**L’annexe B commence à la page 149 du document de référence**

\clearpage

# NORMALISATION DES CPUE COMMERCIALES

Nous avons cherché à générer un indice de l’abondance de la morue du Pacifique à partir des données sur les captures commerciales par unité d’effort, qui ont été normalisées pour la profondeur, le lieu de pêche (régions spatiales définies; figures \@ref(fig:cpue-locality-map-5), \@ref(fig:cpue-locality-map-3)), le mois, le navire et la latitude, lorsque ces renseignements étaient disponibles. Avant d’ajuster un modèle de normalisation, nous avons dû filtrer et manipuler les données disponibles sur les captures et l’effort afin de générer un ensemble de données approprié pour l’ajustement du modèle. Dans les sections qui suivent, nous décrivons ces décisions pour les données « historiques » (1956-1995) et « modernes » (1996--2017). Nous décrivons ensuite notre modèle de normalisation des indices, explorons la contribution des diverses composantes de normalisation et déterminons l’effet de l’inclusion ou de l’omission des interactions espace-temps.

## DÉFINITION DE LA FLOTTILLE DE 1956 À 1995

Les données sur la pêche commerciale du poisson de fond au chalut avant 1991 ont été consignées au moyen d’entrevues à quai et regroupées en combinaisons du lieu de pêche et des sorties. Les données de 1991 à 1995 ont été consignées dans les journaux de bord au niveau des événements de pêche (chalut). Nous avons donc regroupé les données de 1991-1995 au niveau lieu-sortie (ci-après les « sorties » dans la présente annexe) pour les faire correspondre à la résolution des données antérieures. Pour regrouper ces données de 1991-1995, nous avons retiré tous les cas de chalutage de plus de cinq heures, puisqu’il s’agit probablement d’erreurs de saisie de données. Pendant ces périodes, les variables « profondeur », « date » et « lieu » sont disponibles comme covariables du modèle.

## DÉFINITION DE LA FLOTTILLE DE 1996 À 2017

Les données de la pêche commerciale du poisson de fond au chalut de fond de 1996 jusqu’à aujourd’hui ont été enregistrées au niveau des événements de pêche en présence d’observateurs à bord ou par surveillance vidéo. Nous avons traité cet ensemble de données moderne séparément de celui des données historiques pour 1) tirer parti de la résolution des données plus élevée, 2) inclure l’information sur la latitude et l’ID du navire dans notre modèle de normalisation et 3) éviter de supposer une capturabilité constante et une relation entre la CPUE et les covariables de la normalisation entre les principaux changements réglementaires.

Étant donné que nous disposons de données sur les différents navires de cette flottille moderne, et conformément aux analyses antérieures des stocks de poisson de fond du Pacifique, nous avons défini pour l’ensemble des données modernes une « flottille » qui ne comprend que les navires jugés admissibles parce qu’ils répondent à certains critères de captures régulières de morues du Pacifique. Nous suivons l’approche utilisée dans un certain nombre d’évaluations récentes des stocks de poisson de fond de la Colombie-Britannique en exigeant que les navires aient capturé l’espèce dans au moins 100 traits toutes les années d’intérêt et qu’ils aient dépassé un seuil de cinq sorties (sorties ayant enregistré certains individus de l’espèce) pendant au moins cinq ans, entre 1996 et 2017.

## DÉFINITION DES PRÉDICTEURS DU MODÈLE DE NORMALISATION

Pour la profondeur et la latitude, nous avons regroupé les valeurs dans une séquence de tranches pour établir des relations non linéaires entre ces prédicteurs et les CPUE [p. ex. @maunder2004a]. Pour la profondeur, nous avons compartimenté la profondeur du chalut en tranches de 25 m. Pour la latitude, nous avons utilisé des tranches de 0,1 degré. Afin d’obtenir suffisamment de données pour estimer un coefficient pour chaque niveau de facteur, nous avons limité la fourchette des tranches de profondeur à celles qui se situaient dans la probabilité cumulative d’observations positives de 0,1 % à 99,9 %, puis nous avons supprimé tous les niveaux de facteur (pour tous les prédicteurs) contenant moins de 0,1 % de ces observations positives.

Les prédicteurs qui sont traités comme des facteurs dans un modèle statistique ont besoin d’un niveau de référence ou de base — un niveau à partir duquel les autres coefficients de cette variable estiment une différence. Ce niveau de base devient alors la valeur du prédicteur utilisée dans la prévision pour l’indice normalisé. Nous avons choisi le niveau de facteur le plus fréquent comme niveau de base — un choix courant pour ces types de modèles [@maunder2004a]. Par exemple, nous avons établi le mois de base comme étant le mois le plus couramment observé dans l’ensemble de données filtré pour seulement les traits dans lesquels l’espèce a été capturée. Ce choix du niveau de base ne touche que l’interception ou l’ampleur relative de notre indice en raison de la forme de notre modèle (voir ci-après). Cette ampleur relative ne devrait pas avoir d’incidence sur les résultats du modèle d’évaluation des stocks parce que nous avons estimé des capturabilités distinctes pour chaque indice des CPUE commerciales.

## UN MODÈLE DE NORMALISATION DE L’INDICE DU MLGM DE TWEEDIE

Les données sur les CPUE des pêches contiennent à la fois des valeurs nulles et des valeurs continues positives. Diverses approches ont été utilisées dans la documentation sur les pêches pour modéliser ces données. Une approche consistait à ajuster un modèle linéaire généralisé (MLG) de type delta, un modèle qui ajuste les valeurs nulles par rapport aux valeurs non nulles selon une régression logistique (un MLG binomial et un lien logit) et les valeurs positives selon une régression linéaire adaptée aux données log-transformées ou un MLG de type gamma avec un lien logarithmique [p. ex. @maunder2004a; @thorson2013]. La probabilité que les CPUE ne soient pas nulles pour la première composante peut alors être multipliée par les CPUE attendues pour la deuxième composante en vue d’obtenir une estimation inconditionnelle des CPUE. Toutefois, cette approche pose problème sur quelques points :

1. L’approche du MLG Delta ajoute de la complexité en nécessitant l’ajustement et la production de rapports sur deux modèles.

