Responses of epilimnetic and hypolimnetic dissolved oxygen concentrations to variability in thermal structure

VH

2023-10-02

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.3.1
```

```
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
library(lubridate)
library(rLakeAnalyzer)
library(LakeMetabolizer)
library(ggdist)
library(patchwork)
library(MuMIn)
library(nlme)
```

```
##
## Attaching package: 'nlme'
##
## The following object is masked from 'package:lme4':
##
##     lmList
##
## The following object is masked from 'package:dplyr':
##
```

```
##     collapse
```

```
library(ggeffects)


# weather %>%
#   filter(month(date) %in% c(6,7) &
#            year(date) %in% c(2011:2022)) %>%
#   summarise(mean_pressure = mean(air_pressure, na.rm = T))

surface <-
  data.frame(temp = 0:20,
             do = o2.at.sat.base(0:20, baro = 1012)) # mean pressure


do_alt <-
  data.frame(temp = 0:20,
             do_150 = o2.at.sat.base(0:20, altitude = 150),
             do_300 = o2.at.sat.base(0:20, altitude = 300),
             do_450 = o2.at.sat.base(0:20, altitude = 450))


do <-
  read_csv('manuscript/data_to_upload/E-Hazukova_DO_DOC_SC.csv')
```

```
## Rows: 232 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): lake, layer
## dbl  (4): DO_mean, DO_sd, DOC, cond_mean
## dttm (1): date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
sample_data_sub_m <- read_csv("data/processed/Hazukova_et_al_date-level-data-for-models-vertical-profil
  mutate(year = year(date))
```

```
## Rows: 123 Columns: 19
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (2): lake, clim
## dbl  (16): zMix, schmidt, epiVA, hypoVA, temp_mean, temp_min, temp_max, wind...
## date  (1): date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

DO in hypolimnion

```
do_data <-
  do %>%
  filter(!is.na(DO_mean) &
           layer == 'hypo') %>%
  left_join(sample_data_sub_m,
```

```
            by = c('date', 'lake'))



# models

fmnull <- lmer(DO_mean ~ 1 + (1|lake),  REML = F, na.action =  na.omit, data = do_data)
do9s <- lmer(DO_mean ~ scale(diff) + (diff|lake), REML = F, na.action =  na.omit, data = do_data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
do12s <- lmer(DO_mean ~ scale(hypoVA) + (hypoVA|lake), REML = F, na.action =  na.omit, data = do_data)
do13s <- lmer(DO_mean ~ scale(schmidt) + (schmidt|lake), REML = F, na.action =  na.omit, data = do_data
```

```
## boundary (singular) fit: see help('isSingular')
```

```
do14s <- lmer(DO_mean ~ scale(zMix) + (zMix|lake), REML = F, na.action =  na.omit, data = do_data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
do15s <- lmer(DO_mean ~ scale(epiVA) + (epiVA|lake), REML = F, na.action =  na.omit, data = do_data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
AIC(fmnull, do9s, do12s, do13s, do14s, do15s)
```

```
##        df      AIC
## fmnull  3 521.8821
## do9s    6 520.8728
## do12s   6 489.8780
## do13s   6 527.7692
## do14s   6 527.6242
## do15s   6 525.6641
```

```
do12 <- lme(DO_mean ~ scale(hypoVA), random = ~hypoVA|lake, method = 'REML', na.action =  na.omit, data
do12_AR <- lme(DO_mean ~ scale(hypoVA), random = ~hypoVA|lake, correlation = corCAR1(form = ~1|lake), m

AIC(do12, do12_AR)
```

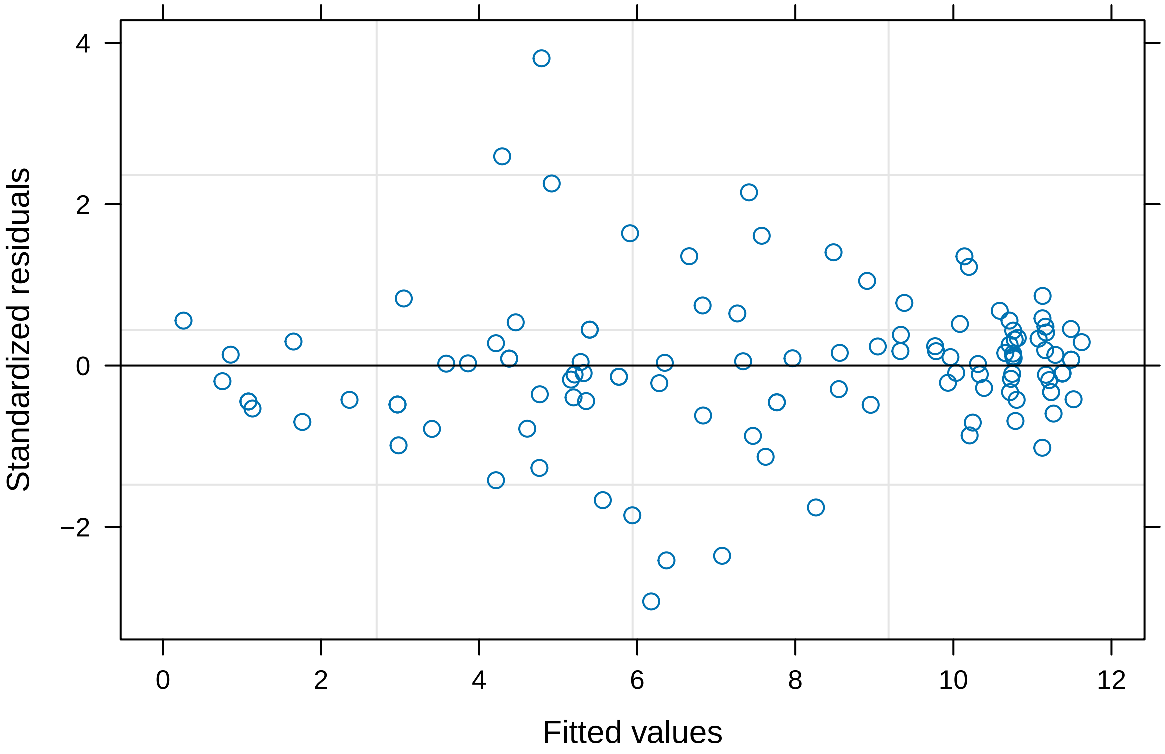
```
##         df      AIC
## do12     6 489.4662
## do12_AR  7 491.4662
```

```
summary(do12)
```

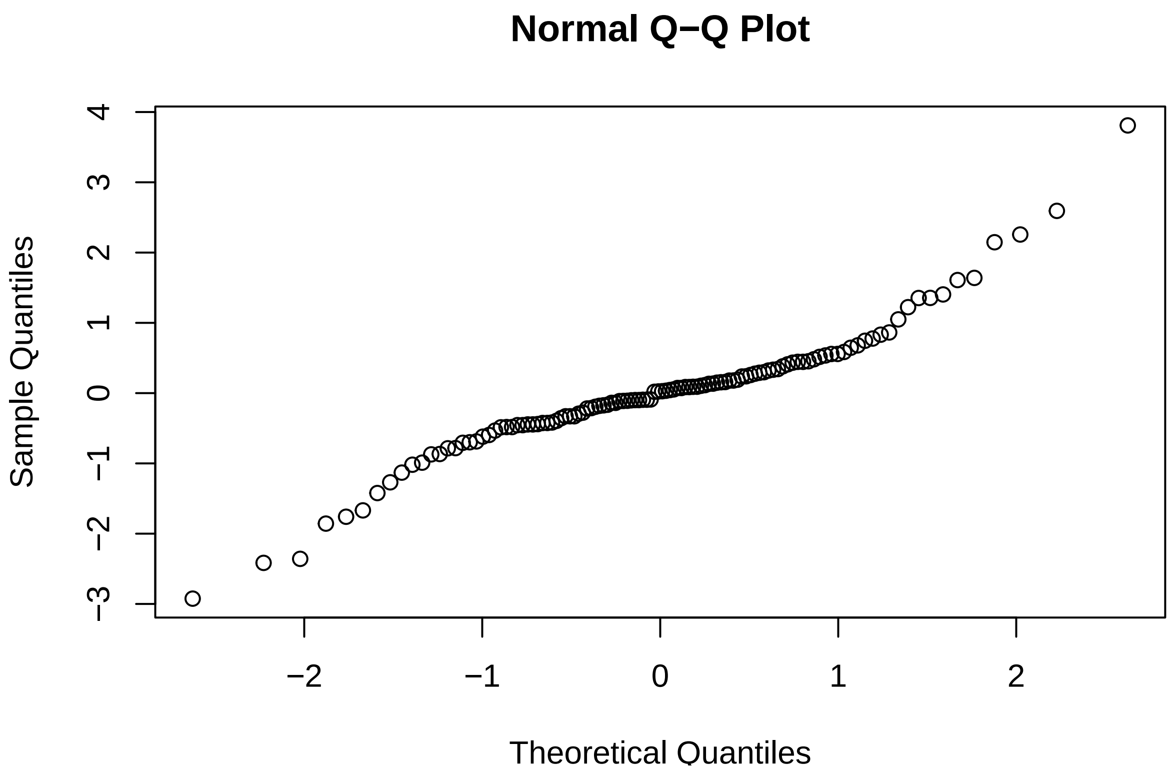
```
## Linear mixed-effects model fit by REML
##   Data: do_data
##        AIC      BIC    logLik
##   489.4662 505.8834 -238.7331
##
## Random effects:
##  Formula: ~hypoVA | lake
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 5.7974200 (Intr)
## hypoVA      0.5755503 -0.914
## Residual    1.5069800
##
```

```
## Fixed effects:  DO_mean ~ scale(hypoVA)
##                   Value Std.Error  DF  t-value p-value
## (Intercept)    7.964701 0.8371757 102 9.513774   0e+00
## scale(hypoVA) 0.971688 0.2765077 102 3.514144   7e-04
##  Correlation:
##               (Intr)
## scale(hypoVA) -0.461
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.92434214 -0.42903832  0.02517175  0.35183899  3.80920609
##
## Number of Observations: 116
## Number of Groups: 13
```

