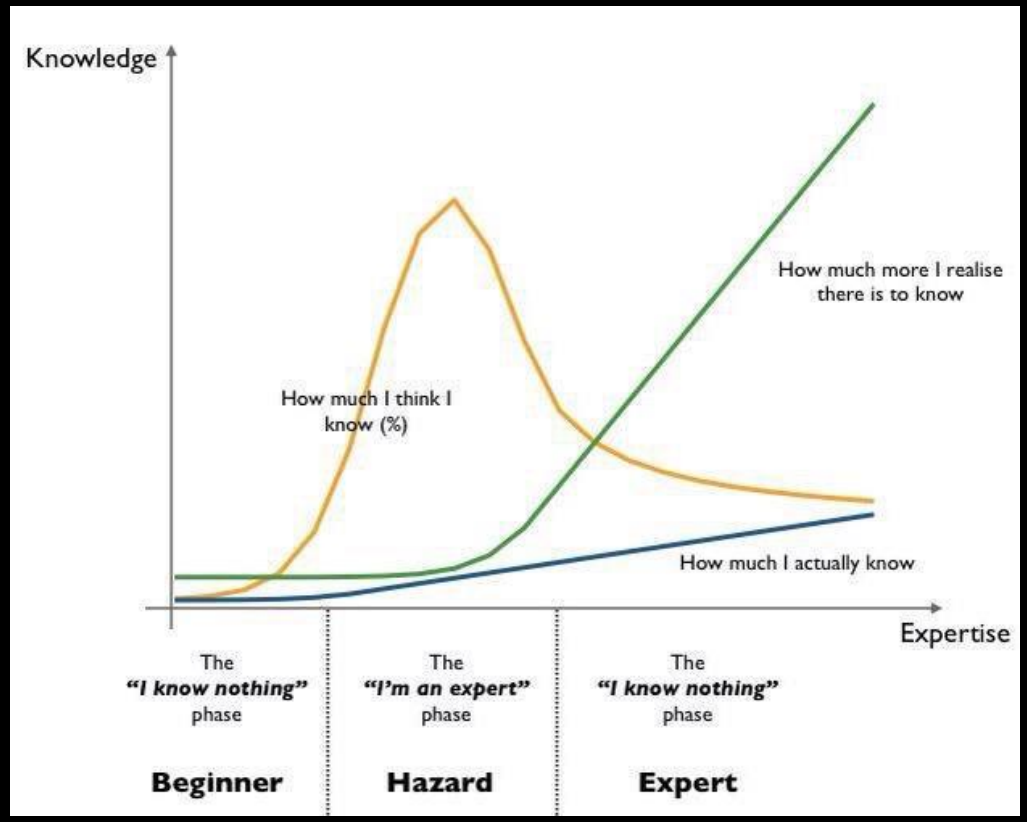


Modelling workflows in R and the R-INLA package

Julie Vercelloni

11/05/2021

Heron Island survey

scientific data

Explore content | Journal information | Publish with us

nature › scientific data › data descriptors › article

Data Descriptor | Open Access | Published: 16 March 2021

Benthic and coral reef community field data for Heron Reef, Southern Great Barrier Reef, Australia, 2002–2018

Chris Roelfsema, Eva M. Kovacs, Kathryn Markey, Julie Vercelloni, Alberto Rodriguez-Ramirez, Sebastian Lopez-Marcano, Manuel Gonzalez-Rivero, Ove Hoegh-Guldberg & Stuart R. Phinn

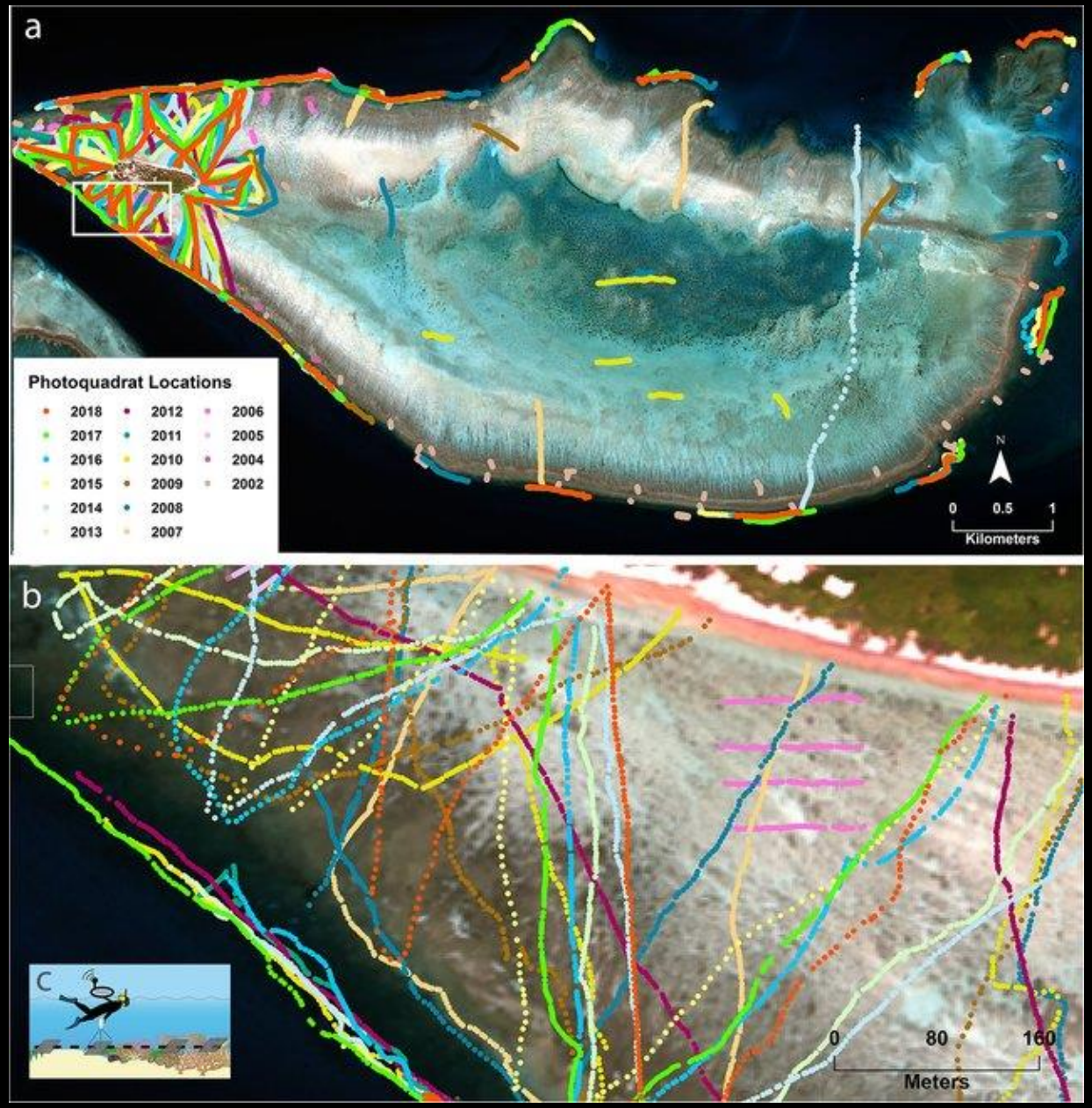
Report | Published: 29 April 2021

Fine-scale time series surveys reveal new insights into spatio-temporal trends in coral cover (2002–2018), of a coral reef on the Southern Great Barrier Reef

Chris Roelfsema, Eva M. Kovacs, Julie Vercelloni, Kathryn Markey, Alberto Rodriguez-Ramirez, Sebastian Lopez-Marcano, Manuel Gonzalez-Rivero, Ove Hoegh-Guldberg & Stuart R. Phinn