1. Dans l’approche typique du MLG Delta, les deux modèles sont ajustés selon des liens distincts et les coefficients ne peuvent donc pas être combinés.

1. L’approche du MLG Delta suppose l’indépendance entre les deux composantes [p. ex. @thorson2017].

1. Il a été démontré que l’approche du MLG Delta n’est pas suffisamment solide par rapport à l’intensité variable de l’échantillonnage (p. ex. dans le temps ou l’espace) [@lecomte2013].

1. Ce qui est peut-être le plus important pour nous, c’est qu’un MLG Delta dans lequel les deux modèles utilisent des liens différents donne un indice final pour lequel la tendance dépend des niveaux de référence spécifiques auxquels les prédicteurs sont fixés [p. ex. @maunder2004a].

La distribution de Tweedie [@jorgensen1987] résout les problèmes ci­dessus [p. ex. @candy2004; @shono2008; @foster2013; @lecomte2013; @thorson2017], mais elle n’a pas été largement utilisée, probablement principalement en raison du coût informatique du calcul de la fonction de densité de probabilité de Tweedie. Récemment, la fonction de densité de Tweedie a été introduite dans le logiciel TMB [@kristensen2016] et peut être adaptée relativement rapidement aux grands ensembles de données et aux modèles comportant de nombreux paramètres d’effets fixes et aléatoires, soit au moyen de modèles TMB codés sur mesure, soit au moyen du progiciel R glmmTMB [@brooks2017].

En plus d’un paramètre moyen, la distribution de Tweedie comporte deux autres paramètres : un paramètre de puissance $p$ et un paramètre de dispersion $phi$. Si $1 > p > 2$, la distribution de Tweedie représente une distribution composée entre la distribution de Poisson ($p = 1$) et la distribution Gamma ($p = 2$) (figure \@ref(fig:cpue­tweedie­ex)). En fait, la distribution de Tweedie est aussi appelée « distribution composée Poisson­Gamma » dans ce cas délimité. Nous notons toutefois que la distribution composée Poisson­Gamma est souvent utilisée pour faire référence à un reparamétrage dans lequel les composantes Poisson et Gamma sont ajustées de sorte qu’elles ne sont pas censées avoir les mêmes coefficients prédictifs, comme c’est le cas dans la distribution de Tweedie [p. ex. @foster2013; @lecomte2013].

Nous adaptons le MLGM (modèle linéaire généralisé à effets mixtes) de Tweedie comme suit :

\begin{align}

(\#eq:cpue­tweedie)

y\_i &\sim \mathrm{Tweedie}(\mu\_i, p, \phi), \quad 1 < p < 2,\\

\mu\_i &= \exp \left(

\bm{X}\_i \bm{\beta}

+ \alpha^\mathrm{locality}\_{j[i]}

+ \alpha^\mathrm{locality-year}\_{k[i]}

+ \alpha^\mathrm{vessel}\_{l[i]}

\right),\\

\alpha^\mathrm{locality}\_j &\sim

\mathrm{Normal}(0, \sigma^2\_{\alpha \; \mathrm{locality}}),\\

(\#eq:cpue­locality­year)

\alpha^\mathrm{locality-year}\_k &\sim

\mathrm{Normal}(0, \sigma^2\_{\alpha \; \mathrm{locality-year}}),\\

(\#eq:cpue­vessel)

\alpha^\mathrm{vessel}\_l &\sim

\mathrm{Normal}(0, \sigma^2\_{\alpha \; \mathrm{vessel}}),

\end{align}

où $i$ représente une sortie (données historiques) ou un trait (données modernes) unique, $y\_i$ représente les captures (kg) par unité d’effort (heures de chalutage), $\bm{X\_i}$ représente un vecteur de prédicteurs à effet fixe (historiques : profondeur, mois; modernes : profondeur, mois, latitude), $\bm{\beta}$ représente un vecteur de coefficients et $\mu\_i$ représente les CPUE attendues pour une sortie ou un trait. Les interceptions des effets aléatoires (symboles $\alpha$) peuvent différer de l’interception globale par lieu $j$ (\(\alpha^\mathrm{locality}\_j\)), lieu­année $k$ (\(\alpha^\mathrm{locality-year}\_k\)) et navire $l$ (\(\alpha^\mathrm{vessel}\_l\)) (uniquement pour l’ensemble des données modernes) et sont limitées par des distributions normales dont les écarts­types respectifs sont indiqués par les paramètres $\sigma$.

On peut alors calculer l’estimation normalisée des CPUE pour l’année $t$, $\mu\_t$, comme suit :

\begin{equation}

\mu\_t = \exp \left(\bm{X}\_t \bm{\beta} \right)

\end{equation}

où $\bm{X\_t}$ représente un vecteur de prédicteurs fixé aux niveaux de référence ($r$), avec l’année fixée à l’année d’intérêt. Comme chacune des interceptions aléatoires $\alpha$ est fixée à zéro, l’indice est prédit pour un lieu, un lieu­année et un navire moyens (pour les données modernes). Nous avons estimé les effets fixes avec une vraisemblance marginale maximale tout en intégrant les effets aléatoires avec le logiciel statistique TMB à l’aide du progiciel R glmmTMB. Nous avons utilisé les erreurs-types ($\mathrm{SE}$) calculées par le logiciel TMB sur $\log (\mu\_t)$ selon la méthode delta. Nous avons ensuite calculé les intervalles de confiance de Wald à 95\% selon la formule $\exp (\mu\_t \pm 1.96 \mathrm{SE}\_t)$.

