Accounting for the spatio-temporal spread of a fungal invasion

SPDE/INLA workshop

Facundo Muñoz
facundo.munoz@cirad.fr
🐦 🐙 famuvie

cirad

https://famuvie.gitlab.io/spdeinla_workshop_2018

Avignon, Nov. 7-9, 2018

Arnaud Dowkiw

Research officer

Facundo Muñoz

Biostatistician

Ash-dieback

1996

© INRA, A.Dowkiw

Un pays touché, même partiellement, apparait en rouge

Symptoms
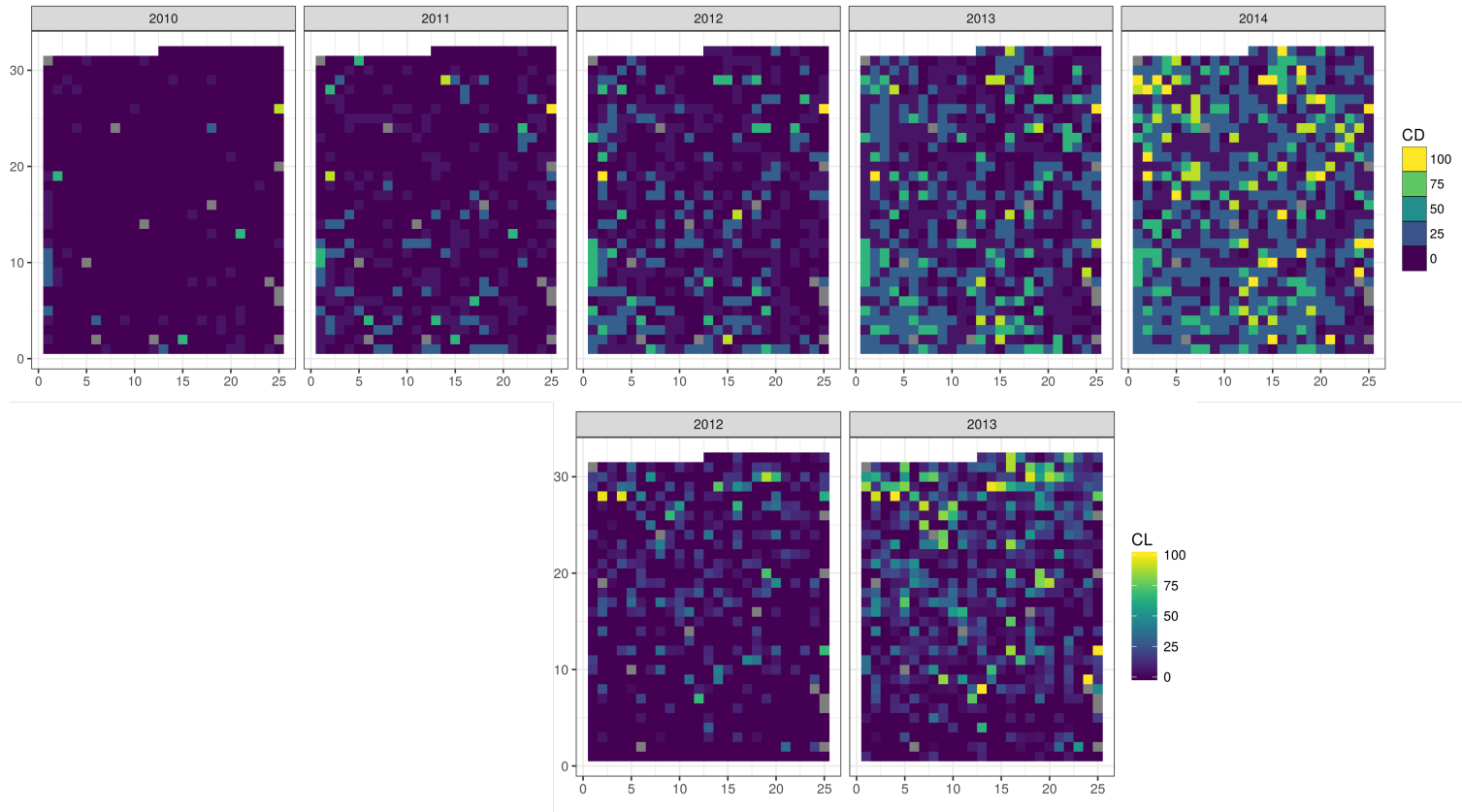
Crown-defoliation (CD)

Collar-lesion (CL)

Study goals

Can we **breed** (more) resilient trees?

