

Stephan Hemri, Dmytro Lisniak und Bastian Klein

# Ermittlung probabilistischer Abflussvorhersagen unter Berücksichtigung zensierter Daten

## Ascertainment of probabilistic runoff forecasts considering censored data

Um die meteorologische Vorhersageunsicherheit in der Abflussvorhersage probabilistisch zu quantifizieren, werden in der Vorhersagepraxis häufig Ensembletechniken eingesetzt. Meteorologische Ensemble-Vorhersagen dienen als Input für die Erstellung von probabilistischen Wasserstands- und Abflussvorhersagen. Einerseits stellen hydrologische Modelle stets eine starke Vereinfachung der Realität dar und andererseits werden für die probabilistischen Vorhersagen meteorologische Ensemble-Modelle verwendet, die in der Regel einen Bias und zu geringe Streuung aufweisen. Dies führt unweigerlich zu einem Bias und zu mangelnder Streuung im hydrologischen Ensemble. Somit muss statistisches Post-Processing verwendet werden, um gut kalibrierte und gleichzeitig scharfe Vorhersagen zu erhalten. Solche Methoden, wie zum Beispiel Bayesian Model Averaging (BMA) oder die im Folgenden verwendete Ensemble Model Output Statistics (EMOS) Methode, haben ihre Ursprünge meist in meteorologischen Anwendungen, werden aber zunehmend auch in der Hydrologie eingesetzt. In einigen hydrologischen Vorhersagesystemen sind die Daten sowohl unter- als auch oberhalb gewisser Schwellenwerte mit großen Unsicherheiten behaftet. Deshalb werden in dem Vorhersagesystem für die zwei Teileinzugsgebiete des Rheins, die in dieser Studie betrachtet werden, Werte unterhalb eines gewissen Schwellenwertes zensiert. Aufbauend auf EMOS wird eine zensierte EMOS Methode vorgestellt, die mit zensierten Daten umgehen kann. Daraufhin wird die zensierte EMOS Methode mit Hilfe unterschiedlicher statistischer Gütemaße über den kompletten Vorhersagehorizont von 1 bis 114 Stunden verifiziert.

**Schlagwörter:** Abflussvorhersage, EMOS Methode, Ensemble-Modelle, probabilistische Vorhersage, Unsicherheitsanalyse, zensierte Daten

When forecasting water levels and river flow, ensemble weather forecasts are frequently used as meteorological input to hydrologic process models. As hydrologic models are imperfect and the input ensembles tend to be biased and underdispersed, the output ensemble forecasts for river runoff typically are biased and underdispersed, too. Thus, statistical post-processing is required in order to achieve sharp and calibrated predictions. Standard post-processing methods like Bayesian Model Averaging (BMA) or Ensemble Model Output Statistics (EMOS), which will be discussed extensively in the following, have their origins in the field of meteorological forecasting. In recent times, these methods are increasingly used in the field of hydrology. In hydrologic forecasting systems, data below or above certain threshold values are subject to increased uncertainty. In this study, two sub-catchments of the River Rhine are considered. The forecasting systems associated with these two catchments use runoff data that are censored below a predefined threshold value. On the basis of EMOS, a censored EMOS method is developed that is able to cope with censored data. Censored EMOS will then be verified over the entire forecast time horizon of 1 to 114 hours using several different statistical measures.

**Keywords:** Censored data, EMOS method, ensemble models, probabilistic forecasts, runoff forecasts, uncertainty analysis

## 1 Einleitung

Im Rahmen des übergeordneten "Seamless Prediction" Forschungsvorhabens der Bundesanstalt für Gewässerkunde (BfG) werden am Institut für Angewandte Mathematik der Universität Heidelberg Verfahren zur probabilistischen Bewertung von Abfluss- und Wasserstandsvorhersagen entwickelt. Diese "Probabilistic forecasting" Arbeiten (PFA) basieren auf der Grundlage der Abfluss- und Wasserstandssimulationen des operationellen Vorhersagesystems der BfG für den Rhein (MEIßNER & RADEMACHER 2010). Eine der Hauptquellen der Unsicherheit von Abfluss- und Wasserstandsvorhersagen ist die meteorologische Unsicherheit über die künftige Entwicklung des Wetters. Um diese meteorologische Unsicherheit in der Vorhersage zu berücksichtigen, werden meteorologische Ensemble-Vorhersagen als Eingangsdaten für das hydrologische Vorhersagemodell verwendet und daraus Abfluss-Ensemble-Vorhersagen berechnet. Mittels statistischem Post-Processing werden diese Abfluss-Ensemble-Vorhersagen bewertet und vor allem auch im Sinne von probabilistischen Vorhersagen verbessert. Demzufolge werden möglichst scharfe Vorhersagen unter der Bedingung guter Kalibrierung angestrebt (GNEITING & RAFTERY 2007). Bei gut kalibrierten Vorhersagen stimmen die theoretischen Wahrscheinlichkeiten von Vorhersageintervallen mit der empirischen Frequenz, mit der die Beobachtungen in das Intervall fallen, überein. Die Schärfe bezieht sich nur auf die Vorhersagen und bezeichnet wie „eng“ ein be-

stimmtes Vorhersageintervall ist, z.B. die mittlere Breite von 90 % Vorhersageintervallen. Im Rahmen der PFA sollen probabilistische Vorhersagen der Teileinzugsgebiete des deutschen Rheinabschnitts betrachtet werden.

Eine Vielzahl von Arbeiten haben sich schon mit statistischem Post-Processing von hydrologischen Ensemble-Vorhersagen beschäftigt. So wurden zum Beispiel Ensemble-Vorhersagen für den Rhein an der deutsch-niederländischen Grenze mittels Bayesian Revision als Post-Processing Methode erfolgreich nachbearbeitet (REGGIANI et al. 2009). MADADGAR et al. (2012) wendeten Copula Techniken an, um Ensemble-Vorhersagen mittels einer bivariaten Verteilung von Vorhersagen und Beobachtungen statistisch nachzubearbeiten. Zur probabilistischen Kombination von mehreren (Ensemble-)Niederschlags-Abfluss-Vorhersagen wurde in letzter Zeit vielfach das von RAFTERY et al. (2005) entwickelte Bayesian Model Averaging (BMA) angewandt. So zeigten AJAMI et al. (2007) oder DUAN et al. (2007), dass die Kombination von hydrologischen Vorhersagen mittels BMA nebst quantitativen Angaben zur Vorhersageunsicherheit auch zu Verbesserungen in Bezug auf deterministische Gütemaße führt. FRALEY et al. (2010) führten eine angepasste BMA Version ein, die mit austauschbaren Ensemblemitgliedern umgehen kann, wie sie in meteorologischen Ensemble-Vorhersagen typischerweise vorkommen. Jüngste Entwicklungen erlauben einerseits die Verwendung flexibler Vorhersagedichteverteilungen (PARRISH et al. 2012, RINGS

et al. 2012) und andererseits die simultane statistische Nachbearbeitung über einen längeren Vorhersagezeitraum (BOGNER et al. 2013, HEMRI et al. 2013). Weitere Alternativen zur statistischen Nachbearbeitung sind der Model Conditional Processor (COCCIA & TODINI 2011, TODINI 2008) sowie der Hydrologic Uncertainty Processor (KRZYSZTOFOWICZ & KELLY 2000). Diese Liste ist nicht abschließend, sondern soll einen Eindruck über die unterschiedlichen Post-Processing Methoden vermitteln.

Eine Besonderheit im BfG Vorhersagesystem für den Rhein stellt das Vorkommen links-zensierter (Modell-)Daten dar. Links-zensiert bezeichnet einen Datensatz, in welchem Werte unterhalb eines bestimmten Schwellenwertes durch den Schwellenwert selbst ersetzt werden. Die Zensierung bzw. Maskierung von Daten bildet in hydrologischen Echtzeitanwendungen ein probates Mittel, um äußerst unsichere oder nicht definierte Daten zu behandeln. Dies kann sowohl den niedrigen Abflussbereich (z.B. bei unsicheren Wasserstands-Abfluss-Beziehungen und im Fall staueregelter Gewässer, wenn Wasserstand und Abfluss entkoppelt sind) als auch Hochwasser (z.B. wenn der Bereich der definierten Wasserstand-Abfluss-Beziehung überschritten wird) betreffen. Das Ziel der vorliegenden Arbeit ist das Post-Processing zensierter Abflüsse. Im Folgenden werden exemplarisch die relativ kleinen Gebiete der Ahr (Pegel Altenahr) und der Wied (Pegel Friedrichsthal) behandelt, da diese einen hohen Anteil zensierter Daten aufweisen. Um obiges Ziel zu erreichen, soll eine nicht zu komplexe, einfach ausbaubare Post-Processing Methode so angepasst werden, dass zensierte Daten vom Modell korrekt verwendet werden können. Die Ensemble Model Output Statistics (EMOS) Methode (GNEITING et al. 2005, THORARINSDOTTIR & GNEITING 2010), die zusammengefasst aus einer normalverteilten Vorhersagedichte mit Mittelwert und Varianz abhängig von den Ensemble-Vorhersagen besteht, ist hierfür sehr geeignet. Deshalb wird im Folgenden das zensierte EMOS entwickelt, das EMOS mit dem Konzept von zensierten Verteilungen kombiniert. Alle hier vorgestellten Analysen wurden mit der statistischen Software R durchgeführt (R CORE TEAM 2012).

In Abschnitt 2 werden die in dieser Studie verwendeten Daten vorgestellt. Zensiertes EMOS wird in Abschnitt 3 sowohl konzeptionell als auch mathematisch detailliert eingeführt. Auf die Ergebnisse in Abschnitt 4 folgen eine Diskussion in Abschnitt 5 sowie eine kurze Zusammenfassung und einige Schlussfolgerungen in Abschnitt 6.

