Predicting probability of heat waves using a CNN

George Miloshevich, Patrice Abry, Pierre Borgnat and Freddy Bouchet

ENS DE LYON

Context and Goals

- Extreme events are rare but impactful
- Sampling heat waves in future climates costly
- Machine learning in climate/weather models
- Training on imbalanced datasets is hard
- We train a CNN to predict a probability

Heat Wave (HW) definitions

- HW: extreme of time averaged running mean 2 meter temperature anomalies over France:

$$A_T(t) = \frac{1}{T}\int_t^{t+T} \frac{1}{|\mathcal{D}|}\int_D \left(T_{2m} - \mathbb{E}\left(T_{2m}\right)\right) \mathrm{d}\vec{r}\mathrm{d}u \quad (1)$$

Duration: $T = 14$ days

Area D - “France”

Data: 8000 years of Plasim

- GCM which models the atmosphere, soil
- SST repeated yearly, fixed climate
- Resolution: 2.8 by 2.8 degrees, 10 levels

Normalized Skill Score (NSS)

The goal: find committor function $P(Y|X)$

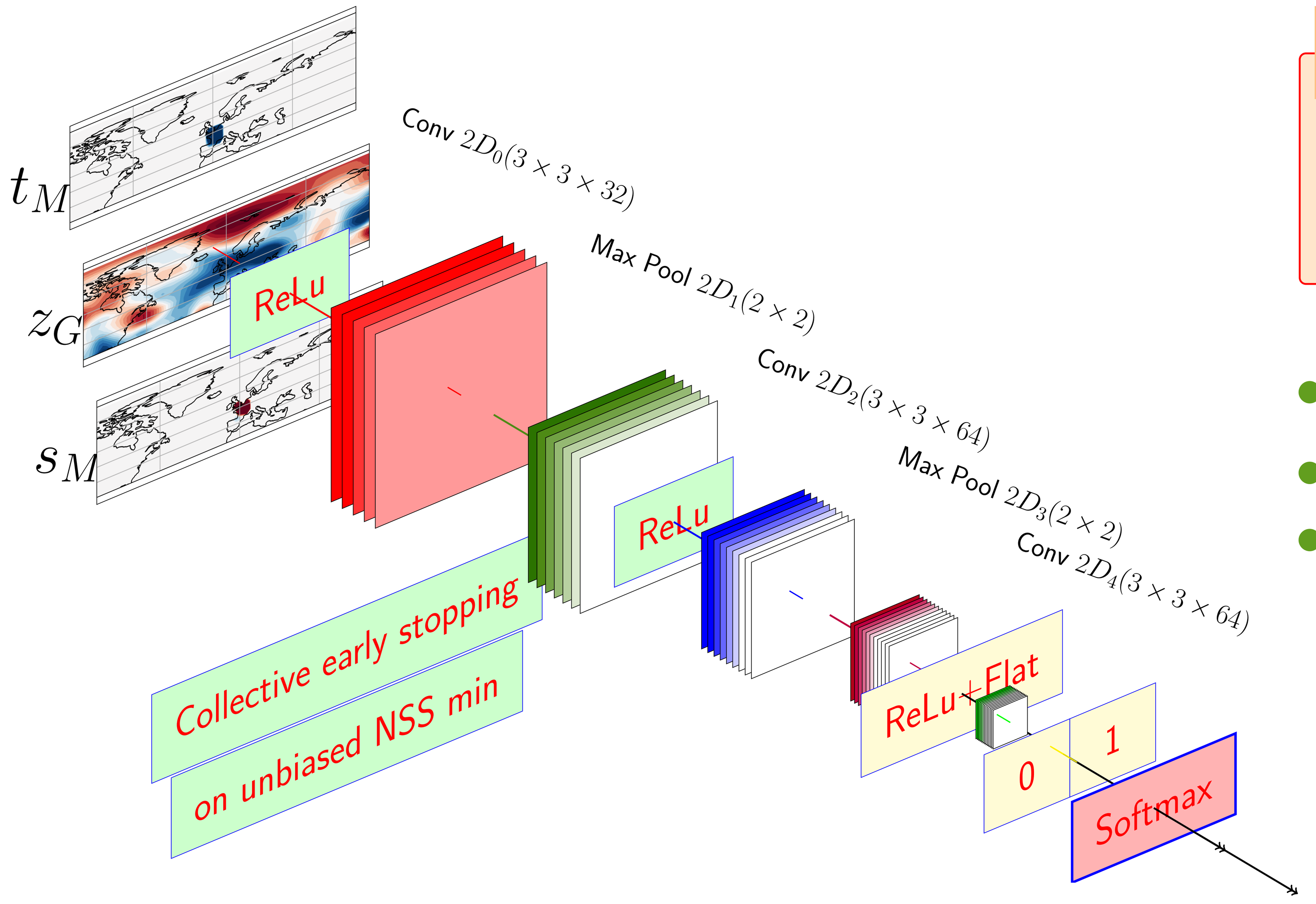
$$\mathbb{P}\left(X = x \text{ and } Y = y\right) = P(Y|X)P(X). \quad (2)$$

