Partially-interpretable neural networks for extreme quantile regression

With application to Mediterranean Europe wildfires

Jordan Richards[1] Raphaël Huser[1]
Emanuele Bevacqua[2] Jakob Zscheischler[2]

[1]King Abdullah University of Science and Technology (KAUST)

[2]Helmholtz Centre for Environmental Research (UFZ)

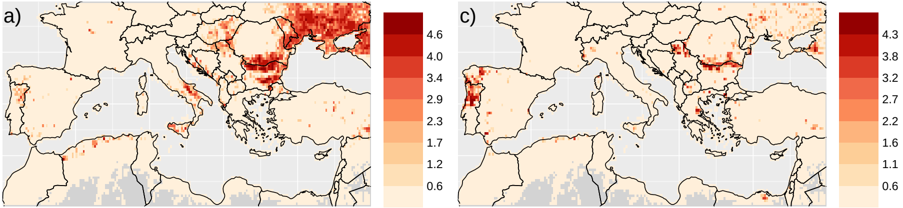
Motivation

- Recent years have seen devastating wildfires in Europe—100s of deaths and millions of km^2 of destroyed land and agriculture
- In 2021, global wildfires contributed to $\approx$ 1760 Mt of carbon emissions
- To mitigate risk, need to identity **drivers** and **high-risk** areas—High quantiles of burnt area are natural risk measures

Extreme quantile regression

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

- Typical quantiles of interest will be larger than previously observed $\Rightarrow$ non-parametric quantile regression likely to perform poorly
- Instead turn to parametric regression using **asymptotically-justified extreme-value (EV) distributions** (e.g., GEV, GPD, Point Process approach) with parameter set $\boldsymbol{\theta}$



Maps of $\log(1 + \text{BA})$: August, 2001 (left) and October, 2017 (right).

Existing approaches

Existing approaches for **parametric extreme quantile regression** represent $\boldsymbol{\theta}$ as linear or additive 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 $d = 38$ predictors

We instead use **deep learning based on neural networks (NNs)** as these methods can

- capture complex structures (e.g., interactions) in $\mathbf{x}$,
- scale well to high dimensions,
- facilitate high predictive accuracy.

Deep learning for extremes

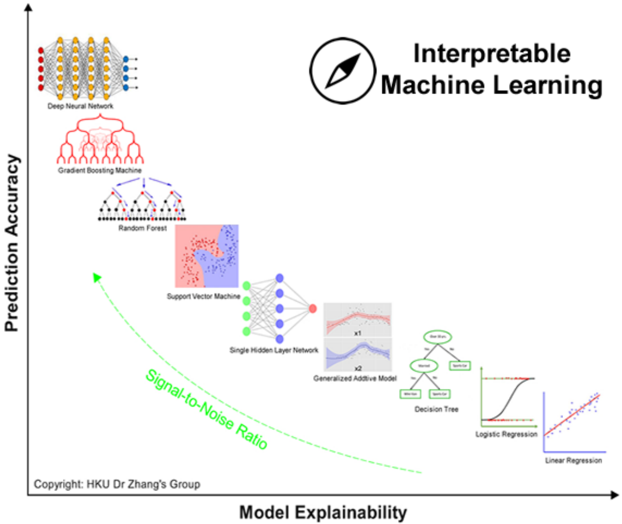
Statisticians generally avoid the use of neural networks.

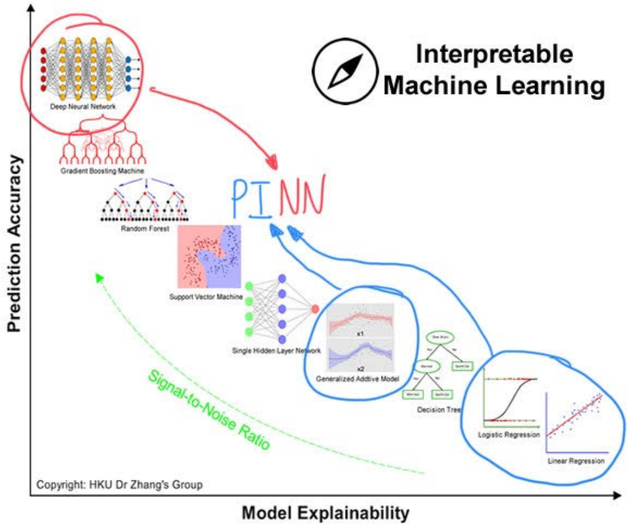
- NNs are a "black box" in the sense that it is difficult/impossible to interpret their output—no good for understanding the drivers of risk
- We extend the partially-linear quantile regression NN of [Zhong and Wang, 2021] and create NNs that are **"partially-interpretable"** (PINNs)
- The effects of some predictors on response are modelled using readily-interpretable functions, while the rest feed a NN

