

Multivariate statistische Methoden als Werkzeug für eine vereinfachte Auswertung abfallwirtschaftlicher Fragestellungen

Katharina Meissl¹, Ena Smidt¹, Johannes Tintner¹ und Manfred Schwanninger²

¹ Institut für Abfallwirtschaft, Department für Wasser, Atmosphäre und Umwelt
Universität für Bodenkultur Wien;

² Department für Chemie, Universität für Bodenkultur Wien

Multivariate data analyses as an evaluation tool for waste management questions

Abstract

Multivariate data analysis is a helpful tool to handle huge data sets. Fourier transform infrared (FTIR) spectroscopy which is applied in waste management research generates many data points for each sample. Multivariate data analysis was used for parameter prediction by means of infrared spectroscopy. Models were developed to predict humic acid contents and respiration activity of composts as well as respiration activity and gas forming potential of mechanically-biologically pretreated (MBT) waste. Furthermore, a classification model was developed to distinguish waste materials like bio-waste compost, MBT-material and old landfill material by means of FTIR spectroscopy. Stabilization effects by in-situ aeration of an abandoned landfill were monitored using FTIR spectroscopy. Cluster analysis was successfully applied to differentiate between initial and aerated samples. The present study gives an overview of the mentioned evaluation tools based on multivariate data analysis.

Zusammenfassung

Bei zahlreichen abfallwirtschaftlichen Fragestellungen wird eine große Datenmenge generiert. Zur Unterstützung der Interpretation werden multivariate statistische Methoden verwendet. Die Infrarotspektroskopie ist eine in der Abfallwirtschaft bereits erprobte Analysemethoden. Zahlreiche Untersuchungen werden mittels Infrarotspektroskopie und multivariaten statistischen Methoden gemacht. Bereits erprobt ist eine Klassifizierung von Infrarotspektren bezüglich ihrer Zugehörigkeit zu den Materialklassen „Bioabfallkompost“, „MBA-Material“ und „Altablagerungs-Material“. Weiters sind Vorhersagemodelle für die Bestimmung des Gehaltes an Huminsäuren und der Atmungsaktivität für Bioabfallkomposte entwickelt worden. Für MBA-Material ist ein vorläufiges Modell für die Vorhersage der Atmungsaktivität und Gasbildung erarbeitet worden. Ein weiteres bereits erprobtes Anwendungsgebiet ist das Monitoring des Belüftungserfolges einer In-Situ Aerobisierung einer Altablagerung basierend auf der Charakteristik der Infrarotspektren. Im Zuge dieser Arbeit ist ein Überblick über diese Anwendungen gegeben.

Keywords

Multivariate Datenanalyse, Klassifizierung, Vorhersagemodelle Fourier Transformation Infrarot Spektroskopie, Multivariate data analysis, classification, parameter prediction, Fourier transform infrared spectroscopy

1 Einleitung

In allen abfallwirtschaftlichen Bereichen werden große Datenmengen generiert. Zur besseren Überschaubarkeit und Unterstützung der Interpretation, kann man sich der multivariaten Statistik bedienen (Esbensen, 2002). Dabei werden die analysierten Daten visuell als Punkte in einem Raum „anschaulich“ dargestellt. Zum Beispiel werden bei der multivariaten statistischen Methode der Hauptkomponentenanalyse Variablen mit vielen Eigenschaften auf einige wenige latente Faktoren reduziert. Weiters ist die Anwendung von multivariaten Verfahren dann vorteilhaft, wenn das Resultat einer Messung nicht ein Einzelwert ist, sondern viele Messpunkte aufgenommen werden (z.B. Infrarotspektrum). Auf der Basis multivariater Statistik können Modelle für verschiedene Fragestellungen entwickelt werden. Zur raschen quantitativen Bestimmung aufwändiger Parameter eignen sich Modelle zur Parametervorhersage aus dem Spektrum. Bei der Verwendung der Infrarotspektroskopie ist die Anwendung der multivariaten statistischen Verfahren für die Auswertung sehr hilfreich. Die Analysenmethode liefert pro Probe über 1000 Datenpunkte. Moron und Cozzolino (2004), Pöllänen et al. (2005) und Zahng et al. (2005) bedienen sich in ihren Arbeiten auch der multivariaten Datenanalyse. Bereits erprobt ist eine Klassifizierung von Infrarotspektren bezüglich ihrer Zugehörigkeit zu den Materialklassen „Bioabfallkompost“, „MBA-Material“ und „Altablagerungs-Material“ (Smidt et al., in press). Weiters wurde über Infrarotspektren ein Modell zur Vorhersage des Gehaltes an Huminsäuren und der Atmungsaktivität für Bioabfallkomposte erstellt (Meissl et al., 2007). Für MBA-Material sind noch nicht validierte Modelle für die Vorhersage der Atmungsaktivität und Gasbildung erarbeitet worden. Ein weiteres Anwendungsbeispiel ist das Monitoring des Belüftungserfolges einer In-Situ Aerobisierung einer Altablagerung basierend auf der Charakteristik der Infrarotspektren (Tesar et al., 2007). Es soll im Zuge dieser Arbeit ein Überblick über die bereits bestehenden Auswertungen und aktuellen Fragestellungen gegeben werden.

