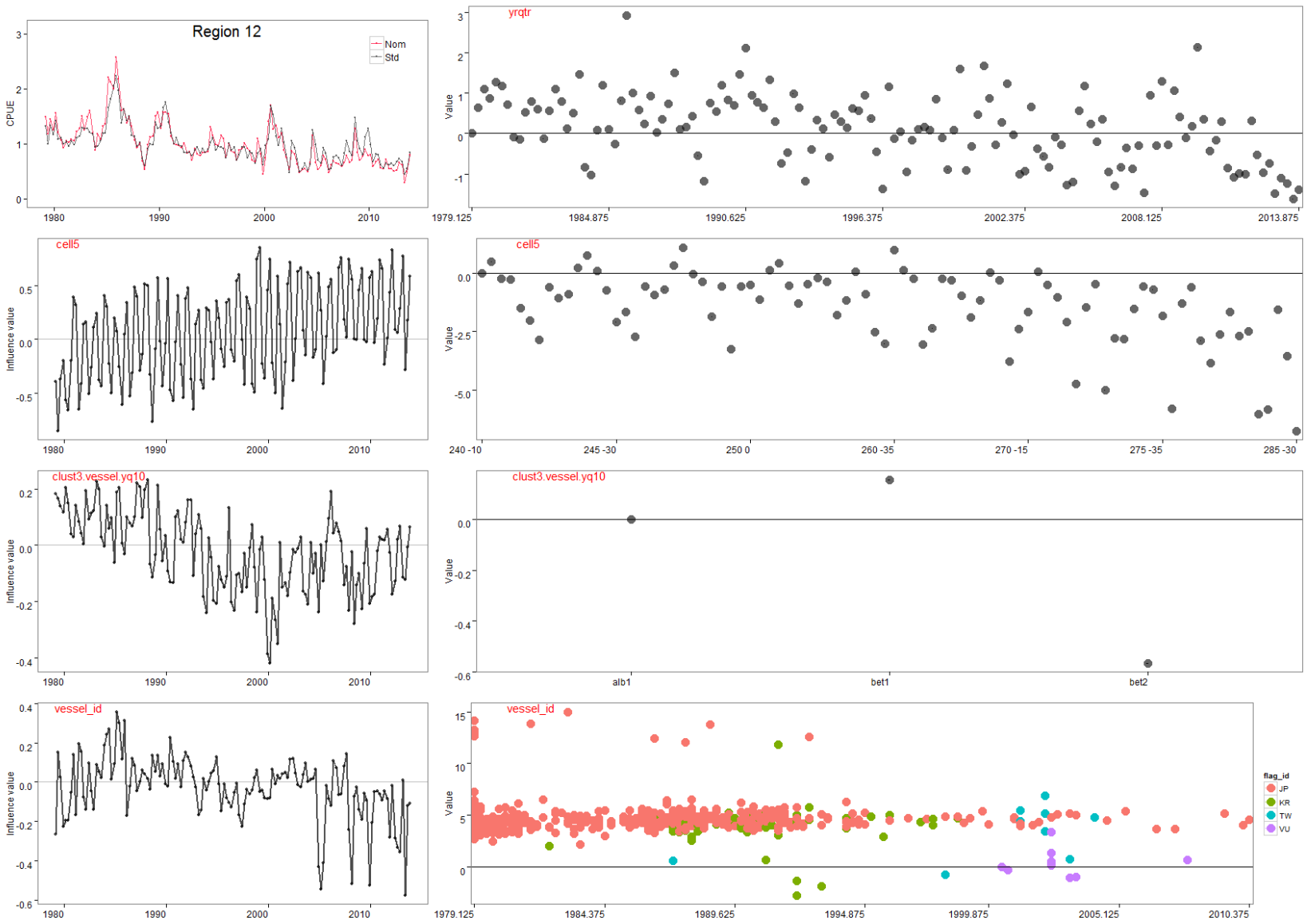


Figure 120: Summary plots of fitted GLM models for region 12, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

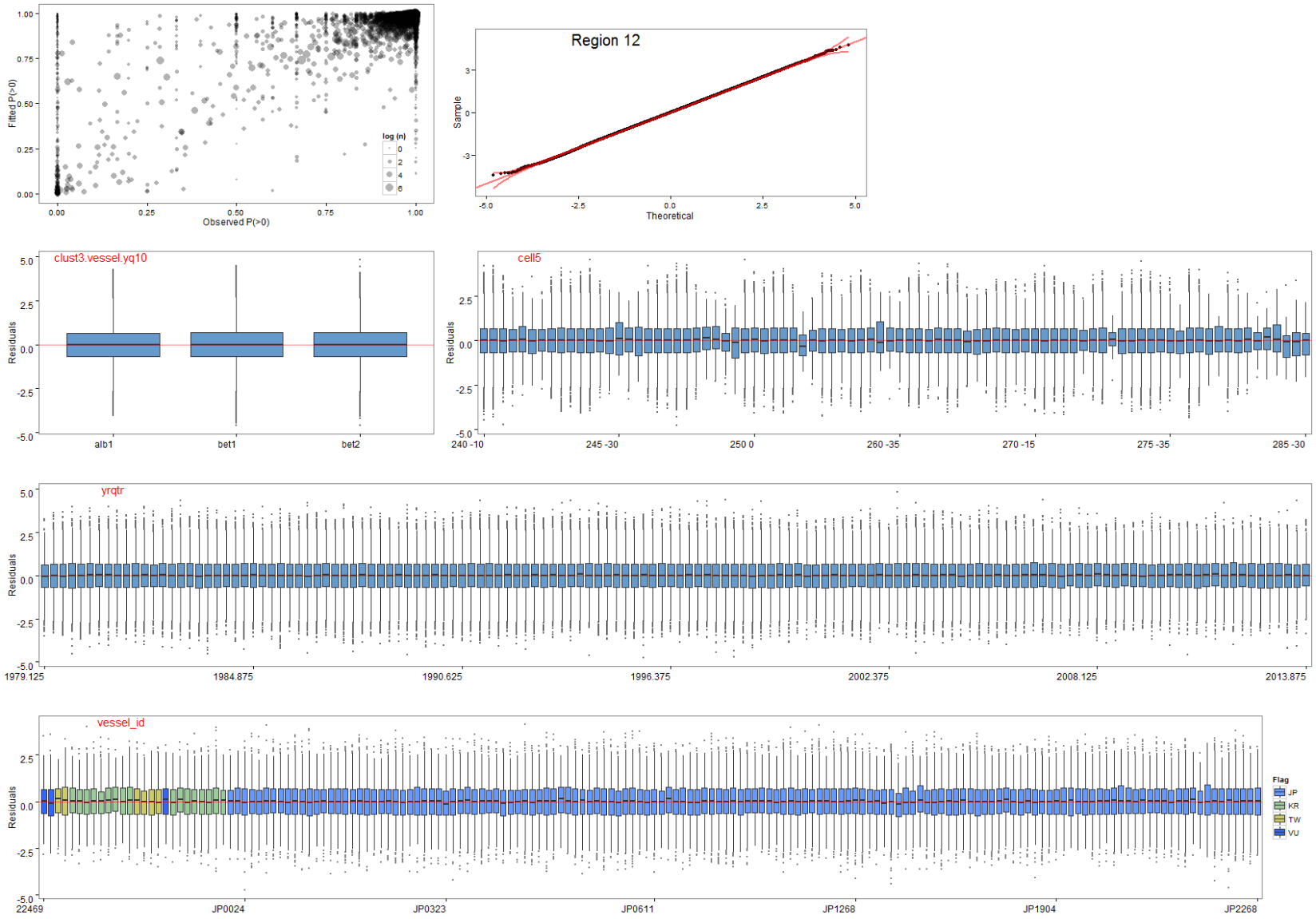


Figure 121: Diagnostic plots of fitted GLM models for region 12, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

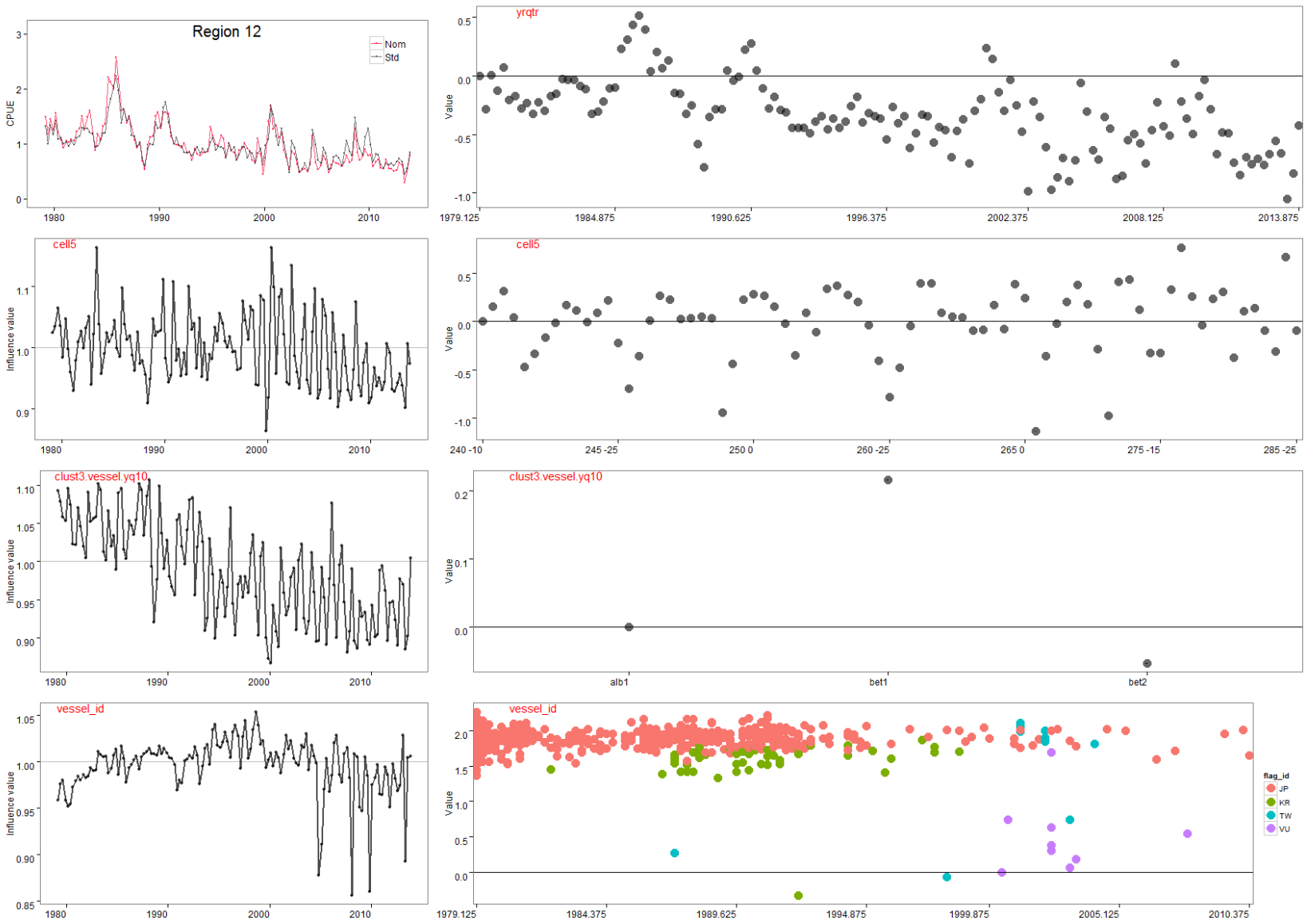


Figure 122: Summary plots of fitted GLM models for region 12, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

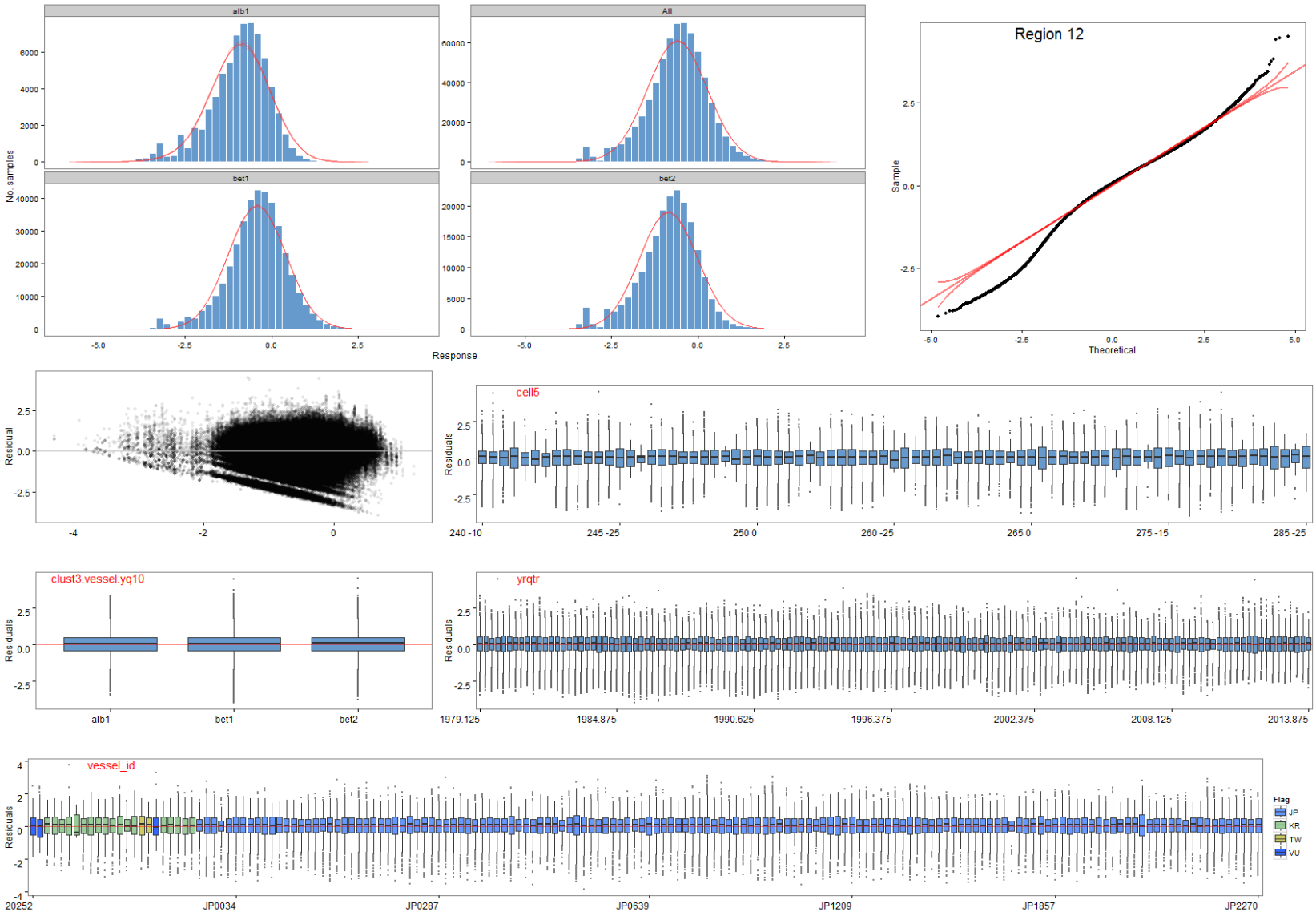


Figure 123: Diagnostic plots of fitted GLM models for region 12, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

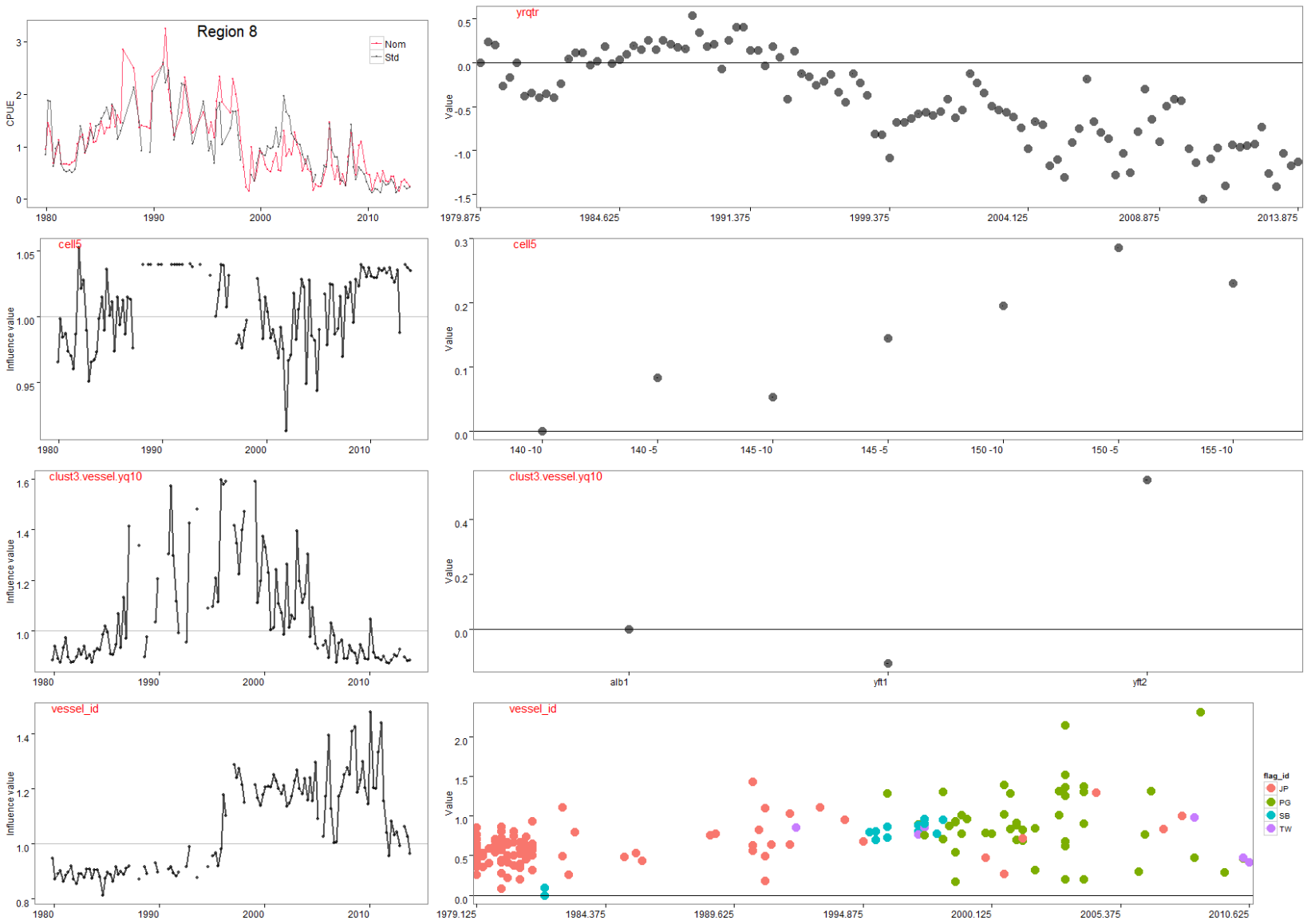


Figure 124: Summary plots of fitted gamma GLM models for region 8, step 1; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

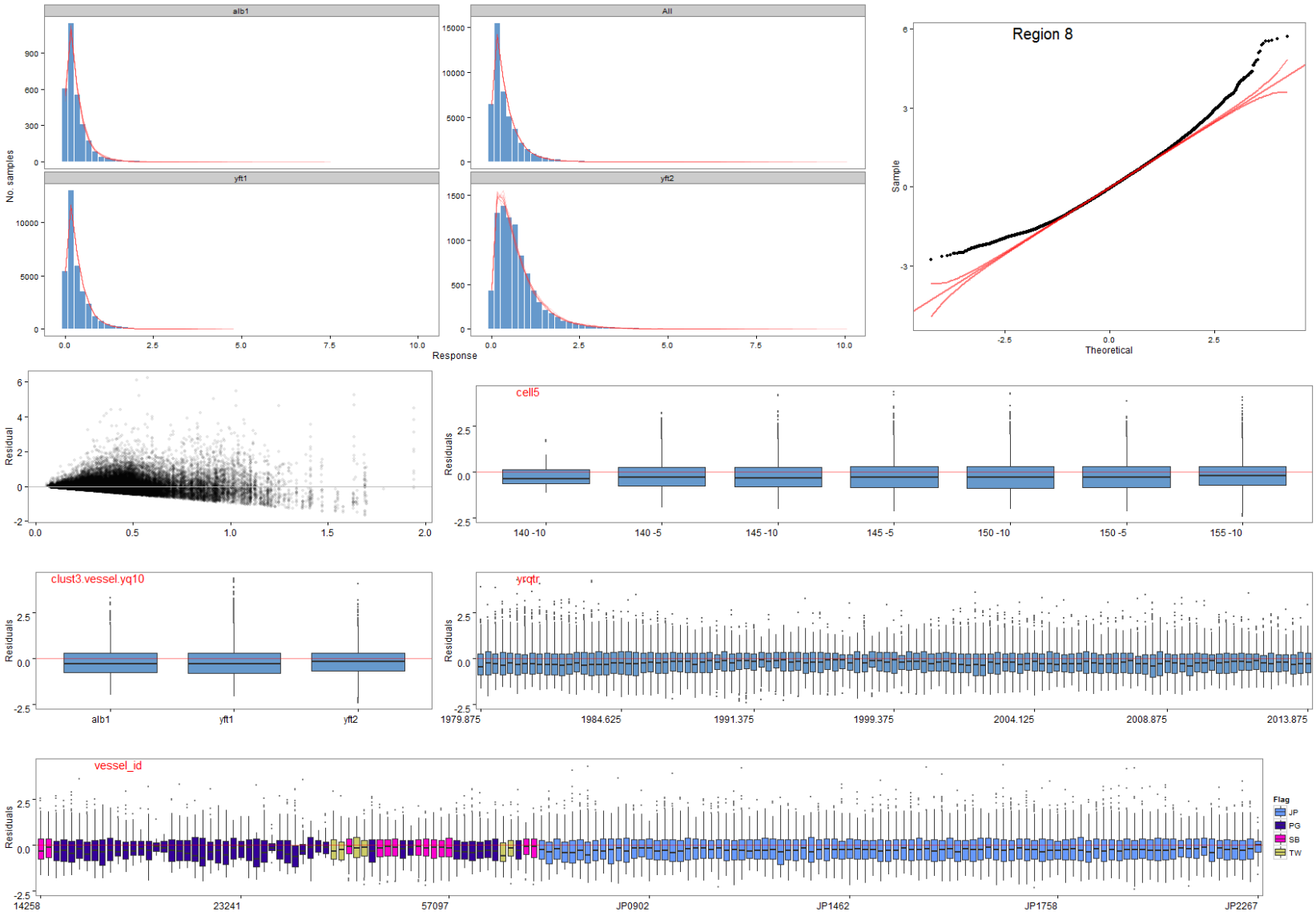


Figure 125: Diagnostic plots of fitted gamma GLM models for region 8, step 1, showing characteristics of the model residuals and comparisons between observed and simulated data.

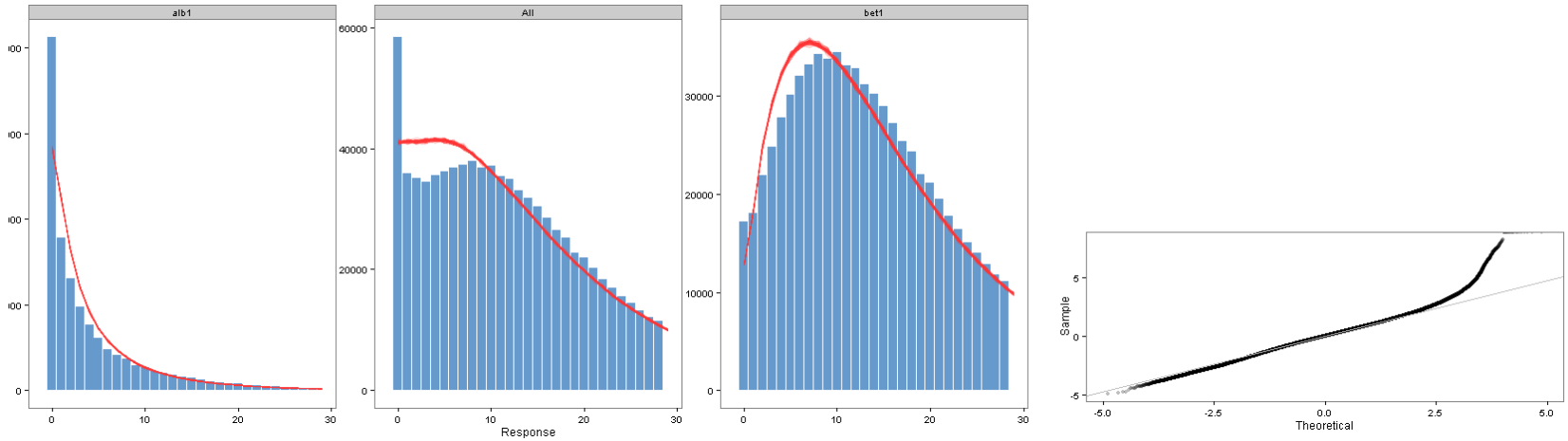


Figure 126: Example comparison of the observed dataset (blue bars) and 20 datasets simulated from the negative binomial GLM model (red lines) for region 11, step 1.

