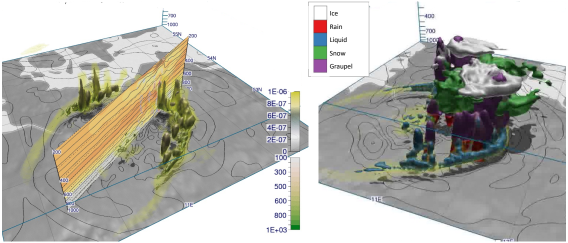
Learning model parameters from observations by combining data assimilation and machine learning

Tijana Janjić, Stefanie Legler, Yvonne Ruckstuhl

[1] Mathematical Institute for Machine Learning and Data Science,
KU Eichstätt-Ingolstadt

MIDS
Mathematical Institute for
Machine Learning and Data Science

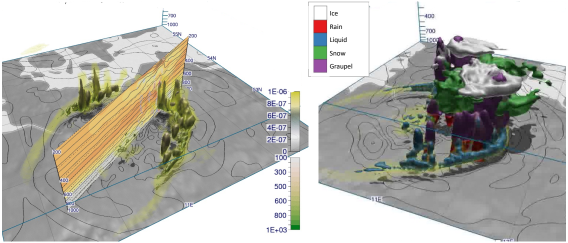
Convective/Storm scale application



Met3D

- High resolution NWP models of atmosphere that incorporate our knowledge of the dynamics and physics.

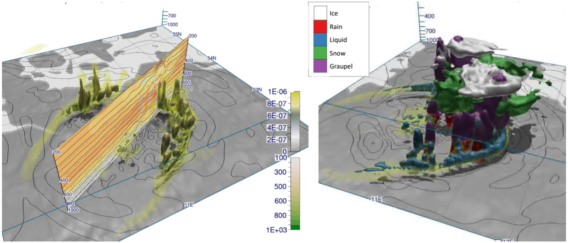
Convective/Storm scale application



Met3D

- High resolution NWP models of atmosphere that incorporate our knowledge of the dynamics and physics.
- For modern geophysical models state vector is of size $10^6 - 10^8$.

Convective/Storm scale application



Met3D

- High resolution NWP models of atmosphere that incorporate our knowledge of the dynamics and physics.
- For modern geophysical models state vector is of size $10^6 - 10^8$.
- Accurate initial conditions are crucial for prediction, even more in the future due to climate change and intensification of the water cycle.

Convective/Storm Scale Data assimilation

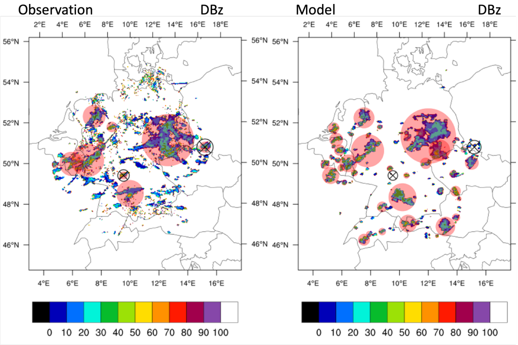
- Problem is highly nonlinear and highly non-Gaussian.
- In addition to global observing system, radar is the primary new observation.
- The state vector $\mathbf{w}_k^b$ at time k consists in addition of prognostic hydrometeors variables (rain, graupel, snow, ...) at all grid points.
- Depending on microphysical scheme in model, even higher dimensional problem with third of the variables (one order of magnitude) that need to be nonnegative.

Convective/Storm Scale Data assimilation

- Problem is highly nonlinear and highly non-Gaussian.
- In addition to global observing system, radar is the primary new observation.
- The state vector $\mathbf{w}_k^b$ at time k consists in addition of prognostic hydrometeors variables (rain, graupel, snow, ...) at all grid points.
- Depending on microphysical scheme in model, even higher dimensional problem with third of the variables (one order of magnitude) that need to be nonnegative.
- Uncertainty quantification is a challenge for both model and observation error.

Uncertainty of geophysical models

- Geophysical models of atmosphere incorporate our knowledge of the dynamics and physics.



- These models are not perfect.
- One of the reasons for model error is limited knowledge of model parameters.