Coral Reefs (2021) | Cite this article

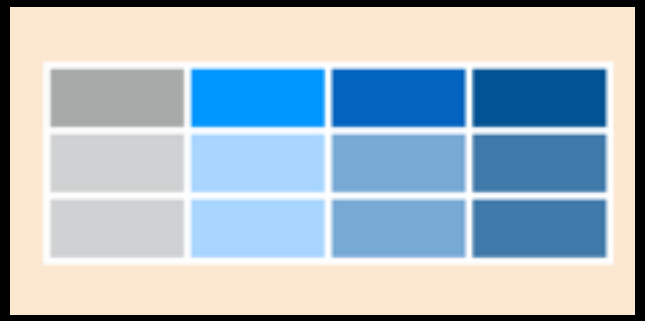
173 Accesses | 9 Altmetric | Metrics



Modelling workflow
STEP1: Data preparation

Heron/script/Data_prep.R

→ Reading the raw data



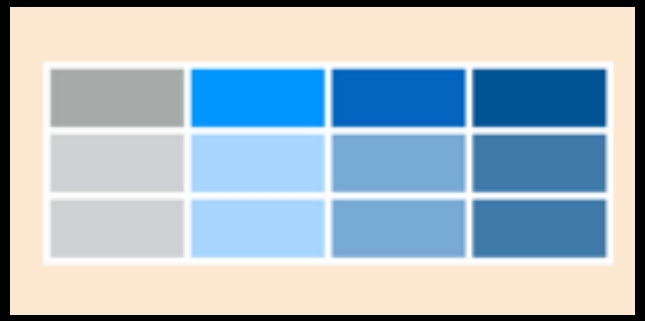
Wide format
45,000 rows and 54 columns

```
> names(t) [1] "X" "transectid" "year" [4] "reef_name"
"reef_type" "sub_region" [7] "transects_reef_section"
"surveyid" "id" [10] "lat" "lng" "ALG_OTH" [13] "Caul" "Chlor"
"Cya_spe" [16] "Discp" "Lobph" "MACR_Cal_H" [19] "MAECBS"
"Pad" "Turbin" [22] "CAL_CCA_DC" "EAM_DHC" "EAM_RB" [25]
"ACR.BRA" "ACR.BRA.B." "ACR.HIP" [28] "ACR.OTH" "ACR.PE"
"BRA_DIG_Ac" [31] "BRA_OTH" "BRA_TAB.Ac" "BRA_TAB.B." [34]
"FAV.MUS" "MASE_OTH" "OTH.HC" [37] "POCI" "POR.BRA" "POR.ENC"
[40] "POR.MASS" "TFP_RDG_Al" "TFP_RDG.B." [43] "TFP_RND_Al"
"ALC.SF" "X.TAPE" [46] "Other" "Unc" "Unk" [49] "MOB.INV"
"OTH.SINV" "GORG" [52] "OTH.SF" "BMA_sand" "Sand"
```

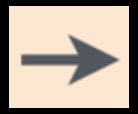
Modelling workflow
STEP1: Data preparation

Heron/script/Data_prep.R

→ Reading the raw data



Wide format
45,000 rows and 54 columns



Long format
1,950,000 rows and 13 columns

```
> library(dplyr)
```

```
➢ Rec_t<- t%>%
➢ gather(key = Comm_benthic,
  Proportion,  12:54)
```

Modelling workflow
STEP1: Data preparation

Heron/script/Data_prep.R

→ Reading the raw data

- Learn about the data

```
> Rec_t %>% group_by(Comm_benthic) %>% distinct()
# A tibble: 1,954,866 x 13 # Groups: Comm_benthic
[43]
> map(Rec_t, ~sum(is.na(.)))
  $X [1] 0
> $transectid [1] 0
> $year [1] 0
> $reef_name [1] 0
  $reef_type [1] 0
  $sub_region [1] 0
  $transects_reef_section
  [1] 0
> $surveyid [1] 0 $id [1] 0
  $lat [1] 0
> $lng [1] 0 $Comm_benthic
  [1] 0 $Proportion [1] 0
```

Modelling workflow
STEP1: Data preparation

Heron/script/Data_prep.R

→ Reading the raw data

- Learn about the data

- Check points

```
> Rec_tally<-Rec.units%>%group_by(splitting.var,year)%>%tally
> wrong_group<-names(which(table(Rec_tally$splitting.var)<=1))
  wrong_group character(0)
```

Modelling workflows
STEP1: Data preparation

Heron/script/Data_prep.R

→ Reading the raw data

- Learn about the data
- Check points
- Happy about the format
- Ensure that everything makes sense to you

→ Save new table calls “data_ready.csv”

Modelling workflow
STEP1: Data preparation

Input

Output

Heron/script/Data_prep.R

Heron/data/Raw

Heron/data/data_ready.csv

→ Reading the raw data

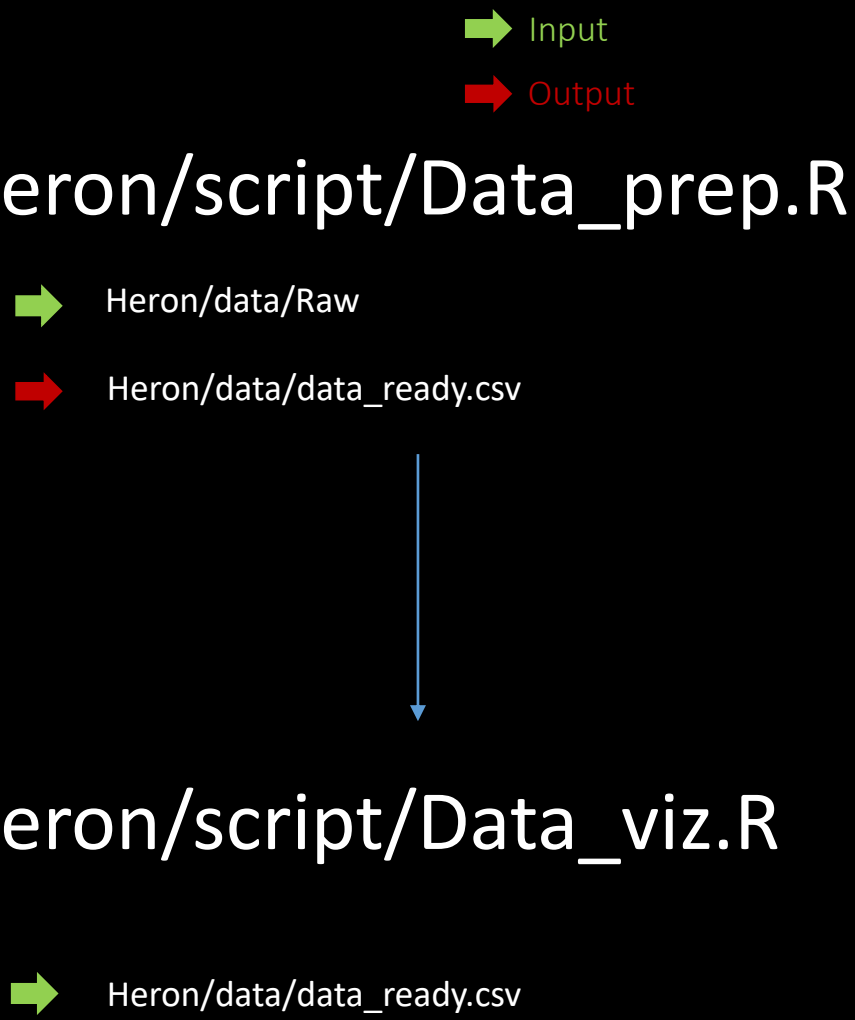
- Learn about the data
- Check points
- Happy about the format
- Ensure that everything makes sense to you

→ Save new table calls “data_ready.csv”

Modelling workflow
STEP2: Data visualization

→ Read data_ready

- Plot data through time



Modelling workflow
STEP2: Data visualization

Input

Output

Heron/script/Data_prep.R

Heron/data/Raw

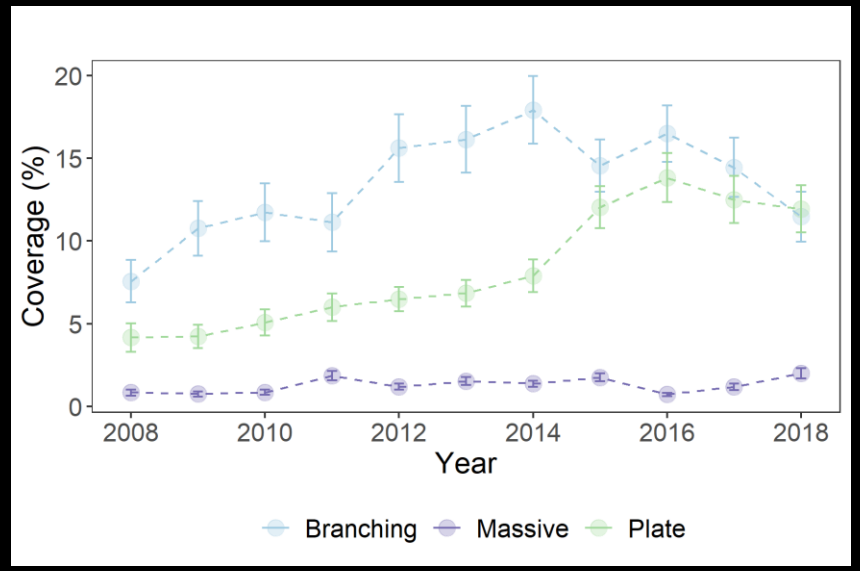
Heron/data/data_ready.csv

→ Read data_ready

- Plot data through time

Heron/script/Data_viz.R

Heron/data/data_ready.csv




```
> ggsave(plot = pheron,width=6, height=4, file =
"Heron/Figure/Traj_heron_full_coral2.png")
```

