WWRP/WGNE Joint Working Group on Forecast Verification Research

Co-Chairs: Marion Mittermaier, UK Met Office
Laurie Wilson, Environment Canada Emeritus

Membership

Marion Mittermaier (UK Met Office) (co-chair)	Laurie Wilson (Env. Canada Emeritus) (co-chair)
Barbara Brown (NCAR)	Barbara Casati (Env Canada)
Pertti Nurmi (FMI)	Martin Göber (DWD)
Caio Coehlo (CPTEC)	Simon Mason (IRI)
Yuejian Zhu (NCEP)	Thomas Haiden (ECMWF)
Manfred Dorninger (Uni Wien)	Jing Chen (CMA)

Aims

Verification component of WWRP, in collaboration with WGNE, WCRP, CBS

- Develop and promote new verification methods
- Training on verification methodologies
- Ensure forecast verification is relevant to users
- Encourage sharing of observational data
- Promote importance of verification as a vital part of experiments
- Promote collaboration among verification scientists, model developers and forecast providers

Outline

- Outreach and Training
- Forecast Demonstration Projects
  - SWFDP
  - Sochi (FROST)
- Research priorities and examples
  - mesoVICT
  - Object methods research
- Thorpex Legacy projects
- Issues
- Future emphases

Outreach and training

- Workshops and tutorials
  - 6IVMW, Delhi, Mar 2014
  - IWTC, Dec 2014
  - Roving tutorial planned fall 2015 Indonesia
- EUMETCAL training modules
  - Available on line
- SWFDP training
  - SWFDP training Jun, Nov 2014 (E.Africa, SW Pacific)
- Verification web page
- Sharing of tools

http://www.cawcr.gov.au/projects/verification/

6IMVW India, March 2014

- 35-40 tutorial students
- 50 posters
- ~20 oral presentations
- Main “legacy” outcome: refereed special issue of MAUSAM
  - Beth Ebert guest editor
  - Expected July, 2015
  - 24 papers, mostly IMD and NCMRWF, but also from Europe and N. America.

JWGFVR

- Tutorial
  - ~35 participants
  - Lectures and exercises
  - Tools to take home
  - Group projects, presented at scientific workshop
- Scientific workshop
  - ~100 participants
  - Talks, posters, keynotes, discussions

Forecast & Research Demonstration Projects

Sydney 2000 FDP

MAP D-PHASE

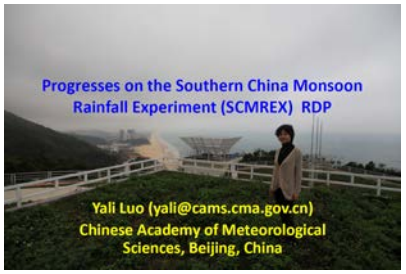
Beijing 2008 FDP/RDP

Typhoon Landfall FDP

SNOW-V10 RDP

Severe Weather FDP

FROST-14 FDP/RDP



SCMREX RDP

SWFDP verification

- Problem: Global centers give deterministic and ensemble model output to African SWFDP projects
  - NO verification for these regions at all
- Some verification done using data in 2010
- Plan to redo for all of Africa
  - GTS stations and possibly non-GTS locations specified by NMS
  - One Year
  - Use the new time series archive of TIGGE data
  - Precipitation, maybe wind.
- Training: So that the African meteorologists can do it.
- Verification funding should be built into all model R&D projects

Finnish Meteorological Institute

XXII Olympic Winter Games
Feb - Mar 2014

Forecast Verification Framework of The Sochi 2014 Winter Olympics

(FROST-2014 ~ Forecast and Research in the Olympic Sochi Testbed)

Pertti Nurmi

WMO WWRP JWGFVR aka Joint Working Group on Forecast Verification Research

Acknowledgements : FMI verification system development team & WMO FROST-2014 Expert Team

pertti.nurmi@fmi.fi

FROST-2014 : FMI Verification - Models – Contributors

Model	Analysis / FC hours	Contributor	FMI Verification	Notes
Deterministic Forecasts				
COSMO-RU - 7 km	00, 06, 12, 18 UTC / + 78 hr	HMC,Russia	***	
COSMO-RU - 2 km	00, 06, 12, 18 UTC / + 42 hr	”	***	
COSMO-RU - 1 km	-	”		to be considered as RDP
GEM - 2.5 km	23 UTC / + 27 hr	Env. Canada	***	
GEM - 1 km	21 UTC / + 25 hr	”	***	
GEM - 0.25 km	00 UTC / + 24 hr	”	***	Processing difficulties
HARMONIE-Sochi - 1 km	00, 06, 12, 18 UTC / + 36 hr	FMI, Finland	***	
NMMB - 1 km	00, 12 UTC / + 24 hr	NOAA,USA	***	
KMA	-	KMA, S-Korea		to be considered as RDP
ARPA (Sochi-mini)	00, 12 UTC / + 72 hr	ARPA SIMC, Italy	***	
INCA	-	ZAMG, Austria		
Joint (also now-casts)	Hourly / + 48 hr	“consensus”	***	
Ensemble Forecasts				
Aladin-LAEF-EPS - 11 km	-	ZAMG,Austria		
GLAMEPS - 11 km	06, 18 UTC / + 54 hr	HIRLAM, Norway	***	
GLAMEPS - updated T	Hourly / + 48 hr	”	***	
HarmonEPS - 2.5 km	06, 18 UTC / + 30 hr	”	***	to be considered as RDP
COSMO-RU-EPS - 2 km	-	HMC,Russia		to be considered as RDP
COSMO-S14-EPS - 7 km	-	ARPA SIMC, Italy		
NMMB-EPS - 7 km	-	NOAA, USA		
Nowcasting				
MeteoExpert - single station	10 minutes / + 4 hr	IRAM,Russia	***	
INCA		ZAMG,Austria		
INTW		Env.Canada		
CARDS		Env.Canada		to be considered as RDP
ABOM		Env.Canada		to be considered as RDP
RW Model (Harmonie-driven)		FMI, Finland	Verification within external CoMoSeF project	

