Partially-interpretable neural networks for high-dimensional extreme quantile regression: With application to U.S. wildfires

Jordan Richards, Raphaël Huser

King Abdullah University of Science and Technology (KAUST)

August 8, 2022

Motivation

- Wildfires cause significant death and damage across the world
- Recent years have seen devastating wildfires in the (west) U.S. - 100s of deaths and millions of acres of destroyed land
- **Frequency** + **severity** to be exasperated by **climate change**
- In 2021, global wildfires contributed to ≈ 1760 megatonnes of **carbon emissions** - High proportion from the U.S.
- To mitigate risk, need to identity **drivers** and **high-risk** areas - Can we do these both simultaneously?

Motivation

- Wildfires cause significant death and damage across the world
- Recent years have seen devastating wildfires in the (west) U.S. - 100s of deaths and millions of acres of destroyed land
- **Frequency** + **severity** to be exasperated by **climate change**
- In 2021, global wildfires contributed to $\approx$ 1760 megatonnes of **carbon emissions** - High proportion from the U.S.
- To mitigate risk, need to identity **drivers** and **high-risk** areas - Can we do these both simultaneously?

Motivation

- Wildfires cause significant death and damage across the world
- Recent years have seen devastating wildfires in the (west) U.S. - 100s of deaths and millions of acres of destroyed land
- **Frequency + severity** to be exasperated by **climate change**
- In 2021, global wildfires contributed to $\approx$ 1760 megatonnes of **carbon emissions** - High proportion from the U.S.
- To mitigate risk, need to identity **drivers** and **high-risk** areas - Can we do these both simultaneously?

Motivation

- Wildfires cause significant death and damage across the world
- Recent years have seen devastating wildfires in the (west) U.S. - 100s of deaths and millions of acres of destroyed land
- **Frequency + severity** to be exasperated by **climate change**
- In 2021, global wildfires contributed to $\approx$ 1760 megatonnes of **carbon emissions** - High proportion from the U.S.
- To mitigate risk, need to identity **drivers** and **high-risk** areas - Can we do these both simultaneously?

Motivation

- Wildfires cause significant death and damage across the world
- Recent years have seen devastating wildfires in the (west) U.S. - 100s of deaths and millions of acres of destroyed land
- **Frequency + severity** to be exasperated by **climate change**
- In 2021, global wildfires contributed to $\approx$ 1760 megatonnes of **carbon emissions** - High proportion from the U.S.
- To mitigate risk, need to identity **drivers** and **high-risk** areas - Can we do these both simultaneously?

Extreme quantile regression

We perform quantile regression with the response taken to be **aggregated burnt area** (BA) for a spatio-temporal grid-box.

- Interested in upper-tails, i.e., most dangerous wildfires
- Typical quantiles of interest will be larger than those previously observed $\Rightarrow$ non-parametric quantile regression perform poorly
- Instead turn to parametric regression using **asymptotically-justified extreme value distributions**
- Three classics: GEV, GPD and PP models - We focus on PP extension as its parameters are **easier to interpret**, but the framework is **applicable for any** of the three

Extreme quantile regression

We perform quantile regression with the response taken to be **aggregated burnt area** (BA) for a spatio-temporal grid-box.

- Interested in upper-tails, i.e., most dangerous wildfires
- Typical quantiles of interest will be larger than those previously observed $\Rightarrow$ non-parametric quantile regression perform poorly
- Instead turn to parametric regression using **asymptotically-justified extreme value distributions**
- Three classics: GEV, GPD and PP models - We focus on PP extension as its parameters are **easier to interpret**, but the framework is **applicable for any** of the three

Extreme quantile regression

We perform quantile regression with the response taken to be **aggregated burnt area** (BA) for a spatio-temporal grid-box.

- Interested in upper-tails, i.e., most dangerous wildfires
- Typical quantiles of interest will be larger than those previously observed ⇒ non-parametric quantile regression perform poorly
- Instead turn to parametric regression using **asymptotically-justified extreme value distributions**
- Three classics: GEV, GPD and PP models - We focus on PP extension as its parameters are **easier to interpret**, but the framework is **applicable for any** of the three

Extreme quantile regression

We perform quantile regression with the response taken to be **aggregated burnt area** (BA) for a spatio-temporal grid-box.

