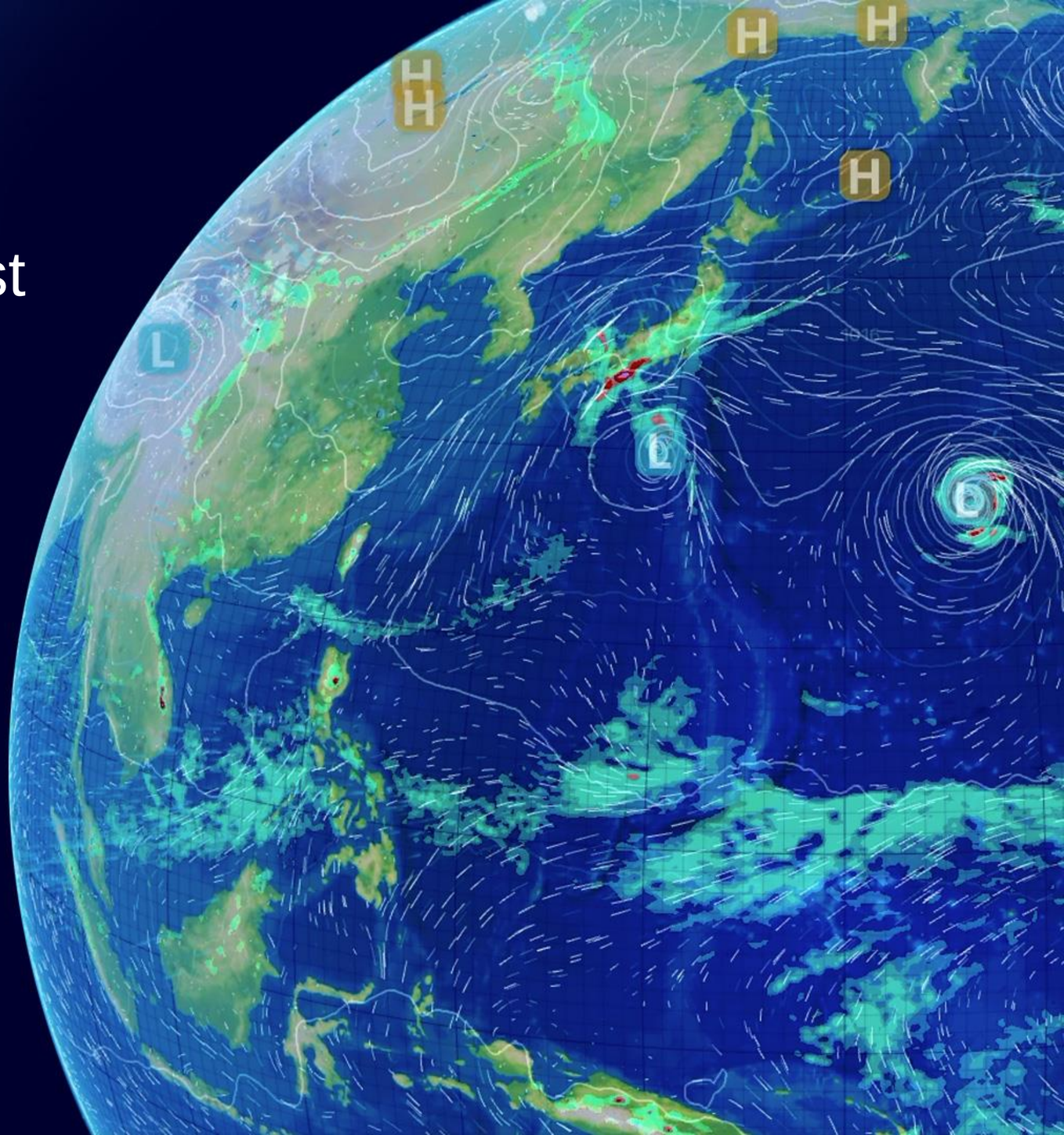
Joint Working Group for Forecast Verification Research highlights



Marion Mittermaier

Met Office

Laurie Wilson

Canada

Outline

- Overview of working group interaction with WMO projects
- Summary of Mesoscale Verification Intercomparison over Complex Terrain (MesoVICT)
- Summary of “Best new user-oriented forecast verification metric” verification challenge
- Summary of workshop and tutorial
- Working group business

WG involvement in WMO projects

- SWFDPs – no training provided in this period; plans to rewrite some of the training material to refocus on hydrology.
- Winter Olympics 2018 (PyeongChang) – WG member on SSC; verification activities minimal compared to Sochi. Exact requirements still unclear.
- UPDRAFT – WG representation on the SSC
- CBS – still looking to appoint a lead centre for the verification of tropical cyclones and implement an exchange of scores. KMA and JMA have applied. WG provided document on recommendations for the evaluation of TCs. This needs to be reworked to form the basis for CBS exchange ahead of next IWTC (2018).