4th FROST-2014 Meeting, Moscow, 29-31 Oct 2014

pertti.nurmi@fmi.fi

FROST-2014 : Weather variables – "Official" Thresholds

	Precipitation	New snow	Low visibility	Wind	Air temperature	Wind chill	Air humidity
Alpine Center	**> 15 mm per 6 h**	**>30 cm per 12 h**	**<20 m on the entire course**	**>17-19 m/s**			
	Type	>15 cm	<20 m on portions of the course	>11 m/s	<-25 C	<-25 C	
	Fact	>5 cm	<50 m on portions of the course	>14 m/s (gusts)			
		>2 cm per 6 h					
Russian National Sliding Centre		**>30 cm per 12 h**			**<-25 C**	**-25 C**	
		>15 cm per 6 h					>85% and Tair=Dew point
		>15 cm per 12 h		>15 m/s			
		Fact	Fact	>13 m/s	Value		<30% and T>4 C
				Crosswind			
K-125, K-95 Ski Jumping Complex				**>4 m/s**			
				variability >90°			
		>2 cm per 2 h	<20 m	>3 m/s			
				>4 m/s (gusts)			
				variability >90°			
					<-20 C	<-20 C	
	Fact				>0 C		
	Type				Sharply rise		
Cross-Country Ski and Biathlon Center					**<-20 C**		
			<50 m	>5 m/s		<-20 C	
Snowboard-Park and Freestyle-Center			<30 m				
	Fact	Fact	<250 m	>7 m/s	<-25 C	<-25 C	
Coastal cluster	Fact	Fact			Value	Fact	Value

Critical Decision point (cancel, rescheduling)
Significant decision point (rescheduling)
Factor to consider

pertti.nurmi@fmi.fi

FROST-2014 : Weather variables – "Non-official" Thresholds

Temperature (^{O}C)	T < -20	-20 ≤ T < -5	-5 ≤ T < -2	-2 ≤ T < 0	0 ≤ T < 2	2 ≤ T < 5	T ≥ 5
Wind speed (m/s)	WS ≥ 3	WS ≥ 4	WS ≥ 5	WS ≥ 7	WS ≥ 11	WS ≥ 15	WS ≥ 19
Horizontal visibility (m)	V < 100	V < 300	V < 1000	V < 10 000			
Precipitation amount	RR < 0.3	RR ≥ 0.3	RR ≥ 1.0	RR ≥ 5.0	RR ≥ 10.0	RR ≥ 15.0	
1-hr and 24-hr (mm)							

