

cercospoRa Earliest Epidemic Onset date analysis for Estimates

Paul Melloy, Rene Heim, Nathan Okole, Facundo Ramón Ispizua Yamati, Anne-Katrin Mahlein

2025-06-22

Setup environment

R libraries

Define helper functions

```
# Read in R function to help interpret gam outputs
source(here("R/interpret_gam.R"))

# Create function to easily return days, or date, after sowing
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)
    }
  }
}

# Test function
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/2025-05-13_fulldf_cc_cercospoRa(in).csv")[, "data" := "canopy_closure"]
eeee <- fread("../data/2025-05-13_fulldf_eo_cercospoRa(in).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(eeeo)
```

```
## [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(eeeo,
  old = c("area", "treatment"),
  new = c("inoculum", "fungicide"))

# Remove date column
eeeo[, date := NULL]
```

Rename variables Join together the two datasets.

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

```
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 1894
# 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"]
```

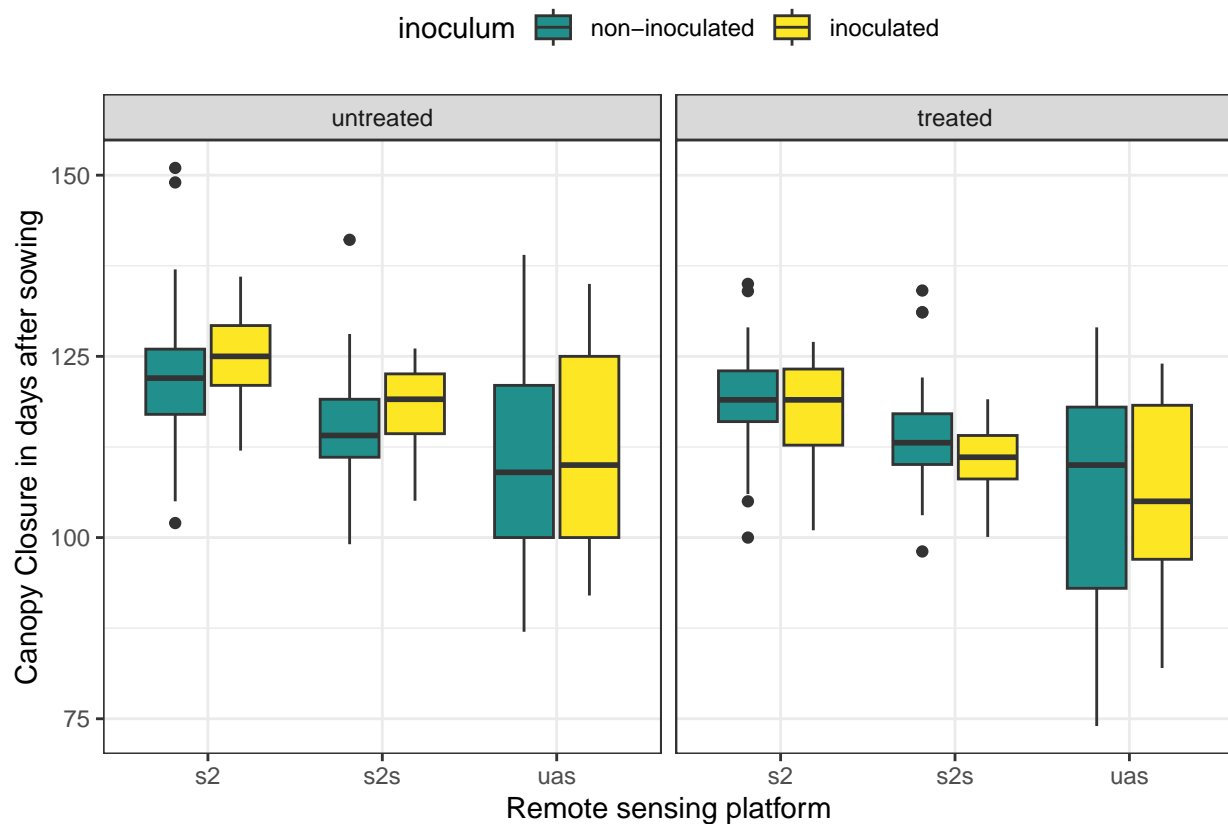
Do a quick check on aov of the data to inspect average differences between inoculation and fungicide treatments.

```
# 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
## -32.486  -4.928  -0.066   5.185  28.652
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    122.6775     0.7145  171.706 < 2e-16 ***
## sources2s       -6.6659     0.8434  -7.903 9.99e-15 ***
## sourceuas     -12.3290     0.8577 -14.375 < 2e-16 ***
## inoculuminoculated  0.4028     0.8305   0.485  0.628
## fungicidetreated  -3.8621     0.6932  -5.572 3.55e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.407 on 732 degrees of freedom
## Multiple R-squared:  0.2457, Adjusted R-squared:  0.2416
## F-statistic: 59.61 on 4 and 732 DF, p-value: < 2.2e-16
```

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.

