30 day rolling average analysis

```
library(ggplot2)
library(mgcv)
```

```
Loading required package: nlme
```

```
This is mgcv 1.9-1. For overview type 'help("mgcv-package")'.
```

```
library(dplyr)
```

```

Attaching package: 'dplyr'
```

```
The following object is masked from 'package:nlme':

    collapse
```

```
The following objects are masked from 'package:stats':

    filter, lag
```

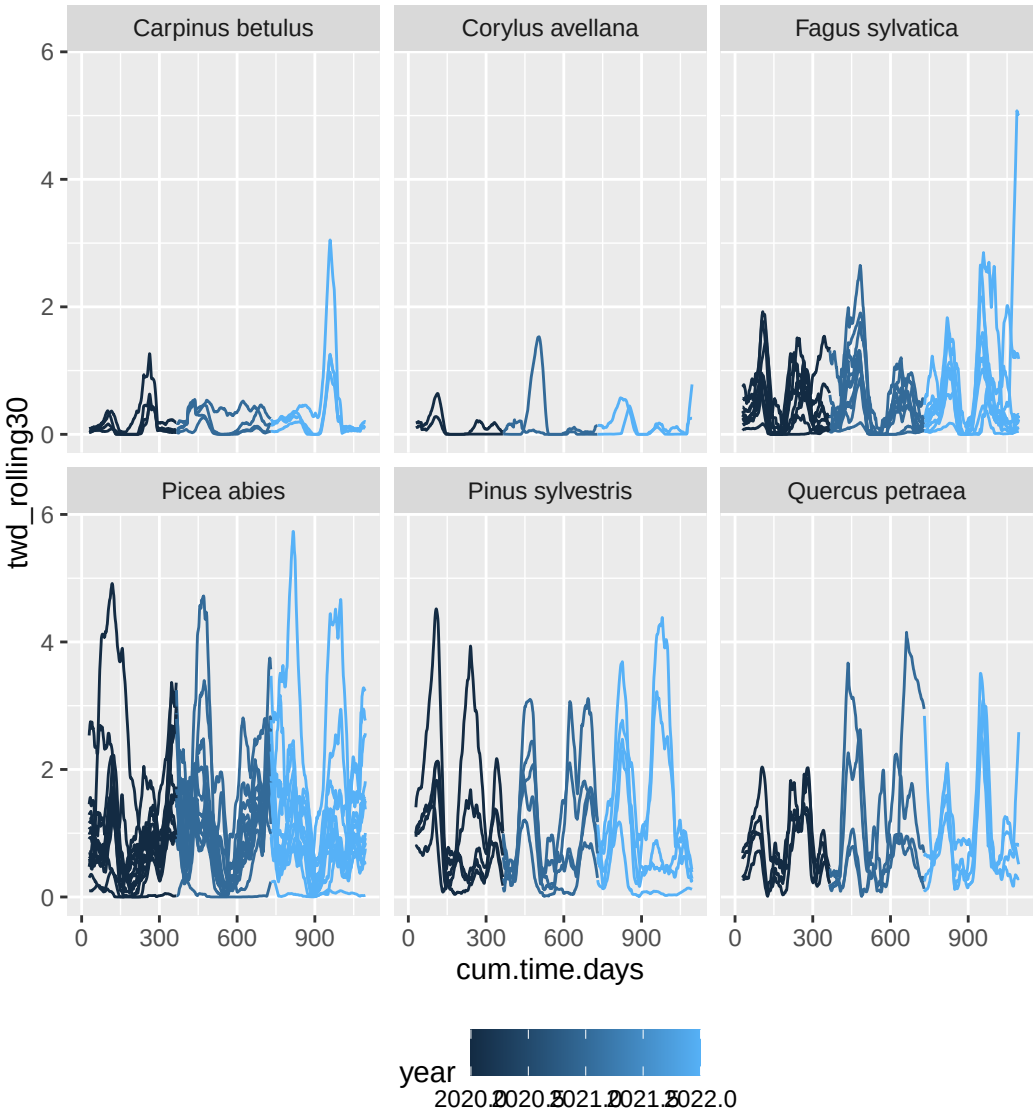
```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
df.ts.series2 <- read.csv("df.ts.series2.csv") %>%
  filter(!species == "Pseudotsuga menziesii")
num.cols <- c("twd", "pr", "at", "ws", "dp", "sr", "lr", "day.of.year")
```

```r
rolling_avg_cols <- colnames(df.ts.series2)[grep("_rolling30", colnames(df.ts.series2))]

rolling_avg_cols <- rolling_avg_cols[!rolling_avg_cols =="day.of.year_rolling30"]

print_time_ycol <- function(ycol){
  ggplot(
    df.ts.series2,
    aes(x = cum.time.days, color = year, group = interaction(year, tree.id))) +
    geom_line(aes_string(y = ycol)) +
    #geom_point(aes(color = year), size = 1)+
    facet_wrap(~species) +
    theme(legend.position='bottom')
}
```