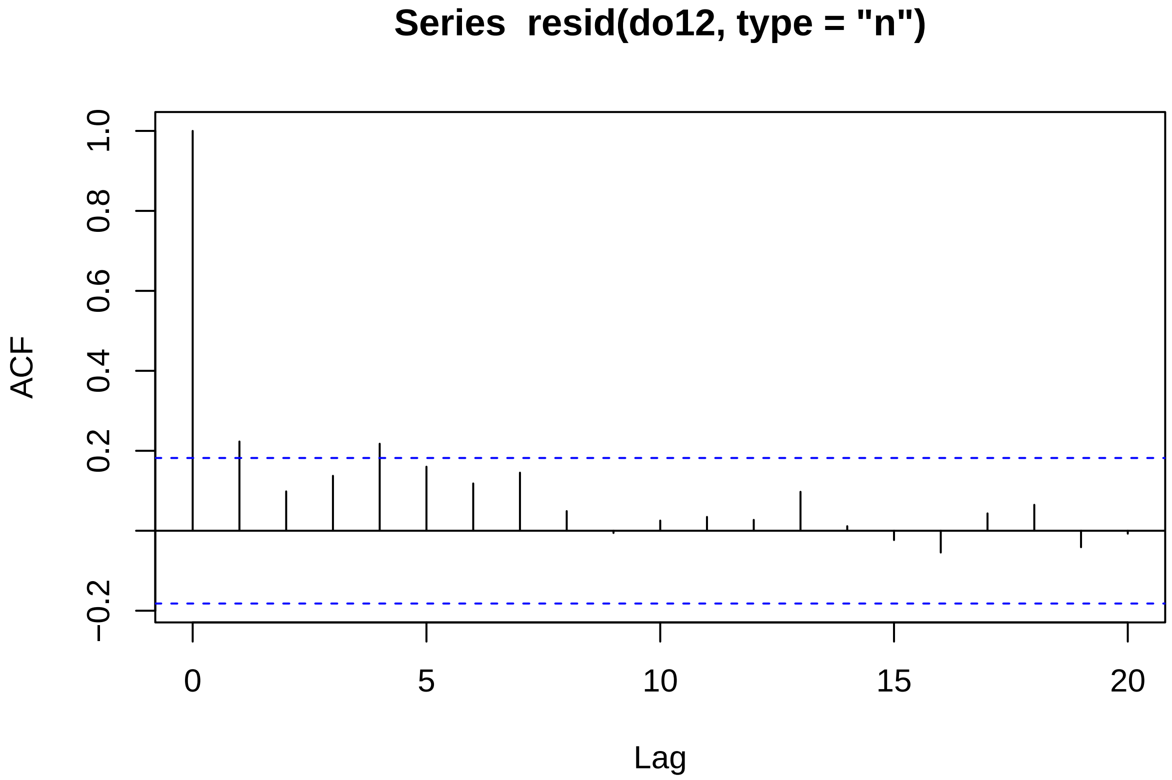
```r
plot(do12)
```



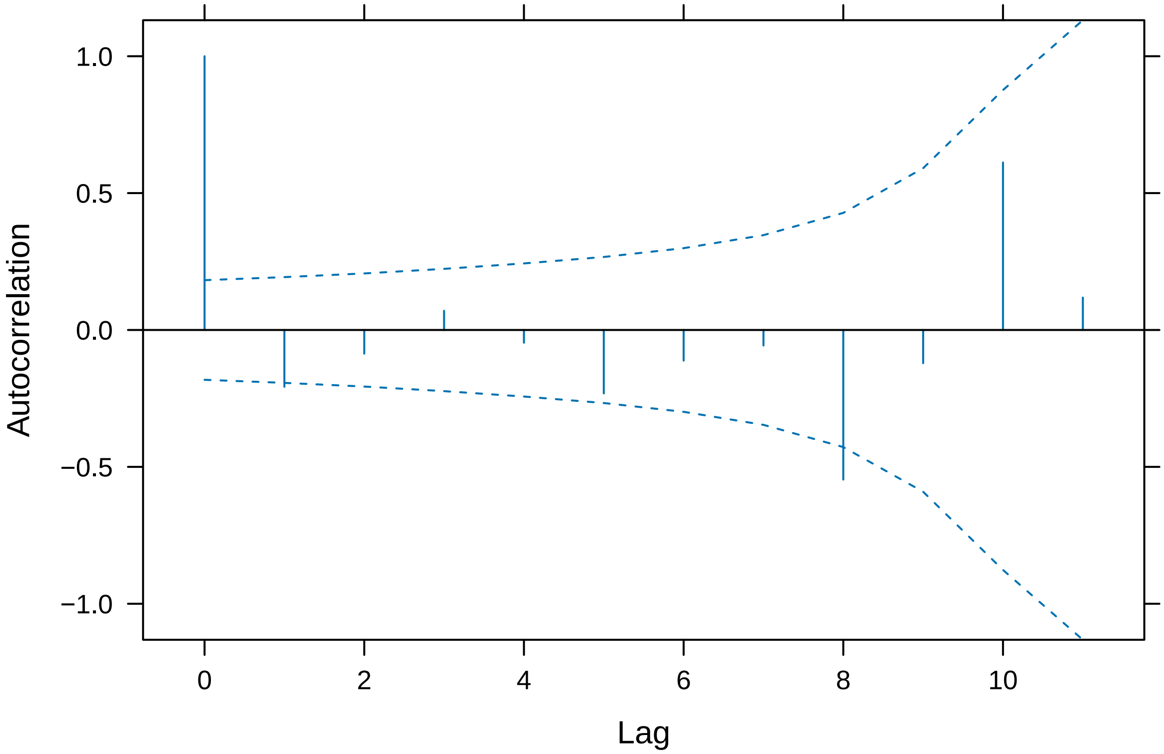
```r
qqnorm(resid(do12, type = 'n'))
```



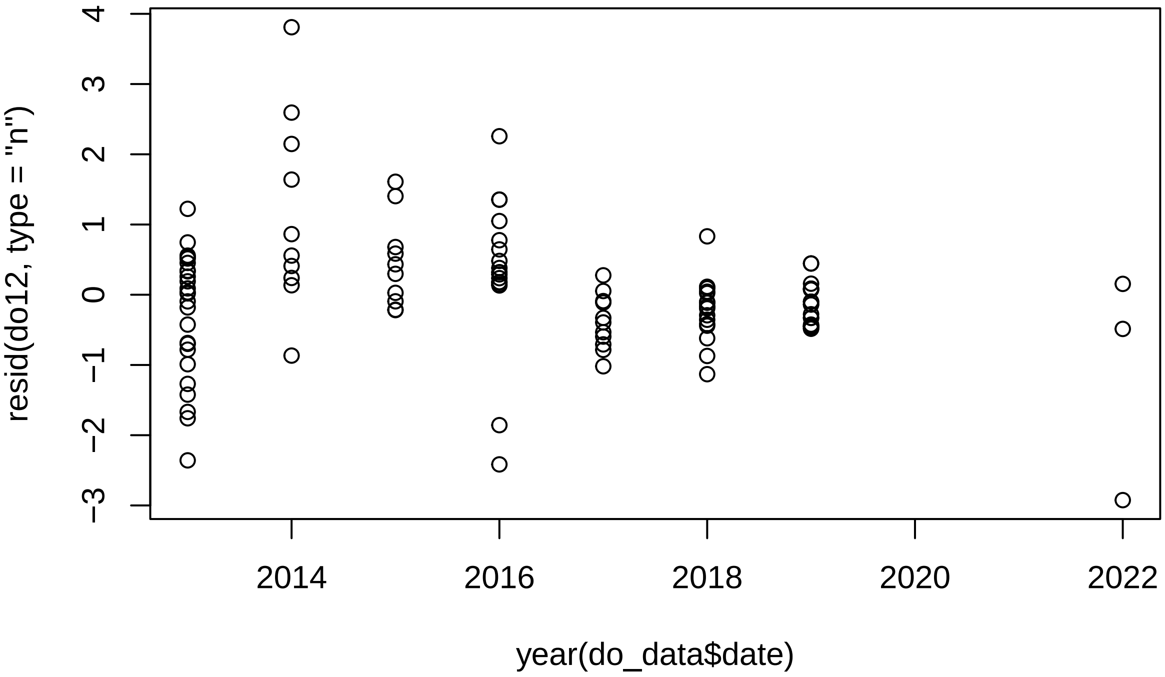
```r
acf(resid(do12, type = 'n'))
```



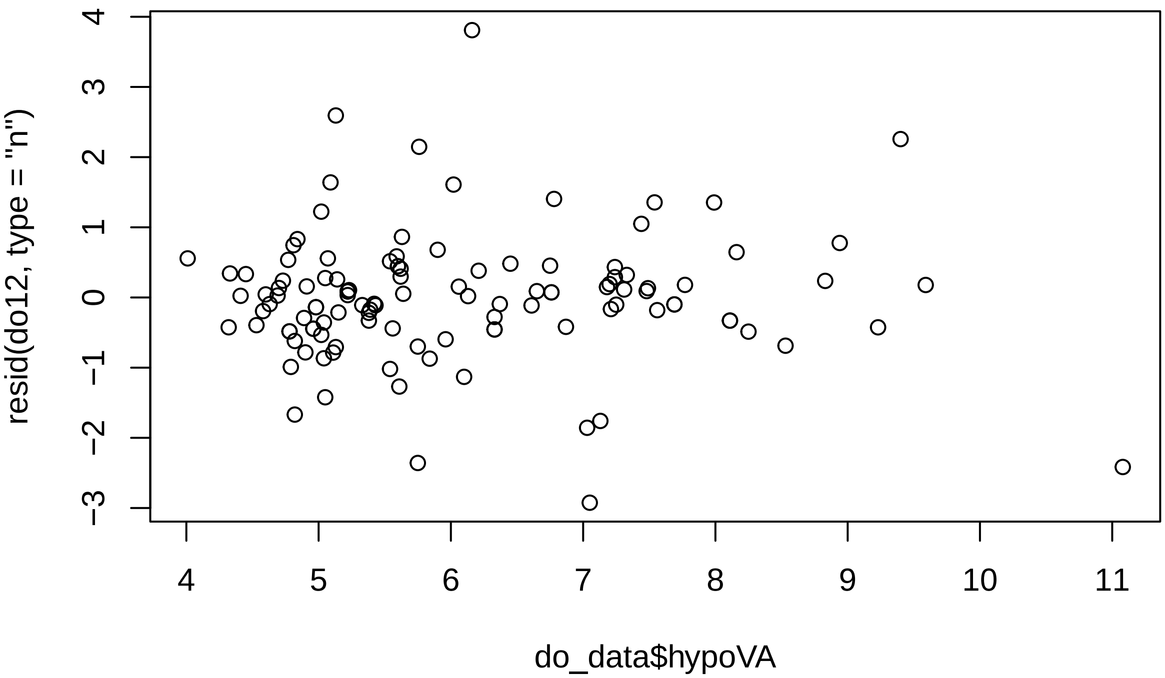
```r
plot(ACF(do12, type = 'n'), alpha = 0.05)
```



```
plot(resid(do12, type = 'n') ~
       year(do_data$date))
```



```
plot(resid(do12, type = 'n') ~
       do_data$hypoVA)
```



```
r.squaredGLMM(do12) # 0.07 / 0.82
```

```
## Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help page.
```

```
##             R2m      R2c
## [1,] 0.07489181 0.819866
```

```
x <-
  do_data %>%
  filter(month(date) %in% c(6,7) & layer == 'hypo') %>%
  group_by(lake) %>%
  summarise(doc_mean = mean(DOC, na.rm = T),
            do_mean = mean(DO_mean, na.rm = T),
            cond_mean = mean(cond_mean, na.rm = T))


cor(x$doc_mean, x$do_mean, use = 'complete', method = 'spearman') #-0.8720779
```

```
## [1] -0.8720779
```

```
cor(x$cond_mean, x$do_mean, use = 'complete', method = 'spearman') #-0.7582418
```

```
## [1] -0.7582418
```