High-impact ⇔ What is hi-impact ? ⇔ Thresholds

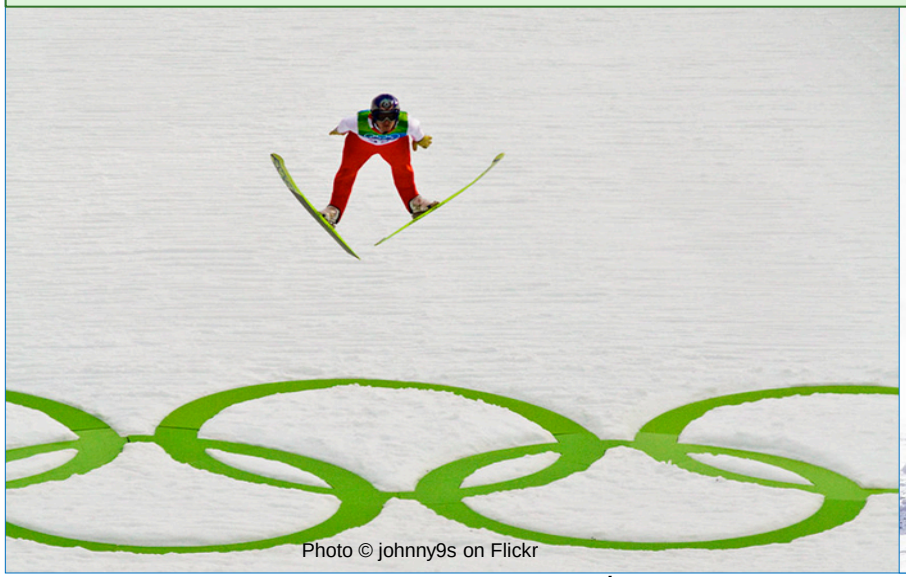


Photo © johnny9s on Flickr

Photo © Prentice Korea on Flickr

4th FROST-2014 Meeting, Moscow, 29-31 Oct 2014

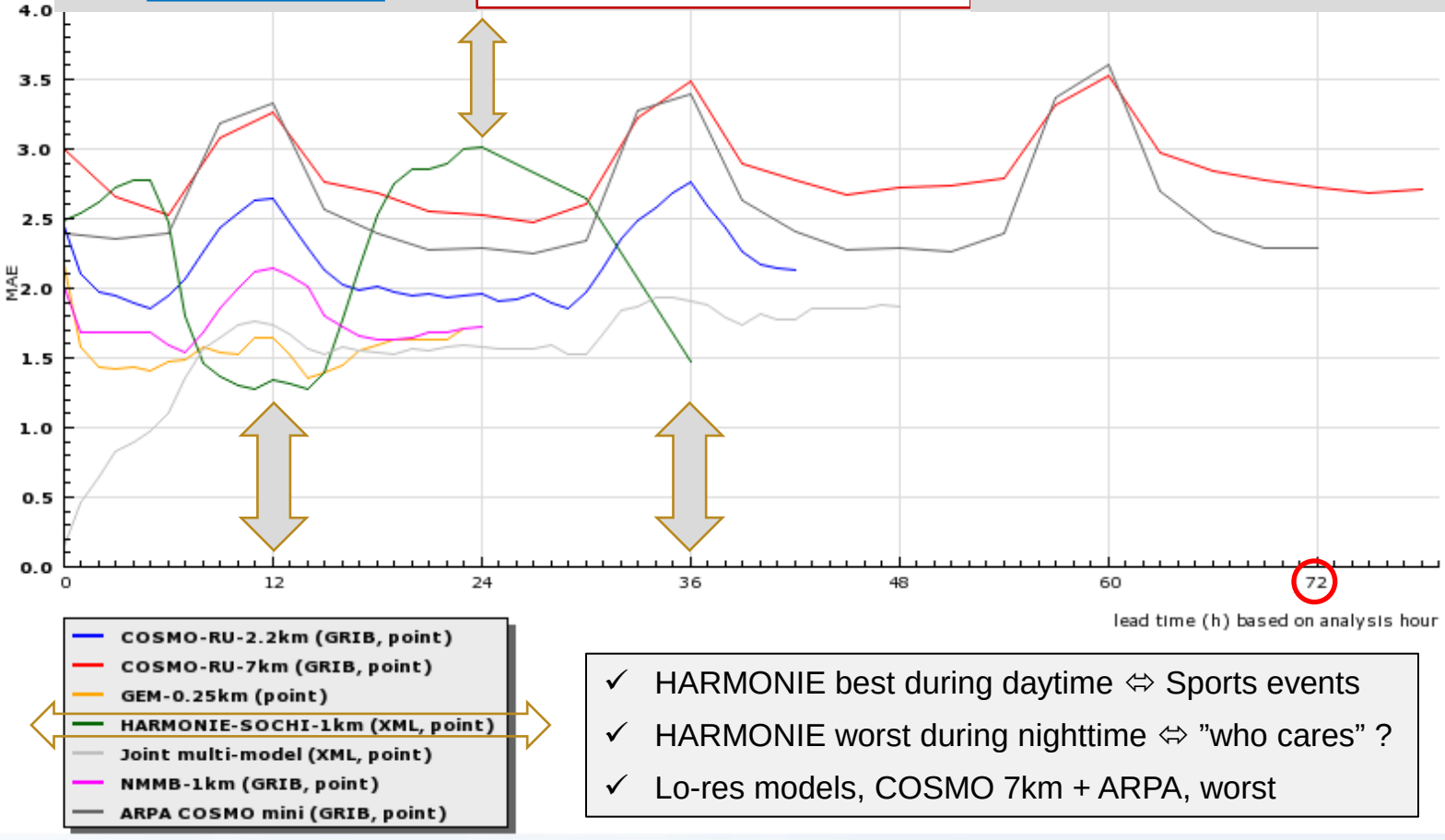
pertti.nurmi@fmi.fi

T
MAE

All stations

temperature, instantaneous
weather stations: 2014-01-15 - 2014-03-16

Determ. models



- ✓ HARMONIE best during daytime ⇔ Sports events
- ✓ HARMONIE worst during nighttime ⇔ ”who cares” ?
- ✓ Lo-res models, COSMO 7km + ARPA, worst

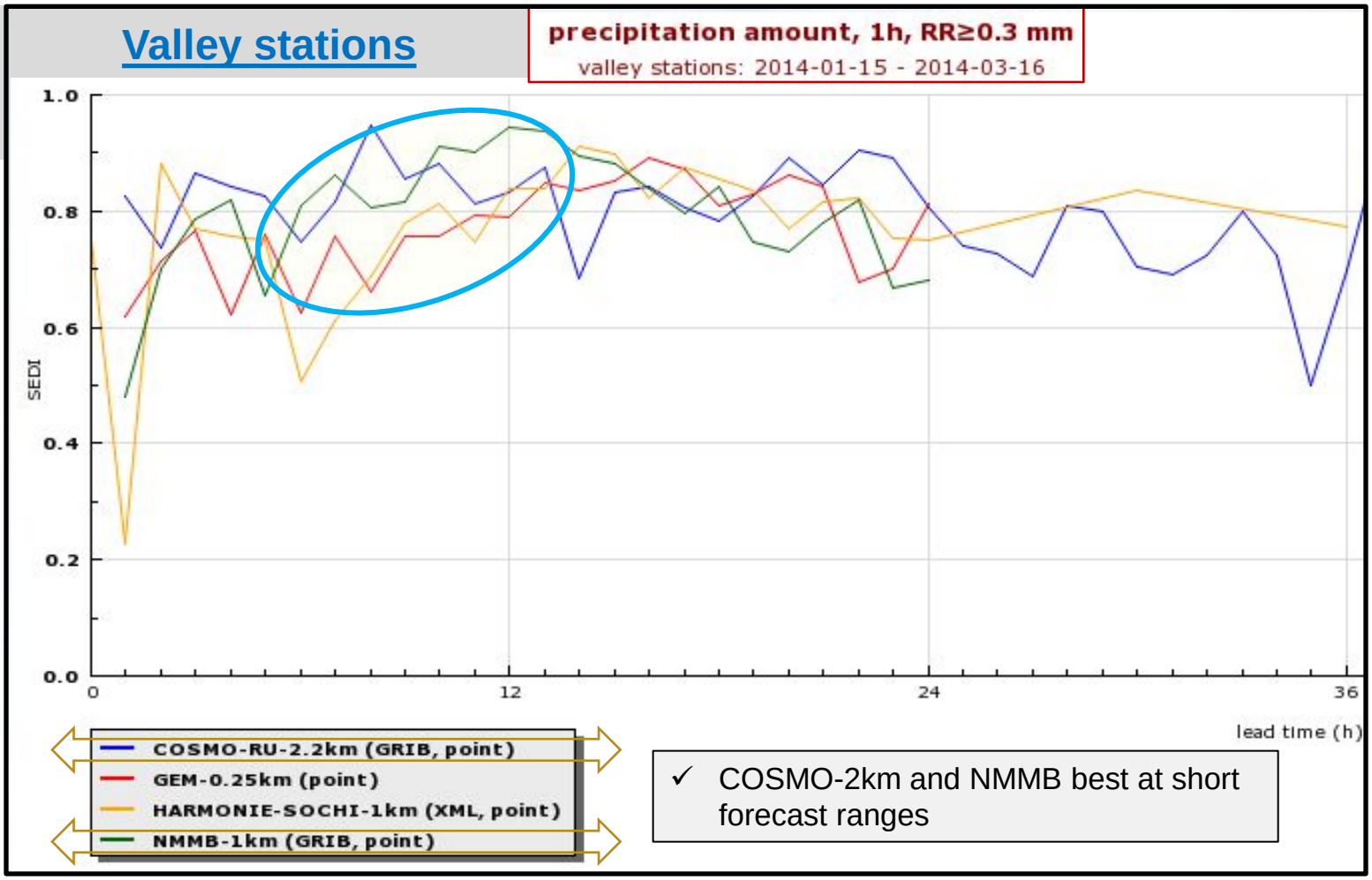
✓ 00UTC forecast runs (available for Olympics forecasters in the morning) used in verification statistics