Logarithmic score is suitable for rare events

$$-S\left[\hat{p}_Y(X)\right] = -\sum_{k=0}^{1} Y_k \log\left[\hat{p}_k(x)\right] \quad (3)$$

$$\mathrm{NSS} = \frac{-\sum_i \overline{p}_i \log \overline{p}_i - \mathbb{E}\left\{S\left[\hat{p}_Y(X)\right]\right\}}{-\sum_i \overline{p}_i \log \overline{p}_i} \quad (4)$$

Network architecture



Training on different fields

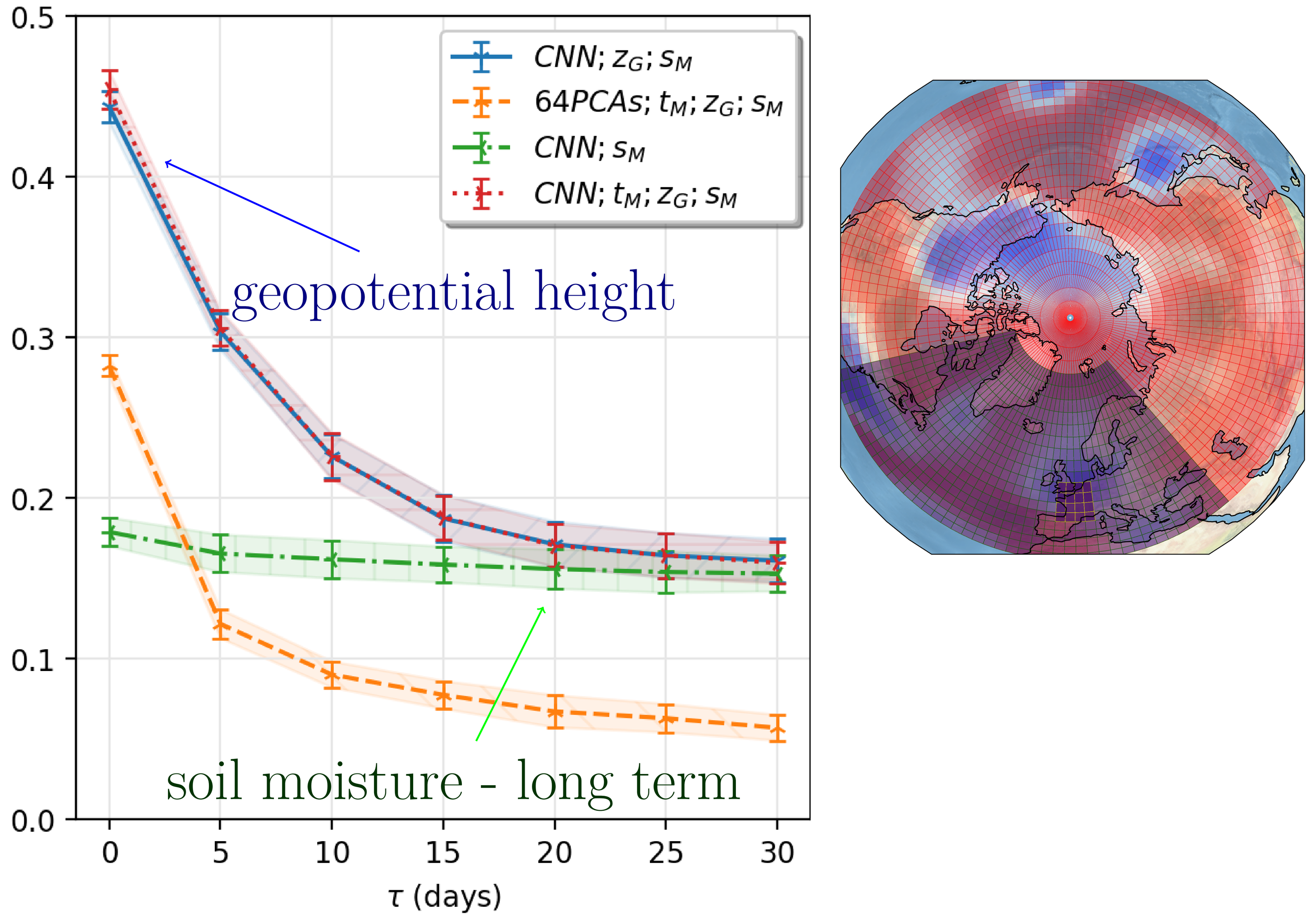


Figure: NSS, 8000 years with 10-fold cross validation vs. lead time τ to the event. CNN outperforms regression on 64 PCA

