Nonparametric Statistics Project

HONEYBEE HEALTH

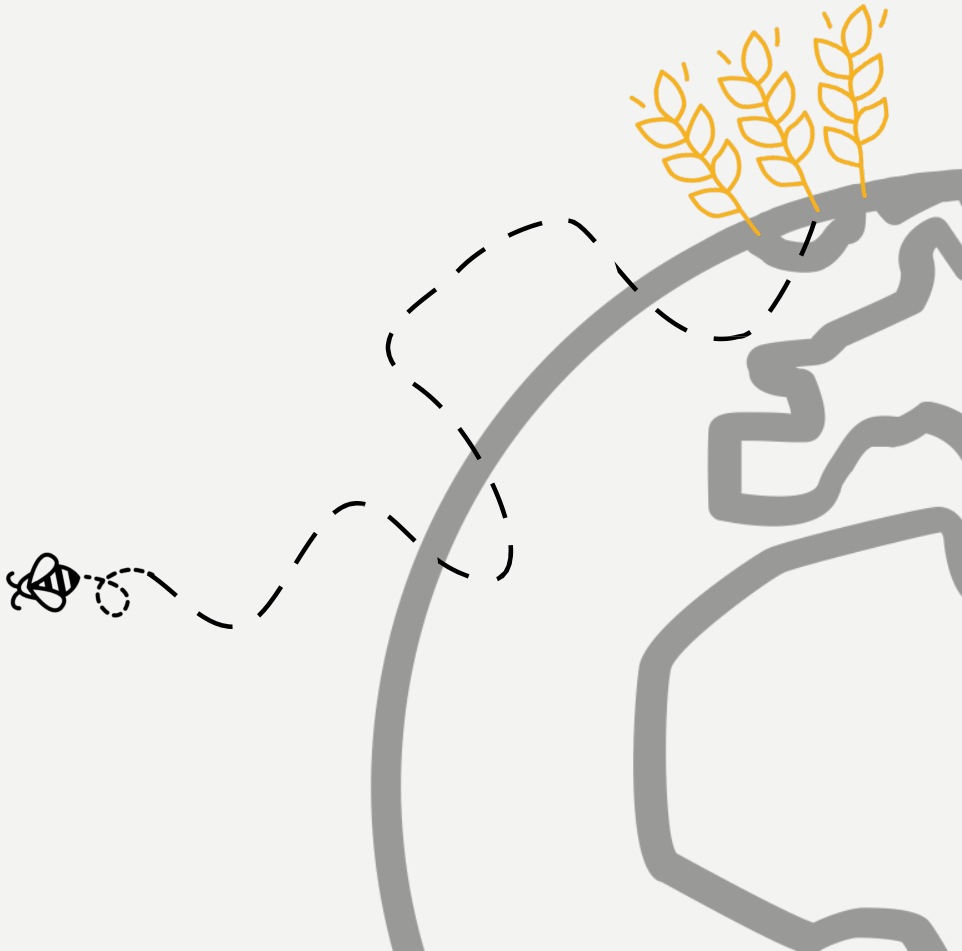
Evaluation of risk factors & strategies to mitigate colony losses

Marta Cerri (5 CFU)
Luca Mainini (5)
Lupo Marsigli (8)
Eugenio Varetti (8)


Why a project on
bee health ?

1/3

Of our global food supply is pollinated by bees



HONEYBEES

are responsible for

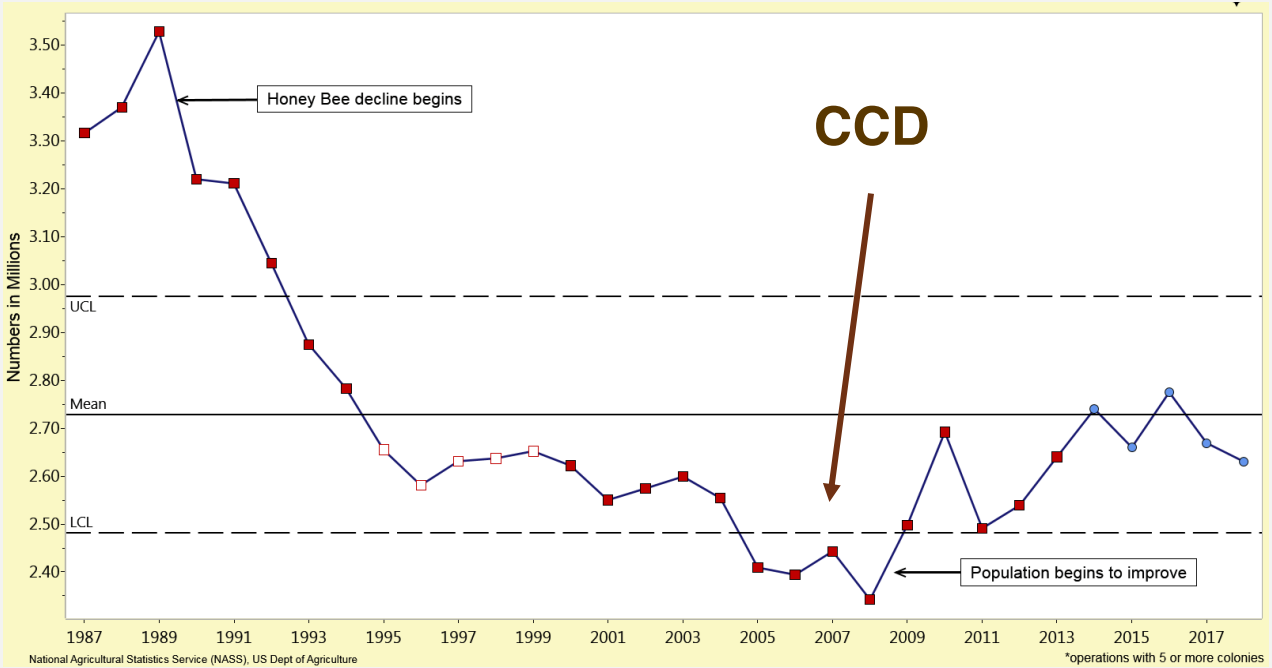
$170 billion*

a year in crops

* $20 Bln only in US (2019, Scott McArt)


WITHOUT BEES
MANY PLANTS
including food crops,
WOULD DIE OFF.

There are significant gaps in our understanding of *honey bee* behaviour



Our project addresses
PUBLIC OPINION
and
GOVERNMENT AUTHORITIES

RESEARCH QUESTIONS

REASONS Is there a relation a between stressors and colony losses?