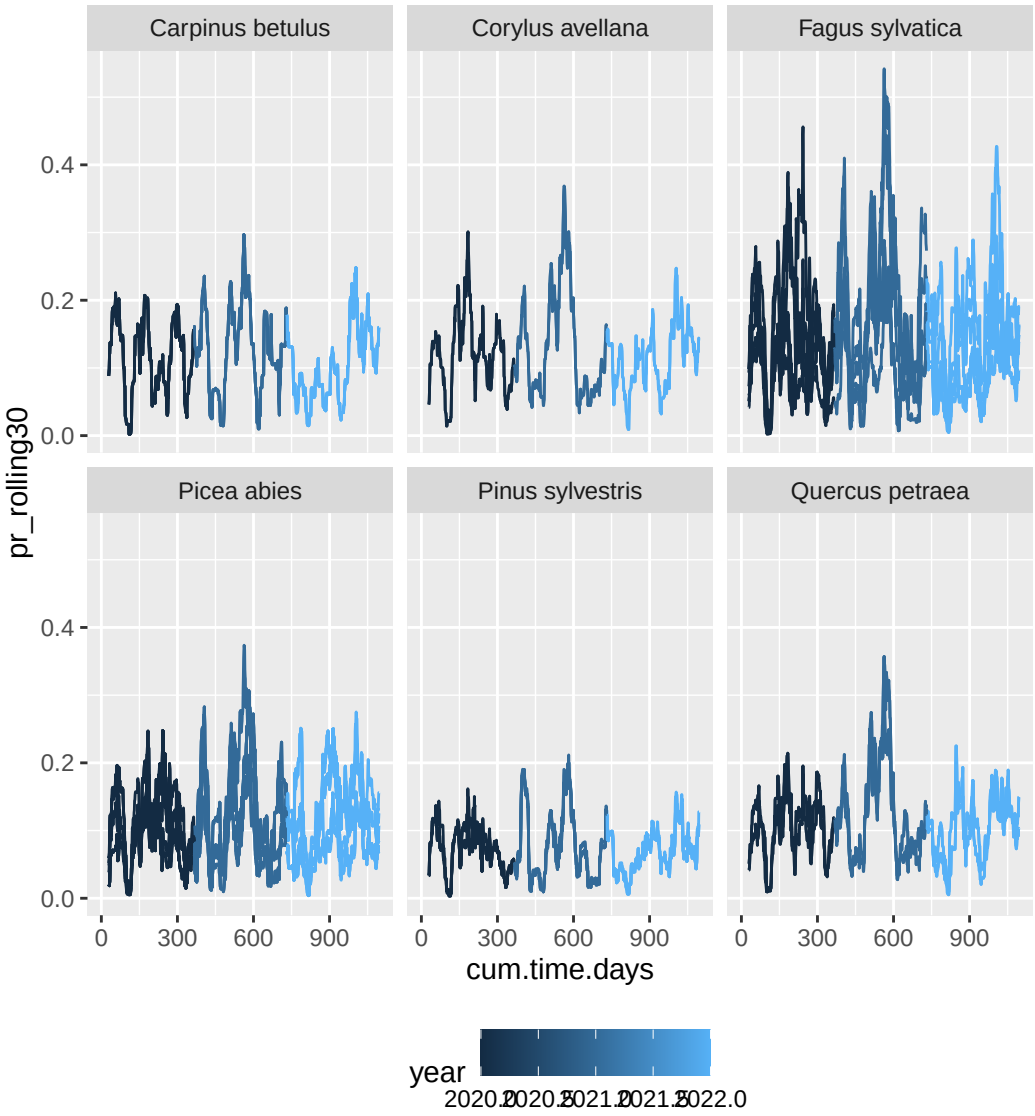
```r
lapply(rolling_avg_cols, print_time_ycol)
```

```
Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
i Please use tidy evaluation idioms with `aes()`.
i See also `vignette("ggplot2-in-packages")` for more information.

[[1]]

Warning: Removed 1160 rows containing missing values or values outside the scale range
(`geom_line()`).
```



