

Mann-Whitney-U-Test: Tiefenstufen

| Variable | Mann-Whitney U-Test (Piezometer.sta) | | | | | | | | | |
|-------------------|---|---------------------|----------|-----------------|-----------------|-----------------|-----------------|---------------------|---------------------|------------------------|
| | Nach Var.: Tiefe Markierte Tests signifikant ab p < .05000 | | | | | | | | | |
| | Rgsumme Gruppe 1 | Rgsumme Gruppe 2 | U | Z | p-Niveau | Z korr. | p-Niveau | Gült. N Gruppe 1 | Gült. N Gruppe 2 | 2* eins. exakt p |
| Al | 71.0000 | 100.0000 | 26.00000 | -1.28038 | 0.200412 | -1.40564 | 0.159833 | 9 | 9 | 0.222419 |
| Ca | 100.0000 | 71.0000 | 26.00000 | 1.28038 | 0.200412 | 1.28038 | 0.200412 | 9 | 9 | 0.222419 |
| Cl | 78.5000 | 92.5000 | 33.50000 | -0.61812 | 0.536500 | -0.61843 | 0.536290 | 9 | 9 | 0.545701 |
| DOC | 56.0000 | 115.0000 | 11.00000 | -2.60491 | 0.009190 | -2.60491 | 0.009190 | 9 | 9 | 0.007775 |
| Fe | 67.0000 | 104.0000 | 22.00000 | -1.63359 | 0.102346 | -1.68294 | 0.092387 | 9 | 9 | 0.113492 |
| K | 93.0000 | 78.0000 | 33.00000 | 0.66227 | 0.507801 | 0.66364 | 0.506923 | 9 | 9 | 0.545701 |
| Mg | 106.0000 | 65.0000 | 20.00000 | 1.81019 | 0.070267 | 1.81019 | 0.070267 | 9 | 9 | 0.077005 |
| Mn | 88.5000 | 82.5000 | 37.50000 | 0.26491 | 0.791082 | 0.26628 | 0.790021 | 9 | 9 | 0.796174 |
| Na | 102.0000 | 69.0000 | 24.00000 | 1.45699 | 0.145121 | 1.45774 | 0.144914 | 9 | 9 | 0.161497 |
| NO3 | 92.0000 | 79.0000 | 34.00000 | 0.57396 | 0.565993 | 0.61319 | 0.539753 | 9 | 9 | 0.604813 |
| pH | 104.5000 | 66.5000 | 21.50000 | 1.67774 | 0.093399 | 1.67948 | 0.093060 | 9 | 9 | 0.093912 |
| Si | 91.0000 | 80.0000 | 35.00000 | 0.48566 | 0.627207 | 0.48566 | 0.627207 | 9 | 9 | 0.666475 |
| SO4 | 86.0000 | 85.0000 | 40.00000 | 0.04415 | 0.964784 | 0.04415 | 0.964784 | 9 | 9 | 1.000000 |
| El. Leitfähigkeit | 97.5000 | 73.5000 | 28.50000 | 1.05963 | 0.289316 | 1.06017 | 0.289067 | 9 | 9 | 0.297326 |

Für Stichprobengröße > 20 schnelle Annäherung der U-Statistik an Normalverteilung, deshalb Ausgabe des „Z korr.“ und p für Normalverteilung

Effekt der Tiefenstufen

| | Kruskall-Wallis | | Mann-Whitney | ANOVA |
|-----------------|-----------------|--------|--------------|--------|
| | H | p | p | p |
| Cl | 0.382461 | 0.5363 | 0.5365 | 0.4841 |
| DOC | 6.785575 | 0.0092 | 0.0092 | 0.0057 |
| NO ₃ | 0.375998 | 0.5398 | 0.5699 | 0.4289 |
| pH | 2.820637 | 0.0931 | 0.0934 | 0.0485 |

Effekt der Flächen

| | Kruskal-Wallis | | ANOVA |
|-----------------|----------------|--------|--------|
| | H | p | p |
| Cl | 15.17355 | 0.0005 | 0.0001 |
| DOC | 3.92982 | 0.1402 | 0.0984 |
| NO ₃ | 12.53475 | 0.0019 | 0.0001 |
| pH | 7.35143 | 0.0253 | 0.0061 |