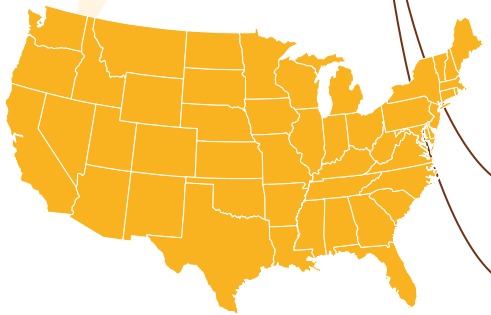
IMPACT Which is the economic impact of the colony losses?

RESILIENCE Which are the most favorable environments?

DATA IMPORT & CLEANING

Data

For each state in US:



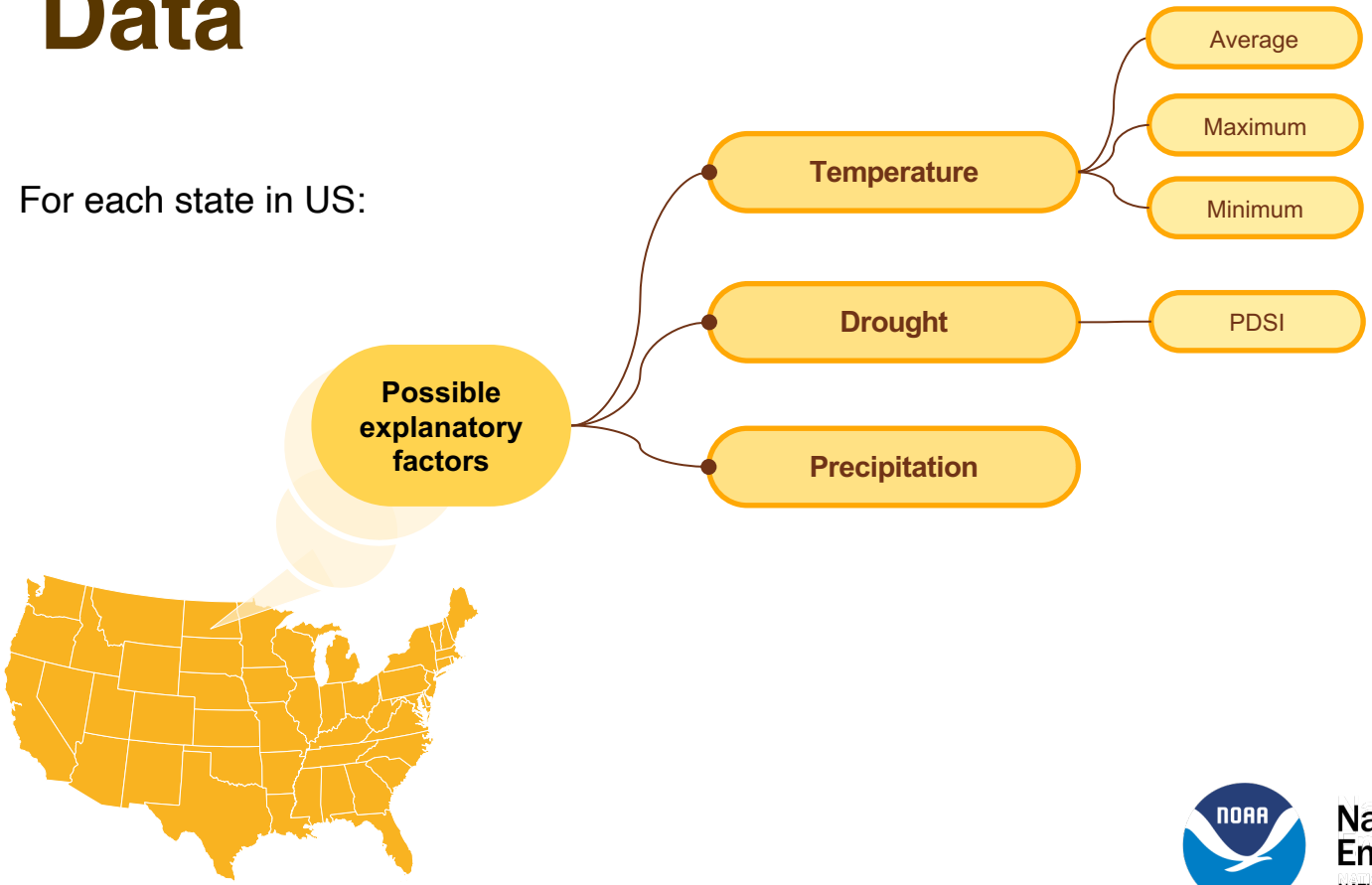
YEAR from 2015 to 2022

- **Q1: January - March**
- **Q2: April - June**
  - Number of colonies — Number of colonies at the beginning of the quarter
  - Max number of colonies — Colonies plus all colonies moved into that state during the quarter
  - Colony lost — Percentage and absolute values of colony lost
  - Colony added — Colony that was either created or purchased
  - Colony renovated — Number of surviving colonies that were renewed
  - Stressors — Percentage of colonies affected by different types of stressors (1 feature each)
    - Diseases
    - Pesticides
    - Other
    - Varroa mites
- **Q3: July - September**
- **Q4: October - December**
- Honey Production — Annual yields per state
- Honey Price — Average annual honey price per state