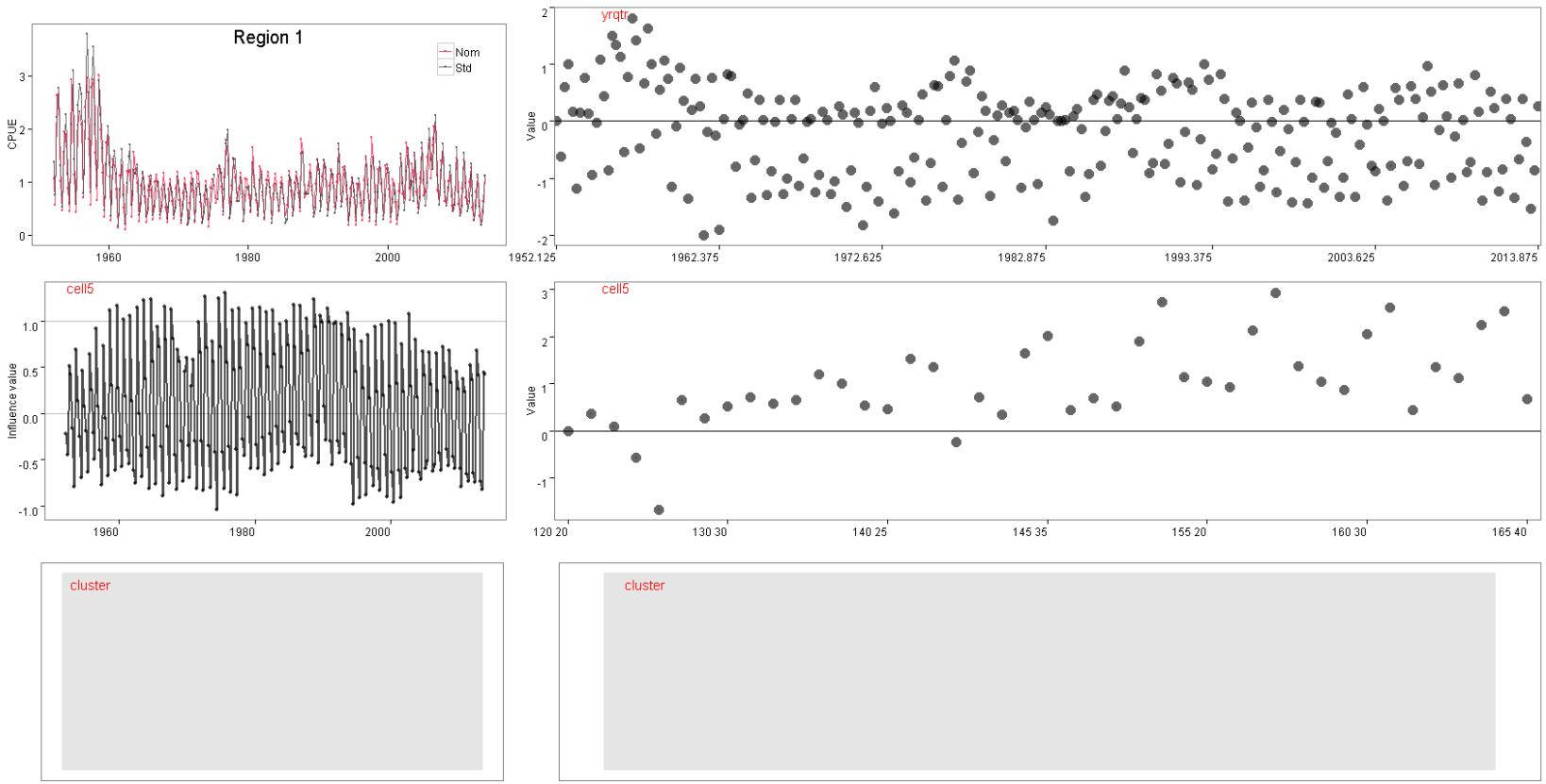


Figure 127: Summary plots of fitted GLM models for region 1, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

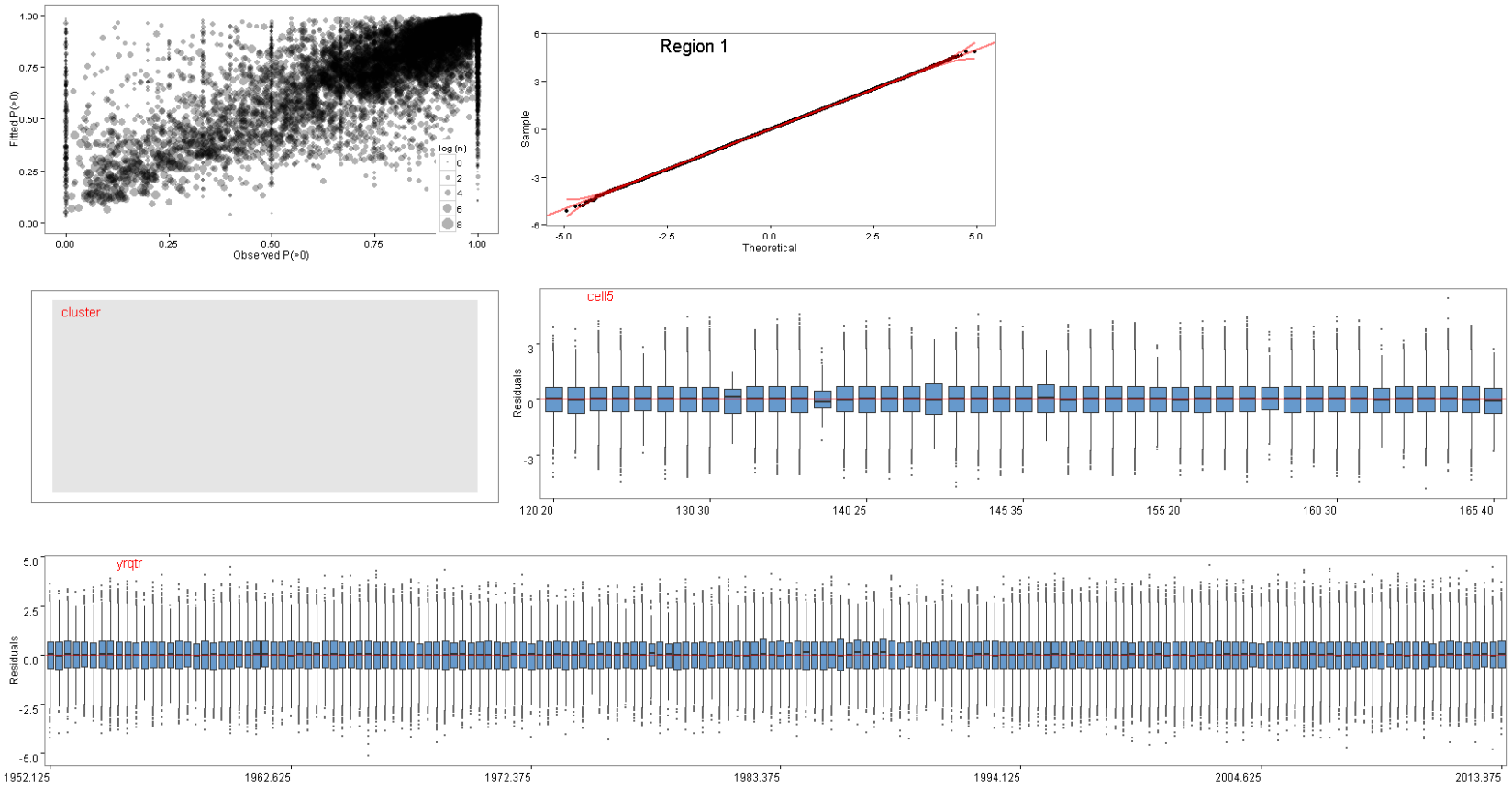


Figure 128: Diagnostic plots of fitted GLM models for region 1, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

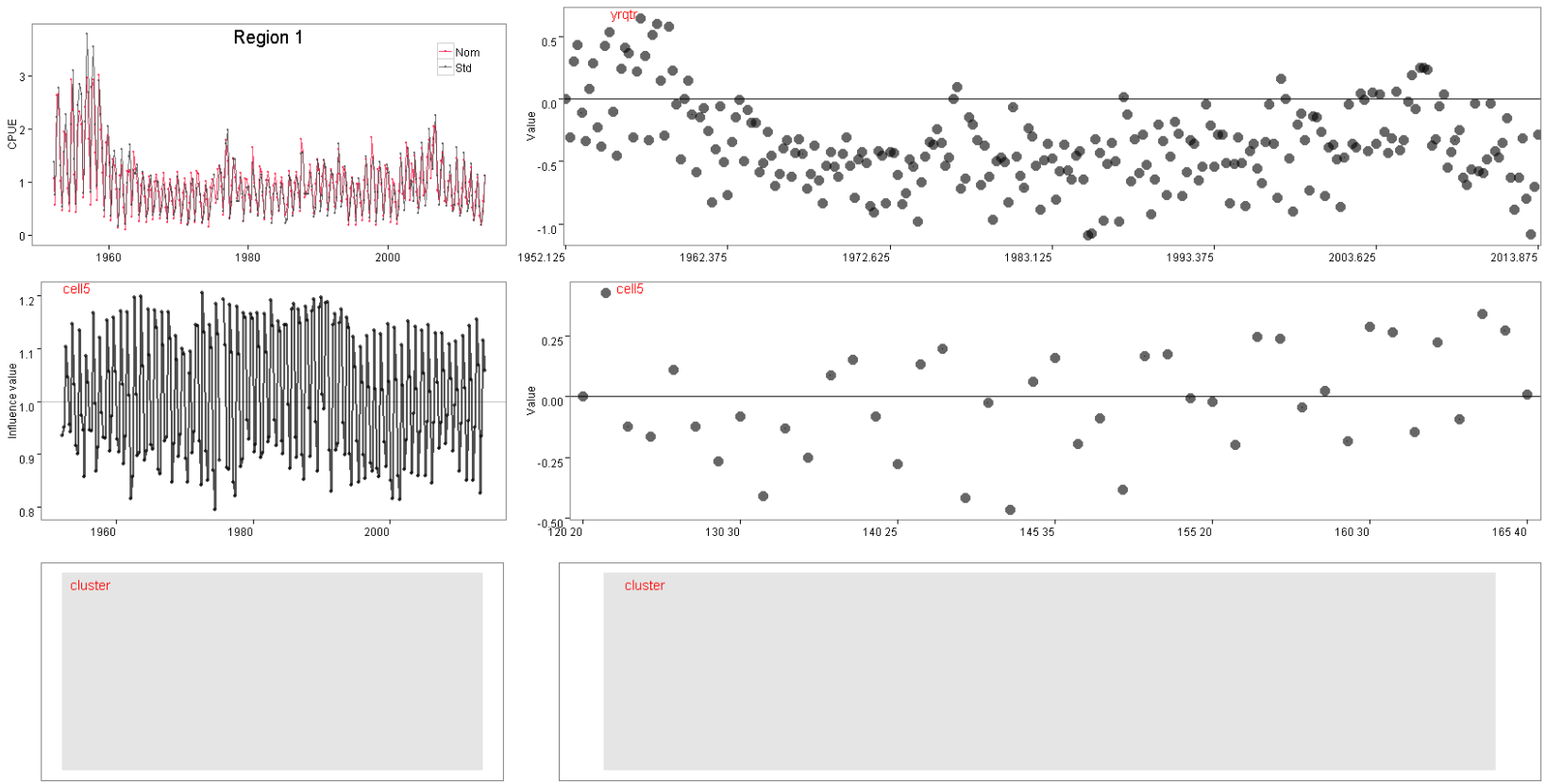


Figure 129: Summary plots of fitted GLM models for region 1, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

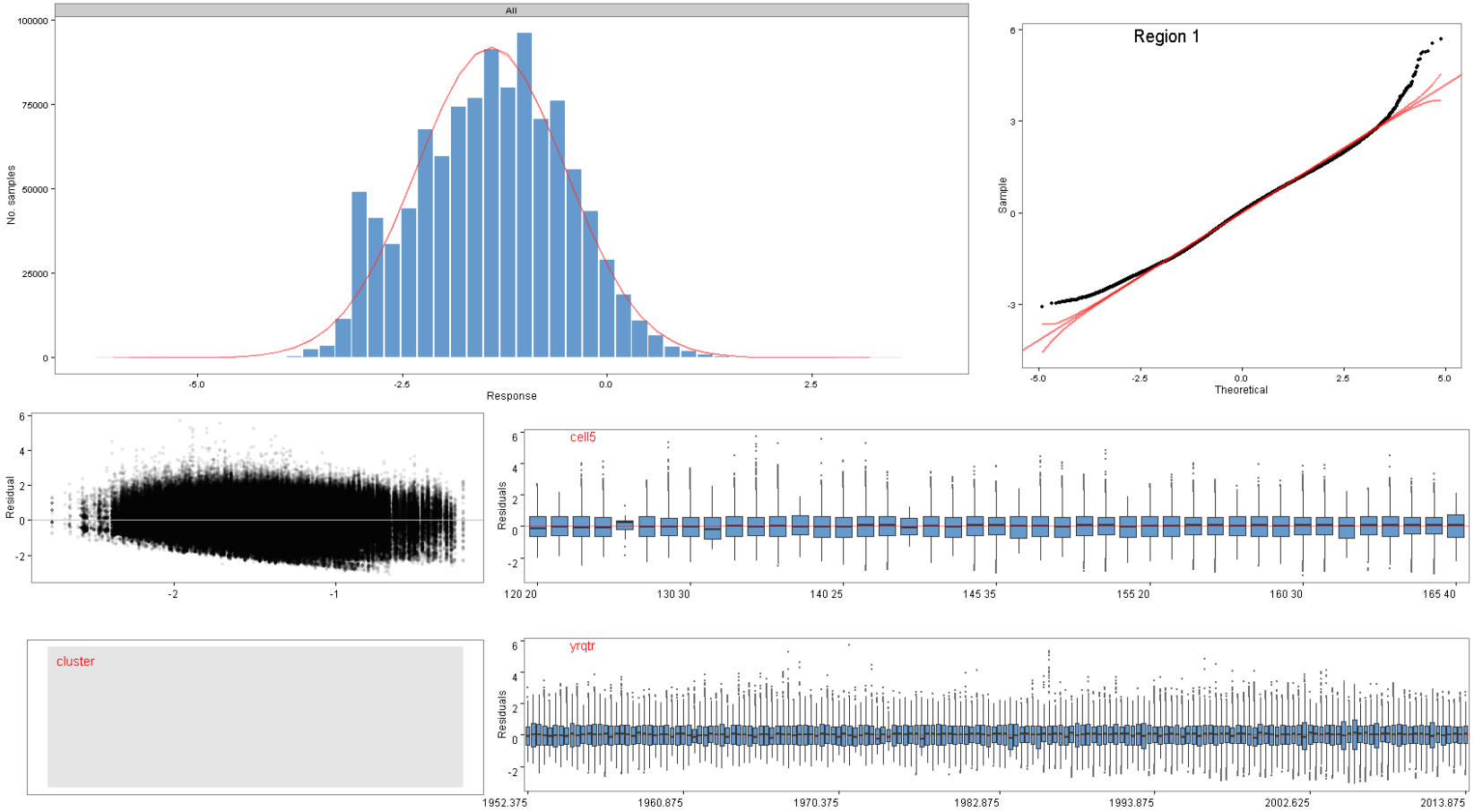


Figure 130: Diagnostic plots of fitted GLM models for region 1, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

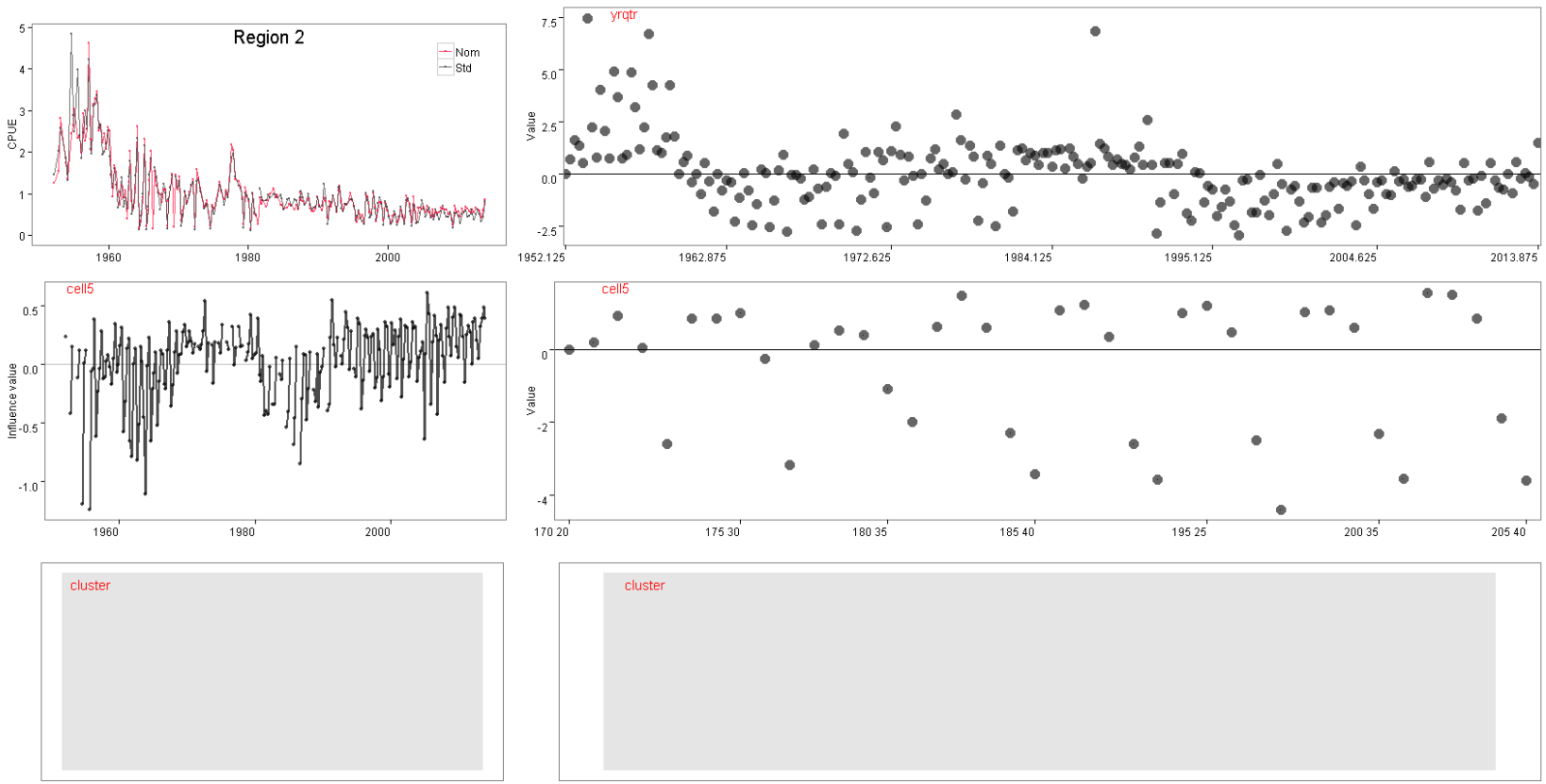


Figure 131: Summary plots of fitted GLM models for region 2, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

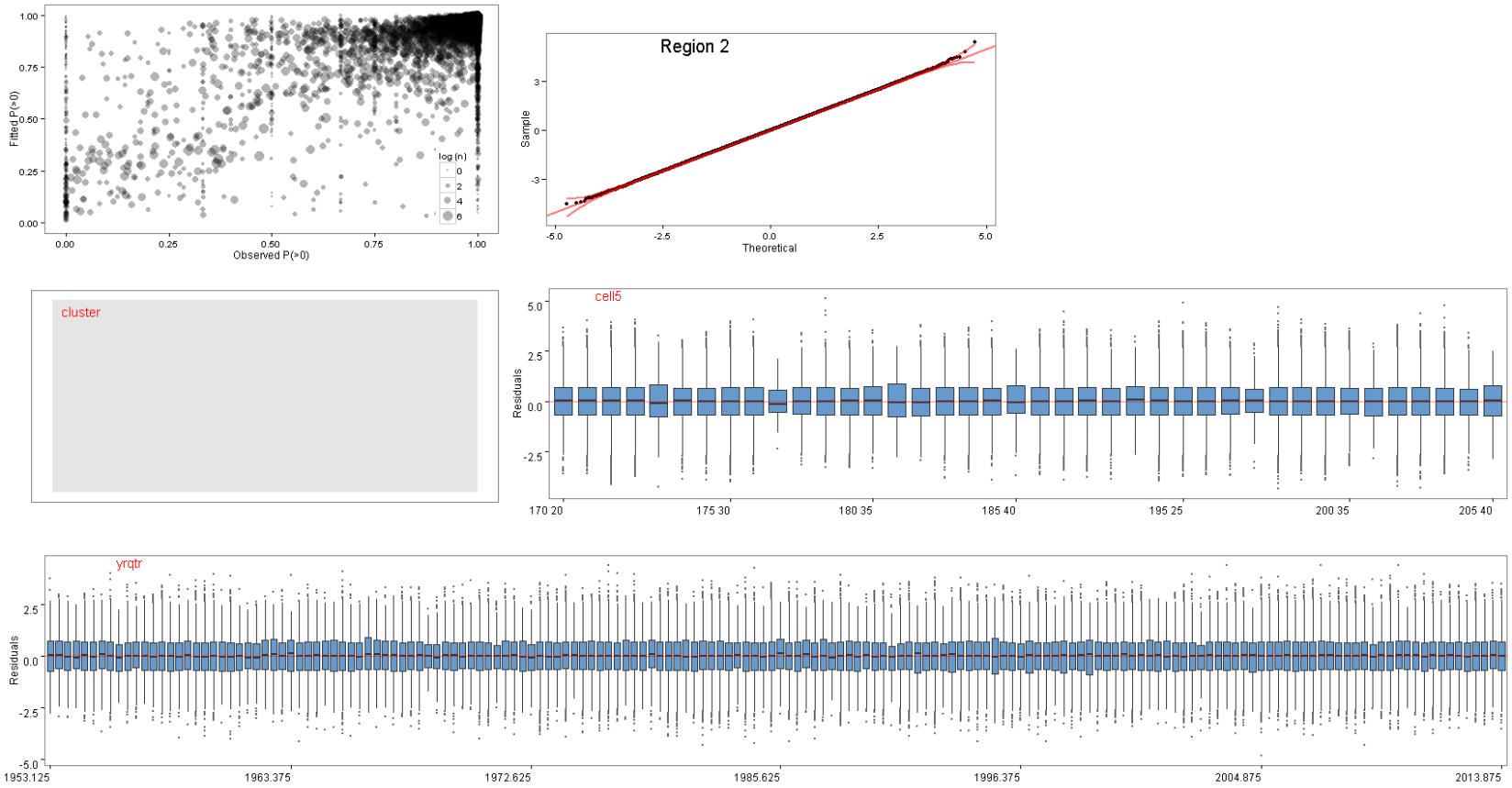


Figure 132: Diagnostic plots of fitted GLM models for region 2, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

