GUY CARPENTER

A catastrophe model for insurance losses due to freeze events using vine copulas

R in insurance 2017

Paris, 8 June 2017

Simos Koumoutsaris
Guy Carpenter

MARSH & McLENNAN COMPANIES

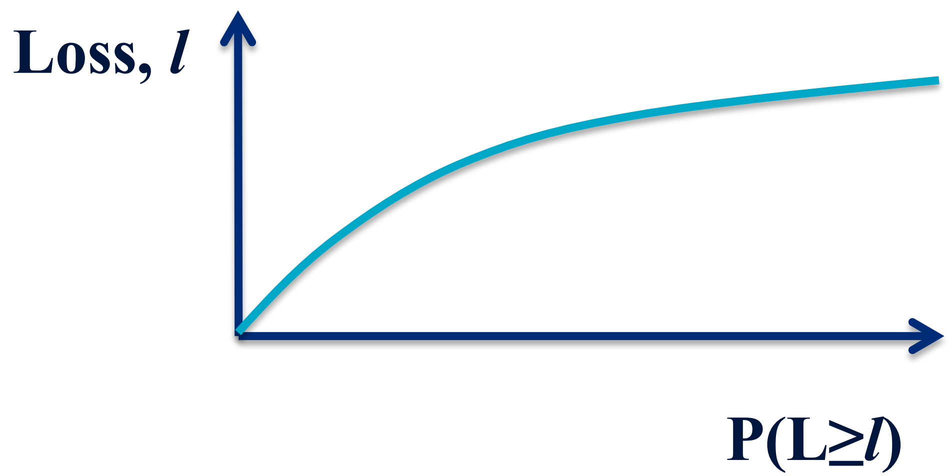
Outline

- Catastrophe modelling
  - What is it?
  - How Vine copulas can be useful in the development of catastrophe models?
- Vine Copulas in R
- Application: a catastrophe model for insurance losses due to freeze events
- Conclusion

Catastrophe modelling

- Goal: generate exceedance probability curves (for a specific portfolio)

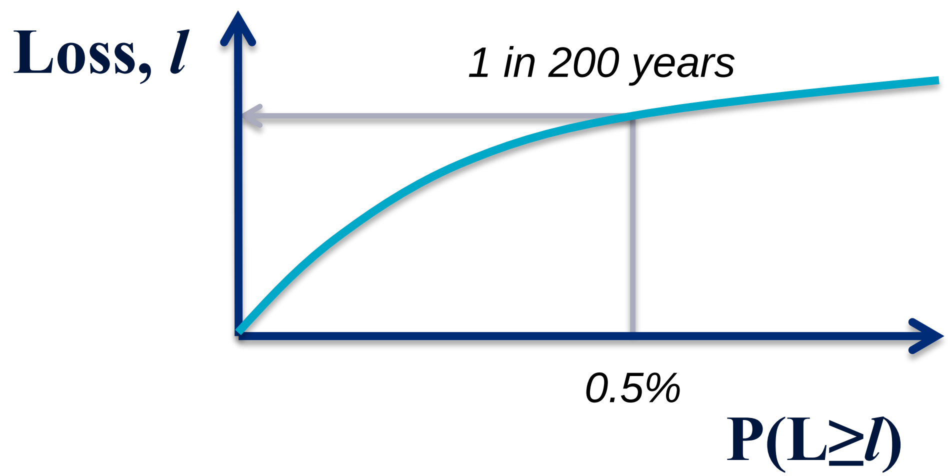
Exceedance Probability (EP) curve



Catastrophe modelling

- Goal: generate exceedance probability curves (for a specific portfolio)
- Insurance focus on extreme events (e.g. 1 in 200 year RP loss)

Exceedance Probability (EP) curve

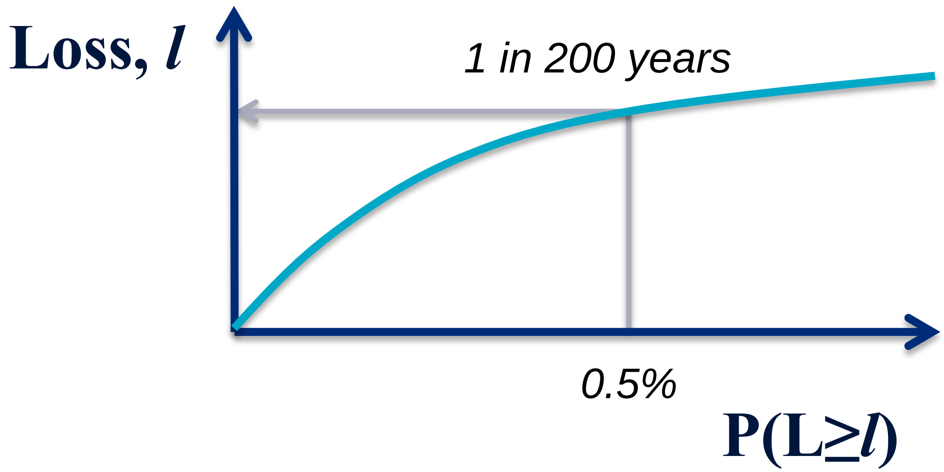


Catastrophe modelling

- Goal: generate exceedance probability curves (for a specific portfolio)
- Insurance focus on extreme events (e.g. 1 in 200 year RP loss)

Create a large sample of synthetic events that are likely to happen

Exceedance Probability (EP) curve



Catastrophe modelling

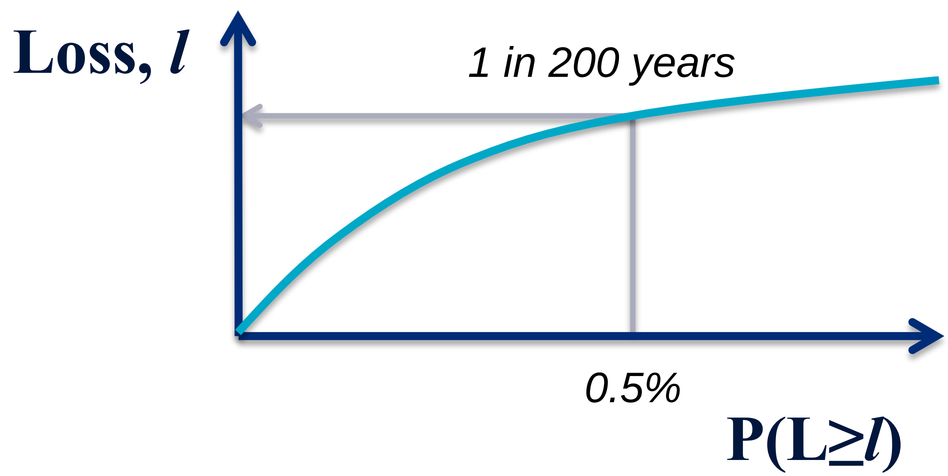
- Goal: generate exceedance probability curves (for a specific portfolio)
- Insurance focus on extreme events (e.g. 1 in 200 year RP loss)

Create a large sample of synthetic events that are likely to happen

⇩

For each event calculate hazard intensity at each location

Exceedance Probability (EP) curve



Catastrophe modelling

- Goal: generate exceedance probability curves (for a specific portfolio)
- Insurance focus on extreme events (e.g. 1 in 200 year RP loss)

Create a large sample of synthetic events that are likely to happen

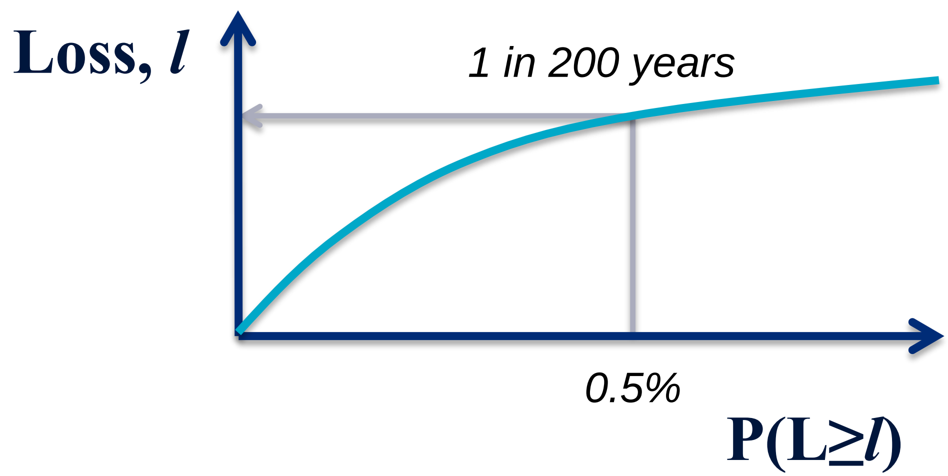
↓

For each event calculate hazard intensity at each location

↓

Based on intensity and exposure at each location calculate damage

Exceedance Probability (EP) curve



Catastrophe modelling

- Goal: generate exceedance probability curves (for a specific portfolio)
- Insurance focus on extreme events (e.g. 1 in 200 year RP loss)

Create a large sample of synthetic events that are likely to happen

↓

For each event calculate hazard intensity at each location

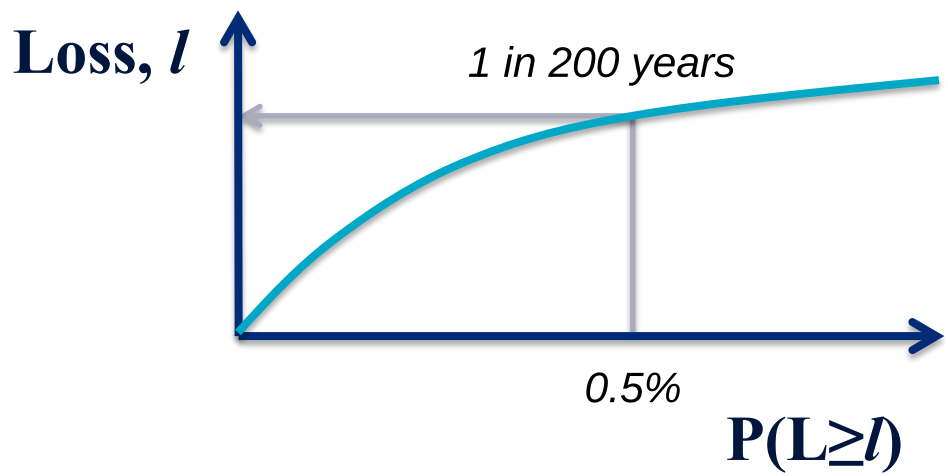
↓

Based on intensity and exposure at each location calculate damage

↓

Calculate aggregate losses (for all risks) and output reports

Exceedance Probability (EP) curve



Catastrophe modelling

- Goal: generate exceedance probability curves (for a specific portfolio)
- Insurance focus on extreme events (e.g. 1 in 200 year RP loss)

Create a large sample of synthetic events that are likely to happen

- Where?
- How big?
- How frequent?

