

# Validation and Verification

Sander Tijm

Input from: Sanchez, Amodei, Stein, Wittman, De  
Rooy, De Leeuw, Arpagaus, Bengtsson

17-05-2010



# Contents

- Purpose of validation and verification
- Validation
- Verification
  - Methods
  - Process
  - Observations
- Known deficiencies

# Purpose

- Show that model is representing weather in correct way (qualitatively)
- Show that model is better than coarse resolution models (management)
- Show that model is better than previous version/other models
- Are processes (in) correct (balance)
- .....

# Validation (1)

- Still, after few years of working alongside each other, not familiar enough with each others work
- Can lead to a duplication of efforts and sometimes to misunderstanding
- Is this a common feeling or only coming from my side?

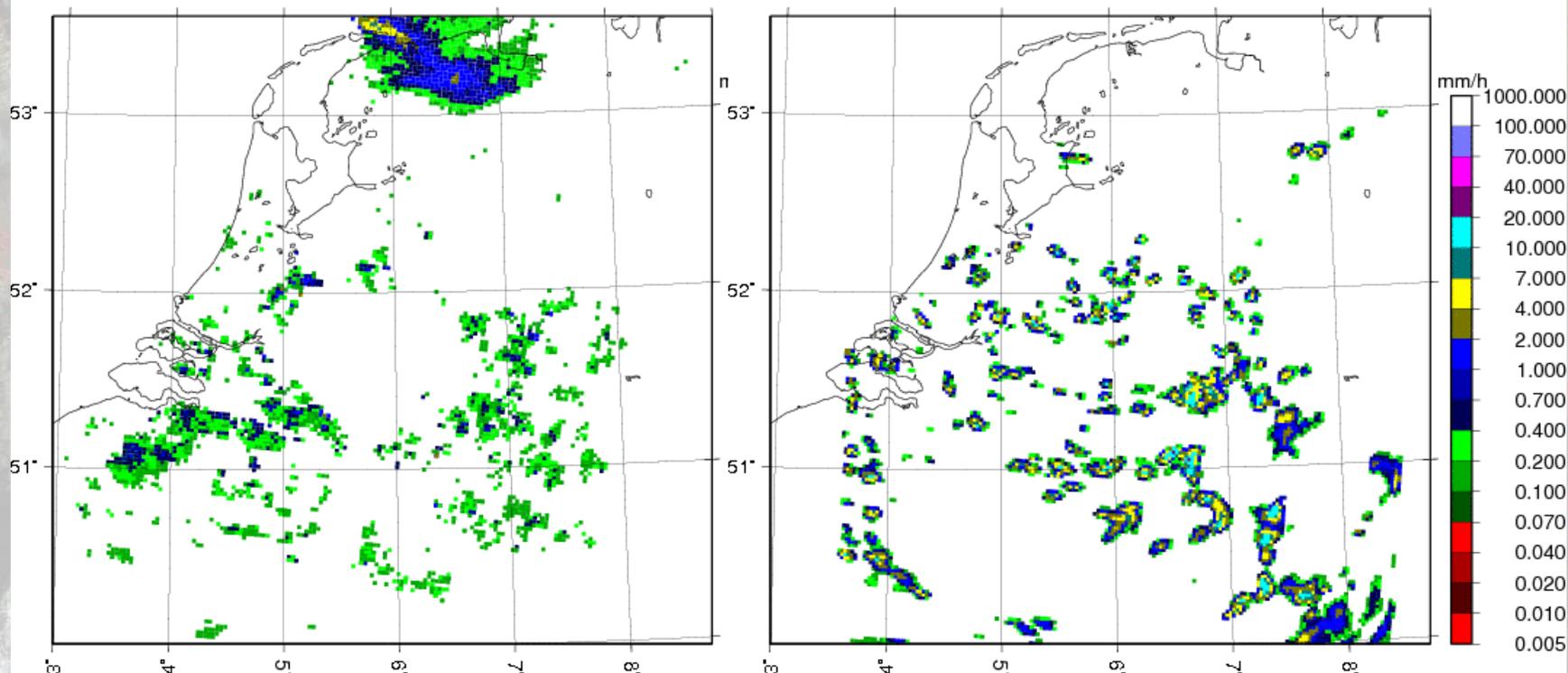
# Validation (2)

- We want to have a model working correctly at 0.5-2.5 km resolution
- How can we show that the model is doing the right thing?
- Testcases (shared, so results can be compared) should be set up and exchanged, part of systematic checks?