WG involvement in projects 2

- SCMREX – South China Monsoon Rainfall Experiment, project started in 2012, and extended to 2021. First BAMS paper published. Working towards enhancing the verification component, which is still very deterministic.
- HIGHWAY/ L. Victoria – has been funded by UK DfID, but primarily for in-country development. Science budget small. Notional verification component. Had hoped for more.
- AvRDP – no news. List of airports was being expanded.
- TLFDP – new verification activities planned for the 3rd phase, including the development of guidance.
- Sub-group to ECMWF TAC – two new additional headline measures proposed
- SRNWP-EPS – Eumetnet funded. Aims to include verification (currently missing!)

WG involvement in WWRP

- PPP – Active contributions to the writing of the verification plan; SOP verification using super sites with lots of instruments, profiling and high temporal sampling, using a variety of methods
- S2S – contributed to the writing of the S2S book chapter, finished in May/June

  - newsletter articles

  - draft plan for the next 2 years

- HIW – evaluation task team activities including the value chain workshop

  - co-opted MesoVICT

  - verification challenge

WG documents/collaborations

- Future of the WGNE request for a document outlining the verification of high-resolution limited area models. We would like to take this forward into a pilot environment.
- Propose we involve SRNWP community to trial a framework over Europe?
- Future plans for the cloud document? Turn into a journal paper? Would this be useful to enhance the profile of WG activities?
- Would welcome a joint project (joined to CMIP6?) looking at the use of spatial methods for the validation of recent climate runs? Partners? Funding?

PPP and YOPP

- Barbara Casati leading verification effort for YOPP, task team produced a report in May titled "Verification of Environmental Prediction in Polar Regions: recommendations for the Year of Polar Prediction".
- Currently working on a second -more technical- document which outlines the YOPP verification primary goals and verification scores / methods to be applied for each YOPP model output variable
- One of the proposed (and accepted) YOPP verification activities is a verification intercomparison (subjective and objective scores) between the major centers/model, focussing on the Special Observing Periods (which are periods with enhanced observations covering the Arctic and Antarctic). for the Arctic it will be Feb-March 2018 and June-July-Aug 2018.
- A second (accepted) proposal is to perform process-based verification for high-frequency NWP outputs which will be produced for a neighbourhood of gridpoints corresponding to super-sites (where there will be high frequency observations).

JWGFVR activities contributing to S2S (C. Coelho)

- Produced forecast verification chapter for the S2S book to be published in 2019 entitled “The gap between weather and climate forecasting: Sub-seasonal to seasonal prediction”. Book Editors: Andrew W. Robertson and Frederic Vitart
Chapter authors: Caio A. S. Coelho, Barbara Brown, Laurie Wilson,
Marion Mittermaier, Barbara Casati.
- Organized S2S verification session in the 7th International Verification Methods Workshop (7IVMW, Berlin, 8-11 May 2017)
- Provided lecture on S2S forecast verification in the tutorial part of the 7IVMW, Berlin, 3-6 May 2017
- Contributed a newsletter article to the S2S project newsletter No 6 released in July 2017 summarizing the 7IVMW, Berlin, 3-11 May 2017
- Provided lectures on S2S forecast verification in the First South American School on Sub-Seasonal Predictability and Prediction, Asunción (Paraguay), 10–14 July 2017 (in Spanish)
- Contributed with a section about the S2S sub-project on verification and products to the S2S project Progress Report (Nov 2013 – July 2017)
- Contributed with preparing the S2S project Phase 2 proposal: JWGFVR activities are planned to be developed under a new sub-project entitled: Research to Operations (R2O) and S2S Forecast and Verification Products Development

“Best new user-oriented forecast verification metric” challenge (HIW activity)

- 17 entries from 11 different countries: Australia 4, USA 3, UK 2, Canada 1, China 1, Germany 1, India 1, Italy 1, Norway 1, Sweden 1 and Ukraine 1.
- Focus on the following aspects: originality, user-relevance, intuitiveness, simplicity & ease of computing, robustness and resistance to hedging. The first two criteria were considered to be the most important to this challenge.
- Judging panel consisted of 9 people, with a number of judges representing the user community. Each judge rated the entries independently using rules defined before the challenge began.