4th FROST-2014 Meeting, Moscow, 29-31 Oct 2014

pertti.nurmi@fmi.fi

RR SEDI

Valley stations



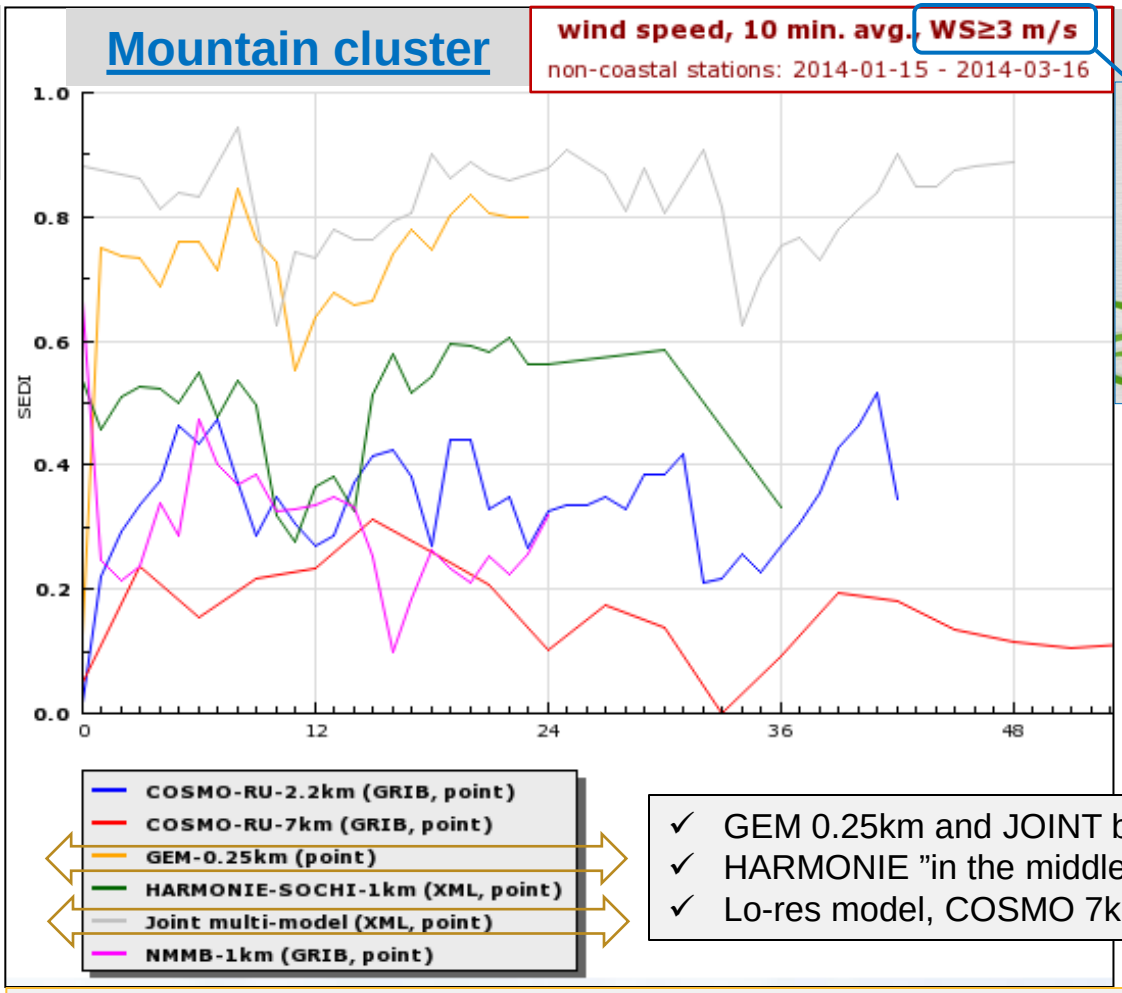
✓ COSMO-2km and NMMB best at short forecast ranges

4th FROST-2014 Meeting, Moscow, 29-31 Oct 2014

pertti.nurmi@fmi.fi

WS SEDI

Mountain cluster



Determ. models

- ✓ GEM 0.25km and JOINT best
- ✓ HARMONIE ”in the middle”
- ✓ Lo-res model, COSMO 7km, worst

