Verification of Categorical Forecasts

Pertti Nurmi
Meteorological Research & Development
Finnish Meteorological Institute

Categorical forecasts of categorical variables

⇔ **Binary** ***(Dichotomous; Yes/No)***

Examples

- ✓ **Rain** vs. no rain
- ✓ **Snowfall** vs. no snowfall
- ✓ **Strong winds** vs. no strong wind
- ✓ **Night frost** vs. no frost
- ✓ **Fog** vs. no fog

Categorical forecasts of categorical variables

Examples

- ✓ Rain
- ✓ Snowfall
- ✓ Strong winds

{ with various thresholds }

⇔ **Binary** ***(Dichotomous; Yes/No)***

⇔ **Multi-category**

MSG Modules - Netscape Browser

http://www.eumetcal.org.uk/eumetcal/verification/www/english/courses/msgcrs/index.htm

Advertisement

User Unknown

Forecast Verification

Issued April 2006

Welcome to the modules on forecast verification. The modules are designed both for users of verification results, who wish to understand what the results really mean, and those who wish to dabble in verification methodology themselves.

There are 4 modules in this course. The introductory module covers general issues about reasons for verifying, and the different types of forecast and observation data used in verification. The other modules are organized by type of forecast. If you are new to the subject of verification, then it is highly recommended that you complete the introductory module, which will help put the other modules into better perspective. If, on the other hand, you already know what a "deterministic forecast of a continuous variable", or a "probability forecast of a categorical variable" is, then feel free to skip the generalities of the introduction and go directly to the modules on the various forecast types.

You can return to this index page at any point by clicking on the icon on the left hand side which looks like this:

Module Index

- Introduction
- Verification of continuous variables
- Verification of probability forecasts
- Verification of categorical forecasts

Issued Jan 2007 !

www.eumetcal.org.uk/eumetcal/verification/www/english/courses/msgcrs/index.htm

Contingency table

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	Hit	False alarm	Fc Yes
No	Miss	Corr. non-event	Fc No
Marginal total	Obs Yes	Obs No	Sum total

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

Artificial data

Date	24h forecast	Observation
Jan 1	0.3	-
Jan 2	0.1	Yes
Jan 3	0.1	-
Jan 4	0.7	-
Jan 5	0.8	-
. . .	. . .	. . .
Dec 27	0.8	Yes
Dec 28	1.0	Yes
Dec 29	0.9	Yes
Dec 30	0.1	Yes
Dec 31	0.1	-

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

Verification history, Tornados in the U.S., 1884

(slightly modified Finley case)

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

$$\frac{2680 + 30}{2800} = 96{,}8\ \%$$

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

96,8 %

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	0	0	0
No	50	2750	2800
obs Σ	50	2750	2800

Never forecast a Tornado

$$\frac{2750 + 0}{2800} = 98,2\ \%$$

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	0	0	0
No	50	2750	2800
obs Σ	50	2750	2800

98,2 % **NO false alarms – NO hits – HEAVY underforecasting !**

Back to the original results:

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

$\frac{30}{50}$ = 60% hits

$\frac{70}{100}$ = 70% false alarms

$\frac{100}{50}$ = 2* overforecasting

96,8 %

Scores ...1

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

"Finley"

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Bias aka **Frequency Bias Index**

B = FBI = (a + b) / (a + c) [~ Total fc Yes / Total obs Yes]

B = 2.00
PC = 0.97

- With B > 1 , the event is overforecast.
- With B < 1 , the event is underforecast.

Range: 0 to oo
Perfect score = 1

Proportion Correct

PC = (a + d) / n [~ (Hits + Correct non-events) / Total sample size]

- Most simple and intuitive performance measure.
- Usually very misleading because rewards correct "Yes" and "No" forecasts equally.
- Can be maximized by forecasting the most likely event all the time.

⇔ Strongly influenced by the more common event.

- **Never for extreme event verification – Remember "Finley" !!!**

Range: 0 to 1
Perfect score = 1

Scores ...2

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

"Finley"

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

B = 2.00
PC = 0.97
POD = 0.60
FAR = 0.70

Probability Of Detection, Hit rate, Prefiguration

POD = H = a / (a + c) [~ Hits / Total obs Yes]

- Sensitive to missed events only, not false alarms.
- Can be artificially improved by overforecasting.
- Complement score Miss Rate, **MR** = 1 – H = c / (a+c)
- **Should be examined together with ...**

Range: 0 to 1
Perfect score = 1

False Alarm Ratio

FAR = b / (a + b) [~ False alarms / Total fc Yes]

- Sensitive to false alarms only, not missed events.
- Can be artificially improved by underforecasting.
- Improving POD is achieved by worsening FAR, and vice versa.

