

# analysis

Paul Melloy

2025-12-05

```
calc_das <- function(d, days, Sd = "2022-03-31"){
  if(missing(days)){
    as.POSIXct(d, format = "%Y-%m-%d") - as.POSIXct(Sd, format = "%Y-%m-%d")
  }else{
    if(missing(d)){
      as.POSIXct(Sd, format = "%Y-%m-%d") + (days * 60 * 60 * 24)
    }
  }
}
calc_das("2022-05-15")
```

```
## Time difference of 45 days
```

```
calc_das(days = 156)
```

```
## [1] "2022-09-03 AEST"
```

## Read in data

```
cclo <- fread("../data/2025-12-02_fulldf_cc_cercospoRa.csv")[, "data" := "canopy_closure"]
eeeo <- fread("../data/2025-12-02_fulldf_eo_cercospoRa.csv")[, "data" := "epidemic_onset"]
```

We will combine the data so we can associate observation locations with canopy closure locations.

## Inspect data

```
colnames(cclo)
```

## Check headings

```
## [1] "x"          "y"          "cc"         "treatment" "area"       "source"
## [7] "data"
```

```

colnames(eeee)

## [1] "x"          "y"          "das"        "treatment" "area"       "source"
## [7] "date"       "data"

```

We need to change the headings so they have the same names and join them.

```

setnames(cclo,
         old = c("cc", "area", "treatment"),
         new = c("das", "inoculum", "fungicide"))

setnames(eeee,
         old = c("area", "treatment"),
         new = c("inoculum", "fungicide"))

# Remove date column
eeee[, date := NULL]

```

**Rename variables** Join together the two datasets.

```
dat <- rbind(cclo, eeee)
```

```
unique(dat$fungicide)
```

**Recheck variables**

```

## [1] "treated"    "untreated"

unique(dat$inoculum)

## [1] "non-inoculated" "inoculated"

# dat[inoculum == "roi", inoculum := "non-inoculated"]
# dat[inoculum == "ino", inoculum := "inoculated"]

unique(dat$source)

## [1] "uas"          "s2"           "s2s"          "observation"

#dat[source == "uav", source := "uas"]

# Relevel fungicide factor
dat[, fungicide := factor(fungicide, levels = c("untreated", "treated"))]
dat[, inoculum := factor(inoculum, levels = c("non-inoculated", "inoculated"))]

```

## Get nearest neighbour

Associate the closest observed epidemic onset date (EO) to the UAS or S2 Earliest Estimated Epidemic Onset (EEEO) date.

```
obs_nearest <- nn2(data = dat[source == "observation", .(x,y)],
                     query = dat[,(x,y)],
                     k = 2)

range(obs_nearest$nn.idx[,1])

## [1] 1 1362

# associate RS estimated EO with closest observation
dat[, obs_das:= dat[source == "observation", das][obs_nearest$nn.idx[,1]]]
```

\*\*\*

## Canopy Closure

```
cclo <- dat[data == "canopy_closure"]
```

Do a quick check on aov of the data to inspect average differences between inoculation and fungicide treatments.

```
# check variance averages
mod <- lm(das ~ source + inoculum + fungicide, data = cclo)
summary(mod)

##
## Call:
## lm(formula = das ~ source + inoculum + fungicide, data = cclo)
##
## Residuals:
##      Min       1Q       Median       3Q      Max 
## -19.3352  -4.5153   -0.0653   4.1109  27.6648 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 93.3294    0.5189 179.845 < 2e-16 ***
## sources2s   -3.2641    0.6263 -5.211 2.35e-07 ***
## sourceuas   -14.7477   0.5808 -25.391 < 2e-16 ***
## inoculuminoculated  0.6300    0.5773   1.091    0.275  
## fungicidetreated  -3.2465    0.4797  -6.768 2.42e-11 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.044 on 858 degrees of freedom
## Multiple R-squared:  0.4823, Adjusted R-squared:  0.4799 
## F-statistic: 199.9 on 4 and 858 DF,  p-value: < 2.2e-16
```

We need to check the residuals for spatial autocorrelation.

```
library(spdep)

## Warning: package 'spdep' was built under R version 4.5.2

## Loading required package: spData

## Warning: package 'spData' was built under R version 4.5.2

## To access larger datasets in this package, install the spDataLarge
## package with: 'install.packages('spDataLarge',
## repos='https://nowosad.github.io/drat/', type='source')'

## Loading required package: sf

## Warning: package 'sf' was built under R version 4.5.2

## Linking to GEOS 3.13.1, GDAL 3.11.4, PROJ 9.7.0; sf_use_s2() is TRUE

# Create a spatial weights matrix (e.g., based on k-nearest neighbors)
coords <- cbind(cclo$x, cclo$y)
knn_weights <- knearneigh(coords, k = 2) # k=2 for small dataset
nb_weights <- knn2nb(knn_weights)

## Warning in knn2nb(knn_weights): neighbour object has 164 sub-graphs

listw <- nb2listw(nb_weights, style = "W", zero.policy = TRUE)

# Perform Moran's I test on OLS residuals
moran.test(residuals(mod), listw)

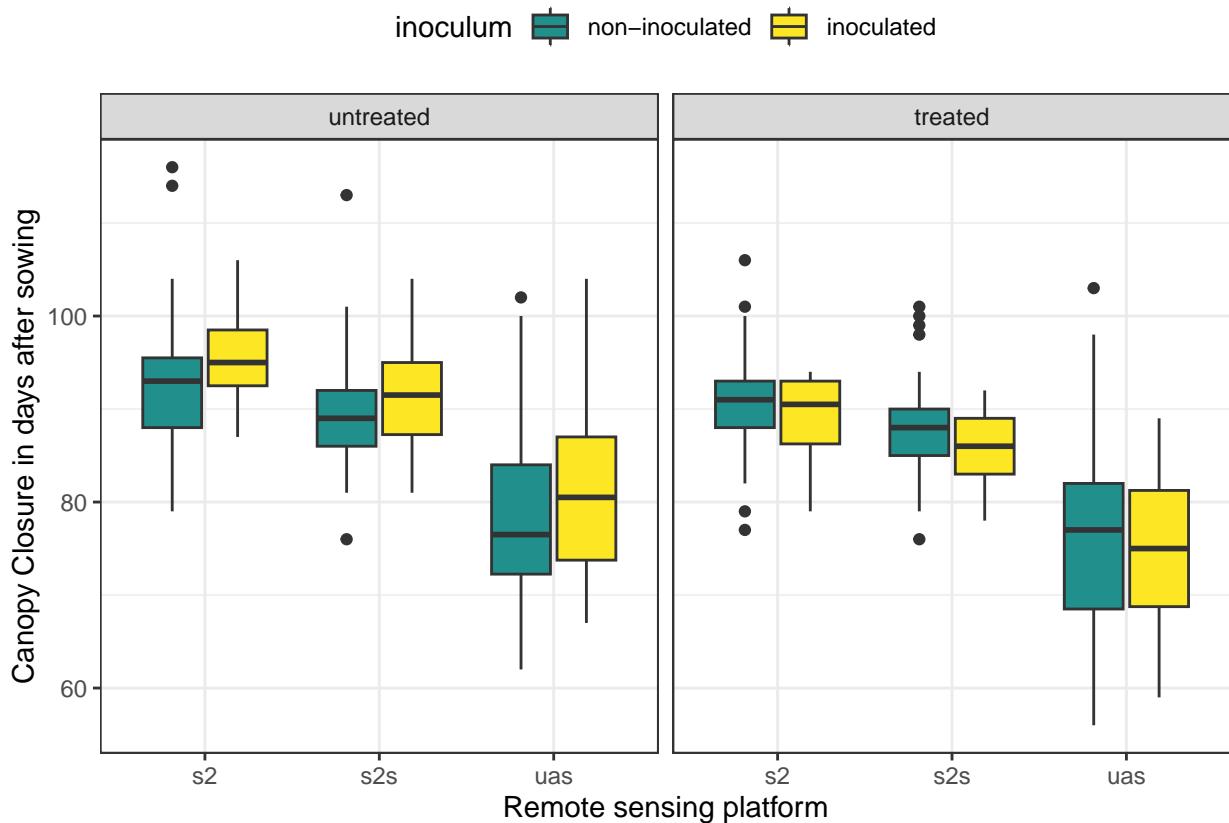
## 
## Moran I test under randomisation
## 
## data: residuals(mod)
## weights: listw
## 
## Moran I statistic standard deviate = 18.698, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##           0.598691044     -0.001160093     0.001029191
```

The moran test indicates a strong spatial auto-correlation. This indicates we should proceed with a model that can handle spatial autocorrelation.