USDA

Data

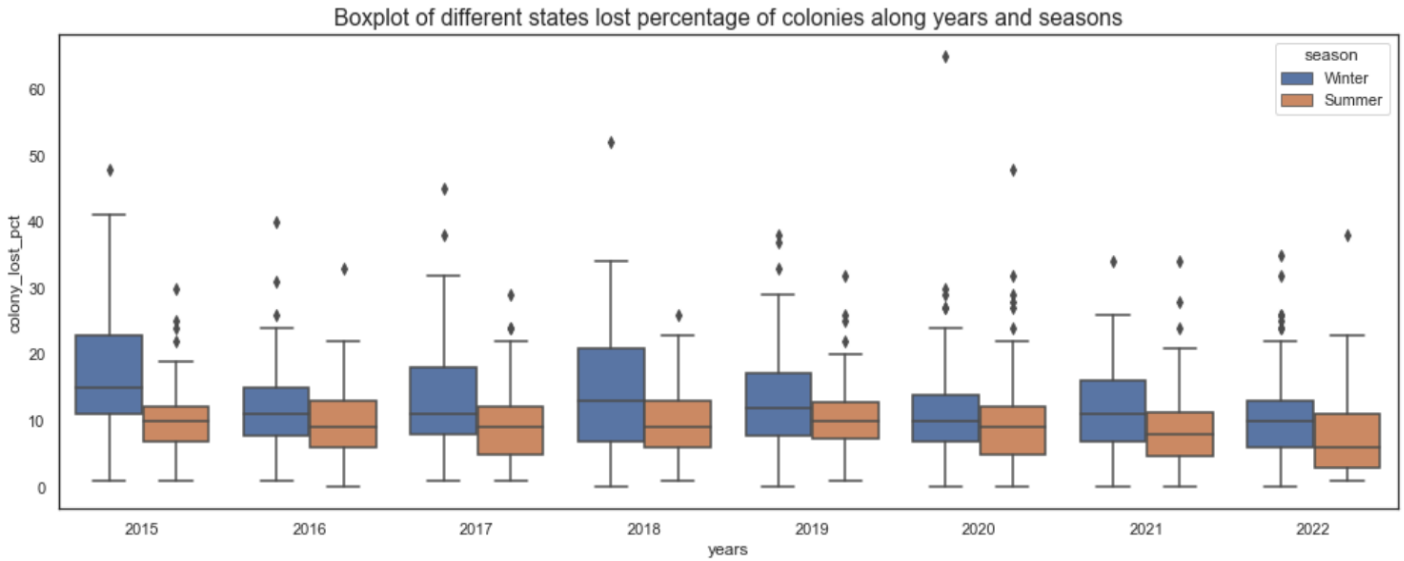
For each state in US:



National Centers for Environmental Information
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

GENERAL TRENDS IN DATA

Losses are higher in winter

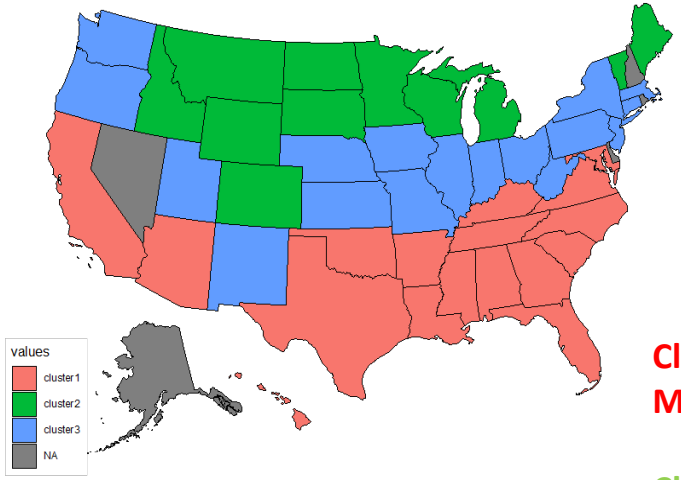


Global Permutational test for functional data: there is statistical evidence of difference in distribution between summer and winter.

INFLUENCE OF THE WEATHER

An apparent paradox: increased colony mortality in warmer states

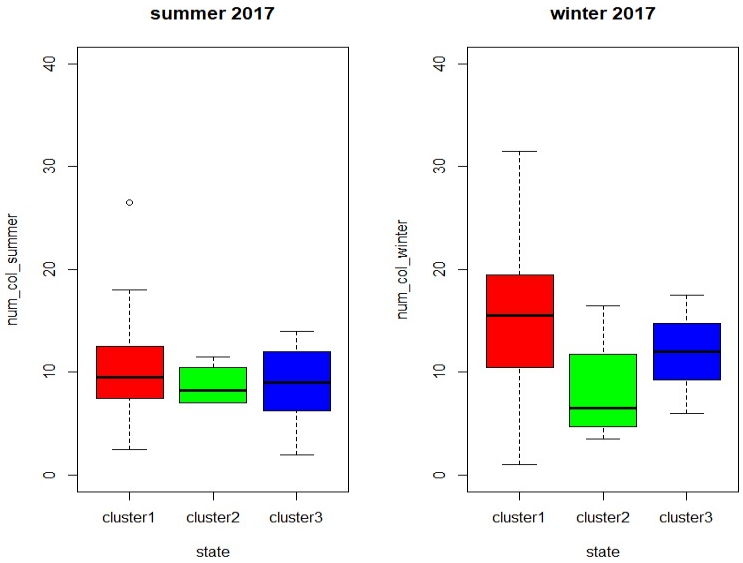
Cluster by minimum temperature



Cluster 1:Subtropical/ Mediterranean

Cluster 2: Alpine

Cluster 3: Continental



Cold winters do not kill a healthy colony

- As temperature drops below 10°C, bees come together to form a "**winter cluster**" (they can stay at 25-30°C).
- During late autumn, **cold spells** with **adequate duration and intensity** **trigger colony wintering** at the right time