- Interested in upper-tails, i.e., most dangerous wildfires
- Typical quantiles of interest will be larger than those previously observed $\Rightarrow$ non-parametric quantile regression perform poorly
- Instead turn to parametric regression using **asymptotically-justified extreme value distributions**
- Three classics: GEV, GPD and PP models - We focus on PP extension as its parameters are **easier to interpret**, but the framework is **applicable for any** of the three

Existing approaches

Existing approaches for **parametric extreme quantile regression** represent $\boldsymbol{\theta}$ as lin. or add. functions of predictors $\mathbf{x} \in \mathbb{R}^d$, i.e., $\boldsymbol{\theta}(\mathbf{x})$

- Linear models are unable to capture non-linear structure so perform poorly for complex problems, e.g., wildfire occurrence and spread
- Spline-based regression models can capture non-linear relationships, but scale poorly to high dimensions - We consider $d = 30$ predictors

We instead use **deep learning** methods as these can **(i)** capture complex structure in $\mathbf{x}$, **(ii)** scale well to **high dimensions** and **(iii)** facilitate **high predictive accuracy**.

Existing approaches

Existing approaches for **parametric extreme quantile regression** represent $\boldsymbol{\theta}$ as lin. or add. functions of predictors $\mathbf{x} \in \mathbb{R}^d$, i.e., $\boldsymbol{\theta}(\mathbf{x})$

- Linear models are unable to capture non-linear structure so perform poorly for complex problems, e.g., wildfire occurrence and spread
- Spline-based regression models can capture non-linear relationships, but scale poorly to high dimensions - We consider $d = 30$ predictors

We instead use **deep learning** methods as these can **(i)** capture complex structure in $\mathbf{x}$, **(ii)** scale well to **high dimensions** and **(iii)** facilitate **high predictive accuracy**.

Partially interpretable neural networks

Statisticians generally avoid the use of neural networks.

- Neural networks (NNs) are "black box" in the sense that it's difficult/impossible to interpret their output - no good for understanding the drivers of risk
- We extend the approach of [Zhong and Wang, 2021] (who propose "partially-linear" NNs) and create NNs that are "**partially-interpretable**" (PINN)
- The effect of some predictors **can be interpreted** whilst the rest feed a neural network to improve **predictive accuracy**

Partially interpretable neural networks

Statisticians generally avoid the use of neural networks.

- Neural networks (NNs) are "black box" in the sense that it's difficult/impossible to interpret their output - no good for understanding the drivers of risk
- We extend the approach of [Zhong and Wang, 2021] (who propose "partially-linear" NNs) and create NNs that are "**partially-interpretable**" (PINN)
- The effect of some predictors **can be interpreted** whilst the rest feed a neural network to improve **predictive accuracy**

Partially interpretable neural networks

Statisticians generally avoid the use of neural networks.

- Neural networks (NNs) are “black box" in the sense that it’s difficult/impossible to interpret their output - no good for understanding the drivers of risk
- We extend the approach of [Zhong and Wang, 2021] (who propose “partially-linear" NNs) and create NNs that are “**partially-interpretable**" (PINN)
- The effect of some predictors **can be interpreted** whilst the rest feed a neural network to improve **predictive accuracy**

Proposed model

Let the response follow $\mathcal{F}(\boldsymbol{\theta}(\mathbf{x}))$ with parameter set $\boldsymbol{\theta}(\mathbf{x}) = (\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \dots)$. Then for all $i = 1, 2, \dots,$

- Split predictor set $\mathbf{x}$ into two **complementary** subsets $\mathbf{x}_{\mathcal{I}}^{(i)}$ and $\mathbf{x}_{\mathcal{N}}^{(i)}$ - "interpreted" and "non-interpreted"
- Let
$$\theta_i(\mathbf{x}) = h_i[\eta_0^{(i)} + m_{\mathcal{I}}^{(i)}(\mathbf{x}_{\mathcal{I}}^{(i)}) + m_{\mathcal{N}}^{(i)}(\mathbf{x}_{\mathcal{N}}^{(i)})],$$
for constant intercept $\eta_0^{(i)} \in \mathbb{R}$ and link $h_i : \mathbb{R} \to \mathbb{R}$
- Interpretable: $m_{\mathcal{I}}^{(i)}$, e.g., linear, spline. Neural network: $m_{\mathcal{N}}^{(i)}$.
- Our framework applies for **any generic parametric distribution** $\mathcal{F}$, e.g., Bernoulli for occurrence, as well as for non-parametric quantile regression.