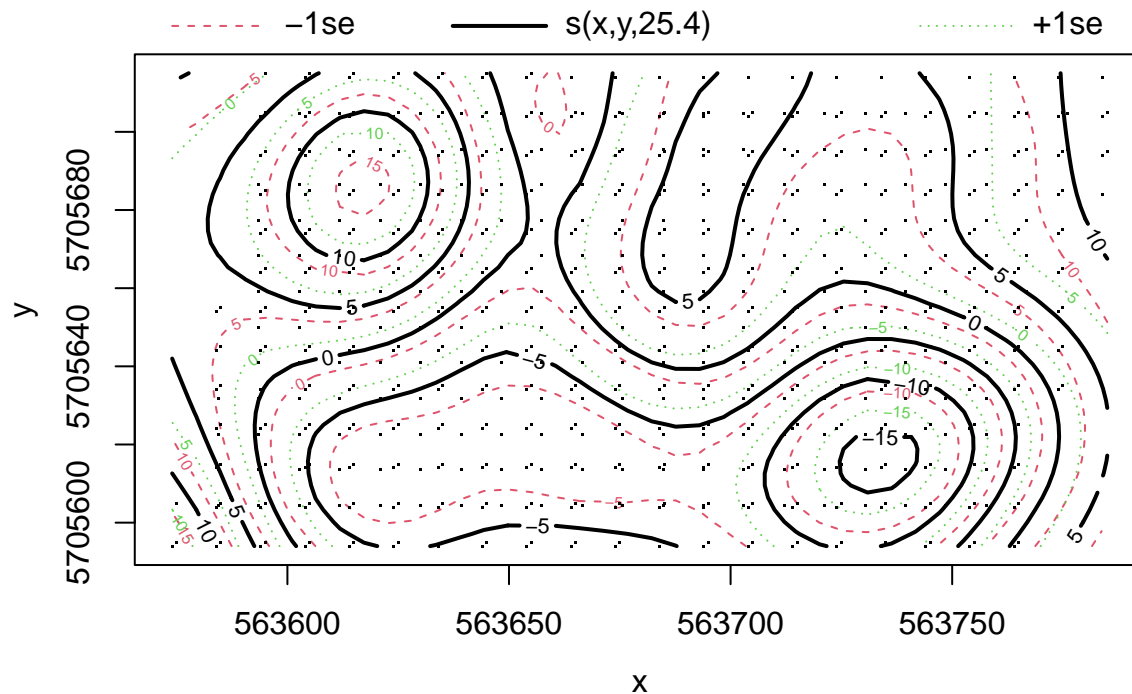
```
# what are the average influences independant of other treatments
mod1 <- gam(das ~ s(x,y) + source + inoculum + fungicide,
            data = cclo,
            method = "REML")
summary(mod1)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## das ~ s(x, y) + source + inoculum + fungicide
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   121.4048    1.0716  113.289  <2e-16 ***
## sources2s      -6.9067    0.6058  -11.400  <2e-16 ***
## sourceuas     -12.2133    0.6224  -19.623  <2e-16 ***
## inoculuminoculated  2.1910    0.8580   2.554  0.0109 *
## fungicidetreated -2.0181    1.9381  -1.041  0.2981
```

```
## ---
## 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) 25.4  28.22 24.8 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.612   Deviance explained = 62.7%
## -REML = 2481.4   Scale est. = 45.297    n = 737
```

Check the spline fit to ensure no over fitting.

```
plot(mod1)
```



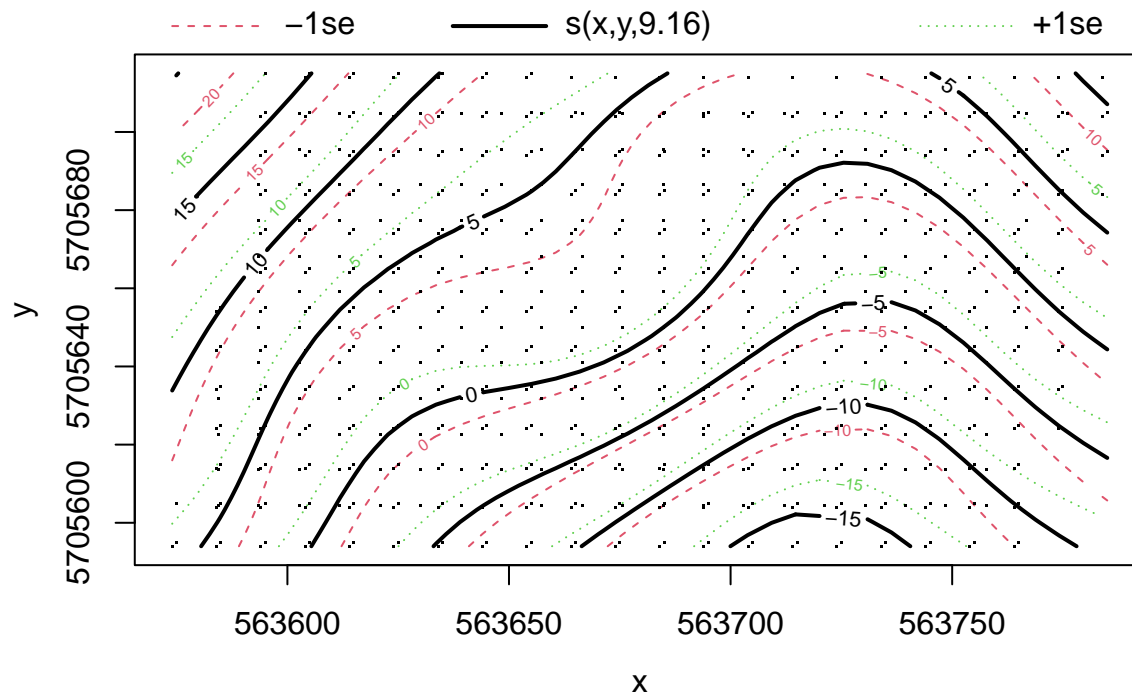
There seems an over-fitting due to the experimental setup. The spatial spline is accounting for too much of the inoculated areas.

