SWP Overview and Goal

The Size-Weighted Proximity (SWP) metric is intended to quantify how strongly each voxel in the parenchyma is influenced by nearby EPVS, accounting for both distance and EPVS size. The question we are looking at is:
How do optical properties change as a function of log(SWP)?

Our goal within this analysis is not only to visualize the relationship, but also to quantify its shape, assess it strength, and determine whether the relationship appear linear or nonlinear. This helps evaluate whether changes in local tissue microstructure (reflected in optical properties) can be predicted from EPVS proximity and size

Begin SWP Relationship Analysis

Here are two figures based on what was presented in the intake meeting. It's clear that there is a non-linear patterns at higher log(SWP) values. Differences between frontal and occipital regions, suggesting a regional effect. In order look further into it we would need to fit a non-linear model. This because using any linear model would underestimate the relationship where it bends upward or downward and provides poor fit. Thus we needed a method that does *not* assume linearity, can adapt to the shape of the data, and still yields interpretable effects and significance tests.

```
file <- "op_vs_swp_16-Oct-2025.xlsx"

sheets <- excel_sheets(file)

swp <- map_dfr(sheets, function(s) {
  read_excel(file, sheet = s) %>%
    rename(log_swp = 1, optical_value = 2) %>%  # rename first two columns consistently
    mutate(sheet_name = s)
})

# Label property + region
swp <- swp %>%
  mutate(
    property = case_when(
      str_detect(sheet_name, "mus") ~ "scattering",
      str_detect(sheet_name, "ret") ~ "retardance"
    ),
    region = case_when(
      str_detect(sheet_name, "front") ~ "frontal",
      str_detect(sheet_name, "occip") ~ "occipital",
      TRUE ~ "combined"
    )
  )

glimpse(swp)
```

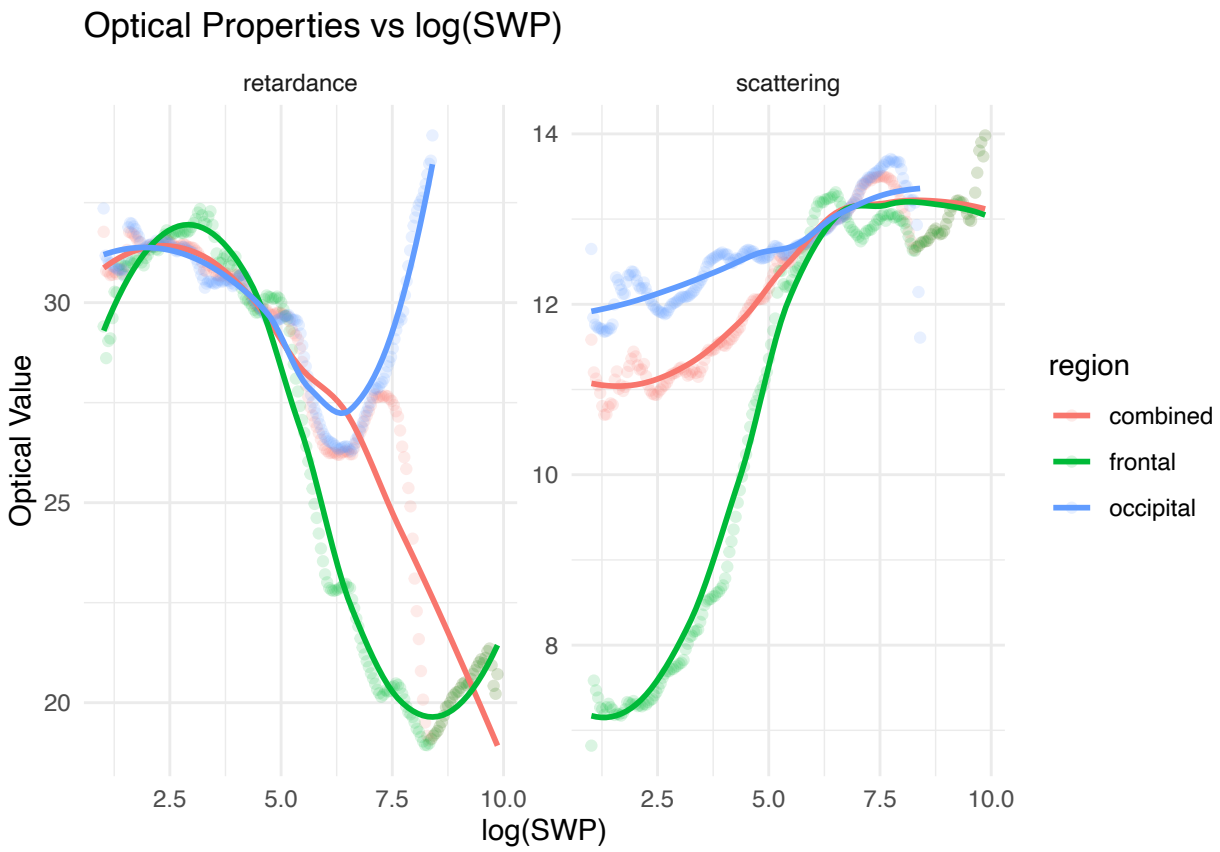
```
## Rows: 1,140
## Columns: 5
## $ log_swp       <dbl> 1.013489, 1.069199, 1.116735, 1.163813, 1.210787, 1.2577~
## $ optical_value <dbl> 11.58480, 11.20085, 11.12777, 11.04941, 10.96003, 10.805~
## $ sheet_name    <chr> "comb_mus", "comb_mus", "comb_mus", "comb_mus", "comb_mu~
## $ property      <chr> "scattering", "scattering", "scattering", "scattering", ~
## $ region        <chr> "combined", "combined", "combined", "combined", "combine~
```

```
table(swp$property, swp$region)
```

```
##
##              combined frontal occipital
##   retardance      190     190       190
##   scattering      190     190       190
```

```
ggplot(swp, aes(x = log_swp, y = optical_value, color = region)) +
  geom_point(alpha = 0.15) +
  geom_smooth(method = "loess", se = FALSE) +
  facet_wrap(~ property, scales = "free_y") +
  theme_minimal() +
  labs(title = "Optical Properties vs log(SWP)", y = "Optical Value", x = "log(SWP)")
```

