P. GRÖNQUIST, C. YAO, T. BEN-NUN, N. DRYDEN, P. DUEBEN, S. LI, **T. HOEFLER**

Deep Learning for Post-Processing Ensemble Weather Forecasts

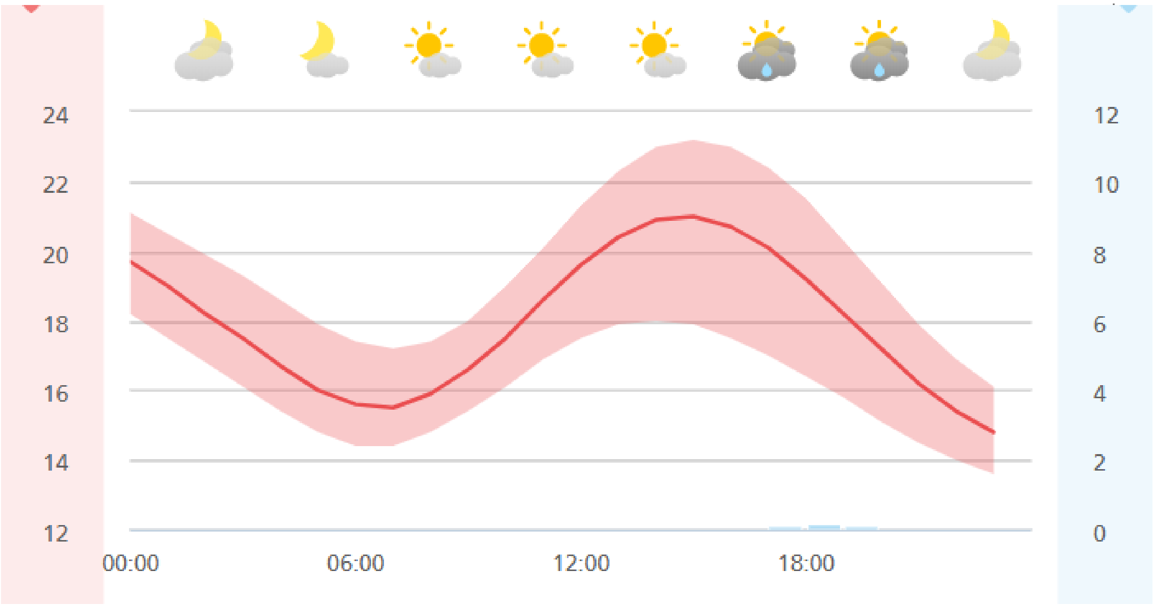
ESIWACE 2020 workshop, virtually anywhere

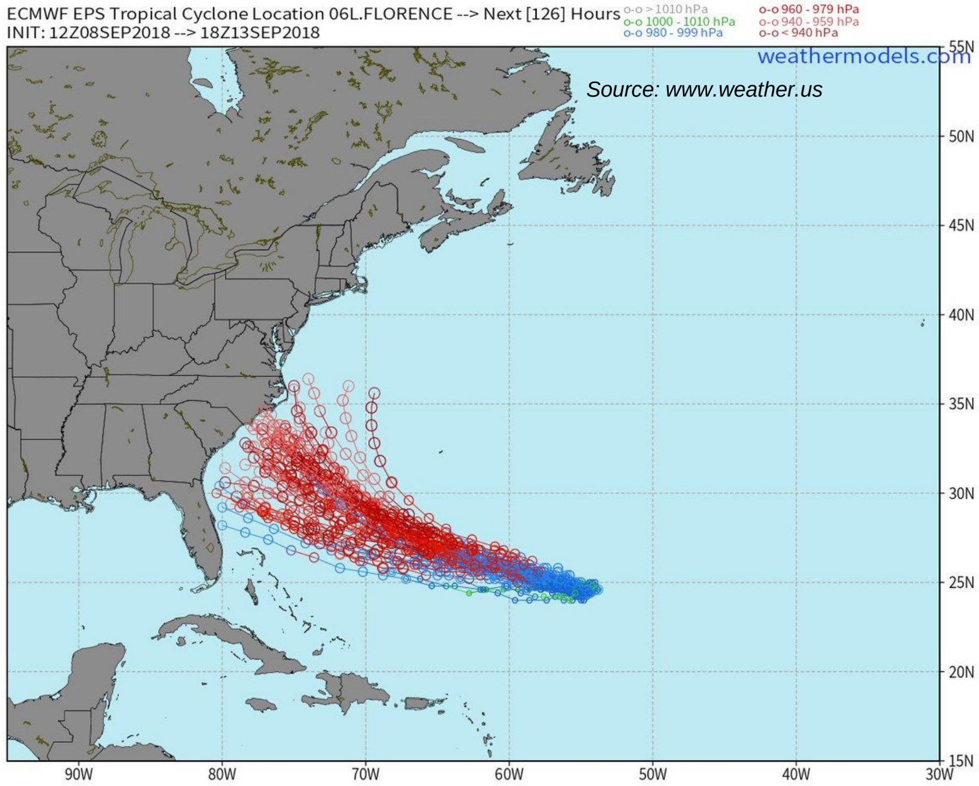




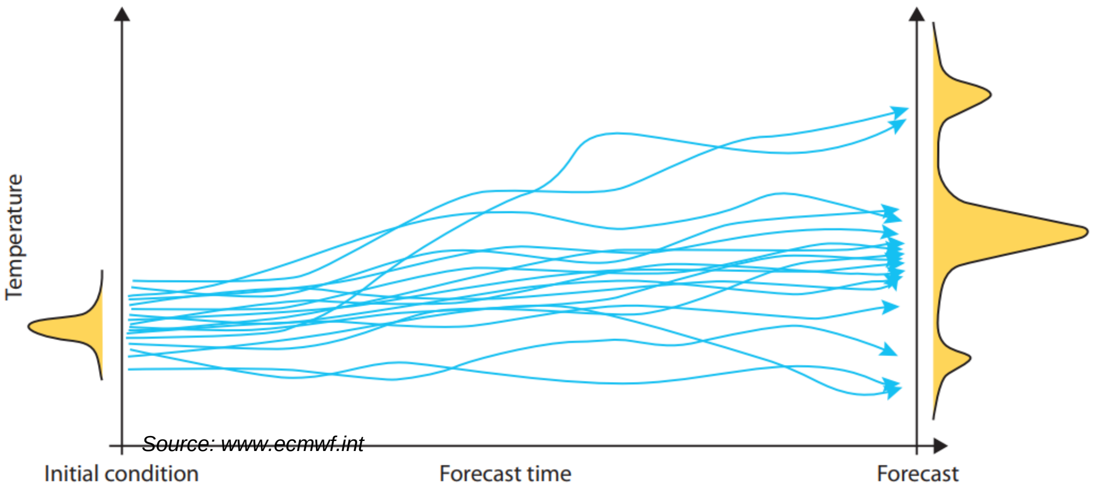
SPCL

Uncertainty in forecasting





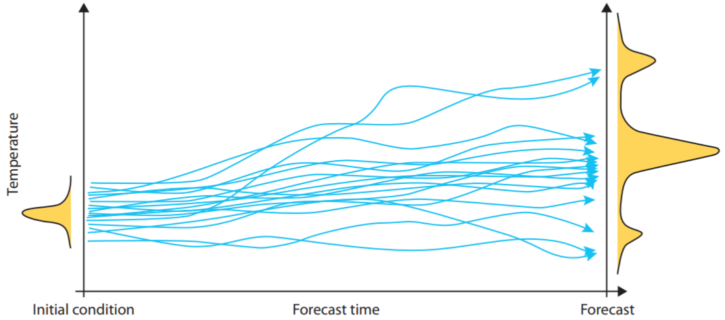
- Weather is a chaotic system
  - Minor perturbations affect the outcome the further into the future we predict
- Solution: Ensemble Prediction Systems – predict weather as a probability distribution
  - Approximated by (stochastic) partial differential equations



