

# Digitalgespräch Folge 34

## Maschinelles Lernen im Umweltmonitoring

Mit Hanna Meyer von der Universität Münster, 14. März 2023

<https://zevedi.de/digitalgespraech-034-hanna-meyer/>

*[Der Vorspann mit Musik und Ausschnitten aus dem Gespräch beginnt.]*

**Marlene Görger [mg]:** Frau Meyer, Sie sind Professorin für Fernerkundung und räumliche Modellierung am Institut für Landschaftsökologie der Uni Münster. Sie sind außerdem Expertin für Umweltmonitoring und maschinelle Lernverfahren, die dabei zum Einsatz kommen können.

**Hanna Meyer [Meyer]:** Kann ich diese Informationen, die ich vom Satellit oder von der Drohne bekomme, übersetzen in das, was uns aus ökologischer Perspektive interessiert? Wir als Mensch können gar nicht so genau überblicken: Wie hängen diese ganzen Parameter zusammen? Und deshalb überlassen wir das einem Algorithmus, diese Zusammenhänge zu finden. Wir wollen die meisten Informationen gar nicht auf zehn Kilometer, auf einen Kilometer oder 10 Meter haben, sondern am liebsten natürlich immer im Zentimeterbereich. Da hat Citizen Science natürlich ein Riesenpotenzial.

**Petra Gehring [pgg]:** Gibt es Daten, die im Prinzip vorhanden sind, an die Sie aber nicht rankommen, weil Behörden oder wie auch immer Ihnen die nicht geben?

**[Meyer]:** Ein Click in der Software oder vielleicht zwei, drei Zeilen Programmcode, und dann habe ich meine schöne bunte Karte. So ein maschinelles Lernmodell zu trainieren, das ist keine Herausforderung. Da müssen wir ganz vorsichtig sein, dass keine Fehlinterpretation letztendlich vom Modell stattfindet, und nachher diese Vorhersagen im Prinzip überhaupt nichts mit der Realität zu tun haben.

*[Der Vorspann endet, das Gespräch beginnt.]*

**[mg]:** Menschliches Verhalten und die technologische Entwicklung haben immer schon kleinere oder größere Veränderungen für die Natur bewirkt. Und das oft in einer Weise, die wir eigentlich nicht wollen können, weil unsere eigene Lebens- und Wirtschaftsgrundlage Schaden nimmt. Trotzdem bleiben diese unklugen Praktiken bestehen. Wo Waldflächen in großem Stil gerodet werden, entstehen auch heute noch dauerhaft Steppen und Wüsten. Land- und Viehwirtschaft, übermäßige Jagd und Fischerei, Industrie und Abfälle machen es zunehmend vielen Arten von Lebewesen in kleinen wie großen Regionen der Welt schwer bis unmöglich, weiterzubestehen. Nun bringt außerdem der Klimawandel überall auf der Welt eingespielte Ökosysteme aus dem Gleichgewicht und gefährdet Lebensräume. Und auch wenn einzelne Regionen schwerer von ökologischen Folgen der Klimaerwärmung betroffen sind als andere, brauchen wir global präzises Wissen darüber, welche Veränderungen sich in der Umwelt vollziehen. Ein Fach, das helfen kann, ökologische Zusammenhänge auch über große räumliche und zeitliche Skalen besser zu verstehen, ist die Umweltinformatik. Dabei steht aus Sicht der Wissenschaftler:in natürlich nicht immer das Abwenden von akuten oder drohenden Umweltproblemen im Vordergrund. Aber das gesellschaftliche Interesse und der Bedarf für möglichst umfassende Grundlagen für politische und wirtschaftliche Entscheidungen ist groß. Wie funktioniert die Beobachtung der

Umwelt mit Methoden der Informatik? Wie arbeiten verschiedene Gruppen zusammen, um Daten zu interpretieren und zu verknüpfen? Und trägt auch künstliche Intelligenz dazu bei, belastbares Wissen zu generieren? Darüber wollen wir heute im Digitalgespräch reden. Mein Name ist Marlene Görger. Ich bin Physikerin und Technikphilosophin am Zentrum verantwortungsbewusste Digitalisierung.

**[pgg]:** Und mein Name ist Petra Gehring. Ich bin Professorin für Philosophie an der TU Darmstadt. Bei uns zu Gast im Digitalgespräch ist heute Frau Professor Hanna Meyer aus Münster. Herzlich willkommen bei uns, Frau Meyer und danke für Ihre Zeit.

**[Meyer]:** Ja, danke auch für die Einladung.

**[mg]:** Frau Meyer, Sie sind Professorin für Fernerkundung und räumliche Modellierung am Institut für Landschaftsökologie der Uni Münster. Sie sind außerdem Expertin für Umweltmonitoring und maschinelle Lernverfahren, die dabei zum Einsatz kommen können. In Ihrer Forschung befassen Sie sich also einerseits mit ökologischen Fragen und untersuchen Zusammenhänge in der Natur. Andererseits hört man schon an den Begriffen, die ich gerade genannt habe: Es geht da um Data Science und mathematische Methoden, die zum Erkunden und Verstehen der Umwelt eingesetzt werden. Ist die Umweltinformatik ein eigenständiges Fach oder eher ein Spezialgebiet, das man sich als normale Informatikerin aussucht?

**[Meyer]:** Ja, ich würde die Umweltinformatik eher als Wissenschaft an der Schnittstelle sehen, also an der Schnittstelle zwischen der inhaltlichen Ausrichtung, also zum Beispiel der Landschaftsökologie auf der einen Seite, und auf der anderen Seite dann aber die stark methodische Ausrichtung, wo wir dann im Bereich der Informatik unterwegs sind oder auch der Datenwissenschaften, wo wir im Prinzip versuchen, diese beiden Bereiche zusammenzubringen und Informationen über die Umwelt zu generieren mit computergestützten Methoden.

**[pgg]:** Wie ist das bei Ihnen persönlich gewesen? Wann sind Sie so reingebogen in dieses Gebiet und vielleicht auch warum?

**[Meyer]:** Ja, ich bin eigentlich gar keine Informatikerin. Jedenfalls habe ich das in meinem Studium nicht so direkt gelernt. Ich habe Geografie studiert in Marburg und habe da aber sehr schnell einen Schwerpunkt auf die Fernerkundung gelegt und hatte dort auch einen sehr stark methodisch ausgerichteten Masterstudiengang, wo wir uns sehr stark mit Methoden der Fernerkundung und mit räumlicher Modellierung und dem Umweltmonitoring beschäftigt haben und da auch schon an der Schnittstelle: Wie können wir eigentlich aus, zum Beispiel, Satellitendaten Informationen über die Umwelt generieren? Was brauchen wir da für Methoden? Wie verwende ich da computergestützte Methoden und letztendlich was brauche ich da auch an Grundlagen aus der Informatik, um da ranzukommen?