2 Material und Methoden

2.1 Materialien

Für die Analysen wurde das Material Kompost luftgetrocknet und in der Scheibenschwingmühle gemahlen. Der mechanisch-biologisch behandelte Restmüll (MBA-Material) und das Altablagerungsmaterial wurden bei 105°C getrocknet und mit der Zentrifugalmühle gemahlen. Alle Materialien wurden nach dem Mahlschritt auf <0,63 mm abgeseibt.

2.2 Analytische Methoden

Für die infrarotspektroskopischen Untersuchungen wurden KBr (Kaliumbromid) Presslinge mit 200 mg KBr und 2 mg Probe für 3 Minuten unter einem Druck von 75 kN/cm² angefertigt und im Transmissionsmodus unter Umgebungsbedingungen gemessen. Dabei wird der Wellenzahlbereich von 4000-400 cm⁻¹ aufgenommen. Es werden 32 Scans gemittelt und gegen den Hintergrund korrigiert.

Die Atmungsaktivität wurde im Sapromat (Voith Sulzer) nach BINNER ET AL. (1998) gemessen. Die Gasspendensumme wird im Inkubationsversuch nach BINNER ET AL. (1998) bestimmt.

Die Huminsäurebestimmung erfolgte nach GERZABEK ET AL. (1993).

2.3 Multivariate Datenanalyse

2.3.1 Hauptkomponentenanalyse (PCA)

Bei einer Hauptkomponentenanalyse versucht man, aus Variablen mit vielen Eigenschaften einige wenige latente Faktoren zu extrahieren, die für diese Eigenschaften bestimmend sind. Mathematisch wird eine Hauptachsentransformation durchgeführt: Man minimiert die Korrelation mehrdimensionaler Merkmale durch Überführung in einen Vektorraum mit neuer Basis. Die Daten liegen dabei als Punktwolke in einem n-dimensionalen kartesischen Koordinatensystem vor. Die erste Achse (1. Hauptkomponente = PC 1) wird so durch die Punktwolke gelegt, dass die Varianz der Daten in dieser Richtung maximal ist. Die Varianz von Daten ist ein Maß für ihren Informationsgehalt. Die zweite Achse (2. Hauptkomponente = PC 2) steht senkrecht auf die ersten Achse. In ihrer Richtung ist die Varianz am zweitgrößten usw. Für die n-dimensionale Datenmatrix gibt es also grundsätzlich n viele Achsen, die aufeinander senkrecht stehen. Sie sind orthogonal. Die Gesamtvarianz der Daten ist die Summe dieser "Achsenvarianzen". Die Scores zeigen die Position der Proben entlang der Hauptkomponenten des Modells. Die Loadings zeigen, durch welche Variablen und wie stark eine Hauptkomponente beeinflusst wird.

2.3.2 Parametervorhersage (PLS-R)

Partial Least Squares Regression (PLS-R) Modelle basieren auf den Hauptkomponenten sowohl der unabhängigen Variablen X, als auch der abhängigen Variablen Y. Die zentrale Idee besteht darin, für die Matrices X und Y die Hauptkomponenten getrennt zu berechnen und ein Regressionsmodell zwischen den Scores der Hauptkomponenten zu erstellen. Eine geeignete Modellvalidierung ist durchzuführen. In dieser Arbeit wurde eine volle Kreuzvalidierung durchgeführt und zur Prüfung der Stabilität des Modells wurde das Datenset geteilt (1:1 und eine 2:1). Bei der Teilung des Datensets wird der

eine Teil als Kalibrationsset und der andere als Testset verwendet und umgekehrt. Wenn das Modell stabil ist, sind die verwendete Anzahl an Hauptkomponenten, das Bestimmtheitsmaß und die Vorhersagefehler bei allen Validierungsvarianten sehr ähnlich.

2.3.3 Clusteranalyse

Bei der Clusteranalyse werden die spektralen Distanzen zwischen allen Spektren berechnet, um wie bei der PCA die Datenstruktur zu erfassen. Die beiden Spektren mit der größten Ähnlichkeit (d.h. die Spektren mit der kleinsten spektralen Distanz) werden in einem Cluster zusammengefasst. Anschließend werden die Distanzen zwischen diesem Cluster und allen anderen Spektren berechnet. Dazu stehen mehrere mathematische Methoden (Ward's Algorithm, Single Linkage, Complete Linkage...) zur Verfügung. Diese Prozedur wird solange wiederholt, bis nur noch ein großes Cluster übrig bleibt. Die Distanzen der Cluster zueinander kann dann interpretiert werden.

3 Ergebnisse und Diskussion

Im Folgenden werden einige Anwendungsbeispiele (Forschungsergebnisse) erläutert.