Vercelloni et al. in review

Modelling workflow
STEP2: Data visualization

Input

Output

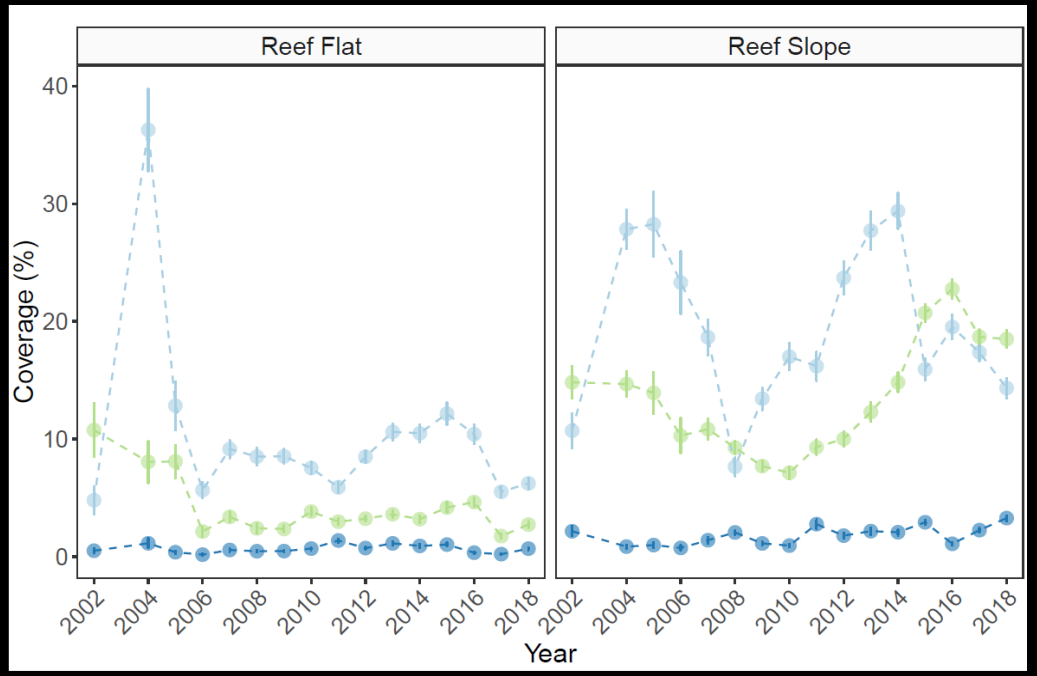
Heron/script/Data_prep.R

Heron/data/Raw

Heron/data/data_ready.csv

→ Read data_ready

- Plot data through time



Heron/script/Data_viz.R

Heron/data/data_ready.csv

Modelling workflows
STEP2: Data visualization

Input

Output

Heron/script/Data_prep.R

Heron/data/Raw

Heron/data/data_ready.csv

→ Read data_ready

- Plot data through time
- Plot data through space

Heron/script/Data_viz.R

Heron/data/data_ready.csv

Modelling workflow
STEP2: Data visualization

Input

Output

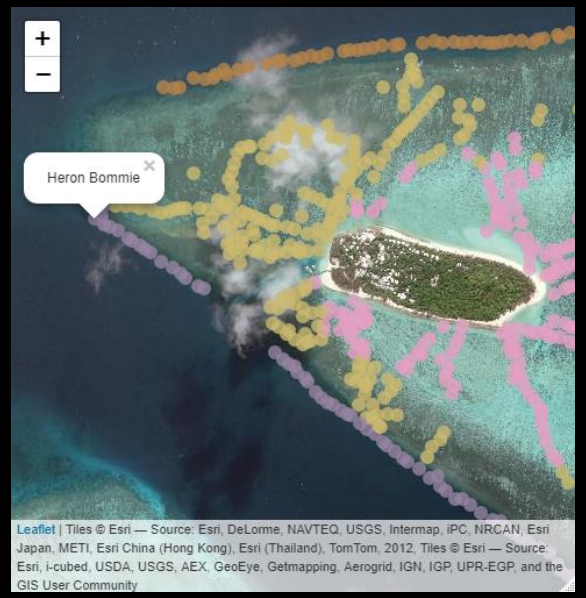
Heron/script/Data_prep.R

Heron/data/Raw

Heron/data/data_ready.csv

→ Read data_ready

- Plot data through time
- Plot data through space




```
> library(leaflet)
> library(mapview)
```