Joint state and parameter estimation

Parameters are not observed

To learn parameters of a numerical model from observations

- Data assimilation
- Machine learning

Joint state and parameter estimation

Parameters are not observed

To learn parameters of a numerical model from observations

- Data assimilation
- Machine learning

Data assimilation:

Augment state vector $\mathbf{x}$ with parameters θ

$$\mathbf{w}_k^a = \begin{bmatrix} \mathbf{x}_k^a \\ \theta_k^a \end{bmatrix}$$

EnKF

Propagation step. Propagate the mean and the covariance with the dynamics between observations. Prior to new observation we have $\mathbf{w}_k^f$ and its covariance $\mathbf{P}_k^f$.

$$\mathbf{w}_k^{f,i} = \mathcal{M}\mathbf{w}_{k-1}^{a,i} + \mathbf{q}_k^i \quad i = 1, \ldots N$$

$$\mathbf{P}_k^f = \frac{1}{N-1}\sum_{i=1}^{N}[\mathbf{w}_k^{f,i} - \mathbf{w}_k^f][\mathbf{w}_k^{f,i} - \mathbf{w}_k^f]^T.$$

Kalman analysis.

$$\mathbf{w}_k^{a,i} = \mathbf{w}_k^{f,i} + \mathbf{K}_k(\mathbf{w}_k^o + r^i - \mathbf{H}_k\mathbf{w}_k^f),$$

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T\mathbf{P}_k^f$$

Derived using $q^i \sim \mathcal{N}(0, \mathbf{Q})$, $r^i \sim \mathcal{N}(0, \mathbf{R})$, $\mathbf{w}_0^f \sim \mathcal{N}(0, \mathbf{P_0^f})$ and all uncorrelated.

Augmented EnKF

Stochastic model for parameters

$$\theta_k^{f,i} = \theta_{k-1}^{a,i} + \mathbf{D}_{k-1}\mathbf{C}^{\frac{1}{2}}\eta^i$$

$\theta_{k-1}^{a,i}$ is the raw analysis value after applying the EnKF

$\theta_k^{f,i}$ the perturbed value that is passed to the model

$\mathbf{D}_{k-1}$ is a diagonal matrix that locally controls the ensemble spread

$\mathbf{C}^{\frac{1}{2}}$ is the error correlation matrix that specifies the correlations within parameter field

$\eta^i \sim \mathcal{N}(0, I)$ is the random realization of the stochastic model.

Example

Can accounting for model error by allowing uncertainty in parameters reduce state error?

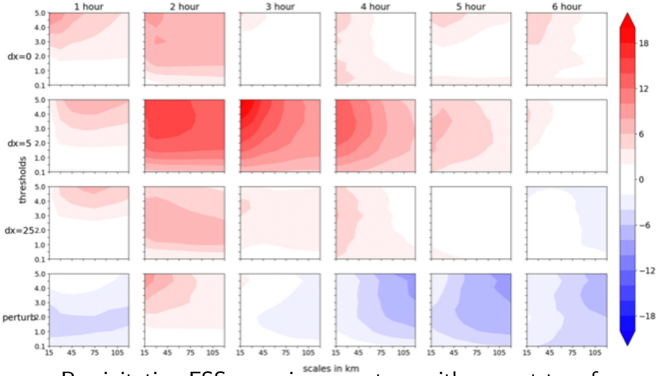
Example

Can accounting for model error by allowing uncertainty in parameters reduce state error?

- Parameter: roughness length (2D field)
- Operational atmospheric model
- All observations including volumes of radar reflectivity assimilated with LETKF
- Stochastic model for roughness length
  - with correlation length scale of either 0, 5, or 25 grid points
  - at each grid point a temporally constant standard deviation of 25% of the original roughness length parameter value
  - we set $z^a_{0_{t-\mathbf{1},i}} \leftarrow \max\left(z^a_{0_{t-\mathbf{1},i}}, z_0^{min_b}\right)$ before the dynamical model is applied (non-negativity)

Ruckstuhl, Y. and T. Janjić, 2020, Combined State-Parameter Estimation with the LETKF for Convective-Scale Weather Forecasting, Mon. Wea. Rev., 148, 1607–1628.

