

## **Alternative statistische Maßzahlen zur Modellbewertung Definition und erste Ergebnisse**

**Ingo Haag**

HYDRON Ingenieurgesellschaft für  
Umwelt und Wasserwirtschaft mbH

**Norbert Demuth**

Landesamt für Umwelt, Wasserwirtschaft  
und Gewerbeaufsicht Rheinland-Pfalz LUWG

# Inhalt

## (1) Zielsetzung

## (2) Übliche Gütemaße

$r^2$	Bestimmtheitsmaß
NSE	Nash-Sutcliffe-Effizienz
logNSE	NSE logarithmierter Werte
Bilanz	

## (3) Alternatives Gütemaß

VE	Volumen-Effizienz
----	-------------------

## (4) Zusammenfassung

## (5) Empfehlungen und Ausblick

# Zielsetzung

## **Statistische Maßzahlen (Gütemaße):**

Bewertung hydrologischer Simulationen und Vorhersagen

- Objektiv
- Schnell / einfach / klar interpretierbar
- Integral (Fehlerart, Zeitbereich, Abflussbereich)
  
- Werden die Kriterien von üblichen Gütemaßen erfüllt?
- Gibt es ein alternatives, besser geeignetes Gütemaß?

⇒ Erste Analysen / Ergebnisse

⇒ Empfehlungen

# Übliche Gütemaße: $r^2$

## Bestimmtheitsmaß – $r^2$

$$r^2 = \left( \frac{\sum(mes_i - \overline{mes}) \cdot (sim_i - \overline{sim})}{\sqrt{\sum(mes_i - \overline{mes})^2} \cdot \sqrt{\sum(sim_i - \overline{sim})^2}} \right)^2$$

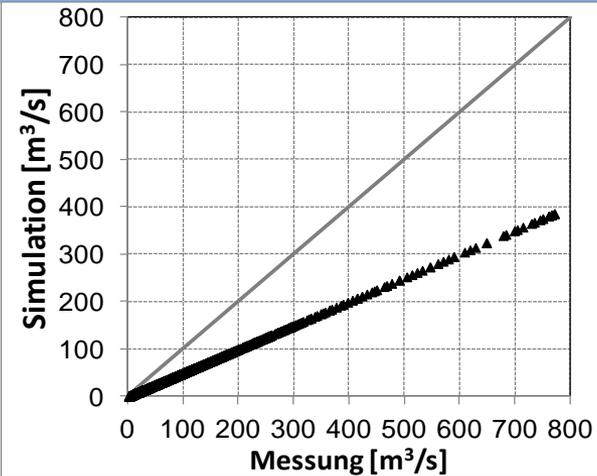
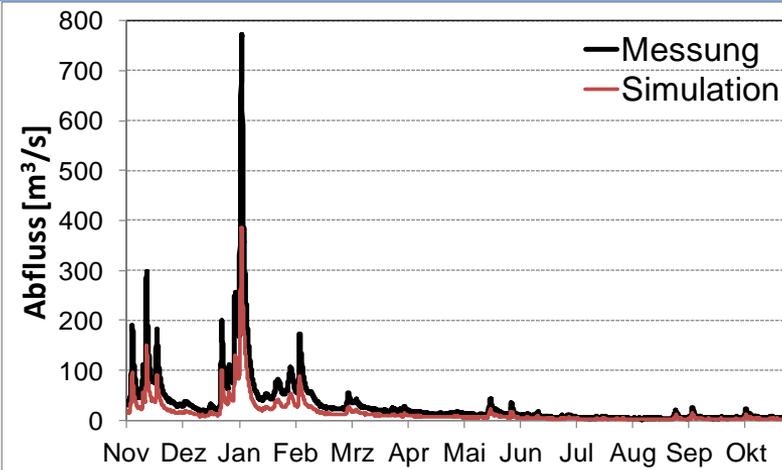
- Kovarianz von Messung und Simulation
- Anteil der durch linearen Zusammenhang erklärten Varianz
- Wertebereich: 0 ... 1  
mit: 1 = perfekte Übereinstimmung

### Hauptprobleme:

- Systematische Fehler nicht berücksichtigt  
(Steigung, Achsenabschnitt der Regression)
- Überbewertung großer Werte

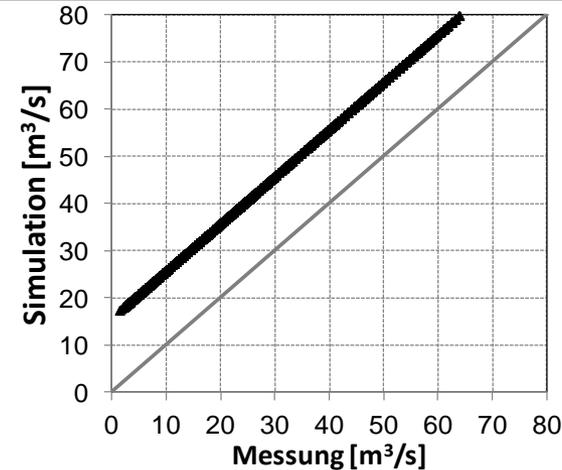
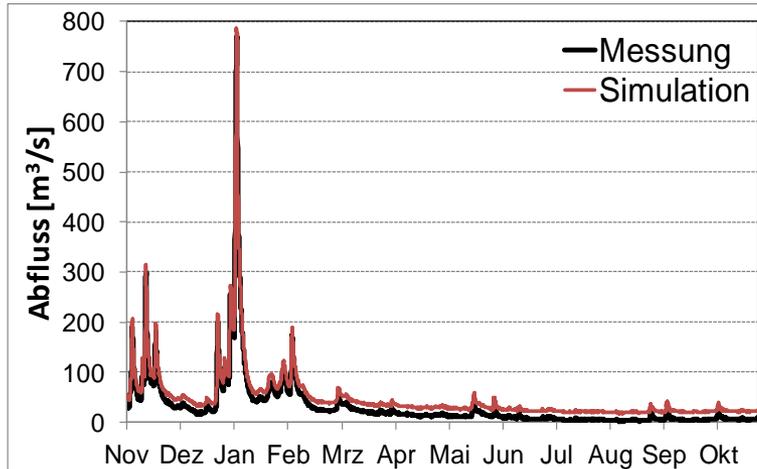
# Übliche Gütemaße: $r^2$

