

6-Wochen-Vorhersage Güte-Skala – Wie entsteht sie und wie ist sie zu verstehen?

Zusätzlich zur bisherigen Vorhersageabbildung wird nun auch die Güte der Vorhersagen mit Hilfe einer 5-teiligen Skala dargestellt. Diese zeigt die Vorhersagegüte abhängig vom Vorhersagezeitpunkt im Jahr. Die Werte ändern sich monatlich, abhängig von dem Monat, in dem der Vorhersagezeitpunkt liegt. Neben der Darstellung der Saisonalität der Güte, bietet sich so ebenfalls eine Vergleichs- und Einordnungsmöglichkeit der Qualität der hydrologischen 6-Wochen-Vorhersage von Pegel zu Pegel.

Vorhersagequalität*:



*abgeleitet aus den Oktober-Vorhersagen des Evaluierungszeitraums 2016-2023

Abbildung 1: Beispiel, wie die Güte der Vorhersage im PDF-Bericht abgebildet wird. Sie ändert sich abhängig von dem Monat in dem der Vorhersagezeitpunkt liegt. Hier handelt es sich um eine Vorhersage aus dem Oktober.

Vorhersagequalität:



Abbildung 2: Legende der Güte-Skala

Einflussfaktoren der Vorhersagequalität

Jede Vorhersage ist mit Unsicherheiten behaftet. Diese ist, wie die Vorhersage selbst, dynamisch. Grundsätzlich gilt eine Zunahme der Unsicherheit über den Vorhersagezeitraum.

Die Einflussfaktoren sind dabei variabel:

Eigenschaften des Einzugsgebietes:

Hydrologische Vorhersagen für größere Einzugsgebiete sind in der Regel genauer, als für kleinere, da sich hier Fehler der numerischen Wettervorhersage und Anfangsbedingungen eher ausgleichen und mehr meteorologische und hydrologische Messdaten für das Gebiet vorhanden sind. Eine Abflussvorhersage in Einzugsgebieten mit längeren Fließzeiten aufgrund von geringerem Gefälle ist ebenfalls einfacher, als in steilen Einzugsgebieten, da die Niederschlagsmessung für einen längeren Zeitraum die Abflussbildung bestimmt und erst später die Unsicherheit der Wettervorhersage eine Rolle spielt. Die Qualität der hydrologischen Modellierung ist außerdem abhängig von weiteren, komplexen Gebietseigenschaften, wie z. B. Karst, wasserbaulichen BfG – Hintergründe der 6-Wochen-Vorhersage Güte-Skala

Maßnahmen oder Speicherregulierungen, die schwieriger exakt im Modell abzubilden sind.

Fehlende oder unsichere Beobachtungsdaten:

Wenige oder fehlerhafte hydrologische und meteorologische Messdaten erschweren es, den Ist-Zustand im Modell korrekt abzubilden. Daten, die nicht ausreichend in der Fläche vorhanden sind, wie zum Beispiel das Wasseräquivalent der Schneedecke, die Verdunstung oder die Bodenfeuchte, werden aus anderen gemessenen Daten berechnet, was ebenfalls zu Unsicherheiten führt. All dies spielt für die Anfangsbedingungen im Modell eine entscheidende Rolle. Unsicherheiten in den Anfangsbedingungen beeinflussen direkt die Berechnungen für die Zukunft.

Wettervorhersage:

Auch die numerische Wettervorhersage, die neben den Messdaten die Basis für die hydrologische Vorhersage bildet, ist mit Unsicherheiten behaftet, welche sich in der Modellkette weiter durchpaust. Abweichungen im räumlichen und zeitlichen Verlauf des vorhergesagten Niederschlages haben einen direkten Effekt auf den Wasserstand am betrachteten Pegel.

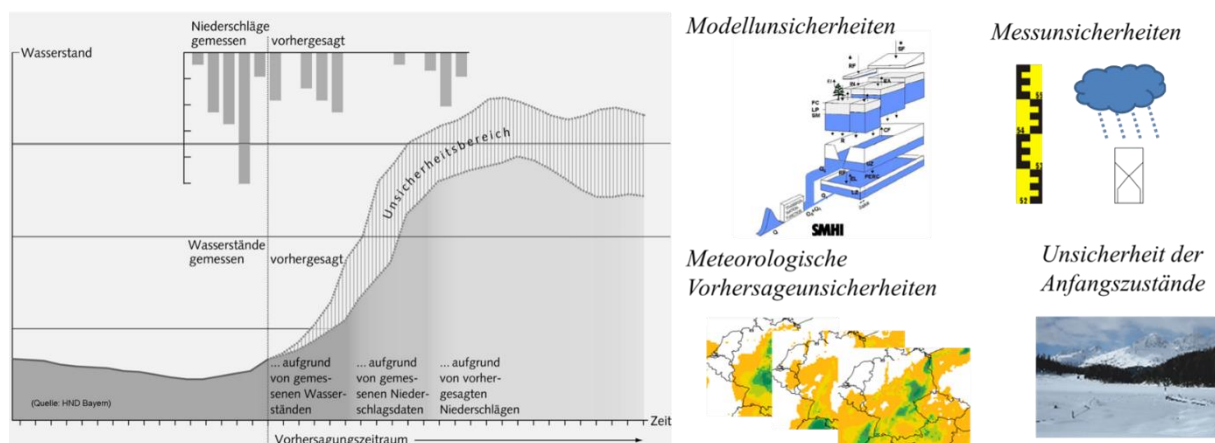


Abbildung 3: Zusammenfassung der Unsicherheiten in der hydrologischen Vorhersage. (Quelle: HND Bayern (links), BfG (rechts))