Catastrophe modelling

- Goal: generate exceedance probability curves (for a specific portfolio)
- Insurance focus on extreme events (e.g. 1 in 200 year RP loss)

Create a large sample of synthetic events that are likely to happen

- Where?
- How big?
- How frequent?

We can use copulas to model separately

➢ The probability of hazard extremes at each location (**marginals**).

➢ The spatial correlation between the locations (**dependence**)

Catastrophe modelling

- Goal: generate exceedance probability curves (for a specific portfolio)
- Insurance focus on extreme events (e.g. 1 in 200 year RP loss)

Create a large sample of synthetic events that are likely to happen

- Where?
- How big?
- How frequent?

We can use copulas to model separately

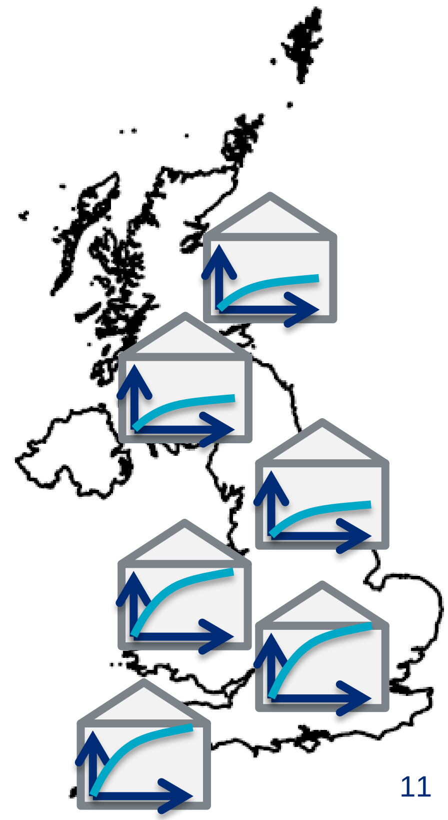
➢ The probability of hazard extremes at each location (**marginals**).

➢ The spatial correlation between the locations (**dependence**)

- dependence in the vulnerability between risks
- due to the spatial structure of hazard

Copulas

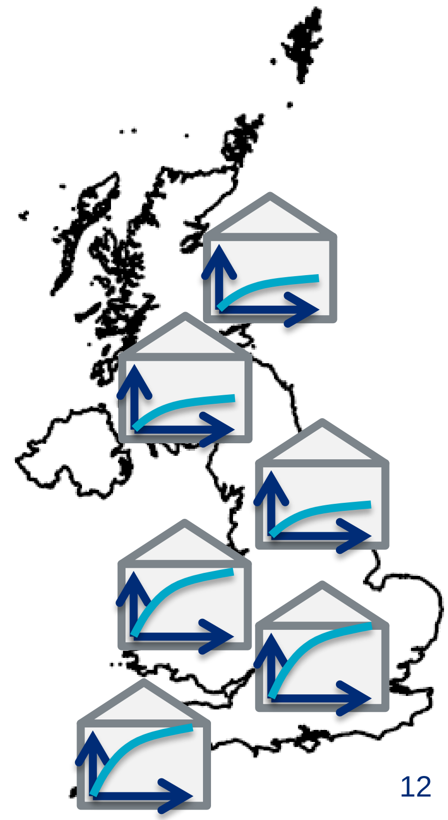
- Copulas are relative simple in 2 dimensions, but it gets increasingly difficult in higher dimensions:
  - The choice of adequate copulas is limited.
  - Standard multivariate copulas either do not allow tail dependence (e.g. multivariate Gaussian) or only have one or two parameters to control tail dependence of all pairs of variables (t-Student and Archimedean multivariate copulas).



Copulas

- Copulas are relative simple in 2 dimensions, but it gets increasingly difficult in higher dimensions:
  - The choice of adequate copulas is limited.
  - Standard multivariate copulas either do not allow tail dependence (e.g. multivariate Gaussian) or only have one or two parameters to control tail dependence of all pairs of variables (t-Student and Archimedean multivariate copulas).

➔ **Vine-Copulas** for higher dimensional data



Vine Copulas

- **Vine-Copulas** are based on a pairwise decomposition of a multivariate model into **bivariate** copulas, where each pair-copula can be chosen independently from the others.

- E.g. in 3 dimensions:

$$f(x_1, x_2, x_3) = f_1(x_1)\, f_2(x_2)\, f_3(x_3)$$ *marginals*

$$\times\, c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}\,(F_2(x_2), F_3(x_3))$$ *uncoditional pairs*

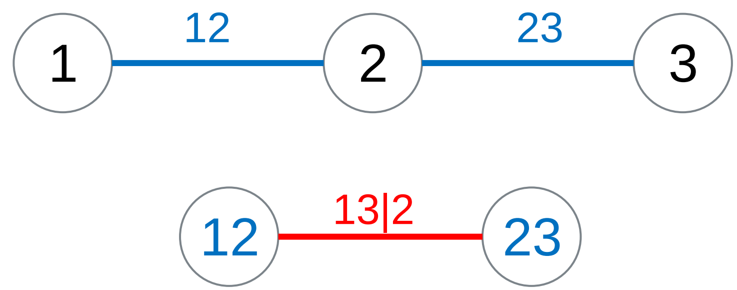
$$\times\, c_{13|2}\,(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))$$ *conditional pairs*

- Vines thus combine the advantages of multivariate copula modelling, that is separation of marginal and dependence modelling, and the flexibility of bivariate copulas

Vine Copulas

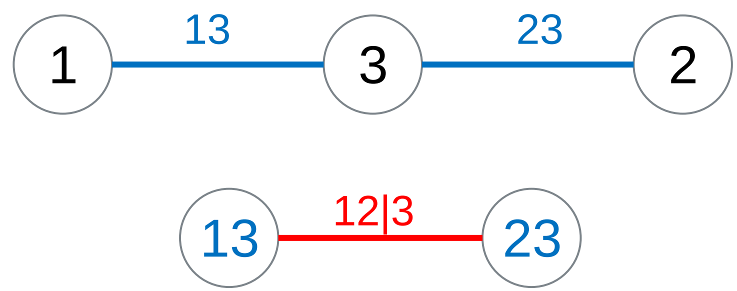
- The decomposition is **not unique.**
- Bedford and Cook (2001) have introduced **a graphical structure called vine structure** which arranges the pair-copulas into trees.
- E.g. in 3 dimensions:

$$f(x_1, x_2, x_3) = f_1(x_1)\, f_2(x_2)\, f_3(x_3)$$
$$\times\, c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}\,(F_2(x_2), F_3(x_3))$$
$$\times\, c_{13|2}\,(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))$$



Or

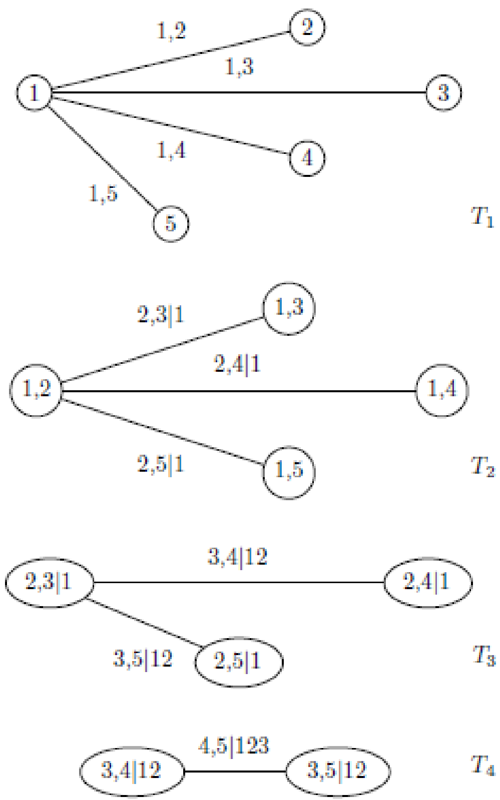
$$f(x_1, x_2, x_3) = f_1(x_1)\, f_2(x_2)\, f_3(x_3)$$
$$\times\, c_{13}(F_1(x_1), F_3(x_3)) \cdot c_{23}\,(F_2(x_2), F_3(x_3))$$
$$\times\, c_{12|3}\,(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3))$$



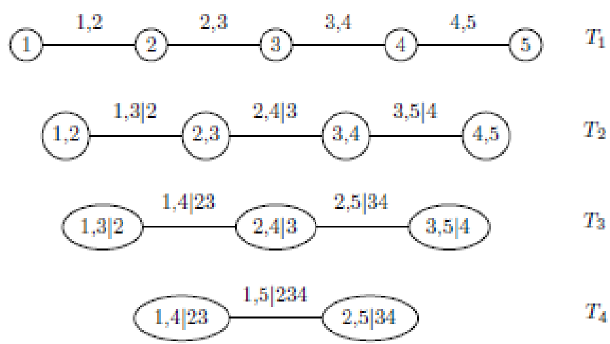
Vine Copulas

- The decomposition is **not unique.**
- Bedford and Cook (2001) have introduced **a graphical structure called vine structure** which arranges the pair-copulas into trees.

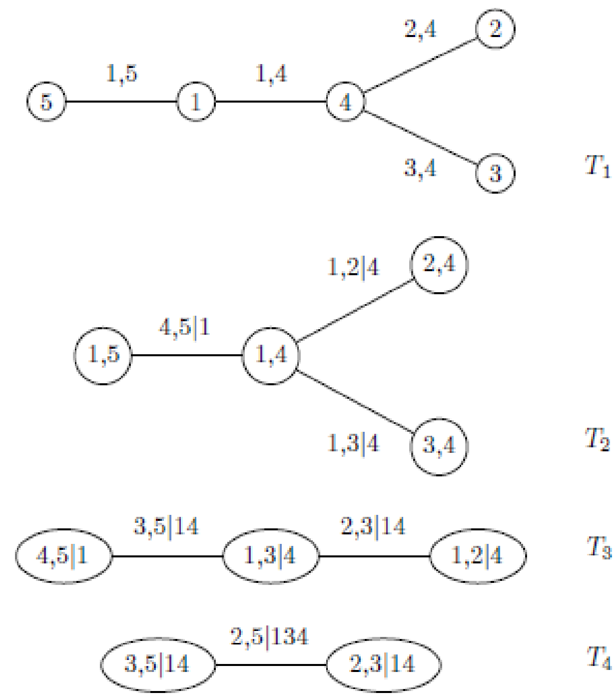
"C(anonical)-Vine"



"D(rawable)-Vine"



"R(egular)-Vine"



Vine Copulas in R

- Packages
  - CDVine (Authors: Schepsmeier and Brechmann)
    - (https://cran.r-project.org/web/packages/CDVine/CDVine.pdf)
  - VineCopula (Authors: Schepsmeier, Stoeber, Brechmann et al.)
    - https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf)
- References
  - http://www.statistics.ma.tum.de/en/research/vine-copula-models/
  - Brechmann E. C. and U. Schepsmeier, Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine, *Journal of Statistical Software*, Vol 52:3 (2013)
  - Brechmann, Statistical inference of vine copulas using the R-package VineCopula, Presentation, May 23 (2013)
  - Schepsmeier (2013) Estimating standard errors and efficient goodness-of-fit tests for regular vine copula models, *PhD Thesis*, Faculty of Mathematics, University of Munich, Germany