- (i.e. is there sufficient **genetic variation** for the **resistance** trait?)

Dataset

Some modelling decisions

- Model both symptoms **separately**

Some modelling decisions

- Model both symptoms **separately**

- Use a Gaussian approximation for the (discrete) variable CD (after transformation)

Some modelling decisions

- Model both symptoms **separately**

- <small>Use a Gaussian approximation</small> for the (discrete) variable CD (after transformation)

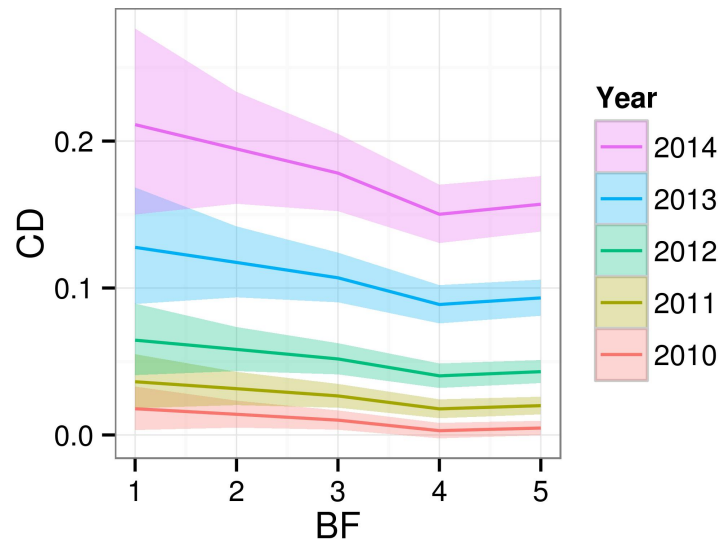- Use a **hurdle** model for CL

Model for (T)CD

Response in year i, location j for the individual k:

$$y_{ijk} = \mathrm{Year}_i + \mathrm{BF} + \eta_{ij} + a_k + \varepsilon_{ijk}$$
$$\boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \tau_\eta^2 \mathbf{Q}(\rho_t, \rho_s)\right)$$
$$\boldsymbol{a} \sim \mathcal{N}\left(\mathbf{0}, \tau_a^2 \mathbf{A}^{-1}\right)$$
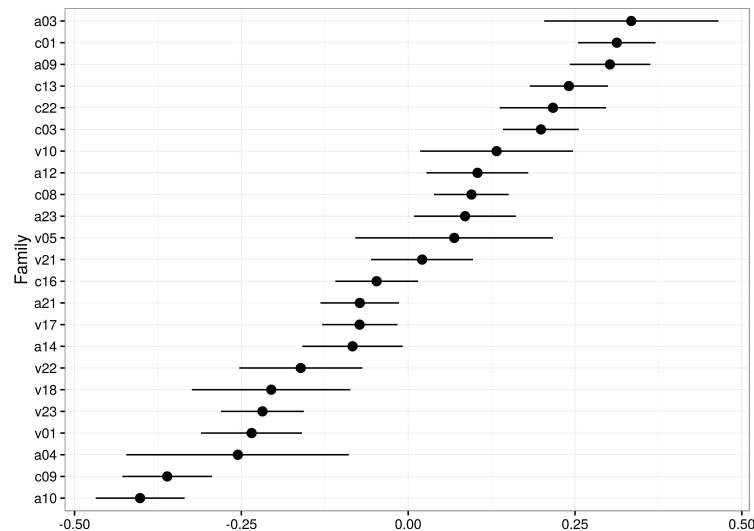$$\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right),$$

