

Statistische Verfahren SS 2016

Projekt 9 - Bodenatmung

Problemstellung:

Herleitung von Prognosemodellen für Bodenatmung im Nationalpark Hainich

Datensatz: (hainich.csv, A. Soe, MPI Biogeochemie)

Der Datensatz enthält Daten zur Bodenatmung (soil.res), d.h. zur Freisetzung von CO₂ aus dem Boden durch die Aktivität von Bodenlebewesen.

Die Bodenatmung hängt von abiotischen Bodenparametern, aber natürlich auch von der biologischen Umgebung, die die Zusammensetzung der Bodenfauna bestimmt, ab. Für die Modellierung der Bodenatmung in Waldklimamodellen werden einfache statistische Modelle zur Vorhersage der Bodenatmung benötigt. Dazu sind im Datensatz folgende potentiellen erklärenden Variablen enthalten:

- temp.0, temp.5, temp.10, temp.15 – Bodentemperatur in 0, 5, 10, 15 cm Tiefe
- smoi – relative Bodenfeuchte (soil moisture)
- lmoi – relative Feuchte der Streuschicht (litter moisture)
- litter.d – Dicke der Streuschicht
- ph.litter, ph.soil0, ph.soil5 – pH-Wert der Streuschicht bzw. des Bodens in 0-5 cm bzw. 5-10 cm Tiefe
- rootdw0, rootdw5 – Trockenmasse der Wurzeln in 0-5 cm bzw. 5-10 cm Tiefe
- rootn0, rootn5 – Stickstoffgehalt der Wurzeln in 0-5 cm bzw. 5-10 cm Tiefe
- rootc0, rootc5 – Kohlenstoffgehalt der Wurzeln in 0-5 cm bzw. 5-10 cm Tiefe
- litter.c, litter.n – Kohlenstoff- bzw. Stickstoffgehalt der Streuschicht
- soiln0, soiln5 – Stickstoffgehalt des Bodens in 0-5 cm bzw. 5-10 cm Tiefe
- soilc0, soilc5 – Kohlenstoffgehalt des Bodens in 0-5 cm bzw. 5-10 cm Tiefe
- litno3, sno30, sno35 – Nitratgehalt der Streuschicht bzw. des Bodens in 0-5 cm bzw. 5-10 cm Tiefe
- litnh4, snh40, snh45 – Ammoniumgehalt der Streuschicht bzw. des Bodens in 0-5 cm bzw. 5-10 cm Tiefe
- litdoc, sdoc0, sdoc5 – Gehalt an gelöstem organischen Kohlenstoff (DOC - dissolved organic carbon) der Streuschicht bzw. des Bodens in 0-5 cm bzw. 5-10 cm Tiefe
- nn – Entfernung der Probestelle zum nächsten Baum (nearest neighbour)

Aufgaben:

- Bestimmen Sie mit Hilfe eines geeigneten Variablenselektionsverfahrens ein Vorhersagemodell für die Bodenatmung.

Simulation:

- Das im ersten Teil ausgewählte Modell ist Ausgangspunkt für die Simulationsaufgabe. Simulieren Sie Pseudo-Beobachtungen der Zielgröße dieses Modells für eine Design-Matrix, die aus zufällig ausgewählten Zeilen des ursprünglichen Datensatzes besteht. Während der Vorwärtsselektion wird mehrfach eine Hypothese der Form

$$H_0: \beta_j = 0$$

überprüft, die im klassischen linearen Modell durch einen F-Test geprüft werden kann.

Bei der Modellwahl durch Vorwärtsselektion wird jedoch jeweils die absolut größte von mehreren möglichen Statistiken betrachtet. Simulieren Sie ausgehend vom in der Simulation bekannten „wahren“ Modell die Häufigkeitsverteilung des Maximums mehrerer F-Statistiken für die Hinzunahme einer weiteren Einflussgröße beim nächsten Schritt der Vorwärtsselektion und diskutieren Sie die Auswirkungen auf die Wahrscheinlichkeit einer irrtümlichen Ablehnung der Null-Hypothese.