Generalised additive models are a reasonable fit here as our dependant variable is an integer and it can handle non-linearity in the co-variates.

## Plot data

```
# plot EO data from each source an treatment
cclo |>
  ggplot(aes(source,das, fill = inoculum)) +
  geom_boxplot() +
  facet_wrap(~fungicide) +
  theme_bw() +
  labs(x = "Remote sensing platform",
       y = "Canopy Closure in days after sowing") +
  theme(legend.position = "top")+
  scale_fill_viridis_d(begin = 0.5)
```



## Model data

### Model 1

What are the average influences independent of other treatments accounting for spatial variability?

We will use Generalized Additive Models (GAM) to model the data, adding a spline to account for the spatial variance across the field.

```
# what are the average influences independant of other treatments
mod1 <- gam(das ~ s(x,y) + source + inoculum + fungicide,
            data = cclo,
            method = "REML")
summary(mod1)
```

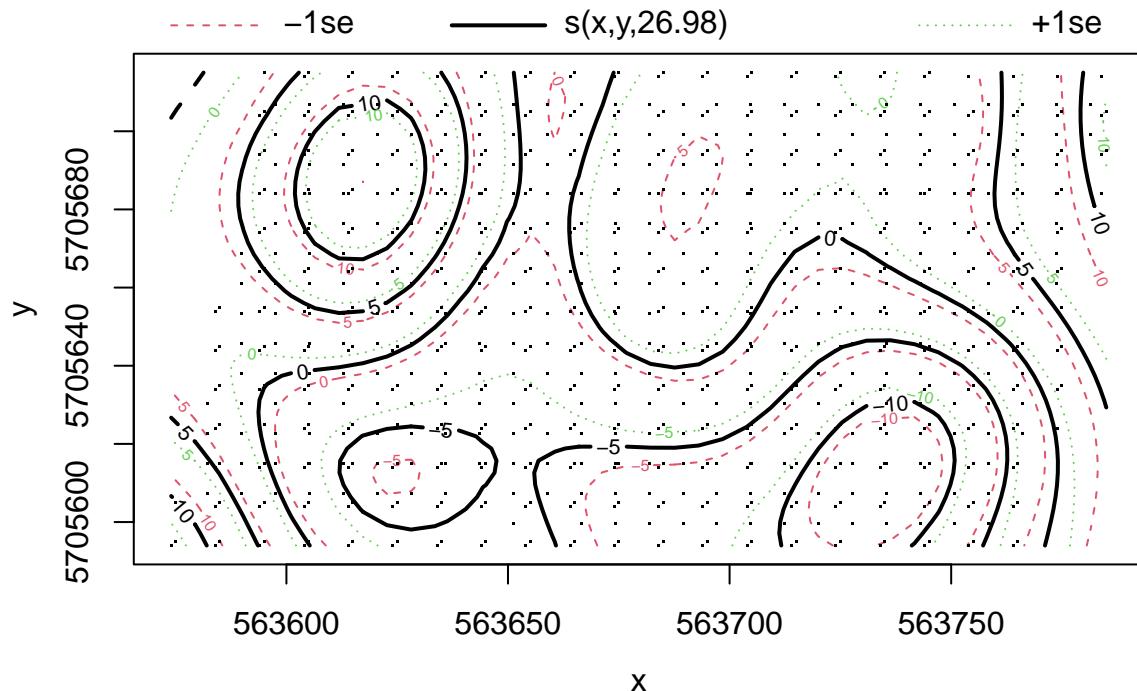
```

##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y) + source + inoculum + fungicide
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)            91.6717    0.6700 136.829 < 2e-16 ***
## sources2s           -3.3518    0.3857 -8.691 < 2e-16 ***
## sourceus          -14.6936    0.3620 -40.590 < 2e-16 ***
## inoculuminoculated   1.5140    0.5264  2.876 0.00413 **
## fungicidetreated     -0.3608    1.1857 -0.304  0.76097
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##      edf Ref.df      F p-value
## s(x,y) 26.98 28.74 48.87 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.803 Deviance explained =  81%
## -REML = 2532.5 Scale est. = 18.776 n = 863

```

Check the spline fit to ensure no over fitting.

```
plot(mod1)
```



There seems an overfitting due to the experimental setup. The spatial spline is accounting for too much of

the inoculated areas.

Lets also check if autocorrelation remains

```
moran.test(residuals(mod1), listw)
```

```
##  
## Moran I test under randomisation  
##  
## data: residuals(mod1)  
## weights: listw  
##  
## Moran I statistic standard deviate = 1.8728, p-value = 0.03054  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic      Expectation      Variance  
##          0.058913716     -0.001160093     0.001028880
```

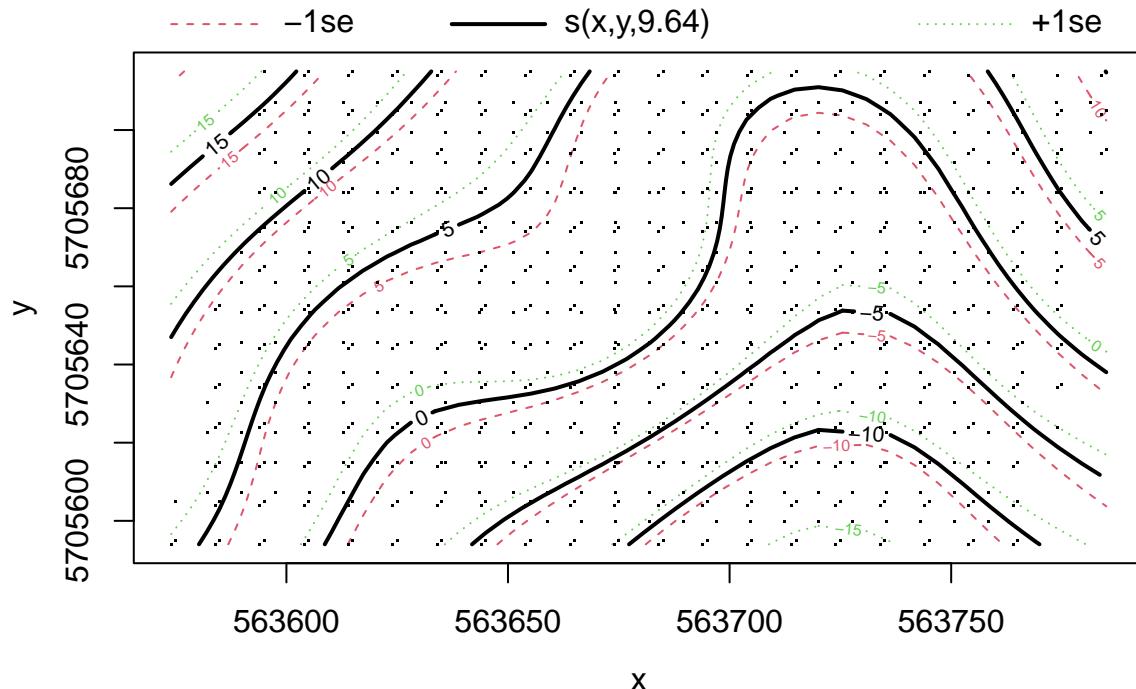
Auto correlation is much less significant, but still significant.