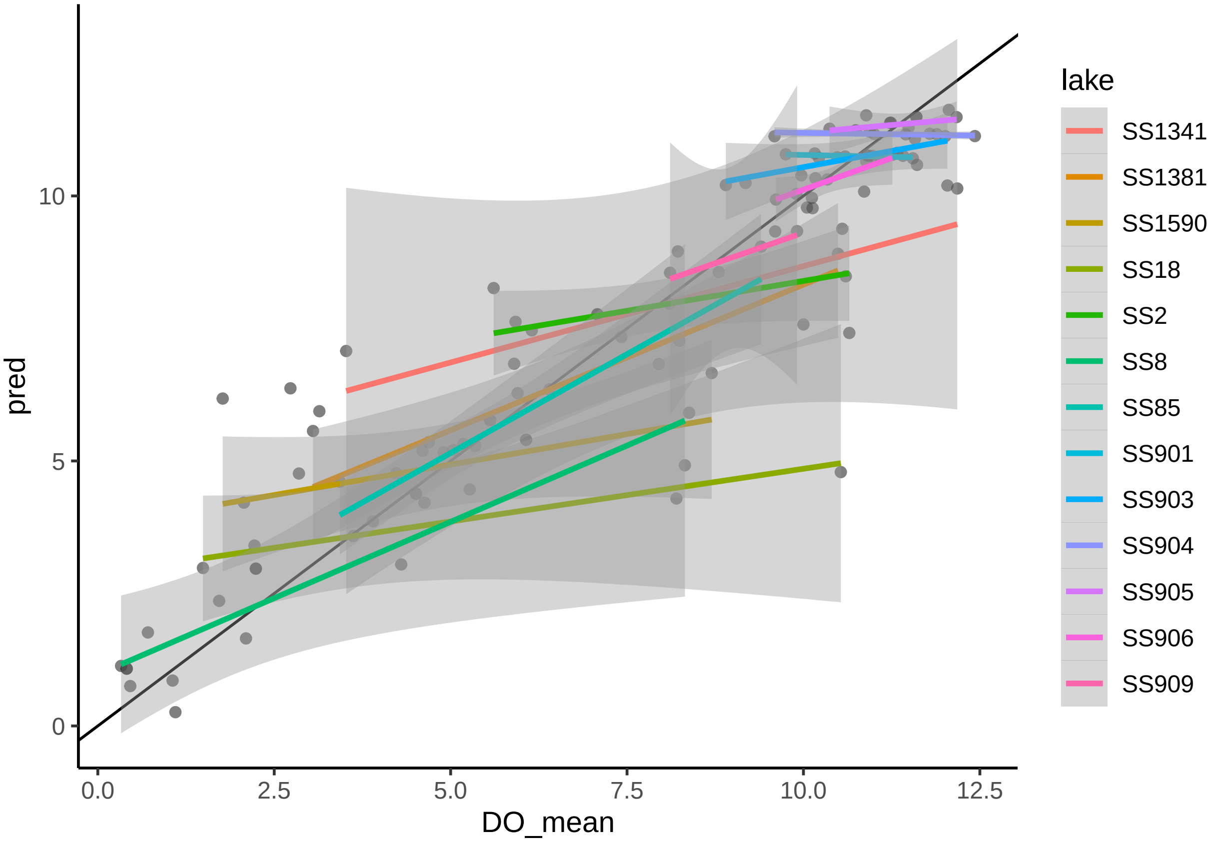
```
cor(x$doc_mean, x$cond_mean, use = 'complete', method = 'spearman') #0.95
```

```
## [1] 0.9546089
```

```
do_data %>%
  filter(!is.na(DO_mean)) %>%
  cbind(., pred = predict(do12)) %>%
  ggplot(aes(x = DO_mean, y = pred)) +
  geom_point(alpha = 0.5)  +
  theme_classic() +
  geom_abline(slope = 1, intercept = 0) +
  geom_smooth(aes(group = lake, color = lake), method = 'lm', size = 1)
```

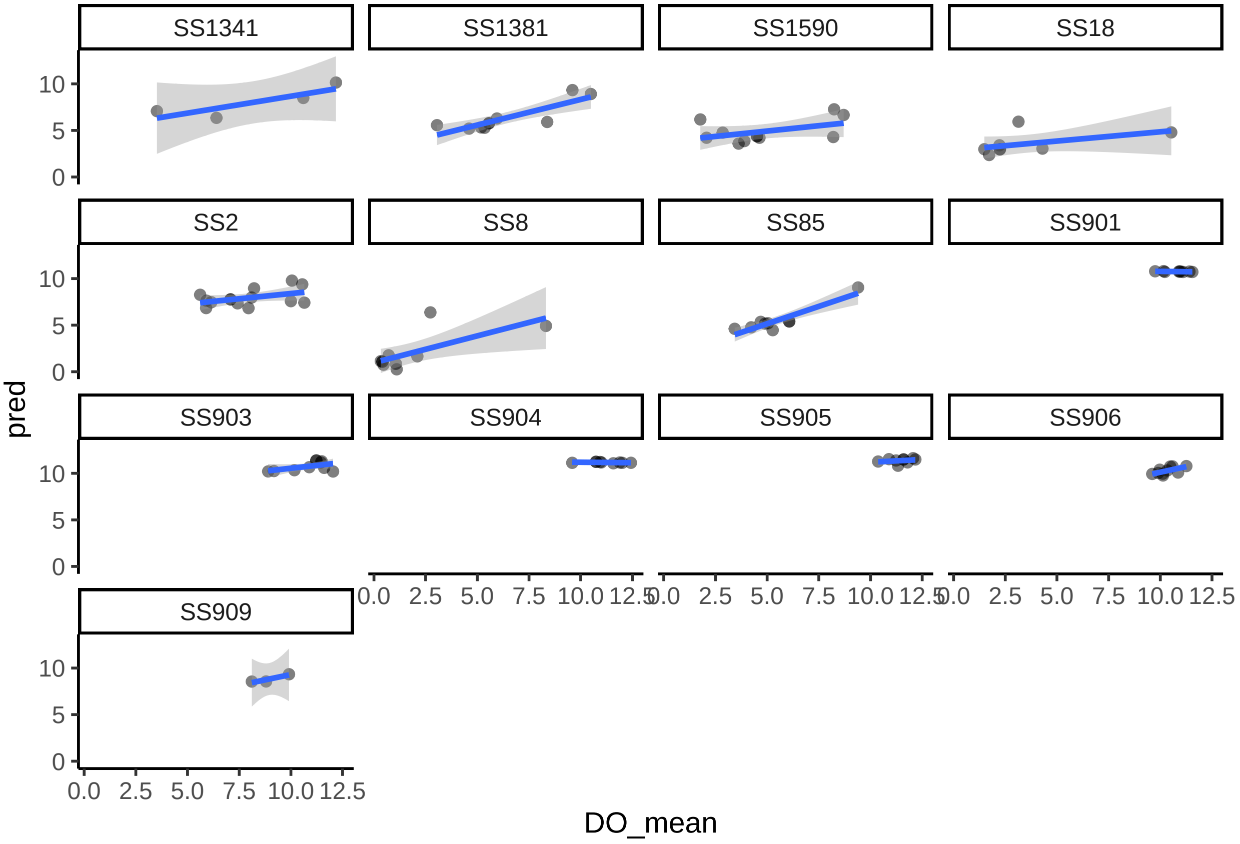
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

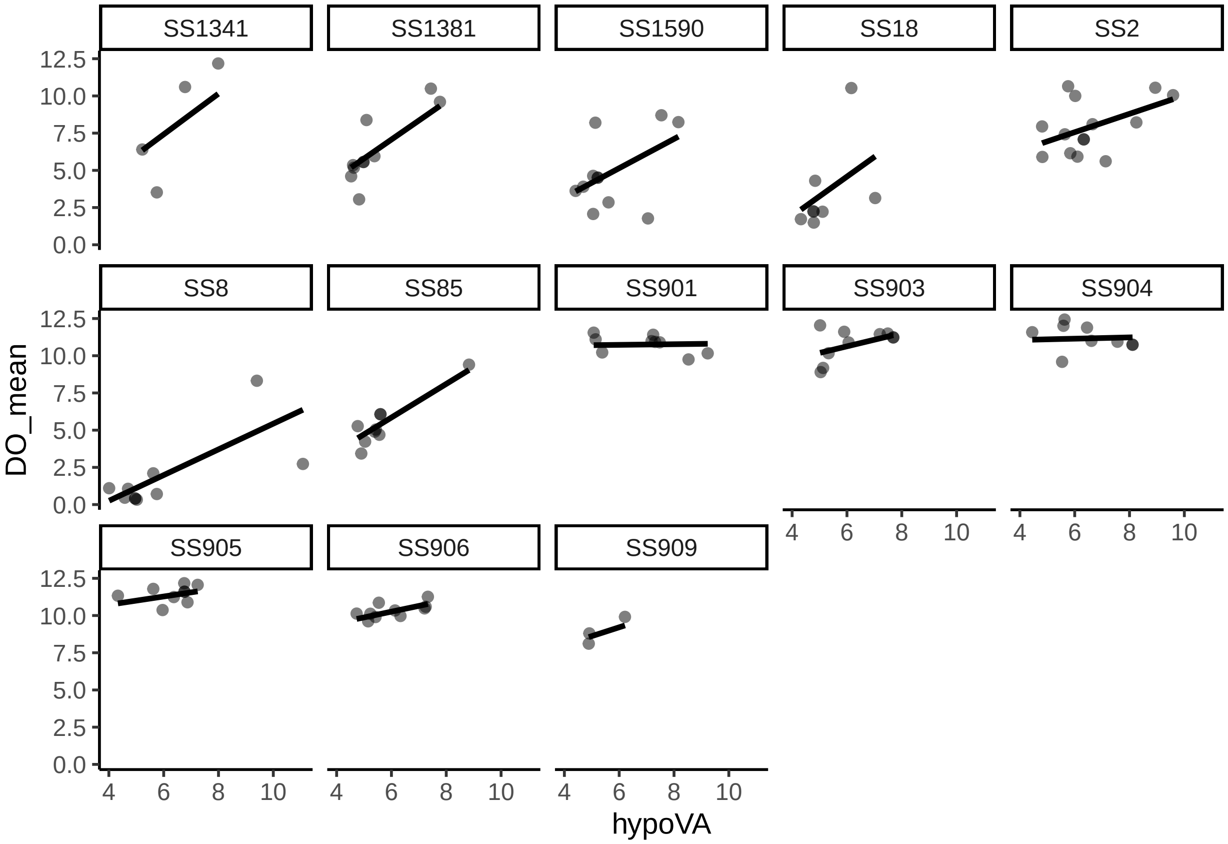


```
do_data %>%
  filter(!is.na(DO_mean)) %>%
  cbind(., pred = predict(do12)) %>%
  ggplot(aes(x = DO_mean, y = pred)) +
  geom_point(alpha = 0.5)  +
  geom_smooth(method = lm) +
  theme_classic() +
  facet_wrap(~lake)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
do_data %>%
  filter(!is.na(DO_mean)) %>%
  cbind(., pred = predict(do12)) %>%
  ggplot(aes(x = hypoVA, y = DO_mean)) +
  facet_wrap(~lake, ncol= 5) +
  geom_point(alpha = 0.5)  +
  theme_classic() +
  geom_line(aes(y = pred), size = 1)
```