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.

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

Non-Inoculated areas on average show earlier canopy closure, compared to inoculated areas.

Remote sensing platform UAS detects average canopy closure 12 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)	118.6848	0.9842	120.592	<2e-16 ***
sources2s	-6.7976	0.7665	-8.868	<2e-16 ***
sourceuas	-12.5466	0.7979	-15.725	<2e-16 ***
inoculuminoculated	1.7934	1.2719	1.410	0.1590
fungicidetreated	3.8048	1.6265	2.339	0.0196 *
sources2s:inoculuminoculated	-0.5816	1.6664	-0.349	0.7272
sourceuas:inoculuminoculated	0.3633	1.6398	0.222	0.8247

```
## ---
```

```
## 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)  9.149   9.859 40.56  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.508   Deviance explained = 51.8%
## -REML = 2540.1   Scale est. = 57.411    n = 737
```

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)                  119.0598      1.0011 118.927  < 2e-16 ***
## sources2s                     -6.8851      0.6775 -10.162  < 2e-16 ***
## sourceuas                    -12.3806      0.6933 -17.859  < 2e-16 ***
## inoculuminoculated             3.8932      1.1394   3.417  0.000669 ***
## fungicidetreated              2.8848      1.7262   1.671  0.095119 .
## inoculuminoculated:fungicidetreated -3.7858      1.6162  -2.342  0.019431 *
## ---
## 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)  9.554   9.96 40.28  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.513   Deviance explained = 52.2%
## GCV = 58.098   Scale est. = 56.872    n = 737
```

There is a significant interaction, between fungicide and inoculum ($P = 0.02$). 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)    118.3986    1.0400  113.842 < 2e-16 ***
## sources2s       -6.9587    0.9593   -7.254 1.05e-12 ***
## sourceuas      -11.5062    0.9718  -11.840 < 2e-16 ***
## fungicidetreated  4.3941    1.8125    2.424  0.0156 *
## inoculuminoculated 1.7748    0.7977    2.225  0.0264 *
## sources2s:fungicidetreated 0.0587    1.3587    0.043  0.9656
## sourceuas:fungicidetreated -1.9278    1.3886   -1.388  0.1655
## ---
## 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) 9.151   9.86 40.76 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.51   Deviance explained =  52%
## -REML = 2539.3   Scale est. = 57.228    n = 737

```

```

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 71.357 <2e-16
## fungicide       1  5.877  0.0156
## inoculum        1  4.950  0.0264
## source:fungicide 2  1.325  0.2663
##
## Approximate significance of smooth terms:
##               edf Ref.df      F p-value
## s(x,y) 9.151   9.860 40.76 <2e-16

```

Fungicide treatment

Overall there was no significant effects of fungicide treated part of the field. The only significant result showed remote sensing methods S2 and UAS resulted in earlier EEEO dates in fungicide treated compared to

untreated.

The fungicide treated area seemed to show thicker canopy and I think this result can be put down to spatial field effects.

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)    121.3754     0.8164 148.676 < 2e-16 ***
## sources2s       -6.6827     1.1124  -6.007 4.62e-09 ***
## sourceuas      -10.9251     1.1424  -9.564 < 2e-16 ***
## inoculuminoculated  5.0669     1.8267   2.774 0.00583 **
## sources2s:inoculuminoculated -0.8151     2.3887  -0.341 0.73314
## sourceuas:inoculuminoculated -1.7939     2.3540  -0.762 0.44654
## ---
## 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) 5.686    5.97 27.15 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.456 Deviance explained = 47.2%
## -REML = 1281.8 Scale est. = 60.189 n = 371
```

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.

<i>Variable</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>Pr(> t)</i>
Intercept	121.375	0.816	0.000
Super S2	-6.683	1.112	0.000
UAS	-10.925	1.142	0.000
Inoculated Area	5.067	1.827	0.006
Super S2 : Inoculated	-0.815	2.389	0.733
UAS : Inoculated	-1.794	2.354	0.447

