Status of Targeted Covariance Inflation and CML Data Assimilation

RealPEP

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

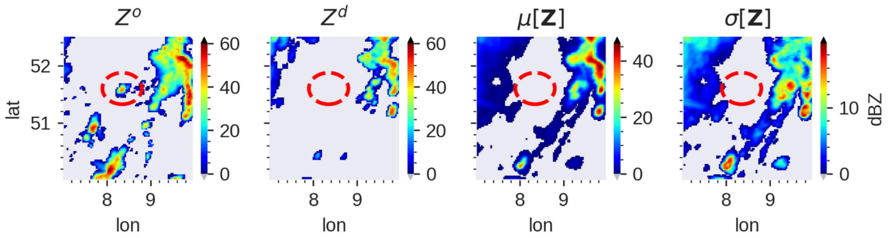
DWD

- K. Vobig (DWD)
- R. Potthast (DWD)
- C. Chwala (KIT)
- J. Polz (KIT)
- U. Blahak (DWD)
- J. Mendrok (DWD)
- ...

Targeted Covariance Inflation

TCI Basics: Motivation & Recap

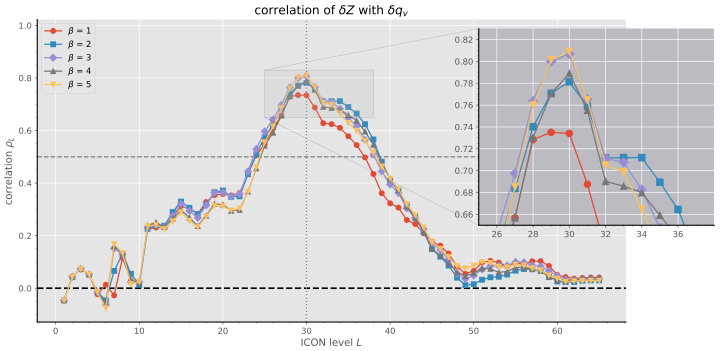
- LETKF data assimilation (DA) may give **small increments** due to **small ensemble** spread – even in the presence of **large discrepancies** between obs./sim. REFLs



- general idea of **targeted covariance inflation** (TCI) for REFLs:
  - assign individual **"virtual" simulated REFL** to **each member** via certain **algorithm/model** → spread is increased → previously discarded observations may be **actively assimilated again** → 🙂
- TCI evolved to **process-oriented** and **conditional** approach
  - conditional → **reduce noise** introduced into system
  - process orientation → accurately **initiate convection**
- TCI **implemented** by pre-processing feedback files (before entering LETKF) and altering simulated REFLs for each member

TCI Basics: Process Orientation

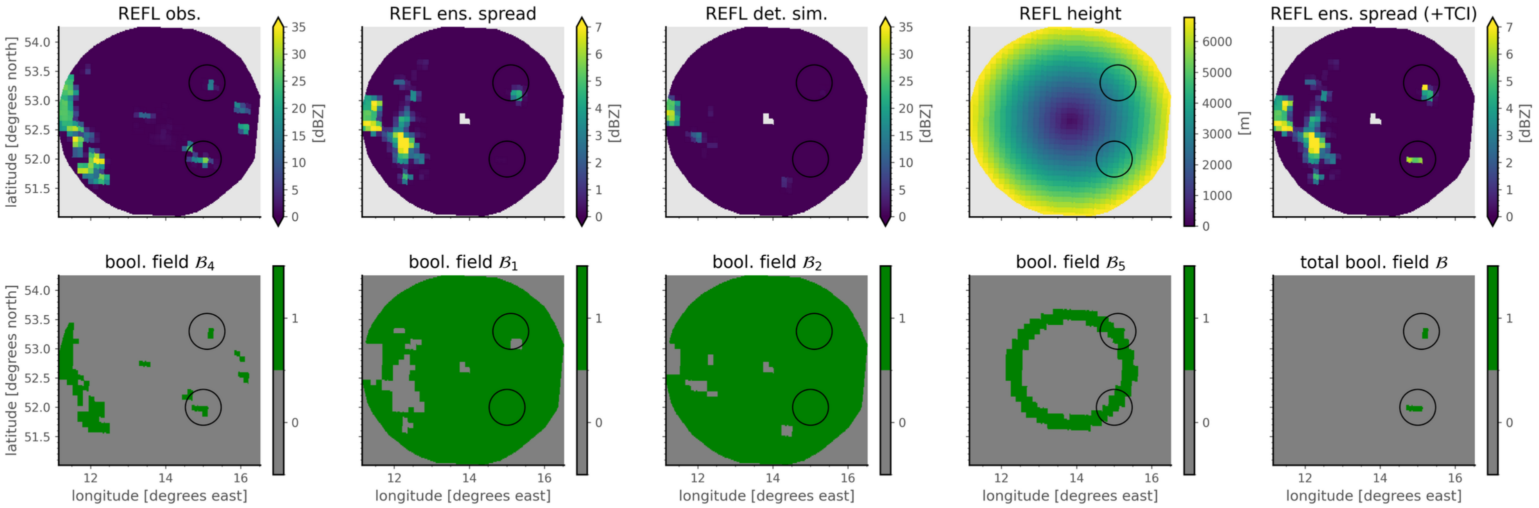
- process-oriented TCI: accurately **trigger convection**/dynamic generation of REFLs
- determine **values for “virtual” REFLs** accordingly
- using **simple linear models** M with **spec. humidity** qv as predictor
  - M: δZi(x,y,h,t) = α * δqvi(x,y,h’,t)



