



ATMOSPHERIC
RESEARCH

RECHERCHE
ATMOSPHERIQUE

SURVEY OF CLIMATE VARIATION
IN THE TROPICS

by
Henry R. Stanski
Lauren J. ...
William R. B...

RESEARCH REPORT NO.
(MSRE) 89-5

July, 1989

Верификация численных и синоптических прогнозов: концепции и проблемы

И.А.Розинкина, Г.С.Ривин



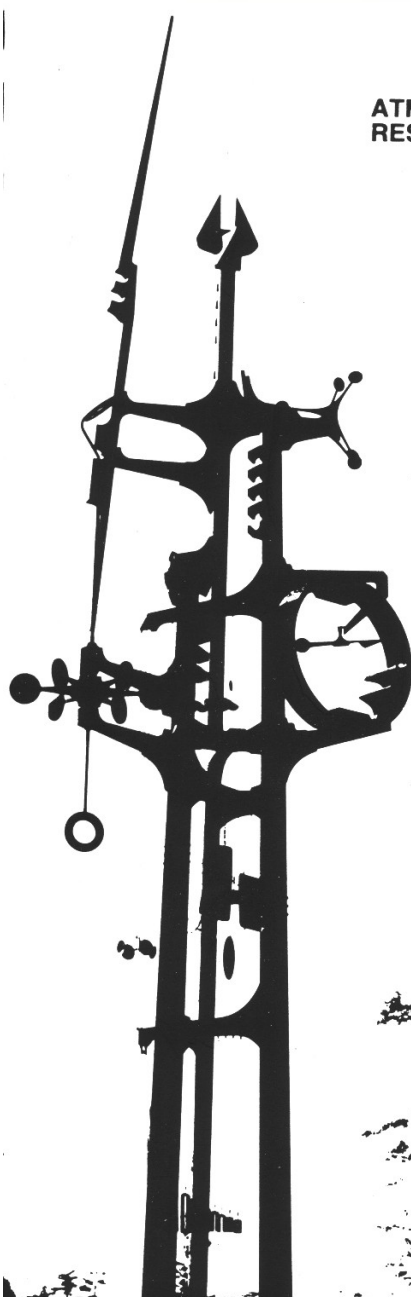
Environment
Canada

Environnement
Canada

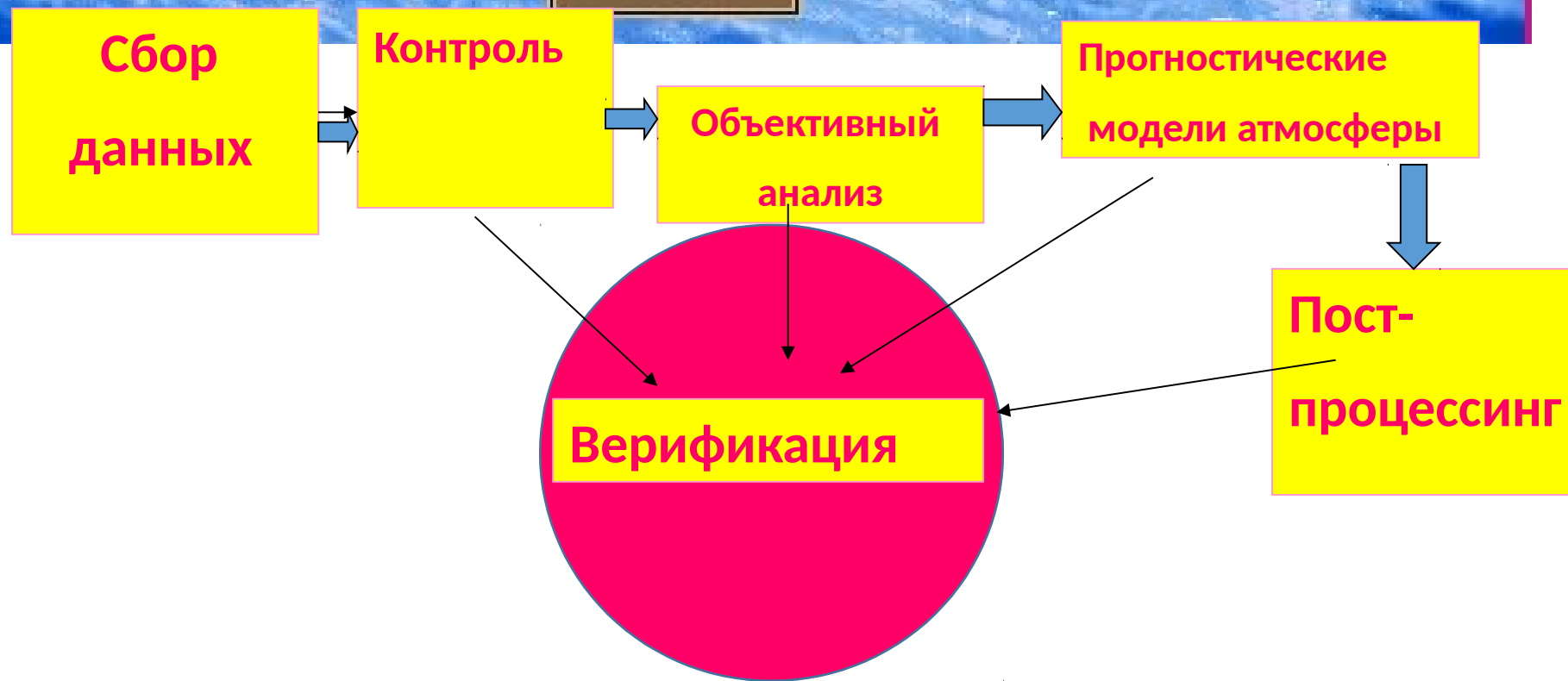
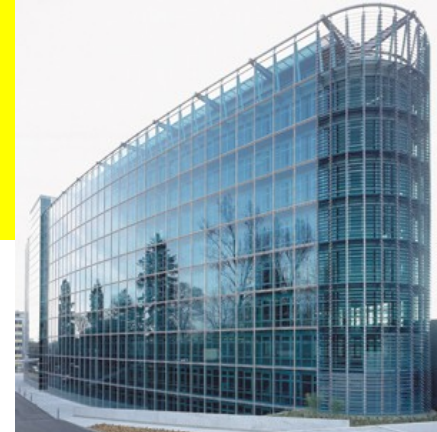
Atmospheric
Environment
Service

Service
de l'environnement
atmosphérique

inna.rozinkina@mail.ru, gdaly.rivin@mail.ru



Технологическая цепочка автоматизированной обработки метеорологической информации





Этапы развития обработки гидрометеорологической информации

Единичное наблюдение



150-200 лет

Синхронные наблюдения над большими областями
Построение метеополей
19 февраля 1855



70 - 80 лет

Построение метеополей свободной атмосферы 1930



50 - 80 лет

Численный анализ и построение полей будущего состояния атмосферы 1950



Интерпретация численных прогнозов для пунктов



300 - 350 лет

30 лет



Что означает верификация?

- Вычисление показателей успешности прогнозов путем сопоставления прогноза и реальности.
- Верификация производится на основе сопоставления:
 - а) значений метеорологических величин
 - б) факта «попадания» в интервалы пороговых значений
 - в) наличия в заданной области метеорологических объектов

Большинство критериев оценок для а) и б) изложено в нормативных документах национальных метеорологических служб и ВМО



Цели и подходы верификации

- Выявление **типичных погрешностей** прогноза предиктантов (помощь при интерпретации прогнозов, важно для развития методов прогноза) – «научный подход»
- Определение **степени достоверности** прогноза (принятие решений при использовании прогноза) – подход «менеджера»
- Определение **сравнительной полезности** прогноза или методики прогнозирования (насколько полезен и **информативен**) - «административный» подход

Предиктант- объект прогноза

Оценки прогнозов делаются для начальника, модельера или пользователя, отсюда «верификация» научная, экономическая и административная



Some general issues... on (operational) verification

**Verification is very much about making questions
– and trying to answer them**

Who ...

... wants to know? Goals of the verification?

What ...

... kind of parameter are we evaluating ? What are its characteristics ?
(e.g., continuous, probabilistic)

... thresholds are we looking at ? – Essentially event-based ⇔ **FROST**

... forecast resolution (e.g. point wise, station groups, area-averages) ?

... are the characteristics of the observations (e.g. quality, uncertainty) ?

How ...

... to present the results (e.g., stratification/ aggregation) ? ⇔ **User impact**

Which ...

...methods and metrics to use (e.g. bias, SEDS, EDI etc...) ?

Кто... интересуется
результатами? Цели?

Что ... требуется
оценить? Каковы
характеристики
предиктанта и
наблюдений?

Как ... это сделать? Какие
методики и критерии?



Хотелось бы добиться от верификации

- Объективной оценки для редких событий
- Минимальной зависимости от пространственно-временных масштабов явления
- Минимальной зависимости от пространственного разрешения вычислений
- Учета как хитов, так и ложных тревог
- Быстрой сходимости на небольших выборках
- Сопровождения мерой доверия

ДО НАСТОЯЩЕГО ВРЕМЕНИ НИ ОДНА СИСТЕМА КРИТЕРИЕВ НЕ УДОВЛЕТВОРЯЕТ ЭТИМ ТРЕБОВАНИЯМ!



**РД 52.27.284-91 Методические указания. Проведение
производственных (оперативных) испытаний новых и
усовершенствованных методов гидрометеорологических
и гелиогеофизических прогнозов**

РД 52.27.284-91

Группа Т58

**РУКОВОДЯЩИЙ ДОКУМЕНТ
МЕТОДИЧЕСКИЕ УКАЗАНИЯ**

**Проведение производственных (оперативных) испытаний новых и
усовершенствованных методов гидрометеорологических и
гелиогеофизических прогнозов**

Дата введения 1992-01-01



WWWRP 2009 - 1 Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP



World
Meteorological
Organization
Weather • Climate • Water
WMO/TD - No. 1485

Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP Models

Revision 2
October 2008



Из нормативных документов ВМО

- *Заблаговременности (6ч, 12ч, etc.)*
- *Сезона (месяца)*
- *Региона*
- *Для ряда явлений – по порогам значений*

Многомерность !