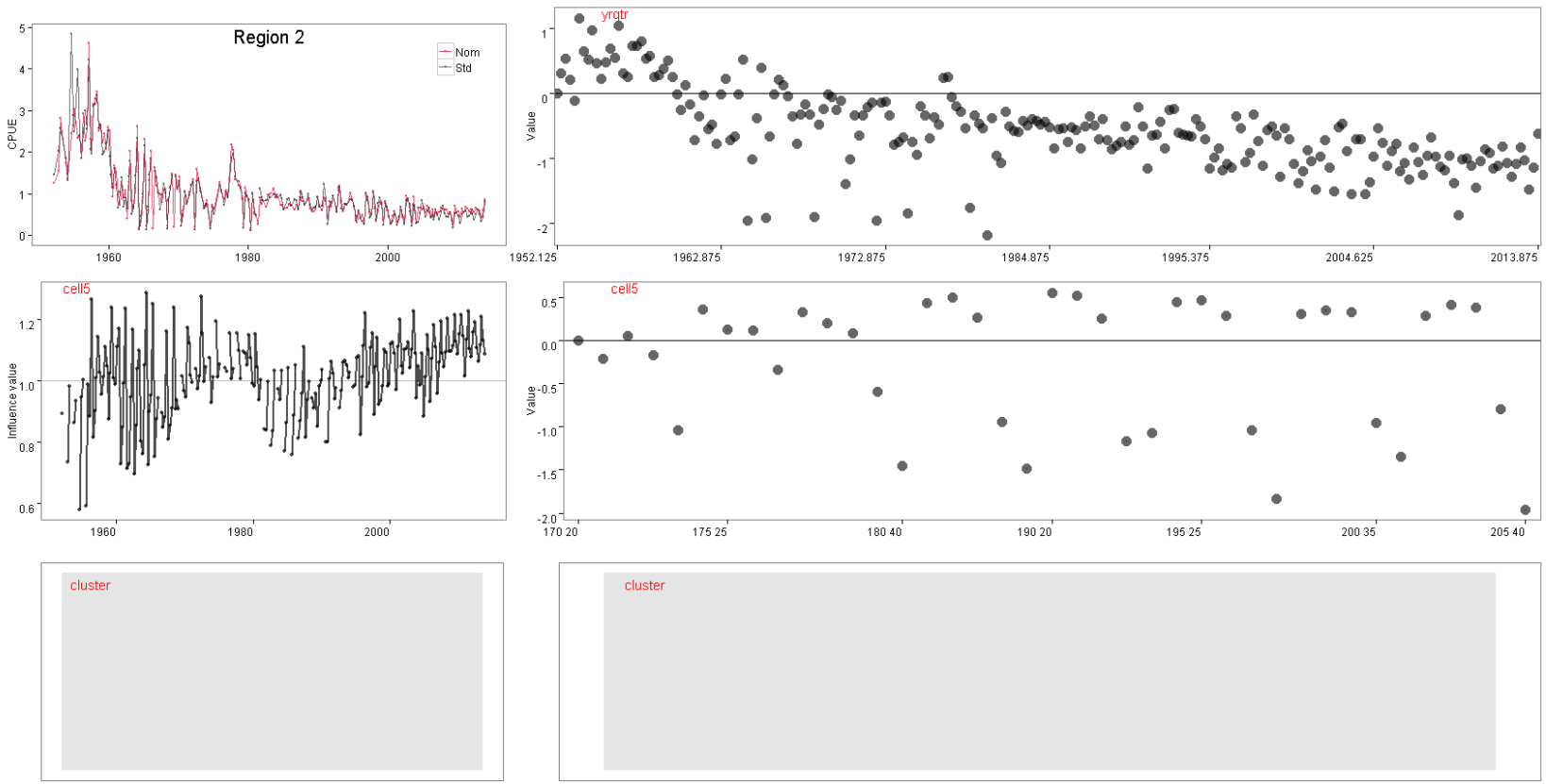


Figure 133: Summary plots of fitted GLM models for region 2, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

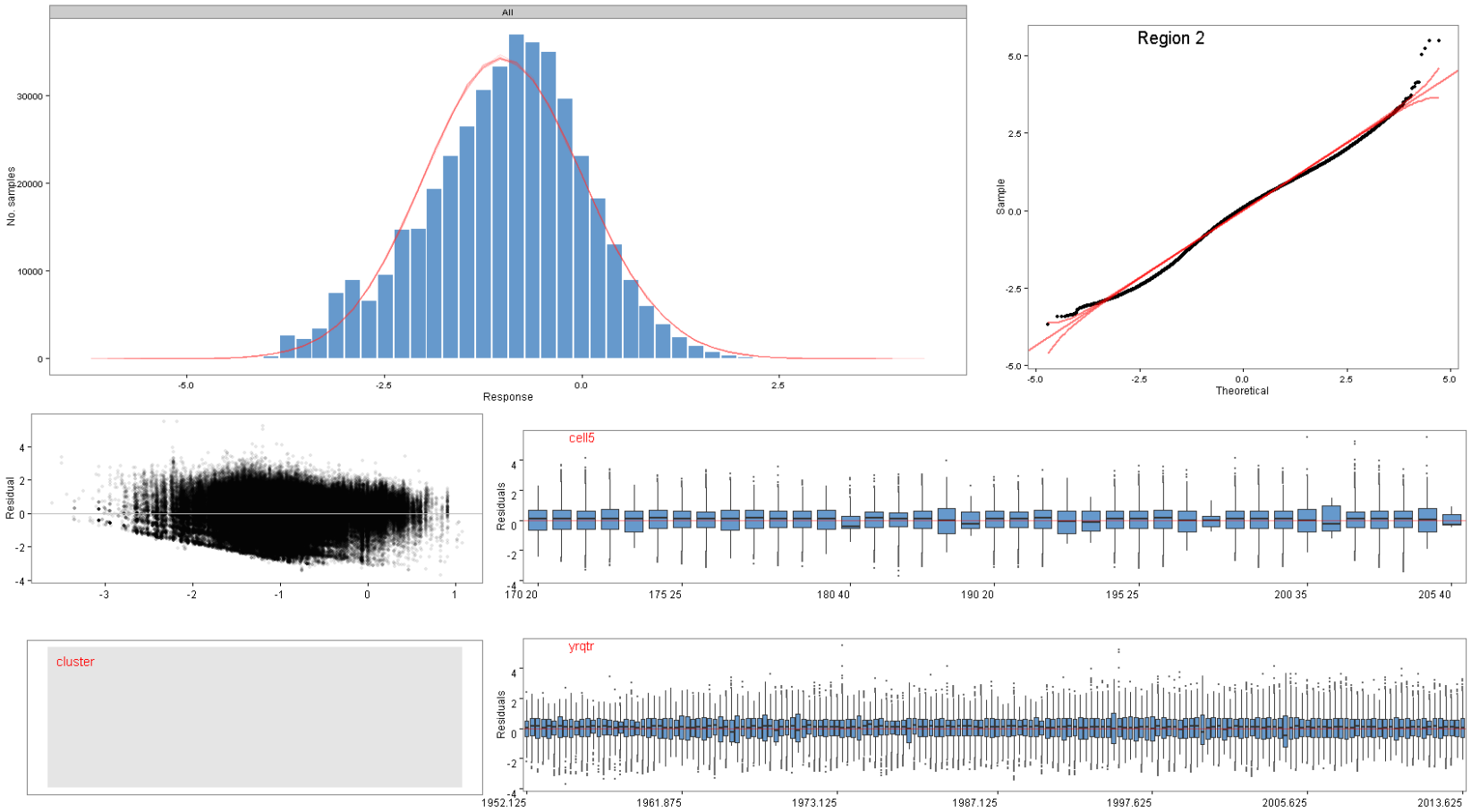


Figure 134: Diagnostic plots of fitted GLM models for region 2, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

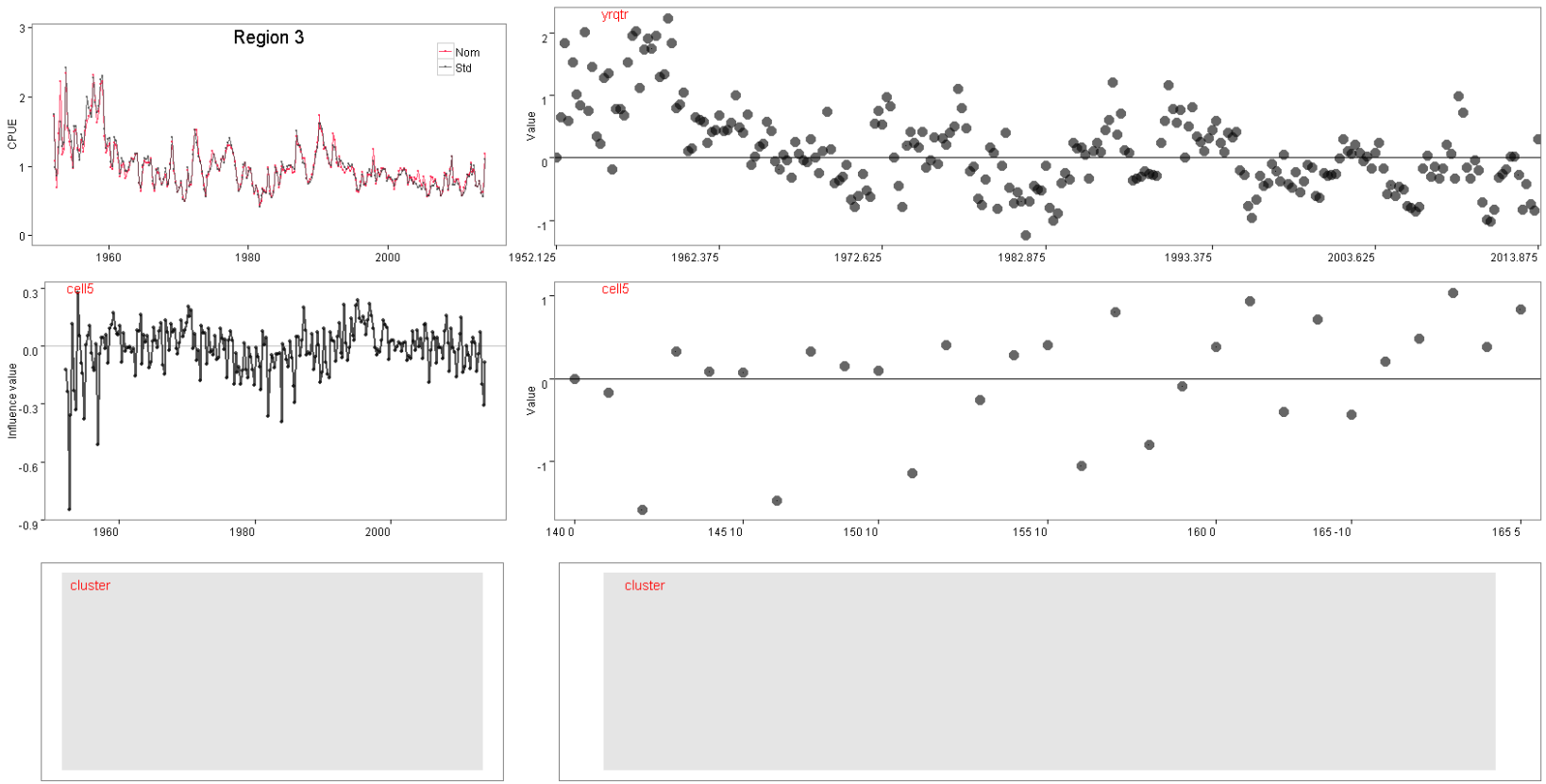


Figure 135: Summary plots of fitted GLM models for region 3, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

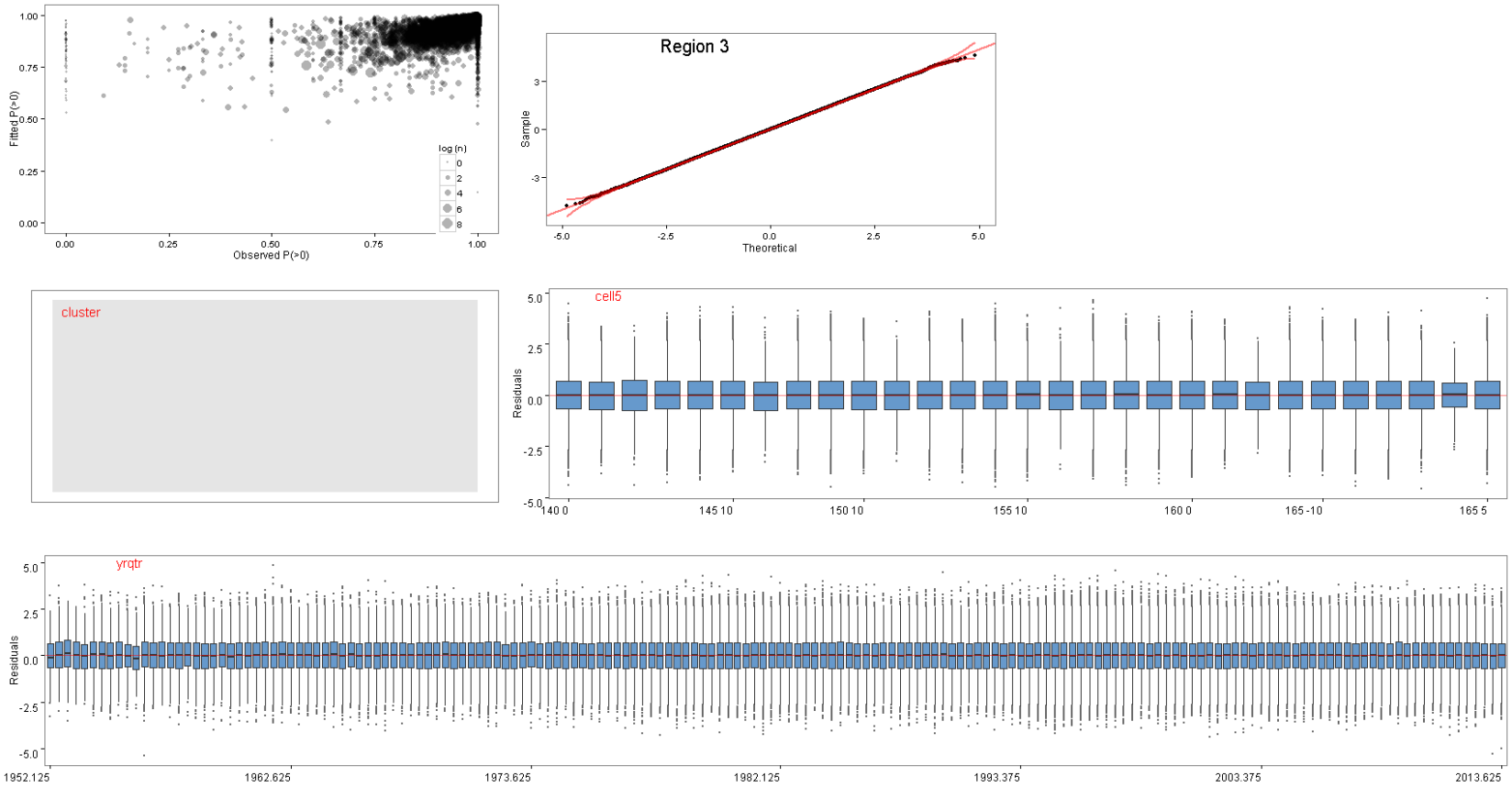


Figure 136: Diagnostic plots of fitted GLM models for region 3, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

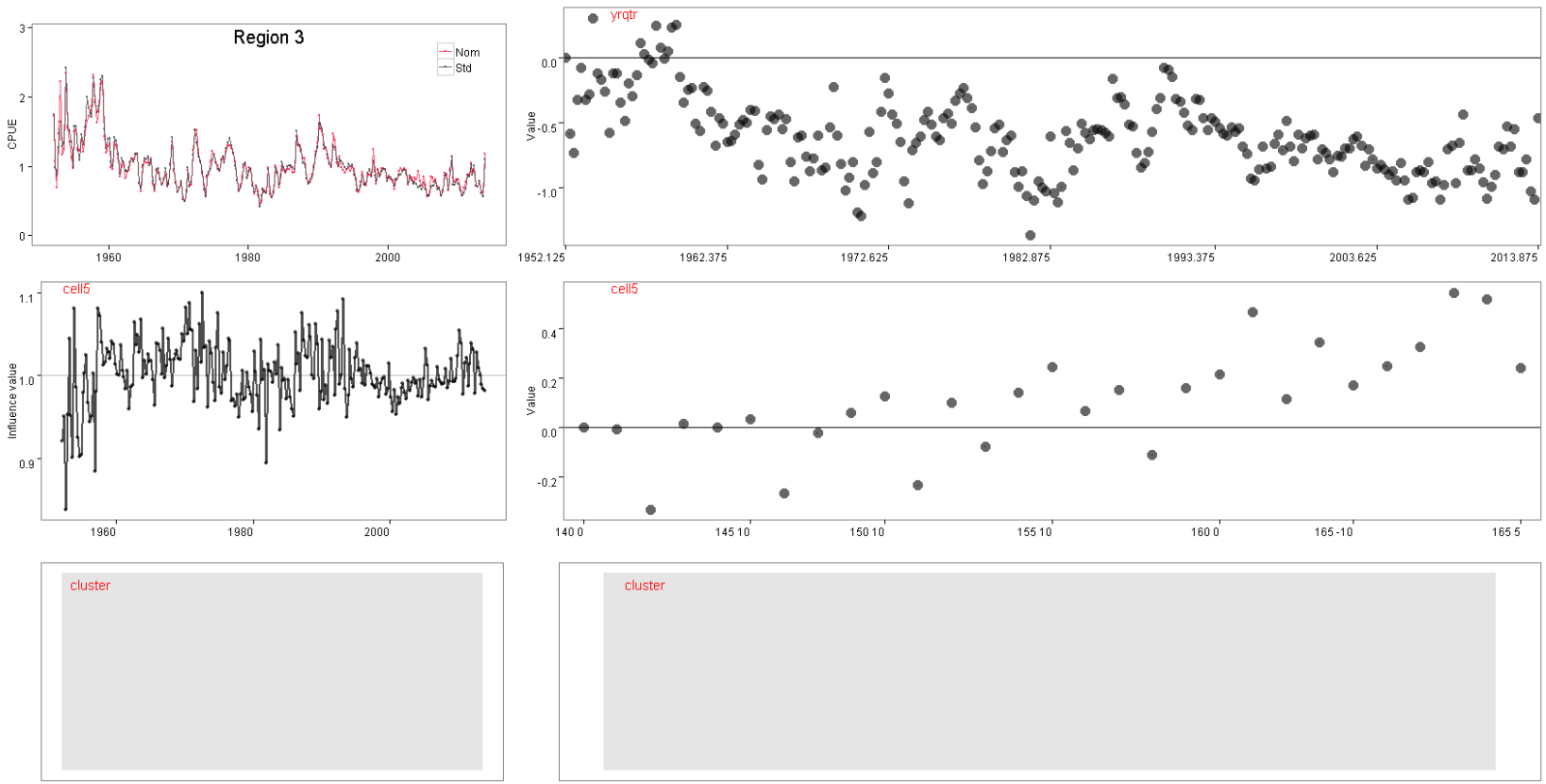


Figure 137: Summary plots of fitted GLM models for region 3, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

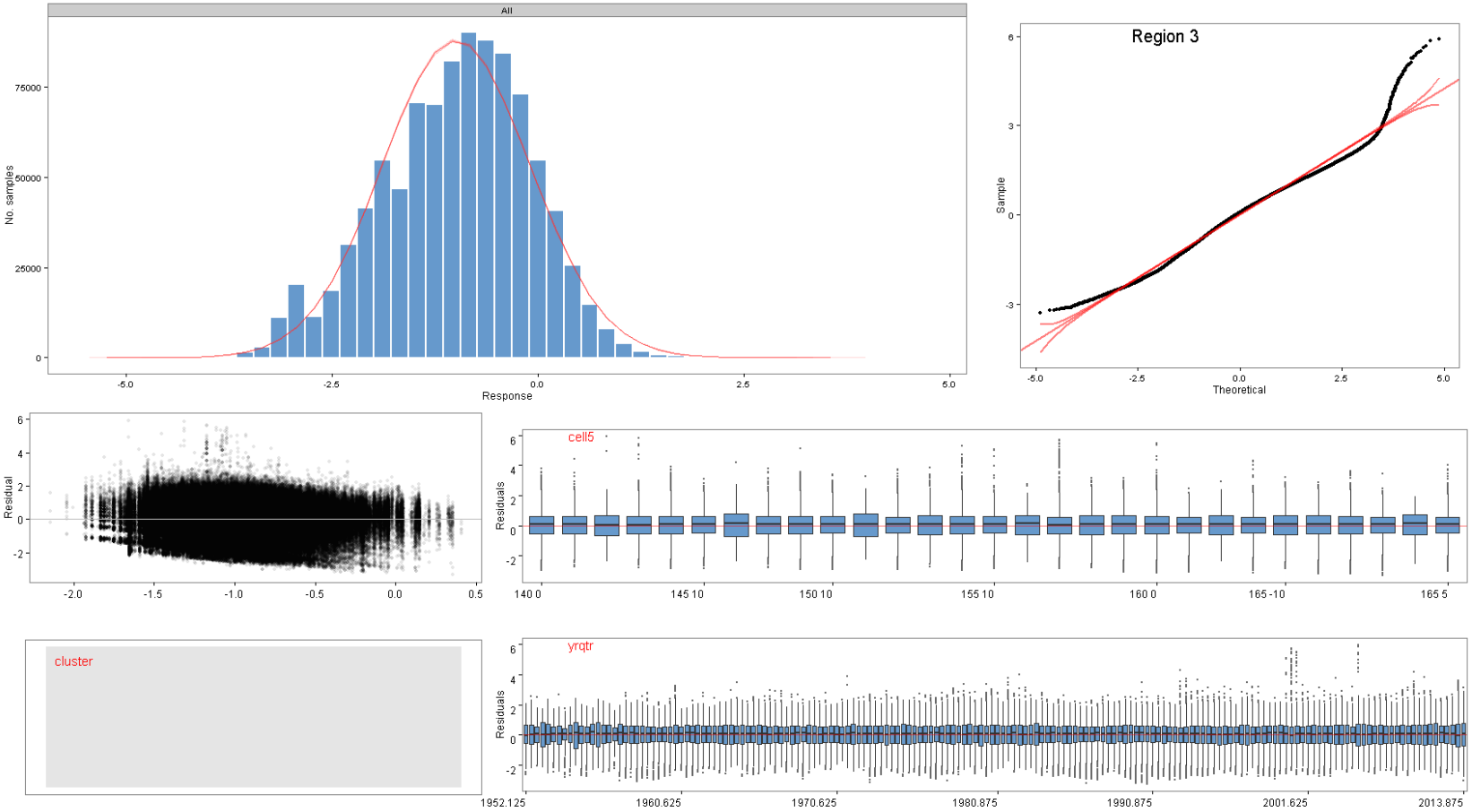


Figure 138: Diagnostic plots of fitted GLM models for region 3, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