ANOVA: Gewichtung der Mittelwerte

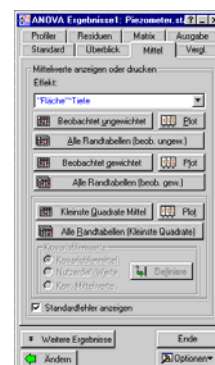
z.B. Berechnung des Mittelwerts der einzelnen Faktoren:

Ungewichtet:

- Mittelung der Mittelwerte aller Zellen ohne Berücksichtigung der u.U. unterschiedlichen Zahl der Wiederholungen
- = unverzerrte Schätzung der Randmittel der Grundgesamtheit

Gewichtet:

- direkte Berechnung aus den Einzeldaten => Gewichtung mit der Zahl der Wiederholungen
- = unverzerrte Schätzung des gewichteten Randmittelwertes der Grundgesamtheit



Nicht-lineare Methoden

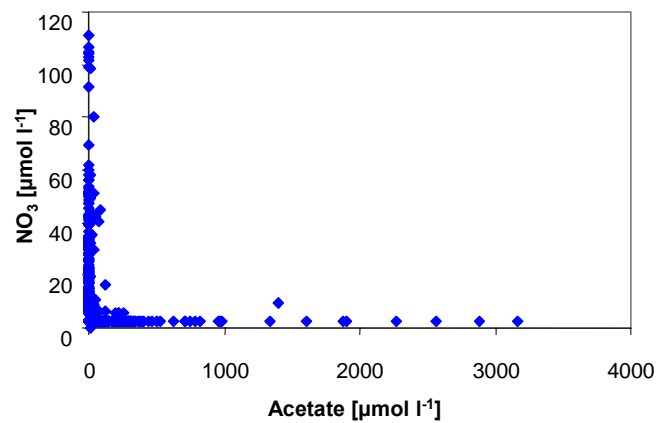
Nicht-lineare Zusammenhänge:

- Sättigungskurven
- Wechselseitiger Ausschluss (Bsp.: Redoxstufen)
- „lokale“ Zusammenhänge
- ...

Probleme nicht-linearer Methoden:

- Hoher Datenbedarf
- Schlechtere Interpretierbarkeit
- Höherer Rechenbedarf (oft iterative Methoden)

Wechselseitiger Ausschluss: NO₃ und Azetat



Eigenschaften (Künstlicher) Neuronaler Netze

- Aufbau aus einer Vielzahl einfacher, \pm identischer Einheiten (= Neuronen)
- Informationsverarbeitung durch Verbindungen unterschiedlicher Gewichtung zwischen Neuronen
- Verknüpfung iterativ optimiert durch "Training,, (Lernen)
- sinnvoll (nur) für große Datensätze
- verteilungsfrei
- nicht-linear

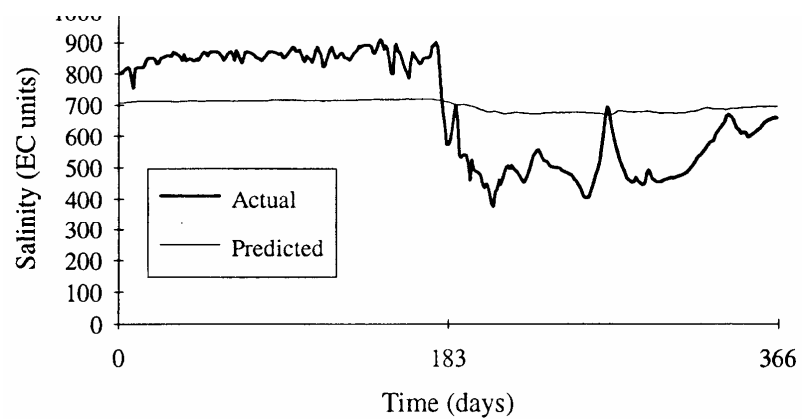
Künstliche Neuronale Netze

| | <u>lineares Verfahren</u> | <u>Künstliches Neuronales Netz</u> |
|---------------------------------------|---|------------------------------------|
| Vorhersage | lineare Regression | Mehrschicht-Perzeptron |
| Visualisierung | Korrespondenzanalyse Hauptkomponentenanalyse | Selbstorganisierende Karte |
| überwachte Klassifizierung | Diskriminanzanalyse | Lernende Vektor-Quantisierung |