R Vine Matrix Object

- RVM = **RVineMatrix**(Matrix=Matrix, family=family, par=par, par2=par2)

  - Matrix: used to describe the vine structure
  - Family: 34 families of bivariate copulas implemented (Gaussian, Student's t, Clayton, Gumbel, Joe, Frank, etc.)
  - Par, Par2: parameters

R Vine Matrix Object

- RVM = **RVineMatrix**(Matrix=Matrix, family=family, par=par, par2=par2)

Parameter estimation:

Sequential estimation of parameters for each bivariate pair

- RVM_SeqEst = **RVineSeqEst**(data, RVM, method="mle") # fast

Maximum likelihood estimation of all parameters jointly:

- RVM_MLE = **RVineMLE**(data, RVMSeqEst) # Starting values using sequential estimation (slow)

R Vine Matrix Object

- RVM = **RVineMatrix**(Matrix=Matrix, family=family, par=par, par2=par2)

Pair copula selection:

- Manually using tools for bivariate analysis (e.g. plots or goodness-of-fit tests: BiCopMetaContour, BiCopGofTest)

- Automatically determine the pair-copula families and parameters using AIC or BIC:

  - RVM_Cop = **RVineCopSelect**(data, Matrix=Matrix, selectioncrit="AIC")

R Vine Matrix Object

- RVM = **RVineMatrix**(Matrix=Matrix, family=family, par=par, par2=par2)

Vine structure selection:

- The method follows an ***automatic strategy*** of jointly searching for an appropriate R-vine tree structure, its pair-copula families and estimating their parameters developed by Dissmann et al. (2013).
  - RVM_Matrix = **RVineStructureSelect**(data)

R Vine Matrix Object

- RVM = **RVineMatrix**(Matrix=Matrix, family=family, par=par, par2=par2)

Simulation:

  – SimData = **RVineSim**(10000, RVM)

R Vine Matrix Object

- RVM = **RVineMatrix**(Matrix=Matrix, family=family, par=par, par2=par2)

Goodness-of-fit:

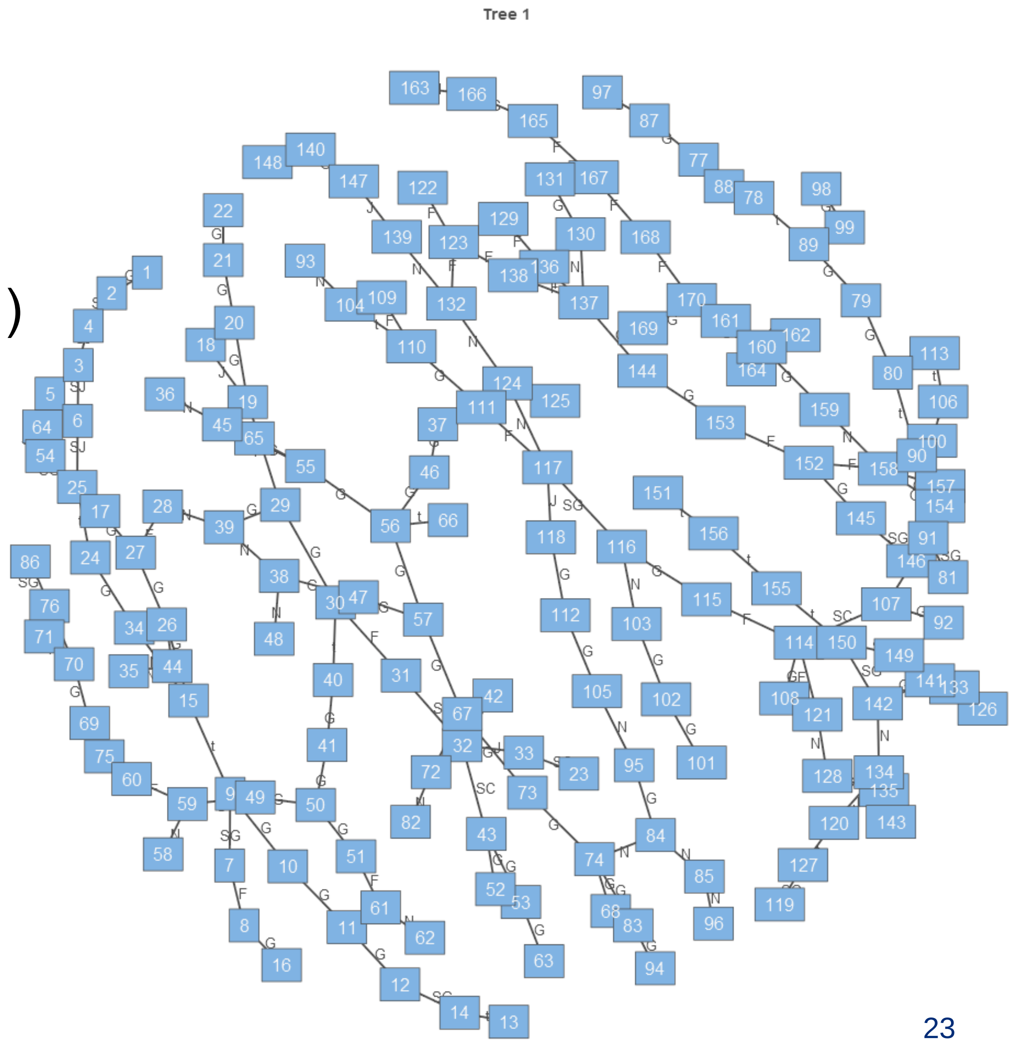
  - goftest = **RVineGofTest**(data, RVM)

R Vine Matrix Object

- RVM = **RVineMatrix**(Matrix=Matrix, family=family, par=par, par2=par2)

R Vine tree plot:

  – **RVineTreePlot**(RVM, tree=1, …)

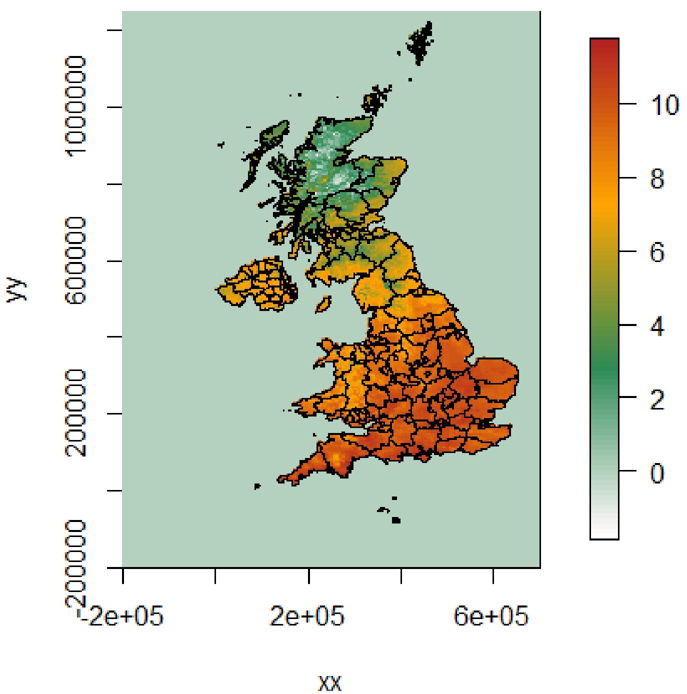


Application: a catastrophe model for insurance losses due to freeze events

- Losses resulting from burst / leaking pipes have a significant impact on the insurance industry
- Total insurance losses £50 – 300 mil *per year* in the last 10 years (ABI).

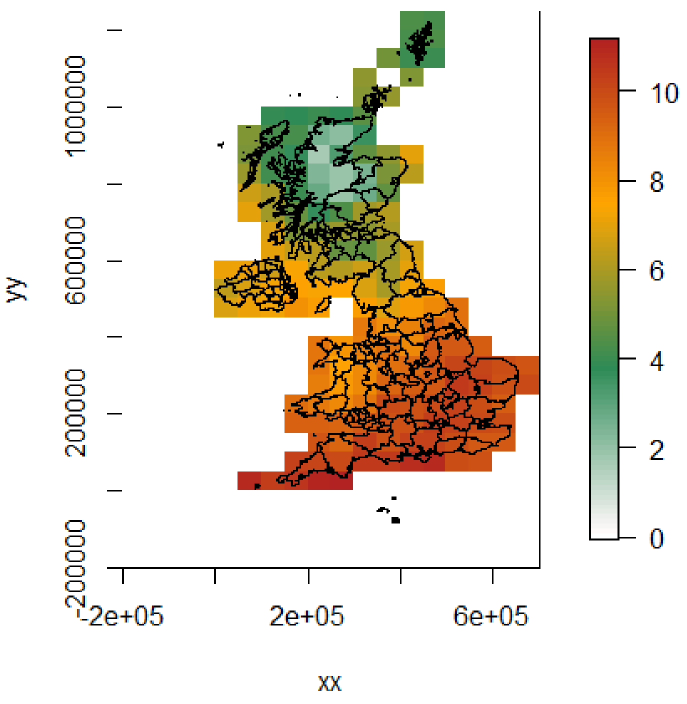
Hazard

- Daily temperature data from the UK Met Office for 51 years (1960-2011)



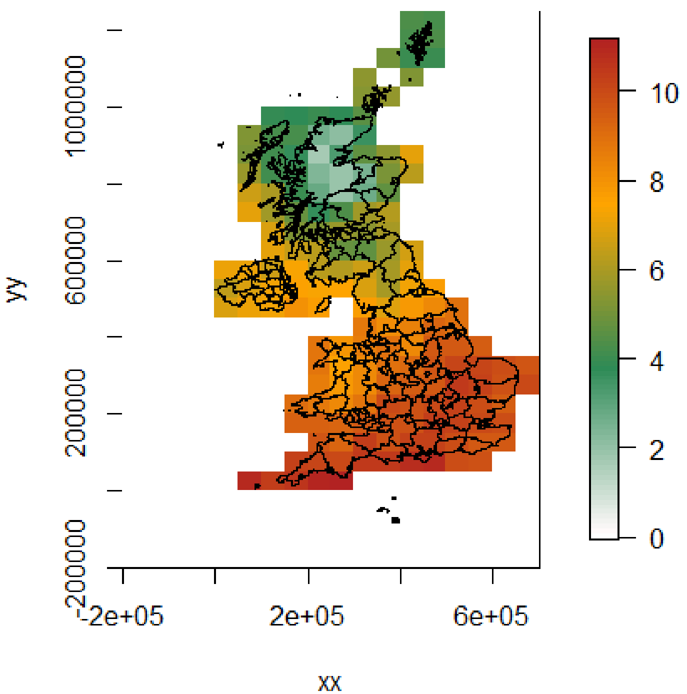
Hazard

- Daily temperature data from the UK Met Office for 51 years (1960-2011).
- Regrid to 50km x 50km (for computational reasons)