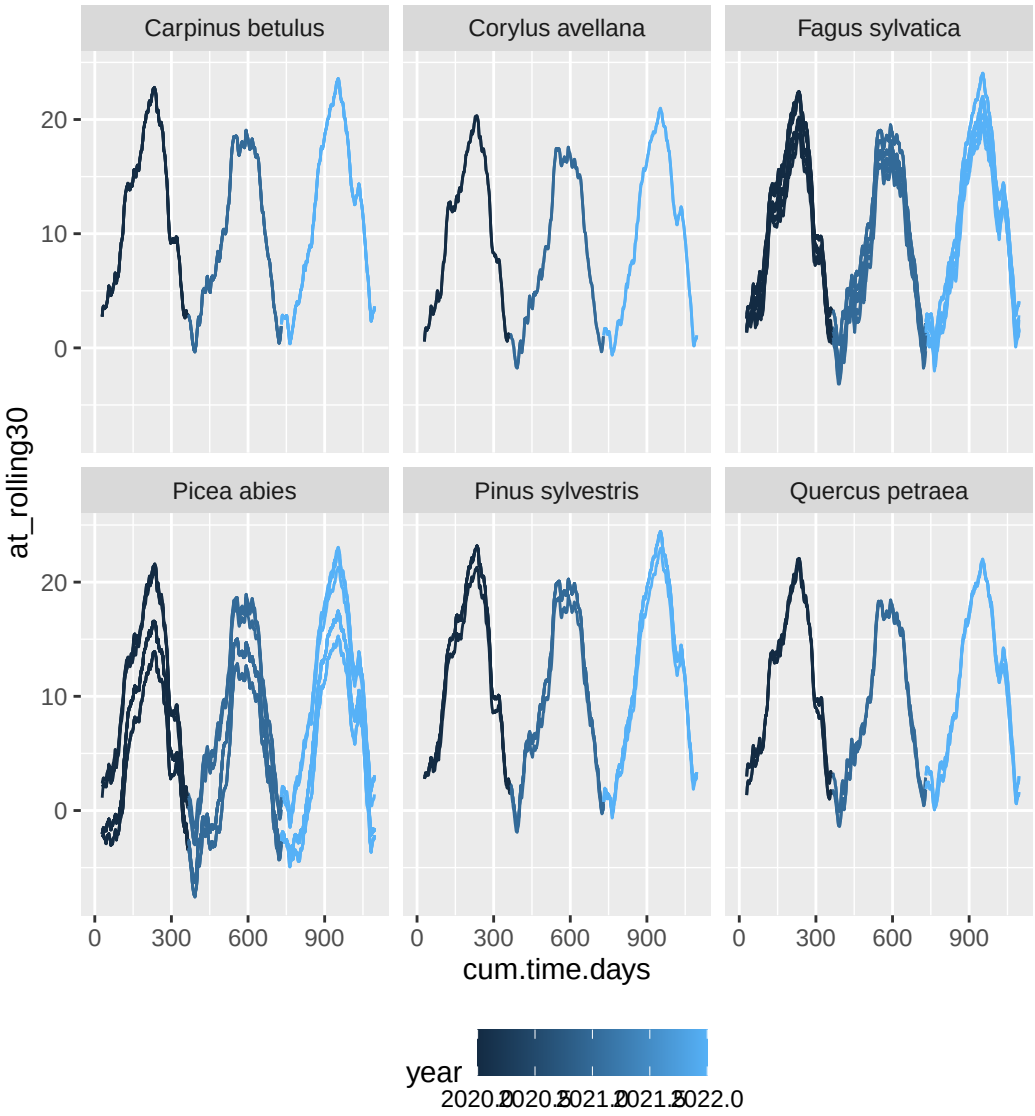

```
[[2]]
```

```
Warning: Removed 1160 rows containing missing values or values outside the scale range
(`geom_line()`).
```