Training (Lernen)

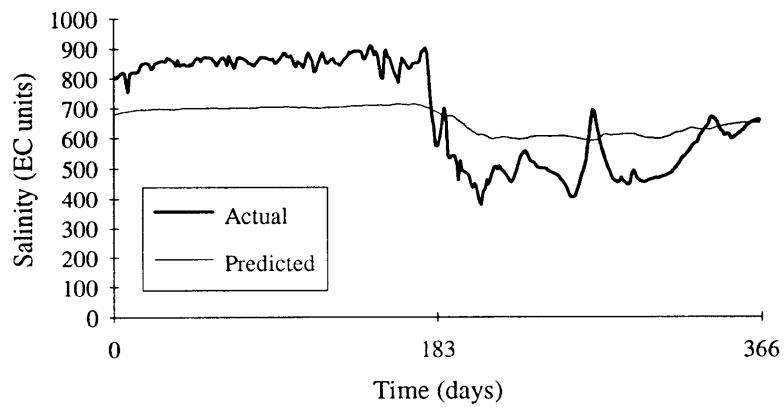
- = Anpassung der Gewichte zur Optimierung der Vorhersage
- „Überwachtes Lernen“ = iterative Optimierung der Relation zwischen Regressoren und Regressanden (= Soll-Output)
- während des Trainings passt sich das Netz zunehmend besser der Struktur des Datensatzes an

Nach 1.000 Trainingszyklen



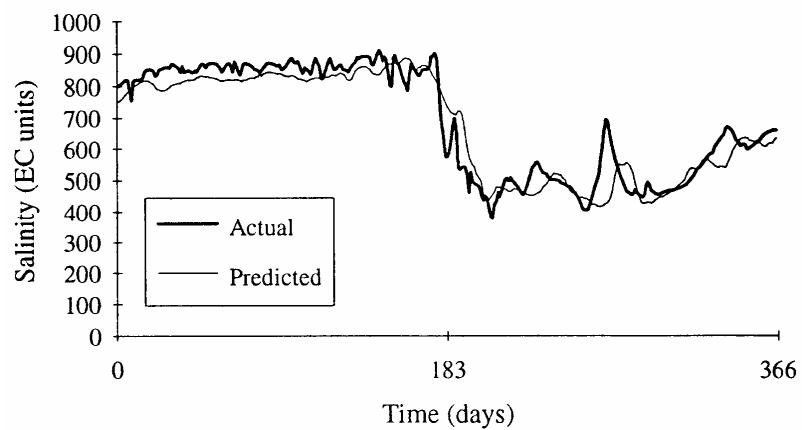
(Maier und Dandy 1995)

Nach 3.000 Trainingszyklen



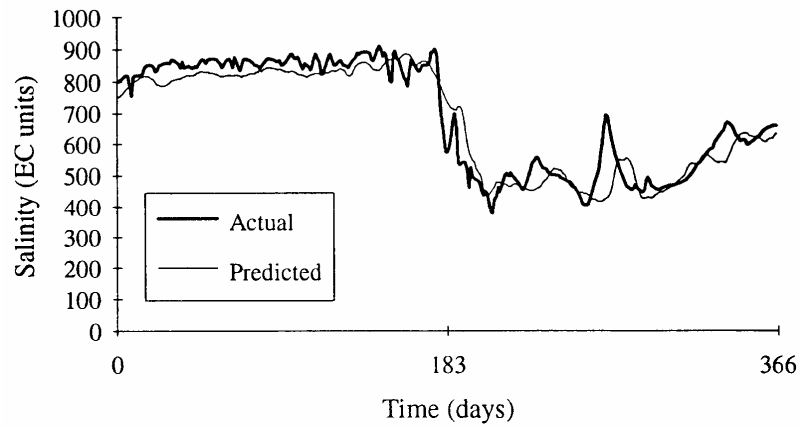
(Maier und Dandy 1995)