3.1 Klassifizierung von Bioabfallkomposten basierend auf ihrer Charakteristik im Infrarotspektrum

Um festzustellen, ob sich die Materialien Kompost, MBA-Material und Altablagerungsmaterial in ihrer spektroskopischen Charakteristik unterscheiden, wurde eine Hauptkomponentenanalyse angewandt. Spektrale Bereiche, die zur Trennung der verschiedenen Abfallmaterialien verwendet wurden, sind 3000-2800, 1800-1394, und 1375-400 cm^{-1} . In diesen Bereichen liegen die Absorptionsbanden der funktionellen Gruppen, die die „chemischen“ Unterschiede der verschiedenen Materialien widerspiegeln. PC 1 in Abbildung 1c zeigt den Einfluss der Organikbanden bei 2920, 2850 und 1640 cm^{-1} sowie den Einfluss der Anorganik (Carbonat bei 1430 und 875 cm^{-1}). PC 2 zeigt Unterschiede bei den Wellenzahlen 1640 und 1170-900 cm^{-1} , die organischen Verbindungen zugeordnet werden können. Diese Unterschiede sind durch die verschiedenen Inputmaterialien bei MBA- und Altablagerungs - Material zurückzuführen und zeigen sich in einer großen Streuung der MBA-Materialien und der Altablagerungen im Scores plot (Abbildung 1a). Die dritte Hauptkomponente (PC 3) trennt die Komposte deutlich von den anderen beiden Materialien. Der Loadings plot (Abbildung 1d) zeigt, dass dafür Wellenzahlenbereiche verantwortlich sind, die vorwiegend organischer Natur sind. Der Pfeil in Abbildung 1b soll zeigen, dass hier eine zeitliche Entwicklung von MBA-Material zu Altablagerungen durch Mineralisierung/ Stabilisierung stattfindet.

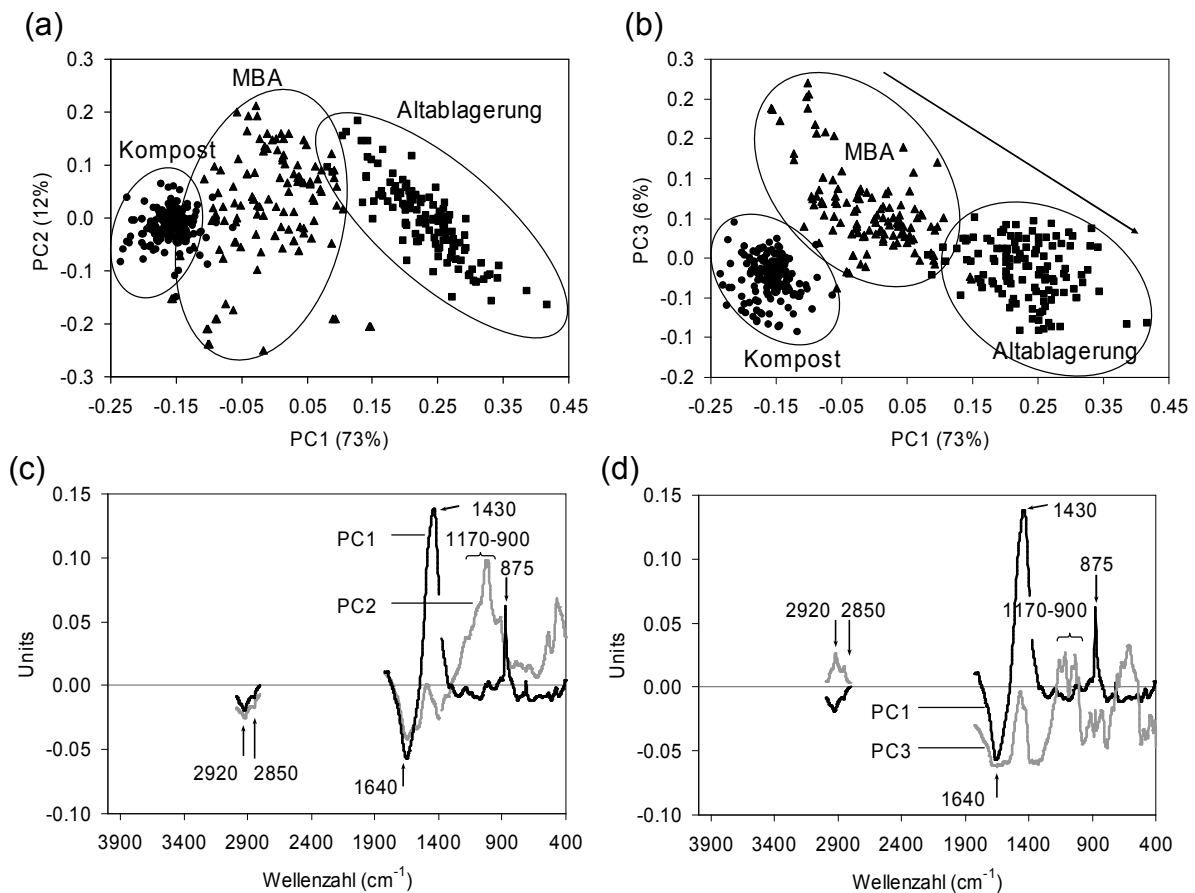


Abbildung 1 PCA der untersuchten Abfallarten: Kompost, MBA-Material, Altablagerungen (a-b) Scores plots, (c-d) zugehörige Loading plots