## 2 Daten

Für die Analyse stehen Daten der Pegel Friedrichsthal (Wied) sowie Altenahr (Ahr) mit einer stündlichen Auflösung zur Verfügung. Die Wied verfügt über ein Einzugsgebiet von 680 km<sup>2</sup>, das der Ahr beträgt 746 km<sup>2</sup>. Deren Lage im Rheingebiet sowie die Topographie sind in Abbildung 1 dargestellt. Diese Auswahl wurde getroffen, da für beide Gebiete bedeutende Anteile der Beobachtungen unterhalb eines global gesetzten Schwellenwertes von 5 m<sup>3</sup>/s zur weiteren Verarbeitung (primär der statistischen Vorhersagekorrektur) im operationellen Vorhersagesystem zensiert sind. Im Falle der Wied sind 55 % der Beobachtungen im betrachteten Zeitraum vom 1.11.2008 bis zum 31.10.2011 zensiert, im Falle der Ahr sogar 72 %. Die statistisch korrigierten Abflussvorhersagen des hydrologischen Modells sind gleichermaßen links-zensiert. Dieser Zensierung muss im Post-Processing Rechnung getragen werden.

Für beide betrachteten Einzugsgebiete stehen hydrologische Vorhersagen des Modells HBV-96 (BERGSTRÖM 1995, LINDSTRÖM et al. 1997) zur Verfügung, die mit unterschiedlichen meteorologischen Vorhersagemodellen angetrieben werden. Wie aus Tabelle 1 zu entnehmen ist, unterscheiden sich die Modelle sowohl in der Anzahl der Mitglieder als auch in den maximalen Vorhersagezeiten. Die Vorhersagen von Temperatur und Niederschlag werden auf die Einzugsgebiete interpoliert und mit diesen Eingangsdaten mittels des HBV-Modells die Abflussvorhersagen mit einer Rechenzeitschrittweite von 1 h berechnet. Hierbei ist zu beachten, dass die betrachteten Einzugsgebiete anhand von Landnutzung und Geländehöhenklassen weiter in Hydrotope ein-

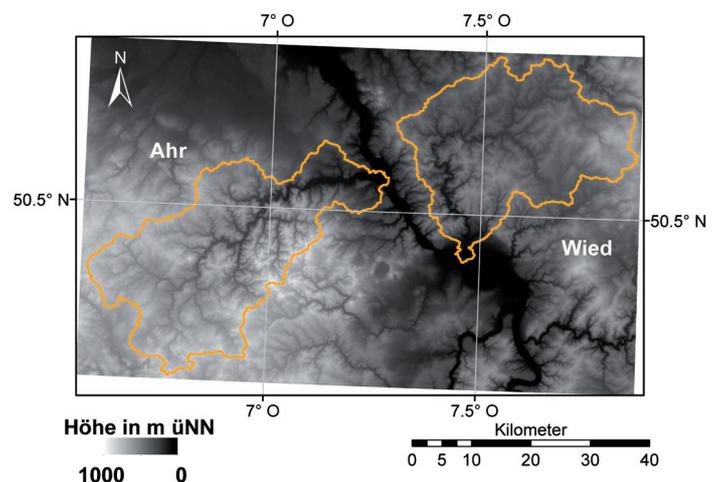
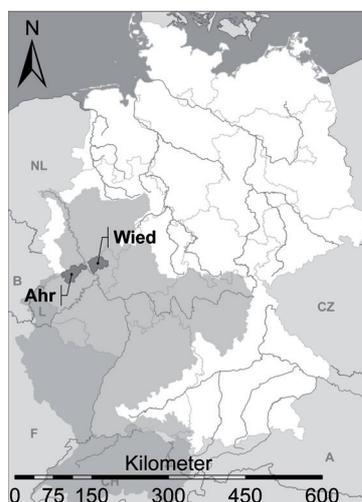
**Tabelle 1**  
Meteorologische deterministische und Ensemble-Inputmodelle<sup>1</sup>  
*Meteorological deterministic and ensemble forcing models*

Name	Anzahl Mitglieder	Vorhersagezeit	Auflösung ~
COSMO-LEPS	16	132 h	10 km
DWD-GME	1 (deterministisch)	174 h	20 km
DWD-MER	1 (deterministisch)	78 h (174 h)	7 km (20 km)

<sup>1</sup> DWD-MER bedeutet, dass das hydrologische Modell bis zu Vorhersagezeit 78 h mit COSMO-EU angetrieben wird und danach von DWD-GME. DWD-MER stands for a model run based on COSMO-EU forcing up to lead-time 78 h and on DWD-GME thereafter.

**Abbildung 1**

Lage der Einzugsgebiete von Wied und Ahr im Rheingebiet sowie flächentreue Projektion der digitalen Höhenmodelle beider Gebiete  
*Location of the catchments of the rivers Wied and Ahr in the Rhine basin and equal-area projection of both areas' digital elevation models*



geteilt sind, auf denen die relevanten hydrologischen Prozesse der Abflussbildung berechnet werden (MEIBNER & RADEMACHER 2010).

Die verwendeten meteorologischen Vorhersagen bestehen aus einem 18 Mitglieder Multi-Modell Ensemble. Dieses setzt sich aus den 16 COSMO-LEPS (MONTANI et al. 2010) Mitgliedern sowie aus den beiden deterministischen Modellen DWD-GME (MAJEWSKI et al. 2002, MAJEWSKI et al. 2012) und DWD-MER zusammen. DWD-MER verwendet für die ersten 78 Stunden die meteorologischen Vorhersagen des COSMO-EU Modells (SCHULZ & SCHÄTTLER 2011, STEPELER et al. 2002) und ab einem Vorhersagehorizont von 79 h die Meteodaten des DWD-GME Modells. Obwohl ab Stunde 79 die Vorhersagen des DWD-GME Modells in zwei Ensemblemitgliedern (DWD-GME, DWD-MER) verwendet werden, sind aufgrund der unterschiedlichen Systemzustände (z.B. Bodenfeuchte) des hydrologischen Modells auch für längere Vorhersagehorizonte unterschiedliche Vorhersagen möglich.

Als Beobachtungsdaten des Abflusses dienen die im operationellen Betrieb über Wasserstands-Abfluss-Beziehungen in Abfluss umgerechneten, gemessenen Wasserstände für den Zeitraum vom 1.11.1998 bis zum 10.1.2013. Die Beobachtungsdaten werden analog zu den Vorhersagedaten für Abflüsse kleiner 5 m³/s zensiert (d.h. auf 5 m³/s gesetzt). Diese Daten dienen der Modellanpassung sowie der Verifikation. Zur Verifizierung von probabilistischen Vorhersagen wird häufig die Klimatologie, also das beobachtete empirische Wissen aus der Vergangenheit, als Vergleichs-Referenz verwendet. Die hier verwendete klimatologische Vorhersage entspricht der empirischen Wahrscheinlichkeitsdichte der Beobachtungen vom 1.11.1998 bis zum 31.10.2008. Die von den atmosphärischen Modellen angetriebenen Abflussvorhersagen werden täglich vom 1.11.2008 bis zum 25.10.2011 um 06:00 GMT+1h Uhr initialisiert und liefern stündliche Vorhersagen bis zu 114 h (COSMO-LEPS) sowie 174 h (DWD-GME und DWD-MER). In dieser Studie werden nur Vorhersagezeiten bis 114 h betrachtet, da das Wegfallen von Ensemblemitgliedern eine detailliertere Betrachtung erfordert (HEMRI et al. 2013). Um die gesamte Zeitspanne zur Verifikation verwenden zu können, werden Paare von saisonalen Trainings- und Verifikationsperioden gemäß Tabelle 2 gebildet. Dieses Auswahlverfahren hat in Tests zu deutlich besseren Verifikationsergebnissen geführt als das in der Meteorologie häufig verwendete fortlaufende Zeitfenster von x Tagen im Vorfeld des Verifikationstages als Trainingsperiode (s. RAFTERY et al. 2005).

**Tabelle 2**  
Beispielpaare von Verifikations- und zugehörigen Trainingsperioden<sup>1</sup>  
*Examples of pairs of verification periods and training periods*

Verifikationsperiode	Trainingsperiode
November 2008	SON 2009, SON, 2010, SO 2011
DJF 2008/2009	DJF 2009/2010, DJF 2010/2011
MAM 2009	MAM 2010, MAM 2011
.	.
.	.
.	.
SO 2011	November 2008, SON 2009, SON 2010

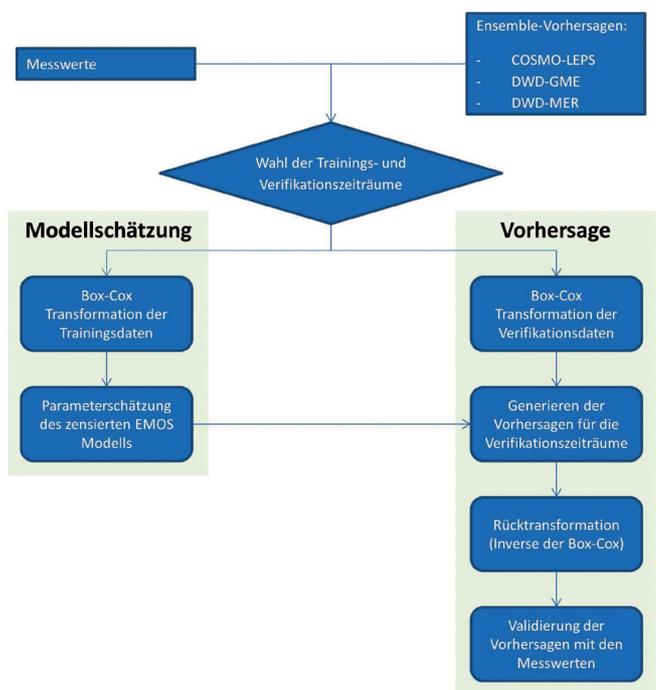
<sup>1</sup> SON steht für September, Oktober, November; SO für September, Oktober; DJF für Dezember, Januar, Februar; MAM für März, April, Mai. SON denotes September, October, November; SO means September, October; DJF stands for December, January, February; MAM for March, April, May.