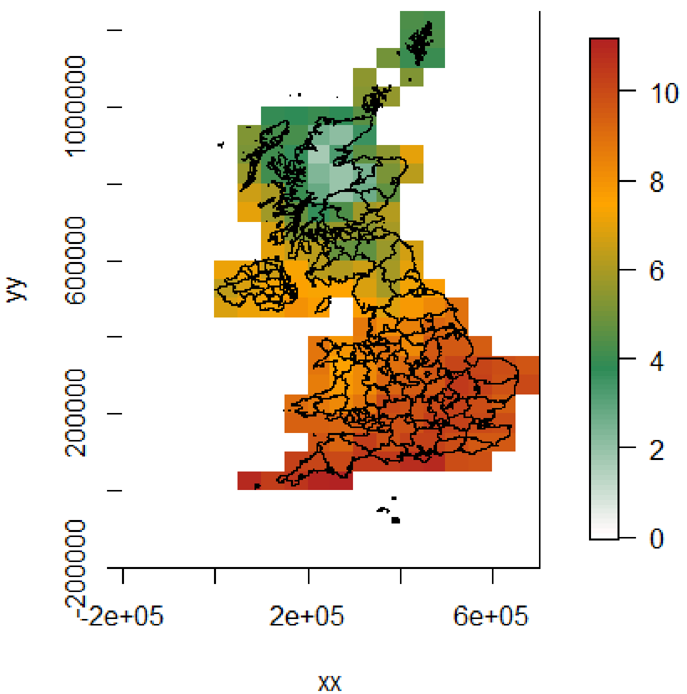
Hazard

- Daily temperature data from the UK Met Office for 51 years (1960-2011).
- Regrid to 50km x 50km (for computational reasons)
- Compute the **annual maximum Air Freezing Index (AFI)**, a commonly used metric for determining the freezing severity of the winter season. It measures the magnitude and duration of air temperature below freezing.



Hazard

- Daily temperature data from the UK Met Office for 51 years (1960-2011).
- Regrid to 50km x 50km (for computational reasons)
- Compute the **annual maximum Air Freezing Index (AFI)**, a commonly used metric for determining the freezing severity of the winter season. It measures the magnitude and duration of air temperature below freezing.

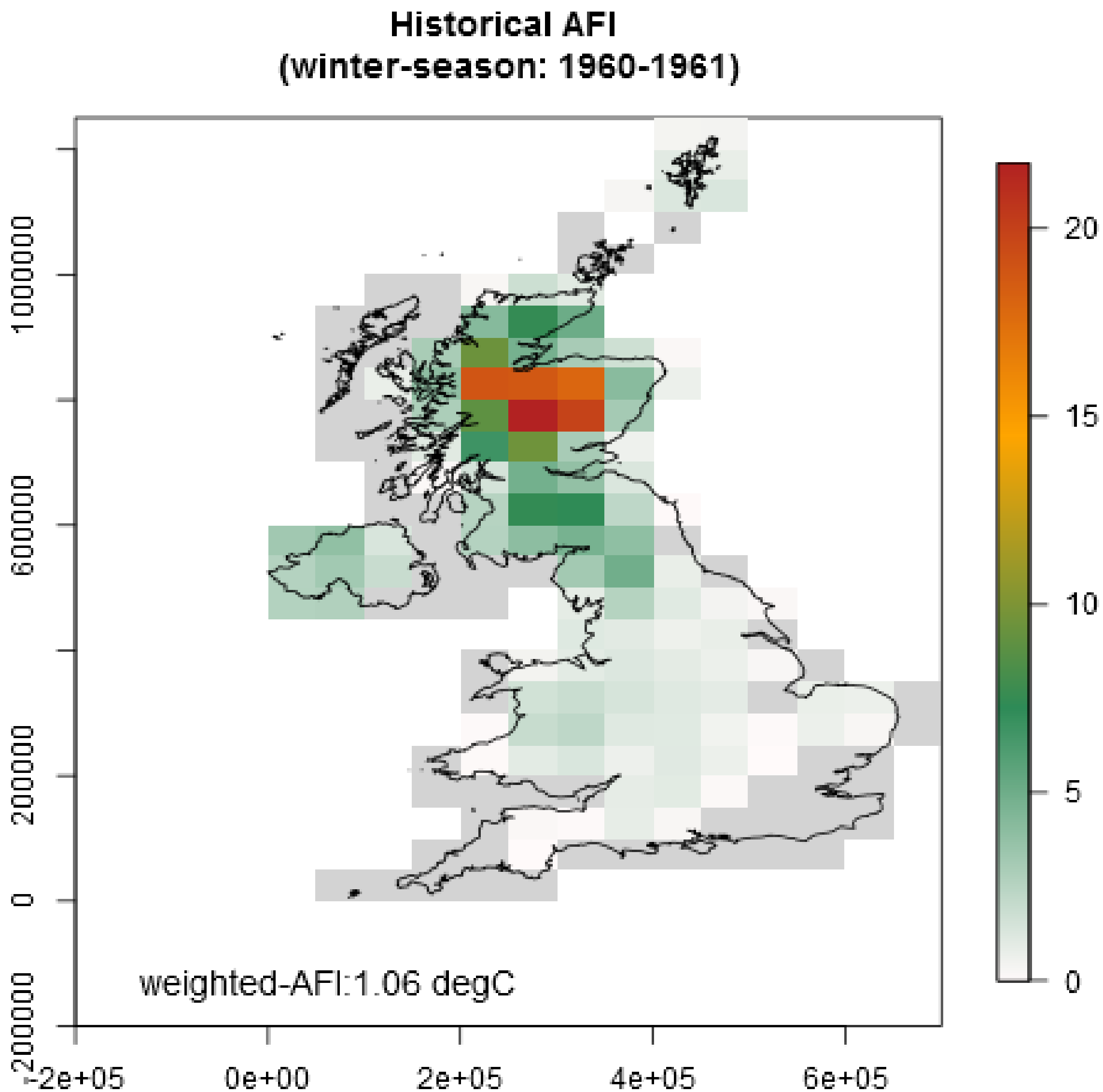


AFI is computed as the cumulative sum of consecutive negative temperatures.

- Several such periods during a winter-season (defined from 1st of July to 31st of June of the next year).
- Take the maximum to represent the winter-season.
- Computed for each grid-box.

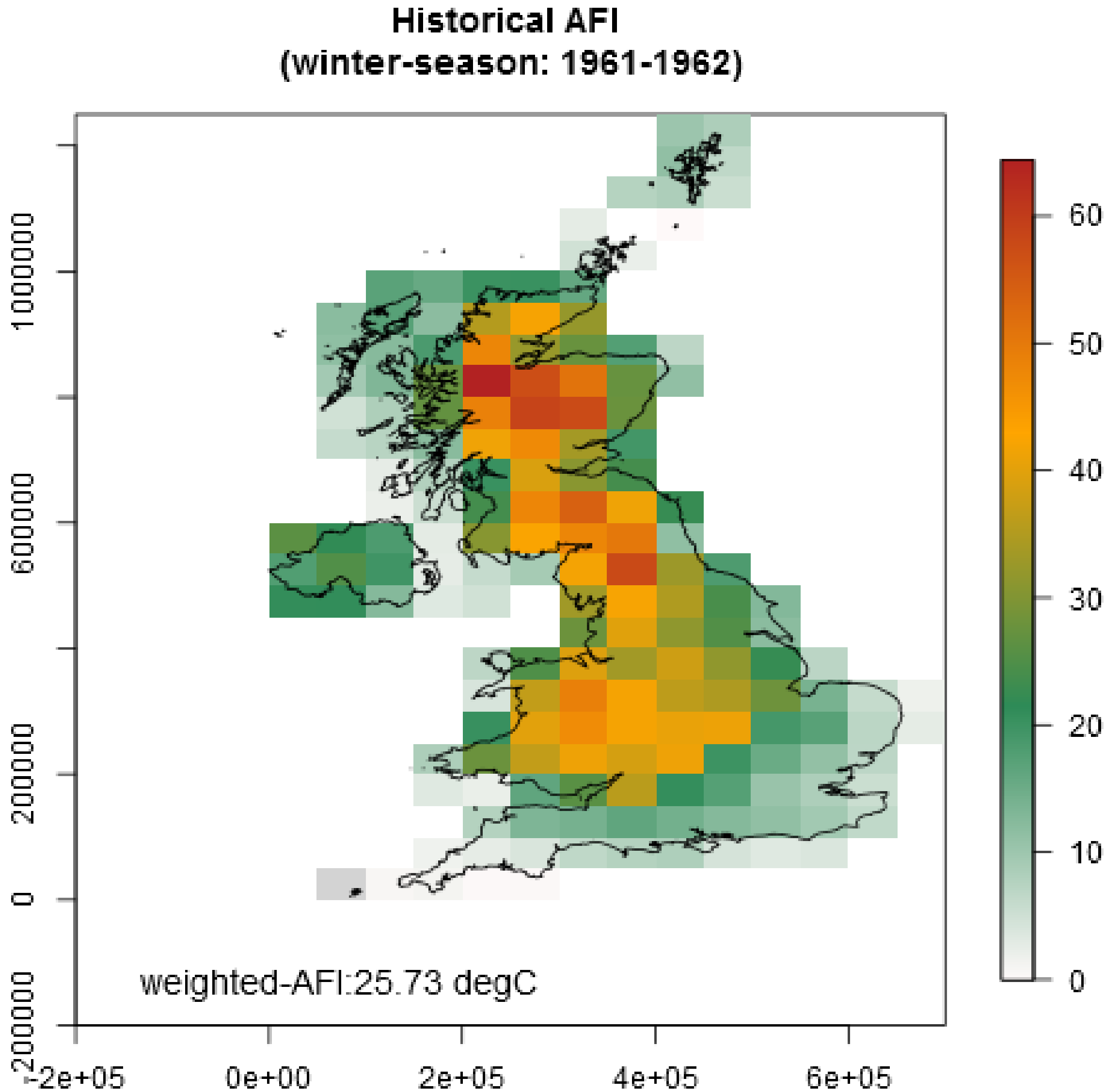
AFI maps

- *1960 (i.e. 1.7.60-31.6.61)*
- wAFI = 1.06 degC
- Weighed AFI (wAFI) weighted over the residential properties
- $wAFI = \frac{\sum_{i=1}^{n=170} nRisks_i \cdot AFI_i}{\sum_{i=1}^{n=170} nRisks_i}$