Berechnung der statistischen Vorhersagegüte

Um trotz der oben genannten Unsicherheiten eine Vorhersage richtig einschätzen zu können, ist es notwendig die Güte zu kennen. Weil es natürlich für die neuste Vorhersage noch keine Beobachtungswerte zum Vergleich gibt, ist es wichtig, den Blick in die Vergangenheit zu werfen. Es gilt die Annahme: Wenn mit dem gleichen Modell in der Vergangenheit gute Vorhersagen berechnet werden konnten, werden auch die zukünftigen Vorhersagen ähnlich gut sein. Diese Annahme kann getroffen werden, da viele Prozesse im hydrologischen System wiederkehrend sind, wie die Schneeschmelze, aber auch saisonale Perioden niedrigerer oder höherer Wasserstände. Diese Prozesse sorgen auch dafür, dass die Vorhersagequalität nicht das ganze Jahr über gleich ist.

Die bei der BfG verwendeten Modelle sind aufgrund ihres Anwendungszweckes primär auf Niedrig- und Mittelwasser kalibriert und können in Folge dessen Hochwasserabflüsse weniger gut abbilden.

Zur Berechnung der statistischen Vorhersagegüte braucht es möglichst viele Vorhersagen der Vergangenheit und hydrologische Messwerte zum Vergleichen. Bei der 6-Wochen-Vorhersage bedeutet das, dass die Boxplots der Vorhersage mit den Wochenmittelwerten der vergangenen Beobachtung verglichen werden.

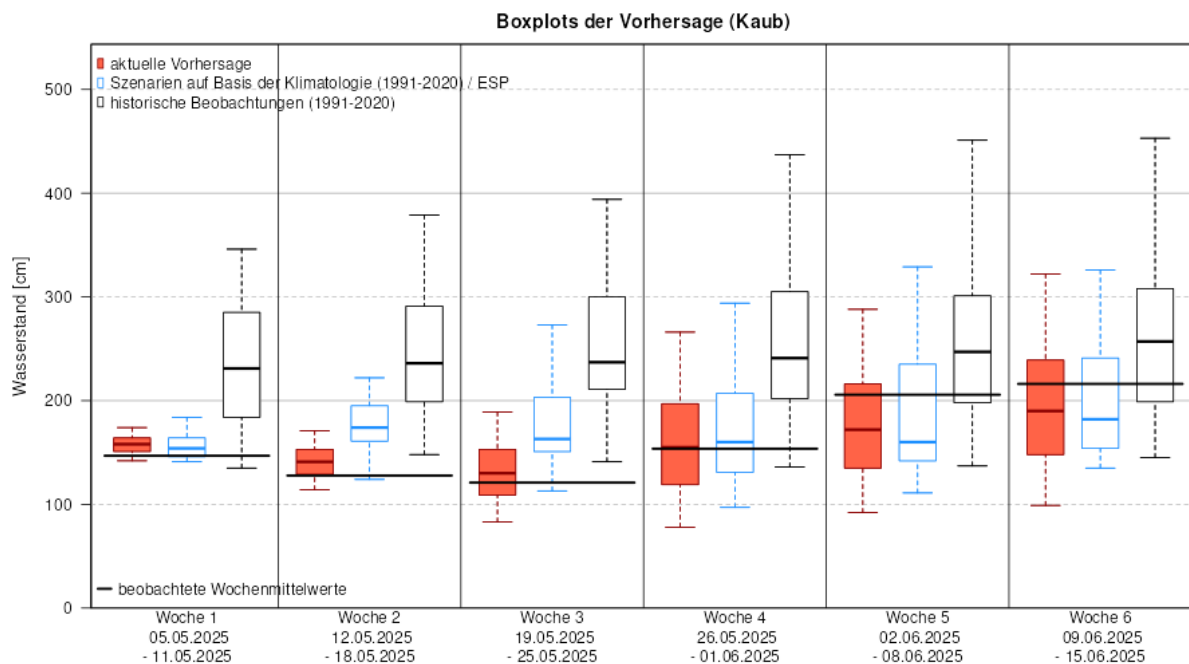


Abbildung 4: Rückschau der Vorhersage am Rheinpegel Kaub vom 05.05.2025. Die Vorhersage liegt bereits vollständig in der Vergangenheit, daher kann Sie mit den bereits bekannten beobachteten Wochenmittelwerten (schwarze Striche) verglichen werden.

Da es sich um eine wahrscheinlichkeitsbasierte Vorhersage handelt, kann nicht einfach eine Wasserstandsganglinie mit einer anderen abgeglichen werden. Für die Gütebetrachtung probabilistischer Vorhersagen gibt es verschiedene Gütemaße, welche unter anderem statistisch abbilden, ob die Messwerte im Bereich der höchsten Wahrscheinlichkeit der Vorhersage liegen und ob das Unsicherheitsband realistisch abgebildet wird.

Da die Berechnung einer absoluten Vorhersagegüte nur die Aussage zulässt, wie weit die Vorhersage im Mittel von der Beobachtung abweicht, aber keine Wertung derselben, wird für die Güte-Skala eine relative Vorhersagegüte berechnet. Dies geschieht auf Grundlage vergangener Vorhersagen von 2016-2023. Hier wird die absolute Vorhersagegüte (die mittlere Abweichung der roten Boxplots in Abbildung 4 von den beobachteten Wochenmittelwerten) mit der absoluten Güte einer sog. Referenzvorhersage verglichen. In diesem Fall wird die Verteilung der historischen hydrologischen Beobachtungen (schwarze Boxplots in Abbildung 4) als

Referenzvorhersage genutzt. Sie bildet das natürliche Verhalten des Flusses in Bezug auf Jahresgang und Maximalabflüsse ab.