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



A Generalized Additive Model (GAM) is a type of statistical model that allows the relationship between a predictor and an outcome to be non-linear, instead of forcing it to be a straight line.

In a normal linear regression, we assume:

- Every unit increase in X changes the outcome by the same amount.
- The relationship is assumed to be a straight line.

But biological processes often aren't linear — especially tissue changes in disease.

A GAM replaces the straight line with a smooth curve:

We used a Generalized Additive Model because we expected the relationship between size-weighted proximity (SWP) and the optical properties to be non-linear.

A GAM allows us to model this relationship flexibly, without forcing it into a straight line or predefined curve.
This lets the data determine the true shape of how tissue optical properties change as proximity to EPVS increases, while still adjusting for differences between brain regions.

```
gam_scatter <- gam(optical_value ~ s(log_swp, k = 6) + region,
                   data = filter(swp, property == "scattering"))

gam_retard <- gam(optical_value ~ s(log_swp, k = 6) + region,
                  data = filter(swp, property == "retardance"))

summary(gam_scatter)
```

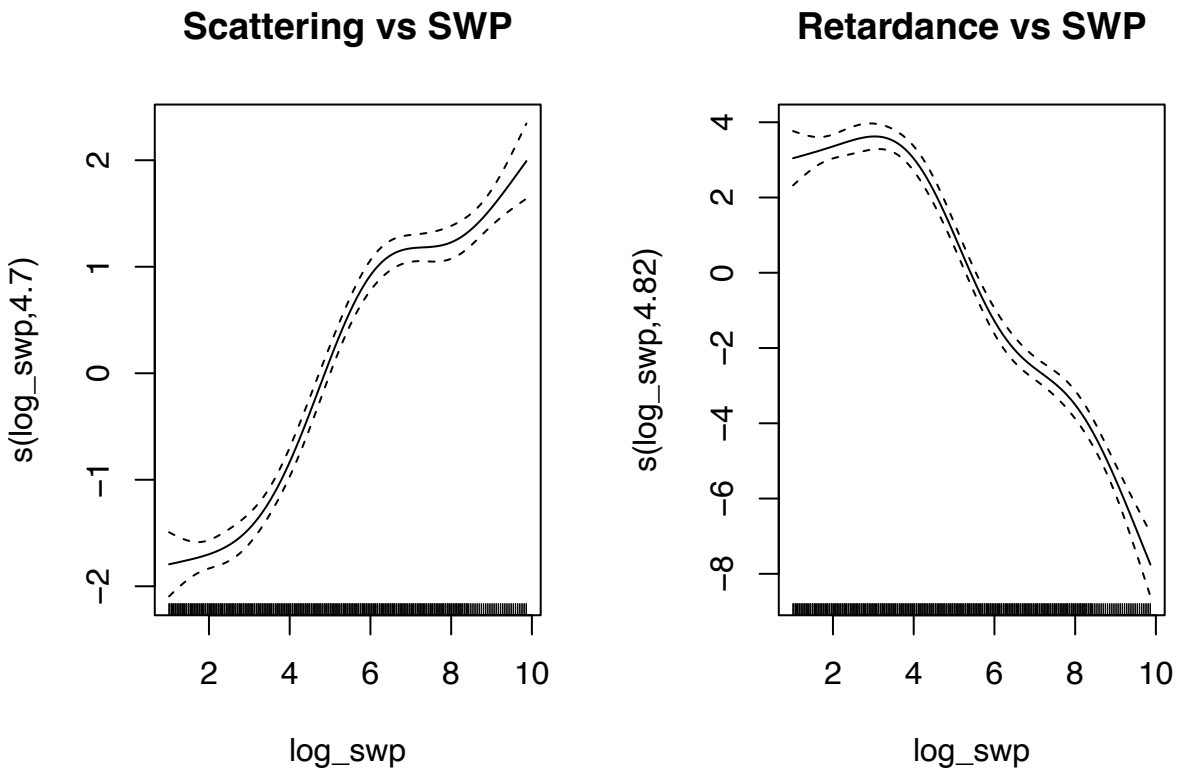
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## optical_value ~ s(log_swp, k = 6) + region
##
## Parametric coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     12.16170    0.06327 192.207  < 2e-16 ***
## regionfrontal   -1.43294    0.08908 -16.085  < 2e-16 ***
## regionoccipital  0.66085    0.09086   7.273 1.19e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df     F p-value
## s(log_swp) 4.699  4.955 243.9  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.744   Deviance explained = 74.7%
## GCV = 0.76424  Scale est. = 0.75392   n = 570
```

```
summary(gam_retard)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## optical_value ~ s(log_swp, k = 6) + region
##
## Parametric coefficients:
```

```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      27.7548     0.1498 185.310  < 2e-16 ***
## regionfrontal    -1.3834     0.2109  -6.560 1.22e-10 ***
## regionoccipital   1.3557     0.2151   6.303 5.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df   F p-value
## s(log_swp) 4.816  4.983 294  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =   0.76   Deviance explained = 76.3%
## GCV = 4.2828  Scale est. = 4.2241    n = 570
```