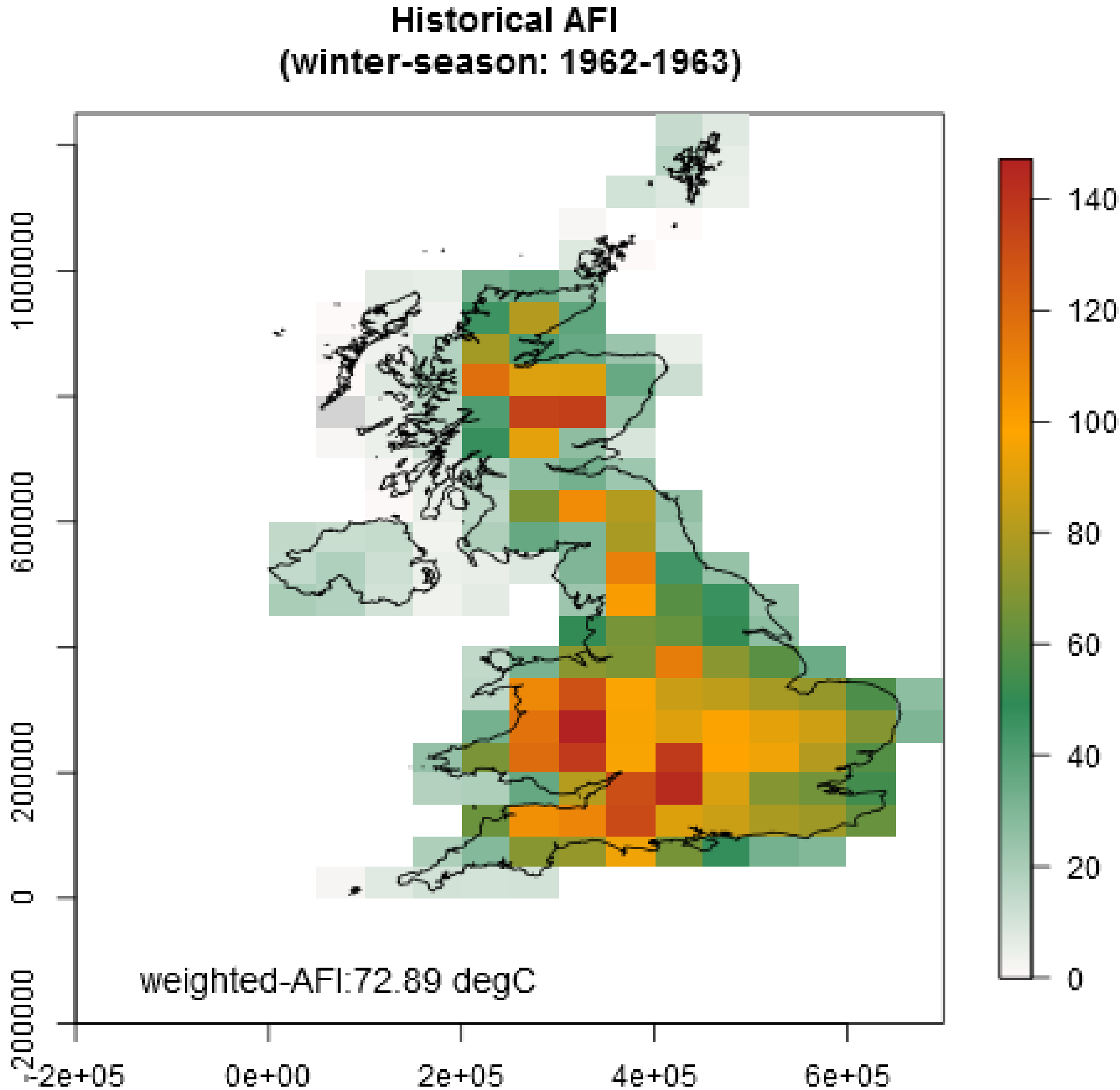
AFI maps

- *1961 (i.e. 1.7.61-31.6.62)*
- wAFI = 25.73 degC



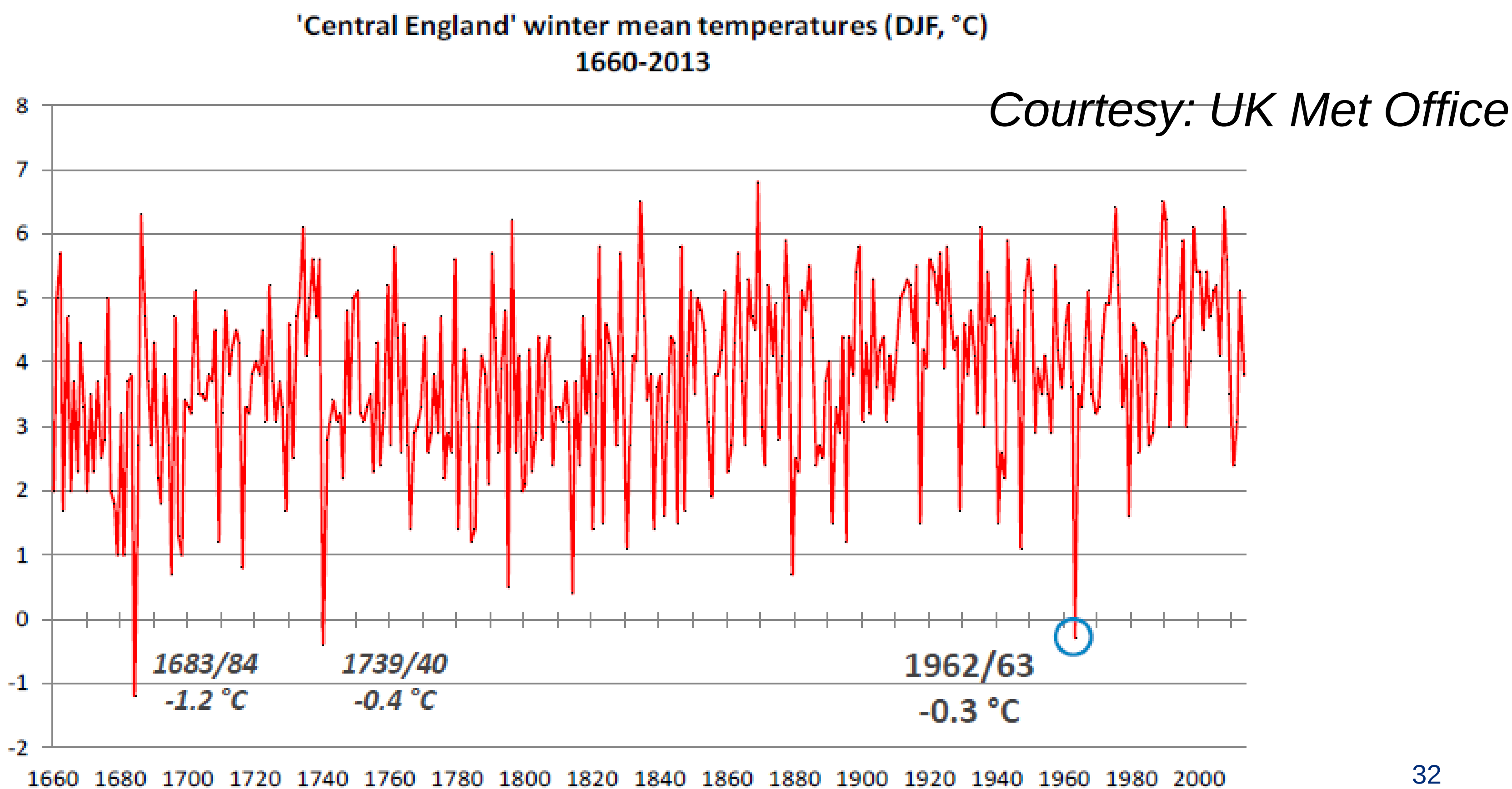
AFI maps

- *1962 (i.e. 1.7.62-31.6.63)*
- wAFI = 72.89 degC



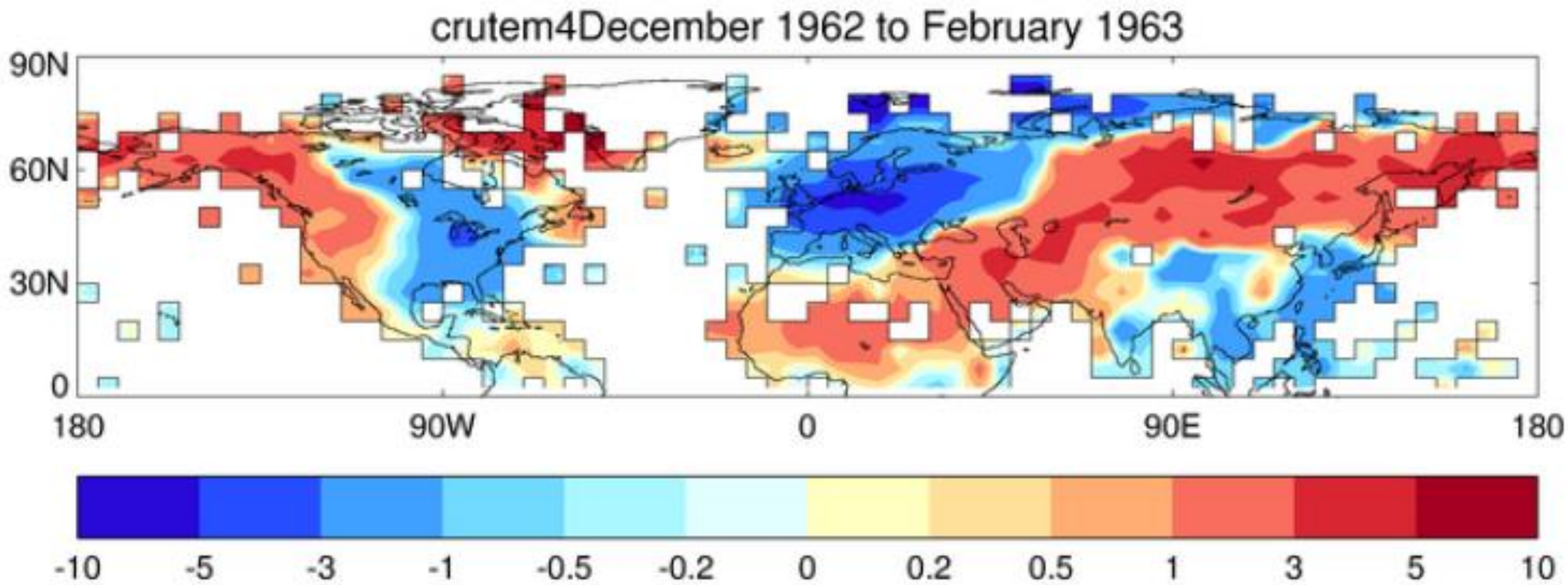
The “Big Freeze” of the winter 1962/1963

- The **winter of 1962–1963** (also known as the **Big Freeze of 1963**) was one of the coldest winters on record in the United Kingdom.
- Notable for its persistence: it started on the 22nd of December and lasted until 4th of March.