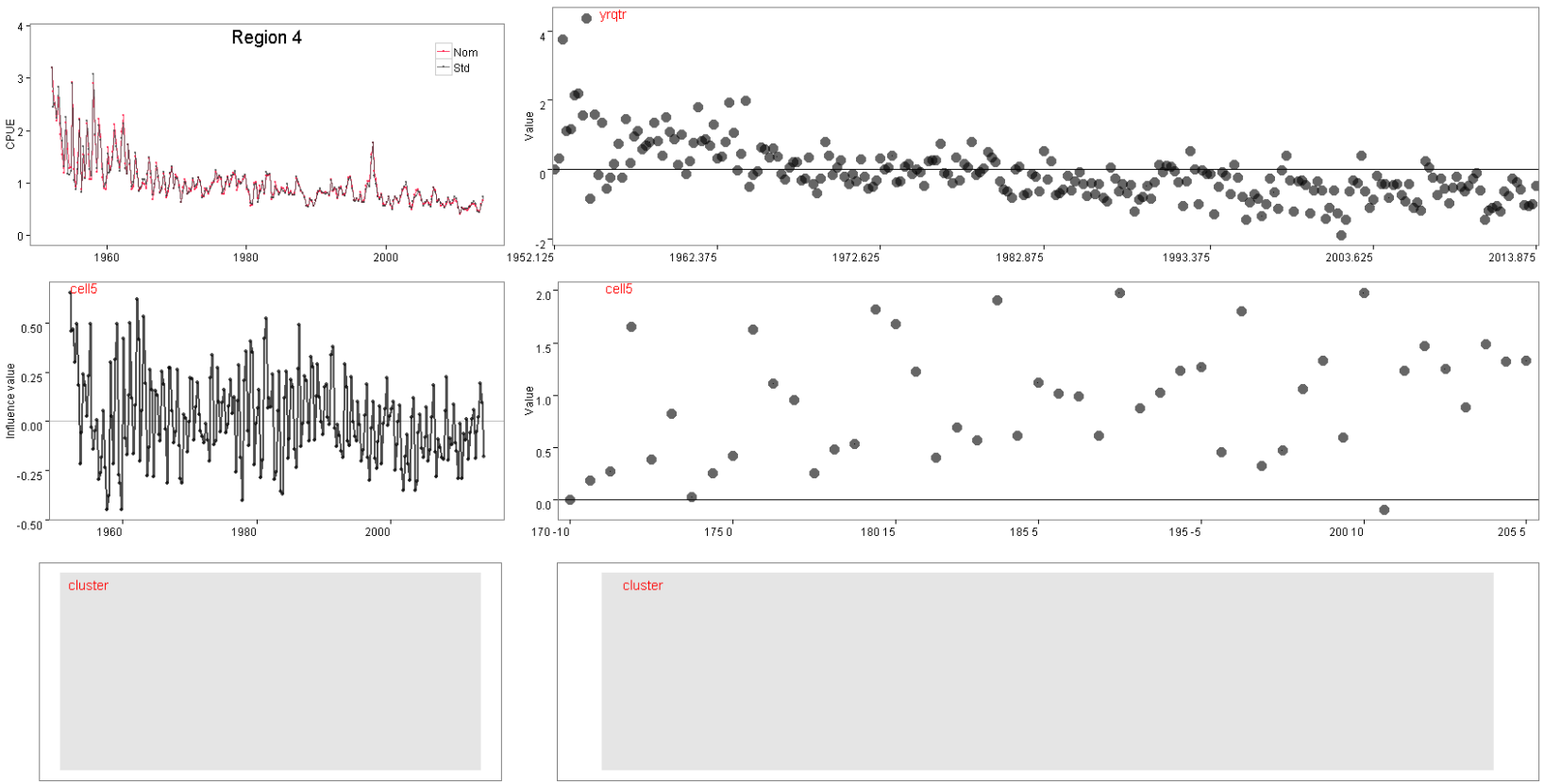


Figure 139: Summary plots of fitted GLM models for region 4, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

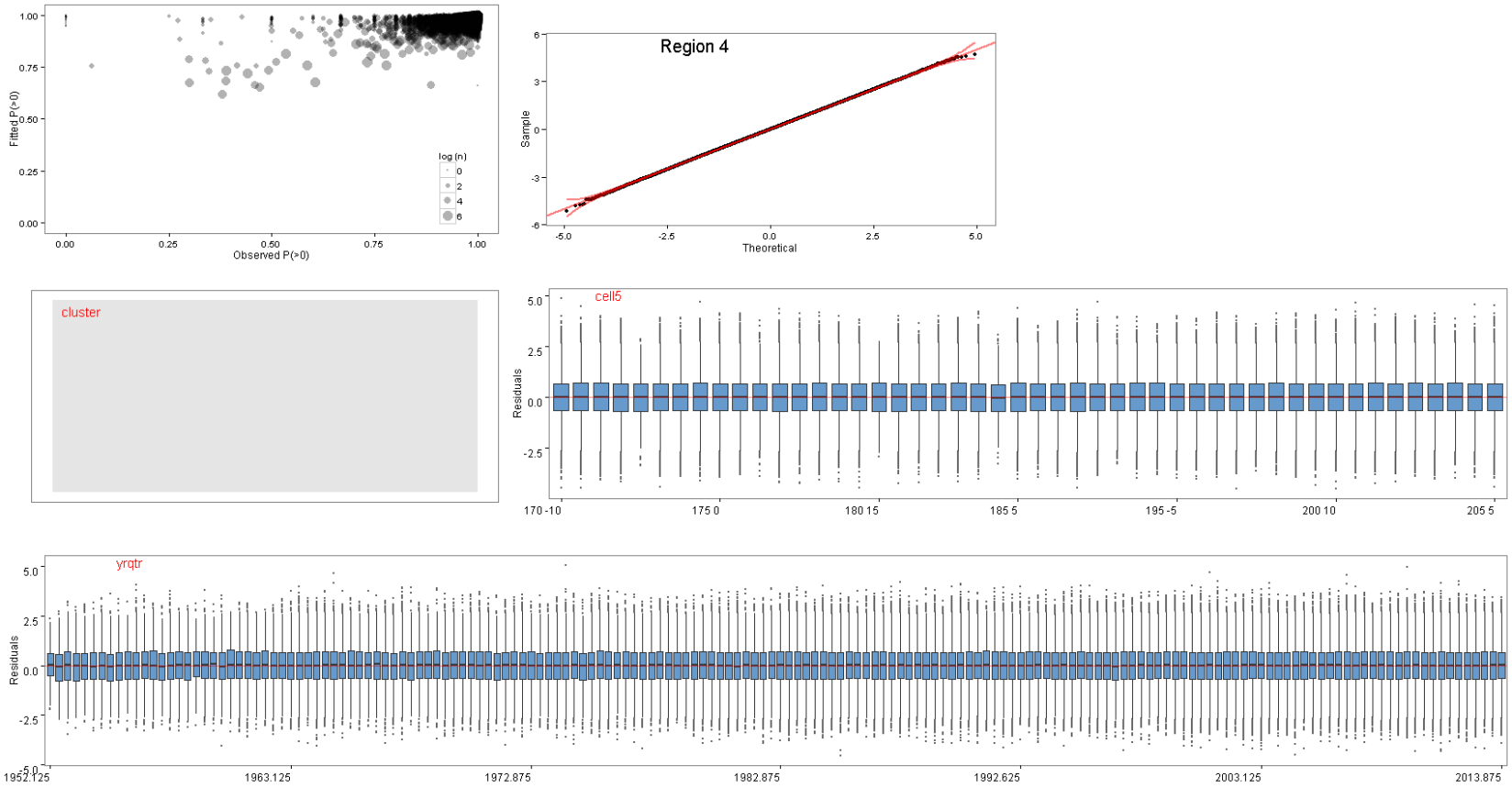


Figure 140: Diagnostic plots of fitted GLM models for region 4, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

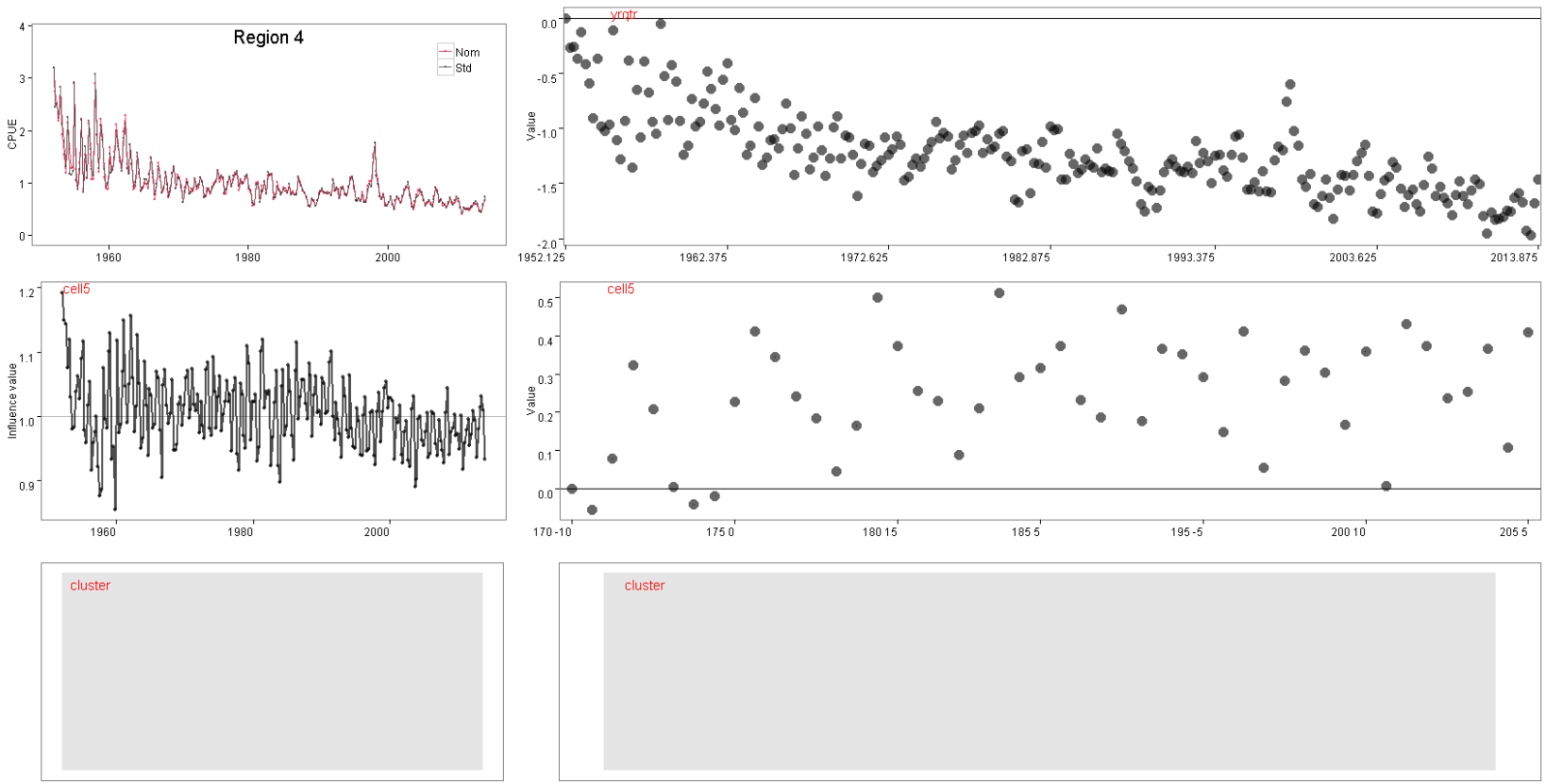


Figure 141: Summary plots of fitted GLM models for region 4, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

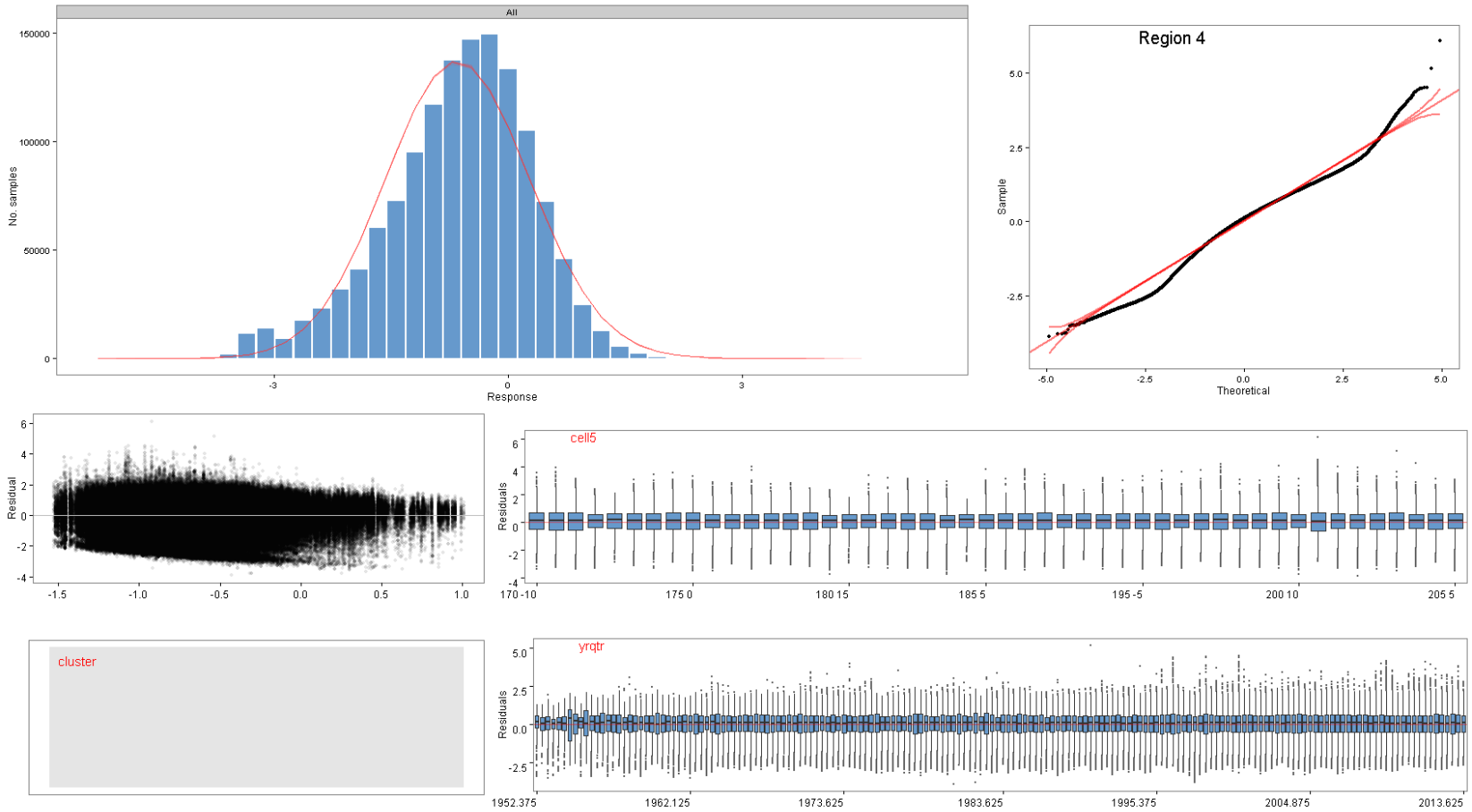


Figure 142: Diagnostic plots of fitted GLM models for region 4, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

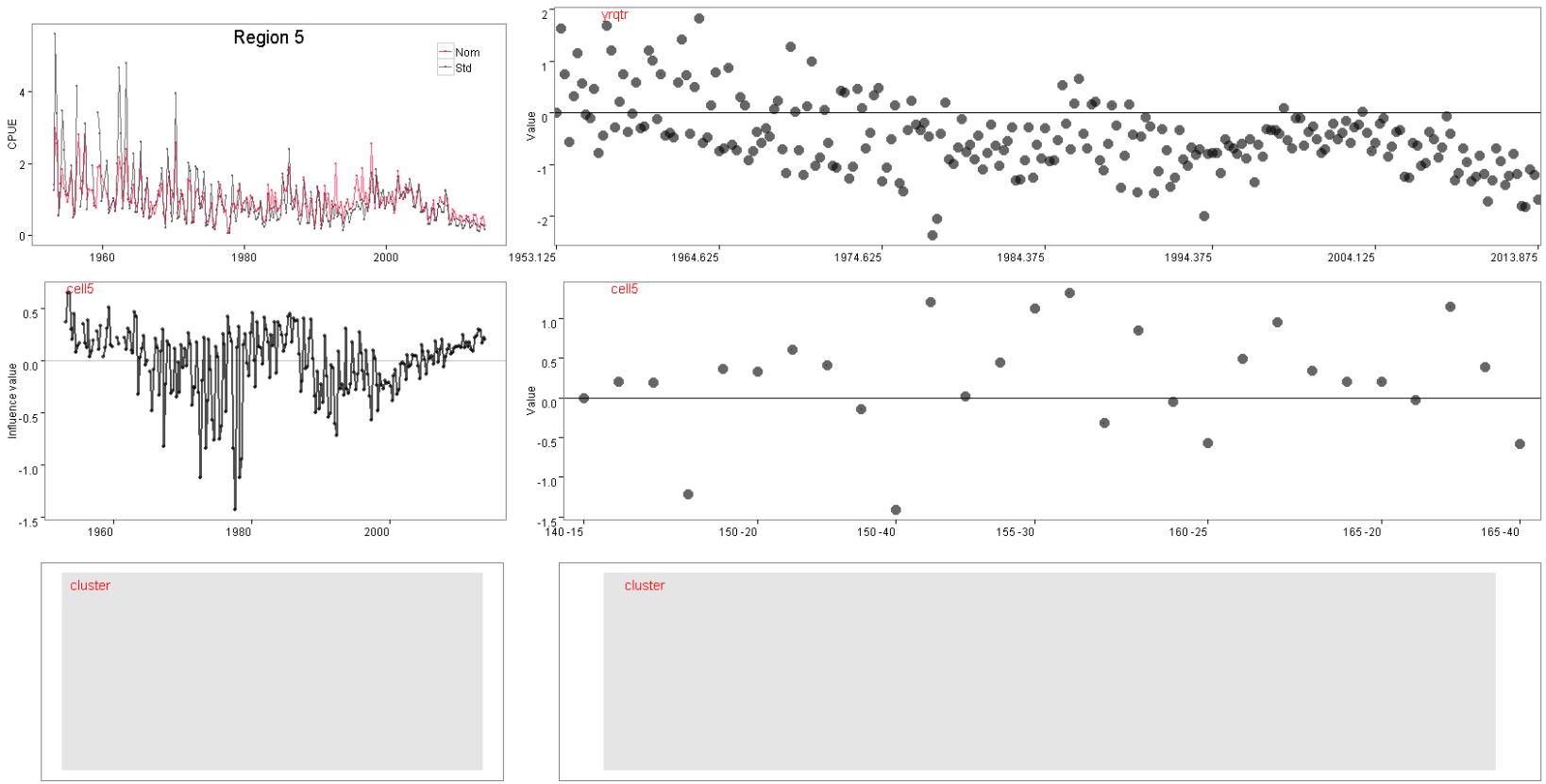


Figure 143: Summary plots of fitted GLM models for region 5, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

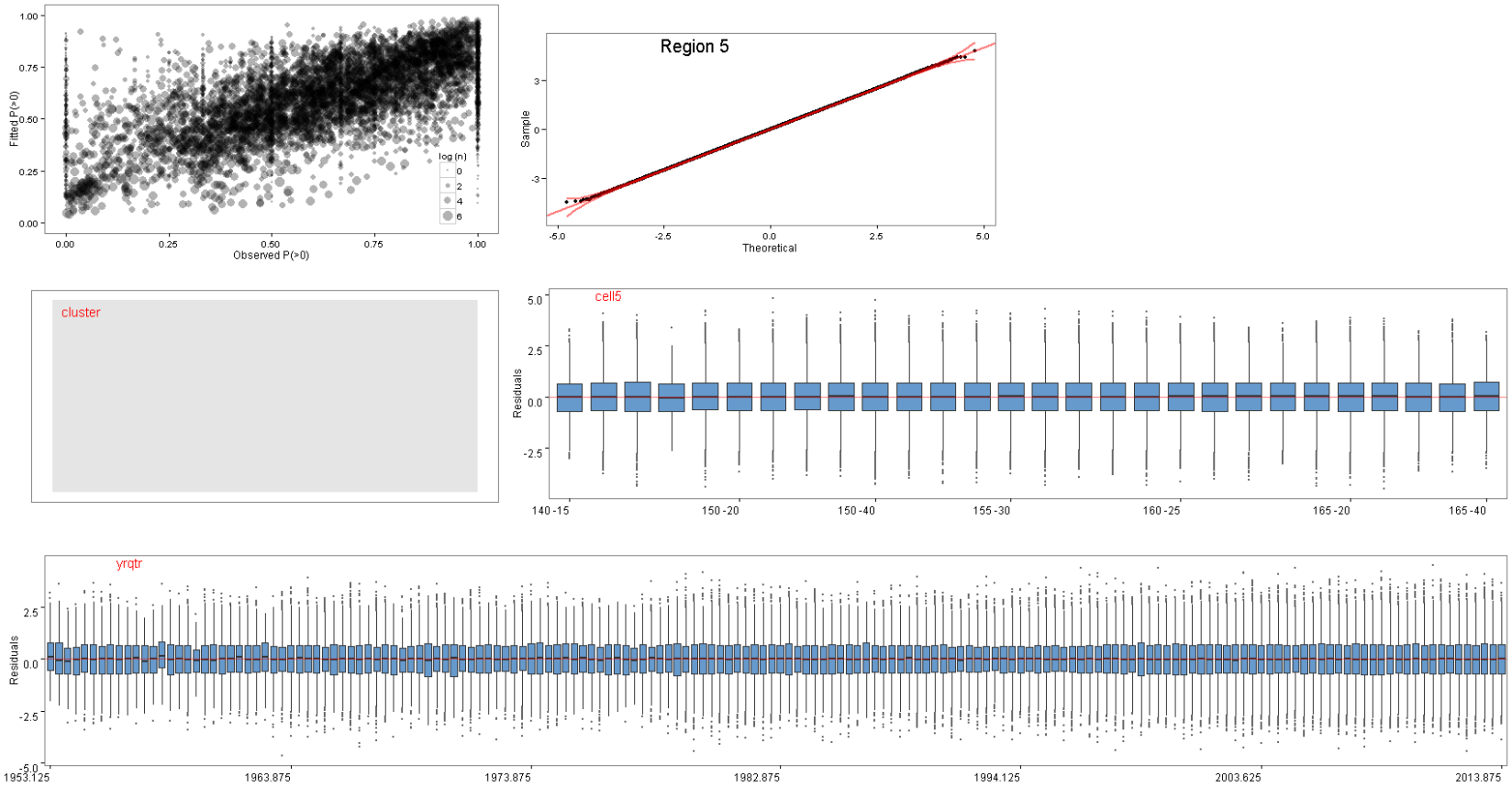


Figure 144: Diagnostic plots of fitted GLM models for region 5, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

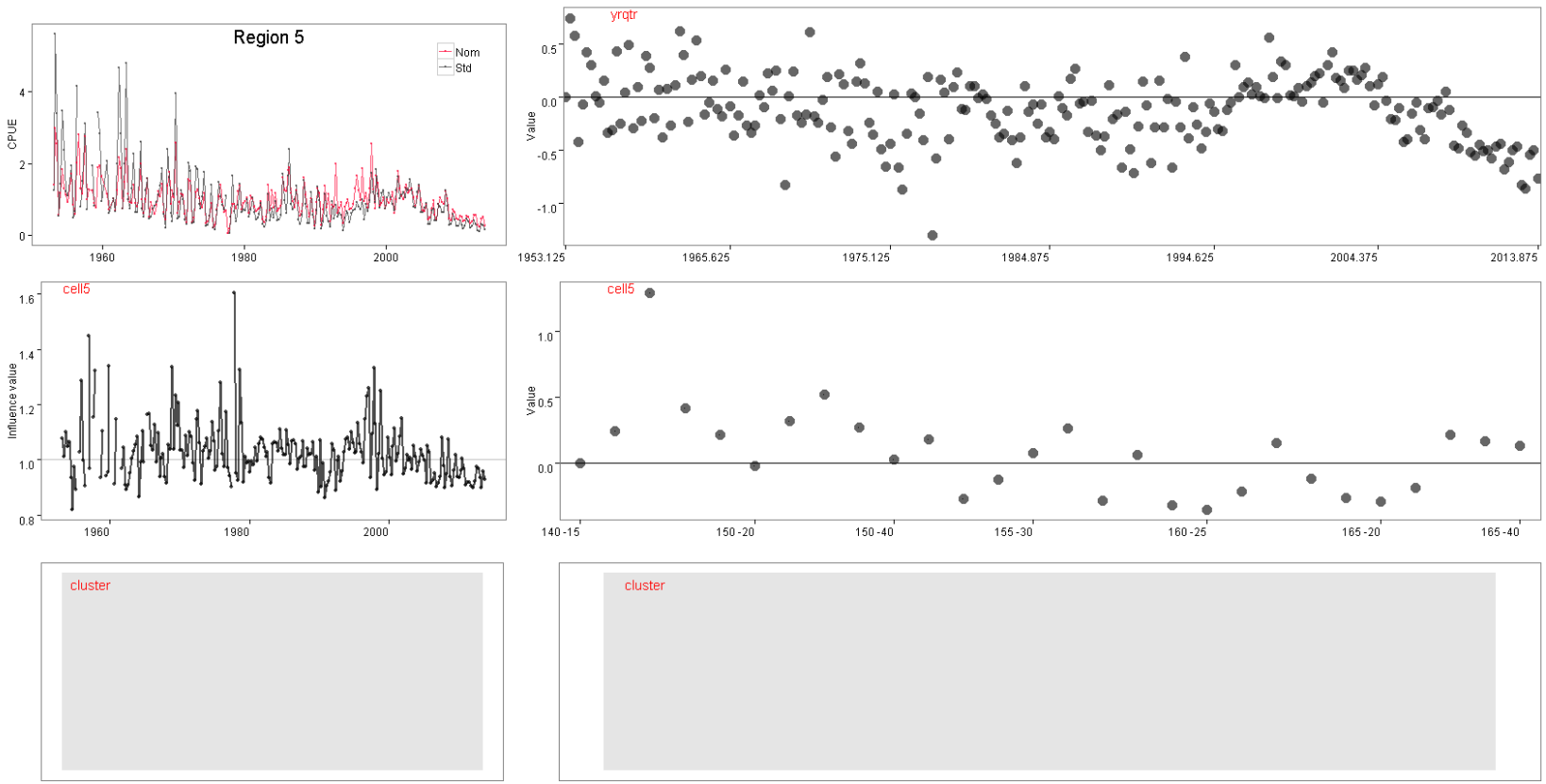




Figure 145: Summary plots of fitted GLM models for region 5, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