Proposed model

Let the response follow $\mathcal{F}(\boldsymbol{\theta}(\mathbf{x}))$ with parameter set $\boldsymbol{\theta}(\mathbf{x}) = (\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \ldots)$. Then for all $i = 1, 2, \ldots,$

- Split predictor set $\mathbf{x}$ into two **complementary** subsets $\mathbf{x}_{\mathcal{I}}^{(i)}$ and $\mathbf{x}_{\mathcal{N}}^{(i)}$ - "interpreted" and "non-interpreted"
- Let
$$\theta_i(\mathbf{x}) = h_i[\eta_0^{(i)} + m_{\mathcal{I}}^{(i)}(\mathbf{x}_{\mathcal{I}}^{(i)}) + m_{\mathcal{N}}^{(i)}(\mathbf{x}_{\mathcal{N}}^{(i)})],$$
for constant intercept $\eta_0^{(i)} \in \mathbb{R}$ and link $h_i : \mathbb{R} \to \mathbb{R}$
- Interpretable: $m_{\mathcal{I}}^{(i)}$, e.g., linear, spline. Neural network: $m_{\mathcal{N}}^{(i)}$.
- Our framework applies for **any generic parametric distribution** $\mathcal{F}$, e.g., Bernoulli for occurrence, as well as for non-parametric quantile regression.

Proposed model

Let the response follow $\mathcal{F}(\boldsymbol{\theta}(\mathbf{x}))$ with parameter set $\boldsymbol{\theta}(\mathbf{x}) = (\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \dots)$. Then for all $i = 1, 2, \dots,$

- Split predictor set $\mathbf{x}$ into two **complementary** subsets $\mathbf{x}_{\mathcal{I}}^{(i)}$ and $\mathbf{x}_{\mathcal{N}}^{(i)}$ - "interpreted" and "non-interpreted"
- Let
$$\theta_i(\mathbf{x}) = h_i[\eta_0^{(i)} + m_{\mathcal{I}}^{(i)}(\mathbf{x}_{\mathcal{I}}^{(i)}) + m_{\mathcal{N}}^{(i)}(\mathbf{x}_{\mathcal{N}}^{(i)})],$$
for constant intercept $\eta_0^{(i)} \in \mathbb{R}$ and link $h_i : \mathbb{R} \to \mathbb{R}$
- Interpretable: $m_{\mathcal{I}}^{(i)}$, e.g., linear, spline. Neural network: $m_{\mathcal{N}}^{(i)}$.
- Our framework applies for **any generic parametric distribution** $\mathcal{F}$, e.g., Bernoulli for occurrence, as well as for non-parametric quantile regression.

Proposed model

Let the response follow $\mathcal{F}(\boldsymbol{\theta}(\mathbf{x}))$ with parameter set $\boldsymbol{\theta}(\mathbf{x}) = (\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \dots)$. Then for all $i = 1, 2, \dots,$

- Split predictor set $\mathbf{x}$ into two **complementary** subsets $\mathbf{x}_{\mathcal{I}}^{(i)}$ and $\mathbf{x}_{\mathcal{N}}^{(i)}$ - "interpreted" and "non-interpreted"
- Let
$$\theta_i(\mathbf{x}) = h_i[\eta_0^{(i)} + m_{\mathcal{I}}^{(i)}(\mathbf{x}_{\mathcal{I}}^{(i)}) + m_{\mathcal{N}}^{(i)}(\mathbf{x}_{\mathcal{N}}^{(i)})],$$
for constant intercept $\eta_0^{(i)} \in \mathbb{R}$ and link $h_i : \mathbb{R} \to \mathbb{R}$
- Interpretable: $m_{\mathcal{I}}^{(i)}$, e.g., linear, spline. Neural network: $m_{\mathcal{N}}^{(i)}$.
- Our framework applies for **any generic parametric distribution** $\mathcal{F}$, e.g., Bernoulli for occurrence, as well as for non-parametric quantile regression.

Estimating $m_{\mathcal{N}}$

We estimate $m_{\mathcal{N}}$ using a neural network (NN):

- Training of neural networks is implemented in the `R` interface to `Keras` (`R` package `pinnEV` forthcoming)
- Loss is (penalised) **negative log-likelihood** for $\mathcal{F}$
- **Different types of NN** can be used depending on structure in $\mathbf{x}$. We compare densely-connected (vanilla), **CNN**, as well as RNN
- Models with simple NNs **outperform** fully-linear/additive models

We estimate extreme quantiles using a **novel point process model**:

- Has three parameters: location q_α, spread $s_\beta > 0$ and shape $\xi \geq 0$
- All describe the properties of the **corresponding block-maxima dist.**

