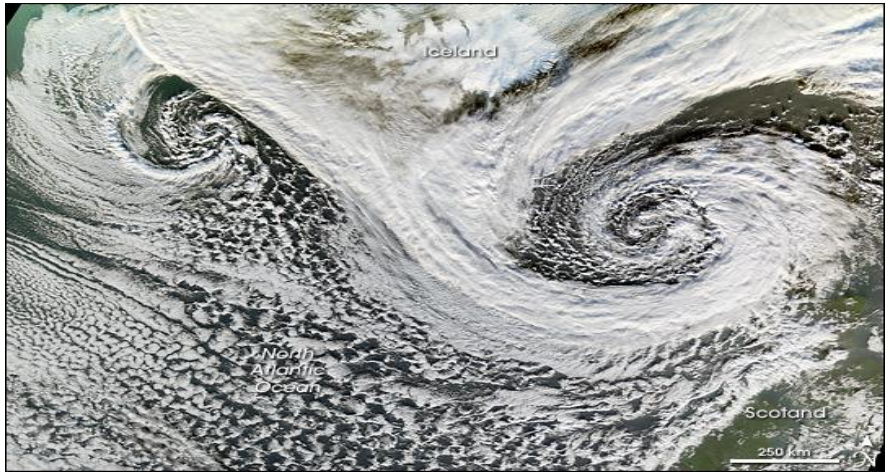
Probabilistic Climate Prediction from Multi-Model Ensembles

David Stephenson and Phil Sansom
Exeter Climate Systems

Acknowledgements:
Giuseppe Zappa, Len Shaffrey, Chris Ferro



ExIStA Forecasting Workshop, 12 June 2012, Exeter

NERC project TEMPEST:

Testing and Evaluating Model Predictions of European Storms

NCAS, Reading*
Len Shaffrey
Tim Woollings
Mike Blackburn

NCEO, Reading
Kevin Hodges

Oxford University
Tim Palmer

ECMWF (Partner)
Thomas Jung

Met Dept, Reading
Brian Hoskins
Helen Dacre

Exeter University
David Stephenson
Phil Sansom

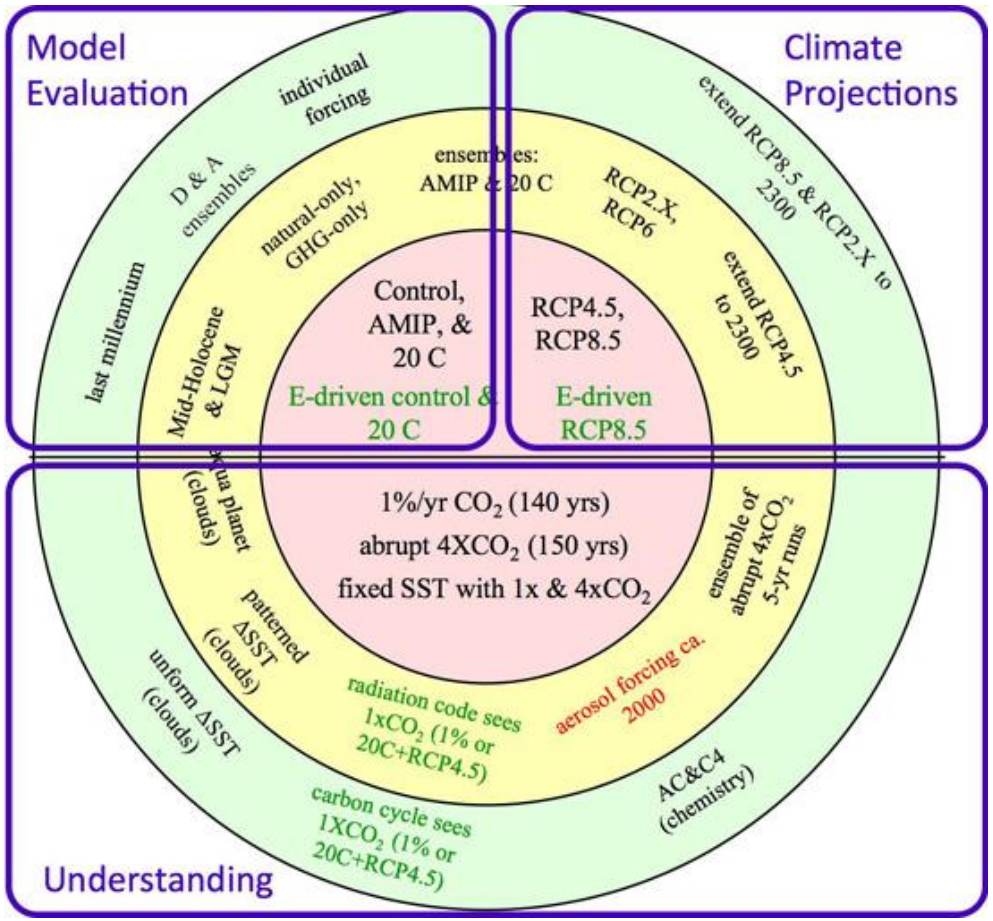
Met Office (Partner)
Simon Brown
Ruth McDonald

**Plus David Brayshaw as Co-I Researcher at Reading*

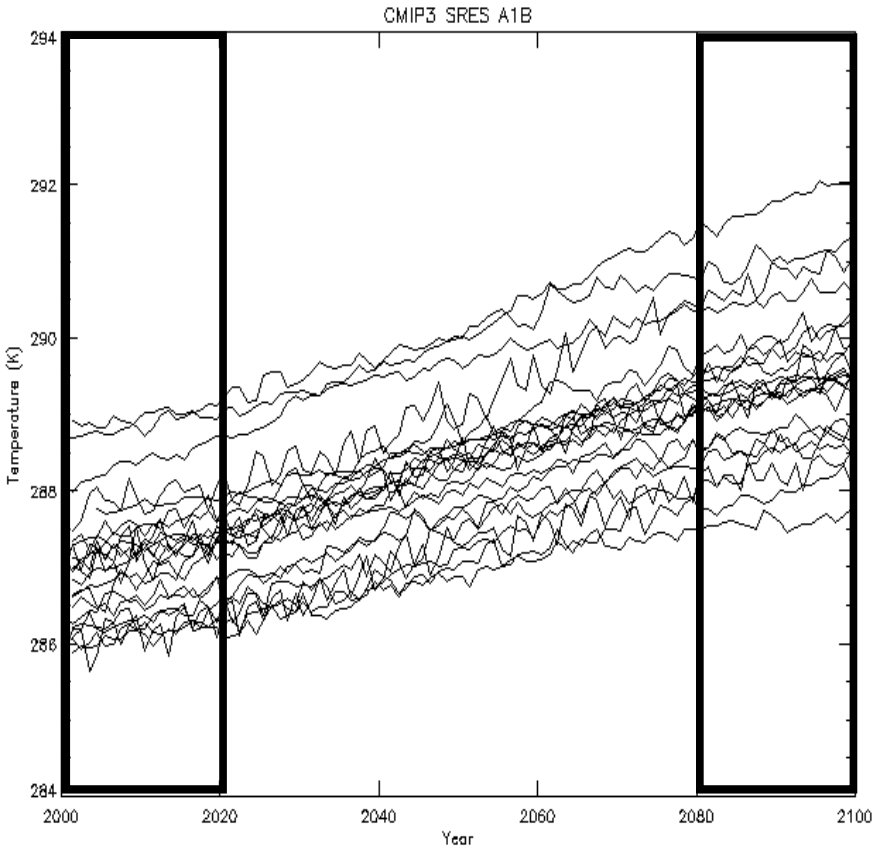
How will European storms respond to future climate change?