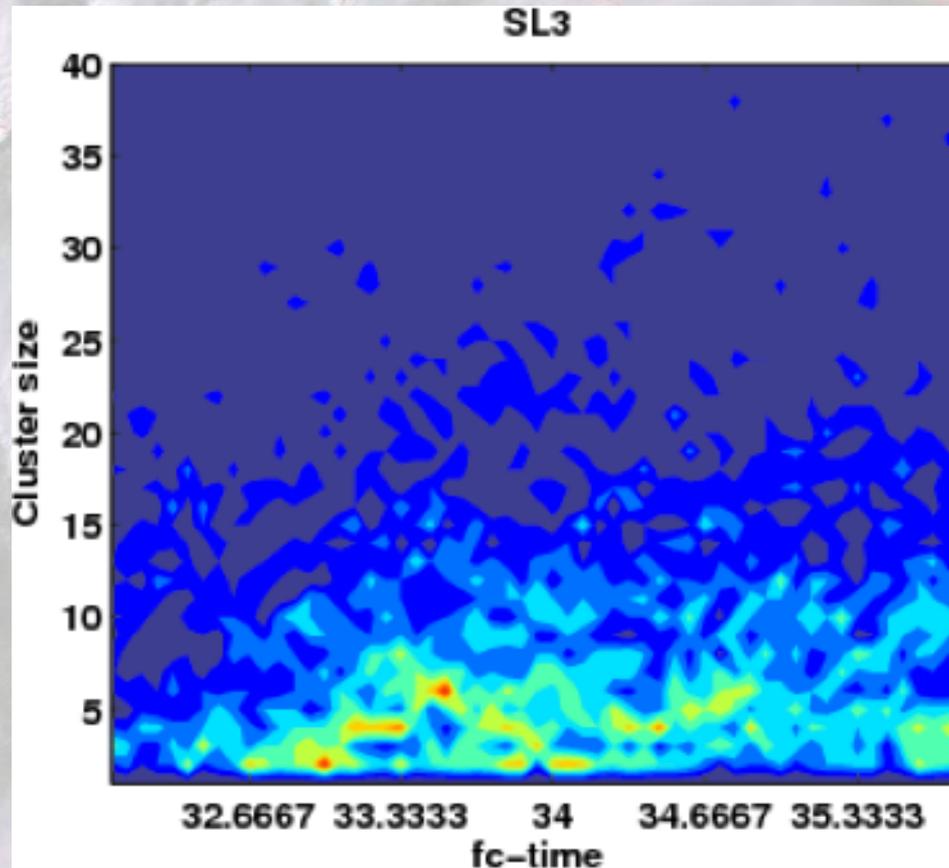
# Impact horizontal diffusion

Radar NL 20060430 1100 UTC

ARO\_2 20060430 1100 UTC



# Impact horizontal diffusion



17-05-2010



# Validation (3)

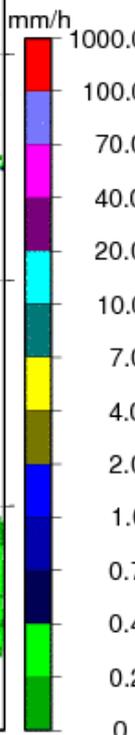
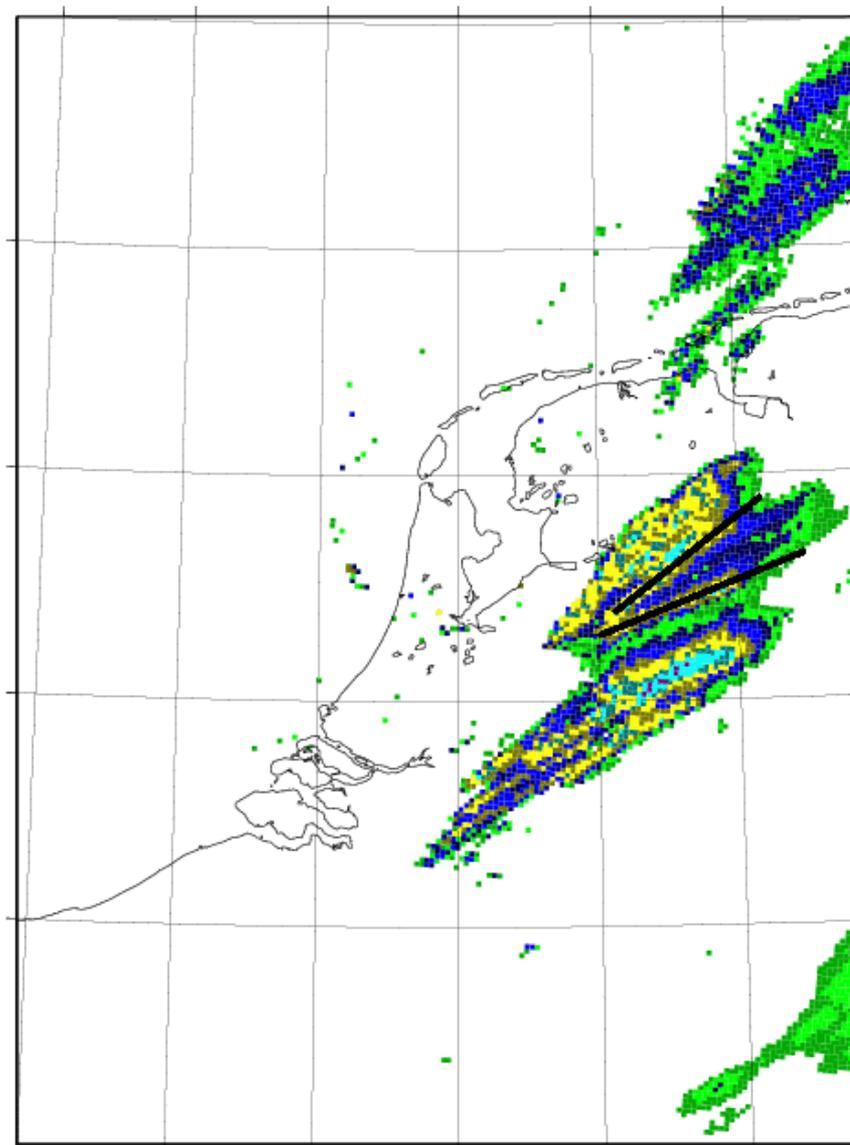
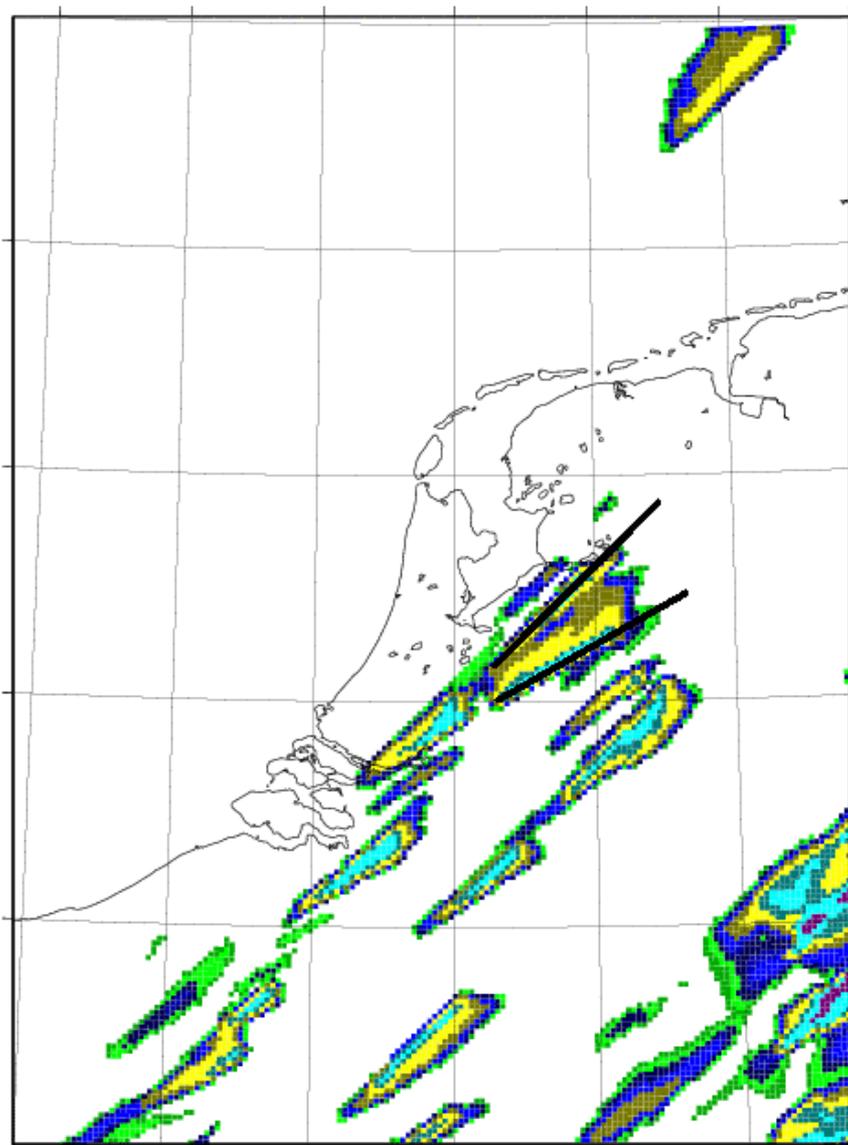
- Compare with turbulence and shallow cumulus world
- They have common cases that are always used for baseline studies
- ARM, FIRE, EUROCS, RICO ....
- Similar testset for deep convection?
- There are such cases (GEWEX working group?), we are not making enough use of them?

# Validation (4)

- How to validate that model has correct behaviour at different resolutions?
- Compare with IOP's, CRM at LES scales?
- Back to idea of validation/verification workinggroup?
- Whole range of convection from weak single cells to MCC's should be tested + more average weather!

AROME NL 20080622 13 UTC

Radar NL 20080622 13 UTC



17-05-2010

nwp central europe

# Verification (1)

- Methods
- Old scores not appropriate (double penalty etc.)
- New verification methods proposed and used a little bit within ALADIN/HIRLAM
- Are we doing enough?