Partially interpretable neural network (PINN) framework

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 $i = 1, 2, \dots,$

- Split predictor set $\mathbf{x}$ into two **complementary** subsets $\mathbf{x}_{\mathcal{I}}^{(i)}$ ("interpretable"), and $\mathbf{x}_{\mathcal{N}}^{(i)}$ ("non-interpretable")
- Let
$$\theta_i(\mathbf{x}) = h_i[m_{\mathcal{I}}^{(i)}\{\mathbf{x}_{\mathcal{I}}^{(i)}\} + m_{\mathcal{N}}^{(i)}\{\mathbf{x}_{\mathcal{N}}^{(i)}\}],$$
for link $h_i : \mathbb{R} \to \mathbb{R}$
- Interpretable: further split into linear/additive terms as $m_{\mathcal{L}}^{(i)}\{\mathbf{x}_{\mathcal{L}}^{(i)}\} + m_{\mathcal{A}}^{(i)}\{\mathbf{x}_{\mathcal{A}}^{(i)}\}$
- Non-interpretable: feed a neural network $m_{\mathcal{N}}^{(i)}$
- All functions estimated simultaneously by minimizing neg. log-likelihood for $\mathcal{F}$, by exploiting variants of stoch. gradient descent using the `R` interface to `Keras`/`Tensorflow`


Interpretable
Machine Learning
Prediction Accuracy
Model Explainability
Deep Neural Network
Gradient Boosting Machine
Random Forest
Support Vector Machine
Single Hidden Layer Network
Generalized Additive Model
Decision Tree
Logistic Regression
Linear Regression
Signal-to-Noise Ratio
Copyright: HKU Dr Zhang's Group



GAM, GLM or NN on the boundary of the parameter space.

Specification of $m_{\mathcal{N}}$

We estimate $m_{\mathcal{N}}$ using a NN:

- There are **no fundamental restrictions** on the complexity of the architecture (size, depth, type, etc.) of this NN$\Rightarrow$We consider complexity ranging from hundreds to tens-of-thousands of parameters
- Different types of NN can be used depending on structure in $\mathbf{x}$. We use densely-connected (standard) and **CNNs**
- Both fully-linear and fully-additive models are often **outperformed** by even the simplest NN

Model comparison—Parameter functional form

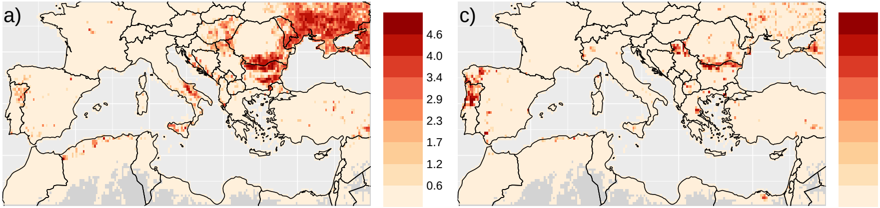
From [Richards and Huser, 2022], who model U.S. wildfire spread:

Table 1: Comparison of $\theta_i(s,t)$ forms. Metrics are averaged over five folds. Values of the loss, AIC and twCRPS are given as the absolute difference to the lowest across all models.