### 3 Methoden

#### 3.1 Einführung zensiertes EMOS

Ensemble Model Output Statistics – EMOS ist eine Standard Post-Processing Methode für Multi-Modell Vorhersagen (GNEITING et al. 2005). Bevor nun auf die Details des verwendeten EMOS Modells eingegangen wird, soll zuerst EMOS im Vorhersageprozess eingeordnet werden. Wie Abbildung 2 zeigt, müssen in einem ersten Schritt Trainings- und zugehörige Verifikationsperioden ausgewählt werden. Die Trainingsperioden dienen der Schätzung der Modellparameter, die Verifikationsperioden der Abschätzung der Modellgüte. In einem etwaigen operationellen Betrieb entsprechen die Verifikationsperioden, in der das im Trainingszeitraum geschätzte Modell unverändert angewendet wird, den Echtzeit-Vorhersagen, wobei dann natürlich die dazugehörigen Beobachtungen zum jeweiligen Zeitpunkt der Vorhersage und somit dem Einsatz des EMOS-Modells noch unbekannt wären.

Im Modellschätzungsprozess werden zuerst sowohl die Trainingsvorhersagen als auch die dazugehörigen Beobachtungen so transformiert, dass die statistischen Modellannahmen erfüllt werden können. In der vorliegenden Arbeit wird eine Box-Cox Transformation (BOX & COX 1964) verwendet, um annähernd normalverteilte Daten zu erhalten. Daraufhin werden die Modellparameter geschätzt. Nach analoger Transformation der Verifikationsvorhersagen werden mit Hilfe der aus den Trainingsdaten geschätzten Modellparameter statistisch aufbereitete probabilistische Vorhersagen für den Verifikationszeitraum generiert. Diese werden darauf wieder rücktransformiert, was einen direkten Vergleich mit den dazugehörigen gemessenen Abflusswerten ermöglicht.

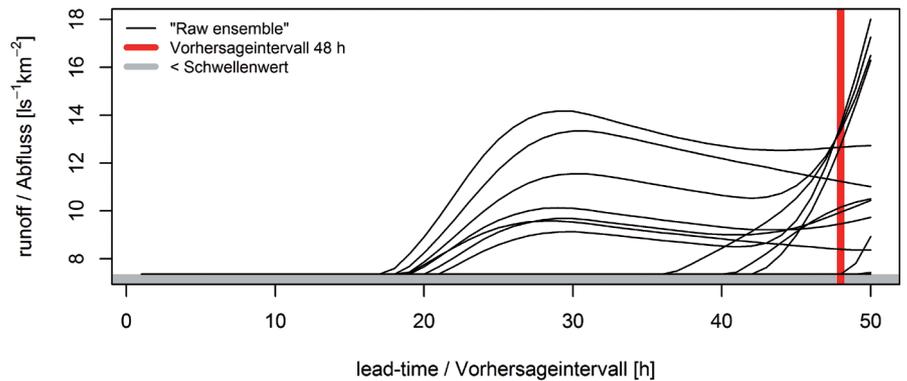


**Abbildung 2**  
Flussdiagramm des Modellschätzungs- und Vorhersageprozesses für zensiertes EMOS  
*Flow chart of the model estimation and forecasting process using censored EMOS*

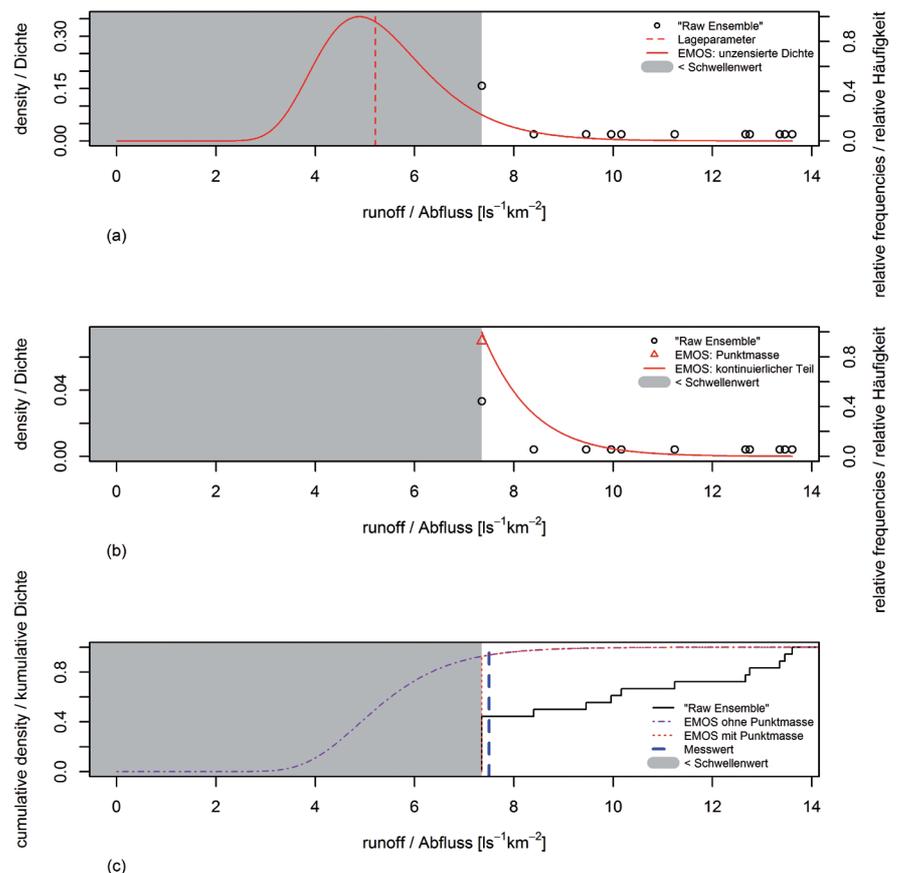
Nun soll das hier verwendete zensierte EMOS Modell konzeptionell erläutert werden. Eine detaillierte mathematische Beschreibung inklusive Literaturverweisen erfolgt in Abschnitt 3.2. Abbildung 3 zeigt beispielhaft die 18 Vorhersagetrajektorien vom 9.5.2009 der Kombination der Ensemblemitglieder aus COSMO-LEPS (16 Mitglieder), DWD-GME (1 Mitglied) und DWD-MER (1 Mitglied) für den mit der Einzugsgebietsgröße normierten Abfluss der Wied am Pegel Friedrichsthal. Die Ensemblemitglieder sind dabei nur identifizierbar, wenn ihr Abflusswert oberhalb des Schwellenwertes von  $5 \text{ m}^3/\text{s}$  ( $7,35 \text{ ls}^{-1}\text{km}^{-2}$ ) liegt, da sie sich ansonsten überdecken. Die in dieser Studie eingeführte zensierte EMOS Methode betrachtet jede Vorhersagezeit einzeln, d.h. zeitliche Abhängigkeiten fließen nur über die unkorrigierten Ensemble-Mitglieder, das "Raw ensemble", nicht aber über die EMOS Parameter, in die EMOS Vorhersage ein. Zur Illustration der zensierten EMOS Vorhersage mit positiver Wahrscheinlichkeit für Werte genau am Schwellenwert wird hier die 48 Stunden Vorhersage verwendet. Diese beinhaltet ein "Raw ensemble", das sowohl aus Werten am Schwellenwert von  $5 \text{ m}^3/\text{s}$  bzw.  $7,35 \text{ ls}^{-1}\text{km}^{-2}$  als auch aus Werten oberhalb des Schwellenwertes besteht.

Abbildung 4(a) zeigt nun einerseits die relativen Häufigkeiten der "Raw ensemble" Werte (Punkte) der obigen 48 h Vorhersage. Hier ist die Häufung von Werten am Schwellenwert zu beachten. Andererseits sind auch der geschätzte Lageparameter und die dazugehörige nicht zensierte EMOS Vorhersagedichte abgebildet. Diese resultieren aus der Anwendung eines unzensierten EMOS Modells mit den aus den Trainingsdaten geschätzten EMOS Parametern auf die Vorhersage-Werte des vorliegenden "Raw ensemble". Aufgrund der Box-Cox Transformation liefert das normalverteilte EMOS Modell eine rechtsschiefe Verteilung. Das EMOS Modell ist so konzipiert, dass in einem Fall mit "Raw ensemble" Werten am Schwellenwert erhebliche Teile der Vorhersagedichte unterhalb des Schwellenwertes zu liegen kommen, womit eine Aussage über die Wahrscheinlichkeitsverteilung unterhalb des Grenzwertes getroffen wird, die in der Realität nicht existiert. Wie Abbildung 4(b) veranschaulicht, werden bei der Anwendung des zensierten EMOS Modells diese Teile zu einer sogenannten Wahrscheinlichkeits-Punktmasse am Schwellenwert integriert. Im vorliegenden Fall liefert das zensierte EMOS

Modell eine Wahrscheinlichkeit von etwas über 90 % für einen Abfluss gleich dem Schwellenwert (in der Realität natürlich kleiner gleich dem Schwellenwert). Die restlichen knapp 10 % der Vorhersage verteilen sich gemäß der daran anschließenden kon-



**Abbildung 3** Beispiel einer "Raw ensemble" Vorhersage von 1–50 h. Jede Trajektorie entspricht genau einem Modell oder Ensemblemitglied aus der Kombination von COSMO-LEPS, DWD-GME und DWD-MER. Man beachte, dass sich die 18 Trajektorien am Schwellenwert überdecken können.  
*Example of a raw ensemble forecast covering lead-times 1–50 h. Each trajectory corresponds to a single forecast model or ensemble member from the combination of COSMO-LEPS, DWD-GME and DWD-MER. Note that the 18 trajectories may overlap at the cutoff value.*



**Abbildung 4** Beispiel einer 48 h zensierten EMOS Vorhersage. Von oben nach unten: nicht zensierte EMOS PDF zusammen mit den relativen Häufigkeiten der "Raw ensemble" Werten; zensierte EMOS PDF mit Punktmasse am Schwellenwert; CDF der unzensierten EMOS Dichte, des "Raw ensemble" und der dazugehörigen Beobachtung  
*Example of a 48 hrs censored EMOS forecast. From top to bottom: non-censored EMOS PDF along with the relative frequencies of raw ensemble values; censored EMOS PDF with point mass at the cutoff value; CDF of uncensored EMOS, censored EMOS, the raw ensemble and the verifying observation*

tinuierlichen Vorhersagedichte auf Werte oberhalb des Schwellenwertes. Diese Mischungsverteilung aus Punktmasse und Vorhersagedichte kann auch wie in Abbildung 4(c) als kumulative Dichte abgebildet werden. Der senkrechte Anstieg von 0 auf über 0,9 am Schwellenwert entspricht hier der Punktmasse. Zum Vergleich sind auch die kumulative Dichte des "Raw ensembles", der entsprechenden unzensierten EMOS Vorhersage und der beobachtete Abflusswert aufgetragen. Dieses Beispiel zeigt die Relevanz der Anwendung spezieller Verfahren zur Berücksichtigung zensierter Daten in der Ermittlung von probabilistischen Abflussvorhersagen.

**3.2 Mathematische Details zu zensiertes EMOS**

Nachdem die Bedeutung der zensierten EMOS Methode in Abschnitt 3.1 konzeptionell vorgestellt wurde, werden nun im Folgenden die mathematischen Details beschrieben. Um die oben beschriebenen statistischen Besonderheiten abzubilden, werden die Daten in einem ersten Schritt mit einer Box-Cox Transformation normalisiert (Abschnitt 3.2.1). Grundlage für die hier entwickelte Methode bildet die in THORARINSDOTTIR & GNEITING (2010) vorgestellte EMOS-Methode für am Schwellenwert trunkierte Vorhersagen (Abschnitt 3.2.2). Trunkiert bedeutet, dass die Verteilungsfunktion (hier Normalverteilung) an einem Schwellenwert abgeschnitten wird, wenn für die betrachtete Variable ein physikalisch begründeter unterer Grenzwert existiert (z.B. Abflüsse kleiner 0 nicht möglich). Da in dem hier betrachteten Fall eine Wahrscheinlichkeit für die Unterschreitung des Grenzwerts existiert (Werte kleiner 5 m³/s), wird die Trunkierung der Verteilungsfunktion durch eine Zensierung ersetzt (SCHEUERER 2013), d.h. dass die Wahrscheinlichkeit für die Unterschreitung als Wahrscheinlichkeitspunktmasse am Grenzwert eingeführt wird (Abschnitt 3.2.3). Auf Grund der großen Schiefe in der Verteilung der Abflüsse an den betrachteten Pegeln können durch die Rücktransformation vom normalen in den realen Raum unrealistisch hohe Abflüsse entstehen. Daher wird die Verteilungsfunktion mit dem Grenzwert des zweifachen Maximums der Klimatologie nach oben begrenzt (Abschnitt 3.2.4). Diese nach oben beschränkte (trunkierte) und nach unten zensierte (Wahrscheinlichkeitspunktmasse) Verteilungsfunktion entspricht der in dieser Studie betrachteten EMOS-Verteilung.

**3.2.1 Box-Cox Transformation**

Abflussdaten sind im Allgemeinen stark rechtsschief verteilt, da Hochwasser seltener auftreten als Mittel- respektive Niedrigwasserabflüsse. Zudem sind Mess- und Prognosefehler bei Hochwasser deutlich größer als bei Mittel- und Niedrigwasserabflüssen. Standard Post-Processing Methoden beruhen aber meist auf der Annahme normalverteilter Vorhersagefehler. Für die hier behandelten Analysen wird die Box-Cox Transformation verwendet, um annähernd normalverteilte Daten zu erhalten:

$$h(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log x, & \lambda = 0 \end{cases} \tag{1}$$

wobei  $x$  die untransformierten Abflüsse bezeichnet und  $\lambda$  der Box-Cox Koeffizient ist (BOX & COX 1964, DUAN et al. 2007). Im Rahmen dieser Studie wird sowohl für die Transformation der Beobachtungen als auch der Vorhersagen dasselbe geschätzte  $\hat{\lambda}$  verwendet. Für die Schätzung wurde die komplette Zeitreihe der Beobachtungen vom 1.11.1998 bis zum 31.10.2008 verwendet. Das heißt,  $\hat{\lambda}$  ist gleich für alle in Tabelle 2 gelisteten Trainings- und Verifikationsperioden. Details zu den Daten finden sich in

Abschnitt 2. Die verwendete Maximum-Likelihood Schätzmethode ist in der Funktion boxcox des R Pakets MASS (VENABLES & RIPLEY 2002) implementiert.

Eine interessante Alternative zur Box-Cox Transformation stellt die log-sinh Transformation dar (WANG et al. 2012). Aufgrund unbefriedigender Ergebnisse in ersten Tests wird hier nicht näher auf diese eingegangen.

**3.2.2 Trunkiertes EMOS**

Wie schon erwähnt, ist EMOS eine Post-Processing Methode für Multi-Modell Vorhersagen (GNEITING et al. 2005). Diese beruht auf den Mittelwerten der einzelnen Ensemble-Modelle sowie auf der Varianz zwischen den Vorhersagen aller Ensemblemitglieder aller Modelle. Da zum Beispiel negative Abflüsse physikalisch unmöglich sind, wird die Vorhersageverteilung auch ohne Zensierung und ohne Transformation auf positive Werte, d.h.  $[0, \infty)$ , beschränkt. Der Einfachheit halber wird nun angenommen, dass sowohl alle Vorhersagewerte als auch alle Beobachtungen oberhalb des Schwellenwertes liegen. Dann kann nach geeigneter Transformation auf eine trunkierte Normalverteilung zurückgegriffen werden. Im Folgenden wird nun eine heteroskedastische EMOS Variante basierend auf einer bei Null trunkierten Normalverteilung vorgestellt (THORARINSDOTTIR & GNEITING 2010). Heteroskedastisch bedeutet, dass die Methode auch eine Heterogenität in der Varianz der Residuen abbilden kann.

Sei  $\bar{r}_k$  das Mittel von Ensemble  $k$ , respektive der Wert vom deterministischen Modell  $k$ . Um eine trunkierte Normalverteilung mit Definitionsbereich  $[0, \infty)$  anwenden zu können, wird das Modell mit den Box-Cox transformierten Daten  $\bar{f}_k = h(\bar{r}_k) - h(d)$ , wobei  $d$  der untere Schwellenwert ist, geschätzt. Beobachtungen werden analog transformiert. Somit ist gesichert, dass alle Werte in  $[0, \infty)$  liegen und der Nullpunkt dem Abfluss am Schwellenwert entspricht. Damit kann die EMOS Vorhersagedichte folgendermaßen geschrieben werden:

$$p(\varphi | \bar{f}_1, \dots, \bar{f}_k) = N^0(\mu, \sigma^2), \tag{2}$$

wobei  $\mu = a + b_1 \bar{f}_1 + b_2 \bar{f}_2 + \dots + b_k \bar{f}_k$  und  $\sigma^2 = c + dS^2$ . Dabei ist

$$S^2 = \frac{1}{\sum_{i=1}^K m_i} \sum_{i=1}^K \sum_{j=1}^{m_i} (f_{i,j} - \bar{f})^2 \tag{3}$$

die Varianz des gesamten Ensembles und

$$\bar{f} = \frac{1}{\sum_{i=1}^K m_i} \sum_{i=1}^K \sum_{j=1}^{m_i} f_{i,j}, \tag{4}$$

wobei  $f_{i,j}$  das Mitglied  $j$  und  $m_i$  die Anzahl Mitglieder des Ensemble-Modells  $i$  bezeichnet. Die Parameter  $a, b_1, \dots, b_k, c, d$  sind alle in  $[0, \infty)$  und werden mittels Minimierung des "Continuous Ranked Probability Scores" (CRPS) über die Trainingsperiode numerisch ermittelt. Der CRPS ist ein weit verbreitetes Verifikationsmaß, das sowohl Kalibrierung wie auch Schärfe einer probabilistischen Vorhersage erfasst. Der CRPS ist gegeben durch:

$$CRPS(P, \varphi) = \int_{-\infty}^{\infty} [P(x) - \mathbb{1}_{[x \geq \varphi]}]^2 dx, \tag{5}$$

wobei  $P$  die kumulative Verteilungsfunktion der Vorhersagen bezeichnet und  $\varphi$  die assoziierte Beobachtung ist (GNEITING & RAFTERY 2007, HERBACH 2000). Der CRPS ist ein negativ orientiertes Gütemaß und wird in der Regel über die Verifikations-

periode gemittelt. Häufig wird der entsprechende "skill score",

$$CRPSS = \frac{CRPS_{ref} - CRPS_{forc}}{CRPS_{ref}}, \quad (6)$$

angegeben, wobei  $CRPS_{forc}$  dem CRPS der Vorhersagen und die Referenz  $CRPS_{ref}$  in dieser Studie dem CRPS der Klimatologie über die Verifikationsperiode entsprechen. Eine perfekte Vorhersage resultiert in einem CRPSS gleich 1, während bei negativem CRPSS die klimatologische Vorhersage vorzuziehen ist. Eine detaillierte Beschreibung der Methode, wie der CRPS über die Trainingsperiode minimiert wird, findet sich in der Arbeit von THORARINS-DOTTIR & GNEITING (2010).

### 3.2.3 Zensiertes EMOS

Die in Abschnitt 3.2.2 eingeführte trunkierte Normalverteilung kann nicht für das Post-Processing der Wied und der Ahr verwendet werden, da ein hoher Anteil der Abflusswerte am Schwellenwert zensiert ist. Dieses Problem kann angegangen werden, indem EMOS mit dem Konzept einer trunkierten Normalverteilung mit Punktmasse am Abschneidepunkt (GNEITING et al. 2004) verbunden wird. Im Unterschied zur trunkierten Verteilung in Abschnitt 3.2.2 wird die Dichte des Bereichs unterhalb des Schwellenwertes von Null (wegen nicht-negativer Abflüsse) nicht auf  $[0, \infty)$  verteilt, sondern als Punktmasse am Schwellenwert aufgefasst. Deshalb wird diese Verteilung im Folgenden punktbeschränkte oder zensierte Normalverteilung genannt. Ist der Schwellenwert Null, so ist die Punktmasse bei Null wie folgt gegeben:

$$p = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^0 \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy, \quad (7)$$

wobei  $\mu$  und  $\sigma^2$  wiederum Funktionen der transformierten Ensemblemitglieder sind. Wie in Abschnitt 3.2.2 sind die transformierten Ensemblemittelwerte gegeben durch  $\bar{f}_k = h(\bar{r}_k) - h(d)$ , wobei  $d$  der untere Schwellenwert ist. Beobachtungen sowie Modellvorhersagen unterhalb  $d$  werden gleich  $d$  gesetzt. Für die transformierten  $\bar{f}_k$  liegt somit der Abschneidepunkt bei 0. Damit ist die EMOS Vorhersagedichte gegeben durch:

$$p(\varphi | \bar{f}_1, \dots, \bar{f}_K) = N^0(\mu, \sigma^2), \quad (8)$$

wobei  $N^0$  hier eine punktbeschränkte Normalverteilung anstelle einer gewöhnlichen trunkierten Normalverteilung bezeichnet. Um auch Punktmassen  $p > 0,5$  am Abschneidepunkt zu ermöglichen, werden einerseits negative Interceptparameter  $a_1$  zugelassen und andererseits ein weiterer Prädiktor eingeführt, nämlich der Anteil Ensemblemittelwerte, die dem Schwellenwert entsprechen (SCHEUERER 2013). Damit stellen sich die Parameter der Normalverteilung wie folgt zusammen: Der Lageparameter  $\mu$  ist gegeben durch

$$\mu = a_1 + a_2 \overline{\mathbb{1}_{\{f=0\}}} + b_1 \bar{f}_1 + b_2 \bar{f}_2 + \dots + b_K \bar{f}_K, \quad (9)$$

wobei

$$\overline{\mathbb{1}_{\{f=0\}}} := \frac{1}{K} \sum_{k=1}^K \mathbb{1}_{\{\bar{f}_k=0\}}$$

den Anteil Ensemblemittelwerte, die exakt dem Schwellenwert entsprechen, bezeichnet. Die Varianz ist gegeben durch  $\sigma^2 = c_1 + c_2 S^2$ , wobei  $S^2$  und  $\bar{f}$  wie in Abschnitt 3.2.2 definiert sind.

Die Parameter  $\mu$  und  $\sigma^2$  werden wiederum mittels CRPS Minimierung über die Trainingsperiode ermittelt. Die Koeffizienten  $a_1$  und  $a_2$  sind reell, während  $b_1, \dots, b_K, c_1, c_2$  nicht-negativ sein müssen.

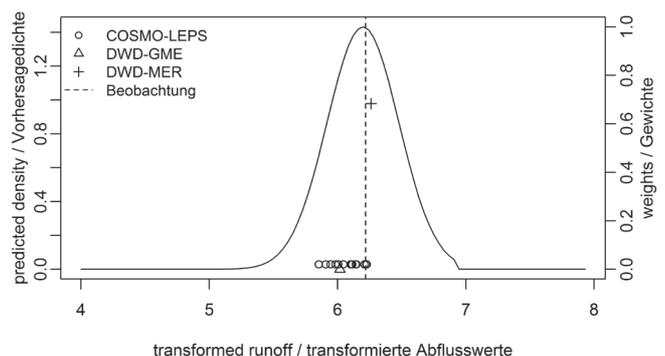
### 3.2.4 Links-zensiertes rechts-trunkiertes EMOS

Die Daten der Wied und der Ahr erfordern eine extreme Box-Cox Transformation, um die Normalverteilungsannahme der verwendeten EMOS Methode zu erfüllen. Dadurch können aber bei der Rücktransformation unrealistisch hohe Werte entstehen. Dies kann verhindert werden, indem die EMOS Modellierung mit einer beidseitig begrenzten – gegen unten zensiert sowie gegen oben trunkiert – Normalverteilung erfolgt. Als oberer Schwellenwert wird hier das Zweifache des Maximums der Klimatologie verwendet. Dementsprechend muss das Modell aus Abschnitt 3.2.3 mit folgender kumulativen Dichtefunktion ersetzt werden:

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{\beta} & , \quad 0 \leq x \leq v \\ 1 & , \quad x > v \end{cases} \quad (10)$$

mit  $\Phi(\cdot)$  als kumulative Dichtefunktion einer Standardnormalverteilung und  $\beta = \Phi\left(\frac{v-\mu}{\sigma}\right)$  als kumulative Dichte am transformierten oberen Schwellenwert  $v$  unter Vernachlässigung der Trunkierung. Zur Illustration werden in Abbildung 5 die EMOS Vorhersagedichtefunktionen für ein Hochwasserereignis der Wied dargestellt. Hier ist zu beachten, dass die Werte auf der Abszisse so transformiert sind, dass der untere Schwellenwert bei Null und der obere bei knapp unter 7 liegt.

Tests mit Vorhersagen für die Ahr haben die Notwendigkeit einer Anpassung der Modellformulierungen für die Parameter  $\mu$  und  $\sigma$  angezeigt, damit trotz Anwendung einer trunkierten Normalverteilung der Erwartungswert der Vorhersagedichte dem gewichteten Mittel der Vorhersagedichten entspricht. Auf eine beim



**Abbildung 5**

Vorhersagedichte einer 21 h Vorhersage für transformierten Abfluss eines Hochwasserereignisses initialisiert um 06:00 Uhr am 13.1.2011. Die Vorhersage wurde mit der naiven EMOS-Version ohne Korrektur für  $\mu$  erstellt. Predicted density of a 21 hrs transformed runoff forecast for a flood event initialized at 06:00 on 13.1.2011. The prediction is obtained by the naive EMOS approach without correction for  $\mu$ .

statistischen Post-Processing häufig durchgeführte Korrektur des systematischen Fehlers (Biaskorrektur) wird aus numerischen Stabilitätsgründen verzichtet. Bei Bedarf kann aber eine Biaskorrektur der EMOS Schätzung vorgeschaltet werden. Der Parameter  $\mu$  wird nun mittels numerischer Optimierung ermittelt und muss folgende Gleichung erfüllen:

$$\mathbb{E}[X|X < v] := \sum_{k=1}^K w_k \bar{f}_k + a \overline{1_{f=0}} = \mu - \sigma \frac{\varphi\left(\frac{v-\mu}{\sigma}\right)}{\Phi\left(\frac{v-\mu}{\sigma}\right)}, \quad (11)$$

wobei  $w_k$  Gewichte darstellen. Die Varianzparameter werden abhängig davon, ob alle Ensemblemitglieder dem unteren Schwellenwert entsprechen oder nicht, unterschiedlich geschätzt. Für erstere wird  $\sigma^2 = c_1 + c_2 S^2$  zu  $\sigma^2 = c_1$ , da  $S^2 = 0$ . Für letztere wird  $\sigma^2 = c_1 + c_2 S^2$  verwendet.

Der Term  $\sigma \frac{\varphi\left(\frac{v-\mu}{\sigma}\right)}{\Phi\left(\frac{v-\mu}{\sigma}\right)}$  korrigiert den Lageparameter  $\mu$  so, dass

die Beschränkung der Verteilung zu keinem systematischen Bias führt.

### 3.3 Verifikationsmaße

Der schon eingeführte CRPS betrachtet das gesamte Spektrum möglicher Vorhersagen und Beobachtungen. Im Falle zensierter Abflussdaten ist es aber von großem Interesse, die Güte der vorhergesagten Wahrscheinlichkeiten für Abflüsse unter  $5 \text{ m}^3/\text{s}$  abzuschätzen. Hierfür kann der Brier Score (Brier 1950) verwendet werden:

$$BS = \frac{1}{n} \sum_{t=1}^n (P[Y_t \leq 5] - \mathbb{1}_{[y_t \leq 5]})^2, \quad (12)$$

der die, über die Verifikationsperiode gemittelte, quadrierte Abweichung zwischen Vorhersagewahrscheinlichkeit und der Realisierung  $- 1$  bei Eintreffen und  $0$  bei Nicht-Eintreffen  $-$  von Vorhersagen unter  $5 \text{ m}^3/\text{s}$  angibt. Der dazugehörige "skill score" ist gegeben durch

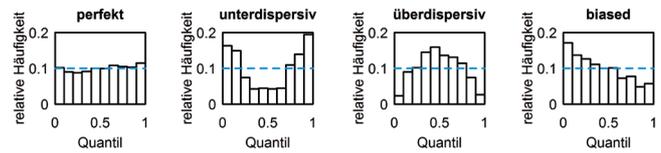
$$BSS = \frac{BS_{ref} - BS_{forc}}{BS_{ref}}, \quad (13)$$

wobei als Referenz eine klimatologische Vorhersage verwendet wird. Eine perfekte Vorhersage führt zu einem BSS von  $1$ , während ein negativer BSS anzeigt, dass die klimatologische Vorhersage überlegen ist.

Neben den schon erwähnten CRPS und BS können Kalibrierung und Schärfe auch getrennt untersucht werden. Die Kalibrierung probabilistischer Vorhersagen wird in der Regel mit dem "Probability Integral Transform" (PIT) untersucht. Das PIT ist wie folgt definiert:

$$z_i = \int_{-\infty}^{y_i} p(y|x_i) dy, \quad (14)$$

wobei  $p(y_i|x_i)$  die Wahrscheinlichkeitsdichte der Vorhersageverteilung für die Zielvariable  $y_i$ , hier Abfluss, bedingt auf das "Raw ensemble"  $x_i$  darstellt. ROSENBLATT (1952) zeigte, dass die  $z_i$  einer perfekt kalibrierten Vorhersage iid  $U(0,1)$  verteilt sind. Das heißt, die Frequenz der Beobachtungen in ein bestimmtes Quantilintervall der Vorhersage zu fallen, entspricht der Breite des Intervalls. Diese Intervalle können gut mit dem PIT Histogramm



**Abbildung 6**

Von links nach rechts: Hypothetische PIT Histogramme für eine gut kalibrierte Vorhersage, eine mit zu kleiner Dispersion, eine mit zu großer Dispersion und eine mit Bias.

*Left to right: Hypothetical PIT histograms for a well calibrated, an underdispersed, an overdispersed, and a biased forecast distribution*

visualisiert werden. Gut kalibrierte Vorhersagen führen zu einem flachen PIT Histogramm, wie in Abbildung 6 dargestellt ist. Andere Formen zeigen hingegen Bias oder Dispersionsfehler, d.h. falsche Streuung der Vorhersage(-verteilung), an.

Die Vorhersageschärfe kann im Falle von unzensierten Vorhersagen einfach abgeschätzt werden. Als Maß kann zum Beispiel der Median der Breite von zentrierten 90 % Vorhersageintervallen über die Verifikationsperiode verwendet werden. Im Falle zensierter Vorhersagen macht dies wenig Sinn, da häufig große Teile des Vorhersageintervalls in der Punktmasse liegen. Die Varianz der Vorhersage stellt hingegen ein Maß dar, das auch bei zensierten Daten verwendet werden kann. Um die kontinuierliche EMOS Vorhersagedichte mit dem "Raw ensemble" vergleichen zu können, wird im Falle der EMOS Dichte die Varianz über die  $\{0.5/18, 1.5/18, \dots, 17.5/18\}$  Quantile berechnet. Die ungewöhnliche Auswahl der Quantile beruht auf der Annahme, dass die Mitglieder des "Raw ensembles" diesen Quantilen einer zugrunde liegenden Verteilung entsprechen sollten.

## 4 Ergebnisse

### 4.1 Verifikation

Im Folgenden soll die zensierte EMOS Methode mit Hilfe der oben genannten Verifikationsmaße evaluiert werden. Alle Werte in diesem Abschnitt verstehen sich als Mittelwerte über den gesamten Verifikationszeitraum, d.h. über alle in Tabelle 2 erwähnten Verifikationsperioden. Um ein Bild der Vorhersagegüte als Kombination von Kalibrierung und Schärfe zu erhalten, soll nun der in Abschnitt 3.2.2 eingeführte CRPSS mit einer klimatologischen Vorhersage als Referenz über die gesamte Verifikationsperiode für die Wied und die Ahr betrachtet werden. Die folgenden Ausführungen gelten sowohl für die Wied als auch die Ahr, da deren CRPSS und BSS Werte in den Abbildungen 7 und 8 ähnliche Verläufe zeigen: Für geringe Vorhersagezeiten bis etwa 10 h sind die CRPSS Werte der EMOS Methode identisch mit den Werten des "Raw ensembles". Danach zeigen sich deutliche Verbesserungen durch das Post-Processing. Aufgrund der Zensierung ist neben des CRPSS' auch die Güte der Prognose von Abflusswerten bis zum Schwellenwert von  $5 \text{ m}^3/\text{s}$  von großer Bedeutung. Dies kann, wie bereits erwähnt, mit Hilfe des BSS' verifiziert werden. Wie aus den Verläufen der BSS Werte ersichtlich ist, kann die Vorhersage zensierter Werte gegenüber dem "Raw ensemble" nicht verbessert werden. Im Falle der Ahr wird der BSS ab einer Vorhersagezeit von 50 h sogar deutlich verschlechtert. Davor sind die Werte für das "Raw ensemble" und EMOS vergleichbar. Des Weiteren fällt auf, dass die Verläufe von CRPSS und BSS bei Weglassen der rechenintensiven Korrektur des Lageparameters  $\mu$  gegenüber der korrekten Methode mit Korrektur nicht verschlechtert werden. In den Abbildungen ist die Me-

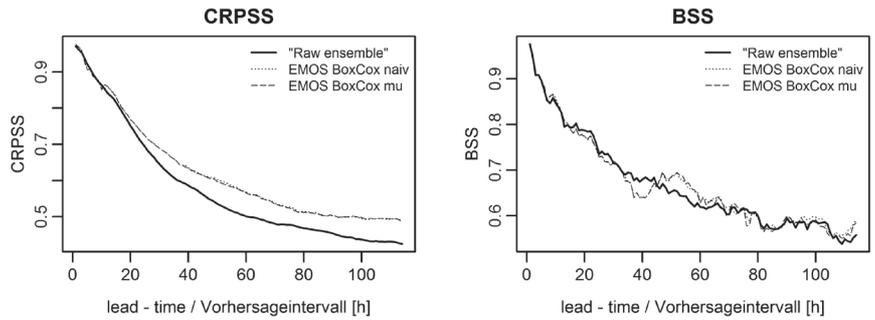
thode ohne Korrektur des Lageparameters als „naiv“ sowie diejenige mit Korrektur als „mu“ gekennzeichnet.

Im Folgenden sollen Kalibrierung und Vorhersageschärfe näher untersucht werden. Wie Abbildungen 9 und 10 zeigen, führt zensiertes EMOS bei beiden Gebieten zu guter Kalibrierung über alle Vorhersagezeiten, während die „Raw ensembles“ unterdispersiv sind. Hier sei angemerkt, dass bei Vorhersagen mit über 10-prozentiger Zensierungswahrscheinlichkeit zensierte Beobachtungen anteilmäßig einem der möglichen Quantilintervallen zugeordnet werden. So wird zum Beispiel eine zensierte Beobachtung bei einer vorhergesagten Zensierungswahrscheinlichkeit von 15 % zu 2/3 dem ersten Dezantil und zu 1/3 dem zweiten Dezantil des PIT Histogramms zu geordnet.

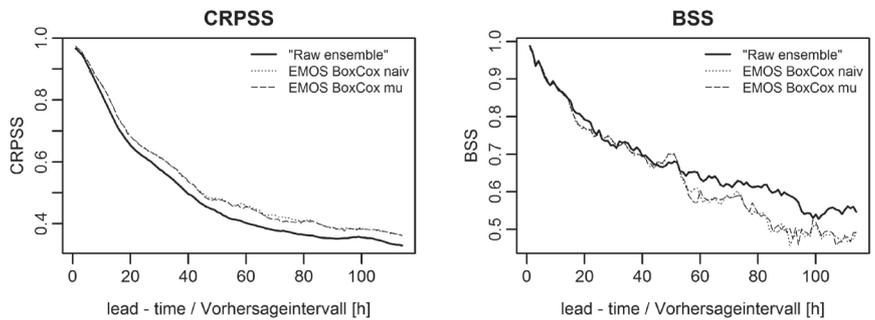
Neben der Kalibrierung soll nun auch die Vorhersageschärfe untersucht werden. Wie in Abschnitt 3.3 erwähnt, wird hierfür die Varianz der Vorhersagen betrachtet. Abbildung 11 zeigt die über die Verifikationstage gemittelten Varianzen gegen die Vorhersageintervalle sowohl für die Wied als auch die Ahr. Da die Varianz stark von Ausreißern beeinflusst wird, wird hier zur robusteren Schätzung ein 0,05 getrimmtes Mittel über die Verifikationstage verwendet. Das heißt, dass für alle Vorhersagezeiten zur Berechnung der mittleren Varianz die Verifikationstage mit den 5 % kleinsten sowie den 5 % größten Varianzen nicht berücksichtigt werden. Bei beiden Gebieten führen sowohl das „Raw ensemble“ als auch zensiertes EMOS zu einer gegenüber einer klimatologischen Vorhersage deutlich verbesserten Vorhersageschärfe. Die „Raw ensembles“ sind leicht schärfer als zensiertes EMOS. Dies ist aber aufgrund ihrer Unterdispersion zu erwarten. Vielmehr ist der nur sehr kleine Unterschied zwischen „Raw ensembles“ und zensiertem EMOS ein Indiz dafür, dass zensiertes EMOS im Mittel zu sehr scharfen Vorhersagen führt.

**4.2 Beispielvorhersage**

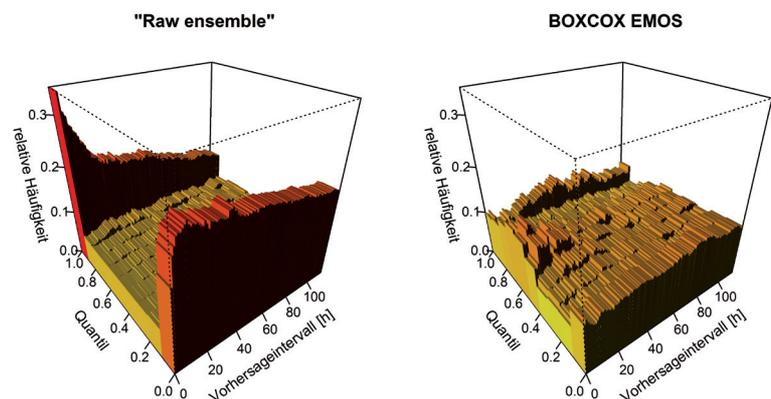
Die Verifikationsresultate aus Abschnitt 4.1 zeigen, dass EMOS in der Lage ist, Post-Processing mit zensierten Abflussdaten durchzuführen. Als Beispiel einer Vorhersage am Übergang von zensiertem zu unzensiertem Abfluss werden in Abbildung 12 die probabilistischen Vorhersagen des „Raw ensembles“ sowie von zensiertem EMOS vom 7.11.2010 für die Wied gezeigt. Neben der guten marginalen Vorhersage der EMOS Variante zeigt dieses Beispiel



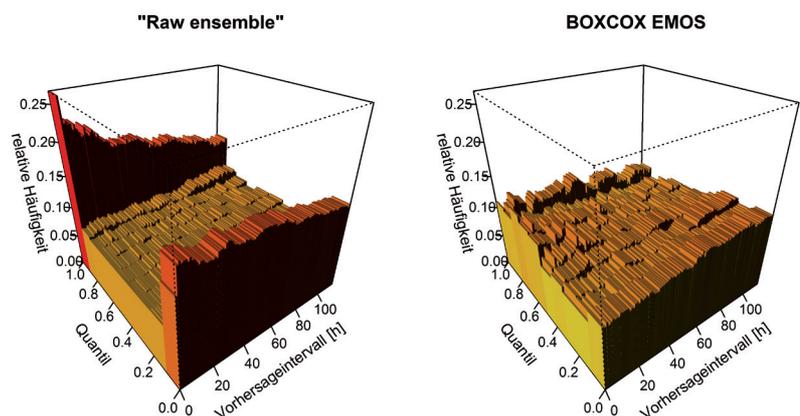
**Abbildung 7**  
CRPSS und BSS für Abflüsse an der Zensierungsgrenze von 5 m³/s für die Wied bei Friedrichsthal  
*CRPSS and BSS for discharge values at the censoring threshold of 5 m³/s for river Wied at Friedrichsthal*



**Abbildung 8**  
CRPSS und BSS für Abflüsse an der Zensierungsgrenze von 5 m³/s für die Ahr bei Altenahr  
*CRPSS and BSS for discharge values at the censoring threshold of 5 m³/s for river Ahr at Altenahr*



**Abbildung 9**  
3D PIT Histogramme für die Wied bei Friedrichsthal. Die EMOS Methode ist mit Korrektur für  $\mu$   
*3D PIT histograms for river Wied at Friedrichsthal. The EMOS method is with  $\mu$  correction*



**Abbildung 10**  
3D PIT Histogramme für die Ahr bei Altenahr. Die EMOS Methode ist mit Korrektur für  $\mu$   
*3D PIT histograms for river Ahr at Altenahr. The EMOS method is with  $\mu$  correction*

auch die Notwendigkeit einer multivariaten EMOS Methode. So weisen die vorhergesagten Quantile Sprünge von Vorhersagezeit zu Vorhersagezeit auf, die physikalisch schwer erklärbar sind.