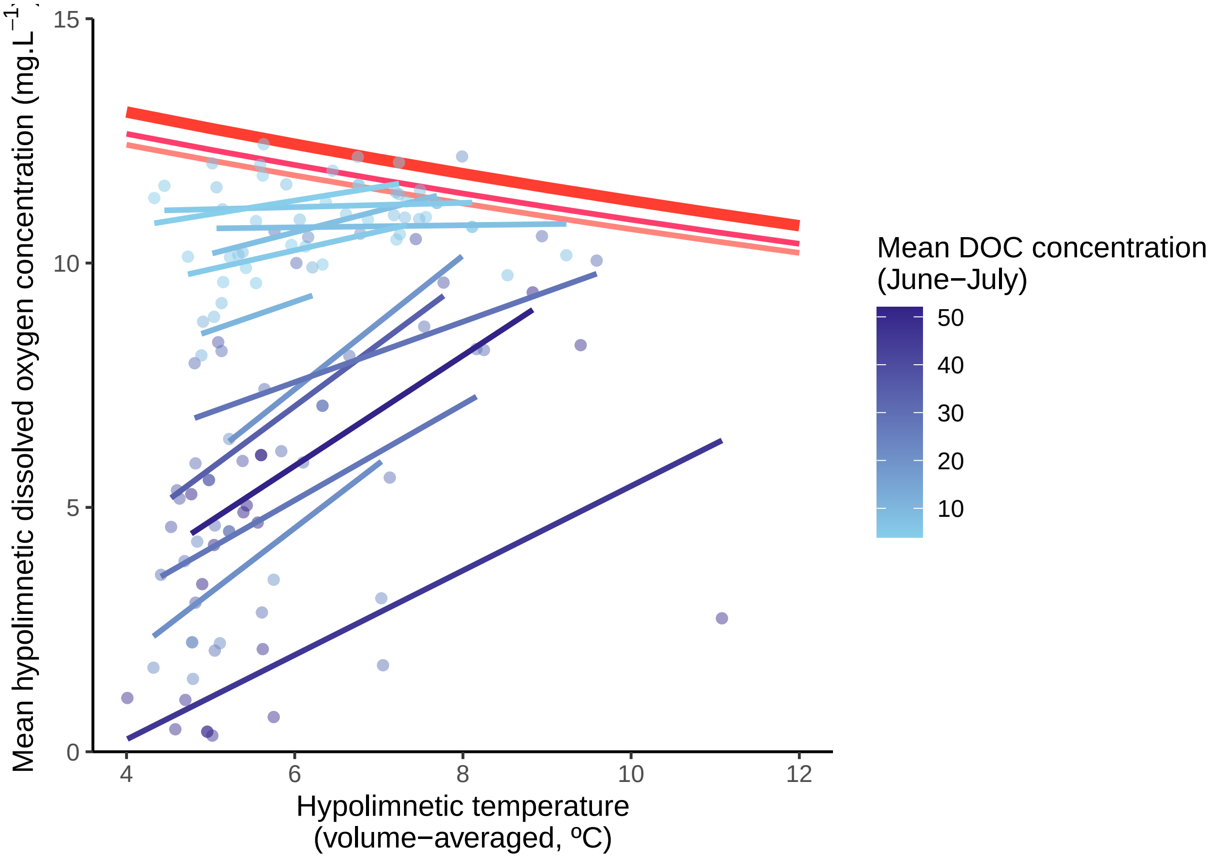


```
do_data_plot <-
  do_data %>%
  filter(!is.na(DO_mean)) %>%
  cbind(.,
        pred = predict(do12)) %>%
  left_join(x,
            by = 'lake')


hypo_do_plot <-
  ggplot() +
  geom_line(data = surface, aes(y = do, x = temp), color = '#FF3D31', linetype = 1, size = 2) +
  geom_line(data = do_alt, aes(y = do_300, x = temp), color = '#FF3D6C', size = 1) +
  geom_line(data = do_alt, aes(y = do_450, x = temp), color = '#FF857C', size = 1) +
  geom_point(data = do_data_plot,
             aes(x = hypoVA, y = DO_mean, color = doc_mean), alpha = 0.5)  +
  geom_line(data = do_data_plot,
            aes(x = hypoVA, y = pred, group = lake, color = doc_mean), size = 1) +
  scale_x_continuous(limits = c(4,12), expand = c(0.05,0)) +
  scale_y_continuous(limits = c(0,15), expand = c(0,0)) +
  scale_colour_gradient(
    'Mean DOC concentration\n(June-July)',
    low = "skyblue",
    high = '#332287',
    guide = "colourbar",
    aesthetics = "colour") +
  labs(x = 'Hypolimnetic temperature\n(volume-averaged, ºC)',
```

```
     y = bquote('Mean hypolimnetic dissolved oxygen concentration (mg.L'^-1*')')) +
  theme_classic()

hypo_do_plot
```

```
## Warning: Removed 12 rows containing missing values (`geom_line()`).
```

```
## Warning: Removed 12 rows containing missing values (`geom_line()`).
## Removed 12 rows containing missing values (`geom_line()`).
```

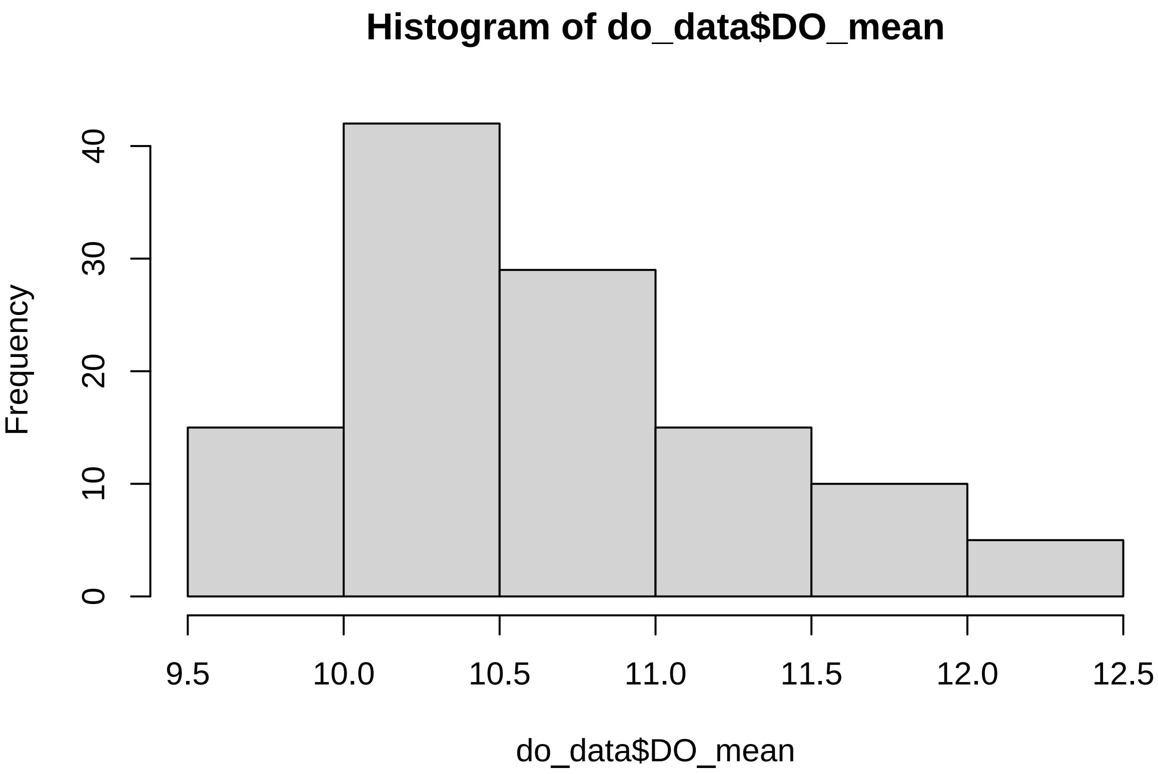


```
ggsave("output/plots/do_mean_lakes_temp.pdf", width = 6.5, height = 5, units = "in")
```

```
## Warning: Removed 12 rows containing missing values (`geom_line()`).
## Removed 12 rows containing missing values (`geom_line()`).
## Removed 12 rows containing missing values (`geom_line()`).
```

Epilimnion DO

```
do_data <-
  do %>%
  filter(!is.na(DO_mean) &
           layer == 'epi') %>%
  left_join(sample_data_sub_m,
            by = c('date', 'lake'))

hist(do_data$DO_mean) # there is no variability in the data (3 units)
```



```
do_data %>%
  group_by(lake) %>%
  summarise(mean(DO_mean), sd = sd(DO_mean))
```

```
## # A tibble: 13 x 3
##    lake   `mean(DO_mean)`    sd
##    <chr>            <dbl> <dbl>
##  1 SS1341            10.9 0.700
##  2 SS1381            10.7 0.455
##  3 SS1590            10.4 0.429
##  4 SS18              10.5 0.616
##  5 SS2               10.5 0.548
##  6 SS8               11.2 0.750
##  7 SS85              10.3 0.385
##  8 SS901             10.9 0.683
##  9 SS903             11.0 0.688
## 10 SS904             10.8 0.743
## 11 SS905             10.6 0.775
## 12 SS906             10.7 0.491
## 13 SS909             10.0 0.350
```

```
fmnull <- lmer(DO_mean ~ 1 + (1|lake),  REML = F, na.action =  na.omit, data = do_data)
do9 <- lmer(DO_mean ~ scale(diff) + (diff|lake), REML = F, na.action =  na.omit, data = do_data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
do12 <- lmer(DO_mean ~ scale(hypoVA) + (hypoVA|lake), REML = F, na.action =  na.omit, data = do_data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
do13 <- lmer(DO_mean ~ scale(schmidt) + (schmidt|lake), REML = F, na.action =  na.omit, data = do_data)
do14 <- lmer(DO_mean ~ scale(zMix) + (zMix|lake), REML = F, na.action =  na.omit, data = do_data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
do15 <- lmer(DO_mean ~ scale(epiVA) + (epiVA|lake), REML = F, na.action =  na.omit, data = do_data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
AIC(fmnull, do9, do12, do13, do14, do15)
```

```
##        df      AIC
## fmnull  3 227.7705
## do9     6 200.7655
## do12    6 224.7892
## do13    6 208.0075
## do14    6 234.4403
## do15    6 167.9976
```

```
do15 <- lme(DO_mean ~ scale(epiVA), random = ~epiVA|lake, control = lmeControl(opt='optim'), method = '
do15_AR <- lme(DO_mean ~ scale(epiVA), random = ~epiVA|lake, correlation = corCAR1(form = ~1|lake), con
AIC(do15, do15_AR)
```

```
##         df      AIC
## do15     6 175.2850
## do15_AR  7 132.9079
```

```
# model with AR is much better

r.squaredGLMM(do15)
```

```
##            R2m       R2c
## [1,] 0.2836831 0.5910069
```

```
r.squaredGLMM(do15_AR)
```

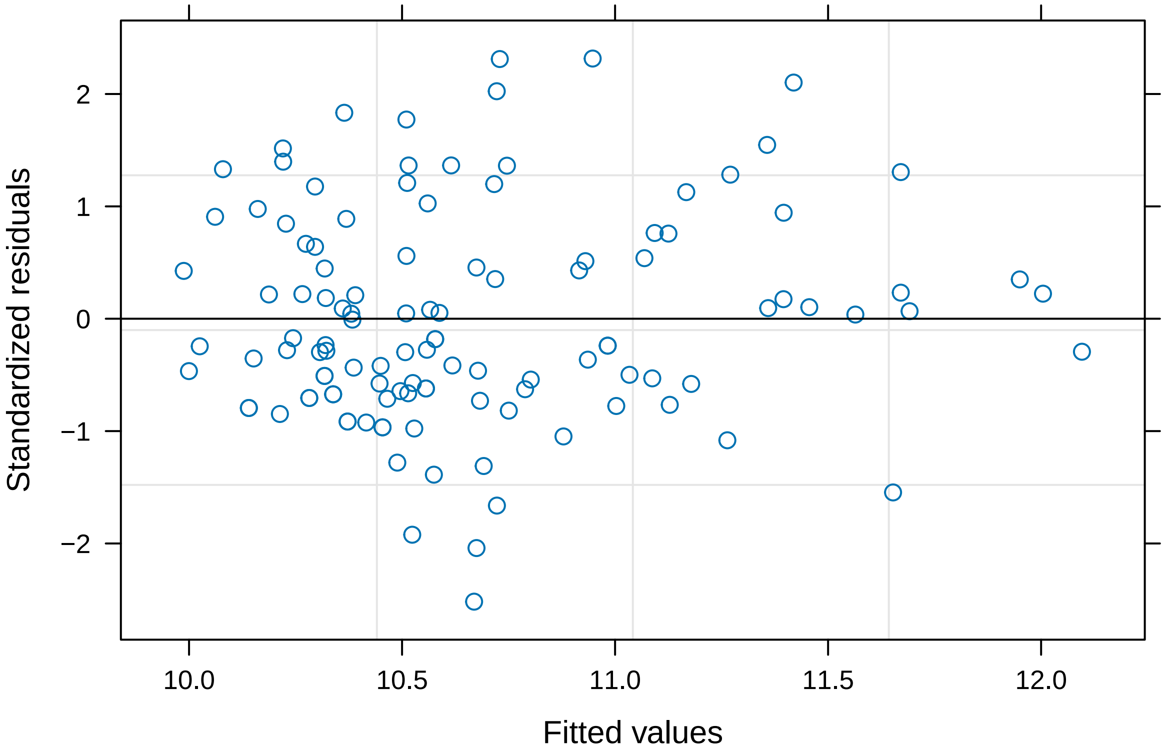
```
##            R2m      R2c
## [1,] 0.2460482 0.555462
```

```
summary(do15)
```

