

Quantifizierung der maximalen Anpassungsgüte von Sensormesswerten

Diana Pietzner¹ und Peter Wagner¹

Abstract: Bei der Ermittlung von Bodenparametern im Labor treten regelmäßig Messfehler auf. Werden Labormessungen durch Sensormessungen ersetzt, so werden diese gewöhnlich anhand von Labormessungen kalibriert. Wie gut kann die Kalibrierung der Sensormessungen maximal sein, wenn die zur Kalibrierung verwendeten Labormessungen mit Messfehlern behaftet sind? Es werden aus bekannten Verteilungen „wahre“ Bodenparameter simuliert. Dann werden diese durch ebenfalls simulierte Laborfehler gemäß der VDLUFA-Dokumentation zur Messunsicherheit verfälscht. Mittels Regression wird untersucht, welche Anpassungsgüte ein Modell erreichen kann, wenn nur die messfehlerbehafteten Werte bekannt sind. Damit wird gezeigt, dass selbst ein perfekter Sensor, der jedoch anhand imperfekter Laborwerte kalibriert wird, keine vollständige Übereinstimmung mit den „wahren“ Bodenparametern liefern kann. Es werden Grenzwerte für die Anpassungsgüte (R^2) angegeben, die nicht überschritten werden können, sofern ein bestimmter Labormessfehler vorliegt.

Keywords: Sensormessung, Anpassungsgüte, Simulation, Labormessfehler

1 Einleitung

Precision farming hat zum Ziel, die Ertragsfähigkeit des Bodens bestmöglich auszunutzen und den Verbrauch an landwirtschaftlichen Produktionsfaktoren in Bezug zu den Erträgen zu optimieren. Damit soll die Wirtschaftlichkeit des Pflanzenbaus erhöht und es sollen negative Umweltbeeinträchtigungen vermieden werden. Das Problem hierbei ist, die Ertragsfähigkeit des Bodens korrekt einzuschätzen. Die Ertragsfähigkeit wird von vielen verschiedenen, räumlich variierenden Faktoren bestimmt und ist deshalb, auch auf engstem Raum, sehr heterogen. Wichtige Faktoren sind z. B. die Bodengehalte an Makronährstoffen, der pH-Wert oder die nutzbare Feldkapazität. Zur Bestimmung dieser Faktoren werden traditionell Bodenproben im Labor analysiert. Der damit verbundene Arbeitsaufwand und die hohen Kosten der Laboranalysen verhindern aber, dass Daten in der notwendigen hohen räumlichen Auflösung gewonnen werden. Um dennoch kostengünstige und hochaufgelöste thematische Bodenkarten zu erstellen, werden in den letzten Jahren zunehmend Bodensensoren verwendet [Vi11]. Bodensensoren liefern allerdings nicht direkt die gewünschten Werte, z. B. Nährstoffgehalte, sondern beispielsweise Spektren, aus denen die Nährstoffgehalte erst abgeleitet werden müssen. Erschwerend kommt hinzu, dass einzelne Sensoren meist keine befriedigenden Ergebnisse liefern. Erst durch mehrere Sensoren und Datafusion sind bessere Ergebnisse zu erwarten.

¹ Martin-Luther-Universität Halle-Wittenberg, Institut für Agrar- und Ernährungswissenschaften,
Landwirtschaftliche Betriebslehre, Karl-Freiherr-von-Fritsch-Straße 4, 06120 Halle (Saale),
diana.pietzner@landw.uni-halle.de, peter.wagner@landw.uni-halle.de

In jüngerer Zeit werden dazu vermehrt Machine-Learning-Verfahren angewandt [CSW18]. Bei diesen Verfahren werden die mit verschiedenen Sensoren auf einem Feld gemessenen Werte meist individuell interpretiert und diese Ergebnisse dann im Training (Lernvorgang) mit im Labor ermittelten Werten, z. B. Nährstoffgehalten, zusammengeführt. Es ist offensichtlich, dass die Ergebnisse dieser Modellierung nicht besser sein können als die Qualität der im Labor bestimmten Werte, die i.d.R. als wahr angesehen werden. Tatsächlich sind die wahren Werte (Nährstoffgehalte) allerdings nur näherungsweise bekannt. Dies ist durch eine erhebliche Messunsicherheit bedingt, die einer Laboranalyse innewohnt.