Ключевые концепции: процесс верификации

- Заранее требуется **определить то, что будем проверять:**

Какой элемент, для какой области, для какого периода времени:

- области должны быть **сравнительно небольшими и географически однородными**, обладать хотя бы одной репрезентативной станцией наблюдений (для проверки прогнозов погодных явлений)
- **Осреднение успешности по всему региону НГМС** – решает НГМС путем осреднения либо по всем станциям, либо по осредненным показателям успешности по выделенным регионам. (имеет значение при неравномерной плотности населения). Можно считать оба варианта
- Для выполнения расчетов успешности синоптических прогнозов часто используют **EXEL- возможности**, если речь идет о небольших территориях. Доступно и можно легко обнаружить причины лучшего/ худшего качества прогнозов

Ключевые концепции:

1. Заранее необходимо **знать пользователя и цель** – для какой цели определяются параметры успешности? *«Определить, как изменилась успешность прогнозов НГМС при осуществлении какого-либо проекта»*
«Определить, насколько изменилась успешность численных прогнозов при изменении переходе на другой способ определения начальных данных»
2. **Нельзя подобрать единственный критерий** для оценки прогнозов. Для получения относительно полной картины об успешности прогнозирования требуется не менее 2-х взаимодополняющих критериев
3. **Прогноз для верификации должен быть проверяем:** например, формулировка должна содержать сведения- какой элемент (предиктант), где и когда ожидается *«Осадки с интенсивностью свыше 20 мм ожидаются 25 марта на юге Узбекистана»* - требуется точно знать, какая область входит в понятие «юг Узбекистана». Либо: «прогнозируется местоположение явления» – требуется установить, что имеется в виду- на станциях, по части



Верификация прогнозов

Непрерывных полей

Непрерывных полей
внутри области фактов
явлений

Количественные критерии
успешности: средние,
среднеквадратические ошибки,
Возможно: коэффициенты
корреляции по пространству и
времени

Величин с разрывами
(фактов явлений)

Базируются на таблицах
«сопряженности»



Оценка непрерывных сеточных полей значений метеоэлементов (численных прогнозов)

- Средняя ошибка
- Средняя абсолютная ошибка
- Среднеквадратическая ошибка
- Коэффициент корреляции аномалий
- Коэффициент корреляции тенденций
- Градиентная ошибка

- Ошибка объектов в целом (глубины и положения центров циклонов)

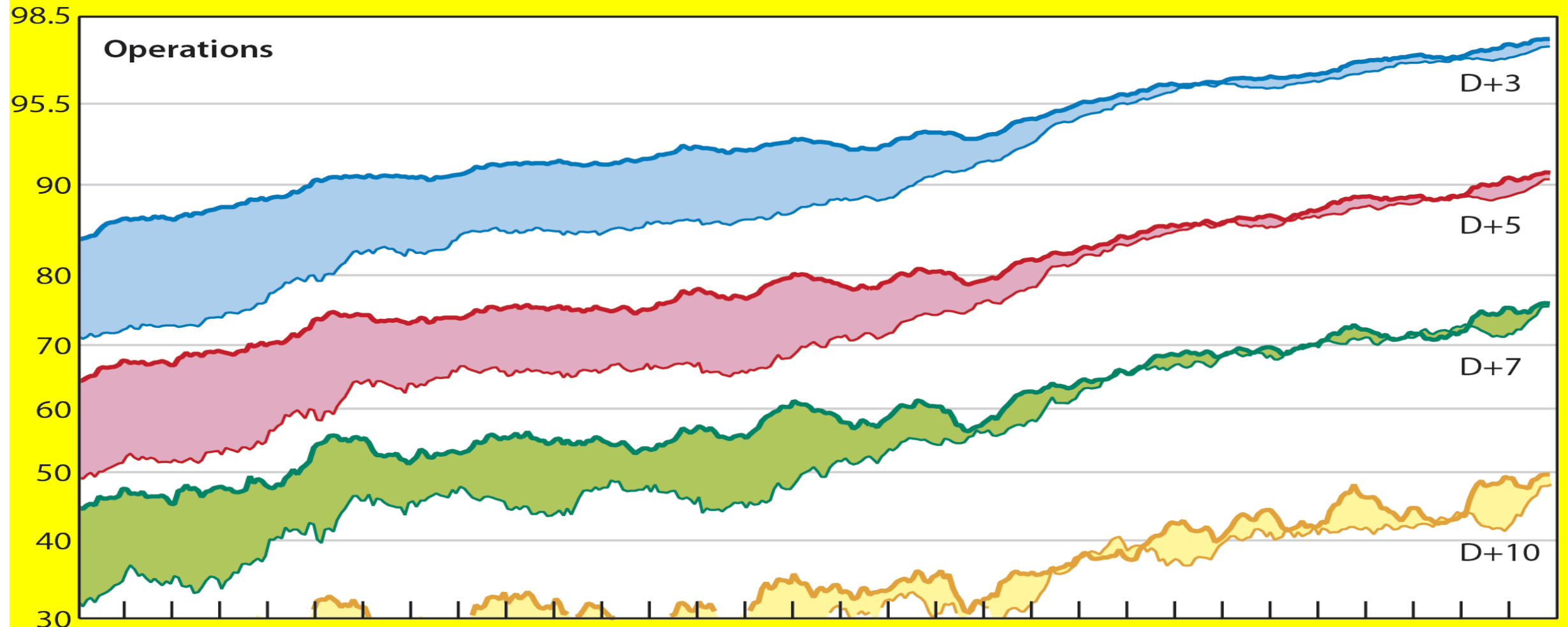


С повышением заблаговременности успешность прогнозов неизменно уменьшается!

Anomaly correlation of 500 hPa height forecasts

— Northern hemisphere

— Southern hemisphere





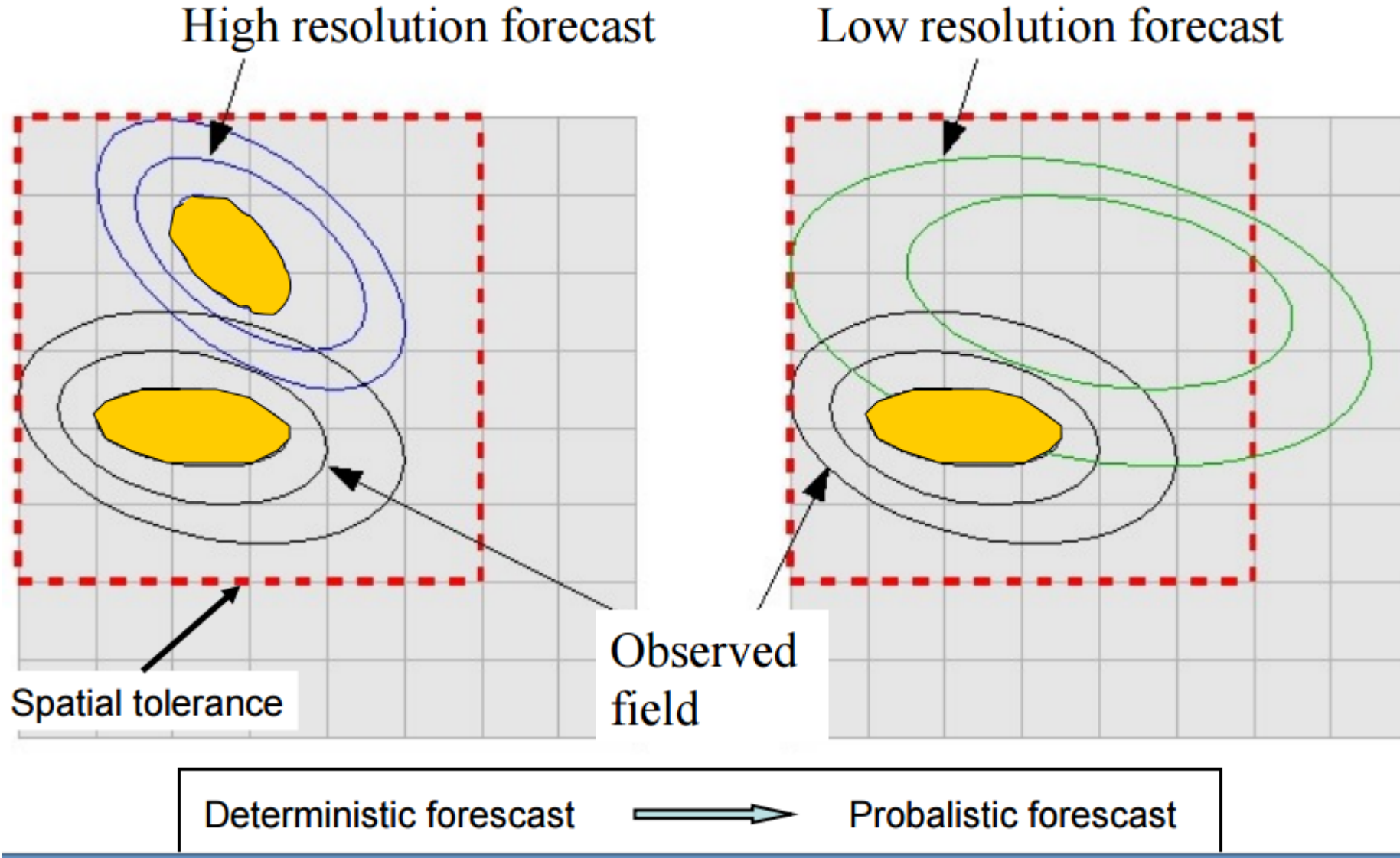
- Категорийные прогнозы

- прогноз явлений

В качестве явления – «попадание» в заданный интервал значений

Важнейший этап – выбор интервалов для оценок

Проблемы верификации прогнозов высокого разрешения



Рекомендации ВМО по оценке прогнозов опасных явлений на основе опыта SWFDP, 2014



Forecast Verification for the African Severe Weather Forecasting Demonstration Projects