Analog zu den Boxplots, wo sich Vorhersage in rot und Referenzvorhersage in schwarz wiederfinden, lassen sich auch die Kuchendiagramme der 5-Klassen-Vorhersage betrachten. Neben den ungleich verteilten Kuchendiagrammen der Vorhersage bildet die Legende mit ihrer Gleichverteilung die Referenzvorhersage ab. Diese ist über das ganze Jahr in 5 gleich wahrscheinliche Kuchenstücke unterteilt, da sich die Grenzen dieser Kuchenstücke, welche in der Tabelle angegeben sind, von Woche zu Woche ändern.

Die unterschiedlichen Vorhersageauswertungen der 6-Wochen-Vorhersage werden mit unterschiedlichen jeweils passenden Gütemaßen analysiert.

Boxplotvorhersage: Continuous Ranked Probability Skill Score (CRPSS)

5-Klassen-Vorhersage: Ranked Probability Skill Score (RPSS)

UnterschreitungsVorhersage: Brier Skill Score (BSS)

Weiterführende Informationen zu den Gütemaßen finden Sie hier: <https://confluence.ecmwf.int/display/FUG/Section+12.B+Statistical+Concepts+-+Probabilistic+Data>

Hintergründe der Güte-Skala

Anhand der Ergebnisse der Vorhersagegüte wird die Güte-Skala (Abbildung 1) erstellt. Sie zeigt, wie die Vorhersagequalität im Monat des Vorhersagezeitpunktes einzuschätzen ist (Abbildungen des Jahresverlaufs finden sich im Anhang).

Für alle verwendeten Gütemaße gilt:

Bei einem Wert von 1 ist eine perfekte Vorhersage erreicht. Bei Werten zwischen 1 und 0 ist die aktuelle Vorhersage das beste Werkzeug, bei 0 kann weder die Vorhersage, noch die Referenzvorhersage als besser eingestuft werden und bei Werten kleiner 0 ist die Referenzvorhersage der Vorhersage überlegen.

Zusätzlich zur Güteberechnung wird statistisch geprüft, wie wahrscheinlich eine Über- oder Unterschreitung des Grenzwertes 0 ist, also ob diese signifikant ist. Ist sie nicht signifikant, kann sie statistisch gesehen zufällig aufgetreten sein und wird damit nicht bestätigt.

Es ergeben sich folgende Kriterien für die Farben der Skala (vgl. Abbildung 1 und Abbildung 2):

Sehr gute Vorhersagequalität (dunkelgrün)

Das Gütemaß zeigt einen Wert größer als 0,75. Diese Grenze ist so gewählt, dass die besten 10-15% der Vorhersagen der Vergangenheit aus dem Evaluierungszeitraum (2016-2023) dies erreichen.

=> Die roten Boxplots haben eine sehr gute Vorhersagequalität.

Gute Vorhersagequalität (hellgrün)

Das Gütemaß zeigt einen Wert signifikant größer als 0,4 und kleiner 0,75.

=> Die roten Boxplots sind den schwarzen Boxplots klar überlegen und zeigen eine gute Vorhersagequalität.

Moderate Vorhersagequalität (gelb)

Das Gütemaß zeigt einen Wert signifikant größer als 0 und kleiner 0,4.

=> Die roten Boxplots sind den schwarzen Boxplots immer noch überlegen, zeigen aber nur eine moderate Vorhersagequalität.

Geringe Vorhersagequalität (orange)

Das Gütemaß zeigt entweder einen Wert über 0, der aber nicht signifikant ist, oder einen ebenfalls nicht signifikanten Wert unter 0. Vorhersage und Referenzvorhersage bieten statistisch gesehen eine gleichgute Vorhersage.

=> Die roten und schwarzen Boxplots zeigen die gleiche Vorhersagequalität.

Schlechte Vorhersagequalität (rot)

Das Gütemaß zeigt einen signifikanten Wert unter 0. Die Referenzvorhersage als Wahrscheinlichkeitsverteilung ist der Vorhersage überlegen.

=> Die schwarzen Boxplots bieten statistisch gesehen die bessere Vorhersagequalität als die roten Boxen.

Überträgt man die Bedeutung der Güte-Skala von der Boxplot-Darstellung auf die Kuchendiagramme der 5-Klassen-Vorhersage, finden sich die roten Boxplots in den Vorhersagediagrammen wieder. Eine Gleichverteilung aller Kuchenstücke, wie sie in der Legende (Abbildung 5) zu sehen ist, ist hier das Pendant zu den schwarzen Boxplots. Die Darstellung ändert sich nicht über das Jahr, da sich die Grenzen der Abflussklassen ändern.