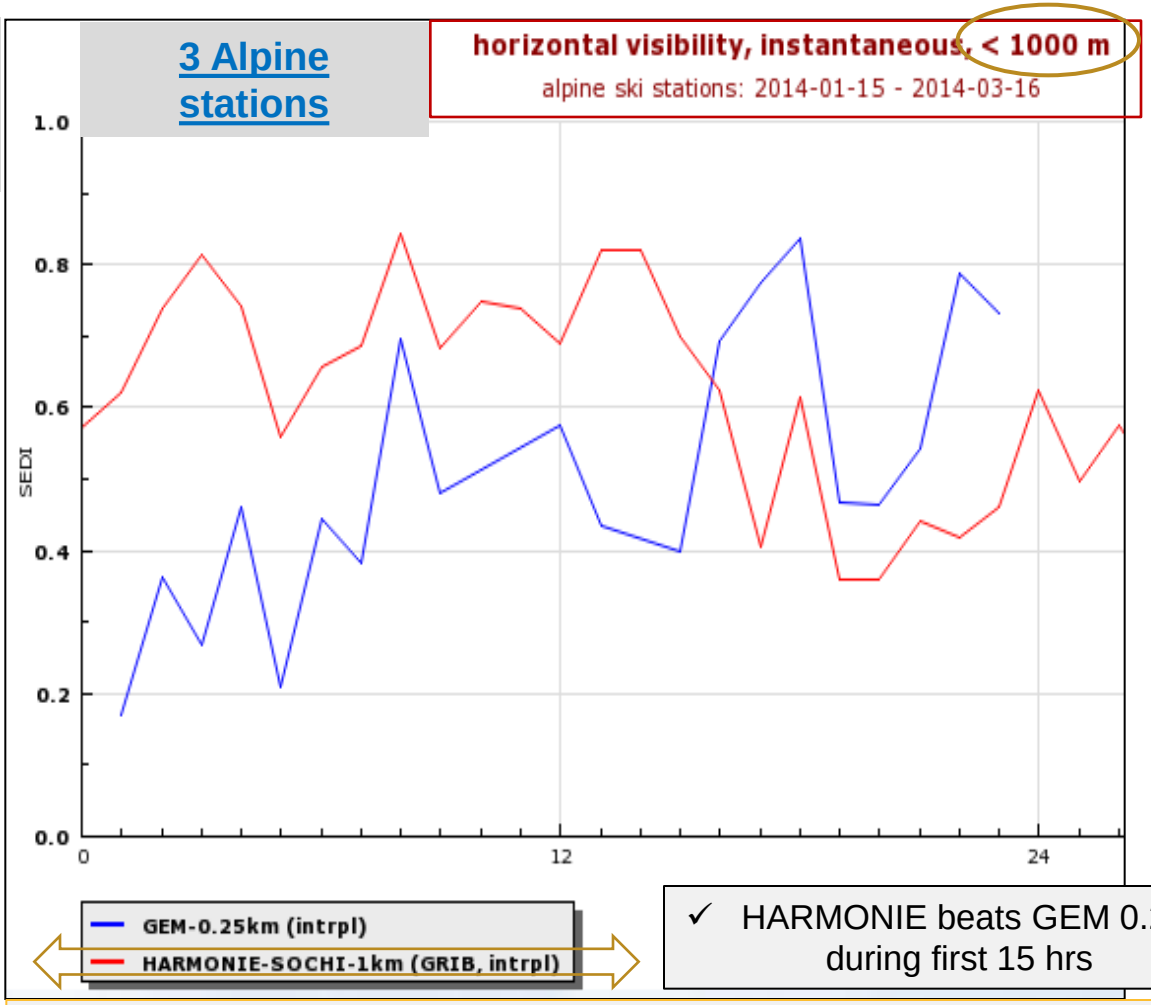
✓ 00UTC forecast runs (available for Olympics forecasters in the morning) used in verification statistics

4th FROST-2014 Meeting, Moscow, 29-31 Oct 2014

pertti.nurmi@fmi.fi

VIS SEDI

3 Alpine stations



✓ HARMONIE beats GEM 0.25km during first 15 hrs

✓ 00UTC forecast runs (available for Olympics forecasters in the morning) used in verification statistics

4th FROST-2014 Meeting, Moscow, 29-31 Oct 2014

pertti.nurmi@fmi.fi

Future actions and activities

- ✓ Quality control and checking of all observations
- ✓ Re-run of statistics after full data sets available
- ✓ Comprehensive diagnostic verification
- ✓ Compare with others' verification results
- ✓ Joint reporting and publishing with WMO FROST-2014 expert group
- ✓ Extension to societal aspects ⇔ The Impact Issue ⇔ SERA group ?
- ❖ Presentation of results:
  - ✓ FROST Fall Meeting, ECAM etc...

Verification research priorities

- High resolution NWP
  - Does the choice of verification method have any relationship to the resolution of the model being verified?
- Ensembles
- Seamless forecasts – nowcasts → short-medium range → sub-seasonal → seasonal → …
- Warnings (intensity, timing, spatial extent, etc.)
- Polar forecasts
- Urban forecasts
- Hazard impacts / user focus

MesoVICT: Ideas, plans, build-up, etc.

Project committee:
Marion Mittermaier, Met Office, UK
Eric Gilleland and Barb Brown, NCAR, USA
Manfred Dorninger, University of Vienna, Austria
Beth Ebert, Bureau of Meteorology, Australia
Laurie Wilson, Environment Canada

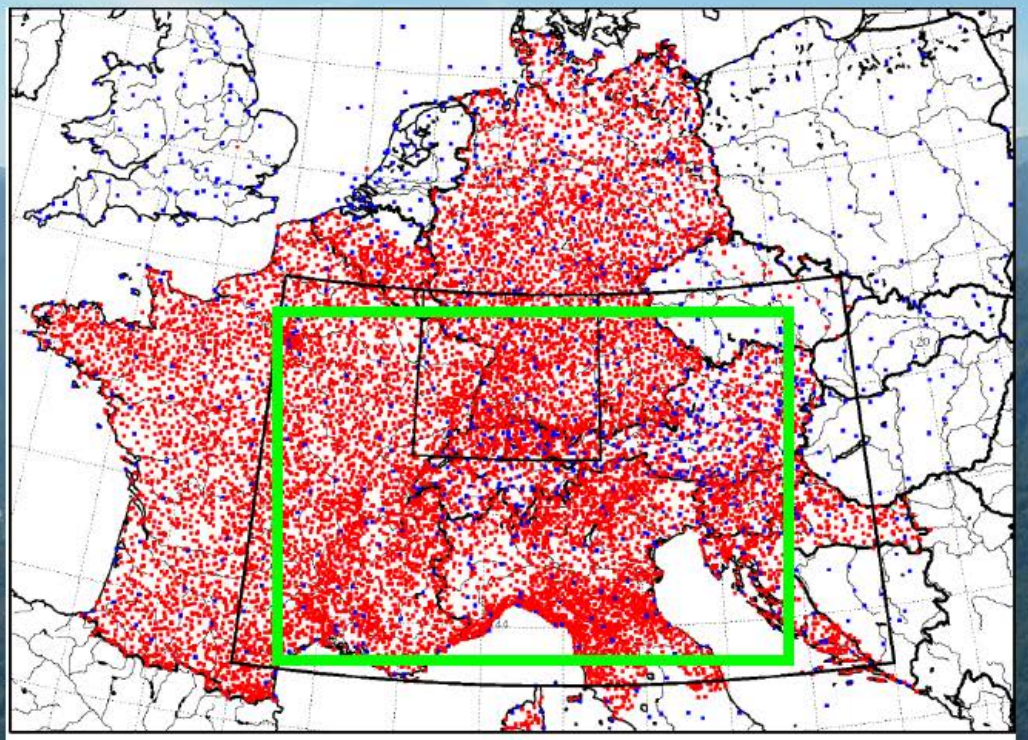
Aims of MesoVICT

How do/can spatial methods:

- Transfer to other regions with complex terrain, and other parameters: wind (speed and direction) and rain?
- Work with ensembles?
- Incorporate observations uncertainty?