IMPACT ON BEEKEEPERS

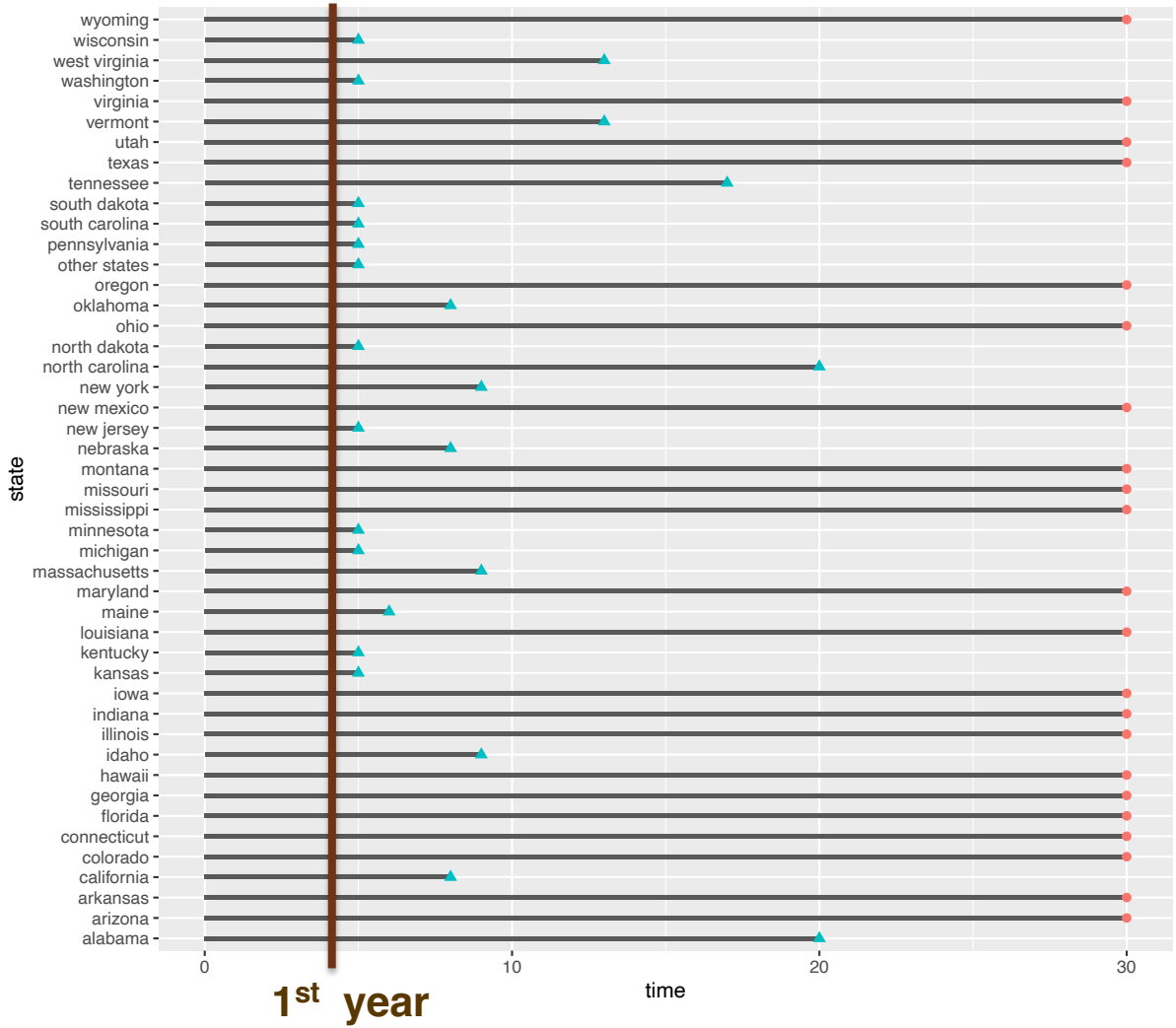
Should government be worried about beekeepers not able to cope with colony losses?

Introduction of a new metric: cumulative "*after-action*" loss

$$AAL_t = AAL_{t-1} + \frac{Colony\ loss_t - Colony\ added_t}{Colony\ max_t}$$

Under a survival-analysis approach,
Time to death = first time when $AA\ Loss > 20\%$

In many states, beekeepers do not cope with first year losses



Dry conditions and fall heatwaves are linked with increased colony mortality

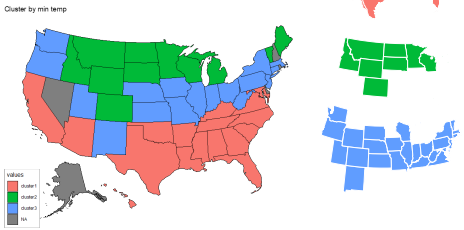
$$h_i(t|X_i) = h_0(t)exp(\beta_1 AverageTemp_i + \beta_2 Precipitation_i + \beta_3 PDSI_i + \beta_4 cluster2_i + \beta_5 cluster3_i)$$

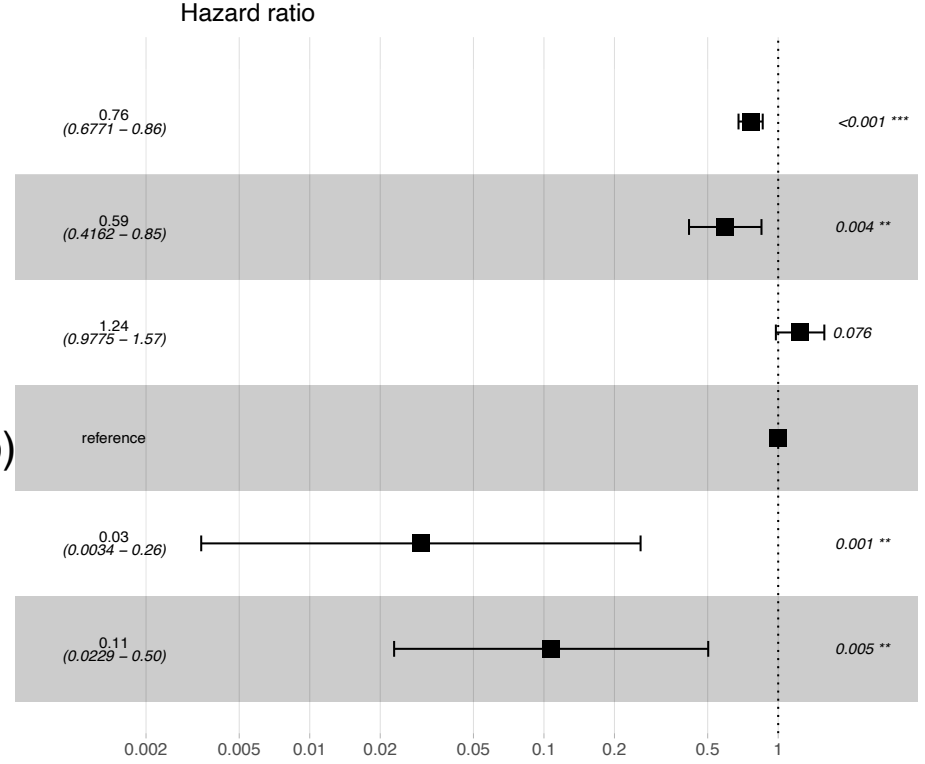
Average Temperature

Precipitation

Palmer Drough Index

Group 1  (min_temp)