**Simulation =  
0,5 x Messung**



$$r^2 = 1,0$$

**Simulation =  
Messung + 0,5 MQ**



$$r^2 = 1,0$$

⇒  $r^2$  erfasst keine systematischen Fehler

# Übliche Gütemaße: NSE

## Nash-Sutcliffe-Effizienz – NSE

$$NSE = 1 - \frac{\sum (sim_i - mes_i)^2}{\sum (\overline{mes} - mes_i)^2}$$

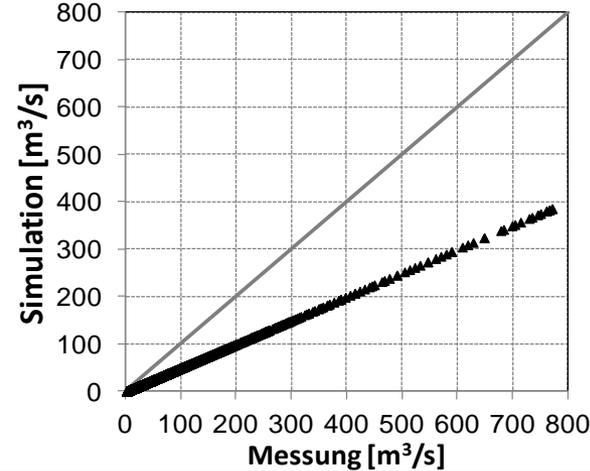
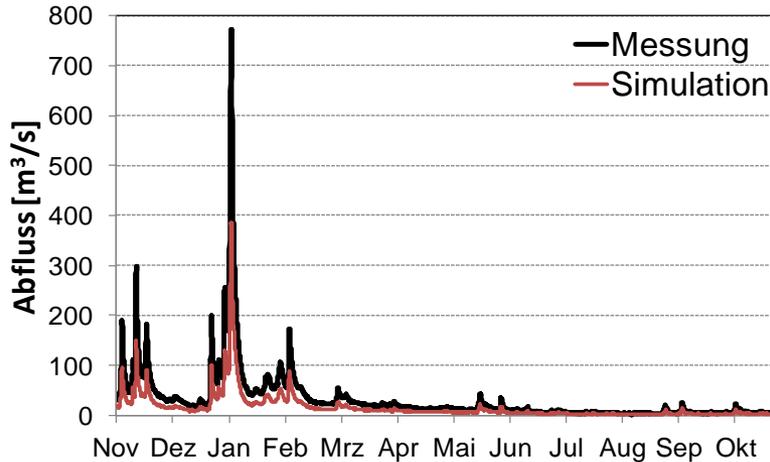
- Verhältnis quadrierte Abweichungen zur Varianz der Messungen
- Mittelwert der Messungen fungiert als Benchmark
- Wertebereich:  $-\infty \dots 1$ 
  - < 0: Schlechter als Mittelwert
  - 1: perfekte Übereinstimmung

### Hauptprobleme:

- Starke Gewichtung großer Werte (ungeeignet für Niedrigwasser)
- Geringe Sensitivität gegenüber systematischen Fehlern (Bilanz)
- Extreme Abhängigkeit von der Varianz (Dynamik, Zeitdauer)

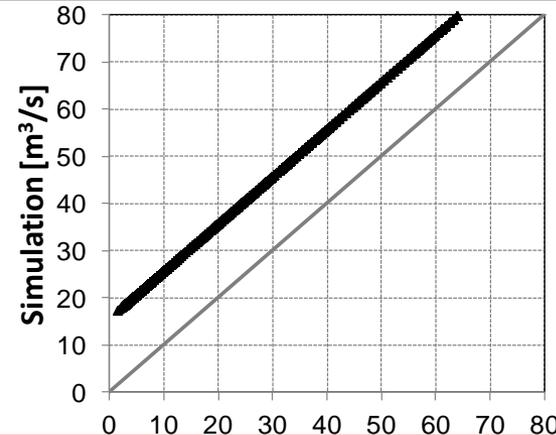
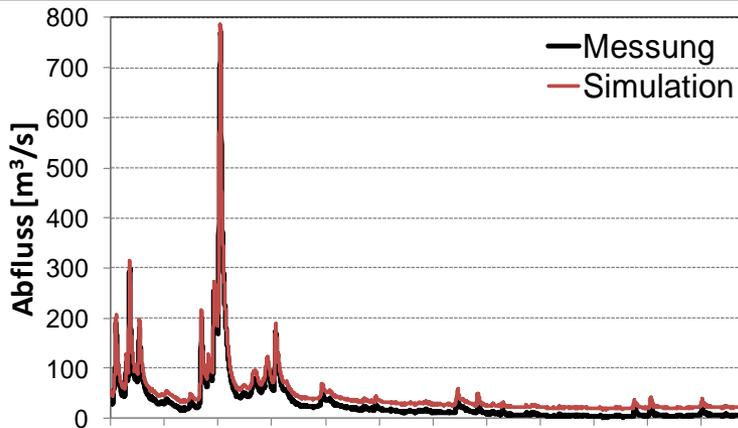
# Übliche Gütemaße: NSE