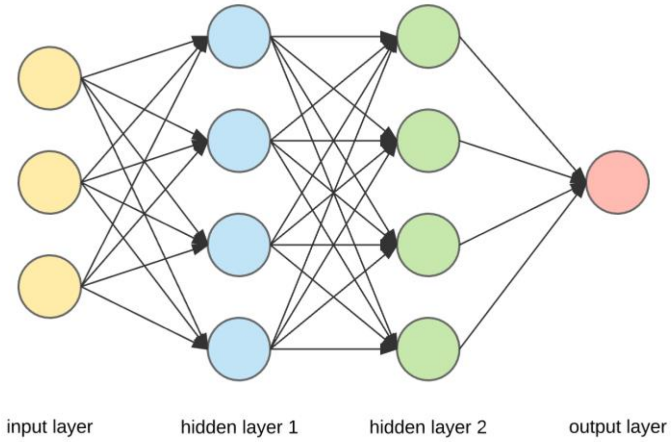
Ensemble Prediction System at ECMWF

- Initial condition uncertainties result from data assimilation

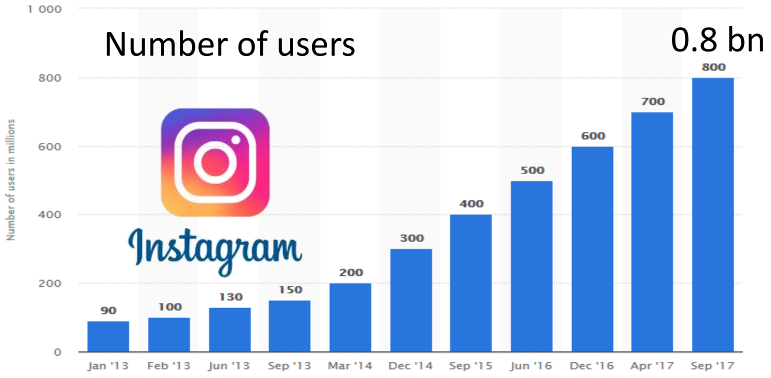
- 51 ensemble members, 1 control (deterministic), 50 perturbed (stochastic)
  - Approximate the highest likely trajectory from output distribution D
  - Lower resolution (9km vs. 18km) in order to fit compute budget
    mostly an economic argument



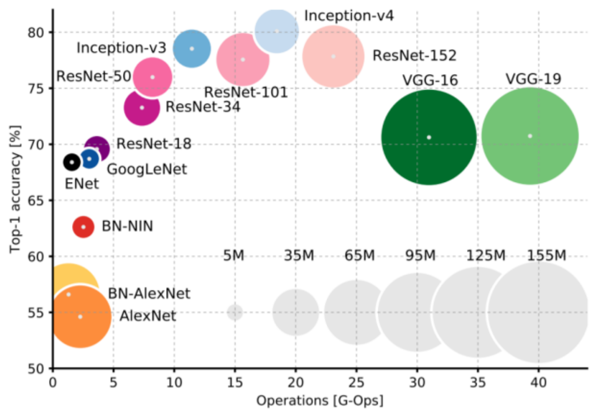
- Next step in the economic argument:
  - Could the number of ensemble members be reduced without sacrificing accuracy?
  - Idea I: predict mean and standard deviation (StdDev) of D from a smaller ensemble
    This may allow us to increase resolution at equal cost – better predictions
  - Can we improve prediction quality by learning from ground truth observations?
  - Idea II: learn (local) model bias from observations
    This may allow us to increase accuracy – better predictions

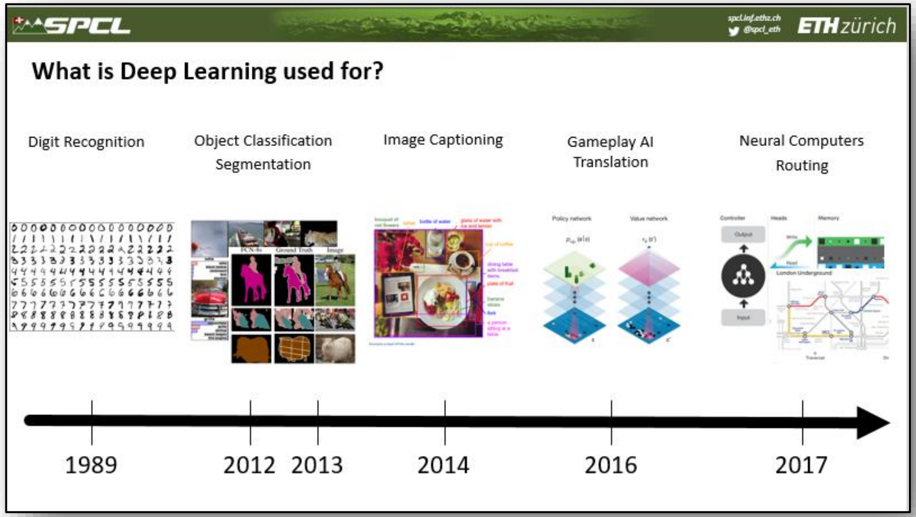


Why machine learning/deep learning?

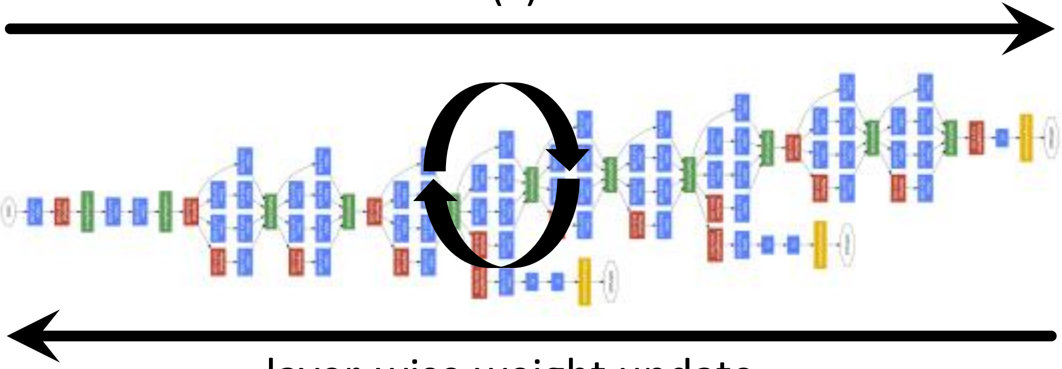


Canziani et al. 2017





f(x)



layer-wise weight update

Cat	0.54	Cat	1.00
Dog	0.28	Dog	0.00
Airplane	0.07	Airplane	0.00
Horse	0.04	Horse	0.00
Bicycle	0.33	Bicycle	0.00
Truck	0.02	Truck	0.00