**SURVEY OF COMMON VERIFICATION METHODS IN
METEOROLOGY**

WMO/TD-No. 358



ATMOSPHERIC
RESEARCH

RECHERCHE
ATMOSPHERIQUE

SURVEY OF COMMON VERIFICATION METHODS
IN METEOROLOGY

by
Henry R. Stanski
Laurence J. Wilson
William R. Burrows

RESEARCH REPORT NO.
(MSRB) 89-5

July, 1989

1989 г.. Одна из первых монографий по
верификации

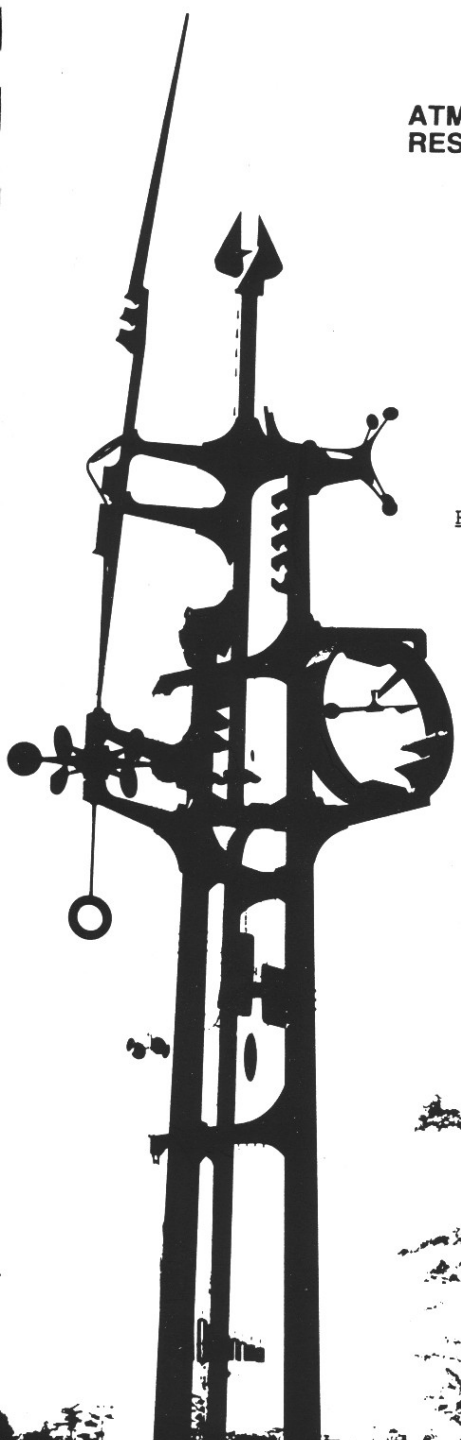


Environment
Canada

Environnement
Canada

Atmospheric
Environment
Service

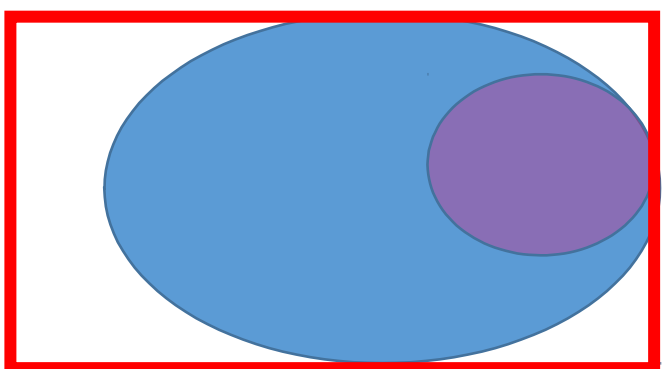
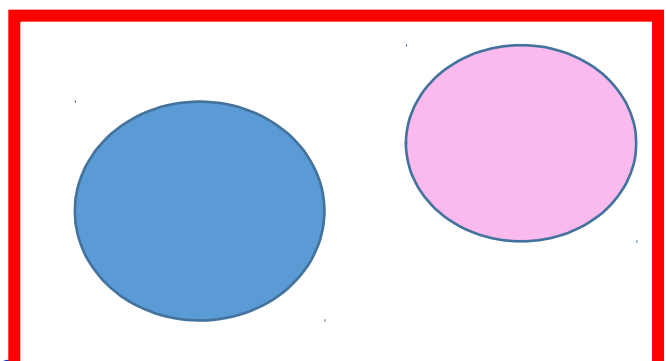
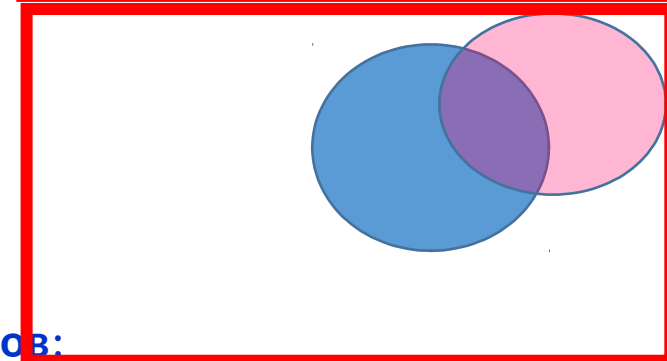
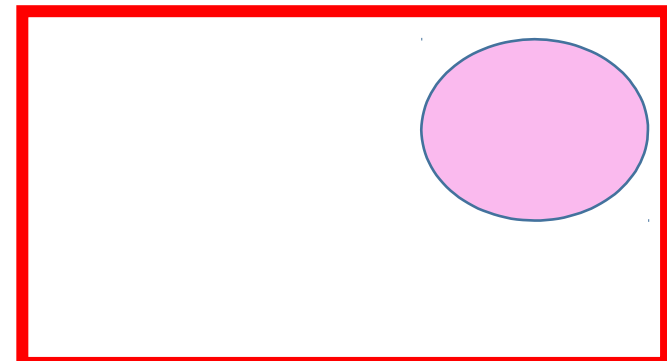
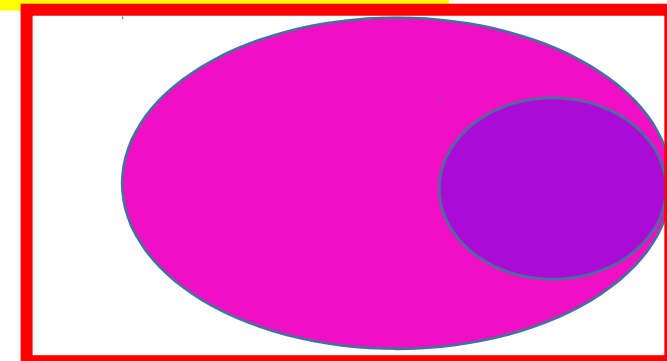
Service
de l'environnement
atmosphérique



Оценка прогнозов в альтернативной форме «категорийных» прогнозов



	прогноз			
		Без явления	с явлением	Всего
факт	Без явления	N_{00}	N_{10}	$N_{00}+N_{10}$
	С явлением	N_{01}	N_{11}	$N_{01}+N_{11}$
	Всего	$N_{00} + N_{01}$	$N_{10} + N_{11}$	N



На анализе таблицы сопряженности построены оценки альтернативных прогнозов:

Предупрежденность явления, оправдываемость прогноза (отсутствия) явления, критерии Пирси-Обухова и т.д.

Верификация категориальных прогнозов

те ИИКСИГ

FOU Y

Finley Tornado Data (1884)



LIEUTENANT JOHN P. FINLEY, SIGNAL CORPS, UNITED STATES ARMY.

John P. Finley

Forecast answering the question:

Observation answering the question:

Will there be a tornado?

Did a tornado occur?

YES
NO



YES
NO

Answers fall into 1 of 2 categories ** Forecasts and Obs are Binary

Верификация категориальных прогнозов

But the contingency table concept is good...

2 x 2 Contingency Table

		Observed		
		Yes	No	Total
Forecast	Yes	Hit	False Alarm	Forecast Yes
	No	Miss	Correct Negative	Forecast No
	Total	Obs. Yes	Obs. No	Total

Example: Accuracy = (Hits+Correct Negs)/Total

Верификация категориальных прогнозов

A Success?



LIEUTENANT JOHN P. FINLEY, SIGNAL CORPS, UNITED STATES ARMY.

John P. Finley

		Observed		
		Yes	No	Total
Forecast	Yes	28	72	100
	No	23	2680	2703
	Total	51	2752	2803

Percent Correct = $(28+2680)/2803 = 96.6\% !!!!$

Верификация категориальных прогнозов

What if forecaster
never forecasted a tornado?



LIEUTENANT JOHN F. FINLEY, SIGNAL CORPS, UNITED STATES ARMY.

John F. Finley

		Observed		
		Yes	No	Total
Forecast	Yes	0	0	0
	No	51	2752	2803
	Total	51	2752	2803

Percent Correct = $(0+2752)/2803 = 98.2\% !!!!$

Examples of Categorical Scores

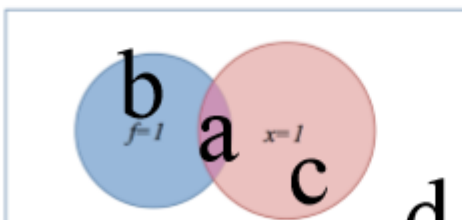
(most based on conditioning)

- Hit Rate (POD_y) = $a/(a+c)$
- POD_n = $d/(b+d) = (1 - \text{POFD})$
- False Alarm Rate (POFD) = $b/(b+d)$
- False Alarm **Ratio** (FAR) = $b/(a+b)$
- (Frequency) Bias (FBIAS) = $(a+b)/(a+c)$
- Threat Score or Critical Success Index = $a/(a+b+c)$

POD
Probability of
Detection

POFD
Probability of
False Detection

Conditional probabilities

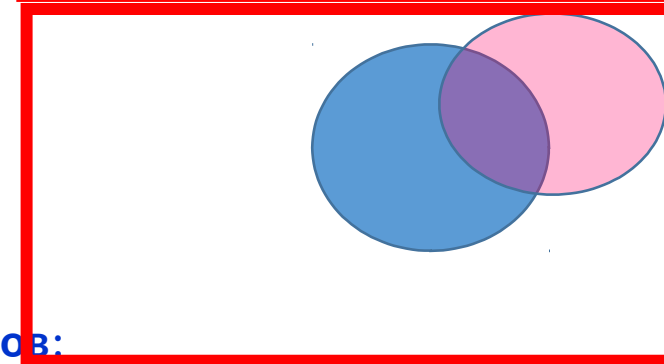
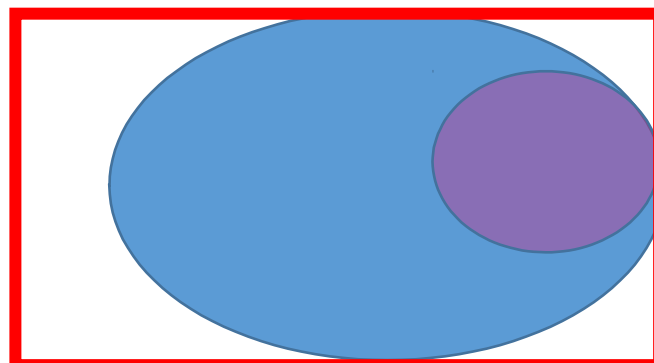
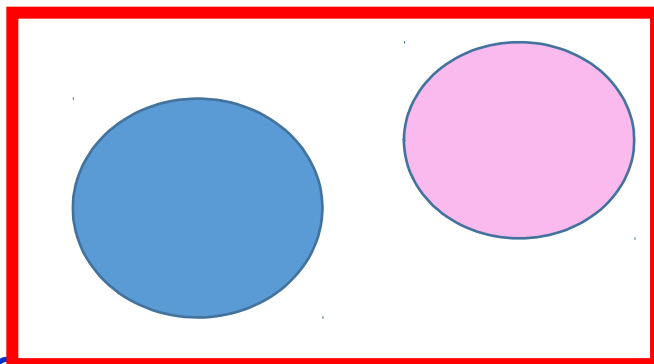
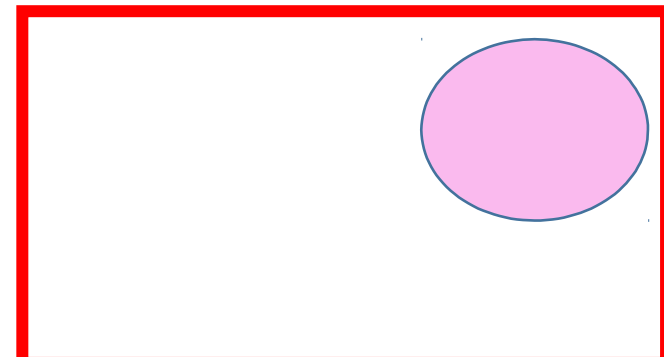
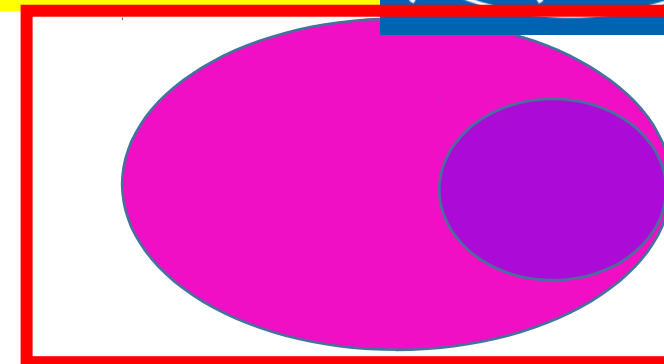


(CSI)

		Observed		
		Yes	No	Total
forecast	Yes	a	b	a+b
	No	c	d	c+d
	Total	a+c	b+d	a+b+c+d

Оценка прогнозов в альтернативной форме «категшорийных» прогнозов

	прогноз			
		Без явления	с явлением	Всего
факт	Без явления	N_{00}	N_{10}	$N_{00}+N_{10}$
	С явлением	N_{01}	N_{11}	$N_{01}+N_{11}$
	Всего	$N_{00} + N_{01}$	$N_{10} + N_{11}$	N



На анализе таблицы сопряженности построены оценки альтернативных прогнозов:

Предупрежденность явления, оправдываемость прогноза (отсутствия) явления, критерии Пирси-Обухова и т.д.



N11 N12 N10

МАТРИЦА СОПРЯЖЕНИЙ

N21 N22 N20

N01 N02 N00

N11 - явление прогнозировалось и наблюдалось;

N12 - явление прогнозировалось, но не наблюдалось («ложные тревоги»);

N10=N11+N12 - число случаев, где прогнозировалось явление;

N21- прогнозировалось отсутствие явления, но оно наблюдалось («пропуск цели»);

N22 - прогнозировалось отсутствие явления и его не наблюдалось;

N20=N21+N22 - число случаев с прогнозом отсутствия явления;

N01=N11+N21 - число случаев с явлением;

N02=N12+N22 - число случаев с отсутствием явления;

N00 - общее число случаев.