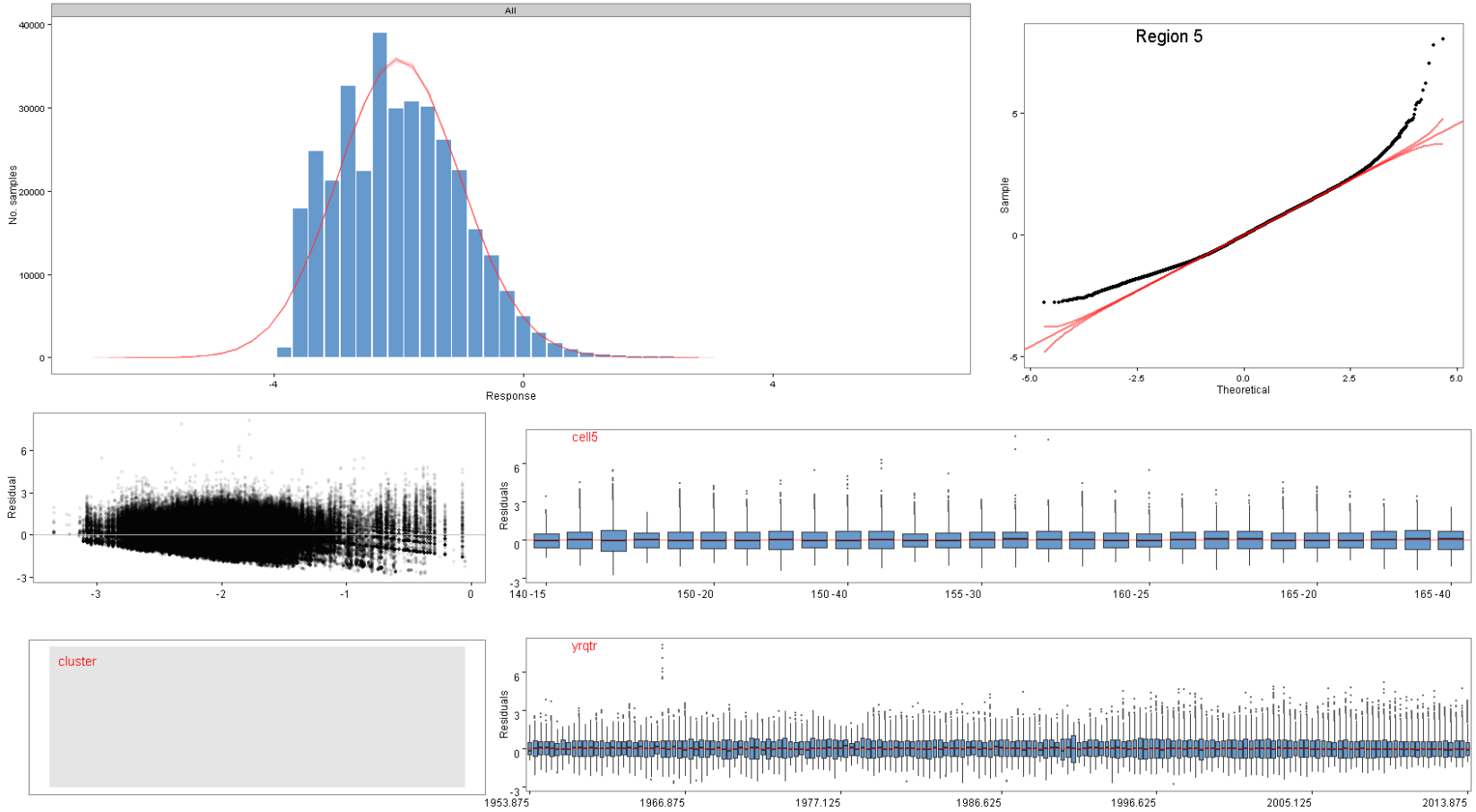


Figure 146: Diagnostic plots of fitted GLM models for region 5, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

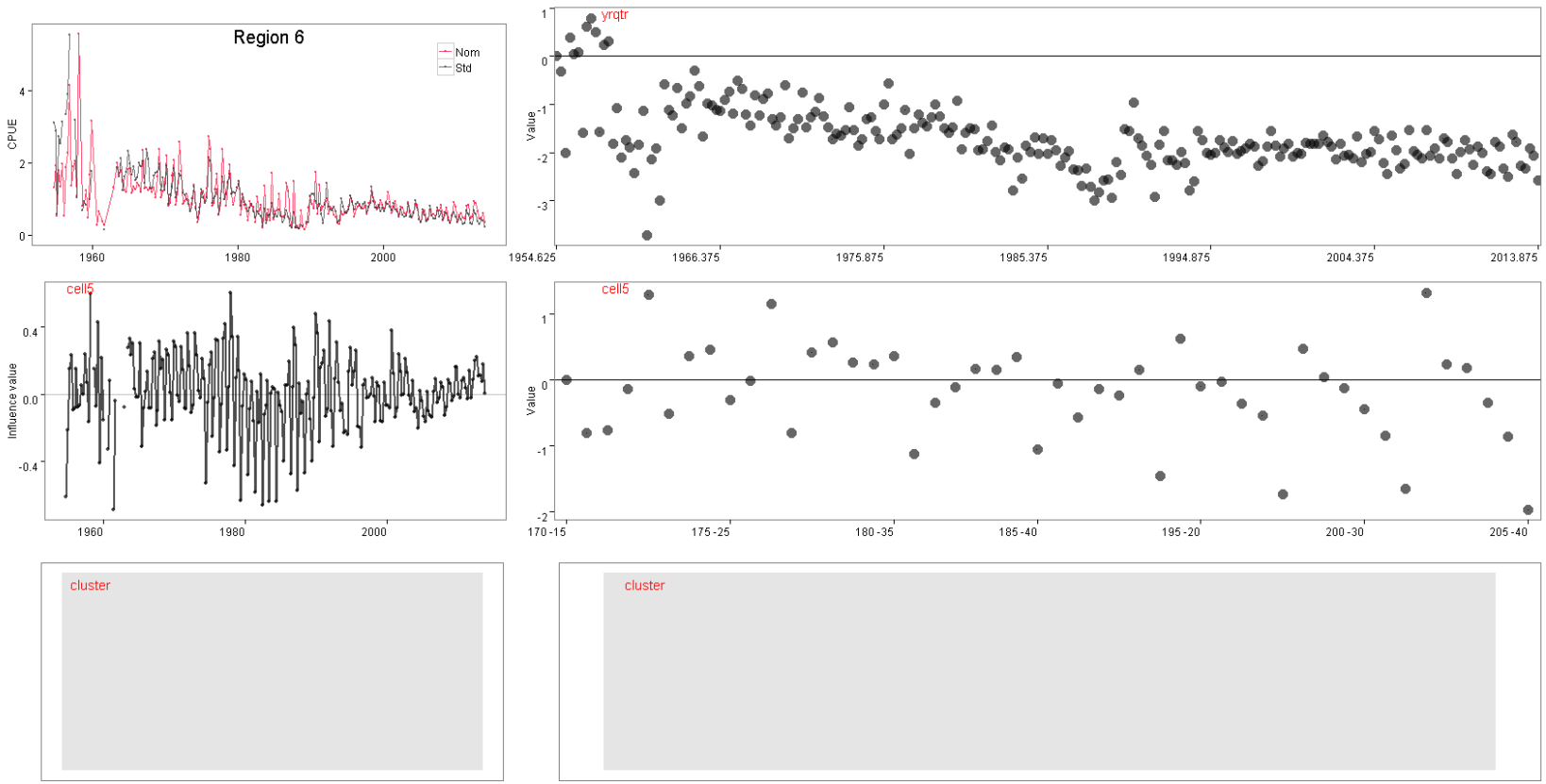


Figure 147: Summary plots of fitted GLM models for region 6, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

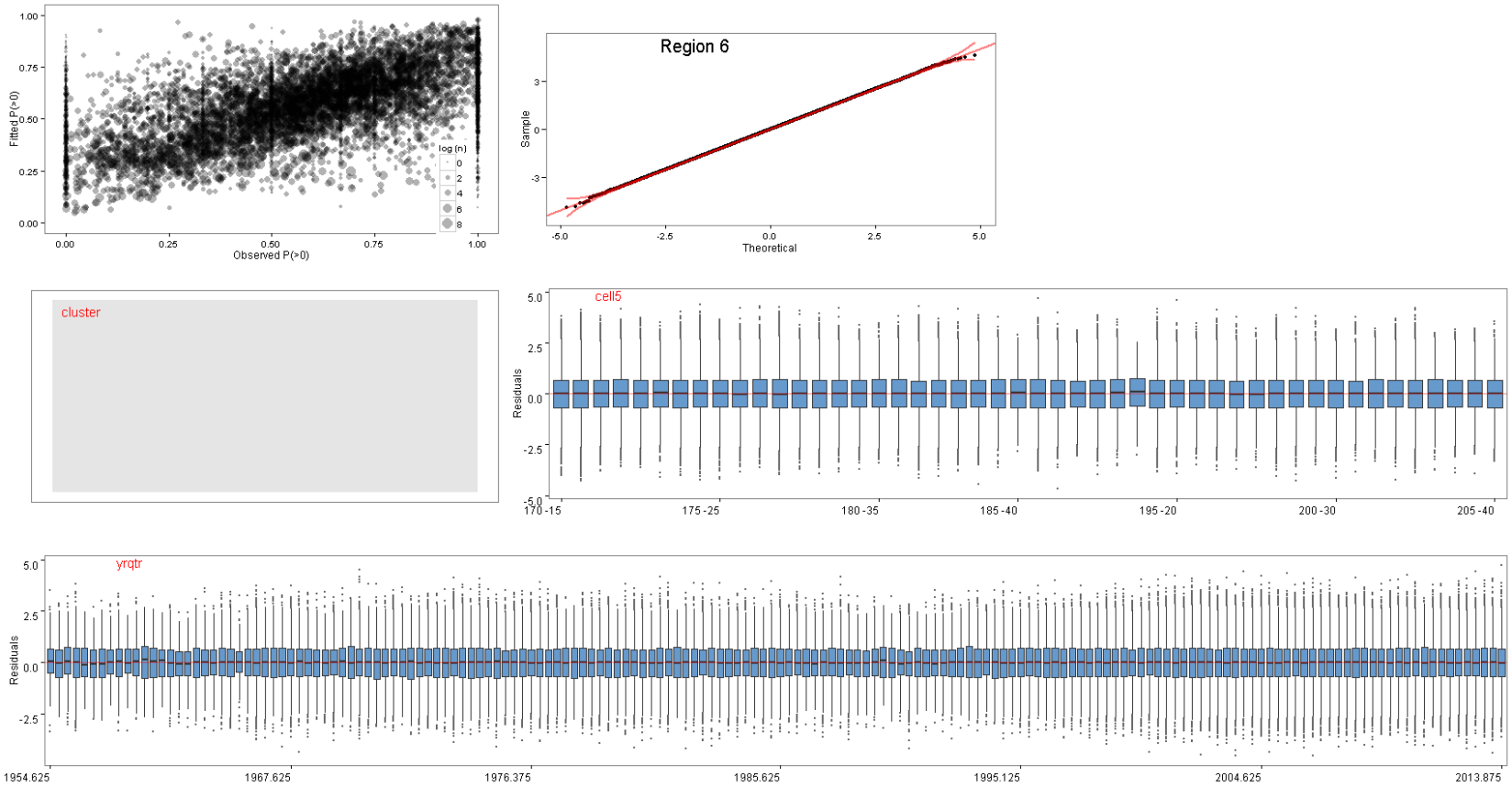




Figure 148: Diagnostic plots of fitted GLM models for region 6, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

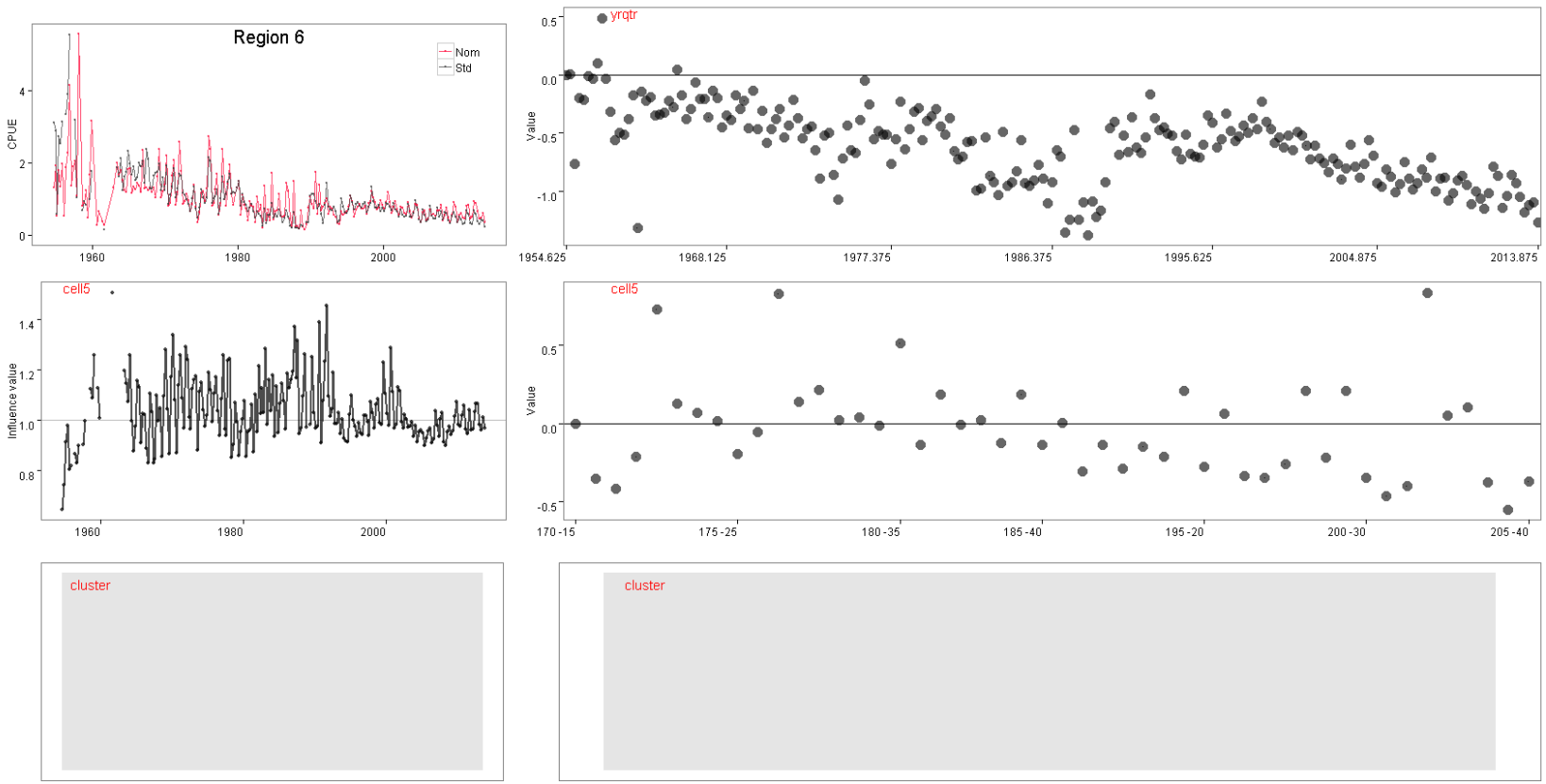


Figure 149: Summary plots of fitted GLM models for region 6, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

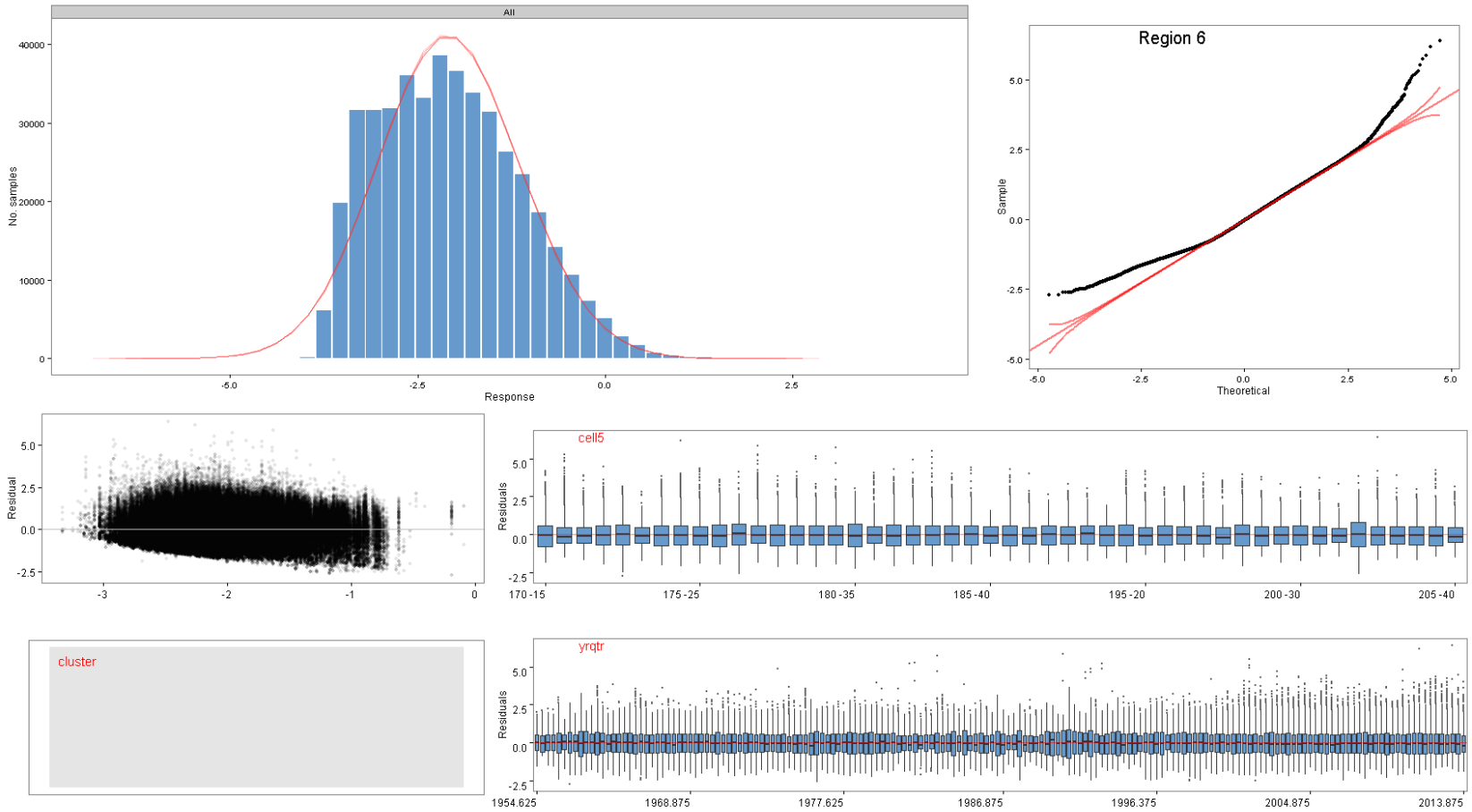


Figure 150: Diagnostic plots of fitted GLM models for region 6, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

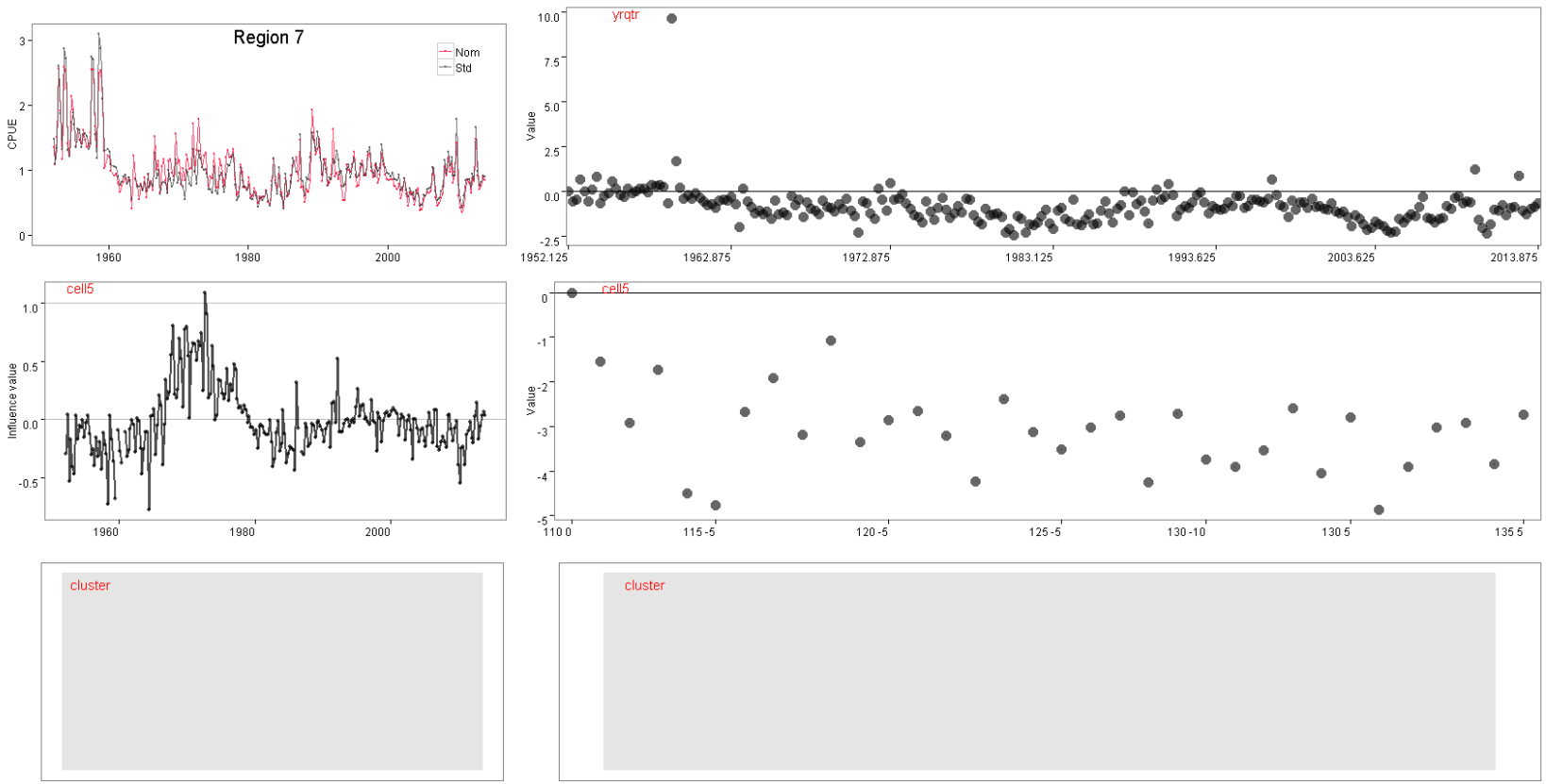


Figure 151: Summary plots of fitted GLM models for region 7, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