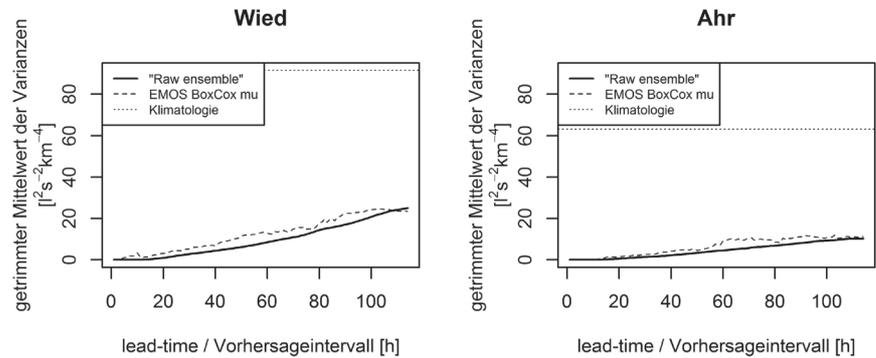
### 5 Diskussion

Gemäß den hier vorgestellten Analysen konnte aufgezeigt werden, dass die probabilistischen Vorhersagen für Wied und Ahr mittels statistischem Post-Processing in Bezug auf die Verifikationsmaße CRPSS und PIT Histogramme verbessert werden können ohne eine deutliche Verschlechterung der Vorhersageschärfe hinnehmen zu müssen. Dies gilt insbesondere für längere Vorhersagezeiten. Dies zeigt, dass Gebiete mit einem hohem Anteil zensierter Datenpunkte mit einer angepassten EMOS Methode, die auf einer Mischung aus Dichtefunktion und Punktmasse beruht, modelliert werden können. Aufgrund der Art, wie zensiertes EMOS Werte unter beziehungsweise über einem Schwellenwert behandelt, stellt sich die Frage, ob zensiertes EMOS auch für Wahrscheinlichkeitsausagen einer Unter- oder Überschreitung eines Schwellenwertes geeignet ist. Bei unzensierten größeren Gebieten wie zum Beispiel der Mosel am Pegel Trier könnte künstlich rechtszensiertes EMOS bei Hochwassersituationen zu Verbesserungen der Wahrscheinlichkeitsvorhersagen führen. Insbesondere bei Messunsicherheiten ist zensiertes EMOS in der Lage, Fragestellungen wie „Mit welcher Wahrscheinlichkeit überschreitet der Abfluss am Pegel Trier in 48 h die Hochwasser-Meldestufe I von 823 m<sup>3</sup>/s“ adäquat zu behandeln. Deshalb sollen in Zukunft auch weitere Teileinzugsgebiete des Rheins sowie anderer Flüsse berücksichtigt werden.

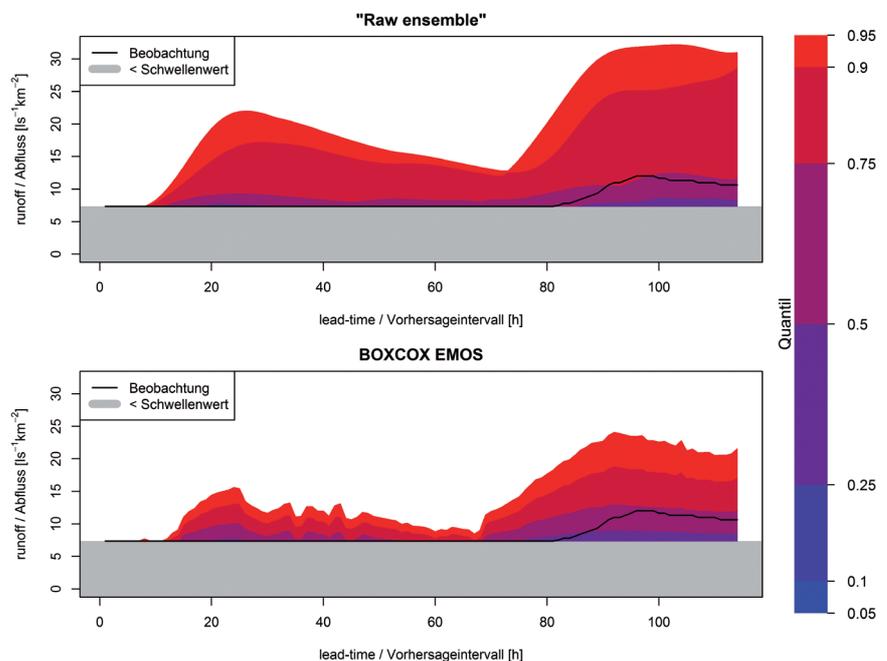
Aufgrund der unbefriedigenden BSS Werte für die Ahr ist aber klar, dass zensiertes EMOS weiterer Verbesserungen bedarf. Des Weiteren sollten auch alternative zensierte Methoden in Betracht gezogen werden. Denn bis jetzt wurde noch keine Alternative zu zensiertem EMOS angewandt. Zensiertes EMOS stellt aber eine vielversprechende Methode dar, da sie relativ einfach ist und die Parameter mit geringem Rechenaufwand geschätzt werden können. Eine Alternative zu zensiertem EMOS könnte eine Art zensiertes BMA darstellen. Dies würde aber eine deutlich komplexere Modellierung voraussetzen. Ein guter Ansatzpunkt hierfür sind schon vorhandene Arbeiten zu BMA basierend auf einer Mischung trunkierter Normalverteilungen (BARAN 2013).

Neben der Erweiterung auf bisher nicht betrachtete Einzugsgebiete des Rheins

sowie Einzugsgebiete anderer Flüsse und auf zusätzliche Ensemblemitglieder sowie die weitere Verbesserung der univariaten zensierten EMOS Methode, sollen in Zukunft auch multivariate EMOS Methoden zum Einsatz kommen. Damit sollen einerseits die Abhängigkeitsstruktur zwischen den Vorhersagezeiten aus den „Raw ensembles“ erhalten bleiben und andererseits Sprünge in den marginalen Vorhersagedichten von Vorhersagezeit zu Vorhersagezeit vermieden werden. Zur Zeit beruhen die hierfür verfügbaren Methoden entweder auf multivariaten Normalverteilungen (HEMRI et al. 2013, PINSON et al. 2009, PINSON &



**Abbildung 11**  
Getrimmtes Mittel der Varianz der Vorhersagen gegen das Vorhersageintervall. Getrimmtes Mittel bedeutet hier, dass der Mittelwert für ein bestimmtes Vorhersageintervall nur aus den Werten zwischen dem 0,05 und dem 0,95 Quantil der Vorhersagevarianzen aller Verifikationstage berechnet wurde.  
*Trimmed mean of the forecast variances versus lead-time. Trimmed mean here denotes that the mean value for a certain lead-time has only been calculated from values between the 0.05 and the 0.95 quantile of the forecast variances of all verification days.*



**Abbildung 12**  
„Raw ensemble“ und zensierte EMOS Vorhersage für die Wied bei Friedrichsthal initialisiert am 7.11.2010 um 06:00 Uhr. Die Quantile des „Raw ensembles“ wurden mittels linearer Interpolation zwischen benachbarten Ensemblemitgliedern berechnet.  
*Raw ensemble and censored EMOS forecast for the Wied at Friedrichsthal initialized on 7.11.2010 at 06:00 hrs. The quantiles of the raw forecasts have been obtained by linear interpolation between adjacent ensemble members.*

GIRARD 2012) oder der aufgrund ihrer Einfachheit sehr attraktiven Ensemble Copula Coupling (ECC) Methode (SCHEFZIK et al. 2013). ECC kann aber nur die Abhängigkeitsstruktur aus dem "Raw ensemble" wiedergeben, was dazu dient, realistische Szenarien zu bilden. Um die Sprünge in den marginalen Verteilungen von einer Vorhersagezeit zur nächsten zu verhindern, muss EMOS selbst angepasst werden. Hierzu stehen unterschiedliche Ansätze zu Verfügung wie z.B. ein gleitendes Fenster von Vorhersagezeiten für die Parameterschätzung, d.h. zuerst werden Vorhersagen von 1 h bis 5 h betrachtet, dann von 3 h bis 7 h usw. Alternativ dazu könnte auch ein gleitendes Mittel der univariaten Parameterschätzungen zu einem zufriedenstellenden Ergebnis führen. Die obigen Ausführungen stellen keine komplette Auswahl aller möglichen multivariaten Post-Processing Methoden dar. Weitere hier noch nicht betrachtete Methoden werden deshalb mit großer Wahrscheinlichkeit in die noch folgenden Arbeiten einfließen.

## 6 Zusammenfassung und Schlussfolgerungen

In der vorliegenden Studie wurde eine EMOS Variante eingeführt, die mit zensierten Vorhersage- respektive Beobachtungsdaten umgehen kann. Die zensierte EMOS Methode wurde über einen Vorhersagehorizont von 1 bis 114 Stunden an den Gebieten Wied und Ahr getestet. Dabei wurden Multi-Modell Ensemblevorhersagen statistisch nachbearbeitet. Für beide Gebiete konnten die probabilistischen Vorhersagen im Sinne der betrachteten statistischen Gütemaße verbessert werden.