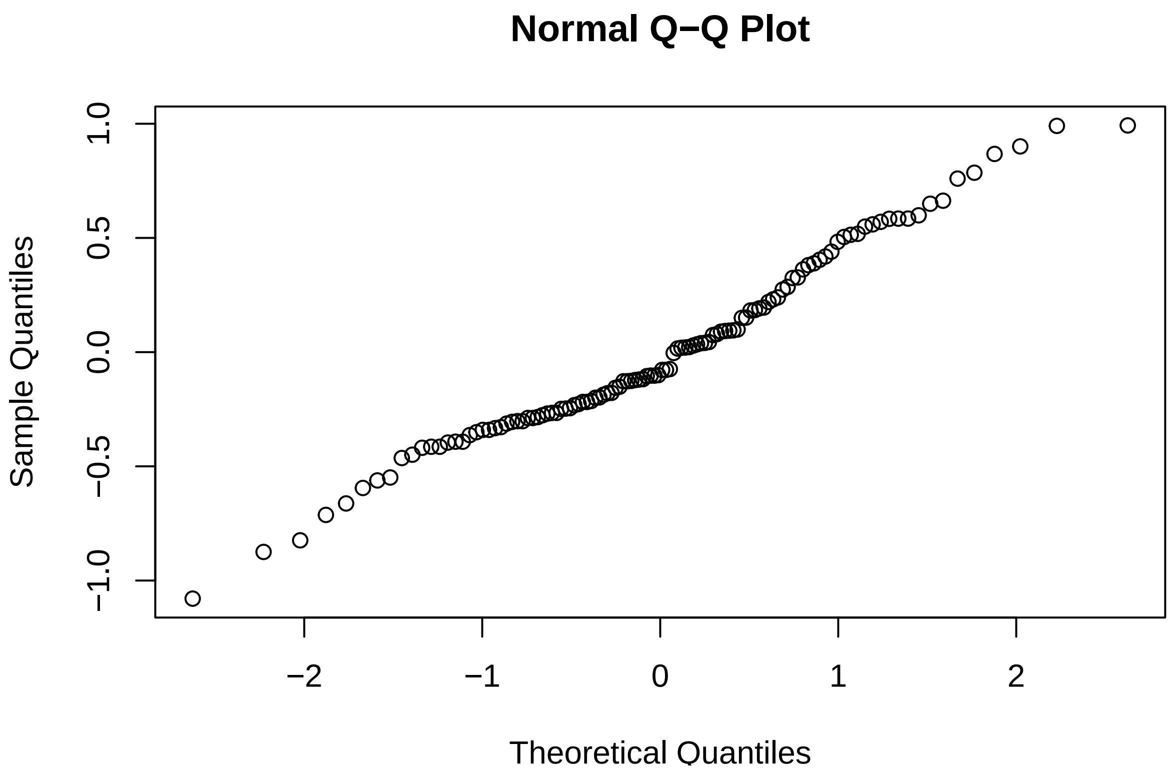
```
## Linear mixed-effects model fit by REML
##   Data: do_data
##       AIC      BIC   logLik
##   175.285 191.7022 -81.6425
##
## Random effects:
##  Formula: ~epiVA | lake
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 1.4549801 (Intr)
## epiVA       0.1296762 -0.999
## Residual    0.4285866
##
## Fixed effects:  DO_mean ~ scale(epiVA)
##                   Value  Std.Error  DF   t-value p-value
## (Intercept)  10.522175 0.06044528 102 174.07768  0.0000
## scale(epiVA) -0.356942 0.10665424 102  -3.34672  0.0011
```

```
##  Correlation:
##              (Intr)
## scale(epiVA) 0.546
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.5179346 -0.6491160 -0.2080517  0.5793135  2.3158681
##
## Number of Observations: 116
## Number of Groups: 13
```

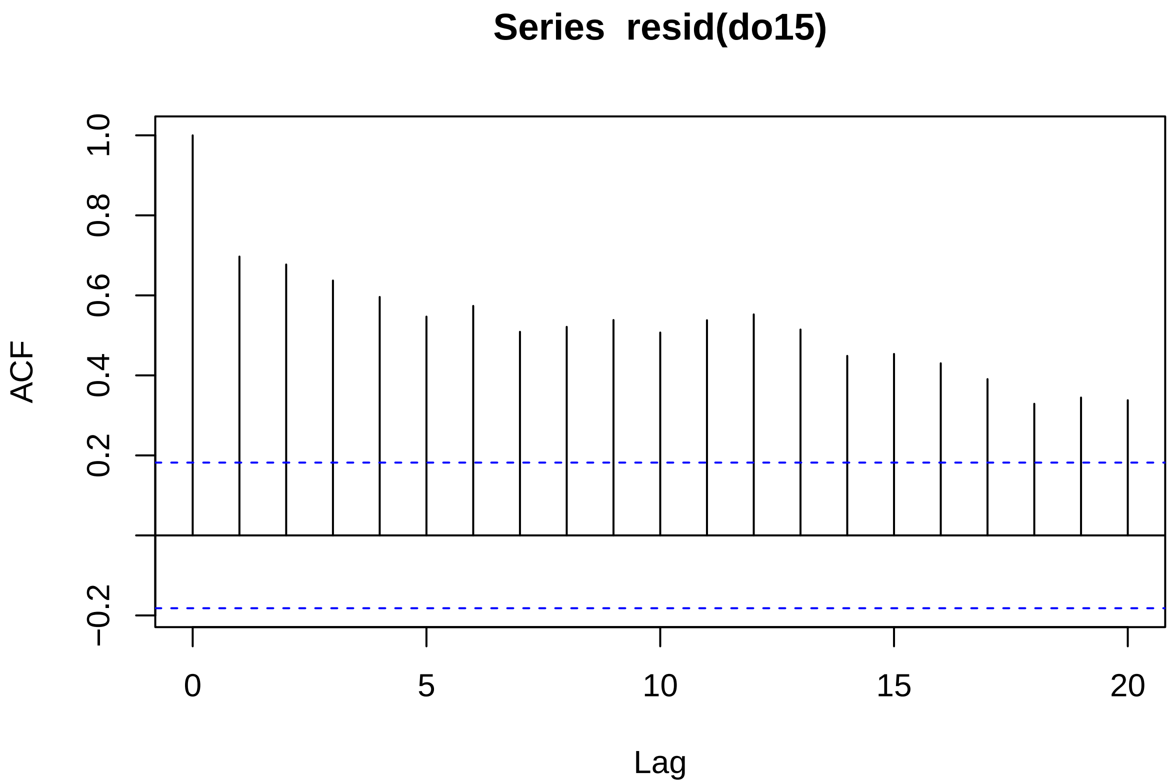
```
plot(do15)
```



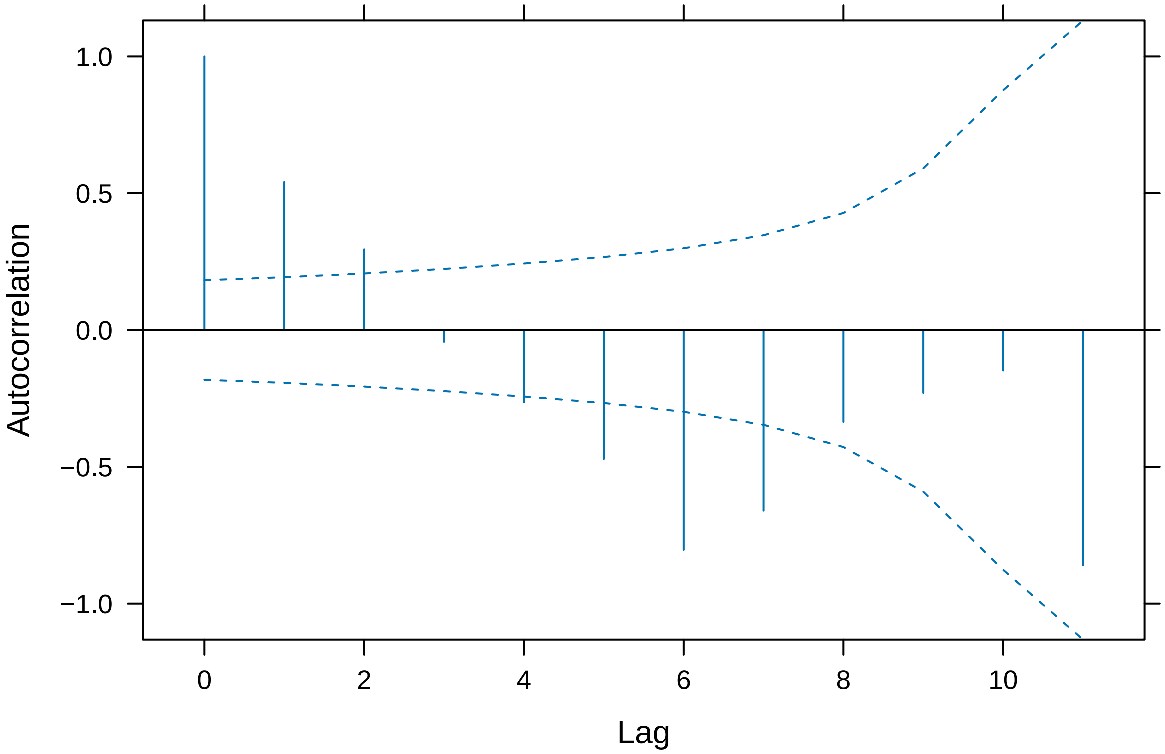
```
qqnorm(resid(do15))
```



```
acf(resid(do15))
```



```
plot(ACF(do15), alpha = 0.05)
```