```
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)   121.5808    0.7528 161.502 < 2e-16 ***
## sources2s     -6.8683    0.9822  -6.993 1.31e-11 ***
## sourceuas    -11.3495    0.9958 -11.398 < 2e-16 ***
## inoculuminoculated 4.1700    1.2041   3.463 0.000598 ***
## ---
## 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) 5.686   5.97 28.24 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.458   Deviance explained = 47.1%
## -REML = 1285.6   Scale est. = 59.955      n = 371
```

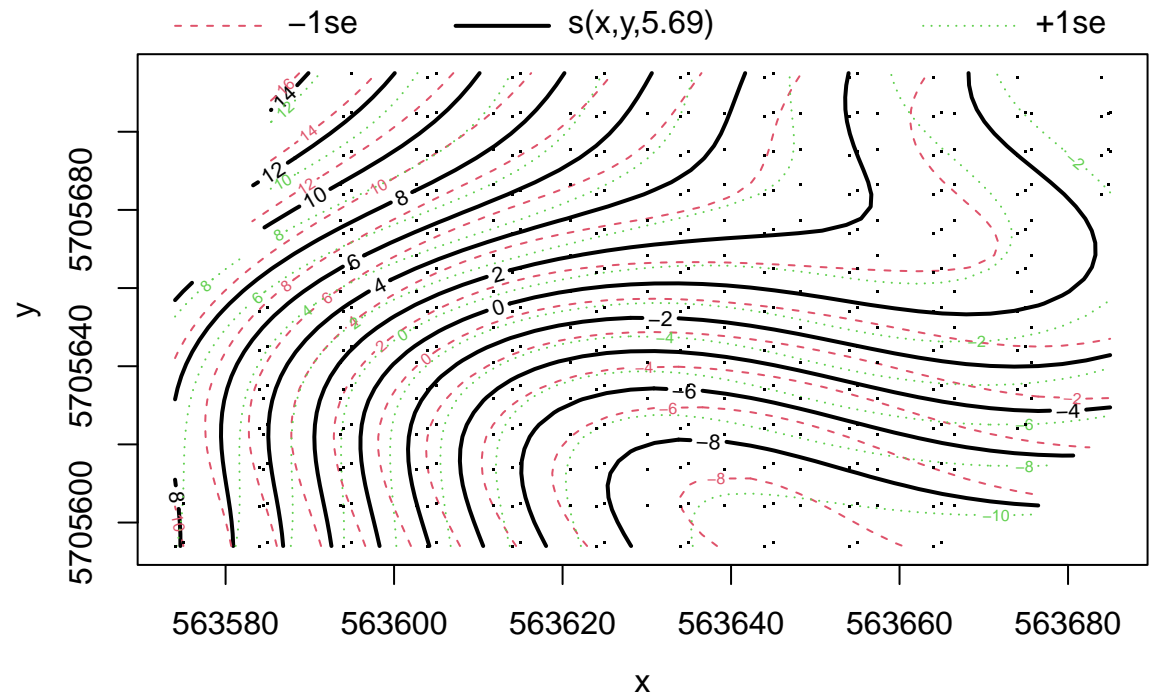
Intercept The intercept here represents: canopy closure dates in inoculated areas determined by sentinel-2 and averages around 121.58 days after sowing.

Remote sensing variables UAS estimated epidemic onset an average of -11.35 days earlier than Sentinel 2.

Inoculation effect Non-inoculated areas were on average 4.17 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.

```
plot(mod6)
```



Spatial effect

Model comparisons

```
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      358.79      21627
## 2      360.79      21663 -2   -36.044 0.2994 0.7414
```

No significant difference, use simpler model mod6.

```

icept <- round(unnname(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 sentinal-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 sentinal-2 data (intercept). P values indicate statistical significance in comparison to the intercept.

<i>Variable</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>Pr(> t)</i>
Intercept	121.5810	0.753	0.000
Super S2	114.7128	0.982	0.000
UAS	110.2318	0.996	0.000
Inoculated Area	125.7508	1.204	0.001

```

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

```

Epidemic Onset

Successful negative prognosis models estimate *earliest* epidemic onset (EEEE) dates prior to actual observed epidemic onset (EO). Failed EEEO dates occur after the actual EO.

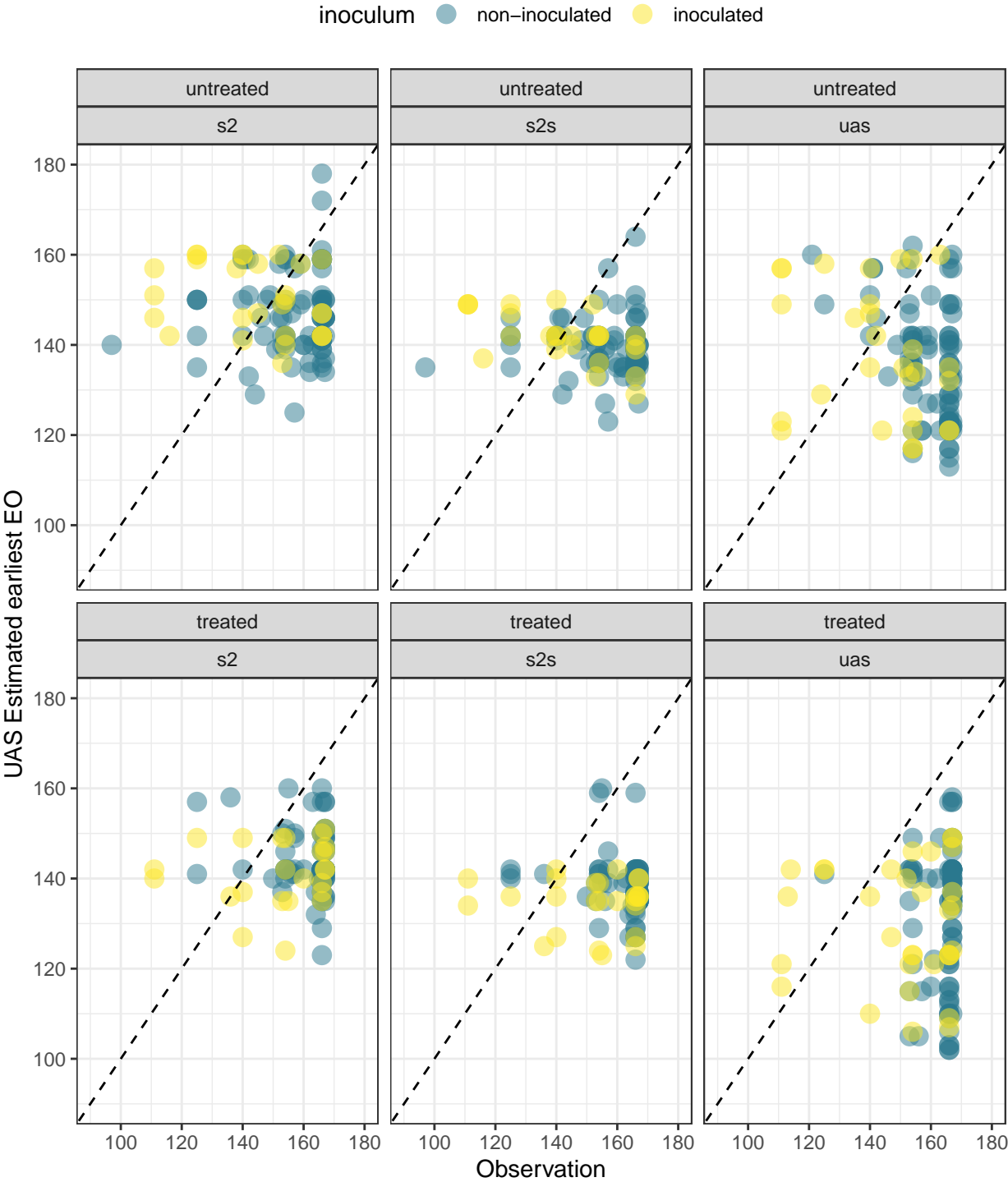
```
eeeeo <- dat[data == "epidemic_onset"]
```