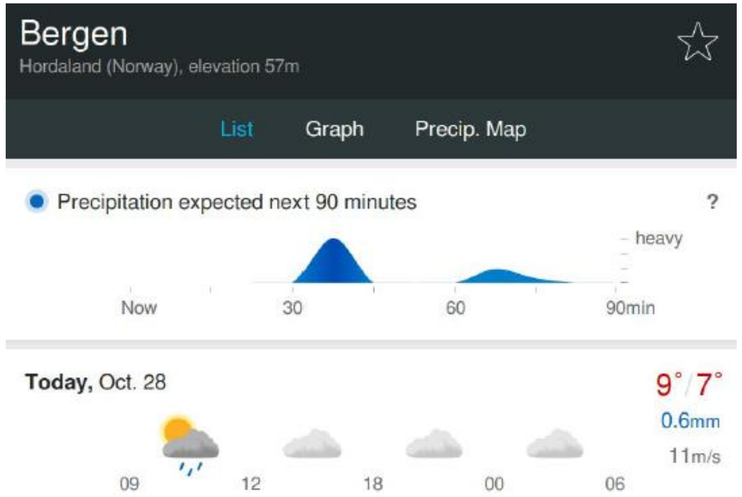
“Best new user-oriented forecast verification metric” challenge

- **Winner: Helge Goessling** from the Alfred Wegener Institute in Germany with his entry “Integrated Ice Edge Error (IIEE) & Spatial Probability Score (SPS)”. The prize was an invited talk at 7th international verification Methods workshop 8-11 May 2017 in Berlin.
- **Runner-up: Dominique Brunet** from Environment Climate Change Canada with “A spatio-temporal user-centric distance”.
- **Joint third: Thomas Nipen and Ivar Seierstad** from the Norwegian Meteorological Institute on the “Rain-free window accuracy”, and **William Wang, Andrew Watkins & David Jones** from the Bureau of Meteorology in Australia, with their “Weighted Percent Consistence” applied to their 3-month seasonal outlooks.
- The top four entries covered four very different user-communities, from marine transport in Arctic waters, to strategic decision makers for e.g. water management, aviation and the public. Other entries presented metrics and applications of interest to users of tropical cyclone forecasts, seasonal hydrological outlooks and ocean wave height.

Equal 3rd place

Rain-free window accuracy

- Thomas Nipen & Ivar Seierstad, Norwegian Meteorological Institute
- Threat score for exceedance of different window lengths



Weighted percent consistence

- William Wang, Andrew Watkins & David Jones, Australian Bureau of Meteorology

$$WPC = \frac{\sum (H + CN) * |anomaly|}{\sum |anomaly|}$$

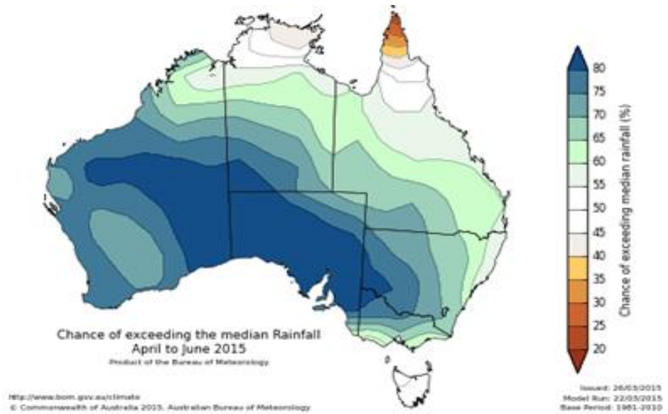


Figure 4 - POAMA model seasonal above median precipitation probability forecast for 2015 April-June.