**[mg]:** Was sind denn das dann für Methoden oder was sind das für Geräte, sage ich mal, mit denen dann Daten gesammelt werden?

**[Meyer]:** Ja, das ist sehr unterschiedlich. Das kommt so ein bisschen auf die Skala an, die wir jetzt betrachten wollen. Also wir können uns das ja so vorstellen, zumindest in der Ökologie, da fangen wir meistens mit, ich nenne das oft die Plottskala oder die Punktskala, an. Das heißt, als Ökologen geht man häufig raus ins Feld, in den Wald, wo

auch immer. Und wir erfassen Informationen über die Umwelt wirklich vor Ort. Und das können wir natürlich rein visuell machen, indem wir beispielsweise zählen, was kommen da für Pflanzen vor. Wir können aber auch hier schon Methoden verwenden, wo es dann in eine automatisierte Auswertung geht. Also wahrscheinlich kennen alle Apps zur Pflanzenbestimmung, wo schon Methoden der künstlichen Intelligenz genutzt werden, um hier automatisiert Daten im Feld zu erfassen, was für Pflanzen vorkommen oder was ich auch für verschiedene Umwelteigenschaften dort habe. Wir können auch Sensoren installieren, die zum Beispiel Informationen über das Klima aufnehmen, und somit bekomme ich eigentlich meistens recht gute Informationen über die Umwelt an diesen einzelnen Punkten. Aber diese Informationen an einzelnen Standorten reicht uns in der Regel nicht, sondern wir wollen Informationen bekommen über Dynamiken, über Veränderungen auf einer viel größeren Skala, also für viel größere Flächen. Und da kommen wir dann relativ schnell in den Bereich der Fernerkundung rein, wo wir verschiedene Sensoren nutzen, die beispielsweise in einer Drohne installiert sind oder auch am Flugzeug oder womit wir viel auch arbeiten, mit Satellitendaten, wo wir dann die Möglichkeit haben, ja flächendeckende Informationen zu generieren und auch zeitlich wiederkehrende Informationen zu generieren. Und womit wir uns dann beschäftigen, ist quasi die Frage: Wie kann ich diese Information, die ich vom Satellit oder von der Drohne bekomme, letztendlich übersetzen in das, was uns aus ökologischer Perspektive interessiert. Also das, was wir dann auch im Feld erkennen, messen oder dort verzeichnen.

**[mg]:** Das heißt, Sie arbeiten eigentlich sozusagen mit zwei Typen von Daten, vielleicht, wenn ich das so sagen darf. Die einen sind Daten, mit denen sie große Flächen erfassen können, und die dann zum Beispiel Bilder sind oder sowas, stelle ich mir vor. Und das andere sind dann im Feld gesammelte oder näher sozusagen an der echten biologischen Situation gesammelte Daten. Und das wird dann aufeinander abgebildet, oder wie kann man sich das vorstellen?

**[Meyer]:** Ja, genauso kann man das sich vorstellen. Also wir kommen im Prinzip nicht drumherum, Referenzdaten im Feld zu erheben, weil wir brauchen einfach die Informationen: Wie sieht die Umwelt draußen aus? Also das können wir alleine mit Satellitenbildern oder Aufnahmen von der Drohne nicht erfassen. Aber die Idee ist, dass wir quasi diese zwei Daten verlinken, also dass wir zum einen die Referenzdaten im Feld uns anschauen und uns aber parallel anschauen: Was sehe ich denn mit dem Satellit oder was sehe ich mit der Drohne, und versuche diese Informationen dann zu verknüpfen, sodass ich die Möglichkeit habe, nicht nur für den Standort, wo ich draußen im Feld unterwegs war, Umweltinformationen zu generieren, sondern dann auch für die gesamte Fläche, die ich dann mit der Drohne befliegen habe oder in meinem Satellitenbild vorfinde. Und diese Verknüpfung ist es letztendlich, wo wir dann auch die Methoden des maschinellen Lernens oder, wenn wir das Buzzword künstliche Intelligenz nutzen wollen, anwenden, weil das, was wir mit dem Satellit oder mit der Drohne bekommen, da bekommen wir letztendlich Informationen über die Reflexion in verschiedenen Wellenlängenbereichen. Und unsere Frage ist jetzt: Wie verknüpfe ich das mit meiner Information, die ich im Feld gemessen habe? Also wie übersetze ich im Prinzip das, was der Satellit sieht oder das, was die Drohne sieht, in das, was uns letztendlich aus ökologischer Perspektive interessiert?

**[pgg]:** Ich stelle mir vor, dass es letztlich für Sie wissenschaftlich wichtig ist, dass Sie möglichst viel, was ja alles mit allem zusammenhängt, in den Blick kriegen. Und insofern stelle ich mir vor, dass die Forschung eine sehr kooperative, kollaborative und

gut koordinierte Forschung wahrscheinlich sein muss, um diese vielen Messungen und Daten und ja auch Auswertungen und vielleicht auch Hypothesen zu der Verknüpfung von konkreten und großflächigen Daten, die Sie haben, dann auch zu teilen mit anderen und das zusammensetzen zu so Mosaikbildern. Gibt's da so eine große Community auf der einen Seite? Und auf der anderen Seite: Teilen Sie Daten? Haben Sie gemeinsame Repositorien? Über die Sie dann auch in der Community reden können?

**[Meyer]:** Genau. Also dieses Teilen der Daten, das ist sehr wichtig, schon alleine, was die Referenzdaten angeht. Ich habe ja gerade schon gesagt: Ohne diese Referenzdaten aus dem Feld geht es nicht. Ohne die kommen wir eigentlich da nicht weiter. Da ist es relativ wichtig, dass wir diese Daten auf Plattformen im Prinzip zusammenstellen, sodass auch andere Forschergruppen die Möglichkeit haben, darauf zuzugreifen. Weil ansonsten ist es so, dass jedes Forschungsteam rausgeht, geht in eine Region, nimmt da Daten auf, und wenn man die nicht teilt, dann bekommt man letztendlich jetzt auch kein Bild, woraus man die Möglichkeit hat, dann auch mal für große Flächen oder vielleicht sogar global nachher Aussagen zu machen. Was da auch ganz wichtig ist, ist, dass man da mit Wissenschaftlern aus verschiedenen Bereichen zusammenarbeitet. Also es sind zum einen die Ökologen, die ganz viel dieses Wissen darüber haben, was wir da an ökologischen Zusammenhängen auch erwarten würden. Und das müssen wir dann auch ganz stark mit dieser methodischen Kompetenz zusammenbringen, um letztendlich diese Stärken, die wir dann auch mit maschinellen Lernverfahren haben, richtig nutzen zu können und richtig einordnen zu können, was unser maschinelles Lernmodell uns dann nachher ausspuckt.