If we plot the estimated vs observed success will be shown by all dots below the dashed 1:1 line.

Plot data

```
# plot EO data from each source an treatment
library(ggplot2)
eeeeo[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(90,180))+
  ylim(c(90,180))+
  geom_abline(slope = 1, intercept = 0, color = "black", linetype = "dashed")+
  scale_color_viridis_d(begin = 0.4)
```

Epidemic onset by spatial location



Model data

Model 1

What are the average influences independent of other treatments?

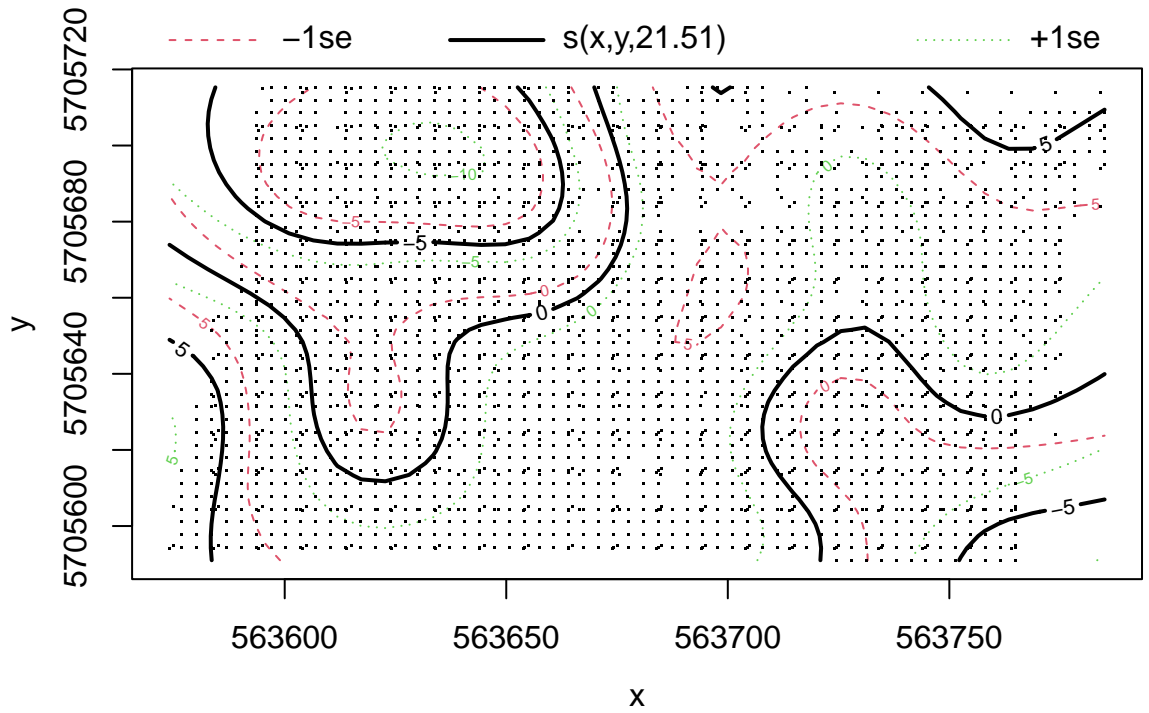
We will use Generalized Additive Models (GAM) to model the data, adding a spline to account for the spatial location of the data.

```
# what are the average influences independant of other treatments
mod1_eo <- gam(das ~ s(x,y) + source + inoculum + fungicide,
               data = eeeo,
               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)    157.8811     0.8626  183.033  <2e-16 ***
## sources2       -11.0949     0.8249  -13.450  <2e-16 ***
## sources2s      -18.2991     0.8094  -22.609  <2e-16 ***
## sourceuas      -22.7860     0.8340  -27.323  <2e-16 ***
## inoculuminoculated -8.5882     0.7693  -11.164  <2e-16 ***
## fungicidetreated  0.5816     1.6840   0.345    0.73
## ---
## 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) 21.51  25.97 7.583  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.406   Deviance explained = 41.2%
## -REML = 10308   Scale est. = 146.16      n = 2631
```