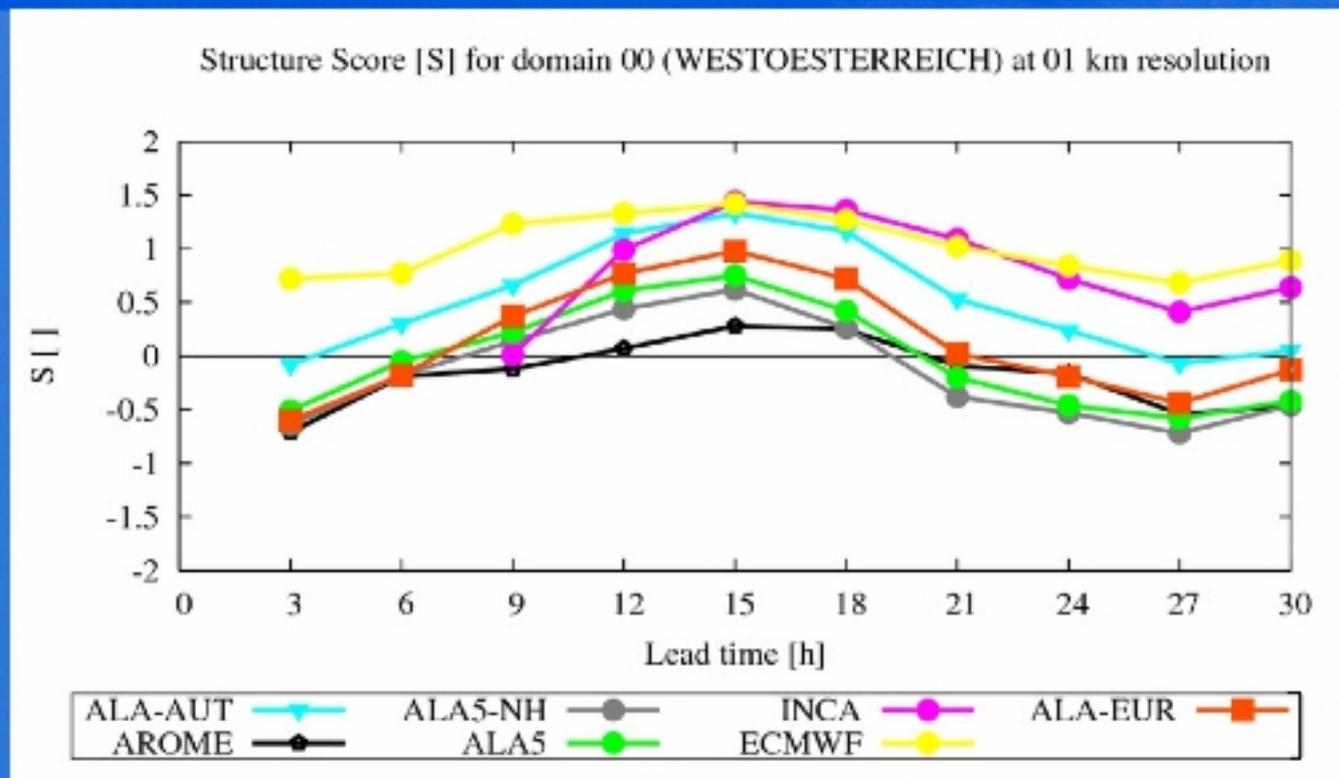
# A Fuzzy Verification Toolbox

Fuzzy method	Decision model for useful forecast
Upscaling (Zepeda-Arce et al. 2000; Weygandt et al. 2004)	Resembles obs when averaged to coarser scales
Anywhere in window (Damrath 2004), 50% coverage	Predicts event over minimum fraction of region
Fuzzy logic (Damrath 2004), Joint probability (Ebert 2002)	More correct than incorrect
Multi-event contingency table (Atger 2001)	Predicts at least one event close to observed event
Intensity-scale (Casati et al. 2004)	Lower error than random arrangement of obs
Fractions skill score (Roberts and Lean 2005)	Similar frequency of forecast and observed events
Practically perfect hindcast (Brooks et al. 1998)	Resembles forecast based on perfect knowledge of observations
Pragmatic (Theis et al. 2005)	Can distinguish events and non-events
CSRR (Germann and Zawadzki 2004)	High probability of matching observed value
Area-related RMSE (Rezacova et al. 2005)	Similar intensity distribution as observed

Ebert, E.E., 2007: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. Meteorol. Appl., submitted.  
Toolbox available at [http://www.bom.gov.au/bmrc/wefor/staff/eee/fuzzy\\_verification.zip](http://www.bom.gov.au/bmrc/wefor/staff/eee/fuzzy_verification.zip)

# Precipitation – Results for Structure

ALARO-1 Working days  
03/03/2010



- higher resolution -> better structure
- significant difference 5km NH – 5km H in Alpine domain
- L component: usable for case studies

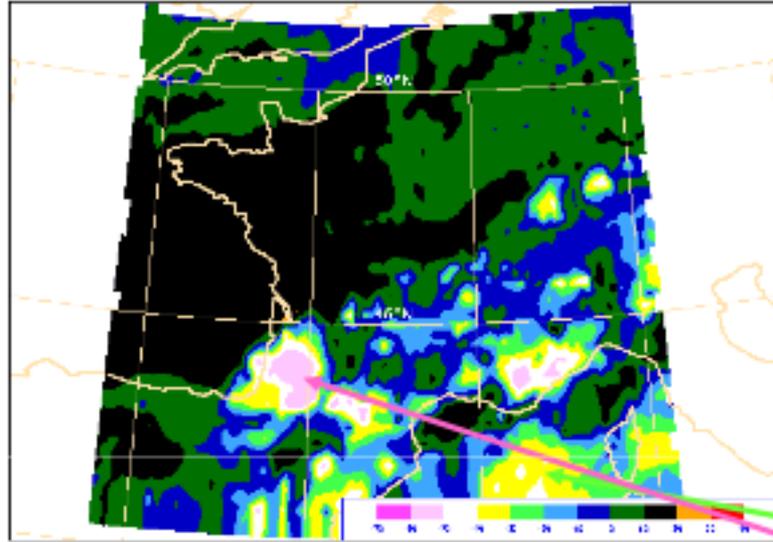
17-05-2010



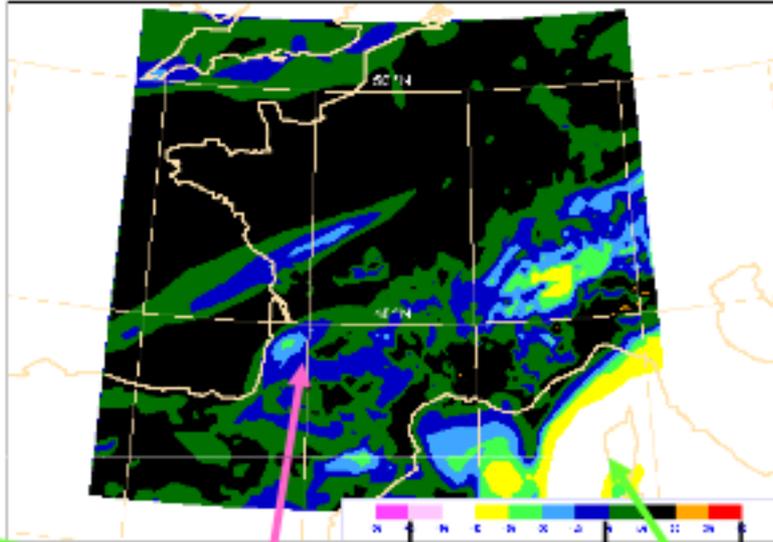
# Simulated satellite images (SSI)

