

CLIMATE CHANGE

40/2023

Abschlussbericht

# Entwicklung eines bundesweiten Hantavirus- Prognosemodells

von:

Orestis Kazasidis, Jens Jacob  
Julius Kühn-Institut, Münster

Herausgeber:

Umweltbundesamt



CLIMATE CHANGE 40/2023

Ressortforschungsplan des Bundesministeriums für  
Umwelt, Naturschutz, nukleare Sicherheit und  
Verbraucherschutz

Forschungskennzahl 3720 48 401 0  
FB001170

Abschlussbericht

# **Entwicklung eines bundesweiten Hantavirus- Prognosemodells**

von

Orestis Kazasidis, Jens Jacob

Julius Kühn-Institut, Münster

Im Auftrag des Umweltbundesamtes

## Impressum

### Herausgeber

Umweltbundesamt  
Wörlitzer Platz 1  
06844 Dessau-Roßlau  
Tel: +49 340-2103-0  
Fax: +49 340-2103-2285  
[buergerservice@uba.de](mailto:buergerservice@uba.de)  
Internet: [www.umweltbundesamt.de](http://www.umweltbundesamt.de)

[f/umweltbundesamt.de](https://www.facebook.com/umweltbundesamt.de)

[t/umweltbundesamt](https://www.twitter.com/umweltbundesamt)

### Durchführung der Studie:

Julius Kühn-Institut (JKI) – Bundesforschungsinstitut für Kulturpflanzen, Institut für  
Epidemiologie und Pathogendiagnostik, Nagetierforschung  
Toppheideweg 88  
48161 Münster

### Abschlussdatum:

März 2023

### Redaktion:

Fachgebiet IV 1.4. Gesundheitsschädling und ihre Bekämpfung  
Anke Geduhn

Publikationen als pdf:

<http://www.umweltbundesamt.de/publikationen>

ISSN 1862-4359

Dessau-Roßlau, September 2023

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

**Kurzbeschreibung: Entwicklung eines bundesweiten Hantavirus-Prognosemodells**

Das durch Rötelmäuse (*Clethrionomys glareolus*, Syn. *Myodes glareolus*) auf den Menschen übertragene Puumala Orthohantavirus (PUUV) ist das häufigste Hantavirus in Europa und hochendemisch im Süden und Nordwesten Deutschlands. Seit 2001, als die Hantavirus-Erkrankung in Deutschland meldepflichtig wurde, haben die jährlichen humanen PUUV-Infektionen in sechs Jahren die Tausendergrenze überschritten. Die Ausbrüche der Infektionen korrelieren stark mit der Mast der Rotbuche (*Fagus sylvatica*) und den Wetterparametern der beiden vorangegangenen Jahre, die das Wachstum der Populationen der Nagetierwirte beeinflussen. Dies ermöglicht die Entwicklung eines auf Wetter- und Phänologie-Daten basierten Prognosemodells für das humane PUUV-Infektionsrisiko.

Durch eine Datentransformation wurde die Frage der Risikobewertung umformuliert. Anstatt der inzidenz- und infektionsbasierten Formulierungen, die in der Literatur üblich sind, wurde das Ausbruchsrisiko eingeführt, eine neue Größe, welche der Prognose von Infektionen und Risikoklassen vorangestellt ist. Diese Größe bestimmt die kreisbezogenen Ausbrüche, d. h., dass die historischen Inzidenzwerten anhand einer kreisspezifischen Schwelle in zwei Gruppen eingeteilt werden: Ausbrüche und nicht-Ausbrüche. Zur Unterscheidung der Ausbrüche wurde ein Support-Vector-Machine-Klassifikator mit einem linearen Kernel verwendet, ein Algorithmus mit hoher Leistung und geringer Modellkomplexität.

Das optimale Modell mit sechs Variablen bewies eine hohe Vorhersagekraft mit über 82% Genauigkeit im Zeitraum 2006–2021. Zwei automatische Programme können die erforderlichen Variablen erfassen und bearbeiten und schließlich die Prognose am Anfang Oktober für das Folgejahr fertigstellen. Für das Jahr 2022 hatte das Modell ein geringes Ausbruchsrisiko für alle Landkreise ermittelt, was zu 100% validiert wurde. Für das Jahr 2023 hat das Modell ein hohes Ausbruchsrisiko für 11 Landkreise in Niedersachsen und Nordrhein-Westfalen und für 41 Landkreise in Mittel- und Süddeutschland ermittelt.

**Abstract: Development of a Germany-wide prediction model for human hantavirus infections**

The Puumala Orthohantavirus (PUUV) is the most common hantavirus in Europe. It is transmitted by bank voles (*Clethrionomys glareolus*, syn. *Myodes glareolus*) and is highly endemic in Southern and Northwestern Germany. Since 2001, when hantavirus disease became notifiable in Germany, the annual human PUUV infections have exceeded one thousand cases in six years. The infection outbreaks correlate strongly with beech (*Fagus sylvatica*) mast and weather parameters from the two previous years, which drive the growth of rodent host populations. This allows the development of a prediction model for human PUUV infection risk based on weather and phenology data.

Applying a data transformation, the question about the risk assessment was reformulated. Instead of the incidence and infection-based formulations that are common in the literature, we introduced the outbreak risk, a new quantity that precedes the prediction of infections and risk classes. This quantity determines district-based outbreaks, i.e., based on a threshold unique for each district its past incidence values are split into two groups: outbreaks and non-outbreaks. A support vector machine classifier with a linear kernel was used to detect the outbreaks, an algorithm with high performance and low model complexity.

The optimal model with six variables demonstrated high predictive power with over 82% accuracy in the period 2006–2021. Two automatic programs can collect and process the required variables and finally generate the prediction in early October for the following year. For 2022, the model had predicted a low outbreak risk for all districts, which was validated with 100% accuracy. For 2023, the model predicted a high outbreak risk for 11 districts in Lower Saxony and North Rhine-Westphalia, and for 41 districts in Central and Southern Germany.

## Inhaltsverzeichnis

Abbildungsverzeichnis.....	8
Tabellenverzeichnis.....	9
Abkürzungsverzeichnis.....	10
Zusammenfassung.....	12
Summary.....	22
1 Grundlagen.....	31
1.1 Einleitung.....	31
1.2 Infektionen und Inzidenz.....	31
1.3 Leistungskennzahlen.....	31
2 Sammeln von benötigten Daten für die Modellierung.....	33
2.1 Rohdaten und Bearbeitung.....	33
2.2 Zusammenfassen von Stadt- und Landkreisen.....	36
2.3 Datensatz.....	37
3 Validierung und Vergleich der bisherigen Modellansätze.....	38
3.1 BW Modell.....	38
3.1.1 Validierung für 2021.....	40
3.2 JKI Modelle.....	41
3.2.1 Extrapolation auf Hessen und Thüringen.....	43
3.2.2 Extrapolation vom JKI Modell-Süd auf Norddeutschland.....	43
3.3 Diskussion und Fazit.....	44
3.3.1 BW Modell.....	44
3.3.2 JKI Modelle.....	44
3.3.3 Vergleich.....	45
3.3.4 Fazit.....	46
4 Entwicklung und Validierung eines neuen, bundesweiten Prognosemodells.....	47
4.1 Entwicklung.....	47
4.1.1 Auswahl der Kreise.....	47
4.1.2 Aufbereitung der Daten.....	48
4.1.2.1 Transformation der Inzidenz und Ausbruchsrisko.....	48
4.1.2.2 Prädiktoren.....	50
4.1.2.3 Risikoklassen.....	52
4.1.3 Klassifizierungsmethode und Inzidenz-Vorhersage.....	53
4.1.4 Feature Selection.....	54

4.1.4.1	Das Argument für die Blühstärke der Buche .....	58
4.2	Optimale Modelle .....	58
4.3	Fazit und Ausblick .....	61
5	Publikation der Prognoseergebnisse auf einer verstetigten Online-Plattform.....	64
5.1	Allgemein .....	64
5.2	Zusätzliche Eigenschaften .....	64
6	Prognose für Jahre außerhalb des Modellierungszeitraums .....	66
6.1	Für das Jahr 2022 .....	66
6.1.1	Prognose .....	66
6.1.2	Validierung .....	68
6.2	Für das Jahr 2023 .....	69
7	Quellenverzeichnis .....	73
A	Landkreise .....	77
A.1	Zusammengefasste Stadtkreise .....	77
A.2	Einbezogene Landkreise.....	77
B	Herkunftsgebiete.....	79
B.1	Herkunftsgebiete der Rotbuche.....	79

## Abbildungsverzeichnis

Abbildung 1:	Rasterung für die Bearbeitung der Wetterdaten. ....	34
Abbildung 2:	Der Laubwald, Landkreise und Herkunftsgebiete. ....	36
Abbildung 3:	Validierung für das BW Modell im Zeitraum 2014–2020. ....	39
Abbildung 4:	Validierung für das BW Modell in 2021. ....	41
Abbildung 5:	Validierung für die JKI Modelle im Zeitraum 2016–2020. ....	42
Abbildung 6:	Vergleich der zwei Modellansätze im Zeitraum 2016–2020. ....	45
Abbildung 7:	Auswahl der Landkreise für die Analyse. ....	48
Abbildung 8:	Histogramme der skalierten jährlichen PUUV-Inzidenz von 2006 bis 2021. ....	49
Abbildung 9:	Die jährlichen Inzidenzwerte in den einbezogenen Landkreisen von 2006 bis 2021. ....	50
Abbildung 10:	Karte der für diese Analyse relevanten Herkunftsgebiete (HKG) der Rotbuche. ....	51
Abbildung 11:	Choroplethenkarte des Laubwaldanteils im Jahr 2018 pro Landkreis. ....	52
Abbildung 12:	Histogramm der jährlichen PUUV-Inzidenz von 2006 bis 2021. ....	53
Abbildung 13:	Schematische Darstellung der Vorhersage-Methode, nach dem Training des SVM-Klassifikators. ....	54
Abbildung 14:	Pseudocode für die Auswertung der Vorhersagekraft. ....	54
Abbildung 15:	Die häufigsten Jahr-Monat-Kombinationen unter den besten Modellen mit zwei Prädiktoren. ....	56
Abbildung 16:	Die mittlere jährliche Genauigkeit für das Ausbruchsrisiko. ....	57
Abbildung 17:	Aussagekraft der beiden finalen Modelle. ....	59
Abbildung 18:	Die Koeffizienten des Support Vector Klassifikators für die beiden finalen Modelle. ....	60
Abbildung 19:	Prognose für 2022 anhand des Modells mit Blühstärke. ....	66
Abbildung 20:	Prognose für 2022 anhand des Modells ohne Blühstärke. ....	67
Abbildung 21:	Prognose für 2023 anhand des Modells mit Blühstärke. ....	69
Abbildung 22:	Prognose für 2023 anhand des Modells ohne Blühstärke. ....	70

## Tabellenverzeichnis

Tabelle 1:	Leistungskennzahlen für die Validierung des BW Modells im Zeitraum 2014–2020. ....	40
Tabelle 2:	Leistungskennzahlen für die Validierung der JKI Modelle im Zeitraum 2016–2020. ....	42
Tabelle 3:	Die im Modell enthaltenen zusammengefassten Stadtkreise..	77
Tabelle 4:	Die für diese Analyse einbezogenen Landkreise. ....	77
Tabelle 5:	Die Herkunftsgebiete (HKG), die die einbezogenen Landkreise enthalten. ....	79

## Abkürzungsverzeichnis

<b>Abkürzung</b>	<b>Bedeutung bzw. Erklärung</b>
<b>a.u.</b>	Willkürliche Einheit (aus dem Englischen “ <u>a</u> rbitrary <u>u</u> nit”)
<b>BF</b>	Laubwald (aus dem Englischen “ <u>B</u> road- <u>l</u> eaved <u>F</u> orest”)
<b>BW</b>	Baden-Württemberg
<b>BW Modell</b>	das Hantavirus-Prognosemodell für Baden-Württemberg – FKZ 01 KI 1101; Monitoring sylvatischer Zoonosen, Eprimos, 2014
<b>BY</b>	Bayern
<b>bzw.</b>	beziehungsweise
<b>CART</b>	<u>C</u> lassification and <u>R</u> egression <u>T</u> ree
<b>CDC</b>	<u>C</u> limate <u>D</u> ata <u>C</u> enter
<b>CLC</b>	<u>C</u> ORINE <u>L</u> and <u>C</u> over
<b>d. h.</b>	das heißt
<b>DWD</b>	Deutscher Wetterdienst
<b>FI</b>	Blühstärke der Rotbuche (aus dem Englischen “ <u>F</u> lowering <u>I</u> ntensity”)
<b>FLI</b>	Friedrich-Loeffler-Institut
<b>ggb.</b>	gegenüber
<b>HE</b>	Hessen
<b>HKG</b>	Herkunftsgebiet(e) der Rotbuche ( <i>Fagus sylvatica</i> )
<b>JKI</b>	Julius Kühn-Institut, Bundesforschungsinstitut für Kulturpflanzen
<b>JKI Modell-Nord</b>	das Puumalavirus-Prognosemodell für Landkreise in Norddeutschland (NI und NW) – FKZ 3713 48 401, JKI und FLI, 2018
<b>JKI Modell-Süd</b>	das Puumalavirus-Prognosemodell für Landkreise in Süddeutschland (BW und BY) – FKZ 3713 48 401, JKI und FLI, 2018
<b>LK</b>	Landkreis
<b>LOOCV</b>	Leave-One-Out-Kreuzvalidierung (aus dem Englischen “ <u>L</u> east- <u>O</u> ne- <u>O</u> ut <u>C</u> ross- <u>V</u> alidation”)
<b>NI</b>	Niedersachsen
<b>NW</b>	Nordrhein-Westfalen
<b>Pr</b>	Monatssumme der Niederschlagshöhe in mm (aus dem Englischen “ <u>P</u> recipitation”)
<b>PUUV</b>	Puumalavirus
<b>RKI</b>	Robert Koch-Institut
<b>RP</b>	Rheinland-Pfalz
<b>SD</b>	Monatssumme der Sonnenscheindauer in Stunden
<b>SK</b>	kreisfreie Städte bzw. Stadtkreis
<b>SM</b>	mittlere Bodenfeuchte unter Gras und sandigem Lehm in Prozent pflanzenverfügbares Wasser (% nFK)
<b>ST</b>	mittlere Bodentemperatur in 5cm Tiefe bei unbewachsenem Boden in °C (aus dem Englischen “ <u>S</u> oil <u>t</u> emperature”)
<b>SVC</b>	Klassifikator basierend auf SVM (aus dem Englischen “ <u>S</u> upport <u>V</u> ector <u>C</u> lassifier”)
<b>SVM</b>	Support Vector Machines
<b>TH</b>	Thüringen
<b>Tmax</b>	Monatsmittel der Lufttemperaturmaxima in 2m Höhe in °C
<b>Tmean</b>	Monatsmittel der Lufttemperatur in 2m Höhe in °C
<b>Tmin</b>	Monatsmittel der Lufttemperaturminima in 2m Höhe in °C
<b>V0</b>	Parameter aus dem Prognosejahr oder aktuellen Jahr
<b>V1</b>	Parameter aus dem Vorjahr

<b>Abkürzung</b>	<b>Bedeutung bzw. Erklärung</b>
<b>V2</b>	Parameter aus dem Vorvorjahr
<b>z. B.</b>	zum Beispiel
<b>zzgl.</b>	zuzüglich

Die Abkürzungen basieren auf englischen Begriffen, um mit den englischen projektbezogenen Veröffentlichungen einheitlich zu sein.

## Zusammenfassung

### Grundlagen

Das Puumala Orthohantavirus (PUUV) ist das häufigste Hantavirus in Europa, wird durch Rötelmäuse (*Clethrionomys glareolus*, Syn. *Myodes glareolus*) übertragen und ist hochendemisch im Süden und Nordwesten Deutschlands. Es kann leichtes bis mäßiges hämorrhagisches Fieber mit renalem Syndrom (Nephropathia epidemica) verursachen. Nach Angaben des Robert Koch-Instituts schwanken die jährlichen humanen PUUV-Infektionen mehrjährig und haben in sechs Jahren seit 2001 die Tausendergrenze überschritten.

Die PUUV-Ausbrüche sind von der Mast der Rotbuche (*Fagus sylvatica*) getrieben. Die zugrundeliegenden Mechanismen für die Übertragungsrate des PUUV auf den Menschen scheinen zu komplex zu sein, um sie direkt zu modellieren, da sie von der Abundanz des Nagetierreservoirs, ihrer PUUV-Prävalenz und der Interaktion zwischen Mensch und Rötelmaus abhängen, welche alle zeitlich schwanken und lokal variieren. Dennoch können Wetterparameter als Prädiktoren für das humane PUUV-Infektionsrisiko verwendet werden, da die Schwankungen der Rötelmauspopulationen stark mit den Wetterparametern der beiden vorangegangenen Jahre korrelieren, während die PUUV-Prävalenz hauptsächlich von der Abundanz von Rötelmäusen abhängt.

Als PUUV-Infektionen werden die gemeldeten humanen Hantavirus-Fälle definiert, die labordiagnostisch vom Puumalavirus verursacht wurden. Die jährliche Inzidenz (Jahresinzidenz) wird als die Anzahl der jährlichen Infektionen pro 100.000 Einwohner definiert.

### Leistungskennzahlen

Die Güte der Vorhersagen wird unter anderem anhand der folgenden Maße bewertet.

- ▶ Die Genauigkeit ist der prozentuale Anteil der richtig vorhergesagten Risikoklassen an den gesamten Vorhersagen eines bestimmten Datensatzes. Diese Größe soll maximiert werden (Maximum=100%, Minimum=0%).
- ▶ Der Pearson-Korrelationskoeffizient wird auf die vorhergesagten und die echten Infektionen bzw. Inzidenzwerte aller Jahre in den Landkreisen angewendet. Das Bestimmtheitsmaß ( $R^2$ ) kann als das Quadrat vom Pearson-Korrelationskoeffizienten angenähert werden.

### Sammeln von benötigten Daten für die Modellierung

Folgende Datensätze wurden gesammelt bzw. digitalisiert:

- ▶ Human-Erkrankungen durch das Puumalavirus, Quelle: SurvStat@RKI 2.0, Zeitraum: 2001–2022, zeitliche Auflösung: wöchentlich bzw. jährlich, räumliche Auflösung: auf Kreisebene.
- ▶ Bevölkerung, Quelle: Eurostat, Zeitraum: 2001–2021, zeitliche Auflösung: jährlich, räumliche Auflösung: auf Kreisebene.
- ▶ Wetterdaten, Quelle: Climate Data Center, Zeitraum: 2000–2022, zeitliche Auflösung: monatlich, räumliche Auflösung: 1km x 1km Raster.

Datensätze: Monatsmittel der Lufttemperatur in °C, Lufttemperaturminima in °C, Lufttemperaturmaxima in °C, Monatssumme der Niederschlagshöhe in mm, Monatssumme der

Sonnenscheindauer in Stunden, mittlere Bodentemperatur in 5cm Tiefe in °C, mittlere Bodenfeuchte in Prozent pflanzenverfügbares Wasser.

Die Raster für jeden dieser Datensätze wurden mithilfe eines Geodaten-Vektorlayers der Kreisgrenzen Deutschlands bearbeitet und die Durchschnittswerte des entsprechenden Parameters auf Kreisebene berechnet.

- ▶ Landbedeckungsdaten, Quelle: CORINE Land Cover, Zeitraum: einzelne Jahre (2000, 2006, 2012 und 2018), zeitliche Auflösung: jährlich, räumliche Auflösung: 100m x 100m Raster.

Aus den Rasterdaten wurde der Laubwaldanteil pro Kreis berechnet. Die Menge des Laubwaldes wurde anhand der Klasse 3.1.1 ("broad-leaved forest") abgeschätzt. Für den absoluten Laubwaldanteil haben wir den Referenzdatensatz CLC2018 verwendet. Für die zeitliche Entwicklung des Laubwaldanteils haben wir die Veränderungen auf Basis der "CLC-Change layers" (CHA - Change mapping) Datensätze für 2000-2006, 2006-2012 und 2012-2018 bewertet.

- ▶ Blühstärke der Rotbuche, Quelle: Dagmar Schneck (AFZ-DerWald und persönliche Mitteilung), Zeitraum: 2000–2022, zeitliche Auflösung: jährlich, räumliche Auflösung: auf der Ebene der Herkunftsgebiete der Rotbuche.

Deutschland ist in 26 Herkunftsgebiete (HKG) für die Rotbuche unterteilt. Aus jedem HKG werden charakteristische Baumbestände ausgewählt, deren Blühintensität einmal im Jahr in den ersten Juniwochen bewertet wird. Die Einzelbewertungen werden über jedes HKG gemittelt. Um die räumliche Auflösung zu erhöhen, rechneten wir die Blühstärke auf Kreisebene um, indem die in jedem Herkunftsgebiet enthaltenen Laubwaldanteile des Kreises berechnet und dann als Gewichte für das gewichtete Mittel der Blühstärke angewandt wurden.

- ▶ Zusätzliche Parameter für die Validierung der bisherigen Modellansätze.

Mittlerer Wert der Wetterparameter in Baden-Württemberg: Dieser wurde, wie vorher für die Berechnung der Kreiswerte beschrieben, ermittelt, wobei die Maske auf das gesamte Bundesland bezogen wurde.

Buchenmast sowie Anteile der Buche und der sonstigen Laubbäume auf Kreisebene in Baden-Württemberg: gesammelt von dem auf der Internetseite [Hanta-Vorhersage](#) implementierten Vorhersagemodell.

- ▶ Vektordaten für die Kreisgrenzen Deutschlands, Quelle: Kreisgrenzen 2017 (Bundesamt für Kartographie und Geodäsie).

### **Validierung und Vergleich der bisherigen Modellansätze**

Zum Zeitpunkt des Projektstarts (November 2020) waren zwei PUUV-Prognosemodelle bzw. Modellierungsansätze für verschiedene Regionen in Deutschland vorhanden. Beide Modellierungsansätze wurden in Python nachgebildet. Mit den selbst erhobenen Datensätzen wurden Vorhersagen retrospektiv für den Zeitraum 2006–2020 erstellt und mit den echten Human-Inzidenzwerten oder Risikoklassen verglichen. Der Vergleich wurde außerhalb des Modellierungszeitraums jedes Modells durchgeführt.

## **BW Modell**

Das generalisierte lineare Modell für Baden-Württemberg (FKZ 01 KI 1101; Monitoring sylvatischer Zoonosen, Eprimos, 2014) wurde als BW Modell abgekürzt. Der Modellierungszeitraum war von 2005 bis 2013. Wir bildeten das Modell nach, das auf der Internetseite [Hanta-Vorhersage](#) implementiert wurde. Die erforderlichen Eingaben waren: auf Kreisebene der Anteil der Buche, der Anteil der sonstigen Laubbäume und die Buchenmast, und auf Bundeslandebene Wetterparameter aus dem Vorjahr: die mittlere Temperatur im September und die Sonnenscheindauer im August und September.

Unsere Validierung im Zeitraum 2014–2020 zeigte eine Tendenz zur Überschätzung der Infektionen, weil die Infektionszahlen für die Ausbruchsjahre im Modellierungszeitraum deutlich höher als im Validierungszeitraum waren. Die größte Abweichung der Inzidenz erschien als Überschätzung im Jahr 2015. Nur in einigen Einzelfällen gab es Unterschätzung der Risikoklasse. Die Ausbruchsjahre (2017 und 2019) waren im Allgemeinen richtig erkannt. Die gesamte Genauigkeit im Validierungszeitraum war 57,6%. Der höchste Wert der jährlichen Genauigkeit war 79% für 2014 und der geringste Wert war 9% für 2015.

Im Jahr 2021 wich das Vorhersagemodell auf der Internetseite [Hanta-Vorhersage](#) vom ursprünglichen Modell ab. Wir haben die gemeldeten Erwartungswerte direkt von der Webseite verwendet, ohne unsere eigene Nachbildung anzuwenden. Das Modell zeigte für dieses Jahr eine hohe Vorhersagekraft mit 76% Genauigkeit für die Risikoklassen (28 richtige Vorhersagen, mit 5 Überschätzungen und 4 Unterschätzungen).

Die Validierung unserer Nachbildung des BW Modells konnte nur so gut erfolgen, wie nah unsere Datensammlung an die Originaldatensätze heranreichte. Eine endgültige Aussage für die Qualität und die Güte der BW Modelle ist deshalb mit Vorsicht zu betrachten. Weder für 2022 noch für 2023 lag auf der Internetseite [Hanta-Vorhersage](#) eine Prognose vor.

## **JKI Modelle**

Der auf CART basierende Modellierungsansatz resultierte in getrennten Modellen für Landkreise in Süd- bzw. Norddeutschland (FKZ 3713 48 401, JKI und FLI, 2018), die als JKI Modell-Süd und JKI Modell-Nord abgekürzt wurden. Beide Modelle ermitteln direkt Risikoklassen: das erste für Baden-Württemberg und Bayern; und das zweite für Niedersachsen und Nordrhein-Westfalen. Der Modellierungszeitraum war 2006–2015 für beide Modelle, wobei die Hälfte der Datenpunkte zufällig als Testdatensatz genutzt wurde. Das JKI Modell-Süd benötigt die maximale Temperatur im August aus dem Vorvorjahr, die Sonnenscheindauer im Mai aus dem Vorjahr und im März aus dem Prognosejahr und den Laubwaldanteil. Das JKI Modell-Nord benötigt die maximale Temperatur im Oktober aus dem Vorvorjahr, den Niederschlag im Januar aus dem Vorjahr und die Sonnenscheindauer und die maximale Temperatur im April aus dem Vorjahr. Alle erforderlichen Eingaben werden auf Kreisebene benötigt.

Beide JKI Modelle wurden im Rahmen dieser Studie im Zeitraum 2016–2020 validiert. Die gesamte Genauigkeit war 47,5% für das JKI Modell-Süd und 55,5% für das JKI Modell-Nord. Die Genauigkeit beider Modelle für alle vier Bundesländer war besser für nicht-Ausbruchsjahre (2016, 2018 und 2020) als für Ausbruchsjahre (2017 und 2019). Das JKI Modell-Süd folgte der allgemeinen Tendenz der echten Inzidenz für alle Jahre richtig. Es bewies eine gute Vorhersagekraft und detektierte beide Ausbruchsjahre im Validierungszeitraum, auch wenn seine regionalen Vorhersagen nicht immer genau waren. Andererseits erklärte das JKI Modell-Nord keine Varianz der Inzidenz in Norddeutschland, was auf eine Überanpassung im Trainingsdatensatz hinweist. Folglich ist das JKI Modell-Nord ungeeignet für Vorhersagen. Seine Genauigkeit erklärt sich durch die geringen Inzidenzwerte im Norden.

Einige Vergleiche, die an den Knoten der Entscheidungsbäume stattfinden, waren unzureichend. Das ist zu erwarten, weil es unklar ist, wie angemessen die Entscheidungsbäume im Validierungszeitraum sind, obwohl sie die Ergebnisse des Modellierungszeitraums anschaulich wiedergeben.

### **Vergleich und Fazit**

Die Vorhersagen des BW Modells konnten mit den Vorhersagen des JKI Modells-Süd für Baden-Württemberg verglichen werden. Im überlappenden Zeitraum 2016–2020 war die Leistung des BW Modells bei jeder Metrik besser (zum Beispiel, war die Genauigkeit der Risikoklassen 63% ggb. 46%). Das JKI Modell-Süd war zuzüglich der Daten aus Bayern zustande gekommen und es wurde erwartet, dass es eine geringere Leistung im Vergleich zu einem BW-spezifischen Modell aufweisen würde.

Aufgrund der Ungenauigkeit der JKI Modelle im Validierungszeitraum und der extremen Spezialisierung des BW Modells auf ein einzelnes Bundesland war es nötig, ein neues Vorhersagemodell zu generieren, das über alle PUUV-hochendemische Gebiete Deutschlands einheitlich ist. Die Grundlage basiert auf Erkenntnissen der Validierungen der beiden bisherigen Modellansätze. Das BW Modell enthielt einen kreisbezogenen Skalierungsfaktor, der Unterschiede in der lokalen Größenordnung der PUUV-Inzidenz berücksichtigte. Unsere Methode baut auf diesem Gedanken auf und erweitert ihn, indem sie die Berechnung der Skalierungsfaktoren vom Vorhersagemodell selbst entkoppelt und sie mit der in den JKI Modellen angewandten Inzidenztransformation kombiniert. Andererseits verwendeten die JKI Modelle Wetterdaten mit höherer räumlicher Auflösung und intrinsisch definierte Schwellenwerte für die Risikoklassen. Schließlich demonstrierten die für die JKI Modelle verwendeten Entscheidungsbäume die Vorteile einer schematischen Darstellung des Modell-Targets und inspirierten unsere Implementierung eines geometrisch motivierten Klassifikationsalgorithmus.

### **Entwicklung und Validierung eines neuen, bundesweiten Prognosemodells**

Die Analyse wurde auf PUUV-Infektionen seit 2006 beschränkt, um Verzerrungen durch Untererfassung zu vermeiden. Jeder Datenpunkt des Datensatzes repräsentiert einen bestimmten Landkreis in einem bestimmten Jahr. Die Parameter des Datenpunktes sind die Prädiktorenwerte in diesem Landkreis in diesem Jahr. Die Labels des Datenpunktes sind die PUUV-Infektionen, -Inzidenz und -Risikoklasse, sowie das PUUV-Ausbruchsrisko, eine eigene Größe dieser Studie.

### **Zusammenfassen von Stadt- und Landkreisen**

Um Unterschiede zwischen dem gemeldeten Wohnort und dem Infektionsort teilweise auszugleichen, haben wir die meisten kreisfreien Städte (Stadtkreise) mit ihrem umfassenden oder umgebenden Landkreis zusammengefasst. Die einzigen Stadtkreise, die getrennt blieben, waren SK Köln, SK Münster und SK Stuttgart, die deutlich großflächig sind.

Die Datenpunkte für die kombinierten Kreise behielten die Bezeichnung des Landkreises. Für die Infektions- und Bevölkerungszahlen wurden die Einzelwerte aggregiert und die Inzidenz neu berechnet. Für die Wetterparameter haben wir die Mittelwerte, gewichtet nach der Fläche des jeweiligen Kreises, diesem zugeordnet. Für den Laubwaldanteil und die Blühstärke der Rotbuche haben wir die Werte auf der Basis des aggregierten Laubwaldes bzw. Laubwaldes pro HKG neu errechnet.

## Auswahl der Kreise

Um die Landkreise auszuwählen, in denen das PUUV tatsächlich endemisch ist, und Einzelfälle aus nicht endemischen Regionen auszuschließen, haben wir in die Analyse nur Landkreise einbezogen, in denen die Gesamtinfektionen  $\geq 20$  und die höchste Jahresinzidenz  $\geq 2$  im Zeitraum 2006–2021 waren. Insgesamt wurden 66 Landkreise anhand dieser Kriterien ausgewählt: 26 Landkreise aus Baden-Württemberg (BW), 16 aus Bayern (BY), 8 aus Hessen (HE), 3 aus Niedersachsen (NI), 10 aus Nordrhein-Westfalen (NW), 1 aus Rheinland-Pfalz (RP) und 2 aus Thüringen (TH). Auf die ausgewählten Kreise entfielen 10.090 humane PUUV-Infektionen, 89,9% der landesweit gemeldeten Gesamtinfektionen in diesem Zeitraum.

## Aufbereitung der Daten

### *Transformation der Inzidenz und Ausbruchsrisiko*

Wir haben zunächst eine Log-Transformation auf die Inzidenzwerte angewendet, um den Einfluss von Nicht-Null-Werten zu erhöhen. Um die Auswirkungen zu berücksichtigen, die das Auftreten einer hohen kreisbezogenen Inzidenz verursachen, haben wir die Inzidenz auf Kreisebene mit zwei Klassen (Bins) diskretisiert, d. h. binarisiert. Für die Binarisierung haben wir eine nicht-überwachte Clusteranalyse der Log-Inzidenz durchgeführt, getrennt für jeden Kreis. Diese Transformation kompensiert die hohe räumliche Inhomogenität der Inzidenz und wandelte ihre Vorhersage in ein Extrema-Detektionsproblem um.