- **model training/selection** based solely on data in the nearest spatio-temporal vicinity of convective events
- **overall idea:** spread of qv “imprinted” onto spread of Z → assim. “favors” members with more humidity → additional increments for humidity qv are produced → model (hopefully) generates qr/qs/qg →EMVORADO simulates REFL
- in progress: **machine-learning** based model→more **flexible/powerful** approach

TCI Basics: Conditions

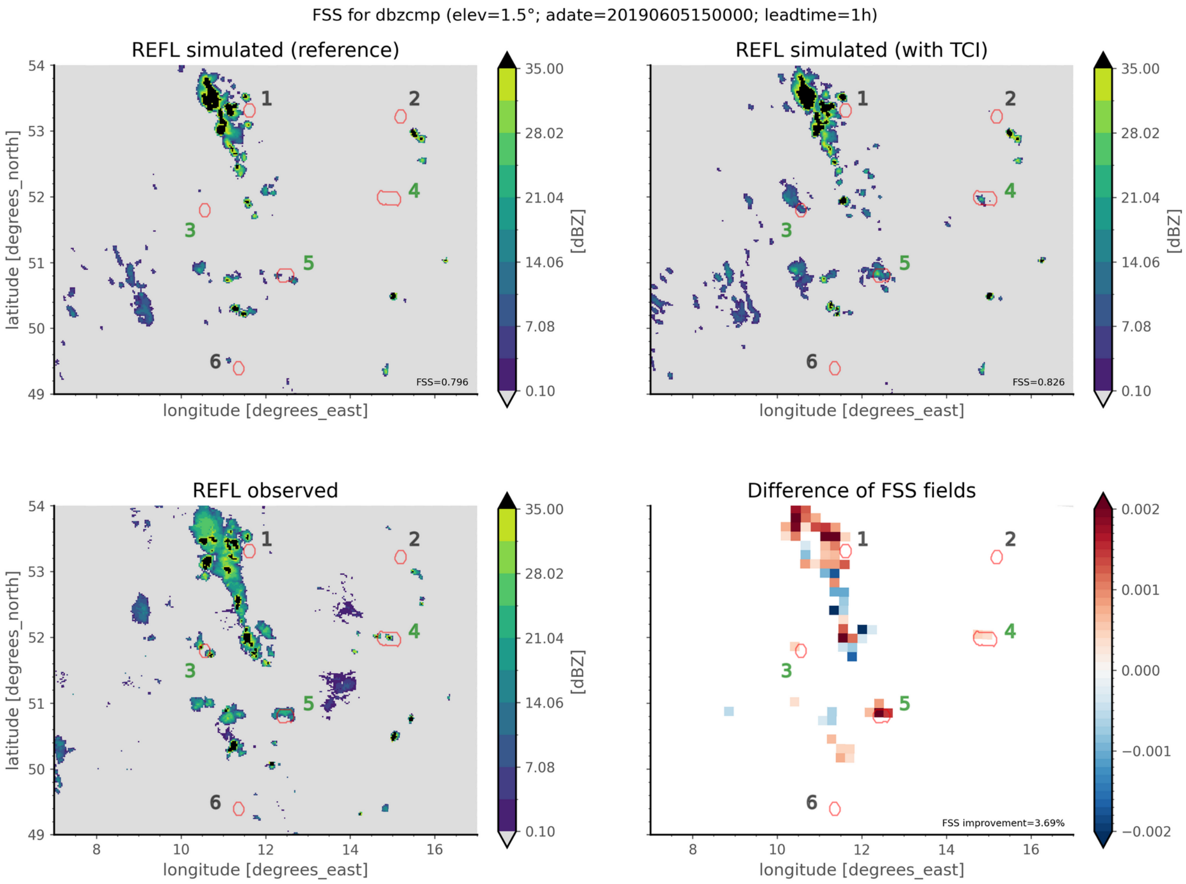
- **conditional TCI** aims to **reduce noise** introduced into the system state
- especially relevant for **fully-cycled DA** runs
- TCI only active at **minimal set of spatial points** (less is more)
- each observation must fulfill a certain **set of conditions** for TCI to be active: missing spread, large enough observation, ...