11 June 2008 18 UTC

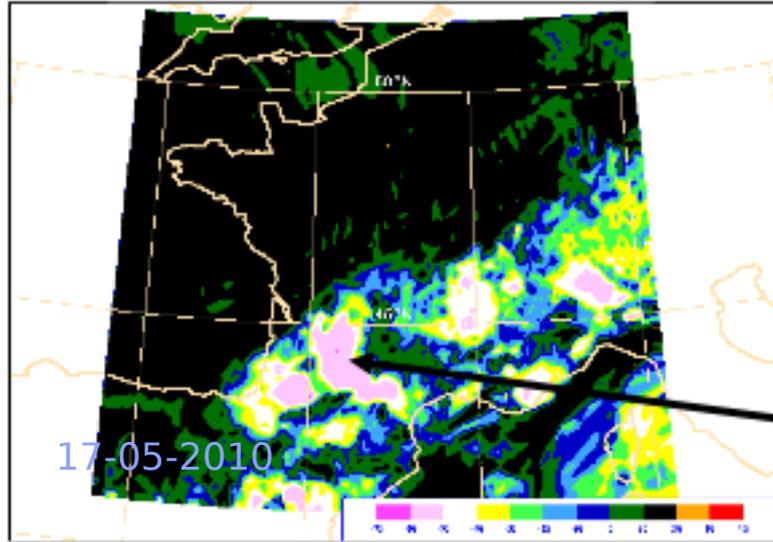
### Observation



### ALADIN



### AROME



213 K 273 K 313 K

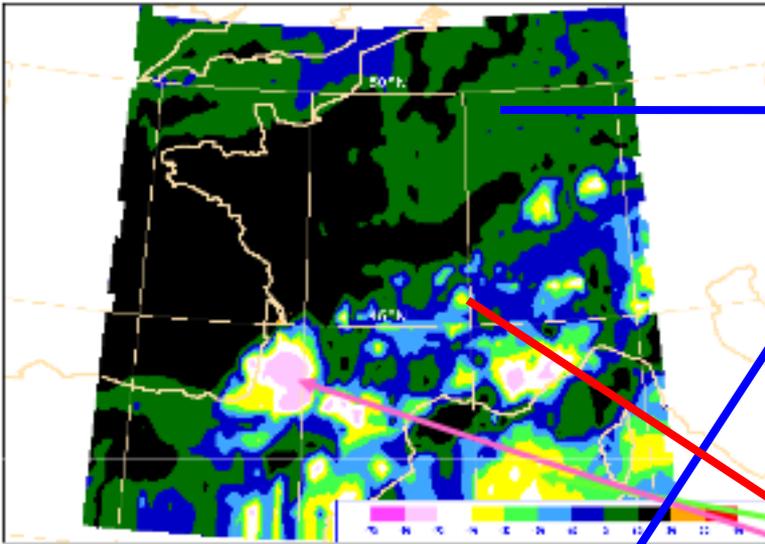
No convective subgrid clouds

Explicit clouds

Correct development of the convection

17-05-2010

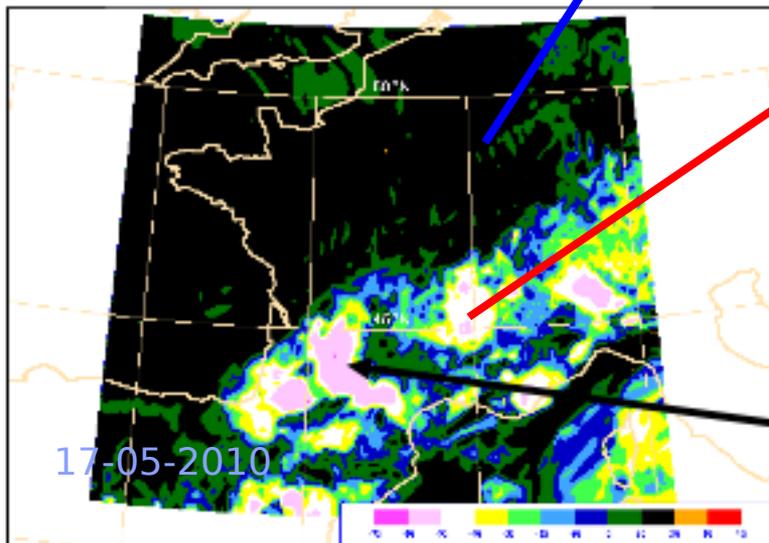
## Observation



Underestimation of low clouds

Overestimation of high clouds,  
overactive deep convection?

## AROME



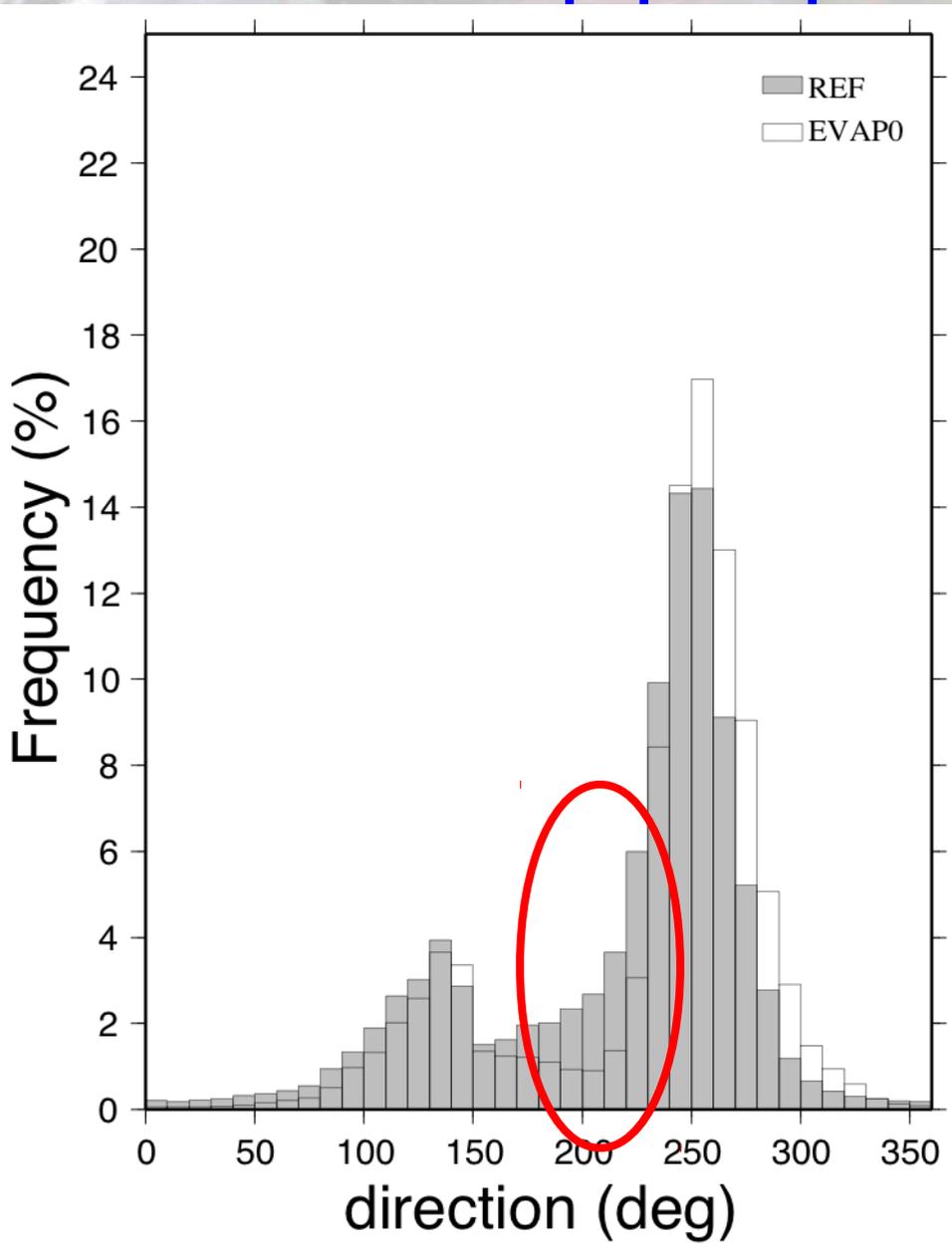
# Verification (2)

- Observations

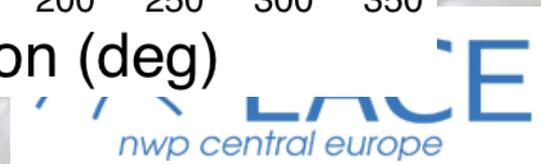
- Obvious: climatological stations, radar, satellite
- Less obvious: obs in cities, obs from cars, obs from airplanes (ModeS) .... Less in quality but higher quantity, with good monitoring of quality valuable data sources
- High time resolution synop stations in different way (distributions)
- Next generation radars for info on hydrometeors
- MSG special products like precipitation intensity and cloud water path. Quality more evenly distributed, restricted to day-light and 60(?)N

Cor

1 1 1



17-05-2010



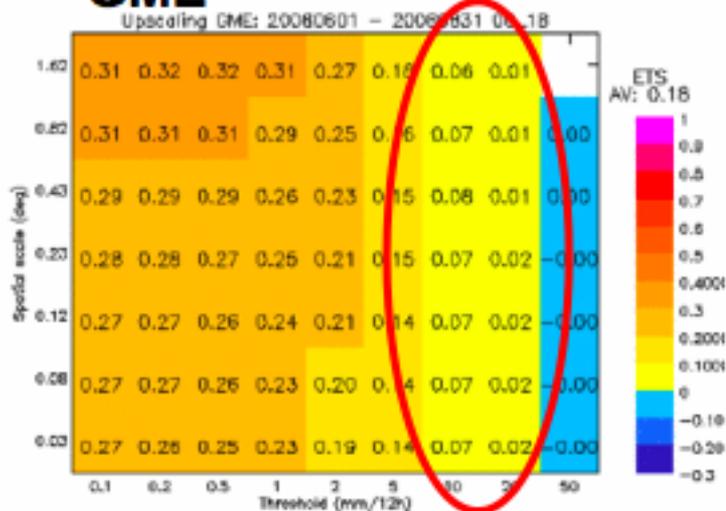
# Verification (3)

- High impact weather is focus of developments (without losing average weather out of sight)
- Verification of extremes difficult, therefore we should aim at improving the highest levels with statistically significant results?

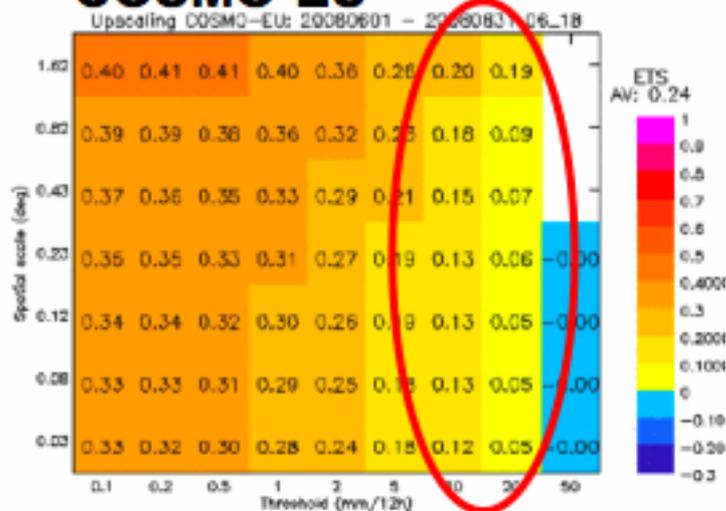


# ETS Upscaling Summer 2008

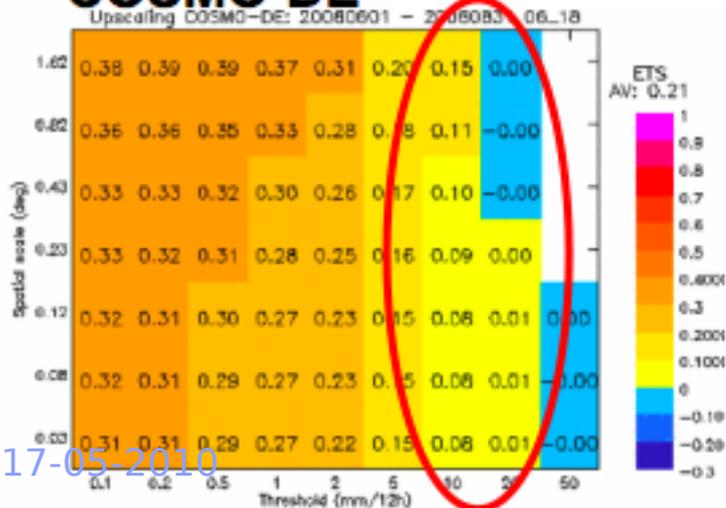
## GME



## COSMO-EU



## COSMO-DE



area: Germany

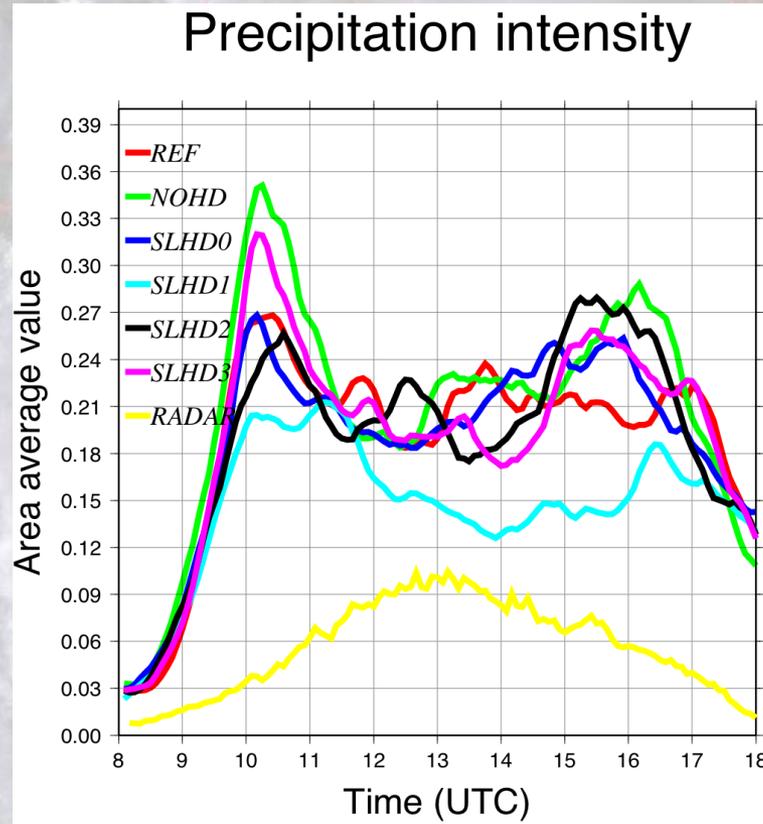
# Verification (4)

- Process
- How to arrive at a mesoscale verification toolbox
- Verification and validation working group?
- Goal: flexible, easy to use, portable mesoscale verification package (WRF-like?) existing of free software?

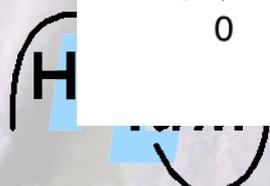
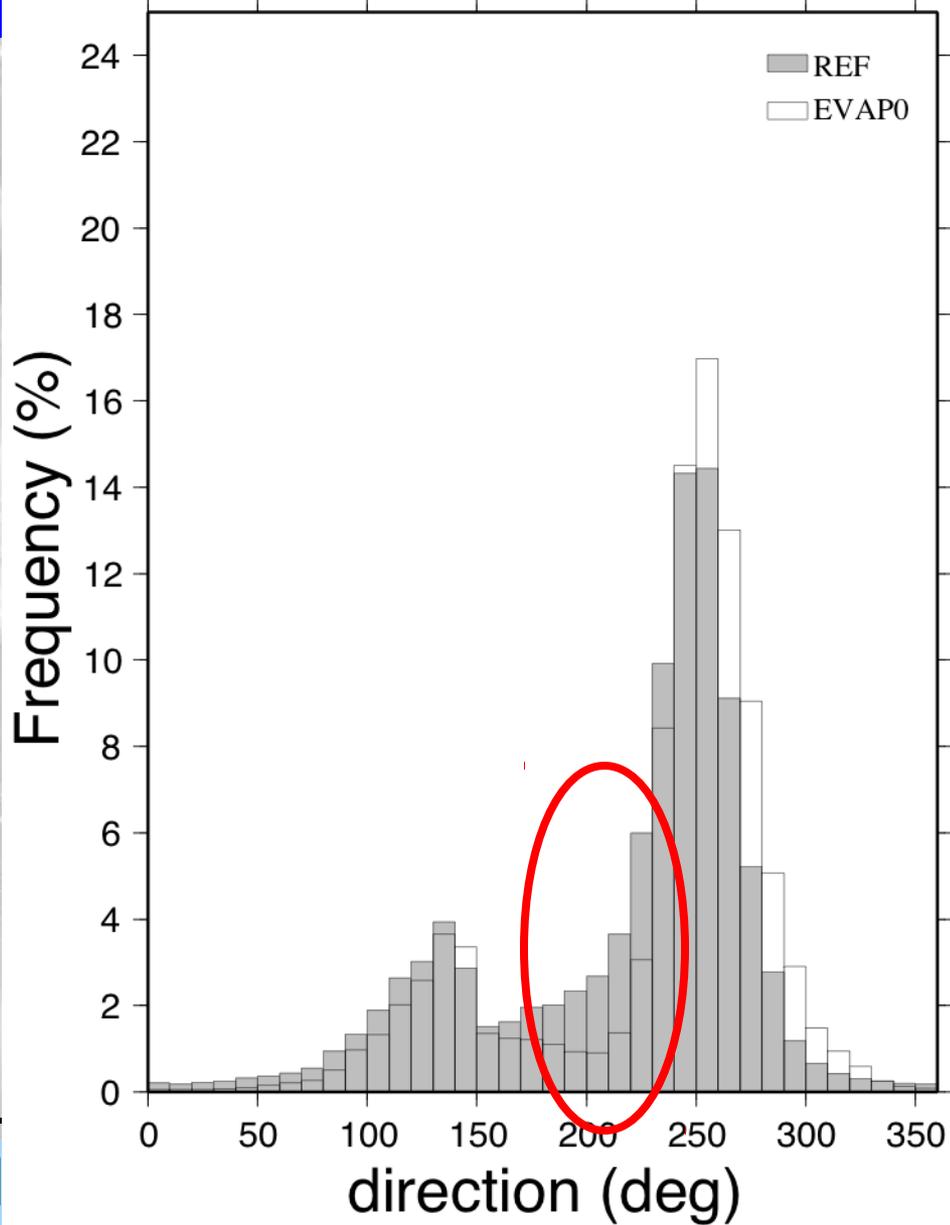
# Known deficiencies

- Daily cycle of convection
- Outflow (how to verify?)
- Low level clouds
- Strength of convection

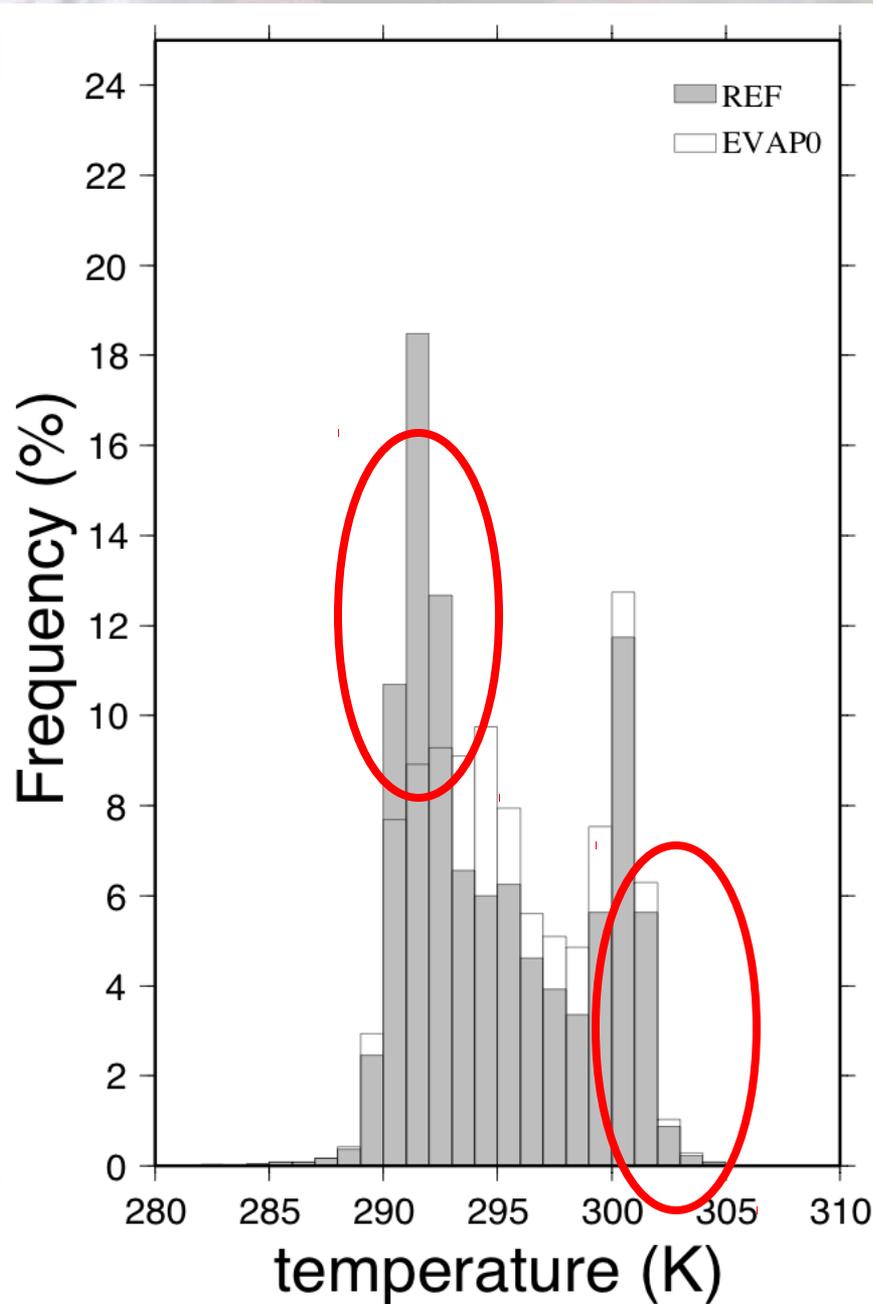
# Daily cycle of convection



# Outgoing Wind Direction



Out



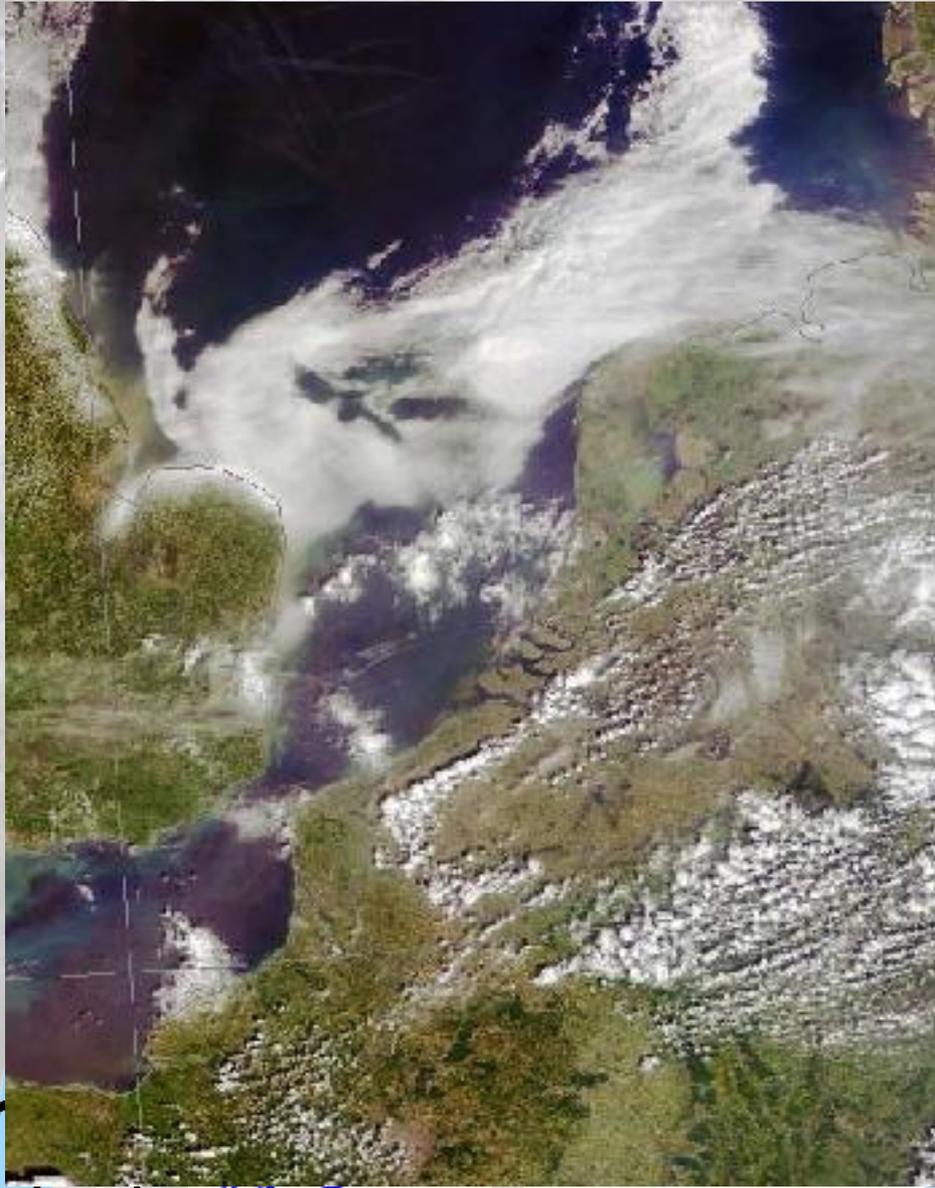
17-05-2010

Hi

ACE

nwp central europe

# Low level clouds



17-05-2010

Hir

CE

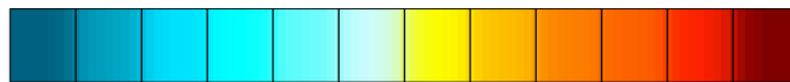
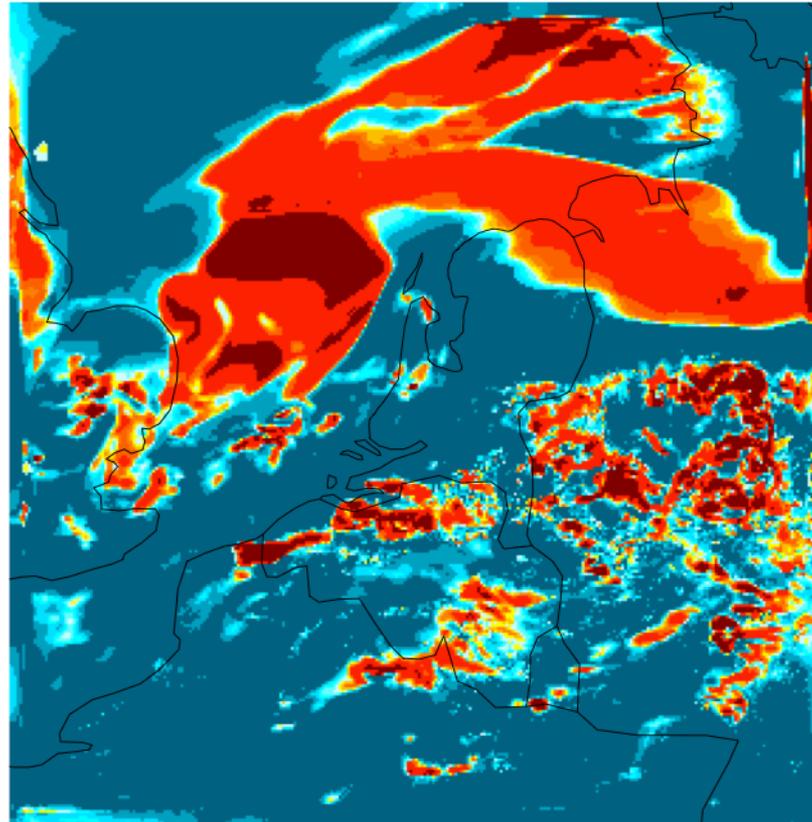
nwp central europe

25

# Low level clouds

EMKN Var002qsat VT=2008051312 (+36h)

Total cloud cover %



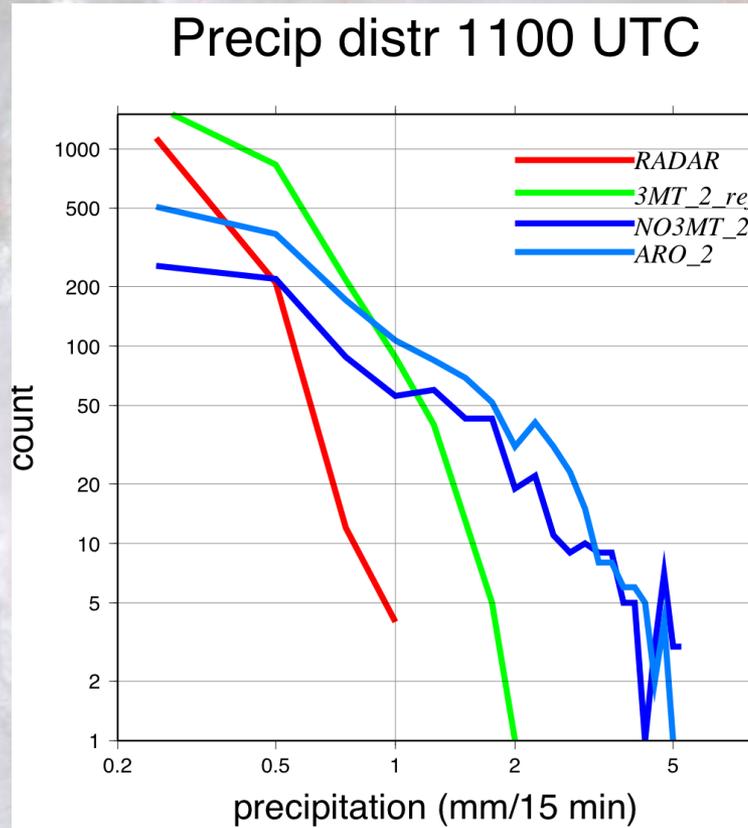
0.00 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

17-05-2010



26

# Strength convection



17-05-2010



27

# Questions

- Are we doing enough (val & ver)
- Are we learning what we want to learn
- Are we doing the right experiments (type and same relevant experiments)
- Are results shared well enough, aware of everything that is going on?

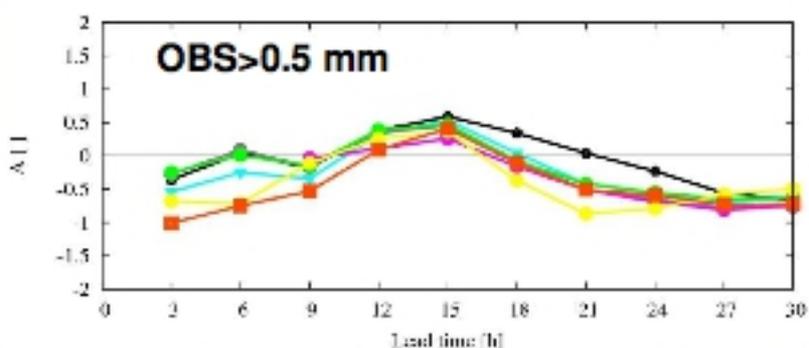
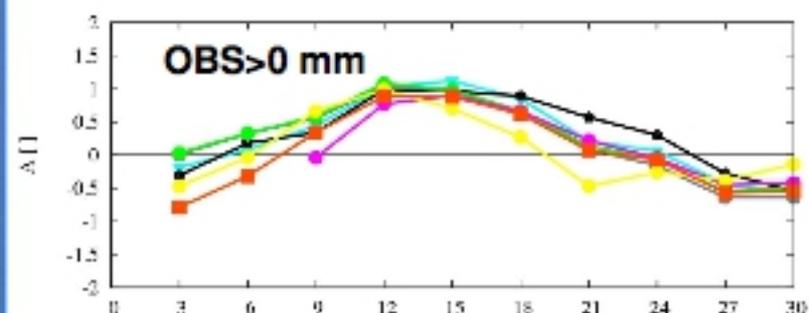
# Final remarks

- Missing: interpretation of results (suitable areas for probabilistic interpr.)
- When looking at results, do not only look at time of interest but also history, may be very different in convection permitting model from one setting to another!

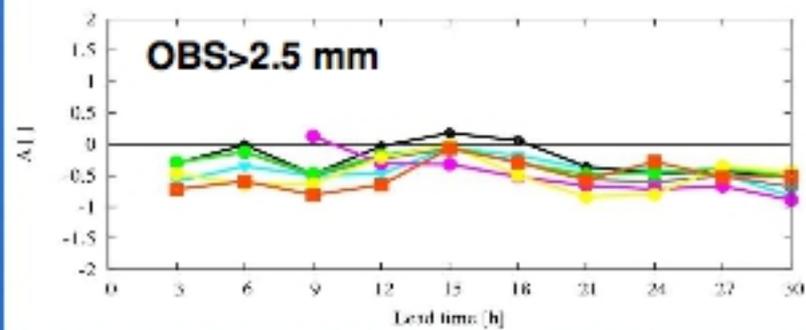
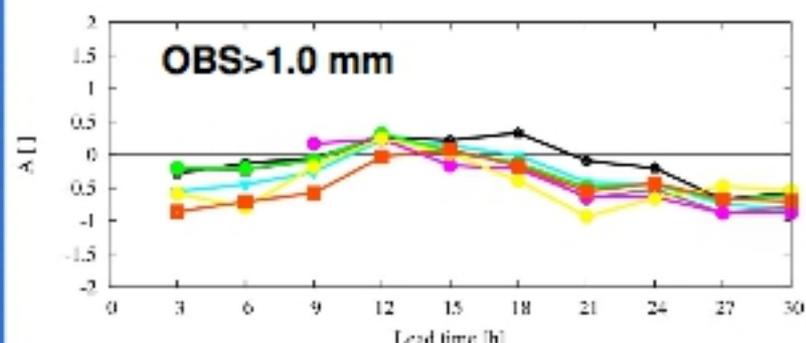
# Precipitation – Results “Alpine domain” II