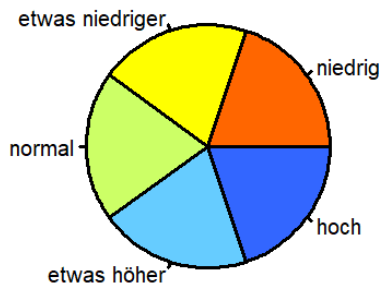


Abbildung 5: Legende der 5-Klassen-Vorhersage

Die Güte-Skala der Vorhersage der Unterschreitungswahrscheinlichkeit in der Web-Applikation hat eine Besonderheit. Hier ist es auch möglich, dass keine der fünf Stufen angezeigt wird. Das hier verwendete Gütemaß (BSS) prüft, wie gut die Unterschreitung des gewählten Schwellenwertes vorhergesagt wurde. Die Güte ist also abhängig vom Vorhersagemonat und vom Schwellenwert. Ist letzterer sehr hoch oder sehr niedrig gewählt, kann es sein, dass keine Güte im Evaluierungszeitraum berechnet werden konnte. Dies liegt dann vor, wenn in den für die Berechnung genutzten Daten entweder in der Referenzvorhersage oder in der Vorhersage keine Werte in diesem Bereich zu finden sind und somit auch keine Aussage zur Qualität getroffen werden kann.

Nutzungshinweise

Zur effektiven Nutzung der Güte-Skala ist folgendes zu beachten:

Die Skala kann im Einzelfall falsch liegen. Das bedeutet, dass auch bei einer grünen Anzeige im Einzelfall die Referenzvorhersage die bessere Vorhersage darstellt. Umgekehrt kann aber auch bei einer orangen oder roten Anzeige die aktuelle Vorhersage die bessere Wahl sein. Die Güteberechnungen sind immer statistische Abschätzungen, die eine Hilfe zu Einordnung auf Basis früherer Vorhersagen liefern.

Ebenso kann die Anzeige täuschen. Eine grüne Klasse bedeutet, dass die aktuelle Vorhersage besser ist, als die Referenzvorhersage. Das kann auch vorkommen, wenn beide Vorhersagen eine geringe absolute Güte haben, da es nur auf das Verhältnis der Beiden Vorhersagen zueinander ankommt.

Trotz aller Einschränkungen gilt:

Die Vorhersage verfügt im Jahresmittel über eine positive relative Vorhersagegüte über alle 6 Wochen. Um hier saisonale Unterschiede zu berücksichtigen, ist die Güte-Skala ein wichtiges Werkzeug, um die Vorhersage abhängig von der Jahreszeit noch effektiver nutzen zu können.

Kontakt:

Bundesanstalt für Gewässerkunde

Am Mainzer Tor 1

56068 Koblenz

Tel.: 0261/1306-0

<https://www.bafg.de/>

Vorhersageteam:

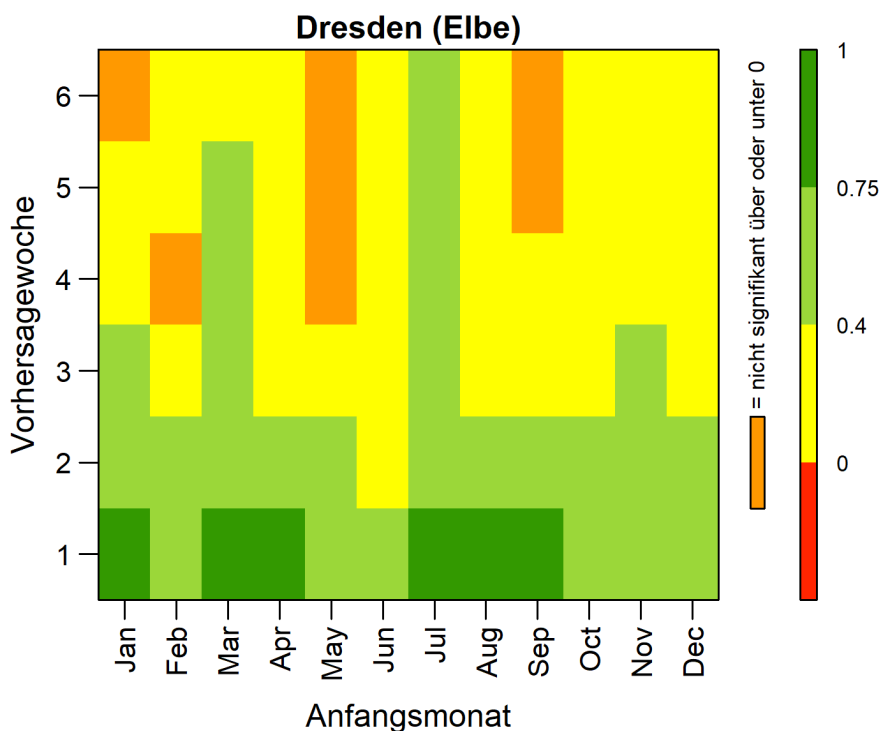
vorhersage@bafg.de

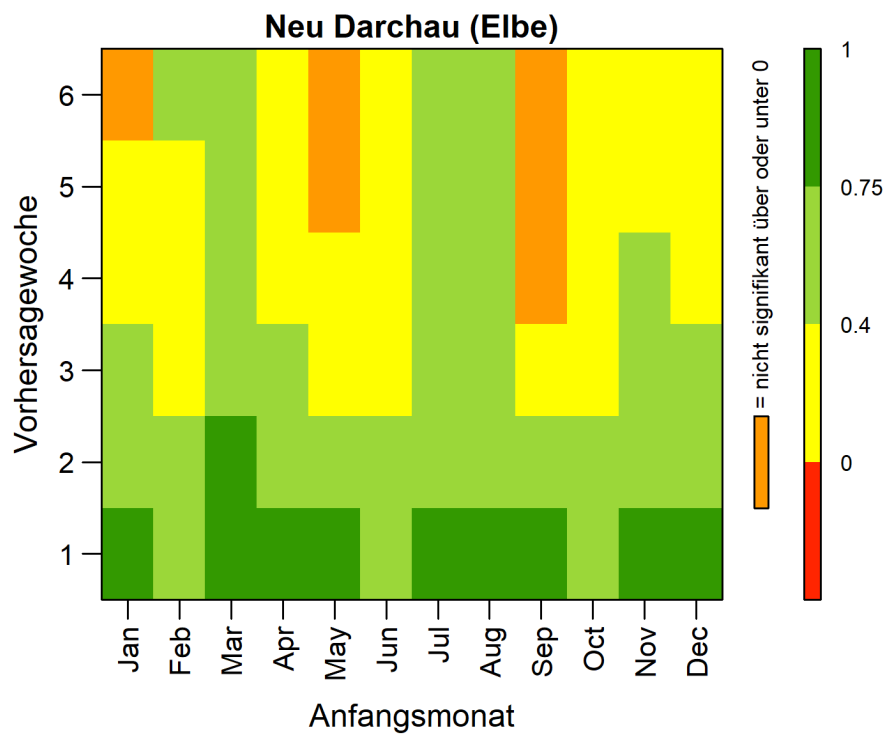
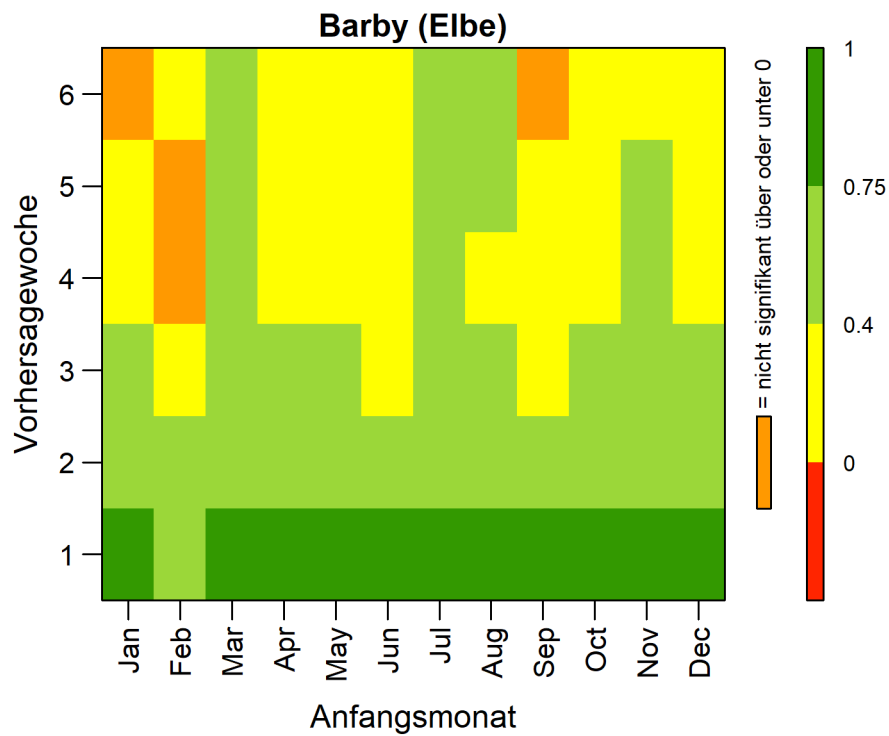
Koblenz, 11.11.2025