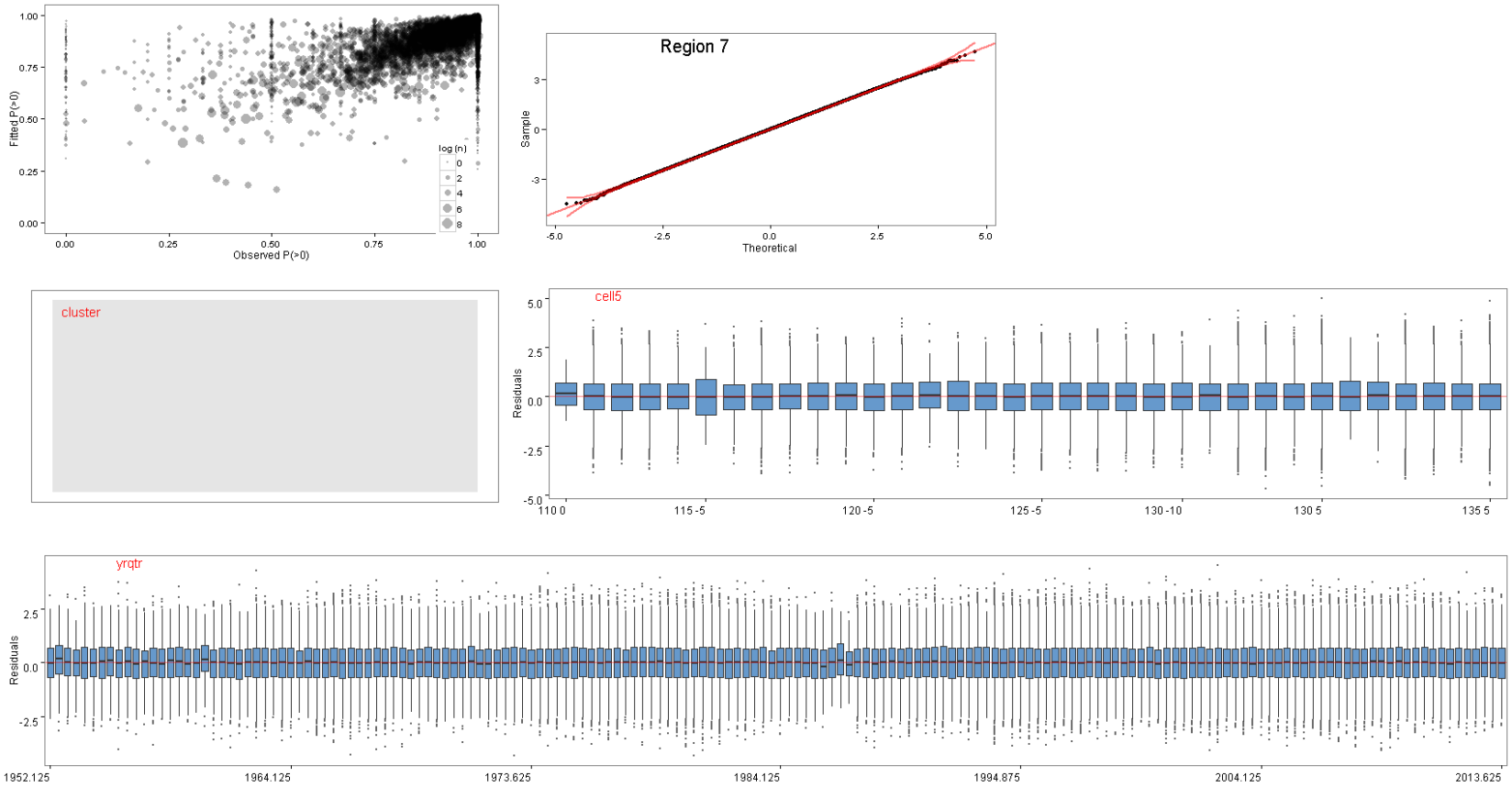


Figure 152: Diagnostic plots of fitted GLM models for region 7, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

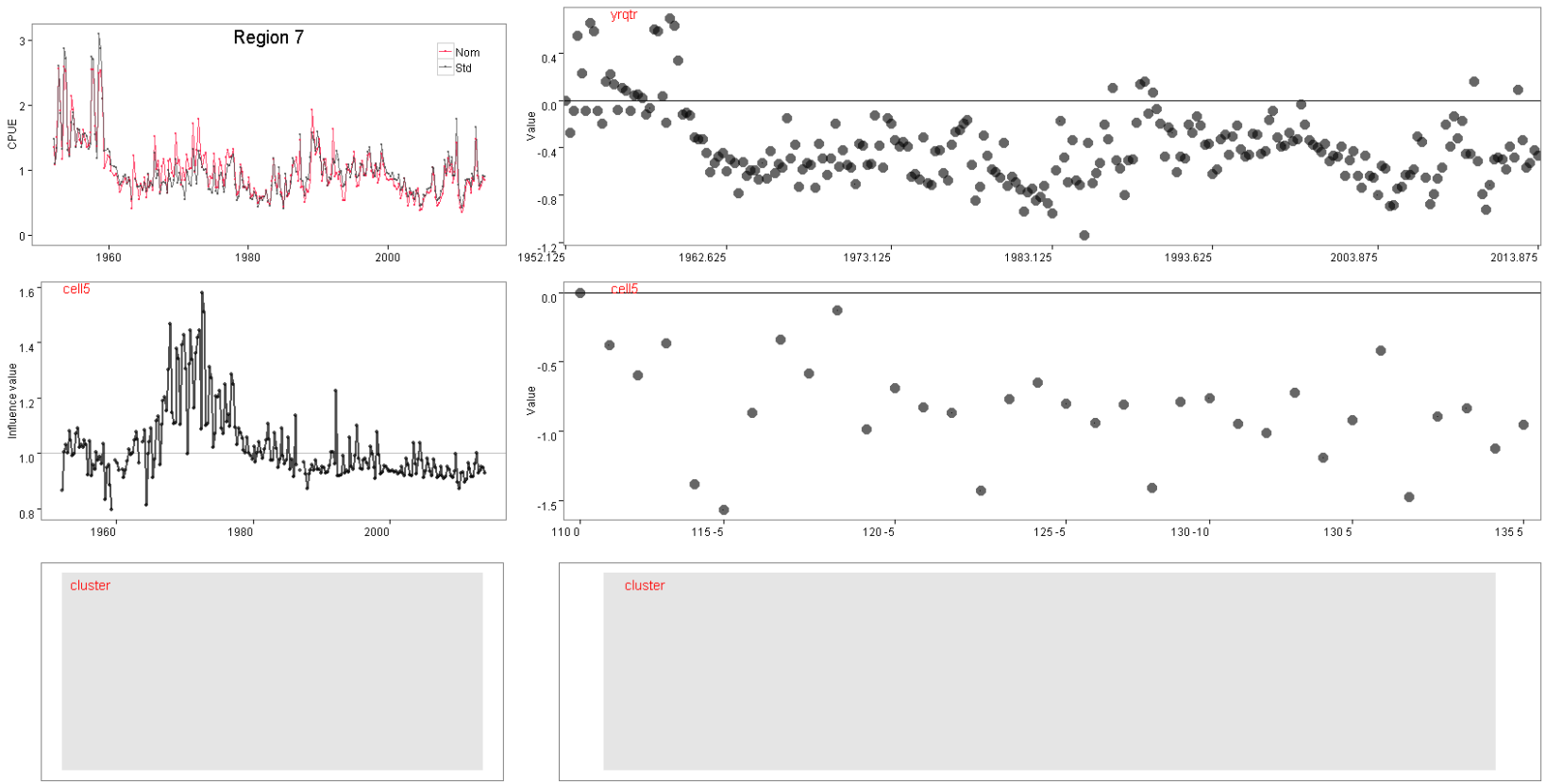


Figure 153: Summary plots of fitted GLM models for region 7, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

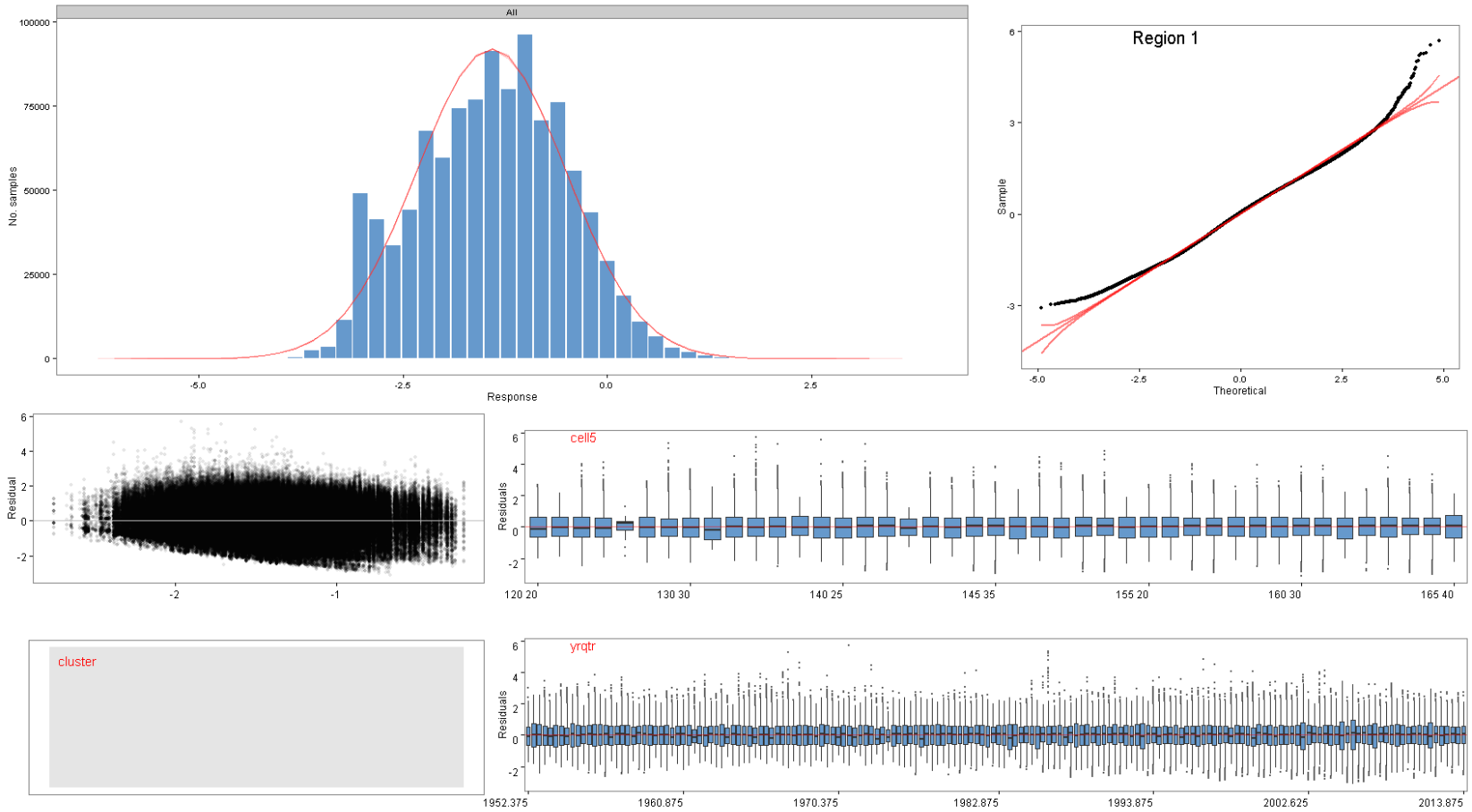


Figure 154: Diagnostic plots of fitted GLM models for region 7, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

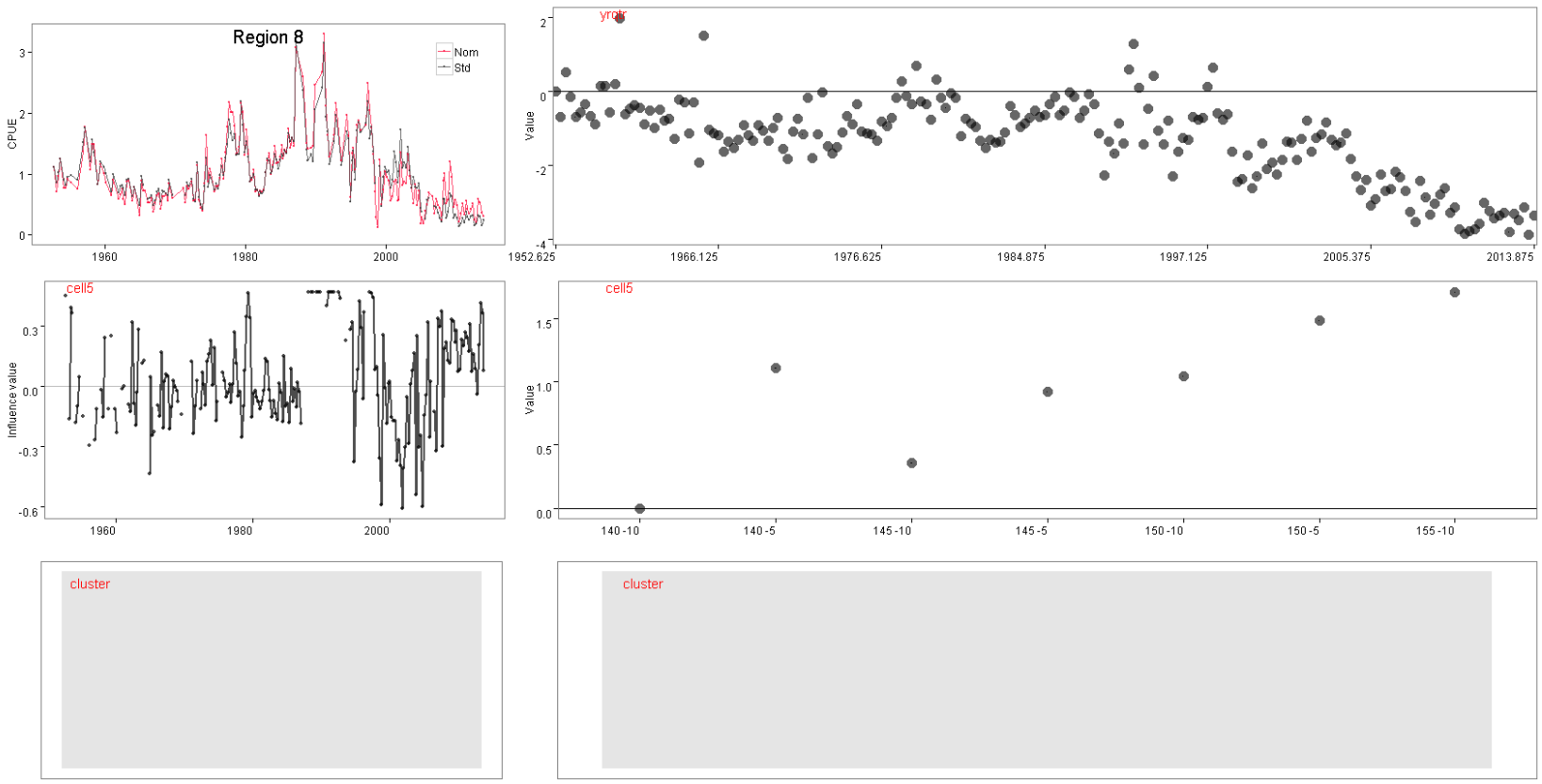


Figure 155: Summary plots of fitted GLM models for region 8, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

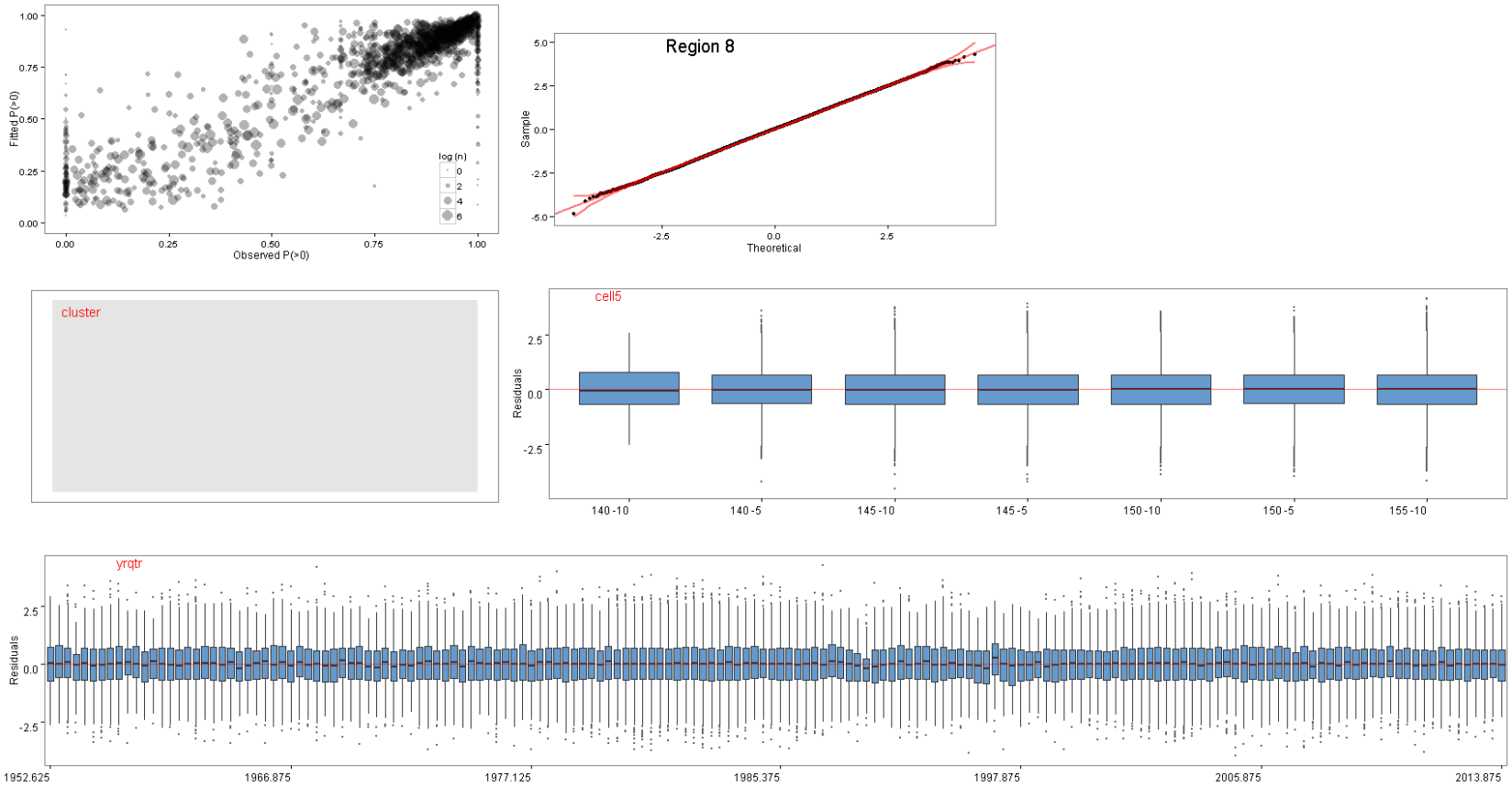


Figure 156: Diagnostic plots of fitted GLM models for region 8, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

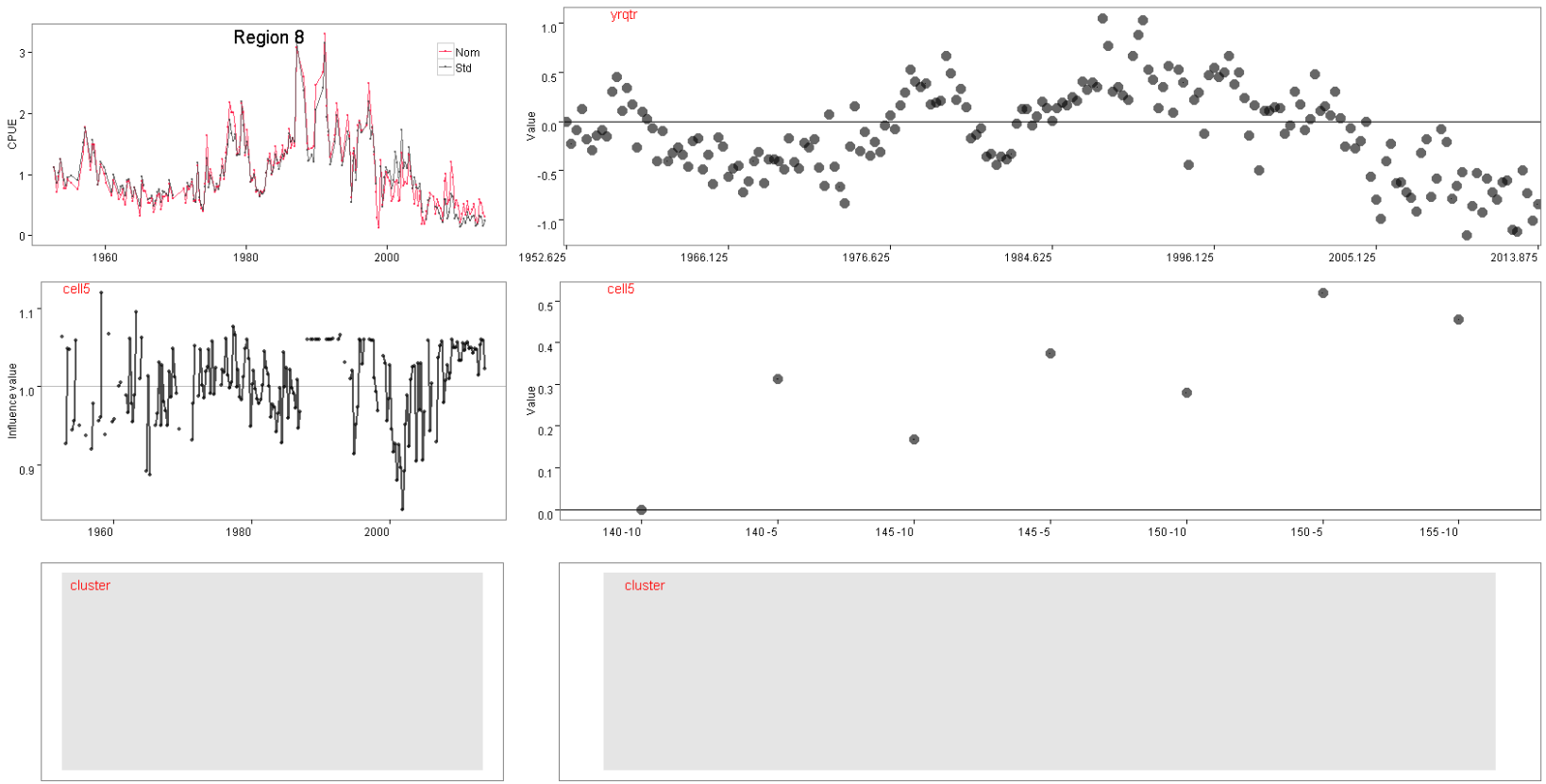


Figure 157: Summary plots of fitted GLM models for region 8, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