Heron/script/Data_viz.R

Heron/data/data_ready.csv

Modelling workflow
STEP2: Data visualization

Input

Output

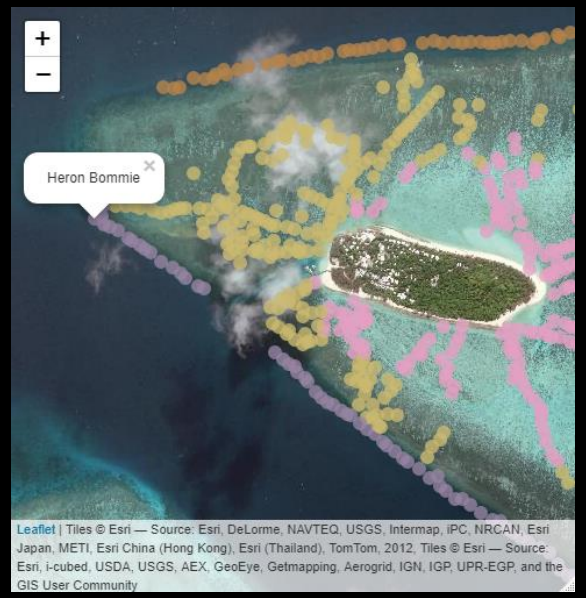
Heron/script/Data_prep.R

Heron/data/Raw

Heron/data/data_ready.csv

→ Read data_ready

- Plot data through time
- Plot data through space

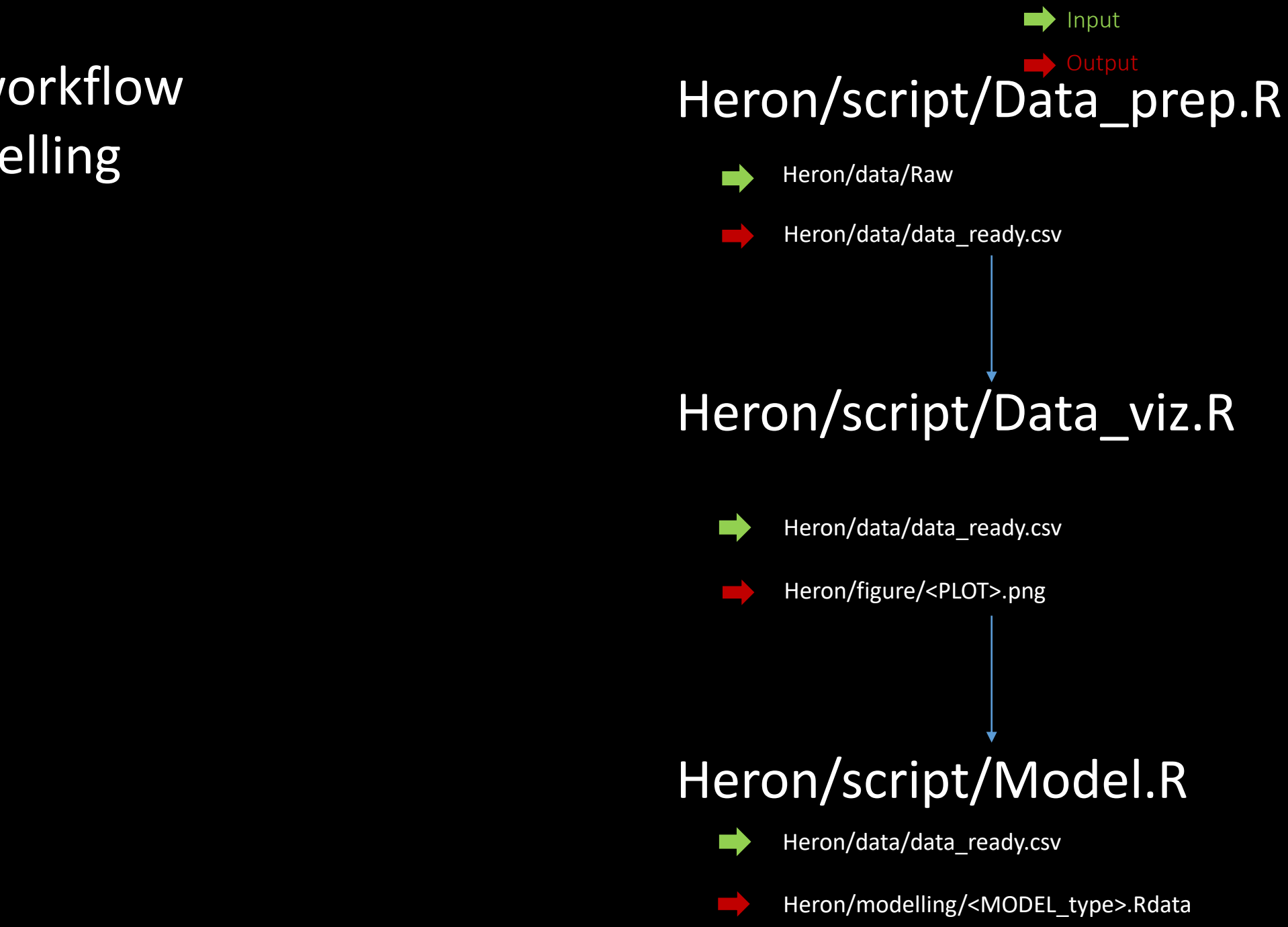


Heron/script/Data_viz.R

Heron/data/data_ready.csv

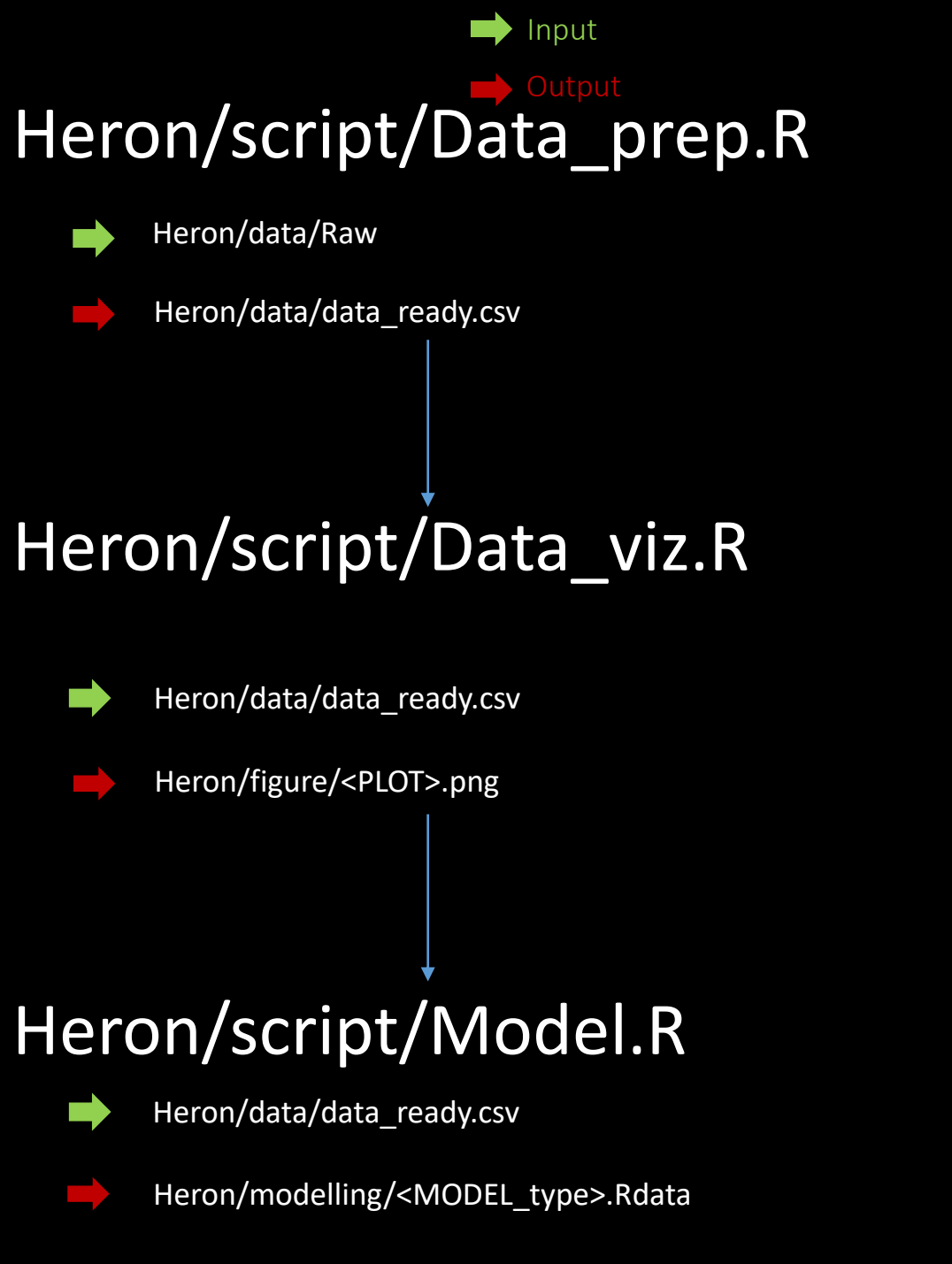
Heron/figure/<PLOT>.png

Modelling workflow
STEP3: Modelling

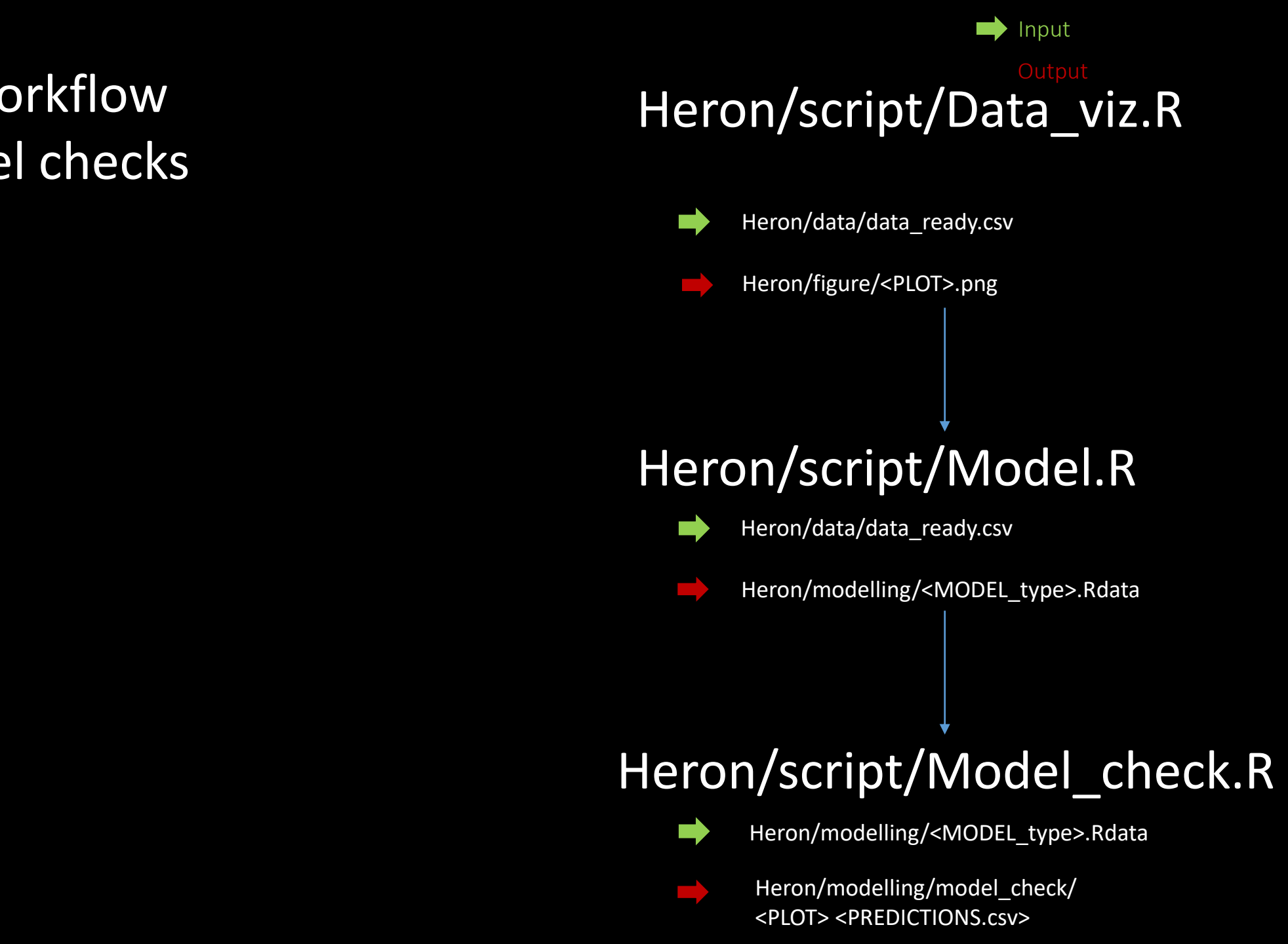


Modelling workflow
STEP3: Modelling

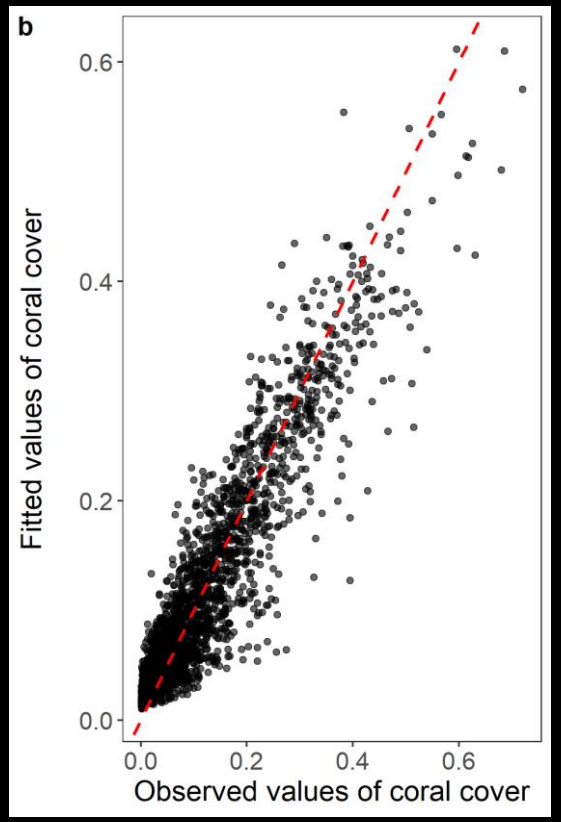
```
###### Run on HPC
> rm(list=ls())

library(INLA)

load("./Heron/Data_prep.Rdata")

rprior<- list(theta = list(prior = "pccor1", param =
c(0,0.9)))

formula <- y ~ -1 + b0 + Geomorph + f(s, model=spde,
group=s.group, + control.group =
list(model="ar1",hyper=rprior)) + f(iidx, model = "iid")

res<-inla()

save.image("./Heron/modelling/Recovery_Plate_model_nested
_output_FINAL.Rdata")
```