À titre de comparaison, nous avons calculé une série chronologique non normalisée à l’aide d’une procédure semblable, mais sans aucune des covariables autres qu’un prédicteur pour chaque année. Ce calcul est semblable au calcul de la moyenne géométrique de la CPUE chaque année, mais en supposant un modèle d’observation de Tweedie au lieu d’un modèle d’observation log-normal qui ne permet pas les zéros.

# RÉSULTATS DE LA NORMALISATION DES CPUE COMMERCIALES

Les données brutes sur les captures et l’effort de la flottille de pêche de la morue du Pacifique affichent des tendances variables au fil du temps dans les deux ensembles de données, données historiques et données modernes (figure \@ref(fig:cpue-catch-effort-ts)). La majorité des événements de pêche positifs pour la morue du Pacifique se sont produits à des profondeurs d’environ 50 m à 200 m (figure \@ref(fig:cpue-depth-hists)). Nous pouvons constater une augmentation du nombre total de combinaisons sortie-lieu depuis 1991, certains changements dans les lieux habituellement pêchés et le début de sorties avec des profondeurs moyennes de pêche plus grandes vers la fin de la période 1956-1995 (figures \@ref(fig:cpue-bubble-plots-hist-3CD), \@ref(fig:cpue-bubble-plots-hist-5ABCD)). Pour les ensembles de données modernes, nous observons certains changements dans la distribution des profondeurs, des latitudes et des mois de pêche au fil du temps, ainsi que des changements dans les navires faisant partie de la « flottille » et dans les

lieux de pêche (figures \@ref(fig:cpue-bubble-plots-modern-3CD), \@ref(fig:cpue-bubble-plots-modern-5ABCD)).

Les modèles de normalisation de l’indice du MLGM de Tweedie sont relativement bien ajustés aux données (figure \@ref(fig:cpue-quantile-residuals)). Les paramètres $p$ de Tweedie avaient tendance à se situer autour de $1.6$, dénotant une distribution à peu près à mi-chemin entre les distributions de Poisson et Gamma et les paramètres $phi$ indiquaient des observations relativement dispersées (tableau \@ref(tab:cpue-pars)). La variabilité était beaucoup plus grande dans les effets aléatoires du lieu et du lieu-année que dans les effets aléatoires des navires (tableau \@ref(tab:cpue-pars)).

Pour la période entre 1956 et 1995, la profondeur et le lieu ont eu un effet modéré sur l’indice normalisé de la CPUE pour la zone 3CD, mais les séries normalisée et non normalisée étaient quelque peu différentes dans la zone 5ABCD (figure \@ref(fig:cpue-index-ts-hist)). La prise en compte de la profondeur ou du lieu a réduit la CPUE dans la zone 3CD dans les années 1970 et l’a augmentée du milieu des années 1980 à 1995. La prise en compte de la profondeur a également réduit un pic de la CPUE dans la zone 3CD au milieu des années 1960. De 1996 à 2017, ce sont la profondeur et la latitude qui ont eu l’effet le plus important sur l’indice normalisé et, encore une fois, un effet plus important dans la zone 3CD que dans la zone 5ABCD (figure \@ref(fig:cpue-index-ts-modern)). La prise en compte de la profondeur ou de la latitude a légèrement abaissé l’indice de la CPUE pendant deux à trois ans avant et après 2010 et cet effet s’est répercuté sur le modèle de normalisation avec toutes les covariables.

La prise en compte des interactions lieu-année a eu peu d’effet sur la forme des indices normalisés, à l’exception d’un léger changement de forme en 2014-2015 pour l’ensemble de données moderne dans la zone 5ABCD (figure \@ref(fig:cpue-int-test-plot)). Le principal effet de l’inclusion des interactions de l’effet aléatoire du lieu-année a été d’augmenter la largeur des intervalles de confiance dans toutes les zones et toutes les périodes. Nous pouvons examiner la contribution de tous les paramètres d’effet fixe et aléatoire au moyen de graphiques de coefficients (historiques : figures \@ref(fig:cpue-coef-plot1), \@ref(fig:cpue-coef-plot2), \@ref(fig:cpue-coef-plot3); modernes : figures \@ref(fig:cpue-coef-plot1-modern), \@ref(fig:cpue-coef-plot2-modern), \@ref(fig:cpue-coef-plot3-modern)).

## INTERACTIONS ESPACE-TEMPS (LIEU-ANNÉE)

Pour tester l’effet de l’inclusion ou de l’omission des interactions espace-temps lorsque de telles interactions sont présentes ou non, nous avons effectué un essai de simulation. Bien qu’un essai de simulation complet comportant de nombreuses itérations stochastiques et une fourchette de valeurs des paramètres dépasse la portée de la présente annexe, nous pensons que cette simple simulation demeure instructive. Nous avons paramétré notre simulation pour qu’elle corresponde approximativement aux paramètres estimés à partir des données observées. Notre simulation comprenait les éléments suivants : 20 années de données; 12 lieux avec leurs effets ($\alpha^\mathrm{locality}\_{j}$) dans l’espace logarithmique, tirés d’une distribution normale avec un écart-type de 0,3 et une moyenne de 0; des effets aléatoires optionnels année-lieu ($\alpha^\mathrm{locality-year}\_{j}$) tirés d’une distribution avec un écart-type de 0,5 et une moyenne de 0; 10 observations par année par lieu; un indice réel connu de la CPUE qui, dans l’espace logarithmique, suivait un processus autorégressif avec une corrélation de 0,3 au retard 1, un écart-type de 1 et une moyenne de 2; et des paramètres de Tweedie de $p = 1.6$ et $\phi = 5$. Nous avons généré des versions de cet ensemble de données avec et sans les interactions lieu-année, puis nous avons ajusté des modèles de normalisation à ces ensembles de données qui permettaient ou ignoraient les interactions lieu-année.