Obwohl die Prozeduren zur chemischen Analyse von Bodenproben genau festgelegt sind (z. B. VDLUFA Methodenbuch), zeigen vergleichende Untersuchungen zur Qualitätssicherung doch erhebliche Abweichungen bei der Analyse „identischen“ Probematerials. Wolf et al. [WJH96] dokumentieren Ergebnisse eines Labortests in den USA in dem 20-50 Labore (je nach zu bestimmendem Analyt) bezüglich der Messunsicherheit verglichen werden. Für Phosphor, Kalium und Nitrat werden relative Standardabweichungen (RSD) von 25 %, 20 % bzw. 15 % ausgewiesen (S. 1615). Ein Vergleich europäischer Labore von COOLS et. al [Ci04] kommt ebenfalls zu relativen Standardabweichungen von über 25 % bei Phosphor (S. 691). In einem weiteren Vergleich US-amerikanischer Labore weisen Jacobsen et. al [Ja02] RSDs für Phosphor zwischen 15 und 31 %, für Kalium zwischen 13 und 16 % und für Nitrat zwischen 14 und 44 % aus (S. 309ff). Für Deutschland wird die Messunsicherheit durch Ringversuche der VDLUFA erhoben und dokumentiert [Mu07]. Für Phosphor wird beispielsweise eine „erweiterte“ Messunsicherheit von $\pm 28,7\%$ für einen P-Gehalt von 3 mg/100 g Boden angegeben, für einen P-Gehalt von 30 mg/100g beträgt die Messunsicherheit $\pm 13,1\%$ (S. 724).

Dieser Beitrag hat zum Ziel zu erklären, welches Bestimmtheitsmaß (R^2) bei der Modellierung von Nährstoffgehalten basierend auf Sensordaten unter Berücksichtigung der Messunsicherheit bei Laboranalysen erreicht werden kann.

2 Daten und Methoden

2.1 Maschinelles Lernen

Algorithmen des maschinellen Lernens [CSW18] erzeugen aus einem Angebot an Daten (Trainingsdaten) ein nicht-parametrisches Modell, das die vorhandenen Daten beschreibt und auch auf neue, gleichartige Daten (Testdaten) angewendet werden kann. Dabei ist im Idealfall die Modell-Anpassung (R^2 , RMSE) für die Testdaten genauso gut wie für die Trainingsdaten. Es kann jedoch auch vorkommen, dass die Anpassung an die Testdaten schlechter bzw. besser ist als die Anpassung an die Trainingsdaten, man spricht dann von Over-Fitting bzw. Under-Fitting. Um eine Aussage über die Stärke des Over- bzw. Under-Fitting zu machen, ist es insbesondere in Simulationsstudien üblich, die Daten in einen

Trainings- und einen Testdatensatz aufzuteilen, in der Regel eine Aufteilung von 70 % zu 30 %.

Im Folgenden sollen drei Verfahren des maschinellen Lernens betrachtet werden: Random Forest (RF), Support Vector Machine (SVM) und ein künstliches neuronales Netz (ANN). Zum Vergleich soll ein lineares Modell dienen, welches auch als Simulationsannahme diene. Zur Umsetzung wurden die R-Pakete randomForest [LW02], e1071 (SVM) [Me19] und neuralnet[FGW19], sowie die Funktion lm aus Base-R verwendet.

