Supplementary materials file: *No tillage outperforms conventional tillage under arid conditions and following fertilization*

Authors: Stavros D. Veresoglou, Junjiang Chen, Xuheng Du, Qi Fu, QingLiu Geng, Chenyan Huang, Xilin Huang, Nan Hu, Yiming Hun, Guolin C. Li, Zhiman Lin, Zhiyu Ma, Yuyi Ou, Shuo Qi, Haitian Qin, Yingbo Qiu, Xibin Sun,Ye Tao, YiLing Tian, Jie Wang, Lingxiao Wu, Ziwei Wu, Siqi Xie, Ao Yang, Dan Yang, Chen Zeng, Ying Zeng, RuJie Zhang

Contents:

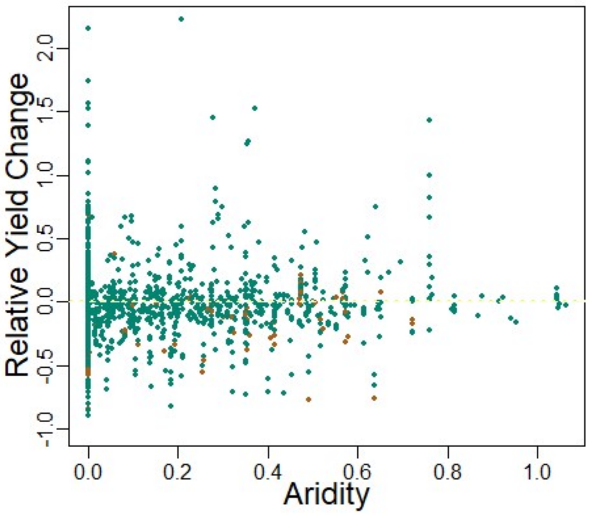

Fig. S1. Scatter plot between relative yield change (*y* – axis) and aridity (*x* – axis) of crops thad did not receive (in green) and those that received (red) fertilization. We overlay the null line for RYC = 0 in yellow.

Table S1 Median quantile regression coefficients and significance (at $P=0.05$; highlighted with a star sign) between *relative yield change* and the parameters yield and aridity for each of the ten crops in the database[†]

Crop	Yield	Aridity
Barley spring	-0.00003*	0.15*
Barley winter	-0.00024*	1.10*
Cotton	-0.00008*	-0.21 *n.s*
Maize	-0.00001*	0.07*
Rice	0.00000 *n.s*	0.04 *n.s*
Sorghum	0.00000 *n.s*	0.63*
Soybean	-0.00003*	0.02 *n.s*
Sunflower	-0.00051*	2.08*
Wheat spring	-0.00002 n.s	0.02 *n.s*
Wheat winter	0.00000 *n.s*	0.04*

[†] We only used entries that contained information on fertilization status

Appendix One: Annotated code on the analysis we carried out

```
#________________________________
#
#      Load Some Libraries
#________________________________
library(randomForest)
library(nlme)
library(performance)
library(vegan)
data<-read.csv("database.csv", sep=",") ##Read the dataset
### we only retain data with fertilization information
data<-data[(data[,29]=="Yes")|(data[,29]=="No"),]



#_____________________________________________________
#
#      Part I. Random Forests - Data Exploration
#_____________________________________________________

mylist<-list()      ### Create an empty list where we store results for
different crops
for(counter in 1:10)### Work with the ten crops
{
data<-read.csv("database.csv", sep=",")
data<-data[data$Crop==levels(data$Crop)[counter],]
data<-data[(data[,29]=="Yes")|(data[,29]=="No"),]
### Reset some factors
data[,29]<-factor(data[,29])
data[,22][data[,22]=="Mixed"]<-""
data[,22]<-factor(data[,22])
data[,9]<-as.numeric(as.character(data[,9]))
data[,9][is.na(data[,9])]<-median(data[,9], na.rm=T)
data$aridity<-1-(data$P / data$E)       ### Calculate aridity
data$aridity[data$aridity<0]<-0
### Focus on a specific set of parameters (forteen) which are eligible for
the model
data2<-data[,c(41, 11, 29, 6,39, 22,50, 51, 52, 9, 53, 47, 54, 15)]
colnames(data2)[1]<-"RYC"  ### Rename Yield to RYC
### This is the main model ####
model <- randomForest(
  formula = RYC ~ .,
  data = data2
)
which.min(model$mse)
sqrt(model$mse[which.min(model$mse)])
mylist[[counter]]<-print(importance(model,type = 2))
}
###############################################
### Process the results from Ranodm Forests ###
###############################################
mynames<-rownames(mylist[[1]])
mylist2<-mylist                         ## Save output under a new name
mylist2<-matrix(unlist(mylist2), 13, 10)## Convert the list to a matrix
rownames(mylist2)<-mynames        ## Name the parameters
scores<-apply(mylist2, 1, mean)   ## Average the importance of the parameters
```

```
sort(scores)                         ## Sort the results
```