Estimating $m_{\mathcal{N}}$

We estimate $m_{\mathcal{N}}$ using a neural network (NN):

- Training of neural networks is implemented in the `R` interface to `Keras` (`R` package `pinnEV` forthcoming)
- Loss is (penalised) **negative log-likelihood** for $\mathcal{F}$
- **Different types of NN** can be used depending on structure in $\mathbf{x}$. We compare densely-connected (vanilla), **CNN**, as well as RNN
- Models with simple NNs **outperform** fully-linear/additive models

We estimate extreme quantiles using a **novel point process model**:

- Has three parameters: location q_α, spread $s_\beta > 0$ and shape $\xi \geq 0$
- All describe the properties of the **corresponding block-maxima dist.**

Estimating $m_{\mathcal{N}}$

We estimate $m_{\mathcal{N}}$ using a neural network (NN):

- Training of neural networks is implemented in the `R` interface to `Keras` (`R` package `pinnEV` forthcoming)
- Loss is (penalised) **negative log-likelihood** for $\mathcal{F}$
- **Different types of NN** can be used depending on structure in $\mathbf{x}$. We compare densely-connected (vanilla), **CNN**, as well as RNN
- Models with simple NNs **outperform** fully-linear/additive models

We estimate extreme quantiles using a **novel point process model**:

- Has three parameters: location q_α, spread $s_\beta > 0$ and shape $\xi \geq 0$
- All describe the properties of the **corresponding block-maxima dist.**

Estimating $m_{\mathcal{N}}$

We estimate $m_{\mathcal{N}}$ using a neural network (NN):

- Training of neural networks is implemented in the `R` interface to `Keras` (`R` package `pinnEV` forthcoming)
- Loss is (penalised) **negative log-likelihood** for $\mathcal{F}$
- **Different types of NN** can be used depending on structure in $\mathbf{x}$. We compare densely-connected (vanilla), **CNN**, as well as RNN
- Models with simple NNs **outperform** fully-linear/additive models

We estimate extreme quantiles using a **novel point process model**:

- Has three parameters: location q_α, spread $s_\beta > 0$ and shape $\xi \geq 0$
- All describe the properties of the **corresponding block-maxima dist.**

Estimating $m_{\mathcal{N}}$

We estimate $m_{\mathcal{N}}$ using a neural network (NN):

- Training of neural networks is implemented in the `R` interface to `Keras` (`R` package `pinnEV` forthcoming)
- Loss is (penalised) **negative log-likelihood** for $\mathcal{F}$
- **Different types of NN** can be used depending on structure in $\mathbf{x}$. We compare densely-connected (vanilla), **CNN**, as well as RNN
- Models with simple NNs **outperform** fully-linear/additive models

We estimate extreme quantiles using a **novel point process model**:

- Has three parameters: location q_α, spread $s_\beta > 0$ and shape $\xi \geq 0$
- All describe the properties of the **corresponding block-maxima dist.**

Estimating $m_{\mathcal{N}}$

We estimate $m_{\mathcal{N}}$ using a neural network (NN):

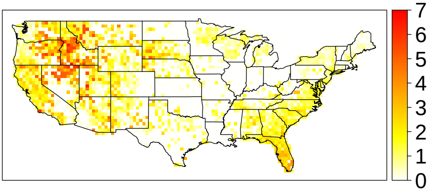
- Training of neural networks is implemented in the `R` interface to `Keras` (`R` package `pinnEV` forthcoming)
- Loss is (penalised) **negative log-likelihood** for $\mathcal{F}$
- **Different types of NN** can be used depending on structure in $\mathbf{x}$. We compare densely-connected (vanilla), **CNN**, as well as RNN
- Models with simple NNs **outperform** fully-linear/additive models

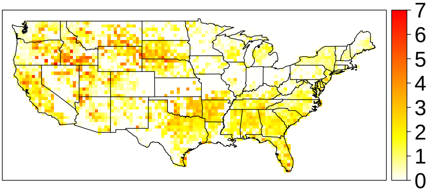
We estimate extreme quantiles using a **novel point process model**:

- Has three parameters: location q_α, spread $s_\beta > 0$ and shape $\xi \geq 0$
- All describe the properties of the **corresponding block-maxima dist.**