NH temperature anomalies

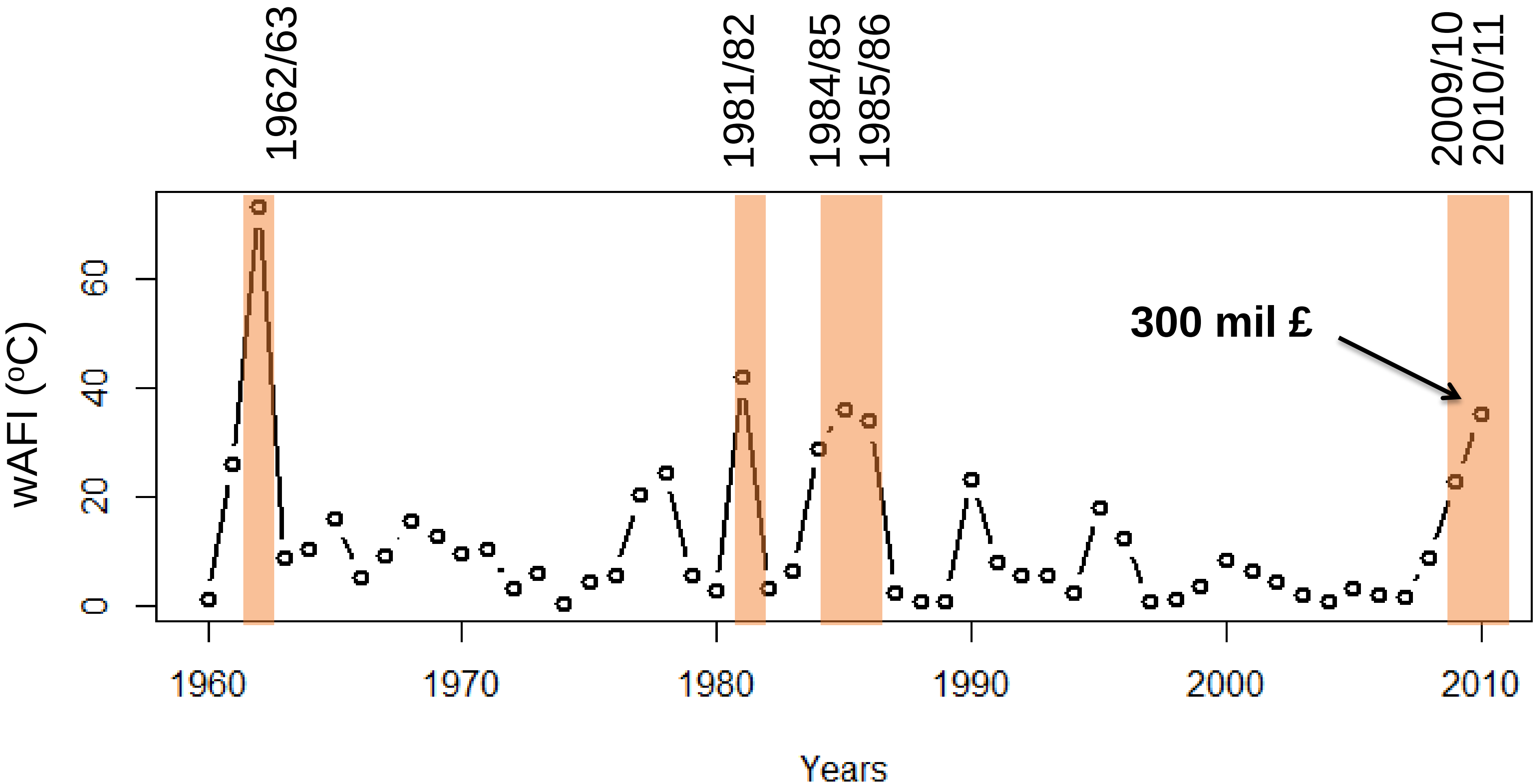
Temperature anomalies – northern Hemisphere



- Very cold over NW Europe (anomalies below -5 degC)
- Very mild over Greenland and northern Canada (+5 degC)

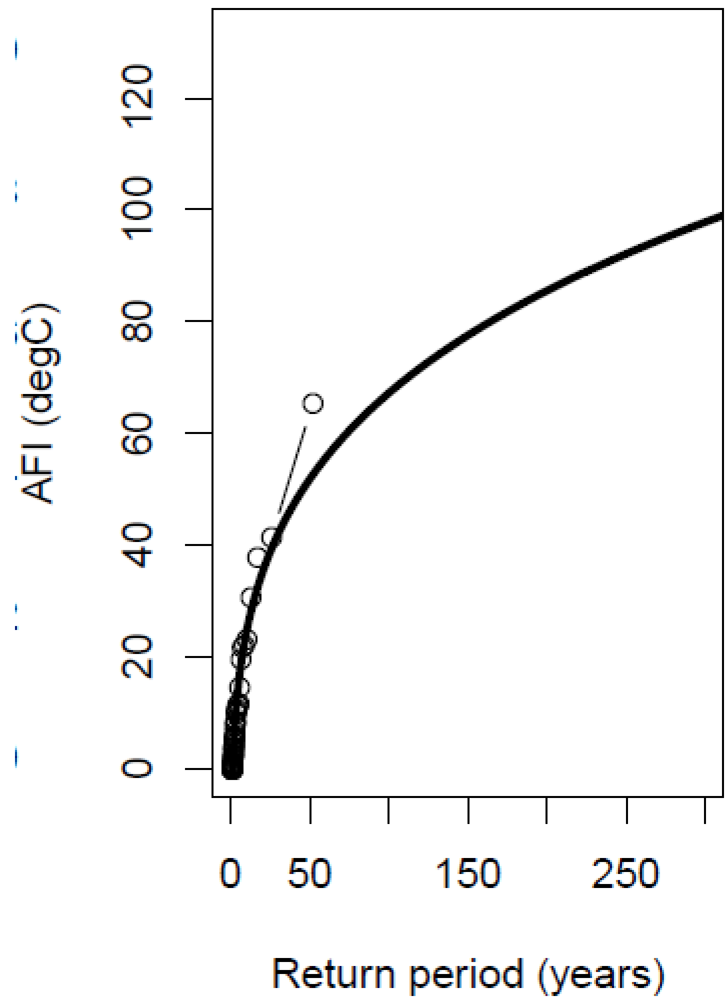
Courtesy John Kennedy, Met Office Hadley Centre

wAFI timeseries – all U.K.

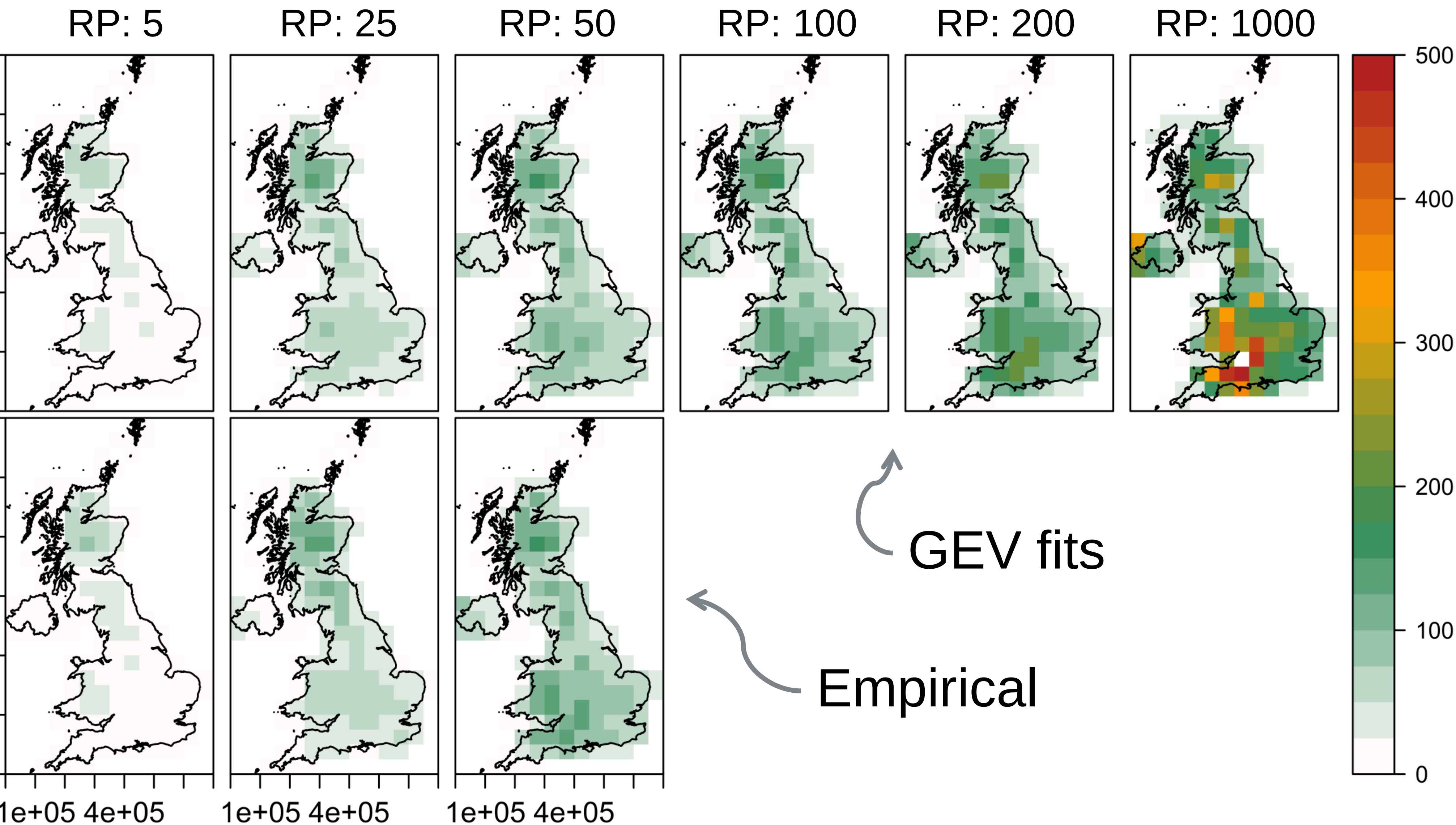


Stochastic hazard generation

- Fit a Generalized Extreme Value (**GEV**) distribution (which includes the Gumbel, the Frechet, and Weibull distributions) *at each grid cell.*
- I use the *Tail-Weighted Maximum Likelihood Estimation* (**TWMLE**) method developed by *Kemp et al*. (2013) to estimate the parameters at each cell.