Simulation =  
0,5 x Messung



NSE = 0,68  
Bilanz = 0,5

Simulation =  
Messung + 0,5 MQ

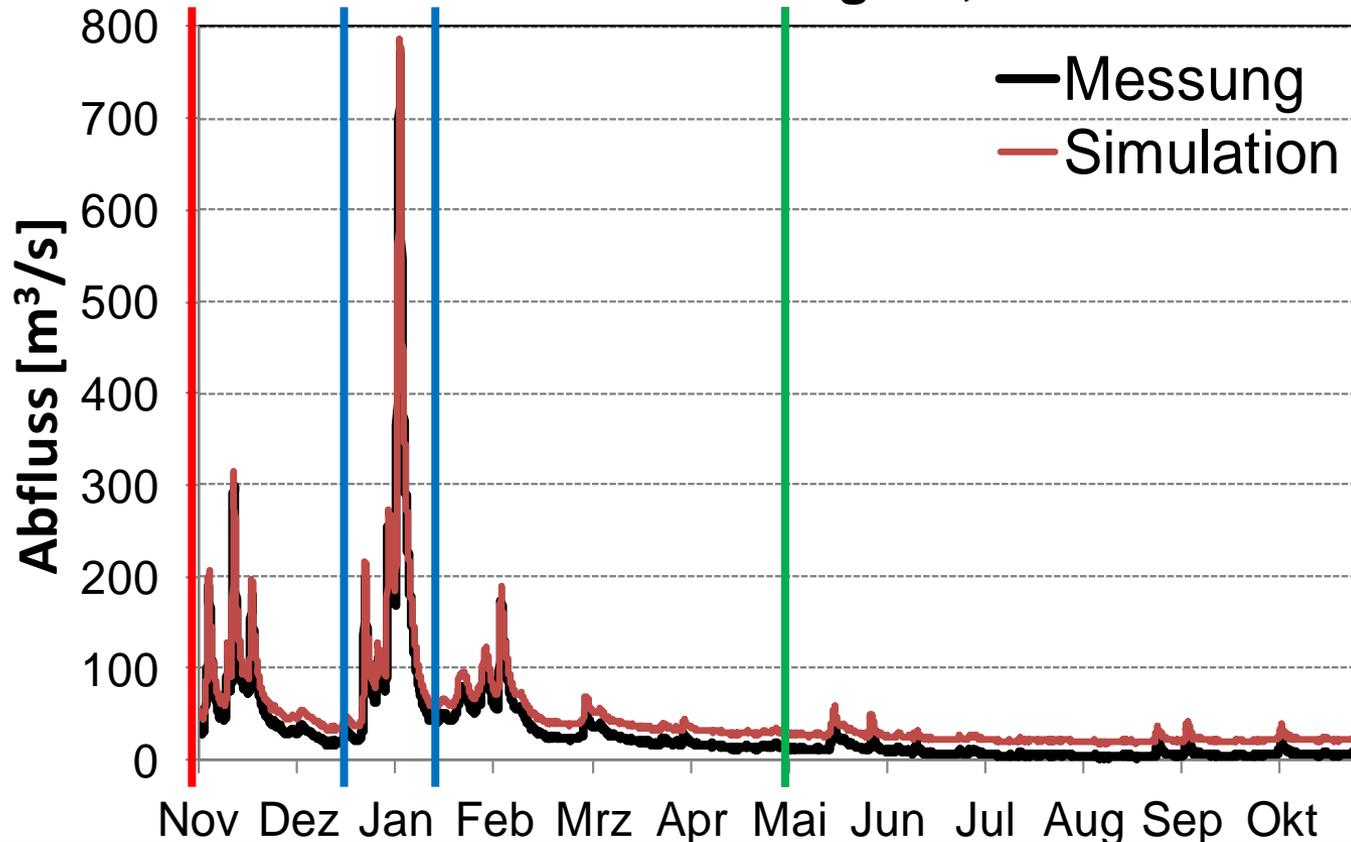


NSE = 0,93  
Bilanz = 1,5

- ⇒ NSE zeigt geringe Sensitivität für systematische Fehler
- ⇒ Kombination mit Bilanz erforderlich

# Übliche Gütemaße: NSE

Simulation = Messung + 0,5 MQ



Gesamtjahr:  
NSE = 0,93

Hyd. Winter:  
NSE = 0,96

Hyd. Sommer:  
NSE = -8,59

HW-Ereignis:  
NSE = 0,99

- ⇒ NSE extrem abhängig von Varianz der Messungen (Dynamik)
- ⇒ Stationäre Zeiten werden generell schlechter bewertet

# Übliche Gütemaße: logNSE

- NSE zu starke Gewichtung hoher Werte
- ⇒ Logarithmische Transformation (logNSE)

$$\logNSE = 1 - \frac{\sum (lg(sim_i) - lg(mes_i))^2}{\sum (lg(\overline{mes}) - lg(mes_i))^2}$$

- ⇒ Stärkere Gewichtung geringer Werte (Niedrigwasser)
- ⇒ Sensitiver gegenüber systematischen Fehlern