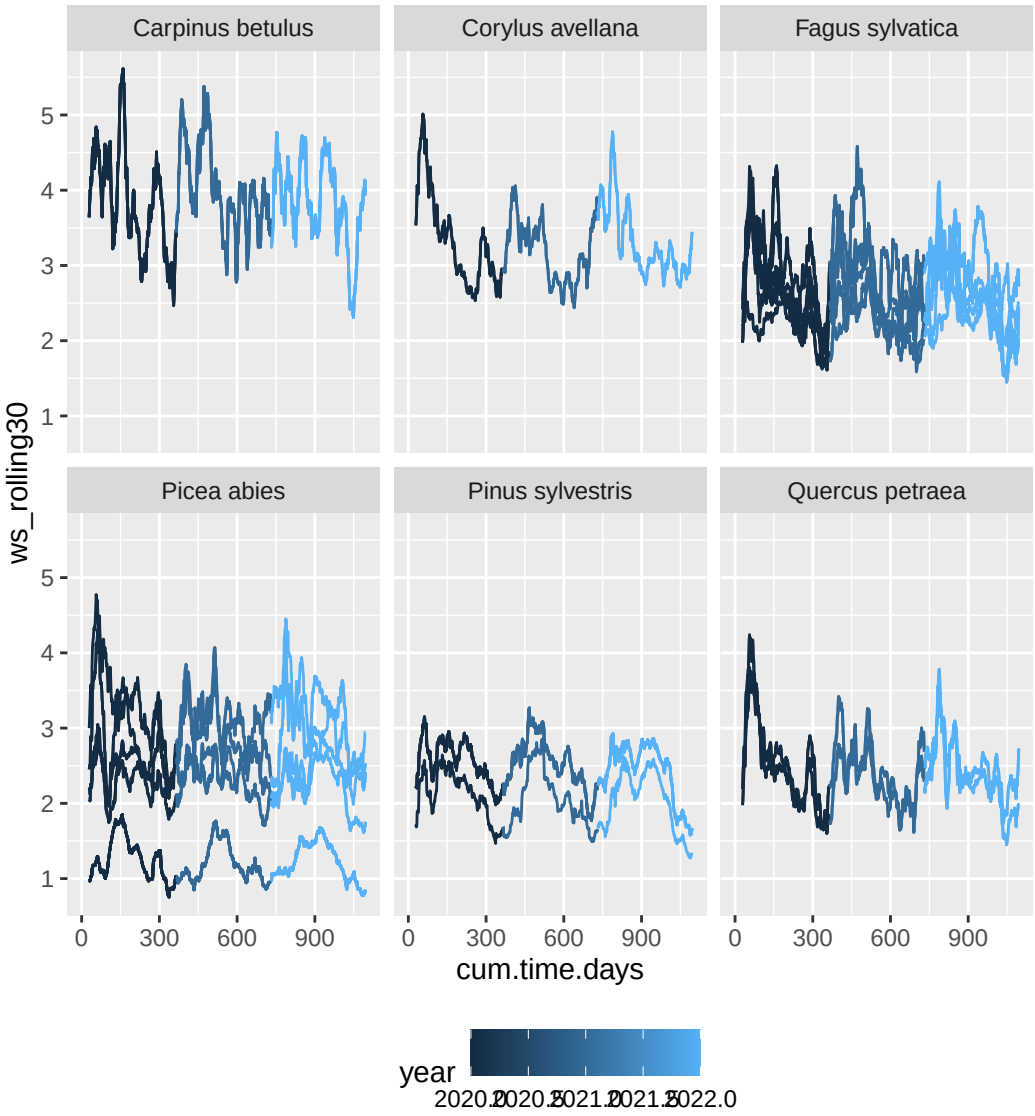

```
[[3]]
```

```
Warning: Removed 1160 rows containing missing values or values outside the scale range
(`geom_line()`).
```