BIO-CHEMICAL STRESSING FACTORS

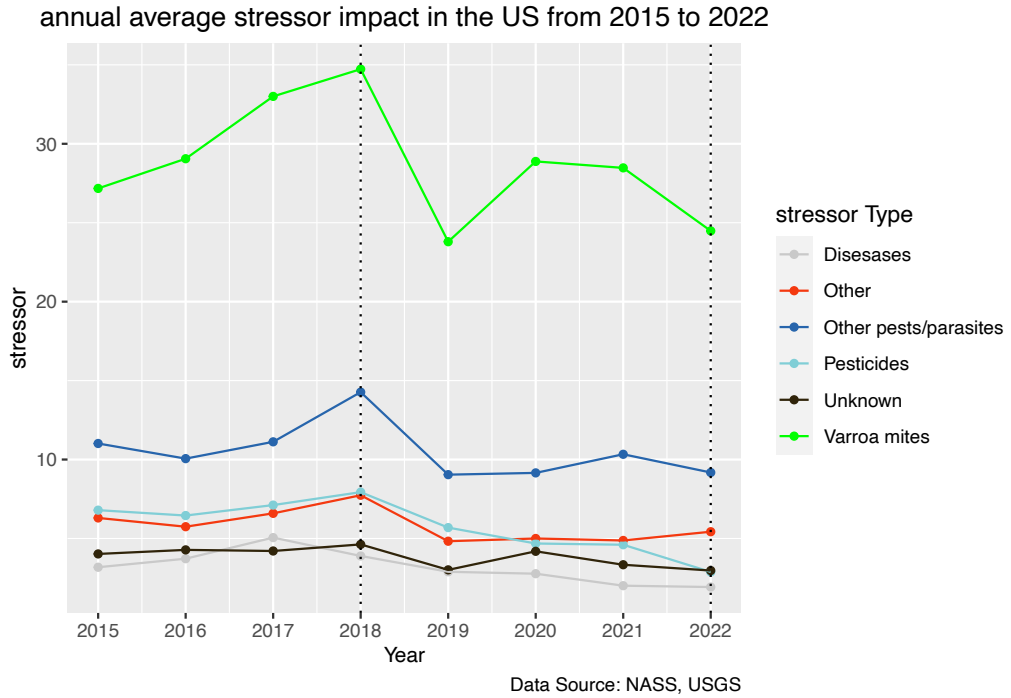
Pesticides seem the main causes of summer losses

0.66 Spearman correlation (Pesticides, Loss %)

0.54 Spearman correlation (Diseases, Loss %)

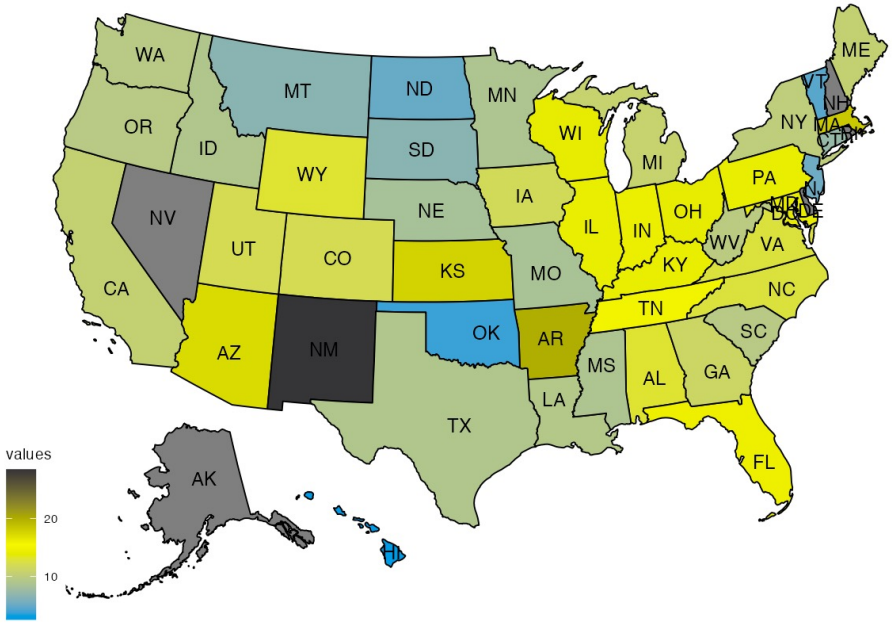
0.41 Spearman correlation (Varroa Mite, Loss %)

Varroa destructor: the most devastating honey bee pest

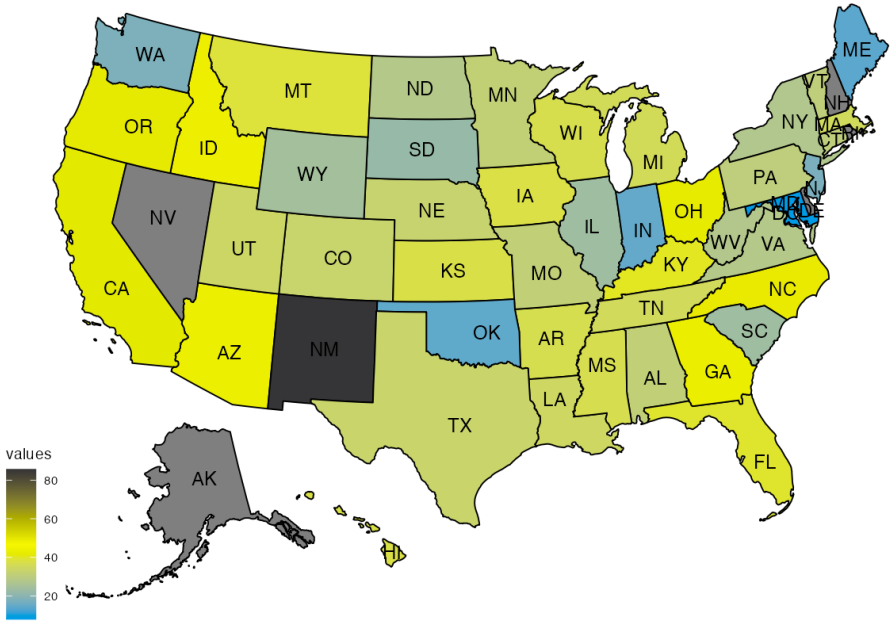


In some years, number of losses and Varroa show the same pattern

Lost colonies (%)



Stressed colonies by Varroa mites (%)