En ce qui concerne les données réelles, l’inclusion des effets aléatoires lieu-année a permis de dégager pour chaque lieu une tendance qui s’écarte légèrement de la tendance globale (figure \@ref(fig:cpue-re-int-ts)). L’omission de ces effets aléatoires lieu-par-année, en revanche, supposait que la tendance de la CPUE est identique en forme et que seule son ampleur changeait d’un lieu à l’autre (figure \@ref(fig:cpue-re-no-int-ts)). L’omission de ces interactions espace-temps peut produire des intervalles de confiance beaucoup trop étroits si les tendances ne sont, en fait, pas identiques dans tout l’espace (figure \@ref(fig:cpue-sim-test-tweedie-glmm-plot)). En outre, le fait de permettre les interactions n’a aucun effet qualitatif sur le rendement du modèle ou sur la couverture si les interactions ne sont pas présentes (figure \@ref(fig:cpue-sim-test-tweedie-glmm-plot)).

L’ajustement d’un modèle adéquat de normalisation spatiotemporelle géostatistique constituerait une solution de rechange à ces effets aléatoires lieu-année [p. ex. @thorson2015b; @monnahan2018]. Pour cette évaluation, nous avons choisi de modéliser la variation spatiale selon les lieux du MPO afin de maintenir la cohérence avec les évaluations antérieures dans cette région. Toutefois, à l’avenir, nous pourrions envisager un modèle de normalisation spatiotemporelle.

\clearpage

```{r cpue-funcs}

source(here::here('R/cpue-functions.R'))

```

```{r cpue-params}

params <- list()

params$species\_proper <- "Pacific Cod"

params$april1\_year <- TRUE

params$area <- c("5[ABCD]+", "3[CD]+")

params$area\_name <- c("5ABCD", "3CD")

params$skip\_single\_variable\_models <- FALSE

```

```{r cpue-run-historic, message=FALSE, warning=FALSE, results='hide'}

params$era <- "historic"

source(here::here("R/cpue.R"))

dfleet\_hist <- dfleet

gg\_cpue\_hist <- gg\_cpue

cpue\_pred\_hist <- predictions

arith\_cpue\_hist <- arith\_cpue

m\_historic <- readRDS(here::here("data/generated/cpue-models-pcod-historic.rds"))

```

```{r cpue-run-modern, message=FALSE, warning=FALSE, results='hide'}

params$era <- "modern"

source(here::here("R/cpue.R"))

dfleet\_modern <- dfleet

gg\_cpue\_modern <- gg\_cpue

cpue\_pred\_modern <- predictions

arith\_cpue\_modern <- arith\_cpue

m\_modern <- readRDS(here::here("data/generated/cpue-models-pcod-modern.rds"))

```

```{r cpue-save}

readr::write\_csv(cpue\_pred\_modern, here::here("data/generated/cpue-predictions-modern.csv"))

readr::write\_csv(cpue\_pred\_hist, here::here("data/generated/cpue-predictions-historical.csv"))

```

```{r cpue-locality-map-5, fig.cap="Lieux du MPO utilisés dans le modèle moderne de normalisation de la CPUE dans la zone 5ABCD.", fig.asp=0.9}

gfplot:::plot\_dfo\_localities(unique(dfleet\_modern[[1]]$locality\_code))

```

```{r cpue-locality-map-3, fig.cap="Lieux du MPO utilisés dans le modèle moderne de normalisation de la CPUE dans la zone 3CD.", fig.asp=0.9}

gfplot:::plot\_dfo\_localities(unique(dfleet\_modern[[2]]$locality\_code))

```

```{r cpue-tweedie-ex, fig.asp=0.7, fig.cap="Exemples de fonctions de densité pour la distribution de Tweedie. Le symbole $\\phi$ (représenté par « phi » sur cette figure) représente le paramètre de dispersion, $p$ représente le paramètre de puissance et $\\mu$ représente la moyenne. Il convient de noter que le pic de densité que l’on voit sur la gauche des graphiques a une valeur de 0 sur l’axe des x."}

plot\_tweedie\_ex <- function(df, max = 15) {

xx <- seq(0, max, length.out = 1000)

out <- plyr::mdply(df, function(mu, power, phi) {

data.frame(x = xx,

y = tweedie::dtweedie(y = xx, mu = mu, power = power, phi = phi))

}) %>%

mutate(phi = paste("phi =", phi)) %>%

mutate(power = paste("p =", power)) %>%

mutate(mu = paste("μ =", mu))

ggplot(out, aes(x, y, colour = mu)) +

geom\_line() +

facet\_grid(power~phi, scales = "free") +

gfplot::theme\_pbs() +

labs(colour = "μ") +

xlab("Value") + ylab("Density") +

coord\_cartesian(expand = FALSE, xlim = c(-0.2, max(out$x))) +

scale\_colour\_brewer(palette = "Dark2") +

scale\_fill\_brewer(palette = "Dark2")

}

df <- expand.grid(power = c(1.2, 1.6, 1.8), mu = c(1, 3, 6), phi = c(0.5, 1, 2))

plot\_tweedie\_ex(df)

```

```{r cpue-catch-effort-ts, fig.cap="Série chronologique brute des captures de morue du Pacifique et des heures totales de chalutage (indépendamment des espèces capturées). Les données antérieures à 1996 sont présentées séparément des données postérieures à 1996.", fig.asp=1.1}

g1 <- make\_catch\_effort\_ts\_plot(dfleet\_hist) +

ggtitle("1956-1995 data")

g2 <- make\_catch\_effort\_ts\_plot(dfleet\_modern) +

ggtitle("1996+ data")

cowplot::plot\_grid(g1, g2, ncol = 1)