2.2 Simulation

Auf einem Schlag des Lehr- und Versuchsgutes in Görzig (Sachsen-Anhalt) wurden 2016 Bodenproben entnommen und Laborwerte unter anderem für den pflanzenverfügbaren Phosphorgehalt in mg/100g (p_{cal}), pH-Wert und den Gesamtphosphorgehalt bestimmt durch Röntgenfluoreszenzmessung in mg/kg Trockensubstanz (p_{rfa}) ermittelt. Die Laborwerte haben folgende Minima und Maxima: p_{cal} : 1,4 | 9,8; p_{rfa} : 362,2 | 733,1; pH: 5,5 | 7,2. Nach Ausschluss fehlender und unplausibler Messungen ergibt sich ein Stichprobenumfang von $n = 190$. Es wurde hier ein Stichprobenumfang von 190 Messpunkten gewählt, um den oben genannten Datensatz nachbilden zu können. In vielen praktischen Anwendungen, so auch bei der Sensorkalibrierung, ist die Anzahl der verfügbaren Datenpunkte beschränkt. Jedoch ist bekannt, dass für größere Stichprobenumfänge eine bessere Anpassung zu erwarten ist. Hier wurde als einfachste Simulationsannahme ein lineares Modell gewählt. In der Praxis sind komplexere Zusammenhänge vorstellbar. Jedoch sollten gerade die Machine-Learning-Modelle in der Lage sein, auch diese Zusammenhänge ohne eine Verschlechterung der Anpassung abzubilden.

Um den Einfluss der relativen Messunsicherheit (RMU) auf die Anpassungsgüte R^2 einer linearen Regression bzw. eines Machine-Learning-Modells zu untersuchen, wurden 100 Datensätze simuliert, die dem Originaldatensatz in ihren Verteilungsannahmen nachempfunden sind. Zunächst wurden die Prädiktoren pH-Wert und p_{rfa} unter einer Normalverteilung simuliert, wobei Mittelwert und Standardabweichung aus den Originaldaten abgeleitet wurden. Für diese Prädiktoren wurde kein Labormessfehler eingefügt. Im Anschluss wurde aus diesen beiden Werten ein Wert für p_{cal} ermittelt, dazu diente eine bivariate lineare Regression auf die Parameter der Originaldatei. Im Anschluss wurden auf die simulierten exakten Werte für p_{cal} eine Messunsicherheit addiert und so Labordaten mit Messunsicherheit simuliert. Hierfür wurden nacheinander verschiedene Annahmen zur Größe der RMU getroffen: Festgelegt wurden die RMU an der Stelle des kleinen p_{cal} -Wertes = 3 mg/100g und des großen p_{cal} -Wertes 40 mg/100g. Folgende Werte wurden in der vorliegenden Simulation angenommen: $RMU_{max} / RMU_{min} = 0,1/0,05 \mid 0,2/0,1 \mid 0,29/0,13 \mid 0,4/0,2$.

Die 100 simulierten Datensätze wurden zufällig aufgeteilt in 70 % Trainingsdaten und 30 % Testdaten. Zunächst wurde für die Trainingsdaten je ein lineares Modell und drei Machine-Learning-Modelle (RF, SVM, ANN) angepasst. Anhand der mittels der

Trainingsdaten ermittelten Modellparameter wurde die Anpassung der Modelle an einem bisher unbekanntem Datensatz überprüft. Sowohl für die Trainingsdaten als auch für die Testdaten werden die Gütekriterien (R^2 , RMSE) angegeben.

3 Ergebnisse

Abbildung 1 a und b stellt Boxplots für R^2 bzw. RMSE dar. Die Regressions-/Machine-Learning-Modelle sind in unterschiedlichen Graustufen dargestellt. Untereinander sind verschiedene Annahmen für die Messunsicherheit getroffen.

Wie zu erwarten, wird R^2 im Mittel kleiner, je größer die simulierte Messunsicherheit ist. Darüber hinaus fällt auf, dass im Random-Forest-Modell ein Overfitting stattfindet, das umso größer ist, je größer die Messunsicherheit wird, was auch zu einer schlechteren Anpassung an die Testdaten als im linearen Modell führt. Für Random-Forest-Modelle führt eine geringe Anzahl von Parametern zu schlechten Anpassungen, da im ersten Schritt 1/3 der Parameter zur Konstruktion des Baumes verwendet wird. Die Ergebnisse des linearen Modells unterscheiden sich darüber hinaus kaum von den Ergebnissen der Machine-Learning-Modelle, das ist auch dadurch zu begründen, dass in der Simulation ein linearer Zusammenhang angenommen wird.