$U=N11/(N21+N10)$ - оправдываемость редкого явления;

$Pred=N11/N01$ - предупреденность явления;

$kLT=N12/N01$ - коэффициент «ложных тревог»;

$ETS=(N11-ar)/(N11-ar+N12+N21)$ - критерий ETS, где

$ar=((N11+N12)*(N11+N21))/N00$; «равновесная оценка угроз»

$BX=(v-v0)/(1-v0)$ - критерий Н.А.Багрова-Хайдке (для редких явлений), где

$v=(N11+N22)/N00$, $v0=(m1+m2)/N00$, $m1=(N10 \times N02)/N00$, $m2=(N20 \times N02)/N00$;

Верификация категориальных прогнозов



Examples of Contingency table calculations

		Observed		
		Yes	No	Total
Forecast	Yes	28	72	100
	No	23	2680	2703
	Total	51	2752	2803

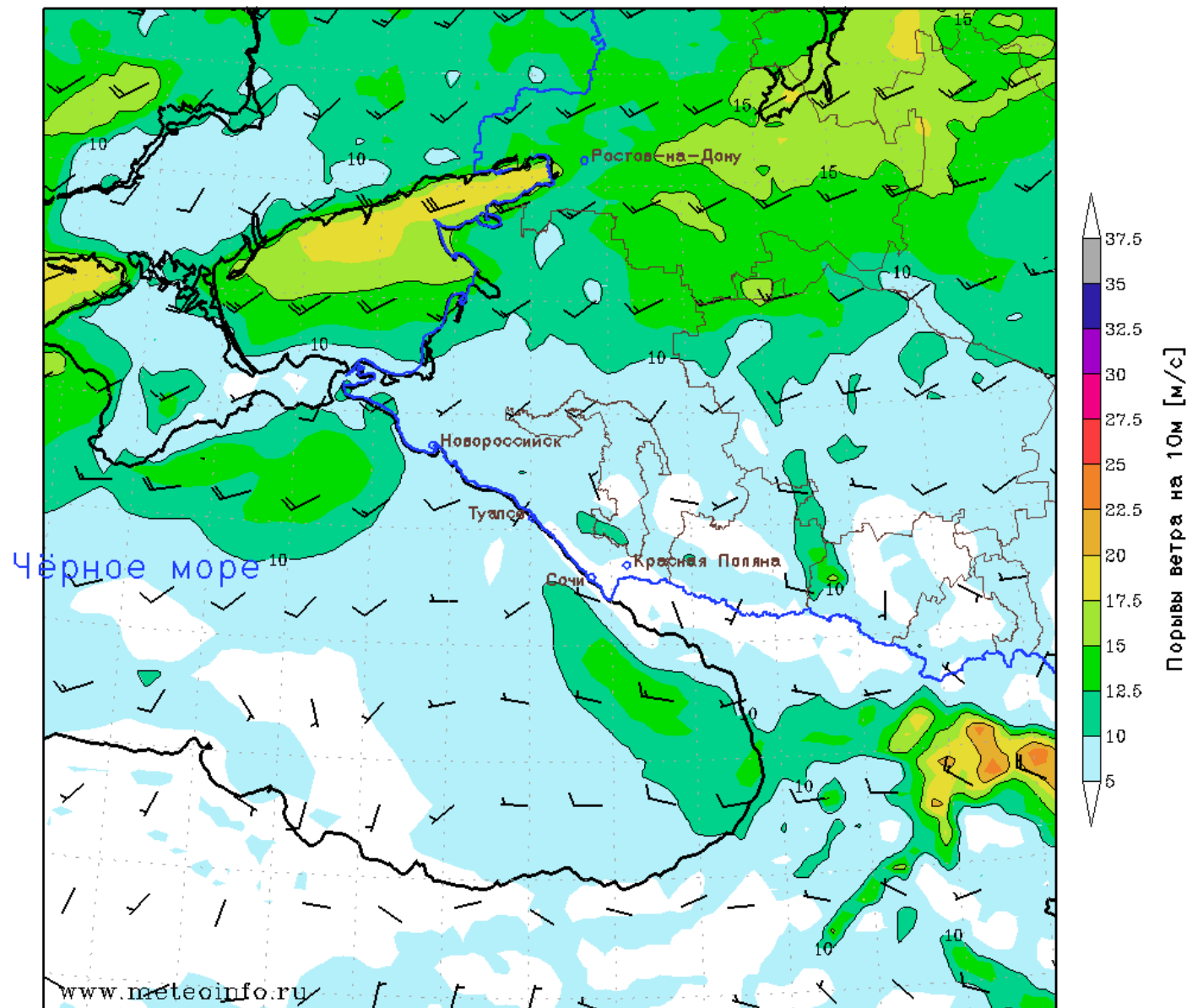
$$\text{Threat Score} = 28 / (28 + 72 + 23) = 0.228$$

