

Das Laufzeitproblem bei stochastischen Projektionsrechnungen

Dr. Stefan Nörtemann

Reisensburg, 21. September 2012



DAA-Workshop für junge Mathematiker 2012

Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
4. Automatisierung & Validierung der Verdichtung
5. COR.FJA CLUSTRED
6. Systemintegration
7. Zusammenfassung

Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
4. Automatisierung & Validierung der Verdichtung
5. COR.FJA CLUSTRED
6. Systemintegration
7. Zusammenfassung

Themenfelder Risikomanagement & Financial Reporting

- **Themenfelder / Einsatzgebiete für den Aktuar / die Aktuarin im Bereich Risikomanagement & Financial Reporting**
 - ◆ Asset-Liability Management-Analysen / ALM-Studien
 - ◆ Risikomanagement im Sinne §64a Versicherungsaufsichtsgesetz (VAG)
 - ◆ Risikokapitalberechnung gemäß Solvabilitätsvorschriften
 - ❖ Solvency II (Säule 1)
 - ◆ Rechnungslegung
 - ❖ HGB-Bilanz / GuV
 - ❖ IFRS
 - ◆ Embedded-Value-Berechnungen (TEV, MCEV)
 - ◆ Produktentwicklung kapitalmarktnaher Produkte
 - ❖ Dynamische Hybrid-Produkte
 - ❖ Variable Annuities

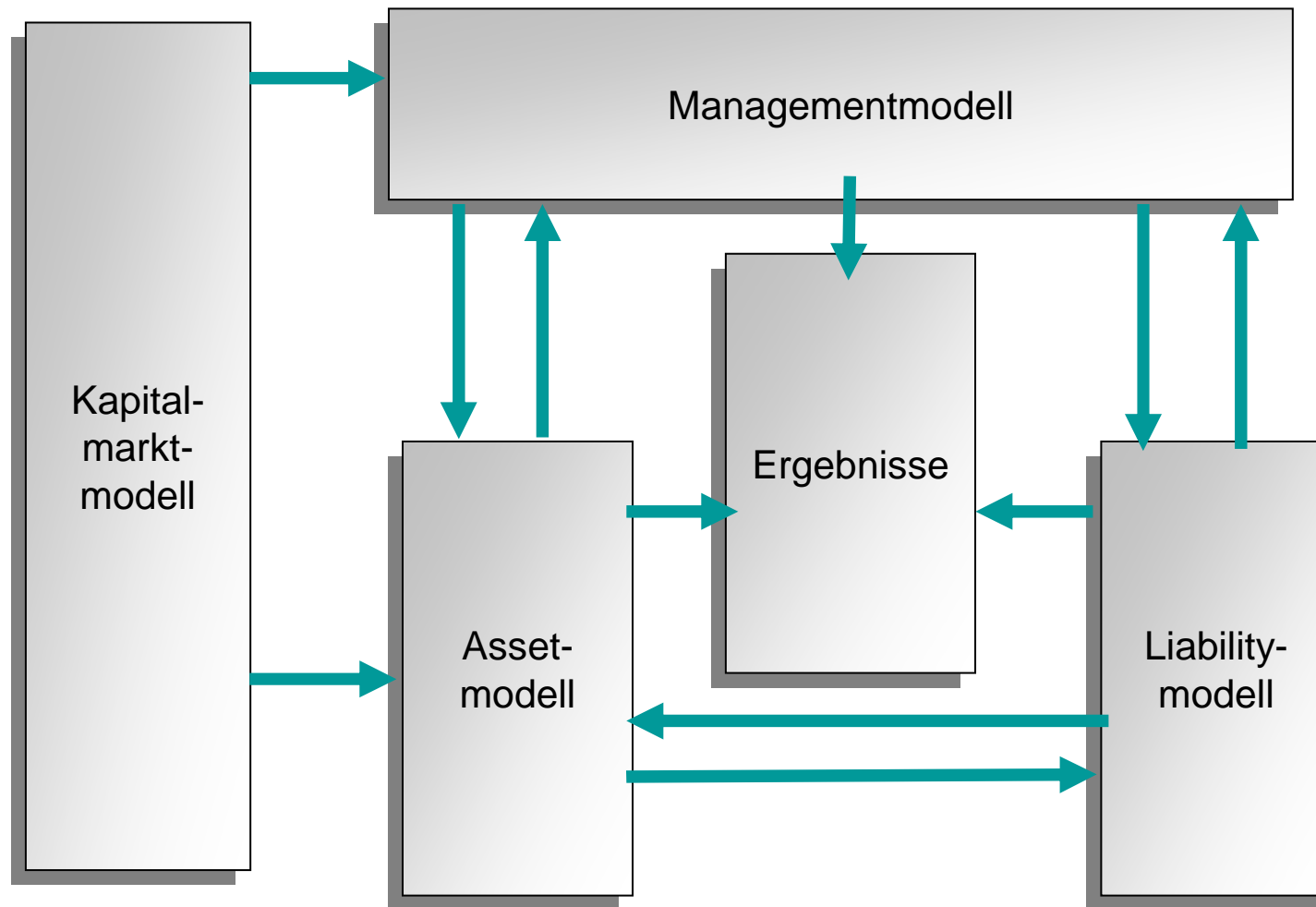
Projektionsrechnungen (1)

- (Fast) allen Themen gemeinsam ist die Notwendigkeit Projektionsrechnungen durchzuführen
- Projektionsrechnungen = Hochrechnung des gesamten Versicherungsunternehmens über einen definierten Zeitraum (Projektionsdauer) ...
- ... auf Basis gewisser (dynamischer) Regeln
- Vorarbeiten:
 - ◆ Modellierung Vermögenswerte (Asset-Model)
 - ◆ Modellierung Verpflichtungen (Liability-Modell)
 - ◆ Modellierung Kapitalmarkt
 - ◆ Modellierung Entscheidungsregeln (Management-Model)
- ...

Projektionsrechnungen (2)

- ...
- Hochrechnung der erwarteten Erträge, Einnahmen, Leistungen, Steuern, Ausschüttungen, etc. (Cashflows)
- Berechnung relevanter Kennzahlen, Bilanzen, etc. auf Basis der projizierten Cashflows
- Analyse der Auswirkungen
- Erarbeitung von Handlungsalternativen
- Überprüfung auf Basis weiterer Projektionsrechnungen

Asset-Liability-Management



ALM-Techniken

- **szenarienbasierte Analysen (What-if-Analysen)**
 - ◆ Berechnung ausgewählter primärer Kennzahlen für ein oder mehrere Szenarien, d.h. Sätze von Input-Daten (Beispiel: Stresstests)
 - **Sensitivitätsanalysen**
 - ◆ Untersuchung der (prozentualen) Änderung der Kennzahlen bei Änderung bestimmter Input-Werte
 - **Break-even-Analysen**
 - ◆ Berechnung der Input-Werte die zur Über- oder Unterschreitung gewisser Schwellwerte einer Kennzahl führen
 - **Monte-Carlo-Simulation**
 - ◆ Approximierung der Wahrscheinlichkeitsverteilung einer Kennzahl durch ihre empirische Verteilung einer Anzahl von Szenarien / Simulationen gemäß der Verteilung der Input-Daten unter Berücksichtigung wechselseitiger Zusammenhänge
- Verteilungen verschiedener Kennzahlen, Momente (Erwartungswert, Varianz, etc.), Risikomaße

Projektionsrechnungen – technische Herausforderung

- Beispiel aus der Praxis:
 - ◆ Lebensversicherungsbestand 3,5 Mio. Verträge
 - ◆ Projektionsrechnung über 100 Jahre (für MCEV)
 - ◆ Monatliche Projektionsschritte
 - ◆ Dynamische Managementregel für
 - ❖ Überschussverwendung
 - ❖ Kapitalanlage
 - ❖ Ausschüttungspolitik
 - ◆ Stochastische Simulation des Kapitalmarktes (20.000 Kapitalmarktpfade)
- Verwendung einer am Markt etablierten Projektionssoftware
- ➔ Problem: Laufzeiten von mehreren Tagen!!

Lösungsansätze

- Für das Laufzeitproblem liegen folgende Lösungsansätze nahe:
 - ◆ Bessere Hardware
 - ◆ Parallelverarbeitung
 - ◆ Geduld
 - ◆ Fachlich getriebene Lösungen
 - ❖ Redundanzvermeidung
 - ❖ Bestandsverdichtung

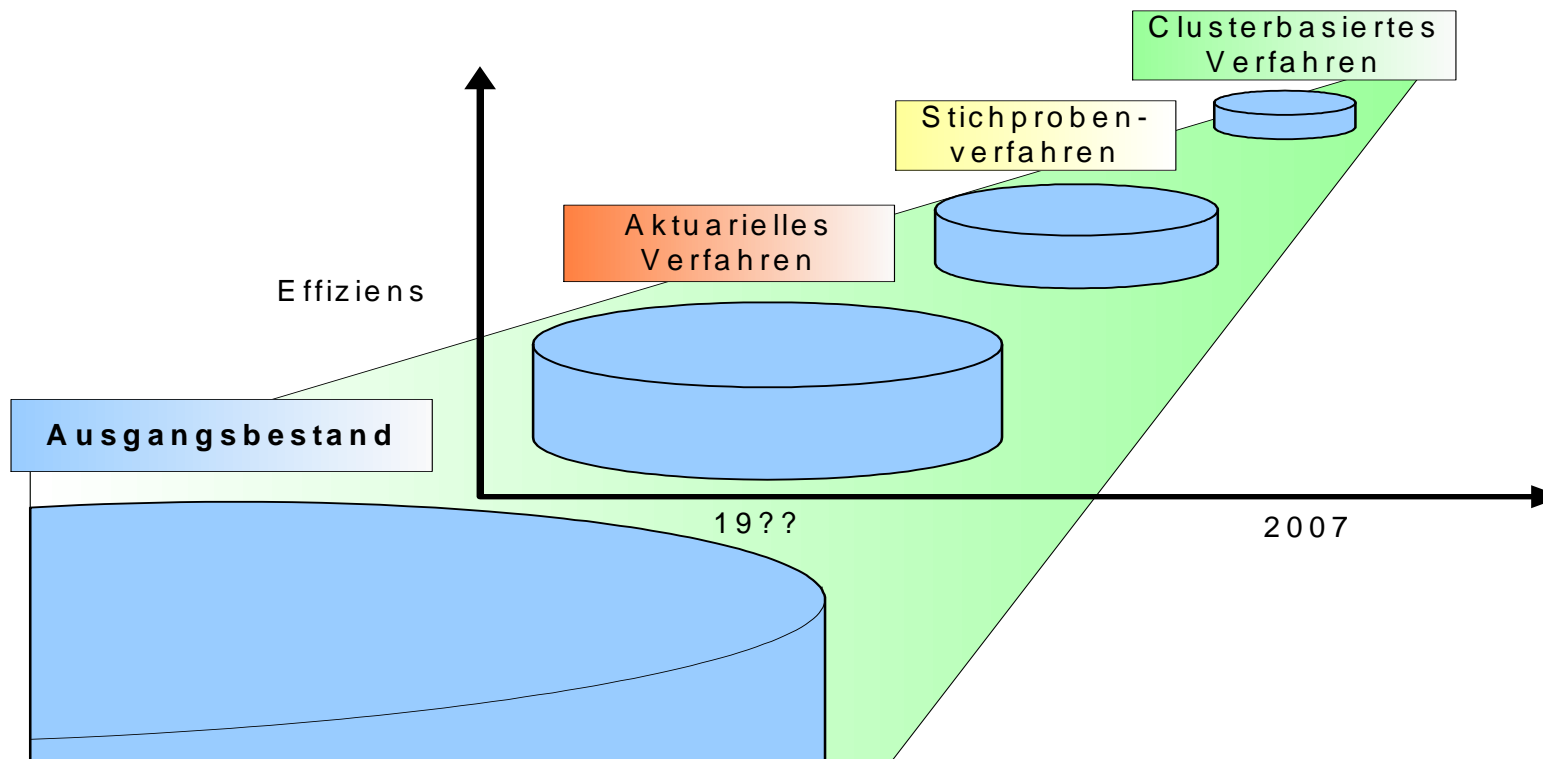
- ➔ „Trick“ der Verdichtung ist die Erzeugung eines Teilbestandes (des Versichertenbestandes), der dieselben (oder ähnliche) Eigenschaften aufweist, wie der Ausgangsbestand.

Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
4. Automatisierung & Validierung der Verdichtung
5. COR.FJA CLUSTRED
6. Systemintegration
7. Zusammenfassung

Verdichtungsverfahren: Entwicklung

- Ziel der Verdichtung: Einen möglichst kleinen Teilbestand zu erzeugen, der die gleichen Eigenschaften besitzt wie der Originalbestand



Verdichtungsverfahren: Aktuarielle Verfahren

Aktuarielle Verfahren: Verfahren bei denen auf der Basis von aktuariellem Expertenwissen, gemäß vorgegebbarer Kriterien verschiedene Verträge zusammengefasst werden

Probleme:

- Unzureichende Verdichtungsgrad – infolgedessen sind die Laufzeitverbesserungen häufig nicht ausreichend
- Kaum quantifizierbare Güte – und damit eine hohe Unsicherheit hinsichtlich der Qualität der Analyseergebnis
- Ineffizienz – mit der Folge eines hohen Aufwandes bei vergleichsweise geringem Nutzen

Verdichtungsverfahren: Stichprobenverfahren (1)

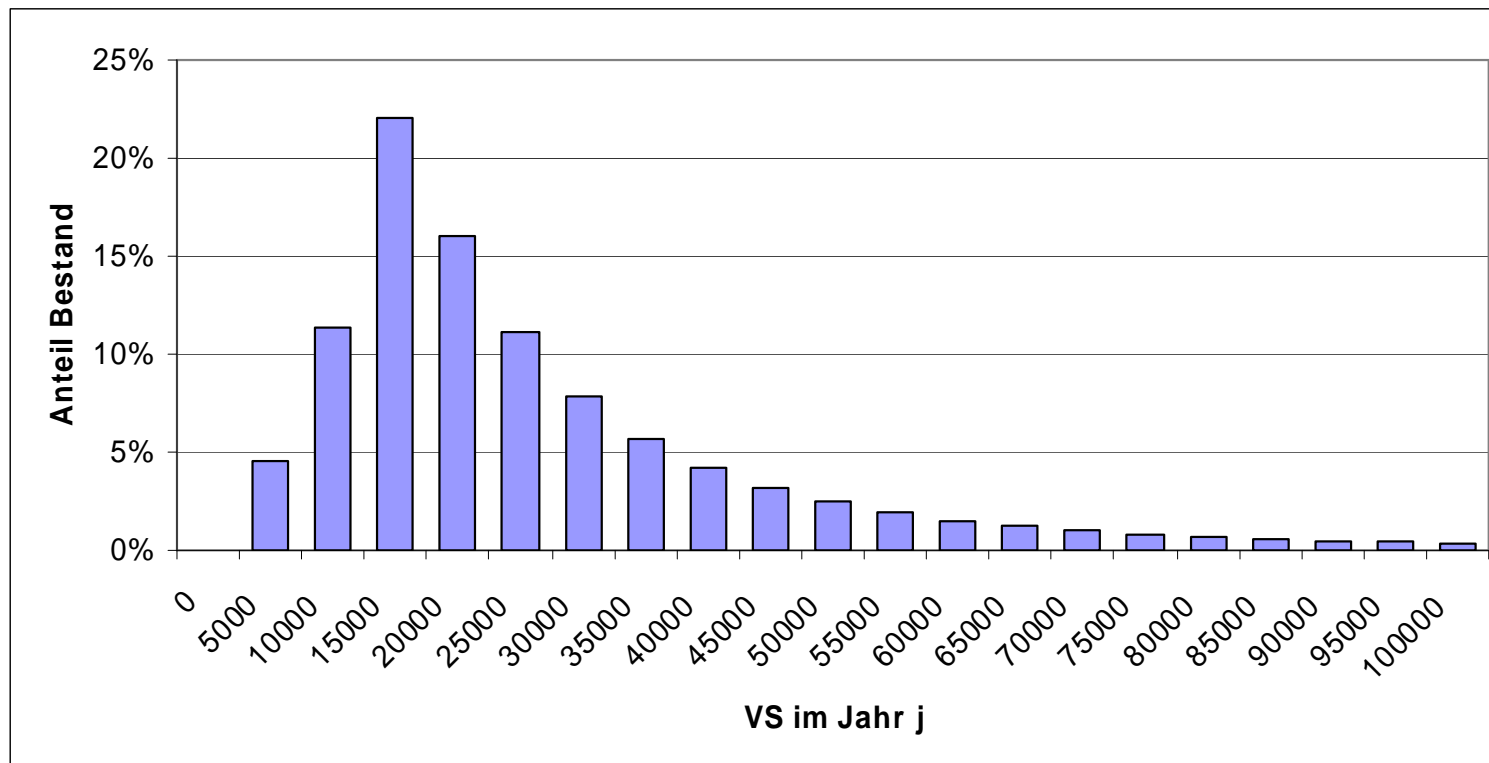
Stichprobenverfahren: qualifizierte Monte-Carlo-Methode (qMCM):
Kombination einer Stichprobenziehung mit einer nachgelagerten
Optimierung

Vorteile:

- Höhere Verdichtungsgrade (als bei aktuariellen Verfahren) –
infolgedessen erhebliche Laufzeitverbesserungen
- Effizienter Einsatz, da nur geringe Vorarbeiten und kein
aktuarielles Vorwissen notwendig sind
- Optimierung des Verdichtungsgrades unter vorgegebener Güte
durch Variationsrechnungen mit unterschiedlichen
Verdichtungsgraden (Stichprobengrößen)
- Einfache Weiterverarbeitung des verdichteten Bestandes, da
Model Points reale Verträge sind, die ohne Einschränkung
aktuariell rechen- und bewertbar sind

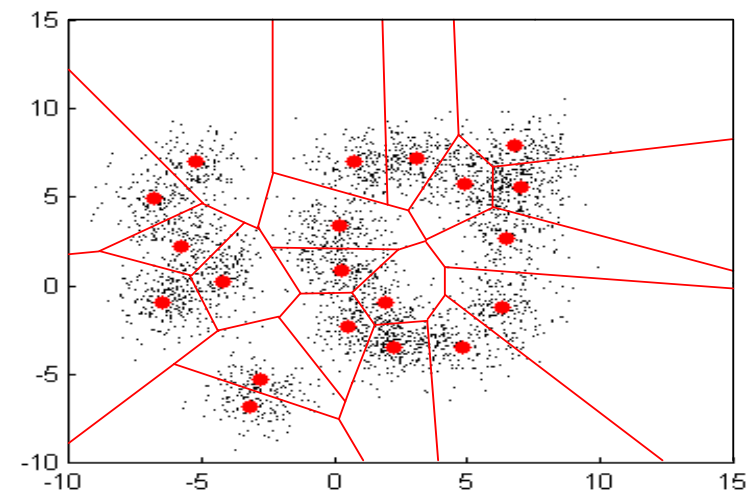
Verdichtungsverfahren: Stichprobenverfahren (2)

- Zu wenige Stichprobenpunkte bei seltenen Vertragskonstellationen (z.B. hoher VS)
- Unnötige viele Stichprobenpunkte bei häufigen Vertragskonstellationen (z.B. mittlere VS)



Verdichtungsverfahren: Cluster-Verdichtung

- **Problem:** Laufzeit (insbesondere bei stochastischen Projektionen)
- **Lösung:** Bestandsverdichtung
- **Ziel der Verdichtung:** Erzeugung eines möglichst kleinen Teilbestandes, der die gleichen Eigenschaften besitzt wie der Originalbestand
- **Eine Lösung:** Clusterbasiertes Verfahren
- **Idee:** Gruppierung ähnlicher Daten in Klassen (Cluster, Klumpen)
- **Ziel:**
 - ◆ möglichst hohe Homogenität innerhalb eines Clusters
 - ◆ möglichst hoher Heterogenität zwischen den Clustern



Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
4. Automatisierung & Validierung der Verdichtung
5. COR.FJA CLUSTRED
6. Systemintegration
7. Zusammenfassung