```

```{r, cpue-depth-hists, fig.asp=0.8, fig.cap="Distribution en profondeur pour les sorties de pêche (rangée du haut) et les événements de chalutage (rangée du bas) qui ont permis ou non de capturer des morues du Pacifique."}

g1 <- gg\_cpue\_modern$depth + ylab("Nombre d’événements de pêche") ("Profondeur") +

labs(fill = "Fishing event\ncaught\nPacific Cod") +

theme(legend.position = c(0.95, 0.95), legend.background = element\_rect(fill = "white"),

legend.justification = c(1, 1)) +

ggtitle("1996+ data") +

scale\_colour\_brewer(palette = "Dark2") +

scale\_fill\_brewer(palette = "Dark2")

g2 <- gg\_cpue\_hist$depth + ylab("Trip count") + xlab("Depth") +

labs(fill = "Trip\ncaught\nPacific Cod") +

theme(legend.position = c(0.95, 0.95), legend.background = element\_rect(fill = "white"),

legend.justification = c(1, 1)) +

ggtitle("1956-1995 data") +

scale\_colour\_brewer(palette = "Dark2") +

scale\_fill\_brewer(palette = "Dark2")

cowplot::plot\_grid(g2, g1, ncol = 1)

```

```{r cpue-bubble-plots-hist-3CD, fig.asp=1.35, fig.cap="Distribution des prédicteurs dans les modèles de normalisation de la CPUE pour l’ensemble de données de 1956 à 1995 dans la zone 3CD. La superficie des cercles extérieurs représente le nombre de combinaisons sortie-lieu pour cette combinaison de la valeur prédictive et de l’année. La superficie et l’ombrage des cercles intérieurs représentent le nombre de combinaisons sortie-lieu pour cette combinaison de la valeur prédictive et de l’année qui ont capturé des morues du Pacifique."}

dfleet\_hist %>%

make\_bubble\_dat(measure.vars = c("month", "locality", "depth")) %>%

filter(area == "3CD") %>%

make\_facet\_bubble\_plot(group = "trips") +

ggtitle("1956-1995 3CD")

```

(ref:caption-cpue-bubble-plots-hist-5ABCD) Même que pour la figure \@ref(fig:cpue-bubble-plots-hist-3CD) but for 5ABCD.

(ref:caption-cpue-bubble-plots-modern-3CD) Même que pour la figure \@ref(fig:cpue-bubble-plots-hist-3CD) but for 1996--2017 3CD.

(ref:caption-cpue-bubble-plots-modern-5ABCD) Même que pour la figure \@ref(fig:cpue-bubble-plots-hist-3CD) but for 1996--2017 5ABCD.

```{r cpue-bubble-plots-hist-5ABCD, fig.asp=1.35, fig.cap="(ref:caption-cpue-bubble-plots-hist-5ABCD)"}

dfleet\_hist %>%

make\_bubble\_dat(measure.vars = c("month", "locality", "depth")) %>%

filter(area == "5ABCD") %>%

make\_facet\_bubble\_plot(group = "trips") +

ggtitle("1956-1995 5ABCD")

```

```{r cpue-bubble-plots-modern-3CD, fig.asp=1.4, fig.width=9, fig.cap="(ref:caption-cpue-bubble-plots-modern-3CD)"}

dfleet\_modern %>% make\_bubble\_dat() %>%

filter(area == "3CD") %>%

make\_facet\_bubble\_plot() +

ggtitle("1996+ 3CD")

```

```{r cpue-bubble-plots-modern-5ABCD, fig.asp=1.4, fig.width=9, fig.cap="(ref:caption-cpue-bubble-plots-modern-5ABCD)"}

dfleet\_modern %>% make\_bubble\_dat() %>%

filter(area == "5ABCD") %>%

make\_facet\_bubble\_plot() +

ggtitle("1996+ 5ABCD")

```

(ref:caption-cpue-quantile-residuals) Histogrammes des résidus de quantiles randomisés [@dunn1996] pour les modèles de normalisation des MLGM des CPUE. Les histogrammes illustrent la distribution de la densité réelle de 10 000 résidus de quantiles randomisés sélectionnés au hasard. Les lignes en pointillés indiquent la densité de probabilité pour une distribution normale avec le même écart­type.

```{r cpue-quantile-residuals, fig.asp=0.8, fig.cap="(ref:caption-cpue-quantile-residuals)"}

stopifnot(length(m\_historic) == 12L)

stopifnot(length(m\_modern) == 16L)

qr1 <- qres\_tweedie(m\_historic[[6]], "cpue")

gqr1 <- plot\_qres\_histogram(qr1) + ggtitle(paste("Historical", params$area\_name[1]))

qr2 <- qres\_tweedie(m\_historic[[12]], "cpue")

gqr2 <- plot\_qres\_histogram(qr2) + ggtitle(paste("Historical", params$area\_name[2]))

qr3 <- qres\_tweedie(m\_modern[[8]], "cpue")

gqr3 <- plot\_qres\_histogram(qr3) + ggtitle(paste("Modern", params$area\_name[1]))

qr4 <- qres\_tweedie(m\_modern[[16]], "cpue")

gqr4 <- plot\_qres\_histogram(qr4) + ggtitle(paste("Modern", params$area\_name[2]))

cowplot::plot\_grid(gqr2, gqr1, gqr4, gqr3, ncol = 2)