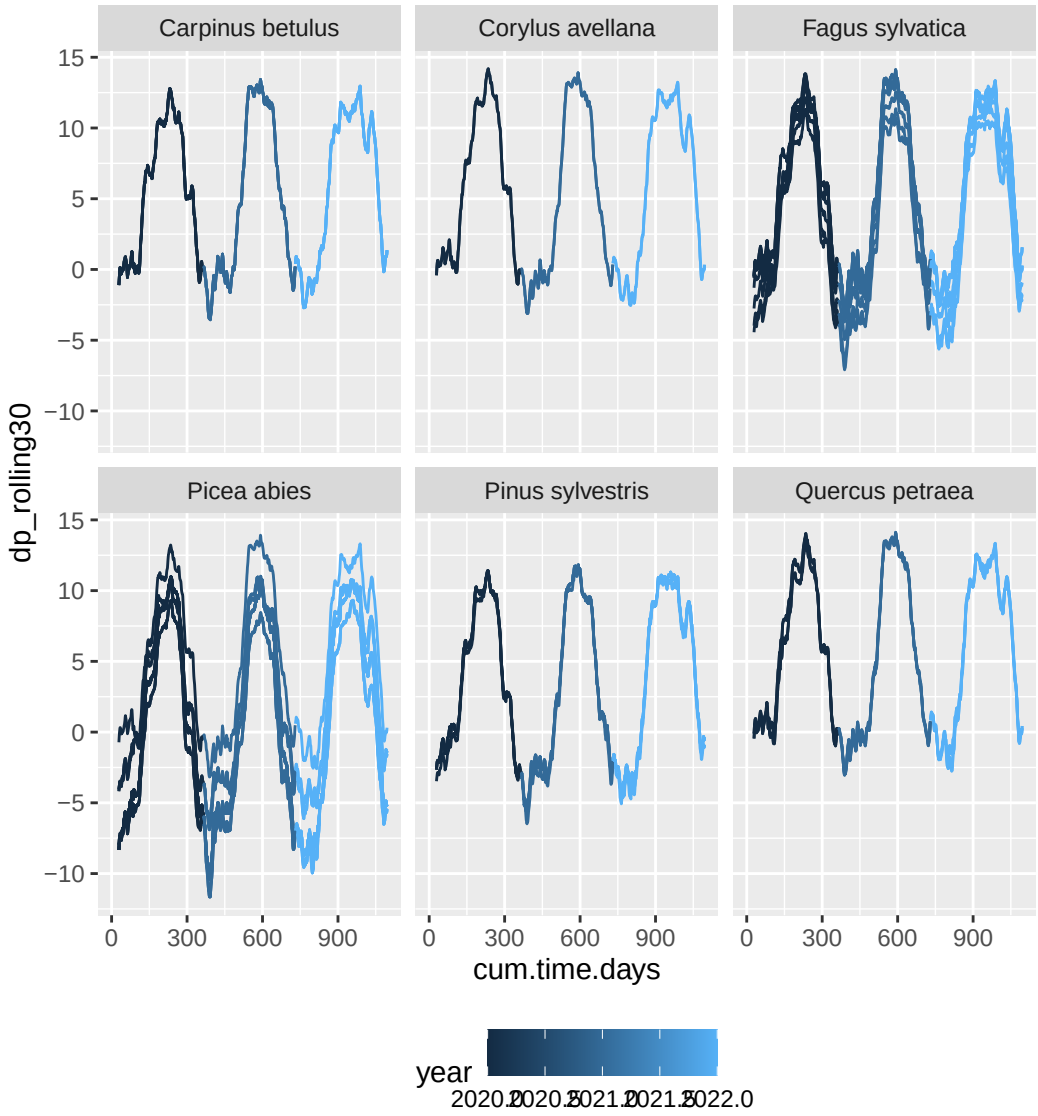

```
[[4]]
```

```
Warning: Removed 1160 rows containing missing values or values outside the scale range
(`geom_line()`).
```



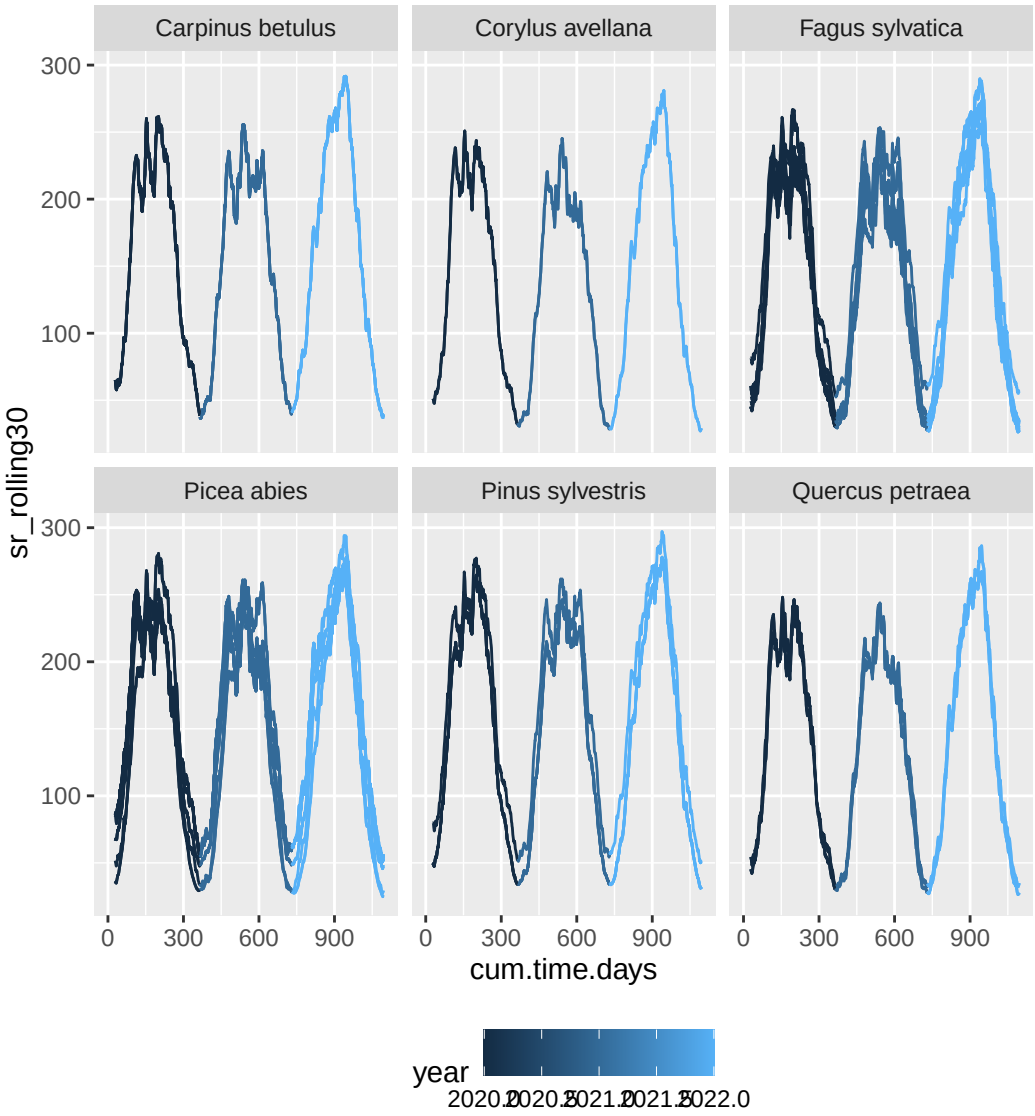

```
[[5]]
```

```
Warning: Removed 1160 rows containing missing values or values outside the scale range
(`geom_line()`).
```



