

Interpretation und Bewertung von probabilistischen Prognosen

Prof. Leonhard Held
Abteilung Biostatistik
Institut für Sozial- und Präventivmedizin
Universität Zürich

27. Mai 2009



Über Prognosen

- “Prognosen sind schwierig, besonders wenn sie die Zukunft betreffen” (zugeschrieben Karl Valentin, Mark Twain, Winston Churchill u.a.)
- “Die beste Art, die Zukunft vorauszusagen, ist, die Zukunft zu erfinden” (Alan Kay, Computerwissenschaftler)
- “Ein Prognostiker ist ein Mann, der in lichten Momenten düstere Ahnungen hat” (Tennessee Williams)



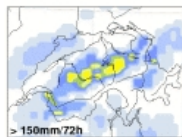
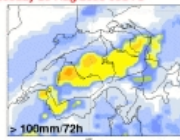
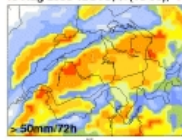
Probabilistische versus deterministische Prognosen

- Champions League Finale:
FC Barcelona - Manchester United
Wer gewinnt?
 - Ein Anbieter bietet für EUR 1 Einsatz eine Auszahlung von EUR 2.55, EUR 3.35, EUR 2.55 bei Sieg, Unentschieden oder Niederlage von Barcelona nach 90 Minuten.
Dies entspricht den Wahrscheinlichkeiten 0.36, 0.28 und 0.36.
 - Deterministische Prognose:
Manchester United gewinnt!

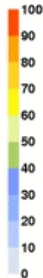
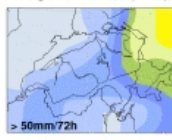


Wetterprognose vom 19. August 2005

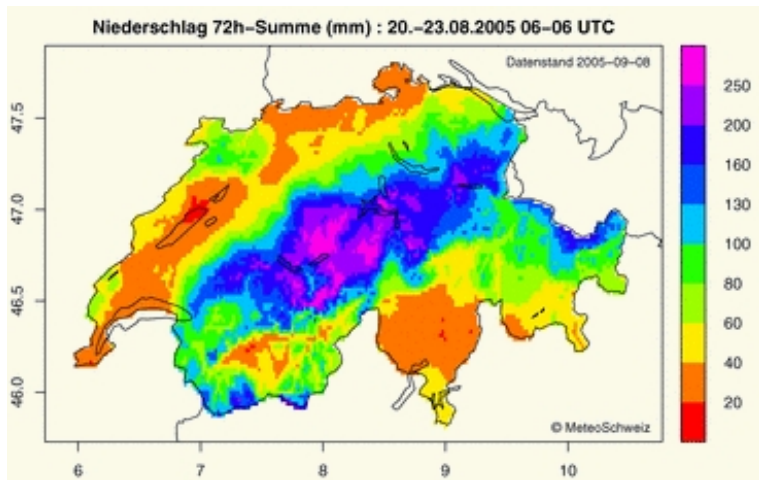
COSMO-LEPS probability forecast: 72h sum of total precipitation
19 Aug 2005 12UTC, t+(18-90), VT: Tuesday 23 Aug 2005 06UTC



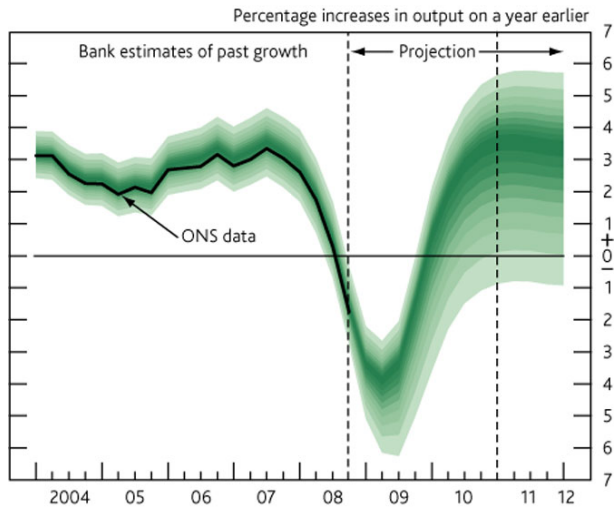
ECMWF EPS probability forecast: 72h sum of total precipitation
19 Aug 2005 12UTC, t+(18-90), VT: Tuesday 23 Aug 2005 06UTC



Niederschlagsmengen 20. bis 23. August 2005



Wirtschaftsprognosen



Gesundheitsprognosen

TIMES**ONLINE**

From The Times

May 7, 2009

The risk of swine flu? I haven't a clue ...