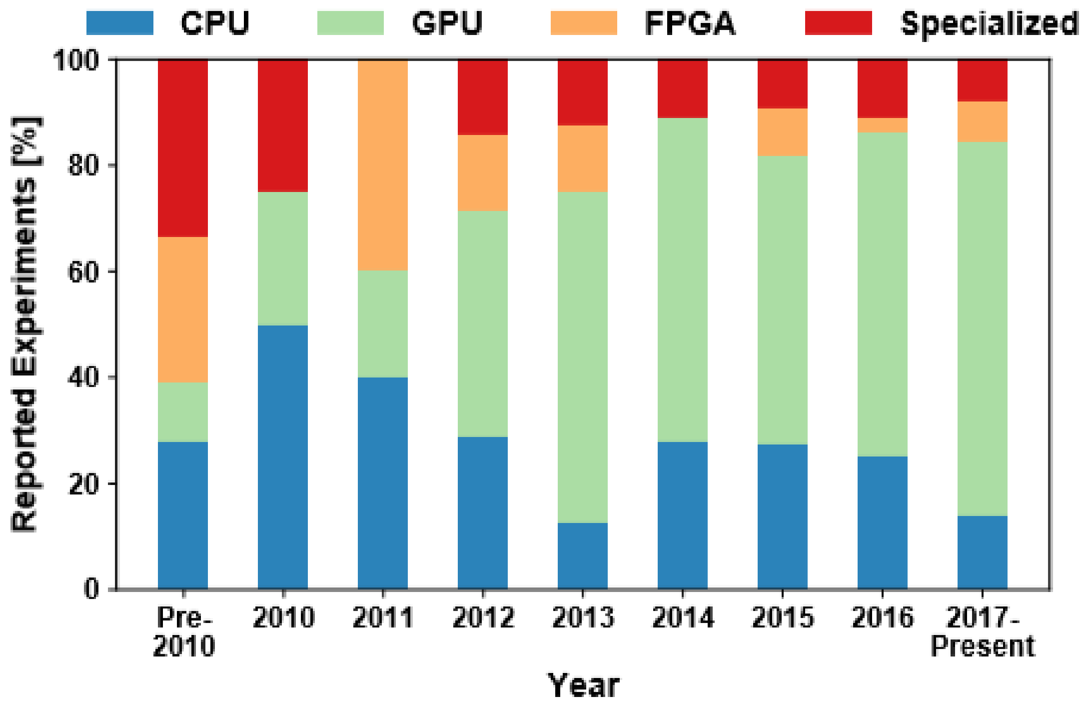
- ImageNet (1k): 180 GB
- ImageNet (22k): A few TB
- Industry: Much larger

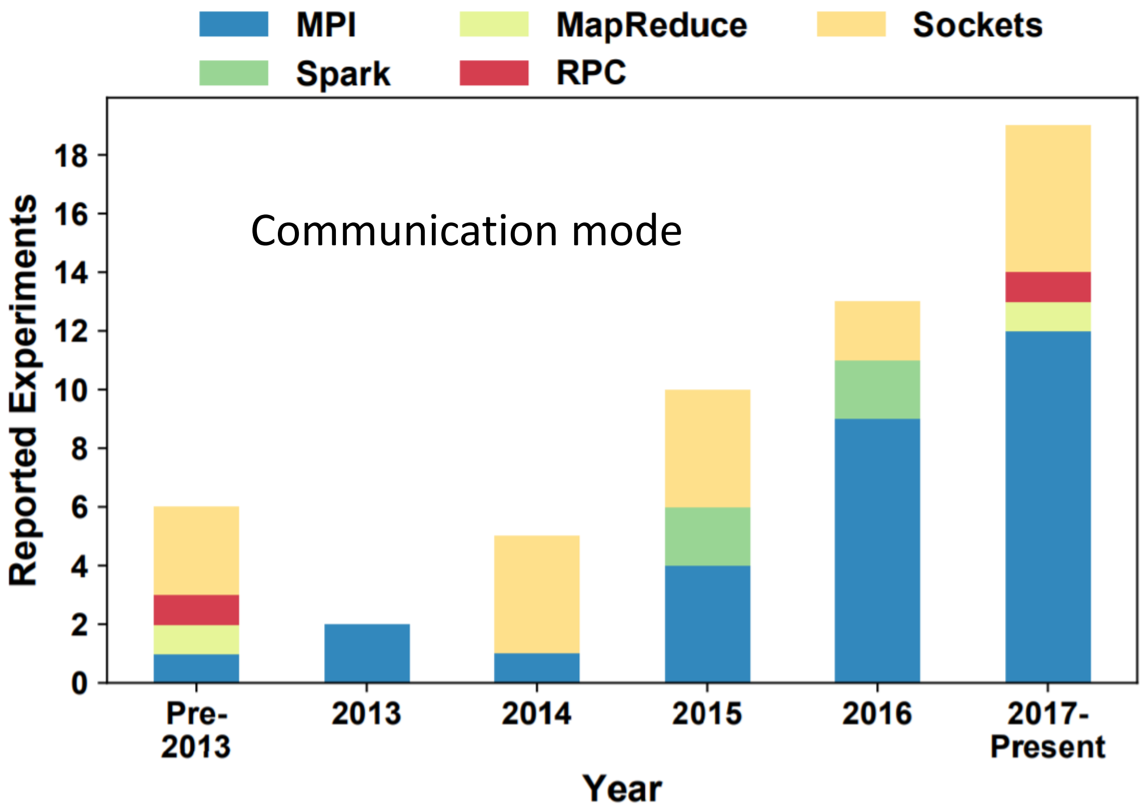
- 100-200 layers deep
- ~100M-2B parameters
- 0.1-8 GiB parameter storage

- 10-22k labels
- growing (e.g., face recognition)
- weeks to train

And everybody is optimizing for it …

The field is moving fast – trying everything imaginable – survey results from 227 papers in the area of parallel deep learning





Deep learning is here to stay – as programming 2.0 or otherwise!

T. Ben-Nun, T. Hoefler: Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis, arXiv Feb 2018

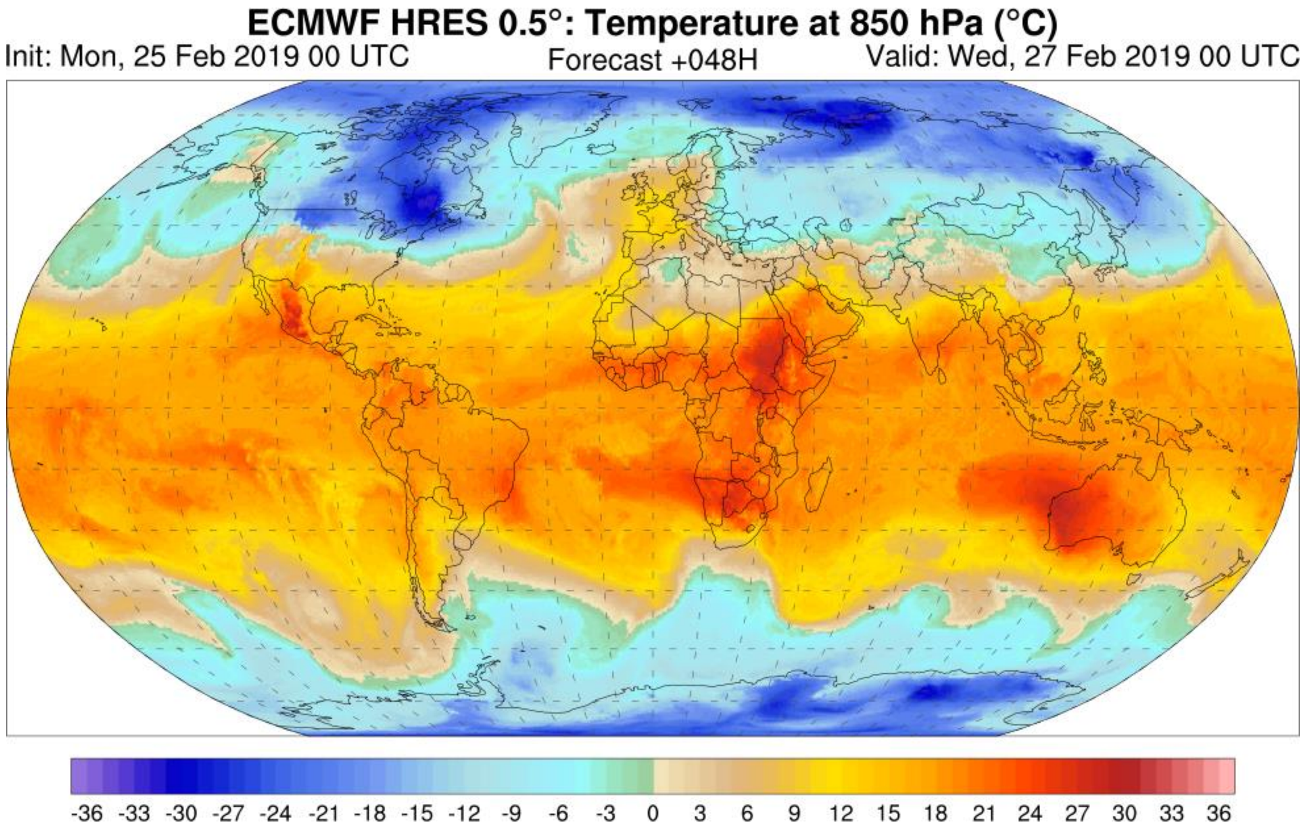
A multi billion dollar (hardware) industry