Modelling workflow
STEP4: Model checks



Modelling workflow
STEP4: Model checks



Vercelloni et al. in review

Input

Output

Heron/script/Data_viz.R

- Heron/data/data_ready.csv
- Heron/figure/<PLOT>.png

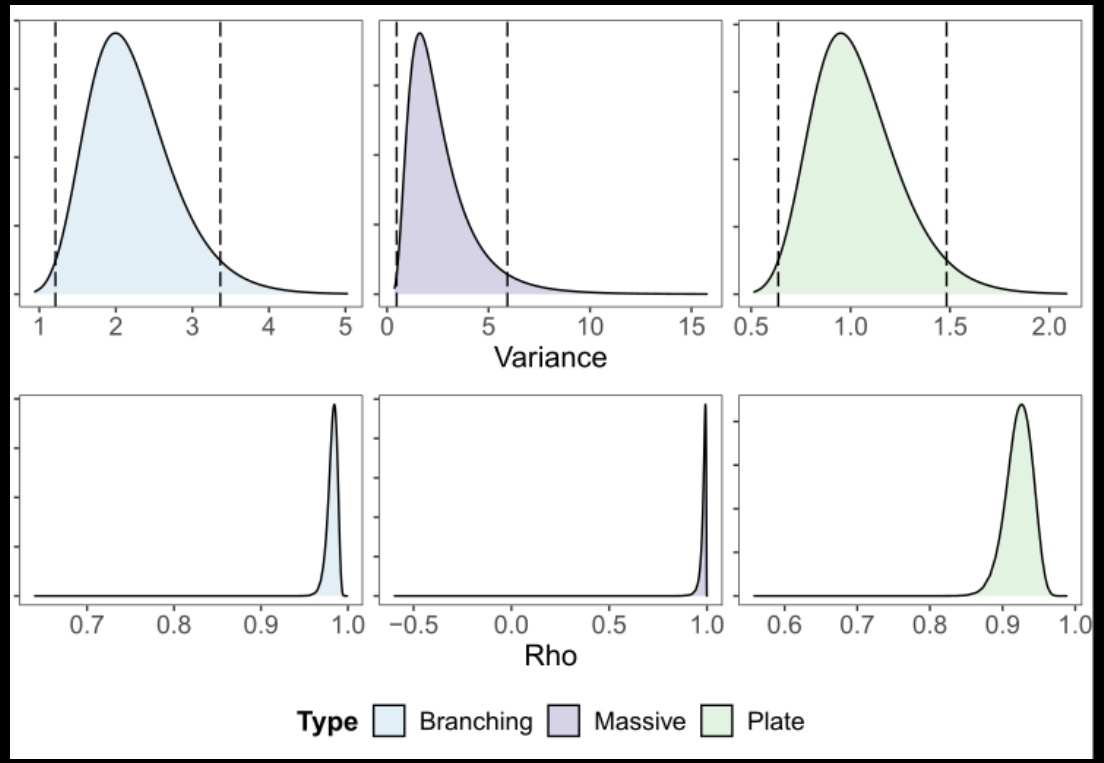
Heron/script/Model.R

- Heron/data/data_ready.csv
- Heron/modelling/<MODEL_type>.Rdata

Heron/script/Model_check.R

- Heron/modelling/<MODEL_type>.Rdata
- Heron/modelling/model_check/ <PLOT> <PREDICTIONS.csv>

Modelling workflow
STEP4: Model checks



Vercelloni et al. in review

Input

Output

Heron/script/Data_viz.R

- Heron/data/data_ready.csv
- Heron/figure/<PLOT>.png

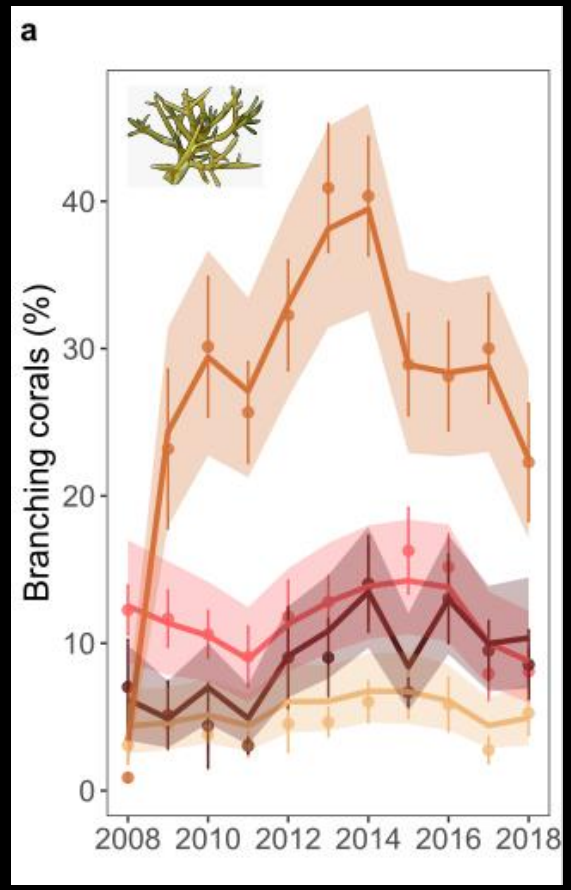
Heron/script/Model.R

- Heron/data/data_ready.csv
- Heron/modelling/<MODEL_type>.Rdata

Heron/script/Model_check.R

- Heron/modelling/<MODEL_type>.Rdata
- Heron/modelling/model_check/
  <PLOT> <PREDICTIONS.csv>

Modelling workflow
STEP4: Model checks



Vercelloni et al. in review

Input

Output

Heron/script/Data_viz.R

- Heron/data/data_ready.csv
- Heron/figure/<PLOT>.png

Heron/script/Model.R

- Heron/data/data_ready.csv
- Heron/modelling/<MODEL_type>.Rdata

Heron/script/Model_check.R

- Heron/modelling/<MODEL_type>.Rdata
- Heron/modelling/model_check/ <PLOT> <PREDICTIONS.csv>

Modelling workflow Summary

```
|   Heron
|
+---data
|   \---Raw
+---figure
+---modelling
|   \---model_check
\---script
```

Modelling workflow Summary

```
|   Heron
|
+---data
|   \---Raw
+---figure
+---modelling
|   \---model_check
\---script
```

```
\---script
```

Heron/script/Data_prep.R

- Heron/data/Raw
- Heron/data/data_ready.csv

Heron/script/Model.R

- Heron/modelling/<MODEL_type>.Rdata
- Heron/modelling/model_check/ <PLOT> <PREDICTIONS.csv>

Heron/script/Data_viz.R

- Heron/data/data_ready.csv
- Heron/figure/<PLOT>.jpeg

Heron/script/Model_check.R

- Heron/modelling/<MODEL_type>.Rdata
- Heron/modelling/model_check/ <PLOT> <PREDICTIONS.csv>

Modelling workflow
STEP3: Modelling

```
###### Run on HPC
> rm(list=ls())

library(INLA)

load("./Heron/Data_prep.Rdata")

rprior<- list(theta = list(prior = "pccor1", param =
c(0,0.9)))

formula <- y ~ -1 + b0 + Geomorph + f(s, model=spde,
group=s.group, + control.group =
list(model="ar1",hyper=rprior)) + f(iidx, model = "iid")

res<-inla()

save.image("./Heron/modelling/Recovery_Plate_model_nested
_output_FINAL.Rdata")
```