Return period maps



RVM for Freeze model

- In the UK Freeze model, the joint multivariate hazard distribution has **170 dimensions** (i.e. cells) and it is decomposed as a product of **14,365** pair copula and marginal densities as follows:

$$f(x_1,\ldots,x_{170}) = \prod_{j=1}^{169}\prod_{i=1}^{169} c_{i,(i+j)|(i+1),\ldots,(i+j-1)}\left(F(x_i \mid x_{i_1},\ldots,x_{i_k}), F(x_j \mid x_{i_1},\ldots,x_{i_k})\right)\cdot\prod_{k=1}^{170} f_k(x_k)$$

- We need to find:
  - The bivariate copulas families for all the 14,365 pairs and their parameters
  - The appropriate RVine tree structure

RVM for Freeze model

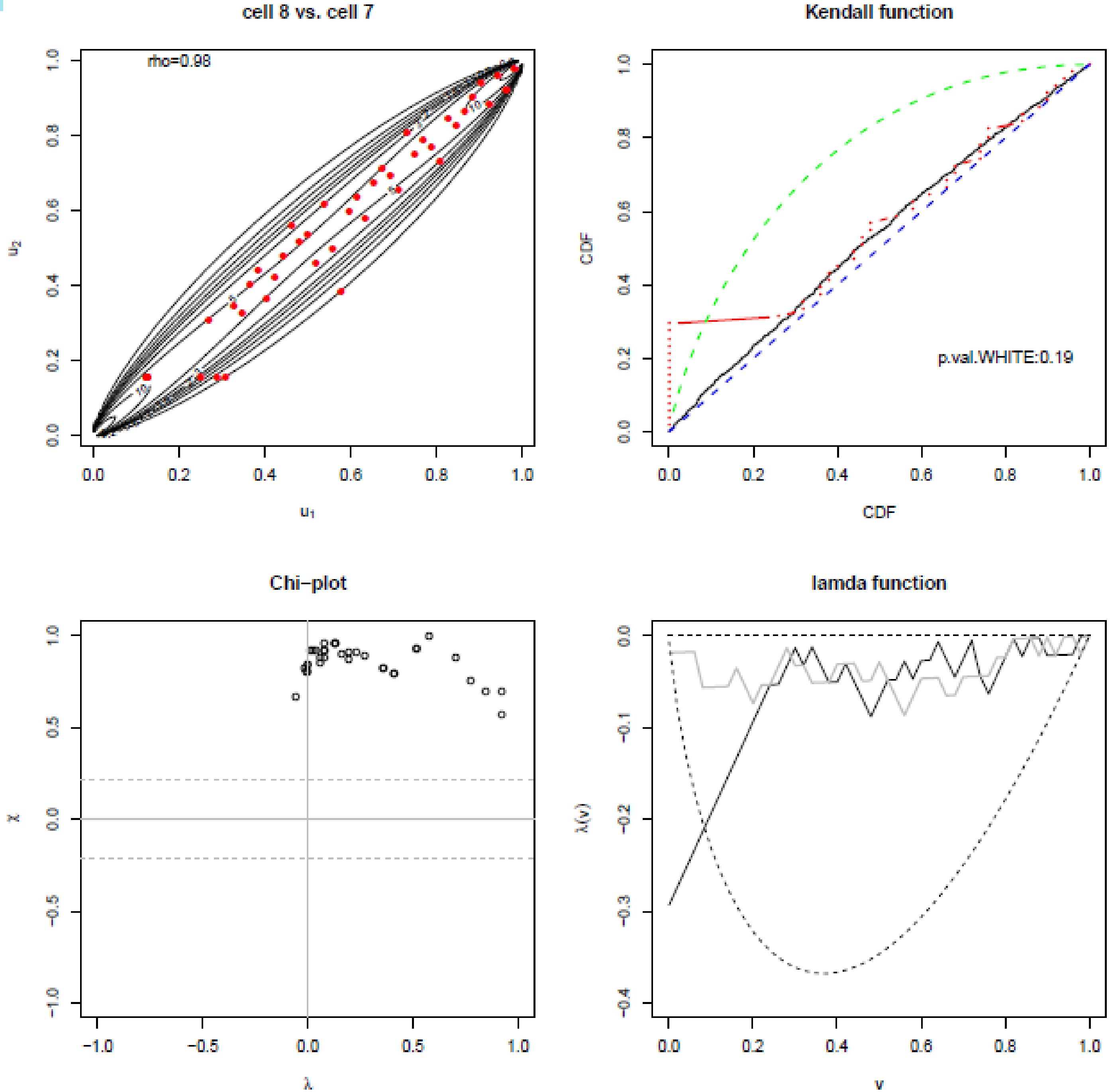
- Fit an RVM model:
  - rvm <- **RVineStructureSelect** (AFI, type = “RVine”, …)
  - ~ 30 min

My laptop:

Processor:	Intel(R) Core(TM) i7-3520M CPU @ 2.90GHz 2.90 GHz
Installed memory (RAM):	8.00 GB (7.87 GB usable)

RVM for Freeze model

- Manually checking the fits between the pair-wise copulas.



RVM for Freeze model

- Fit an RVM model:
  - rvm <- **RVineStructureSelect** (AFI, type = “RVine”, …)
  - ~ 30 min
- Simulate 10K years of events:
  - corrsamples = **RVineSim**(10000, rvm)
  - ~ 5 min

My laptop:

Processor:	Intel(R) Core(TM) i7-3520M CPU @ 2.90GHz 2.90 GHz
Installed memory (RAM):	8.00 GB (7.87 GB usable)

RVM for Freeze model

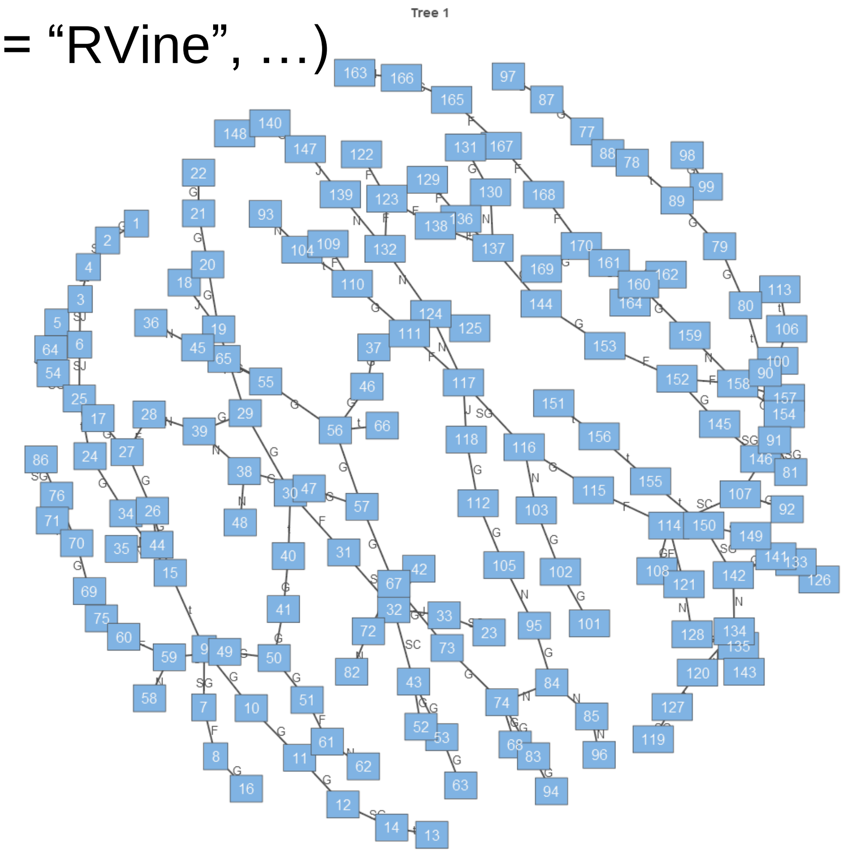
- Fit an RVM model:
  - rvm <- **RVineStructureSelect** (AFI, type = “RVine”, …)
  - ~ 30 min
- Simulate 10K years of events:
  - corrsamples = **RVineSim**(10000, rvm)
  - ~ 5 min
- Goodness-of-fit test:
  - goftest <- **RVineGofTest**(AFI, rvm, …)
  - p.value = 0.725
  - ~ 130 min

My laptop:

Processor:	Intel(R) Core(TM) i7-3520M CPU @ 2.90GHz 2.90 GHz
Installed memory (RAM):	8.00 GB (7.87 GB usable)

RVM for Freeze model

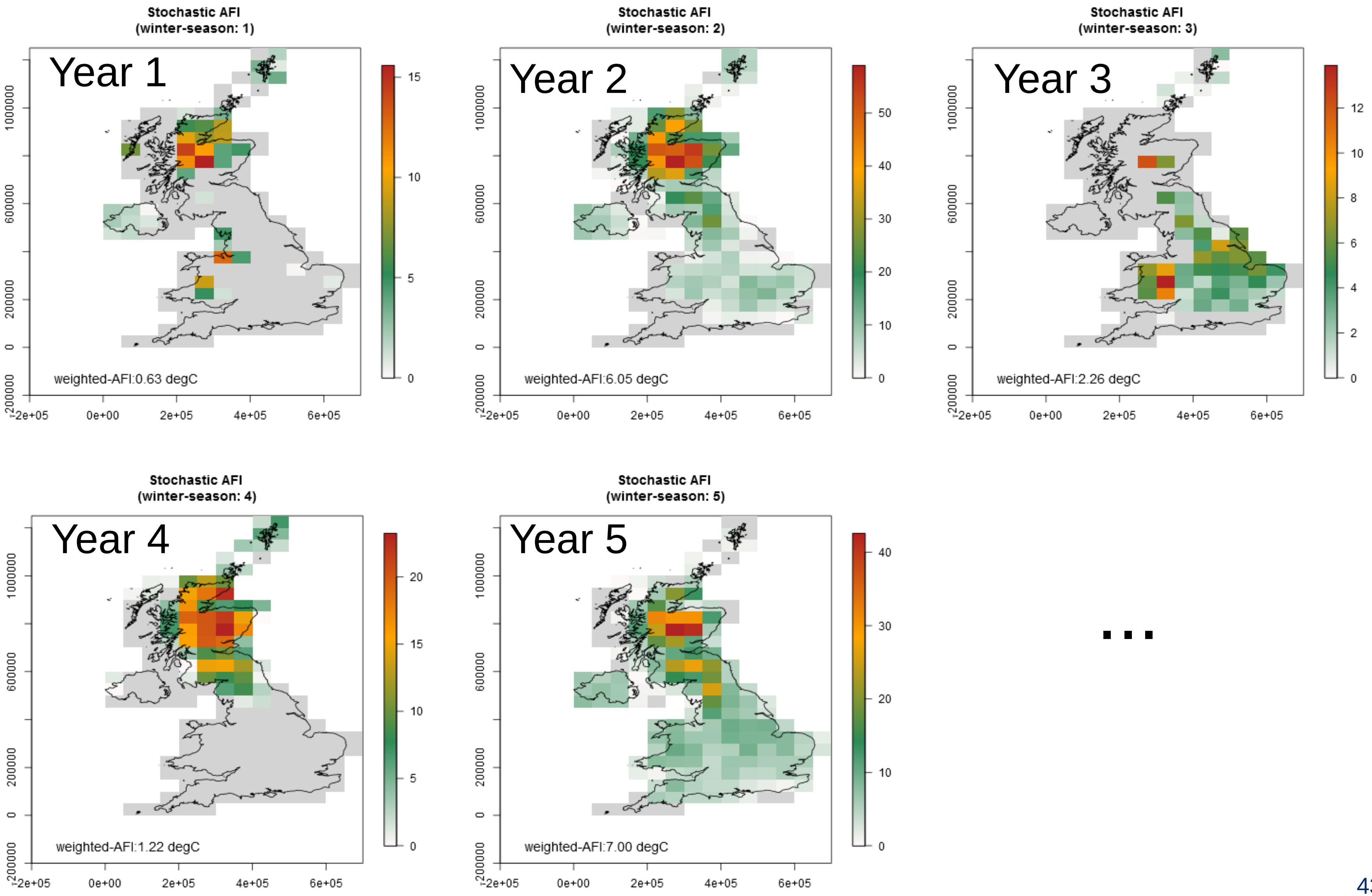
- Fit an RVM model:
  - rvm <- **RVineStructureSelect** (AFI, type = “RVine”, …)
  - ~ 30 min
- Simulate 10K years of events:
  - corrsamples = **RVineSim**(10000, rvm)
  - ~ 5 min
- Goodness-of-fit test:
  - goftest <- **RVineGofTest**(AFI, rvm, …)
  - p.value = 0.725
  - ~ 130 min
- Plot the first level tree:
  - **RVineTreePlot**(rvm, tree=1, …)