**[mg]:** Das heißt, man muss im Grunde schon vorher wissen, ob das stimmen kann, was dann am Ende das maschinelle Lernen ergibt?

**[Meyer]:** Das wäre auf jeden Fall ganz gut. Vielleicht kurz der Hintergrund, also so ein maschinelles Lernmodell zu trainieren und eine schöne bunte Karte von irgendeiner Umweltinformation zu generieren, das ist keine Herausforderung. Also, das ist ein Klick in der Software oder vielleicht zwei, drei Zeilen Programmcode. Und dann habe ich meine schöne bunte Karte, und die alleine ist halt relativ wenig wert, wenn ich nicht ja da gut interpretieren kann, wie diese Vorhersagen, diese Modellvorhersagen, zu Stande gekommen sind und wie diese Lücken gefüllt worden sind. Und wenn ich da einfach das ökologische Wissen, was ich dafür brauche, nicht habe, dann wird das relativ schwierig, das zu bewerten. Und das brauche ich letztendlich aber, um die Qualität von diesem Produkt, was wir dann im Prinzip generieren, richtig einschätzen zu können, dann auch kommunizieren zu können: Wie gut ist eigentlich diese Karte, die ich letztendlich erstellt habe mit diesen Methoden?

**[pgg]:** Hätten Sie da mal Beispiele?

**[Meyer]:** Da gibt es eine ganze Reihe. Wir beobachten nämlich gerade einen relativ großen Trend, dass maschinelle Lernverfahren zusammen mit Fernerkundung dafür verwendet wird, dass ganz viele verschiedene Umweltinformationen sogar auf globaler Skala veröffentlicht werden. Das reicht von Biodiversitätsparametern, also Abundanzen, also Vorkommen von verschiedenen Arten, über Forstparameter, zum Beispiel Baumhöhen, globale Karten davon oder potenzielle Waldbedeckung, und darauf aufbauend Informationen, zum Beispiel über potenzielle Aufforstungsmöglichkeiten, zu generieren. Klimaparameter, Bodeninformationen: Also

was für Bodentypen, was für Bodenarten habe ich global? Wie ist die Temperatur des Bodens usw. Also, das reicht eigentlich in sämtliche Gebiete der Ökologie rein, wo maschinelles Lernen zusammen mit Fernerkundung oder anderen, ja modellierten Datensätzen, zum Beispiel Klimamodelldaten, genutzt wird, um hier, sogar global, letztendlich Informationen über verschiedene Umweltparameter bereitzustellen.

**[pgg]:** Also bei Baumhöhen und Bodentemperaturen: So, ich denke da jetzt an den harten Sommer. Aufforstungsaktivitäten und irgendwie so ein Versuch zu reagieren. Geht das so schnell, dass Ihre Forschung auch so relativ zeitnah dann auch so in Problembewältigung schon weiterhelfen kann? Also können Sie Forstbehörden, nehme ich jetzt mal, unterstützen? Oder ist es doch eher Forschung im Sinne einer Langfristbeobachtung, so dass Sie eigentlich eher nachträglich registrieren, was passiert ist, und das eher dokumentieren?

**[Meyer]:** Also die Informationen, was gerade passiert, die können wir sehr zeitnah bereitstellen. Also wir haben in der Regel Satellitendaten, die bekommen wir zur Verfügung gestellt – also ich weiß gar nicht – vielleicht ein Tag versetzt, vielleicht sogar noch weniger, aber auf jeden Fall sehr zeitnah. Und da haben wir schon die Möglichkeit, dann sehr aktuell auch abbilden zu können oder quantifizieren zu können, wo wir vielleicht geschädigte Waldgebiete haben, wo vielleicht gerade der Borkenkäfer drin ist, und die Wälder sind geschädigt und sterben letztendlich auch ab. Was schwieriger wird, wird diese Früherkennung. Also wenn es damit anfängt: Der Borkenkäfer ist in der Fläche, aber wir sehen das noch nicht am Wald. Also der Wald ist erstmal noch intakt, da wird das Ganze schon schwieriger. Aber erstmal zu erkennen: Hier liegt ein Schaden vor, das können wir sehr aktuell liefern. Da brauchen wir dann aber auch nicht unbedingt Methoden des maschinellen Lernens oder der künstlichen Intelligenz. Das können wir relativ direkt aus den Satellitenbildern ableiten, weil sich Wald in einem schlechten Zustand doch relativ gut abhebt von einem, der gerade nicht befallen ist.

**[mg]:** Wie stehen Sie denn in Kontakt zu den Menschen in der Praxis, nenne ich es jetzt mal, also Forstwirten oder Leuten, die dann auch für bestimmte Regionen zuständig sind oder die pflegen? Nutzen die Ihre Daten? Sind die darauf angewiesen oder wissen die das eigentlich schon selber, was Sie dann da auch in den Satellitenbildern sehen können?

**[Meyer]:** Doch da gibt es auf jeden Fall immer wieder Austausch, dass dann auch angefragt wird, für bestimmte Regionen dann mal Informationen bereitzustellen. Sowohl im Bereich von Waldökosystemen, wir sind gerade aktuell, aber auch in verschiedenen Projekten zum Monitoring von Moorökosystemen unterwegs, wo wir auch relativ viel Austausch haben mit Leuten, die in der Renaturierung auch von Mooren zuständig sind. Ja, wo die Hoffnung besteht, dass wir mit der Fernerkundung da die Möglichkeit haben, eine Erfolgskontrolle von Maßnahmen durchzuführen. Genau.

**[mg]:** Sie hatten ja vorhin auf die Frage nach den Karten, an denen gerade gearbeitet wird, so globale Karten beschrieben, und die Bedeutung der Referenzdaten für die Qualität oder für die Aussagekraft von Fernerkundungsdaten betont. Ist das denn so, dass die Referenzdatenlage für so globale Betrachtung ausreicht, oder wird da nicht weit gesprungen?