Die binäre Klassifikation, die wir „Ausbruchsrisiko“ nennen, führte zu zwei Klassen, die als „geringes Risiko“ (0) und „hohes Risiko“ (1) bezeichnet wurden. Ein lokaler „Ausbruch“ trat dann auf, wenn in einem Jahr die Inzidenz eines Kreises in die Hochinzidenzklasse der gemeldeten Werte eingeordnet wurde. Die Inzidenz in Kreisen in Jahren ohne Ausbruch war null oder wurde im Vergleich zu den gemeldeten Werten in diesem speziellen Kreis als niedrig betrachtet. Von den insgesamt 1.056 Datenpunkten (16 Jahre  $\times$  66 Kreise) wurden 682 als geringes Risiko (65%) und 374 als hohes Risiko (35%) eingestuft.

### *Risikoklassen*

Wir haben drei Risikoklassen ausgewählt, die durch leichte Abfälle im Histogramm der Inzidenzwerte definiert wurden. Der erste Schwellenwert von 2 Infektionen pro 100.000 Einwohner begrenzt die Klasse mit niedrigem Risiko gegenüber der Klasse mit mittlerem Risiko. Der zweite Schwellenwert, der die Klassen mittleren und hohen Risikos trennt, wurde auf 9,5 Infektionen pro 100.000 Einwohner festgelegt.

## Klassifizierungsmethode und Inzidenz-Vorhersage

Wir haben uns nur auf diejenigen Variablenkombinationen konzentriert, die zu einer linearen Entscheidungsgrenze (Hyperebene) für die Einstufung unseres gewählten Targets („Ausbruchsrisiko“) führten. Zur Unterscheidung der Ausbrüche verwendeten wir einen Support-Vector-Machine-Klassifikator (SVC) mit einem linearen Kernel, da er eine hohe Leistung mit einer geringen Modellkomplexität kombiniert, indem er die Entscheidungsgrenze als lineare Gleichung der Eingaben zurückgibt.

Die Ausgabe vom SVC stellt das erwartete Ausbruchsrisiko in jedem Landkreis dar: eine 0 (kein lokaler Ausbruch erwartet) oder eine 1 (lokaler Ausbruch erwartet). Sie kann in die Inzidenz umgerechnet werden, indem der Inzidenzmittelwert aus allen Datenpunkten des jeweiligen Kreises im Trainingsdatensatz zugewiesen wird, die das gleiche Ausbruchsrisiko anzeigten, entweder 0 oder 1. Anschließend ermittelt eine einfache Klassifizierung der vorhergesagten Inzidenz anhand der vordefinierten Schwellenwerte die vorhergesagte Risikoklasse.

## Prädiktoren und Feature Selection

Die Leistung wurde anhand zweier Verfahren bewertet:

- ▶ die Aussagekraft bewertete die Modellleistung an dem Trainingsdatensatz, und
- ▶ die Vorhersagekraft schätzte die Methodenleistung an einem Testdatensatz.

Die Auswertung der Vorhersagekraft erfolgte durch eine Leave-One-Out-Kreuzvalidierung (LOOCV). Ein einzelnes Jahr wurde aus dem Trainingsdatensatz (2006–2021) entfernt, das vollständige Modell wurde anhand der verbleibenden Jahre trainiert und dieses trainierte Modell wurde verwendet, um die Inzidenzwerte für das entfernte Jahr vorherzusagen. Dies wurde für alle sechzehn Jahre im Trainingsdatensatz wiederholt. Abschließend wurde die Prognose für alle Jahre anhand der Genauigkeit des Ausbruchsrisikos ausgewertet.

Aus jedem monatlichen Wetterparameter wurden 24 Einträge für alle Monate der beiden vorangegangenen Jahre als mögliche Prädiktoren für das Modell erstellt. Insgesamt gab es 168 verfügbare Prädiktoren aus den Wetterparametern (7 Parameter x 24 Monate), im Folgenden „Variablen“ genannt.

Weitere mögliche Prädiktoren waren die Blühstärke-Werte der Rotbuche aus dem Vorjahr und dem Vorvorjahr. Bucheckern sind eine Hauptnahrungsquelle für Rötelmäuse, und ihre Verfügbarkeit wurde mit dem Wachstum der Rötelmauspopulationen in Mitteleuropa in Verbindung gebracht. Es wird erwartet, dass die Einbeziehung der Mast die Leistung unseres Prognosemodells verbessert, indem Effekte einbezogen werden, die nicht direkt vom Wetter angetrieben sind. Die Blühstärke im späten Frühjahr und Frühsommer ist ein guter Indikator für die Mast, obwohl eine erhöhte Blüte nicht immer zu einer erhöhten Samenproduktion führt.

Der letzte mögliche Prädiktor in unserem Datensatz war der Laubwaldanteil, der in der Vergangenheit als Prädiktor für das PUUV-Infektionsrisiko in Süddeutschland verwendet wurde.

Die Einschränkung der Anzahl von Prädiktoren für unser endgültiges Modell verringert die Gefahr einer Überanpassung. Daher haben wir ein heuristisches Verfahren zum Durchsuchen des mehrdimensionalen Variablenraums etabliert.

Wir wandten zuerst einen SVC mit einem linearen Kernel für alle Teilmengen mit einer einzigen Variablen an. Die Leistung des Modells mit der Blühstärke aus dem Vorjahr war besser als die des Modells mit der Blühstärke aus dem Vorvorjahr (Genauigkeit 79,6% ggb. 73,2%). Da die beiden Variablen eine starke negative Korrelation aufwiesen, wurde die Blühstärke aus dem Vorvorjahr aus unserem Pool von Prädiktoren entfernt und nur die Blühstärke aus dem Vorjahr beibehalten.

Das Modell, welches nur den Laubwaldanteil berücksichtigte, hatte keine Aussagekraft, weil es immer ein geringes Ausbruchsrisiko ermittelte. Darüber hinaus erhöhte die Einbeziehung des Laubwaldanteils in die Modelle mit einzelnen Wettervariablen die Genauigkeit nur unwesentlich. Deswegen wurde er aus unserem Pool von Prädiktoren entfernt. Das bedeutet nicht, dass er das PUUV-Infektionsrisiko nicht beeinflusst, sondern dass er für unsere spezifische Problemdefinition nicht geeignet war, eine Vorhersage zu verbessern. Wir gehen davon aus, dass die Effekte der Landbedeckungsdaten in die kreisbezogene Inzidenztransformation einfließen, d. h., dass der Laubwaldanteil zwar auf die lokale Größenordnung der PUUV-Inzidenz einwirken kann, aber nicht die Wahrscheinlichkeit eines Ausbruchs selbst beeinflusst.

Für die weitere Suche konzentrierten wir uns auf Teilmengen mit wenigen Prädiktoren. Wir wandten einen SVC mit einem linearen Kern für alle 2-Variablen-Kombinationen aller 168 möglichen Wettervariablen an. Die Blühstärke aus dem Vorjahr wurde separat betrachtet, um

ihren Einfluss quantitativ zu prüfen. Da unter den besten Modellen nur selten Variablen nach September des Vorjahres auftauchten, entfernten wir diese, um das Frühwarnsystem mit möglichst viel Vorlaufzeit auszustatten.

Wir haben die Jahr-Monat-Kombinationen ausgewählt, die in den besten 100 Modellen für alle 2-Variablen-Kombinationen der 147 möglichen Wettervariablen aus den beiden Vorjahren bis zum vorhergehenden September am häufigsten vorkamen. Die Wetterbedingungen im April und im September aus dem Vorvorjahr und im September aus dem Vorjahr tauchten auf Platz 1-3 auf. Dies ist ein starker Hinweis darauf, dass diese Variablen die Wahrscheinlichkeit eines Nagetierausbruchs und folglich eines humanen PUUV-Ausbruchs am stärksten beeinflussen.

Um Mengen mit stark korrelierten Variablen zu vermeiden, bildeten wir die weiteren Teilmengen mit N-Prädiktoren, indem wir die N-Variablen-Kombinationen auswählten, mit genau einer Variable aus jeder der N obersten Jahr-Monat-Kombinationen (N-faches kartesisches Produkt) für Modelle ohne bzw. mit der Blühstärke aus dem Vorjahr. Aus allen Modellen für jede unterschiedliche Anzahl von Prädiktoren bis 6 bevorzugten wir das Modell mit der besten Aussagekraft. Schließlich beurteilten wir anhand der Leave-One-Out-Kreuzvalidierung die Vorhersagekraft der 13 besten Modelle: 6 Modelle ohne und 7 Modelle mit der Blühstärke aus dem Vorjahr.

Wir wählten die Modelle mit fünf Prädiktoren ohne bzw. zuzüglich der Blühstärke aus dem Vorjahr aus. Für die Teilmengen ohne die Blühstärke aus dem Vorjahr war die Auswahl eindeutig, da die Vorhersagekraft bei einer Teilmengengröße von 6 abnahm. Für die Teilmengen mit der Blühstärke aus dem Vorjahr blieb die Voraussagekraft konstant. Dafür wählten wir die Teilmenge mit der kleinsten Größe aus. Eine geringe Modellkomplexität insbesondere ohne hochkorrelierte Prädiktoren ist wichtig, um Überanpassung zu vermeiden und dem „Fluch der Dimensionalität“ zu entgehen.

#### *Das Argument für die Blühstärke der Buche*

Durch alleinige Betrachtung der Aussagekraft an dem Trainingsdatensatz würden wir zum Schluss kommen, dass die Blühstärke aus dem Vorjahr nicht notwendig ist, weil sie die Genauigkeit nicht verbesserte. Der Unterschied wurde jedoch in der Vorhersagekraft an dem Testdatensatz deutlich, weil die Genauigkeit der Modelle mit der Blühstärke aus dem Vorjahr immer höher war. Der Einbezug der Blühstärke aus dem Vorjahr wird voraussichtlich die Vorhersagekraft einer beliebigen Teilmenge von Prädiktoren erhöhen.

Das Modell mit nur der Blühstärke aus dem Vorjahr hatte bereits eine Genauigkeit von 79,6%. Ein Modell mit nur Wetterparametern musste mindestens fünf Variablen enthalten, um dieses Genauigkeitsniveau zu erreichen. Dies ist eine Demonstration der Vorhersagekraft der Blühstärke aus dem Vorjahr. Obwohl die Wetterparameter mit der Blühstärke korreliert sind, führt ihre Einbeziehung zu einem robusteren Modell.

Die Blühstärke ist zeitvariabel und ihre räumliche Verteilung spiegelt die Mastvariabilität in jedem Jahr wider. Damit können räumlich variierende Effekte berücksichtigt werden, die nicht allein durch die Wetterbedingungen erklärt werden können.

#### **Optimale Modelle**

Das Modell mit fünf Wetterprädiktoren zuzüglich der Blühstärke aus dem Vorjahr hatte anhand der LOOCV eine mittlere jährliche Genauigkeit von 82,8% in 16 Jahren in allen untersuchten Landkreisen. Die Prädiktoren aus dem Vorvorjahr waren die maximale Lufttemperatur im April, der Niederschlag im September und die mittlere Lufttemperatur im November, sowie aus dem Vorjahr die Bodenfeuchte im Mai, die Bodentemperatur im September und die Blühstärke.

Da die Erhebung von Blühstärke-Daten auf freiwilliger Basis erfolgt und möglicherweise nicht langfristig gesichert ist, wählten wir auch das optimale Modell ohne Blühstärke aus, um sicherzustellen, dass das Prognosemodell auch ohne Blühstärke funktionsfähig bleiben wird. Anhand der LOOCV hatte das Modell mit nur fünf Wetterprädiktoren eine mittlere jährliche Genauigkeit von 80,5% in 16 Jahren. Die Prädiktoren aus dem Vorvorjahr waren die maximale Lufttemperatur im April und die Sonnenscheindauer im September, sowie aus dem Vorjahr die Bodenfeuchte im April und die Bodentemperaturen im Juni und im September.

In beiden Modellen waren die Wetterprädiktoren aus dem Vorvorjahr universeller, da sie sowohl die Wettereffekte auf die Buchensamenproduktion als auch den direkten Einfluss auf das Wachstum der Rötelmauspopulationen berücksichtigen können. Andererseits sind die Wetterprädiktoren aus dem Vorjahr bodenspezifisch, entweder die Feuchte oder die Temperatur, und stehen in direktem Zusammenhang mit dem Habitat der Nagetiere.

Aus dem Ausbruchsrisiko haben wir die vorhergesagte Inzidenz und die vorhergesagte Risikoklasse umgerechnet. Das Bestimmtheitsmaß der Inzidenz betrug 0,457 für das Modell mit der Blühstärke aus dem Vorjahr (0,418 für das Modell ohne Blühstärke) und die mittlere jährliche Genauigkeit der Risikoklassen betrug 72,2% (71,1%).

Für das endgültige Modell entfielen die Validierungs- und Testdatensätze und alle Datenpunkte wurden in den Trainingsdatensatz aufgenommen, da alle Jahre wichtig sind. Die beiden Modelle haben eine vergleichbare Risikoklassen-Genauigkeit: 83,2% für das Modell mit der Blühstärke aus dem Vorjahr und 83,3% für das Modell mit nur Wettervariablen. Die Wahrheitsmatrizen beider Modelle sind diagonaldominant, mit 178 Überschätzungen und 84 Unterschätzungen für das Modell mit der Blühstärke aus dem Vorjahr; sowie 183 Überschätzungen und 78 Unterschätzungen für das Modell mit nur Wettervariablen über die 16 Jahre hinweg.

Beide Modelle ermittelten ein hohes Ausbruchsrisiko für alle Landkreise in den Jahren 2010, 2012, 2017 und 2019; und ein geringes Ausbruchsrisiko in den Jahren 2008, 2009, 2011, 2013, 2016, 2018 und 2020. Zusätzlich ermittelte das optimale Modell mit der Blühstärke aus dem Vorjahr ein geringes Ausbruchsrisiko in den Jahren 2006 und 2014; wohingegen das Modell ohne die Blühstärke zusätzlich ein hohes Ausbruchsrisiko für alle Landkreise in den Jahren 2015 und 2021 ermittelte. Die restlichen Jahre bestanden aus einer Mischung aus geringem und hohem Ausbruchsrisiko.

### **Fazit und Ausblick**

Dies ist das erste einheitliche Prognosemodell für das humane PUUV-Infektionsrisiko in Deutschland. Es kombiniert eine geringe Komplexität mit einer hohen Vorhersagekraft und ist erweiterbar auf weitere Gebiete Deutschlands, in denen vielleicht zukünftig das PUUV endemisch wird.

Wir führten das Ausbruchsrisiko ein, eine neue Größe, welche der Prognose von Infektionen und Risikoklassen vorangestellt ist. Basierend auf dieser Definition findet in jedem Landkreis in jedem Jahr entweder ein humaner PUUV-Ausbruch statt oder nicht. Unsere Methode berücksichtigt keine quantitativen Auswirkungen der Prädiktoren auf die Inzidenz oder die Infektionen. Stattdessen wird davon ausgegangen, dass die Inzidenz zwischen zwei Extremwerten schwankt, die jeweils einer unterliegenden Wahrscheinlichkeitsverteilung folgen.

Unsere Methode vermeidet die Einbeziehung stark korrelierter Variablen, die zu Überanpassung führen könnten, und weist eine hohe Sensitivität und Präzision auf. Die beiden optimalen Modelle benötigen nur 5 bzw. 6 Prädiktoren um eine hohe Vorhersagekraft mit einer mittleren jährlichen Genauigkeit von über 80% bei der Leave-One-Out-Kreuzvalidierung zu erreichen. Der letzte Prädiktor stammte vom September aus dem Vorjahr, was genügend Zeit bietet, um

Gegenmaßnahmen für ein potentielles Ausbruchsjahr vorzubereiten und Gesundheitsbehörden, Risikogruppen und Ärzte für die Gefährdung durch das Virus zu sensibilisieren.

Neben dem Ausbruchsrisiko hängt die Unsicherheit der Inzidenz und der Risikoklasse von der Berechnung des Erwartungswerts aus den vergangenen Zeitreihen ab. Unser Ansatz geht davon aus, dass die Inzidenzwerte einer Gammaverteilung um jeden der beiden Extremwerte folgen. Die restliche Ungenauigkeit ist auf in den Modellen nicht berücksichtigte Phänomene und auf probabilistische Effekte zurückzuführen, die zu einem Ausbruch führen können, auch wenn die Prädiktorenwerte nicht darauf hinweisen und umgekehrt. Die Wettervariablen aus dem Prognosejahr können nicht für ein Prognosemodell verwendet werden, beeinflussen aber trotzdem sowohl die Rötelmauspopulationen als auch die menschlichen Aktivitäten sowie deren Interaktion. Außerdem ist zu erwarten, dass alle präventiven Maßnahmen das zukünftige Infektionsrisiko beeinflussen, insbesondere reduzieren, vor allem nach der Etablierung dieses systematischen Prognosemodells. Dies sollte bei jeder zukünftigen Validierung oder Modellaktualisierung berücksichtigt werden.

Unser Prognosemodell benötigt keine Daten zur Rötelmauspopulation (z. B. Abundanz und PUUV-Prävalenz) und erfordert daher keine Feldarbeit, wie Monitoring, Fangen und Beprobieren. Dennoch sind solche Daten für die Validierung des Modells unerlässlich, insbesondere um falsche Vorhersagen zu minimieren.

#### **Publikation der Prognoseergebnisse auf einer verstetigten Plattform**

Das Julius Kühn-Institut Münster stellte zwei Programme zur Online-Publikation zur Verfügung: das erste Programm führt die Erfassung und Aufbereitung der für die Prognose erforderlichen Eingaben durch und das zweite Programm erstellt die Prognose anhand der erfassten Daten. Die Prognose-Ausgabe ist im Shapefile-Format und Bild-Format möglich.

Die Programme wurden in Python v3.8.5 generiert und auf Windows 10 getestet. Beide Programme werden über eine einzige Konfigurationsdatei gesteuert. Im einfachsten Fall sollte die nutzende Person nur das Jahr festlegen, für welches das humane PUUV-Infektionsrisiko vorhergesagt werden soll.

Zwei SVC-Modelle stehen zur Verfügung: die optimalen Modelle mit bzw. ohne die Blühstärke aus dem Vorjahr. Um das Modell mit Blühstärke zu verwenden, sollte die nutzende Person die Blühstärke-Daten, die voraussichtlich in der August-Ausgabe des Fachmagazins AFZ-DerWald veröffentlicht werden, manuell erfassen.

Wenn einzelne Blühstärke-Werte fehlen, wird die Prognose für die restlichen Landkreise dennoch erstellt werden. Eine Prognose für die Landkreise mit fehlenden Blühstärke-Werten kann mit dem Modell ohne Blühstärke erstellt werden. Die Inzidenz-Zeitreihen können mit zukünftigen Werten aktualisiert werden und daher die Inzidenz-Vorhersage stetig verbessern. Die Schwellenwerte der Risikoklassen und die Farbpalette für die Bilder der Prognose-Ausgabe können konfiguriert werden. Zusätzlich können die Programme zukünftige Bevölkerungsdaten automatisch erheben und bearbeiten, weitere Landkreise einbeziehen, und ein aktualisiertes SVC-Modell für die Prognose einsetzen.

## **Prognose für Jahre außerhalb des Modellierungszeitraums**

### **Prognose und Validierung für das Jahr 2022**

Mit Wetterparametern aus den Jahren 2020 und 2021 und den Blühstärke-Werten aus dem Jahr 2021 haben wir die Prognose für das Jahr 2022 erstellt. Die beiden Modelle ermittelten für alle Landkreise ein geringes vorhergesagtes Ausbruchsrisiko. Beide Modelle haben nur für 5 Landkreise eine Inzidenz höher als 2 Infektionen pro 100.000 Einwohner vorhergesagt: LK Göppingen, LK Heidenheim und LK Sigmaringen von Baden-Württemberg, sowie LK Freyung-Grafenau und LK Main-Spessart von Bayern. Die übrigen Landkreise hatten ein niedriges vorhergesagtes Risiko.

Die höchste echte Inzidenz in 2022 war 2,55 im LK Freyung-Grafenau (BY) mit 2 Infektionen. Der Landkreis Böblingen (BW) hat 4 Infektionen gemeldet und die Landkreise Karlsruhe (BW) und Wesel (NW) haben jeweils 3 Infektionen gemeldet. Die Genauigkeit der Prognose war 100% für das Ausbruchsrisiko und 93,9% für die Risikoklassen, mit 62 richtigen Vorhersagen und 4 Überschätzungen für Landkreise in BW und BY (LK Göppingen, LK Heidenheim, LK Sigmaringen und LK Main-Spessart).

### **Prognose für das Jahr 2023**

Mit Wetterparametern aus den Jahren 2021 und 2022 und den Blühstärke-Werten aus dem Jahr 2022 haben wir die Prognose für das Jahr 2023 erstellt. Die beiden Modelle ermittelten die gleiche Prognose für 14 Landkreise in BW und BY mit geringem Ausbruchsrisiko; und für 11 Landkreise in NI und NW mit hohem Ausbruchsrisiko. Von beiden Modellen wird eine relativ hohe Anzahl von humanen PUUV-Infektionen für die folgenden Landkreise im Jahr 2023 vorhergesagt: LK Borken, LK Coesfeld, LK Emsland, LK Grafschaft Bentheim, LK Osnabrück (zusammengefasst mit dem SK Osnabrück), LK Recklinghausen, LK Steinfurt, LK Warendorf, LK Wesel, SK Köln und SK Münster.

Das Modell mit der Blühstärke sagte zusätzlich in 41 Landkreisen in Mittel- und Süddeutschland ein hohes Ausbruchsrisiko vorher, während das Modell ohne Blühstärke dort keine Ausbrüche vorhersagte. Dieser Unterschied pflanzte sich in den Risikoklassen fort. Der Unterschied zwischen den beiden Modellen wird durch einen Vorzeichenwechsel bei den Abständen von diesen 41 Datenpunkten zur Hyperebene verursacht. Diese Vorzeichenwechsel können als ein Versatz des Clusters der Datenpunkte betrachtet werden, die sich der Hyperebene nähern oder sie sogar überschreiten, und beweisen die Bedeutung und die Dominanz der Blühstärke für das Modell.

## Summary

### Fundamentals

The Puumala Orthohantavirus (PUUV) is the most common hantavirus in Europe. It is transmitted by bank voles (*Clethrionomys glareolus*, syn. *Myodes glareolus*) and is highly endemic in Southern and Northwestern Germany. The PUUV can cause mild to moderate hemorrhagic fever with renal syndrome (nephropathia epidemica). According to the Robert Koch Institute, the human PUUV infections fluctuate over several years and have exceeded one thousand cases in six years since 2001.

The PUUV outbreaks are driven by the beech (*Fagus sylvatica*) mast. The underlying mechanisms for its transmission rate to humans seem too complex to model directly, as they depend on the abundance of the bank vole populations, their PUUV-prevalence, and the human-bank vole interaction; all of which fluctuate temporally and vary locally. Nevertheless, weather conditions can be used as predictors for human PUUV-infection risk, because the fluctuation in bank vole populations strongly correlates with weather parameters from the two previous years, whereas PUUV-prevalence mainly depends on bank vole abundance.

The PUUV infections are defined as the reported human hantavirus cases that are confirmed by a laboratory to be caused by the Puumala virus. The annual incidence is defined as the number of infections per 100,000 population in a year.

### Performance metrics

The performance of the predictions is evaluated using the following metrics, among others.

- ▶ The accuracy is the percentage of correctly predicted risk classes in the total predictions of a given dataset. This quantity should be maximized (maximum=100%, minimum=0%).
- ▶ The Pearson correlation coefficient is applied to the predicted and the true infections or incidence values of all years in the districts. The coefficient of determination ( $R^2$ ) can be approximated as the square of the Pearson correlation coefficient.

### Data collection

The following datasets were collected or digitized:

- ▶ Human infections caused by the Puumala virus, source: SurvStat@RKI 2.0, period: 2001–2022, temporal resolution: weekly or annual, spatial resolution: at the district level.
- ▶ Population, source: Eurostat, period: 2001–2021, temporal resolution: annual, spatial resolution: at district level.
- ▶ Weather data, source: Climate Data Center, period: 2000–2022, temporal resolution: monthly, spatial resolution: 1km x 1km grid.

Datasets: Mean values of the mean daily air temperature in °C, minimum daily air temperature in °C, maximum daily air temperature in °C, total precipitation in mm, total sunshine duration in hours, mean soil temperature at 5 cm depth in °C, mean soil moisture in percent plant useable water.

The grids for each of these datasets were processed using a vector layer of the district boundaries of Germany and the average values of the corresponding parameters were calculated at the district level.

- ▶ Land cover data, source: CORINE Land Cover, period: individual years (2000, 2006, 2012, and 2018), temporal resolution: annual, spatial resolution: 100m x 100m grid.

The proportion of the broad-leaved forest per district was calculated from the grid data. The amount of broad-leaved forest was estimated using the class 3.1.1 (“broad-leaved forest”). For the absolute proportion of deciduous forest, we used the CLC2018 reference dataset. For temporal trends in deciduous forest proportion, we assessed changes based on the “CLC-Change layers” (CHA - Change mapping) datasets for 2000-2006, 2006-2012, and 2012-2018.

- ▶ Flowering intensity of common beech, source: Dagmar Schneck (AFZ-DerWald and personal communication), period: 2000–2022, temporal resolution: annual, spatial resolution: at the level of the regions of provenance for the beech.

Germany is divided into 26 regions of provenance (abbreviated as HKG, from the German term Herkunftsgebiet) for common beech (*Fagus sylvatica*). Characteristic tree stands are selected from each HKG and their flowering intensity is assessed once a year in the first weeks of June. The individual assessments are averaged over each HKG. To increase the spatial resolution, we recalculated the flowering intensity at the district level by calculating the proportions of deciduous forest of each district included in each region of provenance, and then applying those proportions as weights to the weighted mean of the flowering intensity.

- ▶ Additional parameters for the validation of the previous models.

The mean value of the weather parameters in Baden-Württemberg was determined as previously described for the calculation of the district values, where the entire federal state was applied as mask.

The beech mast and the proportions of beech and other broad-leaved trees at the district level in Baden-Württemberg were collected from the prediction model implemented on the website [Hanta-Vorhersage](#).

- ▶ Vector data for Germany's district boundaries, source: Kreisgrenzen 2017 (Federal Agency for Cartography and Geodesy).

### Validation and comparison of existing models

At the project launch (November 2020), two regional PUUV prediction models were available in Germany. Both models were reproduced in Python. Using the self-collected datasets, retrospective predictions were made for the period 2006–2020 and compared to the true human incidence values or risk classes. The comparison was performed outside the modelling period of each model.

#### The BW model

The generalized linear model for Baden-Württemberg (FKZ 01 KI 1101; Monitoring of sylvatic zoonoses, Eprimos, 2014) was abbreviated as BW model. The modelling period was from 2005 to 2013. We reconstructed the model that was implemented on the website [Hanta-Vorhersage](#). The required inputs were: at the district level, the proportion of beech, the proportion of other broad-leaved trees, and the beech mast, and at the state level, weather parameters from the previous year: the mean temperature in September and the sunshine duration in August and September.

Our validation in the period 2014–2020 showed a tendency to overestimate the infections, because the cases in the outbreak years were significantly higher in the modelling period than in the validation period. The largest incidence deviation appeared as an overestimation in 2015. There were only a few isolated underestimations of the risk class. In general, the outbreak years (2017 and 2019) were correctly identified. The overall accuracy in the validation period was 57.6%. The highest value of the annual accuracy was 79% for 2014 and the lowest value was 9% for 2015.

In 2021, the prediction model on the website [Hanta-Vorhersage](#) deviated from the original model. We used the reported expected values directly from the website without applying our own reproduction. The model showed high predictive power with 76% accuracy of the risk classes (28 correct predictions, with 5 overestimations and 4 underestimations).

Our validation of the BW model was only as good as how close our data were to the original datasets. A definitive assessment for the quality and value of the BW model should therefore be viewed with caution. No prediction for either 2022 or 2023 was available on the [Hanta-Vorhersage](#) website.

### **The JKI models**

The separate CART-based models for districts in Southern and Northern Germany (FKZ 3713 48 401, JKI and FLI, 2018) were abbreviated as JKI Model-South and JKI Model-North, respectively. Both models determine directly risk classes: the first for Baden-Württemberg and Bavaria; and the second for Lower Saxony and North Rhine-Westphalia. The modelling period was 2006–2015 for both models, with half of the observations randomly used as a test dataset. The JKI Model-South requires the maximum temperature in August from two years before the prediction year, the sunshine duration in May of the previous year and in March of the prediction year, and the proportion of deciduous forest. The JKI Model-North requires the maximum temperature in October from two years before, the precipitation in January of the previous year, and the sunshine duration and maximum temperature in April of the previous year. All the data are required at the district level.

Within the framework of this study, both JKI models were validated in the period 2016–2020. The overall accuracy during the validation period was 47.5% for the JKI Model-South and 55.5% for the JKI Model-North. The accuracy of both models for all four federal states was better for non-outbreak years (2016, 2018, and 2020) than for outbreak years (2017 and 2019). The JKI Model-South correctly followed the general trend of the true incidence for all years. It demonstrated some predictive power and detected both outbreak years in the validation period, even though its regional predictions were not always accurate. On the other hand, the JKI Model-North did not explain any variance of the incidence in Northern Germany, an indication of overfitting in the training dataset. Consequently, the JKI Model-North is unsuitable for predictions. Its accuracy is explained by the low incidence values in the North.

Some comparisons at the nodes of the decision trees were inadequate. This is expected because it is unclear how suitable the decision trees are in the validation period, even though they may accurately reproduce the results of the modelling period.

### **Comparison and conclusion**

The predictions of the BW model were compared with the predictions of the JKI Model-South for Baden-Württemberg. In the overlapping period 2016–2020, the performance of the BW model was better for all metrics (for example, the accuracy of the risk classes was 63% versus 46%). The JKI Model-South was created with additional data from Bavaria, and it was expected to have a lower performance compared to a BW-specific model.

Due to the inaccuracy of the JKI models in the validation period and the extreme specialization of the BW model on a single federal state, it was necessary to develop a new prediction model that is uniform across all the PUUV-endemic areas of Germany. The development was based on findings from the validation of the two previous models. The BW model included a district-based scaling factor that accounted for differences in the local magnitude of the PUUV incidence. Our method builds on this idea and extends it by decoupling the calculation of the scaling factors from the prediction model itself and combining it with the incidence transformation used by the JKI models. On the other hand, the JKI models used weather data with higher spatial resolution and applied intrinsically-defined thresholds for the risk classes. Finally, the decision trees used by the JKI models demonstrated the advantages of a schematic representation of the model target and inspired our implementation of a geometrically-motivated classification algorithm.

### **Development and validation of a new, countrywide prediction model**

The analysis was limited to the PUUV infections after 2006 to avoid bias due to underreporting. The years 2006–2021 contain 91.9% of the total reported PUUV cases from 2001 to 2021. Each observation in the dataset represents a specific district in a specific year. The parameters of the observations are the predictor values in that district in that year. The labels of the observations are the PUUV infections, incidence, and risk classes as well as the outbreak risk, a new quantity introduced by this study.

### **Combination of urban and rural districts**

To compensate partially for differences between the reported place of residence and the place of infection, we combined most of the urban districts with their surrounding rural district. The only urban districts that remained separate were SK Köln (Cologne), SK Münster, and SK Stuttgart, whose areas are distinctly large. The naming convention matches that of the German version of SurvStat@RKI 2.0, where LK stands for a rural district (from the German *Landkreis*) and SK stands for an urban district (from the German *Stadtkreis*).

The observations for the combined districts retained the designation of the rural district. For the infections and populations, we aggregated the individual values and recalculated the incidence. For their weather variables, we assigned the mean values weighted by the area of each district. For the deciduous forest proportion and beech flowering intensity, we recalculated the values based on the aggregated deciduous forest and the deciduous forest per HKG, respectively.

### **Selection of the districts**

To select the districts where PUUV is actually endemic and to exclude individual cases from non-endemic regions, we included in the analysis only districts where the total infections were  $\geq 20$  and the maximum annual incidence was  $\geq 2$  in the period 2006–2021. A total of 66 districts were selected based on these criteria: 26 districts from Baden-Württemberg (BW), 16 from Bavaria (BY), 8 from Hesse (HE), 3 from Lower Saxony (NI), 10 from North Rhine-Westphalia (NW), 1 from Rhineland-Palatinate (RP), and 2 from Thuringia (TH). The selected districts accounted for 10,090 human PUUV infections, 89.9% of the total infections reported countrywide during this period.