Agenda

1. Die Herausforderung

2. Verfahren zur Bestandsverdichtung

3. Clusterverdichtung

a. Clusteralgorithmus

b. Merkmalsvektoren

c. Quantitative Ergebnisse

d. Performanz

Agenda

1. Die Herausforderung

2. Verfahren zur Bestandsverdichtung

3. Clusterverdichtung

a. Clusteralgorithmus

b. Merkmalsvektoren

c. Quantitative Ergebnisse

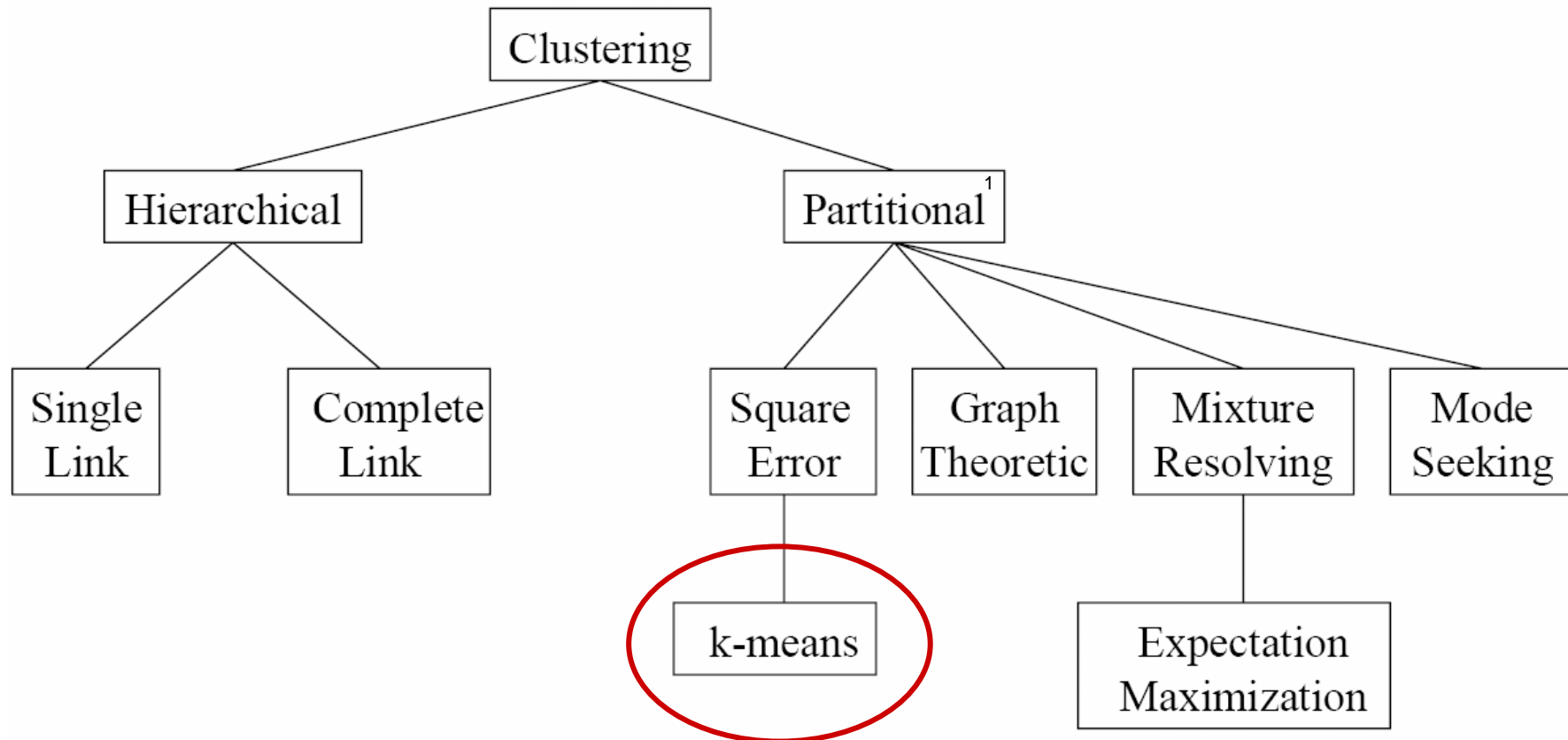
d. Performanz

Clusteralgorithmus: Überblick

Was ist Cluster-Analyse (kurz Clustering)?

- Gruppierung ähnlicher Daten in Klassen (Cluster, Klumpen)
- Die Datenmenge wird partitioniert, mit dem Ziel
 - möglichst hoher Homogenität innerhalb eines Clusters
 - möglichst hoher Heterogenität zwischen den Clustern
- Objekte sind sog. Merkmalsvektoren eines metrischen Vektorraums (i.a.)
- Ähnlichkeit i.d.R. Euklidischer Abstand, angewandt auf die standardisierten Merkmalswerte (also auf $(x-\mu_x) / \sigma_x$)

Taxonomie der Clusterverfahren



Nach Jain, Murty, Flynn, „Data Clustering: A Review“, ACM Computing Reviews, 1999

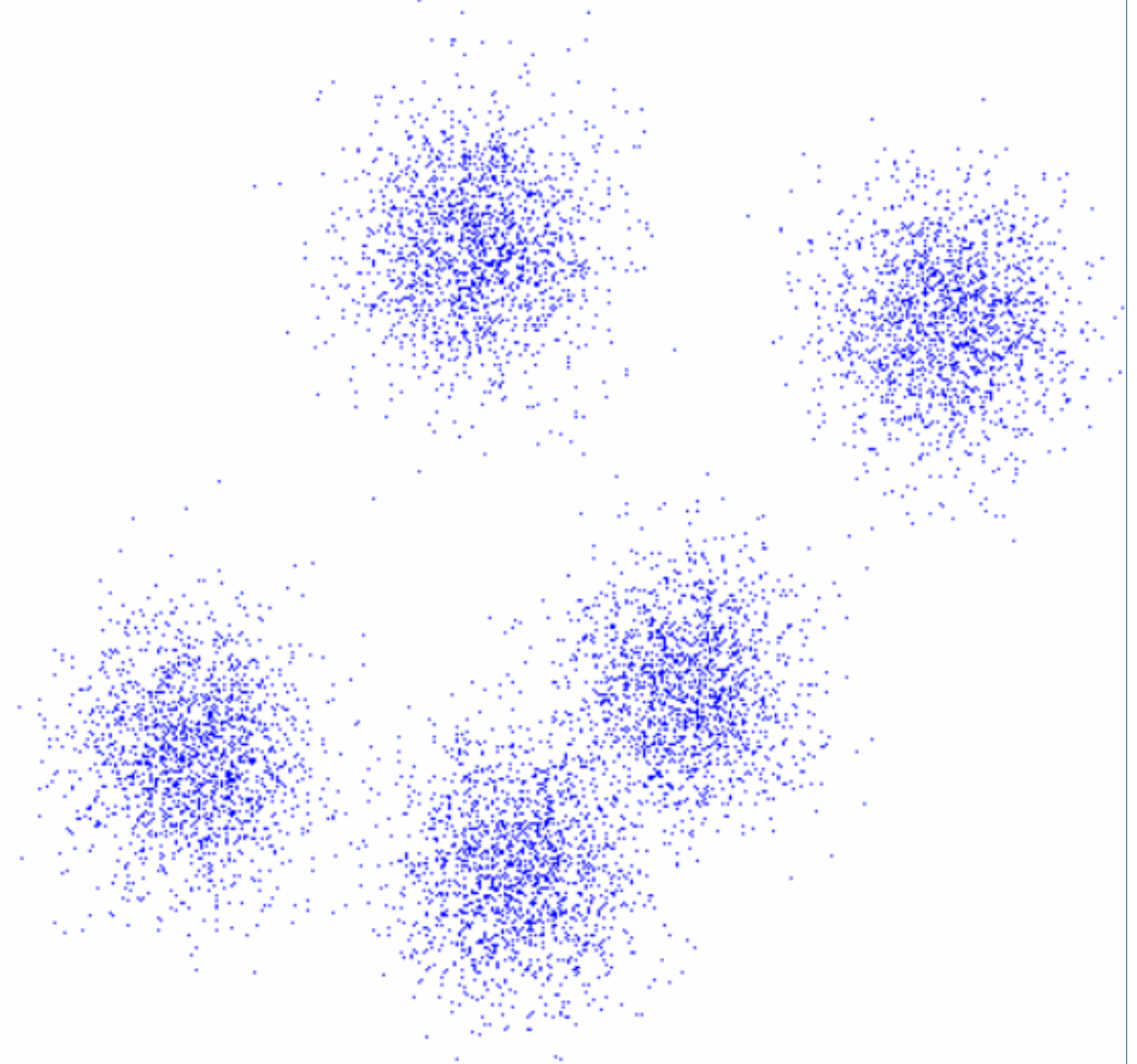
¹Auch „Relocation methods“ genannt

Bestandsverdichtung mit Clustering – Algorithmus

- Algorithmische Formulierung (vereinfacht)
 - 1) Wähle k anfängliche Cluster-Centroide, $c_1, c_2, c_3, \dots, c_k$.
 - 2) Für jeden Datenpunkt x , finde den nächsten (nearest neighbour) Cluster-Centroiden
 - 3) Für jedes Cluster (gegeben durch vorherigen Schritt), aktualisiere den Cluster-Centroiden (bilde den Mittelwert über alle Cluster-Elemente, d.h. dem Cluster zugeordnete Datenpunkte)
 - 4) Gehe zu (2) bis sich die Cluster nicht mehr ändern (z.B. SSE konstant) oder die Änderung nur noch minimal ist (vorgegebenes Epsilon)