**[Meyer]:** Also das kann man sich wahrscheinlich relativ gut vorstellen, dass es Bereiche gibt, wo sehr viele Referenzdaten vorliegen. Also typischerweise ist das Mitteleuropa oder auch in den USA, wo wir eine relativ hohe Dichte an Referenzdaten haben, und dann hörts in anderen Gebieten auch relativ schnell auf. Ich meine, das kann man sich gut vorstellen. Im Himalaya zum Beispiel ist es deutlich schwieriger, da eine ausreichende Menge an Referenzdaten zu generieren, einfach weil es schwieriger ist, dort hinzukommen. Und da könnten wir jetzt wahrscheinlich noch viele, viele andere Bereiche nennen, wo es einfach nicht möglich ist, eine ausreichende Dichte an Referenzdaten zu gewinnen. Aber genau da ist ja auch die Motivation, dass wir dann Fernerkundung nutzen, um auch für solche Bereiche Informationen zu generieren, wo es vielleicht schwierig ist, hinzukommen. Aber da kommen wir natürlich direkt dann auch in die Problematik: Wenn ich keine Referenzdaten über solche Regionen habe, kann ich dann die Fernerkundung und die maschinellen Lernverfahren nutzen, um für diese Bereiche sogenannte Vorhersagen zu machen, also Informationen für diese Region zu generieren? Und das funktioniert im Prinzip nur, wenn ich vergleichbare Gebiete habe, also wo ich Gebiete habe, die sich in ihren Umwelteigenschaften ähnlich verhalten zu der Region, wo ich jetzt keine Referenzdaten habe, so dass ich meinem maschinellen Lernmodell letztendlich beibringen kann, mit was für einer Umwelt ich dort zu rechnen habe, wo ich nicht hinkomme, dadurch, dass es lernen konnte von anderen ähnlichen Gebieten. Das heißt, das muss im Prinzip gegeben sein. Ich muss eine ähnliche Umwelt letztendlich abgedeckt haben, irgendwo in meinen Referenzdaten, sonst wird das relativ schwierig für eine neue Region, eine Vorhersage zu machen, wenn die sich sehr stark unterscheidet in ihren Umwelteigenschaften.

**[mg]:** Gibt es denn Fälle, in denen das schon gut funktioniert, also wo diese globalen Karten gut belastbar sind, in der Hinsicht?

**[Meyer]:** Also relativ gut funktioniert das für Parameter, wo ich den relativ direkten Zusammenhang habe zwischen dem, was man mein Satellit sieht, und dem, was ich im Feld messe. Also ich denke da an verschiedene Parameter aus der Vegetationsökologie, also zum Beispiel ganz einfach: Deckungsgrad der Vegetation oder Biomasse. Das ist etwas, was ich relativ leicht aus Satellitendaten ableiten kann. Und schwieriger wird es dann, je komplexer dann nachher diese Zielvariable oder dieser Zielparameter ist. Also ich denke gerade an eine Studie, wo es um Abundanzen von Nematoden, also kleinen Würmern im Boden, ging. Wenn ich das modellieren möchte, dann kann ich mir relativ gut vorstellen, dass ich eine ganze Reihe an verschiedenen Einflussfaktoren habe, die da auch eine Rolle spielen, wie viele Individuen da letztendlich vorkommen. Und ja, da können wir uns dann auch gut vorstellen, dass natürlich auch die Qualität dieses globalen Produktes dann nachher was ganz anderes ist als für die Parameter, wo ich einfach einen relativ klaren Zusammenhang zwischen meinen Fernerkundungsdaten und meiner Zielvariable habe.

**[pgg]:** Was ist Abundanz? Ob es viele oder wenige Würmer sind?

**[Meyer]:** Ja, genauso können wir das im Prinzip erst mal ausdrücken. Ja.

**[mg]:** Aber Sie befassen sich ja auch mit zeitlicher Modellierung. Wir haben jetzt ja viel über Flächen gesprochen und räumliche Modellierung. Was für Zeiträume haben Sie denn da im Auge? Kann man das so pauschal sagen oder ist das ganz unterschiedlich?

**[Meyer]:** Also wir machen bei uns in der Forschung relativ wenig in Richtung Vorhersage, sondern wir machen eine zeitliche Modellierung in der Regel in dem Sinne, dass wir Lücken füllen wollen. Was auch wieder der Hintergrund ist, dass Referenzdaten ja häufig nicht kontinuierlich aufgenommen werden – außer wir haben natürlich Sensoren, die kontinuierlich Parameter messen, zum Beispiel Klimastation. Aber ansonsten geht man in der Regel raus ins Feld, nimmt Parameter auf und kommt vielleicht ein Jahr später wieder dahin. Und alles, was dazwischen passiert ist, davon wissen wir nichts. Und da können wir uns das ähnlich vorstellen wie bei der räumlichen Vorhersage, wo wir quasi diese räumlichen Lücken füllen wollen, dass wir diese Lücken in der Zeitreihe füllen wollen, um letztendlich zu räumlich kontinuierlichen, aber auch zeitlich kontinuierlichen Umweltinformationen zu kommen.

**[pgg]:** Und wie weit reichen die zurück? Wir können ja beim Wald mal bleiben. Das ist ja eine sehr langfristige Angelegenheit, bis so ein Wald gewachsen ist.

**[Meyer]:** Ja, genau. Also die ersten Satellitendaten, die wir in der Regel nutzen vom Landsat Satellit die sind – jetzt darf ich nichts Falsches sagen – ich glaube 1973, aber die ersten Datensätze, also so weit können wir mit einem System zurückgehen, was uns relativ hochaufgelöste, also räumlich hochaufgelöste Daten liefert, also wo wir einen relativ hohen Detailgrad über die Umwelt bekommen. Und das sind die Daten, mit denen wir dann überwiegend arbeiten, wenn es darum geht, weiter zurückzuschauen, wie sich ein Ökosystem verändert hat.

**[mg]:** Noch mal so ein anderer Aspekt der Zeitlichkeit. Wie lange sind denn die Referenzdaten brauchbar? Gerade wenn man an den Klimawandel denkt: Ist der ein Problem für die Haltbarkeit?

**[Meyer]:** Naja, die Referenzdaten, die wir jetzt aufnehmen, die haben ja erstmal nur Gültigkeit für den Zeitpunkt, wo wir sie aufgenommen haben. Und jetzt können wir uns natürlich die Frage stellen, je nachdem um was für Informationen es sich handelt, wenn wir jetzt an Baumarten zum Beispiel denken, also da gehe ich jetzt davon aus, dass am nächsten Tag immer noch die gleiche Baumart an dieser Stelle steht, dann kann ich natürlich davon ausgehen, dass die dann immer noch Gültigkeit hat. Wenn ich jetzt aber an Lufttemperatur denke, dann hat nicht unbedingt die Messung von jetzt in diesem Moment auch noch morgen zu diesem Zeitpunkt Gültigkeit. Also das hängt dann davon ab, wie zeitlich dynamisch diese Zielvariablen sind, ob die noch Gültigkeit haben oder nicht.

**[pgg]:** Da wird man aber wahrscheinlich auch wieder mit Durchschnittswerten arbeiten müssen. Ich bleib jetzt mal bei den Würmern. Es könnte ja irgendeine Katastrophe kommen, die im Boden die Würmer absterben lässt. Plötzlich sind gar keine mehr da, obwohl man normalerweise prognostizieren würde, was weiß ich, so eine Wurmpopulation bleibt ungefähr so konstant.