Auf Grundlage der in Abbildung 1 dargestellten Hauptkomponentenanalyse wurde ein Klassifizierungsmodell erstellt. Die Definition der Klasse „Kompost“ orientiert sich an den österreichischen Komposten mit getrennt gesammeltem biogenem Inputmaterial, das in einem aeroben, biologischen Prozess behandelt wurde. Die angestrebte Qualität der Organik setzt die getrennte Sammlung der Bioabfallfraktion und eine klare Unterscheidung zu MBA-Material voraus. Basierend auf den Infrarotspektren wurde für jede definierte Klasse ein Modell erstellt. Die Abbildung 2a zeigt die Modelldistanzen der Klasse „Komposte“ und „MBA-Materialien“ zu einander. Als Testset wurden Spektren „unbekannter“ Komposte mit diesem Modell klassifiziert. In Abbildung 2b ist zu erkennen, dass mehrere Komposte aufgrund ihrer Behandlungsverfahren oder Ausgangsmaterialien (z. B. Anaerobrückstände oder Klärschlammkomposte) nicht dem für das Modell definierten Kompostkriterium entsprechen und nicht der Klasse Bioabfallkomposte zugeordnet werden. Von SMIDT ET AL. (in, press) wurden noch weitere Testsetvalidierungen zur Prüfung des Modells durchgeführt.

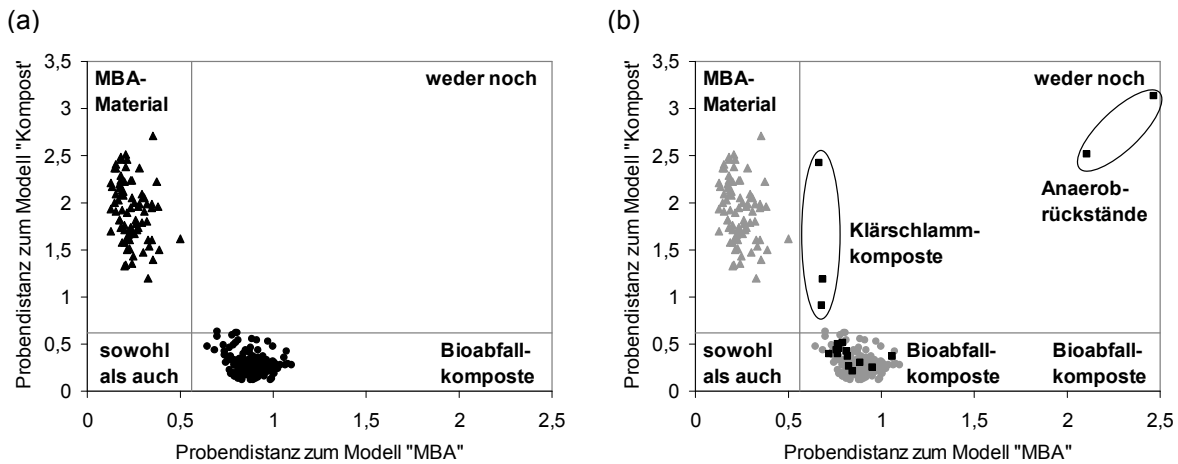


Abbildung 2 (a) Klassifikationsmodell für Bioabfallkomposte und MBA – Material basierend auf ihren Infrarotspektren; (b) Anwendungsbeispiel des Klassifizierungsmodells (unbekannte Proben = dunkle Quadrate)

Dieses Klassifizierungsmodell ermöglicht nun eine einfache und schnelle Zuordnung von unbekanntem Proben zu den definierten Abfallmaterialien.

3.2 Bestimmung des Huminsäuregehaltes und der Atmungsaktivität in Bioabfallkomposten mittels Vorhersagemodell basierend auf der Infrarotspektroskopie

Die quantitative Vorhersage von Parametern aus dem Spektrum ist besonders dann interessant, wenn die Referenzmethode zeitaufwändig, störungsanfällig oder schwierig ist. Es war daher eine Forschungsfrage, den Huminsäuregehalt und die Atmungsaktivität von Komposten mithilfe des Infrarotspektrums und einem Vorhersagemodell zu bestimmen. Zur quantitativen Vorhersage eignet sich die Partial Least Squares – Regression (PLS - Regression). Eine ausführliche Abhandlung der Modellerstellung und Modellvalidierung wurde von MEISSEL ET AL. (2007) publiziert. Im Folgenden sind die aktualisierten Vorhersagemodelle näher erläutert.

Abbildung 3 zeigt den ausgewählten Bereich aus den Infrarotspektren für die quantitative Vorhersage der Huminsäuren. Die Bande bei 1595 cm^{-1} entsteht durch die Humifizierung während des Kompostierprozesses. Diese Bande ist den aromatischen Verbindungen zuzuordnen. Die Carbonylgruppe verschiedener Verbindungen absorbiert im Wellenzahlbereich zwischen 1740 und 1720 cm^{-1} . Diese Verbindungen nehmen mit zunehmender Stabilisierung/ Humifizierung ab. Die Abnahme der entsprechenden Bande im Spektrum ist daher ein weiterer Indikator für den Stabilisierungsprozess.