Courtesy Beth Ebert

2nd place

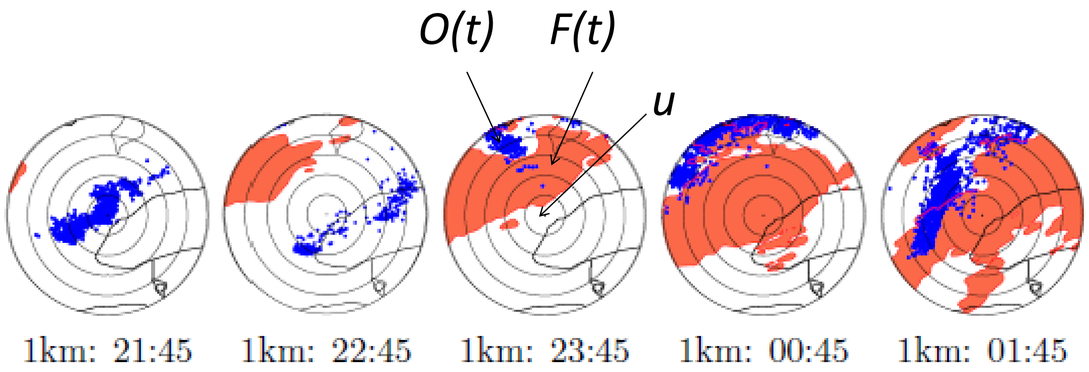
Spatio-temporal user-centric distance

- Dominique Brunet, Environment & Climate Change Canada
- Generalized distance transform

$$D_q(u,A) = \left(\frac{1}{|A|} \sum_{i:a_i \in A} d^q(u,a_i) \right)^{1/q}$$

u = location of user
A = observed or forecast events
q = parameter

- Compare $D_q(u,O(t))$ and $D_q(u,F(t))$
- ΔT that minimizes $|D_q(u,O(t)) - D_q(u,F(t-\Delta T))|$ is the timing error

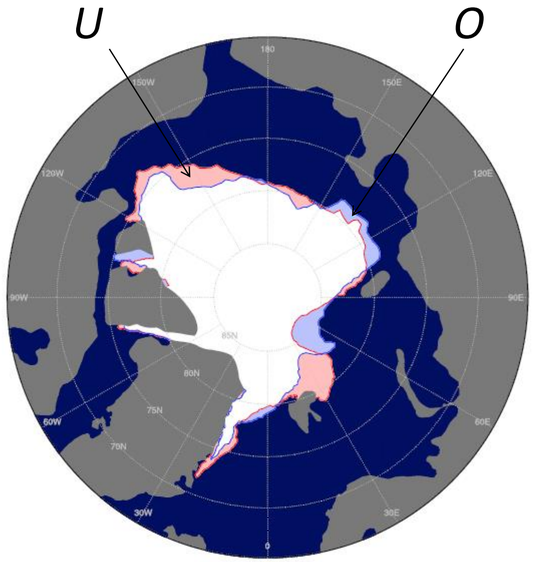


Courtesy Beth Ebert

1st place

Integrated ice edge error (IEEE) and spatial probability score (SPS)

- Helge Goessling, Alfred Wegener Institute
- IIEE = *O* + *U* = area where forecast and “truth” disagree on ice concentration being above or below 15%



- IIEE is a special case of spatial probability score

$$SPS = \int_V (P_f - P_o)^2 \, dV$$

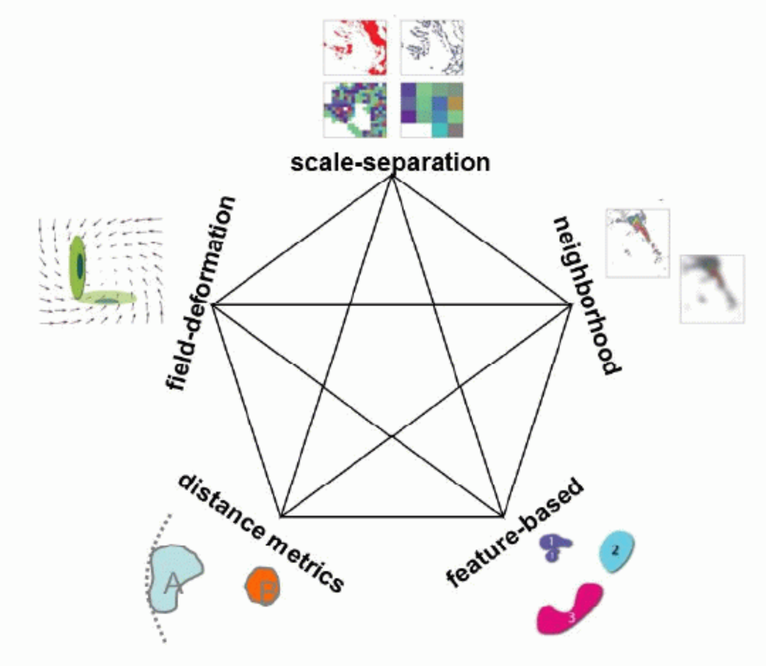
where *V* is the "volume" of interest spanned by *N* spatial dimensions