$$\text{Probability of Detection} = 28 / (28 + 23) = 0.55$$

$$\text{False Alarm Ratio} = 72 / (28 + 72) = 0.720$$



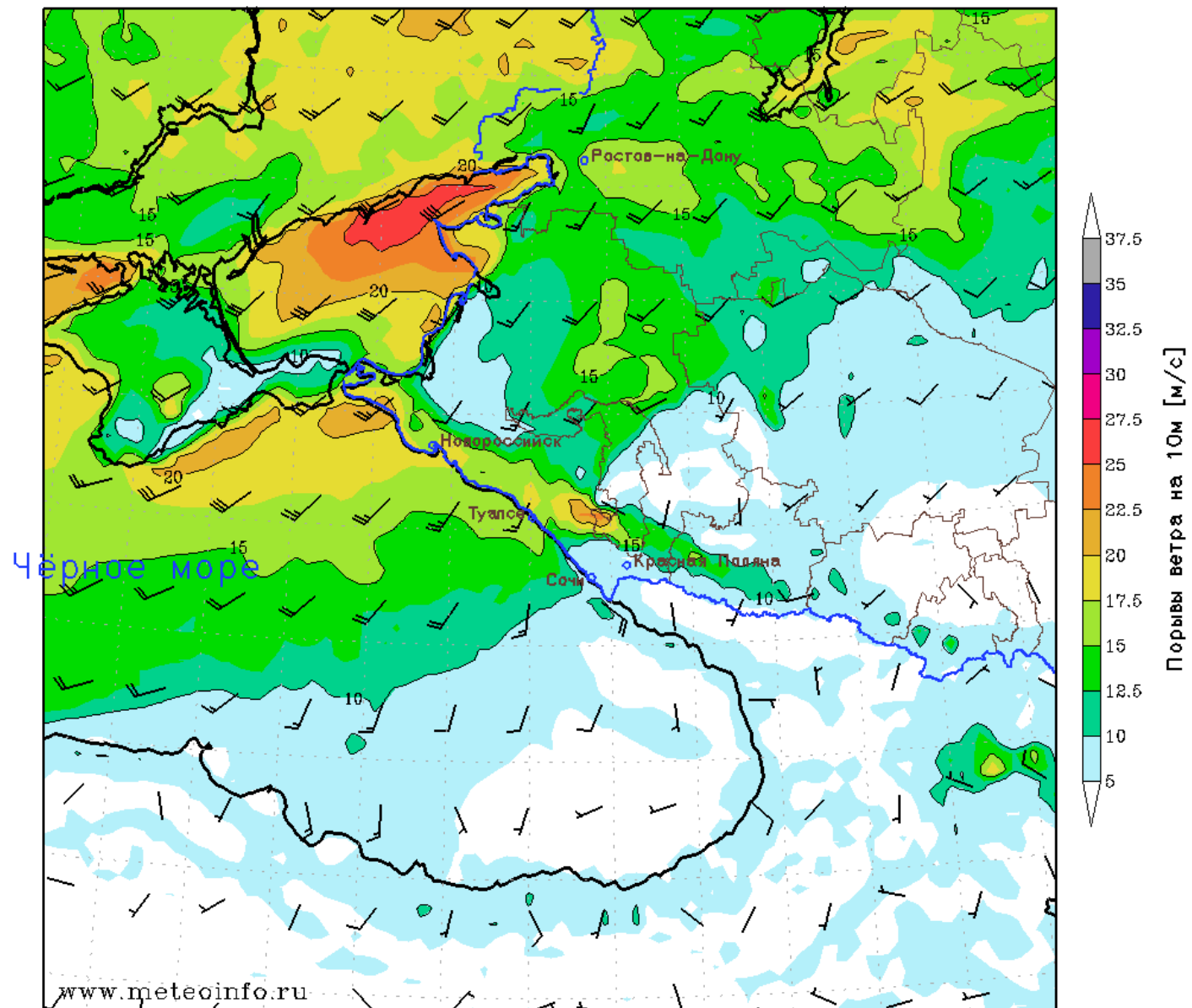
15:00 08дек 2016 (МСК): Ветер на 10м



Прогноз на 12ч. от 03:00 08дек 2016 (МСК)
COSMO-Ru 7км



21:00 08дек 2016 (МСК): Ветер на 10м

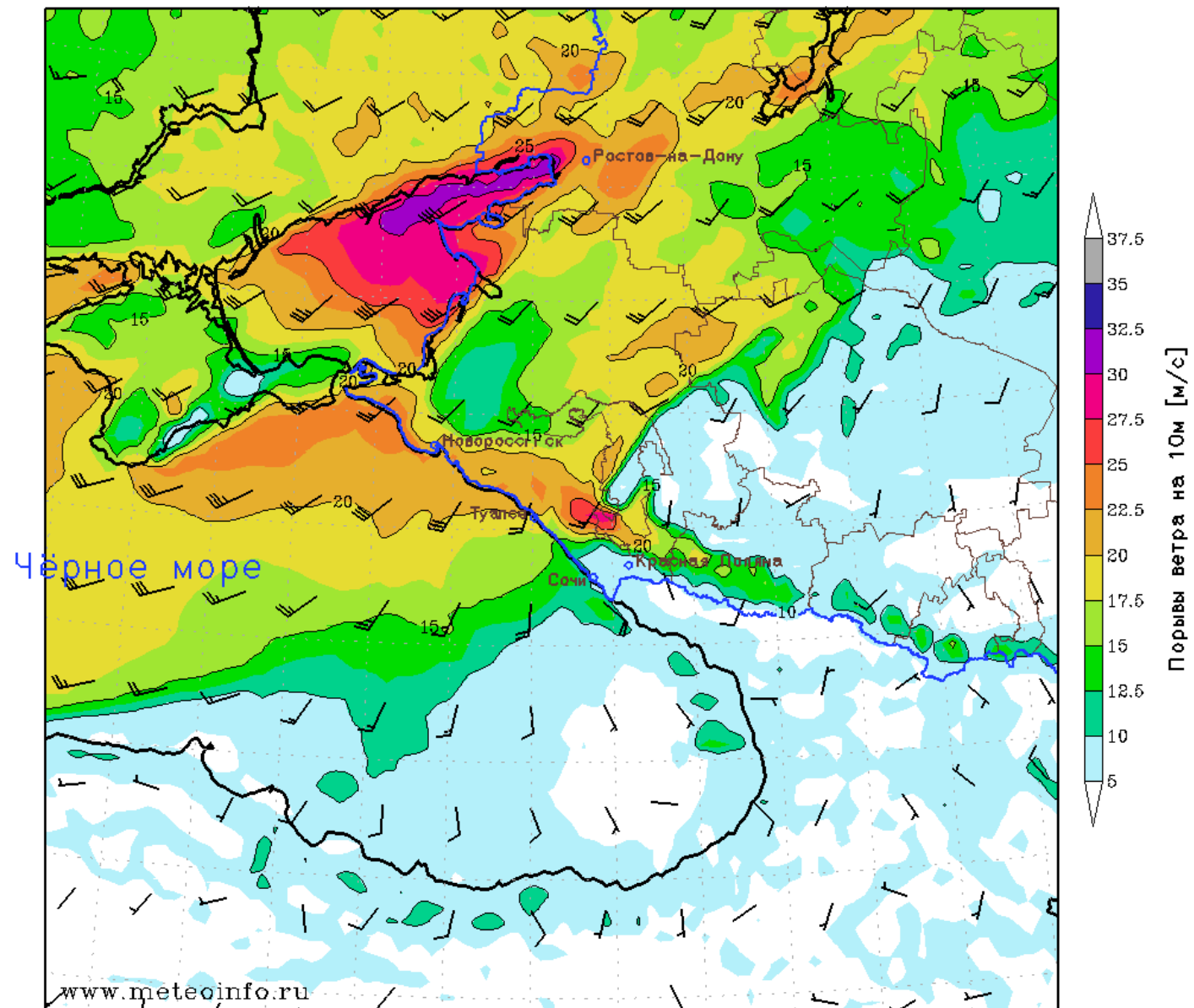


Прогноз на 18ч. от 03:00 08дек 2016 (МСК)
COSMO-Ru 7км

Ветер на 10м



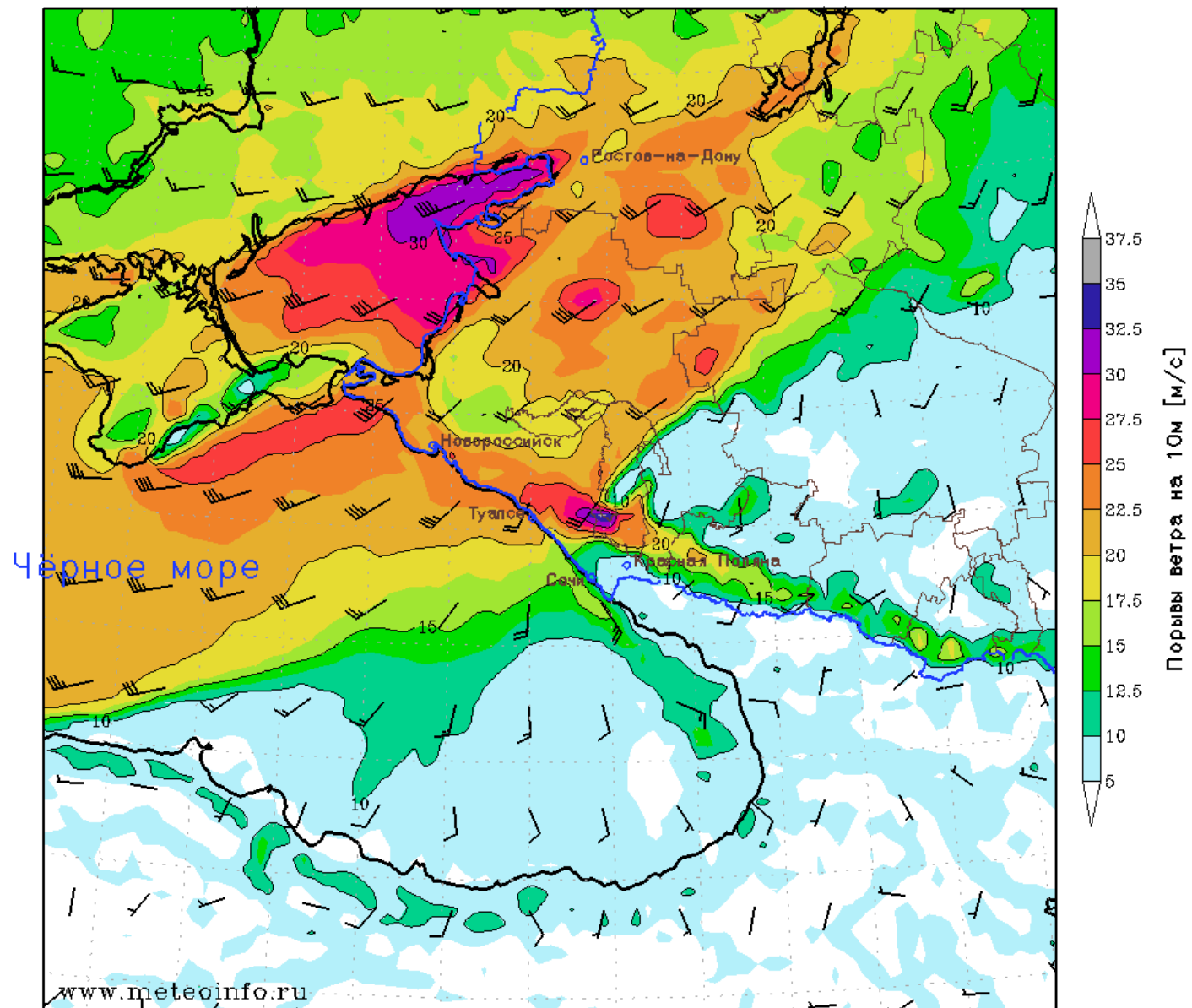
00:00 09дек 2016 (МСК): Ветер на 10м



Прогноз на 21ч. от 03:00 08дек 2016 (МСК)
COSMO-Ru 7км



03:00 09дек 2016 (МСК): Ветер на 10м



Прогноз на 24ч. от 03:00 08дек 2016 (МСК)
COSMO-Ru 7км

Ветер на 10м



- При оценке прогнозов явлений крайне осторожно нужно относиться к использованию в сравнении моделей различной детализации
- При переходах точка-область использование ETS может фиктивно ухудшать успешность прогнозов при повышении их детализации

«EQUITABLE THREAT SCORE APPLIED TO MCS RAINFALL: TO SCALE DEPENDENCE DISAGREEMENT WITH SUBJECTIVE EVALUATIONS» (William A., Gallus, Jr., Isidora Jankov, Steven Aves, Steven Aves Iowa State University Iowa State University)



Оценка мезомасштабных прогнозов: «двойной штраф»

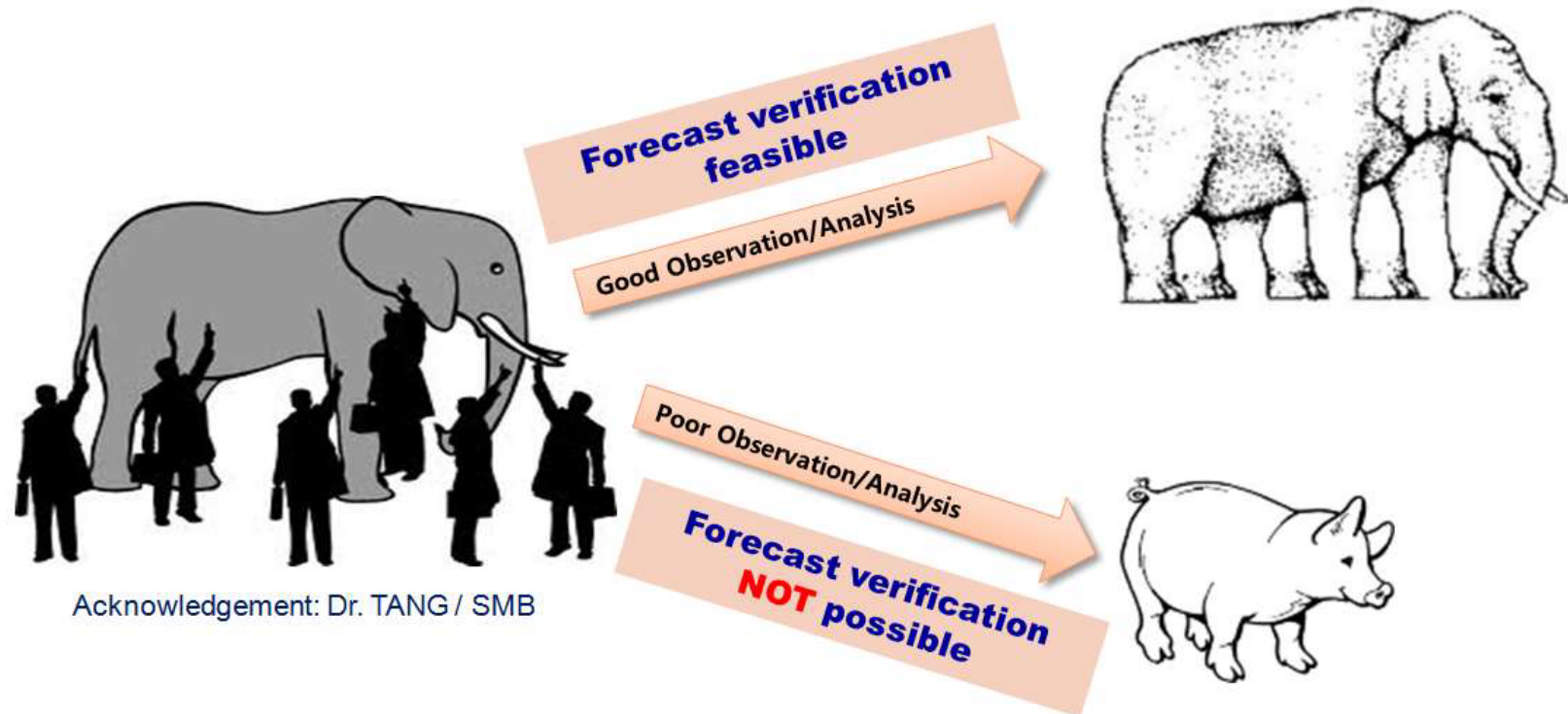
Более грубые модели «лучше» по метрическим показателям, чем более тонкие модели!

- Проблема «Double Penalty» и что есть «истина» в верификации?
- Новые методы верификации в мезомасштабах с конца прошлого века.
- Основной объект - разрывные поля осадков, «головная боль» для наблюдения, моделирования и верификации

Данные для верификации- важней

Observations are THE cornerstone of forecast verification !