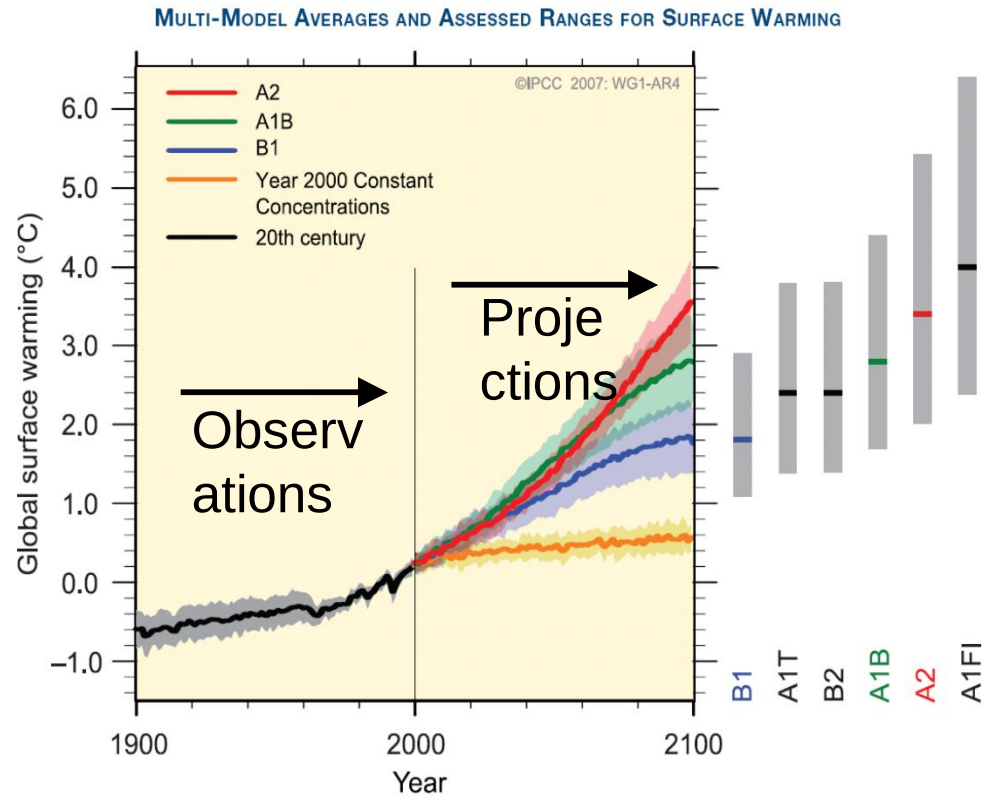
Need simulations from a set of future weather generators …

i.e. a coordinated set of climate model simulations
e.g. CMIP5 Coupled Model Intercomparison Project Phase 5



Multi-Model Ensembles (MME) → climate change projections

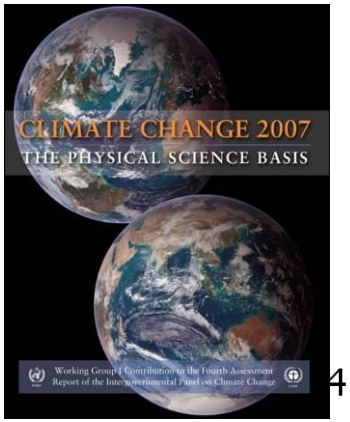




Collins et al. (2012)
Quantifying future climate change,
Nature Climate Change

ipcc
INTERGOVERNMENTAL PANEL ON climate change

Working Group I (WG I) – The Physical Science Basis



CMIP5 Multi-Model Ensemble (MME)

y_{msr} = 30-year wintertime (DJF) mean of track densities from run r=1,é ,R_{ms} of model m=1,é , M=14 under scenario s=Historical (1975-2004) or Future (2069-2098 RCP4.5 emissions scenario)

Climate model m	No. of runs R_{mH}	No. of runs R_{mF}	$w_m = R_{mH}R_{mF}/(R_{mH}+R_{mH})$
BCC-CSM1.1	3	1	0.75
CanESM2	5	1	0.83
CNRM-CM5	5	1	0.83
CSIRO-Mk3.6.0	4	5	2.22
EC-Earth	3	2	1.20
HadGEM2-ES	1	1	0.50
INM-CM4	1	1	0.50
IPSL-CM5A-LR	4	3	1.71
IPSL-CM5A-MR	1	1	0.50
MIROC5	1	1	0.50
MIROC-ESM	3	1	0.75
MPI-ESM-LR	3	3	1.50
MRI-CGCM3	3	1	0.75
NorECM1-M	1	1	0.50
M=14	Sum=38	Sum=23	Sum=13.04

Approaches to statistical inference from MMEs

- Heuristic weighting e.g.

$$\frac{1}{M}\sum_{m=1}^{m=M}\left(\overline{Y}_{mF.}-\overline{Y}_{mH.}\right)$$

- Fixed effects regression models

$$Y_{msr} \sim N(\alpha_m + \beta_s + \gamma_{ms}, \sigma^2) \qquad \text{ANOVA}$$

$$\overline{Y}_{mF.} - \overline{Y}_{mH.} \sim N(\beta_0 + \beta_1 \overline{Y}_{mH.}, \sigma^2) \qquad \text{State-dependent response}$$

e.g. Bracegirdle and Stephenson, Clim Dynamics, 2012

- Bayesian hierarchical models

Stephenson et al., (2012):
Statistical Problems in the Probabilistic Prediction of Climate Change, Environmetrics (accepted)

Heuristic weighting

The climate change response is often estimated using:

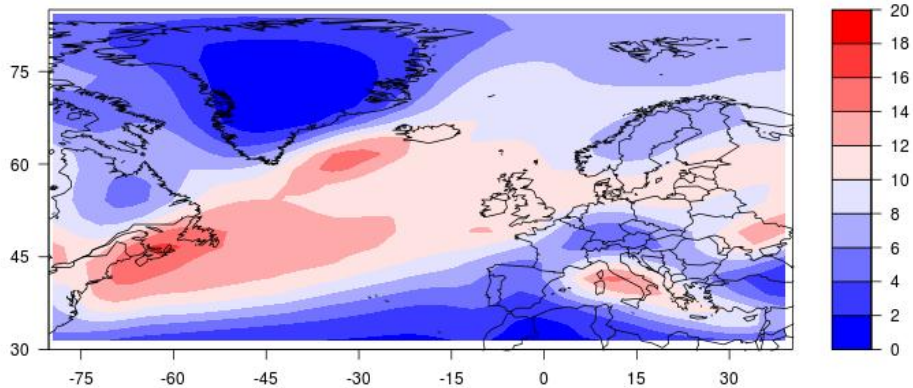
$$\frac{1}{M}\sum_{m=1}^{M}(\overline{y}_{mF.} - \overline{y}_{mH.})$$

This descriptive approach:

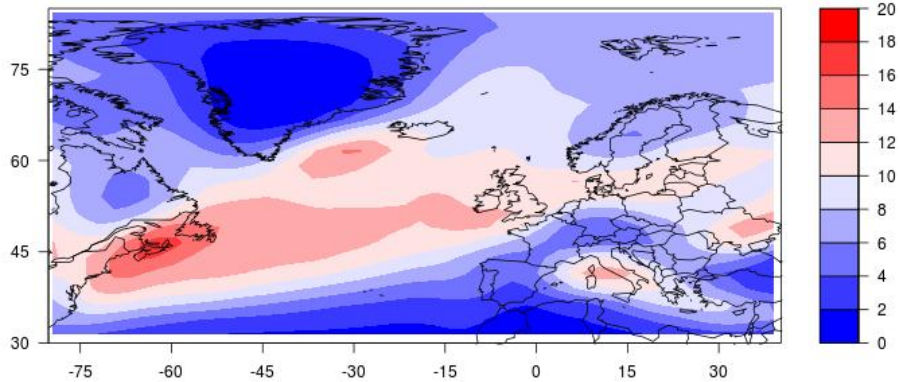
- assumes all climate models are equal (equal weight per climate model rather than equal weight per run)
- is not resistant to outlier model runs and can not identify overly influential runs
- does not quantify uncertainty (e.g. confidence intervals)