```
par(mfrow=c(1,2))
plot(gam_scatter, shade = TRUE, main = "Scattering vs SWP")
plot(gam_retard, shade = TRUE, main = "Retardance vs SWP")
```



We modeled scattering and retardance as smooth functions of SWP and included region as a fixed effect. In both cases, the smooth term was highly significant, confirming a non-linear relationship between SWP and tissue optical properties. Scattering increases and then decreases as SWP increases, consistent with an early increase in extra cellular tissue density followed by late-stage tissue degradation. Retardance decreases and then increases, which suggests initial myelin breakdown, followed by fiber compression or re-packing in more advanced disease. Additionally, both scattering and retardance are higher in the occipital region than in the frontal region, consistent with the known pattern of CAA-related white matter vulnerability. The models is a good fit with the adjusted r-squred value (0.74 and 0.76)

I visualized the relationship between EPVS Size-Weighted Proximity (SWP) and optical properties.

Both scattering and retardance show non-linear dependence on SWP, so linear models are not appropriate.

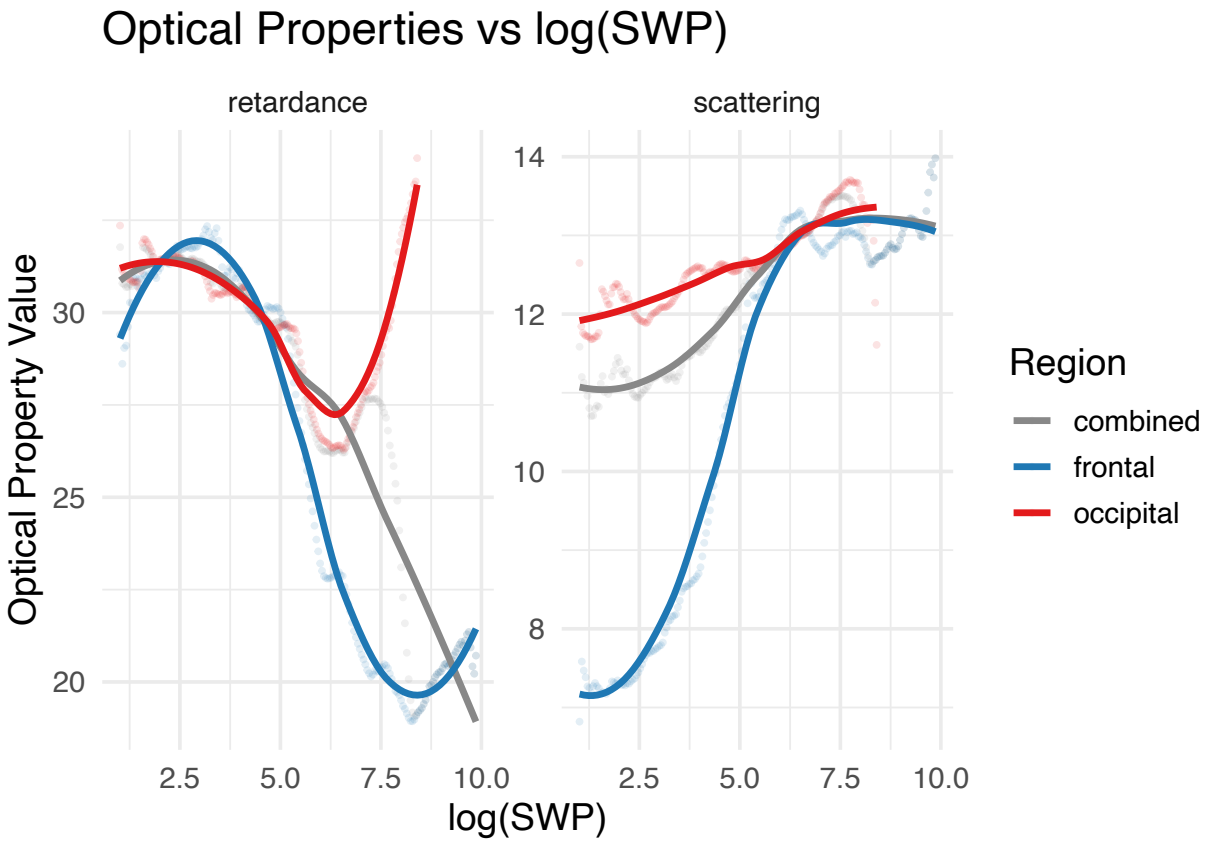
A GAM model captures the shape of the curve, which aligns with the biological interpretation:

- Scattering increases at moderate SWP, likely due to increased cellular density or ECM changes — then drops at high SWP, consistent with late-stage tissue breakdown.
- Retardance decreases initially, due to myelin degeneration — then rises again, suggesting myelin fiber compression in late disease progression.