Application: Data

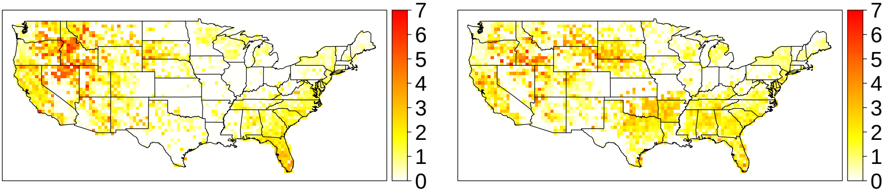
- Monthly burnt area (BA) for the contiguous U.S.
- **Fire Program Analysis** fire-occurrence database
- 1993-2015, March - September. 161 total fields
- $0.5^\circ \times 0.5^\circ$ spatial resolution. 3503 locations, 216713 non-zero values
- Maps of $\log(1+\sqrt{\text{BA}})$. Left: July 2007. Right: July 2012. California wildfires.





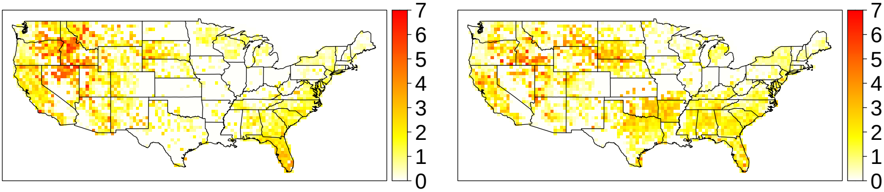
Application: Data

- Monthly burnt area (BA) for the contiguous U.S.
- **Fire Program Analysis** fire-occurrence database
- 1993-2015, March - September. 161 total fields
- $0.5^\circ \times 0.5^\circ$ spatial resolution. 3503 locations, 216713 non-zero values
- Maps of $\log(1+\sqrt{\text{BA}})$. Left: July 2007. Right: July 2012. California wildfires.



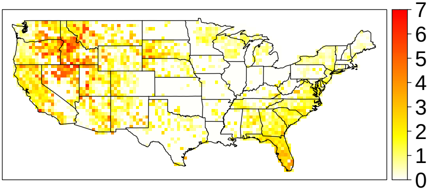
Application: Data

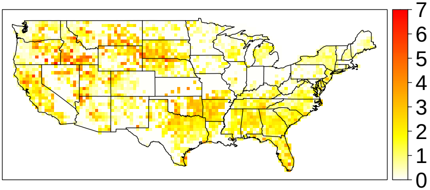
- Monthly burnt area (BA) for the contiguous U.S.
- **Fire Program Analysis** fire-occurrence database
- 1993-2015, March - September. 161 total fields
- $0.5^\circ \times 0.5^\circ$ spatial resolution. 3503 locations, 216713 non-zero values
- Maps of $\log(1+\sqrt{\text{BA}})$. Left: July 2007. Right: July 2012. California wildfires.



Application: Data

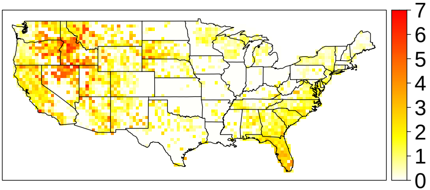
- Monthly burnt area (BA) for the contiguous U.S.
- **Fire Program Analysis** fire-occurrence database
- 1993-2015, March - September. 161 total fields
- $0.5^\circ \times 0.5^\circ$ spatial resolution. 3503 locations, 216713 non-zero values
- Maps of $\log(1+\sqrt{BA})$. Left: July 2007. Right: July 2012. California wildfires.

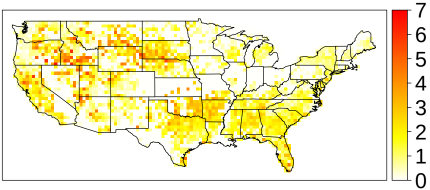




Application: Data

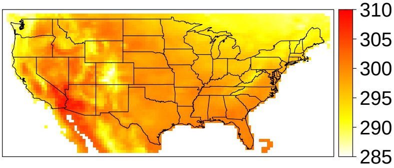
- Monthly burnt area (BA) for the contiguous U.S.
- **Fire Program Analysis** fire-occurrence database
- 1993-2015, March - September. 161 total fields
- $0.5^\circ \times 0.5^\circ$ spatial resolution. 3503 locations, 216713 non-zero values
- Maps of $\log(1+\sqrt{BA})$. Left: July 2007. Right: July 2012. California wildfires.