We will try again refitting with less knots.

```
# what are the average influences independant of other treatments  
mod1a <- gam(das ~ s(x,y, k=11) + source + inoculum + fungicide,  
               data = cclo,  
               method = "REML")
```

Check the spline fit to ensure no over fitting.

```
plot(mod1a)
```



With fewer knots we have obtained a spatial correction for the field gradient.

Differences in canopy closure dates between fungicide treated areas were significant compared to untreated for date of canopy closure.

Non-Inoculated areas showed later canopy closure, of 1.88 days compared to inoculated areas.

Remote sensing platform UAS detected canopy closure -14.75 days earlier.

## Model 2

Is there a significant interaction between measurement *source* and *inoculum*?

```
mod2 <- gam(das ~ s(x,y, k = 11) + source * inoculum + fungicide,
             data = cclo,
             method = "REML")
summary(mod2)

##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 11) + source * inoculum + fungicide
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 89.27964   0.63748 140.050 < 2e-16 ***
## sources2s                  -3.16815   0.50825  -6.234 7.19e-10 ***
## sourceuas                  -14.77743   0.47956 -30.814 < 2e-16 ***
## inoculuminoculated          2.03751   0.85775   2.375  0.0178 *
## fungicidetreated            4.15368   1.02666   4.046 5.69e-05 ***
## sources2s:inoculuminoculated -0.71831   1.12084  -0.641  0.5218
## sourceuas:inoculuminoculated  0.08692   1.01656   0.086  0.9319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df    F p-value
## s(x,y) 9.637 9.973 79.04 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.728  Deviance explained = 73.3%
## -REML = 2638.8  Scale est. = 25.926 n = 863
```

There is no significant interaction in inoculated areas and canopy closure dates from different platforms.

## Model 3

*Inoculum* and *fungicide* interaction?

```
mod3 <- gam(das ~ s(x,y, k = 11) + source + inoculum * fungicide,
             data = cclo)
summary(mod3)
```

```

## 
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 11) + source + inoculum * fungicide
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 89.7556   0.6409 140.040 < 2e-16 ***
## sources2s                  -3.2884   0.4502  -7.305 6.41e-13 ***
## sourceus                   -14.7291   0.4200 -35.070 < 2e-16 ***
## inoculuminoculated          3.6062   0.7188   5.017 6.40e-07 ***
## fungicidetreated            3.2084   1.0754   2.983  0.00293 **
## inoculuminoculated:fungicidetreated -3.3147   1.0163  -3.261  0.00115 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df F p-value
## s(x,y) 9.771 9.989 78.3 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.732 Deviance explained = 73.6%
## GCV = 26.072 Scale est. = 25.596 n = 863

```

There is a significant interaction, between fungicide and inoculum ( $P = 0$ ).  
Inoculated areas in fungicide treated plots show earlier canopy closure date.

#### Model 4

Is there a interaction between RS *source* and *fungicide*?

```

mod4 <- gam(das ~ s(x,y, k = 11) + source * fungicide + inoculum,
             data = cclo,
             method = "REML")
summary(mod4)

```

```

## 
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 11) + source * fungicide + inoculum
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 89.0070   0.6774 131.391 < 2e-16 ***
## sources2s                  -3.2234   0.6419  -5.021 6.26e-07 ***
## sourceus                   -14.0650   0.6004 -23.425 < 2e-16 ***
## fungicidetreated            4.7556   1.1567   4.111 4.32e-05 ***

```

```

## inoculuminoculated      1.8882     0.5117    3.690 0.000238 ***
## sources2s:fungicidetreated -0.1943     0.9049   -0.215 0.830069
## sourceuas:fungicidetreated -1.3391     0.8427   -1.589 0.112405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df    F p-value
## s(x,y) 9.64  9.973 79.12 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.729  Deviance explained = 73.4%
## -REML = 2638  Scale est. = 25.851 n = 863

```

```
anova(mod4)
```

```

##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 11) + source * fungicide + inoculum
##
## Parametric Terms:
##          df      F p-value
## source      2 313.438 < 2e-16
## fungicide   1  16.902 4.32e-05
## inoculum    1  13.617 0.000238
## source:fungicide 2   1.555 0.211705
##
## Approximate significance of smooth terms:
##          edf Ref.df    F p-value
## s(x,y) 9.640  9.973 79.12 <2e-16

```

## Fungicide treatment

Overall the interaction between fungicide and RS source was non significant.

**Model 5** Therefore the fungicide variable will be removed and this area treated as a field replicate.

Is there an interaction between RS platform and inoculated areas on canopy closure?

```

mod5 <- gam(das ~ s(x,y, k = 7) + source * inoculum,
             data = cclo[fungicide == "untreated"],
             method = "REML")
summary(mod5)

```

```

##
## Family: gaussian
## Link function: identity

```

```

## 
## Formula:
## das ~ s(x, y, k = 7) + source * inoculum
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 92.1153   0.5712 161.278 < 2e-16 ***
## sources2s                  -2.9590   0.7823 -3.783 0.000178 ***
## sourceus                   -13.7488   0.7435 -18.491 < 2e-16 ***
## inoculuminoculated          4.6848   1.3030   3.595 0.000363 ***
## sources2s:inoculuminoculated -1.2309   1.7023  -0.723 0.470050
## sourceus:inoculuminoculated -1.3012   1.5374  -0.846 0.397827
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df    F p-value
## s(x,y)  5.779  5.985 46.99 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.667  Deviance explained = 67.5%
## -REML = 1327.1  Scale est. = 30.277 n = 425

```

There is no canopy closure timing interaction between remote sensing platform and inoculated areas.

```

cclosure_estimates <-
  data.table(Variable = c("Intercept","Super S2","UAS","Inoculated Area","Super S2 : Inoculated","UAS : 
  round(summary(mod5)$p.table,3))[,c(1,2,3,5)] |>
  flextable() |>
  align(align = "center", part = "all") |>
  fontsize(size = 8, part = "body") |>
  fontsize(size = 10, part = "header") |>
  italic(italic = TRUE, part = "header") |>
  set_caption(
    "Table ##: Estimated canopy closure in days after sowing (DAS) and the mean
    difference to the non-inoculated sentinel-2 data (intercept).
    P values indicate statistical significance in comparison to the intercept."
  ) |>
  autofit()

cclosure_estimates

```

## results table

```

## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.

```

Table 1: Table ##: Estimated canopy closure in days after sowing (DAS) and the mean difference to the non-inoculated sentinel-2 data (intercept). P values indicate statistical significance in comparison to the intercept.