My laptop:

Processor:	Intel(R) Core(TM) i7-3520M CPU @ 2.90GHz 2.90 GHz
Installed memory (RAM):	8.00 GB (7.87 GB usable)

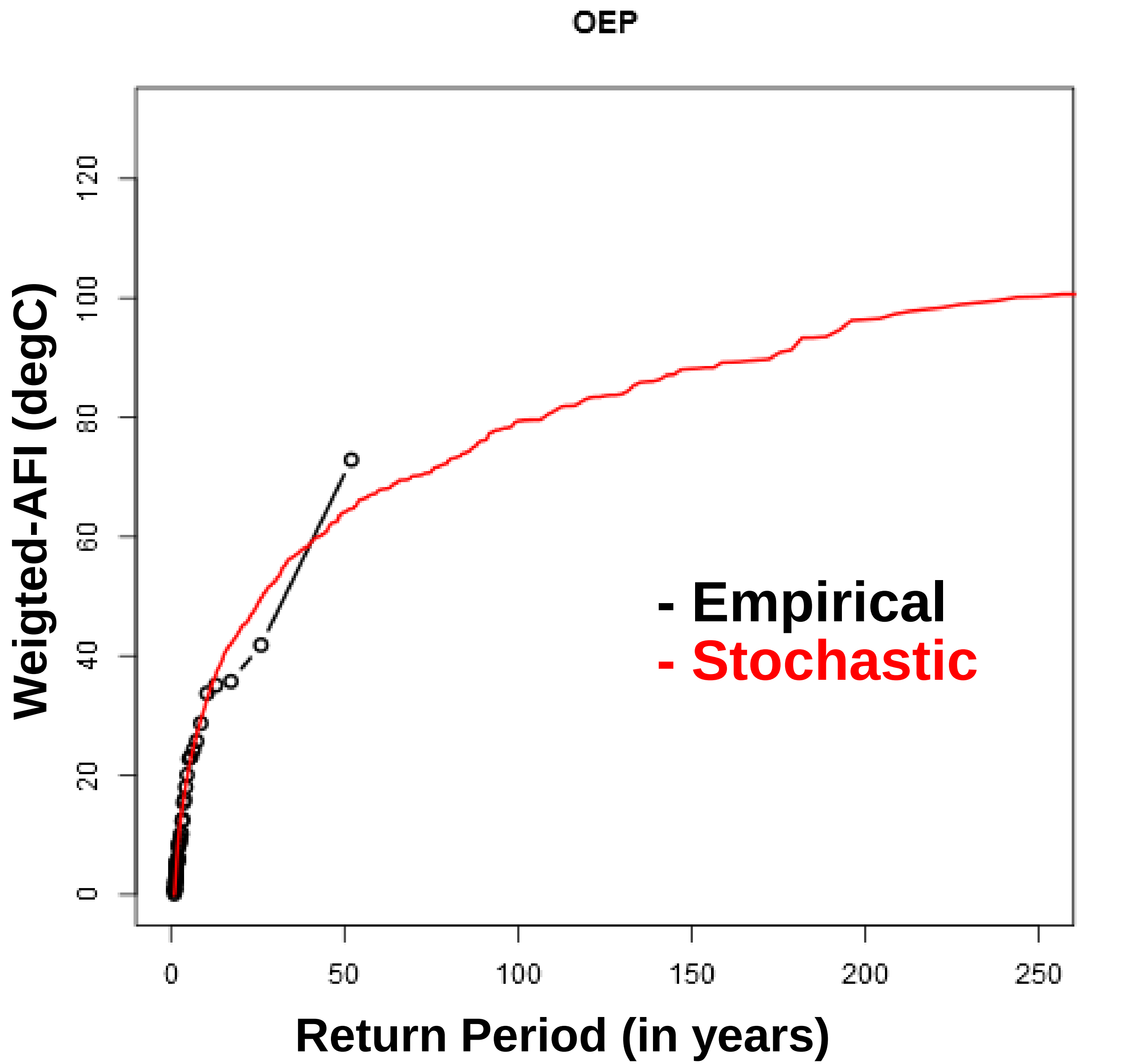
Stochastic set of synthetic events



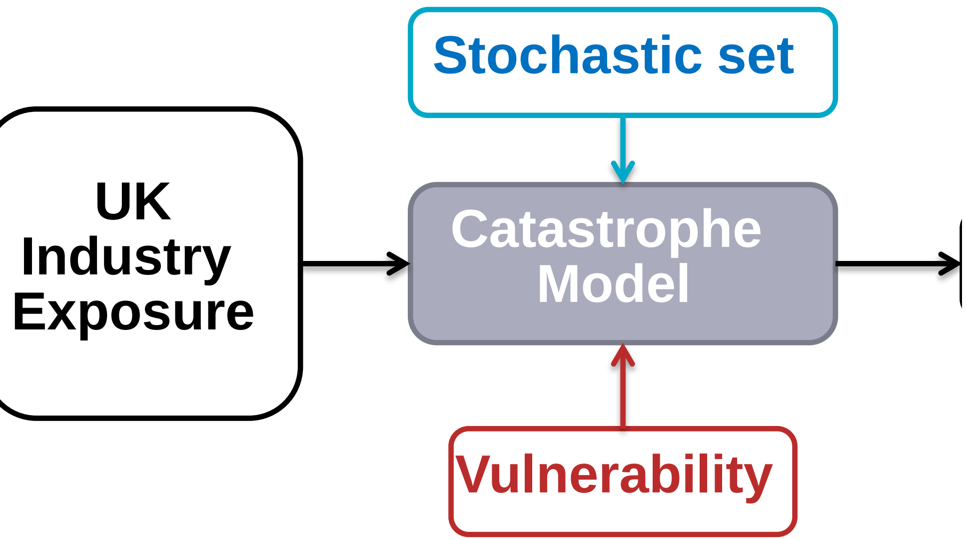
Hazard simulation (weighted AFI)

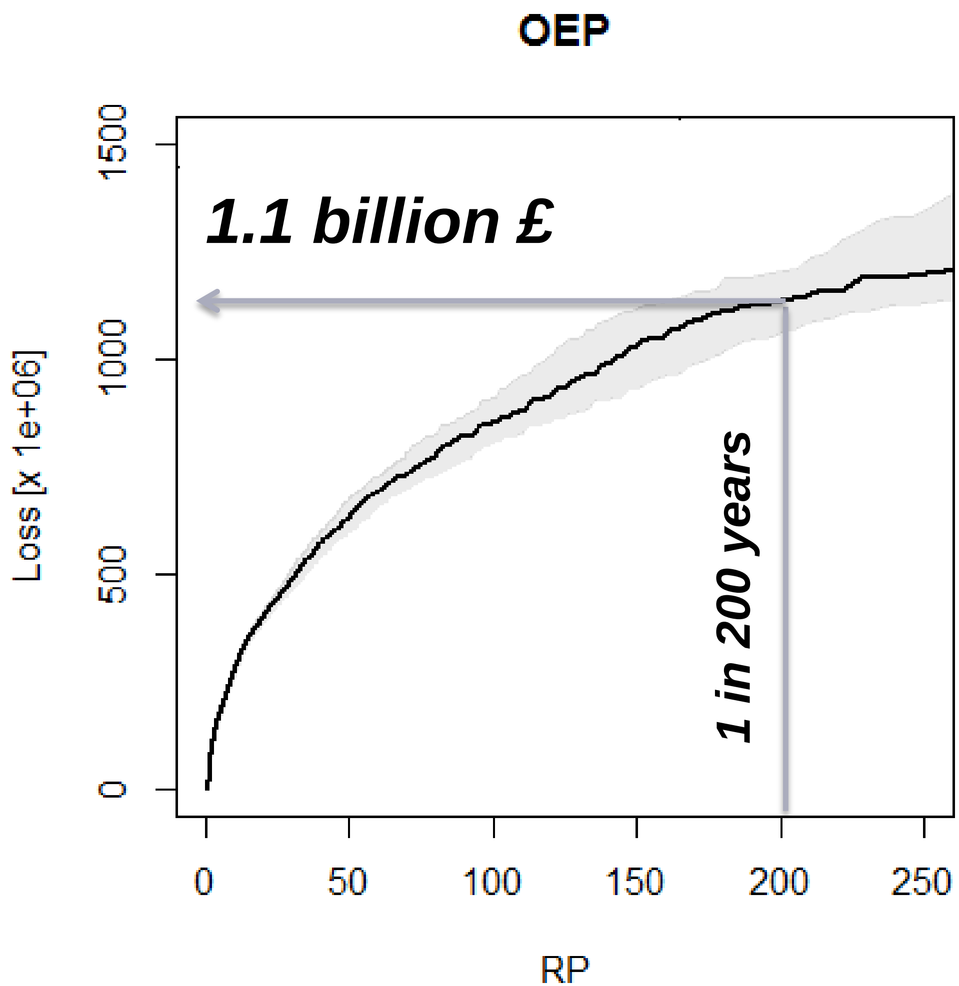
- Winter of 1962/63 (with wAFI of 73 degC) is estimated to be ~ 80-year RP event.

- A 200-year RP event has a weighted AFI of ~ 90degC



Implementation inside a Catastrophe model





Conclusion

- “All models are wrong.. Some are useful”
- Vine copulas can be useful in catastrophe modelling
- Other applications could be:
  - Vulnerability dependence between risks
  - Cross-peril correlations
- The implementation in R is very powerful & easy to use. Many thanks to the authors of the VineCopula and CDVine packages!
- Thank you for your attention!!

Some history of vine models

- Joe (1996) gave a probabilistic construction of multivariate distributions functions based on simple building blocks called pair-copulas.
- Bedford and Cooke (2001) and Bedford and Cooke (2002) organized these constructions in a graphical way called regular vines and gave expression for the joint density.
- Estimation for the Gaussian case was considered in the book by Kurowicka and Cooke (2006).
- Aas et al. (2009) used the PCC construction to construct flexible multivariate copulas based on pair-copulas such as bivariate Gaussian, t-, Gumbel and Clayton copulas and provided likelihood expressions.
- First and second vine workshops took place in Delft in Nov. 2007 and Dec. 2008, a third one took place in Oslo in Dec. 2009. Workshop results are published in Kurowicka and Joe (2011).
- A recent survey about PCC models is Czado (2010).

From “The world of vines”, Czado, (2011)