Model for (T)CD

Response in year i, location j for the individual k:

$$y_{ijk} = \text{Year}_i + \text{BF} + \eta_{ij} + a_k + \varepsilon_{ijk}$$
$$\boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \tau_\eta^2 \mathbf{Q}(\rho_t, \rho_s)\right)$$
$$\boldsymbol{a} \sim \mathcal{N}\left(\mathbf{0}, \tau_a^2 \mathbf{A}^{-1}\right)$$
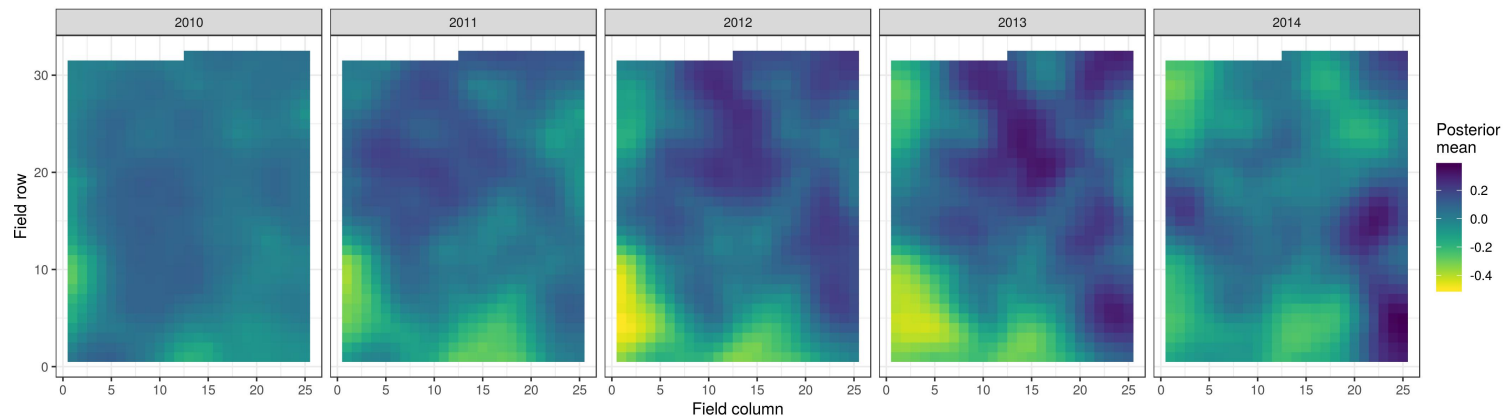$$\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right),$$

Model for (T)CD

Response in year i, location j for the individual k:

$$y_{ijk} = \text{Year}_i + \text{BF} + \eta_{ij} + a_k + \varepsilon_{ijk}$$
$$\boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \tau_\eta^2 \mathbf{Q}(\rho_t, \rho_s)\right)$$
$$\boldsymbol{a} \sim \mathcal{N}\left(\mathbf{0}, \tau_a^2 \mathbf{A}^{-1}\right)$$
$$\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right),$$

Spatiotemporal structure

$$\boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \tau_\eta^2 \mathbf{Q}(\rho_t, \rho_s)\right)$$

- The parametric structure matrix $\mathbf{Q}(\rho_t, \rho_s)$ is the tensor product of a **Matérn** spatial process and a **exchangeable** temporal structure by year

```
f(eta, model=spde,
  group = eta.group,
  control.group = list(model = 'exchangeable'))
```

Spatiotemporal structure

$$\boldsymbol{\eta} \sim \mathcal{N}\big(\mathbf{0}, \tau_\eta^2 \mathbf{Q}(\rho_t, \rho_s)\big)$$