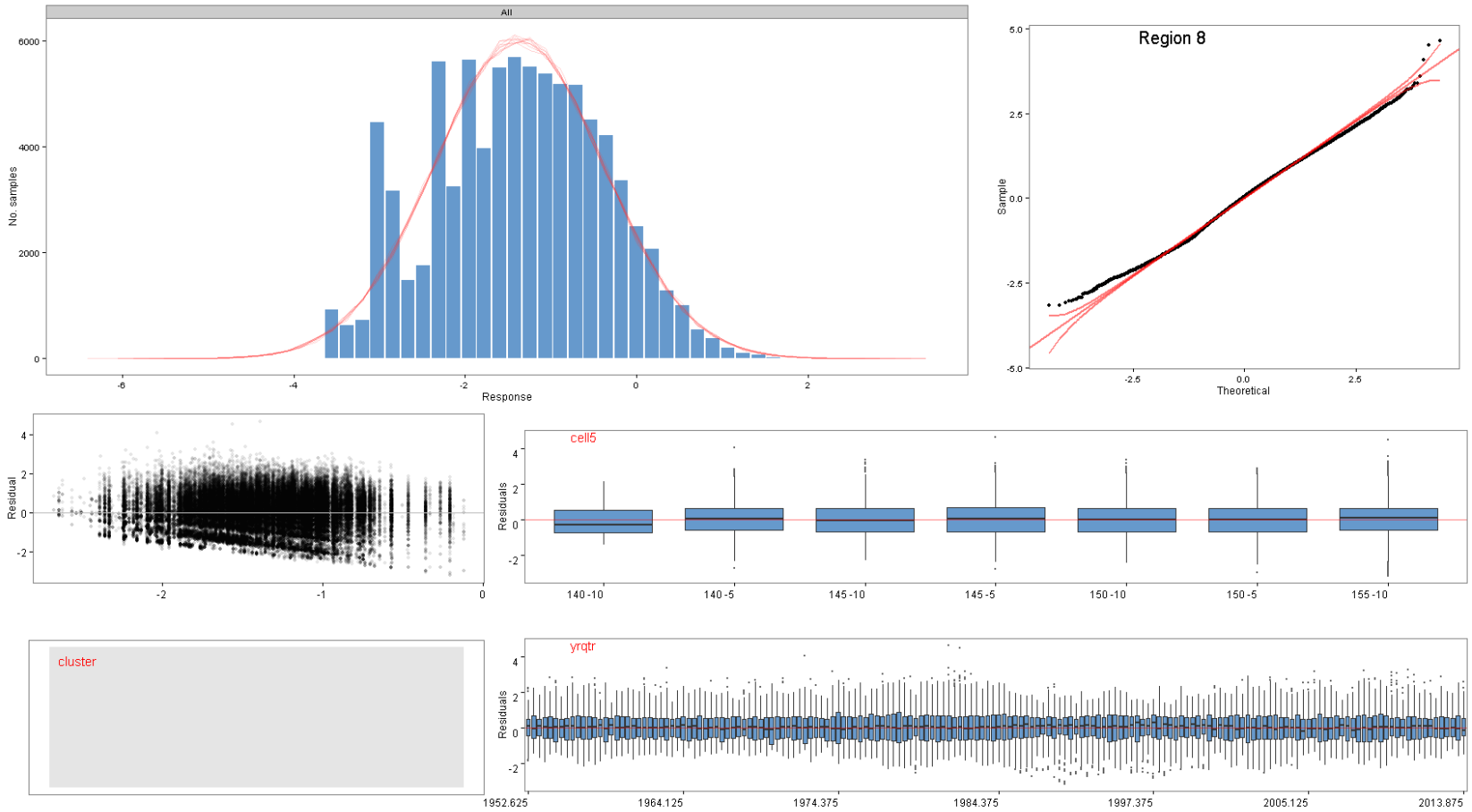


Figure 158: Diagnostic plots of fitted GLM models for region 8, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

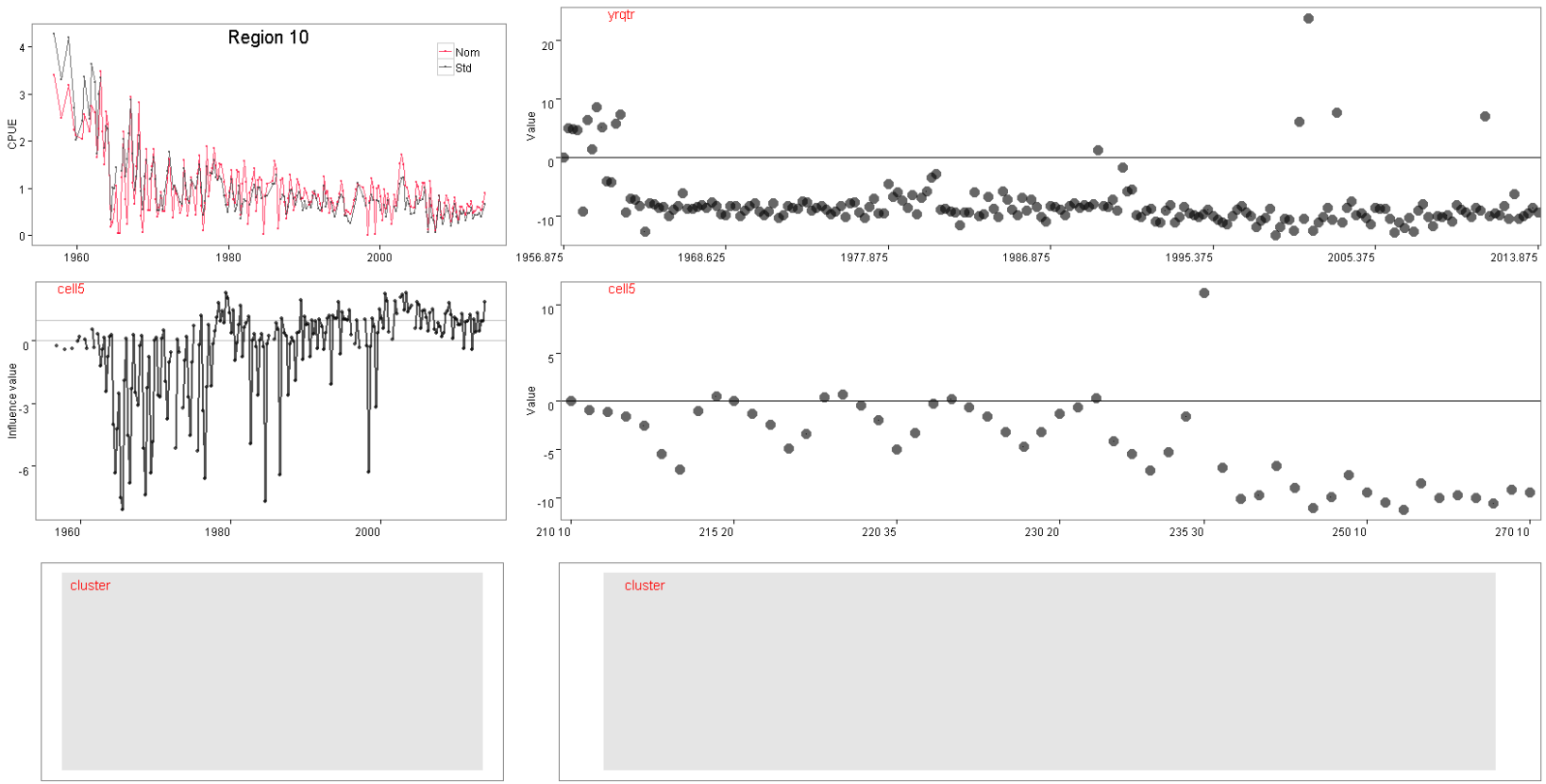


Figure 159: Summary plots of fitted GLM models for region 10, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

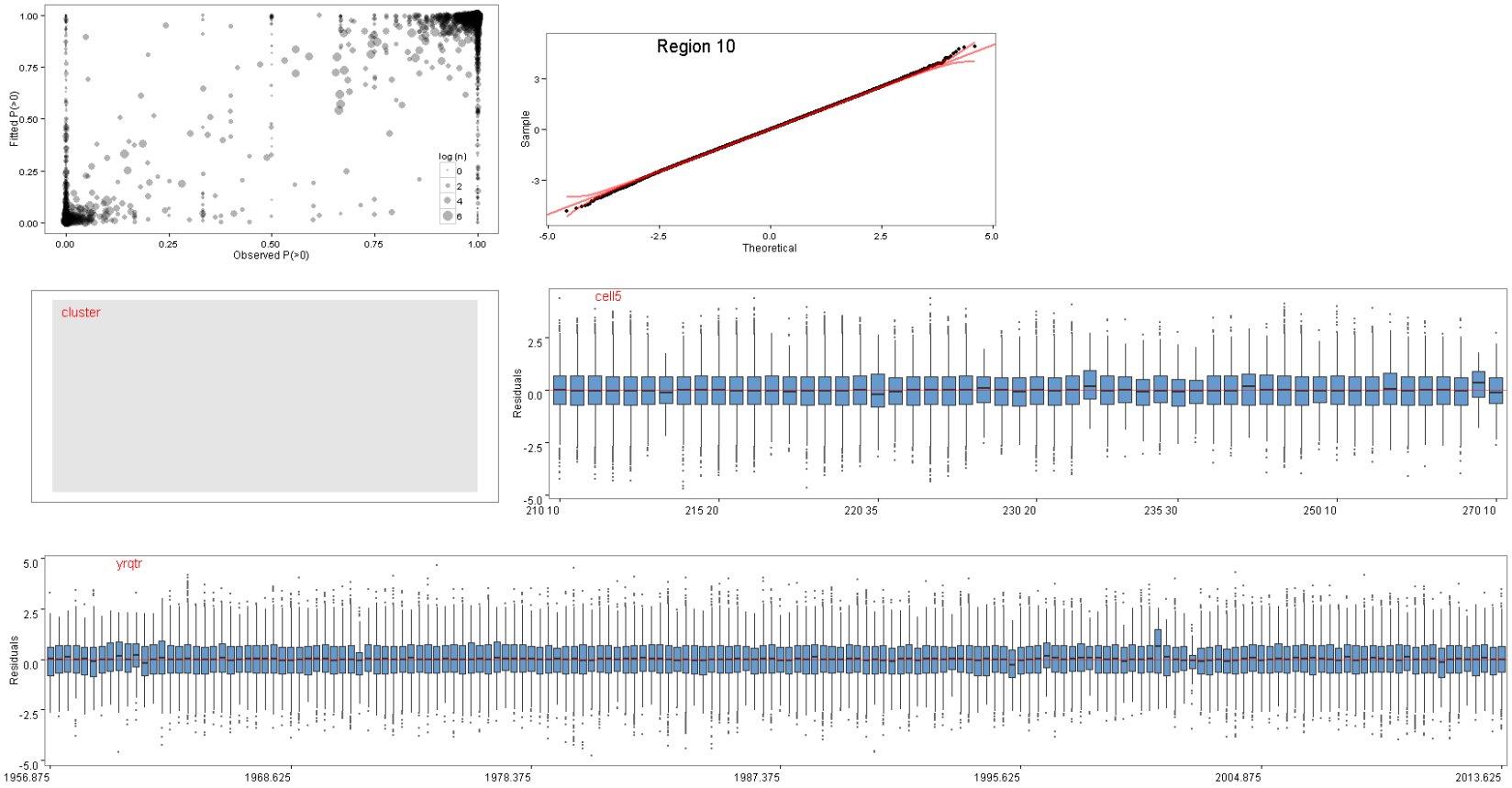


Figure 160: Diagnostic plots of fitted GLM models for region 10, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

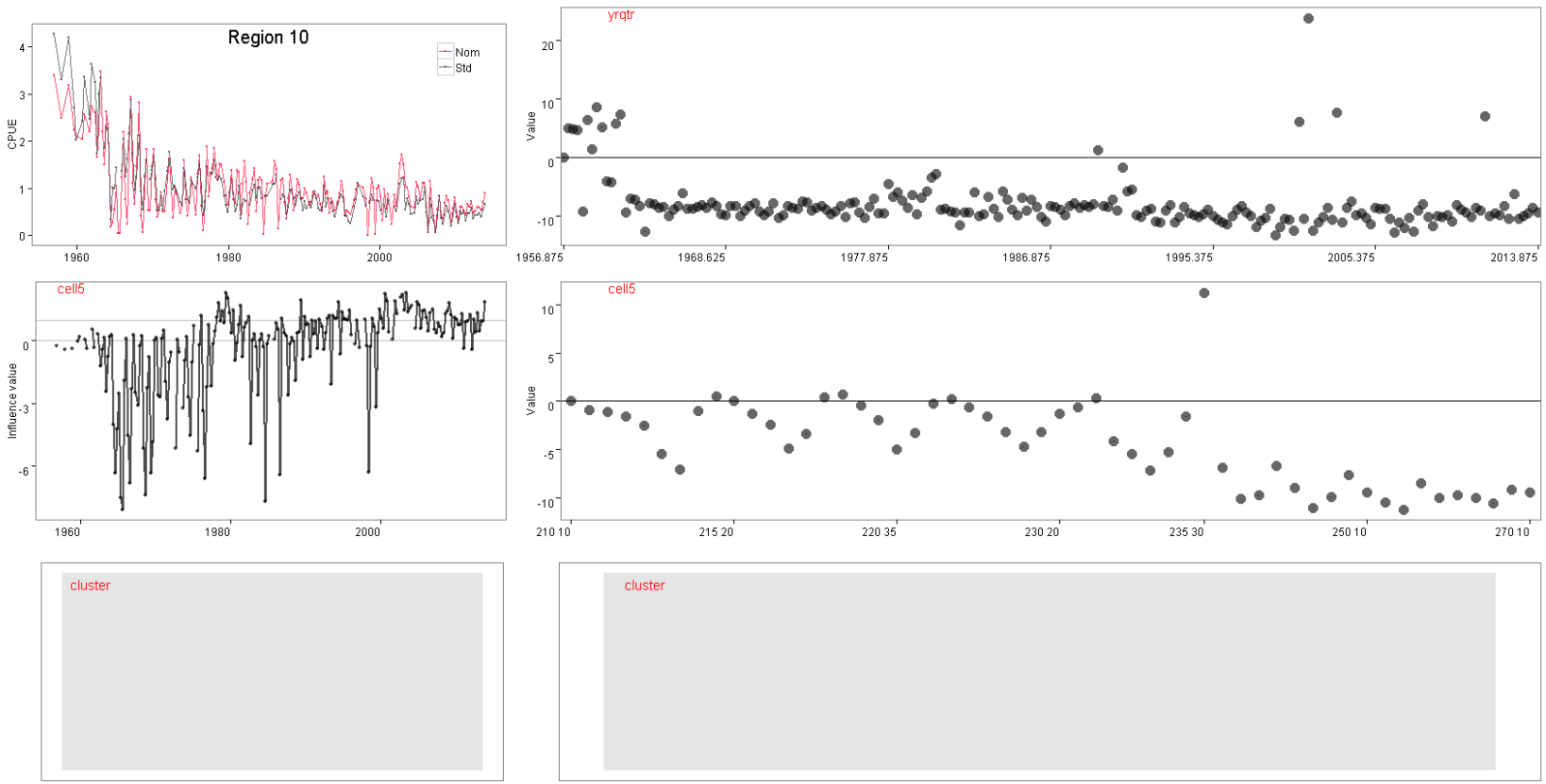




Figure 161: Summary plots of fitted GLM models for region 10, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

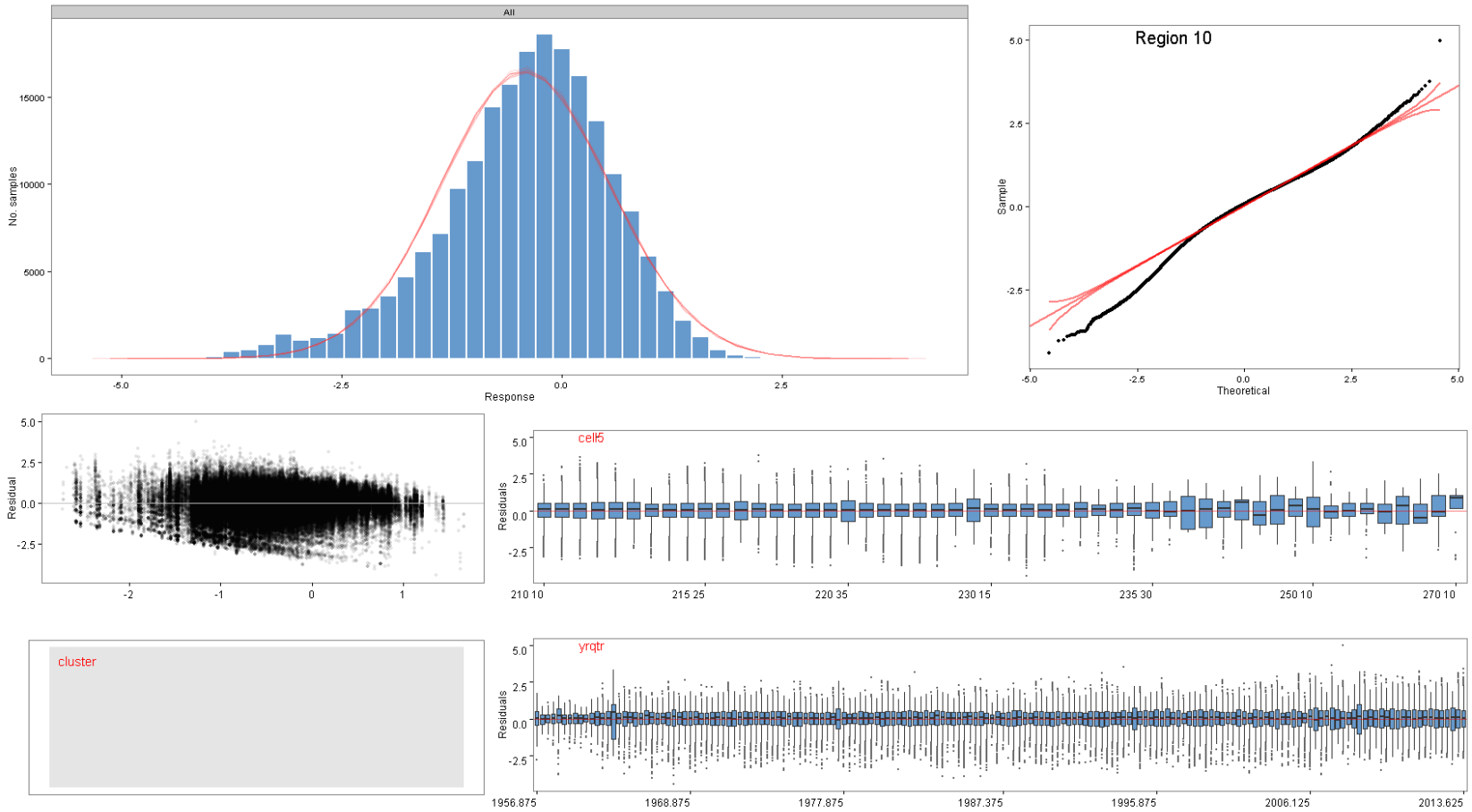


Figure 162: Diagnostic plots of fitted GLM models for region 10, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

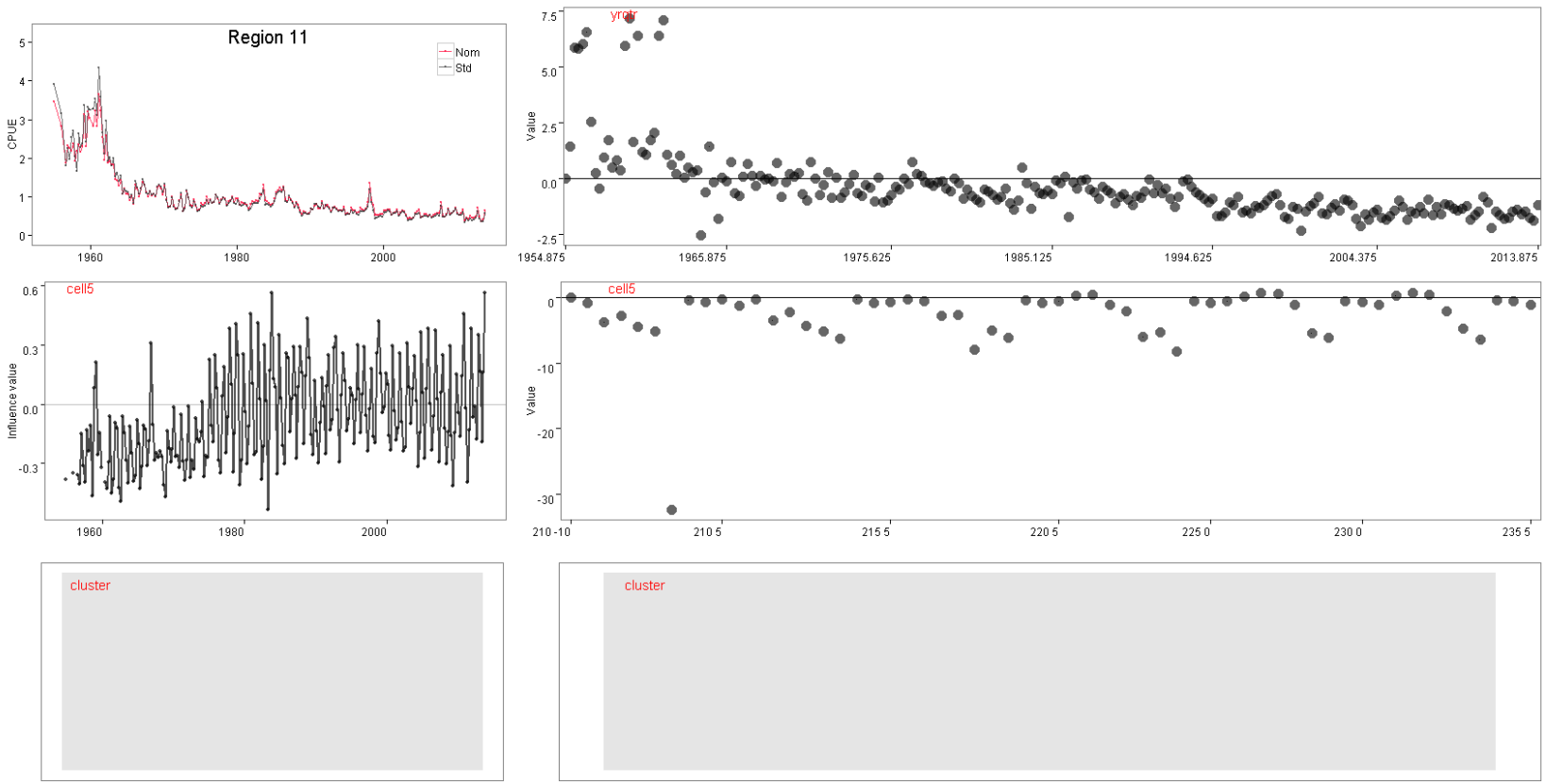


Figure 163: Summary plots of fitted GLM models for region 11, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

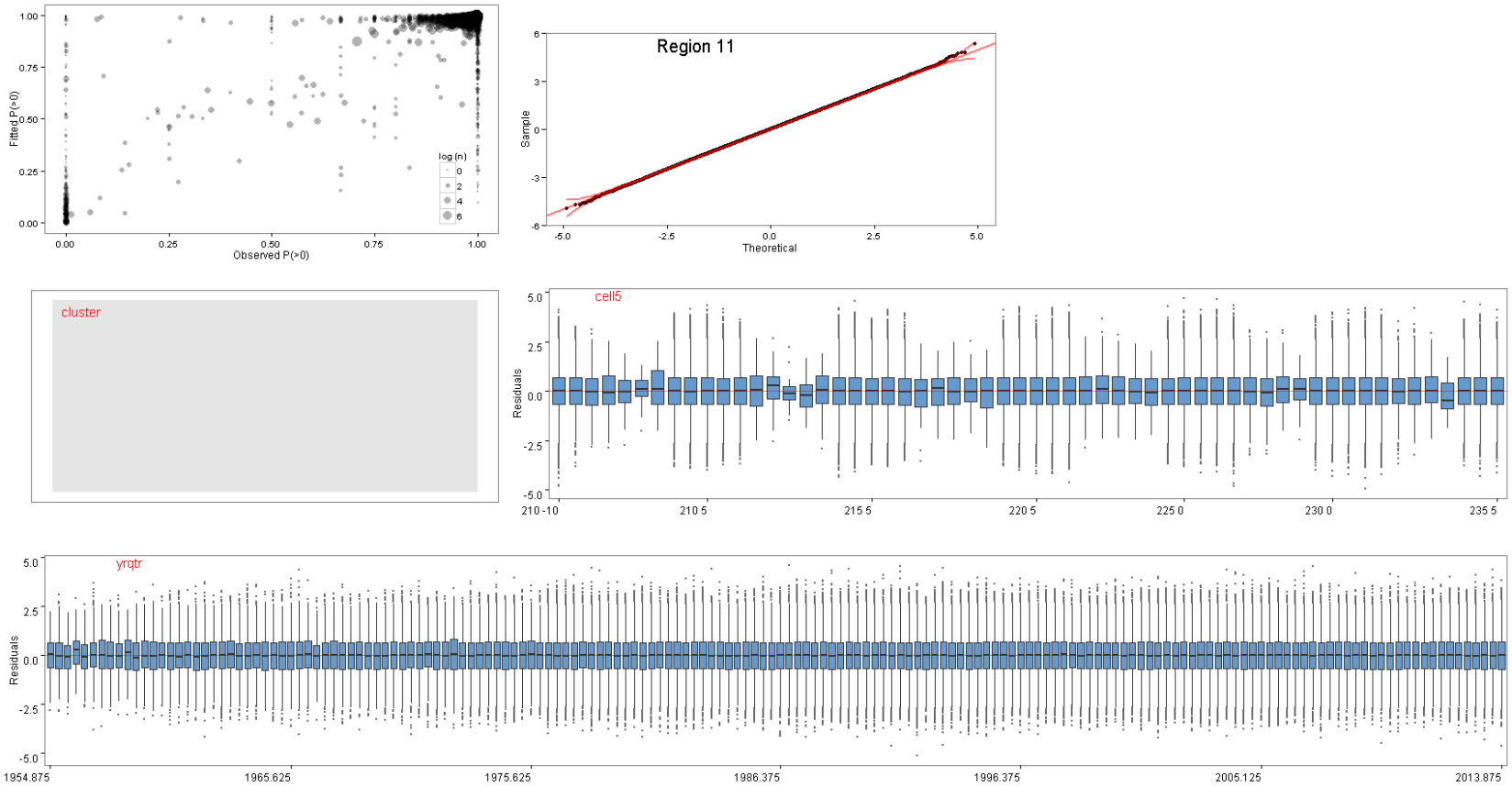




Figure 164: Diagnostic plots of fitted GLM models for region 11, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