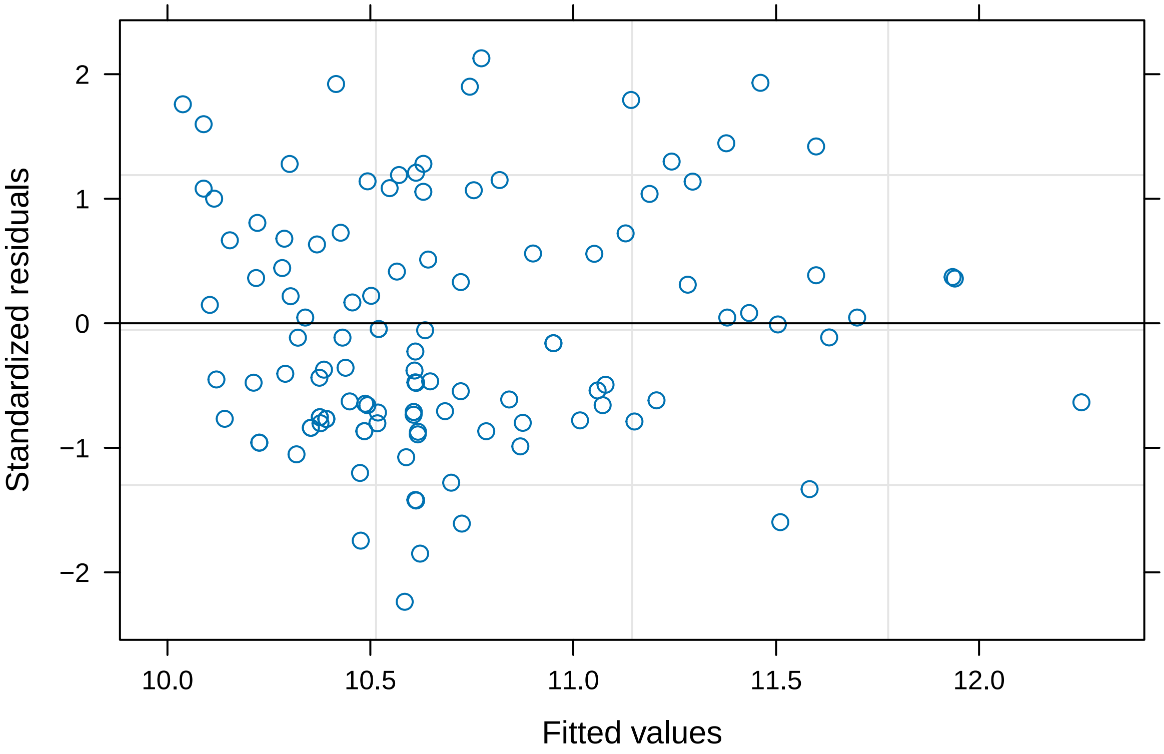


```
summary(do15_AR)
```

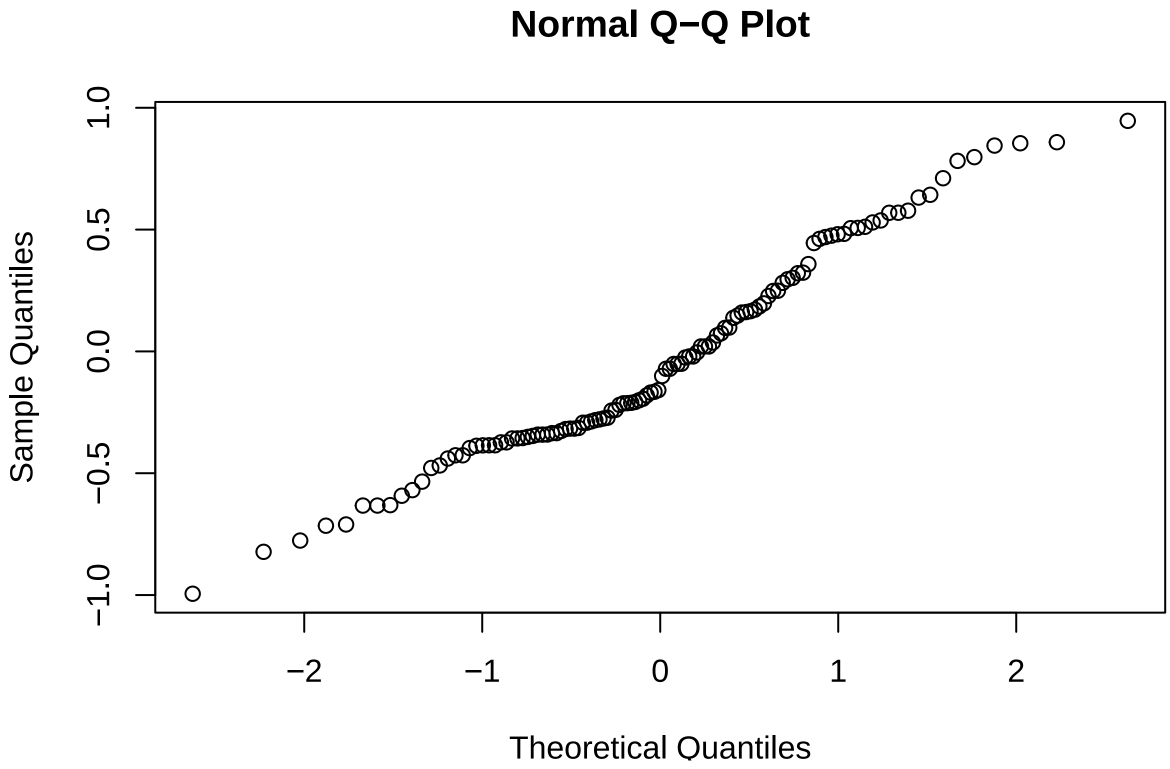
```
## Linear mixed-effects model fit by REML
##   Data: do_data
##        AIC      BIC    logLik
##   132.9079 152.0612 -59.45393
##
## Random effects:
##  Formula: ~epiVA | lake
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 1.5331676 (Intr)
## epiVA       0.1330307 -0.999
## Residual    0.4447643
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~1 | lake
##  Parameter estimate(s):
##       Phi
## 0.6518614
## Fixed effects:  DO_mean ~ scale(epiVA)
##                   Value  Std.Error  DF  t-value p-value
## (Intercept)  10.545849 0.08344453 102 126.3816  0.0000
## scale(epiVA) -0.330891 0.10463630 102  -3.1623  0.0021
##  Correlation:
##              (Intr)
## scale(epiVA) 0.291
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
```

```
## -2.2363404 -0.7680932 -0.2921023  0.5782474  2.1277427
##
## Number of Observations: 116
## Number of Groups: 13
```