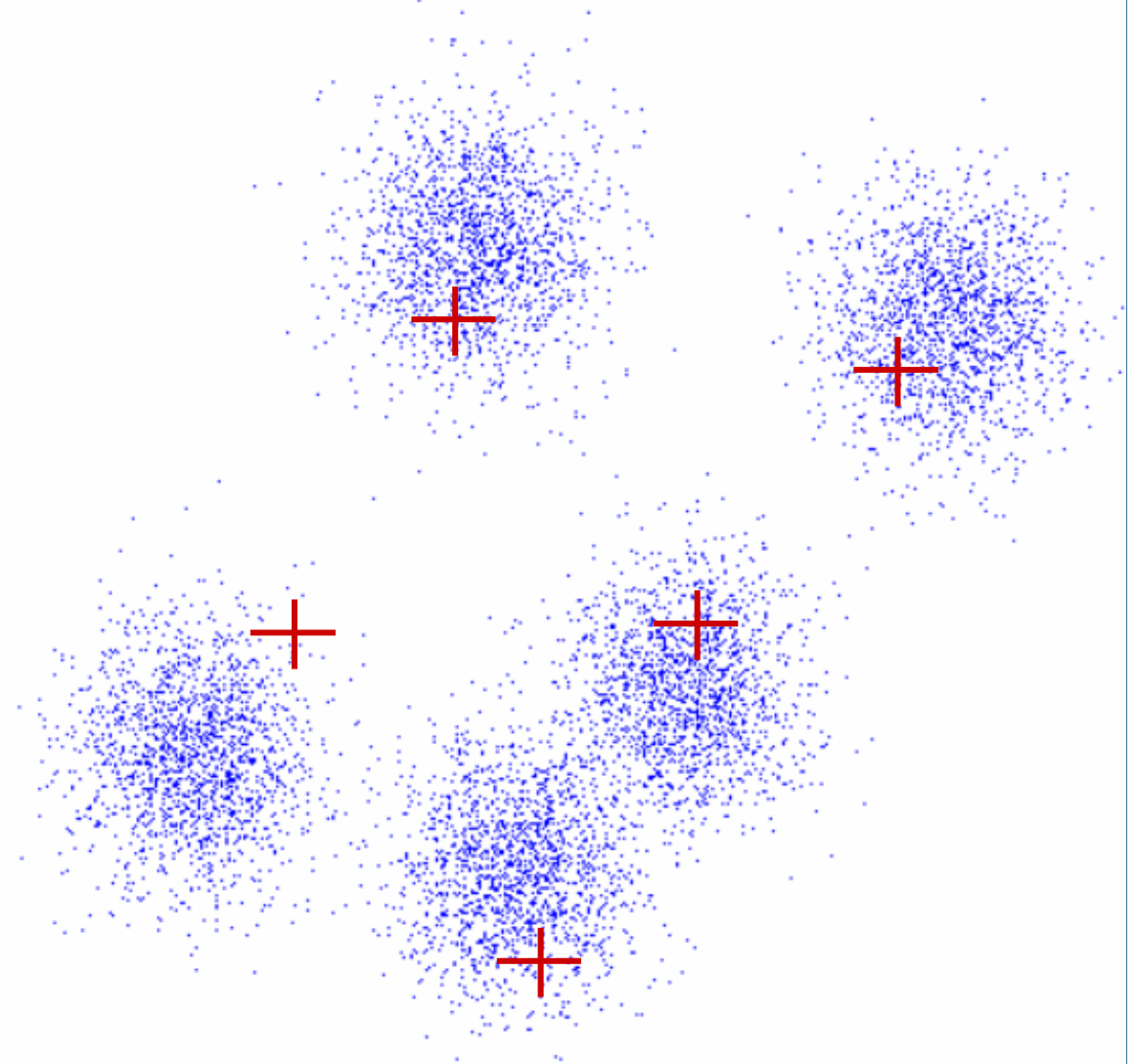
Beispiel

- Wie viele Cluster k ? (z.B. 5)



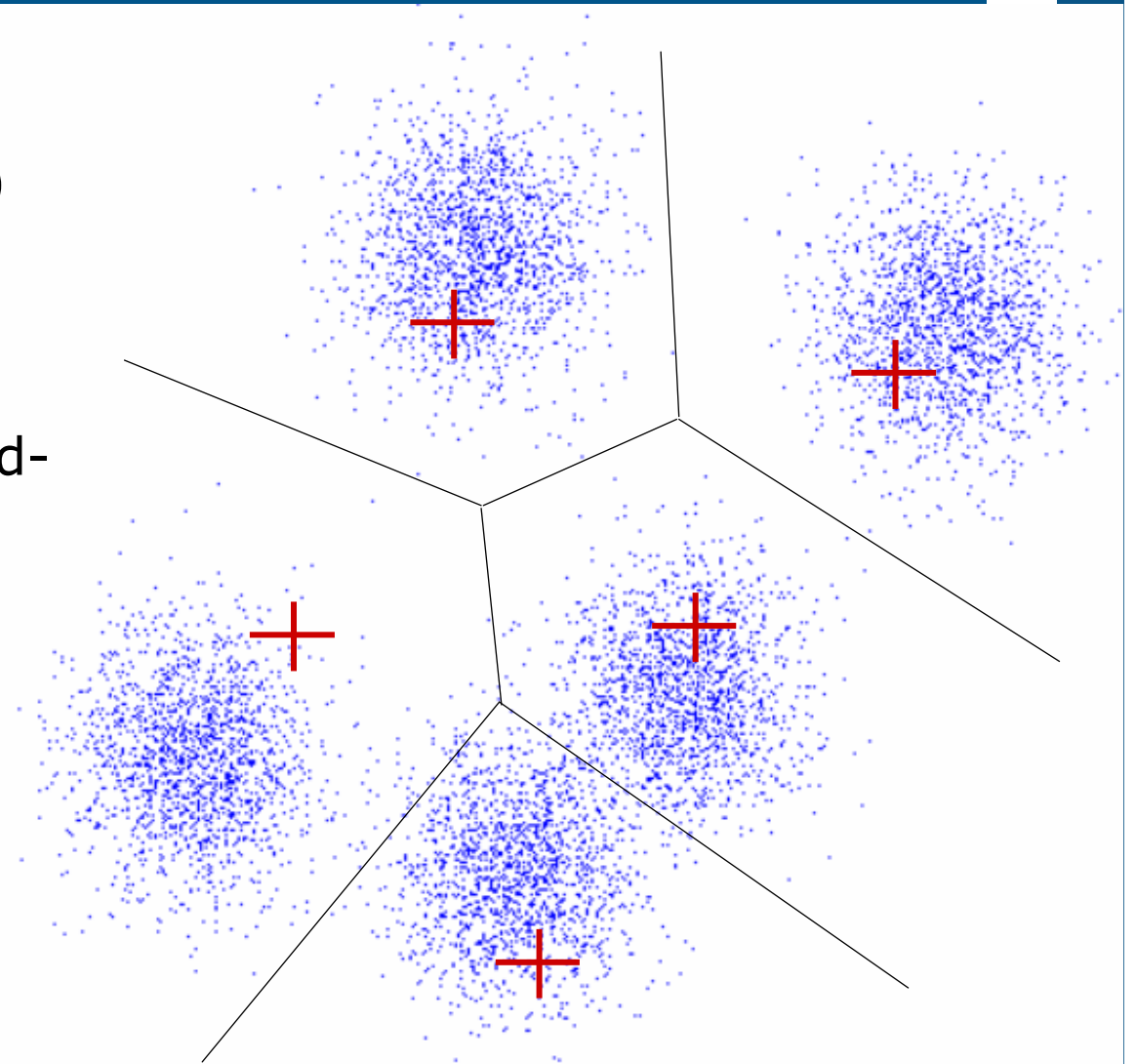
Beispiel

- Wie viele Cluster k ? (z.B. 5)
- Wähle zufällig k Centrioid-Kandidaten



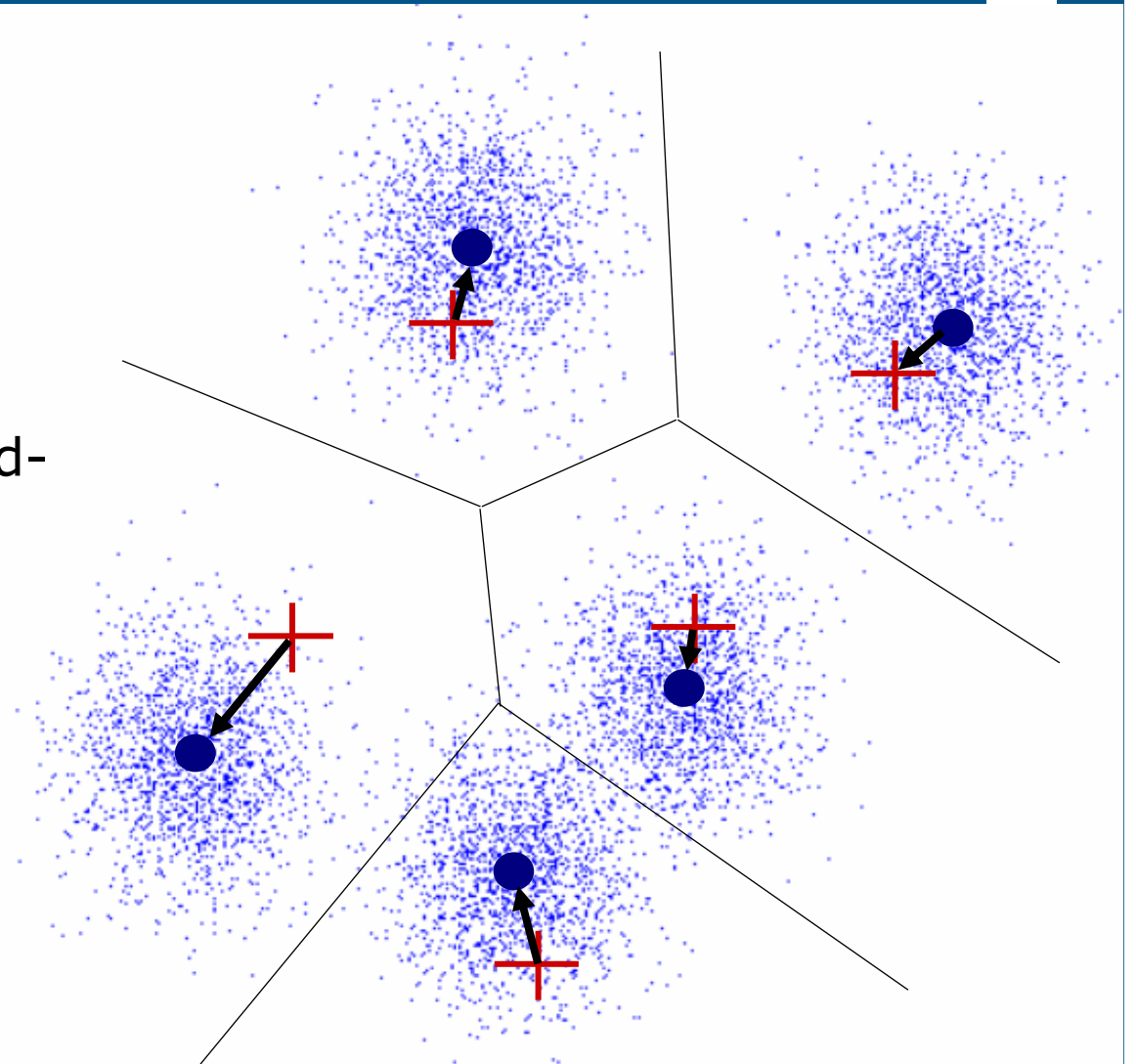
Beispiel

- Wie viele Cluster k ? (z.B. 5)
- Wähle zufällig k Centrioid-Kandidaten
- Jeder Vektor (Datenpunkt) findet seinen nächsten Centrioid-Kandidaten