### **Preparation of the data**

#### *Incidence transformation and outbreak risk*

We first applied a log-transformation to the incidence values to increase the impact of nonzero values. To consider the effects that drive the occurrence of high district-relative incidence, we discretized the incidence at the district level with two bins, i.e., we binarized it. For the

binarization, we performed an unsupervised clustering of the log-transformed incidence, separately for each district. This transformation compensated for the high spatial inhomogeneity of the incidence and converted its prediction into a problem of extrema detection.

The binary classification, which we call „Ausbruchsrisiko“ (“outbreak risk”), resulted in two classes labelled “low risk” (0) and “high risk” (1). A local “outbreak” occurred in a year when the incidence in a district was classified in the high-incidence bin of the past reported values. The incidence in districts for non-outbreak years was zero or considered low relative to the past reported values in that particular district. From the total 1,056 observations (16 years × 66 districts), 682 were assigned low risk (65%) and 374 were assigned high risk (35%).

#### *Risk classes*

We selected three risk classes, defined by slight drops in the histogram of the incidence values. The first threshold of 2 infections per 100,000 population limits the low-risk class. The second threshold, which separates the medium-risk and the high-risk classes, was set at 9.5 infections per 100,000 population.

#### **Classification method and incidence prediction**

We concentrated only on those variable combinations that led to a linear decision boundary (hyperplane) for the classification of our selected target (“outbreak risk”). To detect the outbreaks, we used a classifier based on a support vector machine (SVC) with a linear kernel, because this algorithm combines high performance with low model complexity by returning the decision boundary as a linear equation of the inputs.

The output of the SVC represents the expected outbreak risk in each district: a 0 (no expected local outbreak) or a 1 (expected local outbreak). It can be converted to incidence by assigning the mean incidence from all observations in the training dataset from the district with the same outbreak risk, either 0 or 1. Then, the predicted risk class is determined by a simple classification of the predicted incidence based on the predefined thresholds.

#### **Predictors and feature selection**

The performance was evaluated using two methods:

- ▶ the explanatory power assessed the model performance on the training dataset, and
- ▶ the predictive power estimated the method performance on a test dataset.

The evaluation of the predictive power was performed by a leave-one-out cross-validation (LOOCV). A single year was removed from the training dataset (2006–2021), then the model was trained using the remaining years, and this trained model was used to predict the incidence values for the removed year. This was repeated for all sixteen years in the training dataset. Finally, the prediction for all years was evaluated using the accuracy of the outbreak risk.

From each monthly weather parameter, 24 entries were created for all months of the previous two years as possible predictors for the model. In total, there were 168 available predictors from the weather parameters (7 parameters × 24 months), thereafter called “variables”.

Other possible predictors were the beech flowering intensity of the previous year and from two years before. Beechnuts are a primary food source for bank voles, and their supply has been linked to the growth of bank vole populations in Central Europe. The inclusion of masting is expected to enhance the performance of our prediction model by incorporating effects that are not directly driven by weather. The flowering intensity in late spring and early summer is a good

proxy for masting, although increased flowering does not always translate to increased seed production.

The last possible predictor in our dataset was the proportion of deciduous forest, which has been used in the past as a predictor for the PUUV infection risk in Southern Germany.

Limiting the number of predictors for our final model reduces the risk of overfitting. Therefore, we established a heuristic method to search the multidimensional variable space.

We first applied an SVC with a linear kernel for all subsets with a single variable. The performance of the model with the flowering intensity of the previous year was better than that of the model with the flowering intensity from two years before (accuracy 79.6% versus 73.2%). Since the two variables had a strong negative correlation, the flowering intensity from two years before was removed from our pool of predictors and only the flowering intensity of the previous year remained.

The model with only the proportion of deciduous forest had no explanatory power because it always assigned a low outbreak risk. In addition, the inclusion of the deciduous forest proportion in the models with individual weather variables increased the accuracy only marginally. Therefore, it was removed from our pool of predictors. This does not mean that it does not affect the PUUV infection risk, but that it was not likely to improve the prediction for our specific problem definition. We consider that the effects of land cover data are incorporated into the district-based incidence transformation, i.e., the proportion of deciduous forest may affect the local magnitude of the PUUV incidence but it does not affect the probability of an outbreak.

For the further search, we focused on subsets with few predictors. We applied an SVC with a linear kernel for all 2-variable combinations of all 168 possible weather variables. The flowering intensity of the previous year was considered separately in order to quantitatively assess its influence. Since variables after the September of the previous year rarely appeared among the best models, we removed them in order to provide as much lead-time for the early warning system as possible.

We selected the year-month combinations with the most occurrences in the best 100 models for all 2-variable combinations of the 147 possible weather variables from the previous two years until the previous September. Weather conditions in April and September from two years before, and in September of the previous year emerged in the top three places, providing strong evidence that these variables influence most strongly the likelihood of a rodent outbreak and consequently of a human PUUV outbreak.

To avoid sets with highly correlated variables, we formed the further subsets with N predictors, by selecting the N-variable combinations, with exactly one variable from each of the N best year-month combinations (N-fold Cartesian product) for models without or with the flowering intensity of the previous year. From all the models for each different number of predictors up to 6, we selected the model with the best explanatory power. Finally, using a leave-one-out cross-validation, we assessed the predictive power of the 13 optimal models: 6 models without the flowering intensity of the previous year and 7 models with it.

We selected the models with five predictors without and plus the flowering intensity of the previous year. For the subsets without the flowering intensity of the previous year, the selection was unambiguous because the predictive power decreased at a subset size of 6. For the subsets with the flowering intensity of the previous year, the predictive power remained constant. For this reason, we selected the subset with the smallest size. Low model complexity, especially

without highly correlated predictors, is important to avoid overfitting and the curse of dimensionality.

#### *The argument for the beech flowering intensity*

By observing the explanatory power on the training dataset alone, we would conclude that the flowering intensity of the previous year is not required. However, the difference became apparent in the predictive power on the test dataset, because the mean annual accuracy of the models with the flowering intensity of the previous year was always higher. Including the flowering intensity of the previous year is expected to increase the predictive power of any subset of predictors.

The model with only the flowering intensity of the previous year already had an accuracy of 79.6%. A model with only weather parameters had to include at least five variables to achieve this level of accuracy. This demonstrates the predictive power of the flowering intensity of the previous year. Although the weather parameters correlate with the flowering intensity, their inclusion results in a more robust model.

The flowering intensity is time-variant and its spatial distribution reflects the most variability in each year. This accounts for spatially varying effects that cannot be explained by weather conditions alone.

#### **Optimal models**

Using the LOOCV, the model with five weather predictors plus the flowering intensity of the previous year had a mean annual accuracy of 82.8% in 16 years across all selected districts. The predictors from two years before were the maximum air temperature in April, precipitation in September, and mean air temperature in November; and from the previous year were the soil moisture in May, soil temperature in September, and flowering intensity.

The collection of flowering intensity data in Germany is voluntary and may not be assured in the long term. Therefore, we also selected the optimal model without the flowering intensity to ensure that the prediction model would remain functional without the flowering intensity. Using the LOOCV, the model with only five weather predictors had a mean annual accuracy of 80.5% in 16 years. The predictors from two years before were the maximum air temperature in April and sunshine duration in September; and from the previous year were the soil moisture in April and soil temperatures in June and in September.

In both models, the weather predictors from two years before were universal because they can account for weather effects on beech seed production as well as for their direct influence on bank vole population growth. On the other hand, weather predictors from the previous year are soil-specific, either moisture or temperature, and are directly related to the rodent habitat.

From the outbreak risk, we estimated the predicted incidence and predicted risk class. The coefficient of determination for the incidence was 0.457 for the model with the flowering intensity of the previous year (0.418 for the model without the flowering intensity), and the mean annual accuracy of the risk classes was 72.2% (71.1%).

For the final model, the validation and test datasets were omitted, and all observations were included in the training dataset because all years are important. The two models have comparable risk class accuracy: 83.2% for the model with the flowering intensity of the previous year and 83.3% for the model with only weather variables. The confusion matrices of both models are diagonally dominant, with 178 overestimations and 84 underestimations within these 16 years for the model with the flowering intensity of the previous year; and 183 overestimations and 78 underestimations for the model with only weather variables.

Both models assigned a high outbreak risk for all districts in 2010, 2012, 2017, and 2019; and a low outbreak risk in 2008, 2009, 2011, 2013, 2016, 2018, and 2020. In addition, the optimal model with the flowering intensity of the previous year assigned a low outbreak risk in 2006 and 2014; whereas the model without the flowering intensity additionally assigned a high outbreak risk for all districts in 2015 and 2021. The remaining years consisted of a mixture of low and high outbreak risk.

### **Conclusion and outlook**

This is the first prediction model for human PUUV infection risk across Germany. It combines low complexity with high predictive power and is transferable to other areas of Germany where the PUUV may become endemic in the future.

We introduced the outbreak risk, a new quantity that precedes the prediction of infections and risk classes. Based on this definition, in each year each district will either have an outbreak or not. Our method does not consider any quantitative effects of the predictors on incidence or infections. Instead, incidence is assumed to fluctuate between two extreme values, each following an underlying probability distribution.

Our method avoids including highly correlated variables that could lead to overfitting and it is bound to have high sensitivity and precision. The two optimal models require only 5 and 6 predictors, respectively, to achieve high predictive power with > 80% mean annual accuracy in a leave-one-out cross-validation. The last predictor was from the previous September, which provides sufficient time to prepare countermeasures for a potential outbreak and to raise awareness in health authorities, risk groups and medical practitioners about human PUUV infection risk.

In addition to the outbreak risk, the uncertainty of incidence and risk class depends on the calculation of the expected value from the past time series. Our approach assumes that the incidence follows a gamma distribution around each of the two extreme values. The remaining inaccuracy is due to phenomena disregarded from the models and to probabilistic effects that may lead to an outbreak even if the predictor values do not indicate it, and vice versa. Weather variables from the actual year cannot be used for a prediction model, but still influence both the bank vole populations and the human activities as well as their interactions. In addition, any preventive measures are expected to influence the future infection risk, specifically to reduce it, especially after this systematic prediction model is established. This should be considered in any future validation or model update.

Our prediction model does not require rodent population data (e.g., abundance and PUUV prevalence) and therefore, does not require fieldwork such as monitoring, trapping, and sampling. Nevertheless, such data are essential for validating the model, especially in order to minimize inaccurate predictions.

### **Publication of the prediction results on a robust platform**

The Julius Kühn Institute Münster provided two programs for online publication: the first program performs the collection and preparation of the variables required for the prediction, and the second program generates the prediction based on the collected data. The prediction output is available in shapefile and image formats.

The programs were generated in Python v3.8.5 and tested on Windows 10. Both programs are controlled by a single configuration file. In the simplest case, the user only specifies the year for which the human PUUV infection risk should be predicted.

Two SVC models are available: the optimal models with and without the flowering intensity of the previous year. To use the model with the flowering intensity, the user should manually collect the flowering intensity data, which are expected to be published in the August issue of the magazine AFZ-DerWald.

If any individual flowering intensity values are missing, the prediction for the remaining districts will still be generated. A prediction for the districts with missing flowering intensity values can be made using the model without the flowering intensity. The incidence time series can be updated with future values and therefore, steadily improve the incidence prediction. The risk class thresholds and the color palette for the prediction output images can be configured. In addition, the programs can automatically collect and process future population data, include additional districts, and use an updated SVC model for the prediction.

### **Prediction for years outside the modelling period**

#### **Prediction and validation for the year 2022**

Using weather parameters from 2020 and 2021, and flowering intensity values from 2021, we generated the prediction for 2022. The two models assigned a low predicted outbreak risk for all districts. Both models predicted an incidence higher than 2 infections per 100,000 population for only 5 districts: LK Göppingen, LK Heidenheim, and LK Sigmaringen from Baden-Württemberg, and LK Freyung-Grafenau and LK Main-Spessart from Bavaria. The remaining districts had a low predicted risk.

The highest true incidence in 2022 was 2.55 in the LK Freyung-Grafenau (BY) with 2 infections. The district of Böblingen (BW) reported 4 infections and the districts of Karlsruhe (BW) and Wesel (NW) each reported 3 infections. The prediction accuracy was 100% for the outbreak risk and 93.9% for the risk classes, with 62 correct predictions and 4 overestimations for districts in BW and BY (LK Göppingen, LK Heidenheim, LK Sigmaringen, and LK Main-Spessart).

#### **Prediction for the year 2023**

Using weather parameters from 2021 and 2022, and flowering intensity values from 2022, we generated the prediction for 2023. The two models assigned the same prediction for 14 districts in BW and BY with low outbreak risk, and for 11 districts in NI and NW with high outbreak risk. A relatively high number of human PUUV infections is predicted by both models for the following districts in 2023: LK Borken, LK Coesfeld, LK Emsland, LK Grafschaft Bentheim, LK Osnabrück (combined with SK Osnabrück), LK Recklinghausen, LK Steinfurt, LK Warendorf, LK Wesel, SK Köln (Cologne) and SK Münster.

The model with the flowering intensity predicted that 41 additional districts in Central and Southern Germany will have an outbreak risk, whereas the model without the flowering intensity predicted no outbreaks there. This difference propagated through the risk classes. The difference between the two models is caused by a sign change in the distances of those 41 observations from the hyperplane. These sign changes can be regarded as an offset of the data cluster that approaches or even exceeds the hyperplane, demonstrating the importance and dominance of the flowering intensity for the model.

# 1 Grundlagen

## 1.1 Einleitung

Veränderte Umweltbedingungen, die durch den Klimawandel ausgelöst werden, spielen eine immer größere Rolle bei der Ausbreitung zoonotischer Infektionskrankheiten, indem sie die natürlichen Lebensräume der Tiere verändern, die Nahrungsverfügbarkeit beeinflussen und sogar Veränderungen in der Verbreitung der Arten bewirken. In diesem Rahmen haben wir ein Vorhersagemodell für das humane Infektionsrisiko mit dem Puumala Orthohantavirus (PUUV) in Deutschland entwickelt.

Das PUUV ist das häufigste Hantavirus in Europa, wird vor allem durch Rötelmäuse (*Clethrionomys glareolus*, Syn. *Myodes glareolus*) übertragen und ist hochendemisch im Süden und Nordwesten Deutschlands. Das PUUV kann leichtes bis mäßiges hämorrhagisches Fieber mit renalem Syndrom (Nephropathia epidemica) mit einer Sterblichkeitsrate von 0,1–0,4 % verursachen (Krüger et al. 2011). Die humanen PUUV-Infektionen schwanken mehrjährig. Nach Angaben des Robert Koch-Instituts fand der erste dokumentierte Ausbruch humaner PUUV-Infektionen im Jahr 2007 mit 1.625 Fällen bundesweit statt. Seitdem haben die jährlichen Humanfälle in fünf Jahren die Tausendergrenze überschritten. Die Ausbrüche sind großräumig und von der Mast der Rotbuche (*Fagus sylvatica*) getrieben (Clement et al. 2009; Tersago et al. 2009; Imholt et al. 2015; Reil et al. 2016; Reil, Binder, et al. 2018).

Die zugrundeliegenden Mechanismen für die Übertragungsrate des PUUV auf den Menschen scheinen zu komplex zu sein, um sie direkt zu modellieren, da sie von der Abundanz des Nagetierreservoirs, ihrer PUUV-Prävalenz und der Interaktion zwischen Mensch und Rötelmaus abhängen, welche alle zeitlich schwanken und lokal variieren. Dennoch können Wetterparameter als Prädiktoren für das humane PUUV-Infektionsrisiko verwendet werden, da die Schwankungen der Rötelmauspopulationen stark mit den Wetterparametern der beiden vorangegangenen Jahre korrelieren (Imholt et al. 2015), während die PUUV-Prävalenz hauptsächlich von der Abundanz von Rötelmäusen abhängt (Tersago et al. 2009; Khalil et al. 2014; Reil et al. 2017).

## 1.2 Infektionen und Inzidenz

Als PUUV-Infektionen werden die gemeldeten humanen Hantavirus-Fälle definiert, die labordiagnostisch vom Puumalavirus verursacht wurden. Die jährliche Inzidenz (Jahresinzidenz) wird als die Anzahl der jährlichen Infektionen pro 100.000 Einwohner definiert:

$$\text{Inzidenz} = \frac{\text{Fälle}}{\text{Bevölkerung}/100.000}$$

## 1.3 Leistungskennzahlen

Die Güte der Vorhersagen wird anhand der folgenden Maße bewertet.

- ▶ Die „Wahrheitsmatrix“ (auch „Konfusionsmatrix“ genannt) dient als Überblick der Abweichung zwischen Vorhersage und Realität. Sie ist nur für die Risikoklassen relevant. In unserer Definition repräsentieren die Spalten die vorhergesagten Risikoklassen und die Zeilen die echten Risikoklassen. Die Matrix soll idealerweise diagonal oder zumindest diagonaldominant sein, d. h., dass die Diagonalelemente größer als die Summe der restlichen jeweiligen Zeileneinträge sein sollen.

- ▶ Die Genauigkeit (auch Treffergenauigkeit und Richtigkeit genannt) ist der prozentuale Anteil der richtig vorhergesagten Risikoklassen an den gesamten Vorhersagen eines bestimmten Datensatzes. Diese Größe soll maximiert werden (Maximum=100%, Minimum=0%). Als Baseline (Grundlinie) für die Genauigkeit werden insgesamt drei fiktive Modelle verwendet.
  - Das Nullmodell ermittelt immer die häufigste Risikoklasse im kompletten Datensatz, d. h. die Risikoklasse 1 („niedriges Risiko“). Nach Definition passt das Nullmodell in nicht-Ausbruchsjahren am besten, aber es ist ungeeignet in Ausbruchsjahren. Das Nullmodell hat keine Vorhersagekraft.
  - Das jährliche Mehrheit-Modell besteht aus mehreren jährlichen Nullmodellen. Für nicht-Ausbruchsjahre ist es gleich zum Nullmodell, aber für Ausbruchsjahre ermittelt es die häufigste Risikoklasse in diesem Jahr. Das Mehrheit-Modell erfordert Kenntnis der Inzidenz in allen Landkreisen in Deutschland in diesem Jahr.
  - Das Ausbruch-Modell ermittelt in Ausbruchsjahren die höchste jemals erfasste Risikoklasse für jeden Kreis. Die Ausbruchsjahre sollen extern definiert werden. Wir haben die Jahre mit mehr als 1.000 gemeldeten PUUV-Infektionen als Ausbruchsjahre für ganz Deutschland definiert, d. h. die Jahre 2007, 2010, 2012, 2017, 2019 und 2021 (Robert Koch-Institut). Für nicht-Ausbruchsjahre ist das Ausbruch-Modell auch gleich zum Nullmodell. Das Ausbruch-Modell erfordert Kenntnis der höchsten jemals erfassten Jahresinzidenz im Landkreis und ob im Prognosejahr ein Ausbruch stattfindet.
- ▶ Für die binäre Klassifizierung, zusätzlich zur Genauigkeit, definieren wir die Sensitivität (auch Recall und Trefferquote genannt) als prozentualer Anteil der echten Datenpunkte mit hohem Risiko, die richtig klassifiziert werden; und die Präzision (auch Relevanz und Wirksamkeit genannt) als prozentualer Anteil der ermittelten Datenpunkte mit hohem Risiko, die tatsächlich richtig klassifiziert werden. Diese Größen sollen ebenfalls maximiert werden (Maximum=100%, Minimum=0%).
- ▶ Das F1-Maß (F1-score) wird als das harmonische Mittel von Sensitivität und Präzision definiert. Diese Größe soll maximiert werden (Maximum=1, Minimum=0).

$$\text{F1 – Maß} = \frac{\text{Sensitivität} * \text{Präzision}}{\text{Sensitivität} + \text{Präzision}}$$

- ▶ Der Spearman-Korrelationskoeffizient und der Pearson-Korrelationskoeffizient werden auf die vorhergesagten und echten Infektionen bzw. Inzidenzwerte aller Jahre in den Landkreisen angewendet. Das Bestimmtheitsmaß ( $R^2$ ) kann als das Quadrat vom Pearson-Korrelationskoeffizienten angenähert werden.
- ▶ Ein Tukey's post-hoc Test kann zwischen den unterschiedlichen vorhergesagten Risikoklassen, bzw. zwischen den Blättern des Entscheidungsbaums, anhand der echten Inzidenz durchgeführt werden. Das Signifikanzniveau (p-Wert) ist überall 0,05, soweit nicht explizit anders angegeben. Ein p-Wert höher als das Signifikanzniveau bedeutet, dass die Mittelwerte der vorhergesagten Risikoklassen statistisch nicht signifikant unterschiedlich sind.

## 2 Sammeln von benötigten Daten für die Modellierung

Dieser Abschnitt basiert auf der projektbezogenen Veröffentlichung betitelt “Machine learning identifies straightforward early warning rules for human Puumala hantavirus outbreaks” (Übersetzung: „Maschinelles Lernen identifiziert eindeutige Frühwarnregeln für humane Puumala-Hantavirus-Ausbrüche“). In: Scientific Reports, 2023, 13(1:3585) (Kazasidis und Jacob 2023).

### 2.1 Rohdaten und Bearbeitung

Folgende Datensätze wurden gesammelt bzw. digitalisiert:

► Human-Erkrankungen durch das Puumalavirus (PUUV)

Quelle: SurvStat@RKI 2.0 (Robert Koch-Institut), Letztes Abfragedatum: 20.01.2023

Zeitraum: 2001–2022

Zeitliche Auflösung: wöchentlich bzw. jährlich

Räumliche Auflösung: auf Kreisebene

► Bevölkerung

Quelle: Demographische Veränderung - absoluter und relativer Bevölkerungsstand auf regionaler Ebene (Eurostat), Zugriffsdatum: 28.07.2022

Zeitraum: 2001–2021

Zeitliche Auflösung: jährlich

Räumliche Auflösung: auf Kreisebene

Für jedes Jahr haben wir die Bevölkerung verwendet, die für den 1. Januar dieses Jahres gemeldet wurde. Die Bevölkerung für 2021 wurde auch für die Jahre 2022 (Testdatensatz) und 2023 (Prognose) verwendet, da aktuellere Daten bis zum 25.01.2023 noch nicht zur Verfügung standen.

► Wetterdaten

Quelle: Climate Data Center (CDC) (Deutscher Wetterdienst), Letztes Zugriffsdatum: 25.11.2022

Zeitraum: 2000–2022

Zeitliche Auflösung: monatlich

Räumliche Auflösung: 1km x 1km Raster

Datensätze:

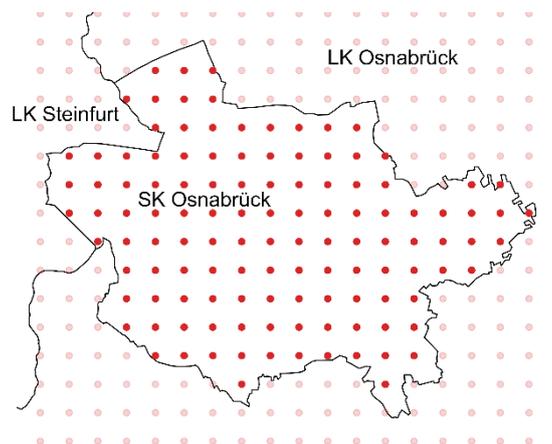
- Monatsmittel der Lufttemperatur (Tmean) in °C
- Lufttemperaturminima (Tmin) in °C
- Lufttemperaturmaxima (Tmax) in °C

- Monatssumme der Niederschlagshöhe (Pr) in mm
- Monatssumme der Sonnenscheindauer (SD) in Stunden
- mittlere Bodentemperatur in 5cm Tiefe (ST) in °C
- mittlere Bodenfeuchte (SM) in Prozent pflanzenverfügbares Wasser (% nFK).

Die Raster für jeden dieser Datensätze wurden mithilfe eines Geodaten-Vektorlayers der Kreisgrenzen Deutschlands (GeoBasis-DE / BKG 2017) bearbeitet und die Durchschnittswerte des entsprechenden Parameters auf Kreisebene berechnet. Zuerst wurde eine Maske basierend auf den im Kreis enthaltenen Rasterzentren definiert. Diese Maske wurde dann auf das Raster angewandt. Jedes Rechteck des Rasters wurde einem einzelnen Kreis zugeordnet, demjenigen, der sein Zentrum enthielt. Abbildung 1 zeigt die Rasterzentren, überlagert mit dem Beispiel Stadtkreis Osnabrück.

### Abbildung 1: Rasterung für die Bearbeitung der Wetterdaten.

Die Mittelpunkte eines 1km x 1km Rasters für die Bearbeitung der Wetterdaten überlagert mit dem Stadtkreis Osnabrück. Anhand des Rasters werden die Mittelpunkte ermittelt, die innerhalb der Grenzen jedes Landkreises liegen ("central point rasterizing"). Im Falle des SK Osnabrück liegen 120 Punkte innerhalb der Kreisgrenzen, die mit dunkelroter Farbe markiert sind. Die restlichen Punkte (außerhalb des SK Osnabrück) sind hellrot markiert. Der typische Verarbeitungsfehler wurde auf etwa 1% geschätzt, was als deutlich geringer angesehen wird als die Unsicherheiten der Raster selbst, die sich aus ungenauen Berechnungen, dem Interpolationsverfahren und aus fehlerhaften oder fehlenden Beobachtungen ergeben.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Hintergrundkarte: (GeoBasis-DE / BKG 2017).

#### ► Landbedeckungsdaten

Quelle: CORINE Land Cover (European Environment Agency), Zugriffsdatum: 29.12.2020

Zeitraum: einzelne Jahre: 2000, 2006, 2012 und 2018

Zeitliche Auflösung: jährlich

Räumliche Auflösung: 100m x 100m Raster

Aus den Rasterdaten wurde der Laubwaldanteil pro Kreis berechnet. Die Menge des Laubwaldes (BF) wurde anhand der Klasse 3.1.1 ("broad-leaved forest") abgeschätzt, die Flächen mit einem Deckungsgrad von mehr als 75% mit laubabwerfenden und immergrünen Laubbaumarten umfasst, mit unter anderem Rein- oder Mischbestände von Buche, Eiche, Hainbuche, Linde, Ahorn, Esche, Pappel und Birke (European Environment Agency - CORINE Land Cover).

Für den absoluten Laubwaldanteil haben wir den Referenzdatensatz CLC2018 verwendet. Für die zeitliche Entwicklung des Laubwaldanteils haben wir die Veränderungen auf Basis der “CLC-Change layers” (CHA - Change mapping) Datensätze für 2000-2006, 2006-2012 und 2012-2018 bewertet, die alle Veränderungen  $\geq 5$ ha abbilden. Anschließend interpolierten bzw. extrapolierten wir auf Basis der Nächster-Nachbar-Methode (nearest-neighbor).

► Blühstärke der Rotbuche

Quelle: Dagmar Schneck, Landesstelle für forstliches Vermehrungsgut, Landesbetrieb Forst Brandenburg, AFZ-DerWald (Schneck 2020, 2021, 2022) und persönliche Mitteilung

Zeitraum: 2000–2022

Zeitliche Auflösung: jährlich

Räumliche Auflösung: auf der Ebene der Herkunftsgebiete der Rotbuche

Deutschland ist in 26 Herkunftsgebiete (HKG) für die Rotbuche (*Fagus sylvatica*) unterteilt (Bundesanstalt für Landwirtschaft und Ernährung 2023a). Aus jedem HKG werden charakteristische Baumbestände ausgewählt, deren Blühintensität (FI) einmal im Jahr in den ersten Juniwochen bewertet wird (Schneck 2019, 2020, 2021, 2022). Für die Bewertung wird eine qualitative 4-Klassen-Skala verwendet (Schneck 2020): 1 steht für kein oder geringes Blühen (0–10%), 2 für schwaches oder mäßiges Blühen (11–30%), 3 für mittleres oder gutes Blühen (31–60%) und 4 für reichliches oder volles Blühen (61–100%). Die Einzelbewertungen werden über jedes HKG gemittelt.

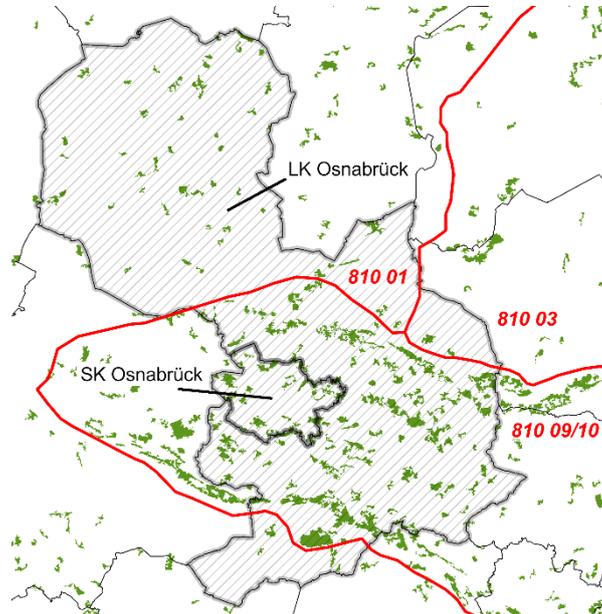
Um die räumliche Auflösung zu erhöhen, rechneten wir die Blühstärke auf Kreisebene ( $FI_{LK}$ ) um. Die in jedem Herkunftsgebiet enthaltenen Laubwaldanteile des Kreises ( $P_{HKG_i}$ ) wurden berechnet und dann als Gewichte für das gewichtete Mittel der Blühstärke angewandt.

$$FI_{LK} = \frac{\sum_{i \text{ in HKG}} (FI_i * P_i)}{\sum_{i \text{ in HKG}} P_i} \text{ mit } P_{HKG_i} = \frac{-BF \text{ vom LK in HKG}_i}{\text{Summe BF vom LK}}$$

Abbildung 2 zeigt ein Beispiel für die beiden Kreise von Osnabrück. Der Stadtkreis Osnabrück ist vollständig im HKG 810 09/10 enthalten. Obwohl das HKG 810 01 mehr als die Hälfte (56,4%) der Gesamtfläche des Landkreises Osnabrück einnimmt, umfasst es nur 20,0% seines gesamten Laubwaldes.

## Abbildung 2: Der Laubwald, Landkreise und Herkunftsgebiete.

Karte der beiden Kreise Osnabrück (Stadt- und Landkreis, Kreisgrenzen grau markiert) zusammen mit den Grenzen der Buchen-Herkunftsgebiete (HKG, Grenzen rot markiert) und Überlagerung mit der Laubwaldklasse aus CLC 2018. Der SK Osnabrück gehört zum HKG 810 09/10 [100% der Fläche und 100% vom Laubwald (BF)], während der LK Osnabrück in die HKG 810 01 (56,4% der Kreisfläche, 20,0% vom BF des Kreises), 810 03 (3,2% der Fläche, 1,3% vom BF) und 810 09/10 (40,4% der Fläche, 78,7% vom BF) unterteilt ist.



Für eine bessere Darstellbarkeit wurde die Standardfarbe #80ff00 der Klasse 3.1.1 in CLC 2018 (Sentinel Hub Public Collections) durch #5c961d ersetzt.

Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Datenquelle: (European Environment Agency). Hintergrundkarten: (GeoBasis-DE / BKG 2017) und (Bundesanstalt für Landwirtschaft und Ernährung 2023b).

### ► Zusätzliche Parameter für die Validierung der bisherigen Modellansätze

- Mittlerer Wert der Wetterparameter in Baden-Württemberg: Dieser wurde, wie vorher für die Berechnung der Kreiswerte beschrieben, ermittelt, wobei die Maske auf das gesamte Bundesland bezogen wurde.
- Buchenmast sowie Anteile der Buche und der sonstigen Laubbäume auf Kreisebene in Baden-Württemberg: gesammelt von dem auf der Internetseite [Hanta-Vorhersage](#) implementierten Vorhersagemodell (Eichner et al.).