... writes a professor of risk. But I'm still sending my daughter in Mexico some Tamiflu

David Spiegelhalter

It could have been designed to make me feel inadequate. I am a professor of risk, and when my daughter Rosie wanted to spend part of her gap year working on a newspaper, she chose, with a true nose for a story, to go to Mexico.

So it is assumed that I know the chances of her, and everyone else, getting or even dying of, swine flu. But I just don't know; risk is such an odd thing - no instrument can measure it but it constantly changes as we find out more information, just as the odds on Barack Obama being President oscillated wildly in the year before the election. What do we really mean by chance and risk anyway?

[...]

Anyway, my gut feeling is that the chances we will see the girl again are looking quite good. But we've sent out Tamiflu just in case.

David Spiegelhalter is Winton Professor of the Public Understanding of Risk at the University of Cambridge. Rosie Spiegelhalter is sticking it out in Mexico

Fussballprognosen

SF Tagesschau vom 5. März 2008:

- “Tschechien holt laut UBS-Analysten den EM-Titel”
- “Schweiz kommt in den EM-Halbfinal”

Ein genauerer Blick zeigt jedoch:

- “Tschechien gewinnt den Final im Penaltyschiessen gegen Italien mit einer **Wahrscheinlichkeit** von 51%”
- “Die Schweiz hat eine sehr gute **Chance** von 75%, im Viertelfinal gegen Deutschland zu triumphieren”



Die UBS-Prognose (Vorrunde)

Erste Runde und Wahrscheinlichkeit für die zweite Runde in %

Gruppe A		Gruppe B	
Schweiz	64	Kroatien	66
Tschechien	56	Deutschland	57
Portugal	45	Polen	39
Türkei	34	Österreich	38
Gruppe C		Gruppe D	
Italien	54	Spanien	75
Niederlande	53	Griechenland	46
Rumänien	48	Schweden	45
Frankreich	45	Russland	35

Fett = qualifiziert

Fett grün = heissester Aussenseiter,

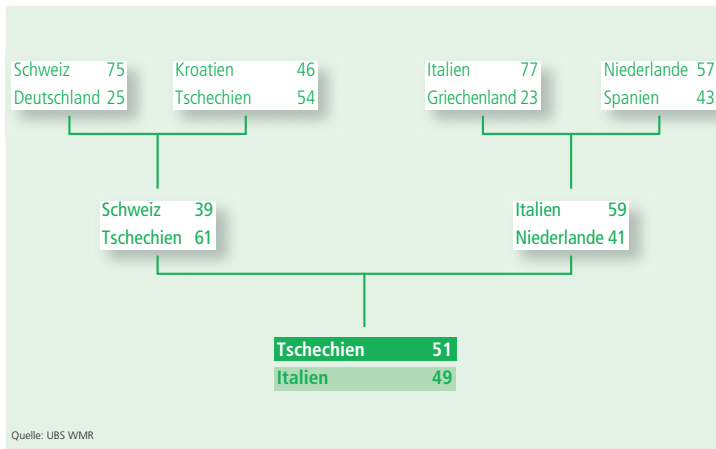
Wahrscheinlichkeit > 45

Quelle: UBS WMR



Die UBS-Prognose (Zweite Runde)

Zweite Runde und Wahrscheinlichkeit zu gewinnen, in %



Forensische Risikokalkulationen

The screenshot shows a web browser window displaying the FOTRES website. The browser's address bar shows the URL <http://www.fotres.ch/>. The website has a navigation menu at the top with links for Home, Contact, Favorites, Print, and Mail this page. A logo for FOTRES (Forensic Operationalized Therapy/Risk Evaluation System) is visible in the top right corner. On the left side, there is a vertical menu with the heading 'PROFECTA' and several sub-items: Language (de, en, nl), What is FOTRES?, FOTRES Login, Manual & Licence, Data Protection, FOTRES Training Courses, Demo Assessment, and Research & Evaluation. The main content area features a large banner image of an open book with a blue overlay containing the text 'FOTRES Forensic Operationalized Therapy/Risk Evaluation System'. At the bottom of the browser window, the status bar shows 'Done' and a copyright notice '© 2008 THREE URY AG'.