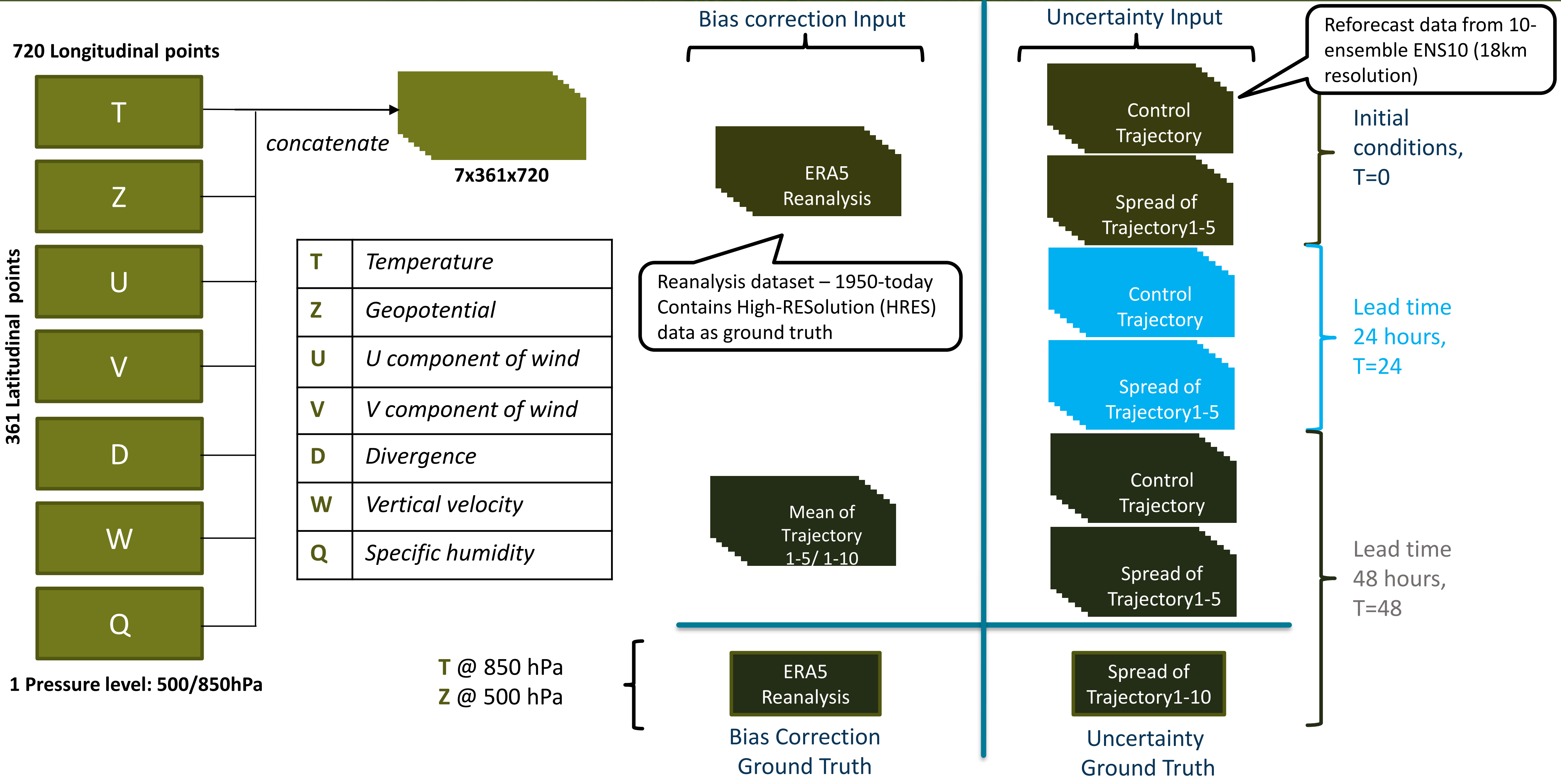
krtkl
Cambricon 寒 武 纪 科 技
Tachyum
adapteva
antmicro
SiFive
Tenstorrent
GRAPHCORE
BITMAIN
cerebras
another brain
thinci
KONIKU
GREENWAVE SYSTEMS
Kneron
KNOWM
LIGHTMATTER
PEZY Computing
LIGHTELLIGENCE
AMD
HUAWEI
Google
NVIDIA
Qualcomm
IBM

intel Nervana
SambaNova SYSTEMS
MYTHIC
TERADEEP
KALRAY
Think Force
NOVUMIND
AMPERE
gyrfalcon technology
Esperanto TECHNOLOGIES
GREENWAVES TECHNOLOGIES
WAVE COMPUTING
ChipIntelli
DEEPVISION
groq
DEEPSCALE
brainchip The Neuromorphic Computing Company

Data Acquisition: Data Selection

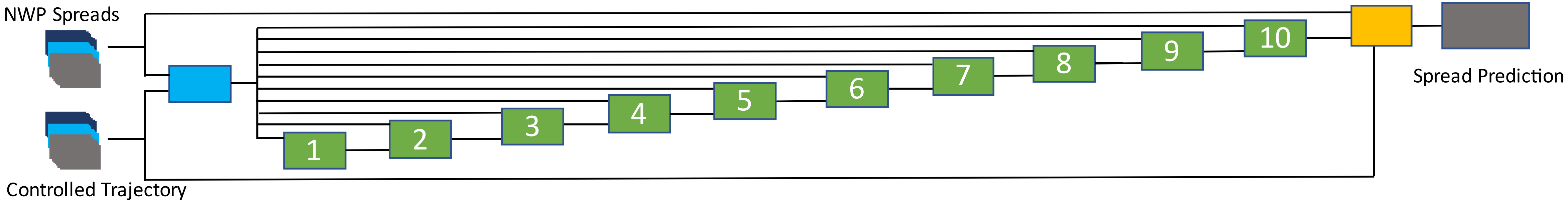
- **Spatial:**
  - 10-member ensembles from ECMWF's hindcasts "ENS10" and "ERA5" reanalysis data – both interpolated on lat/lon grid with 0.5 degree resolution
  - 850 hPa (T850) and 500 hPa (Z500) pressure levels
- **Temporal:**
  - Forecasts available from 0600 and 1800 UTC for each day from 2000-2018
  - Using smallest timestep: 3 hour steps