### ► Vektordaten für die Kreisgrenzen Deutschlands

Quelle: Kreisgrenzen 2017, Bundesamt für Kartographie und Geodäsie (GeoBasis-DE / BKG 2017), Zugriffsdatum: 18.02.2021

## 2.2 Zusammenfassen von Stadt- und Landkreisen

Die von der SurvStat-Anwendung (Robert Koch-Institut) bereitgestellte räumliche Information zu den Infektionen bezieht sich auf den Landkreis, in dem die Infektion gemeldet wurde. Daher entspricht der gemeldete Landkreis in den meisten Fällen dem Wohnort der infizierten Person, der vom Ort der Infektion abweichen kann. Um Unterschiede zwischen dem gemeldeten Wohnort und dem Infektionsort teilweise auszugleichen, haben wir die meisten kreisfreien Städte (Stadtkreise) mit ihrem umfassenden oder umgebenden Landkreis zusammengefasst. Die

Grundannahme war, dass die meisten in Stadtkreisen gemeldeten Infektionen tatsächlich im benachbarten oder umliegenden Landkreis auftraten. Darüber hinaus haben einige Stadt- und Landkreise das gleiche Gesundheitsamt. Die einzigen Stadtkreise, die getrennt blieben, waren SK Köln, SK Münster und SK Stuttgart, die deutlich großflächig sind.

Tabelle 3 im Anhang listet die zusammengefassten Stadtkreise auf, die für diese Analyse relevant sind. Die Datenpunkte für die Kombinationen behielten die Bezeichnung des Landkreises. Für die Infektions- und Bevölkerungszahlen wurden die Einzelwerte aggregiert und die Inzidenz neu berechnet. Für die Wetterparameter haben wir die Mittelwerte, gewichtet nach der Fläche des jeweiligen Kreises, diesem zugeordnet. Für den Laubwaldanteil und die Blühstärke der Rotbuche haben wir die Werte auf der Basis des aggregierten Laubwaldes bzw. Laubwaldes pro HKG neu errechnet.

## 2.3 Datensatz

Jeder Datenpunkt des Datensatzes repräsentiert einen bestimmten Landkreis in einem bestimmten Jahr. Die Parameter des Datenpunktes sind die Prädiktorenwerte in diesem Landkreis in diesem Jahr. Die Labels des Datenpunktes (auch Targets genannt) sind das PUUV-Ausbruchsrisiko (eine eigene Größe dieser Studie, die in Kapitel 4 definiert wird) und die PUUV-Infektionen, -Inzidenz und -Risikoklasse.

In diesem Bericht sind die Begriffe „Testdatensatz“ und „Validierungsdatensatz“, sowie „echt“ und „gemeldet“ synonym und werden austauschbar verwendet. Die Letzteren unterscheiden sich von den Begriffen „vorhergesagt“, „prognostiziert“ und „ermittelt“, die ebenfalls synonym sind und austauschbar verwendet werden. Der Begriff „modelliert“ bezieht sich auch auf eine Klassifizierung durch das Modell, aber er hebt hervor, dass der Datenpunkt im Trainingsdatensatz enthalten ist, d. h., dass er für das Trainieren des Modells selbst verwendet wurde.

### 3 Validierung und Vergleich der bisherigen Modellansätze

Zum Zeitpunkt des Projektstarts (November 2020) waren zwei PUUV-Prognosemodelle bzw. Modellierungsansätze für verschiedene Regionen in Deutschland vorhanden:

- ▶ das generalisierte lineare Modell für Baden-Württemberg – FKZ 01 KI 1101; Monitoring sylvatischer Zoonosen, Eprimos, 2014 (Eichner et al.) – als *BW Modell* abgekürzt und
- ▶ die auf CART basierenden getrennten Modelle für Landkreise in Süd- bzw. Norddeutschland – FKZ 3713 48 401, JKI und FLI, 2018 (Reil, Jacob, et al. 2018) – als *JKI Modell-Süd* und *JKI Modell-Nord* abgekürzt.

Beide Modellierungsansätze wurden in Python (Van Rossum und Drake 2009) nachgebildet. Mit den selbst erhobenen Datensätzen wurden Vorhersagen retrospektiv für den Zeitraum 2006–2020 erstellt und mit den echten Human-Inzidenzwerten oder Risikoklassen verglichen. Der Vergleich wurde außerhalb des Modellierungszeitraums jedes Modells durchgeführt, d. h. von 2014 bis 2020 für das BW Modell und von 2016 bis 2020 für die JKI Modelle.

#### 3.1 BW Modell

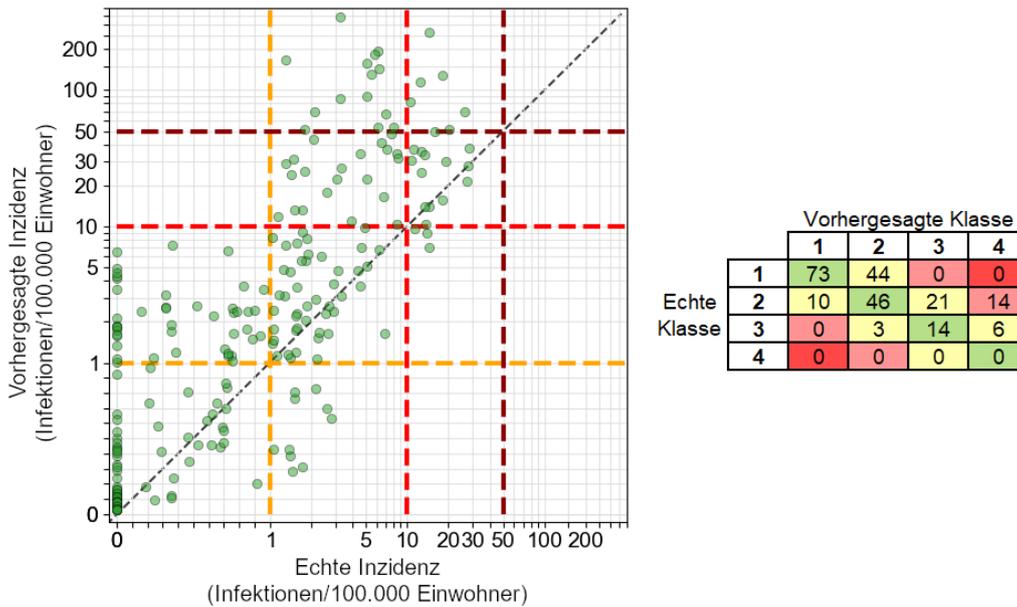
Für Baden-Württemberg wurden mehrere Modelle untersucht, die humane Hantavirus-Infektionen prognostizieren, mit unterschiedlichen Kombinationen von Prädiktoren in einem gemeinsamen Modellierungszeitraum von 2005 bis 2013. Von allen Modellen bildeten wir das sogenannte Herbstmodell nach, das auf der Internetseite [Hanta-Vorhersage](#) implementierte Vorhersagemodell (Eichner et al.). Die erforderlichen Eingaben waren: auf Kreisebene der Anteil der Buche, der Anteil der sonstigen Laubbäume und die Buchenmast, und auf Bundeslandebene Wetterparameter aus dem Vorjahr: die mittlere Temperatur im September und die Sonnenscheindauer im August und September.

Das BW Modell wurde im Rahmen der hier vorliegenden Studie im Zeitraum 2014–2020 validiert (Validierungszeitraum). Der Datensatz enthält 33 Landkreise in 7 Jahren, d. h. insgesamt 231 Datenpunkte. In den meisten Vorhersagen wurden die Infektionen und damit die Inzidenzwerte überschätzt. Die linke Seite der Abbildung 3 zeigt die vorhergesagten Inzidenzwerte gegenüber den echten Inzidenzwerten. Idealerweise würden alle Datenpunkte auf der schwarzen diagonalen Linie  $y=x$  liegen. Die meisten Datenpunkte liegen links (oberhalb) der diagonalen Linie, was zeigt, dass die Inzidenz überschätzt wurde. Die rechte Seite der Abbildung 3 zeigt die Wahrheitsmatrix, die diagonaldominant ist. Die Matrix bestätigt, dass die Überschätzung der Infektionen auch auf die Risikoklassen übertragen wurde (Abbildung 3, rechts; mehr Datenpunkte rechts–oberhalb–der Hauptdiagonale). Es gibt nur einzelne Unterschätzungen (links–unterhalb–der Hauptdiagonale), die meisten für das Jahr 2014 (5 Datenpunkte). Der Datensatz enthält 117 Datenpunkte mit niedrigem echtem Risiko (Risikoklasse 1) und somit betrug die Genauigkeit des Nullmodells 50,6%.

Die vielen Überschätzungen lassen sich dadurch erklären, dass die Infektionszahlen für die Ausbruchsjahre im Modellierungszeitraum deutlich höher als im Validierungszeitraum waren. Es gab sieben klare Ausreißer mit mehr als 400 vorhergesagten Infektionen, die unrealistisch für einen einzelnen Landkreis sind, weil die maximalen jemals gemeldeten jährlichen Infektionen in einem Landkreis unter 200 waren. Diese Ausreißer führen zu vorhergesagten Inzidenzwerten im Bereich von 86 bis 346. All diese Ausreißer treten für das Jahr 2015 auf, welches ein „kurioses“ Jahr war: es gab die klimatische Bedingungen, die auf einen Ausbruch hindeuteten, der jedoch nicht auftrat, zumindest nicht im Ausmaß vergangener oder späterer Ausbruchsjahre.

**Abbildung 3: Validierung für das BW Modell im Zeitraum 2014–2020.**

Die 231 Datenpunkte für die Validierung des BW Modells werden auf zwei verschiedene Arten präsentiert. Links: Die vorhergesagten Inzidenzwerte gegenüber den echten Inzidenzwerten. Für eine bessere Darstellbarkeit sind beide Achsen linear im Bereich [0, 1] und logarithmiert im Bereich [1, 400]. Die farbigen gestrichelten Linien bezeichnen die Schwellenwerte der Inzidenz für die Definition der Risikoklassen: 1, 10 und 50. Die horizontalen Linien definieren die vorhergesagten Risikoklassen und die vertikalen Linien definieren die echten Risikoklassen. Rechts: Die Wahrheitsmatrix kann mit dem linken Plot verglichen werden. Die Hauptdiagonale der Matrix entspricht der Linie  $y=x$  im Plot. Die Zellfarbe unterstützt die Anschaulichkeit. Die grünen Zellen (auf der Hauptdiagonale) sollten große Werte enthalten, während die gelben und die graduell mehr roten Zellen leer sein sollten.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

Tabelle 1 zeigt einige Leistungskennzahlen für die Bewertung der Validierung. Die gesamte Genauigkeit im Validierungszeitraum war 57,6% und damit höher als die des Nullmodells. Der höchste Wert der jährlichen Genauigkeit war 79% für 2014 (fast das Doppelte des Nullmodells) und der geringste Wert war 9% für 2015 (niedriger als die Hälfte des Nullmodells).

Der höchste Wert des Spearman-Korrelationskoeffizienten war 0,8560 für 2019 und der geringste Wert war 0,4462 für 2016. Der höchste Wert des Pearson-Korrelationskoeffizienten im Validierungszeitraum war 0,6082 ( $R^2=0,37$ ) für 2019 und der geringste Wert war 0,2438 ( $R^2=0,06$ ) für 2016.

**Tabelle 1: Leistungskennzahlen für die Validierung des BW Modells im Zeitraum 2014–2020.**

Die absoluten Infektionszahlen werden angegeben, um die Ausbruchsjahre zu erkennen. Die Genauigkeit des nachgebildeten Modells wurde mit der Genauigkeit des Nullmodells verglichen. Für 2014, 2017 und 2019 konnten die originalen Vorhersagen des BW Modells abgerufen werden. Die Genauigkeit des „originalen“ BW Modells ist für diese drei Jahre in Klammern angegeben und ein Proxy für die Qualität der Nachbildung.

Jahr	Infektionen BW	Genauigkeit Nullmodell	Genauigkeit BW Modell	Spearman-Koeff.	Pearson-Korrelationskoeff.
Insgesamt 2014–2020	1763	50,6%	57,6%	0,7857	0,3650
2014	145	39,4%	78,8% (87,9%)	0,7299	0,4798
2015	337	21,2%	9,1%	0,7523	0,4942
2016	39	84,8%	63,6%	0,4462	0,2438
2017	591	27,3%	42,4% (42,4%)	0,7666	0,4627
2018	43	87,9%	69,7%	0,5991	0,5236
2019	549	24,2%	66,7% (69,7%)	0,8560	0,6082
2020	59	69,7%	72,7%	0,6723	0,4705

### 3.1.1 Validierung für 2021

Im Jahr 2021 wich das Vorhersagemodell auf der Internetseite [Hanta-Vorhersage](#) (Eichner et al.) vom ursprünglichen Modell ab. Wir haben die gemeldeten Erwartungswerte direkt von der Webseite (Stand: 01.06.2021) verwendet, ohne unsere eigene Nachbildung anzuwenden. Der Datensatz enthielt 37 Landkreise, statt der 33 Landkreise des ursprünglichen BW Modells. Die vier zusätzlichen Landkreise stammen aus dem Regierungsbezirk Freiburg (LK Emmendingen, LK Lörrach, LK Waldshut und LK Breisgau-Hochschwarzwald) und hatten jeweils eine erwartete Inzidenz  $< 0,5$ . Wir gehen davon aus, dass ihnen im ursprünglichen Modell implizit ein geringes Risiko zugeordnet wurde oder dass diese vier Landkreise zusätzlich in das Modell eingefügt wurden.

Die Validierung für 2021 wurde mit den gesamten Hantavirus-Erkrankung-Fällen (Robert Koch-Institut: [SurvStat@RKI 2.0](#), Abfragedatum: 22.02.2022) und nicht mit nur den PUUV-Fällen ausgeführt, weil wir festgestellt haben, dass sie näher an dem Datensatz liegen, der zum Trainieren des BW Modells für 2021 verwendet wurde.

Abbildung 4 zeigt die Wahrheitsmatrix. Sie ist diagonaldominant mit 5 Überschätzungen und 4 Unterschätzungen. Das Bestimmtheitsmaß ( $R^2$ ) für die Infektionen war 0,61 und für die Inzidenz 0,55. Die Genauigkeit für die Risikoklassen war 76% (28 richtige Vorhersagen), im Vergleich zu 24% des Nullmodells (9 richtige Vorhersagen), 43% des jährlichen Mehrheit-Modells (16 richtige Vorhersagen) und 65% des Ausbruch-Modells (24 richtige Vorhersagen).

**Abbildung 4: Validierung für das BW Modell in 2021.**

Die Wahrheitsmatrix für das Jahr 2021.

		Vorhergesagte Klasse			
		1	2	3	4
Echte Klasse	1	9	0	0	0
	2	1	10	5	0
	3	0	1	9	0
	4	0	0	2	0

Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

**3.2 JKI Modelle**

Der Abschlussbericht des JKI Projektes FKZ 3713 48 401 nannte zwei unterschiedliche Modelle, die direkt Risikoklassen ermitteln: das erste für Baden-Württemberg und Bayern wird als JKI Modell-Süd abgekürzt und das zweite für Niedersachsen und Nordrhein-Westfalen wird als JKI Modell-Nord abgekürzt. Der Modellierungszeitraum war 2006–2015 für beide Modelle, wobei die Hälfte der Datenpunkte zufällig als Testdatensatz genutzt wurde. Das JKI Modell-Süd benötigt die maximale Temperatur im August aus dem Vorvorjahr, die Sonnenscheindauer im Mai aus dem Vorjahr und im März aus dem Prognosejahr und den Laubwaldanteil. Das JKI Modell-Nord benötigt die maximale Temperatur im Oktober aus dem Vorvorjahr, den Niederschlag im Januar aus dem Vorjahr und die Sonnenscheindauer und die maximale Temperatur im April aus dem Vorjahr. Alle erforderlichen Eingaben werden auf Kreisebene benötigt.

Beide JKI Modelle wurden im Rahmen dieser Studie im Zeitraum 2016–2020 validiert (Validierungszeitraum). Der Datensatz für Süddeutschland enthielt 72 Landkreise (35 von Baden-Württemberg und 37 von Bayern) in 5 Jahren, d. h. insgesamt 360 Datenpunkte. Der Datensatz für Norddeutschland enthielt 53 Landkreise (17 von Niedersachsen und 36 von Nordrhein-Westfalen) in 5 Jahren, d. h. insgesamt 265 Datenpunkte.

Die gesamte Genauigkeit im Validierungszeitraum war 47,5% für das JKI Modell-Süd, gegenüber 55,8% für das Nullmodell, und 55,5% für das JKI Modell-Nord, gegenüber 63,0% für das Nullmodell. Die vorhergesagten Inzidenzwerte wurden durch den Mittelwert jeder Risikoklasse im Trainingsdatensatz angenähert. Das Bestimmtheitsmaß für die Inzidenz war 0,321 für das JKI Modell-Süd. Das Bestimmtheitsmaß für das JKI Modell-Nord war negativ (-0,062), was bedeutet, dass dieses Modell keine Varianz der Inzidenz in Norddeutschland erklärte. Das negative Bestimmtheitsmaß für das JKI Modell-Nord weist auf eine Überanpassung („overfitting“) im Trainingsdatensatz hin.

Tabelle 2 zeigt einige Leistungskennzahlen für die Bewertung der Validierung. Die Genauigkeit beider Modelle für alle vier Bundesländer war besser für nicht-Ausbruchsjahre (2016, 2018 und 2020) als für Ausbruchsjahre (2017 und 2019).

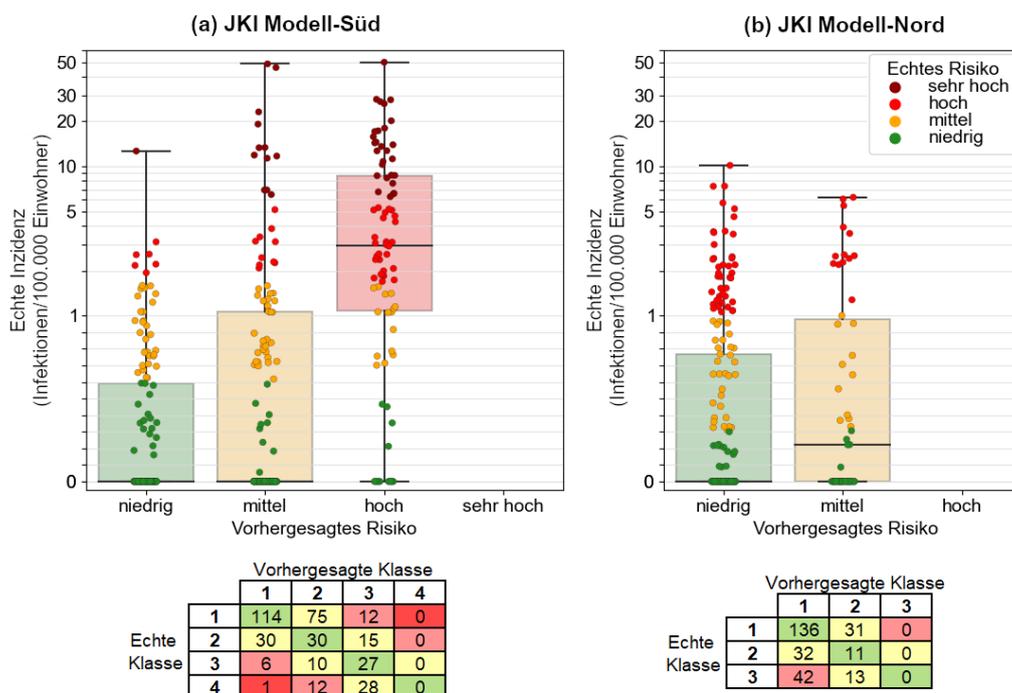
Abbildung 5 zeigt die Box-Plots für die vorhergesagten Risikoklassen gegenüber den echten Inzidenzwerten. Die vorhergesagte Risikoklasse 4 („sehr hohes Risiko“) im JKI Modell-Süd war leer, obwohl es 41 Datenpunkte mit sehr hohem echten Risiko gab (s. letzte Reihe der Wahrheitsmatrix). Die vorhergesagte Risikoklasse 3 („hohes Risiko“) im JKI Modell-Nord war leer, obwohl es 55 Datenpunkte mit hohem echten Risiko gab (s. letzte Reihe der Wahrheitsmatrix).

**Tabelle 2: Leistungskennzahlen für die Validierung der JKI Modelle im Zeitraum 2016–2020.**

Kennzahlen	Baden-Württemberg (JKI Modell-Süd)	Bayern (JKI Modell-Süd)	Niedersachsen (JKI Modell-Nord)	Nordrhein-Westfalen (JKI Modell-Nord)
Genauigkeit	45,7%	49,2%	52,9%	56,7%
Nullmodell Mittelwert	46,3%	64,9%	64,7%	62,2%
2016	71,4%	83,8%	64,7%	69,4%
2017	34,3%	24,3%	52,9%	44,4%
2018	68,6%	81,1%	47,1%	55,6%
2019	31,4%	35,1%	29,4%	33,3%
2020	22,9%	21,6%	70,6%	80,6%

**Abbildung 5: Validierung für die JKI Modelle im Zeitraum 2016–2020.**

Oberer Teil: Die Box-Plots, Unterer Teil: Die Wahrheitsmatrizen. Die Schwellenwerte der Inzidenz für die Definition der Risikoklassen waren: 0,6, 1,7 und 5,4 für das JKI Modell-Süd, bzw. 0,3 und 1 für das JKI Modell-Nord. Für eine bessere Darstellbarkeit ist die y-Achse linear im Bereich [0, 1] und logarithmiert im Bereich [1, 60]. Die Whisker für die Box-Plots waren auf 0 und 100 gesetzt, d. h., dass kein Datenpunkt in dieser Analyse als Ausreißer gekennzeichnet wurde. Die Box jeder vorhergesagten Risikoklasse enthält die mittleren 50% der Inzidenzwerte und der Strich in der Box repräsentiert den Median der Datenpunkte. Die Farbe der Box entspricht der vorhergesagten Risikoklasse. Die Farbe der Datenpunkte entspricht der echten Risikoklasse.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

Obwohl das JKI Modell-Süd eine ähnliche Genauigkeit mit dem Nullmodell für alle Jahre im Validierungszeitraum hatte –außer für das Jahr 2020, in dem das Modell sehr schlecht prognostizierte– war es dem Nullmodell überlegen, weil es der allgemeinen Tendenz der echten

Inzidenz für alle Jahre richtig folgte. Der Tukey's post-hoc Test zeigte, dass die Mittelwerte der vorhergesagten Risikoklassen 1, 2 und 3 signifikant unterschiedlich sind (alle p-Werte < 0,05).

Die Genauigkeit vom JKI Modell-Nord war gleich oder schwächer als die des Nullmodells für alle Jahre. Anders als für das JKI Modell-Süd, konnte das JKI Modell-Nord der allgemeinen Tendenz der echten Inzidenz nicht immer folgen. Der Tukey's post-hoc Test zeigte, dass die Mittelwerte der vorhergesagten Risikoklassen 1 und 2 nicht signifikant unterschiedlich sind (p-Wert=0,1446 > 0,05). Für 2016, 2017 und 2020 war das JKI Modell-Nord mit dem Nullmodell gleich, weil es nur die Risikoklasse 1 („niedriges Risiko“) ermittelte.

### 3.2.1 Extrapolation auf Hessen und Thüringen

Die zwei JKI Modelle wurden auch außerhalb der Regionen, für die sie erstellt wurden, validiert. Die Bundesländer Hessen und Thüringen wurden dafür ausgewählt, weil es hier Nachweise von PUUV Infektionen gibt. Insgesamt waren 24 Landkreise enthalten, 17 aus Hessen und 7 aus Thüringen, in denen mindestens 2 PUUV-Infektionen zwischen 2006–2015 zu verzeichnen waren. Die Extrapolation und Validierung fanden im gesamten Zeitraum 2006–2020 statt, mit insgesamt 360 Datenpunkten.

Das JKI Modell-Süd überschätzte meistens die Risikoklassen (179 richtige Vorhersagen, 155 Überschätzungen und 26 Unterschätzungen). Die mittlere jährliche Genauigkeit war 49,7%. Außer für 2010 und 2012 war die jährliche Genauigkeit schwächer als die des Nullmodells. Der Tukey's post-hoc Test zeigte, dass die Mittelwerte der vorhergesagten Risikoklassen 1 und 2 nicht signifikant unterschiedlich sind (p-Wert=0,9 > 0,05). Trotzdem sind die Mittelwerte der anderen Risikoklassen signifikant unterschiedlich (p-Wert < 0,05).

Das JKI Modell-Nord war fast symmetrisch zwischen Über- und Unterschätzungen der Risikoklasse (197 richtige Vorhersagen, 86 Überschätzungen und 77 Unterschätzungen). Die mittlere jährliche Genauigkeit war 54,7%. Außer für 2010 war die jährliche Genauigkeit schwächer als die des Nullmodells. Der Mittelwert der echten Inzidenz für die vorhergesagte Risikoklasse 1 war höher als der Mittelwert für die vorhergesagte Risikoklasse 2 (0,70 gegenüber 0,46), jedoch anhand vom Tukey's post-hoc Test nicht signifikant unterschiedlich (p-Wert=0,55 > 0,05).

Die Extrapolation der zwei Modelle auf Landkreise in Hessen und Thüringen hatte eine ähnliche Genauigkeit wie in den ursprünglichen Regionen (ähnlich gut oder ähnlich schwach). Das JKI Modell-Nord hatte hier eine bessere mittlere jährliche Genauigkeit als das JKI Modell-Süd. Trotzdem war das JKI Modell-Süd exakter für höhere Inzidenzwerte und höhere Risikoklassen. Wenn die Überschätzungen der Risikoklassen gegenüber den Überschätzungen der Risikoklassen bevorzugt werden, wäre das extrapolierte JKI Modell-Süd daher besser geeignet.

### 3.2.2 Extrapolation vom JKI Modell-Süd auf Norddeutschland

Das JKI Modell-Süd wurde für Vorhersagen in Landkreisen in Niedersachsen und Nordrhein-Westfalen verwendet. Die Genauigkeit im gesamten Zeitraum 2006–2020 war 47,0% (374 richtige Vorhersagen für 795 Datenpunkte insgesamt). Zum Vergleich war die Genauigkeit vom JKI Modell-Nord für diese Datenpunkte 55,7% (443 richtige Vorhersagen). Das Bestimmtheitsmaß für das JKI Modell-Süd in Landkreisen in Niedersachsen und Nordrhein-Westfalen lag bei -1,2.

### 3.3 Diskussion und Fazit

#### 3.3.1 BW Modell

Unsere Nachbildung des BW Modells zeigte eine Tendenz zur Überschätzung der Infektionen. Die größte Abweichung der Inzidenz erschien als Überschätzung im Jahr 2015. Nur in einigen Einzelfällen gab es Unterschätzung der Risikoklasse.

Die Ausbruchsjahre (2017 und 2019) waren im Allgemeinen richtig erkannt. Es gab keine Risikoklasse 4 („sehr hohes Risiko“) für die echten Inzidenzwerte im Validierungszeitraum. Im Modellierungszeitraum 2005–2013 wurden nur 6 Datenpunkte (für 2007 und 2012) in die Risikoklasse 4 eingeordnet. Diese Risikoklasse ist deswegen eher überflüssig.

Die BW Modelle sind sehr von den verwendeten Datensätzen abhängig. Es war nicht möglich, genau die gleichen Datensätze für Wetter-Parameter, Landbedeckung bzw. Vegetation, Buchenmast und Infektionen zu finden, oder das BW Modell nachzurechnen. Es wurde zudem festgestellt, dass die Modellierung die gesamte Anzahl an Hantavirus-Infektionen nutzte und nicht nur die labordiagnostisch nachgewiesenen PUUV-Infektionen, die in dieser Studie zu Verfügung standen und verwendet wurden. Die labordiagnostisch nachgewiesenen PUUV-Infektionen machten in Baden-Württemberg in 2014–2020 rund 60%–80% der gesamten Hantavirus-Infektionen aus.

Die Validierung unserer Nachbildung des BW Modells konnte nur so gut erfolgen, wie nah unsere Datensammlung an die Originaldatensätze heranreichte. Eine endgültige Aussage für die Qualität und die Güte der BW Modelle ist deshalb mit Vorsicht zu betrachten. Nach 2021 basiert die Version des BW Modells auf der Internetseite [Hanta-Vorhersage](#) auf einem neuen Ansatz, für den keine relevante Veröffentlichung in der Literatur gefunden wurde. Das Modell zeigte eine hohe Vorhersagekraft im Jahr 2021. Weder für 2022 noch für 2023 (Stand: 24.03.2023) lag auf der Internetseite [Hanta-Vorhersage](#) eine Prognose vor. Es ist unklar, ob die Vorhersagen in BW zukünftig weitergeführt werden.

#### 3.3.2 JKI Modelle

Das JKI Modell-Süd hatte ein positives Bestimmtheitsmaß von 0,321 und eine angemessene Genauigkeit von 47,5%, gegenüber der Genauigkeit von 55,8% des Nullmodells. In den Jahren 2016 und 2018 (nicht-Ausbruchsjahre) war das JKI Modell-Süd fast gleich dem Null-Modell. Allerdings bewies es Vorhersagekraft und detektierte beide Ausbruchsjahre im Validierungszeitraum (2017 und 2019), auch wenn seine regionalen Vorhersagen nicht immer genau waren. Insgesamt war das JKI Modell-Süd dem Nullmodell überlegen und es hatte nur für 2020 eine deutlich geringere Genauigkeit.

Obwohl das JKI Modell-Nord eine relativ gute Genauigkeit von 55,5% hatte, war sein Bestimmtheitsmaß um null. Folglich ist das JKI Modell-Nord ungeeignet für Vorhersagen. Seine gute Genauigkeit erklärt sich durch die geringen Inzidenzwerte im Norden. Zum Vergleich lag die mittlere Genauigkeit des Nullmodells bei 63,0%.

Nach Betrachtung ausgewählter 3D-Plots des Raums der Modellparameter wurde es festgestellt, dass einige Vergleiche, die an den Knoten der Entscheidungsbäume stattfinden, unzureichend waren, z. B. der Knoten mit der maximalen Temperatur im August aus dem Vorvorjahr im JKI Modell-Süd, oder die Kombination der zwei ersten Knoten mit der maximalen Temperatur im April aus dem Vorjahr und dem Niederschlag im Januar aus dem Vorjahr im JKI Modell-Nord. Andererseits funktionieren manche Vergleiche sehr gut, z. B. der erste Knoten mit der Sonnenscheindauer im aktuellen März im JKI Modell-Süd, oder der letzte Knoten mit der maximalen Temperatur im August aus dem Vorvorjahr im JKI Modell-Süd aber nur für Hessen

und Thüringen, weil er unzureichend für Baden-Württemberg und Bayern war. Das ist zu erwarten, weil die Entscheidungsbäume die Ergebnisse des Modellierungszeitraums anschaulich wiedergeben. Trotzdem ist es unklar, wie angemessen sie außerhalb des Modellierungszeitraums sind, d. h. im Validierungszeitraum.

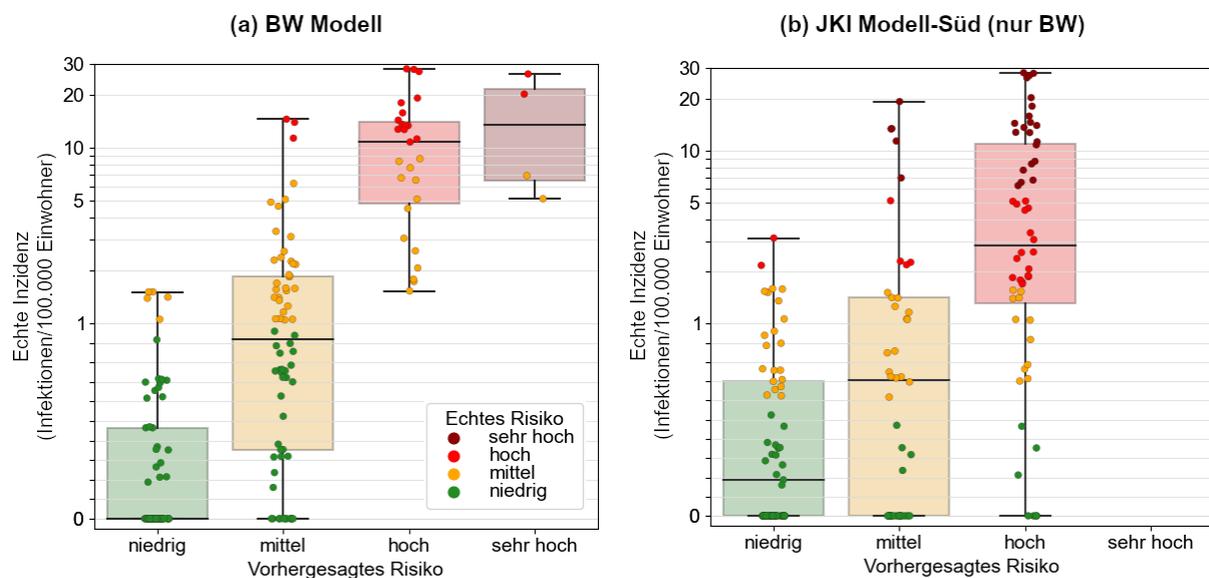
Im Allgemeinen funktioniert die Extrapolation der JKI Modelle auf andere Regionen nicht gut.