- Regional effects (frontal vs occipital) will be tested next, but occipital is expected to show stronger pathology based on previous findings.

Below is a more advanced visualization

```
p1 <- ggplot(swp, aes(x = log_swp, y = optical_value, color = region)) +
  geom_point(alpha = 0.12, size = 0.7) +
  geom_smooth(method = "loess", se = FALSE, linewidth = 1.2) +
  facet_wrap(~ property, scales = "free_y") +
  scale_color_manual(values = c("combined"="#888888", "frontal"="#1f78b4", "occipital"="#e31a1c")) +
  theme_minimal(base_size = 14) +
  labs(title = "Optical Properties vs log(SWP)",
       x = "log(SWP)",
       y = "Optical Property Value",
       color = "Region")
p1
```

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



```r
# Predict for smooth plotting
newdata <- swp %>%
  group_by(region) %>%
  reframe(log_swp = seq(min(log_swp), max(log_swp), length.out = 200), .groups="drop")

newdata$scatter_pred <- predict(gam_scatter, newdata)
newdata$ret_pred <- predict(gam_retard, newdata)
p2 <- ggplot(newdata, aes(x = log_swp, color = region)) +
  geom_line(aes(y = scatter_pred), linewidth = 1.4) +
  geom_line(aes(y = ret_pred), linetype="dashed", linewidth=1.4) +
  scale_color_manual(values = c("combined"="#888888", "frontal"="#1f78b4", "occipital"="#e31a1c")) +
  theme_minimal(base_size = 14) +
  labs(title = "GAM Smooths: Scattering (solid) & Retardance (dashed)",
       x = "log(SWP)",
       y = "Optical Property Value",
       color = "Region")

p2
```