Observations data set

JDC-data: WWRP D-PHASE (FDP, Rotach, et al., 2009, BAMS) and WWRP COPS (RDP, Wulfmeyer, et al., 2008, BAMS), data available: (http://cera-www.dkrz.de/WDCC/ui/Index.jsp)

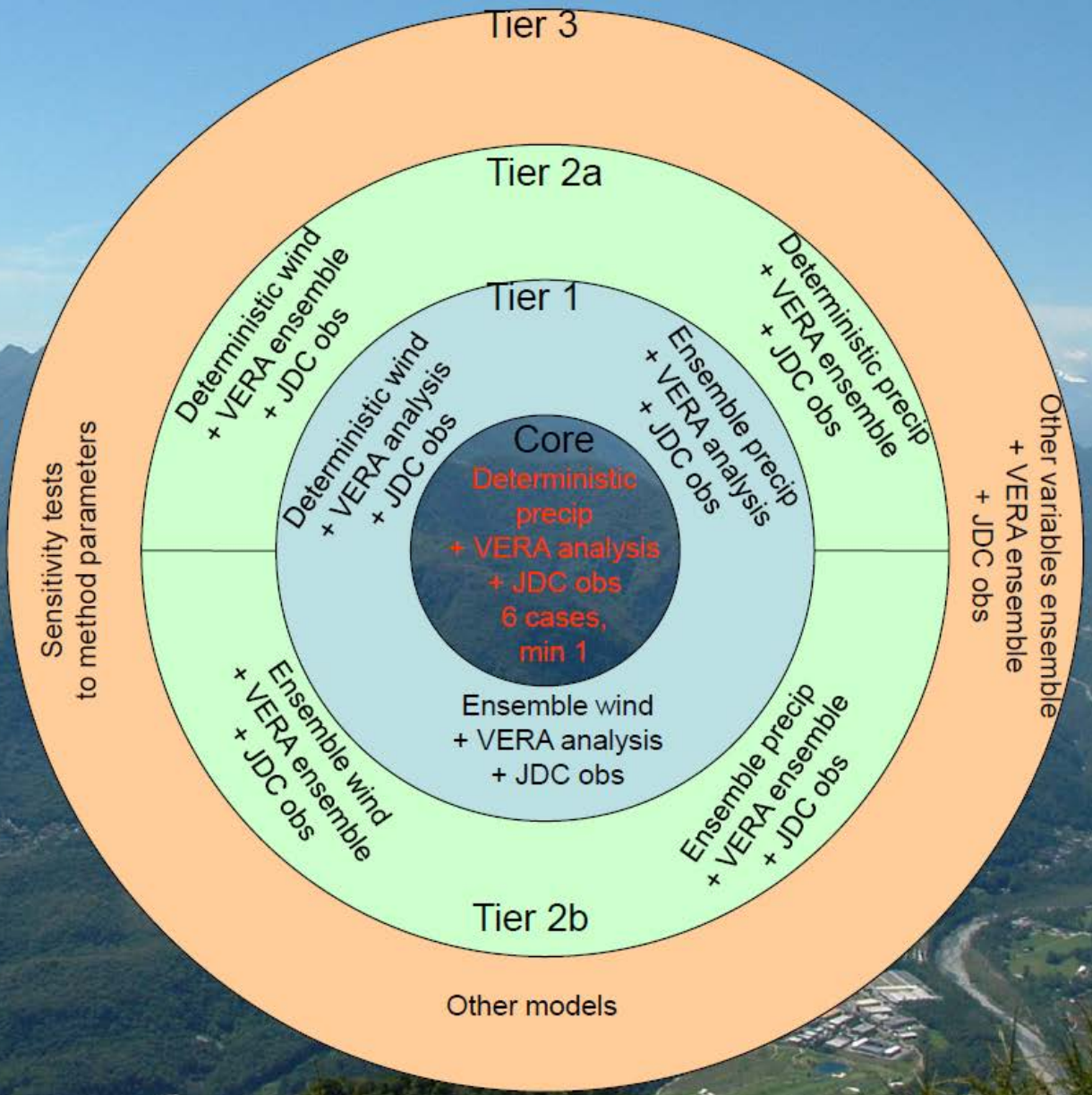


- 32 data providers
- GTS-Stations: 1232
- NGTS-Stations: > 13000
- Mean station distance: GTS: ~ 36km
  GTS+Non-GTS: ~ 12km

Frames: D-PHASE (black, large)
COPS (black, small)
this study (green)

Red: Non-GTS stations
Blue: GTS stations

Experimental design



Models

- From MAP D-PHASE COPS archive
  - Deterministic 2 km COSMO-2 Init-time:
    - Initialised 06 UTC FC-range: 24h
  - Deterministic 2 km CMC-GEM-H Init-time:
    - Initialised 06 UTC FC-range: 18h
  - Ensemble 10 km COSMO-LEPS Init-time:
    - Initialised 12 UTC FC-range:132h
- Invitation for modelling centres to produce re-runs of cases with more up-to-date model configurations (Tier 3), but core experiments to be done using COSMO-2 and CMC-GEM-H.

Outcomes

- Participants must complete the core experiment for at least case 1 to formally be classed participants.
- This requires the provision of hourly verification statistics (following the forecast evolution) for
  - Hourly precipitation (and 6h precipitation)
  - Hourly wind speed and direction
- Participation in subsequent tiers 1-3 is at the discretion of participants, but output should follow the same rules as above.