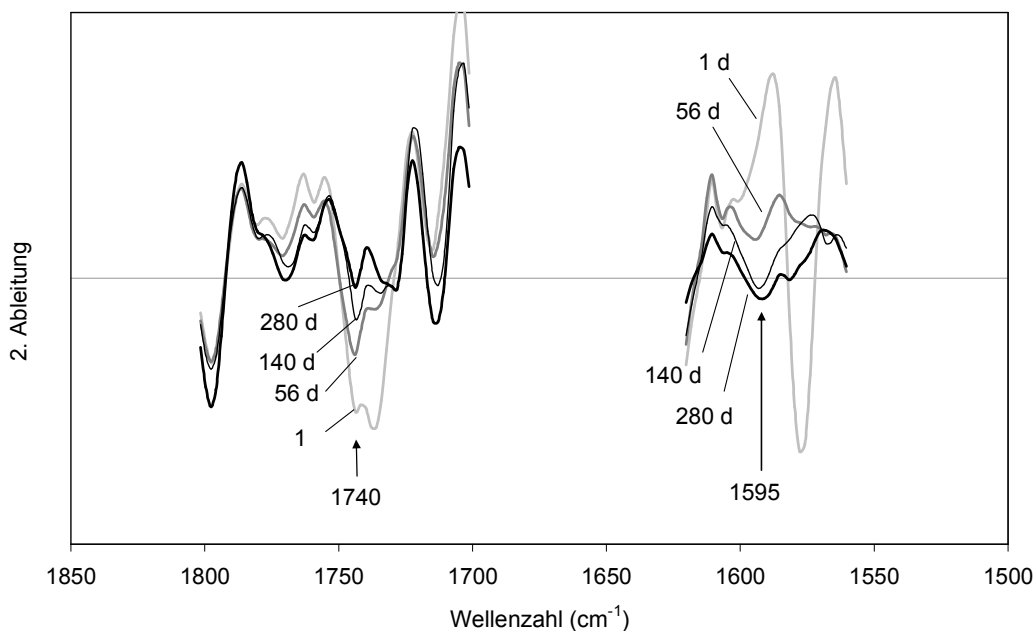


Abbildung 3 Ausgewählte Bereiche der Infrarotspektren (2. Ableitung) für die quantitative Vorhersage der Huminsäuregehalte

Die für die Erstellung des Vorhersagemodells der Atmungsaktivität verwendeten Wellenzahlenbereiche 3000-2800, 1788-1533, 1348-1201 cm^{-1} wurden in Abbildung 4 hervorgehoben. Diese können organischen Verbindungen zugeordnet werden und eine Veränderung durch Stabilisierungs- und Mineralisierungsprozesse während des Kompostierprozesses wird beobachtet. Die Anorganikbanden Carbonat bei 1420 und 875 cm^{-1} , Nitrat bei 1384 cm^{-1} und Silikat bei 1030 cm^{-1} wurden nicht verwendet.

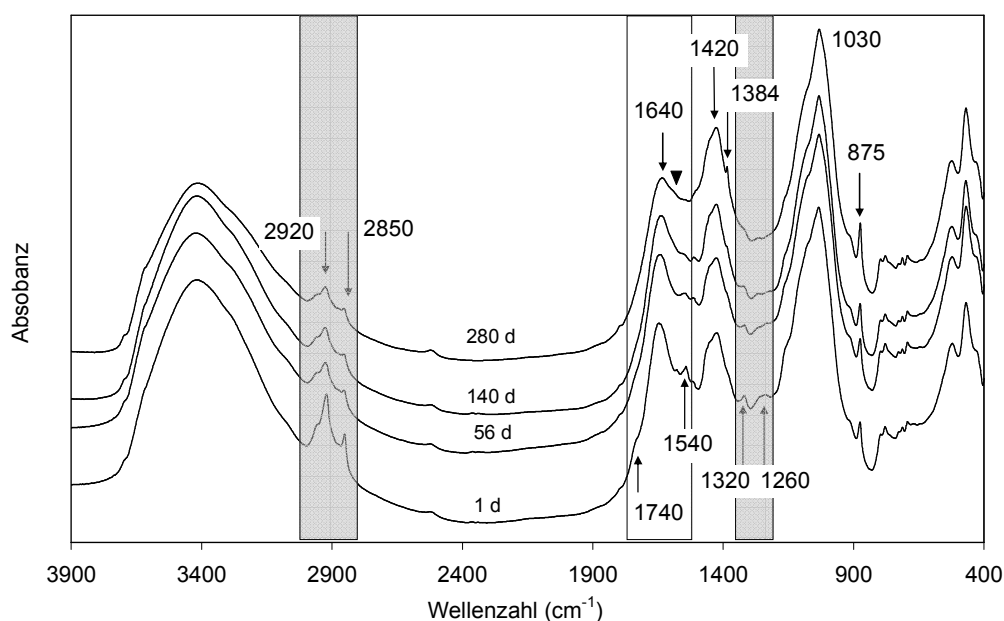


Abbildung 4 Ausgewählte Bereiche der Infrarotspektren für die quantitative Vorhersage der Atmungsaktivität

Abbildung 5 zeigt die Vorhersagemodelle für Huminsäuregehalte und Atmungsaktivitäten. Das Huminsäuremodell (HS - Kompost) beinhaltet 465, das Atmungsaktivitätsmodell (AT_4 - Kompost) 258 Kalibrationsproben. Die Kennzahlen der Modelle sind in Tabelle 1 zusammengefasst.