Input

Output

Heron/script/Data_prep.R

- Heron/data/Raw
- Heron/data/data_ready.csv

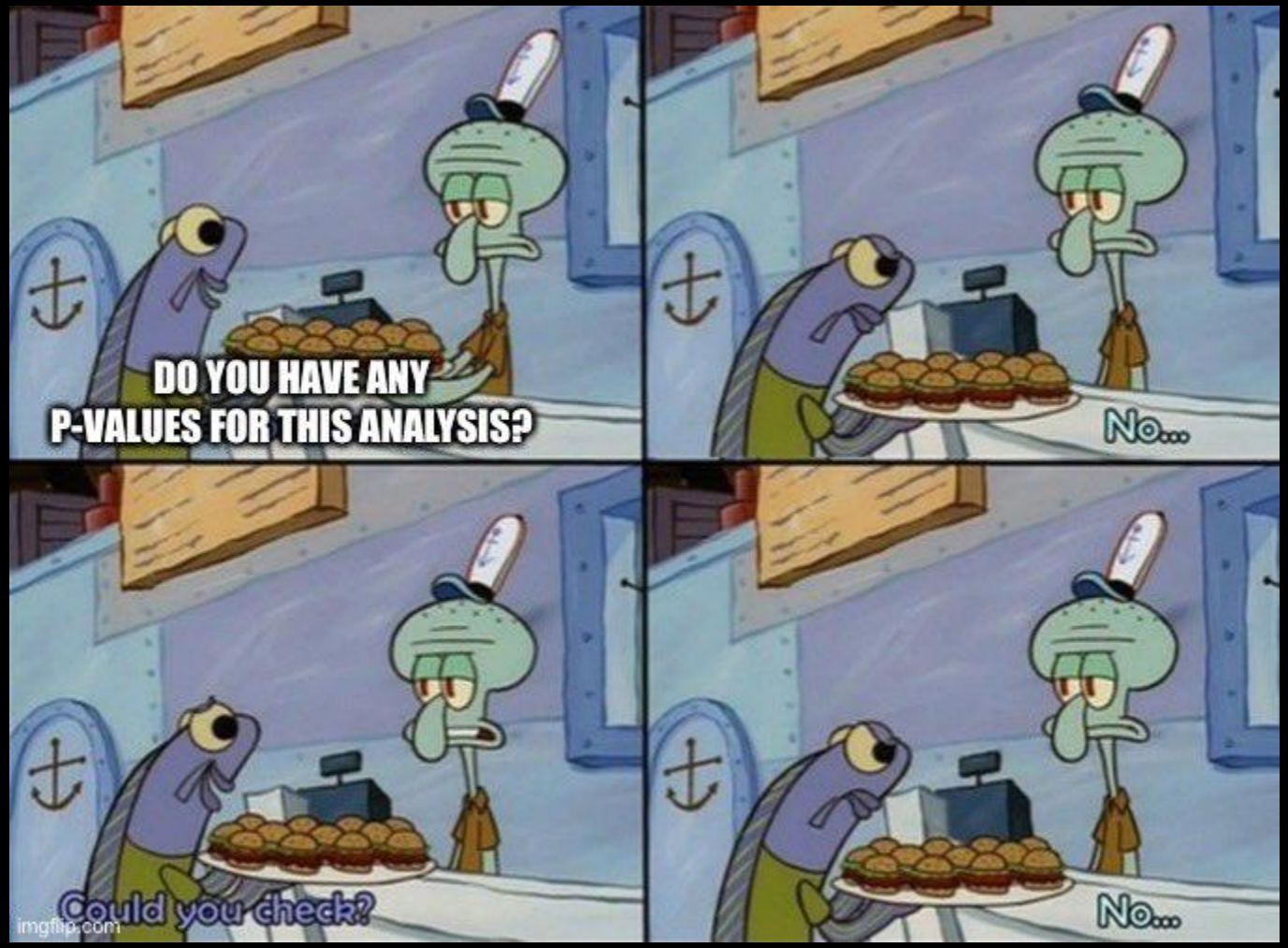
Heron/script/Data_viz.R

- Heron/data/data_ready.csv
- Heron/figure/<PLOT>.jpeg

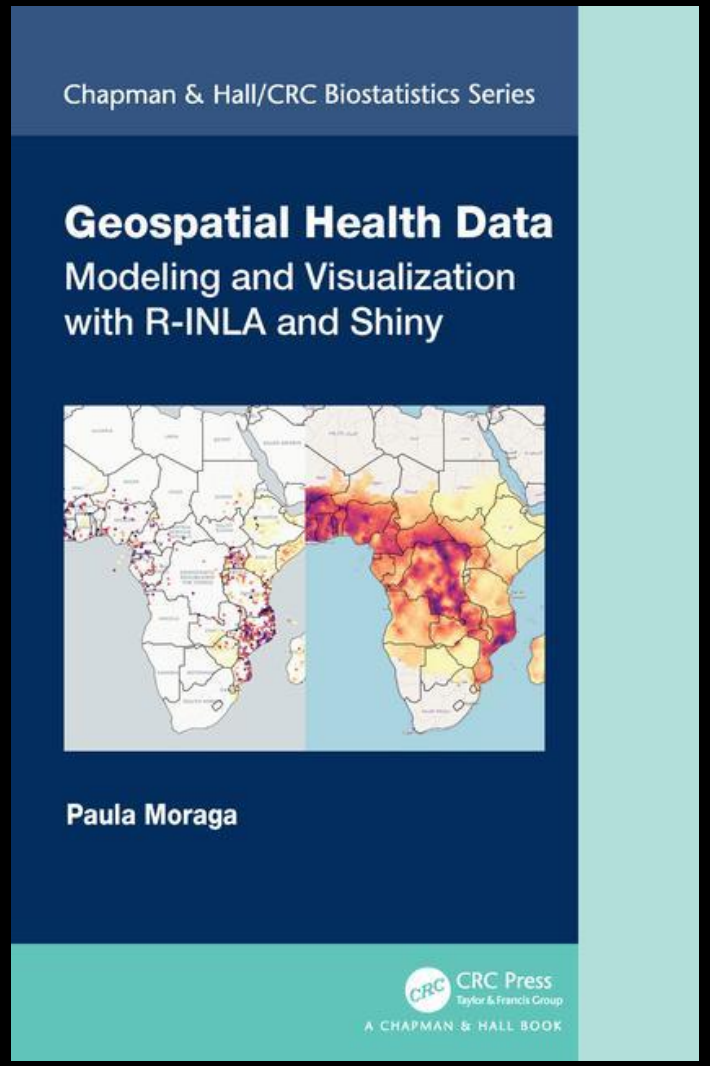
Heron/script/Model.R

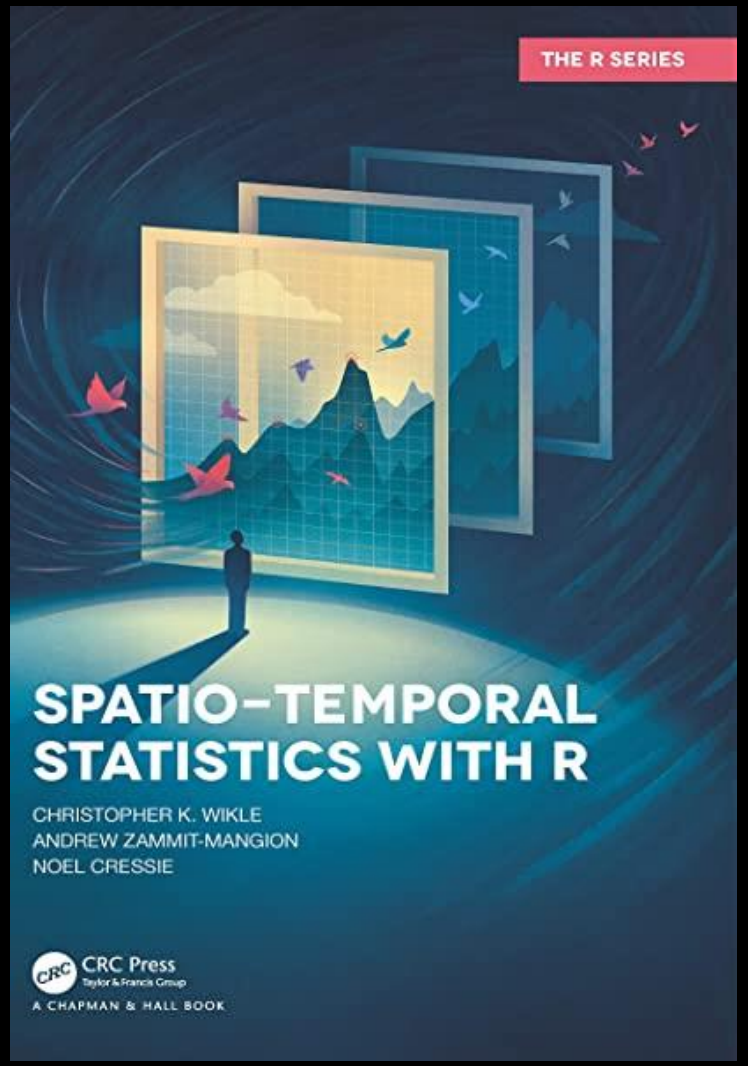
- Heron/modelling/<MODEL_type>.Rdata
- Heron/modelling/model_check/
  <PLOT> <PREDICTIONS.csv>

R-INLA
Integrated Nested Laplace Approximation (INLA)



R-INLA
Integrated Nested Laplace Approximation (INLA)