Probabilistic prediction

Soft-max bounds probabilities to $(0, 1)$ range

$$P\left(Y_n = k \mid \boldsymbol{x}_n, \{w_{k'}\}\right) = \frac{\mathrm{e}^{-x_n^T w_k}}{\sum_{k'=0}^{K-1} \mathrm{e}^{-x_n^T w_{k'}}}, \quad (5)$$

- Y - binary (0: is not HW, 1: is HW):
- HW: above 95 percentile of $A(t)$ $\sim 2.8K$
- $X(\tau)$ - data anomalies at time τ preceding HW:
  - $X_0 = t_M$ - $2m$ temperature anomaly, masked over France
  - $X_1 = z_G$ - $500hPa$ geopotential height anomaly
  - $X_2 = s_M$ - soil moisture anomaly, masked over France

Regional dependence

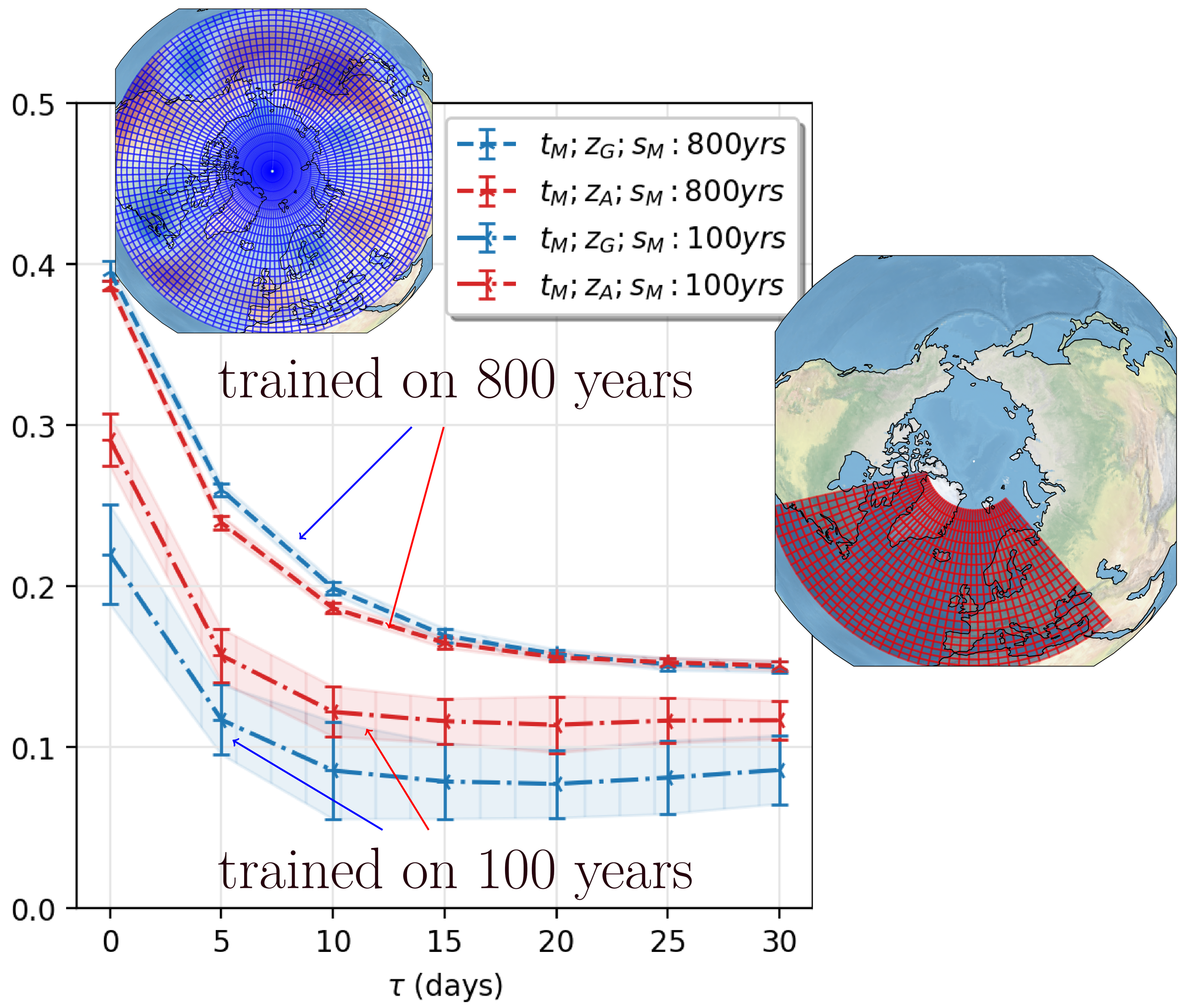


Figure: NSS data reduction vs regional dependence. CNN is trained on North hemisphere or North Atlantic

Key Result