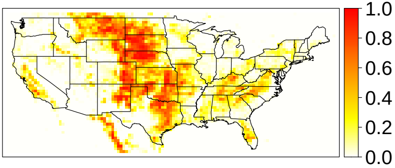




Predictors

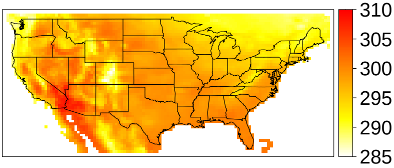
- 10 meteorological variables from **ERA-5 reanalysis on land surface**, e.g., **temperature**, wind-speed components, precipitation
- Land cover maps (**COPERNICUS**) with proportion of grid-cell consisting of one of 18 types, e.g., water, urban areas, **grassland**
- Mean and s.d. altitude
- Left: temp. Right: grassland proportion. July 2007.

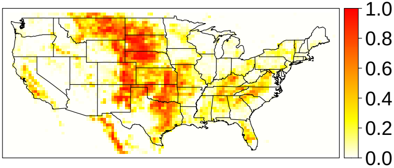




Predictors

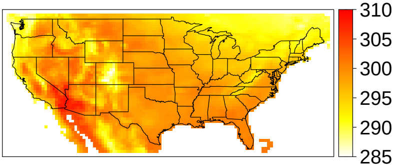
- 10 meteorological variables from **ERA-5 reanalysis on land surface**, e.g., **temperature**, wind-speed components, precipitation
- Land cover maps (**COPERNICUS**) with proportion of grid-cell consisting of one of 18 types, e.g., water, urban areas, **grassland**
- Mean and s.d. altitude
- Left: temp. Right: grassland proportion. July 2007.

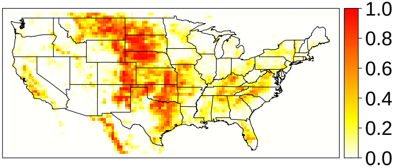




Predictors

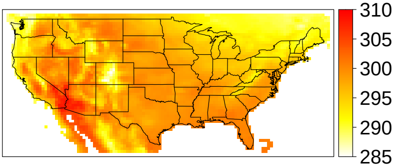
- 10 meteorological variables from **ERA-5 reanalysis on land surface**, e.g., **temperature**, wind-speed components, precipitation
- Land cover maps (**COPERNICUS**) with proportion of grid-cell consisting of one of 18 types, e.g., water, urban areas, **grassland**
- Mean and s.d. altitude
- Left: temp. Right: grassland proportion. July 2007.

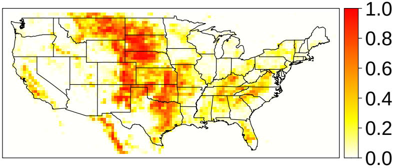




Predictors

- 10 meteorological variables from **ERA-5 reanalysis on land surface**, e.g., **temperature**, wind-speed components, precipitation
- Land cover maps (**COPERNICUS**) with proportion of grid-cell consisting of one of 18 types, e.g., water, urban areas, **grassland**
- Mean and s.d. altitude
- Left: temp. Right: grassland proportion. July 2007.





Model

We model the **occurrence** (not presented) and **spread** of wildfire, separately.

- For the spread, we model the **square-root of strictly positive** BA, i.e., $\sqrt{\text{BA}}|(\text{BA} > 0)$
- **Shape fixed** over space and time - $\hat{\xi} = 0.359\ (0.342, 0.372)$
- Location q_α and spread s_β modelled using **PINN framework**
- **Seven interpreted predictors** - Some linear, some additive - Different for either parameter - Other 23 predictors feed a CNN
- Model uncertainty addressed through **stationary bootstrap** - Results presented as average over 250 samples
- Over-fitting avoided using **validation** techniques

Model

We model the **occurrence** (not presented) and **spread** of wildfire, separately.

- For the spread, we model the **square-root of strictly positive** BA, i.e., $\sqrt{\text{BA}}|(\text{BA} > 0)$
- **Shape fixed** over space and time - $\hat{\xi} = 0.359\ (0.342, 0.372)$
- Location q_α and spread s_β modelled using **PINN framework**
- **Seven interpreted predictors** - Some linear, some additive - Different for either parameter - Other 23 predictors feed a CNN
- Model uncertainty addressed through **stationary bootstrap** - Results presented as average over 250 samples
- Over-fitting avoided using **validation** techniques

Model

We model the **occurrence** (not presented) and **spread** of wildfire, separately.