### 3.3.3 Vergleich

Die Vorhersagen vom BW Modell konnten mit den Vorhersagen vom JKI Modell-Süd für Baden-Württemberg verglichen werden. Beide Modelle bestehen aus vier Risikoklassen, die aber unterschiedlich definiert waren. Die Landkreise Lörrach und Waldshut waren aus unbekanntem Gründen im BW Modell nicht enthalten, deswegen wurden sie auch vom JKI Modell-Süd entfernt. Im überlappenden Zeitraum 2016–2020 (165 Datenpunkte insgesamt) war die Genauigkeit der Risikoklassen vom BW Modell (63%–104 richtige Vorhersagen) besser als vom JKI Modell-Süd (46%–76 richtige Vorhersagen), siehe Abbildung 6. In diesem Zeitraum lag der Pearson-Korrelationskoeffizient für das BW Modell bei 0,66 für die Inzidenzwerte und bei 0,71 für die Risikoklassen. Für das JKI Modell-Süd lag der Pearson-Korrelationskoeffizient bei 0,59 für die Risikoklassen. Beide JKI Modelle ermitteln im Prinzip direkt Risikoklassen aber keine Inzidenzwerte.

#### Abbildung 6: Vergleich der zwei Modellansätze im Zeitraum 2016–2020.

Die Box-Plots für die vorhergesagten Risikoklassen der Landkreise Baden-Württembergs von 2016 bis 2020 gegenüber den echten Inzidenzwerten. Die Risikoklasse-Definition unterscheidet sich zwischen den beiden Modellen (s. Abschnitte 3.1 und 3.2).



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

In Baden-Württemberg war die Leistung des BW Modells bei jeder Metrik besser. Das JKI Modell-Süd war zuzüglich der Daten aus Bayern zustande gekommen und es wurde erwartet, dass es eine geringere Leistung im Vergleich zu einem BW-spezifischen Modell aufweisen würde. Andererseits kann das BW Modell nicht für andere Landkreise direkt verwendet werden, obwohl das Prinzip des generalisierten linearen Modells selbst angewendet werden kann.

### 3.3.4 Fazit

Aufgrund der Ungenauigkeit der JKI Modelle im Validierungszeitraum und der extremen Spezialisierung des BW Modells auf ein einzelnes Bundesland war es nötig, ein neues Vorhersagemodell zu generieren, das über alle PUUV-hochendemischen Gebieten Deutschlands einheitlich ist. Die Entwicklung dieses Modells wird ausführlich in Kapitel 4 beschrieben. Seine Grundlage basiert auf Erkenntnissen der Validierungen der beiden bisherigen Modellansätze. Das BW Modell enthielt einen kreisbezogenen Skalierungsfaktor, der Unterschiede in der lokalen Größenordnung der PUUV-Inzidenz berücksichtigte. Unsere Methode baut auf diesem Gedanken auf und erweitert ihn, indem sie die Berechnung der Skalierungsfaktoren vom Vorhersagemodell selbst entkoppelt und sie mit der in den JKI Modellen angewandten Inzidenztransformation kombiniert. Andererseits verwendeten die JKI Modelle Wetterdaten mit höherer räumlicher Auflösung (auf Kreisebene, im Gegensatz zum BW Modell, welches Wetterparameter auf Bundeslandebene verwendete) und intrinsisch definierte Schwellenwerte für die Risikoklassen. Schließlich demonstrierten die von den JKI Modellen verwendeten Entscheidungsbäume die Vorteile einer schematischen Darstellung des Modell-Targets und inspirierten unsere Implementierung eines geometrisch motivierten Klassifikationsalgorithmus.

## 4 Entwicklung und Validierung eines neuen, bundesweiten Prognosemodells

### 4.1 Entwicklung

Die Struktur des Datensatzes ist in Abschnitt 2.3 definiert. Obwohl die Hantavirus-Erkrankung in Deutschland 2001 meldepflichtig wurde, wurde die Analyse auf PUUV-Infektionen seit 2006 beschränkt, um Verzerrungen durch Untererfassung zu vermeiden. Die Jahre 2006–2021 enthalten 91,9% der gesamten gemeldeten PUUV-Fälle von 2001 bis 2021. Die humane PUUV-Inzidenz wurde als Anzahl der Infektionen pro 100.000 Einwohner unter der Verwendung von Bevölkerungsdaten von Eurostat berechnet (s. Abschnitt 1.1).

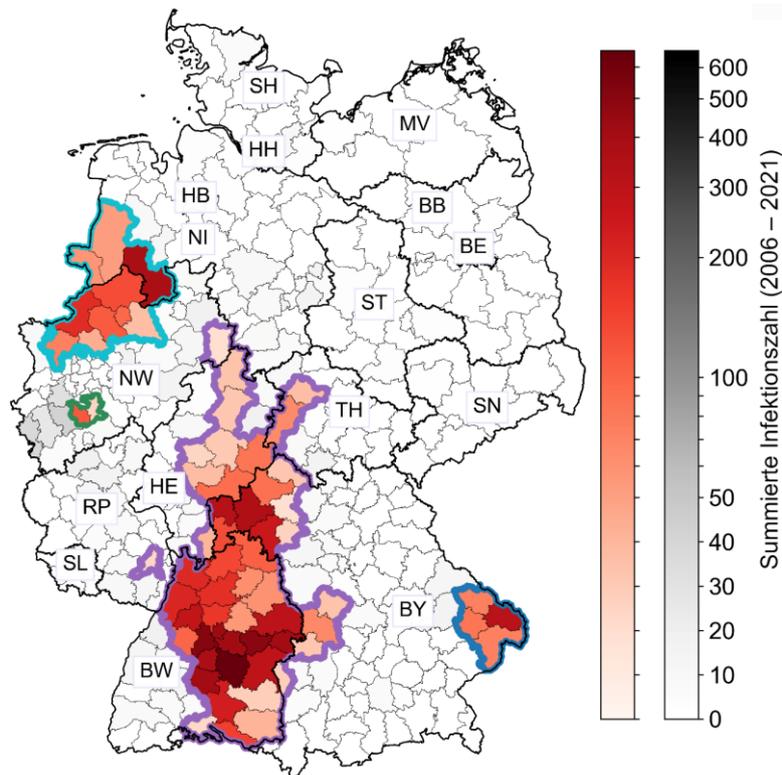
#### 4.1.1 Auswahl der Kreise

Die PUUV-Infektionen werden entsprechend des Wohnorts der infizierten Person gemeldet, der vom Infektionsort/-kreis abweichen kann (Abschnitt 2.2). Um die Landkreise auszuwählen, in denen das PUUV tatsächlich endemisch ist, und Einzelfälle aus nicht endemischen Regionen auszuschließen, haben wir in die Analyse nur Landkreise einbezogen, in denen die Gesamtinfektionen  $\geq 20$  und die höchste Jahresinzidenz  $\geq 2$  im Zeitraum 2006–2021 waren.

Insgesamt wurden 66 Landkreise anhand dieser Kriterien ausgewählt (Abbildung 7 und Tabelle 4 im Anhang). Auf die ausgewählten Kreise entfielen 10.090 humane PUUV-Infektionen, 89,9% der landesweit gemeldeten Gesamtinfektionen in diesem Zeitraum. Die Kreise wurden in vier Cluster gruppiert, die mit dem hypothetischen Rand des Verbreitungsgebiets der westlichen Rötelnmaus-Evolutionslinie (Drewes et al. 2017) übereinstimmten und alle bisher entdeckten genetisch unterschiedlichen PUUV-Kladen (Ettinger et al. 2012; Faber et al. 2019; Hofmann et al. 2022) umfassten. Das erste Cluster (Abbildung 7, blauer Umriss) umfasste 10 Landkreise zwischen Niedersachsen und Nordrhein-Westfalen, an der Grenze zu den Niederlanden, und entsprach der Klade Münsterland/Teutoburger Wald. Das zweite Cluster (Abbildung 7, grüner Umriss) umfasste 2 Kreise im Südwesten Nordrhein-Westfalens, beide aus der Klade Rheinland. Das größte Cluster (Abbildung 7, violetter Umriss) umfasste 50 Landkreise in einem zentralen vertikalen Korridor mit einer Länge von etwa 450km und einer maximalen Breite von etwa 200km, der sich von Hessen und Thüringen über Bayern bis in den Süden Baden-Württembergs an der Grenze zur Schweiz erstreckte. Dieses Cluster umfasste die PUUV-Kladen Spessart/Nordhessen, Schwäbische Alb und Thüringer Wald. Das letzte Cluster (Abbildung 7, blauer Umriss) umfasste 4 Landkreise in Ostbayern an der Grenze zur Tschechischen Republik, mit der PUUV-Molekularklade Bayerischer Wald.

**Abbildung 7: Auswahl der Landkreise für die Analyse.**

Die 66 landesweit einbezogenen Landkreise sind in einem roten Verlauf abhängig von ihrer summierten Anzahl an PUUV-Infektionen in den Jahren 2006–2021 dargestellt. Für eine bessere Darstellbarkeit ist der Farbbalken linear im Bereich [0, 50] und logarithmiert im Bereich [50, 650]. 26 Landkreise sind aus Baden-Württemberg (BW), 16 aus Bayern (BY), 8 aus Hessen (HE), 3 aus Niedersachsen (NI), 10 aus Nordrhein-Westfalen (NW), 1 aus Rheinland-Pfalz (RP) und 2 aus Thüringen (TH). Dicke schwarze Linien trennen die Bundesländer; dicke farbige Linien trennen die vier Cluster der detektierten PUUV-Kladen, wie im Text beschrieben. Weitere Landkreise werden in grauen Farbverläufen mit der gleichen Farbbalkenskalierung dargestellt.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Datenquelle: (Robert Koch-Institut). Hintergrundkarte: (GeoBasis-DE / BKG 2017).

**4.1.2 Aufbereitung der Daten****4.1.2.1 Transformation der Inzidenz und Ausbruchsrisiko**

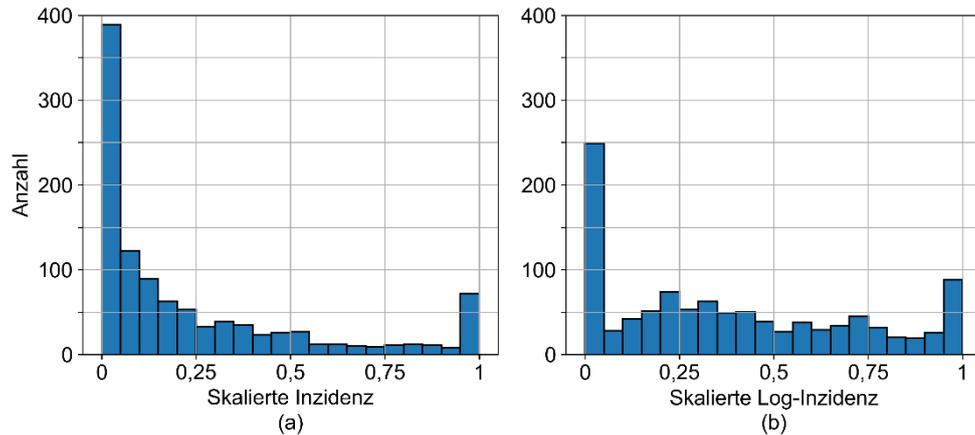
Wir haben zunächst eine Log-Transformation auf die Inzidenzwerte (Esther et al. 2014) angewendet, um den Einfluss von Nicht-Null-Werten zu erhöhen:

$$\text{Log - Inzidenz} = \log_{10}(\text{Inzidenz} + 1)$$

Die Inzidenz variierte räumlich stark zwischen den Kreisen. Um die Auswirkungen zu berücksichtigen, die das Auftreten einer hohen kreisbezogenen Inzidenz verursachen, haben wir die Inzidenz auf Kreisebene diskretisiert. Die auf ihren Maximalwert skalierte Inzidenz für jeden Landkreis zeigte Extremwerte für Minima und Maxima. Etwa 48% aller Datenpunkte lagen im Bereich [0, 0,1) und 8% im Bereich [0,9, 1] (Abbildung 8). Daher haben wir uns spezifisch dafür entschieden, die skalierte Inzidenz mit zwei Klassen (Bins) zu diskretisieren, d. h. sie zu binarisieren.

**Abbildung 8: Histogramme der skalierten jährlichen PUUV-Inzidenz von 2006 bis 2021.**

Die kreisbezogene Inzidenz skaliert auf ihren Maximalwert für jeden der zusammengefassten Landkreise. (a) Die rohe Inzidenz mit hoher positiver Schiefe. (b) Die skalierte Log-Inzidenz mit reduzierter Schiefe, d. h., dass niedrige Inzidenzwerte zu höheren Werten gespreizt werden, was zu gleichmäßigeren Bin-Höhen im Bereich [0,05, 0,95] nach der Transformation führt. Das Histogramm der skalierten Log-Inzidenz zeigt zwei deutliche Spitzen für 0 (Minimalwert im Kreis) und 1 (Maximalwert im Kreis).



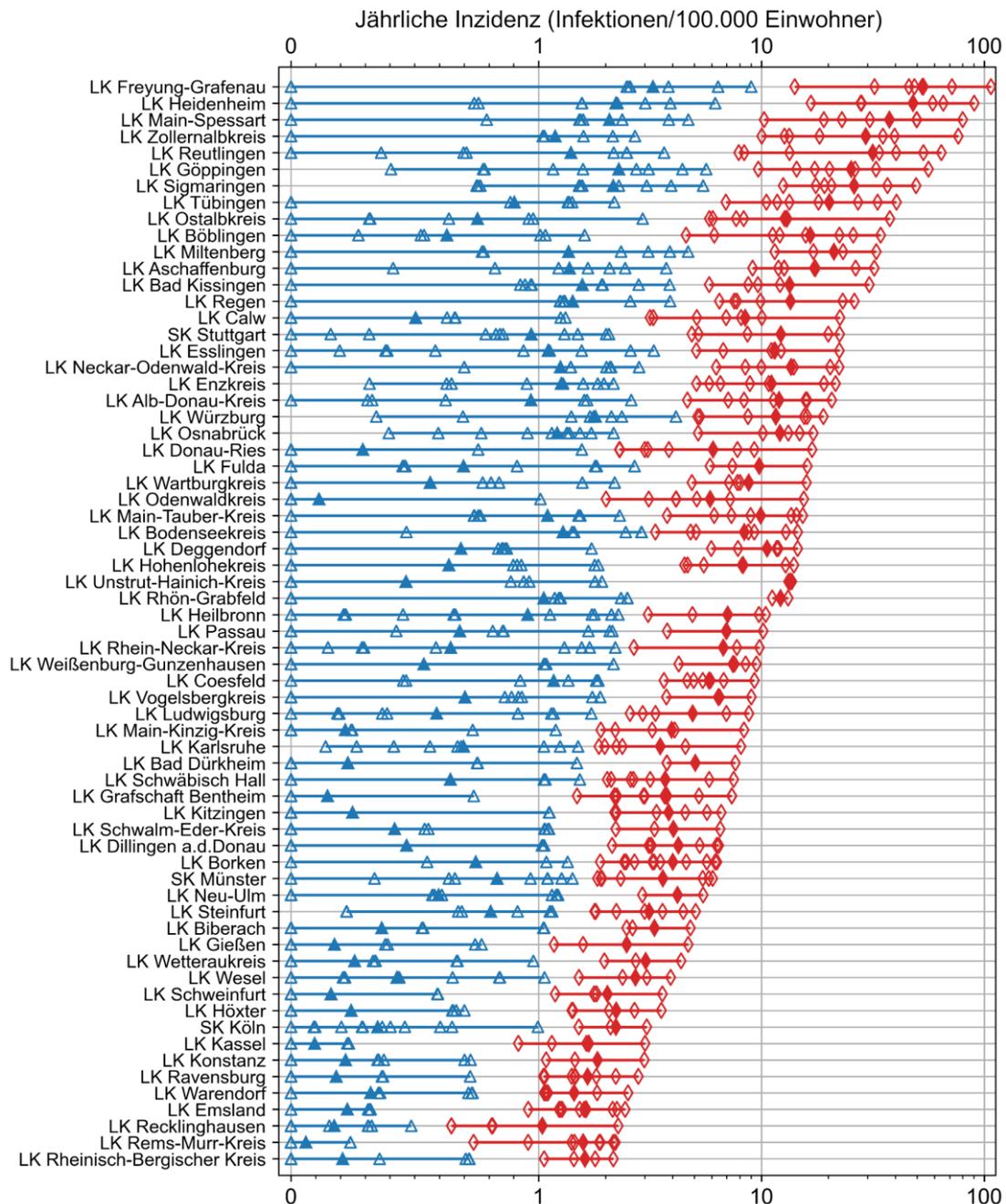
Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Datenquelle: SurvStat (Robert Koch-Institut).

Für die Binarisierung haben wir eine nicht-überwachte Clusteranalyse der Log-Inzidenz durchgeführt, getrennt für jeden Kreis, unter Verwendung der Funktion *KBinsDiscretizer* des Pakets *scikit-learn* (Pedregosa et al. 2011). Unsere gewählte Strategie war das k-Means-Clustering mit zwei Klassen, da es keinen vordefinierten Schwellenwert erfordert und mit der gleichen festen Anzahl von Klassen für jeden Kreis arbeiten kann, indem die Clusterschwerpunkte automatisch entsprechend angepasst werden. Die resultierenden zwei Klassen wurden als „geringes Risiko“ (0) und „hohes Risiko“ (1) bezeichnet. Ein lokaler „Ausbruch“ trat dann auf, wenn in einem Jahr die Inzidenz eines Kreises in die Hochinzidenzklasse der gemeldeten Werte eingeordnet wurde. Die Inzidenz in Kreisen in Jahren ohne Ausbruch war null oder wurde im Vergleich zu den gemeldeten Werten in diesem speziellen Kreis als niedrig betrachtet. Von den insgesamt 1.056 Datenpunkten (16 Jahre × 66 Kreise) wurden 682 als geringes Risiko (65%) und 374 als hohes Risiko (35%) eingestuft. In Datenpunkten der Ausbruchsklasse mit hohem Risiko wurden insgesamt 8.779 Infektionen registriert, 87% der Gesamtinfektionen in dieser Analyse (oder 78% der Gesamtinfektionen in Deutschland von 2006 bis 2021).

Diese binäre Klassifikation, die wir „Ausbruchsrisiko“ nennen, wurde als Zielgröße für unser Modell verwendet. Da das Ausbruchsrisiko kreisbezogen erfolgte, konnte je nach Kreis derselbe Inzidenzwert als geringes Risiko oder hohes Risiko zugeordnet werden (Abbildung 9). Die Transformation kompensiert die hohe räumliche Inhomogenität der Inzidenz und wandelte ihre Vorhersage in ein Extrema-Detektionsproblem um.

**Abbildung 9: Die jährlichen Inzidenzwerte in den einbezogenen Landkreisen von 2006 bis 2021.**

Die 66 Landkreise sind nach der höchsten Jahresinzidenz angeordnet. Die Ausbruchsklasse mit geringem Risiko wird durch blaue Dreiecke (auf der linken Seite des Diagramms) angezeigt. Die Ausbruchsklasse mit hohem Risiko („Ausbruch“) wird durch rote Rauten (auf der rechten Seite des Diagramms) angezeigt. Die gefüllten Dreiecke und Rauten geben den Durchschnittswert für jede Ausbruchsklasse an. Die durchgezogenen Linien markieren den Bereich jeder Ausbruchsklasse. Der Weißraum zwischen der blauen und der roten Linie zeigt die Trennung zwischen den beiden Ausbruchsklassen für jeden Kreis an. Für eine bessere Darstellbarkeit ist die x-Achse linear im Bereich [0, 1] und logarithmiert im Bereich [1, 110].



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Datenquelle: (Robert Koch-Institut).

#### 4.1.2.2 Prädiktoren

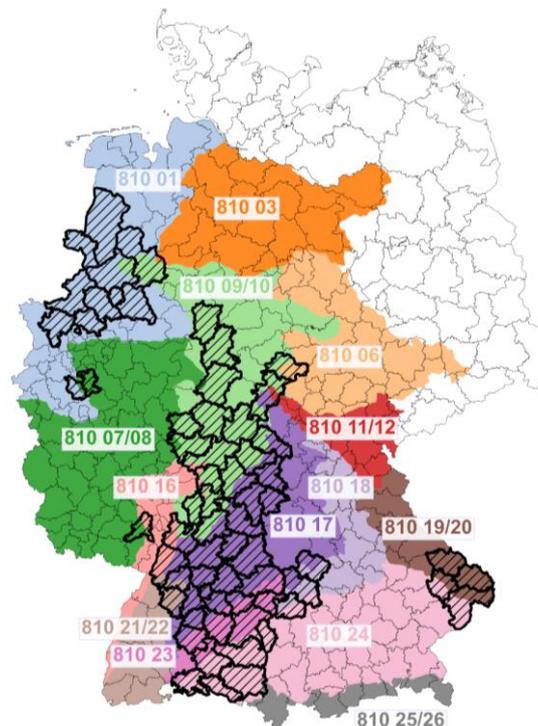
Aus jedem monatlichen Wetterparameter wurden 24 Einträge für alle Monate der beiden vorangegangenen Jahre als mögliche Prädiktoren für das Modell erstellt. Insgesamt gab es 168 verfügbare Prädiktoren aus den Wetterparametern (7 Parameter x 24 Monate), im Folgenden

„Variablen“ genannt. Die Nomenklatur für die Benennung der Variablen folgt dem Format  $Vx_{<Parameter>_{<Monat>}$ , wobei „Vx“ V1 oder V2 sein kann, was jeweils dem Vor- bzw. Vorvorjahr entspricht; <Parameter> ist die Abkürzung des Wetterparameters (siehe Abschnitt 2.1); und <Monat> ist der Zahlenwert des Monats, d. h. von 1 (Januar) bis 12 (Dezember).

Weitere mögliche Prädiktoren waren die Blühstärke-Werte der Rotbuche aus dem Vorjahr und dem Vorvorjahr. Bucheckern sind eine Hauptnahrungsquelle für Rötelmäuse und ihre Verfügbarkeit wurde mit dem Wachstum der Rötelmauspopulationen in Mitteleuropa in Verbindung gebracht (Clement et al. 2010; Reil et al. 2015). Die Buchensamenproduktion ist üblicherweise großflächig synchronisiert (Schauber et al. 2002), was zu sogenannten Mastjahren führt, die mit Wetterbedingungen in Verbindung gebracht wurden (resource matching hypothesis) (Sork 1993; Piovesan und Adams 2001). Die Wetterbedingungen allein können jedoch nicht die gesamte Mastvariabilität erklären (Koenig und Knops 2000) und es wurden andere Selektionshypothesen vorgeschlagen, wie z. B. die Windbestäubungseffizienz und die Räubersättigung (Kelly und Sork 2002). Unabhängig davon wird erwartet, dass die Einbeziehung der Mast die Leistung unseres Prognosemodells verbessert, indem Effekte einbezogen werden, die nicht direkt vom Wetter angetrieben sind. Die Blühstärke im späten Frühjahr und Frühsommer ist ein guter Indikator für die Mast, obwohl eine erhöhte Blüte nicht immer zu einer erhöhten Samenproduktion führt (Konnert et al. 2014). Abbildung 10 zeigt die Herkunftsgebiete (HKG) der Rotbuche, die alle einbezogenen Landkreise enthalten.

#### Abbildung 10: Karte der für diese Analyse relevanten Herkunftsgebiete (HKG) der Rotbuche.

Die in die Analyse einbezogenen Landkreise sind mit dicken Rändern und schwarzen Schraffur zu sehen. Die Blühstärke-Werte für die kombinierten HKG (07. mit 08., 09. mit 10., 11. mit 12., 19. mit 20., 21. mit 22., und 25. mit 26.) entsprechen den Mittelwerten der einzelnen HKG. Nicht abgebildet sind die HKG 02, 04, 05 und 13–15.

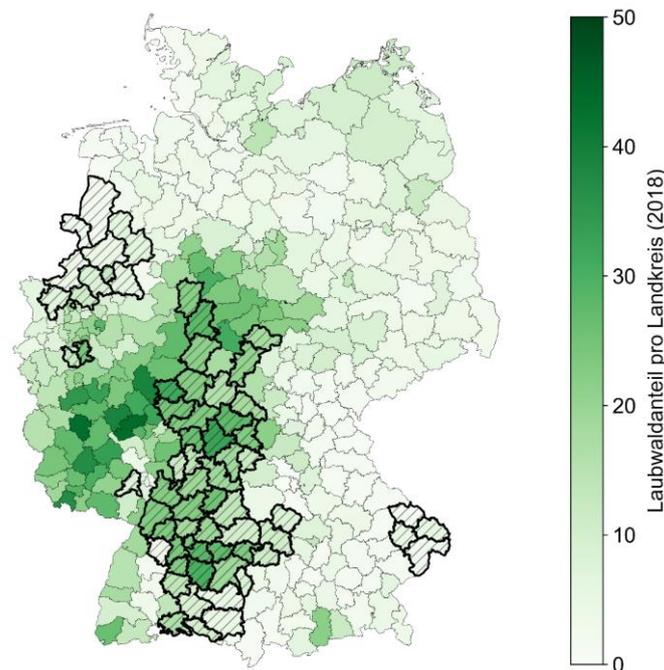


Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Hintergrundkarten: (GeoBasis-DE / BKG 2017) und (Bundesanstalt für Landwirtschaft und Ernährung 2023b).

Der letzte mögliche Prädiktor in unserem Datensatz war der Laubwaldanteil, der in der Vergangenheit als Prädiktor für das PUUV-Infektionsrisiko in Süddeutschland verwendet wurde (siehe Abschnitt 3.2). Abbildung 11 zeigt den Laubwaldanteil pro Landkreis landesweit. Beim Betrachten dieser Karte wird deutlich, dass der Laubwaldanteil unserem Modell wahrscheinlich keinen Mehrwert verleihen kann. Insbesondere die in Abbildung 7 definierten PUUV-Cluster 1 und 4 haben im Vergleich zu den anderen Regionen einen sehr niedrigen mittleren Laubwaldanteil von 5,9% bzw. 3,1%, obwohl sie Kreise mit hohen Inzidenzwerten umfassen, wie LK Freyung-Grafenau (BY, mit der höchsten jemals erfassten Jahresinzidenz von 107 im Jahr 2021, siehe Abbildung 9), LK Regen (BY) und LK Osnabrück (NI).

#### Abbildung 11: Choroplethenkarte des Laubwaldanteils im Jahr 2018 pro Landkreis.

Der Laubwaldanteil wurde anhand des Referenzdatensatzes CLC2018 berechnet. Die in die Analyse einbezogenen Kreise sind mit dicken Rändern und schwarzen Schraffur zu sehen. Der höchste Wert war 32,7% für den LK Main-Spessart (BY) und der geringste Wert 0,7% für den LK Regen (BY). In ganz Deutschland waren die höchsten Werte 43,7% für den LK Rheingau-Taunus-Kreis (HE) und 43,4% für den LK Cochem-Zell (RP) und der geringste Wert 0,3% für den LK Lindau (BY).



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Datenquelle: (European Environment Agency) Hintergrundkarte: Kreisgrenzen (GeoBasis-DE / BKG 2017).

#### 4.1.2.3 Risikoklassen

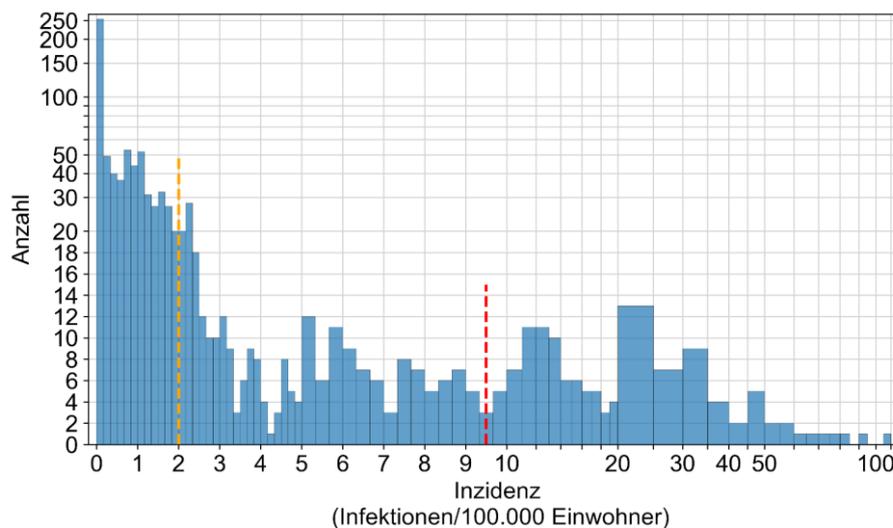
Die Risikoklassen wurden nicht direkt vom Modell benötigt und nur im letzten Schritt angewendet, nachdem die vorhergesagten Inzidenz-Werte generiert wurden. Das Warnsystem sollte jedoch Inzidenz-basierte Risikoklassen implementieren. Daher mussten sie definiert und verwendet werden, um die Leistung des vollständigen Modells zu bewerten.

Wir haben drei Risikoklassen ausgewählt, eine Zahl, die nach der Validierung der bisherigen Modellansätze (Kapitel 2) sinnvoll erschien. Es gab keine exakten Vorgaben für die Schwellenwerte der Risikoklassen. Daher wurden sie durch Begutachtung des Histogramms der Inzidenz für alle Datenpunkte im kompletten Trainingsdatensatz 2006–2021 (Abbildung 12) definiert. Der erste Schwellenwert von 2 begrenzt die Klasse mit niedrigem Risiko gegenüber der Klasse mit mittlerem Risiko. Obwohl der leichte Abfall im Histogramm kein eindeutiger Hinweis ist, stellt dieser Wert einen Kompromiss dar: ein kleinerer Wert wäre wie in der

Vergangenheit für den Fall von vier Risikoklassen vorzuziehen, während ein höherer Wert keine klare Unterscheidung der geringen Inzidenz erlauben würde. Der zweite Schwellenwert, der die Klassen mittleren und hohen Risikos trennt, wurde auf 9,5 festgelegt, wo ebenfalls ein leichter Abfall zu erkennen ist. Sollte eine zusätzliche Risikoklasse gewünscht werden, erscheint ein Wert um 4 als eine sinnvolle Auswahl.

### Abbildung 12: Histogramm der jährlichen PUUV-Inzidenz von 2006 bis 2021.

Dargestellt ist die kreisbezogene Inzidenz für alle einbezogenen Landkreise. Die farbigen gestrichelten senkrechten Linien bezeichnen die Schwellenwerte der Inzidenz für die Definition der Risikoklassen: 2 und 9,5. Die erste Risikoklasse mit Inzidenz im Bereich  $[0, 2)$  enthält 667 Datenpunkte (63% der gesamten 1056), die zweite Risikoklasse mit Inzidenz im Bereich  $[2, 9,5)$  enthält 263 Datenpunkte (25%) und die dritte Risikoklasse mit Inzidenz  $\geq 9,5$  enthält 126 Datenpunkte (12%). Für eine bessere Darstellbarkeit ist die Bin-Breite variabel:  $1/6$  im Bereich  $[0, 5)$ ,  $1/3$  im Bereich  $[5, 10)$ , 1 im Bereich  $[10, 20)$  und 5 im Bereich  $[20, 110]$ . Aus dem gleichen Grund ist die y-Achse linear im Bereich  $[0, 20]$  und logarithmiert im Bereich  $[20, 260]$ , und die x-Achse linear im Bereich  $[0, 10]$  und logarithmiert im Bereich  $[10, 110]$ .