- Essentially a spatial CRPS

Courtesy Beth Ebert

MesoVICT update

- Overview BAMS paper submitted – proposes a reclassification of methods, and the introduction of a new class.
- Good buy-in from COSMO community (with their own complementary project INSPECT)
- Many centres have submitted re-runs.
- Extremely rich observation data set now updated with latest state-of-the-art model forecasts
- Special collection planned for MWR/WAF
- One other publication on MesoVICT so far (Skok and Roberts, QJ 2017).



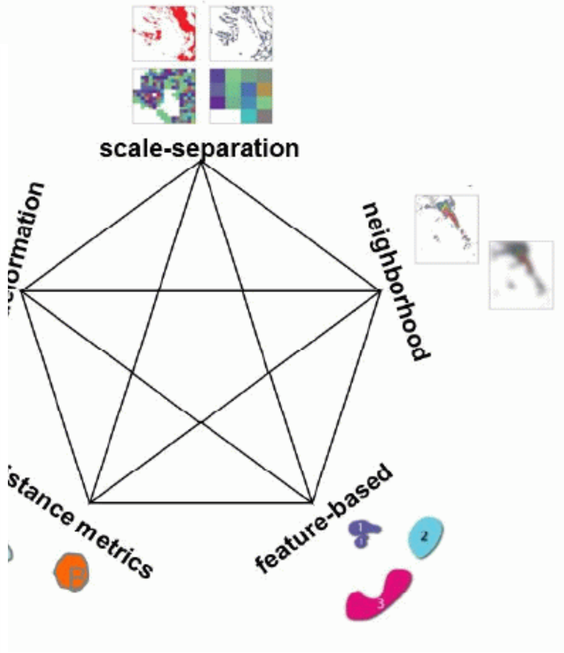
Event	Location	Date	Link
13th EMS/11th ECAM	Reading, UK	9-13 September 2013	https://www.ems2013.net/
1st MesoVICT workshop (kick-off)	Vienna, Austria	2-3 October 2014	http://mesovict.univie.ac.at
15th EMS/12th ECAM	Sofia, Bulgaria	7-11 September 2015	https://www.ems2015.eu
16th EMS/11th ECAC	Trieste, Italy	12-16 September 2016	https://www.ems2016.eu
2nd MesoVICT workshop	Bologna, Italy	21-23 September 2016	http://www.arpae.it/dettaglio_evento.asp?idLivello=32&id=2415
7th International Verification Methods Workshop	Berlin, Germany	8-11 May 2017	http://www.7thverificationworkshop.de/
Planned activities and presentations:			
Webex meeting		January 2018	
EMS annual meeting	Budapest, Hungary	September 2018	
ICAM	Lake Garda	Spring 2019	
EMS annual meeeting	Copenhagen, Denmark	September 2019	
8th International Verification Methods Workshop		2020	Final workshop

Met Office

MesoVICT

- Overview BAMS paper methods, and the introd
- Good buy-in from COS project INSPECT)
- Many centres have sub
- Extremely rich observa now updated with lates state-of-the-art model f
- Special collection planr
- One other publication o far (Skok and Roberts,