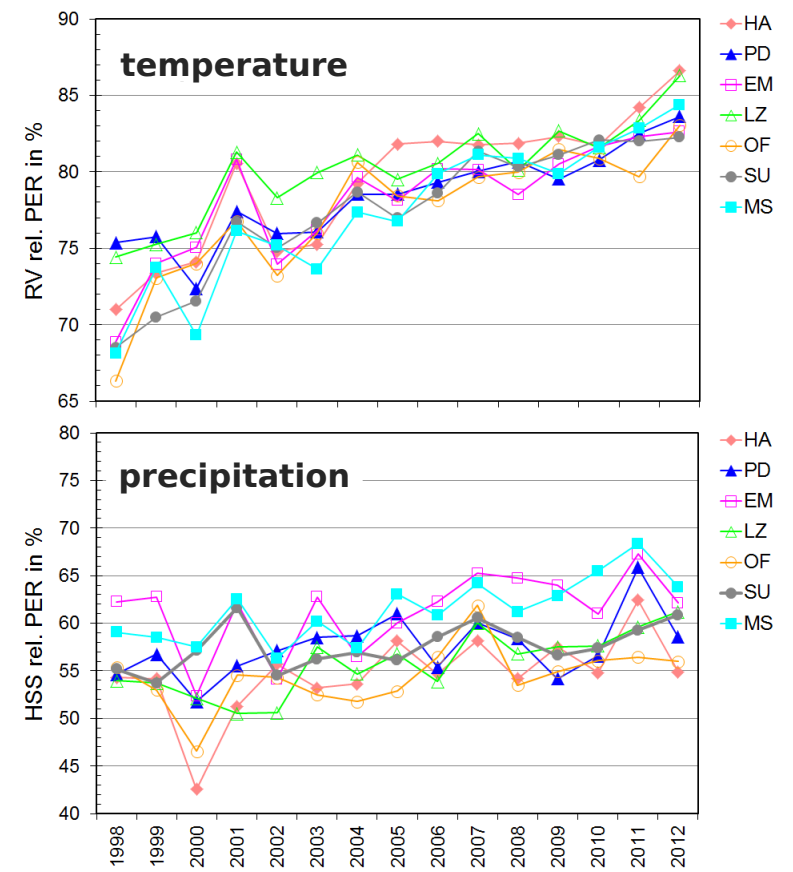
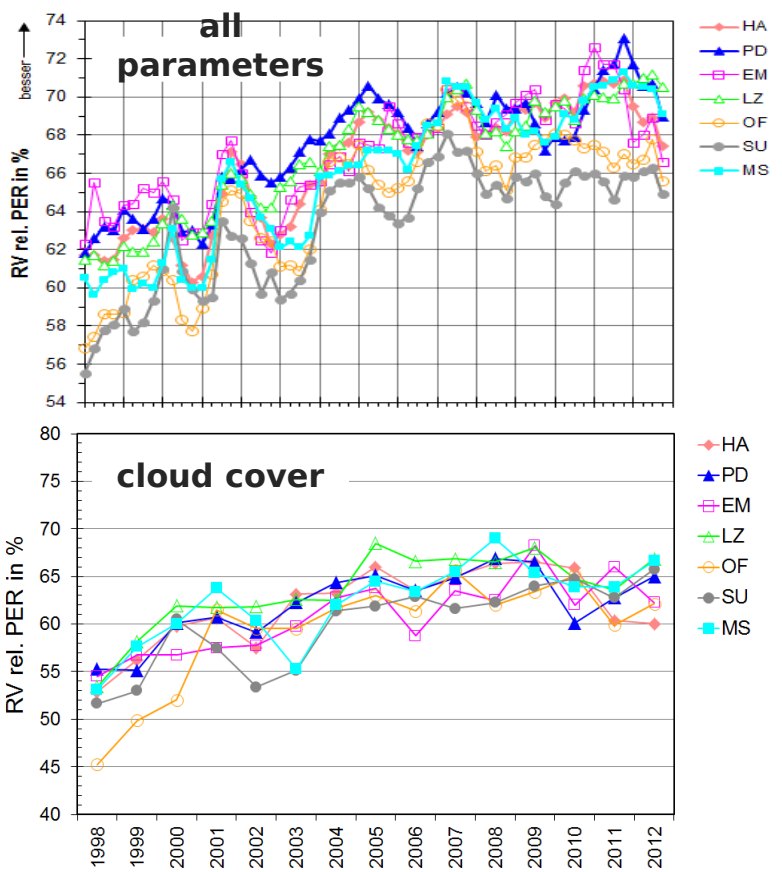
- ✓ Linking of various observing means and data platforms
⇔ Highly relevant in forecast verification
- ✓ Observation needs always to be associated with forecast verification ⇔
Verification may help in identifying observation gaps



Acknowledgement: Dr. TANG / SMB

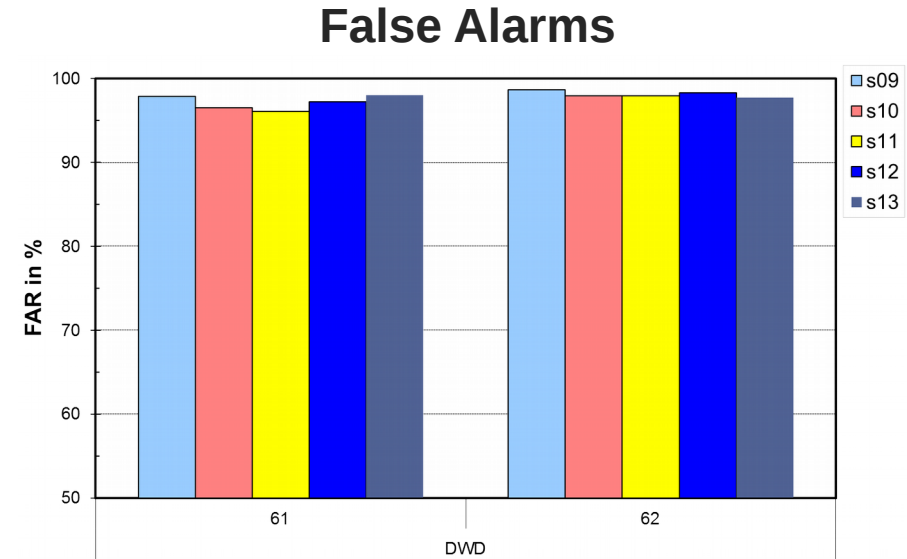
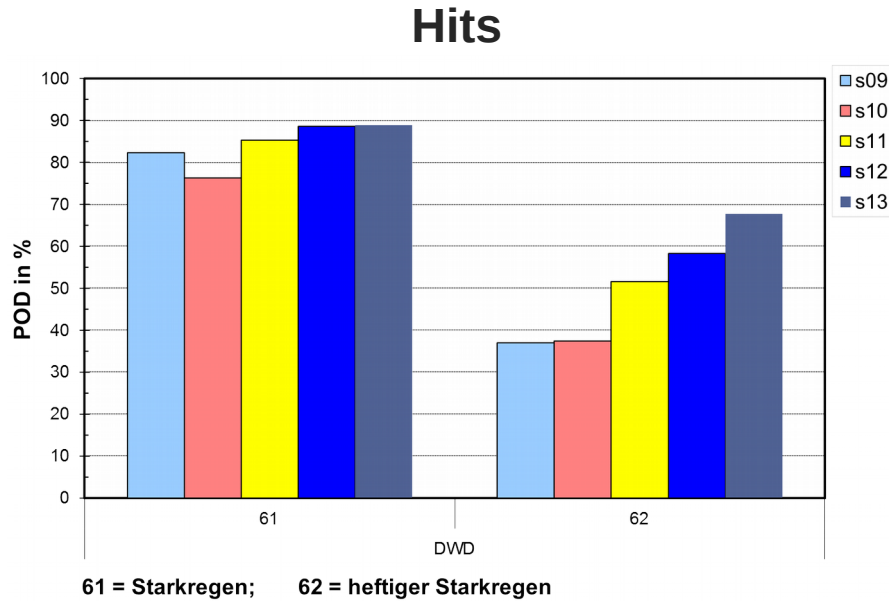


Short-Range forecasts (next day) trend-over years forecasts compared to persistence forecasts)

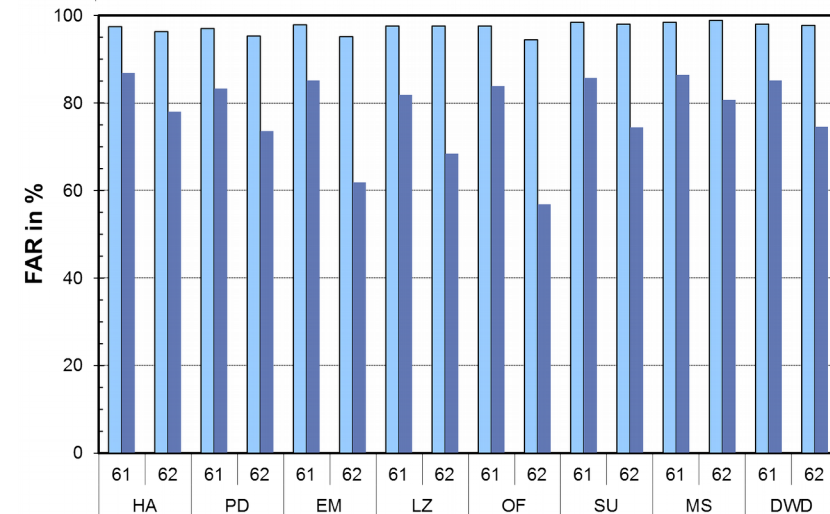




Warning Verification: hourly heavy rain



- Verification based on **station measurements**
- FAR: verification **against stations and a calibrated radar product**





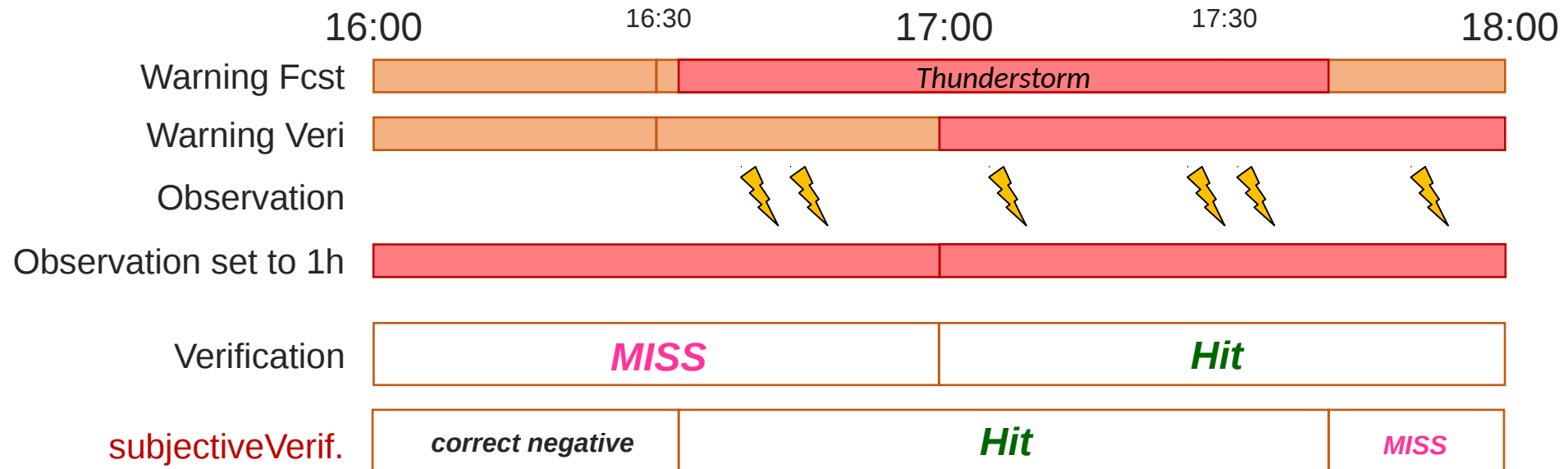
Warning Verification – current setup

DWD county warnings are verified every single hour:

	UTC 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23		
Warnung vor Sturmböen								Böen 18 - 24 m/s																		
Beobachtung (km/h)	9	10	10	11	15	17	22	25	17	19	18	17	16	16	15	14	13	12	12	12	11	14	13	12		
Prüfung Sturmböen	N	N	N	N	N	N	V	T	F	T	T	F	F	F	F	N	N	N	N	N	N	N	N	N		