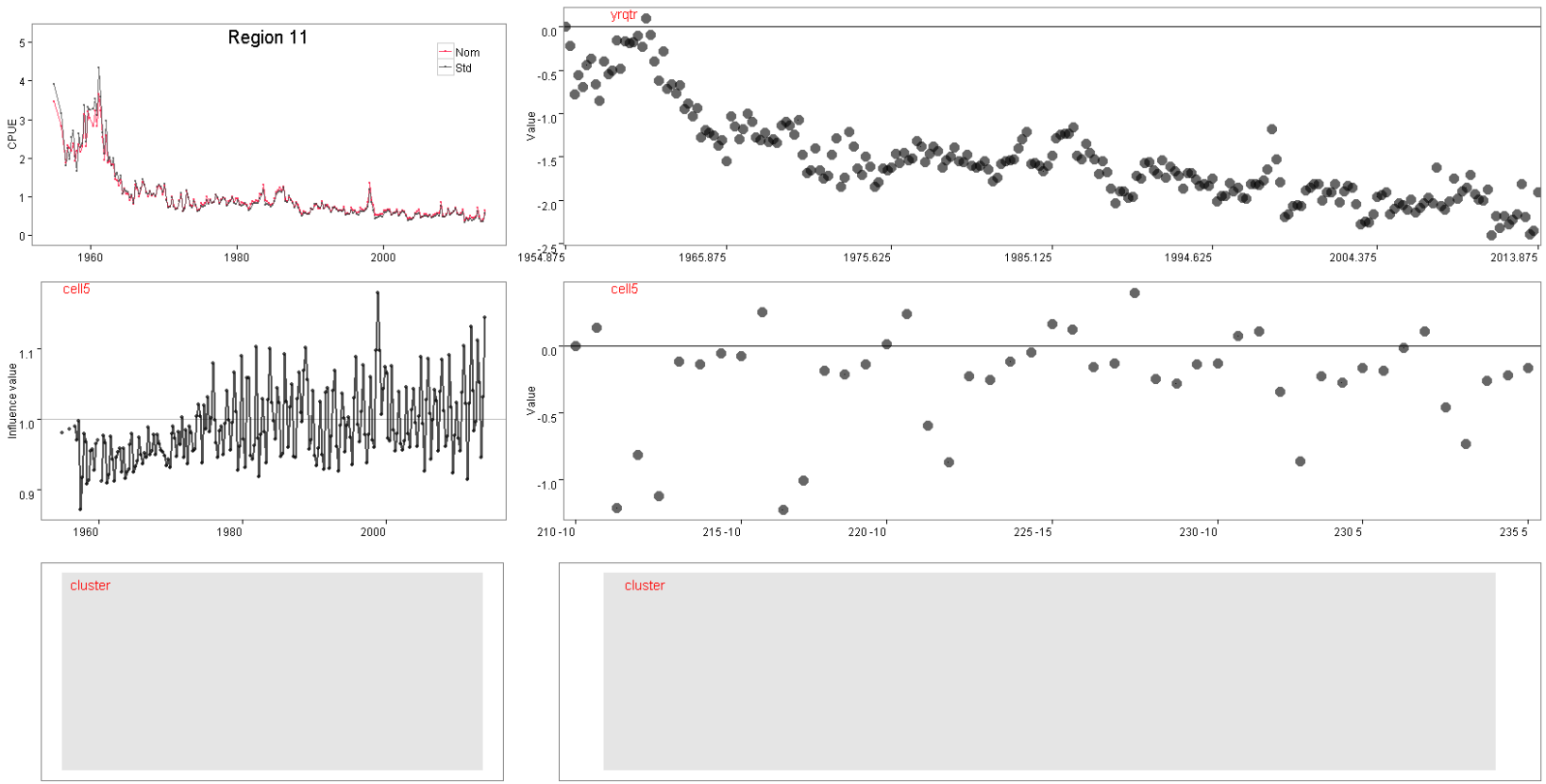


Figure 165: Summary plots of fitted GLM models for region 11, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

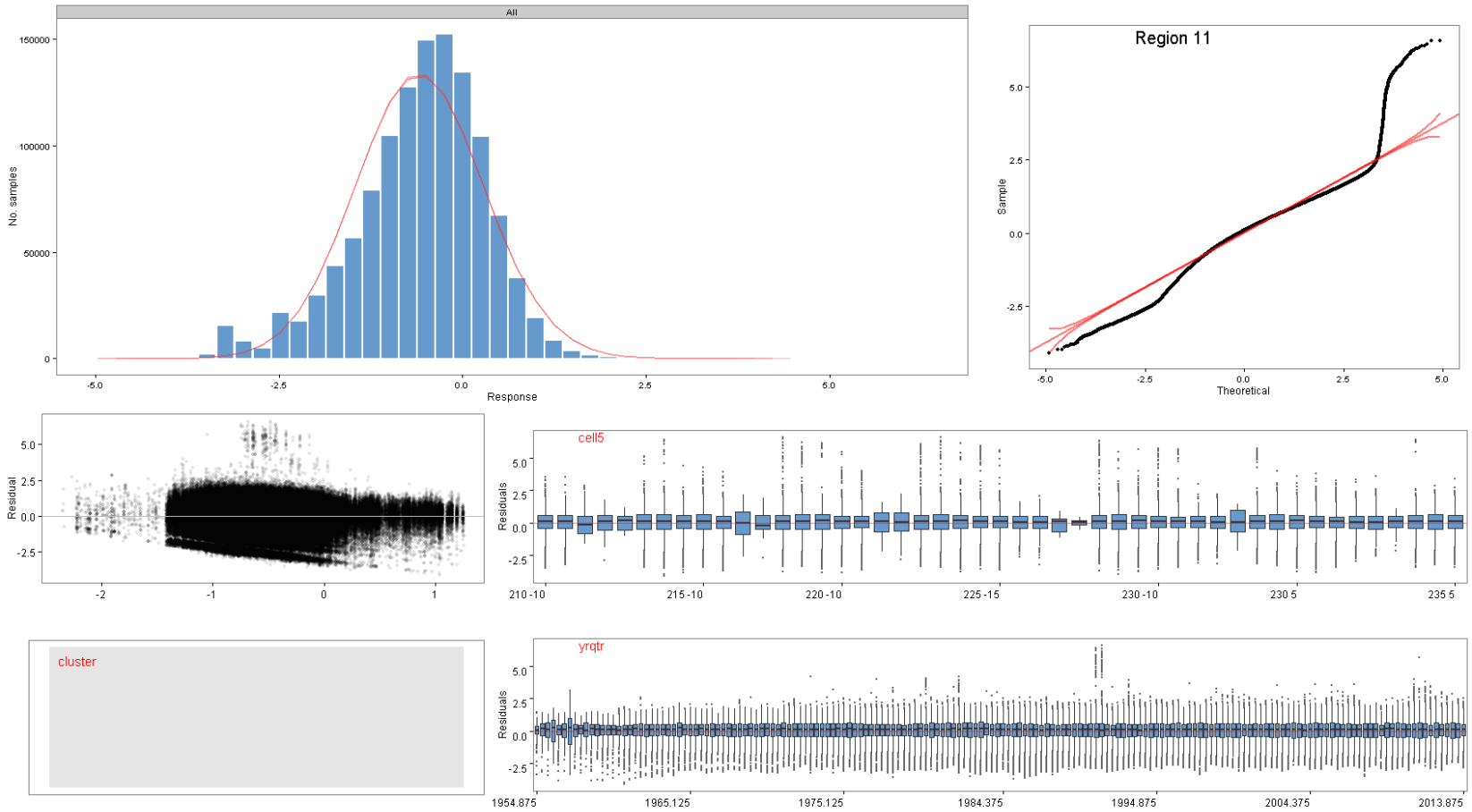


Figure 166: Diagnostic plots of fitted GLM models for region 11, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

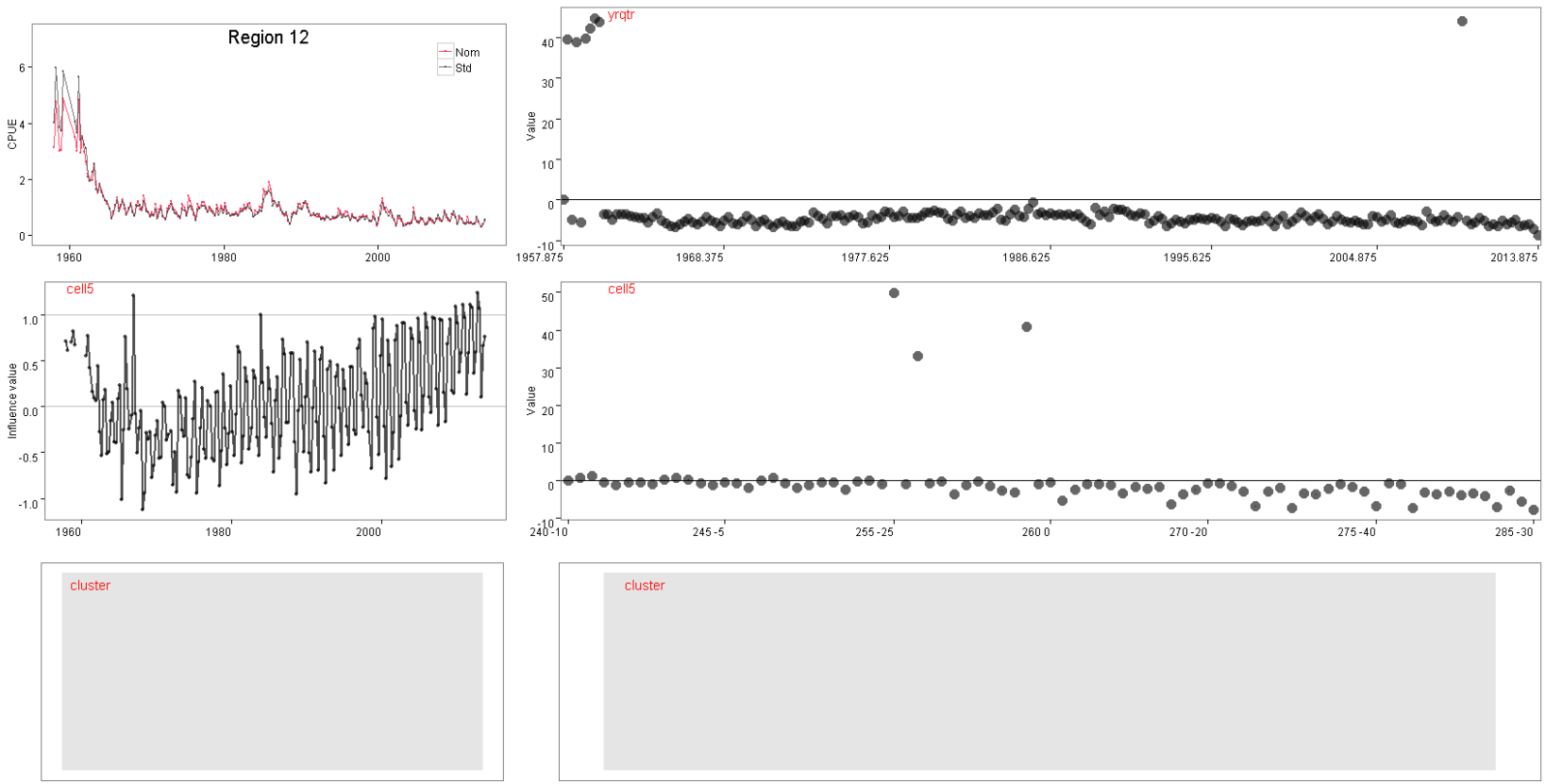


Figure 167: Summary plots of fitted GLM models for region 12, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

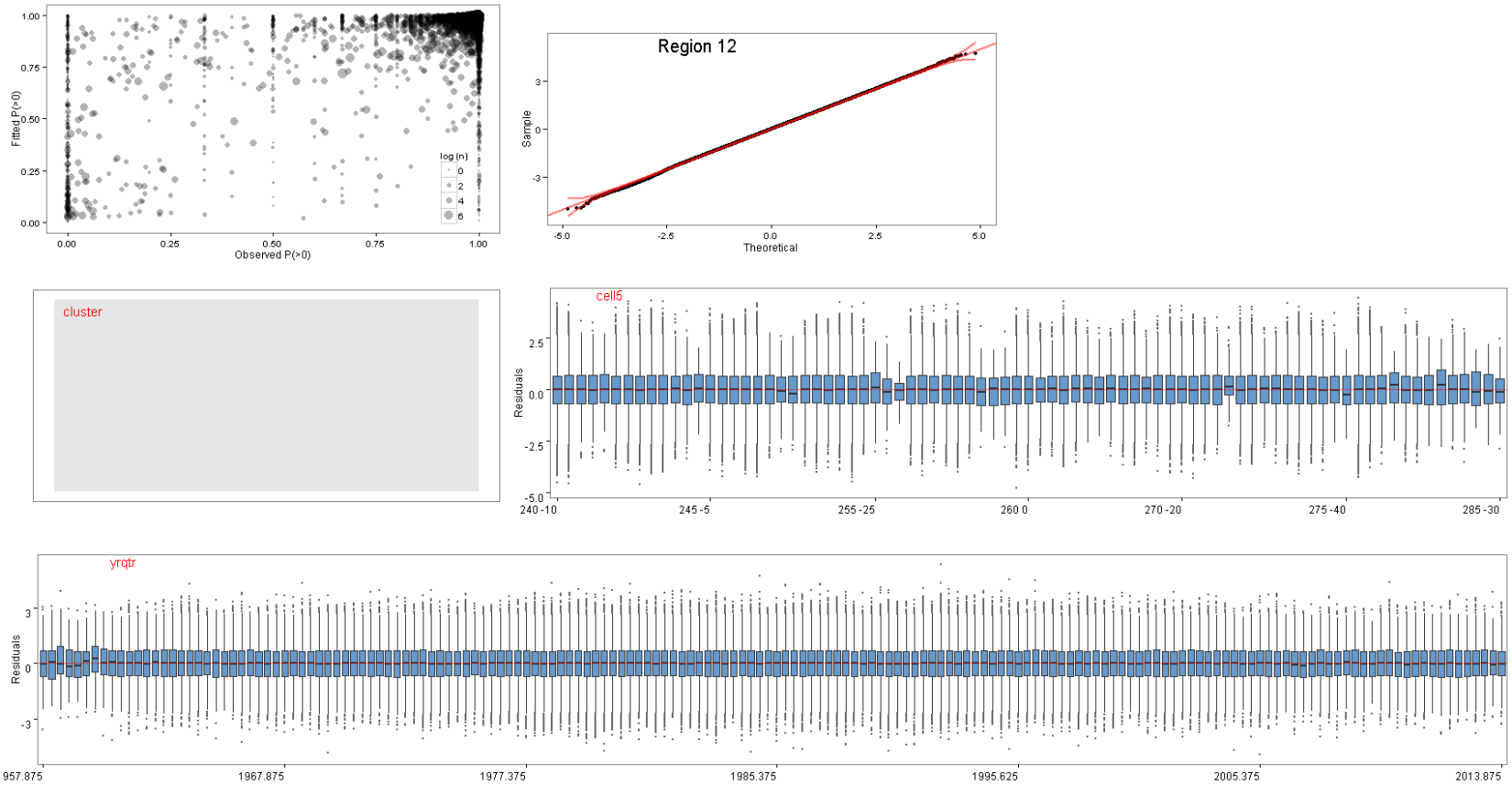


Figure 168: Diagnostic plots of fitted GLM models for region 12, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.

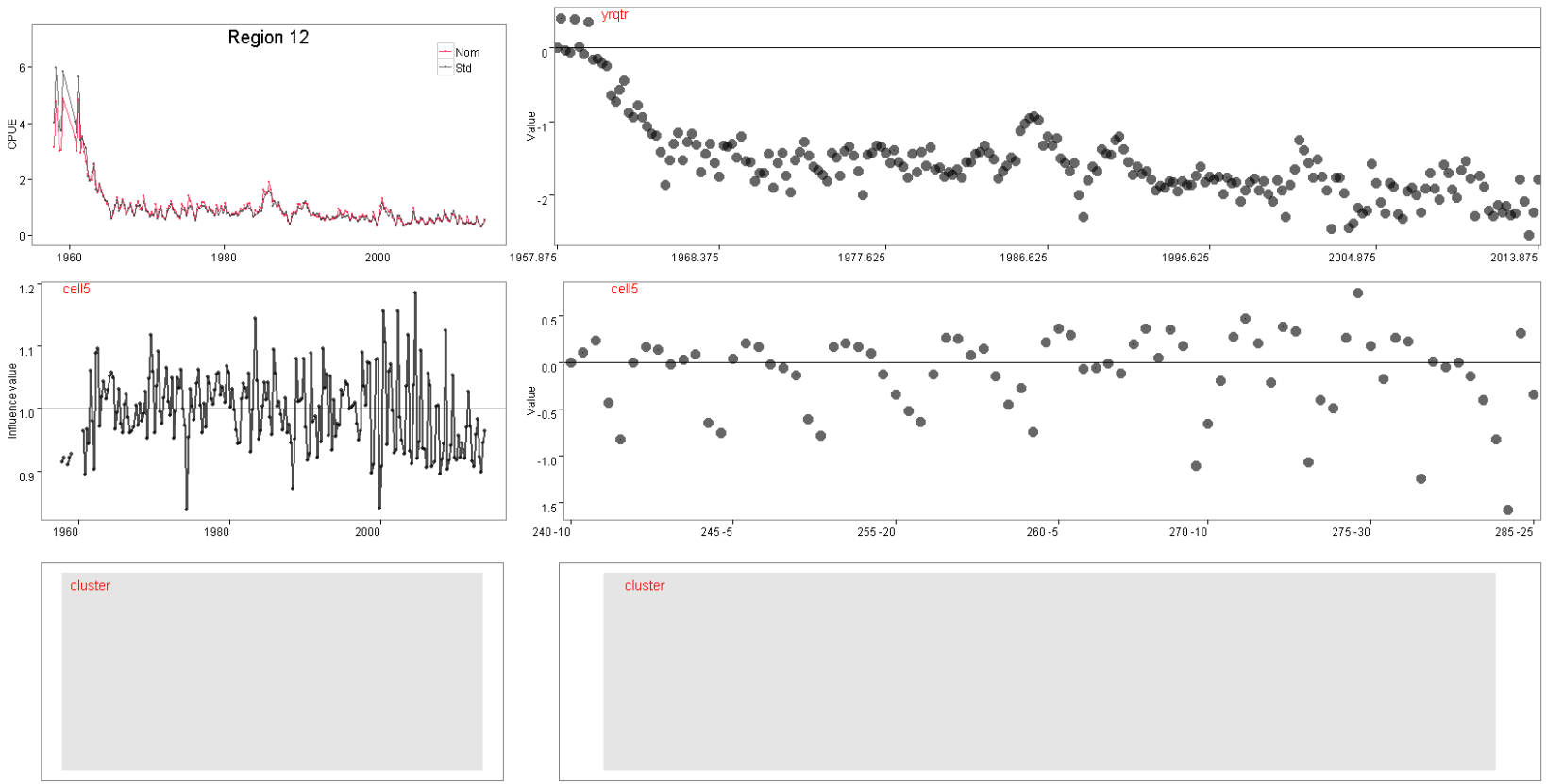


Figure 169: Summary plots of fitted GLM models for region 12, step 4; standardised versus nominal indices (top left), and influence plots (left column) and model coefficients for each explanatory variable (right column).

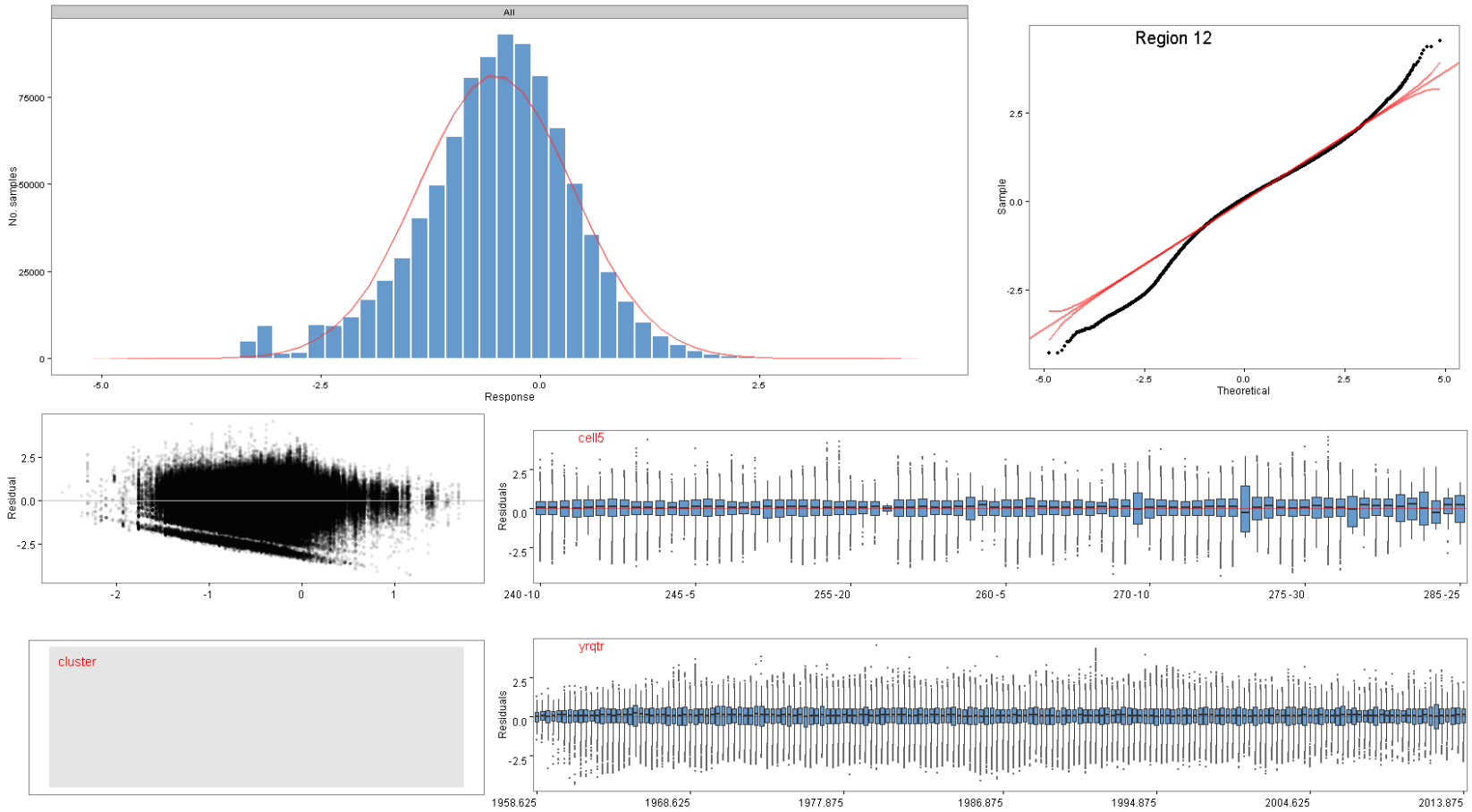


Figure 170: Diagnostic plots of fitted GLM models for region 12, step 4, showing characteristics of the model residuals and comparisons between observed and simulated data.