Nach 9.000 Trainingszyklen

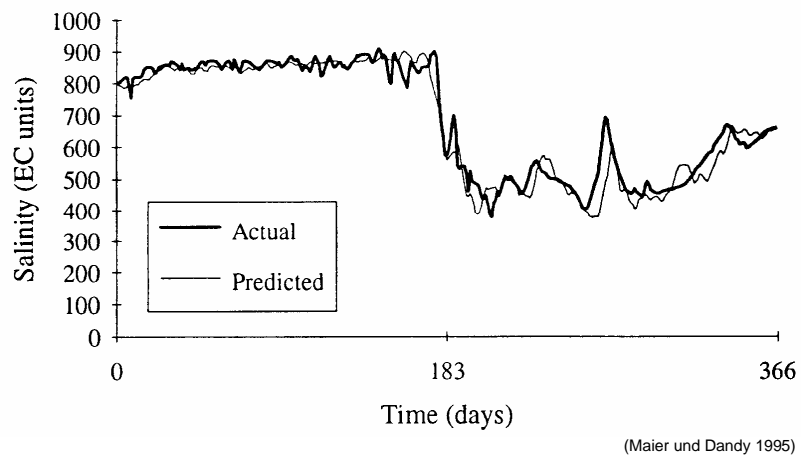


(Maier und Dandy 1995)

Nach 320.000 Trainingszyklen



3.400.000 Trainingszyklen



Overfitting (Overlearning)

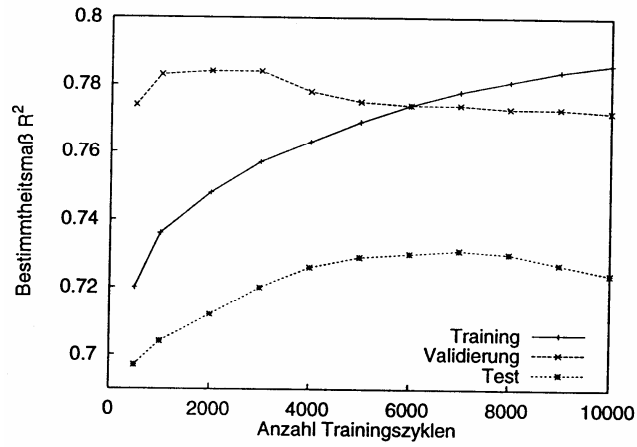
- während des Trainings passt sich das Netz zunehmend besser der Struktur des Datensatzes an
- mit zunehmender Dauer des Trainings werden überwiegend Strukturen des Trainingsdatensatzes gelernt, die aber in der Grundgesamtheit nicht enthalten sind
- Gegenmaßnahme: Kreuzvalidierung

Kreuzvalidierung

Aufteilung des Datensatzes in

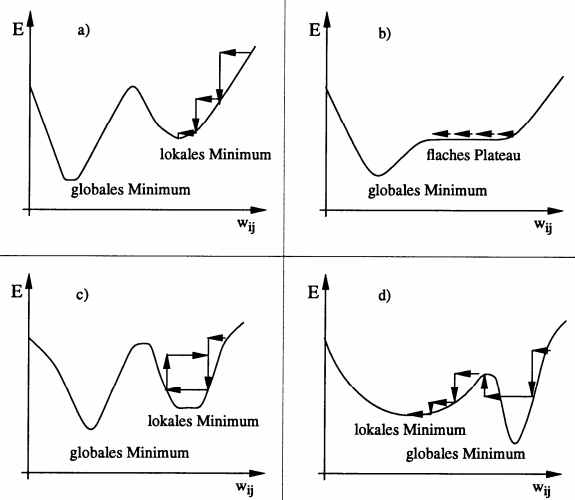
- **Trainingsdatensatz:** verwendet für Training
- **Validierungsdatensatz:** wiederholte Überprüfung des Netzfehlers *während* des Trainings (ohne Anpassung der Gewichte) zur Bestimmung des optimalen Netzes
- **Testdatensatz:** Überprüfung des Netzfehlers *nach* dem Training zur Überprüfung der Generalisierbarkeit des Netzes