Multi-model mean track density (storms/month)

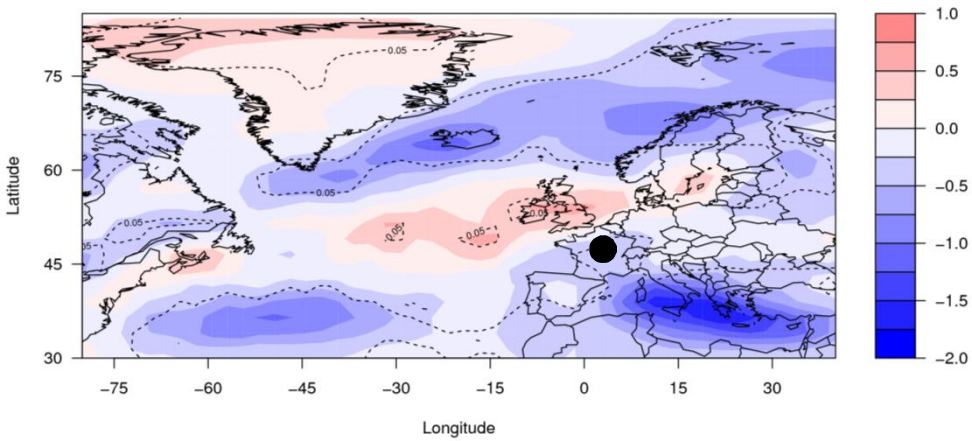
Mean of historical runs

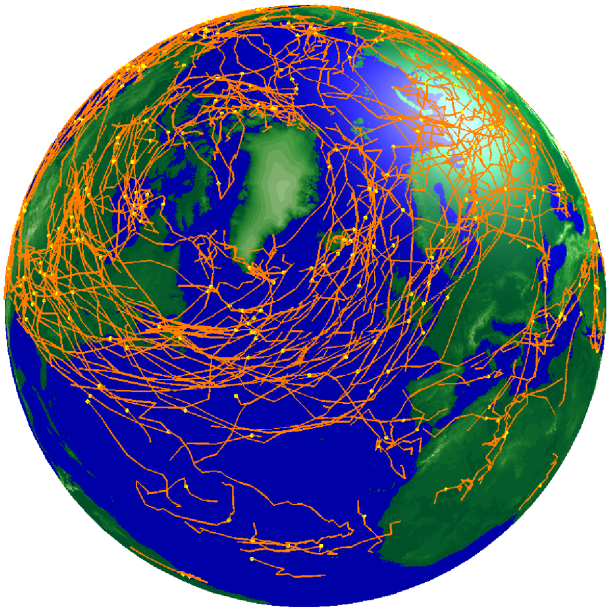


Mean of future runs



Difference





→ Subtle change in mean density of storms

Two-way ANOVA regression model

$$Y_{msr} = \alpha_m + \beta_s + \gamma_{ms} + \varepsilon_{msr}$$

$$\varepsilon_{msr} \overset{iid}{\sim} N\left(0, \sigma^2\right)$$

where $\beta_{mH} = 0$ and $\gamma_{mH} = 0$ for all *m* so that:

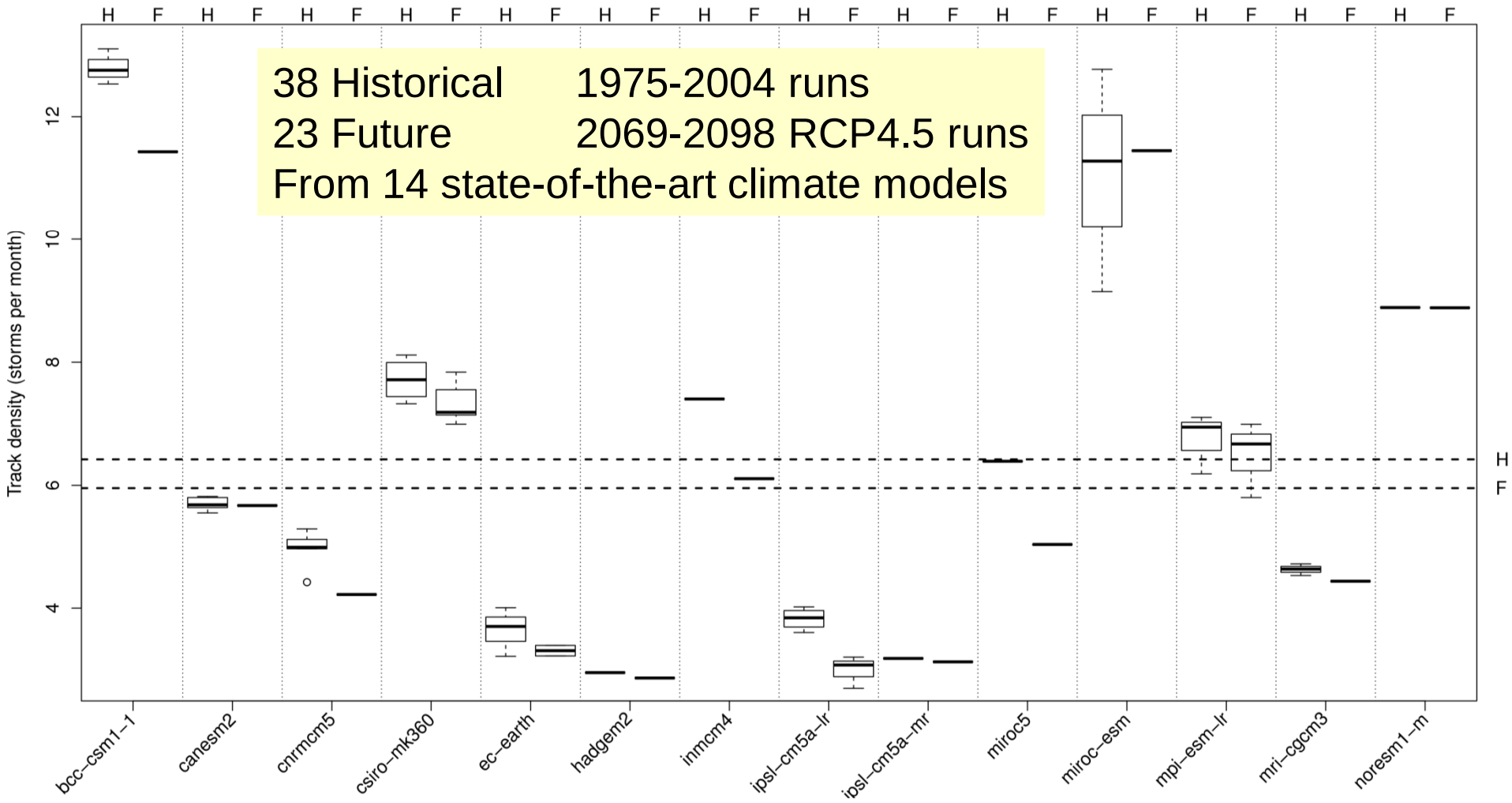
- α_m is the mean historical climate of model *m*;
- β_F is the climate change response of model m;
- γ_{mF} is the model-dependent climate change response;
- ε_{msr} is the natural variability or 'weather noiseó

The point estimate from the ANOVA model is identical to the multi-model mean response using equal weights for all climate models:

$$\Rightarrow \check{\beta}_F = \frac{1}{M}\sum_{m=1}^{M}(\bar{y}_{mF.} - \bar{y}_{mH.})$$