Anhang

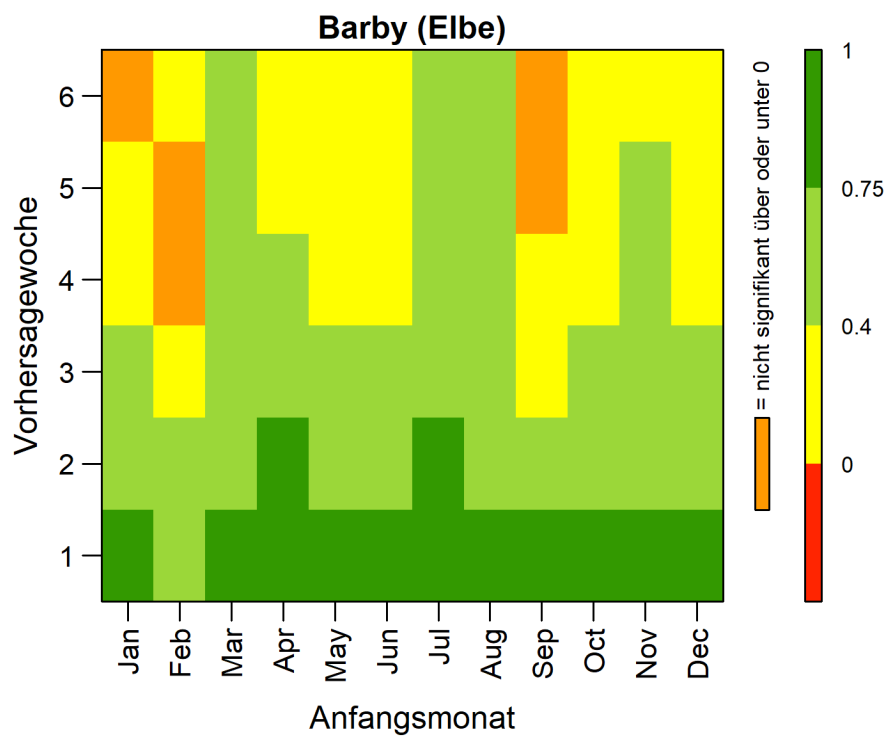
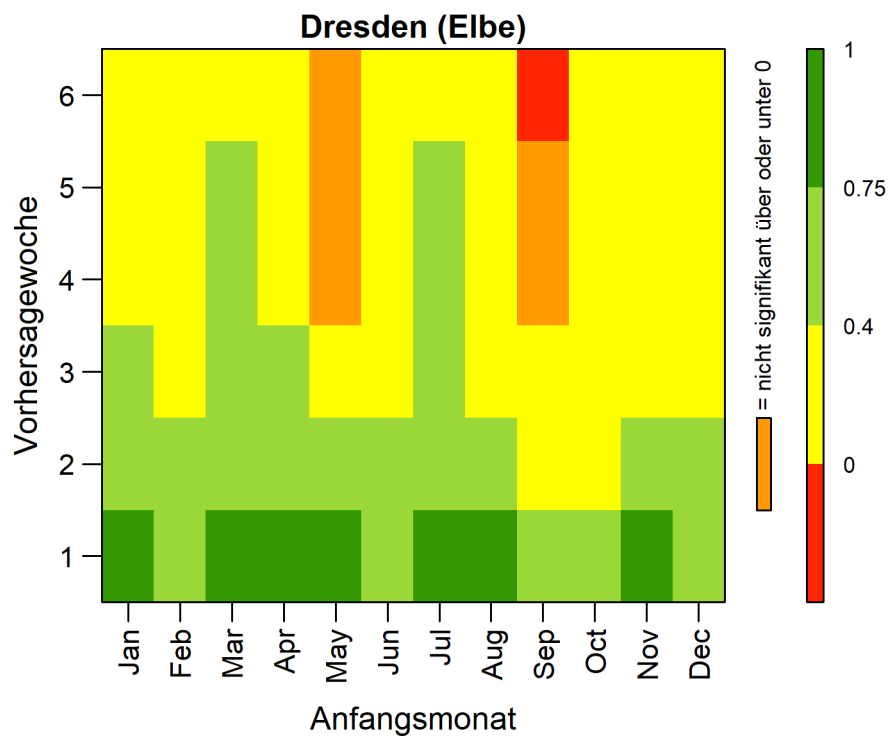
Die folgenden Abbildungen zeigen die Güteklassen für die Boxplot- und 5-Klassen-Vorhersage im Jahresverlauf abhängig von der Vorhersagewoche. Die Farben finden sich in der Legende in Abbildung 2 wieder. Ist ein Wert über Null, aber nicht signifikant, fällt dieser nicht, wie alle anderen Werte zwischen 0 und 0,4 in die gelbe, sondern in die orange Güteklasse. Ebenso verhält es sich mit Werten unter null, die nicht signifikant sind. Diese sind dann statt in der roten Güteklasse, ebenfalls in der orangen (vgl. Absatz „Hintergründe der Güte-Skala“, „geringe Vorhersagequalität (orange)“). Da es möglich ist, dass der gleiche Wert über oder unter null einmal signifikant und einmal nicht signifikant ist, ist in dieser Darstellung die orange Güteklasse nicht in die Skala der anderen Güteklassen eingereiht. Als Berechnungsgrundlage dienen die Vorhersagen der Vergangenheit von 2016-2023.

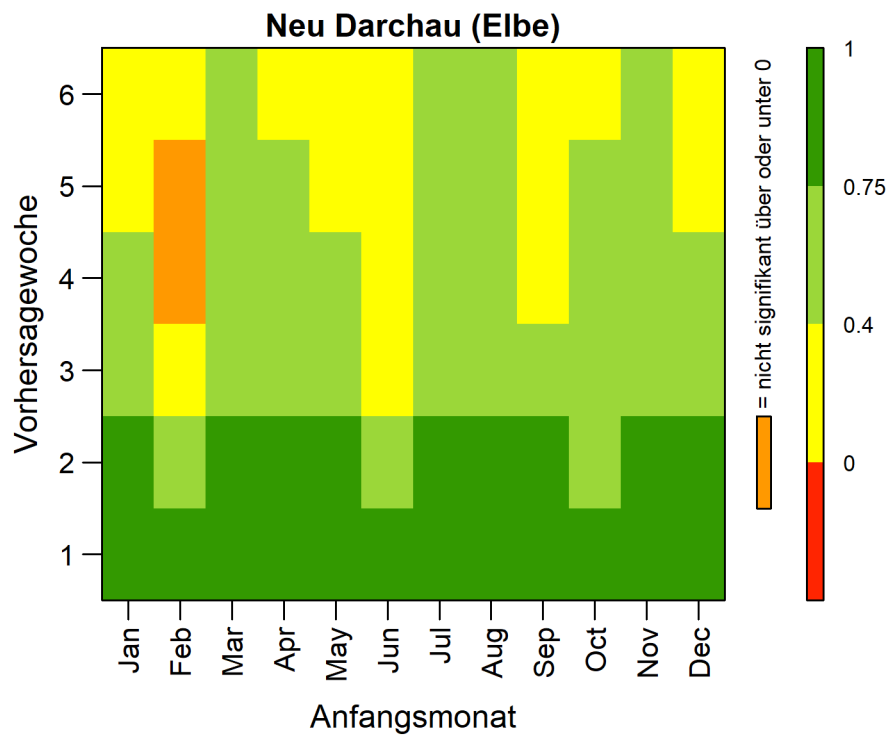
Güteklassen der Boxplots (CRPSS) an der Elbe:



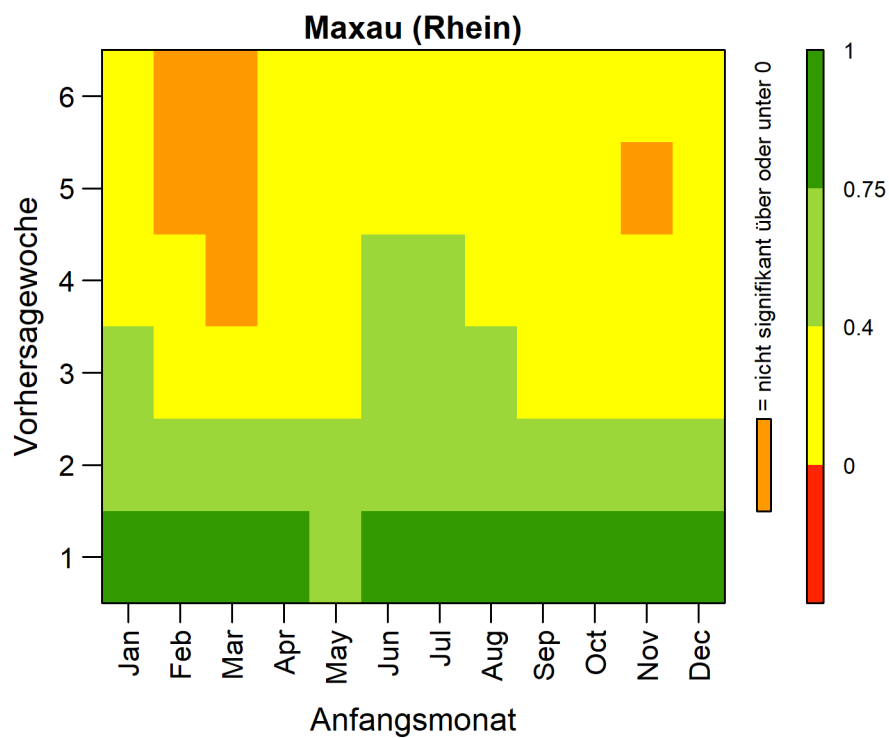


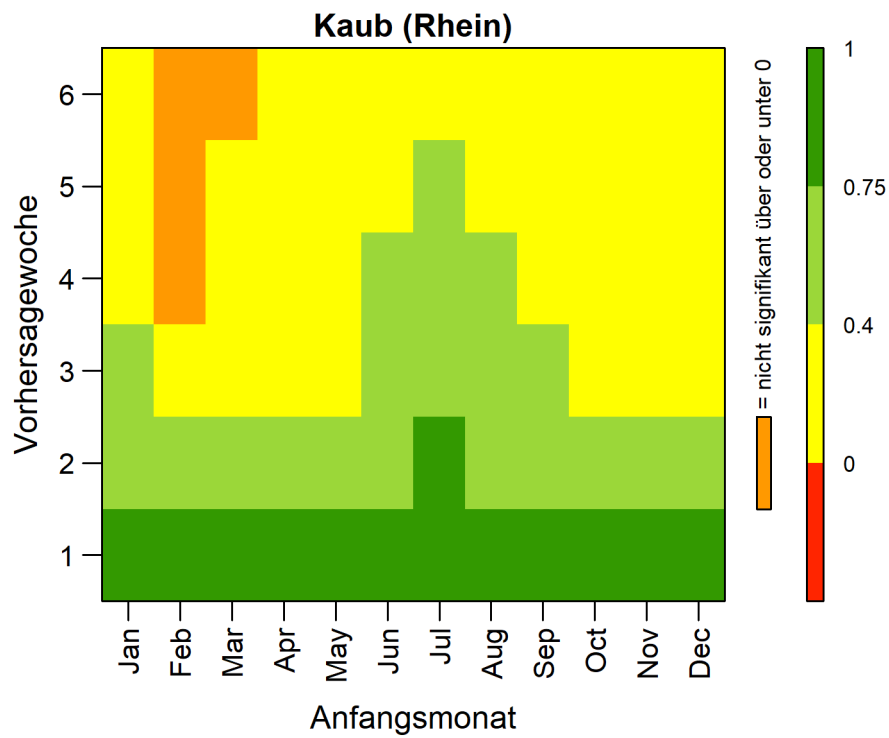
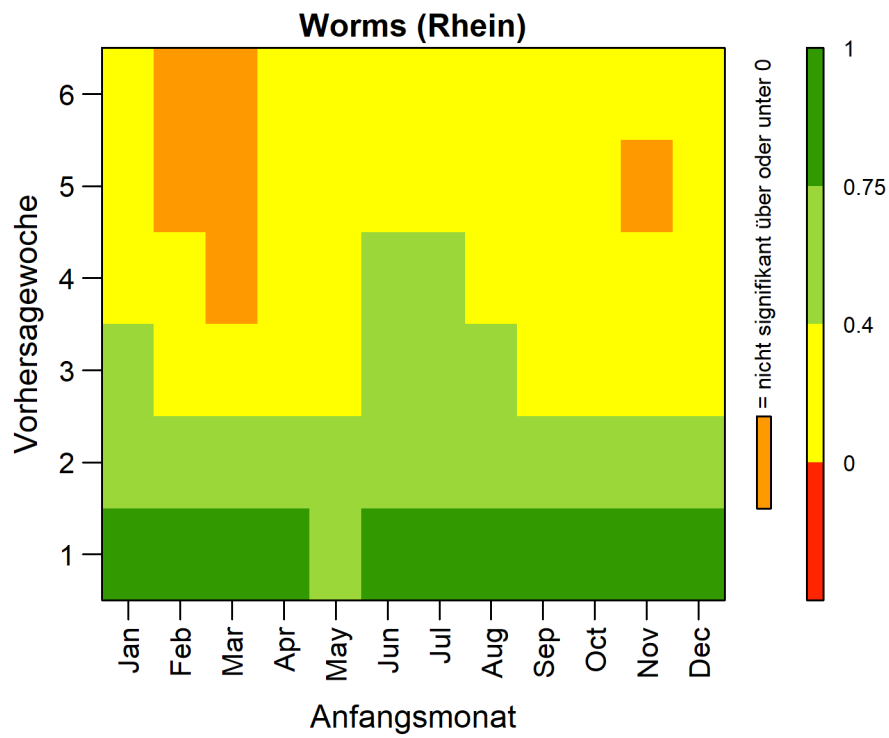
Güteklassen der 5-Klassen-Vorhersage (RPSS) an der Elbe:

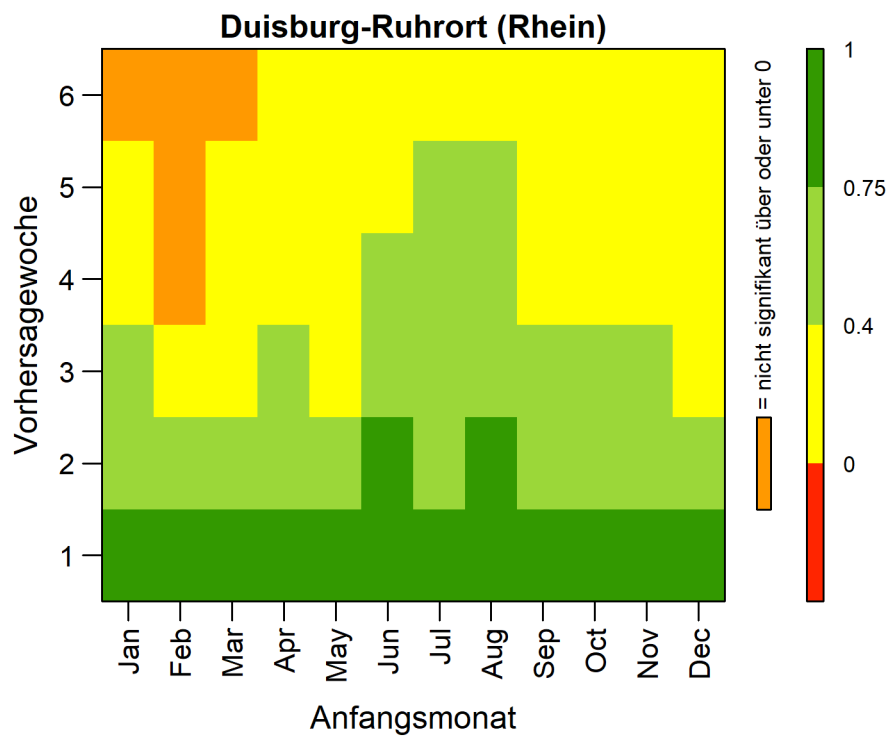
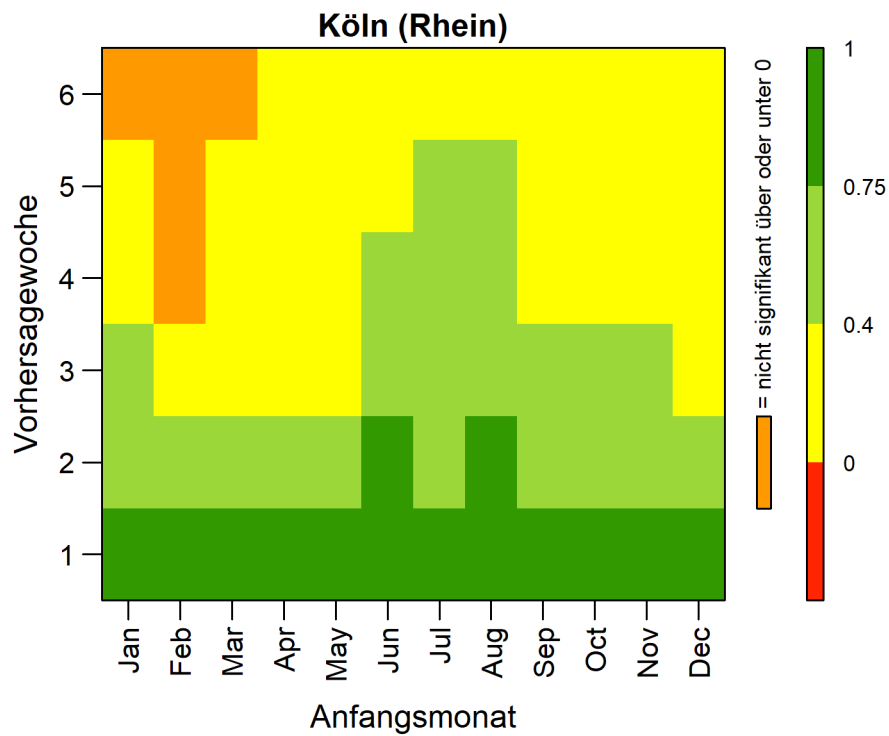




Güteklassen der Boxplots (CRPSS) am Rhein:







Güteklassen der 5-Klassen-Vorhersage (RPSS) am Rhein:

