Fine-tuning probabilistic precipitation forecasts using censored Bernstein Quantile Networks

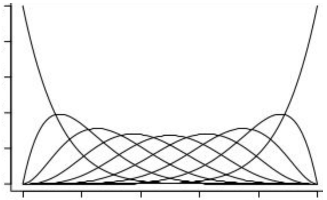
John Bjørnar Bremnes
Norwegian Meteorological Institute

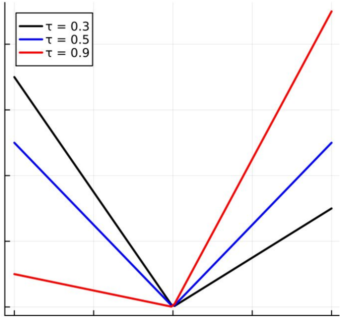
NHR, Machine Learning in Hydrology, Meteorology and related fields, 2025-04-25, Oslo.

Bernstein Quantile Networks

… is a flexible method for distributional regression of continuous variables

- **Bernstein polynomial** as predictive quantile distribution
- **Neural network** to link distribution to input variables
- Estimation by minimising **quantile loss** averaged over predefined quantile levels

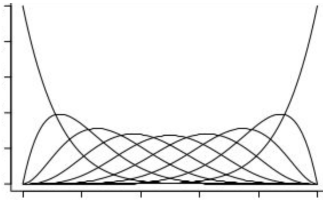


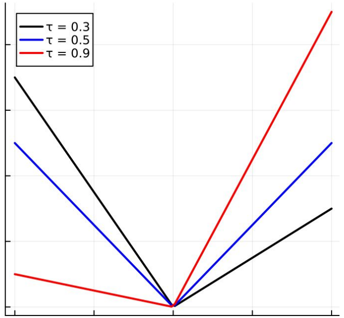


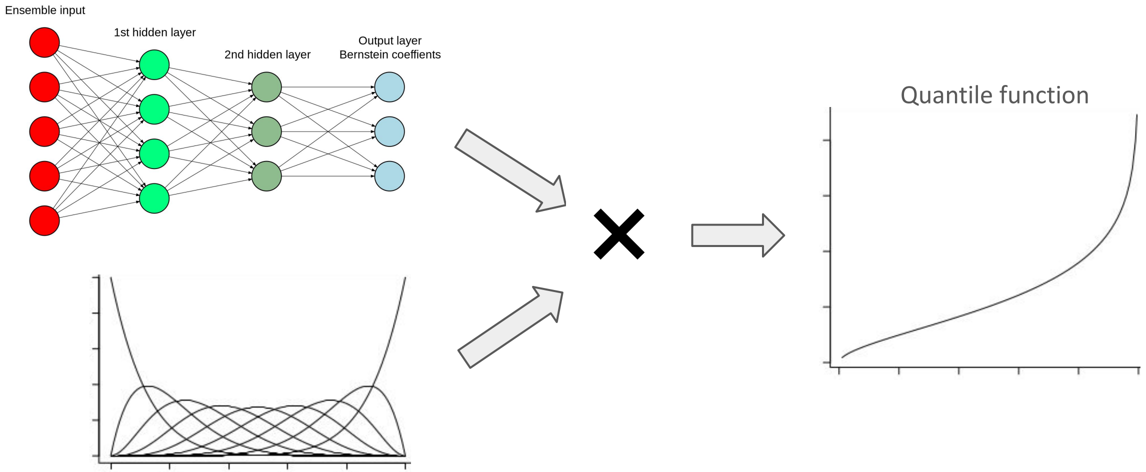
Bernstein Quantile Networks

… is a flexible method for distributional regression of continuous variables

- **Bernstein polynomial** as predictive quantile distribution
- **Neural network** to link distribution to input variables
- Estimation by minimising **quantile loss** averaged over predefined quantile levels