Range: 0 to 1
Perfect score = 0

Scores ...3

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

"Finley"

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

B = 2.00
PC = 0.97
POD = 0.60
FAR = 0.70
PAG = 0.30
F = 0.03

Post agreement

PAG = a / (a + b) [~ Hits / Total fc Yes]

- Complement of FAR (i.e. = 1 – FAR).
- Sensitive to false alarms, not misses.
- Not widely used.

Range: 0 to 1
Perfect score = 1

False Alarm Rate, Probability of False Detection (POFD)

F = b / (b + d) [~ False alarms / Total obs No]

- False alarms, given the event did not occur (obs No).
- Sensitive to false alarms only, not missed events.
- Can be artificially improved by underforecasting.
- **Generally used with H** (or POD) **to produce the ROC score for probability forecasts!**
- **Otherwise rarely used.**

Range: 0 to 1
Perfect score = 0

Frequency Bias, B ~ FBI

Reference: EUMETCAL Verification module

1 / 7

Below are two contingency tables representing two completely different sets of forecasts. The left hand table is for 151 Gale forecasts, and the right hand table is for 2800 forecasts of the occurrence of tornados in the US. Test your understanding of frequency bias using the question below the tables:

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Which of the following statements correctly describes the bias of the two sets of forecasts?

- Gales and tornados are underforecast
- Gales are underforecast and tornados are overforecast
- Gales are overforecast and tornados are underforecast
- Gales and tornados are overforecast

Frequency Bias, B ~ FBI

Reference: EUMETCAL Verification module

1 / 7

Below are two contingency tables representing two completely different sets of forecasts. The left hand table is for 151 Gale forecasts, and the right hand table is for 2800 forecasts of the occurrence of tornados in the US. Test your understanding of frequency bias using the question below the tables:

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Which of the following statements correctly describes the bias of the two sets of forecasts?

- ✗ Gales and tornados are underforecast
- ✓ Gales are underforecast and tornados are overforecast
- ✗ Gales are overforecast and tornados are underforecast
- ✗ Gales and tornados are overforecast

Feedback

Yes, correct: For gales, B=17/26 which is less than 1 while for tornados, B= 100/50 which is greater than 1.

Hit rate *aka* Probability Of Detection, H ~ POD / Proportion Correct, PC

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Question: Determine the PC and the hit rate for the two contingency tables shown above by dragging the correct answer to the appropriate boxes in the table.

	Gale Forecasts	Tornado Forecasts
Hit Rate =		
Proportion Correct =		

0.30
0.40
0.42
0.58
0.60
0.88
0.91
0.97

Reference: EUMETCAL Verification module

Scores … Answer_2

Hit rate *aka* Probability Of Detection, H ~ POD / Proportion Correct, PC

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Question: Determine the PC and the hit rate for the two contingency tables shown above by dragging the correct answer to the appropriate boxes in the table.

correct

	Gale Forecasts	Tornado Forecasts
Hit Rate =	0.58	0.60
Proportion Correct =	0.91	0.97

0.30

0.40

0.42

0.88

Reference: EUMETCAL Verification module

False Alarm Ratio, FAR / False Alarm Rate, F

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Question: Determine the false alarm ratio and false alarm rate for the two contingency tables shown above by dragging the correct answer to the appropriate boxes in the table.

	Gale Forecasts	Tornado Forecasts
False alarm ratio =		
False alarm rate =		

0.02
0.03
0.12
0.30
0.70
0.88

Reference: EUMETCAL Verification module

False Alarm Ratio, FAR / False Alarm Rate, F

Gale forecast	Gale observed Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Question: Determine the false alarm ratio and false alarm rate for the two contingency tables shown above by dragging the correct answer to the appropriate boxes in the table.

correct

	Gale Forecasts	Tornado Forecasts
False alarm ratio =	0.12	0.70
False alarm rate =	0.02	0.03

0.30

0.88

Reference: EUMETCAL Verification module

Scores …4

Categorical Forecasts

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

"Finley"

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

B = 2.00
PC = 0.97
POD = 0.60
FAR = 0.70
PAG = 0.30
F = 0.03
TS = 0.25
ETS = 0.24

Threat Score, Critical Success Index

$$TS = CSI = a / (a + b + c)$$

Range: 0 to 1
Perfect score = 1
No skill level = 0

- **Simple popular measure of rare events**.
- Takes into account hits, false alarms and misses.