ALARO-1 Working days  
03/03/2010

Amplitude Score [A] for domain 00 (WESTÖSTERREICH) at 01 km resolution



Amplitude Score [A] for domain 00 (WESTÖSTERREICH) at 01 km resolution



ALADIN-AUSTRIA (mean=0.47)	
AROME (mean=0.48)	
ALARO5-NH (mean=0.44)	
ALARO5 (mean=0.49)	
INCA06 (mean=0.41)	
ECMWF (mean=0.29)	
ALADIN-EUROPE (mean=0.29)	

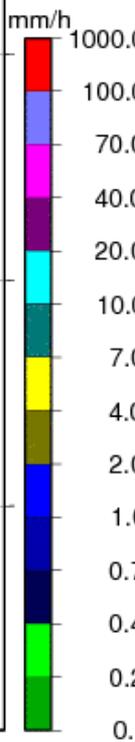
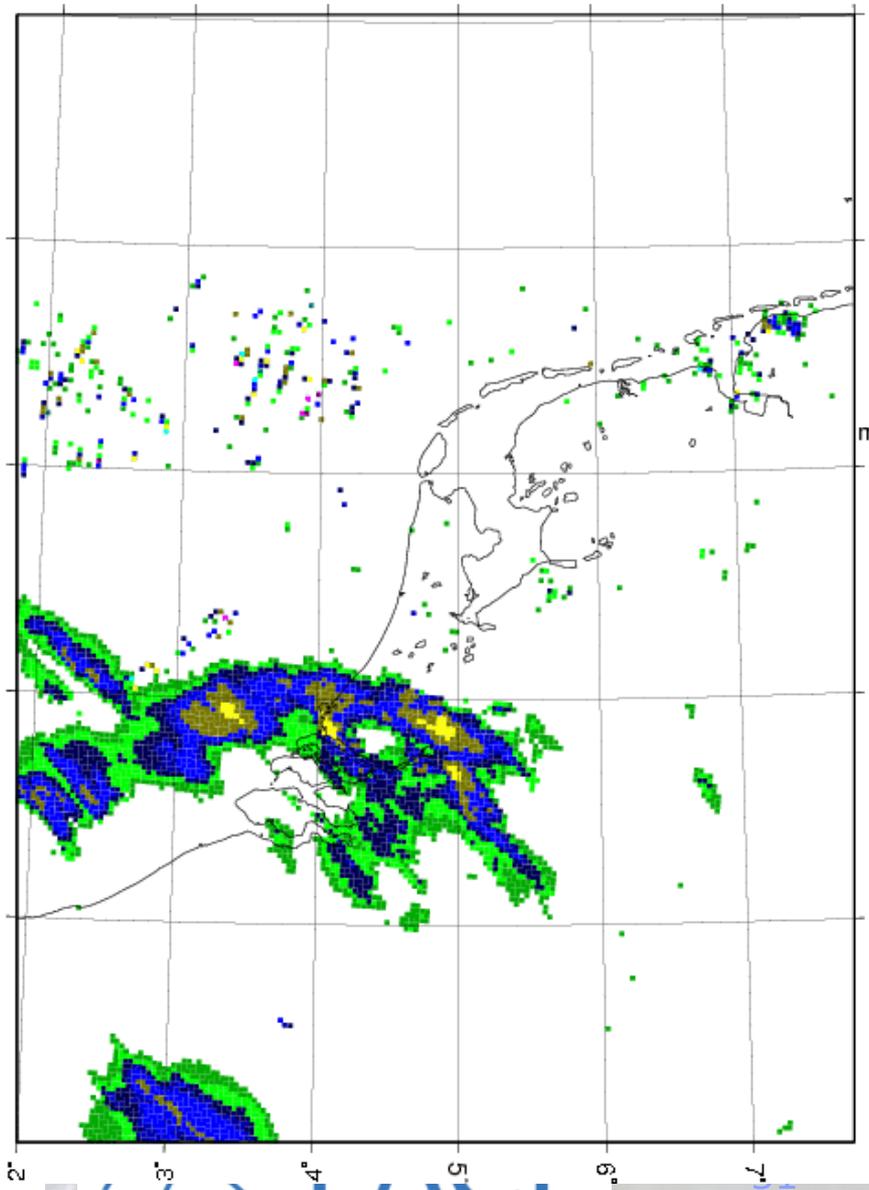
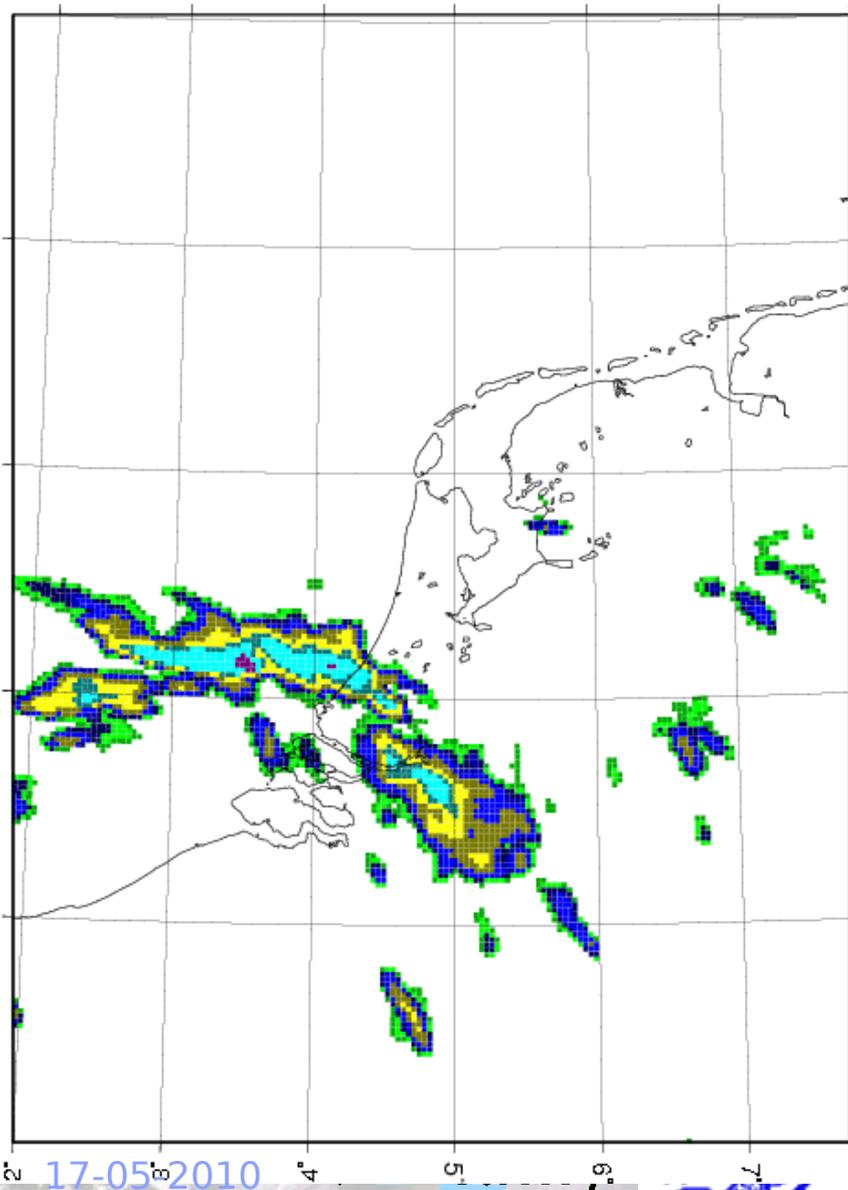
- smaller amplitude for stronger events (domain average, 20.000 grid points)
- better overall performance for high resolution models for stronger events (?)

17-05-2010



AROME NL 20090415 20 UTC

Radar NL 20090415 20 UTC



17-05-2010

nwp central europe