```

```{r cpue-pars, results='asis'}

get\_cpue\_pars <- function(obj, name = "") {

re <- broom.mixed::tidy(obj, effects = "ran\_pars")

re$group <- gsub("\_", "-", re$group)

re <- tibble(Model = name, Parameter = paste(re$group, "SD"),

Estimate = round(re$estimate, 2))

p <- round(stats::plogis(obj$fit$par[["thetaf"]]) + 1, 2)

phi <- round(exp(obj$fit$par[["betad"]]), 2)

bind\_rows(re,

tibble(Model = name, Parameter = c("$p$", "$\\phi$"), Estimate = c(p, phi)))

}

pars <- list()

pars[[1]] <- get\_cpue\_pars(m\_historic[[12]], "Historical 3CD")

pars[[2]] <- get\_cpue\_pars(m\_historic[[6]], "Historical 5ABCD")

pars[[3]] <- get\_cpue\_pars(m\_modern[[16]], "Modern 3CD")

pars[[4]] <- get\_cpue\_pars(m\_modern[[8]], "Modern 5ABCD")

bind\_rows(pars) %>%

knitr::kable(caption = "Écart-type (ET) de l’effet aléatoire et estimations des paramètres de puissance ($p$) et de dispersion ($\\phi$) du modèle d’observation de Tweedie.", booktabs = TRUE, linesep = "", escape = FALSE, format = "pandoc")

```

```{r cpue-index-ts-hist, fig.asp=1.15, fig.cap="Modèles de normalisation de la CPUE pour le chalut commercial. Sur tous les graphiques, la ligne noire et la région ombrée en noir indiquent un indice de la CPUE avec seulement un prédicteur d’année. La ligne colorée et les rubans ombrés en couleur indiquent des indices normalisés par un ou plusieurs prédicteurs. Les trois premières rangées illustrent des modèles de normalisation qui comprennent un seul prédicteur, précisé à droite. L’avant-dernière rangée illustre un modèle de normalisation qui inclut tous les prédicteurs dans un modèle. La dernière rangée illustre un modèle de normalisation qui comprend tous les prédicteurs plus les effets aléatoires lieu-par-année (espace-temps). Les interactions lieu et lieu-navire sont ajustées comme des effets aléatoires et toutes les autres variables comme des effets fixes."}

make\_cpue\_ts\_dat(cpue\_pred\_hist) %>% make\_cpue\_ts\_plot() +

ggtitle("1956-1995 CPUE") +

scale\_colour\_brewer(palette = "Dark2") +

scale\_fill\_brewer(palette = "Dark2")

```

(ref:caption-cpue-index-ts-modern) Même chose que pour la figure \@ref(fig:cpue-index-ts-hist), mais pour les données de 1996 à 2017. Les interactions lieu, navire et lieu-navire sont ajustées comme des effets aléatoires et toutes les autres variables comme des effets fixes.

```{r cpue-index-ts-modern, fig.asp=1.55, fig.cap="(ref:caption-cpue-index-ts-modern)", out.width="5.2in", fig.width=6.5}

make\_cpue\_ts\_dat(cpue\_pred\_modern) %>% make\_cpue\_ts\_plot() +

ggtitle("1996+ CPUE") +

scale\_colour\_brewer(palette = "Dark2") +

scale\_fill\_brewer(palette = "Dark2")

```

```{r cpue-int-test-plot, fig.asp=0.85, fig.cap="Comparaison des séries chronologiques des CPUE normalisées avec un modèle qui n’inclut pas les effets aléatoires lieu-année (espace-temps) (noir/gris) et avec un modèle qui inclut les effets aléatoires lieu-année (en couleur)."}

g1 <- make\_cpue\_ts\_dat\_noint(cpue\_pred\_hist) %>% make\_cpue\_ts\_plot\_noint() +

ggtitle("1956-1995 CPUE")

g2 <- make\_cpue\_ts\_dat\_noint(cpue\_pred\_modern) %>% make\_cpue\_ts\_plot\_noint() +

ggtitle("1996+ CPUE")

cowplot::plot\_grid(g1, g2, ncol = 1)

```

```{r cpue-coef-plot1, fig.asp=1.3, echo=FALSE, fig.cap="Coefficients à effet fixe pour le modèle de normalisation des CPUE commerciales historiques. Les points et les segments épais et minces représentent les moyennes et les intervalles de confiance de Wald à 50\\% et 95\\%."}

toplot <- c(6, 12)

p <- list()

for (i in seq\_along(params$area\_name)) {

p[[i]] <- make\_fe\_plots(m\_historic[[toplot[i]]]) + ggtitle(params$area\_name[i])

}

cowplot::plot\_grid(p[[2]], p[[1]], ncol = 1)

```

```{r cpue-coef-plot2, fig.asp=1, echo=FALSE, fig.cap="Effets aléatoires du lieu pour le modèle de normalisation des CPUE commerciales historiques."}

p <- list()

for (i in seq\_along(params$area\_name)) {

p[[i]] <- make\_re\_plots(m\_historic[[toplot[i]]], re\_names = "locality") +

ggtitle(params$area\_name[i]) + xlab("Random intercept value\n(log space)")

}

cowplot::plot\_grid(p[[2]], p[[1]], ncol = 2)

```

```{r cpue-coef-plot3, fig.asp=1.35, echo=FALSE, fig.cap="Effets aléatoires du lieu-par-année (espace-temps) pour le modèle de normalisation des CPUE commerciales historiques.", fig.width=10.5}

p <- list()

for (i in seq\_along(params$area\_name)) {

p[[i]] <- make\_year\_locality\_plots(m\_historic[[toplot[i]]]) + ggtitle(params$area\_name[i])

}

cowplot::plot\_grid(p[[2]], p[[1]], ncol = 1)

```

```{r cpue-coef-plot1-modern, fig.asp=1.2, echo=FALSE, fig.cap="Coefficients à effet fixe pour le modèle de normalisation des CPUE commerciales modernes. Les points et les segments épais et minces représentent les moyennes et les intervalles de confiance de Wald à 50\\% et 95\\%."}

toplot <- c(8, 16)

p <- list()

for (i in seq\_along(params$area\_name)) {

p[[i]] <- make\_fe\_plots(m\_modern[[toplot[i]]]) + ggtitle(params$area\_name[i])

}

cowplot::plot\_grid(p[[2]], p[[1]], ncol = 2)