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

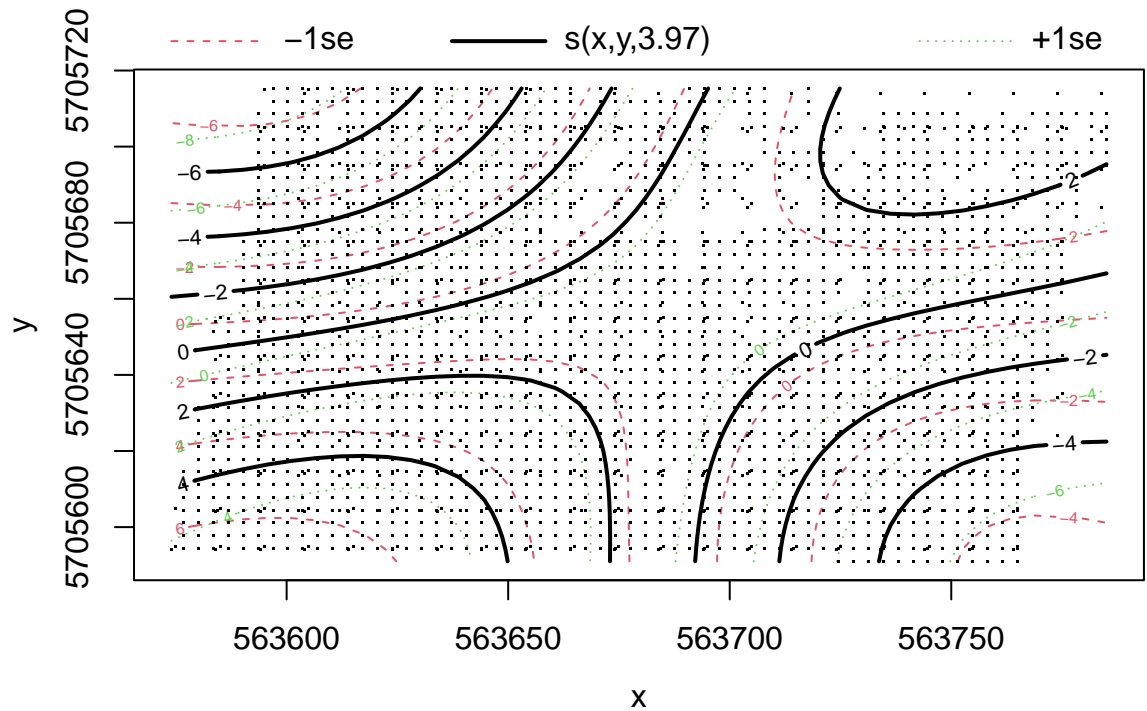
```
plot(mod1_eo)
```



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 = eeeo,
  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 = eeeo,
               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)    157.9318     0.5398  292.583 < 2e-16 ***
## sources2       -14.6797     0.9135  -16.069 < 2e-16 ***
## sources2s      -21.4755     0.8921  -24.072 < 2e-16 ***
## sourceuas      -26.7288     0.9446  -28.297 < 2e-16 ***
## inoculuminoculated -15.2225     0.6507  -23.395 < 2e-16 ***
## fungicidetreated   3.6767     0.9452   3.890 0.000103 ***
## sources2:inoculuminoculated 16.3784     1.9670   8.326 < 2e-16 ***
## sources2s:inoculuminoculated 15.0763     1.9436   7.757 1.24e-14 ***
## sourceuas:inoculuminoculated 15.6368     1.8777   8.328 < 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) 3.976      4 38.98 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.428 Deviance explained = 43%
## -REML = 10235 Scale est. = 140.87 n = 2631
```

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 = eeeo)
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)    156.7012    0.5585  280.573
## sources2       -10.3429    1.2563   -8.233
## sources2s      -17.8611    1.2391  -14.414
## sourceuas      -21.7388    1.2748  -17.053
## inoculuminoculated -16.1428    0.8844  -18.253
## fungicidetreated   6.3687    1.0062   6.330
## sources2:inoculuminoculated 19.9845    2.6479   7.547
## sources2s:inoculuminoculated 19.1624    2.7092   7.073
## sourceuas:inoculuminoculated 18.8040    2.6253   7.163
## sources2:fungicidetreated  -8.9315    1.7896  -4.991
## sources2s:fungicidetreated  -7.3820    1.7472  -4.225
## sourceuas:fungicidetreated -10.6092    1.8510  -5.732
## inoculuminoculated:fungicidetreated 1.7221    1.2750   1.351
## sources2:inoculuminoculated:fungicidetreated -8.0601    3.8605  -2.088
## sources2s:inoculuminoculated:fungicidetreated -7.9383    3.8075  -2.085
## sourceuas:inoculuminoculated:fungicidetreated -5.2576    3.6812  -1.428
##
##               Pr(>|t|)
## (Intercept)    < 2e-16 ***
## sources2       2.85e-16 ***
## sources2s      < 2e-16 ***
## sourceuas      < 2e-16 ***
## inoculuminoculated < 2e-16 ***
## fungicidetreated 2.88e-10 ***
## sources2:inoculuminoculated 6.10e-14 ***
## sources2s:inoculuminoculated 1.94e-12 ***
## sourceuas:inoculuminoculated 1.02e-12 ***
## sources2:fungicidetreated 6.41e-07 ***
## sources2s:fungicidetreated 2.47e-05 ***
## sourceuas:fungicidetreated 1.11e-08 ***
## inoculuminoculated:fungicidetreated 0.1769
## sources2:inoculuminoculated:fungicidetreated 0.0369 *
## sources2s:inoculuminoculated:fungicidetreated 0.0372 *
## sourceuas:inoculuminoculated:fungicidetreated 0.1533
## ---
## 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) 3.982      4 42.6 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.451 Deviance explained = 45.5%
## GCV = 136.12 Scale est. = 135.09 n = 2631