- The parametric structure matrix $\mathbf{Q}(\rho_t, \rho_s)$ is the tensor product of a **Matérn** spatial process and a **exchangeable** temporal structure by year
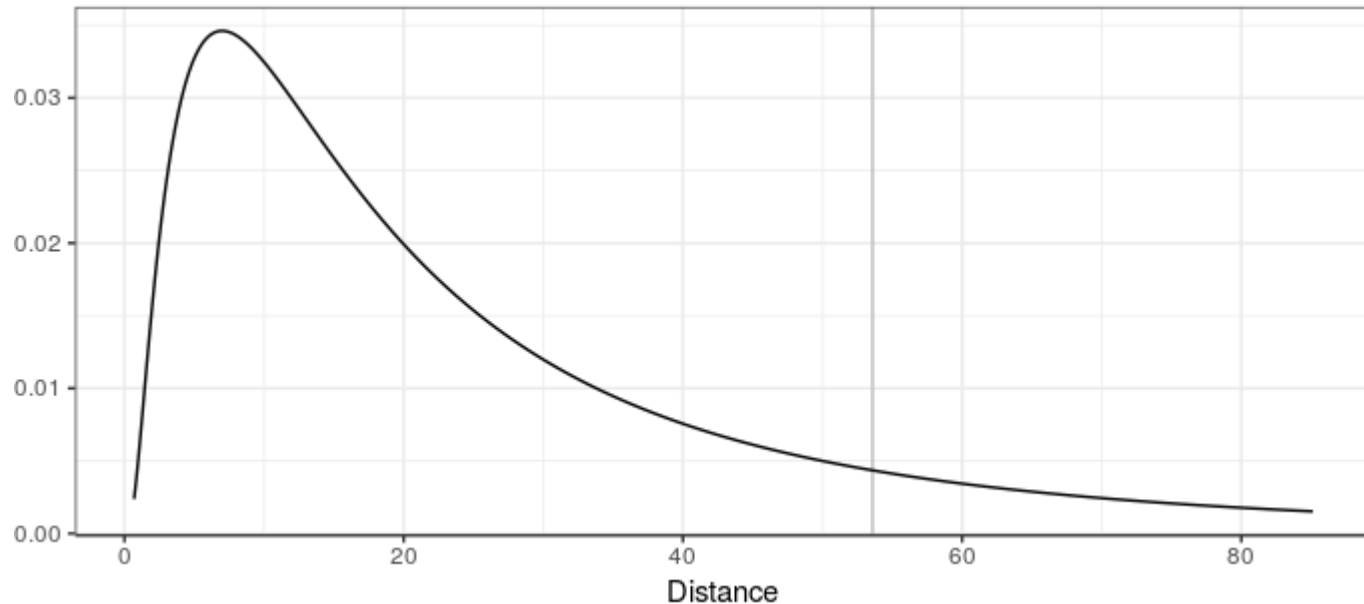
```
f(eta, model=spde,
  group = eta.group,
  control.group = list(model = 'exchangeable'))
```

- The spatial effect need to **integrate to 0** to separate it from the Year effect

```
spde = inla.spde2.matern(..., constr = TRUE)
```
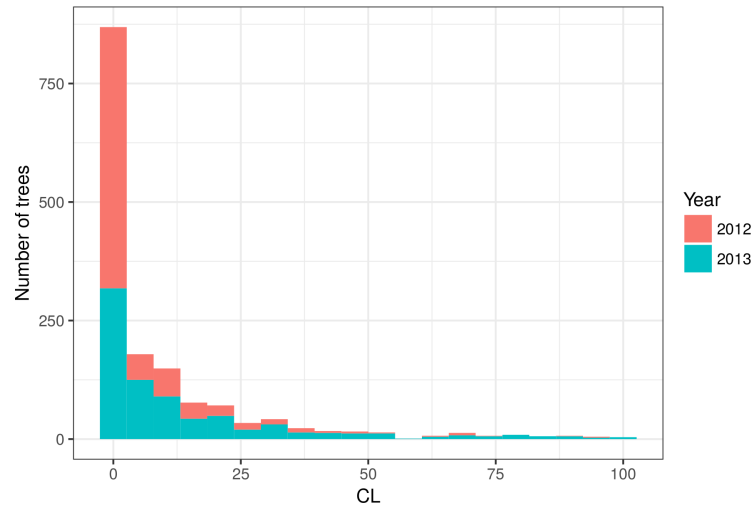
Priors

- Prior spatial range constrained within the field dimensions.

- Other priors are *reasonably* vague. See paper for details.