Promotion of best practice

Recommendations for the verification and intercomparison of QPFs and PQPFs from operational NWP models (2008)

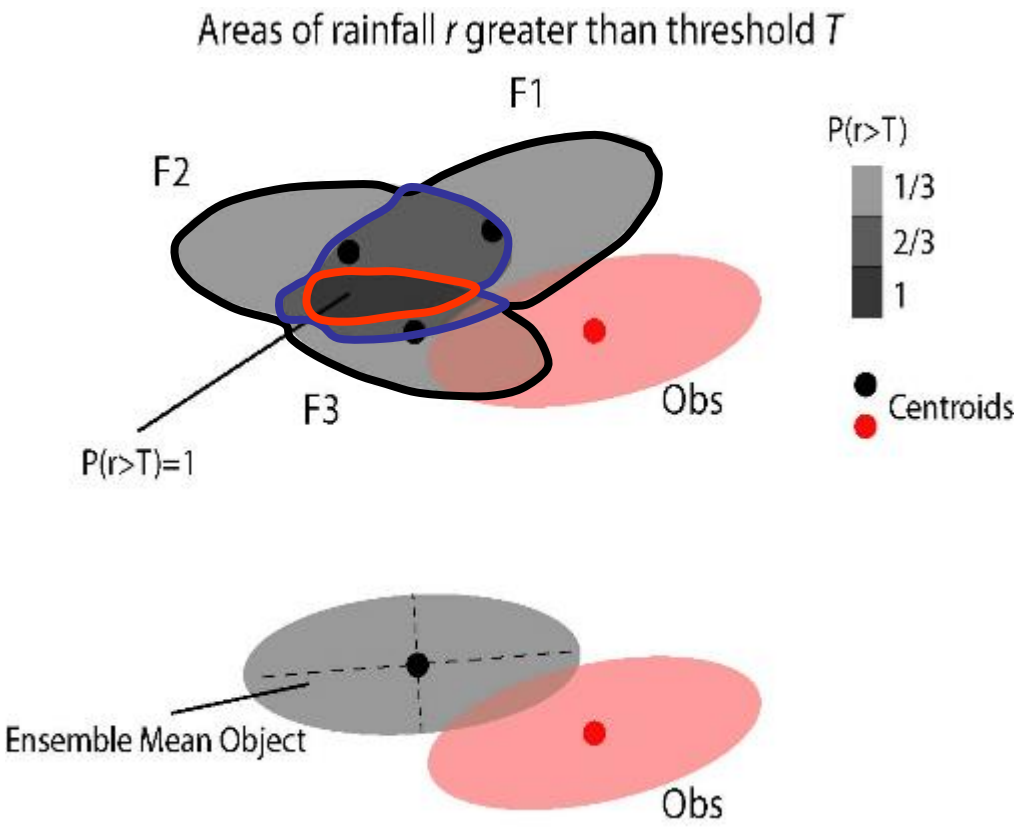
Recommended methods for evaluating cloud and related parameters (2012)

Verification methods for tropical cyclone forecasts (2013)

-edited in response to comments

Suggested methods for the verification of high resolution precipitation forecasts against high resolution limited area observations (Nov, 2013)

Applying spatial methods to ensembles



As probabilities: Areas do not have "shape" of precipitation areas; may "spread" the area

As mean:

Area is not equivalent to any of the underlying ensemble members

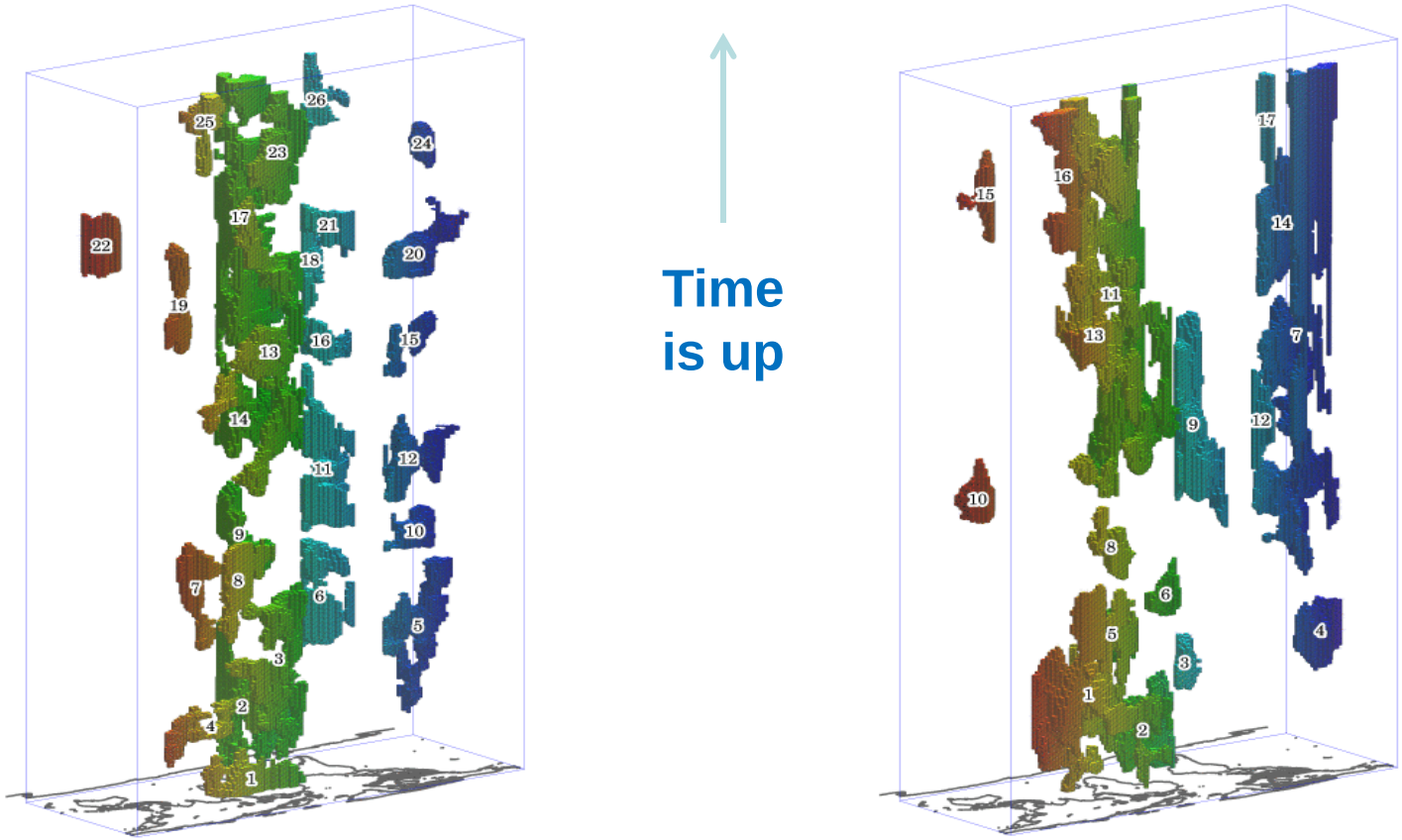
As an ensemble of attributes:

May have many interesting features

Adding the time dimension: MODE-TD

Forecast

Observed



Time is up

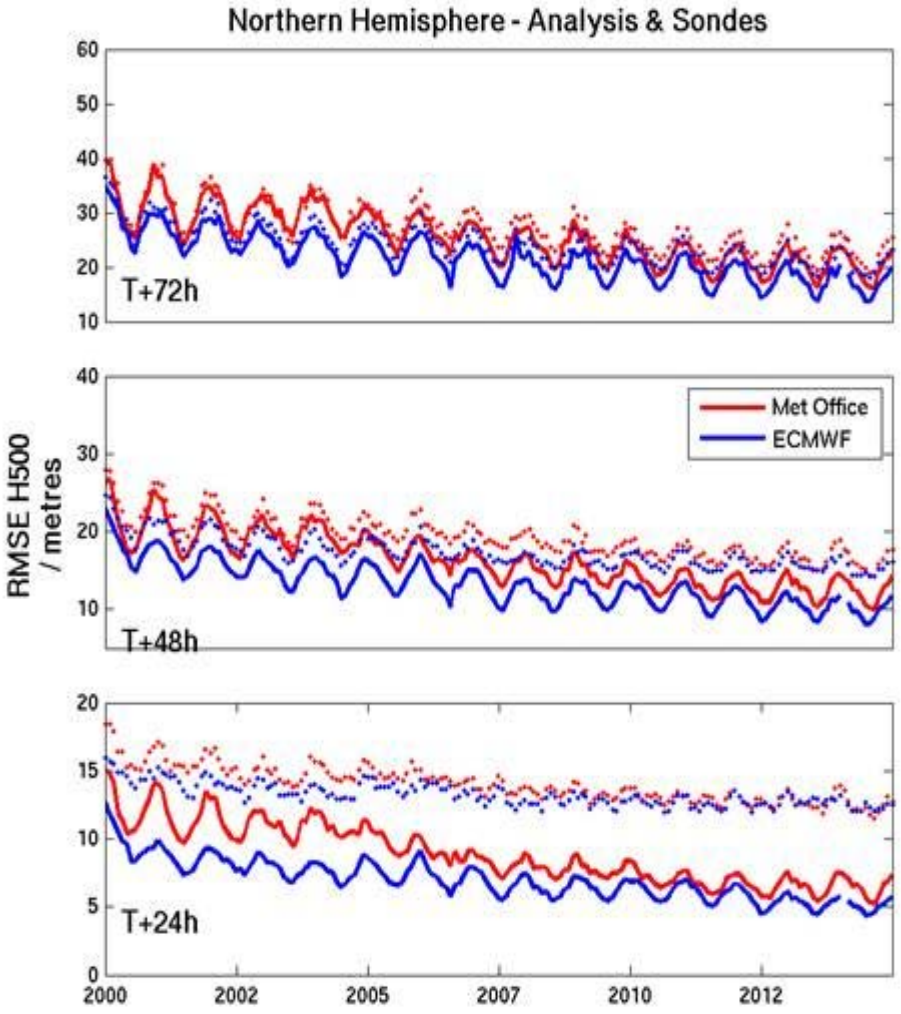
Credit: R. Bullock

Note: This example is an application to climate model output

THORPEX Legacy projects

- High Impact Weather
  - Focus of spatial methods with high resolution spatial data
  - Focus on methods for extremes (EDI, SEDI)
  - Need to consider impact verification and case studies
  - Task team identified – JWGFVR members and others
    - mesoVICT
    - Contest for new user-oriented verification metrics
    - Evaluation of global hazard map (UKMO)
- S2S
  - S2S subproject on verification led by Caio Coehlo
    - Project description on the S2S website
- PPP
  - Issues of data availability
  - Question of grid box vs point verification

Analysis vs points



Issue: Modelers cannot show improvements anymore with respect to 500 mb observations, at 24h.

Verification against the analysis:

Is the improvement with time real?

1. Model tainting of truth data
2. Grid box averaging of obs

Results less relevant to users?

From: Marion Mittermaier

Grid box vs. Point

"Metrics need to be developed to make validation relevant to the real world" (Tolman)

- "What does the model forecast at the verification location?"
- Grid Box
  - Of use to modelers, not fcst users
  - Smooth out sub grid scale, results dependent on grid resolution
    - Analysis – DA system, or BETTER – an independent analysis
    - Upscaled data – for dense observation networks – estimate grid box average
  - Model tainted (usually) – minimize by verifying only where data is
  - Easier
- Point
  - Preferred for results that is useful to users
    - Tends to be where users are, esp in polar
    - Obs error may be important
  - Model-independent (if qc is kept independent)
  - Results valid for data points only
  - Preferred for model comparisons

Next few years

- Promote verification research for high resolution NWP, ensembles, seamless, warnings, polar, urban, hazard impacts
- 7th International Verification Methods Workshop (2017)
- SWFDP verification training; New verification of HIW for Africa
- New FDPs and RDPs (e.g., 2018 Olympics, Lake Victoria, La Plata Basin, …)
- THORPEX legacy projects (PPP, S2S, HIWeather)
- mesoVICT – Focus towards verification methods for hi res models
- Verifying impacts? With SERA