Variable	Estimate	Std. Error	Pr(> t )
Intercept	92.115	0.571	0.000
Super S2	-2.959	0.782	0.000
UAS	-13.749	0.744	0.000
Inoculated Area	4.685	1.303	0.000
Super S2 : Inoculated	-1.231	1.702	0.470
UAS : Inoculated	-1.301	1.537	0.398

```
mod6 <- gam(das ~ s(x,y, k = 7) + source + inoculum,
             data = cclo[fungicide == "untreated"],
             method = "REML")
summary(mod6)
```

## Model 6

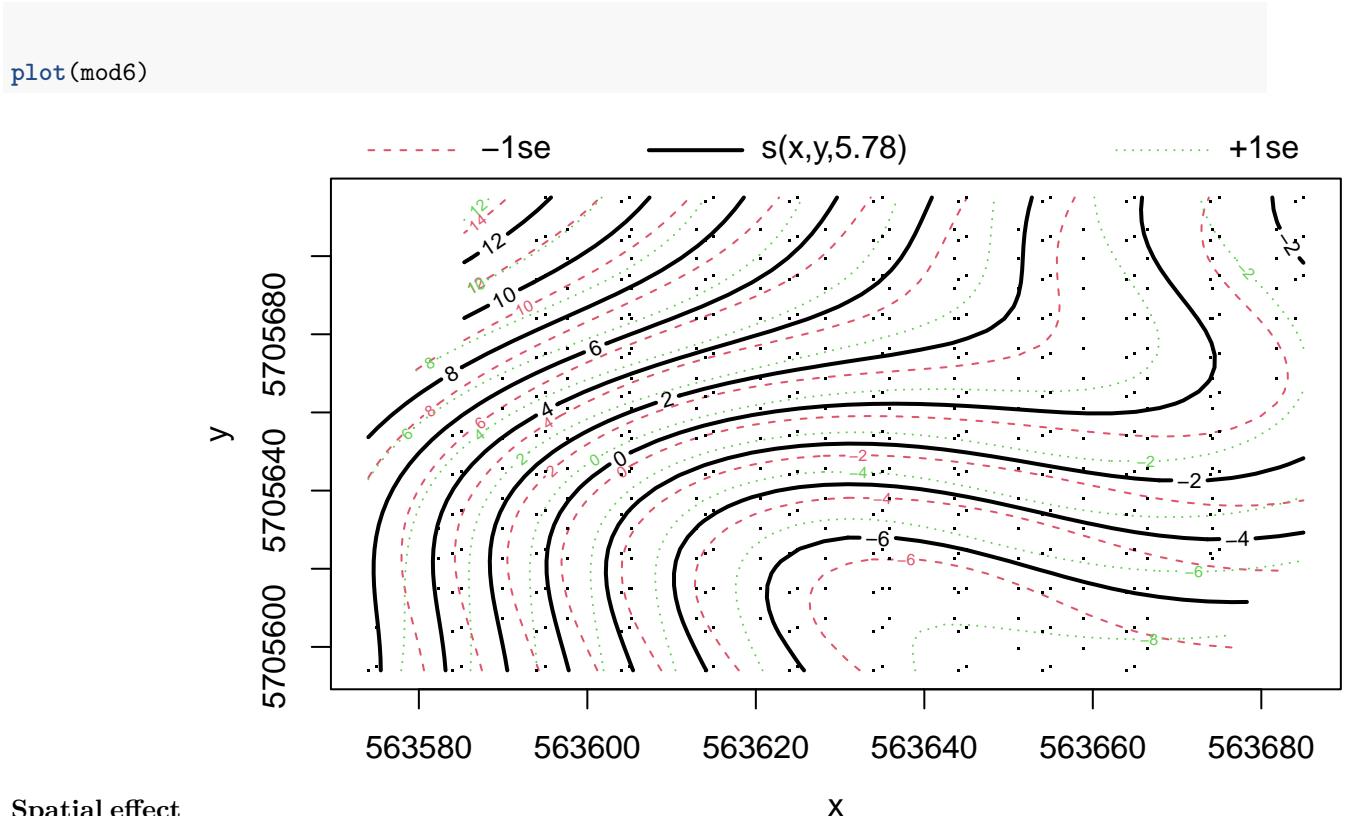
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 7) + source + inoculum
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)            92.3155    0.5255 175.656 < 2e-16 ***
## sources2s           -3.2212    0.6938 -4.643 4.61e-06 ***
## sourceuas          -14.0474    0.6494 -21.631 < 2e-16 ***
## inoculuminoculated   3.7610    0.8039  4.678 3.92e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df F p-value
## s(x,y) 5.779  5.985 47.12 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.668  Deviance explained = 67.5%
## -REML = 1330.1  Scale est. = 30.192 n = 425
```

**Intercept** The intercept here represents: canopy closure dates in inoculated areas determined by sentinel-2 and averages around 92.32 days after sowing.

**Remote sensing variables** UAS estimated canopy closure. an average of -14.05 days earlier than Sentinel 2.

**Inoculation effect** Non-inoculated areas were on average 3.76 days later than non-inoculated areas.

As shown in mod5 there is no interactive effect of source, indicating both remote sensing platforms showed inoculated areas had retarded phenology.



anova(mod5,mod6)

```
## Analysis of Deviance Table
##
## Model 1: das ~ s(x, y, k = 7) + source * inoculum
## Model 2: das ~ s(x, y, k = 7) + source + inoculum
##   Resid. Df Resid. Dev      Df Deviance    F Pr(>F)
## 1     412.85     12511
## 2     414.85     12536 -1.9999  -25.073 0.4141 0.6612
```

No significant difference, use simpler model mod6.

```

icept <- round(unname(mod6$coefficients[["(Intercept)"]]), 3)
tbl_dat <-
  data.table(Variable = c("Intercept", "Super S2", "UAS", "Inoculated Area"),
             round(summary(mod6)$p.table, 3))[, c(1, 2, 3, 5)][
               Estimate != icept, Estimate := Estimate + mod6$coefficients[["(Intercept)"]]]
cclosure_estimates6 <-
  tbl_dat |>
  flextable() |>
  align(align = "center", part = "all") |>
  fontsize(size = 8, part = "body") |>
  fontsize(size = 10, part = "header") |>
  italic(italic = TRUE, part = "header") |>
  set_caption(
    "Table ##: Estimated canopy closure in days after sowing (DAS) and the mean
     difference to the non-inoculated sentinel-2 data (intercept).
     P values indicate statistical significance in comparison to the intercept."
  ) |>
  autofit()

cclosure_estimates6

```

## results table

```

## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.

```

Table 2: Table ##: Estimated canopy closure in days after sowing (DAS) and the mean difference to the non-inoculated sentinel-2 data (intercept). P values indicate statistical significance in comparison to the intercept.

Variable	Estimate	Std. Error	Pr(> t )
Intercept	92.31500	0.526	0
Super S2	89.09448	0.694	0
UAS	78.26848	0.649	0
Inoculated Area	96.07648	0.804	0