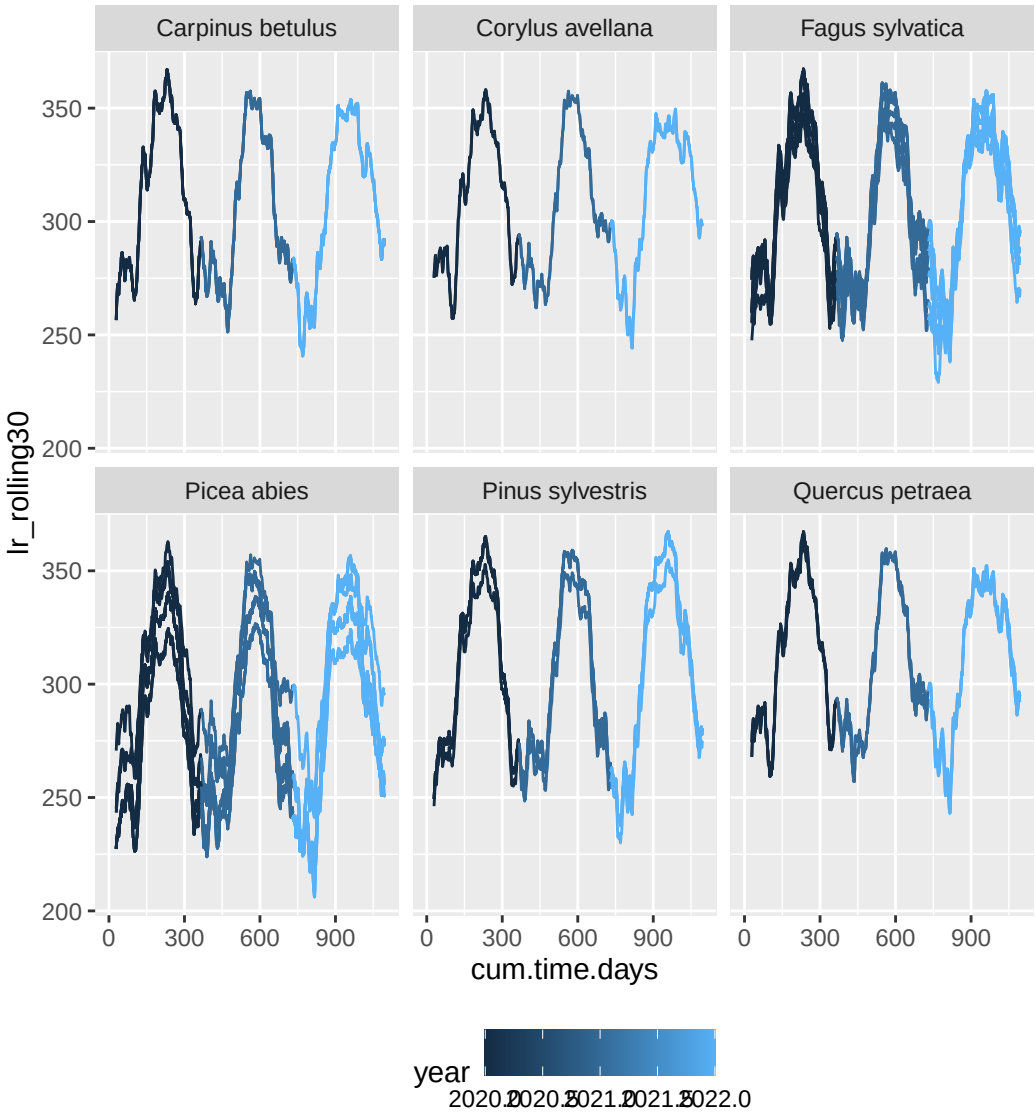

```
[[6]]
```

```
Warning: Removed 1160 rows containing missing values or values outside the scale range
(`geom_line()`).
```




```
[[7]]
```

```
Warning: Removed 1160 rows containing missing values or values outside the scale range
(`geom_line()`).
```




```
rolling_avg_cols
```

```
[1] "twd_rolling30" "pr_rolling30"  "at_rolling30"  "ws_rolling30"
[5] "dp_rolling30"  "sr_rolling30"  "lr_rolling30"
```

```
df.ts.series2$year <- factor(df.ts.series2$year)
gam_model <- gam(twd_rolling30 ~ s(cum.time.days) +  s(pr_rolling30) + s(at_rolling30) + s(ws
```