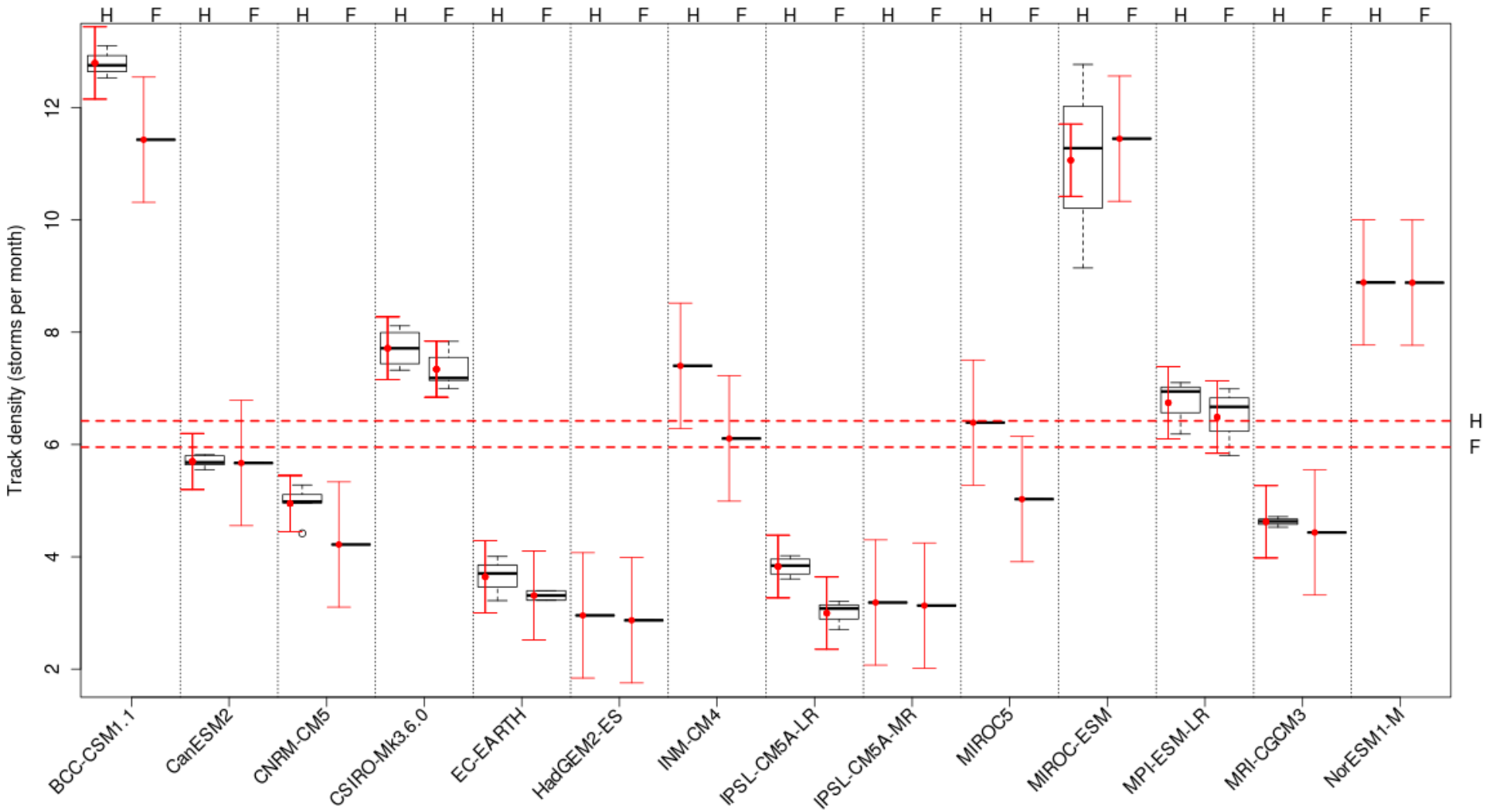
But if the response is very model-dependent, then the real world response might be very different to the MME response!

Track density in central France from CMIP5



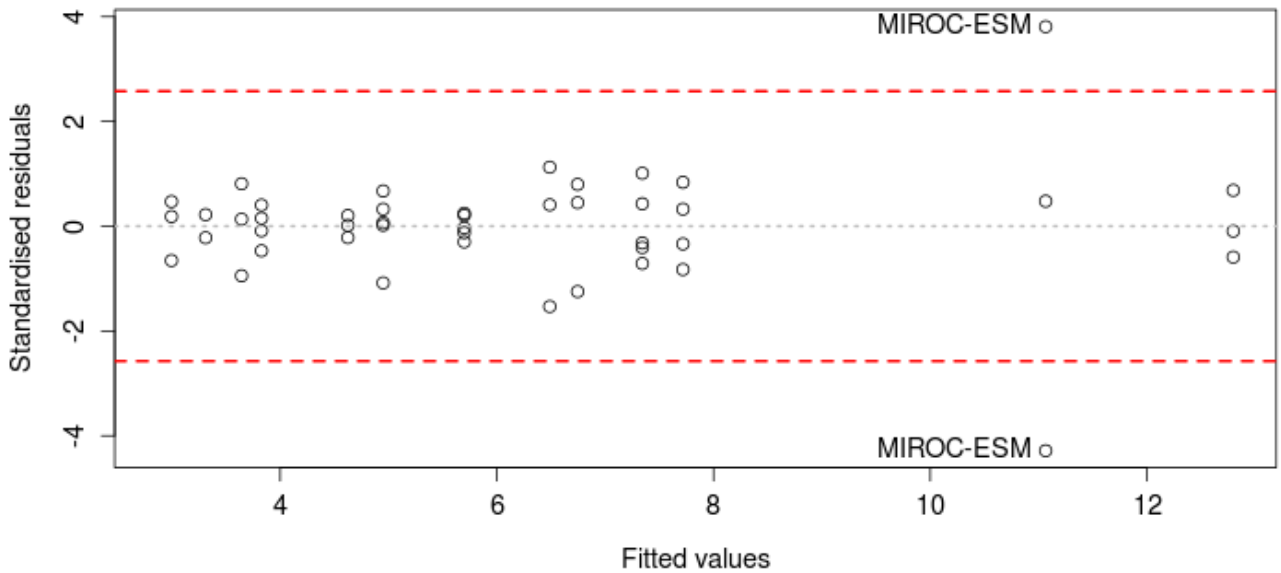
→ The climate change signal is much smaller than spread between models
→ Model spread is greater than natural variability between runs

Fitted values from full framework

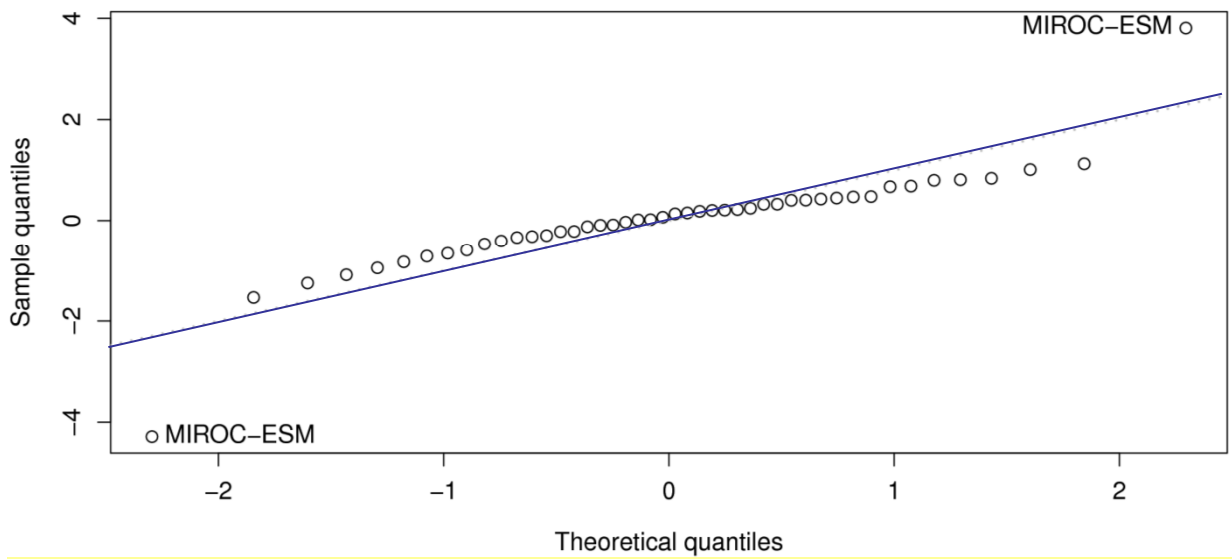


→ Mean response identical to climate model mean
→ Can obtain standard errors as well as means