Institution	Model	Ensemble size (1=det.)	Mesh size km (degree)	Forecast range (hours)	Initial time (UTC)	MesoVICT case (section 4)
MeteoSwiss	COSMO-2	1	2.2 (0.02)	24	06	1-6
Environment Canada	GEM-LAM	1	2.5 (0.0225, 0.0327)	18	06	1-6
ARPA Emilia-Romagna	COSMO-LEPS	16	10 (0.09)	132	12	1-6
Environment Canada	GEM-glb	1	33 (0.297, 0.445)	48	00	1-6
Environment Canada	GEM-LAM-reg	1	15 (0.188, 0.135)	48	00	1-6
Environment Canada	GEM-LAM-hres	1	2.5 (0.0225,0.0327)	48	00	1-6
Met Office	UM12	1	12 (0.11)	30	00	1
Met Office	UM4	1	4 (0.04)	30	00	1
Met Office	UM1p5	1	1.5 (0.015)	30	00	1
Met Office	UM0p3	1	0.3 (0.003)	30	00	1
CETEMPS	WRF-d1	1	9 (0.09)	96	12	1,2,3,4
CETEMPS	WRF-d2	1	3 (0.03)	96 (1) 72 (2,3,4)	12	1,2,3,4
ISPRA	MOL0025	1	2.5 (0.025)	72	00	1-6
ISPRA	BOLAM007	1	7.8 (0.07)	96	00	1-6
ISPRA	BOLAM01	1	11 (0.1)	84	00	1-6
MeteoSwiss	COSMO-1	1	1.1 (0.01)	24	06	1-6
MeteoSwiss	COSMO-E	21	2.2 (0.02)	72	00, 12	1, 2
Russia Hydro Met	COSMO-Ru-E	10	2.2 (0.02)	72	12	1
Russia Hydro Met	CSOMO-Ru-E	51	2.2 (0.02)	72	12	1



Link
https://www.ems2013.net/
http://mesovict.univie.ac.at
https://www.ems2015.eu
https://www.ems2016.eu
http://www.arpae.it/dettaglio_evento.asp?idLivello=32&id=2415
http://www.7thverificationworkshop.de/
Final workshop