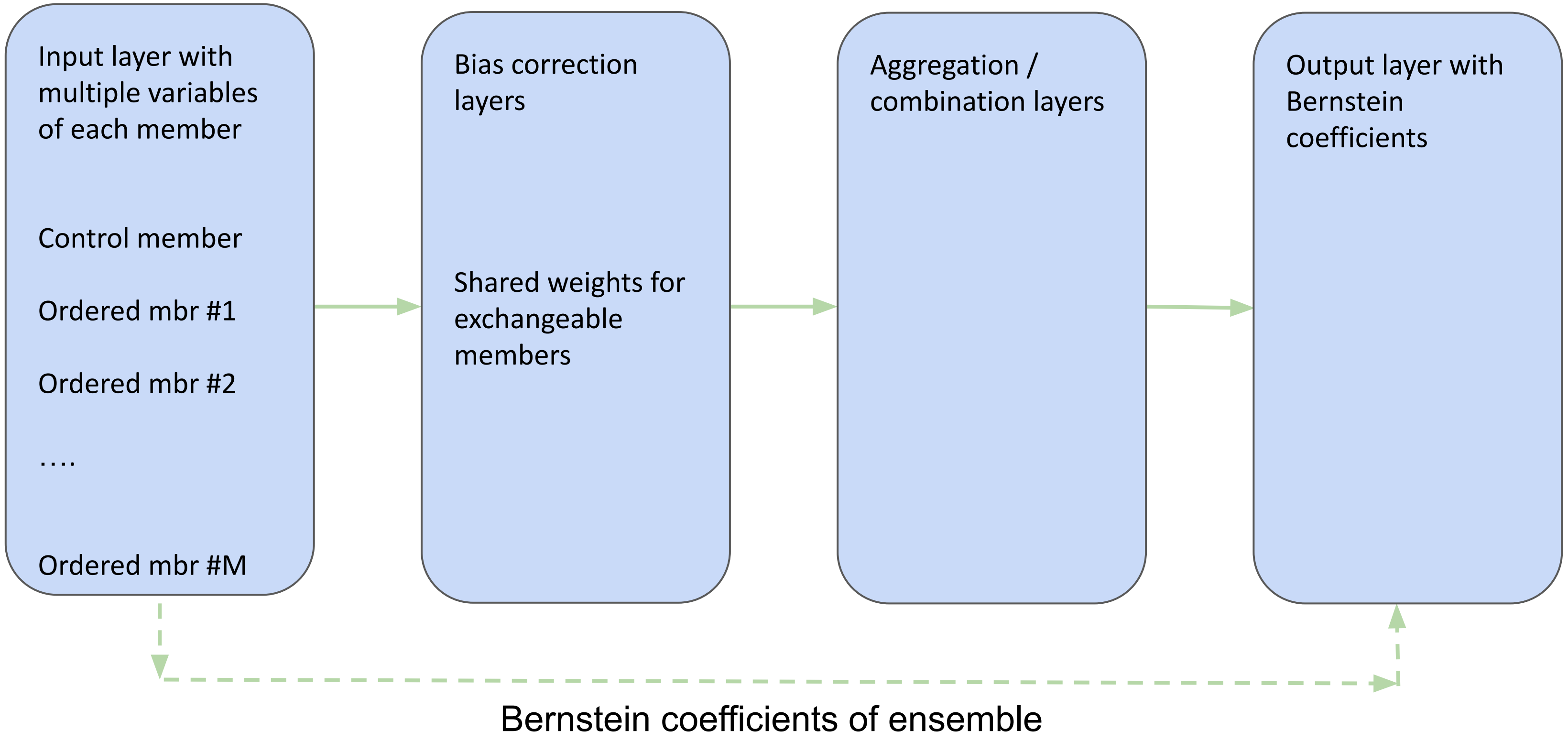




Bremnes (2020)
Schulz and Lerch (2022)
Gneiting et al (2023)
Höhlein et al (2024)

https://github.com/jbbremnes/BQNet.jl

Example of network architecture



Statistical challenges with precipitation

Continuous variable with a (large) point mass at zero

- mixed distributed
- need special attention

Modelling options

- discretisation of precipitation $\rightarrow$ categorical modelling
- separate models for the zeros (discrete) and the positive amounts (continuous)
  - predictive distribution by combination of the two models
- continuous model where zeros are treated as censored values
  - introduce a latent variable with no lower bound
  - adjust loss function
  - truncate at zero (non-positive part = probability of no precipitation)

Censored Bernstein Quantile Networks

Alternative 1: with prob of precip

- make a model for probability of precipitation **p(x)**
  - neural network, logistic regression, use of ensemble/scenarios, climatology etc.
- 1st epoch of BQN training
  - compute quantile loss only over levels τ and cases *x* where $\mathbf{p(x) > 1 - \tau, \forall \tau, x}$
- Remaining epochs
  - compute quantile loss for positive quantiles, i.e. over levels τ and cases *x* where $\mathbf{Q(\tau|x) > 0, \forall \tau, x}$

Censored linear single-quantile regression, Friederichs and Hense (2007)

Censored Bernstein Quantile Networks

Alternative 1: with prob of precip

- make a model for probability of precipitation **p(x)**
  - neural network, logistic regression, use of ensemble/scenarios, climatology etc.
- 1st epoch of BQN training
  - compute quantile loss only over levels τ and cases *x* where $\mathbf{p(x) > 1 - \tau, \forall \tau, x}$
- Remaining epochs
  - compute quantile loss for positive quantiles, i.e. over levels τ and cases *x* where $\mathbf{Q(\tau | x) > 0, \forall \tau, x}$

Alternative 2: without prob of precip

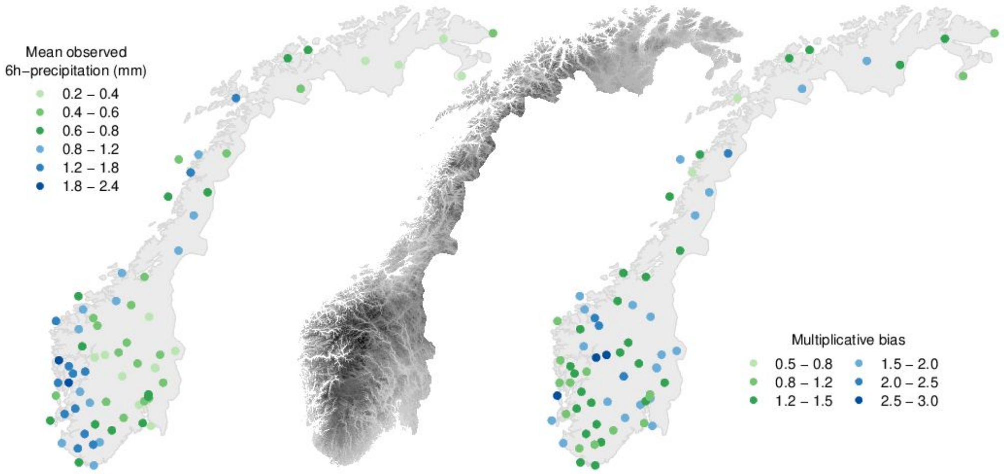
- all epochs
  - compute quantile loss for positive quantiles, i.e. over levels τ and cases *x* where $\mathbf{Q(\tau | x) > 0, \forall \tau, x}$
  - Note! random initialisation of network parameters implies random number of negative quantiles

Censored linear single-quantile regression, Friederichs and Hense (2007)

Example: 6h-precipitation forecasting

Data

- 70 Norwegian stations
- 00+66h ECMWF ENS reforecasts (11 members)
  - ensemble means of total precipitation, convective precipitation, total column cloud liquid water, CAPE, wind at 700 hPa
  - standard deviation of total precipitation
  - probability of precipitation
- training (#22997), validation (#5758) and test (#5633) datasets



Example: 6h-precipitation forecasting

Data

- 70 Norwegian stations
- 00+66h ECMWF ENS reforecasts (11 members)
    - ensemble means of total precipitation, convective precipitation, total column cloud liquid water, CAPE, wind at 700 hPa
    - standard deviation of total precipitation
    - probability of precipitation
- training (#22997), validation (#5758) and test (#5633) datasets

Methods (variants of BQN)

- no censoring
- censoring with probability input (alternative 1) from
    - neural network
    - logistic regression
    - ensemble
    - climatology
- censoring without probability input (alternative 2)
- 2-model approach
    - Neural net for probability of precipitation
    - BQN for positive precipitation amounts

Model selection/tuning and testing framework

- extensive tuning of each method on validation dataset
    - 4320 configurations trained and evaluated
- 5 best configurations of each method re-trained 5 times
    - predictions on test dataset (#5×5×5633)

Results

	BQN method	Quantile Skill Score (%) Overall	Quantile Skill Score (%) Extremes*
	No censoring	-0.35	0.16
Alt. 1	Censoring: Neural net	0.79	1.38
Alt. 1	Censoring: Logistic reg.	0.53	0.33
Alt. 1	Censoring: Ensemble prb	0.53	0.69
Alt. 1	Censoring: Climate	0.45	0.35
Alt. 2	Censoring	0.15	-0.32

2-model approach applied as reference

*) Extremes = 3 most extreme ENS cases for each station, 3×70 = 210 cases

Results

	BQN method	Quantile Skill Score (%) Overall	Quantile Skill Score (%) Extremes*	Brier Skill Score (%) 0 mm	Brier Skill Score (%) 0.05 mm	Brier Skill Score (%) 0.1 mm	Brier Skill Score (%) 10 mm
	No censoring	-0.35	0.16	-215.80	-5.24	-3.54	-0.14
Alt. 1	Censoring: Neural net	0.79	1.38	-2.24	-2.51	-2.65	0.79
Alt. 1	Censoring: Logistic reg.	0.53	0.33	-1.86	-2.04	-2.13	0.13
Alt. 1	Censoring: Ensemble prb	0.53	0.69	-1.38	-1.44	-1.50	0.03
Alt. 1	Censoring: Climate	0.45	0.35	-1.89	-2.00	-2.04	0.44
Alt. 2	Censoring	0.15	-0.32	-3.48	-3.83	-3.98	0.14

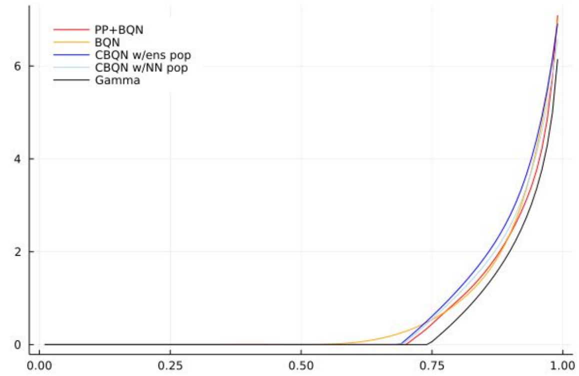
2-model approach applied as reference

*) Extremes = 3 most extreme ENS cases for each station, 3×70 = 210 cases

Example: 6h-precipitation forecasting

Summary

- Ignoring point masses at zero leads to unreliable quantiles
  - in particular for small amounts
  - but overall scores are not much affected
- Various censoring approaches gives about the same overall scores
  - useful to have an estimate of prob of precip
- cBQN about 25% (QSS +66h) better than ECMWF ENS 11-member reforecast
  - Brier skill score for prob of precip > 50%



Extensions to multivariate/scenario generation

Traditional approaches

- re-use of ranks
    - ensemble copula coupling (Bremnes (2007), Schefzik et al (2013))
        - ranks of forecast ensemble
    - Schaake shuffling (Clark et al (2004))
        - ranks of observation (sub)sets

Extensions to multivariate/scenario generation

Traditional approaches

- re-use of ranks
  - ensemble copula coupling (Bremnes (2007), Schefzik et al (2013))
    - ranks of forecast ensemble
  - Schaake shuffling (Clark et al (2004))
    - ranks of observation (sub)sets

Learning dependency structure with deep learning

- learn and fix marginals with cBQN
- create a network to learn dependency structure
  - generative or non-generative for fixed member size
  - optionally with a prior/template
- optimisation by multivariate proper scoring rules

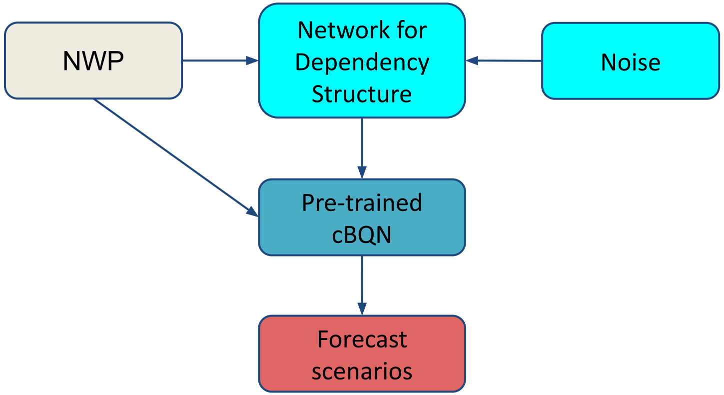
Extensions to multivariate/scenario generation

Traditional approaches

- re-use of ranks
  - ensemble copula coupling (Bremnes (2007), Schefzik et al (2013))
    - ranks of forecast ensemble
  - Schaake shuffling (Clark et al (2004))
    - ranks of observation (sub)sets

Learning dependency structure with deep learning

- learn and fix marginals with cBQN
- create a network to learn dependency structure
  - generative or non-generative for fixed member size
  - optionally with a prior/template
- optimisation by multivariate proper scoring rules



Destination Earth project: On-demand extremes (DE330)

NWP forecasts at hectometric resolution for expected extreme weather in Europe

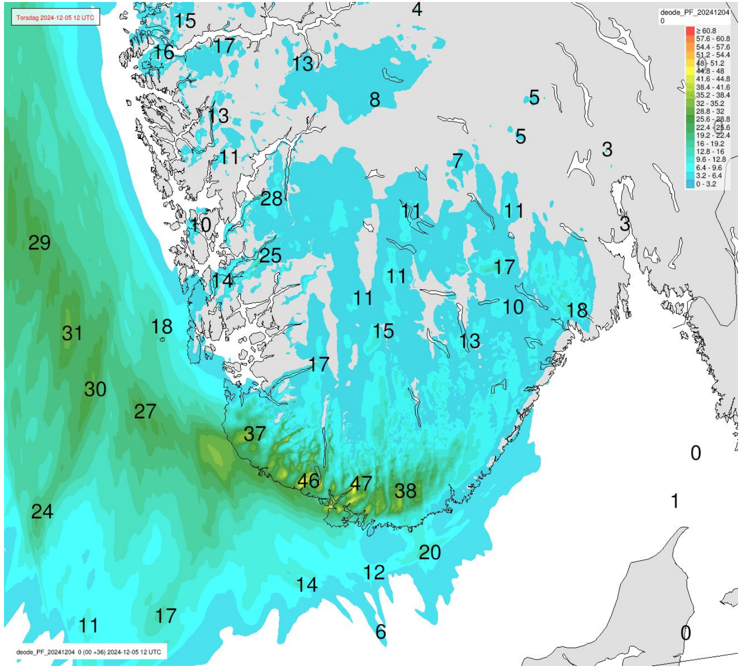
- NWP model run at ~500 m resolution
  - up to 2-3 days ahead
  - only once a day
  - 4.4 km Global DT NWP at boundary
- Domains are determined on-demand
  - proposals by a triggering algorithm
  - non-static domains
- Single deterministic runs

Suggested domains for flooding: 17 March 2025

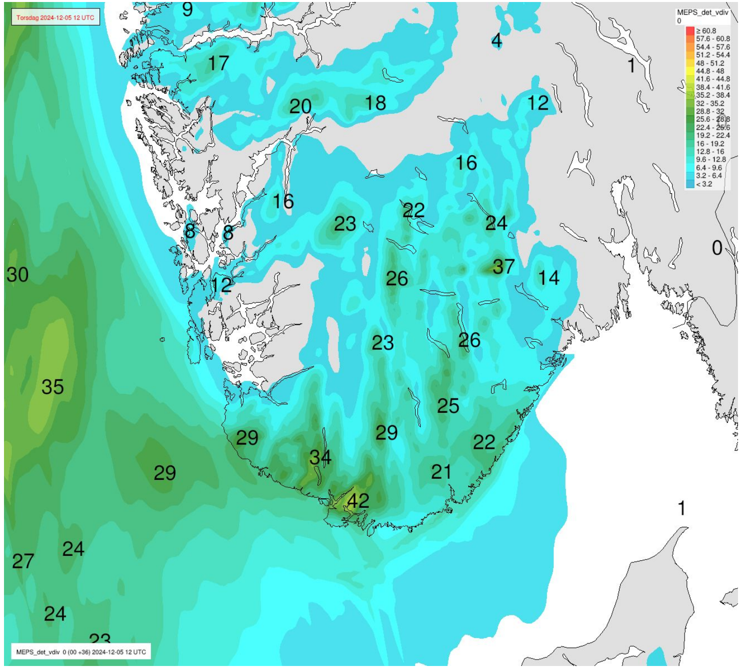
Destination Earth project: On-demand extremes (DE330)

NWP forecasts at hectometric resolution for expected extreme weather in Europe

HARMONIE-AROME 500m



MEPS control 2500m



Destination Earth project: On-demand extremes (DE330)

Uncertainty quantification with cBQN (amongst other)

Training against synops – predicting on 500m grid

- Make one cBQN model for any location and lead time
- Input variables: NWP variables, lead time, orographic info, time of day and year, climate variables(?), geographical coordinates.
- Output: Bernstein quantile function at any collection of points for lead times up to 2 days ahead
    - cBQN model must be evaluated separately for each location and lead time of interest

Destination Earth project: On-demand extremes (DE330)

Uncertainty quantification with cBQN (amongst other)

Training against synops – predicting on 500m grid

- Make one cBQN model for any location and lead time
- Input variables: NWP variables, lead time, orographic info, time of day and year, climate variables(?), geographical coordinates.
- Output: Bernstein quantile function at any collection of points for lead times up to 2 days ahead
  - cBQN model must be evaluated separately for each location and lead time of interest

Downscaling: Global DT (4.4 km) → 500m grid

- Objective: scenarios / ensemble
- cBQN model for marginals
- Separate deep learning model for the spatio-temporal dependency structure
- Forecast scenarios can be provided every day for any domain

Summary

- Censored Bernstein Quantile Networks works well for probabilistic fine-tuning of precipitation forecasts
- BQN can be applied to more or less any non-categorical variable
  - … like streamflow, …
- Current and future work
  - more focus on extremes and distribution tails
  - other network architectures, in particular for gridded data
  - modelling of dependency structures (time, space and between variables)

https://github.com/jbbremnes/BQNet.jl