Residual diagnostics from full model



Dashed lines show 0.5% and 99.5% quantiles of N(0,1)



→ MIROC-ESM is a clear outlier! So throw it out of the MME

Simpler additive model (no interaction term)

$$Y_{msr} = \alpha_m + \beta_s + \varepsilon_{msr}$$

$$\varepsilon_{msr} \overset{iid}{\sim} N\left(0, \sigma^2\right)$$

Can use an F-test to select simpler model having no interactions:

H_0 : $\gamma_{mF} = 0$ for all *m* — *Simpler model with M+2 parameters*

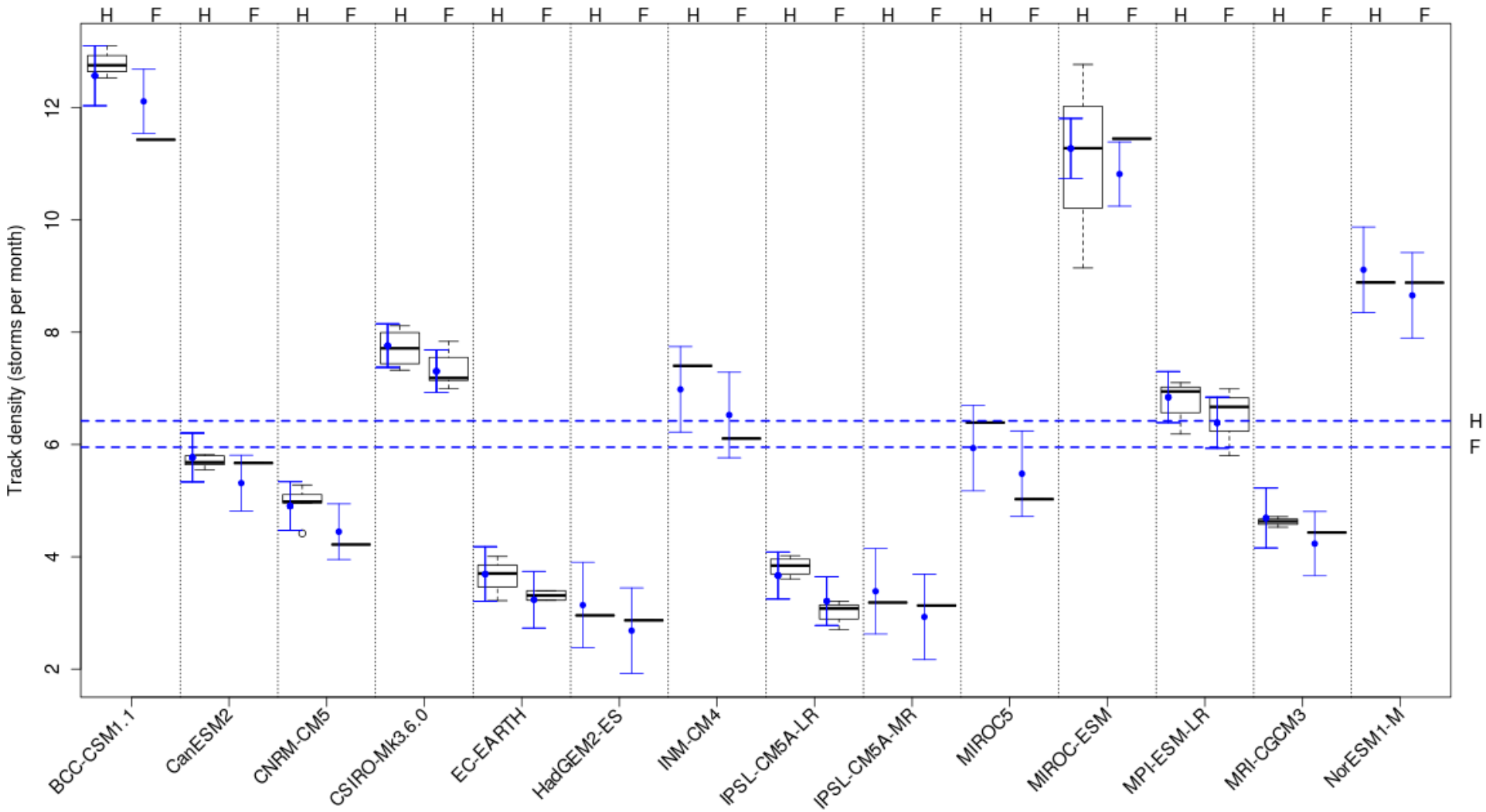
H_a : γ_{mF} Í 0 — *Full model with 2M+1 parameters*

$$\Rightarrow \breve{\beta}_F = \left(\sum_{m=1}^{M} w_m\right)^{-1} \sum_{m=1}^{M} w_m (\bar{y}_{mF.} - \bar{y}_{mH.})$$

$$w_m = \frac{R_{mH} R_{mF}}{R_{mH} + R_{mF}}$$

Weights depend on the number of historical and future runs of each model

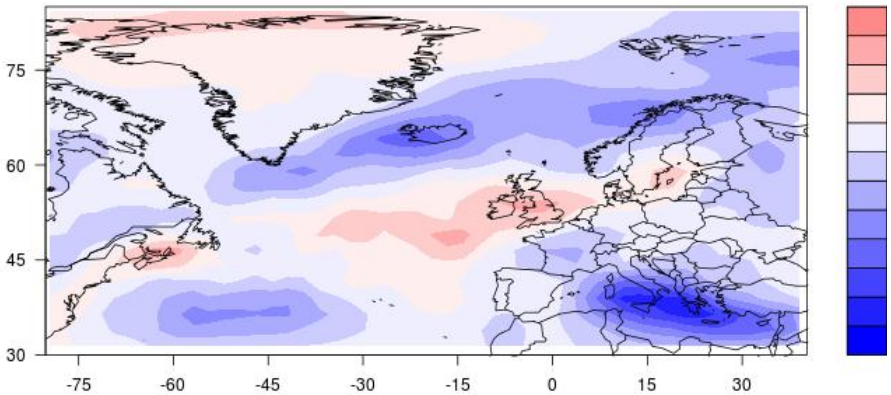
Fitted values from simpler framework



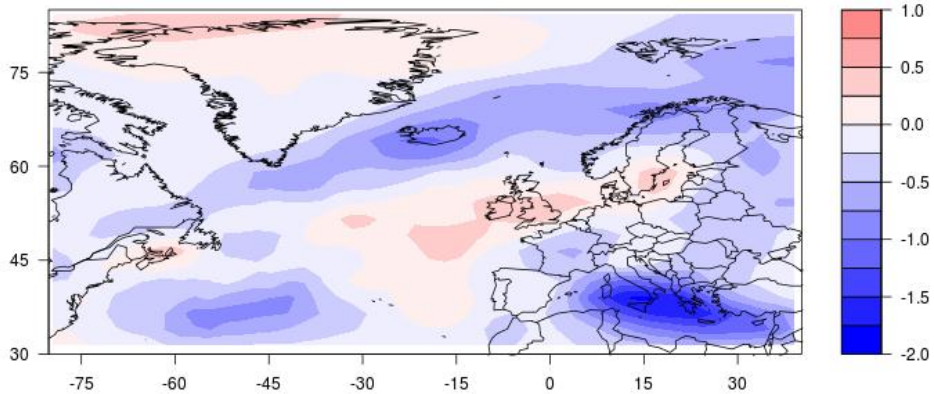
→ No-interaction framework reproduces CMIP5 data well