Bruno de Finetti, 1906-1985



"Alle und immer befinden wir uns - gegenüber allen oder fast allen Dingen - in Ungewissheit"

- Teoria delle Probabilità (1970)
- Theory of Probability (1974, 75)
- Wahrscheinlichkeitstheorie (1981)



Die zwei Arten von Unsicherheit

- **Aleatorische Unsicherheit** ist durch Zufälligkeit verursacht:
 - Unsicherheit durch Messfehler oder prinzipielle physikalische Eigenschaften (fehlende Vorhersagbarkeit)
 - Beispiel: Münzwurf, Würfelwurf, Lotto etc.
- **Epistemische Unsicherheit** ist durch eine (partielle) Unkenntnis über ein *per se* nicht zufälliges Phänomen gekennzeichnet.
 - Unsicherheit ändert sich, wenn weiteres Wissen berücksichtigt wird.
 - Beispiel: Kartenspiel bei (partieller) Kenntnis der Karten der Mitspieler
- Häufig auch Kombinationen von aleatorischer und epistemischer Unsicherheit



Aleatorische oder epistemische Unsicherheit?

- Welche Zahlen soll ich beim Lotto ankreuzen?
- Hat dieser Patient Krebs?
- Werde ich in den kommenden 10 Jahren einen Herzinfarkt erleiden?
- Wird dieser Straftäter in den kommenden 10 Jahren rückfällig werden?



Wahrscheinlichkeiten: Quantifizierung von Unsicherheit

- Wahrscheinlichkeiten dienen der Quantifizierung von Unsicherheit
- Ein mathematischer Calculus existiert unabhängig von der Interpretation
- Aber was ist eigentlich eine Wahrscheinlichkeit?
 - Frequentistische Interpretation
 - Subjektive Interpretation



Frequentistische Interpretation

- **Frequentistische** Interpretation: Wahrscheinlichkeit als Grenzwert der **relativen Häufigkeit** des Eintretens eines Ereignisses bei hypothetischer Wiederholung des Experiments unter identischen Bedingungen.
- Sinnvoll bei aleatorischen Unsicherheiten:
- Wahrscheinlichkeit einer Sechs im SWISS-Lotto ist 1:8,145,060
 - Wahrscheinlichkeit einer Knabengeburt ist 51.2%



Subjektive Wahrscheinlichkeiten

*For over 200 years men had asked the question:
"What is Probability?"*

De Finetti replied: "Probability does not exist"

Lindley (1986)

- **Subjektive** Interpretation: Wahrscheinlichkeit als Mass für die Sicherheit der persönlichen Einschätzung eines Sachverhaltes. <degree of belief>
- Wahrscheinlichkeit ist Ausdruck unserer unzureichenden Information.
→ Sinnvoll bei aleatorischen und epistemischen Unsicherheiten



Bestimmung von subjektiven Wahrscheinlichkeiten

- **<elicitation>**: Herauslocken, Herausholen
- Ein mögliches Verfahren durch Anbieten einer Wette:
 1. Im Falle des Eintretens eines Ereignisses A wird ein Gewinn von CHF x ausbezahlt.
 2. Eine Person ist bereit, maximal CHF y auf das Eintreten des Ereignisses zu wetten,
 - falls A eintritt, wird der Betrag x ausbezahlt
 - falls A nicht eintritt, wird nichts ausbezahlt
 3. Setzt man den **erwarteten Gewinn** $x \cdot p$ mit dem Einsatz y gleich, so ergibt sich die Wahrscheinlichkeit p für das Eintreten des Ereignisses A zu

$$p = y/x$$

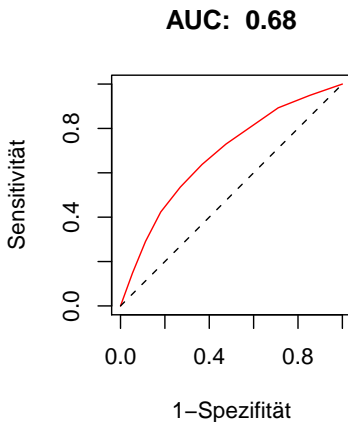
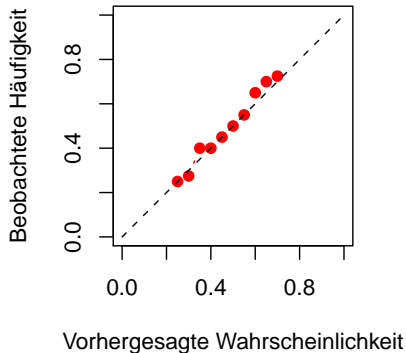


Kalibrierung und Trennschärfe

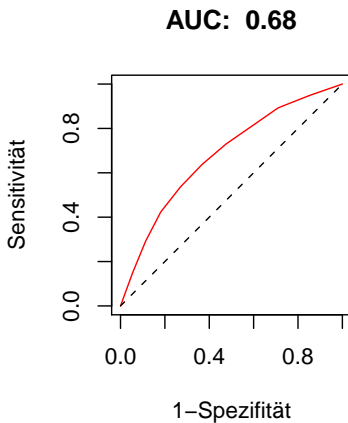
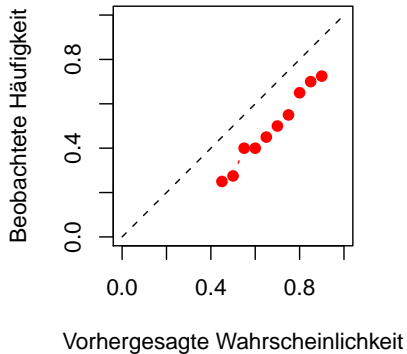
- Zur Bewertung von probabilistischen Prognosen sind verschiedene Aspekte wichtig:
 - **Kalibrierung** beinhaltet, dass von allen Ereignissen mit Prognosewahrscheinlichkeit p im Mittel $p \cdot 100\%$ auch tatsächlich eintreten.
 - **Trennschärfe** <discrimination> ist die Fähigkeit, das Ereignis korrekt zu prognostizieren, beispielsweise durch Festlegung eines Schwellenwertes p_S .
Wird meist durch die AUC <area under the curve> quantifiziert.
- Grafische Bewertung durch **Kalibrierungskurve** und **ROC-Kurve**



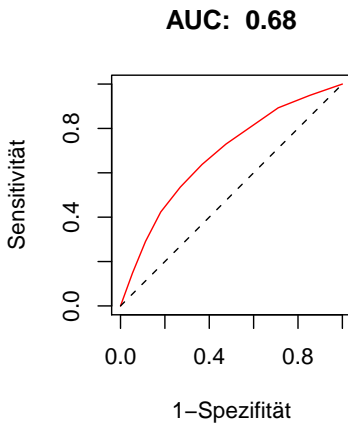
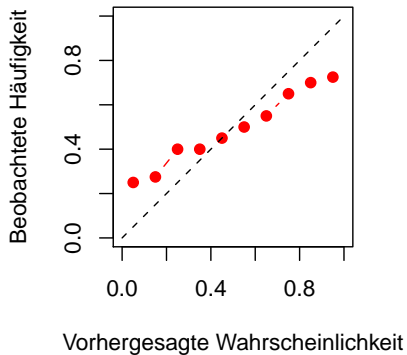
Kalibrierungskurve und ROC-Kurve



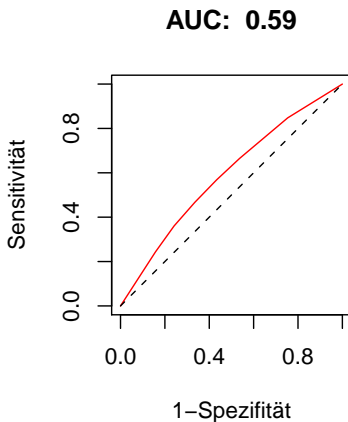
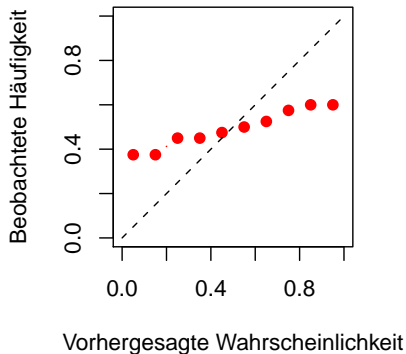
<Overconfidence>



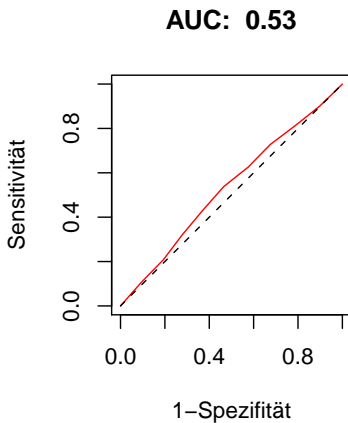
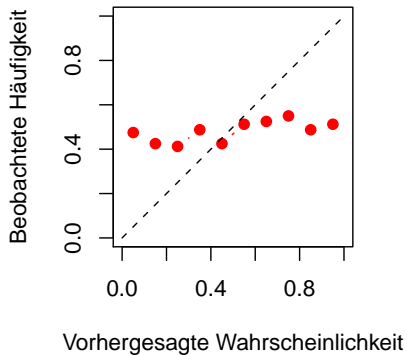
<Over-extremity>



<Over-extremity> und nachlassende Trennschärfe



Fehlende Trennschärfe



Korrekte Bewertungsregeln

- Ziel: Vergleich der Vorhersagen mit den tatsächlich beobachteten Daten mit geeigneten Kriterien
- Leonard Jimmie Savage (1917-71) stellte **korrekte Bewertungsregeln** <proper scoring rules> auf ein neues theoretisches Fundament.



Korrekte Bewertungsregeln

- Der Vergleich einer Prognose p mit dem tatsächlich eingetretenen Ereignis $x \in \{0, 1\}$ wird durch Auszahlung eines Gewinns bewertet.
- Eine **Bewertungsregel** gibt an, wie der Gewinn von p und x abhängt.
- Der **erwartete Gewinn** einer **korrekten Bewertungsregel** ist maximal, wenn das Ereignis $x \in \{0, 1\}$ tatsächlich mit der prognostizierten Wahrscheinlichkeit p eintritt.
- Korrekte Bewertungsregeln berücksichtigen gleichzeitig Kalibrierung und Trennschärfe.



Bewertungsregeln: Beispiele

- Die **lineare Bewertungsregel**

$$S(p, x) = 1 - |p - x|$$

ist nicht korrekt.

- Die **quadratische Bewertungsregel** <Brier score>

$$S(p, x) = 1 - (p - x)^2$$

ist korrekt.

- Die untere Schranke $S = 0.75$ für den <Brier score> erhält man durch die konstante Prognose $p = 0.5$ <fence sitter, hedging>



Beispiel

- Eine bestimmte Vorhersagewahrscheinlichkeit sei $p_1 = 0.7$, eine zweite sei $p_2 = 0.9$, eine dritte $p_3 = 1.0$.
- Falls das Ereignis mit Wahrscheinlichkeit 0.7 eintritt, ist der **erwartete Gewinn** jeweils gleich:

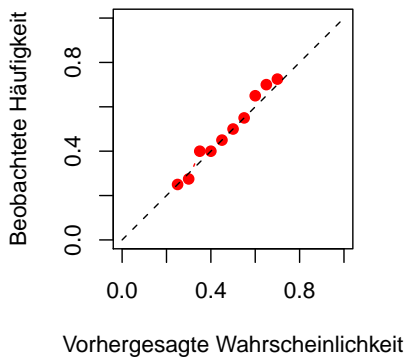
Bewertungsregel	Vorhersage		
	$p_1 = 0.7$	$p_2 = 0.9$	$p_3 = 1.0$
Linear	0.58	0.66	0.70
Quadratisch	0.79	0.75	0.70

- Die lineare Bewertungsregel ermuntert den Vorhersager, unehrlich zu sein!