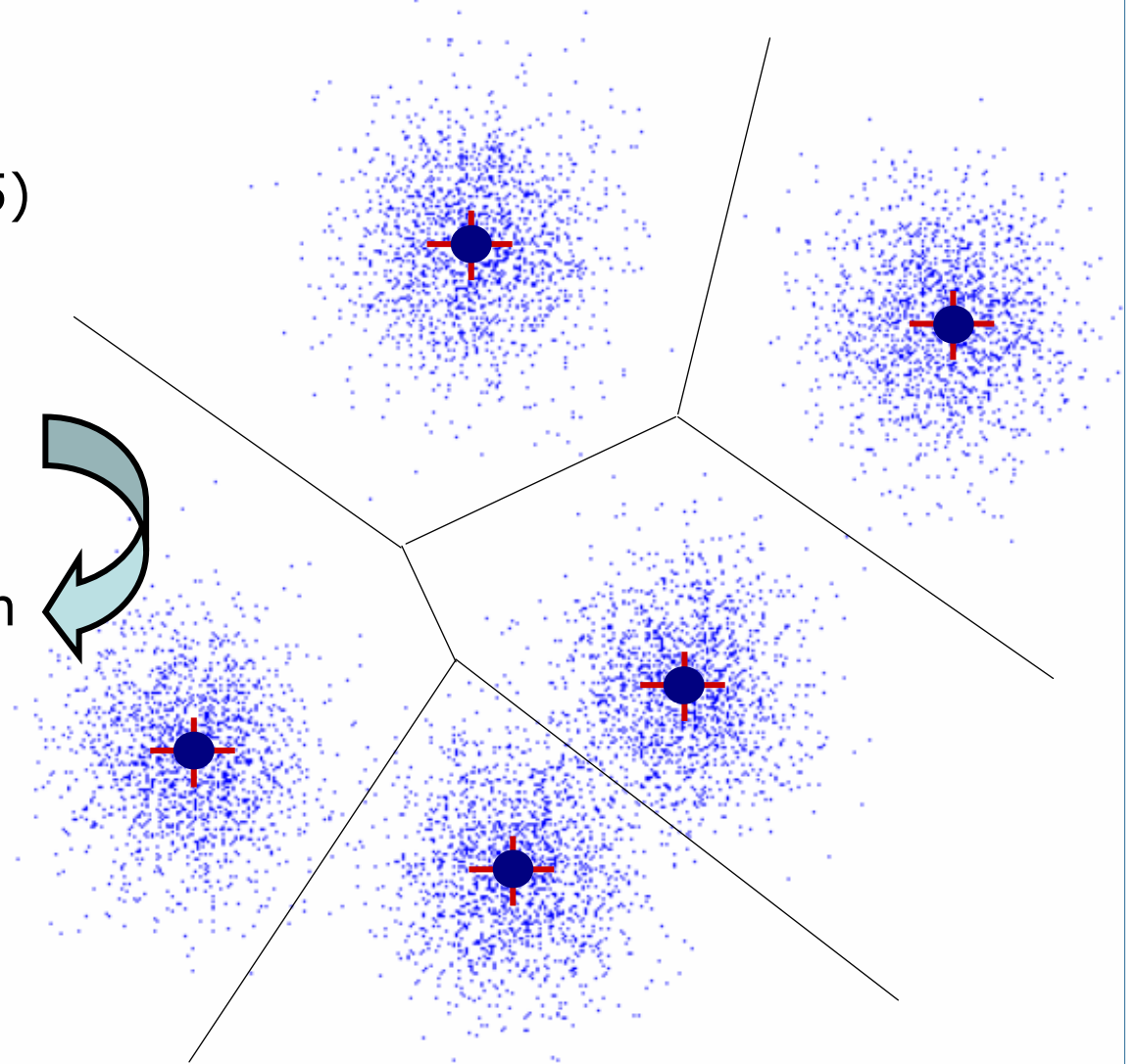
Beispiel

- Wie viele Cluster k ? (z.B. 5)
- Wähle zufällig k Centrioid-Kandidaten
- Jeder Vektor (Datenpunkt) findet seinen nächsten Centrioid-Kandidaten
- Berechne den tatsächlichen Centroiden eines Clusters



Beispiel

- Wie viele Cluster k ? (z.B. 5)
- Wähle zufällig k Centroid-Kandidaten
- Jeder Vektor (Datenpunkt) findet seinen nächsten Centroid-Kandidaten
- Berechne den tatsächlichen Centroiden eines Clusters

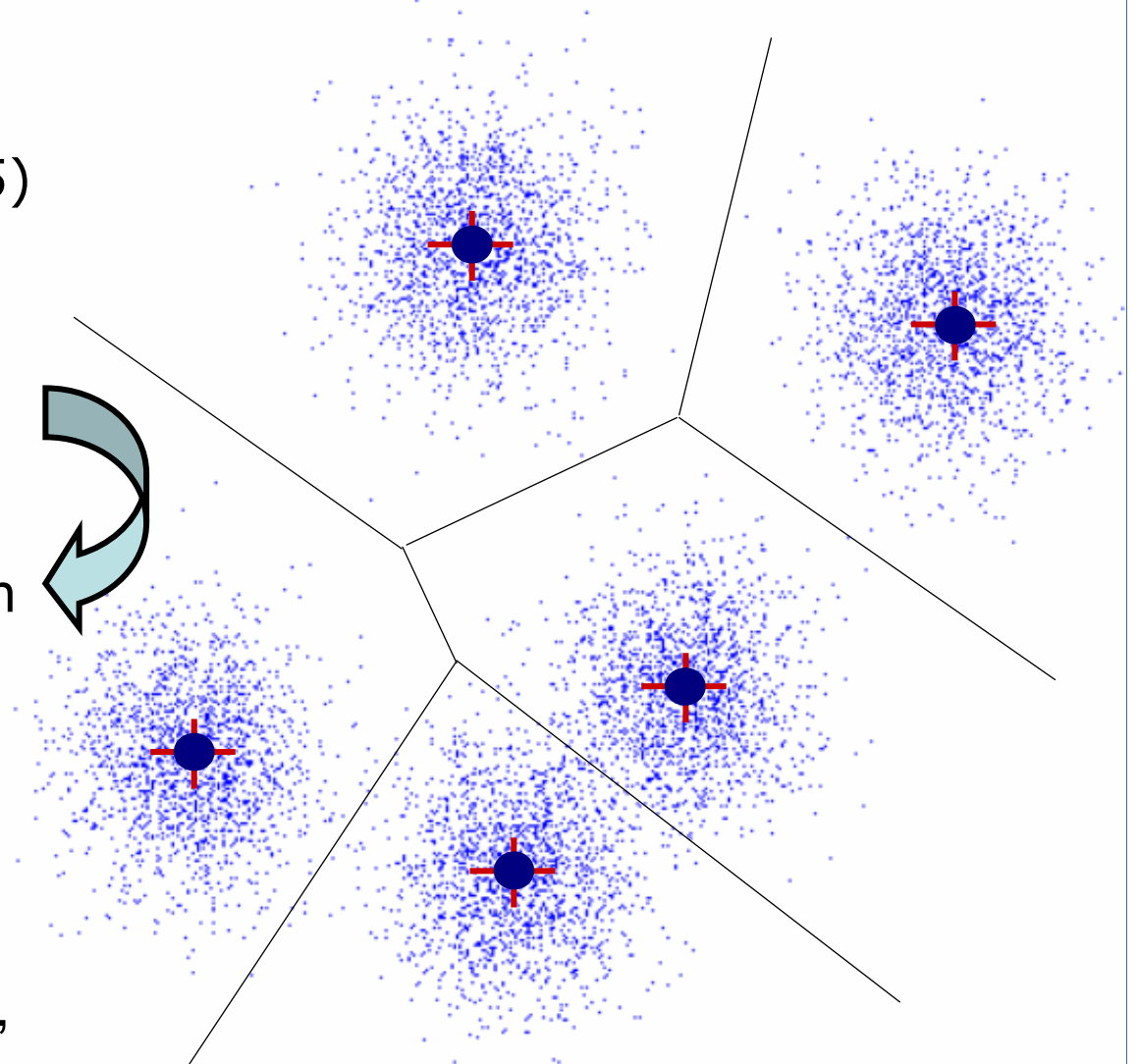


Beispiel

- Wie viele Cluster k ? (z.B. 5)
- Wähle zufällig k Centrioid-Kandidaten
- Jeder Vektor (Datenpunkt) findet seinen nächsten Centrioid-Kandidaten
- Berechne den tatsächlichen Centroiden eines Clusters

• Laufzeitkomplexität
 $O(N \cdot k \cdot d \cdot i)$

d Dimension Merkmalsvektoren,
 i Anzahl Iterationen



Clusteralgorithmus: K-Means

K-Means Cluster-Algorithmus

- „Is by far the most popular clustering algorithm used in scientific and industrial applications“¹
- Lineare Laufzeitkomplexität (für eine Iteration)
- Terminiert (SSE echt monoton fallend, sowie maximal k^N mögliche Clusterbildungen)
- In der Praxis schnelle Konvergenz
 - Eigene Erfahrung 50-150 Iterationen ($N=250.000$, $k=1.000$, $d=5$)
 - Theoretische Resultate: vgl. Literatur²

¹ Pavel Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA, 2002

² David Arthur and Sergei Vassilvitskii. How slow is the k-means method? In SCG '06: Proceedings of the twenty-second annual symposium on computational geometry. ACM Press, 2006

Agenda

1. Die Herausforderung

2. Verfahren zur Bestandsverdichtung

3. Clusterverdichtung

a. Clusteralgorithmus

b. Merkmalsvektoren

c. Quantitative Ergebnisse

d. Performanz

Merkmalsvektoren

Definition Merkmalsvektoren (MVs)