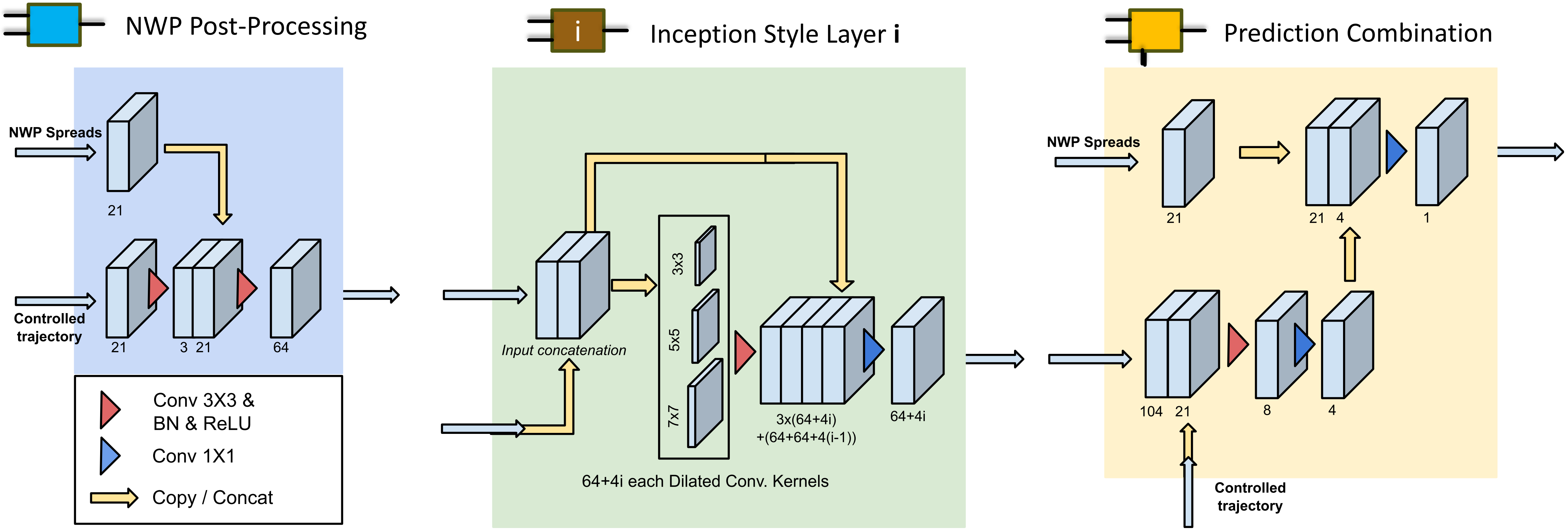




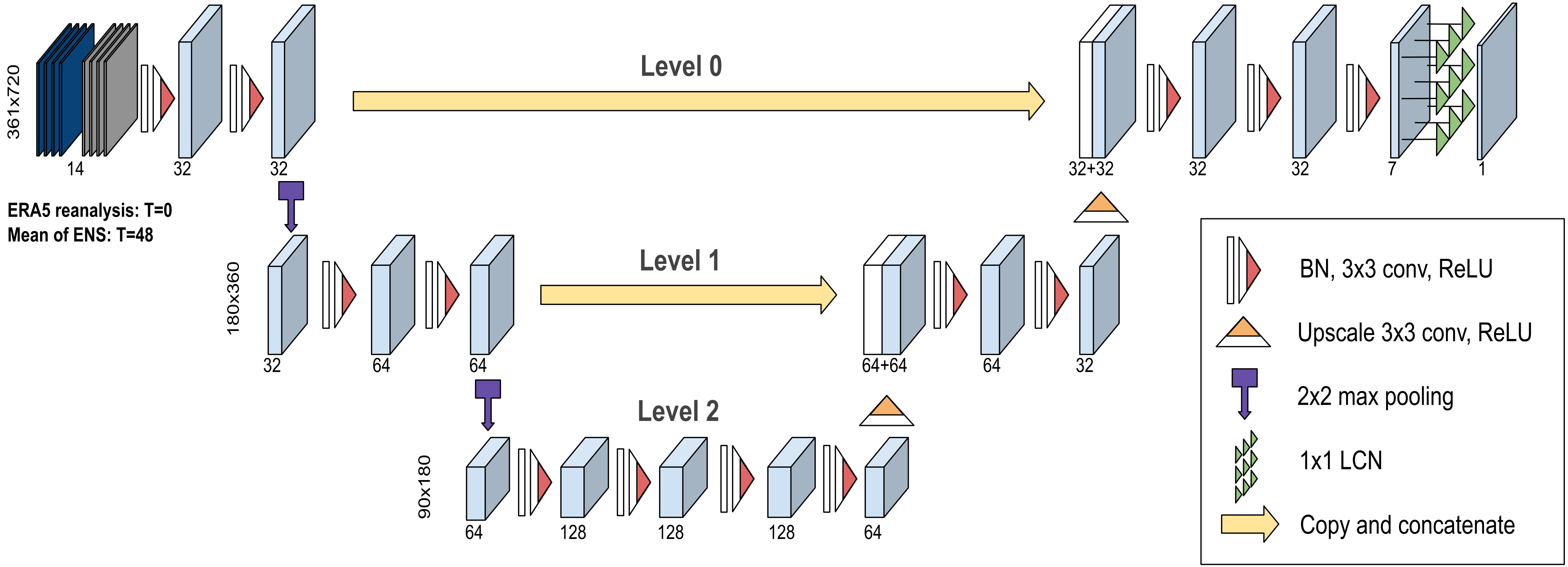
T	*Temperature*
Z	*Geopotential*
U	*U component of wind*
V	*V component of wind*
D	*Divergence*
W	*Vertical velocity*
Q	*Specific humidity*

Uncertainty Quantification Network (based on ResNet)





Bias Correction Network (based on 3D-Unet + LCN)