Mean climate change response from both models

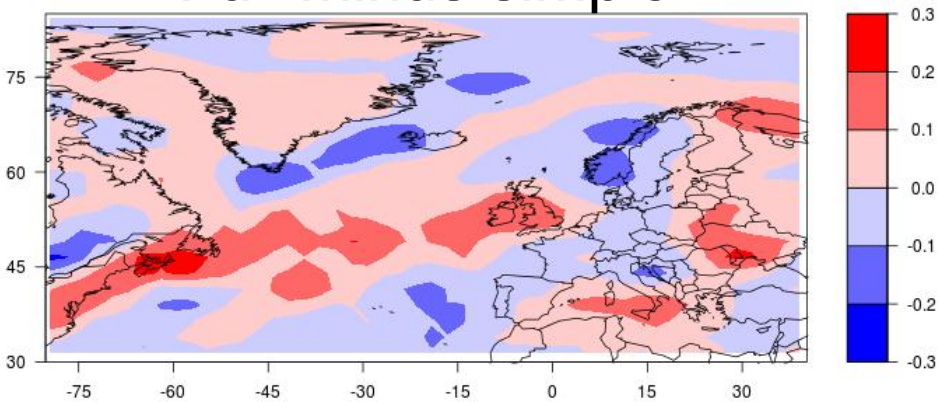
Full model



Simpler model



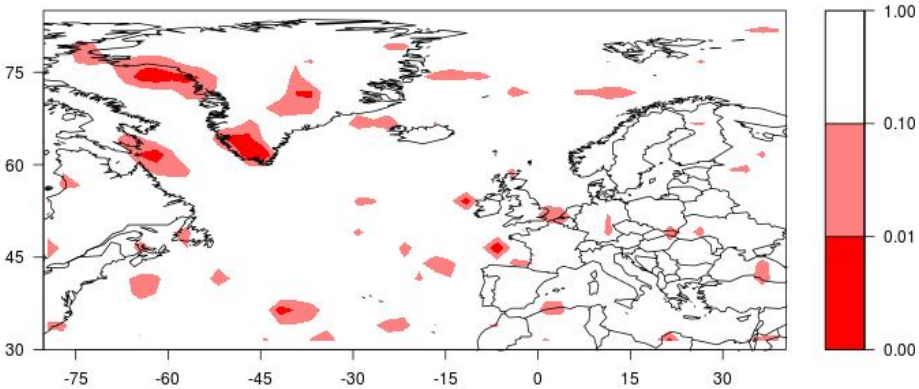
Full minus simple



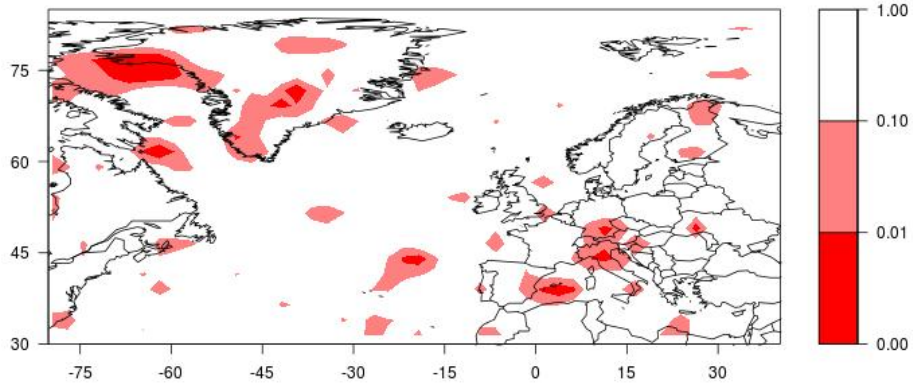
→ The two frameworks give similar results for storm densities

Normality of residuals: Anderson-Darling p-values

Full model



Simpler model

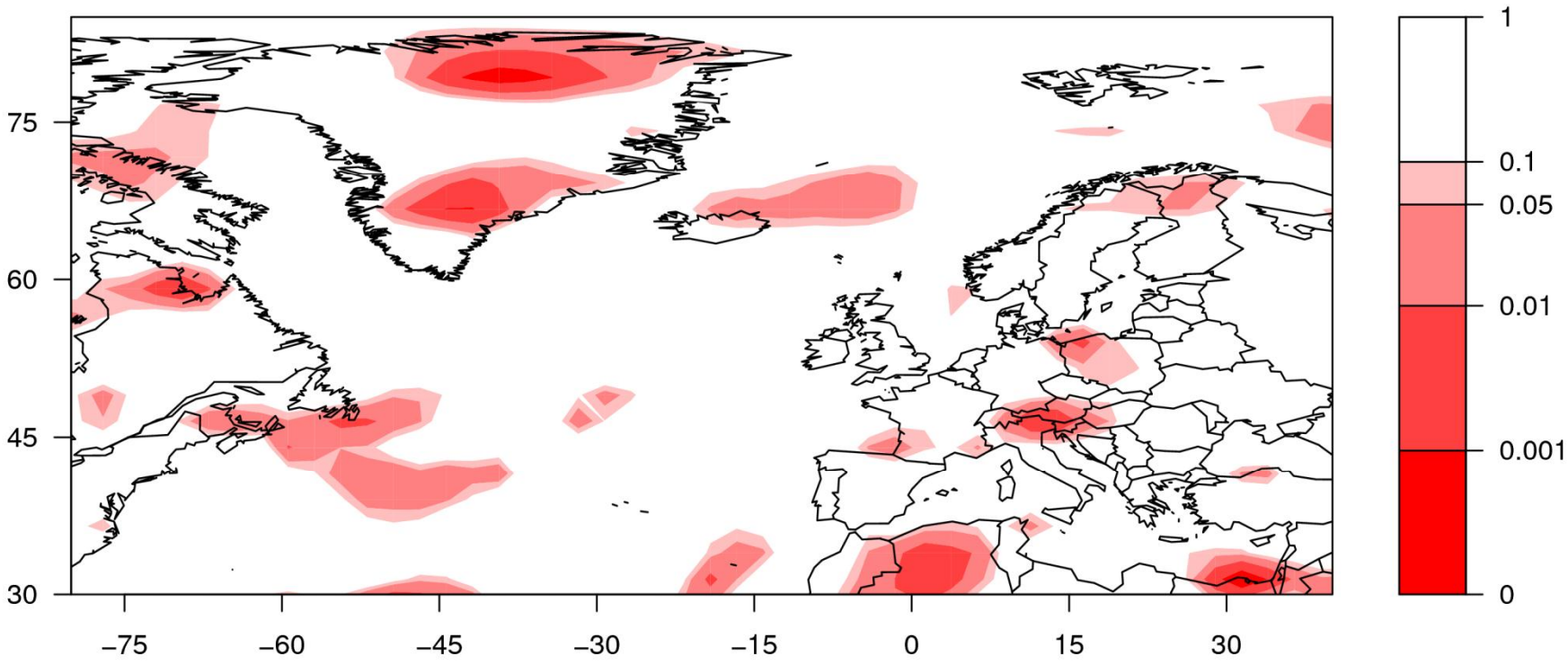


→ Assumption of normal residuals well satisfied for both models

Is the climate change response model-dependent?