Training



(Plötscher 1997)

Lernverfahren: Gradientenabstiegsverfahren



(Zell 1994)

Lernverfahren: Backpropagation

(Werbos 1974, Rumelhart et al. 1986)

Anpassung der Gewichte: $\Delta w_{ij} = \eta \cdot o_i \cdot \delta_j$ $0 < \eta < 1$ $\eta = \text{Lernrate}$

→ Fehler an vorgelagerte Schichten zurückgereicht:

$$\delta_j = \frac{\partial f(\text{net}_j)}{\partial \text{net}_j} \cdot (t_j - o_j) \quad \text{für Ausgabeknoten}$$

$$\delta_i = \frac{\partial f(\text{net}_i)}{\partial \text{net}_i} \cdot \sum_j (w_{ij} \cdot \delta_j) \quad \text{für Knoten der Zwischenschicht}$$

| | |
|-------------------------|--|
| w_{ij} | = Gewicht der Verbindung zwischen den Knoten i und j |
| net_j | = Inputwert des Knotens j |
| $a_j = f(\text{net}_j)$ | = Aktivierungswert des Knotens j |
| o_j | = Outputwert des Knotens j |
| t_j | = Sollwert des Knotens j |

für: $\text{net}_j = \sum_i (o_i \cdot w_{ij})$ $a_j = f(\text{net}_j) = \frac{1}{1 + e^{-\text{net}_j}}$ $o_j = a_j$

=> $\delta_j = o_j \cdot (1 - o_j) \cdot (t_j - o_j)$ für Ausgabeknoten

$\delta_i = o_i \cdot (1 - o_i) \cdot \sum_j w_{ij} \cdot \delta_j$ für Knoten der Zwischenschicht

Lernverfahren: Resilient Propagation

(Riedmiller und Braun 1992, 1993)

Prinzip:

- Vergleich der Steigung der Fehlerfunktion zum aktuellen und zum vorigen Zeitpunkt

- Gewichtsänderung $\Delta w_{ij}(s)$: $w_{ij}(s+1) = w_{ij}(s) + \Delta w_{ij}(s)$

- Betrag der Gewichtsänderung: $\Delta_{ij}(s) = |\Delta w_{ij}(s)|$

$$\Delta_{ij}(s) = \begin{cases} \Delta_{ij}(s-1) \cdot \eta^+ & \text{falls } E'(s-1) \cdot E'(s) > 0 \\ \Delta_{ij}(s-1) \cdot \eta^- & \text{falls } E'(s-1) \cdot E'(s) < 0 \\ \Delta_{ij}(s-1) & \text{falls } E'(s-1) \cdot E'(s) = 0 \end{cases}$$

- 1. Ableitung des Fehlers: $E'(s) = \frac{\partial E(s)}{\partial w_{ij}(s)}$

- Lernrate: $0 < \eta^- < 1 < \eta^+$

Optimierung des Netzes

Die Güte des Netzes (Regression, Vorhersageleistung, Generalisierungsleistung) hängt v.a. ab von:

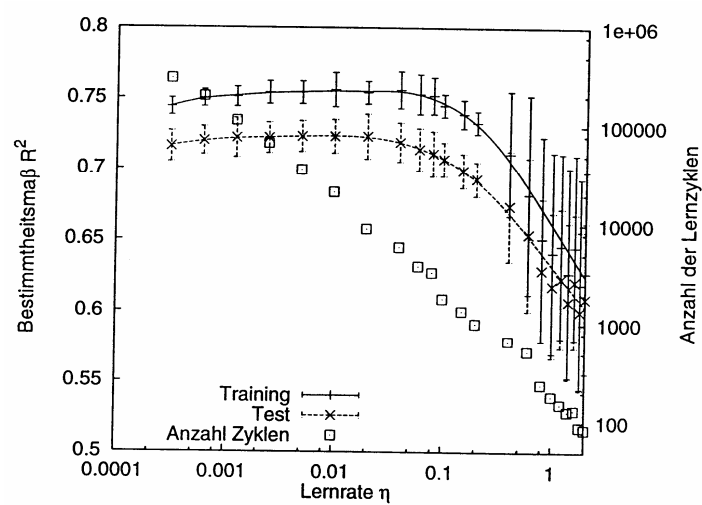
- **Lernverfahren:**

- Art
- Parametern
- Lerndauer

- **Topologie:**

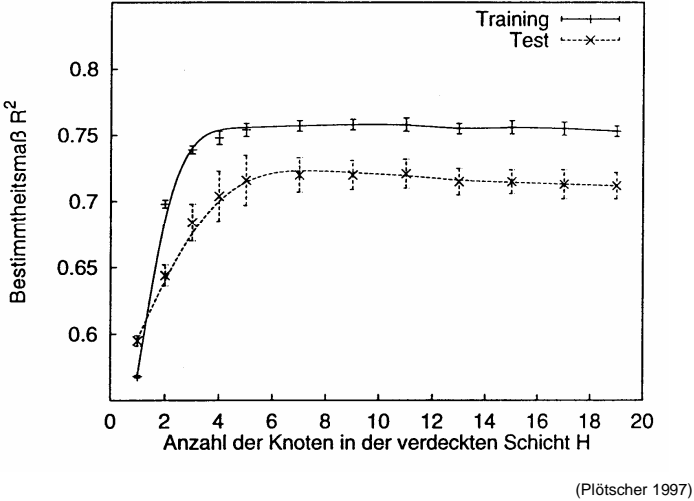
- Wahl der Regressoren
- Anzahl der versteckten Schichten
- Anzahl der Knoten in der/n versteckten Schicht(en)

Lernrate



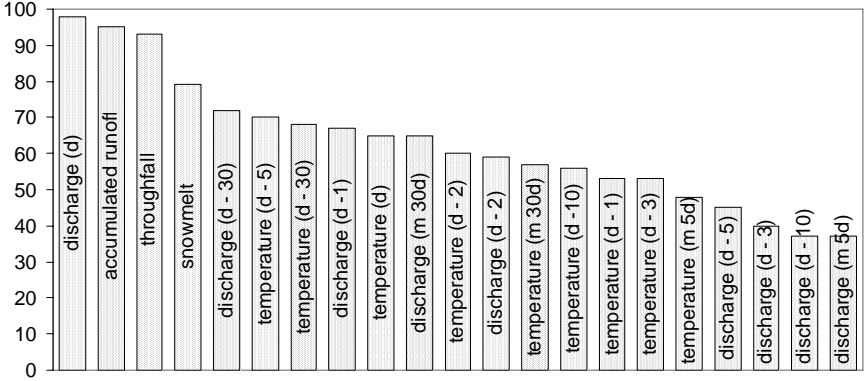
(Plötscher 1997)