Für die Testdaten ergeben sich im linearen Modell für die Messunsicherheiten $RMU_{max} / RMU_{min} = 0,1/0,05 \mid 0,2/0,1 \mid 0,32/0,13 \mid 0,4/0,2$ R^2 -Werte von im Median $0,93 \mid 0,77 \mid 0,57 \mid 0,44$. Für die Modelle SVM und ANN ergeben sich sehr ähnliche Werte, lediglich für das RF-Modell sind die R^2 -Werte noch etwas schlechter. Eine Erhöhung der Anzahl der simulierten Datensätze zeigte in einer weiteren Simulation keine Verbesserung der Werte.

Die RMSE-Werte betragen dagegen im Median $0,34 \mid 0,65 \mid 1,02 \mid 1,35$. Hier ist auffällig, dass die Anpassung des ANN-Modells weniger stark von der Höhe des RMU abhängig ist. Hier beträgt der Median des RMSE auch für den höchsten angenommenen RMU ($0,4/0,2$) lediglich $0,79$.

4 Diskussion

Die vorgestellte Simulation zeigt, dass der Labormessfehler einen wesentlichen Einfluss auf die Modellgüte hat. So wurde in der Simulation ein perfekter linearer Zusammenhang durch das Hinzufügen eines zufälligen Labormessfehlers einer bestimmten Größenordnung verfälscht. Die Simulationsergebnisse im linearen Modell zeigen, dass schon bei einem relativen Labormessfehler von $28,7\% - 13,1\%$, wie er von der VDLUFA beschrieben wird, die Anpassungsgüte R^2 im Median nur noch einen Wert von $0,57$ erreicht. Bei 100 Simulationen wurde maximal ein Wert von $0,78$ für die Anpassungsgüte R^2 beobachtet. Im Vergleich dazu lag die Anpassungsgüte bei den Originaldaten bei $0,68$

und damit um 0,1 über dem Median der Simulation. Dies ist u. a. durch zusätzliches „Rauschen“ der Sensordaten zu erklären, das in der Simulation nicht berücksichtigt wurde. Durch die Simulation wird deutlich, dass aufgrund des vorhandenen Labormessfehlers selbst bei einem perfekten Zusammenhang keine Anpassungsgüte zu erwarten ist, die über 0,78 hinausgeht. Beim Vorliegen weiterer Fehler beispielsweise bei der Probennahme oder Verfahrensfehlern im Labor muss mit einer weiteren Verschlechterung der Anpassungsgüte gerechnet werden.