```

```{r cpue-coef-plot2-modern, fig.asp=0.8, echo=FALSE, fig.cap="Effets aléatoires du lieu et du navire pour le modèle de normalisation des CPUE commerciales modernes." }

p <- list()

for (i in seq\_along(params$area\_name)) {

p[[i]] <- make\_re\_plots(m\_modern[[c(8, 16)[i]]], re\_names = c("locality", "vessel")) +

ggtitle(params$area\_name[i])

}

cowplot::plot\_grid(p[[2]], p[[1]], ncol = 2)

```

```{r cpue-coef-plot3-modern, fig.asp=1.35, echo=FALSE, fig.cap="Effets aléatoires du lieu-par-année (espace-temps) pour le modèle de normalisation des CPUE commerciales modernes."}

p <- list()

for (i in seq\_along(params$area\_name)) {

p[[i]] <- make\_year\_locality\_plots(m\_modern[[toplot[i]]]) + ggtitle(params$area\_name[i])

}

cowplot::plot\_grid(p[[2]], p[[1]], ncol = 1)

```

```{r cpue-re-int-ts, fig.cap="Tendances de l’indice des CPUE propre à un lieu pour un modèle de normalisation qui permet des interactions lieu-année (espace-temps). Les lignes colorées indiquent les estimations propres au lieu et tous les autres prédicteurs sont réglés à leur niveau de base. La ligne noire et le ruban ombré indiquent les CPUE annuelles moyennes globales.", fig.asp=0.8}

if (!file.exists(here::here("data/generated/cpue-re-preds.rda"))) {

stopifnot(length(m\_modern) == 16L)

g1 <- plot\_cpue\_int\_res(m\_modern[[8]], fleet = dfleet\_modern[[1]],

index\_data = cpue\_pred\_modern,

era = "modern") + ggtitle("1996+ 5ABCD")

g2 <- plot\_cpue\_int\_res(m\_modern[[16]], fleet = dfleet\_modern[[2]],

index\_data = cpue\_pred\_modern,

era = "modern") + ggtitle("1996+ 3CD")

stopifnot(length(m\_historic) == 12L)

g3 <- plot\_cpue\_int\_res(m\_historic[[6]], fleet = dfleet\_hist[[1]],

index\_data = cpue\_pred\_hist,

era = "historical")+ ggtitle("1956-1995 5ABCD")

g4 <- plot\_cpue\_int\_res(m\_historic[[12]], fleet = dfleet\_hist[[2]],

index\_data = cpue\_pred\_hist,

era = "historical")+ ggtitle("1956-1995 3CD")

save(g1, g2, g3, g4, file = here::here("data/generated/cpue-re-preds.rda"))

} else {

load(here::here("data/generated/cpue-re-preds.rda"))

}

cowplot::plot\_grid(g4, g3, g2, g1, ncol = 2)

```

```{r cpue-re-no-int-ts, fig.cap="Tendances de l’indice des CPUE propre à un lieu pour un modèle de normalisation qui ne permet pas les interactions lieu-année (espace-temps). Les lignes colorées indiquent les estimations propres au lieu et tous les autres prédicteurs sont réglés à leur niveau de base. La ligne noire et le ruban ombré indiquent les CPUE annuelles moyennes globales.", fig.asp=0.8}

if (!file.exists(here::here("data/generated/cpue-re-preds-noint.rda"))) {

g1 <- plot\_cpue\_int\_res(m\_modern[[7]], fleet = dfleet\_modern[[1]],

index\_data = cpue\_pred\_modern,

the\_formula\_version = "Full standardization minus interactions",

era = "modern") + ggtitle("1996+ 5ABCD")

g2 <- plot\_cpue\_int\_res(m\_modern[[15]], fleet = dfleet\_modern[[2]],

index\_data = cpue\_pred\_modern,

the\_formula\_version = "Full standardization minus interactions",

era = "modern") + ggtitle("1996+ 3CD")

g3 <- plot\_cpue\_int\_res(m\_historic[[5]], fleet = dfleet\_hist[[1]],

index\_data = cpue\_pred\_hist,

the\_formula\_version = "Full standardization minus interactions",

era = "historical")+ ggtitle("1956-1995 5ABCD")

g4 <- plot\_cpue\_int\_res(m\_historic[[11]], fleet = dfleet\_hist[[2]],

index\_data = cpue\_pred\_hist,

the\_formula\_version = "Full standardization minus interactions",

era = "historical")+ ggtitle("1956-1995 3CD")

save(g1, g2, g3, g4, file = here::here("data/generated/cpue-re-preds-noint.rda"))

} else {

load(here::here("data/generated/cpue-re-preds-noint.rda"))

}

cowplot::plot\_grid(g4, g3, g2, g1, ncol = 2)