TCI Case Study: Setup

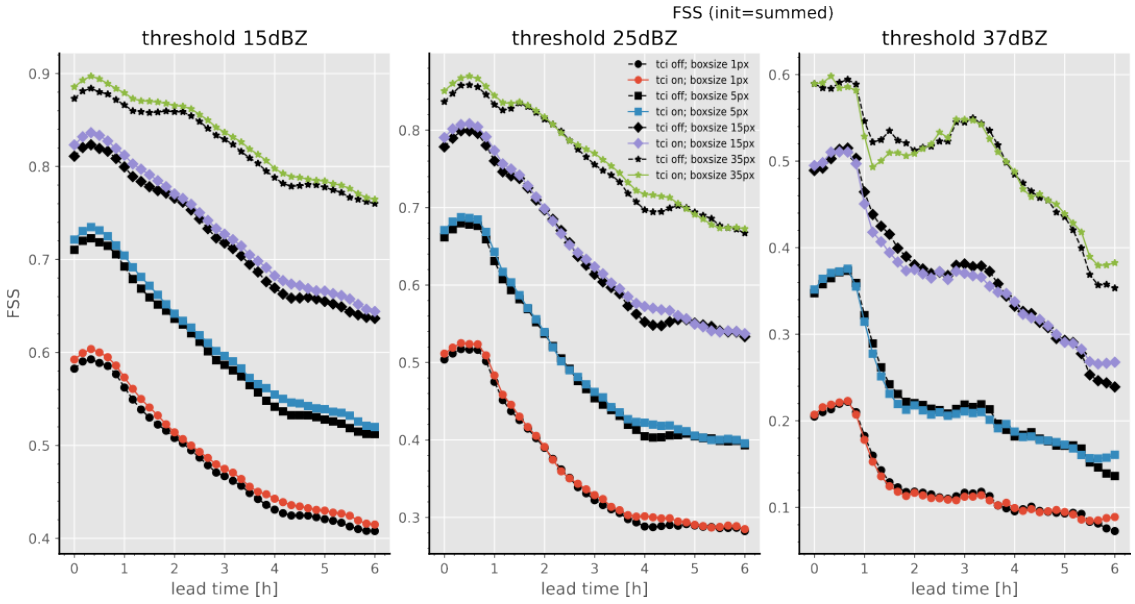
- performed two data assimilation cycles: **“reference” / “TCI” cycle**
  - cycles are based on **operational ICON-KENDA** framework
  - assimilation of **conv.** and 3D **radar** data
  - **Latent Heat Nudging** (LHN) mechanism active
  - **free forecasts** starting every 3h, max. leadtime 6h
  - period: from 2019-06-03 to 2019-06-20
- TCI algorithm is applied:
  - **hourly** at every LETKF assimilation step
  - to **ALL radar data** of German radar network

TCI Case Study: Verification I (REFL Forecasts)

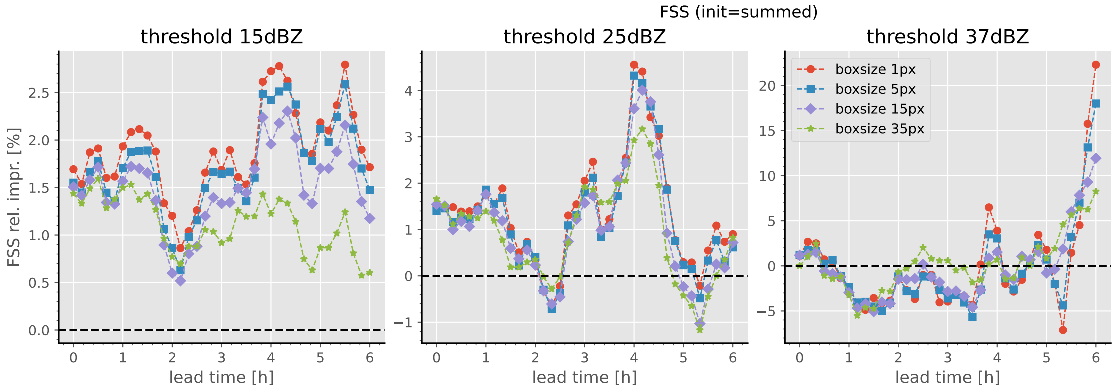


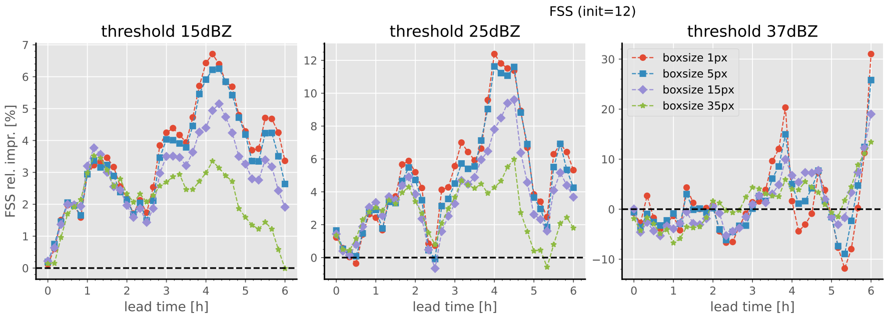
- **forecasts** initialized from reference/TCI ass. cycle
- forecast lead time 1h
- ass. cycles already ran for ~2 days
- depiction of **forecasted REFLs** (“interesting” region)
- source for differences:
  - accumulation effects
  - last assim. at 15 UTC
- result: **accurate, dynamic generation** of REFLs