**[Meyer]:** Genau. Was wir dann letztendlich auch im Prinzip machen, ist so eine indirekte Modellierung. Das heißt, wir sehen ja letztendlich mit dem Satellit nicht, ob da ein Wurm ist oder wie viele Würmer da sitzen, sondern was wir machen, ist, dass wir uns anschauen, was habe ich denn da eigentlich für Umwelteigenschaften? Also, was ist die Geländehöhe, was ist die Exposition vielleicht? Was habe ich da für Bodeneigenschaften? Wie viel Niederschlag bekommt diese Fläche, wie kalt oder wie warm ist sie usw. Und aus diesen Eigenschaften lerne ich im Prinzip, was das

potenzielle Vorkommen von diesen Würmern ist. Also ist diese Umwelt im Prinzip geeignet, dass diese Tiere dort vorkommen können oder nicht, und wie geeignet ist sie? Weil direkt diese Individuen zu zählen, das ist noch nicht mal mit der Drohne danach möglich.

**[mg]:** Was lernt man denn aus Modellprojekten? Sie waren ja auch in Marburg mal beteiligt, oder sind das noch, an einem Projekt, in dem eine Fläche sehr intensiv, sag ich mal, mit Sensoren und anderen Methoden der Datengewinnung ausgestattet war und man sozusagen richtig viel rausholen konnte für eine bestimmte Fläche. Können Sie mal beschreiben, wie so ein Projekt aufgebaut ist, was da alles gemacht wird, vielleicht auch, was da dann geht, was ansonsten im Feld nicht ginge?

**[Meyer]:** Ja, das Projekt, von dem Sie sprechen, das ist ein Projekt, wo ganz verschiedene Wissenschaftler daran beteiligt waren, also Geographen, Biologen, Mathematiker, Informatiker und einige weitere. Und was da der Vorteil ist, ist, dass man sich einfach aus verschiedenen Blickwinkeln, quasi mit einer Fläche oder mit einer übergeordneten Fragestellung, letztendlich beschäftigt. Ja, wo wir die Möglichkeit dann haben, auch ganz verschiedene Expertise da reinzubringen. Und wir haben die Möglichkeit, dadurch, dass wir uns so ein kleines Waldsystem anschauen, wo wir rausgehen können und es uns regelmäßig einfach auch noch mal anschauen können, dass wir die Region relativ gut kennen. Das heißt, wenn wir jetzt hier Methoden des maschinellen Lernens einsetzen, um flächendeckende Informationen zu generieren, da habe ich einfach die Möglichkeit, das Ganze zu validieren, also zu überprüfen, wie gut ist mein Modell überhaupt in der Lage, diese Vorhersagen zu machen. Einfach weil durch diese kleine, vergleichsweise kleine Fläche ich erst mal die Option habe, das Ganze wirklich zu kontrollieren. Und deshalb eignet sich sowas natürlich super, um erst mal Methoden zu entwickeln, die ich dann hoffentlich auch auf Fragestellungen anwenden kann, die über so eine kleine Modellregion letztendlich hinausgehen.

**[mg]:** Wie gut funktioniert denn da die interdisziplinäre Zusammenarbeit?

**[Meyer]:** Also aus dem Projekt kann ich jetzt gar nicht so ganz viel raus erzählen, weil ich da leider nur am Anfang beteiligt war und dann dadurch, dass ich nach Münster gewechselt bin, nicht mehr so intensiv dabei bin. Aber so ganz generell glaube ich, dass es da schon noch zum Teil einiges zu tun gibt. Also weil man doch immer wieder sieht, dass so Projekte im Prinzip scheitern, wenn man da nicht zusammenkommt und sich austauscht. Sowohl wenn ich rein als Ökologe versuche, solche globalen Datensätze über die Umwelt zu generieren, oder ich als reiner Datenwissenschaftler das versuche, ich im Prinzip die Expertise des anderen doch brauche, um da zu zuverlässigen Aussagen zu kommen. Ansonsten ist das vielleicht methodisch perfekt gelöst, aber vielleicht an dem vorbei, was ich eigentlich als Ökologe brauche. Oder es ist von der ökologischen Idee her eine super Herangehensweise, aber mir fehlen vielleicht die Tools oder die Möglichkeiten, um dann letztendlich wirklich das herauszukitzeln, was diese Methode liefern kann. Und deshalb, denke ich, ist das ganz wichtig, dass man da viel mehr zusammenkommt und aus den einzelnen Teildisziplinen die Expertise zusammenbringt, um solche ja doch ehrgeizigen Projekte anzugehen, wie auf globaler Skala letztendlich Umweltinformationen bereitzustellen. Und das Ganze gilt natürlich auf der regionalen Skala, genauso wie auf der globalen Skala. Aber auf der globalen Skala hat es in der Regel dann doch noch mal eine ganz andere Dimension. Und auch viel mehr Aufmerksamkeit wird da erzeugt, sodass es da



vielleicht noch offensichtlicher ist, dass man hier diese verschiedenen Bereiche gut verknüpfen muss.

**[pgg]:** Ich würde Aufmerksamkeit, Sie haben auch eben mal kurz gesagt, das, was Sie am Ende erstellen, ist ein Produkt. Und das ist es ja auch, weil tatsächlich dieses Wissen einerseits natürlich wissenschaftlich hochrelevant ist für verschiedene Disziplinen. Auf der anderen Seite geht es aber auch um Handlungsmöglichkeiten, um Umgang mit ökologischen Fragestellungen, Klimawandel usw. Das Ergebnis – ich stellte mir das jetzt irgendwie mal kartenartig vor – das Sie produzieren, ist das perfektioniert oder optimiert, dass andere Wissenschaftlerinnen und Wissenschaftler damit umgehen können und insofern schon so ein innerwissenschaftliches Produkt zur wissenschaftlichen Weiterarbeit? Oder optimieren Sie Karten auch in Richtung auf so eine generelle Lesbarkeit, dass man sagen kann: Naja, das sagt einfach Leuten was, die auf diese Karte gucken, und man kann jetzt auch, ohne Spezialistin zu sein, die Hauptbotschaften erkennen oder die Hauptpunkte und Unterschiede, die in der Karte gezeigt werden, oder gibt es beides? Gibt es Karten für sozusagen die Fachwelt und dann auch noch vielleicht ein bisschen einfachere, klarere für öffentliche und Politikberatungszwecke? Wie geht Ihre Disziplinen damit um?

**[Meyer]:** Zunächst, um zu verdeutlichen: Also ich habe vorhin auch schon davon geredet, dass wir letztendlich Produkte generieren. Und da sagen Sie richtig, das kann man sich im Prinzip wie eine Karte vorstellen, wobei die Karte letztendlich ja nur die Abbildung von diesem Datensatz oder von diesem Produkt ist, was wir generieren. Dahinter liegt natürlich ein Datensatz, der dann auch ausgewertet werden kann und nicht nur als Abbildung letztendlich präsentiert wird. Es geht uns aber auch darum, diese Daten dann, die wir generieren, zu nutzen, um daraus neue Erkenntnisse zu generieren. Also es geht uns nicht nur darum, dieses Produkt bereitzustellen und das anderen an die Hand zu geben und zu sagen, so, jetzt macht damit, was ihr wollt, sondern letztendlich wollen wir das natürlich auch nutzen, um daraus neue Hypothesen zu generieren, neues Verständnis über Zusammenhänge zu generieren, um letztendlich dann auch mit den Methoden, die wir hier anwenden, zu einem Erkenntnisgewinn in der Ökologie beizutragen.