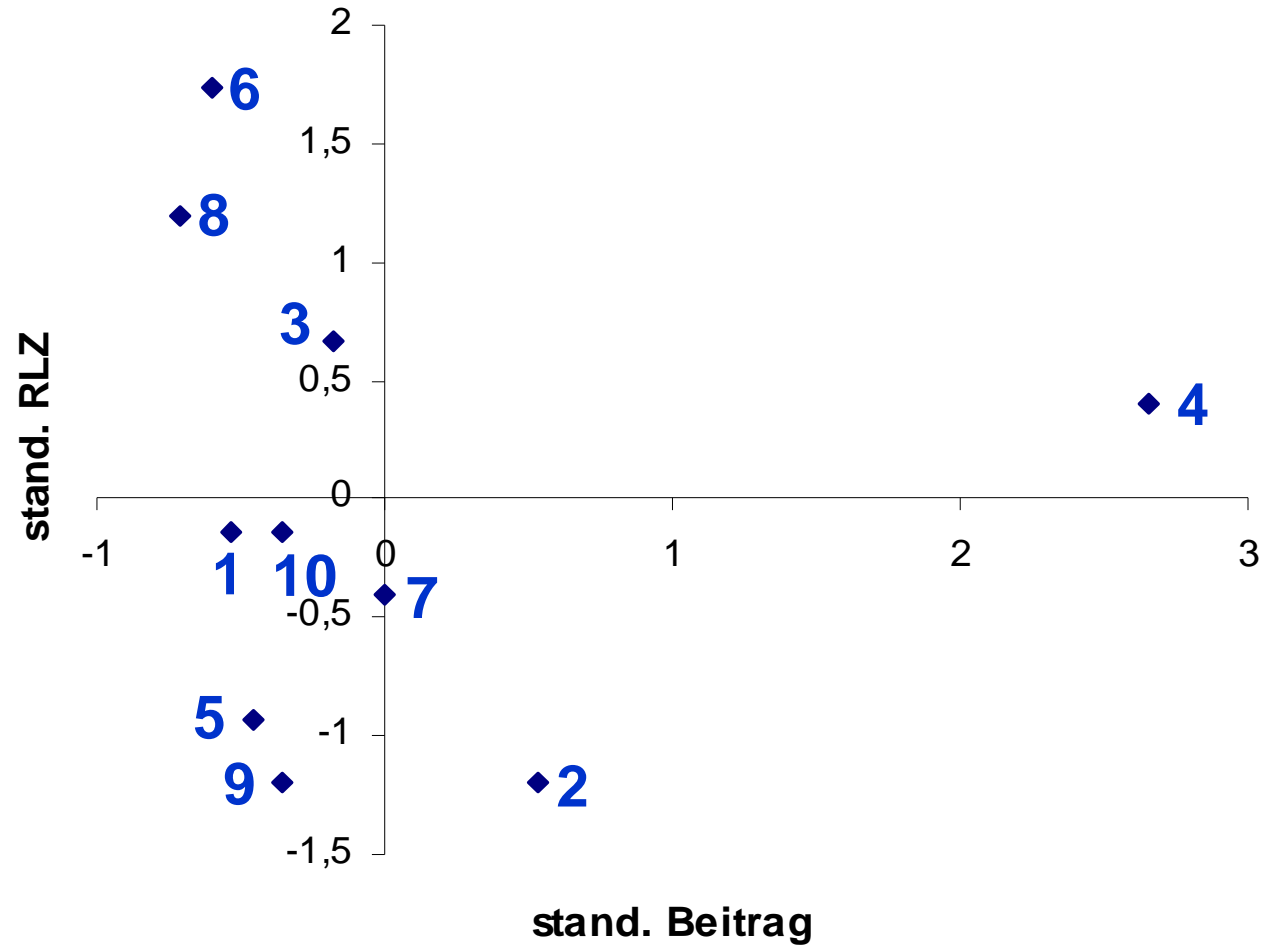
- Merkmalsvektoren *kodieren* einen Vertrag im Eukl. Raum
- Cashflow-Verläufe soll(t)en abgebildet werden
Leistungen, Beiträge, indirekt durch Reserveentwicklung, ...

Bestandsverdichtung mit Clustering

VNr.	Beitrag	Restlaufzeit	stand. Beitrag	stand. RLZ
1	10000	5	-0,532	-0,133
2	40000	1	0,532	-1,200
3	20000	8	-0,177	0,667
4	100000	7	2,659	0,400
5	12000	2	-0,461	-0,934
6	8000	12	-0,603	1,734
7	25000	4	0,000	-0,400
8	5000	10	-0,709	1,200
9	15000	1	-0,355	-1,200
10	15000	5	-0,355	-0,133
Mittelwert	25000	5,5		
Standardabw.	28201,65	3,75		

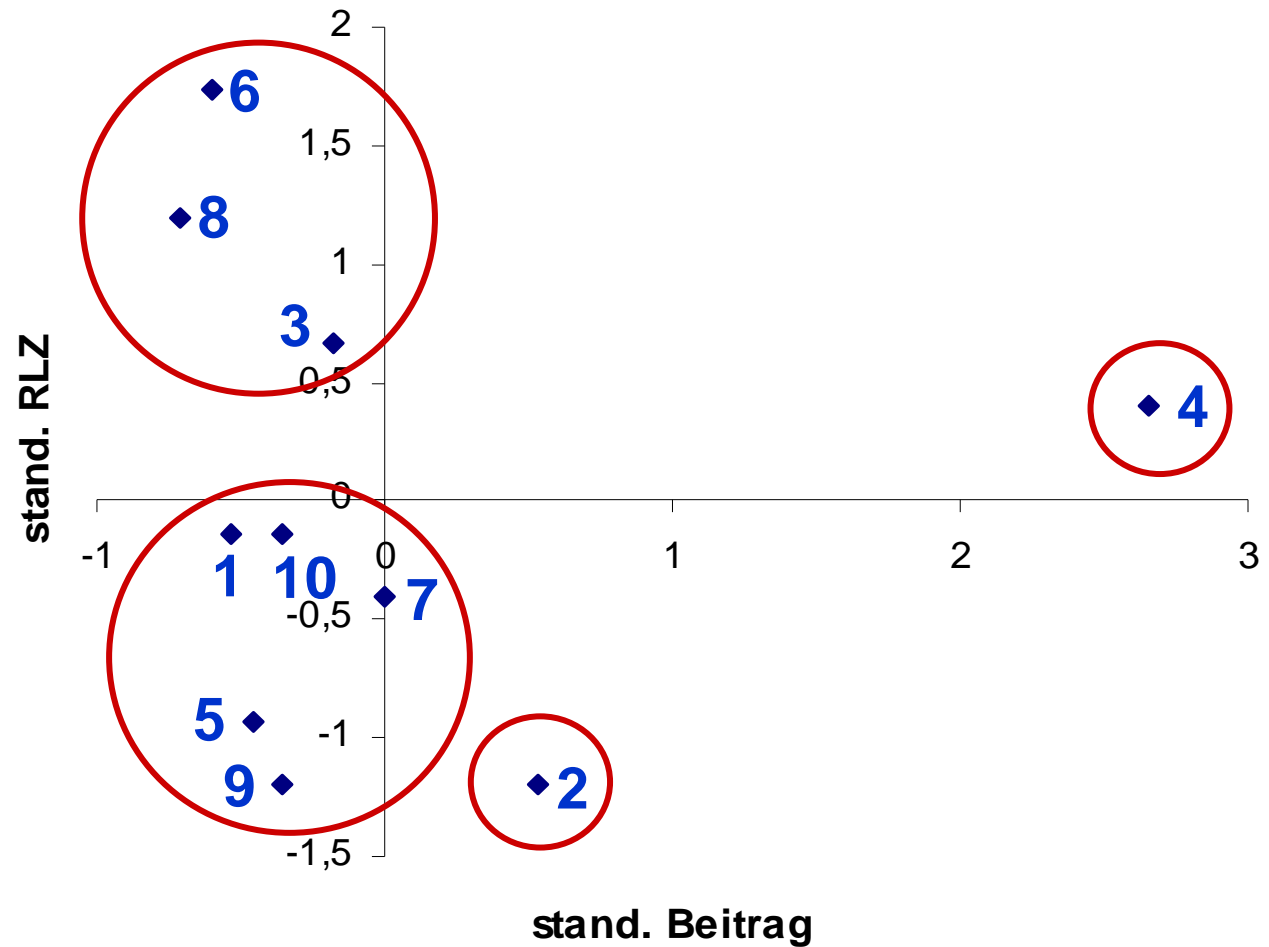
Beispiel

VNr.	Beitrag	RLZ
1	10000	5
2	40000	1
3	20000	8
4	100000	7
5	12000	2
6	8000	12
7	25000	4
8	5000	10
9	15000	1
10	15000	5



Beispiel

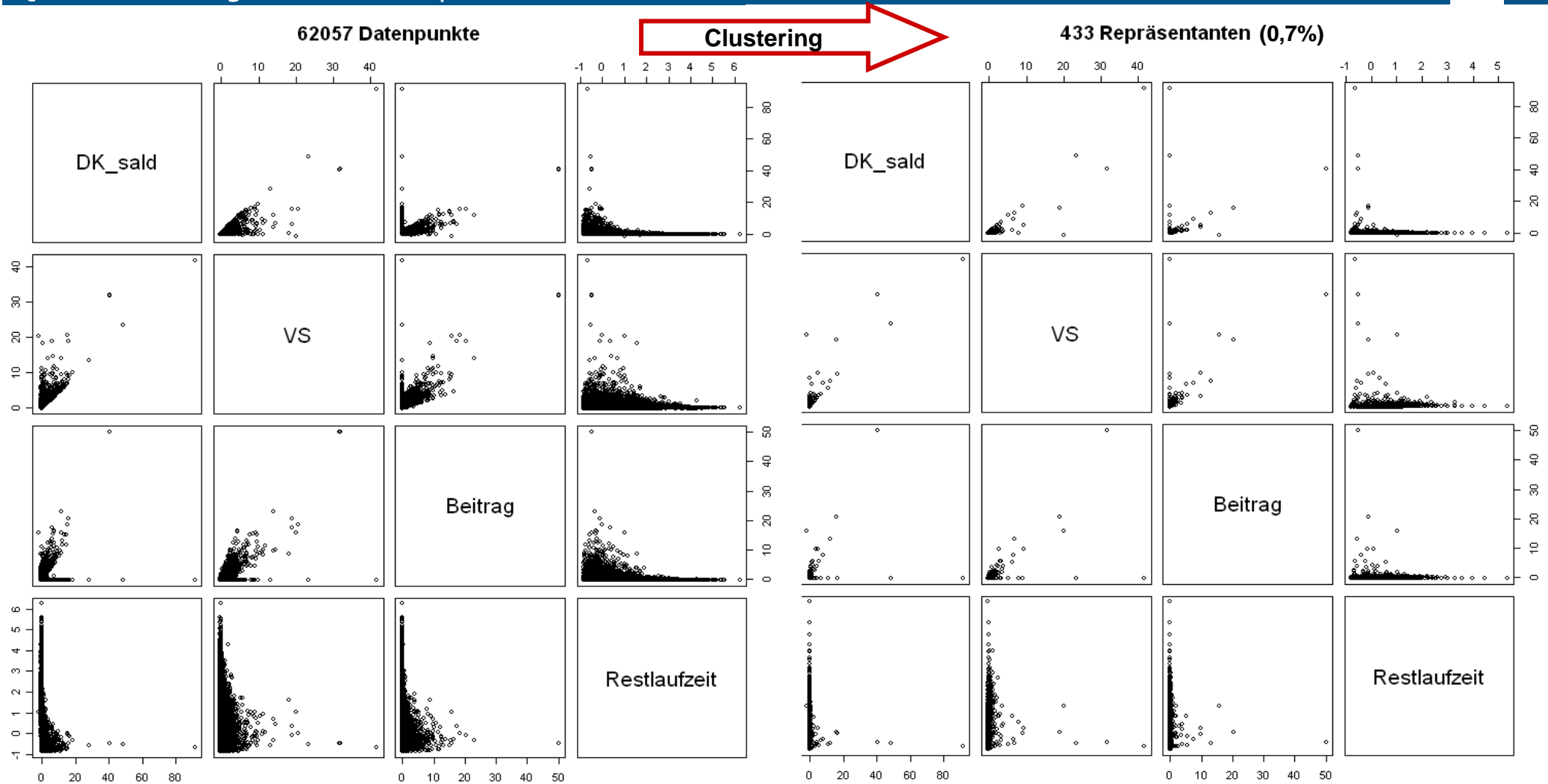
VNr.	Beitrag	RLZ
1	10000	5
2	40000	1
3	20000	8
4	100000	7
5	12000	2
6	8000	12
7	25000	4
8	5000	10
9	15000	1
10	15000	5



Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
 - a. Clusteralgorithmus
 - b. Merkmalsvektoren
 - c. Quantitative Ergebnisse
 - d. Performanz

Quantitative Ergebnisse - Beispiel



→ Clusteralgorithmus erhält Strukturen und berücksichtigt Ausreisser

Agenda

1. Die Herausforderung

2. Verfahren zur Bestandsverdichtung

3. Clusterverdichtung

a. Clusteralgorithmus

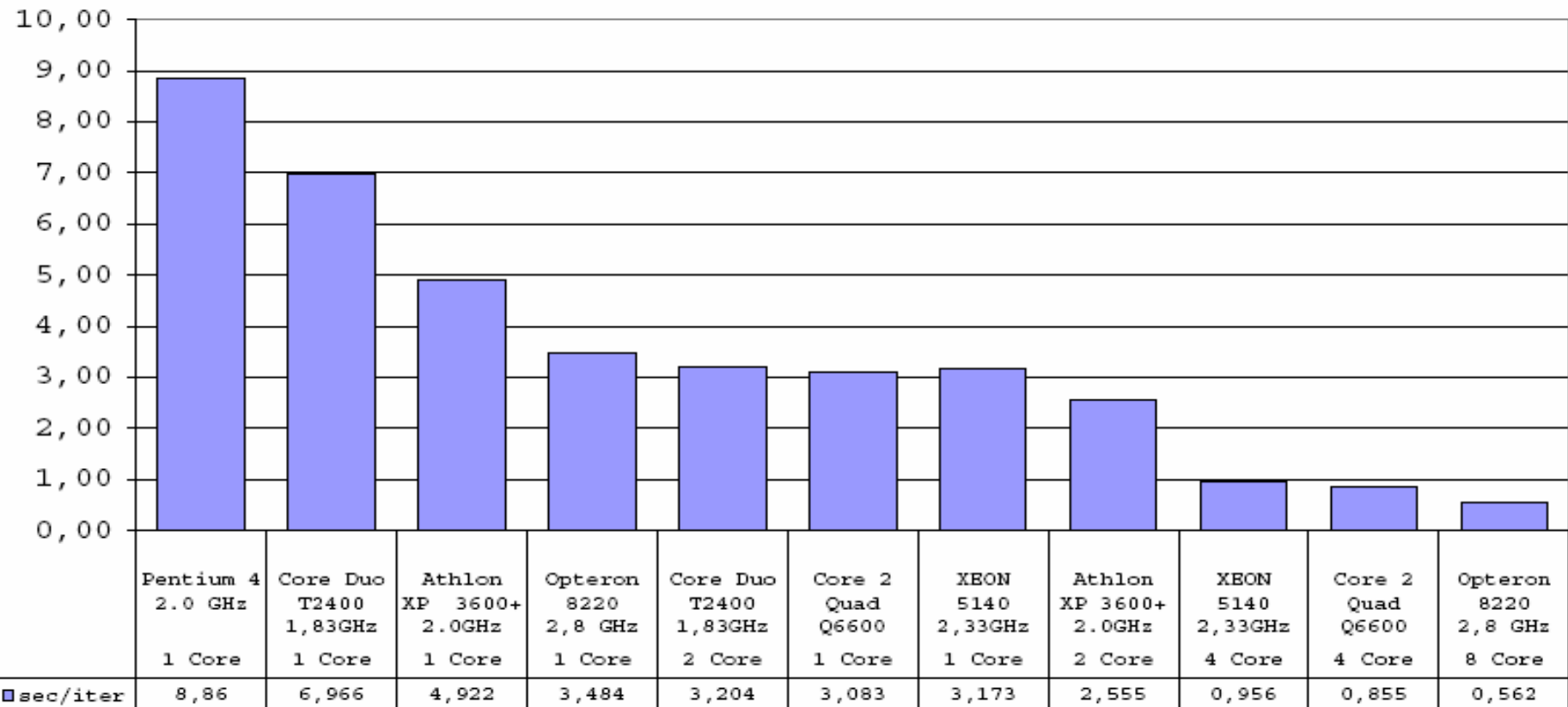
b. Merkmalsvektoren

c. Quantitative Ergebnisse

d. Performanz

Performance Clustering

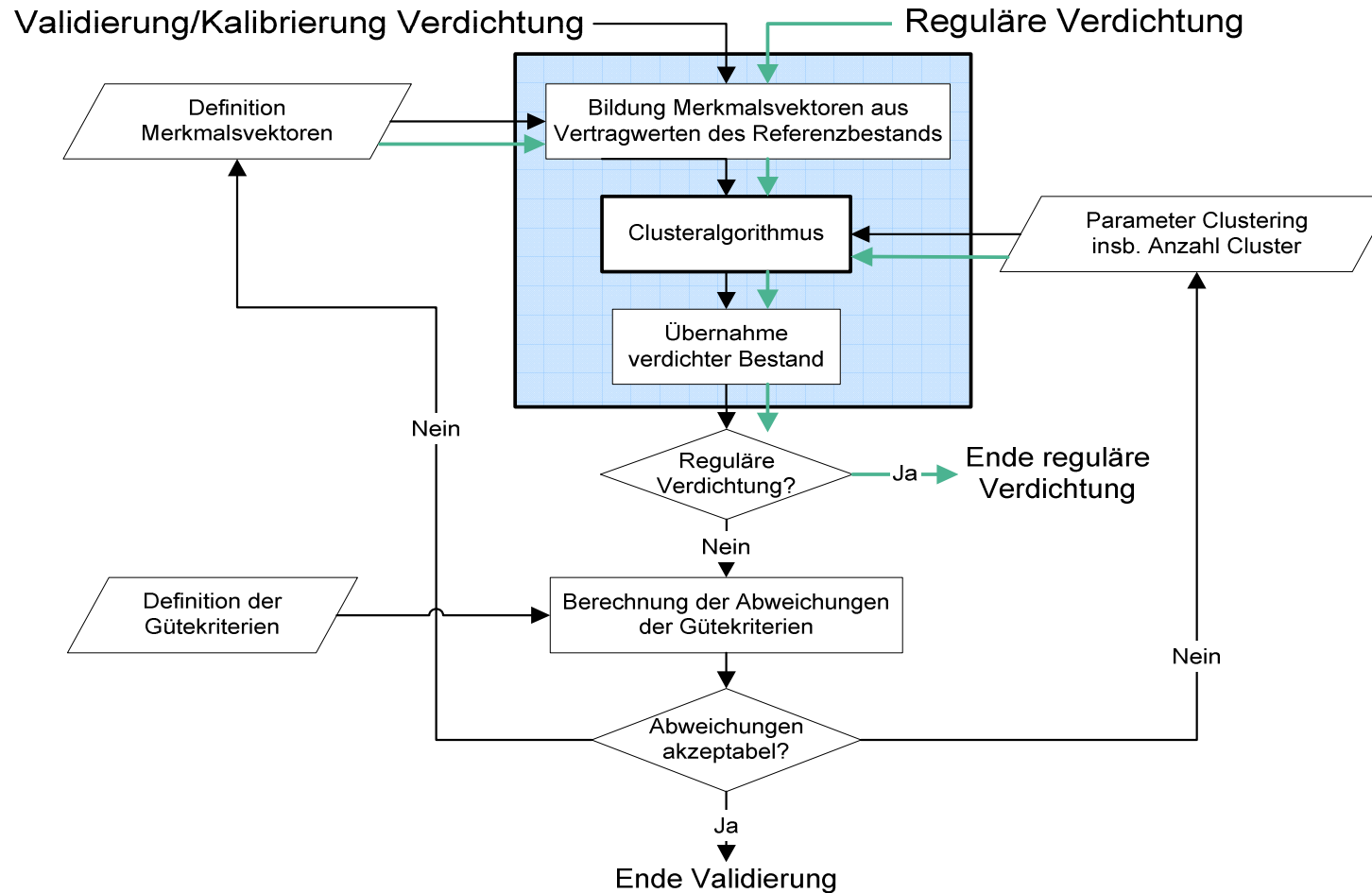
Sekunden pro Iteration (250Tsd Merkmalsvektoren mit Dimension 5). Kleine Balken sind besser.



Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
4. Automatisierung & Validierung der Verdichtung
5. COR.FJA CLUSTRED
6. Systemintegration
7. Zusammenfassung

Prozesse zur Verdichtung mit Clusterverfahren



Validierung der Verdichtung – Zusammenfassung

- Variationsrechnungen mit unterschiedlichen Verdichtungsgraden (Stichprobengrößen)
- Abgleich der Abweichung von Hochrechnungsergebnissen des Originalbestandes mit den Ergebnissen verdichteter Bestände
- Abschluss der Analyse (Abnahme), sobald für einen angemessenen Verdichtungsgrad die Güteanforderungen erfüllt sind
- Dokumentation der Stichprobengröße für die einzelnen Teilbestände (Vorgabe für künftige Verdichtungen)
- Sicherung der Qualität für künftige Versionen durch Abgleich unterschiedlicher Stichproben

Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
4. Automatisierung & Validierung der Verdichtung
5. COR.FJA CLUSTRED
6. Systemintegration
7. Zusammenfassung

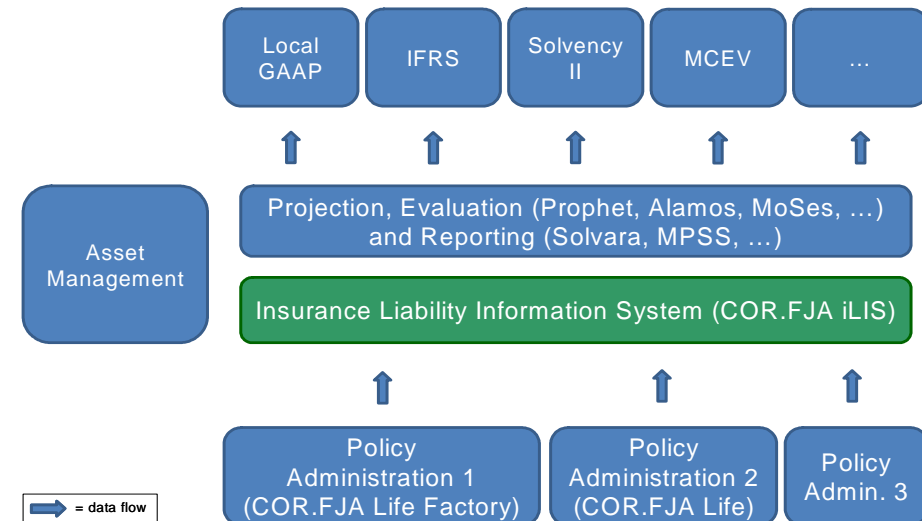
COR.FJA CLUSTRED

- COR.FJA CLUSTRED ist eine Standardsoftwarelösung zur Clusterbasierten Bestandsverdichtung
- COR.FJA CLUSTRED ist standardmäßig integriert in den COR&FJA Produkten
 - COR.FJA Alamos – Asset Liability and model office system
 - COR.FJA iLIS – Insurance Liability Information System
 - COR.FJA CLUSTRED ist darüber hinaus unabhängig von der verwendeten Projektionssoftware einsetzbar !!!

COR.FJA iLIS

■ Vision / Zielbild:

- ◆ COR.FJA iLIS bildet die Basis für eine zentrale Datenhaltung, in der die für die aktuariellen Berechnungen notwendigen Daten – aktuell, in hoher Qualität und revisionssicher – verwaltet werden.
- ◆ COR.FJA iLIS steuert die Prozesse, indem es die benachbarten Systeme anstößt und mit den notwendigen Input-Daten versorgt.

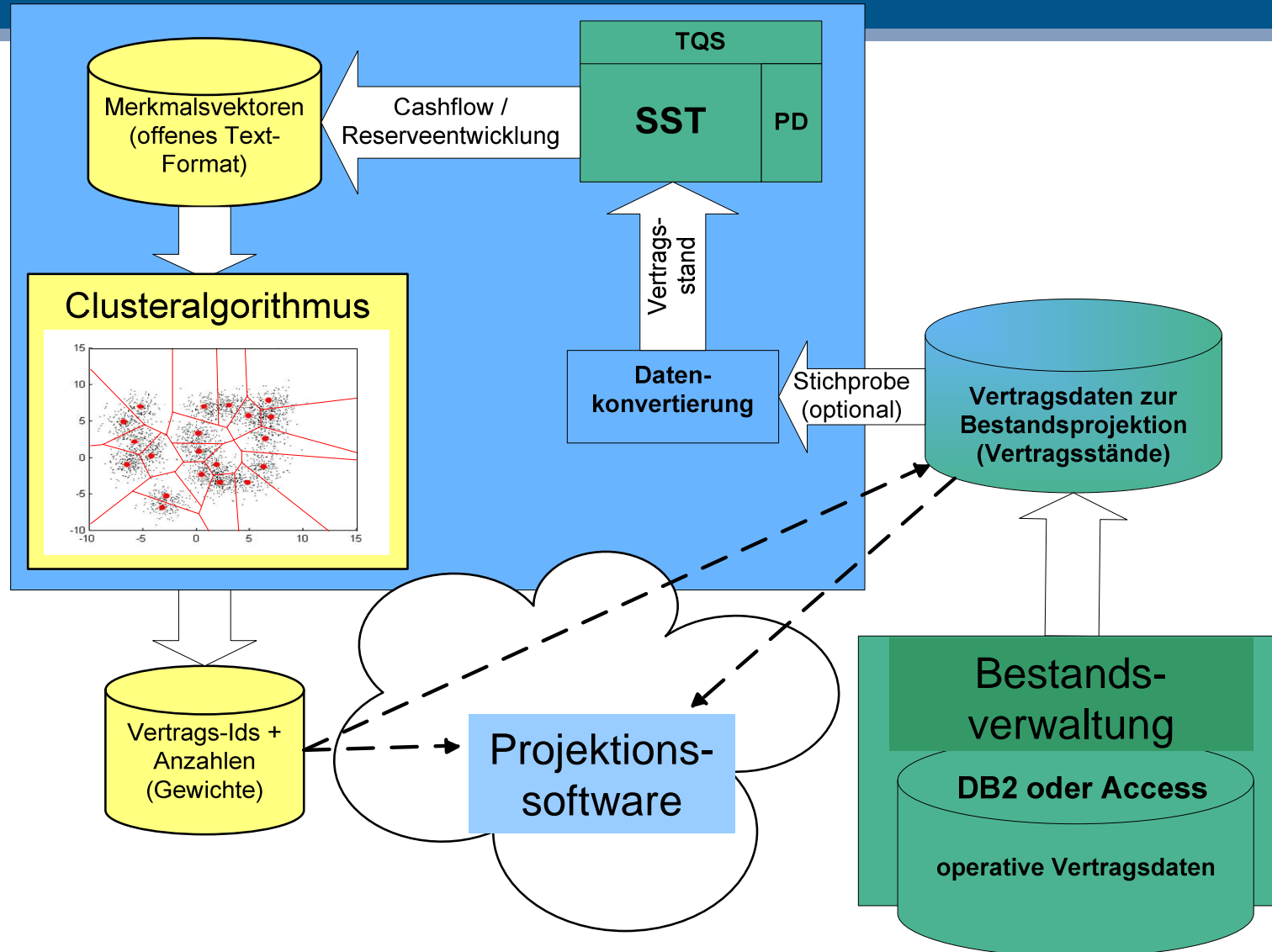


- ➔ Insgesamt gelingt mit COR.FJA iLIS eine Steigerung der Effizienz und der Qualität durch Optimierung der Prozesse im Kontext aktuarieller Berechnungen und Financial Reporting.

Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
4. Automatisierung & Validierung der Verdichtung
5. COR.FJA CLUSTRED
6. Systemintegration
7. Zusammenfassung

Integration Clusterverfahren in Systemlandschaft

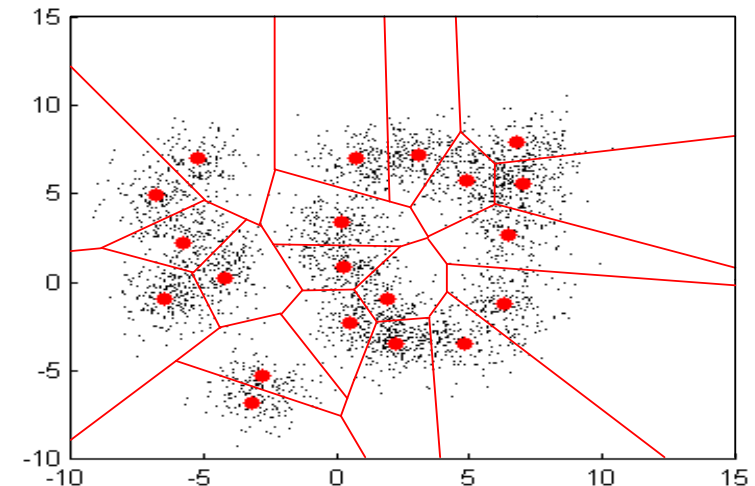


Agenda

1. Die Herausforderung
2. Verfahren zur Bestandsverdichtung
3. Clusterverdichtung
4. Automatisierung & Validierung der Verdichtung
5. COR.FJA CLUSTRED
6. Systemintegration
7. Zusammenfassung

Zusammenfassung Cluster-Verdichtung (1)

- **Problem:** Laufzeit (insbesondere bei stochastischen Projektionen)
- **Lösung:** Bestandsverdichtung
- **Ziel der Verdichtung:** Erzeugung eines möglichst kleinen Teilbestandes, der die gleichen Eigenschaften besitzt wie der Originalbestand
- **COR&FJA-Lösung:** Clusterbasiertes Verfahren
- **Idee:** Gruppierung ähnlicher Daten in Klassen (Cluster, Klumpen)
- **Ziel:**
 - möglichst hohe Homogenität innerhalb eines Clusters
 - möglichst hoher Heterogenität zwischen den Clustern



Zusammenfassung Cluster-Verdichtung (2)

- Höchste Verdichtungsgrade erreichbar (Praxis: 100.000 bis 1 Mio. Verträgen zu 1.000 Modelpoints)
- infolgedessen starke Laufzeitverbesserungen (bis zum Faktor 1.000)
- ➔ z.B. 8 Stunden -> 30 Sekunden
- Struktur des Ausgangsbestandes bleibt erhalten (insbesondere seltene, jedoch wichtige Verträge (z.B. mit hoher VS) sind im verdichteten Bestand angemessen repräsentiert)
- Standard-Lösung, unabhängig von der eingesetzten Projektionssoftware
- Effizienter Einsatz, da nur geringe Vorarbeiten und kein aktuarielles Vorwissen notwendig sind
- Optimierung des Verdichtungsgrades unter vorgegebener Güte durch Variationsrechnungen mit unterschiedlichen Verdichtungsgraden
- Einfache Weiterverarbeitung des verdichteten Bestandes, da Model Points reale Verträge sind, die ohne Einschränkung aktuariell rechen- und bewertbar sind

Kontaktdaten

COR&FJA Deutschland GmbH

A Company of the COR&FJA Group
Domstraße 55 - 73
50668 Köln
Germany

Phone: +49 221 3380-0
Fax: +49 221 9321730
Internet: www.cor.fja.com
E-mail: info@cor.fja.com