Model for CL

For a measurement of CL y_{ijk} taken in year i, at location j for individual k, we assume that

$$\Pr[y_{ijk} = 0] = p_{ijk}, \qquad\qquad 0 < p_{ijk} < 1$$
$$\pi(y_{ijk}|y_{ijk} > 0) = \mathrm{Ga}(a_{ijk}, b_{ijk}), \qquad a_{ijk}, b_{ijk} > 0.$$
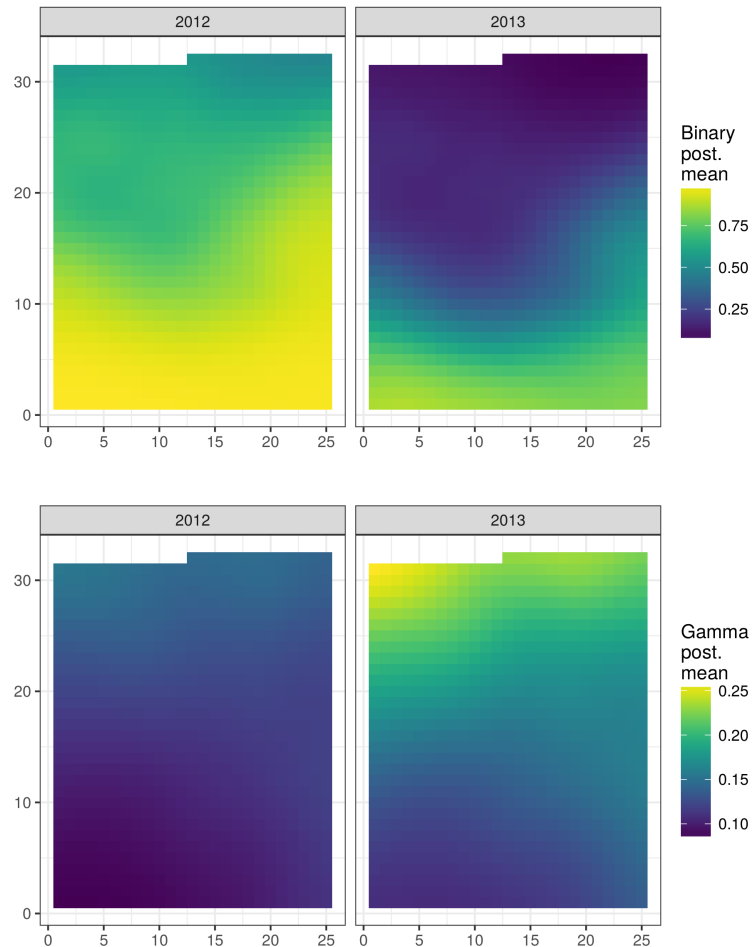
Latent models