$\theta_i(s,t)$	Number of parameters	Training loss	Validation loss	Training AIC	In/Out-sample sMAD ($\times 10^{-2}$)	twCRPS
fully-linear	43	7754	1661	14389	15.8/16.2	253.8
fully-GAM	803	5810	1214	12020	14.2/14.8	203.1
fully-NN	603	0	0	0	6.01/7.41	0
lin+GAM	689	6119	1282	12411	15.3/15.8	211.6
lin+NN	477	2055	428	3859	7.74/9.05	74.0
GAM+NN	743	1776	365	3834	8.37/9.44	64.8
lin+GAM+NN	629	1851	394	3754	7.55/8.98	63.6

Data application

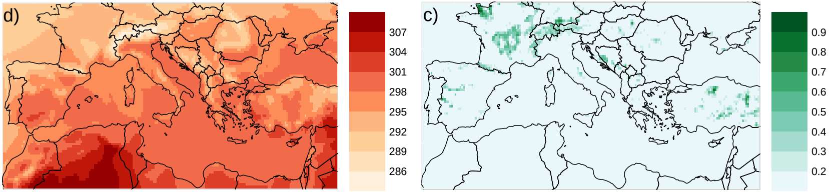
- Monthly burnt area (BA) for **Mediterranean European** wildfires
- FireCCI database, generated by MODIS data
- 2001-2020, June–November
- $\approx$ 10000 locations, $\approx$ 1.2M feasible locations, 102240 non-zero values



Maps of $\log(1 + \text{BA})$: August, 2001 (left) and October, 2017 (right).

Predictors

- Land cover maps (**COPERNICUS**) with proportion of grid cells consisting of one of 21 types, e.g., water, urban areas, **grassland**
- 13 meteorological variables from **ERA-5 reanalysis on single levels**, e.g., **temperature**, wind-speed components, surface pressure
- Mean, and s.d., of altitude and long/lat coordinates



Temperature (left) and grassland proportion (right) for August, 2001.

Model

We model the **occurrence** and extreme **spread** of wildfire separately.

- We use a **logistic** regression model for occurrence probability $p_0(\mathbf{x})$
- For extreme wildfire spread, we model

$$\{BA - u(\mathbf{x})\} | \{BA > u(\mathbf{x}), \mathbf{x}\} \sim \text{GPD}^*(\sigma(\mathbf{x}), \xi > 0; u(\mathbf{x}))$$

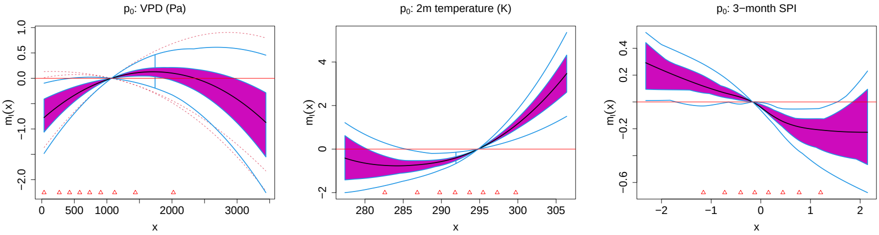
  where $u(\mathbf{x}) > 0$ is a high-threshold and (severity) scale $\sigma(\mathbf{x}) > 0$

- $\text{GPD}^*(\sigma, \xi; u) = \text{GPD}(\sigma + \xi u, \xi)$ is parameterised so that scale σ is **independent of** $u(\mathbf{x})$
- **Shape fixed over space and time with** $\hat{\xi} = 0.322\ (0.280, 0.353)$

Model

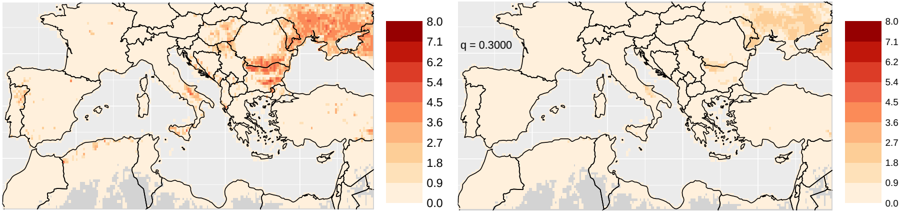
- We **interpret** the effect of vapour pressure deficit (VPD), 2m air temperature and a drought index, three-month SPI, on p_0 and σ, using splines
- $m_{\mathcal{N}}$ is a 5-layered CNN for p_0 ($\approx 14,000$ pars.) and a 4-layered densely-connected network for σ ($720 + 1$ pars.)
- Training/testing/validation used to **reduce over-fitting** and perform model/architecture comparison
- Parameter uncertainty assessed using a **stationary bootstrap**—Results presented as average over 250 samples