Aufgrund der Ergebnisse kann davon ausgegangen werden, dass statistische Nachbearbeitung mit zensiertem EMOS die probabilistischen Abflussvorhersagen für Einzugsgebiete mit zensierten Beobachtungen und Vorhersagen verbessern kann. In Zukunft sollen deshalb folgende Ziele verfolgt werden: Einerseits wurden bis jetzt nur Vorhersagen für zwei Gebiete getestet. Deshalb soll zensiertes EMOS an weiteren, möglichst unterschiedlichen Einzugsgebieten geprüft werden. Andererseits kann die zensierte EMOS Methode sicherlich noch weiter verbessert werden, und vor allem sollten zum Methodenvergleich weitere Post-Processing Methoden entwickelt werden, die mit zensierten Daten umgehen können. Des Weiteren soll künftig das Potenzial von zensiertem EMOS für andere Anwendungen wie Wahrscheinlichkeitsaussagen zur Unter- bzw. Überschreitung von Schwellenwerten ermittelt werden.

### Summary and conclusions

This study has introduced a novel EMOS variant that is able to statistically post-process censored forecast and observed data in a sound way. We have tested the censored EMOS method using the catchments Wied and Ahr over the entire forecast time horizon from 1 to 114 hours. For both catchments probabilistic forecasts could be improved in terms of the statistical verification scores considered.

On the basis of the present study it can be assumed that censored statistical post-processing using censored EMOS leads to runoff forecasts with improved statistical properties in case of catchments with a high proportion of censored forecasts and observations. Therefore, future work should pursue the following aims: On the one hand we have tested censored EMOS only in two catchments. This requires censored EMOS to be tested in

several other catchments differing as much as possible. On the other hand censored EMOS is still under development and can certainly be further improved. Moreover, other methods that are able to cope with censored data should be developed in order to evaluate the benefit of censored EMOS in a sound way. Additionally, the potential of censored EMOS for other applications like probability statements on undercutting or exceeding of predefined thresholds should be evaluated.

### Danksagungen

Die Autoren danken der Bundesanstalt für Gewässerkunde für die Finanzierung dieser Arbeit sowie die Bereitstellung der Daten. Der Dank der Autoren gilt auch den zwei anonymen Gutachtern, die mit ihren Kommentaren zu einer deutlichen Verbesserung dieser Arbeit beigetragen haben. Des Weiteren soll hier T. Gneiting, M. Scheuerer und A. Jordan von der Universität Heidelberg für viele hilfreiche Diskussionen und Anregungen gedankt werden sowie D. Meißner und den Kollegen der Bundesanstalt für Gewässerkunde für die gute Zusammenarbeit. Bedanken möchten sich die Autoren auch bei T. Thorarinsdottir vom "Norwegian Computing Center" für die Bereitstellung ihrer EMOS Skripte.

### Anschriften der Verfasser:

S. Hemri  
CST group  
HITS gGmbH  
Schloss-Wolfsbrunnenweg 35  
69118 Heidelberg  
stephan.hemri@h-its.org

D. Lisniak  
Dr. B. Klein  
Bundesanstalt für Gewässerkunde  
Am Mainzer Tor 1, 56068 Koblenz

### Literaturverzeichnis

- AJAMI, N.K., Q. DUAN & S. SOROOSHIAN (2007): An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. – *Water Resources Research* 43; DOI: 10.1029/2005WR004745
- BARAN, S. (2013): Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components. – Pre-print arXiv:1305.1184 [stat.ME]
- BERGSTRÖM, S. (1995): The HBV model. – In: V.P. Singh (Ed.): *Computer Models of Watershed Hydrology*. – Water Resources Publications. Highland Ranch, CO.
- BOGNER, K., F. PAPPENBERGER & H.L. CLOKE (2013): Model combination and weighting methods in operational flood forecasting. – *Geophysical Research Abstracts* 15, 13629
- BOX, G. & D. COX (1964): An analysis of transformations. – *Journal of the Royal Statistical Society Series B-Statistical Methodology* 26, 211–252
- BRIER, G.W. (1950): Verification of forecasts expressed in terms of probability. – *Monthly Weather Review* 78, 1–3
- COCCIA, G. & E. TODINI (2011): Recent developments in predictive uncertainty assessment based on the model conditional processor approach. – *Hydrology and Earth System Sciences* 15, 3253–3274

- DUAN, Q., N.K. AJAMI, X. GAO & S. SOROOSHIAN (2007): Multi-model ensemble hydrologic prediction using Bayesian model averaging. – *Advances in Water Resources* 30 (5), 1371–1386
- FRALEY, C., A.E. RAFTERY & T. GNEITING (2010): Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. – *Monthly Weather Review* 138 (1), 190–202
- GNEITING, T., K. LARSON & K. WESTRICK (2004): Development of Next-Generation Wind Energy Forecast and Optimization Technologies. – Washington Technology Center, Seattle, WA.
- GNEITING, T. & A.E. RAFTERY (2007): Strictly proper scoring rules, prediction and estimation. – *J. Am. Statist. Ass.* 102, 359–378
- GNEITING, T., A.E. RAFTERY, A.H. WESTVELD & T. GOLDMAN (2005): Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. – *Monthly Weather Review* 133, 1089–1118
- HEMRI, S., F. FUNDEL & M. ZAPPA (2013): Simultaneous calibration of ensemble river flow predictions over an entire range of lead-times. – *Water Resources Research* 49; DOI: 10.1002/wrcr.20542
- HERSBACH, H. (2000): Decomposition of the continuous ranked probability score for ensemble prediction systems. – *Weather and Forecasting* 15 (5), 559–570
- KRZYSZTOFOWICZ, R. & K.S. KELLY (2000): Hydrologic uncertainty processor for probabilistic river stage forecasting. – *Water Resources Research* 36, 3265–3277
- LINDSTRÖM, G., B. JOHANSSON, M. PERSSON, M. GARDELIN & S. BERGSTRÖM (1997): Development and test of the distributed HBV-96 hydrological model. – *Journal of Hydrology* 201(1–4), 272–288
- MADADGAR, S., H. MORADKHANI & D. GAREN (2012): Towards improved post-processing of hydrologic forecast ensembles. – *Hydrological Processes*, Online library; DOI: 10.1002/hyp.9562
- MAJEWSKI, D., D. LIERMANN, P. PROHL, B. RITTER, M. BUCHHOLD, T. HANISCH, G. PAUL, W. WERGEN & J. BAUMGARDNER (2002): The operational global icosahedral-hexagonal gridpoint model GME: Description and high-resolution tests. – *Monthly Weather Review* 130, 319–338
- MAJEWSKI, D., D. LIERMANN & B. RITTER (2012): Kurze Beschreibung des Globalmodells GME (20 km/L60) und seiner Datenbanken auf dem Datenserver des DWD. – *Deutscher Wetterdienst (DWD)*, Offenbach
- MEIßNER, D. & S. RADEMACHER (2010): Die verkehrsbezogene Wasserstandsvorhersage für die Bundeswasserstraße Rhein. – *KW Korrespondenz Wasserwirtschaft* 3 (9), 485–491
- MONTANI, A., D. CESARI, C. MARSIGLI & T. PACCAGNELLA (2010): Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. – *COSMO Technical Report* 19
- PARRISH, M.A., H. MORADKHANI & C.M. DE'CHANT (2012): Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. – *Water Resources Research* 48; DOI: 10.1029/2011WR011116
- PINSON, P., H. MADSEN, H.A. NIELSEN, G. PAPAETHYMIU & B. KLÖCKL (2009): From probabilistic forecasts to statistical scenarios of short-term wind power production. – *Wind Energy* 12, 51–62
- PINSON, P. & R. GIRARD (2012): Evaluating the quality of scenarios of short-term wind power generation. – *Applied Energy* 96, 12–20
- R CORE TEAM (2012): R: A Language and Environment for Statistical Computing. – R: Foundation for Statistical Computing, Wien
- RAFTERY, A.E., T. GNEITING, F. BALABDAOUI & M. POLAKOWSKI (2005): Using Bayesian model averaging to calibrate forecast ensembles. – *Monthly Weather Review* 133 (2), 1155–1174
- REGGIANI, P., M. RENNER, A.H. WEERTS & P.A.H.J.M. VAN GELDER (2009): Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system. – *Water Resources Research* 45; DOI: 10.1029/2007WR006758
- RINGS, J., J.A. VRUGT, G. SCHOUPS, J.A. HUISMAN & H. VERECKEN (2012): Bayesian model averaging using particle filtering and Gaussian mixture modeling: Theory, concepts, and simulation experiments. – *Water Resources Research* 48; DOI: 10.1029/2011WR011607
- ROSENBLATT, M. (1952): Remarks on multivariate transformation. – *Annals of Mathematical Statistics* 23, 470–472
- SCHEFZIK, R., T.L. THORARINSDOTTIR & T. GNEITING (2013): Uncertainty quantification in complex simulation models using ensemble copula coupling. – *Statistical Science* 28, 616–640
- SCHUEERER, M. (2013): Probabilistic quantitative precipitation forecasting using ensemble model output statistics. – *Quarterly Journal of the Royal Meteorological Society*; DOI: 10.1002/qj.2183
- SCHULZ, J.-P. & U. SCHÄTTLER (2011): Kurze Beschreibung des Lokal-Modells Europa COSMO-EU (LME) und seiner Datenbanken auf dem Datenserver des DWD. – *Deutscher Wetterdienst (DWD)*, Offenbach
- STEPPELER, J., G. DOMS & G. ADRIAN (2002): Das Lokal-Modell LM. – *promet* 27 (3/4), 123–128
- THORARINSDOTTIR, T.L. & T. GNEITING (2010): Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. – *Journal of the Royal Statistical Society* 173 (Series A), 371–388
- TODINI, E. (2008): A model conditional processor to assess predictive uncertainty in flood forecasting. – *International Journal of River Basin Management* 6, 123–137
- VENABLES, W.N. & B.D. RIPLEY (2002): *Modern Applied Statistics with S*. Fourth Edition. – Springer, New York; ISBN 0-387-95457-0
- WANG, Q.J., D.L. SHRESTHA, D.E. ROBERTSON & P. POKHREL (2012): A log-sinh transformation for data normalization and variance stabilization. – *Water Resources Research* 48; W05514