Deep learning the *probabilities* of heat waves requires big data but outperforms traditional methods

Composite maps

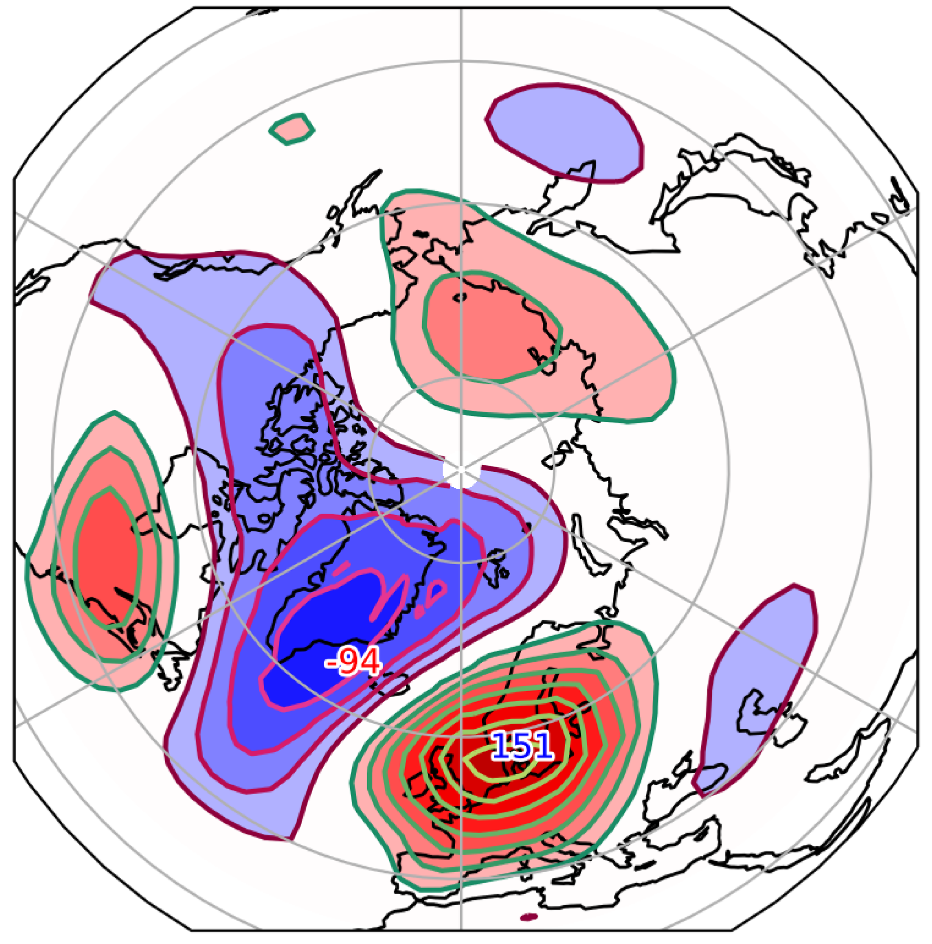


Figure: Composite for $p_{z_G}(\mathrm{HW}|\tau = 0) > 0.67$

Conclusions

- Undersampling: minimal impact on NSS
- The NSS optimal for few CNN layers
- Large dataset needed to learn p_{z_G} well
- p_{s_M} provides long-term prediction skill
- In low data regime masking reduces overfitting
- Extra time or geopotential slices not very useful

What is missing now

- Couple $P(Y|X)$ with the rare event algorithm
- Transfer learning with more complex models
- Dimensional reductions, generative models

Contact Information

- Email: georgemiloshevich(at)ens-lyon.fr