**[pgg]:** Und leistet ein und dasselbe digitale Artefakt beides? Oder müssen Sie die irgendwie ausbuchstabieren?

**[Meyer]:** Ja, da muss man tatsächlich so ein bisschen aufpassen, weil wir reden jetzt hier relativ viel von Methoden des maschinellen Lernens. Und tatsächlich sind diese Methoden relativ gut in der Lage, solche räumlichen Vorhersagen zu machen, also wenn wir es so nennen wollen, diese Produkte zu generieren. Sie sind aber erst mal gar nicht so gut in der Lage, einen Erkenntnisgewinn zu generieren, weil diese Methoden häufig auch als Blackbox bezeichnet werden. Das heißt, wir wissen gar nicht, was genau da passiert. Und das ist auch genau dieser Grund, warum wir sie eigentlich verwenden. Weil die Idee ist: Der Algorithmus lernt selbst diese Zusammenhänge aus den Daten. Das ist ja das Tolle daran. Wir gehen davon aus, wir als Mensch können gar nicht so genau überblicken, wie hängen diese ganzen Parameter zusammen, und deshalb überlassen wir das einem Algorithmus, diese Zusammenhänge zu finden. Aber dadurch, dass dieser Algorithmus in der Lage ist, sehr komplexe nichtlineare Zusammenhänge zu lernen, können wir auch diese gelernten Zusammenhänge nachher gar nicht mehr so gut interpretieren, was letztendlich dazu führt, dass wir zwar häufig dann ein gutes Produkt haben oder optimalerweise ein gutes Produkt

haben, aber wirklich zu verstehen: Was hat dieser Algorithmus da eigentlich gelernt oder was steckt da drin in diesem Modell? Da muss man dann noch mal in neue Methoden reinschauen, und das ist auch ein Feld, was sich gerade im Bereich der Fernerkundung und räumlichen Modellierung gerade so weiterentwickelt, das sogenannte Interpretable Machine Learning, wo es darum geht, auch zu interpretieren: Was hat dieses Modell gerade gelernt und warum sagt es mir jetzt für die eine Region vorher, dass ich hier mit einer Biodiversität von ich weiß nicht, wie viel Arten zu rechnen habe und woanders mit weniger oder mehr. Also dieses Verständnis, was dieses Modell letztendlich gelernt hat, das ist nochmal so ein Forschungsbereich, der gerade auch viel mehr aufkommt und wo ich ja Hoffnung habe, dass uns das auch weiterhilft, um zum einen die Modelle zu verbessern. Dadurch, dass ich in der Lage bin zu verstehen, was da eigentlich gelernt wurde, um dann auch möglicherweise einzugreifen, wenn ich sehe: Oh, das Modell hat völlig an dem vorbei gelernt, was wir eigentlich als Ökologe wissen, wie diese Zusammenhänge aussehen, und dass ich da nochmal hinterfragen kann: Ist dieses Modell überhaupt geeignet? Oder wie kann ich noch mal das Ganze so optimieren, dass das Ganze letztendlich verbessert wird? Und zum Anderen kann ich dann auch diese Modellvorhersagen besser verstehen, was ich dann ja auch brauche, um das Ganze zu kommunizieren. Also ich muss ja irgendwie in der Lage sein, jemandem zu erklären, wenn er jetzt die Managementmaßnahme XY durchführt, was das für einen Effekt für diese Fläche hat. Wenn ich das nicht kann, dann ja, glaube ich, ist das relativ schwierig, das auch zu vermitteln.

**[mgj]:** Jetzt nochmal, ob ich das richtig verstanden habe: Also das bedeutet im Zweifelsfall, das maschinelle Lernverfahren wird angewandt und es findet einen Zusammenhang. Der ist dann auch darstellbar, zum Beispiel als Karte. Aber was dieser Zusammenhang ökologisch bedeutet, ist gar nicht klar. Also ist irgendwas, was für die Maschine nur Sinn macht, und sie müsste uns eigentlich sagen, was sie gelernt hat?

**[Meyer]:** Also was das Modell lernen soll, das geben wir vor als sogenannte Trainingsdaten. Das sind im Prinzip diese Referenzdaten, die wir im Feld erheben. Das heißt, es kann nicht sein, dass irgendwas gelernt wird, was nicht das ist, was wir quasi als Ziel reingesteckt haben. Aber wir wissen nicht genau, wie das gelernt wurde. Dadurch, dass wir dem Modell eine ganze Reihe an sogenannten Selektoren vorgeben, was zum Beispiel unsere Satellitendaten sein können, mit der Aufgabe, finde jetzt den Zusammenhang zwischen diesen Informationen und meiner Referenzgröße, die ich im Feld aufgenommen habe. Und wie genau dieser Zusammenhang aussieht. Das ist etwas, was erstmal für uns unbekannt bleibt. Der Algorithmus versucht nur so gut wie möglich diese Zusammenhänge zu finden, also zu lernen, um auf dieser Grundlage dann die Modellvorhersage zu machen. Aber wie genau diese Zusammenhänge aussehen, das müssen wir dann in einem zweiten Schritt versuchen herauszufinden, wenn es uns darum geht, wirklich zu verstehen, was das Modell gelernt hat, was ich glaube, was wir zwingend brauchen, weil ansonsten die Gefahr besteht, dass das Modell letztendlich gar nicht das gelernt hat, was aus wissenschaftlicher Sicht sinnvoll ist.

**[pgg]:** Das klingt so, als ob dann sowas wie eine wirklich belastbare Karte, also die vielleicht auch wirklich Anwendern, Behörden im politischen Raum, Entscheidern oder sowas gegeben wird, doch eine Menge Vorlauf braucht. Also auch absichernden Korrekturlauf oder Abgleich und nochmaliges Messen vielleicht?