- For the spread, we model the **square-root of strictly positive** BA, i.e., $\sqrt{\text{BA}}|(\text{BA} > 0)$
- **Shape fixed** over space and time - $\hat{\xi} = 0.359\ (0.342, 0.372)$
- Location q_α and spread s_β modelled using **PINN framework**
- **Seven interpreted predictors** - Some linear, some additive - Different for either parameter - Other 23 predictors feed a CNN
- Model uncertainty addressed through **stationary bootstrap** - Results presented as average over 250 samples
- Over-fitting avoided using **validation** techniques

Model

We model the **occurrence** (not presented) and **spread** of wildfire, separately.

- For the spread, we model the **square-root of strictly positive** BA, i.e., $\sqrt{\text{BA}}|(\text{BA} > 0)$
- **Shape fixed** over space and time - $\hat{\xi} = 0.359\ (0.342, 0.372)$
- Location q_α and spread s_β modelled using **PINN framework**
- **Seven interpreted predictors** - Some linear, some additive - Different for either parameter - Other 23 predictors feed a CNN
- Model uncertainty addressed through **stationary bootstrap** - Results presented as average over 250 samples
- Over-fitting avoided using **validation** techniques

Model

We model the **occurrence** (not presented) and **spread** of wildfire, separately.

- For the spread, we model the **square-root of strictly positive** BA, i.e., $\sqrt{\text{BA}}|(\text{BA} > 0)$
- **Shape fixed** over space and time - $\hat{\xi} = 0.359\ (0.342, 0.372)$
- Location q_α and spread s_β modelled using **PINN framework**
- **Seven interpreted predictors** - Some linear, some additive - Different for either parameter - Other 23 predictors feed a CNN
- Model uncertainty addressed through **stationary bootstrap** - Results presented as average over 250 samples
- Over-fitting avoided using **validation** techniques

Model

We model the **occurrence** (not presented) and **spread** of wildfire, separately.

- For the spread, we model the **square-root of strictly positive** BA, i.e., $\sqrt{\text{BA}}|(\text{BA} > 0)$
- **Shape fixed** over space and time - $\hat{\xi} = 0.359\ (0.342, 0.372)$
- Location q_α and spread s_β modelled using **PINN framework**
- **Seven interpreted predictors** - Some linear, some additive - Different for either parameter - Other 23 predictors feed a CNN
- Model uncertainty addressed through **stationary bootstrap** - Results presented as average over 250 samples
- Over-fitting avoided using **validation** techniques

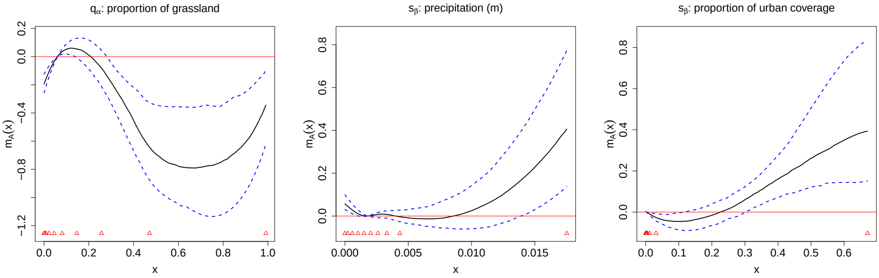
Drivers of extreme wildfire spread

Consider effect on **location** q_α, the **median of the annual maxima dist.** for $\sqrt{\text{BA}}|(\text{BA} > 0)$, i.e., extreme wildfire magnitude

- Linear regression coefficients (given a one s.d. increase):
  - **temperature:** $0.97\ (0.93, 1.37)$,
  - **evaporation:** $0.93\ (0.91, 1.06)$,
  - precipitation: $-0.01\ (-0.03, 0.08)$,
  - proportion of urban coverage: $-0.01(-0.05, 0.05)$
- Effect of **wind-speed** modelled using splines and found to be negligible at this temporal scale

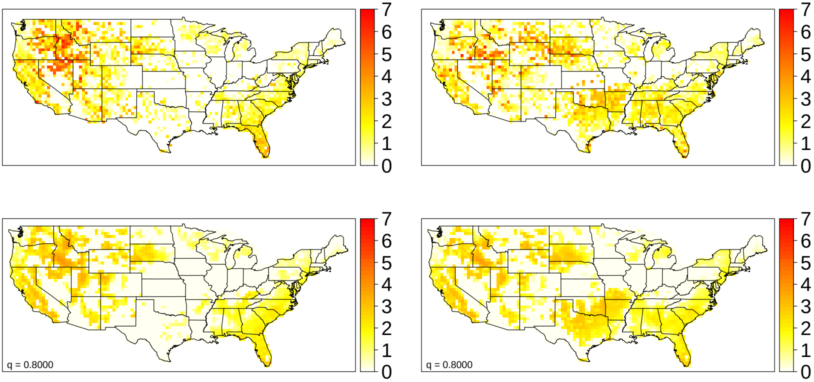
Spline results

Here s_β is the **IQR of the annual maxima dist.** of $\sqrt{BA}|(BA > 0)$. Red triangles are knots, blue dashed lines are 95% confidence envelopes