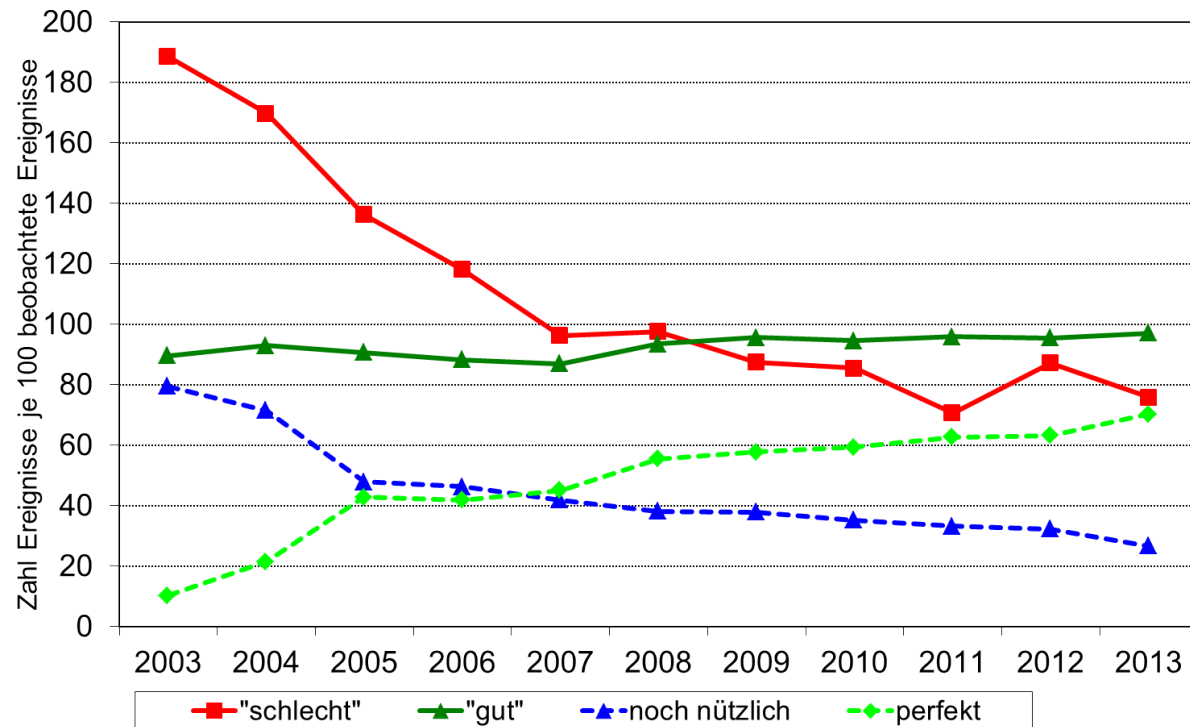
T: Hits, N: Correct Negatives, V: Misses, F: False Alarms

But in reality: warnings and observations are not well defined within fixed hours



Could look like ...

Time in advance: 0 hours



« провалы »

« общая

оправдываемость »

« попадание в 100% »

« поздний выпуск
предупреждений »

- bad** = completely missed or false alarm
- perfect** = hit and short warning length
- still usefull** = warning but a bit too late / hit but warning too long
- good** = **perfect** + **still usefull**

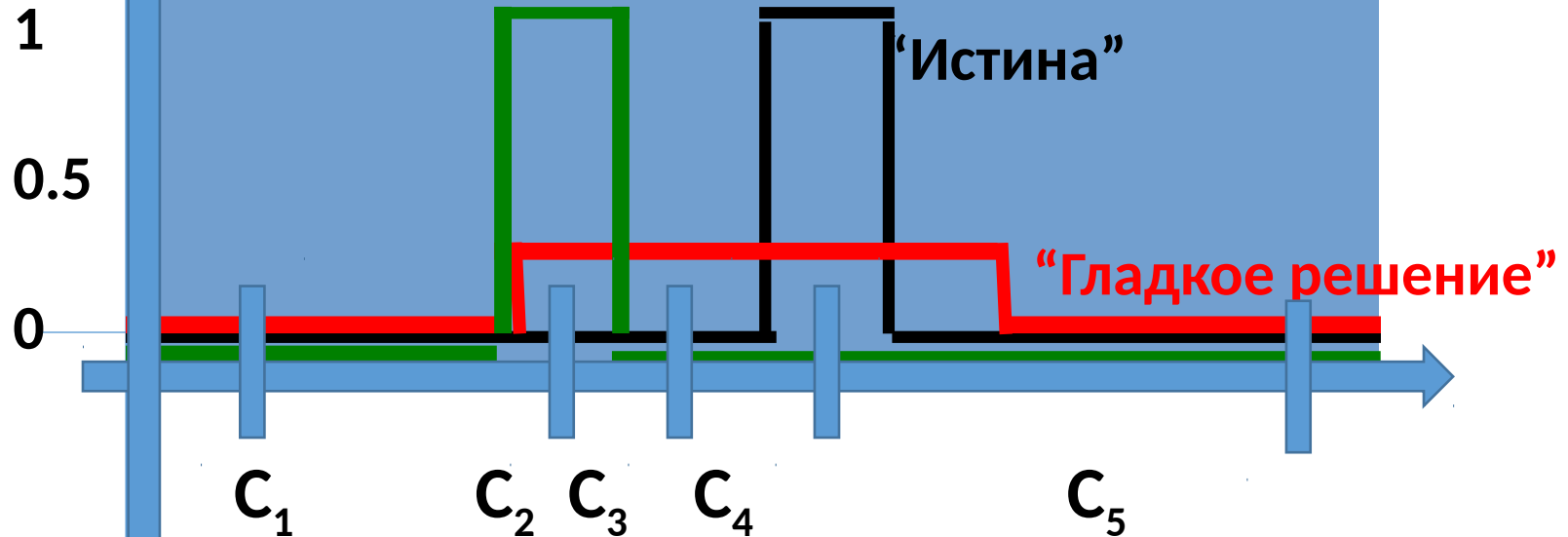


Оценка мезомасштабных прогнозов: «двойной штраф»

Более грубые модели «лучше» по метрическим показателям, чем более тонкие модели!

- Проблема «Double Penalty» и что есть «истина» в верификации?**
- Новые методы верификации в мезомасштабах с конца прошлого века.**
- Основной объект - разрывные поля осадков, «голловная боль» для наблюдения, моделирования и верификации**

ДЕМОНСТРАЦИЯ РЕЗУЛЬТАТА ФОРМАЛЬНОЙ ОЦЕНКИ “ОСАДКОВ” ДЛЯ МЕЗОПРОГНОЗА



Макс. ошибка “мезопрогноза” равна 1

Макс. ошибка “гладкого решения” равна 0.75.

Среднекв. ошибка “мезопрогноза” $\sqrt{(0 + 1^2 + 0 + 1^2 + 0) / 5} = \sqrt{0.4} \approx 0.64$

Среднекв. ошибка “гладкого прогноза”

$$\sqrt{(0 + 0.25^2 + 0.25^2 + 0.75^2 + 0) / 5} = \sqrt{0.1375} = 0.3708$$

Макс. ошибка “мезопрогноза” равна 1

Макс. ошибка “гладкого решения” равна 0.75.

Среднеквадратическая . ошибка “мезопрогноза”

$$\sqrt{(0 + 1^2 + 0 + 1^2 + 0) / 5} = \sqrt{0.4} \approx 0.64$$

Среднеквадратическая ошибка “гладкого прогноза”

$$\sqrt{(0 + 0.25^2 + 0.25^2 + 0.75^2 + 0) / 5} = \sqrt{0.1375} = 0.3708$$

Итак, среднеквадратическая . ошибка “мезопрогноза” в

$$0.64 / 0.37 \approx 1.73$$

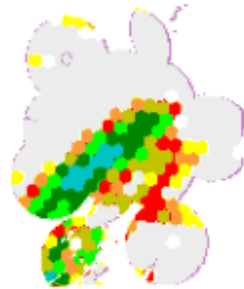
МЕНЬШЕ среднеквадратической ошибки “гладкого прогноза” в то время как для синоптика более полезным является “мезопрогноз”, содержащий более ценную информацию!



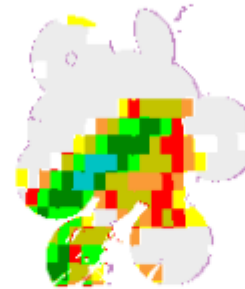
Объектно-ориентированные методы верификации

- Подход основан не на сравнении «точка в точку», а на поиске спрогнозированного объекта в «окне» вблизи наблюдаемого
- «Fuzzy verification»

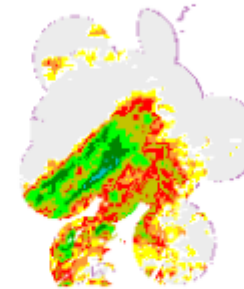
GME/I192F AV: 0.72 MA: 32.0 STD: 1.18



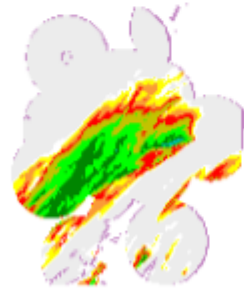
GME/R192F AV: 0.72 MA: 25.4 STD: 1.22



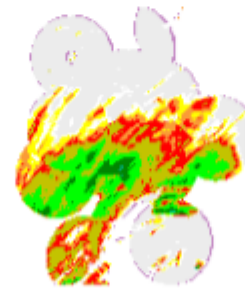
COSMO/LME AV: 0.44 MA: 30.6 STD: 1.07



COSMO/LMK AV: 0.48 MA: 31.0 STD: 1.07



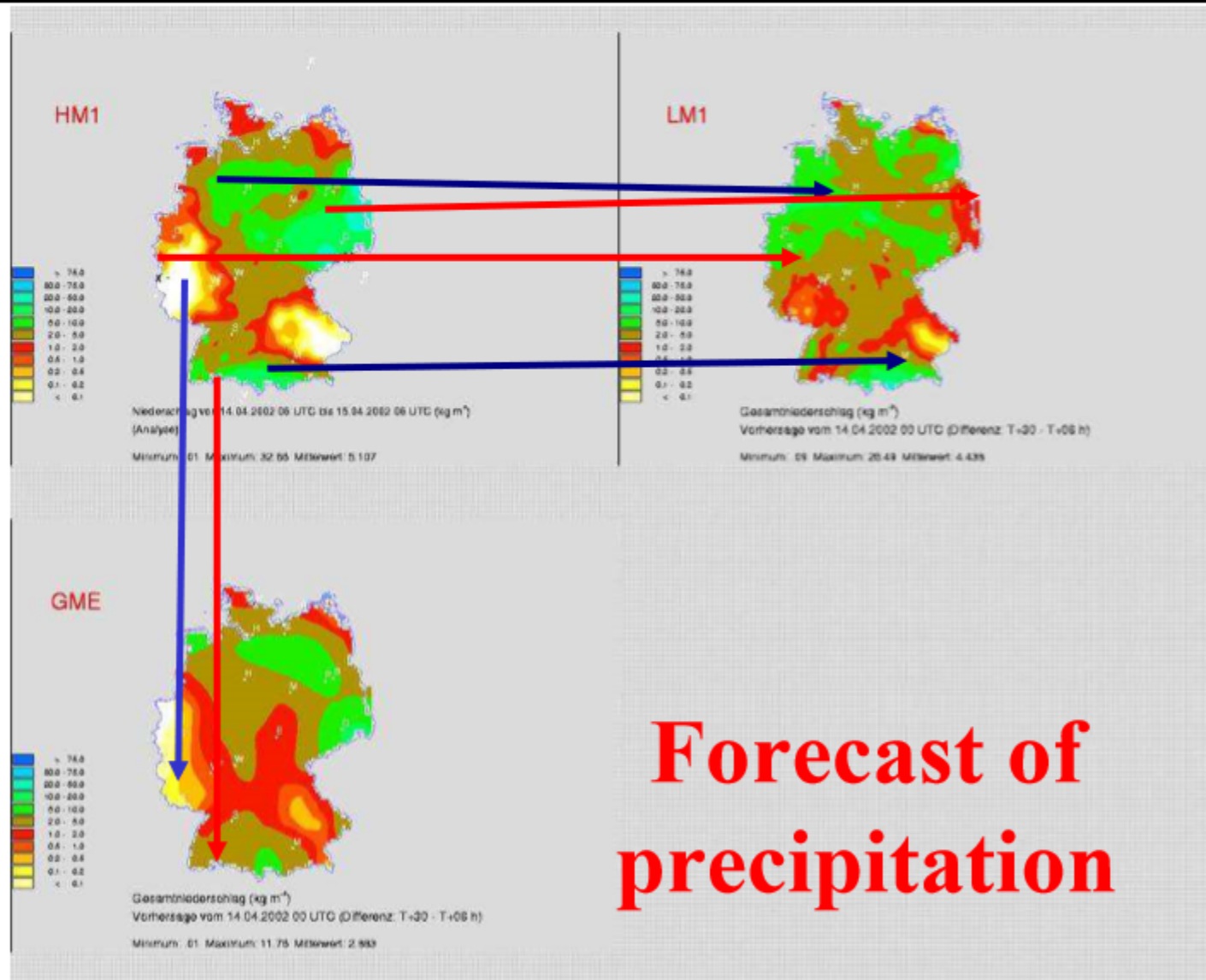
RADAR AV: 0.54 MA: 38.8 STD: 1.19



0.2
0.5
1.0
2.0
5.0
10.0
20.0
50.0
75.0
400.0

Basic data
(example with some
missing observations)