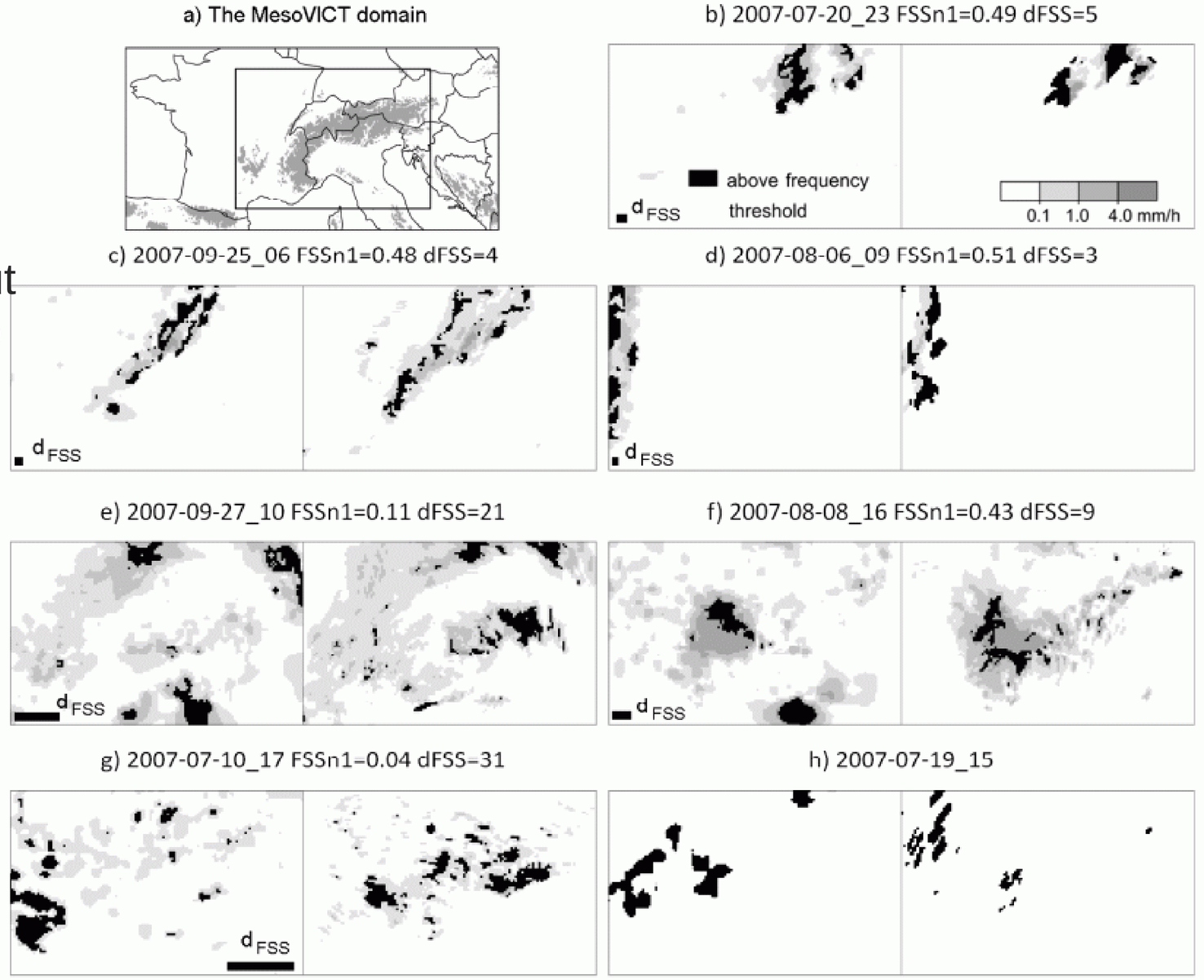
Real cases 2: MesoVICT cases

Skok and Roberts
(workshop pres)

MesoVICT publication exploring the properties of the FSS

Using FSS to determine spatial displacement is very appealing but large precipitation areas can have A disproportionately large impact On the derived spatial error.

1h hourly precipitation. 5% frequency threshold used.



7th international verification methods workshop and tutorial

- Berlin 3-11 May 2017
- Jointly hosted by the Free University of Berlin, the Max-Planck-Institute for Human Development, the Hans-Ertel-Centre for Weather Research (HErZ) and the German Weather Service DWD.
- The **goal** of the workshop is to discuss and promote all aspects of verification methodology research and practice, as applied to weather forecasts and warnings, climate predictions, and their applications. **Special sessions are planned on verification methods for sub-seasonal and longer range forecasts.** Participants are welcome from operational, research and forecast user communities.

Weather information value chain workshop: 7 May 2017

- Organised by Brian Mills and Jeff Lazo, with many JWGFVR members in attendance.
- Brian and Jeff presented a value chain model, which was practically applied to examples selected by the participants.
- Why do we need one? There is a distinction between data, information and knowledge. We need something to explicitly map the value of the information from creation to valuation in economic terms. Such a chain is based on stakeholders rather than processing. For the user the chain should be read from right to left.

Courtesy Jeff Lazo

Tutorial

- Was extended in length in response to feedback from previous tutorials
- 32 students from 26 countries: Argentina, Brazil, Iran, Kazakhstan, Nepal, Serbia, Ukraine, India, Australia, Botswana, Finland, UK, ISA, Spain, Sweden, Burkina Faso, France, Iceland, South Korea, Russia, South Africa, China, Germany, Italy.
- Topics covered include: Introduction to R, basic concepts, data preparation, verification of continuous, categorical and probabilistic/ensemble forecasts, S2S, statistical inference, spatial methods, operational verification systems. A selection of “special topics” were also covered: multivariate verification, observation uncertainty, warnings and extremes, non-standard variables.
- As usual students were split into groups to work on projects. Each group presented their result during the workshop.

Workshop

- Workshop extended in length to accommodate contributions from the seasonal and climate communities.
- Full and varied science programme with sessions on HIW, meta verification, observation uncertainty, user-relevant metrics, methods for ensemble and probabilities, spatial methods, software, S2S, climate and verification studies.
- Over 100 abstracts were submitted with 6 keynote talks by Beth Ebert, Helge Goessling, Frederic Vitart, Henning Rust, Jeff Lazo and Chris Ferro.