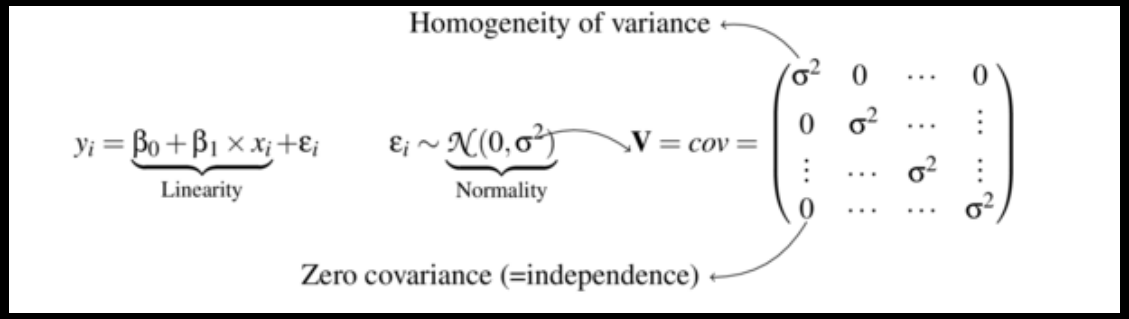
- https://www.paulamoraga.com/book-geospatial/
- https://becarioprecario.bitbucket.io/inla-gitbook/index.html
- https://www.seascapemodels.org/rstats/2017/02/22/spatial-statistics-photos.html
- https://www.seascapemodels.org/rstats/2017/06/18/estimating-popn-decline.html

R-INLA
Integrated Nested Laplace Approximation (INLA)

Spatial autocorrelation

- Tobler's (1970) first law of geography: everything is related to everything else, but nearby things are more related than distant things

- Linear model equation – non-independent errors in model residuals

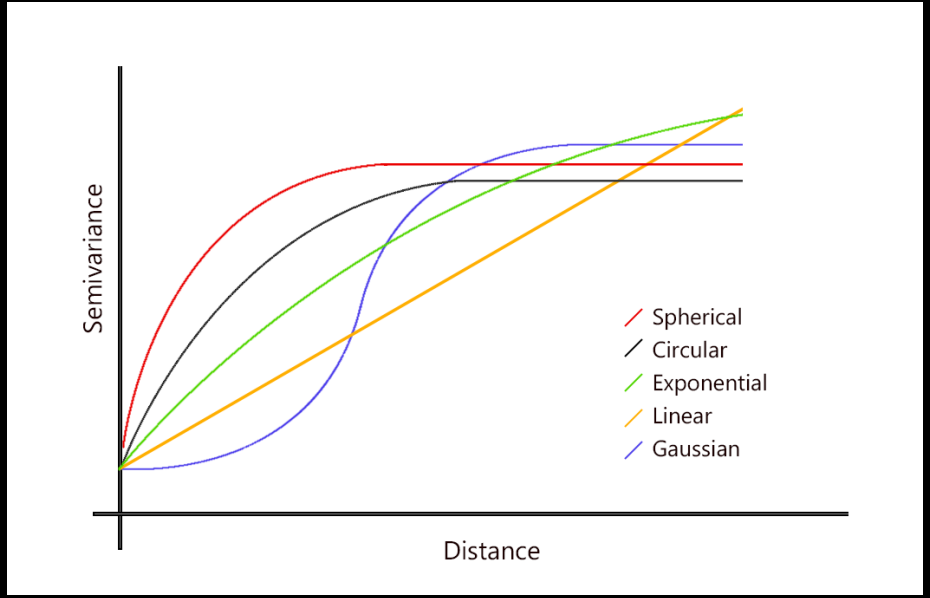


https://www.flutterbys.com.au/stats/tut/tut9.1.html

R-INLA
Integrated Nested Laplace Approximation (INLA)

Spatial autocorrelation

- Tobler's (1970) first law of geography: everything is related to everything else, but nearby things are more related than distant things
- Linear model equation – non-independent errors in model residuals
- Geostatistics – semivariogram (spatial field)

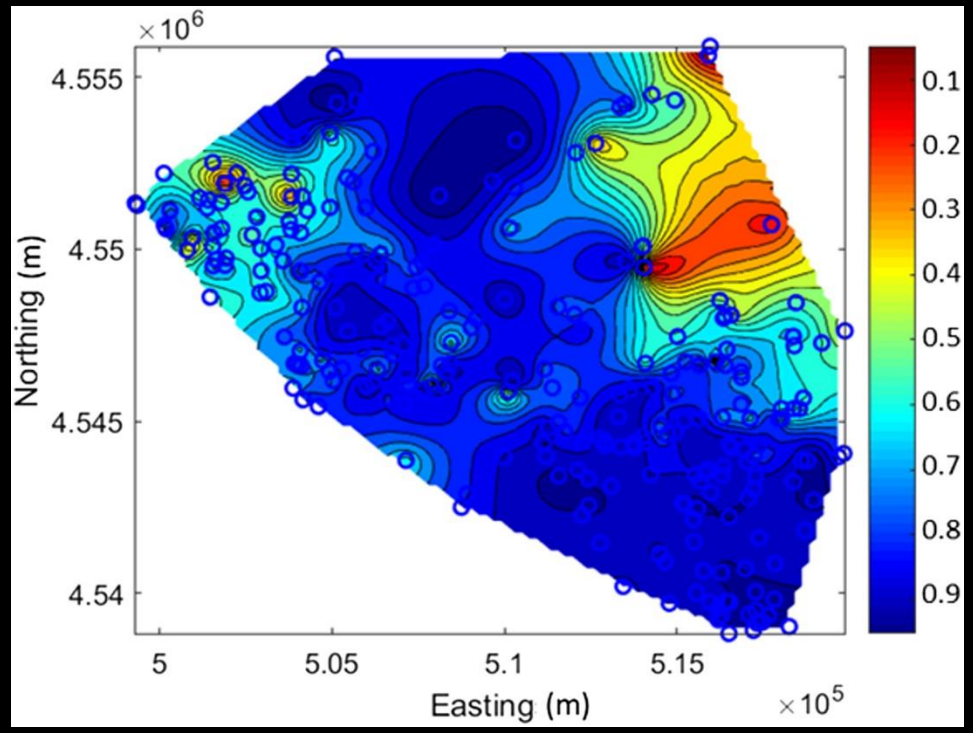


https://gisgeography.com/kriging-interpolation-prediction/

R-INLA
Integrated Nested Laplace Approximation (INLA)

Spatial autocorrelation

- Tobler's (1970) first law of geography: everything is related to everything else, but nearby things are more related than distant things
- Linear model equation – non-independent errors in model residuals
- Geostatistics – semivariogram (spatial field)
- Kriging



https://ascelibrary.org/doi/10.1061/%28ASCE%29HZ.2153-5515.0000464

Modelling Spatial autocorrelation with R-INLA

Spatial model for coral recovery

$$y_{it} \sim Beta(\mu(s_i,t),\phi)$$

$$logit(\mu(s_i,t) = \beta_0 + Geomorphic_i \times \beta_1 + \zeta(s_i,t) + \varepsilon_{it} \quad (1)$$

$$\zeta(s_i,t) = a\zeta(s_i,t-1) + \omega(s_i,t)$$

Spatio-temporal random effect

- Use stochastic partial differential equation (SPDE) to model the spatial field
- Major aims to learn about spatial autocorrelation during the recovery process and recovery trajectories across the entire reef

https://www.paulamoraga.com/book-geospatial/sec-geostatisticaldatatheory.htm

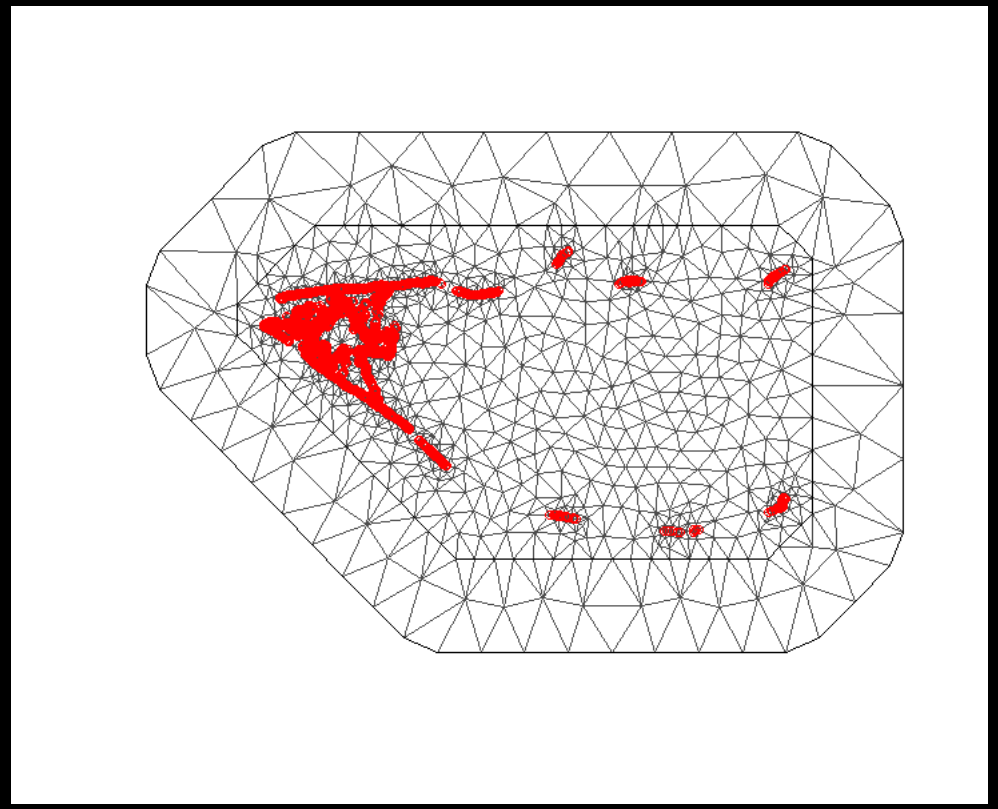
Modelling Spatial autocorrelation with R-INLA