Forecasts of precipitation Start: 19.07.2007 00 UTC W=18-06

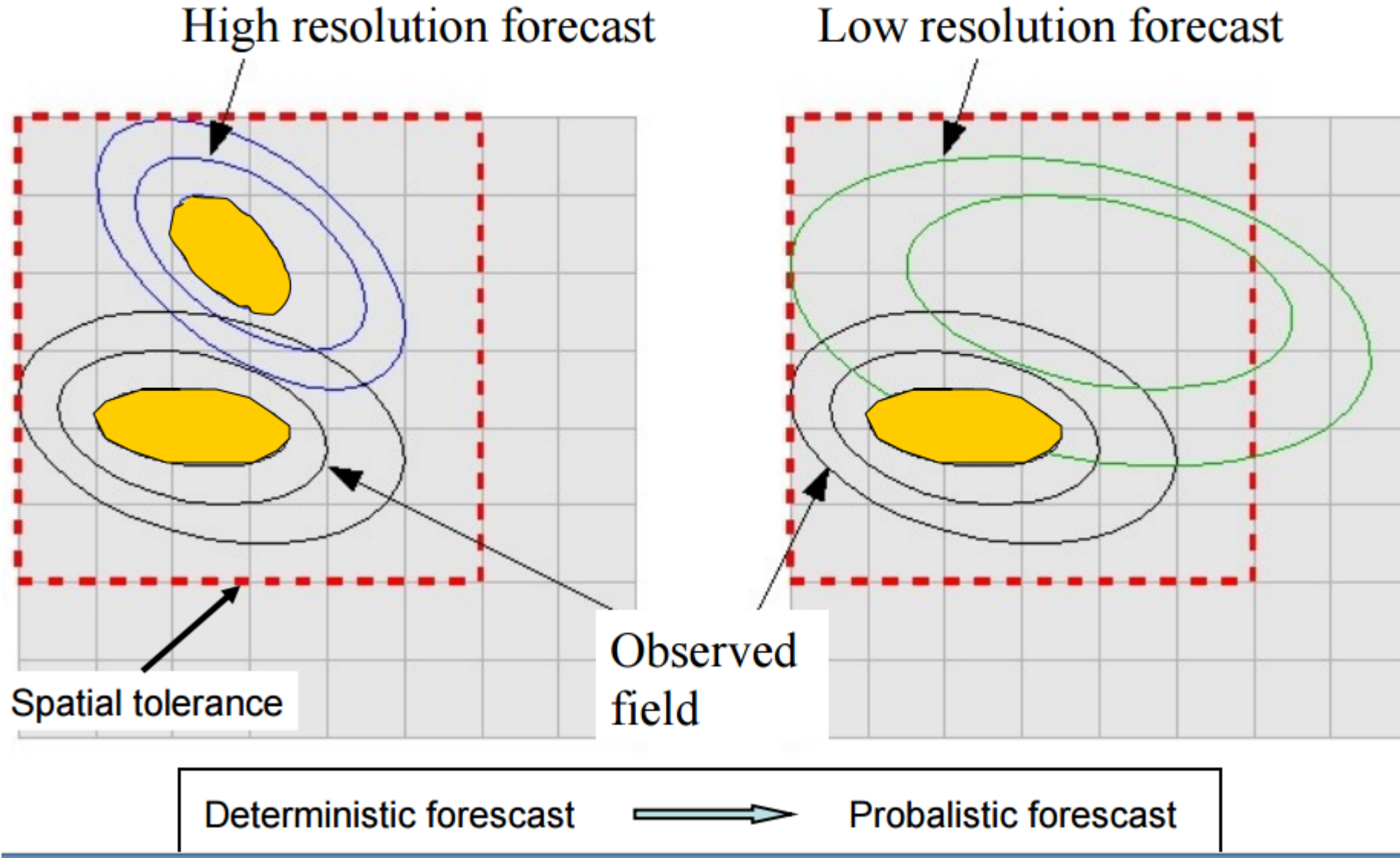


Forecast of precipitation





Проблемы верификации прогнозов высокого разрешения



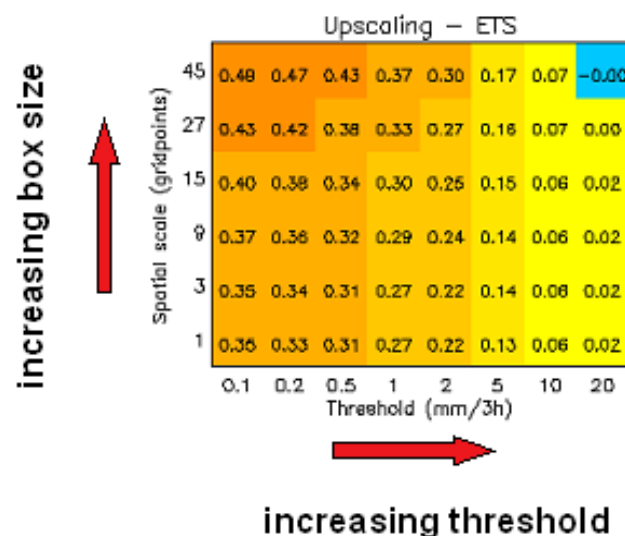
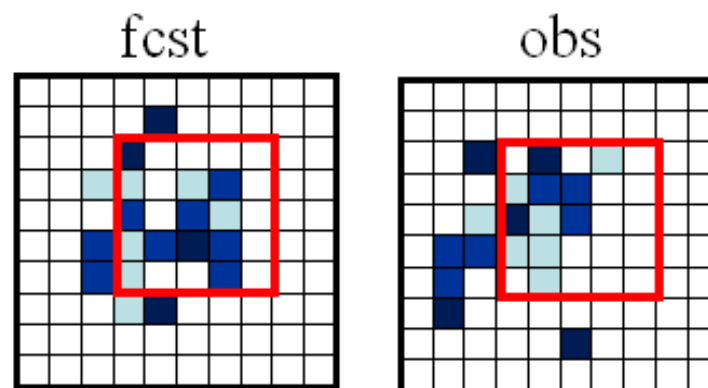


Fuzzy Verification

- „multi-scale, multi-intensity approach“
- „Fuzzy verification toolbox“ of B. Ebert
- Two methods
 - Upscaling (UP)
 - Fraction Skill Score (FSS)
- present output scale dependent
- standard setting: 3h accumulations, Jun-Nov 2007, vs Swiss radar data

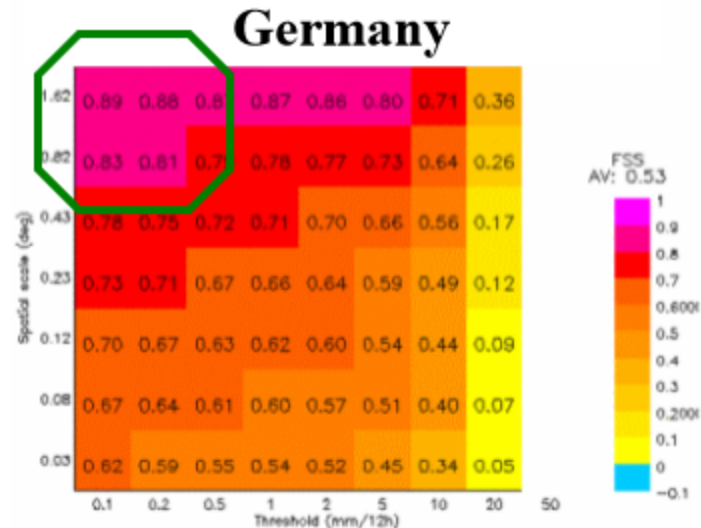
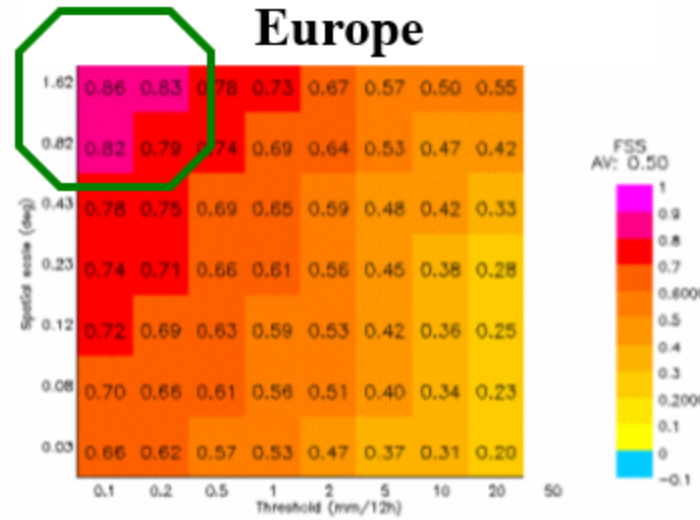
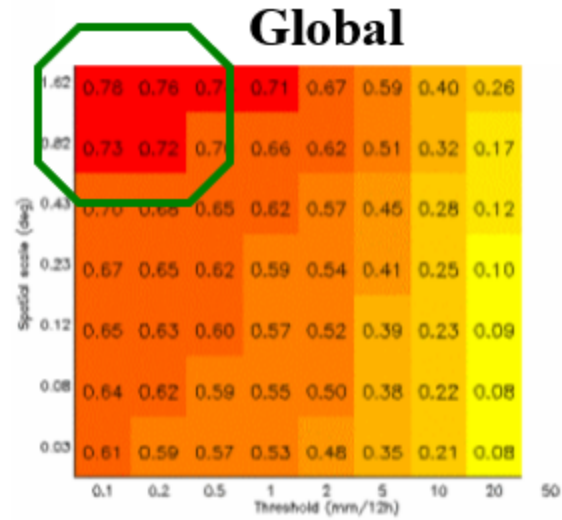
COSMO-2 (2.2km): leadtimes 03-06

COSMO-7 (6.6km): leadtimes 03-06,06-09,09-12,12-15





Deutscher Wetterdienst



Fractions skill score for forecasts of GME, COSMO-EU and COSMO-DE for August 2008, forecast time 06-18 hours



Хотелось бы добиться от верификации

- Объективной оценки для редких событий
- Минимальной зависимости от пространственно-временных масштабов явления
- Минимальной зависимости от пространственного разрешения вычислений
- Учета как хитов, так и ложных тревог
- Быстрой сходимости на небольших выборках
- Сопровождения мерой доверия

ДО НАСТОЯЩЕГО ВРЕМЕНИ НИ ОДНА СИСТЕМА КРИТЕРИЕВ НЕ УДОВЛЕТВОРЯЕТ ЭТИМ ТРЕБОВАНИЯМ!



Спасибо за внимание!