Modell-Topologie

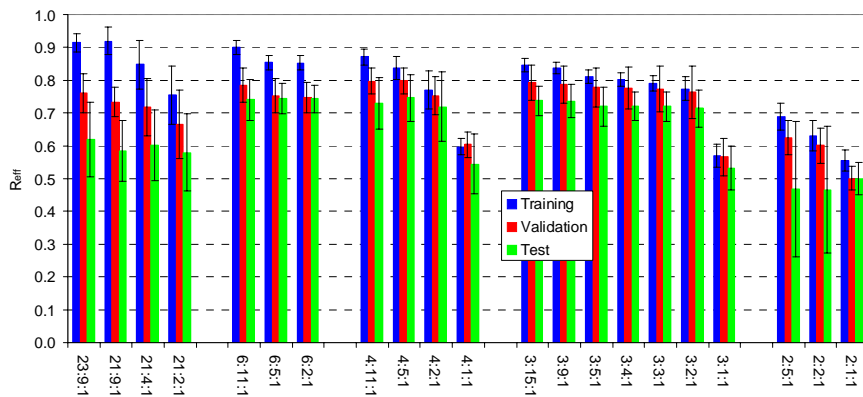


Ranking der Steuergrößen

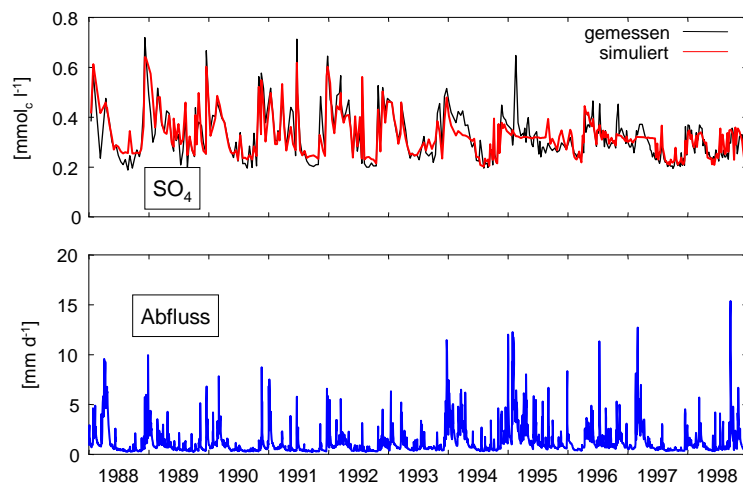
- 100x:
- Training des Netzes
 - Pruning (Löschen) der irrelevanten Regressoren



Optimierung des Netzes



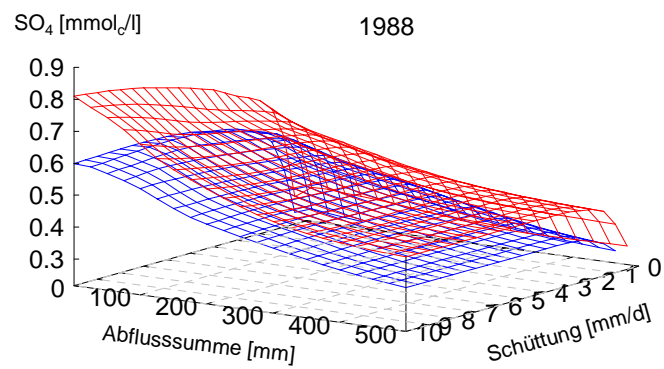
KNN: Sulfat im Lehstenbach



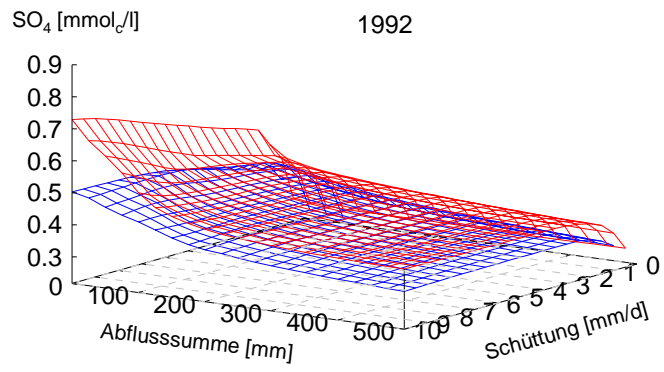
Steuergrößen

- 1. Schüttung:**
Tagesmittelwert des Probenahme-Tages
- 2. Abflusssumme:**
aufaddiert seit letztmaligem Überschreiten einer Schüttung von 0,5 mm/Tag (Basisabfluss)
- 3. Deposition:**
gleitender Mittelwert der Konzentrationen im Bestandesniederschlag