TCI Case Study: Verification II (FSS for REFLs)



- Fractional Skill Scores (FSS) for **REFLs from free forecasts**
- result I: clear FSS improvement for lead times of up to 6h
- result II: positive impact even more pronounced for forecast initialized at 12 UTC, **FSS improvement by up to 10%**





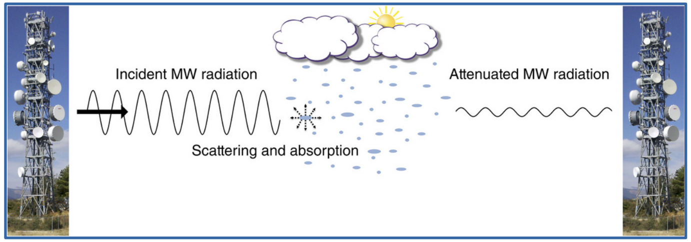
TCI Summary & Outlook

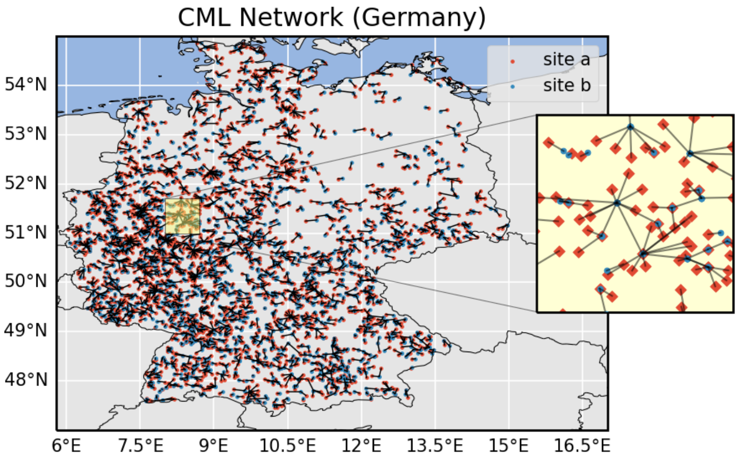
- discussed most recent **conditional and process-oriented TCI**
- overall, TCI **results are positive**
  - TCI leads to accurate generation of **“new” REFL cells**
  - TCI **improves fractional skill scores** (FSS) of REFL forecasts over lead times of 6h and **by up to 10%**
- **paper** discussing the TCI recently submitted:
  - Vobig et al., https://doi.org/10.5194/egusphere-2024-2876, 2024
- next steps:
  - work on **operationalization** of TCI
  - continue work on **machine learning** based TCI

CML Data Assimilation

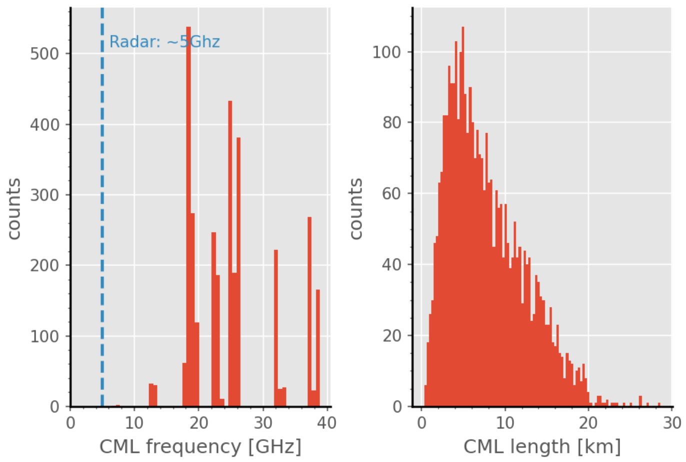
CML Basics: Motivation & Overview

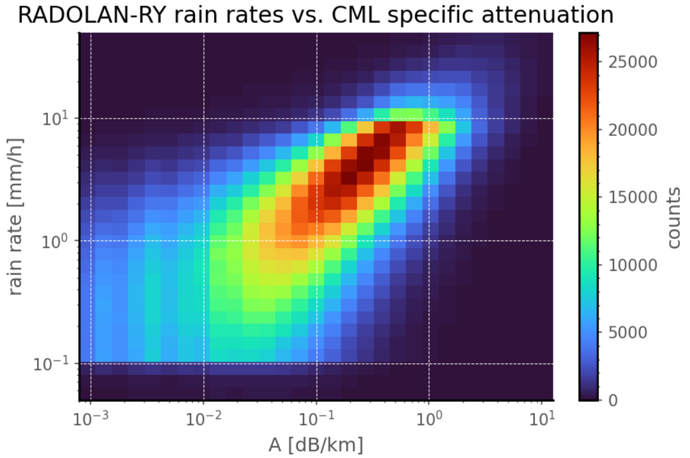
- overall **objective** here: **data assimilation** (DA) of **Commercial microwave link** (CML) data in NWP models for **improving QPF**
  - (How much) does it improve QPF?
  - How does it compare to Radar DA?
- CMLs employed for the **interconnection** of (commercial) **cell phone towers**
- transmitted radiation may be attenuated by, e.g., raindrops → **CML attenuation** carries information about atmospheric conditions between two towers
- ~4000 CMLs in current dataset for June 2019 with resolution of 1min





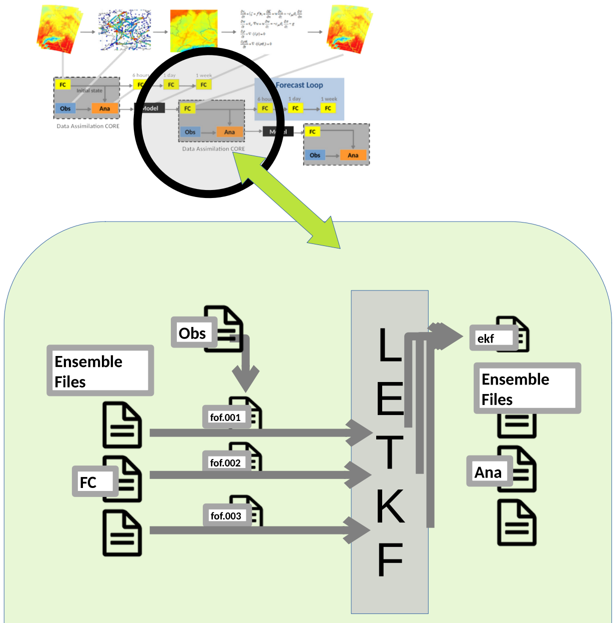
- **CML frequency** above DWD **Radar frequency** → different physics involved!
- use **path-integrated specific attenuation A** (with unit dB/km) for assimilation
- direct **relationship of A with rain rate** via power law
  - also empirically hinted at by “linear” relationship on double log. scale (see plot)
- (very) noisy data for A < ~10^{-2} dB/km → use as **cutoff** for CML DA





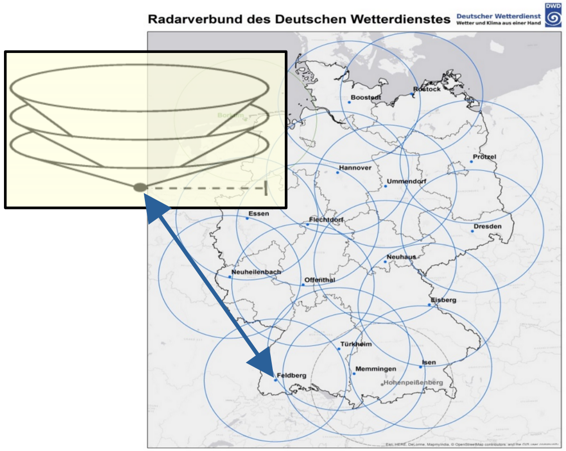
CML Basics: LETKF DA System

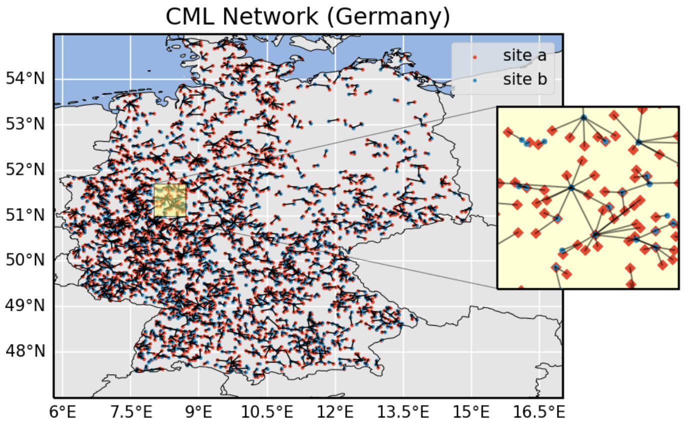
- for a LETKF data assimilation it is necessary to generate **feedback/fof files**
- each (ensemble) fof file contains all data relevant to LETKF assimilation (at specific date)
- particularly, for each observation there has to be a **simulated model equivalent**
- **system** for construction of **CML fof files:**
  - perform all necessary data (pre-)processing steps: EMVORADO calculations, temporal superobbing, etc.
  - implemented (mostly) in Python
  - integrated in BACY → quasi-operational DA exps.



- fof.*: sim. + obs. quantities of ens. members
- LETKF produces increments depending on innovations + Kalman gain

- employ radar forward operator **EMVORADO** for computing simulated **CML attenuations** A
- differences between **Radar and CML**:
  - Radar: 17 stations, many azimuths, few elevations, frequency ~5 GHz
  - CML: ~4000 “stations”/sender, individual azimuth/elevation (only one per station) and frequency within 10 – 40 GHz
- each CML sender is interpreted as a single Radar station with **individual** lat/lon/level, azimuth/elev. of ray, frequency, etc.
- perform EMVORADO run for each **ensemble member** based on ICON-D2 model fields



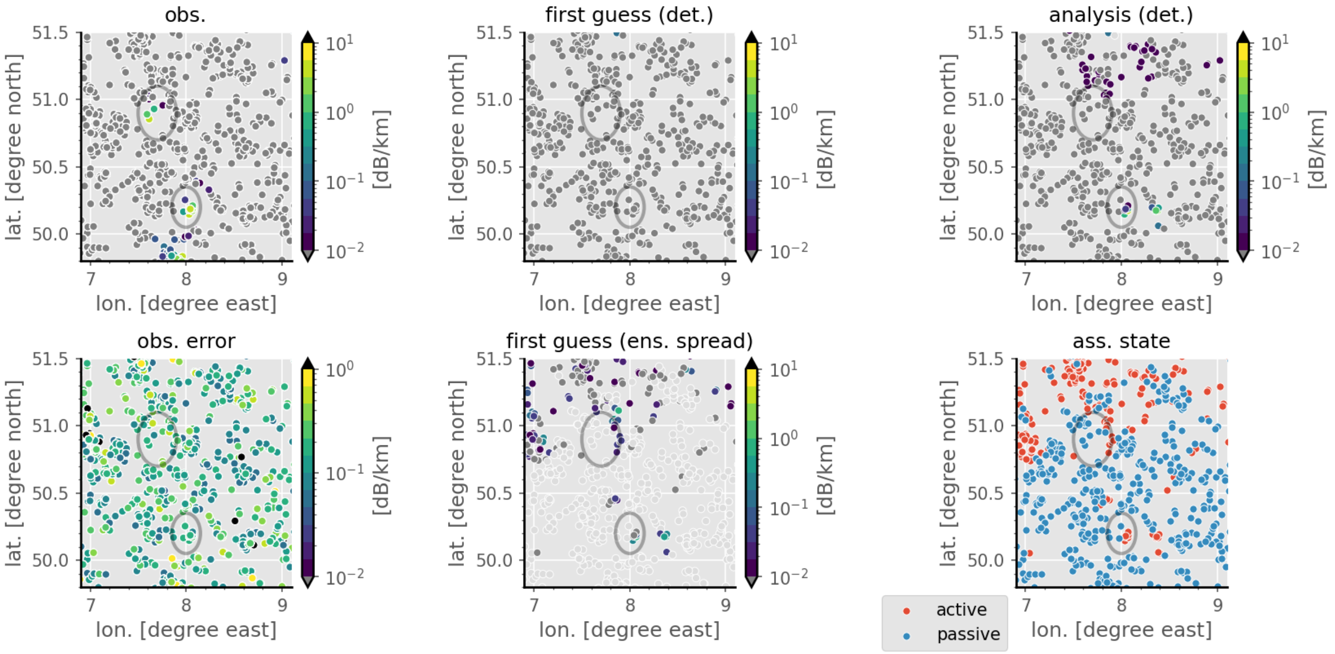


CML Case Study: Setup

- perform single BACY **"core-more runs"**:
  - **single LETKF assim.** followed by **ICON model run**
  - assimilating **ALL available CMLs** at 2019-06-03T12:00
  - **branching off** from **"parent" BACY cycle** during which only conventional data is assim.: no LHN (!), no RADAR assim., etc.
  - study different DA schemes: **conv**, **CML**, **radar**, **conv+CML**, ...
- study LETKF output, ICON increments, model dynamics, and FSS
- eventually zoom into "interesting" regions exhibiting certain properties:
  - **large discrepancies** between obs. and sim. REFLs
  - **sizeable spread** for sim. REFLs
  - **"enough" CML stations**

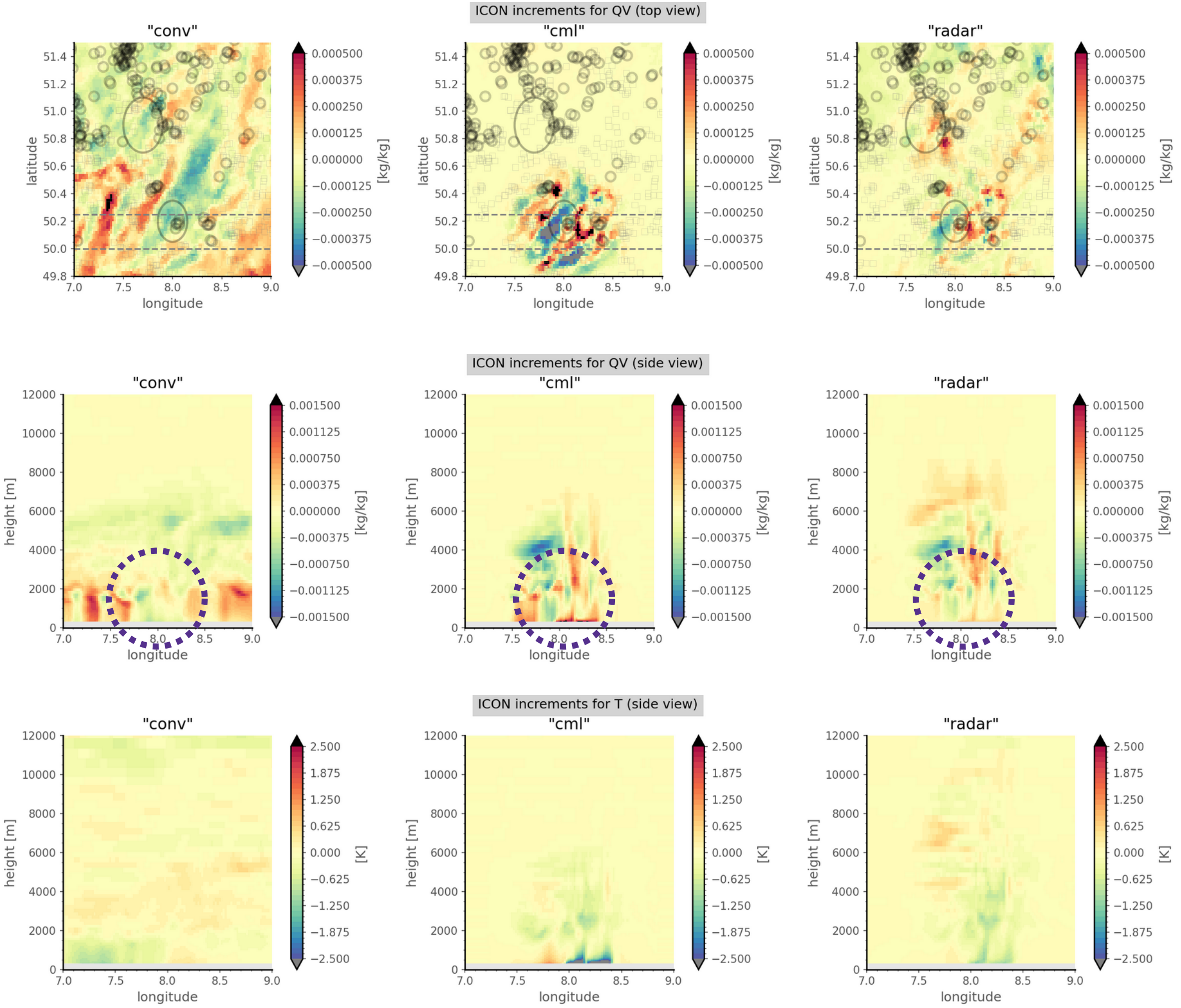
CML Case Study: LETKF Assimilation Results

representation of relevant LETKF assimilation input/output data (from “ekfCML” file)



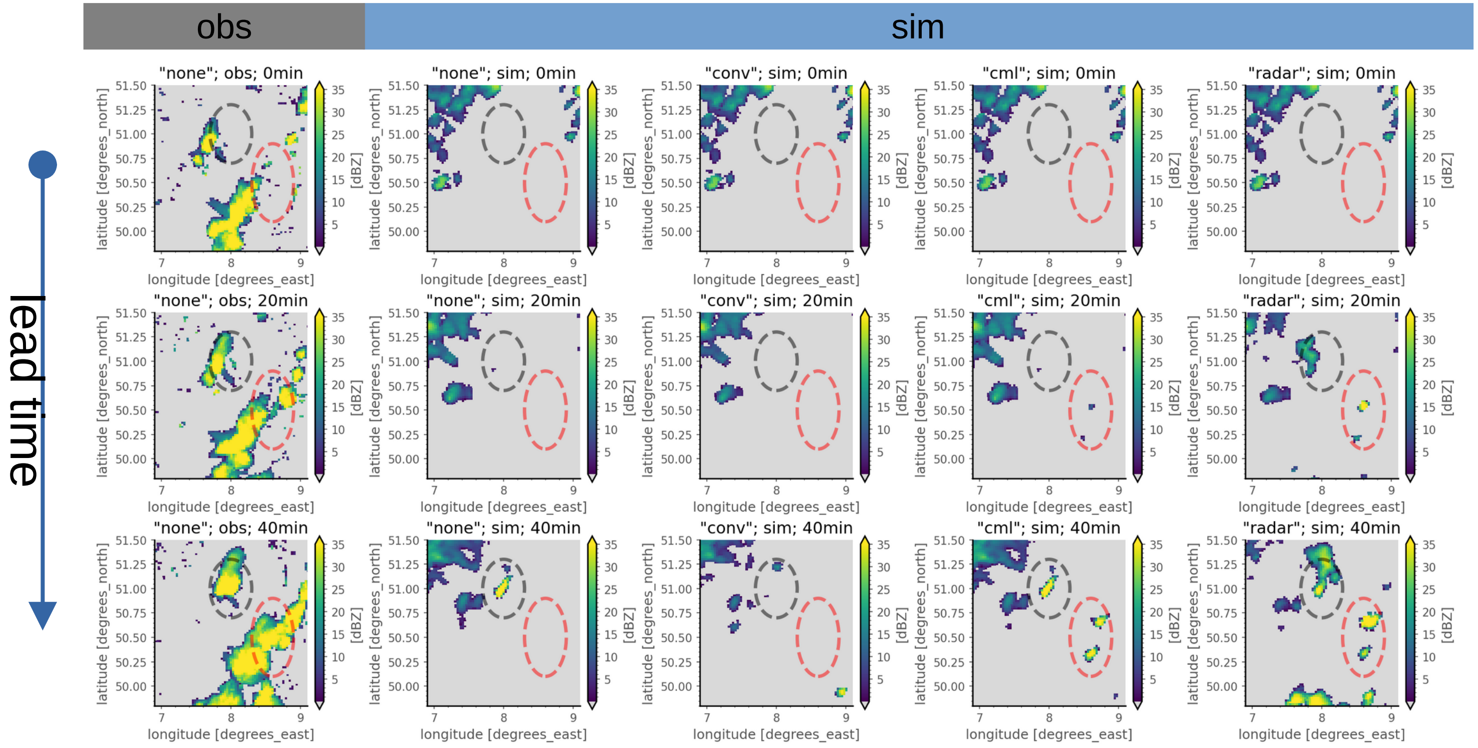
- **only** assimilating **CML** data here
- dynamic obs. error: 1 dB / “CML length”
- first-guess check switched off
- vert. localization: 0.3
- horiz. localization: 16.0