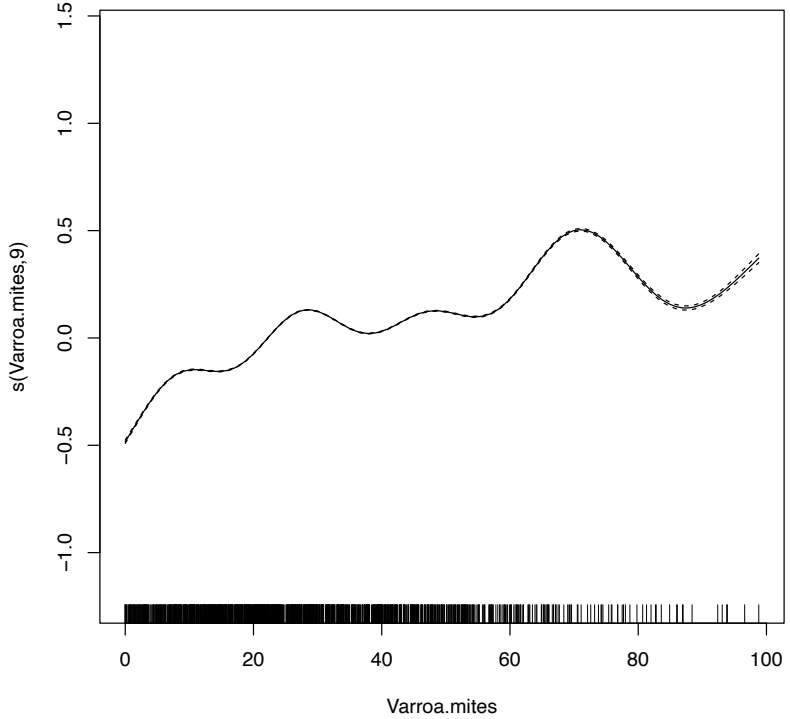


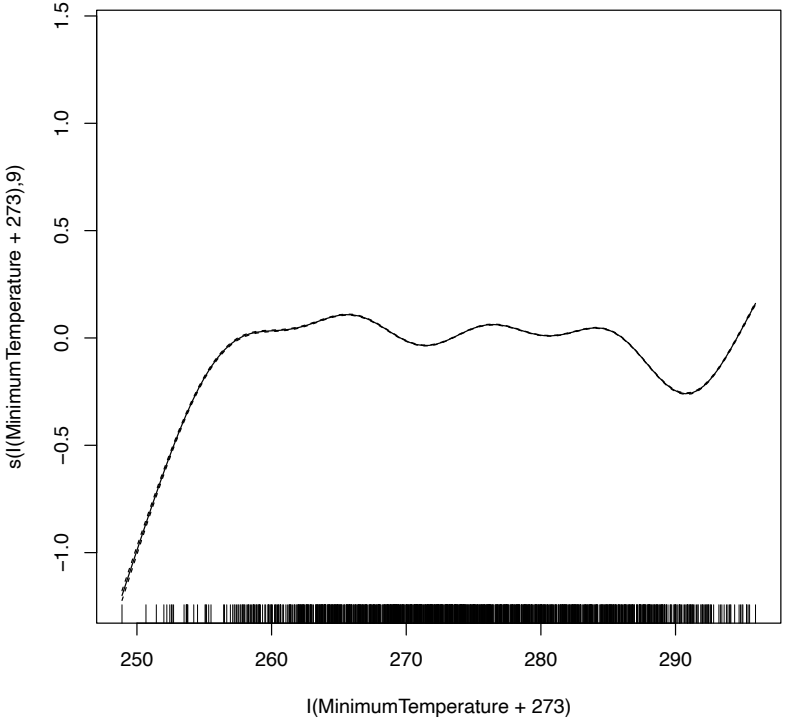
MODELLING OF LOSS - STRESSORS RELATION

Generalized Additive Models

$$n_i \sim Bin(Max\ n_i, \pi_i)$$

$$logit(\pi_i) = f(Stressors_i) + f(MinTemp_i) + f(space, time) + \epsilon_i$$

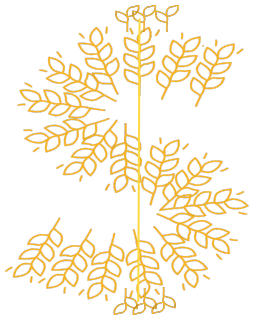




ECONOMIC IMPACT

Penalized Semiparametric Regression Model for Spatial Functional Data

$$z_{ij} = \boldsymbol{w}_{ij}^{T}\boldsymbol{\beta} + f(\boldsymbol{p}_i, t_j) + \epsilon_{ij}$$ state i=1,…,44, time j=1,…,29



$\boldsymbol{p}_i$: $(longitude_i, latitude_i) \in \Omega$ of the centroid of each state
$t_j \in [T1, T2]$, quarters from 2015-Q1 to 2022-Q2
$\boldsymbol{w}_{ij}$: vector of $(Varroa_{ij}, Pesticides_{ij})$ at location $\boldsymbol{p}_i$ and time t_j
$\boldsymbol{\beta}$: vector of 2 parametric coefficients
ϵ_{ij}: iid residuals with mean $= 0$ and constant std

z_{ij}: **amount of money loss in Kdollars for 100 colonies present in location $\boldsymbol{p}_i$ at quarter $\boldsymbol{t}_j$**

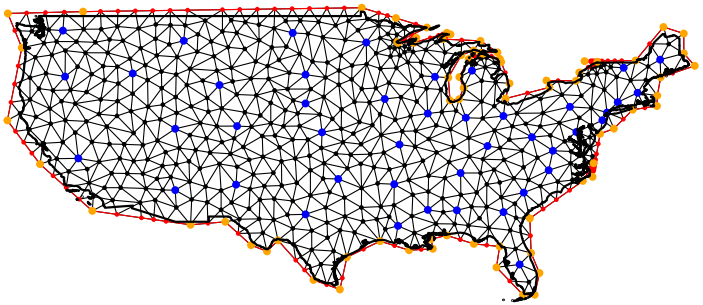
Minimize the following penalized sum of squares criterion:

$$J(f,\beta) = \sum_{i=1}^{44}\sum_{j=1}^{29}\left(z_{ij} - w_{ij}^{T}\beta - f(p_i,t_j)\right)^2 + \lambda_S \int_{T_1}^{T_2}\int_{\Omega}\left(\boldsymbol{\Delta f(p,t)}\right)^2 d\,p\,dt + \lambda_T \int_{\Omega}\int_{T_1}^{T_2}\left(\frac{\boldsymbol{\partial^2 f(p,t)}}{\boldsymbol{\partial t^2}}\right)^2 d\,t\,dp$$