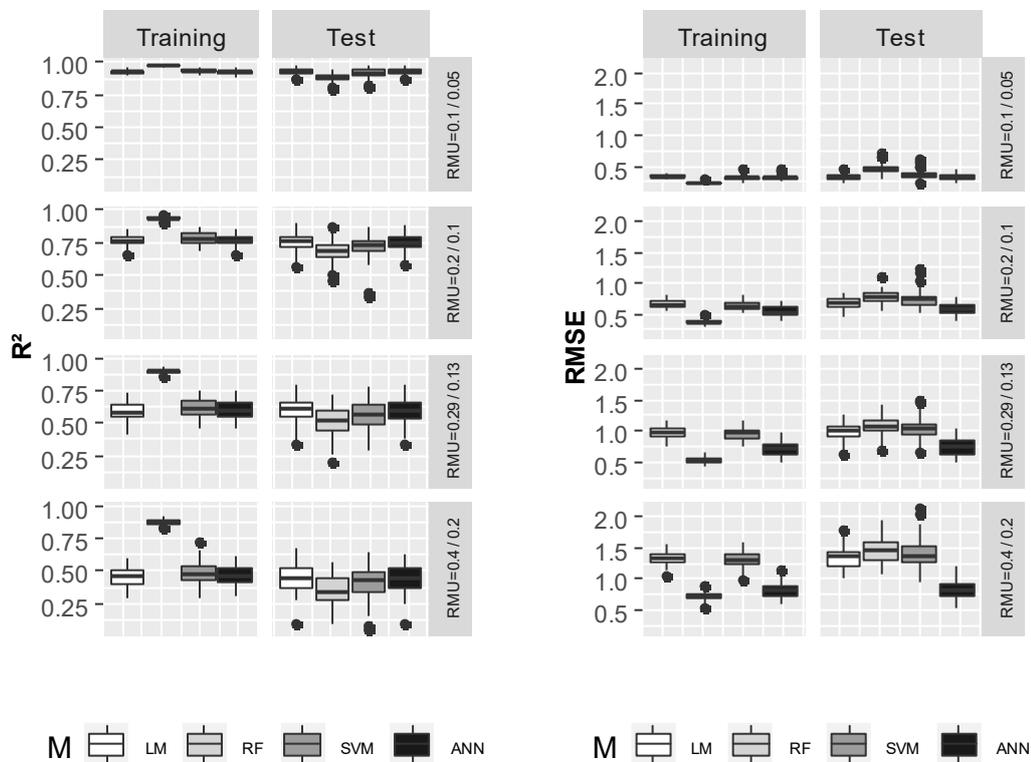


Abb. 1a: R² nimmt mit zunehmender Labormessunsicherheit ab

Abb. 2b: RMSE nimmt mit zunehmender Labormessunsicherheit zu

Insbesondere für die Kalibrierung von Sensordaten mittels Laborwerten gilt es darauf zu achten, den Labormessfehler zu kontrollieren und ggf. Methoden zu finden, um ihn zu minimieren, um eine zuverlässige Anpassung zu erreichen.

Des Weiteren legt die vorliegende Simulation den Schluss nahe, dass Publikationen, die eine Anpassungsgüte von nahezu 1 angeben, mit Vorsicht zu betrachten sind. Zwar verkleinert sich aufgrund des Gesetzes der großen Zahlen der Einfluss des Labormessfehlers je größer der verwendete Datensatz ist, jedoch ist beim Einsatz von Labormessungen die Größe des Datensatzes typischerweise begrenzt. Somit ist nicht auszuschließen, dass es sich bei derart perfekten Ergebnissen um Artefakte, Zufallsfunde oder auch unzulässige Selektionen von Teildaten handelt. Es ist jedoch auch nicht möglich, ein allgemein gültiges, maximal erreichbares R^2 anzugeben. Ein solches kann jedoch für andere Zusammenhänge ebenso durch eine Simulation ermittelt werden.

Literaturverzeichnis

- [CSW18] Chlingaryan, A., Sukkarieh, S. and Whelan, B.: Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture* 151, 61-69, 2018.
- [Co04] Cools, N., Delanote, V., Scheldeman, X., Quataert, P., De Vos, B., and Roskams, P.: Quality assurance and quality control in forest soil analyses: A comparison between European soil laboratories. *Accredit. Qual. Assur.* 9, 688-694, 2004.
- [FGW19] Fritsch, S., Guenther, F. and Wright, M.N.: neuralnet: Training of Neural Networks. R package version 1.44.2, 2019.
- [Ja02] Jacobsen, J. S., Lorbeer, S. H., Schaff, B. E., and Jones, C. A.: Variation in soil fertility test results from selected northern great plains laboratories. *Commun. Soil Sci. Plant Anal.* 33, 303-319, 2002.
- [LW02] A. Liaw and M. Wiener: Classification and Regression by randomForest. *R News* 2(3), 18-22, 2002.
- [Me19] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. and Leisch, F.: e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-1, 2019.
- [Mu07] Munzert, M., Kießling, G., Übelöhr, W. Nätcher, L. and Neibert, K.-H.: Expanded measurement uncertainty of soil parameters derived from proficiency-testing data. *J. Plant Nutr. Soil Sci.* 2007, 170, 722-728, 2007.
- [Vi11] Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J. and Lobsey, C.: Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. In: *Advances in Agronomy*. Vol. 113, 237-282, 2011.
- [WJH96] Wolf, A. M., Jones, J. B., and Hood, T.: Proficiency testing for improving analytical performance in soil testing laboratories: A summary of results from the council's soil and plant analysis proficiency program. *Commun. Soil Sci. Plant Anal.* 27, 1611-1622, 1996.