CML Case Study: LETKF Increments



- LETKF **increments** for QV and T
- reduced 3D to 2D fields via mean along height dimension (→ **top view**) or lat. dimension (→ **side view**)
- result: clear differences to conv. DA become apparent; CML and radar rather similar

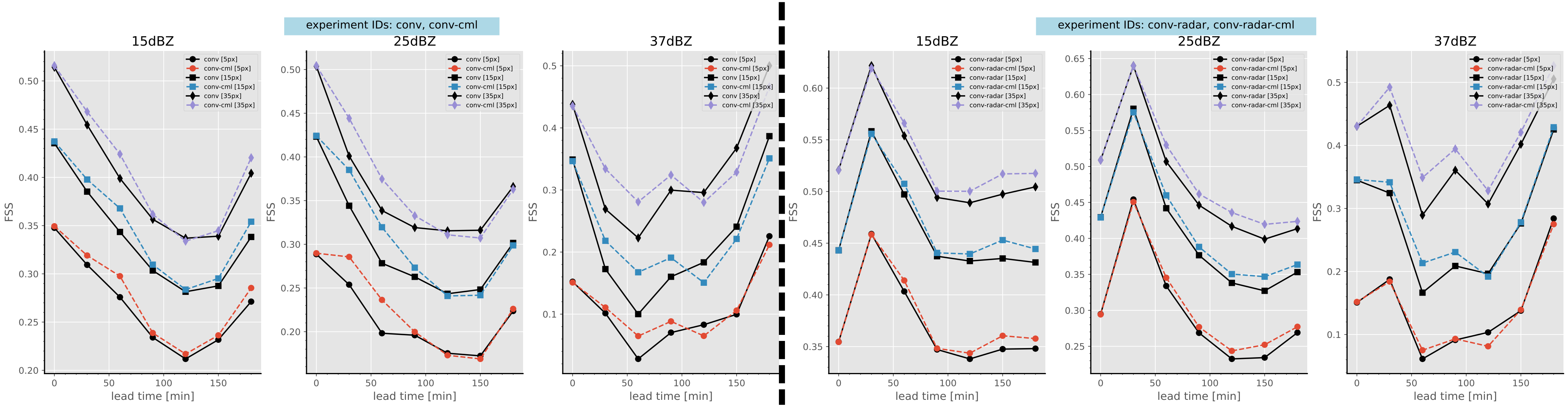
CML Case Study: Model Dynamics (REFLs)



visualization of radar REFLs (dbzcmp) at 1.5°

- accurate **initiation of convection**
- clear **positive impact of CML** DA (w.r.t. conv. DA)
- CML DA **similar to radar DA**
- interesting: conv. data seem to “block” REFL generation

CML Case Study: Fractional Skill Score (REFLs)



fractional skill score (FSS) w.r.t radar REFLs over complete domain

- CML DA consistently **improves FSS by up to about 10%**
- CML DA brings **improvement even on top of conv.+radar** DA
- however, impact of **radar DA much more pronounced** than CML

CML Summary & Outlook

- set up system for **simulating and assimilating CML** data
- case study comparing results of single time DA and subsequent model run for different configurations (**“core-more runs”**)
  - short-term REFL verification shows **accurate initiation of convection**
  - FSS for REFLs **improved by up to 10%**
  - overall, already **clear improvement** for these **non-cycled experiment**
- next steps:
  - finish **paper** on CML data assimilation
  - conduct **longer-term fully-cycled BACY** experiments and study CML impact on FSS and observation error statistics
  - general quality control, spatial thinning/superobbing, bias correction
  - further study impact of parameters like obs. error, localization, etc.

Thank you for your attention!