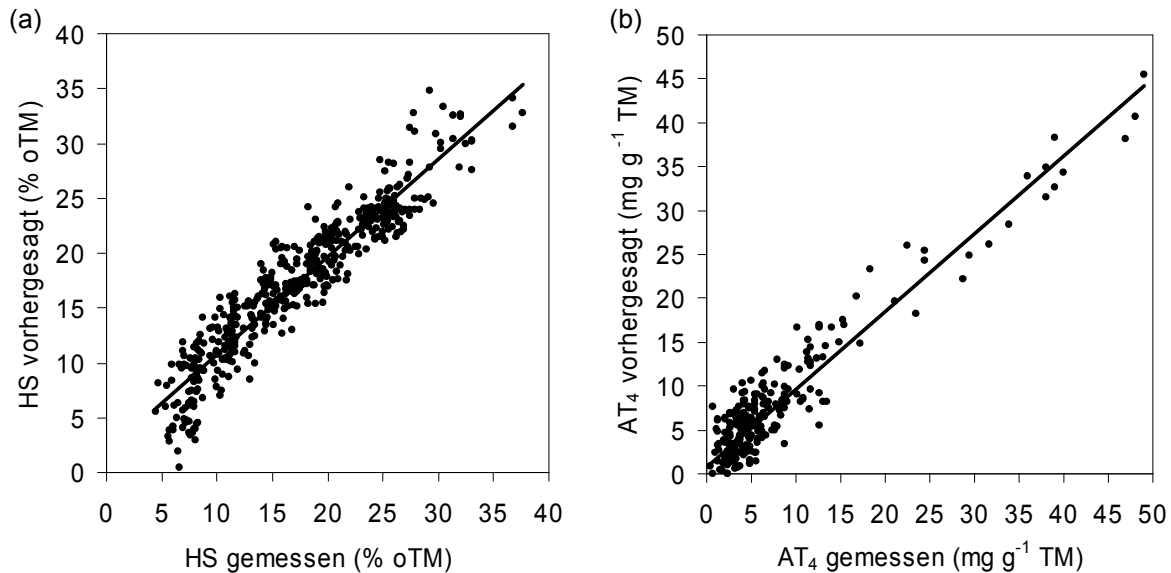


Abbildung 5 Vorhersagemodelle basierend auf einer Partial Least Squares Regression (PLS-R) von (a) den aus Infrarotspektren vorhergesagten und quantitativ bestimmten Huminsäuregehalten in Komposten, und (b) den aus Infrarotspektren vorhergesagten Atmungsaktivitäten und dem quantitativ bestimmten Sauerstoffverbrauch (AT_4)

Tabelle 1 Information zu den Modellen

Modell	HS - Kompost	AT_4 - Kompost
Bereich	0-40 % oTM	0-50 mg O_2 /g TM
Anzahl der verwendeten Proben	465	258
Regressionskoeffizient R^2	88	88
Mittlerer Vorhersagefehler	2,4 % oTM	2,9 mg O_2 /g TM

Die beiden vorgestellten Modelle sind bereits mittels Testsetvalidierung validiert und damit verifiziert.

3.3 Bestimmung der Atmungsaktivität und Gasspendensumme mittels Vorhersagemodell basierend auf der Infrarotspektroskopie

Da immer wieder Probleme bei der Bestimmung der Atmungsaktivität und der Gasspendensumme auftreten (BINNER, 2006), ist eine aktuelle Forschungsfrage, diese beiden biologischen Parameter über das Infrarotspektrum vorherzusagen. In Abbildung 6

werden jeweils die vorläufigen Vorhersagemodelle für die Atmungsaktivität (AT_4) und die Gasspendensumme (GS_{21}) dargestellt. Für die Berechnung wurden die Spektralbereiche 3000-2800, 1788-1533 und 1348-1201 cm^{-1} verwendet. Die Kennzahlen der Modelle sind in Tabelle 2 zusammengefasst.

Tabelle 2 Information zu den Modellen

Modell	AT_4 - MBA	GS_{21} - MBA
Bereich	0-55 mg O_2 /g TM	0-30 NI/kg TM
Anzahl der verwendeten Proben	75	41
Regressionskoeffizient R^2	0.9	0.6
Mittlerer Vorhersagefehler	3,9 mg O_2 /g TM	4,8 NI/kg TM