F-test of $H_0 : \gamma_{mF} = 0$ for all *m* against H_1 : at least one $\gamma_{mF} \neq 0$

Small p-values (< 0.05) indicate suggest model-dependent climate change



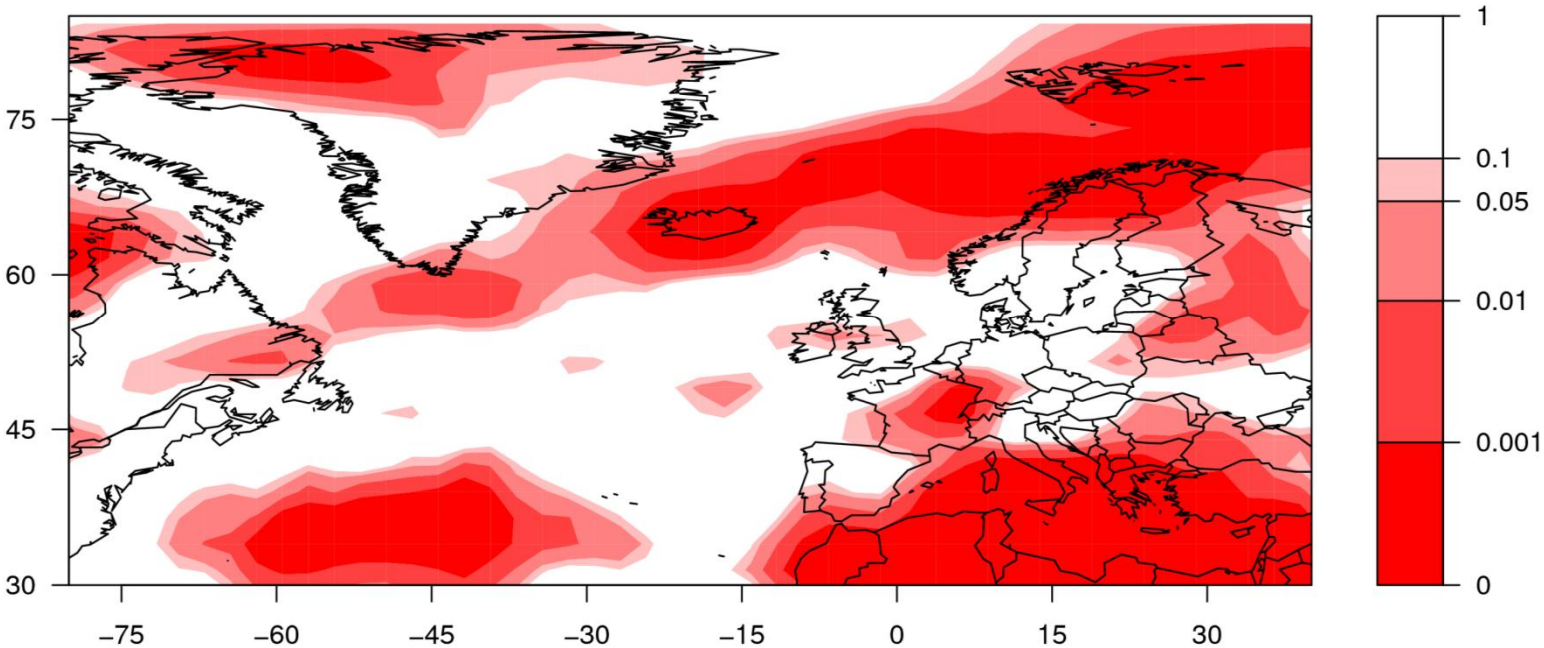
→ Simpler no-interaction model is hard to reject

→ Storm-track response is not overly model-dependent

Where is the the climate change response β'_F significant?

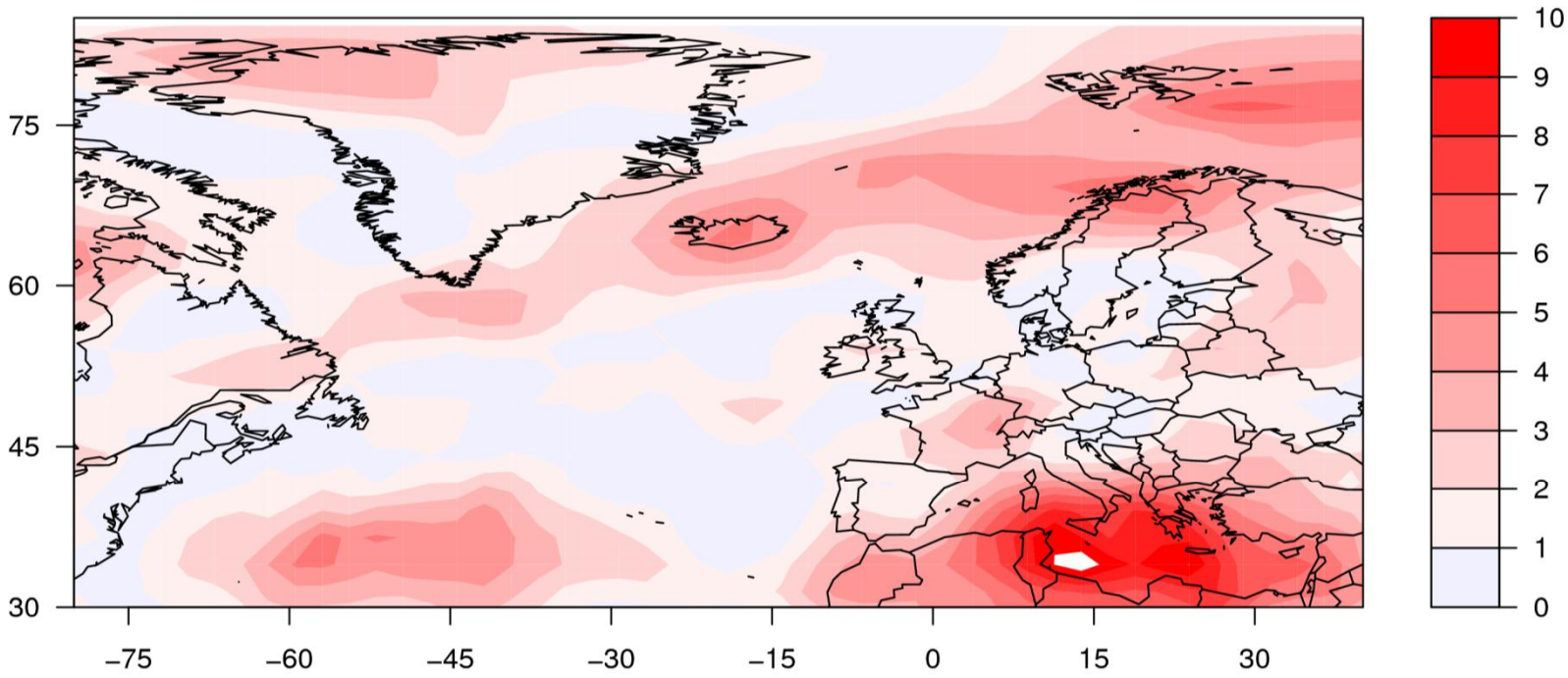
T-test of $H_0 : \beta'_F = 0$ against $H_1 : \beta'_F \neq 0$

Small p-values (< 0.05) indicate significant climate change



→ Mediterranean and northern Scandinavia significant change at 5% level

Climate change signal-to-noise F-ratio

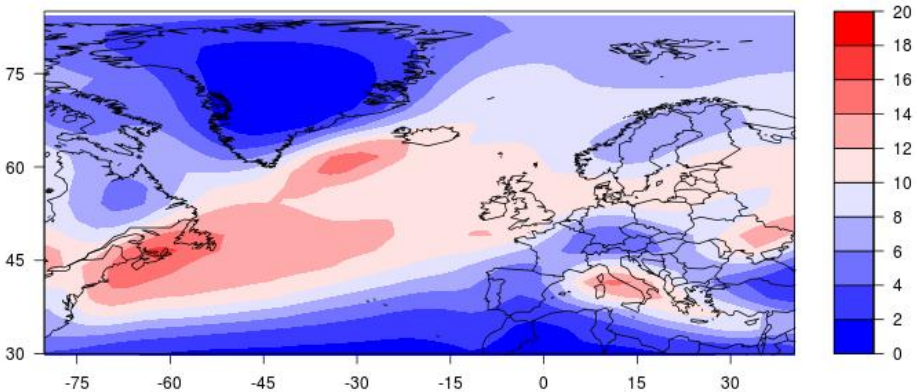


→ the estimated climate change signal is generally much smaller than the estimated natural variability in the climate model run