  ⇔ More balanced than POD or FAR.
- Correct (simple) "no" forecasts not considered.
- Sensitive to climatological frequency of event ⇔ Poorer scores for rare events

Equitable Threat Score, Gilbert Skill Score (GSS)

$$ETS = (a - a_r) / (a + b + c - a_r)$$

where $a_r = (a + b) (a + c) / n$

… is the number of hits due to random forecasts (chance).

⇔ Simple TS may include hits due to random chance.

Range: -1/3 to 1
Perfect score = 1
No skill level = 0

Threat Score *aka* Critical Success Index, TS ~ CSI / Equitable Threat Score, ETS

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Determine the CSI, the number of hits by chance (a_r) and the ETS for the two contingency tables shown above by dragging the correct answer to the appropriate boxes in the table.

	Gale Forecasts	Tornado Forecasts
Threat Score		
Hits by chance =		
Equitable Threat Score =		

0.24
0.25
0.43
0.48
0.54
0.60
1.79
2.93

Reference: EUMETCAL Verification module

NB: Perhaps skip the calculations and go directly to Exercise 4b

Threat Score *aka* Critical Success Index, TS ~ CSI / Equitable Threat Score, ETS

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Determine the CSI, the number of hits by chance (a_r) and the ETS for the two contingency tables shown above by dragging the correct answer to the appropriate boxes in the table.

correct

	Gale Forecasts	Tornado Forecasts
Threat Score	0.54	0.25
Hits by chance =	2.93	1.79
Equitable Threat Score =	0.48	0.24

0.43

0.60

Reference: EUMETCAL Verification module

Threat Score *aka* Critical Success Index, TS ~ CSI / Equitable Threat Score, ETS

	Gale Forecasts	Tornado Forecasts
Threat Score	0.54	0.25
Hits by chance =	2.93	1.79
Equitable Threat Score =	0.48	0.24

Using the results obtained above, select the correct answer to the following question

How did the value of the ETS change with respect to the TS?

- For both gales and tornados, the ETS is smaller than the TS
- The ETS for gales is higher, lower for tornados
- The ETS for gales is lower, but higher for tornados
- The ETS is higher for both gales and tornados

Reference: EUMETCAL Verification module

Threat Score *aka* Critical Success Index, TS ~ CSI / Equitable Threat Score, ETS

	Gale Forecasts	Tornado Forecasts
Threat Score	0.54	0.25
Hits by chance =	2.93	1.79
Equitable Threat Score =	0.48	0.24

Using the results obtained above, select the correct answer to the following question

How did the value of the ETS change with respect to the TS?

- ✓ For both gales and tornados, the ETS is smaller than the TS
- ✗ The ETS for gales is higher, lower for tornados
- ✗ The ETS for gales is lower, but higher for tornados
- ✗ The ETS is higher for both gales and tornados

Feedback

Correct. Actually, the ETS must always decrease because the number correct by chance is subtracted from both numerator and denominator

Reference: EUMETCAL Verification module

Threat Score *aka* Critical Success Index, TS ~ CSI / Equitable Threat Score, ETS

	Gale Forecasts	Tornado Forecasts
Threat Score	0.54	0.25
Hits by chance =	2.93	1.79
Equitable Threat Score =	0.48	0.24

Looking at the table, are the following statements true or false?

	True	False
The number correct by chance is greater for gales than for tornados	?	?
The decrease in the ETS, when compared to the CSI, is greater for the gales than for the tornados	?	?

Reference: EUMETCAL Verification module

Threat Score *aka* Critical Success Index, TS ~ CSI / Equitable Threat Score, ETS

	Gale Forecasts	Tornado Forecasts
Threat Score	0.54	0.25
Hits by chance =	2.93	1.79
Equitable Threat Score =	0.48	0.24

Looking at the table, are the following statements true or false?

	True	False
The number correct by chance is greater for gales than for tornados	✓	X
The decrease in the ETS, when compared to the CSI, is greater for the gales than for the tornados	✓	X

Feedback

Yes, 2.93 vs. 1.79. Tornados are a rare event, so the chance of guessing the occurrence of a tornado correctly is lower.

Feedback

Correct. The ETS is about .06 lower for the gales and only .01 lower for the tornadoes. Since the TS is typically lower for rare events than for more common events for a particular hit rate (note the hit rates are nearly equal), the adjustment for chance forecasts helps offset this systematic tendency.

Scores ...5

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

"Finley"

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Hanssen & Kuiper's Skill Score, True Skill Statistics

$$KSS = TSS = H - F$$

$$= (ad - bc) / [(a+c)(b+d)]$$

Range: -1 to 1
Perfect score = 1
No skill level = 0

B = 2.00
PC = 0.97
POD = 0.60
FAR = 0.70
PAG = 0.30
F = 0.03
TS = 0.25
ETS = 0.24
KSS = 0.57
HSS = 0.39

- **Popular combination skill score of H and F.**
- Measures ability to separate "yes" cases (H) from "no" cases (F).
- For rare events, **d** cell is high => F small => KSS close to POD.
- Related to the Relative Operating Characteristic (ROC) ⬄ Probability forecasts

Heidke Skill Score

$$HSS = \{ PC - \mathit{ref} \} / \{ 1 - \mathit{ref} \}$$

$$= \{ (a + d) / n - [(a+b)*(a+c) + (b+d)*(c+d)] / n^2 \} / \{ 1 - [(a+b)*(a+c) + (b+d)*(c+d)] / n^2 \}$$

Range: - oo to 1
Perfect score = 1
No skill level = 0

Scores ...5

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

"Finley"

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Hanssen & Kuiper's Skill Score, True Skill Statistics

KSS = TSS = H – F

= (ad – bc) / [(a+c) (b+d)]

Range: -1 to 1
Perfect score = 1
No skill level = 0

- **Popular combination skill score of H and F.**
- Measures ability to separate "yes" cases (H) from "no" cases (F).
- For rare events, **d** cell is high => F small => KSS close to POD.
- Related to the Relative Operating Characteristic (ROC) ⇔ Probability forecasts

B = 2.00
PC = 0.97
POD = 0.60
FAR = 0.70
PAG = 0.30
F = 0.03
TS = 0.25
ETS = 0.24
KSS = 0.57
HSS = 0.39

Heidke Skill Score *(in a simplified, calculation-friendly form)*

HSS = 2 (ad – bc) / [(a + c)(c + d) + (a + b)(b + d)]

- **One of the most popular skill measures for categorical forecasts.**
- Measures fractional improvement over random chance.
- Can be compared on different datasets

Range: - oo to 1
Perfect score = 1
No skill level = 0

Hansen & Kuiper's Skill Score *aka* True Skill Satistics, KSS ~ TSS

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Question: Determine the KSS values for the tornado and gale forecasts and fill them in by dragging the correct value to its place in the table below.

	Gale Forecasts	Tornado Forecasts
Hanssen-Kuiper Skill Score =		

0.02

0.03

0.41

0.52

0.56

0.57

Reference: EUMETCAL Verification module

Hansen & Kuiper's Skill Score *aka* True Skill Satistics, KSS ~ TSS

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Question: Determine the KSS values for the tornado and gale forecasts and fill them in by dragging the correct value to its place in the table below.

correct

	Gale Forecasts	Tornado Forecasts
Hanssen-Kuiper Skill Score =	0.56	0.57

0.02

0.03

0.41

0.52

Reference: EUMETCAL Verification module

Scores …6

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

"Finley"

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

Odds ratio

OR = a d / b c

Range: 0 to oo
Perfect score = oo
No skill level = 1

Measures forecasts' probability (odds) to score a hit (H) as compared to making a false alarm (F):

$$OR = [H / (1 - H)] / [F / (1 - F)]$$

– Independent of potential biases between observations and forecasts.

B = 2.00
PC = 0.97
POD = 0.60
FAR = 0.70
PAG = 0.30
F = 0.03
TS = 0.25
ETS = 0.24
KSS = 0.57
HSS = 0.39
OR = 57.43
ORSS = 0.97

Transformation into a skill score, ranging from -1 to +1:

ORSS = (ad – bc) / (ad + bc)

= (OR – 1) / (OR + 1)

– Produces typically very high absolute skill values, due to definition.