```

save_as_docx(cclosure_estimates6, path = "./cclosure_estimates.docx")

```

## Epidemic Onset

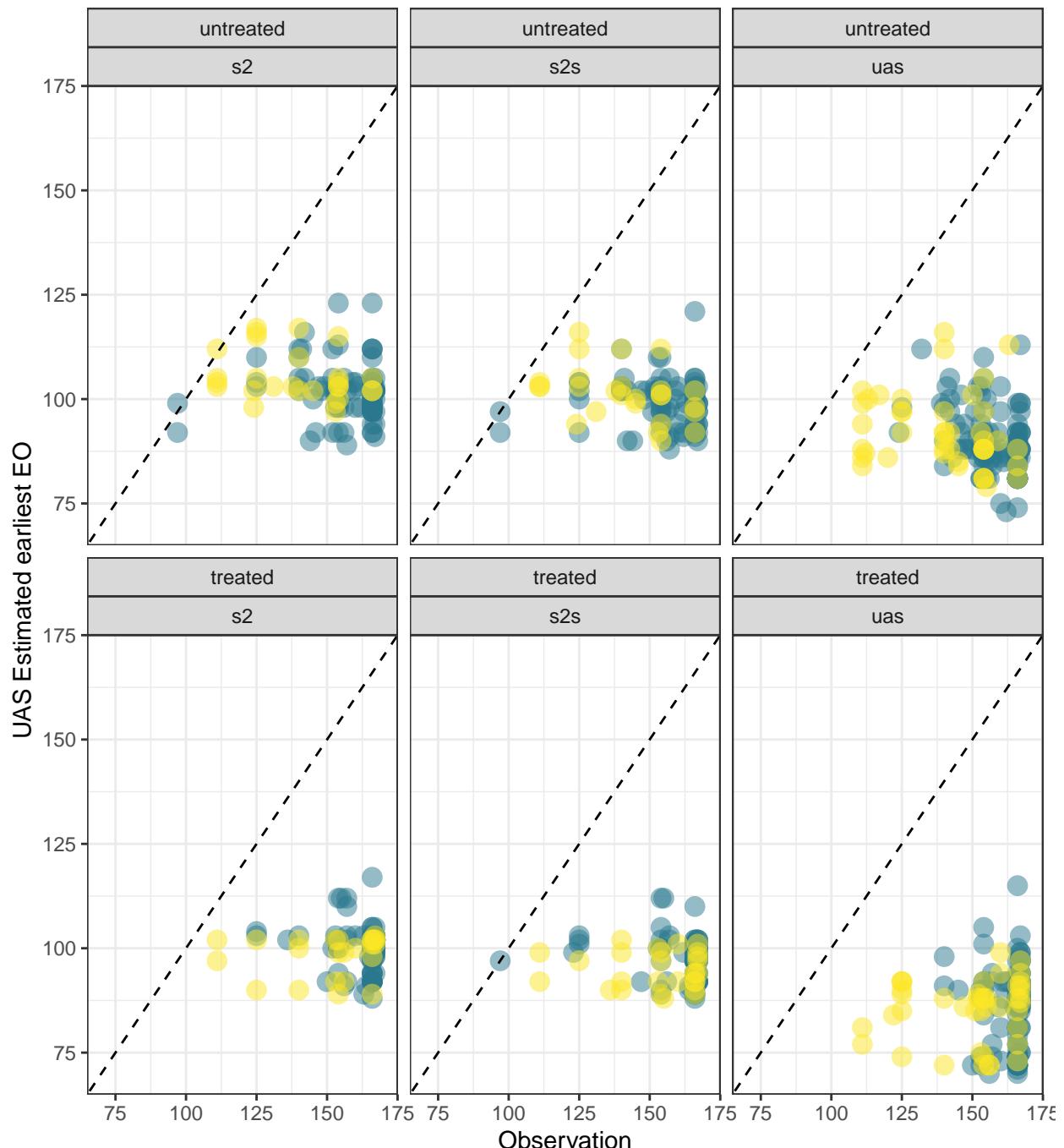
```
eeeo <- dat[data == "epidemic_onset"]
```

### Plot data

```
# plot EO data from each source an treatment
library(ggplot2)
eeeo[source != "observation",] |>
  ggplot(aes(x = obs_das, y = das, color = inoculum)) +
  geom_point(size = 4, alpha = .5, shape = 16) +
  facet_wrap(~fungicide + source) +
  theme_bw() +
  labs(title = "Epidemic onset by spatial location",
       x = "Observation",
       y = "UAS Estimated earliest EO") +
  theme(legend.position = "top") +
  xlim(c(70,170)) +
  ylim(c(70,170)) +
  geom_abline(slope = 1, intercept = 0, color = "black", linetype = "dashed") +
  scale_color_viridis_d(begin = 0.4)
```

## Epidemic onset by spatial location

inoculum   ● non-inoculated   ● inoculated



Model data

Model 1

What are the average influences independent of other treatments?

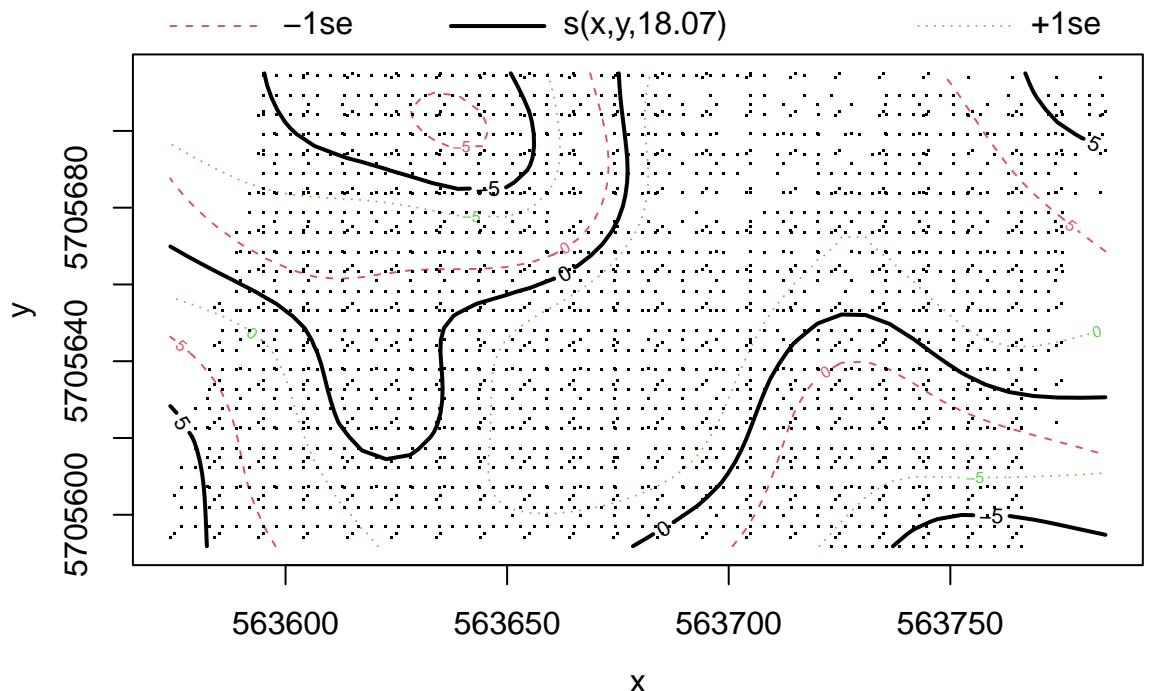
We will use Generalized Additive Models (GAM) to model the data, adding a spline to account for the spatial location of the data.