**[Meyer]:** Ja, genau das sollte es auf jeden Fall. Also ich habe vorhin schon gesagt, einfach diese Karte zu erstellen ist super einfach. Also das ist überhaupt kein Problem, sondern die Schwierigkeit steckt dann letztendlich im Detail. Erstens zu schauen, wie gut ist dieses Produkt oder diese Karte dann nachher, und was wurde denn auch da gelernt? Da müssen wir nämlich ganz vorsichtig sein, dass wirklich da keine Fehlinterpretation letztendlich vom Modell stattfindet und nachher diese Vorhersagen ja im Prinzip überhaupt nichts mit der Realität zu tun haben. Also da gibt es ganz verschiedene spannende Beispiele, wo man nachher festgestellt hat, dass das Modell letztendlich super versucht hat, die Daten zu reproduzieren, aber damit solche Besonderheiten dieser Trainingsdaten gelernt hat, dass es gar nicht mehr anwendbar war für neue, unbekannte Daten. Also es gab so ein schönes Beispiel, wo Bilddaten verwendet worden sind, die alle aus einer Datenbank kamen, und das waren auch Bilddaten von Pferden, und die sind gar nicht im geografischen Raum jetzt unterwegs. In diesen Bilddaten von Pferden war immer in der unteren linken Ecke so ein Source Tech, also wo wahrscheinlich der Fotograf draufgeschrieben war. Und was das Modell letztendlich nachher gelernt hat, das hat man dann sichtbar gemacht, ist, zu erkennen, ob in dem Bild unten in der linken Ecke ein Text geschrieben ist. Und auf dieser Grundlage hat das Modell die Vorhersage gemacht oder die Antwort geliefert: Ja, das muss ein Bild von einem Pferd sein. Und man kann sich relativ gut vorstellen, wenn ein Modell sowas gelernt hat, und ich wende das wieder an auf die gleichen Daten, dann bekomme ich anscheinend ein zuverlässiges Ergebnis. Also, es wird alle Bilder perfekt als Pferd erkennen können. Sobald ich es aber mit einem Bild konfrontiere, was keinen Text in der unteren linken Ecke hat, dann wird es nicht in der Lage sein zu erkennen, dass es sich dabei um ein Pferd handelt. Und genau solche Gefahren müssen wir bei unseren Modellen letztendlich auch ausschließen. Also dass unser Modell anscheinend gut funktioniert, weil es gut unsere Trainingsdaten reproduzieren kann. Aber sobald wir das Setting ändern, also das Modell vor die Aufgabe stellen, mach mir eine Vorhersage für einen neuen Raum, den du noch nicht kennst, dass das komplett scheitert, weil es so diese Trainingsdaten – wir nennen das – überfittet hat, dass das Modell letztendlich nicht generalisierbar war.

**[pgg]:** Gibt es da pfiffige Faustformeln oder so? Wie können Sie irgendwie erspüren, ob da vielleicht die Maschine oder das Modell in eine falsche Richtung läuft?

**[Meyer]:** Ja, also tatsächlich ist der erste Ansatz, dass man einfach erstmal visuell draufschaut. Das heißt, man schaut sich das Ergebnis an und ja, schaut, ob da irgendwelche Muster in dieser Vorhersage, in dieser Modellvorhersage auftauchen, wo man denkt: Okay, das ist jetzt komplett entgegen allem, was ich als Ökologe weiß oder kenne. Dann kommuniziere ich das natürlich am besten auch mit Experten in diesem Gebiet. Und ganz häufig ist es so, dass man, wenn man denen diese Karte zeigt und die gucken drauf und die schauen sich eine Region an, die sie gut kennen, weil sie dort oft unterwegs waren oder selbst dort Daten aufgenommen haben, dass sie sofort sagen: Nee, also da kann irgendwas so überhaupt nicht stimmen. Und dann muss man sich natürlich überlegen: Woran liegt das? Also liegt das daran, dass mein Modell vielleicht wirklich was Falsches gelernt hat? Oder liegt es daran, dass die Information über meine Zielgröße einfach in den Daten nicht drinsteckt? Wo ich dann nochmal überlegen muss, wie ich mein Modell anders gestalten kann. Und letztendlich gibt es aber auch Methoden, um das Ganze so ein bisschen quantitativ einzuschätzen. Zum Beispiel verschiedene Kreuzvalidierungsmethoden, wo ich sage okay, ich lasse jetzt mal aus meinem Modelltraining eine ganze Region raus, also zum Beispiel mal ganz Deutschland raus, trainiere nur mit den Daten der anderen Länder, wende mein

Modell auf Deutschland an, und schaue dann, wie gut kann denn mein Modell wirklich diese Referenzdaten aus Deutschland letztendlich vorhersagen, weil da habe ich ja dann eine Referenz und kann das damit vergleichen. Und da gibt es aber auch noch eine ganze Reihe an Forschungsbedarf, wie ich das wirklich mache, um da vergleichbare Fehlerabschätzung zu bekommen.

**[mg]:** Das ist jetzt eh eine Frage, die ich die ganze Zeit schon im Kopf habe. So aus Laienperspektive könnte man ja denken, es gibt jetzt zwei Möglichkeiten, das Problem zu lösen. Das eine ist sozusagen, die Informatikseite immer besser zu machen. Und das andere wäre ja die Referenzdatenlage zu verbessern, also dichter, vielleicht repräsentativer zu schauen gezielt: Wo bräuchte ich eigentlich Daten und die dann auch dazu holen oder so? Wird beides verfolgt und wenn ja, was halten Sie für den mittelfristig zielführenderen Ansatz?

**[Meyer]:** Ja, das wird beides auf jeden Fall verfolgt. Also wir sehen das auch in Datenbanken, was Referenzdaten angeht, die immer weiterwachsen, also wo immer mehr Daten auch zusammenkommen. Und auf der anderen Seite natürlich ganz viel Entwicklung, was diese Modelle oder diese Modellierungsmethoden angeht. Ich glaube tatsächlich, dass das erstmal fast wichtiger ist, diese Referenzdatenlücken zu füllen. Das heißt, dass ich wirklich schauen muss, dass ich Referenzdaten generiere, die wirklich repräsentativ nachher sind für die Region, für die ich nachher Modellvorhersagen haben möchte, weil das beste Modell oder der beste Modellierungsansatz hilft uns nachher nicht, wenn ich die Information einfach nicht habe. Also letztendlich kann unser maschinelles Lernmodell da auch nicht irgendwie irgendwas Sinnvolles draus machen, wenn es einfach keine Daten zum Lernen bekommen hat. Das heißt also: Ohne diese Daten zum Lernen, diese Referenzdaten oder Trainingsdaten, helfen mir die Modellentwicklungen auch nicht weiter. Aber auf der anderen Seite muss ich natürlich auch parallel daran arbeiten, um dann diese Methoden auch in diese Richtung dann weiter zu verbessern.

**[mg]:** Haben Sie denn oder gibt es denn Anfragen oder einen spürbaren Bedarf, eine Nachfrage nach Karten oder gerade globalen Karten, wo Sie sagen, die kann man im Moment noch nicht zuverlässig generieren?