```

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 USAS 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 USAS 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 = eeeo[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)    156.9405     0.4394 357.169 < 2e-16 ***
## sources2       -10.3432     1.3166  -7.856 7.96e-15 ***
## sources2s      -17.8454     1.2990 -13.738 < 2e-16 ***
## sourceuas      -21.7424     1.3388 -16.240 < 2e-16 ***
## inoculuminoculated -15.9780     0.9834 -16.248 < 2e-16 ***
## sources2:inoculuminoculated 19.9672     2.7747   7.196 1.02e-12 ***
## sources2s:inoculuminoculated 19.1272     2.8392   6.737 2.37e-11 ***
## sourceuas:inoculuminoculated 18.8290     2.7532   6.839 1.20e-11 ***
## ---
## 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) 3.724   3.961 26.77 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.353   Deviance explained = 35.8%
## GCV = 149.59   Scale est. = 148.32     n = 1384
```

Intercept The intercept here represents:

Observed EO dates in non-inoculated areas
and averages around 156.94 days after sowing.

Remote sensing variables Sentinel 2 estimated epidemic onset an average of -10.34 days earlier than observed EO.

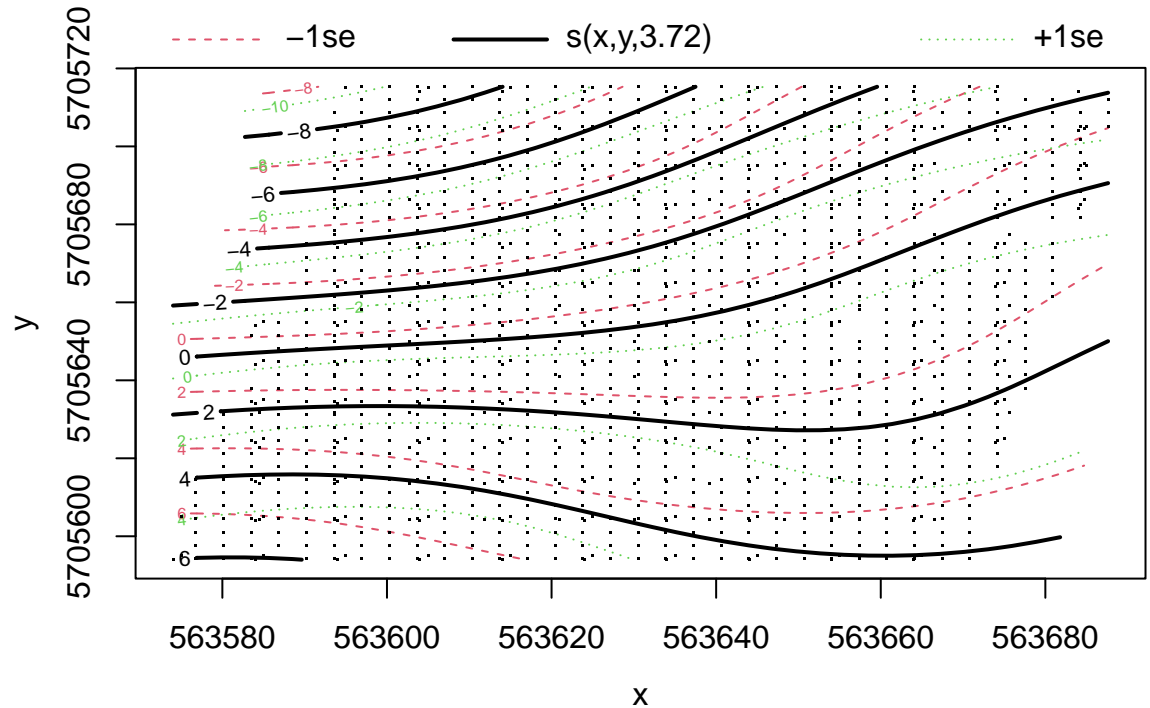
Super Resolution Sentinel 2 estimated earliest epidemic onset an average of -17.85 days earlier than observed EO.

UAS estimated earliest estimated epidemic onset an average of -21.74 days earlier than observed EO.

Inoculation effect Observed EO dates within inoculated areas were on average -15.98 days earlier than non-inoculated areas.

However EEO dates in inoculated areas produced by remote sensing methods S2 , super S2 and UAS were 19.97, 19.13 and 18.83 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)$p.table, 3))[, c(1, 2, 3, 5)] [
      Estimate != icept, Estimate := Estimate + mod6$coefficients["(Intercept)"]]]

eeee_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()

eeee_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.

<i>Variable</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>Pr(> t)</i>
Intercept	156.940	0.439	0
Sentinal-2	146.597	1.317	0
Super Sentinal-2	139.095	1.299	0
UAS	135.198	1.339	0
Observed : Inoculated Area	140.963	0.983	0
S2 : Inoculated	150.587	2.775	0
Super S2 : Inoculated	142.244	2.839	0
UAS : Inoculated	138.347	2.753	0