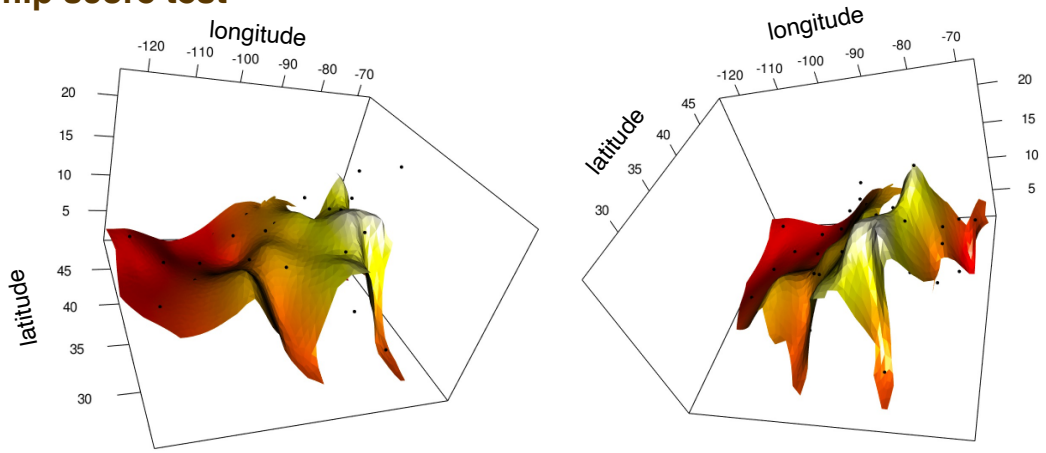
Penalized Semiparametric Regression Model for Spatial Functional Data

- Isotropic smoothing
- Space: Linear Finite Element Basis
- Time: cubic B-spline basis
- Non-tensor product basis to account for how the shape of the domain influences the phenomenon
- Smoothing parameters λ_S, λ_T chosen via GCV
- Inference on $\boldsymbol{\beta}$: **Nonparametric Eigen sign-flip score test**