**[Meyer]:** Ja, ja, einfach flächendeckende Daten. Jeder möchte flächendeckende Daten haben. Man hätte natürlich alles gerne global hochaufgelöst, am besten auf den Meter genau. Ich glaube, da gibt es aber ganz viel, wo das einfach noch nicht möglich ist, weil letztendlich wir ja auch nur gute Modellvorhersagen machen können, wenn diese Informationen irgendwie dann auch in unseren Fernerkundungsdaten stecken, und wir können einfach nicht alles damit modellieren, bzw. aktuell jedenfalls nicht, mit den Daten, die wir haben. Ja, und dann muss man vielleicht da auch einfach kritisch mit umgehen und an der einen oder anderen Stelle auch sagen: Okay, für diesen Parameter oder auch für diese bestimmte Region kann ich einfach keine Vorhersage zur Verfügung stellen. Das geht einfach nicht. Da hilft nichts. Ich muss einfach rausgehen und mir das Ganze vor Ort anschauen, weil es ansonsten einfach nicht möglich ist.

**[pgg]:** Was ist denn die Karte, die Sie als nächstes rausgeben werden die jetzt gerade in Arbeit ist?

**[Meyer]:** Tatsächlich etwas, was mit maschinellem Lernen gar nicht so viel zu tun hat. Woran wir gerade arbeiten, ist Monitoring von Moorökosystemen auf europäischer Skala. Wir wollen uns anschauen, wo wir Moore haben, die, vielleicht bedingt durch den Klimawandel, trockener geworden sind, wo ich Veränderungen beobachte, was etwas ist, was ich relativ schön mit langen Zeitreihen mir anschauen kann. Ich habe ja vorhin schon die Landsat-Satelliten-Zeitreihen genannt, wo wir relativ weit zurückkommen und wo wir da uns anschauen können: Wie hat sich denn eigentlich dieses Moor entwickelt und kann ich da bestimmte Trends erkennen? Was die maschinellen Lernverfahren angeht, bin ich relativ viel gerade eher in dieser Methodenentwicklungsrichtung unterwegs. Das heißt, da habe ich gerade überhaupt keine globale Karte, die ich gerade in nächster Zeit herausgeben werde oder herausgeben möchte, sondern da ist mein Steckenpferd so ein bisschen mehr diese Entwicklung der Methoden, um zu schauen, was muss ich eigentlich machen oder was für Strategien muss ich entwickeln, um diese maschinellen Lernverfahren besser anwendbar zu machen.

**[mg]:** Gibt es denn technologische Entwicklungen, die noch nicht da sind, aber die Sie spannend finden und auf die Sie sich freuen? Bestimmte Typen von Sensoren, oder?

**[Meyer]:** Ja, da wo ich super gespannt bin, sind die sogenannten EnMap Daten. EnMap ist der erste deutsche Hyperspektralsatellit, der uns Reflexion spektral ganz hochaufgelöst bereitstellt. Das heißt, wir bekommen nicht nur die Informationen der Reflexion im roten, grünen, blauen Wellenlängenbereich, sondern wir bekommen wirklich ganz, ganz viele Informationen. Da bin ich mal gespannt, was wir da letztendlich rausziehen können.

**[pgg]:** Gibt es auch Daten, die im Prinzip vorhanden sind, an die Sie aber als Forscherin nicht rankommen, weil Behörden oder wie auch immer Ihnen die nicht gehen?

**[Meyer]:** Ja, also ich meine, es besteht natürlich auch immer der Wunsch, das Ganze so hochaufgelöst, räumlich hochaufgelöst wie möglich zu machen. Das heißt, wir wollen die meisten Informationen gar nicht auf zehn Kilometer, auf einen Kilometer oder 10 Meter haben, sondern am liebsten natürlich immer im Zentimeterbereich. Und abgesehen davon, dass es natürlich schwierig wird, mit in Zentimeter aufgelösten Bereich mit Daten global zu arbeiten, einfach was Rechenkapazität angeht, aber das wäre vielleicht noch eine Sache, die lösbar ist, kommen wir da natürlich dann auch an Grenzen von dem, was dann nachher für uns auch verfügbar ist.

**[pgg]:** Aber es gibt sie irgendwo, höre ich raus.

**[Meyer]:** Wir sind zum Beispiel hier in NRW in der luxuriösen Situation, dass es Befliegungsdaten, also regelmäßige Befliegungsdaten von ganz NRW gibt, was für uns für Forschung und Lehre auch kostenlos zur Verfügung steht, was zum Beispiel in anderen Bundesländern nicht unbedingt der Fall ist. Also da müssten wir dann auch für Forschungsprojekte relativ viel bezahlen, um an diese Daten zu kommen, also da gibt es auf jeden Fall auch da Unterschiede.

**[mg]:** Spielt in Ihrem Feld auch Citizen Science eine Rolle? Also bekommen Sie Daten von, sage ich mal, Laien zur Verfügung gestellt?

**[Meyer]:** Ja, das ist natürlich ein super Stichwort für die Referenzdaten. Wo wir gerade schon gesagt haben, dass das so ein bisschen die Limitierung ist, weil wir einfach genug Referenzdaten brauchen, aus denen unsere Modelle nachher lernen können. Und da hat Citizen Science natürlich ein Riesenpotenzial. Dadurch, dass viele, viele Leute rausgehen können und dann auch mit verschiedenen zum Beispiel Apps zur Tierbestimmung, zur Pflanzenbestimmung, da gibt es ja ganz viel, Daten generieren können, die dann wieder genutzt werden können, also von uns genutzt werden können, um aus diesen Daten zu lernen. Da muss man natürlich zum Teil dann auch aufpassen, was es da für Bias oder für Ungenauigkeiten gibt, je nachdem, wie diese Daten aufgenommen wurden. Also keiner wird regelmäßig das Vorkommen von irgendeinem Gänseblümchen oder so dokumentieren, einfach weil man es kennt und weil es nichts Besonderes ist, sondern wird dann eher auf seltene Arten gehen und das melden und uns zur Verfügung stellen. Aber natürlich hat das ein Riesenpotenzial, weil dadurch ja die Möglichkeit gegeben ist, vielleicht nicht die qualitativ hochwertigsten Daten zu generieren, weil da wahrscheinlich immer noch je nach Technik, die verwendet wird, oder nach Expertise das Ganze vielleicht nicht so genau ist, aber dadurch einfach riesen Datenmengen generiert werden können, mit der Hoffnung, dadurch auch die vorhin erwähnten Lücken füllen zu können, über Gebiete, wo relativ wenig Daten vorliegen.

*[Der Abspann mit Musik beginnt.]*

**[mg]:** Und damit sind wir am Ende dieses Digitalgesprächs angekommen und bedanken uns bei Hanna Meyer von der Universität Münster für die vielen Einblicke und das spannende Gespräch. Viele Grüße! Und wie immer auch vielen Dank an Sie, liebe Zuhörerinnen und Zuhörer, für das Interesse und die Aufmerksamkeit. Wenn Sie mögen, hören wir uns wie immer in drei Wochen wieder, zur nächsten Folge des Digitalgesprächs, dem Podcast von ZEVEDI, dem Zentrum verantwortungsbewusste Digitalisierung.



This work is licensed under CC BY-NC-ND 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>