Im Gegensatz zur Abbildung 8, zeigt diese Abbildung die unskalierte Inzidenz.

Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Datenquelle: (Robert Koch-Institut).

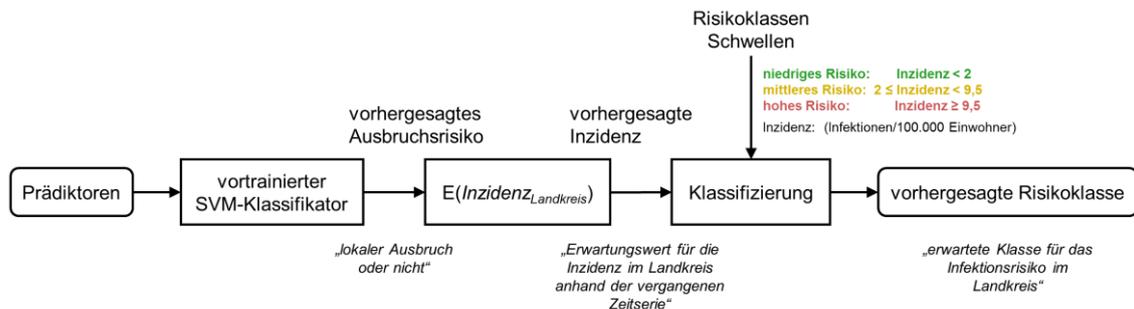
#### 4.1.3 Klassifizierungsmethode und Inzidenz-Vorhersage

Wir haben uns nur auf diejenigen Variablenkombinationen konzentriert, die zu einer linearen Entscheidungsgrenze (Hyperebene) für die Einstufung unseres gewählten Targets („Ausbruchsrisiko“) führten. Zur Unterscheidung der Ausbrüche verwendeten wir einen Support-Vector-Machine-Klassifikator (SVC) (Cortes und Vapnik 1995) mit einem linearen Kernel, da er eine hohe Leistung mit einer geringen Modellkomplexität kombiniert, indem er die Entscheidungsgrenze als lineare Gleichung der Eingaben zurückgibt. Darüber hinaus wurde erwartet, dass SVM weniger empfindlich gegen Ausreißer und Überanpassung ist als andere Klassifikationsalgorithmen für maschinelles Lernen wie z. B. die logistische Regression. Für den gesamten Modellierungsprozess wurde der Regularisierungsparameter C auf 1 gesetzt, das ist der Standardwert in der angewandten Methode SVC des Pakets scikit-learn (Pedregosa et al. 2011) und die Gewichte für beide Klassen des Ausbruchsrisikos wurden ebenfalls auf 1 gesetzt.

Die Ausgabe vom SVC stellt das erwartete Ausbruchsrisiko in jedem Landkreis dar: eine 0 (kein lokaler Ausbruch erwartet) oder eine 1 (lokaler Ausbruch erwartet). Sie kann in die Inzidenz umgerechnet werden, indem der Inzidenzmittelwert aus allen Datenpunkten des jeweiligen Kreises im Trainingsdatensatz zugewiesen wird, die das gleiche Ausbruchsrisiko anzeigten, entweder 0 oder 1. Das bedeutet, dass jeder Kreis im Wesentlichen zwei mögliche Werte für die

vorhergesagte Inzidenz hat: die mittlere Inzidenz aller Ausbruchsjahre (falls das erwartete Ausbruchsrisiko gleich 1 ist) oder die mittlere Inzidenz der übrigen Jahre (falls das erwartete Ausbruchsrisiko gleich 0 ist). Anschließend ermittelt eine einfache Klassifizierung der vorhergesagten Inzidenz anhand der vordefinierten Schwellenwerte die vorhergesagte Risikoklasse.

**Abbildung 13: Schematische Darstellung der Vorhersage-Methode, nach dem Training des SVM-Klassifikators.**



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

#### 4.1.4 Feature Selection

Die Leistung wurde anhand zweier Verfahren bewertet:

- ▶ die Aussagekraft bewertete die Modellleistung an dem Trainingsdatensatz und
- ▶ die Vorhersagekraft schätzte die Methodenleistung an einem Testdatensatz.

Die Auswertung der Vorhersagekraft erfolgte durch eine Leave-One-Year-Out-Kreuzvalidierung (LOOCV), die mit dem Pseudocode in der Abbildung 14 beschrieben ist. Ein einzelnes Jahr wurde aus dem Trainingsdatensatz (2006–2021) entfernt, das vollständige Modell wurde anhand der verbleibenden Jahre trainiert, und dieses trainierte Modell wurde verwendet, um die Inzidenzwerte für das entfernte Jahr vorherzusagen. Dies wurde für alle sechzehn Jahre im Trainingsdatensatz wiederholt. Abschließend wurde die Prognose für alle Jahre anhand einer ausgewählten Metrik ausgewertet.

**Abbildung 14: Pseudocode für die Auswertung der Vorhersagekraft.**

Die Methode ist eine modifizierte Leave-One-Out-Kreuzvalidierung (LOOCV), wobei ein komplettes Jahr als Testdatensatz herausgelassen wird, anstatt eines einzelnen Datenpunkts. Deswegen nennen wir die Methode „Leave-One-Year-Out-Kreuzvalidierung“.

- 1: **FOR** jedes *Jahr* im Trainingsdatensatz **DO**
- 2:     Datenpunkte aus dem *Jahr* vom Trainingsdatensatz entfernen
- 3:     Das SVM-Modell mit dem verbleibenden Trainingsdatensatz trainieren
- 4:     Prognose für die Datenpunkte aus dem *Jahr* erzeugen
- 5:     Prognosen für alle Jahre auswerten

Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

Insgesamt enthielt unser ursprünglicher Datensatz 171 mögliche Prädiktoren (Pool von Prädiktoren: 168 Wettervariablen, die Blühstärken aus dem Vorjahr und dem Vorvorjahr, und der Laubwaldanteil), die zu insgesamt  $2^{171} \approx 3 \cdot 10^{51}$  möglichen Teilmengen unterschiedlicher

Anzahlen von Prädiktoren (Mächtigkeiten) kombiniert werden. Es reichte jedoch aus, die kleinste Anzahl von Prädiktoren für ein Klassifizierungsmodell mit genügender Leistung zu finden, da die Einschränkung der Anzahl von Prädiktoren für unser endgültiges Modell auch die Gefahr einer Überanpassung verringert. Daher haben wir ein heuristisches Verfahren zum Durchsuchen des mehrdimensionalen Variablenraums etabliert, das im Folgenden beschrieben wird.

Wir wandten zuerst einen SVC mit einem linearen Kernel für alle Teilmengen mit einer einzigen Variablen an. Die Variablen wurden auf ihre Minimal- und Maximalwerte skaliert, was die Verarbeitungszeit erheblich verkürzte.

Die Leistung des Modells mit der Blühstärke aus dem Vorjahr war besser als die des Modells mit der Blühstärke aus dem Vorvorjahr (Genauigkeit 79,6% ggb. 73,2%, dies entspricht 68 zusätzlichen korrekt klassifizierten Datenpunkte). Da die beiden Variablen eine starke negative Korrelation aufwiesen (-0,72 Pearson-Korrelationskoeffizient mit einem p-Wert  $\ll 0,001$ ), wurde die Blühstärke aus dem Vorvorjahr aus unserem Pool von Prädiktoren entfernt und nur die Blühstärke aus dem Vorjahr beibehalten.

Das Modell, welches nur den Laubwaldanteil berücksichtigte, hatte keine Aussagekraft, weil es ein immer zu geringes Ausbruchsrisiko ermittelte. Darüber hinaus erhöhte die Einbeziehung des Laubwaldanteils in die Modelle mit einzelnen Wettervariablen die Genauigkeit nur unwesentlich, mit maximal 12 zusätzlichen korrekten Klassifizierungen (+1,1% Genauigkeit). Zum Vergleich erhöhte die Einbeziehung der Blühstärke aus dem Vorjahr die Genauigkeit um mindestens 77 richtige Klassifikationen (+7,3% Genauigkeit). Da der Laubwaldanteil die Modelleleistung nicht signifikant zu verbessern schien, wurde er aus unserem Pool von Prädiktoren entfernt. Das bedeutet nicht, dass er das PUUV-Infektionsrisiko nicht beeinflusst, sondern dass er für unsere spezifische Problemdefinition nicht geeignet war, eine Vorhersage zu verbessern. Ein weiterer Hinweis ist, dass der Laubwaldanteil nicht mit dem Ausbruchsrisiko linear korrelierte (-0,06 Pearson-Korrelationskoeffizient mit einem p-Wert = 0,05, verglichen mit dem Koeffizienten von 0,59 mit einem p-Wert  $\ll 0,001$  für die Blühstärke aus dem Vorjahr). Wir gehen davon aus, dass die Effekte der Landbedeckungsdaten in die kreisbezogene Inzidenztransformation einfließen, d. h., dass der Laubwaldanteil zwar auf die lokale Größenordnung der PUUV-Inzidenz einwirken kann, aber nicht die Wahrscheinlichkeit eines Ausbruchs selbst beeinflusst.

Für die weitere Suche konzentrierten wir uns auf Teilmengen mit wenigen Prädiktoren. Dennoch war eine erschöpfende (vollständige) Suche und die Auswertung der Vorhersagekraft (wie in der Abbildung 14 definiert) für jede mögliche Teilmenge extrem zeitaufwändig und nicht praktikabel. Daher wählten wir die Teilmengen mit der besten Aussagekraft für jede Anzahl von Prädiktoren, die eine um Größenordnungen schnellere Berechnung ermöglichte.

Wir wandten einen SVC mit einem linearen Kern für alle 2-Variablen-Kombinationen aller 168 möglichen Wettervariablen an. Die Blühstärke aus dem Vorjahr wurde separat betrachtet, um ihren Einfluss quantitativ zu prüfen. Von den insgesamt 14.028 Modellen für jedes eindeutige Paar wurden die 100 Modelle mit dem besten F1-Maß behalten und das Vorkommen jeder Jahr-Monat-Kombination in den Variablen gezählt. Da unter den besten Modellen nur selten Variablen nach September des Vorjahres auftauchten, entfernten wir diese, um das Frühwarnsystem mit möglichst viel Vorlaufzeit auszustatten. Abbildung 15 zeigt die aktualisierte Liste der besten Jahr-Monat-Kombinationen aus den 100 Modellen mit dem besten F1-Maß für alle 2-Variablen-Kombinationen aller 147 möglichen Wettervariablen (7 Parameter aus den beiden Vorjahren bis zum vorangegangenen September). Wir haben diesen Vorgang

wiederholt, um die Blühstärke aus dem Vorjahr zu allen Teilmengen (rechte Tabelle) hinzuzufügen.

**Abbildung 15: Die häufigsten Jahr-Monat-Kombinationen unter den besten Modellen mit zwei Prädiktoren.**

Die Jahr-Monat-Kombinationen mit den meisten Vorkommen in den 100 Modellen mit dem besten F1-Maß für alle 2-Variablen-Kombinationen der 147 möglichen Wettervariablen aus den beiden Vorjahren bis zum vorhergehenden September. Insgesamt wurden 10.731 Modelle für jede Liste verwendet.

<i>Modelle mit nur Wettervariablen</i>		<i>Modelle mit der Blühstärke aus dem Vorjahr</i>	
Jahr-Monat-Kombination	Vorkommen	Jahr-Monat-Kombination	Vorkommen
V1_9	95	V2_9	34
V2_4	23	V1_9	29
V2_9	22	V2_4	24
V1_6	15	V2_11	21
V1_4	9	V1_5	14
V2_5	7	V1_6	11
V2_12	5	V2_6	10
V2_8	5		

(a) (b)

Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

Dieselben drei Jahr-Monat-Kombinationen tauchten auf Platz 1-3 beider Listen der Abbildung 15 auf. Dies ist ein starker Hinweis darauf, dass die Wetterbedingungen im V2\_4 (April aus dem Vorvorjahr), V2\_9 (September aus dem Vorvorjahr) und V1\_9 (September aus dem Vorjahr) die Wahrscheinlichkeit einer Rötelmaus-Massenvermehrung und folglich eines humanen PUUV-Ausbruchs am stärksten beeinflussen.

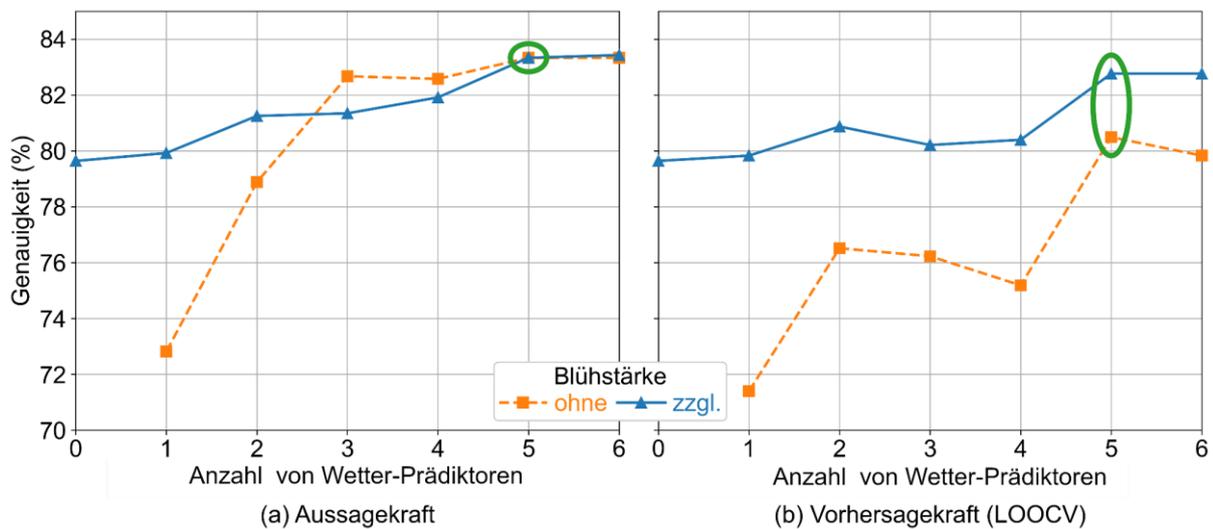
Um Mengen mit stark korrelierten Variablen zu vermeiden, bildeten wir die weiteren Teilmengen mit N-Prädiktoren, indem wir die N-Variablen-Kombinationen auswählten, mit genau einer Variable aus jeder der N obersten Jahr-Monat-Kombinationen (N-faches kartesisches Produkt) für Modelle ohne bzw. mit der Blühstärke aus dem Vorjahr. Aus allen Modellen für jede Mächtigkeit der Teilmengen bis 6 ( $7^N$  Kombinationen, d. h. 7 Wetterparameter für N Jahr-Monat-Kombinationen,  $N \leq 6$ ) bevorzugten wir das Modell mit der besten Sensitivität und einem F1-Maß von mindestens 97% des maximal erreichten für diese Mächtigkeit. Wir nennen dieses Modell „optimal“ für diese Mächtigkeit. Die genauen Kriterien für die Auswahl des optimalen Modells jeder Mächtigkeit waren nicht besonders entscheidend. Eine vergleichbare Leistung wurde für die meisten Variablensets mit einem hohen F1-Maß erwartet, da die Variablen für jede Dimension des kartesischen Produkts stark korreliert waren.

Schließlich beurteilten wir anhand der Leave-One-Year-Out-Kreuzvalidierung (definiert bei der Abbildung 14) die Vorhersagekraft der 13 optimalen Modelle: 6 Modelle ohne und 7 Modelle mit der Blühstärke aus dem Vorjahr. Für diese Modelle wurden die Variablen nicht skaliert, da die unskalierten Stützvektoren („support vectors“) benötigt wurden. Abbildung 16 fasst die Leistung des optimalen Modells für jede Mächtigkeit der Teilmengen anhand der mittleren jährlichen Genauigkeit für das Ausbruchrisiko zusammen.

**Abbildung 16: Die mittlere jährliche Genauigkeit für das Ausbruchsrisiko.**

(a) Die Genauigkeit basierend auf dem Trainingsdatensatz des besten Modells für unterschiedliche Anzahlen von Prädiktoren. Dies ist eine Metrik der Aussagekraft. (b) Der Mittelwert der Genauigkeit für jedes einzelne Jahr nach einer Leave-One-Year-Out-Kreuzvalidierung (LOOCV) für unterschiedliche Anzahlen von Prädiktoren. Dies ist eine Metrik der Vorhersagekraft.

Die orangen gestrichelten Linien mit den rechteckigen Markierungen zeigen die Modelle ohne die Blühstärke aus dem Vorjahr. Die blauen durchgezogenen Linien mit Rautenmarkierungen zeigen die Modelle zuzüglich der Blühstärke aus dem Vorjahr. Die grünen Kreise bezeichnen die optimalen Modelle ohne bzw. zuzüglich der Blühstärke aus dem Vorjahr, welche die finalen Modelle sind.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

Die Aussagekraft stieg mit dem Einbezug weiterer Prädiktoren. Es wird erwartet, dass sich die Anstiegsrate verlangsamt, bis die Genauigkeit ihren Maximalwert erreicht und um diesen herum schwankt. Andererseits ist die Vorhersagekraft keine zunehmende Funktion der Anzahl von Prädiktoren und es wird erwartet, dass sie mehrere lokale Maxima aufweist. Das heißt, dass der Einbezug weiterer Prädiktoren nicht immer zu einer erhöhten Genauigkeit führt.

Wir wählten die Modelle mit fünf Prädiktoren ohne bzw. zuzüglich der Blühstärke aus dem Vorjahr aus, die im folgenden Abschnitt diskutiert werden. Für die Teilmengen ohne die Blühstärke aus dem Vorjahr war die Auswahl eindeutig, da die Vorhersagekraft bei einer Teilmengengröße von 6 abnahm. Für die Teilmengen mit der Blühstärke aus dem Vorjahr blieb die Voraussagekraft konstant. Dafür wählten wir die Teilmenge mit der kleinsten Größe aus. Darüber hinaus war die zusätzliche Variable die Bodenfeuchte vom vorhergehenden Juni (V1\_SM\_6), die stark mit der bereits im Modell enthaltenen Bodenfeuchte vom vorhergehenden Mai (V1\_SM\_5) korreliert (0,65 Pearson-Korrelationskoeffizient mit einem p-Wert  $\ll 0,001$ ) und auch zukünftig korrelieren wird. Eine geringe Modellkomplexität insbesondere ohne stark korrelierte Prädiktoren ist wichtig, um Überanpassung zu vermeiden und dem „Fluch der Dimensionalität“ zu entgehen.

Wir trainierten noch ein Modell mit allen 171 Variablen aus unserem anfänglichen Pool von Prädiktoren. Die Genauigkeit für das Ausbruchsrisiko im Trainingsdatensatz (Aussagekraft) betrug 90,0% und war damit allen bisher durchsuchten Modellen überlegen. Andererseits betrug die mittlere jährliche Genauigkeit nach der LOOCV (Vorhersagekraft) nur 43,1% und war damit allen bisher durchsuchten Modellen unterlegen. Dies ist ein Beweis für die Überanpassung, die bei einer hohen Anzahl von Prädiktoren erwartet wird, die zudem stark miteinander korreliert sind.

#### 4.1.4.1 Das Argument für die Blühstärke der Buche

Aus Abbildung 16 geht hervor, dass die Aussagekraft der Modelle ohne die Blühstärke aus dem Vorjahr zunächst steil ansteigt und um die Aussagekraft der Modelle mit der Blühstärke aus dem Vorjahr schwankt. Durch alleinige Betrachtung dieses Diagramms würden wir zum Schluss kommen, dass die Blühstärke nicht notwendig ist. Der Unterschied wurde jedoch in der Vorhersagekraft an dem Testdatensatz deutlich, weil die mittlere jährliche Genauigkeit der Modelle mit der Blühstärke aus dem Vorjahr immer höher war, was bedeutet, dass der Einbezug der Blühstärke aus dem Vorjahr voraussichtlich die Vorhersagekraft einer beliebigen Teilmenge von Prädiktoren erhöhen wird.

Das Modell mit nur der Blühstärke aus dem Vorjahr hatte bereits eine Genauigkeit von 79,6%. Ein Modell mit nur Wetterparametern musste mindestens fünf Variablen enthalten, um dieses Genauigkeitsniveau zu erreichen. Dies ist eine Demonstration der Vorhersagekraft der Blühstärke aus dem Vorjahr. Obwohl die Wetterparameter mit der Blühstärke korreliert sind, führt ihre Einbeziehung zu einem robusteren Modell, das sich in einer Abnahme der Standardabweichung der jährlichen Genauigkeit von 18,4% (nur die Blühstärke aus dem Vorjahr) auf 15,9% (fünf Wetterprädiktoren zusammen mit der Blühstärke aus dem Vorjahr) zeigte. Gleichzeitig stieg die mittlere jährliche Genauigkeit auf 82,8 % mit 33 zusätzlichen richtigen Vorhersagen in der LOOCV.

Die Blühstärke ist zeitvariabel und ihre räumliche Verteilung spiegelt die Mastvariabilität in jedem Jahr wider. Damit können, wie in 4.1.2.2 diskutiert, räumlich variierende Effekte berücksichtigt werden, die nicht allein durch die Wetterbedingungen erklärt werden können.

## 4.2 Optimale Modelle

Wie im vorherigen Abschnitt diskutiert, wählten wir das Modell mit fünf Wetterprädiktoren zuzüglich der Blühstärke aus dem Vorjahr aus, dessen LOOCV eine mittlere jährliche Genauigkeit von 82,8% hatte, mit 123 Überschätzungen des Ausbruchrisikos (11,6%) und 59 Unterschätzungen (5,6%) in 16 Jahren in allen untersuchten Landkreisen. Die Prädiktoren aus dem Vorvorjahr waren die maximale Lufttemperatur im April (V2\_Tmax\_4), der Niederschlag im September (V2\_Pr\_9) und die mittlere Lufttemperatur im November (V2\_Tmean\_11) sowie aus dem Vorjahr die Bodenfeuchte im Mai (V1\_SM\_5), die Bodentemperatur im September (V1\_ST\_9) und die Blühstärke (V1\_Blühstärke).

Da die Erhebung von Blühstärke-Daten auf freiwilliger Basis erfolgt und möglicherweise nicht langfristig gesichert ist, wählten wir auch das optimale Modell ohne Blühstärke aus, um sicherzustellen, dass das Prognosemodell auch ohne Blühstärke funktionsfähig bleiben wird. Die Leistung wird voraussichtlich geringer sein, aber die Vorhersagekraft ist hoch genug, um es als valides Prognosemodell zu werten. Anhand der LOOCV hatte das Modell mit nur fünf Wetterprädiktoren eine mittlere jährliche Genauigkeit von 80,5%, mit 136 Überschätzungen des Ausbruchrisikos (12,9%) und 70 Unterschätzungen (6,6%) in 16 Jahren. Die Prädiktoren aus dem Vorvorjahr waren die maximale Lufttemperatur im April (V2\_Tmax\_4) und die Sonnenscheindauer im September (V2\_SD\_9) sowie aus dem Vorjahr die Bodenfeuchte im April (V1\_SM\_4) und die Bodentemperaturen im Juni (V1\_ST\_6) und im September (V1\_ST\_9).

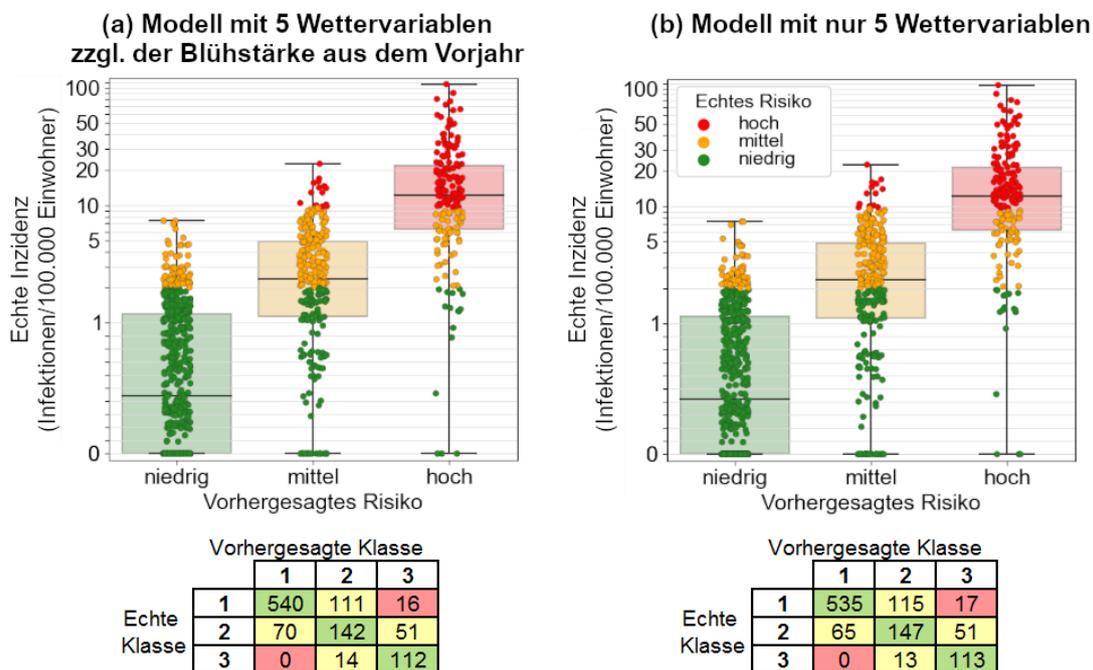
In beiden Modellen waren die Wetterprädiktoren aus dem Vorvorjahr universeller, da sie sowohl die Wettereffekte auf die Buchensamenproduktion als auch den direkten Einfluss auf das Wachstum der Rötelmauspopulationen berücksichtigen können. Andererseits sind die Wetterprädiktoren aus dem Vorjahr bodenspezifisch, entweder die Feuchte oder die Temperatur, und stehen in direktem Zusammenhang mit dem Habitat der Nagetiere.

Aus dem Ausbruchsrisiko haben wir die vorhergesagte Inzidenz und die vorhergesagte Risikoklasse umgerechnet, wie in Abbildung 13 beschrieben. Das Bestimmtheitsmaß der Inzidenz betrug 0,457 für das Modell mit der Blühstärke aus dem Vorjahr (0,418 für das Modell ohne Blühstärke) bei einer prädiktiven Fehlerquadratsumme von 54.530 (58.489) und die mittlere jährliche Genauigkeit der Risikoklassen betrug 72,2% (71,1%). Diese Werte können mit dem perfekten Modell verglichen werden, das 100% Ausbruchsrisiko-Genauigkeit hätte: 0,583 Bestimmtheitsmaß der Inzidenz, 41.920 Fehlerquadratsumme und 81,7% Risikoklassen-Genauigkeit. Da das perfekte Modell unrealistisch ist, kann die Risikoklassen-Genauigkeit auch mit der des Nullmodells (63,2%), des Mehrheits-Modells (68,9%) und des Ausbruch-Modells (71,1%) verglichen werden.

Für das endgültige Modell entfielen die Validierungs- und Testdatensätze, und alle Datenpunkte wurden in den Trainingsdatensatz aufgenommen, da alle Jahre wichtig sind, insbesondere bei hoher Dimensionalität. Abbildung 17 zeigt die Auswertung der Aussagekraft der beiden finalen Modelle.

**Abbildung 17: Aussagekraft der beiden finalen Modelle.**

Oberer Teil: Box-Plots, Unterer Teil: Wahrheitsmatrizen. Die Schwellenwerte der Inzidenz für die Definition der Risikoklassen waren 2 und 9,5. Für eine bessere Darstellbarkeit sind die y-Achsen linear im Bereich [0, 1] und logarithmiert im Bereich [1, 110]. Die Whisker für die Box-Plots waren auf 0 und 100 gesetzt, d. h., dass kein Datenpunkt in dieser Analyse als Ausreißer gekennzeichnet wurde. Die Box jeder vorhergesagten Risikoklasse enthält die mittleren 50% der Inzidenzwerte und der Strich in der Box repräsentiert den Median der Datenpunkte. Die Farbe der Box entspricht der vorhergesagten Risikoklasse. Die Farbe der Datenpunkte entspricht der echten Risikoklasse.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

Die beiden Modelle haben eine vergleichbare Risikoklassen-Genauigkeit: 83,2% für das Modell mit der Blühstärke aus dem Vorjahr und 83,3% für das Modell mit nur Wettervariablen. Die Wahrheitsmatrizen beider Modelle sind diagonaldominant, mit 178 Überschätzungen und 84 Unterschätzungen für das Modell mit der Blühstärke aus dem Vorjahr; sowie 183 Überschätzungen und 78 Unterschätzungen für das Modell mit nur Wettervariablen über die 16 Jahre hinweg. Die zweite Variante resultierte in weniger Unterschätzungen, was für die Risikoeinschätzung vorzuziehen ist. Wie bereits erwähnt hat das Modell mit der Blühstärke aus

dem Vorjahr jedoch eine höhere Vorhersagekraft. Der Tukey's post-hoc Test zeigte an, dass die Mittelwerte aller vorhergesagten Risikoklassen jedes Modells signifikant unterschiedlich sind (alle p-Werte < 0,001).

Abbildung 18 listet die Koeffizienten der beiden SVC für alle Prädiktoren auf. Die Koeffizienten stellen die Koordinaten eines Vektors dar, der orthogonal zur Hyperebene ist, die das geringe und das hohe Ausbruchsrisiko trennt. Daher bedeutet ein positiver Koeffizient, dass, wenn alle anderen Prädiktorwerte konstant bleiben, ein erhöhter Wert des spezifischen Prädiktors den Datenpunkt näher an die Hyperebene oder sogar auf ihre gegenüberliegende Seite (wenn der Punkt ein geringes Ausbruchsrisiko anzeigte) führen würde, oder weiter von der Hyperebene weg (wenn er schon ein hohes Ausbruchsrisiko anzeigte). Die gemeinsamen Prädiktoren der beiden Modelle sind die maximale Lufttemperatur im April aus dem Vorvorjahr (V2\_Tmax\_4) und die Bodentemperatur im September aus dem Vorjahr (V1\_ST\_9). Die Koeffizienten beider Prädiktoren haben für beide Modelle das gleiche Vorzeichen: negativ für V2\_Tmax\_4 und positiv für V1\_ST\_9. Die Prädiktoren aus der zusätzlichen gemeinsamen Jahr-Monat-Kombination sind V2\_Pr\_9 und V2\_SD\_9, jeweils mit positivem und negativem Koeffizienten. Ihre entgegengesetzte Relation mit dem Ausbruchsrisiko wurde aus der negativen Korrelation zwischen den beiden Prädiktoren validiert: -0,41 Pearson-Korrelationskoeffizient mit einem p-Wert << 0,001. Die Koeffizienten der Restwetterprädiktoren (V2\_Tmean\_11, V1\_SM\_4, V1\_SM\_5 und V1\_ST\_6) sind negativ. Der Koeffizient der Blühstärke aus dem Vorjahr ist positiv.

#### **Abbildung 18: Die Koeffizienten des Support Vector Klassifikators für die beiden finalen Modelle.**

Die Koeffizienten wurden auf den Wert von V1\_ST\_9 skaliert und auf vier Dezimalstellen gerundet. Ihre Werte sind zwischen den Modellen aufgrund der unterschiedlichen Dimensionalität auch für die gemeinsamen Prädiktoren nicht direkt vergleichbar.

Innerhalb jedes Modells ermitteln die Koeffizienten eine Schätzung für den Einfluss ("importance") jedes Prädiktors, nach der Multiplikation mit den erwarteten Parameterwerten. Beispielsweise waren die Mittelwerte für die Vorhersage 2023 in ganz Deutschland: 11,3 °C für V2\_Tmax\_4, 35,3 mm für V2\_Nd\_9, 4,90 °C für V2\_Tmean\_11, 64,0 % nFK für V1\_SM\_5, 15,8 °C für V1\_ST\_9, 2,49 für V1\_Blühstärke (Mittelwert über alle HKG), 174 Stunden für V2\_SD\_9, 79,7 % nFK für V1\_SM\_4 und 22,1 °C für V1\_ST\_6, wobei alle Werte auf drei signifikante Stellen gerundet wurden. Mit diesen Werten hat die V1\_ST\_9 die höchste Bedeutung im Modell mit nur den Wetterparametern, etwa 1,7-mal mehr als die V1\_ST\_6, die folgt.