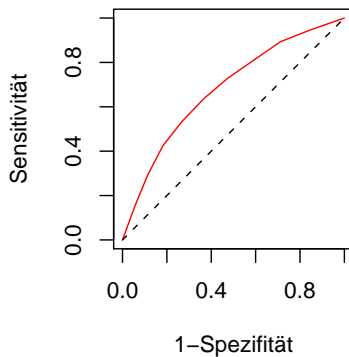


Kalibrierungskurve und ROC-Kurve

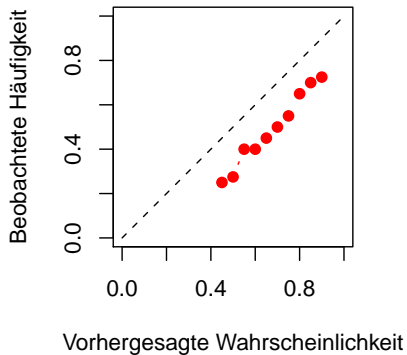
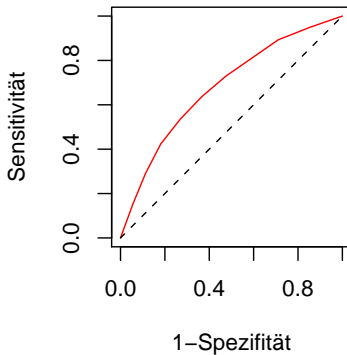
BS: 0.77



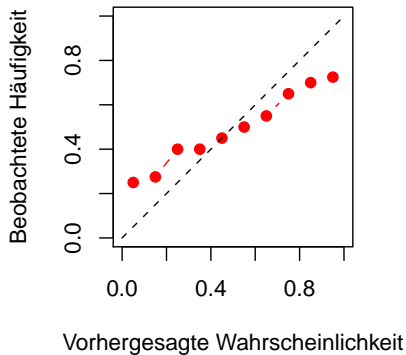
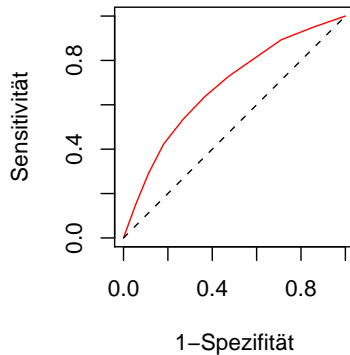
AUC: 0.68



<Overconfidence>

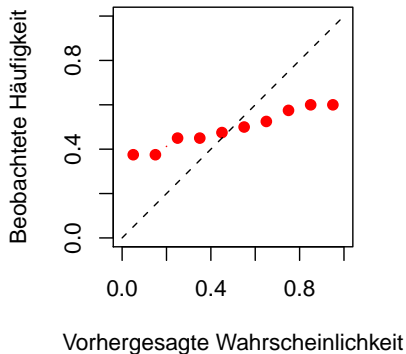
BS: 0.74**AUC: 0.68**

<Over-extremity>

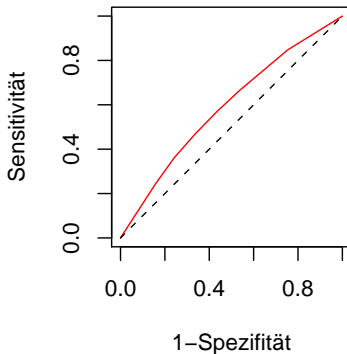
BS: 0.76**AUC: 0.68**

<Over-extremity> und nachlassende Trennschärfe

BS: 0.71

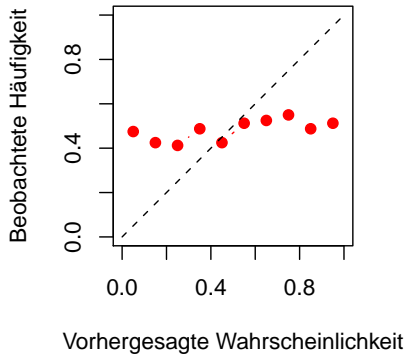


AUC: 0.59

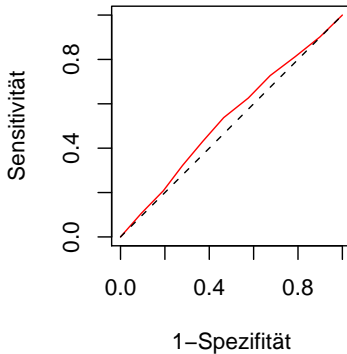


Fehlende Trennschärfe

BS: 0.68



AUC: 0.53



Zusammenfassung

- Probabilistische Prognosen reflektieren unsere Unsicherheit bezüglich zukünftigen Ereignissen.
- Geeignete Kriterien zur Beurteilung der Güte von probabilistischen Prognosen sind Kalibrierung und Trennschärfe.
- Korrekte Bewertungsregeln kombinieren Kalibrierung und Trennschärfe in einer Grösse.

Today, we may be witnessing what future generations might refer to as the transition from point prediction to [...] probabilistic forecasting.

Gneiting (2008)