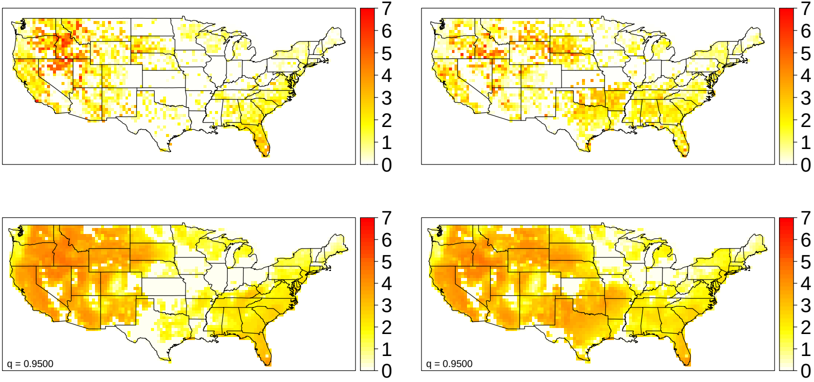
Extreme quantile maps: compound risk

Top: obs. Bottom: estimated q-quantile for $\log(1+\sqrt{\text{BA}})$. Left: July 2007. Right: July 2012.



Extreme quantile maps: compound risk

Top: obs. Bottom: estimated 0.95-quantile for $\log(1 + \sqrt{\text{BA}})$. Left: July 2007. Right: July 2012.



Concluding remarks

- We propose a (very) **flexible framework** for **fitting extreme value models** using **deep learning**
- Combines the **high-predictive accuracy** of neural networks with the **interpretability** of linear and additive models
- Model fits very well to wildfire data, **significantly outperforms** (classical) linear or additive regression models and reveals **new insights** into the **drivers** of extreme wildfires
- Parallel project for Med. with Emanuele Bevacqua and Jakob Zscheischler (UFZ)

Concluding remarks

- We propose a (very) **flexible framework** for **fitting extreme value models** using **deep learning**
- Combines the **high-predictive accuracy** of neural networks with the **interpretability** of linear and additive models
- Model fits very well to wildfire data, **significantly outperforms** (classical) linear or additive regression models and reveals **new insights** into the **drivers** of extreme wildfires
- Parallel project for Med. with Emanuele Bevacqua and Jakob Zscheischler (UFZ)

Concluding remarks

- We propose a (very) **flexible framework** for **fitting extreme value models** using **deep learning**
- Combines the **high-predictive accuracy** of neural networks with the **interpretability** of linear and additive models
- Model fits very well to wildfire data, **significantly outperforms** (classical) linear or additive regression models and reveals **new insights** into the **drivers** of extreme wildfires
- Parallel project for Med. with Emanuele Bevacqua and Jakob Zscheischler (UFZ)

Concluding remarks

- We propose a (very) **flexible framework** for **fitting extreme value models** using **deep learning**
- Combines the **high-predictive accuracy** of neural networks with the **interpretability** of linear and additive models
- Model fits very well to wildfire data, **significantly outperforms** (classical) linear or additive regression models and reveals **new insights** into the **drivers** of extreme wildfires
- Parallel project for Med. with Emanuele Bevacqua and Jakob Zscheischler (UFZ)

Selected references

Richards, J. (2022).
pinnEV: Partially-Interpretable Neural Networks for modelling of Extreme Values.
R package. Will be made available at github.com/Jbrich95/pinnEV.

Richards, J. and Huser, R. (2022).
A unifying partially-interpretable framework for neural network-based extreme quantile regression.
Pre-print. Not available online.

Zhong, Q. and Wang, J.-L. (2021).
Neural networks for partially linear quantile regression.
arXiv preprint arXiv:2106.06225.

Both will be available alongside slides at my website `jbrich95.github.io` (via QR code).

SCAN ME

Thanks for your attention!