##### *Modell mit 5 Wettervariablen zzgl. der Blühstärke aus dem Vorjahr*

Konstante	V2_Tmax_4	V2_Pr_9	V2_Tmean_11	V1_SM_5	V1_ST_9	V1_Blühstärke
-3,5083	-0,7036	0,0102	-0,1515	-0,0889	1	1,4449

##### *Modell mit nur 5 Wettervariablen*

Konstante	V2_Tmax_4	V2_SD_9	V1_SM_4	V1_ST_6	V1_ST_9
4,7835	-0,3727	-0,0164	-0,059	-0,4313	1

Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster.

Obwohl das Jahr aus dem die Datenpunkte stammten nicht in die Modelle einfluss, gruppieren die Modelle die meisten Datenpunkte desselben Jahres in dieselbe Ausbruchsklasse. Dieser Clustering-Effekt wird durch die intrinsische Clusterbildung der Prädiktorenwerte verursacht, die deutlicher bei den Wetterparametern ist, aber auch in der großflächig synchronisierten Blühstärke auftritt.

Beide Modelle ermittelten ein hohes Ausbruchsrisiko für alle Landkreise in den Jahren 2010, 2012, 2017 und 2019 und ein geringes Ausbruchsrisiko in den Jahren 2008, 2009, 2011, 2013, 2016, 2018 und 2020. Zusätzlich ermittelte das optimale Modell mit der Blühstärke aus dem Vorjahr ein geringes Ausbruchsrisiko in den Jahren 2006 und 2014; wohingegen das Modell ohne die Blühstärke zusätzlich ein hohes Ausbruchsrisiko für alle Landkreise in den Jahren 2015

und 2021 ermittelte. Die restlichen Jahre bestanden aus einer Mischung aus geringem und hohem Ausbruchsrisiko.

Die beiden Modelle hatten 168 gemeinsame Fehler mit 128 Überschätzungen und 40 Unterschätzungen. Das Modell mit der Blühstärke aus dem Vorjahr hat 9 zusätzliche Fehler in den Jahren: 2007 (Unterschätzung für den LK Emsland und Überschätzung für den LK Unstrut-Hainich-Kreis), 2014 (6 Unterschätzungen für LK Borken, LK Coesfeld, LK Emsland, LK Grafschaft Bentheim, LK Steinfurt und LK Wesel) und 2015 (Unterschätzung für den LK Freyung-Grafenau). Die zusätzlichen Fehler des Modells ohne die Blühstärke waren in den Jahren: 2006 (2 Überschätzungen für LK Emsland und LK Wesel), 2007 (Überschätzung für den LK Grafschaft Bentheim), 2014 (Überschätzung für den LK Recklinghausen), 2015 (3 Überschätzungen für LK Deggendorf, LK Passau und LK Regen) und 2021 (Überschätzung für den LK Wesel). Bei beiden Modellen wurde kein Bias oder systematischer Fehler festgestellt.

### **4.3 Fazit und Ausblick**

Wir haben ein systematisches, heuristisches Verfahren zur Entwicklung eines Prognosemodells für das humane PUUV-Infektionsrisiko in Deutschland etabliert. Nach ausführlicher Untersuchung des Datensatzes wurden die Daten so transformiert, dass nützliche Informationen daraus gewonnen werden konnten. Daher haben wir die Frage der Risikobewertung umformuliert. Der resultierende Ansatz unterscheidet sich von inzidenz- und infektionsbasierten Formulierungen, die in der Literatur üblich sind.

Dies ist das erste einheitliche Prognosemodell für das humane PUUV-Infektionsrisiko in Deutschland. Es kombiniert eine geringe Komplexität mit einer hohen Vorhersagekraft. Seine Parameter können Einblicke in die ökologischen, biologischen und sozialen Mechanismen hinter den humanen PUUV-Infektionen geben, die wiederum angewendet werden können, um das Modell in einem zukünftigen Schritt zu verfeinern. Das Modell ist auf weitere Gebiete Deutschlands erweiterbar, in denen vielleicht zukünftig das PUUV endemisch wird. Die Methode kann auf weitere Regionen in Mittelwesteuropa übertragen werden, die ähnliche ökologische Parameter aufweisen, wie in den Niederlanden, Belgien, Dänemark und Frankreich.

Der vorliegende Ansatz unterscheidet sich grundlegend von den in Kapitel 2 validierten bisherigen Modellierungsansätzen. Während das BW Modell (Eichner et al.) die Infektionen und die JKI Modelle (Reil, Jacob, et al. 2018) direkt die Risikoklassen ermittelten, führten wir das Ausbruchsrisiko ein, eine neue Größe, welche der Prognose von Infektionen und Risikoklassen vorangestellt ist. Das Ausbruchsrisiko beschreibt das binäre humane PUUV-Infektionsrisiko. Basierend auf dieser Definition findet in jedem Landkreis in jedem Jahr entweder ein Ausbruch statt oder nicht. Unsere Methode berücksichtigt keine quantitativen Auswirkungen der Prädiktoren auf die Inzidenz oder die Infektionen. Stattdessen wird davon ausgegangen, dass die Inzidenz zwischen zwei Extremwerten schwankt, die jeweils einer unterliegenden Wahrscheinlichkeitsverteilung folgen.

Obwohl unsere Methode eventuell nicht das globale Maximum in Bezug auf ein bestimmtes Leistungskriterium ermittelt, vermeidet sie die Einbeziehung stark korrelierter Variablen und weist laut Definition eine hohe Sensitivität und Präzision auf. Die beiden optimalen Modelle benötigen nur 5 bzw. 6 Prädiktoren um eine hohe Vorhersagekraft mit einer mittleren jährlichen Genauigkeit von über 80% bei der Leave-One-Year-Out-Kreuzvalidierung zu erreichen. Der letzte Prädiktor stammte vom September aus dem Vorjahr, was genügend Zeit bietet, um Gegenmaßnahmen für ein potentiell Ausbruchsjahr vorzubereiten und Gesundheitsbehörden, Risikogruppen und Ärzte für die Gefährdung durch das Virus zu sensibilisieren.

Abhängig vom verwendeten Leistungskriterium und der Klassengewichtung kann ein Modell mit einer anderen Teilmenge von Prädiktoren als optimal ausgewählt werden. Vorläufige Ergebnisse deuten auf einen solchen Effekt bei der Verwendung der prädiktiven Fehlerquadratsumme hin. Dennoch vermeidet das entwickelte Verfahren eine Überanpassung und die Einbeziehung stark korrelierter Variablen und es ist hochwahrscheinlich, dass das resultierende Modell eine hohe Sensitivität und Präzision aufweisen wird. Die Vorhersagen selbst werden voraussichtlich nur geringfügig abweichen, hauptsächlich für Landkreise in der Nähe der Hyperebene. Die Unsicherheit des Ausbruchsriskos wird in Kapitel 6 weiter diskutiert.

Nach Betrachtung der Datenpunkte mit geringem und hohem Ausbruchsrisiko in Bezug auf zwei und drei Prädiktoren wählten wir einen Support-Vector-Machine-Klassifikator (SVC) mit einem linearen Kernel als Klassifizierungsmethode. Es gibt jedoch keinen Beweis dafür, dass die beiden Klassen tatsächlich perfekt linear trennbar sind. Unser Ansatz bleibt gültig, solange die Abundanzen der Rötelmauspopulationen zwischen zwei Extrema schwanken. Eine dritte zwischenliegende Klasse für das Ausbruchsrisiko würde zu einer höheren Genauigkeit führen, ihre Unterscheidung war jedoch mit den derzeit verfügbaren Daten nicht möglich. Eine Modellierung mit Anpassung der Hyperparameter von SVC (Regularisierungsparameter C und Klassengewichte) und mit einem radialen Basisfunktion (RBF)-Kernel zeigte keine deutliche Verbesserung der Vorhersagekraft.

Neben dem Ausbruchsrisiko hängt die Unsicherheit der Inzidenz und der Risikoklasse von der Berechnung des Erwartungswerts aus den vergangenen Zeitreihen ab. Unser in Abbildung 13 dargestellter Ansatz geht davon aus, dass die Inzidenzwerte einer Gammaverteilung um jeden der beiden Extremwerte folgen. Für ein hohes Ausbruchsrisiko nähert sich die Gammaverteilung einer Normalverteilung an. Die Annahme einer anderen zugrundeliegenden Wahrscheinlichkeitsverteilung würde zu einem anderen Erwartungswert führen. Beispielsweise ist der Erwartungswert bei einer Gleichverteilung gleich der Mitte des Intervalls für jedes Ausbruchsrisiko. Eine andere Alternative wäre, nur den Mittelwert der letzten Jahre mit demselben Ausbruchsrisiko zu berechnen, was mögliche Niveauveränderungen berücksichtigen kann, z. B. aufgrund von Änderungen im Meldesystem oder im Habitat der Nagetiere. Wir haben uns aufgrund der geringen Datenmenge entschieden, die Berechnung der Inzidenzvorhersage im gesamten Zeitraum 2006–2021 durchzuführen, da es Kreise gibt, die keinen oder nur einen einzigen gekennzeichneten vergangenen Ausbruch in einem begrenzten Datensatz hatten, z. B. ab 2017. Wir zogen das arithmetische Mittel dem Median und dem Modus vor, weil es die Extremwerte berücksichtigt. Diese Extremwerte, die zu der Verteilung gehören und relevant für die Risikobewertung sind, insbesondere für ein hohes Ausbruchsrisiko, würden aber vom Median als Ausreißer gewertet und blieben damit unberücksichtigt. Eine angepasste Errechnung des erwarteten Inzidenzwertes könnte im Einzelfall durchgeführt werden, unter Beachtung der tatsächlichen Auswirkungen und der Relevanz auf Kreisebene der Publikation dieser Prognoseergebnisse.

Ein Teil der Ungenauigkeit ist auf die nicht-überwachte Clusteranalyse der Inzidenz zurückzuführen, welche die Labels vergab. Eine horizontale Verschiebung der Wahrscheinlichkeitsverteilung zwischen einem geringen oder einem hohen Ausbruchsrisiko oder von beiden würde dann zu falschen Labels führen, d. h. ein erhöhter Inzidenzwert, der eigentlich ein neues Niveau für nicht-Ausbruchsjahre bezeichnet, könnte dem hohen Ausbruchsrisiko zugerechnet werden. Dies kann beispielsweise durch eine wesentliche Änderung des Meldesystems verursacht werden, die dazu führen würde, dass die Anzahl der gemeldeten Infektionen näher an der Anzahl der tatsächlichen Fälle liegt. Wir gehen davon aus, dass viele der Unterschätzungen von NI und NW in den letzten Jahren auf keinen lokalen Ausbruch zurückzuführen sind, sondern auf falsche Labels. Eine solche Ungenauigkeit kann nur

durch manuelle Datenbereinigung für die falsch negativen Datenpunkte des Modells reduziert werden, d. h., dass alle falschen Modellklassifikationen genau geprüft und deren Labels gegebenenfalls korrigiert werden sollten. Dieser Schritt müsste mit einer modifizierten Methode zur Berechnung des Erwartungswerts kombiniert werden, wie im vorherigen Absatz erläutert.

Die restliche Ungenauigkeit ist auf in den Modellen nicht berücksichtigte Phänomene und auf probabilistische Effekte zurückzuführen, die zu einem Ausbruch führen können, auch wenn die Prädiktorenwerte nicht darauf hinweisen und umgekehrt. Die Wettervariablen aus dem Prognosejahr können nicht für ein Prognosemodell verwendet werden, beeinflussen aber trotzdem sowohl die Rötelmauspopulationen (Gubler et al. 2001; Vanwambeke et al. 2019) als auch die menschlichen Aktivitäten (Piechotowski et al. 2008) sowie deren Interaktion. Das charakteristischste Beispiel ist der Monat März; die Bodentemperatur (V0\_ST\_3) hat einen Pearson-Korrelationskoeffizienten von 0,39 (p-Wert  $\ll 0,001$ ) mit dem Ausbruchsrisiko. Gleiches gilt für die Wettervariablen nach dem September des Vorjahres, mit dem stärksten Einfluss vom November aus dem Vorjahr; die Sonnenscheindauer (V1\_SD\_11) hat einen Pearson-Korrelationskoeffizienten von 0,35 (p-Wert  $\ll 0,001$ ) mit dem Ausbruchsrisiko. Außerdem ist zu erwarten, dass alle präventiven Maßnahmen das zukünftige Infektionsrisiko beeinflussen, insbesondere reduzieren, vor allem nach der Etablierung dieses systematischen Prognosemodells. Dies sollte bei jeder zukünftigen Validierung oder Modellaktualisierung berücksichtigt werden.

Unser Prognosemodell benötigt keine Daten zur Rötelmauspopulation (z. B. Abundanz und PUUV-Prävalenz) und erfordert daher keine Feldarbeit, wie Monitoring, Fangen und Beprobieren. Dennoch sind solche Daten für die Validierung des Modells unerlässlich, insbesondere um falsche Vorhersagen zu minimieren. Für zukünftige Weiterentwicklungen sollte geklärt werden, wie Blühstärke und Samenproduktion zusammenhängen und welche Bedingungen zu Abweichungen in der Kausalkette Nagetierausbruch–PUUV-Prävalenz–humane PUUV-Inzidenz führen. Antworten auf solche Fragen wären entscheidend für die Entwicklung einer aktualisierten Version des Prognosemodells auf der Grundlage der hier vorliegenden Studie und für die Überprüfung ihrer Schlussfolgerungen und Hypothesen.

## 5 Publikation der Prognoseergebnisse auf einer verstetigten Online-Plattform

### 5.1 Allgemein

Das Julius Kühn-Institut Münster (JKI-MS) stellte zwei Programme zur Online-Publikation zur Verfügung:

- ▶ `PUUV_Prognosemodell_Daten.exe`: Dieses Programm führt die Erfassung und Aufbereitung der für die Prognose erforderlichen Eingaben durch.
- ▶ `PUUV_Prognosemodell_Prognose.exe`: Dieses Programm erstellt die Prognose anhand der erfassten Daten.

Die Programme wurden mit `pyinstaller v5.2` (Cortesi D. et al. 2022) in Python `v3.8.5` (Van Rossum und Drake 2009) generiert und auf Windows 10 getestet. Die verwendeten Pakete waren: `Numpy` (Harris et al. 2020), `Pandas` (McKinney 2010), `Geopandas` (Jordahl et al. 2021), `Matplotlib` (Hunter 2007), `Selenium` (Muthukadan B. et al. 2018), `Beautiful Soup` (Richardson 2022), `SciPy` (Virtanen et al. 2020), `scikit-learn` (Pedregosa et al. 2011), `validators` (Vesterinen K. et al. 2014), `requests` (Reitz K. et al. 2004), `urllib3` (Petrov A. et al. 2020) und `msedge-selenium-tools` (Microsoft Edge Developer Tools Team 2022).

Beide Programme werden über eine einzige Konfigurationsdatei (`.ini`-Format) gesteuert, die mit einem Texteditor geöffnet und bearbeitet werden kann. Im einfachsten Fall muss die nutzende Person nur das Jahr festlegen, für welches das humane PUUV-Infektionsrisiko vorhergesagt werden soll.

Die automatische Datenerfassung erfolgt entweder mit dem Paket `requests` (Reitz K. et al. 2004) oder mit einer eigenständigen Version eines der folgenden Web-Browser: `Edge` (Microsoft 2023), `Chrome` (Google 2023), `Firefox` (Mozilla 2022) oder `PhantomJS` (Hidayat 2018).

Das JKI-MS stellt zwei SVC-Modelle zur Verfügung, die optimalen Modelle mit bzw. ohne die Blühstärke aus dem Vorjahr, wie in Kapitel 4 beschrieben. Um das Modell mit Blühstärke zu verwenden, sollte die nutzende Person noch die Blühstärke-Daten, die voraussichtlich in der August-Ausgabe des Fachmagazins `AFZ-DerWald` (Deutscher Landwirtschaftsverlag GmbH) veröffentlicht werden, manuell erfassen und sie in eine vorgegebene Excel-Tabelle eingeben. Es wurde festgestellt, dass das Modell mit Blühstärke eine höhere Vorhersagekraft hat. Die Einbeziehung des Modells ohne Blühstärke erfolgt jedoch aus zwei Gründen: 1. um eine zweite Prognose zu Vergleichszwecken zu ermöglichen und 2. um sicherzustellen, dass eine Prognose auch dann möglich sein wird, sollten die Blühstärke-Werte unvollständig oder gar nicht vorhanden sein.

Die Prognose-Ausgabe ist im `Shapefile`-Format (`.shp`) und `Bild`-Format (`.png`, `.tiff` und `.jpg`) möglich. In nächstem Kapitel werden die Bilder der Prognose-Ausgabe für die Testfälle 2022 und 2023 vorgestellt.

### 5.2 Zusätzliche Eigenschaften

Wenn einzelne Blühstärke-Werte aus einem HKG fehlen, welche im Modell benötigt werden, wird die Prognose für die restlichen Landkreise dennoch erstellt werden. Eine Prognose für die Landkreise mit fehlenden Blühstärke-Werten kann mit dem Modell ohne Blühstärke erstellt werden.

Falls das Climate Data Center von DWD an eine andere Adresse verlegt wird, kann die neue Adresse in der Konfigurationsdatei eingegeben werden und die automatische Datenerfassung bleibt funktionsfähig. Ändert der DWD die Struktur des Climate Data Center, wird die automatische Datenerfassung nicht mehr funktionieren. In diesem Fall fordert das Programm die nutzende Person auf, die benötigten Raster manuell zu erfassen, sie entsprechend der internen Namenskonvention umzubenennen und in einen bestimmten Ordner zu verschieben, damit die Datenverarbeitung sie nachverfolgen kann.

Die Ausbruchprognose wird in eine Inzidenzvorhersage umgewandelt, die auf den vergangenen Zeitreihen der Inzidenz in jedem Landkreis basiert. Derzeit umfasst der Trainingsdatensatz die Jahre 2006–2021. Weitere zukünftige Werte der echten Inzidenz werden die Vorhersage stetig verbessern. Daher können die vergangenen Zeitreihen aktualisiert werden, z. B. kann die Prognose für das Jahr 2023 auch die Inzidenzwerte des Jahres 2022 berücksichtigen. Das Programm erwartet die jährlich gemeldeten humanen PUUV-Infektionen auf Kreisebene. Die Werte können aus SurvStat@RKI 2.0 (Robert Koch-Institut) automatisch abgefragt werden. Ändert das Robert Koch-Institut die Domain und/oder die Struktur der SurvStat-Anwendung, wird die automatische Datenerfassung nicht mehr funktionieren. In diesem Fall wird die nutzende Person angeleitet, die benötigten Daten manuell zu erfassen, zu formatieren und in eine spezielle Excel-Tabelle einzufügen, damit das Programm sie nachverfolgen kann. Die aktuelle Inzidenzzeitreihe wird als Kontrolle und als zukünftige Referenz ausgegeben.

Die nutzende Person kann die Schwellenwerte der Risikoklassen und die Farbpalette für die Bilder der Prognose-Ausgabe konfigurieren, indem sie die entsprechenden Parameter der Konfigurationsdatei anpasst.

Zusätzlich können die Programme zukünftige Bevölkerungsdaten (Eurostat) automatisch erheben und bearbeiten, weitere Landkreise einbeziehen, und ein aktualisiertes SVC-Modell für die Prognose einsetzen.

## 6 Prognose für Jahre außerhalb des Modellierungszeitraums

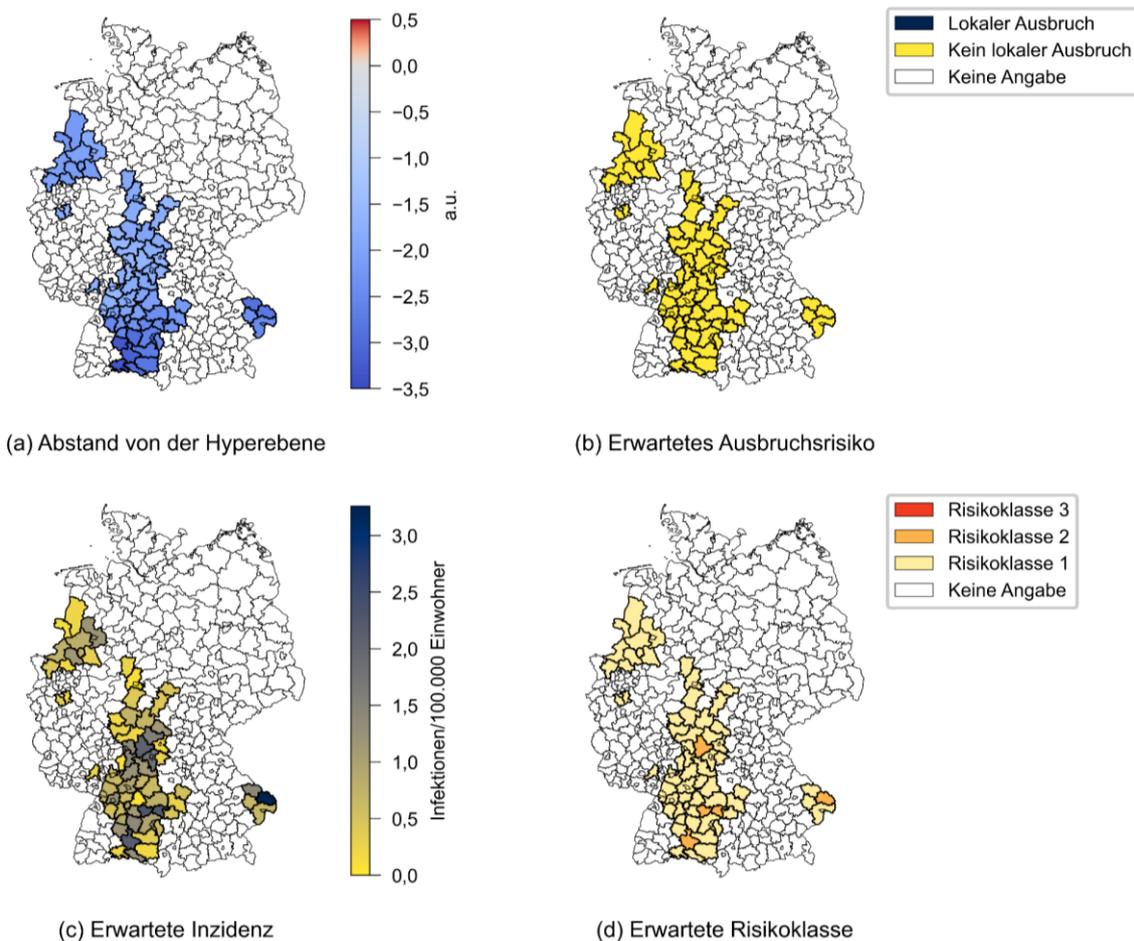
### 6.1 Für das Jahr 2022

#### 6.1.1 Prognose

Mit Wetterparametern aus den Jahren 2020 und 2021 und den Blühstärke-Werten aus dem Jahr 2021 haben wir die Prognose für das Jahr 2022 erstellt. Abbildungen 19 und 20 zeigen die vom Modell generierten Prognosekarten für 2022 anhand des Modells mit bzw. ohne die Blühstärke aus 2021.

#### Abbildung 19: Prognose für 2022 anhand des Modells mit Blühstärke.

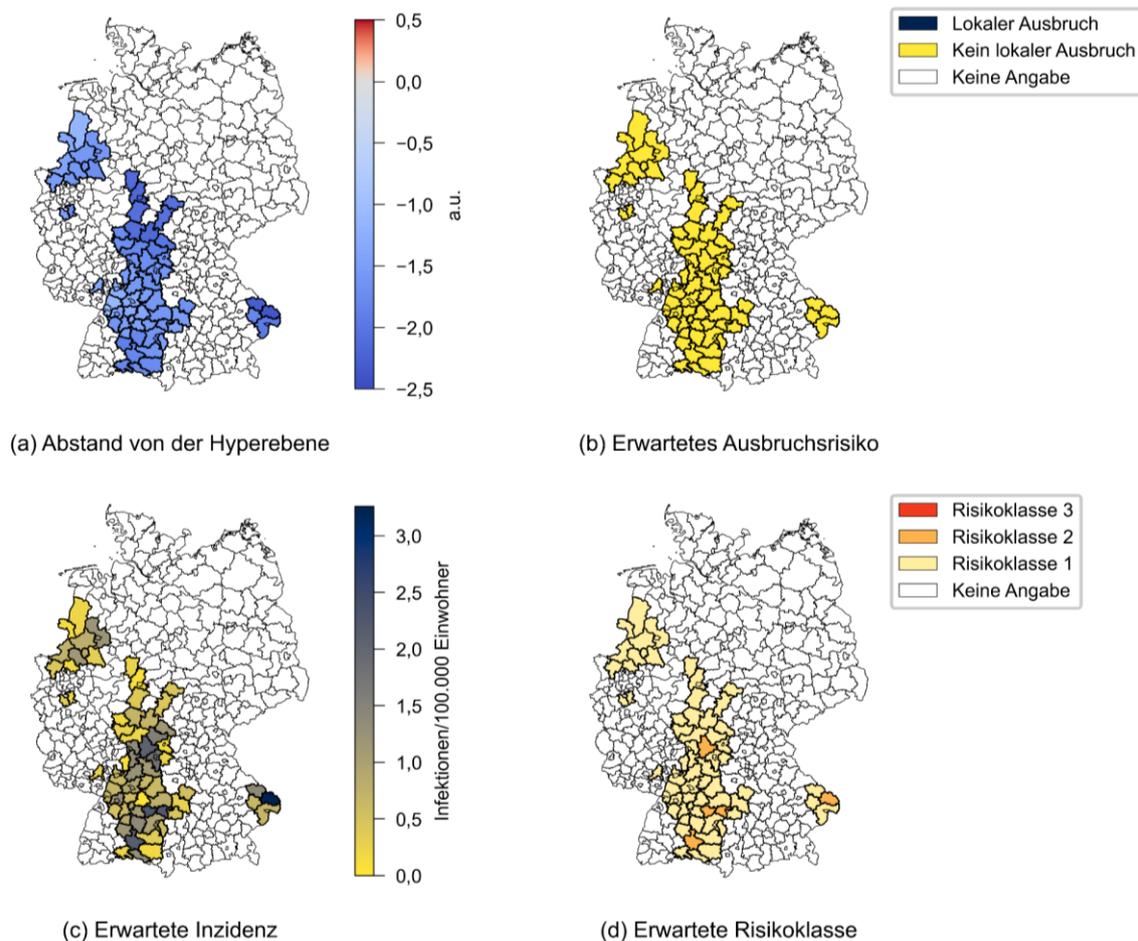
Dargestellt sind die mit dem Programm „PUUV\_Prognosemodell\_Prognose.exe“ generierten Choroplethenkarten für das Jahr 2022. Hier mit dem Modell mit Blühstärke. (a) Der Abstand des Datenpunktes für jeden Landkreis von der vom Support Vector Klassifikator eingepassten Hyperebene. Ein positiver Abstand bedeutet, dass das Modell den Datenpunkt als „Ausbruch“ klassifiziert. (b) Das vorhergesagte Ausbruchsrisiko ergibt sich aus dem Vorzeichen des Abstands von der Hyperebene. (c) Die vorhergesagte Inzidenz wird durch den Mittelwert der vergangenen Inzidenzwerte in jedem Landkreis geschätzt, die das gleiche echte Ausbruchsrisiko anzeigten wie das für dieses Jahr vorhergesagte. (d) Die vorhergesagte Risikoklasse ergibt sich aus der Klassifizierung der vorhergesagten Inzidenz anhand der vordefinierten Schwellenwerte. Die Standardwerte der Schwellen sind 2 und 9,5 Infektionen/100.000 Einwohner.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Hintergrundkarte: (GeoBasis-DE / BKG 2017).

**Abbildung 20: Prognose für 2022 anhand des Modells ohne Blühstärke.**

Die Prognosekarten werden in der Abbildung 19 erklärt. Es werden die Ergebnisse des Modells ohne Blühstärke dargestellt.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Hintergrundkarte: (GeoBasis-DE / BKG 2017).

Die beiden Modelle ermittelten für alle Landkreise das gleiche vorhergesagte Ausbruchrisiko, was zu den gleichen vorhergesagten Inzidenzwerten und vorhergesagten Risikoklassen führte. Beide Modelle haben nur für 5 Landkreise die Risikoklasse 2 („mittleres Risiko“) vorhergesagt: LK Göppingen, LK Heidenheim und LK Sigmaringen von Baden-Württemberg, sowie LK Freyung-Grafenau und LK Main-Spessart von Bayern. Die übrigen Landkreise hatten ein niedriges vorhergesagtes Risiko.

Nur die vorhergesagten Abstände zur Hyperebene unterscheiden sich zwischen den beiden Modellen. Der Datenpunkt, welcher der Hyperebene am nächsten liegt, war beim Modell ohne Blühstärke vom LK Karlsruhe (Abstand=-1,1 a.u.); beim Modell mit Blühstärke vom LK Rhein-Neckar-Kreis (Abstand=-1,6 a.u.). Allerdings sind die Abstände nicht direkt vergleichbar, da die Modelle unterschiedliche Inputräume haben und das Modell mit Blühstärke sogar eine zusätzliche Dimension hat.

Innerhalb eines Modells sind die Abstände der Datenpunkte jedoch gut vergleichbar.

Die Klassifizierung des Ausbruchrisikos ist vertrauenswürdiger je größer der Abstand ist, d. h. je weiter der Punkt von der Hyperebene liegt. Punkte, die nah zur Hyperebene liegen, können durch eine leichte Veränderung eines Prädiktorwertes auf die gegenüberliegende Seite der Hyperebene verschoben werden. Für das Modell ohne Blühstärke, würde das für den LK Karlsruhe passieren, wenn der Wert von V1\_ST\_9 um 7% zunehmen würde (21,127 °C ggb. des

originalen Wertes von 19,745 °C). Für das Modell mit Blühstärke, würde das für den LK Rhein-Neckar-Kreis passieren, wenn der Wert von V1\_ST\_9 um 16% zunehmen würde (22,427 °C ggb. des originalen Wertes von 19,334 °C). Die geringste Zunahme der Blühstärke für eine Veränderung des Ausbruchsrisikos wäre eine mehr als Verdopplung für den LK Rheinisch-Bergischer Kreis, 3,92 ggb. des originalen Wertes von 1,70. All diese Veränderungen sind relativ groß und gelten als unrealistisch. Daher sind diese Vorhersagen als sehr vertrauenswürdig einzustufen.

### **6.1.2 Validierung**

Die höchste echte Inzidenz in 2022 (Robert Koch-Institut: [SurvStat@RKI 2.0](mailto:SurvStat@RKI), Abfragedatum: 24.03.2023) war 2,55 im LK Freyung-Grafenau (BY) mit 2 Infektionen. Der Landkreis Böblingen (BW) hatte 4 Infektionen gemeldet und die Landkreise Karlsruhe (BW) und Wesel (NW) hatten jeweils 3 Infektionen gemeldet. Insgesamt 8 Infektionen (16% der 49 gesamten gemeldeten Infektionen) wurden in Landkreisen gemeldet, die in diese Analyse nicht einbezogen wurden, nämlich 2 Infektionen in SK Berlin (BE), und jeweils 1 Infektion in LK Ahrweiler (RP), LK Lindau (BY), LK Peine (NI), LK Pfaffenhofen a.d.Ilm (BAY), LK Schaumburg (NI) und LK Unna (NW).