– Very little used in meteorological forecast verification.

"Summary"

Reference: EUMETCAL Verification module

Rain forecast	Rain observed		
	Yes	No	fc Σ
Yes	52	45	97
No	22	227	249
obs Σ	74	272	346

~>

Scores

B	= 1.31	TS	= 0.44
PC	= 0.81	ETS	= 0.32
POD	= 0.70	KSS	= 0.53
FAR	= 0.46	HSS	= 0.48
PAG	= 0.54		
F	= 0.17		

Which of the following statements about the verification scores are true.

	True	False
Rain is a frequent event at this station	?	?
Rain was overforecast at this station	?	?
The high frequency of forecasting of rain has led to a high false alarm rate	?	?
The PC is high (0.81) because forecasting for this dry location is easy	?	?
The POD is high (0.7) because forecasting for this dry location is easy	?	?
The forecasts were skilful on average	?	?

"Summary"

Reference: EUMETCAL Verification module

Rain forecast	Rain observed		
	Yes	No	fc Σ
Yes	52	45	97
No	22	227	249
obs Σ	74	272	346

~>

Scores

B = 1.31	TS = 0.44
PC = 0.81	ETS = 0.32
POD = 0.70	KSS = 0.53
FAR = 0.46	HSS = 0.48
PAG = 0.54	
F = 0.17	

Which of the following statements about the verification scores are true.

Statement	True	False
Rain is a frequent event at this station	✗	✓
Rain was overforecast at this station	✓	✗
The high frequency of forecasting of rain has led to a high false alarm rate	✗	✓
The PC is high (0.81) because forecasting for this dry location is easy	✓	✗
The POD is high (0.7) because forecasting for this dry location is easy	✗	✓
The forecasts were skilful on average	✓	✗

Feedback

Correct. Rain occurs with a frequency of only about 20% (74/346) at this station, relatively dry for Finland!

True. The frequency bias is 1.31, greater than 1, meaning overforecasting.

Correct. It could be seen that the overforecasting is accompanied by high false alarm RATIO, but the false alarm rate depends on the observation frequencies, and is low because the climate is relatively dry.

Probably true. The PC gives credit for all those "easy" correct forecasts of the non-occurrence. Such forecasts are easy when the non-occurrence is common.

The POD is high most likely because the forecaster has chosen to forecast the occurrence of the even too often, and has incurred more false alarms too.

Yes. Both the KSS and the HSS are well within the positive range. Remember, the standard for the HSS is a chance forecast, which is easy to beat.

"Summary"

Attached is a contingency table of five months of categorical warnings against gale-force winds, i.e. wind speeds exceeding 14 m/s (left). Compute the specified verification statistics. For reference, corresponding "Finlay" tornado verification statistics are shown (right). Interpret the scores and compare the two.

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

B = (a+b)/(a+c) = ____
PC = (a+d)/n = ____
POD = a/(a+c) = ____
FAR = b/(a+b) = ____
PAG = a/(a+b) = ____
F = b/(b+d) = ____
KSS = POD-F = ____
TS = a/(a+b+c) = ____
ETS = $(a-a_r)/(a+b+c-a_r)$ = ____
HSS = 2(ad-bc)/[(a+c)(c+d)+(a+b)(b+d)] = ____
OR = ad/bc = ____
ORSS = (OR-1)/(OR+1) = ____

2. 00 = B
0.97 = PC
0.60 = POD
0.70 = FAR
0.30 = PAG
0.03 = F
0.57 = KSS
0.25 = TS
0.24 = ETS
0.39 = HSS
57.43 = OR
0.97 = ORSS

"Summary"

Attached is a contingency table of five months of categorical warnings against gale-force winds, i.e. wind speeds exceeding 14 m/s (left). Compute the specified verification statistics. For reference, corresponding "Finlay" tornado verification statistics are shown (right). Interpret the scores and compare the two.

Gale forecast	Gale observed		
	Yes	No	fc Σ
Yes	15	2	17
No	11	123	134
obs Σ	26	125	151

Tornado forecast	Tornado observed		
	Yes	No	fc Σ
Yes	30	70	100
No	20	2680	2700
obs Σ	50	2750	2800

B = (a+b)/(a+c) = 0.65
PC = (a+d)/n = 0.91
POD = a/(a+c) = 0.58
FAR = b/(a+b) = 0.12
PAG = a/(a+b) = 0.88
F = b/(b+d) = 0.02
KSS = POD-F = 0.56
TS = a/(a+b+c) = 0.54
ETS = $(a-a_r)/(a+b+c-a_r)$ = 0.48
HSS = 2(ad-bc)/[(a+c)(c+d)+(a+b)(b+d)] = 0.65
OR = ad/bc = 83.86
ORSS = (OR-1)/(OR+1) = 0.98

2. 00 = B
0.97 = PC
0.60 = POD
0.70 = FAR
0.30 = PAG
0.03 = F
0.57 = KSS
0.25 = TS
0.24 = ETS
0.39 = HSS
57.43 = OR
0.97 = ORSS

Multi-category Events

- Extension of 2*2 to several (k) mutually exhaustive categories
  - ✓ Rain type: rain / snow / freezing rain (k=3)
  - ✓ Wind warnings: strong gale / gale / no gale (k=3)
- Only PC (Proportion Correct) can be directly generalized
- Other verification measures need be converted into a series of 2*2 tables
  - ✓ “Forecast event” distinct from the “non-forecast event”

Forecast	Observed			
	o_1	o_2	o_3	fc Σ
f_1	**r**	s	t	Σf_1
f_2	u	**v**	w	Σf_2
f_3	x	y	**z**	Σf_3
obs Σ	Σo_1	Σo_2	Σo_3	Σ

a = z	b= x+y
c= t+w	d= r+s+u+v

a = v	b= u+w
c= s+y	d= r+t+x+z

a = r	b= s+t
c= u+x	d= v+w+y+z

Generalization of KSS and HSS – measures of improvement over random forecasts:

$$\mathbf{KSS} = \{ \Sigma\, p(f_i, o_i) - \Sigma\, p(f_i)\, p(o_i) \} \;/\; \{ 1 - \Sigma\, (p(f_i))^2 \}$$

$$\mathbf{HSS} = \{ \Sigma\, p(f_i, o_i) - \Sigma\, p(f_i)\, p(o_i) \} \;/\; \{ 1 - \Sigma\, p(f_i)\, p(o_i) \}$$

Exercise_8:
Multi-category event

Clouds forecast	Clouds observed 0 - 2	3 - 5	6 - 8	fc Σ
0 - 2	65	10	21	96
3 - 5	29	17	48	94
6 - 8	18	10	128	156
obs Σ	112	37	197	346

~>

Cloudiness in Finland

No clouds (0-2)	Partly cloudy (3-5)	Cloudy (6-8)
B = 0.86	B = 2.54	B = 0.79
POD = 0.58	POD = 0.46	POD = 0.65
FAR = 0.32	FAR = 0.82	FAR = 0.18
F = 0.13	F = 0.25	F = 0.19
TS = 0.45	TS = 0.15	TS = 0.57

Overall: PC = 0.61 KSS = 0.41 HSS = 0.37

Multi-category contingency table of one year (with 19 missing cases) of cloudiness forecasts (left), and resulting statistics (right), exclusively for forecasts of each cloud category, together with the overall PC, KSS and HSS. **Please examine/ comment:**

Overall skill ? ____________________

____

____

Partly cloudy category ? ____________________

____

____

Answer_8:
Multi-category event

Clouds forecast	Clouds observed 0 - 2	3 - 5	6 - 8	fc Σ
0 - 2	65	10	21	96
3 - 5	29	17	48	94
6 - 8	18	10	128	156
obs Σ	112	37	197	346

~>

Cloudiness in Finland

No clouds (0-2)	Partly cloudy (3-5)	Cloudy (6-8)
B = 0.86	B = 2.54	B = 0.79
POD = 0.58	POD = 0.46	POD = 0.65
FAR = 0.32	FAR = 0.82	FAR = 0.18
F = 0.13	F = 0.25	F = 0.19
TS = 0.45	TS = 0.15	TS = 0.57

Overall: PC = 0.61 KSS = 0.41 HSS = 0.37

Multi-category contingency table of one year (with 19 missing cases) of cloudiness forecasts (left), and resulting statistics (right), exclusively for forecasts of each cloud category, together with the overall PC, KSS and HSS.

Overall skill ?

- Both skill scores relatively high, c. 0.4
- Most (90% of the) cases in ”no cloud” or ”cloudy” category
- Neither score considers relative sample probabilities

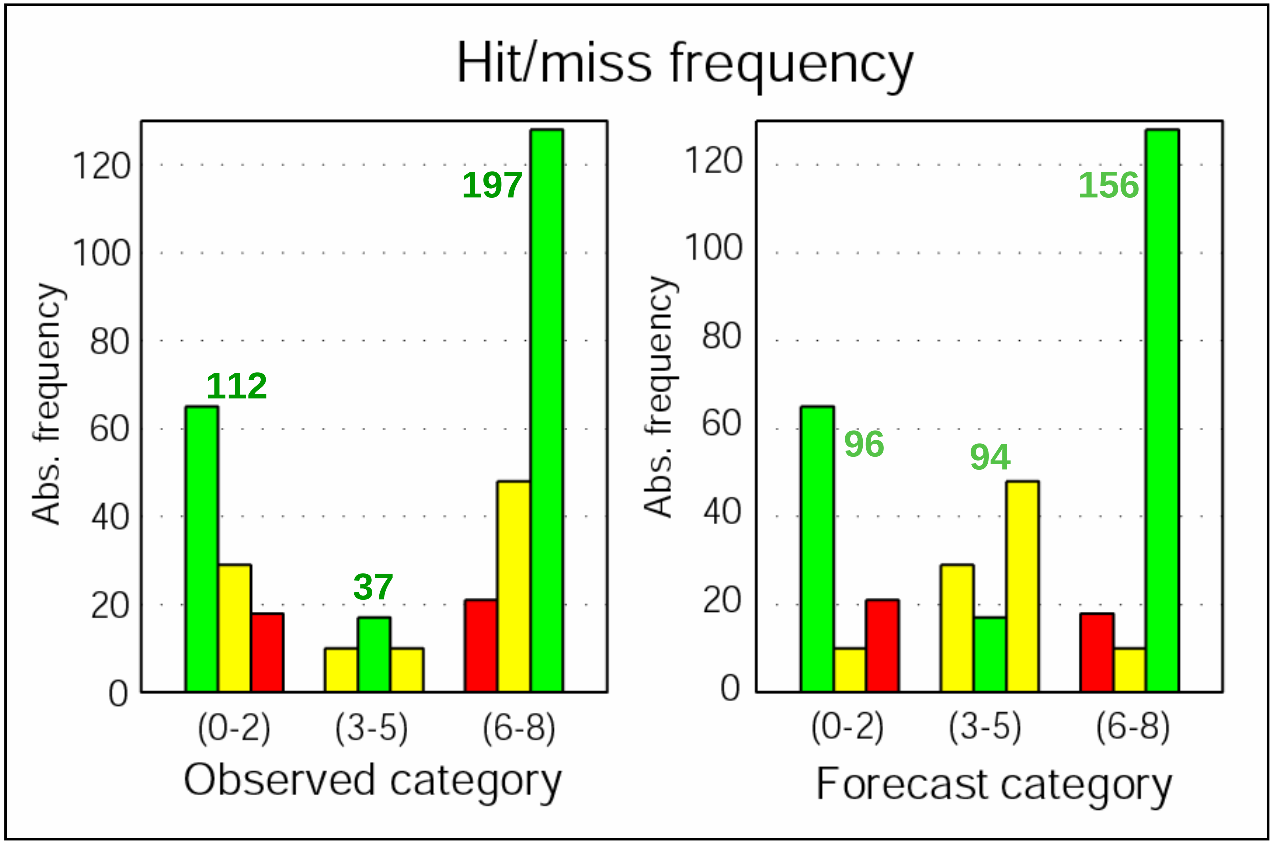
Partly cloudy category ?