Main Simplification:
Areal data are assigned to the centroid of the state

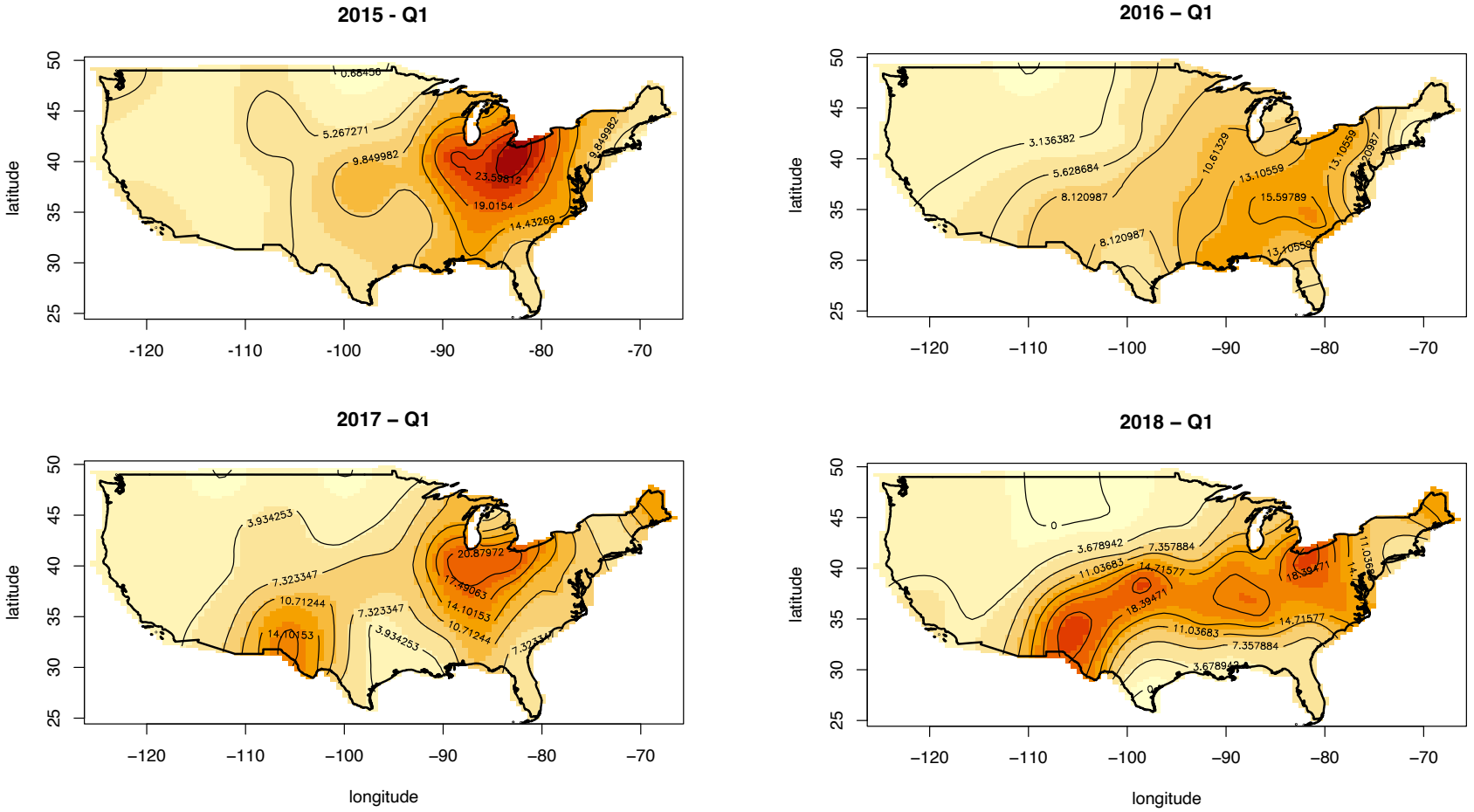


US Mesh

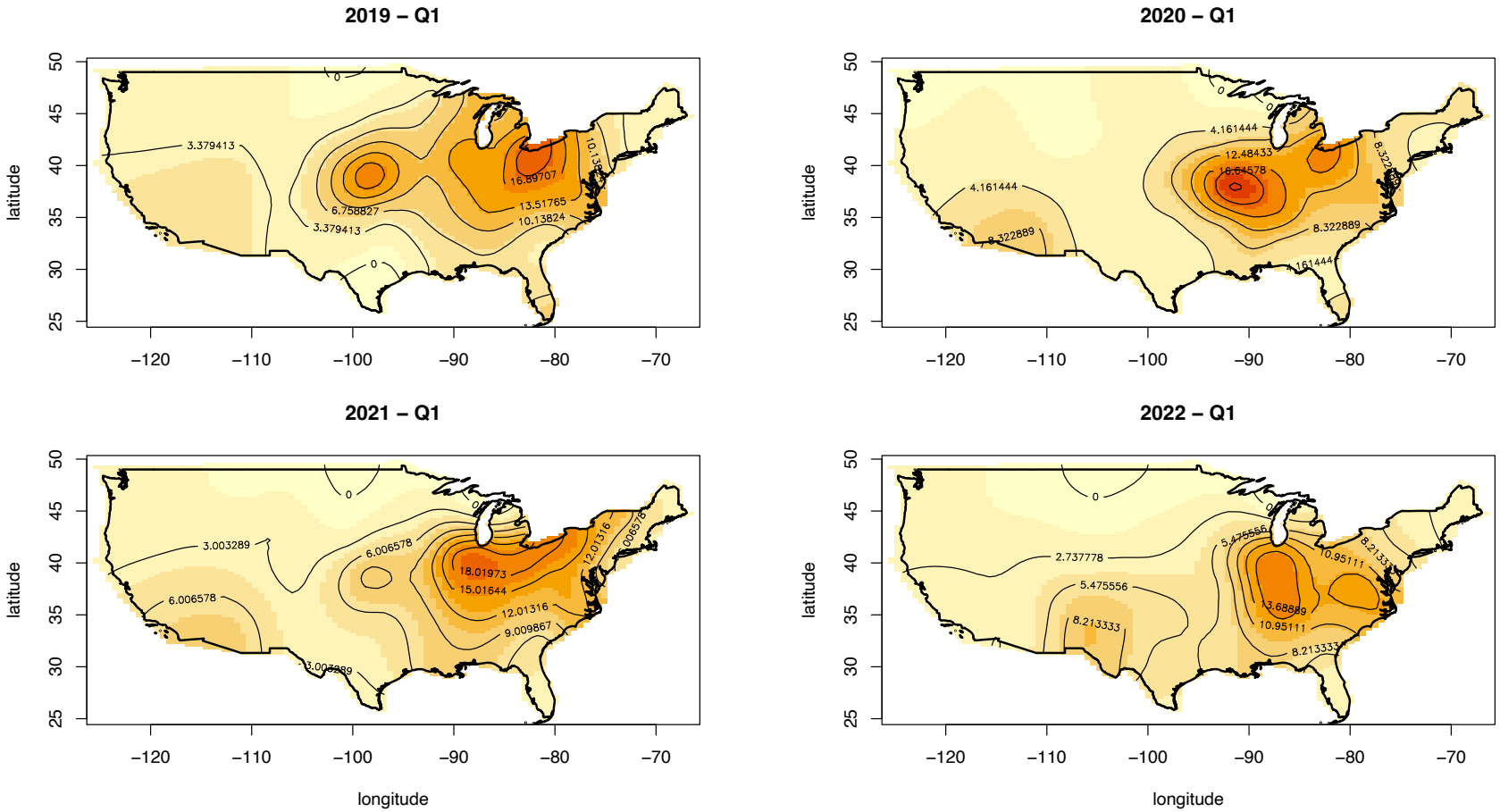


2016-Q1: Estimated spatio-temporal field for the money lost (in Kdollars) every 100 colonies

Overall money loss (k$) every 100 colonies in the quarter reported

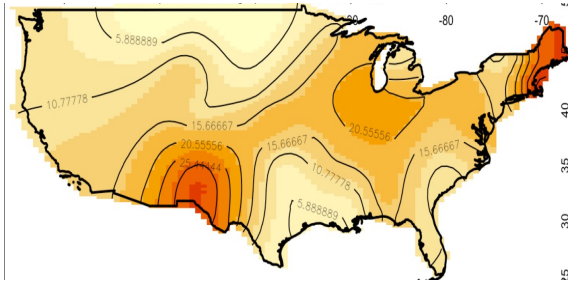


Overall money loss (k$) every 100 colonies in the quarter reported

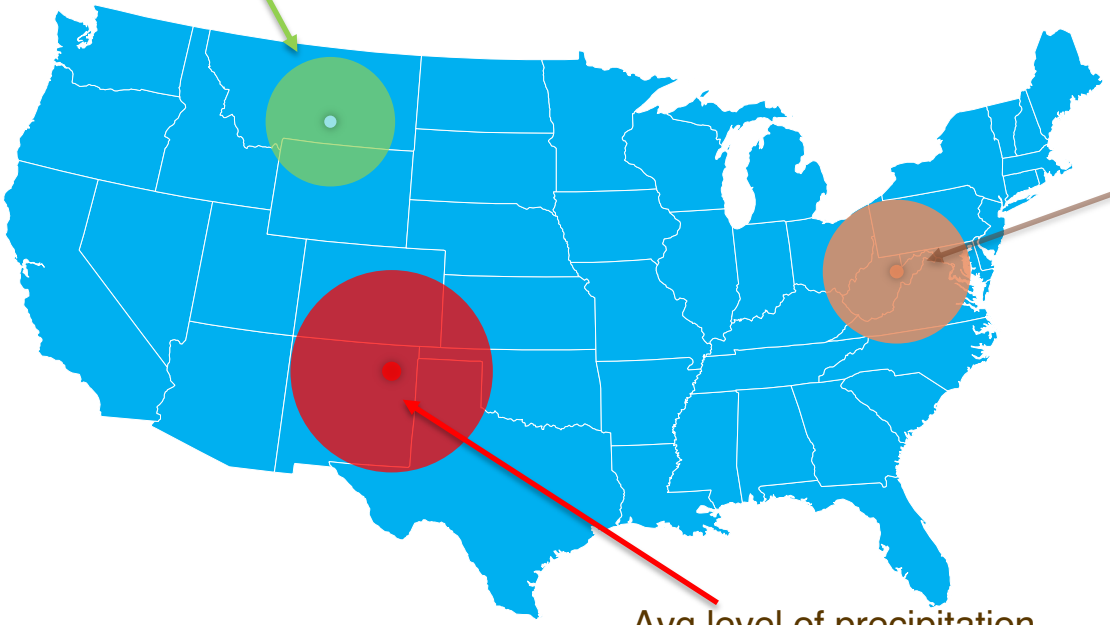


CONCLUSIONS

2017 Colony loss (%)



Good level of precipitation
Cold fall
Medium level of Varroa-Mites



high level of precipitation
mild fall
Medium level of Varroa-Mites

Avg level of precipitation
Warm fall
High level of Varroa-Mites

Reduced-risk zone

High-risk zone

MAIN REFERENCES

- Mara S Bernardi et al. *‘A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province’.*
- Eleonora Arnone et al. *‘Modeling spatially dependent functional data via regression with differential regularization’*
- Federico Ferraccioli, Laura M Sangalli and Livio Finos. *‘Some first inferential tools for spatial regression with differential regularization’.*
- Alain F. Zuur, Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev, Graham M. Smith. *"Mixed Effects Models and Extensions in Ecology with R"*

THANK YOU

together we can...