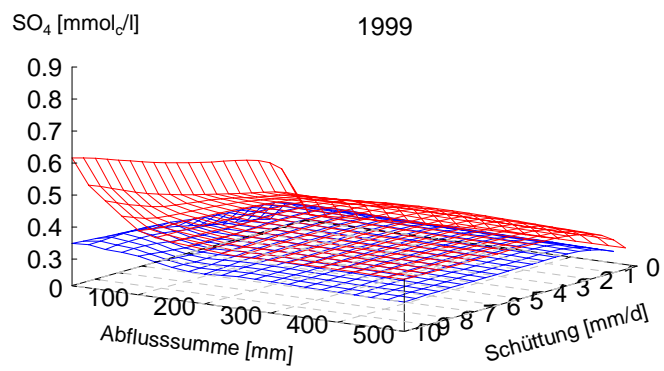
Regressionsflächen



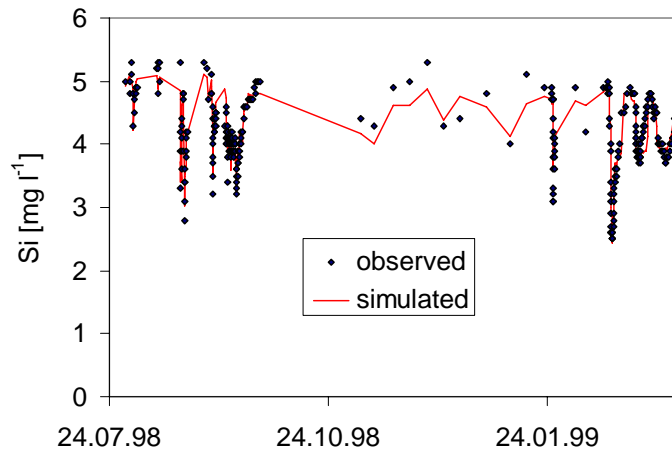
Regressionsflächen



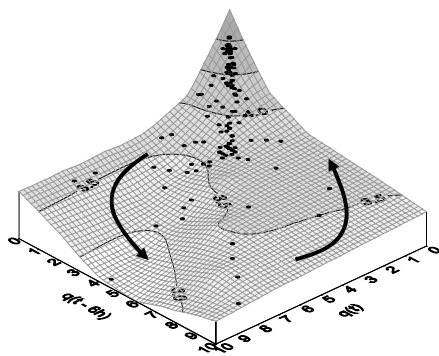
Regressionsflächen



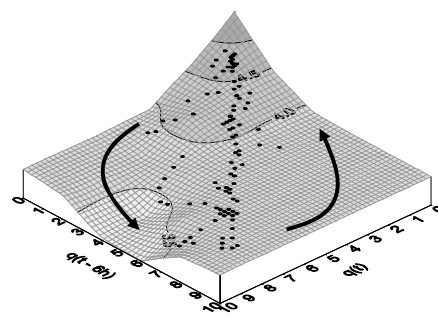
Silizium im Abfluss (Brugga)



Regressionsflächen



geringe Vorfeuchte



hohe Vorfeuchte

Literatur

Monographien:

- Bishop, C.M. (1995): Neural networks for pattern recognition. Oxford Press
- Hecht-Nielsen, R. (1990): Neurocomputing. Addison Wesley
- Ritter, H., T. Martinetz and K. Schulten (1990): Neuronale Netze. Eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke. - Addison Wesley (2. Auflage)
- Rojas, R. (1996): Theorie der neuronalen Netze. Eine systematische Einführung. Springer, Berlin, 4. Nachdruck
- Zell, A. (1994): Simulation neuronaler Netze. Addison Wesley

WWW:

- Sarle, W.S., ed. (1997), Neural Network FAQ, part 1 of 7: Introduction, periodic posting to the Usenet newsgroup comp.ai.neural-nets, URL: <ftp://ftp.sas.com/pub/neural/FAQ.html>

Aufgabe

1. Erstellen Sie in Statistica eine multivariate Regression mit drei ausgewählten Regressoren und einem Regressanden (z-normiert).
2. Bilden Sie diese Regression mit einem 3:2:1-Mehrschicht-Perzeptron in Excel „zu Fuß“ nach. Verwenden Sie dabei den kompletten Datensatz als Trainingsdatensatz.
 - Verwenden Sie die Excel-Funktion „Zufallszahl()“ zur wiederholten Variation der Gewichte, und lassen Sie sich den Korrelationskoeffizienten r^2 ausgeben. Speichern Sie die Werte der Gewichte, wenn sich gegenüber dem vorherigen Durchgang eine Verbesserung ergibt.
 - Verwenden Sie in der zweiten Phase das Backpropagation-Lernverfahren zur weiteren Optimierung des Netzes. Variieren Sie gegebenenfalls die Lernrate.
 - Variieren Sie die Anzahl der Knoten in der versteckten Schicht.