```
# what are the average influences independant of other treatments
mod1_eo <- gam(das ~ s(x,y) + source + inoculum + fungicide,
                 data = eeee,
                 method = "REML")
summary(mod1_eo)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y) + source + inoculum + fungicide
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           157.4373    0.8365 188.211  <2e-16 ***
## sources2            -55.1731    0.7759  -71.105  <2e-16 ***
## sources2s           -58.5320    0.7715  -75.866  <2e-16 ***
## sourceus            -68.0511    0.6731 -101.098  <2e-16 ***
## inoculuminoculated -8.3531    0.7392  -11.300  <2e-16 ***
## fungicidetreated     0.9944    1.5818     0.629     0.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df   F p-value
## s(x,y) 18.07 23.06 5.878 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.88  Deviance explained = 88.1%
## -REML = 8557.5  Scale est. = 126.99 n = 2225
```

Fungicide treated areas on average were not significantly different to untreated for epidemic onset days.

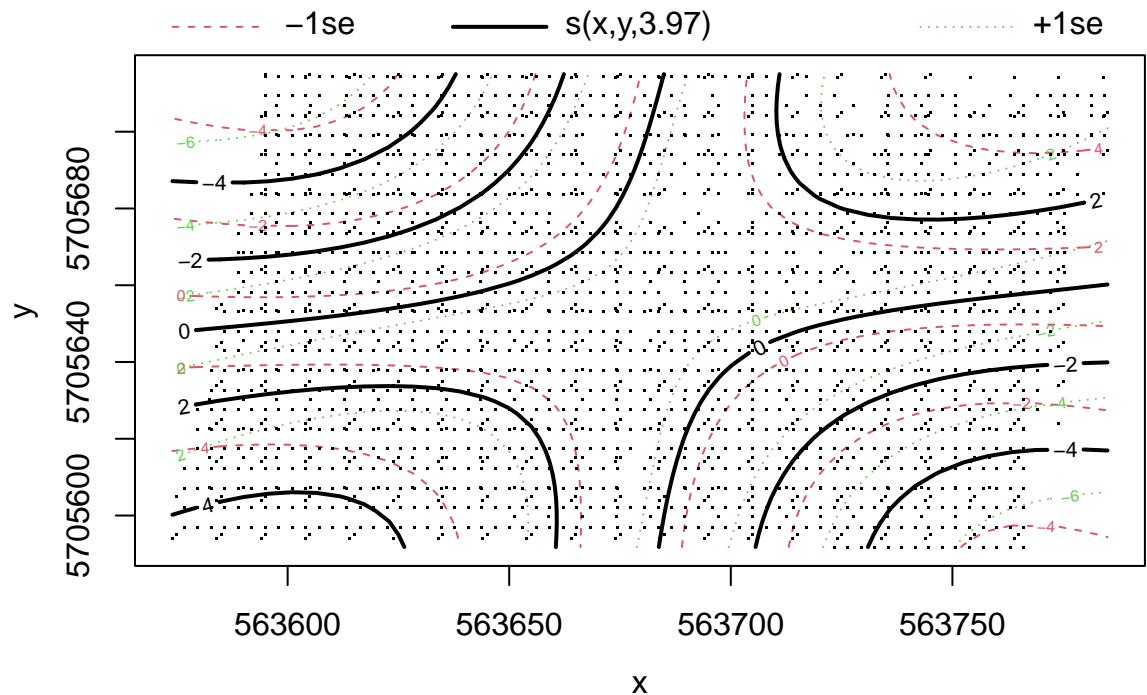
```
plot(mod1_eo)
```



check spatial fit

This looks slightly over fit. We should use fewer knots.

```
mod2_eo <- gam(das ~ s(x,y, k = 5) + source + inoculum + fungicide,
                 data = eeee,
                 method = "REML")
plot(mod2_eo)
```



Model 2

### Model 3

Is there a significant interaction between measurement source and inoculum?

```
mod3_eo <- gam(das ~ s(x,y, k = 5) + source * inoculum + fungicide,
                 data = eeee,
                 method = "REML")
summary(mod3_eo)

##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 5) + source * inoculum + fungicide
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                158.4458   0.5527 286.697 < 2e-16 ***
## sources2                  -58.9725   0.8342 -70.696 < 2e-16 ***
## sources2s                 -62.1738   0.8310 -74.817 < 2e-16 ***
## sourceus                  -72.2618   0.7403 -97.616 < 2e-16 ***
## inoculuminoculated        -16.3661   0.6983 -23.437 < 2e-16 ***
## fungicidetreated          2.6988    0.9300   2.902  0.00374 **
## sources2:inoculuminoculated 17.7096   1.8327   9.663 < 2e-16 ***
## sources2s:inoculuminoculated 16.7150   1.8073   9.248 < 2e-16 ***
## sourceus:inoculuminoculated 17.6428   1.4995  11.766 < 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(x,y) 3.972     4 30.42 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.889  Deviance explained =  89%
## -REML = 8445.5  Scale est. = 116.7 n = 2225
```

There is a significant interaction between inoculated areas observed epidemic onset and remote sensing platform dates.

### Model 4

Three-way interaction?

```
mod4_eo <- gam(das ~ s(x,y, k = 5) + source * inoculum * fungicide,
                 data = eeee)
summary(mod4_eo)
```

```
##
## Family: gaussian
```

```

## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 5) + source * inoculum * fungicide
##
## Parametric coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                   156.7161   0.5806 269.931
## sources2                     -55.1617   1.1468 -48.099
## sources2s                    -58.4878   1.1469 -50.998
## sourceuas                   -67.3544   1.0247 -65.729
## inoculuminoculated          -16.9463   0.9393 -18.042
## fungicidetreated             6.4453    1.0210   6.313
## sources2:inoculuminoculated 20.8792   2.4700   8.453
## sources2s:inoculuminoculated 20.0839   2.5019   8.028
## sourceuas:inoculuminoculated 19.7193   2.0618   9.564
## sources2:fungicidetreated    -7.8803   1.6254  -4.848
## sources2s:fungicidetreated   -7.6006   1.6192  -4.694
## sourceuas:fungicidetreated   -9.9151   1.4425  -6.874
## inoculuminoculated:fungicidetreated 1.0737   1.3623   0.788
## sources2:inoculuminoculated:fungicidetreated -6.8844   3.5755  -1.925
## sources2s:inoculuminoculated:fungicidetreated -6.5413   3.5226  -1.857
## sourceuas:inoculuminoculated:fungicidetreated -4.2520   2.9224  -1.455
##                                     Pr(>|t|)
## (Intercept)                   < 2e-16 ***
## sources2                     < 2e-16 ***
## sources2s                    < 2e-16 ***
## sourceuas                   < 2e-16 ***
## inoculuminoculated          < 2e-16 ***
## fungicidetreated             3.31e-10 ***
## sources2:inoculuminoculated < 2e-16 ***
## sources2s:inoculuminoculated 1.60e-15 ***
## sourceuas:inoculuminoculated < 2e-16 ***
## sources2:fungicidetreated    1.33e-06 ***
## sources2s:fungicidetreated   2.84e-06 ***
## sourceuas:fungicidetreated   8.12e-12 ***
## inoculuminoculated:fungicidetreated 0.4307
## sources2:inoculuminoculated:fungicidetreated 0.0543 .
## sources2s:inoculuminoculated:fungicidetreated 0.0635 .
## sourceuas:inoculuminoculated:fungicidetreated 0.1458
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df   F p-value
## s(x,y)  3.98     4 34.83 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.895  Deviance explained = 89.6%
## GCV = 111.76  Scale est. = 110.76  n = 2225

```

As fungicide treated areas lead to later epidemic onset, however this was not observed with remote sensing platforms.

In the fungicide treated areas S2 showed earlier estimated epidemic onset, super S2 showed earlier estimated epidemic onset, and UAS showed earlier estimated epidemic onset, This was likely due to non-treated areas returning later canopy closure.

Inoculated areas in fungicide treated plots showed for S2 earlier estimated epidemic onset, super S2 showed earlier estimated epidemic onset, and UAS showed earlier estimated epidemic onset, compared to non-inoculated areas.

## Drop Fungicide Treatment

The fungicide variable will be removed, initially I will test the model using fungicide as field replicate, then drop fungicide treated areas from the analysis.