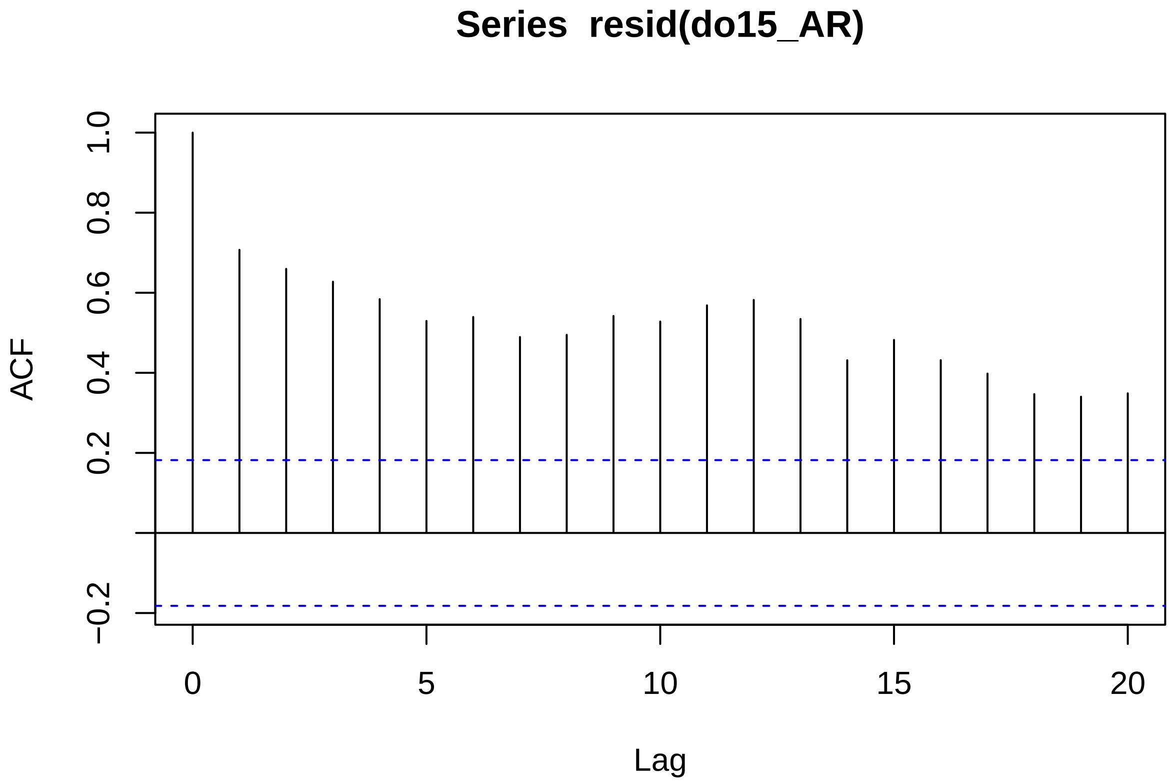
```
plot(do15_AR)
```



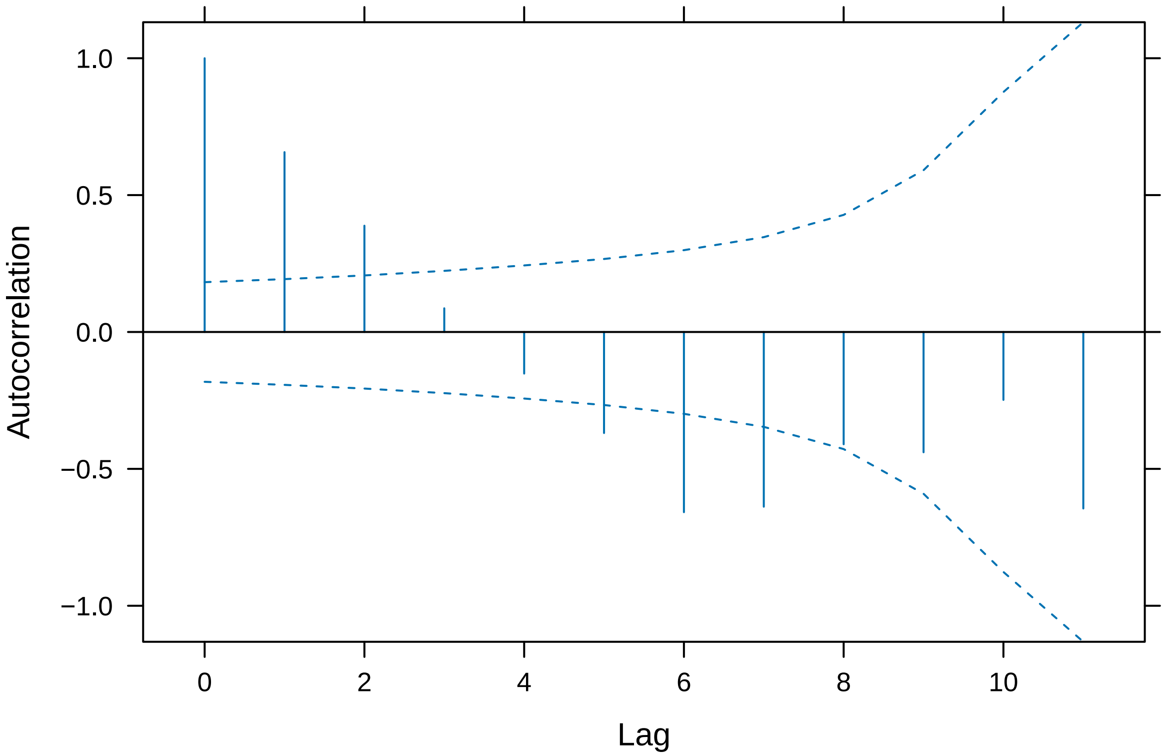
```
qqnorm(resid(do15_AR))
```



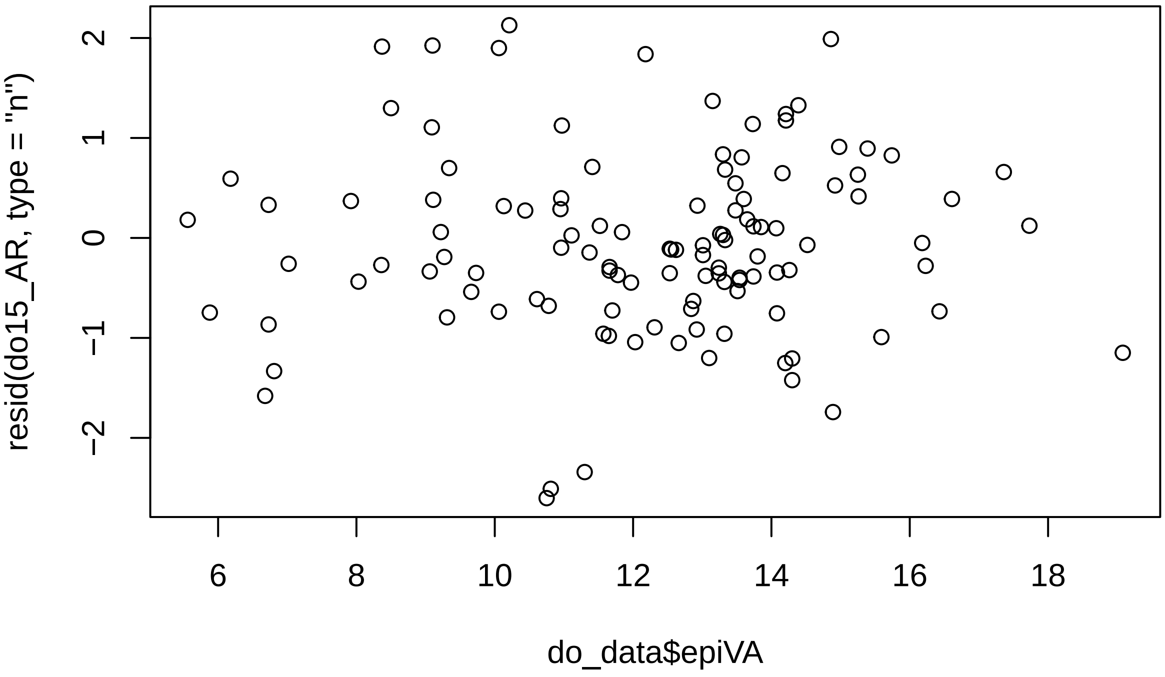
```
acf(resid(do15_AR))
```



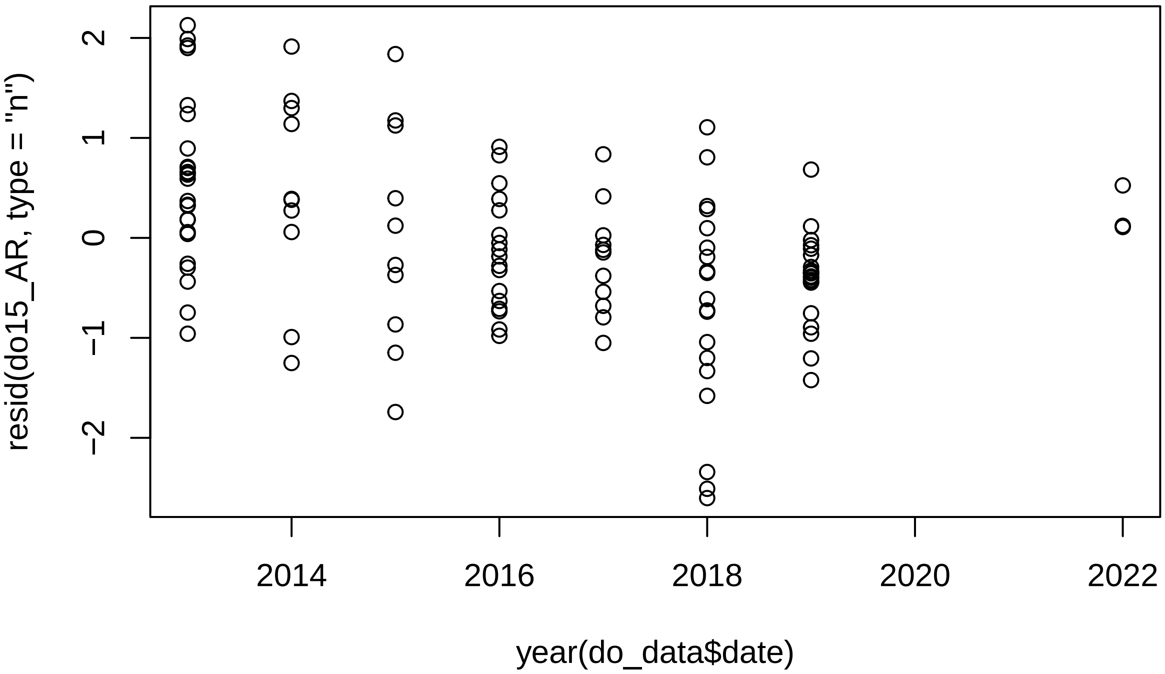
```
plot(ACF(do15_AR), alpha = 0.05)
```



```
plot(resid(do15_AR, type = 'n') ~
       do_data$epiVA)
```

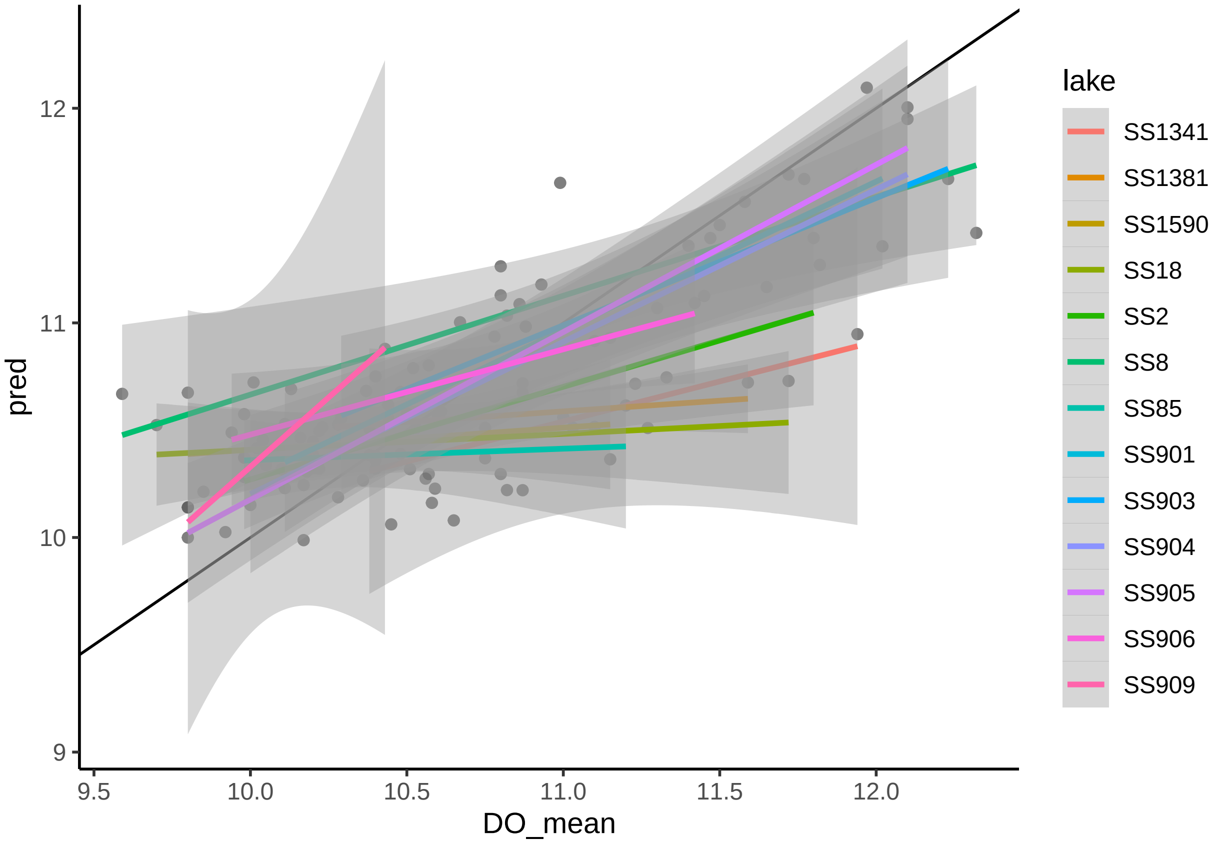


```
plot(resid(do15_AR, type = 'n') ~
       year(do_data$date))
```



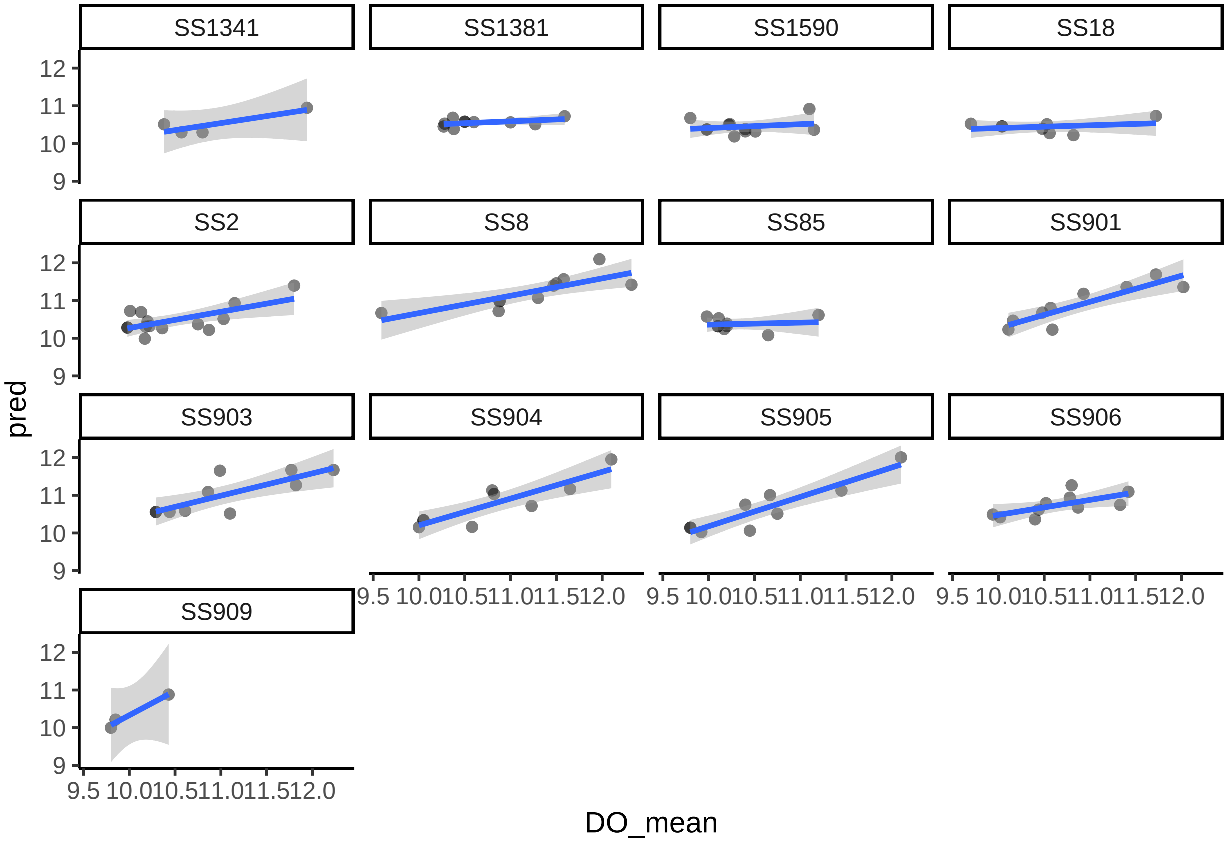
```
x <-
  do_data %>%
  filter(month(date) %in% c(6,7) & layer == 'epi') %>%
  group_by(lake) %>%
  summarise(doc_mean = mean(DOC, na.rm = T),
            do_mean = mean(DO_mean, na.rm = T),
            cond_mean = mean(cond_mean, na.rm = T))

do_data_plot <-
  do_data %>%
  filter(!is.na(DO_mean)) %>%
  cbind(.,
        pred = predict(do15_AR)) %>%
  left_join(x,
            by = 'lake')


do_data %>%
  filter(!is.na(DO_mean)) %>%
  cbind(., pred = predict(do15)) %>%
  ggplot(aes(x = DO_mean, y = pred)) +
  geom_point(alpha = 0.5)  +
  theme_classic() +
  geom_abline(slope = 1, intercept = 0) +
  geom_smooth(aes(group = lake, color = lake), method = 'lm', size = 1)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

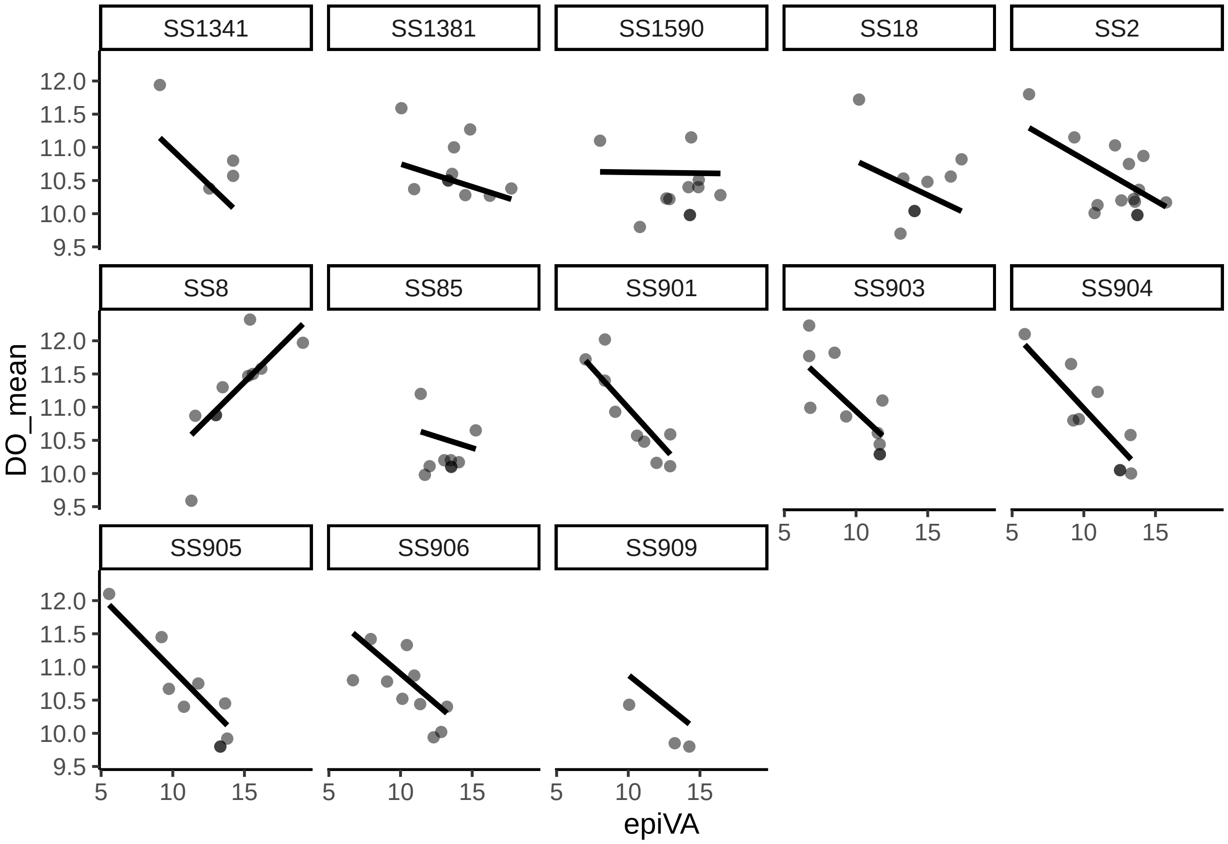