Partial effect plots of fitted Generalized Additive Model (GAM)

```
anova(gam_model)
```

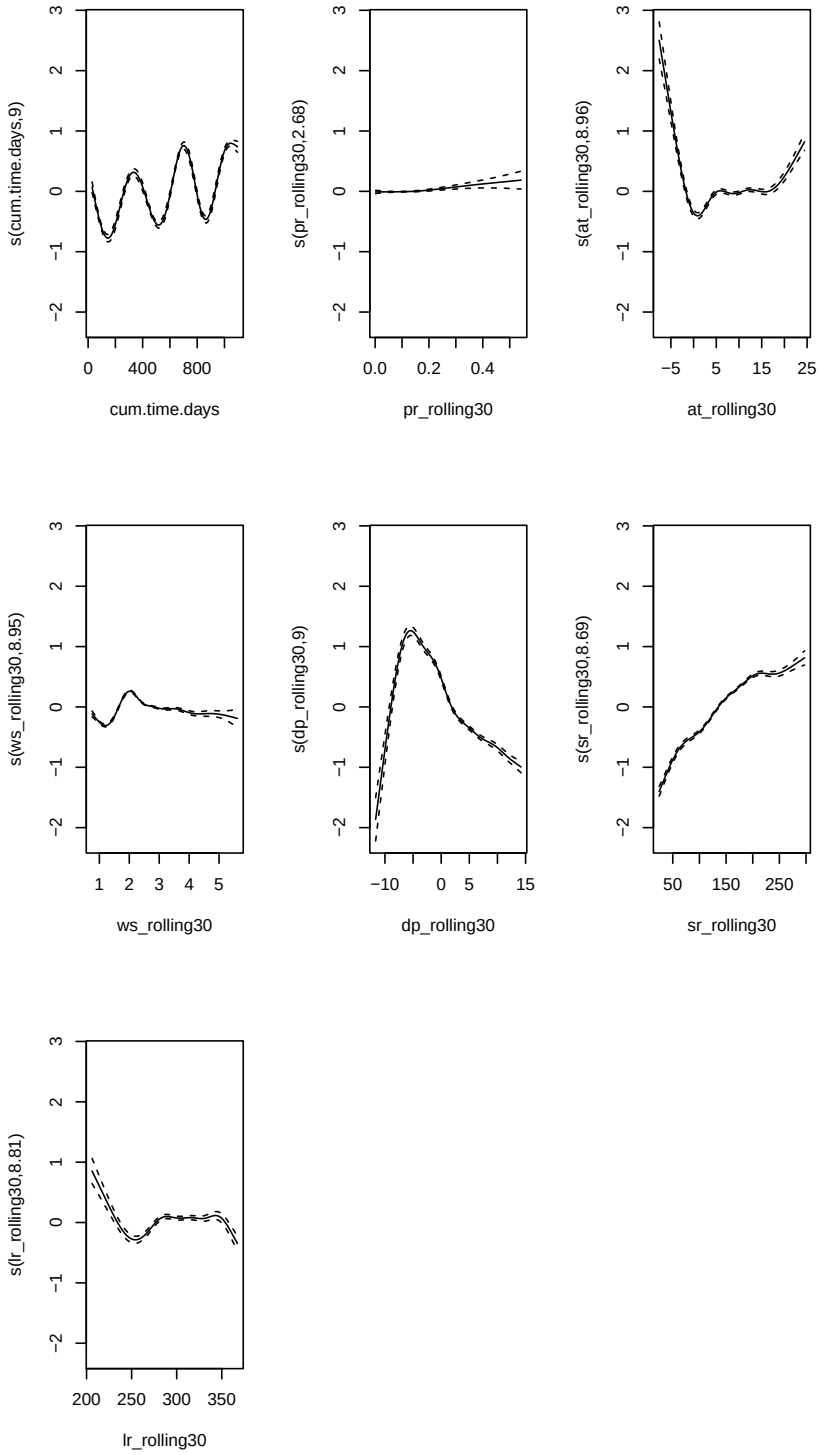
```
Family: gaussian
Link function: identity

Formula:
twd_rolling30 ~ s(cum.time.days) + s(pr_rolling30) + s(at_rolling30) +
    s(ws_rolling30) + s(dp_rolling30) + s(sr_rolling30) + s(lr_rolling30) +
    year + species

Parametric Terms:
        df      F p-value
year     2  36.77  <2e-16
species  5 694.84  <2e-16

Approximate significance of smooth terms:
                   edf Ref.df       F p-value
s(cum.time.days) 9.000  9.000 448.935 < 2e-16
s(pr_rolling30)  2.684  3.470   5.703 0.00038
s(at_rolling30)  8.958  8.999 120.196 < 2e-16
s(ws_rolling30)  8.946  8.999 174.700 < 2e-16
s(dp_rolling30)  8.999  9.000 188.053 < 2e-16
s(sr_rolling30)  8.690  8.973 231.097 < 2e-16
s(lr_rolling30)  8.809  8.991  65.148 < 2e-16
```

```
summary(gam_model)
```

```
Family: gaussian
Link function: identity

Formula:
twd_rolling30 ~ s(cum.time.days) + s(pr_rolling30) + s(at_rolling30) +
    s(ws_rolling30) + s(dp_rolling30) + s(sr_rolling30) + s(lr_rolling30) +
    year + species

Parametric coefficients:
```

```
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.60984    0.02517  24.233  < 2e-16 ***
year2021                 -0.23798    0.02982  -7.981 1.49e-15 ***
year2022                 -0.33800    0.04195  -8.056 8.06e-16 ***
speciesCorylus avellana   0.03687    0.01712   2.154   0.0313 *
speciesFagus sylvatica    0.16766    0.01464  11.455  < 2e-16 ***
speciesPicea abies        0.38443    0.01636  23.503  < 2e-16 ***
speciesPinus sylvestris   0.52730    0.01809  29.145  < 2e-16 ***
speciesQuercus petraea    0.70036    0.01685  41.568  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                   edf Ref.df       F p-value
s(cum.time.days) 9.000  9.000 448.935 < 2e-16 ***
s(pr_rolling30)  2.684  3.470   5.703 0.00038 ***
s(at_rolling30)  8.958  8.999 120.196 < 2e-16 ***
s(ws_rolling30)  8.946  8.999 174.700 < 2e-16 ***
s(dp_rolling30)  8.999  9.000 188.053 < 2e-16 ***
s(sr_rolling30)  8.690  8.973 231.097 < 2e-16 ***
s(lr_rolling30)  8.809  8.991  65.148 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.44   Deviance explained = 44.1%
GCV = 0.31729  Scale est. = 0.31682   n = 42680
```

```r
plot(gam_model, pages = 1, se = TRUE)
```