Spatial model for coral recovery

- Major aims to learn about spatial autocorrelation during the recovery process and recovery trajectories across the entire reef

- Create a mesh to approximate the continuous spatial field

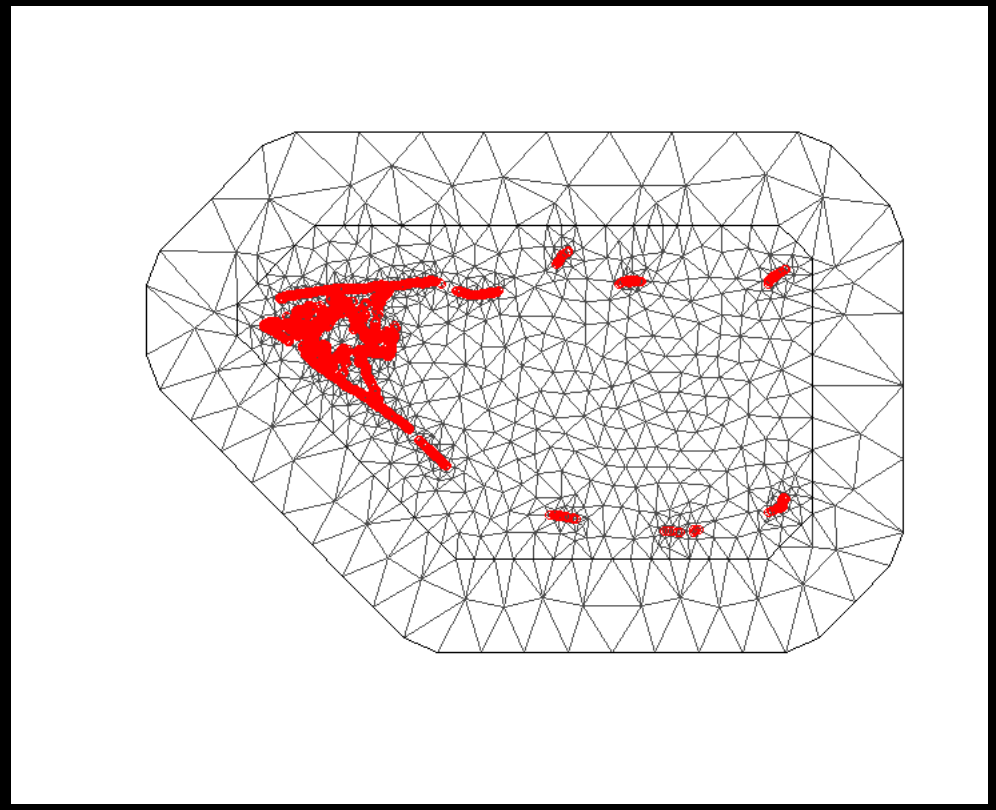
```
> mesh = inla.mesh.2d(loc=cbind(df$lng, df$lat),
>  max.edge = c(1,5)*max.edge,
> cutoff = max.edge/20,
>  offset = NULL)
```



Modelling Spatial autocorrelation with R-INLA

Spatial model for coral recovery

- Major aims to learn about spatial autocorrelation during the recovery process and recovery trajectories across the entire reef
- Create a mesh to approximate the continuous spatial field
- Create objects needed for the spde



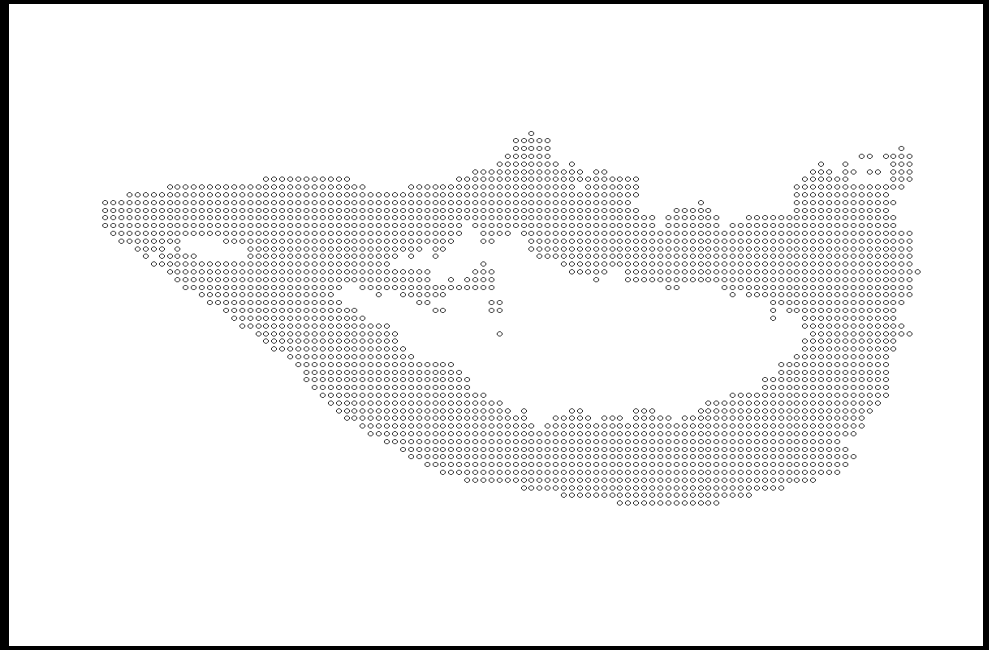
```
spde = inla.spde2.pcmatern(mesh, alpha=2, prior.range = c(.01, .01),
prior.sigma = c(3, 0.01))
```

```
indexs <- inla.spde.make.index ("s", n.spde = spde$n.spde
A <- inla.spde.make.A(mesh = mesh, loc=cbind(df$lng, df$lat))
```

Modelling Spatial autocorrelation with R-INLA

Spatial model for coral recovery

- Major aims to learn about spatial autocorrelation during the recovery process and recovery trajectories across the entire reef
- Create a mesh to approximate the continuous spatial field
- Create objects needed for the spde



```
Ap<- inla.spde.make.A(mesh = mesh, loc= cbind(pred.df[,1], pred.df[,2]))
```

Modelling Spatial autocorrelation with R-INLA

Spatial model for coral recovery

- Major aims to learn about spatial autocorrelation during the recovery process and recovery trajectories across the entire reef
- Create a mesh to approximate the continuous spatial field
- Create objects needed for the spde
- Stack everything

```
stk.e <- inla.stack(data=list(y=df$g.mean),
A=list(A,1,1,1), effects=list(c(indexs, list(b0=1)),
list(Geomorph=df[ ,"Geomorph"]),
list(iidx=df$idsub),list(idyear=df$idyear)), tag="est")


stk.p <- inla.stack(data=list(y=NA), A=list(Ap,1,1,1),
effects=list(c(indexs, list(b0=1)), +
list(Geomorph=pred.df[ ,"Geomorph"]),
list(iidx=pred.df$idsub), list(idyear=pred.df$group)),
tag="pred")

stk.full <- inla.stack(stk.e,stk.p)
```

```
res<-inla(formula, family="beta",data=inla.stack.data(stk.full), +
control.predictor=list(A=inla.stack.A(stk.full),compute=T, link=1), verbose=TRUE,
+ control.results=list(return.marginals.random=TRUE, +
return.marginals.predictor=TRUE), num.threads = 2, +
control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE))
```

Modelling Spatial autocorrelation with R-INLA

Spatial model for coral recovery

- Major aims to learn about spatial autocorrelation during the recovery process and recovery trajectories across the entire reef
- Create a mesh to approximate the continuous spatial field
- Create objects needed for the spde
- Stack everything
- Extract model outputs

```
➢ summary(res_nested)
```

```
➢ library(INLAutils)

➢ # Extract the information on the spatial field
➢  Mod_p2.field <- inla.spde2.result(inla = res_nested, name
  = "s",  spde = spde,  do.transf = T)
```

```
➢ ######################## Plot fitted trajectories
➢ index.est <- inla.stack.index(stack = stk.full,tag="est")$data

➢ df$pred<-res_nested$summary.fitted.values$mean[index.est]
```

```
> index.pred <- inla.stack.index(stack =
stk.full,tag="pred")$data
```

HAPPY WOMEN IN MATHS DAY TO EVERYONE!

https://acems.org.au/women-maths-day-2021