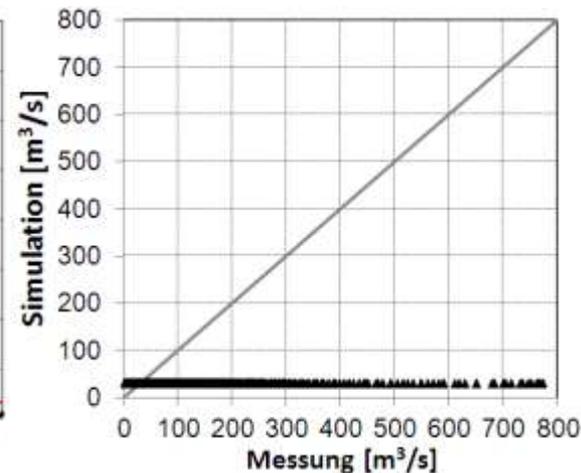
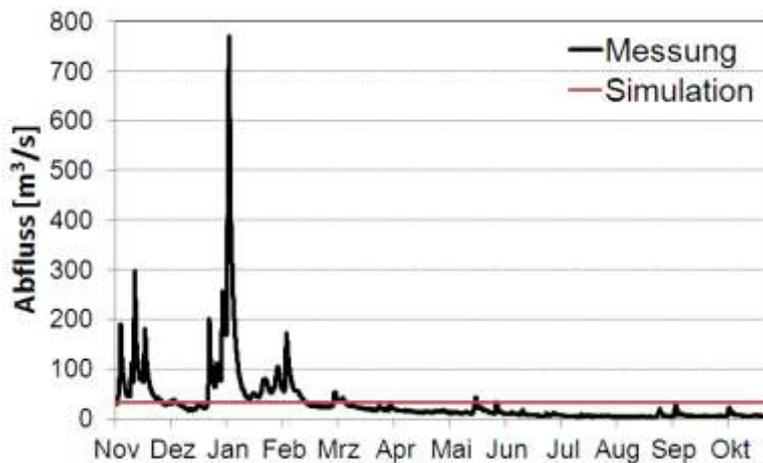
- ⇒ logNSE extrem abhängig von Varianz (Dynamik, Zeitbereich)
- ⇒ Schlechtere Bewertung von stationäre Abflüssen
- ⇒ Schwer interpretierbar

# Übliche Gütemaße: Bilanz

NSE nicht sensitiv gegenüber systematischen Fehlern  
⇒ zusätzlich Betrachtung der Bilanz

$$\text{Bilanz} = \frac{\sum sim_i}{\sum mes_i}$$

Simulation =  
Mittelwert (mes)



Bilanz = 1,0  
NSE = 0,0

- ⇒ Unabhängig von Dynamik
- ⇒ nur als zusätzliches Maß sinnvoll

# Übliche Gütemaße: Zusammenfassung

## Bestimmtheitsmaß $r^2$ :

- Systematische Fehler werden nicht erfasst

## Nash-Sutcliffe-Effizienz NSE:

- Übermäßige Gewichtung hoher Werte
  - ⇒ zusätzlich  $\log NSE$  für Niedrigwasser
- Geringe Sensitivität gegenüber systematischen Fehlern
  - ⇒ zusätzlich Bilanz
- Starke Abhängigkeit von der Varianz (Dynamik, Zeitdauer)

Alternatives Gütemaß?

# Alternatives Gütemaß: VE

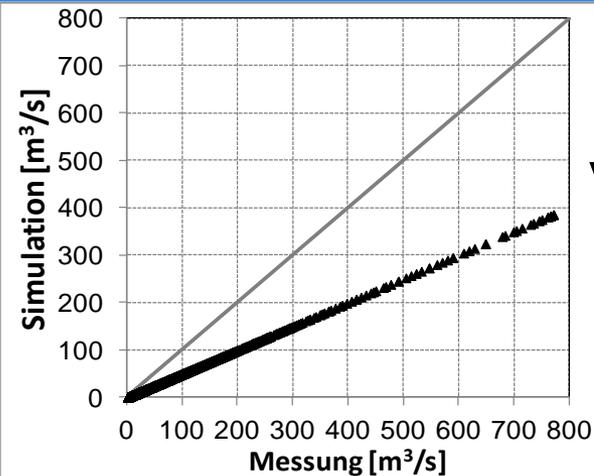
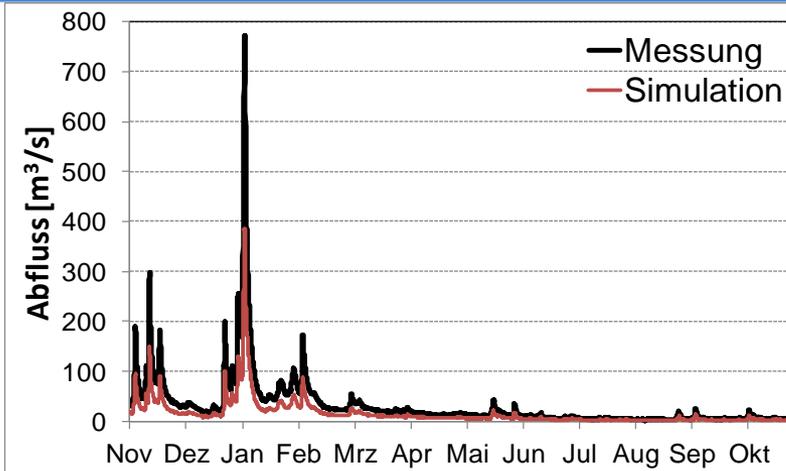
## Volumen-Effizienz VE:

$$VE = 1 - \frac{\sum |Q_{sim} - Q_{gem}|}{\sum Q_{gem}}$$

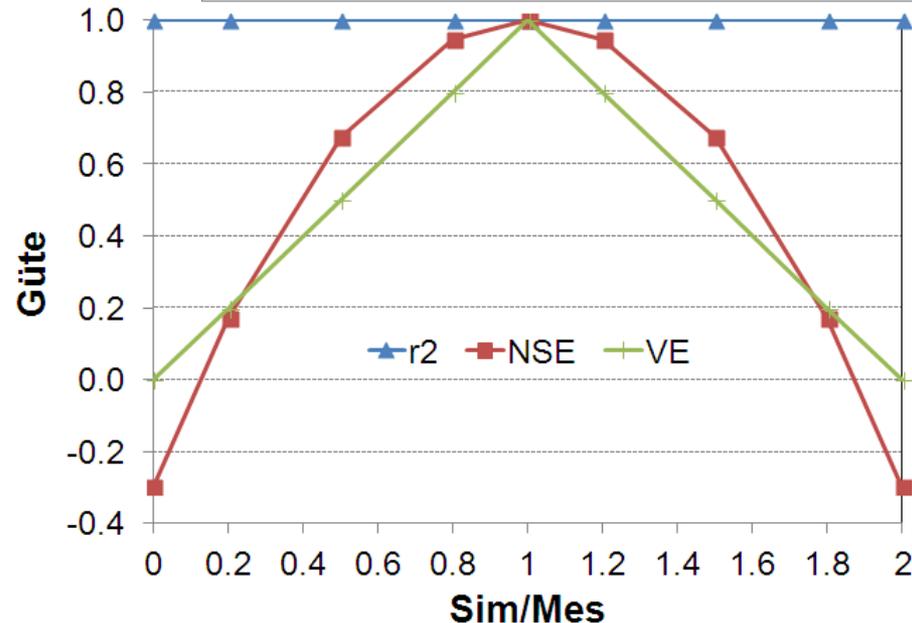
- Verhältnis Abweichungen zu Messungen (relativer Fehler)
- Gleiche Gewichtung jedes Wertepaars
- Anteil des zum richtigen Zeitpunkt simulierten Abflusses (bei korrekter Bilanz)
- Wertebereich:  $-\infty \dots 1$   
1: perfekte Übereinstimmung

# Alternatives Gütemaß: VE

Simulation =  
0,5 x Messung



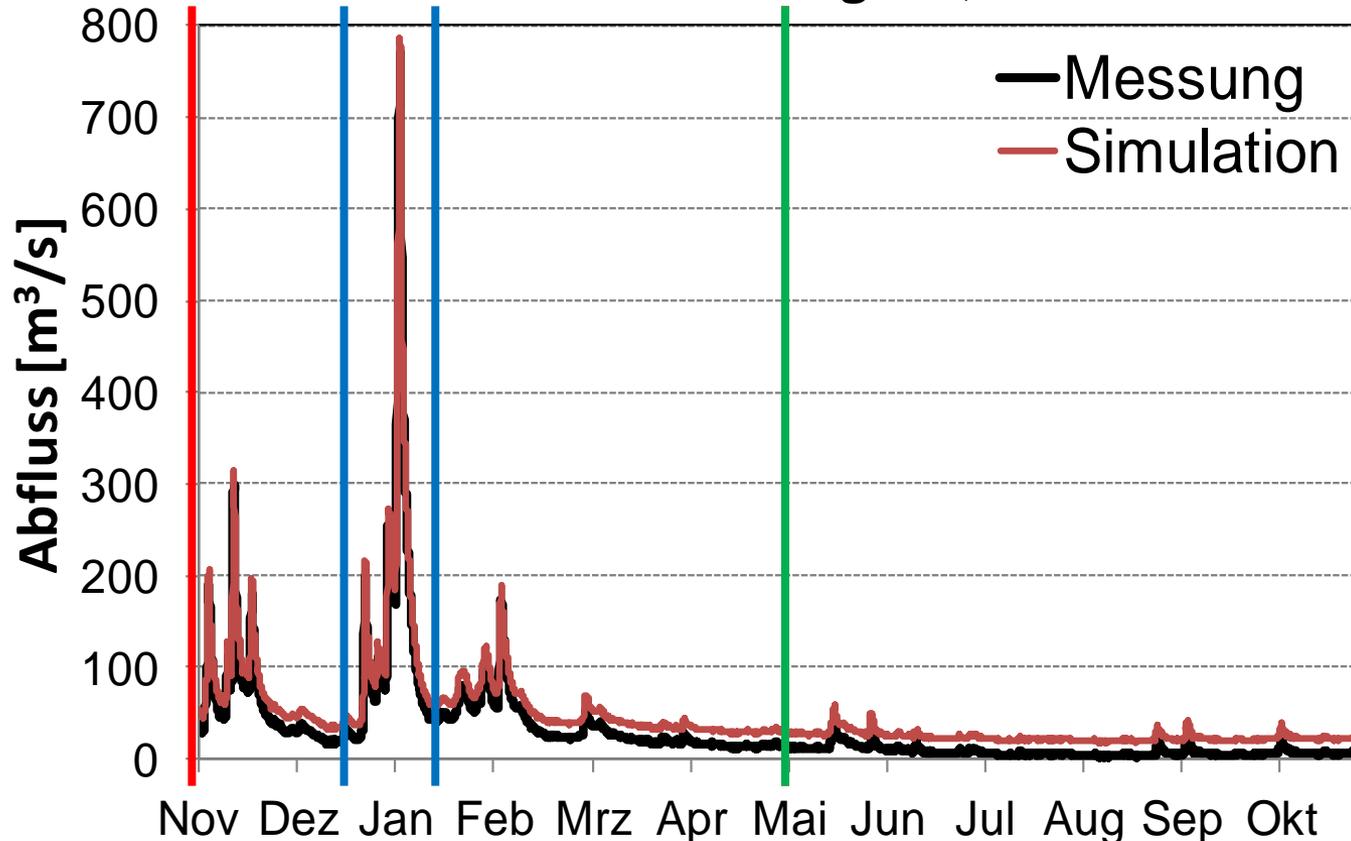
VE = 0,50  
 $r^2 = 1,0$   
NSE = 0,68



- ⇒ VE erfasst systematische Fehler adäquat
- ⇒ Sensitiv für kleine systematische Fehler
- ⇒ Lineare Abhängigkeit von der Bilanz

# Alternatives Gütemaß: VE

Simulation = Messung + 0,5 MQ



Gesamtjahr:

VE = 0,5

NSE = 0,93

Hyd. Winter:

VE = 0,5

NSE = 0,96

Hyd. Sommer:

VE = 0,5

NSE = -8,59

HW-Ereignis:

VE = 0,5

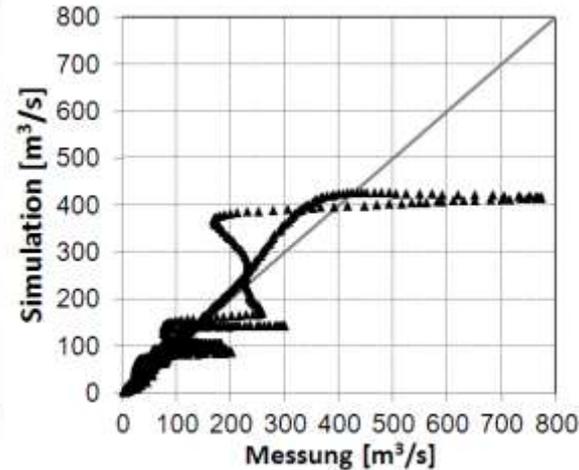
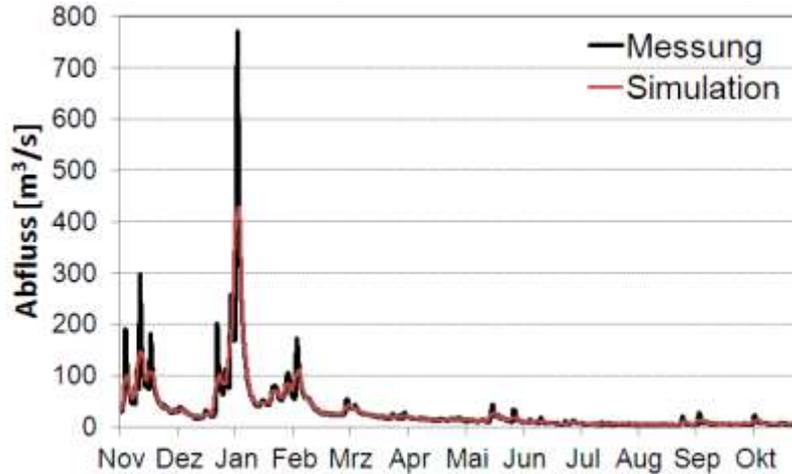
NSE = 0,99

- ⇒ VE ist unabhängig von der Dynamik (Varianz)
- ⇒ Alle Werte werden gleich stark gewichtet

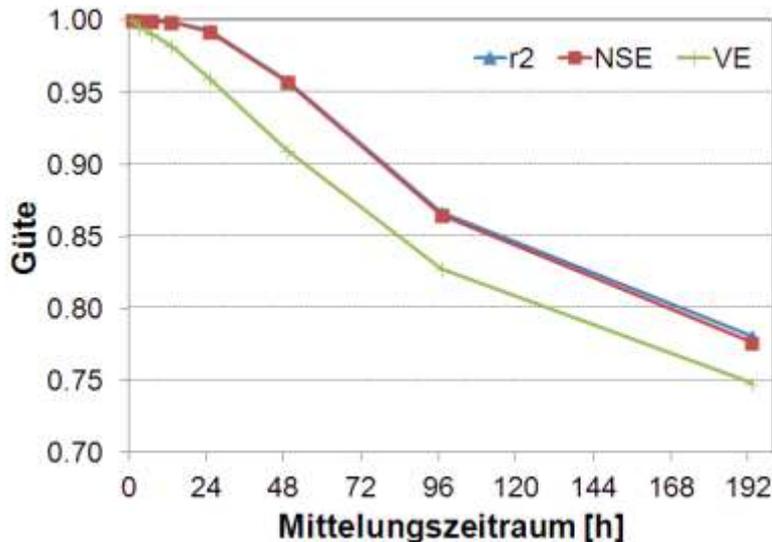
# Alternatives Gütemaß: VE

Dynamik: Gedämpfte Simulation (Bsp. gleitendes Mittel über 4 Tage)

Simulation =  
gleit. Mittel(mes)



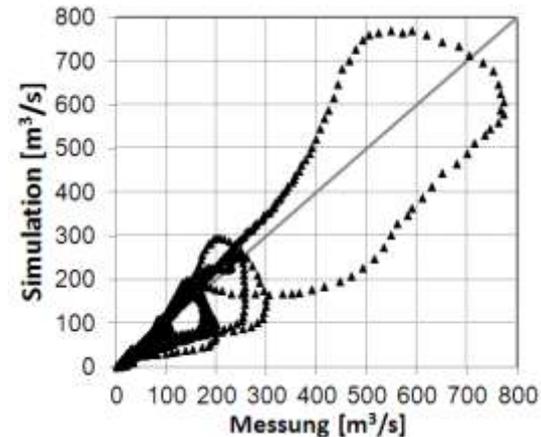
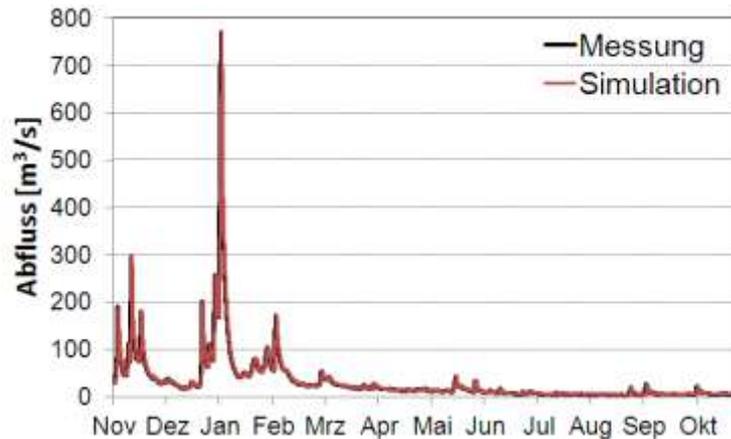
$VE = 0,83$   
 $r^2 = 0,87$   
 $NSE = 0,87$



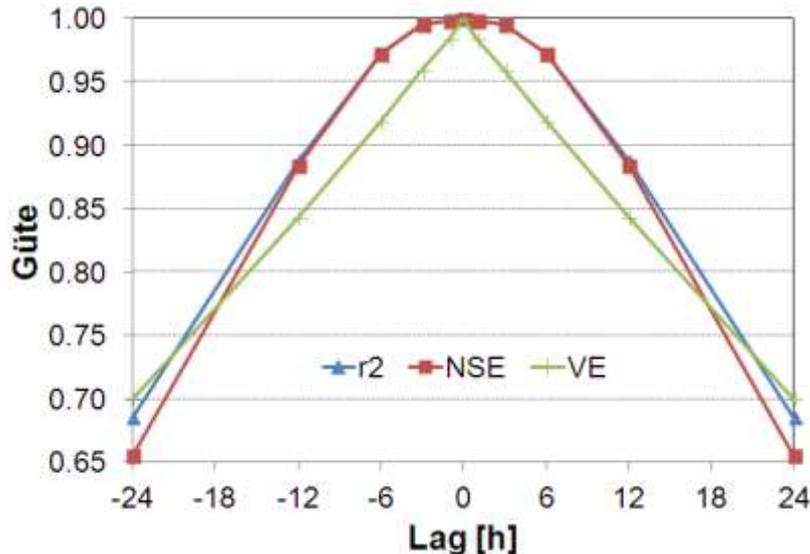
- ⇒ VE sensitiv für geringe Unterschiede in Dynamik
- ⇒ Bei größeren Unterschieden, ähnlicher Verlauf wie NSE,  $r^2$

# Alternatives Gütemaß: VE

Zeitverzug: Lag zwischen Simulation und Messung (Bsp. Lag + 12h)



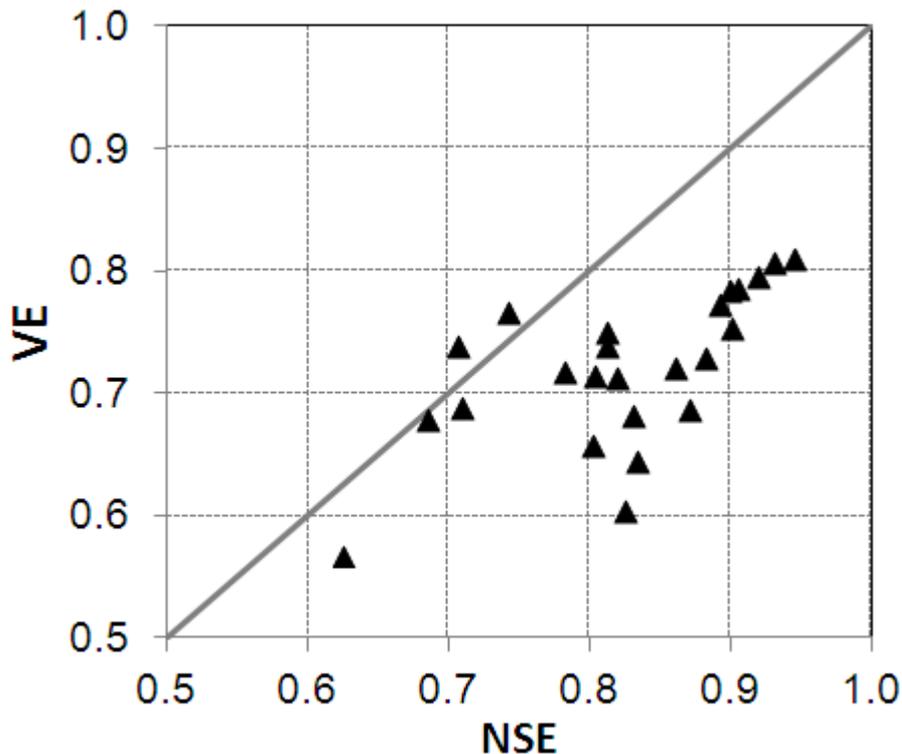
$$\begin{aligned} VE &= 0,84 \\ r^2 &= 0,89 \\ NSE &= 0,88 \end{aligned}$$



- ⇒ VE sensitiv auch bei kleinem Lag
- ⇒ Lineares Verhalten mit zunehmendem Lag

# Alternatives Gütemaß: VE

**Beispiel Nahemodell**  
(24 Pegel)



- Nur grob ähnliche Tendenz
  - VE im Bereich hoher Güte kleiner als NSE
- ⇒ Größere Sensitivität von VE gegenüber kleinen Fehlern

# Zusammenfassung

Nash-Sutcliffe-Effizienz ist nur bedingt geeignet:

- Übergewichtung hoher Werte 
    - Niedrigwasser wird nicht erfasst
  - Geringe Sensitivität gegenüber systematischen Fehlern 
    - Bilanzfehler werden kaum erfasst
  - Extreme Abhängigkeit von der Varianz 
    - Keine Vergleichbarkeit von Pegeln mit unterschiedlicher Dynamik
    - Keine Vergleichbarkeit von Zeitbereichen unterschiedlicher Dynamik
    - Abhängigkeit vom ausgewerteten Zeitbereich
    - Ungeeignet für getrennte Betrachtung von Abflussbereichen
- ✓ zusätzlich  $\log NSE$  für Niedrigwasser
- ✓ zusätzlich Bilanz für systematische Fehler
- ⇒ Abhängigkeit von Varianz bleibt
- ⇒ Nicht integral und einfach

# Zusammenfassung

Volumen-Effizienz ist gut geeignet:

- Gleiche (relative) Gewichtung aller Fehler
    - Integrale, gleich gewichtete Bewertung aller Abflussbereiche
  - Lineare Abhängigkeit von systematischen Fehlern
  - Sensitiv gegenüber anderen üblichen Fehlern
    - Zeitverzug
    - Falsche Dynamik
  - Unabhängig von der Varianz der Messreihe
    - Vergleichbarkeit von Pegeln/Zeiten mit unterschiedlicher Dynamik
    - Unabhängig vom ausgewerteten Zeitbereich
    - Geeignet für getrennte Betrachtung von Abflussbereichen
- ✓ VE erfüllt die Kriterien: **Objektiv, einfach, integral**
- ⇒ **Bislang wenig Erfahrung / Referenz**
  - ⇒ **Keine besondere Sensitivität für Hochwasser**

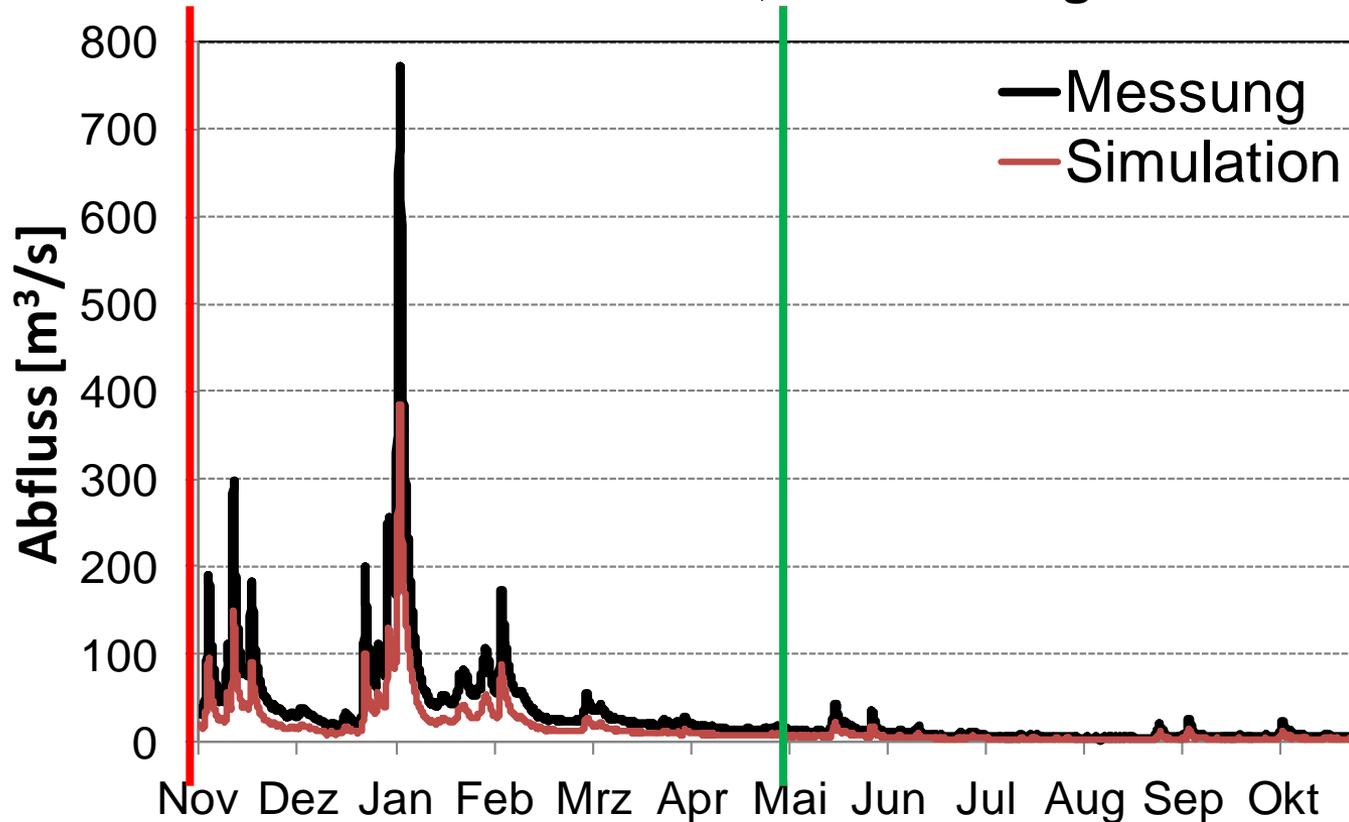
# Ausblick und Empfehlungen

- Weiterhin Verwendung von NSE in Kombination mit logNSE, Bilanz und visuellem Vergleich
- Parallel Ermittlung/Nutzung von VE (Erfahrung und Referenzwerte sammeln)
- Zusätzlich getrennte Betrachtung von Abflussbereichen und ansteigendem Ast von HW (VE scheint hierzu geeignet)
- Systematische Untersuchungen auch im visuellen Vergleich und unter Einbeziehung von Vorhersagetests

**Vielen Dank für Ihre Aufmerksamkeit**

# Übliche Gütemaße: NSE

Simulation = 0,5 x Messung



Gesamtjahr:  
NSE = 0,68

Hyd. Winter:  
NSE = 0,61

Hyd. Sommer:  
NSE = 0,18

- ⇒ NSE extrem abhängig von Varianz der Messungen (Dynamik)
- ⇒ Stationäre Zeiten werden generell schlechter bewertet