```
do_data %>%
  filter(!is.na(DO_mean)) %>%
  cbind(., pred = predict(do15)) %>%
  ggplot(aes(x = DO_mean, y = pred)) +
  geom_point(alpha = 0.5)  +
  geom_smooth(method = lm) +
  theme_classic() +
  facet_wrap(~lake)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

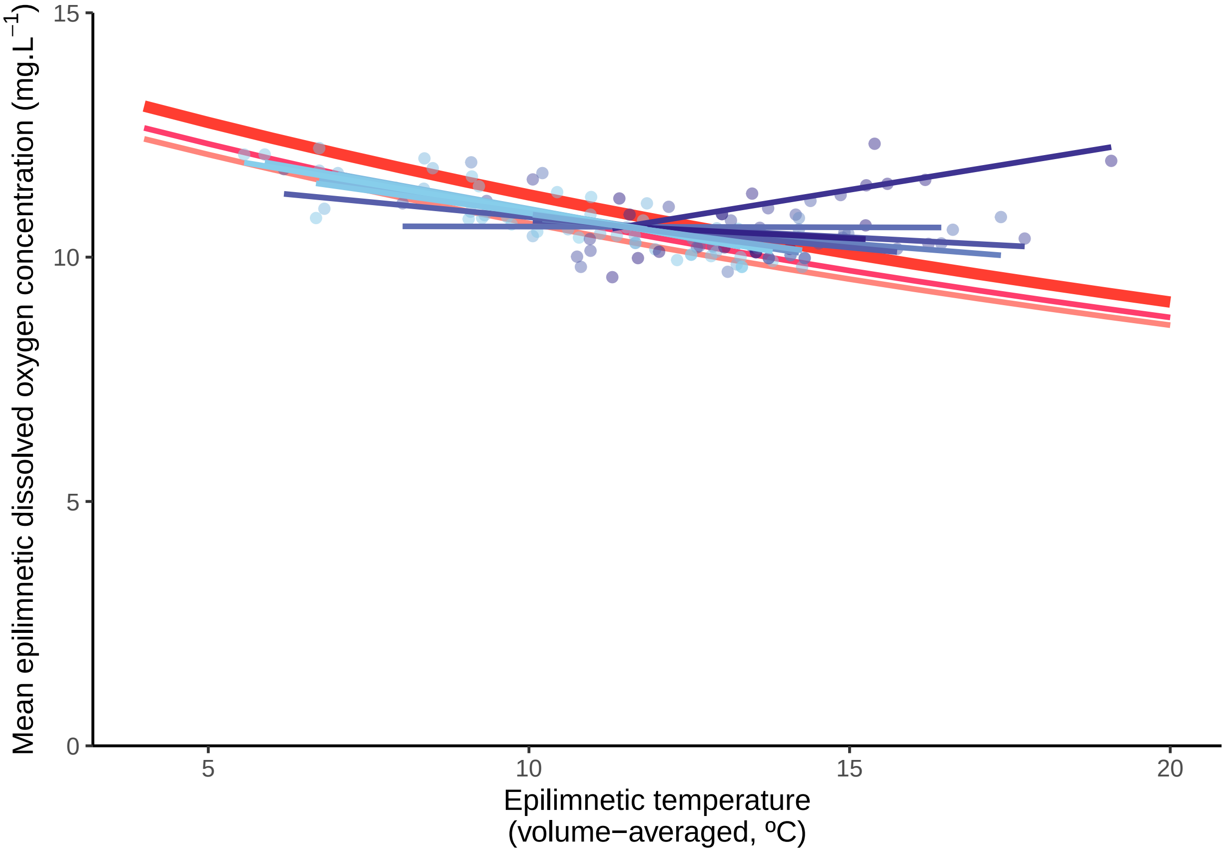


```
do_data %>%
  filter(!is.na(DO_mean)) %>%
  cbind(., pred = predict(do15_AR)) %>%
  ggplot(aes(x = epiVA, y = DO_mean)) +
  facet_wrap(~lake, ncol= 5) +
  geom_point(alpha = 0.5)  +
  theme_classic() +
  geom_line(aes(y = pred), size = 1)
```



```
epi_do_plot <-
  ggplot() +
  geom_line(data = surface, aes(y = do, x = temp), color = '#FF3D31', linetype = 1, size = 2) +
  geom_line(data = do_alt, aes(y = do_300, x = temp), color = '#FF3D6C', size = 1) +
  geom_line(data = do_alt, aes(y = do_450, x = temp), color = '#FF857C', size = 1) +
  geom_point(data = do_data_plot,
             aes(x = epiVA, y = DO_mean, color = doc_mean), alpha = 0.5)  +
  geom_line(data = do_data_plot,
            aes(x = epiVA, y = pred, group = lake, color = doc_mean), size = 1) +
  scale_x_continuous(limits = c(4,20), expand = c(0.05,0)) +
  scale_y_continuous(limits = c(0,15), expand = c(0,0)) +
  scale_colour_gradient(
    'Mean DOC concentration\n(June-July)',
    low = "skyblue",
    high = '#332287',
    guide = "colourbar",
    aesthetics = "colour") +
  labs(x = 'Epilimnetic temperature\n(volume-averaged, ºC)',
       y = bquote('Mean epilimnetic dissolved oxygen concentration (mg.L'^-1*')')) +
  theme_classic() +
  theme(legend.position = 'none')

epi_do_plot
```

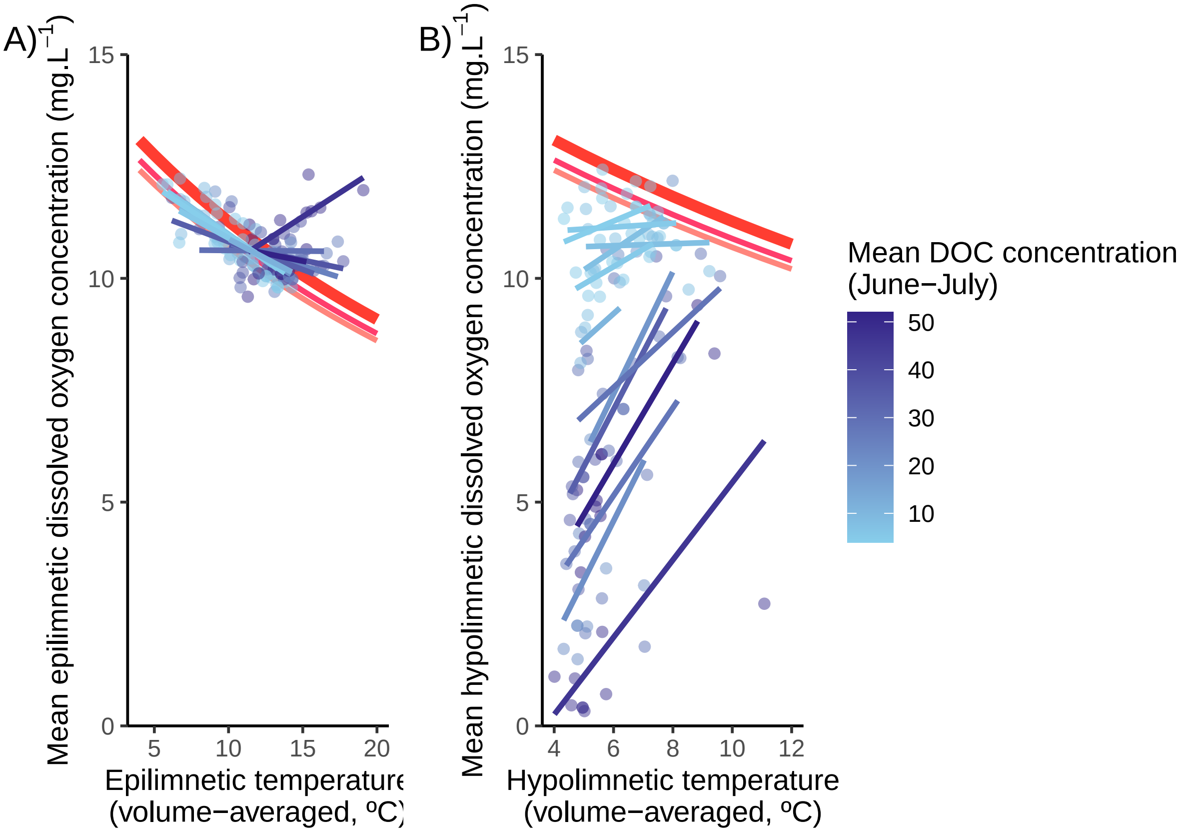
```
## Warning: Removed 4 rows containing missing values (`geom_line()`).
## Removed 4 rows containing missing values (`geom_line()`).
## Removed 4 rows containing missing values (`geom_line()`).
```



```
epi_do_plot + hypo_do_plot &
  plot_annotation(tag_levels = 'A',
                  tag_suffix = ')')
```

```
## Warning: Removed 4 rows containing missing values (`geom_line()`).
## Removed 4 rows containing missing values (`geom_line()`).
## Removed 4 rows containing missing values (`geom_line()`).
```

```
## Warning: Removed 12 rows containing missing values (`geom_line()`).
## Removed 12 rows containing missing values (`geom_line()`).
## Removed 12 rows containing missing values (`geom_line()`).
```


A)
Mean epilimnetic dissolved oxygen concentration (mg.L^{-1})
15
10
5
0
5
10
15
20
Epilimnetic temperature
(volume−averaged, ºC)
B)
Mean hypolimnetic dissolved oxygen concentration (mg.L^{-1})
15
10
5
0
4
6
8
10
12
Hypolimnetic temperature
(volume−averaged, ºC)
Mean DOC concentration
(June−July)
50
40
30
20
10