Drivers of wildfire occurrence



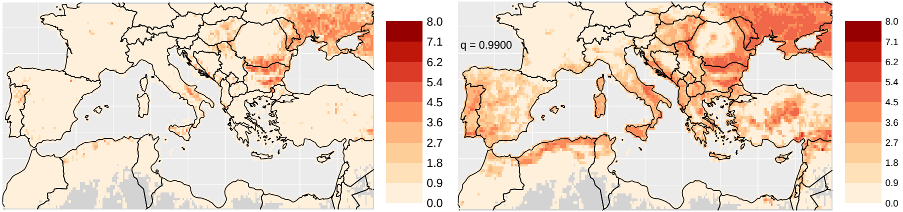
Functional box-plots of estimated $m_{\mathcal{I}}(x)$ for p_0.

Risk assessment



Observed $\log(1+\text{BA})$ (left) and estimated extreme q-quantiles (right) for August 2001.

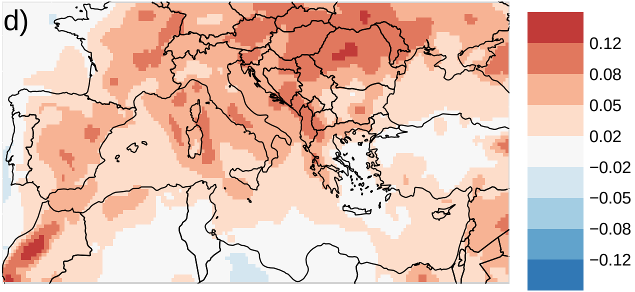
Risk assessment



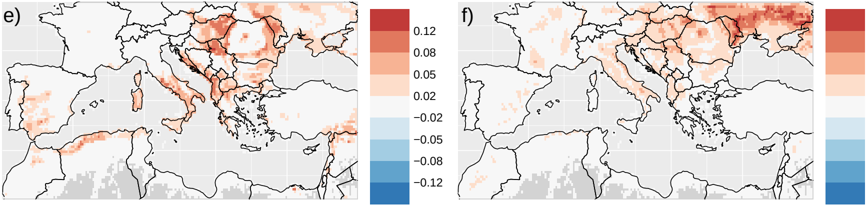
Observed $\log(1 + \mathrm{BA})$ (left) and estimated extreme q-quantiles (right) for August 2001.

Impacts of long-term climate trends

What would the distribution have looked like in August 2001, but with **predicted air temperature values** for 2020? How do the values of p_0 and extreme quantiles change under these conditions?



Estimated trends in August 2m air temperature (K).



Median changes in p_0 (left) and 95% quantile of spread (right).

Summary

- We propose a (very) **flexible framework** for **fitting statistical regression models** that combines the **high-predictive accuracy** of neural networks ("guaranteed" by universal function approximation theorems) with the **interpretability** of linear and additive models
- Model fits very well to wildfire data and reveals **new insights** into the **climatic drivers** of extreme wildfires and **climate change impacts**
- Extreme value and classical statistical models implemented in the R package, *pinnEV*

```
#Define predictor subsets for GPD scale
X=list("X.nn"=X.nn, "X.lin"=X.lin,  "X.add.basis"=X.add.basis)
#Fit the GPD model to Y.train for exceedances above u. Neural network is a three-layered MLP. Model is trained with batch size 50 for 1000 epochs
NN.fit<-GPD.NN.train(Y.train,X, u, type="MLP",   widths=c(10,6,3),
                     n.ep=1000, batch.size=50,init.scale=1, init.xi=0.3)
#Output estimated parameters
out<-GPD.NN.predict(X=X,u=u,model=NN.fit$model)
```

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).
Regression modelling of spatiotemporal extreme U.S. wildfires via partially-interpretable neural networks.
arXiv preprint arXiv:2208.07581.

Richards, J., Huser, R., Bevacqua, E., and Zscheischler, J. (2022).
Insights into the drivers and spatio-temporal trends of extreme mediterranean wildfires with statistical deep-learning.
arXiv preprint arXiv:2212.01796.

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

Thanks for your attention!

SCAN ME

Scan for full details of my research.