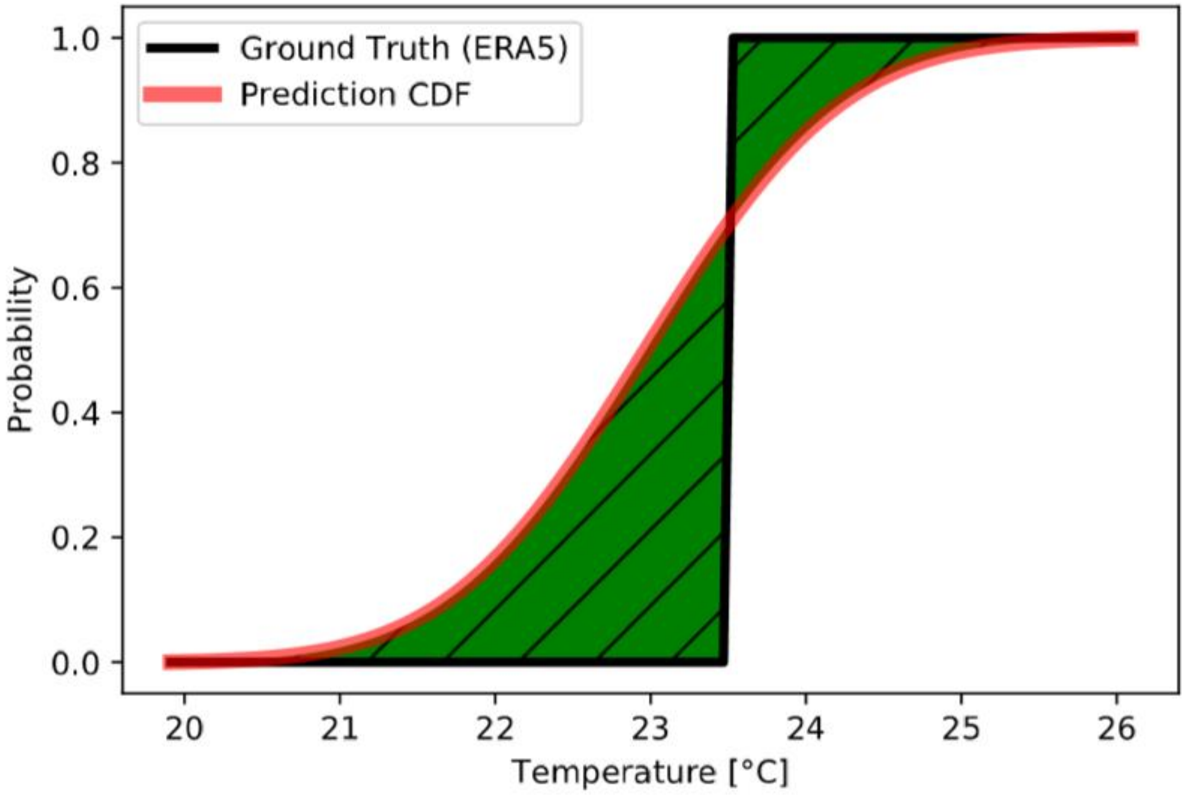


Training: Setup

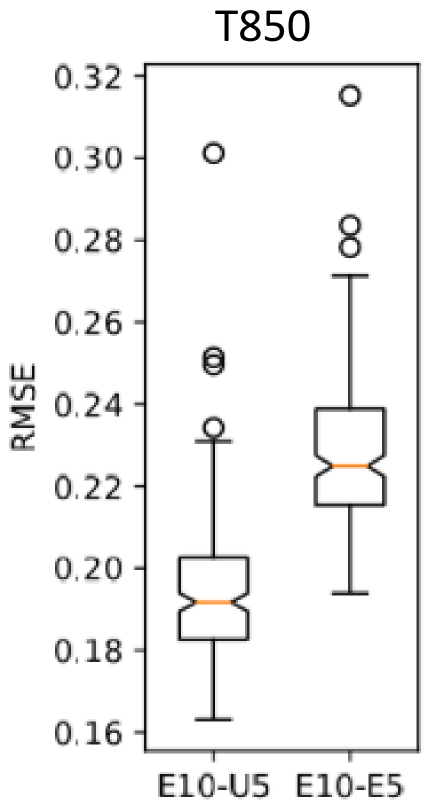
- **Framework**: TensorFlow
  - Default Adam optimizer
  - NVIDIA V100
    Four hours for training
    1/3rd second for inference
  - Batch size 2

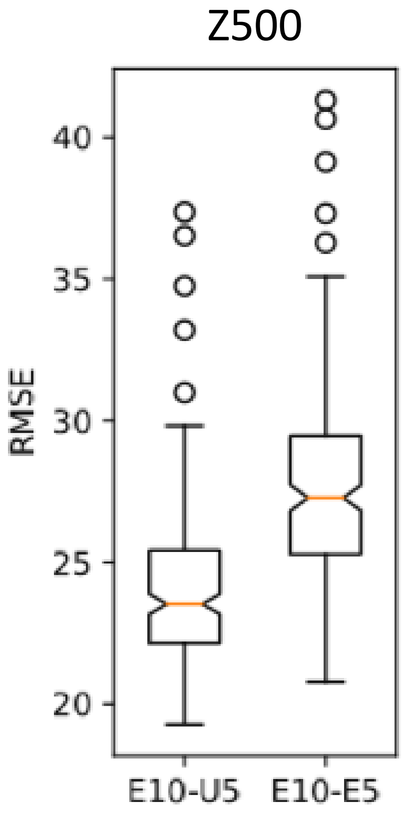
- **Training Loss**: MSE
  - Evaluation on RMSE

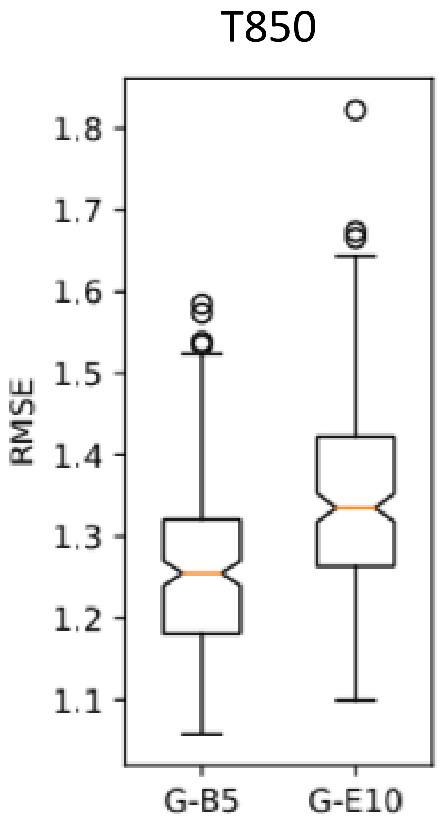
- **Combined training of both models**
  - Loss function $\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} [F(x) - \mathbf{1}_{x>y}]^2 dx$