calling $\mu = E(y|y > 0) = \frac{a}{b}$, we define two linear predictors

$$\text{logit}(p_{ijk}) = \text{Year}_i^{(1)} + \eta_{ij}^{(1)} + a_k^{(1)}$$
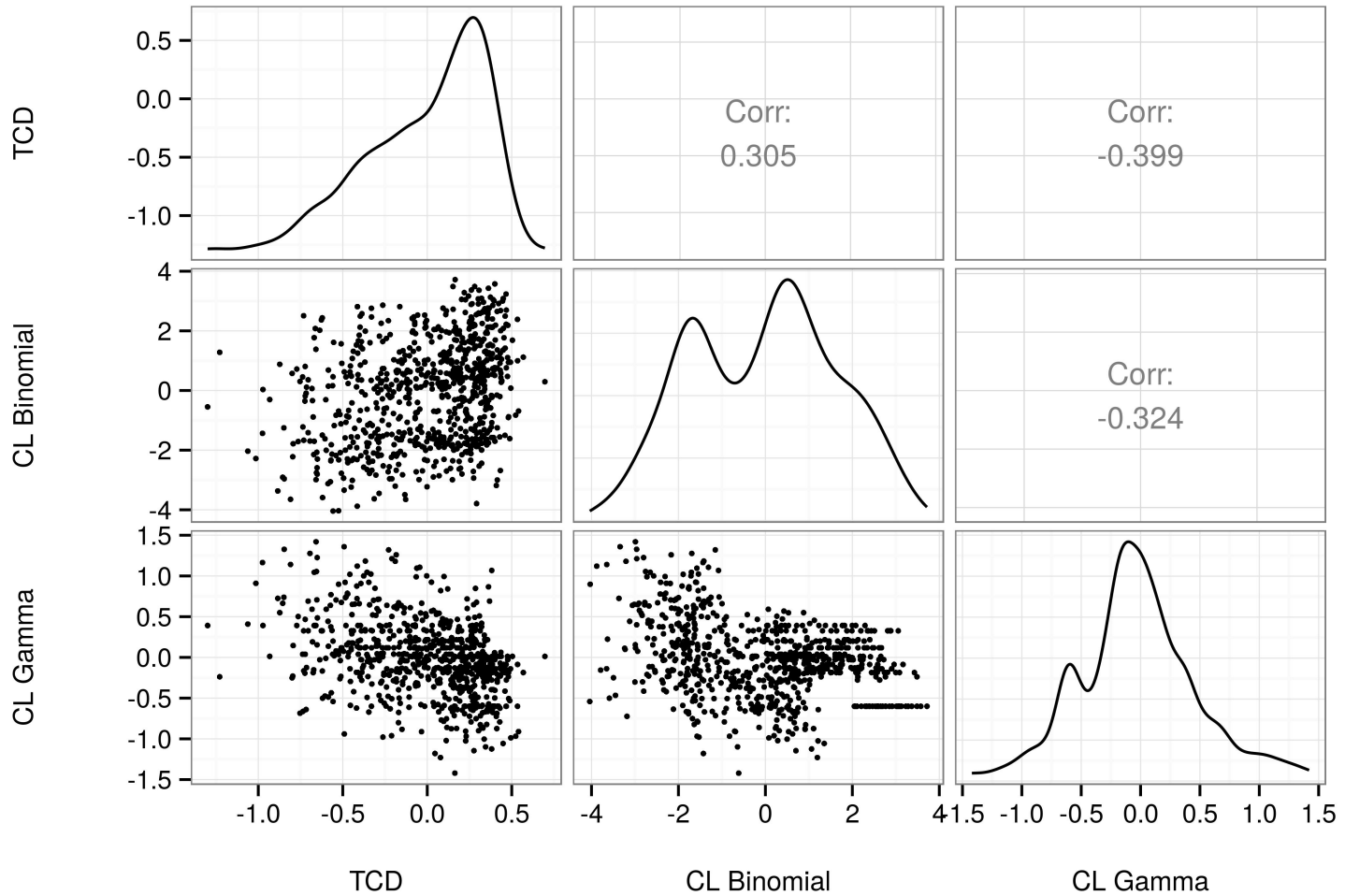$$\log(\mu_{ijk}) = \text{Year}_i^{(2)} + \eta_{ij}^{(2)} + a_k^{(2)}$$

for the binary and continuous components, respectively. The elements in the latent linear predictors are defined as for CD

Posterior mean spatio-temporal effect for CL

Genetic correlations

Conclusions

The spatio-temporal modelling of the disease spread allowed to:

- Control for **differences in exposure** to the pathogen and better identify the **genetic effects**

Conclusions

The spatio-temporal modelling of the disease spread allowed to:

- Control for **differences in exposure** to the pathogen and better identify the **genetic effects**

- **Select** more resilient genotypes and **quantify** the expected improvements in tolerance

Conclusions

The spatio-temporal modelling of the disease spread allowed to:

- Control for **differences in exposure** to the pathogen and better identify the **genetic effects**

- **Select** more resilient genotypes and **quantify** the expected improvements in tolerance

- Learn about the **different nature** of the processes causing both symptoms

Improvements

- Proper **ordered-logistic** model for CD

Improvements

- Proper **ordered-logistic** model for CD

- **Joint model** of both traits for proper inference on **genetic correlations**

Improvements

- Proper **ordered-logistic** model for CD

- **Joint model** of both traits for proper inference on **genetic correlations**

- Correlated genetic effect **across components** of CL

Improvements

- Proper **ordered-logistic** model for CD

- **Joint model** of both traits for proper inference on **genetic correlations**

- Correlated genetic effect **across components** of CL

- **Time-varying** variances

Thank you

References:

- **RisingAshes** R-package with **data** and **code**, + poster and paper:

    https://github.com/famuvie/2016_RisingAshes

- These slides: https://famuvie.gitlab.io/spdeinla_workshop_2018