```
mod5_eo <- gam(das ~ s(x,y, k =5) + source * inoculum,
                 data = eeeeo[fungicide != "treated"])
summary(mod5_eo)
```

## Model 5

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 5) + source * inoculum
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                156.8947   0.4722 332.232 < 2e-16 ***
## sources2                  -55.1329   1.2099 -45.568 < 2e-16 ***
## sources2s                  -58.4678   1.2096 -48.336 < 2e-16 ***
## sourceuas                 -67.3690   1.0823 -62.247 < 2e-16 ***
## inoculuminoculated        -16.8884   1.0362 -16.299 < 2e-16 ***
## sources2:inoculuminoculated 20.8358   2.6042   8.001 3.02e-15 ***
## sources2s:inoculuminoculated 20.0374   2.6377   7.597 6.31e-14 ***
## sourceuas:inoculuminoculated 19.7083   2.1747   9.063 < 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(x,y) 3.49  3.868 16.84 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.866  Deviance explained = 86.7%
## GCV = 124.32  Scale est. = 123.08    n = 1151
```

**Intercept** The intercept here represents:

Observed EO dates in non-inoculated areas

and averages around 156.89 days after sowing.

**Remote sensing variables** Sentinel 2 estimated epidemic onset an average of -55.13 days earlier than observed EO.

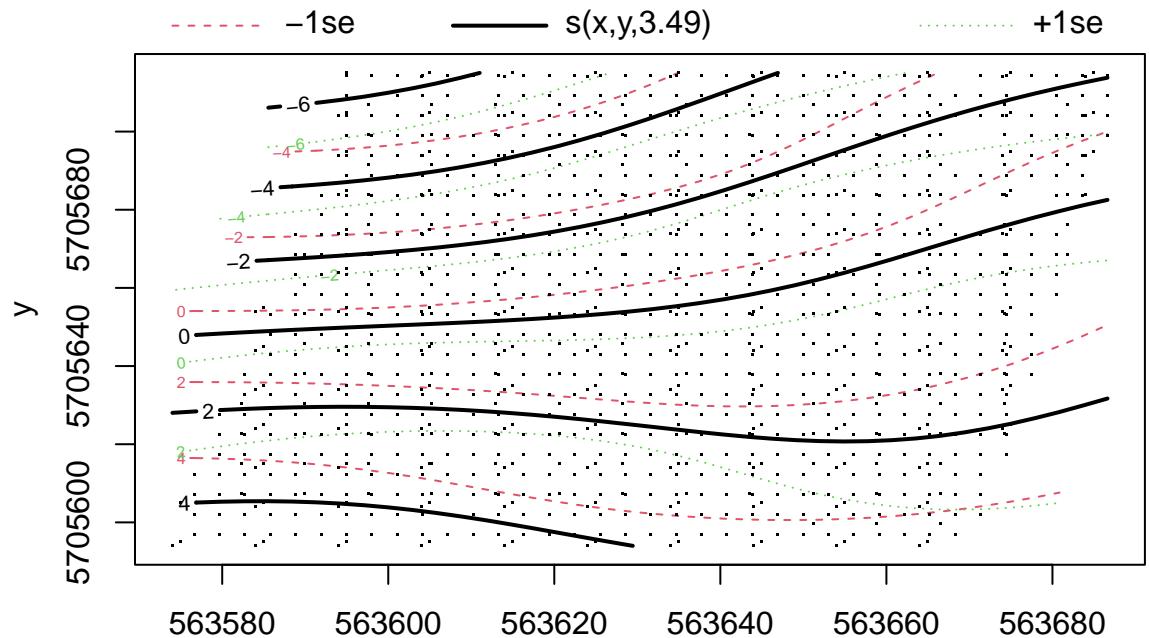
Super Resolution Sentinel 2 estimated earliest epidemic onset an average of -58.47 days earlier than observed EO.

UAS estimated earliest estimated epidemic onset an average of -67.37 days earlier than observed EO.

**Inoculation effect** Observed EO dates within inoculated areas were on average -16.89 days earlier than non-inoculated areas.

However EEEO dates in inoculated areas produced by remote sensing methods S2 , super S2 and UAS were 20.84, 20.04 and 19.71 days later than the observed EO days.

```
plot(mod5_eo)
```



### Spatial Effect

The spatial effect shows northern parts of the field showed earlier observed epidemic onset compared to Southern areas.

```
var_names <- names(mod5_eo$coefficients)  
icpt <- round(unname(mod5_eo$coefficients[["(Intercept)"]]),3)
```

```

adj_coefs <- unname(
  c(mod5_eo$coefficients[["(Intercept)"]],
    mod5_eo$coefficients[["(Intercept)"]] + mod5_eo$coefficients[var_names[2]],
    mod5_eo$coefficients[["(Intercept)"]] + mod5_eo$coefficients[var_names[3]],
    mod5_eo$coefficients[["(Intercept)"]] + mod5_eo$coefficients[var_names[4]],
    mod5_eo$coefficients[["(Intercept)"]] + mod5_eo$coefficients[var_names[5]],
    mod5_eo$coefficients[["(Intercept)"]] + mod5_eo$coefficients[var_names[2]] +
      mod5_eo$coefficients[var_names[5]] + mod5_eo$coefficients[var_names[6]],
    mod5_eo$coefficients[["(Intercept)"]] + mod5_eo$coefficients[var_names[3]] +
      mod5_eo$coefficients[var_names[5]] + mod5_eo$coefficients[var_names[7]],
    mod5_eo$coefficients[["(Intercept)"]] + mod5_eo$coefficients[var_names[4]] +
      mod5_eo$coefficients[var_names[5]] + mod5_eo$coefficients[var_names[7]])
))

tbl_dat <-
  data.table(Variable = c("Intercept", "Super S2", "UAS", "Inoculated Area"),
             round(summary(mod6)$p.table, 3))[, c(1, 2, 3, 5)][
               Estimate != 1, Estimate := Estimate + mod6$coefficients[["(Intercept)"]]]

eeeeo_estimates <-
  data.table(Variable = c("Intercept", "Sentinal-2", "Super Sentinal-2", "UAS",
                         "Observed : Inoculated Area", "S2 : Inoculated", "Super S2 : Inoculated",
                         "UAS : Inoculated"),
             Estimate = round(adj_coefs, 3),
             round(summary(mod5_eo)$p.table[, c(2, 4)], 3)) |>
  flextable() |>
  align(align = "center", part = "all") |>
  fontsize(size = 8, part = "body") |>
  fontsize(size = 10, part = "header") |>
  italic(italic = TRUE, part = "header") |>
  set_caption(
    "Table ##: Estimated epidemic onset in days after sowing (DAS) and the mean
     difference to the non-inoculated observed epidemic onset days (intercept).
     P values indicate statistical significance in comparison to the intercept."
  ) |>
  autofit()

eeeeo_estimates

```

## Results Table

```

## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.

```

Table 3: Table ##: Estimated epidemic onset in days after sowing (DAS) and the mean difference to the non-inoculated observed epidemic onset days (intercept). P values indicate statistical significance in comparison to the intercept.