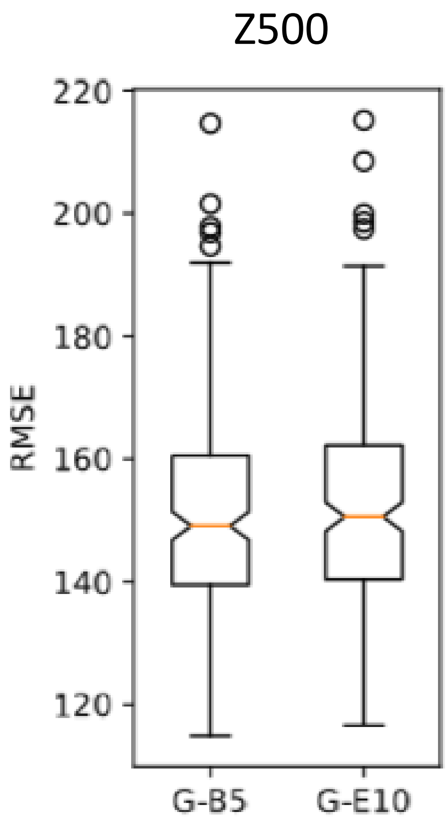


Global RMSE results









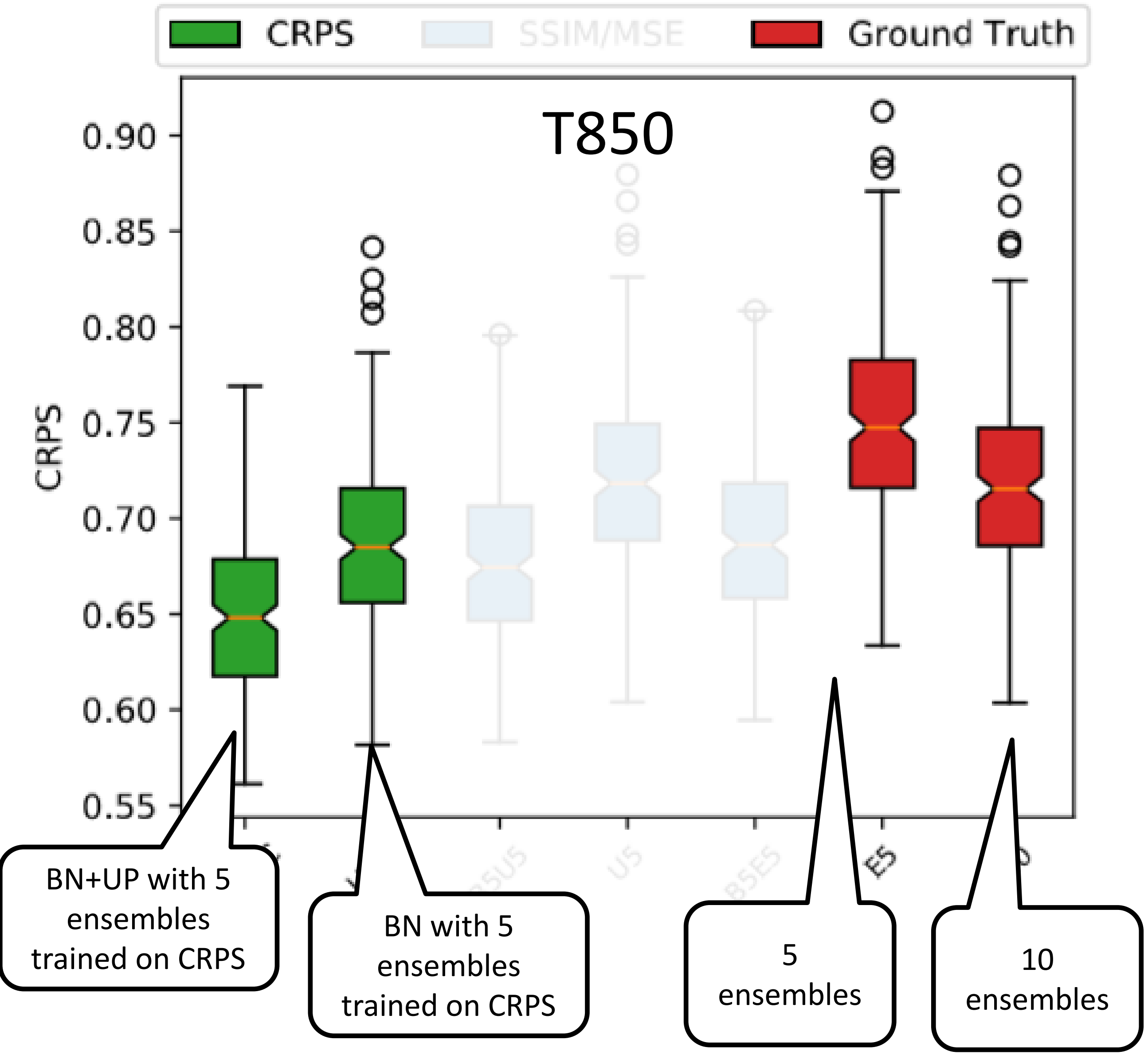
10 ensembles vs. UP with 5 ensembles

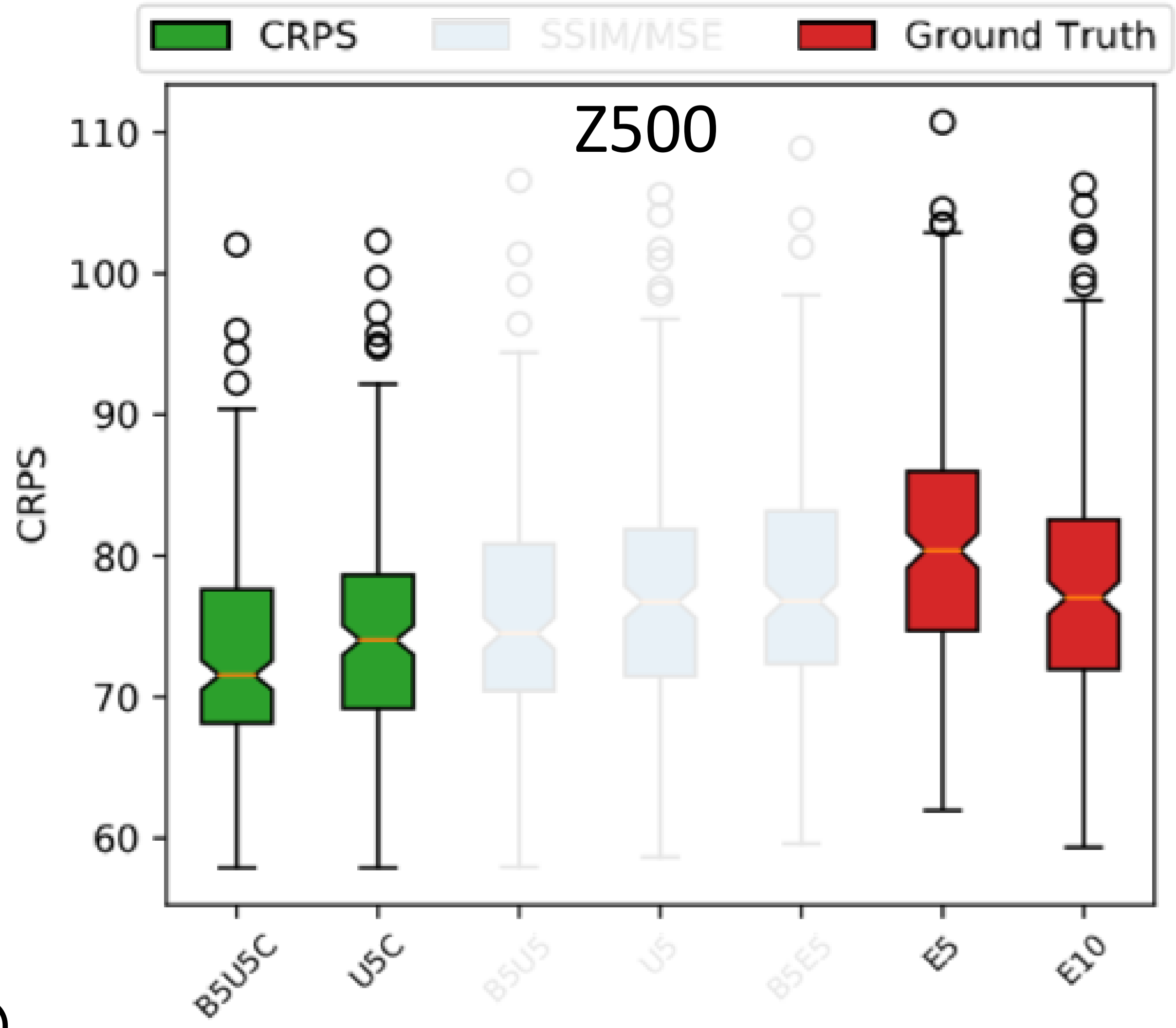
10 ensembles vs. 5 ensembles

ERA5 (ground truth) vs. BN with 5 trajectories

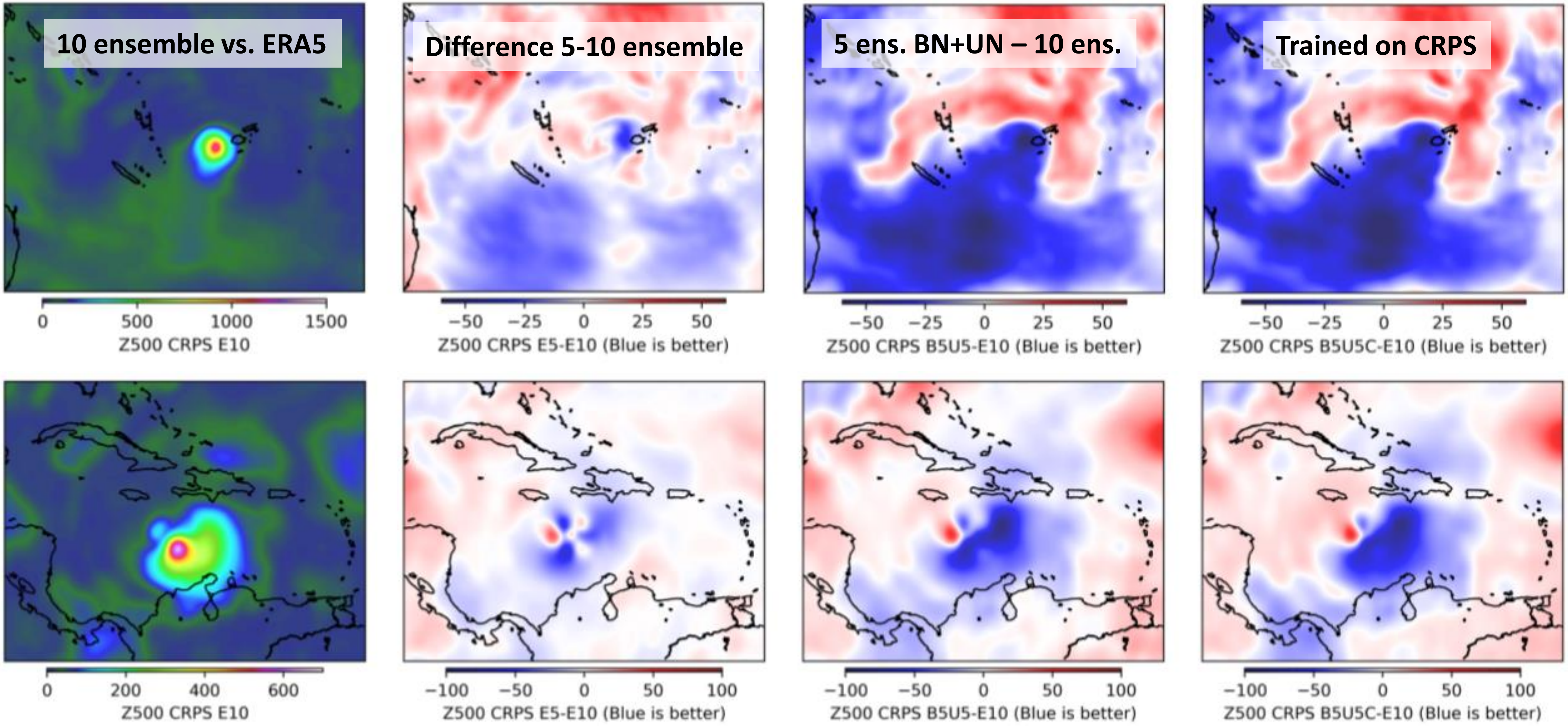
ERA5 (ground truth) vs. 10 trajectories

Global average values for each day 2016-2017)

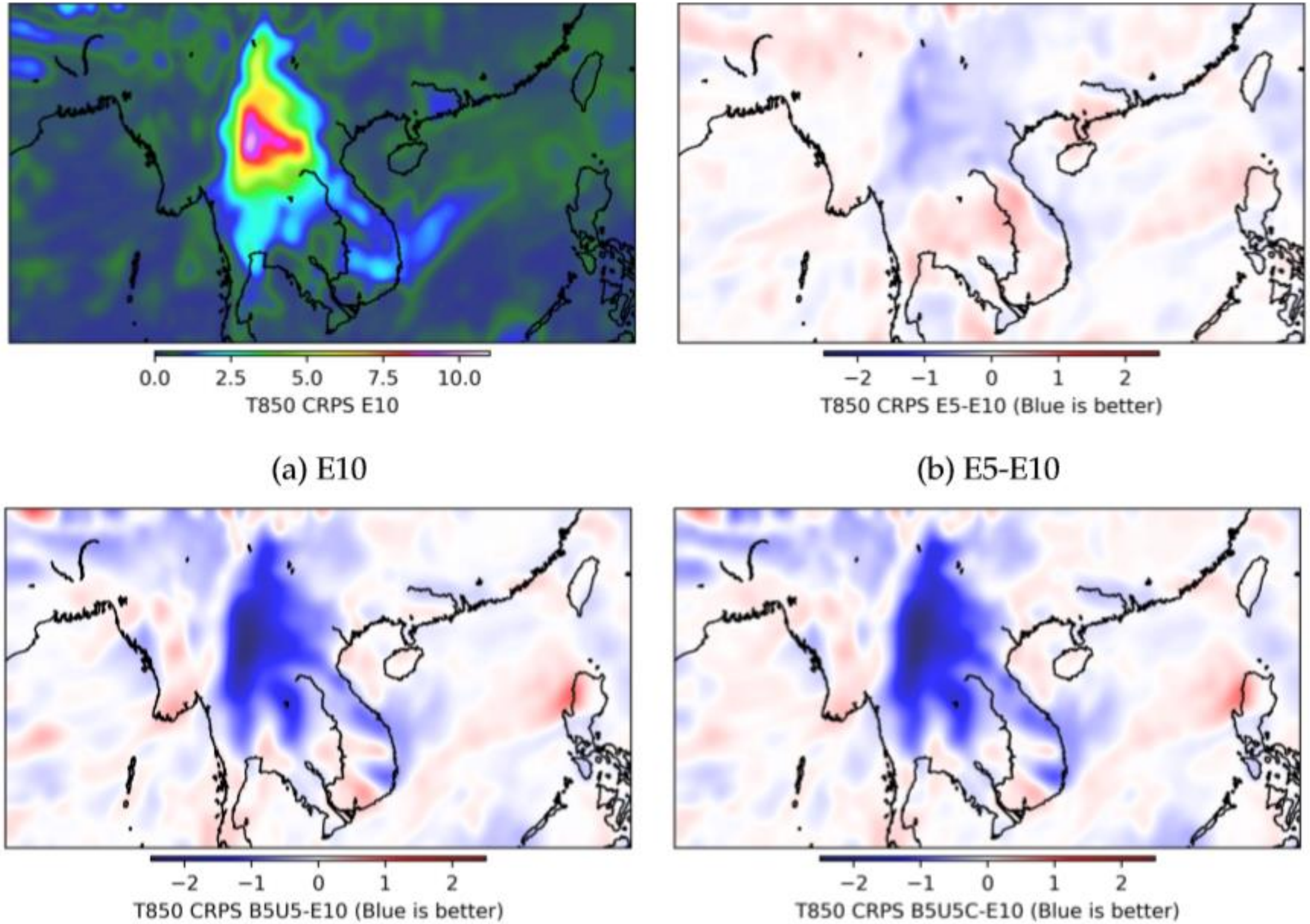




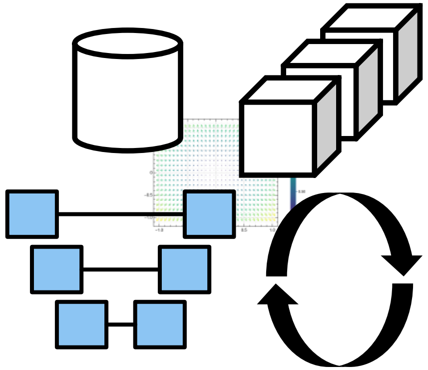
Extreme event: Tropical Cyclone Winston & Hurricane Matthews



Cold wave over Asia



Summary of our preliminary study



- Simple Deep Learning can be used to accelerate forecast pipelines
  - Take advantage of industry efforts to tune hardware and tool-chains
  - An informed approach is **necessary** to ensure improved results

- Using Encoder-Decoder networks for predicting mean and StdDev in ensemble systems yields higher accuracy than using small ensemble statistics
  - Fewer than half of the ensemble members are necessary
  - Accuracy improved with custom operators

- Promising for increasing performance in large-scale settings
  - Needs further investigation!
  - Join us/try yourself: https://github.com/spcl/deep-weather

- Future directions:
  - Larger datasets
  - Custom neural architectures for unstructured grids
  - Integrate into dace tool-chain for further optimization