Met Office

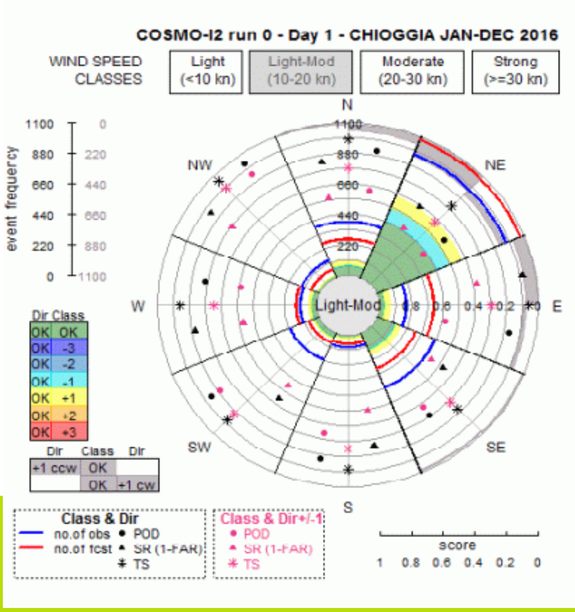
From Ebert et al.

1. Forecast or product to be verified	
What is the forecast or product?	Simulator isochrones
2. User of the verification (complete for each type of user):	
Who is the user?	Fire behaviour analyst
How do they use the forecast?	Decision support
What aspects of forecast quality are important to them (e.g., bias, timing, absence of large errors, etc.)?	Biases in the rate of spread and direction? Right place right time Avoiding misses Can tolerate false alarms
For this user, what are likely to be the most effective ways to describe forecast performance (charts, statistics, verbal description, etc.)?	Maps and summary statistics
Can the user participate in the design of the verification – if so, how?	Absolutely – iterative drafting
3. Characteristics of the forecast	
Variable(s) and units	Where (location) and when (hours, minutes)
Spatial domain	Varies by event – rectangular box surrounding the fire
Point / area / grid (resolution)	Area (polygon)
Forecast range	At least 12 hours. Up to 4 days. In practice (24 hours max)
Temporal resolution of output	Hourly – can do half hourly
Instantaneous / averaged / accumulated	Accumulated
Update frequency	Variable and driven by changes in input data (new weather grids or fire intelligence).
4. Available observations and their characteristics	
Variable(s) and units – same as forecast, or proxy?	Line scans + GPS track from aircraft, GPS track from ground vehicles and some manual maps, satellite

Workshop highlights

- Special collection of papers in Meteorologische Zeitschrift (deadline end Oct, likely to be extended)
- The workshop facilitated several emerging themes:
  - Emphasis on the development of community tools. MET, R packages (SpatialVx, scoringRules, murphydiagram, freva) with web interfaces, running in the cloud.
  - How to plan verification studies
  - New visualisation methods and diagrams: performance “rose”
  - Continued exploration of what to do about extremes, characteristics of scores
  - How do we extract the signal from the noise, especially for model testing
  - Impact of (lack of) observations

From Tesini et al.



From Geer et al.

Summary: four issues in operational R&D verification

1. Type I error due to multiple comparisons:
   - Try to determine how many independent tests *n* are being made (e.g. compute correlation between scores)
     - Paired differences in medium range dynamical tropospheric scores are all quite correlated
     - Paired differences are correlated at different forecast ranges
   - Once *n* is estimated, use a Šidák correction

2. Type I error due to time-correlated forecast error:
   - Chaos experiment used to validate an AR(2) model for correcting time-correlations
   - Note that at forecast day 10, this may not work: long-range time-correlations?

From Geer et al.

Summary: four issues in operational R&D verification

3. Type II error because typical experiments test only small changes in forecast error:
   - 300-400 forecasts are now a minimum requirement for research experiments at ECMWF

4. Are the forecast scores meaningful?
   - Own-analysis scores are accurate in the medium and long-range, for midlatitude dynamical scores
   - In other areas (e.g. tropics, stratosphere, early forecast range) these scores are often measuring something very different from forecast skill
     - Also check observational-based verification

For more detail on issues 1-3 see Geer (2016,Tellus) "Significance of changes in forecast scores"

WG membership

- Several members due to leave the group in December: L. Wilson (co-chair), B. Brown, Y. Zhu and we have one unfilled vacancy.
- Caio Coelho (CPTEC, Brazil) becomes new co-chair in January 2018.
- We have a short list but several people we have approached in the last year have declined.
- We will be approaching people between now and the end of the year to fill the 4 vacancies.
- Remaining members: Marion Mittermaier (MetO, co-chair), Thomas Haiden (ECMWF), Barbara Casati (ECCC), Caio Coelho (CPTEC), Jing Chen (CMA), Chiara Marsigli (ARPA), Manfred Dorninger (U. Vienna), Stephanie Landman (SAWS)

Met Office

Questions?