Variable	Estimate	Std. Error	Pr(> t )
Intercept	156.895	0.472	0
Sentinal-2	101.762	1.210	0
Super Sentinal-2	98.427	1.210	0
UAS	89.526	1.082	0
Observed : Inoculated Area	140.006	1.036	0
S2 : Inoculated	105.709	2.604	0
Super S2 : Inoculated	101.576	2.638	0
UAS : Inoculated	92.675	2.175	0

```
save_as_docx(eeeeo_estimates, path = "./earliest_onsets_estimates.docx")
```

## Model 6

Does the interaction still hold up when fungicide is excluded?

```
mod6_eo <- gam(das ~ s(x,y, k = 5) + source + inoculum,
                 data = eeeeo[fungicide != "treated"])
summary(mod6_eo)

##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y, k = 5) + source + inoculum
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           155.2199    0.4812 322.57  <2e-16 ***
## sources2            -50.6040    1.1402 -44.38  <2e-16 ***
## sources2s           -54.1862    1.1438 -47.37  <2e-16 ***
## sourceuas          -62.6003    0.9987 -62.68  <2e-16 ***
## inoculuminoculated -9.3871    0.8912 -10.53  <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(x,y) 3.381 3.806 14.7  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

## 
## R-sq.(adj) =  0.849  Deviance explained =  85%
## GCV = 140.41  Scale est. = 139.39    n = 1151

```

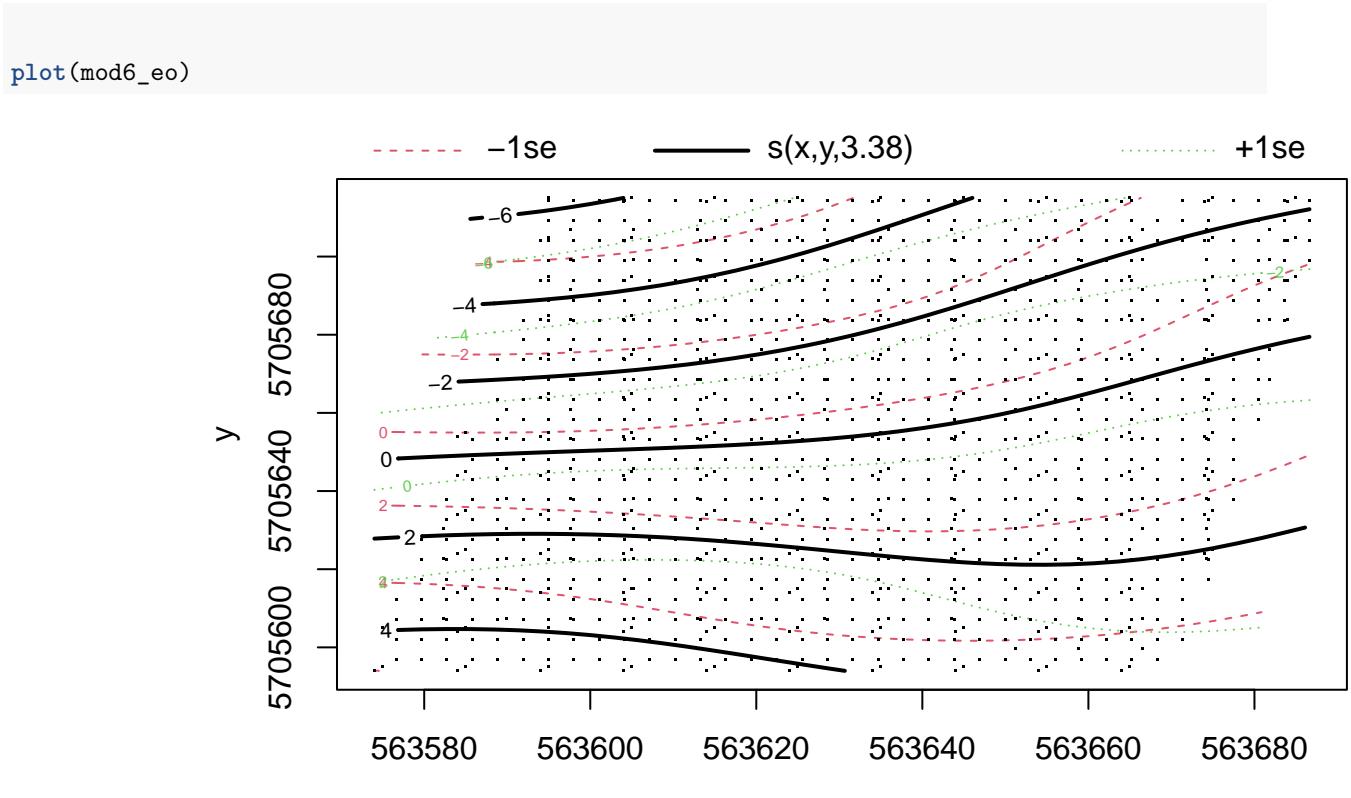
**Intercept** The intercept here represents: *Observed EO dates in non-inoculated areas* and averages around 155.22 days after sowing.

**Remote sensing variables** Sentinel 2 estimated epidemic onset an average of -50.6 days earlier than observed EO.

Super Resolution Sentinel 2 estimated epidemic onset an average of -54.19 days earlier than observed EO.

UAS estimated epidemic onset an average of -62.6 days earlier than observed EO.

**Inoculation effect** Observed EO dates within inoculated areas were on average -9.39 days earlier than non-inoculated areas.



#### Spatial Effect

The spatial effect shows northern parts of the field showed earlier observed epidemic onset compared to Southern areas.

#### Model comparisons

```

anova(mod5_eo,mod6_eo)

## Analysis of Deviance Table
##
## Model 1: das ~ s(x, y, k = 5) + source * inoculum
## Model 2: das ~ s(x, y, k = 5) + source + inoculum
##   Resid. Df Resid. Dev      Df Deviance      F    Pr(>F)
## 1     1139.1    140246
## 2     1142.2    159270 -3.0624   -19025 50.477 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
AIC(mod5_eo,mod6_eo)
```

```

##           df      AIC
## mod5_eo 12.490460 8819.360
## mod6_eo  9.380955 8959.557

```

There is a significant difference between these models and model 5, with the interaction shows a lower AIC.

---

## Proportional successes

What is the proportion of observed EO occuring later than EEEO?

This should show us how successful the negative prognosis model is

```

tbl_prop_success <-
  eeee[source != "observation" & fungicide != "treated",
  list("prop_success" = round(sum(das < obs_das)/NROW(das),3),
       n = .N),
  by = .(source, inoculum, fungicide)][,.(source,
                                             inoculum,
                                             n,
                                             prop_success)] |>
  flextable() |>
  align_text_col(align = "center") |>
  set_header_labels(source = "RS method",
                    inoculum = "Inoculum",
                    prop_success = "proportional success") |>
  fontsize(size = 8, part = "body") |>
  fontsize(size = 10, part = "header") |>
  set_caption("Table ##: Proportion success of RS estimated onset dates occuring prior to spatially pr
tbl_prop_success

```

```

## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.

```

Table 4: Table ##: Proportion success of RS estimated onset dates occurring prior to spatially proximal observed epidemic onset

RS method	Inoculum	n	proportional success
uas	non-inoculated	130	1.000
	inoculated	44	1.000
s2	non-inoculated	99	0.990
	inoculated	27	0.963
s2s	non-inoculated	99	0.990
	inoculated	26	1.000

```
save_as_docx(tbl_prop_success, path = "./prop_success.docx")
```

Lets visualise the densities of the observed and estimated epidemic onset dates

```

eee0[, obs_dif := das - obs_das]

eee0[source != "observation"] |>
  ggplot(aes(x = obs_dif, fill = inoculum)) +
  geom_density(alpha = .5) +
  facet_wrap(~source + fungicide) +
  theme_bw() +
  labs(x = "Difference between observed EO and Earliest estimated EO",
       y = "Density") +
  theme(legend.position = "top")+
  scale_fill_viridis_d()+
  geom_vline(xintercept = 0, linetype = "dashed", color = "black", linewidth = 1)

```