```

```{r cpue-sim-test-tweedie-glmm, message=FALSE, warning=FALSE, results='hide'}

set.seed(99)

N <- 12

nyrs <- 20

locs <- LETTERS[seq\_len(N)]

loc\_effects <- rnorm(N, 0, 0.3)

yr\_effects <- as.numeric(arima.sim(n = nyrs, list(ar = 0.3), sd = 1, mean = 2))

obs\_per\_yr\_loc <- 10

tweedie\_p <- 1.6

tweedie\_phi <- 5

x <- data.frame(loc = rep(locs, each = nyrs \* obs\_per\_yr\_loc),

loc\_ef = rep(loc\_effects, each = nyrs \* obs\_per\_yr\_loc))

x <- mutate(x, yr = rep(seq\_len(nyrs), N \* obs\_per\_yr\_loc),

yr\_ef = rep(yr\_effects, N \* obs\_per\_yr\_loc))

x <- mutate(x, mu = yr\_ef + loc\_ef)

x <- mutate(x, yr\_x\_loc = paste(yr, loc))

x <- mutate(x, y = tweedie::rtweedie(length(mu),

mu = exp(mu), power = tweedie\_p, phi = tweedie\_phi))

# ggplot(x, aes(yr, y)) + geom\_point() + geom\_line(aes(x = yr, y = exp(yr\_ef)))

yr\_loc\_int <- expand.grid(yr = unique(x$yr), loc = unique(x$loc))

yr\_loc\_int <- mutate(yr\_loc\_int, yr\_x\_loc = paste(yr, loc))

yr\_loc\_int <- mutate(yr\_loc\_int, yr\_x\_loc\_ef = rnorm(nrow(yr\_loc\_int), 0, 0.5))

x <- left\_join(x, yr\_loc\_int, by = c("loc", "yr", "yr\_x\_loc"))

x <- mutate(x, mu\_w\_int = yr\_ef + loc\_ef + yr\_x\_loc\_ef)

x <- mutate(x, y\_w\_int =

tweedie::rtweedie(length(mu\_w\_int),

mu = exp(mu\_w\_int), power = tweedie\_p, phi = tweedie\_phi))

# ggplot(x, aes(yr, y\_w\_int)) + geom\_point() + geom\_line(aes(x = yr, y = exp(yr\_ef)))

d1 <- x %>% sample\_frac(1) %>%

reshape2::melt(id.vars = c("loc", "yr", "yr\_ef"), measure.vars = c("y", "y\_w\_int"))

# d1 %>% ggplot(aes(yr, value)) + geom\_point(alpha = 0.7, pch = 21) +

# geom\_line(aes(x = yr, y = exp(yr\_ef))) +

# facet\_wrap(~variable)

d1 <- mutate(d1, year\_factor = as.factor(yr), cpue = value)

d1 <- mutate(d1, yr\_x\_loc = paste(yr, loc))

invisible(capture.output({

m\_int <- plyr::dlply(d1, "variable", function(xx)

gfplot::fit\_cpue\_index\_glmmtmb(xx, formula = cpue ~ 0 + year\_factor + (1|loc) + (1|yr\_x\_loc)))

m\_no\_int <- plyr::dlply(d1, "variable", function(xx)

gfplot::fit\_cpue\_index\_glmmtmb(xx, formula = cpue ~ 0 + year\_factor + (1|loc)))

}))

x1 <- plyr::ldply(m\_int, function(xx) gfplot::predict\_cpue\_index\_tweedie(xx)) %>%

mutate(model = "Space-time interactions")

x2 <- plyr::ldply(m\_no\_int, function(xx) gfplot::predict\_cpue\_index\_tweedie(xx)) %>%

mutate(model = "No space-time interactions")

pp <- bind\_rows(x1, x2)

orig\_dat <- select(x, yr, yr\_ef) %>% unique() %>%

mutate(yr\_ef = yr\_ef - mean(yr\_ef)) %>%

rename(year = yr)

```

```{r cpue-sim-test-tweedie-glmm-plot, fig.cap="Exemple de simulation illustrant l’effet de la modélisation ou non des interactions espace-temps comme effets aléatoires dans un modèle de normalisation de l’indice des CPUE. Le graphique de gauche représente un scénario dans lequel les données ont été générées avec la même tendance pour tous les lieux dans l’espace. Le graphique de droite représente un scénario dans lequel les données ont été générées avec des interactions espace-temps. Les lignes vertes et orange et les régions ombrées représentent les indices estimés des CPUE des modèles qui permettent les interactions espace-temps ou ne permettent pas les interactions espace-temps, avec les intervalles de confiance de 95\\%. La ligne noire pointillée indique la moyenne réelle de la CPUE pour chaque année. Toutes les combinaisons de modèles et de données ont une couverture correcte à 95\\%, à l’exception du modèle sans interaction espace-temps ajusté aux données qui a des interactions espace-temps, dont la couverture est de 55\\%. Il convient de noter que les intervalles de confiance se chevauchent complètement sur le graphique de gauche.", fig.asp=0.45}

pp <- left\_join(pp, orig\_dat, by = "year") %>%

mutate(est\_link = est\_link - mean(est\_link)) %>%

ungroup() %>%

mutate(variable = gsub("^y$", "Données sans interaction espace-temps", variable)) %>%

mutate(variable = gsub("^y\_w\_int$", "Données avec interactions espace-temps", variable)) %>%

mutate(model = forcats::fct\_rev(model))

gg <- pp %>% group\_by(model, variable) %>%

ungroup() %>%

ggplot() +

geom\_ribbon(aes(x = year, ymin = exp(est\_link - 1.96 \* se\_link),

ymax = exp(est\_link + 1.96 \* se\_link), fill = model), alpha = 0.4) +

geom\_line(aes(x = year, y = exp(est\_link), colour = model), lwd = 0.8) +

facet\_grid(~variable) +

geom\_line(data = orig\_dat, aes(x = year, y = exp(yr\_ef)), lwd = 0.9, col = "grey20", lty = 2) +

labs(fill = « Modèle", couleur = "Modèle", "Année", "CPUE") +

gfplot::theme\_pbs() +

theme(legend.position = c(0.02, 0.97), legend.justification = c(0, 1)) +

scale\_colour\_brewer(palette = "Dark2") +

scale\_fill\_brewer(palette = "Dark2")

print(gg)

```

```{r cpue-sim-save}

ggplot2::ggsave(here::here("presentations/figures/cpue-sim.svg"), plot = gg,

width = 8, height = 4)

```

```{r cpue-sim-test-tweedie-glmm-table, results='asis'}

# group\_by(pp, variable, model) %>%

# rename(data = variable) %>%

# summarise(coverage =

# sum(yr\_ef < est\_link + 1.96 \* se\_link & yr\_ef > est\_link - 1.96 \* se\_link) / n()) %>%

# knitr::kable(caption = "TODO", booktabs = TRUE, linesep = "")

```

\clearpage