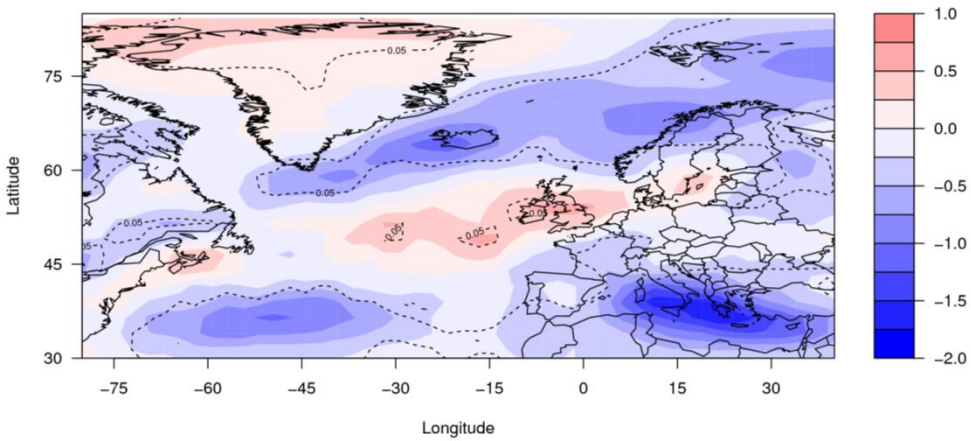
GOOD NEWS for humanity ï BAD NEWS for storm track scientists!

Summary

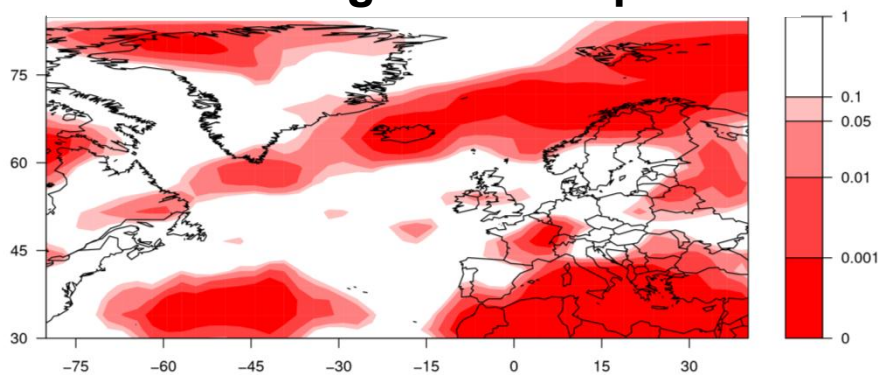
Mean track density in historical runs



Mean climate change response



Statistical significance: p-value



→ Slight increase in the centre of the stormtrack
→ Decrease on the edges especially over the Eastern Mediterranean.

Conclusions

ÅThe multi-model mean estimate of climate change response is equivalent to the mean response estimated from a 2-way linear ANOVA model with model-scenario interactions;

ÅThe ANOVA model can be used to identify outlying climate runs (e.g. the MIROC-ESM climate model runs).

ÅFor storm track density, a simpler more parsimonious model with no interaction term provides a good description of the CMIP5 data ï the climate change response from this model correspond to weights that depend on the number of runs of each climate model;

ÅThe climate change response in storm track density is small compared to natural variability and model spread and not very model-dependent;

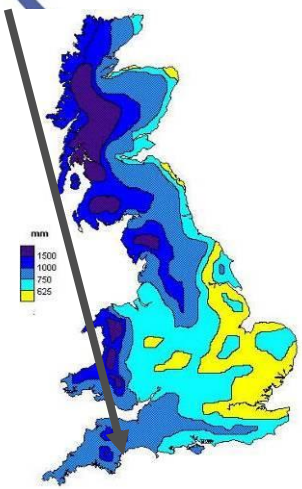
ÅProjected storm track response for end of 21st century:

- ÅEastern Mediterranean: slight decrease in frequency and intensity
- ÅNorthern Scandinavia: slight decrease in frequency and intensity
- ÅUK and central Europe: slight increase in storm frequency and intensity

Thank you for your attention

Professor David B. Stephenson
Exeter Climate Systems
Mathematics Research Institute

Any questions or you want to visit us then please contact:
d.b.stephenson@exeter.ac.uk

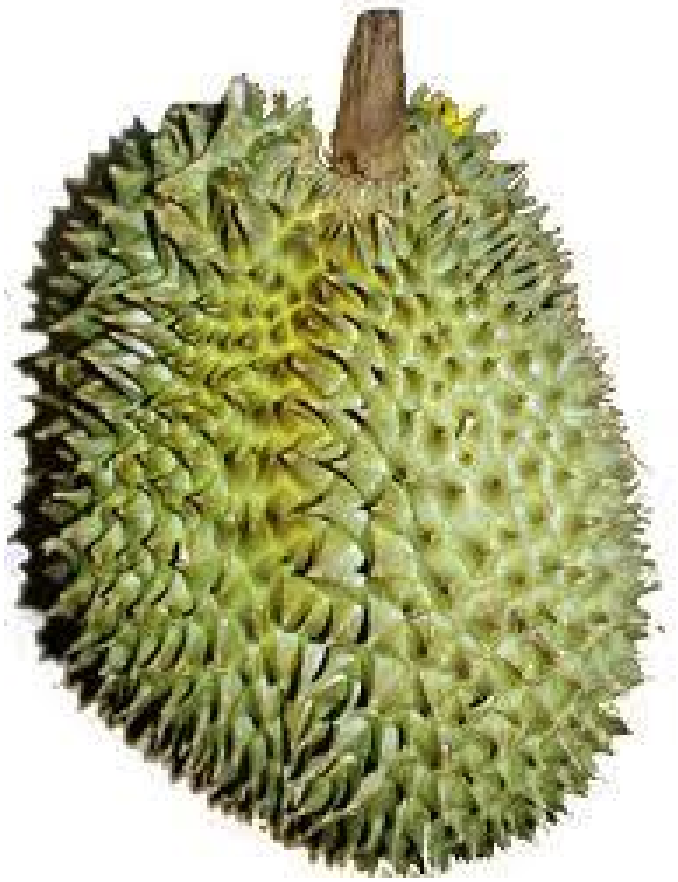


Additional slides for questions etc. ...

A Multi-Model Ensemble is like my fruit bowl ...

A ñfruit bowl of opportunityò{Y_{mrs}; m=1,...M; r=1,...,R_{ms}; s=H,F}
Note: Not a random sample from one homogeneous population
(and it does not include all possible fruit!)

What does reality look like?



An inconvenient truth

actual climate Y ï an observation of the truth Z
It could not have been drawn out of my fruit bowl!

How can we infer properties of this from the fruit in the fruitbowl?

Multi-model means (smoothies)

A multi-model mean is a smoothie - a heuristically weighted average of fruits. It is not an item of real fruit! (important information has been lost by averaging)

→We require modelling frameworks for obtaining samples of real fruit from the posterior distribution p(Y|X) (weather generators?)

Should we use everything in the fruit bowl?

Should we select subsets? How should we weight the fruits?

"All fruit are wrong, but some are tasty" - Granny Smith

Interaction plots