- Very strong overforecasting, B = 2.5
- Numerous false alarms, FAR = 0.8
- Despite of above, poor detection of event, POD c. 0.5

Exercise_8b: Multi-category event, *cont'd…*

Visualization =>

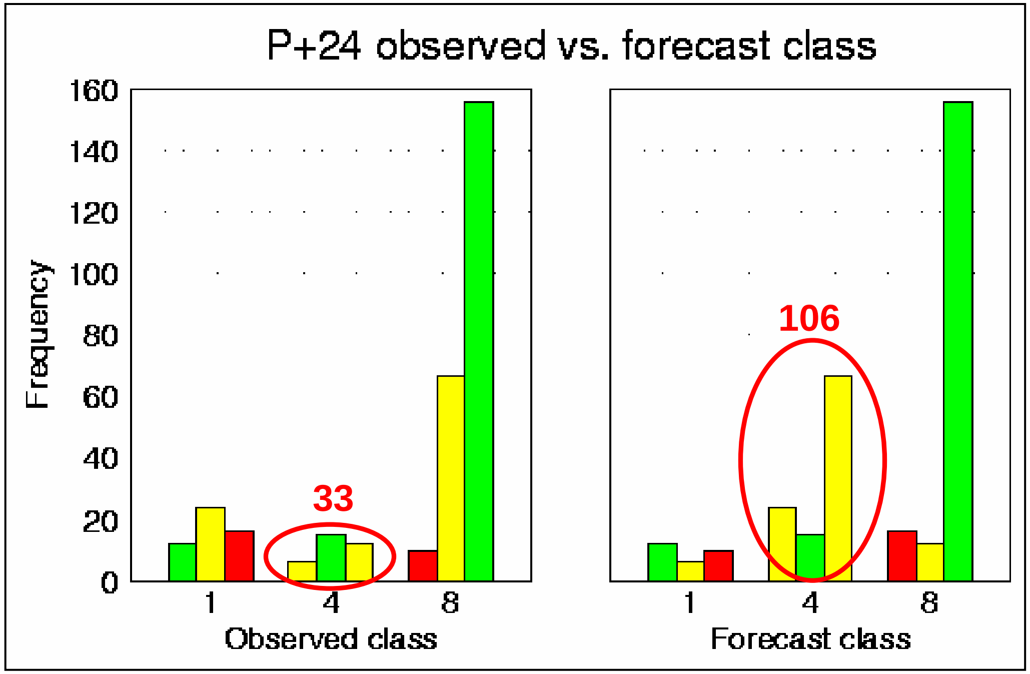
Clouds forecast	Clouds observed			
	0 - 2	3 - 5	6 - 8	fc Σ
0 - 2	65	10	21	96
3 - 5	29	17	48	94
6 - 8	18	10	128	156
obs Σ	112	37	197	346



Previous data transformed into hit/miss bar charts, either given the observations (left), or given the forecasts (right). The green, yellow and red bars denote correct and one and two category errors, respectively.

- ✓ U-shape in observations evident (left)
- ✓ No hint of U-shape in forecast distribution (right).

Exercise_9:
Multi-category event #2



… Fcs from Finland, *again !*

Observation

Forecaster	0–2	3–5	6–9	
1	12	6	10	28
4	24	15	67	106
8	16	12	156	184
	52	33	233	**318**

PC = 57.5
KSS = 24.1

Class	1	4	8
BIAS	0.5	3.2	0.8
TS	17.6	12.1	59.8
POD	23.1	45.5	67.0
FAR	57.1	85.8	15.2

Summary

- Verify a comprehensive set of categorical events
  - ✓ Compile relevant contingency tables
  - ✓ Cover, if possible, multi-category events
  - ✓ Focus on adverse and/or extreme local weather
- "Stratify & Aggregate"

  \+ Compute **FBI**, PC, **POD & FAR**, F, PAG, **TS**, **ETS**, **KSS**, **HSS**
- Additionally, compute OR, ORSS, ROC

Examples

- ✓ **Rain (vs. no rain)**; with various rainfall thresholds
- ✓ **Snowfall**; with various thresholds
- ✓ **Strong winds** (vs. no strong wind); with various wind force thresholds
- ✓ **Night frost** (vs. no frost)
- ✓ **Fog** (vs. no fog)

References

- ✓ **Jolliffe and Stephenson, Eds. (2003): Forecast Verification: A Practitioner's Guide, *Wiley & Sons.***
- ✓ **Nurmi (2003): Recommendations on the verification of local weather forecasts. ECMWF Tech. Mem. 430.**
- ✓ **Wilks (2005): Statistical Methods in the Atmospheric Sciences, Chapter 7** (Forecast Verification)**. *Academic Press.***
- ⇨ www.eumetcal.org.uk/eumetcal/verification/www/english/courses/msgcrs/index.htm
- ⇨ www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html

Thank You

[Pertti Nurmi]

Meteorological Research & Development

Finnish Meteorological Institute