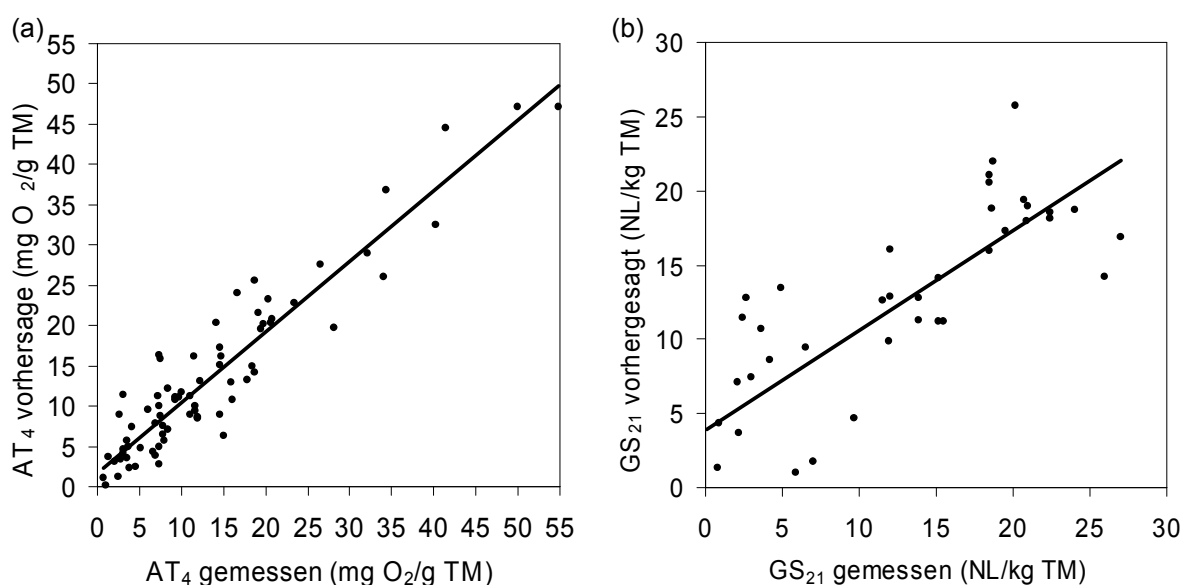


Abbildung 6 Vorhersagemodelle basierend auf einer Partial Least Squares Regression (PLS-R) von (a) den aus Infrarotspektren vorhergesagten und quantitativ bestimmten Atmungsaktivitäten (AT_4) und (b) den aus Infrarotspektren vorhergesagten und quantitativ bestimmten Gasspendensummen (GS_{21}) von MBA-Material

Diese beiden Modelle müssen noch mit zusätzlichen Proben erweitert werden. Weiters wird noch daran geforscht, welche Wellenzahlbereiche des Infrarotspektrums die Vorhersage der Parameter gegebenenfalls noch verbessern. Bei einer Probenzahl von etwa 100 können die Modelle validiert und einem Praxistest unterzogen werden.

3.4 Monitoring des Belüftungserfolges einer In-Situ Aerobisierung einer Altablagerung basierend auf der Charakteristik der Infrarotspektren

Zur Überprüfung des Belüftungserfolges (In-Situ Aerobisierung) einer Altablagerung wurde eine Clusteranalyse der aufgenommenen Infrarotspektren von Proben vor deren In-Situ Aerobisierung (original, I), sowie nach Belüftungsversuchen im Labor (stabilisiert, III und IV) durchgeführt (TESAR ET AL., 2007). Im dargestellten Beispiel wurde für die Erstellung des Modells der spektrale Bereich 3000-2800, 1790-1530, 1350-1200cm⁻¹ und der Ward's Algorithmus herangezogen (Abbildung 7). Die spektroskopische Charakteristik lässt weiters eine Unterscheidung der Länge der Belüftungszeit zu (vgl. Abbildung 7). Das Modell wurde anschließend mit Proben aus der belüfteten Altlast getestet. Die Testproben wurden mittels des erstellten Modells klassifiziert. Das Ergebnis der Klassifizierung wurde durch den Vergleich der Verteilungen stabilitätsrelevanter Parameter (z.B. AT₄, TOC,...) in den beiden Gruppen des Kalibrationssets und des Testsets geprüft. Die Gegenüberstellung der Verteilung für z.B. AT₄ und TOC ergab mittels t-test, dass die beiden Gruppen (original und stabilisiert) sich jeweils signifikant voneinander unterscheiden ($p < 0.001$).