s(cum.time.days,9)
cum.time.days
s(pr_rolling30,2.68)
pr_rolling30
s(at_rolling30,8.96)
at_rolling30
s(ws_rolling30,8.95)
ws_rolling30
s(dp_rolling30,9)
dp_rolling30
s(sr_rolling30,8.69)
sr_rolling30
s(lr_rolling30,8.81)
lr_rolling30

General structure of each plot:

- X-axis: The values of the predictor variable (e.g., pr_rolling30, at_rolling30, etc.).
- Y-axis: The smooth effect f(x) of the predictor on the response. This is not the raw coefficient, but the estimated non-linear transformation.
- Black line: Estimated smooth function.
- Dashed lines: Confidence intervals (usually 95%).

Individual variable interpretations:

pr_rolling30 (precipitation?): Effect is fairly flat → very little to no influence on the response in this range.

at_rolling30 (air temperature?): U-shaped effect → lower values are associated with higher outcomes, middle values suppress it, and higher values raise it again.

ws_rolling30 (wind speed?): Slightly declining influence, but very flat → weak or no effect overall.

dp_rolling30 (dew point?): Positive effect when very low, then peaks and slightly drops → might indicate an optimal dew point range.

sr_rolling30 (solar radiation?): Flat and near-zero → suggests solar radiation has little effect on the response.

lr_rolling30 (longwave radiation?): Slight decline → higher values may slightly decrease the response.