```
save_as_docx(eeeo_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 = eeeo[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)    155.7877     0.4437  351.094 < 2e-16 ***
## sources2       -5.8728     1.2065   -4.867 1.26e-06 ***
## sources2s     -13.8091     1.2024  -11.485 < 2e-16 ***
## sourceuas     -17.3514     1.2173  -14.254 < 2e-16 ***
## inoculuminoculated -10.7212     0.8895  -12.054 < 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)  3.685   3.95 24.88 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.299   Deviance explained = 30.3%
## GCV = 161.76   Scale est. = 160.75      n = 1384
```

Intercept The intercept here represents: *Observed EO dates in non-inoculated areas* and averages around 155.79 days after sowing.

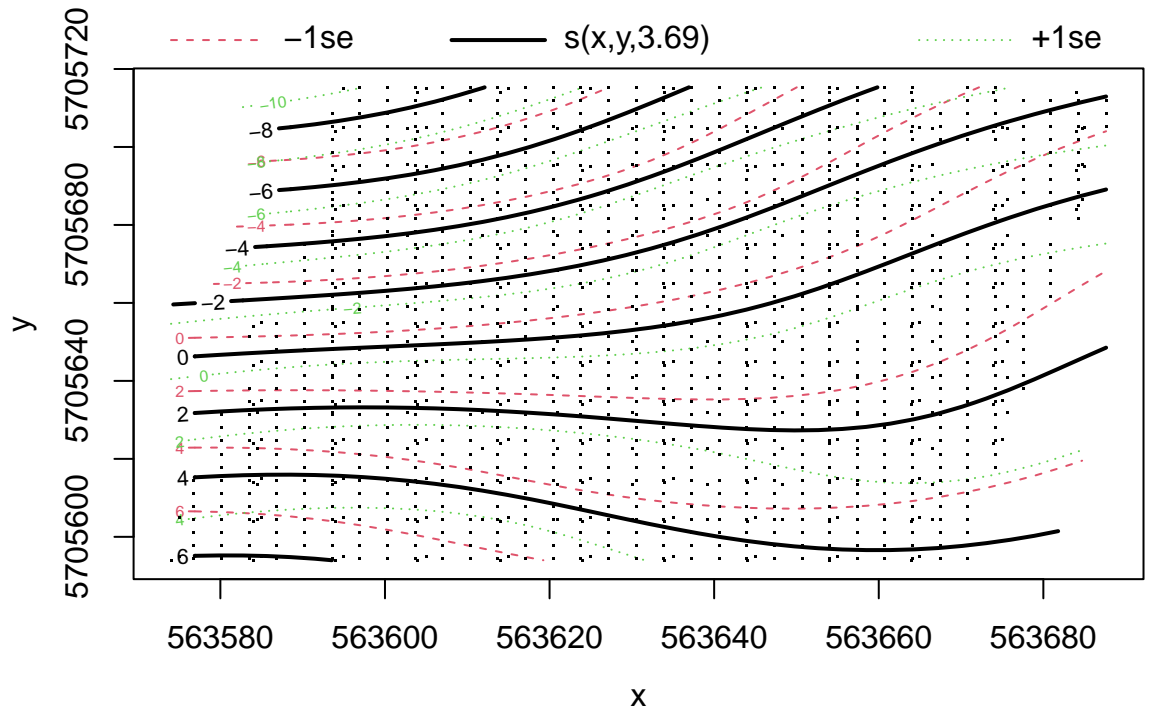
Remote sensing variables Sentinel 2 estimated epidemic onset an average of -5.87 days earlier than observed EO.

Super Resolution Sentinel 2 estimated epidemic onset an average of -13.81 days earlier than observed EO.

UAS estimated epidemic onset an average of -17.35 days earlier than observed EO.

Inoculation effect Observed EO dates within inoculated areas were on average -10.72 days earlier days than non-inoculated areas.

```
plot(mod6_eo)
```



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    1372.0    203541
## 2    1375.1    221076 -3.0117   -17535 39.255 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##           df      AIC
## mod5_eo 12.724298 10860.46
## mod6_eo  9.685314 10968.76
```

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 occurring later than EEEO?

This should show us how successful the negative prognosis model is

```
tbl_prop_success <-
  eeeo[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 occurring prior to spatially proximal observed epidemic onset")

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	93	0.892
uas	inoculated	29	0.517
s2	non-inoculated	96	0.750
s2	inoculated	28	0.464
s2s	non-inoculated	99	0.848
s2s	inoculated	26	0.577

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

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

```
eeeo[, obs_dif := das - obs_das]
```



```

eeeo[source != "observation"] |>
  ggplot(aes(x = obs_dif, fill = inoculum)) +
  geom_density(alpha = .5) +
  facet_wrap(~fungicide + source) +
  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)

```