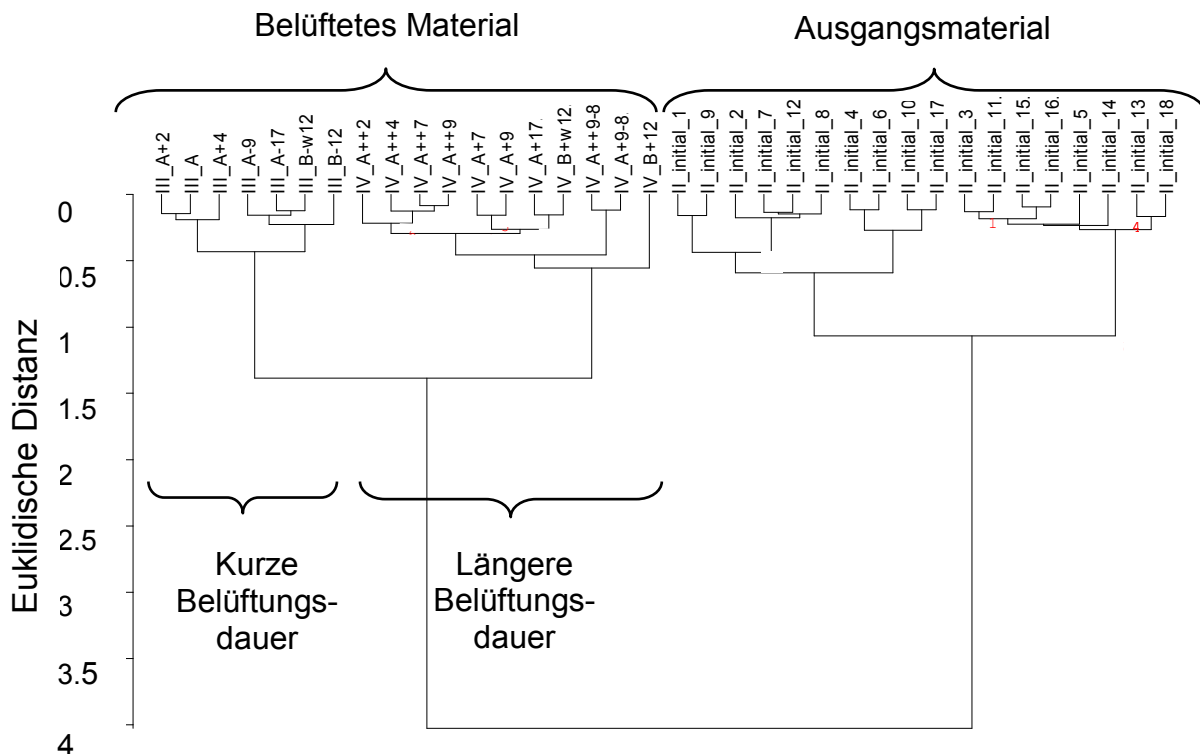


Abbildung 7 Clusteranalyse der Infrarotspektren von unbelüftetem und in-situ belüftetem Altablagerungsmaterial

4 Zusammenfassung

Multivariate statistische Methoden sind ein nützliches Werkzeug, um große Datenmengen leichter interpretieren zu können. Vor allem bei Analysen deren Ergebnis eine Kurve mit vielen Datenpunkten ist, haben sich diese Methoden bewährt. In der vorgestellten Arbeit wurden mit Hilfe dieser Methoden Modelle zur

- Klassifizierung von Kompost, MBA-Material und Altablagerungsmaterial,
- Vorhersage des Huminsäuregehalts und der Atmungsaktivität von Komposten,
- Vorhersage der Atmungsaktivität und der Gasspendensumme von MBA-Material, sowie
- Überprüfung des Belüftungserfolges einer In-Situ Aerobisierung einer Altablagerung entwickelt.

Diese Auswertemethoden sollen auch bei weiteren abfallwirtschaftlichen Fragestellungen und der Interpretation von großen Datensätzen eingesetzt werden.

5 Literatur

- | | | |
|---|------|---|
| Binner, E., Zach, A., Widenrin, M., und Lechner, P. | 1998 | Auswahl und Anwendbarkeit von Parametern zur Charakterisierung des Endproduktes aus mechanisch-biologischen Restmüllbehandlungsverfahren. Wien, Schriftenreihe des Bundesministeriums für Umwelt, Jugend und Familie |
| Binner, E. | 2006 | Beurteilung der Reaktivität von Abfällen - Anwendbarkeit der für mechanisch-biologisch behandelten Restabfall entwickelten Methoden bei anderen Abfällen. In: Matthias Kühle-Weidemeier (Eds.): Abfallforschungstage 2006 „Auf dem Weg in eine nachhaltige Abfallwirtschaft“, 16. – 17. May 2006, Hannover, Germany; Proceedings, Cuvillier Verlag, pp. 63-77 |
| Esbensen K. H. | 2002 | Multivariate Data Analysis -in practice, Alborg University, Esbjerg, 2002 |
| Gerzabek, M.H., Danneber, O., Kandeler, E. | 1993 | Bestimmung des Humifizierungsgrades. In Bodenbiologische Arbeitsmethoden. Schinner F., Öhlinger, R., Kandeler, E., Margesin, R., Eds; Springer Verlag, pp. 107-109 |
| Meissl, K., Smidt, E. und Schwanninger, M. | 2007 | Prediction of humic acid content and respiration activity of biogenic waste by means of Fourier Transform Infra-red (FTIR) spectra and partial least squares regression (PLS-R) models, Talanta 72, pp. 791-799 |

- Moron, A., Cozzolino, D., 2004 Determination of potentially mineralizable nitrogen and nitrogen in particulate organic matter fractions in soil by visible and near-infrared reflectance spectroscopy, *Journal of Agricultural Science* 142, pp. 335-343
- Pöllänen, K., Hakkinen, A., 2005 IR spectroscopy together with multivariate data analysis as a process analytical tool for in-line monitoring of crystallization process and solid-state analysis of crystalline product, *Journal of Pharmaceutical and Biomedical Analysis* 38, pp. 275-284
- Reinikainen, S.P., Rantanen, J., Karjalainen, M., Louhi-Kultanen, M., Nystrom, L.
- Smidt, E., Meissl, K., in press Classification of waste materials using Fourier transform infrared spectroscopy and soft independent modeling of class analogy, *Waste Management*, Schwanninger, M, Lechner, P.
- Tesar, M., Prantl, R., Lechner, P. 2007 Application of FT-IR for assessment of the biological stability of landfilled municipal solid waste (MSW) during in-situ aeration, *Journal of Environmental Monitoring* 9, pp. 110-118
- Zhang, A.D., Zeng, W.X., 2005 Multivariate analysis of infrared spectra for monitoring and understanding the kinetics and mechanisms of adsorption processes, *Applied Spectroscopy* 59, pp. 47-55
- Niemczyk, T.M., Keenan, M.R., Haaland, D.M.

Anschrift der Verfasserin

DI Katharina Meissl

Institut für Abfallwirtschaft, Department für Wasser, Atmosphäre und Umwelt, Universität für Bodenkultur Wien

Muthgasse 107

A-1190 Wien

Telefon +43 -1 381 99 00 - 345

Email: katharina.meissl@boku.ac.at

Website: www.wau.boku.ac.at/abf.html