output ONE: Mean Relative Importance for Parameters

```
                    Crop           Fertilization.CT
               0.0000000                  0.6334872
         Soil.cover.in.NT       pH..surface.layer.
               0.8483506                  1.3397314
                      ST                   Latitude
               2.3157234                 11.7411233
                       P                       Tave
              61.7888271                 66.6038702
                    Tmin Years.since.NT.started..yrs.
              67.8854497                 85.8060288
                 aridity                       Tmax
              89.2161833                145.3988911
              Yield.of.CT
             931.0853455
```

```
#_______________________________________________________
#
#             Part II. General Model
#_______________________________________________________
###Reload data
data<-read.csv("database.csv", sep=",")
data<-data[(data[,29]=="Yes")|(data[,29]=="No"),]
### Reset some factors
data[,29]<-factor(data[,29])
data[,22][data[,22]=="Mixed"]<-""
data[,22]<-factor(data[,22])
data[,9]<-as.numeric(as.character(data[,9]))
data[,9][is.na(data[,9])]<-median(data[,9], na.rm=T)
data$aridity<-1-(data$P / data$E)       ### Calculate aridity
data$aridity[data$aridity<0]<-0
data2<-data[,c(41, 11, 29, 6,39, 22,50, 51, 52, 9, 53, 47, 54, 15)]
colnames(data2)[1]<-"RYC"  ### Rename Yield to RYC
###Fit an overall model with all parameters and check the assumptions
anova(model2<-lme(RYC~Yield.of.CT+aridity+Tmax+Fertilization.CT, data=data2,
random=~1|Crop))
plot(resid(model2)~fitted(model2))             ### no homoscedasticity
data2$Yield.of.CT<-log(data2$Yield.of.CT)      ### transform response
variable and refit
anova(model2<-lme(RYC~Yield.of.CT+aridity+Tmax+Fertilization.CT, data=data2,
random=~1|Crop))
plot(resid(model2)~fitted(model2), ylim=c(-2,2))### no homoscedasticity
anova(model2<-lme(RYC~Yield.of.CT+aridity+Tmax+Fertilization.CT, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT)))
plot(resid(model2, type="normalized")~data2$Yield.of.CT, ylim=c(-2,2),
xlab="Yield data", ylab="residuals of best fit model", pch=21, bg="grey")
hist(resid(model2, type="normalized"), breaks=100, col="grey",
xlab="residuals best fit model", main="")
###############################################################################
### the model now looks ok but we have many non significant variables     ###
### We engage in a manual procedure to remove them                        ###
###############################################################################
anova(model2<-lme(RYC~Fertilization.CT+Yield.of.CT+aridity, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT)))
summary(model2<-lme(RYC~Fertilization.CT+Yield.of.CT+aridity, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT)))
```

output TWO: Statistics of the simplified model

```
> anova(model2<-
lme(RYC~Fertilization.CT+Yield.of.CT+aridity, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT)))
                 numDF denDF   F-value p-value
(Intercept)          1  3973   0.46424  0.4957
Fertilization.CT     1  3973   7.35386  0.0067
Yield.of.CT          1  3973  71.65360  <.0001
aridity              1  3973 232.15776  <.0001
Warning messages:
1: In nlminb(c(oldPars), function(lmePars) -logLik(lmeSt,
lmePars),  :
  NA/NaN function evaluation
2: In nlminb(c(oldPars), function(lmePars) -logLik(lmeSt,
lmePars),  :
  NA/NaN function evaluation
> summary(model2<-
lme(RYC~Fertilization.CT+Yield.of.CT+aridity, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT)))
Linear mixed-effects model fit by REML
 Data: data2
       AIC      BIC    logLik
  10741.93 10785.95 -5363.963

Random effects:
 Formula: ~1 | Crop
        (Intercept) Residual
StdDev:  0.08063092 2.297801

Variance function:
 Structure: Exponential of variance covariate
 Formula: ~Yield.of.CT
 Parameter estimates:
       expon
-0.0001723084
Fixed effects: RYC ~ Fertilization.CT + Yield.of.CT +
aridity
                          Value  Std.Error   DF   t-value
p-value
(Intercept)         -0.10359950 0.05857562 3973 -1.768646
0.077
Fertilization.CTYes  0.11946830 0.05001349 3973  2.388722
0.017
Yield.of.CT         -0.00000536 0.00000063 3973 -8.573548
0.000
aridity              0.23433001 0.01537929 3973 15.236724
0.000
 Correlation:
                    (Intr) Fr.CTY Yl..CT
Fertilization.CTYes -0.789
Yield.of.CT          0.047 -0.121
```

```
aridity             -0.028  0.057 -0.853

Standardized Within-Group Residuals:
          Min            Q1           Med             Q3
Max
-11.515272014  -0.108061224   0.006164442   0.146733124
46.176327731

Number of Observations: 3986
Number of Groups: 10
Warning messages:
1: In nlminb(c(oldPars), function(lmePars) -logLik(lmeSt,
lmePars),  :
  NA/NaN function evaluation
2: In nlminb(c(oldPars), function(lmePars) -logLik(lmeSt,
lmePars),  :
  NA/NaN function evaluation
```

```
hist(resid(model2, type="normalized"), breaks=100, col="grey",
xlab="residuals best fit model", main="")
plot(resid(model2, type="normalized")~data2$Yield.of.CT, ylim=c(-2,2),
xlab="Yield data", ylab="residuals of best fit model", pch=21, bg="grey")
plot(resid(model2, type="normalized")~fitted(model2), ylim=c(-2,2),
xlab="Fitted values", ylab="residuals of best fit model", pch=21, bg="grey")


#We repeated fitting the model on all observations (also those without
fertilization information) in the form of a sensitivity analysis. The result
was as follows:
```

output TWOb: Sensitivity test on the optimal model

```
> anova(model2<-lme(RYC~Yield.of.CT+aridity, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT)))
            numDF denDF   F-value p-value
(Intercept)     1  4391   0.51523  0.4729
Yield.of.CT     1  4391 128.30085  <.0001
aridity         1  4391  50.87115  <.0001
```

```
#________________________________________________________
#
#      Part III. Variation Partitioning
#________________________________________________________
a<-matrix(NA, 8, 2) ### a is a vector of eight elements to store results
a[1,]<-unlist(r2_nakagawa(lme(RYC~Yield.of.CT, data=data2, random=~1|Crop,
weights=varExp(form=~Yield.of.CT))))
a[2,]<-unlist(r2_nakagawa(lme(RYC~Fertilization.CT, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT))))
a[3,]<-unlist(r2_nakagawa(lme(RYC~aridity, data=data2, random=~1|Crop,
weights=varExp(form=~Yield.of.CT))))
a[4,]<-unlist(r2_nakagawa(lme(RYC~Yield.of.CT+Fertilization.CT, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT))))
a[5,]<-unlist(r2_nakagawa(lme(RYC~Yield.of.CT+aridity, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT))))
a[6,]<-unlist(r2_nakagawa(lme(RYC~Fertilization.CT+aridity, data=data2,
random=~1|Crop, weights=varExp(form=~Yield.of.CT))))
a[7,]<-unlist(r2_nakagawa(model2))
a<-a*100*10^5 #### in thousant
### Now we do the calculations manually
```

```
aaa<-a
aaa[4,]<-a[1,]+a[2,]-a[4,]
aaa[5,]<-a[1,]+a[3,]-a[5,]
aaa[6,]<-a[2,]+a[3,]-a[6,]
aaa[1,]<-a[1,]-aaa[4,]-aaa[5,]
aaa[2,]<-a[2,]-aaa[4,]-aaa[6,]
aaa[3,]<-a[3,]-aaa[5,]-aaa[6,]
aaa[7,]<-aaa[1,]+aaa[2,]+aaa[3,]+aaa[4,]+aaa[5,]+aaa[6,]-a[7,]
aaa[8,]<-10^7
l<-aaa[5,]
aaa[5,]<-aaa[6,]
aaa[6,]<-l
rownames(aaa)<-c("Yield (Y)", "Fertilization (F)", "Aridity (A)", "Y+F", 
"F+A", "Y+A", "Y+F+A", "")
aaa[aaa<0]<-0
### And plot the variance partitioning graph
showvarparts(3, paste(rownames(aaa), Xnames=round(aaa[,1], 2), sep="\n= "), 
id.size=0.7, bg=2:4)
par(cex=0.85)
showvarparts(3, paste(Xnames=round(aaa[,1], 2), " ", sep="") , id.size=0.2, 
bg=2:4)
text(-0.10,0.4, "Yield", adj=1, cex=1.2)
text(0.8,0.4, "Fertilization", adj=0, cex=1.2)
text(0.5,-1.4, "Aridity", cex=1.2)

#_____________________________________________________
#
#      Part IV. Some Descriptive statistics
#_____________________________________________________
###Reload data
data<-read.csv("database.csv", sep=",")
data<-data[(data[,29]=="Yes")|(data[,29]=="No"),]
### Reset some factors
data[,29]<-factor(data[,29])
data[,22][data[,22]=="Mixed"]<-""
data[,22]<-factor(data[,22])
data[,9]<-as.numeric(as.character(data[,9]))
data[,9][is.na(data[,9])]<-median(data[,9], na.rm=T)
data$aridity<-1-(data$P / data$E)       ### Calculate aridity
data$aridity[data$aridity<0]<-0
colnames(data)[41]<-"RYC"               ### Rename the response variable
t.test(data$RYC[data[,29]=="Yes" ])      ### is the effect larger than zero?
```

output THREE: One sample *t* test for fertilized trials

```
> t.test(data$RYC[data[,29]=="Yes" ])

        One Sample t-test

data:  data$RYC[data[, 29] == "Yes"]
t = 2.7657, df = 3597, p-value = 0.00571
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.02838098 0.16661963
sample estimates:
 mean of x 
0.09750031
```

(*note that the analysis assumes a normally distributed response variable which is an assumptions that is partially met and should be interpreted cautiously. We rely for this mainly on the large number of entries and do not filter outliers out*)