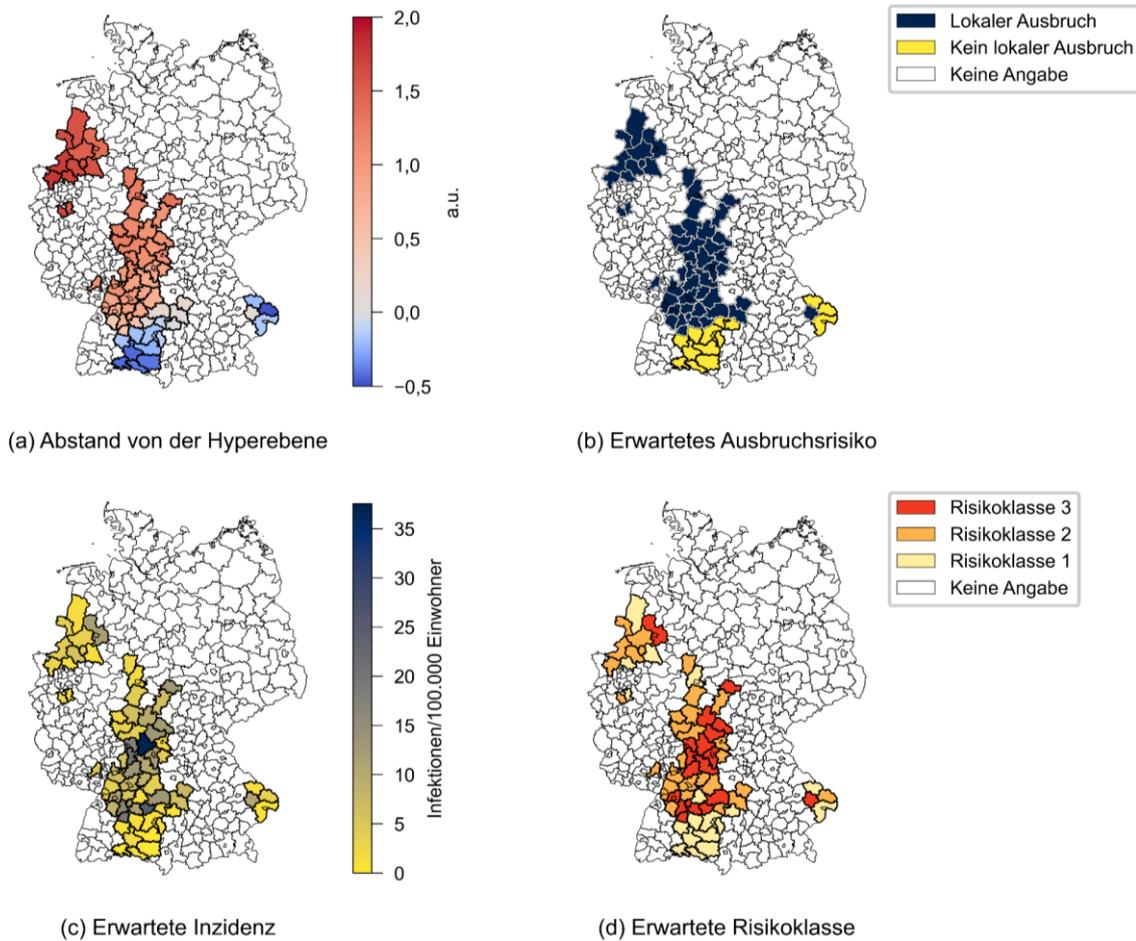
Die Genauigkeit der Prognose war 100% für das Ausbruchsrisiko und 93,9% für die Risikoklassen, mit 62 richtigen Vorhersagen und 4 Überschätzungen für Landkreise in BW und BY (LK Göppingen, LK Heidenheim, LK Sigmaringen und LK Main-Spessart).

## 6.2 Für das Jahr 2023

Mit Wetterparametern aus den Jahren 2021 und 2022 und den Blühstärke-Werten aus dem Jahr 2022 haben wir die Prognose für das Jahr 2023 erstellt. Abbildungen 21 und 22 zeigen die vom Modell generierten Prognosekarten für 2023 anhand des Modells mit bzw. ohne die Blühstärke aus 2022.

### Abbildung 21: Prognose für 2023 anhand des Modells mit Blühstärke.

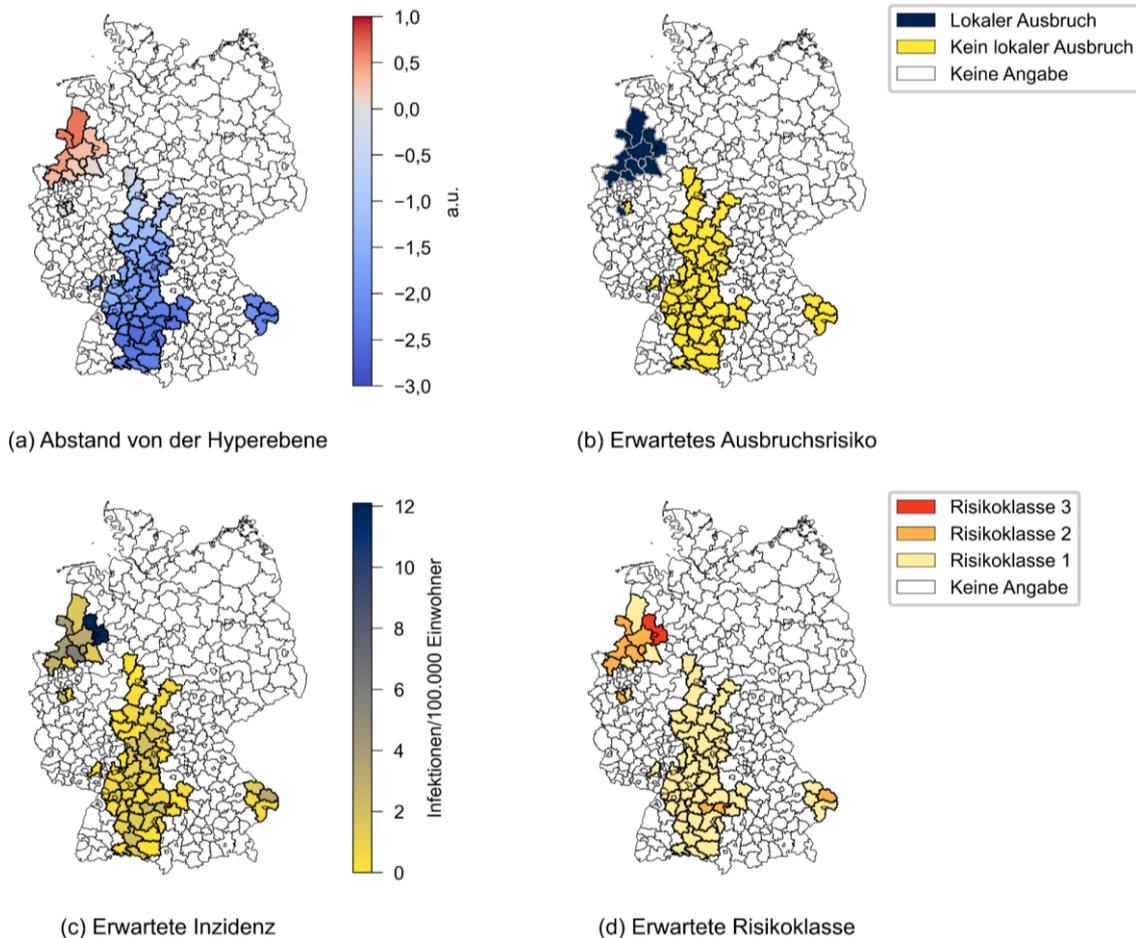
Die Prognosekarten werden in der Abbildung 19 erklärt. Es werden die Ergebnisse des Modells mit Blühstärke dargestellt.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Hintergrundkarte: (GeoBasis-DE / BKG 2017).

**Abbildung 22: Prognose für 2023 anhand des Modells ohne Blühstärke.**

Die Prognosekarten werden in der Abbildung 19 erklärt. Es werden die Ergebnisse des Modells ohne Blühstärke dargestellt.



Quelle: eigene Darstellung, Orestis Kazasidis, Julius Kühn-Institut Münster. Hintergrundkarte: (GeoBasis-DE / BKG 2017).

Die beiden Modelle ermittelten die gleiche Prognose für 14 Landkreise in BW und BY mit geringem Ausbruchsrisiko; und für 11 Landkreise in NI und NW mit hohem Ausbruchsrisiko. Von beiden Modellen wird eine relativ hohe Anzahl von humanen PUUV-Infektionen für die folgenden Landkreise im Jahr 2023 vorhergesagt: LK Borken, LK Coesfeld, LK Emsland, LK Grafschaft Bentheim, LK Osnabrück (zusammengefasst mit dem SK Osnabrück), LK Recklinghausen, LK Steinfurt, LK Warendorf, LK Wesel, SK Köln und SK Münster.

Das Modell mit Blühstärke sagte zusätzlich in 41 Landkreisen in Mittel- und Süddeutschland ein hohes Ausbruchsrisiko vorher, während das Modell ohne Blühstärke dort keine Ausbrüche vorhersagte. Dieser Unterschied pflanzte sich in den Risikoklassen für das Modell mit Blühstärke im Vergleich zum Modell ohne Blühstärke fort: 20 Landkreise sind in die Risikoklasse 2 statt in die Risikoklasse 1, 1 Landkreis ist in die Risikoklasse 3 statt in die Risikoklasse 2 und 17 Landkreise sind in die Risikoklasse 3 statt in die Risikoklasse 1 eingestuft. Der Unterschied zwischen den beiden Modellen wird durch einen Vorzeichenwechsel bei den Abständen von 41 Datenpunkten zur Hyperebene verursacht. Wie bereits erwähnt sind die Abstände nicht direkt vergleichbar. Dennoch können die Vorzeichenwechsel als ein Versatz des Clusters der Datenpunkte betrachtet werden, die sich der Hyperebene nähern oder sie sogar überschreiten, und beweisen die Bedeutung und die Dominanz der Blühstärke für das Modell.

Diese 41 Landkreise sind hauptsächlich in der HKG 09/10 (mit einem Blühstärke-Wert von 2,8), 17 (mit einem Blühstärke-Wert von 2,7), 18 und 23 (jeweils mit einem Blühstärke-Wert von 2) enthalten, und hatten ein mäßiges bis mittleres Blühen im Jahr 2022 im Bereich von 2 bis 2,8. Obwohl es nicht häufig vorkommt, gab es in der Vergangenheit PUUV-Ausbrüche in Jahren mit schwachem oder mäßigem Blühen und daher niedrigen Blühstärke-Werten in bestimmten Landkreisen. So meldeten beispielsweise im Jahr 2007, dem ersten Jahr mit > 1.000 humanen PUUV-Infektionen bundesweit, mehrere Landkreise ihre bisher höchste Jahresinzidenz, obwohl sie 2006 einen Blühstärke-Wert < 2,4 hatten (hauptsächlich verursacht durch den Blühstärke-Wert von 2 im HKG 23), wie z. B. der LK Heidenheim (90,2 Inzidenz), der LK Zollernalbkreis (76,5 Inzidenz) und der LK Reutlingen (64,2 Inzidenz). Andererseits gab es in der Vergangenheit Jahre mit reichlichem oder vollständigem Blühen (>3) in bestimmten Landkreisen, das nicht zu einer hohen Infektionszahl führte. Die charakteristischsten Beispiele waren im globalen Ausbruchsjahr 2017, wobei der Blühstärke-Wert im Vorjahr > 3,5 für alle einbezogenen Landkreise war. Damals hat kein Ausbruch in einigen Landkreisen stattgefunden, wie z. B. im LK Heilbronn (1,7 Inzidenz, 17% der maximal gemeldeten Jahresinzidenz), LK Kitzingen (1,1 Inzidenz, 17% der maximal gemeldeten Jahresinzidenz) und LK Gießen (0 Inzidenz, mit einer maximalen gemeldeten Jahresinzidenz von 4,7 im Jahr 2010).

Das Modell mit Blühstärke ermittelte ein geringes Ausbruchsrisiko immer, wenn der Blühstärke-Wert < 2, und ein hohes Ausbruchsrisiko immer, wenn der Blühstärke-Wert > 3. Die Wetterparameter beeinflussen direkt das Ausbruchsrisiko nur für Landkreise in Jahren mit einem Blühstärke-Wert im Bereich von 2 bis 3. Das Modell ohne Blühstärke hingegen hängt nur von Wetterparametern ab und ermittelte somit sogar ein hohes Ausbruchsrisiko für manche Landkreise in Jahren mit einem Blühstärke-Wert < 2. Es hatte 367 richtige Vorhersagen für ein geringes und 6 richtige Vorhersagen für ein hohes Ausbruchsrisiko, 3 Überschätzungen und 25 Unterschätzungen, wenn der Blühstärke-Wert < 2), wobei das Modell mit Blühstärke allen ein geringes Ausbruchsrisiko ermittelt hatte (370 richtige Vorhersagen, 31 Unterschätzungen).

Ein erhöhter Blühstärke-Wert wird nicht immer in einer erhöhten Samenproduktion resultieren und umgekehrt. Zum Beispiel führte das oben erwähnte mäßige Blühen von 2 für HKG 23 (Schwäbische Alb) im Jahr 2006 zur zweithöchsten jemals gemeldeten Saatgutproduktion (Bundesanstalt für Landwirtschaft und Ernährung 2022) in diesem HKG nach 2001 (4.255 kg im Vergleich zu 4.405 kg im Jahr 2001 und gefolgt von 3.129 kg im Jahr 2016). Andererseits führte das volle Blühen für HKG 23 im Jahr 2018 zu einer Samenproduktion von 886 kg, und damit nur zu 21 % der 2006 gemeldeten Produktion. Die Daten für die Samenproduktion der Rotbuche können jedoch nicht als Prädiktoren genutzt werden, da sie zwischen den HKG oder den Bundesländern nicht konsistent und harmonisiert erhoben werden. Außerdem werden sie später veröffentlicht als die Blühstärke-Werte. Zum Zeitpunkt der Erstellung dieses Berichts lag die Samenproduktion für 2022 noch nicht vor (Bundesanstalt für Landwirtschaft und Ernährung 2022).

Wie bereits im Abschnitt 6.1.1 erwähnt, ist die Klassifizierung des Ausbruchsrisikos vertrauenswürdiger je weiter der Punkt von der Hyperebene des SVC liegt. Der Abstand des Datenpunkts, welcher der Hyperebene am nächsten liegt, war beim Modell ohne Blühstärke -0,003 (a.u.) für den SK Köln; beim Modell mit Blühstärke war er -0,005 (a.u.) für den LK Dillingen a.d.Donau. Diese Datenpunkte könnten bei einer leichten Veränderung eines Prädiktorwertes auf die gegenüberliegende Seite der Hyperebene verschoben werden, was ein anderes Ausbruchsrisiko bedeuten würde. Beim Modell ohne Blühstärke würde der SK Köln zum geringen Ausbruchsrisiko wechseln, wenn irgendeiner der Werte von V2\_Tmax\_4, V2\_SD\_9, V1\_SM\_4 oder V1\_ST\_6 um 0,1% zunehmen oder der Wert von V1\_ST\_9 um 0,1% abnehmen würde. Beim Modell mit Blühstärke würde der LK Dillingen a.d.Donau zum hohem

Ausbruchsrisiko wechseln, wenn der Wert von V2\_Tmax\_4 oder V1\_SM\_5 um 0,2% abnehmen oder der Wert von V1\_ST\_9 um 0,1% zunehmen würde.

Punkte, die viel weiter von der Hyperebene entfernt sind, würden eine wahrscheinlich unrealistisch große Störung erfordern, um das Ausbruchsrisiko zu wechseln. Zum Beispiel, würde der LK Rhein-Neckar-Kreis (Abstand=-1,527 a.u.) beim Modell ohne Blühstärke zum hohen Ausbruchsrisiko wechseln, wenn der Wert von V1\_ST\_9 um 10,5% zunehmen würde, was unrealistisch ist.

## 7 Quellenverzeichnis

- Bundesanstalt für Landwirtschaft und Ernährung (2022): Ernteaufkommen Forstsaatgut. [Internetquelle](#) (24.01.2023)
- Bundesanstalt für Landwirtschaft und Ernährung (2023a): Nationales Inventar forstgenetischer Ressourcen (FGRDEU) - Herkunftsgebiete - Rotbuche. [Internetquelle](#) (25.01.2023)
- Bundesanstalt für Landwirtschaft und Ernährung (2023b): Nationales Inventar forstgenetischer Ressourcen (FGRDEU) - Karte der Herkunftsgebiete, Fagus sylvatica / Rot-Buche [810]. [Internetquelle](#) (25.01.2023)
- Clement, J., Vercauteren, J., Verstraeten, W. W., Ducoffre, G., Barrios, J. M., Vandamme, A. M., Maes, P., Van Ranst, M. (2009): Relating increasing hantavirus incidences to the changing climate: the mast connection. In: International Journal of Health Geographics, 2009, 8:1
- Clement, J., Maes, P., van Ypersele de Strihou, C., van der Groen, G., Barrios, J. M., Verstraeten, W. W., van Ranst, M. (2010): Beechnuts and outbreaks of nephropathia epidemica (NE): of mast, mice and men. In: Nephrology Dialysis Transplantation, 2010, 25(6), S. 1740–1746
- Cortes, C., Vapnik, V. (1995): Support-vector networks. In: Machine Learning, 1995, 20(3), S. 273–297
- Cortesi D. et al. (2022): PyInstaller, Version: 5.2. [Internetquelle](#)
- Deutscher Landwirtschaftsverlag GmbH: Allgemeine Forstzeitschrift für Waldwirtschaft und Umweltvorsorge. AFZ-DerWald. [Internetquelle](#) (25.01.2023)
- Deutscher Wetterdienst: Climate Data Center (CDC). [Internetquelle](#) (12.12.2022)
- Drewes, S., Ali, H. S., Saxenhofer, M., Rosenfeld, U. M., Binder, F., Cuypers, F., Schlegel, M., Rohrs, S., Heckel, G., Ulrich, R. G. (2017): Host-associated absence of human Puumala virus infections in northern and eastern Germany. In: Emerging Infectious Diseases, 2017, 23(1), S. 83–86
- Eichner, M., Brockmann, S., Schwehm, M.: Hanta Vorhersage. [Hanta-Vorhersage Webseite](#) (25.01.2023)
- Esther, A., Imholt, C., Perner, J., Schumacher, J., Jacob, J. (2014): Correlations between weather conditions and common vole (*Microtus arvalis*) densities identified by regression tree analysis. In: Basic and Applied Ecology, 2014, 15(1), S. 75–84
- Ettinger, J., Hofmann, J., Enders, M., Tewald, F., Oehme, R. M., Rosenfeld, U. M., Ali, H. S., Schlegel, M., Essbauer, S., Osterberg, A., Jacob, J., Reil, D., Klempa, B., Ulrich, R. G., Kruger, D. H. (2012): Multiple synchronous outbreaks of Puumala virus, Germany, 2010. In: Emerging Infectious Diseases, 2012, 18(9), S. 1461–1464
- European Environment Agency - CORINE Land Cover: Nomenclature Guidelines. [Internetquelle](#) (25.01.2023)
- European Environment Agency: CORINE Land Cover. [Internetquelle](#) (29.12.2020)
- Eurostat: Population change - Demographic balance and crude rates at regional level (NUTS 3), Online database. [Internetquelle](#)
- Faber, M., Krüger, D. H., Auste, B., Stark, K., Hofmann, J., Weiss, S. (2019): Molecular and epidemiological characteristics of human Puumala and Dobrava-Belgrade hantavirus infections, Germany, 2001 to 2017. In: Eurosurveillance, 2019, 24(32:1800675)
- GeoBasis-DE / BKG (2017): Kreisgrenzen. [Internetquelle](#) (15.03.2021)
- Google (2023): ChromeDriver. [Internetquelle](#) (24.01.2023)

- Gubler, D. J., Reiter, P., Ebi, K. L., Yap, W., Nasci, R., Patz, J. A. (2001): Climate variability and change in the United States: potential impacts on vector- and rodent-borne diseases. In: Environmental Health Perspectives, 2001, 109 Suppl 2, S. 223–233
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T. E. (2020): Array programming with NumPy. In: Nature, 2020, 585(7825), S. 357–362
- Hidayat, A. (2018): PhantomJS. [Internetquelle](#) (24.01.2023)
- Hofmann, J., Loyen, M., Faber, M., Krüger, D. H. (2022): Hantavirus-Erkrankungen: Ein Update. In: Deutsche Medizinische Wochenschrift, 2022, 147(06), S. 312–318
- Hunter, J. D. (2007): Matplotlib: A 2D graphics environment. In: Computing in Science & Engineering, 2007, 9, S. 90–95
- Imholt, C., Reil, D., Eccard, J. A., Jacob, D., Hempelmann, N., Jacob, J. (2015): Quantifying the past and future impact of climate on outbreak patterns of bank voles (*Myodes glareolus*). In: Pest Management Science, 2015, 71(2), S. 166–172
- Jordahl, K., Bossche, J. V. d., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L. J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, Leblanc, F. (2021): GeoPandas, Version: 0.9.0. [Internetquelle](#)
- Kazasidis, O., Jacob, J. (2023): Machine learning identifies straightforward early warning rules for human Puumala hantavirus outbreaks. In: Scientific Reports, 2023, 13(1:3585)
- Kelly, D., Sork, V. L. (2002): Mast Seeding in Perennial Plants: Why, How, Where? In: Annual Review of Ecology and Systematics, 2002, 33(1), S. 427–447
- Khalil, H., Hornfeldt, B., Evander, M., Magnusson, M., Olsson, G., Ecke, F. (2014): Dynamics and drivers of hantavirus prevalence in rodent populations. In: Vector-Borne and Zoonotic Diseases, 2014, 14(8), S. 537–551
- Koenig, W. D., Knops, J. M. H. (2000): Patterns of Annual Seed Production by Northern Hemisphere Trees: A Global Perspective. In: The American Naturalist, 2000, 155(1), S. 59–69
- Konnert, M., Schneck, D., Zollner, A. (2014): Blühen und Fruktifizieren unserer Waldbäume in den letzten 60 Jahren. In: LWF-Wissen, 2014, 74, S. 37–45
- Krüger, D. H., Schonrich, G., Klempa, B. (2011): Human pathogenic hantaviruses and prevention of infection. In: Human Vaccines, 2011, 7(6), S. 685–693
- McKinney, W. (2010): Data structures for statistical computing in Python. In: Proceedings of the 9th Python in Science Conference, 2010, 445, S. 56–61
- Microsoft (2023): Microsoft Edge WebDriver. [Internetquelle](#) (24.01.2023)
- Microsoft Edge Developer Tools Team (2022): msedge-selenium-tools, Version: 3.141.4. [Internetquelle](#)
- Mozilla (2022): geckodriver. [Internetquelle](#) (24.01.2023)
- Muthukadan B. et al. (2018): Selenium, Version: 3.141.0. [Internetquelle](#)
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011): Scikit-learn: Machine Learning in Python. In: Journal of Machine Learning Research, 2011, 12, S. 2825–2830

- Petrov A. et al. (2020): urllib3, Version: 1.25.11. [Internetquelle](#)
- Piechotowski, I., Brockmann, S. O., Schwarz, C., Winter, C. H., Ranft, U., Pfaff, G. (2008): Emergence of hantavirus in South Germany: rodents, climate and human infections. In: Parasitology Research, 2008, 103 Suppl 1, S. 131–137
- Piovesan, G., Adams, J. M. (2001): Masting behaviour in beech: linking reproduction and climatic variation. In: Canadian Journal of Botany, 2001, 79(9), S. 1039–1047
- Reil, D., Imholt, C., Eccard, J. A., Jacob, J. (2015): Beech Fructification and Bank Vole Population Dynamics - Combined Analyses of Promoters of Human Puumala Virus Infections in Germany. In: PLoS One, 2015, 10(7:e0134124)
- Reil, D., Imholt, C., Drewes, S., Ulrich, R. G., Eccard, J. A., Jacob, J. (2016): Environmental conditions in favour of a hantavirus outbreak in 2015 in Germany? In: Zoonoses Public Health, 2016, 63(2), S. 83–88
- Reil, D., Rosenfeld, U. M., Imholt, C., Schmidt, S., Ulrich, R. G., Eccard, J. A., Jacob, J. (2017): Puumala hantavirus infections in bank vole populations: host and virus dynamics in Central Europe. In: BMC Ecology, 2017, 17(1:9)
- Reil, D., Binder, F., Freise, J., Imholt, C., Beyrer, K., Jacob, J., Krüger, D., Hofmann, J., Dreesman, J., Ulrich, R. (2018): Hantaviren in Deutschland: Aktuelle Erkenntnisse zu Erreger, Reservoir, Verbreitung und Prognosemodellen. In: Berliner und Münchener Tierärztliche Wochenschrift, 2018, 131
- Reil, D., Jacob, J., Imholt, C., Ulrich, R. G. (2018): Regionalspezifisches Vorhersagesystem für das Vorkommen gesundheitsgefährdender Nagetiere als Anpassung an den Klimawandel – Abschlussbericht, Forschungskennzahl 3713 48 401. In: 2018, Climate Change(08)
- Reitz K. et al. (2004): Requests, Version: 2.28.1. [Internetquelle](#)
- Richardson, L. (2022): Beautiful Soup, Version: 4.11.1. [Internetquelle](#)
- Robert Koch-Institut: SurvStat@RKI 2.0. [Internetquelle](#)
- Schauber, E. M., Kelly, D., Turchin, P., Simon, C., Lee, W. G., Allen, R. B., Payton, I. J., Wilson, P. R., Cowan, P. E., Brockie, R. E. (2002): Masting by eighteen New Zealand plant species: The role of temperature as a synchronizing cue. In: Ecology, 2002, 83(5), S. 1214–1225
- Schneck, D. (2019): Das Blühen der Waldbäume 2019. In: AFZ-DerWald, 2019, 16, S. 21–23
- Schneck, D. (2020): Zum Blühverhalten der Rotbuche. In: AFZ-DerWald, 2020, 8, S. 25–29
- Schneck, D. (2021): Das Blühen der Waldbäume 2021. In: AFZ-DerWald, 2021, 16, S. 19–21
- Schneck, D. (2022): Das Blühen der Waldbäume 2022. In: AFZ-DerWald, 2022, 16, S. 31–33
- Sentinel Hub Public Collections: CORINE Land Cover nomenclature and colouring scheme. [Internetquelle](#) (25.01.2023)
- Sork, V. L. (1993): Evolutionary Ecology of Mast-Seeding in Temperate and Tropical Oaks (Quercus Spp). In: Vegetatio, 1993, 107/108, S. 133–147
- Tersago, K., Verhagen, R., Servais, A., Heyman, P., Ducoffre, G., Leirs, H. (2009): Hantavirus disease (nephropathia epidemica) in Belgium: effects of tree seed production and climate. In: Epidemiology & Infection, 2009, 137(2), S. 250–256
- Van Rossum, G., Drake, F. L. (2009): Python 3 reference manual. CreateSpace, Scotts Valley, California
- Vanwambeke, S. O., Zeimes, C. B., Drewes, S., Ulrich, R. G., Reil, D., Jacob, J. (2019): Spatial dynamics of a zoonotic orthohantavirus disease through heterogenous data on rodents, rodent infections, and human disease. In: Scientific Reports, 2019, 9(1:2329)

Vesterinen K. et al. (2014): Validators, Version: 0.20.0. [Internetquelle](#)

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., SciPy, C. (2020): SciPy 1.0: fundamental algorithms for scientific computing in Python. In: Nature Methods, 2020, 17(3), S. 261–272

## A Landkreise

### A.1 Zusammengefasste Stadtkreise

**Tabelle 3: Die im Modell enthaltenen zusammengefassten Stadtkreise.**

Die daraus resultierenden aggregierten Daten behielten die Bezeichnung des Landkreises.

Bundesland	Stadtkreis	Landkreis
BW	SK Heidelberg	LK Rhein-Neckar-Kreis
BW	SK Heilbronn	LK Heilbronn
BW	SK Karlsruhe	LK Karlsruhe
BW	SK Pforzheim	LK Enzkreis
BW	SK Ulm	LK Alb-Donau-Kreis
BY	SK Aschaffenburg	LK Aschaffenburg
BY	SK Passau	LK Passau
BY	SK Schweinfurt	LK Schweinfurt
BY	SK Würzburg	LK Würzburg
HE	SK Kassel	LK Kassel
NI	SK Osnabrück	LK Osnabrück
TH	SK Eisenach *	LK Wartburgkreis

\* Der Stadtkreis Eisenach wurde in der Kreisreform 2021 in den LK Wartburgkreis eingegliedert.

### A.2 Einbezogene Landkreise

**Tabelle 4: Die für diese Analyse einbezogenen Landkreise.**

Die Landkreise bzw. Stadtkreise wurden nach zwei Kriterien für den Zeitraum 2006–2021 gefiltert: die Gesamtinfektionen sollten mindestens 20 betragen und die höchste Jahresinzidenz sollte mindestens 2 betragen.

Bundesland	Landkreis	Bundesland	Landkreis
BW	LK Alb-Donau-Kreis	BW	LK Biberach
BW	LK Bodenseekreis	BW	LK Böblingen
BW	LK Calw	BW	LK Enzkreis
BW	LK Esslingen	BW	LK Göppingen
BW	LK Heidenheim	BW	LK Heilbronn
BW	LK Hohenlohekreis	BW	LK Karlsruhe
BW	LK Konstanz	BW	LK Ludwigsburg

Bundesland	Landkreis	Bundesland	Landkreis
BW	LK Main-Tauber-Kreis	BW	LK Neckar-Odenwald-Kreis
BW	LK Ostalbkreis	BW	LK Ravensburg
BW	LK Rems-Murr-Kreis	BW	LK Reutlingen
BW	LK Rhein-Neckar-Kreis	BW	LK Schwäbisch Hall
BW	LK Sigmaringen	BW	LK Tübingen
BW	LK Zollernalbkreis	BW	SK Stuttgart
BY	LK Aschaffenburg	BY	LK Bad Kissingen
BY	LK Deggendorf	BY	LK Dillingen a.d.Donau
BY	LK Donau-Ries	BY	LK Freyung-Grafenau
BY	LK Kitzingen	BY	LK Main-Spessart
BY	LK Miltenberg	BY	LK Neu-Ulm
BY	LK Passau	BY	LK Regen
BY	LK Rhön-Grabfeld	BY	LK Schweinfurt
BY	LK Weißenburg-Gunzenhausen	BY	LK Würzburg
HE	LK Fulda	HE	LK Gießen
HE	LK Kassel	HE	LK Main-Kinzig-Kreis
HE	LK Odenwaldkreis	HE	LK Schwalm-Eder-Kreis
HE	LK Vogelsbergkreis	HE	LK Wetteraukreis
NI	LK Emsland	NI	LK Grafschaft Bentheim
NI	LK Osnabrück	NW	LK Borken
NW	LK Coesfeld	NW	LK Höxter
NW	LK Recklinghausen	NW	LK Rheinisch-Bergischer Kreis
NW	LK Steinfurt	NW	LK Warendorf
NW	LK Wesel	NW	SK Köln
NW	SK Münster	RP	LK Bad Dürkheim
TH	LK Unstrut-Hainich-Kreis	TH	LK Wartburgkreis

Die Namenskonvention entspricht der deutschen Version von SurvStat@RKI 2.0.

Die 12 Stadtkreise der Tabelle 3 sind bereits mit ihrem benachbarten oder umliegenden Landkreis zusammengefasst.

## B Herkunftsgebiete

### B.1 Herkunftsgebiete der Rotbuche

**Tabelle 5: Die Herkunftsgebiete (HKG), die die einbezogenen Landkreise enthalten.**

Kennziffer	Bezeichnung des Herkunftsgebietes
810 01	Niedersächsischer Küstenraum und Rheinisch-Westfälische Bucht
810 03	Heide und Altmark
810 06	Mitteldeutsches Tief- und Hügelland
810 07	Rheinisches und Saarpfälzer Bergland, kolline Stufe
810 08	Rheinisches und Saarpfälzer Bergland, montane Stufe
810 09	Harz-, Weser- und Hessisches Bergland, kolline Stufe
810 10	Harz-, Weser- und Hessisches Bergland, montane Stufe
810 11	Thüringer Wald, Fichtelgebirge und Vogtland, kolline Stufe
810 12	Thüringer Wald, Fichtelgebirge und Vogtland, montane Stufe
810 16	Oberrheingraben
810 17	Württembergisch-Fränkisches Hügelland
810 18	Fränkische Alb
810 19	Bayerischer und Oberpfälzer Wald, submontane Stufe
810 20	Bayerischer und Oberpfälzer Wald, montane Stufe
810 21	Schwarzwald, submontane Stufe
810 22	Schwarzwald, hochmontane Stufe
810 23	Schwäbische Alb
810 24	Alpenvorland
810 25	Alpen, submontane Stufe
810 26	Alpen, hochmontane Stufe

Quelle: (Bundesanstalt für Landwirtschaft und Ernährung 2023a).