Example



Precipitation FSS score in percentage with respect to ref

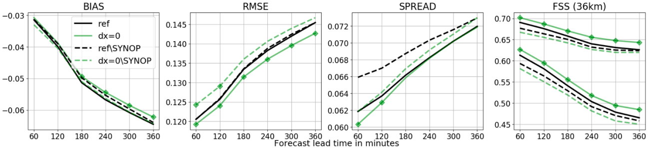
Accounting for model error by allowing uncertainty in parameters can reduce state error.

Ruckstuhl, Y. and T. Janjić, 2020, Combined State-Parameter Estimation with the LETKF for Convective-Scale Weather Forecasting, Mon. Wea. Rev., 148, 1607–1628.

Alternative Algorithms for joint estimation

For data assimilation

- Parameters updated through cross-correllations



Verification against visible Meteosat SEVIRI images (Scheck et al. 2016, 2018).
FSS thresholds 0.3 and 0.5 (Rusckstuhl and Janjic 2020).

- Sampling error and localization
- Stochastic models for parameters needed

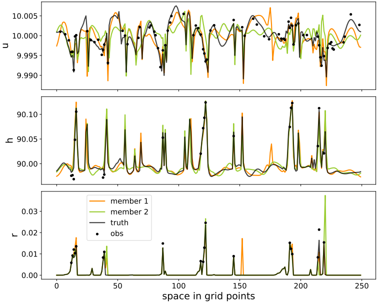
Alternative algorithms ?
Non-Gaussian methods needed for parameter estimation?

Modified shallow water model

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial(\phi + \gamma^2 r)}{\partial x} = \beta_u + D_u \frac{\partial^2 u}{\partial x^2}, \phi = \begin{cases} \phi_c & \text{if } h > h_c \\ gh & \text{otherwise}, \end{cases}$$

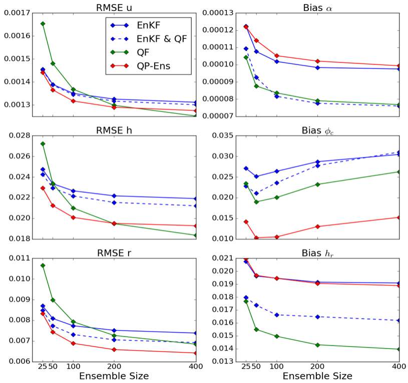
$$\frac{\partial r}{\partial t} + u\frac{\partial r}{\partial x} = D_r \frac{\partial^2 r}{\partial x^2} - \alpha r - \begin{cases} \delta \frac{\partial u}{\partial x}, & \text{if } h > h_r \text{ and } \frac{\partial u}{\partial x} < 0 \\ 0 & \text{otherwise}, \end{cases}$$

$$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} = D_h \frac{\partial^2 h}{\partial x^2}.$$



Wuersch and Craig 2014: A simple dynamical model of cumulus convection for data assimilation research., Meteorol. Z., 23, 483-490.

Non-Gaussian aspects



QPEns (Janjic et al 2014), QF (Hodyss 2011,2012), EnKF (Evensen 2003).

Ruckstuhl Y. and T. Janjic, 2018, Parameter and state estimation with ensemble Kalman filter based approaches for convective scale data assimilation, Q. J. R. Meteorol. Soc., 144:712, 826–841.

NNs

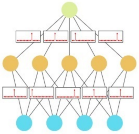
$Input:\ state\ x_t$
- $wind\ u$
- $water\ level\ h$
- $rain\ r$

$\longrightarrow$

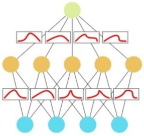
$Neural\ Network$
- $fully\ connected$
- $feedforward$

$\longrightarrow$

$Output:\ model\ parameters\ \theta_t$
- $rain\ removal\ rate\ \alpha$
- $const.\ geopotential\ \phi_c$
- $rain\ threshold\ h_r$



$Ensemble\ of\ Point\ Estimate\ Neural\ Networks$
- $train\ 15\ neural\ networks$
- $4\ linear\ layers$
- $batch-norm\ +\ dropout$
- $activation: ReLU$
- $optimizer: Adam$

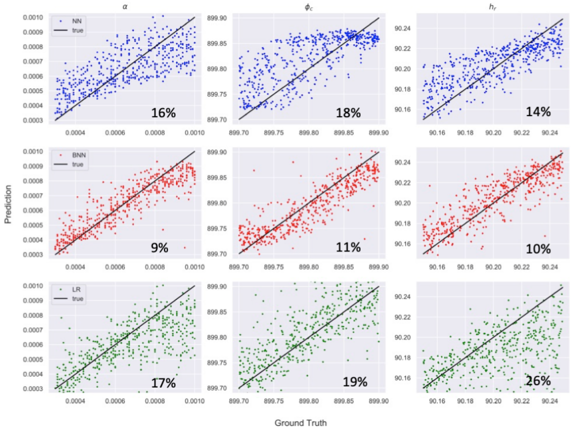


$Bayesian\ Neural\ Network$
- $N(\mu,\sigma)\ for\ each\ weight$
- $4\ linear\ layers$
- $batch-norm$
- $activation:\ LeakyReLU$
- $optimizer:\ SVI$

(figure from Jospin et al. 2020)

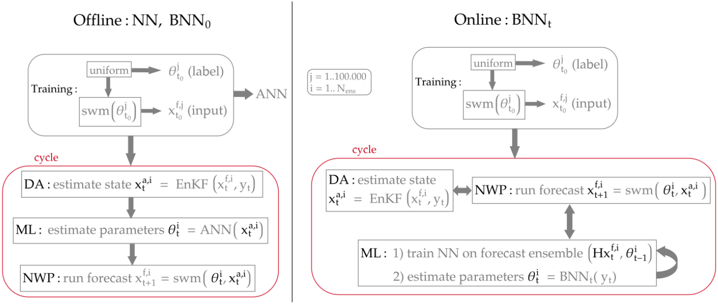
Trained on 100 000 model simulations using random parameter values from the uniform distributions.

NN accuracy



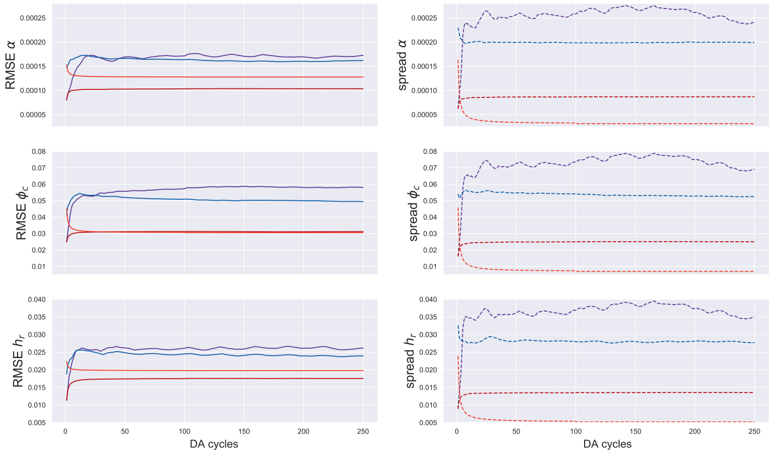
Output of NN (blue dots), BNN (red dots), and LR (green dots) against corresponding ground truths and ideal output (black lines) of 500 samples

DA+NNs for parameters



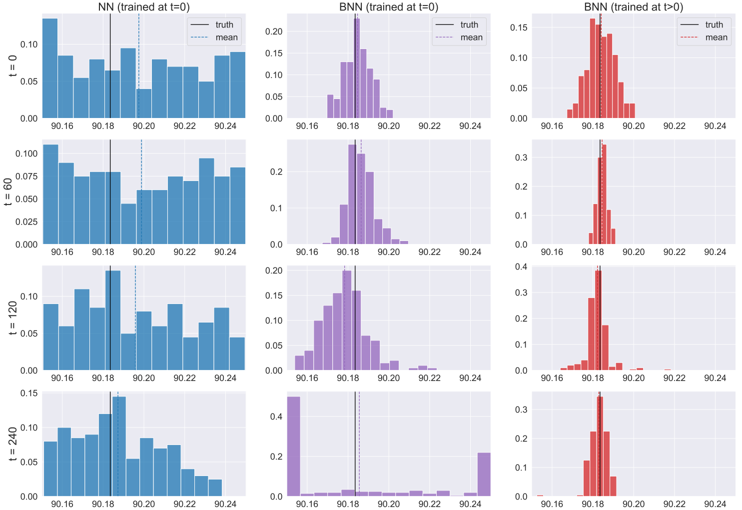
Legler and Janjic, 2022: Combining data assimilation and machine learning to estimate parameters of a convective-scale model. Q.J.R. Meteorol. Soc., 148, 860-874, https://doi.org/10.1002/qj.4235.

Time evolution of errors

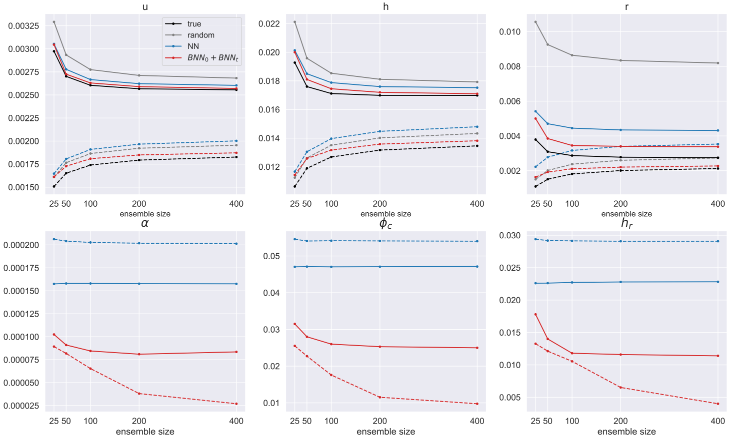


Offline training: BNN_0, NNs; BNN_t training continues online during DA

Statistical Properties

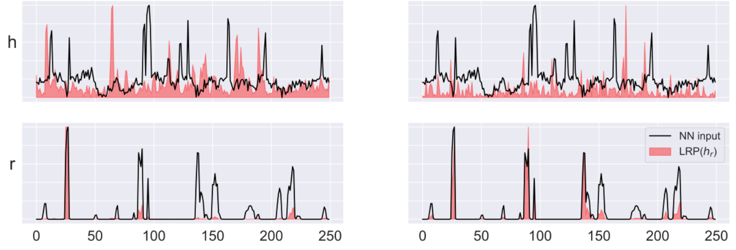


Sensitivity to Ensemble Size



Results are averaged over 100 individual experiments with different ground truth values. True parameter is used in simulation (black), wrong (gray).

LRP

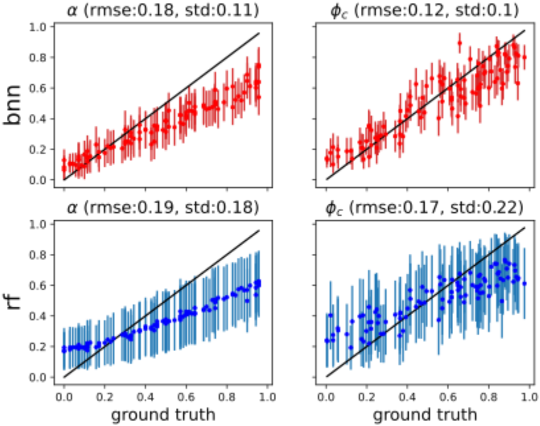


parameter	u	h	r
α	20%	57%	23%
ϕ_c	23%	61%	16%
h_r	21%	63%	16%

parameter	u	h	r
α	16%	59%	25%
ϕ_c	19%	44%	37%
h_r	21%	38%	41%

We follow Toms et al. 2020 to calculate Layer-wise relevance propagation (LRP) map for h_r in case Left: 3 parameters are estimated simultaneously. Right: LRP map when only h_r is estimated.

Uncertainty quantification



Comparison of two ML methods (BNNs and random forest) for estimates of a parameters and their uncertainty in a modified shallow water model (Legler et al. 2022).

Conclusion

- Estimating parameters can reduce prediction errors
- If stochastic model for parameters can be made, EnKF can be used to objectively estimate parameters from data
- Methods that take into account non-Gaussian aspects of parameter estimation can further improve parameter estimates
- Alternatively, Bayesian neural networks and Bayesian approximations of point estimate neural networks are able to estimate model parameters and their relevant statistics
- ML is able to retrieve hidden relationships while DA can provide data-sets for training/inference from sparse and noisy observations (Brajard et al. 2020; Bonavita and Laloyaux 2020; Ruckstuhl et al. 2021, Farchi et al. 2021)
- ML results are comparable, but not better than augmented data assimilation estimation on our test case.